MPA 634  
Data Science for Managers  
Midterm I: Winter 2020

# I. Definitions and Concepts

1. Explain the effect of the scale (standardized) function on variables and explain how scaled variables are used.
2. Compare and contrast filter, mutate, and summarize. Your answer should include an explanation of how each one affects data frames or tibbles.
3. Carefully explain how the whiskers of a boxplot are constructed. How do whiskers help us identify outliers?
4. Explain each of the seven parts of the grammar of graphics by writing a script that illustrates the definition of each part.
5. **Compare** the location, scale, symmetry, and outliers of departure delays using the following violin plot and summary statistics:

A close up of a map

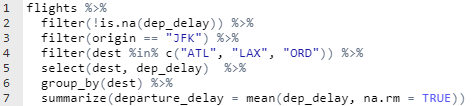
Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Origin | Mean | Median | IQR | SD |
| EWR | 0.6 | -2 | 8 | 8.7 |
| JFK | 0.0 | -2 | 7 | 8.1 |
| LGA | -1.5 | -4 | 7 | 8.3 |

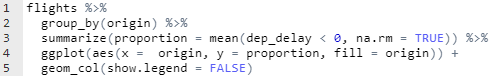
* 1. Location
  2. Scale
  3. Symmetry
  4. Outliers

# II. Line by Line Code Interpretation (Don’t interpret the first line)

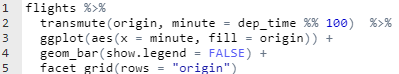
Code Chunk I



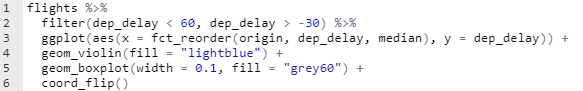
Code Chunk II



Code Chunk III



Code Chunk IV



Code Chunk V

