

## Udacity Data Visualization Project – Airline flight information

Elad Shahar 2018 August

### **Links**

Link to the initial story

<https://public.tableau.com/profile/elad.shahar#!/vizhome/UdacityTableauProjectEladShaharV1/OriginalAirlinesstory>

Link to the updated story after feedback

<https://public.tableau.com/profile/elad.shahar#!/vizhome/UdacityProjectTableauEladShaharV2/Airlinesstory>

### **Summary**

This project I will deep dive into US domestic flight information 2008, dataset downloaded from the Bureau of transportation statistics.

I will analyze arrival delay across months, days, carriers and flight distance as well as cancellation rate.

Finally I provided a useful dashboard for investigating the best airline by performance for specific routes

### **Design and Feedback**

#### **Variable of interest**

The first decision I took After loading the data set was choosing the most relevant variable of interest.

After comparing average departure delay to average arrival delay I saw a very strong correlation between the two. I chose arrival delay to be my main focus as it is the delay that is most critical to travelers.

#### **Visual encodings and Legends**

Per chart

#### **Layout**

The layout for the dashboard is standard. With all worksheet data related to the same datasource so that when a filter is used, it affects all dashboard worksheets. This allows us to drill down and investigate different paths quickly

#### **Hierarchy**

I created a Hierarchy comprised of destination airport and the routes terminating in said destination airport. I created a calculated field named Route which concatenates origin and destination airports into one value.



## **Initial version**

### **Chart 1 – Destination airport by number of arriving flights**

This visualization shows the busiest destination airports. We can clearly see the busiest airports are Atlanta, Chicago and Dallas.

I chose treemap visualization with two visual encodings, the first by size for number of flights. The second for average arrival delay by color

For example we can see that Atlanta is the busiest airport with medium average delay while Chicago (ORD) is slightly less busy but with 50% higher average delay

Feedback A - “as the chart before it is hard to get a feeling for the data from a bar chart. Could the airports be shown on a geographic setting? Also not sure the function of color ”

Feedback B - “Could you show the full names of the airports or their state?”

### **Chart 2 – Average arrival delay by destination airport**

Here I chose average arrival delay by destination airport, showing the airports in a treemap sorted in descending order. We can clearly see average delay per airports and those airports with the longest average arrival delay.

I chose two visual encodings. The first average arrival delay by size showing the airports with highest average delay and the second, number of flights in color.

We can see that the airports with the largest delays are airports with relatively few domestic flights (except for EWR)

Feedback A - “It is clear which airports have longest arrival delay but it is hard to get a feeling for the data from the visualization. Could the airports be shown on a geographic setting?”

Feedback B - “Could you show the full names of the airports or the state they are in?”

### **Chart 3 - Average delay by carrier**

For this visualization I chose packed bubbles, the size of the bubble indicates size (higher average arrival delay) while the color indicates the carrier (airline). The chart shows several carriers with similar delay performance.

Feedback A – “This chart doesn’t provide me real value, many bubbles are close in size and all the colors are making it hard to focus”

Feedback B – “it is very hard to understand the message here, what I take from this chart is that there are many carriers with similarly length average arrival delay.

#### **Chart 4 - cancellation rate versus average arrival delay by carrier**

Following the previous average delay by carrier chart I decided to compare cancellation rate by carrier to average delay. For this reason I chose dual axis for these two values with bar chart for cancellation percentage and a line chart for the average delay.

#### **Chart 5 - Average delay by month**

I chose to show average arrival delay by month to observe seasonal effects.

I selected Unique carrier as color visual encoding with the relevant Legend

Feedback A – “There are too many colors, its confusing”

Feedback B – “The carrier information is not adding any value, if I just want to know which months have higher average delay”

#### **Chart 6 - Average monthly delay by delay reason**

This chart shows monthly average delay in minutes with a breakdown by delay reason.

As some months have much higher flight volume, I added average arrival delay to show performance by month. For example December has the highest average delay with over 16 minutes delay on average.

For this I created new calculated fields holding the delay in minutes (NA will be considered as 0) for each type of delay (Security delay, Carrier delay, NAS Delay, Weather delay, Late aircraft delay).

These values are measure values and visually encoded by color. Also added

Feedback A – “breakdown reason by month is not so interesting, how about by airline?”

Feedback B – “I don’t understand the relationship between the line and the bars. Also shouldn’t the breakdown by delay reasons reach exactly 100%?”

#### **Dashboard 1 - Average delay by month**

The dashboard includes:

1. average delay by destination airport (treemap)
2. total flights by destination airport (treemap)
3. Reasons for delay with average delay by month
4. Average delay versus cancellation rate by carrier

Worksheets are all related to the same datasource

Feedback A – “I don’t find this information useful,”

Feedback B – “the worksheets in the dash board should show different view of the same data but able to paint a clear bigger picture. I don’t feel this dashboard is doing that”

### **Chart 7 - Average delay by route by airline.**

I chose to show average arrival delay for each route (ORIGIN-DESTINATION) by carrier.

The visual encoding best fit here is color, showing each airline flying this route in color with the size of the bars the respective part of the total delay on the route.

Added detail encoding with number of flights.

Also added search field where the route can be typed manually.

It seems a useful tool for a traveler when deciding flight options, especially if timing is important due to connection flights or meetings.

Feedback A – “looks good and I can see the potential use of it ”

Feedback B – “can we see flight cancellations in this chart as well? ”

### **Final Version**

#### **Chart 1 – Destination airport by number of arriving flights**

Following the feedback received on the first chart I decided to change the view to a geographical representation. Showing the busiest airports on the map on the USA. I chose to use size as a visual encoder.

By using the map, it is clear in which parts of the US the largest flight hubs are located and their relative distance.

#### **Chart 2 – Destination airport by average arrival delay**

I chose to show this chart as a geographical map as well. Due to the size of the map the differences in size were not clearly apparent, so for emphasis I chose two visual encodings for the same variable, average arrival delay.

In addition I added number of flights as a detail. Now when hovering over an airport, it is clear if it is a small regional airport or a busy hub.

#### **Chart 3 (NEW) – average delay by distance group**

For the creation of this chart I created a calculated field by binning the distance variable.

In addition I added number of flights as a detail. Now when hovering over an distance bin, it is clear how many flights comprise the bin.

My intention is to investigate correlation between flight distance and average delay. Looking at the results of the bar chart there seems to be no clear correlation. Flights between 2900 and 3100 miles have the highest average delay, however it is important to remember that flights over 1500 miles are much less common compared to shorter flights.

#### **Chart 4 (initially 3) - Average delay by carrier**

Following the feedback I decided on a deeper exploration of average arrival delay by carrier. The first chart is changed from packed bubbles to barchart and shows average arrival delay by carrier with the number of flights as a dual axis line. This way the performance of the carriers is clear as well as their volume of flights.

#### **Chart 5 (initially 4) - cancellation rate versus average arrival delay by carrier**

This chart remains the same but added after chart 4, it is easier to understand.

#### **Chart 6 (Initially 6) – Which carrier has the highest average delay with breakdown of delay reasons**

Considering the feedback I decided to shift from a breakdown of delay reason by month, to a breakdown by carrier. It is also fitting more closely with the narrative of the story.

The chart shows clearly the average delay per airline with a breakdown of delay reasons

I chose dual axis for average arrival delay and measure values of the respective percent by delay reason.

In addition I added number of records to the top chart which shows again the volume of flights handled by the airline.

#### **Chart 7 (New) - Average delay by day of the week**

Considering the differences between months, it made sense to consider this criteria as well.

I chose a simple bar chart and it is clear that Friday and Sunday are the days with highest average delays.

I added number of flights as a detail and interestingly, flights are almost exactly the same among weekdays with the only notable reduction on saturday and partly on Sunday.

#### **Chart 8 (initially 5) - Average monthly delay**

This chart shows monthly average delay in a simple barchart removing the carrier color encoding which confused the viewer and did not add real value.

We can see December has the highest delays very clearly. By adding number of flights as a visual encoding detail. It is easy to evaluate number of flights per month.

#### **Dashboard 9 - Average delay by month**

Following the feedback I have tried to provide more useful and clearer information, the dashboard now includes:

1. average delay by destination airport (geographical map)

2. Average delay by weekday
3. Average delay by month
4. Average delay by carrier with breakdown of delay reason
5. Average delay versus cancellation rate by carrier

Again all worksheets are related to the same datasource

#### **Dashboard 10 (initially chart 7) - Average delay and cancellation rate by route by airline.**

Considering the usefulness of chart I decided to add to the information additional usage in the form of average cancellation rate on the route by airline.

When selecting a route, the data for both average arrival delay by carrier and average cancellation rate by carrier is presented. I thought of adding the cancellation rate as a detail but considering comparison between multiple airlines, it made more sense to add a separate chart below.

All worksheets are related the same datasource.

#### **Resources:**

Link to data

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