**Homework #1: Data Manipulation & Analysis using R**

**Due Date:** Monday; August 14th, 2017, 10 AM Eastern Score: 40

**Method of Submission:** Upload; Sollers online portal with your full name listed

Please make sure you give the codes under each question and necessary screenshots for credits.

This data is about a particular area with variables

----------------

1. YEAR = Year 1947 - 1962
2. PGNP = GNP Deflator 1954=100
3. GNP = Gross National Product in $1,000,000
4. UEM = Unemployed in thousands
5. AF = Armed Forces in thousands
6. POP = Population in thousands
7. EM = Employed in thousands
8. Download this file from the internet link <http://web.pdx.edu/~crkl/ceR/data/longley.txt> and show the file as an object in R. Examine the link before you get this into R. (do not forget to use header = T, nrows=16 in the end of the link). You may want to call this data as “data”.
9. Create a new dataset “data1” by Deleting or Replacing the variable “YEAR” with “STATES” and list them as NJ for the first 5 rows, NY for second 5 rows and PA for third 6 rows.
10. Add another variable “CITY” with a name of a place that you are familiar with e.g., Edison with each row having a different city name.
11. What is the minimum number of employed and average number of employed overall?
12. Which STATES has minimum unemployed and maximum unemployed?
13. Which CITY has min GNP and the max GNP. What else do you observe? Which city you suggest if someone is planning to relocate for good?
14. Create a box plot by plotting AF on State. What is your understanding from the box plot. You will need to give your understanding from the box plot clearly.
15. Create a new dataset “data1” which has the following variables STATE, UEM, EM, POP.
16. Create and add a new variable “diff” which is the difference of POP and EM.
17. Show using scripts as to what you see in the row 7 and column 5.
18. Create a pie chart based on all the three states for UEM and comment about it.
19. Create and show histograms for UN, UEM and POP and comment about your understanding.
20. Create a scatter plot to understand the effect of UEM on GNP. Please provide your understanding from the plot. Please provide a least square fit (using abline) in the scatter plot. What do you see? Linear, non Linear?
21. Create a scatter plot to understand the effect of EM on PGNP. Please provide your understanding from the plot. . Please provide a least square fit (using abline) in the scatter plot. What do you see? Linear, non-linear?
22. Give summary statistics for UEM. Give your understanding as much as possible.
23. Create a dot plot for UEM and rank them with your comments.
24. Explore APPLY, LAPPLY, SAPPLY and TAPPLY on the internet and give your comments about what each of these functions do.
25. What is the total number of UEM and EM for PA.
26. Delete row 16 and show all the remaining 15 rows (as a dataset or with the rcodes and output).