Data Analyst Nanodegree

P6 – MAKE EFFECTIVE DATA VISUALIZATION

James Gallagher | Data Analyst Nano Degree | March 15, 2017

Summary

This data visualization charts 5 different U.S domestic airlines' performance from 20 average of on-time arrivals. The data was collected from RITA.

Design

Exploratory Data Analysis and Cleaning

I downloaded the data from RITA, selecting a dataset that included all domestic flights from all carriers to and from major airports from June 2003 through May 2013. Exploratory data analysis was conducted using **RStudio,** and is detailed in *data/data.html*. While studying the data, I hypothesized that there might be trends in individual airline performance (# arrivals delayed / # total arrivals) over the 10+ year period. I decided that a line chart with multiple series would best show these different trends across different airlines. I first produced a cursory plot to explore the data:
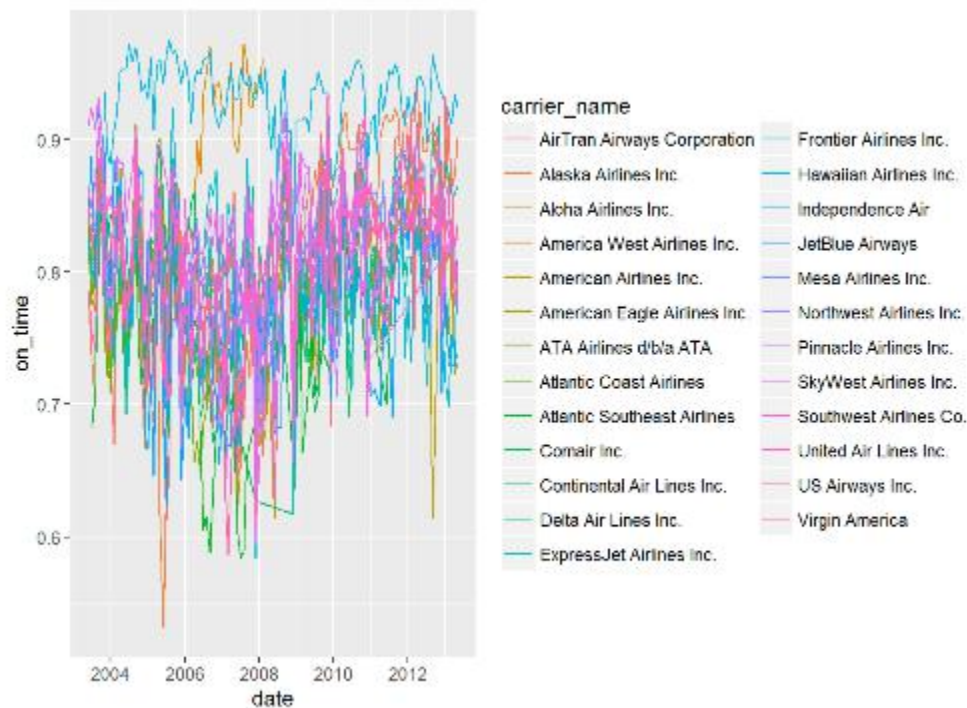
This was clearly too busy! There were 27 airlines, and the line chart was cluttered and ineffective in displaying any distinguishable trends. With the context of providing the reader with airlines that would be most relevant to him or her, I truncated that data to feature only the 5 most active airlines, i.e. the 5 airlines with the highest gross number of flights on a monthly basis. I generated two plots; a redo of the first line chart, and another chart showing total annual flights:
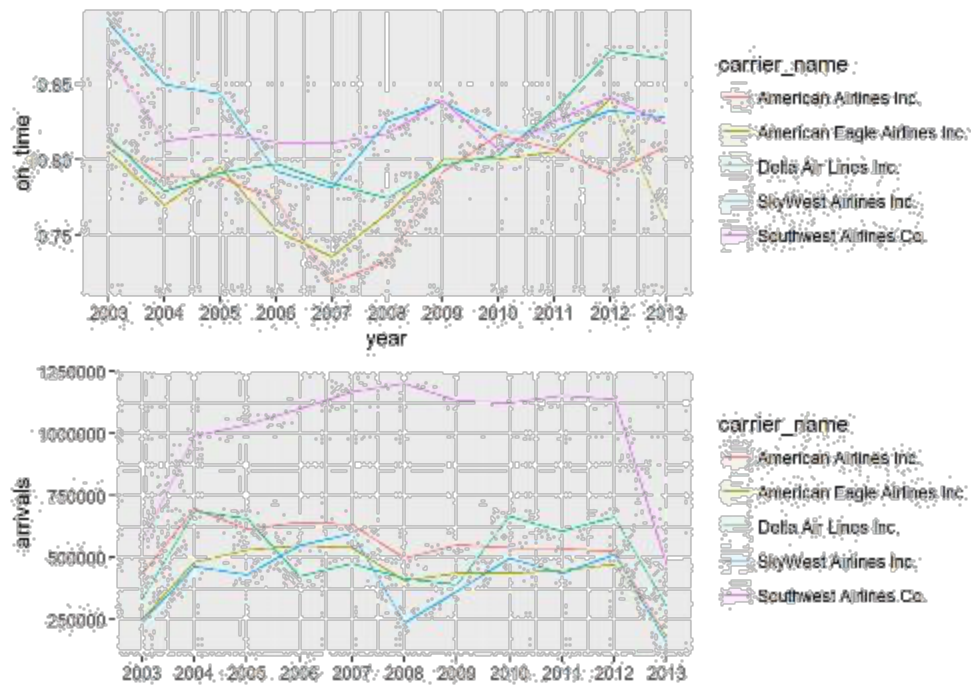
My initial evaluation of these charts were that they were satisfactory in visualizing the different trends of these 5 airlines. It shows how various airlines improved or worsened over time, and which airlines were currently performing the best, as of 2014. It also showed the general trends that all 5 airlines experienced: an aggregate dip in performance from 2006 to 2008, individual peaks from 2010 to 2012, and a more recent drop from 2012 to 2014.

I decided to display the performance trends rather than gross number of flights, as I was more interested in the question, "Which airlines are you more likely to be on time with?"
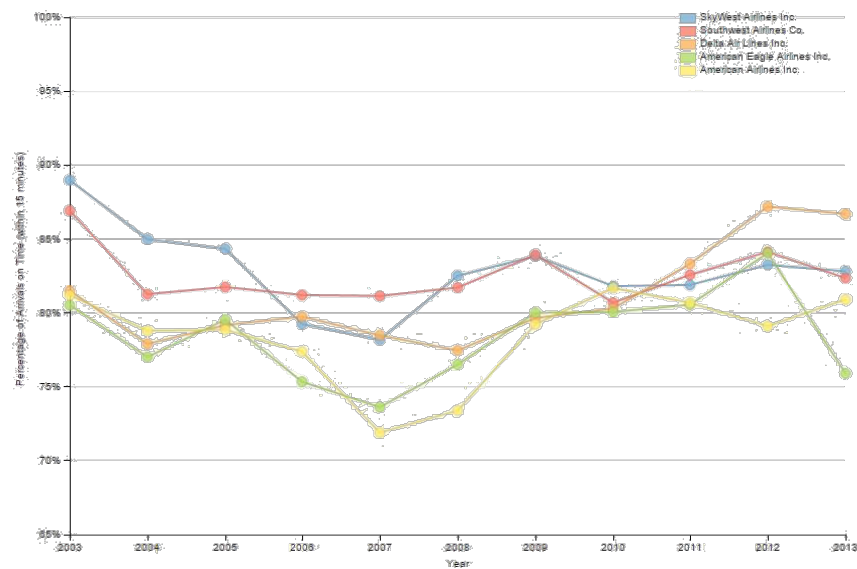
Data Visualization (dimple.js)

I decided to improve upon the initial data visualization as I implemented it with **D3.js** and **dimple.js** in the following ways:

- ✓ Fix scaling in the 'On Time Percentage' axis, particularly setting the maximum at 100% to show what the disparity from the maximum, perfect on-time value.
- ✓ Overlay scatter plot points to emphasize each airline's individual data points at each year.
- ✓ Add transparency, as some areas of the graph (2010-2012) have considerable overlap.

I re-evaluated the design decisions that I made during exploratory data analysis, and considered that a line chart was the best way to represent the trends of each airline over time. Coloring each line series was also a good way to visually encode and distinguish airlines from each other. With the adjustment of the y-axis, I also decided to move the legend to the top right, providing close proximity to the more relevant data points near 2014. I was concerned with 'Lie Factor' and truncating the y-axis minimum at a non-zero value; ultimately, I chose to truncate it just below the lowest 'valley', as I was concerned that a 0-100% scale would have crowded the lines and obfuscated any trends.

This initial iteration can be viewed at *index-initial.html*, or below:

Feedback

I interviewed 3 individuals in person, and asked for their feedback on the data visualization after prompting them with the background information and a small set of questions. Highlighted comments from them are listed below:

Interview #1

I immediately noticed the large dip in in 2007 that 3 of the showcased airlines experienced in performance. On second glance, I saw that the percentage ranged from 65% to 100%. I wonder if it would be more transparent by showing the full scale?

Having the right side of the chart open was nice, in addition to the points being spread apart; it was clear to see who the current "champ" is.

I think the main takeaway is that the competition is still "fair", in the sense that achieving the best performance is still a close match, and fair game.

Interview #2

One thing that I didn't notice was that this was an interactive graphic! Hovering over the points was interesting, but I kind of wanted to see some of the additional data, like how many flights each airline had.

I liked that there were lines. It made it easier to follow each "path" of points. I could see how this would have been a lot more confusing if it were just the circles.

Honestly, after I got a grasp of what the chart was, I looked all the way to the right, because I wanted to see how these airlines were performing today.

Even though the lines and points are colored differently, I think it would be really nice to highlight or emphasize individual airlines when you select them. It's a bit hard to follow a specific trend line.

I don't think that the title matches the y-axis, or vice-versa. It didn't immediately make sense to me what these metrics were.

*I think overall, it looks pretty clean and straightforward.*

Post-Feedback Design

Following the feedback from the 3 interviews, I implemented the following changes:

✓
    I added a mouseover event for the lines, so it would 'pop' it out and emphasize the path. This would allow for better understanding of each individual airline's trend from 2003 to 2014.

✓
    I changed the chart title to be more consistent with the data presented.

✓
    I subdued and muted the grid lines.

✓
    I polished the tooltip variable names to be more natural.

I chose not to include the arrival data and raw numbers. I didn't think that it was the focus of the chart, or had any impact on the understanding of airline on-time arrival rates.

Below is the final rendition of the data visualization:
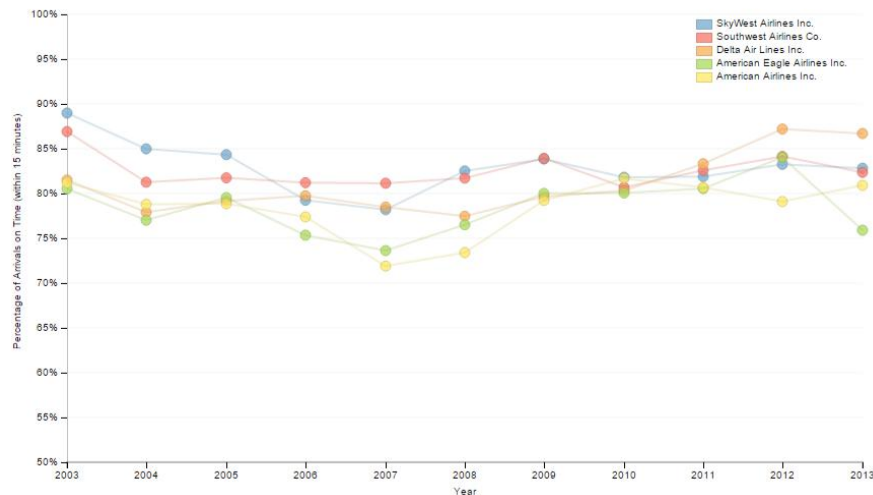
Findings

From the final rendition we can gather that:

Overall, there does not appear to be a clear trend of improvement or deterioration in all flight carriers combined. Almost all carriers tend to achieve a performance between 75-90% each year. Some tend to do well in a given year, while others lag behind.

On-Time Arrival Rates for Top U.S. Domestic Airlines, 2003-2013
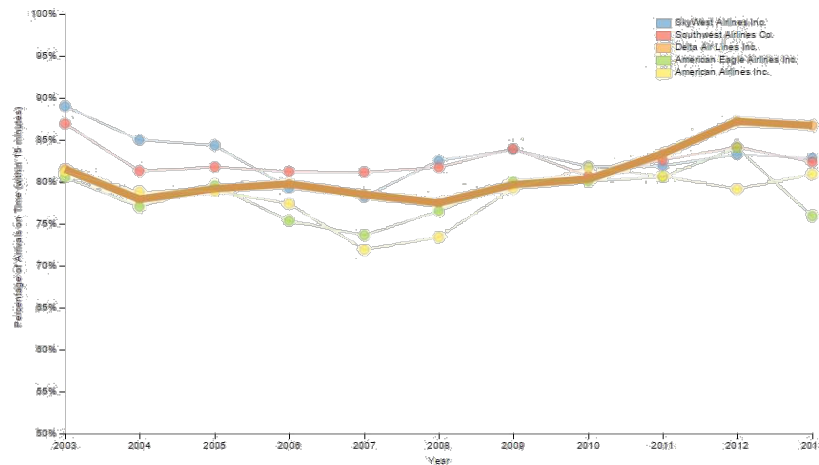
Observing the past trends, Delta Airlines has shown a great trend in improvement and seems to be following a steady upwards trajectory, but has not actually improved every year since 2008. Delta dipped one year and remained essentially the same in 2012 and 2013. But, on asking my interviewers the question –

Which airlines would you choose to get to some place on-time?

all of them agreed that they'd like to put their money on Delta Airlines.

Resources

i.   Airline On-Time Statistics And Delay Causes Data (RITA) – http://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1

ii. Dimple.js Examples - http://dimplejs.org/examples_index.html