Create a Tableau Story Final Project

Neil Seas

December 10, 2017

Contents

Abstract

This document contains the **Summary, Design, Feedback, and Resources** write-ups required by the Udacity Data Analyst Nanodegree as described in the Project Details

Chapter 1

Write-Ups

1.1 Summary

This project investigates cancellations in the year 2008 for airlines in the United States. The data set was combined with geospatial data from Data.gov to show patterns of cancellations on a map of the United States. Further exploration highlights specific airlines that are consistently poor performers in terms of cancellation rate.

Link to version 1
Link to version 2
Link to version 3

1.2 Design

My first design decision was to incorporate geospatial data in order to better visualize how cancellation rates vary across the United States. In the visualizations I used both Number of Flights and Cancellation Rate to reflect the scale of activity. I believe this choice facilitates quickly identifying both the most active airports and airports/airlines with relatively higher cancellation rates.

In the first visualization I chose to use a map to show the cancellation rate by the airport of origin. To make this visualization more clean, I chose to filter the data to only airports with at least 50,000 flights. I chose a diverging color palette because the majority of the cancellation rate range is quite narrow. The information in the tool tips is provided for airport specific detail, while the legends are provided to provide information about the meaning of bubble sizes and color.

For the second visualization I chose a box-plot because I wanted to show both the distribution of cancellation rates for the carriers, but also to highlight any outliers. I chose to also have the size of the dots indicate the size of the carrier (number of flights).

The third visualization provides similar information to the box-plot, but I felt that looking at the information in a scatter-plot really highlighted the higher cancellation rate airlines. I think it also makes clear that there is a general trend of larger airlines (by number of flights) having worse performance in terms of cancellation rate with the clear exception of Southwest.

For the last visualization I returned to using a map to demonstrate that the poor performers had high cancellation rates regardless of airport. When filtered to the six airlines identified in the prior visualizations the impact of seeing mostly red on the map would allow the reader to quickly draw the intended conclusion.

One change incorporated from the feedback I received was to remove information that was not contributing to the desired take-away from the visualization. In the case of the maps, this involved limiting the amount of information being shown (limiting the number of carriers). I also translated carrier codes into carrier names for readability.

1.3 Feedback

The feedback I received centered around making the visualizations more concise. On the maps I was asked to zoom in cluttered regions such as the northeast. Also, limiting the visualization to only carriers of interest. For the box-plot the reviewer pointed out that I was showing redundant information by including color to reflect cancellation rate, which was already clearly reflected by the y-axis. The reviewer also pointed out that carrier codes needed to be changed to actual carrier names.

1.4 Resources

- Data.gov.
- stat-computing.org
- Tableau help