9. DATA VISUALISATION METHODS

Author: Dario Šebalj

In today's data-driven world, the ability to efficiently translate complex datasets into clear, intuitive visualizations is a necessity for organizations that want to use their data effectively. Data visualization goes beyond a purely esthetic representation; it is a fundamental component of business intelligence that helps decision makers identify trends, outliers and patterns hidden in raw data. This chapter introduces different types of visualizations, from simple charts such as bar charts and line charts to more complicated graphical representations such as heat maps and bullet graphs. Each type of visualization serves different purposes and is suitable for different data sets, so it is critical for data analysts to select the appropriate visual to effectively convey the intended message.

Data visualization is the process of transforming information into a visual context, such as a map or chart, and is used to make data easier for the human mind to understand and draw conclusions from. The main goal of data visualization is to facilitate the identification of patterns, trends and outliers in large data sets. Common types of data visualization include charts, tables, maps and dashboards (Brush, 2022; GeeksForGeeks, 2024).

Due to the growing popularity of big data and data analytics projects, visualization is now more important than ever. Companies are increasingly using machine learning to collect huge amounts of data, which can be difficult and slow to process, understand and explain. This can be sped up with the help of visualization, which also makes the information easier to understand for stakeholders and business owners (Brush, 2022).

Before choosing a visualization method, it is important to understand the context of the visualization.

9.1. Understanding of the situation context

Nussbaumer Knaflic (2015) states that understanding and contextualizing is the first and most important step before engaging with data visualization techniques and storytelling methods. Understanding the audience is the key aspect of context. Nussbaumer Knaflic emphasizes the importance of knowing who the audience is, their level of expertise, and what is important to

them. This understanding ensures that data visualization and storytelling are tailored to the audience's needs and preferences, making the information more relevant and engaging.

According to IBM (n.d.), general background information helps the audience understand the significance of a particular data point. For example, if a company's email open rate is below average, we should show how the open rate compares to the industry as a whole to illustrate that there is a problem with this marketing channel for the company. The audience needs to understand how the current performance compares to a specific target, benchmark or other key performance indicators (KPIs) in order to be motivated to take action.

There are three important questions which need to be answered (Nussbaumer Knaflic, 2015; IBM, n.d.):

- **Who:** This involves identifying the audience and understanding their perspective in order to know how the story should be tailored. This ensures that the visualization is aimed directly at the target audience, making it more effective and engaging. For example, while a quarterly annual report may only require a high-level summary data, a financial analyst may need a detailed trend analysis over several years. Deciding on the complexity, level of detail and insights to emphasize depends on who will be looking at the visualization.
- **What:** This refers to the key message or insight to be communicated to the audience. It is about being clear about the action or decision that the data visualization is intended to influence. The context defines the purpose of the visualization. Is it to persuade, inform, explore or confirm? Each purpose can lead to different decisions about the type of visualization and the data points highlighted. For example, a persuasive visualization designed to gain support for a new initiative may focus on different data than a visualization designed to simply inform about past performance.
- **How:** This is about choosing the most appropriate and effective means of communicating the story or insight, taking into account the medium, format and visualization techniques that will best resonate with the target audience. Certain types of datasets also require special visualization. For example, scatter plots are good for showing the relationship between two variables, and line graphs are a good way to show time series data. The visuals should help the audience to understand the main message. An incorrect arrangement of charts and data can have the opposite effect and confuse rather than enlighten the audience.

When it comes to visualization and data analysis, a distinction is made between explorative and explanatory visualization. Exploratory visualization motivates the user to investigate the data or topic more thoroughly in order to make their own discoveries. Explanatory visualization brings the results to the forefront, conveying the author's hypothesis or argument to the reader (Schwabish, 2021).

Having explored the importance of understanding the situational context in which data visualizations are used, it is clear that this foundational knowledge determines the way information is best communicated and perceived by an audience.

The next critical step is to effectively engage the audience. The next subchapter describes strategies for engaging and maintaining the viewer's interest. This includes selecting elements that enhance the visual appeal and readability of the data visualization and ensure that key insights do not go unnoticed. By using techniques that draw the viewer's eye and highlight important data, visualizations can be more than just informative — they can be captivating and compelling.

9.2. Methods to attract attention

When designing data visualizations, it is very important to capture and direct the audience's attention. The interplay between the mechanics of vision and the principles of visual perception determines how effectively a visualization conveys the intended message. Understanding how the human eye perceives visual elements is the first step in creating compelling visualizations.

Approximately 70% of the sense receptors in our bodies are dedicated to vision (Few, 2012).

The eye is preattentively attracted to certain visual attributes, i.e. these attributes are processed quickly and automatically in the visual system without conscious effort. Attributes such as color, size, shape and orientation can be used to highlight critical data points or areas within a visualization and immediately attract the viewer's attention.

According to Schwabish (2021), **Gestalt theory** describes how people typically group visual elements. The word *Gestalt* means *pattern* (Cairo, 2013). It was developed by German

psychologists in the early 20th century. When it comes to creating diagrams and other visualizations, the six principles of Gestalt theory are particularly helpful.

The principle of **proximity** states that our perception groups objects together when they are in close proximity to each other (e.g. Figure 9.1).

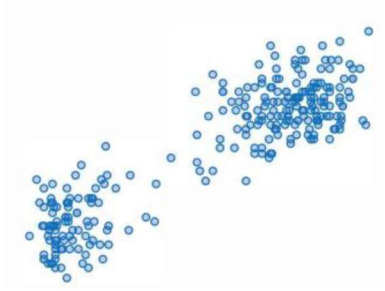

Figure 9.1 Proximity as Gestalt theory principle

Source: Schwabish (2021).

The principle of **similarity** says that the human brain categorizes objects based on their shared attributes such as color, shape, or direction (e.g. Figure 9.2).

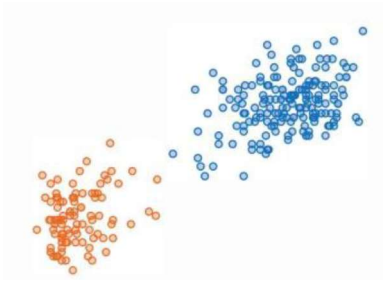

Figure 9.2 Similarity as Gestalt theory principle

Source: Schwabish (2021).

According to the principle of **enclosure**, bounded objects are perceived as a group (e.g. Figure 9.3).

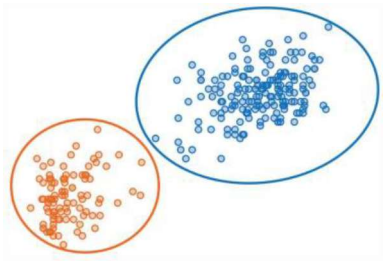

Figure 9.3 Enclosure as Gestalt theory principle

Source: Schwabish (2021).

According to the principle of closure, our brain tends to ignore gaps and fill in missing information in order to form a complete structure. When analyzing a line chart that contains missing data, we tend to mentally fill in the gaps using the simplest approach (e.g. Figure 9.4).

Figure 9.4 Closure as Gestalt theory principle

Source: Schwabish (2021).

The principle of **continuity** suggests that elements aligned in a straight line or a smooth curve are perceived by the viewer as more related than elements that do not lie on the line or curve (e.g. Figure 9.5).

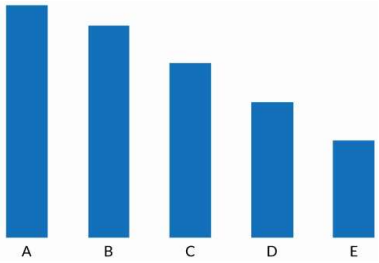

Figure 9.5 Continuity as Gestalt theory principle

Source: Schwabish (2021).

Based on the principle of **connection**, our perception categorizes objects that are connected to each other as belonging to the same group (e.g. Figure 9.6).

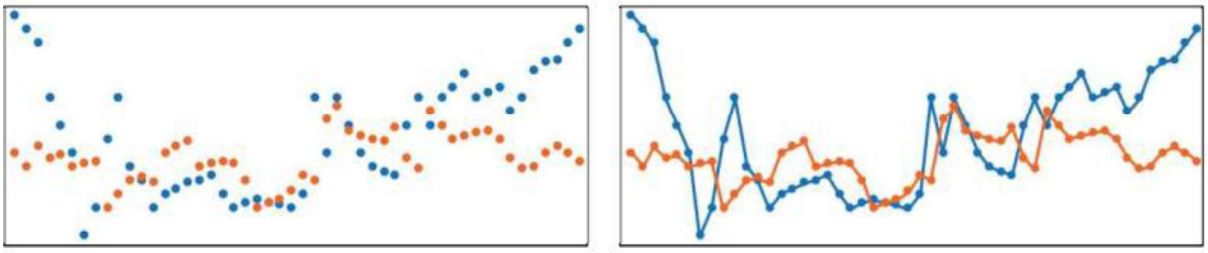

Figure 9.6 Connection as Gestalt theory principle

Source: Schwabish (2021).

There is a very important concept for data visualization, a subset of Gestalt theory called preattentive processing. Schwabish (2021) explains that preattentive features draw our attention to a specific area of a graph or image.

These features refer to the visual qualities that the human visual system can perceive in the early stages of visual processing without conscious attention, and which are usually measured in milliseconds. One of the preattentive attributes used in this book is color (blue) and weight (bold text). The next very popular example is finding a specific number in the number matrix (e.g. Wexler et al., 2017). Figure 9.7 shows the number matrix without (on the left side) and with (on the right side) preattentive attributes.

Figure 9.7 Using preattentive attributes in data visualization

Source: Wexler et al. (2017).

From the left matrix, it takes a long time to guess how many 9s are there. But just one change in the matrix makes a big difference. Only the color was changed – 9s are red and all the other numbers are light gray. Color (in this case, hue) is one of the several preattentive attributes. Figure 9.8 shows examples of some preattentive attributes often used in data visualization.

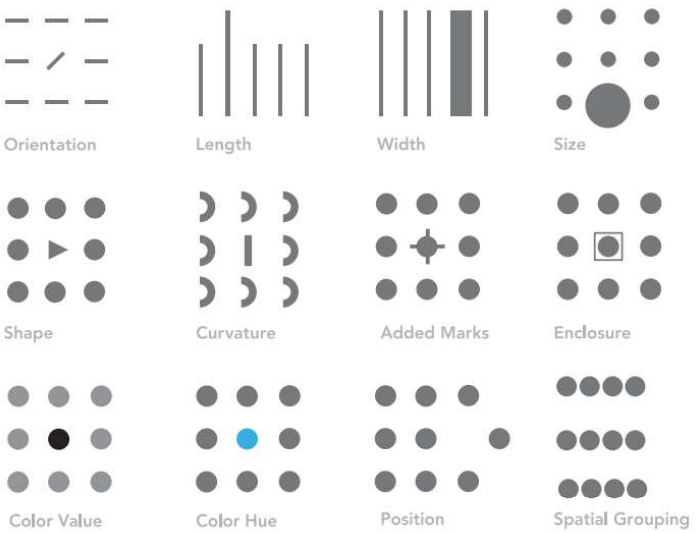

Figure 9.8 Types of preattentive attributes used in data visualization

Source: Wexler et al. (2017).

Preattentive attributes enable viewers to recognize patterns, outliers, or important data points almost instantly. By employing strategies that guide the viewer's eye to the most significant information, data visualizations can significantly enhance the communication and understanding of complex datasets.

Next subchapter will examine how different types of data and the insights they are intended to provide influence the choice of visualization methods. From simple bar charts to more complex heat maps or bullet graphs, choosing the right visualization method is essential in ensuring that the data not only captures attention but also communicates the intended message effectively and accurately.

9.3. Choosing the right visualization method

The first step in selecting the appropriate method is a thorough understanding of the data. What are the key messages we want to convey? What type of data are we dealing with? Are we working with time series data, geographical information or hierarchical structures? The type of data used can have a significant impact on the chosen visualization method. Choropleth maps, for example, are best suited for displaying geographical data, while line charts may be more appropriate for time series data.

The background context, and expectations of the audience also play an important role in the choice of visualization method. A technical audience may appreciate detailed and complex visualizations such as heat maps or network diagrams. A general audience, on the other hand, may find simpler charts, such as bar charts or line charts, more accessible and engaging.

Interactivity is another important aspect to consider. Interactive visuals, such as dynamic dashboards, allow users to explore different levels of data by filtering, zooming and selecting specific elements. This interactivity can lead to deeper insights as users can tailor the visualization to their specific questions.

A visualization method does not always have to be charts. It can also be a table or even simple text. As Few (2012) states, the purpose of tables and graphs is to effectively convey important information and provide the reader with important, meaningful and useful insights.

In the next few sub-chapters, the most popular visualization methods will be briefly explained.

9.3.1. Simple text

Nussbaumer Knaflic (2015) suggests using simple text when there is only one or two numbers to share (Figure 9.9).

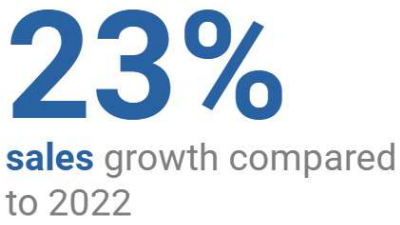

Figure 9.9 Simple text in visualization

Source: Author.

According to Schwabish (2021), this simple text is often referred to as BAN (Big Ass Numbers). They are typically used to draw attention to key metrics or performance indicators and give the viewer immediate access to important information. By highlighting areas that require attention or action, BANs typically aid in decision making by helping users focus their attention on important aspects of the data (Tay, 2024). Although BANs are simple, they can be enhanced with subtle visual elements such as color coding or icons to indicate performance against targets or changes over time. For example, a red downward arrow next to a sales number can immediately indicate a decline, while a green upward arrow signals growth.

9.3.2. Table

Tables are an essential part of data visualization because they provide a structured and clear way to present numerical data. Tables are incredibly useful when it comes to presenting detailed information with precision and clarity, even though they may not have the same visual impact as charts or maps. According to Schwabish (2020), in most cases, they are not intended to provide a quick visual representation of data. Instead, tables are useful when the exact values of the data or estimates need to be shown. While they are not the ideal option for presenting a lot of data or in a small space, a well-designed table can help the reader find specific numbers as well as spot trends and outliers. Few (2012) notes that tables are useful for reference and one-to-one comparisons due to their simple structure and the fact that the quantitative values are expressed as text that we can immediately understand without having to translate it.

Tables should be formatted as follows (Schwabish, 2020; Nussbaumer Knaflic, 2015):

- remove all borders around the table
- lighten the grid lines as much as possible or remove them completely
- clearly separate the header from the body of the table
- align the text in the table and header to the left, and the numbers to the right
- use an appropriate level of data detail (e.g. use numbers with one decimal place if this is sufficient to understand the data)

9.3.3. Bar chart

A bar chart is ideal for displaying numerical values by groups or categories (e.g. if we want to display the number of employees by department). It can be aligned vertically or horizontally. Horizontal alignment (as in Figure 9.10) is recommended if the category names are too long or if there are too many categories. The bar chart in Figure 9.10. shows sales (quantitative data) per region (qualitative data).

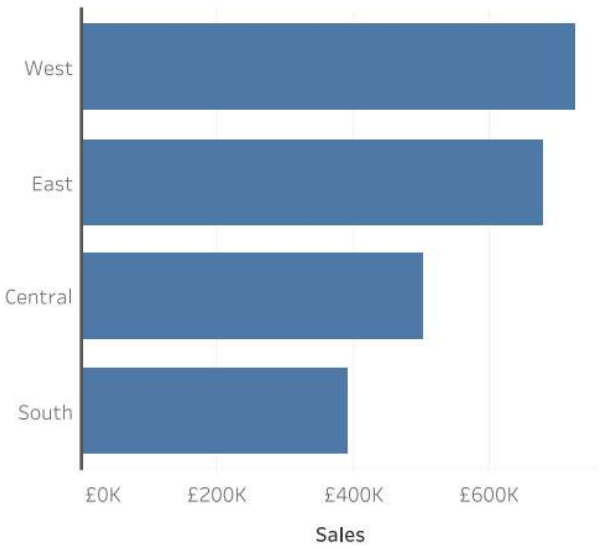

Figure 9.10 Bar chart in visualization

Source: Wexler et al. (2017).

According to Few (2013), the most effective way to represent measures related to discrete items on a nominal or ordinal scale is a bar chart. It is easy to compare individual values by simply comparing the height of the bars. The axis of a bar chart must start at zero. If the axis starts at a value other than zero, this can overemphasize the difference between the bars and distort our perception of the values in the bar chart, which is based on the length of the bars (Schwabish, 2021).

9.3.4. Line chart

A line chart is used to represent the change of a quantitative value, which lies on the y-axis, over time, which is positioned on the horizontal x-axis. Yi (n.d.a) suggests that a line chart should not contain more than five lines. Also, it is not necessary to include a zero baseline for a line chart. It is acceptable to extend the range of the vertical axis to the point where the value changes are most informative if a zero line is neither understandable nor helpful.

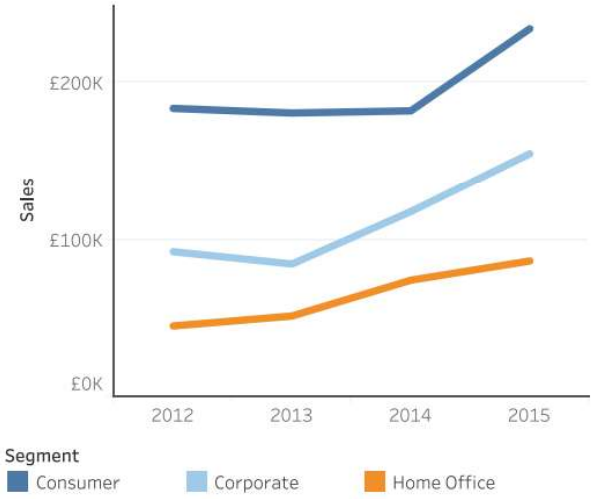

Figure 9.11 Line chart in visualization

Source: Wexler et al. (2017).

The line chart in Figure 9.11 shows the sales (quantitative data) over a period of 4 years and is also broken down by segment.

An area chart (Figure 9.12), which is a variant of the line chart, adds shade between the line and a zero baseline (Yi, n.d.a).

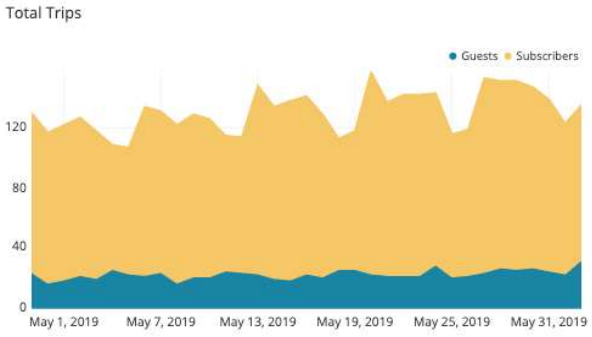

Figure 9.12 Area chart in visualization

Source: Yi (n.d.)

The area chart can be seen as a cross between a line chart and a bar chart, as the values can be interpreted not only by their vertical positions but also by the area shaded between each point and the baseline (Yi, n.d.).

9.3.5. Scatterplot

A scatterplot is used when we want to see if there is a relationship between two quantitative variables. According to The Data Visualisation Catalogue (n.d.), the patterns seen on a scatterplot can be used to interpret the nature of the correlation. These are: positive (values increase together), negative (one value decreases while the other increases) or zero (no correlation).

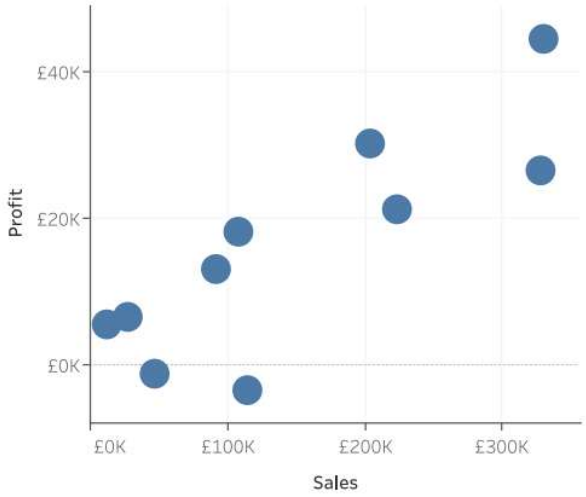

Figure 9.13 Scatterplot in visualization

Source: Wexler et al. (2017).

The scatterplot in Figure 9.13 shows the relationship between profit and sales (both quantitative variables).

According to Yi (n.d.), it is very important to mention that in a scatter plot, just because we see a relationship between two variables, it does not mean that changes in one variable cause changes in the other. This leads to the widely used phrase in statistics: "correlation does not imply causation."

9.3.6. Choropleth map

A choropleth map uses differences in shading or coloring within predefined areas to indicate the values or categories in those areas (Wexler et al., 2017). According to Schwabish (2021),

the color palette on the choropleth map is easy to understand, smaller values correspond to lighter colors and larger values to darker colors.

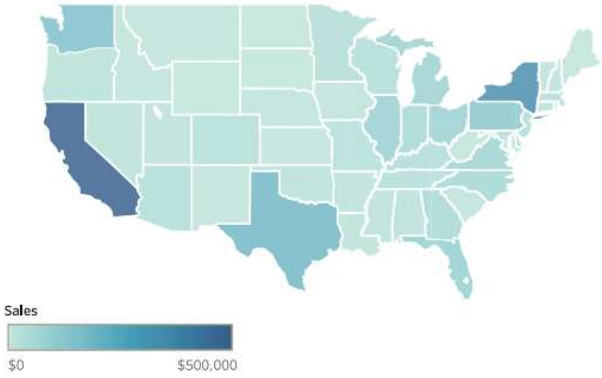

Figure 9.14 Choropleth map in visualization

Source: Wexler et al. (2017).

The choropleth map in Figure 9.14 shows the sum of sales in different states in USA.

9.3.7. Heatmap

A heatmap is a visualization of data in tabular format, where colored cells represent the relative magnitude of the numbers (Nussbaumer Knaflic, 2015).

Since color is a key element of this type of chart, you need to make sure that the color palette you choose matches the data. The most common type of color is a sequential color, where darker colors correlate with higher values and lighter colors correlate with lower values, or vice versa (Yi, n.d.b).

	Region A	Region B	Region C
Category 1			
Category 2			
Category 3			
Category 4			
Category 5			

Figure 9.15 Heatmap in visualization

Source: Author, adapted from Nussbaumer Knaflic (2015).

The heatmap in Figure 9.15 shows different values of some quantitative data (e.g. sales) by category (in rows) and region (in columns).

9.3.8. Bullet graph

A bullet graph is invented by Stephen Few in 2005 (Few, 2013). It is basically a bar chart with a single black horizontal bar representing an actual value, an additional (vertical) marker for a target value (that we want to achieve) and shaded areas in the background representing a scale of success (e.g. poor, good, excellent).

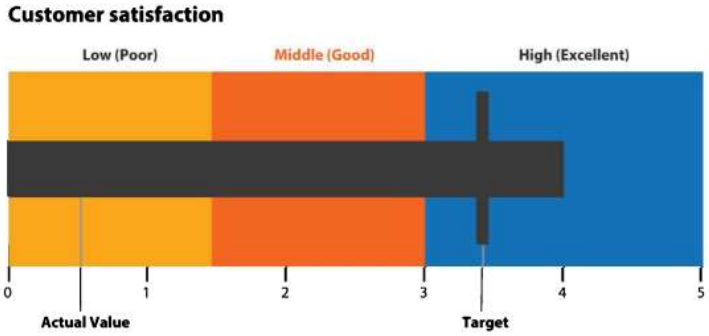

Figure 9.16 Bullet graph in visualization

Source: Schwabish (2021).

The bullet graph in Figure 9.16 shows that we want to achieve a customer satisfaction rating of 3.4 (out of 5). Our current satisfaction rating is 4, which is above the target value. In the background there are three areas of customer satisfaction – low (poor), middle (good) and high (excellent).

Choosing the right visualization method is very important for effective communication, but the design principles that guide these techniques also play a critical role in the clarity and effectiveness of the data presentation. In the next section, we will look at the importance of layout, typography, color schemes, and the strategic use of space, which are critical to making visualizations not only esthetically pleasing, but also easy to understand and interpret.

9.4. The guidelines for good visualization design

Effective visualization design is about enhancing the viewer's ability to understand and interact with data. This involves a careful balance between esthetic elements and functionality, with the choice of color, font and layout playing a crucial role in conveying information clearly and efficiently. Simplicity should also be a guiding principle. But according to Cairo (2013), graphics should not simplify messages. They should clarify them, highlight trends, uncover patterns and reveal realities that were not previously visible.

A common pitfall is overcomplicating a visual with too many elements that confuse rather than clarify. The goal is to make the data accessible and understandable to the target audience and ensure that the visualization serves its purpose, which is to inform and support decision-making.

The clarity and effectiveness of data visualization can be improved by removing unnecessary distractions. Nussbaumer Knaflic (2015) emphasizes that any element that does not add value or directly support the understanding of the data is considered **clutter**. This includes unnecessary grid lines, excessive colors, irrelevant data points, and overly complex chart decorations. These elements can distract attention from the key messages the data is intended to convey. She recommends techniques such as simplifying color schemes, minimizing text and using white space. By using white space strategically, you can create a visual hierarchy that highlights key data points and makes the overall presentation clearer and easier to understand. In addition, the effective use of white space can help create a balanced layout that is less cluttered and more organized.

Color is a very important part of any data visualization. It serves not only to attract attention, but also to organize information and convey meaning effectively. When used correctly, color can greatly enhance the clarity and impact of a visualization. There are several suggestions for the effective use of color in visualizations (Few, 2012; Cairo, 2013; Few, 2013; Nussbaumer Knaflic, 2015; Wexler et al, 2017; Schwabish, 2021; Lidwell, 2023; Interaction Design Foundation, n.d.):

- **Choose appropriate color combination:** using the color wheel, it is possible to create visualizations that are visually balanced and pleasing to the eye. There are several common color combinations (Figure 9.17):
    - **Monochromatic** – one color in different shades.
    - **Analogous** – three colors next to each other on the color wheel. These colors are pleasing to the eye and create harmonious design.
    - **Complementary** – two opposite colors on the color wheel. These colors are contrasting colors and should be used to emphasize something (e.g. increase - green/decline - red).
    - **Triadic** – three equally distant colors on the color wheel. These colors are dynamic and attract attention.

In addition, warmer colors should be used for foreground elements and cooler colors for background elements.

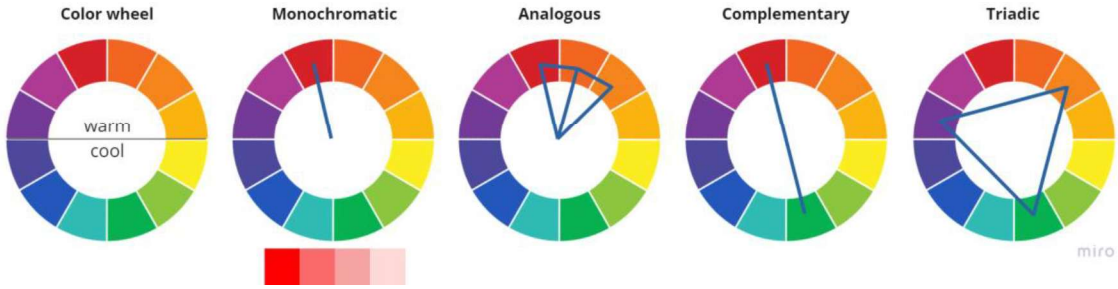

Figure 9.17 Color combinations

Source: Author, adapted from Lidwell (2023).

- **Choose appropriate color schemes:** the choice of color scheme depends on the type of data to be visualized. A **sequential** color scheme is the use of one color from light to dark and is ideal for displaying numerical data that progresses from a low to a high value (e.g. sales by state). A **diverging** color scheme is useful for highlighting values above or below a midpoint (e.g. profit/loss). A **categorical** color scheme is best for categorical data where the colors need to distinguish different groups without implying an order or value (e.g. product categories). Figure 9.18 shows different color schemes used in visualizations.

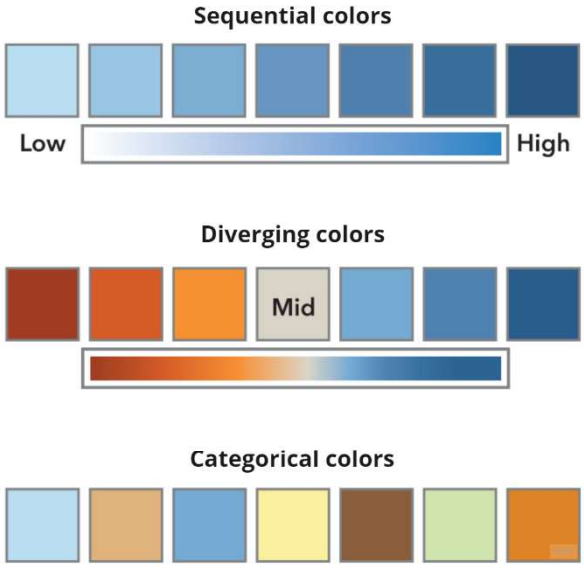

Figure 9.18 Color schemes

Source: Wexler et al. (2017).

- **Use color sparingly:** excessive use of color can cause confusion and make a chart harder to understand. The color palette should be limited to what the human eye can quickly distinguish between, about five different colors.
- **Consider color blindness:** about 8% of men and 0.5% of women are color blind. Avoid color combinations that are difficult for color-blind users to distinguish, such as red and green. Instead of these colors, the better combination is orange and blue.
- **Colors should be consistent:** consistent use of color across multiple visualizations allows the viewer to easily understand and compare data. Once a color scheme has been established for specific data types or categories, it should be maintained in all associated visualizations.
- **Use color to highlight important data:** color can be a strong indicator of where to look. Using a bright or contrasting color can draw attention to key data points or findings, while more neutral colors can be used for less critical information. Some authors suggest that creating a clear, understandable visualization should start with a gray color. All data elements in the chart (e.g. bars in the bar chart or lines in the line chart) should be gray. Then add labels and color only for the elements you want to highlight.
- **Keep it simple:** in the field of data visualization, the KISS principle, an acronym for "Keep It Simple, Stupid", is very relevant and helpful. It refers to the use of simple charts, as complex charts or overly detailed visuals can overwhelm users and make it difficult to recognize the key messages or data points. It also means that you should avoid visual clutter and reduce unnecessary visual components such as overly bright colors, fonts, and grid lines. When applying the KISS principle, the focus should be on the data itself and not on decorative or overly complex design elements.

These principles are crucial to ensuring that data visualizations achieve their primary goal of communicating complex information in a way that is accessible and understandable to all audiences.

It is clear from this chapter that effective data visualization is a critical component in the process of data-driven decision making. Based on an understanding of situational context, this chapter has highlighted the importance of tailoring visualizations to the specific needs of the target audience. Through a detailed examination of different visualization methods and design principles, it is shown that thoughtful visual representation of data is important to enable better understanding and communication of complex information.

REFERENCES

1. Brush, K. (2022). Data visualization. TechTarget [available at: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization, access April 15, 2024]
2. Cairo, A. (2013). The functional art: An introduction to information graphics and visualization. New Riders.
3. Data Visualisation Catalogue (n.d.). Scatterplot [available at: https://datavizcatalogue.com/methods/scatterplot.html, access April 17, 2024]
4. Few, S. (2012). Show Me the Numbers. Analytics Press.
5. Few, S. (2013). Information dashboard design: Displaying data for at-a-glance monitoring. Analytics Press.
6. GeeksForGeeks (2024). What is Data Visualization and Why is It Important? [available at: https://www.geeksforgeeks.org/data-visualization-and-its-importance/, access April 15, 2024]
7. IBM (n.d.). What is data visualization? [available at: https://www.ibm.com/topics/data-visualization, access April 15, 2024]
8. Interaction Design Foundation (n.d.). Color Theory [available at: https://www.interaction-design.org/literature/topics/color-theory, access April 18, 2024]
9. Lidwell, W., Holden, K. & Butler, J. (2023). Universal Principles of Design, 3rd Edition. Quarto Publishing Group USA.
10. Nussbaumer Knaflic, C. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley.
11. Schwabish (2020). Ten guidelines for better tables. Journal of Benefit-Cost Analysis, 11(2), pp. 151-178.
12. Schwabish (2021). Better data visualization: A guide for scholars, researchers and wonks. Columbia university press.
13. Tay, J. (2024). Effective use of BANs. Medium [available at: https://medium.com/@e0373084/eye-catching-bans-88d29632e4fa, access April 12, 2024]
14. Wexler, S., Shaffer, J. & Cotgreave, A. (2017). The big book of dashboards: Visualizing your data using real-world business scenarios. Wiley.