Airline Safety

April 11, 2021

1 Are Airlines Safe to Travel?

Due to recent unfortunate airline crashes, the media has been promoting statistics stating air is no longer a safe way to travel. The news and media outlets have been bombarding the public with reports and figures about the trends of airline safety and that things are not looking good. What was previously thought as the safest way to travel, especially when compared to automobiles, is now being presented as one of the most dangerous to the public. But are any of these claims based on facts?

We have collected data from two different data sources. One from fivethirtyeight.com and one from the Department of Transportation analytics website. The goal of this project will be to evaluate how safe airlines are based on statistics and visualization.

The data from fivethirtyeight.com has data for the following features: - airline: which airline company the data belongs to - avail_seat_km_per_week: Number of seats that move 1 kilometer per week. This is a common metric when evaluating air travel due to varying cabin capacity for multiple airplanes - incidents_85_99 : Number of airline incidents per airline between 1985 and 1999 - fatal_accidents_85_99: Number of fatal accidents per airline between 1985 and 1999 - fatalities_85_99 : Number of fatalities per airline between 1985 and 1999 - incidents_00_14 : Number of airline incidents per airline between 2000 and Jun 2014 - fatal_accidents_00_14: Number of fatal accidents per airline between 2000 and Jun 2014 - fatalities_00_14 : Number of fatalities per airline between 2000 and Jun 2014

[91]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import datetime
import seaborn as sns
```

[25]:
```python
df = pd.read_csv("airline-safety.csv")
df.head()
```

[25]:
```
                 airline  avail_seat_km_per_week  incidents_85_99  \
0             Aer Lingus               320906734                2
1              Aeroflot*              1197672318               76
2  Aerolineas Argentinas               385803648                6
3            Aeromexico*               596871813                3
4             Air Canada              1865253802                2
```

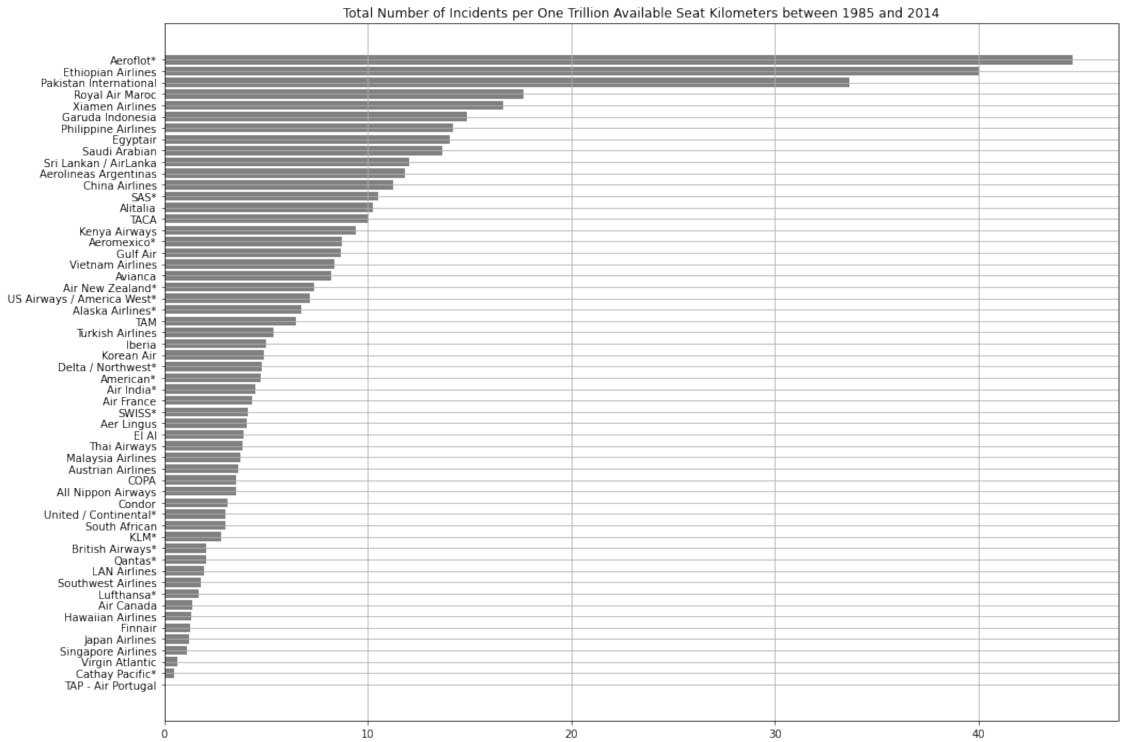
```
   fatal_accidents_85_99  fatalities_85_99  incidents_00_14  \
0                      0                 0                0
1                     14               128                6
2                      0                 0                1
3                      1                64                5
4                      0                 0                2

   fatal_accidents_00_14  fatalities_00_14
0                      0                 0
1                      1                88
2                      0                 0
3                      0                 0
4                      0                 0
```

Using this data, we can generate some statistics to help us better evaluate how safe each airline company is. A common metric used in these evaluations is the number of incidents, accidents, or fatalities per one trillion avialable seat kilometers. We can calculate these metrics by using the following calculations

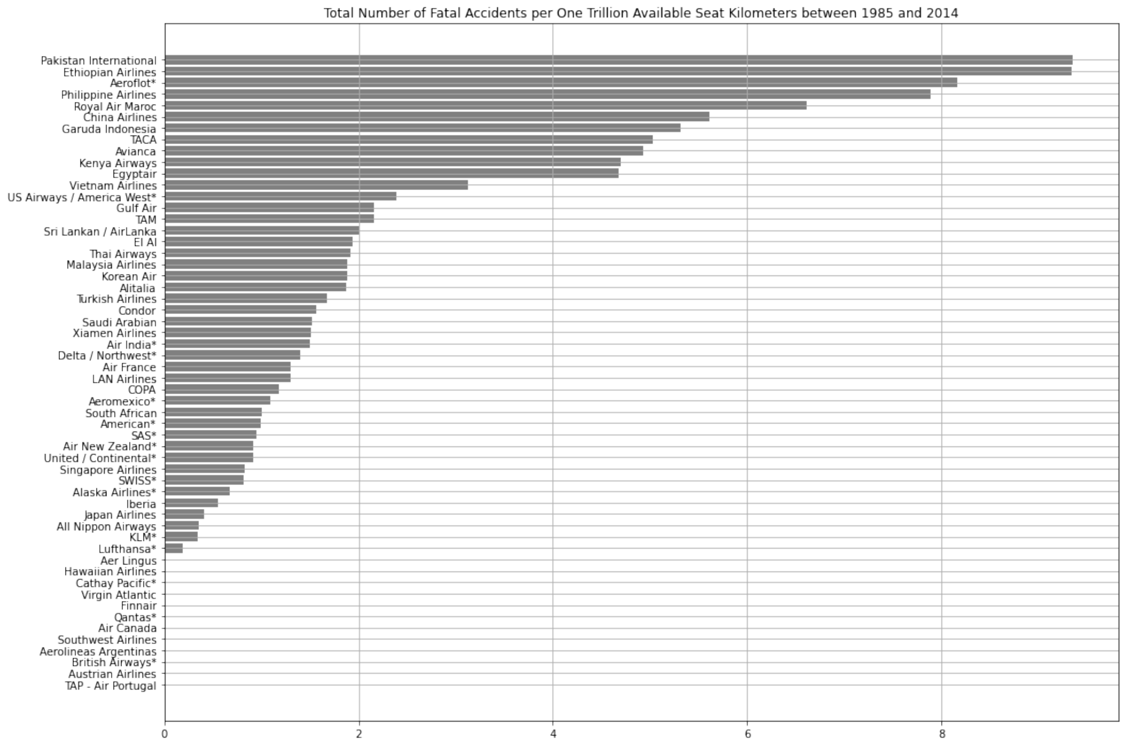
[28]:
```
x = datetime.datetime(2014,6,14) - datetime.datetime(2000,1,1)

df['total_fatailites'] = df['fatalities_85_99'] + df['fatalities_00_14']
df['total_incidents'] = df['incidents_85_99'] + df['incidents_00_14']
df['total_fatal_accidents'] = df['fatal_accidents_85_99'] +␣
 ↪df['fatal_accidents_00_14']

df['trillion_available_seat_km_00_14'] = (df['avail_seat_km_per_week'] * x.days␣
 ↪/ 7) / 10**12
df['trillion_available_seat_km_85_99'] = (df['avail_seat_km_per_week'] * 52 *␣
 ↪15) / 10**12

df['total_trillion_available_seat_km'] = df['trillion_available_seat_km_85_99']␣
 ↪+ df['trillion_available_seat_km_00_14']

df['total_fatalities per trillion_available_seat_km'] = df['total_fatailites'] /
 ↪ df['total_trillion_available_seat_km']
df['total_incidents per trillion_available_seat_km'] = df['total_incidents'] /␣
 ↪df['total_trillion_available_seat_km']
df['total_fatal_accidents per trillion_available_seat_km'] =␣
 ↪df['total_fatal_accidents'] / df['total_trillion_available_seat_km']

df.to_csv("data4.csv")
```

[29]:
```
df = df.sort_values("total_incidents per trillion_available_seat_km")

plt.figure(figsize=(16,12))
```

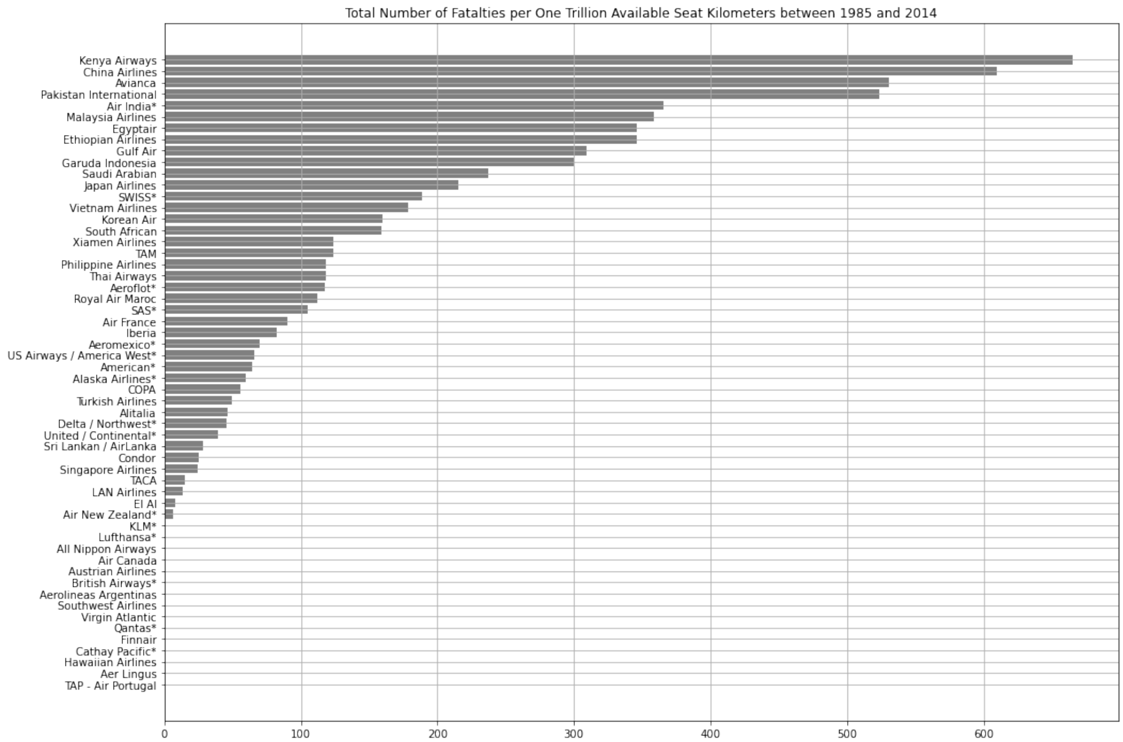
```
plt.barh(y = df["airline"], width = df['total_incidents per␣
↪trillion_available_seat_km'], color = 'gray')
plt.grid()
plt.title("Total Number of Incidents per One Trillion Available Seat Kilometers␣
↪between 1985 and 2014")
plt.show()
```



[30]:
```
df = df.sort_values("total_fatal_accidents per trillion_available_seat_km")

plt.figure(figsize=(16,12))
plt.barh(y = df["airline"], width = df['total_fatal_accidents per␣
↪trillion_available_seat_km'], color = 'gray')
plt.grid()
plt.title("Total Number of Fatal Accidents per One Trillion Available Seat␣
↪Kilometers between 1985 and 2014")
plt.show()
```



[31]:
```python
df = df.sort_values("total_fatalities per trillion_available_seat_km")

plt.figure(figsize=(16,12))
plt.barh(y = df["airline"], width = df['total_fatalities per trillion_available_seat_km'], color = 'gray')
plt.grid()
plt.title("Total Number of Fataltles per One Trillion Available Seat Kilometers between 1985 and 2014")
plt.show()
```

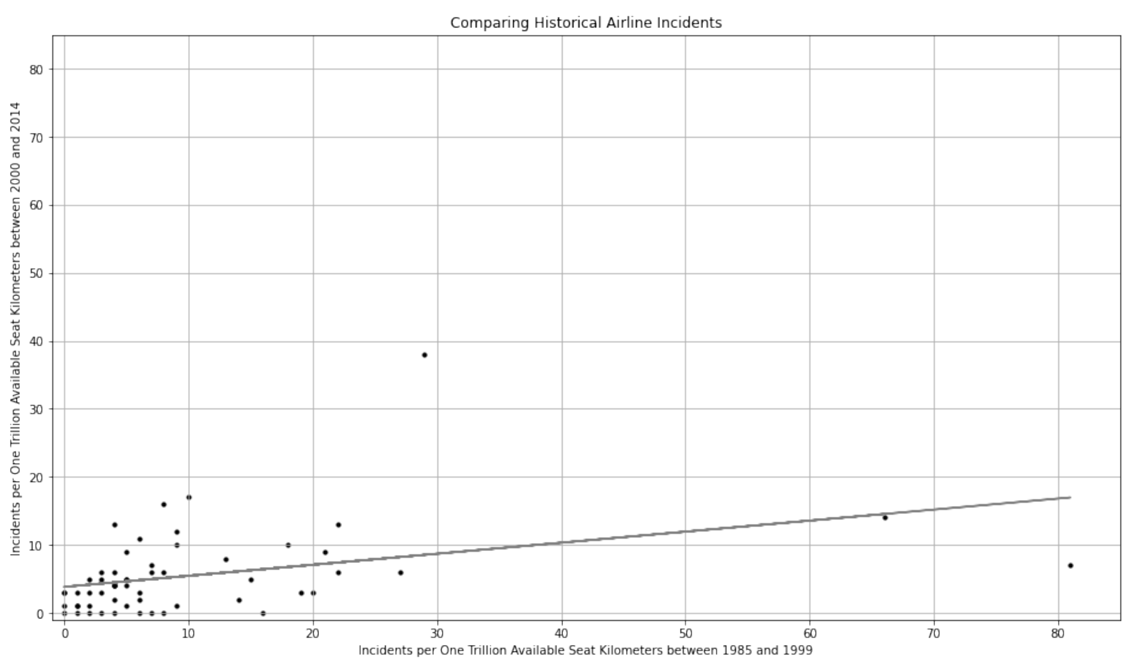


A question commonly asked is whether or not an airlines historical safety can predict how safe the airline will be in the future. Since we have two different time periods of data available, we can test this theory by plotting the data and calculating the regression.

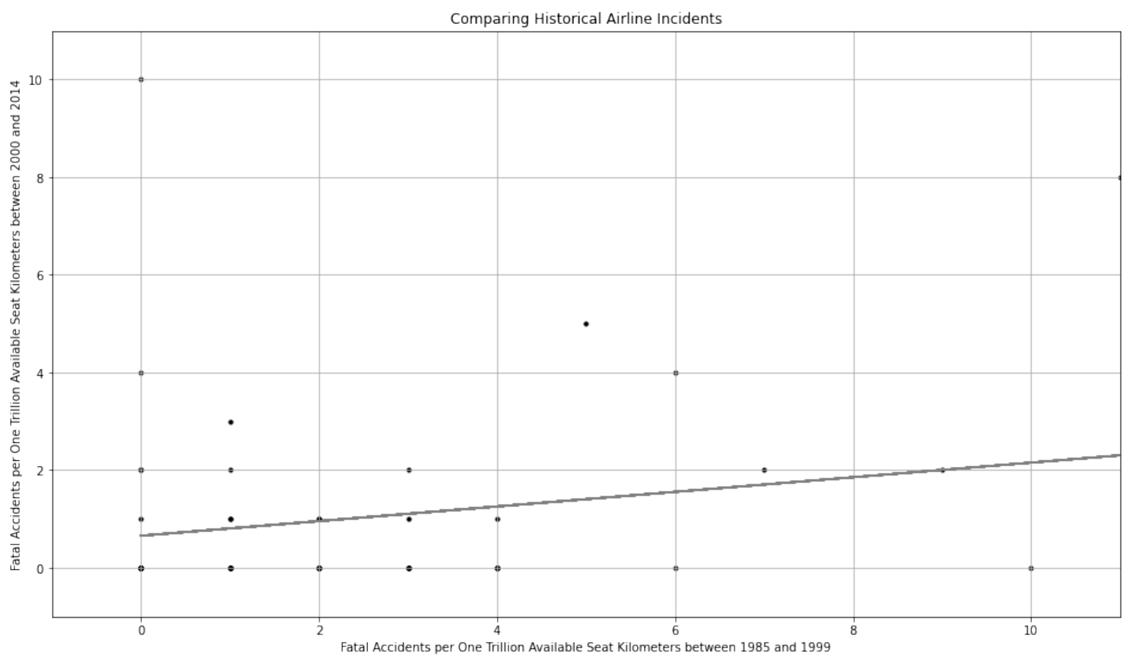
[34]:
```
df['fatailites_per_trillion_available_seat_km_85_99'] = np.
↪round(df['fatalities_85_99'] / df['trillion_available_seat_km_85_99'])
df['incidents_per_trillion_available_seat_km_85_99'] = np.
↪round(df['incidents_85_99'] / df['trillion_available_seat_km_85_99'])
df['fatal_accidents_per_trillion_available_seat_km_85_99'] = np.
↪round(df['fatal_accidents_85_99'] / df['trillion_available_seat_km_85_99'])

df['fatailites_per_trillion_available_seat_km_00_14'] = np.
↪round(df['fatalities_00_14'] / df['trillion_available_seat_km_00_14'])
df['incidents_per_trillion_available_seat_km_00_14'] = np.
↪round(df['incidents_00_14'] / df['trillion_available_seat_km_00_14'])
df['fatal_accidents_per_trillion_available_seat_km_00_14'] = np.
↪round(df['fatal_accidents_00_14'] / df['trillion_available_seat_km_00_14'])

df = df[['airline',
    'incidents_per_trillion_available_seat_km_85_99',
    'fatal_accidents_per_trillion_available_seat_km_85_99',
    'fatailites_per_trillion_available_seat_km_85_99',
    'incidents_per_trillion_available_seat_km_00_14',
```

```
    'fatal_accidents_per_trillion_available_seat_km_00_14',
    'fatailites_per_trillion_available_seat_km_00_14']]

df.to_csv("data2.csv")
```

[46]:
```
plt.figure(figsize=(16,9))

x = df['incidents_per_trillion_available_seat_km_85_99']
y = df['incidents_per_trillion_available_seat_km_00_14']

plt.scatter(x, y, s = 10, color = 'black')
plt.xlabel("Incidents per One Trillion Available Seat Kilometers between 1985␣
 ↪and 1999")
plt.ylabel("Incidents per One Trillion Available Seat Kilometers between 2000␣
 ↪and 2014")

m, b = np.polyfit(x,y,1)
plt.plot(x, m*x + b, color = "gray")

plt.title("Comparing Historical Airline Incidents")
plt.xlim(-1,85)
plt.ylim(-1,85)
plt.grid()
plt.show()
```

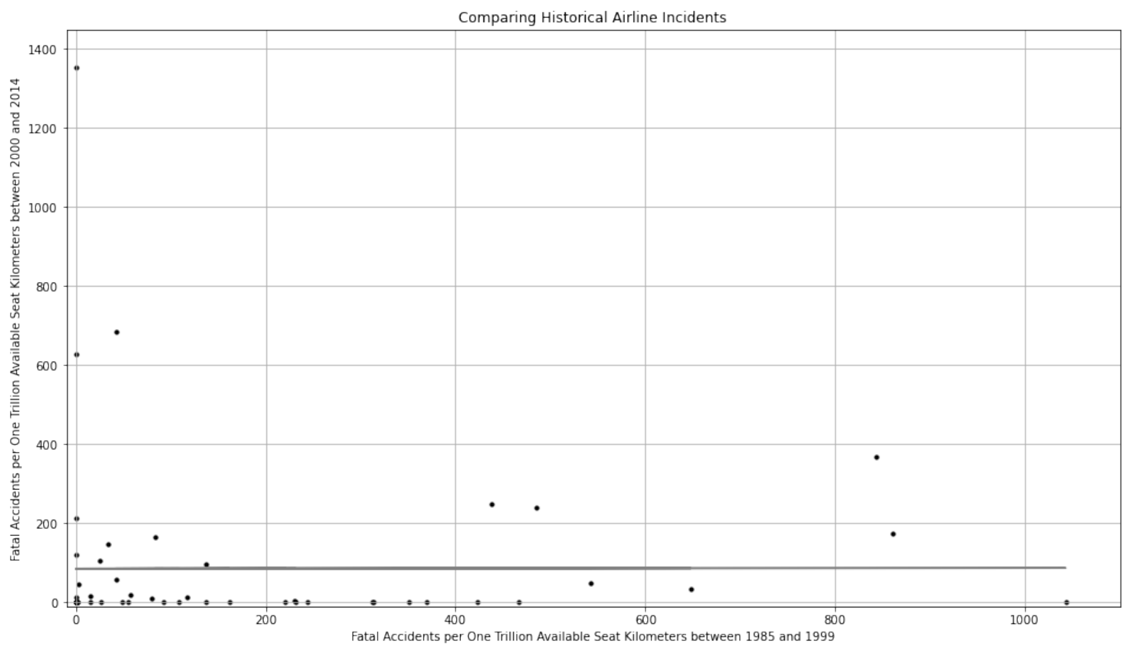


[48]:
```
plt.figure(figsize=(16,9))

x = df['fatal_accidents_per_trillion_available_seat_km_85_99']
y = df['fatal_accidents_per_trillion_available_seat_km_00_14']

plt.scatter(x, y, s = 10, color = 'black')
plt.xlabel("Fatal Accidents per One Trillion Available Seat Kilometers between␣
↪1985 and 1999")
plt.ylabel("Fatal Accidents per One Trillion Available Seat Kilometers between␣
↪2000 and 2014")

m, b = np.polyfit(x,y,1)
plt.plot(x, m*x + b, color = "gray")

plt.title("Comparing Historical Airline Incidents")
plt.xlim(-1,11)
plt.ylim(-1,11)
plt.grid()
plt.show()
```



[51]:
```
plt.figure(figsize=(16,9))

x = df['fatailites_per_trillion_available_seat_km_85_99']
y = df['fatailites_per_trillion_available_seat_km_00_14']
```

```
plt.scatter(x, y, s = 10, color = 'black')
plt.xlabel("Fatal Accidents per One Trillion Available Seat Kilometers between␣
 ↪1985 and 1999")
plt.ylabel("Fatal Accidents per One Trillion Available Seat Kilometers between␣
 ↪2000 and 2014")

m, b = np.polyfit(x,y,1)
plt.plot(x, m*x + b, color = "gray")

plt.title("Comparing Historical Airline Incidents")
plt.xlim(-10,1100)
plt.ylim(-10,1450)
plt.grid()
plt.show()
```



From these graphics, we can see a very loose correlation between the historical number of incidents correlating with the future number of indicents, but the number of histoical fatalities per airline has no correlation.

Lastly, we can investigate how much safer vehicles are compared to airplanes. The DOT has a dataset containing all of the US related traffic data including data for fataltíes, injuries, and crashes. We can try to generate a similar metric to one trillion avaialble seat kilometers for vehicles by using the data in the dataset.

[54]:
```
cars = pd.read_excel("motor-vehicle-safety.xlsx", header=1, index_col=0,␣
 ↪na_values="N").head(4).T
```

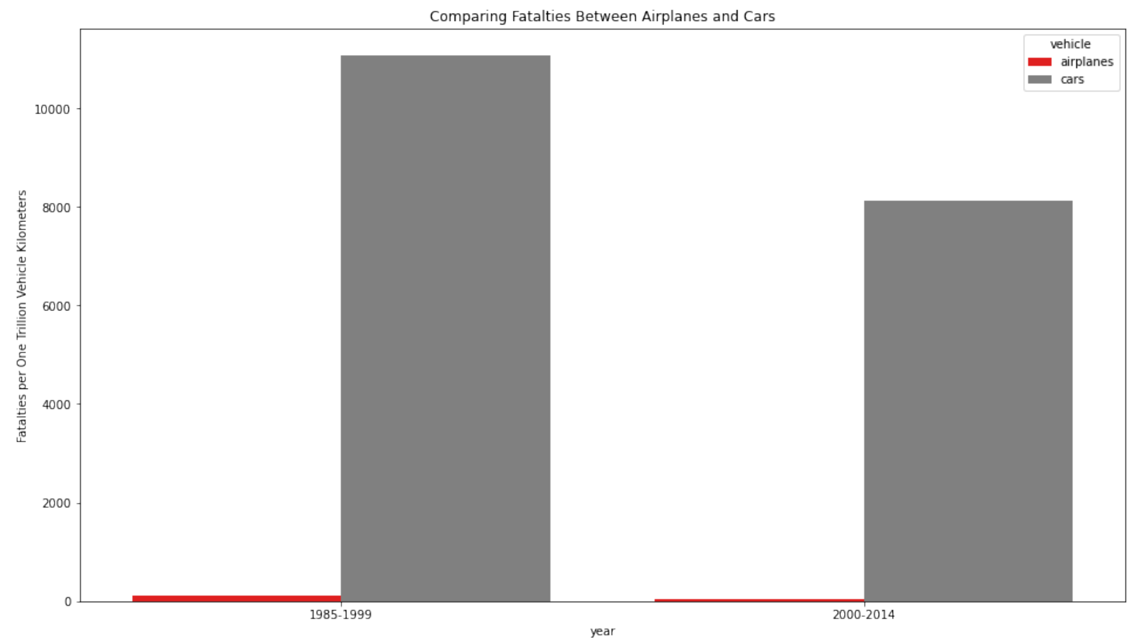
```
cars['Vehicle-km (Trillions)'] = cars['Vehicle-miles (millions)'] * 1.60934 /
↪1000000
cars['Fatailites per Trillion Vehicle-km'] = cars['Fatalities'] /␣
↪cars['Vehicle-km (Trillions)']
cars['Crashes per Trillion Vehicle-km'] = cars['Crashes'] / cars['Vehicle-km␣
↪(Trillions)']
cars['Year'] = cars.index.astype(int)
cars.reset_index()
cars[['Vehicle-km (Trillions)',
      'Fatalities',
      'Injured persons',
      'Crashes',
      'Crashes per Trillion Vehicle-km',
      'Fatailites per Trillion Vehicle-km']]

arr = []

for i in cars['Year']:
    if 1985 <= int(i) <= 1999:
        arr.append("85_99")
    elif 2000 <= i <= 2014:
        arr.append("00-14")
    else:
        arr.append("nan")
```

[89]:
```
cars['grouping'] = arr

tmp = pd.DataFrame(cars.groupby('grouping')['Fatalities'].sum() / cars.
↪groupby('grouping')['Vehicle-km (Trillions)'].sum())
tmp['vehicle'] = ['cars','cars','cars']
tmp['year'] = ['2000-2014','1985-1999','nan']
tmp2 = pd.DataFrame(fptskm)
tmp2['vehicle'] = ['airplanes','airplanes']
tmp2['year'] = ['1985-1999','2000-2014']
x = pd.concat([tmp2,tmp])
x.to_csv("data3.csv", header = 0)
x.columns = ['value','vehicle','year']
x = x[x['year'] != "nan"].reset_index()
x = x[['year','vehicle','value']]
x
```

[89]:

	year	vehicle	value
0	1985-1999	airplanes	104.083550
1	2000-2014	airplanes	53.177797
2	2000-2014	cars	8120.514838
3	1985-1999	cars	11071.457342

[111]:
```
plt.figure(figsize=(16,9))
sns.barplot(x = x['year'], y = x['value'], hue = x['vehicle'], dodge=True,␣
 ↪palette=['red','gray'])
plt.title("Comparing Fataltties Between Airplanes and Cars")
plt.ylabel("Fatalties per One Trillion Vehicle Kilometers")
plt.show()
```



From this graphic, we can see that travelling via airplane is orders of magnitude safer than driving. This is likely due to the fact that airlines follow strict safety protocols and undergo routine maintenance every time before a flight.

1.0.1 References

Bureau of Transportation. (n.d.). *Motor Vehicle Safety Data*. Retrieved Apr 11, 2021, from Bureau of Transportation: https://www.bts.gov/content/motor-vehicle-safety-data

Silver, N. (2014, Jul 18). *Should Travelers Avoid Flying Airlines That Have Had Crashes in the Past?* Retrieved Apr 11, 2021, from FiveThirtyEight: https://fivethirtyeight.com/features/should-travelers-avoid-flying-airlines-that-have-had-crashes-in-the-past/