

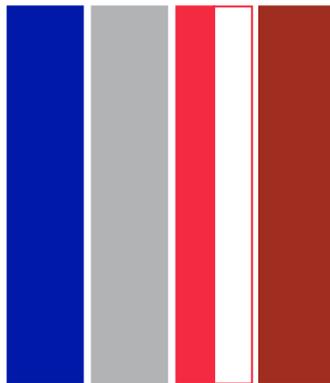
MESTRADO EM CIÊNCIAS DA COMUNICAÇÃO –  
Variante de Estudo dos Media e Jornalismo

# **Jornalismo de Dados: fluxos de trabalho e conteúdos publicados no jornal *Público***

Rafael Pereira Oliveira

# M

2023



Rafael Pereira Oliveira

## **Jornalismo de Dados: fluxos de trabalho e conteúdos publicados no jornal *Público***

Relatório realizado no âmbito do Mestrado em Ciências da Comunicação, orientado pelo Professor Doutor Hélder Bastos

Faculdade de Letras da Universidade do Porto

Setembro de 2023



## Sumário

Declaração de Honra .....	5
Agradecimentos .....	6
Resumo.....	7
Abstract .....	8
Índice de Gráficos .....	9
Índice de Tabelas .....	10
Índice de Figuras.....	11
Lista de abreviaturas e siglas .....	12
Introdução .....	13
Capítulo 1. O Estágio .....	17
Capítulo 2. O Estudo de Caso .....	22
Capítulo 3. O Jornalismo de Dados .....	24
3.1 A relação do Jornalismo e da tecnologia.....	25
3.2 O jornalismo quantitativo.....	27
3.3 O “jornalismo de precisão” e a Reportagem com Auxílio do Computador.....	29
3.4 Evolução e contrastes.....	33
3.5 O dilúvio dos dados .....	35
3.6 “Abertura” e transparência dos dados .....	38
3.7 As definições do jornalismo de dados.....	42
3.8 Os fluxos de trabalho.....	50
3.8.1 Recolha dos dados.....	54
3.8.2 “Limpeza” dos dados .....	55
3.8.3 Análise .....	55
3.8.4 Representação visual dos dados .....	58
3.8.5 Interatividade e formas de narração .....	62
3.9 A ética no jornalismo de dados .....	67
Capítulo 4. Metodologia.....	69
4.1 O estudo etnográfico.....	71
4.2 Análise de conteúdo .....	74
Capítulo 5. Resultados.....	77
5.1 A observação ao jornalismo de dados no <i>Público</i> .....	78
5.1.1 A redação.....	79
5.1.2 Meios disponibilizados .....	81

5.1.3 Atores do jornalismo de dados.....	82
5.1.4 A equipa de jornalismo de dados.....	84
5.1.5 As rotinas de produção.....	91
5.2 Análise de conteúdo.....	100
5.2.1 As potencialidades Web adotadas ao nível da apresentação.....	100
5.2.2 Principais representações visuais de dados e ferramentas de execução.....	103
5.2.3 Transparência, precisão e imparcialidade dos dados.....	106
Conclusão.....	108
Referências Bibliográficas.....	117
Anexos.....	125
Anexo 1 – Diário de campo.....	125
Anexo 2 – Corpo da amostra.....	176
Anexo 3 – Análise de conteúdo.....	178
Anexo 4 - Avaliação qualitativa do orientador de estágio.....	264

## **Declaração de Honra**

Declaro que o presente relatório é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Porto, Setembro de 2023

Rafael Pereira Oliveira

## **Agradecimentos**

O percurso feito ao longo deste trabalho não teria sido possível sem o apoio incondicional dos meus pais. São eles os pilares sobre os quais procuro edificar um futuro cada vez mais risonho. Não poderia deixar de agradecer aos meus amigos e colegas de trabalho que, por várias ocasiões, me motivaram a escrever um dos capítulos mais relevantes da minha vida.

Gostaria de endereçar um agradecimento a toda a redação do jornal Público no Porto pelo acolhimento, amizade e ensinamentos prestados ao longo daqueles três meses. Em especial, enalteço o papel de Rui Barros, enquanto colega, amigo e conselheiro inestimável, que em muito contribuiu para que esta experiência tenha sido possível e superado todas as expectativas. Graças a ele, desvendaram-se novos horizontes profissionais que, ansiosamente, pretendo explorar. Agradecer também ao Ivo Neto por toda a atenção prestada e pelos conselhos preciosos que me ficaram cravados na memória (e também no caderno). Uma palavra de reconhecimento à direção editorial do Público por toda a confiança depositada em mim, nomeadamente a Amílcar Correia, David Pontes e Manuel Carvalho.

Por último, e não menos importante, obrigado ao meu orientador, Professor Hélder Bastos, por toda a paciência, colaboração e disponibilidade prestados ao longo deste tempo.

## Resumo

A tecnologia e a Internet alteraram substancialmente o panorama jornalístico. Num contexto de crescente digitalização, a compreensão, análise e escrutínio das dinâmicas políticas, económicas e sociais dependem cada vez mais de dados. Nos últimos anos, uma disciplina jornalística tem impactado profundamente a maneira como estas informações, frequentemente provenientes de grandes bases de dados, são apresentadas por órgãos de comunicação social: o jornalismo de dados. Este subdomínio jornalístico especializado tem desempenhado um papel crucial na tradução de dados complexos em narrativas acessíveis, permitindo ao público compreender o Mundo à sua volta. Este relatório, que tem como ponto de partida o estágio curricular na redação do Porto do jornal Público, visa investigar as competências profissionais requeridas para operar nesta área, examinar o estado final das peças jornalísticas baseadas em dados e compreender como este órgão de comunicação social português assegura a transparência, precisão e imparcialidade dos dados recolhidos. Assim, adotou-se um estudo etnográfico de observação participante delimitado por um período temporal, que serviu igualmente de orientação para a análise de uma amostragem a fim de averiguar as potencialidades Web empregues, os principais tipos de representação visual de dados e a transparência da informação numérica utilizada. O estudo revelou que a pequena equipa de jornalismo de dados do Público opera num ambiente multidisciplinar, combinando as competências jornalísticas tradicionais com a proficiência nas linguagens de programação, desenvolvimento Web e criação de representações visuais de dados. Além disso, a pesquisa examinou as potencialidades Web, ao nível da execução e apresentação de peças jornalísticas de dados, evidenciando a utilização de ferramentas de livre acesso e a abundância e diversidade de aplicação de recursos interativos. No entanto, questões relacionadas com a transparência foram identificadas como aspetos a ser melhoradas. A dependência de ferramentas de terceiros para a apresentação visual da informação também foi constatada como um desafio.

**Palavras-chave:** Jornalismo de dados; rotinas de produção; análise de conteúdo; estudo etnográfico; jornal Público

## Abstract

Technology and the Internet have significantly transformed the journalistic landscape. In a context of increasing digitalization, the understanding, analysis and scrutiny of political, economic and social dynamics increasingly rely on data. In recent years, a journalistic discipline has profoundly impacted the way this information, often sourced from large databases, is presented by media organizations: data journalism. This specialized journalistic subdomain has played a crucial role in translating complex data into accessible stories, enabling the public to understand better the world around them. This report, which takes as its starting point a curricular internship in the Porto newsroom of the Público newspaper, aims to investigate the professional skills required to operate in this field, examine the final state of data-driven journalistic stories, and understand how this portuguese media outlet ensures data transparency, accuracy, and impartiality. Therefore, an ethnographic study involving participant observation within a defined timeframe was adopted, which also served for the analysis of a sample to assess the employed web capabilities, the primary types of data visualizations and data transparency. The study revealed that the small data journalism team at Público operates in a multidisciplinary environment, combining traditional journalistic competencies with proficiency in programming languages, web development, and the creation of data visualizations. Furthermore, the research examined web capabilities in terms of the execution and presentation of data-driven journalistic pieces, highlighting the use of freely accessible tools and the abundance and diversity of interactive resources. However, issues related to transparency were identified as areas needing improvement. Dependency on third-party tools for data visualization presentations was also recognized as a challenge.

**Keywords:** data journalism; workflow; content analysis; ethnographic study; Público newspaper

## Índice de Gráficos

Gráfico 1 - Contagem dos tipos de representações visuais de dados publicadas .....	103
Gráfico 2 - Comparação entre representações visuais de dados estáticas e interativas .....	104
Gráfico 3 - Total de representações visuais de dados estáticas e interativas por peça	105
Gráfico 4 – Ferramenta utilizada ou autoria das representações visuais de dados .....	105

## Índice de Tabelas

Tabela 1 – Esquema de explicitação dos métodos aplicados .....	70
Tabela 2 – Categorização de análise às potencialidades Web numa peça de jornalismo de dados .....	76
Tabela 3 – Exemplo do processo de análise de conteúdo sobre a transparência, precisão e imparcialidade dos dados utilizados.....	77
Tabela 4 - Quadro de análise à hipertextualidade.....	178
Tabela 5 – Quadro de análise à multimedialidade .....	184
Tabela 6 - Quadro de análise a outras potencialidades Web exploradas.....	234
Tabela 7 – Quadro de análise aos recursos interativos presentes nas representações visuais de dados.....	237
Tabela 8 – Quadro de análise ao tipo de interatividade nas representações visuais de dados .....	242
Tabela 9 – Quadro de análise aos esquemas narrativos - orientações narrativas da história.....	245
Tabela 10 – Quadro de análise aos esquemas narrativos - estética visual da narrativa .....	250
Tabela 11 – Quadro de análise para avaliar a transparência, precisão e imparcialidade dos dados.....	252

## Índice de Figuras

Figura 1 - Esquema do processo de trabalho do jornalismo de dados .....	52
Figura 2 – A pirâmide invertida do jornalismo de dados .....	53
Figura 3 – Fases da recolha e análise de dados do estudo etnográfico .....	73
Figura 4 – Redação do Público no Porto.....	80
Figura 5 – Captura de ecrã da página da equipa de jornalismo de dados no Notion .....	86
Figura 6 – Captura de ecrã do canal de Discord da equipa de jornalismo de dados .....	87
Figura 7 - Dados do cabaz alimentar num repositório do Público na plataforma GitHub .....	128
Figura 8 – Visão geral da interface padrão do RStudio.....	131
Figura 9 – Visualização geral da página Figma do jornal Público .....	136
Figura 10 – Ambiente de trabalho no site do Flourish – na aba “preview” .....	137
Figura 11 – Ambiente de trabalho no Flourish – na aba “data” .....	137
Figura 12 – Exemplo de um gráfico “treemap” construído no Flourish .....	141
Figura 13 – Página inicial da peça “Contas à vida: enfrentar a inflação” .....	148
Figura 14 – Ilustração criada por Rui Barros para a peça da satisfação laboral .....	157
Figura 15 – Capa da edição de 6 de março do jornal Público .....	163
Figura 16 – Trecho de código escrito em linguagem R.....	164
Figura 17 – Exemplo da mensagem personalizada na peça do número de casas vagas .....	170

## **Lista de abreviaturas e siglas**

RAC – Reportagem Assistida por Computador;

API - *Application Programming Interface* (traduzido para português: Interface de Programação de Aplicação);

INE – Instituto Nacional de Estatística;

ADSE – Assistência na Doença aos Servidores Civis do Estado;

RSI – Rendimento Social de Inserção.

## Introdução

Num mundo cada vez mais digitalizado, a ubiquidade dos equipamentos tecnológicos constitui-se como uma prova das profundas alterações que se deram ao nível das relações da sociedade contemporânea nos últimos anos. A chegada da Internet e, posteriormente, da *World Wide Web* potenciou um fluxo incomensurável nos níveis da produção e circulação da informação, através de múltiplos ecrãs, alterando hábitos, comportamentos e paradigmas nas mais diversas áreas.

Estas mutações de natureza tecnológica também se verificaram no Jornalismo, que se deparou com desafios e oportunidades nos primórdios da digitalização das redações e no desenvolvimento da Web. A digitalização da informação ocorreu em vários meios e o advento dos “dados de grande dimensão” trouxe consigo um grande volume, variedade e velocidade na disponibilidade e processamento de dados (Pentzold & Fechner, 2020). Esse contexto fez com que os profissionais do Jornalismo tivessem de se adaptar progressivamente a uma nova forma de aceder e analisar a informação, adotando técnicas, ferramentas e tecnologias que lhes permitissem lidar com a crescente quantidade de dados disponíveis.

Os dados quantitativos, enquanto fonte de informação passível de contribuir para a compreensão, debate, escrutínio e tomada de decisões das conjunturas políticas, económicas e sociais, têm sido usados no Jornalismo desde o final do século XX. Contudo, no início da primeira década deste século assistiu-se a uma transformação na forma como a torrente de informação é recolhida, processada e apresentada pelos média tradicionais nos meios digitais, a fim de “traduzir” grandes quantidades de dados em artefactos jornalísticos inteligíveis para as suas audiências. Exemplos contemporâneos como o do *Wikileaks* ou dos *Panama Papers* evidenciam como essa conjuntura impactou as rotinas e os fluxos de trabalho da produção jornalística e das estruturas organizacionais, exigindo aos profissionais do meio conhecimentos especializados para desvendar as histórias “escondidas”, tendências e padrões em bases de dados de grande dimensão e complexidade.

Mediante este contexto, a informação quantitativa assumiu um papel central na produção e distribuição da informação na era digital (Veglis & Maniou, 2018), originando a popularização de um termo entre profissionais do Jornalismo e académicos: o jornalismo de dados.

Desde 2014 que este tópico tem registado um aumento do número de estudos académicos que se debruçam sobre o tema (Ausserhofer et al., 2020). Embora as suas fundações estejam sobejamente relacionadas com práticas e teorias que remontam à segunda metade do século XX (Coddington, 2015) - como o “jornalismo de precisão” e a Reportagem Assistida por Computador (RAC) -, esta terminologia popularizou-se a partir de 2010. Tal deveu-se ao crescimento exponencial da disponibilidade de dados em formato digital, pela abundância de ferramentas computacionais de análise da informação e pelos formatos inovadores na sua publicação através das plataformas digitais (Treadwell et al., 2016), onde o utilizador pode interagir com a informação através das representações visuais de dados (Parratt-Fernández et al., 2021).

No entanto, não seria expectável que todo e qualquer jornalista compreenda e sumarie o que está por detrás, por exemplo, de três *terabytes* de informação. Ou saiba interpretar e apresentar esses dados aos seus leitores nas plataformas digitais, recorrendo a ferramentas e técnicas especializadas. Neste sentido, novas aptidões e conhecimentos emergiram no seio desta profissão face ao desenvolvimento tecnológico das redações, que passaram a agregar componentes multimédia, de interatividade, narratividade, programação, representação visual de dados e da numeracia e/ou literacia de dados (López-García et al., 2017) nas suas rotinas de produção e publicação de conteúdos.

Por conseguinte, este estudo visa examinar o fluxo de trabalho em jornalismo de dados para compreender as competências profissionais necessárias para operar nesta área. Não obstante, interessa-nos também analisar as potencialidades que a Web proporciona na execução e apresentação de uma peça jornalística deste tipo; as principais representações visuais de dados utilizadas e a forma como são produzidas e, por último, a forma como a transparência, precisão e imparcialidade dos dados é assegurada.

Nesse sentido, realizou-se um estudo de caso, baseado no estágio curricular realizado no âmbito do Mestrado em Ciências da Comunicação da Faculdade de Letras da Universidade do Porto, na redação do Porto do jornal Público. A investigação pretende dar a conhecer e descrever os processos, metodologias e recursos utilizados neste subdomínio jornalístico até à publicação final da história jornalística, de forma a analisar o estado final dessas peças. Para esse efeito, foram formuladas as seguintes quatro questões de investigação e respetivas hipóteses:

1. Quais são as competências necessárias para a prática do Jornalismo de Dados no jornal Público?

H1: As principais aptidões passam por saber consultar bases de dados, recolher essa informação e ter conhecimentos básicos de Estatística para a analisar.

2. Quais as potencialidades que a Web proporciona na execução e apresentação de uma peça jornalística baseada em dados no jornal Público?

H2: Na execução, as principais potencialidades da web neste tipo de peças passam por permitir consultar e trabalhar com ferramentas facilitadoras da recolha e a análise dos dados, enquanto na apresentação incidem sobre a interatividade que a peça proporciona ao leitor, assim como os conteúdos multimédia e/ou representações visuais que a acompanha.

3. Quais são as principais representações visuais de dados a que o Público recorre nas suas peças para apresentar dados e de que forma são produzidas?

H3: Os principais tipos de representação de dados são mapas estáticos e são produzidos a partir de software de terceiros.

4. De que forma a equipa de jornalismo de dados do Público garante a transparência, precisão e imparcialidade dos dados que utiliza?

H4: Refere a metodologia da análise, a fonte e fornece a base de dados utilizada na análise.

No sentido de responder a estas questões, foi adotada uma metodologia que emprega o estudo etnográfico de observação participante, e a análise de conteúdo ao estado final das peças de jornalismo de dados produzidas no período temporal correspondente ao período do estágio: de 2 de janeiro até 31 de março de 2023.

Em suma, o primeiro capítulo deste estudo apresenta uma breve descrição do estágio realizado no jornal Público, dando conta dos processos de produção, dos trabalhos realizados e uma reflexão sobre esta experiência na formação profissional. O segundo capítulo é dedicado ao estudo de caso que será desenvolvido ao longo deste trabalho, sendo a sua escolha justificada a partir da experiência decorrente do estágio.

O terceiro capítulo apresenta a contextualização teórica sobre o jornalismo de dados, explicitando o desenvolvimento histórico desta área, os contrastes com práticas antecessoras, as conjunturas impulsionadoras e características desta área, a ambiguidade da sua definição, as fases do fluxo de trabalho e os desafios éticos que se colocam a esta área.

Já o quarto capítulo explicita a metodologia escolhida para analisar os fluxos de trabalho e as peças de jornalismo de dados publicadas no jornal Público, com a enunciação do período temporal de análise, as questões de investigação, as hipóteses formuladas, os métodos adotados para responder a cada uma das questões e as categorias para a análise da amostra.

O quinto capítulo contempla os dados recolhidos e a discussão dos resultados da investigação. Por último, são apresentadas as conclusões deste trabalho, onde são aprofundados os resultados obtidos e as respostas às perguntas de investigação enunciadas. Esta última secção integra ainda uma visão geral do conhecimento gerado deste estudo, das suas limitações e das possibilidades de investigações futuras.

## Capítulo 1. O Estágio

Inserido no plano curricular do Mestrado em Ciências da Comunicação da Faculdade de Letras da Universidade do Porto, o estágio realizado na redação do Porto do jornal Público, teve a duração de três meses, tendo-se iniciado a 2 de janeiro de 2023 e findado a 31 de março do mesmo ano. A escolha por este órgão de comunicação social para a realização do estágio deveu-se, sobretudo, a três motivos: o interesse em ingressar numa redação que publicasse regularmente trabalhos de referência na área do jornalismo de dados; o reconhecimento e vanguardismo profissionais do Público no panorama jornalístico português e por, pessoalmente, considerar este periódico uma referência da imprensa nacional dada a qualidade e abordagem que empregam na sua produção noticiosa.

Fundado em 1990, o jornal Público desempenhou um papel “acelerador das mudanças tecnológicas” e assumiu um posicionamento inovador no que às novas tecnologias diz respeito: foi o primeiro jornal do Mundo com sistema *Macintosh* (Bastos, 2023, p.55). O seu primeiro diretor foi Vicente Jorge Silva e, desde a sua fundação, o grupo económico *Sonae* é seu detentor. No período de realização de estágio, a direção editorial era composta pelo diretor Manuel Carvalho e pelos diretores-adjuntos Amílcar Correia, Andreia Sanches, David Pontes e Tiago Luz Pedro. Amílcar Correia foi o supervisor da instituição do estágio, tendo a orientação do estágio em jornalismo de dados ficado ao encargo do jornalista Rui Barros. O horário ficou definido para decorrer durante os dias úteis, das 10h00 até às 18h30, com a ressalva de que o mesmo poderia ser ajustado mediante as necessidades do trabalho. Em alguns casos, este horário estendeu-se até mais tarde, tendo essas situações sido compensadas - em termos de horas - no dia seguinte.

O primeiro dia na redação do Porto do jornal *Público* ficou marcado pela apresentação do espaço da redação e dos elementos da direção editorial, nomeadamente o então diretor Manuel Carvalho; diretores-adjuntos Amílcar Correia e David Pontes, o editor da secção *Online*, Ivo Neto, e do jornalista encarregado pela orientação do estágio, Rui Barros. Com este último, foram trocadas algumas impressões iniciais e discutidas propostas de trabalho a desenvolver.

A primeira tarefa pela qual ficamos incumbido consistiu na atualização semanal, até ao final do estágio, dos dados da peça “Conheça o impacto da inflação no seu carrinho de compras<sup>1</sup>”, recorrendo à plataforma *GitHub*, à qual fomos introduzidos no primeiro dia. Durante o período de estágio, foram publicadas quatro peças noticiosas: duas peças de jornalismo de dados, em que uma delas foi realizada colaborativamente com o jornalista Rui Barros, e duas notícias breves - uma na secção “Ípsilon” do jornal, e outra sobre a nomeação do jornalista Rui Barros para os *Sigma Awards 2023* - a mais importante distinção internacional em jornalismo de dados.

Relativamente às peças publicadas de jornalismo de dados, estas contemplam, além do conteúdo textual, a produção própria de um total de 12 gráficos que apresentam visualmente a informação quantitativa utilizada. Além disso, foram realizadas cinco análises de dados, através do *Excel* e do *RStudio*, que foram partilhadas no repositório privado do *GitHub* do Público. Este trabalho consistiu na análise a temas propostos e/ou sugeridos pelo orientador de estágio, entre as quais: a relação dos rendimentos da população portuguesa face ao preço das rendas das casas por município; o fluxo de deslocação dos alunos que ingressaram pela primeira vez no Ensino Superior face ao concelho em que viviam; a evolução do número de famílias reconstituídas em Portugal; o aumento registado, entre fevereiro e março, no preçário dos serviços na Assistência na Doença aos Servidores Civis do Estado (ADSE) e, por último, a análise demográfica às pessoas que assinalaram o Rendimento Social de Inserção (RSI) como principal fonte de rendimento nos Censos 2021. Foi ainda efetuado o trabalho conjunto de limpeza e formatação de uma base de dados sobre os imóveis devolutos do Estado português com o orientador do estágio.

Para cada uma das peças de jornalismo de dados publicadas e análises conduzidas foi realizada a recolha, limpeza, processamento, análise e a representação visual de dados, sendo que o esclarecimento de qualquer dúvida ou ajuste necessário foi devidamente acompanhado pelo orientador do estágio e o editor da secção *Online*.

Apesar da quantidade relativamente reduzida de conteúdos publicados durante este período, isso foi discutido nas primeiras semanas do estágio.

---

<sup>1</sup> Disponível em: <https://www.publico.pt/interactivo/inflacao-portugal> [consultado pela última vez a 3 de julho de 2023]

Por iniciativa própria, partilhamos essa preocupação com o orientador de estágio que, gentilmente, explicou que o ciclo de produção em jornalismo de dados é "diferente" e "mais moroso" em comparação com a maioria das tarefas realizadas pelos restantes colegas da redação, o que vai ao encontro de evidências semelhantes na literatura (Boyles & Meyer (2017: p.435) . O trabalho desenvolvido antes da publicação de uma peça poderia demorar, pelo menos, uma semana até as principais conclusões da análise aos dados serem alcançadas. As duas peças de jornalismo de dados desenvolvidas e publicadas nas plataformas digitais – sendo que uma delas foi publicada na versão impressa e com chamada na capa -, comportaram declarações de uma ampla diversidade de fontes, no sentido de assegurar um trabalho jornalístico rigoroso e aprofundado. A isto acresce ainda a produção própria das representações visuais de dados e, em um dos casos, a publicação do artigo num formato diferenciador recorrendo ao desenvolvimento Web, através do software *Visual Studio Code*, com o auxílio do orientador de estágio. A descrição destes processos serve para evidenciar que a produção de uma peça jornalística deste género pode ser demorada, nomeadamente para alguém que é iniciante nesta área, devido ao conhecimento e de funções relativamente especializadas.

A equipa de jornalismo de dados do jornal *Público* era constituída, essencialmente, por dois jornalistas da redação do Porto - Rui Barros e José Volta e Pinto -, sendo que este último produzia frequentemente outros conteúdos para a secção *Online* do jornal. Além da produção referida, durante estes três meses foram também desenvolvidos outros trabalhos de jornalismo de dados por esta equipa em colaboração com colegas da redação do Porto e de Lisboa. As peças produzidas pela equipa de jornalismo de dados abordam uma diversidade de temáticas, que vão desde a Economia, Habitação ou Educação, e a maioria dos conteúdos comporta uma vertente multimédia e/ou interativa diferente da maioria dos publicados na página Web do *Público*. Esses casos eram observáveis, sobretudo na forma de ilustrar a informação quantitativa ou a estrutura visual da narrativa da história.

A minha familiaridade, embora reduzida, com esta área do Jornalismo foi uma das principais motivações para ingressar neste estágio, que se revelou uma experiência bastante positiva e proveitosa.

Tive a oportunidade de aprender e aprofundar conhecimentos e metodologias de trabalho que me eram desconhecidas, bem como abordar temas pelos quais passei a nutrir especial interesse. Embora seja jornalista há algum tempo na imprensa regional, a experiência no jornal Público revelou-se uma aprendizagem muito enriquecedora pela exigência, rigor e qualidade que norteiam a produção *online* e impressa deste jornal. Os trabalhos desenvolvidos em jornalismo de dados possibilitaram entrevistar diversos agentes da sociedade portuguesa, tais como Eduardo Oliveira e Sousa, então presidente da Confederação dos Agricultores de Portugal, economistas, sociólogos, entre outros testemunhos. Além disso, importa realçar a oportunidade que tive em aprender e aprofundar a utilização de ferramentas de recolha, limpeza, análise e representação visual de dados, assim como publicar peças nas plataformas digitais, através de *softwares* e ferramentas especializadas como o *RStudio*, *GitHub*, *Flourish* e *Visual Studio Code*. Com a exceção da linguagem de programação R, que se emprega no *RStudio*, todas estas ferramentas eram-me desconhecidas, mas a introdução às mesmas, por parte do jornalista Rui Barros, permitiu que a adaptação decorresse de forma progressiva e natural. Este aspeto foi determinante para que as peças publicadas neste âmbito comportassem uma vertente multimédia e de interatividade para ser apresentada aos leitores. Aliás, conforme referido, estas componentes são transversais e intrínsecas à maioria das peças de jornalismo de dados do Público, que também inclui hiperligações numa vasta maioria dos seus textos, seja para redirecionar o leitor para uma outra notícia/reportagem do arquivo do jornal ou para outros sítios Web (por exemplo, a fonte da base de dados utilizada).

Um outro aspeto a destacar tem que ver com o rigor na escrita. O Público possui um livro de estilo que em muito contribui para a elaboração cuidada do texto, primando por uma linguagem precisa, clara, acessível e informativa. Esse guia detalhado estabelece diretrizes para a redação, gramática, pontuação e uso adequado da língua portuguesa e do acordo ortográfico. Também orienta a ética jornalística no sentido de alcançar a imparcialidade nas suas reportagens, instigando os jornalistas a investigar a fundo e a ouvir todas as partes envolvidas na apresentação dos factos. Inicialmente, foi difícil empregar a capacidade de sintetizar e organizar a informação nas peças publicadas.

Contudo, as correções e sugestões construtivas do editor da secção *Online*, Ivo Neto, e do orientador de estágio, Rui Barros, permitiram aprimorar esse aspeto. Em suma, o sentimento de integração e pertença nas dinâmicas de trabalho da redação sempre esteve presente.

Considerando a forma como este órgão de comunicação social propicia a colaboração dos seus estagiários, esta foi uma excelente oportunidade para aprofundar, desenvolver e aprender novas competências profissionais em áreas que me eram relativamente familiares - como o Jornalismo e a linguagem de programação R - e a outras que não me sentia tão confortável, como a Estatística e a produção das representações visuais de dados. Estes dois últimos aspetos, apesar de desafiantes, foram dos mais frutuozos, pois foi possível aprender e aplicar metodologias e técnicas que facilitaram o fluxo de trabalho nesta área. No que concerne às representações visuais de dados, fiquei também munido de um maior traquejo e sensibilidade dada a exigência e cuidados necessários na sua apresentação de forma a não induzir o leitor em erro - fosse através de gráficos de barras, mapas ou até na escolha da paleta de cores que o Público adota nas suas representações gráficas e/ou infografias.

Importa ainda frisar que o orientador do estágio, Rui Barros - o principal jornalista de dados do *Público* -, foi a pessoa com quem lidei mais regularmente, tendo-se revelado um colega de trabalho excepcional que muito contribuiu para que esta experiência tenha decorrido de forma tão positiva. Prova disso foram as várias conversas, trocas de ideias e opiniões, reparos e sugestões, recomendações de leitura e de tutoriais *online*, e o companheirismo prestado ao longo destes três meses. Ainda que pudesse estar atarefado no seu quotidiano profissional, mostrou-se sempre disponível para esclarecer dúvidas, ensinar algumas das técnicas complexas desta área e guiar no processo jornalístico, por exemplo, com a sugestão do melhor ângulo da história conduzida por dados ou das fontes mais apropriadas a contactar. Nesse sentido, importa também frisar que as sugestões do supervisor da instituição do estágio, Amílcar Correia, e do editor da secção *Online* - Ivo Neto - foram determinantes para aperfeiçoar a escrita jornalística.

Ambos serviram como importantes pontes de ligação para conhecer uma diversidade de colegas jornalistas e profissionais da equipa de Infografia das redações do Porto e Lisboa, bem como pelo fornecimento de contactos de testemunhos nas notícias, o que contribuiu para fortalecer a minha lista de contactos, o que, nesta profissão, é bastante relevante. Esta conduta exemplar das relações profissionais na redação foi fulcral para o bom ambiente de trabalho e para o enriquecimento neste processo de aprendizagem que, de antemão, previa que fosse significativo.

Por último, este estágio foi bastante relevante para a escolha do estudo de caso desenvolvido neste projeto, pois permitiu analisar os fluxos de trabalho e as histórias desenvolvidas e publicadas em torno do jornalismo de dados num órgão de comunicação social português.

## **Capítulo 2. O Estudo de Caso**

Conforme mencionado no capítulo anterior, uma parte substancial do trabalho desenvolvido no estágio curricular incidiu na realização de duas peças de jornalismo de dados e cinco análises de dados. No sentido de contribuir para uma compreensão mais ampla da produção dos conteúdos de jornalismo de dados do jornal Público, foram tidos em conta outros seis trabalhos desenvolvidos pela equipa de jornalismo de dados da redação do Porto. Esta escolha é justificada pela pertinência de alargar a representatividade total da nossa amostra, sem comprometer o período temporal definido da análise.

O estudo de caso é uma estratégia de investigação centrada na compreensão das dinâmicas presentes em contextos específicos que permite a testabilidade e a validade empírica, que surgem da ligação íntima com as provas recolhidas (Eisenhardt, 1989), pelo que se adequa aos objetivos desta investigação: compreender os fluxos de trabalho em jornalismo de dados e analisar os conteúdos das peças produzidas e publicadas nesta área. Através de recursos interativos e multimédia, o Público procura tirar o melhor partido das potencialidades da Internet na produção e publicação dos seus artigos.

Tendo em conta esse contexto, os fluxos de trabalho em jornalismo de dados e a apresentação dos conteúdos são duas vertentes que foram aprofundadas durante o percurso deste estágio, constituindo-se como fatores determinantes para a escolha do estudo de caso.

Neste sentido, numa primeira fase, serão analisados os fluxos de produção na redação do Porto, empregando o estudo etnográfico de observação participante. O objetivo é entender as competências necessárias para trabalhar nesta área, explorar as potencialidades Web na criação de reportagens jornalísticas deste tipo, examinar o processo de criação das representações visuais de dados e dos valores éticos que norteiam a produção de conteúdos no núcleo de jornalismo de dados do Público.

Numa segunda fase, proceder-se-á a uma análise de conteúdo ao estado final das peças de jornalismo de dados publicadas com o propósito de identificar e analisar o tipo de potencialidades Web presentes na publicação final das peças, os principais tipos de representações visuais de dados apresentados aos leitores, e a forma como esta equipa procura assegurar a transparência, precisão e imparcialidade dos dados à sua audiência.

### Capítulo 3. O Jornalismo de Dados

O jornalismo de dados possui uma definição ambígua, mas com historial. Algumas das características mais evidenciadas na literatura identificam a sua relação com a informação quantitativa; com a dimensão tecnológica, ao nível da execução e da apresentação da peça jornalística; a transparência dos dados e o tamanho das bases de dados utilizadas para analisar informação. Esta aplicação de dados quantitativos e a utilização de ferramentas tecnológicas no contexto jornalístico não constitui uma novidade. De facto, existem práticas que precedem o jornalismo de dados e que surgiram no contexto da modernização tecnológica das redações, utilizando a informação quantitativa enquanto prática promotora de um jornalismo “mais objetivo” e/ou “próximo do rigor científico”.

O que constitui a diferença entre as práticas deve-se, sobretudo, à evolução tecnológica, social e económica, a que ainda hoje assistimos. A chegada do fenómeno dos “dados de grande dimensão”, da “cultura de transparência”, da abertura das organizações públicas e governamentais, bem como a “democratização” da Internet, são, para vários autores, aspetos determinantes dessa evolução terminológica – primeiro o “jornalismo de precisão”, depois a Reportagem Assistida por Computador (RAC) e, por último, o jornalismo de dados.

O progresso nesta prática jornalística introduziu novas ferramentas, métodos e atores na sua produção, provocando alterações no fluxo do trabalho jornalístico, nas estruturas organizacionais das redações e até ao nível da ética profissional. Um aspeto que se evidencia como consensual é de que o jornalismo de dados é considerado uma prática especializada e multidisciplinar, que requer conhecimentos nas áreas da Estatística, da programação e até ao nível da apresentação gráfica dos dados.

Ao longo das últimas duas décadas, a maioria dos estudos de caso sobre este tema realizaram-se nos Estados Unidos da América e no Reino Unido. A realização de entrevistas semiestruturadas a jornalistas de dados é a metodologia mais adotada, seguindo-se a análise de conteúdo ao texto ou a imagens, embora poucas incidam nas histórias conduzidas por dados (Ausserhofer et al., 2020).

Em Portugal, foi possível identificar cinco investigações relacionadas com este tópico: uma dissertação de mestrado que caracteriza o estado da arte do jornalismo de dados em Portugal (Martinho, 2013); um estudo que procura definir e examinar o que é o jornalismo de dados e as práticas associadas (Martinho, 2014); dois estudos exploratórios sobre as práticas profissionais nesta área (Gehlen & Sousa, 2018; Gehlen, 2019) e uma investigação sobre a evolução histórica de reportagens em jornalismo de dados (Alexandre, 2020). Feita esta nota introdutória, iremos explanar nos capítulos seguintes a revisão literária sobre os principais aspetos do jornalismo de dados.

### **3.1 A relação do Jornalismo e da tecnologia**

Os primórdios da imprensa remontam à época em que começou a ser relatado o "movimento generalizado" do aumento das trocas comerciais na Europa durante os séculos XV e XVI (Rosa, 2016: p.25). Desde a invenção da máquina de impressão tipográfica de Gutenberg até aos primórdios do século XIX, a tecnologia da impressão "não sofreu qualquer mutação de fundo". Foi já durante a primeira metade do século XIX que se deu a revolução industrial da imprensa, tratando-se, essencialmente, de uma "revolução tecnológica" impulsionada pelo conjunto de desenvolvimentos económicos, legislativos e sociais (*Ibidem*, p.51). Os progressos que ocorreram na imprensa, assim como na rádio, enquanto meios de comunicação de massas, são apontados por Rosa (2016) como fatores de uma "transformação profunda da sociedade" e da noção de "espaço público". Compreende-se, por isso, que a tecnologia foi um fator preponderante na organização social e exploração económica dos meios de comunicação, tendo-se estabelecido uma relação histórica e intrínseca no desenvolvimento dos meios e ferramentas em que o Jornalismo hoje nos chega.

O que mais tarde viria a constituir, novamente, uma revolução no modelo de comunicação de massas foi o surgimento das novas tecnologias dos média, como são exemplos o computador e a Internet. McQuail (2010) diz que "a chave para o imenso poder do computador", enquanto "máquina de comunicação" reside no processo de digitalização que permitiram que "informações de todos os tipos e formatos sejam transportadas com a mesma eficiência".

No mesmo sentido, Rosa (2016, p.115) dá conta de que o computador se tornou a “tecnologia de base” dos novos meios de comunicação e, a partir dos anos 50, esta “nova máquina” passou a ser usada no “processamento de dados” por diversas entidades, como os institutos de estatística ou as instituições bancárias. Contudo, o aparecimento da Internet provocou a “revolução conceptual na estrutura de funcionamento dos meios de comunicação social” (*Ibidem*, p.149-156).

O sistema tecnológico do computador foi elementar para o desenvolvimento desta “comunicação assente em redes abertas” e Norbert Wiener (1894-1964) foi quem idealizou o conceito destas “novas máquinas de processamento de informação” enquanto poderosos instrumentos que poderiam realizar um “processo comunicativo global”. Ainda assim, Joseph Licklider (1915-1990), entre a década de 50 e 60, foi quem constituiu a “ideia revolucionária” da “comunicação entre os homens através de computadores ligados em rede” (*Ibidem*). Ou seja, concretizou-se a troca comunicativa direta entre “participantes ativos”, ao invés da perceção mais conservadora de Wiener, que entendia os computadores como “processadores simbólicos de informação”. Já em 1969, a *Arpanet* foi a primeira rede a ser implementada (Rosa, 2016, p.162-164) e, à medida que outras redes foram surgindo, houve a necessidade de se criar uma interligação entre elas. Surge, assim, a invenção do protocolo TCP/IP que possibilitou essa conexão e se constituiu como “denominador mínimo comum de conexão”, no qual assenta a Internet. Mas a sua popularidade só viria a atingir o apogeu com a criação da *World Wide Web*, no início dos anos 90.

Nessa época, Bastos (2023, p.19) dá conta de um crescimento “sem precedentes” das comunicações online e as empresas do ramo do jornalismo viram a Internet como uma “nova oportunidade” para veicularem os seus conteúdos nesta “rede mundial de computadores”. A exponenciação deveu-se, sobretudo, à criação dessa rede virtual e universal, de “tipo cliente/servidor assente na rede física Internet”, fundada por Tim Berners Lee (Rosa, 2016, p. 172). Foi, assim, que a introdução dos primeiros *browsers* comerciais, como o *Netscape*, em 1994, e o *Internet Explorer*, em 1995, originou a “verdadeira explosão” do ciberjornalismo (Bastos, 2023, p.27).

### 3.2 O jornalismo quantitativo

Decorrente destes desenvolvimentos históricos e da evolução nos vários domínios do Jornalismo, nomeadamente a sua profissionalização, especialização e digitalização, compreende-se que a relação da tecnologia com o jornalismo de dados não poderia ser diferente. Aliás, a utilização da informação numérica e/ou quantitativa para reportar acontecimentos já é feita há centenas de anos e precede até os jornais impressos. Rogers, Schwabish & Bowers (2017: p.5) dão conta de que, por exemplo, na viragem do século XVI para o XVII existia já uma contagem semanal do número de mortes e nascimentos na cidade londrina vendida a troco de alguns cêntimos. E, mais recentemente, em 1946, Anderson (2015: p.349-350) relembra que alguns gráficos informativos eram capazes de ilustrar certas conjunturas, como o cenário de crise da imprensa norte-americana, ao mostrar a carência de jornais diários em circulação comparativamente à população norte-americana entre os anos de 1909 e 1940. Tratam-se, pois, de episódios que evidenciam que a utilização da informação quantitativa no Jornalismo não é uma novidade.

Em 2009, Rogers (2011) fundou um blogue do jornal *The Guardian* dedicado ao jornalismo de dados. Num dos seus artigos corrobora essa ideia, dando como exemplo a primeira edição deste órgão de comunicação social, publicada a 5 de maio de 1821, na qual uma reportagem sobre as escolas de Manchester - com informação sobre o número de alunos e custos de cada estabelecimento escolar apresentados numa tabela - foi a manchete do jornal. Para este e vários outros autores (Gray, Bounegru & Chambers, 2012; Knight, 2015; Tabary et al., 2016; Hammond, 2017; Appelgren et al., 2019), esta peça noticiosa constitui-se como o primeiro trabalho de jornalismo de dados, em que a principal fonte foi a informação quantitativa.

Já no caso português, Alexandre (2020: s/p) refere que um dos primeiros exemplos de “reportagens que utilizam dados como fonte primária” é datado de 1868, no extinto *Diário Popular*, no qual o foco da peça jornalística incidu no cálculo do custo final de produção de novas habitações: “O jornal estampa na capa a reprodução de um possível modelo de casa popular para os trabalhadores e a lista dos custos associados à construção das habitações – terrenos, materiais e mão de obra.

A matéria calcula qual seria o custo final das casas, argumenta que as rendas não sairiam por menos do que 45 mil réis e desafia a comissão a mostrar como pretende realizar esse projeto”. Além da apresentação dos números para o leitor, após a “recolha e o tratamento dos dados”, o autor constata que houve um “trabalho de contextualização da informação” e um “questionamento ao poder público”.

O percurso histórico que Anderson (2017: p.98) traça em torno da utilização da informação numérica, situando-a nos primórdios do século XX, explica que o “entusiasmo” sobre os dados deriva da sua associação ao “poderoso valor simbólico” que representam. Nessa época, explica, os dados eram percebidos pela sociedade como um “objeto simbólico e material”, onde as “provas quantitativas” eram capazes de realizar uma “persuasão retórica”. No entanto, as organizações noticiosas desvalorizaram a sua utilização nas reportagens, uma vez que poucos jornais tinham “recursos tecnológicos” para “produzir ou reproduzir estatísticas” e as rotinas de produção jornalística que vigoravam não permitiam a expansão dessa prática. O mesmo esclarece que o Jornalismo do início do século XX estava mais direcionado para “indivíduos, entrevistas, instituições ou cobertura de eventos”, o que levou a que os jornais “não enfatizassem a utilização de ‘dados enquanto contexto’ nas suas notícias”, pois o seu conteúdo deveria servir apenas para “informar” e “não para explicar”. Por isso, considera que a “necessidade pela objetividade”, tal como a Ciência e o Jornalismo a percebem hoje, só surgiu devido a uma “crise da fé empírica” que emerge na sequência da Primeira Guerra Mundial.

Tabary et al. (2016: p.68) explicam que a informação quantitativa do século XIX era “rara e difícil” de obter e que o uso mais recorrente das “provas numéricas” só se estabeleceu durante o século XX também pela necessidade de “progredir através de várias etapas”, nomeadamente “a conceptualização e formulação de hipóteses, a recolha de dados, o teste e o estabelecimento da própria prova através da crítica (ou seja, a colocação dos resultados do teste dentro do quadro teórico ou social dentro do qual as hipóteses foram formuladas)”.

À medida que, progressivamente, a informação se transformou (e transforma) em “uns e zeros”, potenciado pela evolução tecnológica e a digitalização, o Jornalismo começou a recolher, a analisar, a processar e a representar visualmente essa informação, constituindo aquilo que Coddington (2015, p.332) denomina como a “viragem quantitativa” do Jornalismo na forma de reportar acontecimentos.

### **3.3 O “jornalismo de precisão” e a Reportagem com Auxílio do Computador**

O ponto mais frequentemente evocado no percurso histórico da relação do Jornalismo com a informação quantitativa coincide com o “jornalismo de precisão” e da Reportagem com Auxílio do Computador (tradução do inglês “Computer-Assisted Reporting”). São práticas que emergiram nos Estados Unidos da América, na transição entre o final da década de 60 e o início dos anos 70 (Hammond, 2017).

Uma das personalidades mais referenciadas sobre a mudança paradigmática desta “viragem quantitativa” do Jornalismo é atribuída a Philip Meyer: foi um dos primeiros jornalistas a propor a aplicação da metodologia quantitativa das ciências sociais no contexto jornalístico com o intuito de aproximar a profissão às noções de “objetividade e precisão das informações” (Parasie & Dagiral, 2013). No final da década de 60, Meyer terá sido um dos primeiros jornalistas americanos a utilizar o computador como ferramenta para realizar jornalismo de investigação, tendo cunhado o termo “jornalismo de precisão” que marcou a “fusão mais explícita” entre ciências sociais, a documentação de dados e a prática jornalística (Anderson, 2015: p.357).

A fase inicial desta prática incidiu sobretudo na aplicação das metodologias das ciências sociais no Jornalismo (Meyer, 2002), ainda que o termo tenha evoluído posteriormente para Reportagem com Auxílio do Computador (RAC) e tenha dado aso a outras práticas jornalísticas intimamente ligadas à tecnologia e ao pensamento computacional. Meyer (2002: p.79) nunca utilizou explicitamente esse termo e explica que a gíria foi adotada pelos jornalistas da imprensa para se referirem à utilização dos computadores para realizarem “análise e investigação” jornalísticas.

O “seu” jornalismo de precisão trata-se, portanto, da adoção de “métodos e ideais científicos” no processo da comunicação de massas através de “inquéritos, análises de conteúdo, observações, análises estatísticas a partir da recolha e interpretação de dados” ou do teste de hipóteses. No livro *Precision Journalism* (2002), o autor contextualiza que, em tempos, para ser jornalista era necessário “dedicar-se à verdade, ter muita energia e algum talento para a escrita”. No entanto, face à crescente informação que estava disponível, ao contexto profissional cada vez mais exigente e a uma sociedade “progressivamente complexa”, essas aptidões tornaram-se parcas. Por isso, argumenta, o jornalista deveria agora ser capaz de se tornar “um gestor de base de dados, um processador de dados e um analista de dados”. Por entender que o Jornalismo requer “disciplina”, o autor entendeu que a melhor forma de cumprir esse desígnio deveria ser através da aproximação aos métodos científicos, utilizando “ferramentas de recolha e análise de dados” e a “pesquisa disciplinada” por “verdades verificáveis”. No fundo, a sua proposta assemelha-se àquilo que afirma ser um “jornalismo científico”. O jornalismo de precisão tornou-se a prática capaz de “expandir o conjunto de ferramentas” do jornalista para trazer temas que “anteriormente inacessíveis” ao escrutínio jornalístico (Meyer, 2002: p. 235). Por consequência, o computador tornou-se um utensílio fundamental para atingir esse objetivo. Hammond (2017) considera que o livro original de Meyer, datado de 1973, constituiu o início da análise de dados enquanto “parte significativa do reportório jornalístico” e que a sua proposta serviu de “inspiração” para a RAC. Também outros fatores, como a crescente profissionalização no Jornalismo, ao longo da década de 60, e o “aumento do acesso público a registos governamentais” contribuíram para a “nova forma de entender o documento como prova” (Anderson, 2015: p.358).

A crescente disponibilidade dos dados (Anderson, 2015) e a integração dos computadores nas redações (Bastos, 2023), assim como os avanços tecnológicos na capacidade de processamento destas máquinas, possibilitaram aos jornalistas tratar a informação quantitativa de uma forma que antes não era possível (Tabary et al., 2016: p.71).

No caso português, a implementação das tecnologias digitais nas redações só começou na década de 80 (Bastos, 2023: p.52-54), tendo o jornal “O Comércio do Porto” sido o primeiro a “informatizar a sua redação”. Mas, logo no final dessa década, já “todas as redações dos grandes jornais estavam informatizadas”.

Por outro lado, Rogers (2011) considera que o que constituiu a “grande diferença” entre o jornalismo de dados do século XIX, no *The Guardian*, e o do século XXI, incide na forma de como os dados são disponibilizados. Se, inicialmente, eram publicados em “livros bastante caros”, em 2009 passaram a ser disponibilizados em “folhas de cálculo e ficheiros formatados para computadores” (*Ibidem*) capazes de “processar e armazenar informação” (Bastos, 2023: p.26). Ora, Hammond (2017) entende que esta mutação é “basicamente técnica”, pois os tradicionais formatos analógicos de apresentação e compilação de dados foram substituídos pelas novas tecnologias digitais - o que evidencia a relação e evolução histórica entre o jornalismo de dados e a tecnologia.

Mediante este contexto, importa notar que a primeira vez que o computador foi utilizado pelos jornalistas no “processo de reportagem”, terá sido na cobertura das eleições nacionais norte-americanas de 1952 para prever os resultados desse ato eleitoral (Léchenet, 2014: p.5). Foi o canal televisivo *Columbia Broadcasting System* (CBS) a conduzir essa previsão (Felle, 2016: p.87). Após esse episódio, seguiu-se um interregno de 15 anos até os jornalistas voltarem a utilizar o “poder computacional” para as suas investigações (Léchenet, 2014). Foi, então, em 1967, que Meyer utilizou esta ferramenta tecnológica para cobrir os motins na cidade norte-americana de Detroit e “através de um inquérito, realizado com a ajuda de um computador (o IBM 360)”, conseguiu traçar o perfil dos manifestantes, o que viria a culminar na realização de um trabalho jornalístico que lhe valeu um prémio Pulitzer (*Ibidem*).

Retomando a evolução cronológica dos termos, La-Rosa & Sandoval-Martín (2016: p. 1209-1210) referem que o atual jornalismo de dados se trata de uma evolução da RAC e que esta, por sua vez, é “descendente” do “jornalismo de precisão”. Esta variação deve-se pelas diferenças em “várias características” – técnicas e culturais – relativamente ao contexto temporal em que cada uma delas emergiu.

No entanto, a RAC terá surgido só na década de 80, numa altura em que a utilização das bases de dados passou a ficar “estabelecida na produção noticiosa” (Léchenet, 2014; La-Rosa & Sandoval-Martín, 2016). A sua denominação popularizou-se durante a década seguinte, mais precisamente em 1989, coincidindo com o ano da criação do Instituto Nacional da Reportagem com Auxílio do Computador nos Estados Unidos da América (Anderson, 2015: p. 358).

Léchenet (2014) lembra que, nessa época, esta prática incidia sobretudo na “recolha de dados” e “análise estatística” – processos intimamente relacionados com a proposta de “jornalismo de precisão” de Meyer. Mas, apesar dos progressos, a disponibilidade de informações ainda era limitada. Embora os “dados computadorizados” estivessem a ser utilizados nas redações estadunidenses desde o final da década de 60, Parasie & Dagiral (2013) afirmam que só no final da década de 80 é que o uso de dados na investigação jornalística se tornou “cada vez mais comum”.

Face a este contexto, o jornalista poderia realizar a preconização de preconizado de Meyer (2002: p.3): “ler, processar e analisar dados”, sem perder de vista o objetivo da missão que é o de “encontrar os factos, compreendê-los e explicá-los sem perder tempo”.

Ao longo da década de 90, a RAC estabeleceu-se como “conceito único e unificado” pelos profissionais e académicos (Coddington, 2015: p.332). Contudo, o uso e a disponibilidade mais generalizada dos dados computadorizados com a emergência e “explosão” da *World Wide Web*, fez com que outras práticas jornalísticas semelhantes, mas com terminologias diferentes, florescessem, nomeadamente o “jornalismo computacional”, o “jornalismo de programação” e ainda o “jornalismo de dados”, o que levantou sérios desafios à distinção (e definição) entre cada um destes termos (Parasie & Dagiral, 2013; Coddington, 2015).

### 3.4 Evolução e contrastes

Antes de discorrer sobre as distinções entre as práticas que surgiram ao longo do tempo, importa realçar que, atualmente, a definição de “jornalismo de dados” não reúne consenso científico e os próprios profissionais desta área percebem-no de forma distinta (Fink & Anderson, 2015; De Maeyer et al., 2015; Lewis & Waters, 2018; Stalph & Borges-Rey 2018; Stalph, 2018; Ausserhofer et al., 2020). Tal acontece devido ao seu historial caracterizado por “várias arestas irregulares e descontinuidades da prática” ao invés de um “desenvolvimento contínuo de um trabalho quantitativo cada vez mais sofisticado” (Anderson, 2015: p.352).

Coddington (2015) reconhece a ambiguidade entre os diferentes termos e propôs, por isso, um sistema de classificação de três práticas: a RAC, o jornalismo de dados e o jornalismo computacional, em que assinala as principais características e contrastes. Para o âmbito desta investigação, iremos focar na assinalação das diferenças entre a RAC e o jornalismo de dados, na qual essa divergência estabelece-se, sobretudo, por uma “caracterização da mudança da prática” ao longo do tempo (*Ibidem*). O termo “RAC” deixou de ser usado na viragem para o século XXI e, como vimos, a sua invocação está associada à historicidade do jornalismo quantitativo, bem como às metodologias das ciências sociais. Nesse seguimento, alguns autores entendem que o jornalismo de dados se tornou “o termo mais contemporâneo” da RAC, dada a ascensão e difusão da sua prática nas redações a partir do final da década de 2000 (Coddington, 2015; La-Rosa & Sandoval-Martín, 2016). Gynnild (2014: p. 719) aprofunda esta ideia referindo que o termo “jornalismo de dados” popularizou-se aquando da publicação do artigo supramencionado de Rogers (2011), no blogue do *The Guardian*, em 2009, o que fez com que a denominação se tenha disseminado pela Europa e pelos Estados Unidos da América.

Nesse mesmo período, Léchenet (2014) identifica outros fatores que impulsionaram a democratização, disseminação e adoção desse termo, tais como: o “acesso mais barato aos computadores”, o aparecimento e utilização da Web “como fonte e suporte de publicação” e o “movimento dos dados abertos” (tradução do inglês “open-data movement”).

Ainda que comunguem de algumas semelhanças, Coddington (2015: p. 337-343) aponta cinco contrastes entre a RAC e o jornalismo de dados: primeiro incide na “abertura participativa” e “hibridismo entre diferentes domínios” característico no jornalismo de dados, que abarca áreas como a análise estatística, a ciência computacional, a representação visual de dados, o *web design* e a reportagem. A segunda distinção incide na forma como as histórias são “descobertas e contadas”: tal como na RAC, o jornalismo de dados “coloca em primeiro plano a narração da história ao invés da utilização dos dados”, mas prioriza a narração ou exploração da história jornalísticas “utilizando apenas os números como fonte”, não necessitando de “entrevistas e outras práticas jornalísticas” tradicionais. Em teoria, isso permitiria a “qualquer pessoa contar histórias”, desde que consiga “extrair com exatidão o significado dos dados”.

A terceira diferença deve-se ao conceito de “transparência”, em que a inclusão e menção dos métodos e dados utilizados são “elementos centrais” (algo que não se verificava na RAC). A quarta dissemelhança identificada, que é aprofundada no subcapítulo seguinte, diz respeito à temática da “crescente quantidade e tamanho das bases de dados” disponíveis. Este trata-se de um aspeto indissociável da “evolução da velocidade de processamento dos computadores na análise de dados” identificada por Hammond (2017: p.412). O quinto e último contraste que Coddington (2015) identifica tem que ver com o público que assume “um maior e direto envolvimento” na história jornalística produzida, pois uma das finalidades do jornalismo de dados passa por “permitir que a audiência analise e compreenda os dados por si própria”. Para este autor, a “dimensão da informação” e “o ritmo rápido do trabalho” são também aspetos intrínsecos do jornalismo de dados, o que conduz a uma “forma de análise mais exploratória e de grandes volumes de dados” comparativamente à RAC.

La-Rosa & Sandoval-Martín (2016: p. 1210) reforçam explicitamente esta última ideia ao considerarem que a RAC concentrou os seus esforços “na procura e recolha de informações num Mundo onde a informação era escassa”, enquanto o jornalismo de dados surge num contexto em que “a informação na Web é abundante” e dirige o seu foco “no processamento dos dados”.

Uma outra dissemelhança identificada na literatura entre estas duas práticas é a de que o jornalismo de dados tende a ser “mais diversificado” nos seus conteúdos e por aplicar a análise de dados em temáticas desvalorizadas pela RAC, como o entretenimento ou o desporto (Zamith, 2019: p. 476).

Assim sendo, compreende-se que as diferenças assinaladas provêm, uma vez mais, de uma evolução de fatores tecnológicos, económicos e sociais. Mas importa frisar que, tal como Gynnild (2014: p. 719) menciona, a evolução e paulatina adesão ao jornalismo de precisão, que mais tarde se reformulou como RAC, teve um papel preponderante para o atual entendimento do jornalismo de dados, uma vez que construiu as “fundações” para a “abordagem atual das redações (...) no tratamento digital de dados quantitativos”.

### **3.5 O dilúvio dos dados**

Recuperando uma das distinções mencionadas - a da crescente quantidade de dados disponibilizados (Coddington, 2015) -, Rodríguez & Clark (2021: p.74) reconhecem que a evolução tecnológica facilitou a “recolha, armazenamento e disseminação” da informação por parte de vários atores da esfera pública e privada. Um exemplo recente dessa capacidade tem que ver com o surto pandémico da COVID-19, no qual os “dados de grande dimensão” (tradução do inglês “big data”) permitiram às agências governamentais e independentes “compilar estatísticas e seguir tendências em todo o Mundo” (*Ibidem*).

Pentzold & Fechner (2020: p. 735) consideram que a intersecção entre o Jornalismo e a tecnologia é afetada, particularmente, pela “velocidade” a que atualmente os dados digitais são acumulados e analisados, e realçam os três aspetos mais proeminentes e característicos do fenómeno dos “dados de grande dimensão”: o seu (grande) volume, variedade e velocidade. Essa “imensidão” da disponibilidade de dados decorre também da “crescente ubiquidade” dos dispositivos móveis, máquinas “inteligentes”, dados digitais de rastreio, repositórios e arquivos digitais sobre a sociedade, assim como os “fragmentos de atividade social e natural representados por cliques, tweets, gostos, coordenadas GPS”, entre outros (Lewis & Westlund, 2015: p.447-448).

Ou seja, a “quase omnipresença” de dispositivos e registos digitais de informação tornaram-se fontes de informação sobre o próprio utilizador e o meio que o rodeia. Face às “origens e implicações culturais, económicas e políticas” decorrentes deste “fenómeno sociotécnico”, Lewis & Westlund (2015) entendem que os “dados de grande dimensão” não se trata apenas de uma “transição tecnológica do dilúvio da informação”, pois acarretam implicações em várias dimensões da sociedade civil. Já numa perspetiva “estritamente computacional”, os autores definem os “dados de grande dimensão” como um conceito relativo a “bases de dados demasiado grandes para serem processadas pela memória e *software* de um computador normal”, embora a sua definição possa “bastante flexível” e assumir “diferentes significados, em diferentes contextos, para objetivos distintos”.

A origem do fenómeno dos “dados de grande dimensão”, segundo Tandoc & Oh (2017: p. 1002), remonta à área do Marketing, onde “a informação sobre consumidores e o mercado é compilada em massa”, tendo como propósito direccionar as estratégias comerciais das empresas. Por conseguinte, essa quantificação dos dados e as “novas formas” de retirar ilações, através da crescente capacidade computacional e do armazenamento de informações, permitiram a expansão para outras áreas. No que ao Jornalismo diz respeito, esta “imensidão de informação” incorpora “ideias emergentes, atividades e normas” que podem ser encontradas e recolhidas a partir de “bases de dados, algoritmos, métodos computacionais, processos e perspetivas ligadas à quantificação” (Lewis & Westlund, 2015). Esta conjuntura *per se* “não é boa, nem é má” para o Jornalismo, dependendo, por isso, da forma, dos objetivos e com que interesses a aplicação dos “grandes dados” é implementada (*Ibidem*).

Partindo da ideia de que o Jornalismo procura “dar sentido” à grande quantidade de informação disponível na Web, os desenvolvimentos para cumprir esse objetivo impactaram as formas de saber (epistemologia) e de fazer (especialização), assim como para o valor (economia) e valores (ética) dos profissionais do jornalismo. A menção a estes quatro aspetos “distintos, mas interrelacionados” (Lewis & Westlund, 2015) é verificável em acontecimentos marcantes do jornalismo de dados.

A investigação de Landert & Miscione (2017: p.14) sobre a peça jornalística de dados “Afghan War Logs” - que parte das informações proveniente da Wikileaks – do jornal *The Guardian*, mostra que os jornalistas tiveram de compreender os dados e saber o que fazer com os mesmos (epistemologia) dada a sua “complexidade e tecnicidade”. Seguiu-se o tratamento, seleção e apresentação (especialização) dos “incidentes mais importantes”. Na sequência da investigação e notícias produzidas, este órgão de comunicação social angariou valor e credibilidade (economia) junto dos seus leitores por ter tratado a informação de forma responsável (ética), visto que optaram por não incluir “informação sensível”, como “nomes de informadores” ou “elementos que pudessem colocar em risco” os agentes envolvidos. Assim, os quatro conceitos avançados por Lewis & Westlund (2015) - epistemologia, especialização, economia e ética – verificaram-se na prática.

No seguimento do progresso tecnológico, Sandoval-Martín & La-Rosa (2018: p.204) enumeram seis “elementos diferenciadores” do jornalismo de dados associados ao “fenómeno tecnológico, social e cultural” dos “dados de grande dimensão”: 1) a “massificação das ferramentas para trabalhar com os dados graças ao software gratuito”; 2) a importância dos “dados abertos” e a reivindicação por uma maior transparência; 3) o trabalho colaborativo com acesso público à informação e análise utilizadas; 4) o trabalho em equipa com grupos profissionais multidisciplinares; 5) a representação visual dos dados, e 6) o acesso facilitado aos dados decorrente da “cultura de transparência” e dos “dados abertos”. Para os autores estas práticas constituem uma “revalorização da profissão”, dada a aplicação de “novas ferramentas informáticas” para o processamento e tratamento de “enormes quantidades de dados”, que não existiam na altura do jornalismo de precisão ou da RAC.

Também Colussi & Gomes-Franco e Silva (2017: p. 166) entendem que é a partir desta relação com os “dados de grande dimensão” que se constitui o que é hoje entendido como “jornalismo de dados”. Embora isso não exclua a possibilidade de o “jornalismo tradicional” utilizar “dados brutos como fonte de informação”, no caso do jornalismo de dados o que ressalta é o “volume de dados trabalhados” pelos seus profissionais.

Conforme mencionado, a importância do jornalismo de dados nesta era do “dilúvio da informação” reside, sobretudo, na capacidade dos jornalistas em “providenciar contexto e clareza” e “encontrar a verdade na quantidade crescente de conteúdos digitais no mundo” (Gray, Bounegru & Chambers, 2012). “Encontrar sentido” nos dados de grande dimensão, nomeadamente em “dados não estruturados”, afigura-se hoje como o “principal objetivo” para quem lida com este tipo de informação (*Ibidem*).

A menção aos episódios da Wikileaks, os vazamentos de Edward Snowden, em 2013, ou ainda os recentes *Panama Papers* são marcos deste fenómeno contemporâneo, que Gray & Bounegru (2019: p.2) destacam na segunda edição do *The Data Journalism Handbook*. Estes acontecimentos, em que grandes quantidades de informação foram vazadas, promoveram uma “consciencialização mais ampla” sobre a prática do jornalismo de dados. O caso das revelações de Snowden, por exemplo, contribuiu para que os cidadãos compreendessem o “aparato de vigilância transnacional” que estava a ser conduzido por governos e empresas, em que a sociedade estava a ser “reconhecida” a partir dos seus “dispositivos, aplicativos e plataformas geradoras de dados” (*Ibidem*).

### **3.6 “Abertura” e transparência dos dados**

O advento dos “dados de grande dimensão” é indissociável de outro aspeto comumente mencionado no atual jornalismo de dados e que o distingue das práticas antecessoras: os “dados abertos” (tradução do inglês “open data”) e a “cultura de transparência”.

De acordo com Lesage & Hackett (2014: p.45) a história dos “dados abertos” está intimamente relacionada com a da Informática, sobretudo com o movimento inicial de *software* livre de patentes no final da década de 90 (Coddington, 2015: p.332-333). Mas também se deve pela “longa e bem documentada tradição académica” da revisão por pares na investigação científica. Da mesma forma, a “abertura” ou “transparência” dos dados é frequentemente associada à ideia original da “utilização pretendida com a Internet na sua fundação”, se a percebermos como “instrumento de transparência pública” sobre a informação detida por governos e instituições (Brolcháin et al., 2017).

Importa, contudo, compreender primeiramente o que são “dados”. Appelgren & Salaverría (2018) entendem que se trata de informação que é “frequentemente apresentada de forma desestruturada”, podendo ser obtida de diversas formas, nomeadamente: “pedidos de acesso de informação, recolha automatizada a partir de sítios Web, fugas de informação, subscrição a bases de dados (...)”, entre outros.

Já Tabary et al. (2016: p.68) definem “dados” como “abstrações que tornam possível a articulação de um problema” e acrescentam que o tratamento deste tipo de informação, no contexto jornalístico atual, requer “competências estatísticas, informáticas e de representação gráfica” para a formulação de “suposições contextualizadas”. Se esse conjunto de condições prévias não for assegurado, a análise quantitativa poderá resultar na “apresentação de uma abstração como uma manifestação da realidade”, o que, por outras palavras, se traduz numa análise enviesada ou incorreta da informação quantitativa. Neste sentido, os dados podem ser entendidos como “um fragmento oculto de informação à espera de ser descoberto” e a sua análise como um meio de “descobrir verdades ocultas” (Anderson, 2015: p. 360). Por norma, estes “fragmentos ocultos de informação” são os “dados em bruto” (tradução do inglês “raw data”), que são entendidos como “símbolos que ainda não foram interpretados” e que só se tornam em “informação” quando “estruturados e contextualizados”, por exemplo, numa história jornalística (*Ibidem*).

Por conseguinte, essa informação só se torna em “conhecimento” quando é “interpretada, significativa e acionável” (Baack, 2015: p.4) para os leitores. Além disso, os dados são também entendidos como uma “base factual” para a análise; uma forma de “tentar minimizar os riscos” de reportar erradamente acontecimentos e uma fonte de informação para “contrariar a influência dos assessores de imprensa” na agenda mediática (Lesage & Hackett, 2014: p. 47). De outro ponto de vista, Cushion et al. (2017: p.1198-1199) entendem que os dados são “uma poderosa ferramenta democrática”, sob a qual o Jornalismo deve ter um papel a desempenhar ao nível da “descoberta, tradução e interpretação”.

Há, no entanto, o “pré-requisito sistémico” que Appelgren & Salaverría (2018) destacam: a transparência por parte das instituições públicas no fornecimento da informação. Sem a garantia prévia dessa condição, tornar-se-ia difícil os jornalistas aceder aos dados que, aliás, se encontram muitas vezes no “centro das lutas do poder político”.

Por sua vez, Baack (2015: p. 4) adota a definição de “dados abertos” a partir da *Open Knowledge Foundation*, que os designa como “os dados que podem ser acedidos, modificados e partilhados por qualquer pessoa, para qualquer propósito, sem restrição”. Em termos técnicos, a “abertura dos dados” refere-se à supressão de entraves, na qual as bases de dados devem ser “completas, divulgadas em tempo útil, acessíveis, capazes de serem lidas por máquinas e disponibilizadas em formatos abertos”. O autor entende que os dados deveriam ser “informação estruturada” e passível de ser “analisada, editada e combinada com outros dados”, de forma a contribuir para o “conhecimento aberto”. Para que tal seja exequível, a “informação em bruto” deve ser facultada pelas organizações governamentais, permitindo que os “intermediários especializados” possam “refinar” a informação e facilitem a compreensão (*ibidem*). Caso esse processo não venha a ser realizado por profissionais especializados, muitos desses dados - mesmo que o seu acesso esteja disponível - poderão nunca chegar ao entendimento do cidadão comum, criando “um fosso” entre a “proliferação emergente de dados públicos abertos” e a capacidade da sociedade em “aceder e compreender as informações mais relevantes” (Brolcháin et al., 2017, s/p).

Recuando ao contexto histórico do movimento dos “dados abertos”, o primeiro regulamento na União Europeia sobre o acesso a documentos públicos foi redigido em 2001 (Appelgren & Salaverría, 2018: p.987) e, ainda que vários países tenham adotado leis de transparência, o desempenho da legislação e a perceção do público sobre a transparência das entidades públicas variam bastante.

Appelgren & Salaverría (2018) realçam que a diretiva da Informação do Setor Público (ISP), adotada em 2003 pela Comissão Europeia, foi um incentivo importante para a “reutilização de dados do sector público”, permitindo a “criação de novos serviços por parte de terceiros”, assim como o acesso dos jornalistas aos dados públicos digitais.

Contudo, o objetivo dessa promoção de acessibilidade digital a informação é dissonante com aquilo que os jornalistas se deparam, encontrando, frequentemente, “obstáculos burocráticos” que se desdobram em “negociações prolongadas” e, em certos casos, os dados são fornecidos, mas num formato ilegível para o computador ou certos *softwares* (*Ibidem*). Esta observação é pertinente, pois Brolcháin et al. (2017: s/p) indicam que o jornalismo de dados “começa com a exploração de bases de dados em plataformas acessíveis”, sobre as quais os governos são os “maiores produtores” desse tipo de informação. Ora, se esses dados públicos estão inacessíveis ou ilegíveis, o jornalismo de dados carecerá da sua “matéria-prima”.

A “cultura de transparência” é, por isso, frequentemente evocada como um princípio fundamental nesta área e há até quem sugira que pode mesmo vir a substituir a objetividade enquanto “nova norma jornalística”, sobretudo na era digital (Chadha & Koliska, 2015: p.215-216), visto que a ênfase desta prática passa por “tornar visível” as decisões, métodos e fontes utilizadas na produção de uma peça conduzida por dados. Alguns autores (Lesage & Hackett, 2014; Weber et al., 2018) argumentam que a transparência, bem como os “dados abertos”, significa também “dar a possibilidade ao público de analisar a informação em bruto por si mesmo” e retirar as próprias conclusões.

Em jornalismo de dados, isso significa “explicar o processo de recolha, análise e apresentação” da informação, permitindo que os utilizadores “verifiquem os dados do trabalho do jornalista”, constituindo uma “estratégia de gestão qualitativa e norma ética” da profissão (Weber et al., 2018: p.203). Investigações académicas sobre este assunto dão conta de que, por exemplo, a mera inclusão da hiperligação dos “dados em bruto” utilizados já é considerado um “ritual de transparência” (Tandoc & Oh, 2017).

A cultura “por uma maior transparência e abertura dos dados”, a explicação dos métodos utilizados ou a publicação dos “dados em bruto”, permitirão assim a reprodutibilidade dos resultados e conclusões obtidos a partir da informação utilizada pelo jornalista, o que prova a “evolução na cultura da transparência” no jornalismo (Sandoval-Martín & La-Rosa, 2018: p.204).

Parasie & Dagiral (2013: p.854) consideram que o jornalismo de dados contemporâneo melhorou a forma como o Jornalismo contribui para a democracia, sobretudo numa época em que a informação é abundante. Já Borges-Rey (2020: p.926-927) afirma que, quando os governos locais se envolvem com a “dataficação da vida social” e permitem o escrutínio às suas bases de dados, constrói-se uma “constante intermediação” entre os ideais de transparência e de abertura, amplificando a “responsabilidade social” dos jornalistas de dados para com as comunidades locais. Em última análise, a garantia desse acesso contribuirá para uma “compreensão mais ampla e aprofundada” dos problemas da sociedade, assim como para o escrutínio dos governantes democraticamente eleitos.

Importa realçar que também as redes sociais podem ser uma fonte alternativa ou complementar às bases de dados públicas. Gehrke & Benetti (2020) entendem que, em alguns casos, essas plataformas podem servir de “fonte documental estatística”, da qual a “extração massiva de dados” pode servir de “matéria-prima para análises”, ampliando a variedade de fontes utilizadas.

### **3.7 As definições do jornalismo de dados**

Considerando o contexto histórico do jornalismo quantitativo, a evolução decorrente entre as diferentes práticas, o advento da grande quantidade dos dados e da importância da transparência e acessibilidade a esta informação por parte das instituições públicas, Anderson (2017) considera que a profissão “inquietantemente híbrida” do jornalismo de dados se situa hoje num “novo terreno”. A hibridez apontada por este autor é, como vimos, um pilar do jornalismo de dados que, por conseguinte, torna a sua definição complexa e provoca “tensões nas estruturas e culturas organizacionais” (Stalph, 2020: p.10). Contudo, a sua definição é relevante para que a investigação académica possa ser “coesa e coerente”, permitindo analisar o seu significado profissional e cultural (Coddington, 2015: p. 332), e a estabelecer “limites e identidades profissionais” (Wright & Doyle, 2019: p.1815).

Conforme referimos, no início do século XXI, o surgimento do "jornalismo de dados" marcou o início de uma nova fase em que a ampla disponibilidade de dados *online*, combinada com a utilização de ferramentas especializadas, possibilitou que um número crescente de pessoas trabalhasse com dados de uma forma que antes não havia sido possível (Gray, Bounegru & Chambers, 2012). A sua popularização no final da primeira década do século XXI, foi potenciada nos Estados Unidos da América, pelo caso da *Wikileaks*, e no Reino Unido na sequência do "escândalo" das despesas do parlamento britânico, (Stalph & Borges-Rey, 2018: p.1078-1079), o que levou a que alguns média tradicionais – como o *The Guardian* ou *The New York Times* – apostassem no desenvolvimento desta área nas suas redações, criando "unidades especializadas". Essa decisão terá influenciado outros órgãos de comunicação social em todo o Mundo (*Ibidem*). Também devido à sua disseminação geográfica, o seu desenvolvimento e conceptualização não é uniforme (Wright et al., 2019: p.1299).

Os jornais generalistas portugueses só começaram a produzir peças de jornalismo de dados a partir do início da década de 2010 (Alexandre, 2020), entre os quais o jornal Público se destacava pela qualidade de algumas reportagens inovadoras. De acordo com este autor, mais recentemente outros órgãos de comunicação social portugueses começaram a produzir peças deste tipo, mas não existe ainda uma "grande periodicidade na produção desse conteúdo (...)". Isso deve-se ao facto de não existirem equipas exclusivamente dedicadas à produção deste tipo de "conteúdo interativo e de qualidade", que exige "tempo" e "profissionais com diferentes conhecimentos, como programação, análise de dados e design gráfico". Existem somente "um ou dois jornalistas de dados nas redações" que, na sua maioria, são "autodidatas" ou provêm de outras áreas profissionais.

Mas o que é o jornalismo de dados? Knight (2015), por exemplo, considera que jornalismo de dados é tudo aquilo que engloba "uma história cuja fonte primária é a informação numérica" ou que contenha "um elemento substancial de dados ou representações visuais", permitindo "incorporar e captar o maior número de exemplos" de peças de jornalismo de dados. Partindo dessa premissa, Gynnild (2014: p. 719) argumenta que o jornalismo de dados se trata de uma prática jornalística que opera "apenas com dados abertos".

Já Martinho (2014) entende que a designação do termo é justificada por a sua “matéria-prima principal” serem dados, independentemente de a narração jornalística seguir uma orientação “para contar histórias, para os processos ou para as bases de dados”. No mesmo sentido, Léchenet (2014) argumenta que se trata da “utilização mais alargada dos dados no Jornalismo” e que tende a mostrar o “processo da história” através de representações visuais e do “acesso direto” à fonte.

Já Howard (2014: p. 5) refere que o jornalismo de dados é “contar histórias com números ou encontrar histórias neles” e que o tratamento dessa informação pode servir como “fonte complementar” aos “testemunhos humanos, oficiais ou dos especialistas”. Por outro lado, há quem considere se trata de uma prática que lida apenas com “dados abertos”, ou seja, os “dados que estão disponíveis *online* e que podem ser analisados através de ferramentas de fonte aberta e de livre acesso” (Gynnild, 2014: p.719).

Ainda que Landert & Miscione (2017: p.15) considerem que a informação quantitativa desempenha uma função determinante na “credibilização da notícia”, a história jornalística “é muito mais do que uma simples coleção de informações factuais”. No livro *Factfulness*, Rosling et al. (2019: p.200) corroboram esta afirmação ao mencionar que os números “nunca contarão a história completa”, pois “o Mundo não pode ser compreendido sem números” e, da mesma forma, “não pode ser compreendido só com números”.

Nesse seguimento, compreende-se que Bradshaw (2012: s/p) considere que simplificar o jornalismo de dados como o “jornalismo que é feito com dados” é redutor. Entende, por isso, que esta prática comporta “novas possibilidades” que podem auxiliar o jornalista na análise e narração de uma história com dados complexos, recorrendo, por exemplo, a “gráficos interativos” ou ao “uso da programação” para “automatizar processos de recolha e de combinação de informações”.

Os métodos de recolha de dados, mas sobretudo da “tradução da informação numérica” em “histórias de simples compreensão”, levam a que Brolcháin et al. (2017: s/p) percecionem o jornalismo de dados como um “subdomínio” da “narração de histórias conduzidas por dados” (tradução do inglês “data-driven storytelling”).

Baseando-se no primeiro debate em torno do jornalismo de dados - organizado pelo Centro Europeu de Jornalismo, em 2010, (pouco tempo depois da reportagem *Afghan War Logs*) -, os autores partilham a definição que os jornalistas participantes propuseram: o jornalismo de dados mais não é do que “o processo que inclui a recolha de dados, filtragem de dados, representações visuais e a publicação” da notícia. Acrescentam que a “narração de histórias conduzidas por dados” pode ser entendida como o “processo de tradução” da análise de dados em “histórias simples e lógicas que podem ser compreendidas” pelo cidadão comum.

Também Colussi & Gomes-Franco e Silva (2017: p.167, *as cited in Bertocchi, 2013*) entendem que o jornalismo de dados se trata do “processo jornalístico” que começa pela recolha de dados e a sua “curadoria” culminando com a “representação visual” num formato entendível para as audiências através das “interfaces digitais”. Estas técnicas do jornalismo de dados são equiparadas pelos autores ao jornalismo de investigação, pois envolvem “formas e técnicas de recolha de notícias” a partir de bases de dados “complexas e inexploradas”, que originam notícias, promovendo a “democratização do fluxo de informação” e que pode contemplar a criação e apresentação de infográficos (Cushion et al., 2017: p. 1199). Embora Stalph (2018) tenha presumido que os dados não têm necessariamente de ser apresentados sob a forma de representações visuais, podendo ser apresentados “em forma textual”, a sua investigação mostrou que as representações gráficas são uma “característica chave” do jornalismo de dados e que podem, inclusive, “melhorar as histórias” que considera serem “distintas” das do “jornalismo tradicional”.

Por sua vez, Robles & López (2021: p.647) entendem que o jornalismo de dados continua a comportar os mesmos “elementos clássicos”, aptidões e princípios do jornalismo tradicional. No entanto, assinalam que um dos contrastes reside nos “processos computacionais”, que exigem “conhecimentos básicos” em Estatística ou em linguagem de programação. Segundo os mesmos, a aprendizagem e aplicação destas competências permitem ao jornalista “procurar por notícias nos números”, conduzir “entrevistas às bases de dados” e, posteriormente, contextualizar a informação.

Tendo em conta esta intersecção multidisciplinar, Sandoval-Martín & La-Rosa (2018: p.195) consideram que o jornalismo de dados se trata de uma “proposta jornalística que emergiu da Web”, que comunga do “rigor metodológico das ciências sociais para a extração, padronização e tratamento de dados”. Tal constatação alude às influências do jornalismo de precisão e da RAC e sustentam a ideia de que o jornalismo de dados se trata de um “produto da evolução do jornalismo a par da evolução tecnológica e social” (Martinho, 2014: p.72).

Ao perceber o jornalismo de dados como “o processo através do qual os jornalistas dão sentido aos dados (...) para encontrar ângulos e pormenores para a história”, Tandoc & Oh (2017: p. 1002) entendem esta é uma “nova área de especialização” nas redações, em que o jornalista produz a história “sobre o conjunto de dados” e fornece “mais contexto” e as “conclusões mais significativas da análise” através de recursos multimédia e componentes visuais que facilitam a compreensão da informação.

Decorrente desta capacitação tecnológica e do cruzamento de conhecimentos de outras áreas, compreende-se que o desenvolvimento e aplicação de determinados *softwares* e ferramentas digitais - como a Interface de Programação de Aplicações (API's, do inglês “Application Programming Interface”) que permite ao computador “captar informações de fontes *online*” (Bradshaw, 2014: p. 211); as “ferramentas visuais”, como o *Tableau* (Felle, 2016), e outros recursos interativos (Zamith, 2019) – expandiram as possibilidades de como a história jornalística baseada em dados pode ser trabalhada, narrada e apresentada.

Felle (2016: p.88) entende que, para a audiência, estas mutações terão tornado o jornalismo “mais envolvente e interativo”, possibilitando aos leitores aprofundar a informação “até ao nível individual ou local”, convocando a ideia da granularidade e do volume da informação disponível *online*. Também o “design visual” e as “ferramentas tecnológicas” constituem-se como “instrumentos centrais” para a apresentação de uma peça de jornalismo de dados por serem capazes de integrar um “elevado nível de interatividade” para que as suas audiências possam “criar e encontrar a sua própria história nos dados” (Appelgren, 2018: p. 309).

Parratt-Fernández et al. (2021: p.9) resumem na sua proposta de definição referindo que o jornalismo de dados se trata de “um processo de análise a grandes bases de dados para apresentar notícias com dados relevantes”, mas também de “um produto que permite ao utilizador (...) interagir com as notícias através da representação visual”.

Neste “mundo da narrativa multimédia”, o jornalismo de dados parece também estar interligado com a “interpretação das suas audiências, apresentação das visualizações de dados e na análise de grandes volumes de dados” (Du et al., 2019: p.181). Quer isto dizer que o processo jornalístico transcende a interpretação e análise de números ou estatísticas, procurando assim “traduzir” as grandes quantidades de dados com a criação de representações visuais apelativas e interativas, sobre assuntos de interesse público. A investigação de Charbonneau & Gkouskou-Giannakou (2015: p.260) refere mesmo que o discurso em torno do jornalismo de dados não está somente relacionado com o “volume ou o teor qualitativo da informação revelada”, mas também com “a forma como a informação é apresentada para ser lida”. Nesta lógica, o jornalismo de dados é também entendido como “uma oportunidade comercial inovadora” ao ter em conta a criação de “conteúdos distintos” nas plataformas online (De-Lima-Santos & Salaverría, 2021: p.9).

A análise de Weber et al. (2018) às narrativas jornalísticas baseadas em dados identificou sete atributos que resumem os principais aspetos até aqui mencionados e que procuram contribuir para a sua definição: 1) os dados como um “aspeto central” da história jornalística; 2) os “diversos objetivos comunicativos” da narrativa; 3) as novas relações entre a componente textual e a visual na narrativa jornalística, assim como 4) as novas estruturas e formas de *design* das narrações jornalísticas no contexto digital, 5) a constatação de diferentes níveis e tipos de interatividade, 6) a importância crescente da transparência e de acesso às fontes utilizadas, e 7) a disponibilização da análise conduzida pelo jornalista. Em suma, Ausserhofer et al. (2020) reconhecem que cada definição de jornalismo de dados comporta aspetos distintos, seja enquanto processo ou produto, mas realçam que há denominadores comuns, nomeadamente por se tratar de um processo jornalístico caracterizado por “desenvolver histórias conduzidas por dados através da análise a grandes bases de dados com métodos quantitativos e computacionais”.

Essas histórias comportam uma “forma especial” de apresentação com as representações visuais de dados interativas, em que o propósito incide em usar “informação estruturada” para a criação de notícias.

A tónica desta definição é também partilhada por Zamith (2019: p.473) que reitera duas das características mais evidenciadas: o foco na informação quantitativa em torno da construção da narrativa jornalística, e a representação visual dos dados na história. No entanto, para Martinho (2014), mais importante do que alcançar um consenso sobre a sua definição, a atenção deve recair no paradigma das “novas abordagens” que o jornalismo de dados incorpora, nomeadamente “o crescimento da quantidade de dados disponíveis (...); os avanços tecnológicos que possibilitam a democratização de formas mais acessíveis e rápidas de tratamento e visualização de dados” e o “aumento da literacia digital da população”.

Considerando a multiplicidade de definições, compreende-se que o jornalismo de dados é um “termo genérico” e contemporâneo, que emerge no contexto da *World Wide Web*, e abarca uma “diversidade crescente de ferramentas, técnicas e abordagens” para contar histórias a partir de dados e “fornecer informações e análises” que ajudem a compreender questões relevantes da atualidade (Gray, Bounegru & Chambers, 2012). Ainda assim, a definição de Howard (2014) parece-nos ser das mais sensatas, pois entende que o jornalismo de dados se pode resumir à aplicação da ciência de dados – aqui entendida como “o estudo da extração de conhecimentos a partir de dados” – no Jornalismo, onde técnicas como a recolha, limpeza, organização, análise, representação visual e publicação dos dados criam os “artefactos jornalísticos”. O autor considera, por isso, que o jornalismo de dados comporta, essencialmente, três etapas: a recolha e verificação dos dados como “fonte credível” de informações; a aplicação de técnicas e conhecimentos estatísticos para “interrogar” os dados e, por fim, a apresentação dos “resultados” através de representações visuais da informação. Pode ainda ser incluída uma quarta dimensão relacionada com o arquivamento e publicação dos “dados brutos” utilizados nas investigações jornalísticas, assim como “a metodologia e o código utilizados nas análises que levaram às conclusões publicadas” (*Ibidem*: p.4).

Dada a sintetização e abrangência desta definição, que abarca a generalidade dos aspetos mencionados, será a que tomaremos como fio condutor da nossa investigação.

Um último aspeto característico desta área jornalística “altamente especializada” (Boyles & Meyer, 2017: p.435) é o da colaboração com atores externos (Ureta & Ferreras-Rodríguez, 2021). Tomando como ponto de partida o caso dos *Panama Papers*, que se revelou “o maior vazamento na história do jornalismo de dados” com “11,5 milhões de registos financeiros e legais”, que se traduziam em 2,6 *terabytes* de informação (Gray & Bounegru, 2019: p.94) - uma quantidade de informação que representa “1500 vezes mais” do que a da *Wikileaks* (Landert & Miscione, 2017: p.19) - , o modelo e as redes colaborativas internacionais surgem como uma mais-valia para agilizar a pesquisa e análise de grandes quantidades de informação (Gray & Bounegru, 2019).

Não obstante, a colaboração entre diferentes agentes trouxe “mudanças de comportamentos nas redações”, nas quais se incluem “novas relações entre jornalistas e a sociedade civil”, assim como o envolvimento de “atores periféricos” que auxiliam no processo de trabalho (De-Lima-Santos & Mesquita, 2021a: p. 1418). Coddington (2015) esclarece mesmo que, embora haja o cuidado na “seleção editorial e no discernimento profissional na análise e apresentação de dados”, os jornalistas de dados reconhecem que as competências para “analisar e extrair significado” dos dados também existem “fora da profissão” e até “entre o público”.

De-Lima-Santos & Mesquita (2021: p.1418-1423) entendem que os “agentes externos ao jornalismo”, que têm vindo a conquistar o seu lugar em equipas de jornalismo de dados, provêm tipicamente do ramo tecnológico e vêm munidos de um pensamento de “cultura aberta”, enquanto “dinâmica estrutural e cultural”. Neste sentido, compreende-se que exista uma “cultura cooperativista”, dada a integração de diferentes áreas (como as ciências sociais e a tecnologia), onde “profissionais qualificados” operam numa lógica de “difusão de conhecimentos” e entrelaçada para resolver problemas. É, nesse sentido, que floresce um “trabalho frequentemente conjunto” entre “jornalistas e tecnólogos” na apresentação de novas alternativas graças à “flexibilidade e interatividade” das plataformas digitais dos média (*Ibidem*).

Boyles & Meyer (2017: p.435) afirmam mesmo que a produção em jornalismo de dados exige mais trabalho de equipa do que qualquer outro modelo tradicional do Jornalismo, pois os seus profissionais desempenham “funções altamente especializadas”.

Alguns desses atores periféricos ao jornalismo incluem programadores, desenvolvedores de *software*, analistas de dados, “hackers” (Lewis & Usher, 2014) empresas do ramo tecnológico, infografistas (Stalph, 2020), especialistas correspondentes (Borges-Rey, 2016) ou iniciativas da sociedade civil (De-Lima-Santos & Mesquita, 2021a: p.1430). A fusão entre jornalistas e profissionais do ramo tecnológico exige “um esforço significativo, coordenado e sustentado” (Lewis & Usher, 2014) e, ainda que se constatem conflitos éticos, Baack (2018: p.688) considera que os jornalistas de dados e os tecnólogos “complementam-se” e “partilham as mesmas aspirações” em torno do interesse público.

### **3.8 Os fluxos de trabalho**

As fases do fluxo de trabalho em jornalismo de dados são, naturalmente, influenciadas pela natureza metodológica das práticas antecessoras. Por isso, trabalhar nesta área requer “aptidões especializadas” para elaborar este tipo de reportagens - como a literacia de dados, estatística e design visual (De-Lima-Santos & Mesquita, 2021b: p.122) -, assim como “poderosas e sofisticadas ferramentas” de análise de dados, o que evidencia a necessidade de uma “especialização considerável” dos seus profissionais (Ureta & Ferreras-Rodríguez, 2021: p. 52).

O primeiro volume do livro *The Data Journalism Handbook*, da autoria Gray, Bounegru & Chambers (2012), refere que essas ferramentas incorporaram “vários aspetos e campos de conhecimento” que cruzam o uso dos dados, enquanto fonte primária, com a conceção aprofundada de representações visuais e interativas, a linguagem de programação, novas formas de narração e *news apps*.

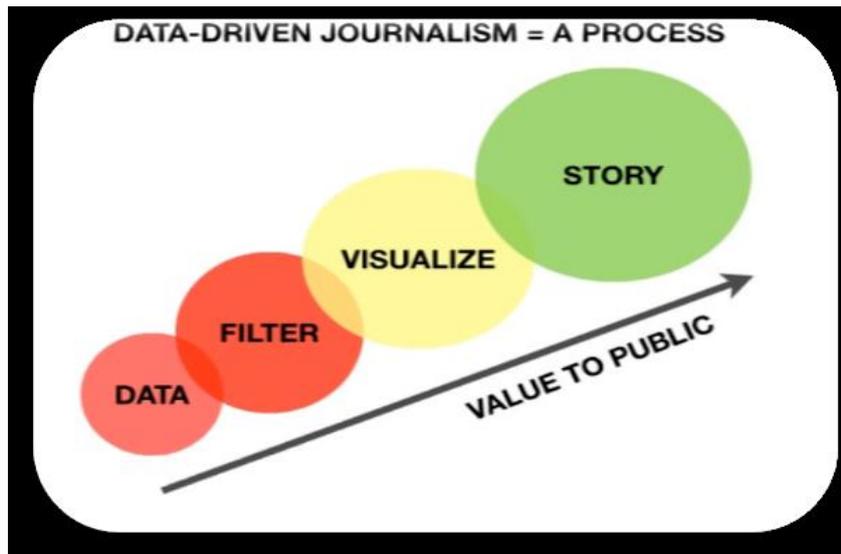
Assim, existem, pelo menos, quatro áreas de conhecimento que se evidenciam: o jornalismo, a programação, a análise de dados e estatística, e o design visual (Araújo, 2019: p.86). Ojo & Heravi (2017: p. 21) reforçam que, de forma a assegurar a produção de “boas histórias jornalísticas baseadas em dados”, são necessários profissionais com “competências de computação e programação, de análise de dados e representação visual de dados”, além das competências “tradicionais” do Jornalismo.

No mesmo sentido, Bastos (2006) dá conta de que, no contexto do ciberjornalismo, o jornalista deve possuir um “domínio alargado de múltiplas capacidades”, nas quais se incluem a produção textual e audiovisual de conteúdos, a inclusão de hiperligações, construção de páginas Web ou ainda o fornecimento de interfaces que permitam aos utilizadores recorrer a bases de dados diversas.

Face a este contexto de competências profissionais especializadas, Wright & Doyle (2019: p.1813-1814) consideram que existem dois contextos profissionais na produção de jornalismo de dados. O primeiro, que é “mais frequente” nas “organizações com recursos”, consiste numa “pequena equipa de profissionais” dedicada a projetos de investigação de maior envergadura, em que executam “diferentes partes do processo jornalístico de dados”, existindo a hipótese de se recorrer à “abordagem colaborativa” – com a contratação de atores externos à redação. O segundo passa por os jornalistas de dados fazerem a maioria ou todos os processos de trabalho de forma individual e quase autónoma.

No que concerne às fases da sua operação - como a recolha, limpeza, organização, análise, visualização e publicação dos dados (Howard, 2014) -, estas concretizam-se através de um fluxo que, mediante a sua evolução, valorizam a história jornalística (Martinho, 2014). A autora portuguesa dá a conhecer dois esquemas do fluxo de trabalho. O primeiro esquema, da autoria de Mirko Lorenz, ilustra que conforme os dados vão sendo trabalhados nas diferentes fases, o valor da história ganha cada vez maior relevância para o público. A sua ilustração (*Figura 1*) terá sido apresentada numa “mesa-redonda” sobre *Data Driven Journalism*, organizada pelo Centro Europeu de Jornalismo (*Ibidem*).

Figura 1 - Esquema do processo de trabalho do jornalismo de dados

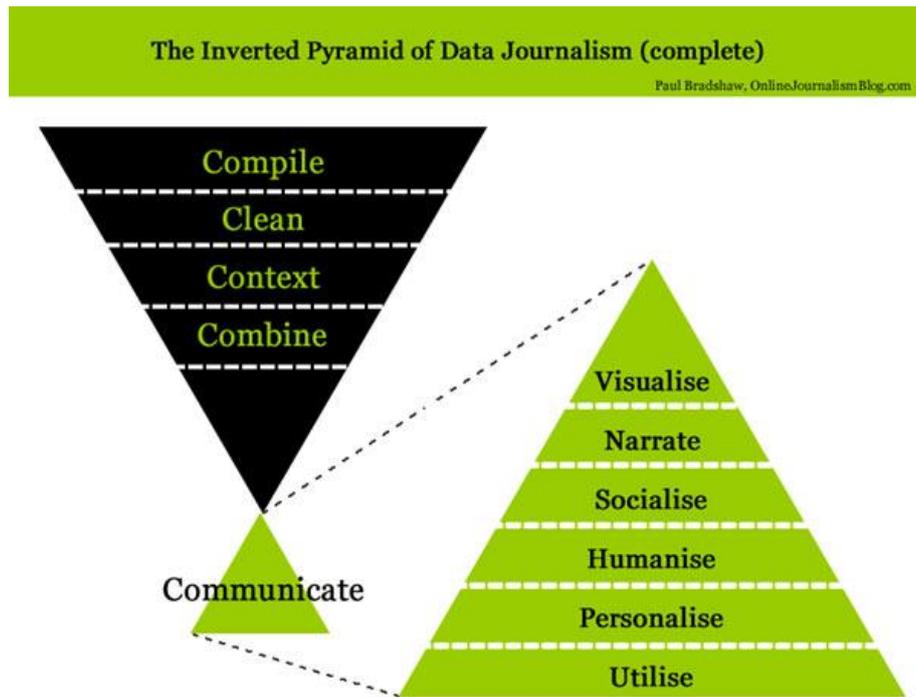


Fonte: Martinho (2014)

Este esquema mostra que a primeira fase do processo incide na obtenção e recolha dos dados, seguindo-se a sua filtragem ou seleção, a criação da representação visual dos dados e, por último, a construção e publicação da história jornalística. Ainda que seja simplista, a figura demonstra que o valor da história jornalística cresce consoante o processo do fluxo de trabalho. Uma proposta mais elaborada, complexa e referenciada como “das mais difundidas” (Robles & López, 2021: p.640), que Martinho (2014) também apresenta é a da pirâmide invertida do jornalismo de dados, que Paul Bradshaw publicou em “Online Journalism Blog”<sup>2</sup>.

<sup>2</sup> Disponível em: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism> [Consultado pela última vez a 10 de junho de 2023]

Figura 2 – A pirâmide invertida do jornalismo de dados



Fonte: Martinho (2014)

Nesta readaptação da clássica pirâmide invertida do Jornalismo (*Figura 2*), Martinho (2014) explica que Bradshaw procurou catalogá-la em quatro fases: a compilação dos dados, a sua limpeza, contextualização e combinação. “Após a última fase, o autor acrescenta a comunicação dos dados (que aqui significa a ‘descodificação’ para o público), que originará uma nova pirâmide. Esta apresenta o que deve acontecer com o produto final da pirâmide anterior: visualização, narração, socialização, humanização, personalização e utilização” (*Ibidem*: p.71).

No mesmo sentido, Brolcháin et al. (2017) explicitam que, após a pesquisa e recolha dos dados, segue-se “a filtragem da informação”, na qual apenas os “dados importantes” são extraídos para serem “transformados em visualizações”. Quando a informação estiver processada e ilustrada de uma forma entendível, a “história relevante” constrói-se em torno dos dados.

Gupta et al. (2016: p.62) acrescentam que todas estas etapas são de “natureza iterativa”, o que significa que “sempre que houver necessidade de refinar ou de melhorar” é possível regressar à etapa anterior e refazer todo o processo, e que o último passo neste fluxo de trabalho é o da publicação nos meios de comunicação, após testar e verificar a “usabilidade e funcionalidade” dos gráficos interativos.

### **3.8.1 Recolha dos dados**

O ponto de partida do processo de trabalho no jornalismo de dados é, portanto, a recolha dos dados ou a pesquisa por bases de dados acessíveis ou “abertas” (Weber et al., 2018: p. 194). Lewis et al. (2020) explicam que este processo significa “saber que existem registos públicos e como obtê-los”, assim como procurar por “números adequados” para serem analisados em folhas de cálculo ou bases de dados. Identificada a fonte, é necessário obter estes dados, existindo para esse efeito vários métodos. Gray & Bounegru (2019: p.37) dão conta de cinco formas diferentes de o fazer: a recolha manual, na qual os dados são inseridos manualmente numa planilha; a extração de informações a partir de PDF’s, convertendo-os em dados estruturados para análise; a busca de documentos junto de fontes humanas ou através da Lei de Acesso à Informação; a extração de dados através da linguagem de programação, utilizando *scripts* ou programas específicos para extrair dados de documentos ou páginas Web; a automação usando as interfaces de programação (API’s) para a recolha direta dos dados a partir das fontes.

Vale a pena notar que estes últimos dois métodos de recolha de dados são, frequentemente, associados às potencialidades computacionais, como o *scraping* (Borges-Rey, 2020: p.930) e as API’s (Bradshaw, 2014). O *web scraping* pode ser entendido como “o método que permite a extração e a reformulação de informações que possam ser utilizadas pelos investigadores” utilizando “pequenos trechos de código, escritos em linguagem de programação como Python, Ruby ou PHP”, enquanto as API’s são interfaces de programação fornecidas pelas próprias bases de dados *online*, páginas e aplicações Web para recolher uma grande quantidade de dados garantindo a transversalidade da sua execução em qualquer tipo de computador físico (Gray, Bounegru & Chambers, 2012; Rosa, 2016: p.125).

A vantagem do *scraping* comparativamente às API's, é que permite executar a recolha e agregação de dados a partir de qualquer página Web, mesmo que não exista uma API integrada para aceder aos “dados em bruto” (Gray, Bounegru & Chambers, 2012). Há, contudo, questões éticas quanto à sua utilização (Howard, 2014: p.76; Lewis et al., 2020: p. 19).

Ojo & Heravi (2017: p.4) acrescentam que a informação das bases de dados, páginas Web ou de documentos pode ser recolhida através das “formas de medição”, como são exemplo os questionários, estudos e a manutenção de registos, em que a intenção principal é a de “gerar dados úteis para análise”. Este tipo de dados é tipicamente disponibilizado por organizações públicas ou privadas, mas raramente são disponibilizados no “formato desejado e necessário” (Gupta et al., 2016).

### **3.8.2 “Limpeza” dos dados**

Para que a informação possa ser analisada e, posteriormente, se produzam as representações visuais, os dados necessitam de ser “tratados” ou “limpos”, o que requer “*software* e inteligência humana” para “padronizar a ortografia, a pontuação e a representação” de forma a obter “contagens exatas” a partir da informação recolhida (Lewis et al., 2020: p.17). Esta fase pode ser entendida como uma introdução à análise dos dados, uma vez que começam a ser excluídos os “dados desnecessários” (Gupta et al., 2016), em que “técnicas”, como a análise e filtragem, são utilizadas para transformar a informação num “formato adequado” e estruturado.

Essa “análise” mais não é do que a “criação de uma estrutura para obter significados e conhecimentos a partir dos dados, ordenando-os em categorias”. Já a “filtragem” diz respeito à seleção dos dados que interessam para o passo subsequente: a análise da informação quantitativa.

### **3.8.3 Análise**

No seguimento da perspetiva de Gupta et al. (2016: p.63), a análise dos dados implica identificar e abordar “questões específicas em torno de um evento, tendência ou problema”, que resultarão na verificação das “hipóteses” formuladas previamente:

“O jornalismo de dados tem como propósito apoiar e provar a hipótese formulada através da evidência dos dados. Por isso, extrair conhecimento a partir dos dados é muito importante. Isto é possível através da transformação dos dados num contexto matemático, aplicando métodos estatísticos e de extração da informação para discernir padrões. Estes métodos estatísticos envolvem a ordenação, médias, variâncias e correlações para ordenar e estabelecer comparações relativas e relacionais entre os elementos” (Gupta et al., 2016: p.63).

Gehrke & Benetti (2020) entendem que a formulação de hipóteses é o “ponto de partida” dos projetos de jornalismo de dados, podendo os dados “confirmar ou refutar” essa hipótese. Todavia, isto exige que o jornalista compreenda “numeracia, estatística e a informação quantitativa” de forma a evitar erros e conseguir “comunicar os dados com precisão” (Lewis et al., 2020: p.18). Além do mais, essas competências permitem que o jornalista de dados desenvolva “pensamento crítico” sobre “como os dados são definidos e obtidos”, e a capacidade de detetar “estatísticas tendenciosas ou manipuladas” – um ponto que será aprofundado mais à frente. Por estas razões, os conhecimentos básicos em numeracia e a estatística descritiva não são opcionais - “são tão essenciais como a ortografia e a gramática” (*Ibidem*).

Lewis et al. (2020: p.17-18) explicam ainda que a análise dos dados pode ser realizada recorrendo a “folhas de cálculo, linguagens de programação, sistemas de gestão de bases de dados ou representações visuais”, que permitirão encontrar “pontos comuns” ou “anomalias” (tradução do inglês “outliers”). Neste contexto, a identificação desses “padrões” ou valores incomuns é um dos principais métodos utilizados para compreender as bases de dados mais complexas (Gehrke & Benetti, 2020).

Sem surpresa, compreende-se que a “manipulação” das bases de dados nesta fase e as ferramentas de trabalho utilizadas exijam “altos níveis de especialização”, que se expandem para “áreas desconhecidas” da maioria dos jornalistas, como a codificação e a programação (Boyles & Meyer, 2017: p.428). A utilização de códigos de programação e modelos estatísticos permitem que o jornalista de dados compreenda rapidamente, através das análises quantitativas aos dados, “um cenário mais amplo” sobre o evento que estão a reportar (Gehrke & Benetti, 2020).

Aliás, a programação pode mesmo ser transversal a todas as fases do fluxo de trabalho: no tratamento dos “dados em bruto” e/ou na transformação da estrutura dos dados (Lindgren, 2021: p.123), assim como na recolha e representação visual dos dados (Lewis et al., 2020: p.17).

Lewis et al. (2020) constataam que a importância da aprendizagem desta competência é controversa: alguns profissionais consideram que as competências na área da programação “são úteis, mas não necessárias para a prática do jornalismo de dados”, enquanto outros “confiam na linguagem de programação R”, existindo ainda quem considere o *Excel* “suficiente”. Esta questão evidencia, uma vez mais, a ambiguidade e dificuldade em estabelecer os limites do jornalismo de dados com outras áreas, como a do jornalismo computacional (*Ibidem*). Para Coddington (2015: p.332-333), embora haja semelhanças na forma de processar os “dados de grande dimensão”, o jornalismo computacional está mais relacionado com a vertente tecnológica que engloba algoritmos e processos de automação da informação. Ainda assim, reconhece que há uma “interação crescente entre programadores e jornalistas”, sendo que estes últimos estão a adotar as “capacidades técnicas e normas culturais da programação” no seu trabalho. Essa adaptação pode ser vista como um aspeto positivo se tivermos em conta que “diferentes linguagens de programação partilham linhagens sobre o modo como os dados podem ser reescritos através de múltiplas formas, estruturas e formatos de ficheiros para fins criativos e consequentes” (Lindgren, 2021: p.151). Ou seja, o código escrito, que é um “recurso textual intermediário” facilitador da “produção de sentido”, poderá ser utilizado em futuros projetos, bastando reescrevê-lo, editá-lo ou adaptá-lo aos propósitos pretendidos pelo jornalista de dados (*Ibidem*).

O estudo conduzido por Rogers, Schwabish & Bowers (2017: p.11-12) mostra que vários jornalistas consideraram a análise dos dados como a fase mais relacionada com as “aptidões especializadas”. No entanto, as competências para a criação da representação visual dos dados são percecionadas como “ainda mais especializadas”.

### 3.8.4 Representação visual dos dados

As representações visuais de dados, ou visualizações de dados (tradução do inglês “data visualization”), são entendidas como uma pedra angular da narrativa das histórias jornalísticas conduzidas por dados (Weber et al., 2018: p.191), constituindo-se como “prática fundamental” que está associada à área do design e dos valores jornalísticos (Coddington, 2015: p. 335-339).

Tabary et al. (2016: p.69) suportam-se nas investigações de Friendly & Denis (2000) e de Michael Friendly (2005) para explicar que o aumento de utilização de dados quantitativos foi despoletado pelo “desenvolvimento do pensamento estatístico” e pelo “aparecimento da representação visual de dados”, que resultam da evolução dessas práticas ao longo dos séculos XVII e XVIII. Historicamente, a visualização estatística da informação, que registou altos e baixos na sua utilização até ao ano de 1960, ganhou uma renovada atenção pelo desenvolvimento e disseminação da tecnologia (*Ibidem*).

À semelhança das infografias, as representações visuais de dados podem ser designadas como “artefactos híbridos multimodais que juntam números, palavras, imagens e design num conjunto coerente” representados através de “gráficos, diagramas, mapas, cronologias ou uma combinação destas” (Weber et al., 2018: p.191-192, 194). No entanto, Cunha (2020: p.531) realça que há um contraste entre as representações visuais de dados e as infografias. Enquanto as infografias são utilizadas para “comunicar uma ou mais mensagens específicas a partir de uma representação gráfica”, a representação visual de dados permite ao leitor “analisar, explorar e descobrir informação”, originando “novas formas de comunicação” através de “gráficos interativos” que usam bases de dados “desde o início até à sua apresentação final”.

Ainda assim, importa notar que as infografias - no contexto jornalístico - são uma prática que antecede o jornalismo de dados, remontando, pelo menos, aos finais do século XIX, sendo também entendidas como “a produção de gráficos, quadros, mapas e outras ilustrações factuais” (Knight, 2015: p.56).

No contexto digital e online, a infografia, “que já havia revolucionado a informação no jornalismo impresso”, teve uma repercussão assinalável por estar “mais próxima do utilizador” com um aspeto “mais atrativo e de simples compreensão” (Pereira, Oliveira & Zamith, 2013: p.781). Este recurso visual pode também ser entendido como uma “ferramenta de análise de dados”, visto que a introdução de ferramentas interativas possibilitou uma “mudança no paradigma da visualização dos dados” e auxilia na compreensão da informação (*Ibidem*).

Mas Córdoba-Cabús & García-Borrego (2020: p.88, *as cited in* Ilisnky, 2012) denotam que as infografias são também distintas das representações gráficas de dados por combinarem “desenhos artísticos e técnicos”, incorporando uma “quantidade menor” de informação. Neste sentido, ainda que haja pontos comuns entre as práticas, consideramos que a perspetiva de Cunha (2020) que realça as “novas formas de comunicação”, os “gráficos interativos” e a sua conceção a partir das bases de dados, aliada a esta última ideia, é a que mais se adequa ao entendimento da representação visual de dados.

Alexandre (2020) afirma também que a “grande novidade” do jornalismo de dados é caracterizada não apenas pela “possibilidade de contar histórias complexas com dados”, mas também por “permitir que o leitor interaja e explore as informações em representações visuais interativas”. Além das infografias, os mapas são também utilizados há muito tempo no Jornalismo para ilustrar certos eventos (McAdams, 2014: p.191). Mais uma vez, a interatividade adotada neste recurso visual permitiu aos jornalistas mostrar “conexões que não poderiam ser reveladas só pela narrativa noticiosa ou pela representação visual de dados complexos” (Usher, 2020: p.253).

Estes recursos de representação visual da informação quantitativa desempenham várias funções na “fase de reportagem” para, por exemplo, “identificar temas e questões” ou “identificar casos anómalos” e também na “fase de publicação” ao ilustrar “de forma mais convincente” o ângulo da história jornalística (Gray, Bounegru & Chambers, 2012). Entende-se, por isso, que este processo não deve ser considerado como “algo a tratar depois de a história estar praticamente escrita”.

As representações visuais de dados servem, portanto, para “tentar comunicar informações complexas que, de outra forma, seriam difíceis de transmitir aos leitores” (Veglis & Maniou, 2018: p.36). Costumam ser representadas através diagramas, mapas ou gráficos que permitem aos leitores “compreender a análise ou personalizar os dados” através de recursos interativos (Lewis et al., 2020: p.17), atestando o argumento de que a “inovação” é mais provável de se verificar nesta fase do fluxo de trabalho (Lowrey et al., 2019b: p.2145).

Na maioria das histórias conduzidas por dados, as representações visuais “complementam a narrativa que é transmitida através do texto”. Mas, em alguns casos, podem assumir o papel central da história, ao passo que o texto, normalmente limitado, “complementa ou explica a visualização” (Veglis & Bratsas, 2017: p.115).

Gupta et al. (2016: p.70) notam que para a comunicação da representação visual dos dados ser eficaz é necessário “escolher o diagrama ou gráfico adequado”. Além dessa sensibilidade, a extração de dados, a programação, escrita e design são uma “combinação de competências necessárias” para criar “representações visuais de dados com alta qualidade” e, por norma, jornalistas e designers não possuem a capacidade de “converter o conteúdo jornalístico numa forma visual”, o que requer “cooperação e fusão de diferentes disciplinas” (Smit et al., 2014: p.351-352). Também Usher (2020) refere que é importante “conhecer a forma como os dados são recolhidos”, uma vez que a sua posterior apresentação e representação visual pode induzir a “falsas suposições”.

Neste seguimento, as soluções tecnológicas de terceiros conseguiram “democratizar a produção noticiosa” dos jornalistas de dados que não são programadores, “sem comprometer a integridade ou a qualidade jornalística das visualizações baseadas em dados” permitindo, assim, que principiantes “criem histórias com dados de forma rápida e fácil” (De-Lima-Santos et al., 2021: p. 157-158).

Alguns exemplos de ferramentas de terceiros para a representação visual de dados são o *Datawrapper* ou o *Flourish*, que fornecem indicações úteis e sofisticadas, embora a sua utilização também acarrete riscos devido às “dependências do fornecedor da ferramenta”, podendo vir a ser descontinuadas pelo fornecedor e “quebrar” a ligação à representação visual publicada (Heravi et al., 2022: p.2085).

Ainda que Lewis et al. (2020: p.18) considerem que os jornalistas não tenham de saber como criar representações gráficas dos dados, sublinham que estes profissionais devem, pelo menos, saber como apresentar a informação.

Este é um ponto importante, visto que os profissionais das redações julgam que os leitores se sentem “mais atraídos” por esta “forma inovadora” de apresentação dos conteúdos (Cunha, 2020: p.547), embora Link et al. (2021: p.590) tenham evidenciado que o esforço aplicado na criação e apresentação de infografias interativas, informação estatística e representações visuais de dados parece não compensar aos níveis da “credibilidade da mensagem” e da “experiência de leitura”. Ainda assim, Alexandre (2020) nota que quando as “visualizações interativas” abordam “assuntos pertinentes” podem “atrair novos e antigos leitores”.

Numa outra perspetiva, Henke et al. (2020: p.313-314) referem que a inclusão de “fontes científicas, informações estatísticas e a representação visual dessas informações” melhoram a compreensão dos leitores e que isso verifica-se, sobretudo, nas representações visuais de dados que facilitam o “processamento da informação”, podendo ser utilizadas para contrariar “a utilização excessiva de dados estatísticos”, tornando a informação “facilmente acessível e compreensível”.

Já no caso das visualizações de dados estáticas, ou seja, desprovidas de interatividade, Paul Mena (2023) sugere que essas podem ter “um efeito na redução das perceções erróneas” em determinados grupos de leitores de notícias.

Há também estudos que sugerem que os jornalistas optam por “experiências visuais simples e minimalistas”, devido ao reduzido contacto (ou conhecimento) que têm com as “funcionalidades interativas”, pelas “pressões do ciclo de produção” noticioso e pela “necessidade de as adaptar para várias plataformas”, como é o caso dos telemóveis que “transformaram os hábitos e as exigências de consumo das audiências” (Anderson & Borges-Rey, 2019: p.1263). Neste sentido, os “formatos visuais estáticos” tendem a tornar-se “mais frequentes” que os “gráficos e mapas clicáveis”.

Esta introdução da interatividade na representação visual de dados, através de *softwares* específicos ou das potencialidades de ferramentas permite que haja “um maior nível de interação” na construção da narrativa e da relação com o público, o que potencia a “hipótese de os leitores se envolverem com as histórias” (Felle, 2016: p.88).

### **3.8.5 Interatividade e formas de narração**

Conforme tem vindo a ser descrito, as fases que constituem a construção de uma história jornalística baseada em dados contêm elementos que podem contribuir para uma “maior vivacidade aos olhos do leitor”, simplificando a compreensão da informação quantitativa e estatística (Henke et al., 2020). Por essa razão, muitos jornalistas empenham-se em apresentar este tipo de histórias através de formatos interativos (Howard, 2014: p.16).

Borges-Rey (2016: p.841) sublinha que “apesar das reivindicações a favor da autoridade jornalística, em detrimento da competência informática”, o jornalismo de dados criou uma disrupção na prática “bastante normativa” do Jornalismo. Isso deveu-se à introdução progressiva do “pensamento computadorizado” na “produtividade e reflexividade dos jornalistas tradicionais”, assistindo-se a uma “substituição progressiva da narração linear por formas mais interativas e envolventes para a experiência do utilizador, na qual se oferecem conteúdos dinâmicos com multicamadas, multiplataformas, e ligados a bases de dados”. Ora, esta mudança acarreta uma “influência substancial” no fluxo de trabalho, pois os jornalistas de dados seguem agora uma “orientação para criar utilidade” que permita à audiência “tirar as suas próprias conclusões e entendimentos” de forma eficaz (Coddington, 2015).

Para uma compreensão mais aprofundada sobre a forma de apresentar as histórias jornalísticas conduzidas por dados no contexto digital, importa, primeiro, compreender que tipos de histórias podem ser criadas a partir da informação quantitativa.

A investigação conduzida por Rogers, Schwabish & Bowers (2017: p.6, 13-14) ao estado do jornalismo de dados, identifica três tipos: 1) as histórias que são enriquecidas pelos dados, em que a informação quantitativa mais relevante é utilizada para verificar a reportagem subjacente, sem que tenham um papel central; 2) as que utilizam os dados para a investigação, onde o jornalista “expõe informação” ou revela “a história escondida nos dados”, normalmente a partir de bases de dados de maior dimensão, o que se pode traduzir numa produção mais morosa dada a necessidade de combinar um conjunto de competências e conhecimentos; 3) e as histórias que “explicam os dados”, nas quais o jornalista utiliza “ferramentas de dados rebuscadas” para “transmitir e reduzir a complexidade do significado por detrás dos números” aos seus leitores.

No que concerne à forma estética e/ou visual da narrativa da história, Duarte (2021: p.138) entende que o episódio desolador do 11 de setembro de 2001, nos Estados Unidos da América, terá despoletado o “dinamismo da interação exploratória” das narrativas visuais no jornalismo *online*. Segundo esta autora, a infografia digital revelou-se a “nova narrativa visual interativa” capaz de apresentar a informação de forma “não linear, num formato hipertextual” e de ser gerida pelo leitor, que se transforma em “utilizador” mediante o seu interesse pelo assunto. Já para Weber et al. (2018: p.200) a principal diferença entre as narrativas textuais e as que integram representações visuais de dados interativas é que as segundas contemplam “potenciais narrativas múltiplas integradas numa única visualização de dados”. Ou seja, “têm a capacidade de incentivar os utilizadores a procurarem outras histórias por si próprios”, constituindo “uma forma moderna de jornalismo dialógico”.

Tomando como ponto de partida a proposta metodológica de análise da autoria de Zamith (2008), sobre o aproveitamento das potencialidades ciberjornalísticas na Internet, podemos argumentar que a interatividade, hipertextualidade e multimedialidade são três aspetos presentes nas peças interativas de jornalismo de dados.

Isso verifica-se uma vez que: 1) na área da interatividade insere-se o “conteúdo jornalístico multimédia interativo” que permite “ao visitante/utilizador interagir, fazendo com que (...) esse conteúdo possa ser diferente de utilizador para utilizador (...)”; 2) na área da hipertextualidade, são introduzidas as “hiperligações extra ou intratextuais” de fontes documentais, cronologias e/ou infografia, e as “hiperligações genéricas”; e 3) na da multimedialidade por abarcar recursos como a fotografia ou desenho, diaporama, infografia estática ou dinâmica, áudio, e os vídeos com ou sem som (*Ibidem*). Todos estes elementos são passíveis de ser identificados numa história de jornalismo de dados, publicada no contexto digital, na qual a informação quantitativa assume “uma função semiótica” (Weber et al., 2018: p. 203).

Para Segel & Heer (2010) a interatividade é parte intrínseca da história jornalística conduzida por dados, pois permite ao leitor utilizar a “narrativa principal” como “ponto de partida” para a exploração personalizada dos dados. No mesmo sentido, Appelgren & Jönsson (2021: p.767) veem a interatividade como “a possibilidade de o utilizador manipular a forma representativa dos objetos e conteúdos, e obter *feedback*”. Ou seja, a interatividade “coloca o controlo nas mãos dos utilizadores”, permitindo-lhes explorar “diferentes opções e combinações” e proporcionar uma “compreensão mais profunda e completa de uma história (...)” (McAdams, 2014: p.191).

A partir da sua pesquisa, Veglis & Bratsas (2017: p.113-117) identificaram três tipos de interatividade aplicados às representações visuais de dados interativas: a “interatividade de transmissão”, que se aplica às representações visuais de dados mais simples permitindo ao utilizador visualizá-la e fornecendo “algumas explicações adicionais” nos elementos que a compõe através de um “pop-up” de informação; a “interatividade consultiva”, que inclui as várias visualizações interativas que permitem a consulta de várias perspetivas dos mesmos dados, possibilitando ao utilizador escolher “uma seleção existente de informações pré-geradas numa comunicação bidirecional”, como num mapa ou cronologia; e a “interatividade de registo”, que se traduz na capacidade da representação visual interativa possibilitar a introdução de dados do utilizador, que resultem em alterações à representação visual em tempo real.

McAdams (2014) realça, no entanto, que a interatividade nas representações visuais de dados só deve ser usada se tornar a história “mais interessante”, de “fácil compreensão” ou permitir ao jornalista “incluir mais informação sem sobrecarregar os leitores” e, por isso, as opções de interatividade devem ser “óbvias, aliciantes e fáceis de utilizar”.

Já Segel & Heer (2010) consideram que existem “muitos graus de interatividade possíveis”, embora os recursos interativos mais comuns sejam os “botões de navegação, destaques e/ou detalhes ao passar o cursor, a filtragem, a pesquisa, a pesquisa detalhada, o zoom (...)”, entre outros. Estes autores introduzem ainda uma questão pertinente: o equilíbrio entre as “histórias orientadas pelo autor” (tradução do inglês “author-driven stories”) e as “histórias orientadas para o leitor” (tradução do inglês “reader-driven stories”). Enquanto a primeira abordagem funciona melhor para “contar histórias” ou realizar “uma comunicação eficiente” e é desprovida de recursos interativos, as histórias orientadas para o leitor apresentam “níveis elevados de interatividade” e incluem “ferramentas de análise visual” que apoiam o “diagnóstico de dados, a descoberta de padrões e a formação de hipóteses”. Da mesma forma que o texto pode “clarificar os elementos visuais”, mas “originar desorganização”, também a interatividade pode funcionar como um “chamariz” para o utilizador, mas “desvirtuar a mensagem pretendida pelo autor”. Nesse sentido, os autores propõem três esquemas narrativos a ser aplicados nas peças jornalísticas que utilizam dados:

- 1) a estrutura do “Martini Glass” (ou copo de Martini), que começa com uma “abordagem orientada pelo autor”, apresentando “perguntas, observações ou artigos escritos para introduzir a visualização”, e, após a introdução textual, a representação visual dos dados segue a “fase orientada para o leitor”, permitindo a exploração interativa dos dados. Esta estrutura “assemelha-se a um copo de Martini, com a haste a representar a narrativa de um único caminho conduzida pelo autor e a boca alargada do copo a representar os caminhos disponíveis e possíveis através das interações conduzidas e orientadas pelo leitor”;

- 2) a apresentação do “*slideshow* interativo”: incorpora a interação a meio da narrativa, sendo mais eficaz para comunicar “bases de dados complexas”. Este modelo permite ao leitor “explorar melhor determinados pontos” antes de avançar para a fase seguinte da história. Contudo, este esquema funciona muitas vezes como a anterior, pois comunica “as mensagens pretendidas pelo autor antes de conduzir o utilizador a interagir com o ecrã”;
- 3) já a “história de pesquisa” é a que segue mais uma orientação para o leitor, apresentando “um tema geral” que permite ao utilizador escolher as categorias específicas do tema para revelar “mais detalhes” ou histórias secundárias. Esta estrutura permite que o utilizador escolha quais histórias são contadas e quando, mas exige “uma quantidade significativa de trabalho de criação para determinar os possíveis tipos de interação (...), as histórias candidatas e os detalhes” (*Ibidem*).

Para Segel & Heer (2010) as histórias conduzidas por dados parecem ser mais eficazes quando a sua interação é limitada, permitindo ao utilizador “explorar os dados sem se afastar demasiado da narrativa pretendida” pelo autor. Por sua vez, Weber et al. (2018: p.196) entendem que “a estrutura narrativa mais simples” passa pela combinação da narrativa textual (história conduzida pelo autor) com as representações visuais de dados. Mas há também novas formas nativas do digital para construir a narrativa visual das histórias conduzidas por dados. Uma delas é o “*scrollytelling*”, em que a história se desenrola à medida que o utilizador percorre a página e surgem representações visuais de dados em determinados momentos (*Ibidem*), e outra é a “*news app*” (Howard, 2014: p.16-17), que se trata de uma aplicação *online* (sem a necessidade de instalar *software*) ou funcionalidade interativa que, mais do que contar uma história, conta “a história do utilizador” através da personalização dos dados, que se aproxima da proposta de “interatividade de registo” avançada por Veglis & Bratsas (2017).

“As *news apps* ajudam a dar sentido a grandes quantidades de dados para pessoas que precisam de compreender um assunto complexo, mas que não possuem a literacia digital para manipular os próprios dados em bruto” (Howard, 2014: p.16-17).

A questão da interatividade no jornalismo *online* é um assunto controverso para certos autores, uma vez que o seu uso poderia ser “mais e melhor contextualizado”, aprofundado e aproveitado (Dick, 2014: p.504; McAdams, 2014: p.191; Appelgren & Jönsson, 2021: p.769), ainda que seja percebida como um aspeto positivo (Link et al., 2021: p.590).

### **3.9 A ética no jornalismo de dados**

Os valores e ideais do jornalismo clássico - como a objetividade, o rigor e a imparcialidade (McQuail, 2010) – são transversais ao jornalismo de dados (Lewis & Westlund, 2015; Tandoc & Oh, 2017; Robles & López, 2021). Todavia, tal como vimos em capítulo anterior, a questão da transparência e da acessibilidade à informação utilizada são aspetos relevantes nesta área.

É difícil refutar a ideia de que a utilização de dados na reportagem jornalística trouxe novos “desafios éticos” (Lewis et al., 2020: p.19) à profissão, sobretudo ao nível do seu fluxo de trabalho (Bradshaw, 2014: p.203). Nesse sentido, a ética e a conduta profissional são aspetos essenciais para os jornalistas; são normas que “conduzem as escolhas e os define enquanto profissionais” (Lewis & Westlund, 2015).

Assim sendo, Lesage & Hackett (2014: p.52) consideram que deve existir “moderação” ao perceber a informação quantitativa (ou os dados) como “fonte de objetividade” que conduz a um “jornalismo mais objetivo e sem vieses” (Sandoval-Martín & La-Rosa, 2018). A objetividade, neste contexto, está relacionada com o “processo de recolha de notícias baseado em factos” (Veglis & Maniou, 2018) e pela utilização de números para transmitir a “sensação de factualidade” ao leitor (Du et al., 2019: p.182) e, com isso, fortalecer a “objetividade jornalística” (Parasie & Dagiral, 2013).

A advertência não é desprovida de sentido, pois vários autores (Parasie & Dagiral, 2013; Lugo-Ocando & Brandão, 2016; Tabary et al., 2016, Cushion et al. 2017; Zamith, 2019; Stalph & Heravi, 2021) mostraram que a principal fonte dos dados utilizados nestas peças é proveniente de organismos institucionais, na sua maioria governamentais. Isso pode constituir um cenário adverso para os meios de comunicação social e a influência que têm sobre o público (Zamith, 2019).

Para Zamith (2019), a “dependência contínua de fontes institucionais” e a “notável raridade da recolha própria de dados” por parte dos meios de comunicação, pode conduzir a uma “maior manipulação” por partes das fontes oficiais e “limitar o sucesso” da função de vigilância do jornalista. Aliás, Veglis & Maniou (2018: p.39) referem que é “prática comum” os funcionários governamentais apresentarem dados favoráveis ao invés daqueles que podem revelar “lacunas (...) nas despesas ou nos investimentos nacionais”.

No mesmo sentido, Lowrey et al. (2019b: p.2145) notam que é “pouco provável” os jornalistas de dados questionarem as agências governamentais sobre a estrutura e a categorização utilizada para construir as bases de dados disponíveis para consulta. Por essa razão, Bradshaw (2014: p.105) sugere que a “consideração ética mais básica” a ter em conta nesta área é a do rigor, sublinhando que o contexto da informação utilizada é, igualmente, “uma parte vital do processo de jornalismo de dados”. A partir destas perspetivas, compreende-se o porquê de Tong & Zuo (2021) argumentarem que “o conceito de objetividade é inaplicável ao jornalismo de dados”, pois as histórias apresentam “apenas uma versão da realidade revelada a partir dos próprios dados”.

Deve existir, por isso, um certo ceticismo (Bradshaw, 2012: p.206), sentido crítico (Lesage & Hackett, 2014; Lowrey et al., 2019a: p.80) e escrutínio (Cushion et al., 2017: p.1213) nos dados usados para reportar acontecimentos, sendo que este é um aspeto “tão importante quanto as competências técnicas e práticas” da profissão (Tong & Zuo, 2021: p. 164-166).

Uma outra preocupação ética identificada tem que ver com os jornalistas que fazem o *scraping* de páginas Web, devendo estar “cientes das implicações legais e éticas da compilação de dados que os autores não pretendiam que fossem agregados” (Lewis et al., 2020: p.19) e por essa prática de recolha massiva de informação poder prejudicar o funcionamento normal da página Web (Bradshaw, 2014). O dilema entre “liberdade” e “responsabilidade” na publicação de dados que contenham informações pessoais e que constituam um risco à privacidade dos cidadãos (Craig et al., 2017), e o rigor na recolha, metodologia, análise e apresentação dos dados (Gray & Bounegru, 2019) são outras das principais questões éticas destacadas.

Por último, importa notar que a questão da colaboração multidisciplinar neste subdomínio jornalístico, pode comportar um “choque cultural” entre os atores de diferentes áreas, uma vez que, por exemplo, os “hackers” e/ou tecnólogos defendem a partilha, abertura e descentralização total dos dados utilizados como um dos seus principais princípios (Bradshaw, 2014), o que pode provocar tensões e alterar a “cultura e estrutura organizacionais” (Stalph, 2020) no seio das redações.

#### **Capítulo 4. Metodologia**

Após esta contextualização teórica, importa explicar a metodologia adotada nesta investigação. Considerando que o estágio realizado no jornal Público incidiu sobre o jornalismo de dados e foi exequível participar, desenvolver e acompanhar a produção e publicação de peças jornalísticas desse tipo com a equipa responsável, assim como realizar análises de dados sobre determinadas temáticas, os resultados dessa experiência constituem importantes contributos para uma compreensão mais aprofundada sobre os fluxos de trabalho e a cultura natural do jornalismo de dados num órgão de comunicação social português.

Para o efeito desta investigação adotou-se o estudo etnográfico, assumindo o papel de participante-observador, e a análise de conteúdo. Para ambos os métodos, definiu-se o período temporal coincidente com a duração do estágio, que decorreu entre o dia 2 de janeiro e 31 de março de 2023. Conforme mencionado, este estudo de caso visa analisar o fluxo de trabalho em jornalismo de dados e o estado final das peças produzidas pela equipa de jornalismo de dados do Público, da redação do Porto.

Atendendo às dimensões que cada questão de investigação comporta, estabeleceu-se um esquema (*Tabela 1*) que explicita o(s) método(s) aplicado(s) a cada uma das questões de investigação a fim de responder às hipóteses supramencionadas.

Tabela 1 – Esquema de explicitação dos métodos aplicados

Questão de investigação	Hipótese	Método(s) aplicado(s)
1. Quais são as competências necessárias para a prática do jornalismo de dados no jornal Público?	H1: As principais aptidões passam por saber consultar bases de dados, recolher essa informação e ter conhecimentos básicos de estatística para a analisar.	Estudo etnográfico
2. Quais as potencialidades que a Web proporciona na execução e apresentação de uma peça jornalística de dados no jornal Público?	H2: Na execução, as principais potencialidades da web neste tipo de peças passam por permitir consultar e trabalhar com ferramentas facilitadoras da recolha e a análise dos dados, enquanto na apresentação incidem sobre a interatividade que a peça proporciona ao leitor, assim como os conteúdos multimédia e/ou representações visuais que a acompanha.	Ao nível da execução: estudo etnográfico;  Ao nível da apresentação: análise de conteúdo.
3. Quais são as principais representações visuais de dados a que o Público recorre nas suas peças para apresentar dados e de que forma são produzidas?	H3: Os principais tipos de representação visual de dados são mapas estáticos e são produzidos a partir de <i>software</i> de terceiros.	Análise quantitativa
4. De que forma a equipa de jornalismo de dados garante a transparência, precisão e imparcialidade dos dados que utiliza?	H4: Refere a metodologia da análise, a fonte e fornece a base de dados utilizada na análise.	Estudo etnográfico e análise de conteúdo.

#### 4.1 O estudo etnográfico

A aplicabilidade do estudo etnográfico é útil para compreender práticas contemporâneas (Anderson, 2015), como é o caso do jornalismo de dados, e trata-se de um método que contribui para a “descoberta de novas realidades” (Palomo et al., 2019: p.1285) nas investigações académicas. A origem etimológica da palavra “etnografia” está relacionada com o dar a conhecer um povo (do grego, *ethnós*) através da escrita (grafia)<sup>3</sup>.

Segundo Cramer & McDevitt (2004), a prática da etnografia implica a utilização exaustiva de anotações sobre os acontecimentos, podendo ser incluídas entrevistas com membros do grupo. Este método exige um “longo período de imersão” de forma a compreender plenamente as ações, pensamentos, sentimentos e atitudes dos membros desse grupo. No caso específico do Jornalismo, o estudo etnográfico pode servir para clarificar o papel e a integração de certos atores nas redações (Ausserhofer et al., 2020), sendo que a sua aplicabilidade passa por “contar uma história o mais intimamente possível do ponto de vista do grupo que está a ser estudado” (Cramer & McDevitt, 2004).

Considerando o desconhecimento de estudos passados que tenham aplicado este método para investigar o fluxo de trabalho do jornalismo de dados em Portugal, optamos por empregar esta abordagem. Tradicionalmente, as designações mais frequentes nos estudos etnográficos são a de observação participante ou não-participante. A observação participante permite que o investigador conheça e interaja com os membros do grupo, o que pode ser bastante frutuoso para a recolha de informações e compreensão dos fluxos da produção jornalística (De-Lima-Santos & Salaverría, 2021). Cramer & McDevitt (2004) avançam com uma proposta na qual o etnógrafo pode envolver-se nas rotinas do grupo estudado em quatro dimensões: enquanto participante integral, participante-observador, observador-participante ou observador integral. Para este estudo, adotamos o papel de participante-observador, uma vez que reconhecemos “o papel de observação do grupo em estudo”, embora participemos “plenamente na cultura ou nas atividades desse grupo” (*Ibidem*).

---

<sup>3</sup> Definição consultada a partir do “Priberam Dicionário” online: <https://dicionario.priberam.org/etnografia> [consultado pela última vez a 17 de julho de 2023].

Atendendo que o estágio foi orientado por um jornalista de dados da referida entidade, foi-lhe comunicado verbalmente a intenção de realizar este estudo etnográfico. Para garantir que a identidade de certos atores permaneça confidencial, optamos por usar termos como "Jornalista A, Jornalista B, (...)" para nos referir a profissionais que não estivessem envolvidos no núcleo de jornalismo de dados, nem ocupassem cargos de editor ou na direção editorial do jornal. Importa referir que a aplicação deste método foi possível devido ao registo diário das principais atividades desenvolvidas e observadas, e das conversas informais com os membros da equipa de jornalismo de dados do Público, que decorreram maioritariamente no interior da redação. No entanto, atendendo ao grau do nosso envolvimento, nem sempre foi possível registar tudo o que foi testemunhado. As anotações do diário de campo foram realizadas, sobretudo, no final de cada dia ou, quando possível, no momento. Os passos seguidos neste estudo etnográfico foram influenciados pela metodologia adotada na tese de doutoramento de Jerónimo (2014), embora com algumas adaptações.

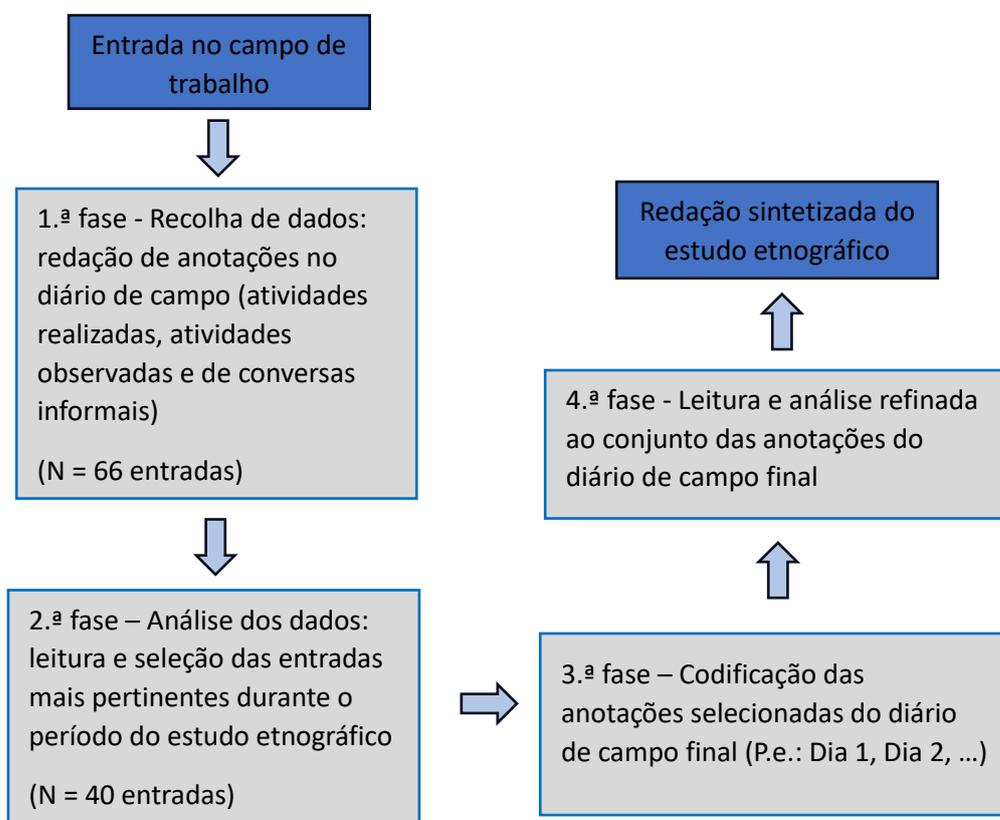
De 2 de janeiro a 31 de março, foram despendidas, no mínimo, 40 horas semanais na redação. No que respeita à observação propriamente dita, esta teve de ser balanceada ao longo do tempo, considerando a necessidade das tarefas que tivemos de executar. Importa clarificar que, enquanto investigador, estávamos já familiarizados com o ambiente e as rotinas de produção de uma redação jornalística e que possuíamos conhecimento prévio de uma das ferramentas que poderia ser necessária usar na produção de peças jornalísticas deste tipo: o *RStudio*.

À semelhança de outros estudos (De-Lima-Santos & Salaverría, 2021), desempenhar um papel ativo neste contexto etnográfico permitiu-nos compreender aprofundadamente os desafios, fluxos e rotinas de produção da equipa de jornalismo de dados. A *Figura 3* ilustra as fases da recolha e de análise dos dados deste estudo etnográfico, começando pela redação diária de anotações sobre as atividades realizadas, dinâmicas observadas e conversas informais, o que resultou num total de 66 entradas reunidas no diário de campo (N = 66).

Posteriormente, as anotações foram relidas e procedeu-se a uma seleção dos dias mais pertinentes, excluindo-se aqueles em que não se verificou qualquer interação relevante com o grupo (dez) e os que não continham qualquer registo (seis), perfazendo um total de 40 entradas e/ou registos diários (N = 40).

De forma a facilitar a pesquisa pelas observações selecionadas, optamos por codificar os dias que correspondem, na realidade, a cada uma das entradas do estudo etnográfico. Ou seja, o “Dia 1” corresponde ao primeiro dia do estudo etnográfico, o “Dia 2” ao segundo dia e assim sucessivamente. Atendendo à extensão do diário de campo final, que pode ser consultado na íntegra na secção “Anexos”, procedemos a uma releitura e análise refinada das observações, dinâmicas e comportamentos destacados sobre este grupo. Esse processo conduziu à redação final do estudo etnográfico, que contempla um texto sucinto e identificativo das situações descritas com base na codificação correspondente de cada observação.

Figura 3 – Fases da recolha e análise de dados do estudo etnográfico



Fonte: Elaboração própria influenciada pela metodologia adotada por de De-Lima-Santos & Salaverría (2021)

## 4.2 Análise de conteúdo

No sentido de enriquecer esta investigação foi também realizada uma análise de conteúdo às peças de jornalismo de dados publicadas pelo jornal Público, entre 2 de janeiro e 31 de março de 2023. Desta forma, podemos investigar aprofundadamente as potencialidades Web presentes no estado final de peças jornalísticas deste tipo, as representações visuais de dados incorporadas e a forma de como é assegurada a transparência, precisão e imparcialidade dos dados utilizados.

A análise de conteúdo é definida por Bardin (2016) como um “conjunto de técnicas de análise das comunicações” que, através de “procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens”, visa obter “indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção” dessas mensagens. A autora entende que a aplicabilidade deste método é muito vasta e adaptável a diferentes contextos, desde que exista “qualquer veículo de significados de um emissor para um recetor”. Importa referir a descrição que Bardin (2016) faz sobre as três fases da organização desta técnica: a pré-análise, a exploração do material, e o tratamento dos resultados obtidos e interpretação. Em suma, a primeira fase (pré-análise) consiste na escolha dos documentos a ser analisados, a formulação das hipóteses e dos objetivos que permitam elaborar os indicadores para fundamentar a interpretação final. A segunda fase passa pela “aplicação sistemática das decisões tomadas” da fase anterior e abrange as “operações de codificação, decomposição ou enumeração”. Por último, o tratamento dos resultados obtidos e interpretação, consiste em transformar os “resultados brutos” em “significativos”, permitindo criar “quadros de resultados, diagramas, figuras e modelos” que sumariam e destaquem as informações resultantes da análise, interpretando-as à luz dos objetivos enunciados (*Ibidem*). Neste sentido, a primeira fase da análise de conteúdo passou pela observação, consulta e recolha da amostra, tendo sido identificadas 10 peças jornalísticas produzidas e/ou atualizadas por algum membro da equipa de jornalismo de dados da redação do Porto do Público (N = 10) e que fossem ao encontro da definição de jornalismo de dados considerada no subcapítulo 3.7 **As definições do jornalismo de dados.**

O corpo da amostra pode ser consultado em *Anexo 2 – Corpo da amostra*, na secção “Anexos”, onde constam: código da peça atribuído a cada um dos artigos, hiperligação das peças de jornalismo de dados do Público, autor(es) da peça, título da peça, a data de publicação e/ou atualização e respetivo título da notícia. A codificação das peças segue a ordem cronológica da sua primeira publicação.

Procedeu-se a uma análise de conteúdo qualitativa no sentido de apurar quais as potencialidades Web utilizadas ao nível da apresentação da peça, e como a equipa de jornalismo de dados do Público garante a transparência, precisão e imparcialidade dos dados que utiliza nestas peças. A análise quantitativa foi adotada para testar se “H3: os principais tipos de representação visual de dados são mapas estáticos e são produzidos a partir de software de terceiros”.

De forma a analisar as potencialidades Web na apresentação de uma peça de jornalismo de dados aos seus leitores, foi estabelecida uma tabela de análise que agrupa cinco potencialidades ciberjornalísticas identificadas na revisão de literatura, estabelecendo um modelo híbrido de análise às unidades semânticas extraídas e de observações face às categorias pré-selecionadas. As potencialidades consideradas para esta fase são: hipertextualidade e multimedialidade (Zamith, 2008); recursos interativos nas representações visuais de dados (Segel & Heer, 2010); o tipo de interatividade presente nas representações visuais de dados (Veglis & Bratsas, 2017); a dos esquemas narrativos adotados, que foi subdividida na dimensão da orientação da narrativa da história (Segel & Heer, 2010) e da estética visual da narrativa (Howard, 2014; Weber et al. 2018). Foi ainda adicionada uma secção dedicada a “outras potencialidades exploradas”, a fim de incluir algum elemento que não tenha sido identificado previamente. A seleção por estas categorias deve-se ao facto de as representações visuais de dados serem um elemento central do jornalismo de dados (Weber et al., 2018; Coddington, 2015) e comportarem recursos interativos com diferentes níveis de interatividade (Segel & Heer, 2010; McAdams, 2014), e também por os esquemas narrativos serem considerados elementos das potencialidades Web (Howard, 2014; Borges-Rey, 2016; Veglis & Bratsas, 2017). No total, são 24 as categorias conceptuais de análise, conforme ilustrado na *Tabela 2*.

Relativamente à análise quantitativa utilizada para avaliar o tipo predominante de representações visuais de dados e a ferramenta de criação ou autoria das mesmas, registamos para cada peça da amostra o tipo específico de representação visual empregado, distinguindo entre formatos estáticos e interativos (como mapas, gráficos de linha, gráficos de barras, entre outros), o total de representações visuais de dados estáticas e interativas por peça e a ferramenta de criação ou a autoria das representações visuais de dados.

Tabela 2 – Categorização de análise às potencialidades Web numa peça de jornalismo de dados

<b>Potencialidades Web na apresentação de uma peça jornalística de dados</b>		
<b>Potencialidades Web</b>	<b>Dimensão</b>	<b>Subdimensão</b>
1.1- Hipertextualidade	1.1.1 - Hiperligações intratextuais	-
	1.1.2 - Hiperligações extratextuais	-
	1.1.3 - Hiperligações genéricas	-
1.2 - Multimedialidade	1.2.1 - Fotografia ou desenho	-
	1.2.2 - Diaporama	-
	1.2.3 - Infografia/gráfico/mapa estático	-
	1.2.4 - Infografia/gráfico/mapa dinâmico	-
	1.2.5 - Áudio	-
	1.2.6 - Vídeo com som	-
	1.2.7 - Vídeo sem som	-
1.3 Outras potencialidades exploradas		
1.4 - Recursos interativos nas representações visuais de dados	1.4.1 - Botões de navegação	-
	1.4.2 - Destaques ou detalhes ao passar o cursor	-
	1.4.3 - Filtragem	-
	1.4.4 - Pesquisa ou pesquisa detalhada	-
	1.4.5 - Zoom	-
1.5- Tipo de interatividade nas representações visuais de dados	1.5.1 - Interatividade de transmissão	-
	1.5.2 - Interatividade consultiva	-
	1.5.3 - Interatividade de registo	-
1.6 - Esquemas narrativos	1.6.1 - Orientações narrativas da história	1.6.1.1 - Estrutura Martini Glass
		1.6.1. 2 - Slideshow interativo
		1.6.1.3 - História de pesquisa
	1.6.2 – Estética visual da narrativa	1.6.2.1 - Scrollytelling
		1.6.2.2 - News app

Por último, adotamos a análise qualitativa para examinar como é assegurada a transparência, precisão e imparcialidade dos dados nestas peças da amostra. Para tal, estabelecemos a seguinte categorização de análise a partir do subcapítulo 3.6 **“Abertura” e transparência dos dados** da revisão da literatura: descrição da metodologia ou referência a dados; referência explícita à fonte dos dados e a disponibilização da base de dados utilizada na análise (Lesage & Hackett, 2014; Chadha & Koliska, 2015; Tandoc & Oh, 2017; Weber et al., 2018; Sandoval-Martín & La-Rosa, 2018).

Tabela 3 – Exemplo do processo de análise de conteúdo sobre a transparência, precisão e imparcialidade dos dados utilizados

<b>Código da peça</b>	<b>Descrição da metodologia/referência a dados</b>	<b>Fonte dos dados</b>	<b>Base de dados utilizada na análise</b>
<b>Peça 1</b>	" (...) O cabaz de produtos da Deco Proteste é definido com base na estrutura de consumo das famílias do INE. Este conjunto de 63 produtos inclui, por exemplo, leite, queijo, manteiga e fiambre, um pacote de arroz, de farinha, de massa e de açúcar, além de 1 kg de carne de peru e a mesma quantidade de frango, carapau, pescada, cebola, batata, cenoura, banana, maçã ou laranja. É calculado o valor médio por produto em todas as lojas de Portugal onde está disponível, obtendo assim o valor médio daquele produto em dado dia. Para cada dia analisado, é calculado o valor total do cabaz (com base no valor médio por produto)."	Identificação no primeiro gráfico: "Fonte: DECO Proteste"; Identificação no final do artigo: "Em conjunto com: DECO Proteste".	Não foi identificada

Através da conjugação destes métodos, procuramos dissipar as eventuais fraquezas de um estudo etnográfico de observação participante. Algumas investigações anteriores já utilizaram o estudo etnográfico e de entrevistas com jornalistas de dados, bem como a análise de conteúdo a peças de jornalismo de dados submetidas a competição (Zamith, 2019), mas não identificamos nenhum estudo de caso que tenha usado ambos para a investigação das práticas diárias e da produção de conteúdos do jornalismo de dados em Portugal.

## **Capítulo 5. Resultados**

Neste capítulo será explanado, em primeiro plano, o estudo etnográfico de observação participante realizado à equipa de jornalismo de dados do Público.

As entradas do diário de bordo utilizado para descrever e investigar as rotinas de produção podem ser encontradas na secção “Anexos”, em “Anexo 1 – Diário de campo”. De seguida, são apresentados os resultados da análise de conteúdo às peças de jornalismo de dados produzidas ao longo do período temporal definido.

A redação continua a ser o espaço de eleição para a produção noticiosa, ainda que os meios tecnológicos permitam hoje que esse processo possa ocorrer de forma remota ou no exterior. Traquina (2007: p.126) afirma que as notícias “são o resultado de processos de interação social” que ocorrem não só com as fontes dos jornalistas, mas também entre os próprios jornalistas. Entende-se, portanto, que este “espaço comum” dos jornalistas – a redação – potencie a “troca de experiências, saberes, truques, anedotas”, uma vez que “faz parte de um processo decisivo de formação de consenso” e do espírito de camaradagem que influencia o trabalho jornalístico (*Ibidem*). Interessa-nos, por isso, investigar os fluxos de trabalho em jornalismo de dados em Portugal, procurando aferir como estas notícias são construídas na redação e os fatores que influenciam a sua produção. Para esse efeito, tomamos como estudo de caso o jornal Público que, anualmente, tem sido distinguido pelo Observatório do Ciberjornalismo<sup>4</sup> com algumas menções a trabalhos de jornalismo de dados.

### **5.1 A observação ao jornalismo de dados no Público**

A unidade de jornalismo de dados do Público é composta por dois jornalistas: Rui Barros e José Volta e Pinto. Ambos trabalham na redação localizada no Porto e fazem parte da secção “Online” do jornal. Assumem um modelo de trabalho híbrido, em que o trabalho presencial e o remoto é intercalado mediante as necessidades da empresa, dos colaboradores e a escala mensal do jornal. De acordo com os registos apurados numa consulta à página Web do Público, a produção de conteúdos por parte destes dois profissionais em jornalismo de dados começou em março de 2020, quando a pandemia de Covid-19 ainda vigorava. Até à data deste estudo, encontravam-se no ativo os mesmos membros.

---

<sup>4</sup> As distinções anuais do Observatório do Ciberjornalismo podem ser consultadas em: <https://obciber.wordpress.com> [consultado pela última vez a 20 de julho de 2023]

Este pequeno núcleo dedica-se, sobretudo, à produção de peças jornalísticas para a página Web do *Público*, sendo que a informação numérica é o “fio condutor” das suas histórias, podendo as mesmas ser adaptadas no formato impresso. A unidade de jornalismo de dados não é completamente autónoma na sua produção, estando subordinada aos editores de diferentes secções, sobretudo, a do *Online* e de Economia, onde são discutidos os temas, as perspetivas e diferentes ângulos da história jornalística.

Rui Barros é o jornalista do *Público* que assume o maior protagonismo no jornalismo de dados, contando com diversos trabalhos assinados em nome próprio<sup>5</sup>. Esta observação foi também confirmada pelo então diretor do jornal, Manuel Carvalho, assim como por outros colegas da redação durante conversas informais.

José Volta e Pinto também realiza peças de jornalismo de dados, mas assume um papel menos ativo comparativamente ao seu colega. Uma das razões que podem justificar o seu menor envolvimento tem que ver com o papel que desempenha na cobertura noticiosa diária *online* do jornal. Neste ponto, importa realçar que outros colegas – jornalistas, infografistas e especialistas – também colaboram, em diversas ocasiões, com a equipa de jornalismo de dados, podendo os mesmos se encontrar na redação do Porto ou de Lisboa.

### **5.1.1 A redação**

A redação do *Público* no Porto está localizada na Rua Júlio Dinis, nº 270, Bloco A, no terceiro piso do edifício. Este piso é ocupado por um outro órgão de comunicação social (*Rádio Nova*<sup>6</sup>), que opera em instalações separadas, sendo que a receção e o refeitório são os únicos espaços físicos partilhados entre as entidades. Além disso, um estúdio de gravação também é utilizado por ambas, embora esse fosse o espaço predileto do *Rádio Nova*. O acesso à redação no Porto só é possível de se efetuar depois de a pessoa responsável da secretaria atender no intercomunicador do prédio.

---

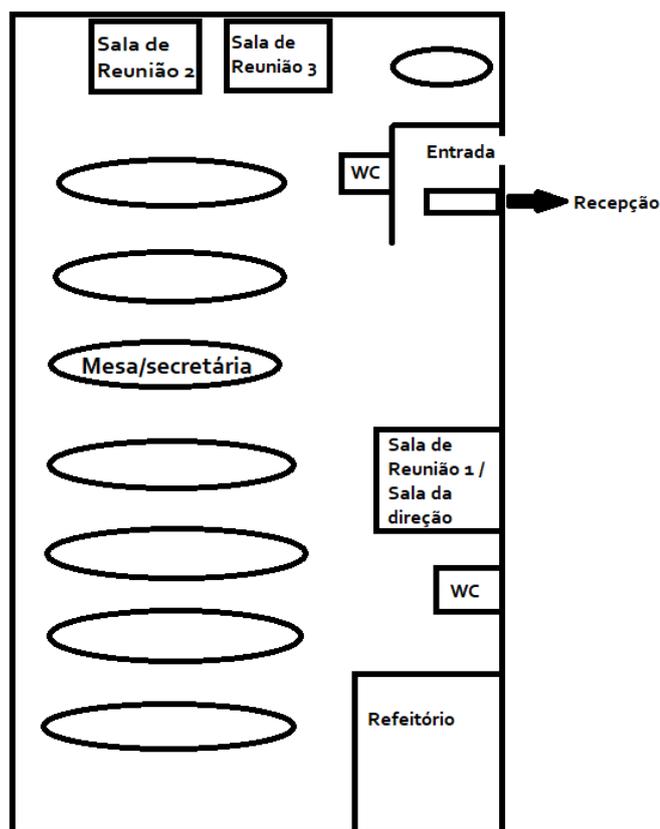
<sup>5</sup> Os trabalhos do jornalista Rui Barros no *Público* podem ser consultados em: <https://www.publico.pt/autor/rui-barros> [consultado pela última vez a 20 de julho de 2023]

<sup>6</sup> Mais informações sobre a *Rádio Nova* e a sua localização podem ser consultadas em: <http://www.radionova.fm/info> [consultado pela última vez a 20 de julho de 2023]

Nesse aparelho, encontra-se um pequeno autocolante com o logótipo do jornal, com a letra “P” a vermelho, identificando facilmente o botão correspondente. Nos primeiros dias, a nossa identificação foi feita vocalmente, dando conta do nosso propósito. Posteriormente, a pessoa responsável passou a reconhecer-nos pela câmara do intercomunicador. Depois de sermos identificados, a porta principal do prédio é desbloqueada e dirigimo-nos até a um elevador que nos leva até ao terceiro piso.

O espaço da redação é amplo, segue uma geometria retangular e conta com bastante iluminação, quer pela luz natural ou artificial. É dividida em três zonas principais: a da redação, a sala da direção, que podia também ser utilizada como sala de reuniões, e o refeitório.

Figura 4 – Redação do *Público* no Porto



Legenda: esboço da planta da redação

A redação é composta por sete secretárias de grande dimensão, espaçadas entre si, e os jornalistas distribuem-se em torno delas.

Na medida do possível, estas secretárias – com cerca de oito metros de comprimento - eram organizadas pelas respetivas editorias ou também conhecidas como “ilhas” (Rui Barros, conversa informal, 2 de janeiro de 2023).

“Ainda não estão aqui muitos jornalistas, mas para perceberes um pouco a dinâmica vou-te mostrar como estamos distribuídos. Aquela mesa ali do fundo é a ‘ilha’ do Desporto, depois temos a da Cultura e da Fugas, segue-se a do ‘P3’, e temos aqui a do Online que a direção também usa às vezes, como podes ver. Depois esta secretária é a ‘ilha’ da Sociedade, naquela ali estão os jornalistas de Economia e da Política e, por último, temos o ‘Azul’ e é onde também costumam estar os fotojornalistas. Ah, aquela mesa ali do fundo, ao lado destas salas de reuniões, é onde estão os comerciais e o pessoal da informática”. (Rui Barros, conversa informal, 2 de janeiro de 2023, 10h15).

Os jornalistas trabalham lado-a-lado, podendo ter colegas atrás ou diante de si. A ampla dimensão do espaço para acomodar os jornalistas e a iluminação são dois aspetos que sobressaem numa primeira observação. Contudo, no decorrer da investigação, foi possível averiguar que, quando a maioria dos jornalistas se encontrava na redação, sobretudo às quartas e quintas-feiras à tarde, o espaço parecia ser mais diminuto. A isso, acresce o facto de que, por várias ocasiões, tivemos de mudar de local de trabalho e deslocarmo-nos entre “ilhas” até encontrar um lugar vago, pois existia uma planta que identificava o respetivo lugar de cada jornalista. Esta medida pretendia contribuir para uma melhor comunicação entre cada editoria do jornal. Este aspeto, em certa medida, limitou o nosso espectro de observação.

Além disso, existiam cerca de seis televisões, colocadas a meia-altura, e uma outra no fundo da redação, junto à secção do Desporto. Por norma, estas televisões eram utilizadas para acompanhar o noticiário ou os acessos à página Web do jornal, em que era possível verificar um *ranking* dos artigos mais visualizados e partilhados daquele dia.

### **5.1.2 Meios disponibilizados**

Os meios disponibilizados na redação eram bastante satisfatórios. Computadores fixos, ecrãs, ratos, teclados estavam disponíveis em todas as secretárias para a utilização de qualquer colaborador.

Ainda assim, foi possível observar que uma vasta maioria dos jornalistas opta por utilizar o computador portátil, pois tratava-se de um dispositivo tecnologicamente mais desenvolvido, leve, prático e que poderia ser transportado para qualquer local da redação. O caso não era diferente na equipa de jornalismo de dados. Aliás, foi-nos sugerido que levássemos o computador portátil pessoal, uma vez que seria necessário instalar certos *softwares* que os computadores fixos poderiam não suportar devido às especificidades dos programas. Os dois membros da equipa de jornalismo de dados utilizavam igualmente um computador portátil para trabalhar e aparelhos periféricos próprios, como teclado, rato e auscultadores. O único aparelho periférico a que recorriam na redação era a um monitor, que funcionava como o ecrã principal de trabalho.

No nosso primeiro dia, a ligação à Internet do jornal foi prontamente estabelecida, tendo sido instalada uma “Virtual Private Network” (VPN) para assegurar uma conexão segura. Foi também criado um endereço eletrónico para uso próprio através dos serviços informáticos do jornal. Das observações e contacto com outros jornalistas, foi possível perceber que os custos de transporte e, em certos casos, de alojamento, para efeitos do trabalho jornalístico, eram suportados pela empresa, não existindo qualquer distinção entre estagiários e jornalistas.

### **5.1.3 Atores do jornalismo de dados**

Conforme mencionado, o principal jornalista da secção de jornalismo de dados do Público é Rui Barros. Ele próprio se identifica como sendo um “jornalista de dados”. Importa, contudo, notar que a hibridez constatada nesta área pode levar a que estes profissionais se denominem com outros termos - como “jornalista programador” ou “jornalista informático” (Parasie & Dagiral, 2013; Tabary et al., 2016; Hermida & Young, 2017; Gray & Bounegru, 2019).

José Volta e Pinto é o segundo jornalista alocado à equipa de jornalismo de dados, embora não se tenha denominado como “jornalista de dados” em qualquer ocasião, dado que a sua produção neste domínio era significativamente mais diminuta e menos frequente do que a de Rui Barros.

Atendendo que as peças de jornalismo de dados eram, sobretudo, publicadas na página Web do Público, esta equipa estava enquadrada na secção *Online* do jornal. À data, os editores desta secção, que se encontravam distribuídos entre a redação do Porto e de Lisboa, eram: Ivo Neto, Pedro Rios, Mariana Adam, Patrícia Jesus, Ana Maria Henriques e Pedro Esteves. A equipa do *Online* era composta por 16 elementos identificados na ficha técnica pelas funções que desempenhavam: Inês Chaíça (última hora), Miguel Dantas (última hora), Filipa Almeida Mendes (última hora), José Volta e Pinto (última hora), Sofia Neves (última hora), Aline Flor (áudio), Ruben Martins (áudio), Carolina Pescada (multimédia), Carlos Costa (multimédia), Teresa Pacheco Miranda (multimédia), Joana Bougard (editora multimédia), Joana Gonçalves (multimédia), Pedro Guerreiro (editor de redes sociais), Patrícia Campos, Lucas Freitas, Rui Barros (jornalista de dados).

De entre estes profissionais, somente dois (José Volta e Pinto e Rui Barros) nos foram apresentados como membros do núcleo de jornalismo de dados. Além disso, existia também a equipa da Infografia que, ocasionalmente, colaborava e auxiliava este grupo e era composta por Célia Rodrigues (coordenadora), Cátia Mendonça, José Alves, Francisco Lopes e Gabriela Pedro. Importa clarificar que, desde o primeiro dia, fomos integrados na “unidade de jornalismo de dados” do jornal, enquanto jornalista estagiário.

Normalmente, o jornalista Rui Barros encontrava-se na redação em três ou quatro dias da semana. Às terças e quintas-feiras de manhã, não estava presente na redação, pois lecionava numa instituição de Ensino Superior na região do Porto, sendo que, nesses dias, dava entrada na redação por volta das 13h45. Este jornalista era um dos poucos que, na redação do Porto, possuía um horário flexível, não existindo qualquer referência à sua hora de entrada e/ou de saída na escala do jornal, que era partilhada através de uma planilha do *Google Sheets* com a restante redação.

Ainda que se tenha verificado esta flexibilidade no regime laboral, este jornalista costumava entrar ao serviço por volta das 09h45. Quanto ao horário de saída, era bastante incerto.

Durante a realização deste estudo, foi possível observar que a sua saída estava dependente dos trabalhos que tivesse em curso, podendo até, ocasionalmente, trabalhar a partir de casa ou na redação depois do horário de expediente nesse(s) projeto(s). O regime híbrido era adotado várias vezes por este profissional, intercalando dessa forma o trabalho presencial com o teletrabalho.

Quem também seguia um regime híbrido era José Volta e Pinto. No entanto, este jornalista cumpria com um turno estabelecido na escala e as suas horas de entrada e saída poderiam variar, tendo cumprido diferentes horários durante o período observado, tais como: das 11h00 às 21h00, das 14h00 às 22h00 ou das 16h00 à 00h00. Quanto à equipa de Infografia, não nos foi possível aferir os seus horários de entrada e saída, nem o regime de trabalho que adotavam, uma vez que a vasta maioria destes profissionais se encontrava na redação de Lisboa.

#### **5.1.4 A equipa de jornalismo de dados**

Inicialmente, o jornalista Rui Barros apresentou-nos o núcleo de jornalismo de dados através de duas aplicações digitais e *online*: o *Notion* e o *Discord*. O *Notion* trata-se de uma aplicação *online* semelhante a uma agenda física<sup>7</sup>. Aqui, a equipa de jornalismo de dados do *Público* geria e organizava os seus trabalhos, ideias de futuros projetos, instruções, boas práticas e as suas principais formas de trabalho. A página criada a partir deste aplicativo (*Figura 5*) denominava-se de “PÚBLICO Data Unit” e era composta por cinco elementos: Rui Barros (jornalista de dados e criador da página), José Volta e Pinto (jornalista), Rafael Oliveira (investigador e estagiário), Ivo Neto (editor da secção *Online*) e Pedro Rios (editor da secção *Online*). O acesso a esta página só era possível através de convite.

---

<sup>7</sup> Uma explicação mais detalhada sobre a aplicação “Notion” pode ser encontrada em: <https://www.notion.so/pt-br/product> [consultado pela última vez a 22 de julho de 2023]

Na página da aplicação estavam disponíveis para consulta quatro secções: 1) a dos “Links úteis”, que incluía as subpáginas “Projetos” (ideias e/ou projetos que estavam a ser desenvolvidos), “Trabalhos realizados” (base de dados dos trabalhos publicados no *Público*), “Design Kit do Público” (um pequeno manual de princípios elementares do sistema de design do PÚBLICO), um ficheiro PDF intitulado “Cores Público – Infografia.pdf”, que continha os códigos hexadecimais de cores utilizadas para a criação de gráficos e/ou infografias, e “Tangram” (manual com lista de componentes de código para a construção de peças a ser publicadas na página Web do jornal).

A segunda secção, intitulava-se de “Código”. Trata-se de uma hiperligação para aceder à página do *GitHub* do “*Público Data Unit*”, onde se encontram repositórios de análises de dados de trabalhos realizados e/ou por publicar e outros ficheiros úteis, como alguns códigos-fonte utilizados, documentos com dados geoespaciais para a criação de mapas, entre outros. Em suma, o *GitHub* trata-se de “uma plataforma de alojamento de código para controlo de versões [de projetos] e colaboração”, permitindo aos utilizadores “trabalhar em conjunto a partir de qualquer lugar” (Github, s/d). O terceiro segmento era relativo às “Tarefas”, onde era possível apontar as atividades necessárias a realizar e, por último, o “Calendário” que se repartia em duas subsecções – uma que continha a hiperligação com a escala de horários do jornal, e a de “repositórios equipas de dados”, onde se encontrava a hiperligação do *GitHub* de outros órgão de comunicação social, nomeadamente da *Reuters*, *Los Angeles Times* e *The Pudding*.

Figura 5 – Captura de ecrã da página da equipa de jornalismo de dados no *Notion*



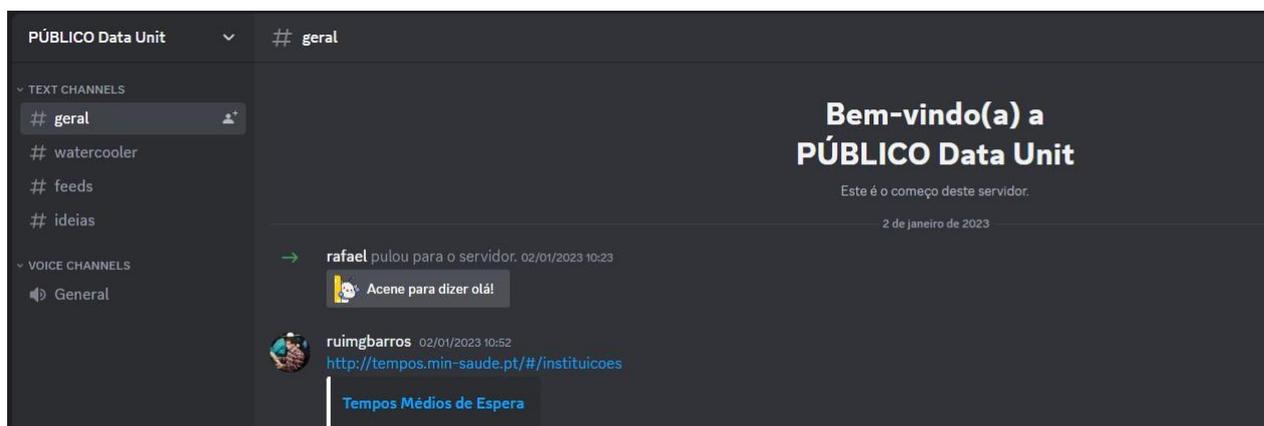
Legenda: Página inicial na aplicação “Notion” da equipa de jornalismo de dados do Público.

Por sua vez, o *Discord*, que se trata de um canal de comunicação textual ou por voz<sup>8</sup>, servia para a equipa comunicar mais facilmente entre si e partilhar conteúdos que pudessem ser úteis para a criação de trabalhos. O canal de *Discord* (*Figura 6*) estava organizado em duas partes: os “canais textuais” e o “canal por voz”. No total, existiam quatro canais textuais: o “#geral”, que servia para a comunicação textual sobre qualquer assunto relacionado com o trabalho, o “#watercooler” para o convívio informal da equipa; o “#feed” que contemplava uma atualização semanal e mensal de boletins informativos sobre jornalismo de dados; e a secção “#ideias”, onde eram partilhados trabalhos de jornalismo de dados de outros órgãos e ideias de futuros projetos. Existia somente um canal para a comunicação vocal, sendo o mesmo suficiente.

<sup>8</sup> Uma explicação detalhada sobre o “Discord” pode ser encontrada em: “O que é o Discord?” - <https://discord.com/safety/360044149331-what-is-discord> [consultado pela última vez a 22 de julho de 2023]

Este grupo era composto por Rui Barros (criador do canal), Ivo Neto (editor da secção *Online*), José Volta e Pinto e por nós, onde cada um poderia assumir um nome de utilizador diferente, mas que era facilmente associável à pessoa em questão. Tal como no *Notion*, o acesso a este canal privado só é possível através do convite do criador.

Figura 6 – Captura de ecrã do canal de *Discord* da equipa de jornalismo de dados



#### 5.1.4.1 Percursos

Sem margem para surpresas, o núcleo de jornalismo de dados do Público é composto na sua totalidade por jornalistas com formação no Ensino Superior. Rui Barros e José Volta e Pinto são licenciados em Ciências da Comunicação, pela Universidade do Minho e pela Faculdade de Letras da Universidade do Porto, respetivamente, na vertente de Jornalismo. Durante as conversas informais, Rui Barros explicou que iniciou um ciclo de estudos no Mestrado em Ciências da Comunicação em 2016, mas optou por congelar a sua matrícula no segundo ano curricular devido à incompatibilidade de horários. Contudo, este jornalista teve um percurso escolar “diferenciado” considerando as várias áreas que aprendeu (Gehlen, 2019), e que nos acabou por confirmar durante as conversas informais.

“No [Ensino] Secundário, frequentei o curso de Ciências e Tecnologias. Tinha boas notas a Matemática e, quando terminei o 12.º ano, os meus pais queriam que seguisse algo relacionado com a área da Engenharia, mas não tinha muito interesse nisso. Sempre gostei de jornais, de notícias e quando chegou o momento de escolher, disse-lhes que ia seguir Jornalismo” (Rui Barros, conversa informal, 31 de março de 2023, [Dia 66](#)).

Ao encontrar-se neste limbo entre as ciências exatas e as sociais, o seu percurso académico desvendou-lhe a fórmula que lhe permitiria fundir as duas áreas do seu interesse. Numa das aulas que frequentou, descobriu a utilidade e aplicabilidade que a matemática poderia ter no jornalismo. Foi por essa altura que descobriu projetos como o blogue de jornalismo de dados do *The Guardian* e o *FiveThirtyEight* que lhe abriram novos horizontes profissionais.

“Safava-me bem com os números, mas não era uma área que me cativasse muito porque não entendia bem qual era a utilidade de algumas operações tão complexas para o meu dia-a-dia. Acho que isso é um dos problemas com a matemática. Na universidade é que isso me foi explicado de melhor forma, principalmente na área da Estatística. Depois partiu de mim descobrir e aprender fazer mais coisas. Comecei por tentar perceber como é que certas peças de jornalismo de dados eram feitas e a replicar o que faziam em alguns trabalhos, sempre tive essa curiosidade. Com o tempo, fui aperfeiçoando certos aspetos e adquirindo mais experiência. (...)” (Rui Barros, conversa informal, 31 de março, [Dia 66](#)).

O seu percurso e aproveitamento escolar, em áreas como a Matemática e Estatística, permitiram-lhe desenvolver aptidões e conhecimentos em outras áreas, como a da representação visual de dados e da programação. Ainda antes de ingressar na Rádio Renascença, este profissional foi colaborador no jornal universitário “ComUM Online” e, numa breve pesquisa à sua página Web<sup>9</sup> pessoal, podemos constatar que durante a licenciatura também participou na construção do website deste jornal universitário.

Rui Barros esteve na secção *Online* da Rádio Renascença como estagiário durante três meses, tendo depois sido convidado a ingressar a redação com um vínculo contratual nessa empresa. Durante cerca de três anos, foi o jornalista de dados daquele órgão, onde publicou vários trabalhos de jornalismo de dados que ainda hoje são possíveis de consultar<sup>10</sup>. À semelhança do que o estudo de Porlezza & Splendore (2019: p.1245) indica sobre as pessoas interessadas em tornar-se jornalistas de dados, Rui Barros aprendeu novos conhecimentos e técnicas de forma autodidata no desenvolvimento de projetos inovadores nesta área. Em fevereiro de 2020, deu entrada no jornal Público.

---

<sup>9</sup> A página Web pessoal de Rui Barros está disponível em: <https://ruimgbarros.com/projects/> [consultado pela última vez a 20 de julho de 2023]

<sup>10</sup> Os trabalhos do jornalista Rui Barros na Rádio Renascença podem ser consultados em: <https://rr.sapo.pt/artigo/80191/ru-barros> [consultado pela última vez a 24 de julho de 2023]

Por sua vez, José Volta e Pinto, de 25 anos, apresenta um percurso académico idêntico, mas com algumas diferenças. Durante a sua licenciatura, colaborou nos jornais universitários “JornalismoPortoNet” e “Jornal Universitário do Porto”, tendo assumido neste último as funções de editor da secção de Desporto. No decurso do ano da conclusão da licenciatura, estagiou durante três meses no *Público*, tendo sido integrado na redação também em fevereiro de 2020. Durante as conversas informais foi possível constatar que este jornalista teve, pelo menos, uma unidade curricular durante a sua licenciatura que abrangeu “noções básicas” de programação. Além disso, tem vindo a frequentar alguns cursos *online* de jornalismo de dados a fim de melhorar as suas aptidões na área.

“– Olá, posso sentar-me aqui para almoçar? (José Volta e Pinto)

- Olá, claro que sim. Tu és o outro colega da equipa de jornalismo de dados, não é? (Rafael Oliveira).

– Sim, sou o Volta! Tu deves ser o Rafael, certo? O Rui [Barros] disse-me que tu escolheste fazer um estágio em jornalismo de dados aqui. É verdade? (José Volta e Pinto).

- Sim, acho que é uma área interessante. Nunca tive a oportunidade de a explorar muito, embora já seja jornalista há algum tempo. Decidi aproveitar esta ocasião para aprender algo mais. (Rafael Oliveira)

- Claro, fizeste bem. Acho interessante teres escolhido esta área. Por acaso acho que nunca tivemos um estagiário em jornalismo de dados por aqui. (José Volta e Pinto)

- Pois, acredito. Diria que é uma área relativamente recente. Tu tiveste alguma unidade curricular na licenciatura sobre jornalismo de dados? (Rafael Oliveira)

- Concretamente em jornalismo de dados, não. Tivemos uma cadeira ou outra que incluía algumas noções na área da linguagem da programação e marcação, como o HTML e CSS. O resto da minha ‘formação’ no jornalismo de dados tem sido através de alguns cursos *online* que tenho frequentado sozinho ou com o Rui”. (José Volta e Pinto, conversa informal, 19 de janeiro de 2023, [Dia 14](#)).

Considerando o percurso escolar e académico destes dois jornalistas, pudemos verificar que a equipa de jornalismo de dados do *Público* é composta por dois profissionais recém-licenciados (Boyles & Meyer, 2017: p.434), munidos de conhecimento nas áreas da programação, matemática e estatística. Além disso, continuaram a apostar na sua formação na área do jornalismo de dados, de forma autodidata ou pela frequência em cursos *online*.

#### 5.1.4.2 Conceitos construídos

No que concerne à percepção sobre o que é o jornalismo de dados, à exceção da própria equipa, da direção do jornal e de alguns editores, poucos jornalistas da redação do Porto revelaram compreender ou interessar-se sobre subdomínio jornalístico. Nas várias interpelações desenvolvidas e observadas, os jornalistas, nomeadamente mais jovens, mostravam-se curiosos sobre o tema. Contudo, após uma breve explicação sobre o conceito, as técnicas empregues e a alusão às evidências numéricas, o interesse parecia desvanecer. Esta hipotética, mas repentina, apatia sobre o assunto foi já observado noutros estudos que evidenciaram o desconforto e até receio da numeracia por parte de jovens alunos e jornalistas (De Mayer et al., 2015; Lewis et al., 2020).

Rui Barros surge como o profissional mais imbuído e interessado no jornalismo de dados. Uma das observações que evidenciam o seu envolvimento pela área decorre de uma outra conversa informal com este jornalista que entende que os *Sigma Awards*, uma competição que premeia anualmente os melhores trabalhos de jornalismo de dados, são comparáveis ao “Prémio Pulitzer do jornalismo de dados”.

Para ele, o jornalismo de dados é alicerçado, sobretudo, a partir de evidências numéricas que servirão o propósito de contar uma história jornalística, depois de formuladas as hipóteses, tal como é evidenciado na literatura (Gehrke & Benetti, 2020).

“No jornalismo de dados podemos partimos de uma ideia geral, da qual temos um “palpite”. Pode ser algo do senso comum, de uma conversa de café ou com os teus amigos. A partir dessa noção que normalmente começa com “eu acho que isto é assim” ou “há uma ideia generalizada de que isto é assim”, formulamos uma hipótese que poderá dar aso à história jornalística. Um aspeto que torna a história muito interessante é o nível da granularidade dos dados. Ou seja, quanto mais pormenorizados forem os dados, mais interesse o leitor vai ter porque vai relacionar com a história e a informação vai dizer respeito à sua rua, freguesia ou concelho. Mas isso também levanta questões éticas, como deves entender. Por vezes temos de perceber que, por muito boa que aquela história pudesse ser, não podemos utilizar informação mais sensível por causa da questão da privacidade dos dados e assim. Aliás, o próprio INE não publica alguns desses dados” (Rui Barros, conversa informal, 26 de janeiro de 2023, [Dia 19](#)).

Neste sentido, é possível constatar que a evidência numérica é uma das principais características do jornalismo de dados desenvolvido no jornal Público.

Contudo, poderíamos assinalar que, por exemplo, na secção de Economia, a evidência numérica é igualmente uma das principais fontes utilizadas, mas os meios e ferramentas em que o jornalismo de dados se desenvolve são díspares. A recolha de dados, a seleção da informação, a análise conduzida, a representação visual dos dados, a estética narrativa construída para a história, bem como conhecimentos, ferramentas e inclusão de fontes humanas para o propósito da peça jornalística, são elementos que compõem o processo de construção de uma história conduzida por dados neste órgão de comunicação social.

### **5.1.5 As rotinas de produção**

Relativamente às rotinas de produção na equipa de jornalismo de dados, os primeiros dias de observação permitiram compreender que se tratava de uma área em que a maior parte do tempo é despendido diante do computador. O ciclo de trabalho só era momentaneamente interrompido pela interação com algum colega, pelo contacto com alguma fonte, para o esclarecimento e/ou pedido de informações a agentes internos ou externos à redação, e pelas pausas de almoço ou pequenos intervalos no final da tarde.

Este trata-se de um trabalho que, independentemente da fase em que o jornalista de dados se encontre (ver subcapítulo 3.8 **Os fluxos de trabalho**), requer um nível considerável de concentração. Foi possível constatar que os restantes jornalistas da redação tinham saídas mais regulares para realizar reportagens e/ou cobertura de eventos do que os da equipa de jornalismo de dados. A nossa introdução ao fluxo de trabalho da equipa de jornalismo de dados começou pela apresentação conduzida por Rui Barros a alguns dos trabalhos publicados recentemente, os projetos que estavam em curso e propostas de trabalho a ser desenvolvidas ([Dia 1](#)).

Desde o primeiro dia de observação até ao último, ficamos encarregues de atualizar, semanalmente, os dados que compunham a peça “Conheça o impacto da inflação no seu carrinho de compras” ([Dia 1](#), [Dia 9](#), [Dia 14](#), [Dia 34](#), [Dia 39](#), [Dia 66](#)), e que acabaria por dar aso a uma peça jornalística conduzida por dados ([Dia 5](#), [Dia 16](#)). Fomos também introduzidos a uma das principais ferramentas de trabalho que esta equipa utiliza nos seus projetos: o *GitHub*.

O jornalista Rui Barros indicou ([Dia 1](#)) também que a linguagem de programação R e o programa *RStudio* seriam igualmente ferramentas importantes para a condução de análises de dados ([Dia 3](#), [Dia 4](#), [Dia 5](#), [Dia 10](#), [Dia 15](#), [Dia 19](#), [Dia 22](#), [Dia 48](#), [Dia 49](#), [Dia 50](#), [Dia 52](#), [Dia 54](#), [Dia 56](#), [Dia 57](#), [Dia 59](#), [Dia 62](#)) e na elaboração de gráficos e/ou representações visuais de dados ([Dia 3](#), [Dia 4](#), [Dia 19](#), [Dia 49](#), [Dia 52](#), [Dia 54](#), [Dia 57](#), [Dia 59](#), [Dia 62](#)). O *RStudio* é operado através da linguagem R, que se trata de uma linguagem de programação especializada para a análise estatística de dados e para a construção de gráficos e outras representações visuais de dados (Crawley, 2013). Este *software* serviu ainda como ferramenta para a recolha de dados, através da implementação de *scripts* para esse efeito ([Dia 2](#), [Dia 42](#), [Dia 45](#), [Dia 47](#), [Dia 48](#), [Dia 57](#)), e para criar relatórios de análise que pudessem ser apresentados e/ou reproduzidos ([Dia 4](#), [Dia 5](#), [Dia 13](#), [Dia 15](#), [Dia 19](#), [Dia 48](#), [Dia 49](#), [Dia 50](#), [Dia 52](#), [Dia 54](#), [Dia 59](#), [Dia 61](#), [Dia 62](#)). Esta foi uma ferramenta regularmente utilizada durante a nossa observação participante, principalmente no último mês. Podemos ainda constatar uma destreza assinalável da utilização desta ferramenta por parte do jornalista Rui Barros ([Dia 4](#)). O tempo de observação deste estudo permitiu constatar que a utilização rotineira e a experiência adquiridas ao longo do tempo com este *software* são fatores determinantes para que a sua utilização seja fluída ([Dia 52](#), [Dia 66](#)). Não obstante, a utilização deste *software* requer conhecimentos básicos de Estatística, matemática e, sobretudo, de linguagem de programação R. Um dos aspetos interessantes possíveis de evidenciar com este *software* prendeu-se como a necessidade de adotar “boas práticas” das análises de dados conduzidas ([Dia 4](#)):

“Isto, no fundo, é um pouco como nas investigações académicas. No jornalismo de dados, assim como na academia, há uma preocupação com a reprodutibilidade dos resultados. Então, para isso, convém enunciarmos por escrito, no próprio documento da análise, sobre o que se trata o trabalho, quais as hipóteses que queremos testar, de onde é que vêm os dados, que categorias foram excluídas e o porquê, entre outras coisas (...)” (Rui Barros, conversa informal, [Dia 4](#)).

Ainda que o jornalista Rui Barros tenha elencado as vantagens do *RStudio* comparativamente a outros *softwares* de análise de dados, como a sua rapidez, automação, reaproveitamento do código e reprodutibilidade dos resultados ([Dia 3](#)), o *Excel* e *Google Sheets* também são ferramentas utilizadas nas fases iniciais do fluxo de trabalho em jornalismo de dados, sendo indicadas para a formatação e “limpeza” de bases de dados ou ficheiros ilegíveis para a análise imediata da informação ([Dia 41](#), [Dia 49](#), [Dia 50](#)), podendo servir ainda para realizar análises rápidas e cálculos matemáticos simples ([Dia 22](#)).

A existência de bases de dados ilegíveis para serem analisadas nestes *softwares* ([Dia 3](#), [Dia 41](#), [Dia 49](#)), vai ao encontro de evidências constatadas na literatura (Appelgren & Salaverría, 2018). Tal facto exige que o jornalista de dados se debruce durante horas na “limpeza” e transformação da informação num formato compatível e legível para os *softwares* de análise de dados, revelando-se um dos processos mais demorados e extenuantes na rotina de produção ([Dia 41](#), [Dia 49](#), [Dia 50](#)), tendo o mesmo sido afirmado pelos jornalistas deste núcleo ([Dia 41](#)).

Ao longo do período da nossa participação, fomos também confrontados com diversas dificuldades durante a análise ([Dia 2](#), [Dia 3](#), [Dia 4](#), [Dia 5](#), [Dia 13](#), [Dia 15](#), [Dia 19](#), [Dia 42](#), [Dia 66](#)). Em grande parte desses casos, a pesquisa conduzida na Internet ([Dia 10](#), [Dia 49](#), [Dia 52](#), [Dia 54](#), [Dia 59](#)), em fóruns de comunidade da linguagem R ([Dia 52](#), [Dia 59](#)), na secção de “Ajuda” no *RStudio* ([Dia 49](#)) e a colaboração do jornalista de dados Rui Barros permitiram ultrapassar obstáculos, sobretudo no que respeita aos cálculos estatísticos ([Dia 4](#), [Dia 5](#), [Dia 19](#), [Dia 46](#), [Dia 52](#), [Dia 57](#), [Dia 66](#)), código escrito em linguagem R ([Dia 4](#), [Dia 19](#), [Dia 45](#), [Dia 52](#), [Dia 54](#), [Dia 59](#)) e na criação de representações visuais de dados no *RStudio* ([Dia 3](#), [Dia 4](#), [Dia 9](#), [Dia 19](#), [Dia 46](#), [Dia 49](#), [Dia 57](#), [Dia 59](#)).

No entanto, para que os dados possam ser analisados é necessário proceder primeiro à sua recolha ([Dia 2](#), [Dia 3](#), [Dia 10](#), [Dia 13](#), [Dia 14](#), [Dia 19](#), [Dia 22](#), [Dia 42](#), [Dia 45](#), [Dia 47](#), [Dia 48](#), [Dia 57](#), [Dia 59](#), [Dia 62](#)). Um dos processos primordiais observados na consulta a bases de dados passa pela compreensão do contexto em que os dados foram recolhidos, através da metainformação disponível na própria página das fontes ([Dia 13](#), [Dia 39](#), [Dia 41](#), [Dia 54](#)) ou do esclarecimento direto com as fontes que os recolheram ([Dia 2](#), [Dia 22](#)). Vimos também que a equipa de jornalismo de dados tem por hábito consultar as agências governamentais de estatística nacionais e internacionais como fontes de informação, entre as quais o Instituto Nacional de Estatística (INE) ([Dia 13](#), [Dia 24](#), [Dia 36](#), [Dia 39](#), [Dia 41](#), [Dia 42](#), [Dia 45](#), [Dia 47](#), [Dia 48](#), [Dia 56](#)), a Segurança Social ([Dia 61](#), [Dia 62](#)), o site da ADSE ([Dia 49](#)), Direção-Geral de Estatísticas da Educação e Ciência ([Dia 3](#)), Arquivo Histórico-Parlamentar ([Dia 2](#)) e Eurostat ([Dia 10](#), [Dia 22](#), [Dia 47](#)). Verificou-se a colaboração na obtenção de dados com entidades privadas, como o caso da Deco Proteste ([Dia 1](#), [Dia 5](#), [Dia 9](#), [Dia 10](#), [Dia 14](#), [Dia 34](#), [Dia 39](#), [Dia 50](#), [Dia 66](#)), e a consulta a bases de dados reunidas por agências não-governamentais, como a “Our World in Data” ([Dia 1](#)). A esta lista de fontes de informação, acresce ainda a reunião própria de dados por parte do Público, conforme se observou no questionário que deu ao caso à criação de uma página Web agregadora de vários testemunhos, de um gráfico e uma série de reportagens ([Dia 24](#), [Dia 50](#), [Dia 66](#)).

As rotinas de produção de jornalismo de dados são notoriamente marcadas por estes três primeiros processos – a recolha, limpeza e análise da informação numérica – que se realizam de forma bastante autónoma ([Dia 1](#), [Dia 2](#), [Dia 3](#), [Dia 52](#), [Dia 54](#)), ainda que exista a troca de sugestões, por exemplo, de fontes a consultar ([Dia 5](#), [Dia 57](#), [Dia 61](#)) e de correções à análise conduzida ([Dia 4](#), [Dia 46](#), [Dia 54](#), [Dia 59](#)). Ainda sobre este processo, a equipa de jornalismo de dados procura assegurar que nenhum dos projetos de análise se perde, independentemente da sua fase de conceção ou publicação. Para esse efeito, o *Público* conta com uma página no *GitHub* ([Dia 1](#)) com diversos repositórios de trabalhos e ficheiros que facilita a consulta ou partilha de projetos entre os profissionais desta equipa ([Dia 15](#), [Dia 34](#), [Dia 35](#), [Dia 49](#), [Dia 52](#), [Dia 54](#), [Dia 57](#)).

Para cada um dos projetos de jornalismo de dados que estão a ser ou já foram desenvolvidos, é criado um repositório próprio que, por norma, comporta os seguintes elementos: o ficheiro de análise do *RStudio*, o relatório de análise em ficheiro HTML, as bases de dados ou outros ficheiros utilizados ([Dia 15](#), [Dia 61](#)) – como *scripts* ou *shapefiles* –, podendo ainda incluir uma breve descrição sobre o propósito de cada repositório ([Dia 61](#)).

No decurso deste estudo, observamos que as dinâmicas do fluxo de trabalho desta equipa não são exclusivas da sua própria vontade ([Dia 5](#), [Dia 6](#), [Dia 41](#), [Dia 49](#)). Ou seja, pelo menos em duas ocasiões ([Dia 41](#), [Dia 49](#)) verificamos que jornalistas de outras secções do jornal solicitaram a este “departamento” que conduzissem análises de dados a informações que tiveram acesso.

Uma das vertentes mais desafiantes constatadas na literatura (Rogers, Schwabish & Bowers, 2017) também se verificou durante o período deste estudo: a representação visual dos dados. Nesta fase do trabalho, pudemos observar e entender que a criação de gráficos ([Dia 3](#), [Dia 6](#), [Dia 9](#), [Dia 10](#), [Dia 13](#), [Dia 14](#), [Dia 28](#), [Dia 32](#), [Dia 33](#), [Dia 36](#), [Dia 49](#), [Dia 57](#)) ou mapas ([Dia 4](#), [Dia 10](#), [Dia 27](#), [Dia 28](#), [Dia 32](#), [Dia 34](#), [Dia 36](#), [Dia 56](#), [Dia 59](#)) é um processo em jornalismo de dados bastante especializado, que requer um elevado grau de compreensão e sensibilidade na escolha do tipo de gráfico, cor e estilo mais apropriados para a representação visual da informação. Ainda antes deste elemento visual ser incorporado numa peça para ser publicada *online*, a sua conceção, que se consubstancia em gráficos ou mapas, pode *per se* constituir-se como um elemento de análise da informação ([Dia 54](#), [Dia 59](#)), conforme Pereira, Oliveira & Zamith (2013) indicam, revelando novas percepções ([Dia 59](#)). Esse aspeto foi recorrente na utilização do *RStudio*.

Embora este *software* possibilite a criação de representações visuais de dados, não foi observado o seu uso para a inclusão de gráficos em nenhuma das peças publicadas durante o período de observação. Por outro lado, notamos que a sua utilização - para produzir gráficos e/ou mapas - é predominante na elaboração de relatórios de análise de dados ou, por exemplo, para “visualizar a informação” nos diferentes níveis geográficos das Nomenclatura das Unidades Territoriais de Portugal.

A produção dos relatórios de análise no *RStudio* serve para que outros membros da equipa de jornalismo de dados ou profissionais do mesmo meio possam aceder, compreender e reproduzir a análise realizada ([Dia 4](#), [Dia 5](#), [Dia 52](#), [Dia 54](#), [Dia 61](#), [Dia 66](#)). Para este efeito, o *RStudio* possibilita “transformar” o documento de análise num ficheiro HTML, através de uma linguagem e funções próprias que o *software* compreende: o *RMarkdown* e o “*knitr*”, respetivamente ([Dia 52](#), [Dia 54](#), [Dia 59](#), [Dia 66](#)). Esse ficheiro HTML pode, posteriormente, ser enviado por endereço eletrónico à/ao colega que solicitou a análise da informação ([Dia 54](#)).

No que concerne à fase da criação das representações visuais de dados que serão publicadas, o *Público* possui um “livro de estilo” próprio ([Dia 6](#)), com objetivos específicos, que visa uniformizar gráficos e mapas, através do qual a equipa de jornalismo de dados se norteia. Esta fase do fluxo de trabalho requer um nível considerável de dedicação e tempo despendidos ([Dia 27](#)) na sua produção. Em certas ocasiões, assistimos à colaboração entre a equipa de jornalismo de dados com colegas da área da Infografia ([Dia 16](#), [Dia 32](#), [Dia 33](#)) e também à utilização da Internet por soluções ou ficheiros úteis para esse efeito ([Dia 4](#), [Dia 54](#), [Dia 57](#), [Dia 59](#), [Dia 61](#)). Durante este estudo verificamos, pelo menos, uma ocasião ([Dia 27](#)) onde ocorreu a colaboração e participação de agentes externos à redação para elaborar representações visuais de dados (De-Lima-Santos & Mesquita, 2021, p. 1418).

Para a criação das representações visuais de dados publicadas *online*, o núcleo de jornalismo de dados recorre ao site e às funcionalidades do *Flourish* ([Dia 6](#), [Dia 10](#), [Dia 28](#)), que permite importar a base de dados com a qual se opera ([Dia 9](#)), possibilitando ao utilizador escolher a informação que pretende destacar, selecionar diversos tipos de gráficos ([Dia 14](#), [Dia 28](#)), estilizá-los de acordo com o “livro de estilo” ([Dia 9](#), [Dia 16](#), [Dia 33](#)) e incorporar elementos interativos ([Dia 14](#)). Relativamente aos desafios constatados neste ponto, decorrente da nossa participação e envolvimento, foi possível observar que o jornalista de dados Rui Barros se mostra ciente dos mesmos por ter demonstrado sensibilidade, conhecimentos e experiência numa das fases mais importantes do trabalho ([Dia 9](#), [Dia 14](#), [Dia 15](#), [Dia 16](#), [Dia 24](#), [Dia 27](#), [Dia 32](#), [Dia 35](#), [Dia 38](#), [Dia 46](#), [Dia 56](#)). O mesmo profissional evidenciou possuir conhecimentos na área da ilustração e design ([Dia 24](#), [Dia 38](#), [Dia 47](#)).

Sublinhe-se também que o diálogo e a colaboração, formal ([Dia 47](#)) e/ou informal ([Dia 33](#)), com a equipa da Infografia ocorre com regularidade nesta fase do trabalho, assim como na fase subsequente à publicação final da peça jornalística ([Dia 47](#)).

Outra das evidências observadas nas rotinas de produção tem que ver com o desenvolvimento Web das histórias conduzidas por dados. O nosso nível de participação nesta vertente foi condicionado devido à inexperiência e desconhecimento em operar com linguagens de marcação e de programação utilizadas pela equipa, nomeadamente HTML, CSS e Javascript ([Dia 22](#), [Dia 28](#)).

O maior envolvimento constado neste âmbito decorreu num trabalho jornalístico, publicado a 6 de março de 2023 ([Dia 47](#)), que comportou uma narrativa visual em *scrollytelling* ([Dia 22](#), [Dia 28](#), [Dia 34](#), [Dia 35](#)) e vários processos de desenvolvimento Web e componentes interativas ([Dia 27](#), [Dia 28](#), [Dia 34](#), [Dia 38](#), [Dia 39](#), [Dia 41](#)). Esta peça acabaria por ser coassinada por dois elementos da equipa de jornalismo de dados ([Dia 47](#)), algo que também se verifica em outros trabalhos ([Dia 36](#), [Dia 47](#)) e em colaborações externas a este núcleo ([Dia 24](#), [Dia 50](#), [Dia 56](#), [Dia 66](#)). Antes dessa peça em *scrollytelling* ser publicada, foi discutido com o jornalista Rui Barros a forma como a história seria desenvolvida visualmente no *online* ([Dia 22](#), [Dia 27](#)). Face a essa intenção, tivemos a oportunidade de assistir e participar numa aula informal conduzida pelo mesmo, dirigida à unidade de jornalismo de dados, realizada nas instalações da redação ([Dia 28](#)). Salvo o nosso caso, os jornalistas Rui Barros e José Volta e Pinto possuíam conhecimentos e competências nesta área ([Dia 14](#), [Dia 28](#)), embora Rui Barros evidencie uma maior habilidade e entendimento na linguagem de programação, comprovando-se com os trabalhos que desenvolveu durante o período de observação.

O *Visual Studio Code* é a ferramenta de trabalho primordial utilizada para desenvolver peças que comportam uma estrutura técnica e estética diferenciadora ([Dia 1](#), [Dia 22](#), [Dia 24](#)) na Web. Mediante a nossa observação, vimos que este *software* foi utilizado pelo jornalista Rui Barros para implementar e desenvolver componentes, a partir de uma estrutura e linguagem de componente *front-end* gratuita – o *Svelte* ([Dia 28](#), [Dia 35](#), [Dia 41](#)) –, que ficariam ao dispor dos elementos da equipa de jornalismo de dados para reutilização em projetos futuros.

Evidenciamos que a utilização do *Visual Studio Code* permite “paginar” a peça de raiz, pré-visualizar e testar a mesma antes de ser publicada ([Dia 9](#), [Dia 35](#)). À semelhança do *RStudio*, o *Visual Studio Code* também é um *software* de licença gratuita ([Dia 27](#)).

Sem surpresas, a redação da história jornalística é um processo igualmente importante para os membros deste núcleo ([Dia 1](#), [Dia 9](#), [Dia 10](#), [Dia 35](#), [Dia 38](#), [Dia 41](#)), o que vai ao encontro do “talento para a escrita” e das aptidões jornalísticas tradicionais evidenciadas na literatura (Meyer, 2002; Ojo & Heravi, 2017; Robles & López, 2021). Em várias ocasiões, o jornalista Rui Barros, o editor Ivo Neto, o diretor-adjunto Amílcar Correia e outros jornalistas davam conselhos de melhoramento e clareza do texto, fornecendo também contactos que pudessem servir de testemunhos para a peça jornalística ([Dia 5](#), [Dia 6](#), [Dia 24](#), [Dia 59](#), [Dia 61](#)).

No quotidiano na redação foi ainda possível observar que, dos dois elementos da equipa de jornalismo de dados, Rui Barros é o que assume um papel mais presente na condução de análises, da criação das representações visuais e ainda no desenvolvimento Web em projetos alicerçados em evidências numéricas. Devido a esses fatores, este jornalista encontrava-se frequentemente ocupado ([Dia 9](#), [Dia 13](#), [Dia 32](#), [Dia 41](#), [Dia 49](#), [Dia 56](#), [Dia 61](#)) despendendo várias horas diante do computador e, por vezes, até fora do horário de expediente ([Dia 35](#)), ainda que, tal como indicamos, tivesse um horário flexível. Face a este volume de trabalho, o esclarecimento de dúvidas com o mesmo, no imediato, não ocorria com regularidade. Por outro lado, quando os trabalhos se encontravam-se na fase da criação das representações visuais de dados ([Dia 14](#), [Dia 15](#), [Dia 16](#), [Dia 32](#), [Dia 35](#), [Dia 38](#), [Dia 59](#)), prestes a ser publicados ([Dia 16](#), [Dia 46](#)) ou quando este profissional pareceu estar com um menor volume de trabalho ([Dia 1](#), [Dia 3](#), [Dia 14](#), [Dia 19](#), [Dia 28](#)), o seu nível de interação e envolvimento ocorria com maior frequência.

Face à dimensão das atividades realizadas e do nosso envolvimento com esta equipa, pudemos notar que o jornalismo de dados se trata de um trabalho que se pode tornar frustrante e desafiador em qualquer uma das suas fases – desde a recolha dos dados até à criação das representações visuais da informação analisada.

Isso deve-se, essencialmente, às competências necessárias para superar obstáculos e dificuldades técnicas, bem como o tempo despendido. Em diversas ocasiões, a pesquisa na Internet por esclarecimentos revelou-se uma ferramenta importante para solucionar algumas dessas dificuldades. No mesmo sentido, a leitura da metainformação disponível nas bases de dados, o contacto com as instituições que reuniram os dados ou com fontes que podem servir de testemunhos para compreender o contexto da informação são estratégias eficientes, mas que também levam o seu tempo.

Não obstante, a publicação de uma história jornalística baseada em dados traduz-se num sentimento de gratificação, sobretudo quando os pares profissionais reconhecem o trabalho realizado ([Dia 16](#), [Dia 36](#), [Dia 47](#)) por esta equipa e pelo alcance que a história atinge nos seus leitores ([Dia 16](#), [Dia 47](#)).

Em suma, é possível afirmar que as competências do principal jornalista de dados do Público (Rui Barros) são notórias em campos como a Estatística, matemática, programação, representação visual da informação, redação textual e desenvolvimento Web. Além disso, o jornalista é sensível às questões éticas e deontológicas inerentes à profissão do Jornalismo e também neste subdomínio jornalístico. Ao longo deste estudo, foi possível notar que este jornalista procurou incentivar e desenvolver as nossas competências em ambas dimensões: na das competências profissionais com a participação e acompanhamento no fluxo de trabalho em jornalismo de dados, assim como recomendação de leituras ([Dia 2](#), [Dia 9](#)), enquanto as questões éticas e deontológicas foram transmitidas em conversas informais ([Dia 19](#), [Dia 45](#)), assim como as correções e/ou sugestões de melhoramento. Por último, notamos que o fluxo de trabalho em jornalismo de dados é mais demorado do que a publicação de conteúdos de outras secções do jornal dada as suas exigências técnicas e conceptuais ([Dia 19](#)).

## 5.2 Análise de conteúdo

### 5.2.1 As potencialidades Web adotadas ao nível da apresentação

No que concerne às potencialidades Web apresentadas nas peças que compõem a amostra desta investigação, a análise de conteúdo permitiu compreender que, ao nível da hipertextualidade, a equipa de jornalismo de dados opta por incluir notícias do próprio jornal sobre assuntos relacionados com a peça em questão através de hiperligações intratextuais. Apenas em quatro ocasiões foram identificadas hiperligações no “miolo do texto” que redirecionam o leitor para páginas externas, tratando-se de duas páginas de bases de dados (Peça 2 e Peça 7), uma outra sobre um relatório publicado pela Organização Internacional do Trabalho (Peça 7) e uma sobre uma música na plataforma *Spotify*.

No que concerne às hiperligações extratextuais, este é um recurso recorrente para fornecer o contacto dos autores da notícia, hiperligações de partilha em redes sociais, sobretudo, Facebook, Twitter e LinkedIn, mas também o Pinterest e WhatsApp (Peça 2, Peça 9). Além disso, este tipo de hiperligação é utilizado para destacar visualmente notícias recomendadas, identificar a secção editorial da peça e também para o envio do artigo através de e-mail, guardá-lo na conta de utilizador do site do Público e/ou para comentar a peça. A utilização de hiperligações genéricas só se verificou em um caso da amostra (Peça 7).

Relativamente à multimedialidade, verificamos uma ausência da utilização de diaporamas, enquanto a utilização de imagens ou desenhos também não é frequente na página das peças de jornalismo de dados: apenas uma utilizou fotografia (Peça 2), outra recorreu à ilustração (Peça 7) e uma outra a um modelo tridimensional animado de um carrinho de compras (Peça 1). Por outro lado, as representações visuais de dados estáticas e dinâmicas são bastante recorrentes (Peça 1, Peça 2, Peça 3, Peça 6, Peça 7 e Peça 9), existindo, em alguns casos, a presença de ambas (Peça 1, Peça 7 e Peça 9). A análise conduzida permitiu compreender que nas peças que não utilizam infografias, gráficos ou mapas estáticos/dinâmicos, outras potencialidades ciberjornalísticas são aplicadas, tais como o áudio e os vídeos com ou sem som (Peça 3, Peça 4, Peça 5, Peça 8 e Peça 10).

Em nenhum caso se verificou a presença de todos os elementos da dimensão da multimedialidade e apenas um (Peça 3) combina o vídeo sem som, tratando-se da pré-visualização da reportagem multimídia, com uma representação visual de dados dinâmica.

Outras das potencialidades ciberjornalísticas exploradas pelos autores das peças recaem principalmente na componente estética ou visual, como são exemplo a utilização de emojis; o uso de sublinhados, negritos ou traçados animados com cores expressivas (tons avermelhados) para destacar dados ou informações; excertos de texto com um tamanho de letra maior que o texto da reportagem; a adoção de um estilo estético diferenciado (“Peça 4: O fundo desta página é preto, evidenciando-se uma diferença estética com a maioria dos artigos identificados nesta amostra.”); a combinação entre áudios e sua transcrição síncrona (p.e.: Peça 4); a legendagem dos vídeos com som (p.e.: Peça 5); a opção por ativar ou desativar o som dos vídeos e áudios ou a revelação de subcamada(s) de informação após interagir com alguma representação visual de dados interativa (p.e.: “gráfico de árvore” da Peça 2 e mapa dinâmico da Peça 9).

Quanto aos recursos interativos presentes nas representações visuais de dados, descobriu-se que metade da composição da amostra contempla a potencialidade de revelar detalhes ou destaques ao passar o cursor (Peça 1, Peça 2, Peça 3, Peça 6 e Peça 7), independentemente da representação visual se tratar de um gráfico ou mapa interativo (p.e.: Peça 7). A segunda ferramenta interativa mais identificada trata-se dos botões de navegação que se encontram em mapas dinâmicos (Peça 6 e Peça 9), assim como em gráficos (Peça 2) e na estrutura da página Web (Peça 3). A filtragem dos dados é o terceiro recurso mais evidenciado (Peça 1, Peça 2 e Peça 6), assim como o zoom nos mapas (Peça 6 e Peça 9) e gráficos (Peça 2). A pesquisa ou pesquisa detalhada dos resultados foi o recurso interativo menos evidenciado, verificando-se apenas em dois artigos (Peça 1 e Peça 6).

Nos tipos de interatividade presentes nas representações visuais de dados, verificamos que a interatividade de transmissão é a mais evidente (Peça 1, Peça 2, Peça 3, Peça 6, Peça 7 e Peça 9), possibilitando ao leitor consultar informações complementares através de “pop-ups” ao passar o cursor ou clicando em determinados pontos da representação visual de dados. No entanto, este tipo de interatividade também é conjugado com a interatividade consultiva (Peça 1, Peça 2, Peça 6 e Peça 9) e também com a interatividade de registo (Peça 3 e Peça 6).

A interatividade consultiva ocorre quando: o leitor explora os dados disponíveis na própria representação visual de dados (Peça 1); clica para ampliar gráficos (Peça 2) ou mapas (Peça 6 e Peça 9) e/ou seleciona categorias de informação pré-definidas (Peça 6). Por sua vez, a interatividade de registo é a menos frequente, verificando-se apenas em duas peças (Peça 3 e Peça 6) nas quais a inserção ou preenchimento de dados provocam alterações em tempo real (Peça 6) ou subsequentes (Peça 3) à representação visual de dados.

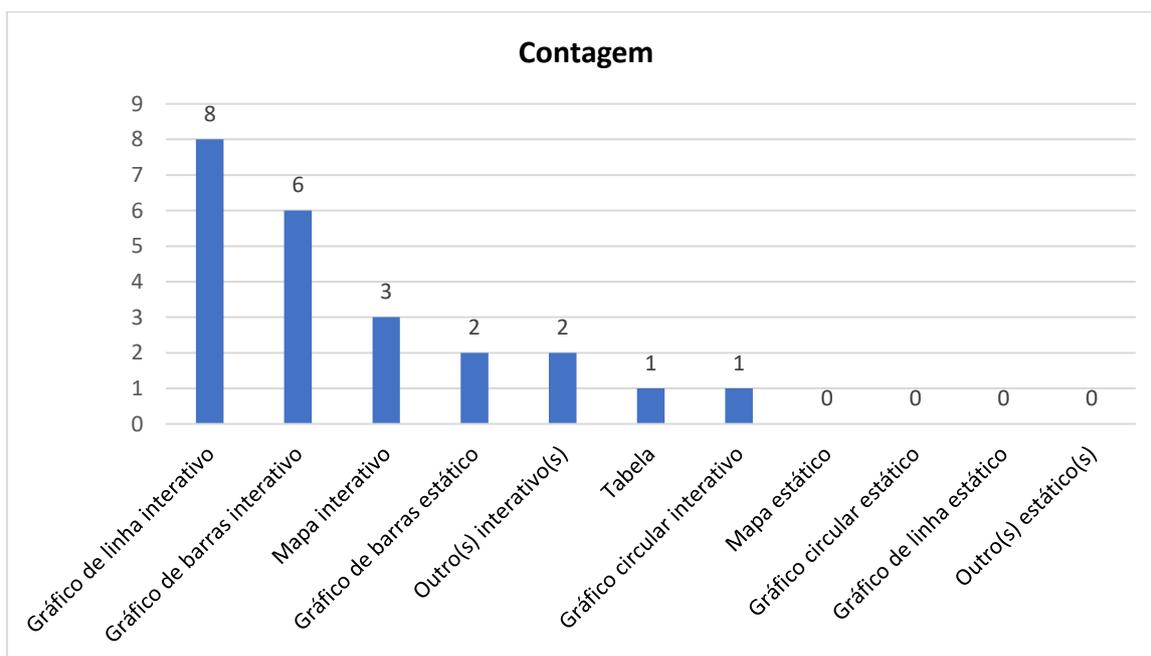
Na dimensão dos esquemas narrativos, de forma a analisar a orientação narrativa, apenas foram consideradas as peças que contemplavam, pelo menos, uma representação visual de dados, pois trata-se de um elemento necessário tal como é indicado pelos autores (Segel & Heer, 2010). Nesse sentido, compreendemos que a orientação narrativa mais frequente é a da estrutura de Martini-Glass, sendo que o autor assume a condução inicial da narrativa da história, apresentando “perguntas, observações ou artigos escritos para introduzir a visualização” (*Ibidem*), tal como foi evidenciado nas Peça 1, Peça 2, Peça 6, Peça 7 e Peça 9. Em alguns casos a pergunta (Peça 2, Peça 6, Peça 7 e Peça 9) surge no próprio título da notícia, seguindo-se a representação visual de dados que permite ao leitor “a exploração interativa dos dados”, observando-se assim a “fase orientada para o leitor”. Consideramos que a Peça 3 é a única que comporta uma estrutura de “história de pesquisa”, uma vez que esta página Web permite que o leitor escolha e navegue pelas histórias secundárias, através das hiperligações que dão acesso aos testemunhos na íntegra, e das reportagens publicadas. Não verificamos qualquer tipo de *slideshow*. Por último, sobre a estética visual da narrativa, vimos que o *scrollytelling* foi adotado nas Peça 4, Peça 5, Peça 7, Peça 8 e Peça 10.

Com a exceção da Peça 7, não contemplam qualquer tipo de representação visual, embora haja a menção a dados e a inclusão de outras ferramentas multimédia, como áudios e vídeos com e sem som. As Peça 3 e Peça 6 foram incluídas na categoria de *newsapp*, pois contempla funcionalidades interativas *online* que contam “a história do utilizador” (Howard, 2014) permitindo a inserção de informações do leitor, que levam à personalização dos dados apresentados em gráficos ou mapas, e que auxiliam a compreensão e manipulação de grandes quantidades de dados, como é o caso da Peça 6.

### 5.2.2 Principais representações visuais de dados e ferramentas de execução

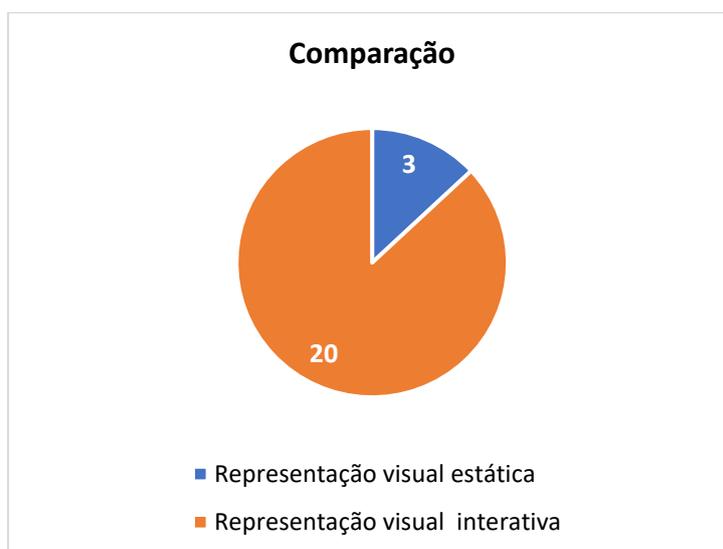
No total das 10 peças da amostra, foram identificadas 23 representações visuais de dados, das quais o gráfico de linha interativo (34,8%), o gráfico de barras interativo (26,1%) e o mapa interativo (13%) são as mais utilizadas (*Gráfico 1*). Não foi encontrada qualquer evidência de utilização de mapas estáticos, gráficos circulares ou de linha estáticos. Neste sentido, podemos afirmar que, entre dia 2 de janeiro e 31 de março de 2023, o gráfico de linha interativo foi o tipo de representação visual de dados mais utilizado nas peças publicadas pela equipa de jornalismo de dados do Público.

Gráfico 1 - Contagem dos tipos de representações visuais de dados publicadas



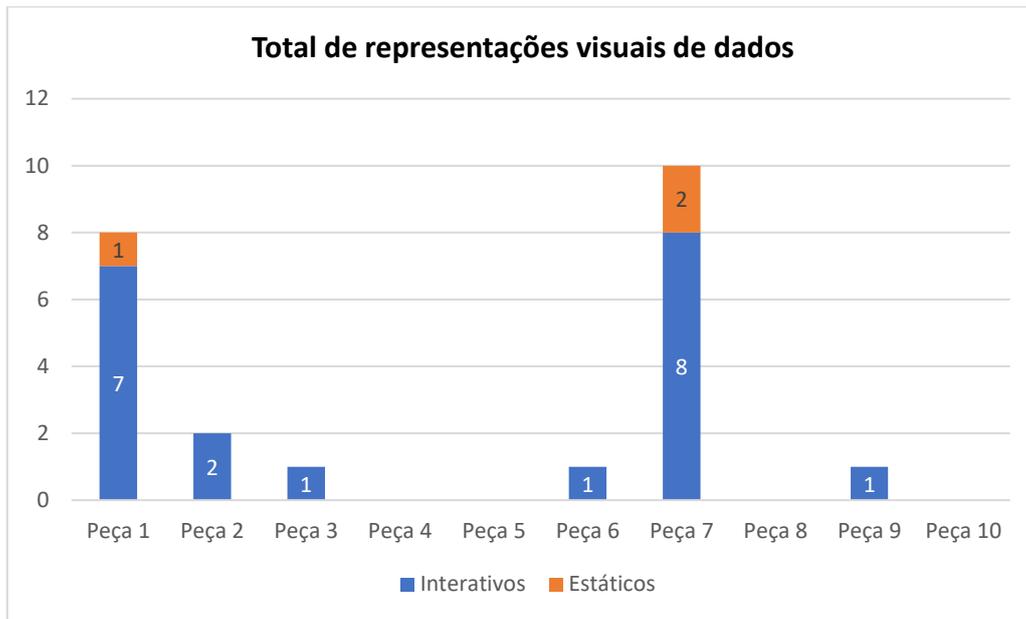
Em complemento a esta análise, verificamos que 87% do total das representações visuais de dados comportam algum nível de interatividade, enquanto 13% delas são estáticas (*Gráfico 2*). Essa análise reforça a ideia de que os recursos interativos, identificados no subcapítulo anterior sobre as potencialidades Web, são frequentemente utilizados e adotados nas representações visuais da informação apresentadas.

Gráfico 2 - Comparação entre representações visuais de dados estáticas e interativas



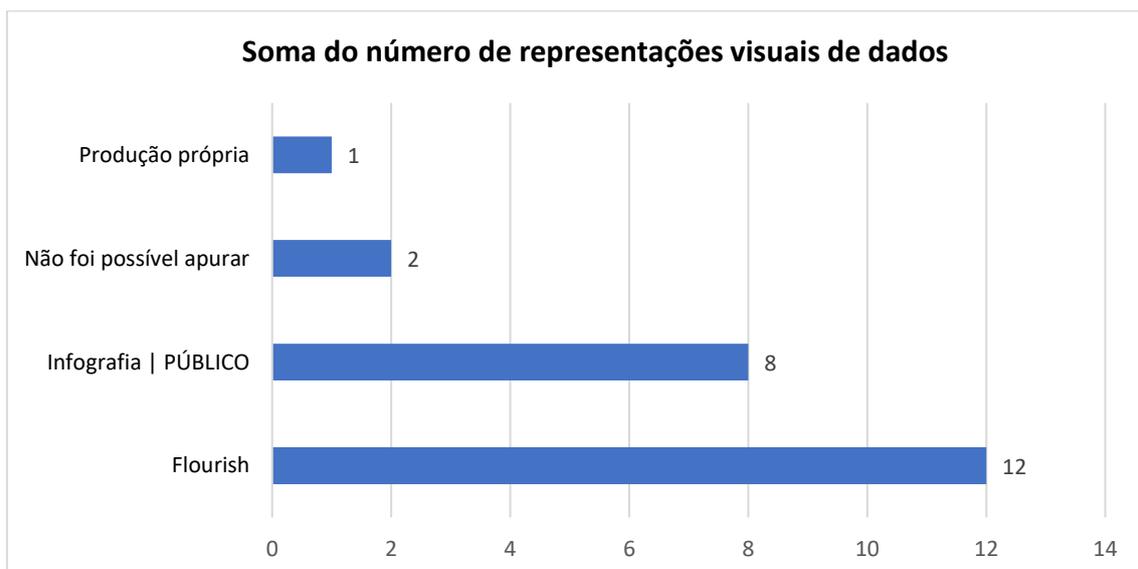
Por último, notamos que a presença de representações visuais de dados estáticas só ocorre nas peças em que existe um número substancial de gráficos ou mapas interativos (*Gráfico 3*). Além disso, 40% das 10 peças analisadas não contemplaram qualquer tipo de representação visual de informação numérica. Não obstante, devemos notar que essas quatro peças são reportagens que provêm do questionário da Peça 3 (“Contas à vida: enfrentar a inflação”), o que revela uma aproximação às práticas antecessoras do jornalismo de dados – a RAC (Meyer, 2002).

Gráfico 3 - Total de representações visuais de dados estáticas e interativas por peça



No que concerne à ferramenta de trabalho utilizada para a produção das representações visuais de dados, foi possível apurar que a plataforma online *Flourish* foi a mais utilizada (*Gráfico 4*). Das 23 representações visuais de dados identificadas, 12 delas (52,1%) foram criadas a partir desta ferramenta, o que significa que mais de metade delas foram desenvolvidas a partir deste serviço de terceiros.

Gráfico 4 - Ferramenta utilizada ou autoria das representações visuais de dados



Já a equipa de Infografia do *Público* foi o segundo elemento identificado com o maior número de representações visuais de dados criadas (oito), o que representa 34,8% do total da amostra. Em duas representações visuais de dados não foi possível identificar a ferramenta utilizada ou o autor da sua elaboração, pois não existia qualquer menção a esse elemento. Apenas uma das 23 representações visuais de dados foi assinalada como sendo de produção própria da equipa de jornalismo de dados, pois o primeiro parágrafo dessa notícia (Peça 6) refere que os valores apresentados são de um “mapa do PÚBLICO” e é assinado por dois dos elementos da equipa de jornalismo de dados (Rui Barros e José Volta e Pinto).

### 5.2.3 Transparência, precisão e imparcialidade dos dados

Para compreender como o núcleo de jornalismo de dados garante a transparência, precisão e a imparcialidade dos dados utilizados nas suas peças, a análise de conteúdo às peças da amostra permitiu compreender que estes profissionais utilizam a composição do título e descrição das representações visuais de dados, assim como o próprio corpo textual da notícia, como meios para explicar a metodologia aplicada nas análises dos dados. As seguintes unidades semânticas extraídas da nossa amostra evidenciam isso mesmo:

- 1) **Título e descrição da representação visual de dados** - Peça 2: *“Descrição do primeiro gráfico: O preço médio atingiu um novo máximo a 18 de janeiro de 2023. Com base em 11 marcas, este produto está a custar, em média, mais de dois euros nos supermercados; Título e descrição do segundo gráfico: Até novembro de 2022, mais de 50% do arroz importado veio da América do Sul [título]; “Outros tipos de arroz vendidos em Portugal não registaram um aumento tão expressivo. A maioria (quase 70%) é proveniente de países “imunes” aos efeitos da guerra” [descrição]”.*
- 2) **Texto da notícia** – Peça 6: *“[Texto da reportagem]: Para quem está habituado a navegar pelos sites de compra e arrendamento de casas, os valores apresentados no mapa do PÚBLICO podem parecer baixos em relação à realidade do mercado – exceção feita a algumas regiões do país.*

*Uma primeira razão para este desfasamento é muito directa: as páginas de compra e arrendamento de imóveis apresentam o valor que o proprietário pretende obter e não o que fica firmado nos contratos de arrendamento, tendencialmente mais baixos. Por sua vez, os dados do mapa reflectem o preço dos contratos assinados e que são utilizados na definição das rendas do Programa de Apoio ao Arrendamento (PAA). (...)”.*

Algumas representações visuais de dados contêm blocos textuais de contextualização da informação, integrados num formato narrativo visual de *scrollytelling*, e a referência explícita a “notas” ou “notas metodológicas”, que se encontram no rodapé, como é exemplo a Peça 7: “[2.º bloco textual do scroller] “O indicador da alta satisfação assume diferentes pesos em cada um dos Estados-membros da UE, assim como na Noruega e na Suíça. Malta, Letónia, e Suíça são os países onde se registam os valores mais elevados: mais de 65% dos entrevistados estavam “altamente satisfeitos” com o seu emprego.”, ao qual acresce uma nota (“Nota: A base de dados reúne as respostas de 27 países da União Europeia, aos quais acrescentam a Noruega e a Suíça.”).

Relativamente à identificação da fonte dos dados, o rodapé das representações visuais de dados é utilizado para esse efeito, tal como foi assinalado na Peça 9: “Fonte identificada no rodapé da representação visual de dados (“Fonte: Censos 2021”)”. A menção à fonte da informação numérica pode também ser realizada:

- a) **no texto da notícia:** Peça 4 - “Miguel é um dos 238 mil trabalhadores que, em Portugal, acumulam dois ou mais empregos, segundo dados do Instituto Nacional de Estatística (INE) relativos ao terceiro trimestre de 2022. Há um ano eram menos 30 mil, agora representam quase 5% do total da população empregada”);
- b) **em hiperligações intratextuais:** Peça 3 – hiperligação intratextual em “Contribua com o seu caso e ajude-nos a pintar o gráfico” que remete o leitor para o questionário, que formulará a base de dados própria do jornal”;
- c) **em hiperligações extratextuais:** Peça 10 - “hiperligação extratextual em “Contas à vida: enfrentar a inflação”, que remete o leitor para a página onde pode ser encontrado a hiperligação para o formulário e a base de dados própria no gráfico circular (...)”.

- d) **nas notas metodológicas das representações visuais de dados:** Peça 6 – “[Nota do rodapé]: ‘(...) São apresentados dados para cada município ou freguesia para o qual o INE disponibiliza dados (...)’.”

Importa realçar que consideramos que o gráfico circular presente na Peça 3 terá sido construído com base na resposta do inquérito, estabelecendo-se como uma base de dados própria do jornal, embora não exista qualquer menção. Partindo desse pressuposto, a Peça 4, Peça 5, Peça 8 e Peça 10 apresentam dados dos próprios entrevistados, mas observou-se a menção a informações quantitativas que carecem da identificação da fonte. Um desses casos pode ser encontrado na Peça 8, no segundo e terceiro blocos textuais do *scroller*: ([2.º bloco textual do *scroller*]: "Apenas 13% dos agricultores em Portugal trabalham a tempo inteiro na própria exploração"; [3.º bloco textual do *scroller*]: "Aproximadamente dois terços ocupam menos de 50% do horário laboral em culturas agrícolas").

## **Conclusão**

O estágio realizado na redação do Público no Porto, entre o dia 2 de janeiro e 31 de março de 2023, permitiu desenvolver diversas competências no campo do jornalismo. Esta experiência proporcionou uma imersão direta e aprofundada no ambiente cultural e profissional da equipa de jornalismo de dados, que procuramos aproveitar da forma mais proveitosa possível, tendo sido possível contribuir para um projeto jornalístico inovador e focado em aproveitar os recursos tecnológicos, as potencialidades da interatividade e de colaboração próprias do meio digital. Durante este período, observamos, aplicamos e aprimoramos as competências necessárias para a análise e apresentação de dados numéricos, contribuindo para a criação de conteúdos jornalísticos relevantes.

De forma a responder à primeira pergunta de investigação (Quais são as competências necessárias para a prática do jornalismo de dados no jornal Público?), concluímos que a hipótese 1 é rejeitada, pois as principais competências desta equipa de jornalismo de dados - composta, na prática, por dois jornalistas – extravasam a consulta por uma base de dados, a recolha da informação e os conhecimentos básicos em Estatística para a analisar.

Ainda que essas competências se tenham verificado, aptidões nas áreas da linguagem de marcação, programação, desenvolvimento Web e da representação visual de dados são essenciais para que o jornalista opere com tecnologias e *softwares* especializados na recolha, “limpeza” e análise de dados de forma a produzir representações visuais de dados e extrair informações relevantes para as comunicar em narrativas acessíveis e atrativas para os leitores. Esta constatação vai ao encontro da parte teórica desta investigação (Tabary et al., 2016), que evidencia igualmente a importância das “competências tradicionais” do Jornalismo (Ojo & Heravi, 2017; Robles & López, 2021), como a escrita proficiente.

Na área da programação, a linguagem R e o *RStudio* são tecnologias de livre acesso frequentemente utilizadas por esta equipa para manipular e analisar grandes quantidades de dados. Estas ferramentas permitem recolher dados de grande dimensão, recorrendo a técnicas e linhas de código, como o *scraping*; executar a “limpeza” e seleção dos dados a analisar; realizar cálculos estatísticos; visualizar os dados em gráficos ou mapas e produzir relatórios de análise que podem ser partilhados e reproduzidos pelos seus pares. Para este último efeito, o *GitHub* é a ferramenta predileta para partilhar o progresso de um projeto de análise de dados em curso, pois permite que alterações e sugestões sejam introduzidas e atualizadas remotamente, bem como para disponibilizar recursos ou documentos úteis para esta equipa. Já o *Excel* e o *Google Sheets* são ferramentas utilizadas para executar, sobretudo, a limpeza de bases de dados, podendo, ocasionalmente, ser usados para efetuar cálculos estatísticos mais usuais, como a média, percentagens ou proporções. Ainda na área da programação, o *Visual Studio Code* é outra ferramenta utilizada. Esta serve para desenvolver, nos artigos publicados no meio digital, estruturas narrativas visuais diferenciadoras ou componentes e recursos distintos. Operar com este *software*, também gratuito, exige um profundo conhecimento de linguagens de marcação, como HTML e CSS, bem como na linguagem de programação de desenvolvimento web, como o JavaScript e *frameworks* de desenvolvimento Web que compilam os seus componentes, como o *Svelte*.

Já o *Flourish* é a plataforma *online* a que esta equipa (e os elementos da secção da Infografia do jornal) recorre para a produção visual de gráficos e mapas de forma a comunicar informações de maneira eficaz e envolvente. Além disso, o *Flourish* possui uma variedade de modelos e opções de personalização das representações visuais de dados, permitindo que os mesmos sejam incorporados em sites. Mediante esta descrição das ferramentas usadas no fluxo de trabalho em jornalismo de dados, verificamos a existência de um nível elevado de conhecimentos especializados em áreas que a maioria dos jornalistas não está familiarizada, como as diferentes linguagens de programação (Boyles & Meyer, 2017) e a representação visual de dados (Rogers, Schwabish & Bowers, 2017). Verificamos também na prática que, aquando da utilização do *RStudio*, a linguagem de programação R tornou-se transversal a todas as fases do fluxo de trabalho antes de a peça ser publicada – desde a recolha até à representação visual dos dados (Lewis et al., 2020; Lindgren, 2021).

Relativamente à segunda pergunta de investigação (Quais as potencialidades que a Web proporciona na execução e apresentação de uma peça jornalística de dados no jornal Público?), a hipótese formulada é confirmada. Primeiramente, foi possível concluir que, ao nível da execução, existe uma interligação com a primeira questão, pois a utilização das ferramentas tecnológicas elencadas são, maioritariamente, de livre acesso (Gynnild, 2014), comprovando a presença de um dos “elementos diferenciadores” do jornalismo de dados: a disseminação abundante de “ferramentas para trabalhar os dados com software gratuito” (Sandoval-Martín & La-Rosa, 2018). A sua utilização facilita o fluxo de trabalho destes profissionais e, além disso, a consulta em fóruns ou páginas Web para o esclarecimento de dúvidas também pode ser considerada como uma potencialidade do ciberespaço por auxiliar na produção dos artefactos jornalísticos deste tipo (Sandoval-Martín & La-Rosa, 2018; De-Lima-Santos & Mesquita, 2021b).

Ao nível da apresentação e publicação dos conteúdos assinados pelos elementos desta equipa, constatamos a presença de um vasto conjunto de potencialidades interativas exploradas, como são exemplo os diferentes tipos de hiperligações empregues, os conteúdos multimédia, as orientações e estéticas visuais da narrativa, entre outras potencialidades descritas.

O mesmo sucede-se nas representações visuais de dados que comportam recursos interativos de navegação, exploração e consulta por mais informações, através do cursor, que revela detalhes ou destaques, dos botões de navegação, da filtragem dos dados ou do zoom em mapas e gráficos.

A implementação destes recursos interativos permitiu ainda verificar a existência dos três tipos de interatividade identificados na parte teórica: a de transmissão, a consultiva e a de registo (Veglis & Bratsas, 2017). Assim sendo, a hipótese é confirmada, mas importa notar que o uso de hiperligações genéricas foi preterido, assim como o das imagens ou desenhos. Já a aplicação de diaporamas e de *slideshows* interativos para a representação visual de dados não se verificou em qualquer caso durante este período. Por outro lado, os casos em que não se verificou a presença de qualquer infografia, gráfico ou mapa – de natureza estática ou interativa - outras potencialidades Web foram aplicadas, como a estética narrativa visual do *scrollytelling* ou a *news app*.

Em suma, ao empregar formas de narração inovadoras, hiperligações, conteúdos multimédia - como imagens, vídeos ou gráficos interativos – e representações visuais de dados interativas, este subdomínio jornalístico espelha algumas das potencialidades transformadoras que o advento da Internet e da *World Wide Web* trouxeram à forma de produção, consumo e apresentação final dos artefactos jornalísticos. Face à ausência de alguns recursos identificados, podemos concordar com a ideia de que as potencialidades Web poderiam ser “mais e melhor” aprofundadas e aproveitadas (Dick, 2014; McAdams, 2014; Appelgren & Jönsson, 2021), mas vale a pena ressaltar que: 1) o núcleo de jornalismo de dados é bastante reduzido no seu número de profissionais afetos a este subdomínio jornalístico; 2) as pressões do ciclo de produção e publicação de notícias podem afetar a exploração mais aprofundada dessas potencialidades Web; e 3) a equipa é confrontada com a necessidade de adaptar estes conteúdos para diferentes ecrãs e dispositivos (Anderson & Borges-Rey, 2019).

No que diz respeito à terceira pergunta de investigação (Quais são as principais representações visuais de dados a que o *Público* recorre nas suas peças para apresentar dados e de que forma são produzidas?), a hipótese formulada é parcialmente rejeitada, pois não foi identificado o uso de qualquer tipo de mapa estático nas representações visuais de dados publicadas, e é parcialmente confirmada visto que o *Flourish* se trata de uma ferramenta de terceiros e foi a mais utilizada para conceber as representações visuais de dados identificadas. Assim sendo, a análise conduzida permitiu concluir que os principais tipos de representação visual de dados publicados durante o período definido foram os gráficos de linha interativos.

Contrariamente ao que a literatura sugere sobre os “formatos visuais estáticos” serem “mais frequentes” que os “gráficos e mapas clicáveis” em jornalismo de dados (Anderson & Borges-Rey, 2019), as representações visuais de dados publicadas por esta equipa do *Público* são, maioritariamente, interativas. Das 23 representações visuais de dados identificadas, 87% apresentavam algum grau de interatividade, permitindo ao leitor explorar as informações disponibilizadas através dos recursos interativos empregues. A análise conduzida permite ainda compreender que os mapas interativos são o terceiro tipo de representação visual de dados mais frequente, precedido pelo gráfico de barras interativo, não tendo sido identificado qualquer mapa estático. No total, apenas três representações visuais de dados estáticas foram assinaladas e surgem quando se verifica um número substancial (mais de cinco) de gráficos ou mapas interativos.

No que concerne à ferramenta utilizada para a criação destes gráficos e mapas, a plataforma *Flourish* foi a mais identificada nas representações visuais de dados publicadas, o que confirma parcialmente a hipótese formulada. Neste sentido, é possível compreender que o uso desta ferramenta de terceiros para criar as representações visuais de dados, comportam riscos devido às “dependências do fornecedor da ferramenta”, podendo ser descontinuadas pelo serviço e “quebrar” a ligação ao gráfico ou mapa publicados (Heravi et al., 2020: p.2085). Não obstante, o núcleo de jornalismo de dados conta com a colaboração da equipa de Infografia, que se constitui como o segundo elemento mais recorrente na criação das representações visuais de dados.

Foi também assinalado um caso em que a própria equipa de jornalismo de dados produziu uma representação visual de informações numéricas. Estas evidências mostram que embora o *Flourish* – uma plataforma de terceiros – seja a principal ferramenta utilizada neste âmbito, o jornal *Público* tem recursos próprios para apresentar visualmente essa informação.

Por último, a quarta pergunta de investigação (De que forma a equipa de jornalismo de dados do *Público* garante a transparência, precisão e imparcialidade dos dados que utiliza?) levou-nos rejeitar a hipótese formulada, uma vez que a equipa de jornalismo de dados do *Público* nem sempre refere a metodologia da análise adotada, a fonte da informação e em quase nenhuma ocasião é facultado o acesso direto à base de dados utilizada na análise das peças publicadas.

Esta constatação leva a entender que a equipa de jornalismo de dados do *Público* ainda não adotou a prática dos “dados abertos” como uma norma estabelecida e que significa “dar a possibilidade ao público de analisar a informação em bruto por si mesmo” e retirar as próprias conclusões (Lesage & Hackett, 2014; Weber et al., 2018). Do mesmo modo, é possível depreender que não existe ainda uma evidente “cultura de transparência”, que se pode traduzir na mera inclusão da hiperligação dos “dados em bruto” utilizados (Tandoc & Oh, 2017).

Contudo, sublinhe-se que em grande parte das peças evidenciou-se a contextualização dos dados, a fonte, as notas metodológicas e explicações das representações visuais de dados através das notas de rodapé, nos títulos ou em descrições. Além disso, Rui Barros mostrou possuir um sentido crítico apurado quanto à precisão dos cálculos realizados, metodologias adotadas e na criação das representações visuais de dados (Gray & Bounegru, 2019), como se pode ler no seguinte excerto:

“Sobre o trabalho dos rendimentos, reparei numa coisa: não faz sentido estares a calcular o número de pessoas de cada freguesia que recebem, por exemplo, o Rendimento Social de Inserção pelo total de pessoas que recebem isso a nível nacional. A intenção do trabalho é perceber que percentagem de pessoas de cada freguesia recebe esse subsídio face à sua população. Ou seja, tens de calcular com base no total da população da freguesia e não pelo total de pessoas que recebem o Rendimento Social de Inserção em todo o país.” (Rui Barros, conversa informal, [Dia 52](#)”).

Uma outra situação verificou-se quanto às questões éticas na utilização de dados que podem comprometer a privacidade dos cidadãos (Craig et al., 2017), tal como o mesmo refere:

“(…) Um aspeto que torna a história muito interessante é o nível da granularidade dos dados. Ou seja, quanto mais pormenorizados forem os dados, mais interesse o leitor vai ter porque vai relacionar com a estória e a informação vai dizer respeito à sua rua, freguesia ou concelho. Mas isso também levanta questões éticas, como deves entender. Por vezes temos de perceber que, por muito boa que aquela história pudesse ser, não podemos utilizar informação mais sensível por causa da questão da privacidade dos dados e assim. Aliás, o próprio INE não publica alguns desses dados” (Rui Barros, conversa informal, [Dia 19](#)).

Este profissional está também ciente dos riscos sobre a recolha de dados (Bradshaw, 2014) a partir de técnicas como o *scraping*:

“(…) Depois de colocares os códigos dos indicadores do INE que queres descarregar, como o do indicador geral, o da dimensão temporal e o da dimensão geográfica, tem aqui uma parte do código do *script* que diz “sleep”. O que isto faz é, no intervalo entre 1 e 10 segundos, dar uma pausa a sacar os dados. E tu perguntas-me: mas porquê dar essa pausa, se vai demorar mais tempo? Pois, imagina que corríamos o *script* sem dar pausas: o servidor do INE ia receber imensos pedidos de nossa parte para obter a informação. Isso podia prejudicar o funcionamento normal da página e não é ético prejudicar qualquer página ou o acesso à mesma, por parte de outras pessoas, só porque queremos muito aqueles dados (…)” Rui Barros, conversa informal, [Dia 45](#)).

Por último, na dimensão da imparcialidade dos dados, vale a pena notar que a equipa de jornalismo de dados possui uma abordagem diversificada das fontes consultadas e utilizadas, como são exemplo a colaboração com a Deco Proteste ou a criação de uma base de dados própria. No entanto, a maioria dos dados são provenientes de fontes governamentais, como o INE, os Censos ou o Eurostat.

Conclui-se, então, que o jornalismo de dados do *Público* é realizado por uma pequena equipa multidisciplinar com competências em áreas da linguagem de programação, de marcação, desenvolvimento Web, Estatística e design das representações visuais de dados, além das competências jornalísticas tradicionais. Entre estas competências, a capacidade de planeamento e organização são igualmente aspetos determinantes, bem como o tempo necessário para produzir conteúdos de jornalismo de dados – tal como se evidenciou no estudo etnográfico.

Além disso, esta investigação confirmou que a Web oferece um vasto leque de oportunidades que são aproveitadas para a execução e apresentação inovadora de peças jornalísticas de dados, entre as quais se destaca a abundância de ferramentas de acesso gratuito e a capacidade de explorar fóruns e recursos online. No entanto, surgiram desafios notáveis relacionados com a transparência, a precisão e a imparcialidade dos dados utilizados pela equipa de jornalismo de dados do Público. Embora se note esforços ao nível da contextualização da informação, da referência às fontes e se apresentem notas metodológicas, a prática dos "dados abertos" ainda não é uma norma estabelecida. A divulgação de bases de dados para permitir que o público faça a sua própria análise e conclusões dos dados é uma área que pode ser melhorada. Também a dependência de ferramentas de terceiros, como o *Flourish*, para criar representações visuais de dados, apresenta desafios.

No entanto, este estudo realça a importância do papel de jornalismo de dados como uma disciplina em constante evolução, capaz de aproveitar as potencialidades Web para contar histórias ricas e interativas, enquanto mantém um compromisso contínuo com a qualidade e a integridade jornalística, que norteiam este órgão de comunicação social.

Uma das fraquezas deste estudo é que se baseia no estágio realizado na redação do Público no Porto, durante um período específico, o que significa que a pesquisa pode não captar completamente a diversidade das práticas e abordagens dos autores do jornalismo de dados ao longo do tempo e em diferentes contextos dentro da organização. Embora o estudo tenha identificado questões relacionadas com a transparência, precisão e imparcialidade no jornalismo de dados do Público, não foi possível realizar uma análise aprofundada desses aspetos devido à falta de acesso às fontes de dados subjacentes, nem acompanhar todos os processos de tomada de decisão. Um futuro caminho de investigação poderá passar por abordar esta limitação através da realização de entrevistas detalhadas com os membros da equipa de jornalismo de dados e editores ou em compreender se o livro de estilo deste órgão comporta ou comportará alguma secção dedicada a este subdomínio do Jornalismo.

Uma área promissora para futuras investigações poderá passar pela análise do impacto e da eficácia das representações visuais de dados interativas em termos de engajamento e na compreensão da informação por parte do leitor. Isso poderia envolver a realização de estudos de caso detalhados ou questionários sobre peças jornalísticas específicas que utilizam representações visuais de dados interativas e recolher o *feedback* dos leitores de forma a avaliar como esses recursos contribuem para a experiência e compreensão da história.

## Referências Bibliográficas

- Alexandre, I. A. R. (2020). História Ilustrada do Jornalismo de Dados em Portugal: Exemplos de reportagens pioneiras que usaram dados como fonte primária de informação. In R. B. Simões, M. B. Marques, & J. Figueira (Eds.), *Media, informação e literacia: rumos e perspetivas* (pp. 511-534). (Conferências & Debates Interdisciplinares; Vol. 5). Imprensa da Universidade de Coimbra.
- Anderson, C.W. (2015). Between the Unique and the Pattern. *Digital Journalism*, 3(3), 349–363. <https://doi.org/10.1080/21670811.2014.976407>
- Anderson, C. W. (2017). Social survey reportage: Context, narrative, and information visualization in early 20th century American journalism. *Journalism*, 18(1), 81–100. <https://doi.org/10.1177/1464884916657527>
- Anderson, B., & Borges-Rey, E. (2019). Encoding the UX: User Interface as a Site of Encounter between Data Journalists and Their Constructed Audiences. *Digital Journalism*, 7(9), 1253–1269. <https://doi.org/10.1080/21670811.2019.1607520>
- Appelgren, E. (2018). An Illusion of Interactivity. *Journalism Practice*, 12(3), 308–325. <https://doi.org/10.1080/17512786.2017.1299032>
- Appelgren, E., Lindén, C., & Van Dalen, A. (2019). Data Journalism Research: Studying a Maturing Field across Journalistic Cultures, Media Markets and Political Environments. *Digital Journalism*, 7(9), 1191–1199. <https://doi.org/10.1080/21670811.2019.1685899>
- Appelgren, E., & Jönsson, A. M. (2021). Engaging Citizens for Climate Change—Challenges for Journalism. *Digital Journalism*, 9(6), 755–772. <https://doi.org/10.1080/21670811.2020.1827965>
- Appelgren, E. & Salaverría, R. (2018). The Promise of the Transparency Culture. *Journalism Practice*, 12(8), 986–996. <https://doi.org/10.1080/17512786.2018.1511823>
- Araújo, A. C. (2019). Gender and Numbers: Using Data from International Women’s Day Coverage on the Sites of Three Major Brazilian Newspapers. *Brazilian Journalism Research*, 15(1), 74–101. <https://doi.org/10.25200/BJR.v15n1.2019.1061>
- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2020). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, 21(7), 950–973. <https://doi.org/10.1177/1464884917700667>
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715594634>
- Baack, S. (2018). Practically Engaged. *Digital Journalism*, 6(6), 673–692. <https://doi.org/10.1080/21670811.2017.1375382>
- Bardin, L. (2016). *Análise de Conteúdo*. São Paulo: Almedina Brasil. Edições 70.
- Bastos, H. (2006). Ciberjornalismo: dos primórdios ao impasse. *Comunicação e Sociedade*, 9(10), 103–112. [https://doi.org/10.17231/comsoc.9\(2006\).1158](https://doi.org/10.17231/comsoc.9(2006).1158)

- Bastos, H. (2023). *História do Ciberjornalismo em Portugal. Os primeiros vinte e cinco anos.* Editora LabCom.
- Borges-Rey, E. (2016). Unravelling Data Journalism: A study of data journalism practice in British newsrooms. *Journalism Practice*, 10:7, 833-843. <https://doi.org/10.1080/17512786.2016.1159921>
- Borges-Rey, E. (2020). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*, 21(7), 915–932. <https://doi.org/10.1177/1464884917693864>
- Boyles, J. L., & Meyer, E. (2017). Newsrooms accommodate data-based news work. *Newspaper Research Journal*, 38(4), 428–438. <https://doi.org/10.1177/0739532917739870>
- Bradshaw, P. (2012). What is Data Journalism?. In Gray, J., Bounegru, L. & Chambers, L. (Eds). *The Data Journalism Handbook 1* (s/p). Disponível em: <https://datajournalism.com/read/handbook/one>
- Bradshaw, P. (2014). Data Journalism. In Zion, L., & Craig, D. (Eds.), *Ethics for Digital Journalists: Emerging Best Practices*, (p. 202-219). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9780203702567>
- Brolcháin, N. Ó., Porwol, L., Ojo, A., Wagner, T., Lopez, E. T., & Karstens, E. (2017). Extending Open Data Platforms with Storytelling Features. Association for Computing Machinery. <https://doi.org/10.1145/3085228.3085283>
- Coddington, M. (2015). Clarifying Journalism’s Quantitative Turn. *Digital Journalism*, 3:3, 331-348. <https://doi.org/10.1080/21670811.2014.976400>
- Colussi, J., & Gomes-Franco e Silva, F. (2017). Do jornalismo de dados à narrativa hipermídia: um estudo de caso dos jornais brasileiros Folha e Estadão. *index.Comunicación*, 7(3), 163–186. Disponível em: <https://indexcomunicacion.es/index.php/indexcomunicacion/article/view/318>
- Cunha, R. (2020). Journalism, Data Visualization, and Perception about Readers. *Brazilian Journalism Research*, 16(3), 526–549. <https://doi.org/10.25200/BJR.v16n3.2021.1309>
- Cushion, S., Lewis, J., & Callaghan, R. (2017). Data Journalism, Impartiality And Statistical Claims. *Journalism Practice*, 11, 1198 – 1215, <https://doi.org/10.1080/17512786.2016.1256789>
- Chadha, K., & Koliska, M. (2015). Newsrooms and Transparency in the Digital Age. *Journalism Practice*, 9(2), 215–229. <https://doi.org/10.1080/17512786.2014.924737>
- Charbonneau, J., & Gkouskou-Giannakou, P. (2015). “Data Journalism”, an investigation Practice? A glance at the German and Greek cases. *Brazilian Journalism Research*, 11(2), 244–267. <https://doi.org/10.25200/bjr.v11n2.2015.855>
- Córdoba-Cabús, A., & García-Borrego, M. (2020). Visualisations as a critical information source for data journalism. Analysis of the typology, interactivity, and functions in the 2019 Data Journalism Awards. *Doxa Comunicación. Revista Interdisciplinar De Estudios De Comunicación Y Ciencias Sociales*, (31), 87–105. <https://doi.org/10.31921/doxacom.n31a>
- Craig, D., Ketterer, S., & Yousuf, M. (2017). To Post or Not to Post: Online Discussion of Gun Permit Mapping and the Development of Ethical Standards in Data Journalism. *Journalism & Mass Communication Quarterly*, 94(1), 168–188. <https://doi.org/10.1177/1077699016684796>

- Cramer, J. M. & McDevitt, M. (2004). Ethnographic Journalism. In Iorio, S.H. (Ed.), *Qualitative Research in Journalism: Taking It to the Streets* (1st ed.) (pp. 127-144). Routledge.  
<https://doi.org/10.4324/9781410609557>
- Crawley, M. (2013). *The R Book*. 2nd Edition. Wiley. Disponível em:  
<https://www.cs.upc.edu/~robert/teaching/estadistica/TheRBook.pdf>
- De-Lima-Santos, M., & Mesquita, L. (2021a). Data Journalism Beyond Technological Determinism. *Journalism Studies*, 22(11), 1416–1435.  
<https://doi.org/10.1080/1461670x.2021.1944279>
- De-Lima-Santos, M., & Mesquita, L. (2021b). Data Journalism in favela: Made by, for, and about Forgotten and Marginalized Communities. *Journalism Practice*, 17(1), 108–126.  
<https://doi.org/10.1080/17512786.2021.1922301>
- De-Lima-Santos, M., & Salaverría, R. (2021). From Data Journalism to Artificial Intelligence: Challenges Faced by La Nación in Implementing Computer Vision in News Reporting. *Palabra Clave*, 24(3), 1–40. <https://doi.org/10.5294/pacla.2021.24.3.7>
- De-Lima-Santos, M., Schapals, A. K., & Bruns, A. (2021). Out-of-the-box versus in-house tools: how are they affecting data journalism in Australia?. *Media International Australia*, 181(1), 152–166. <https://doi.org/10.1177/1329878X20961569>
- De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., & Cam, F. L. (2015). Waiting for Data Journalism. *Digital Journalism*, 3(3), 432–446. <https://doi.org/10.1080/21670811.2014.976415>
- Dick, M. (2014). Interactive Infographics and News Values. *Digital Journalism*, 2(4), 490–506.  
<https://doi.org/10.1080/21670811.2013.841368>
- Du, Y. R., Zhu, L., & Cheng, B. K. L. (2019). Are Numbers Not Trusted in a “Post-Truth” Era? An Experiment on the Impact of Data on News Credibility. *Electronic News*, 13(4), 179–195.  
<https://doi.org/10.1177/1931243119883839>
- Duarte, A. G. (2021). Infografia digital: O género jornalístico que emergiu do trauma para preannunciar o rumo de uma nova cultura visual mediática. *Media & Jornalismo*, 21(39), 135-160.  
[https://doi.org/10.14195/2183-5462\\_39\\_7](https://doi.org/10.14195/2183-5462_39_7)
- Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, 14(4), 532–550. <https://doi.org/10.2307/258557>
- Felle, T. (2016). Digital watchdogs? Data reporting and the news media’s traditional ‘fourth estate’ function. *Journalism*, 17(1), 85–96. <https://doi.org/10.1177/1464884915593246>
- Fink, K. & Anderson, C. W. (2015). Data Journalism in the United States. *Journalism Studies*, 16:4, 467-481, <https://doi.org/10.1080/1461670X.2014.939852>
- Gehlen, M. & Sousa, J. (2018). *Jornalismo de Dados em Portugal: um estudo exploratório sobre práticas jornalísticas especializadas*. Repositório institucional da Universidade Nova de Lisboa.  
<http://hdl.handle.net/10362/63438>
- Gehlen, M. (2019). *As práticas do Jornalismo de Dados em Portugal*. (s.n.) Relatório apresentado à Universidade Fernando Pessoa como parte dos requisitos para o cumprimento do programa de Pós-Doutoramento em Ciências da Comunicação.  
<http://hdl.handle.net/10284/7375>

- Gehrke, M., & Benetti, M. (2020). Twitter as a News Source in Data Journalism. *Brazilian Journalism Research*, 16(3), 410–431. <https://doi.org/10.25200/BJR.v16n3.2021.1277>
- GitHub. (s/d). Olá, Mundo. GitHub Docs. <https://docs.github.com/pt/get-started/quickstart/hello-world>
- Gupta, K., Sampat, S., Sharma, M., & Rajamanickam, V. (2016). Visualization of election data: Using interaction design and visual discovery for communicating complex insights. *JeDEM - EJournal of EDemocracy and Open Government*, 8(2), 59–86. <https://doi.org/10.29379/jedem.v8i2.422>
- Gray, J., Bounegru, L. & Chambers, L. (2012). *The Data Journalism Handbook 1*. European Journalism Centre. Disponível em: <https://datajournalism.com/read/handbook/one>
- Gray, J., & Bounegru, L. (2019). *The Data Journalism Handbook 2*. European Journalism Centre and Google News Initiative. Disponível em: <https://datajournalism.com/read/handbook/two>
- Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, 15(6), 713–730. <https://doi.org/10.1177/1464884913486393>
- Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism*, 18(4), 408–424. <https://doi.org/10.1177/1464884915620205>
- Henke, J., Leissner, L., & Möhring, W. (2020). How can Journalists Promote News Credibility? Effects of Evidences on Trust and Credibility. *Journalism Practice*, 14(3), 299–318. <https://doi.org/10.1080/17512786.2019.1605839>
- Heravi, B. R., Cassidy, K., Davis, E., & Harrower, N. (2022). Preserving Data Journalism: A Systematic Literature Review. *Journalism Practice*, 16(10), 2083–2105. <https://doi.org/10.1080/17512786.2021.1903972>
- Hermida, A., & Young, M. (2017). Finding the data unicorn. *Digital Journalism*, 5(2), 159–176. <https://doi.org/10.1080/21670811.2016.1162663>
- Howard, A. B. (2014). *The art and science of data-driven journalism*. New York: Tow Center for Digital Journalism. <https://doi.org/10.7916/D8Q531V1>
- Jerónimo, P. (2014). *Ciberjornalismo de proximidade: A construção de notícias online na imprensa regional em Portugal*. Tese de Doutoramento. Porto: Faculdade de Letras da Universidade do Porto.
- Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16:1, 55-72. <https://doi.org/10.1080/14682753.2015.1015801>
- Landert, D., & Miscione, G. (2017). Narrating the stories of leaked data: The changing role of journalists after Wikileaks and Snowden. *Discourse, Context and Media*, 13-21. <https://doi.org/10.1016/j.dcm.2017.02.002>
- La-Rosa, L., & Sandoval-Martín, T. (2016). The Transparency Law's insufficiency for Data Journalism's practices in Spain. *Revista Latina de Comunicación Social*, 71, 1208-1229. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=5779773>

- Lesage, F., & Hackett, R. A. (2014). Between Objectivity and Openness—The Mediality of Data for Journalism. *Media and Communication*, 2(2), 42–54. <https://doi.org/10.17645/mac.v2i2.128>
- Lewis, N. P., & Waters, S. (2018). Data Journalism and the Challenge of Shoe-Leather Epistemologies. *Digital Journalism*, 6(6), 719–736. <https://doi.org/10.1080/21670811.2017.1377093>
- Lewis, N. P., McAdams, M., & Stalph, F. (2020). Data Journalism. *Journalism & Mass Communication Educator*, 75(1), 16–21. <https://doi.org/10.1177/1077695820904971>
- Lewis, S. C., & Usher, N. (2014). Code, Collaboration, And The Future Of Journalism. *Digital Journalism*, 2(3), 383–393. <https://doi.org/10.1080/21670811.2014.895504>
- Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3), 447–466. <https://doi.org/10.1080/21670811.2014.976418>
- Léchenet, A. (2014). Global Database Investigations: The role of the computer-assisted reporter. Reuters Institute Fellowship Paper. University of Oxford - Reuters Institute for the Study of Journalism. Disponível em: <https://reutersinstitute.politics.ox.ac.uk/our-research/global-database-investigations-role-computer-assisted-reporter>
- Lindgren, C. A. (2021). Writing With Data: A Study of Coding on a Data-Journalism Team. *Written Communication*, 38(1), 114–162. <https://doi.org/10.1177/0741088320968061>
- Link, E., Henke, J., & Möhring, W. (2021). Credibility and Enjoyment through Data? Effects of Statistical Information and Data Visualizations on Message Credibility and Reading Experience. *Journalism Studies*, 22(5), 575–594. <https://doi.org/10.1080/1461670x.2021.1889398>
- López-García, X., Rodríguez-Vázquez, A., & Pereira-Fariña, X. (2017). Technological skills and new professional profiles: Present challenges for journalism. [Competencias tecnológicas y nuevos perfiles profesionales: Desafíos del periodismo actual]. *Comunicar*, 53, 81–90. <https://doi.org/10.3916/C53-2017-08>
- Lowrey, W., Broussard, R., & Sherrill, L. A. (2019a). Data journalism and black-boxed data sets. *Newspaper Research Journal*, 40(1), 69–82. <https://doi.org/10.1177/0739532918814451>
- Lowrey, W., Sherrill, L., & Broussard, R. (2019b). Field and Ecology Approaches to Journalism Innovation: The Role of Ancillary Organizations. *Journalism Studies*, 20(15), 2131–2149. <https://doi.org/10.1080/1461670X.2019.1568904>
- Lugo-Ocando, J., & Brandão, R. (2016). Stabbing News. *Journalism Practice*, 10(6), 715–729. <https://doi.org/10.1080/17512786.2015.1058179>
- Martinho, A.P. (2013). *Jornalismo de Dados: Contributo para uma Caracterização do Estado da Arte em Portugal*. [Dissertação de Mestrado, ISCTE]. Repositório institucional do ISCTE – Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/8329>
- Martinho, A. P. (2014). Jornalismo de dados: caraterização e fluxos de trabalho. *Revista Exedra – Número 9 de 2014 - Comunicação e Ciências Empresariais*. Disponível em: <http://exedra.esec.pt/wp-content/uploads/2015/04/n9-B3.pdf>

- Mena, P. (2023). Reducing misperceptions through news stories with data visualization: The role of readers' prior knowledge and prior beliefs. *Journalism*, 24(4), 729–748.  
<https://doi.org/10.1177/14648849211028762>
- Meyer, P. (2002). *Precision Journalism. A Reporter's Introduction to Social Science Methods*. – 4th ed., ISBN: 978-0-7425-1088-3
- McAdams, M. (2014). Multimedia Journalism. In Zion, L., & Craig, D. (Eds.), *Ethics for Digital Journalists: Emerging Best Practices*, (p. 187-201). Routledge, Taylor & Francis Group.  
<https://doi.org/10.4324/9780203702567>
- McQuail, D. (2010). *McQuail's mass communication theory*. 6th edition. SAGE. ISBN 0-7619-6547-5.
- Ojo, A. & Heravi, B. (2017). Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences. *Digital Journalism*. 6. 1-26.  
<https://doi.org/10.1080/21670811.2017.1403291>
- Palomo, B., Teruel, L., & Blanco-Castilla, E. P. (2019). Data Journalism Projects Based on User-Generated Content. How La Nacion Data Transforms Active Audience into Staff. *Digital Journalism*, 7(9), 1270–1288. <https://doi.org/10.1080/21670811.2019.1626257>
- Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*, 15(6), 853–871. <https://doi.org/10.1177/1461444812463345>
- Parratt-Fernández, S., Mayoral-Sánchez, J & Mera-Fernández, M. (2021). “The application of artificial intelligence to journalism: an analysis of academic production”. *Profesional de la información*. <https://doi.org/10.3145/epi.2021.may.17>
- Pentzold, C., & Fechner, D. (2020). Data journalism's many futures: Diagrammatic displays and prospective probabilities in data-driven news predictions. *Convergence*, 26(4), 732–750.  
<https://doi.org/10.1177/1354856519880790>
- Pereira, F., Oliveira, L., & Zamith, F. (2013). A utilização da infografia nas secções de ciência e tecnologia de quatro jornais on-line nacionais. In Martins, M.L., & Veríssimo, J., *Comunicação global, cultura e tecnologia: livro de atas, 8º Congresso SOPCOM*. 780-787.  
<https://doi.org/10.34624/sopcom.v0i0.15511>
- Porlezza, C. & Splendore, S. (2019). From Open Journalism to Closed Data: Data Journalism in Italy, *Digital Journalism*, 7:9, 1230-1252. <https://doi.org/10.1080/21670811.2019.1657778>
- Robles, F. A., & López, P. (2021). Driving the Closest Information. *Local Data Journalism in the UK*. *Journalism Practice*, 15(5), 638–650. <https://doi.org/10.1080/17512786.2020.1749109>
- Rodríguez, J., & Clark, A., M. (2021). Big Data and Journalism: How American Journalism is Adopting the Use of Big Data. *Novum Jus*, 15(1), 69–89.  
<https://doi.org/10.14718/NovumJus.2021.15.1.4>
- Rogers, S. (2011). Data journalism at the Guardian: what is it and how do we do it?. *The Guardian*, 28 de julho de 2011. Disponível em:  
<https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism>

- Rogers, S., Schwabish, J., & Bowers, D. (2017, setembro). Data Journalism in 2017: The Current State and Challenges Facing the Field Today. Google News Labs. Disponível em: <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>
- Rosa, A. M. (2016). A comunicação e o fim das instituições: das origens da imprensa aos novos media. Lisboa: Media XXI.
- Rosling, H., Rosling, O., & Rönnlund, A. R. (2019). Factfulness – Factualidade - Dez razões pelas quais estamos errados acerca do mundo - e porque as coisas estão melhor do que pensamos. 10.ª edição. Temas e Debates - Círculo de Leitores.
- Sandoval-Martín, M. T., & La-Rosa, L. (2018). Big Data as a differentiating sociocultural element of data journalism: the perception of data journalists and experts. *Communication & Society*, 31(4), 193-208. <https://doi.org/10.15581/003.31.4.193-209>
- Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. <https://doi.org/10.1109/tvcg.2010.179>
- Smit, G. J., De Haan, Y., & Buijs, L. (2014). Visualizing News. *Digital Journalism*, 2(3), 344–354. <https://doi.org/10.1080/21670811.2014.897847>
- Stalph, F. (2018). Classifying Data Journalism. *Journalism Practice*, 12(10), 1332–1350. <https://doi.org/10.1080/17512786.2017.1386583>
- Stalph, F. (2020). Evolving data teams: Tensions between organisational structure and professional subculture. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720919964>
- Stalph, F., & Borges-Rey, E. (2018). Data Journalism Sustainability. *Digital Journalism*, 6(8), 1078–1089. <https://doi.org/10.1080/21670811.2018.1503060>
- Stalph, F. & Heravi, B. (2021). Exploring Data Visualisations: An Analytical Framework Based on Dimensional Components of Data Artefacts in Journalism. *Digital Journalism*. 1-23. <https://doi.org/10.1080/21670811.2021.1957965>
- Tabary, C., Provost, A.-M., & Trottier, A. (2016). Data journalism’s actors, practices and skills: A case study from Quebec. *Journalism*, 17(1), 66–84. <https://doi.org/10.1177/1464884915593245>
- Tandoc, E. C., & Oh, S. (2017). Small Departures, Big Continuities?. *Journalism Studies*, 18(8), 997–1015. <https://doi.org/10.1080/1461670x.2015.1104260>
- Tong, J., & Zuo, L. (2021). The Inapplicability of Objectivity: Understanding the Work of Data Journalism. *Journalism Practice*, 15(2), 153–169. <https://doi.org/10.1080/17512786.2019.1698974>
- Traquina, Nelson (2007). *O que é Jornalismo?* (2.ª Edição). Quimera Editores.
- Treadwell, G., Ross, T., Lee, A., & Lowenstein, J. K. (2016). A Numbers Game: Two Case Studies in Teaching Data Journalism. *Journalism & Mass Communication Educator*, 71(3), 297–308. <https://doi.org/10.1177/1077695816665215>

- Ureta, A. L., & Ferreras-Rodríguez, E. (2021). The potential of investigative data journalism to reshape professional culture and values. A study of bellwether transnational projects. *Communication & Society*, 34(1), 41–56. <https://doi.org/10.15581/003.34.1.41-56>
- Usher, N. (2020). News cartography and epistemic authority in the era of big data: Journalists as map-makers, map-users, and map-subjects. *New Media & Society*, 22(2), 247–263. <https://doi.org/10.1177/1461444819856909>
- Veglis, A., & Bratsas, C. (2017). Towards A Taxonomy of Data Journalism. *Journal of Media Critiques*, 3(11), 109–121. <https://doi.org/10.17349/jmc117309>
- Veglis, A., & Maniou, T. A. (2018). The Mediated Data Model of Communication Flow: Big Data and Data Journalism. *Kome*, 6(2), 32–43. <https://doi.org/10.17646/kome.2018.23>
- Weber, W., Engebretsen, M., & Kennedy, H. (2018). Data stories. Rethinking journalistic storytelling in the context of data journalism. *Studies in Communication Sciences*, 18(1), 191–206. <https://doi.org/10.24434/j.scoms.2018.01.013>
- Wright, S. M., & Doyle, K. (2019). The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies*, 20(13), 1811–1827. <https://doi.org/10.1080/1461670x.2018.1539343>
- Wright, K., Zamith, R., & Bebawi, S. (2019). Data Journalism beyond Majority World Countries: Challenges and Opportunities. *Digital Journalism*, 7(9), 1295–1302. <https://doi.org/10.1080/21670811.2019.1702472>
- Zamith, F. (2008). Uma Proposta Metodológica para analisar o Aproveitamento das Potencialidades Ciberjornalísticas da Internet. *Observatorio (OBS\*)*, 2(2). <https://doi.org/10.15847/obsOBS222008109>
- Zamith, R. (2019). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. *Digital Journalism*, 7(4), 470–489. <https://doi.org/10.1080/21670811.2018.1554409>

## Anexos

### Anexo 1 – Diário de campo

**Dia 1**, 2 de janeiro de 2023, segunda-feira, 10h10. Após a apresentação à direção do jornal e ao espaço, reunimo-nos com o jornalista de dados, Rui Barros, na Sala de Reuniões 2 a fim de discutir ideias de projetos a ser desenvolvidos, assim como as principais ferramentas ali utilizadas para a condução do trabalho nesta área, nomeadamente a linguagem R e o *RStudio*<sup>11</sup>.

O *RStudio* trata-se de um “ambiente de desenvolvimento integrado (IDE)” gratuito que opera sob a linguagem de programação R, sendo que a sua interface inclui “uma consola, um editor de realce de sintaxe que suporta a execução direta do código e ferramentas para a visualização de dados, histórico, depuração e gestão de espaços de trabalho”<sup>12</sup>. Em suma, esta ferramenta computacional permite realizar análises de dados, cálculos estatísticos, criar *scripts*, tabelas, gráficos, entre outras funções, através da linguagem R. O R trata-se de uma linguagem de programação especializada para a análise estatística de dados e para a construção de gráficos e outras representações visuais de dados (Crawley, 2013).

O jornalista acrescenta que as principais funções de código que iremos utilizar para análise serão as mais conhecidas, como “*filter()*”, “*select()*”, “*arrange()*”, “*mutate()*”, “*summarize()*” e “*group\_by()*”, enquanto que, para visualizar os dados, seria utilizada o pacote “*ggplot*”.

Durante a conversa, apontamos algumas temáticas que gostaríamos de tratar e Rui Barros refere que há também alguns projetos que gostaria de ver realizados ao longo dos próximos meses, como uma peça que desse a conhecer o estado das obras previstas no projeto “Ferrovia 2020”, os tempos de espera nas unidades hospitalares do Serviço Nacional de Saúde, a mortalidade nas estradas portuguesas, a análise às principais palavras ou assuntos que são discutidos no Parlamento português, a

---

<sup>11</sup> Os conhecimentos básicos para operar com este programa foram adquiridos previamente pela recomendação feita pelo jornalista encarregue da orientação do estágio (Rui Barros).

<sup>12</sup> Uma descrição mais detalhada sobre o *RStudio* está disponível em: <https://posit.co/products/open-source/rstudio/> [consultado pela última vez a 28 de julho de 2023]

atualização do “Prometómetro<sup>13</sup>”, entre outros. “As ideias que trazes são interessantes, mas é necessário encontrar o ângulo certo para a estória de forma a convencer que tem potencial. Quero com isto dizer que não basta termos os dados e apresentá-los. Tens um bom exemplo disso com o trabalho que apresentamos no final do ano passado, o “Somos 8 mil milhões”<sup>14</sup>, em que a partir dos dados da população mundial criamos uma peça interativa, que envolve e capta a atenção do leitor, apresentando várias perspetivas da informação reunida. Entretanto, queria que assumisses a atualização semanal do cabaz dos alimentos” (Rui Barros).

A peça “Conheça o impacto da inflação no seu carrinho de compras”<sup>15</sup> é-nos apresentada como sendo um “tracker” semanal (ou registo semanal) da evolução do preço do cabaz de produtos essenciais. Este artigo no site do *Público* contém dados sobre o preço do cabaz alimentar desde o dia 5 de janeiro de 2022 e é atualizado semanalmente desde 2 de março de 2022, tendo como objetivo acompanhar o impacto da inflação no custo de vida em Portugal. Rui Barros explica que os dados são fornecidos semanalmente, às quintas-feiras, pela Deco Proteste, através de um ficheiro Excel, com embargo até às 12h00 de sexta-feira.

A inserção dos dados faz-se a partir de um repositório do *GitHub* que, através de alguns ficheiros e à distância de um clique, atualiza automaticamente os dados da notícia publicada na página do *Público*. Rui Barros explica que a atualização semanal dos dados é efetuada em cinco ficheiros CSV distintos e com finalidades próprias: “prices\_total.csv” (ficheiro que contém o preço semanal do cabaz alimentar); “prices\_category.csv” (variação do preço nas categorias de mercearia, carne, peixe, laticínios, frutas e legumes, congelados); “prices\_product.csv” (variação do preço face ao período homólogo); “prices\_product\_last\_week.csv” (variação do preço face à semana anterior) e o “update.csv” (data de atualização dos dados). Com a exceção deste último, todos os ficheiros continham a data de registo dos dados, os valores totais, a diferença total e percentual do preço dos produtos. Além disso, este jornalista

---

<sup>13</sup> O artigo está disponível em: <https://www.publico.pt/interactivos/prometometro/> [consultado pela última vez a 26 de julho de 2023]

<sup>14</sup> O artigo está disponível em: <https://www.publico.pt/interactivo/8-mil-milhoes> [consultado pela última vez a 27 de julho de 2023]

<sup>15</sup> O artigo está disponível em: <https://www.publico.pt/interactivo/inflacao-portugal> [consultado pela última vez a 27 de julho de 2023]

teve o cuidado de assegurar a coerência semântica e gramatical do texto com a criação de uma variável para inserir o artigo definido “o” ou “a”, de forma a identificar corretamente o produto alimentar (**Erro! A origem da referência não foi encontrada.**). “Quando clicares em ‘update’ os dados vão ser atualizados automaticamente na página do artigo, o que faz com que os gráficos de linhas, que registam a evolução, se alterem, assim como excertos do texto que identificam o alimento que registou o maior aumento do preço naquela semana ou face ao período homólogo. Aqui o problema é que tens de estar atento aos dados que inseres, pois tem de ser feito manualmente” (Rui Barros, conversa informal).

Quando terminássemos esta tarefa, fomos incentivados a procurar autonomamente por informação na Internet sobre algumas das ideias por nós propostas, tais como o consumo de antidepressivos em Portugal. A pesquisa pelo assunto, através do motor de pesquisa da *Google*, levou-nos até ao site “Our World in Data”, onde Portugal se destacava como um dos países europeus com o maior consumo diário de antidepressivos por cada 1000 pessoas. Após ter sido comunicada esta “descoberta”, Rui Barros assinalou a necessidade de “encontrar um foco”, uma forma de “vender” a notícia e a história a partir dos dados.

Figura 7 - Dados do cabaz alimentar num repositório do Público na plataforma GitHub

	article	produto	date_begin	price_begin	date_end	price_end	diff	var_txt
1								
2	a	Cebola	2022-02-23	1.05	2023-04-19	1.86	0.81	77
3	o	Arroz carolino	2022-02-23	1.14	2023-04-19	1.98	0.84	73
4	a	Polpa de tomate	2022-02-23	0.89	2023-04-19	1.51	0.62	70
5	a	Cenoura	2022-02-23	0.77	2023-04-19	1.28	0.62	70
6	o	Azeite virgem	2022-02-23	4.65	2023-04-19	7.27	2.62	56
7	os	Flocos de cereais	2022-02-23	1.82	2023-04-19	2.81	0.99	54
8	o	Açúcar branco	2022-02-23	1.11	2023-04-19	1.69	0.58	52
9	o	Carapau	2022-02-23	3.32	2023-04-19	4.99	1.67	50
10	a	Batata vermelha	2022-02-23	0.91	2023-04-19	1.36	0.45	49
11	a	Couve-coração	2022-02-23	1.04	2023-04-19	1.54	0.50	49

Legenda: Ambiente de trabalho no *GitHub*: repositório dos dados com a maior subida percentual em relação ao período homólogo (variáveis da esquerda para a direita: artigo definido, nome do produto, data inicial, preço relativo à data inicial, data atual, preço relativo à data atual, diferença do preço, variação percentual do preço).

**Dia 2, 3 de janeiro de 2023, terça-feira. 10h05.** Nas primeiras horas da manhã, a redação conta com poucos jornalistas. Rui Barros é um dos profissionais presentes e, apercebendo-se da nossa chegada, convida-nos a sentar perto dele. No ecrã que se encontra a utilizar é possível visualizar uma peça jornalística de um outro órgão de comunicação social que exemplifica uma das ideias abordadas na reunião do dia anterior.

“Lembraste daquilo que falamos ontem sobre o Parlamento português? Pronto, nesta peça o que fizeram foi pegar em todas as atas disponíveis e realizaram uma análise textual para encontrar os assuntos ou as palavras que dominaram cada ano na Câmara dos Representantes. Tenta encontrar o registo dos debates parlamentares de Portugal para fazer algo idêntico. Para isso, tens de pensar e criar uma forma eficaz de descarregar essa informação, porque manualmente vai levar uma eternidade”. (Rui Barros, conversa informal).

Nesta altura, apercebemo-nos de que seria necessário recorrer a algum tipo de *software* ou *script* que permitisse descarregar a informação disponível no Arquivo Histórico-Parlamentar automaticamente. Contudo, não sabíamos de que forma o fazer. Constatada essa dificuldade, Rui Barros começou por sugerir o contacto com a instituição supramencionada para saber se os documentos poderiam ser facultados. Na chamada telefónica efetuada, foi-nos explicado que a única forma de os descarregar seria manualmente, através da própria página. Comunicamos o que nos foi dito ao colega. “Pois, isso é um problema... Faz uma pesquisa e vê se consegues descobrir como podemos descarregar isso. Posso já adiantar que, muito provavelmente, terás de usar ou criar um *script* que te vá buscar automaticamente cada um dos documentos. Vai funcionar como um *scraper*. Está na altura de pesquisares na Internet e trabalhar com o R” (conversa informal, Rui Barros).

Embora não estivéssemos familiarizados com este *software*, tínhamos já um conhecimento básico da linguagem de programação R e da interface do *RStudio*, na ótica do utilizador. Seguidas as instruções de Rui Barros, uma pesquisa pela Internet levou-nos a conhecer outros *softwares* que poderiam ajudar a alcançar o pretendido. Mas não fomos bem-sucedidos. No final desse dia, Rui Barros abordou-nos para saber o ponto da situação:

“- Então? Conseguiu encontrar alguma coisa? (Rui Barros)

- Estive a procurar e a experimentar algumas ferramentas que encontrei, mas não está muito fácil... (Rafael Oliveira)

- Ok, sem problema. Quando estiveres farto de tentar resolver esse assunto, deixo-te aqui um livro que será bastante útil para realizar a análise textual (“Text Mining with R: A Tidy Approach”, de Julia Silge & David Robinson<sup>16</sup>). É muito acessível, também está disponível *online* e tem um tutorial que ajuda a perceber para que serve cada função” (Rui Barros). (conversa informal, Rui Barros).

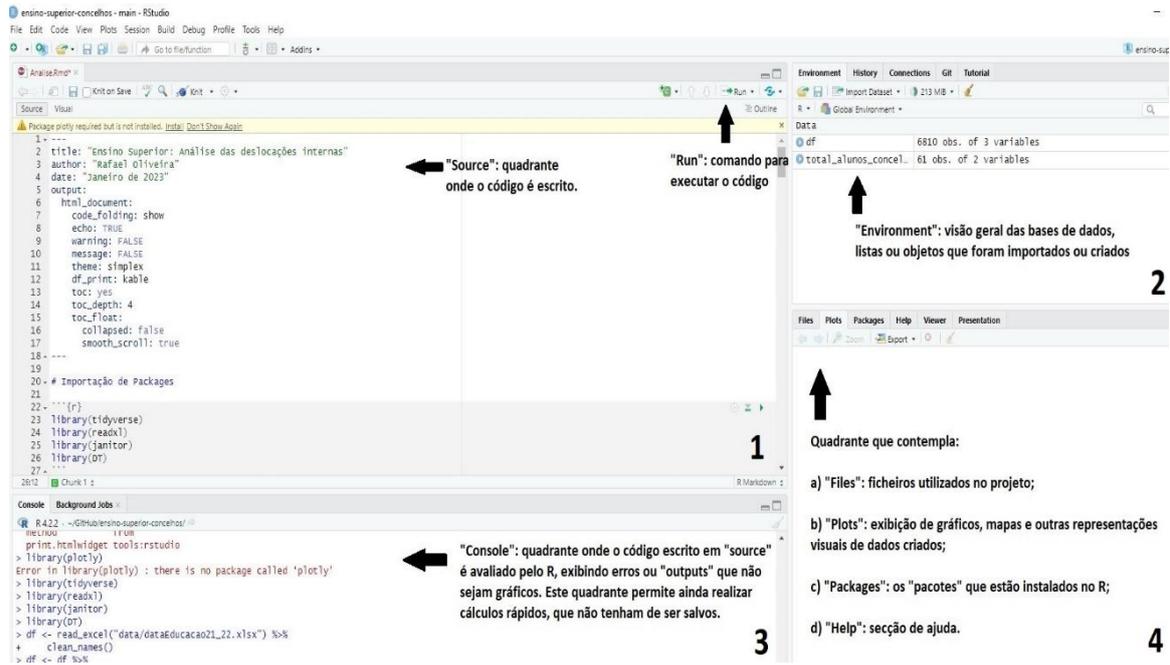
Antes da despedida, o colega sugeriu ainda que, no dia seguinte, procurasse encontrar dados que mostrassem a proveniência dos alunos portugueses que ingressaram no primeiro ano do Ensino Superior, para tentar descobrir quais os concelhos “mais atrativos” para os novos alunos.

**Dia 3**, 4 de janeiro de 2023, quarta-feira. Novo dia traz novos desafios. A nossa manhã foi dedicada à procura de dados sobre o Ensino Superior. Na página da Direção-Geral de Estatísticas da Educação e Ciência, foi possível encontrar e descarregar uma base de dados que continha a informação necessária. Com a base de dados na nossa posse, passamos para o *RStudio* a fim de analisar a informação. O primeiro passo na utilização deste *software* passa por criar um “novo projeto” e um ficheiro que permita escrever o código para instalar e carregar os “pacotes” (tradução do inglês “packages”) de forma a analisar os dados. Isto é efetivado através da redação de linhas de código em um dos quatro painéis disponíveis.

---

<sup>16</sup> O livro “Text Mining with R: A Tidy Approach” também está disponível online em: <https://www.tidytextmining.com/index.html> [consultado pela última vez a 27 de julho de 2023]

Figura 8 – Visão geral da interface padrão do *RStudio*



Legenda: Explicação de cada um dos quadrantes do *RStudio* - 1) "Source": quadrante onde o código é escrito e "Run": comando para executar o código; 2) "Environment": visão geral das bases de dados, listas ou objetos que foram importados ou criados; 3) "Console": quadrante onde o código escrito em "source" é avaliado pelo R, exibindo erros ou "outputs" que não sejam gráficos. Este quadrante permite ainda realizar cálculos rápidos que não tenham de ser salvos. 4) "Quadrante que contempla: a) "files": ficheiros utilizados no projeto; b) "plots": exibição de gráficos, mapas e outras representações visuais de dados criados; c) "packages": os "pacotes" que estão instalados no R; d) "Help": secção de ajuda.

Após instalar e carregar os pacotes necessários para a importação e análise de dados, o documento deve ser importado para o *RStudio*, através de um comando de código próprio. Há, contudo, a necessidade de verificar se a base de dados se encontra corretamente formatada. Quer isto dizer que, na eventualidade de uma base de dados ter linhas ou colunas em branco, será necessário proceder à sua "limpeza". Esta fase no fluxo de trabalho permite eliminar linhas ou colunas vazias da base de dados que podem comprometer a análise. Também é possível excluir variáveis que sabemos, de antemão, que não serão utilizadas.

Rui Barros explica que este processo pode ser realizado no *RStudio*, através de código escrito, ou em outros *softwares* de análise de dados menos sofisticados e mais intuitivos para o utilizador, como *Excel* ou o *Google Sheets*. Mas, então, porquê utilizar

a linguagem R e o *RStudio*? “A vantagem do *RStudio*, comparativamente ao *Excel*, *SPSS* ou *Google Sheets*, é que se precisares de alterar alguma coisa ou recuperar alguma parte do processo de seleção ou da análise dos dados, só tens de ir atrás e modificar o que queres no código. É muito rápido. Além disso, muito do código que escreves e utilizas pode ser reaproveitado para outros projetos. O *RStudio* também te permite fazer rapidamente gráficos, é gratuito e é o mais indicado para analisar grandes bases de dados. O *Excel* nesses aspetos é mais limitado”. (Rui Barros, conversa informal).

Com a seleção das variáveis efetivada, damos seguimento à análise de dados, procurando aferir o número total de alunos que ingressaram nas licenciaturas e mestrados integrados. A meio da tarde, já com breves conclusões retiradas a partir dos cálculos e análises efetuados, somos interpelados pelo jornalista Rui Barros. “Vai haver agora uma reunião da equipa do Online, acho que deverias estar presente para seres introduzido, conheceres os editores de Lisboa e os restantes colegas”. (Rui Barros, conversa informal). Esta reunião por videoconferência decorreu através da plataforma *Microsoft Teams*, com cerca de duas dezenas de participantes que deram a conhecer os trabalhos que estão a desenvolver. Após a discussão de alguns trabalhos em curso, os editores dão-nos as boas-vindas à equipa e mostram-se disponíveis para qualquer esclarecimento ou proposta de trabalho.

Já perto do final do dia, aproveitamos o momento para esclarecer com Rui Barros algumas dúvidas conceptuais sobre o projeto em mãos. O mesmo explicou que o objetivo do trabalho passava por perceber, através dos dados e, posteriormente, com a visualização dessa informação num mapa, quais os municípios que registaram um maior fluxo de entrada e saída de alunos, sendo necessário definir um critério para essa medição. “Tens de formular uma questão de investigação que será o teu ponto de partida, por exemplo: é possível que Lisboa e Porto sejam os principais focos de afluência? Acredito que já tenhas chegado a essa conclusão, portanto, o mais importante agora é calcular o peso relativo de cada concelho. Ou seja, saber que percentagem de alunos do concelho X, Y e Z vieram estudar para o Porto, por exemplo” (Rui Barros, conversa informal).

**Dia 4**, 5 de janeiro de 2023, quinta-feira. Rui Barros não se encontra na redação de manhã, sendo previsível a sua chegada por volta das 13h45. Aproveitamos a manhã para criar uma base de dados no *RStudio* que permitisse fazer os cálculos sugeridos pelo colega. No entanto, ainda subsistiam algumas dúvidas sobre como operar com este software, assim como efetuar os cálculos e a forma de visualizar essa informação num mapa. Quando Rui Barros chega, aproveitamos para lhe expor a situação. No *RStudio*, o jornalista rapidamente escreve as linhas de código que permitiram montar uma nova base de dados que agrega o número de alunos de cada município, o número de alunos que foi estudar para cada concelho, e o peso relativo calculado para cada região. De forma a calcular o peso relativo, Rui Barros explica que é necessário dividir o número de alunos do seu concelho de origem pelo número total de alunos que o outro concelho acolheu. Quanto à parte de visualizar essa informação num mapa, foi-nos passada uma hiperligação que continha um documento *shapefile*. “Um *shapefile* basicamente é um conjunto de dados de vetor com muitos pontinhos e, todos juntos, formam aquelas linhas que delimitam as localizações geográficas dos concelhos, freguesias e assim. Em Portugal, somos divididos pelas NUT’s (Nomenclatura das Unidades Territoriais) e aqui interessa-nos aquela que divide Portugal por concelhos. Vou-te passar esse ficheiro.” (Rui Barros, conversa informal). O colega, entretanto, sai para almoçar. Embora já tenhamos o ficheiro em mãos e o tenhamos carregado no *RStudio*, subsistem algumas dúvidas sobre como visualizar a informação, tornando-se um desafio difícil de superar.

Com parte da tarde ocupada a tentar resolver este problema, Rui Barros procura saber o ponto de situação. Após ler as linhas de código escritas, o jornalista de dados mostra-se agradado com as tentativas realizadas para visualizar a informação num mapa, mas alerta para a necessidade de adotar “boas práticas” na análise de dados.

“Isto, no fundo, é um pouco como nas investigações académicas. No jornalismo de dados, assim como na academia, há uma preocupação sobre a reprodutibilidade dos resultados. Então, para isso, convém enunciarmos por escrito, no próprio documento da análise, sobre o que se trata o trabalho, quais as hipóteses que queremos testar, de onde é que vêm os dados, que categorias foram excluídas e o porquê, entre outras

coisas. Além disso, para eu acompanhar o teu processo é mais fácil carregares isto no *Git*. Vou-te mostrar como se faz isso” (Rui Barros, conversa informal).

Em conjunto, foram formulados os objetivos e as hipóteses a testar sobre o projeto e é-nos feita a introdução a uma nova funcionalidade do *RStudio*: o *Git*. Esta ferramenta trata-se de um sistema de controlo de versões distribuídas que permite que um grupo de pessoas possa trabalhar no mesmo projeto ao mesmo tempo, sem que a sua versão seja comprometida pelo trabalho de outro colega. Desta forma, é possível registar o histórico de edições de qualquer participante no projeto ou descarregar rapidamente a versão do projeto que pretendemos. À saída do trabalho, Rui Barros informa que amanhã estará em teletrabalho, mas que a comunicação se pode realizar através do *Discord* ou por chamada telefónica.

**Dia 5**, 6 de janeiro de 2023, sexta-feira. Sexta-feira é dia de atualização dos dados do cabaz alimentar e, após isso estar concluído, Rui Barros chama-nos para uma chamada no canal de *Discord*.

“- Olá, Rafael, tudo bem? Olha, estive a ver os dados desta semana e acho que há aqui um assunto interessante para pegares. Não sei se reparaste, mas no espaço de um ano o preço do arroz carolino aumentou 80%... (Rui Barros)

- Sim, por acaso tinha reparado nisso. É um pouco estranho, não é? (Rafael Oliveira).

- Sim, é muito estranho. Como é que um produto que é feito em Portugal aumentou tanto? Ok, estamos a viver um período de inflação, mas porque é que o arroz carolino subiu tanto e, por exemplo, o tomate não? Tenta descobrir isso, acho que pode dar uma história gira. Vou falar aqui com um colega para ver se me arranja o contacto de um agricultor que produza arroz e já te digo qualquer coisa.” (Rui Barros). (conversa informal através de uma chamada de voz no *Discord*).

Para realizar este trabalho jornalístico, Rui Barros sugeriu contactar, em primeiro lugar, a fonte dos dados (Deco Proteste) para perceber como eram realizados os cálculos para o preço médio do arroz carolino nos hipermercados, pois poderia tratar-se de um “acaso estatístico”. De seguida, foi encetados uma série de contactos por via telefónica e endereço eletrónico a fim de trocar algumas impressões e agendar entrevistas.

Por volta das 16h30, regressamos ao projeto do Ensino Superior, mas as dúvidas voltaram a surgir, o que comprometeu o desenvolvimento da análise. Tentou-se esclarecer as dúvidas com Rui Barros, através da partilha de ecrã, mas o nosso computador pessoal era pouco eficiente quando várias ferramentas de trabalho eram usadas em simultâneo (por exemplo, *Google Chrome*, *Discord*, *RStudio*), complicando qualquer tipo de interação. Neste sentido, Rui Barros procurou saber se José Volta e Pinto poderia ajudar com a dúvida que tínhamos, mas o mesmo revelou não saber resolver rapidamente a situação e, além disso, estava encarregue pela atualização “Ao Minuto” na página do jornal. Foi, então, sugerido pelo colega que anotasse as dúvidas no próprio documento da análise e que, posteriormente, seriam esclarecidas.

Quanto à tentativa de aceder aos discursos do Parlamento foi dito ao Rui Barros que não havíamos conseguido, devido à complexidade da tarefa que, na semana seguinte, foi também por ele reconhecida.

No decorrer desta primeira semana do estudo, José Volta e Pinto esteve presente na maioria dos dias na redação, mas a interação com o mesmo não foi além das primeiras apresentações. No mesmo sentido, não se evidenciou qualquer interação do jornalista na área do jornalismo de dados.

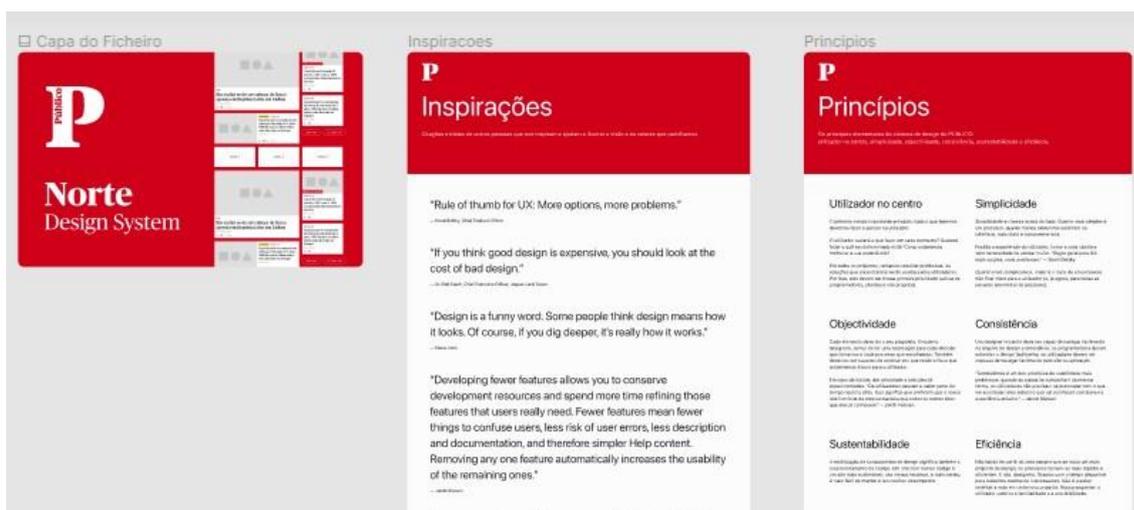
**Dia 6**, 9 de janeiro de 2023, segunda-feira, 11h30. Com dois projetos em mão – a análise ao do Ensino Superior e a inflação do preço do arroz carolino -, o diretor-adjunto, Amílcar Correia, solicitou-nos, por *e-mail*, a redação de uma notícia sobre uma exposição artística em S. João da Madeira. Ainda que fora do âmbito deste estudo, foi a partir deste artigo que tivemos o primeiro contacto com o *backoffice* do *Público*, onde aprendemos a incorporar a imagem de destaque da notícia, galeria de fotos e hiperligações com o editor do *Online*, Ivo Neto. Aproveitamos o início da tarde para realizar uma breve entrevista telefónica com a curadora da exposição e a hora seguinte para redigir o texto.

Já na reta final do dia, avançamos com a entrevista ao orizicultor, através do contacto que Rui Barros nos forneceu. Após termos explicado e comentado com o colega as justificações dadas pelo entrevistado, o mesmo tirou um tempo para explicar que os gráficos nas peças publicadas no site do *Público* seguem também um “livro de estilo”,

nos quais são adotadas cores específicas, tipos e tamanhos de letra, a fonte dos dados, título e subtítulo, e a utilização da vírgula como separador decimal.

“No *Público* utilizamos o *Flourish* para criar os gráficos publicados *online*. Sei que nunca trabalhaste com isto, mas é uma ferramenta bastante completa, intuitiva e a página deles tem um bom tutorial para cada tipo de gráfico a utilizar. (Rui Barros, conversa informal). A paleta de cores utilizada pelo *Público* para a criação de gráficos foi-nos disponibilizada através de um ficheiro PDF partilhado no *Notion*. Além disso, o colega também nos remeteu uma hiperligação para a página *Figma* do *Público* que elenca os princípios orientadores do sistema de design do *Público*, nomeadamente: utilizador no centro, objetividade, simplicidade, consistência, sustentabilidade e eficiência.

Figura 9 – Visualização geral da página *Figma* do jornal *Público*



Legenda: A página *Figma* do *Público*, onde são descritos os princípios elementares do sistema de design *online* do jornal.

**Dia 9**, 12 de janeiro de 2023, quinta-feira. Os últimos dias foram dedicados à realização de entrevistas, redação do texto e a criação dos gráficos da peça sobre o aumento do preço do arroz carolino. De facto, o *Flourish* é uma ferramenta intuitiva e acessível, mas que também se pode tornar complexa.

Para esta peça, o processo da criação do primeiro gráfico resumiu-se à importação dos dados da Deco Proteste e à personalização do mesmo seguindo o exemplo de trabalhos anteriormente publicados. A página do *Flourish* comporta duas abas: uma dedicada à pré-visualização do gráfico, onde é possível editar esteticamente o gráfico, e a outra para a importação dos dados e respetiva edição e/ou seleção das variáveis a ser incluídas (ver Figura 10 e Figura 11).

Figura 10 – Ambiente de trabalho no site do *Flourish* – na aba “preview”

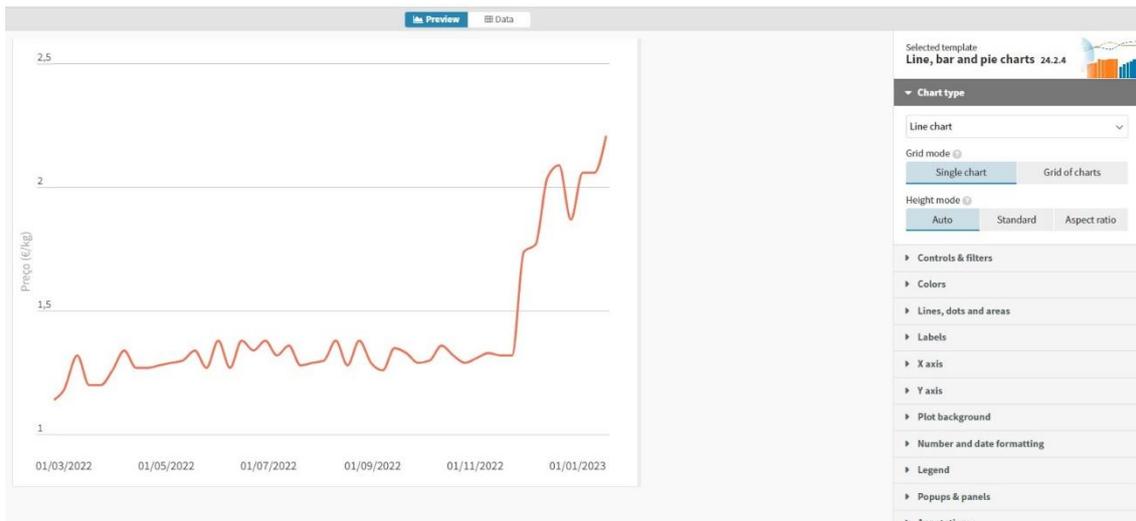


Figura 11 – Ambiente de trabalho no *Flourish* – na aba “data”

The screenshot shows the 'data' tab of the Flourish interface. It features a data table with the following columns: 'Data', 'ano', and 'Preço do arroz carolino'. The table contains 16 rows of data. To the right of the table is a settings panel for data visualization:

- Upload data: (button)
- SELECT COLUMNS TO VISUALISE:
  - Labels/time: (B)
  - Values: (C)
  - Charts grid: (purple)
  - Row filter: (blue)
  - Info for custom popups: (A)

O jornalista Rui Barros encontra-se focado num outro projeto que, aparentemente, lhe tem vindo a ocupar uma parte substancial do seu tempo. Numa das visitas ao seu

posto de trabalho, encontramos no seu ecrã uma página com o título “Contas à vida: enfrentar a inflação”, na qual é possível verificar um aspeto diferenciador.

“- Olá, Rui. Desculpa a curiosidade, mas o que é isso? (Rafael Oliveira)

- É a minha dor de cabeça dos últimos tempos [risos]. Não, é um trabalho que estamos a preparar para lançar em breve, só que o desenvolvimento Web disto não está a ser muito fácil. Mas o que precisavas? (Rui Barros) (conversa informal, Rui Barros).

Aproveitamos o momento para lhe perguntar se já tinha tido a oportunidade de rever o texto por nós redigido. A resposta é afirmativa, mas são feitas várias observações sobre o texto, evidenciando a sua sensibilidade relativamente à organização de ideias e da clareza do texto. “O texto precisa de umas afinações. O teu primeiro parágrafo tem de responder logo à questão: Porquê que o arroz carolino já custa o dobro? Tens aqui uma série de fatores, mas o texto está extenso... Temos de organizar e sintetizar bem a informação para o leitor ficar logo agarrado à história. Antes de escreveres, faz o exercício de contar a história a um amigo teu”. (Rui Barros, conversa informal).

Já no final do dia, Rui Barros entrega-nos um novo livro (“*The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*”). “Esse livro é um guia do *The Wall Street Journal*. É muito bom para compreender quais gráficos devemos utilizar e em que situações. Além disso tem algumas regras para a utilização das cores, os limites dos eixos e assim. Acho que te pode ajudar bastante no futuro” (Rui Barros, conversa informal). O jornalista pede ainda para atualizarmos os dados da Deco Proteste, por volta das 21h00, em casa, uma vez que outro órgão de comunicação social já tinha publicado essa informação. Chegados a casa, consultamos o ficheiro *Excel* providenciado pela Deco Proteste, mas apresentava alterações na sua estrutura normal, o que levantou problemas com o código escrito por Rui Barros: parte dos dados enviados não estavam a coincidir com a componente textual do artigo atualizado. O jornalista referiu que, em breve, iria corrigir a situação.

**Dia 10**, 13 de janeiro de 2023, sexta-feira. A conceção dos gráficos para a peça do arroz carolino estão em curso, mas Rui Barros refere que a sua preocupação ainda incide no texto. Constata a necessidade de explicar o “aumento exponencial do preço” registado a partir de novembro de 2022. Após contactarmos uma das fontes consultadas, a

justificação surge e é integrada no texto. Face a estas últimas alterações, Rui Barros relembra-nos a necessidade de “agrupar melhor as ideias”, de “responder de imediato à pergunta” e de “dizer mais, mas com menos palavras” (Rui Barros, conversa informal). Ao longo da tarde, procuramos aprimorar o texto, assim como as representações visuais construídas no *Flourish*, a partir dos dados da Deco Proteste e do Eurostat. Relativamente a esta última instituição, foram consultados os dados sobre os níveis de importação e exportação de arroz por sugestão de uma das fontes entrevistadas.

Por volta das 17h00, procuramos avançar na análise do Ensino Superior. Após consultarmos vários tutoriais na Internet, conseguimos encerrar o dia com a criação de um mapa com o peso relativo de cada concelho.

**Dia 13**, 18 de janeiro de 2023, quarta-feira. Os últimos dias foram passados a ultimar pormenores visuais sobre os gráficos da peça do arroz carolino, como a cor dos eixos (X e Y). Entretanto, o texto foi revisto por Rui Barros e enviado para o editor Ivo Neto, que o enquadrou na secção de “Economia”.

Rui Barros está em teletrabalho e continua atarefado com o projeto que tem em mãos. Através de uma chamada de voz no *Discord*, contamos-lhe o ponto de situação das análises, mas informa-nos que, até ao momento, não lhe foi possível rever as dúvidas apontadas no projeto dos dados do Ensino Superior, nem aos gráficos elaborados sobre o trabalho do arroz carolino. Neste seguimento, pede-nos para avançarmos com uma outra análise de dados que cruzasse os rendimentos médios de cada concelho com o preço das rendas. É-nos indicado o site do INE para recolher esses dados. Contudo, a busca por esta informação no INE não foi uma tarefa fácil, dada a complexidade das várias bases de dados sobre o assunto. Face a isto, o jornalista de dados aconselhou ler primeiramente a “metainformação” associada ao(s) indicador(es) selecionado(s).

**Dia 14**, 19 de janeiro de 2023, quinta-feira, 11h00. Os dados da Deco Proteste estão atualizados e o resto do período da manhã será dedicado à recolha dos dados sobre a nova proposta de trabalho.

À hora de almoço, por volta das 13h30, tivemos oportunidade de dialogar com o jornalista José Volta e Pinto, ficando a conhecer um pouco melhor o seu percurso

académico e profissional e algumas das suas competências na área do jornalismo de dados.

“– Olá, posso sentar-me aqui para almoçar? (José Volta e Pinto)

- Olá, claro que sim. Tu és o outro colega da equipa de jornalismo de dados, não é? (Rafael Oliveira).

– Sim, sou o Volta! Tu deves ser o Rafael, certo? O Rui [Barros] disse-me que tu escolheste fazer um estágio em jornalismo de dados. É verdade? (José Volta e Pinto).

- Sim, acho que é uma área interessante. Nunca tive a oportunidade de a explorar muito, embora já seja jornalista há algum tempo. Decidi aproveitar esta ocasião para aprender algo mais. (Rafael Oliveira)

- Claro, fizeste bem. Acho interessante teres escolhido esta área. Por acaso acho que nunca tivemos um estagiário em jornalismo de dados por aqui. (José Volta e Pinto)

- Pois, acredito. Diria que é uma área relativamente recente. Tu tiveste alguma unidade curricular na licenciatura sobre jornalismo de dados? (Rafael Oliveira)

- Concretamente em jornalismo de dados, não. Tivemos uma cadeira ou outra que incluía algumas noções na área da linguagem da programação e marcação, como o HTML e CSS. O resto da minha ‘formação’ no jornalismo de dados tem sido através de alguns cursos *online* que tenho frequentado sozinho ou com o Rui”.

Por volta das 14h00, Rui Barros chega à redação e dedicou parte da sua tarde a indicarnos algumas melhorias a implementar nos gráficos criados no *Flourish*: um gráfico de linha (com os dados semanais da Deco Proteste) e um gráfico de mapa de árvore (tradução do inglês “treemap graph”) a partir da informação do Eurostat relativamente às toneladas de importação e exportação deste produto em 2021. Este último gráfico é normalmente utilizado para representar dados absolutos de forma hierárquica, semelhante a uma árvore, com ramificações, utilizando formas geométricas (ver Figura 12).

Figura 12 – Exemplo de um gráfico “treemap” construído no *Flourish*



De forma que mais informação fosse incorporada nos gráficos, Rui Barros explica que o *Flourish* permite que estes sejam interativos. A título de exemplo, ao passar o cursor do rato no gráfico de linha, torna-se possível visualizar o preço do arroz carolino em diferentes períodos. Após a implementação e acerto de pormenores, os gráficos encontram-se prontos para ser incorporados com o texto no *backoffice*, através do *embed* [termo de HTML utilizado para incorporar ficheiros multimédia, como vídeo, áudio, gráficos, entre outros].

**Dia 15**, 20 de janeiro de 2023, sexta-feira. O dia foi maioritariamente dedicado à análise dos dois projetos propostos, embora o do Ensino Superior tenha perdido fulgor, pois não conseguimos calcular as distâncias entre cada município através do *RStudio*, conforme solicitado por Rui Barros. Perto do final da tarde, recebemos uma mensagem textual do colega no *Discord* a dar conta de que os gráficos do trabalho do arroz carolino precisam de uma última alteração.

“[17h50] Rui Barros: Vamos ter de fazer umas alterações aos gráficos. Basicamente tens de eliminar o título e a legenda e meter o *iframe* dentro do código que te enviei acima. Acho que tenho de te explicar isso melhor na segunda-feira. Temos de fazer um *iframe* dentro de um *iframe* e meter o gráfico entalado dentro do que te enviei.

[17h52] Rafael Oliveira: Ok, já apaguei o título e a legenda, mas o resto não percebi muito bem o que tenho de fazer...

[17h53] Rui Barros: Deixa, eu explico-te o processo na segunda-feira, é mais fácil. Falei agora com o Pedro Rios e a peça entra na segunda.” (conversa textual no *Discord*, Rui Barros).

Em jeito de despedida, informamos o jornalista de que carregamos nos respetivos repositórios do *GitHub* as análises conduzidas até ao momento para serem revistas pelo próprio.

**Dia 16**, 23 de janeiro de 2023, segunda-feira. Por volta das 10h45 relembramos Rui Barros, através do *Discord*, sobre a necessidade de ajustar os pormenores identificados na passada sexta-feira e que gostaríamos de perceber o processo. Contudo, dada a urgência de publicar a peça e a sua ausência da redação, o jornalista avançou com as correções, tendo solicitado os títulos e subtítulos que estavam definidos anteriormente no gráfico. Após a sua inserção, remeteu-nos duas hiperligações com os gráficos e respetivos títulos e subtítulos para serem incorporados através do *backoffice*.

“[10h50] Rui Barros: (hiperligação 1), (hiperligação 2). Todas as alterações que fizeres no *Flourish* ficam aqui. Mas título, fonte e afins precisamos de mudar à mão. Temos de arranjar uma solução mais eficiente para isto, mas foi o resultado da conversa que tive com a equipa da Infografia” (conversa informal, Rui Barros). Nos primeiros momentos, não percebemos o que era suposto realizar, mas rapidamente foi-nos explicado que teríamos de colocar as duas hiperligações enviadas no artigo que se encontrava já no *backoffice* para ser publicado.

Após a leitura final por parte da editora do *Online*, Mariana Adam, a peça do arroz carolino foi publicada na página Web do *Público*. Às 12h30 desse dia, Rui Barros informou-nos que era a peça mais lida e alertou-nos para possíveis desfechos do trabalho jornalístico, como mostra a seguinte conversa textual no *Discord*:

“[12h58] Rui Barros: Já recebeste comentários a insultar?

[12h59] Rafael Oliveira: Vi a publicação e os comentários no Facebook há pouco... Mas não notei nada de grave.

[12h59] Rui Barros: Terás os comentários no artigo e no teu e-mail. Não sei se o teu e-mail aparece público, mas há sempre sugestões de correção.

[13h01] Rafael Oliveira: Pois, por acaso acho que o meu e-mail não está público. Mas consigo aceder a isso?

[13h02] Rui Barros: Vai para o e-mail dos editores. Se for algo importante, depois irá ter ao teu.

(...)

[16h12]: Rui Barros: Há a correção de um leitor: não há 0,80 cêntimos. Há 0,80 euros que são 80 cêntimos. Mudás isso?

Rafael Oliveira: Sim, feita a correção é só guardar e sair? Não preciso de dizer “editado a tal horas”?

[16h16] Rui Barros: Vou ver com o Ivo. Aqui é uma coisa pequena... mas deixa ver.

[16h30] Rui Barros: Não se mete aí. É uma edição super pequena.” (conversa informal no *Discord*, Rui Barros).

Ainda neste dia, Rui Barros questionou-nos sobre quais haviam sido as dificuldades constatadas na análise do Ensino Superior e, após as termos descrito, disse-nos que no dia seguinte procuraria esclarecer as mesmas. Pouco depois das 18h00, Rui Barros atualiza-nos sobre o acesso ao artigo publicado: nove mil visualizações e estava entre as notícias mais lidas do dia. Neste momento, confessamos-lhe sentir pressão por publicar um artigo com tal alcance.

“(...) [18h14] Rui Barros: Ahaha, como te disse, fica pior com a receção direta de comentários no e-mail, e com provedores e tudo mais.

Rafael Oliveira: Pois, isso é complicado de se lidar...

Rui Barros: É esta vida... vai haver sempre alguém a encontrar uma coisa mal feita e não deves desvalorizar isso. Mas esta profissão é complicada por isso mesmo.

Rafael Oliveira: Entendo, os olhos estão em cima de vocês. Vou ter isso em consideração

Rui Barros: Sim, mas não te deixes afetar por comentários que possam fazer. Fizeste um bom trabalho (...)” (conversa informal através do *Discord*, Rui Barros).

Decorrente desta conversa, foi transmitido que gostaríamos de abordar alguns assuntos para futuros trabalhos de jornalismo de dados. Em resposta, foi sugerido que apresentássemos uma “pequena metodologia” sobre o tema e que os mesmos seriam discutidos no dia seguinte.

**Dia 19**, 26 de janeiro de 2023, quinta-feira. Nos últimos dias tivemos a oportunidade de esclarecer algumas dúvidas com Rui Barros sobre o trabalho do Ensino Superior, bem como alguns trabalhos que propusemos. Aproveitamos assim para conduzir a respetiva recolha, seleção, análise e representação visual dos dados no *RStudio*.

No dia anterior, Rui Barros indicou que um dos assuntos parecia ter indicadores interessantes e que iria procurar conversar com uma colega jornalista sobre esse tema: o número de famílias reconstituídas e monoparentais. Após ter tido essa conversa, sem nela tivéssemos participado, foi-nos dito que os dados recolhidos podiam ser “mais complexos do que aparentavam”, uma vez que a definição de família reconstituída e monoparental poderia ter diferentes leituras. Quanto à análise aos dados do Ensino Superior, o jornalista entendeu que o impasse identificado – calcular as distâncias entre municípios - poderia ser difícil de resolver, tendo, por isso, essa análise ficado suspensa.

No final do dia, Rui Barros sentou-se junto a nós para ter uma conversa sobre como encontrar um tema que pudesse tornar-se um trabalho jornalístico conduzido por dados:

“Em jornalismo de dados podemos partimos de uma ideia geral, da qual temos um ‘palpite’. Pode ser algo do senso comum, de uma conversa de café ou com os teus amigos. A partir dessa noção que normalmente começa com ‘eu acho que isto é assim’ ou ‘há uma ideia generalizada de que isto é assim’, formulamos uma hipótese que poderá dar aso à história jornalística. Um aspeto que torna a história muito interessante é o nível da granularidade dos dados. Ou seja, quanto mais pormenorizados forem os dados, mais interesse o leitor vai ter porque vai relacionar com a estória e a informação vai dizer respeito à sua rua, freguesia ou concelho. Mas isso também levanta questões éticas, como deves entender. Por vezes temos de perceber que, por muito boa que aquela história pudesse ser, não podemos utilizar informação mais sensível por causa da questão da privacidade dos dados e assim. Aliás, o próprio INE não publica alguns desses dados” (Rui Barros, conversa informal).

Decorrente deste diálogo, foram-nos propostas novas tarefas a explorar, nomeadamente os dados sobre a satisfação laboral em Portugal, a atualização do “Prometómetro” e uma análise à principal fonte de rendimento da população portuguesa, cruzando dados demográficos que mostrassem a percentagem de

população jovem, em idade ativa e de reformados. Aproveitamos o momento para expor uma preocupação sentida quanto ao ritmo de produção de peças noticiosas.

“- Destas três ideias, diria que a que está mais encaminhada para se executar é este da satisfação laboral. Essa dos rendimentos e da demografia é um projeto que tenho em mente há algum tempo e era giro mostrar contrastes do país, identificando a freguesia mais jovem do país com e a que tem mais idosos. Acho que era giro e também era uma oportunidade para sair e ir fazer reportagem no terreno. (Rui Barros)

- Sim, acho que esses dois trabalhos são muito interessantes para desenvolver. Vou tratar disso. Uma coisa que queria comentar contigo e que me está a preocupar é a velocidade com que tenho vindo a publicar peças... achas que estou pouco preparado ou devia acelerar mais o ritmo de produção? (Rafael Oliveira)

- Como assim? (Rui Barros)

- Tenho falado com alguns colegas e apercebi-me que publicam mais peças do que eu. Mas não sei se isso é normal por estar numa área diferente da deles. (Rafael Oliveira)

- Mas com quais colegas? (Rui Barros)

- Foi em conversas com uma colega da secção do Azul e outra do P3. (Rafael Oliveira)

- Certo, mas tu próprio já percebeste que são áreas diferentes [risos]. Repara, o trabalho que fazemos em jornalismo de dados exige mais tempo: temos de recolher dados, seleccionar a informação, analisar, fazer gráficos, entre outras coisas. Portanto, não te preocupes com isso, a sério. Acho que estás num ritmo normal para quem está nesta área (Rui Barros). (conversa informal).

**Dia 22**, 31 de janeiro de 2023, terça-feira. Ao longo dos últimos dias, o nosso trabalho foi dedicado à pesquisa, recolha e análise de dados através do Eurostat. Na página desta organização de produção de dados estatísticos da Comunidade Europeia, encontramos uma base de dados que reunia respostas de 29 países europeus sobre os seus níveis de satisfação laboral (altamente satisfeito, medianamente satisfeito, nada satisfeito).

Exportamos o documento *Excel* e, ao calcular as percentagens das respostas de cada país, percebemos que Portugal tinha a menor percentagem de trabalhadores “muito satisfeitos”. Algumas dúvidas surgiram relativamente ao contexto em que este

questionário havia sido realizado, mas foram rapidamente clarificadas através de um e-mail com a referida instituição.

Relativamente à análise que fizemos, esta foi executada através do *Excel*, pois “nem sempre temos de ir para o *RStudio*” (Rui Barros, conversa informal). Procedemos, então, ao cálculo da percentagem de respostas de cada um dos níveis de satisfação face ao número total de inquiridos de cada país e avançamos na análise da mesma forma com outras variáveis relacionadas com esta base de dados, tais como os setores de atividade com as menores percentagens de alta satisfação, o número médio de horas trabalhadas em cada país, o salário médio bruto, o tipo de vínculos laborais ou o ganho médio mensal entre mulheres e homens.

Numa conversa breve, Rui Barros indica que este projeto pode comportar uma narrativa visual que já havia desenvolvido anteriormente para outros trabalhos de jornalismo de dados: o *scrollytelling*. Segundo o mesmo, este formato narrativo é utilizado, geralmente, para tornar o conteúdo na Web mais interativo e envolvente, ao passo em que o jornalista conduz o leitor pela história, apresentando visualmente “os dados mais importantes”, complementando-os com pequenos trechos. No decorrer deste diálogo, o jornalista de dados refere que gostaria que assumíssemos o processo de desenvolvimento Web desta história jornalística, assim como a criação das representações visuais dos dados a ser integradas. “Gostava que fosses tu a tentar desenvolver esta componente do *scroller*. Sei que não estás à vontade com linguagem de programação, nem tens conhecimentos nessa área, mas eu vou-te ajudar” (Rui Barros).

**Dia 24**, 2 de fevereiro de 2023, quinta-feira, 11h15. Nos últimos dias, encetamos contactos e realizamos entrevistas com personalidades de diversas áreas, desde a Economia à Sociologia, visando responder à questão que viria a ser o título deste trabalho: “Portugal tem dos trabalhadores mais insatisfeitos da Europa. Porquê?”.

Hoje o dia será dedicado à transcrição das entrevistas e à redação do texto. Numa consulta rápida à página Web do *Público* encontramos uma nova peça publicada: “Especial Multimédia: Os leitores do Público contam como enfrentam a inflação”. Quase ao mesmo tempo, recebemos uma mensagem no *Discord* de Rui Barros com a

hiperligação desse artigo e escreve de seguida: “Já vou conseguir falar contigo com mais tempo!” (Rui Barros, conversa informal no *Discord*).

Ainda que não estivéssemos completamente a par do trabalho que este jornalista estaria a desenvolver, já havíamos visto no seu ecrã de computador uma pré-visualização idêntica a esta, que integrava um vasto conjunto de testemunhos que teriam respondido ao inquérito<sup>17</sup> lançado em novembro de 2022 pelo jornal *Público*. Na hiperligação consultada, além das várias dezenas de testemunhos, encontramos duas reportagens multimédia, coassinadas por Joana Gonçalves (texto e vídeo) e Rui Barros (desenvolvimento).

Ambas reportagens que saíram hoje (“Miguel tornou-se estafeta para cumprir o sonho do filho” e “Ana ‘implorou’ por um quarto para continuar a estudar em Lisboa”) apresentam um estilo diferente da maioria das peças publicadas no site: o fundo da página é preto, a peça é introduzida com uma conversa telefónica entre a jornalista e a/o entrevistada/o, e a narrativa segue em *scrollytelling* intercalando fotografia, vídeo, texto e a apresentação textual de alguns dados do INE.

Neste momento, compreendemos que Rui Barros havia ficado responsável pelo desenvolvimento Web e design destas duas reportagens. Contudo, parece-nos discutível catalogar estas duas peças como trabalhos de jornalismo de dados. Embora exista a referência a alguns números, nenhum dos trabalhos comporta uma representação visual dos mesmos. Neste sentido, estas peças são facilmente associáveis a uma prática antecessora à do jornalismo de dados: a RAC, que decorre a partir da realização de inquéritos (Meyer, 2002; Léchenet, 2014), transformando testemunhos em dados. Isso mesmo pode ser evidenciado na “página principal”, que apresenta o conjunto total de testemunhos enviados e as reportagens realizadas, em que um gráfico exhibe o total de respostas recebidas, bem como o número mensal de respostas e a proporção de inquiridos que indicaram se a sua situação financeira tinha melhorado, piorado ou mantido face aos últimos seis meses.

---

<sup>17</sup> O formulário do inquérito que serviu para a construção desta peça e reportagens subsequentes pode ser encontrado em: <https://www.publico.pt/interactivo/aumento-custo-vida> [consultado pela última vez a 1 de agosto de 2023]

Figura 13 – Página inicial da peça “Contas à vida: enfrentar a inflação”

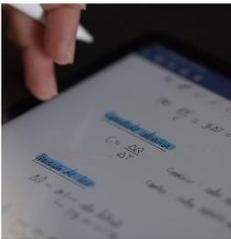
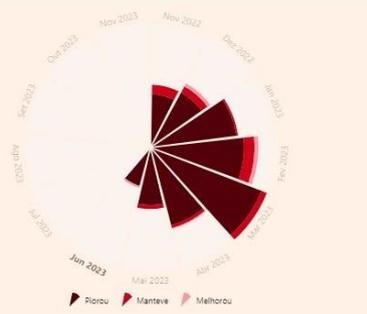
359 leitores já partilharam a sua história

## Contas à vida: enfrentar a inflação

Em Junho das 11 pessoas que deram o seu testemunho ao PÚBLICO, 91% revelavam que, em comparação com os últimos seis meses, a sua situação financeira tinha piorado.

f t in

Contribua com o seu caso e ajude-nos a pintar o gráfico



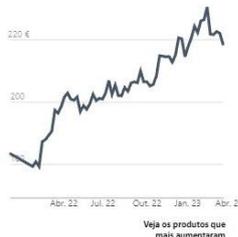
**Reportagem**  
**Francisco paga 695€ numa residência privada. "Um contrato de arrendamento protegia-me"**  
Carolina Pescada e Cristiana Faria Moreira

“  
**Em minha casa quase não entra carne nem peixe porque as reformas não acompanham a inflação**  
Rosária Mateus  
69 anos, Lisboa  
30 de Jun. de 2023



### Evolução do preço do cabaz de produtos essenciais

Cabaz de produtos essenciais elaborado pela Deco Proteste que reúne 63 produtos.



“  
**Fazemos contas à vida para comprar roupa ou calçado para os mais novos**

Cláudia Campos  
45 anos, Lisboa  
13 de Jun. de 2023

“  
**Tenho menos dinheiro disponível para comprar produtos de melhor qualidade e**

Dia 27, 7 de fevereiro de 2023, terça-feira. No decorrer dos últimos dias, Rui Barros indicou a data de 6 de fevereiro para fecharmos o texto da nova peça que estamos a preparar. Todavia, houve contratempos na obtenção de respostas por parte de algumas fontes, tendo esse limite sido prolongado até hoje.

Após informarmos da conclusão do texto, o jornalista de dados explica-nos que definiu essa meta, pois o desenvolvimento Web e a criação dos gráficos para esta história jornalística poderiam tomar uma parte substancial do nosso tempo. No final da tarde, encetamos um diálogo, onde Rui Barros explicou a principal intenção da narrativa visual da história: “prender” o leitor, ou seja captar a sua atenção. “Neste caso, imagino que possa partir de uma constatação mais geral dos dados sobre as percentagens dos países europeus com um gráfico, e depois passarias para o caso específico de Portugal, onde mostras a ideia principal da análise: Portugal é o país com menos trabalhadores satisfeitos face aos outros países” (Rui Barros, conversa informal).

Durante a tarde, apercebemo-nos de que Rui Barros está junto de uma pessoa que ainda não havíamos visto pela redação. No final do dia, o colega indica-nos que, no dia seguinte, gostaria de realizar uma “aula informal” connosco e com José Volta e Pinto. O intuito passa por dar a conhecer os “princípios básicos das três tecnologias mais utilizadas em desenvolvimento Web”: HTML, CSS e Javascript. Para esse efeito, solicita que instalemos o *Visual Studio Code* – um editor de código-fonte gratuito utilizado por desenvolvedores Web.

À saída, aproveitamos para lhe questionar sobre quem era a pessoa que estava junto a ele e é-nos dito que se tratava de um professor da Universidade do Porto, que se dedica à investigação de dinâmicas do território e que o estava a auxiliar na construção de um mapa interativo que iria figurar um artigo a ser publicado brevemente. Face a esta dinâmica, foi possível identificar a presença de um “agente externo” na redação para a construção e colaboração de um trabalho a ser desenvolvido pela equipa de jornalismo de dados (De-Lima-Santos & Mesquita, 2021, p. 1418).

**Dia 28**, 8 de fevereiro de 2023, quarta-feira, 10h45. Iniciamos a construção das representações visuais de dados no *Flourish*. Entretanto, José Volta e Pinto e Rui Barros já se encontram na redação. Rui Barros dirige-se até ao nosso posto de trabalho para nos avisar que nos iremos reunir em instantes na Sala de Reunião 1.

Cada um leva o seu computador portátil para a sala. Rui Barros, que assume neste contexto uma postura mais pedagógica, mas num tom menos formal, começa por perguntar qual é o nosso nível de conhecimento em desenvolvimento Web, nomeadamente com HTML, CSS e Javascript.

“- Com HTML acho que estou à vontade, mas não é à vontadezinha [risos]. CSS ainda me recordo de algumas coisas, mas Javascript não me lembro de quase nada. Entre as três, também acho que é a mais complexa. (José Volta e Pinto)

- Sim. Javascript é mais exigente, requer mais algum conhecimento, concentração e assim, mas há aí umas coisas novas que facilitam o trabalho que já vos vou explicar. E tu, Rafael? (Rui Barros)

- Nunca tive contacto com essas coisas. Já ouvi falar delas, mas nunca mexi com isso. (Rafael Oliveira)

- Ok, sem problema. A ideia desta reunião é dar a conhecer ou relembrar algumas noções básicas de cada uma destas linguagens para que no futuro possamos desenvolver trabalhos autonomamente e diferentes do habitual. (Rui Barros). (conversa informal entre Rui Barros, José Volta e Pinto e Rafael Oliveira).

Ao longo da manhã, foram-nos elencados os fundamentos de cada uma destas vertentes do desenvolvimento Web. A partir das anotações feitas num documento Word, apontamos algumas das “tags” mais utilizadas em HTML, tais como: “head” (tudo o que não aparece visualmente no site, servindo para adicionar metadados), “body” (tudo o que é visível na página), “time” (horário do Google e a hora que o utilizador vê), “divs” (as chamadas “caixas” que serão muito utilizadas), “img” (imagem), “src” (hiperligação para a imagem) ou “alt” (descrição da imagem para portadores de deficiência visual, sendo que a sua utilização dá melhores *ratings* à página no index da Google).

De seguida, fomos introduzidos a alguns fundamentos de CSS, que serve para criar “a parte ‘estética’ do trabalho”, nomeadamente a posição do texto ou do conteúdo multimédia, bem como a sua largura, altura, margens, entre outros aspetos.

Por volta das 13h00, Rui Barros sugere que façamos uma pausa para almoço e que, a seguir, retomáramos esta sessão aprofundando o Javascript. Nesta altura foi-nos endereçado o convite para almoçarmos juntos, mas não nos foi possível estar presentes.

Retornados do almoço, por volta 14h00, voltamos a reunir na mesma sala. Rui Barros explicou que, dentro de uma hora, tinha uma reunião agendada, pelo que iria tentar sintetizar a restante informação. Num primeiro momento, deu-nos a conhecer algumas componentes que podiam ser instaladas no *Visual Studio Code* de forma a facilitar o trabalho. Embora não estivéssemos familiarizados com o assunto, percebemos que algumas destas extensões instaladas poderiam poupar bastante tempo, uma vez que apresentavam, automaticamente, sugestões de linhas de código para escrever ou anotações de passos a seguir. Com o tempo a encurtar, Rui Barros passou à apresentação do Javascript que se trata de uma “linguagem de programação que permite implementar coisas mais complexas na Web” (Rui Barros).

Ao reconhecer a complexidade desta linguagem, Rui Barros sugere façamos o tutorial do *Svelte*, disponível online, até ao ponto “Logic”. A partir dos apontamentos retirados, registamos que o *Svelte* se trata de um “compilador de HTML, CSS e Javascript” que permite “construir interfaces e aplicativos na Web de forma mais fácil e intuitiva”. Neste seguimento, Rui Barros explica-nos que pretende que a construção da peça que estamos a realizar possa ser feita a partir do *Visual Studio Code*, recorrendo ao *Svelte* para criar as componentes do *scrollytelling* da história jornalística.

Terminada a reunião, recorreremos ao papel e à caneta para desenhar um rascunho da narrativa visual que a história poderia seguir, retratando os principais gráficos que gostaríamos de incluir, nomeadamente um gráfico de barras e um mapa da Europa. De seguida, recorreremos ao *Flourish* para criar essas representações visuais e realizamos os primeiros passos do tutorial de *Svelte*.

**Dia 32**, 14 de fevereiro de 2023, terça-feira. Ao longo dos últimos dias, estivemos a “fechar” o texto da peça da satisfação laboral, na qual foi integrada mais uma fonte. Entretanto, foram também feitos avanços significativos na produção das representações visuais de dados, contando já com o mapa e o gráfico de barras inicialmente previstos.

No final do dia partilhamos com Rui Barros os gráficos produzidos. De uma forma geral, o jornalista mostrou-se agradado com o resultado. No entanto, referiu que o mapa geográfico da Europa, que contava com uma gradação da cor de forma a evidenciar os países com maior e menores níveis de satisfação laboral, não o convencia completamente.

“Este é um dos primeiros gráficos que vai surgir na peça, certo? Pelo que percebo, a tua intenção aqui é mostrar que Portugal é o país com menos trabalhadores altamente satisfeitos com o seu trabalho. Há outros gráficos que podiam mostrar isso, mas entendo a opção pelo mapa. No entanto, olhando para isto, não vejo imediatamente que Portugal é o país com menor percentagem. Ou seja, acho que tem de haver um maior contraste nas cores para mostrar que Portugal está em último. Vamos combinar o seguinte: vou-te dar o contacto da Célia, da equipa da Infografia, e vou-lhe explicar a situação. Amanhã dá-lhe um toque a ver se ela te pode ajudar com isto. Desculpa, mas

estou um pouco em caos com o trabalho das rendas que estou a fazer com o Volta [José Volta e Pinto]” (Rui Barros, conversa informal).

**Dia 33**, 15 de fevereiro de 2023, quarta-feira, 10h10. Chegamos à redação, procuramos resolver autonomamente a situação assinalada no dia anterior pelo jornalista de dados, mas os resultados são insatisfatórios. Logo de seguida ligamos com Célia Rodrigues, coordenadora da equipa de Infografia. É-lhe explicado o projeto que temos em curso, assim como o problema identificado com a ilustração do mapa. Uma vez que a conta do *Flourish* é partilhada com a equipa de Infografia do *Público*, a colega rapidamente consegue verificar o mapa criado a partir da redação de Lisboa. Sugere, então, dois caminhos possíveis: procurar na Internet pelo código da cor utilizada no mapa e aplicar “tons mais claros e escuros da mesma família que essa cor” para criar um maior contraste, ou fazer uma divisão de cores por séries. A primeira sugestão - de escurecer e clarear os tons da cor - parece resultar e, de facto, a compreensão visual do mapa melhorou. Entretanto, por volta das 15h00, Rui Barros entra em contacto connosco no *Discord* para saber como correu a conversa com a colega da Infografia. Através de uma captura de ecrã, mostramos o resultado, que mereceu a sua aprovação.

As restantes horas do dia foram dedicadas ao tutorial online do *Svelte* e despendemos cerca de 20 minutos à leitura do livro emprestado por Rui Barros (“*The Wall Street Journal Guide to Information Graphics*”) para retirar alguns apontamentos de forma a otimizar os gráficos que seriam incluídos ao longo da narrativa.

**Dia 34**, 16 de fevereiro de 2023, quinta-feira, 10h35. Os dados enviados pela Deco Proteste desta semana carecem da tabela que faz a comparação com a semana anterior. Após avisarmos Rui Barros sobre essa situação, pede que liguemos com a pessoa responsável que, entretanto, nos faz chegar o documento com os dados completos. A atualização completa da informação está efetuada.

Ontem, depois de sairmos da redação, dedicamos algum tempo à leitura do livro de boas práticas para a representação visual de dados, que nos ajudou a identificar alguns aspetos que necessitavam de ser limados e uma parte substancial da manhã é dedicada a essa tarefa.

Durante a tarde procuramos avançar com o desenvolvimento Web desta peça no *Visual Studio Code*, mas fomos confrontados com várias dúvidas e dificuldades. Por volta das 17h00, deslocamo-nos até ao posto de trabalho de Rui Barros a dar conta dessas adversidades. O mesmo explica-nos que faltava, em primeiro lugar, instalar um *software* para dar início à estrutura da página. Enquanto o computador procede com a instalação demorada das componentes que faltavam, Rui Barros mostra-nos o trabalho que está a desenvolver em conjunto com José Volta e Pinto. Trata-se de um artigo com um mapa interativo que mostra o custo de arrendar uma casa nas freguesias de Portugal com base no rendimento da população. O trabalho ainda não está concluído e são trocadas algumas impressões sobre o mesmo, evidenciando-se alguns problemas, como o traçado geográfico de algumas freguesias e o preço da habitação em cada uma delas.

“– Em que freguesia vives, Rafael? (Rui Barros)

- Vivo em Santa Maria de Lamas, fica no concelho de Santa Maria da Feira.  
(Rafael Oliveira)

- Ok, deixa-me procurar aqui no mapa... Pois, a tua freguesia nem sequer aparece. Mas estes valores das freguesias à volta, achas que batem certo com os preços de renda que se praticam por lá?

- Por acaso não. O preço da renda de um T2 no concelho da Feira é um pouco superior a esse (Rafael Oliveira)

- Pois... Isso é um problema. (Rui Barros)”. (conversa informal).

Concluída a instalação do *software* necessário, Rui Barros indica que irá carregar alguns componentes no *GitHub* para que, amanhã, possamos iniciar a construção da narrativa visual e acrescenta que não estará na redação, mas que se encontrará disponível *online*. Ainda antes de abandonarmos a redação, Rui Barros e José Volta e Pinto dialogam sobre um trabalho em curso:

“- Volta, já conseguiste falar com o professor? Queria ver se ainda esta semana lançávamos a peça. (Rui Barros)

- Vou agora mesmo falar com ele e amanhã já te envio o texto. (José Volta e Pinto)

- Qual é o trabalho que estão a fazer? (Rafael Oliveira)

- É este aqui das rendas que acabei de te mostrar agora.” (Rui Barros) (conversa informal).

Face a este descobrimento, percebemos que esta será mais uma peça elaborada em colaboração, por Rui Barros e José Volta e Pinto, e que contará com um mapa interativo de dados, assim como uma fonte humana.

**Dia 35**, 17 de fevereiro de 2023, sexta-feira. Chegados à redação, tal como habitualmente fazemos, damos os bons-dias aos jornalistas presentes. Rui Barros pede para consultarmos a mensagem que nos deixou no *Discord* durante a noite. A mensagem diz o seguinte:

“Rui Barros (16 de fevereiro, 23:48): Olá Rafael. Acabei de atualizar o *Git* para receber o *scroller* e os *embeds* do *Flourish*. Amanhã atualiza o código do teu lado e olha para a pasta ‘*routes/+page.svelte*’. Acho que te simplifiquei a vida, apenas tens de editar o que está dentro de ‘*scroller\_inicial*’. Honestamente, a parte do *scroller* acho que o gráfico só precisa de título e fonte. Os outros deixei para depois usares no corpo do texto. Mas amanhã vê. Vais, claro, ter de ir ao *Flourish* e tirar de lá os títulos, notas e fontes que puseste, isso passa a ser escrito dentro do código e não no *Flourish*, como fizemos no trabalho do arroz. Para atualizares os gráficos no teu código tens que: 1) ir ao gráfico, editar e depois fazer ‘*export & publish*’ no *Flourish* (canto superior direito) e, 2) entrar dentro da *story* do *Flourish* e voltar a atualizar lá, fazendo ‘*export & publish*’ (canto superior direito). Amanhã falamos melhor que já não consigo pensar grande coisa, abraço” (Rui Barros, conversa informal no *Discord*).

Procuramos seguir as instruções do jornalista, mas, mais uma vez, encontramos dificuldades em compreender os passos a seguir no *Visual Studio Code*. Ainda que estivéssemos no mesmo espaço, Rui Barros envia-nos uma mensagem novamente pelo *Discord*, por volta das 16h00, procurando saber como estava a decorrer o processo.

A partir dessa hora até ao final do dia, diversos pormenores técnicos e dúvidas foram discutidos e esclarecidos com este jornalista de dados. O mesmo mostrou-nos, no ecrã do seu computador, como se constrói a narrativa visual da história em *scroller* e como se procede à incorporação do texto, gráficos e de “notícias recomendadas” que seriam contemplados ao longo da narrativa. Neste momento, apercebemo-nos de que estava a ser construída de raiz a página onde esta peça jornalística seria publicada, assumindo quase o papel de “paginador” Web (desenvolvedor Web), escrevendo código que permitisse incorporar texto, hiperligações e gráficos através de linguagem de marcação

e programação, como HTML, CSS e *Svelte*, no *Visual Studio Code*. Após esta explicação, assumimos as rédeas para colocar o texto, hiperligações, gráficos e notícias recomendadas.

Este trabalho estendeu para lá do horário habitual e, conforme já havíamos mencionado, exigiu um nível considerável de concentração diante do computador. Além disso, Rui Barros alertou para o facto de o texto estar “bastante grande” e que deveríamos “reler com olhos de ver”, editar e seleccionar as partes mais importantes.

“Já te tinha dito isso no teu texto do arroz, mas às vezes és um bocado prolixo. Ao contrário do que nos dizem, a verdadeira economia é a economia da atenção. Temos de ser capazes de agarrar o leitor e manter a sua atenção. Se o texto se desvia ou fica demasiado longo, se tem frases demasiado complexas ou até redundantes, o leitor vai-se embora. Portanto, temos de evitar ao máximo isso. (...) (Rui Barros, conversa informal no *Discord*).

No final do dia, informamos o colega que iríamos terminar de “paginar” o resto da peça em casa e foi-nos dito que, entretanto, fez o *upload* de um *embed* para integrar os gráficos do *Flourish*. Terminado este processo, subtemos a versão para o repositório do projeto no *GitHub* para que o jornalista verificasse se estava tudo em conformidade.

**Dia 36**, 20 de fevereiro de 2023, segunda-feira. Poucos minutos depois de chegarmos à redação, ouvimos um comentário de congratulação a Rui Barros por parte do diretor-adjunto, David Pontes, sobre a mais recente peça publicada na página Web do *Público*: “Onde é que consigo arrendar uma casa com o meu salário?”.

O trabalho é assinado pelos jornalistas Rui Barros, José Volta e Pinto e por Aitor Varea Oro, arquiteto, investigador do grupo Morfologias e Dinâmicas do Território do Centro de Estudos em Arquitetura e Urbanismo da Universidade do Porto e coordenador do programa “Habitar Porto”, que havia estado na redação há alguns uns dias. Este artigo contempla um mapa colorido de Portugal Continental e comporta algumas opções que o utilizador pode seleccionar e preencher, evidenciando a interatividade do mesmo. As opções permitem que o utilizador escolha em que tipo de imóvel quer viver (apartamento, casa em banda, geminada ou isolada), a área do imóvel (equivalente a um T0, T1, T2, T3 ou T4) e um espaço onde poderia escrever o seu salário líquido em euros.

É possível observar que, depois desse mapa, existe uma nota metodológica sobre como os valores de referência dos imóveis foram calculados a partir dos dados que o INE disponibiliza para cada freguesia ou município. Acresce a isso ainda um texto que explica que o desfasamento dos valores de arrendamento no mapa do *Público* se deve às “diferenças na forma como os valores são calculados”. Em suma, o texto refere que as páginas de compra e arrendamento de imóveis mostram o valor que o proprietário deseja e não o que fica firmado nos contratos de arrendamento, que são “tendencialmente mais baixos”. Nesse sentido, este mapa reflete “os preços dos contratos de arrendamento firmados, usados para definir rendas acessíveis pelo Programa de Apoio ao Arrendamento (PAA)”. Como complemento, o investigador Nuno Travasso, da Faculdade de Arquitetura da Universidade do Porto, faz também uma explicação sobre o mecanismo que estabelece os valores de referência de arrendamento para o INE fazer a análise por freguesias.

Durante este dia, executamos algumas melhorias no texto, procurando encurtá-lo, e nos gráficos, tendo remetido a parte textual para a revisão do editor Ivo Neto.

**Dia 38**, 22 de fevereiro de 2023, quarta-feira. Ainda com o desenvolvimento Web da peça sobre a satisfação laboral em mãos, procuramos corresponder ao repto lançado por Rui Barros para tentar implementar uma componente que incorporasse notícias na peça, mas não fomos bem-sucedidos.

Por volta das 10h30, recebemos uma notificação no *Discord*. Trata-se de uma mensagem de Rui Barros que nos pede para alterar o título de um gráfico que, a seu ver, não correspondia com os dados utilizados. Poucos minutos depois, o mesmo colega envia-nos uma ilustração (Figura 14) que iria servir como imagem de destaque da peça na página Web. Expressamos o nosso agrado e recebemos uma resposta do mesmo dizendo: “Estou a ficar um diretor de arte talentoso”. Neste momento, percebemos que a ilustração enviada havia sido criada pelo próprio Rui Barros.

Figura 14 – Ilustração criada por Rui Barros para a peça da satisfação laboral



Já por volta das 15h30, somos interpelados pelo editor do *Online*, Ivo Neto, que nos alertou para a necessidade de adequar o uso das citações e de encurtar o texto. Esse reparo foi transmitido a Rui Barros, assim como a tentativa falhada para criar uma componente em *Svelte* que incorporasse as notícias “recomendadas”. Em resposta, o colega pede que carreguemos o texto no *Visual Studio Code* e que, quando concluíssemos os reparos do editor, enviássemos esse ficheiro atualizado e a componente que tentamos desenvolver para incorporar notícias recomendadas.

**Dia 39**, 23 de fevereiro de 2023, quinta-feira, 10h30. Depois de termos despendido cerca de 30 minutos com a inserção dos dados semanais da Deco Proteste, abrimos o *Visual Studio Code* para atualizar o texto da peça da satisfação laboral, já revisto e editado por Ivo Neto.

Entretanto, encetamos mais uma tentativa para desenvolver a componente que permitisse incorporar notícias e o resultado parece promissor. Contudo, parece-nos que o código não está correto, pois Rui Barros, durante a aula informal, tinha dito que a intenção da criação destes componentes em *Svelte* passava pela sua reutilização e importação para outros projetos, algo que não parecia acontecer com o que executamos. Nesta altura, decidimos partilhar o resultado possível, carregando-o no *GitHub*, mas não obtivemos uma resposta até ao final do dia por parte do colega.

Face ao exposto, decidimos avançar com a consulta a bases de dados do INE sobre a principal fonte de rendimento dos portugueses para tentar dar início a um novo projeto, visto que este estava prestes a ser concluído.

**Dia 41**, 27 de fevereiro de 2023, segunda-feira, 10h00. Chegamos à redação, Rui Barros encontra-se a trabalhar no seu computador pessoal. Perguntamos se nos podemos sentar ao lado da sua secretária. O colega procura confirmar se uma jornalista da secção do *Online*, que costuma ocupar aquele lugar, viria hoje à redação.

“- Olha, Jornalista A, sabes se a Jornalista B vem hoje? (Rui Barros)

- Deixa-me ver aqui na agenda... Não, ela hoje não vem. Está a trabalhar a partir de casa. (Jornalista A)

- Ok, obrigado. Sim, Rafael, senta-te aí. Olha e desculpa não te ter respondido no outro dia, estava numa reunião, vi a tua mensagem, mas depois esqueci-me completamente de responder. (Rui Barros)”. (conversa informal).

Este jornalista conta-nos que teve a oportunidade de consultar a componente em *Svelte* que procuramos criar e, embora reconheça a tentativa encetada, explica que a forma como a executamos não estava correta. Nesse seguimento, pede-nos que atualizemos o projeto no nosso computador, uma vez que corrigiu essa componente e estava já pronta a utilizar. Após termos incorporado as notícias recomendadas ao longo do texto, aproveitamos para rever o estado final da peça. De seguida, acedemos à plataforma online do INE para dar continuidade à consulta de bases de dados sobre um dos trabalhos sugeridos há tempos por Rui Barros – o dos rendimentos das famílias portuguesas. Todavia, somos interrompidos pelo colega.

“- Rafael, estás a fazer algo muito importante? (Rui Barros)

- Mais ou menos... estava a consultar no INE uns dados sobre aquele assunto da principal fonte de rendimentos das famílias para tentar avançar com isso. (Rafael Oliveira)

- Ok, boa, mas queria que largares isso por enquanto e me desses aqui uma ajuda. Vai ser um trabalho chato, mas precisava que me ajudasses a limpar aqui uma base de dados, porque tenho de fazer isto tudo à mão e sozinho vai demorar uma eternidade. Se formos dois, é mais rápido, pode ser? (Rui Barros).

Aceitamos o repto e, após nos ter sido enviada a hiperligação por e-mail para aceder ao documento no *Google Sheets* de forma a trabalharmos de forma colaborativa e simultânea no mesmo ficheiro, perguntamos de que forma o podíamos ajudar e o que era necessário fazer. “Basicamente, a jornalista C arranhou esta base de dados em formato PDF e pediu-me para analisar os dados. O problema é que ao converter o PDF para *Excel*, ou *Google Sheets* neste caso, a tabela ficou desformatada. Há linhas que ficaram sobrepostas, colunas que ficaram vazias... está uma confusão tremenda. Sem ter isto “limpo”, não dá para analisar os dados. Eu já adiantei uma boa parte do início, mas vê primeiro como estou a fazer. Vou continuar a partir da linha onde estou e tu podes começar da última linha da tabela para cima. Encontramo-nos a meio do documento, se tudo correr bem, está bem? [risos]. (Rui Barros, conversa informal).

O processo de “limpeza” desta base de dados demorou, aproximadamente, três horas. Revelou-se uma tarefa extenuante, uma vez que o nosso olhar e atenção estavam focados numa folha de cálculo e precisávamos de assegurar que nenhuma informação era eliminada despropositadamente. Durante este tempo, foram trocados alguns comentários com o jornalista Rui Barros sobre em qual linha nos encontrávamos e se não havia uma forma mais fácil de executar este trabalho.

“- Há uma forma mais fácil, sim. É chamamos uma terceira pessoa [risos]. Olha deixa ver se temos sorte... Oh Volta, estás muito ocupado? (Rui Barros).

- Por acaso estou aqui a tratar de uma coisa que me pediram, mas o que precisavas? (José Volta e Pinto).

- Se tens coisas para fazer, deixa estar. Ia convidar-te a juntares a mim e ao Rafael a fazer aquilo que menos gostas em jornalismo de dados [risos] (Rui Barros).

- O quê? Limpar alguma base de dados? É que nem penses [risos]. Da última vez que fizemos isso estivemos mais de três horas à volta daquilo, Rui”. (José Volta e Pinto).

À medida que o “trabalho de limpeza” ia sendo realizado, estes jornalistas explicavam que a base de dados que estávamos a tratar era substancialmente mais pequena do que a última com que tiveram de lidar há uns meses para outro projeto, ainda que se tratasse de uma tarefa maçadora. Já depois das 19h00, concluímos a limpeza conjunta da base de dados e procedemos a uma breve revisão para averiguar se alguma linha ou

coluna tinha sido apagada. Rui Barros agradece a ajuda e sugere que entremos um pouco mais tarde ao serviço no dia seguinte e também nos informa que o trabalho da satisfação laboral ia ser revisto por “desks” (um grupo de profissionais encarregue por rever, com maior minúcia, os textos produzidos pelos jornalistas) e que seria publicado esta quinta-feira.

**Dia 42**, 28 de fevereiro de 2023, terça-feira. Após nos ter sido remetida a revisão efetuada por aquele grupo de trabalho de revisão textual, procedemos à atualização final do texto. Seguimos com a recolha dos dados sobre a principal fonte de rendimento da população residente em Portugal com 15 e mais anos de idade. Contudo, o INE não nos permite descarregar o ficheiro com os dados das 3092 freguesias. Falamos com Rui Barros sobre este constrangimento e explica-nos que será necessário utilizar um *script* para descarregar os dados pretendidos, utilizando o catálogo de indicadores do INE. O jornalista explica-nos que “quando se trata de uma base de dados muito grande”, esta instituição estatística disponibiliza na sua página uma API que permite criar e/ou exportar a grande quantidade de informação que pretendemos, bastando indicar os indicadores a analisar<sup>18</sup>.

Rui Barros explica que possui esse *script*, que criou a partir da linguagem R, e que nos iria disponibilizá-lo, mas que, primeiro, teria de explicar como o mesmo funciona. O jornalista sugere a próxima sexta-feira para realizar essa explanação connosco.

**Dia 45**, 3 de março de 2023, sexta-feira, 15h00. Ontem, escrevemos uma breve notícia a dar conta de que o reportório do jornalista de dados do *Público* – Rui Barros – foi nomeado para os Sigma Awards 2023.

Hoje, tal como combinado, deslocamo-nos até ao seu posto de trabalho para que nos explicasse como funciona este *script* para descarregar os dados que pretendíamos obter. “Então, basicamente este *script* é corrido no *RStudio*. O que ele vai fazer é um *scrapping* à página do INE com os dados que queremos obter. Ou seja, tu não conseguiste descarregar aqueles dados diretamente do INE porque a informação é demasiado grande para o servidor do INE processar tudo de uma única vez e para

---

<sup>18</sup> O serviço de API do INE está disponível em: [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_api&INST=322751522](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_api&INST=322751522) [consultado pela última vez a 3 de agosto de 2023].

exportar e descarregar num único ficheiro Excel. Aliás, o ficheiro Excel provavelmente nem aguenta com a quantidade dos dados. Então, já há algum tempo desenvolvi este *script* que permite buscar essa quantidade imensa de informação, mas há aqui questões éticas importantes. Depois de colocares os códigos dos indicadores do INE que queres descarregar, como o do indicador geral, o da dimensão temporal e o da dimensão geográfica, tem aqui uma parte do código do *script* que diz “sleep”. O que isto faz é, no intervalo entre 1 e 10 segundos, dar uma pausa a sacar os dados. E tu perguntas-me: mas porquê dar essa pausa, se vai demorar mais tempo? Pois, imagina que corríamos o *script* sem dar pausas: o servidor do INE ia receber imensos pedidos de nossa parte para obter a informação. Isso podia prejudicar o funcionamento normal da página e não é ético prejudicar qualquer página ou o acesso à mesma, por parte de outras pessoas, só porque queremos muito aqueles dados. Além disso, podíamos mesmo ser bloqueados de aceder à página do INE, porque a cibersegurança do site podia interpretar isso como um ataque. Pronto, depois no fim, tem aqui este trecho de código que te vai juntar todos os dados descarregas numa única base de dados no *RStudio*. Mas olha que isto é capaz de te demorar algumas horas...”. (Rui Barros, conversa informal).

Depois de termos seguido estes passos, e termos sido sensibilizados para questões que nunca nos havia ocorrido, vimos que a obtenção destes dados iria, de facto, demorar algumas horas a ser concluída, e o dia de trabalho já se aproximava do fim.

O jornalista explicou-nos ainda que este *script* permite “verificar a percentagem de informação que já foi descarregada”, num dos quadrantes do *RStudio*, e pausar o seu processamento, bastando clicar em “stop”. Para retomar a transferência dos dados bastava verificar na pasta, onde a informação estava a ser armazenada, o número/código da última freguesia descarregada (este número/código é único para cada uma das freguesias, também conhecido como “dicofre”) e retomar o *script*.

Entretanto, damos conta que a peça da satisfação laboral ainda não foi publicada e Rui Barros explica-nos que a mesma estava a ser avaliada pelo editor da secção de Economia, embora não soubesse de mais pormenores.

**Dia 46**, 5 de março de 2023, domingo, 18h15. Apesar de hoje ser domingo e não estar previsto realizar alguma tarefa, recebemos uma chamada telefónica de Rui Barros. Durante a conversa, o jornalista explica que tem dúvidas sobre um dos gráficos produzidos sobre a satisfação laboral, que apresentava os três setores de atividades em Portugal com a maior percentagem de insatisfação registada.

Depois de termos explicado como havíamos feito os cálculos no *Excel*, Rui Barros considera que aquela não era a melhor forma de representar visualmente os dados, visto que não se estabelecia comparação com outros países da amostra. Atendendo a isso, pediu-nos que retirássemos esse gráfico, a partir do *Visual Studio Code*, em casa. Solicitou-nos também que ligássemos com David Pontes, diretor-adjunto, para que se retificasse um pormenor identificado no texto, visto que a notícia iria ser publicada na versão impressa de amanhã, segunda-feira.

**Dia 47**, 6 de março de 2023, segunda-feira, 10h00. A peça sobre a satisfação laboral em Portugal está publicada na página Web<sup>19</sup> e fez chamada de capa na edição impressa do jornal (ver *Figura 15*). A peça na página Web encontra-se assinada por nós e Rui Barros, que ficou encarregue, sobretudo pela parte do desenvolvimento Web da peça, embora não se verifique essa menção. Ainda que tenha sido um processo moroso até à publicação, a narrativa visual interativa construída para a Web ficou do nosso agrado e vários profissionais presentes congratularam-nos pela peça, nomeadamente Rui Barros, Ivo Neto e David Pontes. Por parte da equipa de Infografia, a coordenadora Célia Rodrigues fez alguns reparos por e-mail, nomeadamente sobre as cores utilizadas nos primeiros gráficos da narrativa visual.

---

<sup>19</sup> A peça produzida está disponível em: <https://www.publico.pt/interactivos/portugueses-mais-insatisfeitos-trabalho-europa/> [consultada pela última vez a 3 de agosto de 2023]

Figura 15 – Capa da edição de 6 de março do jornal *Público*



Legenda: No quadrado vermelho, encontra-se a chamada de capa sobre a peça realizada.

Como seria expectável, a notícia na versão impressa do jornal contempla somente uma das dez representações visuais produzidas para o *online* e o texto foi reduzido.

Rui Barros partilhou a hiperligação desta peça de jornalismo de dados nas suas redes sociais, tendo inclusive pedido o nosso nome de utilizador no *Twitter* de forma a nos identificar como autor da notícia e acrescentando a *hashtag* "#DDJ", que representa "data-driven journalism" ou "jornalismo conduzido por dados". Depois de uma consulta à sua conta nesta rede social e no *LinkedIn*, compreendemos que Rui Barros tem uma atividade regular nestas plataformas, sobretudo na partilha de trabalhos seus, de

colegas jornalistas do *Público* ou de outros projetos internacionais relacionados com jornalismo de dados.

A meio da tarde, Rui Barros partilhou connosco alguns outros trabalhos de referência deste domínio, nomeadamente do *New York Times* e do *The Washington Post*.

Ainda neste dia, por volta das 17h00, o *Público* celebrou o seu 33.º aniversário. Os jornalistas e demais colaboradores presentes na redação do Porto reuniram-se para cantar os parabéns e um brindar a esta efeméride, contando com a presença do diretor do jornal, Manuel Carvalho.

Durante este dia, executamos o *script* no *RStudio* para recolher os restantes dados do INE, mas ainda não conseguimos descarregar toda a informação, tendo ficado a 62% do total. No final da tarde, Rui Barros diz-nos que a nossa peça foi a décima mais lida do dia, contando com cerca de 4100 acessos e informa que amanhã estará a trabalhar a partir de casa.

**Dia 48**, 7 de março de 2023, terça-feira, 12h00. O descarregamento dos dados a partir do INE está concluído. A base de dados agora reunida contém imensas linhas de informação e sentimo-nos assoberbados pela quantidade dos dados existentes.

Procuramos, então, simplificar a base de dados original, a partir do *RStudio*, calculando a população total de cada freguesia com 15 ou mais anos de idade, criando-se uma segunda base de dados para esse efeito. O *RStudio* permite criar várias bases de dados a partir do código que escrevemos, bastando redigir, por exemplo, as seguintes linhas de código para obter o total da população residente de cada freguesia:

Figura 16 – Trecho de código escrito em linguagem R

```
df_totais <- df %>%  
  filter(sexo=="HM") %>%  
  group_by(dicofre) %>%  
  mutate(total_pop_freguesia=sum(valor)) %>%  
  select (-sexo)
```

Legenda: Exemplo de linhas de código escritas no *RStudio* para criar uma nova base de dados

Neste caso, o “df\_totais” será a nova base de dados que vamos criar a partir do “df”, que se trata da base de dados original. Para obtermos os totais de população por cada freguesia, filtramos a informação pelos sexos masculino e feminino (“filter(sexo==”HM”)), agrupamos pelo código único de cada uma das freguesias (“group\_by(dicofre)”) e, através da função “mutate”, criamos uma variável (“total\_pop\_freguesia”) que será soma dos valores populacionais de cada território (“sum(valor)”). No fim (“select (-sexo)”), removemos uma variável que já não nos interessará, visto que os valores somados entre os sexos estarão reunidos na nova variável “total\_pop\_freguesia”.

As restantes horas do dia foram dedicadas a compreender as análises que poderíamos realizar a partir da base de dados, estabelecendo algumas perguntas de partida, tais como: “quais são as freguesias com mais jovens e mais idosos em Portugal?”, “quais as freguesias que têm mais pessoas a trabalhar?” ou ainda “qual é a freguesia que tem mais pessoas a receber reformas ou pensões?”. Importa notar que estas questões foram possíveis de formular dada a informação que esta base de dados continha, contemplando variáveis como a idade, principal fonte de rendimento, sexo e a localização.

**Dia 49**, 8 de março de 2023, quarta-feira, 14h30. Durante a manhã procuramos responder a algumas das perguntas formuladas no dia anterior.

Quando surgiam dúvidas sobre que função utilizar no código para obter o pretendido, recorremos ao Google ou à própria secção de “Ajuda” do *RStudio* que nos permitia compreender as possibilidades de cada função. Contudo, outras dúvidas técnicas surgiram, às quais não conseguimos obter respostas na Internet devido à sua especificidade, sobretudo no que concerne à criação de gráficos, o que nos levou a interpelar o jornalista de dados, Rui Barros.

Quando lhe transmitimos o ponto de situação através do *Discord* (devido à sua ausência da redação), Rui Barros pede que carreguemos a análise num novo repositório do *GitHub* e diz que tem uma outra tarefa que gostaria que realizássemos,

pois encontrava-se a trabalhar em dois projetos, sendo que um deles teria de ser publicado já amanhã.

Através de uma chamada de voz nesta plataforma, o colega refere que a jornalista C, da redação de Lisboa, descarregou uma base de dados sobre os aumentos no preçário dos serviços da ADSE e que precisava que “alguém olhasse para isto” e retirasse as “principais conclusões”. “Eu já dei uma vista de olhos no ficheiro, mas é novamente o problema da outra vez. Não está num formato adequado para analisar no R, portanto isso vai obrigar a fazer uma nova base de dados que agregue a informação toda. Além disso, vais ver que a informação está separada por páginas e isso não dá para analisar assim... Tenta resolver isso, analisa os dados e segue aquele guia de boas práticas que te referi para montar uma análise: indicar os passos que foram dados, os objetivos da análise, a análise realizada, os gráficos e depois as principais conclusões” (Rui Barros, conversa informal).

Depois desta explicação e de acedermos ao documento, apercebemo-nos que a tarefa seria idêntica àquela que realizamos em colaboração com Rui Barros há umas semanas: “limpar” manualmente os dados no *Google Sheets* e reuni-los numa base de dados legível para ser analisada no *RStudio*.

**Dia 50**, 9 de março de 2023, quinta-feira. No seguimento do trabalho “Contas à vida: enfrentar a inflação”, encontramos um novo trabalho publicado na página Web do *Público*, assinado por Joana Gonçalves (texto e vídeo), Pedro Rios (entrevista telefónica) e Rui Barros (desenvolvimento), intitulado: “Nuno tem uma vinha: ‘Os preços aumentam, mas não recebemos mais’ ”.

À semelhança das duas reportagens anteriormente publicadas, esta segue o mesmo design, incorporando elementos multimédia - áudio, imagens, vídeo e o destaque de alguns dados estatísticos baseados na informação semanal da Deco Proteste-, mas sem qualquer tipo de representação visual da informação numérica.

O dia de trabalho é dedicado à continuação da “limpeza” e montagem de uma nova base de dados com os preços da ADSE no *Google Sheets*. Tal como já havia sido evidenciado numa tarefa anterior, esta tarefa revelou-se exaustiva e consumiu uma parte substancial do tempo, dado o rigor necessário neste procedimento. Além disso,

foram encontradas algumas células que não continham o preço afeto ao procedimento médico.

Decidimos excluir esses casos e anotando-os no documento onde seria realizada a análise. Ultrapassadas essas questões técnicas, conseguimos montar a base de dados e procedemos às primeiras análises no *RStudio*. O principal objetivo desta análise já havia sido evidenciado pela Jornalista C: descobrir quais áreas e/ou procedimentos médicos ou tratamentos que sofreram um maior aumento relativamente ao mês anterior.

**Dia 52**, 13 de março de 2023, segunda-feira, 14h15. A análise dos dados solicitada está concluída. Neste momento, evidenciamos uma maior destreza em operar com este *software* e com a linguagem R para analisar dados, realizar cálculos estatísticos simples e representar visualmente os dados.

A principal operação realizada nesta análise passou por calcular a variação percentual e absoluta dos preços dos vários procedimentos, uma vez que o propósito era o de descobrir as maiores variações nos preços. Para calcular a variação absoluta do preço, bastava subtrair o preço de março de 2023 com o que estava em vigor em fevereiro de 2023, obtendo a diferença absoluta entre os dois valores. Já para a variação percentual, realizamos a seguinte operação: “preços de fevereiro – preços de março/preços de fevereiro x 100”.

Após termos dado conta da conclusão deste trabalho a Rui Barros, o colega solicita que criemos um documento HTML, a partir do *RStudio*, para que a análise possa ser consultada pela jornalista C, uma vez que a colega não estaria familiarizada com a linguagem R. Há, por isso, a necessidade de incluir a análise realizada, explicitando os vários passos realizados, desde a seleção até às principais conclusões. No fundo, este ficheiro servirá como um “relatório da análise” executada.

Neste seguimento, Rui Barros sugere que consultemos na Internet como se cria um documento HTML a partir do *RStudio*, a fim de ser partilhado. Pede-nos também que este relatório inclua um índice com as fases mais importantes da análise, nomeadamente os objetivos, a metodologia adotada, os cálculos realizados, os gráficos e as principais conclusões. Acrescenta que o *RStudio* utiliza uma linguagem simples de

marcação – *RMarkdown* – que, através da função *knitr*, irá converter o nosso texto nesse documento HTML. Percebemos assim que o *RMarkdown* serve, essencialmente, para transformar as análises de dados em documentos, relatórios e apresentações.

Ao consultarmos a Internet e os fóruns dedicados à comunidade que utiliza a linguagem R, conseguimos dar início à tarefa solicitada. Já no final do dia, Rui Barros pede para irmos até ao seu posto de trabalho e adverte-nos para um erro cometido na análise do trabalho dos rendimentos da população que, entretanto, havíamos partilhado com ele no *GitHub*.

“Sobre o trabalho dos rendimentos, reparei numa coisa: não faz sentido estares a calcular o número de pessoas de cada freguesia que recebem, por exemplo, o Rendimento Social de Inserção pelo total de pessoas que recebem isso a nível nacional. A intenção do trabalho é perceber que percentagem de pessoas de cada freguesia recebe esse subsídio face à sua população. Ou seja, tens de calcular com base no total da população da freguesia e não pelo total de pessoas que recebem o Rendimento Social de Inserção em todo o país.” (Rui Barros, conversa informal).

**Dia 54**, 15 de março de 2023, quarta-feira, 12h00. Durante as primeiras horas da manhã voltamos a consultamos um fórum da comunidade de linguagem R para apurar como melhorar o aspeto do nosso relatório de análise e o converter num documento HTML. O processo de criação deste trabalho permite fazer pré-visualizações do relatório final de forma a verificar se as linhas de código escritas correspondem a uma apresentação aprazível. No fim, para transformar o ficheiro *RMarkdown* em HTML, carregamos uma biblioteca do *RStudio* que ainda não havíamos utilizado: o “*knitr*”. Este pacote integra a computação para a elaboração de relatórios, incorporando o código, texto, análise, resultados e a discussão num único ficheiro.

O resultado é partilhado com Rui Barros, novamente através do *GitHub*, e o mesmo mostra-se agradado com o documento e análise realizados. Não obstante, deixou alguns apontamentos no código escrito por nós, pois algumas linhas de código podiam ser simplificadas ou, por outras palavras, podíamos ter “escrito menos”. De seguida, solicita que contactemos a jornalista C para dar a conhecer as principais conclusões retiradas. Após termos solicitado o seu contacto, a colega atende-nos, mas refere que

está a caminho de uma conferência de imprensa e pede que enviemos o documento de análise para o seu e-mail.

Depois de almoço, por volta das 14h00, regressamos à redação e o calor faz-se sentir. Decidimos dar seguimento à análise da principal fonte de rendimentos da população portuguesa. No dia anterior a este, Rui Barros sugeriu focar a análise os dados relativos ao Rendimento Social de Inserção, uma vez que era um assunto que, ocasionalmente, era discutido por alguns atores da política nacional. Após termos calculado devidamente os valores percentuais por cada freguesia, conforme o colega havia reparado, consideramos que visualizar os dados no mapa de Portugal seria importante para compreender onde se localizavam as freguesias do país com a maior percentagem de pessoas que assinalaram o Rendimento Social de Inserção (RSI) como principal fonte de rendimento. Contudo, encontrar um ficheiro *shapefile* que contivesse todas as freguesias do país não foi possível.

Surgiu-nos, entretanto, uma dúvida: como é que estes dados foram recolhidos? Após uma pesquisa à fonte original (INE), percebemos que os dados tinham sido obtidos a partir dos Censos 2021 e que, de facto, existia uma pergunta sobre qual havia sido a principal fonte de rendimento de cada pessoa, com 15 ou mais anos. Decidimos, por precaução, incluir essa contextualização na nota metodológica da análise no *RStudio*.

**Dia 56**, 17 de março de 2023, sexta-feira. No final do dia anterior decidimos incluir uma nova variável à análise: a do fundo de desemprego. A intenção passava por perceber se poderia existir uma correlação entre estes dois tipos de rendimentos nas freguesias, o que poderia evidenciar as regiões onde o contexto económico-social era desfavorável.

A meio da manhã, numa consulta à página Web do *Público*, encontramos um novo trabalho assinado pelos jornalistas Rui Barros e Luísa Pinto: “Quantas casas vagas há na minha freguesia? Veja o mapa”<sup>20</sup>. Nesta altura, não tínhamos qualquer indicação que Rui Barros estaria a desenvolver um novo projeto, embora, várias vezes, este jornalista se encontre ocupado e sem disponibilidade para esclarecer dúvidas ou rever as análises que realizamos.

---

<sup>20</sup> A peça noticiosa está disponível: <https://www.publico.pt/interactivos/mapa-vagos-freguesias/> [consultado pela última vez a 5 de agosto de 2023].

Quanto ao novo artigo, trata-se de uma peça com um mapa interativo de Portugal, onde é possível consultar quantos alojamentos estavam vagos, em 2021, nas 3092 freguesias de Portugal a partir dos dados do INE. Contrariamente ao primeiro mapa interativo, que se debruçou também sobre a temática da habitação, os limites geográficos de cada uma das regiões são mais realistas. Por outro lado, à semelhança desse mapa anterior, se o ampliarmos e clicarmos numa das freguesias, surge uma pequena mensagem personalizada com alguns destaques a cor e um gráfico de barra única, indicando o seguinte texto: “Nesta freguesia, [número da percentagem] % dos alojamentos familiares estavam, em 2021, vagos, sendo que X % estavam à venda ou para arrendar [identificado com a cor laranja] e Y % estavam vazios (cor rosa). Os restantes Z % eram residências habituais ou secundárias [identificado a azul-claro]”. A opção pelas cores neste pequeno trecho coincide com as cores da barra (ver *Figura 17*), permitindo ao utilizador facilmente identificar as percentagens correspondentes à mensagem. Relativamente ao mapa, conta com uma pequena legenda da paleta de cores utilizada para identificar as freguesias de todo o país com menos e mais casas vagas.

Figura 17 – Exemplo da mensagem personalizada na peça do número de casas vagas

FIGUEIRA DE CASTELO RODRIGO  
Vermiosa

80

Alojamentos vagos que não estavam nem para arrendar ou vender em 2021

Nesta freguesia, 27% dos alojamentos familiares estavam, em 2021, vagos, sendo que 4% estavam à venda ou para arrendar e 23% estavam vazios. Os restantes 73% eram residências habituais ou secundárias.



A peça conta ainda com um texto vincadamente influenciado pela análise dos dados realizados, sendo perceptível que existiu um trabalho de investigação a fontes documentais sobre a temática das “casas vagas”. Após a leitura da notícia na íntegra, apercebemo-nos que o mapa utilizado poderá servir, justamente, para a análise que estamos a conduzir.

**Dia 57**, 20 de março de 2023, segunda-feira, 14h20. A análise dos preços da ADSE ainda não foi discutida com a jornalista C. Apesar disso, Rui Barros indica-nos que o documento já foi consultado e que estaria previsto realizar-se hoje uma reunião por

videoconferência. Contudo, a colega não se encontra disponível e Rui Barros sugere enviar-lhe novamente um e-mail sobre o assunto. Até ao final deste estudo, não recebemos qualquer resposta por parte da colega.

Aproveitamos o momento para rever a análise realizada até ao momento sobre os rendimentos e solicitamos a Rui Barros o mapa das freguesias que tinha utilizado no seu mais recente projeto. Sem hesitações, o colega disponibiliza-nos o ficheiro *shapefile*. Após conduzir a análise sobre os dois rendimentos que decidimos focar (RSI e fundo de desemprego), Rui Barros sugere que a inclusão de outros dados – os da população desempregada –, uma vez que “quem recebe o fundo de desemprego até pode auferir um bom salário”.

Para o descarregamento dessa informação, recorremos novamente ao *script* utilizado. O resto do dia é dedicado à formulação de cálculos e gráficos no *RStudio* com os dados que dispomos para tentar evidenciar a existência de uma correlação positiva entre os dois rendimentos. Conseguimos perceber que, de facto, existia uma correlação positiva. Mas, atendendo à chamada de atenção do colega, compreendemos que seria mais realista verificar a mesma com os dados da população desempregada.

A partir deste dia, a análise passou a ser atualizada diariamente no repositório do *GitHub*. Quanto ao descarregamento da nova base de dados, ficou a 70% da sua conclusão.

**Dia 59**, 22 de março, quarta-feira. Ontem conseguimos obter a totalidade dos novos dados e hoje iremos proceder à sua análise e redigir as principais conclusões. Faltava-nos ainda visualizar os dados no mapa de Portugal, mas após uma breve pesquisa na Internet, conseguimos produzir essa representação visual, o que permitiu acrescentar novas perceções à conclusão da análise.

Já perto do final da tarde, demos a conhecer as principais conclusões a Rui Barros. A conversa originou várias sugestões por parte do colega, nomeadamente criar um mapa colorido que cruzasse as variáveis da população desempregada e do RSI.

Recorrendo à ferramenta que habitualmente usamos, a pesquisa na Internet, o jornalista encontrou um dos sites dedicados à linguagem R<sup>21</sup> que mostrava a representação visual de dados que cruza duas variáveis, através da gradação das cores de cada uma, o que iria permitir identificar as zonas onde o contexto social e económico seria bastante desfavorável. Foi também sugerido criar “grupos etários”, de forma a ser mais facilmente perceptível no mapa onde se encontrava a população jovem, adulta e idosa. Uma das conclusões da análise parece ter despertado a atenção de Rui Barros: nas Áreas Metropolitanas do Porto e de Lisboa há uma maior representação de mulheres do que de homens que referem ter o RSI como principal fonte de rendimento. Por outro lado, se traçássemos um eixo vertical a dividir Portugal, parece haver mais freguesias no Interior em que o sexo masculino predomina. Neste seguimento, sugeriu falarmos com algum investigador da área, no sentido de tentar compreender o porquê deste fenómeno. Entretanto, outras colegas jornalistas juntaram-se à conversa e trocaram-se algumas impressões sobre a análise. A Jornalista D passou-nos o contacto de um “investigador especialista” sobre a questão da pobreza em algumas regiões do país.

Atendendo à aproximação do final deste estudo, Rui Barros reconhece que será “bastante complicado” publicar o trabalho. Nesse sentido, pediu-nos que nos próximos dias concluíssemos a análise com o contributo do especialista recomendado pela colega e que criássemos um “ficheiro-modelo de *RMarkdown*” para que a equipa de jornalismo de dados pudesse partilhar futuras análises. Para esse feito, sugeriu uma leitura *online* mais aprofundada sobre *RMarkdown* e de um novo formato associado a ele, que tinha visto pela Internet, o “quattro”.

**Dia 61**, 24 de março de 2023, sexta-feira. Após termos contactado e trocado algumas impressões com o investigador sugerido pela Jornalista D, o mesmo sugeriu falar com uma outra investigadora especialista na área da pobreza, adequação do rendimento e de políticas de rendimento mínimo. As conversas conduzidas com ambos intervenientes resultaram em apontamentos e possíveis caminhos de investigação da

---

<sup>21</sup> O artigo que o jornalista Rui Barros que serviu de exemplo pode ser encontrado em: <https://cran.r-project.org/web/packages/biscale/vignettes/biscale.html> [consultado pela última vez a 5 de agosto de 2023]

análise. Essas anotações, bem como o contributo de cada um dos investigadores, foram incluídas no próprio documento de análise no *RStudio*.

Ao longo do dia, conseguimos concluir a análise de dados e criamos uma secção explicativa, no repositório do *Público* no *GitHub*, sobre como utilizar documentos *shapefile* no *RStudio*. Este repositório foi criado através da sintaxe básica de gravação e formatação da plataforma *GitHub*, que permite criar uma descrição sobre cada novo projeto. Isso auxilia a uma compreensão imediata sobre o que se trata aquela secção para quem a consultar. Neste caso em específico, denominamos o repositório de “*shapefiles-Portugal*”, onde carregamos três ficheiros *shapefile*: das freguesias, dos concelhos e dos distritos de Portugal. Concebemos também uma breve descrição sobre o que era cada um dos ficheiros, integrando títulos, subtítulos, uma tabela explicativa de cada uma das variáveis da *shapefile*, e trechos de código de linguagem R que explicavam como carregar e utilizar estes ficheiros no *RStudio*.

No final do dia, apresentamos o repositório a Rui Barros. Embora ainda não tivesse tido a oportunidade de rever a análise conduzida aos rendimentos (devido à extensão da mesma), sugeriu ainda consultar uma base de dados que tivesse informação sobre o número de beneficiários do RSI – a da Segurança Social. Quanto à criação do repositório com os mapas de Portugal, o colega mostrou-se bastante agradado e considerou que, no futuro, seria algo bastante útil para a equipa de jornalismo de dados.

“Como já te tinha dito, acho que vai ser impossível conseguirmos publicar o quer que seja dessa análise antes de ires embora, mas dá uma vista de olhos nos dados da Segurança Social para ver se a informação bate certo ou tem dados mais atualizados. Gostava ainda que trocássemos algumas impressões com a Jornalista C, mas já deves ter percebido que ando sempre atarefado e com coisas para fazer... Ainda assim, prometo que irei olhar para a análise que fizeste e depois dou-te algum *feedback* sobre a mesma” (Rui Barros, conversa informal).

Dada a aproximação da data final do nosso estudo, sugerimos a Rui Barros, José Volta e Pinto e Ivo Neto a marcação de um almoço para a próxima segunda-feira.

**Dia 62**, 27 de março de 2023, segunda-feira, 11h00. Após encontrarmos os dados de RSI no site da Segurança Social, apercebemo-nos que esta página não permite descarregar a informação como pretendíamos.

A instituição utiliza o *Microsoft Power BI*, integrado na secção das “Estatísticas” para apresentar visualmente os dados, mas o ficheiro para os descarregar não contempla a totalidade da informação que é possível consultar nos gráficos. Além disso, a Segurança Social disponibiliza os dados do RSI somente por distrito e os dados destoam dos do INE, pois refere-se a “beneficiários do RSI”. Por esse motivo, decidimos não cruzar estes dados com os da análise até aqui realizada.

Optamos, assim, por utilizar a informação para uma análise suplementar, tendo recorrido à coleta manual da informação num ficheiro *Excel*. A partir daqui, seguimos o processo semelhante que norteou o nosso fluxo de trabalho nas últimas semanas: importamos o documento *Excel* no *RStudio*, “limpamos” a base de dados, efetuamos cálculos para encontrar os distritos com maior e menor número de beneficiários de RSI, criamos representações visuais a partir dos dados e análises realizados, e anotamos as principais conclusões.

Com a aproximação da hora do almoço, Rui Barros informa-nos que Ivo Neto não poderá vir connosco, pois teve um contratempo. O almoço entre nós, Rui Barros e José Volta e Pinto decorreu de uma forma natural e orgânica. Os principais tópicos de conversa incidiram sobre o estado atual do jornalismo, questões políticas que marcavam a atualidade e os planos que tínhamos para futuro. Neste momento, sentimo-nos integrados neste núcleo de jornalismo de dados. No regresso à redação, Rui Barros presenteou-nos com a oferta de um livro “Factfulness”, de Hans Rosling, Anna Rosling Rönnlund e Ola Rosling, num gesto de companheirismo.

**Dia 66**, 31 de março de 2023, sexta-feira. Último dia na redação. Ao longo da semana, estivemos a ajustar pormenores estéticos do documento de análise aos rendimentos da população portuguesa no *RStudio*. Ontem, procedemos à última atualização semanal dos dados da Deco Proteste.

Hoje, excecionalmente, notamos que o ritmo de trabalho foi diferente, visto que a maioria das tarefas que nos haviam sido designadas encontravam-se concluídas.

Lamentamos apenas o facto de não termos conseguido realizar o “documento-modelo” em *RMarkdown* para futuras análises de dados para a equipa de jornalismo de dados.

A meio da tarde, damos conta da publicação de uma nova reportagem integrada na série de trabalhos “Contas à vida: enfrentar a inflação”. Esta segue uma estrutura semelhantes às reportagens anteriores, não se evidenciando qualquer tipo de gráfico nem menção à fonte dos dados mencionados.

Na conversa de despedida, realizada por volta das 18h30, Rui Barros referiu ter gostado da nossa prestação nos últimos meses, tendo enaltecido, sobretudo, a nossa evolução em operar com o *RStudio* e na redação de textos.

“- Atualmente, acho que poucos jornalistas estarão tão à vontade a analisar dados neste programa como tu. (Rui Barros).

- Obrigado, Rui. Mas ainda há ali algumas dúvidas que vou tendo, sobretudo nos cálculos mais estatísticos e assim. Acho que isso também se deve por não estar muito confortável com números... (Rafael Oliveira)

- Entendo e sim, houve ali um reparo ou outro que tive de fazer, mas não foi nada de grave, porque fomos a tempo de resolver [risos]. Mas também há quanto tempo já não lidavas com números? (Rui Barros)

- Pois... Se excluir a cadeira que tive no ano passado na faculdade, já lá vão alguns anos que não lidava com números. (Rafael Oliveira)

- Lá está. É uma questão de nos habituarmos e também há leituras que te posso recomendar que te ajudarão a lidar melhor com a Estatística. Mas, como deves ter percebido, a maioria são cálculos relativamente simples de fazer. Temos é de perceber se são os cálculos ou operações adequadas para o contexto em que os estamos a analisar. A minha principal sugestão é de continuares a apostar na tua formação nesta área, sobretudo na linguagem de programação, com o Svelte, e desenvolvimento Web. Acho que reconheces que foi o ponto em que tiveste mais dificuldades, o que é normal porque nunca lidaste com isso. Como deves saber, a análise de dados é uma área muito requisitada pelas empresas atualmente. Ainda assim, se gostas desta área, dedica-te ao jornalismo de dados e tenta aprender essas competências nos teus tempos livres. Eu costumo dizer que me tornei um desenvolvedor Web de forma involuntária, porque precisava de apresentar as histórias conduzidas por dados de forma apelativa e diferente. Então, tudo o que aprendi e vou aprendendo nessa área é de forma autodidata. No Secundário, frequentei o curso de Ciências e Tecnologias. Tinha boas notas a Matemática e, quando terminei o 12.º ano, os meus pais queriam

que seguisse algo relacionado com a área da Engenharia, mas não tinha muito interesse nisso. Sempre gostei de jornais, de notícias e quando chegou o momento de escolher, disse-lhes que ia seguir Jornalismo. É verdade que me safava bem com os números, mas não era uma área que me cativasse muito porque não entendia bem qual era a utilidade de algumas operações tão complexas para o meu dia-a-dia. Acho que isso é um dos problemas com a matemática. Na universidade é que isso me foi explicado de melhor forma, principalmente na área da Estatística. Depois partiu de mim descobrir e aprender fazer mais coisas. Comecei por tentar perceber como é que certas peças de jornalismo de dados eram feitas e a replicar o que faziam em alguns trabalhos, sempre tive essa curiosidade. Com o tempo, fui aperfeiçoando certos aspetos e adquirindo mais experiência.” (Rui Barros, conversa informal).

## Anexo 2 – Corpo da amostra

Código da peça	Hiperligação das peças de jornalismo de dados do Público	Autor(es) da peça	Data de publicação e/ou atualização da peça	Título da peça
Peça 1	<a href="https://www.publico.pt/interactivo/inflacao-portugal">https://www.publico.pt/interactivo/inflacao-portugal</a>	Rui Barros	Atualização semanal (análise compreende o período de 02/01/2023 até 31/03/2023)	Conheça o impacto da inflação no seu carrinho de compras
Peça 2	<a href="https://www.publico.pt/2023/01/23/economia/noticia/arroz-carolino-ja-custa-quase-dobro-2035410">https://www.publico.pt/2023/01/23/economia/noticia/arroz-carolino-ja-custa-quase-dobro-2035410</a>	Rafael Oliveira	23/01/23	Porque é que o arroz carolino já custa quase o dobro?
Peça 3	<a href="https://www.publico.pt/interactivos/inflacao-precos-poupanca">https://www.publico.pt/interactivos/inflacao-precos-poupanca</a>	PÚBLICO (Rui Barros) e José Alves	02/02/2023	Contas à vida: enfrentar a inflação
Peça 4	<a href="https://www.publico.pt/interactivos/miguel-estafeta-emprego-uber-inflacao-poupanca/">https://www.publico.pt/interactivos/miguel-estafeta-emprego-uber-inflacao-poupanca/</a>	Joana Gonçalves (texto e vídeo), Rui Barros (desenvolvimento)	02/02/2023	Miguel tornou-se estafeta para cumprir o sonho do filho
Peça 5	<a href="https://www.publico.pt/interactivos/ana-estudante-universidade-propinas-renda-">https://www.publico.pt/interactivos/ana-estudante-universidade-propinas-renda-</a>	Joana Gonçalves (texto e vídeo), Rui Barros	02/02/2023	Ana “implorou” por um quarto para continuar a estudar em Lisboa

	<a href="#">lisboa-inflacao-poupanca</a>	(desenvolvimento)		
Peça 6	<a href="https://www.publico.pt/interactivos/mapa-rendas-onde-consigo-viver/">https://www.publico.pt/interactivos/mapa-rendas-onde-consigo-viver/</a>	Rui Barros, José Volta e Pinto, Aitor Varea Oro (MDT-CEAU-FAUP, da Universidade do Porto)	20/02/2023	Onde é que consigo arrendar uma casa com o meu salário?
Peça 7	<a href="https://www.publico.pt/interactivos/portugueses-mais-insatisfeitos-trabalho-europa/">https://www.publico.pt/interactivos/portugueses-mais-insatisfeitos-trabalho-europa/</a>	Rafael Pereira Oliveira, Rui Barros	06/03/2023	Portugal tem dos trabalhadores mais insatisfeitos da Europa. Porquê?
Peça 8	<a href="https://www.publico.pt/interactivos/nuno-pequeno-produtor-palmela-supermercado-inflacao-poupanca/">https://www.publico.pt/interactivos/nuno-pequeno-produtor-palmela-supermercado-inflacao-poupanca/</a>	Joana Gonçalves (texto e vídeo), Pedro Rios (entrevista telefónica), Rui Barros (desenvolvimento)	09/03/2023	Nuno tem uma vinha: “Os preços aumentam, mas não recebemos mais”
Peça 9	<a href="https://www.publico.pt/interactivos/mapa-vagos-freguesias/">https://www.publico.pt/interactivos/mapa-vagos-freguesias/</a>	Luísa Pinto, Rui Barros	17/03/2023	Quantas casas vagas há na minha freguesia? Veja o mapa
Peça 10	<a href="https://www.publico.pt/interactivos/habitacao-renda-clara-dividir-casa/">https://www.publico.pt/interactivos/habitacao-renda-clara-dividir-casa/</a>	Joana Gonçalves (texto e vídeo), Joana Bourgard (entrevista telefónica), Rui Barros (desenvolvimento)	31/03/2023	Aos 60 anos, Clara vai dividir casa para “não cair no abismo”

### Anexo 3 – Análise de conteúdo

Tabela 4 - Quadro de análise à hipertextualidade

Código da peça	1.1.1 - Hiperligações intratextuais	1.1.2 - Hiperligações extratextuais	1.1.3 - Hiperligações genéricas
Peça 1		No total foram identificadas seis hiperligações extratextuais: uma no nome do autor da peça, que remete para uma página onde é possível consultar o endereço eletrónico do jornalista e as peças realizadas pelo mesmo; três hiperligações para partilhar a peça em redes sociais identificadas pelo logótipo correspondente (Facebook, Twitter e LinkedIn) e outras duas hiperligações referentes ao modelo tridimensional "desenhado por AdrianXY" e página do mesmo ("AdrainXY").	
Peça 2	Nove hiperligações identificadas no "miolo do texto", das quais oito são hiperligações internas de notícias do jornal <i>Público</i> : [1.ª hiperligação: "(...) As sementes de arroz, à venda nas cooperativas agrícolas, são importadas maioritariamente a partir dos dois maiores produtores na União Europeia (UE): <u>Espanha</u> (...)"; 2.ª hiperligação: "(...) e <u>Itália</u> . (...)"; 3.ª hiperligação: "(...) E <u>Portugal não foi exceção</u> . No Mondego, houve "uma quebra até 20%" na produção de arroz (...)"; 4.ª hiperligação: "(...) A juntar a esta espiral de custos, há ainda o <u>conflito armado na Ucrânia</u> (...)"; 5.ª hiperligação: "(...) A guerra provocou um aumento substancial no custo da energia, com o <u>preço do gás a subir de</u>	No total foram identificadas 12 hiperligações deste tipo: cinco para partilhar a peça nas redes sociais identificadas pelo logótipo correspondente (Facebook, WhatsApp, Twitter, LinkedIn e Pinterest), três para enviar o artigo através de e-mail, guardar o artigo na conta de utilizador do <i>Público</i> e para comentar a peça, e outras quatro hiperligações extratextuais embutidas ao longo do texto, que são sugestões de notícias publicadas pelo jornal relacionadas com o tema da inflação, nomeadamente nos setores da alimentação e do gás.	Identificada uma hiperligação genérica entre as nove hiperligações intratextuais: Apenas uma é de entidade externa ("De acordo com o <u>portal dos preços dos combustíveis</u> " da Direcção-Geral de Energia e Geologia (DGEC) (...)).

	<p>forma acentuada (...); 6.ª hiperligação: “(...) e <u>a electricidade</u> a atingir máximos históricos. (...)”; 7.ª hiperligação: “(...) <u>Também na indústria contam os aumentos</u> no custo de transporte, do embalamento que “aumentou muito” devido ao custo do plástico (...)”; 8.ª hiperligação: “(...) Para entender o contraste de preços entre tipos de arroz, o director-geral da ANIA aponta duas razões: a inflação da energia e do gás estar “muito concentrada” na UE, devido ao <u>conflito armado</u>, e o facto de que (...)”.]</p> <p>Apenas uma é de entidade externa (“De acordo com o <u>portal dos preços dos combustíveis</u>” da Direcção-Geral de Energia e Geologia (DGEC) (...)”) e remete para uma página da Direcção-Geral de Energia e Geologia, onde é possível consultar o preço dos combustíveis.</p>		
Peça 3	Hiperligação que permite ao leitor participar num inquérito realizado pelo jornal <i>Público</i> (“Contribua com o seu caso e ajude-nos a pintar o gráfico”).	<p>Três hiperligações distintas para partilhar a peça em três redes sociais identificadas pelo logótipo correspondente (Facebook, Twitter e LinkedIn).</p> <p>Diversas hiperligações surgem embutidas nas imagens de artigos e/ou títulos que surgem ao longo da página, permitindo ler centenas de testemunhos, consultar quatro reportagens realizadas a partir do inquérito do jornal e ver artigos relacionados com o tema da inflação.</p>	
Peça 4		Identificadas seis hiperligações extratextuais: três de partilha da reportagem em redes sociais identificadas pelo logótipo	

		correspondente (Facebook, Twitter e LinkedIn), duas no nome dos autores do artigo (Joana Gonçalves e Rui Barros) e uma peça embutida a meio do texto ("Contas à vida: enfrentar a inflação") relacionada com a reportagem, visto que se trata de um dos testemunhos que respondeu ao questionário lançado por este órgão.	
Peça 5		Foram identificadas sete hiperligações extratextuais, das quais três são de partilha da peça em redes sociais (Facebook, Twitter e LinkedIn), duas no nome dos autores (Joana Gonçalves e Rui Barros) e duas em peças embutidas, servindo como recomendação de notícias e leitura de um outro testemunho ("Contas à vida: enfrentar a inflação" e "O meu contrato de renda termina nos próximos meses e receio ter de procurar casa noutra cidade" )	
Peça 6		Identificadas sete hiperligações extratextuais: uma na temática da peça ("Habitação em Portugal"); três em cada nome dos autores da peça (José Volta e Pinto, Rui Barros e Aitor Varea Oro (MDT-CEAU-FAUP, da Universidade do Porto)) e a hiperligação de partilha em três redes sociais (Facebook, Twitter e LinkedIn)	
Peça 7	Identificadas 15 hiperligações intratextuais.  13 hiperligações embutidas no "miolo do texto", referentes a notícias relacionadas do <i>Público</i> : [1.ª hiperligação: "(...) No ano em	Fora do texto da peça, foram identificadas nove hiperligações extratextuais: três delas são recomendações de notícias relacionadas "fora do texto", três hiperligações de partilha	Identificada duas hiperligações genéricas, de entre as 15 elencadas nas hiperligações intratextuais, que

	<p>que os institutos nacionais de estatística de cada país realizaram o inquérito da satisfação em contexto profissional, o mundo laboral confrontava-se ainda com o <u>"choque" da pandemia de covid-19. (...)</u>"; 2.ª hiperligação: "(...) Mas, para <u>Elísio Estanque</u>, a insatisfação na região portuguesa "já era crescente desde a crise e o período da troika". (...)" ; 3.ª hiperligação: "(...) Mas Portugal não é o país que paga os salários mais baixos da União Europeia. O problema, diz <u>Ricardo Paes Mamede</u>, reside também no poder de compra. (...)" ; 4.ª hiperligação: "(...) No entanto, para compreender o que afecta a generalidade dos trabalhadores, o economista Ricardo Paes Mamede vê outros "potenciais factores explicativos", como o "<u>número muito elevado de horas trabalhadas</u>" e a qualidade da gestão. (...)" ; 5.ª hiperligação: "(...) E o "desencantamento" é evidente nos <u>profissionais com licenciaturas, mestrados ou doutoramentos. (...)</u>" ; 6.ª hiperligação: "(...) Ao analisar quem assinalou estar pouco satisfeito com a condição profissional, <u>as mulheres empregadas</u> colocam Portugal nos lugares cimeiros. (...)" ; 7.ª hiperligação: "(...)O investigador Elísio Estanque conta que a <u>desigualdade salarial entre os dois sexos</u> em Portugal "é das maiores da União Europeia (...)" ; 8.ª hiperligação: "(...)a divisão de tarefas no seio familiar "<u>continua a pesar mais na mulher</u>". (...)" ; 9.ª hiperligação: "(...) Noutros países, a distribuição de tarefas "é mais efectiva", diz Gabriel Leite Mota, que identifica a "<u>difícil ascensão profissional</u>" (...)" ; 10.ª hiperligação: "(...) e o <u>"machismo"</u> em contexto laboral</p>	<p>em redes sociais (Facebook, Twitter e LinkedIn), duas no nome respetiva dos autores (Rafael Pereira Oliveira e Rui Barros) e uma identificativa da temática da peça ("Trabalho").</p>	<p>levam o leitor até à página de um relatório publicado pela Organização Internacional do Trabalho, a segunda redireciona o leitor até à página do <i>Eurostat</i>, onde é possível consultar a base de dados original utilizada nesta peça.</p>
--	--	--	---

	<p>como outros potenciais motivos. (...); 11.ª hiperligação: "(...) <u>"Temos os funcionários públicos com salários estagnados há quase 20 anos e com perdas de poder de compra que chegam a atingir os 20% nesse período. (...)"</u>; 12.ª hiperligação: "(...) Se olharmos para 2021, junta-se uma pressão brutal sobre <u>serviços subfinanciados</u> para lidar com situações decorrentes da pandemia", relembra. (...); 13.ª hiperligação: "(...) Mas a situação também não é animadora para <u>quem criou o seu próprio posto de trabalho</u> e emprega, pelo menos, uma pessoa. (...)"</p> <p>As outras duas hiperligações redirecionaram para páginas de entidades externas: " (...) Um "mecanismo político" realçado no <u>relatório da Organização Internacional do Trabalho</u> (Trabalho Digno em Portugal 2008-2018: Da crise à recuperação) é o do salário mínimo.", que redireciona o leitor para o relatório referido, e "Por outro lado, <u>os dados reunidos pelo Eurostat</u> "não são uma surpresa (...)", remete o leitor para a base de dados do Eurostat utilizada na peça.</p>		
Peça 8	<p>Identificadas quatro hiperligações embutidas no "miolo do texto". [1.ª hiperligação: "(...)Encostado a uma velha oliveira à entrada da Quinta da Saudade, Nuno explica que quando enviou o testemunho ao <u>Contas à Vida(...)</u>"; 2.ª hiperligação: "(...)Para o jovem de 27 anos, <u>o impacto da inflação no carrinho de compras</u> não se reflecte, muitas vezes, em mais receita para o produtor. E isso, assegura, (...)" ; 3.ª hiperligação: "(...)De acordo com a avaliação da Deco Proteste, organização</p>	<p>Foram identificadas oito hiperligações extratextuais, entre as quais: três de partilha da peça em redes sociais (Facebook, Twitter e LinkedIn), três no nome dos autores (Joana Gonçalves, Pedro Rios e Rui Barros) e duas de peças embutidas - uma de leitura na íntegra de outro testemunho ("A ida ao supermercado tornou-se um momento de aflição constante - Miguel, Lisboa, 35 anos") e a outra de uma</p>	

	<p>portuguesa de defesa do consumidor, o preço do cabaz de produtos essenciais <u> aumentou 24% no último ano. (...)</u>].</p> <p>Todas são páginas de notícias publicadas no site do Público e relacionadas com a temática da peça.</p>	<p>peça relacionada ("Contas à vida: enfrentar a inflação").</p>	
Peça 9	<p>Identificadas três hiperligações intratextuais no corpo do texto: [1.ª hiperligação: "(...) enquadrado no pacote legislativo Mais Habitação, que o Governo tem ainda em discussão pública até dia 24 de Março. (...); 2.ª hiperligação: "" A vereadora da Habitação da Câmara Municipal de Lisboa, Filipa Roseta, já disse que não pretende empenhar muitos esforços na execução desta medida, porque elas não são prioritárias e vão aumentar a (...); 3.ª hiperligação: "Esta medida é também uma daquelas que o Presidente da República, Marcelo Rebelo de Sousa, já questionou se seria concretizável."]. Todas são referentes a notícias do <i>Público</i>.</p>	<p>Estão presentes sete hiperligações extratextuais: uma referente à secção da notícia ("Habitação"), duas no nome dos autores (Luísa Pinto, Rui Barros) e quatro de partilha em redes sociais (Facebook, Twitter, LinkedIn e WhatsApp).</p>	
Peça 10	<p>Uma hiperligação intratextual foi identificada em "(...) A música - <u>De uns tempos para cá</u> - foi lançada em 2006 (...). Hiperligação remete o leitor para a música "De uns tempos para cá", de Chico César, na plataforma do <i>Spotify</i>.</p>	<p>Sete hiperligações extratextuais identificadas, das quais três são dos autores da peça (Joana Gonçalves, Joana Bourgard, Rui Barros), três de partilha da peça nas redes sociais (Facebook, Twitter e LinkedIn) e uma de um artigo embutido relacionado com esta peça ("Contas à vida: enfrentar a inflação").</p>	

Nota: as hiperligações intratextuais identificadas estão identificadas com sublinhado.

Tabela 5 – Quadro de análise à multimedialidade

<b>Código da peça</b>	<b>1.2.1 - Fotografia ou desenho</b>	<b>1.2.2 - Diaporama</b>	<b>1.2.3 – Infografia / gráfico/ mapa estático</b>	<b>1.2.4 - Infografia/ gráfico/ mapa dinâmico</b>	<b>1.2.5 - Áudio</b>	<b>1.2.6 - Vídeo c/ som</b>	<b>1.2.7 - Vídeos/ som</b>
Peça 1	Modelo tridimensional de um carrinho de compras		O artigo conta com uma tabela estática, intitulada "Produtos que mais aumentaram", que estabelece uma lista de 10 produtos ordenados pelo maior aumento na "taxa de crescimento" face a 23 de fevereiro de 2022, apresentando ainda os valores de cada um dos produtos nos dois períodos temporais em análise.	Presença de um gráfico de linhas interativo, intitulado de "Evolução do preço do cabaz de produtos essenciais", e com descrição ("O cabaz de produtos essenciais da Deco Proteste reúne 63 produtos, nos quais se incluem produtos como frango, pescada, cebola, banana, arroz, esparguete, leite, queijo ou manteiga."). A linha é roxa e é possível consultar com o cursor o preço do cabaz em diferentes períodos e a data correspondente. No final			

			<p>do gráfico é indicada a fonte dos dados ("Fonte: Deco Proteste"), o autor da concepção do gráfico ("Infografia   PÚBLICO").</p> <p>Presença de seis gráficos de linhas dinâmicos que corresponde m a distintas categorias de produtos. Contêm um título principal ("Variação de preço por tipo de produto"), uma descrição ("Taxa de crescimento, em percentagem , do total pago pelos mesmos produtos face ao dia 23 de Fevereiro de 2022") e a identificação textual - com um emoji representativo - de cada uma das categorias acima dos</p>			
--	--	--	--	--	--	--

				<p>gráficos. A linha é vermelha e, à semelhança do anterior, é possível consultar com o cursor a percentagem de variação de preço por cada tipo de produto. Em cada um dos gráficos é apresentada a "percentagem de variação face a 23 de Fevereiro de 2022" e uma pequena lista de produtos abaixo de cada categoria ("Mercearia: Inclui despesas em produtos como feijão, grão, óleo alimentar, azeite, sal, atum, massa, fiambre, arroz, pão ou cereais (entre outros."); "Carne: Inclui a despesa em carne como lombo de porco sem osso, frango,</p>			
--	--	--	--	--	--	--	--

				febras, costelas de porco, bife de peru, novilho e perna de peru." ; "Peixe: Inclui a despesa em peixes como bacalhau graúdo, dourada, salmão, pescada fresca, carapau, peixe espada preto, robalo e perca." ; "Lacticínios: Inclui despesas em produtos como queijo curado, ovos classe M/L, iogurtes, manteiga, queijo flamengo ou leite meio- gordo." ; "Frutas e legumes: Inclui despesas em produtos como maçãs, bananas, tomates, couve-flor, alface, bróculos, cenoura, batata, alho (entre outros)." ; "Congelados: Inclui			
--	--	--	--	--	--	--	--

				despesas em produtos como douradinhos de peixe, ervilhas ultracongeladas e medalhões de pescada."). O final do gráfico indica a fonte dos dados ("Fonte: Deco Proteste"), o autor da conceção do gráfico ("Infografia   PÚBLICO").			
Peça 2	Fotografia de uma mão a segurar um punhado de arroz.			Presença de um gráfico de linha interativo, intitulado de "Evolução do preço do arroz carolino desde Fevereiro de 2022", e com descrição ("O preço médio atingiu um novo máximo a 18 de Janeiro de 2023. Com base em 11 marcas, este produto está a custar, em média, mais de dois euros nos supermercados"). A linha			

				<p>é laranja e com o cursor é possível consultar a evolução semanal do preço do arroz carolino desde 23 de fevereiro de 2022 até 18 de janeiro de 2023. No final do gráfico é indicada a fonte dos dados ("Fonte: Deco ") e o autor da concepção do gráfico e os créditos, bem como o software utilizado para a criação do mesmo ("PÚBLICO   Base Flourish").</p> <p>Presença de um gráfico de árvore (do inglês "treemap") dinâmico com o título "Até novembro de 2022, mais de 50% do arroz importado veio da América do Sul" e com</p>			
--	--	--	--	---	--	--	--

				<p>uma descrição ("Outros tipos de arroz vendidos em Portugal não registaram um aumento tão expressivo. A maioria (quase 70%) é proveniente de países "imunes" aos efeitos da guerra."). No final do gráfico é indicada a fonte dos dados ("Fonte: Eurostat ") e o autor da conceção do gráfico e o software utilizado para a criação do mesmo ("PÚBLICO   Base Flourish"). Este gráfico é composto de representações geométricas circulares de diferentes tamanhos e cores, sendo representativos de quatro continentes: América do Sul, América</p>			
--	--	--	--	---	--	--	--

				<p>do Norte, Europa e Ásia. Ao passar o cursor neste círculos, o nome de cada um dos continentes é novamente exibido e, abaixo, num tamanho de letra menor, surge um valor percentual representativo do peso das importações de arroz em Portugal de cada continente. Dentro de cada um destes círculos, é possível encontrar as mesmas formas geométricas, com diferentes tamanhos, e apresentam o nome de certos países do respetivo continente. Ao passar o cursor por cima, o nome do país é exibido num "pop-up", contendo as seguintes</p>			
--	--	--	--	--	--	--	--

				<p>informações: "Nome do país", "Peso das importações de arroz em Portugal", "[valor percentual] até novembro de 2022".</p> <p>Alguns destes círculos são bastante pequenos, mas, ao clicar dentro do círculo, é possível visualizar com maior detalhe os países que compõem a percentagem total do continente. Neste passo, surge uma pequena seta que remete o leitor para o plano original do gráfico, evidenciando-se uma dinâmica de interatividade.</p>			
Peça 3				<p>Presença de um gráfico circular interativo, sem título, no topo da página que ilustra, por</p>			<p>As quatro reportagens integradas nesta página inicial contemplam</p>

				<p>cada mês, o número de respostas obtidas no inquérito sobre a situação financeira. Essa informação é representada numa escala de cor, com uma legenda que surge abaixo do gráfico: "Piorou" (a vermelho-escuro), "Manteve" (vermelho), "Melhorou" (rosa).</p>			<p>am um pequeno excerto de vídeo sem som.</p>
Peça 4					<p>A peça é introduzida com o áudio e transcrição escrita da conversa telefónica entre a jornalista e a pessoa entrevistada.</p>	<p>Identificados quatro vídeos com som.</p> <p>O primeiro deles, com cerca de 55 segundos de duração, é introduzido e sobreposto com um excerto textual destacado ("Cerca de 238 mil trabalhadores, em Portugal, acumulam dois ou mais empregos"), seguindo-se o título da peça</p>	

						<p>("Miguel tornou-se estafeta para cumprir o sonho do filho") e subtítulo ("Para garantir as aulas de piano do filho, Miguel faz entregas à noite. São 84 euros a mais no orçamento de quem já tem de fazer contas à vida"). Este vídeo mostra o entrevistado a sair de sua casa com um capacete na mão para ir trabalhar como estafeta, carregando uma mala de uma empresa de entrega de refeições e alimentos online às costas. As imagens são intercaladas com paisagens noturnas, o seu filho a tocar piano e o entrevistado a trabalhar.</p>	
--	--	--	--	--	--	--	--

						<p>Após a introdução e desenvolvimento textual da reportagem, o leitor pode encontrar um vídeo legendado com som, com dois minutos e seis segundos de duração, que complementa o testemunho da pessoa entrevistada e dá a conhecer a habitação do entrevistado, a sua família e o segundo ofício que possui (estafeta) para fazer face às despesas decorrentes da inflação. Ainda que esta opção não seja evidente, ao clicar com o botão direito do rato, o leitor poderá ativar a opção "mostrar todos os controlos", permitindo controlar o volume do</p>	
--	--	--	--	--	--	--	--

					<p>som, colocar o vídeo em pausa, ocupar a totalidade do ecrã, navegar pela barra temporal do vídeo, transferir o vídeo e aumentar ou diminuir a sua velocidade de reprodução. Nos últimos segundos do vídeo, surge uma pequena mensagem na parte inferior, com uma seta apontada para baixo, que diz "continuar a ler".</p> <p>O terceiro vídeo, com cerca de 30 segundos, mostra o entrevistado a subir e a ligar uma mota, de mala às costas, para dar início ao seu turno noturno de estafeta. O vídeo é sobreposto com as duas seguintes</p>	
--	--	--	--	--	---	--

					<p>frases destacadas:  "Em 2022, um assistente técnico da função pública recebia 757 euros brutos" e  "Desde Janeiro de 2023, Miguel passou a receber 861 euros".</p> <p>O último vídeo, também com 30 segundos de duração, mostra o entrevistado a trabalhar, subindo para a sua mota e conduz em direção a uma rua sem saída.  Surgem outros dois destaques textuais:  "Miguel paga 84 euros de mensalidade por uma aula de piano por semana" e  "Com as entregas como estafeta Miguel recebe, em média, 450 euros extra por mês".</p>	
--	--	--	--	--	--	--

Peça 5					<p>À semelhança do artigo anterior, esta peça abre com o áudio de uma conversa telefónica entre a jornalista e a pessoa entrevistada e existe a transcrição do áudio.</p>	<p>Foram identificados quatro vídeos com som.</p> <p>O primeiro vídeo, com cerca de 15 segundos, surge após a conversa telefónica e é sobreposto com um pequeno excerto textual e o título da peça ("Até Janeiro de 2023, mais de 105 mil estudantes tinham pedido bolsa de estudo no ensino superior"; título da peça: "Ana "implorou" por um quarto para continuar a estudar em Lisboa"; subtítulo da peça: "Com um rendimento de 150 euros mensais, um aumento de 15 euros na renda foi suficiente para tornar a estadia em Lisboa incomportáv</p>	
-----------	--	--	--	--	---	---	--

					<p>el. Durante seis meses, Ana teve de recorrer às poupanças para continuar a estudar na capital"). O vídeo tem o som ambiente do local onde que a entrevistada se encontra, sendo possível ouvir a mesma a folhear um livro.</p> <p>O segundo vídeo com som, de um minuto e 22 segundos, é o mais extenso na sua duração, surgindo a seguir à introdução textual da reportagem. O vídeo possui legendagem do testemunho da pessoa entrevistada que dá a conhecer a situação pessoal, contando com excertos de vídeo da faculdade</p>	
--	--	--	--	--	---	--

					<p>que frequenta. À semelhança do vídeo com som identificado na Peça 4, o leitor pode ativar a opção "mostrar todos os controlos". Este é o único vídeo que permite ativar essa função e nos instantes finais surge uma pequena mensagem com uma seta para baixo ("continuar a ler").</p> <p>O terceiro vídeo, de 10 segundos, mostra a entrevistada de costas, sentada, e ouve-se o som ambiente de alguns carros a passar. O vídeo é sobreposto com dois excertos textuais destacados ("Nunca tantos estudantes tinham</p>	
--	--	--	--	--	--	--

					<p>pedido bolsa de estudo no ensino superior" e "Só alunos de famílias com rendimento per capita inferior a 8.962 euros podem receber bolsas de ação social").</p> <p>O último vídeo, com cerca de 12 segundos, mostra a entrevistada a caminhar em direção a umas escadas, sendo audível os seus passos e algumas vozes de fundo a conversar. O vídeo é sobreposto com dois excertos de texto: "Ana pagava 390 euros por um quarto em Lisboa" e "Para bolseiros o custo da residência pode ser reembolsado na totalidade".</p>	
--	--	--	--	--	---	--

Peça 6			<p>Identificado um mapa dinâmico colorido de Portugal Continental. O título do mapa pode ser considerado o título da peça ("Onde é que consigo arrendar uma casa com o meu salário?") e conta com um subtítulo ("A crise da habitação não afecta todo o país da mesma forma. Arrendar casa custa mais em algumas zonas, especialmente e nas grandes cidades. Veja em que freguesias ou municípios consegue viver com os seus rendimentos").</p> <p>No que concerne ao mapa este é introduzido com a frase "Selecione</p>			
-----------	--	--	--	--	--	--

				<p>as opções que se aplicam ao seu caso e indique o seu rendimento" e é visível uma descrição explicativa no interior desta representação o visual de dados ("Passe o cursor no mapa para ver quanto do seu rendimento mensal seria necessário para conseguir viver em determinado concelho ou freguesia"). Acima do mapa, existem três categorias, das quais duas são de seleção pré-definida ("Quero viver num [apartamento, casa em banda, casa geminada ou casa isolada]", "Com a área equivalente a um [T0, T1, T2, T3, T4]) e uma de preenchimen</p>			
--	--	--	--	--	--	--	--

				<p>to manual ("O meu rendimento é de [valor numérico] euros líquidos por mês"). O mapa ilustra algumas freguesias da Área Metropolitana do Porto e de Lisboa, assim como de alguns município do continente, e contém uma escala de cor como legenda ((Rendimento gasto em renda (em percentagem)"). No final do mapa existe uma nota medológica que pode ser consultada: "Nota: os dados deste mapa reflectem a renda de mercado que, segundo o Programa de Apoio ao Arrendamento, teria uma habitação com as características seleccionada</p>			
--	--	--	--	--	--	--	--

				<p>s no primeiro semestre de 2022. São apresentados dados para cada município ou freguesia para o qual o INE disponibiliza dados. Este valor foi calculado considerando a mediana do preço por m<sup>2</sup> dos novos contratos de arrendamento, a área bruta privativa e o “coeficiente de qualidade e conforto”, seguindo as regras definidas no Programa de Arrendamento Acessível. Em termos de áreas, foram consideradas as seguintes:</p> <p>T0 - 35 m<sup>2</sup>;  T1 - 52 m<sup>2</sup>;  T2 - 72 m<sup>2</sup>;  T3 - 91 m<sup>2</sup>;  T4 - 105 m<sup>2</sup>.</p> <p>Já no chamado “coeficiente de qualidade e conforto”, foi considerada, em todos os</p>			
--	--	--	--	--	--	--	--

				<p>casos, uma habitação não mobilada, com cozinha equipada, lugar de estacionamento, estado de conservação considerado “bom” e classe energética ‘C’. Os resultados apresentados servem apenas como referência, uma vez que os valores medianos não reflectem os múltiplos factores que influenciam a renda numa determinada região”.</p>			
Peça 7	<p>Ilustração de um homem tristonho a caminhar e que carrega algumas frutas num saco, fazendo alusão à ideia de Portugal ter dos trabalhad</p>		<p>Identificados dois gráficos de barras estáticos: o primeiro deles trata-se de um gráfico de barras horizontal, intitulado "Média de horas semanais</p>	<p>No total, foram identificadas oito representações visuais de dados interativas, entre as quais: seis gráficos de barras de orientação horizontal, sendo que um destes trata-se de um gráfico</p>			

	<p>ores mais insatisfeitos da Europa.</p>		<p>de trabalho efectivo em 2021" e com descrição ("Em Portugal trabalhou-se, em média, 38 horas semanais. Os Países Baixos foi onde se registou menos horas de trabalho efectivo, e a Grécia onde se trabalhou mais (40 horas por semana)"). Cada uma das barras ilustra os valores totais de horas trabalhadas e o nome do país. Há também a indicação da média de horas trabalhadas na UE, no traço descontínuo vertical. Portugal surge</p>	<p>de barras horizontal agrupado, um mapa e um gráfico de dispersão.</p> <p>A primeira representação visual de dados trata-se de um gráfico de barra horizontal, com o título "Satisfação laboral na União Europeia em 2021", e contém uma pequena legenda colorida que identifica o nível de satisfação laboral (azul-ciano: alta; amarelo: média; roxo: baixa; laranja: nenhuma destas; cinzento: não respondeu). Além disso, existe um pequeno excerto textual ao lado do gráfico, que acompanha o <i>scroll</i> do leitor ("Em 2021, mais de 197 milhões de</p>			
--	---	--	--	---	--	--	--

		<p>com a cor laranja, enquanto outros países seguem um tom verde. A fonte (Eurostat), autor (PÚBLICO) e ferramenta de criação (Base <i>Flourish</i>) são mencionados no rodapé do gráfico.</p> <p>A segunda representação visual de dados estática é também um gráfico de barras vertical, que ilustra o "Ganho médio mensal entre mulheres e homens", contando com descrição abaixo do título</p>	<p>trabalhadores na União Europeia foram entrevistados sobre o nível de satisfação laboral. Os dados reunidos pelo Eurostat mostram que 43,8% das pessoas estavam "muito satisfeitas" com a condição profissional".</p> <p>) Acima da única barra deste gráfico surge uma pequena informação ("Média da UE (27 países))" e na linha é possível identificar a média das respostas dos inquiridos, expressa em valores percentuais, sendo possível ver em detalhe os mesmos ao passar o cursor. No final do gráfico é identificada a fonte ("Fonte:</p>			
--	--	--	---	--	--	--

			<p>("Os homens empregados por conta de outrem receberam mais 223,6 euros mensais do que as mulheres no ano de 2021"). Existe ainda uma legenda, que identifica as "mulheres" com a cor laranja, e os homens com a cor verde, bem como uma nota: "Nota: dados referentes apenas a Portugal Continental"). Por fim, a fonte dos dados é identificada (Fonte: GEP - Ministério do Trabalho,</p>	<p>Eurostat"), assim como os créditos e a ferramenta utilizada para criar o gráfico ("PÚBLICO   Base Flourish").</p> <p>De seguida, surge um mapa com diferentes graduações de cor azul-ciano, com o título "Variação do nível de satisfação na Europa" e com a descrição textual ao lado ("O indicador da alta satisfação assume diferentes pesos em cada um dos Estados-membros da UE, assim como na Noruega e na Suíça. Malta, Letónia e Suíça são os países onde se registam os valores mais elevados: mais de 65% dos entrevistados estavam</p>			
--	--	--	--	--	--	--	--

			<p>Solidariedade e Segurança Social (2021)), o autor (PÚBLICO) e a ferramenta utilizada (Base Flourish).</p>	<p>"altamente satisfeitos" com o seu emprego"). À semelhança do anterior, é identificada a mesma fonte, os créditos e a ferramenta utilizada para criar o mapa. Existe ainda uma nota que refere "Nota: A base de dados reúne as respostas de 27 países da União Europeia, aos quais acrescem a Noruega e a Suíça". O mapa permite ao leitor passar o cursor e consultar os diferentes níveis de "alta satisfação" nas diferentes regiões.</p> <p>Volta a surgir um gráfico de barras semelhante ao primeiro, com a mesma legendagem e opções de consulta, e existem duas</p>			
--	--	--	--	---	--	--	--

				<p>barras: uma identificando a média da UE (primeiro gráfico) e uma segunda, abaixo da primeira, com os valores de Portugal. O título é "Em Portugal, apenas 21,6% dos trabalhadores estavam muito satisfeitos" e a descrição do lado esquerdo diz: "Em contraste com a média da comunidade europeia, Portugal tem menos de um quarto dos trabalhadores felizes com a sua profissão (21,6%). A maioria respondeu estar "mais ou menos satisfeita", enquanto 10,6% das pessoas expressaram a sua insatisfação". O gráfico contém a</p>			
--	--	--	--	---	--	--	--

			<p>mesma fonte, autor e ferramenta utilizada para a sua concepção.</p> <p>Ao fazer <i>scroll</i> pela página, surge um novo gráfico de barras horizontal, com 29 barras em ordem decrescente, cada uma correspondendo a países europeus. O título é "Portugal é o país com a menor percentagem de trabalhadores muito satisfeitos", e é acompanhado pela seguinte descrição textual: "Esse valor mostra que Portugal é o país com a menor percentagem de trabalhadores muito satisfeitos quando comparado com os outros</p>			
--	--	--	---	--	--	--

				<p>países. A vizinha Espanha, por exemplo, ultrapassa a média da UE: mais de metade dos entrevistados (54%) disseram estar altamente satisfeitos". A média da UE é apresentada num traçado descontínuo vertical, identificado a média na base do gráfico. 28 destas barras tem a cor azul-ciano, sendo que Portugal surge com uma cor distinta dos restantes países (laranja). Ao passar o cursor no gráfico, o leitor pode verificar os níveis de "alta satisfação profissional" em cada uma das regiões. O gráfico volta a identificar a mesma fonte, autor</p>			
--	--	--	--	---	--	--	--

				<p>e ferramenta.</p> <p>Posteriormente, surge um gráfico bastante semelhante ao último, com 29 barras em ordem decrescente, representando cada país da amostra. "E é o segundo país com mais trabalhadores insatisfeitos" é o título deste gráfico, acompanhado por uma nova descrição textual ao lado ("No indicador da insatisfação, o país volta a surgir nas piores posições. É o segundo da lista com a maior percentagem de trabalhadores insatisfeitos, tendo quase o dobro da média europeia (5,8%). É apenas superado</p>			
--	--	--	--	--	--	--	--

				<p>pela Bulgária (11,8%). Por outro lado, a Itália é o país com menor percentagem"). O valor percentual da média da UE volta a ser apresentado num traço descontínuo vertical, sendo o mesmo discriminado na base do gráfico. Tal como no anterior, todas as barras são de cor roxa, com a exceção de Portugal com a cor laranja. Novamente, o gráfico identifica os mesmos elementos supramencionados.</p> <p>Após a componente de <i>scrollytelling</i>, a peça segue o estado "normal" e a primeira representação visual de dados dinâmica no corpo da reportagem</p>			
--	--	--	--	--	--	--	--

				<p>é um gráfico de dispersão, intitulado "Salário médio bruto anual em comparação com o nível de satisfação" e com a descrição "Em 2021, Portugal tinha o nível mais baixo de satisfação profissional e era o 10.º país com o salário médio bruto anual mais baixo". Este gráfico cruza duas variáveis distintas, que estão identificadas no eixo do X e do Y: a percentagem de trabalhadores muito satisfeitos (eixo Y) e o salário médio bruto anual (eixo X). Cada um dos países está representado por um círculo preenchido por uma cor. Consoante a sua posição</p>			
--	--	--	--	--	--	--	--

				<p>neste gráfico, existe uma graduação na cor de cada país, que é o resultado da relação entre as duas variáveis (p.e.: Portugal está a vermelho escuro, por ser o país com menos trabalhadores muito satisfeitos e dos que paga pior, enquanto Malta está a verde-escuro por ser o país com a maior percentagem de trabalhadores satisfeitos, embora não seja dos que paga melhor). Para ajudar a perceber essa relação, existe uma nota apontada com uma seta, identificando Portugal, explicando que "Portugal é o país com a menor percentagem</p>			
--	--	--	--	--	--	--	--

				<p>de alta satisfação, embora não seja o que paga pior". O leitor, ao passar o cursor por cada um destes pontos coloridos, é-lhe exibido um "pop-up" com as seguintes informações: "nome do país", "percentagem de trabalhadores muito satisfeitos:" e "Salário médio bruto anual:". Há ainda uma nota ("Nota: A amostra conta com 26 países da UE. Os Países Baixos não foram considerados por utilizarem uma metodologia diferente) e a identificação da mesma fonte, autoria e ferramenta utilizado para o gráfico.</p> <p>Surge um</p>			
--	--	--	--	--	--	--	--

				<p>outro gráfico de barras horizontal, com o título "Dez países com maior percentagem de contratos de duração limitada" e a descrição "Mais de 15% dos trabalhadores portugueses tinham contratos com duração limitada em 2021". Este gráfico é composto por 10 barras, correspondentes ao valor percentual de 10 países, com a indicação do valor da média da UE num traço descontínuo na vertical, apresentando o valor no topo do gráfico ("Média da UE: 14,2%"). Nove das barras são azuis-claras, sendo que Portugal volta a surgir destacado com a cor laranja. Ao</p>			
--	--	--	--	--	--	--	--

				<p>passar o cursor em cada uma delas, surge um "pop-up" com a seguinte informação: Percentagem de contratos com duração limitada, país e valor percentual. São identificados a fonte, autor e ferramenta da criação do gráfico.</p> <p>O último trata-se de um gráfico de barras horizontais agrupado, que identifica "Cinco países da UE com mais microempresas em 2020" e tem a seguinte descrição: "Portugal é o quinto Estado-membro com mais microempresários. A presença de grandes empresas no país é residual (menos de</p>			
--	--	--	--	---	--	--	--

				<p>0,2%). O gráfico contém cinco barras, correspondentes a cinco países distintos, com tons alaranjados, e está ordenada de forma decrescente. Ao passar o cursor por cada uma das "secções" que compõem a barra, o leitor pode consultar os valores de cada país em quatro grupos de informação: Micro empresas (0 a 9 pessoas empregadas) , Pequenas empresas (10 a 49 pessoas empregadas) , Médias empresas (50 a 249 pessoas empregadas) e Grandes empresas (250 pessoas empregadas ou mais).</p>			
Peça 8					Tal como na Peça 4 e Peça	Presença de cinco vídeos com som.	

					<p>5, este artigo abre com um áudio, revelando o diálogo telefónico entre o jornalista e a pessoa entrevistada. A conversa está transcrita na íntegra.</p>	<p>O primeiro deles, com cerca de 30 segundos, surge após a conversa telefónica e é sobreposto por um excerto textual que acompanha o <i>scroll</i> do leitor ("Em Março de 2023, o preço do cabaz alimentar atingiu o valor mais elevado do último ano: 230,38 euros"), seguindo-se o título da peça ("Nuno tem uma vinha: "Os preços aumentam, mas não recebemos mais" ") e subtítulo ("Nuno, Ricardo e Epiménio, pequenos produtores do concelho de Palmela, garantem que a inflação que se faz sentir no carrinho de compras não se traduz em</p>	
--	--	--	--	--	--	---	--

					<p>receita para os agricultores" ). O vídeo tem o som ambiente do local, onde é possível ver uma pequena quinta vinha, um jovem a trabalhar, uma casa de parede branca e uma pessoa mais velha.</p> <p>O segundo vídeo tem a duração de três minutos e 20 segundos, sendo o mais extenso. Esta componente multimédia surge após a introdução textual da reportagem. O vídeo possui legendas do testemunho da pessoa entrevistada que dá a conhecer a situação da vinha familiar que é mantida por "gosto" e "tradição", contando com excertos da entrevista e imagens do</p>	
--	--	--	--	--	--	--

					<p>trabalho desenvolvido na vinha. À semelhança dos vídeos com som identificados na Peça 4 e Peça 5, o leitor pode ativar ou desativar o som, e pode também ativar a opção "mostrar todos os controlos", com o clique direito do rato, sendo o único vídeo identificado nesta peça que permite essa função. Nos segundos finais, surge uma mensagem indicativa, com uma seta, que diz "continuar a ler".</p> <p>O terceiro vídeo, com aproximadamente 30 segundos, mostra o entrevistado a caminhar pela vinha e ouve-se o som ambiente das suas</p>	
--	--	--	--	--	--	--

					<p>passadas, o vento e o chilrear de pássaros. O vídeo é sobreposto com dois excertos textuais destacados com um sublinhado vermelho ("Apenas 13% dos agricultores em Portugal trabalham a tempo inteiro na própria exploração" e "Aproximadamente dois terços ocupam menos de 50% do horário laboral em culturas agrícolas").</p> <p>O quarto vídeo, com cerca de 15 segundos, mostra uma casa com paredes brancas e, ao fundo, vê-se uma queimada. É audível o ladrar de canídeos e o chilrear de pássaros. O vídeo é</p>	
--	--	--	--	--	---	--

						<p>sobreposto com um excerto: "As frutas e legumes foram a categoria de bens alimentares com o maior aumento percentual de preço". Logo de seguida, surge o último vídeo, com um homem de chapéu e óculos de sol, a mirar a paisagem à sua volta, ouvindo-se pássaros e o cacarejar de uma ave. O vídeo tem cerca de 12 segundos e é sobreposto por duas frases: "Em Junho de 2022, Epiménio vendeu a batata a 25 cêntimos por quilo a uma distribuidora" e "No mesmo mês, os consumidores portugueses pagavam 1,03€ euros pelo quilo da</p>	
--	--	--	--	--	--	--	--

						batata nas grandes superfícies".	
Peça 9			Após ser efetuado um zoom no mapa dinâmico identifica do no ponto seguinte, observa-se a presença de um gráfico de barra única agrupado e estático, que contempla informações personalizadas para cada localidade e geográfica.	Identificado um mapa dinâmico de Portugal Continental e das regiões autónomas, subdividido por freguesias, com diferentes tonalidades de cor (desde o rosa-claro até à cor preta). O título do mapa pode ser considerado o nome da peça "Quantas casas vagas há na minha freguesia? Veja o mapa" e conta com subtítulo ("Com uma distribuição desigual no país, os alojamentos vagos que podem em teoria vir a ser alvo de arrendamento forçado surgem nas áreas do país em perda populacional, mas também em			

				<p>núcleos urbanos").</p> <p>No cabeçalho do mapa há uma breve nota instrutória ("Clique no mapa para fazer zoom e ver quantos alojamentos vagos há em determinada freguesia."), assim como uma legenda sobre a escala de graduação da cor: "Alojamentos vagos, sem ser para venda ou arrendamento", que vai dos 200 aos mais de 2400, e a fonte é mencionada no rodapé da representação visual ("Fonte: Censos 2021"). Ao clicarmos num ponto aleatório do mapa, é feito um zoom automático e surge uma pequena barra no cabeçalho do mapa,</p>			
--	--	--	--	--	--	--	--

				<p>indicando o município a que pertence a freguesia, o nome da respetiva freguesia, o número de alojamentos vagos que que não estavam nem para arrendar ou vender em 2021 naquela localidade, e um gráfico de barra única ao lado. Este gráfico apresentado é precedido de um texto ("Nesta freguesia, 11% dos alojamentos familiares estavam, em 2021, vagos, sendo que 4% estavam à venda ou para arrendar e 6% estavam vazios. Os restantes 89% eram residências habituais ou secundárias.") e para cada um dos valores há uma cor estabelecida, que</p>			
--	--	--	--	---	--	--	--

				<p>corresponde à representação dos dados no gráfico. (P.e.: (...)"4% estavam à venda ou para arrendar [cor laranja]", "e 6% estavam vazios [rosa-choque].", "Os restantes 89% eram residências habituais ou secundárias [azul claro]". Com este zoom, surge, no canto inferior direito do mapa, um botão ("ver país completo") que permite regressar ao estado original do mapa.</p>			
Peça 10					<p>O áudio desta peça é apresentado no mesmo estilo identificado nas Peças 4, 5 e 8 desta amostra.</p>	<p>Foram identificados dois vídeos com o som. O primeiro, que surge logo após a transcrição textual da conversa telefónica entre a jornalista e a entrevistada, mostra algumas imagens do</p>	<p>Observou-se a presença de dois vídeos sem som que surgem após o vídeo com som de maior duração. O primeiro vídeo, com</p>

					<p>interior da habitação e pode-se ouvir uma música brasileira, sob a qual a entrevistada procura explicar o significado da letra. No final, é possível ver uma janela, onde passam algumas pessoas e a entrevistada fecha o estore interior. Este vídeo tem cerca de 45 segundos, é sobreposto com um excerto textual destacado ("O valor mediano das rendas em Portugal atingiu um novo máximo no final de 2022"), título da peça ("Aos 60 anos, Clara vai dividir casa para "não cair no abismo") e subtítulo ("Um aumento de 150 euros na renda</p>	<p>cerca de 15 segundos , mostra a silhueta de uma pessoa a passar junto a uma janela no interior da sua casa à noite, enquanto no exterior passa uma carrinha com as luzes ligadas. O vídeo é sobreposto com dois excertos ("Em 2022, foram celebrados mais de 90 mil novos contratos de arrendamento habitacional" e "O maior número de que há registo") e o vídeo repete-se até avançarm</p>
--	--	--	--	--	---	---

					<p>desequilíbrio ou o orçamento de Clara, que vive há mais de dez anos no mesmo apartamento. No final de 2022, o valor mediano dos novos contratos de arrendamento foi o mais elevado de sempre").</p> <p>O segundo vídeo com som é o que tem a maior duração, com um minuto e 41 segundos, explicando a situação pessoal da entrevistada face às dificuldades que vivencia com o aumento da renda da sua habitação. Durante o vídeo, a cara da pessoa nunca é revelada, mostrando apenas silhuetas suas ou pormenores da sua casa (como um relógio, o papel de</p>	<p>os para a fase seguinte da narrativa.</p> <p>Após três parágrafos de texto, surge o segundo vídeo sem som, com cerca de seis segundos. É semelhante ao primeiro identificado nesta secção, sendo visível a sombra de uma mulher com óculos enquanto fala com alguém. Existem duas sobreposições textuais ("O valor mediano das rendas atingiu, no final de 2022 o preço mais elevado de</p>
--	--	--	--	--	---	--

					<p>parede do quarto, alguns alimentos). Nos instantes finais deste vídeo, surge uma pequena mensagem com uma seta apontada para abaixo ("continuar a ler"). A semelhança das peças 4, 5 e 8, esta também permite ativar controlos do vídeo, com o botão direito do rato e conta com legendas.</p>	<p>sempre" e "Um aumento de 10,6% em relação ao ano anterior") e o vídeo repete-se em ciclo.</p>
--	--	--	--	--	---	--

Tabela 6 - Quadro de análise a outras potencialidades Web exploradas

Código da peça	1.3 - Outras potencialidades exploradas
Peça 1	Utilização de emojis para representar os tipos de alimentos. Utilização de sublinhados e negritos para destacar determinadas informações e dados sobre o aumento absoluto e percentual do cabaz de produtos essenciais, bem como da categoria de produto que mais aumentou.
Peça 2	
Peça 3	<p>Utilização de sublinhados que destacam: número total de pessoas que responderam ao inquérito, o mês em análise e a percentagem de inquiridos que responderam que a sua situação financeira nos últimos seis meses tinha piorado.</p> <p>Na mesma página assiste-se a uma integração de vários testemunhos (nos quais é identificado o primeiro e último nome, idade, localização, uma frase de destaque da sua resposta ao questionário e, em alguns casos, fotografia) e reportagens, com um pequenos excertos de vídeo sem som.</p>
Peça 4	<p>O fundo desta página é preto, evidenciando-se uma diferença estética com a maioria dos artigos identificados nesta amostra.</p> <p>O áudio que dá início à reportagem conta com uma transcrição na íntegra e, conforme o diálogo se desenrola entre ambos, as palavras, que inicialmente estavam num tom cinza, destacam-se com uma cor branca.</p> <p>O áudio e o vídeo de maior duração, estão legendados e permitem que o leitor escolha entre ativar ou desativar o som. Caso opte por desativar, os vídeos não perdem sentido, demonstrando a ubiquidade desta peça.</p> <p>A peça embutida a meio do texto ("Contas à vida: enfrentar a inflação") surge com um traçado vermelho, que é característico da Peça 3.</p> <p>Após a conversa telefónica inicial, a meio e no final da peça, alguns excertos textuais com informação quantitativa são destacados</p>

	através de um tamanho de letra maior e contêm um sublinhado vermelho, sobrepondo-se aos vídeos, excepto o de maior duração.
Peça 5	<p>À semelhança da peça anterior, o fundo da página é preto, o único áudio identificado possui transcrição em sincronia com a conversa telefónica e permite que o leitor ative ou desative o mesmo.</p> <p>A opção de ativar ou desativar o som só se verificou mais uma vez ao longo da peça, no segundo vídeo de maior duração que conta com legendagem.</p> <p>Os excertos textuais destacados ao longo da peça surgem com um sublinhado avermelhado e com um tamanho de letra maior.</p> <p>A última hiperligação extratextual identificada surge com uma animação, como se de um <i>sticker</i> se tratasse, e o texto "ler mais testemunhos" é sublinhado com um traçado vermelho quando o cursor passa por cima deste excerto.</p>
Peça 6	
Peça 7	A peça começa com uma narrativa visual de <i>scrollytelling</i> , com representações visuais de dados, revelando ao leitor uma perspetiva geral do assunto que, posteriormente, é comparada com um país e dá a conhecer as perspetivas realçadas pelo autor, conduzindo a narrativa até ao início da reportagem propriamente dita. Além disso, esta componente do <i>scroller</i> é acompanhada por descrições textuais, do lado esquerdo dos gráficos, o que permite compreender facilmente a informação.
Peça 8	<p>Esta peça segue o mesmo estilo estético da página Web da Peça 4 e Peça 5: o fundo da página é preto, o áudio possui transcrição em sincronia com a conversa telefónica; há a opção de ativar ou desativar o som; os excertos textuais destacados ao longo da peça, que se sobrepõem aos vídeos mais curtos com som, têm um sublinhado avermelhado e um tamanho de letra maior.</p> <p>As duas hiperligação extratextual identificadas têm uma animação à sua volta,</p>

	<p>a primeira com um fundo branco, funcionando como um lembrete e com o texto "ler mais testemunhos" sublinhado com um traço vermelho, quando o cursor passa por cima, e o "Contas à vida: enfrentar a inflação", com um traçado vermelho a rodear esse destaque e um sublinhado a vermelho em "Leia os testemunhos dos leitores do PÚBLICO".</p>
Peça 9	<p>Com o mapa ampliado, surge um gráfico de barra única revelando novas informações no cabeçalho do mapa, sem interferir na totalidade com a visualização do mesmo.</p>
Peça 10	<p>Tal como foi mencionado, esta peça segue o mesmo estilo estético da Peça 4, 5 e 8: o fundo da página é preto, o áudio possui transcrição em tempo real e sincronizado com a conversa telefónica; há a opção de ativar ou desativar o som; os excertos textuais destacados ao longo da peça, que se sobrepõem ao vídeo mais curtos com som e aos dois sem som, e a presença de um sublinhado avermelhado e um tamanho de letra maior.</p> <p>Uma das hiperligações extratextual identificadas têm uma animação à sua volta ("Contas à vida: enfrentar a inflação"), com um fundo esbranquiçado, traçado vermelho a rodear esse destaque e um sublinhado a vermelho em "Leia os testemunhos dos leitores do PÚBLICO".</p>

Tabela 7 – Quadro de análise aos recursos interativos presentes nas representações visuais de dados

<b>Código da peça</b>	<b>1.4.1 - Botões de navegação</b>	<b>1.4.2 - Destaques/detalhes ao passar o cursor</b>	<b>1.4.3 - Filtragem</b>	<b>1.4.4 - Pesquisa ou pesquisa detalhada</b>	<b>1.4.5 - Zoom</b>
Peça 1		No primeiro gráfico de linha identificado o leitor pode consultar o preço semanal do cabaz de produtos essenciais e a data. Nos restantes seis gráficos de linha, o leitor pode consultar a variação do preço nas seis categorias de produtos e a data correspondente.	No primeiro parágrafo do texto, o leitor pode selecionar entre uma lista de datas (P.e.: "Entre (primeira data opcional) 05 de Janeiro de 2022 e (segunda data opcional) 29 de Março de 2023, o preço do cabaz de produtos essenciais da DECO Proteste ficou mais caro 38,29 euros - ou seja, um aumento na ordem dos 20,40%"), alterando a estrutura textual que dá a conhecer a diferença total e percentual dos preços entre as datas selecionadas.	A partir do primeiro gráfico de linha e dos recursos interativos que o integram, o leitor pode fazer uma pesquisa simples, consultando os preços do cabaz de produtos essenciais, de 5 de janeiro de 2022 até 29 de março de 2023. O mesmo sucede nos restantes seis gráficos de linha, permitindo pesquisar a variação percentual do tipo de produto alimentar a partir de 25 de janeiro de 2023. Relativamente à tabela apresentada, ao nível da pesquisa, esta permite consultar os 10 produtos que sofreram um maior aumento face	

				a 23 de fevereiro de 2022.	
Peça 2	Quando explorado pelo leitor, o gráfico de árvore possibilita verificar em detalhe os valores percentuais de importação de arroz dos diferentes países, clicando nos círculos "maiores", os dos continentes. Posteriormente, do lado superior direito da representação visual, surge uma pequena seta que permite ao leitor retornar ao gráfico original.	No gráfico de linha o leitor pode consultar o preço do arroz carolino e o preço ao passar o cursor na linha laranja. O gráfico de árvore possibilita que o leitor verifique o peso das importações de cada continente e também dos países.	Ainda que de forma simples, o gráfico de árvore permite que o leitor filtre os resultados pelos países dos quatro continentes representados, o que permite perceber o peso das importações de cada país.		O gráfico de árvore comporta uma componente de zoom, possibilitando ao leitor verificar com maior detalhe os países e os valores que compõem o peso de importação de arroz de cada continente.
Peça 3	O final desta página apresenta uma opção que permite ao leitor "ler mais testemunhos" e navegar por entre as várias respostas obtidas ao inquérito.	Ao passar o cursor em "Contribua com o seu caso e ajude-nos a pintar o gráfico", surge um traçado a vermelho que rodeia essa frase e no gráfico circular, ao lado, o último mês em análise também é destacado com este traçado. No gráfico circular, é possível verificar o número de respostas por níveis. Da mesma forma, ao passar o cursor nas frases destacadas dos testemunhos nesta página, o fundo fica com um tom			

		<p>avermelhado e surge uma seta vermelha animada que permite ler o testemunho na íntegra". A seta apresenta diferentes formatos consoante a extensão da frase destacada.</p> <p>Um traçado vermelho também surge em "ler mais testemunhos" no final da página.</p>			
Peça 4					
Peça 5					
Peça 6	<p>O mapa dinâmico permite ao leitor clicar em determinada localização geográfica, o que leva a que um zoom automático seja efetuado nessa zona. Após essa ação, o leitor encontra uma opção de navegação ("ver mapa completo") que lhe permite regressar ao estado original da representação visual dos dados e verificar outras zonas.</p> <p>Com o zoom aplicado, para percorrer as diferentes localidades do mapa (p.e. de Norte a Sul do país ou do litoral ao interior), é necessário clicar nas</p>	<p>Ao passar o cursor pelas diferentes localidades, uma pequena mensagem personalizada surge no canto superior direito do mapa, que identifica o nome da freguesia/município, a percentagem do rendimento dedicado à renda, o valor mensal estimado da renda e o valor que a renda teria de diminuir para alcançar a taxa de esforço de 27% na localidade em que o cursor se encontra.</p>	<p>A parte superior do mapa contempla três categorias, das quais duas permitem que o leitor filtre os resultados por 1) tipologia da casa (apartamento, casa em banda, casa geminada ou casa isolada] e 2) a área da habitação (T0, T1, T2, T3, T4).</p>	<p>Embora o mapa não contemple uma barra de pesquisa pelo nome da localidade, a terceira categoria na parte superior do mapa permite que o leitor digite manualmente o valor do seu rendimento líquido mensal. Essa ação, em conjunto a seleção das categorias anteriormente descritas, leva as zonas do mapa fiquem com diferentes cores (verde escuro: 0 a 10% do rendimento gasto em renda até ao vermelho escuro, que representa mais de 45% do</p>	<p>Ao clicar em qualquer localidade do mapa, mesmo naquelas que não existem dados, essa zona do mapa é automaticamente ampliada, facilitando a consulta por localidades geográficas de menor dimensão.</p>

	localidades abaixo, acima ou ao lado para efetuar essa navegação.			rendimento gasto em renda), permitindo-lhe pesquisar as localidades do seu interesse.	
Peça 7		<p>As oito representações visuais de dados identificadas no ponto 1.2.4 apresentam informações complementares ao passar o cursor por cima, que, por vezes, se revelam redundantes para a compreensão dos dados dos gráficos e mapa. De forma a identificar os gráficos que possuem este nível de interatividade, iremos identificá-los pelo título correspondente: 1) Satisfação laboral na União Europeia em 2021; 2) Variação do nível de satisfação na Europa; 3) Em Portugal, apenas 21,6% dos trabalhadores estavam muito satisfeitos; 4) Portugal é o país com a menor percentagem de trabalhadores muito satisfeitos; 5) E é o segundo país com mais trabalhadores insatisfeitos; 6) Salário médio bruto anual em comparação com o nível de satisfação; 7) Dez países com maior percentagem de contratos de duração limitada; 8)</p>			

		Cinco países da UE com mais microempresas em 2020.			
Peça 8					
Peça 9	<p>Ao efetuar o zoom automático em determinada zona do mapa, surge posteriormente a opção de navegação ("ver mapa completo"), o que permite regressar ao estado original do mapa. O mapa, após ser ampliado, não permite uma navegação livre do leitor (p.e. ver uma freguesia do Norte do país e depois passar para uma freguesia da Região Autónoma dos Açores). Para navegar por outras localidades, o leitor terá de regressar ao mapa original ou clicar em zonas geográficas que o levem até à localidade pretendida.</p>				O mapa dinâmico conta com a opção de zoom, permitindo consultar em detalhe os dados de todas as freguesias de Portugal.
Peça 10					

Nota: Apenas foram consideradas nesta análise as peças que continham, pelo menos, uma representação visual de dados.

Tabela 8 – Quadro de análise ao tipo de interatividade nas representações visuais de dados

Código da peça	1.5.1 - Interatividade de transmissão	1.5.2 - Interatividade consultiva	1.5.3 - Interatividade de registo
Peça 1	<p>O primeiro gráfico de linha permite ao leitor visualizar e consultar, desde 5 de janeiro de 2022 até 29 de março de 2023, os preços semanais do cabaz de produtos essenciais e, através do cursor, surge um "pop-up" que indica o preço e a data correspondente a cada um dos pontos.</p> <p>Os seis gráficos seguintes são semelhantes ao primeiro: permitem que o leitor realize a mesma ação interativa, mas com um espectro temporal limitado: de 23 de fevereiro de 2022 até 29 de março de 2023 e que apenas permite consultar as percentagens de variação a partir de 25 de janeiro de 2023.</p>	<p>Considerando os seis dos sete gráficos de linha, é possível apontar um nível de interatividade consultiva, uma vez que as representações visuais de dados, que se seguem ao primeiro gráfico, permitem ao leitor explorar diferentes perspetivas por cada categoria de produto, discriminando a flutuação percentual dos preços de cada uma.</p>	
Peça 2	<p>As duas representações visuais de dados identificadas são características da interatividade de transmissão, pois são simples de consultar e ambas apresentam informações</p>	<p>O gráfico de árvore apresenta um certo grau de interatividade consultiva, pois permite que o leitor explore em detalhe, através do zoom, uma subcamada de informações que</p>	

	complementares através de “pop-up's” que surgem ao passar o cursor.	revelam outras perspetivas subjacentes à história principal da peça.	
Peça 3	O único gráfico identificado no topo desta página permite ao leitor visualizar o número total de respostas obtidas ao questionário desde novembro de 2022 e, recorrendo ao cursor, é possível verificar o número de respostas mensais e quantos referiram se a sua condição financeira piorou, manteve ou melhorou em comparação com os últimos seis meses. O texto ao lado esquerdo do gráfico apresenta, de forma textual, o número total de respostas obtidas até ao momento.		
Peça 4			
Peça 5			
Peça 6	É possível evidenciar a interatividade de transmissão neste mapa dinâmico, visto que, ao passar o cursor pelas unidades territoriais, surge um “pop-up” com informações personalizadas de cada uma das localidades no canto superior direito.	A interatividade consultiva também é contemplada neste mapa, visto que as seleções das categorias na parte superior permitem ao utilizador consultar diferentes perspetivas da informação reunida (tipo de casa e as dimensões da habitação). Além disso, o zoom do mapa também se constitui como um elemento deste tipo	Relativamente à interatividade de registo, esta também está contemplada no mapa dinâmico, pois permite a seleção de uma lista pré-definida de opções e a introdução de dados próprios que alteram, em tempo real, a representação visual.

		de interatividade.	
Peça 7	Os oito gráficos identificados com destaques/detalhes ao passar o cursor, no ponto 1.4.2., e descritos em 1.2.3, mostram que estas representações visuais dos dados comportam somente uma interatividade de transmissão, surgindo "pop-ups" com informações complementares, não se verificando qualquer nível significativo de interatividade consultiva ou exploratória dos dados, nem de interatividade de registo.		
Peça 8			
Peça 9	A interatividade de transmissão neste mapa dinâmico evidencia-se ao clicarmos na localidade pretendida, surgindo no cabeçalho do mapa um gráfico de barra e novos dados personalizados para cada uma das freguesias.	A interatividade consultiva é apenas evidenciada pelo facto de o mapa ter a potencialidade de ser ampliado, sendo um elemento que se integra neste tipo de interatividade.	
Peça 10			

Tabela 9 – Quadro de análise aos esquemas narrativos - orientações narrativas da história

Código da peça	1.6.1.1 - Estrutura Martini Glass	1.6.1. 2 - Slideshow interativo	1.6.1.3 - História de pesquisa
Peça 1	<p>A estrutura narrativa deste artigo segue uma abordagem orientada pelo autor. Embora a introdução textual surja apenas após a primeira representação visual dos dados, o primeiro parágrafo, que permite ao leitor escolher as datas que quer estabelecer a comparação, evidencia uma narrativa guiada pelo autor da peça, que visa dar a conhecer o aumento dos preços, os produtos considerados no cabaz e a forma como se obtém o "valor médio" . De seguida, o leitor pode explorar os dados de forma interativa, consultando as taxas de variação nas diferentes categorias ou os preços e datas do cabaz alimentar.</p>		
Peça 2	<p>A narrativa desta história segue a estrutura do “Martini Glass”, visto que se inicia com uma abordagem orientada pelo autor. Desde logo o título da peça ("Porque é que o arroz carolino já custa quase o</p>		

	<p>dobro?") introduz uma pergunta, o lead noticioso faz observações ("Se está entre os portugueses que não dispensam o arroz carolino no prato, certamente já reparou no disparar do preço deste alimento. Há um ano, o preço de uma embalagem de arroz carolino rondava pouco mais de um euro, agora, o mesmo produto custa o dobro. (...)" e é seguido por um gráfico de linha que ilustra o aumento deste produto alimentar. Esta primeira representação visual de dados permite a exploração interativa da informação, o que condiz com a fase orientada pelo leitor, verificando-se o mesmo no gráfico perto do final do artigo.</p>		
Peça 3			<p>Esta peça segue vincadamente uma orientação orientada para o leitor, pois ao apresentar o tema geral, o leitor pode escolher e navegar as histórias secundárias que pretende consultar através das hiperligações que dão acesso aos testemunhos na íntegra e reportagens</p>

			publicadas.
Peça 4			
Peça 5			
Peça 6	<p>Esta peça segue uma estrutura Martini Glass, visto que parte da apresentação de uma pergunta formulada pelo autor ("Onde é que consigo arrendar uma casa com o meu salário?"), assim como apresentando um pequeno contexto ("A crise da habitação não afecta todo o país da mesma forma. Arrendar casa custa mais em algumas zonas, especialmente nas grandes cidades. Veja em que freguesias ou municípios consegue viver com os seus rendimentos."). De seguida, é apresentado o mapa dinâmico que possibilita a exploração das categorias disponíveis através das interações conduzidas e orientadas pelo leitor. Após essa exploração, os autores voltam a assumir a condução da narrativa do leitor com um texto jornalístico e explicativo sobre o desfasamento dos valores apresentados no mapa do <i>Público</i></p>		<p><b>Não foi considerado</b> que esta peça apresente uma estrutura narrativa de história de pesquisa. Embora o mapa dinâmico possibilite a consulte de detalhes e seja orientado para a utilização do leitor, nenhuma "história secundária" é revelada.</p>

	em relação à realidade do mercado ou nos sites de compra e arrendamento de casas.		
Peça 7	A narrativa desta peça segue uma estrutura de Martini Glass, visto que parte de uma abordagem orientado pelo autor e o próprio título desta peça é uma pergunta ("Portugal tem dos trabalhadores mais insatisfeitos da Europa. Porquê?"). Embora as representações visuais de dados não permitam ao leitor assumir a "condução da história", o narrador (autor da peça) alarga o espectro da mesma através da apresentação de um conjunto de fatores e testemunhos que visam responder à pergunta enunciada.		
Peça 8			
Peça 9	A narrativa desta peça segue nitidamente a estrutura de Martini Glass, visto que o próprio título apresenta uma pergunta e o subtítulo faz uma breve contextualização guiada pelo autor. Logo de seguida, é apresentado o mapa dinâmico, que		

	<p>permite ao leitor interagir com os diferentes pontos do mapa e consultar informações complementares sobre cada freguesia. De seguida, o autor volta a assumir a condução da narrativa com um texto explicativo e de análise dos dados utilizados.</p>		
Peça 10			

Nota: Apenas foram consideradas nesta análise as peças que continham, pelo menos, uma representação visual de dados.

Tabela 10 – Quadro de análise aos esquemas narrativos - estética visual da narrativa

Código da peça	1.6.2.1 - Scrollytelling	1.6.2.2 - News app
Peça 1		
Peça 2		
Peça 3		Esta página interativa pode ser considerada como uma "news app", pois conta as histórias dos testemunhos recolhidos a partir do inquérito de uma forma personalizada (testemunho na íntegra, com ou sem fotografia, reportagem).
Peça 4	Esta peça jornalística segue uma estrutura de <i>scrollytelling</i> , exigindo que o leitor faça <i>scroll</i> dentro da página para que se revele a peça num todo. Contudo, não existe qualquer tipo de representação visual de dados. O leitor para ler a peça tem de percorrer a conversa telefónica, um excerto textual, que vai subindo consoante o <i>scroll</i> , e um vídeo sem som, revelando finalmente o título e o texto da peça. O mesmo volta a suceder quando o leitor encontra o vídeo com som e os restantes dois vídeos sem som sobrepostos por destaques textuais.	
Peça 5	A estrutura de <i>scollytelling</i> identificada e descrita na Peça 4 segue o mesmo modelo que esta.	
Peça 6		Consideramos esta peça como uma "news app", pois a funcionalidade interativa integrada (o mapa dinâmico) possibilita que o leitor personalize as categorias que quer consultar, introduza os seus próprios dados e, no fim, é-lhe exibida uma representação visual de dados personalizada.

Peça 7	A parte inicial desta peça segue a narrativa visual de <i>scrollytelling</i> , revelando ao leitor, através do <i>scroll</i> , um conjunto de dados, gráficos e informações textuais que o leva até à parte textual do trabalho jornalístico.	
Peça 8	A estrutura da narrativa visual de <i>scrollytelling</i> identificada e descrita na Peça 4 e Peça 5 segue o mesmo modelo que nesta peça.	
Peça 9		
Peça 10	A estrutura da narrativa visual em <i>scrollytelling</i> identificada e descrita na Peça 4, 5 e 8 segue o mesmo modelo que nesta peça.	

Nota: As peças da amostra que não contém qualquer informação seguem uma estrutura narrativa visual normal às peças publicadas na página Web do *Público*.

Tabela 11 – Quadro de análise para avaliar a transparência, precisão e imparcialidade dos dados

Código da peça	Descrição da metodologia/referência a dados	Fonte dos dados	Base de dados utilizada na análise
<b>Peça 1</b>	<p>" (...) O cabaz de produtos da Deco Proteste é definido com base na estrutura de consumo das famílias do INE.</p> <p>Este conjunto de 63 produtos inclui, por exemplo, leite, queijo, manteiga e fiambre, um pacote de arroz, de farinha, de massa e de açúcar, além de 1 kg de carne de peru e a mesma quantidade de frango, carapau, pescada, cebola, batata, cenoura, banana, maçã ou laranja.</p> <p>É calculado o valor médio por produto em todas as lojas de Portugal onde está disponível, obtendo assim o valor médio daquele produto em dado dia. Para cada dia analisado, é calculado o valor total do cabaz (com base no valor médio por produto)."</p>	<p>Identificação no primeiro gráfico: "Fonte: DECO Proteste";</p> <p>Identificação no final do artigo: "Em conjunto com: DECO Proteste".</p>	
<b>Peça 2</b>	<p>Descrição do primeiro gráfico: "O preço médio atingiu um novo máximo a 18 de Janeiro de 2023. Com base em 11 marcas, este produto está a custar, em média, mais de dois euros nos supermercados.";</p> <p>Título e descrição do segundo gráfico: "Até novembro de 2022, mais de 50% do arroz importado veio da América do Sul" [título];</p> <p>"Outros tipos de arroz vendidos em Portugal não registaram um aumento tão expressivo. A maioria (quase 70%) é proveniente de países "imunes" aos efeitos da guerra" [descrição].</p>	<p>Identificação no primeiro gráfico: "Fonte: DECO Proteste";</p> <p>Identificação no segundo gráfico: "Fonte: Eurostat".</p>	
<b>Peça 3</b>	<p>"Em Março , das 98 pessoas que deram o seu testemunho ao PÚBLICO, 91% revelavam que, em comparação com os últimos seis meses, a sua situação financeira tinha piorado."</p>	<p>Base de dados própria: hiperligação intratextual em "Contribua com o seu caso e ajude-nos a pintar o gráfico" remete o leitor para o questionário.</p>	

<p><b>Peça 4</b></p>	<p>[1.º bloco textual do scroller]: "Cerca de 238 mil trabalhadores, em Portugal, acumulam dois ou mais empregos";</p> <p>Texto da reportagem: "(...) Miguel é um dos 238 mil trabalhadores que, em Portugal, acumulam dois ou mais empregos, segundo dados do Instituto Nacional de Estatística (INE) relativos ao terceiro trimestre de 2022. Há um ano eram menos 30 mil, agora representam quase 5% do total da população empregada. (...)";</p> <p>[2.º bloco textual do scroller]: "Em 2022, um assistente técnico da função pública recebia 757 euros brutos";</p> <p>[3.º bloco textual do scroller]: "Desde Janeiro de 2023, Miguel passou a receber 861 euros";</p> <p>[4.º bloco textual]: "Miguel paga 84 euros de mensalidade por uma aula de piano por semana";</p> <p>[5.º bloco textual]: "Com as entregas como estafeta Miguel recebe, em média, 450 euros extra por mês";</p>	<p>[No texto da notícia: Instituto Nacional de Estatística]: "Miguel é um dos 238 mil trabalhadores que, em Portugal, acumulam dois ou mais empregos, segundo dados do Instituto Nacional de Estatística (INE) relativos ao terceiro trimestre de 2022. Há um ano eram menos 30 mil, agora representam quase 5% do total da população empregada"). Base de dados própria do questionário da Peça 3 (terceiro, quarto e quinto blocos de texto do scroller).</p>	
----------------------	---	---	--

<p><b>Peça 5</b></p>	<p>[1.º bloco textual do scroller]: "Até Janeiro de 2023, mais de 105 mil estudantes tinham pedido bolsa de estudo no ensino superior";</p> <p>Subtítulo da reportagem: "Com um rendimento de 150 euros mensais, um aumento de 15 euros na renda foi suficiente para tornar a estadia em Lisboa incomportável. Durante seis meses, Ana teve de recorrer às poupanças para continuar a estudar na capital.";</p> <p>Texto da reportagem: "(...) Senta-se num muro já no exterior do edifício. É daqui que conta como por causa de um aumento de 15 euros na renda do quarto onde vivia quase teve de abandonar Lisboa e o curso de Direito. “Apesar de não parecer um aumento significativo, para mim não dava”, justifica. Ana é um dos 70.324 estudantes que recebem bolsa de estudo no ensino superior. A pensão de sobrevivência do pai, que morreu em 2013, é o único rendimento mensal que tem garantido. São 150 euros dos quais apenas 50 sobriam depois de pagas propinas e passe. Tem a certeza de que sem este apoio não conseguiria ter iniciado a licenciatura.</p> <p>“É mesmo uma grande ajuda”, garante. “Reduz o preço dos transportes – em vez de pagar os 30 euros do passe, pago 16 euros – e dá-me a possibilidade de ter o complemento de alojamento da Universidade de Lisboa, que são 285 euros, se estiver num quarto, ou a totalidade do custo, se estiver numa residência universitária”, explica.(...) Por isso mesmo, concorreu ao complemento de alojamento. Sabia que não era prioritária por morar relativamente perto da capital, mas, apesar de não lhe ter sido atribuída uma vaga numa residência, recebeu os 285 euros para apoio no arrendamento de um quarto na cidade. (...) “Por acaso, tive sorte, porque na altura encontrei duas casas. Uma delas era 350 euros, tudo com despesas incluídas, e a outra era 375 euros. Já ficava apertado para o meu orçamento, mas, tendo em conta os benefícios, acabei por escolher essa última. (...) O problema surgiu em Janeiro de 2022, quando a senhoria a notificou de um aumento de 15 euros na renda.”</p> <p>[2.º bloco textual do scroller]: "Nunca tantos estudantes tinham pedido bolsa de estudo no ensino superior";</p> <p>[3.º bloco textual do scroller]: "Só alunos de famílias com rendimento per capita inferior a 8.962 euros podem receber bolsas de ação social";</p> <p>Texto da reportagem: "(...) Ana começou o ano</p>	<p>Base de dados própria decorrente do questionário da Peça 3 e dados a partir da entrevista realizada.</p> <p>Alguns dados carecem da identificação explícita da fonte, tais como: "(...) Ana é um dos 70.324 estudantes que recebem bolsa de estudo no ensino superior (...)" ; assim como as informações do segundo e terceiro blocos textuais do scroller.</p>	
----------------------	--	--	--

	<p>lectivo a receber um complemento de 1600 euros anuais. “Entretanto, a minha bolsa diminuiu, por causa dos rendimentos de um trabalho que tive no Verão. Passou para o valor mínimo: 872 euros. E pagar uma renda de casa que aumentou, com uma bolsa diminuída, foi muito difícil”, desabafa. Para que um aluno tenha acesso ao apoio de uma bolsa da ação social é uma condição obrigatória que o agregado familiar apresente um rendimento per capita inferior a 8962 euros. (...) Em Setembro de 2022 completou a candidatura ao complemento de alojamento e foi contactada pelos serviços de acção social da Universidade de Lisboa para uma entrevista, habitual nestes casos. (...)”</p> <p>[4.º bloco textual do scroller]: "Ana pagava 390 euros por um quarto em Lisboa";</p> <p>[5.º bloco textual do scroller]: "Para bolseiros o custo da residência pode ser reembolsado na totalidade";</p> <p>Texto da reportagem: "(...) Paga 77 euros pela residência, mas o valor é-lhe devolvido todos os meses, juntamente com a bolsa da Direcção-Geral de Ensino Superior, que, entretanto, aumentou para 1100 euros. Feita a divisão por dez meses, a mensalidade fixa-se nos 110 euros. (...)”.</p>		
--	--	--	--

<p><b>Peça 6</b></p>	<p>[Nota em rodapé do mapa]"Nota: os dados deste mapa reflectem a renda de mercado que, segundo o Programa de Apoio ao Arrendamento, teria uma habitação com as características seleccionadas no primeiro semestre de 2022. São apresentados dados para cada município ou freguesia para o qual o INE disponibiliza dados. Este valor foi calculado considerando a mediana do preço por m<sup>2</sup> dos novos contratos de arrendamento, a área bruta privativa e o “coeficiente de qualidade e conforto”, seguindo as regras definidas no Programa de Arrendamento Acessível. Em termos de áreas, foram consideradas as seguintes: T0 - 35 m<sup>2</sup>; T1 - 52 m<sup>2</sup>; T2 - 72 m<sup>2</sup>; T3 - 91 m<sup>2</sup>; T4 - 105 m<sup>2</sup>. Já no chamado “coeficiente de qualidade e conforto”, foi considerada, em todos os casos, uma habitação não mobilada, com cozinha equipada, lugar de estacionamento, estado de conservação considerado “bom” e classe energética ‘C’. Os resultados apresentados servem apenas como referência, uma vez que os valores medianos não reflectem os múltiplos factores que influenciam a renda numa determinada região."</p> <p>[Texto da reportagem]"Para quem está habituado a navegar pelos sites de compra e arrendamento de casas, os valores apresentados no mapa do PÚBLICO podem parecer baixos em relação à realidade do mercado – excepção feita a algumas regiões do país. Uma primeira razão para este desfasamento é muito directa: as páginas de compra e arrendamento de imóveis apresentam o valor que o proprietário pretende obter e não o que fica firmado nos contratos de arrendamento, tendencialmente mais baixos.</p> <p>Por sua vez, os dados do mapa reflectem o preço dos contratos assinados e que são utilizados na definição das rendas do Programa de Apoio ao Arrendamento (PAA). Calculado o valor de referência, tendo em conta as características e localização do imóvel, o Governo define que a renda acessível tem de ser 20% inferior. Da parte dos inquilinos, o valor a pagar não pode ser superior a 35% do rendimento médio mensal bruto.</p> <p>O investigador Nuno Travasso, da Faculdade de Arquitectura da Universidade do Porto (FAUP), considera que existe um “problema do mecanismo” que estabelece estes valores de referência de arrendamento.</p> <p>Em algumas zonas, como Porto e Lisboa, os dados</p>	<p>[Nota do rodapé]: "(...) São apresentados dados para cada município ou freguesia para o qual o INE disponibiliza dados (...)"</p>	
----------------------	--	--	--

	<p>são suficientes para o INE fazer a análise por freguesias, permitindo desenhar um retrato mais ou menos em linha com a realidade. No entanto, no resto do território, como acontece, por exemplo, em Braga, a mediana é retirada por município. Juntando o número insuficiente de contratos e a disponibilização de apenas um valor mediano para estas regiões, o método utilizado pode apresentar falhas ou distorções. Grande parte do interior do país não tem dados, nem sequer à escala de município.</p> <p>Esses dados, segundo Nuno Travasso, “dão pouca informação sobre o que se passa numa zona, e podem passar a ideia de que o problema não é tão grave quanto é”. Mas o investigador vai além do mapa e sublinha que a situação é “mais preocupante”, por estes registos suportarem “um programa que dá subsídios”, correndo-se o risco de uma ferramenta que visa rendas acessíveis contribuir para inflacionar mais o mercado em várias zonas.</p> <p>O especialista dá o exemplo da mediana do preço por metro quadrado de um município com um centro em que há rendas muito altas, mas com freguesias limítrofes com valores muito mais baixos: o valor utilizado para o cálculo vai subvalorizar a resposta no centro e pode ficar acima do valor de mercado para as zonas circundantes, podendo levar proprietários a subirem aí as rendas em função da escala do PAA que avalia todo o concelho."</p>		
<p><b>Peça 7</b></p>	<p>[1.º bloco textual do <i>scroller</i>] Em 2021, mais de 197 milhões de trabalhadores na União Europeia foram entrevistados sobre o nível de satisfação laboral. Os dados reunidos pelo Eurostat mostram que 43,8% das pessoas estavam "muito satisfeitas com a condição profissional.</p> <p>[2.º bloco textual do <i>scroller</i>] "O indicador da alta satisfação assume diferentes pesos em cada um dos Estados-membros da UE, assim como na Noruega e na Suíça. Malta, Letónia, e Suíça são os países onde se registam os valores mais elevados: mais de 65% dos entrevistados estavam "altamente satisfeitos" com o seu emprego.", ao qual acresce uma nota ("Nota: A base de dados reúne as respostas de 27 países da União Europeia, aos quais acrescem a Noruega e a Suíça.").</p> <p>[3.º bloco textual] Em contraste com a média da comunidade europeia, Portugal tem menos de um quarto dos trabalhadores felizes com a sua</p>	<p>A instituição Eurostat é mencionada em quase todas as representações visuais de dados e na hiperligação intratextual "(...) Por outro lado, os dados reunidos com o Eurostat (...)".</p> <p>A única exceção é a do gráfico "Ganho médio mensal entre mulheres e</p>	<p>Presença de uma hiperligação intratextual que remete o leitor para a página do Eurostat, onde se encontra a base de dados utilizada.</p>

	<p>profissão (21,6%). A maioria respondeu estar "mais ou menos satisfeita", enquanto 10,6% das pessoas expressaram a sua insatisfação.</p> <p>[4.º bloco textual ] Esse valor mostra que Portugal é o país com a menor percentagem de trabalhadores muito satisfeitos quando comparado com os outros países. A vizinha Espanha, por exemplo, ultrapassa a média da UE: mais de metade dos entrevistados (54%) disseram estar altamente satisfeitos.</p> <p>[5.º bloco] No indicador da insatisfação, o país volta a surgir nas piores posições. É o segundo da lista com a maior percentagem de trabalhadores insatisfeitos, tendo quase o dobro da média europeia (5,8%). É apenas superado pela Bulgária (11,8%). Por outro lado, a Itália é o país com menor percentagem".</p> <p>No texto da peça: "Em 2021, mais de quatro milhões de pessoas revelaram que trabalhar em Portugal está longe de ser sinónimo de felicidade. Baixos salários, vínculos precários, carga horária excessiva, gestões autoritárias e estagnação profissional estão entre as principais justificações num mercado de trabalho onde poucos se sentem felizes. No ano em que os institutos nacionais de estatística de cada país realizaram o inquérito da satisfação em contexto profissional, o mundo laboral confrontava-se ainda com o "choque" da pandemia de covid-19. Mas, para Elísio Estanque, a insatisfação na região portuguesa "já era crescente desde a crise e o período da troika".</p> <p>Nota no gráfico de dispersão: ("Nota: A amostra conta com 26 países da UE. Os Países Baixos não foram considerados por utilizarem uma metodologia diferente");</p> <p>Nota no gráfico de barras vertical: ("Nota: dados referentes apenas a Portugal Continental.);</p>	<p>homens", em que a fonte se encontra no rodapé da representação visual de dados: "GEP - Ministério do Trabalho, Solidariedade e Segurança Social (2021)".</p>	
--	--	---	--

<p><b>Peça 8</b></p>	<p>"[1.º bloco do scroller] Em Março de 2023, o preço do cabaz alimentar atingiu o valor mais elevado do último ano: 230,38 euros";</p> <p>No texto da reportagem: "(...) Encostado a uma velha oliveira à entrada da Quinta da Saudade, Nuno explica que quando enviou o testemunho ao 'Contas à Vida', quis 'partilhar a realidade dos pequenos produtores', que diz serem uma das 'vítimas da inflação' e 'de um aproveitamento por parte das grandes superfícies' (...) ";</p> <p>[2.º bloco textual do scroller]: "Apenas 13% dos agricultores em Portugal trabalham a tempo inteiro na própria exploração";</p> <p>[3.º bloco textual do scroller]: "Aproximadamente dois terços ocupam menos de 50% do horário laboral em culturas agrícolas";</p> <p>No texto da reportagem: "De acordo com a avaliação da Deco Proteste, organização portuguesa de defesa do consumidor, o preço do cabaz de produtos essenciais aumentou 24% no último ano. A 2 de Março de 2022 o custo total dos 63 produtos que compõem este cabaz, nos quais se incluem tomate, cebola, batata, leite, frango ou pescada, fixava-se nos 185,17 euros. Um ano depois, os portugueses pagam mais 45 euros pelo mesmo conjunto de bens: 230,38 euros. A batata-vermelha está entre os dez produtos com maior crescimento percentual do preço, 61% face ao mesmo mês de 2021. A polpa de tomate ocupa o terceiro lugar do pódio. Em doze meses passou de 88 cêntimos para 1,65 euros.";</p> <p>[4.º bloco textual do scroller]: "As frutas e legumes foram a categoria de bens alimentares com o maior aumento percentual de preço";</p> <p>[5.º bloco textual]: "Em Junho de 2022, Epiménio vendeu a batata a 25 cêntimos por quilo a uma distribuidora";</p> <p>[6.º bloco textual]: "No mesmo mês, os consumidores portugueses pagavam 1,03€ euros pelo quilo da batata nas grandes superfícies".</p>	<p>Base de dados própria a partir do questionário e alguns dados recolhidos através da entrevista.</p> <p>Hiperligação intratextual mencionando DECO Proteste ("De acordo com a avaliação da Deco Proteste, organização portuguesa de defesa do consumidor, o preço do cabaz de produtos essenciais aumentou 24% no último ano. A 2 de Março de 2022 o custo total dos 63 produtos que compõem este cabaz, nos quais se incluem tomate, cebola, batata, leite, frango ou pescada, fixava-se nos 185,17 euros. Um ano depois, os portugueses pagam mais 45 euros pelo mesmo conjunto de bens: 230,38 euros.").</p> <p>Constata-se a presença de dados que não foi identificada a sua fonte,</p>	
----------------------	--	--	--

		nomeadamente no segundo e terceiro blocos textuais do <i>scroller</i> .	
--	--	---	--

<p><b>Peça 9</b></p>	<p>Texto da peça: "No dia 19 de Abril de 2021, o Instituto Nacional de Estatística contabilizou como vagos 375.118 fogos habitacionais que estavam desocupados por motivo de falecimento do anterior morador, porque aguardavam partilha de herdeiros, obras de beneficiação, ou ainda outras situações não especificadas. Trata-se dos imóveis que, não sendo residência habitual, também não estão classificados como residência secundária ou uso sazonal, nem estão colocados no mercado para venda ou arrendamento. São, assim, os imóveis que podem vir a ser abrangidos pelo "regime de arrendamento obrigatório de casas devolutas", enquadrado no pacote legislativo Mais Habitação, que o Governo tem ainda em discussão pública até dia 24 de Março.</p> <p>Graças ao esforço censitário feito pelo INE há quase dois anos, podemos perceber onde é que estão as "casas sem gente" a que governantes, políticos e opinião pública tantas vezes fazem alusão, apontando-as também como uma solução necessária para aumentar o número de fogos no mercado da habitação. Nos Censos 2021, o INE encontrou 348.097 casas vagas, mas já colocadas no mercado de arrendamento – pelo que estas não podem ser inseridas na medida agora proposta pelo Governo. É a soma destas duas parcelas – os fogos vagos para venda ou arrendamento e os vagos por outros motivos – que permite chegar ao número de 723.215 imóveis não ocupados como residência habitual. Se inserirmos os dados apurados pelo INE no mapa do território nacional, percebemos que o fenómeno é expressivo, se distribui de modo muito desigual e que pode ter causas muito variadas. Na proposta de lei que está em consulta pública, o Governo propõe-se avançar com o arrendamento forçado nas casas desocupadas que não sejam "casas de férias", "casas de emigrantes ou de pessoas deslocadas por razões de saúde e razões profissionais ou formativas" ou ainda "casas cujos proprietários estão num equipamento social como um lar ou estão a prestar cuidados permanentes como cuidadores informais". Na impossibilidade de saber quais os imóveis que cumprem ou não estes requisitos, resta perceber, onde se localizam estas propriedades que, em teoria, poderão vir a ser abrangidos por esta medida.</p> <p>Os dados apurados revelam que uma grande parte dos fogos vagos surge nas áreas do país onde há</p>	<p>Fonte identificada no rodapé da representação visual de dados ("Fonte: Censos 2021").</p>	
----------------------	---	--	--

	<p>perda populacional – um sinal do despovoamento e da falta de procura destes territórios. Mas também há uma grande concentração de casos nas áreas metropolitanas e em zonas de pressão urbanística.</p> <p>Na verdade, fazendo uma análise por freguesias, como aquela que o PÚBLICO efectuou, percebe-se que o problema parece estar concentrado num número restrito. Das 3092 freguesias portuguesas, só 60 têm mais de mil fogos vagos que não estavam para venda ou arrendamento. Se a análise for feita em termos absolutos, vamos encontrar o maior número de casos em freguesias eminentemente urbanas, como é o caso da União das Freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória, no Porto – com um total de 3382 fogos vagos, seguida de muito perto da freguesia de Santo António dos Olivais, em Coimbra, com 3285 fogos vagos.</p> <p>Se a análise incidir em termos proporcionais – número de fogos vagos face ao património edificado existente –, vamos encontrar casos dramáticos, como o da freguesia de Pinzio, no concelho de Pinhel, que dos 478 fogos existentes, 203 estão vagos – atingindo uma taxa de 42,47%. Esta taxa só é superada, a nível nacional, pela freguesia de Ilha, no município de Santana, na região Autónoma da Madeira, que tem uma taxa de fogos vagos de 46,1% – mas que apenas dispõe de 193 imóveis."</p>		
<p><b>Peça 10</b></p>	<p>[1.º bloco textual do scroller]: "O valor mediano das rendas em Portugal atingiu um novo máximo no final de 2022";</p> <p>Subtítulo da reportagem: "Um aumento de 150 euros na renda desequilibrou o orçamento de Clara, que vive há mais de dez anos no mesmo apartamento. No final de 2022, o valor mediano dos novos contratos de arrendamento foi o mais elevado de sempre.";</p> <p>Texto da reportagem: "(...)A renda do apartamento onde vive desde o regresso do irmão ao Brasil, há mais de dez anos, subiu, pela primeira vez, em Janeiro e vai voltar a aumentar em Julho. "Eu sempre paguei 300 euros e nunca falhei um mês, mas os preços nesta zona foram aumentando e em Dezembro o senhorio contactou-me com uma proposta de aumento, ajustado ao mercado", explica. (...) Clara não escondeu o desconforto e informou o proprietário de que não conseguiria pagar o valor que lhe foi inicialmente proposto. Acabaram por chegar a um</p>	<p>Hiperligação extratextual em "Contas à vida: enfrentar a inflação" que remete o leitor para a página onde pode ser encontrado a hiperligação para o formulário e a base de dados própria no gráfico circular.</p> <p>Identificados dados a partir da entrevista realizada.</p>	

	<p>“acordo”: um aumento de 150 euros, faseado. Nos primeiros seis meses do ano a renda fixa-se nos 400 euros e a partir do segundo semestre terá de pagar 450 euros, se quiser manter a casa.</p> <p>“Fazendo as contas na ponta do lápis”, aos 60 anos e com um ordenado próximo dos 850 euros, sozinha não consegue suportar o aumento desta despesa. Pagos todos os custos fixos, sobram-lhe menos de 100 euros por mês e teme que qualquer imprevisto possa fazer ruir um orçamento, “equilibrado à beira do abismo”. (...) Sentada num sofá preto colocado, provisoriamente, no quarto onde dorme, Clara desdobra o último recibo de vencimento. No verso da página lêem-se anotações a lápis, “renda: 400 euros; luz: 66 euros; água: 22 euros; internet: 69 euros”. No fim da lista, “dois cartões de crédito: 300 euros”. (...) Hoje a dívida chega aos 1.500 euros. (...);</p> <p>[2.º bloco textual do scroller]: "Em 2022, foram celebrados mais de 90 mil novos contratos de arrendamento habitacional";</p> <p>[3.º bloco textual do scroller]: "O maior número de que há registo";</p> <p>[4.º bloco textual do scroller]: "O valor mediano das rendas atingiu, no final de 2022 o preço mais elevado de sempre";</p> <p>[5.º bloco textual do scroller]: "Um aumento de 10,6% em relação ao ano anterior";</p> <p>Texto da reportagem: "(...)Nesta batalha tem encontrado alguns aliados. Está sempre atenta “ao que a cidade tem para dar” e este ano conseguiu um talhão do projeto "Horta à Porta”, para produção agrícola. Paga 60 euros anuais para utilizar o espaço comunitário. (...)"</p>	<p>Os dados mencionados em todos os blocos textuais do scroller carecem de identificação da fonte.</p>	
--	---	--	--

## Anexo 4 - Avaliação qualitativa do orientador de estágio

**RAFAEL OLIVEIRA (02/01/2023 a 31/03/2023)**

### 1- Avaliação do estagiário

Factores	Avaliação			
	1	2	3	4
Assiduidade				X
Pontualidade				X
Interesse				X
Progressão na Aprendizagem				X
Conhecimento da Profissão				X
Relacionamento				X

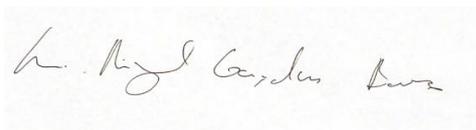
### 2- Avaliação Global

Insuficiente	1	
Suficiente	2	
Bom	3	
Muito Bom	4	X

### 3- Descrição das actividades desenvolvidas

O Rafael desenvolveu todas as tarefas que lhe foram propostas de uma forma totalmente bem-sucedida. Além disso, desde o primeiro dia que demonstra uma capacidade notável de aprender e evoluir na área do jornalismo de dados, tendo demonstrado capacidade de análise de dados e de encontrar potenciais histórias numa base de dados.

Faço, deste modo, uma avaliação muito positiva da sua actividade.



Porto, 31 de Março de 2023