MPA 634  
Data Science for Managers  
Final Exam: Winter 2020

# Definitions and Concepts

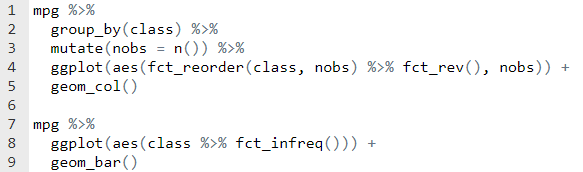
1. First define and explain distribution of a numerical random variable, then define each of the four characteristics of a distribution, and finally detail how you measure and detect each of the four characteristics.
2. Explain the difference between hypothesis generation and hypothesis confirmation.
3. Explain the difference between mutating and filtering joins.
4. Compare and contrast functions and loops. How are they related to the map functions?
5. Explain why a tibble is a list but not all lists are tibbles.

# Line by Line Code Interpretation

Please explain the following code chunks. By referring to line numbers, carefully explain the code and describe what it is accomplishing. Please avoid just repeating the R commands that are in the code.

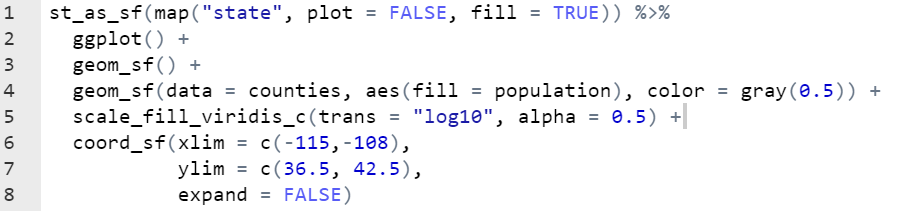
## Code Chunk I: Pareto Charts

Here are two alternative ways to create Pareto charts. Please explain lines 2 – 5 and 8- 9 then compare and contrast the two different approaches to creating Pareto charts.



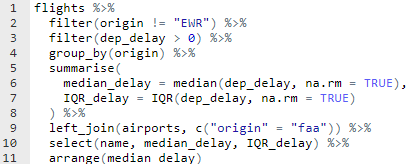
## Code Chunk II: Geospaital

Interpret lines 1 – 8 of the following script that creates a map of Utah with counties shaped by the size of their population



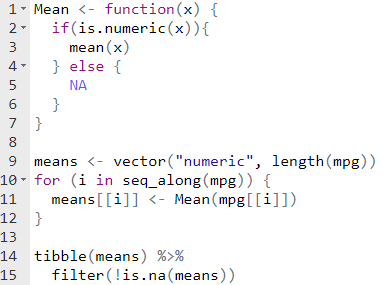
## Code Chunk III: dplyr transformation

Explain lines 2 – 11



## Code Chunk IV: Functions and Iteration.

Explain lines 1 – 5, 9 – 11, and 14-15



## Code Chunk V: Communication Graphics

The presidential tibble has four variables initially:

* name of president
* start date of administration
* end date of administration
* political party

Interpret lines 2 - 12

