PLSC 473: American Judicial Behavior

Tutorial One

Introduction

The purpose of these tutorials is to begin to develop your skills at using statistical software to conduct empirical research. These tutorials are self-guided, and will not be submitted or graded. Rather, they are an opportunity to develop the skills you'll need to complete the research modules / short papers for the class.

The plan is simple: follow the instructions below. Text that appears in `typewriter` font denotes specific software commands or objects (variables, data frames, etc.). If no command(s) are specified, you can complete the task however you would like. Also, be aware that many of the things I'm asking you to do here are for pedagogical purposes only, and are not necessarily the most efficient way to accomplish the tasks in question.

Hints

1. I recommend that you use R and RStudio for the tutorials in this class. R has a somewhat steep learning curve. Some good resources (of various sorts) are:

   - Quick-R (this one is especially useful, in my opinion),
   - The R Cookbook,
   - William King's R tutorials for beginners,
   - Various R4Stats posts (e.g., this amazingly handily reference page),
   - R-Bloggers,
   - R Cheatsheets,
   - The R page on StackOverflow,
   - The Penn State R User Group,
   - And, of course, the mothership.

2. When – not if – you get error messages, a simple but useful thing to do is to put the error message in quotes (e.g., "`Error in match.fun(FUN) : object 'U' not found`"), along with the letter "R," into our old friend Google. It is almost impossible to be the first person ever to have encountered a given error message.

Tutorial

This tutorial focuses solely on data entry and data management. You're first asked to enter some data and conduct some transformations of those data, and second to read in some data from a web source.

Part I

Here are some data on five hypothetical survey respondents:

Name	Sex	Height (in inches)	Hair Color
Maria	Female	69	Black
Sage	Female	60	Red
Dani	Male	68	Grey
Jessica	Female	64	Black
Michael	Male	73	Brown

1. Create a *vector* containing the names of the respondents.
2. Create a *factor variable* containing the sex of the respondents.
3. Create a *numeric variable* containing the height of the respondents.
4. Create a *factor variable* containing the hair color of the respondents.
5. Create a single *data frame* (called `AllData`) by combining all four existing variables.
6. Add a variable to the data frame (called `HeightInFeet`) that records each respondent's height in feet (that is, their height in inches divided by 12).
7. *Sort* the data so that it orders respondents from shortest to tallest.
8. Create a second data frame (called `Females`) of all five variables, selecting only the female respondents.
9. Create a third data frame (called `Under56`) of all five variables, selecting only respondents less than 66 inches tall.
10. Create a fourth data frame (called `NoSex`) that retains all five observations, but deletes (drops) the *Sex* variable.
11. Create a fifth data frame (called `SexOnly`) that retains only the respondent's names and their sexes.
12. Using the names as the "key," *match-merge* the last two data frames you created (the ones in steps 10 and 11 above). When you are finished, you should have a data frame that has the same five variables as in `AllData`.

Part II

The course github repo contains a copy of the U.S. Supreme Court Justices Database; it is filename `Justices-Database-2013.csv`, and it is a comma-separated values file. The corresponding codebook is `Justices-Database-Codebook-2013.pdf`.

Your task in this part is to read these data into RStudio, twice.

1. First, download the data and read the file in "locally."
2. Second, read the data file directly from its github URL, at

   https://raw.githubusercontent.com/PrisonRodeo/PLSC473-git/master/Data/Justices-Database-2013.csv.

3. Hint: Be sure you have installed the `RCurl` package, and take a look at the R code from the second day of class (at around line 156) for an example of how to do this.