Airlines Safety Stats

Thip Rattanavilay, Master of Data Science

Bellevue University

DSC640

Author Note

The topic detail has been selected on personal interest and the data has been collected from online resources mentioned in the reference section.

Project Task 1: Dashboard

Your first task is to create an internal dashboard for your peers and data science management team that outlines the facts – what are the stats and what are the trends? Is there any supplemental data that you can use to support that air travel is still in fact the safest? Is there anything politically going on that would cause this type of media attention to be at a peak – remember, this is for an internal review by your peers and management – and will likely spark a lot of discussion for how you approach the next level of discussion with your executive leadership team. Is there anything to show sales are down or are headed that way? Do the safety incidents appear to be in a specific geographic area or by a specific airline every time? Do some analysis of the data you have and look for other sources to see what you can find to help inform your internal team.

Source Datasets:

1. Main data: Airline Safety

   Data for 56 airlines that were in the global top 100 as of December 2012 and which have operated continuously since Jan. 1, 1985.

   Codebook:

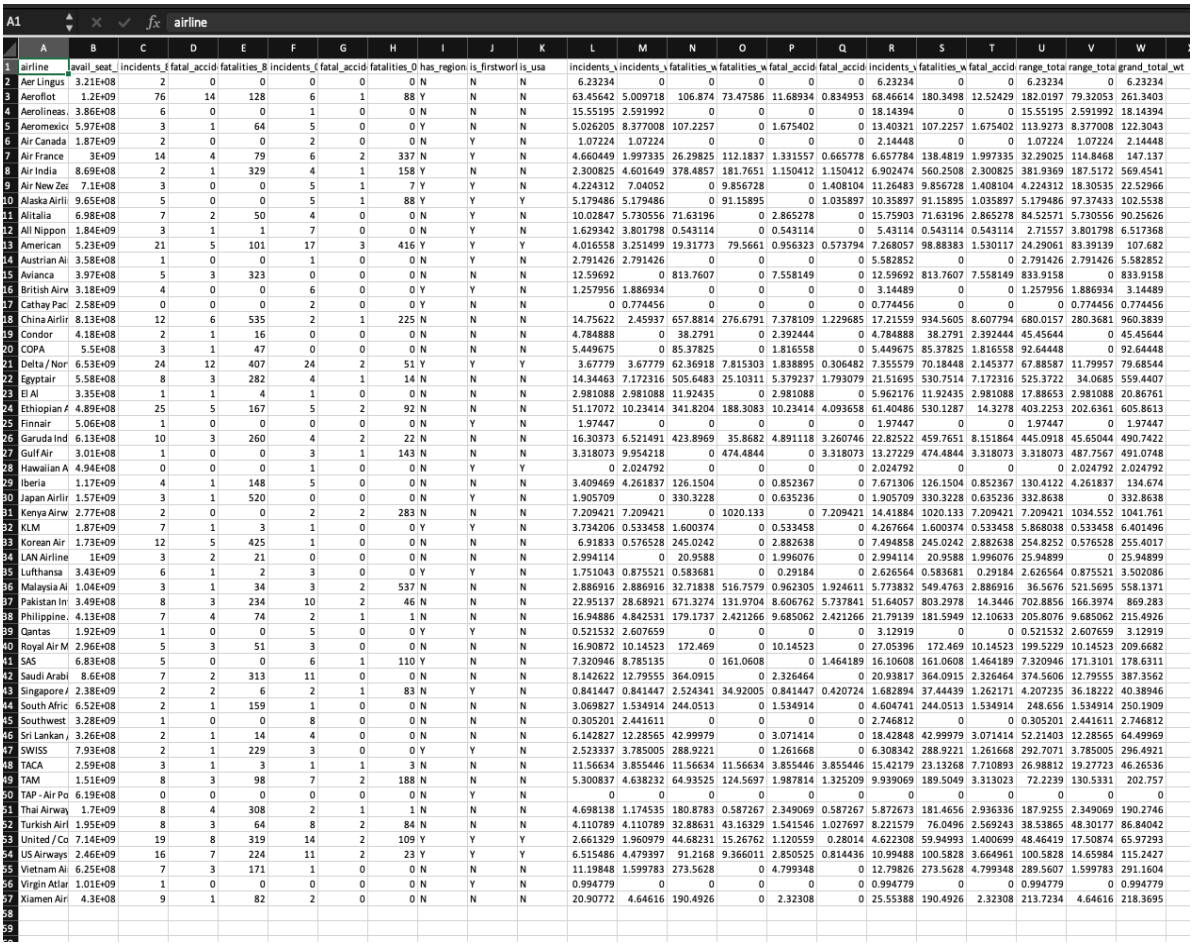
Header	Definition
airline	Airline (asterisk indicates that regional subsidiaries are included)
avail_seat_km_per_week	Available seat kilometers flown every week
incidents_85_99	Total number of incidents, 1985–1999
fatal_accidents_85_99	Total number of fatal accidents, 1985–1999
fatalities_85_99	Total number of fatalities, 1985–1999
incidents_00_14	Total number of incidents, 2000–2014
fatal_accidents_00_14	Total number of fatal accidents, 2000–2014
fatalities_00_14	Total number of fatalities, 2000–2014

2. Supplemental data: Auto Fatalities

   Data showing number of people killed and injured in fatal collisions.

Codebook:

Header	Definition
Year	Year of incident
Deaths	Number of fatalities
VMT_Vehicle_Miles_Travelled_bn	Vehicle travelled miles in billion
Fatalities_Per_100_Million_VMT	Number of fatalities per 100 million Vehicle Miles Travelled
Population	US population
Fatalities_Per_100000_population	Fatalities per 100k population



airline	avail_seat_	incidents_t	fatal_accid	fatalities_8	incidents_(	fatal_accid	fatalities_0	has_region	is_firstworl	is_usa	incidents_v	incidents_v	fatalities_w	fatalities_w	fatal_accid	fatal_accid	incidents_v	fatalities_w	fatal_accid	range_total	range_total	grand_total_wt
Aer Lingus	3.21E+08	2	0	0	0	0	0	N	N	N	6.23234	0	0	0	0	0	6.23234	0	0	6.23234	0	6.23234
Aeroflot	1.2E+09	76	14	128	6	1	88	Y	N	N	63.45642	5.009718	106.874	73.47586	11.68934	0.834953	68.46614	180.3498	12.52429	182.0197	79.32053	261.3403
Aerolineas_	3.86E+08	6	0	0	1	0	0	N	N	N	15.55195	2.591992	0	0	0	0	18.14394	0	0	15.55195	2.591992	18.14394
Aeromexico	5.97E+08	3	1	64	5	0	0	Y	N	N	5.026205	8.377008	107.2257	0	1.675402	0	13.40321	107.2257	1.675402	113.9273	8.377008	122.3043
Air Canada	1.87E+09	2	0	0	2	0	0	N	Y	N	1.07224	1.07224	0	0	0	0	2.14448	0	0	1.07224	1.07224	2.14448
Air France	3E+09	14	4	79	6	2	337	N	Y	N	4.660449	1.997335	26.29825	112.1837	1.331557	0.665778	6.657784	138.4819	1.997335	32.29025	114.8468	147.137
Air India	8.69E+08	2	1	329	4	1	158	Y	N	N	2.300825	4.601649	378.4857	181.7651	1.150412	1.150412	6.902474	560.2508	2.300825	381.9369	187.5172	569.4541
Air New Zea	7.1E+08	3	0	0	5	1	7	Y	Y	N	4.224312	7.04052	0	9.856728	0	1.408104	11.26483	9.856728	1.408104	4.224312	18.30535	22.52966
Alaska Airli	9.65E+08	5	0	0	5	1	88	Y	Y	Y	5.179486	5.179486	0	91.15895	0	1.035897	10.35897	91.15895	1.035897	5.179486	97.37433	102.5538
Alitalia	6.98E+08	7	2	50	4	0	0	N	Y	N	10.02847	5.730556	71.63196	0	2.865278	0	15.75903	71.63196	2.865278	84.52571	5.730556	90.25626
All Nippon	1.84E+09	3	1	1	7	0	0	N	Y	N	1.629342	3.801798	0.543114	0	0.543114	0	5.43114	0.543114	0.543114	2.71557	3.801798	6.517368
American	5.23E+09	21	5	101	17	3	416	Y	Y	Y	4.016558	3.251499	19.31773	79.5661	0.956323	0.573794	7.268057	98.88383	1.530117	24.29061	83.39139	107.682
Austrian Ai	3.58E+08	1	0	0	1	0	0	N	Y	N	2.791426	2.791426	0	0	0	0	5.582852	0	0	2.791426	2.791426	5.582852
Avianca	3.97E+08	5	3	323	0	0	0	N	N	N	12.59692	0	813.7607	0	7.558149	0	12.59692	813.7607	7.558149	833.9158	0	833.9158
British Airv	3.18E+09	4	0	0	6	0	0	Y	Y	N	1.257956	1.886934	0	0	0	0	3.14489	0	0	1.257956	1.886934	3.14489
Cathay Pac	2.58E+09	0	0	0	2	0	0	Y	N	N	0	0.774456	0	0	0	0	0.774456	0	0	0	0.774456	0.774456
China Airlir	8.13E+08	12	6	535	2	1	225	N	N	N	14.75622	2.45937	657.8814	276.6791	7.378109	1.229685	17.21559	934.5605	8.607794	680.0157	280.3681	960.3839
Condor	4.18E+08	2	1	16	0	0	0	N	N	N	4.784888	0	38.2791	0	2.392444	0	4.784888	38.2791	2.392444	45.45644	0	45.45644
COPA	5.5E+08	3	1	47	0	0	0	N	N	N	5.449675	0	85.37825	0	1.816558	0	5.449675	85.37825	1.816558	92.64448	0	92.64448
Delta / Nor	6.53E+09	24	12	407	24	2	51	Y	Y	Y	3.67779	3.67779	62.36918	7.815303	1.838895	0.306482	7.355579	70.18448	2.145377	67.88587	11.79957	79.68544
Egyptair	5.58E+08	8	3	282	4	1	14	N	N	N	14.34463	7.172316	505.6483	25.10311	5.379237	1.793079	21.51695	530.7514	7.172316	525.3722	34.0685	559.4407
El Al	3.35E+08	1	1	4	1	0	0	N	N	N	2.981088	2.981088	11.92435	0	2.981088	0	5.962176	11.92435	2.981088	17.88653	2.981088	20.86761
Ethiopian A	4.89E+08	25	5	167	5	2	92	N	N	N	51.17072	10.23414	341.8204	188.3083	10.23414	4.093658	61.40486	530.1287	14.3278	403.2253	202.6361	605.8613
Finnair	5.06E+08	1	0	0	0	0	0	N	Y	N	1.97447	0	0	0	0	0	1.97447	0	0	1.97447	0	1.97447
Garuda Ind	6.13E+08	10	3	260	4	2	22	N	N	N	16.30373	6.521491	423.8969	35.8682	4.891118	3.260746	22.82522	459.7651	8.151864	445.0918	45.65044	490.7422
Gulf Air	3.01E+08	1	0	0	3	1	143	N	N	N	3.318073	9.954218	0	474.4844	0	3.318073	13.27229	474.4844	3.318073	3.318073	487.7567	491.0748
Hawaiian A	4.94E+08	0	0	0	1	0	0	N	Y	Y	0	2.024792	0	0	0	0	2.024792	0	0	0	2.024792	2.024792
Iberia	1.17E+09	4	1	148	5	0	0	N	N	N	3.409469	4.261837	126.1504	0	0.852367	0	7.671306	126.1504	0.852367	130.4122	4.261837	134.674
Japan Airlir	1.57E+09	3	1	520	0	0	0	N	Y	N	1.905709	0	330.3228	0	0.635236	0	1.905709	330.3228	0.635236	332.8638	0	332.8638
Kenya Airw	2.77E+08	2	0	0	2	2	283	N	N	N	7.209421	7.209421	0	1020.133	0	7.209421	14.41884	1020.133	7.209421	7.209421	1034.552	1041.761
KLM	1.87E+09	7	1	3	1	0	0	Y	Y	N	3.734206	0.533458	1.600374	0	0.533458	0	4.267664	1.600374	0.533458	5.868038	0.533458	6.401496
Korean Air	1.73E+09	12	5	425	1	0	0	N	N	N	6.91833	0.576528	245.0242	0	2.882638	0	7.494858	245.0242	2.882638	254.8252	0.576528	255.4017
LAN Airline	1E+09	3	2	21	0	0	0	N	N	N	2.994114	0	20.9588	0	1.996076	0	2.994114	20.9588	1.996076	25.94899	0	25.94899
Lufthansa	3.43E+09	6	1	2	3	0	0	Y	Y	N	1.751043	0.875521	0.583681	0	0.29184	0	2.626564	0.583681	0.29184	2.626564	0.875521	3.502086
Malaysia Ai	1.04E+09	3	1	34	3	2	537	N	N	N	2.886916	2.886916	32.71838	516.7579	0.962305	1.924611	5.773832	549.4763	2.886916	36.5676	521.5695	558.1371
Pakistan In	3.49E+08	8	3	234	10	2	46	N	N	N	22.95137	28.68921	671.3274	131.9704	8.606762	5.737841	51.64057	803.2978	14.3446	702.8856	166.3974	869.283
Philippine_	4.13E+08	7	4	74	2	1	1	N	N	N	16.94886	4.842531	179.1737	2.421266	9.685062	2.421266	21.79139	181.5949	12.10633	205.8076	9.685062	215.4926
Qantas	1.92E+09	1	0	0	5	0	0	Y	Y	N	0.521532	2.607659	0	0	0	0	3.12919	0	0	0.521532	2.607659	3.12919
Royal Air M	2.96E+08	5	3	51	3	0	0	N	N	N	16.90872	10.14523	172.469	0	10.14523	0	27.05396	172.469	10.14523	199.5229	10.14523	209.6682
SAS	6.83E+08	5	0	0	6	1	110	Y	N	N	7.320946	8.785135	0	161.0608	0	0.1464189	16.10608	161.0608	1.464189	7.320946	171.3101	178.6311
Saudi Arabi	8.6E+08	7	2	313	11	0	0	N	N	N	8.142622	12.79555	364.0915	0	2.326464	0	20.93817	364.0915	2.326464	374.5606	12.79555	387.3562
Singapore /	2.38E+09	2	2	6	2	1	83	N	Y	N	0.841447	0.841447	2.524341	34.92005	0.841447	0.420724	1.682894	37.44439	1.262171	4.207235	36.18222	40.38946
South Afric	6.52E+08	2	1	159	1	0	0	N	N	N	3.069827	1.534914	244.0513	0	1.534914	0	4.604741	244.0513	1.534914	248.656	1.534914	250.1909
Southwest	3.28E+09	1	0	0	8	0	0	N	N	N	0.305201	2.441611	0	0	0	0	2.746812	0	0	0.305201	2.441611	2.746812
Sri Lankan /	3.26E+08	2	1	14	4	0	0	N	N	N	6.142827	12.28565	42.99979	0	3.071414	0	18.42848	42.99979	3.071414	52.21403	12.28565	64.49969
SWISS	7.93E+08	2	1	229	3	0	0	Y	Y	N	2.523337	3.785005	288.9221	0	1.261668	0	6.308342	288.9221	1.261668	292.7071	3.785005	296.4921
TACA	2.59E+08	3	1	3	1	1	3	N	N	N	11.56634	3.855446	11.56634	11.56634	3.855446	3.855446	15.42179	23.13268	7.710893	26.98812	19.27723	46.26536
TAM	1.51E+09	8	3	98	7	2	188	N	N	N	5.300837	4.638232	64.93525	124.5697	1.987814	1.325209	9.939069	189.5049	3.313023	72.2239	130.5331	202.757
TAP - Air Po	6.19E+08	0	0	0	0	0	0	N	Y	N	0	0	0	0	0	0	0	0	0	0	0	0
Thai Airway	1.7E+09	8	4	308	2	1	1	N	N	N	4.698138	1.174535	180.8783	0.587267	2.349069	0.587267	5.872673	181.4656	2.936336	187.9255	2.349069	190.2746
Turkish Airl	1.95E+09	8	3	64	8	2	84	N	N	N	4.110789	4.110789	32.88631	43.16329	1.541546	1.027697	8.221579	76.0496	2.569243	38.53865	48.30177	86.84042
United / Co	7.14E+09	19	8	319	14	2	109	Y	Y	Y	2.661329	1.960979	44.68231	15.26762	1.120559	0.28014	4.622308	59.94993	1.400699	48.46419	17.50874	65.97293
US Airways	2.46E+09	16	7	224	11	2	23	Y	Y	Y	6.515486	4.479397	91.2168	9.366011	2.850525	0.814436	10.99488	100.5828	3.664961	100.5828	14.65984	115.2427
Vietnam Ai	6.25E+08	7	3	171	1	0	0	N	N	N	11.19848	1.599783	273.5628	0	4.799348	0	12.79826	273.5628	4.799348	289.5607	1.599783	291.1604
Virgin Atlar	1.01E+09	1	0	0	0	0	0	N	Y	N	0.994779	0	0	0	0	0	0.994779	0	0	0.994779	0	0.994779
Xiamen Air	4.3E+08	9	1	82	2	0	0	N	N	N	20.90772	4.64616	190.4926	0	2.32308	0	25.55388	190.4926	2.32308	213.7234	4.64616	218.3695

A1 fx airline

	A	B	C	D	E	F	G	H	I
1	airline	has_region	is_firstworl	is_usa	avail_seat_	type	period	count	count_wt
2	Aer Lingus	N	N	N	3.21E+08	Incident	1985-1999	2	6.23234
3	Aeroflot	Y	N	N	1.2E+09	Incident	1985-1999	76	63.45642
4	Aerolineas	N	N	N	3.86E+08	Incident	1985-1999	6	15.55195
5	Aeromexic	Y	N	N	5.97E+08	Incident	1985-1999	3	5.026205
6	Air Canada	N	Y	N	1.87E+09	Incident	1985-1999	2	1.07224
7	Air France	N	Y	N	3E+09	Incident	1985-1999	14	4.660449
8	Air India	Y	N	N	8.69E+08	Incident	1985-1999	2	2.300825
9	Air New Zea	Y	Y	N	7.1E+08	Incident	1985-1999	3	4.224312
10	Alaska Airli	Y	Y	Y	9.65E+08	Incident	1985-1999	5	5.179486
11	Alitalia	N	Y	N	6.98E+08	Incident	1985-1999	7	10.02847
12	All Nippon	N	Y	N	1.84E+09	Incident	1985-1999	3	1.629342
13	American	Y	Y	Y	5.23E+09	Incident	1985-1999	21	4.016558
14	Austrian Ai	N	Y	N	3.58E+08	Incident	1985-1999	1	2.791426
15	Avianca	N	N	N	3.97E+08	Incident	1985-1999	5	12.59692
16	British Airw	Y	Y	N	3.18E+09	Incident	1985-1999	4	1.257956
17	Cathay Pac	Y	N	N	2.58E+09	Incident	1985-1999	0	0
18	China Airlir	N	N	N	8.13E+08	Incident	1985-1999	12	14.75622
19	Condor	N	N	N	4.18E+08	Incident	1985-1999	2	4.784888
20	COPA	N	N	N	5.5E+08	Incident	1985-1999	3	5.449675
21	Delta / Nor	Y	Y	Y	6.53E+09	Incident	1985-1999	24	3.67779
22	Egyptair	N	N	N	5.58E+08	Incident	1985-1999	8	14.34463
23	El Al	N	N	N	3.35E+08	Incident	1985-1999	1	2.981088
24	Ethiopian A	N	N	N	4.89E+08	Incident	1985-1999	25	51.17072
25	Finnair	N	Y	N	5.06E+08	Incident	1985-1999	1	1.97447
26	Garuda Ind	N	N	N	6.13E+08	Incident	1985-1999	10	16.30373
27	Gulf Air	N	N	N	3.01E+08	Incident	1985-1999	1	3.318073
28	Hawaiian A	N	Y	Y	4.94E+08	Incident	1985-1999	0	0
29	Iberia	N	N	N	1.17E+09	Incident	1985-1999	4	3.409469
30	Japan Airlir	N	Y	N	1.57E+09	Incident	1985-1999	3	1.905709
31	Kenya Airw	N	N	N	2.77E+08	Incident	1985-1999	2	7.209421
32	KLM	Y	Y	N	1.87E+09	Incident	1985-1999	7	3.734206
33	Korean Air	N	N	N	1.73E+09	Incident	1985-1999	12	6.91833
34	LAN Airline	N	N	N	1E+09	Incident	1985-1999	3	2.994114
35	Lufthansa	Y	Y	N	3.43E+09	Incident	1985-1999	6	1.751043
36	Malaysia Ai	N	N	N	1.04E+09	Incident	1985-1999	3	2.886916
37	Pakistan In	N	N	N	3.49E+08	Incident	1985-1999	8	22.95137
38	Philippine	N	N	N	4.13E+08	Incident	1985-1999	7	16.94886
39	Qantas	Y	Y	N	1.92E+09	Incident	1985-1999	1	0.521532
40	Royal Air M	N	N	N	2.96E+08	Incident	1985-1999	5	16.90872
41	SAS	Y	N	N	6.83E+08	Incident	1985-1999	5	7.320946
42	Saudi Arabi	N	N	N	8.6E+08	Incident	1985-1999	7	8.142622
43	Singapore A	N	Y	N	2.38E+09	Incident	1985-1999	2	0.841447
44	South Afric	N	N	N	6.52E+08	Incident	1985-1999	2	3.069827
45	Southwest	N	N	N	3.28E+09	Incident	1985-1999	1	0.305201
46	Sri Lankan	N	N	N	3.26E+08	Incident	1985-1999	2	6.142827
47	SWISS	Y	Y	N	7.93E+08	Incident	1985-1999	2	2.523337
48	TACA	N	N	N	2.59E+08	Incident	1985-1999	3	11.56634
49	TAM	N	N	N	1.51E+09	Incident	1985-1999	8	5.300837
50	TAP - Air Po	N	Y	N	6.19E+08	Incident	1985-1999	0	0
51	Thai Airway	N	N	N	1.7E+09	Incident	1985-1999	8	4.698138
52	Turkish Airl	N	N	N	1.95E+09	Incident	1985-1999	8	4.110789
53	United / Co	Y	Y	Y	7.14E+09	Incident	1985-1999	19	2.661329
54	US Airways	Y	Y	Y	2.46E+09	Incident	1985-1999	16	6.515486
55	Vietnam Ai	N	N	N	6.25E+08	Incident	1985-1999	7	11.19848

Transformed Data Set for Power BI:

R is used for transformation of data and Tablu or Power BI is used to build the dashboard. Transformed datasets are as follows:

AirlineSafety – Weighted incident, fatal accident and fatalities are calculated based on per billion '*available seat km per week*'. The table is in wide format.

AirlineSafetyLong – AirlineSafety data transposed to a long format for easy reporting.

AutoStats – Summary of auto fatalities for the given timeframe.

Fatalities – Fatalities for the same period for airlines and auto. This for comparison in the visual.

Considerations:

Following airlines are flagged as first world countries based on their coutry of core operations: *Air Canada, Air France, Air New Zealand, Alaska Airlines, Alitalia, All Nippon Airways, American, Austrian Airlines, British Airways, Delta / Northwest, Finnair, Hawaiian Airlines, Japan Airlines, KLM, Lufthansa, Qantas, Singapore Airlines, SWISS, TAP - Air Portugal, United / Continental, US Airways / America West, Virgin Atlantic*.

Dashboard:

Intention of the dashboard is to expose the truth behind the current negative buzz on air travel through historical facts and stats. From a design perspective, colorblind safe pallet has been used and simple high impact chart types have been used to maintain clarity.

1. **Fatalities by Airlines**

   This scatter plot plots fatalities by airlines (color coded) for the two time periods (1985-1999, 2000-2014). The intention is to find correlation between the fatalities across the two periods.

There is no predictable nature of the number of incidents for the airlines between the two periods. Air Kenya had no fatalities in ’85-’99 period but has very high rate in the other period (’00-’14). Avianca on the other hand had high fatality during ’85-’99 but got better in the second half. China and Pakistan airlines has high rate in both halves. But basically, there is no pattern.

2. **Incidents by Airlines**

The intention is to find correlation between the incidents across the two periods.

We can see that there is a modest correlation between the two periods.

The two major outliers in the chart are Pakistan International Airlines and Ethiopian Airlines, which have had a persistently high rate of incidents. A third outlier, Russia’s Aeroflot, had an extraordinarily high number of reported incidents in ’85-’99. But many of these incidents are attempted hijackings around the time of the breakup of the Soviet Union. Contrastingly it has relatively low number in the ’00-’14 period. Even if we exclude these the outliers, there is still a positive correlation.

The statistical inference from this chart is, some airlines are slightly safer to fly than others.

3. **Top 5 airlines by Total Incidents**

This horizontal bar chart lists the top 5 airlines in terms of overall incidents. Noticeably, none of the 5 airlines are from the first world countries. Therefore, safety measure and technical deficit could be a reason behind the incidents.

The takeaway from this graph is, it cannot be generalized that air travel is unsafe. It might very well have a relationship between airlines and country of operations.

4. **Fatalities Trend: First World vs Rest**

This clustered graph compares the count of fatalities between first world countries and the rest of countries over the two periods. Although in both cases, the overall count decreased in ’00-’14 compared to ’85-’99, first world numbers are considerably lower than the rest in both cases. This chart re-established the fact that there is stark difference between air travel quality of countries. Air travel is not at all unsafe for some countries.

5. **Incidents Trend: First World vs Rest**

   This chart is almost identical to the prior graph. The inference is also very similar. The only noteworthy point is, first world countries did not improve much in terms of incidents. The high air traffic could be a reason behind the high number of incidents in first world countries. But in comparison with the rest, the numbers are still pretty low.

6. **Motor Vehicle Deaths (US) vs Airlines Deaths (First World)**

   This visual use supplemental data of auto fatalities in the USA for the same two periods (’85 - ’99, ’00 - ‘14). The airlines fatalities are filtered for US airlines and compared to the auto numbers. Although there are decreasing trend of incidents in both travels, auto casualties are consistently far greater than air travel. If the facts in the data is to be considered as truth then this visual clearly shows that just by considering number of fatalities, travelling by road is far more dangerous than travelling by air.

Reference : Data Set

FiveThirtyEight, Airline Safety, Retrieved from https://github.com/fivethirtyeight/data/tree/master/airline-safety

Motor vehicle fatality rate in U.S. by year, Retrieved from https://en.wikipedia.org/wiki/Motor_vehicle_fatality_rate_in_U.S._by_year