Udacity Tableau Project
Amit Kumar Shankar
Iteration – Improvements After First Review

Step One:

Chose the Flights data set from https://docs.google.com/document/d/1w7KhqotVi5eoKE3I_AZHbsxdr-NmcWsLTIiZrpxWx4w/pub?embedded=true .

I tried importing all the csv files into Tableau, but the free version only allows for 15,000,000 rows of data. Therefore, I chose to analyze the two most recent datasets: 2007 and 2008 flight data.

I created a union of 2007 and 2008 dataset and then an intersection with the carrier's dataset to extract the proper names for each unique airline carrier.

For feedback purposes, I showed the dashboards and storyline to a friend.

Calculated Field:

Since the dataset contains 'Year', 'Month' and 'Dayof Month' fields, I combined those fields together to create a Reference_Date field. First, converted the fields to string, then concatenated them using '-' and parsed it using the DATEPARSE function. This resulted in a proper date and time. Since the time was not needed, applied the DATE function to only keep the DATE in Reference_Date field.

DATE(DATEPARSE("YYYY-MM-dd",STR([Year])+"-"+STR([Month])+"-"+STR([Dayof Month])))

Note:

The story board best renders on a chrome browser in full screen mode. I had to decrease my screen resolution to 75% to make screen shot images for documentation.

Initially, I had also included the Tableau Workbook for each of the dashboard and story but the file upload on Udacity is only 500mb, so I removed them. These files can be downloaded directly from Tableau Public Page by following the respective links in the discussion.

I also had to remove the 2007.csv and 2008.csv datafiles from folder as they are more than 500mb each. The 2007.csv and 2008.csv can be directly downloaded from http://stat-computing.org/dataexpo/2009/the-data.html .

Iteration based on first Udacity Review:

Udacity Reviewer - Thank you for pointing out that the titles were missing from graphs in Story. Initially, the titles were present in the worksheets but disappeared when dragged onto the story board. However, embedding a worksheet in dashboard and then embeding the dashboard showed the titles in Story board.

Also, thanks for text tile tip. I added the text in my dashboard and then embedded the dashboard to the Story board.

Good workflow I learnt is :
create worksheet -> embed worksheet to dashborad and polish your presentation-> embed dashboard to storyboard

Final Workproduct:

https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Story_v4/ExplorationofFlightsDataset-Story-Final

Description About Files

carriers.csv - data file with information on carriers
Documentation.doc - project documentation for review in doc format
Documentation.pdf - project documentation for review in pdf format

Step Two:

Summary: I explored the Flights dataset for 2007 and 2008. I was interested in finding more information about delays and how they varied across different carriers. Also, when did most delays occur as well.

Design Choices:

Visualization 1: Udacity_Tableau_Project_Dashboard_01_v1

Link:

https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Dashboard_01_v1/Dashboard1
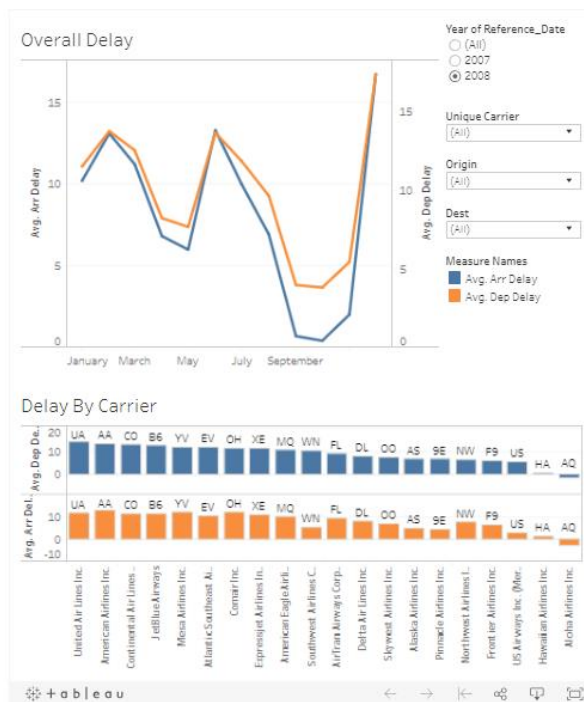

I tried to create the above dashboard to show the differences in delay times by month and carriers. I tried to use line graphs and bar graphs because in this case they show differentiation well, but from **feedback** found the dashboard was 'too busy'. Therefore, I created the next dashboard to only focus on average delays by month and day for exploration.

Visualization 2: Udacity_Tableau_Project_Dashboard_01_v2

Link:

https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Dashboard_01_v2/Dashboard2
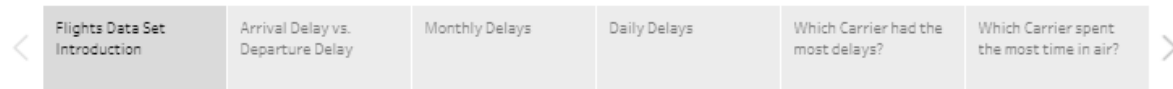
Image:
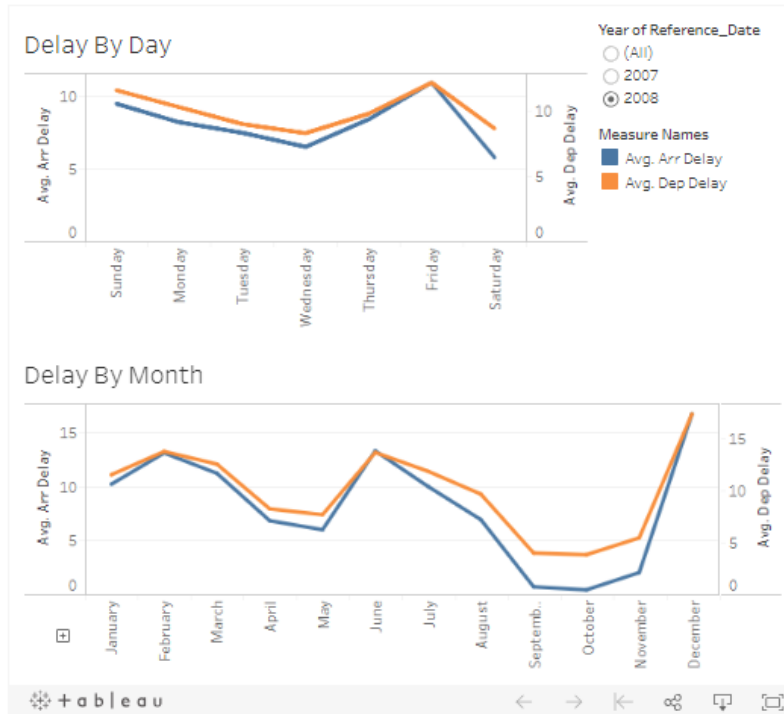
The above dashboard showed the difference in delay times across weekdays and months. I chose the line graph to show the differences in arrival and departure delay. The feedback was positive – the viewer was able to clearly identify the days/months of peak delays.

Visualization 3: Udacity_Tableau_Project_Story_v3

Link:

https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Story_v4/ ExplorationofFlightsDataset-Story-Final
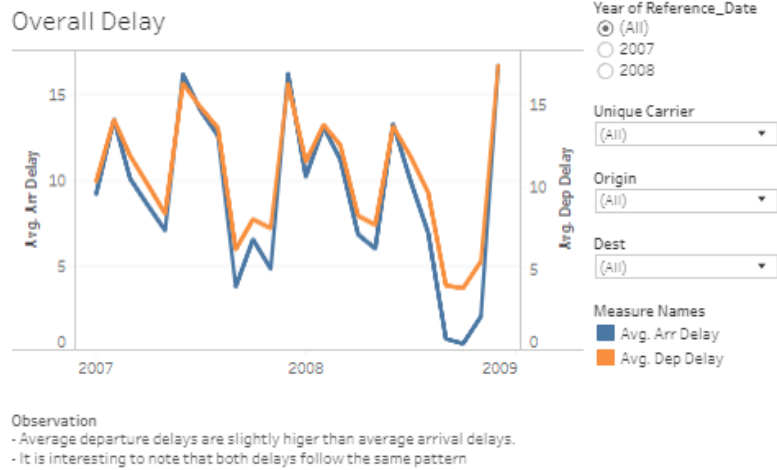

I created the Introduction slide to give some background information to the user about the dataset and the limitations of the free version of the Tableau software. Feedback – the user wanted to know what the presentation was about, and the Introduction slide answered that question and set the expectations of the presentation.
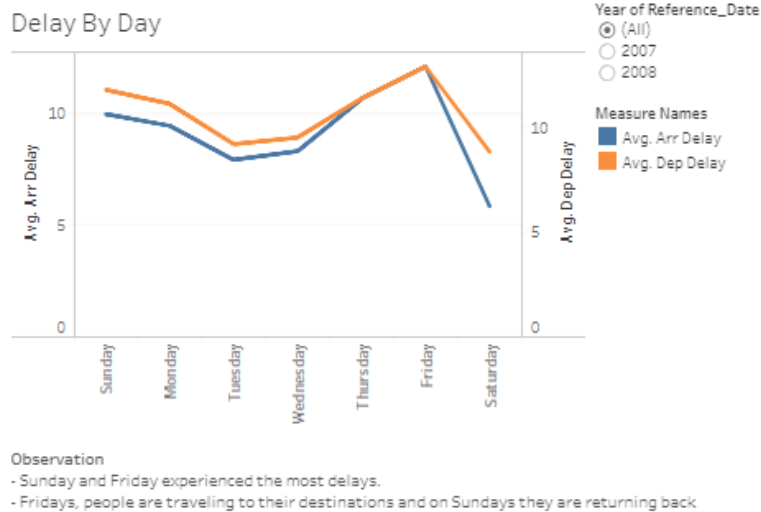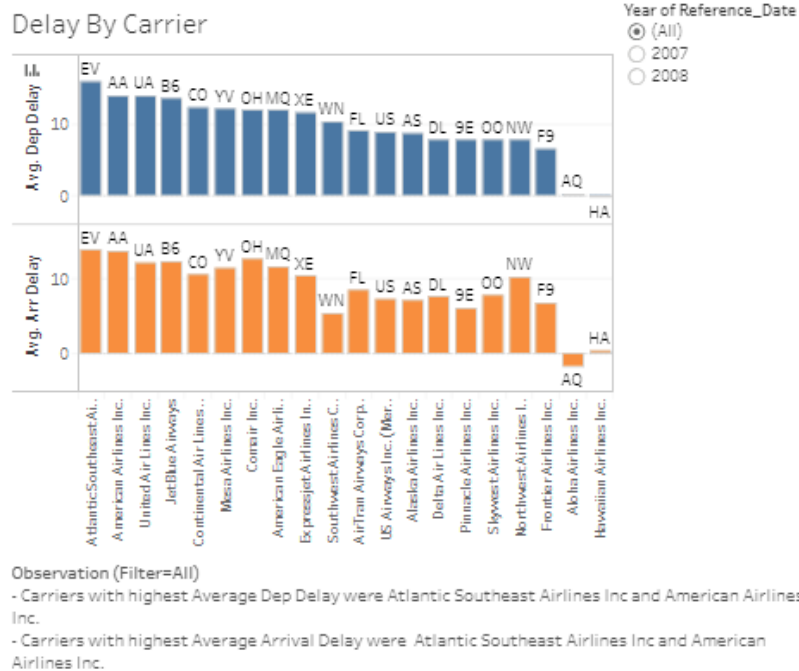
This slide shows that the average departure delay is slightly higher than the average arrival delay. I chose a line graph because it shows a clear difference between the two measures. Feedback – The user understood that the orange line was slightly higher than the blue line, but the user did not understand what the average arrival delay and average departure delay meant. So, I had to explain what the arrival and departure delays meant. Once the user understood the difference between delays, the graph really helped add perspective that one was slightly higher than the other. The user also pointed out that both delays follow the same pattern and I explained that the reason behind that was if carriers came late, they would also leave late as well.

Delay By Month

Measure Names
- ■ Avg. Arr Delay
- ■ Avg. Dep Delay

Observation
- February (Valentines Month), June (Summer) and December (Holiday Season) experienced most delays
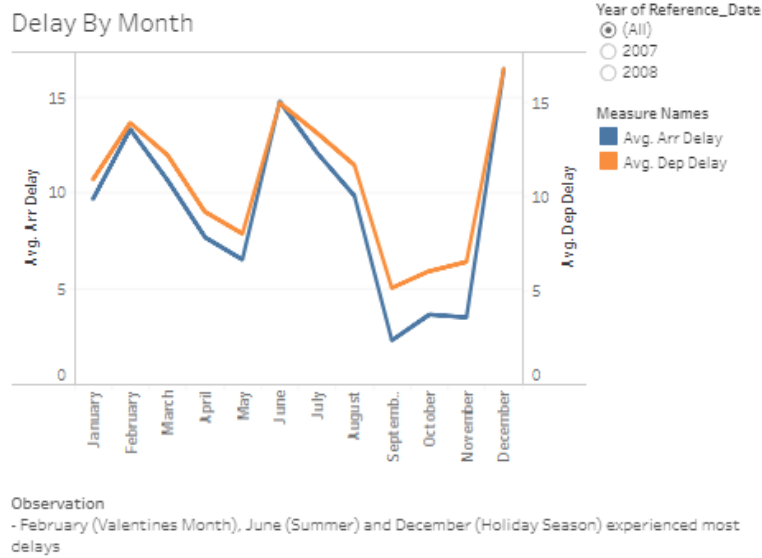
The above line graph shows delays by months. I wanted to see which months had the most delays. I chose a line graph because it is easier to identify trends. Feedback – the user was easily able to identify the peak delay months as February, June and December. The user added that June was the beginning of the Summer and December was the holiday season.

Delay By Day

Year of Reference_Date
- (All)
- 2007
- 2008

Measure Names
- Avg. Arr Delay
- Avg. Dep Delay

Observation
- Sunday and Friday experienced the most delays.
- Fridays, people are traveling to their destinations and on Sundays they are returning back
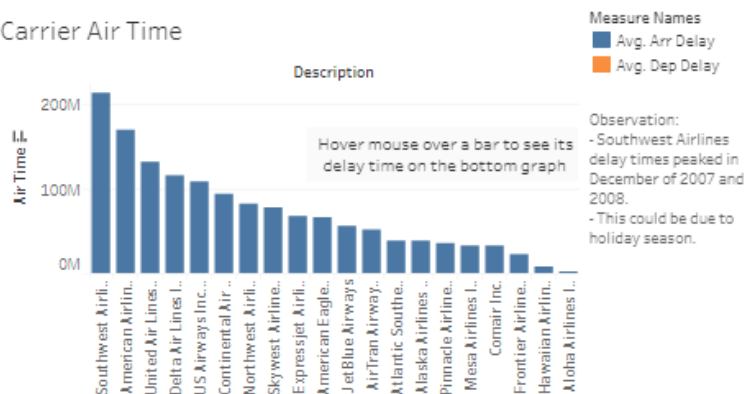
The above line graph shows the delays by days. I chose a line graph to identify any trends. Feedback – The user was clearly able to point out that more delays occurred on Sunday and Friday. The user also pointed that she users Southwest Airlines frequently and the carrier gives a discount for traveling on Tuesday's. This makes sense, delay is minimum on Tuesday and delay could also be directly related to number of people using an airline carrier.

Delay By Carrier

Year of Reference_Date
- ⦿ (All)
- ○ 2007
- ○ 2008

Observation (Filter=All)
- Carriers with highest Average Dep Delay were Atlantic Southeast Airlines Inc and American Airlines Inc.
- Carriers with highest Average Arrival Delay were Atlantic Southeast Airlines Inc and American Airlines Inc.
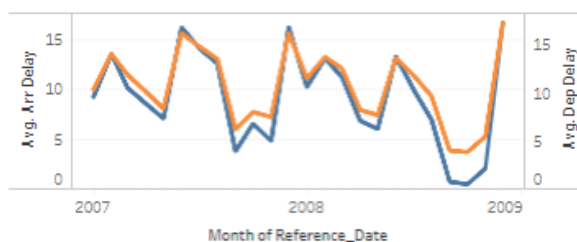
I used a bar graph to explore carriers with highest delays. I prefer bar graphs because datapoints can be ordered in a descending order to show order in data. Once disadvantage of having two bar graphs is that I had to order in one separately while explaining the graph to the user. Feedback – the user was clearly and easily able to identify the carrier with most delays.

Carrier Air Time

Measure Names
- Avg. Arr Delay
- Avg. Dep Delay

Description

Hover mouse over a bar to see its delay time on the bottom graph

Observation:
- Southwest Airlines delay times peaked in December of 2007 and 2008.
- This could be due to holiday season.

User Defined Unique Carrier Delay

I used a bar graph to explore which carriers spent the most time in the air and combined the carrier with a arrival and delay graph. This allows the user to inspect each of the carrier's arrival and delay times.

Feedback – I recreated this after the Udacity review that pointed out that it was difficult to understand the relationship between airtime and delays. Since, the bottom graph is dynamic and depends on choosing a carrier from the above bar graph, the user is able to gain a better understanding of delays. For example, Southwest Airlines experienced peak delays in Decembr of 2007 and 2008.

Step Three – Find a Story

The story can be found at
https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Story_v3/ExplorationofFlightsDataset-Story

The average departure delays are slightly higher than average arrival delays. Average delays were high in February, June and December. Average delays were high on Sundays and Fridays. Atlantic Southwest Airlines had the most arrival and departure delays followed by American Airlines. Southwest Airlines spent the most time in air and covered the most distance.

Step Four – Create your visualizations

Created visualizations as shown in Step 2 and hosted them on Tableau Public Site.

**Step Five – Get Feedback**

Feedbacks are included in Step 2.

Step Six – Document Feedback and Improve Visualization

Creating the story board was a little difficult because I had included text box in my slides for brief explanations. I had to adjust the text box multiple times in full screen mode, so it flowed properly with the graphs. There are multiple iterations of the story board on Tableau Public.

I had to recreate the dashboard below.
https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Dashboard_01_v1/Dashboard1
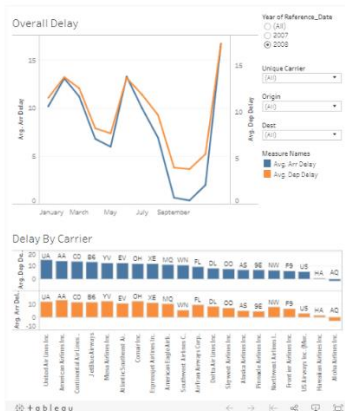

The user had given a feedback that the dashboard was 'too busy'. I had to ask myself what I was trying to convey. I chose to focus on only one idea which was the delay time by period (month/days) instead of focusing on delay times by period and carrier.

The refined dashboard is as follows:

Link:
https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Dashboard_01_v2/Dashboard2

Image:
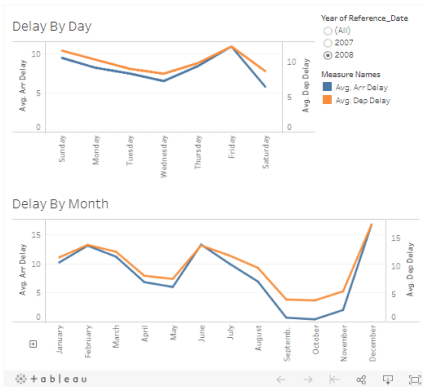
The new feedback user was that the user clearly understood what I was trying to convey.

The final version of the project is the storyboard at

https://public.tableau.com/profile/amit.kumar.shankar#!/vizhome/Udacity_Tableau_Project_Story_v4/ExplorationofFlightsDataset-Story-Final