Exploring CO2 Emissions from Automobiles

Andrew Mashhadi

Introduction

Carbon dioxide (CO2) emissions is often recognized as one of the main drivers in global climate change. Countries today are urgently trying to reduce their annual emissions in an attempt to prevent any further impacts of climate change. Although automobiles are not soley responsible for the current state of carbon dioxide levels, there is little debate that they significantly contribute to the total annual CO2 emissions. In this paper, I will use *Principal Component Analysis* (PCA) and *Factor Analysis* with multiple linear regression to explore emissions data in an attempt to bring out strong patterns between various vehicle attributes and CO2 emissions, and to characterize the key factors that affect emission levels. Additionally, I will evaluate the overall fit of the models, assess the corresponding predictive performance, and investigate the relationship between the main factors and the expected CO2 emissions.

Data

The dataset [2] used in this project was originally taken from the Canadian Government's official open-access website, and was collected over a 7 year long period. The data included over 7000 observations and 10 original variables. Each observation is associated with an independent automobile, and the data includes a variety of vehicle attributes with an associated CO2 emission (measured in grams per kilometer). The vehicle attributes consisted of *Make*, *Model*, *Vehicle Class*, *Engine Size*, *Cylinders*, *Transimission*, *Fuel Type*, *City Fuel Consumption*, *Highway Fuel Consumption*.

Data Cleaning and Feature Engineering

We can see above that we originally have 9 explanatory variables and 1 response variable (CO2 Emissions). However, I should note that many of the original categorical features contain a plethora of categories, which would vastly increase the dimension of the feature space when dummy coding is applied to our variables prior to modeling. Throughout the exploratory data analysis, many of these original categorical variables were split up, or re-categorized, in an attempt to limit the number of categories from a single feature and to combine any categories based on general commonalities.

The `Make` variable originally consisted of over 40 different categories, so I replaced it with a new variable, `Economic Class`, containing 3 categories describing the the luxury status of the vehicle's make. The `Model` variables consisted of over 2000 different categories. Therefore, I decided to remove this feature from consideration since there would simply be too many categories compared to the number of observations. Since the `Drive` variable had over 5 categories describing the "Drive" of the vehicle, I re-categorized the variable such that only a "Two-Wheel Drive" category and an "All-Wheel Drive" category remained. The `Vehicle Class` variable consisted of over 15 different categories describing the type of vehicle it is (Sedan, Van, SUV, etc.), so I also re-categorized this variable to describe the approximate size of the vehicle with the categories: "Compact", "Mid-Size", "Full-Size", and "Large Vehicle". For simplicity, I re-categorized the `Fuel Type` variable such that each observation can be grouped within a "Regular Gasoline" category or a "Not-Regular Gasoline" category.

For the last categorical variable, `Transmission`, I ultimately decided to split up the two independent peices of information the original variable presented. Each observation was given a specific acronym that effectively described the type of transmission ("Manual" or "Automatic") and the number of gears (from 1-10 gears). Therefore, I split this variable into two new variables: (1) `Transmission Type` with two categories for "Manual" and "Automatic" vehicles and (2) `Number of Gears` as a numerical variable containing the total number of gears in the vehicles transmission. Note that I set `Number of Gears = 0` for vehicles with a Continuously Variable Transmission (CVT) since there are technically no established gears in a CVT.

Exploratory Data Analysis

With the cleaned dataset, I explored the one-way distributions for the numerical variables and the one-way frequency tables for the categorical variables (See Figure 1 in the Appendix). Notice that all of the continuous variables except for `Engine Size` and `Cylinders` have approximately normal distributions, while the variables `Engine Size` and `Cylinders` seem to have right (positive) skewed distributions. Also notice that *all* of the cleaned categorical variables demonstrate large sample sizes ($n_i > 100$) in each category. I found that the distribution of the response variable, `CO2 Emissions`, clearly demonstrated normality with an estimated mean of about 250 g/km and an estimated standard deviation of 59 g/km.

Next, I explored the two-way relationships between each vehicle feature and the CO2 Emissions (Figure 2). Box-plots were used for the categorical variables to potentially illustrate general differences in the distribution of CO2 emissions when conditioned on each category, and scatter-plots of the CO2 emissions against the explanatory variable were used for the numerical features. As shown in Figure 2, all of the categorical variables indicate significantly different CO2 emissions means (with similar distributions) for each category of the associated vehicle feature. Additionally, all of the numerical features except for the `Number of Gears` feature demonstrated a clear, positive, linear relationship with the CO2 emissions. After visual examination, it appears that the `Number of Gears` may have a non-linear relationship with the CO2 emissions.

Lastly, I investigated the relationships between all of cleaned variables in the dataset. The Variance Inflation Factors (VIF) and the standard Pearson's Correlations plot is presented in Figure 3 of the Appendix. Notice that all of the VIF values are approximately less than or equal to 5, so I had no reason to believe any negative effects from multicollinearity would impact the models. In agreement with the two-way plots, the correlation plot indicates that most of the numeric variables have a relatively large correlation with `CO2 Emissions`. Although the two fuel consumption variables have relatively large VIF values (as we would expect) and many of the other numeric vehicle features demonstrate relatively large correlations with one another, I should note that PCA and Factor Analysis may potentially alleviate some of the negative effects from collinearity when applied before modeling.

Methodology

As previously mentioned, the goal of this project is to use *PCA* and *Factor Analysis* with multiple linear regession to explore the data for strong patterns and key factors associated with the automobile features, and to examine their abiltiy in predicting CO2 emissions.

Before conducting any analyses, I performed a random 80/20 split on the original data to construct a training set and testing set. Since many of the variables used are not in the same units, I also scaled both sets using the estimated means and standard deviations from *only* the training set.

Using only the training set, I applied principal component anlysis (PCA) to our automobile attributes to reduce the dimensions of the data while minimizing information loss. I investigated the loadings of the first few components in search of any patterns or trends. Based on the location of the "elbow" in the corresponding scree plot and the variance explained by each principal component, I chose an optimal number of components $k^*_{PCA} < p$ where p is the total number of explanatory variables. After linearly transforming the training data into principal components, the first k^*_{PCA} components were used to fit a multiple linear regression model. Linear model assumptions were checked, residual diagnostics were examined, and the significance of each coefficient was tested. It is worth noting that the optimal number of components used as input was made such that the first k^*_{PCA} components should explain a sufficient amount of the sample variance in the training data without significant impacting its relationship with CO2 emissions.

I then applied factor analysis to the training data, using the *Maximum Likelihood* method. To help with the interpretation of the factors, I also applied a *varimax* rotation to "spread out" the squares of the loadings on each factor in hope of finding groups of large and negligible coefficients in any column of the rotated loadings. The loadings of the first few factors were investigated and interpreted in order to characterize the key factors found among the original automobile features. Again, I used the location of the "elbow" in the corresponding scree plot and the variance explained by each factor to choose an optimal number of factors, $k^*_{FA} < p$. With the loadings from the first k^*_{FA} factors, I then generated factor scores using the weighted-least-squares method shown in class:

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'_{\mathbf{z}} \hat{\mathbf{\Psi}}_{\boldsymbol{z}}^{-\mathbf{1}} \hat{\mathbf{L}}_{\mathbf{z}})^{-1} \hat{\mathbf{L}}'_{\mathbf{z}} \hat{\mathbf{\Psi}}_{\boldsymbol{z}}^{-\mathbf{1}} \mathbf{z}_j \quad \text{for } j = 1, 2, ..., n$$

These factor scores were then used as input for another multiple linear regression model. Again, the linear model assumptions were checked, residual diagnostics were examined, and the significance of each coefficient was tested. Note, again, the optimal number factors used as input was made such that the first k^*_{FA} factors should explain a sufficient amount of the sample variance in the training data without significant impacting its relationship with CO2 emissions.

Both linear regression models were evaluated and compared on the test dataset using R-Squared and their Root-Mean-Square Error (RMSE). The R-Squared value provides a measure of the variance in CO2 emissions explained by the k^*_{PCA} components or the k^*_{FA} factors. The RMSE provides a measure of prediction error in the same units of the response variable (g/km).

A Note On Correlation Structure

The methodology described above does not describe how the categorical variables are encoded prior to applying PCA, Factor Analysis, or linear regression modeling. Traditionally, one-hot encoding (also known as *Treatment* Dummy Coding) is applied to the categorical variables so that the variable may be used appropriately in our our analyses and models. One-hot encoding effectively creates a new dichotomous variable (1 or 0) for each category in the associated variable. Normally, this approach is perfectly fine before applying a machine learning model such as linear regression. However, when applying PCA or Factor Analysis, using an estimated correlation matrix, **R**, from Pearson's correlations is not technically appropriate because Pearson's correlations assume that the variables are continuous and follow a multivariate normal distribution. Therefore, Pearson's correlation should not be used on dichotomous or ordinal variables (although it is often performed) [1].

Although I used one-hot encoding to perform PCA, factor analysis, and linear regression modeling as described above, I also wanted to try using a more appropriate estimate for the correlation of the categorical variables in the dataset. Therefore, I estimated the correlation matrix, **R**, using a mix of different correlation methods. The numerical variables still used Pearson's correlation coefficient, but the two-class (dichotomous) categorical variables used tetrachoric correlation [3]. I treated the other multi-class variables, `Vehicle Class` and `Economic Class`, as ordinal (also known as *polytomous*) variables, and used polychoric correlation [3]. Naturally, the vehicle size was used for the ordering of the `Vehicle Class` variable, and general price range was used for the ordering of the `Economic Class` variables ("Economy" is cheapest, and "Luxury" is most expensive). Polyserial or biserial correlation was used to estimate the elements of **R** that represented the correlation between a continuous variable and an ordinal, or dichotomous, variable. The analyses and modeling results from using standard one-hot encoding with Pearson's correlations are presented along with the results from using the mixed correlations in the next section.

Results & Interpretation

Principal Component Analysis

After one-hot encoding the categorical variables, there were a total of 13 predictor variables in the training data. Using *only* Pearson's correlation, I calculated the 13 sample principal components from the training data and assessed the variance explained by each component. Since the corresponding scree plot (Figure 4) shows an elbow occuring right before the 6th sample principal component, I decided to keep only the first $k^*_{PCA} = 6$ components. Together, the components explained approximately 93% of the sample variance in the training data. Each predictor's weights for the first 6 components are displayed in Table 1. Notice that the larger weights of PC1 (≈ 0.5) are exclusively associated with variables like `Engine Size`, `Cylinders`, and `Fuel Consumption` levels. Therefore, PC1 may be interpreted as an average measure of engine power, or engine output. Additionally, the relatively larger weights associated with `Large Vehicle`, `Regular Fuel Type`, and `Two-Wheel Drive` in PC2 indicates that PC2 may represent an aggregate measure for regular, non-premium, large passenger vehicles. For instance, this could represent a typical middle-class family oriented car, since they are typically larger, two-wheel drive vehicles that do not require special fuel. Although PC3 appears to be heavily dominated by the `Number of Gears` variable, the 3 remaining principal components may also be interpreted in a similar fashion as PC1 and PC2.

As mentioned in the methodology section, I also reperformed PCA using a combination of different correlation estimates depending on whether the variable is numerical, ordinal, or dichotomous. Therefore, *without* one-hot encoding, I used the mixed-type correlation estimates to calculate the sample principal components. Since one-hot encoding was not performed, there were only 10 predictor variables in the training data before applying PCA. Using the elbow in the scree plot (Figure 4), I again decided to keep the first 6 components. Together, the components explained over 95% of the sample variance in the training data! The corresponding weights can be seen in Table 2. Again, the heavier weights on `Engine Size`, `Cylinders`, and `Fuel Consumption` levels indicate that PC1 may be interpreted as an average measure of engine power, or engine output. However, PC2 has a different interpretation now. The `Vehicle Class` and `Fuel Type` variables have larger positive weights while the `Economic Class` variable has a large negative weight. This implies that PC2 may be interpreted as the difference between a vehicle's overall size and its general status. For brevity, I do not interpret the remaining components, however, they may also be interpretted as aggregate measures of vehicle features. Note that in this case we can see that none of the components seem to be dominated by any single predictor.

Factor Analysis (MLE Method)

As before, I first used *only* the Pearson's correlations from the One-Hot encoded training data to compute the 13 factor loadings, and assessed the variation explained by each factor. After examining the scree plot (Figure 5), I decided to keep only the first $k^*_{FA} = 8$ factors. Together, these 8 factors were shown to explain approximately 69% of the variation in the training data. The factor loadings are presented in Table 3. Interestingly, `Engine Size`, `Cylinders`, and `Fuel Consumption` levels all load highly on the first factor, so it is certainly possible that this first factor is also an abstract measure of engine strength, power, and output. For the second factor, the loading corresponding to the `Premium Economic Type` variable is relatively large and positive, while the loading for `Regular Fuel Type` variable has a relatively large negative value. Since premium vehicles generally require premium gasoline, this second factor seems to represent the contrast between standard vehicles and premium vehicles, with a heavy emphasis on the vehicle's fuel type. We should note that many of the remaining factors appear to be dominated by a single predictor variable as highlighted in Table 3. Any remaining factors that are not dominated by a single predictor variable may be interpreted in a similar fashion as the first and second factors.

Again, I reperformed factor analysis using a combination of different correlation estimates depending on whether the variable is numerical, ordinal, or dichotomous. Furthermore, *without* one-hot encoding, I used the mixed-type correlation estimates to calculate the factor loadings. Note that there was only 10 predictor variables in the training data in this case, since one-hot encoding was not performed beforehand. Using the elbow in the scree plot (Figure 5), I again decided to keep the first $k^*_{FA} = 8$ factors. In this case, the factors were shown to explain approximately 79.3% of the sample variance in the training data! The corresponding weights can be seen in Table 4. Once again, we can see that `Engine Size`, `Cylinders`, and `Fuel Consumption` levels all load highly on the first factor. As before, I believe this first factor may represent an abstract measure of engine strength, power, and output. Similarly, for the second factor, we can see the loading that corresponds to the `Economic Class` variable is a significantly large positive value, while the loading for the `Regular Fuel Type` variable has a significantly large negative value. This second factor may represent the contrast between standard vehicles and premium vehicles, where fuel type is used to distinguish the two classes. From the larger loading values of `Drive` and `Transmission`, it appears that the third factor may represent an abstract measure of how "sporty" the vehicle is, because most pure sports cars are manual and two-wheel drive. The remaining factors that are not heavily dominated by a single predictor variable may be interpreted in a similar fashion.

Linear Regression

I used the two sets of principal components and the two sets of factors scores discussed above as inputs for four different linear regression models. In all four models, most (if not all) of the components or factors were found to be significant, and each set of components or factors demonstrated a clear linear relationship with the CO2 emissions. The residual diagnostics plots (such as Figure 6) from each model indicated that the linear model assumptions were approximately satisfied in all four models. Additionally, no outliers or influential points were found in the training data. Table 5 contains the estimated coefficients. In all four models, we found that the coefficient of the first component, or factor, is a relatively large positive value. This implies that as the abstract measure of engine output (described above) increases, we expect the CO2 emissions to increase significantly.

Using the test data, the RMSE and R-Squared for each model is shown in the table below:

Linear Model Inputs	RMSE (g/km)	R-Squared
PCA w/ One-Hot Encoding and Pearson's	19.6	0.887
PCA w/ Mixed Correlations	20.6	0.874
FA w/ One-Hot Encoding and Pearson's	19.3	0.890
FA w/ Mixed Correlations	19.3	0.891

From the RMSE values above, it's clear that all four models demonstrate excellent predictive performance. Since the R-Squared values are so high, we know that the linear models using the optimal number of components, or factors, were each able to explain a significant amount of the variance in the CO2 emissions.

Conclusion & Discussion

Using PCA and Factor Analysis with different approaches to correlation estimation, I was able to successfully calculate and interpret the corresponding component weights and factor loadings. Although the loadings and the associated variance explained changed between each analyss, the interpretation of the first component or factor remained similar across all four analyses. The remaining components and factors, however, varied by each analysis.

Additionally, the optimal number of components (k^*_{PCA}) and factor scores (k^*_{FA}) were generated for the training set, and the linear models were successfully fit to the associated set of components and estimated factors. From the diagnostic tools and the testing results, we found that each model displayed a great fit and demonstrated an excellent ability to predict expected CO2 emissions. In this case, it was found that using *either* PCA or factor analysis for dimensionality reduction led to similar predictive performances. However, we found that the factor scores led to slightly lower RMSE scores and higher R-Squared values on our test set. Additionally, it was found that using the polychoric and tetrachoric correlations for non-numeric variables did not significantly affect the linear model's performance.

References

[1] UCLA Office of Advanced Research Computing. *How can I perform a factor analysis with categorical (or categorical and continuous) variables?* URL: `https://stats.oarc.ucla.edu/stata/faq/how-can-i-perform-a-factor-analysis-with-categorical-or-categorical-and-continuous-variables/`. (accessed: 03.09.2023).

[2] Debajyoti Podder. *CO2 Emission by Vehicles.* URL: `https://www.kaggle.com/datasets/debajyotipodder/co2-emission-by-vehicles`. (accessed: 03.09.2023).

[3] William Revelle. *mixedCor: Find correlations for mixtures of continuous, polytomous, and dichotomous variables.* URL: `https://www.rdocumentation.org/packages/psych/versions/2.2.9/topics/mixedCor`. (accessed: 03.09.2023).

Appendix

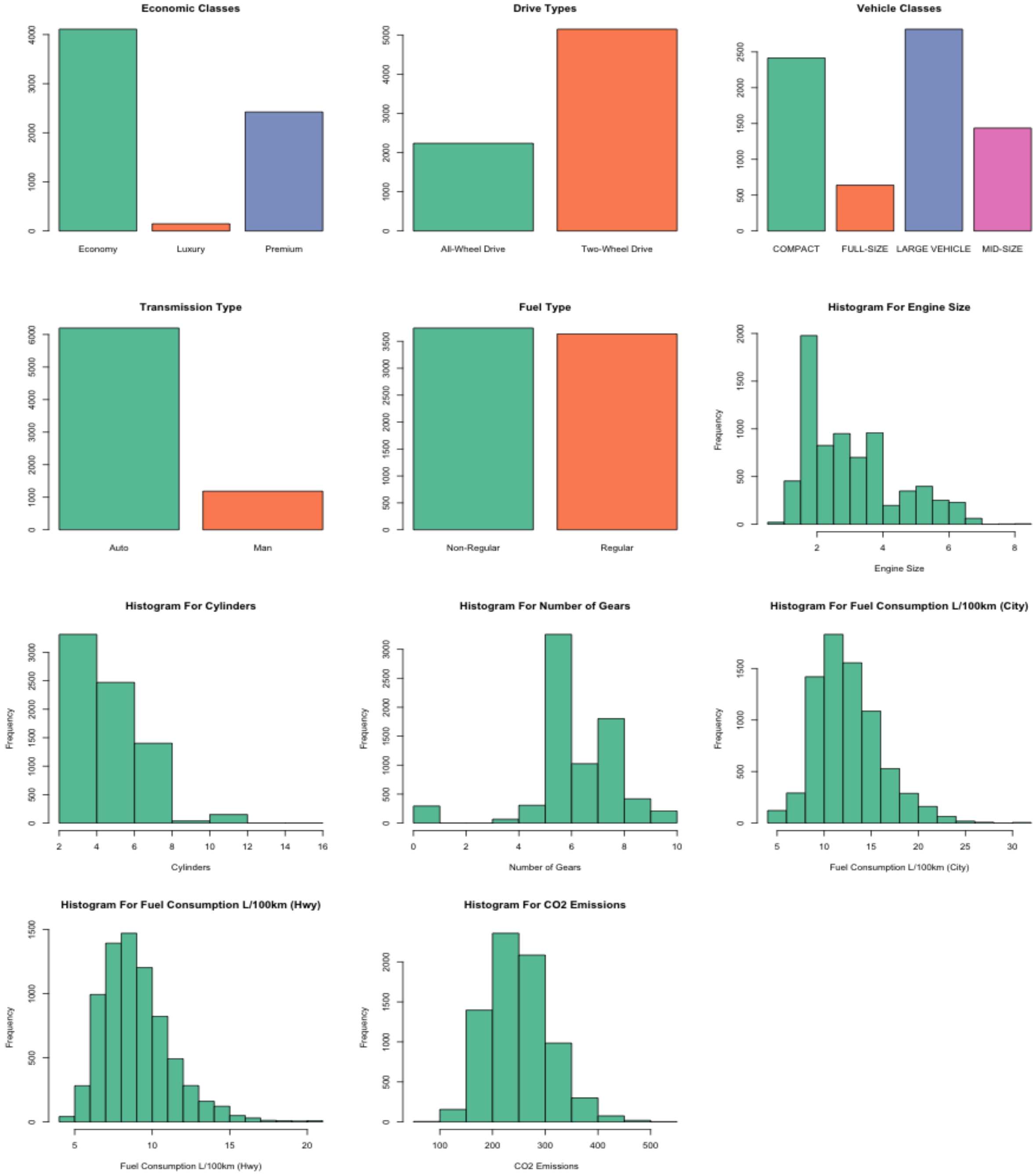


Figure 1: One-way Distributions and Frequencies

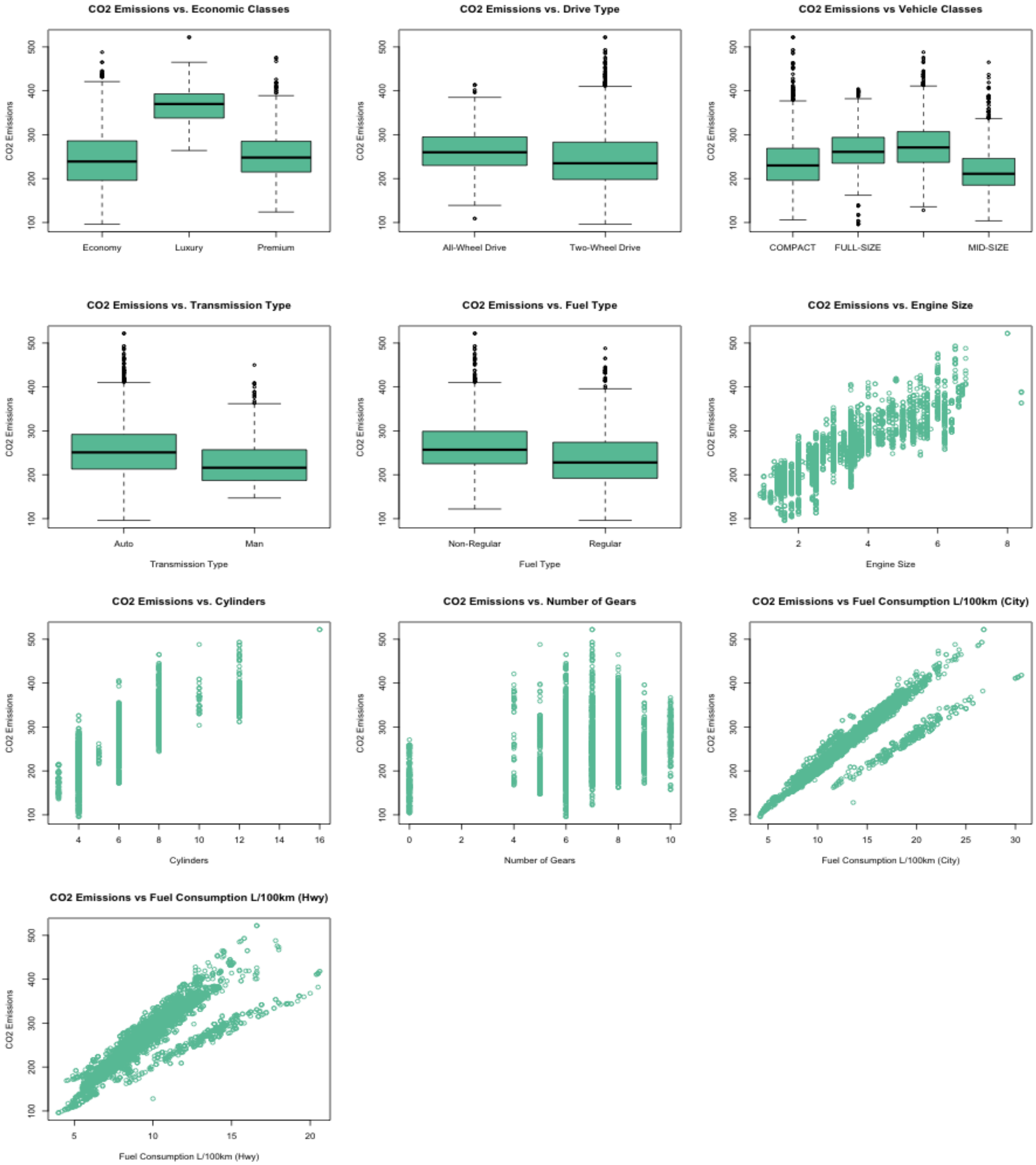
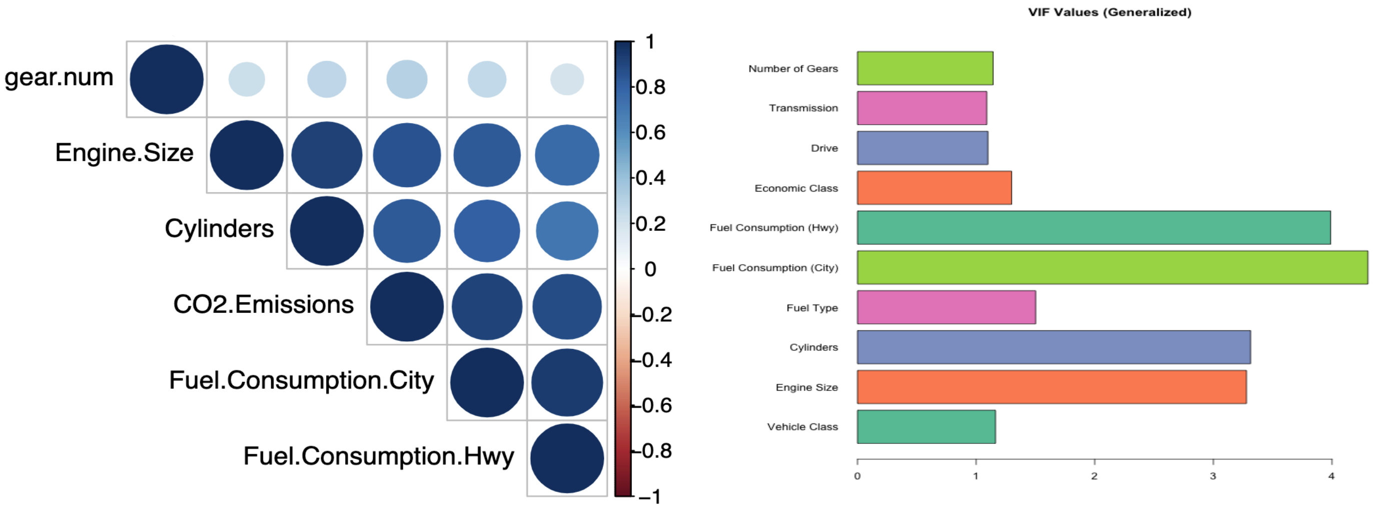


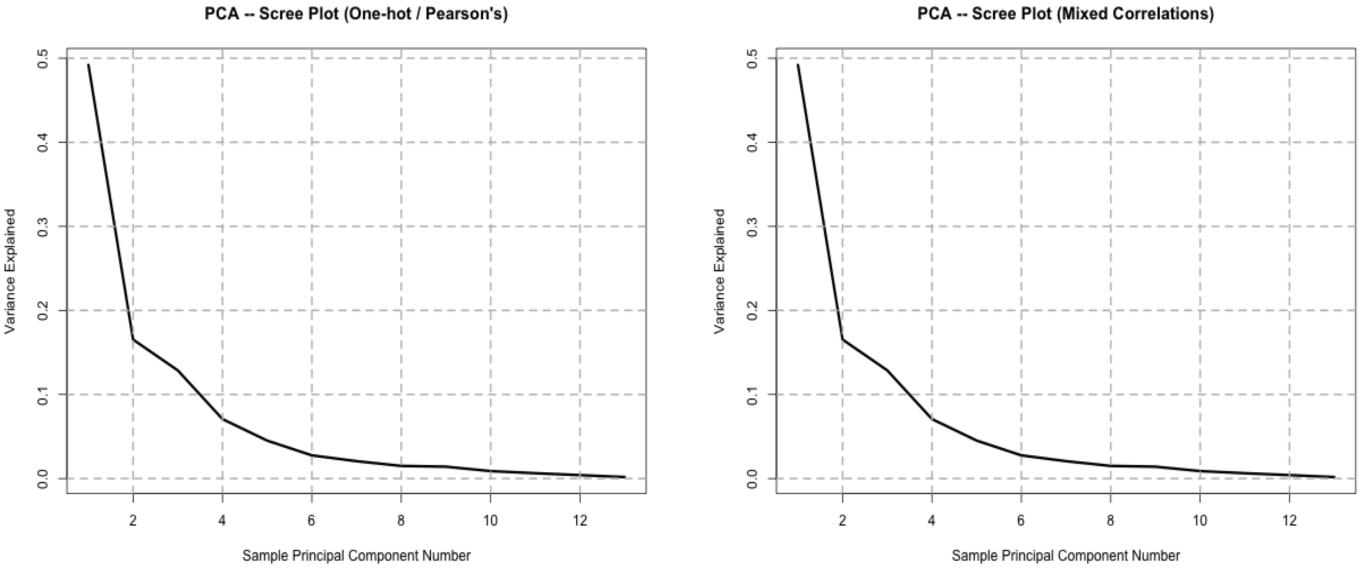
Figure 2: Two-way Relationships



(a) Correlation Plot

(b) VIF Values

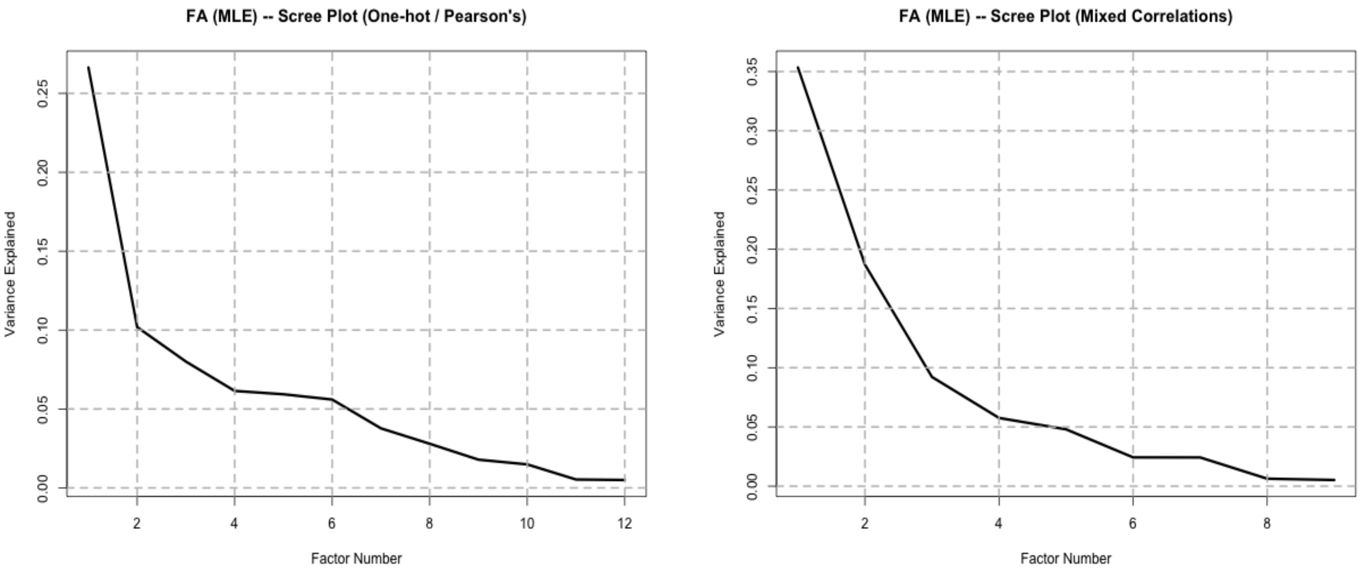
Figure 3: Correlation and Multicollinearity Assessment



(a) One-Hot Encoding / Pearson's Correlations

(b) Mixed Correlations

Figure 4: PCA Scree Plots



(a) One-Hot Encoding / Pearson's Correlations

(b) Mixed Correlations

Figure 5: Factor Analysis (MLE) Scree Plots

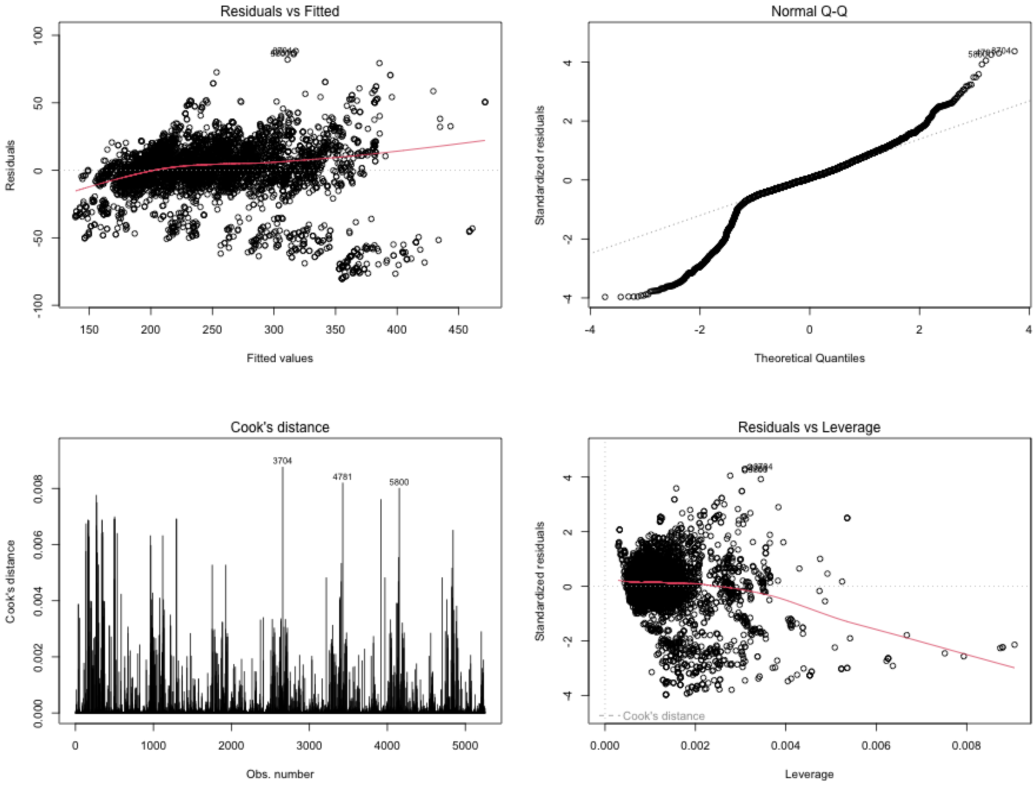


Figure 6: Residual Diagnostics and Outliers Plots Example (PCA w/ Regression)

Variable Name	PC1	PC2	PC3	PC4	PC5	PC6
Vehicle.Class.FULL.SIZE	0.0141	0.0613	-0.0348	-0.0767	-0.0367	-0.0270
Vehicle.Class.LARGE.VEHICLE	0.0647	**0.3906**	-0.0180	**0.5160**	-0.1368	0.4060
Vehicle.Class.MID.SIZE	-0.0555	0.1701	-0.0662	-0.2006	0.0003	-0.1646
Engine.Size	**0.4872**	0.0641	0.0780	-0.2918	-0.3426	0.1096
Cylinders	**0.4785**	0.0466	0.0091	**-0.4433**	-0.2486	0.1764
Fuel.Type.Regular	-0.10005	**0.5293**	0.0416	0.2172	**-0.5150**	-0.0404
Fuel.Consumption.City	**0.4994**	0.0478	0.0864	0.1770	0.2791	-0.2643
Fuel.Consumption.Hwy	**0.4764**	0.0642	0.1647	**0.4044**	**0.3213**	-0.1067
Econ.Class.Luxury	0.0236	0.0183	-0.0049	-0.0512	-0.0225	-0.0502
Econ.Class.Premium	0.0070	0.2399	-0.3010	-0.1990	**0.4964**	**0.6463**
Drive.Two.Wheel.Drive	-0.0474	**0.6356**	-0.2009	-0.2761	0.2454	**-0.387**
Transmission.Type.Man	-0.0457	0.1435	-0.0090	-0.0530	0.1114	-0.2949
Number.Of.Gears	0.1872	-0.1983	**-0.9058**	0.2063	-0.1868	-0.1560

Table 1: Principal Component Weights (One-Hot Encoding / Pearson's Correlations)

Variable Name	PC1	PC2	PC3	PC4	PC5	PC6
Vehicle.Class	-0.0086	**0.4091**	0.2681	**0.7704**	0.0454	**0.3602**
Engine.Size	**0.4407**	0.1482	-0.1941	0.0490	-0.0098	-0.2482
Cylinders	**0.4448**	0.0440	-0.1862	0.0661	-0.0705	**-0.3135**
Fuel.Type	-0.2840	**0.5183**	-0.0528	-0.1058	0.1428	**-0.3330**
Fuel.Consumption.City	**0.4548**	0.1271	-0.1917	-0.0462	0.0458	0.2235
Fuel.Consumption.Hwy	**0.4296**	0.1979	-0.1423	-0.0887	-0.0073	**0.3360**
Econ.Class	0.1638	**-0.5570**	0.2838	0.1427	**-0.3427**	0.1149
Drive	-0.1050	-0.2945	-0.5182	**0.5926**	0.0638	**-0.3552**
Transmission.Type	-0.2177	-0.1728	**-0.5670**	-0.0792	**0.3289**	**0.5292**
Number.Of.Gears	0.2247	-0.2412	**0.3546**	0.0278	**0.8606**	-0.1315

Table 2: Principal Component Weights (Mixed Correlations)

Variable Name	L1	L2	L3	L4	L5	L6	L7	L8
Vehicle.Class.FULL.SIZE	0.0604	0.0003	0.0746	**0.7405**	0.0271	0.0552	0.0016	0.0262
Vehicle.Class.LARGE.VEHICLE	0.1999	-0.2798	**0.5929**	-0.4489	-0.1454	0.4574	0.0753	-0.0423
Vehicle.Class.MID.SIZE	-0.1324	-0.0077	**-0.7625**	-0.1213	0.0043	0.0018	0.0078	-0.0494
Engine.Size	**0.9216**	-0.0018	0.0269	0.0723	0.1074	0.1385	-0.0031	0.1053
Cylinders	**0.8659**	0.1264	0.0281	0.1114	**0.3129**	0.1152	0.0362	0.1266
Fuel.Type.Regular	-0.2264	**-0.5746**	0.0181	-0.0056	-0.1189	0.0835	**0.6555**	-0.2666
Fuel.Consumption.City	**0.9293**	0.0090	0.1439	0.0291	0.0892	0.0737	-0.1032	0.1479
Fuel.Consumption.Hwy	**0.8900**	-0.0186	0.2307	-0.0869	0.0276	0.1621	-0.1410	-0.0190
Econ.Class.Luxury	0.2182	-0.0564	-0.0300	0.0385	**0.7688**	-0.0210	-0.0453	0.0493
Econ.Class.Premium	-0.0196	**0.9151**	-0.0422	0.0161	-0.0890	0.0557	-0.0864	0.1345
Crive.Two.Wheel.Drive	-0.0638	0.0481	-0.1430	0.0367	0.1178	-0.3333	-0.0420	-0.0728
Transmission.Type.Man	-0.1221	-0.0570	0.0077	-0.0578	-0.0238	**-0.5487**	-0.0066	-0.0800
Number.Of.Gears	0.1681	0.2310	0.0527	0.0394	0.0589	0.1684	-0.1079	**0.4536**

Table 3: MLE Factor Loadings (One-Hot Encoding / Pearson's Correlations)

Variable Name	L1	L2	L3	L4	L5	L6	L7	L8
Vehicle.Class	0.0172	-0.1780	-0.0688	**0.9727**	-0.0202	-0.1288	0.0079	-0.0017
Engine.Size	**0.9442**	0.0381	-0.0139	0.0173	0.0822	-0.1217	-0.2033	-0.0108
Cylinders	**0.9149**	0.1632	0.0411	-0.0433	0.0909	-0.1307	-0.2360	0.2189
Fuel.Type	0.2883	**-0.8914**	-0.0928	0.1442	-0.1953	-0.0457	-0.0968	0.0596
Fuel.Consumption.City	**0.9348**	0.0982	-0.0729	0.0108	0.0998	-0.0448	0.2144	-0.1735
Fuel.Consumption.Hwy	**0.8837**	0.0438	-0.1600	0.0567	0.0411	-0.0737	**0.4239**	-0.0206
Econ.Class	0.0224	**0.9614**	0.0199	-0.1056	0.1320	-0.1292	-0.0724	0.0505
Drive	0.0797	0.0790	**0.9679**	-0.0693	-0.0593	0.2040	-0.0206	0.0032
Transmission.Type	0.1903	-0.0823	0.2289	-0.1458	-0.1049	**0.9339**	-0.0030	-0.0040
Number.Of.Gears	0.1450	0.2469	-0.0641	-0.0207	**0.9504**	-0.1005	0.0031	0.0003

Table 4: MLE Factor Loadings (Mixed Correlations)

	Dependent variable:			
	CO2.Emissions			
	(PCA w/ One-Hot)	(FA w/ One-Hot)	(PCA w/ Mixed)	(FA w/ Mixed)
PC1/F1	28.698^{***}	52.115^{***}	27.533^{***}	51.204^{***}
	(0.147)	(0.295)	(0.194)	(0.307)
PC2/F2	9.664^{***}	4.947^{***}	6.7644^{***}	-1.226^{*}
	(0.804)	(0.529)	(0.333)	(0.503)
PC3/F3	0.0004	14.689^{***}	-6.328^{***}	-12.594^{***}
	(0.372)	(0.495)	(0.519)	(0.441)
PC4/F4	6.932^{***}	-3.5205^{***}	6.932^{***}	4.866^{***}
	(0.387)	(0.464)	(0.387)	(0.296)
PC5/F5	2.399^{***}	0.845^{*}	2.4714^{***}	4.636^{***}
	(0.489)	(0.379)	(0.418)	(0.256)
PC6/F6	3.633^{***}	10.774^{***}	7.429^{***}	-2.586^{***}
	(0.641)	(0.322)	(0.566)	(0.261)
PC7/F7		2.308^{***}		0.966^{***}
		(0.440)		(0.249)
PC8/F8		5.719^{***}		2.198^{***}
		(0.226)		(0.196)
Constant	241.767^{***}	236.033^{***}	266.8727^{***}	232.776^{***}
	(0.926)	(0.686)	(0.9161)	(1.601)
Observations	5,253	5,253	5,253	5,253
Adjusted R^2	0.883	0.888	0.873	0.889
Residual Std. Error	20.27	19.84	21.13	19.72
F Statistic	$6{,}588^{***}$	$5{,}187^{***}$	$5{,}990^{***}$	$5{,}257^{***}$
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Linear Regression Coefficients for Each Model