Data Visualization Nanodegree:

Build a Data Visualization project - by Shashank Salian

Topic: Flight Delays and Cancellations

Links:

- **Visualization 1:**
  https://public.tableau.com/app/profile/shashank.s4018/viz/Airlines_delay/delay_dashboard?publish=yes

- **Visualization 2:**
  https://public.tableau.com/app/profile/shashank.s4018/viz/Airlines_delay/Delay_Reasons?publish=yes

- **Visualization 3:**
  https://public.tableau.com/app/profile/shashank.s4018/viz/Airlines_delay/Ratio_of_delayed_flights_per_Airline?publish=yes

Summary:

Calculated Fields:

Year_Month = DATEPARSE('yyyyM',[Year]+[Month])

Total_Delay =
[Departure Delay]+[Airline Delay]+[Air System Delay]+
[Arrival Delay]+[Security Delay]+[Late Aircraft Delay]
+[Weather Delay]

Total_Positive_Delay =
IIF([Total_Delay]>0,1,Null)

delay_ratio = COUNT([**Total_Positive_Delay**])/COUNT([flights])

+ve_Arrival_Delay =
IIF([Arrival Delay]>0,1,Null)

and similarly calculated +ve delay for other delay reasons as well

__Visualization 1__:

Dashboard with bar charts. The top bar chart shows Average delay time for various destination airports for a given Origin Airport. The bottom bar chart displays the number of delayed flights for each airline for a given Origin Airport (where total delay > 0 i.e. __Total_Positive_Delay__).

__Design Choice__:
Bar charts have been used over Pie charts as the slices of the pie chart may get thinner for eg: if an Airline contributes significantly less number of delayed flights compared to others. There are three filters: Origin airport (single select), destination airport (multiselect), Airline (multiselect).

__Insight__:
Among the flights originating from JFK, the flights to CVG have the highest average delay, however, there are just 12 flights flown from JFK to CFG.
Jet Blue Airways accounted for the highest amount of flight delays among all flight delays for flights originating from JFK airport.

Visualization 2:

Line chart showing how much each reason contributes to the delays over the course of the year.

__Design Choice:__
Here I converted the Year_Month to continuous to show how delay reasons move with months. And for a continuous variable, a line chart is a good option.

__Insight__:
Here we can see the delay numbers for all reasons tend to rise around June, again there is a rise around December. This could possibly be due to summer and winter holiday seasons respectively. Interestingly, there is also a rise in weather related delays around February.

Visualization 3:

Scatter plot showing total counts of flights vs delay ratio
(i.e. count total positive delays / total number of flights) per Airline.

Design Choice:

The Airline name has been color coded and available as a label on hovering. There is also a Month filter (multiselect) so we can select specific month/s or select all months to see overall plot on a yearly level. Here, I have used a gradient based color palette since the color blind palette has limited colors. And as an add-on information I have encoded the Count of total_positive_delay for each airline as a tooltip which can be seen while hovering each point on the scatterplot.

Insight:

Here we can see that Spirit Airlines has the highest delay ratio i.e. number of delayed flights for the total number of flights it operated in 2015. Southwest Airlines Co. operates the highest number of flights compared to any other airline but the ratio of delayed flights is far less for Southwest Airlines Co. when compared to Spirit airline.

Resources:

DATEPARSE() to create Year_Month from Year and Month
https://help.tableau.com/current/pro/desktop/en-us/data_dateparse.htm