R for Data Analysis and Visualization

ECON 396 (Fall 2017)

TR 10:30-11:45, Webster Hall 112

Jonathan Page

2017-06-01

Contents

Syllabus

Office Hours

Monday 2-3 PM and Tuesday 3-4 PM, or by appointment, Saunders 509, jrpage at hawaii dot edu.

Student Learning Objectives

1. To be familiar with standard techniques for visualizing data, including heat maps, contour plots, (look at the list in the Data Visualization book)
2. To be able to transform raw data into formats suitable for analysis
3. To be able to perform basic exploratory analysis
4. To be able to create data visualizations

There is no prerequisite for this course.

Resources

Required

Introductory Statistics with Randomization and Simulation: Available as a free PDF (https://www.openintro.org/stat/textbook.php?stat_book=isrs) or for $8.49 on Amazon.

R Graphics Cookbook

Recommended:

ggplot2 Cheatsheet

Course Requirements

Grades for this course will be based on weekly assignments (30%), project assignments (30%), the project proposal (5%), the final project deliverable (15%), and final project presentation participation (20%).

Weekly assignments (30%)

Weekly assignments are short R excercises. Each exercise should take no longer than 15 minutes. You will typically be given time to complete the exercise in class the day the assignment is given. The assign-

ment will be in the form of R Markdown file (*.Rmd). You will submit the completed assignments via classroom.google.com by the following class period.

Individual Project

Project assignments (30%)

Each week, leading up to the project proposal, you will be given an assignment that is designed to provide you with an organized workflow for approaching new data science projects. Project assignments are submitted via classroom.google.com, with the exception of the two presentations

Project proposal presentation (5%)

This presentation should be less than 2 minutes. You simply need to communicate the core question your project seeks to answer and the dataset(s) you will be using to answer this question.

Final project (15%)

The final project will be an R Markdown document which communicates your project question, the data you used, and your results.

Final project presentation participation (20%)

Your final project participation grade is based on a combination of your own presentation and the feedback you provide to your classmates.

Schedule

The following schedule is tentative and subject to change.

Week 1

- **R** Intro to R and RStudio; Histograms, scatterplots, summary statistics
- **Topic** Data sources overview
- **Project Assignment** Indentify interesting datasets and questions

Week 2

- **R** read_csv, dplyr basics, heatmaps, hexbins
- **Topic** Anscombe's Quartet
- **Project Assignment** Choose question and dataset for your project

Week 3

- **R** ggplot facets, bubble plots, transparency
- **Topic** Effective Data Visualization
- **Project Assignment** Write description of your question

Week 4

- **R** geom_smooth, abline, vline, hline, ggTimeSeries
- **Topic** Time series analysis
- **Project Assignment** Write description of your dataset(s)

Week 5

- **R** ggplot2 extensions, scatterplot matrix (GGally)
- **Topic** JunkCharts Trifecta Checkup
- **Project Assignment** Create 2 descriptive plots of your datasets(s)

Week 6

- **R** Boxplots, violin plots
- **Topic**
- **Project Assignment** Write a description of the data cleaning required for your project

Week 7

- **R** geom_spoke, maps
- **Topic**
- **Project Assignment** Write a description of your planned approach

Week 8

- **R** geom_area, geom_ribbon
- **Topic** Project Proposal Description
- **Project Assignment** Work on project proposal presentation

Week 9

- **R** jitter, rug, aesthetics
- **Project Assignment** Present project proposal (<2 Minutes)

Week 10

- **R** themes, ggthemes extension, labels, color scales
- **Topic**
- **Project Assignment** Work on final project

Week 11

- **R** polar coordinates, ggradar extension
- **Topic**
- **Project Assignment** Work on final project (cont.)

Week 12

- **R** word/text analysis
- **Topic**
- **Project Assignment** Work on final project (cont.)

Week 13

- **R** gganimate
- **Topic**
- **Project Assignment** Work on final project (cont.)

Week 14

- **R** git and GitHub for R
- **Topic**
- **Project Assignment**

Week 15

- **R** networks, geomnet extension
- **Topic** Final Project presentations
- **Project Assignment**

Other Resources

There are many useful resources you should be aware of while going through this course:

Statistics

Variance Explained

R for Data Science - Grolemund and Wickham

Visualization

FlowingData

Junk Charts

Courses

Gary King - Quantitative Research Methodology

John Stasko - Information Visualization

Jenny Bryan - Data wrangling, exploration, and analysis with R

Books

Econometrics in R

Using R for Data Analysis and Graphics

Papers

Embedded Plots

Part I

R Tutorials

Chapter 1

Reading data

Chapter 2

How to Judge Visualizations

Bibliography