**Exploratory Analysis of Aircraft Crashes between 1908 and 2008**Data Visualization

Final Project

**By**

Mikhail Terentev

Marina Furman

Wei Chen

Julia Celley

**Project Objectives**

Using aircraft crash data from between 1908-2008, we will look for patterns in crashes and linked fatalities to try and ascertain the safest methods or guidelines for air travel.

**Motivations**

Aircraft today play a large role in our society, benefiting people in many ways:

* Fast national and international travel
* Global cargo transportation
* Faster mail delivery
* Military transportation
* Conducive to greater globalization of industry

Throughout the history of aircraft development speed safety and reliability have all made massive advances. With all the aviation advances of the past century accidents and fatalities are still common enough to cause concern for some travelers. We seek to explore crash data to find what aircraft or operating organizations may be of more or less concern.

**Dataset**

Airplane Crashes and Fatalities Since 1908  
<https://www.kaggle.com/cgurkan/airplane-crash-data-since-1908>

**Column Descriptions:**

* **Date** - Date of accident, in the format - January 01, 2001
* **Time -** Local time, in 24 hr. format unless otherwise specified
* **Airline**/Op - Airline or operator of the aircraft
* **Flight** # - Flight number assigned by the aircraft operator
* **Route** - Complete or partial route flown prior to the accident
* **AC Type -** Aircraft type
* **Reg** - ICAO registration of the aircraft
* **cn / ln** - Construction or serial number / Line or fuselage number
* **Aboard** - Total aboard (passengers / crew)
* **Fatalities** - Total fatalities aboard (passengers / crew)
* **Ground** - Total killed on the ground
* **Summary** - Brief description of the accident and cause if known

The Aviation Crash Database records hundreds of details surrounding more than 68,000 accidents. This database is open and accessible to the public, ripe for information mining to answer a multitude of questions, and search for insights that could lead to improvements in aviation safety. The Aviation Crash Database is an excellent candidate for exploration using Information Visualization techniques.

**What data did we get?**

There is a total of 5,268 crashes reported in this table, from 1908 to 2009.

For each accident recorded, we know:

* date, time and location.
* the operator, flight number, route and type of airplane.
* the number of people that were aboard, the number of fatalities among them, and the number of fatalities on the ground.
* a brief summary of the accident.

**Data Cleaning**

We have performed some data cleaning.

The data set has 22 missing values in Abroad, 12 in Fatalities and 22 in Ground. Now, since the dataset has a total of 5268 records these 56 is only about 0.9% percent of the data. So these Null values are omitted.

To best visualize the data the summary section of the dataset was altered in excel to separate words to columns, and then common/filler words were scrubbed manually.

**Tools and Analyses Used**

* Performed **Exploratory Analysis**
* Used **Tableau** to visualize our findings for presentation
* Presentation of comparison of crashes and outcomes are around:
  + Aircraft Type
  + Airlines/ other operating organizations
  + Accidents over time
  + Cause of crashes

**Data Visualization**

**Crash Trend Analysis by year and month**

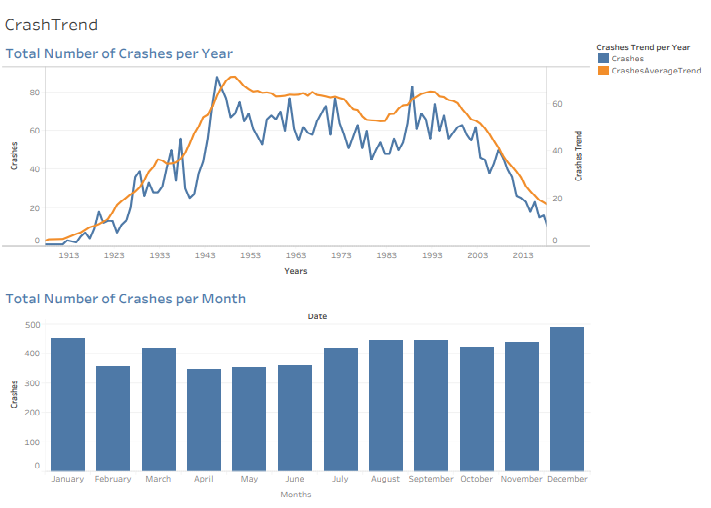
* To analyze if amount of crashes depends on month
  + was created TotalNumberOfCrashesPerMonth worksheet in Tableau
  + for demonstration was used bar chart
  + to represent **columns** was use Month from date of crashes

MONTH(Date)

* + to represent **rows** was created new calculated field to show how many crashes were happened per month

COUNT(MONTH([Date]))

* To analyze amount of crashes per year
  + was created TotalNumberOfCrashesPerYear worksheet in Tableau
  + for demonstration was used combination of two charts line and smoothed line
  + line shows real data calculated as COUNT(YEAR([Date]))
  + smoothed line is Moving Average for COUNT(YEAR([Date]))
* These worksheets were combined into dashboard CrashTrend were we can compare and analyze the results



:

In the first visualization we see aircraft crash trends over the whole history of manned flight. The greatest number of crashes occurred between 1943 and 1990, which follows the growth of commercial flight, and lowers as safety measures improve over time, leading to the modern era of relatively safe air travel.

As we can see from the second visualization the number of crashes month to month do not vary enough to imply a trend.

**Crashes by Operator**

This is yet another interesting question to be asked which airline company has seen lot of fatalities. We will create a Operator causing Aircraft diagram for this analysis and filter the 30 Top Aircraft Operators which has seen more than 500 fatalities in total.

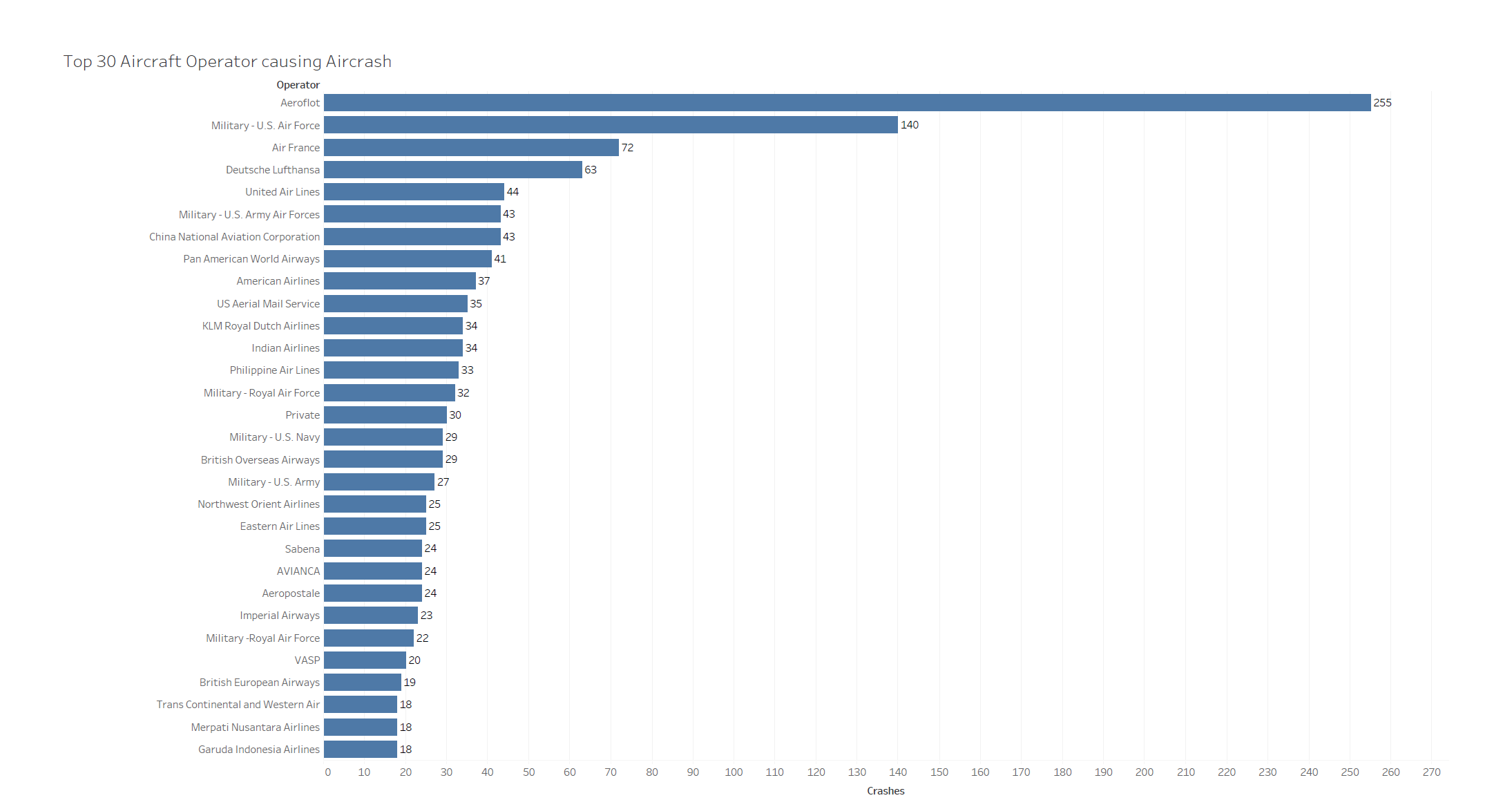
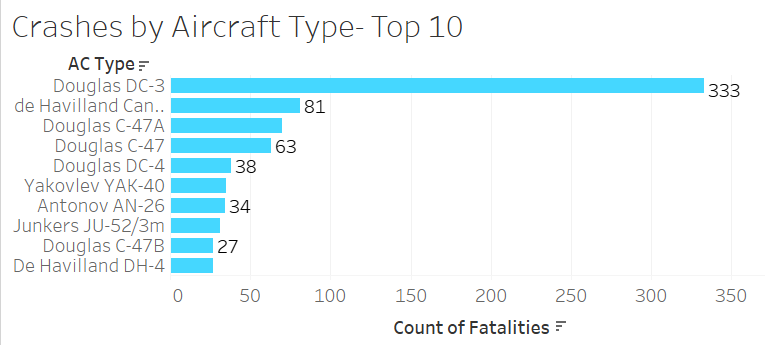


Figure 1. Air crash flight operators

A majority of the flight operators are Aeroflot, US Air Force, , Air France and Lufthansa, as seen from figure above. As can be seen this analysis falls in line with the analysis done in step above. United Airlines and US Air Force both operate of USA which has seen the most number of fatalities. It interesting to see the cause of fatalities happening in Russia, our guess is it will be mostly due to an route issues.

**Crashes by Aircraft Type**

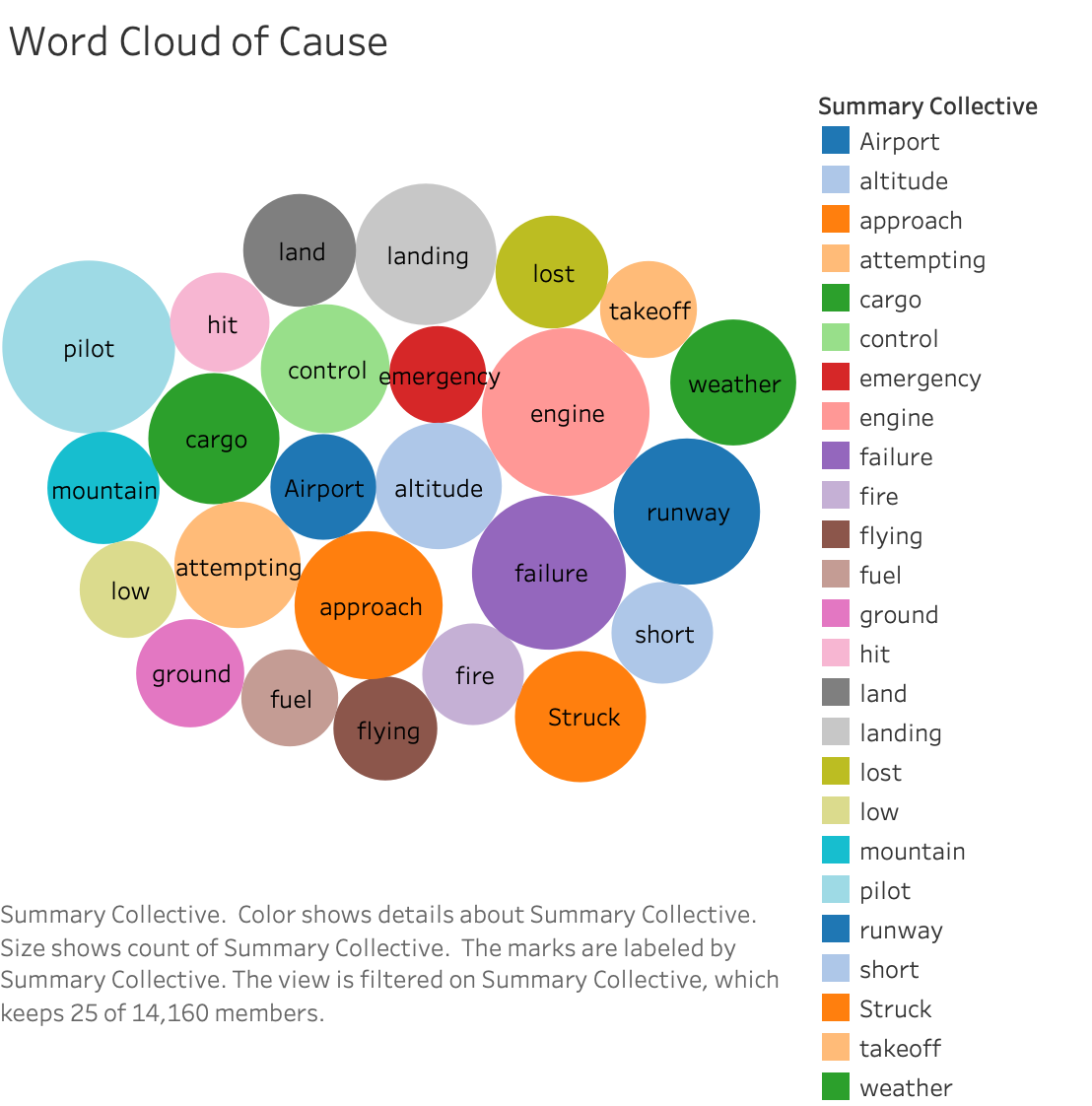
The following visualization shows that Douglas DC-3 had the highest number of crashes. We trend to believe this type of aircraft is less safe than other, however, we can’t conclude that as the dataset does not provide the information about total number of flights using this type of aircraft during the same time period.



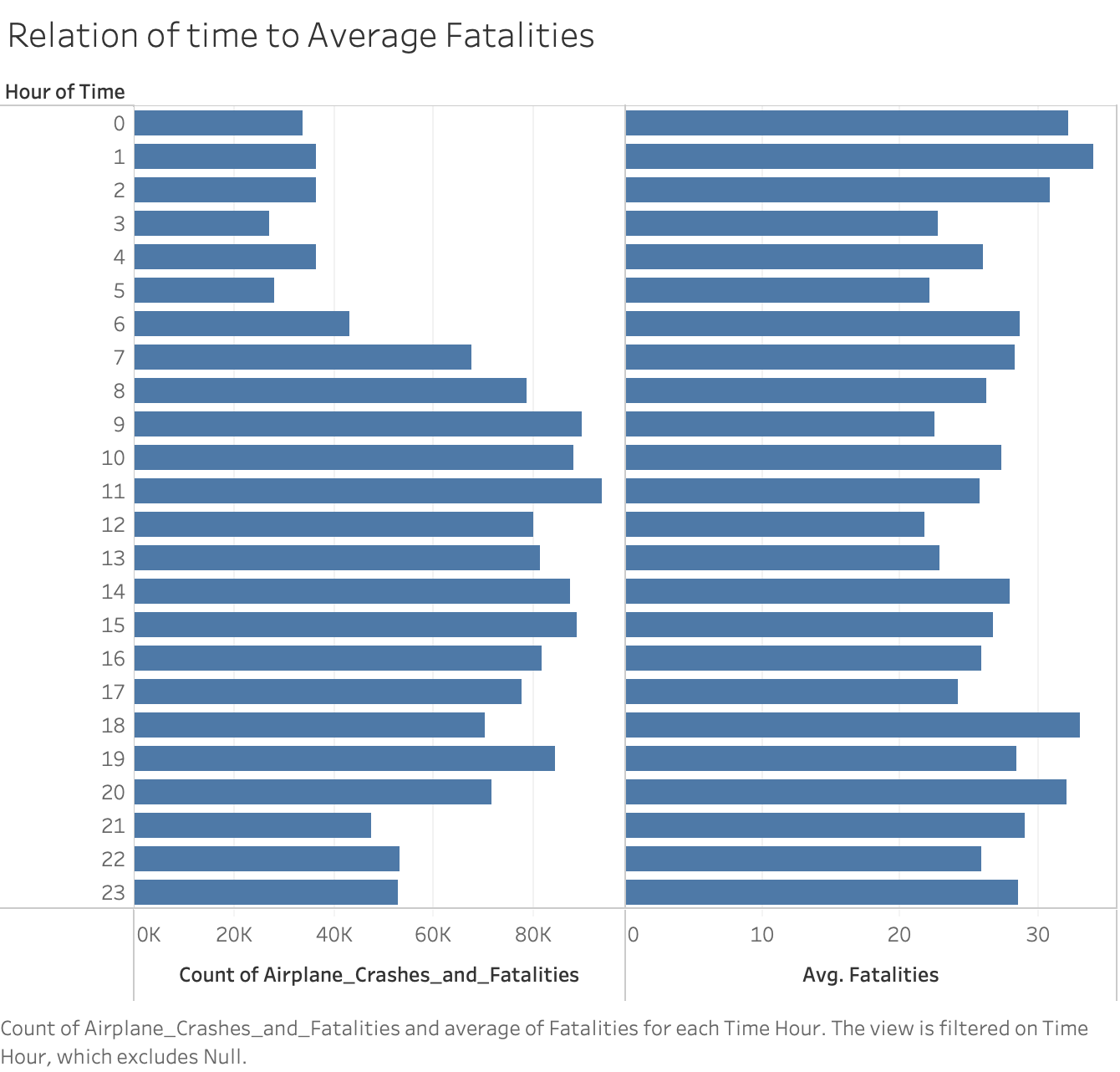
**Fun fact about DC-3: (WIKI)**

The Douglas DC-3 is a propeller-driven airliner which had a lasting effect on the airline industry in the 1930s/1940s and World War II. It was developed as a larger, improved 14-bed sleeper version of the Douglas DC-2.

Perhaps unique among prewar aircraft, the DC-3 continues to fly in active commercial and military service as of mid 2018, more than eighty years after the type's first flight in 1935. There are still small operators with DC-3s in revenue service and as cargo aircraft.

****

The Word Cloud of Cause captures the most common word from the summarized reports included in the airplane-crash dataset.



With the Relation of Time to Average Fatalities we can see that while crashes are rarer in “red-eye” hour flights, the average fatalities are higher in comparison to the corresponding number of crashes, however the data alone does not indicate why.

**Summary**

Overall, we found that flight crashes increased most between 1943-1990. While crash data alone does not explain this trend, this is inline with the increasing availability and popularity of air travel. Over the past 30 years the instances of aircraft accidents have decreased drastically so there isn’t much to be concerned about as a commercial traveler. However, if you are concerned about crashes, it’s easier to find survivors in daylight than darkness which may explain the higher levels of fatalities in night flights over day flights.

As for aircraft reliability; Aeroflot and Douglas Dc-3 type have faced the most air crashes. In this study, we have analyzed the more than 100 years data on air craft crashes. We have shown in our detailed analysis that while there are many factors that can contribute to aircraft failure and passenger/crew fatalities, that overall reliability, especially of commercial aircraft, is currently very high.