This dataset contains information of over 7,000,000 flights in the United States, including carriers, arrival, departure delays, and reasons for delays in 2008. This dataset can be found here: http://stat-computing.org/dataexpo/2009/2008.csv.bz2

I have also merged an additional dataset to get the specific airline for each flight found here: http://stat-computing.org/dataexpo/2009/carriers.csv.

I decided to concentrate on the airline, departure delay, the day of the flight, and the scheduled time of departure.

In this investigation, I wanted to look at the average flight delays with regards to the day of the week, the time of the scheduled departure, and the specific airline.

After cleaning the data, I explored a number of different variables including the reasons for departure delays, average departure delays depending on the time of day or day of the week, and also the departure delays separated by airlines. I wanted to know when the best time to schedule a flight would be and further, what airline would be the best choice with regards to expectation of delays. I used line plots and bar plots to visualize this information.

When plotting the different airlines I had to use a color palette that included 20 different colors and I tried to make them as distinguishable as possible. I concluded that the worst weekday to fly is Friday and Sunday while Wednesday and Saturday were the most reliable. Furthermore, earlier flights, from 6 to 9 am usually experience shorter delays than later flights in the evening, like 5 to 9 pm.