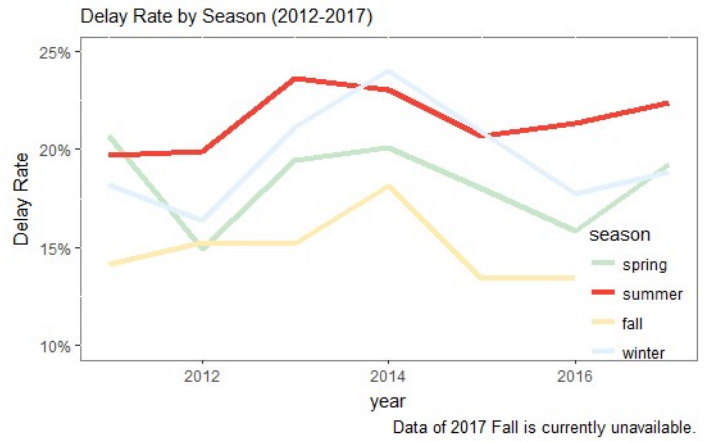
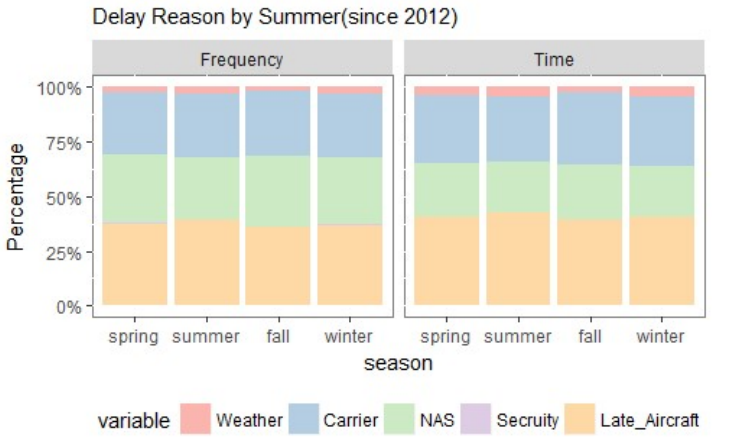
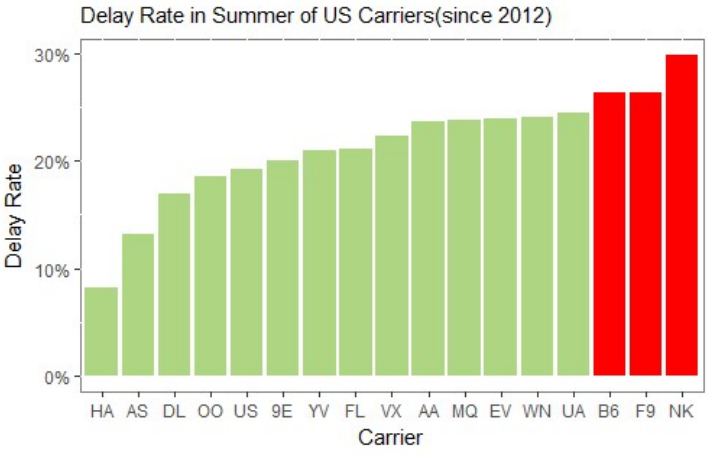
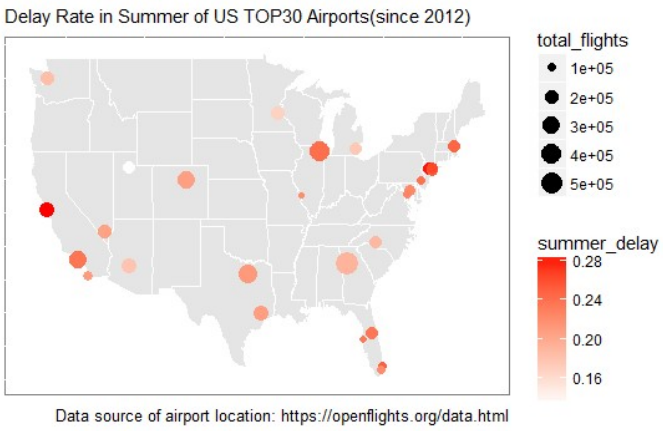
High delay rate in summer: why and how to avoid it?









My visualization consists of four plots. The first plot describes the phenomenon that delay rate in summer is higher than other three seasons. By calculating the sum of total delays over the sum of total arrival flights, I got the delay rate for different seasons. Then since every year's delay situation may be fluctuating, I draw a line chart of different season's delay rate of years 2012 to 2017. And to highlight the data of summer, I choose an eye-catching color for it.

To keep the data updated, all analysis of plot2-4 are based on data since 2012. The second plot analyses the reasons of summer's high delay rate. That is to calculate the fraction of delays or time of delay caused by different reasons. However, the stacked bar plot shows only slight difference both for the frequency of delay and lasting time of every delay. And this plot shows that stacked bar plot might not be suitable to reveal such trivial differences.

The 3rd and 4th plots describe the delay rate in summer in terms of different carriers and major US cities. The delay rate is calculated by same practice of plot 1 and grouped by carriers or cities. With these references, people can make decision to try to avoid some airlines or cities in summer. For carriers, a bar plot clearly shows different delay rates of carriers. Carriers with top 3 delay rate is highlighted by red color. For cities, in order to let viewers discover the distribution of cities with high delay rate, a scatter point plot based on US(mainland) map is drawn. And we can find the cities near the sea are likely to suffer from high delay rate in summer, especially for cities in northeast and southwest coastlines. In the plot, the size of point stands for the number of total arrival flights of an airport, and the redness measures the delay rate in summer.

For last plot, data from package maps is used for drawing the map of US, but all the plots are done by ggplot2. Types of plots include bar chart, line chart, polygon chart and point chart. Package dplyr and reshape2 are used for manipulating the data. Package scales is used to control the scale of axes.