

CMPSC 390
Data Analytics
Fall 2017

Lab 6: Political Science Exploratory Data Analysis

Save this lab assignment to: labs/lab6

Objectives

To learn how to ask exploratory questions while using data from a political science area. To provide the necessary analysis to be able to answer those questions, combining skills, methods and theory from data analytics. To extract meaningful correlations in the data on political cultures in the nature of civil legal representation.

Reading Assignment

Please review Chapter 5 and then read chapters 7, 8, and 13 in the course book. Also, please review your notes from the lab discussion of the data which you will utilize for the completion of this lab. Please use the included paper with your lab data entitled, *Funding Poverty Law in an Era of Decentralization*, by Brian M. Harward of Allegheny College.

Groupwork

You are to work in a group of not more than three people for this lab. Be sure to discuss each of the tasks and proceed after the group has come to a complete agreement. **Each person is to turn in his or her own report and code, however all lab partners should be listed in the submission.**

Political Science Data Analysis

In this lab you will conduct your own exploratory data analysis on a data set provided by Dr. Harward. In particular, you should follow the following steps:

1. Your data set has been included in this lab as well as the article which describes the data of the columns. See Sections **Research Methods** and **Results** of the article for discussion about each of the columns. In these sections, you will also learn of potential variations which exist between types of data in the columns.
2. Generate five questions about the included data set. It is expected that you will select questions which are centered around exploration of variation/covariation between columns of your selected data. In addition, your response should produce visualizations (involving the ggplot function) to conclusively answer your research questions.
3. Open the csv data file using the following command.

```
poliSciData <- read.table(file.choose(), header = TRUE, sep = ",")
```

4. Due to the large size of the data table, you are to reduce the size of your working data by creating new tables which are relevant to each of your questions. For example, you are to use the TIDY commands and techniques which were discussed in class to create these tables.
5. Utilize your skill set learned so far to write an R program that provides **comments** and the **code** for the **solutions** to your questions.
6. In the nature of exploration, update your questions from step 2, if necessary, and generate new questions which may be more useful for your analysis.

Part 1: Presentation on Data Selection and Questions

During the lab session next Friday, 20th of October your group will give a short (5-10 minute) presentation on the first two steps described above. Specifically, your presentation should discuss the data set you have selected and the types of questions you generated with the justification for each. Although you do not need to have the solutions by this time, you should select the approach you will take to find solutions to each of your questions. Please use slides in your talk and ensure that each group member contributes equally to the presentation.

Part 2: Analysis and Solution Generation

During the second part of the lab you will finalize your solutions and your analysis and write a report with your findings and your reflection. Your final deliverable for this lab include:

1. **Your Political Science Data Analysis:** In addition to the R code that provides your solutions, you are to write a report that includes the following:
 - (a) A brief explanation of the data and its headers. Remember to review the article's **Research Methods** and **Results** sections.
 - (b) **Five (5) questions about your data set.** Also, justify your selection of these particular questions (how asking these questions maybe helpful).
 - (c) The answers to your five questions including any appropriate transformations and visualizations that were produced during research.
 - (d) Your final analysis including your input on modifying your original questions and generating new questions if relevant.
- (a) Based on Dr. Harward's provided data and article, discuss the processes to make political decisions to fund states' legal services. In addition, discuss how such decisions are affected. Reflect on potential faults in this system (as you understand them) and how data analysis may be able to identify such faults.

Important Details

Lab directory structure: Make sure you have placed your submission materials for this lab into labs/lab6 directory in your Bitbucket repository (cs390f2017-billb).

Note: Please remember to include your name on everything you submit for the class. Although the instructor collects your work from Bitbucket, each work must be graded outside of the Bitbucket directory and so without adding your name, the instructor will be unable to award you credit for your work.

Required Deliverables

By the 27th October, submit electronic versions of the following deliverables through your Bitbucket repository (`cs390f2017-billb`) by correctly using appropriate Git commands, such as `git add -A`, `git commit -m 'your message'` and `git push`. When you have finished, please ensure that the Bitbucket Web site has your pushed work. Please contact the instructor if you have any questions about assignment submission.

1. The R source code that you used to answer your questions.
2. Report that includes the data descriptions, exploratory questions, answers to your questions and your reflection.
3. Presentation slides.

References

Harward, Brian M. "Funding Poverty Law in an Era of Decentralization." *Journal of Poverty* 12.4 (2008): 387-410.