

Reddit Data Viz Challenge - August

tidyversatile

8/18/2018

TSA Claims Data, 2007-2017

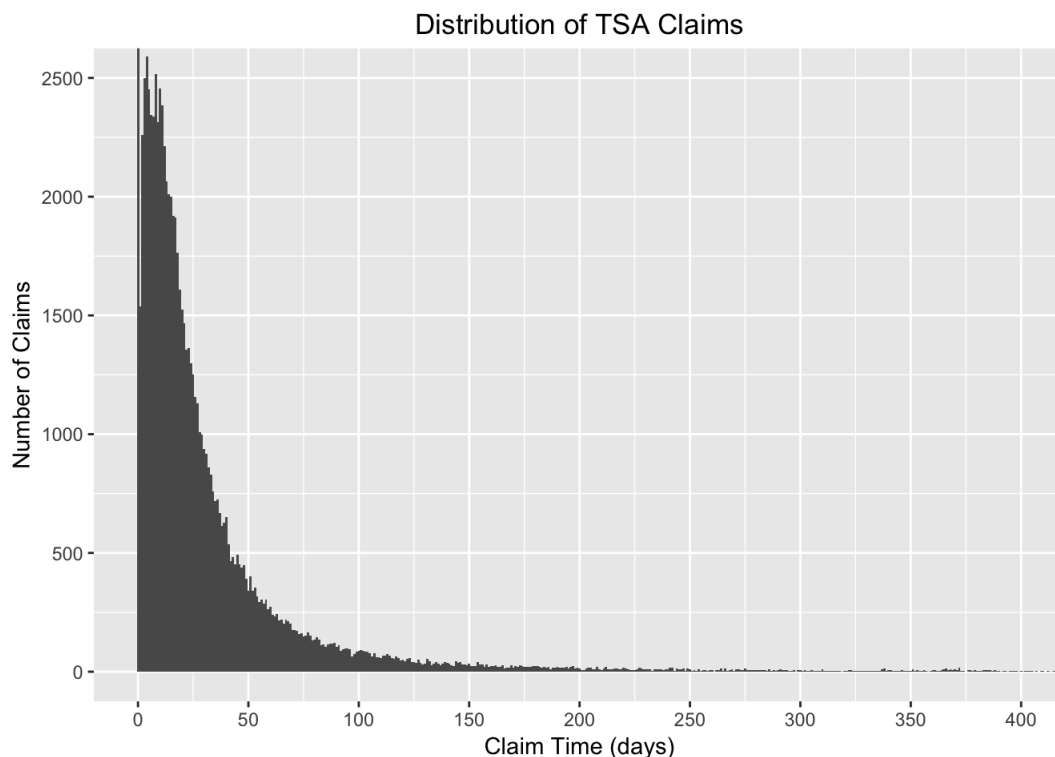
Objective

A few key questions guided my analysis:

1. Does waiting longer to file a TSA claim have an impact on its disposition?
2. Which items are most likely to result in a successful TSA claim?
3. At which airports do the most incidents occur?
4. At which airports do incidents take place that are most likely to result in an approved TSA claim?

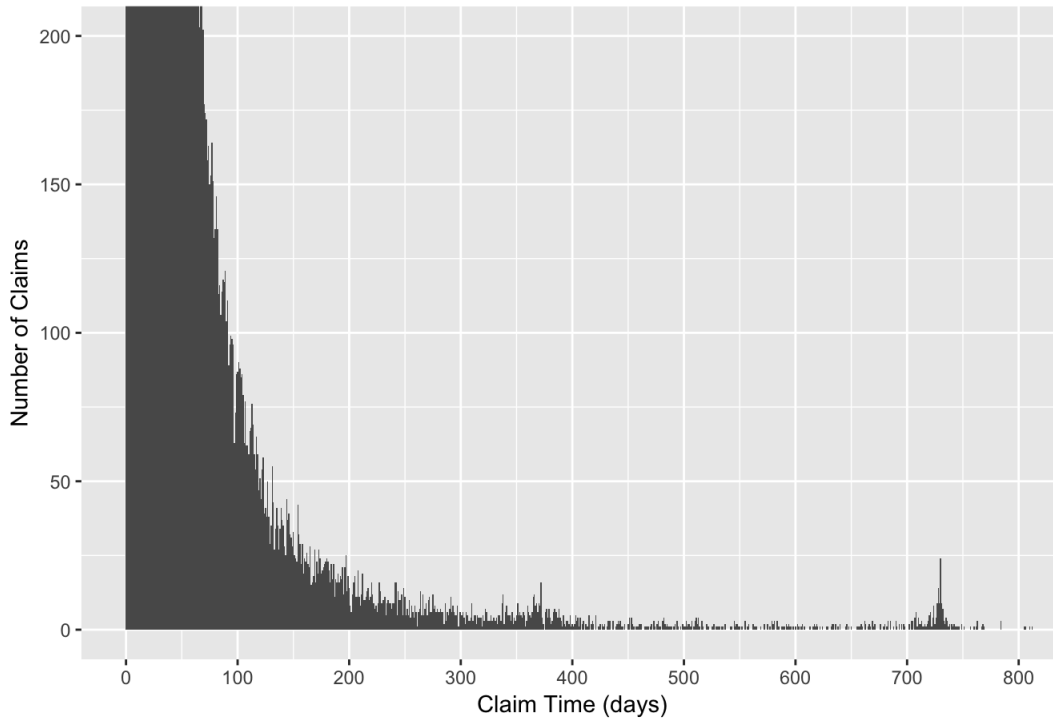
Does waiting longer to file a claim impact its disposition?

I first explored the distribution of reporting times and the proportion of dispositions for each reporting time. To do that, I calculated the difference between the Date Received and Incident Date variables, measured in days. Then, I filtered out cases with a negative Claim Time as they were likely data entry errors. For example, the TSA received a claim on August 2nd, 2007, for an incident reported to take place 4 months later, on December 7th, 2007. If this record were legitimate, then I would be more concerned about flying alongside psychics than with conducting a valid analysis. However, as I have yet to experience any clairvoyants in my air travels, I treated these records as errors, either by the person filling out the claim form, or by the person entering it into the TSA database.



The vast majority of claims were filed within 50 days of the incident. However, it was interesting to note that there is a tiny, but noticeable, uptick in the number of claims filed about two years after the date of the incident. There is a smaller but similar pattern around the one year mark:

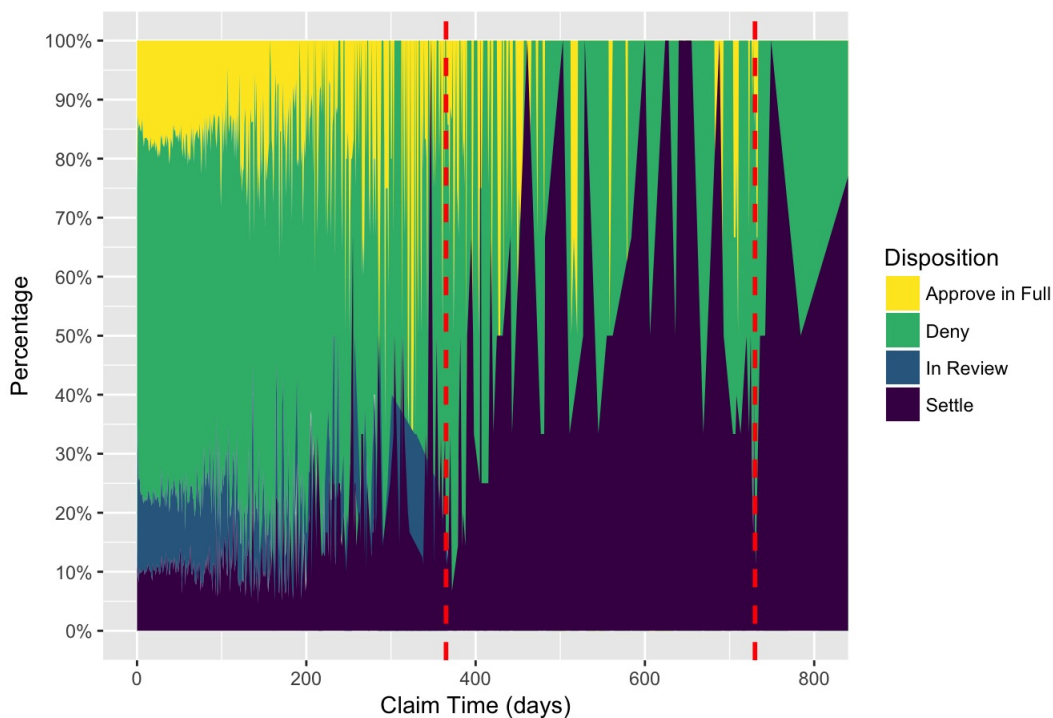
Distribution of TSA Claims



This could be the result of approximation by the person filing the claim. If they do not recall the exact date of the incident, perhaps they decide “oh, it was about two years ago”.

Next, I calculated and plotted the proportion of cases that were dispositioned as “Approve in Full”, “Settle”, “Deny”, and “In Review”. I selected these dispositions because they are consistent across every year, and make up 98% of all Dispositions. I inserted lines at the one-year and two-year mark for reference:

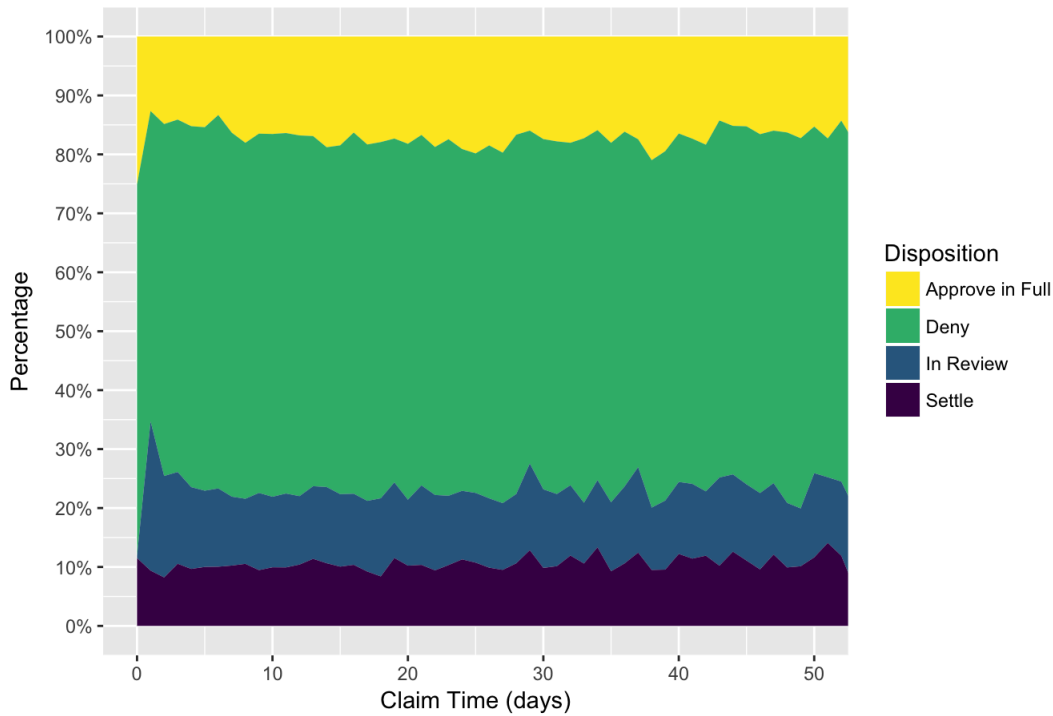
Proportion of Dispositions for Each Claim Time



The data suggest that claims which are filed more than 200 days after the reported incident have an extremely unpredictable chance of being approved in full. However, those patient enough to wait a full year to report an incident are much more likely to see their claim result in a settlement. It is also interesting to note that there are spikes in the denial percentage for claims that are reported around 1 year and 2 years after the incident date.

Since the vast majority of claims were reported within 50 days, I also wanted to focus on that subset:

Proportion of Dispositions for Each Claim Time

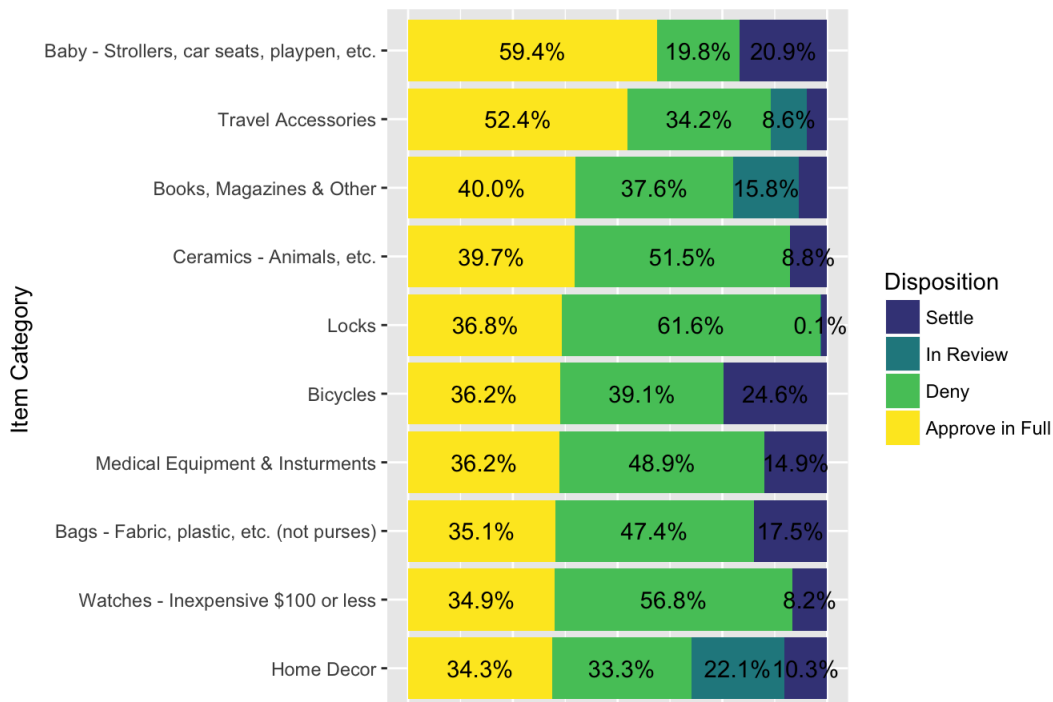


For the majority of cases, there doesn't seem to be any advantage to filing a claim sooner rather than later, or vice versa. Disposition percentages are stable.

Which Item Categories result in highest approval percentage?

Next, I explored Item Category data. Specifically, I wanted to know the top ten Item Categories with respect to approval percentage. I followed a similar approach to the first analysis, filtering out the irrelevant dispositions, grouping the data by Item Category and calculating the total number of records for each Item Category/Disposition combination, as well as their proportions. Then, I identified and plotted the ten categories with the highest approval percentages. I was disappointed to find that ageism clearly runs rampant within the TSA - claims for baby-related items have the highest chance of being approved

Item Category Dispositions

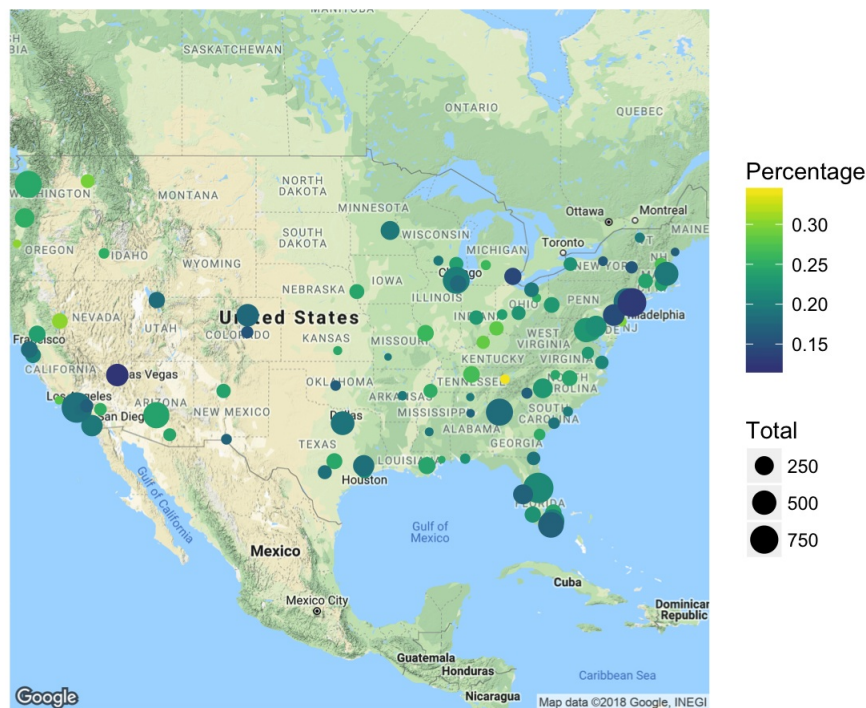


Which airports report the most claims per passenger? Which airports report the highest approval percentage?

Finally, I wanted to explore disposition percentages by airport. I selected the relevant dispositions and calculated disposition totals and percentages for each airport. Then I joined my data with a dataset containing airport coordinates, plotting the airports on a map of the US.

The size of each point represents the number of claims and the color represents the approval percentage. *Note: I searched for a long time for airport traffic data to correctly normalize the number of claims, since airports that have more people flying in and out are likely to have more TSA claims filed, but I could only find data on the top 40 US airports.*

TSA Claim Approval Percentage at US Airports



TSA claims for incidents that took place at many Midwest airports have relatively high approval percentages. This could mean that the actions of TSA agents at these airports justify more claims being approved. Perhaps the Midwestern stereotype of hospitality is unfounded. Approval percentages are low at most of the major hubs, and Las Vegas in particular caught my eye. TSA claims for incidents that took place at McCarran International Airport have one of the lowest chances of being approved in the country. Perhaps this is unsurprising - after all, the House always wins.