

[POLS 4150] Problem Set 1

100 points

Professor L. Jason Anastasopoulos

Date posted: January 22, 2017

Date due: February 2nd, 2017

Submission Instructions

Please submit your problem set by Thursday, February 2nd before class. You can submit your homework at any time using the Dropbox link below.

When you submit your homework, please:

- Make sure that the filename containing the answers to the questions is in the following format: `LastName-ps1`
- Make sure that the filename containing your R-code is in the following format: `LastName-ps1.r`
- Submit anytime before Thursday, February 2nd at 12:30pm using this link: <https://www.dropbox.com/request/9IwjYUcei6i00TtM3Zeh>

Variables names and descriptions for gavotes.csv dataset

county	GA county names.
totalvotes	total number of votes cast in the county.
trumpvotes	total number of votes cast for Donald Trump in the county.
clintonvotes	total number of votes cast for Hillary Clinton in the county.
johnsonvotes	total number of votes cast for Hillary Clinton in the county.

Question 1 (50 Points Total)

2016 Election Data in Georgia Counties

The purpose of this exercise is to familiarize you with R and to make sure that you understand how to do basic things in R. We will be dealing with real 2016 election county-level election results from Georgia, which have recently been released thanks to a former UGA undergrad, Stephen Pettigrew at Harvard.

The dataset that you will download for this exercise can be found here: [gadata.csv](#).

For this part of the assignment, in addition to a word processing document, I would like you to submit an R code file using the following naming convention: “**LastName-ps1.r**”. For example, if I were to turn in the code file for this problem set, the file name that I would turn in would be named “**anastasopoulos-ps1.r**”.

The code should be clearly demarcated by question using hashtags “#”. For example:

```
# Question 1.1
library(foreign)
gadata = read.csv("path-to-data/gadata.csv")
attach(gadata)
# Question 1.2
...

```

1.1 Load the data into R using two methods: (1) using the full path of the data file and (2) setting your working directory to the where your file is located. Call the data that you load into R “gadatar.” Use the `dim()` function to print the number of rows and columns the dataset. (5 Points)

1.2 Create three new variables:

- **pct.trump** – proportion of county vote share that Trump received.
- **pct.clinton** – proportion of county vote share that Clinton received.
- **pct.johnson** – proportion of county vote share that Johnson received.

Find and report the mean, median, maximum and minimum values for each of these variables. (10 Points)

1.3 Find the counties that had the highest vote share for Trump, Clinton and Johnson, respectively. Do a little research on the county that had the highest vote share for Johnson and explain, in three sentences or less, why that county had the highest vote share for Johnson. (10 Points)

1.4 Create a histogram of the proportion of county vote share that Trump received. Label the x-axis with your name and label the title of the histogram “Vote Share for Trump in GA.” What does this histogram tell you about support for Trump in GA counties? (10 Points)

1.5 Create two plots: (1) Trump county vote share v. Clinton county vote share; (2) Trump county vote share v. Johnson county vote share. Put your name as the title of your plot and label the x and y axes as “% Trump”, “% Clinton” etc. In one sentence, what does the plot of Trump v. Clinton vote share tell you about the relationship between these two variables? What about the relationship between Trump v. Johnson vote share? (15 Points)

Question 2 (50 Points Total)

Opinions about the Affordable Care Act in Athens–Clarke County

Your class has been approached by Congressman Hice’s office. His team wants to understand how voters in Athens–Clarke County feel about the Affordable Care Act. The office would like you to present them with a research plan which describes your approach for gathering voter preferences on this topic. Staff has offered a relatively short time frame and a modest budget for the project, but they indicate that there may be some flexibility if you argue a compelling case for the need for more time or more money to complete the work. Please construct a proposal for the Congressman’s office that focuses on the following items:

- 2.1 What is the population of interest in this study and what sampling frame will you use for your study? *(10 points)*
- 2.2 You have decided to collect the data via a survey and need to work out the details. Some of the answers to the questions below will help you sort out how you will conduct your study.
 - (a) How will you choose the people that will take this survey from the population that you defined and why? *(10 points)*
 - (b) Choose a method that you would like to use to administer the survey (ex. land line, cell phone, mail, in-person, internet)? Please describe the mechanisms you will use to implement your survey method (for instance, how will you get paper surveys to people and how will you collect them/follow-up on non-responses?). Identify the advantages and disadvantages of the method that you chose versus other possible methods. *(15 points)*
 - (c) What biases might your data collection method be subject to? What problems could bias create for the interpretation of results? *(15 points)*