

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

List of Figures

List of Tables

Chapter 1

Introduction

This is my dissertation that drives me into my imposter syndrome so deeply
that it's insane!!

Chapter 2

Literature Review

2.1 Overview

Visualization can be used in a vastly different ways, with this in mind, we will start with the vocabulary of visualization. There are two general umbrellas of visualizations - Data Visualization (Scientific, Information and Infographic) and Visual Analytics (Used to analyze data). In which Scientific, the visualization of scientific data that have close ties to real-world objects with spatial properties. Information visualization represent data sets that don't have inherent spatial components. Infographic are combining various statistics and visualizations with a narrative. Visual Analytics are using visualizations to analyze data, which include more formal statistical tests. The data are visuals, such as fMRI brain images and other pieces in medical fields.

A list of Data Visualizations can be broken down into the following:

- One-Dimensional (i.e. Linear)
- Two-Dimensional (i.e. Geospatial)
- Three-Dimensional (i.e. Response Surfaces)

- Temporal
- Multi-Dimensional
- Trees/Hierarchical
- Network

Each type in the above list have a

Visualizations of data are important for exploratory data analysis (EDA) along with model diagnostics. Plots for EDA are a useful tool for guiding an analyst in discovering the relationships between variables in their data. In the case when using plots in model diagnostics, plots help analysts determine whether or not the model is 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 in a way that will visual discovery process.

Visualizations have become more effective in more recent years due to the pandemic and the Johns Hopkins University COVID-19 Dashboard Insert the Dashboard, for most of the world, we were glued to our computers, TVs and phones. As a result, we watched as the dashboard changed in real time to adapt to the needs of the users. In part to data growing and changing, the dashboard as well as the visualizations were needed in a condense platform. The need to be concise along with vastly informative can be a bit of a struggle when it comes to data visualizations. The human brain is capable of only taking in a set amount of data at a time from a table or a paragraph. The

space of infographics have been a much better way of looking at data on a creative scale. While this may be a way of seeing the data in a nice way, infographics miss the interactive piece about data that many people would like to explore.

2.2 Visual Inference

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

2.2.1 Applications

Two general applications in areas of visual inference have developed since the work of (**Buja?**). These applications are actual methodology and methodology based on protocols. Actual methodology applications are used with alternative and corresponding null hypotheses, which perform visual inference tests to show many participants of different backgrounds with lineups.

2.3 Parallel Coordinate Plot Visualization

Parallel coordinate plots have been implemented in analysis software since the mid 1980s ((**Inselberg?**), (**Wegman?**)). Parallel Coordinate Plot Visualization, also called parallel plot, is an established field of high dimensional visualization of xy coordinate analysis. Parallel coordinate plots have such a unique structure that will allow for data structures with n-dimensions. As data grows in exploratory data analysis, the use of data visualizations will