

Christopher Anderson
Professor Williams
February 21, 2021

DSC 680 Applied Data Science

Project 3 Check-In/Milestone 2

Any surprises from your domain from these data?

So far there are not too many surprises from the data that I am working with, other than the number of airline fatalities for the year 2017, which was extremely low at only 44 from 10 crashes. That is almost mind blowing when you stop to think about it! Certainly any number of fatalities are horrific, but the fact that only 44 lives were lost in airline crashes that year definitely stood out to me when looking over the data from years 2015 through 2020.

Is the dataset what you thought it was?

Going into this last project, I was expecting it to be much easier to find readily available downloadable datasets for the airline fatality information for years 2015 through 2020. And while to some extent that is true — as the years 2015 through 2019 weren't too difficult to attain from ASN (Aviation Safety Network) — I ended up doing a lot of my work manually — which seems to be a trend in my projects, unfortunately. The 2020 airline fatality data could have been scraped from the <https://aviation-safety.net/database/dblist.php?Year=2020> page, and automation could really make all of this a lot easier if there was a lot of data to scrape; however, with just a bit more than 300 rows,

it was easier and faster to copy and paste the data from the web page's table into a spreadsheet.

Have you had to adjust your approach or research questions?

Luckily, I have not had to adjust my research questions nor my approach, and this is most likely because this project is an extension or addition to one that we took on in a previous course, DSC 640 Data Presentation and Visualization. I wanted to focus in on the presentation and visualization elements for this final project, and I was very curious about how the fatality numbers looked for the most recent years after the 1985 through 2014 timeframe that were part of the 640 work.

Is your method working?

Yes, my method is working with no issues, and I plan to present updated visuals and graphics representing these additional years (2015 - 2020) of this extended study. Will the trend of decreasing airline fatalities continue for this five year stretch compared to the previous years, or will there be some of an upward trend? It will be most interesting to see things charted out.

What challenges are you having?

One of the biggest challenges I had was one that I touched on earlier, and that was getting compiled airline fatality and crash data for the year 2020 from Aviation Safety Network (ASN), the primary resource used for the project data. I am convinced that is only because that data is compiled and presented in the format that I was used to seeing for the other years sometime later in the year, and it just hadn't been done yet this early in 2021 for the 2020 year. As discussed, I copied and pasted the table of individual crashes from 2020 from their website into my own spreadsheet and was able to do the summary and count functions easily.