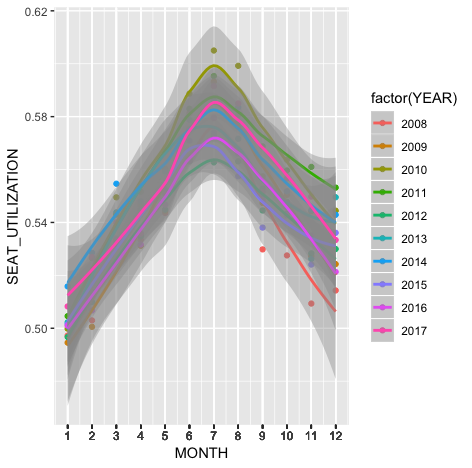
**Introduction:**

The goal of the project is to predict empty seats for a given airline by month of the year. With all the Data cleaning and wrangling done, I tried different approaches for the statistical analysis of the data which is explained in detail in this document.

**Overall Industry Trends:**

To observe the monthly trends across the entire US domestic airline industry (151 airlines), I aggregated the seat utilization by month and year and plotted it by month against the seat utilization.





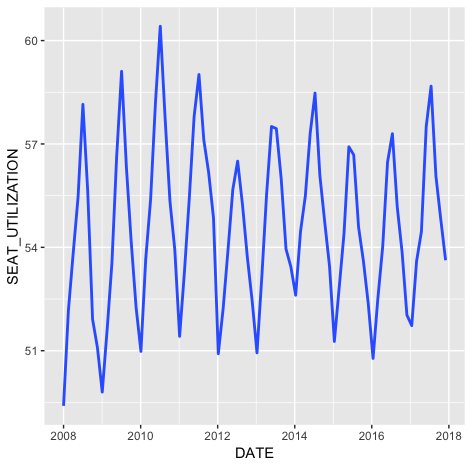
As we can infer from this plot, across all years on average, the months of June, July and August had the best seat utilization and hence the lowest number of empty seats.

January had the most number of empty seats. Also we see that the trends have been pretty consistent over the years with the peaks being in the same months.

Next, a Time Series plot of the data with date on the X axis and Seat Utilization on the Y axis shows us the same trends as well.

The beginning and end of each year comparatively had the most number of empty seats (lowest seat utilization), whereas the middle of the year had the least number of empty seats (maximum seat utilization)

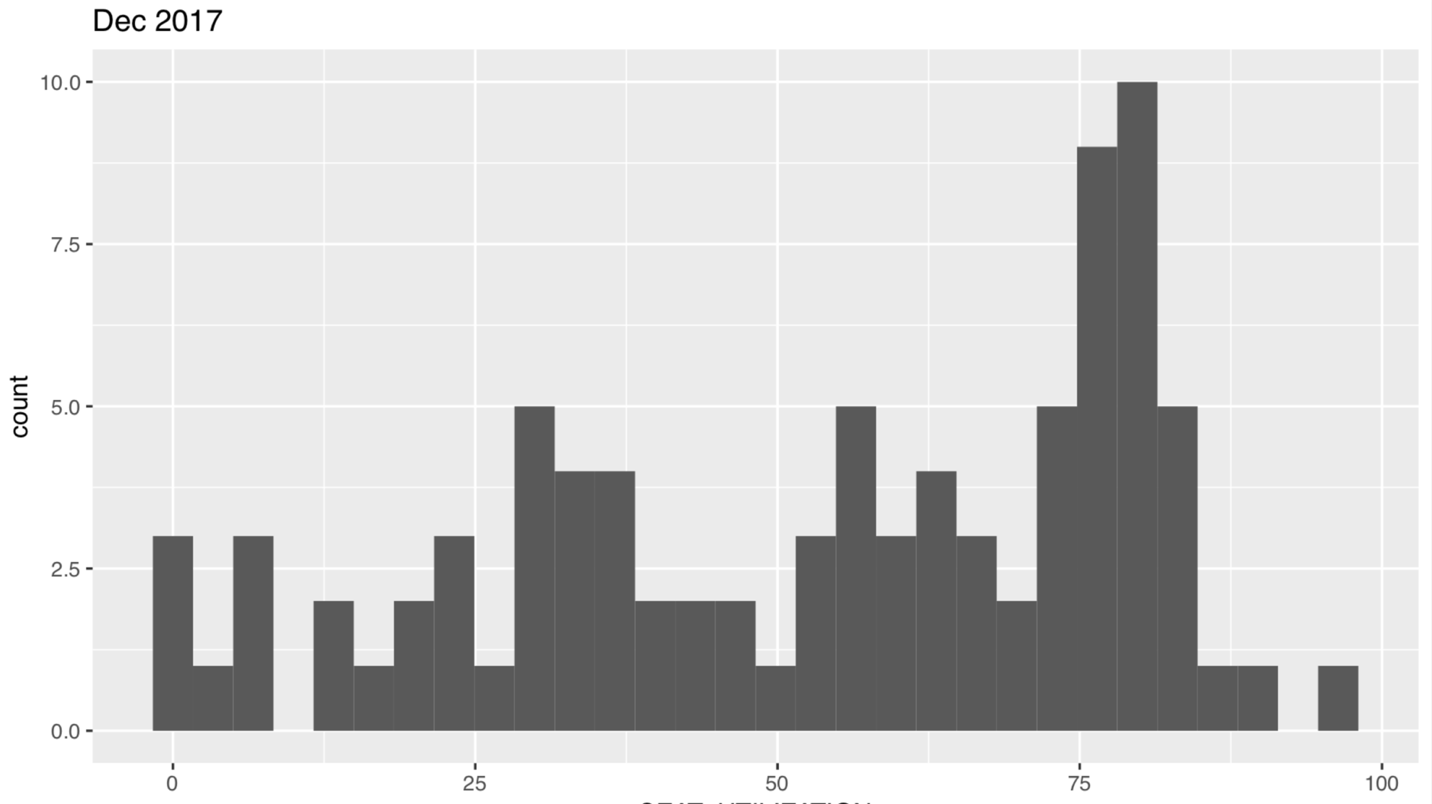


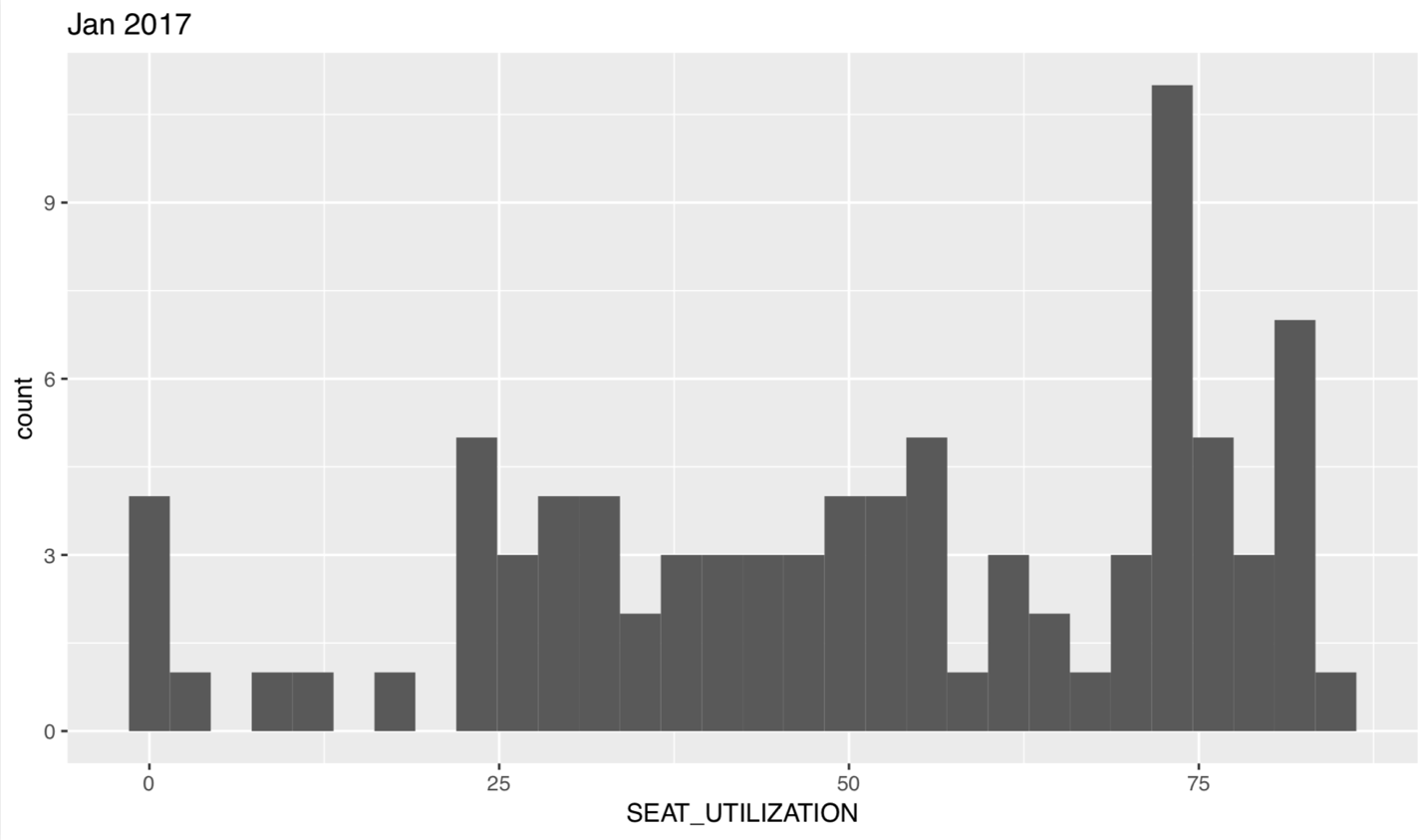
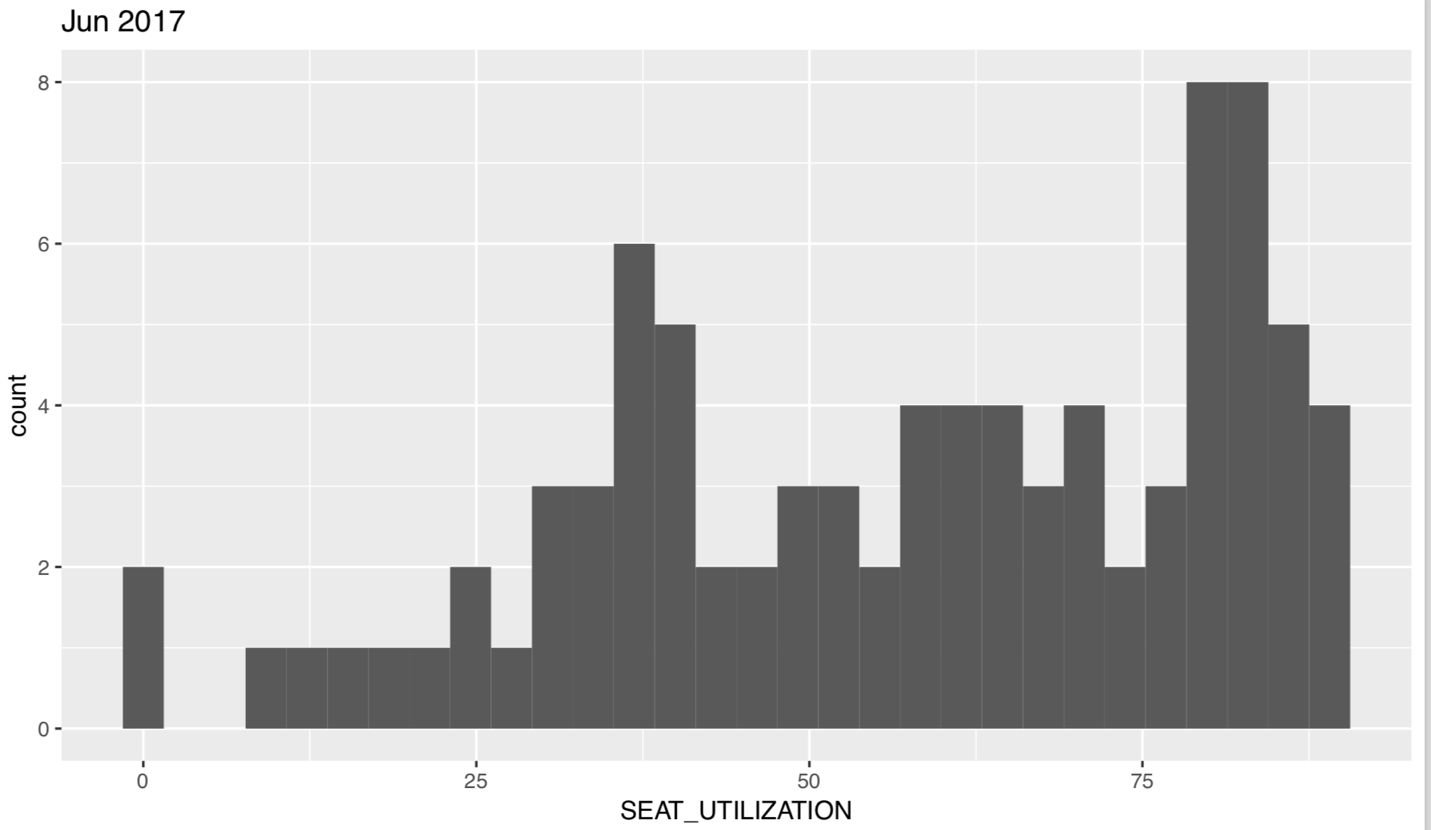


**Monthly Trends By Airline:**

A Histogram plot by month and year would be a good way of representing the number of airlines in each range of seat utilization in every month across every year. These plots are very descriptive of the trends we saw at the overall industry level. Below is the code and a few sample plots of different months.



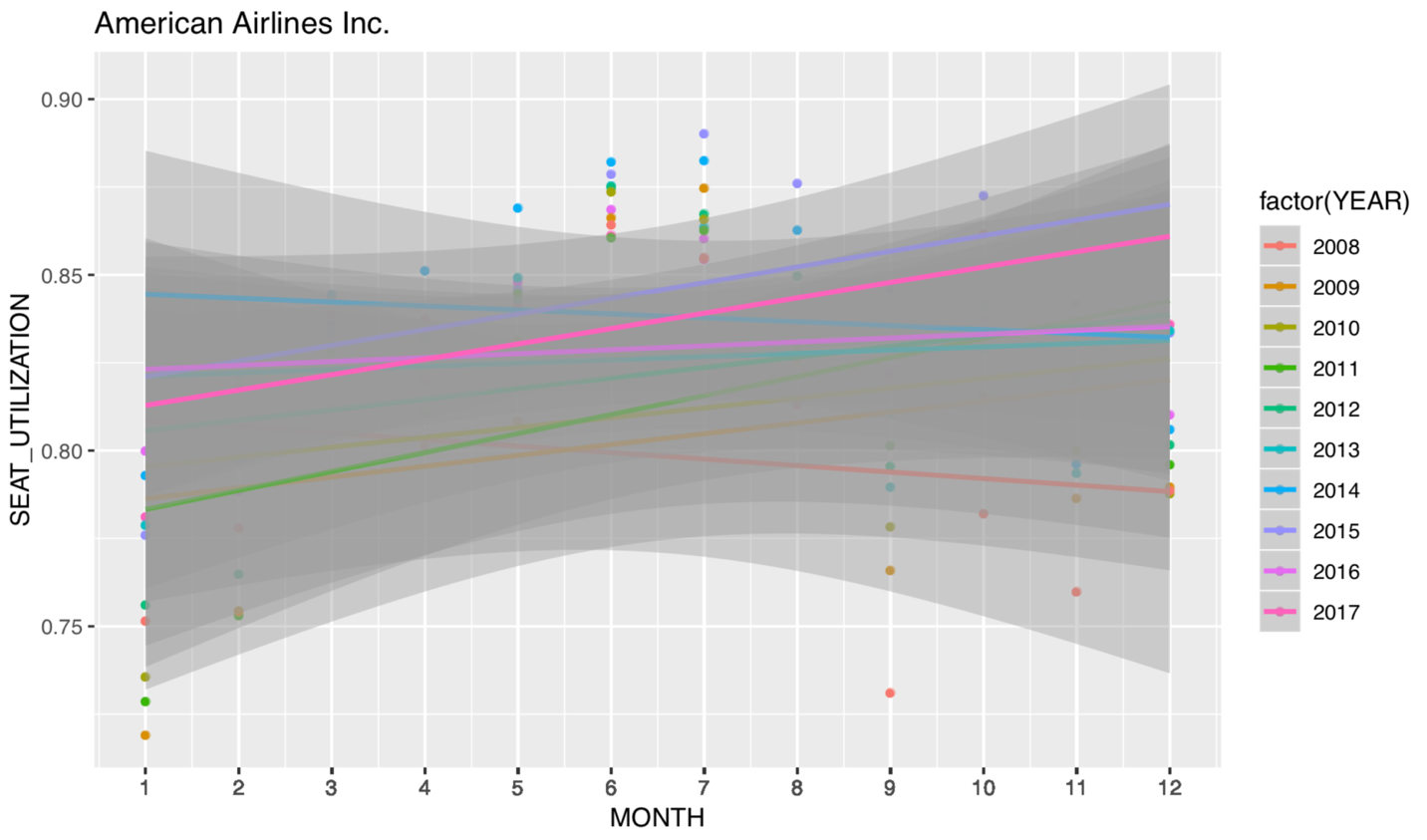




For 2017, Looking at the number of airlines that had greater than 75% seat utilization, June was the best month, while January and December had lower seat utilization.

Every airline was also individually plotted by month and seat utilization, with geom\_smooth, an upward slope was noticed throughout the year with a lot of high data points in the middle of the year.





There were also airlines with straight or downward slopes as well, majority of those airlines were lesser known airlines with comparatively lesser number of flights.

Overall, similar monthly trends were observed across the different approaches.