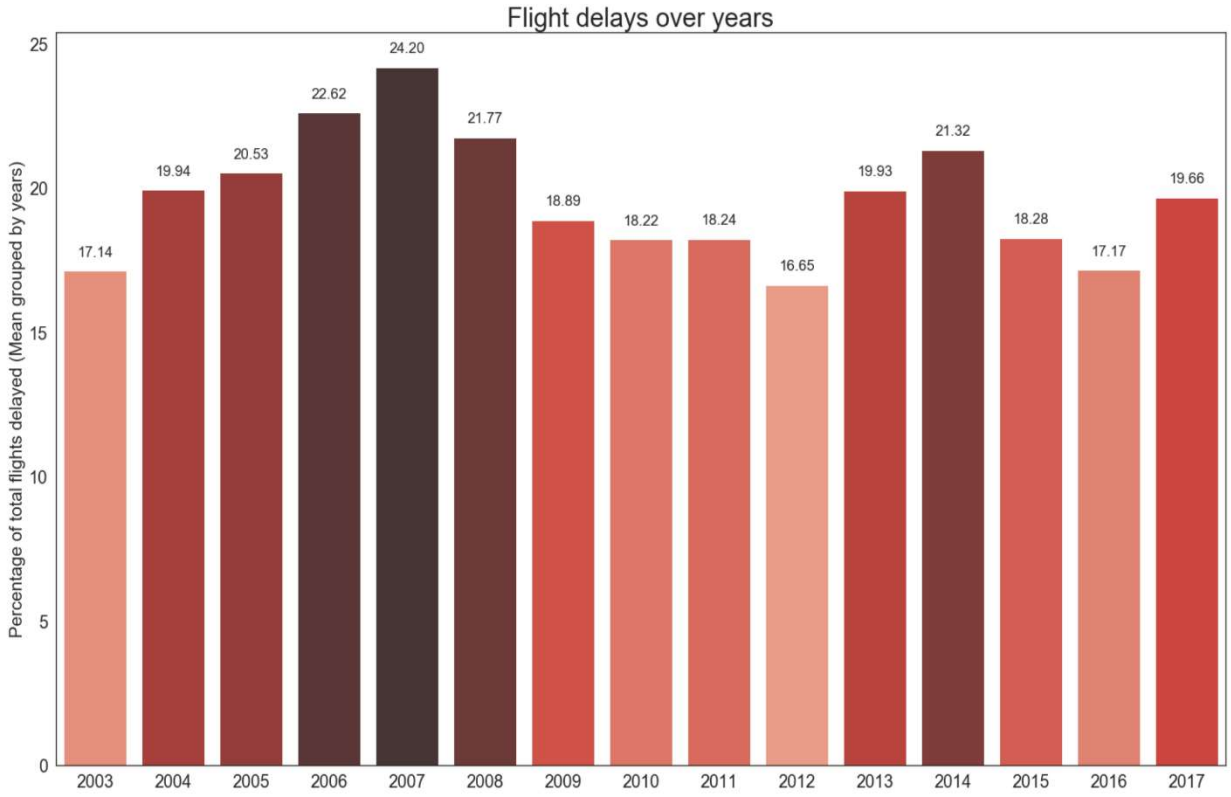
Weather and security delay most of our flights- Myth or Truth?

Most of us believe that weather is the main reason why your flight is delayed?

Let's dig into data to find out more about it. To start with, let us begin by taking data for flights from 2003 to 2017 at all the significant airports(~380) of US. Shockingly, percentage delays have not improved much over the years. Can you believe even with all the advancing technology hype, even in the year 2017, almost 20% of the total flights were delayed. In below graph darker the color worse the percent delays.

Group all the delays by year, sum those up, find the percent of delays per arrivals and take its mean to plot it on the Y-axis, while the X-axis serves as the time period in years.



**Created in python using seaborn library.*

But, what can technology do against powerful nature? The graph below shows the number of flights delayed due to various reasons. Aviation System Performance Metrics(ASPM) categorizes delays as below.

1. Carrier delay (within the control of the air carrier) examples are : aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, cargo loading, computer, outage-carrier equipment, crew legality (pilot or attendant rest), fueling, handling disabled passengers, late crew, lavatory servicing, maintenance, removal of unruly passenger, slow boarding or seating, stowing carry-on baggage, weight and balance delays.

2. NAS (National Aviation System) Delay includes non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control, etc.

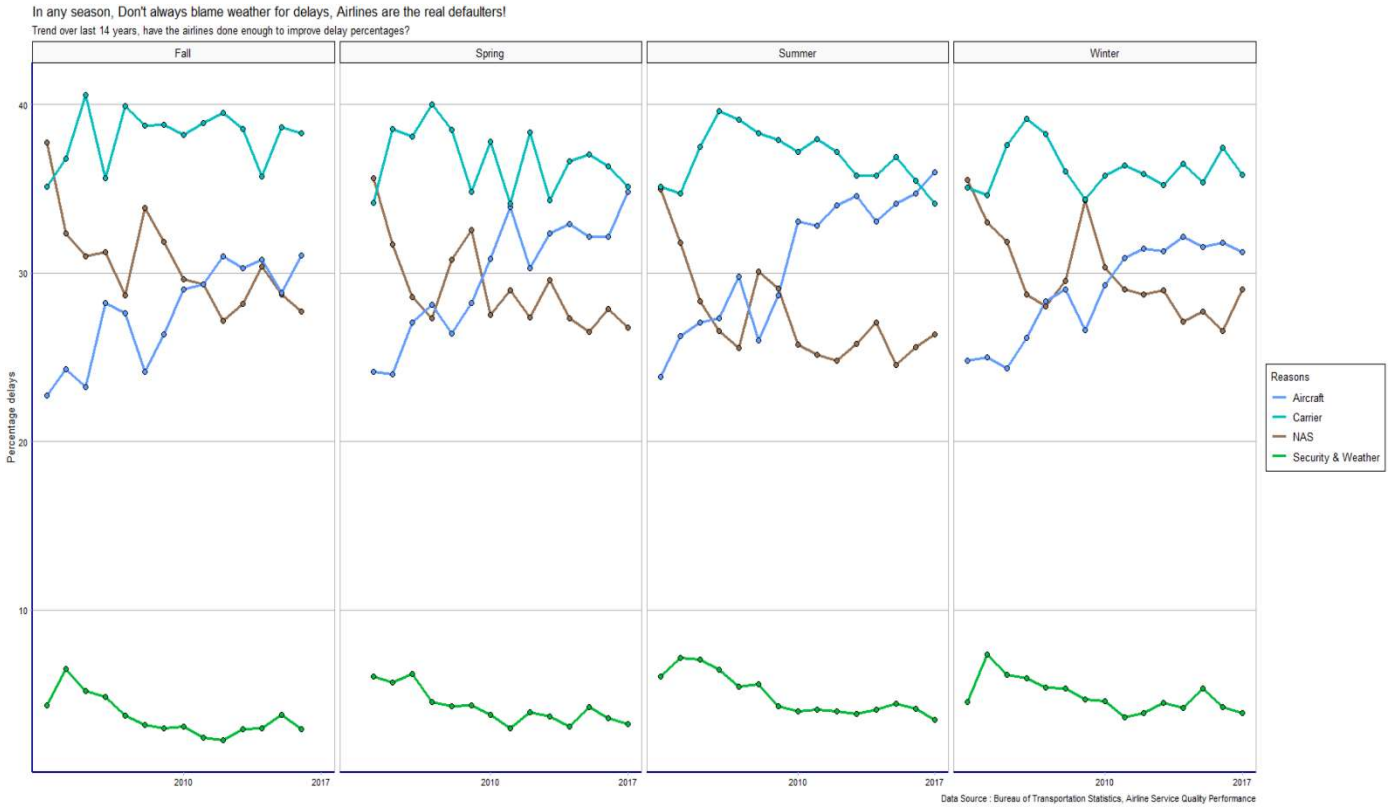
3. Aircraft Delay is due to other aircraft delays.

4. Security delay is caused by the evacuation of a terminal or concourse, re-boarding of aircraft because of the security breach.

5. Weather delay is caused by extreme or hazardous weather conditions that are forecasted or manifest themselves on point of departure, enroute, or on point of arrival.

Using Months columns, I added a new column to identify the season as Fall, Spring, Summer, and Winter. After doing all the complicated code stuff of aggregations, facet wrapping, axis scaling, and aesthetics manipulation in R and ggplot2 library, below is a summarized elegant graph.

Let's look at the trend in delays caused in different seasons So that we can say Summer flights are not often delayed!



**Created in R using ggplot2 library.*

Much to our surprise, of total delays, the number of delay due to weather & security (contributes only 0.5%) are less than 5%! In all the seasons, the delay caused by weather does not show any change. It is the operations at the airport which causes most of the delays. The operational efficiency of airports should be improved to reduce the delays.

Note: I imported the data file given for homework and cleaned the data in Tableau. Cleaned data file is attached with the code repository zip file.

While exploring in data in tableau for cleaning, I joined data from an external source(https://openflights.org/data.html) which gave longitude and latitude to plot airports on the map with the data given in homework.(HomeworkData <LeftJoin> GeoData). Left join is used to not miss any data from homework set. Below is how it looks.