

Introduction to Data Science

Daniel Gutierrez, Data Scientist
Los Angeles, Calif.

Course Outcomes

- Ability to perform data visualization
- In conjunction with EDA, be able to product simple plots involving the transformed data sets
- Produce exploratory plots
- Produce expository plots

Lesson Objectives

- Learn to use `hist()`, `boxplot()`, `barplot()`, density plots, scatterplots with `plot()`, `qqplot()`, heatmaps with `image()`
- Explore big data visualization techniques: random sample, `smoothscatter()`, count bins with `hexbin()` and `plot()`
- Techniques for additional variables: color, size of data point, plot symbols
- Missing value plots
- Correlation plots with `pairs()`
- Expository plots with axis labels, legends, titles, multiple panels
- Create plot PDF and image files

Data Visualization

- Why do we use graphs
 - To understand data properties
 - To find patterns in data
 - To suggest modeling strategies
 - To “debug” analyses
 - To communicate results

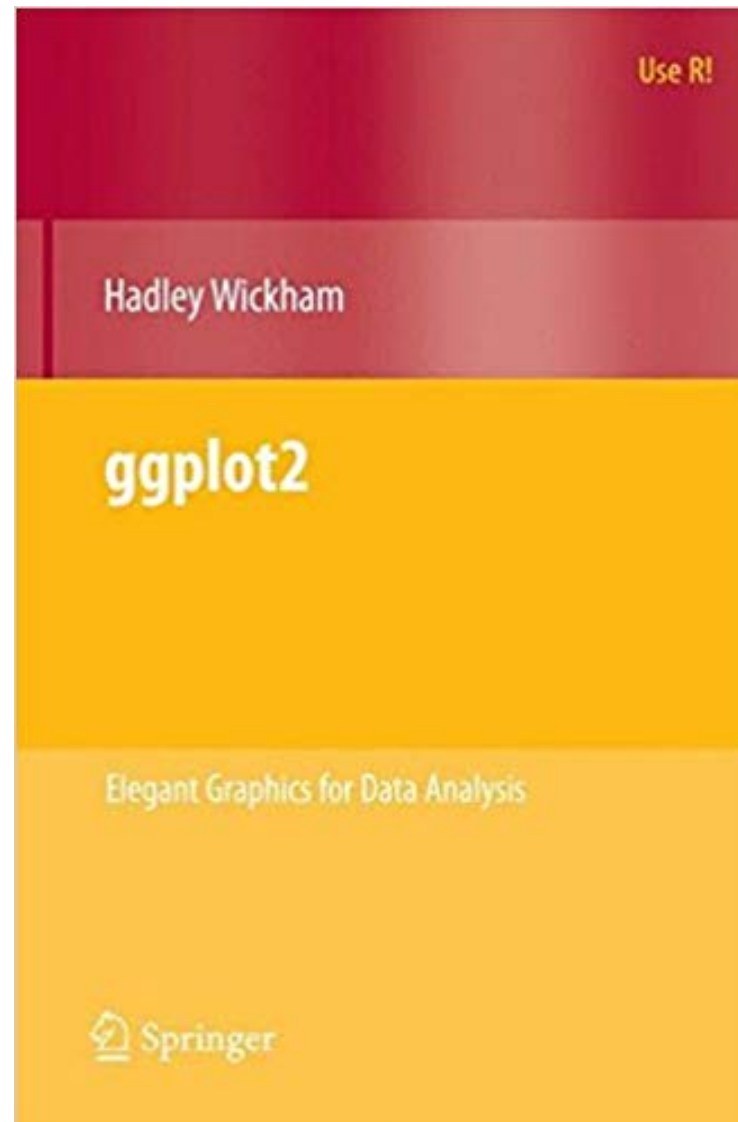
Data Visualization

- Exploratory graphs
 - They are made quickly
 - A large number are made
 - The goal is for personal understanding
 - Axes/legends are generally not cleaned up
 - Color/size are primarily used for information

Data Visualization

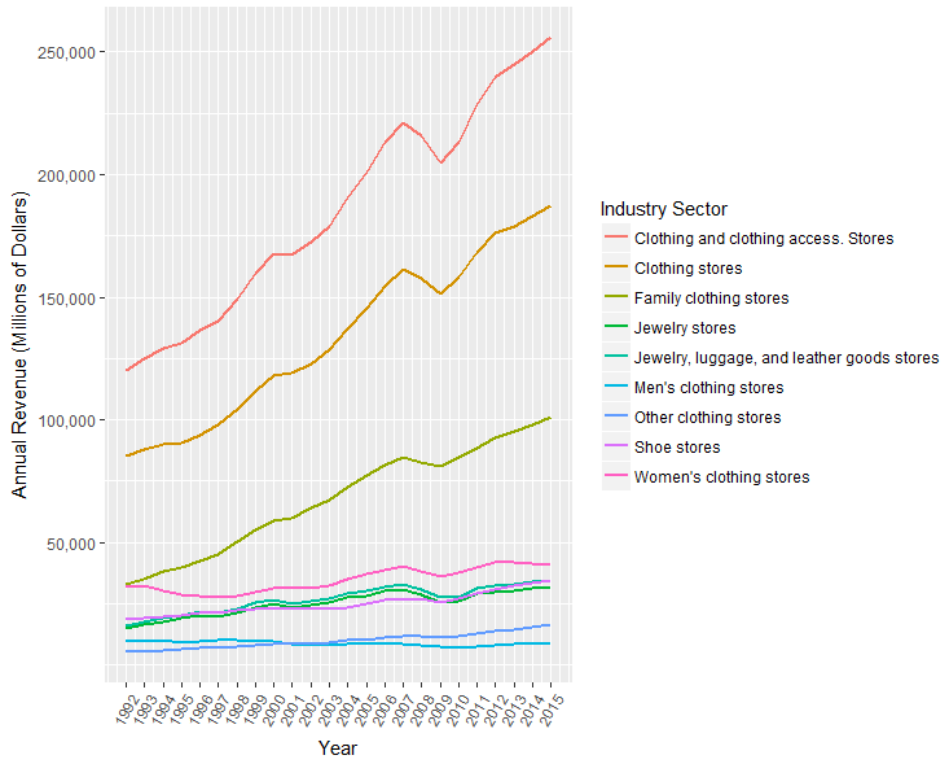
- Expository graphs
 - The goal is to communicate information
 - Information density is generally good
 - Color/size are used both for aesthetics and communication
 - Expository figures have understandable axes, titles, and legends

Data Visualization

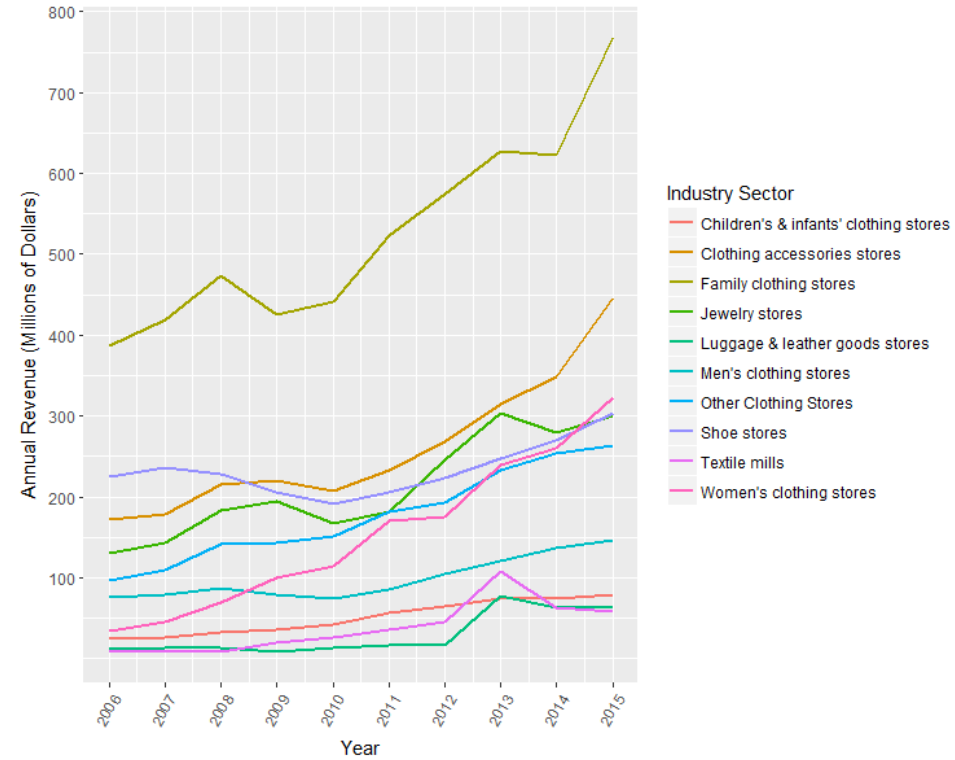


Data Visualization

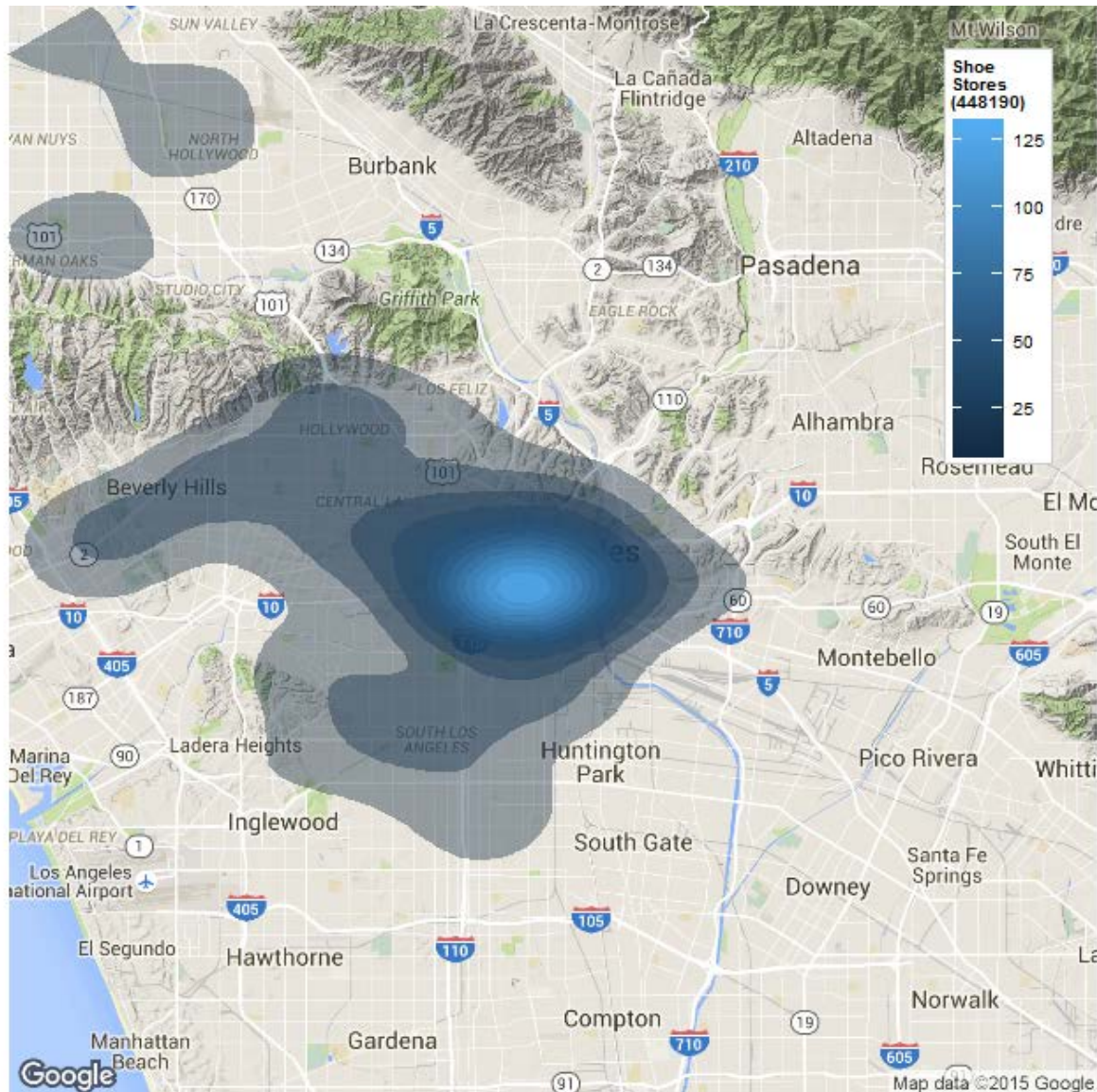
US Apparel Revenue by Industry Sector (1992-2015)



LA Apparel Revenue by Industry Sector (2006-2015)



Data Visualization



Code modules

- WEEK 8-1 Code module – Histograms
- WEEK 8-2 Code module – Boxplots
- WEEK 8-3 Code module – Barplots
- WEEK 8-4 Code module – Density plots
- WEEK 8-5 Code module – Scatterplots
- WEEK 8-6 Code module – QQ-plots
- WEEK 8-7 Code module – Heatmaps
- WEEK 8-8 Code module – Missing value plots
- WEEK 8-9 Code module – Expository plots
- WEEK 8-10 Code module – Creating PDF and PNG

Summary

- In WEEK 8 of Introduction to Data Science, we built up our toolbox of data visualization methods in order to gain familiarity with a data set.
- Now that you've seen base R graphics, you'll want to explore the `ggplot2` package.
- The methods discussed represent a small sample of available techniques. As you progress as a data scientist, you'll pick up more data visualization methods that will help out in this step of the data science process.