Predictive Modeling Exercises

Kessiena Ofunrein, Jordan Pflum, Jennifer Robinson, and Katelyn Vincent

8/8/2020

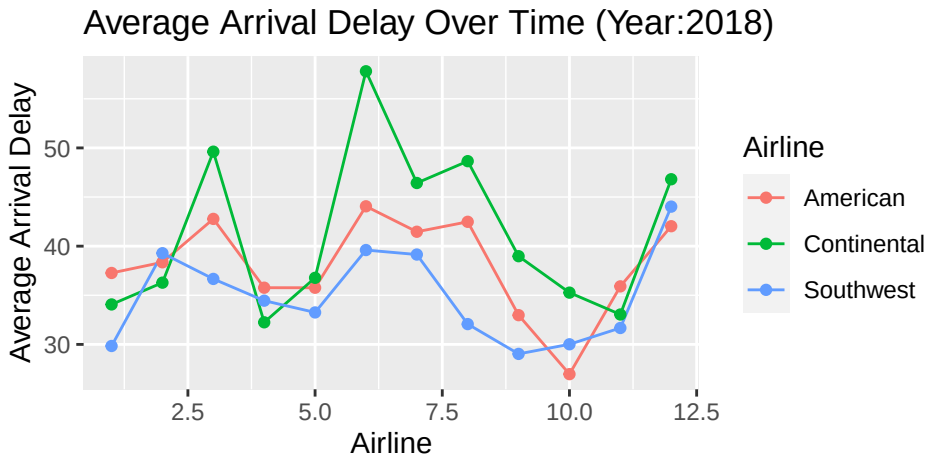
Question 1

Question 2

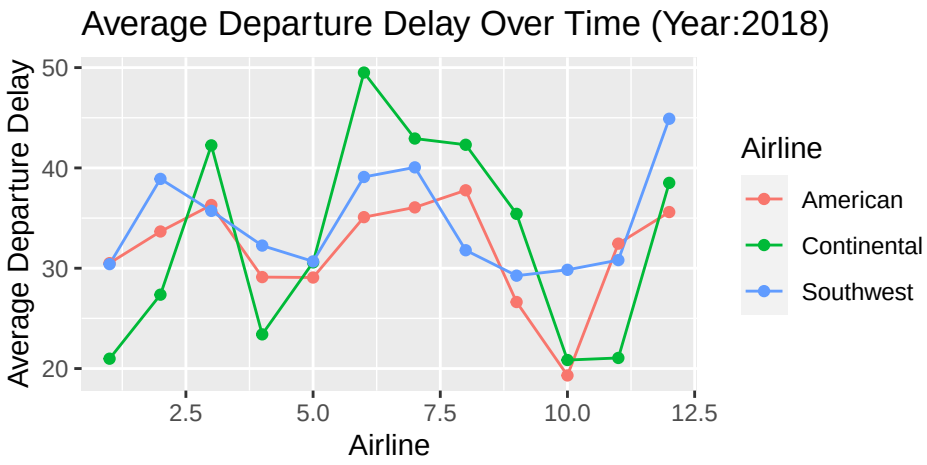
The figures below portray a story about the flights arriving in and departing from the Austin Bergstrom International Airport in 2008. We looked at traits of flights and airlines from a few different perspectives to convey an accurate image of what the airport and travelers experienced that year. This analysis of the data is extremely useful for a few different potential clients. Firstly, travelers are the sustenance of this industry and are impacted by the factors we are analyzing. By providing them with some of these graphs, travelers will be more informed about some things they are likely to experience temporally, geographically, or with a specific airline. Secondly, airport officials would benefit greatly from our analysis because it is advantageous to know what one is likely to experience with each airport. Airports are businesses with reputations and standards. Notoriety for delayed flights is very useful information for preparations and strategic planning.

The line graph figures below are a time series analysis of how the length of flight delays changes throughout the year 2018. The graphs are relatively similar and have peaks during similar times. They are separated by arriving and departing flights, essentially whether the plane was late coming in or late coming out. It is to be expected that a change in one would impact an airport's schedule thus resulting in changes in the other. For both graphs there are peaks in months like March, June, and December. It is easy to deduce the cause of that could be holidays or related to school schedules like Spring Break. For a more concrete picture further analysis should be conducted. The airlines graphed here have more flights than the average airline and the delays are considered long delays because they are in the 75th percentile.

```
## `summarise()` regrouping output by 'Month' (override with `.groups` argument)
```

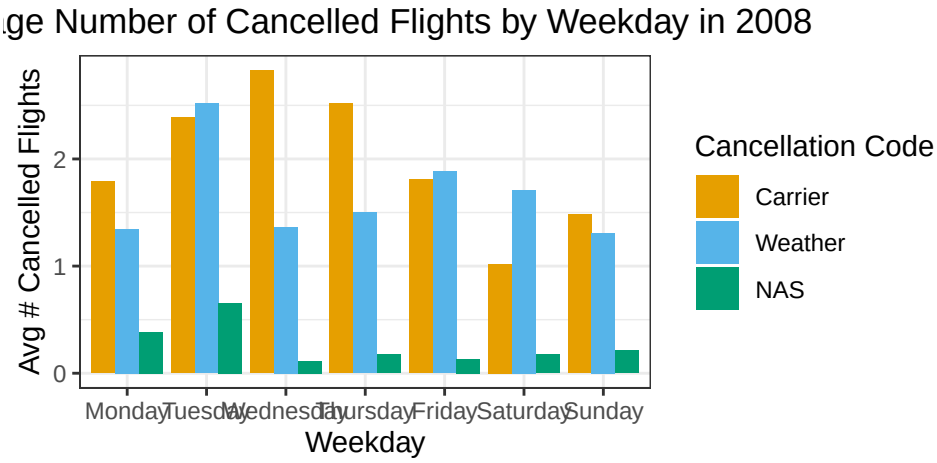



```
## `summarise()` regrouping output by 'Month' (override with `.groups` argument)
```



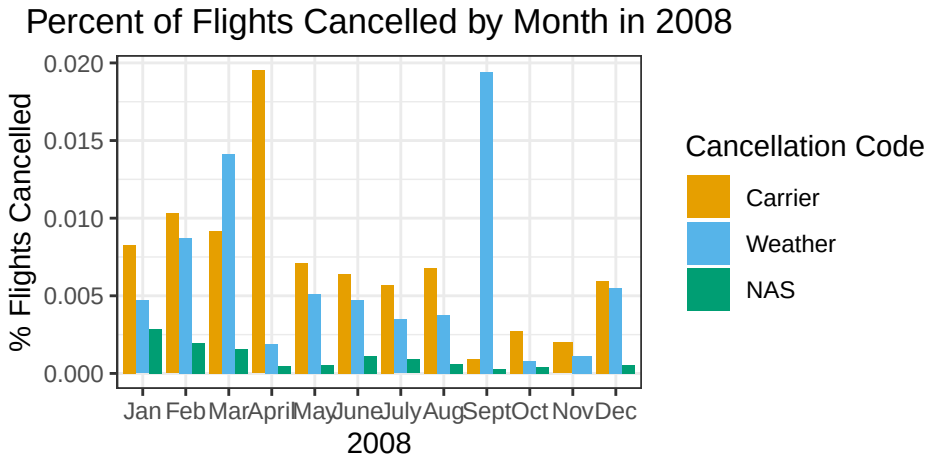
As we transition our focus from how flights are delayed for longer periods of time seasonally, we can also look at the different types of cancellations and delays. The graph below enables us to compare three types of cancellation codes: Carrier, Weather, and NAS. This plot is clearly able to show that NAS delays do not account for many delays and truly do not compare to weather or carrier delays.

```
## `summarise()` regrouping output by 'CancellationCode' (override with `.groups` argument)
```



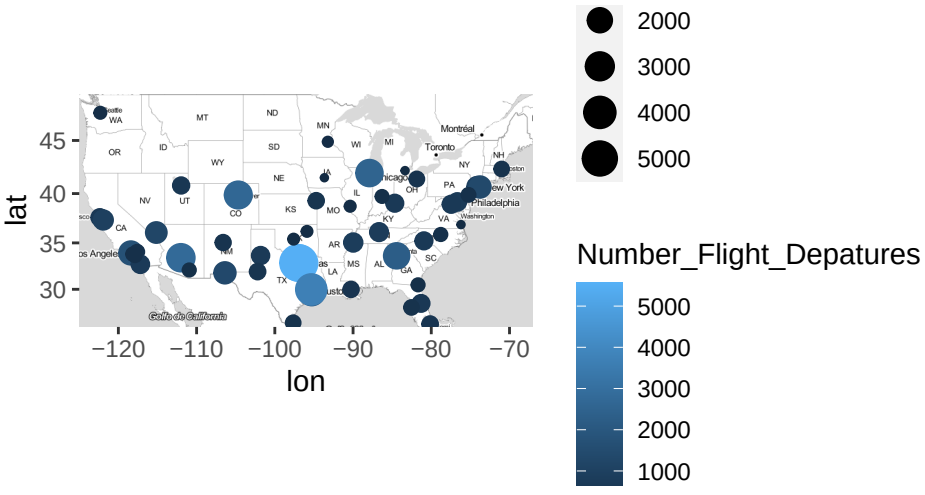
In the graph below, we look at a similar trend but over months so we can look at a seasonal picture of the flights. Some interesting anomalies are found in April and September. Based on knowledge of seasonal changes in weather, one could conclude that the significant flux in weather delays in September could be explained by hurricane season. As for April, carrier delays could increased significantly because American Airlines canceled a lot of flights due to not meeting government regulation standards.

```
## `summarise()` regrouping output by 'CancellationCode' (override with `.groups` argument)
```

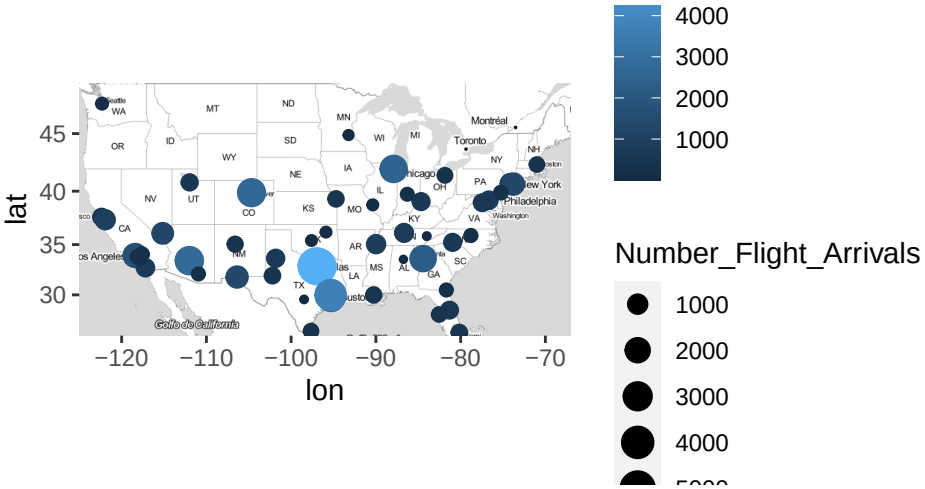


Strategic planning is a fundamental component of how airports sustain themselves. Knowing that an airport or region is notorious for delays or that they are notorious for this seasonally, plans can be made to accommodate to yield the best options for travelers. A simple way to summarize this information in a plot is through geographical mapping. Figure _____ shows this; airports with larger volumes of flights arriving in or departing from Austin have larger and brighter circles. An airport official should show more concern for a large dot that is lightly shaded, as this airport has a lot of flights outgoing to the Austin airport so schedules in these airports are integral to an efficient airport.

```
## Source : http://tile.stamen.com/toner-lite/4/2/5.png
## Source : http://tile.stamen.com/toner-lite/4/3/5.png
## Source : http://tile.stamen.com/toner-lite/4/4/5.png
## Source : http://tile.stamen.com/toner-lite/4/5/5.png
## Source : http://tile.stamen.com/toner-lite/4/2/6.png
## Source : http://tile.stamen.com/toner-lite/4/3/6.png
## Source : http://tile.stamen.com/toner-lite/4/4/6.png
## Source : http://tile.stamen.com/toner-lite/4/5/6.png
## `summarise()` ungrouping output (override with `.groups` argument)
```

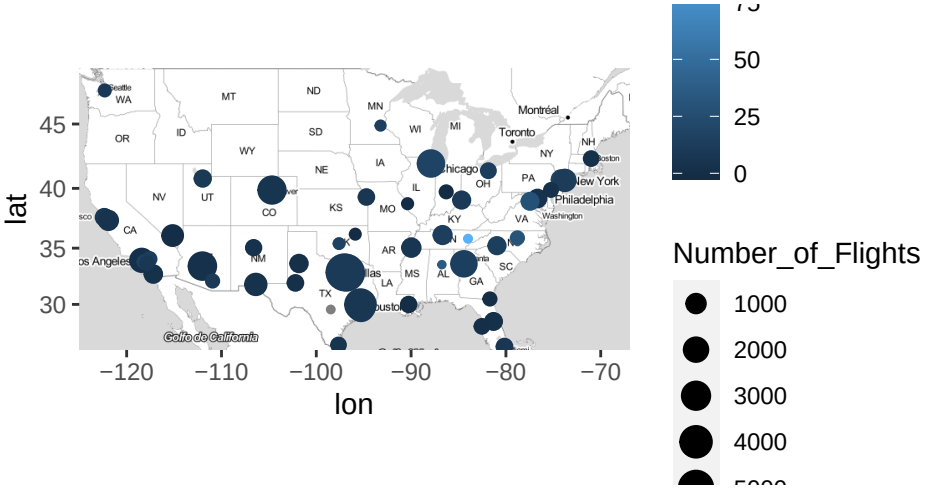



```
## `summarise()` ungrouping output (override with `.groups` argument)
```

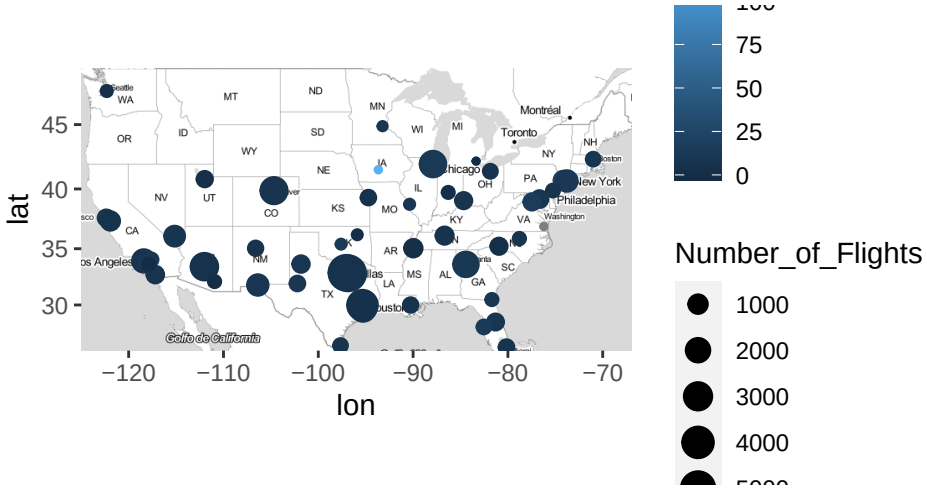


Now that we have an understanding of where the most frequented airports are, we can now look at a similar map that tells us how timely these airports' flights are. Similar to the most frequented locations, in the graphs below we see the locations with the most airports still have larger circles, but lighter colors represent longer delays on average for the year. We know from previous graphs the length of delays is somewhat seasonal, but on average we see that the more frequented airports do not have relatively long delays. We actually notice in states like Iowa there is not a lot of flights from here but they have a significantly higher average length of delay. We see something similar for flights arriving in Tennessee.

```
## `summarise()` ungrouping output (override with `.groups` argument)
```




```
## `summarise()` ungrouping output (override with `.groups` argument)
```



Question 3

Question 4

Question 5

Question 6