

## Data Source and data sets

WorldBank Data:

<https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2020&start=1970&view=chart>

Raw data file name:

1. API\_IS.AIR.PSGR\_DS2\_en\_csv\_v2\_4499051.csv

Power BI Query Name:

1. WorldBankAirTransportPassengers

Aviation Safety Network Data:

<https://www.geckoboard.com/blog/6-data-visualization-techniques-to-display-your-key-metrics/>

Raw data file name:

1. airline-safety.csv
2. table\_02\_14\_102521.xlsx
3. table\_02\_09\_102621.xlsx

Power BI Query Name:

1. airline-safety
2. US General Aviation Safety Data
3. US Air Carrier Safety Data

## Data transformations

airline-safety dataset did not need much transformation. It contains data for 56 international airlines. These airlines were in the global top 100 list as of 2012. Data shows incidents, fatal accidents and fatalities spanning 30 years starting from 1985 ending 2014. Data for each category is divided to 1985-1999 and 2000 to 2014. We added a new column to the table. The new Column, "Ranking" was calculated based on the value of available seat per week. Those with more than 2 million were ranked top, those between 1 and 2 million were ranked Middle and those below 1 million were ranked bottom.

US General Aviation Safety Data was transposed, filtered and transformed into a 10 column table. It has the fatalities and incident counts for all US airlines from 1970 to 2020.

WorldBankAirTransportPassengers dataset contained data for number of passengers traveled worldwide from 1960 to 2020. The data set has this data for every country in the world as well as a total for the world. We transformed the table to two columns-Year and Passenger traveled worldwide. We used this transformed table to create a metric for the presentation. The details of the metrics are described below.

US Air Carrier Safety Data dataset was reduced 9 columns. Column “Nonfatal accidents” was computed by subtracting Fatal accidents column from Total accident column.

## Data Presentation Plan

We placed a picture of an air disaster as an attention grabber. The question below the picture is aimed to persuade the reader to read on to get the answer to the question posed.

We chose the font and the font color to make the words more readable and give a sense of calm because the topic could be disturbing-specially with the picture on top. This makes the beginning on the story dramatic and attracts curiosity.

The 1<sup>st</sup> paragraph is aimed to give an overview of-something the reader already knows and that we can relate it to the story. The 2<sup>nd</sup> paragraph gives some examples and leads into why we are making this presentation.

The graphs and their description tell the story. Each graph will try to prove airline travel is safe and the conclusion will be to assure them that the data we presented should be trusted.

## Data Presentation

### Choice of visualization

We used a bar chart to show the airline rankings to show airlines side-by-side and also distinguish between top, medium, bottom airlines. The two bubble charts that follow, show the fatalities and incidents during 1985-1999 and 2000-214 periods. We highlighted the airlines with high number of fatalities and incidents showing by size that the ranking and number of incidents/fatalities are not correlated. The color scheme to show the rank of airline was maintained throughout so as to not distract the viewer.

The stacked line chart shows 5 metrics showing fatalities and number of fatalities far below total accidents. We used the dark and light blue colors and variations of red to show the other metrics to make the contrast easier to decipherable.

The simple line chart showing the trend in traveling worldwide getting interrupted only by the COVID pandemic. We removed the x and y axis labels and lines because all the data that is needed in on the graph.

## Overall Findings

Overall, we did not see a relevance between fatalities and accidents with airline travel and the public’s opinion of air travel. It would be worth studying other travel options such as bus, rail, ship, personal and vehicle. It would also be worth looking at the trend f each option with respect to each other counting geopolitical, and other environmental influences.

## Ethical considerations

The number of available seats in the airline-safety data set does reflects the ranking of the airline based only on how large an airline they are as opposed to how good they are. Some airlines are conglomerates

(i.e. delta and united), and some are national airlines (i.e. China Air). Other than ranking, the relationship between number of fatalities for larger airlines and smaller airlines can be skewed due to the number of flights these airlines have-the higher the number of flights, the more possibilities of accidents.