Homework #1, due Tuesday, Sept. 19, 11:59pm

Submit the completed assignment as two files: 1) an Rmarkdown (.Rmd) and 2) compiled .html or .pdf output file with code showing. (Do not use echo = FALSE in the code chunks.) Be sure that all of the plots appear in the output file. If there is a problem with the output file and we need to ask you to resubmit it, points will be deducted. Your code should be easy for someone else to read. To that end, use human readable variable names, and add comments where appropriate. Code should also be efficient (within reason -- don't use 20 lines of code to accomplish something that could be done in one line.)

1. Recreate the following graph using base graphics.[1] Data is provided in Files/Data/StatePop.csv

A CENTURY OF POPULATION CHANGE: SHARE OF POPULATION IN EACH STATE AND CHANGE IN RANK

---
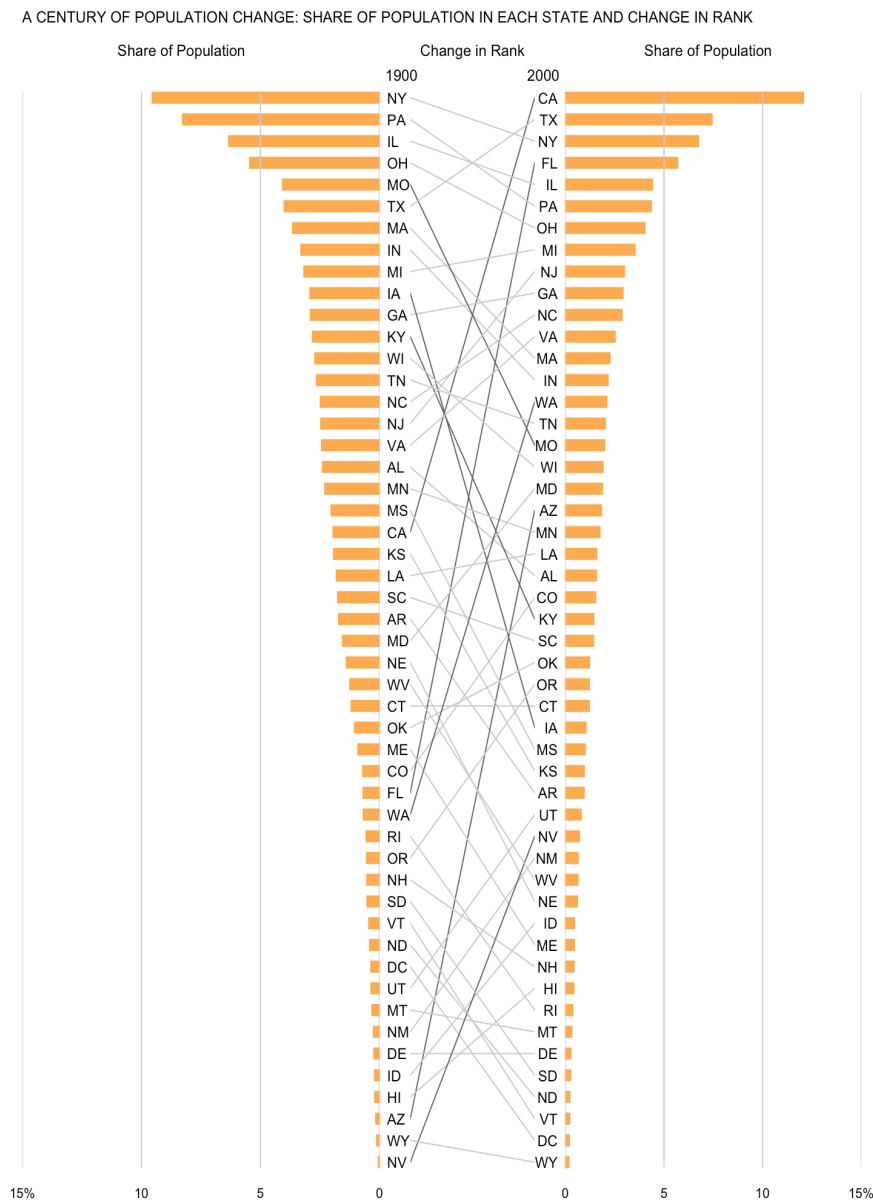
[1] This is a modified version of a graph that appears on p. 348 of Jorge Camões' *Data at Work* (2016).

2. Design a graph using either base graphics or ggplot2 that shows "Change in Rank" only, improving on the middle panel above. That is, use a graphical element that is easier to decode than slope.

3. Use the **movies** dataset in the **ggplot2movies** package for the following:

a) Use histograms, boxplots, and density curves to analyze the distribution of the budget variable. Comment on features using the list on page 29 as a guide. You may subset as necessary to get better views of different ranges of the data.

b) The **rating** variable provides the average movie rating while the **r1-r10** variables provide information about the distribution of ratings (see http://www.stat.cmu.edu/~rnugent/PCMI2016/IMDBMovies.pdf for more details.) Averages are limited in the amount of distribution they provide since we don't know the distribution... for example, an average of "5" might result from a movie that everyone thought was average or a movie that half thought was terrible ("1") and half thought was fantastic ("9"). Use the **r1-r10** columns to get a sense of the distributions of ratings graphically. Do the spread of these distributions vary? You may answer based on a small subset of the data.