NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Uncertainty Visualization using Hypothetical Outcome Plots

Ana Patrícia Gonçalves Pereira

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

UNCERTAINTY VISUALIZATION USING HYPOTHETICAL OUTCOME PLOTS

by

Ana Patrícia Gonçalves Pereira

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Professor Mijail Naranjo-Zolotov

August 2022

DEDICATION

To all those who have always supported me and believed in me, a big thank you!

ACKNOWLEDGEMENTS

This Dissertation would not have been possible without the support of several amazing people in my life to whom I am extremely grateful.

To my advisor, Professor Mijail Naranjo-Zolotov, for all the advices and help he gave me. Thank you for your patience, challenges and support throughout this journey.

To Guilherme, for always being there through the good times and the bad, for all the patience and support even in the most difficult times, for always pushing me forward and believing in me even when I didn't.

To my parents and friends, thank you for being there!

To my work colleagues, for always supporting me and, of course, all the contributions to this Dissertation.

A big thank you to you all, without your help none of this would be possible!

ABSTRACT

Data-driven decision-making is crucial for any business. With an increasing interest in Business Intelligence, Data Visualization is playing a major role in decision-making processes. For making a well-informed and accurate decision, it is important to understand uncertainty in the data visualizations. Uncertainty visualizations improve the way users understand the data, as well as the confidence in their conclusions. An important type of uncertainty visualizations is the Hypothetical Outcome Plots (HOPs), which allow the audience to gain an intuitive idea of uncertainty through animated sequences of random draws from a distribution, leading to a more accurate understanding and decision.

This document intends to detail a proof-of-concept by carrying out a comparison of static visualization vs. HOPs in terms of efficiency and accuracy of results interpretation for Wayne Enterprises (fictional name) forecasting projects, in particularly the ones related with product launches and product loss of exclusivity. Wayne Enterprises is a world-leading supplier of advanced analytics, technological services and clinical investigation solutions for the life sciences industry. For that objective, it was built two prototypes using Python to support the proof-of-concept execution. A between-group experiment was carried out with 40 members of the German consulting team of Wayne Enterprises, where half answered a survey based on static visualizations and the other half based on HOPs. From this experiment, it is possible to conclude that HOPs can achieve similar results that static visualizations, with people taking the decision in less than half of the time when visualizing a HOP. Thus, it is possible to improve Wayne Enterprises decision-making process by accelerating it with Hypothetical Outcome Plots.

KEYWORDS

Data Visualization; Uncertainty; Decision-Making; Hypothetical Outcome Plots; Python

RESUMO

A tomada de decisões baseada em dados é crucial para qualquer negócio. Com um interesse crescente em *Business Intelligence*, a Visualização de Dados está a desempenhar um papel importante nos processos de tomada de decisão. Para se tomar uma decisão bem informada e precisa, é importante compreender a incerteza nas visualizações de dados. As visualizações de incerteza melhoram a forma como os utilizadores compreendem os dados, bem como a confiança nas suas conclusões. Um tipo importante de visualizações de incerteza é o *Hypothetical Outcome Plots* (HOPs), que permite ao público obter uma ideia intuitiva da incerteza através de sequências animadas de desenhos aleatórios de uma distribuição, conduzindo a uma compreensão e decisão mais precisas.

Este documento pretende detalhar uma prova de conceito através da realização de uma comparação entre visualizações estáticas e HOPs em termos de eficiência e exactidão de interpretação de resultados para projectos de *forecast* da *Wayne Enterprises* (nome fictício), em particular os relacionados com lançamentos de produtos e perda de exclusividade de produtos. A Wayne Enterprises é um líder mundial de análises avançadas, serviços tecnológicos e soluções de investigação clínica para a indústria das ciências da vida. Para esse objectivo, foram construídos dois protótipos utilizando Python para apoiar a execução da prova de conceito. Foi realizada uma experiência entre grupos com 40 membros da equipa de consultoria alemã da *Wayne Enterprises*, onde metade respondeu a um inquérito baseado em visualizações estáticas e a outra metade com base em HOPs. A partir desta experiência, é possível concluir que os HOPs podem alcançar resultados semelhantes aos das visualizações estáticas, com as pessoas a tomarem a decisão em menos de metade do tempo quando visualizam um HOP. Por conseguinte, é possível melhorar o processo de tomada de decisão da Wayne Enterprises, acelerando-o com *Hypothetical Outcome Plots*.

PALAVRAS-CHAVE

Visualização de dados; Incerteza; Tomada de decisões; *Hypothetical Outcome Plots*; *Python*

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

DSRM	Design Science Research Methodology
HCI	Human-Computer Interaction
HOPs	Hypothetical Outcome Plots
InfoVis	Information Visualization
LoE	Loss of Exclusivity
MS	Multiple Sclerosis
SPCs	Supplementary Protection Certificates

1. INTRODUCTION

Nowadays, the use of data visualization to support decision-making is increasingly common (Hullman et al., 2019) and its reliance on data has highlighted the need of conveying uncertainty in data visualizations (Grewal et al., 2021; L. Padilla et al., 2020).

Uncertainty can be found everywhere, some examples are economic uncertainties caused by the absence of accurate measurements, meteorological uncertainties driven by forecasting inaccuracy, health diagnostic uncertainties resulting from imprecision, and sensor data uncertainties caused by inaccuracy, error, or incomplete measures (Kamal et al., 2021). In summary, uncertainty can be a result of statistical error, ambiguity, risk information complexity, unawareness of important information, and so forth (Gray et al., 2022; Pang et al., 1997; Politi et al., 2007).

Uncertainty visualization usually concerns the presentation of visualised data together with additional information such as accuracy, error, and other source factors that affect comprehension and interpretation (Jena et al., 2020); and can be associated with results showed in terms of confidence intervals, probability density functions, boxplots, means and standard deviations, among others (Kirchner et al., 2021).

As modern analytical approaches become more complex, the identification and management of uncertainty information across the whole visualization pipeline is becoming increasingly challenging (Jena et al., 2020). Some authors (Han et al., 2011; Politi et al., 2007) point out that some individuals can be pleased with uncertainty identification, while other individuals, can be overwhelmed by uncertainty information when it is not conveyed in an effective way. Hence, the main challenge is to identify and understand the unpredictability of the information given and to efficiently display it (Kamal et al., 2021).

The section 1.1. will address the link between uncertainty and decision making, as well as identify the problem covered by this dissertation.

1.1. PROBLEM IDENTIFICATION AND STUDY RELEVANCE

Although there are many visualization techniques, one key point has frequently been overlooked: the uncertainty. Uncertainty visualizations improve the way users understand the data (Greis et al., 2018). The main objective is not just to share the actual data, but also its underlying uncertainties (Grewal et al., 2021), minimizing the errors in judgment (Kamal et al., 2021) and fostering uncertainty-conscious decisions, where the public recognizes the hazards and benefits of certain decisions, and modulates the reliance in its conclusions (Correll et al., 2018).

Designers frequently think that uncertainty is difficult to deal with because of practical issues in building visualization and modelling uncertainty with regard to decision-making (Kamal et al., 2021). Further suggested explanations why a visualization designer might be reluctant to introduce uncertainty are the concern that information about uncertainty implies unreasonable precision in estimates, the believe that uncertainty's existence is common knowledge, thinking that non-expert public will not comprehend the information, and that conveying uncertainty will result on a less credible message (Fischhoff, 2012).

Providing just point estimates might give the audience a false sense of precision, making them think that an estimate or forecast is more precise than it actually is (Fernandes et al., 2018). By communicating the likelihood that a point estimate can differ, uncertainty visualizations allow the audience to make more informed decisions (Hullman et al., 2019). Yet, uncertainty visualizations can make estimates more ambiguous and, consequently, more difficult to predict reactions to (Hullman, 2020). In addition to ambiguity, uncertainty can also lead to misinterpretation (Hullman, 2016; Kale et al., 2018), for instance, error bars encoding confidence intervals or standard errors are often misinterpreted (Belia et al., 2005; Hoekstra et al., 2014), possibly due to the fact that these frequency statistics are misunderstood as showing the likelihood of an estimate and not the variability of the process that produced the estimate (Kale et al., 2018). A potential solution to mitigate misinterpretations of uncertainty in 2D data is the display of hypothetical outcome plots (HOPs) (Kale et al., 2019). Hypothetical Outcome Plots are an animated sequence of random draws from a distribution (Kale et al., 2018; Roy et al., 2018; Y.-S. Kim et al., 2019; L. Padilla et al., 2020), that are known to lead to more accurate interpretation and performance (Kale et al., 2020). For these reasons, one of the objectives of this dissertation is to build HOPs (Hypothetical Outcome Plots will be further explained in section 2.2.2.).

This dissertation will be developed in a company, from now on designated as “Wayne Enterprises”, inserted in the healthcare sector. Wayne Enterprises is a world-leading supplier of advanced analytics, technological services and clinical investigation solutions for the life sciences industry. The company is made up of different departments, each one having a panoply of projects available. The focus will be two types of projects developed by the German Consulting department. These projects are based on forecasting analysis and are related to pharmaceutical product launches and product loss of exclusivity (LoE). In order to develop these projects, the employees working on the project have to research/gather the data, clean it, do the analysis and in the middle and end of the project make different decisions. Since decision-making is a major part of this type of projects, the company would like to improve the process, in order to make more informed, accurate and quicker decisions.

Decision-making frequently relies on the analysis and assessment of large volumes of data for which information visualization has proven to be a powerful approach (Griethe & Schumann, 2005). Several studies have demonstrated that conveying uncertainty can enable people to make better decisions than when uncertainty is not presented (e.g., (Gkatzia et al., 2016; S. L. Joslyn & LeClerc, 2012; Jung et al., 2015; LeClerc et al., 2016; Nadav-Greenberg & Joslyn, 2009)). Nevertheless, convey uncertainty in data, clearly and accurately, is still an open challenge (Grewal et al., 2021).

1.2. Study Objectives

The main objective of this dissertation is to conceive a proof-of-concept experiment using two prototypes to investigate the impact of Hypothetical Outcomes Plots in the decision-making process in a professional environment. With this objective and the company's goals in mind, the following Research Question was formulated: "How can Hypothetical Outcome Plots perform better than static visualizations when interpreted by experienced visualization users?".

In order to better address the research question, the following sub-questions were formulated:

- What is a Hypothetical Outcome Plot?
- What are the benefits of using HOPs?
- What are the disadvantages of a Hypothetical Outcome Plot?
- Does experience in visualization have an effect on viewers accuracy?
- What is the relationship between time and the interpretation of different plots?
- Does the representation of more than one figure at the same time have an influence on the viewer's interpretation?

To accomplish this study's goal and address the Research Question, the steps below were defined:

1. Identify the requirements for building a HOP;
2. Understand what type of visualizations are used in Launch and LoE projects at Wayne Enterprises;
3. Develop the HOPs' prototypes using Python
4. Use the prototypes to support the proof-of-concept execution;
5. Evaluate the proof-of-concept.

1.3. DISSERTATION STRUCTURE

This document is divided in seven major sections, as described below:

- The first section is the introduction. Here, is where the background of the dissertation is explained, the study problem is identified, and the study objectives are defined.
- The literature review is the second section and is where the theoretical research is carried out. This section addresses two key topics: data visualization, and uncertainty. Each one has two sub-sections: Information Visualization Principles, and Scientific Visualization Principles are the sub-sections of Data Visualization, and Uncertainty Representation and Hypothetical Outcome Plots are the sub-sections of Uncertainty.
- In the third section, the chosen methodology is explained, as well as the development of the prototypes and proof-of-concept.
- The fourth section contains the results of the experiment and is where they are evaluated
- In the fifth section is where the results presented in the fourth section are discussed.
- The conclusions drawn are fully explained in the sixth section.
- The seventh and last section is where the limitations found are explained and the recommendations for future work are made.

2. LITERATURE REVIEW

2.1. DATA VISUALIZATION

Visualization is "the process of transforming data, information and knowledge into visual form making use of humans' natural visual capabilities" (Gershon et al., 1998). It connects the human visual system and the information system, aiding in identifying images, formulating hypotheses and extracting insights from data sets, as well as contributing to scientific research and forecasting (Hrabovskyi et al., 2020) and supporting the decision-making processes (Kinkeldey et al., 2017).

There are two types of visualization: Information Visualization (InfoVis) and Scientific Visualization. The first one comprises concept visualization, which is usually abstract in nature; while Scientific Visualization encompasses precise visualizations of world and is used to clarify well known phenomena (Nagel, 2006). Uncertainty can be observed in both types of visualization (Kinkeldey et al., 2017), therefore, the next subsections will go into more detail on the principles of each type of visualization.

2.1.1. Information Visualization Principles

Edward Tufte (2006), also known as "the da Vinci of data" (Chen, 2017), described six fundamental principles for information visualization:

- **"Comparisons":** This principle claims that visuals should show comparisons, i.e., the main point in a figure is to enable a person to make intelligent and appropriate comparisons.
- **"Causality, Mechanism, Structure, Explanation":** Thoughts about cause and effect might be provoked simply by collecting data, therefore, it is important that visuals represent causality between data
- **"Multivariate Analysis":** In the beginning, the cause-and-effect analysis is bivariate, nonetheless quickly becomes multivariate. The Multivariate Analysis principle defends that thinking about data should not be two-dimensional, multivariate thinking is needed and for that is necessary to show more than one or two variables.
- **"Integration of Evidence":** in order to transmit clear information, it is important to integrate multiple elements (text, numbers, images, diagrams) in the visual representation and present all relevant evidence.

- **“Documentation”:** is a key quality control tool, since credibility is given by sources’ quality and integrity.
- **“Content Counts Most of All”:** The most effective way to convey information is to ensure the quality, relevance, and integrity of the content.

In summary, Tufte advocates thinking about proof, which includes data description, performing multivariate comparisons, comprehending causality, multiple evidence integration, and detailing analysis.

Other authors focus their work on effective visualizations, one example is Ying Zhu (2007), who raises two key points regarding InfoVis: the definition of effective visualizations and its measurement. In order to determine the effectiveness of data visualization, this author uses the following criteria:

- **“Accuracy”:** sets the connection between visualization and data, by declaring that visual component attributes must correspond to data element attributes, and the visualization structure must correspond to the dataset structure
- **“Utility”:** This criterion defines the link between the visuals and the tasks. The first must support users in achieving the objective of the second in order to be considered effective.
- **“Efficiency”:** describes the association amid visualizations and users. The efficiency criterion states that effective visuals must decrease the cognitive burden for a particular task on non-visual depictions, i.e., the visualization must be effortless to learn and enhance the efficiency of the task.

Although Ying Zhu defines these criteria, also points out that they are highly impacted by the domain knowledge of users, visualization experience, and visual-spatial ability; and measuring the effect of these drivers remains a major challenge.

2.1.2. Scientific Visualization Principles

We live surrounded by visuals and the need of creating effective scientific visualizations has been increasing. Nevertheless, it is common to see information incorrectly represented. Stephen Midway (2020) defined ten effective data visualization principles:

- **"Diagram First":** This principle states that we should give priority to the content we want to communicate, view and project it before we make the visualization.

- **"Use the Right Software":** In order to have effective visuals, normally, strong knowledge of one or more software is required. This usually means learning a new software or deepen expertise of a familiar software.

- **"Use an Effective Geometry and Show Data":** Geometry is a depiction of data in various ways, and most of them belong to these categories:

  - Amounts (or comparisons) – Geometries shall be employed only if data have no distributional information or uncertainty associated. Although Cleveland dot plots and heatmaps can be used, bar plots are often displayed;

  - Compositions (or proportions) – May take a wide range of geometries such as stacked or clustered bar plots, pie chart, stacked density plots, mosaic plots and tree maps;

  - Distributions – Geometries for this category demonstrate high data density and are often an underused class of visuals (boxplot, histogram, violin plots, and density plots);

  - Relationships – Considered the "workhorse" of geometries, since the so popular scatterplot is included, as well as other figures of the coordinate data x and y.

  Also, this principle recommends to show the data, i.e., data can be appended and shown to provide the context for the geometry, even though a geometry may be the focal point of the image.

- **"Colors Always Mean Something":** Color represents information and its use in visualizations can be extremely powerful. The majority of visualizations apply color in one of these ways:

  - Sequential – vary from light to dark in one or two shades

  - Diverging – contain two consecutive schemes representing two different extremes, often with one white or neutral color in between

  - Qualitative - the use of different and non-related colors to communicate the qualitative differences of the group

  It is worth not forgetting some technical recommendations in the use of color:

  - The colors used should enable conversion to a gray scale without losing information

- Colors may be associated with symbols, line styles, and other design features
- Effective color schemes for color-blind users should be used
- Color transparency can be used so that the color information is memorized, yet is not visually overwhelming or superior to other design features.

- **"Include Uncertainty":** Not including uncertainty in a visualization can be misleading, however there are two main challenges in including it: failure to include uncertainty and its misinterpretation. Although uncertainty is rarely seen in visuals, it is usually easy to include it in the majority of software programs and can be in the shape of standard geometries, namely error bars and shaded intervals (polygons). Common metrics of uncertainty are standard deviation, standard error, confidence intervals, and credible intervals. Expressing uncertainty is important, readers must be acquainted with uncertainty metrics and how to interpret them, but the figure author also has to choose the proper uncertainty measure so the message can be well interpreted.
- **"Panel, when Possible (Small Multiples)":** Small Multiples is an approach that repeats a figure with a change in one variable, in order to highlight differences and make the data comparable.
- **"Data and Models Are Different Things":** A model must be fully described to guarantee reproducibility and a model visualization needs to be described in the figure legend or referred somewhere within the paper in order to enable the reader to locate all the information about what the model visual is depicting.
- **"Simple Visuals, Detailed Captions":** Legends must clarify the geometries used, i.e., should provide the best possible explanation for the visualizations and representations used.
- **"Consider an Infographic":** Infographics usually embed text, images, and other graphical components, while figures generally focus on representations of data and models. Although infographics score high on memorability and figures are moving in the direction of infographics, converting all figures into infographics is not advised.
- **"Get an Opinion":** The most successful visualizations are those that the audience engage with, thus seeking external reviews is encouraged.

Other authors also added some rules or guidelines, in order to make a more effective visual. Some examples are: "know the audience", "do not trust defaults" (Rougier et al., 2014), focus on viewing patterns or details, choose relevant axis intervals, and integrate wider datasets in a significant manner (Kelleher & Wagener, 2011).

2.2. UNCERTAINTY

Most often, visualization methods presume that the depicted data is error-free; yet this is seldom the case. Since the error is intrinsic to data, it cannot be ignored in visualizations (Kamal et al., 2021), uncertainty cannot be ignored.

It is proven that uncertainty information improves the user understanding of the data (Greis et al., 2018), as well as the decision making process (Fernandes et al., 2018; Gkatzia et al., 2016; S. L. Joslyn & LeClerc, 2012; Jung et al., 2015; Kayongo et al., 2021; LeClerc et al., 2016; Nadav-Greenberg & Joslyn, 2009). One example is the experiment of Joslyn and LeClerc (2013) where individuals made better cost-effective decisions when provided with uncertainty in a weather forecast than those who were provided with weather forecasts alone. Uncertainty is present in our day-to-day life, it can be observed not only in weather forecasts, but also in traffic simulations, GPS, politic polls, games, and so on; but what is uncertainty exactly?

There are several definitions of uncertainty: in the first publications, it was defined as statistical (provided by estimated mean and standard deviation, or a true distribution of the data), errors (a difference, or an absolute valued error between data estimations) and range (an interval where the data should be present, although it is not possible to quantify it in statistical or error definitions) (Pang et al., 1997); while in a most recent article was defined as the likelihood that actual data or the model predictions may assume on a range of possible values (Hullman, 2020). Despite the many definitions, it seems to be consensual that there are several sources of uncertainty. Pang (1997) stated that uncertainty may appear during any phase of the visualization development (acquisition, transformation, or visualization), and Bonneau (2014) complemented by also covering sampling, quantization and interpolation.

According to Kirchner (2021), it appears that there is an understanding of the distinction between two primary causes of uncertainty: stochastic and epistemic. Stochastic uncertainty is caused by underlying variability in natural and human systems (Kirchner et al., 2021), and cannot be reduced (Kirchner et al., 2021; Therón Sánchez et al., 2019). It is also called statistical uncertainty, aleatoric uncertainty, type A uncertainty, variability uncertainty, irreducible uncertainty, and objective uncertainty (Therón Sánchez et al., 2019). Stochastic uncertainty occurs mostly in scientific fields and is generally linked to objective knowledge derived from general knowledge or unique observations (Therón Sánchez et al., 2019). An epistemic uncertainty, on the other hand, is caused by a lack of knowledge/information and can theoretically be reduced (Kirchner et al., 2021; Therón Sánchez et al., 2019). It is also called systematic uncertainty, type B uncertainty, state of knowledge, reducible uncertainty, and subjective uncertainty (Therón Sánchez et al., 2019). Epistemic uncertainty is particularly linked to decision-making processes and, therefore, can be found either in scientific investigation (commonly related to hypothesis testing) or in humanities investigation (related to contested theories or events) (Therón Sánchez et al., 2019).

Other publications address additional uncertainty sources, such as language imprecision (Ascough Ii et al., 2008), decision uncertainty (Ascough Ii et al., 2008; Peterson, 2006), and ambiguity (Warmink et al., 2010). Table 1 represents the taxonomy of uncertainty sources based on Kirchner's (2021) article.

Table 1 – Uncertainty sources' taxonomy (Kirchner et al., 2021)

Main Type	Sub-type	Specific cause
Stochastic	Natural variability	Inherent randomness
		Natural variation
	Human variability	Moral judgements
		Behavior
		Institutional
		Technological breakthroughs
Epistemic	Unreliability	Inexactness
		Missing observations
		Practically unquantifiable
	Structural uncertainty	Contradictory evidence
		Reducible/irreducible ignorance
		Indetermination
	System understanding	Process
		Cause
		Effect
	Linguistic uncertainty	Vagueness
		Ambiguity
		Sub-specificity
		Indetermination of theoretical terms
		Context dependency

2.2.1. Uncertainty Representation

Uncertainty is frequently divided into two categories: the one that can be directly quantified and the one that can't. The first category can be described by mathematical terms and can be employed to forecast future situations, while the second category occurs if a model exhibits variability or error which the analyst can neither predict nor quantify (L. M. Padilla et al., 2020). Following this line of thought, representation approaches can be categorized in quantification or visualization, where quantification deals mainly with the uncertainty of modelling data via multiple mathematical processes, and visualization involves displaying data uncertainty visually (Kamal et al., 2021). Below, figure 1 systematizes the two approaches, quantification and visualization approaches, and the popular techniques under them.

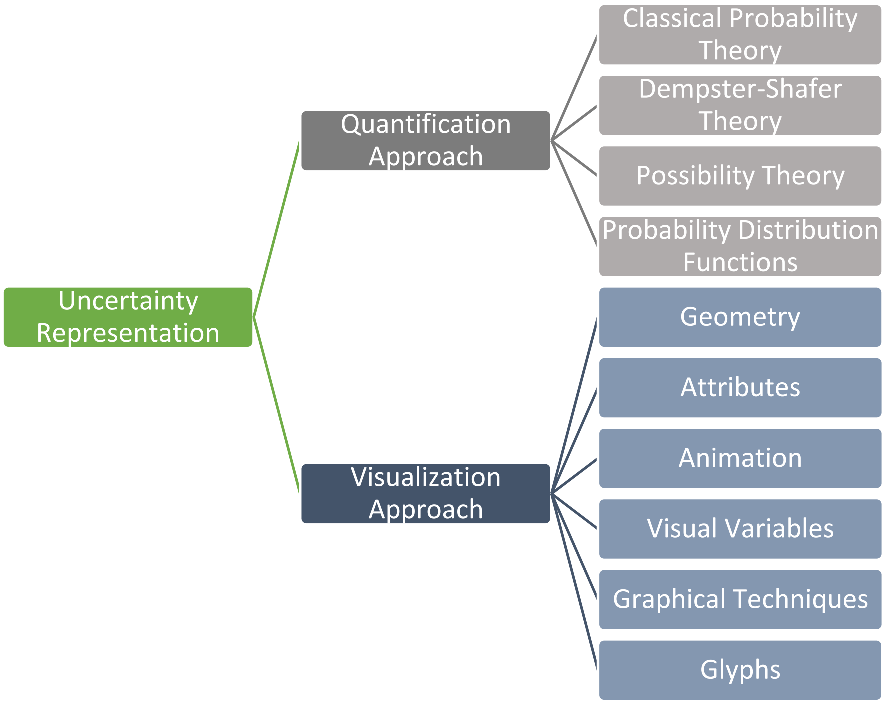


Figure 1 - Approaches to uncertainty representation (Kamal et al., 2021)

The quantification approach will not be further addressed, since is beyond the scope of this dissertation. Focusing on visualization approaches, figure 1 comprises six techniques:

- **Geometry:** added to express continuous data. Some examples of techniques are: contour lines, streamlines, isosurfaces, volumes, and modification of the existing geometry through distortion, scaling or rotation (Kamal et al., 2021; Pang et al., 1997). Glyph is another technique for adding geometry, nevertheless it is only added in discrete locations (Kamal et al., 2021; Pang et al., 1997).

- **Attributes:** uncertainty can be displayed through the modification of attributes such as shading, lighting, color, and embellishments (e.g.: textures) (Kamal et al., 2021; Pang et al., 1997).

- **Animation:** is a strong, innate and unambiguous technique of showing uncertainty to the audience while not cluttering the visualization as several parameters exist in the animation to which uncertainty can be applied (Kamal et al., 2021). Some of these parameters are: speed, duration, blinking, motion blur, and range (Pang et al., 1997). Hypothetical Outcome Plots are one of the uncertainty visualization techniques. The section 2.2.2. will cover this topic in more depth.

- **Visual Variables:** are frequently employed to display uncertainty (Kamal et al., 2021) and some examples are: location, size, color hue, color value, color saturation, orientation, grain, arrangement, shape, fuzziness, transparency, and brightness (Kamal et al., 2021; MacEachren et al., 2012).

- **Graphical Techniques:** traditional and most popular way of depicting large amounts of data in an understandable way, and include, for example, box plots, scatter plots, violin plots, and histograms (Kamal et al., 2021).

- **Glyphs:** geometrically plotted specifiers that encode data values via their shape and/or color (Pang et al., 1997). Glyphs attributes such as size, shape, color, and orientation can be employed to represent data in the display (Kamal et al., 2021). Some applications of glyphs are natural phenomena representation (e.g. hurricane motion on a map), flow visualization, vector visualization, interpolation, and radiosity (Kamal et al., 2021).

These techniques aim to represent the uncertainty information (Kamal et al., 2021) and are valuable to those who need to make informed decisions based on imperfect data (Pang et al., 1997).

2.2.2. Hypothetical Outcome Plots

Uncertainty static visualizations often have the problem that the audience can perceive some feature of the display as deterministic (Wilke, 2019). This problem can be overcome by displaying the uncertainty using animation, specifically using Hypothetical Outcome Plots (HOPs) (Hullman et al., 2015). HOPs (figure 2) are an animated sequence of random draws from a distribution (Kale et al., 2018; Roy et al., 2018; Y.-S. Kim et al., 2019; L. Padilla et al., 2020; Phelan et al., 2019), that allow the audience to make an impression of the uncertainty intuitively while watching (L. Padilla et al., 2020), leading to more accurate understanding (Hullman et al., 2015). Figure 2 represents an example of a bar and line HOPs, where translucent bars/lines represent the movement, and some example frames to better understand what HOPs show.

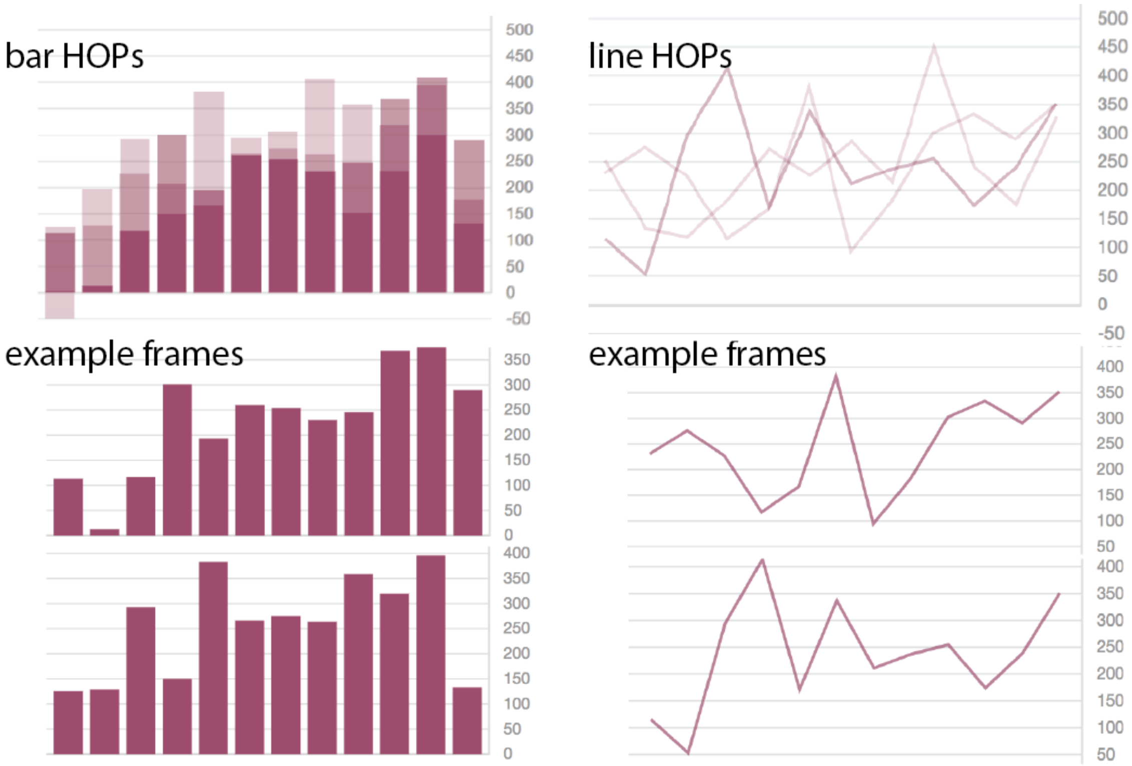


Figure 2 – Example of bar and line HOPs where the single frames are made translucent and merged into one depiction to reflect the HOPs movement (Phelan et al., 2019)

In order to clarify the definition of a HOP, an example of Wilke (2019) will be used. If someone wants to buy a chocolate bar, usually does not concern with the mean taste rating and related uncertainty of particular chocolate bars, rather may wish to find out, for example, what to expect to taste better when randomly picks up a Canadian and an American chocolate bar. Imagining that this person, Joe, has a large dataset containing specialist ratings of chocolate bars, classified on a one to five scale, for chocolate bars produced in several countries such as: Switzerland, Canada, Austria, the United States, Belgium and Peru, then it is possible to answer that question, since Canada and the USA are present in the dataset. Assuming that Joe selects a Canadian bar and an American bar at random from the dataset, compares their ratings, registers the result, and then repeats this several times, then, he could represent this visually, by cycling through multiple of these random draws and displaying the relative ranking of the two bars for each draw as in figure 3. In Figure 3, each vertical green bar depicts the rating for a chocolate bar, and each panel displays a comparison between two bars chosen at random, one from a Canadian manufacturer and one from an American manufacturer; and in a real HOP the visualization would move between the different plot panels rather than show them side by side.

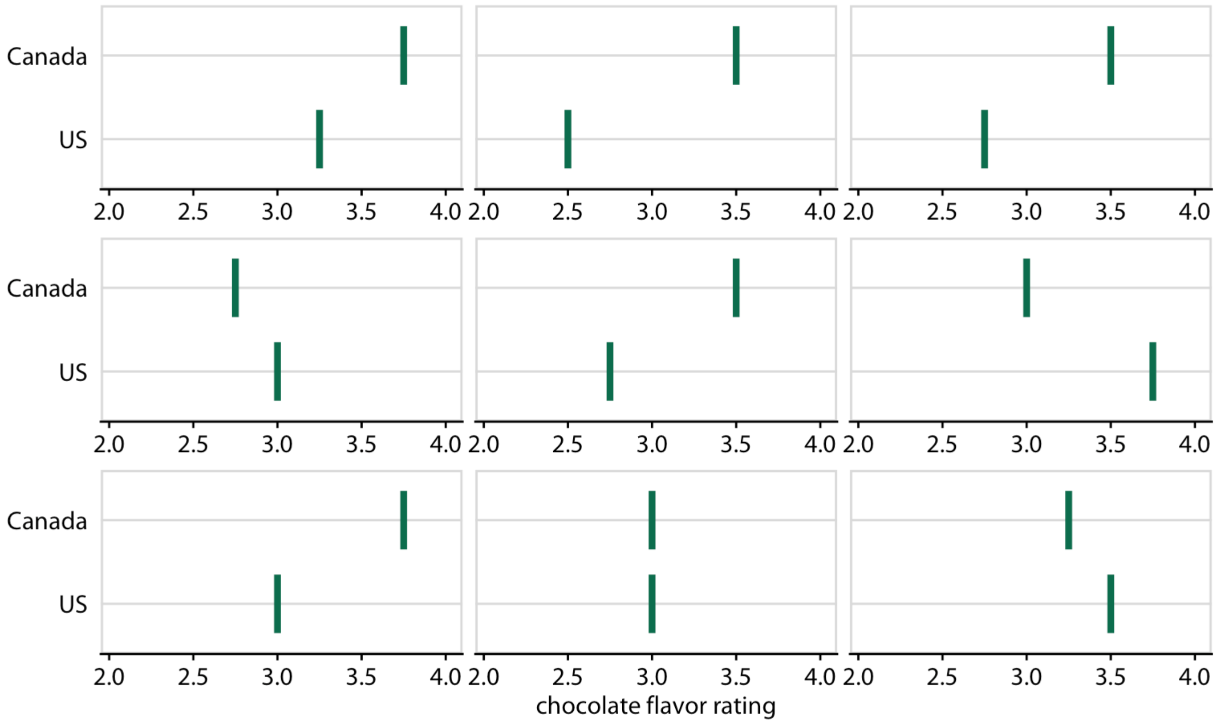


Figure 3 – Representation of a HOP for chocolate bar ratings (Wilke, 2019)

Although Hypothetical Outcome Plots are not feasible in a printed environment they can be highly successful in online environments or presentations (Wilke, 2019). When building a HOP, it is important to ensure that the results displayed reflect the true distribution of possible outcomes, if not, the Hypothetical Outcome Plot can mislead the audience (Wilke, 2019). Possible solutions to this problem are selecting a large number of results, whereby sample biases are not likely to occur, or checking that the results presented are appropriate (Wilke, 2019). Besides this, another important aspect when building a HOP is the frames switch, i.e., how should the switch occur, hard or smooth? Kale et al.

(2018) showed that smooth transitions make it more difficult to judge on the probabilities depicted, and, following this line of thought, Wilke (2019) suggested that HOPs should have fast transitions or an animation style in which the results fade in and out rather than distorting from one to the other.

Hypothetical Outcome Plots are a logical approach for uncertainty visualization when the basic visual encodings are already complex and difficult to understand, so that adding other encodings or glyphs might not be effective (Zhang et al., 2021).

Jessica Hullman (Hullman et al., 2015) identified some drawbacks, such as: introduction of sampling error, since the reader examines a limited set of frames, obtaining an inaccurate idea of the full distribution; and the reduction but not elimination of visual integration's difficulty. Nevertheless, several advantages have been recognized (Hullman et al., 2015; Kale et al., 2018; L. Padilla et al., 2020):

- Allowing the audience to think about individual results in finite ways instead of infinite ways abstracted over entire distributions
- No need of adding new marks or new encodings, and, therefore, not being required that viewers understand those marks/encodings;
- Force the audience to be aware of uncertainty in their understanding of data;
- Force viewers to acknowledge that unlikely outcomes are covered by the distribution;
- Strongly enhance multivariate probability estimation versus conventional static uncertainty visuals;
- Support decision-making.

Furthermore, a couple of studies about visualizations found that HOPs could lead to better estimates than error bars (Hofman et al., 2020; Hullman et al., 2015; Kale et al., 2018), as well as other approaches such as static ensembles and violin plots (Kale et al., 2018). Another advantage of Hypothetical Outcome Plots is that they allow for more accurate interpretations also among individuals who are not experienced in statistics and have had little instruction in using the plots (Hullman et al., 2015; Kale et al., 2018).

Overall, when weighing the advantages and drawbacks, is clear that advantages win.

3. METHODOLOGY

The proposed methodology for this dissertation is the Design Science Research Methodology (DSRM), developed by Peffers et al. (2007), which will guide in the development and assessment of two artifacts designed to address specific problems of Wayne Enterprises.

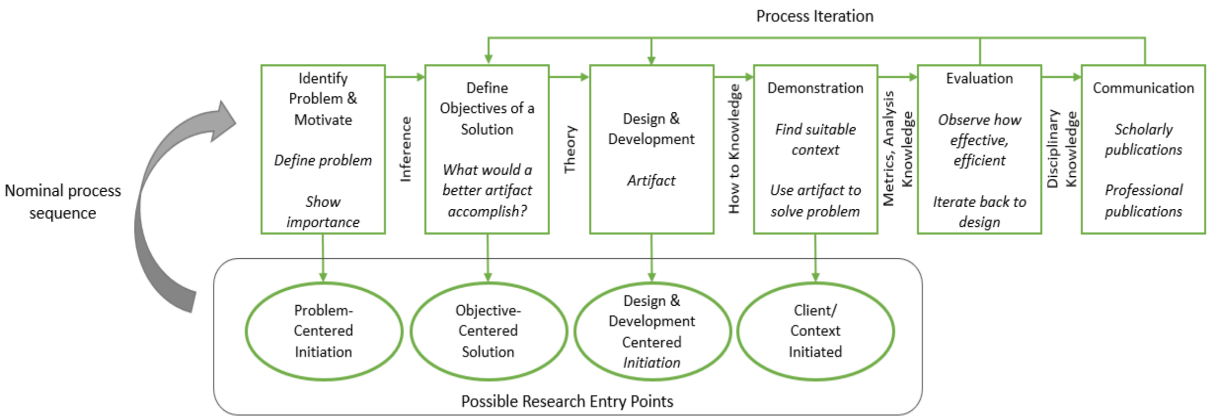


Figure 4 – DSRM Process Model (Peffers et al., 2007)

This approach presents six different steps of implementation:

1. Problem Identification and Motivation;
2. Definition of the Objectives for a Solution;
3. Design and Development
4. Demonstration;
5. Evaluation;
6. Communication

The choice of methodology for this dissertation is supported by the purpose of the same, i.e., the artifacts that will be created (HOPs prototypes) seek to tackle a problem and improve the decision-making process.

The Problem Identification and Motivation stage requires defining the research problem and the solution's value rationale (Peffers et al., 2007). This step was already done in sections 1.1. and 1.2. of this document. Also, it is possible to find in section 1.2. the step 2: Definition of the objectives for a solution.

In step 3 (Design and Development) is where the artifacts are created, which could be any conceived objects where an experiment contribution is incorporated into the design. To create the artifacts, it is necessary first to determine the wished functionalities and their structure, and afterwards to create the concrete artifacts (Peffers et al., 2007). In this step will be developed the entire process that will lead to the creation of the artifacts.

The aim of the Demonstration step is to show the usage of the artifacts to address the problem, which can include using them in experimentation, simulation, case study, proof, or another suitable activity (Peffers et al., 2007). Here is where the HOPs prototypes will support the proof-of-concept execution, being compared against static visualizations through a survey.

Regarding the Evaluation, it is required to notice and gauge the extent to which the artifacts provide an answer to the problem (Peffers et al., 2007). This step is performed in section 4 of this document, and is where the results of the experiment are evaluated.

Finally, the Communication will be done through this dissertation, which will disclose the identification of the problem and development of the solution.

As mentioned, the next sections will address steps 3 and 4, focusing on the development of the prototypes as well as the comparison process against static plots.

3.1. DESIGN & DEVELOPMENT

3.1.1. Context of the Prototypes

In order to perform the Design & Development step, it is necessary to know why launches and loss of exclusivity are so important for pharmaceutical industries. For that, it is important refer the life cycle of pharmaceutical products (figure 5).

The life of a pharmaceutical product starts with its discovery of the product and consequent development. Normally, this phase takes about 12 to 15 years (Haque & Ratemi, 2017), and is where the discovery phase, pre-clinical trials and clinical trials (Phases I, II and III) take place (Gaessler &

Wagner, 2019). The objective of this phase, besides the discovery of a new product, is to get marketing authorization, so that the product can be commercialized, as well as the patent covering the active substance of the medicine, giving exclusive rights. Once the product receives the patent protection and the marketing authorization, begins the market exclusivity period (Gaessler & Wagner, 2019). The duration of this period is set by the patent and data exclusivity periods. Data exclusivity concerns the timeframe in which the results of clinical trials are not eligible to be used by generic players to obtain further marketing approval. Since clinical trials are expensive, data exclusivity raises barriers to entries and is therefore a cause of market exclusivity regardless of patent protection (Gaessler & Wagner, 2019). In addition, there is also the supplementary protection certificates (SPCs). These certificates are optional and, in case the protection is granted, can delay loss of exclusivity by a maximum of five years (Gaessler & Wagner, 2019).

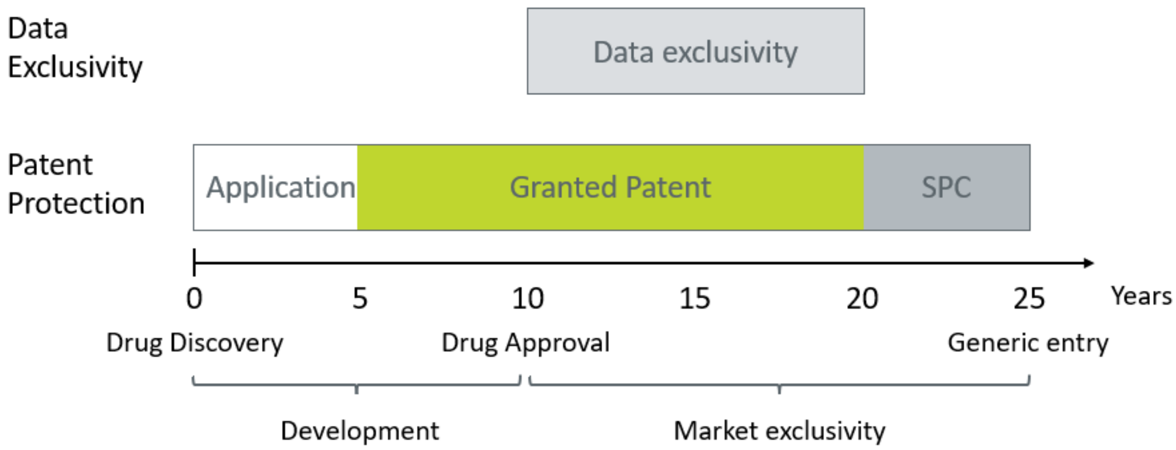


Figure 5 – Representation of pharmaceutical products life cycle (Gaessler & Wagner, 2019)

Once market exclusivity expires, generic manufacturers tend to step into the market, reducing the price and hence the margins and profitability of the originator (Gaessler & Wagner, 2019; Higgins et al., 2020). Aiming to delay generic competition, companies frequently attempt to extend the duration of market exclusivity by working on additional improvements and obtaining further patents during the life cycle of the drug.

Forecasts are important throughout all phases of the product lifecycle, nevertheless, this project focus on launch and LoE forecasts. Companies main interest in launch forecasts is to study how the future market will behave and to assess the overall future sales potential of the new product, while for LoE forecasts the main objective is to understand the impact of generics entry in the market. Some of the key questions that pharmaceutical companies seek answers to are:

- What is the expected share of decrease for the originator? (LoE);
- What share of generic sales can be expected? (LoE);
- How fast is the generics effect taking place? (LoE);
- How does the price develop after generics launch? (LoE);
- What uptake pattern can be expected for the new product? (Launch);
- What is the typical uptake pattern for similar products? (Launch);
- How will controlled access and risk management impact uptake and peak share? (Launch);
- What persistence (or therapy drop-out) and compliance rate can be expected? (Launch);
- What would be price per day? (Launch);
- What is the market and product potentials? (Launch).

Now that the launches and loss of exclusivity processes are explained, it is possible to build the prototypes. The section below (3.1.2.) explains the process of building the prototypes.

3.1.2. Prototypes Development

3.1.2.1. First Prototype (One Product)

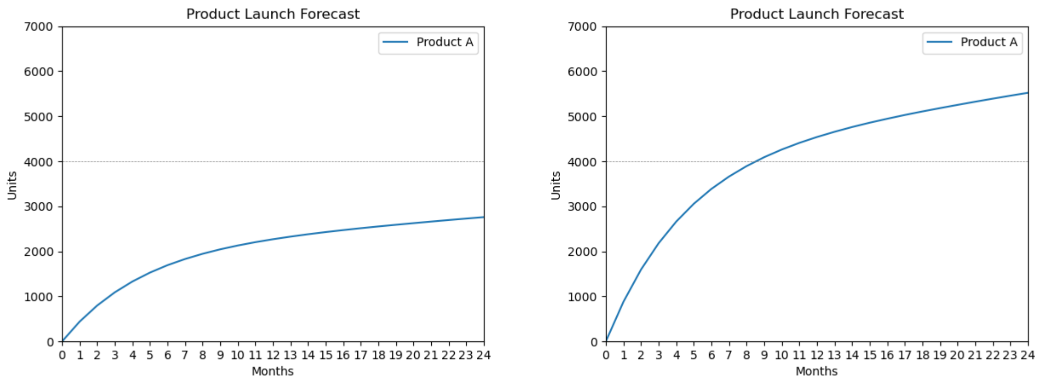
To build the first prototype, it helps having a case scenario, which in this case it will be a hypothetical one. Below is the case scenario's description and the standard procedure:

- **Case Scenario**: Wayne Enterprises has a client – Umbrella Corporation – that is preparing to launch a new product (Product A) on the market of Breast Cancer. The Umbrella Corporation has asked Wayne Enterprises to develop a forecast in order to understand how successful this product can be. For this, Wayne Enterprises will need to provide a list of analogous molecules in oncology market, and from this list select approximately two to three molecules in total based on key selection criteria in alignment with Umbrella Corporation. After analyzing each molecule, Wayne enterprises will develop the launch forecast for the new product.

Assuming that the process described above was done and all the data was gathered, it is possible to start developing the prototype. To do that, it is necessary to know and understand the type of graphics that are used in forecast projects at Wayne Enterprises, in order not to introduce bias in the study. After a brief informal poll with Wayne Enterprises' employees that currently do Launch forecasts, it was clear that the most commonly used type of graph in these forecast projects is line charts. Therefore, the prototype needed to be an animated line chart.

The HOP prototype was developed with Python, in particular with Matplotlib[1] and Imageio[2] libraries. First it was used Matplotlib to build the different plots with the possible outcomes, and only after all plots were obtained, it was used Imageio to combine all of them. The HOP was obtained by repeatedly sampling 24-month sets of units sold for Product A, showing these figures in line charts, and animating the shifts between frames. Animations were rendered at 0.5 frames per second (=500ms), since this frame rate tends to perform best according to L. Padilla (2020), with no transition animations defined (e.g., fading).

Since it is not possible to display the animated prototype in this document, below, in figures 6, are the different frames that integrate this HOP prototype.



[1] https://matplotlib.org/stable/index.html
[2] https://imageio.readthedocs.io/en/stable/

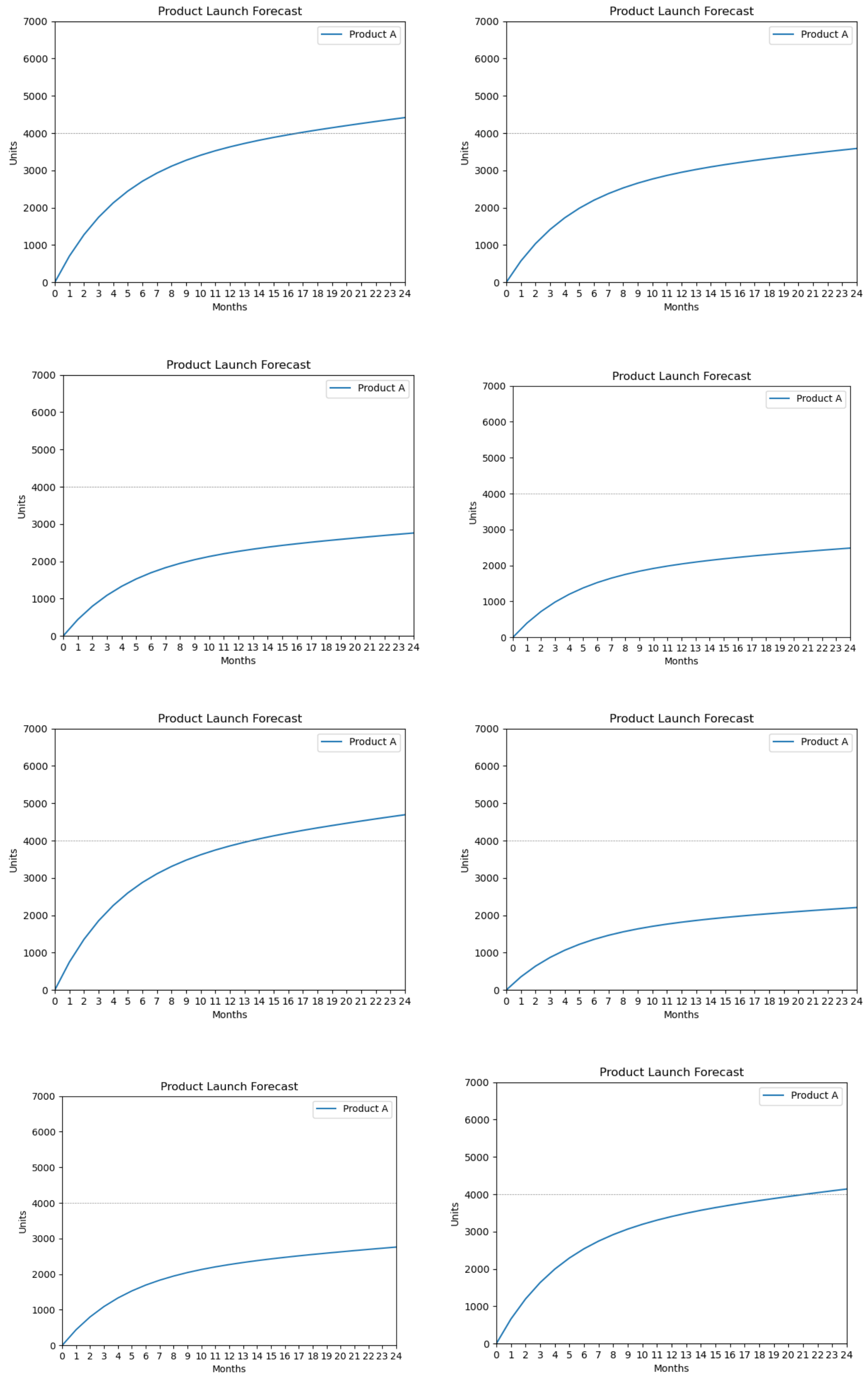


Figure 6 – Frames of Product A's HOP

3.1.2.2. Second Prototype (Two Products)

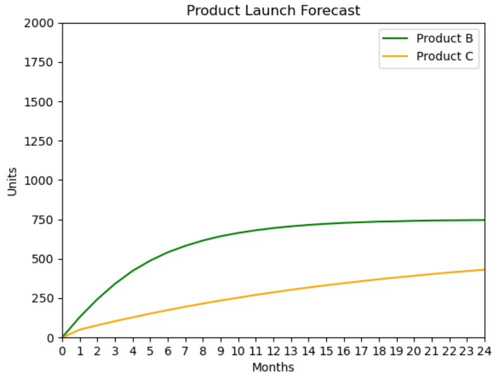
Besides building a Hypothetical Outcome Plot for Umbrella Corporation's product (figure 6), another one was built (figure 7), inspired by Hullman's (Hullman et al., 2015) experiment. This second prototype contains two different products with the objective to verify if having more than one product/line could impact the overall perception.

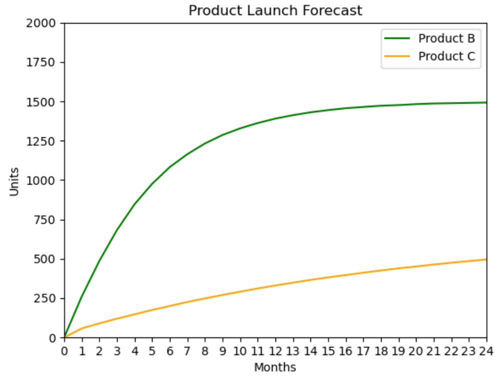
Below is the hypothetical case scenario's description as well as the standard procedure:

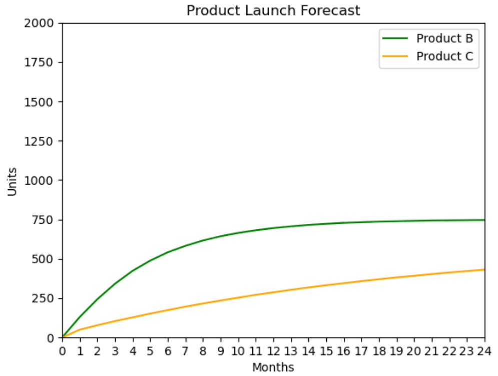
- **Case Scenario**: Wayne Enterprises has a client – Umbrella Corporation – that is preparing to launch a new product (Product B) on the market of Multiple Sclerosis (MS). However, the Umbrella Corporation learned that the competition will also launch a new product for the same market (Product C). Therefore, the Umbrella Corporation has asked Wayne Enterprises to develop a launch forecast for both products in order to understand how successful their own product can be, and if there is any chance of being overtaken by the competitor's product. For this, Wayne Enterprises will need to provide a list of analogous molecules in MS market, and from this list select approximately two to three molecules in total for each product based on key selection criteria in alignment with Umbrella Corporation. After analyzing each molecule, Wayne enterprises will develop the launch forecast for the new products.

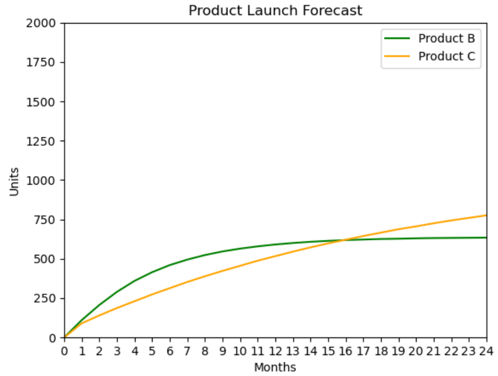
Like the first prototype, the second is also an animated line chart, developed with Python, using Matplotlib and Imageio libraries. Matplotlib was used first to create the distinct plots with the possible results, and only when all the plots were obtained, Imageio was employed to combine all of them. The HOP was generated by repeatedly sampling sets of 24-month units sold for Product A and Product B, plotting these numbers in line graphs, and animating the shifts between frames. Animations were rendered at 0.5 frames per second (=500ms), since, as mentioned in the previous section, this frame rate tends to perform better according to L. Padilla (2020), without setting transition animations (e.g. fading).

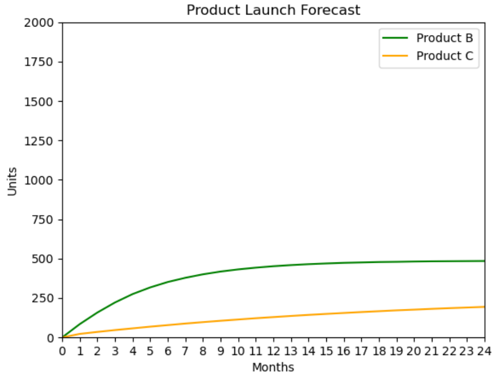
Given that it is not possible to display the animated prototype in this document, below, in figure 7, are the distinct frames that comprise this HOP prototype.

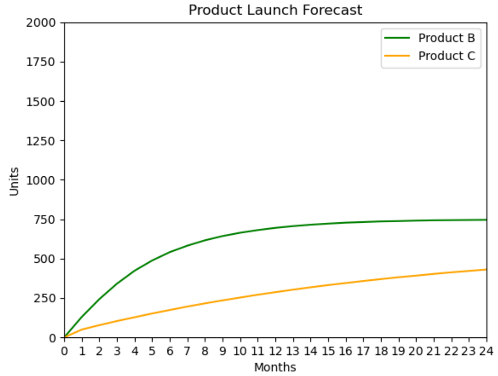

Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Product Launch Forecast
Product B
Product C
Units
Months
0
250
500
750
1000
1250
1500
1750
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

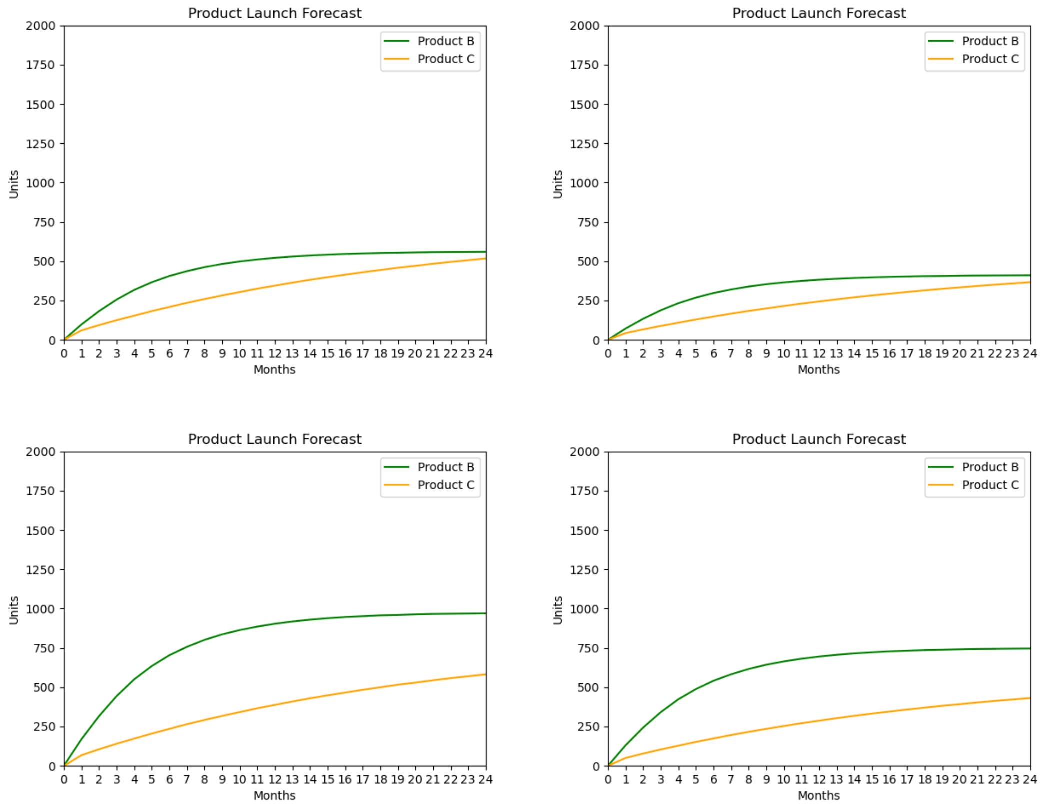


Figure 7 – Frames of two products' HOP

3.2. PROOF-OF-CONCEPT DEMONSTRATION

As described in the beginning of section 3, the Demonstration step shows the usage of the artifacts to address the problem, which, in this case, includes using it in the proof-of-concept. Before describing the proof-of-concept, it is important to understand the general process of a Human-Computer Interaction (HCI) experiment.

In the beginning, HCI research metrics relied on human performance measures from human factors and psychology, i.e., it was based on a task-centered model (Lazar et al., 2017). With this model, certain tasks such as performance speed, error rate, learning time, retention over time, and user satisfaction could be separated, quantified, and measured (Lazar et al., 2017; Shneiderman et al., 2016). Although such measures do not include motivation, cooperation, social engagement, reliability, and empathy, they remain the core metrics to assess interface usability and continue to be of interest today (Lazar et al., 2017; Shneiderman et al., 2016).

According to Shneiderman (2016), the HCI research methods most commonly employed are observations, field studies, surveys, usability studies, interviews, focus groups, and controlled experiments. Such approaches can be used in empirical investigation which can be divided in three categories: descriptive, relational, and experimental (Rosenthal & Rosnow, 2008).

The descriptive category includes observations, surveys and focus groups, and is based on building a precise picture of what is happening. The relational category allows the researcher to recognize relationships among numerous factors, while the experimental category allows the detection of causal relations. These three categories are not entirely unrelated, common research projects comprise a mix of two or even the three categories (Lazar et al., 2017). Table 2 contains the comparison between the three categories.

Table 2 – Comparison between Descriptive, Relational and Experimental Researches (Lazar et al., 2017)

Type of research	Focus	General Claims	Typical Methods
Descriptive	Outline a scenario or a number of events	A is occurring	Observations, field studies, focus groups, interviews
Relational	Recognize the relationships among several variables	A is connected with B	Observations, field studies, surveys
Experimental	Recognize the reasons behind a scenario or group of events	A is responsible for B	Controlled experiments

In order to do an experiment, investigators start by developing a research hypothesis. Usually, an experiment needs a minimum of one null hypothesis and one alternative hypothesis. Effective experimental research relies on clear research hypotheses that detail the dependent variables to be observed and the independent variables to be controlled. A dependent variable can be defined as the result or effect that the investigators are interested in, while the independent variable represents the agents that the investigators are keen to study or the potential "cause" of the dependent variable's change. In general, the aim of the experiment is to assess if the null hypothesis can be rejected or if the alternative hypothesis can be accepted.

After a hypothesis is constructed, is important to choose the design. There are different forms of experimental design: the between-group, the within-group and the split-plot designs.

In a between-group design, each subject only deals with one experimental condition and the number of subject groups matches the number of experimental conditions. Oppositional, the within-group design demands that each subject has contact with several experimental conditions and only a single group of subjects is required for the whole experiment. When compared, the between-group design is cleaner, prevents the learning effect, and is less prone to be impacted by tiredness and frustration, however, it is weaker because of the high loudness of individual differences. Also, the large number of participants that are usually required can be considered a drawback. In contrast, the within-group design successfully isolates individual differences and requires fewer participants. Nevertheless, the within-group design is more susceptible to learning effects and tiredness.

The third design, split-plot, is a mix of between-group and within-group elements, i.e., a single or more independent variable is researched using a between-group method and the other variables are researched sing a within-group method. This design can only be chosen if there is more than one independent variable. (Lazar et al., 2017). Regardless of which method is chosen, it is essential to consider the participants, the nature of the application, and the tasks examined, to be sure it is the most appropriate method.

Once the method is chosen and the experiment is done, is required to analyze the data collected. There are two major categories of tests: parametric and nonparametric. To choose a parametric test, some assumptions need to be met (Cooper et al., 2006; Lazar et al., 2017):

- The observations are required to be independent;
- Observations shall be taken from normally distributed populations;
- The data variance collected from different groups should be equal or approximately equal;
- Scales of measurement must have at least one interval to enable arithmetic operations to be performed.

When these assumptions are not fulfilled, nonparametric tests shall be used instead. Nonparametric tests are employed to test hypotheses containing nominal and ordinal data, and although these tests have less rigorous requisites regarding the data, they are not supposition free (Cooper et al., 2006; Lazar et al., 2017).

There are multiple parametric and nonparametric tests, here are some examples of the most used:

- T-test;
- ANOVA;
- Regression;
- Correlation (Pearson, Spearman's rho);
- Chi-square

The proper choice of statistical analysis methods is essential to avoid erroneous conclusions. The type of data gathered and the design of the experiment establish the most suitable test to be employed. The table below represents the recommended statistical techniques by Cooper et al. (2006).

Table 3 – Recommended statistical techniques by Cooper et al. (2006)

<table>
<tr><th rowspan="2">Measurement scale</th><th rowspan="2">1 sample Case</th><th colspan="2">2 Samples Tests</th><th colspan="2">k-Samples Tests</th></tr>
<tr><th>Related Samples</th><th>Independent Samples</th><th>Related Samples</th><th>Independent Samples</th></tr>
<tr><td rowspan="2">Nominal</td><td>Binomial</td><td rowspan="2">McNemar</td><td>Fisher's Exact Test</td><td rowspan="2">Cochran Q</td><td rowspan="2">χ^2 for k samples</td></tr>
<tr><td>χ^2 one-sample test</td><td>χ^2 two-samples test</td></tr>
<tr><td rowspan="4">Ordinal</td><td rowspan="3">Kolmogorov-Smirnov one-sample test</td><td>Sign test</td><td>Median test</td><td rowspan="4">Friedman two-way ANOVA</td><td>Median extension</td></tr>
<tr><td rowspan="3">Wilcoxon matched-pairs test</td><td>Mann-Whitney U</td><td rowspan="3">Kruskal-Wallis one-way ANOVA</td></tr>
<tr><td>Kolmogorov-Smirnov</td></tr>
<tr><td>Runs test</td><td>Wald-Wolfowitz</td></tr>
<tr><td rowspan="2">Interval and Ratio</td><td>t-test</td><td rowspan="2">t-test for paired samples</td><td>t-test</td><td rowspan="2">Repeated-measures ANOVA</td><td>One-way ANOVA</td></tr>
<tr><td>Z test</td><td>Z test</td><td>n -way ANOVA</td></tr>
</table>

For the experiment described in this document, it won't be possible to fulfill the parametric assumption, meaning that a non-parametric test needs to me chosen. Having in consideration table 3 and other reasons that will be explained further in this document, Fisher's Exact Test was the chosen test. Fisher's exact test evaluates the null hypothesis of independence by employing the hypergeometric distribution of the figures (H.-Y. Kim, 2017), being used mainly in 2 × 2 contingency tables but may be expanded to bigger contingency tables (H.-Y. Kim, 2017; Robertson & Kaptein, 2016). Although this test mainly used in small samples, it can be applied to all sample sizes (H.-Y. Kim, 2017).

3.2.1. Hypotheses and Questionnaire

Normally, an experiment starts by defining a research hypothesis. Therefore, before executing the proof-of-concept, it is necessary to define a hypothesis. Below are the null hypothesis (H_0) and the alternative hypothesis (H_1 and H_2) for the prototypes experiment:

H_0: There is no difference in uncertainty interpretation between a static plot and a HOP.

H_1: HOPs allow to interpret uncertainty with more accuracy.

H_2: HOPs allow to interpret uncertainty faster than static visualizations.

Considering these hypotheses and Wayne Enterprises employees, the between-group design was chosen. The experiment consists in two groups of participants answering a survey. Each group has 20 randomized participants and all of them belong to the German Consulting department of Wayne Enterprises.

There are two surveys, one with the two prototypes (figures 6 and 7) and other without them. In the last case, the survey without the prototypes, it was used static plots instead (figures 8 and 9), so the prototypes could be compared against them and the hypotheses be tested. To build these static plots, Python was used, in specific the Matplotlib library. The lines in figures 8 and 9 represent the mean of the product, and the shading area represents the confidence level (99%).

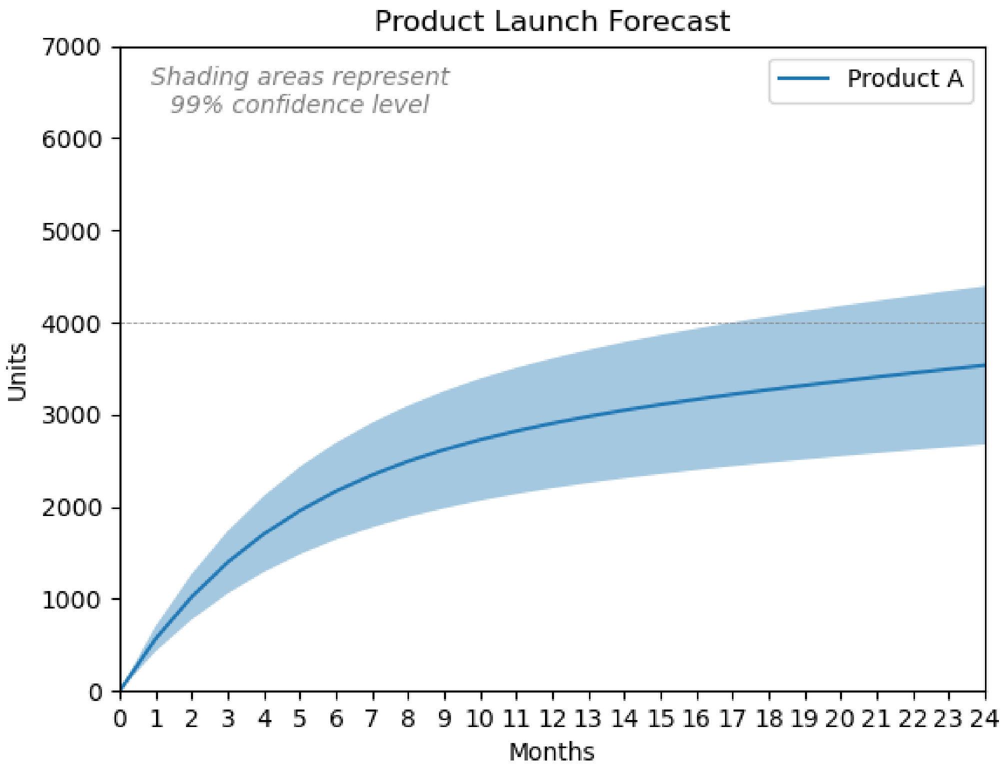


Figure 8 – Static plot representing Product A launch forecast

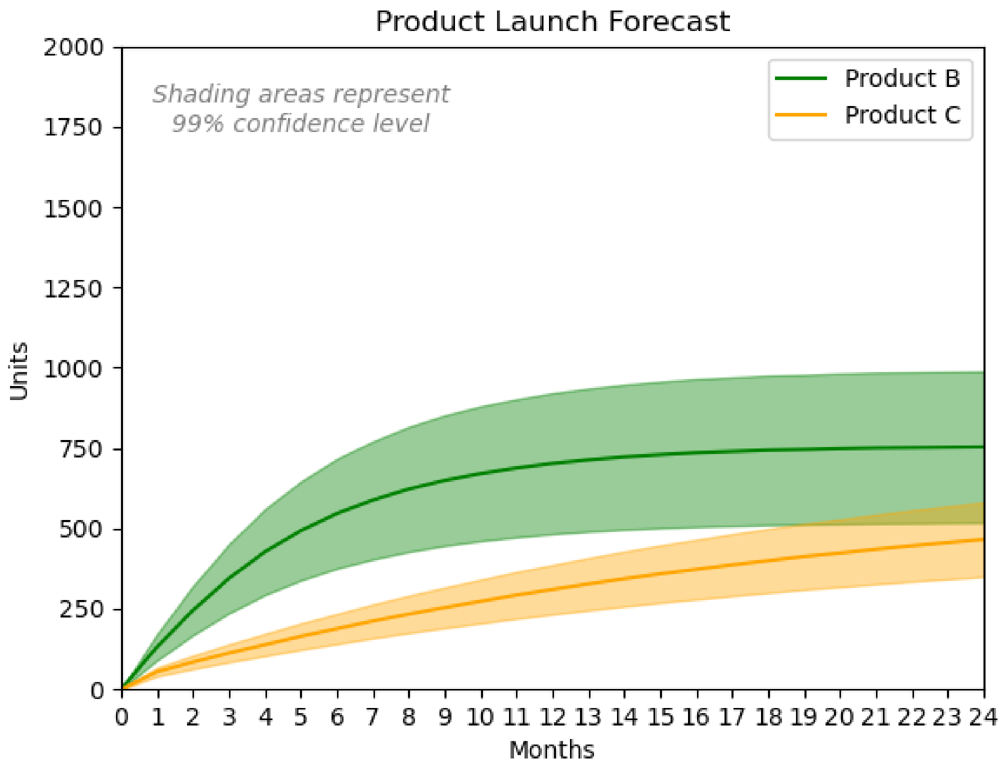
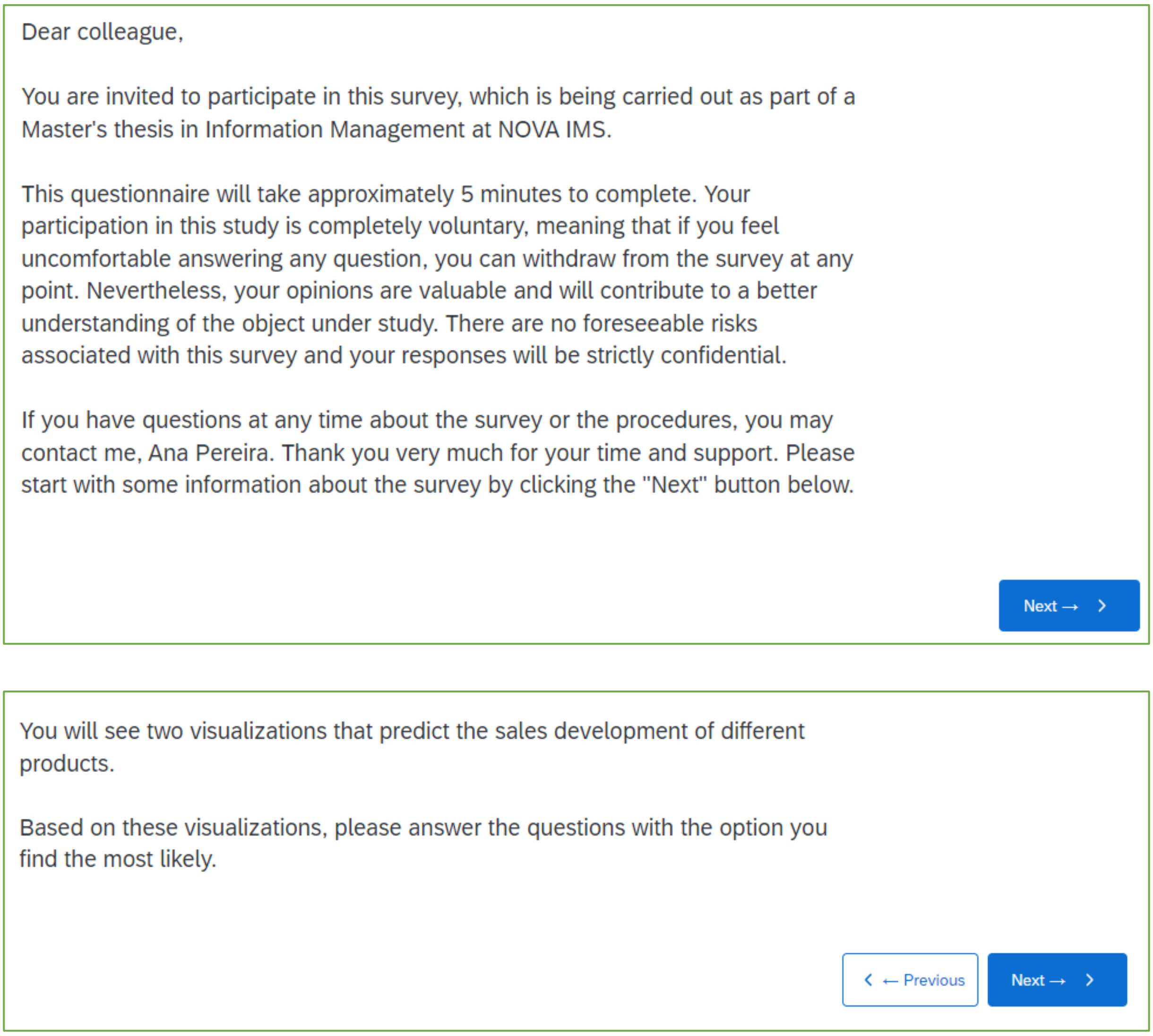


Figure 9 – Static plot representing Product B and Product C launch forecasts

Both surveys have the same introduction (figure 10) and consist in the same four questions:

1. How likely is that one year (month 12) after Product A’s launch the sales are over 4 000 units?
2. How likely is that in month 16 after launch, Product C has the same or more units sold than Product B?
3. How old are you?
4. How many years of experience with visualizations have you?


Dear colleague,

You are invited to participate in this survey, which is being carried out as part of a Master's thesis in Information Management at NOVA IMS.

This questionnaire will take approximately 5 minutes to complete. Your participation in this study is completely voluntary, meaning that if you feel uncomfortable answering any question, you can withdraw from the survey at any point. Nevertheless, your opinions are valuable and will contribute to a better understanding of the object under study. There are no foreseeable risks associated with this survey and your responses will be strictly confidential.

If you have questions at any time about the survey or the procedures, you may contact me, Ana Pereira. Thank you very much for your time and support. Please start with some information about the survey by clicking the "Next" button below.

Next →

You will see two visualizations that predict the sales development of different products.

Based on these visualizations, please answer the questions with the option you find the most likely.

← Previous

Next →

Figure 10 – Introduction of the surveys

Questions 1 and 2 are for figures 6/8 and 7/9, respectively. In both questions, the participants can only choose one of the following options: impossible, very unlikely, unlikely, neutral, likely, very likely, certain (as exemplified in figure 11). To use Fisher's Test on these questions it is necessary to have nominal variables (as explained in table 3), therefore, the participants' answers were transformed into binary, i.e., correct or incorrect answer.

Figure 11 – Question 1 of Static Plots Survey

Questions 3 and 4 are demographic questions, and in order to increase the response rate, participants are requested to answer through predefined intervals as shown in figures 12 and 13.

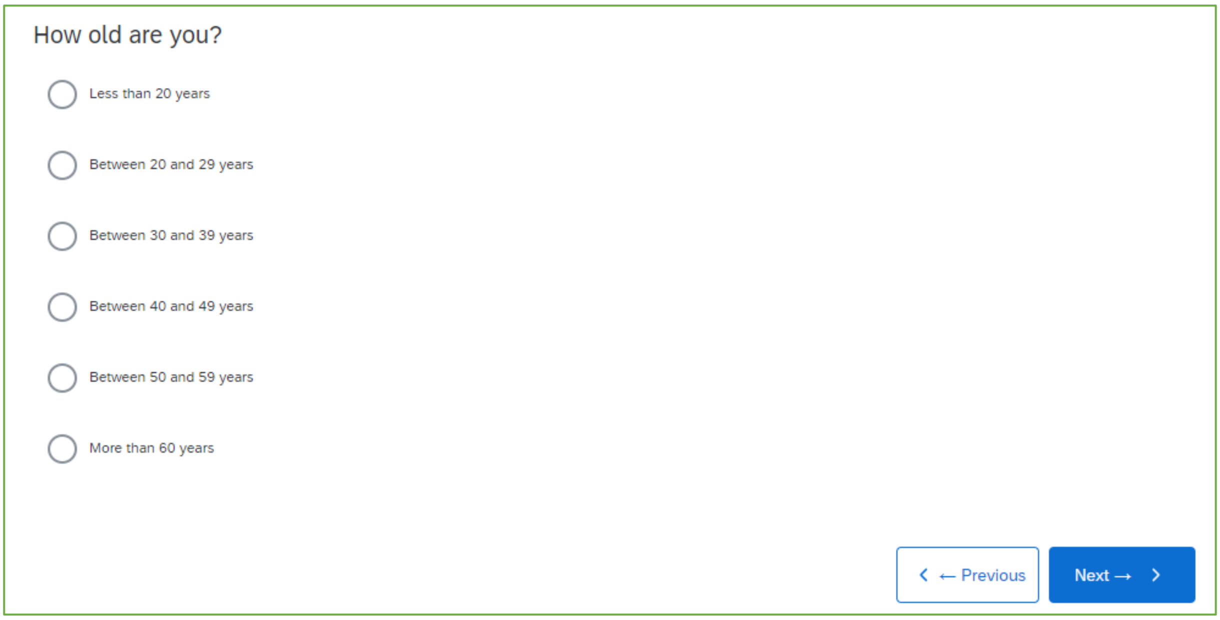


Figure 12 – Question 3 (demographic) of both surveys

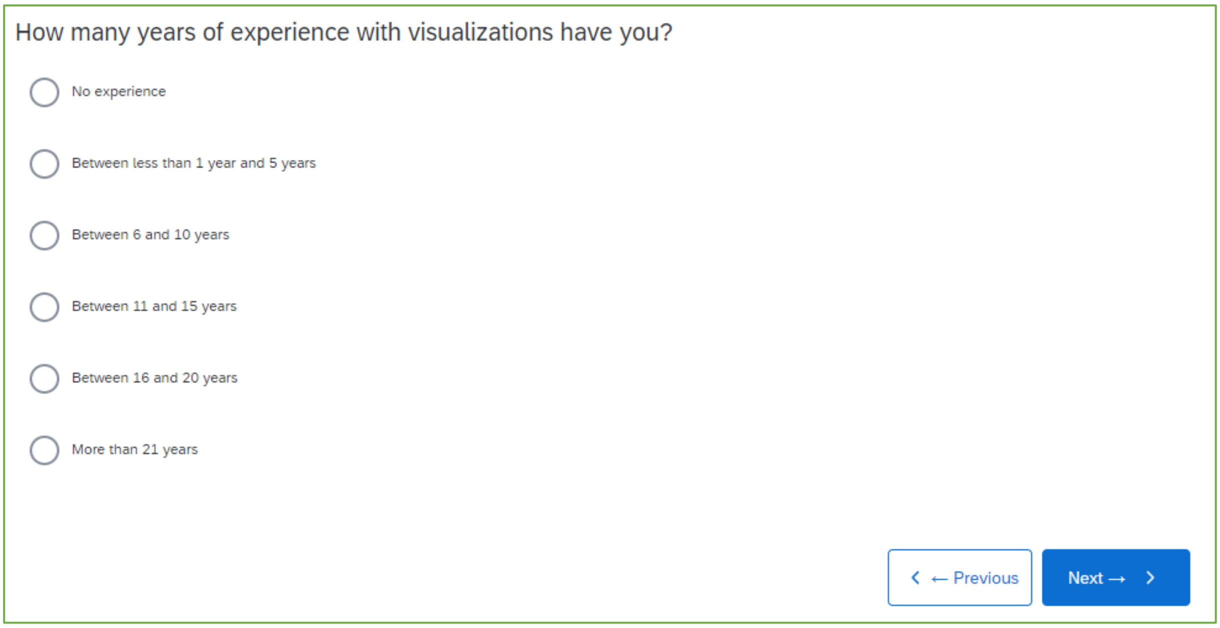


Figure 13 – Question 4 (demographic) of both surveys

The between-group design was chosen because it is cleaner and prevents the learning effect. Although this design can have high loudness of individual differences, in this experience this situation is avoided since the participants belong to the same team and have equivalent knowledge about plots and statistics.

4. RESULTS

4.1. SAMPLE DESCRIPTION

A total of 40 employees from the German Consulting department of Wayne Enterprises were recruited to complete the surveys, being randomly and equally split (20 to each survey). From these 40, 34 answered, who were reduced to 29 participants after cleaning the dataset. While the initial participants split was done to evenly distribute them to one or the other survey, once the dataset was cleaned, the HOPs survey gathered 13 participants, whereas the static plot survey gathered 16 participants.

Both surveys' participants have a similar age distribution (figures 14 and 15), with more than half of the participants having between 20 and 29 years old.

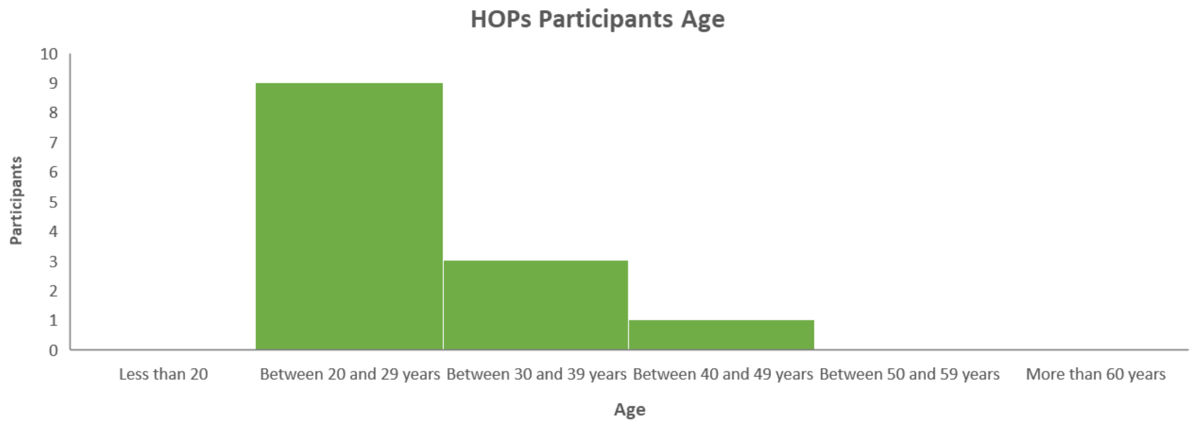


Figure 14 – Age from HOPs survey participants

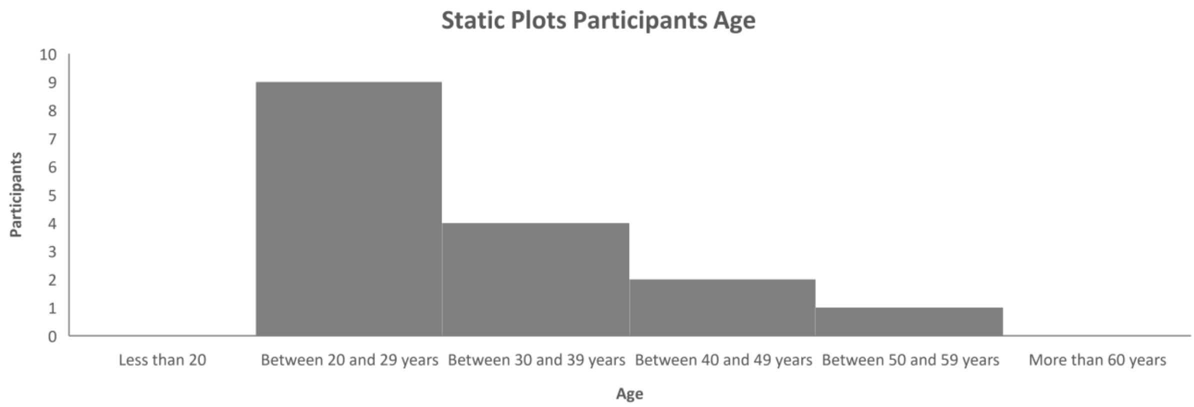


Figure 15 – Age from Static Plots survey participants

Regarding visualizations experience, in both surveys, the majority lays on between less than one year and 10 years. Nevertheless, the static plots survey group has a diverse experience, with at least one participant in each interval, while HOPs survey participants is between less than 1 year and 10 years, or more than 21 years of experience.

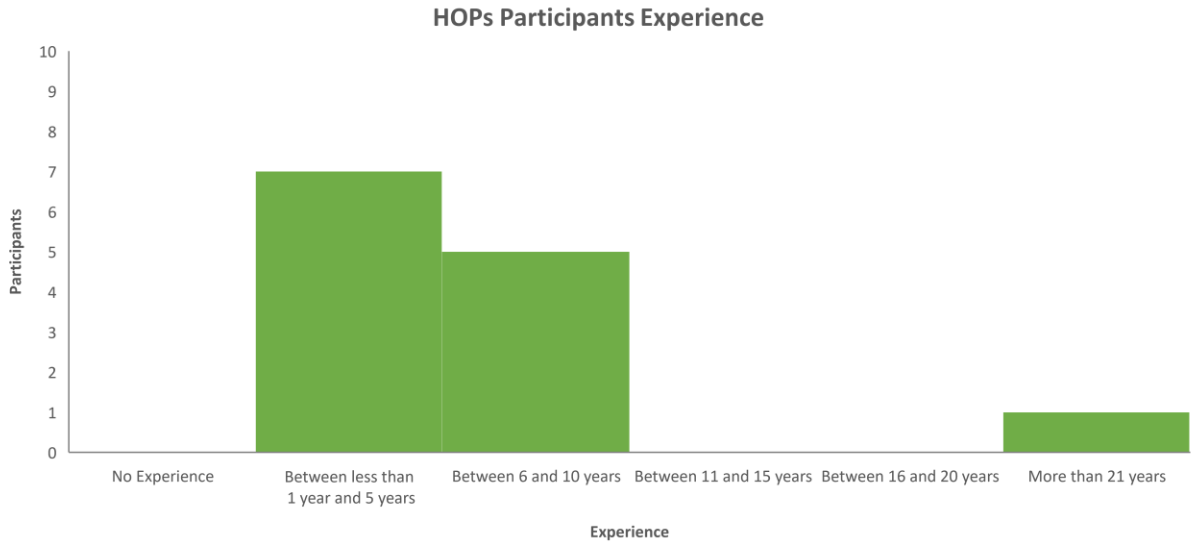


Figure 16 – Experience from HOPs survey participants

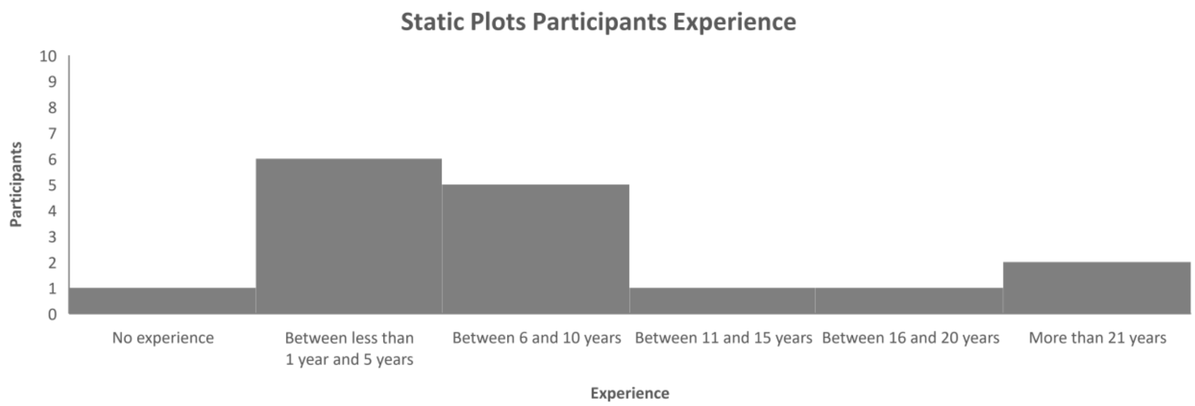


Figure 17 – Experience from Static Plots survey participants

4.2. ACCURACY AND TIME

The accuracy and time were the first measures to be analyzed, in order to get an overview of the results. The accuracy was calculated for questions 1 and 2 of both surveys, being measured in a scale of 0% to 100%. As referred before, the answers (impossible, very unlikely, unlikely, neutral, likely very likely, certain) were transformed into correct or wrong answers. Therefore, the accuracy percentage was calculated by dividing the sum of correct answers by the total of answers given in that specific question and survey. Five ranges of accuracy (Table 4) were defined: very low, low, medium, high, and very high.

Table 4 – Accuracy ranges definition

Accuracy ranges	Values
Very Low	0% - 20%
Low	21% - 40%
Medium	41% - 60%
High	61% - 80%
Very High	81% - 100%

According to the results obtained (figure 18), in both questions, the accuracy in Static Plots survey is higher than in the HOPs survey. The Static Plots survey had high accuracy, with the same accuracy percentage in both questions (75%), while the HOPs survey achieved low accuracy (31%) in the first question, and high accuracy (62%) in the second. This means that Static Plots survey obtained an average of 75%, which is considered high, and HOPs survey obtained an average of 46,5% which is medium.

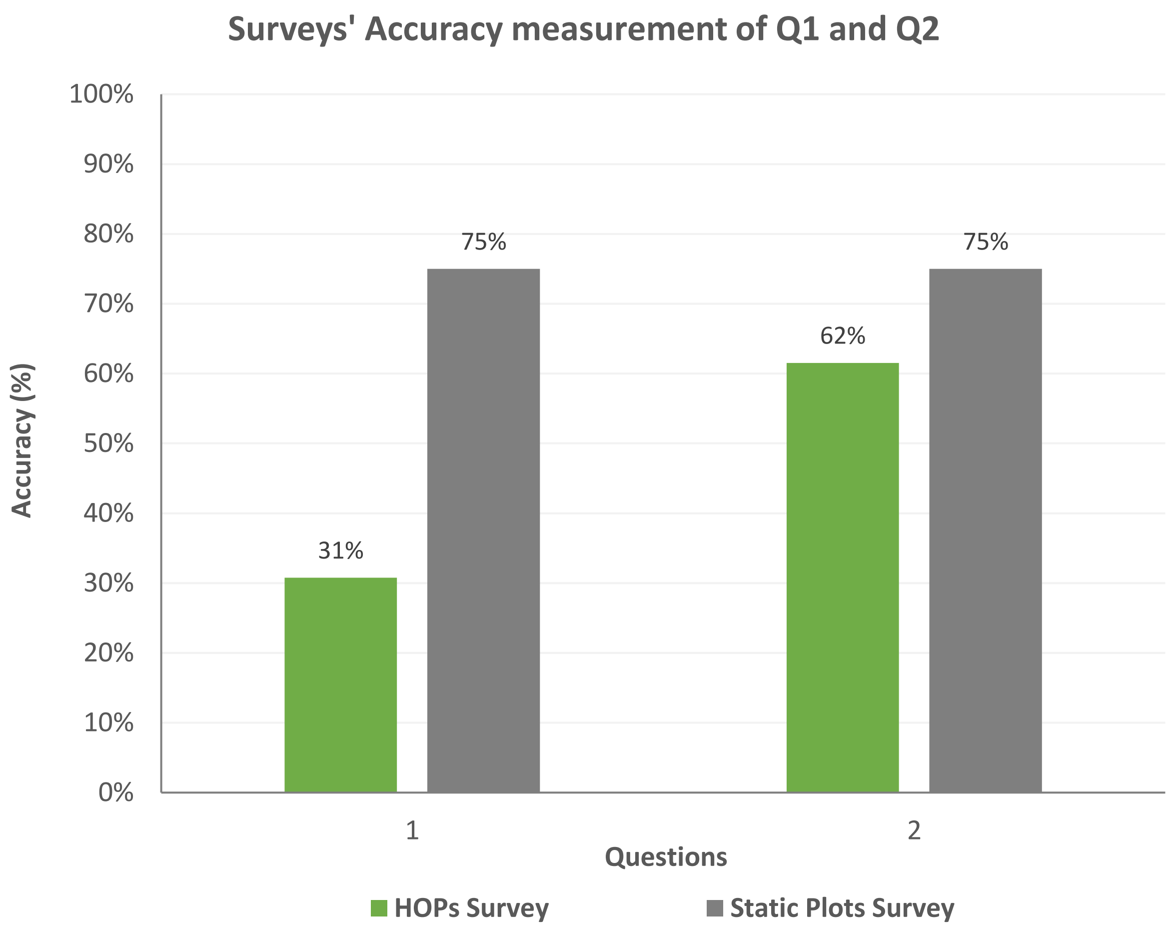


Figure 18 – Accuracy measurement of questions 1 and 2 from HOPs, and Static Plots Surveys

Regarding the time, it was observed the opposite scenario. The HOPs' survey participants completed the survey in less time than the participants of the Static Plots survey. In fact, as it is possible to observe in figure 19, HOPs' survey participants took on average less than half the time that the Static Plots survey participants took, i.e. 2 minutes and 21 seconds versus 4 minutes and 50 seconds, respectively.

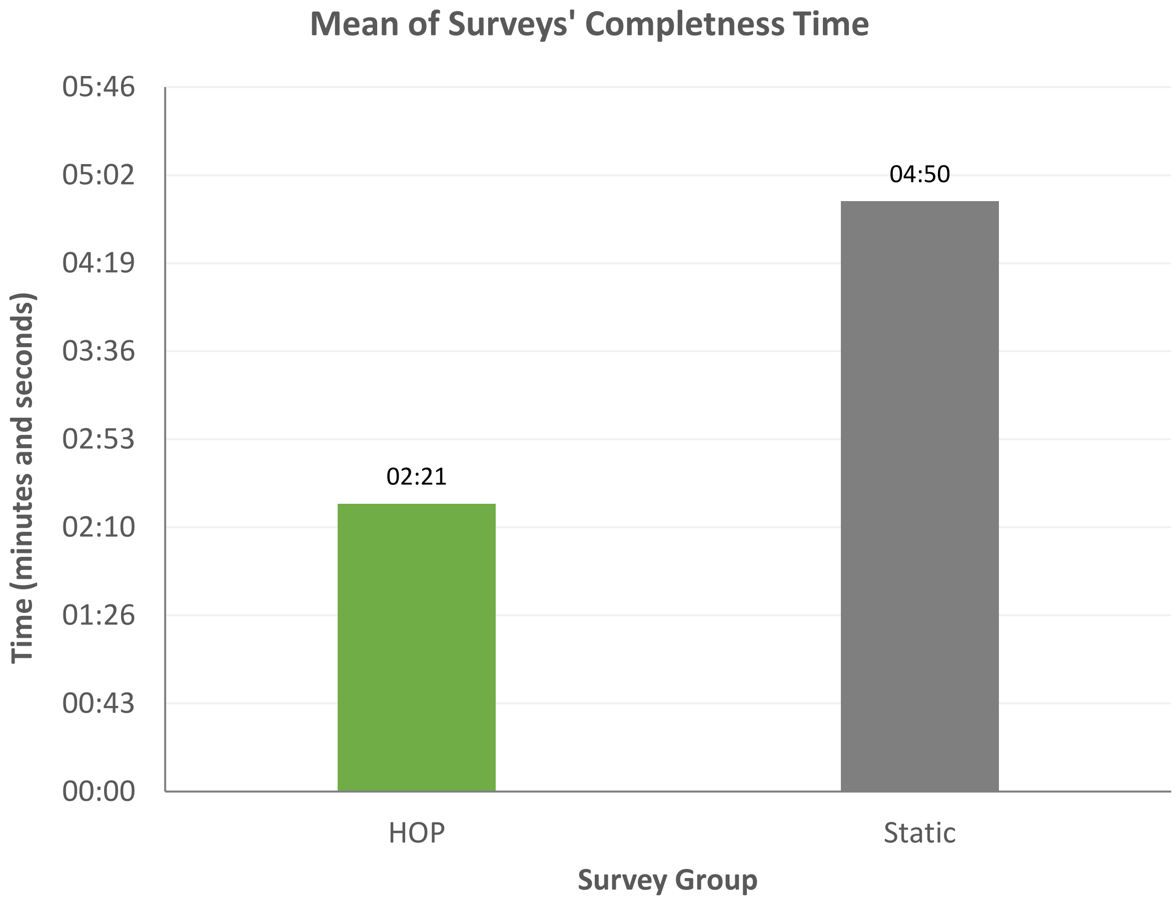


Figure 19 – Mean of total time taken to complete the surveys

4.3. Hypothesis Testing & Experience Impact

To test the null hypothesis (H_0), a Fisher's Exact Test was performed for each question. This statistical test was chosen because a normal distribution could not be assumed and the sample size was small. Furthermore, to perform this test the variable needs to be nominal, which is precisely what we have since the answers given were converted into binary answers (correct or incorrect). The objective is to check if there is a significant difference between the two groups, i.e., a difference between the static visualizations and the Hypothetical Outcome Plots interpretations. The significance tests were performed using IBM® SPSS® Statistics software, with $\alpha=0.05$. For both questions, the first step was to build a crosstabulation table (tables 5 and 6), so the Fisher’s test could be performed (results in table 7).

Table 5 – Survey group * Question 1 answers crosstabulation

		Question 1 answers		Total
		Incorrect	Correct	
Survey Group	HOP	9	4	13
	Static	4	12	16
Total		13	16	29

Table 6 – Survey group * Question 2 answers crosstabulation

		Question 2 answers		Total
		Incorrect	Correct	
Survey Group	HOP	5	8	13
	Static Plots	4	12	16
Total		9	20	29

Table 7 – Fisher's Exact Test results

	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Question 1	0.027	0.022
Question 2	0.688	0.353

Since for question 1 the p-value is less than 0.05, the null hypothesis can be rejected. Therefore, for this question, there is sufficient evidence to state that there is a significant difference between the static visualization and the HOP. With this result and the information present in figure 18, we can also reject the alternative hypothesis (H_1) which considered that HOPs allow to interpret uncertainty with more accuracy, since according to figure 18, the Static Plots Group answered to question 1 with 75% of accuracy, while the HOPs Group answered with only 31%.

However, for question 2, the Fisher's Exact Test results are the opposite, i.e., p-value is higher than 0.05, meaning that the null hypothesis cannot be rejected. Thus, for this question, there is not sufficient evidence to state that there is a significant difference between the static visualization and the HOP.

The difference between the two results leads to the question "did the participants' experience impact the answers to the questions?". To answer this question, four Kendall's Tau correlations were performed using IBM® SPSS® Statistics software. The Kendall's Tau is a non-parametric correlation coefficient which can be used to evaluate and test correlations between ordinal variables (Bolboaca & Jäntschi, 2006; Brossart et al., 2018).

Before performing the correlation, the experience was divided in three levels: low, medium, and high (table 8).

Table 8 – Accuracy ranges definition

Experience Levels	Experience years
Low	0 - 5
Medium	6 - 15
High	>16

After the experience levels definition, the correlation was performed. Tables 9 and 10 represent the Kendall's Tau correlation between experience and the answers to questions 1 and 2 for HOPs group and Static Plots group, respectively.

Table 9 – Kendall's Tau correlation between experience and Q1/Q2 answers for HOPs group

	Question 1 answers	Question 2 answers
Correlation Coefficient	0.146	0.138
Sig. (2-sided)	0.603	0.622
Sig. (1-sided)	0.301	0.311
N	13	

Table 10 – Kendall's Tau correlation between experience and Q1/Q2 answers for Static Plots group

	Question 1 answers	Question 2 answers
Correlation Coefficient	0.577	0.369
Sig. (2-sided)	0.019	0.133
Sig. (1-sided)	0.009	0.066
N	16	

As referred, the Kendall's tau-b correlation was run to determine the relationship between experience level and the answers of questions 1 and 2 for each group. For questions 1 ($\tau_b = 0.146$, $p = 0.603$) and 2 ($\tau_b = 0.138$, $p = 0.622$) from HOPs group and question 2 ($\tau_b = 0.369$, $p = 0.133$) from Static Plots group there was no correlation between the experience and answers given. However, there was a positive correlation between experience level and question 1 answers from Static Plots group, which was statistically significant ($\tau_b = 0.577$, $p = 0.019$).

5. DISCUSSION

In order to study the impact of Hypothetical Outcomes Plots in the decision-making process in a professional environment, two prototypes were built to support the proof-of-concept experiment conceived. This experiment consisted of two groups answering a survey, where one group observed the two prototypes (HOPs group) and the other group observed static figures (Static Plots group). The main objective of this experiment was to refute the null hypothesis and, consequentially, validate the alternative hypothesis, as well as enable to answer the research question formulated in section 1.2. ("How can Hypothetical Outcome Plots perform better than static visualizations when interpreted by experienced visualization users?").

Starting with the null hypothesis ("There is no difference in uncertainty interpretation between a static plot and a HOP"), according to the results in section 4, the null hypothesis can be rejected for question 1, but not for question 2. If the null hypothesis for question 1 was rejected, can the alternative hypothesis H_1 be validated ("HOPs allow to interpret uncertainty with more accuracy")? Unfortunately, no. According to section 4.2., the Static Plots group answered question 1 with higher accuracy than HOPs group. This interesting result might be due to several reasons. First, in the literature reviewed, the majority uses lay auditions, whereas this experiment uses experienced participants. Furthermore, the difference between the "impossible", "very unlikely", "unlikely", "neutral", "likely", "very likely", and "certain" intervals may have a higher subjectivity in HOPs survey, since it is more difficult to fix the values. Also, the fact that Hypothetical Outcome Plots' group accuracy increased from the first question to the second, suggests that a learning effect might have happened, i.e. since the participants are not used to perform analyses on moving images, it is likely that the first question was a "shock" and the second, because it was no longer a "novelty", was easier to answer correctly, indicating some sort of learning of how to interpret a Hypothetical Outcome Plot. Moreover, according to Hullman et al. (2015), people tend to perform more accurate judgements when the HOPs have more than one variable, which is precisely what happened on question 2 (question with two products), where HOP's group increased its performance when compared to question 1. Additionally, another reason for these results might be the experience in forecasts, which leads us to the experience impact.

Results show no correlation between experience level and the answers given of questions 1 and 2 from HOPs group, and question 2 from Static Plots group. However, there was a statistically significant correlation between experience and question 1 answers from Static Plots group. Overall, can be assumed that experience in visualization didn't impact the observed differences between the two Fisher's tests. Despite the importance of these results, it is important to remember that they derived from the question "How many years of experience with visualizations have you?", that is, the question focused on visualization experience, excluding other types of experience that may have influenced the participants' answers such as, for example, experience in statistics, or experience in forecasting.

One of the research sub-questions was “What is the relationship between time and the interpretation of different plots?”. Section 4.2. addressed this question by displaying the average time that each group took to complete the survey. Based on these results, it is possible to say the Hypothetical Outcome Plots’ group showed greater efficiency (the less time, the greater the efficiency) when compared to Static Plots group, since HOPs’ group took on average less than half the time that the Static Plots group took. In general, this efficiency topic has a huge impact on companies, it is what drives companies forward, because, as we normally hear, “time is money”, and money is (almost) everything in the business world.

6. CONCLUSIONS

The main purpose of this dissertation was to better understand how Hypothetical Outcome Plots could improve the decision-making process at Wayne Enterprises, in particular, the decisions made in launch/loss of exclusivity forecast projects. In order to accomplish this, first was conducted a literature review about data visualization, uncertainty, and Hypothetical Outcome Plots. The HOPs information gathered was vital for the development of the two prototypes, since it has enhanced the understanding of what is a HOP, the best practices to build one, as well as the main advantages and disadvantages. The prototypes were used as a support to the proof-of-concept, which consisted in the elaboration of two surveys with the objective of refuting the null hypothesis that stated: "there is no difference in uncertainty interpretation between a static plot and a HOP".

The two surveys were identical, both with four questions, being the first two regarding a specific visualization (static or HOP) and the other two questions regarding the participants' age and visualization experience years. The first two questions results were statistically different. While in the first question a statistical difference between the two surveys was observed, in the second no significant difference was observed. The difference between the results obtained from questions 1 and 2 can have several reasons. Nevertheless, it is very likely that all or most of them are related to the HOPs group, because in terms of accuracy, the Static Plots group kept the same accuracy from the first question to the second question while the HOPs group increased the percentage of accuracy significantly. This increase may suggest that there was a learning effect, i.e., since the participants were not used to perform analyses on moving images, it is likely that the first question was a "shock" and the second, because it was no longer a "novelty", was easier to answer correctly, indicating some sort of learning of how to interpret a HOP.

Another interesting result was the correlation between visualization experience and the answers given to questions 1 and 2. Overall, the visualization experience had no impact on the answers. Also, the time analysis gave important information. The HOPs' survey participants completed the survey in less than half the time that the Static Plots survey participants, showing greater efficiency.

Although the proof-of-concept results didn't provide evidence that HOPs can lead to a more accurate decision than a static plot, it was demonstrated that a decision made with a HOP can be similar to one made with a static plot. And if we consider that the HOP decision took less than half the time than the static plot decision, it is possible to state that the decision-making process of Wayne Enterprises can be improved by using HOPs.

7. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The biggest limitation that has been identified in this dissertation is the sample size, the data would have been significantly better if a larger sample size had been used. Nevertheless, it is important to remember that this study had a team from a real company as basis, and in the real world is difficult, if not nearly impossible, to find a team with more than 100 individuals and that the majority is willing to answer a survey.

Another limitation found was the virtual environment. Since the population of the study was the German Consulting Team, the participants' majority was not based in Portugal, and, therefore, the survey had to be taken online. If it was a face-to-face survey, other type of questions could have been made, and even an eye tracking machine could have been used, allowing further analyses.

In terms of future work, exploring the theoretical learning effect observed in the HOPs' survey participants might be a point of interest. In addition to checking whether there is indeed a learning effect, it would also be interesting to see if with more questions the percentage of accuracy tends to increase (as occurred from question 1 to 2), to stabilize, or if it varies.

Another suggestion would be to investigate the impact of the number of lines on visualization, i.e, the Hypothetical Outcome Plots' group results improved from question 1 to question 2, and the first HOP had only one line/product, and the second HOP had two lines/products, does the quantity of lines impacts how people perceived a HOP? That is something worth looking into.

Finally, my last recommendations would be, if possible, to have a bigger sample size, and count the time by question rather than the total survey.

8. BIBLIOGRAPHY

Ascough Ii, J., Maier, H., Ravalico, J., & Strudley, M. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modelling*, *219*(3–4), 383–399.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389.

Bolboaca, S.-D., & Jäntschi, L. (2006). Pearson versus Spearman, Kendall's tau correlation analysis on structure-activity relationships of biologic active compounds. *Leonardo Journal of Sciences*, *5*(9), 179–200.

Bonneau, G.-P., Hege, H.-C., Johnson, C. R., Oliveira, M. M., Potter, K., Rheingans, P., & Schultz, T. (2014). Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization* (pp. 3–27). Springer.

Brossart, D. F., Laird, V. C., & Armstrong, T. W. (2018). Interpreting Kendall's Tau and Tau-U for single-case experimental designs. *Cogent Psychology*, *5*(1), 1518687.

Chen, H. M. (2017). Information visualization. *Library Technology Reports*, *53*(3), 0024–2586.

Cooper, D. R., Schindler, P. S., & Sun, J. (2006). *Business research methods* (Vol. 9). Mcgraw-hill New York.

Correll, M., Moritz, D., & Heer, J. (2018). *Value-suppressing uncertainty palettes*. 1–11.

Fernandes, M., Walls, L., Munson, S., Hullman, J., & Kay, M. (2018). *Uncertainty displays using quantile dotplots or cdfs improve transit decision-making*. 1–12.

Fischhoff, B. (2012). Communicating uncertainty fulfilling the duty to inform. *Issues in Science and Technology*, *28*(4), 63–70.

Gaessler, F., & Wagner, S. (2019). Patents, data exclusivity, and the development of new drugs. *Review of Economics and Statistics*, 1–49.

Gershon, N., Eick, S. G., & Card, S. (1998). Information visualization. *Interactions*, *5*(2), 9–15.

Gkatzia, D., Lemon, O., & Rieser, V. (2016). Natural language generation enhances human decision-making with uncertain information. *ArXiv Preprint ArXiv:1606.03254*.

Gray, A., Wimbush, A., de Angelis, M., Hristov, P. O., Calleja, D., Miralles-Dolz, E., & Rocchetta, R. (2022). From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information. *Mechanical Systems and Signal Processing*, *165*, 108210.

Greis, M., Joshi, A., Singer, K., Schmidt, A., & Machulla, T. (2018). *Uncertainty visualization influences how humans aggregate discrepant information*. 1–12.

Grewal, Y., Goodwin, S., & Dwyer, T. (2021). *Visualising Temporal Uncertainty: A Taxonomy and Call for Systematic Evaluation*. 41–45.

Griethe, H., & Schumann, H. (2005). *Visualizing uncertainty for improved decision making*. *20*.

Han, P. K., Klein, W. M., Lehman, T., Killam, B., Massett, H., & Freedman, A. N. (2011). Communication of uncertainty regarding individualized cancer risk estimates: Effects and influential factors. *Medical Decision Making*, *31*(2), 354–366.

Haque, S., & Ratemi, E. S. (2017). Drug development and analysis review. *Pharmaceutical Chemistry Journal*, *50*(12), 837–850.

Higgins, M. J., Kronlund, M. J., Park, J. M., & Pollet, J. (2020). *The role of assets in place: Loss of market exclusivity and investment* (No. 0898–2937). National Bureau of Economic Research.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164.

Hofman, J. M., Goldstein, D. G., & Hullman, J. (2020). *How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results*. 1–12.

Hrabovskyi, Y., Brynza, N., & Vilkhivska, O. (2020). Development of information visualization methods for use in multimedia applications. *EUREKA: Physics and Engineering*, *1*, 3–17.

Hullman, J. (2016). *Why evaluating uncertainty visualization is error prone*. 143–151.

Hullman, J. (2020). Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 130–139. https://doi.org/10.1109/TVCG.2019.2934287

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 903–913. https://doi.org/10.1109/TVCG.2018.2864889

Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS One*, *10*(11), e0142444.

Jena, A., Engelke, U., Dwyer, T., Raiamanickam, V., & Paris, C. (2020). *Uncertainty visualisation: An interactive visual survey*. 201–205.

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied*, *18*(1), 126.

Joslyn, S., & LeClerc, J. (2013). Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, *22*(4), 308–315.

Jung, M. F., Sirkin, D., Gür, T. M., & Steinert, M. (2015). *Displayed uncertainty improves driving experience and behavior: The case of range anxiety in an electric car*. 2201–2210.

Kale, A., Kay, M., & Hullman, J. (2019). *Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths*. 1–14.

Kale, A., Kay, M., & Hullman, J. (2020). Visual reasoning strategies for effect size judgments and decisions. *IEEE Transactions on Visualization and Computer Graphics*.

Kale, A., Nguyen, F., Kay, M., & Hullman, J. (2018). Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 892–902.

Kamal, A., Dhakal, P., Javaid, A. Y., Devabhaktuni, V. K., Kaur, D., Zaientz, J., & Marinier, R. (2021). Recent advances and challenges in uncertainty visualization: A survey. *Journal of Visualization*, 1–30.

Kayongo, P., Sun, G., Hartline, J., & Hullman, J. (2021). Visualization equilibrium. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 465–474.

Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, *26*(6), 822–827.

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152–155.

Kim, Y.-S., Walls, L. A., Krafft, P., & Hullman, J. (2019). *A bayesian cognition approach to improve data visualization*. 1–14.

Kinkeldey, C., MacEachren, A. M., Riveiro, M., & Schiewe, J. (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: Systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, *44*(1), 1–21.

Kirchner, M., Mitter, H., Schneider, U. A., Sommer, M., Falkner, K., & Schmid, E. (2021). Uncertainty concepts for integrated modeling-Review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling & Software*, *135*, 104905.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

LeClerc, J. E., Ehlers, A. P., Agrawal, N., Alfonso-Cristancho, R., Flaxman, A. D., & Flum, D. R. (2016). *Understanding clinical and non-clinical decisions under uncertainty: A scenario-based survey*.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2496–2505.

Midway, S. R. (2020). Principles of effective data visualization. *Patterns*, 100141.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209–227.

Nagel, H. R. (2006). *Scientific visualization versus information visualization*. 8–9.

Padilla, L., Kay, M., & Hullman, J. (2020). *Uncertainty visualization*.

Padilla, L. M., Powell, M., Kay, M., & Hullman, J. (2020). Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations. *Frontiers in Psychology*, *11*.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *The Visual Computer*, *13*(8), 370–390.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Phelan, C., Hullman, J., Kay, M., & Resnick, P. (2019). *Some Prior (s) Experience Necessary*.

Politi, M. C., Han, P. K., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695.

Robertson, J., & Kaptein, M. (2016). *Modern statistical methods for HCI*. Springer.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis*.

Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten Simple Rules for Better Figures. *PLoS Computational Biology*, *10*(9), e1003833. https://doi.org/10.1371/journal.pcbi.1003833

Roy, P. T., Ricci, S., Cuenot, B., & Jouhaud, J.-C. (2018). Sounding spider: An efficient way for representing uncertainties in high dimensions. *ArXiv Preprint ArXiv:1808.01217*.

Shneiderman, B., Plaisant, C., Cohen, M. S., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson.

Therón Sánchez, R., Benito Santos, A., Santamaría Vicente, R., & Losada Gómez, A. (2019). *Towards an uncertainty-aware visualization in the digital humanities*. *6*(3), 31.

Tufte, E. (2006). *Beautiful Evidence*. Graphics Press LLC.

Wilke, C. O. (2019). *Fundamentals of data visualization: A primer on making informative and compelling figures*. O'Reilly Media.

Zhang, D., Adar, E., & Hullman, J. (2021). Visualizing Uncertainty in Probabilistic Graphs with Network Hypothetical Outcome Plots (NetHOPs). *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 443–453.

Zhu, Y. (2007). *Measuring effective data visualization*. 652–661.