Scripting Assignment 1: Exploratory Data Analysis Concept Demonstration

Points: 30
Due: Tuesday, February 7 by the start of class

Create an R notebook that illustrates a single concept from the exploratory data analysis topics we have covered so far. Students will work in groups that will be selected in class and each group will work on a different concept. Suggested topics are listed here but groups can also come up with their own concept as long as they get approval from me by Tuesday the 2nd .

1. Differences between mean and median for data distributions
2. Sensitivity of mean, variance, and skew to outliers
3. The sensitivity of IQR compared to variance
4. The impact of different power transforms on positively and negatively skewed datasets
5. Changes in central tendency, spread and symmetry metrics when log transforms are applied
6. Sensitivity of histogram appearance to the choice of origin and the size of the bins
7. Changes in boxplot appearance for different data distributions

I am not expecting an extensive analysis or proof on each topic. Rather the intent is to create a concise illustration that can quickly demonstrate to the reader one basic point about how different exploratory data analysis techniques behave.

Your Notebook should include at least three plots, clear narration of what you are showing in complete sentences and additional comments in your R script as you deem necessary. You should upload both your rmd file and your final html file to the GitHub repository for Assignment 1.

Grading:
10 points: General approach
10 points: Clarity of writing and script
10 points: Quality of visuals