Exploratory Data Analysis Notes

Coursera Course by John Hopkins University

INSTRUCTORS: Dr. Jeff Leek, Dr. Roger D. Peng, Dr. Brian Caffo

Contents

Lesson 5: ggplot2 <3   4

Lesson 6: Hierarchical Clustering   5

Lesson 7: K-Means Clustering & Dimension Reduction   5

Lesson 8: Working with Color in R Plots   5

Quiz 3 Scribbles   5

Case Studies   5

Quiz 4 Scribbles   5

Course Project 2   5

Intro

Instructor's Note

"*This course covers the essential exploratory techniques for summarizing data. These techniques are typically applied before formal modeling commences and can help inform the development of more complex statistical models. Exploratory techniques are also important for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. We will cover in detail the plotting systems in R as well as some of the basic principles of constructing data graphics.*
We will also cover some of the common multivariate statistical techniques used to visualize high-dimensional data.

All the best,

Roger Peng"
Welcome to Exploratory Data Analysis (EDA)

Introduction

Exploratory Data Analysis with R Book

The Art of Data Science

Installing R on. . .

Windows

Mac

Installing R Studio (Mac)

Setting Your Working Directory on. . .

Windows

Mac

Lesson 1: Graphs

Principles of Analytic Graphics

Lesson with `swirl()`: Principles of Analytic Graphs

Exploratory Graphs

Lesson with `swirl()`: Exploratory Graphs

Lesson 2: Plotting

Plotting Systems in R

Base Plotting System

Base Plotting Demonstration

Lesson with `swirl()`: Plotting Systems

Lesson with `swirl()`: Base Plotting System

Lesson 3: Graphics Devices

Graphics Devices in R

Lesson with `swirl()`: Graphics Devices in R

Quiz 1 Scribbles

Lesson 6: Hierarchical Clustering

Part 1

Part 2

Part 3

Lesson with `swirl()`: Hierarchical Clustering

Lesson 7: K-Means Clustering & Dimension Reduction

K-Means Clustering (Part 1)

K-Means Clustering (Part 2)

Lesson with `swirl()`: K Means Clustering

Dimension Reduction (Part 1)

Dimension Reduction (Part 2)

Dimension Reduction (Part 3)

Lesson with `swirl()`: Dimension Reduction

Lesson with `swirl()`: Clustering Example

Lesson 8: Working with Color in R Plots

Part 1

Part 2

Part 3

Part 4

Quiz 3 Scribbles

Case Studies

Clustering Case Study

Air Pollution Case Study

Lesson with `swirl()`: CaseStudy

Quiz 4 Scribbles