MPA 634  
Data Science for Managers  
Final Exam: Fall 2019

# Definitions and Concepts

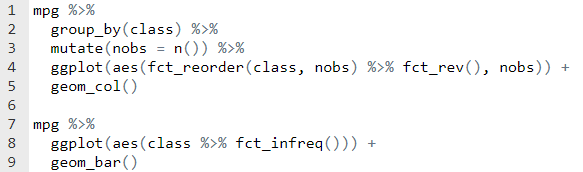
1. First explain distribution of a random variable, then define each of the four characteristics of a distribution, and finally detail how you measure and detect each of the four characteristics.
2. Explain the difference between hypothesis generation and hypothesis confirmation. How is this concept related to training and testing sets?
3. Define factors and then explain how they are related to order and level. Use an example to illustrate your answer.
4. Carefully define residuals and then explain how they are related to the idea of explained and unexplained variation.
5. Explain why a tibble is a list but not all lists are tibbles.

# Line by Line Code Interpretation

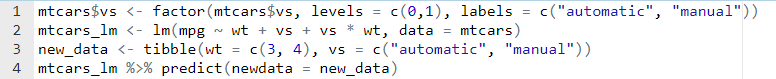
Please explain the following code chunks. By referring to line numbers, carefully explain the code and describe what it is accomplishing. Please avoid just repeating the R commands that are in the code.

Code Chunk 1: Pareto Charts

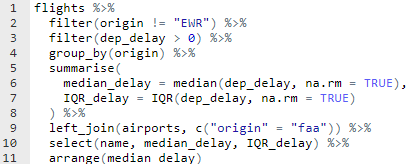
Here are two alternative ways to create Pareto charts. Please explain each line of code and then compare and contrast the two different approaches to creating Pareto charts.



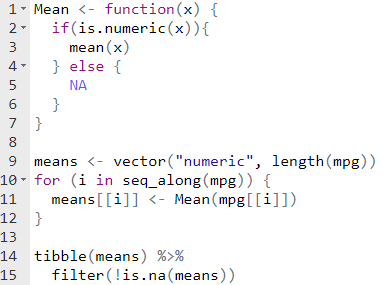
Code Chunk II: Analysis of Covariance



Code Chunk III: Transformations using dplyr



Code Chunk IV: Functions and Iteration



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

