

STA 141 - Exploratory Data Analysis and Visualization

Derek L. Sonderegger

August 22, 2019

Contents

Preface	5
1 Introduction	7
1.1 Initial thoughts.	7
1.2 Amazing Graphics	7
1.3 Bad Graphs	7

Preface

These are the lecture notes for STA 141 - Exploratory Visual Data Analysis. This course is intended to teach students how to think critically about problems, examine data that can provide answers, and create graphs that are insightful, and ask follow-up questions to the visual analysis. Also, because disinformation can be disguised to seem credible, students will also be exposed to various data visualization tricks and statistical malarky that propagandists.

This course draws information from several websites and books.

- Alberto Cairo's book *The Truthful Art*. Alberto Cairo is the Knight Chair in Visual Journalism at the School of Communication of the University of Miami (UM), where he heads specializations in infographics and data visualization. You can find more about him at thefunctionalart.com. Also at his website, his blog highlights data visualizations in the news. Many of the graphs I will shown in this class are featured here.
- Claus O. Wilke has a book *Fundamentals of Data Visualization*. He also has a nice R package called *cowplot* that aids in making publication ready graphs using R's `ggplot2`.
- *Calling Bullshit* This is a 3-credit course taught at Univeristy of Washington. Their goal is to teach students to recognize bullshit provide another scientist a reason why a claim is bullshit.
- Amelia McNamara's SDS 136 course. Amelia is an Assistant Professor at University of St Thomas and I've been influenced by her presentations at national conferences. These are her notes from a data visulization course that she teaches. She is also pretty fun to follow on twitter.

Chapter 1

Introduction

1.1 Initial thoughts.

- What is data?
- Why visualize it?
 - See relationships that raw data obscure.
 - Cognitive work to translate raw numbers into context between other data points are already done for you.
 - Good graphics translate particular numerical relationships into physical relationships which our brains are really good at processing.

1.2 Amazing Graphics

- UK Drug Poison These are the data from the UK related to drug overdose and misuse. Farther down the page gives a similar graph on suicide.
- Migration Patterns in Europe.

1.3 Bad Graphs

- While the New York Times is generally really quite good, this graph is quite misleading. Another view of the data is more fair, although the still depressing.
- I feel compelled to show a bad 3-d Excel graphic as well.
- Reuter's infamous gun deaths in Florida chart.

- An amusing case where a pie chart is ridiculous. This is the result of a survey that asks what pizza toppings are liked. In particular, a person can pick more than one topping and so the percentages don't sum to 100%.