Problem Set 1
Quantitative Political Methodology (U25 363)
Due: February 6, 2018

Instructions
- Please show your work if possible. You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask me directly or come to office hours.
- Your homework should be submitted electronically on the course Github page.
- This problem set is due at the beginning of class on Wednesday, February 6, 2019.
- No late assignments will be accepted.
- Total available points for this homework is 100.

Question 1 (8 points)
Go to the GSS Web site, sda.berkeley.edu/GSS/ and click on "GSS – with 'No Weight' as the default (SDA 4.0)". By entering LDCGAP, find a summary of responses to the question, "Turning to international differences, do you agree or disagree: Present economic differences between rich and poor countries are too large."
   a) (2 points) What is the mean response?
      2.22

   b) (2 points) What was the most common response?
      (2) Agree

   c) (4 points) Is you answer in (b) a descriptive statistic or an inferential statistic? Explain.

      (b) is an inferential statistic b/c it is an inference made about the sample

Question 2 (8 points)
In the 2016 presidential election, an exit poll sampled 1,941 of the 2,808,605 people who voted in Missouri. The poll stated that 57.23% of respondents reported voting for the Republican candidate, Donald Trump. Of all 2,808,605 voters, 56.8% voted for Trump.
   a) (2 points) For this exit poll, what was the population?
      Missouri voters, 2,808,605 people

   b) (2 points) For this exit poll, what was the sample?
      1,941 voters

   c) (2 points) Identify a statistic.
      57.23% of respondents reported voting for Trump

   d) (2 points) Identify a parameter.
      56.8% of all voters voted for Trump

Question 3 (24 points)
Which scale of measurement (nominal, ordinal, or interval) is most appropriate for:
- a) (2 points) Educational attainment (less than high school, some high school, high school, some college, college degree, graduate or professional degree): **ordinal**
- b) (2 points) Race (White, Black or African American, Asian, . . .): **nominal**
- c) (2 points) Letter grades: **ordinal**
- d) (2 points) Statewide murder rate (number of murders per 1000 population): **interval**
- e) (2 points) Distance (in miles) commuted to work: **interval**
- f) (2 points) Hair color (Blond, Brunette, Red, Black): **nominal**
- g) (2 points) Number of people you have known who volunteered for the Obama campaign: **Nominal**
- h) (2 points) Partisan affiliation (Republican, Democrat, Green,....): **nominal**
- i) (2 points) Zip code: **interval**
- j) (2 points) Religious affiliation (Catholic, Protestant, Jewish, Muslim, Buddhist, other): **nominal**
- k) (2 points) Government spending on environment (up, same, down): **ordinal**
- l) (2 points) GPA (4.00, 3.00, 2.00, etc): **interval**

Question 4 (14 points)
Suppose you are comparing the age of 20 men and 20 women. For the men, you get the following responses (in years): 56, 60, 50, 26, 45, 35, 41, 43, 34, 42, 37, 39, 33, 28, 52, 48, 27, 20, 44, 32. (put these #s, into a vector, make it an object in R)

For the women, you get the following responses (in years): 47, 49, 20, 46, 43, 44, 45, 60, 57, 28, 55, 27, 25, 50, 52, 48, 23, 42, 33, 59.
- a) (12 points) In R, make a box plot for both men and women. Make sure the mean and in- terquartile range are clearly marked for each box plot. (Make sure you attach a page with the code and figure to the back of this homework.)

RCode:

```
rm(list =ls(m))
ls()
rm(list=c("m","w")
rm(m)
rm(w)

# had to remove things from my environment

c(56, 60, 50, 26, 45, 35, 41, 43, 34, 42, 37, 39, 33, 28, 52, 48, 27, 20, 44, 32)
m <- c(56, 60, 50, 26, 45, 35, 41, 43, 34, 42, 37, 39, 33, 28, 52, 48, 27, 20, 44, 32)

c(47, 49, 20, 46, 43, 44, 45, 60, 57, 28, 55, 27, 25, 50, 52, 48, 23, 42, 33, 59)
w <- c(47, 49, 20, 46, 43, 44, 45, 60, 57, 28, 55, 27, 25, 50, 52, 48, 23, 42, 33, 59)
print(w)

boxplot (m, xlab="Men", ylab="age", main="")
boxplot (w, xlab="Women", ylab="age", main="")
```
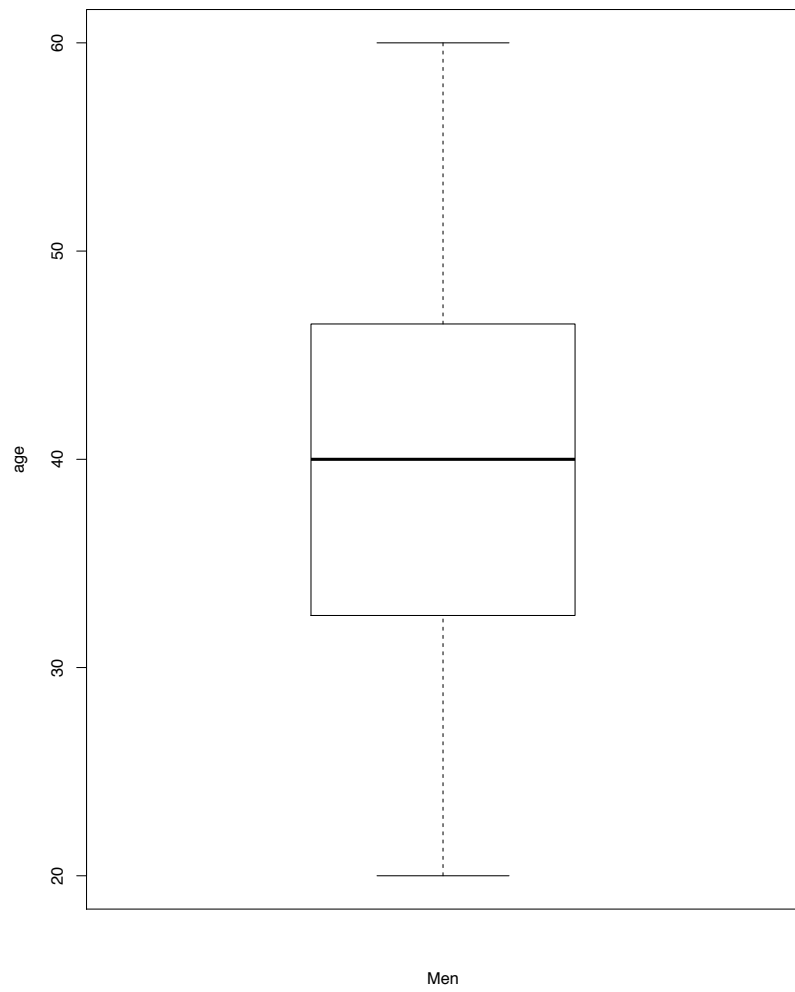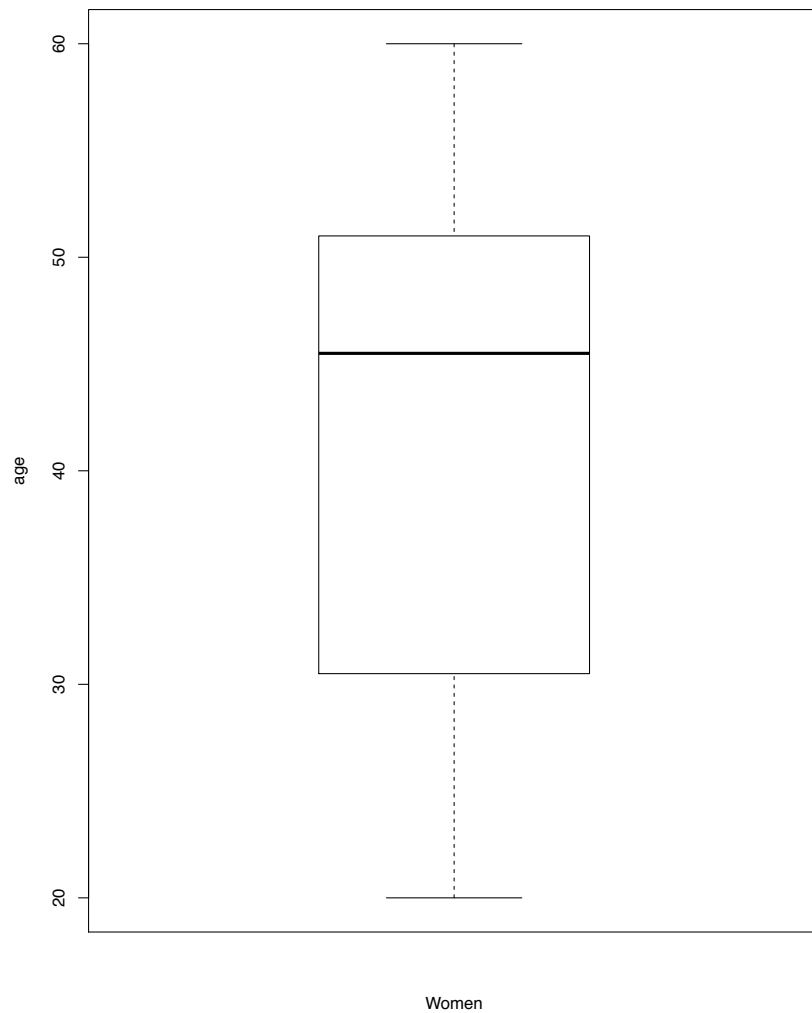
b) (2 points) Is this likely to be a random sample of the American female population? Why or why not?
   a. **No; while the sampling methods aren't explicit in the question, given that the sample size is only 20 women it's not likely random in the sense that it doesn't represent the American female population, and it would be hard to say anything about a mean age of an entire population with such a small sample size**

Question 5 (24 points)
Identify each variable as discrete or continuous. If you think that more than one answer might be correct, justify your answer.
   a) (2 points) Attitudes toward legalization of marijuana (favor, neutral, oppose): **discrete**
   b) (2 points) Number of political parties in a country: **discrete**
   c) (2 points) Religious affiliation (Catholic, Jewish, Protestant, Muslim, ...): **discrete**
   d) (2 points) Choice of candidate a person will vote for: **discrete**
   e) (2 points) Distance (in miles) commute to work: **continuous**
   f) (2 points) Years of school completed (0, 1, 2, ...): **discrete** (if scale doesn't allow for values b/w years, i.e. whole years) or **continuous** (i.e. someone can have completed 0.1 years of school)
   g) (2 points) Number of people you have known who volunteered for the Clinton campaign: **discrete**
   h) (2 points) Partisan affiliation (Republican, Democrat, Green,....): **discrete**
   i) (2 points) Attitude towards the health care reform (favor, oppose, neutral): **discrete**
   j) (2 points) Political ideology (very liberal, somewhat liberal, moderate, somewhere moderate, somewhat conservative, conservative): **discrete**
   k) (2 points) Government spending on environment (up, same, down): **discrete**
   l) (2 points) GPA (4.00, 3.00, 2.00, etc.): **continuous**

Question 6 (10 points)
Assume that the probability of an American high school graduate participating in sports is 0.45. Further assume that of those who participate in sports, about 34% play an instrument. Again, assume that the probability of a high school graduate attending college is 0.69 (which is the case for 2015 high school graduates according to Bureau of Labor Statistics). Estimate the following probabilities (show your work):

   a) (2 point) that a randomly chosen American high school graduate does not participate in sports.

   1-.45=.55, thus the probability that an American high school graduate does not participate in sports is .55

   b) (4 points) that a randomly chosen American high school graduate both participates in sports and plays an instrument.

   P (Sports and Instrument) = .45 * .34 = .153, or 15.3%

   c) (4 points) that a randomly chosen American high school graduate participate in sports and attends to college.

   P (Sports and College) = .45 * .69 =.3105 or 31.1%

Question 7 (12 points)
Find an article that uses a statistic in the news and attach it to your homework. Please comment on the following: What kind of underlying data is used to calculate the statistic? Is it discrete or continuous? Is it nominal, ordinal, or interval? Is the statistic inferential or descriptive?

https://www.nytimes.com/2018/06/15/business/men-unemployment-jobs.html?rref=collection%2Ftimestopic%2FBureau%20of%20Labor%20Statistics&action=click&contentCollection=timestopics&region=stream&module=stream_unit&version=latest&contentPlacement=6&pgtype=collection

- "In 1950, 4 percent of prime age men were not working or looking for work, today that number is 11%"
- Data: from the Bureau of Labor statistics
- Discrete, people (or households)
- Nominal (employed, working but searching for a job, neither working nor looking for work)
- "In 1950, 4 percent of prime age men were not working or looking for work, today that number is 11%"
- This statistic is descriptive b/c while it is based on the sample/households in the U.S., it's making a statement about men in general