UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTERS THESIS

INCREASING THE QUALITY OF A DASHBOARD BY FOLLOWING THE COLOR COMPOSITION THEORY

Ljubljana, September 2022

TINA LEKŠE

AUTHORSHIP STATEMENT

The undersigned Tina Lekše, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Increasing the Quality of a Dashboard by Following the Color Composition Theory, prepared under supervision of Jurij Jaklič,

D E C L A R E

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, September 27th, 2022 Author's signature: _________________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

sl. – Slovene

BI – Business intelligence, sl. Poslovna inteligenca

DQ – Data quality, sl. Kvaliteta podatkov

ELT – Extract, load, transform, sl. Zajem, nalaganje, preoblikovanje

ELTL - Extract, load, transform, load, sl. Zajem, nalaganje, preoblikovanje, nalaganje

ETL – Extract, transform, load, sl. Zajem, preoblikovanje, nalaganje

IBCS – International Business Communication Standards, sl. Mednarodni standardi poslovne komunikacije

ICT – Information and Communication Technology, sl. Informacijska in komunikacijska tehnologija

PaP – Preattentive Perception, sl. Pred-pozorno zaznavanje

SWD – Storytelling with Data, sl. Pripovedovanje zgodb s podatki

UX – User Experience, sl. Uporabniška izkušnja

INTRODUCTION

Data visualization is the key factor of business intelligence end analytics. It is important to be critical with information, but to understand the data in front of us, we need to use the correct visualizations. Designing the right visual representation can be a challenge, however the aftermath of using the wrong one can be a significant loss for a company (Lavalle, Maté, Trujillo, & Rizzi, 2019).

Business Intelligence (BI) is the combination of business analytics, data infrastructure, data mining, and data visualization. In short, it helps businesses make data-driven decisions based on their analytics. Data-driven decisions help reduce errors, eliminate inefficiencies, and open up opportunities that the company might otherwise have missed. BI originated in the 1960s as a system for exchanging information within companies. With the development of computer systems, BI transformed into a framework for making decisions and providing insights from available data. Today, BI's solutions can provide self-service analysis, which means that individuals can generate their own reports, and represent the difference between human-driven decisions and data-driven success - on condition that the data is analyzed correctly. BI is therefore a framework that provides a comprehensive view of an organization to help them make better decisions (Tableau Software, LLC., 2021).

BI contains different sets of processes and activities that help an organization improve its operations. These processes range from data mining, reporting, performance benchmarking, descriptive analytics, querying, and statistical analysis to data visualization, visual analysis, and data preparation (Ranawaka, 2021).

Most business decisions today can only be considered trustworthy if they are backed up by data. The main problem with this is that we first need to understand our data if we want to gain insights from it. This is where data visualization tools come in. Data visualization tools help business leaders make data-driven decisions, and those decisions can only be made if the data is properly presented. There are a number of visualization tools to support data visualization (Adair, 2020).

Dashboards are a visualization tool that combines multiple types of visuals, like tables and charts, ideally providing a 360-degree view of the phenomenon being analyzed. To design a dashboard, users should specify their goals and precisely delineate the information that is to be displayed.

A visualization goal can be a composition, a sequence, a relationship, a comparison, a cluster, a distribution, a trend, or a spatial representation (Lavalle, Maté, Trujillo, & Rizzi, 2019). There are three basic principles: show the data, reduce clutter, and integrate the text and the chart. Data visualization has a form and a function. With regard to form, static visualizations provide all information at once and are not active or moving. Interactive visualizations allow information to be transferred between the figure and the user.

Explanatory visualizations, on the other hand, put the main results in the foreground - the most important findings. Finally, exploratory visualizations help the user to independently study the data set with a data set to discover the results for themselves (Schwabish, 2014).

Color theory is both the science and the art of using colors. It explains how people perceive colors and the visual effects of mixing, matching, or contrasting colors. Additionally, this term covers the study of what messages colors convey and which methods are used to reproduce colors. In color theory, colors are arranged on a color wheel and divided into 3 categories: primary colors, secondary colors, and tertiary colors (Decker, 2016).

Next to color perception, the human brain is exceptionally good at filling in the blanks in an image and creating a whole that is greater than the sum of its parts. That is, for example, why we see faces in things like tree leaves or cracks in the pavement. While the brain can connect, differentiate, or see things that are not really there to make understanding your surroundings easier, all those ideas are the underlying notions of the Gestalt principles of visual perception. There are a number of iterations of the principles, depending on the context we use them in, and they are also used in data visualization, as they help the designer of the visual understand how the human brain will interact with the dashboard at hand (Chapman, 2017).

The problem regarding visualization or dashboard builds in modern business intelligence is that even though people try to choose the correct types of visualization according to their data, most fail to consider the color composition behind the entire project. If we use the wrong color, whitespace, or size, we can quickly loose the readability and intelligibility of the message, as well as risking the misinterpretation of the data at hand. In short, data is the foundation, but if we ignore the user experience (UX) and interface – which are directly connected to color and composition – the data visualizations could be interpreted incorrectly, which could lead to a loss of profit for the business. And while numbers are the basis of the decisions, the way in which we visualize them could lead to a sequence of underlying decisions, based only on the human perception of it (Burnay, Bouraga, & Faulkner, 2020; Mark, 2022).

The overall effect that visualizations have may be hard to measure, but the bottom line is that if business decisions are based upon the research and visualization of data, why is there no set of rules that can warn users of mistakes that may cost the company an extraordinary amount of money? If the users preparing the data visualization dashboard do not follow at least some design principles, the data provided will be harder to interpret or understand as it is meant to be, which will in turn make it harder to decide on the correct course of action. Even if the color and visualization type are chosen correctly, the dashboard design may be the main contributing factor which will take the focus away from the important data findings, and consequentially lead to rash decisions based on the wrong visual. This is why the incorporation of color theory, design theories, and the choosing of the correct visualization may all have a large effect on the final data interpretation.

The purpose of this master's thesis is to show the influence of color theory and composition when working on visualizations and dashboards in business intelligence development. Based on the literature review and the results of the empirical analysis, the purpose is to understand how color and composition influence the understanding of visualizations by individuals and use that understanding to define the guidelines. If we do not follow a set of guidelines, every model we build will have the same issues and obstacles with readability. The end result of this thesis is to define the guidelines and explain the effect they have on our comprehension of the material. The purpose is therefore to help define the creation of visualization dashboards, while keeping in mind their readability and ease of interpretation. All of this will have a great impact on the ability of companies to make informed decisions.

In my master's thesis, I will try to answer the following research questions:

- How does color affect the interpretation of visualization?
- What is the role of color and composition in BI development?

The goal of this master's thesis is primarily to use the critical literature review to understand the concepts behind color theory, dashboards, visualizations, and composition. The rules and guidelines that already exist will be gathered, then they may be applied to a data visualization model. The goal of the empirical part is to use interviews and pre-made dashboards to understand how individuals' precept the quality and information in different dashboards. The end goal is to combine the critical literature review and methodological part, in order to deliver a set of guidelines to follow, when we are building a dashboard in BI.

The master's thesis will be based upon the theoretical part, which includes a critical literature review. Scientific sources will be dissected and analyzed to uncover how human perception is connected to color theory and composition.

Based on the literature review provided, the methodology will be as follows. First, I will use the design principles mentioned beforehand and create examples of good and bad visualizations in Microsoft Power BI. I will focus on using the same data set while creating a dashboard that will give the user as much information while presenting it clearly, while also creating a second dashboard, which will give either wrong information or the visualizations will be unclear for interpretation. I will do an experiment where users will have both dashboards presented and will give their feedback on what they learned from each of them. Lastly, I will do a comparison of the feedback provided, and gather what users find good or bad in the dashboards presented.

Based on the results of the methodology part, I will try to formulate guidelines, which can be used to design dashboards while maximizing the UX by following the color composition theory.

1 BUSINESS INTELLIGENCE

Business Intelligence (BI) helps companies and users uncover insights, which are in turn the base of their strategic decisions. In short, BI helps companies make data-driven decisions based on their analysis. Such a system helps reduce mistakes, eliminate inefficiencies, and open opportunities, which the company might have otherwise missed (Microsoft, n.d.).

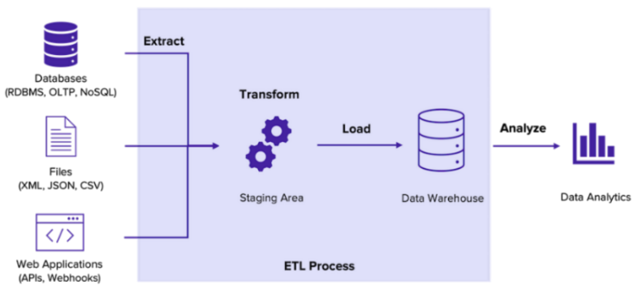
BI emerged around the 1960s as a system of sharing information across organizations. Alongside the development of computer systems, BI transitioned to decision-making and providing insights from the data available. Today, BI solutions provide self-service analysis, which means that individuals can generate their own reports, and represent the difference between human-driven decisions and data-driven success - on condition that the data is analyzed correctly (Tableau Software, LLC., 2021).

1.1 Business Intelligence Processes

Business intelligence contains different sets of processes and activities which help an organization improve their operations. These processes range from data mining, reporting, performance benchmarking, descriptive analytics, querying, and statistical analysis, to data visualization, visual analysis, and data preparation.

The majority of BI functionality could be described through an ETL process, which is pictured in figure 1 below. ETL stands for extract, transform, load, and has been around for years. The process describes the data flow in a format, that stays the same no matter the technology used. With the introduction of cloud data warehouses, the process is being modified into ELT (extract, load, transform) as a newer approach for combining data, which is sometimes also called ELTL (extract, load, transform, load). No matter which process is used, however the three key steps stay the same. In the extract phase, we gather and transfer all your data from various data sources. In the transformation phase, we clean and process the data, so it is suitable for the needs of our data storage. Lastly, the load stage is where we load all the data we have transformed and deemed usable into our storage destination, which will later be the data source for our business intelligence tool (Ranawaka, 2021).

Figure 1: ETL Process



Source: Ranawaka (2021).

1.2 BI Data Visualization Tools

In today's world, there is a high volume of data being generated daily, and most of it is stored and archived unless the data gets visualized. In such cases, the business decisions of the company then stem from the inputs they get out of dashboards. The main problem behind gaining insights from our data, is that we have to understand it. Data visualization is the visual presentation of data to convey information to the end user. Data visualization tools help company leaders make data-driven decisions and that can only be made if the data is shown correctly. To help with data visualization, there are a number of BI tools available. In general, BI tools are used for ETL processing, transformations, and analysis of data, which in turn help make business decisions. Within this thesis, the BI tools are referred to as visualization tools, as it is the intent of the thesis to compare different dashboards created with a BI tool. One of the most well-known and used BI tools is Microsoft's Power BI (Adair, 2020; Brush, 2020).

Power BI is a business analytics service by Microsoft. As any other BI visualization tool, Power BI is the connection between data and the end user. It can be used for both data modelling as well as building dashboards and showing the said data. One of the reasons for its wide usage is that the basic version is free of charge and provides a strong entry point into the world of data visualization.

Power BI is used to find insights within an organization's data. The data can come from various points, that can either be pre-loaded or directly queried to the servers where they host. It can also connect to smaller data sets like csv files, while being updated with cloud connectors frequently. Next to the data connectors, Power BI gets continuous updates of its interface and its data visualization options palette, with the option of importing community made custom visuals. The reports generated can then be shared with other Power BI users within the organization.

The data models created from Power BI can be used in numerous ways by companies, including to tell stories through charts and data visualizations and to examine "what if" scenarios within the data. Power BI reports can also be used for real-time analysis of the data, prediction, and a number of other ways to improve data-driven business value (Scardina, 2018).

There is a number of effective BI tools to choose from, for example Tableau, Pyramid, and QlikView (Adair, 2020; Pyramid Analytics, 2021).

1.3 Dashboards

Business intelligence dashboards are visualization tools that that manage sets of data by grouping them on a single screen. Dashboards neatly present data analysis results while allowing the creator to select which pieces of data they want to show the end user. The main goal when creating a dashboard is for it to be purpose-built while being the main user interaction pillar when making business decisions (Tableau, 2021).

A dashboard is therefore a combination of selected visualizations that present different pieces of data, that show important pieces of information as a big picture. It is the stakeholders most important tool in decision-making, and it is the next step in data analysis, because it diminishes the need for several smaller spreadsheets.

According to Few (2006), the main differentiation between dashboard types stems from their role-oriented categorization. A dashboard can either be strategic, analytical, or operational. The most commonly known ones are strategic, which cover the well-known "executive dashboard". The strategic dashboard is primarily focused on covering a bigger picture of the data, so that the executive level users can quickly get insights and make business decisions based on them. The analytical dashboard presents itself as a tool for further data analysis and not only as a general overview. These types of dashboards usually have more complex visualizations, a larger need for comparisons, and a broader context of use. Lastly, the operational dashboard is similar to the strategic one in terms of simple visualizations, but their use is needed for real time data analysis, for example the health monitoring of machines or operations. Errors shown on the operational dashboard will usually be corrected as fast as possible by other employees (Few, 2006).

1.4 Data Visualization

Data visualization is the presentation of data in a graphical format, out of which the users gather information. It is a way of visually communicating data to the end users, keeping in mind that you must convey as much data as possible in a limited space (Makulec, 2022). Data visualization is a process of taking raw data and transforming it into graphical representations, and the visualization itself can either be a chart, graph, diagram, picture, or even video (Runkler, 2012). Different visualization tools offer users different types of visualizations to choose from, but compiled below is a list of the major types:

- Area chart
- Line chart
- Bar chart
- Column chart
- Simple text
- Pie chart
- Doughnut chart
- Gauge chart
- Funnel chart
- Scatter chart
- Waterfall chart
- Maps
- Slicer chart
- Tree map
- Matrix chart
- Table

Different types of visualizations are usually used for representing different types of data. For example, if you are trying to show the relationship of parts to a whole, people tend to choose pie charts; if you are showing a process through stages you use a funnel chart. Tables work best when considering individual values and only in cases where precise data is available, and the data is precise. Graphs on the other hand feature patterns, trends, and a large sequence of values at once. Each one has its advantages and disadvantages, but you should always be aware of what type of data you have available, and what business questions you are trying to answer through the visualization (Microsoft, 2021; Few, 2011; Rogowitz, Treinish, & Bryson, 1996).

Simple text

When it comes to showing one number, simple text is commonly the most effective way to share data. The numbers in any type of visualization can always be paired with text for further explanation, if the number itself is enough for the user to understand the situation, refrain from using unnecessary graphs.

Tables

Tables as a visual are a great way to communicate with a variety of users, as each one can look at the rows and columns that correspond the best to what they are searching for. However, what we have to keep in mind is that the data is the centre of the visualization, and not the table itself.

Graphs

Where tables communicate their contents with text and numbers, graphs do the same with color and form. The human visual system is much faster at processing such information over raw numbers. Graphs can, therefore, speed up how quickly we process information, as long as they are designed correctly.

Bars

Bar graphs are one of the most commonly used types of visualizations and are consequently sometimes perceived as “boring” and not visually appealing. Because of their popularity, bar charts are also the easiest to interpret and understand, as the logic behind them is something users meet almost daily. The main problem with bar charts, however, is that people tend to misuse them or set the context up wrong, as the most important rule is that bar charts always have a zero baseline, making comparison fair and easily understandable.

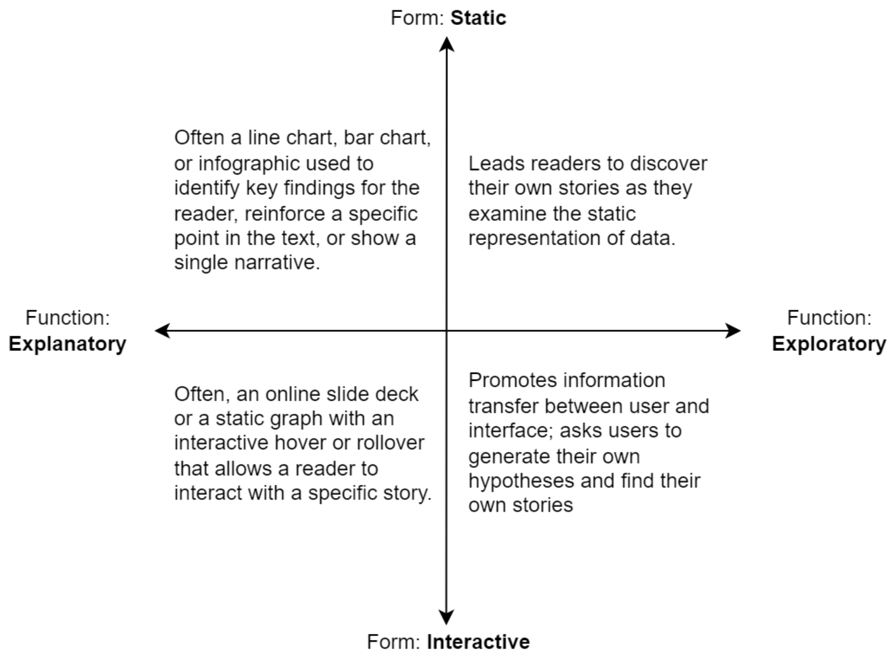
Pie charts and donut charts

Pie and donut charts have a quality that other graphs do not: they ascribe quantitative value to two-dimensional space. When the segments of both charts are similar in size, they are hard to distinguish, and if the user decides to include 3D effects, it makes the graphs entirely unreadable. Both charts are used for showing percentages of a whole, but in most cases that can be done better with either a graph or a bar chart.

There is a number of charts and graphs users can choose from to present the data they have, yet the only important factor when choosing the visualization type is whichever will be the simplest for the audience to comprehend (Nussbaumer Knaflic, 2015).

There are three basic principles for visualization: show the data, reduce the clutter, and integrate the text and the graph. Data visualization has form and function, as depicted in the graph below. With regard to form, static visualizations provide all information at once and are not active or moving. Interactive visualizations allow information to be transferred between the figure and the user. Explanatory visualizations, on the other hand, put the main results in the foreground - the most important findings. Finally, exploratory visualizations help users interact with a data set to uncover the findings themselves. All of this is pictured in figure 2.

Figure 2: Form and Function of a Visualization



Source: Schwabish (2014).

For BI analysts who want readers to apprehend results quickly and accurately, presentation matters. Effective visualizations show the data to tell the story, reduce clutter to keep the focus on the important points, and integrate the text with the graphs to transfer information efficiently (Schwabish, 2014; Miglietti, 2022).

Most users in BI are fairly competent when it comes to reading or understanding what different numbers and key performance indicators mean. They have been taught math and language for their entire lives. The issue arises when it comes to the combination of both. The entire concept of visualization revolves around storytelling with the data at hand. This is where data visualization techniques that involve the understanding of color, layout, and context come to be so important (Nussbaumer Knaflic, 2015).

1.5 Effective Communication through Chart Elements

For easier understanding of chart parts and how they can help make your visual more effective in communicating, Dougherty deconstructs a chart. The deconstruction is a helpful tool when trying to understand parts of a whole, which can later on help us understand how the entire visual translates data to information.

In the context of a line chart, Dougherty explains that a title is the most important element of any chart, as it presents a very short text description of what type of information we are going to be getting from the visualization. If the headline is effective, the reader will immediately understand the point we are trying to convey, yet we can also utilize a subtitle, which can explain a more technical side of the chart.

As well as the title, labels and annotations on a chart are used to add context to the data series presented. This, combined with the labels of the x and y axis, can provide an even deeper understanding of the data setting for the user, while omitting the need for additional visuals for explanation. If this is not enough, or could be accompanied with additional notes, data sources, or credits, for the data to appear more credible, you can also add those for context of where the data came from and its process or analysis.

In BI tools, there is also the option of a tooltip, which can be used for providing additional insights into the data, as the tooltip values change according to the setting, we are viewing at the moment (Dougherty & Ilyankou, 2021).

2 VISUALIZATION PRINCIPLES

Data visualization can be explained as a two-way street; both a tool for analysis and one for communication. If we are focusing on the second part, it is a tool that presents data, with the main goal of grabbing attention and helping engagement. The intention of any visualization is therefore to increase the understanding of data of the user, irrespective of the technical abilities of the author (Evergreen & Metzner, 2013).

2.1 Importance of Visualization

With the rise in data collection and generation, the need to correctly understanding the data has been rising in recent years. The necessity to interpret complex and voluminous data stems from the need to make effective strategic and business decisions. Big data is processed and loaded using various methods, and in the end, presented through various types of approaches and techniques. Data visualization takes the enormous sets of data that are available and unintelligible, and converts it into something easy to comprehend - converts it into knowledge. In today’s world, however, many scientists or users in general do not realize just how much effective data visualization affects data comprehension. In recent years, however, the number of those who understand the importance of the mastery of data visualization is on the rise (Anouncia, Gohel, & Vairamuthu, 2020; Midway, 2020).

Data visualization uses visual information presentation to share information quickly and effectively with all audiences. As well as presenting information, we achieve the ability to interpret the understand the data faster and easier, which leads to faster decision making. Data visualization also increases the speed of understanding the next steps and the speed at which we can act, and in turn achieve success (Brush, 2020).

2.2 Design Theories

2.2.1 Color Theory

Color theory is both the science and art of using color. It explains how humans perceive color; and the visual effects of how colors mix, match, or contrast with each other (Robinson, 2020). Color theory also involves the messages colors communicate; and the methods used to replicate color. In color theory, colors are organized on a color wheel and grouped into 3 categories: primary colors, secondary colors and tertiary colors (Decker, 2016). The theory is the study of color from both scientific and subjective perspectives to understand both how it influences human perception and how it can be utilized to influence better decision-making in communication and design (RevUnit, 2022).

The elements that can influence our visualization building but also human perception the most will be explored in detail in the following paragraphs.

2.2.1.1 Color Anatomy

The anatomy of color can be explained as the aspects making up color, as well as how humans perceive it. Color consists of what aspects make up color, and how humans view it. It consists of hue, saturation, and lightness value (also known as luminance or luminosity), as depicted in figure 3.

Figure 3: Color Anatomy

Source: RevUnit (2022).

Most people understand color as hue, as it is what differentiates how it will be perceived and processed. Saturation refers to the brightness of a certain hue, from the darkest to the lightest variation that a certain hue occupies. Most visualizations should use a lower saturation (duller, whiter colors), as that is what makes important data stand out in higher saturation. Lastly, lightness or luminance is the use of shades or light – variances of whites and blacks.

2.2.1.2 Color Harmony

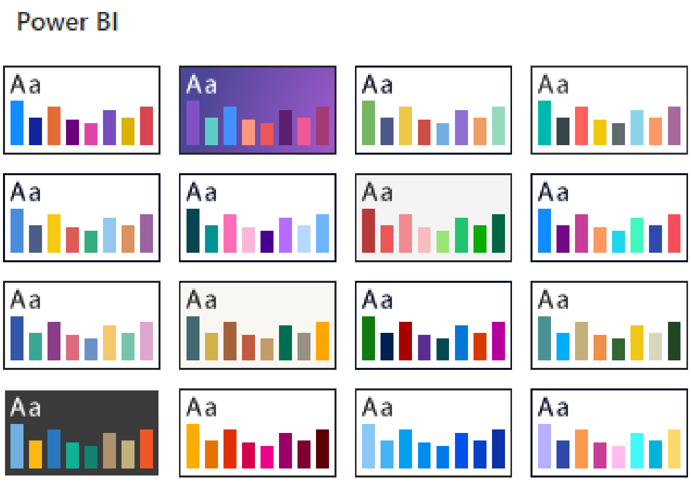
This principle in color theory implies that certain colors (or hues) can be used in a combination that will either produce visual dissimilarity, or cohesion. When we apply this to visualization, we should always keep this principle in mind as the basis for generating our color palette.

When it comes to color harmony in data visualization, we usually select a main color, either based on the data type or the company primary colors and build a color palette on top of that. In the palette selection process, we choose the colors that are well-matched with the primary color tone (Kargin, 2022).

2.2.1.3 Color Schemes

There is a variety of color schemes users can choose, and this also corresponds to the color palette we can chose in most BI tools, as it will be predefined. The Power BI predefined themes (or color schemes), which can be seen in figure 4, range from having different hues and backgrounds, to only differing in in saturation.

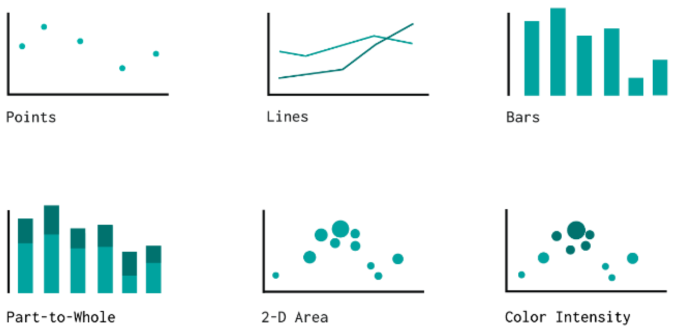
Figure 4: Power BI Predefined Themes



Source: Microsoft (2021).

We can split the schemes into roughly three categories: monochromatic, analogous, and complementary. In practice, the monochromatic palette in figure 5 works best in scenarios where we are trying to emphasize that the data varies only in its degree but not its kind.

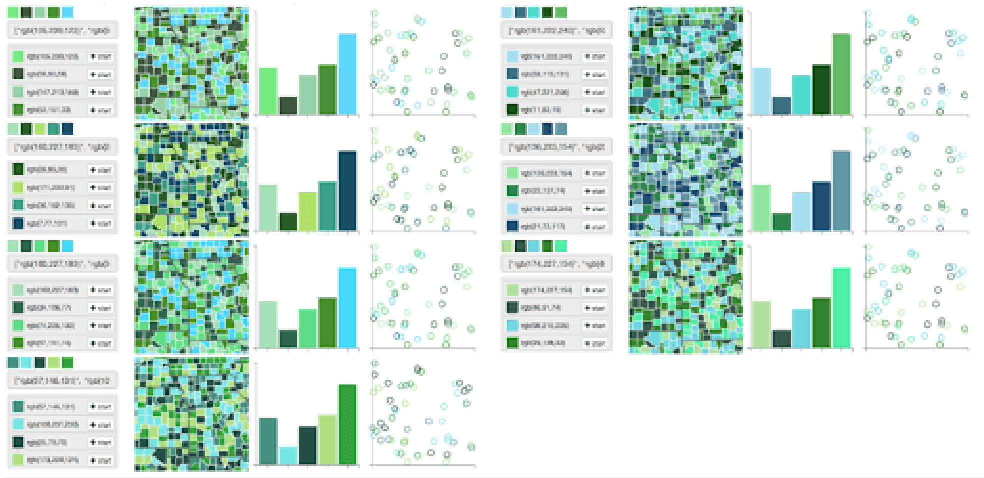
Figure 5: Monochromatic Palette



Source: RevUnit (2022).

The analogous pairings, pictured in figure 6, are colors that are positioned next to each other on the color wheel. In this case, people innately understand that the data differs, but it is closely related, because of the color similarity.

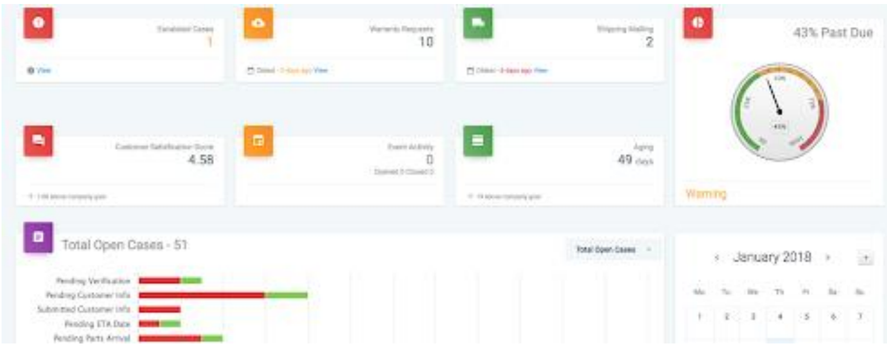
Figure 6: Analogous Pairings



Source: RevUnit (2022).

Lastly, the complementary color pairings, pictured in figure 7, are colors on different sides of the color wheel, which will conversely represent the maximum differentiation possible. This is closely related to colors which we understand mean earnings and losses or losses, or in general, a positive or negative impact.

Figure 7: Complementary Color Pairings



Source: RevUnit (2022).

2.2.1.4 Color Accessibility

When it comes to different people representing our user base, we can soon conclude that color blindness affects more people than most dashboard developers take into consideration. In raw numbers, an estimated of 300 million people in the world have color vision deficiency – that is 1 in 12 men, and 1 in 200 women (Clinton Eye Assoiates, n.d.). The most commonly known color blindness is red-green (seen in figure 8 below) or blue-yellow color blindness, which are also the most used colors in dashboards, with green and red commonly representing positive or negative values.

Figure 8: Color Accessibility

Source: RevUnit (2022).

2.2.1.5 Consistency

The last part is something that most BI tools do automatically, but depending on the data, values can sometimes become mixed up. Being consistent with one color always having only the one value associated with it is associated with them is very important for maximizing user understanding. Because humans tend to group objects together based on color, having multiple object colored red will translate as one same group, even if that is indicated otherwise with labels or a color legend (RevUnit, 2022).

2.2.1.6 Choosing Color for Data Visualization

Color is one of the key elements when it comes to data visualization, as it is most often the tool through which we highlight the story we are trying to tell, as well as being able to distract from or hide a visual purpose.

In the context of data visualization, we usually employ one of three major types of color palette: qualitative, seen in figure 9, employ three major types of color palettes – qualitative, sequential, and diverging. A qualitative palette is used when the variables we want to portray are categorical – that is to say they have a distinct label without ordering, for example gender or race. Having a qualitative palette lets us easily differentiate the values between each other, as they are different in hue (Tackels, n.d.; Yi, 2019; Bruner, Postman, & Rodrigues, 1951).

In contrast, having numerical values or values that have an order, we can utilize the sequential palette. In this case we assign the values color that varies in lightness or saturation. Using this type of palette, we usually associate lighter colors with lower values, while higher ones have a darker color and a bigger impact (Kashyap, 2020).

Figure 9: Qualitative Color Palette

Source: Nightingale (2019)

When the value at hand has a central value – for example a zero, we can use the diverging palette. A simple example would be a Likert scale, where people chose their answer on a scale from strongly agree to strongly disagree, with neutral being the middle value (Yi, 2019).

An important part of using color in data visualization is how we use color, and as well as how much of it we use. We need to use color to create associations, just as most things that surround us in our daily life do – we use color to trigger an association, with green meaning good, and red automatically being linked to bad. Good practice also suggests that we should use a single color to show continuous data and contrasting colors to show comparison. Nevertheless, if there is too much color – or not enough – the importance of information can quickly be lost. Therefore, we must also use color to make information stand out (Tackels, n.d.).

2.2.2 Color Correction in Data Visualization

In Berinato's book, he states if the designer of a visualization has the time to focus and improve on only one thing, they should focus on color. Afterwards, he defined a set of rules, which provide a theoretical framework to reflect on when designing visuals. Firstly, using less colors - hues is what he believes improves visuals the most. Because this limits our color palette, we should try to group data together with the same color, but we should also do so in a smart way, which adds meaning. Second, he urges to use gray, as it provides less of a contrast to a white background as black does. This translates gray elements to being background information instead of a main focus point, which does not draw the eyes away from the data in the same way that stronger colors may. Next, Berinato states that we should be aware of complement or contrast. When the data we are showing is similar, we can employ complementary colors, or the same colors with a lightness change between them. However, when presenting opposing information, we should use contrasting colors. Fourthly, he focuses on labels and text. While using colors in labels could provide an important insight in some cases, we should use that only when necessary, and when we believe it will provide crucial information. While useful, coloring text can distract users instead of being an aid for understanding. Lastly, he states that when building visuals, we should always think about how, not which. This means that when building a visual, instead of using our time to select which colors we should use, we should focus more on how we are going to use the colors in a smart way. Using color saturation and lightness can create variation and draw the end-users focus, while still not overwhelming them with color information. He also emphasizes that the visualization designers should use colors that will improve data retention and draw focus instead of only focusing on colors the company wants to use (Berinato, 2019).

2.2.3 Edward Tufte's Design Principles

In 1982, an American statistician Edward Tufte published a book called *The Visual Display of Quantitative Information*. The contents of the work were a leap in the right direction, as he gives examples of dismal statistical graphics throughout the entire book. But he also set in place a number of principles which are the basis of good visualization even today. He lists a set of reasons that explain the principles of inferior graphical work.

1. Lack of Quantitative Skills of Professional Artists: This refers to artists having next to zero experience with statistics, which eventually leads to "beautifying of data" which compromises statistical integrity.
2. The Doctrine that Statistical Data is Boring: The second doctrine refers to designers assuming that because statistics are "boring", they should use more decoration and make the evidence more convincing by inflating values.
3. The Doctrine that Graphics are Only for the Unsophisticated Readers: The last doctrine refers to artists who strongly believe that using text is too complicated for the general public to comprehend, which eventually also leads to unnecessarily beautification of graphs to interest the readers.

The two other equally important formulae that Tufte also originated were "the chart lies" and "the chart junk", or in other words, the lie factor, and the data-ink ratio. The lie factor can be calculated as the ratio between the size of effect shown in a graphic and the size of effect in data, and therefore translates as the graphical ability to represent the data (or in our case, either exaggerate or underrate). The data-ink ratio, on the other hand, tells us the non-redundant information in a chart. The higher the ratio, the less “useless” data is shown on screen. It is the ratio between non-erasable core ink of a graphic and the total ink used in the graphic (Sigdel, 2020; Tufte, 2001).

In short, he set the six well-known principles of graphical integrity:

1. The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented.
2. Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.
3. Show data variation, not design variation.
4. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
5. The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data. Graphics must not quote data out of context.
6. Graphics must not quote data out of context (Ferrara, 2017)

2.2.4 Using Color in Charts

In his 2008 article, Few defined the practical rules for using color in charts. He states that color can be used in powerful ways to enhance data displays, but only when one understands how to use it correctly. He sets down six practical rules for the use of color. Firstly, he emphasises the use of color in context. We perceive everything in a relative way, which means that all we see, including color, changes based on the surrounding context. This is usually shown with the known example of the small grey square, where we perceive the same shade of grey in four different ways because of the background changes, as seen in figure 10 below.

Figure 10: Color in Context

Source: Few (2008).

In practice this refers to the fact that the usage of color has to be consistent, and that the background color should not change if we want the user to have the same perception of data by color differentiation.

Secondly, Few urges readers to use color meaningfully and with restraint, which suggest that we should only use color when it is needed for a communication goal, and not as decoration. We should also keep in mind that we should use a spectrum of colors when using it for differentiation in data meaning. Speaking of the color spectrum, his third rule implies that projects should have a defined palette of colors for particular purposes. He emphasizes that most data should use soft and natural colors, while highlighted information should use brighter or darker color to stand out more. Another thing Few mentions is that non-data components (of for example tables and graphs) should be only visible enough to perform their role, but not to distract from the important data we are trying to show.

Few emphasizes that we should guarantee inclusion for color-blind people, so that they can distinguish the data groups as well. The most wide-spread type of color blindness in the green-red type, so that is usually the one we should try to avoid. Finally, we should never use "useless" visual effects in any type of visualization, for example shadows or light effects (Few, 2008).

2.2.5 Gestalt

The human brain is exceptionally good at filling in the blanks in an image and creating a whole that is greater than the sum of its parts. That is why we see faces in things like tree leaves or sidewalk cracks. The Gestalt principles of visual perception explain how our brains try to build structure around us by default, in order to help us understand out surroundings. One of the most influential works that influenced the Gestalt theory was published by Max Wertheimer in 1923. Even though there are a number of variations of how to apply the principles depending on the context, they have also evolved in regard to data visualization (Chapman, 2017; UserTesting, 2019).

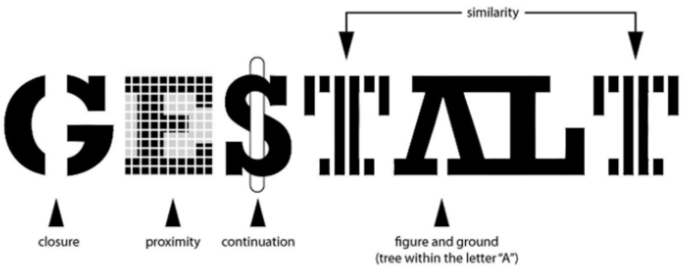
In short, the Gestalt principles will help any UX designer help determine which elements are the most effective in a given situation, how to drive attention to a specific point of focus, and how to – at the highest level – design products that solve the user's needs (Ahmadi, 2020; Bauer, 2022).

Listed below are seven Gestalt principles that apply to modern design:

1. Similarity – if we see similar things, we group them together.
2. Continuity - elements that are arranged on a line or curve are perceived to be more related than elements not on the line or curve.
3. Closure - when we look at a complex arrangement of visual elements, we tend to look for a single, recognizable pattern.
4. Common region - when objects are located within the same closed region, we perceive them as being grouped together.
5. Focal point - whatever stands out visually will capture and hold the viewer's attention first.

6. Proximity – when things are closer together, we tend to perceive them as related.
7. Symmetry (ger. *pragnanz*) - your brain will perceive ambiguous shapes in as simple a manner as possible.

Figure 11: Gestalt



Source: Ahmadi (2020).

2.3 Storytelling with Data

In 2015, Cole Nussbaumer Knaflic published a book called *Storytelling with Data* (SWD), which was based on her BI experience in visualization. SWD is split into steps which dashboard or visualization engineers should follow to maximize the results. Cole divided the process of SWD into 5 steps, which can be applied in any visualization building process in order to maximize the knowledge transfer while keeping the context of the user in mind.

The importance of context

Here, Cole focuses on the fact that we firstly have to learn who our audience is and what we need them to know. She emphasizes the importance of understanding situational context – the target audience and communication mechanisms. If we understand that clearly, problems in communicating information to them are reduced, while downsizing the number of iterations needed to communicate the data clearly.

Choosing an effective visual

SWD focuses on explaining which types of graphs and bars exist, and when to (or not to) use each one. The bottom line that is emphasized with all types, however, is that color or form should never compete with data for your attention.

Clutter as an enemy

When it comes to seeing, any piece of information on a blank page we see takes up cognitive load in the user brain. If we apply this when building dashboards or visualizations, we understand that every part of the page is something our user will have to focus on and clarify for themselves. Keeping this in mind, we need to de-clutter our visuals, so that the data transfer is as easy and as fast as possible.

Focusing the audience’s attention

Here, Cole focuses on how people see and how we can use that to our advantage when we try to explain the data. She also covers color as a strategic tool and the basics of gestalt principles.

Thinking like a designer

The concept of ‘form follows function’ that sprung up in the 19th century is unavoidable in architecture as well as in any sort of visual communication – even in data visualization. Cole emphasizes that we firstly have to think about the function of the data (what we want people to be able to do with the data) in order to build the form (visualizations). We also have to keep in mind the accessibility of the visualizations for people with disabilities, the most common being color blindness (Nussbaumer Knaflic, 2015).

2.4 Perception of Color

The link between color and emotion is very common. The most frequent one may be that the color red is linked to anger, no matter where the person is from or how they were raised. In every language, there are idioms connected to colors and emotion, for example “being green with envy” for being jealous or “having the blues” for being sad or depressed (English Club, n.d.).

Through a survey and ongoing research, they released the results seen in figure 5 in 2020, which show the proportion of emotions selected for each color term and their average intensities.

Figure 12: Most Frequent Color-Emotion Associations in 30 Countries

Color term	Associated emotions	Average %	Color term	Associated emotions	Average %
Red	Love	68%	Pink	Love	50%
	Anger	51%		Joy	41%
	Pleasure	33%		Pleasure	40%
	Hate	29%		Amusement	36%
Orange	Joy	44%	Purple	Pleasure	25%
	Amusement	42%		Interest	24%
	Pleasure	33%		Pride	24%
Yellow	Joy	52%		Admiration	24%
	Amusement	40%	White	Relief	43%
	Pleasure	32%		Contentment	30%
Green	Contentment	39%	Grey	Sadness	48%
	Joy	34%		Disappointment	41%
	Pleasure	34%		Regret	31%
	Relief	33%	Brown	Disgust	36%
	Interest	31%	Black	Sadness	51%
Blue	Relief	35%		Fear	48%
	Contentment	34%		Hate	41%
	Interest	27%		Anger	32%
Turquoise	Pleasure	35%		Guilt	30%
	Relief	34%			
	Joy	32%			
	Contentment	31%			

Source: Mohr (2018).

As stated in an article written by Mohr and Jonauskaite, all participants found it easy to establish a link between color and emotion. It depended on the colors chosen, but the simplest to connect were red, black, and yellow, while brown and purple were a tad harder to link. An important realization is that the color to emotion correlation is not a one-to-one relationship but many-to-many, as users selected multiple emotions for one color, and in turn, several colors were linked to one emotion. One of the other observations that were made is that mostly positive emotions were connected, while only dark colors were associated with negative ones. The only exception being the color red, which has the emotional variance of anger and hate to love and passion (Mohr & Jonauskaite, 2022).

Color, as defined above, plays a major role in the analysis and communication of information. As most BI software is equipped with a default color scheme, we have a set of graphs and charts that utilize the exact same colors, yet the same color map applied to numerous sets of data can quickly get monotonous and confusing. In the modern world, we usually employ two color spaces - the absolute (defining color in the terms of human perception) and non-absolute (defining color based on an input/output device – for example RGB or CMYK). It was also established in the late 20^{th} century that the human eye displays greater sensitivity to luminance differences than to hue (light versus color). Therefore, the human eye can get stimulus from different color spaces, different luminance variations, and most importantly different color maps, but dashboard creators expect that the end user will understand the visualization either way (Zeller & Rogers, 2020).

Colors in general have the ability to aid our communication in visualization or distract from it. Our brain reacts to color by a process called preattentive processing (PaP). It is a process of gaining visual input to the retina, and processing the information such as color, shape, and orientation. As a result, the user is drawn to the variety of shape and color that constructs the world around us. In visualization practice, we can use this knowledge with using a color palette that will shift the user's focus to what the important parts of the data are and let them skip over non-data objects (for example axis colors) (RevUnit, 2022). Next to color and perception theories, there is also a number of subjective dimensions we can implement into a visualization that can change its comprehensiveness and effect. These dimensions are aesthetics, style, playfulness, and vividness, which cover attractiveness, brand, experimentation, and memorability (Joshi, 2018).

2.5 Data Visualization Rules

In any learning path, there is a certain amount of time people will use in order to get better at a certain skill. In data visualization, there is a similar learning curve, but as people can create visuals without much thought, most do not try to find a set of rules to follow when doing so. The black box of data visualization could be omitted by following a set of rules in data visualization. According to Muth (2020), the rules set by Tufte and others, are a helping hand to understanding how people can build upon their skillset and quality when it comes to data visualization. She continues by saying that where there is a set of rules, it becomes a framework which you can discuss and debug, but most of all, you can eliminate bias. While rules do eliminate uniqueness and creativeness, they are also a great supporting tool when starting off or creating basic visuals in any BI tool.

One of the main things to focus on is the lack of experience in firstcomers to visualization, which can result in misuse of visual types. As there is a difference in using a pie chart and a bar chart, there is also a difference between a stacked bar chart and a bar chart as well, and the use case behind them changes according to the data we are trying to show. However, one of the things that Muth (2020) insists on, is that "if a rule prevents you from improving your data visualization, ignore it". However, this only comes into play when you already know your way around data visualization. If you have not been creating dashboards for a while, you also might not know what improvement means, and how to implement it. One of the rules that are highlighted is that data visualizations should be self-evident. This means that when someone looks at them, they should see your point themselves, without the need of additional verbal explanation. This is where additional BI helpers like annotations, labels, and clear text come into play. It is also where it will be the most obvious if the visual chosen to represent data is a correct choice.

As a bottom line, Muth's (2020) debate expresses those rules can benefit everyone that designs data visuals, but at the same time, they can also have a downside, where they block creativity and innovation. Secondly, the same rules that have materialized from Tufte's, Few's, or any other theories of data visualization, tend to have at least some part of contradictions in them. Lastly, she defines the quality of a data visualization to be subjective, as it can only be judged based on the rules that were set for it - and those differ person to person, case to case (Muth, 2020).

3 DESIGN PRINCIPLES IN DASHBOARD CREATION

While business intelligence helps companies make data-driven decisions based on the data exploration, the high volume of data being produced daily makes BI relatively unused, as the most part of the stored data is never utilized in any sort of analysis (Adair, 2020). This may be due to the lack of time or resources, but it also coincides with the BI need of the understanding of data in order for it to be visualized. Data can be visualized in a number of ways, but in BI, data visualizations are usually grouped in a dashboard, which provides an overview on a single screen. The main goal of dashboards is that they are purpose-built, as it may be used as the key tool for decision making in a business (Brush, 2020; Tableau, 2021).

Data visualization is in other words visual communication of data to the end users via a variation of chart types and color usage. If done properly, data visualization turns unintelligible sets of data into knowledge (Midway, 2020; Solis, 2019). When building dashboards, users need to be aware of the type of visual they are using, in regard to the type of data they are trying to present, while keeping in mind that the visual also had to be the easiest for the end-audience to comprehend (Nussbaumer Knaflic, 2015). Effective data visualization shows the data and tells the story, while reducing the clutter around it to keep the focus on the point we are trying to emphasize (Schwabish, 2014). When using any type of visual, we should employ the use of titles, labels, and tooltips to show a deeper meaning of the data at hand (Dougherty & Ilyankou, 2021).

Behind data visualization, there are a number of color, design, and storytelling theories. Color theory - as a science of using color and understanding how humans perceive it - can be broken down into several sub-categories (RevUnit, 2022). Color anatomy differentiates between hue, saturation, and luminosity. Color harmony implies that colors can be used in such a way that we will perceive them as cohesive or not. The color schemes in data visualization can be predefined, but they are not always optimally picked in relation to data presentation (Kargin, 2022). Lastly, color accessibility is what most people tend to dismiss, even though more than 7% of the population has some sort of color blindness (Clinton Eye Assoiates, n.d.). Intertwined with data visualization, all of the components of color theory have an impact on the colors we utilize in dashboards, with most theorists putting emphasis on using less hues and using lower saturation in the color palette. In regard to the choice of color, we also have to keep in mind that the human brain tends to group things together if they are the same color (or hue), which should always be considered when showing data that may not be connected (Decker, 2016). Berinato also emphasized that when we are building the color palette, we should use the colors for data comprehension and not company satisfaction (Berinato S. , 2019).

After data visualization made a breakthrough in the scientific world, people like Tufte and Few started setting ground rules for use of color and design. While Tufte set down the rules of graphical integrity, which emphasize that the representation should be clear, detailed, and show data variation and not design variation (Tufte, The Visual Display of Quantitative Information, 2001), Few continued with the use of color in even more depth. He defined that people see color in context, and we should also utilize it as such in data visuals. His most important rule was that colors should be used meaningfully and with restraint – only when it is used for a communication goal (Few, 2008). Next to these rules, the Gestalt design principles try to explain how the human brain perceives design, and how the human brain will automatically group things together based on either proximity, color, size, or position, which makes effective dashboard design even more complex (Ahmadi, 2020).

For centuries, people have used visual tools to help explain data to one another via charts, maps, and diagrams. One of the underlying reasons might be the fact that the human brain retains visual information more readily than spoken information. Data visualizations can be a powerful tool for communicating large amounts of data quickly and straightforwardly. In order for the information to be transferred as intended, however, the visualization has to be effective, and more importantly, easily comprehendible.

As written in the preceding chapter, people process information differently based on a number of factors, but even subconsciously, our brains tend to group, define, and unlink information with ease, which in turn effects our understanding. If we, for example only follow the Gestalt principle for proximity, clusters of data will automatically be grouped together, even if that was not the creator's intent.

Next to the guidelines set by a number of authors on different topics of data visualization, the IBCS (International Business Communication Standards) keeps an updated list of areas on to which business communication has to comply with, in order to meet their standards. The areas on which they focus are abbreviated in their "SUCCESS formula" (IBCS, 2022), which is an acronym for say, unify, condense, check, express, simplify, and structure. The main message the IBCS is trying to convey is that reports should convey messages, to avoid becoming a data collection. Unify is connected to the gestalt theory, where things that look the same tend to get grouped together (Ahmadi, 2020), which is commonly expressed in many different visualization practices. The condense and structure parts also comply with gestalt principles. Lastly, IBCS pushes forward that visual integrity should be held to a high standard, which we can ensure by presenting information in a comprehensible manner and by avoiding biased visualizations (IBCS, 2022).

In order for visualizations and therefore dashboards be effective and comprehensible, we have to define a set of guidelines to follow, which will increase the usability and clarity. The guidelines have to cover topics from visualization, design, and color requirements, while keeping a storytelling narrative as the main message.

Visualization type requirements that we should try to implement in order to maximize the dashboard quality and data comprehension:

- Utilize form follows function theory: what the function of the data is (what we want people to be able to do with the data) in order to construct the form (visualizations) (Nussbaumer Knaflic, 2015).
- Use additional text or labels as explanation if needed (Dougherty & Ilyankou, 2021; Nussbaumer Knaflic, 2015).
- Minimize heavy borders or axis colors in every type of visualization (Nussbaumer Knaflic, 2015; Few S. , 2012).
- Choose chart types that will correspond best to our data (Microsoft, 2021; Annesley, 2020).

The design requirements, which correspond with color, and placement are the following:

- Do not rely just on color: using visual cues such as text or icons can help users understand more than just color grouping (RevUnit, 2022).
- Use colors for data comprehension, not visual beauty (Berinato, 2019; Few, 2008)

- Group things together visually following gestalt principles (proximity, similarity, etc.) (Ahmadi, 2020; Betzendahl, 2020; Chapman, 2017).

The color requirements, which define the color scheme and hue usage:

- Limiting the use of color: the color scheme should be limited to 10 or fewer colors (Careri, 2022; Microsoft, 2021; Stone, 2006).
- Use hue only when necessary and use greys and blacks where color is not important (axis, average lines) (Few, 2008; Nussbaumer Knaflic, 2015).

The dashboards should also follow a storytelling narrative, where the main message of the dashboard should be clearly visible (Nussbaumer Knaflic, 2015).

Following rules is the basis of any practice, not just data visualization. However, as Muth (2020) emphasized, rules tend to block creativity. Rules should therefore be taken as a guide, and not a checklist, especially when the user is well versed in dashboard design and understands which elements will maximize data comprehension and retention, while shifting the focus to the main point of the data, and not only visual additions (Muth, 2020).

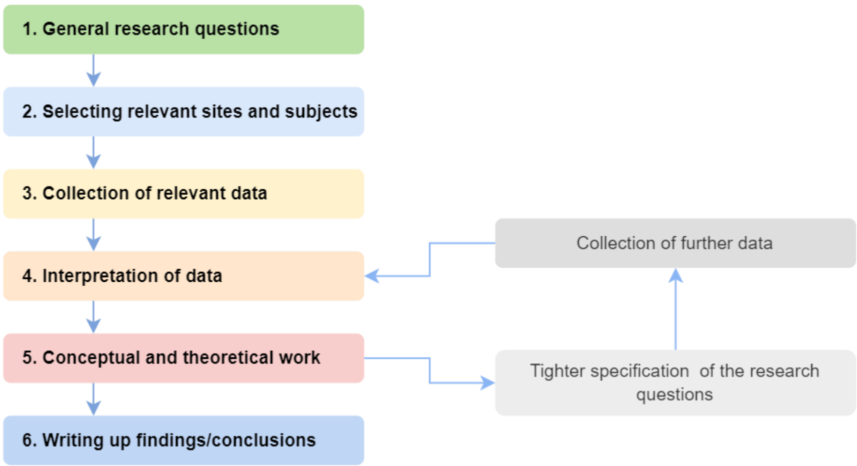
4 METHODOLOGY

Based on the literature review provided and the goals of the thesis, the methodology will be as follows. First, this thesis will use the design principles mentioned beforehand and create examples of good and bad visualizations in Microsoft Power BI, as it is the most commonly used business intelligence visualization tool, as well as a range of functions to work with (Adair, 2020). I will focus on using the same data set while creating a dashboard that will give the user the same amount of information while presenting it clearly, while also creating a second dashboard, the visualizations of which will be unclear for interpretation or distorted by the use of color or composition. I will then conduct a semi-structured, in-depth interview, where users will have both dashboards presented to them and will give their feedback on what they learned from each of them, and what their perception of the visualization or dashboard in general was. Lastly, I will conduct a comparison into both sets of feedback provided, and gather what users find good or bad in both dashboards.

4.1 Qualitative Data Collection

The research strategy used for the methodology part of this thesis is qualitative research. Qualitative research is inductive in nature, as we use the research model to generate theories. The qualitative research outline includes six steps, which we followed in order to generate a successful research model (see figure 13).

Figure 13: Outline of the Main Steps of Qualitative Research



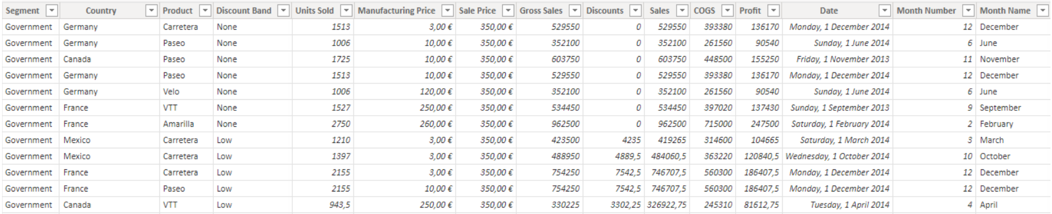
Adapted from Bell & Bryman (2011).

As reliability and validity are highly important criteria in the assessment of the quality of research, many people argue that qualitative research is not a valid way to collect data, as it is hard to. To avoid the pitfalls of qualitative research and comply with the aforementioned requirements, I employed a number of tools to help keep the standards as high as possible. Firstly, the external reliability of the research was maximized, as the study can be replicated. The main issue with external reliability in any qualitative research is that the social setting cannot be replicated. The actual groups and dashboard creation, though, can be replicated, because we used predefined sets of people that can be found anywhere. Additionally, external validity as to the degree to which the findings can be generalized across social settings, was of the utmost importance, as the use of both the color theory and design implications for the dashboard creation can be used in a variety of use cases (Bell & Bryman, 2011).

4.2 Development of Dashboards and Visualizations

After doing a preliminary critical literature review, I used the color, design, and visualization rules written about in chapter three to build up my dashboard. For the purpose of this experiment, I used generated data from Power BI. The default data set consists of financial data from September 1st of 2013 to the first of December 2014. The data set included attributes like country, price, units sold, profit, date, and product information, but the goal was for the data set to be neutral and non-recognizable, so that any emotional attachment would not play a role in data recognition and retention.

Figure 14: Financial Data set



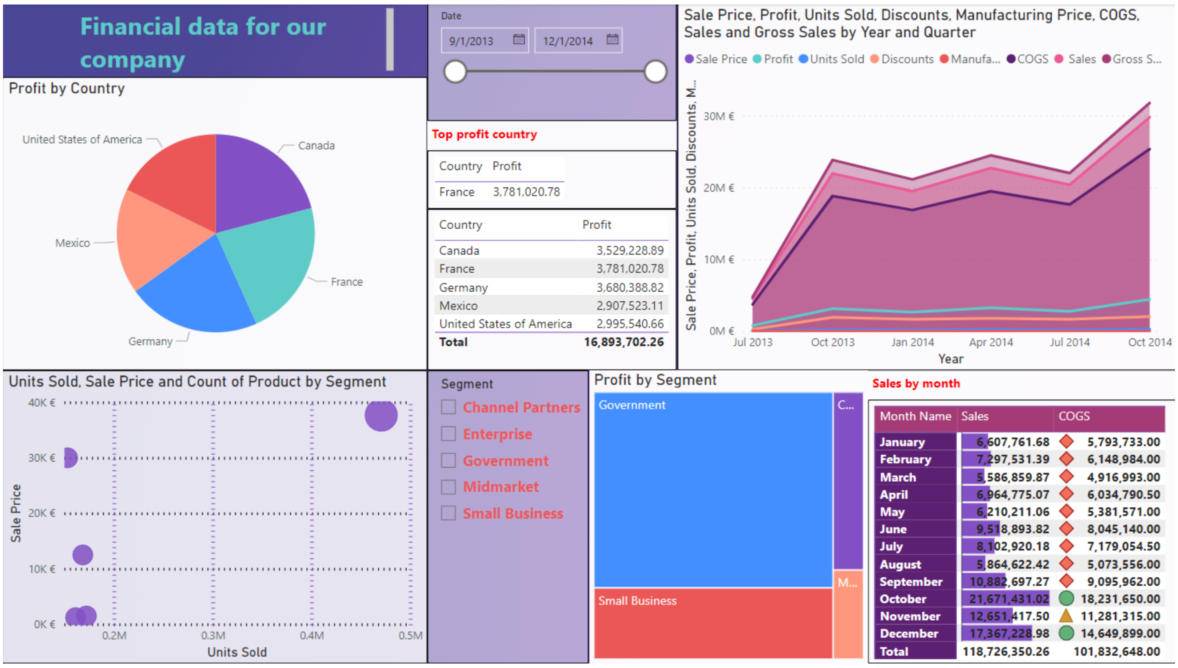
Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name
Government	Germany	Carretera	None	1513	3,00 €	350,00 €	529550	0	529550	393380	136170	Monday, 1 December 2014	12	December
Government	Germany	Paseo	None	1006	10,00 €	350,00 €	352100	0	352100	261560	90540	Sunday, 1 June 2014	6	June
Government	Canada	Paseo	None	1725	10,00 €	350,00 €	603750	0	603750	448500	155250	Friday, 1 November 2013	11	November
Government	Germany	Paseo	None	1513	10,00 €	350,00 €	529550	0	529550	393380	136170	Monday, 1 December 2014	12	December
Government	Germany	Velo	None	1006	120,00 €	350,00 €	352100	0	352100	261560	90540	Sunday, 1 June 2014	6	June
Government	France	VTT	None	1527	250,00 €	350,00 €	534450	0	534450	397020	137430	Sunday, 1 September 2013	9	September
Government	France	Amarilla	None	2750	260,00 €	350,00 €	962500	0	962500	715000	247500	Saturday, 1 February 2014	2	February
Government	Mexico	Carretera	Low	1210	3,00 €	350,00 €	423500	4235	419265	314600	104665	Saturday, 1 March 2014	3	March
Government	Mexico	Carretera	Low	1397	3,00 €	350,00 €	488950	4889,5	484060,5	363220	120840,5	Wednesday, 1 October 2014	10	October
Government	France	Carretera	Low	2155	3,00 €	350,00 €	754250	7542,5	746707,5	560300	186407,5	Monday, 1 December 2014	12	December
Government	France	Paseo	Low	2155	10,00 €	350,00 €	754250	7542,5	746707,5	560300	186407,5	Monday, 1 December 2014	12	December
Government	Canada	VTT	Low	943,5	250,00 €	350,00 €	330225	3302,25	326922,75	245310	81612,75	Tuesday, 1 April 2014	4	April

Source: Own work.

The Bad Dashboard Example

After the data set was imported, I chose one of the more colorful Power BI themes (Bloom) as users usually believe that more color will generate better attention. The data set was then used to generate visualizations on the dashboard. As most users firstly stick to the pie chart, I started off with that as well.

Figure 15: Bad Dashboard Example in Power BI



Source: Own work.

For the dashboard, seen in figure 15, what I included was the following:

- Profit by country (pie chart),
- Top profit country (text box and table),
- Profit by countries (table),
- Units sold and their sales price, split by segment, while the bubble size shows the count of products (scatterplot),

- Profit by segment (tree map),
- Sales and COGS by month (text box and table),
- Sale price, profit, units sold, manufacturing price, cogs, sales, and gross sales by date (area chart),
- Date slicer (slicer),
- Segment picker (slicer),
- Title (text box).

While creating the visuals, I tried to include as much information as possible, while covering a range of visuals. I also included some slicers and the report title. To dissect the issues with this dashboard, we split them up into the three categories we defined beforehand (visualization type requirements, design requirements, color requirements).

The visualization type requirements for the dashboard were not implemented in the following key points:

- Profit by country (pie chart) – a pie chart would be useful to show parts of a whole, but for profit-data this is not the right visualization. The pie chart parts are also too similar to each other for the user to differentiate between them.
- Top profit country (text box and table) – we can use a title option for the table itself, but in general this could and should be shown either as text or as a card visualization, as it only shows one value (the top performer).
- Profit by countries (table) - if the original pie chart was a table or a bar chart, this part would not be needed as it repeats the same information.
- Units sold and their sales price, split by segment, while the bubble size shows the count of products (scatterplot) – the visualization is “too hard” to understand for a regular user, but there is also no need for it to be a scatterplot. It could be presented as a bar chart, but as units sold and sale price together equals profit, we could also just show profit by segment.
- Profit by segment (tree map) – as the scatterplot could be made into profit by segment, this is a useless visualization as well, as it shows the exact same data (as the scatterplot).
- Sales and COGS by month (text box and table) – the table itself is an okay selection for the visualization, but the color scheme of the table and the conditional formatting behind it is what makes the data fall into the background, as the user only focuses on the rows, column headers, and data bars.
- Sale price, profit, units sold, manufacturing price, cogs, sales, and gross sales by date (area chart) – the area chart is overpopulated, and therefore the information value is lost.
- Date slicer (slicer) – the date hierarchy is usually better shown in a dropdown menu, where you can pick either the year or month inside a year. It is harder to achieve when dealing with a time span of over a year.
- Segment picker (slicer) – the segment picker is okay, but the background opacity and text color are a bad choice.

- Title (text box) – the title stands out, which is okay, but the textbox size is too small, which produces a slider on the right side of it.

The design requirements in the aforementioned dashboard were not executed in the following points:

- The visualizations are not equipped with enough visual helpers for them to be understood easily. A great example is the pie chart that is missing at least the values if not percentages as parts of a whole.
- All the visualizations are too close together – there is no usage of whitespace on the page, which makes grouping by proximity impossible. The similarity when it comes to the color scheme is also non-existent, as the same color does not represent any distinct value.

The color requirements of the dashboard were not utilized properly in the following ways:

- The color scheme utilizes more than 10 colors on one page, which makes the reading of it very hard, as all the visual stimuli are overwhelming.
- There is hue differentiation but not saturation or luminosity, which makes it hard to read. The lines and titles are not variances of black and grey which also goes against color theory principles.

Data quality framework requirements that were not used properly in the dashboard:

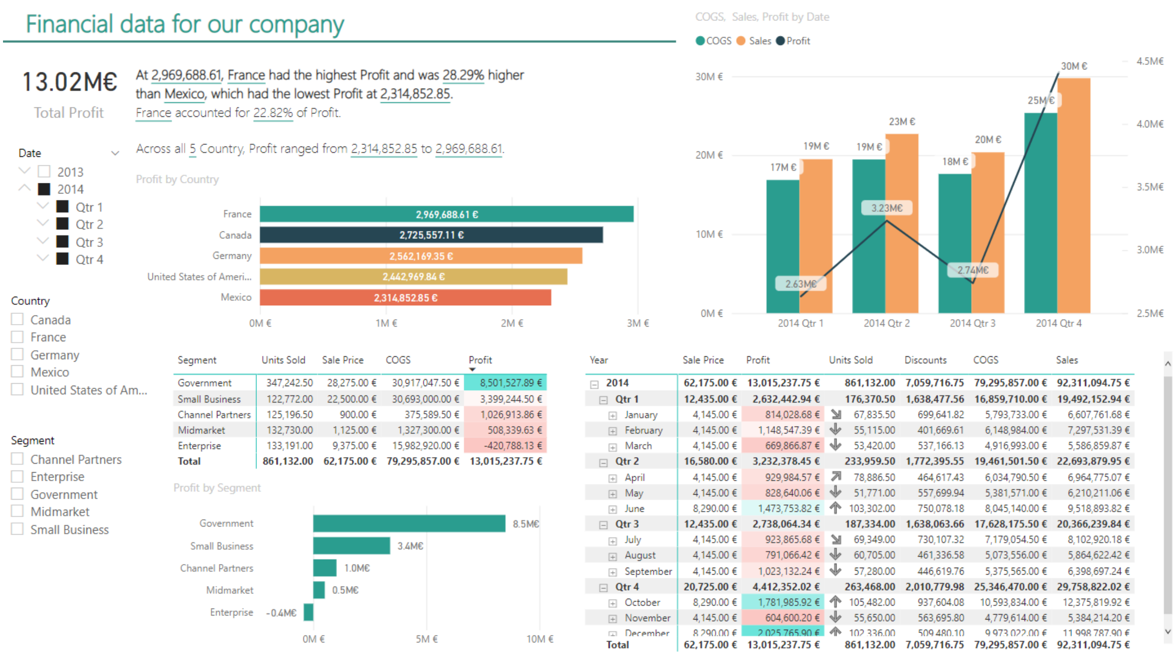
- Representational DQ was not followed, as the interpretability of the dashboard is quite low, as well as the ease of understanding and the consistency of the representations.
- Contextual DQ's dimensions of the amount of data used were most transparently ignored in the area line chart, as the amount of data is not comprehensible.
- Intrinsic DQ's accuracy is ignored wherever we use rounded numbers without showing the exact number as well.

Accessibility DQ is not an issue in either dashboard, as the data is available to all users.

The Good Dashboard Example

The second dashboard, which can be seen in figure 16, was developed by following the structure and guidelines written in the summary of the theoretical part as close as possible, while using the same data set and showing the same KPI's. One of the main things that were changed is the color theme, with the main focus on using a white background to eliminate any focus that the purple one might have shifted.

Figure 16: Good Dashboard Example in Power BI



Source: Own work.

For this dashboard, the following was included:

- Total profit of the company (card),
- Highest profit by country text box with dynamic values (text box),
- Profit by country (bar chart),
- COGS, sales, profit by date (line and stacked column chart),
- Units sold, sales Price, COGS, and profit by segment (table),
- Profit by segment (bar chart),
- Sale price, profit, units sold, discounts, cogs, sales by date hierarchy (matrix),
- Date slicer (slicer),
- Country (slicer),
- Segment (slicer),
- Title (text box).

The visuals chosen can be compared by separating the information into segments. Firstly, the profit, which is the most common measure used in both dashboards, has been used in the following way in the good dashboard example:

- Total profit is shown as a card visual, as the most important piece of information,
- Profit information is dynamically explained for any type of user via a text box,
- The profit elements used in the bad example (tree map, text box, tables, charts) are replaced with visuals that are more easily explained and inspected, while also making

use of the Power BI visual title option, which omits the need for additional textboxes for titles, while also making them stand out less (supportive information).

The date was used in another slicer option:

- The slicer is in a list slicer form, which can cover years, quarters, months, or dates (date hierarchy), and makes the date selection easier than using a slider.

The business segments were now shown differently:

- Segments are shown in a table, ordered by profit, while using conditional formatting for the background colors. The highest profit is colored green, while the lowest stands in a soft-red hue.
- The second segment visual is a bar chart, which clearly shows the leaders and their profit.

Main financial data by date was now shown in a matrix table by year hierarchy, using conditional formatting to shift focus to profit changes.

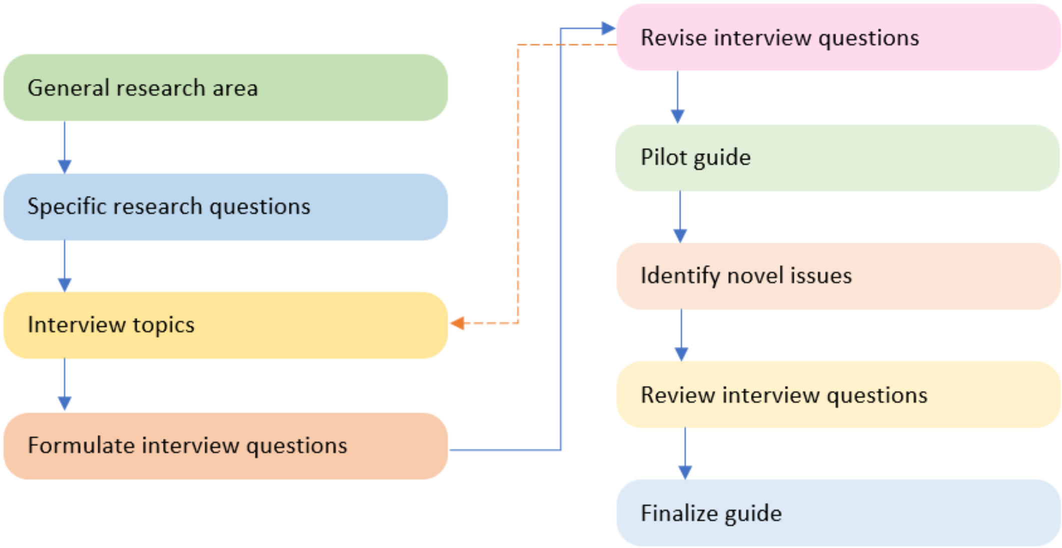
4.3 Interview Guide Development

In qualitative interviewing there is less strict requirement for structure, as any wandering away from the main topic is actually considered positive. This is due to the fact that it gives us insight into what the interviewee deems important. According to Bell and Bryman, the main goal of the qualitative interview is seeing the interviewee's point of view, based on which the final assumptions and theories are generated. The time frame of a qualitative interview is not predefined, the entire interview is much more flexible, and its flow corresponds to the answers given by the interviewee (Bell & Bryman, 2011).

For this use case, we conducted semi-structured interviews. Generally, semi-structured interviews have an interview guide which corresponds to specific topics, but the interviewee has a lot of room in how to reply to the questions asked. The questions are also not necessarily ordered, and may also not be asked, if the interviewee already answered them in another context. However, we still need some structure for us to be able to compare the interviewee's answers between different groups and to be able to draw up conclusions and theories (Bell & Bryman, 2011; Crinson & Leontowitsch, 2006).

The basics of drawing up an interview guide start with preparing a number of topic areas and then ordering them in a way to give them a natural flow, while also being flexible enough for the questions to be turned around or switched. We also have to formulate the questions in such a way that will help answer our research questions, but still having them be wide enough for the interviewee to state their point of view without being too restricted. An important factor is also the use of comprehensible language, which has to be relevant to the people included in the study. Lastly, we have to record general "face sheet" information of general kind to be able to contextualize people's answers – for example their age, gender, and employment status. The formulation of the questions themselves flowed through the guide depicted in Bell and Bryman's book, which can be seen in the figure 17.

Figure 17: Formulating Questions for an Interview Guide



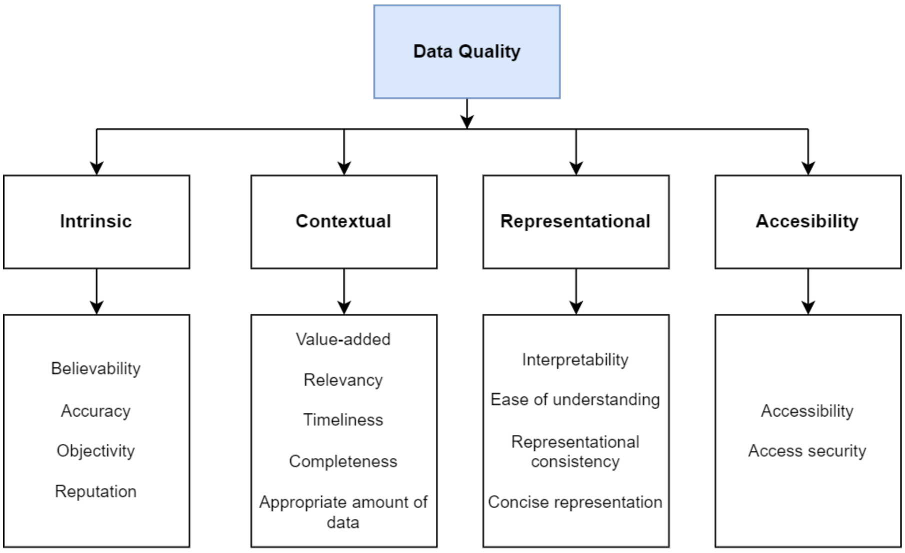
Adapted from Bell & Bryman (2011).

Lastly, the interviews were recorded and transcribed, so that the interviewer could keep eye contact and observe any behavioral changes of the interviewee without accidentally leaving out any important information (Bell & Bryman, 2011).

As we were exploring dashboard quality, we had to have a measurement type through which quality could be defined. There are a number of frameworks that discuss data quality (DQ), and this master's thesis employs the Wang and Strong Data quality framework, pictured in figure 18. The authors argue that that improving data quality only by accuracy is not enough, so they have decided on all the different dimensions of data quality that are important to consumers and grouped them into four categories – intrinsic, contextual, representational, and accessibility.

As per the authors, data quality needs to have the trait of 'fit for use', which translates to taking the clients viewpoint, as the end user will be the one to use and judge the product. But as development characteristics are not the same as use characteristics, they developed their framework, which employs the most valuable data quality dimensions, as determined by the users in their research. Each category of the framework is a hierarchical construct that groups together a number of data quality dimensions.

Figure 18: Data Quality Framework



Adapted from Wang and Strong (1996).

Firstly, intrinsic DQ entails that the data has quality in its own right. It achieves that by being believable, accurate, objective, and having a good reputation – it is, in other words, a trustworthy source. The data consumers stated that accuracy and objectivity alone are not satisfactory enough for data to be considered high quality. Secondly, contextual DQ implies that the data has to be considered within the context of use. The authors sorted value-added, relevancy, timeliness, completeness, and the amount of data that is appropriate for use as the DQ dimensions that correspond to this category. Thirdly, representational DQ means that the data is presented in a way that makes it easy to interpret and understand, while keeping the qualities of being consistent and concise. It falls into the purview of this dimension, for example, that the data should not be presented in a foreign language for example. Lastly, the accessibility DQ implies that the data is accessible by the user, while making sure that it is still kept secure. As stated by the authors, the improvement of data quality comes from the understanding of what data quality means to consumers.

In the context of this thesis however, not all four dimensions of the Wang & Strong model come into use. When researching color, composition, and design of the dashboards, intrinsic and accessibility DQ do not offer a scale on which DQ could be defined, as they deal with data accuracy, reputation, objectiveness, and data access. Therefore, the only two DQ dimensions that were needed to be included in the interview guide were representational and contextual (Wang & Strong, 1996).

Through the literature review, we defined a general research area, which deals with color, composition, design, and DQ. With specific research questions predefined in the introductory part, regarding the effect of color in the interpretation of visualizations, and the role of color and composition in BI development, we had to set the interview topics in order to start formulating questions. During the literature review, it was clear what the authors stated as the most important guidelines or rules to follow when creating any visualization, but as this thesis has a broader scope of dashboards and not only visualizations, but we also had to cover a wider set of topics. The DQ framework was used as the basis upon which the different topics would have impact measured.

As Bell and Bryman stated, getting answers that may not fully reference the initial question is encouraged, and it gives the interviewer a deeper insight into what the interviewee may be experiencing or noticing during the interview. Using this in describing the dashboard gives the guide only a general structure, as the questions can be moved, omitted, or changed, depending on the interviewee reactions and answers. Because of this, the interview guide had to be structured into general sets, that derived from the theoretical literature review. The literature review gave us general interview topics, which were intertwined with the research questions posed in the introduction part of the thesis. After the topics of general information, dashboard, data/data quality, color, and possible corrections of the dashboard were set, the specific interview questions could be constructed (Bell & Bryman, 2011).

The interview guide, which can be seen in the appendices, was built in sets of questions, which went from broad to more specific, to try not to narrow the user down too fast. The first set contained general information about the interviewee and their experience with BI, to have an insight of which events, experiences, or education might have had an effect on their dashboard understanding. The following set dealt with dashboards, to see if color and composition had any effect on their initial understanding (RevUnit, 2022; Berinato S. , 2019; Kosara, 2014; Schwabish, 2014). Next, the set of data and data quality had to dealt with contextual and representational DQ, to see which elements had effect on them and how (Wang & Strong, 1996). The set that followed was strictly defined around the choice of color, and their awareness of its use in the dashboard (Bartram, Patra, & Stone, 2017; Few, 2008; Kargin, 2022). The last set posed questions of possible corrections to the dashboard.

Utilizing the chart in figure 18, the interview questions that were initially set were tested with a BI developer who gave initial understanding into which questions were useful and gave insight into the comprehension of data and design, and which could either be answered in a different way or should not be posed at all. One of the main things that they emphasized was that the initial reaction of the dashboard was what they felt was something worth noting. Through iteration of the questions, and the identification of the issues, the revised question set of the interview guide was finalized.

4.4 Interview Guide Question Sets

The first set of questions included the interviewees work or educational experience with not only BI tools, but also their understanding and usage of graphs and data in general. All the users had to have some experience with data or its graphical representation, as it was the basis of getting information from it, and it also represented a group of people that would ultimately come across dashboards in their work experience.

The second set included a series of questions which showed the interviewees understanding of the dashboard and recorded their initial reaction to what they saw on the screen. The last question in the set encouraged the interviewees to explain their trail of thought when seeing the dashboard for the first time. This part had a focus on contextual DQ.

The data and data quality question set tested their understanding of the data and their ability to find information on the dashboard itself, with sub-questions to how easy the information was to locate or understand. This set revolved heavily on representational DQ, as they had to understand and interpret data on the dashboard.

The fourth set was based on color understanding and used their subjective attitude towards the color selection. It also required them to think if the colors impacted their information retention. Lastly, the corrections of the dashboard question set were placed to see how deep their understanding of dashboard building goes, and if the color or visualization selection on the dashboard should be changed or not. Both of these were intertwined with contextual and representational DQ.

The guide was a general structure, which could be deviated from if any of the interviewees already answered a certain question in regard to another topic. The most important thing was to get insight into topics that revolved around the initial purpose set in the introduction. As the main goal of dashboards in to show a general overview on a certain topic, while showing a collection of different data visualizations, the main idea is to show quality data which in turn produces quality business decisions. The DQ aspect of the created dashboards will therefore be measured through the Wang and Strong DQ Framework.

4.5 Semi-structured Interview and Experiment Execution

To understand which color theory and dashboard design concepts influence the comprehension of data in a given setting, semi-structured qualitative interviews were held with six people that were divided into two groups. To gather the maximum amount of information but have the information be comparable, the two groups each had three members with a similar level of experience in data visualization and dashboards. One group was presented with a dashboard example that did not conform to the color and design principles mentioned in chapter three, while the other followed the rules of color usage and graph placement thoroughly.

As seen in table 1 below, the two testing groups were each comprised of three people. The first user (beginner) had little to no experience with BI dashboards, while still being computer literate. The second group member (casual user) was an average computer user who has dealt with data before but was not used to looking at graphs every day, yet still had some general knowledge regarding BI. The third user (advanced user) in both groups was someone that dealt with BI daily and knew what to look for, and where. All the interviewees had no knowledge about the main purpose or goal of the experiment, or the data set used.

Table 1: List of Interviewees

Interviewee	Age	Job Status	Level of Experience
Group 1			
Interviewee 1 (IN1)	28	System Administrator	Beginner
Interviewee 2 (IN2)	27	Business Informatics Student	Casual User
Interviewee 3 (IN3)	31	Data Architect/BI Lead	Advanced User
Group 2			
Interviewee 4 (IN4)	25	Office Administration Worker	Beginner
Interviewee 5 (IN5)	27	Business Informatics Student	Casual User
Interviewee 6 (IN6)	48	Business Analyst	Advanced User

Source: Own work.

To make the users comparable, they were all roughly in the same age group, with similar IT skill sets and educational backgrounds. The advanced users both had roughly the same length of working experience in BI.

The execution of the experiment/interview was mainly aimed at studying two primary factors: the initial reaction (emotional component) and the data processing (based on how hard information is to find, and the users' reaction to the dashboard itself).

5 ANALYSIS AND RESULTS

The following chapter will describe the execution of the methodology, present the compared dashboards, and discuss the results of the interviews. As written in the previous chapters, I formed two groups of three people, with each group then being exposed to a different dashboard. One dashboard followed the color and design best practices, while the other did not. The people selected into both groups had to have similar level of experience with BI, so that the groups could be compared. All of the participants allowed recording of the interviews and provided vital insights for this thesis.

5.1 Findings

The following subchapters will discuss the findings from the semi-structured interviews. The interview guide was made in an attempt to get answers to my research questions. Analysis was done into the comparison of participants within one group, as well as between groups. To produce the results, I asked users to look at and use the dashboards presented, and then give their comments on them, while also posing them the structured questions prepared beforehand. At the end, they got to see the other dashboard and comment on the difference or general likeness. The following subchapters will include summaries and quotes from the interviews and will be divided as per the interview guide that can be found in the appendices. The interviewee demographic is shown in Table 1.

The main links I was searching for were based around the guidelines I set as the standards for good visualization creation. They were the limitation of color usage (limiting the number of hues), form following function (the visualization type being dictated by what people need to see), using gestalt as the basis of grouping and placement (grouping things together visually), and using the chart types that do correspond best to the data at hand.

5.1.1 Demographics

When I started talking with the interviewees, we firstly discussed their backgrounds and experience with BI and data in general. In a wider sense, all the interviewees had seen graphs or charts before, and have dealt with data, in either a classroom setting or at work. There were 6 interviewees, with an average interviewee age of 31 years. As seen in figure 19, the interviewees had a range of BI experience behind them. The average amount of BI experience was 2.83 years, while the 4 interviewees with previous BI knowledge, had an average of 4.25 years of BI experience between them. The least BI-experienced interviewees had seen dashboards in context of other tools but had not yet used any BI tools themselves. The interviewees all had a higher education degree, having completed either a bachelor's or master's degree.

The interviewees were split randomly into the two groups which determined if they saw the good or bad dashboard example first. Before seeing the dashboard, every interviewee got an explanation of what type of data they are going to be looking at (business financial data), and that the questions were going to be based on the dashboard.

Figure 19: Interviewee Demographics and Experience

Number of Interviewees: 6
Average interviewee age: 31.17
Number of Male interviewees: 4
Number of Female interviewees: 2

Average years of experience: 2.83

First dashboard they saw	Interviewee	Age	Education field	BI Experience years	BI Experience level
Bad	IN1	28	Telecommunication	0.00	Basic
Bad	IN2	27	Business informatics	2.00	Medium
Bad	IN3	32	Computer science	6.00	Expert
Good	IN4	25	Economics	0.00	Basic
Good	IN5	27	Business informatics	2.00	Medium
Good	IN6	48	IT organisation	7.00	Expert

Source: Own work.

5.1.2 Initial Reaction

When they were first presented with the dashboard, the responses between the interviewees were different. Interviewees one to three (who saw the bad dashboard example first) were mostly stunned for a second. Afterwards, the least experienced one (IN1) said that they simply felt like there was a lot of clutter on the dashboard. Meanwhile, both IN2 and IN3 commented on how purple everything was and expressed their disgust with the dashboard color selection. IN3 also laughed as a part of the initial reaction, and commented that even though it may look bad, it was not the worst dashboard example they had ever.

Interviewees four to six, which were first exposed to the good dashboard example, had a less pronounced initial reaction to the dashboards. The least experienced (IN4) commented that it looked nice, and that they liked the choice of color, as the colors were not too loud or garish, while implying that most people usually go for really bright or strong colors when working with graphs. Interviewees 5 and 6 however had a less positive reaction to the dashboard. IN5 said that they first saw the bottom right matrix, which made them feel like there was simply "too much information on the page", which made them a bit confused as to "what it is trying to let them know". IN6, being the most experienced one in the second group, commented that they saw a lot of information on the page, and the composition made them felt like it was cluttered, and not well organized at all. They also commented a similar thing to IN5, that they feel like the dashboard does not tell a cohesive story.

5.1.3 First Impressions

After the interviewees looked through the data and dashboards as a whole, they were asked what the first thing they may have noticed on the dashboard was. IN1 said they firstly noticed the pie chart, as it was the "most clean and isolated visualization on the page" but continued by saying that the next thing he noticed was how unusable it actually was, having no data labels, while all the slices were roughly the same size. IN2 and IN3 both noticed the middle table of top profit country first, and they both agreed that the reason was the white background of the table, and its center position on the board.

When it came to the good dashboard, IN4 and IN6 both looked at the profit by country first, with it also being the second thing IN5 saw. According to them, they were randomly looking at the bottom right when the dashboard was opened and were a bit startled at the amount of data in the matrix, which led them to focus on that first.

No matter what the interviewees saw first, the first piece of information they gathered from the dashboard was the same data in 66,7% of the cases. IN1, 2, 3, and 4 all firstly gathered which country had the highest profit as the first piece of numeric information they got from the dashboard. The focus of IN5 and IN6 went to the top left card visual of total profit first, and afterwards to the top profit country. IN1 through 3 agreed that they looked at the top profit country first, as it was the most normal piece of data available for collection on the page, while IN5 and IN6, who saw the good dashboard first, agreed that they always look at KPI's or card visuals first, as they are usually the indicators of what the dashboard is about.

All of the interviewees had the same path, through which they went when inspecting the dashboard. All of them agreed that their attention flowed from graphs, which were big and colorful, to tables, which had fewer colors and less overall impact.

5.1.4 Data and Data Quality

The next concept we discussed was data. All of the interviewees were first asked if they could tell me which country had the highest profit, and if the information was easy or hard to find.

IN1 to 3 all had a problem finding the data. Even though the data was in the center of the screen, they were all initially drawn towards the pie chart with no labels. After noticing the table in the middle, they all sorted it by profit, as it was unsorted beforehand, and told me the answer. They agreed that finding the data was not very hard to find, but it was not easy either. In contrast, all of the interviewees that saw the good dashboard first found the data in a split second and were shocked when I asked if it was easy to or hard to find, as if I might have been joking.

The second question in data identification was the determination of the segment with the highest profit. All the interviewees could tell me the segment straight away, but the numbers of how big it is were easy to find only for interviewees 4 to 6. The least experienced in the first group had some issues, as they did not know they could hover over the visual for additional information, while IN2 and IN3 saw the numbers, but continued on to say that having an area type visual in that proportion may be useful only to show that the Government segment is the biggest, while any additional information would be better shown as a table or as a bar chart.

Afterwards, they were asked, if they could define which quarter in which year had the highest profit. Here, most of the interviewees that saw the bad dashboard first were the first to answer quickly and correctly. Even though the color scheme was throwing them off, they all agreed that it is visible because of the conditional formatting in the bottom right table. On the other hand, IN4 to 6 had some more issues with finding the information. IN4 was the first to answer correctly, even though the information was gathered from the line chart. IN5 and 6 both tried to find the information in the bottom right matrix but did not notice the scroll option in the matrix, which led them to answer with the wrong year at first. After they drilled up the matrix, so it showed year-quarter and not year-quarter-month hierarchy, they got the information straight away, that the highest one was Q4 in 2014.

The last question in the data category was oriented towards what kind of insight they got from the data. The answers to the question "did the number of units sold impact the profit" varied. IN1 was the only one that said yes right away, while IN2, 4, and 5 all answered with "I am not sure, because the data does not point me towards an answer". IN3 and 6 however, both said that the data does not imply a correlation between profit and units sold. In this context, IN4 was the only one that said that even though they are not sure what the data says, in pure logical or economic sense, the number of units sold should have an impact on profit.

At this point, IN2 and 3 emphasized again that the use of purple was *too much*.

5.1.5 Use of Color

After we discussed data, we went on to color. The first question was a more general one, where we talked about what their general impression of the colors used on the dashboard was.

The interviewees that first saw the bad dashboard example all said, that the first thing they disliked about it is the use of purple. IN1 said that they are bothered by the background, as it is "too much at once". He also stated that the text on the white-background boxes is hardly readable, because of the transparency. On the other hand, IN1 also stated that apart from the purple, the colors seem okay, as there seems to be a limited number of colors. In contrast, IN2 said that there should be a color palette implication, where one color means one thing, so that the user does not mix things up that easily (referring to using the same colors for differentiation between segments and countries). Generally speaking, IN2 said that there are too many colors, and that the visuals chosen combined with the colors make it not readable and not comprehensive. Similarly, IN3 also stated that there are too many colors used, and that the purple is a good choice at all. One of the main things IN3 emphasized was that the "red-purple combination was an eyesore" and that "everything is just purple". He also stated that the line chart colors are too similar to each other, which makes it hard to read.

In contrast, IN4, who first saw the good dashboard said that the colors are toned down, playful, and all-together a great choice. Meanwhile IN5 and IN6 both stated that the colors chosen are not likeable, but it is more of a personal color preference, as they also agreed that they are okay as far as the visuals are concerned. They both also said that they do not really have a strong opinion on the color choice.

The next part I discussed with the interviewees was if the colors on the dashboard or visuals shifted their attention, where all the interviewees agreed that they did. While IN4, IN5, and IN6 said that they feel like their attention was affected by the colors, IN1 to 3 stated that it shifted their attention significantly and made them "run in a maze with their eyes over the dashboard".

Lastly, I asked them if they felt like the colors chosen impacted their understanding of the data, where the group answer was also affirmative again. While the first group of interviewees agreed that it did impact their understanding, they all went on to explain that it impacted them in a negative way – they either had a hard time understanding the visualization, or they drew up wrong conclusions based on color. Contrary to the first group, IN4 to 6 all agreed that the color usage made the data easier to understand, and they all highlighted the use of conditional formatting in the matrix, as it makes finding higher or lower numbers easier.

5.1.6 Corrections for the Dashboard

All of the interviewees were then asked if there was anything they thought should be changed, and if so – what. IN1 started off by wanting to change the visual types according to the data we are trying to show. One of the things that bothered him the most was the tree map, as he felt it was "absolutely useless" if you are trying to get any more information other than which segment sold the most. All of the interviewees in the first group agreed that the line chart should be changed, as it contains too much data, which makes it unreadable, and they also agreed that having too many similar colors on the line chart is not of much help either. Another thing that they all noticed was that the pie chart did not give any useful data, as the numbers could be found on another table, where they all agreed it should be transformed into one useful visual, and not two impractical ones. IN2 started with saying that the first thing they would change is the background color from purple to white. Other than that, the biggest issue was the wrong choice of visuals, and pointlessness of the pie chart. The only thing that IN3 added to the changes of the bad dashboard was the placement, where they stated that the filters should be moved together, and the amount of color should be reduced. They also stated that it might be good to split the dashboard into multiple ones, for each category that the person using it might find beneficial.

When it came to the second group, the interviewees were not as confident with the answers, but tried to list a couple of things that they thought would make it easier for them to read the data. IN5 and IN6 both agreed that the matrix should be drilled up when first opening the dashboard, and that there should be even less color used in the graphs, at least the country bar charts. IN4 stated that the addition of a middle value in the conditional formatting would be good, but other than that, the dashboard and the visual selection did not receive any additional change requests by the second interviewee group.

5.1.7 Comparison to the Second Dashboard

After completing the analysis and discussion about the first dashboard the interviewees were presented with, they were shown the other one as well. The group that firstly saw the bad dashboard had a very happy and pleasurable first reaction when they saw the second dashboard. The main thing that all three interviewees talked about is that the matrix could have been drilled up at first, as it makes it look a bit crammed with data at first sight. After taking a moment to look around the dashboard, IN1 stated that they like the colors as they seem toned down and soft, compared to the bad dashboard. They also said that “it is much easier to get information from the dashboard, for example, the top profit country is clearly visible right away”. IN2 and IN3 commented on the filter positions, and said they found it “nice that they are grouped together”. IN3 remarked that the second dashboard “gives me the information I was missing beforehand”. They also said that the top recap text is a key element, that gives a general overview and should be used more often. Generally, all three liked the colors a lot more, because they were easier to look at and made them feel calm. IN3 also added that the second dashboard is easier to read and follows a structure that gives you a sense of what is going on.

The initial reaction of the second group was a lot different. All three interviewees were startled at first. IN4 started with “oh dear” when they first saw the dashboard. When asked what they thought about the second dashboard, all three stated that they are taken aback by the purple. They were also amazed how differently the exact same data can be shown. IN4 stated that the first thing they noticed was how little added value the pie chart brings because of the lack of values. IN5 said that everything on the dashboard seems “messy and unclear. Every graph or visual selection is wrong”. They also commented that it looks like someone “tried to make a bad dashboard intentionally”. Both IN5 and IN6 agreed that the tree map and pie chart hold zero value. IN6 commented that even though the dashboard is not suitable for everyday use, it gives them a feeling of being more organized because he can see the border lines of the graphs. Afterwards, they also said that nothing is actually visible or useful on the dashboard, and the lack of sorting is painful to look at. Generally, all three interviewees agreed that the bad dashboard example is a horrid way of representing the data. When asked if their opinion on the first one changed after they saw the second one, all three also concurred that the good dashboard example now seems like a great one, even though they found some mistakes or issues with it at first.

Lastly, the general impression interviewees had with the bad dashboard example was that "someone tried to show a lot of data on one page but did not make any sense whilst doing that", while the good dashboard impressions were along the lines of "well-chosen visuals according to the data used".

5.2 Analysis

As a general overview for easier comparison, the summary of findings from the interviews is gathered in the table 2.

Table 2: Summary of Analysis

	Bad dashboard example (IN1-3)	Good dashboard example (IN4-6)
Initial Reaction	Stunned for a couple of seconds. Expressed disgust with the overuse of purple and the general color selection.	Almost no reaction. Generally pleasing, a bit cluttered with information because of the amount of data in the matrix.
First Impressions	First thing they saw was either the pie chart or the top profit country table. First piece of information they gathered was which country has the highest profit. Attention flowed from graphs to tables.	First thing they saw was either profit by country or the matrix. First piece of information they gathered was the total profit card visual or the top profit country. Attention flowed from graphs to tables.

(Table continues)

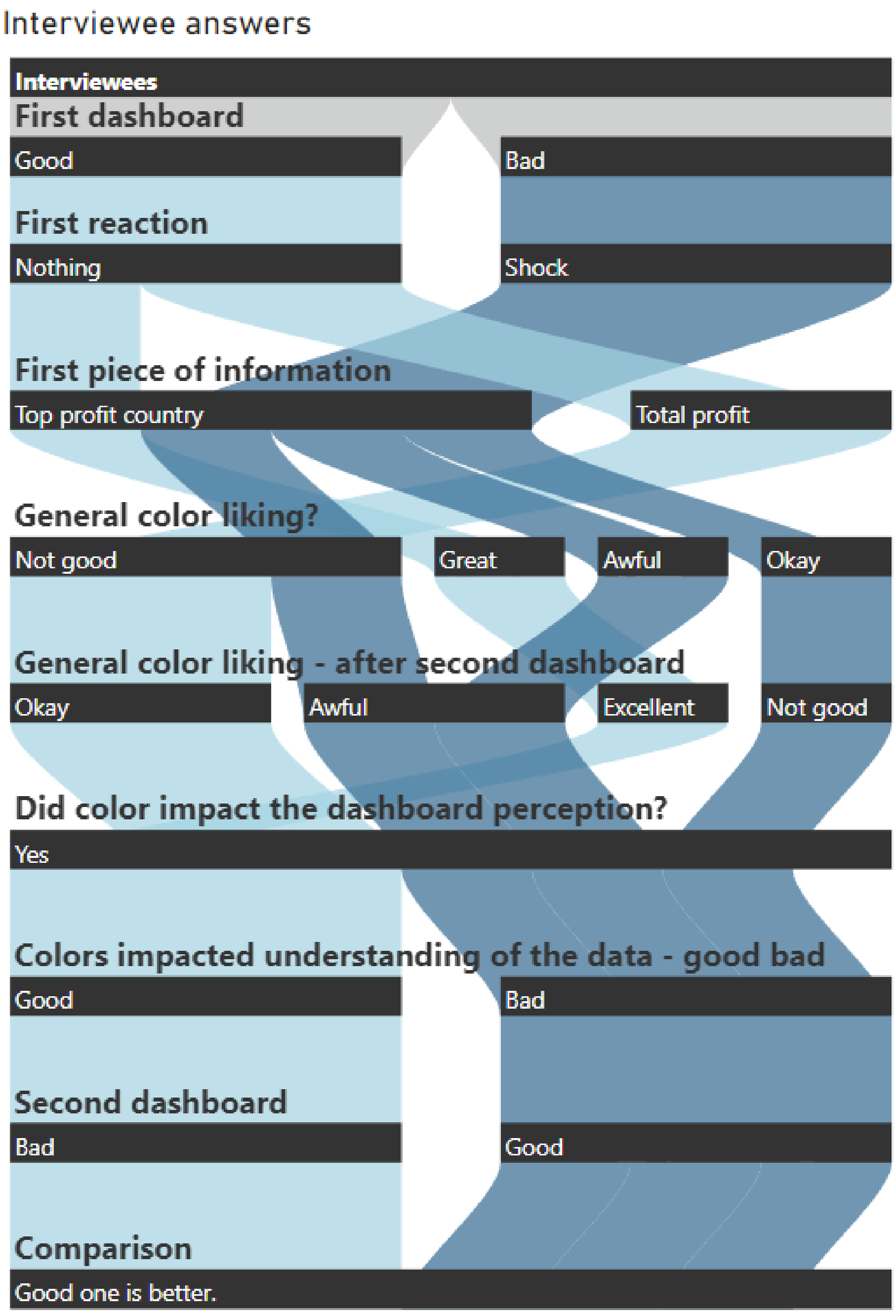
Table 2: Summary of Analysis (continued)

	Bad dashboard example (IN1-3)	Good dashboard example (IN4-6)
Data and Data Quality	There were issues with finding the data, as they were drawn to graphs with no value labels first. Issues with data being unsorted. Finding data was not very hard, but it took some adjustments. Easy to find the name of the largest segment by profit but had a harder time finding the numbers behind. Highest-profit month was easy to find due to the conditional formatting.	Finding data for highest profit country and segment were very easy and straightforward. Had some issues with identifying the top profit month, because the matrix had to be scrolled through. After it was drilled up, the data was more easily found.
Use of Color	Main thing they all disliked is the purple theme – too much purple at once. Hardly readable, colors do not imply meaning. Too many colors used. Colors are also too similar to each other. Visuals shifted their attention and impacted their understanding for the worse.	Nice and toned-down colors. Some personal color distaste, otherwise, they are not too much or badly picked. No strong opinion on the choice of color. Visuals shifted their attention and impacted their understanding, but they agreed it was a good impact, which made the dashboard easily readable – emphasis on the conditional formatting in the matrix.
Corrections for the Dashboard	Change of the visualization types according to which type of data we are trying to show. Less data shown in the line chart. Color theme and palette changes are critical. Some placement changes could also be beneficial.	Drilled up matrix, even less color used in the graphs. Addition of a middle value conditional formatting for easier data comprehension.
Comparison to the Second Dashboard	Pleasurable reaction to the second dashboard. Matrix should be drilled up at first. Likeness of the color palette, calm and nice to look at. Easier to gather information from. Has more key information and it is presented in a better way.	Startled at first, when they saw the bad dashboard example. The purple theme is too much. Wrong visualization types for the data shown. No clear story, not suitable for everyday use. Unpleasant way of data representation.

Source: Own work.

Afterwards, I also built a parallel sets chart, which can be seen in figure 20, that shows how the interviewee's answers went, and how they differed according to the dashboard they were looking at. Even though the interviewees answers differed during the interview, the end answer was the same with all. No matter which dashboard they saw first, they all agreed that the *Good dashboard* was better.

Figure 20: Parallel Sets Chart



Source: Own work.

6 DISCUSSION

Through the interview process we are able to gain valuable insight into how color and composition affect data quality gathered from business intelligence dashboards. The following chapter will be organized around topics that came up naturally in the conversation, while being backed with literature and subjective interpretation.

6.1 Reactions

The interviewees were all exposed to the dashboards under the same circumstances, having never seeing them before, while also not knowing what the main research topic of my thesis was. Even though they all knew they were about to see financial data for a company, the initial reactions differed. Interviewees that were shocked after seeing the bad dashboard example all showed signs of experiencing unpleasant surprise based on the amount of purple (or color in general). Likewise, when the second group saw the bad dashboard example for the first time, they all had a very similar reaction to it. Based on the initial reaction and debate insights I gathered from the interviewees afterwards, it is safe to say that the color and the amount of it was the main thing that provoked that reaction. The color of the background was the main thing their eyes were drawn to, and it was also the first thing they would all change if they were given the chance of fixing the dashboard. Equally, they all agreed that the color selection was what made the dashboard less readable and the data harder to understand. In that context, this applies to representation DQ, which means that the data was not presented in a way that made it easy to interpret and understand (Wang & Strong, 1996).

In contrast to the bad dashboard example, with the interviewees who initially saw the good dashboard example, it was more so a matter of personal preference whether or not they liked the color palette chosen. No interviewee stated that there was too much color used or that it hurt their eyes. The main comment was that they would have chosen some different colors, because of their taste. However, the comments the second interviewee group made were mostly in reference to the number of hues chosen, as IN5 and IN6 both believed that there should be a limit, at least when it came to the bar chart showing profit by country. When the first group saw the good dashboard example, none of them had any complaints or comments on the color selection. They all agreed that the colors were toned down and pleasurable to look at, even for a longer period of time. What they all emphasized more than once was that the white background is the best thing you can use in any dashboard, as it makes it much better for daily use, and it strains your eyes less.

Both groups of interviewees, no matter which dashboard we were discussing, agreed that the choice of color greatly impacted their attention. They also concurred that they all firstly shifted their attention towards bigger, more colorful graphs, and later to the tables and text fields. No matter how the page was structured and where on the dashboard the visualizations were placed, their attention always went from graphs to charts to tables and text. However, interviewees from the first group stated that the color impacted their understanding of the data in a negative way, because it made information harder to find or obscured it in a way. While the second interviewee group also said it impacted their understanding, they were all in unison when answering that it made the experience better or data easier to find, which was mostly accredited to the conditional formatting inside the matrix.

When comparing the color comments from both groups and both dashboards, the main takeaway is that the color intensity and overuse of it is what bothers them all, regardless of their BI experience, yet the hue selection is a personal choice, and not something that impacts their liking of the dashboard or ability to take away information from it. Additionally, the choice of color did not matter when it came to their attention shifting, because they all automatically went from graphs to tables. Although the color choice did not affect their attention shifts in different ways, the search for information was more than clearly affected by it. As even the least skilled BI users stated, the color choice in the bad dashboard example made the information hard to find, while finding it in the good one was easy as the color was used as a helpful tool and not only as decoration (Microsoft, 2015; Few, 2008).

Secondly, the interviewees were asked about data quality, and their ability to find information from the dashboards. As predicted during the design process of the bad dashboard, the main issue they found with the visualization types were the lack of data labels, or the general messiness of the data. Every member of the first group noticed how little value the pie chart added, as it only showed country names with no data labels. Even though you can hover over the slices to see the numbers, they all said that the profit values by country were too close together for you to get any additional information from it, and it could be omitted fully and replaced with only a table or bar chart (Nussbaumer Knaflic, 2015; Tufte & Schmieg, The Visual Display of Quantitative Information, 1985).

When we moved on to gathering data from the dashboards, the differences started showing between both dashboards and their quality. While the interviewees that first saw the good dashboard had little to no issues finding data for profit by country or segment leaders, the first group struggled. Interestingly, the second group had bigger issues finding the month and year with the largest profit, as they relied on the matrix at first and not the bar chart which showed it by line and label, while the first group used the conditional formatting and found the answer rather quickly. The answers I got from the interviewees show that there were representational and contextual data quality issues in both dashboards, as they could not find answers to my questions quickly or answered them wrongly (Wang & Strong, 1996). However, the second group of interviewees found a quick way of fixing the search speed, as they drilled up the matrix, and found the correct answers rather fast.

In this part of the interview, the interviewees answers differed because of the visualization types used, the labelling of units on them, and the placement of the visuals on the dashboard. While the good dashboard example was much easier to use, and used conditional formatting as a helpful tool, having too much data drilled down and expanded too far proved to be more of an issue than a feature providing informational value. A very big factor, which the interviewees pointed out, was that the sorting of data was very helpful. Only in one case was it detrimental for finding information. Even though the bad dashboard example had countries with profits in a table, they found it relatively useless, as they had to sort it by profit to find the top one, while them being sorted alphabetically was an inconvenience. Next to the table, all the interviewees agreed that the area charts (pie chart, tree map) should be omitted and replaced with something else. While the pie chart can be useful to show data, the bar chart that replaced it in the good dashboard example was clearly the better option, which helped the second group find the answers straight away.

If we compare the commentary and observations from both groups again, we can see that neither dashboard was of good data quality, but the good dashboard example still proved to be more useful and easier to work with. When presenting data, all six agreed that the visualizations have to be chosen correctly according to the type of data we are using. Additionally, it should also be ordered by a number that gives us maximum insight value (Microsoft, 2015; Nussbaumer Knaflic, 2015).

One of the points where the answers of the interviewees differed the most was in correcting the dashboards according to what they thought would make them better. While the second group only commented on using less colors in the center bar chart and drilling up the matrix, the first group had a variety of comments. They covered positioning, color selection, and change of visuals according to which type of data they are showing. The crucial thing to all three was the change of the purple background to white, and the replacing of pie and tree maps. One additional thing they all pointed out was too much information in one visual (line chart), which decreases its value and readability. They brought up color in regard to the background as well as the specific colors used in the graphs, which made it harder for information to be gathered from them. Nevertheless, one interviewee from each group made a comment about the dashboard needing to tell more of a story, as they seem "a bit messy". This later changed for the person that first saw the good one, and the bad one afterwards, as it changed their perspective as to how usable the good dashboard is.

An interesting observation came from the difference in the interviewees' backgrounds or experience in other aspect that are not directly connected to BI. One of the interviewees that had a degree in economics seemed to use the general knowledge they gathered with data insights and drew up similar conclusions to the user that had the most experience with BI, as they also had an economic background, but in regard to their work experience. Another interviewee that did not have a lot of experience in BI tried to tie findings and insights to what they knew from dashboards they might have seen in other contexts, that had nothing to do with sales data. In this regard, I believe that the users background experience, either work or studies, can have an impact in data comprehensions and understanding, even if they are inexperienced in BI.

Lastly, the interviewees were asked to comment on the comparison between both dashboards. Before they had a chance to debate however, the interviewees from the second group showed shock and repulsion after being confronted with the bad dashboard, which is the same reaction the first group had as well. This points to the fact that no matter when they saw the bad example, the reaction was the same, it yet again revolved around the use of purple, as it is a lot of information to process with no usable data gotten from the color itself (Kashyap, 2020).

Similarly to the second group's answers, the first one commented on how the matrix should have been drilled up at first, as it might be presented as too much information at first. On the other hand, all of the interviewees in the first group said they liked the colors in the second one and commended the color choices for all the visuals and conditional formatting. The second group also pointed out how the bad dashboard example used the wrong types of visuals and did not tell a story in any way.

6.2 Implications for the Theory

Through the process of interviewing and observation, we can see the correlation between the theoretical models and practical responses of the interviewees. From all aspects – color, design, and data quality – the interviewees reacted in a mostly predictable way.

When it comes to color, we can firstly point out Few's design principles. Where his theory underlines that color must be used meaningfully and with restraint, the bad dashboard example clearly ignores that. And even though the inexperienced BI users have never seen Few's work, they all pointed out how to color usage was a distraction and made the readability bad instead of being a helpful tool, like it should be used as said in *Storytelling with Data* (Nussbaumer Knaflic, 2015). Nevertheless, in the good dashboard example, the color scheme was soft and composed of mostly natural colors, which makes people less tense when reading through the data. Another thing that Few emphasized and was used well was the non-data components either being invisible or barely visible, which was true in the good dashboard example (Few, 2008).

Secondly, we can imply the usage of color through the interviewee's comments based on the color theory being used in data visualization. Most literature implied that the usage of different colors or hues should be as minimal as possible. In the bad dashboard example, the colors were not harmonious, and there were a lot of hues used – both in the background as well as the visuals, which made users more confused and the data harder to understand. On the other hand, the color palette in the good dashboard example was minimal, and used saturation or lightness instead of hues, which made it more clear and easily readable, according to the users. Nonetheless, the good dashboard example should have used even fewer hues when it came to the bar chart of the profit by countries (RevUnit, 2022; Yi, 2019).

When it comes to design, we can find implications to the Gestalt theory and Tufte's design principles. Tufte set his rules of graphical integrity, and one of them (the second rule that states that clear, detailed labeling should be used to defeat graphical distortion and ambiguity) was more than clearly breached in the bad dashboard example, which the users also commented that it made the dashboard unusable. Meanwhile, the good dashboard example used labels where necessary, and used the appropriate amount of text so that the user understood the context without spending a large amount of time trying to decipher the meaning (Tufte, The Visual Display of Quantitative Information, 2001; Ferrara, 2017). Gestalt was used in a number of ways in both dashboards. One of the interviewees pointed out that he firstly thought that the bad dashboard was better planned out, because the visualizations had boxes around them, but later on discovered that that might not be the case. This overlaps with the Gestalt principle of common region, where your brain perceives things to be organized together, because of the box around them. On the other hand, another interviewee said that the color choices for segments and countries in the bad dashboard made it seem like they were showing the same thing, because of the same palette usage – the similarity principle of gestalt (Ahmadi, 2020; Chapman, 2017). Regarding the design of the dashboards, they were built as strategic dashboards, which provide an overview of the general numbers, where we could see that too much information in the drilled down matrix made it seem too cluttered. When building strategic dashboards, limiting the amount of information is critical (Few, 2006).

Lastly, we can see the implication to the data quality framework throughout both dashboards. The most important and utilized ones in the dashboards were contextual and representational data quality. While both dashboards were presenting the same data set, the value-added and the amount of data used was immensely better introduced in the good dashboard example. As well as the contextual DQ, the ease of understanding, and representational consistency was a big factor of success for the good dashboard, in comparison to the bad one (Wang & Strong, 1996).

Although my findings coincide with multiple theories over various sections, they would benefit from further qualitative research and an ongoing improvement cycle for the dashboards. Within the scope of research, this thesis contributes towards the understanding and better utilization of color and design in data visualization.

6.3 Practical Implications

Firstly, according to the findings of this paper, the selection of color should be minimized to a small number of hues and utilize saturation or luminosity in order to differentiate data in selected visuals. However, the selection of hues relies heavily on the user – or the designer's – preference and is not conditioned or limited to a certain spectrum of color that should be used. As mentioned in the literature review, the main restriction when choosing the hues is how saturated the color is, but we also have to keep in mind that the color palette should be inclusive while keeping in mind color blindness. As well as the selection of the color palette, I found that all the users prefer having white or a neutral light color as the background, because it does not take the focus away from the data in any way, but lets the visuals be the center point of the dashboard.

Next, the placement of the elements or the design of the dashboards has to be taken into consideration before building the report. People tend to group things together if they are alike (Chapman, 2017), which makes it harder for people to focus if the similar blocks are not placed together. Considering the gestalt principles, grouping similar things together makes users perceive them as a whole, so the placement of filters together or KPI cards together is also one of the more important factors when building the dashboards. When it comes to storytelling, using it in a general dashboard may be a hard task, but the good dashboard example utilized a dynamic text box for a general overview of the data set, which the users found useful, but at the same time it surprised them, as it is not something they see frequently (Nussbaumer Knaflic, 2015).

In context of DQ, the two most important aspects to consider when designing a dashboard are contextual and representational DQ, while also keeping track of intrinsic and accessibility DQ where applicable. Most importantly, value-added and the amount of data shown can either encourage or deter people from using a dashboard and making a rational decision based on the data it shows. The findings were shortly summarized blow.

Table 3: Summary of Findings

Use of Color	Minimized to a small number of hues, utilizing saturation or luminosity. Liking of color is a personal preference. Neutral or white color as the background, for lesser distraction.
Dashboard Design	Grouping similar items together (example: visuals with the same data, filters, KPI cards).
Visuals	Choosing the right visual according to the data is a key element. Users can find data faster and understand it better.
Storytelling	Can employ the use of dynamic text for additional data insights.
Data Quality	Consider contextual and representational data quality. Value-added and the amount of data on the dashboard are critical.
General findings	For a visual to be effective, it has to have a clear design strategy and color palette, as well as using the correct data and type of graphs.

Source: Own work.

Through the short summary in table 3, one of the main findings is that the liking of color is mostly a personal preference, even though we can influence color comprehension through the utilization of color harmony principles (Kargin, 2022). Through the interview, we could see that at first the number of colors may not have been emphasized as the main issue in the visuals. Nevertheless, using a lesser number of hues was generally connected to better data understanding and even liking the dashboard more. Additionally, the use of a neutral or white background of the dashboard was universally welcomed. Even so, using data for anything other than better data comprehension seemed to not have a big effect on general liking of the dashboard. As far as dashboard design goes, the gestalt principle of grouping things together to be perceived as presenting the same information proved to be important for all the interviewees. A key part of contextual and representational DQ was the selection of the correct visual type, no matter the BI experience of the interviewee. Next to correct visual selection, the visual also has to be simple enough for the end user to understand. This, however, is only applicable to strategic or operational dashboards, as analytical tend to offer more complex visualizations for further data analysis. As far as BI capabilities of dynamic text and tooltips go, the interviewees agreed it added to data comprehension. In the context of color and composition having effect on dashboard quality, we could omit intrinsic and accessibility DQ, as they could not really be applied in the context of this research.

Table 4: Impact of Interview Topics on Likeness and Data Comprehension

		Impact on data quality dimensions	
Question set	*Specific topic*	Contextual DQ	Representational DQ
Use of color	Small number of hues	Low	High
	Likeness of color	Medium	Low
	Neutral colored background	High	High
	Color for comprehension not beauty	Low	Medium
Dashboard design	Grouping similar items	High	High
	Utilizing text or icons	Medium	High
Visuals	Correct visual type selection	High	High
	Form follows function	Low	Medium
	Minimizing color where unnecessary	Low	High
Storytelling	Using dynamic text	Low	High
	Using tooltips	None	High

Source: Own work.

In table 4, the darker the blue color is, the bigger the effect the topic has on DQ.

CONCLUSION

There is an ever-growing number of organizations who use BI and its visualizations for decision making, but there is also a certain percentage of visualizations that make us misinterpret the data or give us close to zero added value. In order for a company to make successful data-driven decisions, the basis for our decisions has to be clear and understandable. For a visualization to be effective, it has to have a clear design strategy and color palette, as well as using the correct data and type of graph to interpret the aforementioned data set. In order for us to achieve that, there is a number of factors based on which the decision should be made. Generally, most papers used in the literature review try to define guidelines or a set of rules which emphasize color, design, the correct visual type regarding data, or a summary of those. However, strict rules on what to watch out for when building dashboards are somewhat infrequent.

In regard to the theoretical implications of the thesis, the visualization, design, and color requirements for dashboards that were defined at the end of the theoretical part were partially confirmed. When talking about visualization requirements, one of the main things that was frequently emphasized in literature, was the importance of the correct selection of the visual type in regard to the data it is conveying (Microsoft, 2021). While additional text and labels were not distracting for the users, they did not seem to be a highly valuable addition in contextual DQ, as it was what users noticed last in the dashboard. In the relatively short period of time, they had to explore the dashboard the text helped with quick explanations. If they were to use the dashboard longer, it might be more valuable, but in this setting, users tried to explain the data contexts to themselves. Conditional formatting did prove beneficial, but only when used in color, as icons such as arrows needed more context to be explained well (Dougherty & Ilyankou, 2021). Lastly, when talking about color, the implication of using less hue and more saturation and luminosity proved to be key in data comprehension, which confirmed all the underlying theories of using color only for data comprehension and not beautification (Berinato, 2019; Few, 2008; Tufte, The Visual Display of Quantitative Information, 2001). The main thing that was only partially confirmed is that the color selection would either be liked or disliked by all the interviewees, as it showed to be a very subjective decision. In this context, it is important to select colors that work harmoniously together, while the primary or main hue selection is the designer's choice. The main thing to consider in color selection is only that certain colors are perceived as good or bad (green/red) because of context we see daily (Mohr & Jonauskaite, 2022).

The practical implications for this paper stem from the findings of which aspects influence DQ in dashboards the most. As the most influential aspects on DQ are the neutral-colored background and a small number of hues, the main suggestion when building dashboards in the future would be to choose a main hue, around which a harmonious color palette should be built. We should also consider the limitations of the emotional effect and visual disabilities (Kargin, 2022). The majority of color used on the dashboard should be variations of a single hue with differentiation coming from lightness or saturation, while the main hue should not have high saturation to begin with. In the context of data, the correct visual selection is key to keeping data quality high. Having a visual hierarchy and grouping things together will also ensure that the user does not get lost in the dashboard. Another important influence was the utilization of text and icons, which had a high representational DQ impact, as data explanation through various additional tools proved to help with data comprehension. Lastly, telling a story should always be crucial in any dashboard creation considering representational DQ, based on which any business decisions could be made in the future.

The purpose of this paper was to show the influence of color theory and composition when working on visualizations and dashboards in business intelligence development. Additionally, we tried to identify how color and composition influence data comprehension while trying to come up with a set of guidelines. This was achieved through a critical literature review and qualitative research, which lead to results summarized in table 4. These results in turn represent a set of guides to follow when building up the visualizations. The main focus of the paper however dealt with visualizations and color theory in the context of dashboards, as a big-picture overview of a large set of information. As the dashboard combines a set of visuals, it should also combine a set of rules behind them, to increase contextual and representational DQ. This leads to higher DQ and better data-driven business decisions that omit bias and mistakes due to the wrong visual or color selection.

The goal of this research was to understand the concepts behind color theory, dashboards, visualizations, and composition. The goal of the empirical part was to understand how individuals perceive the quality and information in different dashboards. The overall goal was to deliver a set of guidelines to follow when we are building a dashboard in BI. All three goals were reached, as the literature review played a key role in setting up the foundation, on top of which the empirical research was built. Through qualitative exploration intertwined with the theoretical part, the results were able to be summarized as key influences, shown in table 4, which can be a starting point when building dashboards in any BI visualization tool.

The limitations of this paper mostly encompassed the size of the sample and the depth of the research. The study could be expanded with iterations of the dashboard, through which we could explore additional aspects and record the reaction people have to different dashboard compositions. In further research, the contents of this paper could also be divided into separate studies, with each focusing on a separate field – color, composition, data quality, and storytelling. In conclusion, there are a number of factors to consider when building a dashboard, but what most people forget about is that next to the accurate visual selection, data and placement have a strong effect on data understanding and retention. Through the use of color, design, and storytelling theories, people can vastly increase the quality of their dashboards, which can in turn lead to better decision making, and faster issue recognition, based purely on data being presented in a correct way.

The trends in data visualization are continuing to evolve daily. From inclusion to better and faster visualization options, it is crucial for data visualization experts to keep track of the changes and try to improve on every part of their work. As dashboards usually present the key link between businessmen and data, they have to have high DQ while showing a big-picture overview. Even though this may be harder to achieve, following a set of rules on what effects DQ in dashboards the most can result in better data-driven business decisions. In a small world of big data this is what can give you an edge over the competition.

REFERENCE LIST

1. Adair, B. (2020). *The Best Data Visualization Tools*. Retrieved October 19, 2021, from SelectHub: https://www.selecthub.com/business-intelligence/data-visualization-tools/

2. Ahmadi, N. (2020, June 17). *Gestalt Principles in UX Design*. Retrieved January 21, 2022, from UX Planet: https://uxplanet.org/gestalt-principles-in-ux-design-2e0f423bfcb5

3. Annesley, T. M. (2020, September). Bars and Pies Make Better Desserts than Figures. *Clinical Chemistry, 56*(9), 1394–1400. Retrieved May 14, 2022

4. Anouncia, S. M., Gohel, H. A., & Vairamuthu, S. (2020). *Data Visualization* (1 ed.). Singapore: Springer Singapore. doi:https://doi-org.nukweb.nuk.uni-lj.si/10.1007/978-981-15-2282-6

5. Bartram, L., Patra, A., & Stone, M. (2017). Affective Color in Visualization. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1364–1374. Retrieved October 5, 2021, from https://doi.org/10.1145/3025453.3026041

6. Bauer, S. (2022, January 9). *Continuously Improving User Experience*. Retrieved February 3, 2022, from PowerBI.tips: https://powerbi.tips/2022/01/continuously-improving-user-experience/

7. Bell, E., & Bryman, A. (2011). *Business Research Methods.* Oxford: Oxford University Press.

8. Berinato, S. (2019). *Good charts workbook: Tips, tools, and exercises for making better data visualizations.* Harvard Business Review Press. Retrieved August 12, 2022

9. Betzendahl, L. (2020, July 6). *Utilizing Gestalt Principles to Improve Your Data Visualization Design*. Retrieved January 4, 2022, from Viz Zen Data: https://vizzendata.com/2020/07/06/utilizing-gestalt-principles-to-improve-your-data-visualization-design/

10. Bruner, J. S., Postman, L., & Rodrigues, J. (1951). Expectation and the Perception of Color. *The American Journal of Psychology*, 216–227. Retrieved October 16, 2021, from https://doi.org/10.2307/1418668

11. Brush, K. (2020, February). *Data visualization*. Retrieved June 21, 2022, from TechTarget: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

12. Burnay, C., Bouraga, S., & Faulkner, S. (2020). User-Experience in Business Intelligence - A Quality Construct and Model to Design Supportive BI Dashboards. *Research Challenges in Information Science*, 174-190. doi:10.1007/978-3-030-50316-1_11

13. Careri, W. (2022, February 8). *Designing for Neurodivergent Audiences*. Retrieved May 14, 2022, from Nightingale: https://nightingaledvs.com/designing-for-neurodivergent-audiences/

14. Chapman, C. (2017). *Eploring the Gestalt Priciples of Design*. Retrieved October 19, 2021, from Toptal: https://www.toptal.com/designers/ui/gestalt-principles-of-design

15. Clinton Eye Assoiates. (n.d.). *Color Blindness*. Retrieved April 2, 2022, from Clinton Eye Assoiates: https://www.clintoneye.com/color-blindness.html

16. Crinson, I., & Leontowitsch, M. (2006). *Semi-structured, narrative, and in-depth interviewing, focus groups, action research, participant observation*. Retrieved April 26, 2022, from HealthKnowledge: https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1d-qualitative-methods/section2-theoretical-methodological-issues-research

17. Decker, K. (2016). *The fundamentals of understanding color theory*. Retrieved October 19, 2021, from 99designs: https://99designs.com/blog/tips/the-7-step-guide-to-understanding-color-theory/

18. Dougherty, J., & Ilyankou, I. (2021). *Hands-On Data Visualization: Interactive Storytelling From Spreadsheets to Code* (1st ed.). O'Reilly Media. Retrieved August 10, 2022

19. English Club. (n.d.). *Colour Idioms*. Retrieved April 2, 2022, from English Club: https://www.englishclub.com/vocabulary/idioms-colour.htm

20. Evergreen, S., & Metzner, C. (2013, December). Design Principles for Data Visualization in Evaluation. *New Directions for Evaluation*, 140. Retrieved January 21, 2022

21. Ferrara, E. (2017, April 22). *Edward Tufte's Principles of Graphical Integrity*. Retrieved January 12, 2022, from Fox School of Business: https://community.mis.temple.edu/mis5208sp18/2017/04/22/edward-tufte-s-principles-of-graphical-integrity/

22. Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data.* O'Reilly. Retrieved 2022

23. Few, S. (2008, February). *Practical Rules for Using Color in Charts.* Retrieved from Perceptual Edge: https://nbisweden.github.io/Rcourse/files/rules_for_using_color.pdf

24. Few, S. (2011, June). The Chartjunk Debate. *Perceptual Edge*. Retrieved August 3, 2022, from http://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

25. Few, S. (2012, November 8). *Telling Compelling Stories with Numbers*. Retrieved January 21, 2022, from Analytics: https://analyticsconsultores.com.mx/wp-content/uploads/2019/03/Telling-Compelling-Stories-with-Numbers-Stephen-Few-Perceptual-Edge-2012.pdf

26. IBCS. (2022, January 28). *IBCS Standards 1.2*. Retrieved September 4, 2022, from IBCS: https://www.ibcs.com/ibcs-standards-1-2/

27. Joshi, A. (2018). *Design Principles.* Retrieved January 21, 2022, from Department of Compuer Science - University of Huston: http://www2.cs.uh.edu/~ceick/UDM/COSC3337-DV2.pdf

28. Kargin, K. (2022, February 14). *Color Harmony in Data Visualization*. Retrieved August 16, 2022, from Medium: https://medium.com/global-maksimum-data-information-technologies/color-harmony-in-data-visualization-2610f9dab390

29. Kashyap, A. (2020, December 28). *8 Rules for optimal use of color in data visualization*. Retrieved April 26, 2022, from Towards Data Science: https://towardsdatascience.com/8-rules-for-optimal-use-of-color-in-data-visualization-b283ae1fc1e2

30. Kosara, R. (2014, July 23). *Putting Data Into Context*. Retrieved April 2, 2022, from https://eagereyes.org/basics/putting-data-into-context

31. Lavalle, A., Maté, A., Trujillo, J., & Rizzi, S. (2019). Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework. *2019 IEEE 27th International Requirements Engineering Conference (RE)*, 109-119. doi:10.1109/RE.2019.00022

32. Makulec, A. (2022, August 2). *Starting Out in Data Visualization Today*. Retrieved August 4, 2022, from Nightingale: https://nightingaledvs.com/starting-out-in-data-visualization-today/

33. Mark, N. L. (2022, June 21). *Seven principles of UX design to improve your business data visualizations*. Retrieved July 22, 2022, from Medium: https://medium.com/select-from-data/seven-principles-of-ux-design-to-improve-your-business-data-visualizations-part-2-b4c94754fa09

34. Microsoft. (2015, September 29). *The Art and Science of Effective Dashboard Design*. Retrieved February 4, 2022, from Microsoft Power BI Blog: https://powerbi.microsoft.com/en-us/blog/the-art-and-science-of-effective-dashboard-design/

35. Microsoft. (2021, September 9). *Tips to optimize visual colors in Power BI reports*. Retrieved April 2, 2022, from Microsoft PowerBI: https://docs.microsoft.com/en-us/power-bi/guidance/report-tips-visual-colors

36. Microsoft. (2021, November 9). *Visualization types in Power BI*. Retrieved January 21, 2022, from Microsoft documentation: https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a#funnel-charts

37. Microsoft. (n.d.). *What is business intelligence?* Retrieved September 2, 2022, from Microsoft: https://powerbi.microsoft.com/en-us/what-is-business-intelligence/

38. Midway, S. R. (2020). Principles of Effective Data Visualization. *Patterns, 1*(9). doi:https://doi.org/10.1016/j.patter.2020.100141.

39. Miglietti, C. (2022, February 3). *Effective Data Visualizations Should Focus on Narrative, Not Numbers*. Retrieved March 4, 2022, from Nightingale: https://nightingaledvs.com/effective-data-visualizations-should-focus-on-narrative-not-numbers/

40. Mohr, C., & Jonauskaite, D. (2022, February 8). *Why Links Between Colors and Emotions May Be Universal*. Retrieved April 2, 2022, from Psychology Today: https://www.psychologytoday.com/us/blog/color-psychology/202202/why-links-between-colors-and-emotions-may-be-universal

41. Muth, L. C. (2020, November 27). *What to consider when considering data vis rules*. Retrieved August 10, 2022, from Lisa Charolette Muth: https://lisacharlottemuth.com/datavisrules

42. Nussbaumer Knaflic, C. (2015). *Storytelling with data.* Hoboken, New Jersey: John Wiley & Sons, Inc.

43. Pyramid Analytics. (2021). *Why Pyramid*. Retrieved October 19, 2021, from Pyramid Analytics: https://www.pyramidanalytics.com/

44. Ranawaka, M. (27. July 2021). *ETL vs ELT: The Difference is in the How*. Pridobljeno 19. October 2021 iz Panoply Blog: https://blog.panoply.io/etl-vs-elt-the-difference-is-in-the-how

45. RevUnit. (2022, March 29). *The Role of Color Theory in Data Visualization*. Retrieved April 2, 2022, from RevUnit: https://www.revunit.com/post/the-role-of-color-theory-in-data-visualization

46. Robinson, A. C. (2020). *Choosing Colors*. Retrieved April 26, 2022, from PennState University: https://www.e-education.psu.edu/maps/l5_p5.html

47. Rogowitz, B., Treinish, L., & Bryson, S. (1996, May). How Not to Lie with Visualization. *Computers in Physics*, 268-273. doi:https://doi.org/10.1063/1.4822401

48. Runkler, T. A. (2012). Data Visualization. In T. A. Runkler, *Data Analytics* (pp. 35-54). Wiesbaden: Vieweg+Teubner Verlag. Retrieved June 22, 2022

49. Scardina, J. (2018, December). *Microsoft Power BI*. Retrieved October 19, 2021, from TechTarget: https://searchcontentmanagement.techtarget.com/definition/Microsoft-Power-BI

50. Schwabish, J. A. (2014). An Economist's Guide to Visualizing Data. *The Journal of Economic Perspectives, 28(1),* 209–233. Retrieved October 5, 2021, from http://www.jstor.org/stable/43193723

51. Sigdel, R. (2020, September 10). *Improve Your Visualization Skills Using Tufte's Principles of Graphical Design*. Retrieved January 10, 2022, from Nightingale: https://medium.com/nightingale/improve-your-visualization-skills-using-tuftes-principles-of-graphical-design-3a0f40a53a2c

52. Solis, J. (2019). Data Visualization Is King. *The Journal of Private Equity, 22*(3), 102–107. Retrieved June 22, 2022, from https://www.jstor.org/stable/26864425

53. Stone, M. (2006, Jauary 17). *Choosing Colors for Data Visualization.* Retrieved May 12, 2022, from Perceptual Edge: https://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf

54. Tableau. (2021). *What You Need To Know About Business Intelligence (BI) Dashboards*. Retrieved January 21, 2022, from Tableau: https://www.tableau.com/learn/articles/business-intelligence/bi-dashboards#:~:text=Business%20intelligence%20dashboards%20are%20information,single%20screen%20for%20snapshot%20overviews.

55. Tableau Software, LLC. (2021). *Business Intelligence: What It Is & Its Importance*. Retrieved October 19, 2021, from Tableau: https://www.tableau.com/learn/articles/business-intelligence

56. Tackels, D. (n.d.). *7 Best Practices for Using Color in Data Visualizations*. Retrieved April 26, 2022, from Sigma Computing: https://www.sigmacomputing.com/blog/7-best-practices-for-using-color-in-data-visualizations/

57. Tufte, E. (2001). *The Visual Display of Quantitative Information.* Cheshire: Graphics Press.

58. Tufte, E., & Schmieg, G. M. (1985, November). The Visual Display of Quantitative Information. *American Journal of Physics*, 1117-1118. Retrieved May 14, 2022

59. UserTesting. (2019, April 10). *7 Gestalt principles of visual perception: cognitive psychology for UX*. Retrieved April 2, 2022, from UserTesting: https://www.usertesting.com/blog/gestalt-principles

60. Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5-33. Retrieved June 8, 2022, from http://mitiq.mit.edu/Documents/Publications/TDQMpub/14_Beyond_Accuracy.pdf

61. Yi, M. (2019, Octiber 24). *How to Choose Colors for Your Data Visualizations*. Retrieved April 26, 2022, from Nightingale: https://medium.com/nightingale/how-to-choose-the-colors-for-your-data-visualizations-50b2557fa335

62. Zeller, S., & Rogers, D. (2020, May 21). *Visualizing Science: How Color Determines What We See*. Retrieved April 2, 2022, from EPS: https://eos.org/features/visualizing-science-how-color-determines-what-we-see

APPENDICES

Appendix 1: Povzetek (Summary in Slovene language)

Človeštvo trenutno živi v dobi množičnih podatkov, hkrati pa je to doba manjka informacij. Slednje je najbolj razvidno skozi kupe podatkov, ki jih zbiramo na dnevni ravni - a v pomanjkanju kadra, ki bi znal iz le-teh razbrati bistvo, in katerega bi lahko posamezniki ali podjetja uporabili za dobre poslovne odločitve. Razločitev med impulzi, ki nam lahko podajo osnovo za odločitev, in šumom podatkov, je težka, včasih skoraj nemogoča, a ravno zato se razvija tudi področje poslovne analitike in vizualizacij podatkov.

Vizualizacija podatkov je ključni dejavnik poslovne analitike. V dobi množičnih podatkov je pomembno, da smo do le-teh kritičnih, vendar bomo iz podatkov dobili pravilne informacije le takrat, ko bodo uporabljene pravilne vizualizacije. Oblikovanje prave vizualizacije je pogosto težavna naloga, uporaba napačne vizualizacije pa lahko podjetju povzroči znatne izgube.

Nadzorne plošče v poslovni analitiki so vizualizacijsko orodje, ki grupira posamezne vizualizacije podatkov na enem zaslonu. Nadzorne plošče lahko preprosto predstavljajo rezultat analize podatkov, kreatorju pa omogočajo izbiro skupkov podatkov, ki jih želi prikazati končnim uporabnikom.

Problem vizualizacij oziroma nadzornih plošč v poslovni analitiki se kaže v njihovi izgradnji. Poslovni analitiki so sicer vedno bolj vešči pri izbiri pravilnega tipa vizualizacije glede na podatke, vendar večina zanemari vidik barve ali kompozicije. Ob uporabi napačne barve, prevelike količine barv, ali pomanjkanja nepokritega prostora okoli vizualizacij, se lahko hitro izgubi berljivost in razumljivost sporočila, s tem pa tvegamo napačno interpretacijo podatkov. Čeprav so podatki osnova vsake nadzorne plošče, lahko zanemarjanje uporabniške izkušnje privede do napačnih interpretacij, kar lahko posledično povzroči zmanjšanje dobička za podjetje. Številke so osnova za sprejemanje odločitev, način njihove vizualizacije pa vpliva na človeško dojemanje in posledično odločitve.

Namen magistrskega dela je prikaz vpliva barvne teorije in kompozicije med kreiranjem vizualizacij in nadzornih plošč v razvoju poslovne analitike. Po začetnem kritičnem pregledu literature in empirični analizi, je bil namen ugotoviti, kako barva in kompozicija vplivata na razumevanje in interpretacijo vizualizacij pri posameznikih, in to razumevanje uporabiti za opredelitev smernic. Če ne sledimo naboru smernic, bo imel vsak model, ki ga zgradimo, enake težave in ovire glede berljivosti. Cilj magistrskega dela je opredeliti smernice in razložiti njihov vpliv na naše razumevanje vpliva barv in kompozicije na kvaliteto nadzornih plošč. Cilj je torej pomoč pri definiranju ustvarjanja vizualizacijskih nadzornih plošč, pri čemer je treba upoštevati njihovo berljivost in enostavno interpretacijo. Vse to bo imelo velik vpliv na sposobnost podjetij pri sprejemanju odločitev na podlagi informacij. Primarni cilj magistrskega dela je pregled teoretičnega dela literature za lažje razumevanje konceptov barvne teorije, nadzornih plošč, vizualizacij, in kompozicije. Cilj empiričnega dela je bil uporaba intervjujev in predpripravljenih nadzornih plošč za razumevanje posameznikovega zaznavanja kakovosti in informacij pri različnih nadzornih ploščah. V okviru magistrskega dela sem izvedla delno strukturirani intervju, povezan z eksperimentom, kjer sem šestim uporabnikom predstavila dve različni nadzorni plošči, ter opazovala njihovo reakcijo skozi uporabo. Ena izmed nadzornih plošč je sledila splošnim smernicam, ki so izhajale iz sinteze literature, druga pa je predstavljala napake, ki jih uporabniki ponavljajo pri izdelavi nadzornih plošč.

Iz eksperimenta izhajajo naslednje ključne ugotovitve. Pri kreiranju nadzornih plošč je pomembno, da se glede na tip podatkov izbere pravilna vrsta vizualizacije. Pogojno oblikovanje znotraj vizualizacij se je izkazalo za uporabno, vendar zgolj kadar je uporabljeno kot barvna shema in ne v obliki ikon. Slednje se je izkazalo za manj pregledno, saj uporabniki niso vedeli, s čim naj ikone primerjajo, oziroma kaj je njihova izhodiščna točka. Pri izbiri barv je ključnega pomena uporaba manj različnih vrst barv, saj se s tem izboljša preglednost in berljivost nadzorne plošče. Ena izmed ugotovitev pri izbiri barv je bila subjektivna izbira, saj se je všečnost barvne palete spreminjala od osebe do osebe, važno je bilo le, da je končna vizualizacija pregledna. S tega stališča se moramo pri kreiranju nadzornih plošč omejiti zgolj na to, da so barve, ki jih izberemo, harmonično povezane med seboj, hkrati pa ne smemo pozabiti na težave barvne slepote, z najbolj pogosto rdeče-zeleno barvno slepoto.

V praksi lahko vsak poslovni analitik, ki kreira nadzorne plošče, uporabi ugotovitve kot smernice glede na to, kateri dejavniki bolj ali manj vplivajo na kakovost podatkov na njej. Ključni dejavniki, ki so imeli največji vpliv na kakovost podatkov, so bili nevtralna barva ozadja, manjša količina različnih barv in manjša saturacija glavnih barv. Prav tako je izbira pravilne vrste vizualizacije kritična za razumljivost informacij. Poleg vizualne razporeditve in hierarhije vizualizacij lahko analitiki uporabijo tudi besedilo, s katerim lahko uporabnikom približajo razumevanje informacij.

Za bolj poglobljene ugotovitve in rezultate bi morali povečati vzorec izvedbe intervjujev oziroma eksperimentov, prav tako pa bi lahko študijo razdelili na iteracije posameznih nadzornih plošč. Vsak aspekt raziskave bi lahko raziskali posebej, torej barvo, postavitev in pripovedovanje zgodb skozi vizualizacije. Pri izdelavi nadzornih plošč obstaja več različnih dejavnikov, ki vplivajo na njeno kvaliteto. Veliko uporabnikov pozabi, da je poleg kvalitetnih podatkov pomembna tudi izbira barvne palete, kot tudi izbira tipa vizualizacije in postavitev le-te. Z uporabo barvne teorije, oblikovanja in pripovedovanja zgodb, lahko močno povečamo kakovost nadzornih plošč, kar lahko posledično privede do boljšega odločanja in hitrejšega prepoznavanja težav, ki temeljijo zgolj na pravilni predstavitvi podatkov.

Appendix 2: Interview Guide

Initial Information
Demographics
Gender
Age
Background information (introduction)
Level of Education
Title/Responsibilities
Experience with BI
Dashboard
Definition of a dashboard, the data set
What is the first thing you might have noticed in the dashboard? (Implications to form follows function and color usage)
What is your initial reaction? (Implications to color usage and data collection)
What is the first piece of data you gathered from the dashboard? (Implications to gestalt principles and the use of color)
Can you explain trail of thought flowed next?
Data & data quality
What data can you get from the dashboard, is it easy to find
Could you tell me which country has the highest profit? (Implications to type of visual chosen and color usage, data quality – representational DQ)
Is the information easy to find? (Implications to color usage and placement – gestalt, data quality – representational DQ)
Which segment has the highest profit? (Implications to data quality – representational DQ)
Which month/year/quarter had the highest profit? (Implications to data quality – contextual, representational DQ)
Did the number of units sold impact the profit? (Implications to data quality – contextual, representational DQ)
Color
Use of color in the dashboard
What do you think about the colors used? (Implications with color theory and emotional theory)
Do you think that the colors shifted your attention? (Implications to the limitation of color usage)
Do you think the colors impacted your understanding of the data? (Implications to using color only where necessary)
Corrections of the dashboard
Do you think there is anything on the dashboard that should be changed?
What would you change? How would you change it?

Source: Own Work.