

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

© Year, Denise Renee Bradford

## DEDICATION

Dedicated to...

## ACKNOWLEDGMENTS

Thank you to all my people!

## Table of Contents

<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 UNL thesis fields</b>	<b>1</b>
<b>2 Introduction</b>	<b>2</b>
2.1 Audience Considerations . . . . .	4
2.2 Data Considerations . . . . .	19
2.3 Audience-Data Interactions . . . . .	30
2.4 Dashboard Design . . . . .	35
2.5 Conclusion . . . . .	38
<b>3 Chapter Paper on Rural Shrink Smart Manuscript submitted to Journal of Data Science Special Issue</b>	<b>40</b>
3.1 Abstract . . . . .	40
3.2 Introduction . . . . .	40
3.3 Data Description . . . . .	40
3.4 Dashboard Design Considerations . . . . .	40
3.5 Guiding Design Principles . . . . .	40
3.6 Dashboard Design Process . . . . .	40

3.7	Discussion . . . . .	41
3.8	Future Work . . . . .	41
3.9	Conclusions . . . . .	41
<b>4</b>	<b>Chapter 2 Stuff</b>	<b>42</b>
4.1	Introduction . . . . .	42
4.2	Dashboard Design Considerations . . . . .	42
<b>5</b>	<b>Tables, Graphics, References, and Labels</b>	<b>43</b>
5.1	Tables . . . . .	43
5.2	Figures . . . . .	43
5.3	Footnotes and Endnotes . . . . .	43
5.4	Cross-referencing chapters and sections . . . . .	43
5.5	Bibliographies . . . . .	43
5.6	Anything else? . . . . .	43
	<b>Conclusion</b>	<b>44</b>
<b>A</b>	<b>The First Appendix</b>	<b>45</b>
<b>B</b>	<b>The Second Appendix, for Fun</b>	<b>47</b>
	<b>Colophon</b>	<b>48</b>
	<b>References</b>	<b>53</b>

## List of Figures

2.1	Visual Reference model by Card . . . . .	13
2.2	Grammar of Graphics Diagram of Wickham and Wilkinson's work	27
2.3	van Wijk Simple Visualization Model . . . . .	28



## List of Tables

## Chapter 1

### UNL thesis fields

Placeholder

## Chapter 2

### 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 based on the "Perceptual foundations in the design of visual displays" by [aspillaga1996] not all visual displays have equal perceptual value. 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 limita-

tions of the human perceptual system and are common to most potential consumers of the visualization: the sine illusion [vanderplas2015] 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 useful? 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.

We wish to discuss and analyze the exploration of using density plots or categorical plots as ”best” plots in the field. We will analyze statistical data and graphics alongside dashboard design, starting with Steven Few’s principles. It would be worthwhile to investigate the development of dashboards from static graphical representations in a practical setting to dynamic representations in an application setting. We would like to see the application of research to practice.

Consider a novel concept in research can be less practical or useful in the field more often than not. 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. We will adopt the concept of dynamically introducing the application to stakeholders with a static representation and

migrating them into an active application. Once this dynamic application is utilized or migrated to that space, it is necessary to comprehend why categorical representations are more critical than density representations in the field. In regards to speaking about the same data being represented, most of the data is then categorized. Is it because categorical shot charts in tables are perceived to be digestible or because density plots become cluttered and difficult to comprehend? Where do we draw the line between what is viewed as an impediment to the development of applicable graphics methodology?

## 2.1 Audience Considerations

Several factors, including prior experience, context, and personal preferences, can influence the human perception of density and categorical plots.

In general, density plots represent the distribution of continuous variables, while categorical plots represent the distribution of categorical or discrete variables. The interpretation of density plots typically involves understanding the shape of the distribution and the presence of any outliers or skewness. On the other hand, categorical plots aim to show the frequency or count of observations in each category.

When comparing density plots and categorical plots, it's essential to consider the data analysis type. For continuous data, a density plot may provide a more informative representation of the distribution. A categorical plot may provide a more intuitive representation of the distribution of categorical data.

However, for various reasons, some people may prefer one type of plot over the other. For example, some may find density plots more visually appealing, while others may prefer categorical plots for their simplicity and ease

of interpretation. Ultimately, the choice between the two types of plots will depend on the context and purpose of the analysis and the preferences of the individual interpreting the data.

Do we comprehend why we utilize the graphics we do? Is there a reason we adopt the perception of a previously used graphic? How do we determine whether there are superior graphical representations when the human brain has determined that this is what I comprehend? How do we develop statistical methodologies based on effective testing of statistical graphics on human perceptions of excellent and good? How do we determine the line distinguishing a density plot from a categorical plot? Consequently, how is this decision made outside of academia and research? In one way, research has been conducted on the visual or human perception of categorical plots.

In addition to graphical elements alongside numerical values and plots such as Parallel Coordinate Plots (PCPs). We wish to discuss and analyze the exploration of using didn't report plots or categorical plots as density plots in the field. We will analyze statistical data and graphics alongside dashboard design, starting with Steven Few's whites. It would be worthwhile to investigate the development of dashboards from static graphical representations in a practical setting to dynamic representations in an application setting. This is the application of research to practice.

What may be considered an exciting thought process in research that is less practical or useful in the field? 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. We will adopt the concept of dynamically

introducing the application to stakeholders with a static representation and migrating them into an active application. Once this dynamic application is utilized or migrated to that space, it is necessary to comprehend why categorical representations are more critical than density representations in the field. We're speaking about the same data being represented. However, most of the data is then categorized. Is it because categorical shot charts in tables are perceived to be digestible or because density plots become cluttered and difficult to comprehend? Where do we draw the line between what is viewed as an impediment to the development of applicable graphics methodology?

### 2.1.1 Perception

To read reality from images is to solve a problem: a series of extremely difficult problems that persist throughout an individual's lifetime. Errors are illusions. 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] Human perception plays a direct role in the area of visualization and graphics. Data Analysis tasks closely resemble the cognitive process known as sensemaking. Tukey and Wilk highlight the role of cognitive processes in their initial descriptions of Exploratory Data Analysis (EDA) [@tukey1966]. Mallows and Walley list psychology as one of four areas likely to support a theory of analysis [@mallows1980]. Data analyses rely on the mind's ability to learn, analyze, and understand. Assigning meaning 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.

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.

Perceptual principles refer to how the human brain processes and interprets visual information. 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.

Some examples of perceptual principles include:

- Proximity: Objects close to each other are perceived as related or grouped.
- 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 are just a few examples of the many perceptual principles that designers use to create compelling visual displays. These principles are based on cognitive psychology and understanding how the human brain processes visual information.

Visual inference uses our ability to detect graphical anomalies. However, the idea of formal testing remains the same in visual inference – with one exception: The test statistic is now a graphical display compared to a “reference distribution” of plots showing the null.

### **2.1.2 Expertise**

Subject matter graphics are visual representations of information and data that help communicate complex concepts, ideas, and information to audiences. This expertise is related to knowledge of the topic or subject being presented in the graphic. A subject matter expert is someone who has a deep understanding of how people perceive and interpret visual information. Subject matter experts (SMEs) have deep knowledge and experience in a specific field, such as science, finance, or medicine. SMEs provide the necessary information and insights to form the graphic’s basis. They ensure that the content presented is accurate, relevant, and informative. They use their human perception and cognition knowledge to design graphics that effectively communicate information to the intended audience. Perception experts in graphics consider a range of factors, including color, shape, size, texture, and spatial relationships, to create graphics that are visually appealing and easy to understand.

Perception experts in graphics may use different theories and principles to inform their design decisions. For example, they may use Gestalt principles

to create visually harmonious and organized graphics. They may also use the principles of visual hierarchy to guide the audience's attention to the essential information in the graphic.

Perception experts in graphics may also conduct user research to understand how different audiences perceive and interpret visual information. They may use eye-tracking and user testing techniques to evaluate the effectiveness of different graphic designs.

A perception expert plays a crucial role in designing graphics that effectively communicate information to the intended audience. Their knowledge of human perception and cognition, as well as their expertise in graphic design, allows them to create visually appealing and easily understood graphics.

### **2.1.3 Engagement with the data**

<!-- This is a good paragraph – but I'm not sure it's in the right place. --> 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. <!-- What about organizations which aren't profit-driven? They still need to understand the relationship between different variables in the dataset. --> Additionally, data analysis can support scientific research by helping researchers to test hypotheses, develop theories, and gain a deeper understanding of complex

phenomena. Ultimately, the goal of data analysis is to turn data into actionable insights and information that can inform and improve decision making.

Data Scientists and other analytic professionals often use interactive visualization in the dissemination phase at the end of a workflow, during which findings are communicated to a wider audience. Digital tools are critical to data science and analytics workflows, and current practice spans the following:

- *Data Analysis Tools* - R, Pandas and SAS
- *Data Warehousing Services* - MySQL, MongoDB, or Amazon Redshift
- *Machine Learning Libraries* - scikit-learn or Apache MLlib

A typical process typically consists of the following general stages:

1. *Discovery*: Formulating an exciting question and determining the data necessary to answer it.
2. *Acquisition*: Locating, organizing, and preparing data to be accessible to the chosen analysis environment.
3. *Exploration*: Investigating and analyzing the data set to collect insights and understand the data
4. *Modeling*: Building fitting and validating a model that can explain the data set and the observed phenomena
5. *Communication*: Disseminating the results to stakeholders in reports, presentations, and charts.

Visualizing Complex Data is representing complex, multi-dimensional data in a graphical or pictorial form to make it easier to understand and interpret.

The goal is to turn large, intricate data sets into visually appealing and intuitive representations, such as charts, graphs, maps, and other types of visualizations, to help identify patterns, trends, and relationships that may not be immediately apparent from raw data. This process can help to communicate complex information effectively and make data-driven decisions.

William Cleveland’s subcycle plots:

- glyph maps and binned graphics emerging from big data visualization efforts.
- glyphs and other plots have been embedded in maps

Bertin’s Semiotic of Graphics is a seminal work in the academic study of visualization. Glyphmaps have been developed as a tool for tracking climate and climate change data (Wickham, Hofmann, Wickham, & Cook, 2012); (hobbs2010?). (Schulz, Nocke, Heitzler, & Schumann, 2013) define two abstractions for the design of visualizations:

$$Data + Task = Visualization$$

$$Data + Visualization = Task$$

These abstractions demonstrate dependence between the data, visual representation, and the task. The more the user interacts with the visualization, they gain knowledge. The interactions allow the user to control their understanding by providing the flexibility to create new views that help them go

beyond just the visual representation [Kiem2008]. The field of information visualization is continually adapting to changes with the big data revolution.

Data Scientists and Statisticians have produced more graphics since the pandemic's start. The reasons someone will create a graphic or dashboard may include but are not limited to understanding raw data structures to analyze model assumptions and present predictions, along with displaying key performance metrics of business logic. These goals help work to navigate and are best served by quick-and-dirty representations of the data, while highly polished graphics may be more useful in other situations. It is valuable and essential to convey data correctly, meaning that we need to understand how graphics are perceived on a dashboard on a general level. Previous research by Tukey focused on graphics as a tool for exploratory analysis. Tukey describes in *Exploratory Data Analysis* [Tukey1966] that pictures are often used to display data in a more enhanced version than a table. Tukey outlines detailed the types of different graphics and in which situations to utilize these graphics. The article - "External cognition: how do graphical representations work?" by Scaife and Rogers [Scaife1996] critique the disparate literature on graphical representations, focusing on four representative studies. In general, this will help in the psychology of the perceptual experience.

The visual reference model developed by Card et al. [Card] describes and identifies the three phases of the visualization process.

While External Cognition describes the advances in graphical technology and how little had been done in the work of the cognitive framework of the discipline, the following citations by Ware attempt to develop the necessary guidelines that are useful for the work done by the perceptual experience.

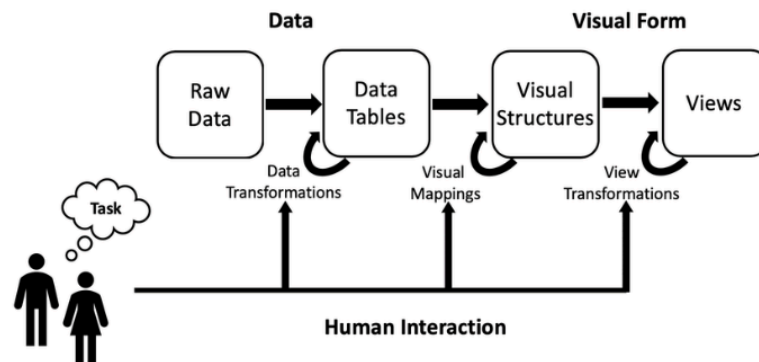


Figure 2.1: Visual Reference model by Card

Colin Ware “Information Visualization: Perception for Design” (Ware, 2004) there are four stages of visualization

- The collection and storage of data itself
- The preprocessing design to transform the data into something we can understand
- The display hardware and the graphics algorithms produce the image on the screen.
- The human perceptual and cognitive system

The overlapping understanding in the field, while Ware takes the process one step further not just to allow the end user to understand the outcomes but to curate the outcomes with a visual perception of the data that makes the cognitive load easier for the end-user.

A dashboard has much information related to tabular data from multiple sources. Design should be recorded a produced with content that will allow for reproducibility. A dashboard should have some content related to interaction with the user. This interaction can be in multiple forms: toggling through the

selection of variables to display uniformly to the use of interaction on the graph and allowing a user to understand best what is being shown in the diagram.

The entire dashboard/Interface should have a human perception piece that is useful for the user to comprehend and use. For example, the dashboard/interface could be more practical if the user is visually overwhelmed.

- Identified the highly relevant from a dashboard design perspective (O'Donnell & David, 2000)
1. Information systems give interaction and feedback
  2. Type of presentation format to be used
  3. Differences in the amount of information load.

Information load is essential, as dashboards must provide the right decision cues without overwhelming the user with excess information.

"A decision cue is a feature of something perceived that is used in the interpretation of perception" [choo2009], where perception is an inferential process as objects in the environment can only be perceived indirectly through available information that has been sensed by the individual [brunswik1952].

Visual complexity and information utility are required. Visual complexity refers to the "degree of difficulty in providing a verbal description of an image ([heaps1999], [olivia2004]).

- Graphs are more suitable for spatial tasks (ex., for comparing a set of values) (Vessey & Galletta, 1991);(Umanath & Vessey, 1994);(Vessey, 1994)

- Graphs reduced the negative influence of information overload
- Graphs produced better correlation estimates and decreased time on a task (schulzand?);(Booth & Siegler, 2006)
- Self-organizing maps and multidimensional scaling did not significantly outperform tabular representations (Huang et al., 2006)

The purposes of a dashboard:

1. Consistency
2. Monitoring
3. Planning
4. Communication

Card stated "Use interactive visual representations of abstract, non-physically based data to amplify cognition." [card1999]

Visual perception involves two elements - the perceptual and conceptual gist. The perceptual gist refers to the process of the brain when it determines the image properties that provide the structural representation of a scene, like color and texture. The conceptual gist refers to the scene's meaning, which is improved after the perceptual information is received ([@friedman1979]; [olivia2004]).

Visual complexity might increase with the quality and range of objects and with varying material and surface styles [heylighen1997].

Repetitive and uniform patterns and existing knowledge of the objects in the scene reduce visual complexity [olivia2004].



If the guidelines on our visual information load can be related to statistical terminology, we can consider this the fisher information of visual information load. This load can be used as a metric for balancing the amount of information rather than overloading the consumer.

### Visual Data Mining

For data mining to be effective, it is important to include humans in the data exploration process and combine the flexibility, creativity, and general knowledge of the computational power of today's computers. Visual data exploration aims at integrating humans in the data exploration process, applying their perceptual abilities to the large data sets available in today's computer systems [Kiem2002].

The visual data exploration process can be seen as a hypothesis generation process:

- The visualizations of the data allow the user to gain insight into the data and come up with new hypotheses
- Along with verification of the hypotheses

The main advantages of visual data exploration over automatic data mining techniques from statistics or machine learning are:

- visual data exploration can efficiently deal with highly inhomogeneous and noisy data.
- visual data exploration is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters.

Visual Exploration Paradigm, also known as MGV (Massive Graph Visualizer), is an integrated visualization and exploration system for massive multidigraph navigation [Abello2002]. MGV usually follows a three-step process:

- overview first
- zoom and filter
- details-on-demand

The user identifies interesting patterns and focuses on one or more of them. Note that visualization technology does not only provide the base visualization techniques for all three steps but also bridges the gap between the steps.

Visualization Technique Classification:

- Standard 2D/3D displays such as bar charts and x-y plots
- Geometrically transformed displays, such as landscapes and parallel coordinates, as used in a scalable framework
- Icon-based displays such as needle icons and star icons as used in MGV
- Dense pixel displays such as the recursive pattern and circle segments techniques and the graph sketches as used in MGV
- Stacked displays, such as treemaps or dimensional stacking

Interaction and distortion techniques allow users to interact directly with the visualizations.

- Projection as used in the Grand Tour System
- Filtering as used in Polaris
- Zooming as used in MGV and scalable framework

- Linking and Brushing as used in Polaris and the scalable framework

### Design Theory in Information System

The knowledge is distinguished as the fifth of five types of theory:

1. Analyzing & describing
2. Understanding
3. Predicting
4. Explaining and predicting
5. Design and action

A definition of information systems that are suitable for our purposes concerns: "the effective design delivery use and impact of information technology in organizations and society [Avison1995].

The two paradigms characterize much of the research in the Information systems discipline:

- *Behavioral Science Paradigm* - seeks to develop and verify theories that explain or predict human or organizational behavior (roots in natural science research methods).
- *Data Science Paradigm* - seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts (roots in engineering and the sciences of the artificial) (Simon, 1996).

Technology and behavior are not dichotomous in an information system. They are inseparable [Lee2000]. Information technology (IT) artifacts are broadly defined as:

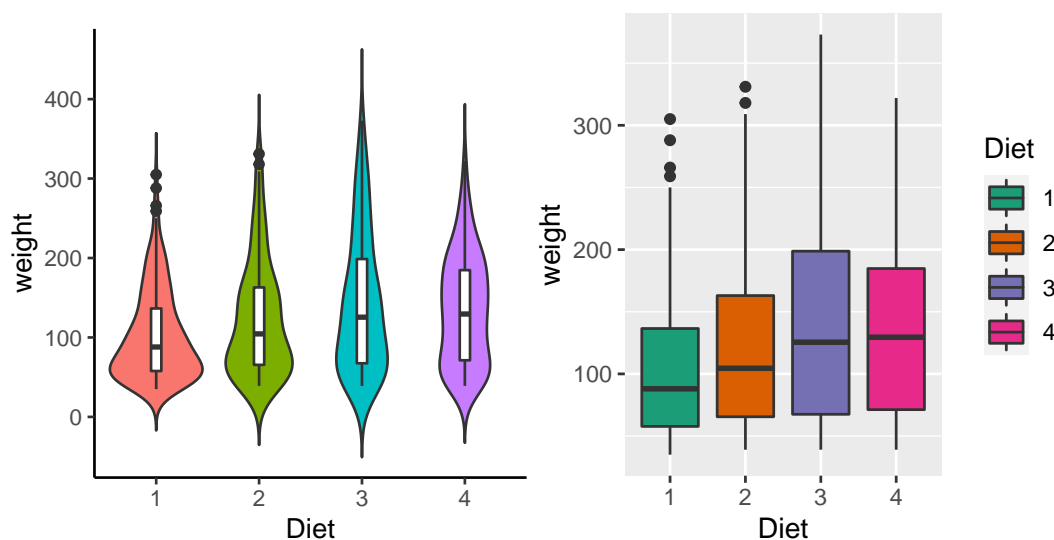
- constructs (vocabulary & symbols)
- models (abstractions & representations)
- methods (algorithms & practices)
- instantiations (implemented & prototype systems)

These are concrete prescriptions that enable IT researchers and practitioners to understand and address the problems inherent in developing and successfully implementing information systems within organizations ([@march1995]; [@nunamaker1991]).

## 2.2 Data Considerations

Exploratory Data Analysis (EDA) analyzes and summarizes a dataset to discover patterns, trends, and insights. It is a crucial step in the data analysis process and is often used to identify which variables are essential, what the data looks like, and what the underlying structure of the data is. EDA is typically done using various techniques, such as visualizations, statistical summaries, and data transformations.

John Tukey was the first to organize the collection and methods associated with philosophy into Exploratory Data Analysis (EDA). John Tukey, creator of stem-and-leaf plot, boxplot-resistant smooth, and the violin plot (also known as rootgram) who taught us to utilize these methods to organize and demonstrate EDA. 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 in EDA:

1. Graphical exploration looking for patterns or displaying fit.
  - The method demonstrates things about data that are not understood by a single numeric metric. 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.

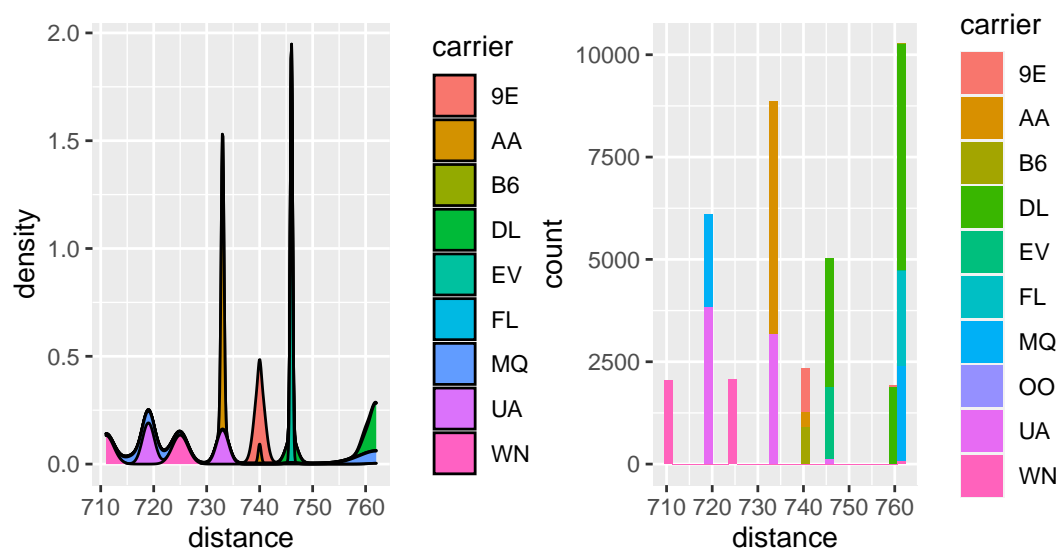
4. The mostly known parts of EDA but is done in the way of accessing fit of the data. This is taught in every statistics 101 class. The growth of machine learning and prediction methods have now used residuals more in the toolbox to assessing the best prediction models.

- The idea generally is to determine the deviations in the data from a general pattern by looking at the data from the fit of the data.}

```
## 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, Dashboards are interactive interfaces that display data visually to provide insights and support decision-making. Dashboards can be used to monitor key performance indicators, track progress over time, and identify patterns and trends in data. They often display real-time data and can be customized to show the most relevant data to the user.

Interactive graphics, are central to EDA [unwin1999]. Beyond the limitations of static statistical displays, interactive graphics enable visualizations to advance alongside the analysis. User interaction and direct manipulation are required for dynamic graphics to reach their full potential (cook1995; unwin1999). The connection between EDA and dashboards is that EDA is the process of preparing and understanding the data, which is the first step for building a dashboard, as the data has to be cleaned, transformed, and analyzed to be used efficiently on the dashboard. EDA results can be used to identify the most relevant data and metrics to include in the dashboard and to design the visualizations that will be used to display the data. And also, the EDA process can be used to identify the outliers, patterns, trends, and insights that will be useful to show in the dashboard to support decision-making.

Visualizations of data are essential for exploratory data analysis (EDA) along with model diagnostics. Plots for EDA are a valuable tool for guiding an analyst in discovering the relationships between variables in their data. When using plots in model diagnostics, plots help analysts determine whether or not the model is an appropriate way to model. 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. This will be important to know before starting a linear model analysis. Much of our general understanding is from introductory statistics courses. The basic understanding can be formalized to visualize the discovery process.

### 2.2.1 Variable Types

Not all variables should be used to create all types of charts - variable type informs chart structure. Categorical Data Visualization is the process of visualizing data that can be divided into distinct categories or groups. Categorical data are non-numeric and often represented by words, labels, or symbols, such as gender, product type, color, etc.

Visualizing categorical data helps uncover patterns and relationships between categories and can provide insights into the data distribution. Categorical data visualization techniques include bar charts, pie charts, histograms, stacked bar charts, and others. The choice of visual representation will depend on the data's nature and the insights being sought. The goal of categorical data visualization is to communicate the information effectively and make it easier to understand and interpret.

Friendly detailed using SAS with hands-on experiments to present categorical data analysis visually [friendly2014]. Researchers have used PCPs to visualize categorical data. Beygelzimer, Perng, Ma and Hellerstien created a fast ordering categorical data analysis algorithm that helped visualization, where their algorithms helped organize the original parallel coordinate plots clearer ([beygelzimer2001]; [ma2001]). Hammock plots are modified versions of similar coordinate plots invented by Schonlau to visualize categorical data [schonlau2003]. His design replaces coordinate polygons with rectangles to present the number. Treemaps are modified to support categorical data visualization.

CatTree gives a hierarchical categorical data visualization with interaction [kolatch2001]. Fernstad developed an interactive system combining parallel



coordinates, tables, and scatterplot matrices for an overview explorative analysis. Thoroughly research categorical data visualization to support algorithm understanding. The novel contingency wheel presented by [alsallakh2011] supports visual analytics in categorical data, and he measured association based on Pearson's residuals and used visual abstraction based on elements frequency.

High-Dimensional Data Visualization represents complex data sets with many variables or features (also known as high-dimensional data). The goal is to find effective ways to express such data in a form that allows easy understanding, analysis, and interpretation. This is typically achieved by reducing the dimensionality of the data, for example, by projecting the data onto a lower-dimensional space or by aggregating the data in some way. The resulting visualizations can then reveal patterns, relationships, and other insights that would otherwise be difficult to detect from the raw data. For example, high-dimensional data visualization techniques include scatter plots, parallel coordinate plots, heat maps, and many others.

A popular research area in visualization since high-dimensional data is always fuzzy to mining. Direct visualization includes geometric visualizations:

- scatterplots: use dots in coordinate to present data points.
- parallel coordinates: present each dimension as axes, and every data item intersects dimensions as a polygon line at a particular position
- RadViz/ PolyViz
- GridViz

Besides these traditional geometric visualization methods, iconographic

displays like human faces and star glyphs used funny ways to present multivariate data. Hierarchical methods are used widely in parallel coordinates, which give analysts an intuitive view of clustering information [fua1999]; [johansson2005]. Rearrange the dimensions by dimension similarity on parallel coordinates, circle segments, and recursive patterns [ankerst1998]. Guo used an interactive feature selection method to help users identify interesting subspaces from high-dimensional data sets [guo2003].

### 2.2.2 Grammar of Graphics

Frame as a way to easily change between appropriate forms of presentation for a given variable or set of variables.

It doesn't help us decide which is better - for that we need user testing.

The grammar of graphics (gg of ggplot2) is a theory that is well-defined for creating statistical graphics with work from Wilkinson [Wilkinson1999] and Hadley Wickham [ggplot2]. The Grammar of Graphics is a framework for understanding the structure of statistical graphics developed by Leland Wilkinson. 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.

Grammar of graphics is defined as the framework which follows 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.
- **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. For example, 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.

### 2.2.2.1 Interactive Graphics

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’ [bostock2011]. Visualizations are more than just a picture. They are now a tool that facilitates analytic activity through different modes of interaction [yi2007]. Visualization is context-free, as it can mean different things to different people depending on the situation [parsons2014].

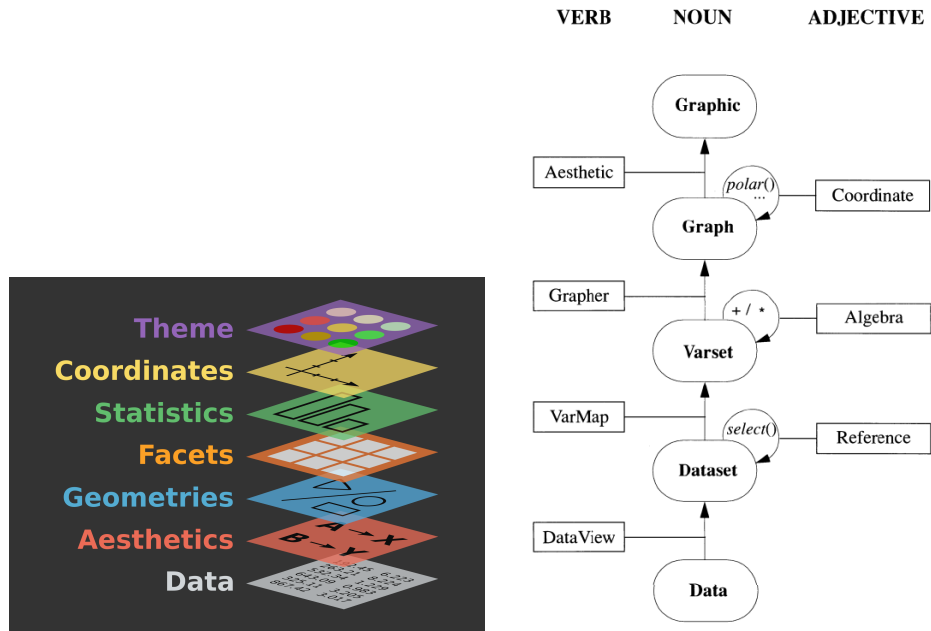


Figure 2.2: Grammar of Graphics Diagram of Wickham and Wilkinson's work

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. It is a commonly used tool in Exploratory Data Analysis (EDA), which is the initial step in the data analysis process. The van Wijk model can be used to represent the flow of data from data sources, through intermediate processing stages, to the final visualization of results. van Wijk's simple visualization model shows how insights are generated as the human participates in a feedback loop between reading and interacting with visualization [van2005].

This model is also context-free, allowing for the focus to be on the feedback loops between visualization and the user.

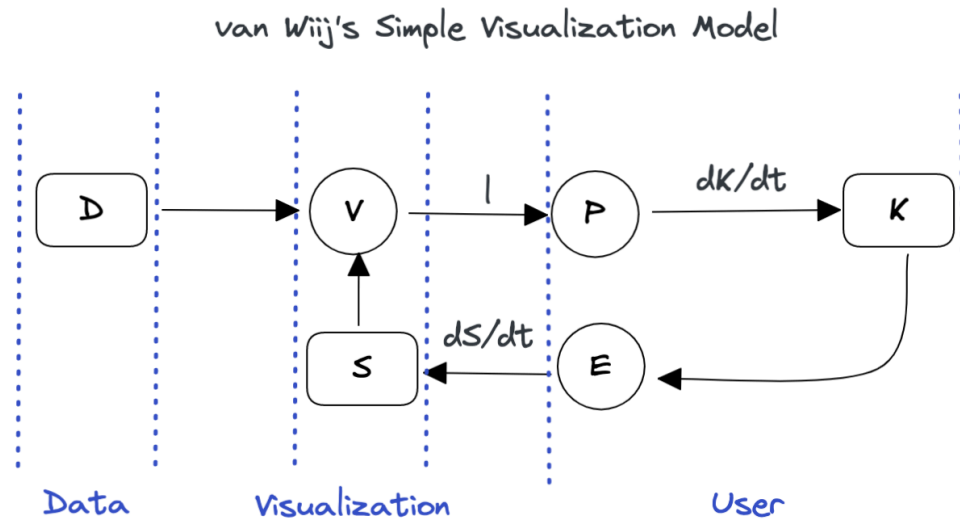


Figure 2.3: van Wijk Simple Visualization Model

Interaction allows the user to define what data they see and how they see it, creating a dialogue between the user and the system. Theories behind visual representation include:

- graphical comprehension ((Cleveland & McGill, 1984))
- preattentive processing ((Ware, 2012))
- gestalt theory ((Few, 2009))
- graphical excellence ((**tufte2001?**))

Theories behind the manipulation of visualizations include but are not limited to:

- cognitive fit ((Vessey & Galletta, 1991))
- visual perceptual approaches ((**baker2009?**))
- human information processing

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.

Interactive visualization is a commonly used tool in Exploratory Data Analysis (EDA), which is the initial step in the data analysis process. The goal of EDA is to gain a high-level understanding of the data and to identify patterns, relationships, and anomalies in the data.

Visualizations have become more effective in recent years due to the pandemic and the Johns Hopkins University COVID-19 Dashboard [JHPHDDashboard] [Dashboard](<https://coronavirus.jhu.edu/map.html>). We were glued to our computers, TVs, and phones for most of the world. As a result, we watched the dashboard change in real-time to adapt to the users' needs. In part to data growth and changing, the dashboard, as well as the visualizations, were needed in a condensed platform. The need to be concise and vastly informative is a struggle regarding data visualizations.

The human brain can only take in a set amount of data from a table or a paragraph. [references?](#) The space of infographics has been a much better way of looking at data on a creative scale.[@stat\_graph\_hist] While this may be a way of seeing the data in a friendly way, infographics need to include an interactive piece of data that many people would like to explore.

The relationship between the van Wijk Simple Visualization Model and Human-Computer Interaction (HCI) lies in the area of data visualization. The van Wijk Model provides a framework for understanding how information and

data flow through a system to be displayed to the user. In the context of HCI, the model helps to understand how data is processed, transformed and presented to the user in a way that is intuitive, informative and engaging. 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.

}

### 2.2.3 Multivariate Data Displays

- Difficulty of visualizing multivariate data
- Different approaches - briefly mention
  - encoding with shape/color/etc.
  - tours
  - pcps
  - linked graphics (interactivity)

This section doesn't need to solve the problem - it just introduces it

## 2.3 Audience-Data Interactions

User analysis is a crucial step in the design of user interfaces, especially in Human-Computer Interaction (HCI) and User Experience (UX) design. It involves studying the users of a system to understand their needs, goals, and behaviors. The purpose of user analysis is to create interfaces that are easy to use, efficient, and effective for the intended audience.

Several methods can be used to conduct user analysis:

- Interviews: This involves conducting in-depth interviews with users to understand their needs, goals, and workflows.
- Surveys involve sending out surveys to users to gather data on their needs and preferences.
- Observations: This involves observing users as they complete tasks with the interface to understand their behaviors and workflows.
- Personas: This is a method of creating fictional characters that represent the different types of users; it helps to understand the users' needs and goals.
- Scenarios: This is a method of creating stories that describe how a user might interact with a system; it helps to understand the context and the user's needs.
- Focus groups: This is a method of gathering a small group of users and facilitating a discussion to understand the users' needs and goals.
- Contextual inquiry: This is a method of visiting the users in their work environment and observing them while they work; it helps to understand the context and the user's needs.

The results of user analysis can be used to inform the design of the interface, including the layout, navigation, and functionality. It can also be used to identify areas of the interface that are confusing or difficult to use and to make recommendations for improvements. User analysis is an iterative process, and it should be done in multiple stages of the design process to ensure that the final product is tailored to the users' needs.



### 2.3.1 Testing static graphics

Brief overview of user studies that compare different types of graphics, accuracy, etc.

Static Visualization is commonly used in the communication phase of data science workflows, and data scientists sometimes use them as part of the analysis. For example, John Tukey’s EDA methods are currently known and well-vetted in the field. However, Satyanarayan et. al began to address 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 [satyanarayan2016]. Vega-Lite has been directly implemented in R via the ‘ggvis’ package using the same - albeit slightly lower-level. Several protocols and methods can be used to test the design of a dashboard or other visual display. Some of these include:

- Usability testing: This involves having users interact with the dashboard and providing feedback on its usability, including how easy it is to navigate, understand, and use.
- Cognitive walkthrough: This involves having experts in human-computer interaction evaluate the dashboard, focusing on the cognitive processes required to use it effectively.
- Eye-tracking: This involves using technology to track the users’ gaze and interactions, to understand the user’s focus and attention on the dashboard elements.
- A/B testing: This involves creating two versions of the dashboard, each with slightly different design elements, and comparing the results to see

which design is more effective.

- Surveys: This involves asking users to complete a survey that measures their satisfaction and understanding of the dashboard and to provide feedback on improvements.
- Heat maps: This involves using heat maps to track where users are clicking on the dashboard to identify which elements are being used the most and which are being ignored.
- Card sorting: This is a method to understand the users' mental models and how they would like to organize and categorize the data; it is helpful to understand how to structure the dashboard navigation.

These are just a few examples of the many protocols and methods that can be used to test the design of a dashboard. The selection of the appropriate way will depend on the specific goals of the testing and the resources available.

Remember that the user's feedback is crucial in the testing process; it will provide the necessary insight to improve the design and make it more efficient.

### 2.3.2 Testing interactive graphics

Overview of testing methods for interactive graphics - talk aloud, measuring user engagement, etc. Testing interactive graphics using human perception principles in psychology involves considering various aspects of human perception and cognition to evaluate the effectiveness and user experience of the graphics. Here are some steps you can follow to test your interactive graphics:

Usability testing: Conduct a usability test to assess the ease of use and accessibility of the graphics. This includes testing for navigation, user control, and overall user experience.

Perception of visual elements: Evaluate how the visual elements in your graphics, such as color, contrast, size, and position, impact the perception of the information being presented.

Cognitive load: Assess the cognitive load on the user by evaluating how easily they can process and understand the information being presented. Factors such as complexity, amount of information, and type of information presented can impact cognitive load.

Attention allocation: Observe where the user's attention is directed while interacting with the graphics. This can help identify areas that may require improvement or modification to better engage the user's attention.

Memory retention: Evaluate the user's ability to retain and recall information presented in the graphics. This can help determine the effectiveness of the design in supporting memory retention.

User feedback: Obtain user feedback through surveys, interviews, or focus groups to gain insights into the user experience and identify areas for improvement.

It is important to keep in mind that human perception and cognition can vary greatly between individuals and can be influenced by a range of factors, such as age, culture, and personal experience. Testing with a representative sample of your target audience can help ensure that your interactive graphics are optimized for a wide range of users.

## 2.4 Dashboard Design

Given that the audience has limitations, there are design constraints around the data, and the ability of the audience to successfully use the graphical displays of the data, what can we take from this body of research that applies to more complicated sets of graphics?

How do we maintain user attention, desire to explore, and accurately communicate the data through the medium of an interactive data dashboard?

A dashboard is a visual display of the essential information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance [a few]. Dashboards have particular characteristics:

- Achieve specific objectives
- Fits on a single computer screen
- Information can be displayed in multiple mediums (web browser or mobile device)
- Can be used to monitor information at a high level

Dashboards can present various statistical data, such as financial performance, website traffic, or customer engagement metrics. They allow users to quickly and easily understand complex data sets using visual elements such as charts, graphs, and tables to display the information. Additionally, statistics can be used to analyze data presented on a dashboard, providing insights into trends and patterns that can inform decision-making.

While a dashboard can be handy, it may be worth describing that a poorly designed dashboard will not be used. A dashboard should be concise, clear, and intuitive when displaying components in combination with a customized list of requirements of users.

Much of the work done within statistical research and dashboard design involves collaboration with other researchers and users. While this may be the best for the growth of the discipline, one will find that working with collaborators with non-STEM backgrounds. Dashboards can help understand and support many data types in essential business objectives. There are many different ways to label and utilize dashboards in different kinds.

Dashboards are cognitive tools that should be used to improve understanding of data, which should help people visually find relationships, trends, patterns, and outliers. Most importantly, dashboards should leverage people's visual cognitive capabilities.

EDA refers to methods and procedures for exploring the data space to learn about a data set. By analogy, exploratory modeling analysis (EMA) refers to methods and procedures for exploring the space of models which may be fit to a data set.

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 [unwin2003].

- PCPs enable the display of multi-dimensional data in two-dimensional space.

- There must be some loss of information, but this can be partly counteracted by varying the order of the axes.
- Interactivity is valuable for reordering the axes flexibly and fast.
- Interaction is valuable for dealing with the dense mass of lines produced by large data sets.

Being able to select subgroups of cases, highlight the chosen lines, and switch between different subgroups all assist in interpreting the otherwise intricate displays which arise.

Cowan suggested that the average person can only hold two to six pieces of information in their attention [Cowan2001]. People can develop detailed understandings of reality, which is infinitely complex.

Cognitive structures consist of mental models and their relationships ([Rumelhart1976], [Carley1992], [Jonassen1996]). A schema is a mental model containing a breadth of information about a specific object or concept. Schemas are organized into semantic networks based on their relationships to other schemas [Wertheimer1938], [Rumelhart1976]. This arrangement helps the brain process its experiences instead of storing every sensory observation; the brain only needs to maintain its schemas, which are good summaries of all previous observations. Some "memories" may even be complete recreations built with a schema [Bartlett1932], [Klein2007]

Wixon introduce "contextual design" as a systems development method in which the researcher partners with the user at the user's place of work to "develop a shared understanding" of the user's activities, and they define contextual inquiry as the first part of the broader process [Cowan2001]. Con-

textual inquiry is the data collection step of the field research element of the contextual design method, and it emphasizes four essential principles:

1. The context of the activity being performed by the user
2. The partnership between the researcher and the participant
3. The spoken verification that the investigator's interpretation of the activity matches the user's
4. The focus of the study is central to the approach taken by the interviewer

Kandal conducted what might be considered a contextual interview study similar to ours in that they analyzed data scientists' self-reported work processes [Kandel2012]. They propose three main archetypes that data scientists may be classed into the following:

- Hackers: who build processes chaining together multiple programming languages of different types (analytical, scripting, and database languages) and use visualization in various environments.
- Scripters: who perform most of their analysis in an analytical environment (e.g., R or Python) and execute the most complex statistical modeling of the types but who do not perform statements their ETL
- Application Users: who performed most or all of their work in an application such as Excel or SPSS and, like scripters, relied on others (namely, their organizations' IT departments) for ETL.

## 2.5 Conclusion

In this dissertation, I address this question of design for dashboards, as well as tools which can be used to support the display of multivariate data

in an interactive context. Chapter 1 presents a thorough review of the literature regarding graphical and human computer interaction/UI-UX methods. Chapter 2 will explore the process of designing dashboards for public use through parallel coordinate plots as a central component to data exploration to make decisions. Chapter 3 focuses on graphical methods for multidimensional categorical variables and visualization methods have for growth. We conclude with a Shiny application that facilitates a better understanding of the possible forms a parallel coordinate plots in exploratory data analysis can take by accommodating a through examination through variables and structural changes to the parallel coordinate plot with a click of a mouse. Chapter 4 further explores multidimensional categorical data visualizations and develops an approach to using parallel coordinate plots to assess predictive model. We identify visual indicators for parameters in different models and extend the connection between parallel coordinate plots of binary tables and odds ratios to include logistic regression models with categorical variables.



## **Chapter 3**

**Chapter Paper on Rural Shrink Smart Manuscript  
submitted to Journal of Data Science Special Issue**

Placeholder

### **3.1 Abstract**

### **3.2 Introduction**

### **3.3 Data Description**

### **3.4 Dashboard Design Considerations**

### **3.5 Guiding Design Principles**

### **3.6 Dashboard Design Process**

#### **3.6.1 Dashboard Components**

#### **3.6.2 Initial Draft**

#### **3.6.3 Redesign**

### **3.7 Discussion**

### **3.8 Future Work**

### **3.9 Conclusions**

## Chapter 4

### Chapter 2 Stuff

Placeholder

#### 4.1 Introduction

#### 4.2 Dashboard Design Considerations

##### 4.2.1 Guiding Design Principles

## Chapter 5

### Tables, Graphics, References, and Labels

Placeholder

#### 5.1 Tables

#### 5.2 Figures

#### 5.3 Footnotes and Endnotes

#### 5.4 Cross-referencing chapters and sections

#### 5.5 Bibliographies

#### 5.6 Anything else?

## Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

### More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

## Appendix A

### The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readability and/or setup.

#### In the main Rmd file

```
library(knitr)
library(palmerpenguins)
library(tidyverse)
library(nycflights13)
data(flights)

library(ggpcp)
library(ggplot2)
library(dplyr)
data(nasa)

library(scales)
library(datasets)
data("ChickWeight")
```

In Chapter 5:

## Appendix B

### The Second Appendix, for Fun



## Colophon

This document is set in [EB Garamond](#), [Source Code Pro](#) and [Lato](#). The body text is set at 11pt with *lmr*.

It was written in R Markdown and  $\LaTeX$ , and rendered into PDF using [huskydown](#) and [bookdown](#).

This document was typeset using the XeTeX typesetting system, and the [University of Washington Thesis class](#) class created by Jim Fox. Under the hood, the [University of Washington Thesis LaTeX template](#) is used to ensure that documents conform precisely to submission standards. Other elements of the document formatting source code have been taken from the [Latex](#), [Knitr](#), and [RMarkdown templates for UC Berkeley's graduate thesis](#), and [Dissertate: a LaTeX dissertation template to support the production and typesetting of a PhD dissertation at Harvard, Princeton, and NYU](#)

The source files for this thesis, along with all the data files, have been organised into an R package, `xxx`, which is available at <https://github.com/xxx/xxx>. A hard copy of the thesis can be found in the University of Washington library.

This version of the thesis was generated on 2023-03-01 12:43:16. The repository is currently at this commit:

The computational environment that was used to generate this version is as follows:

```
## - Session info -----
## setting value
## version R version 4.1.0 (2021-05-18)
## os      macOS Big Sur 10.16
## system  x86_64, darwin17.0
## ui      X11
## language (EN)
## collate en_US.UTF-8
## ctype   en_US.UTF-8
## tz      America/New_York
## date    2023-03-01
## pandoc  2.11.4 @ /Applications/RStudio.app/Contents/MacOS/pandoc/ (via rmarkdown)
##
## - Packages -----
## package      * version date (UTC) lib source
## assertthat    0.2.1   2019-03-21 [2] CRAN (R 4.1.0)
## backports     1.4.1   2021-12-13 [1] CRAN (R 4.1.0)
## bookdown      0.29.1   2022-09-18 [1] Github (rstudio/bookdown@4890be2)
## broom         0.8.0   2022-04-13 [2] CRAN (R 4.1.2)
## cachem        1.0.6   2021-08-19 [1] CRAN (R 4.1.0)
## callr         3.7.0   2021-04-20 [2] CRAN (R 4.1.0)
## cellranger    1.1.0   2016-07-27 [2] CRAN (R 4.1.0)
## cli           3.4.1   2022-09-23 [1] CRAN (R 4.1.2)
## colorspace    2.0-3   2022-02-21 [1] CRAN (R 4.1.2)
## crayon        1.5.2   2022-09-29 [1] CRAN (R 4.1.2)
## DBI           1.1.2   2021-12-20 [1] CRAN (R 4.1.0)
## dbplyr        2.1.1   2021-04-06 [2] CRAN (R 4.1.0)
## devtools      2.4.4   2022-07-20 [1] CRAN (R 4.1.2)
```

```

## digest          0.6.30 2022-10-18 [1] CRAN (R 4.1.2)
## dplyr           * 1.0.10 2022-09-01 [1] CRAN (R 4.1.2)
## ellipsis        0.3.2  2021-04-29 [2] CRAN (R 4.1.0)
## evaluate        0.18   2022-11-07 [1] CRAN (R 4.1.2)
## fansi           1.0.3  2022-03-24 [1] CRAN (R 4.1.2)
## farver          2.1.1  2022-07-06 [1] CRAN (R 4.1.2)
## fastmap         1.1.0  2021-01-25 [2] CRAN (R 4.1.0)
## forcats         * 0.5.1  2021-01-27 [1] CRAN (R 4.1.0)
## fs              1.5.2  2021-12-08 [1] CRAN (R 4.1.0)
## generics        0.1.3  2022-07-05 [1] CRAN (R 4.1.2)
## ggpcp           * 0.2.0  2022-03-22 [1] Github (heike/ggpcp@8c02618)
## ggplot2         * 3.3.6  2022-05-03 [2] CRAN (R 4.1.2)
## glue            1.6.2  2022-02-24 [1] CRAN (R 4.1.2)
## gtable          0.3.0  2019-03-25 [2] CRAN (R 4.1.0)
## haven           2.5.0  2022-04-15 [2] CRAN (R 4.1.2)
## hms             1.1.1  2021-09-26 [1] CRAN (R 4.1.0)
## htmltools       0.5.3  2022-07-18 [1] CRAN (R 4.1.2)
## htmlwidgets    1.5.4  2021-09-08 [1] CRAN (R 4.1.0)
## httpuv         1.6.6  2022-09-08 [1] CRAN (R 4.1.2)
## httr           1.4.3  2022-05-04 [1] CRAN (R 4.1.2)
## jsonlite       1.8.3  2022-10-21 [1] CRAN (R 4.1.2)
## knitr          * 1.41   2022-11-18 [1] CRAN (R 4.1.2)
## labeling       0.4.2  2020-10-20 [2] CRAN (R 4.1.0)
## later          1.3.0  2021-08-18 [1] CRAN (R 4.1.0)
## lifecycle      1.0.3  2022-10-07 [1] CRAN (R 4.1.2)
## lubridate      1.8.0  2021-10-07 [1] CRAN (R 4.1.0)
## magrittr       2.0.3  2022-03-30 [1] CRAN (R 4.1.2)
## memoise        2.0.1  2021-11-26 [1] CRAN (R 4.1.0)
## mime           0.12   2021-09-28 [1] CRAN (R 4.1.0)
## miniUI         0.1.1.1 2018-05-18 [1] CRAN (R 4.1.0)
## modelr         0.1.8  2020-05-19 [2] CRAN (R 4.1.0)

```

```

## munsell          0.5.0   2018-06-12 [2] CRAN (R 4.1.0)
## nycflights13    * 1.0.2   2021-04-12 [1] CRAN (R 4.1.0)
## palmerpenguins * 0.1.0   2020-07-23 [1] CRAN (R 4.1.0)
## pillar          1.8.1   2022-08-19 [1] CRAN (R 4.1.2)
## pkgbuild        1.3.1   2021-12-20 [1] CRAN (R 4.1.0)
## pkgconfig       2.0.3   2019-09-22 [2] CRAN (R 4.1.0)
## pkgload         1.3.0   2022-06-27 [1] CRAN (R 4.1.2)
## prettyunits     1.1.1   2020-01-24 [2] CRAN (R 4.1.0)
## processx        3.5.3   2022-03-25 [2] CRAN (R 4.1.2)
## profvis         0.3.7   2020-11-02 [1] CRAN (R 4.1.0)
## promises        1.2.0.1 2021-02-11 [1] CRAN (R 4.1.0)
## ps              1.7.0   2022-04-23 [2] CRAN (R 4.1.2)
## purrr           * 0.3.5   2022-10-06 [1] CRAN (R 4.1.2)
## R6               2.5.1   2021-08-19 [1] CRAN (R 4.1.0)
## RColorBrewer     1.1-3   2022-04-03 [1] CRAN (R 4.1.2)
## Rcpp             1.0.9   2022-07-08 [1] CRAN (R 4.1.2)
## readr           * 2.1.2   2022-01-30 [1] CRAN (R 4.1.2)
## readxl          1.4.0   2022-03-28 [2] CRAN (R 4.1.2)
## remotes         2.4.2   2021-11-30 [1] CRAN (R 4.1.0)
## reprex          2.0.1   2021-08-05 [2] CRAN (R 4.1.0)
## rlang           1.0.6   2022-09-24 [1] CRAN (R 4.1.2)
## rmarkdown       2.16    2022-08-24 [1] CRAN (R 4.1.2)
## rstudioapi      0.14    2022-08-22 [1] CRAN (R 4.1.2)
## rvest           1.0.2   2021-10-16 [1] CRAN (R 4.1.0)
## scales          * 1.2.0   2022-04-13 [2] CRAN (R 4.1.2)
## sessioninfo     1.2.2   2021-12-06 [1] CRAN (R 4.1.0)
## shiny           1.7.3   2022-10-25 [1] CRAN (R 4.1.2)
## stringi         1.7.8   2022-07-11 [1] CRAN (R 4.1.0)
## stringr         * 1.4.1   2022-08-20 [1] CRAN (R 4.1.2)
## tibble          * 3.1.8   2022-07-22 [1] CRAN (R 4.1.2)
## tidyr           * 1.2.0   2022-02-01 [1] CRAN (R 4.1.2)

```

```

## tidyselect      1.2.0   2022-10-10 [1] CRAN (R 4.1.2)
## tidyverse      * 1.3.1   2021-04-15 [2] CRAN (R 4.1.0)
## tzdb           0.3.0   2022-03-28 [1] CRAN (R 4.1.0)
## urlchecker     1.0.1   2021-11-30 [1] CRAN (R 4.1.0)
## usethis        2.1.6   2022-05-25 [1] CRAN (R 4.1.2)
## utf8           1.2.2   2021-07-24 [2] CRAN (R 4.1.0)
## vctrs          0.5.1   2022-11-16 [1] CRAN (R 4.1.2)
## withr         2.5.0   2022-03-03 [1] CRAN (R 4.1.2)
## xfun           0.35    2022-11-16 [1] CRAN (R 4.1.2)
## xml2           1.3.3   2021-11-30 [1] CRAN (R 4.1.0)
## xtable         1.8-4   2019-04-21 [2] CRAN (R 4.1.0)
## yaml          2.3.6   2022-10-18 [1] CRAN (R 4.1.2)
##
## [1] /Users/dbradford4/Library/R/x86_64/4.1/library
## [2] /Library/Frameworks/R.framework/Versions/4.1/Resources/library
##
## -----

```

## References

Placeholder

- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.
- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.
- Huang, Z., Chen, H., Guo, F., Xu, J. J., Wu, S., & Chen, W.-H. (2006). Expertise visualization: An implementation and study based on cognitive fit theory. *Decision Support Systems*, 42(3), 1539–1557.
- O'Donnell, E., & David, J. S. (2000). How information systems influence user decisions: A research framework and literature review. *International Journal of Accounting Information Systems*, 1(3), 178–203.
- Schulz, H.-J., Nocke, T., Heitzler, M., & Schumann, H. (2013). A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2366–2375.

- Simon, H. A. (1996). The sciences of the artificial 3rd ed. *MIT Press Cambridge*.
- Umanath, N. S., & Vessey, I. (1994). Multiattribute data presentation and human judgment: A cognitive fit perspective. *Decision Sciences*, 25(5-6), 795–824.
- Vessey, I. (1994). The effect of information presentation on decision making: A cost-benefit analysis. *Information & Management*, 27(2), 103–119.
- Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1), 63–84.
- Ware, C. (2004). *Information visualization: Perception for design*. San Francisco, CA: Morgan Kaufmann Publisher.
- Ware, C. (2012). *Information visualization: Perception for design*. Morgan Kaufmann.
- Wickham, H., Hofmann, H., Wickham, C., & Cook, D. (2012). Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, 23(5), 382–393.