POL 204b: Homework 1
Due October 14, High Noon

October 8, 2014

This assignment is tentatively due Tuesday, October 14 at 12:00p. Please:

1. Ask for help if you need it.
2. Work together on the R code...
3. ..but make sure you "get it".
4. Submit everything well organized and legible.
5. Include your code as David asks you to. Email it to him and me.

Exercises

1. For the following, use the dataset of states attached as an appendix. Using the **Percent Retired** variable (the percent of the state's population aged 65 or older), calculate or graph as appropriate all of the following, by hand.
   - Mean, Median, Range, Standard Deviation and Five-Number Summary
   - Stem and Leaf, Histogram, Box Plot
2. Using R, produce the same graphs and statistics, but, for the Birth Rate variable.
   - 50% points bonus if you can do this using only sorting calls, control statements (`if, else`), and arithmetic (rather than, say `mean(...)`).
   - Another potential 50% bonus if you roll all of this into a homemade function called `SumDat(...)` that will store all of these into a single object, that looks nice when printed.
3. Analyze the data in ICPSR study # 22408 using R. Start with the STATA (.dta) format and use the `read.dta` from `foreign`. Remember from bootcamp that loading packages works generally as: `library(<package.name>)`, where you replace `<package.name>` with the name of the package. In this case, `library(foreign)`.

   Produce the same statistics and plots for this dataset as the last two questions.

4. Convince me that the analyst who coded this into Stata the first time "got it right" when converting from ASCII – *convince me that you can read in the ASCII file for at least one of these variables* and then test for the equivalence of the two datasets. I recommend looking into `read.fwf` to start with.

5. Examine the education in single years variable in this dataset (V133). Make a new ordinal variable that describes a R's level of education using the following rule:

   - $Edu < 4 \rightarrow "Low"$
   - $4 \leq Edu < 9 \rightarrow "Medium"$
   - $9 \leq Edu \rightarrow "High"$

   Do you learn anything? Make a cross tab of this new variable with the ordinal variable that the coder created internal to the dataset. How well do they match?

6. Produce the same graphs and statistics as in the previous problem, analyzing number of political figures assassinated, using ICPSR study # 5208. Then examine the variable for group type, propose a smaller number of summary categories for that variable (and why this smaller number makes sense), make those categories, and examine the distribution of this new variable. Which of your types of groups attempt the most assassinations? Which types are most successful?

7. Optional: Obtain a dataset of congressional elections and construct histograms similar to those presented in class for two or three years. Try breaking down the analysis, perhaps comparing histograms for different regions or for incumbents and non-incumbents. To do so, you might need to make a new variable. For example, you might create a dummy variable "South", coded '1' for Southern districts and '0' for non-Southern districts. Once you have the new variable, a plot from the `lattice` package can make some really powerful plots.

8. Using the 1948 National Election Study, make a two-way table of partisanship and ethnicity/race, calculate appropriate percentages, and discuss your results. OPTIONAL AND IN NO WAY REQUIRED: If interested, compare with 1998. Or, examine some aspect of attitudes by age, and look for generational effects.

   Try to have fun and learn. If you wish, try some creative alternatives to the graphs presented in class. Just tell me what David and I are supposed to learn from your plot.

Data for Problem #1

State	PctRetired	BirthRate
Alabama	13.2	1.4
Alaska	6.2	1.66
Arizona	13.6	1.64
Arkansas	14.1	1.4
California	10.9	1.56
Colorado	10	1.53
Connecticut	13.8	1.26
Delaware	13.4	1.39
District of Columbia	12	1.45
Florida	17.8	1.31
Georgia	9.9	1.68
Hawaii	13.8	1.56
Idaho	11.6	1.56
Illinois	12.1	1.5
Indiana	12.4	1.4
Iowa	14.8	1.27
Kansas	13.2	1.44
Kentucky	12.6	1.34
Louisiana	11.6	1.53
Maine	14.6	1.06
Maryland	11.6	1.44
Massachusetts	13.6	1.28
Michigan	12.4	1.35
Minnesota	12.2	1.27
Mississippi	12.2	1.57
Missouri	13.5	1.35
Montana	13.6	1.21
Nebraska	13.5	1.44
Nevada	11.9	1.56
New Hampshire	12.3	1.16
New Jersey	13.3	1.37
New Mexico	12.2	1.49
New York	13	1.36
North Carolina	12.4	1.5
North Dakota	14.7	1.2
Ohio	13.3	1.4
Oklahoma	13.3	1.42
Oregon	12.9	1.32
Pennsylvania	15.5	1.17
Rhode Island	14.5	1.16
South Carolina	12.5	1.44
South Dakota	14.3	1.4
Tennessee	12.6	1.4
Texas	10.3	1.67
Utah	8.9	2.1
Vermont	13	1.02
Virginia	11.5	1.43
Washington	11.5	1.35
West Virginia	15.3	1.01
Wisconsin	13.1	1.28
Wyoming	12	1.22