Project Proposal Update 1 (3/27)

Feedback:

I initially had a data in pdf form which I was able to figure out by converting the files into .csv files. Everything else looks fine for now.

Summary:

Initially, I just analyzed one of my datasets which was for the average temperatures of NYC between 1869 to 2018. This time I have chosen the month of 'JUNE' for all my analysis. The average temperatures and found out that there are some outliers in the temperatures. The mode average temperature was around 72 degrees Fahrenheit. Median was around 71.1. The minimum average temperature recorded in month of June is 64.2 while the maximum average temperature was 76.2 Fahrenheit. The range of average temperatures therefore is 12.0 Fahrenheit.

I also plotted a boxplot to further evaluate the outliers for the average temperatures in JUNE. 50% of the average temperatures for JUNE are between 70-73. Histogram confirms that too. The variance in the temperatures was 4.67 and standard deviation around 2.167. From the data, the 75^{th} percentile temperature was 72 Fahrenheit.

From all these statistics, we can see that there aren't many outliers in our data, and data seems to be clean for the most part. Further, I can take different Months, and track down their trends, and then compare different years as well. Another analysis which I would include in the future updates is that I will divide my data into separate parts. For example, from 1869 – 1950 would be one data set, and 1950-2018 would be second data set. Then I will compare the trend of first 100 years to the latter time period to evaluate any correspondence to the carbon dioxide emissions which have skyrocketed in the past 50 years