DATA SCIENCE, DASHBOARDS, AND THE WAY IT WORKS WITH STATISTICS

by

Denise Renee Bradford

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Susan R. VanderPlas, Ph.D

Lincoln, Nebraska

Month, Year

DATA SCIENCE, DASHBOARDS, AND THE WAY IT WORKS WITH STATISTICS

Denise Renee Bradford, Ph.D.

University of Nebraska, Year

Adviser: Susan R. VanderPlas, Ph.D

Here is my abstract. *(350 word limit)*

COPYRIGHT

DEDICATION

Dedicated to...

ACKNOWLEDGMENTS

Thank you to all my people!

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Statisticians use graphs in almost every stage of their work: we create charts when we get new data, to explore what we have and identify potential problems and opportunities. We fit models based on relationships between variables which are often identified visually. We identify problems with those models based on residual plots and other visual diagnostics. When our modeling work has been completed, we present our results to interested parties using visual displays, because non-statisticians often find it easier to understand data and models through an intuitive visual medium rather than through the mathematical formulae which underlie the statistical work.

Given the wide range of uses for graphs and visual data displays in statistical modeling, it is unsurprising that some graphs are more useful for specific applications such as exploratory analysis, and are unsuitable for other applications, such as presentation to an outside group. In addition, we know that not all visual displays have equal perceptual value (Aspillaga, 1996). The best graphics are designed to account for both the features of the dataset and the features of the intended audience. Some design constraints stem from limitations of the human perceptual system and are common to most potential consumers of the visualization: the sine illusion (VanderPlas & Hofmann, 2015b) affects anyone with binocular depth perception, and color recommendations are built around the specific characteristics of the human retina. Other design constraints are due to the audience's experience level: are they used to working with data? Do they understand specialized techniques such as principal component analysis to the point where a plot of factor loadings might be a useful visual display? When we create visualizations for public consumption we have to consider both perceptual factors and the target audience's domain knowledge. In this introduction, we explore previous research related to the construction of interactive and static visual displays for different audiences and consider the implications of this research when designing interactive data displays such as dashboards.

Most research in statistical graphics has been done on static graphics; usually, research also strips away all but the most essential contextual information. As a result, it can be hard to generalize this research to practical applications, where the contextual information surrounding the data is critical and the chart does not just exist in a vacuum.

In the ”real world”, however, conventions and familiarity often win out over best practice validated by perceptual experiments.

For example, in sports, many coaches desire printable diagrams containing all necessary and valuable information on a single page. As data in sports becomes more prominent, extensive, and collected, this information must be refined.

Thus, in addition to the experimental evidence, we must consider the human element: how to introduce new graphical concepts to stakeholders, and the considerations involved in encouraging stakeholders to adopt these improved graphics. Let us first consider the audience characteristics that affect the selection of graphics. Then, we will engage with considerations based on the data to be displayed. Finally, we will consider the interactions between the audience and the data: how graphics are tested, amended, and hopefully eventually adopted into common use.

1.1 Audience Considerations

Several factors, including perception, attention, and expertise, can influence our desire and ability to read and engage with data visualization.

Perception and attention are crucial cognitive processes that allow us to interpret and make sense of data visualizations. Perception refers to the manner in which we interpret and organize sensory information from our environment, whereas attention refers to the capacity to selectively focus on particular aspects of this information.

Our ability to perceive and pay attention to pertinent features, such as patterns, trends, and relationships, is essential for comprehending data visualizations. This is especially true when working with unfamiliar or complex data sets, as our ability to focus on pertinent information becomes more difficult.

In addition to perception and focus, domain-specific knowledge is essential for understanding and interacting with data visualizations. Expertise in a particular field can enable individuals to better interpret and comprehend the significance of the presented data, as well as identify potential biases or errors in the visualization.

In conclusion, the ability to perceive and interact with data visualizations requires a combination of perceptual and attentional processes, as well as domain-specific knowledge, in order to interpret and comprehend the presented information.

The term "data visualization" dates back to the 2nd century A.D. drawings and other visual representations were used to investigate the world and record historical events in ancient societies. Throughout human history, data visualization has significantly contributed to invention and discovery (**Crapo2000?**). The introduction of computer technology dramatically changed the visual representation of data. Using computer-based graphical data visualization, data analysts have become faster and more precise. Data visualization has become an integral component of research in numerous disciplines, such as algorithms, human perception, animation, computer vision, etc. The origin of data visualization being a sub-category, it is regarded as "the science of the visual representation of data," (**friendly2009?**). "Data Visualization: a successful design process" defined data visualization as "the representation and presentation of data that exploits our visual perception abilities to amplify cognition," Kirk (2012). This suggests that data visualization involves the exploitation of human visual perception in addition to the presentation of data. Assigning meaning to visualization is not a statistical or computational step but a cognitive one. Each step in the data analysis process is part of a more extensive mental process.

1.1.1 Perception

Human perception is an essential component of data visualization. Colin Ware suggests that perception can significantly enhance both the content and quantity of displayed information, Ware (2012). Perception refers to the organization, interpretation, and conscious experience of sensory data. Perception is also defined as “the process of recognizing (being aware of), organizing (gathering and storing), and interpreting (binding to knowledge) sensory information,” Ward, Grinstein, & Keim (2010). Ward et al. explain the notion of perception as the following: > Simply put, perception is the process by which we interpret the world around us, forming a mental representation of the environment. This representation is not isomorphic to the world, but it’s subject to many correspondence differences and errors. The brain makes assumptions about the world to overcome the inherent ambiguity in all sensory data, and in response to the task at hand.

Human visual perception is a highly complex and subjective process, and the efficacy of a visualization in communicating objective understanding depends on a vast array of subtle factors, Reuter, Tukey, Maloney, Pani, & Smith (1990). Furthermore, certain situations present unique challenges and lead to systematic errors; can these provide insight into how the brain solves the problem of which objects are represented by which images in general (**gregory1968?**). Additional to the general process of human visual perception process and efficacy of a visualization in communicating, short-term memory is crucial to the effectiveness of statistical graphics. Research indicates that our short-term memory can only store a limited amount of information at any given time. Therefore, designers must present data in a manner that is simple to comprehend and remember.

Untrained analysts can and do “analyze” data with only their natural mental abilities - The mind performs its data analysis-like process to create detailed understandings of reality from bits of sensory input. In a later chapter, we will show how a utilizing parallel coordinate plot is one method for achieving a simple design to comprehend. These plots enable viewers to compare multiple variables concurrently, thereby reducing cognitive load and making it easier to identify patterns and trends.

Examining the Gestalt principles, which describe how our brain organizes and interprets sensory information to form coherent patterns and objects, is one way to gain a deeper understanding of how perception functions.

1.1.2 Gestalt Principles

Our visual interpretation of the world is a major factor in why humans perceive the world and its objects as organized, regular, and simple shapes, schemas, figures, or forms. The theory of Gestalt has philosophical and psychological roots that date back to the late 1800s. Gestalt therapy is founded on the notion that overall perception is contingent upon the interaction of numerous factors. These principles include proximity, similarity, closure, continuity, symmetry, and figure-ground relationships formally outlined in “Principles of Gestalt Psychology”, Koffka (2013).

Gestalt principles include:

- Proximity: Objects close to each other are perceived as related or grouped.

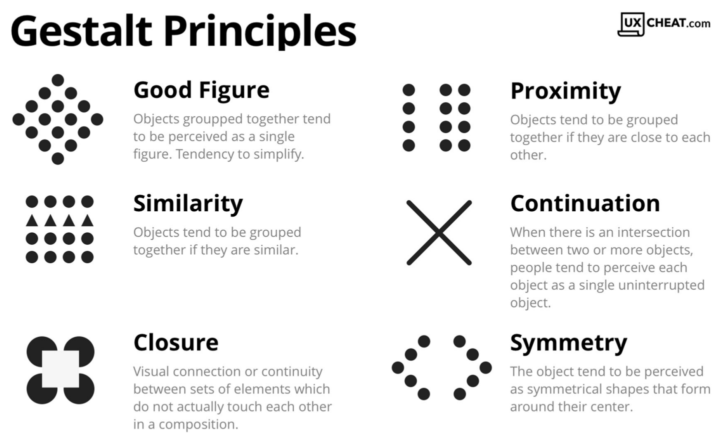


Figure 1.1: Gestalt Principles with Examples

- Similarity: Objects that are similar in some way (e.g., shape, color, size) are perceived as related or grouped.
- Continuation: The human eye follows lines and patterns, so designers can use this principle to guide the viewer's gaze through a display.
- Closure: The human brain tends to complete incomplete figures or patterns, so designers can use this principle to create the illusion of missing information.
- Figure-Ground: The human brain separates the foreground (figure) from the background (ground), so designers can use this principle to create visual hierarchy and emphasis.
- Contrast: The human eye is drawn to high-contrast areas, so designers can use this principle to create emphasis and hierarchy.
- Symmetry and Balance: The human eye finds symmetry and balance visually pleasing, so designers can use this principle to create a sense of harmony and order.

These principles are based on cognitive psychology and understanding how the human brain processes visual information. By applying these principles to dashboard design, designers can create visual arrangements that make it easier for viewers to understand the relationships between data elements. For example, proximity can be used to group related elements together, while symmetry can be used to create balance and harmony in the overall layout of the dashboard. At its most basic, the entire form is perceived (or emerges to our visual pathways) as opposed to its component parts.

All of this suggests that our brain frequently perceives things differently than what is actually present. If you are familiar with optical illusions or the famous "gorilla in the crowd" experiment, you are aware that we do not always process everything in our visual field. There is simply too much information for our brains to process, and if we tried to interpret it all, we would be rendered paralyzed. We organize the world according to Gestalt principles and pre-attentive attributes so that it is familiar, makes sense, and is easy to process. These principles guide how people perceive and make sense of the world around them, and they play a critical role in designing effective visual displays, such as dashboards. Understanding the Gestalt principles can also cast light on how information is processed and stored in short-term memory and attention's critical role in this process.

1.1.3 Attention and Memory

The Gestalt principles describe how humans perceive and organize visual information into meaningful patterns and structures to help the brain to effectively process and organize incoming visual information, making it easier to attend to and remember. It's worth exploring the efficacy of short-term memory when developing effective dashboards.

Short-term memory (STM), also known as working memory, is the stage of temporary storage and processing where the majority of memory retention effort is expended. According to Alan Baddeley's Working memory: Theories, models, and controversies, STM is a limited-capacity system prone to interference and decay, Baddeley (2012). Selective attention is essential for the maintenance of STM because it allows us to filter out irrelevant information and concentrate on what is essential, Cowan (2001).

Visual aids such as charts and diagrams can improve short-term memory by allowing us to encode and retain information more effectively, according to research, Alvarez & Cavanagh (2004). Consequently, utilizing visual aids such as charts can be advantageous for enhancing our short-term memory. Furthermore, annotations can also help aid short-term memory. By adding annotations, such as notes or highlights, to information we are trying to remember, we can improve our recall of the information later on, Alvarez & Cavanagh (2004).

According to the Feature Integration Theory (FIT), STM is composed of two stages: pre-attentive processing and focused attention A. Treisman (1998). Parallel and independently, the brain processes the physical characteristics of an object, such as its color, shape, and orientation, during pre-attentive processing. However, focused attention is required to bind these features into a coherent object representation in STM. STM can be improved through various strategies, such as rehearsal, chunking, and elaboration Oberauer (2009). For example, by repeating a phone number several times or breaking it down into chunks of two or three digits, we can increase the likelihood of it being stored in

STM. Similarly, by elaborating on the information we want to remember, such as creating mental associations or visual images, we can enhance its retention in STM Bui & Myerson (2014).

STM is a dynamic and malleable cognitive system that is crucial to our daily lives. Understanding the mechanisms underlying STM and how to improve it can have significant implications for learning, memory, and the treatment of memory disorders. By analyzing the relationship between attention and working memory, we can gain insight into how we construct meaning from the information in our environment.

1.1.4 Constructing Meaning

Gestalt psychology suggests that humans actively construct meaning by organizing information into patterns and wholes Wertheimer (1938). Both top-down and bottom-up processing are involved in the process of meaning construction. Bottom-up processing entails analyzing sensory data from the environment and constructing perceptions based on this data. Top-down processing is the influence of prior knowledge, expectations, and context on the perception and interpretation of incoming sensory data.

Together, top-down and bottom-up processing facilitate the encoding and retrieval of information in the context of short-term memory. Selective attention, the ability to focus on relevant information while ignoring irrelevant information, is an example of top-down processing that aids in the encoding and retrieval of information in short-term memory Cowan (2010).
According to the feature integration theory, the perception of objects involves both the bottom-up analysis of individual features and the top-down pro-

cessing of higher-level features in order to form a complete perception A. M. Treisman & Gelade (1980).

The Gestalt principles of perception emphasize the significance of bottom-up and top-down processing in constructing meaning from sensory data. Both types of processing are involved in encoding and retrieving information, which has significant implications for understanding how short-term memory works.

1.1.5 Expertise

However, creating effective graphics is not a simple task, and proficiency in this area is required to create high-quality visualizations. This essay will discuss the contributions to the field of graphics made by research on psychological processes, automaticity, readily available information, and practice effects.

Cognitive Processes - the way we think about and approach a task.

As we become more proficient in a particular skill, we develop more complex and efficient mental models or schemata, a heuristic technique to encode and retrieve memories, the majority of typical situations do not require much strenuous processing. These mental models help us to organize information in a meaningful way, and to quickly identify and solve problems related to the task. This process is known as cognitive restructuring and is facilitated by developing domain-specific knowledge Ericsson & Lehmann (1996). For example, a basketball coach is able to quickly recognize patterns and positions on a court that are common in basketball, which allows them to make decisions more quickly and accurately than a novice coach or player.

Automaticity - the ability to perform a task without conscious effort or attention

As our proficiency in a task increases, our performance becomes more automatic, thereby freeing up cognitive resources for other tasks. The development of procedural knowledge, which is the ability to perform a series of steps or actions in a particular order, facilitates this process Fitts & Posner (1967). For example, a well-trained quarterback can throw a ball without looking at the wide receiver because their throwing movements have become automatic.

Information Readily Available - the way we process information related to a task

As our proficiency increases, we can recognize and retrieve pertinent information more rapidly and precisely than a novice. This is made possible by the creation of domain-specific knowledge structures that allow us to retrieve pertinent information from memory quickly Chase & Simon (1973). For example, a medical expert can quickly identify signs and diagnose a patient using their knowledge of disease symptoms and risk factors.

Practice Effects - extensive practice and experience

Practice effects are the performance enhancements that result from repeated practice. These gains are frequently most significant at the outset of practice, but gradually diminish as the individual approaches their performance ceiling Anderson (1982). The development of procedural knowledge and automaticity, which allow for more efficient and accurate task performance, facilitates the effects of the practice.

The contributions of research on psychological processes, automaticity, readily available information, and practice effects to the field of graphics have significant implications. Expertise is required to create high-quality graphics, which requires a thorough understanding of design principles and the capacity to work quickly and efficiently. The use of automatic processing and domain-specific knowledge can aid designers in processing and deciding on design elements efficiently and quickly. The creation of standardized design templates, workflows, and other tools can aid in enhancing the efficiency and effectiveness of the design process. Design skills can be improved through deliberate guidelines, and educators must focus on developing skills that can be applied in a variety of contexts.

As the significance of graphics in numerous fields continues to rise, the demand for specialists in this area will only intensify. By comprehending the contributions of research on psychological processes, automaticity, readily accessible information, and practice effects, designers, educators, and trainers can develop more effective approaches to graphics design and education. This can help ensure that the graphics used to convey complex information are clear, concise, and effective, making it easier for individuals to comprehend and interpret the required information.

1.1.6 Engagement with the data

The goal of data analysis is to extract meaningful insights, patterns, and knowledge from data. The process of data analysis involves collecting, cleaning, transforming, and modeling data, followed by the use of statistical and machine learning methods to uncover patterns and relationships within the data. The end goal of data analysis is to support decision making and provide a basis for informed action. Data analysis can help organizations to better understand their customers, market trends, and operational performance. Additionally, data analysis can support scientific research by helping researchers to test hypotheses, develop theories, and gain a deeper understanding of complex phenomena. Ultimately, data analysis aims to turn data into actionable insights and information that can inform and improve decision-making.

1.2 Data Considerations

John Tukey was the first to organize the collection and methods associated with philosophy into Exploratory Data Analysis (EDA). Previous research by Tukey focused on graphics as a tool for exploratory analysis. In "Exploratory Data Analysis," Tukey wrote that graphics and charts often display data with more enhanced understanding than a table, Tukey & Wilk (1966). Tukey outlines detailed the types of different graphics and in which situations to utilize these graphics. He was a strong advocate for the importance of EDA as a crucial first step in the data analysis process and emphasized the need for visualization and interactive techniques to understand patterns and relationships in data.

Tukey's Principles of EDA have become a cornerstone in the field of statistics and have been adopted by data professionals in various industries. Tukey's principles have enabled data professionals to understand complex data sets better and make more informed decisions by emphasizing the importance of visual exploration, data characterization, and model critique. In this way, Tukey's Principles have revolutionized our data analysis approach and become the foundational framework for EDA.

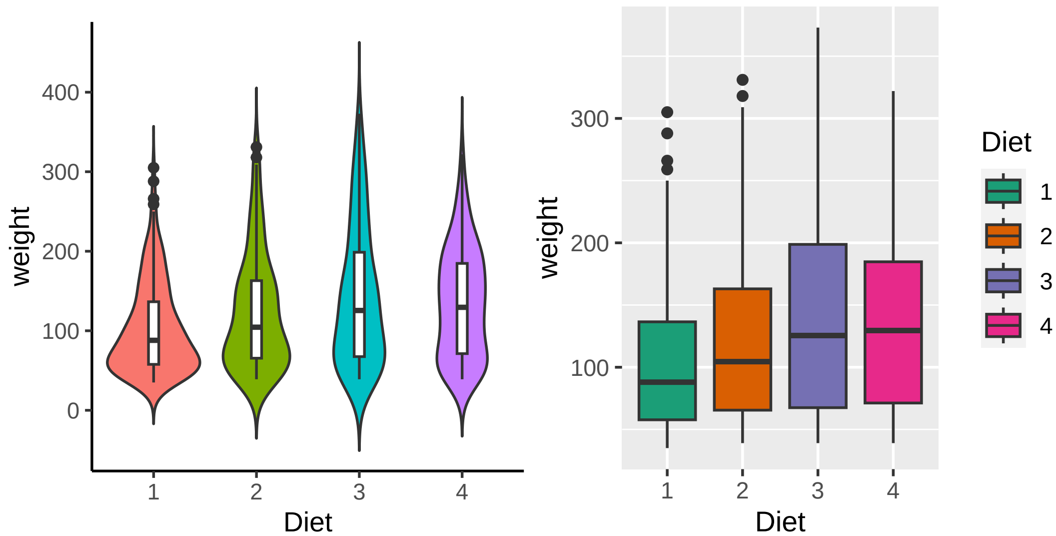
Tukey's Principles in EDA:

1. Graphical exploration, looking for patterns or displaying fit, the method demonstrates things about data that a single numeric metric does not understand. This has been useful in graphing the data before you develop summary statistics.

2. Describing the general patterns of the data. This step should be insensitive to outliers. In general, think about the types of resistant measures (i.e., median or mean). This step is making sure to determine data patterns.

3. The natural scale/state that the data are at their best. This will be the step at which the scale of data can be helpful for analysis. The reexpressing data to a new scale by taking the square root or logarithmic scale.

4. The mostly known parts of EDA but is done in the way of accessing fit of the data. This is taught in every statistics 101 class. The growth of machine learning and prediction methods have now used residuals more in the toolbox to assessing the best prediction models.



Data visualizations are an integral part of the EDA process, enabling analysts to discern patterns and relationships in the data that would otherwise be difficult to discern from tabular data alone. Through data visualization, analysts can quickly identify trends, outliers, and other patterns that may be missed through numerical analysis alone. Moreover, visualizations facilitate the communication of findings to non-technical stakeholders, allowing them to comprehend complex data sets more efficiently. Through visualizations, analysts can also identify potential issues or biases in the data, resulting in better decisions and models. Thus, visualizations play a crucial role in the EDA process by enabling analysts to more effectively explore, comprehend, and communicate data-derived insights. During the initial EDA stage, an analyst may find that a variable or a covariate is directly related to the dependent variable when looking at a correlation heatmap or a scatterplot. The basic understanding can be formalized to visualize the discovery process.

The field of graphical communication, which is directly related to EDA, semiology, and their use in touch, has been a valuable tool and extension of the EDA thoughts that Tukey expressed. One of the fundamental principles of semiology is the relationship between signifier and signified, in which a visual element (the signifier) represents a particular meaning or concept (the signified), (**barthes1972?**). Another essential concept in semiology is using syntax and semantics to convey meaning in graphic communication effectively. This includes both the syntax and semantics of a graphic's visual elements, Monmonier (1985).

Using color to represent data on maps is an example of successful graphical communication utilizing semiology. By using different colors to represent different data points, viewers can comprehend patterns and relationships in the data quickly and easily. Jacques Bertin writes in "Semiology of Graphics" that color can be used to "emphasize a point, distinguish one category from another, or establish a relationship between two points", Monmonier (1985). In addition, Bertin explains that the use of color can help overcome language barriers, making it easier for the audience to comprehend the presented information.

The application of semiology in graphical communication is not devoid of obstacles. One difficulty is the possibility of misinterpretation, in which viewers may assign a different meaning to a visual element than was intended, Monmonier (1985). Another concern is the possibility of cultural differences in interpretation, in which a visual element may have a different meaning in one culture versus another, Norman (2013).

Despite these obstacles, semiology in graphical communication remains an indispensable tool for effectively conveying information. By understanding semiology principles and syntax and semantics' role in graphical communication, designers can create compelling visual representations that convey information clearly and concisely.

By utilizing visual elements such as bars and lines to represent data, graphs can make complex information more understandable to viewers. For instance, a line graph can be used to illustrate the change in the value of a stock over time, making it easier for investors to identify trends and patterns. Leland Wilkinson writes in his book “The Grammar of Graphics” that “graphical methods are not only superior to other forms of communication, but also superior to numerical or verbal methods for certain types of data and reasoning,” Wilkinson (2012).

It proposes that any statistical graphic can be broken down into a set of essential components, or “grammar,” that can be combined in different ways to create a wide range of visualizations, following a layered approach to describe and construct visualizations or graphics in a structured manner.

The components of the grammar of graphics include:

- Data: The raw data being visualized represents a set of observations or values.
- Aesthetic Mappings: The mapping of data variables to visual properties such as position, color, shape, and size.
- Scales: The mapping of data values to visual values, such as mapping a numerical value to a bar height.
- Geometries: The basic shapes representing the data, such as points, lines, bars, and histograms.

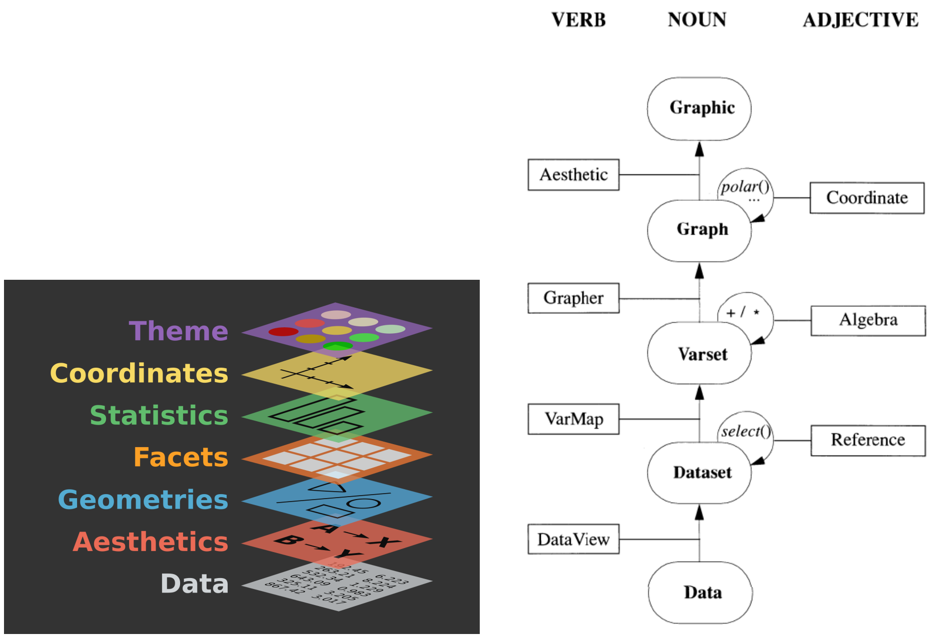


Figure 1.2: Grammar of Graphics Diagram of Wickham and Wilkinson's work

- Facets: The plot division into multiple subplots, each representing a different subset of the data.

For example, a bar chart can be created by mapping a categorical variable to the x-axis, mapping a numerical variable to bar heights, and using rectangular bars as the geometry. Moreover, mapping two numerical variables can create a scatter plot to the x and y positions and use points as the geometry. Finally, the "Grammar of Graphics" provides a systematic way of thinking about visualizations, making it easier to choose the appropriate visual representation for a given dataset.

Michael Friendly, a leading expert in data visualization, has utilized the principles of the grammar of graphics to develop innovative teaching methods that make complex data visualization concepts more accessible to a broader audience. Friendly has investigated the origins and development of graphic techniques, tracing their evolution from antiquity to the present. He used SAS with hands-on experiments to present categorical data analysis visually, Friendly (2014). By emphasizing the role of graphical methods in scientific discovery, Friendly has helped promote his use in various disciplines, from the natural sciences to the social sciences and beyond. In his book "Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization," Friendly provides a comprehensive overview of the key milestones in the evolution of statistical graphics, including the contributions of pioneers like William Playfair, Charles Minard, and John Tukey, Friendly & Denis (2001).

As discussed regarding semiology, Tukey's Exploratory Data Analysis (EDA), and the introduction of the Grammar of Graphics, we should be mindful that a well-constructed graphic can be misleading out of context. Compelling graphics can be a powerful tool for communicating complex information, making numerical accuracy, engagement, correct decision-making, and accurate predictions crucial.

In the context of data visualization, numerical accuracy refers to the precision and correctness of the numerical data displayed in graphics. Accurate graphics can assist users in comprehending complex numerical data and making more informed decisions, Cardoso, Leite, & Aquino (2016).

Engagement in the context of data visualization is the extent to which viewers are drawn to and interested in the displayed data. Engaging graphics can encourage users to interact with and explore data further, resulting in a more thorough comprehension of the data.

Correct decision making refers to the capacity of data visualization to enable users to make informed and precise decisions based on the presented information. Clear, accurate, and well-designed graphics can help users recognize patterns and insights, resulting in more effective decision making.

Correct predictions refers to the capability of graphics to accurately forecast or predict future events or outcomes. For accurate predictions, data visualization must include trustworthy data, sound statistical models, and efficient visualization techniques.

"Graphical Tests for Power Comparison of Competing Designs" by Hofmann et al. presents a graphical method for comparing the power of two or more competing designs in an experimental study (**hofmann2012?**). The article demonstrates that the graphical method is a useful tool for comparing the effectiveness of various experimental designs. It enables researchers to visualize and compare the effectiveness of different designs in an intuitive and straightforward manner.

Two methods for measuring the particle size distribution in a chemical process were compared in one study. The study evaluated both designs under various operating conditions and compared their power using a graphical method. The results demonstrated that one design was more effective at detecting differences in operating conditions than another.

Static Visualization is commonly used in the communication phase of data science workflows, and data scientists sometimes use them as part of the analysis. John Tukey's EDA methods are currently known and well-vetted in the field. However, Satyanarayan et al. addressed this by introducing a high-level grammar of graphics called "Vega-Lite," which presents a set of standardized linguistic rules for producing interactive information visualizations using a concise JSON format for data to be represented by the grammar Satyanarayan, Moritz, Wongsuphasawat, & Heer (2016). Vega-Lite has been directly implemented in R via the `ggvis` package using the same - albeit slightly lower-level.

Understanding cognitive load is crucial for designing compelling data visualizations, as it influences how users perceive, process, and remember the data presented in the visualization. When designing visuals, it is essential to consider the cognitive load they may place on the viewer. Cognitive load is the amount of mental effort required to process information, and minimizing it can enhance a graphic's effectiveness. In addition, displaying as much raw data as possible while minimizing cognitive load can improve the graphic's clarity and precision. Here are some general guidelines for making better graphics with works from Few ((**few2012?**)), Tufte ((**tufte1983?**)), and Cairo (Cairo (2016)):

1. Keep it simple - Avoid overwhelming the viewer with too much information at once by employing a clear and concise design with minimal distractions.

2. Use visual hierarchy - Utilize size, color, contrast, and placement to highlight important information and direct the viewer's focus.

3. Choose appropriate charts - Choose the chart type that best illustrates the data and facilitates comprehension.

4. Label clearly - Use labels that are clear and concise for axes, legends, and other essential information to avoid confusion.

5. Use data-to-ink ratio - Focus on the data by minimizing the amount of non-data ink, such as decorative elements or excessive grid lines.

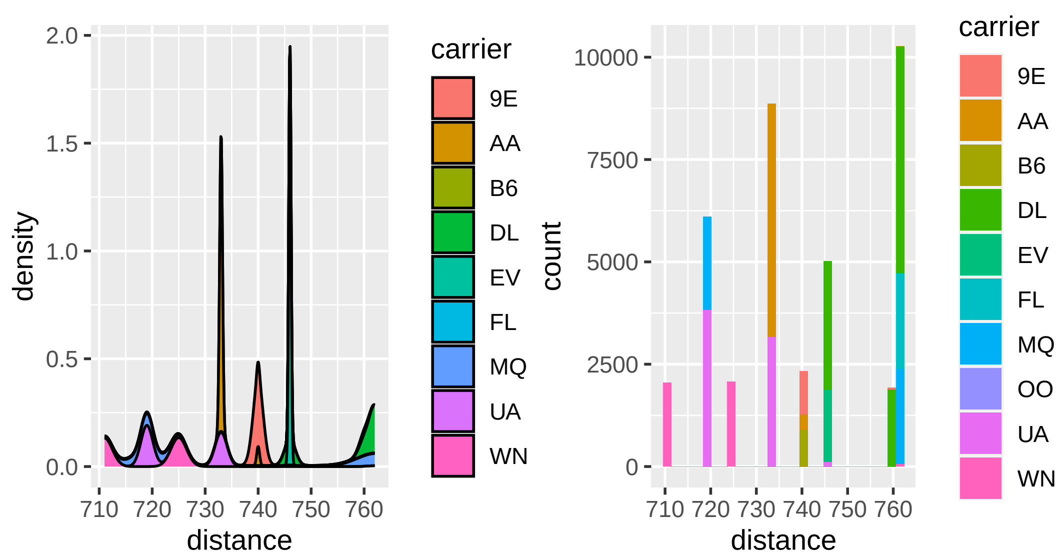
6. Avoid distortion - Use appropriate scaling and avoid distortions to ensure that the graphics accurately represent the data.

7. Provide context - Add context to assist the viewer in comprehending the significance of the data and its relevance to the topic.

```
## Warning: Groups with fewer than two data points have been
## dropped.

## Warning: Removed 1 rows containing missing values
## (`position_stack()`).

## `stat_bin()` using `bins = 30`. Pick better value with
## `binwidth`.
```



On the other hand, Interactive graphics provide a more dynamic and engaging way to explore and analyze complex data sets than traditional static visualizations. By allowing users to manipulate and explore data in real-time, interactive graphics can reveal hidden patterns and relationships that may be difficult to discern in static visualizations, making them a valuable tool for data analysis.

1.2.0.1 Interactive Graphics

As previously mentioned, theories behind visual representation include - graphical comprehension (Cleveland & McGill (1984)), preattentive processing (Ware (2012)), gestalt theory (Few (2009)), and graphical excellence (Edward R. Tufte (2001)). Interactive graphics offer a number of advantages when analyzing complex data sets, and technological progress has played a crucial role in making these tools more accessible and widely adopted, especially in interactive graphics. Compared to static visualizations, interactive graphics offer a more engaging and dynamic way to explore and analyze complex data sets. To gain a deeper understanding of the data, users are able to modify parameters, zoom in on specific regions, and rapidly explore various variables. In addition, interactive graphics offer a more intuitive method of communicating findings to non-technical users, making them a valuable asset for data-driven decision making. Interactive graphics are excellent for EDA; they are designed for exploring rather than presenting information (and more) and can be obtained by directly querying the graphic, Unwin, Volinsky, & Winkler (2003). Overall, interactive graphics are a potent data analysis tool, allowing analysts to gain a deeper understanding of complex data sets and to make more informed decisions.

The area of interactive graphics is still very much a work in progress despite existing as a field of research since the late 1960s. Developments are driven partly by new technology, such as `d3` (Bostock, Ogievetsky, & Heer, 2011).

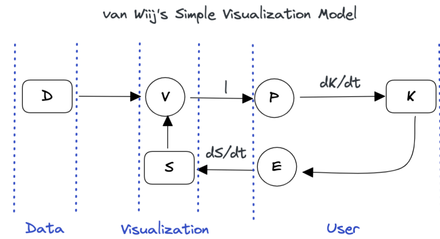


Figure 1.3: van Wiij Simple Visualization Model

Visualizations are more than just a picture. They are now a tool that facilitates analytic activity through different modes of interaction (Yi, Kang, Stasko, & Jacko, 2007). Visualization is context-free, as it can mean different things to different people depending on the situation (Parsons & Sedig, 2014).

In recent years, interactive graphics that enable users to manipulate and explore visualizations in real-time have grown in popularity and align with the van Wijk Simple Visualization Model framework. The van Wijk Simple Visualization Model is a diagrammatic representation that provides a simple and effective way to understand and visualize the flow of information and data through a system. van Wiij's simple visualization model shows how insights are generated as the human participates in a feedback loop between reading and interacting with visualization, Van Wijk (2005). It is a commonly used tool in EDA, the initial step in the data analysis. The van Wijk model can represent data flow from data sources, through intermediate processing stages, to the final visualization of results. This model is also context-free, allowing for the focus to be on the feedback loops between visualization and the user. The model helps to identify the various steps involved in the visualization process, from the collection and processing of data to the presentation of results. By doing so, it supports the design of more effective and user-friendly visualizations, which can enhance the overall user experience.

While the van Wijk Simple Visualization Model provides a valuable framework for designing compelling visualizations, it should not be used in isolation. Instead, effective data visualization requires a thorough understanding of human perception, cognition, and Human-Computer Interaction (HCI) principles. Later in this review, we will explore the fundamental principles of Human-Computer Interaction and how they can be applied to the design of compelling interactive graphics.

Based on the above best practices on the concept of cognitive load in graphics, the theory of manipulation of visualizations provides a set of guidelines and best practices for designing interactive graphics that minimize cognitive load and facilitate practical data analysis.

As interactive visualizations play a more significant role in information systems, designers must know what tasks, visual representations, and interaction techniques are available and how they work to facilitate analytical reasoning. They must decide on the most effective visual representation without being able to estimate every user's ability to read and interpret the visualization.

An influential framework developed by Vessey et al refers to the degree to which a person's cognitive abilities match the cognitive demands of a task, Vessey & Galletta (1991), in the field of cognitive load research has been improved by Heer and Bostock, providing examined how the complexity of interactive visualizations influences users' cognitive load, Heer & Bostock (2010). The authors discovered that more intricate visualizations tend to increase cognitive load, especially for users with lower visual literacy. They recommended that designers consider the cognitive load of interactive visualizations and strive to reduce complexity whenever possible.

While the direct relationship to traditional mobile devices, such as phone screens, will not be discussed, it is noteworthy to account for the impact of larger screen mobile devices, such as iPads and other portable devices, concerning a user's cognitive abilities. Eissele et al investigated the cognitive load of mobile interactive visualizations, Eissele, Weiskopf, & Ertl (2009). The authors discovered that the limited input options and small screen size of mobile devices can increase cognitive load in interactive visualizations. To reduce cognitive load on mobile devices, they suggested that designers employ suitable visual encoding techniques and simplify interactions.

Lastly, Toyama et al. investigated the impact of interactive elements on cognitive load in visualizations, Toyama, Sonntag, Orlosky, & Kiyokawa (2015). Based on the nature of the task and the user's familiarity with the interactive features, the authors discovered that interactive features can both increase and decrease cognitive load. They suggested that designers should evaluate the cognitive load caused by the interactive elements they include in visualizations

These studies suggest that designers should carefully consider the cognitive load implications of interactive graphics and strive, whenever possible, to reduce complexity. Simplifying interactions and implementing suitable visual encoding techniques can reduce cognitive load, especially on mobile devices. In addition, designers should evaluate the cognitive load caused by the interactive features they include in visualizations.

1.3 Audience-Data Interactions

Effective design of interactive graphics necessitates a comprehensive knowledge of Human-Computer Interaction (HCI) and User Experience (UX) design principles to ensure that the visualizations are engaging, informative, intuitive, and user-friendly.

Designing an effective interactive dashboard involves much more than simply selecting a set of visualizations and arranging them on a page. It requires a comprehensive knowledge of Human-Computer Interaction (HCI) and User Experience (UX) design principles and the ability to apply these principles to data analysis and visualization. Before designing an interactive dashboard, it is crucial to have a thorough understanding of the underlying HCI and UX frameworks and the specific needs and preferences of the intended audience. By understanding the principles of HCI and UX design, it is possible to create interactive dashboards that facilitate effective data analysis and offer a seamless and engaging user experience. This can result in greater engagement with the data, more profound insights, and better-informed decisions.

1.3.0.1 Human-Computer Interaction (HCI)

Human-Computer Interaction (HCI) refers to the study of the interaction between humans and computers. It encompasses the design, evaluation, and implementation of computer systems that are intuitive, efficient, and effective to use.

In the 1970s and 1980s, HCI researchers began to concentrate on developing user interface design theories and methodologies. The "GOMS" model, created by Card, Moran, and Newell in 1983, was an influential framework