

# Flight data visualization using Tableau

*Author : Deepak Singh*

*Date : 04/17/2018*

## Link to my Tableau Workbook

[https://public.tableau.com/profile/deepak.singh1476#!/vizhome/Flights\\_Delay\\_Causes\\_2008\\_V2/Story](https://public.tableau.com/profile/deepak.singh1476#!/vizhome/Flights_Delay_Causes_2008_V2/Story)

## Summary

This dataset is obtained from the RITA website which contains information about flight delays and performance. The dataset I used is for 2008. I used two additional datasets **carrier.csv** and **airports.csv**. The datasets can be found in the following website:

<http://stat-computing.org/dataexpo/2009/the-data.html>

The dataset contained approximately 7 million observations along with 29 columns. I cleaned the data using python in jupyter notebook. After cleaning the data, I joined the data with **carrier.csv** – to get the carrier names and the resultant dataset I joined with **airports.csv** to get the city and state information for the airport code.

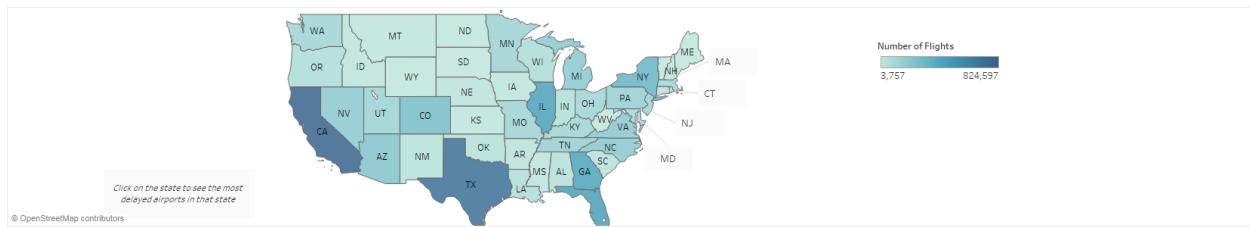
I have created the visualizations in Tableau that answer the following questions:

1. What are the busiest states and cities in those states where the flights originate?
2. What are the cities with most delays when all the airports are considered and what are the causes that contribute for these delays?
3. What are the carriers that cause the most average delays when all the flights are considered?
4. What is the overall Monthly trend for the several types of delays? Is there any correlation among the causes for delays?

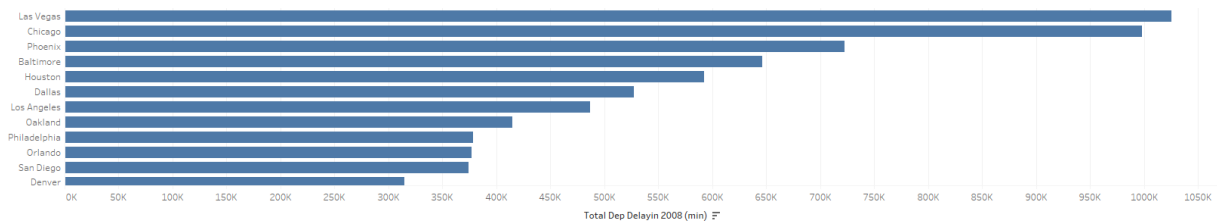
## Design

My initial design choice for the first point of the story was to see which states have the most traffic in terms of flights originating from them. Also, parallelly I wanted to break down even further by airports to see which are the most delayed airports within those states. For this reason, I created a dashboard with filled map where the intensity of the color represents the number of flights. Also, I created a bar chart with total delays for the cities. Here we can also filter by state and year to drill down even further.

## Flights Origin



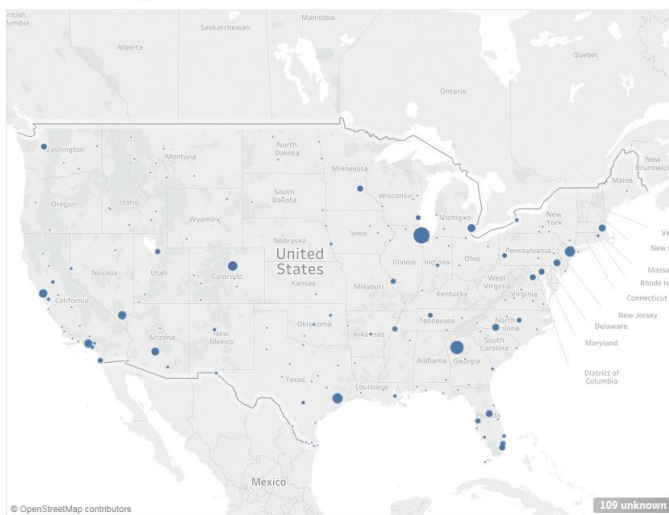
## Most Delayed Departure from City



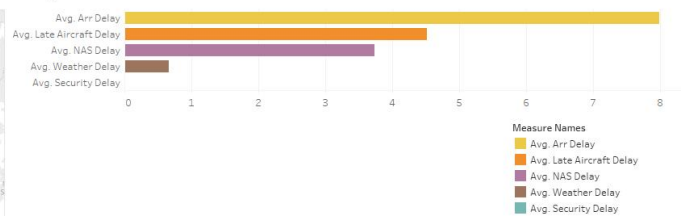
## Slide 1

My second design consideration was that I was wondering if there's any way to make a dashboard with geolocation plot for all the airports and causes for delays for those airports. So, in my 2nd plot I did just that. The size and the darkness of the points showed a redundant plot of how much delay was happening in an airport. The more the delay, the bigger and darker the point is. The bar chart in that dashboard gives the breakdown of the delays caused at the airport (when we click on the airport).

## Most Delayed City in 2008

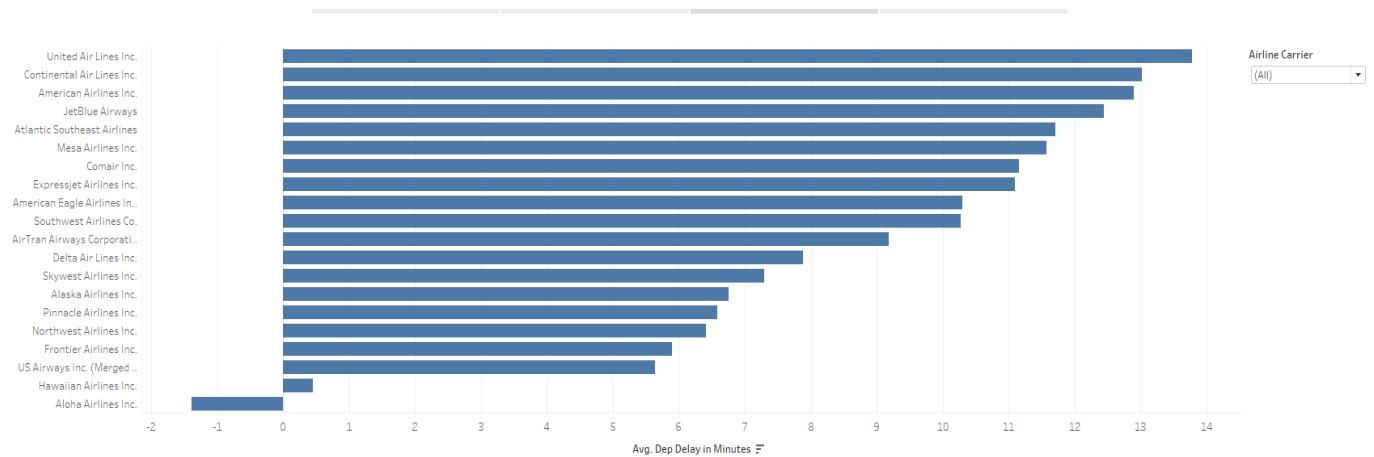


## Delay Breakdown



## Slide 2

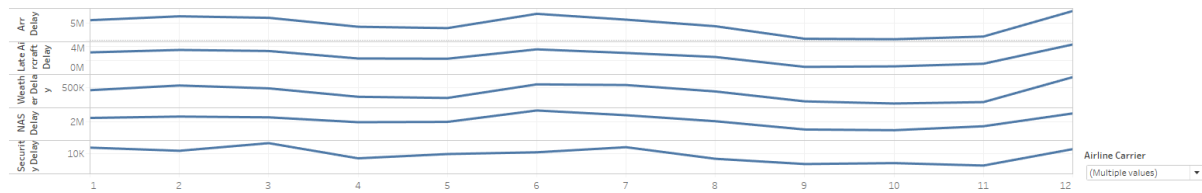
Now my next target was to see if there's a way to find out what is the average delay of the airline carriers when all the airports and all the years were considered. I created a bar chart for this information sorting in the descending order



### Slide 3

Following from the above, I wanted to check the monthly trend of several types of delays for 2008. For this, I created a line chart. Looking deeper in the trend lines, we can see that the Arrival, weather and late air craft delay are somewhat correlated and dependent. This intrigued me to explore are the airline carriers contributing to this? To provide the drill down capability I created the line chart for several types of delays for airline carriers. Drilling down on delays in terms of airline carriers we see the delay caused by Southwest and American Airlines follow the similar trend. This shows most of the Weather and aircraft related delays are caused by Southwest and American Airlines and hence the overall trend seems to be correlated. December seems to be the busiest month in terms of delays

Overall Delay trend - Monthly



Monthly Delay Trend w.r.t Airline Carrier



### Slide 4

## Feedbacks:

I received these set of feedback:

- 1. The title of the story could be clearer** – I changed the Title to “Causes for Delayed Flights in 2008” from “Flights\_Delay\_Causes\_2008”
- 2. The map of the US does not display the various states particularly clearly-** I used different settings from Map Layer option in Maps to rectify this
- 3. The legends were not displayed properly-** I changed the position of the legends and renamed the title to Number of flights in slide 1
- 4. On the 3<sup>rd</sup> slide, not easy to Interpret Dep Delays because it is in millions-** I changed the aggregation to look at the average departure delay in minutes for the airline carriers.
- 5. Corrected some of the punctuation errors and spelling errors**