FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Visual narratives supported by dynamic infographics: a case study in the sports domain

Pedro Manuel Santos Queirós

Mestrado em Engenharia Informática e Computação

Supervisor: Alexandre Miguel Barbosa Valle de Carvalho

Second Supervisor: Sérgio Sobral Nunes

July 25, 2023

Visual narratives supported by dynamic infographics: a case study in the sports domain

Pedro Manuel Santos Queirós

Mestrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: Prof. Augusto de Sousa

External Examiner: Prof. Tiago Araújo

Supervisor: Prof. Alexandre Valle de Carvalho

July 25, 2023

Resumo

O ser humano possui uma maior capacidade de interpretação e compreensão quando os dados são apresentados de forma visual e memorável. No caso concreto de dados temporais, uma história pode ser criada com o intuito de apresentar a informação temporal de forma dinâmica e marcante, permitindo que os pontos chave sejam transmitidos à audiência. ZOS é uma plataforma e serviço reconhecido que disponibiliza diversos conteúdos desportivos a diversos meios de comunicação. Possui uma das maiores bases de dados futebolísticas do mundo com informação respetiva a jogadores, competições, clubes e até jogos disputados. Assim, diversos dados históricos estão presentes nesta base de dados levantando um novo problema: representar visualmente e corretamente estes dados de forma que sejam facilmente percetíveis. Para isso, torna-se fundamental apresentar a informação de forma simples, atrativa e intuitiva aos utilizadores. Adicionalmente, as visualizações deverão de ser capazes de contar uma história e adicionar dimensões implícitas aos dados originais, convidando os utilizadores a tirar as suas próprias conclusões. Este trabalho tem como principal objetivo estudar processos de representação visual para dados históricos de forma a conseguir enfatizar a comunicação da narrativa visual criada, ou seja, contar uma história. O dinamismo é necessário para conseguir atrair os utilizadores e capturar a sua atenção. As diferentes formas de visualização podem ser utilizadas dependendo do tipo de dados a representar. No entanto, um conjunto fundamental de características é partilhado para se criar e contar uma história com sucesso. Neste trabalho, é proposto a utilização de infografias dinâmicas, como gráficos de barras, de linhas ou circulares animados, para apresentar a informação histórica aos utilizadores. O dinamismo e atratividade necessários são atingidos através do uso de animações. Estas animações podem ser introduzidas ao controlar o tempo e velocidade de forma a ajustar do ritmo de mudança da informação apresentada. Tudo isto é possível uma vez que os dados históricos estão intrinsecamente associados ao tempo e à sua evolução. Um protótipo foi desenvolvido para validar a solução proposta. O protótipo permite a criação e customização dos três tipos de infografias dinâmicas mencioandas anteriormente. Um procedimento experimental foi também desenvolvido e conduzido de forma a avaliar a eficácia das infografias dinâmicas na transmissão de informação. Para além disso, a sua usabilidade foi também testada. Os resultados mostraram que os gráficos de barras, de linhas e circulares animados foram capazes de representar e transmitir corretamente a informação aos utilizadores. Contudo, melhorias podem ser efetuadas de forma a transmitir corretamente toda a informação apresentada. Em relação à usabilidade, as três infografias dinâmicas obtiveram bons resultados, sendo o gráfico de barras animado o melhor, seguido pelos gráficos circular e de linhas animados. Novamente, melhorias podem ser efetuadas para melhorar as respetivas usabilidades. Assim sendo, os resultados validaram a hipótese proposta e forneceram a informação necessária para responder a todas as questões de investigação levantadas.

Palavras-chave: Visualização de informação, Narrativa, Animação, Infografias

Classificação ACM: Computação centrada no ser humano → Visualização → Domínios de aplicação de visualização → Visualização de informação

Abstract

Humans tend to understand data better when it is presented visually and memorably. Regarding temporal data, a story can be created so that the temporal information is shown in a dynamic and outstanding form, passing the critical points of the story to the audience. ZOS is a trusted service and platform that provides sports data to several media outlets. It has one of the most extensive football databases in the world, with information regarding players, competitions, clubs, and matches. Much historical data is present in all that information available, raising a new problem: to visually represent it in a way that is easy to understand. It is necessary to present that information to the users in an attractive, simple, and intuitive way so that every user can understand what is being shown. Furthermore, the visualizations should also tell a story and add some implicit dimensions to the original data, inviting viewers to draw conclusions. The main goal of this work is to study processes of visual representation of historical data regarding football in a way that emphasizes the communication of a visual narrative or, in other words, tells a story. Dynamism is required to attract users and keep their attention. Different forms of visualization can be used depending on the information to be shown, but a set of essential characteristics should be shared so that a story can be told. This work proposes using dynamic infographics, *e.g.*, bar race charts, line race charts and pie race charts, to present historical information to the users. A story can be told, and the dynamism and attractiveness necessary are accomplished using animations. Animations can be introduced by controlling time and velocity to change and adjust the change rate, which means changing the information being shown. These changes are possible since historical data is time oriented. A prototype was developed to validate the proposed solution. The prototype allows the creation and customization of the three dynamic infographics types mentioned. An experimental procedure was also developed and conducted to evaluate the dynamic infographics' effectiveness and efficacy in conveying information. In addition, their usability was also tested. The results showed that the bar race, line race, and pie race charts could correctly represent and convey the information to the viewers. Nevertheless, improvements can be made to correctly convey all the information. Regarding usability, the three dynamic infographics had a good score, with the bar race chart being the best, followed by the pie and line race charts. Once again, improvements can be made to improve their usability. Therefore, the results validate the proposed hypothesis and provide the necessary information to answer all the research questions raised.

Keywords: Information visualization, Storytelling, Animation, Infographics

ACM Classification: Human-centered computing → Visualization → Visualization application domains → Information visualization

Acknowledgements

I want to express my sincere gratitude and appreciation to all those who have supported me throughout the college journey. Their constant support, guidance, and encouragement have been invaluable, and I am truly grateful for their presence in my life.

I want to thank my supervisors, Alexandre Valle de Carvalho and Sérgio Nunes, for their expertise and guidance during this work. The weekly meetings were pivoted to receive constant feedback, clarify some doubts, and keep the motivation levels high. I am sincerely thankful for the opportunity given and for your belief in my capabilities while keeping challenging me.

I also want to show my gratitude to all the ZOS members for their support, collaboration, and feedback while conducting this work in the company facilities. Their welcoming reception helped to create a good working environment.

A particular word for the ZOS owners, Pedro Dias and Marco Sousa, for believing in my capabilities and ensuring I was well integrated into the company. Thank you for providing the necessary support to conduct this work, from the technological equipment, to access the facilities and the data required for the research.

I am also indebted to my friends who have stood by me during this journey. Thank you for all of your support and adventures that we shared, for the late-night Discord sessions and experiences lived. You made this journey special.

I want to express my heartfelt gratitude to my family, especially my parents and brother, for their constant support, love, and encouragement throughout my life. Your continuous and unlimited support, belief in my abilities, and sacrifices have been my pillar of strength. Thank you for always being there for me, believing in my dreams, and giving me the conditions to accomplish them.

Lastly, I also want to express my heartfelt gratitude to my girlfriend for your love, patience, support, and constant motivation. Your presence in my life has brought joy, balance, and inspiration. Thank you for being my rock and for always being there, cheering me on every step of the way.

Thank you all from the bottom of my heart.

Pedro Queirós

“Success is walking from failure to failure with no loss of enthusiasm.”

Winston Churchill

Contents

List of Figures

List of Tables

Listings

Chapter 1

Introduction

Information is growing in an exponential way [34], creating the need to present it in a simple format to the readers so that they can understand what is being shown [3]. Several companies deal frequently with the question of how they should present information to their users in a more appealing and effective way. This is the case of ZOS, Lda., a company that has a considerable volume of historical information that needs to be presented to their users in an attractive way. Several new techniques have been developed to deliver data in multiple visual formats and it is vital to choose the most effective one according to the context where it will be applied.

1.1 Context and Motivation

Being founded in 2003, ZOS, Lda. (ZOS) is a trusted service and platform provider for several media outlets [2]. Focusing on the sports world, the company has a lot of data collected over the years to ensure that it is always up to date. One of their most extensive databases has information about the origins of football, major leagues, clubs, and players, as well as minor or regional leagues. This level of detailed information is accomplished by continuously keeping in touch with several institutions and their feeds, making the company one of the leading enterprises in the industry.

After gathering all of the information, it becomes critical to present it to the users so they can consume it. Furthermore, the company always aims to use new methods and techniques to present the information to its users properly. Infographics are widely used in the sports domain since they are simple and attractive, allowing the viewer to quickly identify and analyze what is being shown [36]. They are commonly used to display statistical information and make comparisons between players. ZOS pretends to present its large volumes of information, especially historical information, in a more dynamic format capable of attracting and engaging its users.

1.2 Problem Identification

Data is constantly being produced and stored, growing exponentially due to the evolution of technology, which leads to a global digitalization effect [32]. Consequently, domains such as Artificial Intelligence, Data Analysis and Information Visualization are progressively becoming critical to address the problems of processing the data, give meaning and further use to the data and present it so that readers can understand what is being shown, draw a conclusion from it, or even visually analyze it [14].

In the case of time-evolving information visualization, it is not only essential to present the data correctly but also to be able to tell a story in which the reader becomes engaged [26]. Several techniques were developed over the years to accomplish these characteristics [3]. For the purpose of this work, the problem consists of identifying the available techniques and selecting the most suited one according to defined specific criteria regarding the working context, ensuring the success of the representation.

Two main categories emerge regarding visual representation techniques for time-evolving data [33]:

- Static representations, which can be simpler and more objective since they usually only represent one single variable;
- Dynamic representations, capable of representing more variables at the same time with the evolution over time clearly expressed. They can attract and engage the readers [22], resulting in a better storytelling effect. Despite that, it also introduces complexity to what the reader sees, increasing the risk of becoming overwhelming and confusing [46].

Then, which visual representation techniques can be considered the most adequate and effective for telling a story about time-evolving data?

1.3 Hypothesis

Given the previous problem, this work formulates the following hypothesis:

Dynamic infographics techniques, such as bar race, line race, and pie race, are suitable to present historical information as animation would achieve the necessary dynamism while adjusting the rate of change would help control the speed of the narrative display.

In this type of visual representation, the historical information can be represented progressively, allowing the viewer to keep up with the evolution of the data presented while a story is being told. The necessary dynamism can be achieved using animations. Furthermore, the rate of change could be controlled by adjusting the time and velocity of those animations. This control is fundamental to ensure that viewer's attention is kept across the story and to avoid confusion.

1.4 Research Objectives

The previous hypothesis can be divided into the following set of research questions:

RQ1: What makes a visual representation to be considered a good storytelling artifact?

RQ2: What techniques can be considered more effective to present temporal data?

RQ3: Does animation engage the viewers without overwhelming them?

1.5 Methodology

To prove the hypothesis defined in Section 1.3, this work defines the following methodology:

a) study and characterize techniques that are considered suitable for visual storytelling of time-evolving data in particular for the sports domain. A deeper understanding is needed to select the appropriate visualization techniques. Hence, a literature review needs to be conducted to better understand the existing visualization techniques and their advantages and disadvantages. Furthermore, the literature review also allows finding out the key features of those visualizations to ensure their objective is accomplished.

b) cumulatively with a), assess good practices that have already been successfully applied by the industry.

c) develop a solution based on previous work. After finishing the tasks set in a) and b), it becomes clear which techniques are appropriate for the defined problem.

d) develop a framework to create visualizations systematically according to the methods selected in the solution. The techniques selected should allow storytelling to be attractive and engaging to the viewers. Since historical information is the focus, it is not only essential to be strict and accurate in the representation but also to captivate readers to be engaged in the story. By doing this, the risk of overwhelm and boredom is reduced. The ZOS group provided all the information required by the prototype through their API. It contained information regarding the history of football collected over the years. It is intended that the framework should allow the creation of several visualizations from which editors can select the ones considered more suitable, saving time and increasing efficiency.

e) design the protocol for a user experience test, conduct the test, retrieve results, analyze them and reach conclusions. The goal of the test is to evaluate the effectiveness, simplicity and intuitiveness of the platform and of the selected techniques by the professionals of the company and its users. The results were analyzed to verify the usefulness of the solution and its capabilities to address the problems raised.

1.6 Document Structure

This section briefly describes the structure of this document.

Chapter 2, named State of the Art in Storytelling and Dynamic Visualizations, contains an analysis of the state of the art of relevant topics for the context of this work. In particular, it addresses the fundamental concepts of information visualization, its pipeline, and the different taxonomies. Moreover, it also presents the visual analytics domain, the characteristics of presentation, the storytelling approach, and how animations are used to tell an attractive and engaging story. Finally, it also discusses the advantages and disadvantages of dynamic infographics.

Chapter 3, named Problem Statement, presents the problem being addressed in-depth, along with the requirements that a solution needs to meet, its scope and the relation to the research questions. The proposed solution is also explained.

Chapter 4, named Framework for the Creation of Race Chart Visualizations, details the prototype developed, its data flow, and the technology stack used. The prototype comprises two major components: the front end, which displays a user interface and the race charts, and the back end, which requests and processes the required data. Together, they compose a simple interface to create the race chart visualizations.

Chapter 5, named Experiments, explains the experimental methodology developed and applied to test the prototype and the respective proposed solution. It details the general process, procedure, collected data, variables, metrics, and success criteria alongside the questions defined and what they are testing.

Chapter 6, named Results, presents the results of the experimental methodology conducted. The results for each dynamic chart type are presented with the respective usability. In addition, it discusses the results obtained and what they mean, concluding at the end.

Finally, Chapter 7, named Conclusions, explains the main findings of the previous chapters, adding the final remarks of this work. Moreover, it proposes future work to be made on the topic addressed.

Chapter 2

State of the Art in Storytelling and Dynamic Visualizations

This chapter starts by addressing general concepts and definitions regarding the Information Visualization domain. It is divided into several sections: Section 2.1.1 presents the information visualization pipeline, while Section 2.1.2 presents several taxonomies to classify the different techniques in this domain. Furthermore, Section 2.1.3 details the characteristics of visual analytics, while the next Section, 2.1.4, focuses on the presentation of information. Then, Section 2.1.5 presents storytelling as a way of presenting information, and Section 2.1.6 addresses the use of animations to capture viewers' attention and engage them. Finally, Section 2.1.7 summarizes this chapter by analyzing several dynamic charts according to the previous sections.

2.1 Information Visualization

Card *et al.* defined information visualization as *"... the use of interactive visual representations of abstract data to amplify cognition"* [9]. For the authors, visualizations, through the use of 2D and 3D representations and animations, could be very efficient in displaying information, allowing users to navigate, explore and interact with the information.

Plaisant, when describing information visualization, focused on compactness, which means the ability to graphically represent large amounts of data in a way that allows the viewer to be more efficient in discovering, making decisions, and explaining the information presented [37].

Friendly and Denis summarized the history of information visualization from the early maps and diagrams to the computed visualizations nowadays [18]. According to their work, information visualization began in 6200 BC with the oldest known map. Several maps were developed over the years, and in the 17th and 18th centuries, measurement and theory started to take place, and several new graphical forms were developed. In the first half of the 19th century, modern data graphics began to be created, leading to a golden age of information visualization. In the second half of the century, unparalleled beauty and several innovations happened. The period between 1900 and 1949 is known as the modern dark ages since the developments in this domain slowed down. Then, after 1950, the intersection and collaboration between several fields of research, *e.g.*, computer science and data analysis, provided new developments supported by the growth of computational power that would culminate in today's ability to visually represent large volumes of data effectively.

Arboleda and Dewan stated that information visualization takes advantage of the fact that humans are visual creatures, meaning that they depend a lot on what their eyes can see [3]. With this in mind, successful visualization techniques take the most out of the human visual perception system, presenting information in a format that can be easily captured by the human eyes, relying then on the large bandwidth they possess when transmitting that information to the brain.

According to Healey *et al.*, preattentive processing, referring to cognitive operations that take place before focusing attention on any particular region of the image, is fundamental in information visualization [20]. The authors also referred that, typically, tasks that can be performed on large multi-element displays in 200 milliseconds or less are considered preattentive since eye movements take at least 200 milliseconds to initiate. Furthermore, they also concluded that preattentive processing could deliver visual information in a single glimpse.

Van Dam *et al.* highlighted the contrast between machine and human capabilities. While machines are extremely good at simulations, data filtering, and data reduction, humans are experts in using their highly developed pattern-recognition skills to look through the results for regions of interest, features, and anomalies. In comparison to programs, humans are especially good at seeing unexpected and unanticipated emergent properties [52].

Information visualization should be distinguished from scientific visualization, despite being under the visualization domain [9]. For Card *et al.*, information visualization is used with abstract data, while scientific visualization uses scientific data that is often physically based. Furthermore, in this later type, data is defined in reference to space coordinates, making it easier to visualize intuitively [9].

Chen mentioned the necessity that information visualization researchers have in defining the semantics of visual displays as an integral part of the design [11]. In contrast, scientific visualization researchers need to select a reference system with a degree of complexity that may vary from a well-defined theory of a scientific phenomenon to an initial exploration of an emerging phenomenon.

Despite the nature of the data used, Tufte defined the following requisites to successfully build clear, precise, and efficient representations [49]:

- Present the data without distorting it;
- Induce the viewer to focus on the data, avoiding distractions, *e.g.*, graphic design, methodology used;
- Present information in a small space;
- Keep coherency when presenting large volumes of data;
- Define several levels of detail, from a broad overview to the fine structure;

- Follow a clear purpose: description, exploration, tabulation, or decoration;
- Integrate descriptions regarding the dataset, *e.g.*, statistical and verbal.

Maletic *et al.* suggested five dimensions to take into consideration while developing a visualization [30]:

- Tasks - why is the visualization needed?
- Audience - who will use the visualization?
- Target - what is the data source to represent?
- Representation - how to represent it?
- Medium - where to represent the visualization?

Mackinlay also defined expressiveness and effectiveness as two critical factors in transforming data into a visual representation [28]. Expressiveness is the capability of the visual metaphor to represent all the desired information. Effectiveness refers to how well that visual metaphor represents the information, taking the most out of the output medium and the human visual system. These two factors can be applied to both 2D and 3D representations. Marcus *et al.* considered that a solid data characterization is fundamental to fulfilling expressiveness and effectiveness [31]. Moreover, the authors also believed that the power of visualization is derived from its semantic richness, simplicity, and level of abstraction, considering the amount of information presented to the viewers to avoid overwhelming them.

2.1.1 Pipeline

Due to the systematic approach when mapping data into visual representations, Card *et al.* defined a reference model of the mapping process [9] (Figure 2.1).

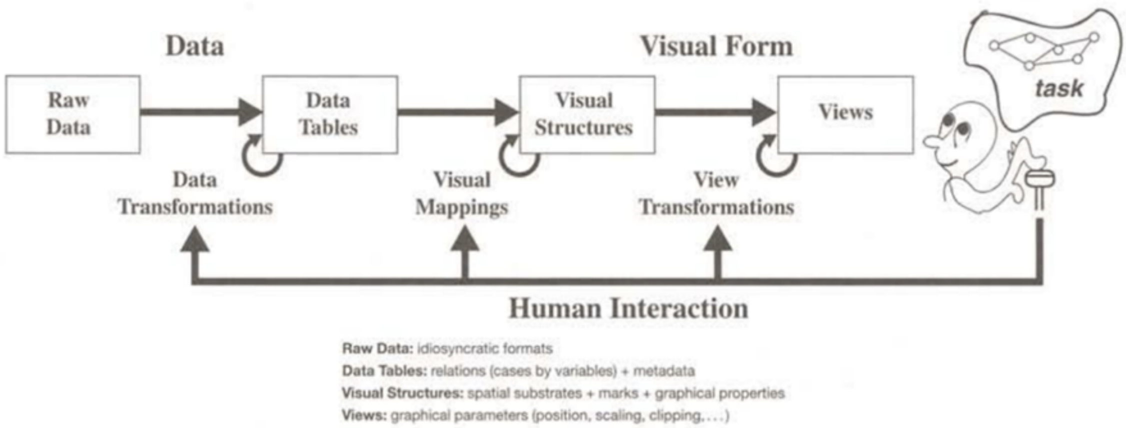

Figure 2.1: Reference model of mapping data into visual representations [9].

The authors defined the data flowing from left to right in this reference model. *Raw data*, which is data in some idiosyncratic format, is the start of the model. Then *Data Transformations* map the *Raw Data* into *Data Tables*, defined as relational descriptions of data extended to include metadata. *Visual Mappings* transform the *Data Tables* into *Visual Structures*, combining spatial substrates, marks, and graphical properties. In the last part, *View Transformations* are used to create *Views* of the *Visual Structures* by specifying graphical properties, *e.g.*, position, scaling, and clipping. These operations are controlled and adjusted through *Human Interaction* by using user-operated controls.

Chi made some modifications to the Card *et al.*'s pipeline (Figure 2.2) [12].

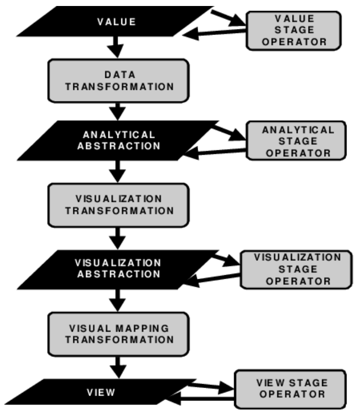

Figure 2.2: Information visualization data state reference model [12].

The author considered four data stages:

1. Value, which is the raw data;
2. Analytical abstraction, which is the metadata;
3. Visualization abstraction, which is the information visually displayed on the screen using a visualization technique;
4. View, which is the end-product of the visualization mapping.

Furthermore, three processing steps connect these data stages:

1. Data transformation, which generates some form of analytical abstraction from the value (usually by extraction);
2. Visualization transformation, which takes an analytical abstraction and further reduces it into some form of visualization abstraction that can be visualized;
3. Visual mapping transformation, which takes information that is in a visualizable format and presents a graphical view.

Fry, in his book, stated that the creation of information visualization includes seven stages [19]:

1. Acquisition - obtain the data to be used;
2. Parsing – provide structure for the data's meaning and order it into categories;
3. Filtering – remove irrelevant data;
4. Mining – apply methods from statistics or data mining to identify patterns;
5. Representation – choose the basic visual model, *e.g.*, bar graph, list, or tree;
6. Refining – improve the basic representation to make it more straightforward and more engaging;
7. Interaction – add interaction methods to control what features are visible.

In addition, the author referred to the dynamism that characterizes this process, meaning that in each scenario, the number of stages used varies. Sometimes only some of the stages will be used, while at other times, all seven stages will be part of the process, making it essential to adapt it to each context and situation [19].

Iliinsky and Steele defined three simple steps for the visualization process [46]:

1. Formulate the question;
2. Gather the data;
3. Apply the visual representation.

The authors pointed out the importance of defining the question or set of questions that guide the viewer to the topic and context within which the data present is most meaningful. Then, it is crucial to gather the data that will be used. Often, it is easier to take data that is already available. Finally, it is necessary to decide which visual representation should be applied to represent the data gathered to portray it better. During this process, the authors highlight one commonly overlooked talent needed to capture the viewer's attention and focus: storytelling [46].

2.1.2 Taxonomies

Shneiderman suggested the Visual Information Seeking Mantra as a helpful starting point to guide for designing user interface: overview first, zoom and filter, then details-on-demand [43]. Furthermore, the author also suggested a task by data type taxonomy, considering seven data types: one-dimensional, two-dimensional, and three-dimensional data, temporal and multi-dimensional data, and tree and network data. Seven tasks are combined in this taxonomy: overview, zoom, filter, details-on-demand, relate, history and extract.

Card and Mackinlay constructed a data-oriented taxonomy [8], which was subsequently expanded [9], dividing the visualization field into seven subcategories: scientific visualization, GIS, multi-dimensional plots, multi-dimensional tables, information landscapes and spaces, node and link, trees, and text transforms.

Regarding the presentation of complex quantitative information, Tufte divided the graphical representations into four categories: data maps, time series, space-time narrative designs, and relation graphics [49].

According to Card *et al.* and Ware, data has the following characteristics [9, 54]:

- Number of dependent variables;
- Number of independent variables;
- Type of each variable:
  - Scalar, vector (2D, 3D, 4D), tensor or more complicated structure;
  - Discrete or continuous;
  - Nominal, ordinal, interval, or ratio.

Tweedie focused on the types of data that can be visually represented: data values, data structure, and metadata [51]. Data values refer to the type of data mentioned above. Data structure refers to data with some structure, *e.g.*, file hierarchies and rectilinear structure. Metadata refers to data about data values or data structure.

Tory and Möller defined a higher-level taxonomy to classify and incorporate a more significant number of visualization techniques based on the design models and not only on data types, as most visualization taxonomies [48] (Table 2.1). In this taxonomy, the authors considered two criteria:

- Whether the model assumes the object of study is discrete or continuous;
- How much the visualization designer chooses display attributes (spatialization, timing, color, transparency).

The authors gave particular importance to spatialization, mentioning the categorization made by Tamara Munzner, which divides spatialization as given or chosen [40]. Tory and Möller added a new category called constrained as the middle term between the two types mentioned previously. Furthermore, the authors also denoted that scientific visualization tends to work with continuous data with a given spatialization, while information visualization tends to represent discrete data with a chosen spatialization.

Tory and Möller also defined two low-level taxonomies: one for continuous models and another for discrete models [48] (Table 2.2 and Table 2.3).

Table 2.1: High-level visualization taxonomy, illustrated by examples [48].

	Display Attributes: **Given**	Display Attributes: **Constrained**	Display Attributes: **Chosen**
Continuous	Images (*e.g.*, medical) Fluid / gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc.) Globe – distribution data (*e.g.*, elevation levels)	Distortions of given / continuous ideas (*e.g.*, flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analyses
Discrete	Classified data / images (*e.g.*, segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (*e.g.*, city locations)	Distortions of given / discrete ideas (*e.g.*, 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (*e.g.*, file structures) Arbitrary multi-dimensional data (*e.g.*, employment statistics)

Table 2.2: Low-level taxonomy of continuous models [48].

# Independent Variables \ Data Structure	**Scalar**	**Vector**	**Tensor**	**Multi-variate**
1D	- Line graph			Combine scalar, vector, & tensor methods
2D	- Colour map - Isolines	- LIC - Particle traces - Glyphs		
3D	- Volume rendering - Isosurfaces		- Tensor ellipsoids	
nD	Multiple 1D, 2D, or 3D views			

Table 2.3: Low-level taxonomy of discrete models [48].

Number of Variables	Variable Types	Example Techniques
2D	1 Dep. + 1 Indep. variable	- Scatter plot - Bar chart
3D	1 Dep. + 2 Indep. or vice versa	- 3D scatter plot - 3D bar chart
nD	Any number of Dep. and Indep. variables	- Charts + colour - Multiple views - Glyphs - Parallel coordinates

Both taxonomies are based on the properties mentioned before. Regarding the continuous models, the taxonomy considers the data structure (scalar, vector, tensor, or multi-variate) and the number of independent variables, affecting the number of dimensions of the representation. Regarding the discrete models, the taxonomy divides the representations according to the number of variables and their types.

2.1.3 Visual Analytics

Thomas and Cook defined visual analytics as *"... the science of analytical reasoning facilitated by interactive visual interfaces"*, according to the panel of the National Visualization and Analytics Center (NVAC) [47]. Furthermore, the authors also highlighted the multidisciplinary character of this science, identifying four focus areas:

1. Analytical reasoning techniques that a lot a deeper insight and support to the users;

2. Visual representations and interaction techniques that exploit the human eye's broad bandwidth into the brain, leading to a better exploration and understanding of a large amount of information;

3. Data representations and transformations to ensure the conversion of conflicting and dynamic data;

4. Techniques to support the production, presentation, and dissemination of the analytical results extracted.

Keim *et al.* described the basic idea beyond visual analytics: information is visually represented to interact with the human to gain insight, draw conclusions, and make better decisions [23]. One of the most famous examples is John Snow's representation, depicting cholera cases in London in 1855 (Figure 2.3), which was crucial in identifying a public fountain as the source of the epidemic [11].

Figure 2.3: John Snow's dot map of cholera deaths [11].

Moreover, Keim *et al.* explained visual analytics as an intersection between two methods to solve analytical problems: automatic analysis and visualization [24]. The automatic analysis uses the power of computers to solve analytical problems that involve large amounts of data, taking advantage of the growing computational power. However, they cannot solve every problem, getting trapped in local optima. Visualization uses human creativity and intuition to solve this kind of problems, being limited to the amount of data that is humanly possible to process. Visual analytics combines both approaches, using the strengths of each method and reducing the weaknesses.

Cui also denoted the constant interaction between humans and machines in the visual analytics process [14], represented in Figure 2.4.

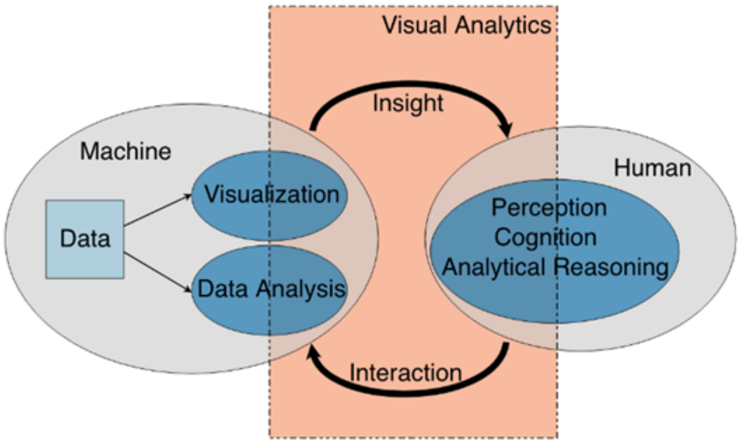

Figure 2.4: Visual analytics as the overlap between machine and human capacity [14].

In the above figure, the author described the interaction through visual representation between human judgment (perspective skills, cognitive reasoning, and domain knowledge) and algorithmic data analyses.

Furthermore, inspired by Shneiderman's Mantra (Section 2.1.2), Cui defined a mantra for visual analytics: analyze/overview first, interaction and visualization repeatedly, insights into data [14]. This mantra describes the following six steps of the visual analytics process:

1. Preprocess (clean, transform, integrate) the data;
2. Apply algorithmic analysis methods to the data;
3. Visualize the data after applying the algorithms with the appropriate visualization techniques;
4. Generate insightful knowledge through human capacities;
5. Make new hypotheses and integrate the newly generated knowledge into the analysis and visualization;
6. Regenerate an updated visualization based on the changes of the previous step.

Battle and Heer referred to visual analytics as Exploratory Visual Analysis (EVA), highlighting the process needed to perform data analytics on the dataset [5]. The authors described EVA as a high-level analysis goal that can be precise by exploring a specific hypothesis, or vague, by exploring the data and trying to find something interesting about it, having visualization as the primary output medium for the exploration. Furthermore, due to the dimension and complexity of the data, the exploration task is divided into sub-tasks, which can be divided into sub-tasks, and so on, until a simple sub-task is defined.

2.1.4 Presentation

Mackinlay described presentation as the use of visual representations to communicate [29]. The author gave William Playfair's time-series representation in 1786 as an example of one of the first uses of visualization in representation (Figure 2.5).

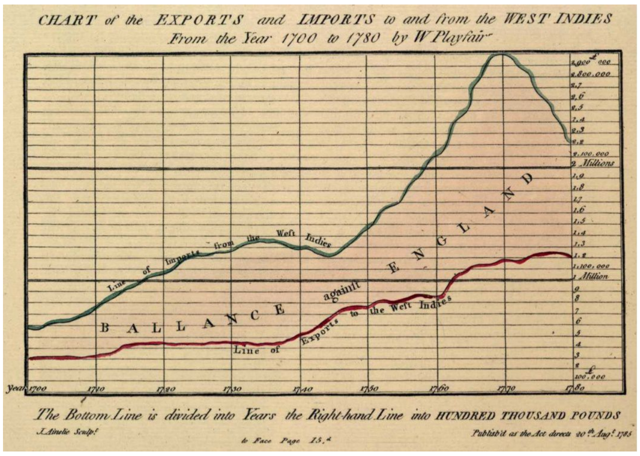

Figure 2.5: William Playfair's chart of the exports and imports from the West Indies by England as one of the first uses of visualization in representation [44].

Spence focused on the presentation's spatial and temporal layout component [45]. The author pointed out the role of presentation as a middleman between representations and the user, allowing them to view and interact with the visual display.

Chittaro defined presentation as how to manage and organize the information in the available space [13]. The author emphasized the necessity of using the available space effectively to ensure that information is presented in a more meaningful and understandable format.

Kosara and MacKinlay stated that the goals and approaches of presentation differ from those in analysis described in the previous section [25]. The authors described the primary purpose of presentation as getting the point across or explaining a finding.

Iliinsky and Steele also mentioned the importance of distinguishing the goals and procedures of presentation and exploration used in analysis [46]. The authors synthesized the main differences between presentation and exploration in Table 2.4.

According to Iliinsky and Steele, in exploration, an analyst does not know what the data will show, having to find different correlations and connections to extract knowledge from it [46].

Table 2.4: Exploration VS Presentation [46].

	Exploration	**Presentation**
Characteristics	Data is surprising. Data may have outliers. Data is likely to move unpredictably. Viewers control interaction.	Data is well known to the presenter. Data has been cleaned. The viewer is passive.
Goals/Procedures	Analyze multiple dimensions. Change mappings many times. Look for trend and holes.	Present fewer dimensions to make a point. Walk through dimensions clearly. Highlight critical points. Group points together to show trends and motion.

Furthermore, the data may have outliers since it wasn't previously processed. In contrast, in presentation, presenters are experts in their data, aiming to highlight the critical points that support the core ideas to be transmitted to the viewers. Moreover, errors were already cleaned from the dataset. Despite distinguishing exploration and presentation, the authors mention that most tools mix the two, *e.g.*, Microsoft Excel [46].

2.1.5 Storytelling

"Storytelling is the cornerstone of human experience" [3].

Figueiras described the longevity of stories in human history, being used by man to entertain, educate and instill moral values since the beginning of time [17]. In addition, the author stated that stories, in comparison to other ways of presenting information, can prevail due to their high power to help assimilation and retention, being more compelling.

Arboleda and Dewan described stories as a way of conveying information since they are more compelling and ease the understanding of facts [3]. Furthermore, the authors defined stories as a sequence of steps, where the timeline is the dominant element.

One of the most compelling exemplars of visual storytelling is Charles Minard's depiction of the Russian campaign of Napoleon's army [25]. The representation shows the size of the diminishing army on its way to Moscow and the even more staggering losses on its retreat (Figure 2.6).

Ma *et al.* also described a story as a sequence of causally related events [27]. Additionally, the authors also identified some important features common to all the good stories:

- They take time to unfold as their pace matches the audience's ability to follow them;
- They capture the audience's attention;

- They leave a lasting impression.

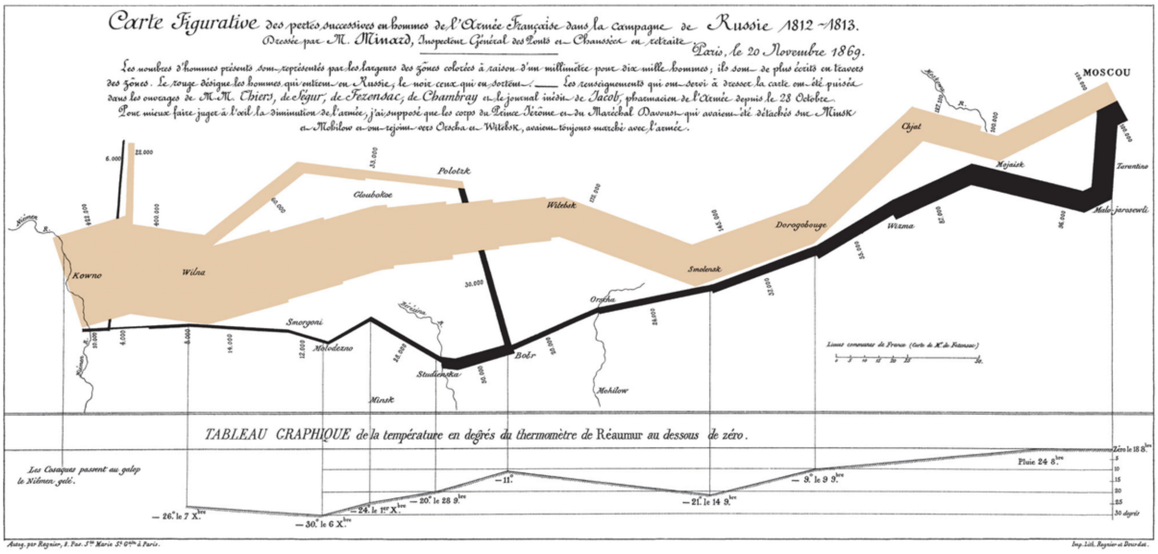

Figure 2.6: Napoleon's Russia campaign by Charles Minard [25].

Furthermore, stories make information more available to episodic memory, which is the part of the memory used for remembering sequences of events [27].

Kosara and MacKinlay defined a story as an ordered sequence of steps containing words, images, visualizations, video, or a combination of both [25]. The authors emphasized the causality transmitted by stories to the users since they are often ordered by time, meaning that previous events may influence future events.

Segel and Heer suggested two approaches to storytelling: author-driven and reader-driven stories [42]. Author-driven stories don't contain any interactivity with the audience, following a linear ordering of scenes to pass a specific message. In contrast, reader-driven stories are more dynamic, allowing the viewers to interact freely and draw their interpretations.

Segel and Heer also identified three common structures based on these approaches: martini glass structure, interactive slideshow, and drill-down story [42]. The martini glass structure begins with an author-driven approach, starting with a broad introduction and narrowing to state a particular point. Then, it opens up interaction with a reader-driven approach. This structure resembles a martini glass, with the stem representing the author-driven phase and the widening mouth of the glass representing the reader-driven phase (Figure 2.7). The interactive slideshow has a typical slideshow format, incorporating mid-narrative interaction within the limits of each slide. It is a more balanced structure between author-driven and reader-driven approaches, allowing the viewer to explore particular points before moving to the next step in the story. This structure works well with both complex datasets and narratives. Finally, the drill-down story uses the reader-driven approach, presenting a general theme and then allowing the viewers to choose what they would like to deepen among particular instances.

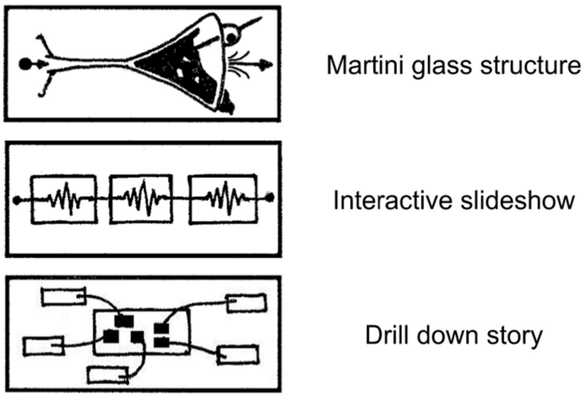

Figure 2.7: Narrative visualization structures [42].

Regarding the audience, Kosara and MacKinlay systemized three typical scenarios for storytelling [25]:

- Self-running presentations for a large audience;
- Live presentations;
- Individual or small-group presentations.

According to the authors, self-running presentations for a large audience is the scenario that media faces, where a presentation is shown to several people independently, aiming to get the point across and explaining it sufficiently to the viewers, *e.g.*, a film clip. In the live presentations scenario, a live speaker is in front of an audience, interacting with it by answering their questions. Furthermore, the presentation can be adapted according to several signs given by the audience. Finally, individual or small-groups presentations can be similar to live presentations. However, in this case, the audience is much smaller. Moreover, the goal is not only to disseminate information but also to collect and condense additional knowledge. The presentation should be flexible to allow a strong interaction between the presenter and the audience to fulfill its goals.

In addition, Arboleda and Dewan complemented the perspective of Kosara and MacKinlay by defining three practical scenarios [3]:

- The presenter is showing findings to an audience without previous knowledge of the information shown;
- The presenter is doing a routine presentation to an audience with some knowledge of what is being shown;
- The presenter shows the information to a small audience with expert domain knowledge.

Zhi *et al.* focused on two particular forms of integrating narratives with visualizations to successfully build a story: linking and layout [55]. Linking refers to using explanatory text linked with its visual counterpart and vice versa. Layout refers to the spatial arrangement of text and visual elements to enable effective communication of the story. Figure 2.8 illustrates an example of a vertical layout and linking, where the explanatory visualization elements are highlighted when hovering over the first sentence of the second paragraph. At the same time, Figure 2.9 shows the same example but with a horizontal layout.

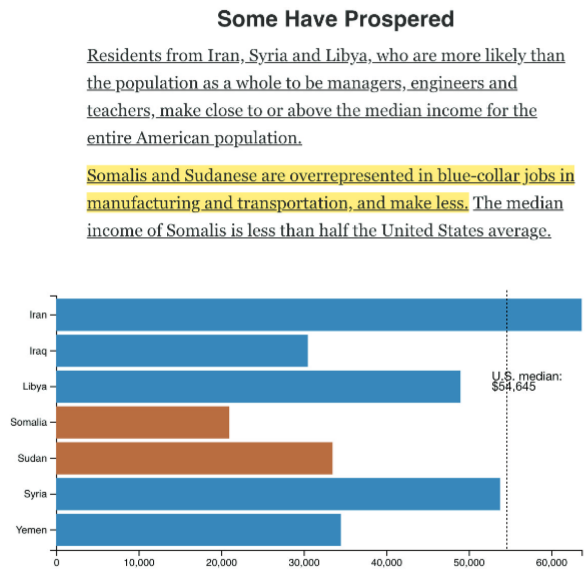

Figure 2.8: Linking example with a vertical layout [55].

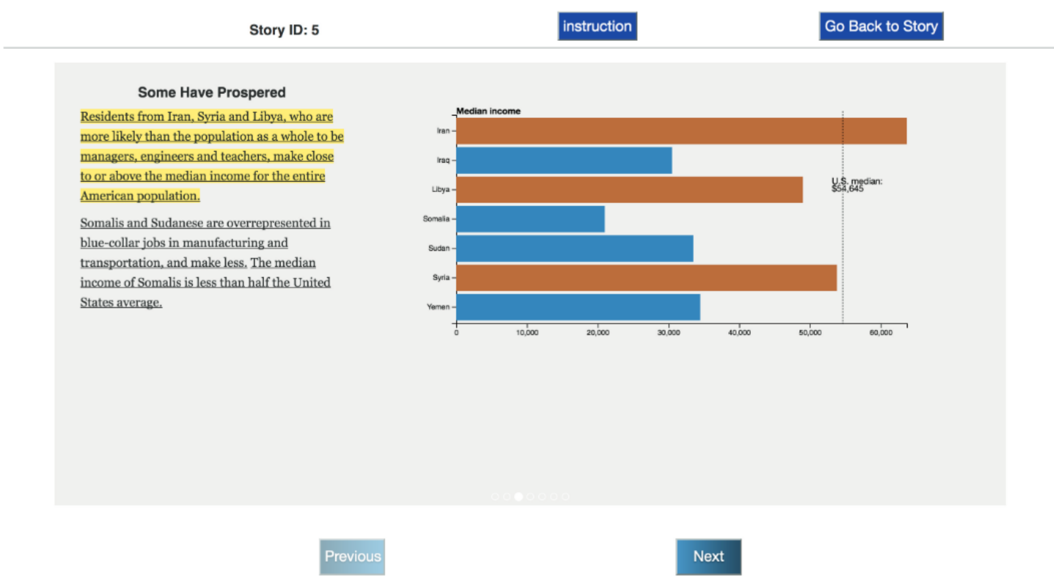

Figure 2.9: Linking example with a horizontal layout [55].

The study conducted by the authors showed that a horizontal layout supported better the viewers' comprehension and that linking improved their engagement and reduced the confusion created when visualizing several elements.

Wang *et al.* compared the effectiveness and engagement of three storytelling techniques: data comics, infographics, and illustrated texts [53]. Illustrated texts are less memorable than natural scenes, something that can be changed by using embellishments. Unnecessary text and closely integrating text and pictures can distract the reader. Data comics focus more on abstract data, being compatible with large audiences with different formats, and more engaging and easier to understand for non-expert readers. Infographics stimulate attractiveness and memory by using embellishments, delivering both overview and detail.

The authors created a two-dimensional design space that can be used to classify these different techniques (Figure 2.10).

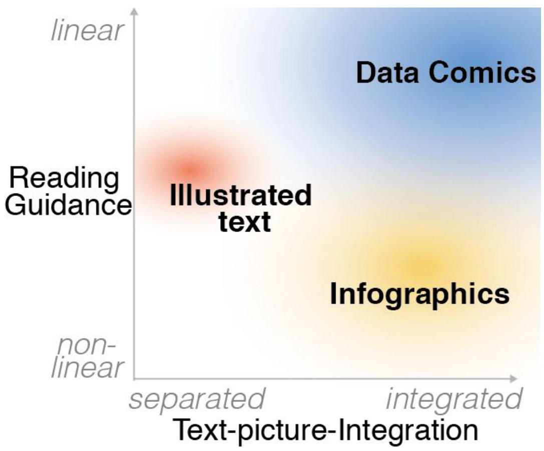

Figure 2.10: Design space to classify different visualization techniques [53].

In this design space, the horizontal axis, called *Text-picture-Integration*, refers to the spatial distance between the verbal message and the visualization. The vertical axis, called *Reading Guidance*, measures how strongly the reader is guided while reading the story. Data Comics guide the reader through the whole story and have text directly connected to the visual elements. Illustrated texts also guide the readers, despite providing less reading guidance to them. Often the text and the visual elements are less connected. Infographics also have their visual elements strongly connected with the text. However, they allow readers to explore the story freer by having less reading guidance.

The survey conducted by Jiang *et al.* inside the infographic's domain showed that line graphs and bar graphs are the preferred types for representing time series data [22]. Furthermore, dynamism is often introduced to retain viewers' attention, matching the fast pace of their lives. Bar graphs are intrinsically connected with this dynamism introduced in visualization.

2.1.6 Animation

Kriglstein *et al.* referred to animation as a technique that aims to create the illusion of movement to show, in a transparent way, the changes in a presentation [26]. Furthermore, animation could also be used in analysis tasks.

Iliinsky and Steele described animation as a rapid succession of images [46]. According to the authors, there is a simple idea behind animation: if an image in a two-dimensional space is good, then a moving image should be better. Furthermore, the growing use of animation in information visualization is justified by the increasing use of programming languages, such as Java and JavaScript, on the Web.

Animation benefits from the research and evolution of several fields, from traditional cartoons to educational psychology [46, 26]. Two main points of educational psychology impact the animation field: the phenomenon of change blindness and the ability of humans to track several moving targets. The first one states that identifying changes in scenes is very difficult since humans tend to concentrate only on restricted areas in their field of vision [39]. This can be addressed by drawing the viewer's attention to specific changes. As for the second one, Pylyshyn stated that humans could only track around four similar targets easily [38]. Cavanagh and Alvarez even stated that the eyes give up, tracking only four or five objects and labeling other movements as noise [10].

Tversky *et al.* defined two principles of effective animation [50]:

1. Congruence, stating that the structure and content of the external representation should correspond to the desired structure and content of the internal representation, *e.g.*, effective external visualization of routes will be based on turns since routes are conceived of as a series of turns;

2. Apprehension, stating that the structure and content of the external representation should be readily and accurately perceived and comprehended, *e.g.*, in the routes example, exact angles of turns and lengths of roads are not important since people represent angles and lengths in gross categories.

Regarding interaction, the authors referred that it could be used to overcome the difficulties of perception and comprehension, where stopping, starting, and replaying an animation could allow reinspection, focused on specific parts and actions [50].

Heer and Robertson defined guidelines as specific recommendations for adhering to Tversky *et al.*'s principles [21]. Under the congruence principle defined, the following guidelines:

- Maintain valid data graphics during transitions – avoid uninformative animation and consider the relation between axes and the data marks during transitions to ensure viewers' mental models are congruent with the semantics of the data;

- Use consistent semantic-syntactic mappings – similar semantic operators should have suitably similar transitions across different types of data graphics to help understand;

- Respect semantic correspondence – ensure that syntax does not violate semantics; otherwise, it can lead to poor interpretations;
- Avoid ambiguity – avoid ambiguous semantics across transitions.

Under the apprehension principle, the authors defined the following guidelines:

- Group similar transitions – according to the Gestalt Principle of Common Fate [35], objects that undergo similar visual changes are more likely to be more perceptually grouped, helping viewers to understand that those objects are undergoing the same operation simultaneously;
- Minimize occlusion – object occlusion during transitions makes it more difficult to track;
- Use simple transitions – multiple simultaneous transitions or complicated transitions lead to an increase in cognitive load;
- Use staging for complex transitions – some transitions are inherently complex. In this case, they should be divided into simple sub-transitions to be more easily tracked;
- Make transitions as long as needed, but no longer – transitions should be long enough for accurate change tracking. Taking too much time can lead to longer task times and diminished engagement [6, 41]. Robertson *et al.* recommended transition times of around one second, which can be reduced for minimal movement animations [41].

Furthermore, Heer and Robertson also proposed a taxonomy of transition types to better identify and distinguish the several animations used [21]. They identified seven types:

- View transformation, where the viewpoint is changed, *e.g.*, panning or zooming;
- Substrate transformation, where the spatial substrate in which marks are embedded change, *e.g.*, axis rescaling;
- Filtering, selecting what element should be visible;
- Ordering, where the ordinal data dimensions are rearranged, *e.g.*, sorting the points;
- Timestep, which applies temporal changes to data values, *e.g.*, a transition from values of different years;
- Visualization change, which consists of changing the visual mappings applied to the data, *e.g.*, change from a bar chart to a pie chart;
- Data schema change, where the data dimensions being visualized change.

Regarding animations used in presentations, Zongker and Salesin defined nine principles to better create effective presentations [56]:

- Make all movements meaningful to avoid losing the audience's attention;
- Avoid instantaneous changes since that smooth transitions are more natural and create the feeling of continuity;
- Reinforce structure with transitions by using more showy effects to highlight critical points that will contrast with subtle transitions;
- Create a large virtual canvas by using panning and zooming, addressing the limited space available for the information on the screen;
- Smoothly expand and compress detail by making the impression that the screen is a window onto a vast space or a kind of magnifying glass to see the presentation at different levels;
- Manage complexity through overlays, presenting only the necessary information on the screen at each point;
- Do one thing at a time to avoid confusing the audience with several transitions happening at the same time;
- Reinforce animation with narration, having the two simultaneously;
- Distinguish dynamics from transitions. Dynamics refer to the most natural use of animation (depicting change over time), and transitions refer to all the other uses of animation.

According to the empirical research by Kriglstein *et al.*, animations are very effective when presenting time-oriented data, improving the viewers' capability of tracking changes [26]. However, that effectiveness depends on the amount of data presented since too much data will get viewers lost. Moreover, interaction, animation speed, and the presentation's context and design also impact the animations' success.

Although animations can be very effective, they should be carefully planned since they always introduce a burden of complexity to the viewers, and that complexity should pay off [46, 26].

2.1.7 Summary

An analysis of animated charts, particularly bar race charts, line race charts, and pie race charts, was conducted to better understand their advantages and disadvantages. All the charts analyzed presented time-oriented data from different domains, *e.g.*, sports, economy, and world history. The conclusions are summarized in the following Table 2.5.

The animated charts allow a straightforward presentation of time-oriented data. These charts rely on the dynamic of change to capture the viewer's attention and engage it with the story that is being presented. Complementary information, *e.g.*, text boxes, can be used to complement the information shown on the graph. It takes advantage of the linking and layout features, helping the viewer to keep track of changes. These advantages are in line with the current literature described before. Additionally, it is possible to extract implicit knowledge from the information presented.

Table 2.5: Advantages and disadvantages of animated charts.

Advantages	**Disadvantages**
Easy to understand what is being presented. The addition of dimensions to extract implicit information. Complementary information can be used. Dynamic of change. Easy to capture the viewer's attention.	Confusing and difficult to track all the changes when several animations happen simultaneously. Details are lost with a high rate of change. Details are lost with a low rate of change. Variables with less value will not be presented if much information is displayed.

In other words, when having some previous knowledge of the domain, it becomes possible to relate it with the information presented in the chart, adding new dimensions to the visualization.

The disadvantages of the animated charts are associated with the data size, which is also highlighted in the literature. Due to the constraints of the chart, variables with lower values may be missed by the viewer or may not even be depicted when there is a lot of information being presented. Furthermore, several changes at the same time create confusion. The rate of change is a feature that should be carefully handled. With a high rate of change, the details of the information will be lost in the presentation. However, with a low rate of change, those details can also be lost since the viewer will be expecting some change, which will only take place after a larger period. In this latter case, the dynamism will be lost, and the viewers will have the feeling of a static representation. Controlling time and velocity in the representation could be a solution to address these rate of change problems. When having a low rate of change, the speed of the animations could be higher so that the passage of time in the chart will be faster to keep the effect of an animated representation. In contrast, when having a high rate of change, the velocity of the animations could be lower so that the passage of time in the chart will be slower to avoid the loss of details.

Chapter 3

Problem Statement

This chapter details the problem previously described in Section 1.2 and presents the proposed solution. Section 3.1 details problem specificities addressed in this work while Section 3.2 presents the research questions intended to be answered at the end of this process. Next, Section 3.3 describes the requirements that a solution should comply with to address the problem defined entirely, while Section 3.4 presents the scope of this work. Section 3.5 details the proposed solution alongside its data flow, while Section 3.6 summarizes the problem and solution key points.

3.1 Problem

Data is growing exponentially since it is continuously being produced and stored by platforms [34], raising a new major challenge to them: how to present the information to users correctly. Furthermore, attracting and engaging the user without overwhelming it with too much information also becomes essential [26]. In the ZOS case study, there is a lot of historical and statistical data on sports up to the present. For this type of information, it becomes mandatory to tell a visual story in a way that attracts the user and keeps it engaged without distorting the original data.

Two main categories emerge regarding visual representation techniques for time-evolving data: static representations, which can be simpler and more objective (as they usually represent one single variable), and dynamic representations, capable of representing more variables simultaneously with the evolution over time clearly expressed [33]. Moreover, dynamic representations can attract and engage the readers, resulting in a better storytelling effect [22]. Despite that, it also introduces complexity to what the reader sees, increasing the risk of becoming overwhelmed and confused [46].

From the plethora of existing presentation/visualization techniques in these two categories [3], it becomes vital to select the proper ones according to the context in which they will be used. In the ZOS case, it will be the sports domain. Moreover, it is also critical to create visualizations that can be consumed simply and intuitively. At the same time, the under-the-hood process should be automatic and systematic to distinguish stories across the multiple entities of the information repositories, preventing design editors, which are responsible for creating the online content, from the need to create each story presentation.

3.2 Research Questions

The problem identified in the previous section, regarding on how to tell a visual story in an attractive and engaging format, can be divided into the three following research questions:

RQ1: What makes a visual representation a good storytelling artifact? Identifying the core characteristics that make visual representations a good format to tell a story is essential, so the users become attracted and engaged.

RQ2: What visualization techniques can be considered more effective for presenting time-evolving data? It is crucial to identify the most effective methods to present temporal data so that the viewers can easily understand what is being shown and how it is changing over time.

RQ3: Does animation engage the viewers without overwhelming them? Understand if animations, in the context of temporal data, help the viewers to quickly understand the variables that are changing over time while capturing their attention and interest.

3.3 Solution Requirements

The solution needs to fulfill the following requirements to address the problem described in Section 3.1 thoroughly and to answer the research questions presented in the previous section:

D1: Implement the visual techniques and animations selected, *e.g.*, bar races, line races and pie races to assess to which extent the representations created are attractive and engaging. Implementing the visual techniques and animations selected allows the identification of the core characteristics that make a visual representation a good storytelling artifact and the evaluation of the best techniques for presenting temporal data. Furthermore, it also enables the assessment of the role of animations.

D2: Provide customization of the representations created, *e.g.*, adjust the narrative velocity so that the rate can be adapted to the amount of information that is displayed. This customization allows the evaluation of animations' effect on the user.

D3: Provide a process to easily create visual representations so that editors can be more efficient and save time in the process. Creating different graphical representations enables the identification of the more effective ones in presenting temporal data.

D4: Access an API from which the source data can be retrieved to be used in the representation. This way, a wide range of time-evolving data can be used to create various narratives that will allow the evaluation of the effectiveness of the selected visual representations.

D5: Allow the editor to select the data used in the representation.

D6: Have an intuitive interface so that every professional can use it.

3.4 Scope

Considering the limited time for developing a solution prototype, it becomes relevant to define the priority assigned to each solution requirement regarding the prototype implementation, experiments, and result analysis. Based on their relevance in the development of a capable prototype to be used to validate or refute the hypothesis proposed in this work, the requirements with higher priority were considered the following:

1. The creation of dynamic infographics by using animations to attract and engage the users;
2. The systematic process of creating those infographics intuitively.

Regarding the end-users of a solution, they can be divided into two groups:

1. The editors that create the dynamic infographics. They are the first ones to judge the result;
2. The users that consume information through visual representations in several platforms, *e.g.*, websites.

3.5 Proposed solution

From the literature review conducted in Chapter 2, it became clear that infographics supported by bar, line, and pie charts are one of the best techniques to represent temporal data visually. Furthermore, they have a good storytelling effect and can be combined with animations to introduce dynamism in the visualization and better represent the evolution of data over time. With this in mind, a solution was researched to allow the users to create and customize dynamic infographics, *e.g.* bar races, line races, and pie races, so that they can capture the audience's attention and interest by telling an engaging, dynamic visual narrative. As mentioned in Section 2.1.7, dynamic infographics can correctly present the information to the viewer and easily capture its attention. Furthermore, it was considered the proposed solution to encompass customization so that it provides a systematic process to create dynamic infographics:

- Customize the different elements that complement the information presented, *e.g.* title, and labels;
- Customize the velocity of the narrative so that it can be adapted according to the information being presented.

The solution proposed follows the conceptual data flow model presented in Figure 3.1, which was defined based on the two pipelines presented in Section 2.1.1.

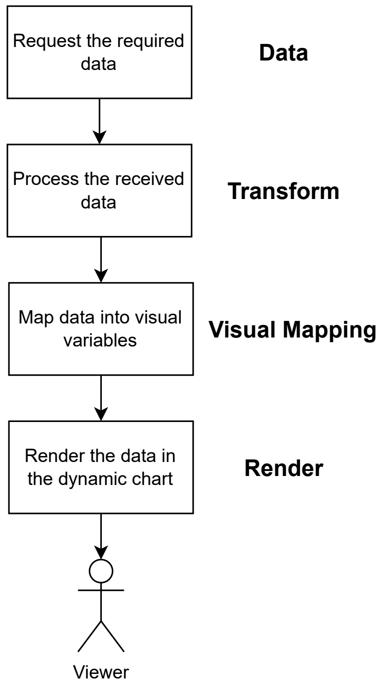

Figure 3.1: Conceptual data flow model of the proposed solution.

The conceptual data flow model starts by requesting the required data for later use in the dynamic infographic. Then, the received data is processed and transformed so that it can be mapped into visual variables in the next stage. Finally, these variables are rendered, and the dynamic infographic is presented to the viewer.

3.6 Summary

Section 3.1 started by presenting and explaining the problem addressed by this work. In short, it consists of telling a visual story in an attractive and engaging format for the users without distorting the original data. Next, Section 3.2 presented the research questions this work intends to answer. In summary, they consist of understanding what makes a visual representation a good storytelling artifact, what visualization techniques can be considered more effective for presenting time-evolving data, and if animation engages the viewers without overwhelming them. Then, Section 3.3 addressed the requirements that a solution should comply with to address the problem entirely. These requirements are based on the creation of different types of dynamic infographics, *e.g.*, bar races, line races, and pie races, that can be customized, *e.g.*, adjusting the narrative velocity, and that communicate with an API to obtain the necessary data for the representations. All combined, it provides an easy process to create visual representations. Section 3.4 took a deeper look at the scope of the problem, defining the features with higher priority in the implementation of a solution and identifying the two main groups of end-users that will benefit from it. Finally, Section 3.5 described the proposed solution, which consists of creating dynamic charts that can be customizable so that the information is presented correctly and attractively to the viewers. A conceptual data flow model was also presented.

Chapter 4

Framework for the Creation of Race Chart Visualizations

This chapter describes how the proposed solution to create engaging infographics using animation was implemented in a prototype. Section 4.1 focuses on the prototype implementation, features, and the technology stack used and Section 4.2 summarizes the main outcomes.

4.1 Prototype Implementation

The prototype was developed based on the solution requirements mentioned in Section 3.3. It is divided into two parts: the first is responsible for rendering the three dynamic chart types with the respective customization, and the second is responsible for requesting the required data to the API. This division into two components allows the separation between requesting and processing the data and mapping them into visual variables and rendering, leading to a more efficient process of creating dynamic charts. In addition, the prototype has a simple interface so that every editor can create and customize the desired dynamic chart. With this in mind, the prototype implemented follows the architecture illustrated in Figure 4.1. The system architecture is divided into two major components: the front end, composed of the user interface which allows the creation of dynamic infographics, and the back end, which is responsible for retrieving the data for the representations and forwarding it to the front end. These two components are addressed in detail in the following sections.

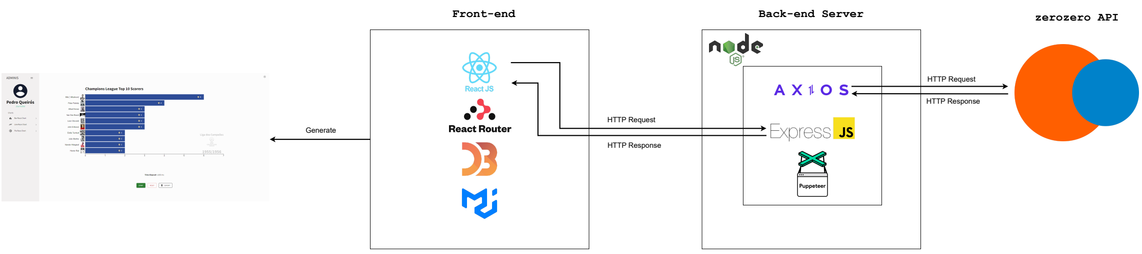

Figure 4.1: Prototype's architecture.

4.1.1 Front end

Several enabling technologies support the front-end component, namely ReactJS[1], React-Router[2], and Material-UI[3]. ReactJS was selected due to successful previous experiences with it and its popularity nowadays. Moreover, it allows a fast development of user interfaces which is an advantage due to the time constraints for this work. Combined with ReactJS, Material-UI was used, as it provides pretty, simple, and intuitive elements for any user interface. Three pages are available to the user, each corresponding to each dynamic chart type: bar race, line race, and pie race. These three pages are supported by four routes, which were created through the React-Router:

- / or /barRaceChart - is the default route, showing a page where a bar race chart can be created (Figure 4.2);

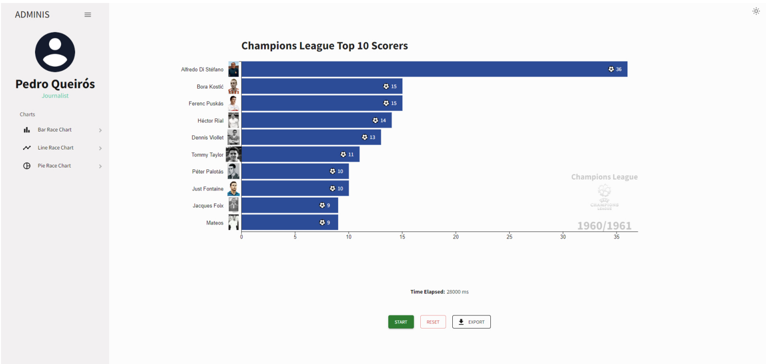

Figure 4.2: Bar race chart page.

- /lineRaceChart - shows a page where a line race chart can be created (Figure 4.3);

Figure 4.3: Line race chart page.

- /pieRaceChart - shows a page where a pie race chart can be created (Figure 4.4);

[1] https://reactjs.org
[2] https://reactrouter.com/en/main
[3] https://mui.com

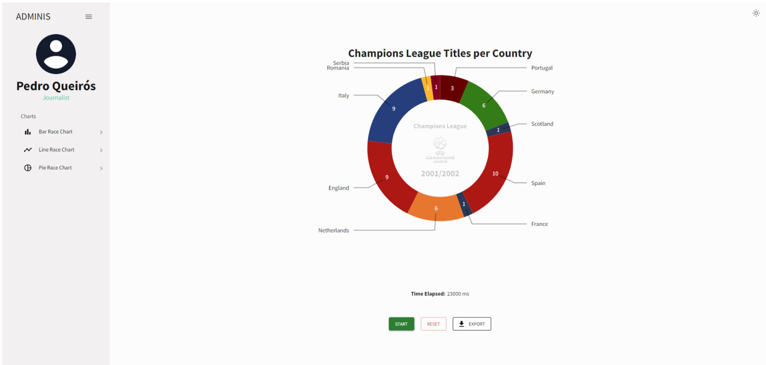

Figure 4.4: Pie race chart page.

As for the creation of the dynamic charts, after some research and feedback on several options available, the following JavaScript libraries that can be combined with ReactJS were tested:

- Chart.js[4], a free and simple library for making HTML-based charts, having eight built-in chart types, e.g., scatter plot, line chart, and bar chart;
- React-Vis[5], a library developed by Uber Eats, which provides a set of react components to render common data visualization charts, e.g., line charts, bar charts, and heat maps;
- Recharts[6], a chart library that provides simple React components to create SVG charts, e.g., bar charts, line charts, and scatter charts.

A small dataset of simple values updated over time was used to test these libraries. The test consisted of creating a dynamic chart based on the small dataset and then judging the result. After the experiments, it became clear that the libraries mentioned above offer limited options for animated charts. The libraries were built to create standard static charts, *e.g.*, bar, line, and pie charts, offering plug-and-play options. They only supported basic animations, which were limited according to the requirements established for this work. Each previous library was built based on another JavaScript library, D3.js[7]. D3.js is a JavaScript library for manipulating documents based on data, working with HTML, SVG, and CSS, which allows the creation of several types of animated charts from scratch with a high level of customization. The same experiment with D3.js confirmed that it was the most suitable library for the prototype.

An editor user can select which dynamic chart type wants to visualize in the section highlighted in Figure 4.5 with the number 1. There are also three buttons available to the user, highlighted with the number 2 in the same figure: one to start the chart's narrative; another to reset the chart's narrative; and the final to export the dynamic chart in video format. A dark/light mode option is also available, identified by the number 3.

[4] https://www.chartjs.org
[5] https://uber.github.io/react-vis
[6] https://recharts.org/en-US
[7] https://d3js.org

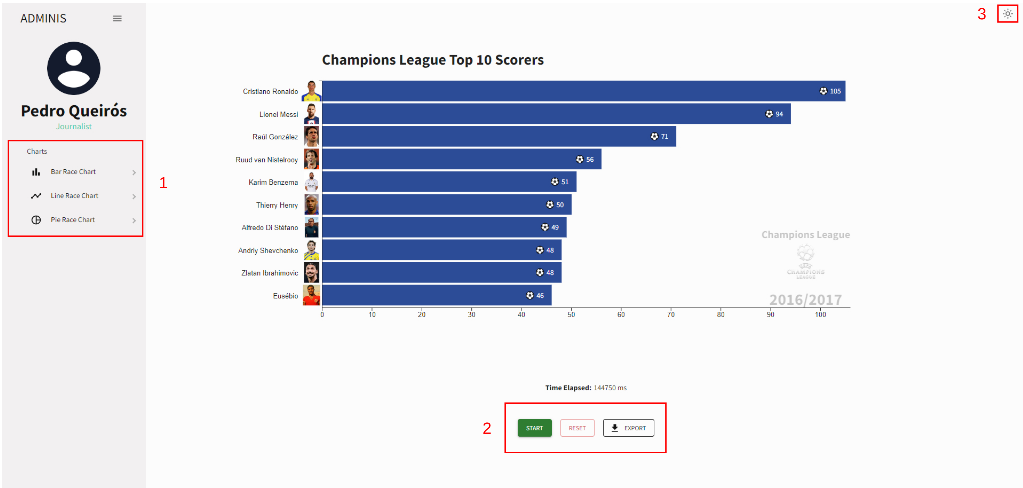

Figure 4.5: Prototype's interface highlighting the selection of the dynamic chart desired (1); the option to start, reset, or export the narrative (2); and the dark/light mode option (3).

Besides displaying the selected dynamic chart to the editor, the platform also allows the customization of the dynamic chart presented, identified by the number 4 in Figure 4.6.

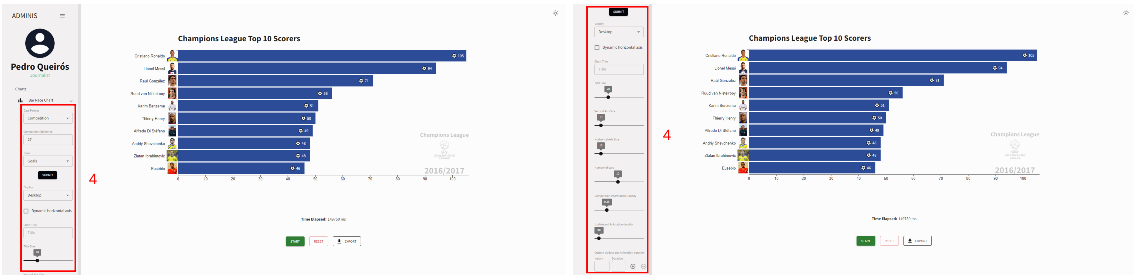

Figure 4.6: Prototype's interface highlighting the customization available to the user.

Furthermore, the following customizations are available for each type of dynamic chart:

- Define the required data - important to choose the information that is going to be displayed;
- Select the type of display, which is available in two options: desktop and mobile - important to adapt the display format according to where the chart is going to be used so that no information is lost;
- Define the title - it describes the representation and explains to the viewer what will be displayed;
- Define the title size - ensures that the title has the proper size to be read and draw attention;
- Define the size of the text in the axis (or labels in the case of the pie race chart) - ensures that the complement information given by the axis and labels is visible;

- Define the opacity of the identifier of what is being shown - by adjusting the opacity, the dynamic chart content can sometimes overlap the information given by the identifier without losing any information;

- Define the animation duration - crucial to adjust the overall velocity of the narrative displayed so that the user's attention is kept without losing any information;

- Define animation duration according to specific intervals - crucial to change the velocity of the narrative according to the rate of change of the information displayed. It allows an increase or decrease the velocity after specific timestamps of the visualization.

Moreover, other customizations are specific and available depending on the chart type. Regarding the racing bar chart, it is possible to define the number of bars displayed and select whether the x-axis is more dynamic or static. The editor can choose if both axes are more dynamic or static in the racing line chart. An axis is more dynamic if the minimal value represented changes as the data is updated. In contrast, an axis is more static if the minimal value remains the same as the data is updated. Finally, in the pie chart, it is possible to define the thickness of the line that connects the labels to the respective slices. These properties complement the information displayed on the dynamic chart to ensure no information is lost.

The following figures illustrate the three types of dynamic charts (bar race chart in Figure 4.7, line race chart in Figure 4.8, and pie race chart in Figure 4.9) before being customized and after being customized, according to the parameters previously mentioned:

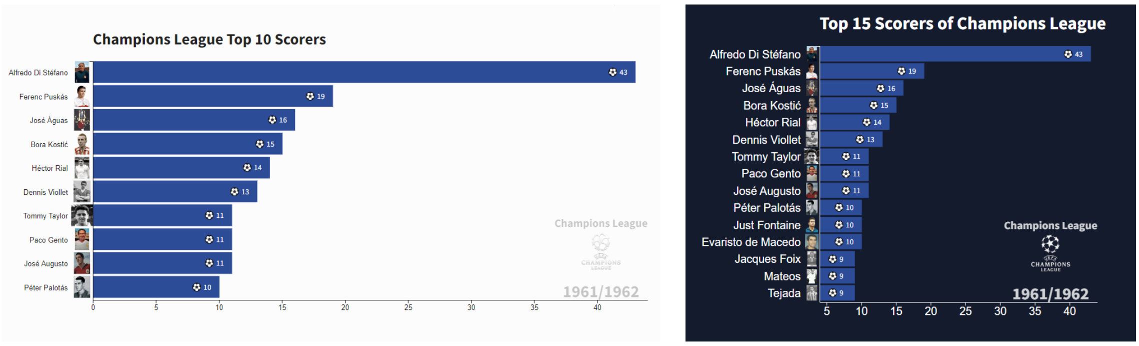

Figure 4.7: Original bar race chart VS Customized bar race chart.

4.1.2 Back end

One of the most critical requirements for the back end is the need to have a run-time environment that is able to use zerozero's API to retrieve the required data in order to build dynamic charts. NodeJS[8] was selected as the enabling technology due to its simplicity and popularity [15]. Moreover, there was already some previous experience with the technology, which is compatible with

[8] https://nodejs.org/en

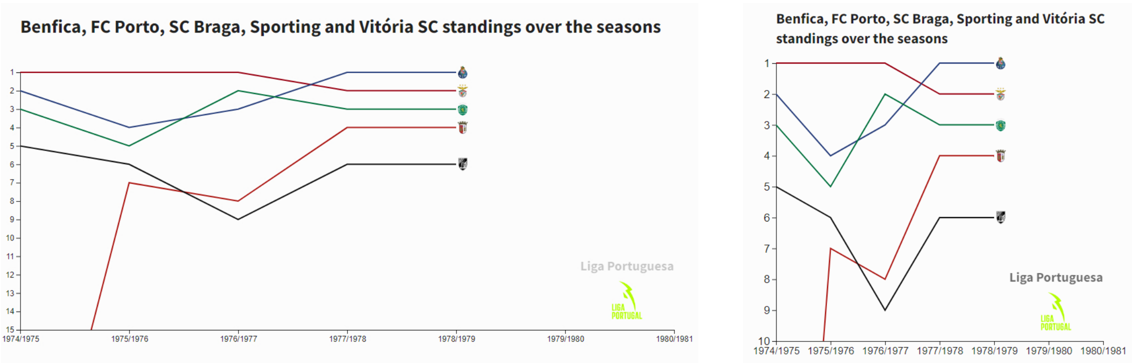

Figure 4.8: Original line race chart VS Customized line race chart.

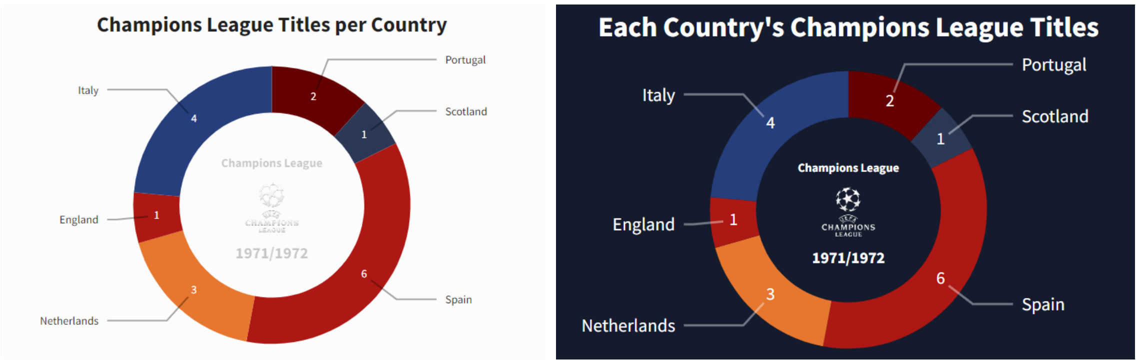

Figure 4.9: Original pie race chart VS Customized pie race chart.

ReactJS. Within the server, the Axios[9] library and the Express.js[10] framework are also enabling technologies. The Axios library requests the information from zerozero's API and provides it for use in the dynamic charts. The Express.js framework creates and handles the required endpoints to receive and respond to the requests made from the front end. The following endpoints, also presented in Table 4.1, were created to handle the features available to the users (for more details, consult Appendix A.1):

- data - accepts only GET requests;
  - /racingBarChart - gets the required information for the racing bar chart;
  - /racingLineChart - gets the required information for the racing line chart;
  - /racingPieChart - gets the required information for the racing pie chart.
- export - receives only POST requests;
  - / - renders and exports a race chart as a video file in MP4 format;
- props - receives GET and POST requests to get and save the dynamic chart's properties.
  - /barRaceChart - gets/saves the properties of the racing bar chart;

[9] https://axios-http.com
[10] https://expressjs.com

– /lineRaceChart - gets/saves the properties of the racing line chart;

– /pieRaceChart - gets/saves the properties of the racing pie chart.

Table 4.1: Back-end endpoints.

	Endpoint	**Request Type**	**Parameters/Payloads**
Data	/data/racingBarChart	GET	dateFormat id event
	/data/racingLineChart	GET	dateFormat id events
	/data/racingPieChart	GET	id
Export	/	POST	JSON object
Properties	/props/racingBarChart	GET	-
		POST	JSON object
	/props/racingLineChart	GET	-
		POST	JSON object
	/props/racingPieChart	GET	-
		POST	JSON object

An adaption of the Render D3 Video[11] tool was made regarding the video export feature. The tool uses the Puppeteer[12], a Node library that provides a high-level API to control headless Chrome or Chromium over the DevTools Protocol. In this headless Chrome, the time used for the animations' duration and update rate is updated according to the frame rate defined, in this case, 60 frames per second. Screenshots of the HTML element containing the chart are taken to get all the necessary video frames. After taking each screenshot, the time is updated according to the number of frames taken. When exporting the video, the URL always includes 'render-d3-video' in the query parameters. This way, it is possible to identify that the prototype is running through the Puppeteer and start controlling the time. After taking all the screenshots necessary, the video is rendered using the FFMPEG[13] framework.

[11] https://github.com/russellsamora/render-d3-video
[12] https://developer.chrome.com/docs/puppeteer
[13] https://ffmpeg.org

4.1.3 Data Flow

The data flow starts on page load. The required data is requested by the back-end server, which in turn requests the required data to the zerozero API. The API responds with the resulting data, which is processed before being sent to the front end. These operations of requesting the required data and processing it compose the previously presented **Data** and **Transform** stages of the conceptual data flow (Figure 3.1), as illustrated by Figure 4.10. After receiving the processed data, the front end maps it to the respective visual variables and finishes the data flow process by rendering the dynamic chart. These last two steps correspond to the **Visual Mapping** and **Render** stages of the conceptual data flow.

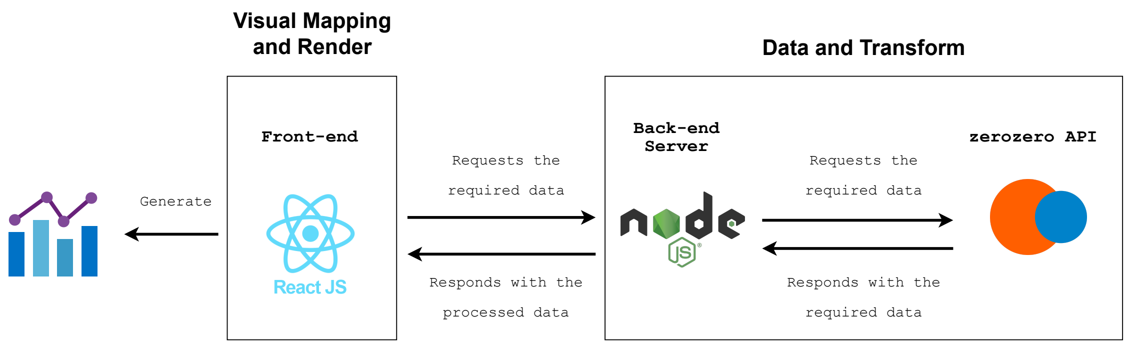

Figure 4.10: Prototype's data flow.

4.2 Summary

This chapter presented the prototype implementation based on the solution considered adequate for the problem described in Chapter 3. Section 4.1 detailed the prototype that implements the proposed solution, where an interface to create the dynamic charts is introduced and explained alongside its data flow, architecture, and technologies used. It is divided into two major JavaScript components: front end, which has as enabling technology ReactJS, and back end, which has as enabling technology NodeJS. The first component displays a user interface with a page for each dynamic chart type, while the second one requests and processes the required data to be used in the dynamic charts. The data flows from the back end to the front end until being presented to the users. Customizations are also available to complement the information presented. Finally, the dynamic charts can be exported as a file in MP4 format for use across several platforms, e.g., social media and websites.

Chapter 5

Experiments

This chapter presents in more detail the experimental procedure defined to evaluate the effectiveness and efficacy of dynamic charts. Section 5.1 explains the methodology defined for the experiments, while Section 5.2 describes the specific decisions applied to the methodology defined. Finally, Section 5.3 syntheses the major points of the experimental procedure developed.

5.1 Methodology

A detailed experiment procedure was developed to evaluate the effectiveness and efficacy of dynamic charts to validate or refute the proposed hypothesis that dynamic charts can tell a story while keeping viewer attention and engagement.

In the methodology developed, there are two groups of participants:

- ZOS employees, since these participants are responsible for creating the content for the zerozero website, being the first ones to judge the visualizations.
- zerozero community, since these participants are the final consumers of the created visualizations. This group has a more extensive diversity of participants with different backgrounds.

5.1.1 General Process Description

With respect to the test design, a dataset is first selected for each chart type (bar race chart, line race chart, and pie race chart). Each dataset addresses historical information regarding football statistics since it is the data type available through the zerozero's API. However, historical data from other different domains, e.g., finance, can be used [1]. After, an editor customizes each chart type created according to the data it represents. Moreover, it is necessary to analyze each chart type created to formulate the questions participants should answer during the test. The correct answer for each question is also defined. When running the experiment, participants visualize the chart race as videos and later answer the formulated questions. While doing so, metrics are retrieved and, together with the answers, are analyzed to draw conclusions. Participants belonging to the zerozero community are reached through email, while ZOS employees are reached through message in the company's workspace.

5.1.2 Procedure

For each participant, the following procedure is followed:

1. Explain the experiment's purpose, the participant rights, the procedure, and what data is viewed during the test;
2. Ask demographic information;
3. Present the dynamic chart and allow the participant to analyze it freely and for whichever time needed;
4. Present questions regarding the presented visualization. The participant can consult the dynamic chart to answer the questions at any point. For each question, the underlying logical reasoning is also asked;
5. Ask the participant to answer a form regarding its experience when using the visualization and its characteristics;
6. Ask the participant to fill out a SUS (System Usability Scale) form;
7. Submit the experiment form.

The test completion criteria is met when the participant fills all the mandatory fields in the experiment form and submit it. Regarding the required equipment, a computer, tablet, or smartphone with Internet access is required.

5.1.3 Collected Data

At the start of the experiment, demographic data is collected: age and gender. In addition, the knowledge in the information visualization domain and the knowledge in the football domain are also collected. Furthermore, the answer to each question is also retrieved alongside the respective logical reasoning. Moreover, the answers to the feedback questions and the SUS form are also retrieved.

5.1.4 Variables and Metrics

For the collected data mentioned above, the errors for each question and the errors in the respective logical reasoning are measured in terms of correct answers (either correct or incorrect) and valid logical reasoning (either valid or invalid). These measures help to understand if the information is correctly displayed and conveyed to the participants. While the errors for each question are objective and quantitative, the errors for logical reasoning are subjective and qualitative, being evaluated by the person responsible for the tests. The answers to the feedback questions and the SUS form are also subjective and qualitative, allowing to better understand the dynamic charts' usability and advantages.

5.1.5 Success Criteria

As for the success criteria, high success rates, which are the percentage of correct responses in the total responses per question, for the answers and logical reasoning evidence the dynamic charts' effectiveness and efficacy in telling a story and conveying information. High scores in the SUS and experience form demonstrate the usability and intuitiveness of the dynamic charts.

5.2 Test form

The experiment form is divided into three parts: pre-test questionnaire, questionnaire, and post-test questionnaire, presented in detail in the following sections.

5.2.1 Pre-test Questionnaire

In the form, an introduction is made to the participants so that they can be informed about what is asked. This introduction is composed of an informed consent form divided into the following parts:

- **Purpose of the study -** explains the purpose of the research and in what context this work is being done;
- **Study procedure -** explains the procedure described in Section 5.1.2. In addition, it is also specified that the data collected is stored in EU (European Union) space and the period during which the data is being held;
- **Duration -** gives an estimate of the duration of the test, synthesizing the two main parts: analyze the dynamic chart and answer the questions.
- **Voluntary participation -** informs the participant that the participation is purely voluntary, meaning the participant can withdraw during the study at any time, and any data collected until that point is deleted. It also informs that by moving to the next step in the form the participant is consenting with these terms.

In addition, the following information, alongside the demographic data previously mentioned in Section 5.1.3, is asked to better assess the participant population:

- Knowledge in the information visualization domain (5-point Likert Scale);
- Knowledge in the football domain (5-point Likert Scale).

This information helps to understand if there is any correlation between the previous knowledge that the participant has in the two domains and the results obtained on visualizing the dynamic charts. After moving to the following step, the dynamic chart in video format is presented so that the participant can freely analyze it and get familiar with it.

5.2.2 Questionnaire

The questionnaire part asks the participants questions about the video that was previously presented. Moreover, the video can be visualized whenever required to answer the questions. The datasets chosen for each chart type were selected based on the criteria consisting of the amount of information displayed and the corresponding time span. Datasets with large amounts of data distributed across a long period should be used so that the dynamic chart's ability to tell a story can be correctly evaluated. With this in mind, there are three datasets:

- **Champions Leagues Top 10 All-Time Scorers**, which is used for the bar race chart - this dataset was chosen due to the amount of data that it contains. It has information regarding the best scorers of the competition from its beginning in 1955 until 2023 (598 data points for each of the 4973 players present in the dataset). In addition, the top 10 list is continuously being updated, with periods with a few changes and others with several changes simultaneously. With these characteristics, this dataset is perfect for evaluating the bar race chart technique to present and convey large amounts of information and its evolution;
- **Standings of Benfica, FC Porto, SC Braga, Sporting, and Vitória SC over the seasons**, which is used for the line race chart - this dataset was chosen due to the extensive time interval in it. It has information about the five clubs from the start of the competition in 1934 until 2023 (89 data points for each of the 5 clubs represented). Moreover, the clubs' classification constantly changes from one season to another, with particular seasons where the clubs weren't funded or were in the second division. These constant changes lead to lines crossing each other. These characteristics make the dataset adequate to evaluate the line race chart's ability to present and convey large amounts of information while telling its story;
- **Champions League Titles per Country**, which is used in the pie race chart - this dataset contains a large amount of information regarding the distribution of Champions League titles per country. It has data from an extensive period, from 1955 until 2022 (68 data points for each of the 10 countries represented). Furthermore, it incorporates comparisons between the whole and a segment. These characteristics compose a proper dataset to evaluate the pie race chart's ability to tell a story while correctly presenting and conveying the information regarding the whole and a part.

A set of questions was elaborated for each chart type displaying data from a specific dataset. Furthermore, several time features are also evaluated as the underlying dimension in the dynamic charts is time itself:

- **Existence of a data element -** Does a data element exist at a specific time?
- **Temporal location -** When does a data element exist in time?
- **Time interval -** How long is the time span from the beginning to the end of the data element?
- **Sequence -** In what order do data elements appear?

Hence, for the bar race chart, the following questions were developed:

1. Which competition is being represented? - Tests the dynamic chart's ability to explain what is being presented;
2. How many seasons are presented in the chart? (Consider that a season is composed of year/year. Example: 2022/2023) - Tests the dynamic chart's ability to correctly represent and convey a time interval;
3. Who is the all-time top scorer of the competition given? - Tests the dynamic chart's ability to correctly represent and convey temporal location;
4. Who was the competition's all-time leading scorer by the end of the 2009/2010 season? - Tests the dynamic chart's ability to correctly represent and convey the existence of a data element;
5. At any point, Lionel Messi was the isolated all-time leading scorer of the competition represented in which season? If yes, during each season? (Consider that a season is composed of year/year. Example: 2022/2023. Also, consider that isolated means that no other player has the same goals as Lionel Messi) - Tests the dynamic chart's ability to correctly represent and convey the existence of a data element;
6. At any point, Pippo Inzaghi had more scored goals than Eusébio? - Tests the dynamic chart's ability to correctly represent and convey temporal location;
7. How many seasons did Ferenc Puskás stay in the top 10 of the best scorers of the competition? (Consider that a season is composed of year/year. Example: 2022/2023) - Tests the dynamic chart's ability to correctly represent and convey a time interval.

After the pilot tests, it became clear that clarifying what composes a season and what isolated means in the football context was necessary. In addition, question 5 was remade to remove the ambiguity presented on it. The original question had the following structure: "Lionel Messi was the all-time leading scorer in which season?".

Following are the line race char questions:

1. Which competition is being represented? - Tests the dynamic chart's ability to explain what is being presented;

2. How many seasons are presented in the chart? (Consider that a season is composed of year/year. Example: 2022/2023) - Tests the dynamic chart's ability to correctly represent and convey a time interval;

3. Who was the champion in the season of 1978/1979? (Consider that a champion is a team that is in the 1st place in the classification) - Tests the dynamic chart's ability to correctly represent and convey the existence of a data element;

4. What was the worst classification of Sporting in a season? - Tests the dynamic chart's ability to correctly represent and convey temporal location;

5. What was the final classification of the clubs represented in the season 2004/2005? (Answer in the following format: team X - 1º, team Y - 3º, etc.) - Tests the dynamic chart's ability to correctly represent and convey the existence of several data elements;

6. Who was the team with more consecutive championships? (Consider that a team wins a championship when it is in 1st place in the classification) - Tests the dynamic chart's ability to correctly represent and convey temporal location together with a time interval.

Again, the pilot tests showed the requirement to clarify what composes a season and when a team becomes champion or wins a championship.

Finally, the pie race chart questions are:

1. Which competition is being represented? - Tests the dynamic chart's ability to explain what is being presented;

2. How many seasons are presented in the chart? (Consider that a season is composed of year/year. Example: 2022/2023) - Tests the dynamic chart's ability to correctly represent and convey time intervals;

3. When did a Portuguese team win the competition represented for the first time? (Consider that a season is composed of year/year. Example: 2022/2023) - Tests the dynamic chart's ability to correctly represent and convey temporal location;

4. Which country has more titles in the competition represented? - Tests the dynamic chart's ability to correctly represent and convey temporal location;

5. To which country does the winner of the 2000/2001 seasons belong? - Tests the dynamic chart's ability to correctly represent and convey the existence of a data element;

6. A German team won for the first time a title in the competition represented before or after a Dutch team? - Tests the dynamic chart's ability to correctly represent and convey sequence between data elements.

Once again, the pilot tests revealed the need to clarify what composes a season.

5.2.3 Post-test Questionnaire

The post-test questionnaire comprises the experience form, containing the feedback statements, and the SUS form. The feedback questions were created to better assess the advantages of the dynamic charts using a 5-stage Likert scale. The feedback statements are the following:

1. The chart was able to tell a story;
2. Animations helped visualize the data presented over time;
3. Animations introduced dynamism to the chart;
4. Animations captured my attention;
5. Animations engaged me in the data presented;
6. Regarding the data presented, a static representation would achieve the same results as the animated representation;
7. Animations introduced unnecessary complexity to the visualization;
8. Animations make it harder to follow a single data point.

The SUS form is composed of 10 statements:

1. I think I would like to use this representation;
2. I found the representation unnecessarily complex;
3. I thought the representation was easy to use;
4. I think that I would need the support of a technical person to be able to use this representation;
5. I found the various functions in the representation were well integrated;
6. I thought there was too much inconsistency in this representation;
7. I would imagine that most people would learn to use this representation very quickly;
8. I found the representation very cumbersome to use;
9. I felt very confident using the representation;
10. I needed to learn a lot of things before I could get going with this representation.

5.2.4 Video Format

The charts could be provided to participants in two formats: video or incorporated in zerozero's website as HTML elements. Both allow the participant to reset the animations to the beginning. However, in video format, it is easier to jump to specific timestamps. Furthermore, making the charts available on the website as HTML elements and JavaScript scripts raises confidentiality concerns since stealing the underlying data used to render each chart becomes possible, as the zerozero API directly provides the required data. The information that ZOS has stored in its databases is its most valuable asset. Its priority is to limit the amount of information that can be scrapped from their website. Considering the previous, charts were made available in video format.

5.3 Summary

This chapter presented the experimental procedure followed to verify the effectiveness and efficacy of dynamic charts in presenting information in an attractive and more engaging format. Section 5.1 detailed the methodology adopted, explaining the group of participants, the general process, the procedure, the collected data, variables, metrics, and success criteria. In short, in the methodology developed, there are two groups of participants: ZOS employees and zerozero community. During the experimental procedure, demographic information is asked of the participants, and the respective dynamic chart is presented in video format. Afterward, questions are given regarding the video visualized with a SUS and experience form. The errors in each answer and respective logical reasoning are measured. The success criteria were defined as high success rates and high scores in the SUS and experience forms. Section 5.2 explained the decisions regarding the datasets used and the questions defined for each one. The datasets were chosen based on the amount of information and corresponding time span. Datasets with large amounts of data across extensive periods would better test the dynamic charts' ability to represent and convey the information. Since the underlying dimension in the dynamic charts is time, the set of questions elaborated was based on the following characteristics: the existence of a data element, temporal location, time interval, and sequence of data elements.

Chapter 6

Results

This chapter presents the results of the experiment described in Chapter 5. Section 6.1 describes the results of each type of dynamic chart (bar race chart, line race chart, and pie race chart), while Section 6.2 discusses those results, interprets them, and draws conclusions.

6.1 Results

The first section of the questionnaires served as an introduction to the current work, explaining the experiment procedure and what data would be retrieved, namely non-identifiable personal data, such as age, gender, and knowledge in the information visualization and football domain. Since the questionnaires for each dynamic chart type were independent, they were analyzed individually, with each one having its participant population. However, some overlap between participant populations may have occurred.

6.1.1 Bar Race Chart

In the bar race chart questionnaire, 40 subjects participated in the experiment and filled up the respective questionnaire forms. From those 40 responses, 4 were considered invalid since those participants answered the questions based on their memory of the dynamic chart, which was not the goal of this experiment. As previously said, the experiment was not to be based on memory. With this in mind, 36 responses were kept and considered valid for the purpose of the investigation. Of those 36 participants, 31 were male (86%), while the remaining 5 were female (14%). In addition, the participants were between 19 and 49 years old, with a substantial amount being 22-23 (44%).

To better understand the influence of knowledge in the information visualization and football domain, the correlation between this knowledge and the percentage of correct answers and correct logical reasoning from the participants was calculated. The participants evaluated their knowledge on a scale from 1 to 5. Although percentages are continuous variables, due to the limited number of questions, the percentage of correct answers and correct logical reasoning can only be within a set of values, making them discrete instead of truly continuous. With this in mind, these variables were treated as ordinal variables, making the Spearman rank correlation coefficient [16] the appropriate method for evaluating the correlation between the variables. Assuming that the correlation existed, two hypotheses were formulated:

- H_0 - The null hypothesis, which states that the variables are not correlated;
- H_1 - The alternative hypothesis, which states that the variables are correlated.

The results of the Spearman rank correlation are presented in Table 6.1.

Table 6.1: Spearman correlation results between the knowledge in information visualization and football and the percentage of correct answers and logical reasoning on the bar race chart.

		% Correct Answers	**% Correct Logical Reasoning**
InfoViz Knowledge	Correlation Coefficient	0.124	0.076
	P-value	0.471	0.661
Football Knowledge	Correlation Coefficient	0.054	-0.051
	P-value	0.753	0.769

Information visualization knowledge had a correlation coefficient of 0.124 with the percentage of correct answers and a correlation coefficient of 0.076 with the percentage of correct logical reasoning, indicating a very weak correlation between the variables. For an alpha value of 0.05 (α = 0.05), the null hypothesis can be rejected if the p-value is less than 0.05 ($p < 0.05$). In this case, with p-values of 0.471 and 0.661, the null hypothesis assumption cannot be rejected, meaning that the variables are not correlated. The football knowledge had a correlation coefficient of 0.054 and -0.051 with the percentage of correct answers and the percentage of correct logical reasoning, showing a very weak correlation between variables. For the same alpha value of 0.05, the p-values of 0.753 and 0.769 showed that the null hypothesis cannot be rejected, indicating that the variables are not correlated. In short, the participants' knowledge of the information visualization and football domain did not influence their responses in the questionnaire regarding the answers to the questions and the respective logical reasoning. The questions described in Section 5.2.2 were made to the participants. For each question, it was calculated the percentage of correct answers and correct logical reasoning to evaluate the effectiveness and efficacy of the bar race chart. The results are presented in Table 6.2.

Table 6.2: Percentage of correct answers and correct logical reasoning for each question in the bar race chart questionnaire.

	% Correct Answers	**% Correct Logical Reasoning**
Question 1	97.22	97.22
Question 2	47.22	88.89
Question 3	100.00	100.00
Question 4	100.00	94.44
Question 5	80.56	75.00
Question 6	91.67	88.89
Question 7	27.78	80.56

According to the table, most of the questions' percentages of correct answers and logical reasoning were high. However, in questions 2 and 7, the percentages significantly differed between correct answers and correct logical reasoning, evidencing the difference between the participants' answers and their reasoning. Question 5 had lower correct answers and logical reasoning percentages than the others.

In the experience form, feedback statements described in Section 5.2.3 were also presented to the participants and answered by them on a 5-stage Likert scale. These statements helped to better understand the effectiveness and efficacy of the bar race chart shown alongside its advantages and disadvantages. For each statement, the average, median, minimum, maximum, first quartile, third quartile, and standard deviation were calculated, as presented in Table 6.3 and in Figure 6.1.

Table 6.3: Average, median, minimum, maximum, 1st quartile, 3rd quartile and standard deviation of the feedback statements of the bar race chart questionnaire.

	Average	Median	Minimum	Maximum	1st Quartile	3rd Quartile	Standard Deviation
Statement 1	4.69	5.00	4.00	5.00	4.00	5.00	0.47
Statement 2	4.56	5.00	3.00	5.00	4.00	5.00	0.61
Statement 3	4.67	5.00	4.00	5.00	4.00	5.00	0.48
Statement 4	4.33	4.00	3.00	5.00	4.00	5.00	0.68
Statement 5	4.33	4.00	3.00	5.00	4.00	5.00	0.59
Statement 6	1.83	2.00	1.00	4.00	1.00	2.00	0.97
Statement 7	1.78	2.00	1.00	4.00	1.00	2.00	0.90
Statement 8	2.33	2.00	1.00	4.00	1.00	4.00	1.24

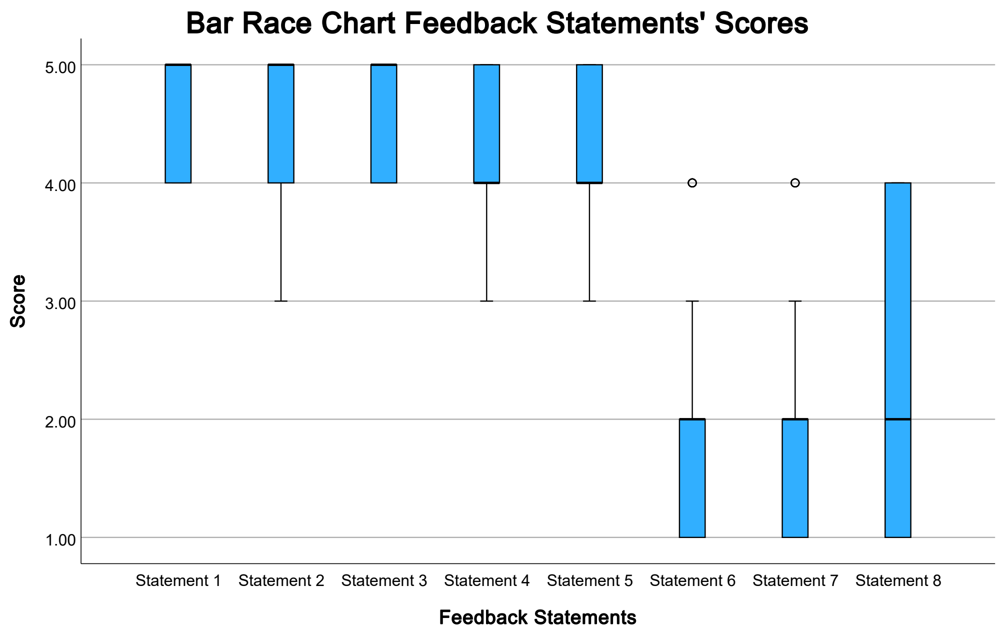

Figure 6.1: Boxplot of the feedback statements of the bar race chart questionnaire.

According to the table and the figure presented, two groups of feedback statements were identified: statements from 1 to 5 and statements from 6 to 8. The first group had higher averages and medians, while the second group got lower values in these two metrics. The first group also had their values in a smaller range, as shown by the minor standard deviations and the boxes' sizes in Figure 6.1. The second group had its values in a slightly bigger range since it had more considerable standard deviations. It also had outliers, particularly in feedback statements 6 and 7, as shown by Figure 6.1. In addition, in the same figure, it becomes evident that feedback statement 8 had a broader range of values.

6.1.2 Line Race Chart

For the line race chart questionnaire, 40 participants filled up and submitted the form. As in the previous Section 6.1.1, 4 responses were considered invalid, being kept 36 responses. Furthermore, the participant population followed the same characteristics as the participant population of the bar race chart questionnaire, with 29 males (81%) and 7 females (19%) within the same range of age.

The correlation between the knowledge in the information visualization and football domain and the percentage of correct answers and logical reasoning of the participants was calculated to understand if that knowledge influenced their responses. The participants evaluated their expertise on a scale from 1 to 5. Since the variables were the same as the ones in Section 6.1.1, the Spearman rank correlation coefficient was the appropriate method for evaluating the variables' correlation. The same two hypotheses were also formulated. The results of the Spearman rank correlation are presented in Table 6.4.

Table 6.4: Spearman correlation results between the knowledge in information visualization and football and the percentage of correct answers and logical reasoning on the line race chart.

		% Correct Answers	**% Correct Logical Reasoning**
InfoViz Knowledge	Correlation Coefficient	-0.012	-0.027
	P-value	0.946	0.875
Football Knowledge	Correlation Coefficient	-0.081	-0.110
	P-value	0.640	0.522

Considering the information visualization knowledge, it had a correlation coefficient of -0.012 with the percentage of correct answers and a correlation coefficient of -0.027 with the percentage of correct logical reasoning. These values indicate that the correlation is very weak. For an alpha value of 0.05 and the p-values of 0.946 and 0.875, the null hypothesis assumption cannot be rejected, meaning there is no correlation between the variables. Regarding the knowledge in the football domain, it had a correlation coefficient of -0.081 and -0.110 with the percentage of correct answers and correct logical reasoning, respectively. Once again, it shows a very weak correlation between the variables. For the same value alpha value of 0.05, the p-values of 0.640 and 0.522, respectively, indicate that the null hypothesis assumption cannot be rejected, meaning there is no correlation. To summarize, the participants' knowledge of the information visualization and football domain did not influence their answers to the presented questions and the respective logical reasoning. The questions described in Section 5.2.2 were made to the participants. Again, for each question, the percentage of correct answers and correct logical reasoning was calculated to evaluate the effectiveness and efficacy of the line race chart. The results are presented in Table 6.5.

Table 6.5: Percentage of correct answers and correct logical reasoning for each question in the line race chart questionnaire.

	% Correct Answers	% Correct Logical Reasoning
Question 1	97.22	97.22
Question 2	63.89	100.00
Question 3	100.00	97.22
Question 4	88.89	94.44
Question 5	80.56	83.33
Question 6	97.22	86.11

According to the presented table, the percentages of correct answers and correct logical reasoning were high. Nevertheless, in question 2, the percentages significantly differed between correct answers and correct logical reasoning, evidencing the difference between the participants' answers and their reasoning.

The experience form presented to the bar race chart questionnaire participants was also presented to the participants of the line race chart. Again, the subjects answered the feedback statements on a 5-stage Likert scale and had the goal of helping understand the effectiveness, efficacy, advantages, and disadvantages of this type of dynamic chart. The average, median, minimum, maximum, first quartile, third quartile, and standard deviation were calculated, as presented in Table 6.6 and Figure 6.2.

Table 6.6: Average, median, minimum, maximum, 1st quartile, 3rd quartile and standard deviation of the feedback statements of the line race chart questionnaire.

	Average	Median	Minimum	Maximum	1st Quartile	3rd Quartile	Standard Deviation
Statement 1	4.56	5.00	2.00	5.00	4.00	5.00	0.69
Statement 2	4.31	5.00	1.00	5.00	4.00	5.00	0.95
Statement 3	4.56	5.00	2.00	5.00	4.00	5.00	0.65
Statement 4	4.39	4.50	2.00	5.00	4.00	5.00	0.77
Statement 5	4.17	4.00	1.00	5.00	4.00	5.00	0.97
Statement 6	2.14	2.00	1.00	5.00	1.00	3.00	1.15
Statement 7	2.08	2.00	1.00	5.00	1.00	2.00	1.08
Statement 8	2.44	2.00	1.00	5.00	1.00	4.00	1.30

According to the table and figure presented, it was possible to identify two main groups regarding the feedback statements: the first composed of statements 1 to 5, and the second composed of statements 6 to 8. The first group had a higher average, with statements 1 and 3 having the highest average of 4.56. The second group had a lower average, with statement 7 having the lowest average of 2.08. Regarding the median, the first group also had higher values than the second group. Despite the difference in the average and median, every statement had a maximum of 5.00 and a minimum of 1.00 or 2.00, which led to the occurrence of outliers represented in Figure 6.2. In the same figure, it is also visible that the first group of feedback statements had their values in a smaller range than the second group. This is also visible in Table 6.6 through the lower standard

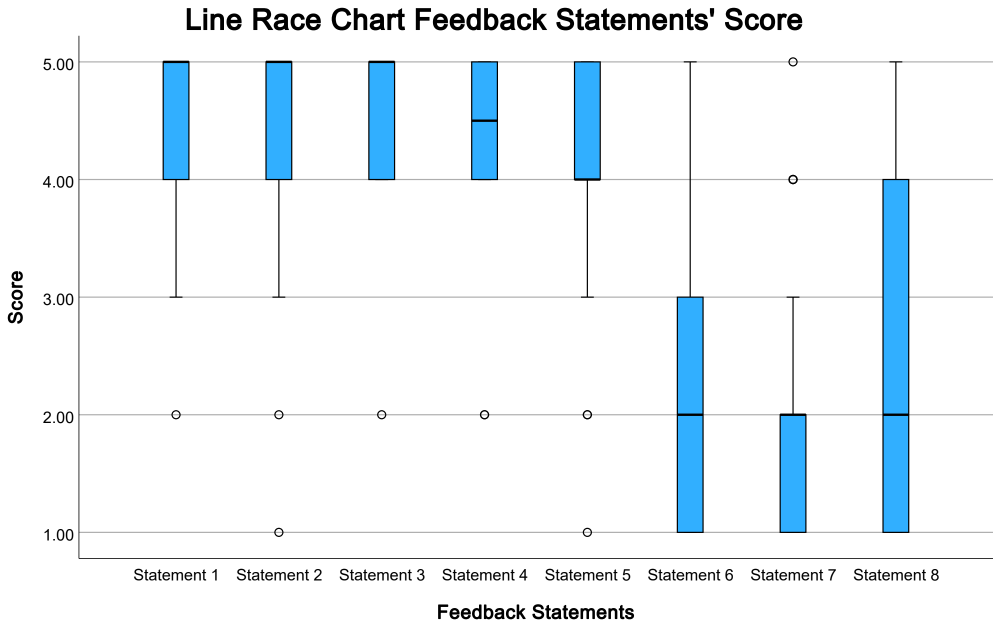

Figure 6.2: Boxplot of the feedback statements of the line race chart questionnaire.

deviations in the first group. Statement 8 had a higher range of values, from 1.00 to 5.00, and a higher standard deviation, 1.30.

6.1.3 Pie Race Chart

In the pie race chart questionnaire, 40 participants filled up and submitted the form. From those 40 responses, 2 were considered invalid. With this in mind, 38 responses were kept. The participant population shared the same characteristics as the participant population of the other two dynamic charts questionnaires, with 31 males (82%) and 7 females (18%) aged between 20 and 48 years old.

Once again, the correlation between the knowledge in the information visualization and football domain and the percentage of correct answers and logical reasoning of the participants was calculated to better understand its influence on the participants' responses. The subjects evaluated their expertise on a scale from 1 to 5. Considering that the variables were the same as the ones in Section 6.1.1 and 6.1.2, the Spearman rank correlation coefficient was the appropriate method for evaluating the variables' correlation. The same two hypotheses were also formulated. The results of the Spearman correlation are presented in Table 6.7.

The knowledge in the information visualization domain had a correlation coefficient of 0.067 and -0.116 with the percentage of correct answers and correct logical reasoning, respectively. These values indicate a very weak correlation. For an alpha value of 0.05, the p-values of 0.689 and 0.489 indicate that the null hypothesis assumption cannot be rejected, meaning that the variables are not correlated. The knowledge in the football domain had a correlation coefficient of

Table 6.7: Spearman correlation results between the knowledge in information visualization and football and the percentage of correct answers and logical reasoning on the pie race chart.

		% Correct Answers	% Correct Logical Reasoning
InfoViz Knowledge	Correlation Coefficient	0.067	-0.116
	P-value	0.689	0.489
Football Knowledge	Correlation Coefficient	-0.052	-0.276
	P-value	0.754	0.093

-0.052 and -0.276 with the percentage of correct answers and correct logical reasoning, respectively. These values indicate that the knowledge in the football domain has a very weak correlation with the percentage of correct answers and a weak correlation with the percentage of correct logical reasoning. For the same alpha value of 0.05, the p-values of 0.754 and 0.093 show that the null hypothesis assumption cannot be rejected, meaning that the variables are not correlated. In essence, the participants' knowledge of the information visualization and football domain did not influence their answers to the questions and the respective logical reasoning. The questions described in Section 5.2.2 were made to the participants. The percentage of correct answers and correct logical reasoning was again calculated to evaluate the effectiveness and efficacy of the pie race chart. The results are presented in Table 6.8.

Table 6.8: Percentage of correct answers and correct logical reasoning for each question in the pie race chart questionnaire.

	% Correct Answers	% Correct Logical Reasoning
Question 1	100.00	100.00
Question 2	57.89	94.74
Question 3	100.00	94.74
Question 4	100.00	97.37
Question 5	76.32	71.05
Question 6	97.37	97.37

As shown by the table presented, the percentages of correct answers and correct logical reasoning were high in the majority of the questions. However, in question 2, the percentages significantly differed between correct answers and correct logical reasoning, evidencing the difference between the participants' answers and their reasoning. Furthermore, compared with the other questions, question 5 was highlighted due to its lower correct answers and logical reasoning percentages.

Once more, the same feedback questions presented to the bar and line race chart questionnaire participants were also presented to the participants of the pie race chart questionnaire. The participants answered on a 5-stage Likert scale. For each feedback statement, the average, the median, the minimum, the maximum, the 1st quartile, the third quartile, and the standard deviation were calculated. The results are presented in Table 6.9 and Figure 6.3.

As in the bar race chart and line race chart, two groups of feedback statements were identified: the first, from statements 1 to 5, and the second, from statements 6 to 8. The statements in the first group had higher averages, with statement 1 standing out with an average of 4.47. Statements in

Table 6.9: Average, median, minimum, maximum, 1st quartile, 3rd quartile and standard deviation of the feedback statements of the pie race chart questionnaire.

	Average	Median	Minimum	Maximum	1st Quartile	3rd Quartile	Standard Deviation
Statement 1	4.47	5.00	2.00	5.00	4.00	5.00	0.73
Statement 2	4.11	4.00	1.00	5.00	4.00	5.00	1.03
Statement 3	4.39	4.00	2.00	5.00	4.00	5.00	0.68
Statement 4	4.29	4.00	1.00	5.00	4.00	5.00	0.90
Statement 5	4.05	4.00	1.00	5.00	4.00	5.00	0.93
Statement 6	2.29	2.00	1.00	5.00	2.00	3.00	0.93
Statement 7	2.34	2.00	1.00	4.00	2.00	3.00	0.97
Statement 8	2.50	2.00	1.00	5.00	2.00	3.00	1.20

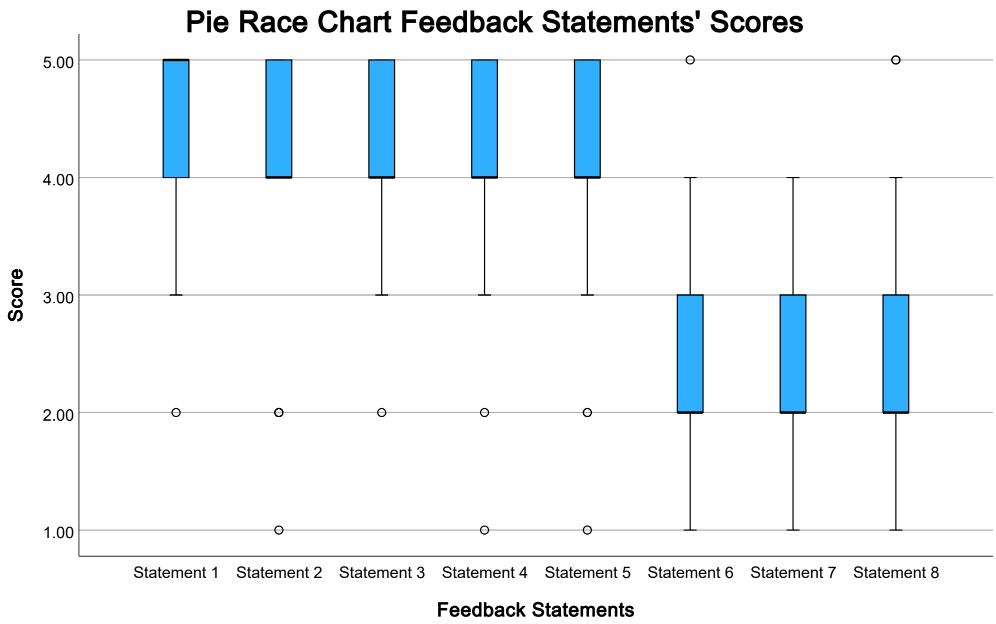

Figure 6.3: Boxplot of the feedback statements of the pie race chart questionnaire.

the second group had lower averages, with statement 6 having the lowest average. Regarding the median, the first group also had statements with higher values, whereas statement 1 had the highest value again. The second group had statements with the lowest median value of 2.00. Despite the differences in the average and median, both groups of statements had similar minimums of 1.00 or 2.00, and maximums of 5.00 or 4.00, in the case of statement 7. These minimums and maximums led to the outliers represented in Figure 6.3 and to the slightly higher standard deviations in Table 6.9, where the statement 1 and 8 stand out with 1.03 and 1.20, respectively.

6.1.4 System Usability Scale (SUS)

In the final section of each questionnaire, a SUS form was presented to the users so that they could evaluate the usability of the respective dynamic chart. For each response, it was calculated the SUS score by summing the score contributions from each item, as Brooke states in his work [7]. According to the author, the score contribution of odd questions is calculated by subtracting 1 from the scale contribution. For the score contribution of even questions, their score is calculated by subtracting the scale contribution to 5. In the end, the sum of the scores is multiplied by 2.5 to be in a range from 0 to 100. After calculating the score of each response, the average, median, minimum, maximum, first quartile, third quartile, and standard deviation were calculated for each dynamic chart type. The results are presented in Table 6.10 and Figure 6.4.

Table 6.10: Average, median, minimum, maximum, 1st quartile, 3rd quartile and standard deviation of the SUS score of each dynamic chart type.

	Average	**Median**	**Minimum**	**Maximum**	**1st Quartile**	**3rd Quartile**	**Standard Deviation**
Bar Race Chart	80.21	83.75	45.00	100.00	72.50	92.50	16.90
Line Race Chart	75.63	80.00	25.00	100.00	65.00	93.13	21.52
Pie Race Chart	75.99	76.25	35.00	100.00	63.13	87.50	16.45

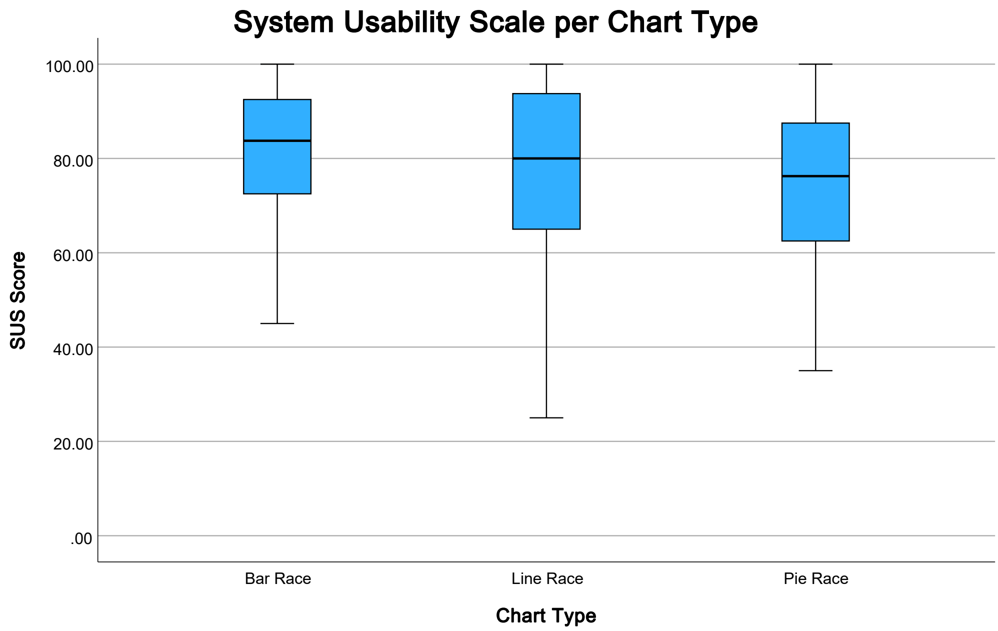

Figure 6.4: Boxplot of the SUS score of each dynamic chart type.

The bar race chart was the chart type with the highest average and median SUS score values, with 80.21 and 83.75, respectively. In contrast, the line race chart had the lowest average of 75.63, while the pie race chart had the lowest median of 76.25. The line race chart also had the lowest minimum, with 25.00. The three types of dynamic charts had the same maximum of 100.00. The bar race chart and the pie race chart had SUS scores in similar values, as shown in Figure 6.4 and by the standard deviation values of 16.90 and 16.45, respectively. The line race chart had a wider range of SUS scores, as illustrated in the same figure and by the standard deviation of 21.52.

6.2 Discussion

Regarding the bar race chart, the high percentages on both correct answers and logical reasoning of questions 3, 4, and 6 indicate that the dynamic chart correctly represented and conveyed the information to the viewer when considering the existence of a data element and temporal location. Question 5 had lower percentages indicating a margin for improvement to ensure that every data element is correctly represented. The results of question 1 show that what was being represented was evident to the viewers. The questions 2 and 7 lower percentages of correct answers can be justified by the fact that the participants had difficulties when counting seasons: the majority of them had correct logical reasoning, but their answers were always off by one. For instance, from 2021/2022 to 2022/2023, two seasons have passed; however, 2022-2021 = 1 or 2023-2022 = 1, which lefts one season off the counting. This math error led to the lower percentages of correct answers. Even so, the high percentages of logical reasoning evidence that the time intervals were also correctly displayed and presented in the bar race chart. Considering the feedback statements in the experience form, the higher the value given by the participants in statements 1 to 5, the better, while in statements 6 to 8, the lower the value, the better. The high averages of statements 1 to 5 indicate that the bar race chart was able to tell a story and present the data over time. In addition, it also shows that the animations introduced dynamism while capturing the viewer's attention and engaging it in the story. The high medians, minimums, and maximums and the low standard deviations evidence that most participants had similar experiences with the dynamic chart. The low averages of statements 6 to 8 show that a static representation couldn't achieve the same experience. Furthermore, it also reveals that the animations did not introduce complexity to the representation and did not make it harder to follow a single data point. The low medians and minimums also emphasize these characteristics. In contrast, the high maximums and higher standard deviations indicate a wide range of experiences within the participant population and that these features can be improved.

As for the line race chart, the high percentages in questions 3, 4, and 5 reveal that this dynamic chart also correctly represented the information regarding the existence of a data element and time interval. Once again, question 1 expresses the capability of the line race chart to indicate what is being represented. Question 5 had higher percentages but lower in comparison to the previously mentioned questions, highlighting the improvements that can be made to correctly represent every data point in time. Question 2 had a low percentage of correct answers due to the math error when counting the seasons. Despite that math error, the logical reasoning of all the participants was correct, emphasizing the capability of correctly representing time intervals. Once more, in the feedback statements, the higher the value given by the participants in statements 1 to 5, the better, while in statements 6 to 8, the lower the value, the better. The high averages of statements 1 to 5 evidence the ability of the line race chart to tell a story and that its animations helped to visualize the data over time while introducing dynamism, capturing the viewer's attention, and engaging it. The high medians also evidence this ability. However, the lower minimums, alongside the standard deviation and the outliers, show the divergence of experiences of the participants and the margin of progress to improve the dynamic visualization. The low averages and median of statements 6 to 8 reveal that the same experience could not be achieved with a static representation and that the animations did not introduce unnecessary complexity or make it harder to follow a single data point. Nevertheless, the high maximums, along with the standard deviations and outliers, exhibit the divergence of participants' experiences again, indicating improvements can be made.

Considering the pie race chart, the high percentages of questions 3, 4, and 6 indicate the correct conveyance of information to the viewer regarding temporal location, time interval, and sequence. In addition, question 1 percentages of 100% highlight the ability to present to the viewer what is being shown so that it can quickly understand it. Question 2 had a low percentage of correct answers caused by the math error previously mentioned when counting the seasons. Once more, the high percentage of correct logical reasoning shows that the viewers were able to correctly visualize and understand the time interval, despite the math error. Question 5 had lower percentages than the other questions (not considering question 2), which expresses the difficulty that some participants had when trying to find the existence of a data element. In this dynamic chart, there is a large margin of improvement in correctly conveying the information to the viewers. Once again, the higher the value given by the participants in feedback statements 1 to 5, the better, while in feedback statements 6 to 8, the lower the value, the better. The high averages and medians of questions 1 to 5 exhibit the ability of the pie race chart to tell a story and use animations to present data over time while attracting and engaging the viewer. Nonetheless, the low minimums, alongside the standard deviations and the outliers, show the different experiences that the participants had. In statements 6 to 8, the low averages and medians indicate that the same experience could not be achieved with a static representation and that animations did not introduce unnecessary complexity or make it harder to follow a single data point. Despite that, the high maximums, in combination with higher standard deviations and the presence of outliers, show again the different experiences that the participants had when using the dynamic chart.

The results obtained using the Spearman rank correlation evidenced that the participants' knowledge in the information visualization and football domain had no influence on their results in any dynamic chart. With this in mind, it is possible to conclude that no previous knowledge is needed to visualize and understand the information presented in the dynamic charts tested. As for the usability of the three dynamic chart types, the bar race chart had better results with the highest average, median, minimum, and maximum. The second best was the pie race chart, and the third was the line race chart. According to the scale defined by Bangor *et al.* [4], the three dynamic chart types had good usability with a margin for improvement since they had averages between 75 and 80 and medians between 76 and 84. Moreover, the high standard deviations of each dynamic chart type indicate a large spectrum of participants' usability experiences, supporting that improvements can be made in each one. The pie race chart, surprisingly, had better results than the line race chart despite being a less common type of dynamic chart.

In short, the bar, line, and pie race chart can correctly represent and convey information to the viewers regarding the existence of a data element, temporal location, time interval, and sequence. Moreover, they can also tell a story and represent the data over time while being attractive and engaging for the viewer. They also have good usability. Nevertheless, some improvements can be made to boost their usability and ability to represent all the necessary information.

6.3 Summary

This chapter presented the results of the experiment conducted. Section 6.1 described the results for the bar, line, and pie race charts, introducing the metrics measured and calculated to evaluate the effectiveness and efficacy of the dynamic charts to convey information. Then, in Section 6.2, the results obtained were discussed, and conclusions were drawn from them. In short, the three types of dynamic charts could correctly represent and convey the information to the viewers while having good usability. However, improvements can be made to increase their usability alongside their ability to correctly convey every data point.

Chapter 7

Conclusions

7.1 Final Remarks

In a time where data is growing exponentially, it becomes mandatory to present it, especially time-evolving data, in a simple and attractive format to the readers. Hence, it matches their fast pace lives. With this in mind, it is essential to correctly present the time-evolving data in a format capable of telling a story while engaging the viewer and capturing its attention. Nevertheless, the data needs to be presented without overwhelming the viewer. In the ZOS case study, there is a lot of historical and statistical data on sports up to the present that needs to be displayed in a way that attracts users and keeps them engaged without distorting the original data. Two visual representation techniques emerged to represent this data type: static representations, which are simpler and more objective, and dynamic representations, which clearly express the evolution over time.

The literature review on storytelling and dynamic visualizations explored the fundamental concepts of information visualization, with prominence on its pipelines and taxonomies, according to the data type represented. Furthermore, the research revealed the two leading roles of information visualization: visual analytics, which consists of using visualizations to analyze data, and presentation, which focuses on communicating information to the audience. Moreover, it also revealed that storytelling is a specific domain inside the presentation, where it is crucial the correctly tell a story of the events that occurred so that no information is lost and the audience becomes engaged. It takes advantage of the fact that stories have existed since the beginning of humanity. In addition, animations can be used in storytelling to emphasize the story's engagement and add attractiveness to it. Nevertheless, they must be used carefully; otherwise, it becomes confusing and overwhelming for the viewer. The rate of change can be adjusted to address this problem so that the velocity of the narrative complies with the amount of information displayed, avoiding the loss of details or the viewer's attention.

After carefully analyzing the problem statement and what a possible solution should accomplish, the solution requirements were defined, which consisted of carefully studying dynamic infographics, *e.g.*, bar race chart, line race chart, and pie race chart, that can be customized according to the information presented. Two groups of solution end-users were identified: the editors that create the visualizations, being the first ones to judge the final result, and the users that consume information through the visualizations on several platforms, *e.g.*, social media and websites. With this in mind, a solution is proposed that combines bar, line, and pie charts with animations to create the previously mentioned dynamic charts. This way, the best techniques to represent temporal data visually are combined with animations to introduce dynamism and keep the viewers engaged and attracted without distorting the original data. A dynamic visual narrative can be created, and customization is encompassed in order to complement the information presented and adjust the narrative velocity.

A prototype that implements the solution was developed based on JavaScript. It can be divided into two components: the back end in NodeJS and the front end in ReactJS. The prototype has a simple interface that allows editors to create and customize three types of dynamic charts: bar, line, and pie races. Moreover, it allows exporting the created charts in video format to be used across several platforms.

An experimental procedure was developed so that the proposed solution could be tested. In this procedure, the participants would measure their knowledge in the information visualization and football domain on a scale from 1 to 5 and answer some questions about the dynamic chart presented, explaining their logical reasoning. Then, the participants would fill feedback statements in a 5-stage Likert scale and a System Usability Scale (SUS) form. The experiment results revealed that the three types of dynamic charts could correctly represent and convey information regarding the existence of a data element, temporal location, time interval, and sequence without needing previous knowledge. Despite that, improvements can be made to correctly represent and convey every data point to the viewer. Despite the wide range of experiences of the participants, the three types of dynamic charts were capable of telling an attractive and engaging story by using animations without creating unnecessary complexity or confusion. Surprisingly, in contrast with the literature review, the animations didn't make it harder to follow a single data point. Finally, the three dynamic charts reached a good usability score, with scores between 75 and 80, being the bar race chart the best, followed by the pie and line race charts. Once more, improvements can be made to improve their usability.

With the results obtained, it is now possible to answer the proposed research questions:

RQ1: What makes a visual representation a good storytelling artifact? To be a good storytelling artifact, a visual representation must be simple, attractive, and engaging. The several events that compose the story need to be clearly expressed and represented so the viewer can understand every part without missing any information. Animations are an excellent option to capture the viewer's attention by introducing dynamism in the narrative. However, it is important to adjust their velocity to avoid being a distracting element.

RQ2: What visualization techniques can be considered more effective for presenting time-evolving data? Three types of dynamic visualization techniques can effectively represent time-evolving data: bar race chart, line race chart, and pie race chart. The high percentages of correct answers and correct logical reasoning evidence the effectiveness of these techniques. The results also proved that static representations could not achieve the same effect as these dynamic representations when considering time-evolving data. Regarding usability, the bar race chart proved to be the best visualization technique, followed by the pie and line race charts.

RQ3: Does animation engage the viewers without overwhelming them? The conducted literature review revealed that animations capture the viewers' attention at the risk of overwhelming them if too much change occurs in the representation. The results of the conducted experience proved that animations engaged the viewers without overwhelming them or creating unnecessary confusion. The rate of change and narrative velocity must be selected and adapted carefully according to the data that needs to be presented.

After answering the research questions, it is now possible to validate or refute the hypothesis *"Dynamic infographics techniques, such as bar race, line race, and pie race, are suitable to present historical information as animation would achieve the necessary dynamism while adjusting the rate of change would help control the speed of the narrative display."* The three types of dynamic charts presented the historical information in an attractive and engaging format while keeping the viewer's attention and representing the evolution of the data over time. Animations were used to achieve these results, introducing dynamism into the visualization. Furthermore, they also proved to effectively convey the information to the audience.

In conclusion, the proposed hypothesis is validated.

7.2 Future Work

Even though the research questions were answered and the proposed hypothesis was validated, some points still need to be further investigated, explored, and tested. The work explored the potential of dynamic infographics to correctly present information to the viewers while capturing their attention and expressing the changes that the data suffered over time. The three visualization techniques explored were tested regarding their effectiveness and efficacy in conveying the information and usability. Unfortunately, each technique only had participant populations of 40 subjects, where some overlap may have occurred, meaning some participants may have contributed to more than one dynamic chart type questionnaire. A larger sample of participants would increase the results' accuracy and confirm the previously drawn conclusions.

The three visualization techniques were only compared regarding their usability. Nevertheless, comparing their efficiency and efficacy would be a significant addition so that conclusions could be drawn concerning the best visualization technique of the three proposed. A shared dataset, alongside the respective questions, must be built to accomplish that. In addition, the time to answer each question also needs to be measured. This way, it is ensured that each visualization produced has the same variables, allowing to establish comparisons.

The developed work creates a new question: "How does the rate of change influence the effectiveness and efficacy of a dynamic visualization". Although the rate of change was part of the proposed solution and implemented on the prototype, it was not tested. With this in mind, further investigation of the influence of the rate of change on the viewers can be conducted. Moreover, the limits of the rate of change can also be explored: at each velocity value, the viewer's attention gets lost, or the visualization becomes messy and complex. Exploring the balance between the rate of change and the amount of information presented is also crucial.

Finally, the framework developed is a starting point for a potential tool to help editors easily create dynamic infographics that can be used across several domains, from football to finance to world history.

References

[1] Investopedia. https://www.investopedia.com. (Accessed on 06/06/2023).

[2] ZOS. https://www.zos.pt. (Accessed on 01/05/2023).

[3] S.A. Arboleda and A. Dewan. Unveiling storytelling and visualization of data. *Proceedings of the 14th SC@ RUG*, 2017:38–42, 2016.

[4] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, May 2009.

[5] L. Battle and J. Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38(3):145–159, 2019.

[6] P. Baudisch *et al.* Phosphor: Explaining transitions in the user interface using afterglow effects. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, page 169–178, New York, NY, USA, 2006. Association for Computing Machinery.

[7] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, Nov 1995.

[8] S.K. Card and J. Mackinlay. The structure of the information visualization design space. In *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, pages 92–99, 1997.

[9] S.K. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualisation: using vision to think*. Morgan Kaufmann Publishers Inc., 1999.

[10] P. Cavanagh and G.A. Alvarez. Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7):349–354, 2005.

[11] C. Chen. Information visualization. *WIREs Computational Statistics*, 2(4):387–403, 2010.

[12] E.H. Chi. A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pages 69–75, 2000.

[13] L. Chittaro. Visualizing information on mobile devices. *Computer*, 39(3):40–45, 2006.

[14] W. Cui. Visual analytics: A comprehensive overview. *IEEE Access*, 7:81555–81573, 2019.

[15] A. Deitcher. Simplicity and performance: Javascript on the server. *Linux J.*, 2011(204), Apr 2011.

[16] Y. Dodge. Spearman rank correlation coefficient. In *The Concise Encyclopedia of Statistics*, pages 502–505. Springer New York, New York, NY, 2008.

[17] A. Figueiras. How to tell stories using visualization. In Ebad Banissi, Mark W. McK. Bannatyne, Francis T. Marchese, Muhammad Sarfraz, Anna Ursyn, Gilles Venturini, Theodor G. Wyeld, Urska Cvek, Marjan Trutschl, Georges G. Grinstein, Vladimir Geroimenko, Sarah Kenderdine, and Fatma Bouali, editors, *18th International Conference on Information Visualisation, IV 2014, Paris, France, July 16-18, 2014*, page 18. IEEE Computer Society, 2014.

[18] M. Friendly and D.J. Denis. Milestones in the history of thematic cartography, statistical graphics, and data visualization. *URL* `http://www.datavis.ca/milestones`, 32:13, 2001.

[19] B. Fry. *Visualizing Data*. O'Reilly Media, 2008.

[20] C.G. Healey, K.S. Booth, and J.T. Enns. High-speed visual estimation using preattentive processing. *ACM Trans. Comput.-Hum. Interact.*, 3(2):107–135, Jun 1996.

[21] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007.

[22] W. Jiang, J. Wu, G. Sun, Y. Ouyang, J. Li, and S. Zhou. A survey of time series data visualization methods. *Journal of Quantum Computing*, 2(2):105, 2020.

[23] D.A. Keim, F. Mansmann, D. Oelke, and H. Ziegler. Visual analytics: Combining automated discovery with interactive visualizations. In Jean-François Jean-Fran, M. R. Berthold, and T. Horváth, editors, *Discovery Science*, pages 2–14, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[24] D.A. Keim, F. Mansmann, and J. Thomas. Visual analytics: How much visualization and how much analytics? *SIGKDD Explor. Newsl.*, 11(2):5–8, May 2010.

[25] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[26] S. Kriglstein, M. Pohl, and C. Stachl. Animation for time-oriented data: An overview of empirical research. In *2012 16th International Conference on Information Visualisation*, pages 30–35, 2012.

[27] K.-L. Ma, I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis. Scientific storytelling using visualization. *IEEE Computer Graphics and Applications*, 32(1):12–19, 2012.

[28] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, Apr 1986.

[29] J. Mackinlay. Opportunities for information visualization. *IEEE Computer Graphics and Applications*, 20(1):22–23, 2000.

[30] J.I. Maletic, A. Marcus, and M.L. Collard. A task oriented view of software visualization. In *Proceedings First International Workshop on Visualizing Software for Understanding and Analysis*, pages 32–40, 2002.

[31] A. Marcus, L. Feng, and J.I. Maletic. 3d representations for software visualization. In *Proceedings of the 2003 ACM Symposium on Software Visualization*, SoftVis '03, page 27–ff, New York, NY, USA, 2003. Association for Computing Machinery.

[32] A.U. Mentsiev, M.V. Engel, A.M. Tsamaev, M.V. Abubakarov, and R.S.-E. Yushaeva. The concept of digitalization and its impact on the modern economy. In *Proceedings of the International Scientific Conference "Far East Con" (ISCFEC 2020)*, pages 2960–2964. Atlantis Press, 2020.

[33] W. Muller and H. Schumann. Visualization methods for time-dependent data - an overview. In *Proceedings of the 2003 Winter Simulation Conference, 2003.*, volume 1, pages 737–745 Vol.1, 2003.

[34] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih. Big data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4):431–448, 2018.

[35] S.E. Palmer. *Vision Science: Photons to Phenomenology*. A Bradford Book. MIT Press, 1999.

[36] C. Perin, R. Vuillemot, C.D. Stolper, J.T. Stasko, J. Wood, and S. Carpendale. State of the art of sports data visualization. *Computer Graphics Forum*, 37(3):663–686, 2018.

[37] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, page 109–116, New York, NY, USA, 2004. Association for Computing Machinery.

[38] Z.W. Pylyshyn. *Things and Places: How the Mind Connects with the World*. Bradford books. Mass., 2007.

[39] R.A. Rensink, J.K. O'Regan, and J.J. Clark. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5):368–373, 1997.

[40] T. Rhyne, M. Tory, T. Munzner, M. Ward, C. Johnson, and D.H. Laidlaw. Information and scientific visualization: separate but equal or happy together at last. In *IEEE Visualization, 2003. VIS 2003.*, pages 611–614, 2003.

[41] G. Robertson, K. Cameron, M. Czerwinski, and D. Robbins. Animated visualization of multiple intersecting hierarchies. *Information Visualization*, 1(1):50–65, 2002.

[42] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[43] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In B. B. Bederson and B. Shneiderman, editors, *The Craft of Information Visualization*, Interactive Technologies, pages 364–371. Morgan Kaufmann, San Francisco, 2003.

[44] I. Spence. William playfair and the psychology of graphs. In *Proceedings of the American Statistical Association*, pages 2426–2436, Aug 2006.

[45] R. Spence. Presentation. In *Information Visualization: An Introduction*, pages 111–171. Springer International Publishing, Cham, 2014.

[46] J. Steele and N. Iliinsky. *Beautiful Visualization: Looking at Data through the Eyes of Experts*. Theory in practice series. O'Reilly Media, 2010.

[47] J.J. Thomas and K.A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[48] M. Tory and T. Moller. Rethinking visualization: A high-level taxonomy. In *IEEE Symposium on Information Visualization*, pages 151–158, 2004.

[49] E.R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 1992.

[50] B. Tversky, J.B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.

[51] L. Tweedie. Characterizing interactive externalizations. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, page 375–382, New York, NY, USA, 1997. Association for Computing Machinery.

[52] A. Van Dam, A.S. Forsberg, D.H. Laidlaw, J.J. LaViola, and R.M. Simpson. Immersive VR for scientific visualization: a progress report. *IEEE Computer Graphics and Applications*, 20(6):26–52, 2000.

[53] Z. Wang, S. Wang, M. Farinella, D. Murray-Rust, N. Henry-Riche, and B. Bach. Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[54] C. Ware. *Information Visualization: Perception for Design*. Interactive Technologies. Elsevier Science, 2019.

[55] Q. Zhi, A. Ottley, and R. Metoyer. Linking and layout: Exploring the integration of text and visualization in storytelling. *Computer Graphics Forum*, 38(3):675–685, 2019.

[56] D.E. Zongker and D.H. Salesin. On Creating Animated Presentations. In D. Breen and M. Lin, editors, *Symposium on Computer Animation*. The Eurographics Association, 2003.

Appendix A

Prototype

The prototype source code is available at `https://git.fe.up.pt/teses/2023/pedro_queiros_2023` as well as the videos used during the phase of the experiment.

A.1 Dynamic Charts API

The following endpoints of the dynamic charts API developed, mentioned in Section 4.1.2, accept POST requests, receiving a JSON object as payload:

- /export - an example of the JSON object received is presented in Listing A.1;

```json
{
  "route": "barRaceChart",
  "chartDiv": "div.barChart",
  "width": 1280,
  "height": 720,
  "port": 3000,
  "firstSpeed": 500,
  "buffer": 500,
  "customTimes": {
    "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 125000,
    "61fb0377-a723-4362-b8b0-0c4aa772743c": 45000
  },
  "costumSpeeds": {
    "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 500,
    "61fb0377-a723-4362-b8b0-0c4aa772743c": 250
  }
}
```

Listing A.1: Example of a JSON object received by the */export* endpoint.

- /props/barRaceChart - an example of the JSON object received is presented in Listing A.2;

```
{
  "width": 60,
  "height": 70,
  "dynamicXAxis": false,
  "dateFormat": "competition",
  "id": 27,
  "event": "goals",
  "title": "Champions League Top 10 Scorers",
  "titleSize": 28,
  "yAxisSize": 13,
  "xAxisSize": 13,
  "numberBars": 10,
  "opacity": 0.25,
  "speed": 500,
  "customTimes": {
    "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 125000,
    "61fb0377-a723-4362-b8b0-0c4aa772743c": 45000
  },
  "customSpeeds": {
    "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 500,
    "61fb0377-a723-4362-b8b0-0c4aa772743c": 250
  }
}
```

Listing A.2: Example of a JSON object received by the */props/barRaceChart* endpoint.

- /props/lineRaceChart - an example of the JSON object received is presented in Listing A.3;

```
{
  "width": 60,
  "height": 70,
  "dynamicXAxis": true,
  "dynamicYAxis": false,
  "dateFormat": "competition",
  "id": 3,
  "teams": "9,4,16,15,18",
  "title": "Standings of Benfica, FC Porto, Sporting, SC Braga and SC
      Vitoria over the seasons",
  "titleSize": 28,
  "yAxisSize": 12,
  "xAxisSize": 12,
  "opacity": 0.25,
  "speed": 500,
  "customTimes": {
    "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 125000,
```

```
      "61fb0377-a723-4362-b8b0-0c4aa772743c": 45000
    },
    "customSpeeds": {
      "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 500,
      "61fb0377-a723-4362-b8b0-0c4aa772743c": 250
    }
  }
```

Listing A.3: Example of a JSON object received by the */props/lineRaceChart* endpoint.

- /props/pieRaceChart - an example of the JSON object received is presented in Listing A.4.

```
  {
    "width": 50,
    "height": 50,
    "id": 27,
    "title": "Champions League Titles per Country",
    "titleSize": 28,
    "innerLabelSize": 15,
    "outerLabelSize": 15,
    "lineTickness": 2,
    "opacity": 0.25,
    "speed": 500,
    "customTimes": {
      "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 125000,
      "61fb0377-a723-4362-b8b0-0c4aa772743c": 45000
    },
    "customSpeeds": {
      "807afd41-c3f9-4289-a4b2-17cc2c00bc82": 500,
      "61fb0377-a723-4362-b8b0-0c4aa772743c": 250
    }
  }
```

Listing A.4: Example of a JSON object received by the */props/pieRaceChart* endpoint.