A visual study of passenger car vehicle configurations and their impact on fuel economy and CO2 emissions in Canada

March 2021

CMS7505-2021

Carl Wilson [U0370630]

u0370630@unimail.hud.ac.uk

1. INTRODUCTION

Transportation is a significant contributor to global greenhouse gas [GHG] emissions, accounting for around 44% of total global GHG's in 2017 (Mengpin Ge, 2020) of which CO2 is a major portion [around 74%] due to burning fossil fuels. Fuel economy is a measure distance covered for fuel used, with higher fuel economy allowing a vehicle to travel further for the same fuel burn. Thus, CO2 emissions are inextricably linked to vehicle fuel economy, the higher the fuel economy – the lower the CO2 emissions.

Passenger vehicles are a consumer commodity item, with a wide variety of options available to the customer. These options create a diverse spectrum of vehicles with equally varied fuel consumption figures and thus CO2 emissions.

The purpose of this report will be to visually examine some of the key choices available to the consumer in selecting a passenger car vehicle with respect to fuel consumption and CO2 emissions; empowering the customer as a key stakeholder to make informed decisions regarding vehicle configuration when considering fuel economy and CO2 emissions.

2. THE DATA

The "CO2 Emissions by Vehicles" dataset was taken from Kaggle (Podder, 2020) and contains a mix of categorical and continuous variables relating to passenger car configurations, such as: body type, engine size, number of cylinders and type of transmission and fuel consumption and CO2 emissions for new model vehicles in Canada over a seven year period. A complimentary file is also available which provides additional information on interpreting the data, such as *fuel type* coding's.

Categorical Variables: Make, Model, Vehicle Class, Cylinders, Transmission and Fuel Type.

Continuous Variables: Engine Size(L), Fuel Consumption City (L/100 km), Fuel Consumption Hwy (L/100 km), Fuel Consumption Comb (L/100 km) Fuel Consumption Comb (mpg) CO2 Emissions(g/km).

3. DATA CLEANING & PRE-PROCESSING

The data was complete and contained no missing values. The Transmission and Fuel Type variables required some further pre-processing.

```
In [3]:  1  MASTER_IMPORT.head()
Out[3]:
```

	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	CO2 Emissions(g/km)
0	ACURA	ILX	COMPACT	2.0	4	AS5	Z	9.9	6.7	8.5	33	196
1	ACURA	ILX	COMPACT	2.4	4	M6	Z	11.2	7.7	9.6	29	221
2	ACURA	ILX HYBRID	COMPACT	1.5	4	AV7	Z	6.0	5.8	5.9	48	136
3	ACURA	MDX 4WD	SUV - SMALL	3.5	6	AS6	Z	12.7	9.1	11.1	25	255
4	ACURA	RDX AWD	SUV - SMALL	3.5	6	AS6	Z	12.1	8.7	10.6	27	244

To help make the *Fuel Type* variable more meaningful the *Fuel Type* code was converted to something more relatable such as *Premium Gasoline*. Similarly, the *Transmission* variable which was provided as transmission type and number of gears within the transmission was split into two categorical variables, one for the number of gears known as *Gears* and one for the type of transmission such as *Manual* or *Automatic* which was stored under *Transmission*.

4. RESULTS

For presenting the relationship between multiple continuous variables, scatter plots can be used, with additional axes in the form of marker colours, sizes and shapes to add additional categorical variables. It is recommended to only capture one additional categorical variable per chart – as too many can start to create a cluttered chart with too much information for the observer to make use of (Knaflic, 2015).

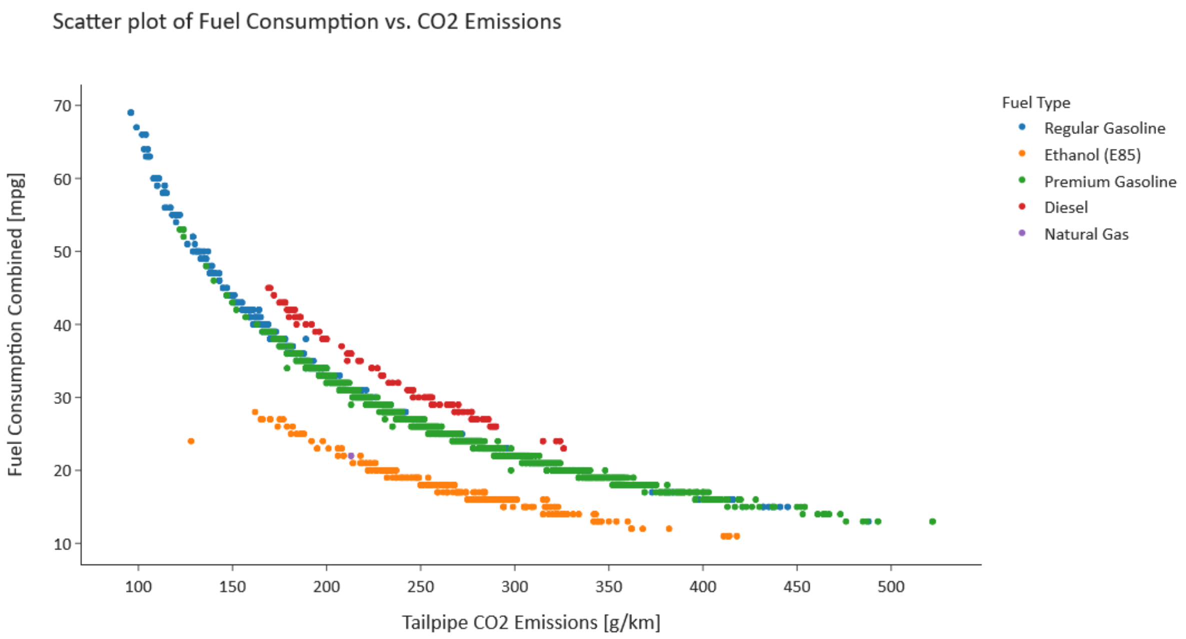

Fig1: Scatterplot of C02 Emissions vs Fuel Consumption

Fig1 illustrates the correlation between fuel consumption and CO2 emissions for all vehicles in the dataset. The colours represent the different fuel types. Using the colour axis in this way shows the clear difference between the three major fuel types: Gasoline, Ethanol and Diesel. The diesel vehicles clearly have the higher fuel consumption for the same CO2 emissions. The colour axis in this instance is very useful, however can only be used for categorical variables with few unique values [in this instance: 5], otherwise the plot becomes busy and unreadable (Iliinsky & Steele, 2011). For example, using the *Vehicle Class* variable which has 16 unique values – the chart would not be as useful.

For multiple categorical variables against a continuous variable, boxplots can be used – nicely showing the distribution of a continuous variable against said categorical variable. With a categorical variable on the x-axis, a continuous on the y-axis and the colour axis again used for a categorical variable – a lot of information can be easily conveyed through a boxplot. *Fig 2* illustrates the fuel consumption for different vehicle classes. In this instance a second categorical variable, using the colour axis has not been used; as to do so would make the plot busy to the point where it would not be useful – this is in keeping with using colour sparingly (Knaflic, 2015).

A third type of useful visualisation is a heatmap, where either up to three continuous variables or two categorical variables and a continuous variable can plotted to create a grid, with a continuous colour gradient for the colour or Z-axis.

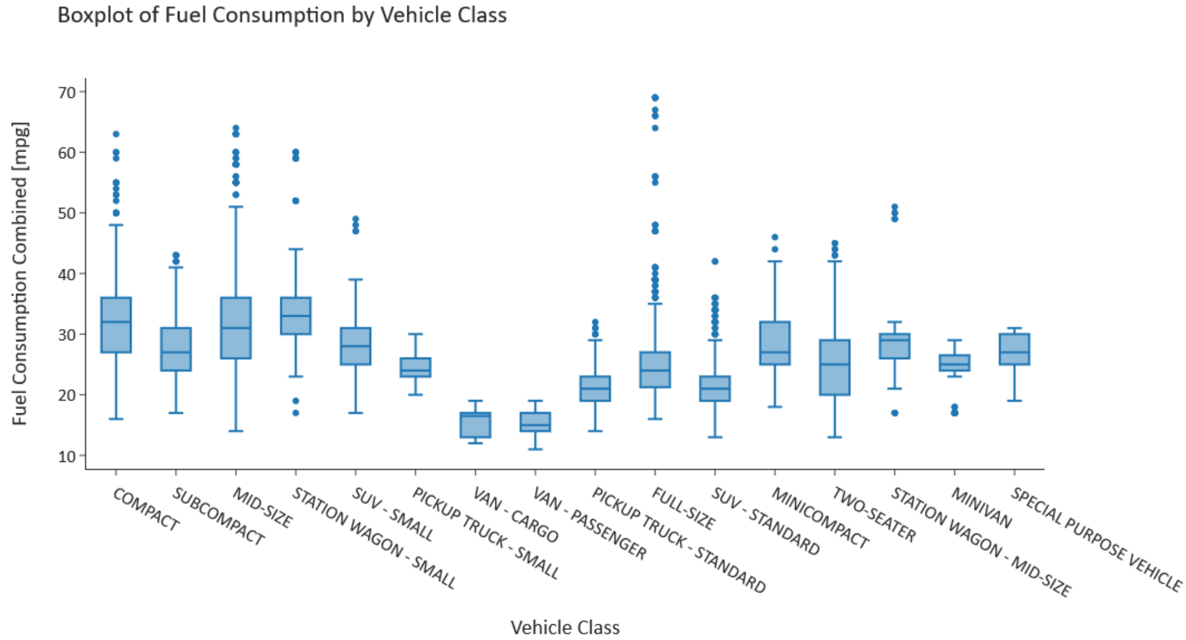

Fig2: Boxplot of Fuel Consumption by Vehicle Class.

In *Fig 3, Engine Size* and *Transmission Type* have been plotted against each other with the colour gradient being set to the fuel consumption. The aggregation method used here was "max" as this will allow the consumer to see the maximum attainable fuel economy for a given configuration.

It nicely illustrates that smaller engines have higher fuel consumption figures and that Continuously Variable and Automated Manual transmissions also appear to have the best fuel economy for a given engine size.

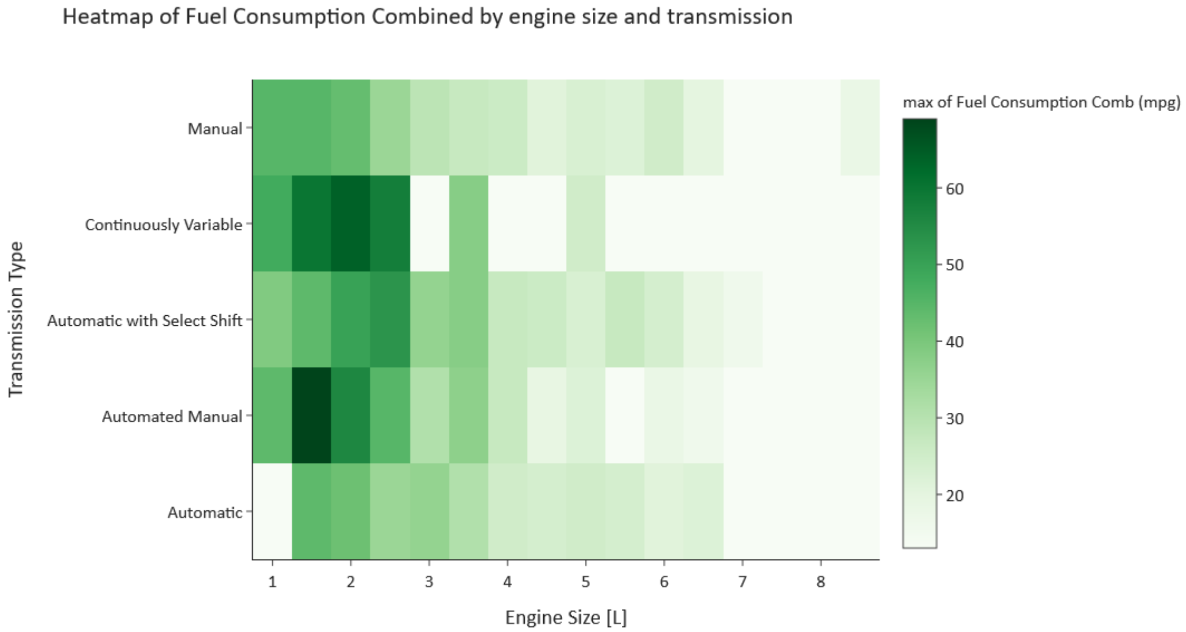

Fig3: Heatmap of fuel economy by engine size and transmission type.

By varying the saturation level on the colour axis for *Fig3* the eye is naturally drawn to the combinations for best fuel consumption (Knaflic, 2015). This is because saturation naturally carries some level of quantification, however different colours would not – for example using a full colour spectrum. If a full spectrum was to be used, we might assume red is *bad* and green is *good* but would struggle with blues. This is because we don't naturally have an order for the colour spectrum (Iliinsky & Steele, 2011). Different colours also mean different things to different people of varying countries and cultures.

Each of the plots have specifically been designed to have minimal clutter in terms of gridlines, borders and labels. The background colour is specifically the same as the page, reducing the burden on the observer (Knaflic, 2015) and allowing the critical information to stand out against the page for the observer.

5. CONCLUSION

The solutions presented within this report clearly and succinctly allow the customer to understand which fuels will give the highest fuel economy and the trade-off with CO_2 emissions, with diesel offering the highest fuel economy for a given CO_2 emission figure and natural gas giving the lowest. The customer can examine which types of vehicle offer the highest fuel economy and that there are economical options across most vehicle classes. It finally shows the types of transmission which offer the greatest fuel economy and that smaller engine sizes give the highest fuel economy and thus lowest CO_2 emissions.

These charts allow the customer to make an informed choice about the type and configuration of vehicle they would like in order to maximise fuel economy and minimise the CO_2 emissions and contribution to global GHG emissions.

6. REFERENCES

Iliinsky, N. P. N., & Steele, J. (2011). *Designing data visualizations*. O'Reilly. http://hud.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE9Is05KTk5NMUkHTSGbGBkZJyQYpFiaJZinA-tQ8CXw7kKWfr6WPo0lIOBODC2xrDHjGELpMUQ9WVsLnUjNKU6ArLfUguwDsU8Als7kFMJmZMwN7YgZGLAx8JiYeoS7xwM5wSLyzYwh4d5eJpTHo3BhT2KFPMD43A3dxfiLQXBMTIIFU07gJMrCmgrYfCDEwpeYJM3AjHRcowiDrAl5uAWQrgBZ2KpRlFoO2REI3Uooy6Li5hjh76CIZHQ8dnYkHNnjAV8mZWxiAz-MyNBZjYAH2_FMlGBRSgbktMQl8KHuaiYVpsqWJWbJxkrGFeYphinmSabIkgypRRkoRqU6agQsyZApCMgwsJUWlqbIMzEA9cuihCABVLYPc

Knaflic, C. N. (2015). *Storytelling with data: a data visualization guide for business professionals* (1st ed.). Wiley. http://hud.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwjR3LSsNAcKjtRUF8Ft_kIL1oy2az2TSCSKktPWg9GPEY8lhrQFNobcGf9VucWZMmFoVeArM7-2Anu_PYnRkAi7dYc-lMkAGyvvaLNJVpmaFwHCVDVwk3cmwRh5FOROcO7927jvCeK_CVu8a8zuJW8PYe_OHrQF6r9MrxM88zdZNk1-tkZRJi8HTrowLs_UNhTUxUqq8T2rUOc1CvEUyHxDIvEwqDirIFMm-Zly3sh71hYzqPdNuu99CYtZJROp4oKlC-nhJhfUxmqjHHcYuZXOXbfw1qHP99VoXNotbvdhY7UatVrGCsliSxprj_pNtV5PxMpz5Clo4ynywBXNo5YNs4UhbaKsfkv2AuWQkmfJS2N6bjANdeCPyU-G1_G2qKnDB2oKLSXdjKU08Y2Um0B_VHehygo4mmI4OMyQa9dd2Hi37P6w6apY79zELlL6-1VYdqOk7VARimcJgincYOuYi52w7aMcqMlhMGzIqkewjnq_R4tBraMayjQGT_mFhOoEpEPIU1bHK2TKpvmVrKdQ

Mengpin Ge, J. F. (2020). *4 Charts Explain Greenhouse Gas Emissions by Countries and Sectors*. WRI. Retrieved 26/02/2021 from https://www.wri.org/blog/2020/02/greenhouse-gas-emissions-by-country-sector

Podder, D. (2020). *CO2 Emissions by Vehicles* Kaggle.com. https://www.kaggle.com/debajyotipodder/co2-emission-by-vehicles

7. APPENDIX [Python code]

8.
```
#libraries required to wrangle and examine data
import pandas as pd

import plotly.express as px
import plotly.io as pio
pio.templates.default="simple_white"
```

9. Import and examine data

10.
```
MASTER_IMPORT = pd.read_csv("CO2 Emissions_Canada.csv")
```

11.
```
MASTER_IMPORT.head()
```

12.
```
MASTER_IMPORT.tail()
```

13.
```
MASTER_IMPORT.info()
```

14. Tidy Data

15.
```
MASTER=MASTER_IMPORT.replace({"X":"Regular Gasoline","Z":"Premium Gasoline","D":"Diesel",
                      "E":"Ethanol (E85)","N":"Natural Gas"})
```

16.
```
MASTER["Gears"]=MASTER["Transmission"].str.strip().str[-1]
MASTER["Gears"]=MASTER["Gears"].replace({"V":"N/A","0":"10"})

CAT_GEARS=pd.CategoricalDtype(["4","5","6","7","8","9","10","N/A"],ordered=True)
MASTER["Gears"]=MASTER["Gears"].astype(CAT_GEARS)
```

17.
```
for i in MASTER.index:
    if MASTER.loc[i,"Transmission"].__contains__("AM")==True:
        MASTER.loc[i,"Transmission"]="Automated Manual"
    if MASTER.loc[i,"Transmission"].__contains__("AV")==True:
        MASTER.loc[i,"Transmission"]="Continuously Variable"
    if MASTER.loc[i,"Transmission"].__contains__("AS")==True:
        MASTER.loc[i,"Transmission"]="Automatic with Select Shift"

MASTER=MASTER.replace({"M5":"Manual","M6":"Manual","M7":"Manual",
                      "A4":"Automatic","A5":"Automatic","A6":"Automatic","A7":"Automatic",
                       "A8":"Automatic","A9":"Automatic","A10":"Automatic"})

CAT_TRANS=pd.CategoricalDtype(["Automatic","Automated Manual","Automatic with Select Shift",
                              "Continuously Variable","Manual"],ordered=True)
MASTER["Transmission"]=MASTER["Transmission"].astype(CAT_TRANS)
```

18.
```
#sort the dataframe by year > country > sex > age into ascending order
MASTER=MASTER.sort_values(["Transmission","Gears","Engine Size(L)","Cylinders"],ascending=[True,True,True,True])
```

19. Plots

20.
```
FIG1 = px.scatter(MASTER,x="CO2 Emissions(g/km)",
                  y="Fuel Consumption Comb (mpg)",
                  color="Fuel Type",
                 title="Scatter plot of Fuel Consumption vs. CO2 Emissions",
                 width=1080,height=610)

FIG1.update_layout(font=dict(family="Calibri",size=16),
                  xaxis_title="Tailpipe CO2 Emissions [g/km]",
                  yaxis_title="Fuel Consumption Combined [mpg]")
FIG1.show()
```

21.
```
FIG2 = px.box(MASTER,
             y="Fuel Consumption Comb (mpg)",
             x="Vehicle Class",
             title="Boxplot of Fuel Consumption by Vehicle Class",
             width=1080,height=610)

FIG2.update_layout(font=dict(family="Calibri",size=16),
                  xaxis_title="Vehicle Class",
                  yaxis_title="Fuel Consumption Combined [mpg]")
FIG2.show()
```

```
22. FIG3 = px.density_heatmap(MASTER,
                             y="Transmission",
                             x="Engine Size(L)",
                             z="Fuel Consumption Comb (mpg)",
                             histfunc="max",
                             color_continuous_scale="Greens",
                              title="Heatmap of Fuel Consumption Combined by engine size and transmiss
    ion",
                              width=1080,height=610)

    FIG3.update_layout(font=dict(family="Calibri",size=16),
                      xaxis_title="Engine Size [L]",
                      yaxis_title="Transmission Type")

    FIG3.show()
```