

**Project Title:**

Airplane crash data from 1908 to 2009

**Team Members:**

Schehrbano Khan, Prem Ganesh Elango, Suvrangshu Ghosh, Amitava Samaddar

**Project Description/Outline:**

Based on two data sets our team found, we are trying to understand the air safety trend over the years. Our data set includes air crashes from 1913 to 2009 and another data set with information about the total number of flights per year from 1970 to 2017.

**Research Question to Answer:**

1. What trends exist for flight safety and fatalities based on operator and time?
2. What percentage of total flights that operated crashed?
3. What was the survival rate of the crashes?
4. What were the most and least accident-prone carriers in the time frame?
5. What type of manufactures (Boeing/Lockheed) has the most number of crashes based on the data we have available?
6. Are there any differences between different types of planes i.e. commercial vs private vs cargo etc?

**Data Sets to be used:**

-Crash data set: <https://opendata.socrata.com/Government/Airplane-Crashes-and-Fatalities-Since-1908/q2te-8cvq>

-Total flight data set: <https://data.worldbank.org/indicator/IS.AIR.DPRT?end=2016&start=1970&view=chart>

**Rough breakdown of Tasks:**

1. Each important column assigned to each member of the team to plot with other columns and gain familiarity with the data set.
2. Suv spearheaded the data cleanup providing us with a clean data set requiring minimal further cleanup to report important data visually
3. Schehrbano will generate graphs and perform data cleanup pertaining to column fatalities. With input from the rest of the team, she will also handle the general structure of the power point to create a flow of the data set
4. Prem will code some key visualizations to understand the data set and being one of the more experienced coders amongst us, help us write more concise data
5. Amitava will generate important visualizations about categories of airlines vs crashes (passenger carriers, military, mail)