**Assessing the Impact of Banning Internal Combustible Engines (ICE) on California Emissions**

DATA 621: Fall 2020 Final Project

Zach Alexander, Sam Bellows, Donny Lofland, Joshua Registe, Neil Shah, Aaron Zalki

Table of Contents

[Abstract 3](#_Toc59211233)

[Key Words 4](#_Toc59211234)

[Introduction 5](#_Toc59211235)

[Literature Review 6](#_Toc59211236)

[Methodology 8](#_Toc59211237)

[Experimentation and Results 9](#_Toc59211238)

[Discussions and Conclusions 13](#_Toc59211239)

[References 15](#_Toc59211240)

[Bibliography 15](#_Toc59211241)

# Abstract

# Key Words

Carbon-Dioxide, ICE, Emissions, Vehicle, California

# Introduction

Climate change—the impact of manmade activities on greenhouse gases and their role in changing the climate—has emerged as a top priority for most Americans in 2020 (Nadia, 2020). While the global scientific community has had consensus since the 1990s, the economic impacts of climate change have been thrust to the forefront with the NRDC reporting nearly $1.9 trillion dollars in losses by 2100 if no action is taken (Hope & Alberth, 2008). Such economic and devastating loss is most apparent in California, the most populous state in the United States, where the record heat has pushed extreme summer temperatures to 2.5 degrees warmer than that of 1970, and combined with drought, has made the state a seasonal matchbox for wildfires. In fact, during the 2020 year, California has suffered from six of it’s largest and most devastating wildfires (Rust & Barboza, 2020).

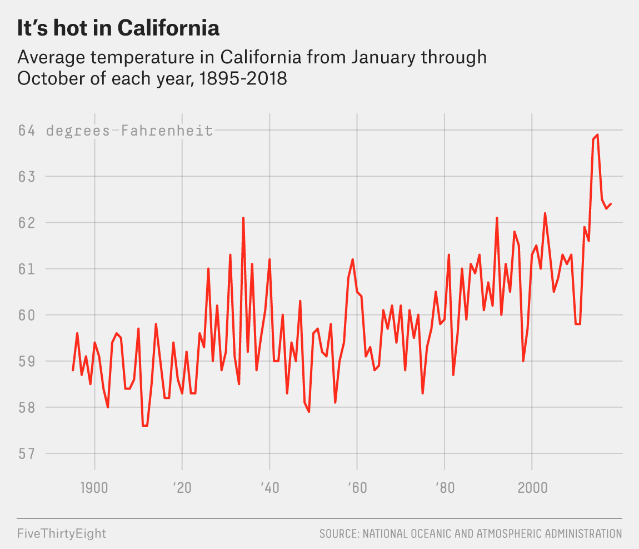


Figure California Temperatures (Christie Aschwanden, 2018)

With such dire consequences on the horizon, California Governor Gavin Newsom signed an executive order in September 2020 to ban all new internal-combustible engine vehicles (ICEs) by 2035, and an outright ban on sales by 2045 (Newsom, 2020). California has long been on the forefront of clean energy, having maintained its own stricter gasoline standard CARBOB, a zero-emission vehicle (ZEV) standard, a separate low-carbon fuel standard incentive on all power generation (LCFS) and even tighter vehicle emission standards (CAA Section 909), with the aforementioned setting nation-wide vehicle standards and also earning the ire of the 2020 EPA administration (Tabuci, 2018). While the transition to zero emission or electric vehicles is more pronounced in Europe and not the United States, Newsom’s outright ban serves as the strongest salvo for any United States state on combatting climate change.

Given the recentness (at the time of this report’s writing) of Newsom’s ban, there is an underlying question of what impact the ban of ICE vehicles would have on meaningful reductions in carbon dioxide emissions in California. Our team seeks to help shed light on different scenarios, and more importantly quantify said carbon dioxide emissions, using generalized linear models on vehicle data set. Such a model would not only be powerful for assessing California’s progress towards a lower carbon future but also serve as a tool for policy makers in other states, to evaluate the efficacy of similar program.

# Literature Review

Emissions is an active area of research, and both the industry and scientific community have a rich history of empirical, policy and simulations to analyze tail pipe emissions. Some researchers focused on point representation of cars and focused on their actual physical characteristics to build generalized linear models or stochastic models parameterized by air resistance, rolling resistance and other physical aspects (Fontaras & Panagiota, 2011). Alternatively, others built a simplified linear model using variables such as mass of car, engine out-put and fuel type, showed that smaller passenger cars using diesel, had less overall CO2 emissions than their diesel or larger European counterparts (G. Mellios, 2011).

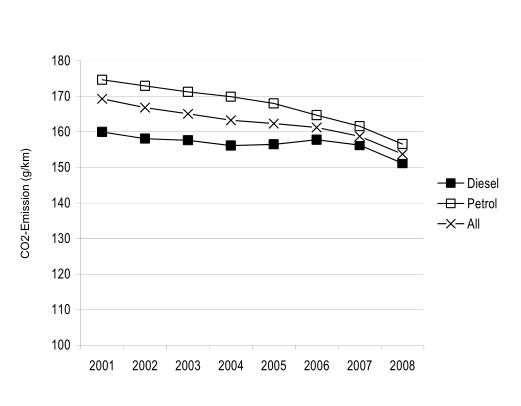


Figure 2 European ICE emissions (G. Mellios, 2011)

Researchers have also shifted from a single-engine point car model to more aggregate systems, that consider traffic conditions or city characteristics as a whole (Kevin R. Gurney, 2012)—one such study reinforced that ICE engine CO2 emissions account for 90% of macro emissions within most cities (Toshiko, 2003). (J.A. Paravantis \*, 2006) incorporated changes in the overall automotive fleet between in Greece as a baseline for CO2 reduction. (Kii, 2020) took the societal aspect of emissions even further, by incorporating changing population dynamics and road infrastructure into CO2 emission modeling, and concluded that technology and a slower growing population, and thus less transportation, would drive emission reductions.

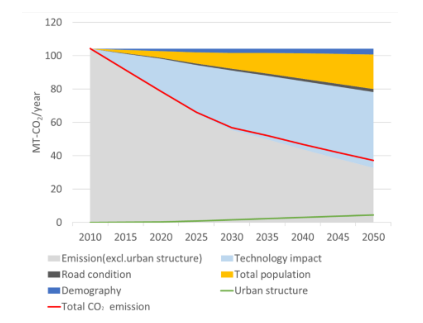


Figure 3: CO2 Reductions (Kii, 2020)

The authors of this paper have found that most studies are localized on a single country or continent, and therefore might not reflect the diversity of automotive manufactures, which while CO2 emission reduction is ubiquitous, ideological differences might impact individual car design. Therefore, the authors wanted to focus on the California market, which as mentioned before is diverse and the largest in the US.

# Methodology

In order to effectively quantify carbon dioxide emissions over the next fifteen years based on policy changes in California, we initially set out to create a meaningful multiple linear regression model that could estimate the amount of carbon emissions (g/km) for a wide array of ICE vehicles. To do this, we had to first identify features that account for the variance in these emissions. Linear regression models offer insight into the relationship between the input variables (x) and the single output variable (y). For our purposes, this was an effective initial technique to determine which factors were correlated with carbon emissions from ICE vehicles and to help estimate our single output of carbon emitted (g/km) given different vehicle features.

The data used to identify these features and help build our multiple linear regression model was originally released on the Government of Canada website but was also posted on Kaggle. We quickly found that the dataset included a large proportion of car makes and models that were built by companies such as Ford, Chevrolet, BMW, and Mercedes Benz, but also included lesser known companies that sell to niche markets, including Bugati and Smart. Additionally, there was a wide variety of vehicle classes, with the most common being classified as small sport utility vehicles (SUVs), mid-size, and compact vehicles. Fuel type was quite ubiquitous across vehicle classifications, with most either requiring regular or premium gasoline. However, there was a small percentage that ran on ethanol or diesel. Most vehicle transmissions fell under the category of “automatic select” or “automatic”, however, there was also a fairly large proportion of manual transmissions with 5 or 6 gears. Most cars contained 4-cylinder engines, but there was a range from 2 to 16 cylinders.

When evaluating kurtosis of our continuous variables in the dataset, we found that fuel consumption was slightly right skewed. Therefore, we subjected these features to Box-Cox transformations to normalize values, where our lambda value ranged from -0.4 to 0.2. Based on these outputs, we were able to create more normal distributions for these features, which helped provide flexibility for our future linear regression models.

After we subjected our data to feature transformations, we combined these new transformed variables with our original nominal variables. With one compiled dataset, we then split the data into a training and testing set – 80 percent was denoted to the training dataset and 20 percent was denoted to a holdout testing set.

Initially, we noticed from a preliminary linear regression model that fuel consumption (in miles per gallon) and engine size (in liters) were quite predictive of carbon dioxide emissions. Therefore, we added additional features of vehicle class, cylinder engine composition, transmission type, and fuel type to the model, subjected it to stepwise selection, and yielded an R2 value of about 0.98.

# Experimentation and Results

We compiled summary statistics on our data set to better understand the data before modeling and sought out to obtain distribution profiles of each of the variables within the dataset. Figure 4 represents histrograms that describe the shape of the data. It was noted that the datasets all were almost normally distributed but were skewed right for all of the variables. This required performing a Boxcox transformation in order to normalize these distributions and allow us to better satisfy linearity assumptions associated with linear regression models.

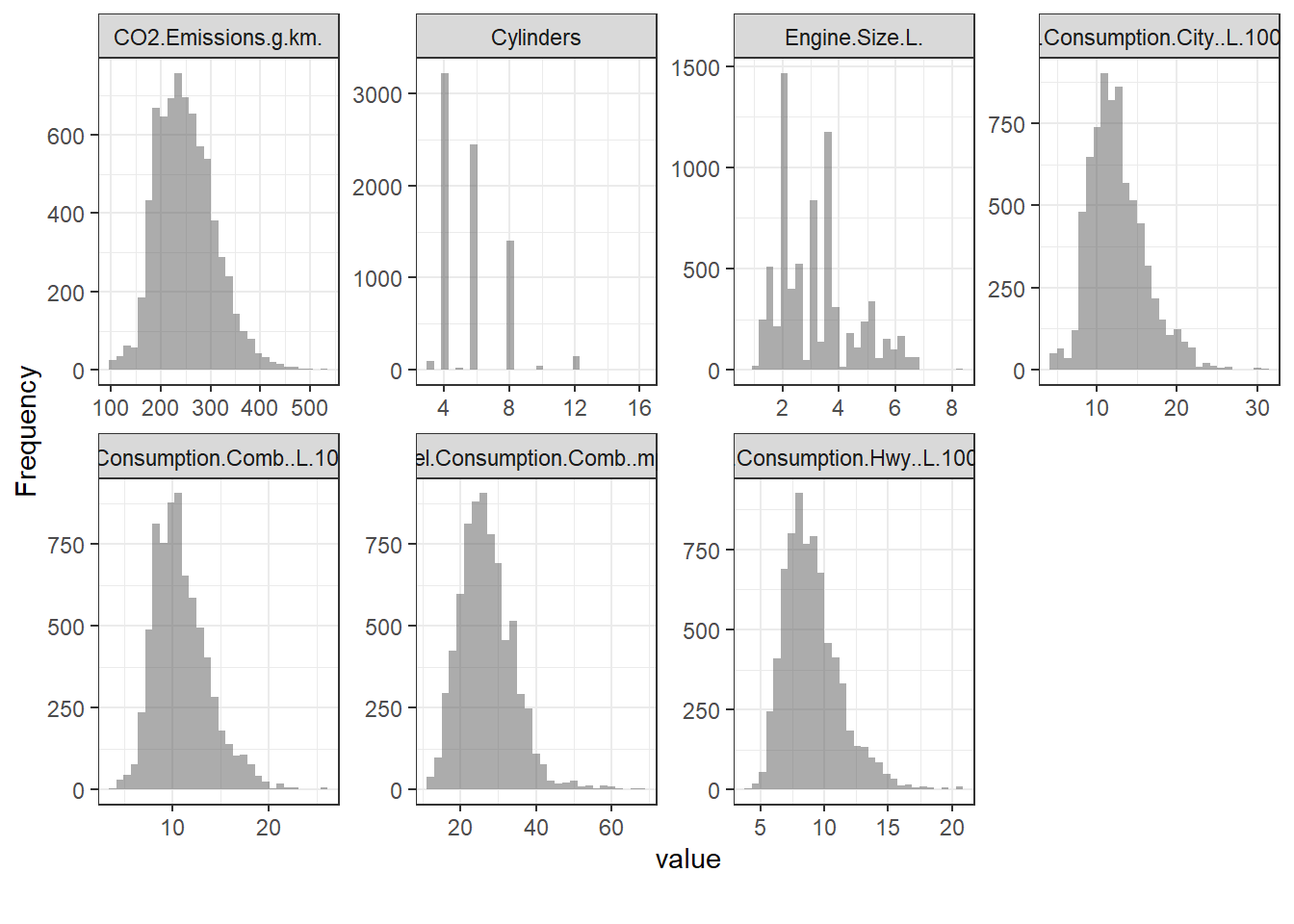


Figure - Distribution of Numeric Variables

Additionally, there were four variables that were considered non-numeric. These parameters are: “Make”, “Vehicle Class”, “Transmission”, and “Fuel Type”. Figure 5 presents a barplot of these variables and their prevalence in the dataset. We observed that Ford, Cheverolets, and BMWs were amongst the top 3 car manufacturers prevalent and small to mid size SUVs were the most common vehicle class. 6-speed and 8-speed automatic transmissions were heavily present and fuel type “X” and “Z” are the most common. In order to consider these systems within our linear regression, we apply dummifying techniques that binarize the data into 0’s and 1’s.

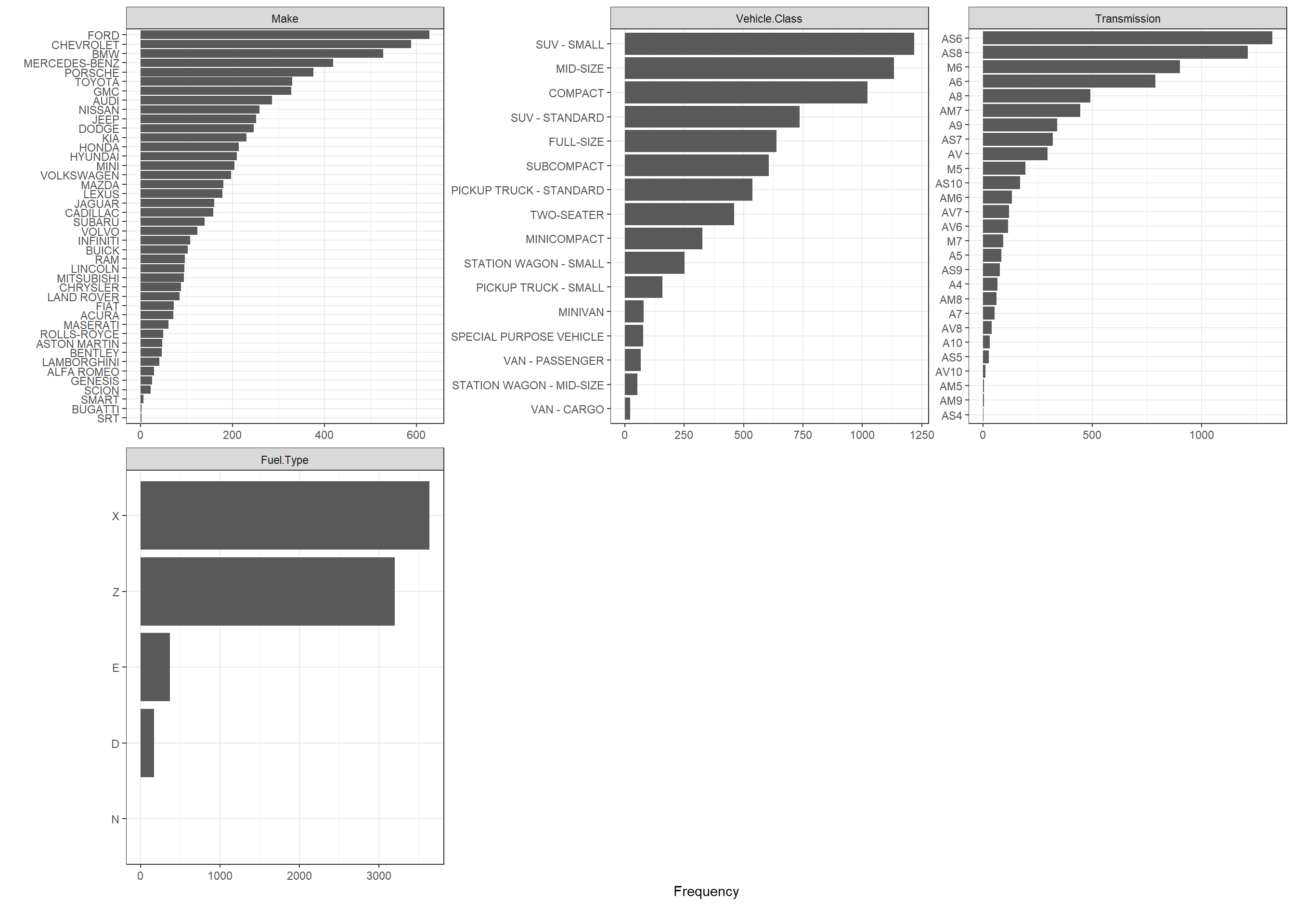


Figure - Quantitative description of Qualitative Variables

Next, In order to obtain a preliminary assessment of the relationships in our dataset to what our target variable will be (CO2 emissions). Figure 6 shows the individual relationships between CO2 emissions and each of the numeric variables within the dataset. It was observed that all numeric variables show some monotonic relationship with CO2. Some of the relationships appear linear such as *Cylinders, Engine Size, Fuel Consumption City* and *Fuel Consumption Hwy.* A relationship with *Fuel Consumption Combined* is also present in the form of what appears to be a polynomial-like function. Relationships like this require transformations in order to satisfy the assumptions of linearity in the regression problem. Additionally, Figure 7 is shows a correlation plot that identifies relationships between all numeric variables. This is important to identify because of the potential of multi-colinearity within the dataset. The presence of multi-colinearity can increase the variance also known as variance inflation. A variance inflation factor can be calculated for the dataset in order to potentially eliminate variables that have high VIFs.

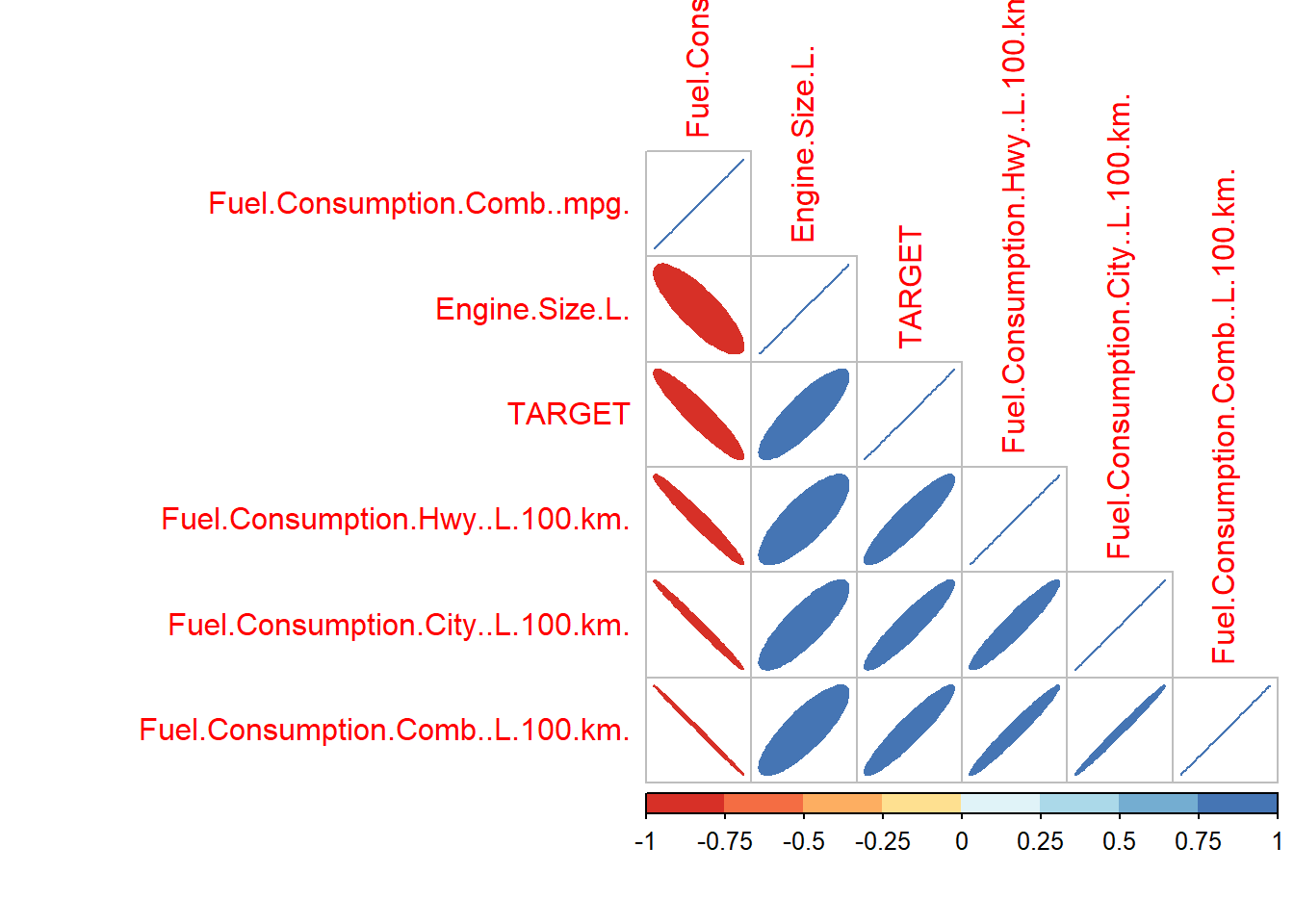
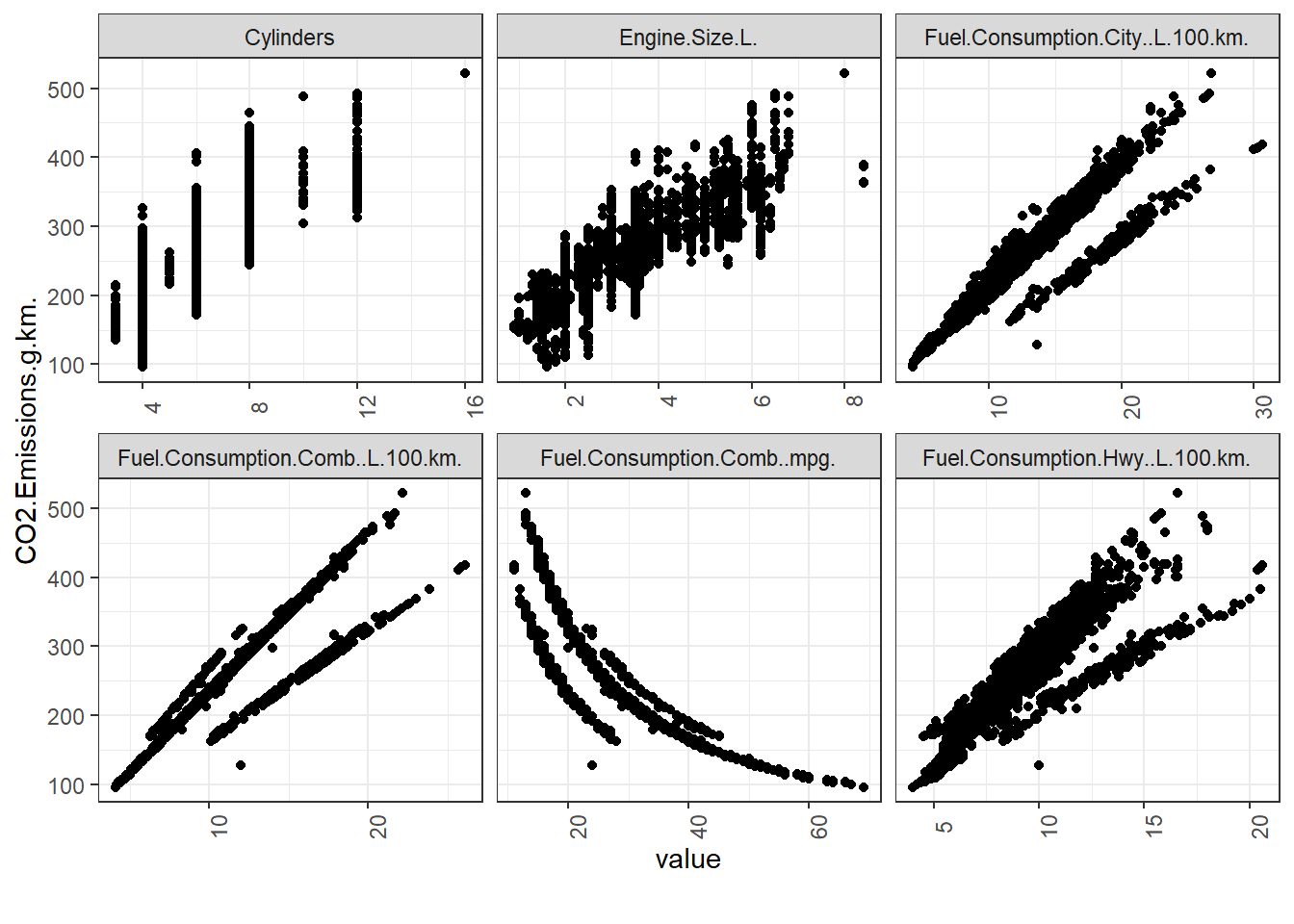
 

Figure - Correlation Matrix of all Variables

Figure - Numeric Variables vs CO2 Emissions

In order to transform the data, boxcox transforms with optimal lambdas determined by R were used to transform all of the variables in the dataset. Figure 9 shows these transformed variables and in comparison to the untransformed data, these appear to follow a more gaussian-like distribution. Finally, imputation on the dataset for missing information was not needed as shown in Figure 8. This figure represents amount of sparsity within the dataset and fortunately this Kaggle dataset did not have any missing information and no imputation was necessary.

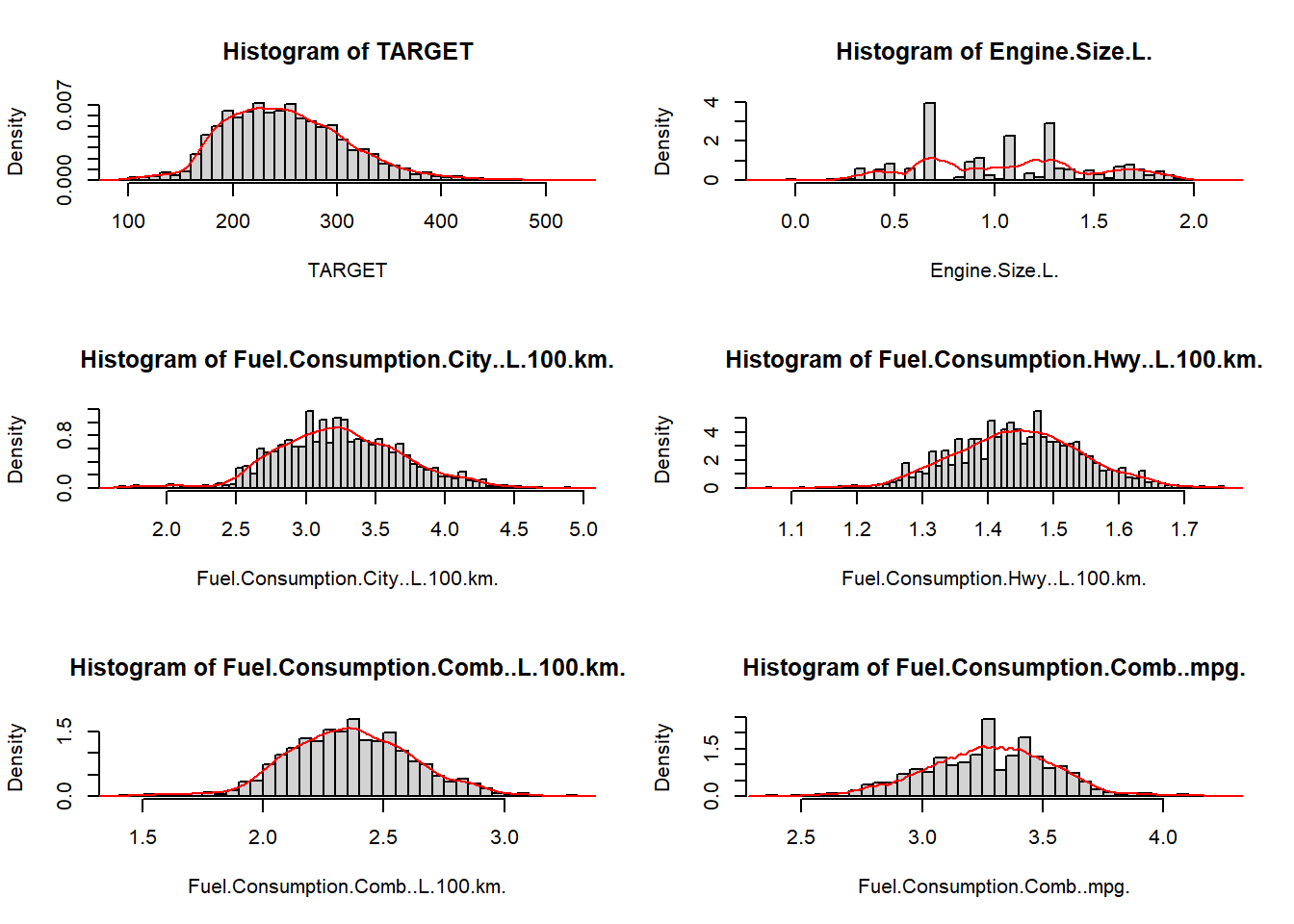
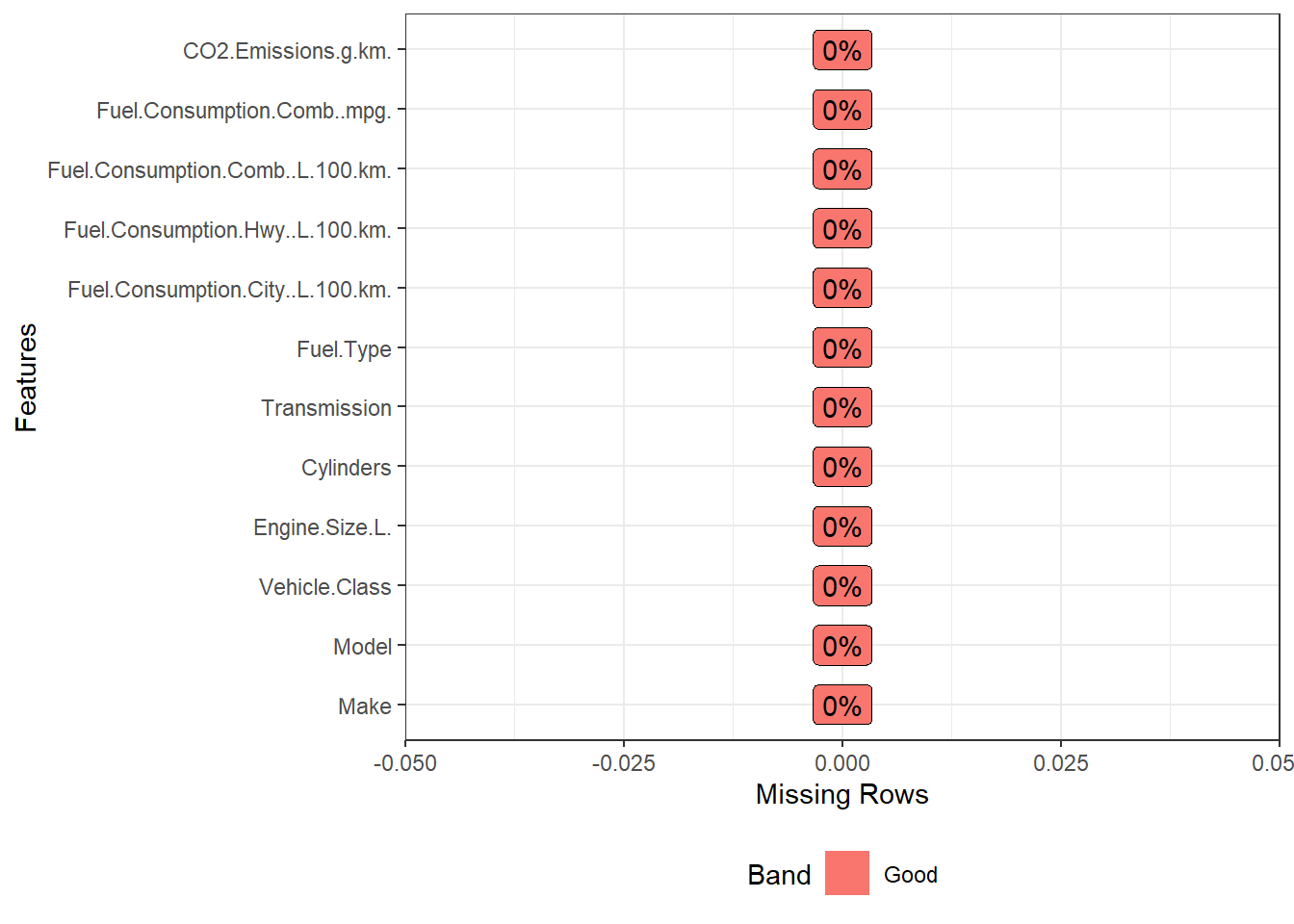


Figure - Distribution of Transformed Variables

Figure - Missing Variables



# Discussions and Conclusions

**Table 1: Model Results**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RSE | Adj. R2 | F. Statistic |
| Model 1 | 19.32 | 0.891 | 1.207x104 |
| Model 2 | 7.577 | 0.983 | 6830 |
| Model 3 | 7.577 | 0.983 | 6967 |

The results were consistent between the training and the testing sets for both model 2 and model 3. Because of the simplicity of model 3 with StepAIC and the reduction of variables considered via stepwise selection, no compromise to the performance was made with the exception of a slightly higher F statistic implying slightly higher statistical significance although negligible. Train/test evaluation was not done for model 1 since model 2 and 3 performed significantly better on the training sets.

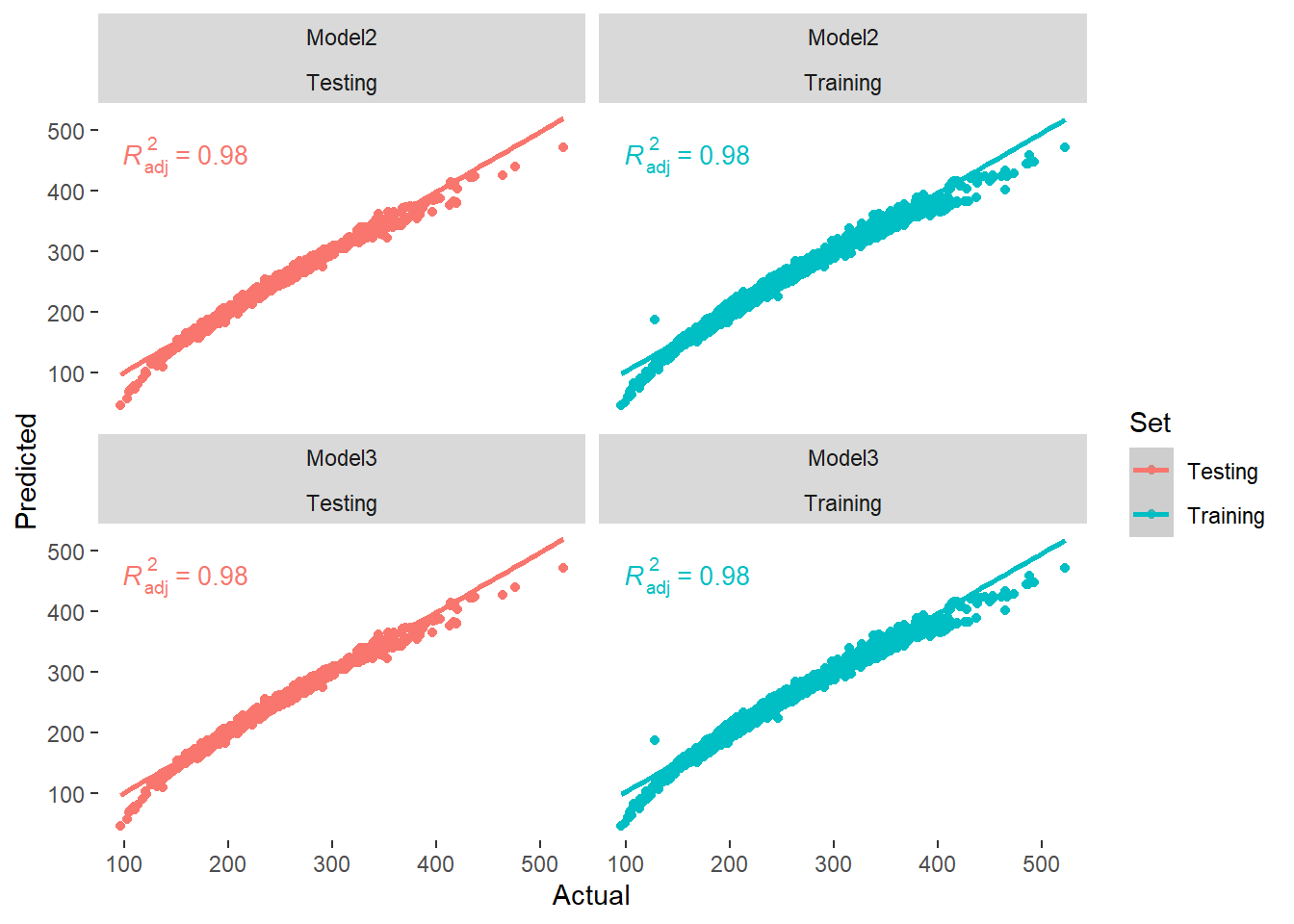


Figure - Test/Train comparison for Model Performance

Finally, Variable importance was observed for all variables and Figure 11 shows the top 5 predictors of each model. Model 1 differed significantly in the variable that were most important in predicting CO2 emissions. The largest predictor was engine size, however for model 2 and 3, they consistently determined that whether or not the vehicle had the specific fuel types was the largest determining factor of CO2 emissions.

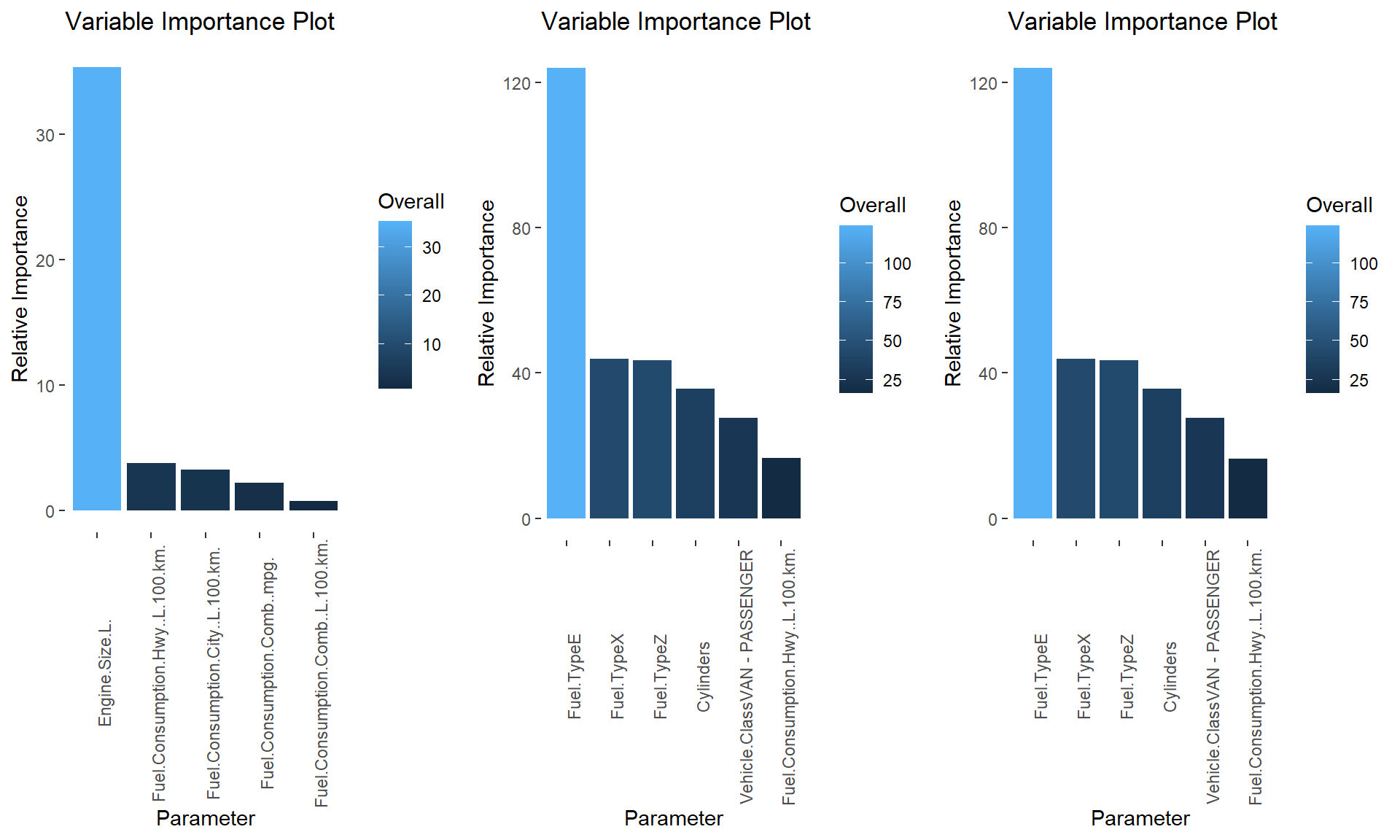


Figure - Variable Importance for Models

# References

# Bibliography

1. Christie Aschwanden, A. M.-J. (2018, November 16). *Why California’s Wildfires Are So Destructive, In 5 Charts*. Retrieved from 538: https://fivethirtyeight.com/features/why-californias-wildfires-are-so-destructive-in-5-charts/
2. Fontaras, G., & Panagiota, D. (2011). The evolution of European passenger car characteristics 2000–2010 and its effects on real-world CO 2 emissions and CO 2 reduction policy . *Energy Policy*.
3. G. Mellios, S. H. (2011). Parameterisation of fuel consumption and CO 2. *JRC Scientific and Technical Reports*.
4. Hope, C., & Alberth, S. (2008). *The Cost of Climate Change: What We’ll Pay if Global Warming.* NRDC.
5. J.A. Paravantis \*, D. G. (2006). Trends in energy consumption and carbon dioxide emissions. *Technology in Forecasting and Societal Change*.
6. Kevin R. Gurney, I. R. (2012). Quantification of Fossil Fuel CO 2 Emissions on the Building/Street. *Environmental Science and Technology*.
7. Kii, M. (2020). Reductions in CO 2 Emissions from Passenger Cars in Japan under Population and Technology. *Sustainability* .
8. Nadia, P. (2020, Feb 20). *Climate Change Rises as a Public Priority. But It’s More Partisan Than Ever.* Retrieved from New York Times: https://www.nytimes.com/interactive/2020/02/20/climate/climate-change-polls.html
9. Newsom, G. (2020, September 23). *Executive Department State of California: N-79-20*. Retrieved from https://www.gov.ca.gov/wp-content/uploads/2020/09/9.23.20-EO-N-79-20-Climate.pdf
10. Rust, S., & Barboza, T. (2020, September 13). *How Climate Change is fueling record-breaking California wildfires, heat and smog*. Retrieved from Los Angeles Times : https://www.latimes.com/california/story/2020-09-13/climate-change-wildfires-california-west-coast
11. Tabuci, H. (2018, April 2). *New York Times*. Retrieved from Calling Car Pollution Standards ‘Too High,’ E.P.A. Sets Up Fight With California: https://www.nytimes.com/2018/04/02/climate/trump-auto-emissions-rules.html
12. Toshiko, N. (2003). Energy Modeling on Cleaner Vehicles in Japan. *Journal of Cleaner Production*.