My research area is flight delays.

﻿- Flight delays are common occurrences in the airline industry, and these events have enormous business disadvantages.

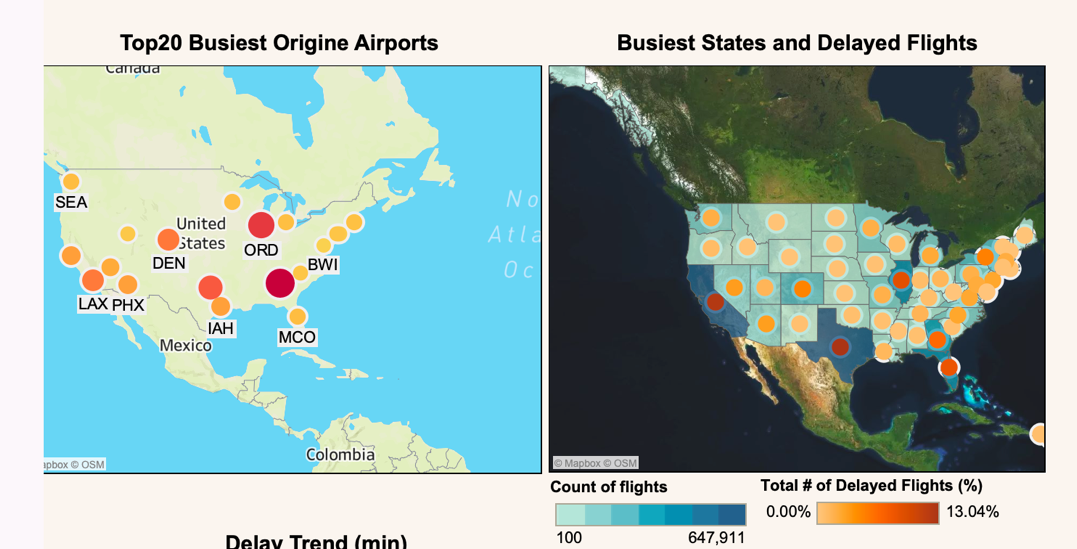
- This study uses a dataset published by the US Department of Transportation’s (DOT) Department of Transportation Statistics that surveyed domestic flights operated by a major airline in 2015.

During the EDA, the understanding of delay mechanisms or background can support developing a delay prediction model in the next part. The aim is to explore the various factors that cause delays.

------------------------------------------------------------------------

In this part, I’m going to explain briefly the finding of EDA about flights delay.

We can highlight or filter airlines that you want to choose.

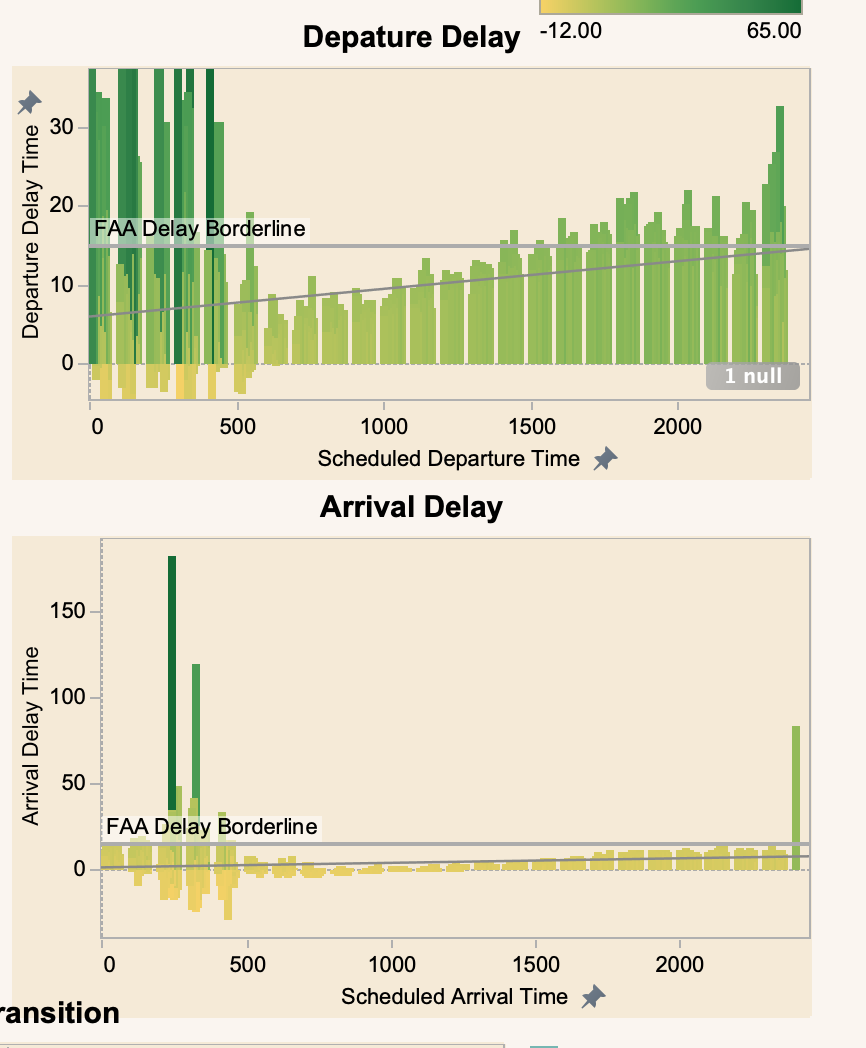


Here, there are a total of 13 infographic data.

The first geographical map shows the top 20 airports with the most flights. Airports with many flights are shown in darker red and size of circle. From this figure, it can be seen that a great number of flights have departed from ATL and ORD. Furthermore, it can be judged that there are many flights at the following three airports, DFW, DEN, and LAX. From this, we can see the main airports used in the northern, eastern, central, and western areas.

The second geographic data shows the total number of flights in each state with a blue color bar. Also, we can see the states with the most delayed flights on circle mark with yellow to red circle mark.

We can find a large number of delayed flights in Texas and California, followed by Florida, Georgia, and Illinois.

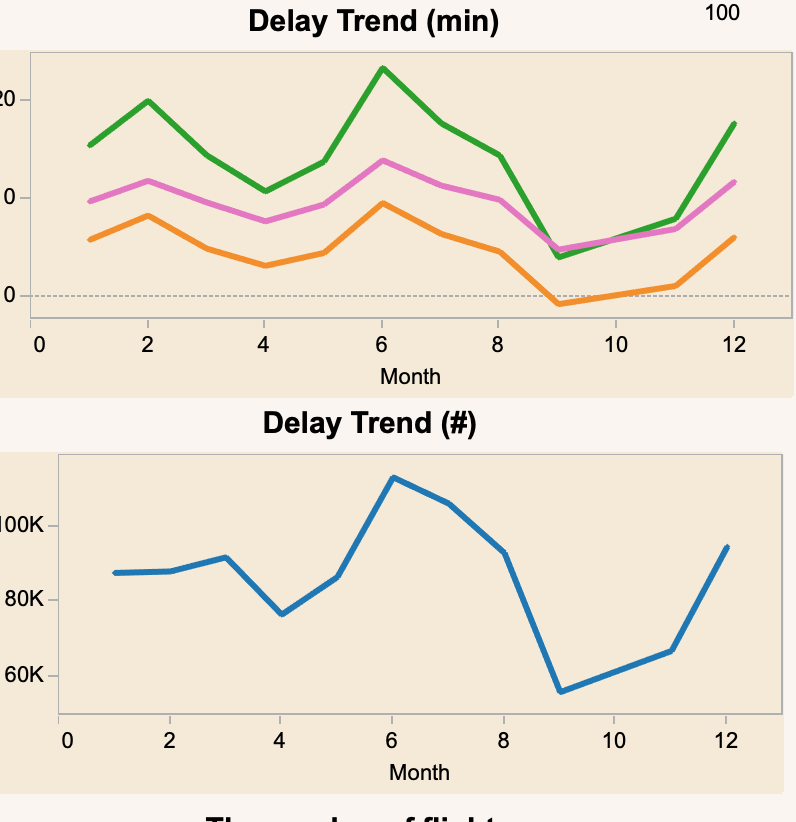


The two bar graphs show the average delay time at the scheduled departure time and arrival time. The Federal Aviation Administration (FAA) considers a flight to be delayed when it is 15 minutes later than its scheduled time. So, the Borderline is placed at 15 minutes. Departure delays of more than 15 minutes increase after 3 PM and continue to increase from 5 PM to 4 AM. It also shows that late flights and early morning flights can cause very long delays of 30 minutes or more.

On the other hand, between 5 AM and 11 PM, it arrives within a delay of 15 minutes from the scheduled time, but from 12 AM to 4 AM, a delay of 15 minutes or more is likely to occur.

From this, it is shown that a delay of about 20 minutes at the time of departure is unlikely to affect the delay of the arrival time.

However, it can be inferred that it is difficult to avoid the delay of the arrival time by the long delay at the time of departure.



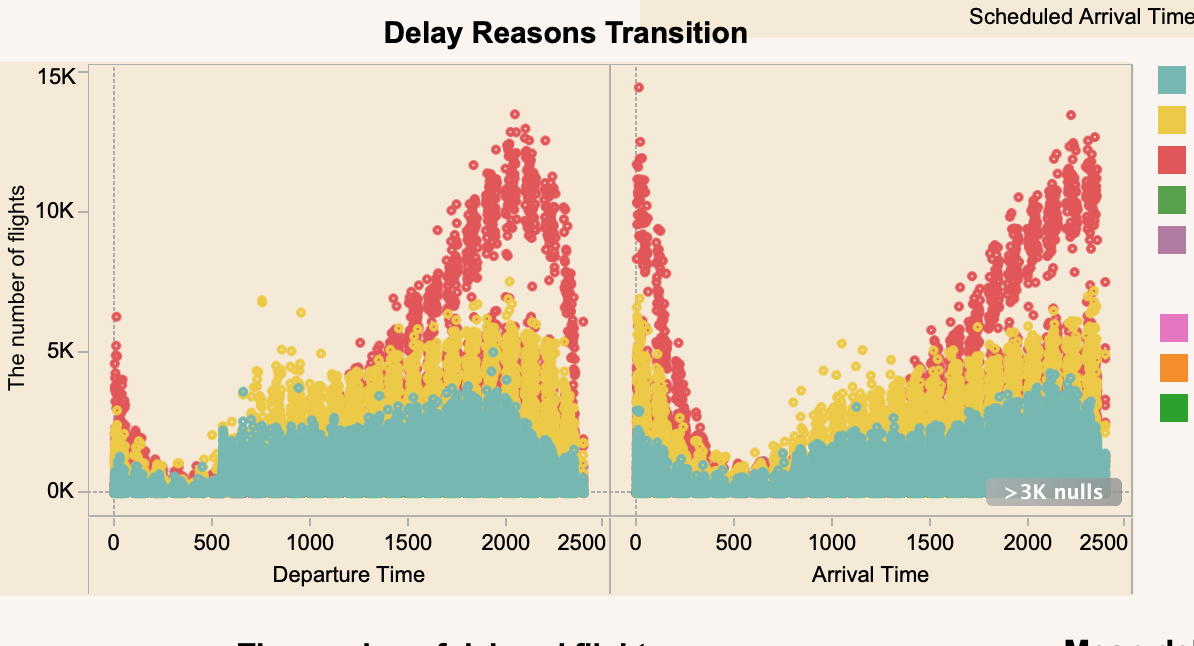
The line graph above shows the trend of average delay times per month.

We can see that the delay time tends to be longer in February and June.

The line graph below shows the trend in the number of delayed flights per month.

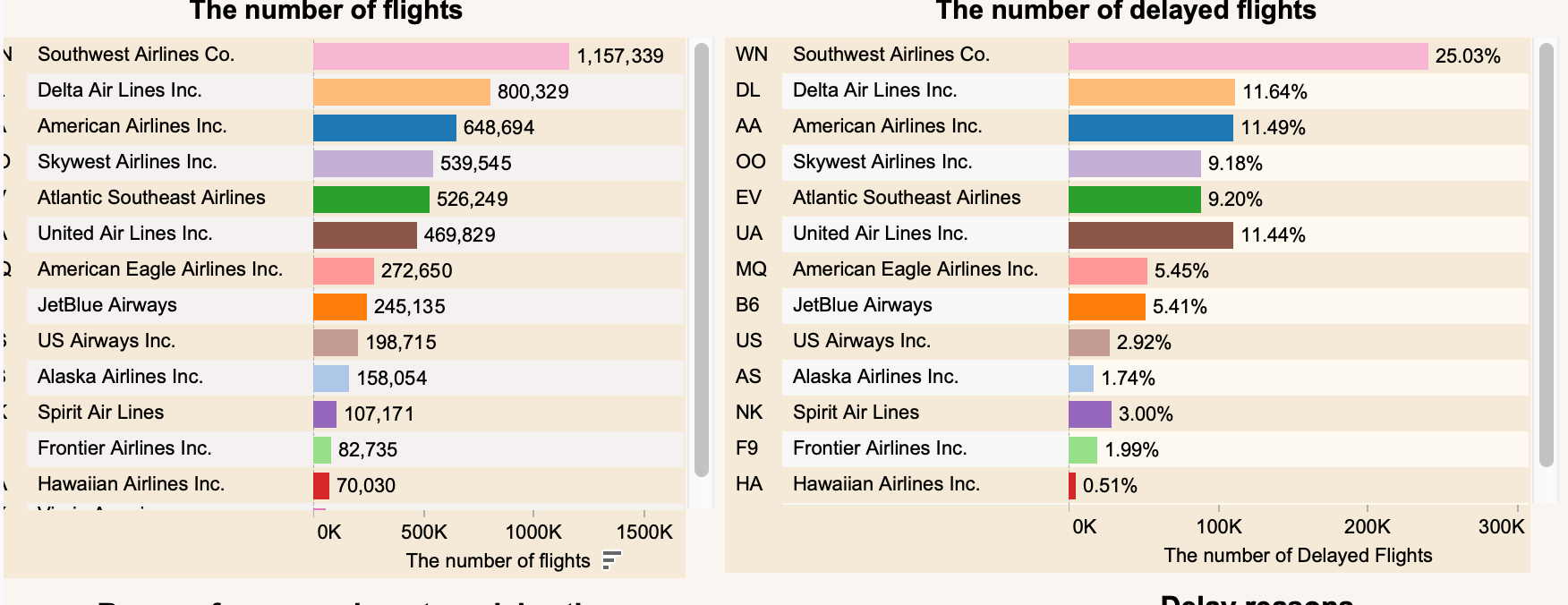
The number of flights delayed is high between June and August, and it is highly likely that flights will be similarly delayed from December to March.

------------------------------------------------------------------------



This scatter plot shows the number of delayed flights at departure and arrival time and the reason for the delay.

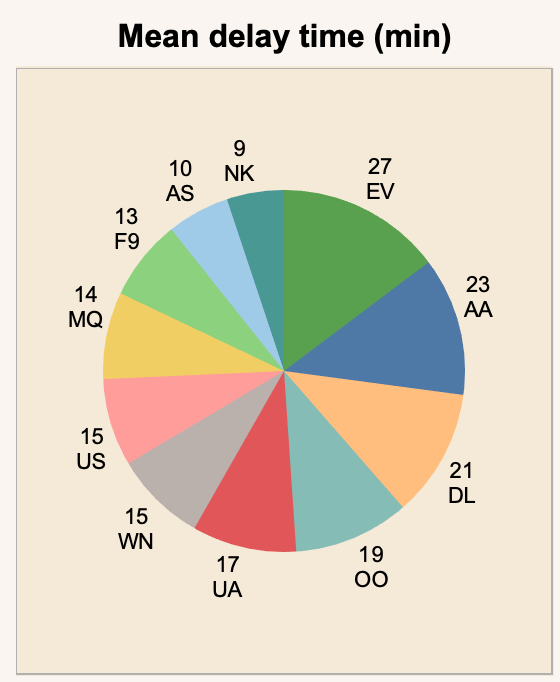
The delay due to Late Aircraft is noticeable for a long time, but it can be seen that the delay due to Airline is more outstanding from 5AM to 12PM.



The next two bar graphs show the total number of flights and the proportion of delayed flights for each airline.

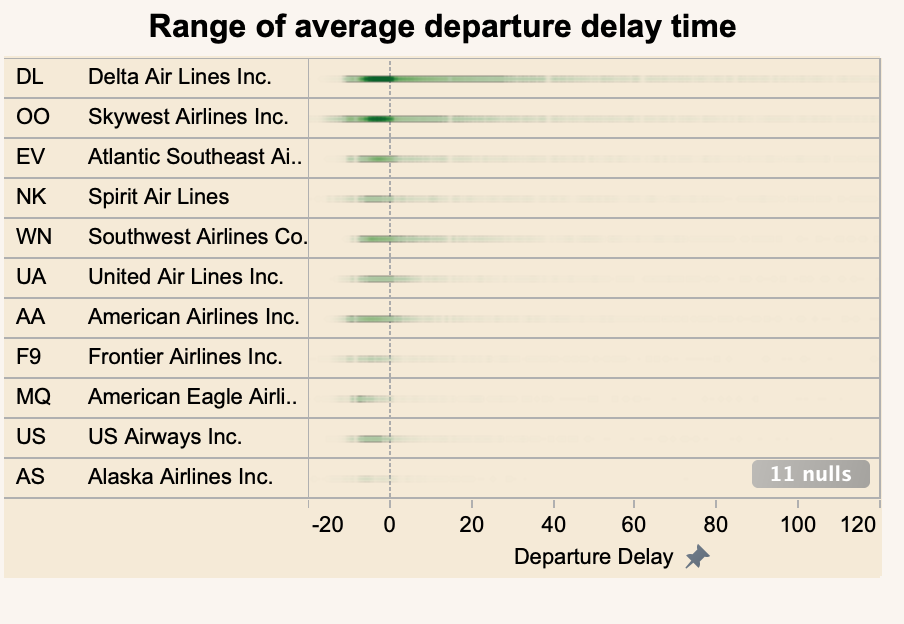
WN has the largest number of flights and proportionally the largest proportion of delayed flights. On the other hand, UA has fewer flights than DL and AA, but the proportion of delayed flights is the same as DL and AA.

------------------------------------------------------------------------



Next, I explain what a pie chart describes. The pie chart shows the average delay time for each airline. Interestingly, there is no big difference in delay time, and it can be implied that all airlines try to suppress the delay in about 20 minutes.

------------------------------------------------------------------------

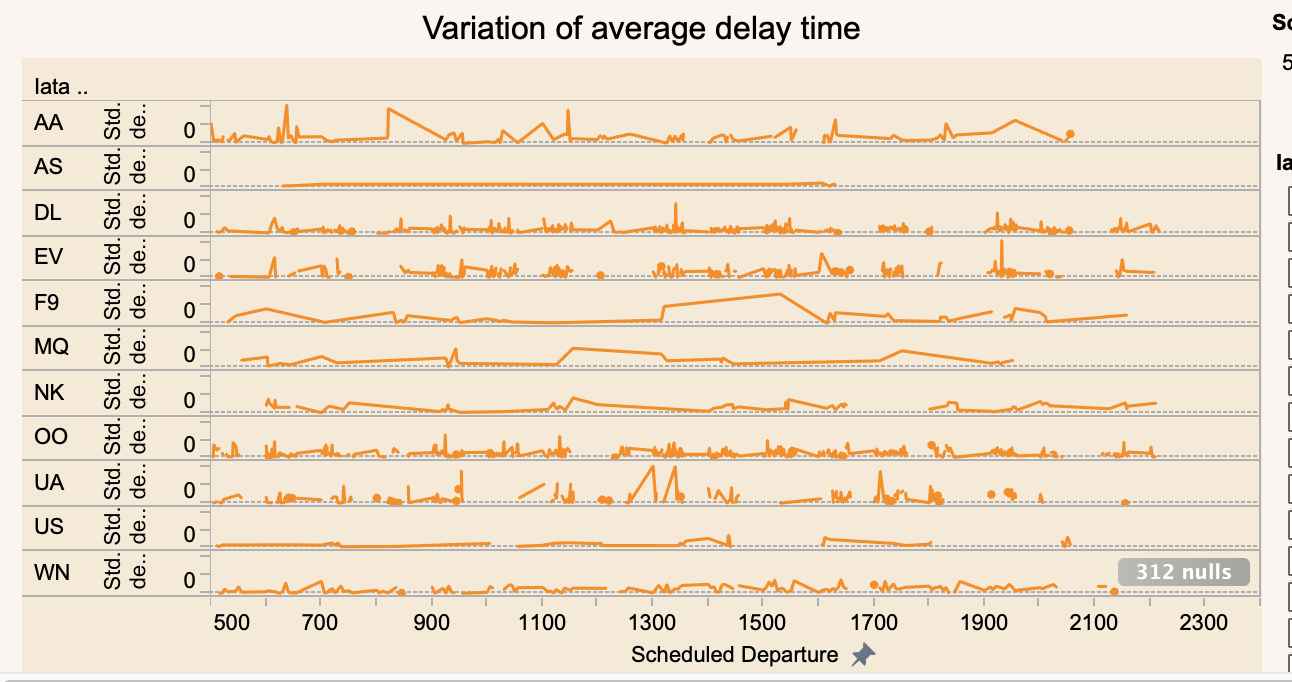


For a closer look, let's look at the graphs that show each airline and the average delay time.

Most airlines realize the departure several minutes before the scheduled departure time, and it is conspicuous that the departure is within 20 minutes even if it is delayed. It turns out that Southwest Airlines, Delta, United, and American Airlines are often delayed by more than 20 minutes, but they are likely to depart within an hour.

We see that while all mean delays are around 20 minutes, this low value is a consequence of the fact that a majority of flights take off on time or before scheduled time. However, we see that occasionally, we can face really large delays that can reach a few hours.

------------------------------------------------------------------------



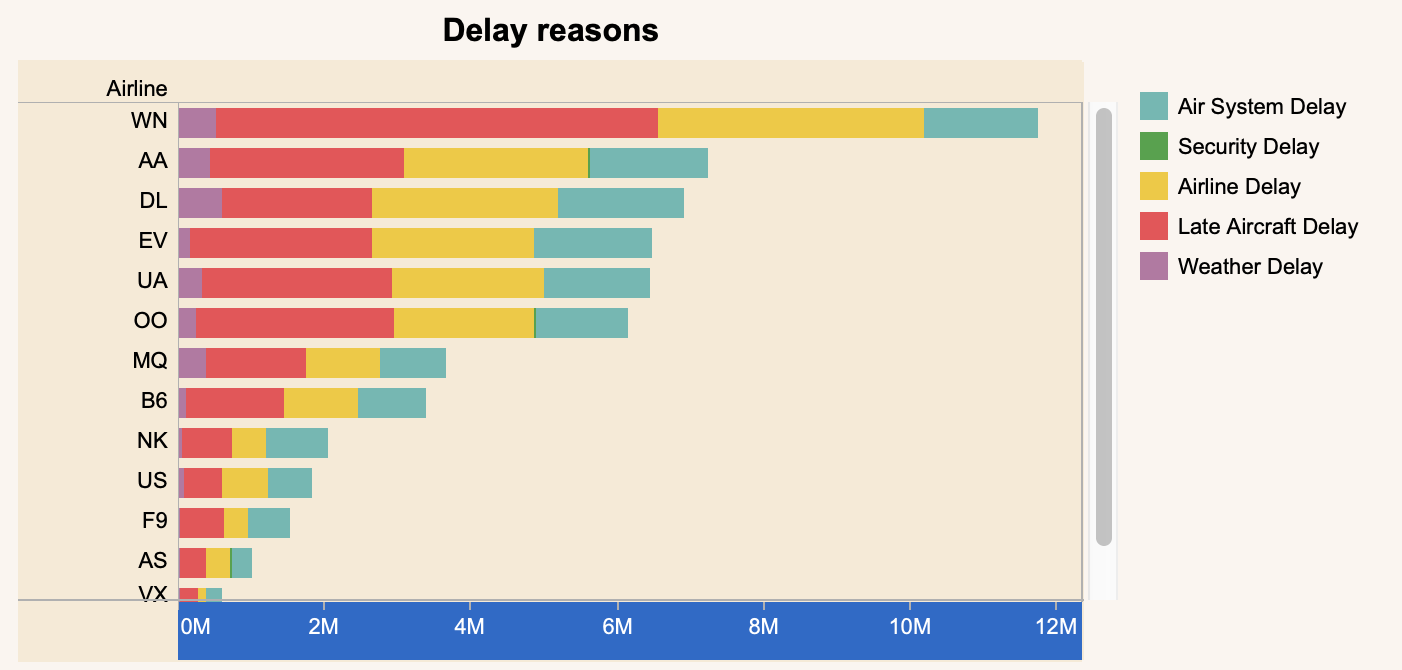
Next, we will look at the variation in the average delay time for each scheduled flight time for the four companies with the most flights.

The yellow line explains the standard deviation of delay time.

It can be seen that South West Airlines have relatively small variations in a delay time in scheduled departure times.

On the other hand, for United, American and Delta Airlines, the variation in delay time is quite large.

------------------------------------------------------------------------



Finally, the bar chart shows the number of delayed flights on each airline caused by the five delay reasons. We can see that the reasons for delays for each airline tend to be different.

Delays are fairly rare for security reasons, while the predominant delay reasons are late aircraft, airlines, and air systems.

------------------------------------------------------------------------

The previous part dealt with an exploration of the dataset, with the aim of understanding some factors of the delays registered by flights.

> The top 5 busiest airports are ATL, ORD, DFW, DEN, and LAX.

> The top 5 busiest states are TX, CA, FL, IL, and GA.

> The departure delays tend to increase between 5 PM to 4 AM.

> The arrival delays from 5 AM and 11 PM are highly likely to be within 15 minutes.

> The flight delay time is longer in February and June.

>The number of flight delays is higher from June to August and December to March.

> The late aircraft delay is noticeable for a long time, but airline delay is more outstanding from 5 AM to 12 PM.

> A quarter of WN flights are delayed.

> UA with fewer flights has the same proportion of delayed flights as DL and AA with many flights.

> All mean delays are around 20 minutes, however, we occasionally can face really large delays that can reach a few hours.

> Compared to United, American, and Delta Airlines, WN has fewer variations in delay time.

> Delays by security reasons are fairly rare, while the predominant delay reasons are late aircraft, airlines, and air systems.