DATA EXPLORATION PROJECT

Name: Aman Roy, **Student ID:** 35011203, **Teaching Associate:** Maxim Srour.

Project Title: Comparison of fuel consumption between planes and cars

Contents

Table of figures

Section 1: Introduction

Problem Description

There has been a recent trend to develop more electric vehicles in order to reduce fuel consumption and emissions, although developing electric vehicles is a step in the right direction, there are many other modes of transport that consume large amount of fuel and emit more co2 gas that a single car could do in a whole month, specifically planes where the amount of fuel consumption and emission is so high that cars could run for months given the same allocation of fuel. This is why I want to compare cars and planes on different parameters to assess the difference between them.

Motivation

I have a keen interest in cars and I follow most model manufacturers, I the last few years many car manufacturers have stopped production of their ice-vehicles, while the main focus is on most four wheeled vehicles, different modes of transport like planes have been completely ignored which consume high quantity of fuels. This is the reason I want to compare them on the basis of different parameters.

Questions

1. How much fuel does a single flight consume and how does it compare it to a car, which mode of transport is more efficient when compared against parameters like cost and cargo capacity.
2. How much distance is travelled by a plane or a car in their life time, how does it compare when you consider all the cars vs all the planes in the world for their fuel consumption and co2 emissions, which mode is more economical?

Section 2: Data wrangling/checking

1. 2022 Fuel Consumption Ratings

This dataset shows tabular data in csv with ~1K rows x 15 columns. This dataset provides model-specific fuel consumption ratings and estimated carbon dioxide emissions for new light-duty vehicles for retail sale in Canada in 2022.

URL: https://www.kaggle.com/datasets/rinichristy/2022-fuel-consumption-ratings

It did not have any null/missing values but had some outliers which were found by calculating the mean and standard deviation of the dataset, with the threshold set to 3 the z-score was calculated and any point exceeding this threshold was dropped, in total 7 such outliers were dropped out.

2. Petrol/Gas Prices Worldwide

This dataset shows the fuel prices world wide for year 2022, containing columns like price per litre for different countries in the world, mostly we would like to focus on the price for certain countries and could be used in comparison analysis.

URL: https://www.kaggle.com/datasets/zusmani/petrolgas-prices-worldwide

This dataset did not have any null values or duplicate values, because it is showing prices for different countries, no outlier detection was needed for this dataset.

3. Australian Vehicle Prices

This dataset shows information about the cars that were available in Australia during 2023, it has information about fuel type, distance in kilometres, fuel consumption and price. It contains information about all the different car manufacturers and model names.

URL: https://www.kaggle.com/datasets/nelgiriyewithana/australian-vehicle-prices/data

This dataset had many columns with null values so I calculated its percentage and dropped the rows having less than three percent null values and for the columns which had high number of null values I replaced it with its mode, there were 28 duplicate values in the dataset so I also dropped them. There were few outliers in the dataset to find them I used the IQR method, they were removed from the dataset

4. Domestic Airlines - Top Routes and Totals

This dataset shows the information on flight national flight paths inside Australia via Regular Public Transport (RPT) air service, data contains city pair, month for passengers carried, aircraft trips, distance between two airports, revenue passenger per kilometre and available seat kilometres and seats

URL: https://www.data.gov.au/data/dataset/domestic-airlines-top-routes-and-totals

This dataset had no duplicate values but had some null values for which I have replace them with the mode of that particular column, as this dataset is primarily going to be used to show the distance between airports of Australia, I decided to look at some outliers in the dataset using IQR aid z-score method and replaced them with the median of that column.

5. Airline Fuel Cost and Consumption

This dataset contains information about all major airlines fuel consumption rates for every month, it also contains international cost, cost per gallon of fuel provided for aircrafts in each subsequent year.

URL: https://www.transtats.bts.gov/fuel.asp?20=E

This dataset had a few columns with duplicate values, which were replaced with the mode of that column, there were no duplicate values in the dataset, as this dataset contained prices for fuels I didn't find any significant outliers in the dataset.

6. Fuel Consumption of the 50 Most Used Passenger Planes

This dataset shows the fuel consumption and co2 emissions for 50 of the most used passenger planes in the world. It includes various sheets that cover an overview of the dataset, specific metrics like Specific Absorption Rate (SAR), payload diagrams, reliability curves, emission calculations, and methodologies. It has both fuel consumption and co2 emission attributes for all the different planes.

URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/4CYNKA

This dataset contained some missing values which were replace by the mean of the column, apart from that there were no duplicate values in the dataset, there were no significant outliers present in the dataset, this dataset contains normalized values of fuel consumption per kilometre per kilogram.

7. Airplanes Avfare 2024 Q1

This dataset shows average ticket fare price for different American airports for 2024 Q1

URL: https://www.transtats.bts.gov/AverageFare/

This dataset contained no missing values, no duplicate values and no outliers

Section 3: Normalization

To compare and do visual analysis I had to first convert cars dataset to similar metrics as planes dataset as the plane's dataset was already normalized

For dataset 1 fuel conversion:

- Fuel consumption Conversion (L/100 km to kg/km/seat)
- 1L of petrol ≈ 0.74kg approx.
- Fuel consumption (kg/km) = Fuel consumption Conversion (L/100 km) x 0.74 / 100
- For seat I assume that all cars have minimum 5 seats available
- Fuel consumption (kg/km/seat) = Fuel consumption (kg/km)/seats

There was no need for normalization of co2 emission as it was already in g/kg which was alright for comparison

For fuel prices for airplanes, I assumed that from dataset 11 the average cost per passenger for a standard journey was represented if the name of the city was the origin of the flight and joined it with dataset 6

For cars I joined dataset 3 with 1 and normalized by extracting the columns fuel consumption and seats then I calculated the total fuel consumption for a 500km journey using this formula

- Total fuel consumption = fuel consumption(L/100km) x distance(km)/100
- I calculated the travel cost by assuming fuel price to be 1.84 AUD [12]
- Travel cost (AUD) = Total fuel consumption X fuel price (1.84 AUD)
- Then I normalized the travel cost per seat by dividing the total cost by the number of seats in the vehicle:
- Car cost per seat = travel cost / seats

Section 4: Data Exploration

All the visualizations were created in R

To answer the first question, first we need to filter our datasets bases on some parameters, as there are over 50 different planes and over 700 unique car models, we cannot compare them all against each other, so first let's try to find the most fuel-efficient planes and cars from our datasets.

First look at the amount of fuel planes and cars consume per kilometre per seat

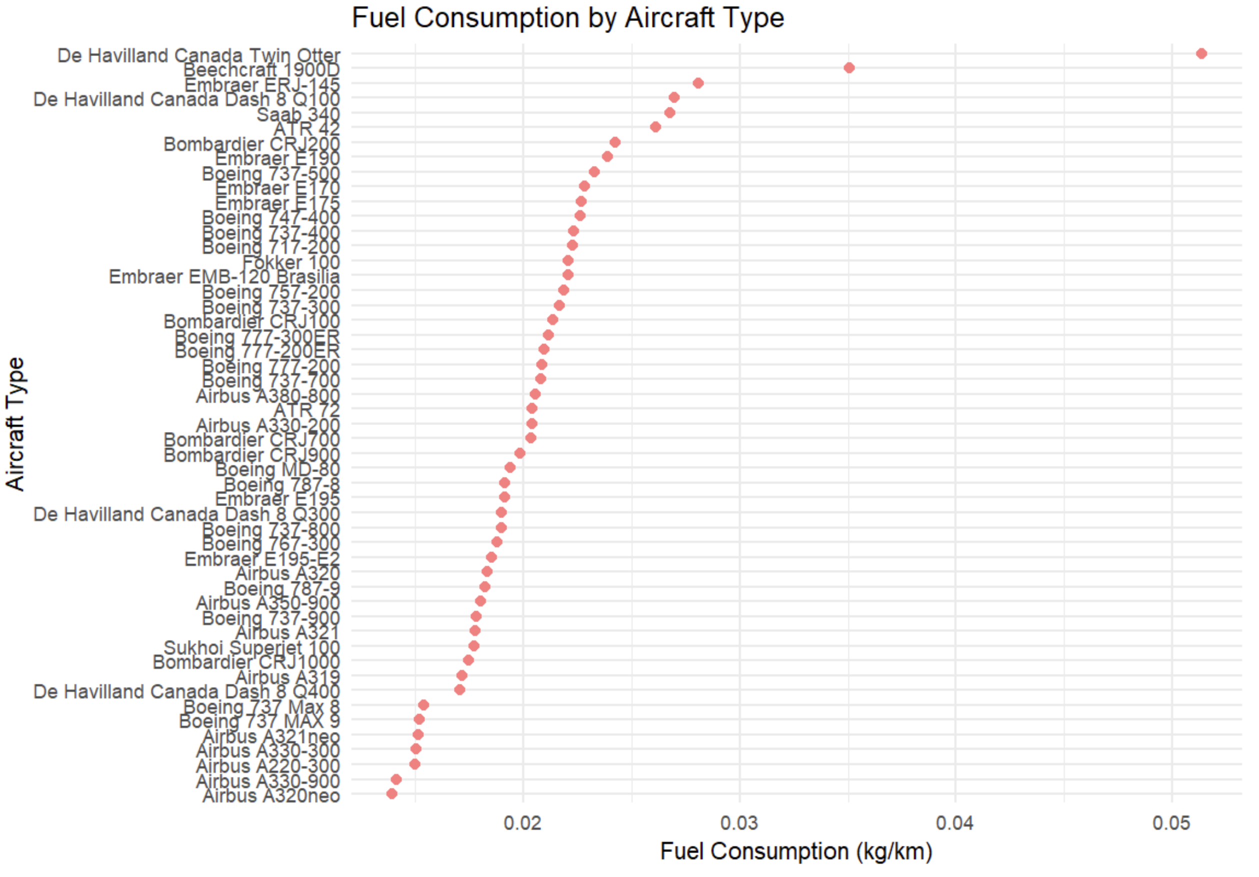

Figure 1

Judging by this graph we see that there are some planes that are very fuel efficient and others not as much, with the help of this graph we can select the five planes that consume the least amount of fuel per kilogram

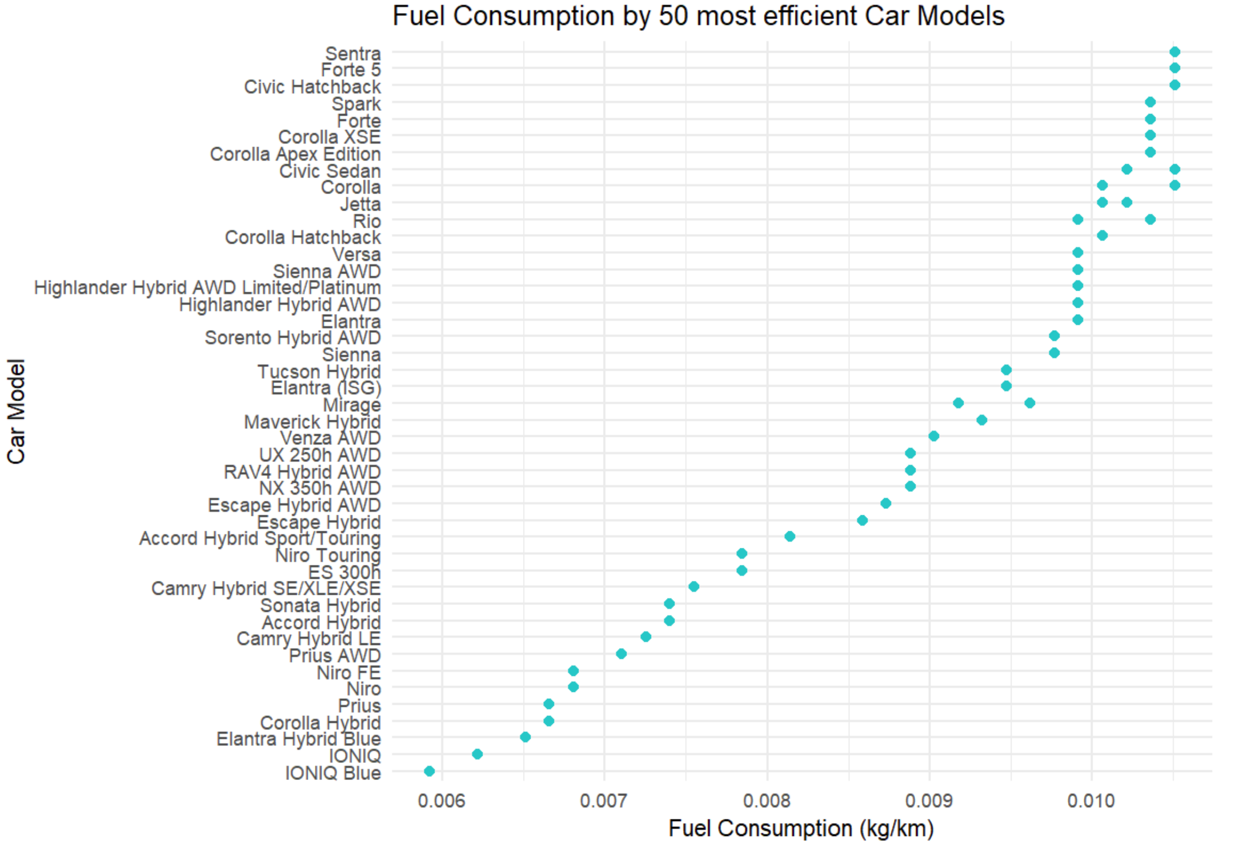

Figure 2

In the cars dataset we had a large number of cars so to make the graph more accessible we chose the 50 most fuel-efficient cars from the whole dataset and compare them, here we can see that there is a lot more variability in this graph

Now we want to compare these vehicles against each other while comparing cost per seat and fuel consumption

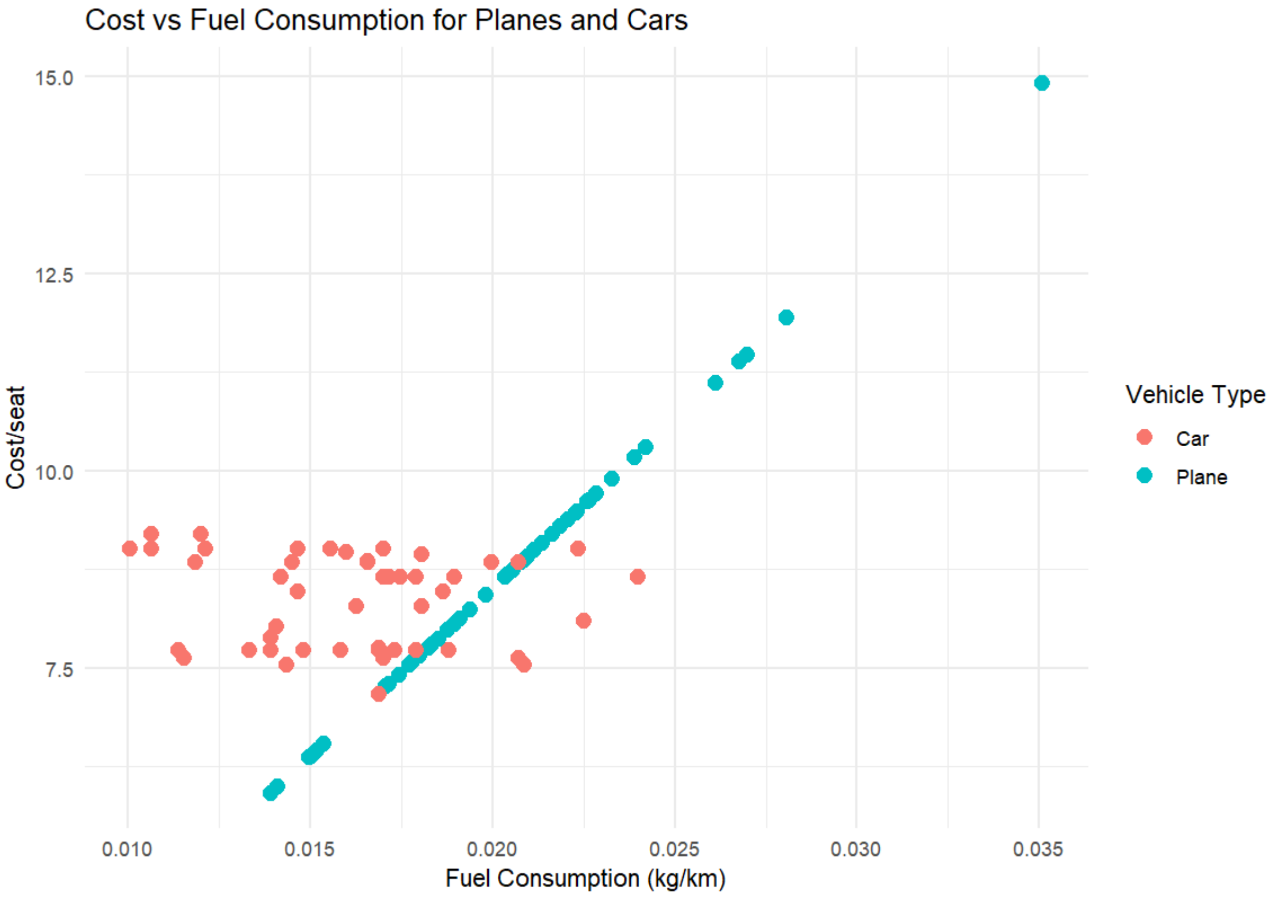

Figure 3

This figure (3) show that there is a positive corelation between the fuel consumption and cost/seat in the planes data which is pretty consistent throughout where as the cars are situated in a specific area when looking at cost, we can see that majority are between 7.5 to 9 dollars per seat and even the highest fuel consumption being around 0.025 kg/km. This shows that while there are some aircrafts which are more efficient than cars it also scales up in a clear direction.

Now I also want to check how much carbon is emitted when compared against the amount of fuel consumed by the top 50 planes and cars from figure (4) we can see that the cars have a higher carbon emission rate when compared to planes but that is because the data used for this graph is calculated for each seat inside that vehicle and the maximum seats in a car can only be 7 but for planes it can go as high as 500 with average seat in a plane being 274. The insight that we get from figure (4) is that planes are more efficient in co2 emission when compared to cars.

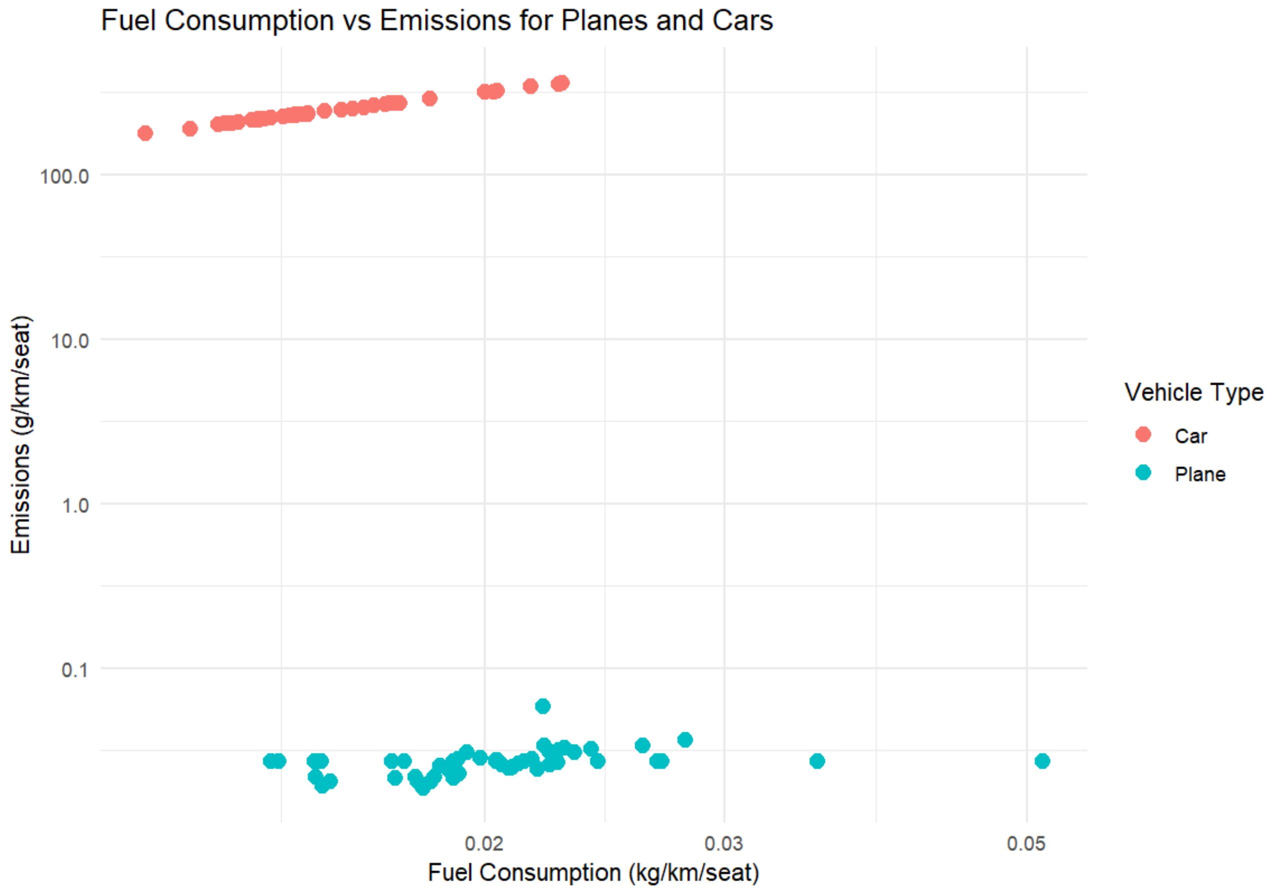

Figure 4

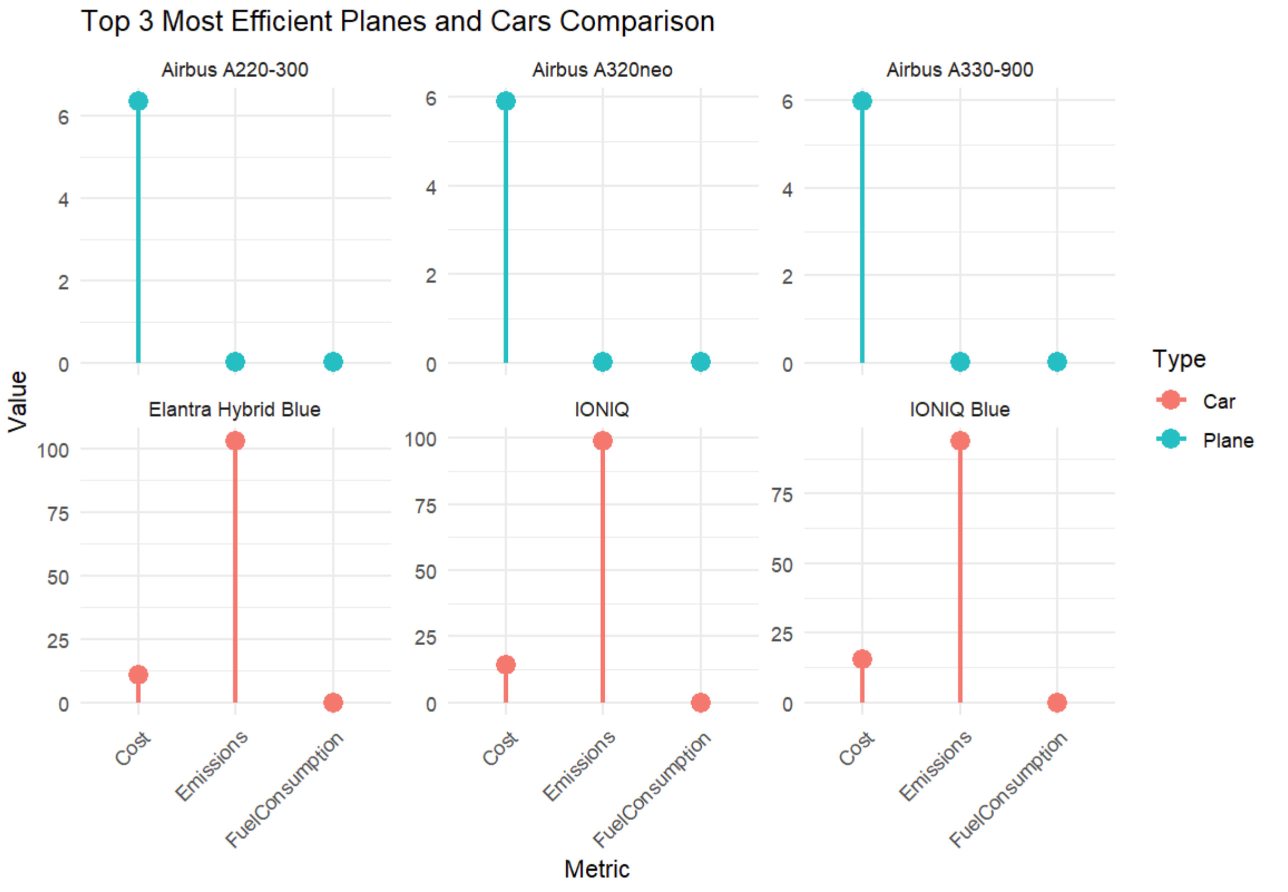

Figure 5

Now with the help of figure (5) we can find the top 3 most efficient cars and planes when compared with these metrics, surprisingly all the planes are an airbus models while all the cars were from Hyundai company.

Judging by the above figures you may think that it's better to travel by air, I mean it's more cost effective, gives out less emissions and consumes less fuel than the cars but that's not true at all this is because the unit conversion for all the values in these figures in kg/km/seat which means for a single person in the plane and car, keeping in mind that a typical plane carries on average around 150 passengers a car on average 5 and in total a car travels around 200,000 miles on average[8] and a plane around 52 million miles in its lifetime[9]   .

Keeping that in mind lets compare how much distance does an average plane or car travel in their lifetime

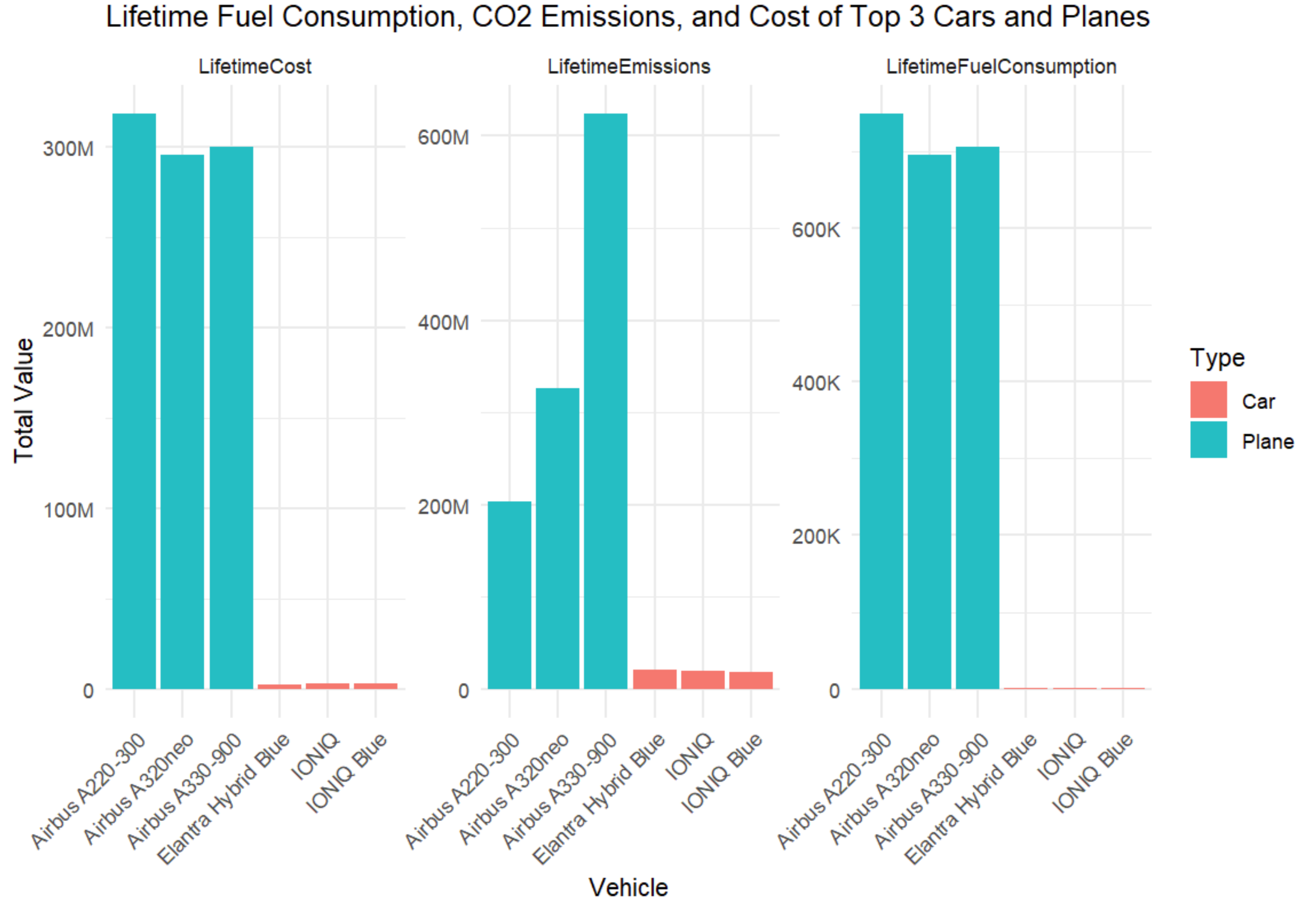

Figure 6

Looking at figure (6) we can say that there is a huge difference between the cost, fuel consumption and co2 emission between cars and aircrafts, an average aircrafts total value goes very high when compared to cars and while for cars it looks like a very small value those numbers are still pretty high.

This still does not give us the full picture as the difference between the number of planes and cars in the world is astronomically high, there are around 25000 active planes in the world [10] and around 1.27 billion internal combustion engines in the world [7] and when we take that in account, we get figure (7)

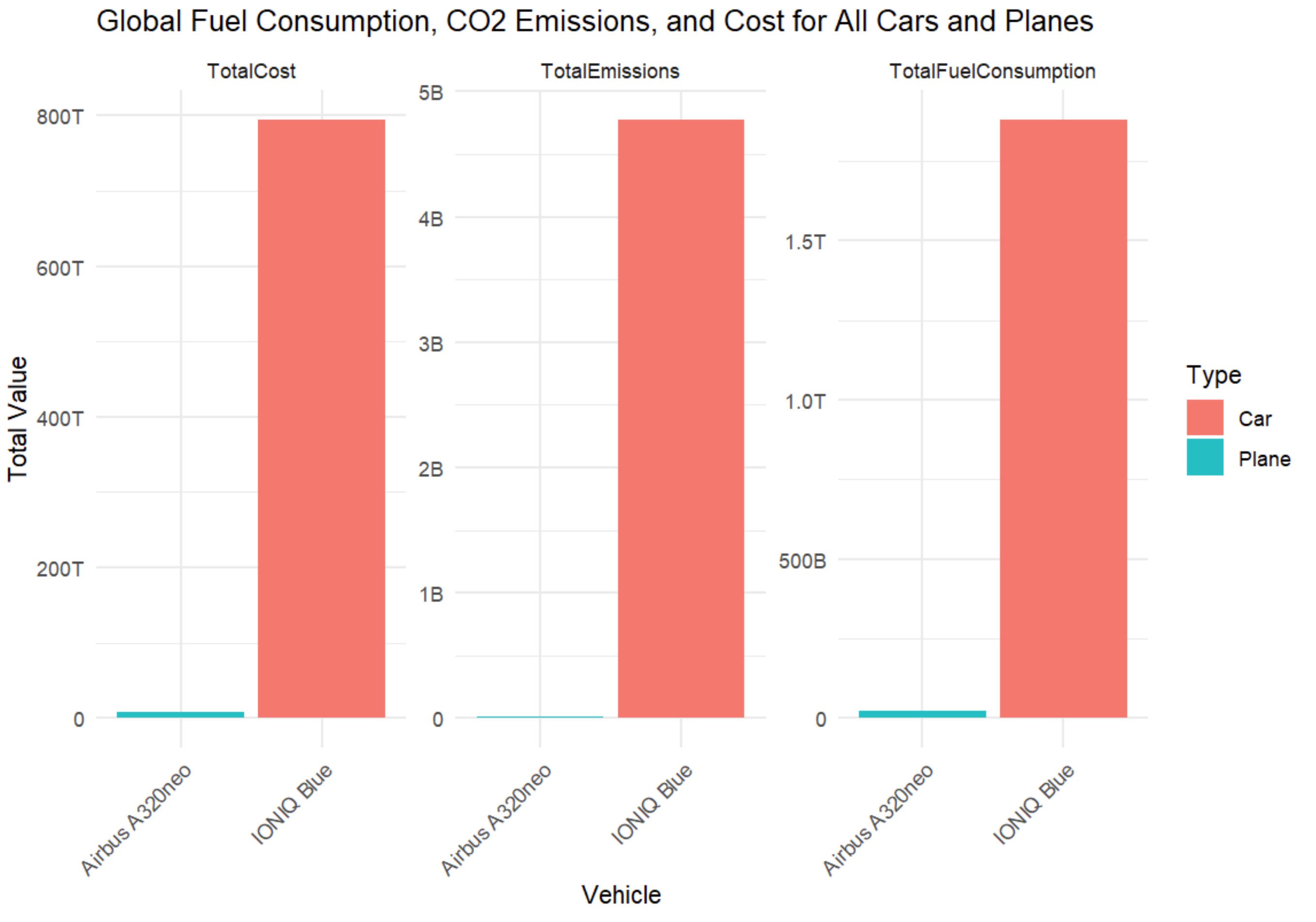

Figure 7

As you can see due to the sheer number of cars in the world, it clearly out shadows the impact caused by planes throughout the world.

Section 5: Conclusion

To answer the first question the mode of transport which is better in terms of environment and economy its cars as even though there were some planes which were more economical than cars the chances of you getting on that one is pretty slim, whilst a person can choose the car which they want to drive, from figure(3) we can see that cars are clustered around in a zone whereas planes have a positive corelation which goes low and high as well.

For the second question for a single vehicle the amount of distance travelled by it in its lifetime the metrics will always look bad if we compare a car and a plane because for obvious reasons a plane will have travelled more distance than a car and so it will have higher values of cost, emission and fuel consumption and thus make car more economical than even the most efficient plane in the market, but when we consider all the planes and cars in the world the difference is quite huge and clearly aircrafts fare much better as their amount is so much smaller that cars.

Section 6: Reflection

During this project I learned about various methods for data representation, gather insights and wrangle data join different parts of datasets and normalize them.

I think I should have used some more complex visualizations but I believe that the graphs and charts I used are good enough that anyone could look at them and instantly be able to understand the meaning behind it and also follow the trends that the graphs are trying to show

I tried to visualize maps where I would compare different journeys from Melbourne to other cities and show those metrics as well but in a static graph they didn't look as good as I thought they would so I decided not to add them thinking a better implementation could be done in the DVP.

I was also not able to include the parameter called cargo space because the datasets which I found were not suitable enough to normalize and join with the main dataset, cargo capacity for the car dataset was only in cubic feet and the airplane dataset had max take-off weight and max-landing weight which I could not successfully convert them in the correct format for them to make sense, so I decided not to use them.

After working on this project, the most important thing I understood how much value conversions and normalization are important for your data and having any mistake and significantly affect the result. I also tried to keep the proper and continuous colour scheme and tried not to shrink the graphs too small. I learned a lot from this project but I still believe there were things that could have been done better

Section 7: Bibliography

[1] Kühn, Marius, 2024, "Fuel Consumption of the 50 Most Used Passenger Aircraft", https://doi.org/10.7910/DVN/4CYNKA, Harvard Dataverse, V3

[2] Domestic Airlines - Top Routes and Totals. (2016). data.gov.au. Retrieved August 4, 2024, from https://www.data.gov.au/data/dataset/domestic-airlines-top-routes-and-totals

[3] Christy, R. (2022). 2022 fuel consumption ratings. Kaggle. Retrieved August 4, 2024, from https://www.kaggle.com/datasets/rinichristy/2022-fuel-consumption-ratings

[4] Airline Fuel Cost and Consumption (U.S. Carriers - Scheduled). (2024, May). Bureau of Transportation Statistics. Retrieved August 4, 2024, from https://www.transtats.bts.gov/fuel.asp?20=E

[5] Petrol/Gas prices worldwide. (2022, June 24). https://www.kaggle.com/datasets/zusmani/petrolgas-prices-worldwide/data

[6] Australian vehicle prices. (2023, November 27). https://www.kaggle.com/datasets/nelgiriyewithana/australian-vehicle-prices

[7] Marc, O. (2023, September 16). *"X-Change: Cars" – just another report predicting the extinction of the "ICE Age."* Autoevolution. https://www.autoevolution.com/news/x-change-cars-just-another-report-predicting-the-extinction-of-the-ice-age-221158.html

[8] Corporation, P. (2023, February 2). *How many miles does a car last?* https://www.progressive.com/answers/how-many-miles-does-a-car-last/

[9] Today, J. C. S. F. U. (2012, November 19). Ask the Captain: How far does a jet fly during its lifetime? *USA TODAY*. https://www.usatoday.com/story/travel/columnist/cox/2012/11/19/ask-the-captain-how-far-does-a-jet-fly-during-its-lifetime/1712269/

[10] Statista. (2024, April 18). *Projected aircraft fleet size 2020-2034*. https://www.statista.com/statistics/282237/aircraft-fleet-size/

[11] *Average domestic airline itinerary fares*. (n.d.). https://www.transtats.bts.gov/AverageFare/

[12] Edlington, A. (2024, March 22). Petrol prices around Australia – 2023 report. *Savvy*. https://www.savvy.com.au/car-loans/petrol-prices-around-australia/