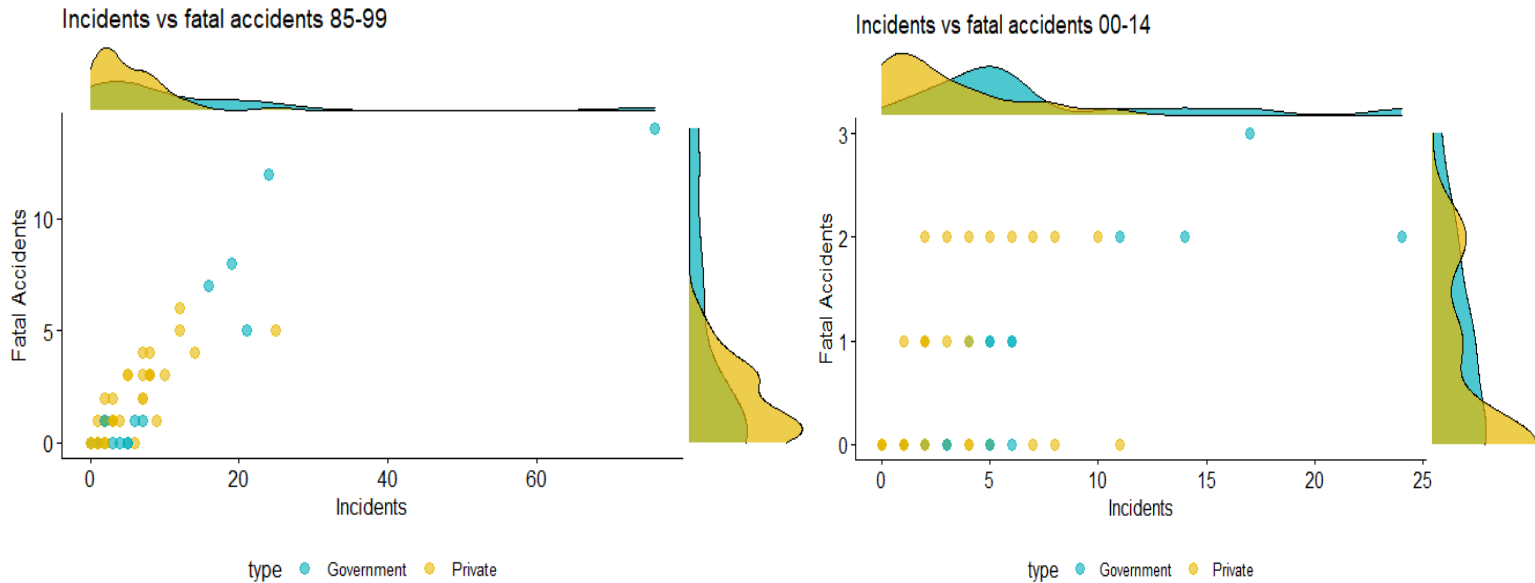


Overall, there is no relation between fatalities from one period to another but incidents are slightly correlated.

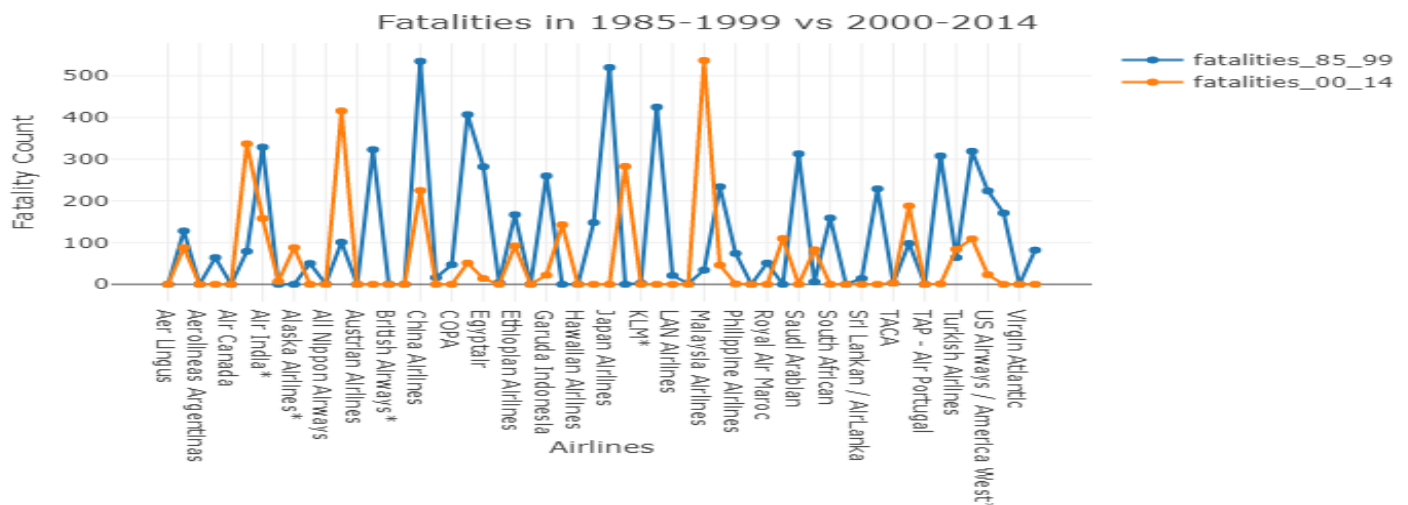
To gain detailed inferences about the data and the relationship among the variables, the following visualizations were produced:

The scatterplot below shows the incidents which resulted into fatal accidents for both the periods and the distribution of that between government and private owned airlines.

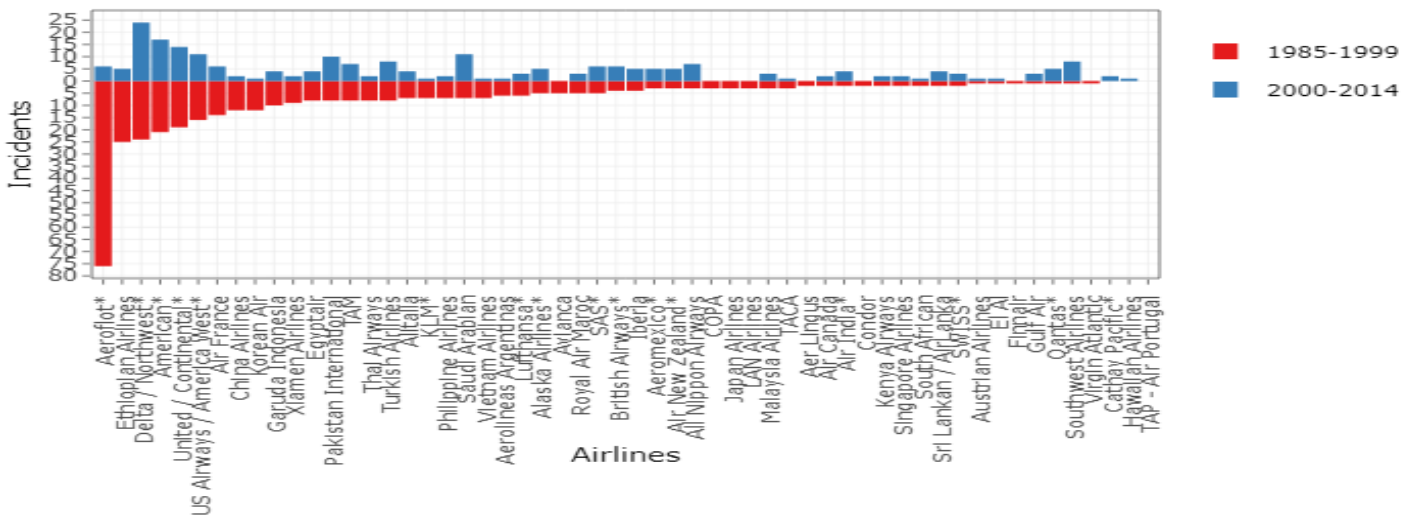
The plot shows that nationalized airlines tend to have more number crashes both in terms of incidents and fatal accidents and during both periods.



The above plot shows the distribution of fatalities over the two period. There are some airlines which contradicts our theory Eg: **Malaysian Airlines** which had a pretty good track record in the first period doomed to the most dangerous airline by the end of the second period whereas **Japan Airlines** which was apparently the most risky airlines in the first half improved significantly in the second period with 0 fatalities.



Airline incidents 1985-1999 vs 2000-2014

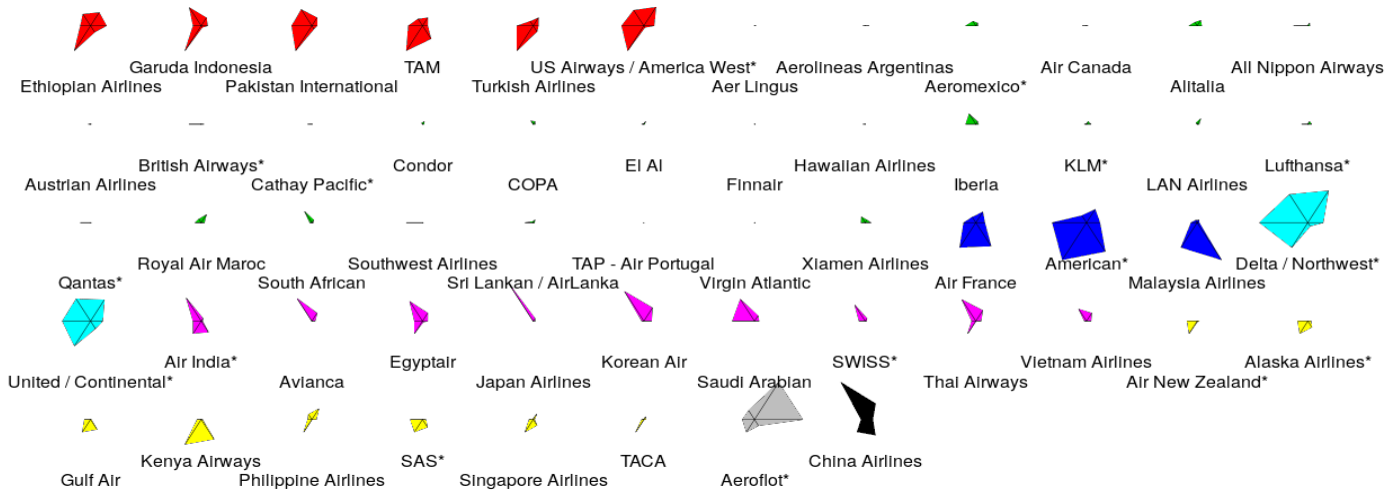


The distribution of incidents over the two time periods concur with our theory to an extent. Most of the airlines which had a good track record in terms of the number of incidents show the same trend in the second time period as well with few exceptions like **Aeroflot**.

Data Analysis and modelling:

K-Means Clustering:

To find groups within the data based on the available features we used K-means clustering algorithm. The data was divided into 8 clusters based on feature similarity.



The number of clusters (K) was decided based on the value of total within SS

Cluster 1: Airlines which have Highest Fatal Accidents among all the airlines (excluding American airlines) in Phase 2, with the number of fatalities significantly reduced in Phase 2, when compared to Phase 1.

Cluster 2: Airlines with significant number of Incidents and Fatalities in Phase 1, with incidents and fatalities highly reduced (nearly zero) in Phase 2.

Cluster 5: Airlines with all the incident, fatal accidents and fatality rates reduced in phase 2 from Phase 1. Supposedly the safest group.

Cluster 6: Airlines with very low incident rates in Phase 1 but have significantly higher rates in Phase 2. (except for Philippines airlines, which seems to be an outlier for the group where the accident rates have reduced).

Cluster 3: Airlines with highest fatality rates in Phase 2. Probably had met with huge accidents and hence **one of the most dangerous groups.**

Cluster 4: Airlines with high fatality rates in Phase 1, which significantly reduced in Phase 2. (These airlines can be trusted, as they have probably ensured to take safety measures and improve their overall flight safety).

Cluster 7: Aeroflot: Highest number of accidents in Phase 1. Have improved in their safety in Phase 2. Seems to have improved.

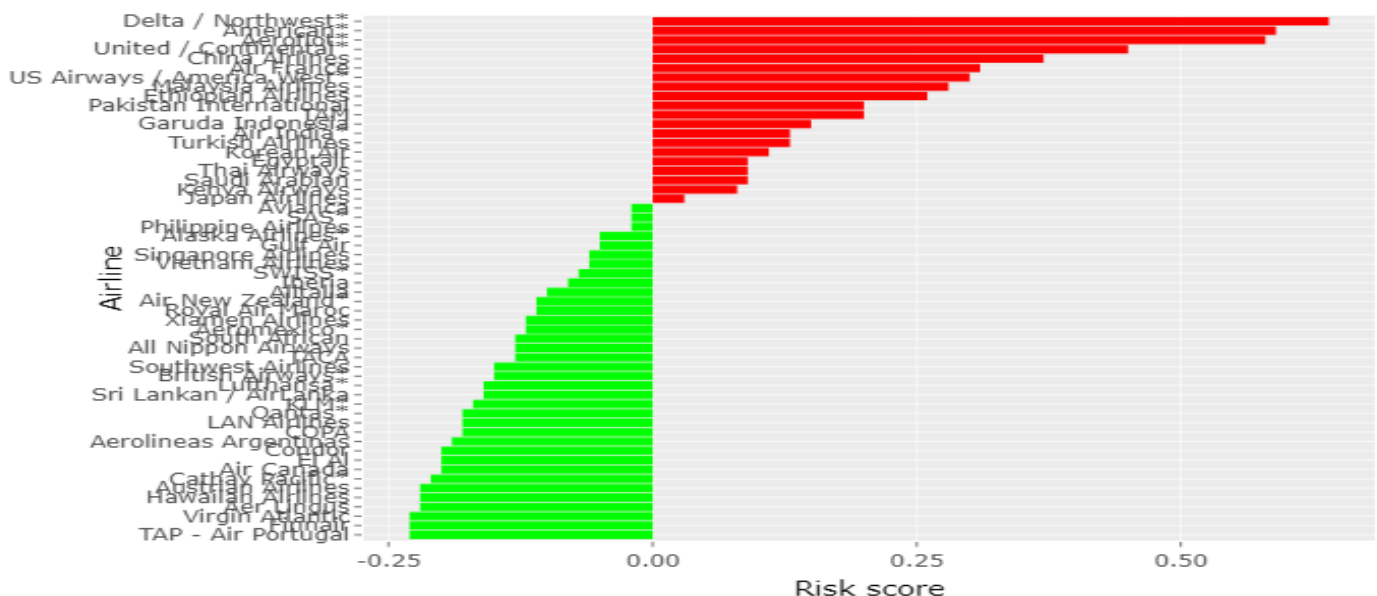
Cluster 8: China airlines has the number of highest fatalities over the period of 1985 to 1999.

Risk Score Calculation:

To quantitatively measure the risk of airlines and compare between we calculated a risk score by combining the 3 variables after scaling and giving appropriate weightages. The score for both periods was then combined to generate an overall score. The score was calculated through the following steps:

1. Calculating Harmless_Incidents: Incidents-Fatal Accidents
2. Standardizing Harmless_Incidents, Fatal_Accidents and Fatalities
3. Degree of harmfulness can be determined in the order: Fatalities > Fatal_Accidents > Harmless_Incident. So providing following weights to the columns:
Harmless_Incidents-25%
Fatal_Accidents-35%
Fatalities - 40%
4. Calculating Risk Score: avg(Harmless_Incidents, Fatal_Accidents , Fatalities)

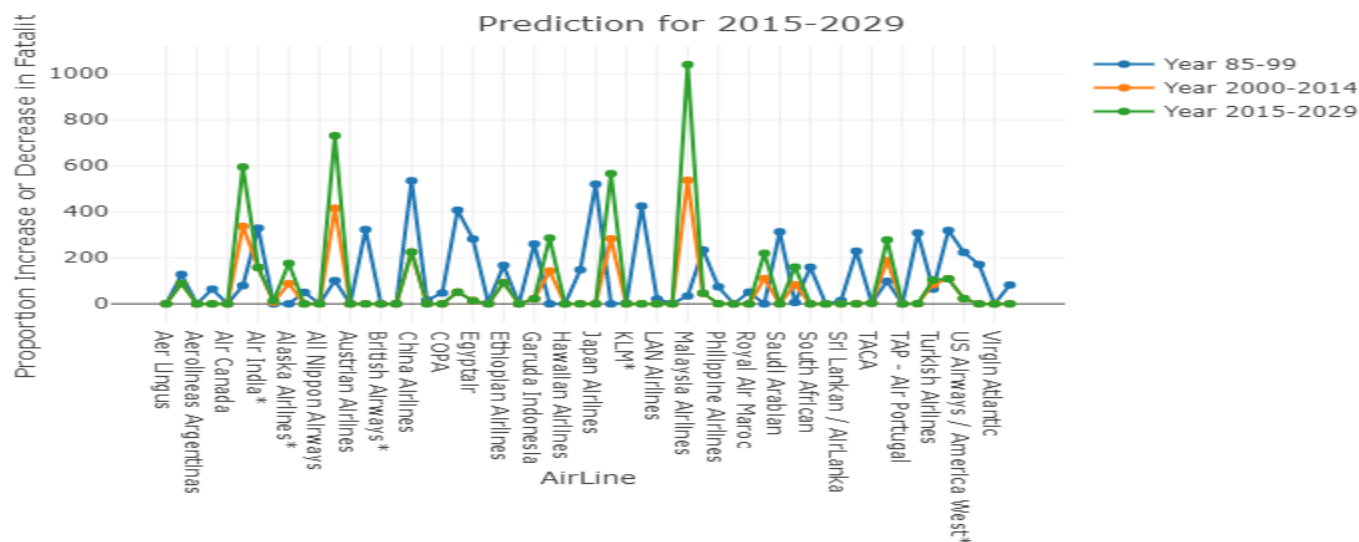
Positive Risk Score indicate a bad track record meaning the airline is unsafe and negative Risk score tells that the airline is relatively safe.



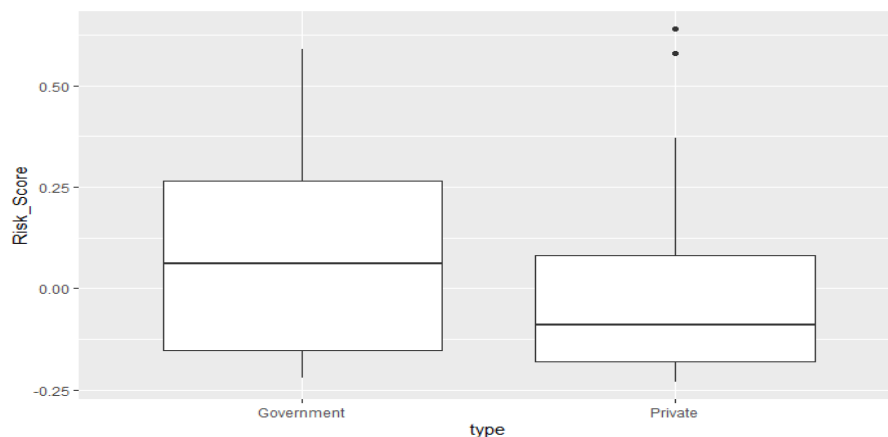
Prediction:

Below plot shows the prediction of the fatalities for 2015-2029 time period.

Malaysian airline has the high possible fatalities based on the previous data.



Inference:



If you conduct a regression on an airline's safety score from 2000-2014 on the basis of its safety score from 1985-1999 we can infer that only a small proportion is predictable based on the track record and can be ignored.

However, plotting the overall Score against the nationalized and Privatized airlines the mean risk score of privatized airlines is relatively small.

Conclusion:

Based on the analysis done it can be said that Private airlines are much safer than Government owned airlines. Malaysian Airlines which had a pretty good track record but ended up to be a most dangerous airline by the end of the second period. Whereas, Japan Airlines which was apparently the riskiest airlines in the first half improved significantly in the second period with no fatalities.

From Risk score generated we can say that TAP-Air Portugal is the safest airline, while the Delta/Northwest airlines is most un-safe airline.

Finally, for further precise conclusion over "Which Airlines is the safest to travel?" we have to take into account on other useful details like Total Passengers travelled, Number of Journey, Total Journey Hours, Type of accident (i.e. Human Error, Mechanical, Terrorist, etc.), Airplane type (i.e. Jumbo, mid-size, small, etc.) and Service Status of the Airlines (i.e. Active, Shutdown).

References:

- [1] Nate Silver. "Should Travelers Avoid Flying Airlines That Have Had Crashes in the Past?". <https://fivethirtyeight.com/features/should-travelers-avoid-flying-airlines-that-have-had-crashes-in-the-past/>
- [2] Catherine Hurley. "K-means Clustering". Maynooth university.