Short Analysis of Airplane Crashes

This article is a visual analysis of airplane crashes between 1969 and 2009. data.world keeps a detailed database about the airplane crashes and gives the opportunity to make an in-depth analysis for anyone interested in the subject. Although the data started from 1908, I preferred to analyze the modern era of flight and decided to take the above-mentioned time domain. There will be no zeppelins to analyze, but the results will be closer to reflect the effectiveness of the modern-day aerospace safety standards.

Wikipedia's definition for aviation safety is: "Aviation safety means the state of an aviation system or organization in which risks associated with aviation activities, related to, or in direct support of the operation of aircraft, are reduced and controlled to an acceptable level."

In order to increase the importance of the aviation safety, I tried to concentrate on the following issues for the selected time frame:

- How many planes crashed?

- How many people were on board?

- How many people survived?

- How many people died?

- Crash numbers grouped by the make/model of aircraft (separately).

Visual exploratory analysis of the results using Pandas and Seaborn give a better picture about the incidents.

Data Cleaning

The first step is to import and clean the data using pandas before exploring the cause and effect relationship in the air disasters - which is not an easy task.

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
air = pd.read_csv('Airplane_Crashes_and_Fatalities_Since_1908.csv')
```

```python
air.tail(3)
```

	Date	Time	Location	Operator	Flight #	Route	Type	Registration	cn/In	Aboard	Fatalities	Ground	Summary
5265	06/01/2009	00:15	AtlantiOcean, 570 miles northeast of Natal, Br...	Air France	447	Rio de Janeiro - Paris	Airbus A330-203	F-GZCP	660	228.0	228.0	0.0	The Airbus went missing over the AtlantiOcean ...
5266	06/07/2009	08:30	Near Port Hope Simpson, Newfoundland, Canada	Strait Air	NaN	Lourdes de BlanSablon - Port Hope Simpson	Britten-Norman BN-2A-27 Islander	C-FJJR	424	1.0	1.0	0.0	The air ambulance crashed into hills while att...
5267	06/08/2009	NaN	State of Arunachal Pradesh, India	Military - Indian Air Force	NaN	Mechuka for Jorhat	Antonov An-32	NaN	NaN	13.0	13.0	0.0	The military transport went missing while en r...

```python
air.shape
```
```
(5268, 13)
```

As shown in the above figure, there are 5268 rows and 13 columns (attributes) of data. Some of the rows define accidents before 1969 and also we will have to get rid of some rows, which have NaNs.

```python
air.isnull().sum()
```
```
Date             0
Time          2219
Location        20
Operator        18
Flight #      4199
Route         1706
Type            27
Registration   335
cn/In         1228
Aboard          22
Fatalities      12
Ground          22
Summary        390
dtype: int64
```

The results for the NaN value count of all the columns tell us to delete some columns. Also, for the current analysis, some columns will not be needed, and they will be deleted.

```
drop_set = ['Flight #', 'Route', 'Registration', 'cn/In']
air.drop(drop_set, axis = 'columns', inplace = True)
```

```
air.shape
```

```
(5268, 9)
```

Next step will be to make a new column, by extracting the year value from the date column.

```
air['year'] = air.Date.str.extract(r'([0-9][0-9][0-9][0-9])', expand=True)
```

```
air.tail(3)
```

	Date	Time	Location	Operator	Type	Aboard	Fatalities	Ground	Summary	year
5265	06/01/2009	00:15	AtlantiOcean, 570 miles northeast of Natal, Br...	Air France	Airbus A330-203	228.0	228.0	0.0	The Airbus went missing over the AtlantiOcean ...	2009
5266	06/07/2009	08:30	Near Port Hope Simpson, Newfoundland, Canada	Strait Air	Britten-Norman BN-2A-27 Islander	1.0	1.0	0.0	The air ambulance crashed into hills while att...	2009
5267	06/08/2009	NaN	State of Arunachal Pradesh, India	Military - Indian Air Force	Antonov An-32	13.0	13.0	0.0	The military transport went missing while en r...	2009

How Many Planes Crashed?

First, dtype of the new column "year" is converted to integer, then rows containing the data before 1969 were deleted, and we are ready to proceed to the analysis with 2994 rows of data about the air accidents. Much less from the starting value, but there will be no zeppelins, and the results will be closer to reflect the effectiveness of the modern-day aerospace safety standards.

```
air['year'] = air['year'].astype(str).astype(int)
```

```
air = air[air['year'] >= 1969]
```

```
air.shape
```

```
(2994, 10)
```

The figure says that, globally, 2994 airplane crashed between the years 1969 and 2009.

The graph below shows the total number of accidents for the same time period. 1972 was an outlier with 104 accidents and there is data for half of 2009 with 30 accidents, still the trend for the number of accidents is downwards.

```
air.groupby('year')['Date'].count().plot(figsize = (9, 6));
```
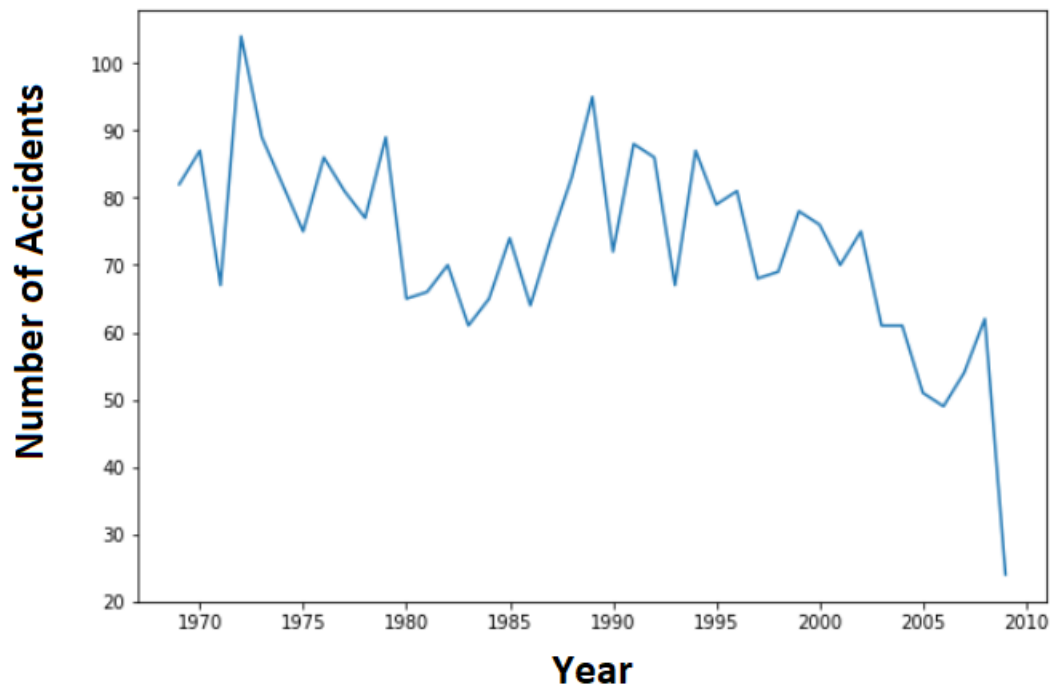

How Many People Were on Board?

Between 1969 and 2009, we have data for 2988 accidents and total number of passengers and flight crew members onboard during the accident is 98677.

Maximum number was 644, which belongs to a disaster in Tenerife, Spain, where two jumbo jets crashed on the runway.

```
air[air.Aboard ==644]
```

	Date	Time	Location	Operator	Type	Aboard	Fatalities	Ground	Summary	year	Survived
2963	03/27/1977	17:07	Tenerife, Canary Islands	Pan American World Airways / KLM	Boeing B-747-121 / Boeing B-747-206B	644.0	583.0	0.0	Both aircraft were diverted to Tenerife becaus...	1977	61.0

The average number of people in the aircraft at the time of the accident is 33. Both large and small aircraft were listed, which is shown by the standard deviation (std).

```
air.Aboard.sum()
```

98677.0

```
air.Aboard.describe()
```

```
count    2988.000000
mean       33.024431
std        53.155189
min         0.000000
25%         5.000000
50%        12.000000
75%        35.000000
max       644.000000
Name: Aboard, dtype: float64
```

How Many People Died?

Between 1969 and 2009, we have data for 2993 fatalities information for accidents involving airplanes and total number of victims involving passengers, flight crew members onboard and people on the ground, is 68714.

Maximum number was 583, which again belongs to the disaster in Tenerife, Spain, which was mentioned above.

The average number of fatalities is ~ 23. Both large and small aircraft were listed, which is shown by the standard deviation (std).

```
air.Fatalities.sum()
```

68714.0

```
air.Fatalities.describe()
```

```
count    2993.000000
mean       22.958236
std        40.415736
min         0.000000
25%         3.000000
50%         8.000000
75%        23.000000
max       583.000000
Name: Fatalities, dtype: float64
```

The graph below gives the annual total fatalities as a result of an aircraft accident. Although there are jumps in the values, the trend is downwards, pointing to the importance of safety precautions.

```
air.groupby('year')['Fatalities'].sum().plot(figsize = (15, 10));
```
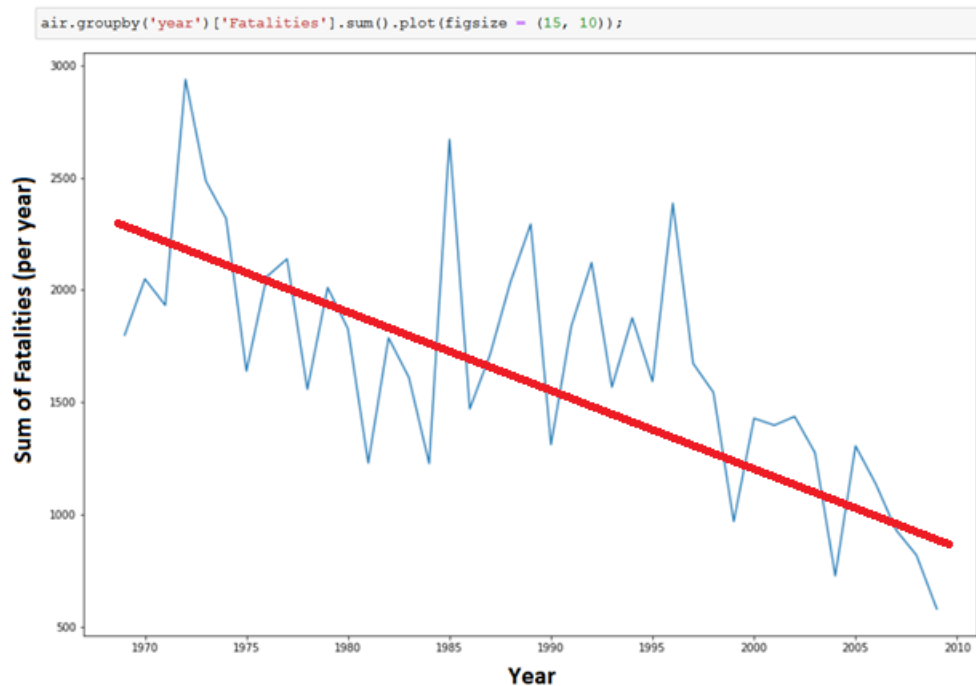

How Many People Survived?

Between 1969 and 2009, we have data for 2988 survival information for accidents involving airplanes and total number of survivors involving passengers and flight crew members onboard is 29988.

Maximum number of survivors was 516, which again belongs to a hijack incident of a jumbo jet in Japan. Only the captain of the aircraft was killed by the hijacker.

```
air[air.Survived ==516]
```

	Date	Time	Location	Operator	Type	Aboard	Fatalities	Ground	Summary	year	Survived
4645	07/23/1999	11:25	Tokyo, Japan	All Nippon Airways	Boeing B-747	517.0	1.0	0.0	Two minutes after taking off from Haneda Airpo...	1999	516.0

The average number of survivors is ~ 10. Both large and small aircraft were listed, which is shown by the standard deviation (std).

```
air.Survived.sum()
```
29988.0

```
air.Survived.describe()
```
```
count    2988.000000
mean       10.036145
std        35.572453
min         0.000000
25%         0.000000
50%         0.000000
75%         3.000000
max       516.000000
Name: Survived, dtype: float64
```

**Crash numbers grouped by the make/model of aircraft.**

Finally, I want to compare the statistics of the two major airplane manufacturing companies, namely Boeing and Airbus. I also included Fokker and Embraer to the comparison; as these two companies are leaders in the production of regional aircraft.

I do not want to make a mistake here. Both Airbus and Boeing produce very high quality and extremely safe aircraft. I tried to reach some company-originated data, but both companies tend to avoid safety comparisons when selling their aircraft to airlines or comparisons on product quality. Most aircraft dominating the companies' current sales have good safety records. Older model aircraft, which were first flown during the 1960s, 1970s, and 1980s, have had higher rates of fatal accidents (Wikipedia link). You can also reach a report from this link.

In our case, older model aircraft are defined as Airbus 300 and 310, and Boeing 707 and 727. Accident statistics also prove this case.

I made another data frame using and cleaning the data from the dataset. Basically, I removed much older Boeing models' accidents (only 7xx series planes were considered). The plot below shows the number of accidents of the companies, which manufacture bigger airplanes (capable of carrying more than 50 passengers) between 1969 and 2009.

```
sns.barplot(x='Make', y = 'Accidents', data = dm);
```
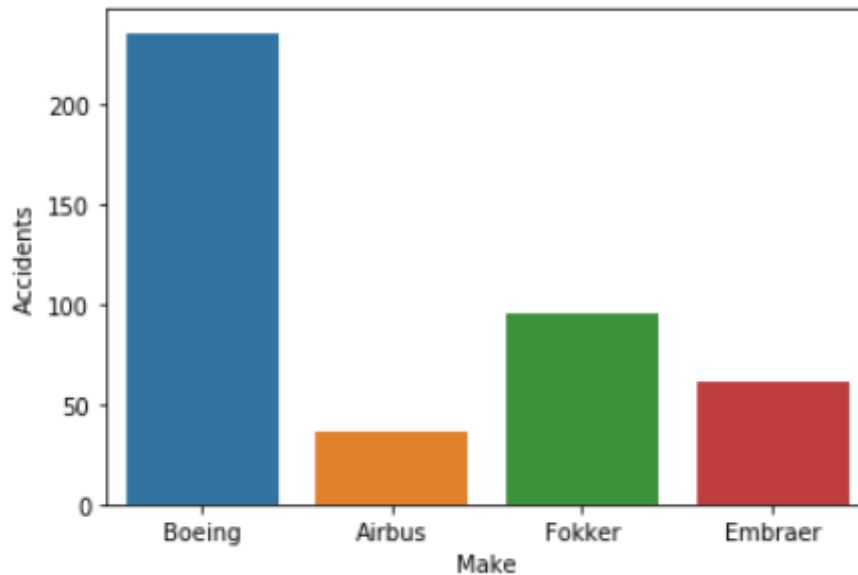

Also, details about the accident numbers of the Boeing and Airbus aircraft are shown in the plots below. Remember that, Boeing 737 is in service since 1968, the number built is more than 10,580 and there are many different types including military versions. The highest accident number may be misleading due to being the most popular aircraft in the commercial flight.

```
df.plot.bar();
```
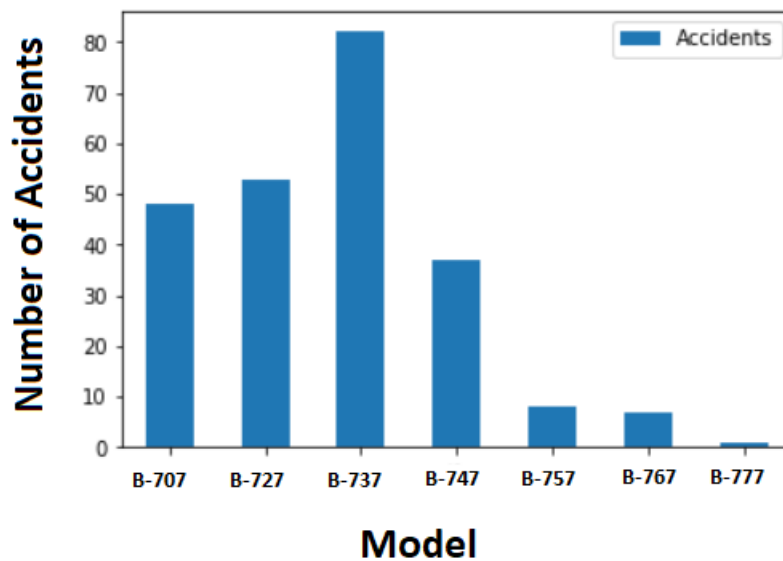
```
da.plot.bar();
```
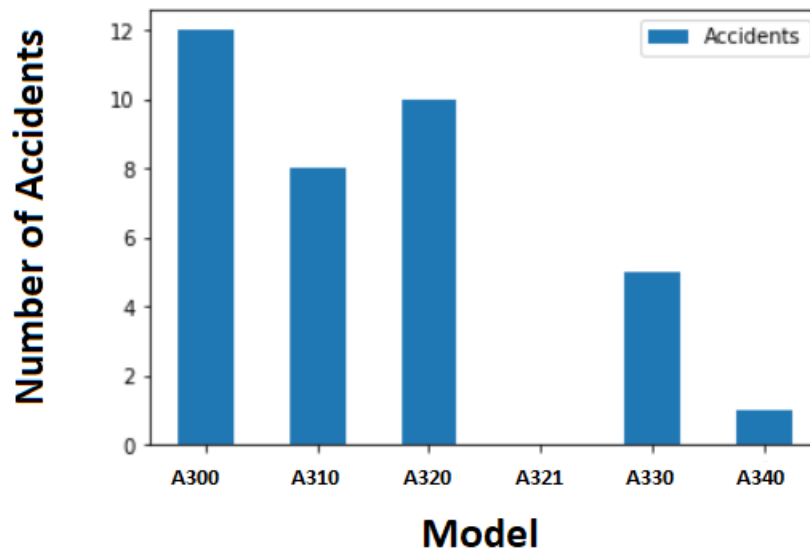

Conclusion

In this article, I analyzed the fatal and non-fatal airplane accident and incidents between 1969 and 2009, using the public data provided by data.world. The analysis provides evidence that:

- Total number of accidents were 2994. 1972 was the year with maximum number of crashes (104) and 2009 had the minimum number (30, only half of the year is covered).

- Total number of people onboard during the accidents were 98677 and total number of victims was 68714 (including passengers, flight crew and people on the ground).

- The highest number of people killed in a single accident is 583; which is a crash involving two jumbo jets.

- Totally, 29988 people survived the accidents.

- There is downward slope in annual total fatalities as a result of an aircraft accident, pointing to the results of increased aviation safety precautions.

- Finally, crash numbers were grouped by the make/model of aircraft. Four major producers were compared to get an idea about the safety of them. Older model aircraft, which were first flown during the 1960s, 1970s, and 1980s, have had higher rates of fatal accidents, but the newer models experience much less accidents.

Thank you for reading. I am planning to publish a more detailed follow-up article on this subject soon.

You can access to this article and similar ones here.