SOC 4015/5050: Problem Set 01 - Descriptive Statistics

Christopher Prener, Ph.D.

Fall 2018

Directions

Complete all of the following questions using the data from the
testDriveR package. Your well-formatted R Notebook source (the
.Rmd file), associated output, and project structure should be up-
loaded to your assignments repository by 4:15PM on Monday, Septem-
ber 17th, 2017.

Analysis Development: Create a Project Folder System

1. Using RStudio, add an R Project to the *existing* directory in your
   assignments repository named PS-01.

2. In the Files tab, add a new folder to the PS-01 project named docs.

3. Create a new text file for your README.md. Save it in the top level
   of your PS-01 project as README.md. In the body of your README.md
   file, use Markdown formatting to write a sentence or two describ-
   ing the purpose of this project.

4. Create a new notebook and save it within that docs/ subdirectory
   you just created.

5. Expand the YAML heading as you have done previously.[1]

6. Remember to use RMarkdown syntax to create your assignment
   notebook. Make sure it has an introductory section, a section for
   loading packages, a section for loading data, and a section for
   parts 1 and 2 below. These sections should be second-level head-
   ings (e.g. ## Introduction). Within each part, use third level
   headings to designate question numbers (e.g. ### Question 9).

7. When you are done, "knit" your document by clicking the Knit
   button in the toolbar at the top of the notebook.

This initial section follows the project
workflow that was handed out during
this week's lecture and is available in
the lecture-03 repo!

[1] Remember that there is some starter
notebook syntax in the lecture-02
repository that contains syntax to get
you going!

Part 1: Data Cleaning

Use the `gss14` data frame saved in the `testDriveR`, make the following changes using techniques covered in Lecture-02 and Lab-01.

8. Keep only the following variables - `ID_`, `SEX`, `HRS1`, `INCOME06`, `RACE`, `WRKSTAT`

9. Keep only observations for respondents who work full time, i.e. where `WRKSTAT == "Working fulltime"`

10. Rename each of the six remaining variables so that they are more intuitive for you.

Part 2: EDA and Descriptive Statistics

Using your cleaned data, perform some exploratory data analyses and calculate the most appropriate descriptive statistics for the given variables. Statistics you should consider are: mode, median, mean, standard deviation, range, and inter-quartile range. These questions draw on both Lectures 2 and 3.

11. Create an appropriate plot for exploratory data analysis of the `SEX` variable and calculate all appropriate descriptive statistics. Include a frequency table using `janitor` as well.

12. Create an appropriate plot for exploratory data analysis of the `RACE` variable and calculate all appropriate descriptive statistics. Include a frequency table using `janitor` as well.

13. Create an appropriate plot for exploratory data analysis of the `HRS1` variable and calculate all appropriate descriptive statistics.

14. Create a scatter plot of last year's income by hours worked for full time[2] workers with the points colored using the `RACE` variable.

15. Create a full set of descriptive statistics for all of the variables in your *cleaned* data set.

[2] *Hint:* Wrapping this variable reference in `as.numeric(INCOME16)` when you specify the aesthetic mapping will clean-up the axis labels.

Rubric

Individual Questions

Part 1		Part 2	
Question	Points	Question	Points
8	3	11	3
9	3	12	3
10	2	13	3
		14	2
		15	1
Points Possible	8		12

Note: Partial credit possible

Project Organization

Category	Details	Points
Excellent	`PS-02` organized following workflow without error	5
Good	Minor concerns with organization	4.5
Improvement Needed	Significant concerns with organization	3
Unsatisfactory	No organizational strategy used	0
Points Possible		5

Notebook Formatting & RMarkdown

Category	Details	Points
Excellent	Syntax used appropriately & without error	5
Good	Minor concerns with syntax use	4.5
Improvement Needed	Significant concerns with syntax	3
Unsatisfactory	No RMarkdown used	0
Points Possible		5

Literate Programming

Category	Details	Points
Excellent	Narrative throughout with great detail	5
Good	Some narrative with inconsistent detail	4.5
Improvement Needed	Limited narrative with little detail	3
Unsatisfactory	No narrative included	0
Points Possible		5