

# Tableau Project

## Links

- 1) Link to the first version  
[https://public.tableau.com/profile/yanyan.liu#!/vizhome/hflights\\_0/Dashboard2](https://public.tableau.com/profile/yanyan.liu#!/vizhome/hflights_0/Dashboard2)
- 2) Link to the final version  
[https://public.tableau.com/profile/yanyan.liu#!/vizhome/hflights\\_tableau\\_projects/Story1](https://public.tableau.com/profile/yanyan.liu#!/vizhome/hflights_tableau_projects/Story1)

## Summary

The data set used is the flights data set in hflights R package, which contains all flights departing from Houston in 2011, totaling 227496 records. From the visualization, we can see four top carriers with large amount of records in the datasets are: XE, CO, WN and OO. Among the four reasons that leads to airline cancellations, weather is on the top list, which occurs mostly in February. Departure delay and arrival delay in minutes positively correlate with each other, which is not a surprise, and minutes in delays tend to top on Thursday .

## Design

- 1) For graphing the flight records number attributed by each carrier, a bar chart is used. A bar chart is the best choice when we want to study single categorical feature. Here the single categorical feature is flight record count. Month is used as a filter, and thus we can see the record for each carrier by month.
- 2) Similar to the record count, the cancellation code is a categorical feature, I used bar chart again to see the count in each category. Of the four categories, I selected the top two categories (weather, carrier) to study in detail. To compare the change in each category month by month, I put the graph for weather on top, and the graph for carrier on bottom. I used different colored bar for the two different categories. Unique carrier is used as a filter to see the cancellations of each carrier month by month.
- 3) To see the correlation between departure delay and arrival delay, a scatter plot is used. Filters are used on both carriers and month to let the reader to study the detail. Each dot represents one record, and record on different week days are represented using different colors. Feature delay and arrival delay show a positive correlation as is seen from scatter plot, but there is a lot overplotting. An aggregated dual plot for departure delay and arrival delay for each carrier, using month as a filter.

- 4) To study more in detail about departure delay and arrival delay, strip plot is used to show the delay on each day of the week filtered by month. This is a busy graph, but it reveals completely the arrival delay and departure delay information. The month filter allows readers to do more detailed study.

#### **Changes made after collecting the feedback**

- 1) For my first version, the review suggested that I should apply the filter to the whole dashboard if possible. I used this suggestion on every dashboard if that is applicable.
- 2) For my first version, the review also suggested dual axis on departure delay and arrival delay. I took her suggestion and used dual axis on one of my departure delay and arrival delay graph.

## **Resources**

<https://github.com/hadley/hflights> for feature information

<https://www.youtube.com/watch?v=CTjASQXnNzA> for how to make dual axis graph

## **Feedback**

Link:

<https://discussions.udacity.com/t/tableau-project-review-help-needed/333304>

Copied from the feedback:

*Please find the following observations,*

*1 - You can use " Dashboard Action" to apply interactive filters to show the relationship between two worksheets on your dashboard.*

*2 - for 2nd worksheet, I think you can use "Duel Axis" to compare Departure Delay and Arrival Delay*