LUND UNIVERSITY
School of Economics and Management

*Department of Informatics*

# Data-driven *Biased* Decision-making?

## Exploring the landscape between dashboards, visualization literacy and decision bias

Master thesis 15 HEC, course INFM10 in Information Systems

Authors: Kristoffer Bergram
Brian Ochan

Supervisor: Odd Steen

Examiners: Niklas Holmberg
Miranda Kajtazi

# Data-driven *Biased* Decision-making? Exploring the landscape between dashboards, visualization literacy and decision bias

**AUTHORS**: Kristoffer Bergram and Brian Ochan

**PUBLISHER**: Department of Informatics, Lund School of Economics and Management, Lund University

**PRESENTED**: June 2018

**DOCUMENT TYPE**: Master Thesis

**NUMBER OF PAGES**: 67

**KEY WORDS**: dashboard, data visualizations, decision bias, visualization literacy, representative heuristic, interpretation accuracy, dashboard-based data visualizations, field experiment, VLAT, base rate neglect, ecological correlation

**ABSTRACT**:
Data quantities and their sources have amplified over the years and so has the trend to employ dashboard-based data visualizations into the hands of a wider audience of end-users. By selecting four of the most common data visualization formats and combining these into a dashboard this thesis quantitively explored the relationship between similarity features of dashboard-based data visualizations, interpretation accuracy and systematic errors in decision-making i.e. decision biases as defined by Kahneman and Tversky (1974). By sampling 87 business practitioners through a double-blind randomized field experiment conducted at a large IT-company in Sweden, the objective of this thesis was to gauge the nature and extent of the relationship between dashboard-based data visualizations, interpretation accuracy and decision biases. The results of the field experiment did not suggest a relationship between similarity features of dashboard-based data visualizations and decision biases. The relationship between peoples' ability to interpret these data visualizations and decision biases was more nuanced, suggesting no overall bias while a difference between two natural groups with a *high* and *low* degree of interpretation accuracy could be demonstrated. The discussion highlights the implications of quantitatively analyzing systematic errors or decision biases that may arise inside the expanding territory of dashboard-based data visualizations.

# Acknowledgements

We want to express our gratitude to the organization that enabled us to produce this research. Although these people will remain confidential, special thanks should be extended to the liaison at the organization, the managers and all the participating employees that openly gave us their time and attention.

A special thanks to our families, loved ones and friends who have constantly encouraged and challenged us during this intellectual journey. We also want to thank our supervisor for compelling us to make sense when describing the nature of this research.

*This publication is part of my research work at Lund University, thanks to a Swedish Institute scholarship. / Brian*

# Contents


1. Introduction ........................................ 1
  - 1.1 Background ........................................ 1
  - 1.2 Problem Area and Delimitations ........................................ 2
  - 1.3 Research Questions ........................................ 3
  - 1.4 Purpose ........................................ 4

2. Theoretical Background ........................................ 5
  - 2.1 Dashboard-based Data Visualizations ........................................ 5
    - 2.1.1 Data Visualizations ........................................ 5
    - 2.1.2 Dashboards ........................................ 7
    - 2.1.3 From DSS to BI and Analytics ........................................ 8
  - 2.2 Decision-making and Biases ........................................ 9
    - 2.2.1 The Representative Heuristic ........................................ 10
  - 2.3 Theoretical Summary ........................................ 15

3. Methodology ........................................ 17
  - 3.1 Research Design ........................................ 17
  - 3.2 Development of Instruments ........................................ 19
    - 3.2.1 Development of the Dashboard-based Data Visualizations ........................................ 19
    - 3.2.2 Data, Context, Attributes and Units ........................................ 21
    - 3.2.3 Development of the Questions related to the Dashboards ........................................ 23
  - 3.3 Measurements ........................................ 25
    - 3.3.1 Two Dashboards (Declining and Growing) ........................................ 25
    - 3.3.2 Interpretation Accuracy (Low and High) ........................................ 26
    - 3.3.3 Decision Bias ........................................ 28
  - 3.4 Field Sampling Procedure ........................................ 29
  - 3.5 Sampled Participants ........................................ 29
  - 3.6 Data Collection Procedure ........................................ 30
  - 3.7 Statistics ........................................ 31
  - 3.8 Research Ethics and Quality ........................................ 32

3.8.1 Ethical Guidelines ........ 32
3.8.2 Debriefing ........ 33
3.8.3 Research Quality ........ 33

4. Data Analysis and Results ........ 35
4.1 Dashboard groups (Declining/Growing) & Decision Bias ........ 35
4.2 Interpretation Accuracy Groups & Decision Bias ........ 38

5. Discussion ........ 42
5.1 Limitations and Critical Considerations ........ 43
5.2 Implications for Research ........ 45
5.3 Implications for Practice ........ 47
5.4 Further Research ........ 47

6. Conclusion ........ 49

Appendix 1. Descriptive Statistics on the Interpretation Accuracy questions ........ 49
Appendix 2. Extraneous Analysis of Interpretation Accuracy questions ........ 51
Appendix 3. Data Collection Email (*in Swedish*) ........ 52
Appendix 4. Debriefing Email ........ 53
Appendix 5. Questionnaire Growing Dashboard Group ........ 54
Appendix 6. Questionnaire Declining Dashboard Group ........ 58

References ........ 62

# Figures

**Figure 2.1** 10th century time-series graph from De cursu per zodiac. ........................................6

**Figure 2.2** Example of illusion caused by attribute substitution. . ..............................................11

**Figure 2.3** Linear approximation of mean likelihood ranks plotted against mean ranks of similarity or "Representativeness" in the Tom W. study.. ..........................................................14

**Figure 2.4** Stylized example of personal spending categories and spending over time. ...............16

**Figure 3.1** Research design with two randomized groups with the dashboards shown in stylized formats along with the two natural groups ......................................................................18

**Figure 3.2** The twelve data visualization formats of the data visualization literacy assessment test. ..........................................................................................................................20

**Figure 3.3** The two versions of the data visualizations that were used for the dashboards...........23

**Figure 3.4**. The two dashboard-based data visualizations of that were used in the study............26

**Figure 4.1** Proportion of participants that had attended university for more than two years and the gender distribution of the Declining (left) and Growing (right) Dashboard groups. .....................36

**Figure 4.2** Scatterplots of the two Dashboard Groups' Average Ranks correlated with the Correct Base Rates. ...................................................................................................................38

**Figure 4.3** Proportion of participants that had attended university for more than two years. .......39

**Figure 4.4** Scatterplots of the two natural groups' Average Ranks correlated with the Correct Base Rates. ...................................................................................................................41

# Tables

**Table 2.1** Judged base-rates of nine graduate specialization areas along with likelihood and similarity rankings ........................................................................................................................13

**Table 4.1** Average rankings of the eleven main distribution centers in the dashboards for the two Dashboard groups and Pearson product moment correlations with the correct base rates ............37

**Table 4.2** Average rankings of the eleven main distribution centers in the dashboard from the two natural groups (Low/High) Interpretation Accuracy ..............................................................40

# 1. Introduction

*Failure is an amazing data-point that tells you which direction not to go* – Payal Kadakia

## 1.1 Background

In our everyday lives, the decisions we make, the conclusions we reach, and the explanations we offer are often based on our judgments of the likelihood of uncertain events (Kahneman & Tversky, 1972). This is still true in the context of business: Which marketing campaign contributed the most the increase in sales? Did the installation of the new machine produce the decline in this year's production? Does the investment in a new Business Intelligence (BI) software explain the increase in profits? These are all questions that managers and business practitioners have to make judgements on, often without having all the necessary data readily available. Even if all the required data for such complex decisions were available, it is still challenging to extract, transform, load and display this data in way that aids our decision-making (Phillips-Wren et al., 2015). In the era of Big Data, dashboard-based data visualizations are becoming more of a norm as organizations traverse from intuition-based decision-making to data driven decision making (Abbasi et al., 2016).

According to Yigitbasioglu and Velcu (2012), many information systems (IS) such as Enterprise Resource Planning (ERP), applications for diverse performance scorecards and Business Intelligence (BI) software compete for the attention of different kinds of business practitioners. This process often conspires to what is known as information overload. Dashboards are a common way to tackle this issue as they make extensive use of data visualizations to help managers quickly find which areas worth further detailed analysis and which do not (Raschke & Steinbart, 2008). However, the tradition of consolidating data into a dashboard for only mangers or Business Intelligence (BI) specialists is disappearing as these tools move into the hands of generic business practitioners (Negash & Gray, 2008). As data quantities and their sources have amplified over the years, people's propensity to explore this data, find meaningful information from this data and visually represent the findings have also increased (Lee, Kim & Kwon, 2017b).

Yigitbasioglu and Velcu (2012) highlight that as a tool, dashboard-based data-visualization is expected to improve decision-making by amplifying human cognition and capitalizing on people's perceptual capabilities. While several scholars highlight that these tools are well received across many different organizations and that the interest around them is growing (e.g. Negash & Gray, 2008; Yigitbasioglu & Velcu, 2012; Abbasi et al., 2016), this paper is focused on potential adverse effects of dashboard-based data visualizations. A vast body of research contends that our human cognition has a strong tendency to make systematic errors in judgements under uncertainty (Hayes et al., 2014; Heath & Tversky, 1991; Kahneman & Tversky, 1972, 1973; Roumbanis, 2017; Tversky & Kahneman, 1974). Business decisions about uncertain events, such as the ones eluded to earlier certainly depend upon human cognition. Griffith et al. (2008) argue that technology can be thought of as any physical tool external to human cognition that is also used to aid human cognition. While these technological aids or Information System (IS) artifacts serve to amplify decision-making within organizations, this research will explore the landscape between certain decision aids such as dashboard-based data visualizations and systematical error in decision-making. Through their seminal research, Tversky and Kahneman (1974) laid the foundations for what would become one of the most influential research programs in behavioral science: The heuristics and biases approach (Mussweiler & Strack, 1999). The objective of this study is to extend Tversky and Kahneman’s (1974) research on heuristics and biases to the context of dashboard-based data visualizations inside a large-scale IT organization and explore whether comparable results will be obtained. The results of this research may problematize the expanding usage dashboard-based data visualizations and produce implications for both research and practice.

## 1.2 Problem Area and Delimitations

Data quantities and their sources have amplified over the years and so has the trend is to employ dashboard-based data visualizations into the hands of non-expert business practitioners (Negash & Gray, 2008; Lee et al., 2017b). Non-expert or generic business practitioners simply mean that these decision aids (dashboard-based data visualizations) are more frequently being used by a wider audience rather than just data analysts and specific managers. According to Arunachalam et al. (2002), researchers and system designers have been working on the problem of how to visualize data in an optimal way since the 1920’s. This area of research has produced many tenants and guidelines for how to optimally display quantitative data i.e. visualizing it for potential decision-makers (Bera, 2016; Kovalerchuk, 2001; Lurie & Mason, 2007; MacGregor & Slovic, 1986).
But poor design choices of data analysts and designers can negate the potential effectiveness of data visualizations as decision aids by causing the viewers to form erroneous initial perceptions (Raschke & Steinbart, 2008). Rather than erroneous perceptions, this study concerns whether systematic errors in decision-making arise when business practitioners are making decisions based on dashboard-based data visualizations in a real-world setting, even though they have a clear understanding of these visualizations. By selecting four of the most common data visualization formats and combining these into a dashboard this study will explore the tendency of business practitioners to make biased decisions as defined by Tversky and Kahneman (1974) when interpreting this dashboard.

The delimitation of this research concerns whether a particular decision bias known as base rate neglect can be detected when business practitioners make decisions based on four specific data visualization formats (bar chart, line chart, pie chart and bubble chart) in the context of real-world department meetings at a large IT-organization in Sweden. This intersection is what will be quantitatively explored in this study.

Boy et al. (2014) highlight that when data visualizations are designed or when evaluations of new visualization systems are being conducted, it is important to be able to pull apart the potential effectiveness of the data visualization and the actual ability of users to understand them and use them as a decision aid. To take this concern into consideration, a randomized, double-blind field experiment will be conducted at a large IT organization in Sweden where business practitioners make interpretations and judgements based on two independent dashboards during their department meetings. The degree of decision bias will then be analyzed for these two independent groups. Afterwards, two natural groups will be constructed: One with a Higher level of interpretation accuracy and one with a Lower level of interpretation accuracy. The degree of decision bias will then also be analyzed in relation to these two natural groups. However, it is important to highlight Bhattecherjee's (2012) sentiment that exploratory research such as this study may not lead to the most exact understanding of the target problem but may be valuable in scoping out both the nature and extent of the problem. The target problem of this research concerns the relationship between dashboard-based data visualizations, interpretation accuracy and decision biases.

## 1.3 Research Questions

The focal point of this exploratory study has a narrow scope where four common data visualizations are combined into a dashboard to explore one specific decision bias among business practitioners working at a large IT-organization in Sweden. The first objective is to explore whether similarity features of dashboard-based data visualizations can influence the tendency of business practitioners to make biased decisions i.e. systematical errors in decision-making as defined by Tversky and Kahneman (1974) and secondly, to explore the relationship between how accurately dashboard-based data visualization are interpreted and potential decision biases. This research will investigate the following two questions:

> Is there a relationship between decision bias and the similarity features of dashboard-based data visualizations such as declining or growing data categories?
>
> Is there a relationship between decision bias and the ability to accurately interpret the data visualizations of a dashboard?

In the above context other applicable synonyms would be "association" or "link" etc. as the notion of a "relationship" is defined in a quantitative setting.

## 1.4 Purpose

The purpose of this research is to quantitatively explore the relationship between dashboard-based data visualizations, their interpretation and potential decision biases. If graphical similarity features of data visualizations have an impact on how accurately dashboards are interpreted or how biased judgements they tend to produce, this has implications for both research and practice. In this context, similarity features emerge when the data visualizations of a dashboard are consistently ordered from highest to lowest, lowest to highest or whether a trend happens to be increasing or declining, thereby creating impressions of growth or decline in the overall dashboard.
Some of these features are under the influence of data analysts such as ordering of data categories, while an increasing or declining trend cannot be manipulated to the same degree by the data analysts' design choices. The reason this area is important to investigate is because dashboards are increasingly being used as a decision aid in more general contexts. Given the task-specific usage and appearance of dashboards, if erroneous decisions are made based on dashboard-based data visualizations, these errors have widely different consequences. Today, dashboards are heavily relied upon in many organizations and are at least employed across domains such as marketing, sales, finance, education and cyber security (Aljohani et al., 2018; Krush et al., 2013; McKenna et al., 2016; Skorka, 2017). Large-scale organizations are also moving this needle. Apple has now revealed a new "privacy dashboard" where the target audience is their whole European customer segment in the effort of complying with Europe's General Data Protection Regulation or GDPR (Hern, 2018). Given this wide degree of adoption, the intersection of decision biases and dashboards calls for further scientific attention.

The intended knowledge contribution is to scope out the nature and extent of a target problem i.e. the relationship between dashboard-based data visualizations, interpretation accuracy and decision biases. Depending on the results of this study, further reasons will be generated for researching this important domain. In order to immerse deeper into this area, several concepts such as data visualizations, dashboards, decision-making and potential heuristics and biases associated to human judgment will need further examination. These important concepts and their theoretical properties will be highlighted in the next chapter.

# 2. Theoretical Background

This chapter focuses on the theoretical background relevant to this research. In the first section, the concept of dashboard-based data visualizations will be broken down into its basic parts: data visualizations and dashboards. This is followed by a brief outline of evolution of their organizational usage setting, known as Decision Support Systems (DSS), Business Intelligence (BI) and Analytics. The second section concerns the key tenants of human decision-making and the cognitive processes known as heuristics and biases. The third section summarizes the key theoretical aspects of this explorative study.

## 2.1 Dashboard-based Data Visualizations

The notion of dashboard-based data visualizations will now be broken down further so that the concepts of "data visualization" and "dashboard" can be explained. The section begins with a brief trace of the historical view of data visualizations later uniting them into their application in the IS discipline. Finally, in light of that prior information, the dashboard concept will be addressed.

### *2.1.1 Data Visualizations*

Data visualization is a suitcase-term used for describing the results and the process of creating a visual representation of data (Card et al., 1999). In this paper, "data visualizations" refer to the results, such as charts, graphs or diagrams rather than the process of making them. In the Handbook of Data Visualization, Friendly (2008) traces the earliest seeds of visualization as having arisen in geometric diagrams, in tables of the positions of stars and other celestial bodies, and in the making of maps to aid in geographical navigation and exploration. He underscores what appears to be the earliest graphical depictions of quantitative information in the 10th (or possibly 11th) century "multiple time-series graph" for the changing positions of seven heavenly bodies over the zodiac. Figure 2.1 shows this humble yet complex beginning of data visualizations.

As part of the appendix of a text for monastery schools, the graphs exact relation to the underlying data is somewhat hard to discern. Funkhouser (1936) notes that the horizontal axis appears to have been chosen for each planet individually since the periods cannot be reconciled while the movements of the sun also appears disconcertingly wavy. Tufte (1983) adds that the diagram featured in Figure 2.1. on the next page is an enigmatic and lonely wonder in the history of data visualization since the next surviving data graphic does not appear until 800 years later.

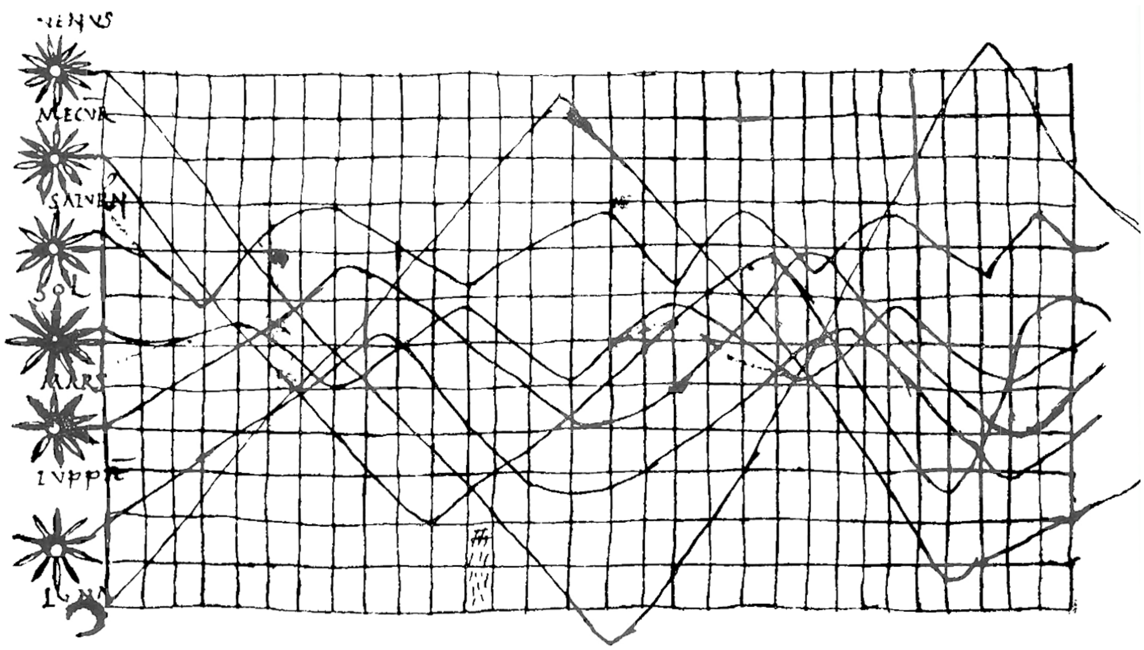


***Figure 2.1*** 10th century time-series graph from De cursu per zodiac. Adopted from "A note on a tenth century graph" by Funkhouser, 1936, Osiris, 1, p. 261. Published by Saint Catherines Press, University of Chicago Press & History of Science Society

Today, visual displays provide the broadest bandwidth channel from a computer to a human in terms of perception (Ware, 2013) and graphic displays of information are an important link in the design of user/ machine interfaces. According to Baker et al. (2009), the need for effective information visualization is high. We live in an environment where managers and knowledge workers need to make decisions based off their ever-increasing information sources and the business decision-makers of today face the task of sorting through the jungle of data created by IS. As Tegarden (1999) explains, visualization technologies allow the business decision-maker to separate the "wheat from the chaff." Many scholars seem to converge on the belief that there is not 'one' optimal format for data visualizations, but that the effectiveness of the visualization depends on the type of task that has to be performed (Boy et al., 2014; Speier, 2006; Tufte, 1983; Vessey, 1991).

Vierck (1981) states that visualization is one method being employed to manage big amounts of data and consequently aid in decision-making within organizations. Amer and Ravindran (2010) correspondingly agree that graphical displays that visualize data are now being widely used as decision aids in many different computing environments. Pushing this argument even further, Segel and Heer (2010) argue that data visualizations are regularly promoted for their ability to reveal stories within data, that could potentially aid in decision making.

Scholars also seems to converge on the fact that data visualizations aid in decision making, with one caveat. Meyer et al. (1999) argue that for information visualizations to aid decision support, the information structure visualized must be consistent with the decision-makers mental representation of a decision problem, also known as task-specific. This notion is widely covered by Vessey (1991) in what is now referred to as the Cognitive Fit Theory and adding credence to this notion, Yigitbasioglu and Velcu (2012) concur that users arrive at decisions very much based on cognitive processes and dashboards are expected to further improve decision-making by augmenting cognition and capitalizing on human's perceptual capabilities.

### *2.1.2 Dashboards*

Dashboards are nothing new, and in many instances the term "dashboard" refers to an Executive Information System (Few, 2006). According to Yigitbasioglu and Velcu (2012), the information dashboard is an outlet for businesses to disclose critical performance metrics and measures. They tie the dashboard to Kaplan and Norton's (1996) concept of the balance scorecard. In a similar manner, Few (2006) draws the same comparison, relating a dashboard to the balance score card, particularly as a new approach to management that involves identification, monitoring and the use of key performance indicators.

There is no one uniform definition of a dashboard as Few (2006) notes, but for this study, his definition of a dashboard will be adopted. Few (2006, p. 34) defines a dashboard as a "visual display of the most important information needed to achieve one or more objectives that is consolidated and arranged on a single screen, so the information can be monitored at a glance." However, it should be noted that this is an idealized definition because if a dashboard in fact had all of the most important information that was needed to make certain decisions, those decisions would arguably be good candidates for computerization and automation (Griffith et al., 2008). Simon (1960) points out that when a decision is very routine, when the constraints of the decision and the possible choice options are well known beforehand, that decision fits the criteria for automation. Dashboards certainly do not fit that description. There would not be much of a reason to create dashboards if the decisions that they were to assist were good candidates for automation. Rather, the information that dashboards provide are meant to aid managers and business practitioners in more complex decision tasks that are not easily automated. There are many uses of dashboards and Few (2006) puts forward a couple. First, he argues that dashboards offer a unique and powerful means for organizations to present their information and data. Secondly, he positions dashboards as communication aids within organizations. Likewise, Bera (2016) looks to dashboards as a tool to assist their users in visually identifying trends, patterns, and anomalies in order to make effective decisions while Yigitbasioglu and Velcu (2012) argue that dashboards can be viewed as data driven DSS, providing information in a certain format to the decision-maker.

Today, dashboards are heavily relied upon in many organizations and are employed across several domains such as marketing, sales, finance, education and even cyber security to mention but a few (Aljohani et al., 2018; Krush et al., 2013; McKenna et al., 2016; Skorka, 2017). As numerous organizations have implemented data driven decision-making processes, embedding this kind of analytical processing into an organizations culture can advance its competitiveness (Baysal et al., 2013). Much as dashboards offer a multitude of advantages as discussed earlier, they often fall short of their potential and ultimately intentions (Few, 2006). In the earlier section on visualization, we mentioned that data visualizations are sometimes promoted for their ability to reveal stories in data, a sentiment shared by Segel and Heer (2010) who term this as 'narrative visualizations', that is visualizations intended to convey stories. This is where the true power of visualizations combined into a dashboard begins to manifest. As Pappas and Whitman (2011) argue, combining different visual formats such as tables, graphics and key performance indicators into a dashboard is a step in the right direction to aid decision-makers in making fact-based decisions. However, as shown by both Spier (2016) and Vessey (1991), this is not a straightforward process and organizations have spent several decades on addressing this challenge.

### *2.1.3 From DSS to BI and Analytics*

Now, the organizational setting under which both data visualizations and dashboards converge will be outlined. Decision-making is a complex process in any organization, a sentiment echoed by Hall (2008). The need to use computerized systems to aid decision-making stretches back to the 1970's as seen in Hosack et al.'s (2012) *History of Decision Making* even though Power (2008) places this in the mid-1960s, attributing it to the development of minicomputers and distributed computing systems. According to Hall (2008), these computerized systems that aid decision-making are commonly referred to as "DSS". Both Holsapple (2008) and Watson and Wixom (2007) argue they were the first applications designed to support decision-making. Arnott (2006) and Arnott and Pervan (2014) place DSS in the area of IS devoted to supporting and improving people's decision-making processes. DSS's according to Holsapple (2008) are defined in terms of the roles they play in decision processes. A simple definition of a DSS according to Holsapple (2008, p. 22) is "a system that somehow assists in decision-making by processing and representing knowledge in ways that allow decision making to be more productive, agile, innovative and reputable".

This takes us to the domain of BI and Analytics, a somewhat fancier and more nuanced term to DSS. According to Chaudhuri et al. (2011), as the costs of data acquisition and data storage have continued to plummet throughout the years, BI software inevitably have become widely adopted by many organizations. They claim that it is extremely difficult to find a successful enterprise that has not leveraged BI technology for its business today. Business Intelligence has its roots firmly implanted in DSS. Going back to Hosack et al.'s (2012) outline of the evolution of DSS, BI began to take ground in the early 2000s. This can also be seen in Arnott and Pervan (2014) *The genealogy of the DSS field*. Negash (2004) adds credence to the notion that BI has its roots in DSS, albeit in one or more disguised forms.

Defining BI as a collection of decision support technologies for the enterprise, Chaudhuri et al. (2011) posit BI enables executives, managers and analysts to make better and faster decisions. However, Chen et al. (2012) extend the relationship between BI and Analytics, together calling them Business Intelligence and Analytics (BI&A) although Arnott and Pervan (2014) say this label did not gain widespread traction as a DSS movement until the early 2000s. From Arnott and Pervan (2014) *The genealogy of the DSS field* Business Analytics or "BA" can be understood to have branched off BI.

Chen et al. (2012) postulate three stages of BI&A evolution as discussed subsequently. Firstly, BI&A 1.0 is characterized by its data-centric approach. This stage according to Chen et al. (2012) has its roots in the database management field. They explain that it relies heavily on data collection, extraction and analysis technologies. The data at this level is mostly structured and sourced from a multitude of legacy systems. Secondly, they explain that BI&A 2.0 relies heavily on the advent and popularity of the internet and the web as a data source. (Chen et al., 2012). A key characteristic of this stage they explain is the reliance on unstructured web content pulled from. Finally, they place the advent of BI&A 3.0 on the adoption of real time data collection from devices such as mobile phones and smart appliances.

As Lee et al. (2017b) argue, it's clear to see that as data quantities and their sources have amplified over the years, peoples' propensity to explore this data, find meaningful information that aids decision-making has increased. One way to make sense of all this collected data and information is through BI&A. Today, DSS and BI&A have become firmly ingrained in many organizations mostly because data is the most important resource an organization has at its disposal to facilitate decision-making (Hall, 2008; Holsapple, 2008; Negash, 2004). Chen et al. (2012) argues that one of the key capabilities of BI&A is dashboards and data visualizations. As mentioned by Zhang and Whinston (1995), as more and more data is becoming available, the primary challenge is the presentation of this collected data in a comprehensible form to support decision-making. These tools are meant to aid decision-making but as the next section will highlight, human decision-making turns out to be a territory with many pitfalls.

## 2.2 Decision-making and Biases

To understand decision-making and the biases associated with it, an outline of the decision-making process, decisions and biases and is needed. The research area of decision-making has probably been around longer than the disciplines of management and leadership (Bennet & Bennet, 2008). As explained by Boland (2008), decision making is concerned with the process of assessing alternate courses of action and making a choice among them. He also notes that this is a rather complex task fraught with several difficulties. There is no common conception of what exactly is constituted by a "decision". One can problematize and highlight the complexity of the concept by asking who, when, how, should etc., while also associating the outcome of decisions to desired objectives (Churchman, 1968).

While this kind of problematization certainly has a place, it is outside the scope of this paper. In this context, two definitions will suffice: Simon (1960 cited by Holsapple, 2008 p. 26) in the *Handbook of Decision Support Systems* defines a decision as “a choice about action” and Fishburn (1964 cited by Holsapple, 2008 p. 26) refers to a decision as “the choice for the strategy for action”. These two definitions of what constitutes a decision will be the focal point of this paper. As definitions they are well aligned with one of the most influential areas of research in behavioral science: The biases and heuristics approach (Mussweiler & Strack, 1999). In this research area, decisions are also examined as an outcome or effect.

Despite the significant contribution from this area of research to many other scientific fields, one of its main criticisms is that the effects on decision-making are given more attention rather than the underlying processes. However, this criticism is not equally valid for all biases and heuristics according to Mussweiler and Strack (1999). The originators of the biases and heuristics approach: Tversky and Kahneman (1974) outline that many decisions are based on beliefs concerning the likelihood of uncertain events and people rely on several judgmental heuristics to reduce complex decision tasks of prescriptions and predictions. In this context, judgmental heuristics are comparable to a cognitive process, principle or as a mental rule-of-thumb. The biases and heuristics approach explores the territory that economist Herbert A. Simon coined as “bounded rationality” and in 2002, Daniel Kahneman was awarded the Nobel Prize in Economic Sciences for his and Amos Tversky’s research concerning human judgement and decision-making under uncertainty (Nobel Foundation, n.d.; Kahneman, 2003). One of the main tenants of their research is that these judgmental heuristics are very useful in general while they can also lead to severe and systematical errors in decision-making (Tversky & Kahneman, 1974). These systematical errors are what define decision biases. Some of these systematical errors are conserved in what is known as Prospect Theory (Tversky & Kahneman, 1981). However, this latter theory is outside the scope of this thesis since the focal point is on the decision biases that arise from a specific heuristic known as Representativeness.

### 2.2.1 The Representative Heuristic

Let’s start by following Kahneman’s (2003) outline of what a judgmental heuristic is. As a cognitive process, heuristics are intimately connected to the notion of attribute substitution. The general notion is that when people are confronted with a difficult cognitive task they sometimes solve an easier one (task) instead. In more technical terms, a decision is mediated by a judgmental heuristic when a person assesses a specified target attribute of a decision by substituting a related heuristic attribute that just comes across more readily to mind. Figure 2.2 on the next page serves as a good illustration of how attribute substitution works.

***Figure 2.2*** Example of illusion caused by attribute substitution. Adopted from “Maps of Bounded Rationality: Psychology for Behavioral Economics” by Kahneman, 2003, The American Economic Review, 93, 5, p. 1460. Photo by Shoham, 2003.

Now, which horse is larger? Intuitively, the horse at top of the picture seems larger but they are in fact the exact same size. The target attribute here is two-dimensional size on a page yet most of us have a strong tendency to use our heuristics for three-dimensional size. Kahneman (2003) explains that a 3D-impression is what comes to mind for people that are not thoroughly trained in making these kinds of judgments (such as professional painters and photographers).

The representativeness heuristic works the same way but in a different context. It refers to the use of representativeness or “similarity” as a heuristic attribute to judge the probability or likelihood of an outcome. Tversky and Kahneman (1974, p. 1124) summarize the representative heuristic by outlining that people are often concerned with the following kinds of questions about uncertain events or processes: What is the likelihood that object A belongs to class B? What is the likelihood that process A will produce event B? If A turns out to be highly representative or similar to B, people tend to decide that the likelihood that A has produced event B is high. Such decisions are said to be relying on the representative heuristic. As a heuristic, it is used across a wide array of decision tasks (Kahneman & Tversky, 1972).

According to Busenitz (1999), the representative heuristic is among one of the most widely referenced heuristics in psychology literature. In general terms, decision-makers using this heuristic are willing to develop broad and sometimes very detailed generalizations about a thing or process based on only a few attributes of that thing or process. There is a lot of empirical evidence supporting that many kinds of decisions are made based on representativeness (Kahneman & Tversky, 1972; Tversky, 1977; Busenitz, 1999; Kahneman, 2003).

The following example provides a good illustration of how the representative heuristic works in practice and the decision bias associated with it. In one study, Kahneman and Tversky (1973) had three groups of participants (base rate, likelihood and similarity). The results of this study are summarized in Table 2.1. on the next page. The first group (*base rate$_1$*) was asked to write down their best guesses about the percentage of students that were enrolled in 9 different graduate specialization areas. Kahneman and Tversky (1973, p. 238) presented the second and third group (*likelihood$_2$ and similarity$_3$*) with the following description of a person:

> Tom W. is of high intelligence, although lacking in creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive…

The second group gained additional information that this description had been authored by a psychologist based on a personality test that Tom had taken five years earlier, when he was in high school. The participants were then asked to rank each graduate specialization based on the likelihood that Tom was now studying them i.e. they were asked to make a prediction. The third group was just asked to rank the nine areas by how similar Tom was to the typical graduate student in each group without receiving the additional information regarding where the description came from.

***Table 2.1*** Judged base-rates of nine graduate specialization areas along with likelihood and similarity rankings

| Graduate Specialization Area | 1. Average judged base rate (in %) | 2. Average likelihood rank | 3. Average similarity rank |
|---|---|---|---|
| Business Administration | 15 | 4.3 | 3.9 |
| Computer Science | 7 | 2.5 | 2.1 |
| Engineering | 9 | 2.6 | 2.9 |
| Humanities and Education | 20 | 7.6 | 7.2 |
| Law | 9 | 5.2 | 5.9 |
| Library Science | 3 | 4.7 | 4.2 |
| Medicine | 8 | 5.8 | 5.9 |
| Physical and Life Sciences | 12 | 4.3 | 4.5 |
| Social Science and Social Work | 17 | 8.0 | 8.2 |
| *Correlations between group rankings* | ***- 0.65*** | ***0.97*** | |

***Note:*** Adopted from the findings of Kahneman and Tversky (1973, p. 238)

Kahneman (2003) later summarized these findings in the following way: In these situations, people make predictions about likelihood in essentially the same way ($r = 0.97$) as they judge similarity, i.e. the participants made decisions regarding likelihood based on how representative the graduate specializations were of Tom's description. Figure 2.3 on the next page highlights this strong association with a scatter-plot and shows an almost perfect linear approximation of the relationship between judged likelihood and judged similarity. That is, if the description of Tom happened to be similar to the participants' stereotype of a student in a certain specialization, they therefore decide the likelihood to be high that Tom also belongs in that category. These finding also demonstrate these predictive decisions are very different from the judged base rates ($r = - 0.65$) and this is very well aligned with the theory of the representative heuristic.

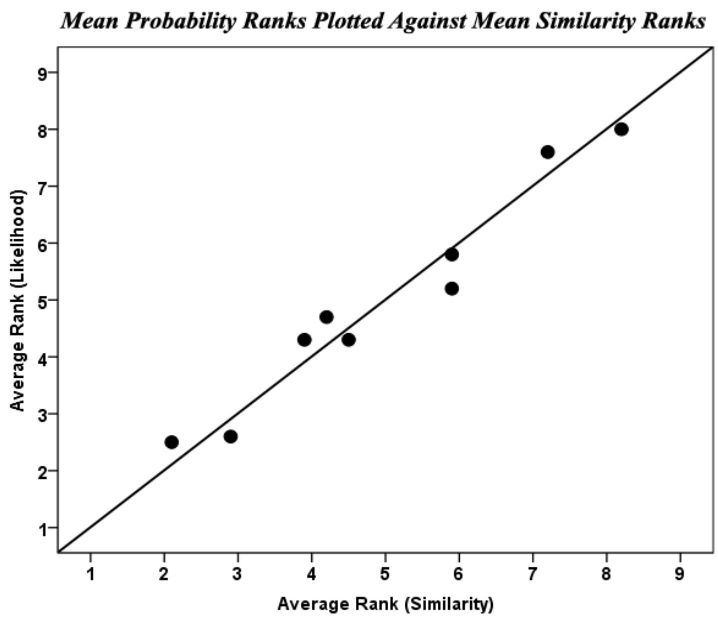



***Figure 2.3*** Linear approximation of mean likelihood ranks plotted against mean ranks of similarity or "Representativeness" in the Tom W. study. Adopted and modified from "A Perspective on Judgment and Choice – Mapping Bounded Rationality" by Kahneman, 2003, The American Psychologist, 58, 9, p. 709.

However, Kahneman (2003) also highlights that this is an example of a decision bias called base-rate neglect that violates statistical logic. A description based on unreliable information (such as a five-year-old personality test) must be given little credibility and decisions made in absence of reliable evidence must revert to base rates. In other words, there is no valid evidence that suggests that Tom is likelier to become a computer scientist than a humanities teacher. In fact, the base rates suggest that a humanities teacher is a much likelier alternative. This rationale suggests that decisions based on likelihood should be highly correlated with the corresponding base rates while this example shows that when people make decisions based on the representative heuristic, the opposite is true. Again, this decision bias is known as base rate neglect and is important to the scope of this paper. It is worth mentioning that a wide array of decision biases has been demonstrated in connection to the representative heuristic. Some of these include; misconceptions about correlations, insensitivity to the prior likelihood of outcomes, illusion of validity and insensitivity to sample size etc. (Tversky & Kahneman, 1973, 1974).

Furthermore, Tversky and Kahneman (1974) suggest that the confidence people have in their prescription and predictions depends primarily on the degree of representativeness, i.e. the degree of similarity between the potential outcome and the input information, with very little regard to the factors that could limit the accuracy of the decision. That is, people express greater confidence in their decision to categorize Tom as a future computer scientist when the description of Tom matches their stereotype of a computer scientist. The reason why the these findings are so thoroughly outlined is because they are central to the research questions, design and results of this particular study.

Our contention is that the representative heuristic and previously mentioned decision biases that arise from it should be explored in the context of real-world business practitioners that make decisions based on dashboard-based data visualizations. Some of the bottlenecks of human decision-making have now been articulated.

## 2.3 Theoretical Summary

To summarize the literature on the topics relevant to this research, it is important to note that dashboard-based data visualizations are heavily relied upon today in many organizations and are employed in a wide array of business functions such as marketing, sales, security and production (Krush et al, 2013; Van Der Heijden, 2013; McKenna et al, 2016; Skorka, 2017). The decisions that visualization tools aim to assist have varying degrees of complexity, ranging from relatively few information cues and simple interpretation, to situations where several decision aids are needed to tackle the complexity of a decision task (Speier, 2006). Within organizations there is a rather long tradition dating back to the 1970's to augment or aid the decision-making of business practitioners with computerized systems (Hosack et al., 2012).

However, human decision-making is mediated by several judgmental heuristics, some of which can cause serious biases when we make certain decisions (Tversky & Kahneman, 1974, Kahneman, 2003). As previously mentioned, the participants in the Kahneman and Tversky (1973) study were asked to rank nine graduate specializations based on the likelihood that Tom was now studying them. On a more abstract level, this kind of question fits the general description of: What is the likelihood that object A belongs to class B? The general finding in situations such as these is that people tend to judge likelihood not on base rates, but by the similarity between object A and class B. The following example aims to highlight how this theoretical background ties together the concepts of dashboards, data visualizations, the representative heuristic and the potential decision biases that can arise in this context. To illustrate: Let's say you decided to monitor your personal spending by using some very simple dashboard-based data visualizations.

In this dashboard, you had a bar chart representing your spending categories from the smallest to largest: Ranging from Gifts, Cloths, Recreation, Transportation, Food and the largest one being Housing. Next to this data visualization, you had a line chart representing your total spending over time, let's say a year, see Figure 2.4.1. below.

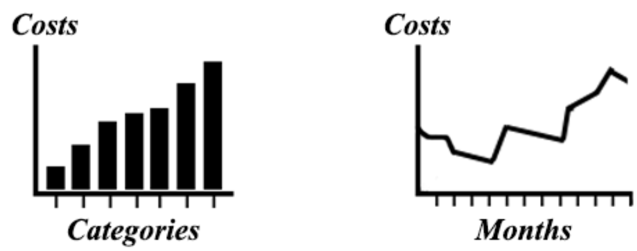



***Figure 2.4*** Stylized example of personal spending categories and spending over time.

If you were now asked to decide: Which spending category do you think has contributed the most to the yearly change in costs? What would you pick, a large or a small category? In this case, the statistically sensible answer would be the largest category: housing. This answer is in line with the base rates. Kahneman (2003) highlights that the statistical logic here is that in the absence of any other reliable piece of evidence, decisions such as these must revert to base rates. In this example, the costs of each spending category would be the base rates. If you were asked to rank the cost categories in order of the likelihood that they produced the yearly cost change, the statistical logic is the same: The evidence in the above dashboard suggests that the ranking should go from the largest to the smallest. Failing to do this is a decision bias known as base rate neglect.

As the earlier section outlined, prior research has shown that people consistently tend to neglect base rates when they use the representative heuristic to make decisions. Instead, similarities between objects and classes seems to be the prevailing feature that determines such rankings rather than the base rates (Kahneman & Tversky, 1973).

This paper will now explore whether similarity features of the data visualizations inside a dashboard, such as declining or growing data categories contribute to comparable decisions biases. As mentioned previously, when data visualizations are used as means to augment decision-making, it is important to be able to pull apart the potential effectiveness of the data visualization and the actual ability of users to understand them and use them as a decision aid (Boy et al., 2014). Therefore, this study will also explore the relationship between peoples' ability to interpret dashboard-based data visualizations and their tendency to make biased decisions, such as neglecting base rates. How this cross-disciplinary area was explored is described in the next chapter.

# 3. Methodology

This chapter will provide a detailed account of how this study was carried out. First, a brief outline for the choice of research design is made, an exploratory field experiment. An explanation of the development of the research instruments; the dashboards along with its accompanying data, contexts attributes and units as well as the data collection instrument then follows. The measurements section then illuminates the variables: dashboards, decision bias and interpretation accuracy. The field sampling and data collection procedure are then explained and finally the research quality and the ethical considerations are delineated.

A field experiment was conducted to investigate the research questions expressed in section 1.3 above. The overall methodological challenge was to balance two key concepts: internal and external validity. Based on Shaughnessy, Zechmeister, & Zechmeister's (2012) outline of these two concepts, the challenge was to balance the degree to which differences in performance on our dependent variable could be attributed clearly and unmistakably to the effect of our independent variable while ensuring that the findings also applied to other practitioners, businesses and conditions beyond the scope of this specific context. Field experiments are a typical way for researchers to increase the external validity of their research in real-world settings while maintaining a high degree of internal validity (Bhattacherjee, 2012; Shaughnessy et al., 2012). Since the purpose of this research is to investigate the relationship between dashboard-based data visualizations and decision bias among non-expert business practitioners, a field experiment was chosen as the optimal methodology.

## 3.1 Research Design

The design of the field experiment had two independent groups and two natural groups. Here, the independent variable was the 2-level between-subjects' groups where participants were randomly assigned in a double-blind procedure. Each of these groups were asked the same questions in relation to one of two dashboards. Each dashboard had four data visualizations
(bar chart, line chart, pie chart and bubble chart). Choices regarding these data visualization formats, underlying data categories and the questions associated with them will be covered in the section on Development of Instruments. The first group (*declining*) received a dashboard with the four different visualizations where the line chart showed an overall decline in costs and all the other data visualizations were ordered in similar way that mirrored this decline. In the second group (*growing*), the data visualizations were again ordered to mirror the line diagram but now in a way that highlighted an overall increase in all the data categories, instead creating a growing impression. The data visualizations of these two dashboards had similar graphical features, see Figure 3.1. on the next page to compare the two dashboards.

The continuous measure of Interpretation accuracy was mainly used to create two natural groups. The dependent variable that was measured in each of the two independent groups and the two natural groups was decision bias. The theoretical concepts behind these individual variables and how the measures are related to one another will be further described in the Measurements section. The rationale behind this rather crude design is because of the exploratory nature of this research. Bhattacherjee (2012) highlights how exploratory research often aims to scope out the magnitude or extent of a particular phenomenon while also generating some initial ideas about that phenomenon. This design allowed us to gauge the relationship between three constructs: Dashboard-based Data Visualization, Decision Bias and Interpretation Accuracy in a descriptive fashion. The reason why Interpretation Accuracy was an important aspect to capture and measure was because earlier research has demonstrated that individual differences in visual/cognitive ability, data literacy, or even other intellectual aptitudes such as design knowledge can mediate both decision-making performance and biases (Spier, 2006; Raschke & Steinbart, 2008; Morewedge & Kahneman, 2010; Lee et al., 2017b).

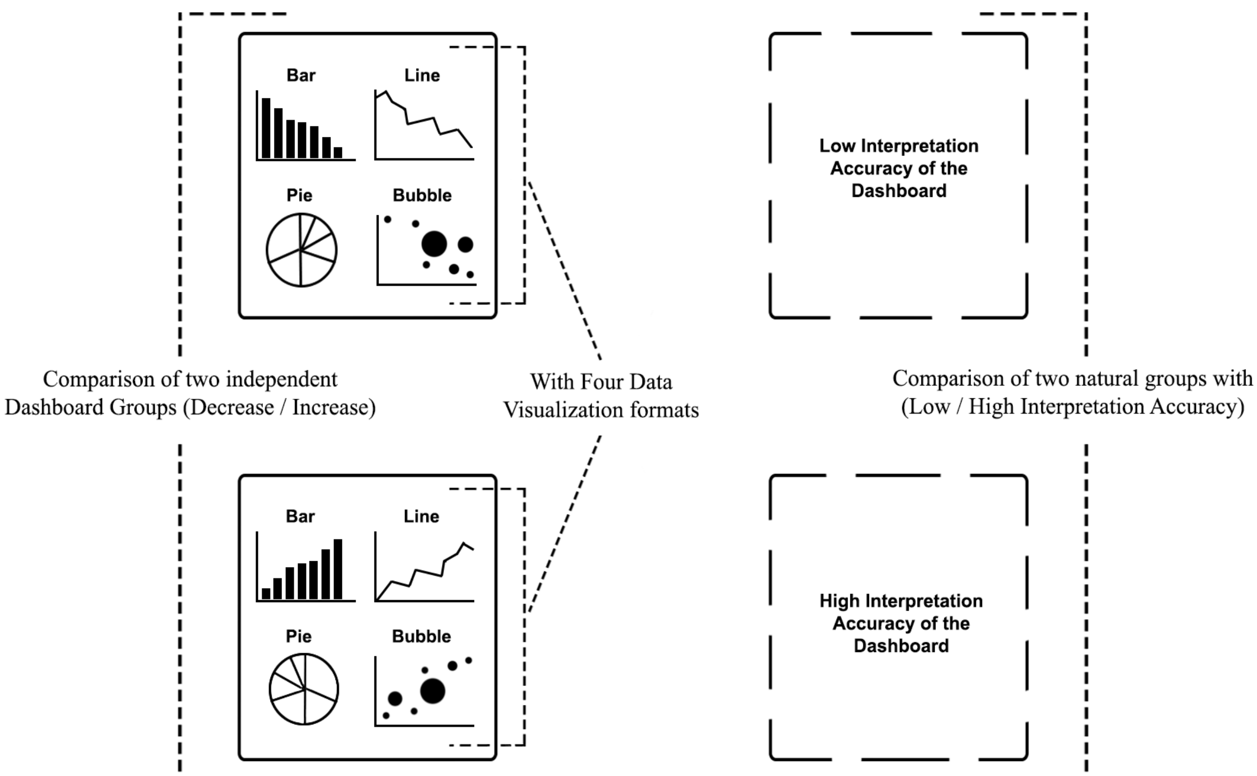



***Figure 3.1*** Research design with two randomized groups with the dashboards shown in stylized formats along with the two natural groups with Low and High degree of Interpretation Accuracy

As mentioned earlier, the dependent measure in the research design was the degree of decision bias. It should also be noted that each of the dashboards in Figure 3.1. on the former page were double-blind randomized groups and the data visualization formats were counter-balanced with a random starting order procedure. The latter procedure was conducted to address the balancing of practice effects and other potential confounds. This was solved by assigning a number to each of the data visualizations and then generating a random order from Random.org's (2018) "Random Integer Generator". This procedure created four different starting orders for the data visualizations in each of the dashboards. The reason for random starting order procedures are to mediate potential practice effects across the whole sample in repeated measure designs (Shaughnessy et al., 2012). Even though the four data visualizations were not really a repeated measure (since all of them were shown at the same time in one dashboard) the questions asked in relation to the dashboard were. Therefore, the order of the of the questions that were related to each of the four data visualizations followed this random starting order.

## 3.2 Development of Instruments

This section concerns the development of the data collection instrument. Initially, a digital dashboard was intended i.e. using local workstations and laptops at the offices as the delivery system for the field experiment. This would also have allowed for "time" as a control measure i.e. having the exact time it took for each participant to answer each specific question. However, due to severe rendering issues on these computers that often resulted in the dashboard being far too small, a paper-based questionnaire was used as an instrument instead. This and the measure of "time" will be further discussed in a later section of the paper. Concerning the development of the data collection instrument, the section below has three parts: First, how the context of the dashboard was modified for this study, second, the practical development of the dashboard-based data visualizations and third the development of the questions associated to these dashboards.

### *3.2.1 Development of the Dashboard-based Data Visualizations*

While this section is fairly focused on data visualizations, since they are often what comprise a dashboard, it is important to highlight that the independent variables of this study are defined at the level of a dashboard i.e. consisting of several data visualizations as shown in Figure 3.1. in the earlier section. Since there is no clear agreement over how exactly a dashboard should look and what it should do, a common conception of a reliable and comparable dashboard is scarcely available (Yigitbasioglu & Velcu, 2012). The same is generally true for data visualizations such as graphs, charts and diagrams (Mayer, 2000; Speier, 2006). Although design guidelines have certainly been developed over the years for both dashboards and data visualizations (e.g. Tufte, 1983; Arunachalam et al., 2002; Maheshwari & Janssen, 2014), the endemic usage of dashboards and the accompanying complexities of data visualization would generally create a very new and untested instrument. Instead, the decision was made to take advantage of previous research in order to yield more reliable and comparable results, which is a warranted concern in the context of dashboard-based data visualizations (Yigitbasioglu & Velcu, 2012).

Given the increased usage of these decision aids, several thoroughly validated data visualization formats have been developed (Lee et al., 2017b) . Through a Google Research Award, the latter scholars systematically developed a visualization literacy assessment test known as the VLAT, tailored for non-expert users. In Figure 3.2. the 12 data visualizations that make up this test are shown.

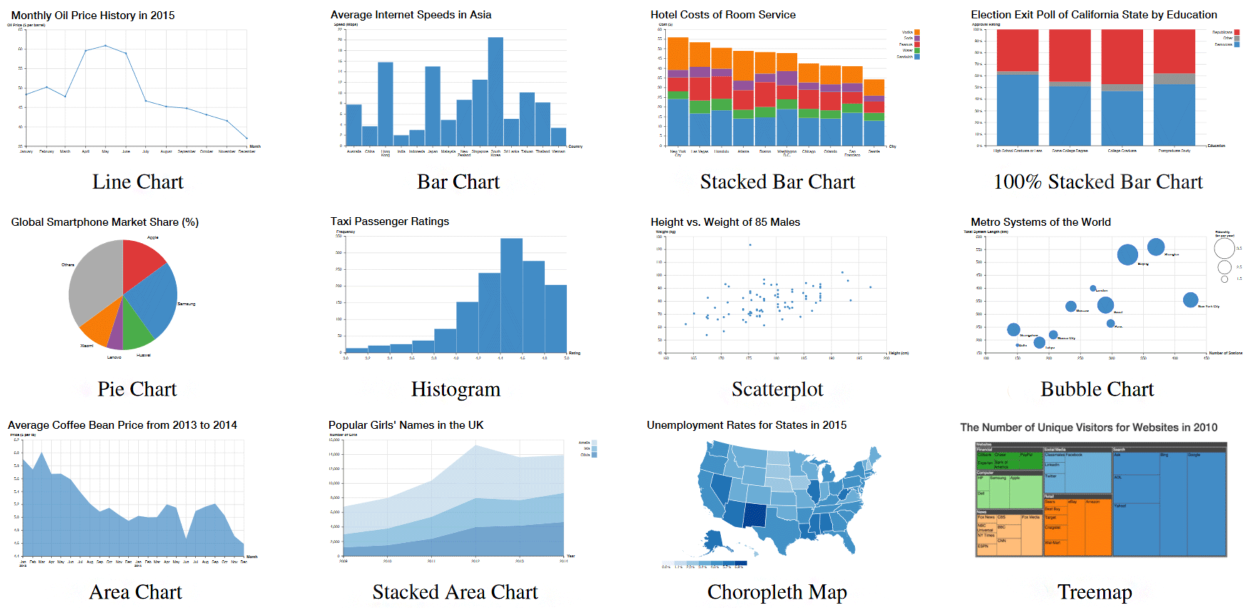



***Figure 3.2*** The twelve data visualization formats of the data visualization literacy assessment test. Adopted from "VLAT: Development of a Visualization Literacy Assessment Test" by Lee, Kim & Kwon, 2017, IEEE Transactions On Visualization and Computer Graphics, 23, 1, p. 555.

The earlier research that had generated the VLAT extended two major contributions to this study. First, a battery of previously tested and validated visualizations formats and second, a list of previously tested and content validated questions associated to each specific data visualization, see section 3.3.2. From these twelve formats, four were chosen for this particular context. The reason why these four data visualization formats of the VLAT were adopted for this research is two-fold. First, these four formats (bar chart, line chart, pie chart and bubble chart) were among the most frequently occurring formats from three sources, either school curriculums (up to twelfth grade), data visualization authoring tools or news outlets (Lee et al., 2017b). Given their frequent usage, these formats are both relevant and more likely to occur as decision aids in many contexts. Second, the software vendor market and IS research concerning both data visualizations and dashboards tend to reflect the lack of consensus regarding both their usage and appearance (Mayer, 2000; Speier, 2006; Yigitbasioglu & Velcu, 2012). This is not a benefit when trying to generate reliable and generalizable results.

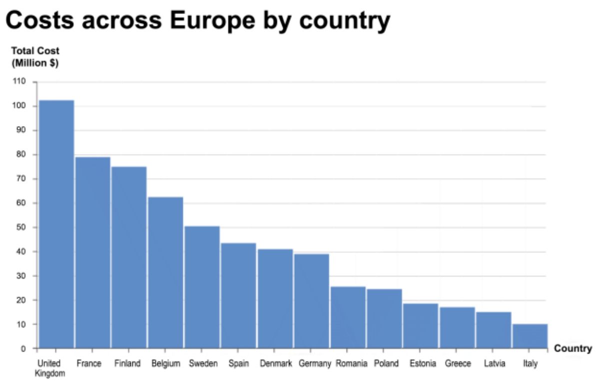
To step out of this ditch and not invent more completely new visualization formats, the decision was made to rely on these four visualization formats of the VLAT. Reducing the number of visualizations and parameters also delineated and reduced the complexities of the instrument. However, some modifications still had to be made to tailor these data visualizations for the context of this study. The purpose was to explore the relationship between decision bias and similarity features in dashboard-based data visualizations, such as declining or increasing data categories and whether the interpretation accuracy of dashboard-based data visualizations was associated to decision bias. Since a dashboard is a visual display of the most important information needed to achieve or assess some specific objective, these four data visualizations needed a coherent context, attributes and units.

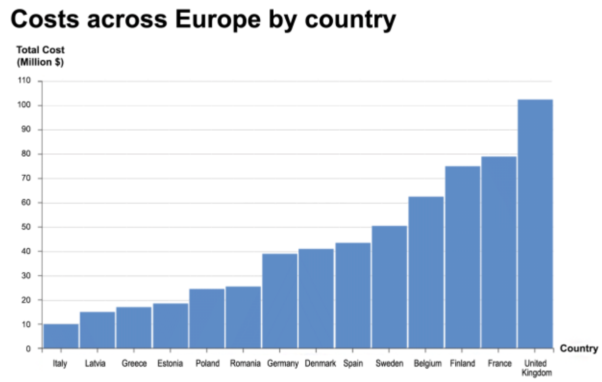
### 3.2.2 Data, Context, Attributes and Units

The authors of the test, Lee et al. (2017b) outline that every data visualization included in the VLAT was based on 12 different datasets published in news outlets. The VLAT itself was created to assess a non-expert user's literacy regarding data visualizations. Because of this purpose, the authors wanted to avoid any familiarity of the underlying data's context. Therefore, twelve very different contexts were used in the VLAT, ranging from coffee bean price to taxi passenger ratings, see Figure 3.2. A dashboard featuring coffee bean prices and taxi passenger ratings is rather odd no matter the business context. However, Lee et al., (2017b) argue that the context of the visualized data is very important when dealing with the users' interpretation or literacy because users might be familiar with the common context of a "car" but not with the specific attributes of "acceleration" or the units of "0 -100 km/h".

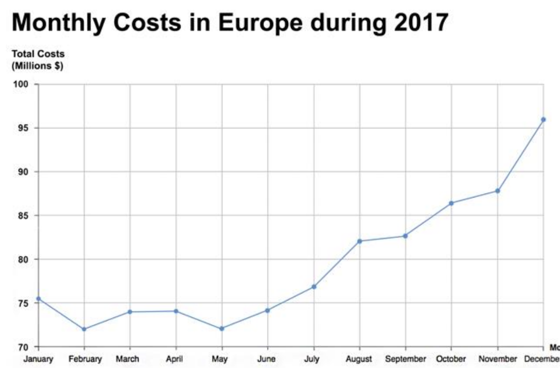
The above argument was also considered when the data visualizations were modified for this study. The following changes were made: The underlying data had to be synthesized to attenuate certain features, such as declining or increasing data categories to create similarities between the data visualizations of the dashboard. Now, dashboard solutions also often come with many features, such as interactive drill down capabilities (Yigitbasioglu & Velcu, 2012). Therefore, a context with several levels of depth was needed. The chosen context was the geographical region of Europe because we knew that the vast majority of the business practitioners in the sample were going to be Swedish. The school system in Sweden has had a tradition of teaching its population about the geography of Europe from sixth grade and onwards (Swedish National Agency for Education, 2018). The assumption here was that people in the sample would understand attributes such as western Europe, United Kingdom, London etc., Finally, the units in each data visualization also needed consistency. The unit that was used in all the data visualizations were costs in US $. There were two reasons for this choice. First, Tufte's (1983) guidelines for displaying quantitative data are a commonly cited source in the context of data visualizations according Amer and Ravindran (2010). When investigating biases caused by visual illusions in graphical presentations, Amer (2005) highlights that one of these guidelines underscores that monetary measurements (such as $) are much preferred compared to any other unit for time-series graphs i.e. line charts.

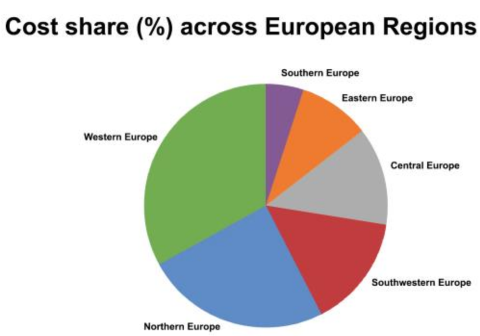
Second, earlier research on decision biases suggests that articulating or “framing” decisions as losses or gains have effects on people’s judgements (Tversky & Kahneman, 1981). It was therefore important to keep the contexts and units consistent between the four data visualizations in the dashboard. This work produced two iterations of each data visualization format. Figure 3.3. below shows a comparison of these final iterations.

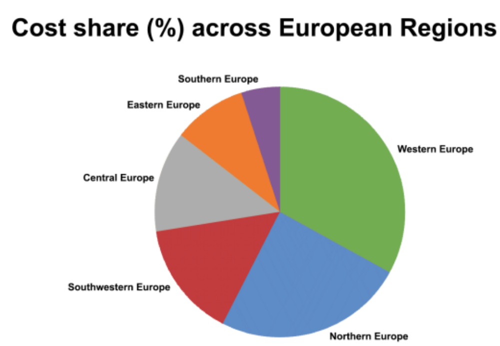






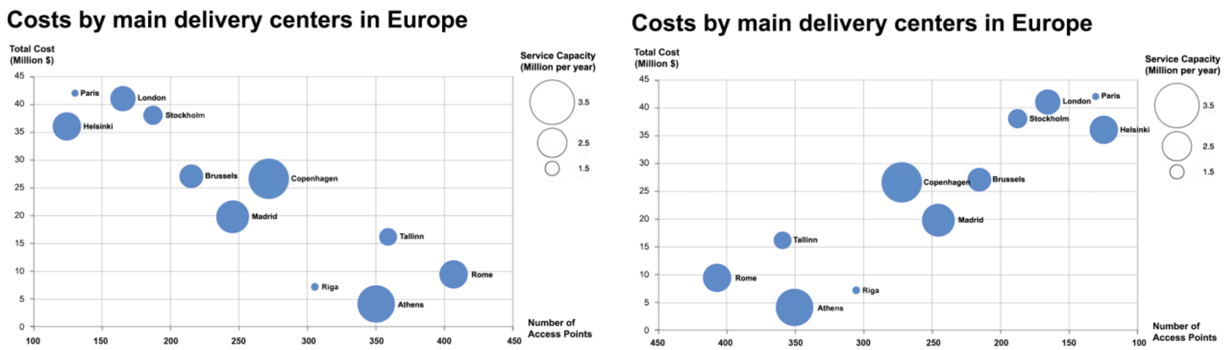



***Figure 3.3*** The two versions of the data visualizations that were used for the dashboards. The formats of each data visualization are based on VLAT: Development of a Visualization Literacy Assessment Test" by Lee, Kim & Kwon, 2017, IEEE Transactions On Visualization and Computer Graphics, 23, 1, p. 555

### *3.2.3 Development of the Questions related to the Dashboards*

As mentioned in Section 3.2.1, a list of previously tested and content validated questions associated to each specific data visualization from the VLAT was already available. Lee et al. (2017b) outline that the VLAT consists of 53 multiple-choice test items that cover twelve data visualization formats. The authors refer to these tasks as measures of the reading and interpretation of visually represented data. These data visualization tasks had several themes such as Retrieve Value, Find Extremum, Determine Range, Characterize Distribution, Find Anomalies, Find Clusters, Find Correlations/Trends, and Make Comparisons. A panel of visualization domain experts and visualization researchers validated the appropriateness of these data visualization questions through the calculation of a content validity ratio.

However, the multiple-choice format is based on a timed test procedure (Lee et al., 2017a) The ecological validity of receiving a dashboard full of 25-second multiple choice items is rather low. Instead, the decision was made to let the business practitioners take the time they needed for each task. Due to the limited resources of this study, every single item of the VLAT could not be adopted as a measure. After the pilot-testing of the data collection instrument was finished a total of eight questions related to the four data visualizations remained. Together, these eight questions touched on every dimension of the four data visualizations that could be interpreted in the dashboard.

The questions that constituted the measure of Interpretation Accuracy were related to four visualization formats (*bar chart, bubble chart, line chart and pie chart*). The operationalization of the measures will be described in the next section. These questions were limited to two themes of tasks: Retrieving a value and determining a range of values. The rationale here is simple. These tasks enabled users to read and interpret the visually represented data which was a key tenet of this exploratory field experiment. All of these questions were intellective so they all had correct answers that could be found in each data visualization (Lee et al., 2017a). One question asked the participants to retrieve a value, such as “What is the combined cost share (%) of Eastern and Central Europe?” while another was focused on determining a range such as ‘What is the range of total costs between the European countries?”. Again, the criteria for these eight questions was that the participants had to make a judgement on every dimension of the four data visualizations that could be interpreted in the dashboard. This is why there is only one question related the Pie chart where the size of a slice is the only dimension for the data values while a bar chart has two dimensions: The size of each bar and the individual category that each bar represents.

Descriptive statistics related to each of the questions can be found in Appendix 1. The final instrument containing all of the questions can be found in Appendix 5 and Appendix 6 for the respective dashboard groups. Although the interpretation of different data visualization formats lies outside the scope of this explorative study, a mixed between-within ANOVA was computed for the two Dashboard groups to ensure that the visualization formats were interpreted in a similar fashion between these groups, see Appendix 2.

The final question related to the dashboard-based data visualizations concerned the measure of Decision Bias. As already articulated Section 2.2, the participants in the Kahneman and Tversky (1973) study were asked to rank nine graduate specialization based on the likelihood that a student (Tom W.) was now studying them. On a more abstract level, this kind of question fits the general description of: What is the likelihood that object A belongs to class B? And so, do several business-related questions such as: Which marketing campaign contributed most to the increase in sales? The general finding in situations such as these is that people tend to judge likelihood not on base rates that should be used to determine the likelihood but by the similarity between object A and class B (Kahneman & Tversky, 1974). The reason the second chapter (Theoretical Background) outlined one of their studies in a fairly detailed fashion was because both the design and analysis served as a benchmark for the operationalization of Decision Bias in this exploratory study.

The final question was built on Kahneman and Tversky’s (1973, p. 239) question to one their experimental groups but now based on the attributes of the data visualizations in the dashboard:

> “Please rank the eleven main delivery centers in order of the likelihood that they produced the yearly cost change from January to December of 2017
> (from ***11 = Highest likelihood*** to ***1 = Lowest likelihood***)”

This might seem like a contrived question, but it allowed for a measure of how sensitive people were to the actual data-points in the dashboard as opposed to graphical features of similarity. As Pappas and Whitman (2011) argue, allowing for these kinds of fact-based decisions is essential in the context of dashboards. By providing these kinds of rankings, the participants were making decisions that were well aligned with the second definition provided in the Theoretical chapter: “a strategy for choice of action”.

## 3.3 Measurements

We will now illuminate the path between the key concepts of this research such as dashboard-based visualizations, interpretation accuracy and decision bias and their respective constructs i.e. operationalization. The general problem with concepts is that many of the underlying phenomena of interest are often fuzzy and imprecise (Recker, 2012). This problem is what the operationalization of each concept will address. This study had the following measures, with the first being the independent variable *Dashboard* with the two groups Declining/Growing, the second being a continuous variable (*Interpretation accuracy*) that was later broken down into two natural groups of Low/High and the third (*Decision bias*) as the dependent variable based on the judged mean ranks of the respective groups. Background and demographic variables were put at the end of the questionnaire per Shaughnessy's et al. (2012) recommendations.

### *3.3.1 Two Dashboards (Declining and Growing)*

Data visualizations and their aggregate format as dashboards have many properties that are important to highlight. This research was concerned with specific features in the context of dashboard-based data visualizations. On the next page, Figure 3.4 shows a full view of the two levels of the independent variable that constituted the two conditions of the field experiment. What is noteworthy about these two dashboards is that they have clear similarities and differences. The data visualization formats are obviously the same. The difference between them is that all the data categories have been inversed so that the dashboards either emphasize a declining order or a growing order among these categories. This is true for three of the four formats since a pie chart cannot really have a declining or growing trend. In other words, all of the formats have been inverted between the two conditions. For the pie, bar and bubble chart, this means that all of the data categories still have the same values, but the categories are shown in the opposite order. The line diagram still has the same range of values, but the values are different between the two conditions. The connection to earlier research works in the following manner: Kahneman and Tversky (1972) has studied how representative or similar ordering of categories show that people tend to consistently decide the more similar categories as more likely to produce an event (that also happen to be similar or representative of the order of those categories), even though it is not according to some objective measure of probability or base rates. What is explored with these two dashboards is weather business practitioners tend to rank these data categories as more likely to have produced another event, just because of graphical similarities between the data visualizations.

The question here is if the participants tended to rank the bubbles (delivery centers) in the bubble chart by the likelihood that they produced the yearly cost change in the line chart. So, weather this ranking was performed from the bottom to the top or in the opposite order for any of these two dashboards is connected to the measure of Decision Bias in section 3.3.3., but first another measure needs explanation.

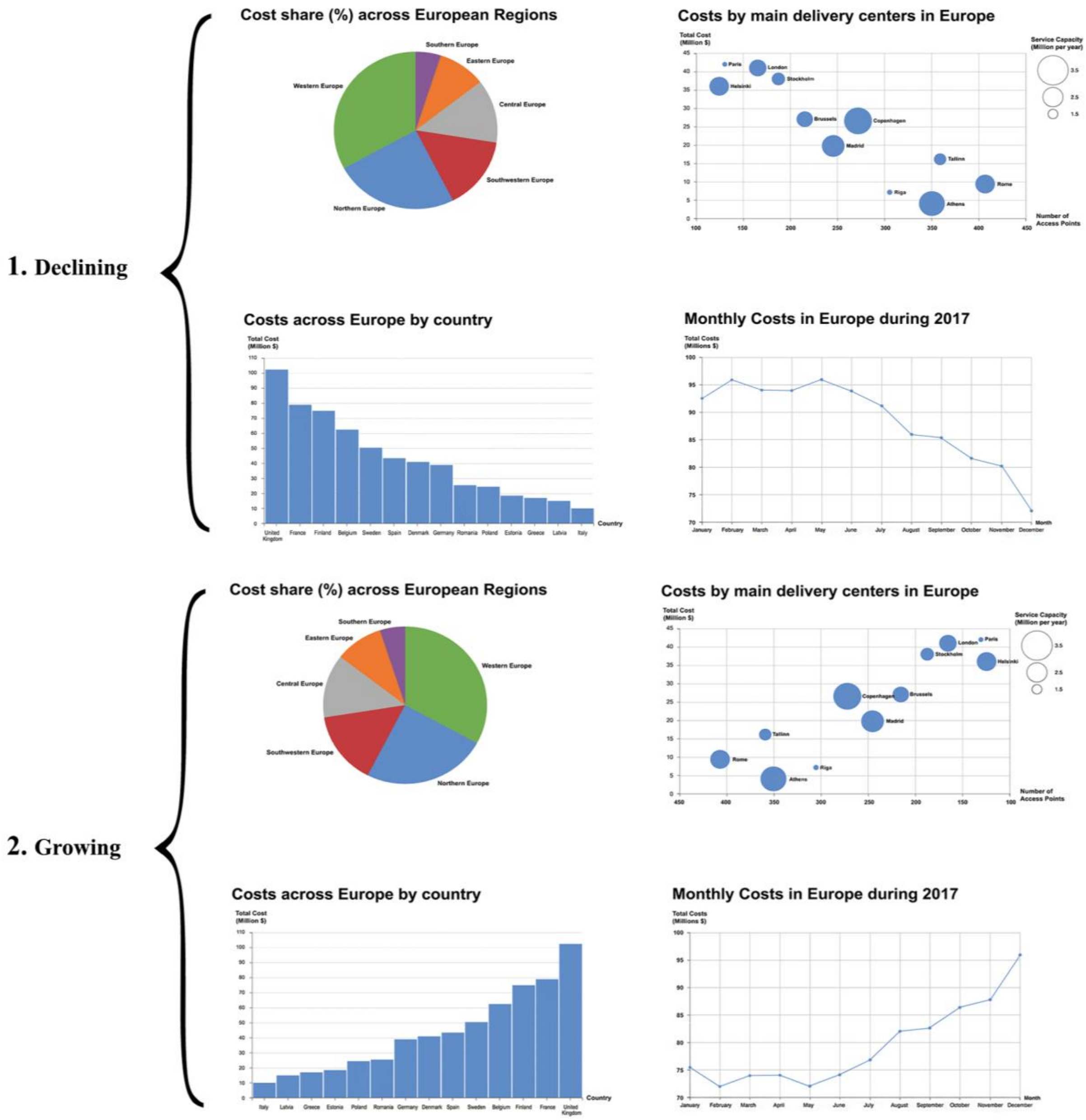



***Figure 3.4.*** The two independent dashboard-based data visualizations that were used in the study

### *3.3.2 Interpretation Accuracy (Low and High)*

The operationalization of Interpretation Accuracy follows a similar path to what Spier (2006) advocates for the measure "decision accuracy". For linguistic purposes, we chose to call this measure "interpretation" rather than "decision" accuracy since the answers provided by each participant was an interpretation of a dashboard-based data visualization rather than a decision.

It is important to note that if the following question is asked in relation to the bar chart: What is the total cost in Finland? This question had an optimal answer that could be found in the bar chart. Spier's (2006 p. 1122) operationalization relies on the following formula:

1- ((*optimal answer - participant's answer*) / *optimal answer) = Decision Accuracy*

A concrete example could be: 1- ((*50 - 40)* / 50*)* = 0*.80.* As the above formula shows, this gives a quotient of the optimally achieved score, i.e. a degree of how close the participants were to the optimal answer. However, Spier (2006 p. 1123) then proceeds to average these scores across the decision tasks to form an overall percent of optimal achieved. Theoretically, one could then underestimate the first decision task by 50% and then, overestimate the next by 50%. At face value, according to Spier's (2006) conception the average "decision accuracy" of these two decision tasks would then be 100%. Our rationale was that this would not really capture any construct of "accuracy", so the following modifications were made to increase the construct validity:

((*optimal answer - participant's answer*) / *optimal answer) * 100 = interpretation error %*

To illustrate: ((50 - *40)* / 50*) * 100 = 20 %*. These errors were then averaged, and this average was then subtracted from 100, i.e. the optimal score if all answers were correct. This measure is what constitutes *interpretation accuracy* and gives a much more precise picture of the actual accuracy of the answers. An error ceiling was also used with this measure for the following reason: If a participant provided an answer of 500 in the previously mentioned example: ((50 - *500)* / 50*) * 100 = - 900 %*. This would seriously skew the average Interpretation Accuracy of that participant. However, it is reasonable to argue that if the optimal interpretation yields 100 %, then for the sake of balance, the highest error should also be 100%. This can also be seen as a method of handling outliers in the data. Now, if a participant would underestimate the first decision task by 50% and then overestimate the next by 50%, the average "interpretation accuracy" of these two decision tasks would be 50%. This operationalization therefore captures how close the participant's answers were (on average) to the optimal answers, i.e. how close the participants were to the optimally achieved scores as Spier (2006) puts it. Although this measure had benefits compared to Spier's (2006) conception, there is one trade-off. This measure of interpretation accuracy captures how close one is (on average) to the optimal interpretation, while sacrificing any information about over-/underestimations in the tasks. Spier (2006) also used quite similar but not identical data visualizations or decision tasks so pilot-tests were still performed, see section .3.6. for more information.

Finally, this measure was collapsed into two groupings, using the median as the cut-point to form the two natural groups of Low and High Interpretation Accuracy. As mentioned earlier, for the sake of construct validity, the questions that were used in this study were all based on the previously validated data literacy assessment test or (VLAT).

### 3.3.3 Decision Bias

The operationalization of Decision Bias is modelled on the method and analysis employed by Kahneman and Tversky (1973). This gained both a legitimate methodology while also making the results of this study more comparable to studies of decision-making in other contexts with other populations. Other studies, such as Kahneman and Tversky (1973) and Spier (2006) sampled students as opposed to business practitioners. As outlined in section 3.2.2., in the final question related to the dashboard, the participants were asked to rank eleven delivery centers in the Bubble chart based on the likelihood that they produced the cost change showed in the Line chart. The rationale behind the attributes (*delivery centers*, *service capacity* and *access points*) that are shown in the bubble chart is the same as Kahneman and Tversky (1973) outline for the description of the student Tom. W. that was mentioned in the Theoretical chapter. The description is meant to leave only unreliable evidence about was Tom is likely to study. The same is true for these dashboards regarding all the attributes. They only reliable evidence for ranking these delivery centers are their costs or base rates. The question is whether the participants will make decisions based on representativeness i.e. base their rankings of these attributes on other unrelated but similar attributes from the other data visualizations and thereby neglect the relevant information or base rates.

Emulating Kahneman and Tversky's (1973 p. 238) approach, the average ranks were then computed for each delivery center. The Pearson product moment correlations were computed between these average ranks and the correct base rates. Correct base rates simply mean that based on the dashboard-based data visualizations it would have been in line with the base rates to assign the delivery center that particular rank, see section 2.4. for a more detailed explanation. The association between the mean ranks and the correct base rates could now be gauged in three quantitative ways. First, the correlation coefficient is a measure of association by itself. Second, by squaring these correlations and thereby getting the proportion of variance accounted for or $r^2$, allowing for the associations to be compared with a percentage % measure (Aron et al., 2014, p. 512). The third way and final demarcation for a potential decision bias to be detected within any of the groups was the linear approximation of the relationship between the mean ranks and the correct base rates. If these approximations were negative, this would mean that people (on average) tended to rank the delivery centers in the opposite order of their correct base-rates hence constituting clear evidence of a decision bias within that group.

## 3.4 Field Sampling Procedure

In line with Bhattacherjee´s (2012) recommendations, a pilot study *(n = 27)* was conducted through a convenience sample before the final instrument was used on the intended participants. After the results had been collected and examined from the pilot study ,a cluster sample was used to select participants from the target population. Scheaffer et al. (1971, p. 266) defines a cluster sample as a probability sample in which each sampling unit is a collection or cluster of units. A cluster sampling procedure was considered as an effective selection procedure since a viable sample frame listing each unit of analysis, i.e. a list of every employee at the business was hard to obtain while a sample frame listing clusters in this case departments was easier to create. This department list was extracted from the organization's intranet and saved as a .csv file. The departments in this list were then randomly sampled through R Studio's sample function. Each department manager was then contacted to schedule a date for the field experiment. Out of the nineteen department managers that were contacted, fourteen departments agreed to participate in the field experiment. Out of the five remaining department managers: Two did not respond to the request while two departments were not located on the main campus and were therefore excluded for logistical reasons and one of the departments was excluded because most of its team members could only participate remotely. The context that was used to sample each department was either their daily morning meeting or a weekly meeting where the whole team was gathered. These contexts were consciously chosen because they made the external conditions of each data collection procedure very similar to one another. The sample was collected over a period of 7 days yielding a total sample of *87* business practitioners.

## 3.5 Sampled Participants

The sample (*n = 87*) consisted of business practitioners that were employees at a large information technology company in Sweden. For ethical reasons the organization where these business practitioners worked will not be named in this paper, though some background facts will be mentioned. The organization had thousands of employees based in more than 50 countries and an annual turnover of more than a billion US $. The supply chain of the organization allows their 70,000+ partners to distribute their products and services in more than 170 countries.

The selection of participants was meant to be representative of generic business practitioners i.e. non-expert users that work in a business where dashboard-based visualizations are used as a decision aid at varying degrees. This was why a probability sample based on departments was used instead of a convenience sample that would have been focused on the degree of dashboard usage within the organization. Since the unit of analysis in this research concerned generic business practitioners a convenience sample would have excluded participants based on criteria that would have been complex to validate before the sampling procedure, such as the degree of dashboard usage.

The actual field experiments took between 13 to 24 minutes to complete. The sample of participants consisted of 74% males and 26% females. This accurately reflected the gender distribution across the entire organization according to their own data. Three people did not respond to this question. Again, for reasons of confidentiality the organization will not be named in this paper. 95% of the sample reported that they had spent more than 2 years at the university with 2 people not responding to this question and 88% of the sampled participants reported Sweden as their nationality. The average work experience of all the business practitioners in the sample was 9.8 years.

## 3.6 Data Collection Procedure

The data was collected through a questionnaire from a sample of business practitioners in Sweden (see Section 3.5) even though a digital version was earlier intended and explored. A paper-based questionnaire was used as the instrument for this exploratory field experiment. The questionnaire included a page with instructions, a dashboard for the specific experimental group – the declining or growing dashboard and a page of questions (see Appendix 5 and 6).

For each of these groups, a total of 52 questionnaires were printed. These were then randomized through Random.org's (2018) "Random Integer Generator" by generating a sequence of the total number of questionnaires. Stacking the questionnaires according to this random sequence ensured that neither the participants nor the researchers were aware of which condition the participants would end up in once the questionnaires were handed out. To reach the participants, a company-wide organizational chart was used from which various managers leading different teams were randomly selected. It is important to mention that this was not a convenient sample as opposed to the one used during our pilot study. As mentioned in section 3.4. an email was sent out to each of these managers requesting for a time slot during their teams' morning meetings to carry out the data collection procedure, see Appendix 3. Once all responses for participation in the data collection were sourced, a pilot-test was conducted ($n = 27$) three days before the instrument was finalized and used for the main data collection procedure. The pilot-test exposed some flaws in the first outline such as unclarity regarding some of the questions. Especially the question concerning the rankings of the distribution centers was rephrased in preparation for the main data collection.

During the morning meetings, a time slot was scheduled to carry out the data collection. First, a brief introduction and instruction was given to the participants, see section 3.8.1. for more information about the ethics of this study. Each participant was then handed a questionnaire from the randomized stack together with a pen after the instructions were given. Each participant then answered the questions sitting down in either the departments conference room (if the meeting was held there) or at their individual desks (if the meeting was held at their departments office environment). Each data collection procedure was timed, see section 3.5. However, its significant to note that this was not an indication of how long each individual participant took to complete the field experiment, but rather of how long it took the slowest / last individual to answer all the questions in each of these departments.

Once the data collection procedure was completed a debriefing email was also sent out to all the sampled departments. This and other ethical aspect will be addressed in the last section. Before these ethical considerations the preparations for the analysis of the collected data will be addressed.

## 3.7 Statistics

The next logical step after the data collection was the analysis of the collected data. However, before the data could be statistically analyzed in any software, it had to be prepared in a digital form. We used Bhattacherjee (2012) outlined steps of data preparation; data coding, data entry, missing values and data transformation. However, no data transformations were performed due to the descriptive rather than inferential nature of this research.

The data coding step involved the development of a variable codebook for the variables in the questionnaire. This allowed for a uniform description, understanding and meaning of each data point collected from the data collection instruments. Next, the data was entered into an .xls file using Excel and this constituted the data entry step. Based off the variable codebook, the paper-based responses were entered into the data file. Excel was chosen for this task because it allowed for simpler use of the formulas connected to the measures. As Bhattacherjee (2012) postulates, it is inevitable that there will be missing values in any empirical dataset. This study was no exception and some participants did not answer certain questions because of unknown reasons. The missing values were manually replaced by 9999 as per Bhattacherjee (2012) recommendations and then flagged as missing values in SPSS. Missing values can be handled by listwise deletion which would mean a loss of all responses that had any missing values. Pairwise handing was used consistently in this study in agreement with Pallant's (2013) instructions.

Once all the data had been captured and computed according the formulas in the Measurement section, it was imported to SPSS, which was the main data analysis tool. Before the data analysis was conducted the data it was first screened in accordance with Pallant's (2013) recommendations. The earlier mentioned "error ceiling" that was used for the measure of Interpretation Accuracy was a sufficient step to handle potential outliers in the dataset. Descriptive statistics were first computed for the whole sample, highlighting demographic characteristics like age, gender distribution and average work experience among the tested business practitioners. Next, Tversky and Kahneman's (1973) analysis was emulated to explore the degree of decision bias by computing the Pearson product moment correlations between the mean ranks of the two dashboard groups and the correct base rates shown in the dashboards. To add further depth and credence to the findings, interpretation accuracy was then divided with the median as the cut-point to explore the relationship between High and Low interpretation accuracy and the degree of decision bias in these two groups.

## 3.8 Research Ethics and Quality

This section concerns the bedrocks of ethical research as laid out by Bhattacherjee (2012) such as voluntary participation, informed consent, harmlessness, confidentiality, transparency of analysis and reporting. Due to the nature of this study the aspects of full disclosure and debriefing has its own sub-section. Finally, a brief summary of the concerning the quality of this research is provided.

### *3.8.1 Ethical Guidelines*

The subsequent ethical guidelines were thoroughly followed throughout this research process. In line with both Bhattacherjee's (2012) and Recker's (2012) recommendations, prior to each data collection procedure, all the participants were informed both verbally and in written form that:

- Their participation was completely voluntary
- The answers they provided would be confidential
- They were free to abort their participation at any time, without any explanation to anyone
- The answers they provided would be used for a MSc. thesis in Informatics

We took time to ask openly if there were any participants who did not want to be part of the sample before each data collection began. Each questionnaire also had a consent question appended to the first page, see Appendix 5 and 6. The participants were also informed that the intention of the analysis and reporting was not to compare either individuals or organizational departments against each other. All respondents were made aware that the answers they provided would be anonymous. The structure of the data collection procedure made it possible for us (as researchers) to identify whether somebody had participated but it would have been impossible for us to verify retrospectively which questionnaire a particular participant filled out. No personally identifiable information was captured during the data collection procedure.

Yet, one ethical wrinkle need further attention. Due to one of the phenomena under investigation in this study i.e. Decision Bias, participants could not be fully informed of the scope without also contaminating the validity of the results. Participants were only informed that the area of inquiry concerned dashboards and decision-making. Therefore, the participants could not participate under the principle of transparent full disclosure. To mitigate this, we provided the participants with a full debriefing in line with Shaughnessy's et al. (2012) recommendations. This ethical consideration is given more attention in the debriefing section.

### 3.8.2 Debriefing

To preserve internal validity of this research, only a few people were privy to full scope of the research area that was explored through the field experiments. These people were mainly respective team managers and project leads. Each manager was informed that this measure was taken to preserve the validity of the study and that everybody included in the sample would be informed that the scope of the study not only concerned decision-making but specific decision biases. A full debrief email was sent out once the data collection procedure was complete. By following the recommendations of Shaughnessy et al. (2012) this email detailed the full scope of the research process and the area of research while also inviting participants for further feedback or questions if they were interested, see Appendix 4.

### 3.8.3 Research Quality

Most of the aspects that concern the quality of this thesis are implicit in each step of the research process and should be judged by potential peer-reviewers. However, the intention of this section is to provide a brief summary of the strengths and weaknesses that make up the quality of this research such as its validity, reliability and generalizability. As mentioned earlier, in line with Yigitbasioglu and Velcu (2012), there is no clear consensus over how exactly a dashboard should look or what it should do, and this tailored usage of dashboards and their accompanying data visualizations make them very complex to operationalize as scientific instruments. For the sake of replication and comparison of results, many steps were taken in this study to transparently and systematically develop the dashboard-based data visualizations that were finally used, see section 3.2. The questions that were asked in relation to these dashboards were firmly rooted in prior research (e.g. Lee et al., 2017b) and they were included to ensure that the participants understood all task-specific aspects (such as categories, units, legends, sizes, ranges etc.) of the dashboards. This latter point also connects to the construct validity of the measurement Interpretation Accuracy. The operationalization of this measure was rooted in earlier research conducted by Speier (2006) while some sensible changes were made to this measurement, see section 3.3.2. The ranking question intended to measure the construct of Decision Bias was modelled on the methods and analyses employed by Kahneman and Tversky (1973 p. 238) whose body of work was later awarded the Nobel Prize in Economic Sciences by Sveriges Riksbank (Royal Swedish Academy of Sciences, 2002). This is an argument about legitimacy of methodology (as researchers, we do not expect any awards).

The field experiment was also administered according to a double-blind randomization of the two dashboard groups in that, neither the researchers or the participants knew which condition the participants were in) and the data visualizations and their accompanying questions were counter-balanced according to a random starting order procedure across the sample to mitigate potential practice effects as laid out by Shaughnessy et al. (2012). To further increase the ecological validity and relevance of these findings a probability sample of generic business practitioners was drawn from a large IT organization as opposed to a convenience sample.

While it is very important to secure as many aspects as possible regarding validity, reliability and generalizability it is important to highlight the following point. Due to the nature of this research i.e. a field experiment exploring the relationship between dashboard-based data visualizations, interpretation accuracy and decision bias in a real-world setting, Bhattecherjee's (2012) sentiment that exploratory research may not lead to the most rigorous and accurate understanding of the target problem but may be worthwhile in scoping out both the nature and extent of the problem. This was why a rather crude research design was used with only two independent groups and two natural groups.

Even though many steps were taken in this research design to keep a high degree of internal validity, the general principle of a field experiment is that the researchers have control over the assignment to the "treatment" i.e. a specific dashboard while forgoing some control over the treatment itself (Humphreys & Weinstein, 2009). Some individual characteristics of the participants and the precise manner and context in which the "treatment" was applied are more likely to take on values given by "nature" rather than being set at the discretion of the researchers. This is true in the context of this research since the all data was collected in the field i.e. inside an actual organization at the offices of business practitioners.

The vocabulary of such a cross-disciplinary domain was also hard to balance in this research. The used terminology in this research stretches between experimental, quasi-experimental and survey research while being conducted in a real-world setting. Based on the argument provided in the beginning of this chapter, the decision was made to verbally label this research as a field experiment according to Shaughnessy et al. (2012) and Bhattacherjee's (2012) definition of that method.

Recker (2012) and Bhattacherjee (2012) highlights that generalizability or external validity points to whether observed associations and operations of the study can be repeated or generalized in equal settings such as populations, other organizations, contexts and times. Granted that our study was carried out at a large IT company in Sweden, we cannot with certainty ensure the generalizability of this study across dashboard-based data visualizations in general (their tailored and task-specific usage and appearance also contributes to this fact). The empirical claims of this study will only preserve their truth in a rather narrow context where four popular data visualizations were used in a dashboard to explore one specific decision bias arising from the representative heuristic among business practitioners working a large-scale IT organization. The intended knowledge contribution was to scope out the nature and extent of a target problem i.e. the relationship between dashboard-based data visualizations, interpretation accuracy and decision biases. Yet, the results outlined in the next chapter could generate further reasons for researching this cross-disciplinary domain.

# 4. Data Analysis and Results

The first section of data analysis and results concerns the two independent Dashboard groups (Declining/Growing). The section highlights several background and demographic statistics that are relevant to the comparability of the two Dashboard groups and then their degree of Decision Bias. The second section concerns the Decision Bias of two natural groups i.e. (Low/High) Interpretation Accuracy. That section also starts by outlining the relevant background variables of those two groups then their degree of Decision Bias. To ensure that the Interpretation Accuracy questions were interpreted in a similar fashion between the two independent Dashboard groups a mixed between within ANOVA was conducted. Since this is outside the scope of the research questions, descriptive and inferential statistics regarding the interpretation Accuracy questions are featured in Appendix 1 and 2.

## 4.1 Dashboard groups (Declining/Growing) & Decision Bias

The Declining Dashboard group (n = 41) and the Growing Dashboard group ($n$ = 46). The Declining group now had an Interpretation Accuracy percentage of ($M$ = 94.06, $SD$ = 5.89) while the Growing group had an Interpretation Accuracy of ($M$ = 91.86, $SD$ = 10.33).

The age of the participants in the Declining Dashboard group was ($M$ = 34.7, $SD$ = 9.7) and ($M$ = 34.2, $SD$ = 9.6) for the other group. The reported amount of work experience (in years) among the participants in the Declining Dashboard group was ($M$ = 9.8, $SD$ = 9.2) and ($M$ = 9.7, $SD$ = 8.8) for the participants in the Growing Dashboard group.

Figure 4.1 on the next page shows the proportion of participants who reported to have attended university for at least two years and the gender distribution between the two Dashboard groups.

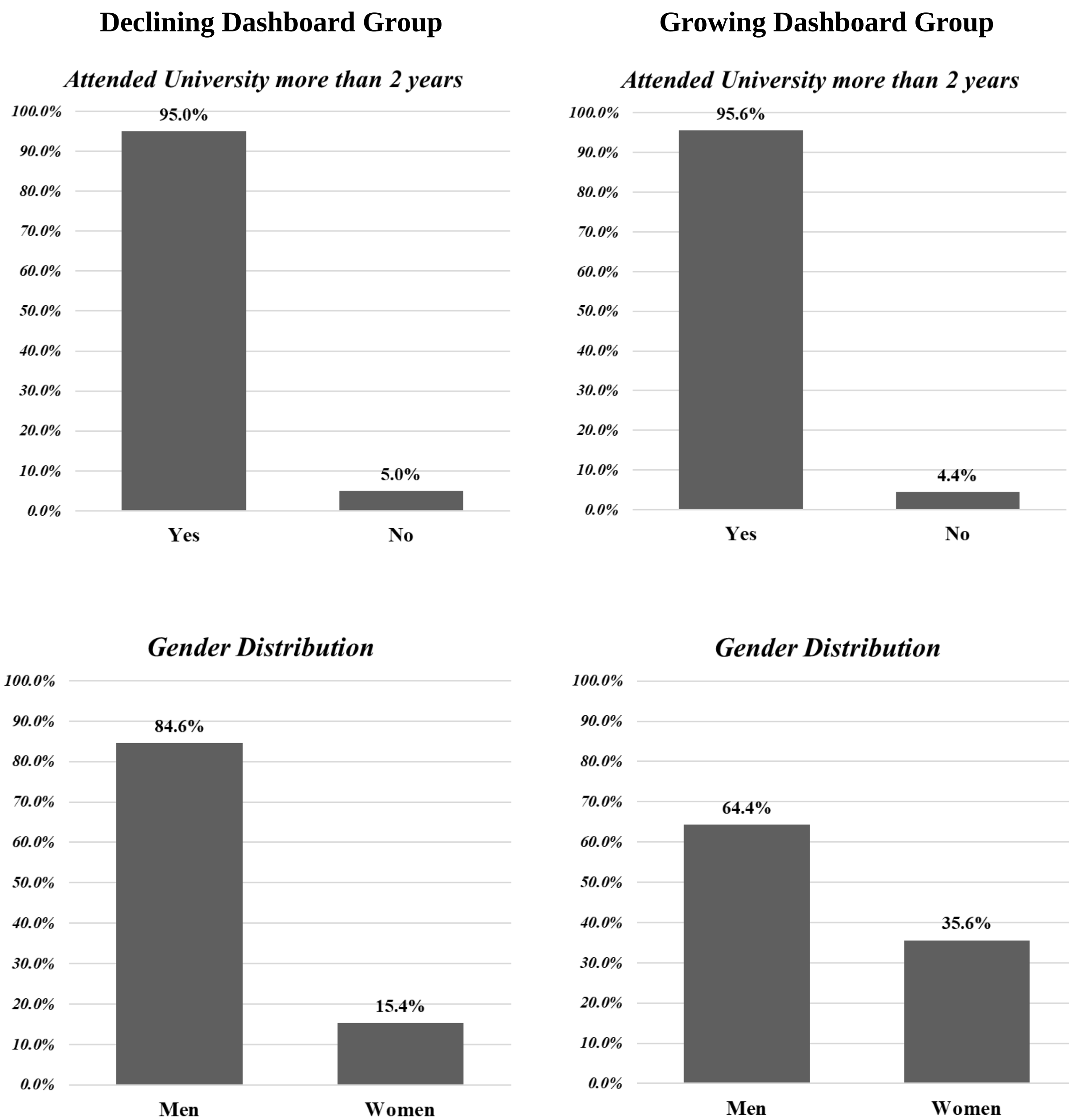



***Figure 4.1*** Proportion of participants that had attended university for more than two years and the gender distribution of the Declining (left) and Growing (right) Dashboard groups.

Table 4.1. below shows the correct base rate compared to the average ranks from the two Dashboard groups.

***Table 4.1*** Average rankings of the eleven main distribution centers in the dashboards for the two Dashboard groups and Pearson product moment correlations with the correct base rates

| Main Delivery Centers in Dashboards | Correct Base Rates $_A$ | Average Ranks in the Declining Dashboard Group | Average Ranks in the Growing Dashboard Group |
|---|---|---|---|
| Athens | 1 | 4.87 | 4.30 |
| Brussels | 7 | 6.21 | 6.23 |
| Copenhagen | 6 | 6.44 | 6.20 |
| Helsinki | 8 | 6.62 | 7.28 |
| London | 10 | 7.41 | 7.65 |
| Madrid | 5 | 6.26 | 5.70 |
| Paris | 11 | 6.38 | 7.60 |
| Riga | 2 | 4.62 | 4.38 |
| Rome | 3 | 5.05 | 4.85 |
| Stockholm | 9 | 6.79 | 6.73 |
| Tallinn | 4 | 5.05 | 5.65 |
| *Correlations between group's average ranks and correct base rates* | | ***r = 0.888**** | ***r = 0.975**** |

***Note:*** $_A$ Correct base rates simply means that based on the dashboard-based data visualizations it would have been correct to assign the delivery center that particular rank

The proportion of variance accounted for ($r^2$) of the Declining Dashboard Group is 78,9 % and 95,0 % for the Growing Dashboard group. On the next page, Figure 4.2 display these values in scatter-plots highlighting that both of these groups had positive linear associations between their rankings and the correct base rates.

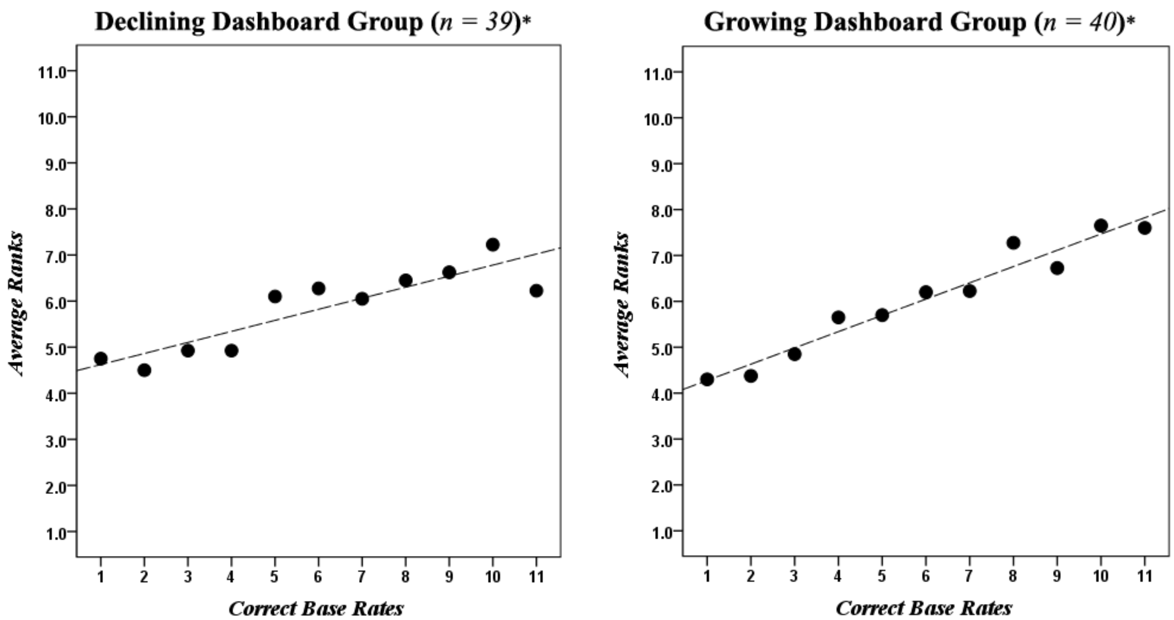



***Figure 4.2*** Scatterplots of the two Dashboard Groups' Average Ranks correlated with the Correct Base Rates. *The Declining Group (left) had 2 missing values and the Growing Dashboard Group (right) had 6 missing values.

## 4.2 Interpretation Accuracy Groups & Decision Bias

In this section the measure of Interpretation Accuracy was divided into two groups with the median (= *96.82*) as the cut point, yielding one LOW (n = 44) and one HIGH group (n = 43).
The LOW group now had an Interpretation Accuracy percentage of (M = 87.56, SD = 9.56) while the HIGH group had an Interpretation Accuracy of (M = 98.35, SD = 0.72).

The age of the participants in the Low Interpretation Accuracy group was (*M* = 35.6, *SD* = 8.6) and (*M* = 33.2, *SD* = 10.5) for the other group. The reported amount of work experience (in years) among the participants in the Low group was (*M* = 10.7, *SD* = 9.2) and (*M* = 8.9, *SD* = 8.7) for the participants in the High group.

Figure 4.3 on the next page shows the proportion of participants who reported to have attended university for at least two years and the gender distribution between the two natural groups.


***Figure 4.3*** Proportion of participants that had attended university for more than two years and the gender distribution of the Low (*left*) and High (*right*) Interpretation Accuracy groups.

Just as in the last section, Table 4.2 below displays the correct base rates compared to the average ranks but from the two natural groups Low/High Interpretation Accuracy with the Pearson product moment correlations for these relationships.

***Table 4.2*** Average rankings of the eleven main distribution centers in the dashboard from the two natural groups (Low/High) Interpretation Accuracy

| Main Delivery Centers in Dashboards | Correct Base Rates $_A$ | Average Ranks in the Low Interpretation Accuracy Group | Average Ranks in the High Interpretation Accuracy Group |
|---|---|---|---|
| Athens | 1 | 5.63 | 3.51 |
| Brussels | 7 | 5.95 | 6.49 |
| Copenhagen | 6 | 6.78 | 5.85 |
| Helsinki | 8 | 6.78 | 7.13 |
| London | 10 | 7.1 | 7.97 |
| Madrid | 5 | 6.4 | 5.54 |
| Paris | 11 | 6.1 | 7.92 |
| Riga | 2 | 4.55 | 4.44 |
| Rome | 3 | 5.25 | 4.64 |
| Stockholm | 9 | 6.2 | 7.33 |
| Tallinn | 4 | 5.2 | 5.51 |
| *Correlations between group's average ranks and correct base rates* | | ***r = 0.700**** | ***r = 0.989***** |

***Note:*** $_A$ Correct base rates simply means that based on the dashboard-based data visualizations it would have been correct to assign the delivery center that particular rank

The proportion of variance accounted for ($r^2$) of the Low Interpretation Accuracy Group is 49,0 % and 97,8 % for the Growing Dashboard groups. On the next page, Figure 4.4 display these average ranks plotted against the correct base rates. While there is now more of a difference between the groups, the linear approximations of the relationship between the average rankings and the correct base rates are still positive i.e. both groups did not show an overall decision bias by neglecting the base rates in the dashboards.

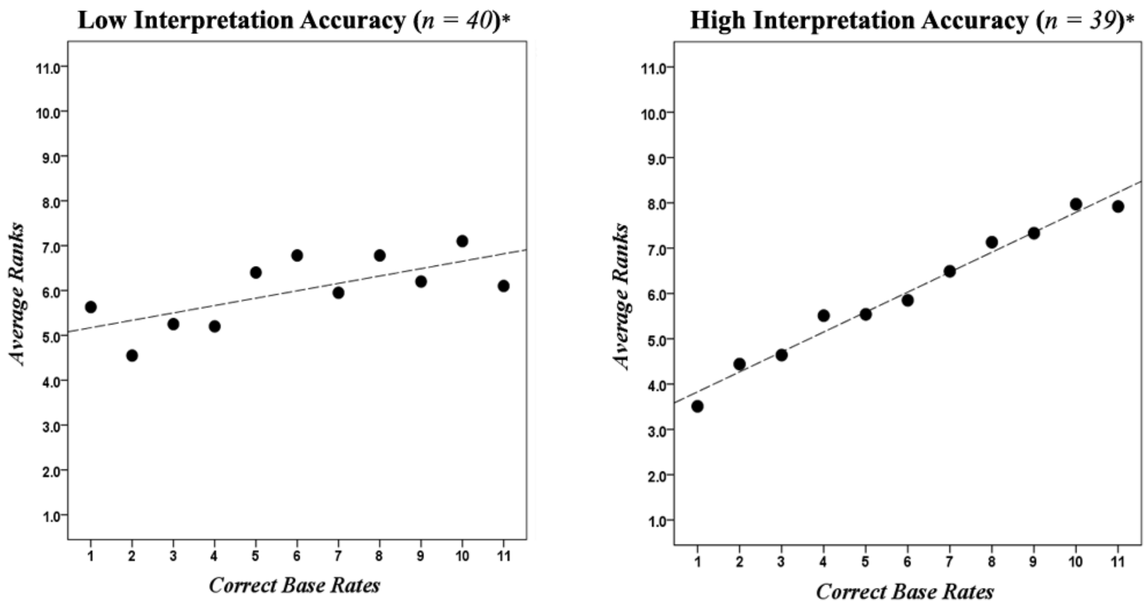



***Figure 4.4*** Scatterplots of the two natural groups' Average Ranks correlated with the Correct Base Rates. *The Low Interpretation Accuracy Group (left) had 4 missing values and the High Interpretation Accuracy Group (right) also had 4 missing values.

# 5. Discussion

This chapter will first provide an outline of the results. Next, a discussion follows concerning the critical considerations and limitations of this study. The implications of the results for both research and practice will be then addressed and in the last section future research will also be discussed.

The first section of the results concerned descriptive statistics between the two independent dashboard groups (Declining/Growing) and their degree of decision bias. All the demographic and background variables appear to be well balanced across the two groups except the gender distribution, see Figure 4.1 in the earlier chapter. This imbalance is probably explained by the fact that the double-blind randomization procedure was conducted across the whole sample. The department meetings usually had an imbalanced gender distribution and some departments (although they were few) had more women than men. Since the randomized stack of the questionnaires sometimes had sequences where one of the conditions would come up several times in a row – this could explain the gender imbalance. Given the rather small sample sizes, and the fact that there were few women, these descriptive statistics are explained by as few as five women randomly ending up in the other group. However, gender is a standard demographic variable and to our knowledge there is no research that suggests that either decision biases or accurate interpretations of data visualizations are well explained by a person's gender.

Table 4.1 and Figure 4.2 describe the degree of bias of the two Dashboard groups. These results are related to the fist research question of this study. The average rankings of the eleven distribution centers in the dashboards are quite similar between these two groups (the proportion of variance accounted for or $r^2$ of the Declining Dashboard Group is 78,9 % and 95,0 % for the Growing Dashboard group. The caeteris paribus interpretation of the $r^2$ value here would be that the correct base rates respectively explain 78,9 % and 95,0 % of the variation of the average ranks in these groups. Both groups correlation coefficients and the linear approximations in the scatter-plots (see Figure 4.2) suggests the average rankings of the participant are not biased in relation to the correct base rates of the dashboards. The average rank and the correct base rates covariate – If the bubble chart shows that a particular distribution center has a high cost, then the participants in both Dashboard groups tended to rank the likelihood as high that that particular distribution center contributed to the yearly cost change in the line chart. The same tended to be true for low costs. This is in line with statistical logic, see section 2.4. for a refresh of the underlying rationale of that logic. Still, these results do not suggest much of a relationship between decision bias and the similarity features of the dashboards and this is true for both dashboard groups i.e. Declining or Growing data categories. This answers the first research question of this thesis.

The second section of the results had the same structure as the former with the difference being that it concerned descriptive statistics between the two natural groups (Low - / High Interpretation Accuracy) and their degree of decision bias. This section is related to the second research question.

It is noteworthy that Interpretation Accuracy was a continuous measure that was divided into two equal groups with the median as the cut point. Interestingly enough, all of the demographic and background variables appear to be well balanced across the two groups again, except the gender distribution, see Figure 4.3 in the earlier chapter.

Table 4.2 and Figure 4.4 describe the degree of bias of the two natural groups. These results are a bit more inconclusive compared to the Dashboard groups. The proportion of variance accounted for of the Low Interpretation Accuracy group is 49,9 % and 97,8 % for the High Interpretation Accuracy group. Both groups correlation coefficients and the linear approximations in the scatter-plots shown in Figure 4.4 suggests the average rankings of the participant are not biased in relation to the correct base rates of the dashboards. In reference to section 3.3.3. on the measure on Decision Bias, if the linear approximation of the relationship between the mean ranks and the correct base rates would be negative, that would constitute evidence of a decision bias known as base rate neglect. This did not happen in either of the two Interpretation Accuracy groups but the average ranks of the Low group showed a poorer covariation with the correct base rates, see Figure 4.4. The line in this scatter-plot is fairly close to horizontal but still positive. However, based on the earlier definition neither of the groups show any clear evidence of decision bias but these results are more inconclusive. The $r^2$ value drops from having accounted for 97,8 % of the variation in the High Interpretation Accuracy group to 49,9 % in the Low Interpretation Accuracy group. The large drop in the $r^2$ value does suggest a relationship between decision bias and the ability to accurately interpret the data visualizations of a dashboard but these results are quite inconclusive. This rather nuanced paragraph answers the second research question.

In summary, the answer to the first research question illuminated the link between similarity features of dashboard-based data visualizations and the tendency of business practitioners to make biased decisions. The answer to the second research question focused the relationship between how accurately dashboard-based data visualization are interpreted and potential decision biases. While these two questions have now been answered, the understanding and meaningfulness of these results are connected to the limitations of this research.

## 5.1 Limitations and Critical Considerations

Now that the results of this study have been run through, it is important to again echo the sentiment of Humphreys and Weinstein (2009). Even though many steps were taken in this research design to keep a high degree of internal validity, the general principle of a field experiment is that we have control over the assignment to the "treatment" i.e. a specific dashboard by a double-blind randomization procedure. By conducting the research in the field i.e. during department meetings in an operating organization, some control over the treatment itself is also lost. The dependent variable (Decision Bias) can therefore take on values given by “nature” rather than being set at the discretion of us as researchers. We took this into consideration by also examining the relationship between decision bias and the two natural groups of High and Low interpretation accuracy. The format of this study i.e. field experiments looks to increase the ecological validity of the findings (Bhattacherjee, 2012; Shaughnessy et al., 2012). However, the natural groups are rather quasi-experimental given their lack of randomization and the independent groups would be well complimented with a control group to reach the bar of a true experiment in terms of validity. We would argue that this research design works but it becomes a relatively crude approach. One should take caution when interpreting the results of a field experiment since all potential confounds were not controlled for and the statistics of this particular study are descriptive rather than inferential.

Regardless of all the other potential confounds inside an organization such as the personal life of each participant, their work experience or educational background etc. the most salient confound in this study is “time”. One could reasonably argue that the reason why people in the High Interpretation Accuracy group managed to get their rankings closer to the correct base rates than the Low Interpretation Accuracy group is not because there is a relationship between interpretation accuracy and decision bias but because of the time they took to make their decisions. Maybe slow decision-makers are more accurate and therefore less biased? If the individual time to completion or even the time it took to complete each question related to the dashboard would have been captured as a variable, this could have been used as a control measure in a statistical analysis. This feature of monitoring elapsed time existed in some experiments and survey software, but the rendering issues that were experienced during the development of the instrument were too critical to ignore which was why the paper-based questionnaire was chosen as the best possible trade-off. However, it can also be noted that “time” was not mentioned as a confounding factor or control variable in the experimental design of the Kahneman and Tversky (1973) study that was used as a model for this one. Yet, conditions were set up to accommodate a similar and non-interruptive environment for all the business practitioners that participated, everybody received the instructions and then answered the questions sitting down in either the departments conference room (if the meeting was held there) or at their individual desks (if the meeting was held at their departments office environment).

Another limitation concerns the endemic usage and appearance of dashboards in organizations. The theoretical chapter outlined that there is no uniform definition neither in practice nor research of either a dashboard or a data visualization (Few, 2008; Mayer, 2000; Speier, 2006; Yigitbasioglu & Velcu, 2012). This made it both complex and time consuming to devise the data collection instruments and to operationalize the concept of dashboard-based visualizations. Although many steps were taken in this study to transparently and systematically develop the dashboard-based data visualizations that were used, one could still argue that these were a serious limitation. One could argue that making decisions based on costs in different European locations based on a bar -, line -, pie and bubble chart is a rather contrived situation lacking in ecological validity.
A digital dashboard would probably have been a more ecological situation but devising an instrument that would act as a properly designed dashboard while also capturing the necessary data was not a viable option.

The criticism concerning the lack of ecological validity in this study could very well be true but given how different dashboards are in both their usage and appearance, it is hard to gauge exactly how true this criticism would be. This line of argument would also miss the overall aim of this study.

This study intended to capture the sentiment of the increasing trend of visualizing data and then consolidating them into a dashboard for non-expert users (Negash & Gray, 2008; Lee et al., 2017b). The intention was to move this format into the hands of generic business practitioners, with varying degrees of experience in using dashboard-based visualizations as decision aids and to scope out their tendency to commit a specific decision bias. Interviews could have provided more depth, but decision biases are largely unconscious cognitive processes and are therefore hard to gauge through interviews. An increased sample size would have made the results of this study more generalizable while the culture of the specific organization sets this limit. The particular large-scale IT organization that was sampled in this study operates in a complex market and the employees are arguable highly educated and skilled in intellective tasks, as the Results chapter indicated. In section 3.1. we highlighted that earlier research has demonstrated that individual differences in a wide array of domains can influence peoples' degree of interpretation accuracy which means that the results obtained at this particular organization might be quite endemic. This implication will be further explored in a later section.

## 5.2 Implications for Research

An implication about the methodology and analysis that this research was modelled upon exits. The experimental study that was thoroughly outlined in the theoretical chapter (e.g. Table 2.1.) is reviewed as a seminal example of how the representative heuristic can be demonstrated in Kahneman and Tversky (1974). This article has more than 11 000 citations on the Web of Science and more than 45 000 citations on Google Scholar (Google Scholar, 2018). However, this approach of computing the Pearson product moment correlations for mean ranks is somewhat controversial for two reasons. Most scholars and their respective books on data analysis stress that these correlations require interval scales (Bhattacherjee, 2012; Pallant, 2013, Aron, et al., 2014). While this is a mathematical truth, it is a requirement that is often violated in social sciences. Aron et al. (2014) concedes that scholars within the behavioral sciences often treat scales with more than ten levels as interval-scales. Even though the ranking scale of this study had eleven levels while it was also modeled after the analyses of two scholar's (e.g. Tversky and Kahneman, 1973; 1974) with an impeccable academic record, there is another controversy. These analyses have to fall into the category that Vogt and Johnson (2011) calls Ecological Correlations i.e. correlations between variables that are based on grouped data such as averages. Furthermore, the Ecological Fallacy refers to the error of drawing false conclusions about individuals based only on data from groups. Even though the above definition falls dangerously close to some of the decision biases outlined in the literature, we have not seen any other researchers put up this red flag regarding their analysis. The Pearson product moment correlations between the judged base rates and estimated likelihood was rather clear-cut in the Kahneman and Tversky (1973) study ($r$ = - 0.65) demonstrating a clear tendency of base rate neglect. The correlation between the similarity and likelihood groups ($r$ = 0.97) was also very clear. This research process has made us sensitive to the fact that obtaining high correlations from average ranks is far more probable than when individual ranks are computed. This red flag could be viewed as a wrinkle on some of the prior empirical evidence associated to the representative heuristic. For the sake of rigor, we used three statistical measures in this field experiment to adjudicate weather any of the groups showed a tendency of decision bias, see section 3.3.3. Again, these results did not suggest much of a relationship between decision bias and the similarity features embedded in the dashboard-based data visualizations.

If one would be inclined to compare the results of earlier research where people make biased decisions based on similarities between textually represented classes and objects such as graduate specializations and personality descriptions, as shown in Kahneman and Tversky (1973), with the results obtained from this study: The participants seemed a lot less persuaded to infer likelihood based on the similarities of dashboard-based data visualizations. This can be viewed as an argument “for” instead of “against” dashboards. Kahneman (2003) argues that people are not accustomed to think hard and are therefore content to trust in the most plausible choice that comes to mind. However, this might be true for many kinds of decision tasks in life but in the context of organizational decisions based on dashboard-based data visualizations the results of this study do not lend too much credence to Kahneman’s (2003) argument. This is somewhat intuitive in light of Pappas and Whitman’s (2011) assertion for combining different visualizations into a dashboard as a step in the right direction to aid decision-makers in making fact-based decisions. As mentioned in section 1.4, this study intended to extend Tversky and Kahneman’s (1974) research on the representative heuristic and decision biases to the context of dashboard-based data visualizations inside a large-scale organization and explore whether similar results would be obtained.

Boy et al. (2014) highlight that when data visualizations are designed or when evaluations of new visualization systems are being conducted, it is important to be able to pull apart the potential effectiveness of the data visualization and the actual ability of users to understand it and using it as a decision aid. This point was taken into consideration by also examining the relationship between decision bias and the two natural groups of High and Low interpretation accuracy. While the results were rather inconclusive, they did suggest that the group of people with a high ability for interpreting the data visualizations had lower degrees of bias. It could be claimed that the theory regarding the representative heuristic would be somewhat inflated if it can be partially explained by peoples’ inability to interpret the decision task. However, the strength of the theory is that it is based on very conclusive empirical results. This study was explorative and did not yield very conclusive results and is therefore not the best challenge to that theory.

This thesis also latently took up Speier’s (2006) call for academicians and researchers to explore and assess how broader selection of data visualization formats influence decision accuracy.

As researchers, we extended this call to four popular formats in the context a dashboard, see Appendix 1 and 2. From a practical perspective the results partially suggest that it could be important to determine users' ability to interpret data visualization formats before designing dashboards. The endemic nature of dashboards restricts this in practice, but it would arguably be wise of data analysts and visual designers to carefully consider and test what visualizations to deploy in potential dashboards.

## 5.3 Implications for Practice

Holsapple (2008) states that "making decisions" is the most frequently occurring activity in any organization, a sentiment echoed by Hammond et al. (1998) who also posit that making decisions is the most important job of any business executive. Today, DSS are heavily relied on employed in many organizations to assist and aid in decision making. As mentioned in the purpose section of this thesis, if clear evidence for different degrees of decision bias due to irrelevant similarity features inside a dashboard could be demonstrated, this would problematize the expanding usage dashboard-based data visualizations.

As was mentioned in the introduction of this thesis, Apple recently revealed a new "privacy dashboard" where the target audience is their whole European customer segment in the effort of complying with Europe's General Data Protection Regulation or GDPR (Hern, 2018). If decision biases could be demonstrated in a context such as at the offices of highly skilled and educated employees at a large IT-organization, that finding would arguable have implications for a wider user audience. As researchers, we had an underlying anticipation that there might be a clear relationship between similarity features of dashboard-based data visualizations and decision biases. However, the results of this study did not demonstrate any such finding. This statement is true in a fairly narrow context where four popular data visualizations were combined into a dashboard to explore one specific bias arising from the representative heuristic. This was explored at the offices of business practitioners at a large-scale IT organization. The purpose was to scope out the nature and extent of a potential relationship between dashboard-based data visualizations, interpretation accuracy and decision biases. Although inconclusively, the results of this field experiment suggested that a workforce with a high ability for interpreting intellective decision-tasks such as those featured in the VLAT, seems to be less likely to show tendencies of systematic errors in decision-making in this organizational context.

## 5.4 Further Research

To drive investigations of this knowledge domain forward, we would suggest the following: The design that was used in this study was rather crude. As was eluded to earlier, the measure that was based on the Kahneman and Tversky (1973) study can definitely be further optimized. The major drawback of this research design is that it does not allow for individual control variables, which would be a great advantage to further increase the validity of future IS studies in general. The two independent groups and the two natural groups fulfill their purpose, but the latter groups are rather quasi-experimental given their lack of randomization and the independent groups would be well complimented with a control group to move closer to the validity of a true experiment. The research design of this study works, but at a fairly crude level. For future studies we would recommend parametric alternatives such as multiple linear regression (MLR) or analysis of covariance (ANCOVA). Such analyses would yield more precision while offering a more scientifically main-stream interpretation. However, that would also require creative scholars, data analysts or UI designers to come up with new better ways of operationalizing different kinds of systematic errors in decision-making. This is a rather complex area of research, pot-marked with difficulties. Devising the research instruments, randomly selecting different departments and then accessing them was a fairly time-consuming aspect of this study. If one could leverage the right IS infrastructure, those processes would have been less complex with a digital or online data collection instrument. In section 3.1. we highlighted that earlier research has demonstrated that individual differences in a wide array of domains could influence peoples' degree of interpretation accuracy. In summary, this means that the results obtained at this particular organization might be quite endemic. Exploring whether these results in fact are specific to this particular organization would also be a worth-while project. Extending this research design to another population such as students or people with a different degree of knowledge regarding statistics or data visualizations would create an informative comparison.

Concerning future research, a good starting point for both scholars or practitioners would be a widely used dashboard solution that could be manipulated and tested in several conditions. In this study, we tried to offer a transparent roadmap for that process in the methodological chapter. Here, dashboard-based data visualizations were manipulated in way so that several data visualizations had similar graphical features that could potentially augment peoples' tendencies to use the representative heuristic and thereby make biased decisions based on the dashboard. However, this is only one potential context where dashboards might not be fully optimal as decision aids. KPI:s can create an anchoring effect and thereby bias important decisions and same is true for the graphical color choices of data visualizations (Bera, 2016; Van Der Heijden, 2013). Informative research is being produced in this area already, but a more systematic approach would probably be instructive. If a consensus could be reached about a more common dashboard solution, a taxonomy of potential decision biases could be investigated and hopefully safeguarded against.

Although rather inconclusive, the results of this study suggest a relationship between the ability to interpret data visualizations and the tendency to make biased decisions. There might also be potential for the vendors of data visualization authoring tools to more proactively test and ensure that their user audience has an accurate understanding of the features that are provided with these analytical tools.

# 6. Conclusion

It reasonable to argue that some monetary resources have been sacrificed to conceive this thesis. However, all of the managers and the people that chose to participate in this study were both open and generous with their time and attention. That same sentiment can in fact be extended to the entire organization where this research was conducted. It is generally difficult to get access to around a hundred business practitioners for roughly 20 minutes each when their organization has a turnover of several hundred thousand US $ per employee, as was the case with this particular organization. Field experiments are relatively rare mostly because of the difficulties that are related to manipulating treatments and controlling for extraneous effects in the field setting and quantitative explorative studies are also rarity (Bhattacherjee, 2012). With that information in hand, the objective of this thesis was to produce a unique knowledge contribution to the IS research field and organizations by to extending Tversky and Kahneman's (1974) seminal research on decision biases to the context of dashboard-based data visualizations inside a large-scale IT-organization and explore whether similar results would be obtained.

Rather than problematizing the expanding usage of dashboards as decision aids, the results of this thesis partially became an argument in favor for visually represented information compared to textual information. Yet, several confounding factors of the field experiment could partly explain this difference. Given the rather high education level of the sample, these participants might simply be better at assessing outcomes defined by likelihood than other participants would have been. The time it took for each individual to make up their mind regarding these dashboard-based data visualizations might have been another factor. The attempt to operationalize similar graphical features between the data visualizations of the dashboards to induce the heuristic of representativeness did not illuminate any clear relationship to decision biases in this context. This result however, answered the first research question. The result about the relationship between decision bias and the ability to accurately interpret the data visualizations of a dashboard are more nuanced. Neither of these two natural groups showed a clear tendency of bias but the results of the two natural groups were quite different, suggesting that there might be a relationship between decision bias and interpretation accuracy in this context. This latter result pertains to the second research question. The purpose of this research illuminated but a small fraction of the landscape between dashboards, data visualizations, decision biases and interpretation accuracy.

Given their task-specific usage and appearance, systematic errors or biases regarding the decisions that are made based on dashboard-based data visualizations has widely different consequences. In the context of marketing, money might be unnecessarily spent, in accounting, important laws might be broken and if the dashboard concerns a nuclear power-plant, lives might be lost. As the usage of dashboard-based data visualizations continues to expand, more research and development concerning how to improve these decision aids becomes increasingly important for our society.

# Appendix 1. Descriptive Statistics on the Interpretation Accuracy questions

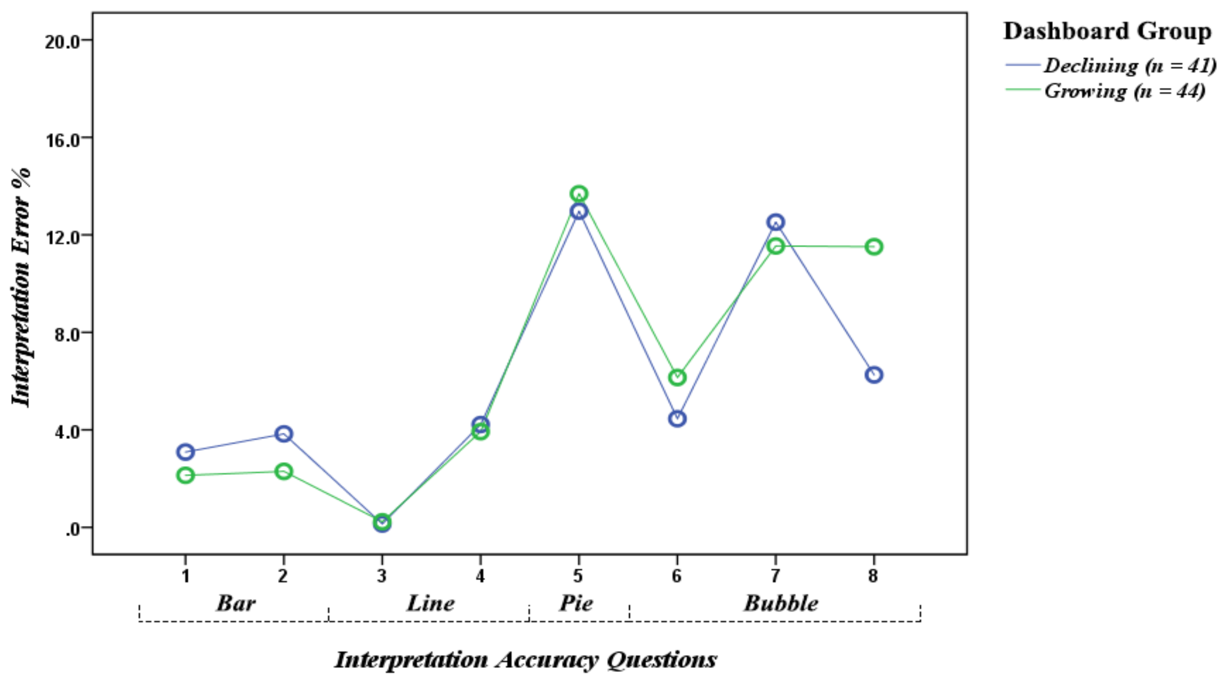
Descriptive Statistics on the eight Questions used for the measure of Interpretation Accuracy ($n = 85$) i.e. two missing values

| ***Questions*** | ***Mean Interpretation Error %*** | ***Std. Deviation %*** |
|---|---|---|
| 1. Bar - Retrieve Value | 2.66 | 9.49 |
| 2. Bar - Decide Range | 5.19 | 17.90 |
| 3. Line - Retrieve Value | 0.22 | 0.59 |
| 4. Line - Decide Range | 4.07 | 15.54 |
| 5. Pie – Retrieve Value | 13.24 | 15.18 |
| 6. Bubble - Retrieve Value | 5.42 | 16.90 |
| 7. Bubble - Retrieve Value | 14.01 | 27.30 |
| 8. Bubble - Decide Range | 11.06 | 25.94 |

***Note:*** Together, these questions touched on every dimension of the four data visualizations that could be interpreted in the dashboard. The Task categories are based on "VLAT: Development of a Visualization Literacy Assessment Test" by Lee, Kim & Kwon, 2017, IEEE Transactions On Visualization and Computer Graphics, 23, 1, p. 558.

# Appendix 2. Extraneous Analysis of Interpretation Accuracy questions

To ensure that the visualization formats were interpreted in a similar fashion between the two independent groups, a mixed between-within analysis of variance was performed to assess the impact of the eight questions and the Dashboard Groups on the Interpretation Error %. The analysis was conducted according to Pallant's (2013, p. 284) recommendations. The normality assumption was relaxed due to ($n > 30$), Levene's test was not significant while the Box M statistic was significant, violating the equality of the covariance matrices. In this analysis the eight questions were the independent repeated measure and the independent Dashboards (Declining/Growing) were the between group factor. There was no significant interaction between Interpretation Errors and the Dashboards, Wilks' Lambda = 0.98, $F$ (7, 77) = 0.27, $p$ = .96, with a partial eta squared of .02. There was a significant main effect for Interpretation Error, $F$ (7, 77) = 13.55, p = .000, partial eta squared = .55. The main effect for the Dashboard Groups was not statistically significant: $F$ (1, 83) = 0.14 p =.71. These results are well received because they highlight that there is no interaction between the measure of Interpretation Error % and the two dashboards and two groups tended to have similar degrees of interpretation error. The significant result could be expected since the questions were related to four different data visualization formats in the dashboards. This latter result highlights that the participants had varying degrees of interpretation error % depending on the individual format of the data visualization and the particular question.




***Appendix 2***. Mixed between-within ANOVA of interpretation error % by each of the eight questions of the two Dashboard Groups

# Appendix 3. Data Collection Email (*in Swedish*)

**Brian Ochan**

**From:** Brian Ochan
**Sent:** den 25 april 2018 11:22
**To:** [redacted]
**Cc:** [redacted] Kristoffer Bergram
**Subject:** Besöka Standup/Morgonmöte

Hej [redacted]
Våra namn är Brian & Kristoffer och vi samlar just nu in data för vårt exjobb här på [redacted] Vi kommer genomföra små fältexperiment relaterade till dashboards och beslutsfattning. Vi hoppas kunna låna ert teams uppmärksamhet under 5-10 miuter under EN av era standups/morgonmöte under vecka 18 och 19.

Av etiska skäl vill vi också nämna att samtliga svar kommer att vara anonyma, deltagandet är fullständigt frivilligt och vår forskningsfråga rör ej ert teams prestationer. Tanken är att vi kommer förbi under ett möte när teamet är samlat och efter att ni är klara så kommer alla få en kort instruktion och ett randomiserat papper eller en url adress och sedan genomföra en enkät som tar mellan 2-5 minuter.
Eftersom ert team har blivit slumpmässigt utvalt så skulle vi verkligen uppskatta ert organisatoriska och vetenskapliga bidrag.

Vi hoppas ni kan tillgodose våra önskemål.

Bästa Hälsningar

Brian Ochan och Kristoffer Bergram

# Appendix 4. Debriefing Email

**Brian Ochan**

**From:** Brian Ochan
**Sent:** den 18 maj 2018 17:40
**To:** [redacted]
**Cc:** Kristoffer Bergram; [redacted]
**Subject:** Debriefing (Data driven biased decision-making?)

Hi!

If you have been answering some questions related to dashboards recently, we want to thank you for your participation in our data collection procedures. Now that the data collection is finished we can give the full picture of this research. The goal of this MSc. thesis was to explore the potential relationship between decision biases (systematic errors in decision-making) and some specific features in dashboard-based data visualizations such as declining or increasing data categories.

It is important for us to provide you with full transparency. As participants, you were informed that the questionnaire examined the relationship between dashboards and decision-making. This information was provided to preserve the quality and validity of the results by not influencing or contaminating your answers. By not giving you full disclosure, we made sure that any decision biases we wanted to capture would manifest themselves naturally - based on the dashboards.

In the field experiment, there were three randomized groups. One group received a questionnaire that showed an increasing trend of all the data, and the second group received a questionnaire that showed a decreasing trend of all the data. These dashboards were designed to affect the ranking preferences in the last question to measure whether a clear bias could be detected. The first 8 questions were a control measure to help determine that everybody could correctly interpret each of the 4 data visualizations on the dashboard. The last question concerned potential biases in ranking preferences. The third group had a mixture of increasing and declining data visualizations and served as a control group.

We are currently doing our data analysis and hope to be done within the next two weeks. If you have any more questions or feedback, you can find us in [redacted] We will have a presentation of the thesis and the findings on 14th June 2018 in the [redacted] at time yet to be decided.

You are all welcome to join us and thanks once again. Please feel free to share this email with your teams that participated.

# Appendix 5. Questionnaire Growing Dashboard Group

This survey has two parts: In the first part you will be asked 9 questions related to a dashboard with four data visualizations. In the final question you will be asked to make a decision where you rank your preferences. The second part of the survey consists of 5 demographic questions. The dashboard on the next page is not directly linked to [redacted] or its current partners.

All the answers you provide will be analyzed and used for a MSc. degree project at Lund University focused on dashboards and decision-making.

Please give all your answers individually. Your participation is completely voluntary, you can abort your participation at any time, without any explanation and all your answers will be processed anonymously.

I have read and understood the above information ☐

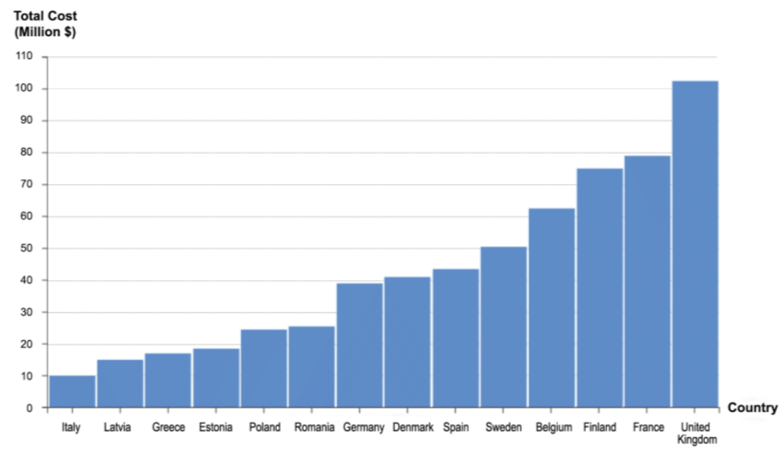
We thank you for participation.

***Kristoffer Bergram & Brian Ochan***

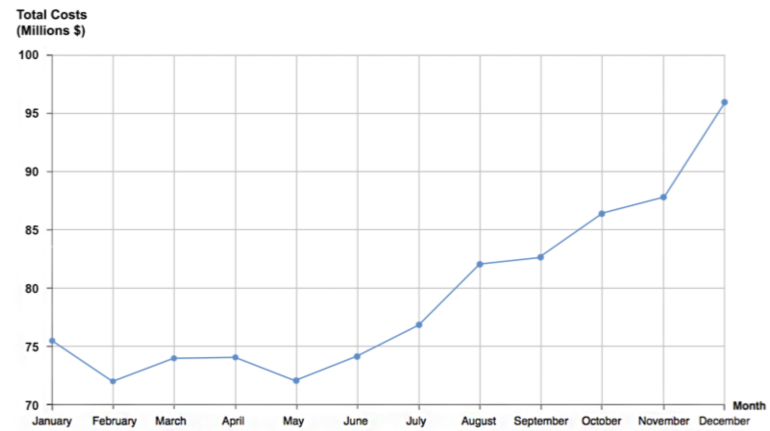
***Lund University, Department of Informatics***

**Note: The different visualizations on this dashboard had random starting orders**

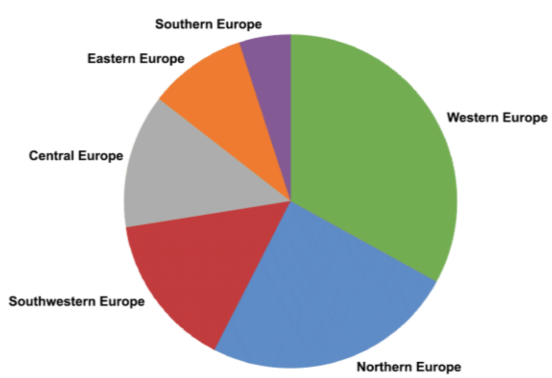
## Costs across Europe by country



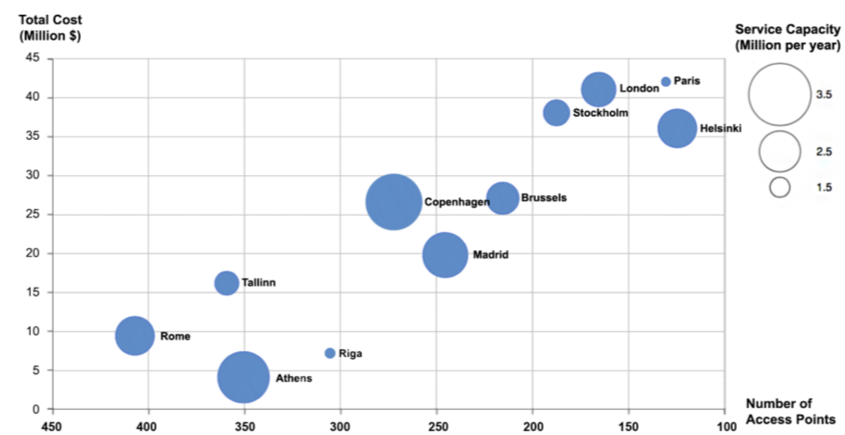

## Monthly Costs in Europe during 2017




## Cost share (%) across European Regions




## Costs by main delivery centers in Europe



**1.** What is the total cost in Finland? ANSWER: ___________

**2.** What is the range of total costs between the European countries? ANSWER: _____ *to* ____

**3.** What is the total cost of February during 2017? ANSWER: ___________

**4.** What is the total range of monthly costs during 2017? ANSWER: _____ *to* ____

**5.** What is the combined cost share (%) of Eastern and Central Europe? ANSWER: ___________

**6.** What is the total cost of the Brussels delivery center? ANSWER: ___________

**7.** How many access points does the main delivery center with the highest service capacity have? ANSWER: ___________

**8.** What is the range of total costs between the main delivery centers? ANSWER: _____ *to* ____

You are now asked to make a decision based on your judgement of the dashboard…

**9.** Please rank the eleven main delivery centers in order of the likelihood that they produced the yearly cost change from January to December of 2017 (*from* ***11 = Highest likelihood*** *to* ***1 = Lowest likelihood****)*

ANSWER:

☐ **Athens**

☐ **Brussels**

☐ **Copenhagen**

☐ **Helsinki**

☐ **London**

☐ **Madrid**

☐ **Paris**

☐ **Riga**

☐ **Rome**

☐ **Stockholm**

☐ **Tallinn**

**10.** Your current age: **______ years.**

**11.** Your gender: **Woman** ☐ **Man** ☐ **Other** ☐

**12.** Your nationality: ______________________________

**13.** During your career, how many years have you worked: **_____ years of work experience.**

**14.** Have you studied at a University for more than 2 years: **Yes** ☐ **No** ☐

# Appendix 6. Questionnaire Declining Dashboard Group

This survey has two parts: In the first part you will be asked 9 questions related to a dashboard with four data visualizations. In the final question you will be asked to make a decision where you rank your preferences. The second part of the survey consists of 5 demographic questions. The dashboard on the next page is not directly linked to [redacted] or its current partners.

All the answers you provide will be analyzed and used for a MSc. degree project at Lund University focused on dashboards and decision-making.

Please give all your answers individually. Your participation is completely voluntary, you can abort your participation at any time, without any explanation and all your answers will be processed anonymously.

I have read and understood the above information ☐

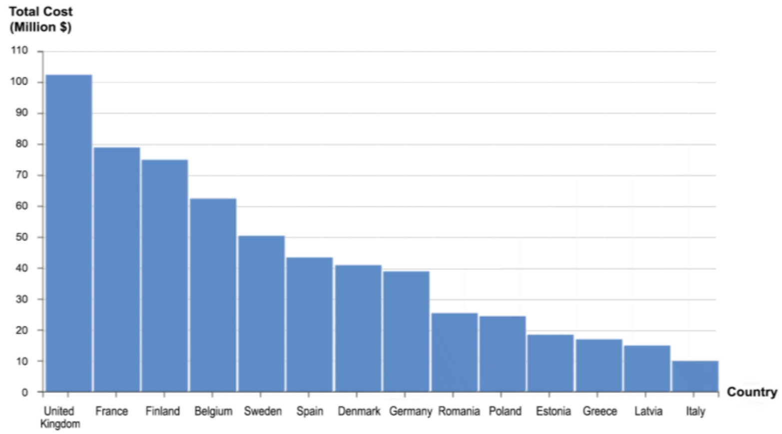
We thank you for participation.

***Kristoffer Bergram & Brian Ochan***

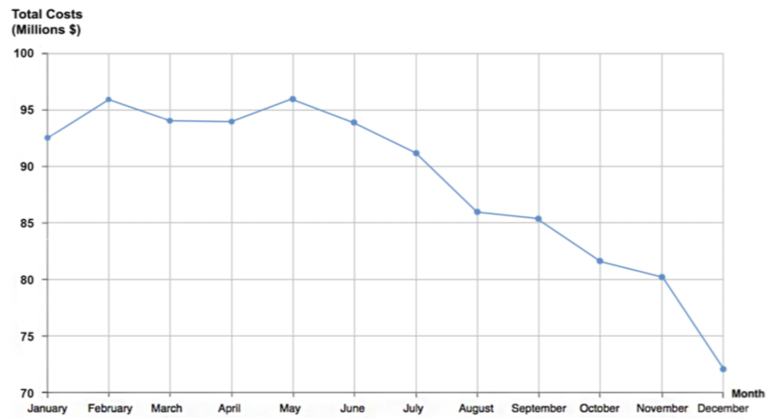
***Lund University, Department of Informatics***

**Note: The different visualizations on this dashboard had random starting orders**

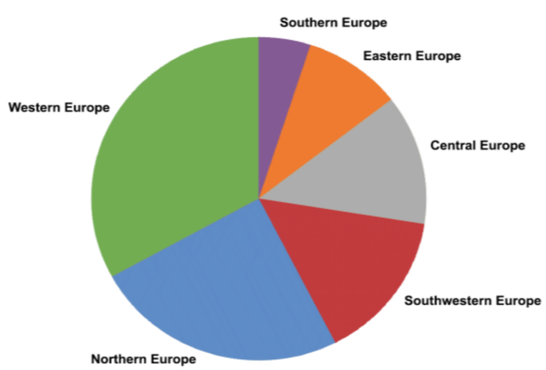
## Costs across Europe by country



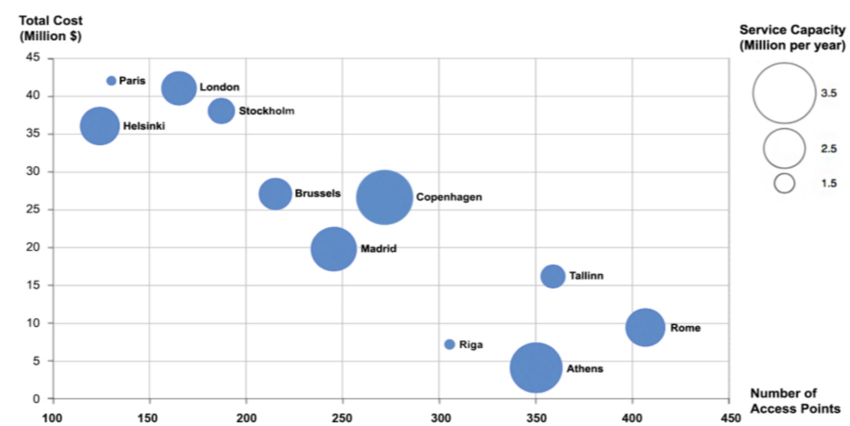

## Monthly Costs in Europe during 2017




## Cost share (%) across European Regions




## Costs by main delivery centers in Europe



**1.** What is the total cost in Finland? ANSWER: ___________

**2.** What is the range of total costs between the European countries? ANSWER: _____ *to* ____

**3.** What is the total cost of February during 2017? ANSWER: ___________

**4.** What is the total range of monthly costs during 2017? ANSWER: _____ *to* ____

**5.** What is the combined cost share (%) of Eastern and Central Europe? ANSWER: ___________

**6.** What is the total cost of the Brussels delivery center? ANSWER: ___________

**7.** How many access points does the main delivery center with the highest service capacity have? ANSWER: ___________

**8.** What is the range of total costs between the main delivery centers? ANSWER: _____ *to* ____

You are now asked to make a decision based on your judgement of the dashboard…

**9.** Please rank the eleven main delivery centers in order of the likelihood that they produced the yearly cost change from January to December of 2017 (*from* ***11 = Highest likelihood*** *to* ***1 = Lowest likelihood****)*

ANSWER:

- ☐ **Athens**
- ☐ **Brussels**
- ☐ **Copenhagen**
- ☐ **Helsinki**
- ☐ **London**
- ☐ **Madrid**
- ☐ **Paris**
- ☐ **Riga**
- ☐ **Rome**
- ☐ **Stockholm**
- ☐ **Tallinn**

**10.** Your current age: **______ years.**

**11.** Your gender: **Women** ☐ **Man** ☐ **Other** ☐

**12.** Your nationality: ______________________________

**13.** During your career, how many years have you worked: **_____ years of work experience.**

**14.** Have you studied at a University for more than 2 years: **Yes** ☐ **No** ☐

# References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17*(2), i – xxxii.

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2018). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*.

Amer, T. S. (2005). Bias Due to Visual Illusion in the Graphical Presentation of Accounting Information. *Journal of Information Systems*, *19*(1), 1-18.

Amer, T. S., & Ravindran, S. (2010). The Effect of Visual Illusions on the Graphical Display of Information. *Journal of Information Systems*, *24*(1), 23-42.

Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal, 16*(1), 55-78.

Arnott, D., & Pervan, G. (2014). A critical analysis of decision support systems research revisited: the rise of design science. *Journal of information Technology, 29*(4), 269-293.

Aron, A., Aron, E., & Coups, E. J. (2014). *Statistics for psychology*. Essex, England: Pearson Education.

Arunachalam, V., Pei, B. K. W., & Steinbart, P. J. (2002). Impression Management with Graphs: Effects on Choices. *Journal of Information Systems*, *16*(2), 183-202.

Baker, J., Jones, D. R., & Burkman, J. (2009). Using visual representations of data to enhance sensemaking in data exploration tasks. *Journal of the Association for Information Systems, 10*(7), 533.

Baysal, O., Holmes, R., & Godfrey, M. W. (2013). Developer dashboards: The need for qualitative analytics. *IEEE software, 30*(4), 46-52.

Bennet, A., & Bennet, D. (2008). The decision-making process in a complex situation. In *Handbook on Decision Support Systems 1* (pp. 3-20): Springer.

Bera, P. (2016). How colors in business dashboards affect users' decision making. *Communications of the ACM, 59*(4), 50-57.

Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices*: USF Tampa Bay Open Access Textbooks Collection.

Boland, R. J. (2008). Decision making and sensemaking. In *Handbook on Decision Support Systems 1* (pp. 55-63): Springer.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE transactions on visualization and computer graphics*, *20*(12), 1963-1972.

Busenitz, L. W. (1999). Entrepreneurial Risk and Strategic Decision Making:It's a Matter of Perspective. *The Journal of Applied Behavioral Science*, *35*(3), 325-340.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. London, UK: Academic Press.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88-98.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, *36*(4), 1165-1188.

Churchman, C. W. (1968). *Challange to Reason*. New York: McGraw-Hill.

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol, California: O'Reilly Media.

Friendly, M. (2008). A brief history of data visualization. In *Handbook of data visualization* (pp. 15-56): Springer.

Funkhouser, H. G. (1936). A Note on a Tenth Century Graph. *Osiris*, *1*, 260-262.

Google Scholar. (2018). judgment under uncertainty 1974. Retrieved from: https://scholar.google.se/scholar?hl=sv&as_sdt=0%2C5&q=judgment+under+uncertainty+1974&btnG=

Griffith, T. L., Northcraft, G. B., & Fuller, M. A. (2008). Borgs in the org? Organizational decision making and technology. In *The Oxford handbook of organizational decision making*: Oxford University Press.

Hall, D. J. (2008). Decision makers and their need for support. In *Handbook on Decision Support Systems 1* (pp. 83-102): Springer.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). The hidden traps in decision making. *Harvard business review*, *76*(5), 47-58.

Hayes, B. K., Hawkins, G. E., Newell, B. R., Pasqualino, M., & Rehder, B. (2014). The role of causal models in multiple judgments under uncertainty. *Cognition*, *133*(3), 611-620.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of risk and uncertainty*, *4*(1), 5-28.

Hern, A. (2018). What is GDPR and how will it affect you? *The Guardian*. Retrieved from: https://www.theguardian.com/technology/2018/may/21/what-is-gdpr-and-how-will-it-affect-you

Holsapple, C. W. (2008). Decisions and knowledge. In *Handbook on Decision Support Systems 1* (pp. 21-53): Springer.

Hosack, B., Hall, D., Paradice, D., & Courtney, J. F. (2012). A look toward the future: decision support systems research is alive and well. *Journal of the Association for Information Systems, 13*(5), 315.

Humphreys, M., & Weinstein, J. M. (2009). Field experiments and the political economy of development. *Annual Review of Political Science, 12*, 367-378.

Kahneman, D. (2003). A Perspective on Judgment and Choice - Mapping Bounded Rationality. *American Psychologist*, *58*(9), 697–720.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454.

Kahneman, D., & Tversky, A. (1973). On the Psychology of Prediction. *Psychological review*, *80*(4), 237-251.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: translating strategy into action*: Harvard Business Press.

Kovalerchuk, B. (2001). *Visualization and decision-making using structural information.* Paper presented at the Proceedings of International Conference on Imaging Science, Systems, and Technology (CISST'2001).

Krush, M. T., Agnihotri, R., Trainor, K. J., & Nowlin, E. L. (2013). Enhancing organizational sensemaking: An examination of the interactive effects of sales capabilities and marketing dashboards. *Industrial Marketing Management*, *42*(5), 824-835.

Lee, S., Kim, S. H., & Kwon, B. C. (2017a). Supplemental Material for IEEE InfoVis 2016 Paper. VLAT: Development of a Visualization Literacy Assessment Test. Retrieved from: https://www.bckwon.com/pdf/vlat_sup.pdf

Lee, S., Kim, S. H., & Kwon, B. C. (2017b). VLAT: Development of a Visualization Literacy Assessment Test. *IEEE transactions on visualization and computer graphics, 23*(1), 551-560.

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. *Journal of Marketing*, *71*(1), 160-177.

MacGregor, D., & Slovic, P. (1986). Graphic representation of judgmental information. *Human-Computer Interaction*, *2*(3), 179-200.

Maheshwari, D., & Janssen, M. (2014). *Dashboards for supporting organizational development: principles for the design and development of public sector performance dashboards.* Paper presented at the Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance.

Mayer, J. (2000). Performance with tables and graphs: effects of training and a Visual Search Model. *Ergonomics, 43*(11), 1840-1865.

McKenna, S., Staheli, D., Fulcher, C., & Meyer, M. (2016). *Bubblenet: A cyber security dashboard for visualizing patterns.* Paper presented at the Computer Graphics Forum.

Meyer, J., Shamo, M. K., & Gopher, D. (1999). Information structure and the relative efficacy of tables and graphs. *Human Factors, 41*(4), 570-587.

Morewedge, C., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences, 14*(10), 435-440.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, *35*(2), 136-164.

Negash, S. (2004). Business intelligence. *The communications of the Association for Information Systems, 13*(1), 54.

Negash, S., & Gray, P. (2008). Business Intelligence. In *Handbook on Decision Support Systems 2: Variations* (pp. 175-193). Berlin, Heidelberg: Springer Berlin Heidelberg.

Nobel Foundation. (n.d.). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2002. Retrieved from: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/

Pallant, J. (2013). *SPSS SURVIVAL MANUAL - A step by step guide to data anlysis using IBM SPSS*. Maidenhead: Open University Press/McGraw-Hill.

Pappas, L., & Whitman, L. (2011). *Riding the technology wave: Effective dashboard data visualization.* Paper presented at the Symposium on Human Interface.

Phillips-Wren, G. E., Iyer, L. S., Kulkarni, U. R., & Ariyachandra, T. (2015). Business Analytics in the Context of Big Data: A Roadmap for Research. *CAIS, 37*, 23.

Power, D. J. (2008). Decision support systems: a historical overview. In *Handbook on Decision Support Systems 1* (pp. 121-140): Springer.

Random.org. (2018). RANDOM.ORG - Interger Generator. Retrieved from: https://www.random.org/integers/

Raschke, R. L., & Steinbart, P. J. (2008). Mitigating the effects of misleading graphs on decisions by educating users about the principles of graph design. *Journal of Information Systems, 22*(2), 23-52.

Recker, J. (2012). Scientific Research in Information Systems. In *Scientific Research in Information Systems* (pp. 162): Springer.

Roumbanis, L. (2017). Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups. *Social studies of science, 47*(1), 95-116.

Royal Swedish Academy of Sciences. (2002). Daniel Kahneman. [*In Swedish*] Retrieved from: https://www.kva.se/sv/priser/pristagare/daniel-kahneman-2

Scheaffer, R. L., Mendenhall III, W., & Ott, R. L. (1971). *Elementary Survey Sampling* (6th ed.). Belmont, California: Thornson Higher Education.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics, 16*(6), 1139-1148.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Research Methods in Psychology* (9th ed.). New York: McGraw-Hill.

Simon, H. A. (1960). *The New Science of Management Decision*. New York: Harper & Brothers Publishers.

Skorka, A. (2017). Successful Dashboard Implementation in Practice: How to Overcome Implementation Barriers and Ensure Long-term Sustainability. *International Journal of Market Research, 59*(2), 239-262.

Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies, 64*(11), 1115-1131.

Swedish National Agency for Education. (2018). Geography. [*In Swedish*] Retrieved from: https://www.skolverket.se/laroplaner-amnen-och-kurser/grundskoleutbildning/grundskola/geografi

Tegarden, D. P. (1999). Business information visualization. *Communications of the AIS, 1*(1es), 4.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tversky, A. (1977). Features of similarity. *Psychological review, 84*(4), 327-352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice *science, 211*(30), 453-457.

Van Der Heijden, H. (2013). Evaluating dual performance measures on information dashboards: Effects of anchoring and presentation format. *Journal of Information Systems*, *27*(2), 21-34.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219-240.

Vierck, R. K. (1981). Decision support systems: an MIS manager's perspective. *MIS quarterly*, 35-48.

Vogt, W. P., & Johnson, B. (2011). *Dictionary of statistics & methodology : a nontechnical guide for the social sciences*. Thousand Oaks, Calif.: SAGE.

Ware, C. (2013). *Information Visualization: Perception For Design* (3rd ed.). USA: Morgan Kaufman & Elsevier.

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer, 40*(9).

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems, 13*, 41-59.

Zhang, P., & Whinston, A. B. (1995). Business information visualization for decision-making support-a research strategy.