Homework 2

Overview:

In this assignment, you'll be using R/RStudio and the GSS data set (2014) to report bivariate statistics for two dependent variables of interest. These variables must be different than the ones I present in the example below (unless you've discussed it with me). Remember, I want you to select variables that interest you, that way you could use these analyses for your final paper. Finally, the only bivariate relationship you will report for this analysis is a **correlation** between two variables.

What you will need:

The following can be downloaded from our course website

- **hw2_script.R**
- **GSS2014_final.csv** aka GSS Data Set (2014)
- **LABELS_script.R**
- **GSS Variables and Descriptions.pdf**
- **GSS_Codebook.pdf**

Tasks:

- Open up the Homework 2 script/code file (**hw2_script.R**). This is the file you'll be working from.
- Set your working directory
- Load the required packages/libraries
- Load the data set (**GSS2014_final.csv**).
- Load the cleaner/labels file (**LABELS_script.R**).
- Select two (2) interval/continuous dependent variables that interest you. To do so, you will need to look through the list of variables (**GSS Variables and Descriptions.pdf**). Once you've selected two variables, locate them in the codebook (**GSS_Codebook.pdf**) to figure out how those variables are coded.

What you will turn in:

- A brief paragraph describing:
  - A sentence or two about the data set you're using for these variables.
  - For each variable, a sentence or two describing the variable name and its description.
- A table with that lists the correlation between the variables, and a description/interpretation of the relationship (use the Cohen 1988 cutoffs to figure out the size of the correlation: less than or equal to .29 is small, between .30 and .49 is moderate, and greater than or equal to .50 is large).
- The text of your R code/script on a separate page.

Example Homework 2

1: Data Set Description

The data for this assignment come from the General Social Survey administered in 2014. The data set has 2,538 observations, with individuals as the unit of analysis.

2: Variable Descriptions

The two outcome variables I selected were **POLVIEWS** and **OPPSEGOV**. The variable **POLVIEWS** is a continuous measure of conservatism that ranges from 1 to 7, with higher scores on this variable indicating higher levels of conservatism. The text of the variable is as follows:

> "We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal–point 1–to extremely conservative– point 7. Where would you place yourself on this scale?"

The variable **OPPSEGOV** is a continuous measure on the importance of civil disobedience that ranges from 1 to 7, with higher scores indicating that the respondent believes civil disobedience is more important. The text of the variable is as follows:

> "That citizens may engage in acts of civil disobedience when they oppose government actions."

3: Correlation Table

$Variable$	$POLVIEWS$	$OPPSEGOV$
$POLVIEWS$	1.000000	
$OPPSEGOV$	-0.1659262	1.000000

The correlation between **POLVIEWS** and **OPPSEGOV** is weak and negative (and significant, since $p < .001$). This indicates that as one variable increases, the other decreases. People with higher levels of conservatism (i.e. higher values of **POLVIEWS**) believe that civil disobedience against the government is not important (i.e. lower levels of **OPPSEGOV**).

4: R Code/Script for Homework 2

```
setwd("/Users/burrelvannjr/Dropbox/methods_csuf")

library(psych)
library(Hmisc)
library(RCurl)

DATA1<-read.csv("resources/data/GSS2014_cleaned_nm.csv",header=TRUE,sep=",")
source("resources/data/LABELS_script.R")

DATA1$polviews
DATA1$oppsegov

cor.test(DATA1$polviews, DATA1$oppsegov)
```