Project Name: Factors Contributing to the Sales of Electric Vehicles in the United States

Huma Meer

1. Business Question and Case

1.1. Business Question: What are the factors that have an impact on the sales of electric vehicles (EV) in the US?

1.2. Business Case: EVs have the potential to transform the transportation sector, which is currently one of the highest carbon-emitting sectors in the United States (EPA). A global comparison study has recently shown that CO2 emissions from Battery Electric Vehicles are 66% lower in Europe, 60% lower in the US and 45% lower in China, as compared to gasoline powered vehicles (ICCT). Studying the factors that influence EV sales would be of particular interest to countries, especially the United States, that are hastening to get ahead in the race for transitioning to EVs from gasoline powered cars in their effort to reach the United Nations Sustainable Development Goal of Clean Energy (United Nations). The stakeholders who have a vested economic and environmental interest in the contributing factors of EV sales in the US include: the government, charging station operators, consumers, and manufacturers in the EV market. The US government can use the finding of this study regarding the influence of tax rebate on EV sales to inform the state and federal level incentivization policies. The Pew Research Center claims that EV sales dropped by 3.2% in 2020, as compared to 2019, largely due to the phaseout of federal tax credits on some of the most popular EV models (DeSilver). If this effect of tax rebate on EV sales is in infact supported by this study, then the state and local governments can confidently use tax credit as a tool to impact EV sales. Additionally, EV manufacturers and charging station operators can also use the findings of this study to make more informed investment decisions; if it is found that EV Sales are higher in areas with more charging stations or where the oil price per gallon is high, then EV dealerships may invest more in marketing for EV sales in those areas.

2. The Analytics Question: Do factors such as government incentives (ZEV mandates and EV tax credits by state), the political affiliation of the state (Democrat or Republican), oil price per gallon, charging infrastructure (the number of charging stations per state), median household income and and Green Score have an impact on the sale of Electric Vehicles in the US, on a state level? If so, then what is the extent to which these variables cause the variance in EV Sales?

2.1. Outcome Variable of Interest: Cumulative number of Sales of Electric Vehicles (sum of Battery EVs and Plug-in Hybrid EVs) from 2011 to September 2021 by State. This is a count or discrete variable because it is in units of EVs sold.

2.2. Main Predictors: The main set of predictors were chosen based on the business understanding of the factors that may influence EV sales in the US. The main predictors include two binary variables to represent the government incentivization policies: *ZEV* represents if the state has adopted the Zero Emissions Vehicle Mandate and *Tax* represents if the state gives tax rebate for purchasing an EV. The continuous variable *Gas* represents the price of gasoline per gallon. *Solar,* which represents the percentage of solar installations per household, was included to gauge if conscious and climate-friendly choice of individual households has any impact on EV sales. Additionally, the continous variable *portstopop* represents the total number of EV charging ports in a state per 100,000 people. Moreover, *PoliticalParty20* represents the political affiliation of the state based on the 2020 elections and *Income* represents the median household income in the state.

2.3. Analytics Goals and Criteria: Our first analytics goal is interpretability; we aim to create a model, the results of which can be easily explained to and understood by government officials, policymakers, EV manufacturers, consumers and the like. Our second goal is inference; we aim to test if the variables: *ZEV, Tax, Gas, Solar, PoliticalParty20,* and *Income* have an impact on EV sales. Particularly, our null-hypothesis is that the effect of *ZEV, Tax, Gas, Solar, PoliticalParty20,* and *Income* is zero, and our alternative hypothesis is that the effect of these variables is not zero. Additionally, we chose 'dimension reduction' as our overarching criteria for model selection because of dimensionality issues in our model (explained in 5.4).

3. Data Set Description: The data used for this study has been extracted from EVAdoption, a data repository website that collects data regarding trends and patterns in the EV Industry. The original data on the website has 51 observations which includes 50 states and the District of Columbia. However, the District of Columbia was removed from the dataset because of several missing values. Our final data has 50 observations, 11 predictors, 1 outcome variable which is EV Sales, and 1 unique identifier of each observation which is the name of the state. None of the observations has any missing value. All the variables included in the dataset have been described in detail in Figure 1, along with their types/units and the rationale for choosing them. Moreover, the data source is described in more detail in the appendix.

4. Descriptive Analytics

4.1. Descriptive Statistics of Key Variables: An initial exploratory data analysis of the outcome variable, EV Sales (*EVS*), surfaced certain interesting patterns that informed the predictive modeling section of our study. Firstly, we found that the mean EV Sales per state, 43,000 units, is much higher than the median sales, 15,000 with an alarmingly high standard deviation of 131,000 units as shown in Figure 2. This suggests that a few observations (like California and Florida) are pulling the mean high. Secondly, we found, in Figure 3, that on average, EV sales are 4 times higher in states that offer a tax rebate (approximately 40,000 units) for purchasing an EV, as compared to states that do not offer a tax rebate (approximately 10,000 units.)
Lastly, we also found some interesting differences between states with Democratic and Republican party majority shown in Figure 4. Particularly, we found that tax rebate is given in 40% of Democratic states (11 of the 27 Democratic states) as compared to only 4% of the Republican states (1 of the 22 Republican states), the median household income is higher, on average, in Democratic states as compared to Republican states and the average number of EV charging ports and EV models is also higher in Democrat states as compared to Republican states.

4.2. Correlation and Covariation Analysis: A correlation analysis of our data helped us identify the direction and the strength of the relationship between our outcome and predictor variables, as well as highlighted how the predictors vary against each other. The most important insights that we derived from the correlation matrices were that Median Household Income and Gasoline Price per Gallon is strongly and positively correlated with the sale of EVs. This shows that the sales of EVs is generally higher in states with high median household income and high oil prices. Moreover, Median Household Income is also positively correlated with Green Score which suggests that wealthier states are better able to invest in 'green' infrastructure.

We also created several scatter plots with trend lines to better visualize the relationship between the log of *EVS* and predictor variables. *EVS* was logged because it is a discrete variable which made visualization difficult. Log-transforming *EVS* allowed us to view the trends on a magnified scale. The scatter plots corroborated the results of the correlation matrix. We found that EVS is positively and strongly correlated with *Gas*, *Income, Green Score, EV Models* and *portstopop*. The ANOVA (Figure 18) and boxplots (Figure 15, 16 and 17) show that, on average, the sales of EV is higher in Democratic States as compared to Republican States, States which have adopted the ZEV Mandate as compared to states without the ZEV Mandate and States that offer tax rebate for purchasing an EV as compared to those that don't. Moreover, the difference in average sales across the aforementioned categories is statistically significant.

4.3. Data Pre-Processing and Transformations: The first transformation involved the 'ZEV/ZEV Mandate' in the raw data, which contained 3 levels of value which are 0, 1, and 2. We re-coded ZEV to 2 levels. The first level was changed to 'ZEV' if the state has adopted California's Low Emissions Vehicle (LEV) standards, Zero Emissions Vehicle Standard (ZEV), or both. The second level is 'No' if the state has adopted neither LEV nor ZEV standards. The second transformation was done to create the variable, charging ports per 100,000 people or *portstopop*. This was done by using the *Total Ports* (per state) and a variable representing the *population* per state. This allowed us to avoid the data scale imbalance which would have incurred if we had used the *Total Ports* variable.

5. Modeling Methods and Model Specifications

5.1. Initial Model Specification: The first model specification is a plain OLS model which is estimated using the 'Generalized Linear Model', with an initial set of 11 predictors. The outcome variable used is *EVS*, which is a count variable representing the number of EVs sold in a state. The predictors include the binary variable *Tax*, categorical variables *ZEV* and *PoliticalParty20*, as well as the continuous variables *Truck, Income, GreenScore, Solar, Electricity,* and *Models*. The details of each of these variables can be found in Figure 1. We used GLM with Gaussian Distribution instead of the basic OLS estimation formula to enable cross-validation and model comparison which can be done using only a glm object.

5.2. Initial OLS Model Results: The initial model did not perform too well with only two significant variables, i.e, *Gas* and *Solar*, an extremely high 10-Fold Cross Validation Error of approximately 155,000 units of EVs and an R-squared value of 65%. The results indicate that on average, keeping all other variables constant, a one-dollar increase in price of oil per gallon increases the EV sales by approximately 120,000 units and a 1% increase in solar installations per household leads to an approximate 67,000 unit increase in EV Sales. Even though the model resulted in a much lower residual deviance as compared to the null deviance, the extremely high error value does not make the aforementioned predictions too reliable. This result was expected because the outcome variable needs to be log-transformed in order to fit the OLS estimation and OLS testing needs to be performed to find a better model for this analytics question.

5.3. Assumption Tests: The OLS Assumption testing, described in detail in Figure 31, revealed that the outcome variable, *EVS*, is not normally distributed because it is a discrete variable, in units of EVs sold, and it is truncated at 656 units. This is also evident in the histogram in Figure 32 in which the data is highly skewed to the right. The Shapiro-Wilk test is significant in Figure 34 and rejects the hypothesis of normality because the p-value is less than 0.05. Moreover, the errors are normally distributed as shown in the QQ-plot in Figure 35 in which most of the data aligns with the qq-line and the residual plot in Figure 36 also shows an even, cloud-like distribution. Multicollinearity is a cause of concern for this model because, even though the individual Variance Inflation Factors for all variables are less than 10, the Condition Index is 90, in Figure 37, which is way higher than the threshold of 50. This indicates that the model has high overall multicollinearity. Additionally, we tested the linearity of the relationship between the outcome and predictor variables using scatter plots (Figure 7-14) which indicated that EVS has a linear relationship with all predictor variables except *Solar*. Figure 13 suggests that there exists a polynomial relationship between *Solar* and *EVS* where *EVS* dramatically increases at lower values of *Solar* and then curves downwards and upwards for higher values of *Solar*. Moreover, no serial correlation or error dependence was found in our model as shown by the Durbin-Watson test.The average of the errors is 3.3e-13 which is very close to zero. Finally, Error variance is not constant, as suggested by the significant Breusch Pagan test and the Residuals vs. Fitted plot which shows that errors are not spread evenly, confirming the presence of heteroskedasticity.

5.4. Model Candidates and Rationale: To improve upon our initial OLS model we estimated the specifications using Ridge Regression (Figure 23-26), Principal Component Analysis (Figure 27a, 28a, 29a, and 30a) and Partial Least Squares Regression (Figure 27b, 28b, 29b, and 30b). The driving factor behind this combination of models is that our OLS model suffered from severe multicollinearity and dimensionality issues which made the model prone to high variance and instability when tested on new data. This can be credited to the fact that it had an alarmingly low degrees of freedom with 50 observations and 11 predictors, which can cause the model to be overfit. Dimension reduction methods like Ridge, PCR and PLSR allowed us to reduce dimensionality by shrinking the predictor coefficients without having to drop any variables. Essentially, we deliberately introduced bias in the model to reduce variance, while also improving the predictive accuracy of the model.

5.5. Model Specification Candidates and Rationale: Our model specification selection was also largely influenced by the issues that surfaced during OLS assumption testing. The first model specification included log-transforming our outcome variable, *EVS*. Again, this was done to transform *EVS*, which was a count variable, to a continuous variable with normal distribution. Secondly, we added a quadratic term of *Solar* because it was obvious from the scatter plot that *Solar* and *EVS* do not have a linear relationship. It can be argued that adding a quadratic variable to our specification would further increase dimensionality in our model but after running several CV tests we noticed that our final model has a lower RMSE with the quadratic variable as compared to this model without the quadratic Solar variable. Lastly, we used Stepwise Selection as a variable method on the OLS model with the log-transformed EVS, the quadratic Solar variable along with the initial set of 11 predictors. This resulted in a smaller set of variables (all significant at the 10% level) became our third specification which we used to estimate the Ridge, PCR and PLSR models. This allowed us to identify an optimal model with a smaller set of predictors which include *ZEV, Truck, Models, GreenScore, PoliticalParty20, Solar,* and *Solar ^2,* which we will call the 'Stepwise Selected' predictors. Each of these sets of specifications was progressively engineered, i.e, we did not make all the changes in one go. We estimated each level of specification with each model (OLS, Ridge, PCR/PLSR) and evaluated the CV at each level. This chart illustrates our model methods and specifications more clearly.

5.6. Cross-Validation Testing and Final Model Selection: We tested all 12 combinations of models and specifications using the 10-Fold Cross Validation method because we had a very small sample and using other methods like Random Splitting would give us highly unstable CV values. All the 8 combinations are included in Chart 2 with the CV RMSE mentioned under each method. For our PCR/PLSR models, we compared the CV values of each component at every level of specification and chose the component that explained at least 70% of the variance in the outcome variable with a tolerable CV value, keeping in line with our objective of reducing dimensionality. The component chosen for comparison in each model is also mentioned under each model. The final model that we chose, based on our comparison of the CV errors, was the Ridge regression model which was modeled using the log-transformed EVS and the smaller set of predictors identified by the Stepwise Selection Method. This model yielded the lowest CV RMSE of 0.728.

6. Analysis of Results

Ridge Regression with Stepwise Selected variables using the log-transformed EVS is our final model as it provided the lowest CV RMSE. We chose the Lambda which resulted in the lowest possible RMSE of 0.112, which is very close to zero. This suggests that our Ridge model is very close to the OLS model making our prediction more interpretable. Additionally, since this model employs only the Stepwise Selected variables which were all selected using the 10% significance level threshold, the variables in this model are statistically significant. Moreover, the direction of the effect that each variable in our final model is robust, as shown in this table, which depicts that the signs of the variables do not change regardless of the model we choose.

6.1. ***Interpretation and Analysis of Quantitative Predictors (all interpretations are "on average and holding everything else constant"):*** The number of EV models available in a state has a positive effect on EV Sales, with a 22% increase in sales with every additional model introduced. The linear effect of the Percentage of solar installations per household is negative, with a 1% increase in solar installation per household leading to a 19% decrease in EV sales. The quadratic effect of Solar on EV sales, on the other hand, is positive. This shows that households generally do not invest in purchasing EV cars as they begin installing 'green' technology like solar panels in their homes. However, as they increase their consumption of solar technology in their households, they also become more inclined to purchase an electric vehicle. This result does not come as a surprise because ‘green’ technology is still a relatively new innovation and many households are still weary of utilizing it. It is when customers realize the cost savings incurred by using solar energy as opposed to the traditional fuel, they may become more comfortable with making more environmentally friendly purchasing decisions. GreenScore has a positive impact on EV sales, with a 1-point increase in green score leading to a 2.1% increase in EV Sales. This suggests that states that have more environmentally friendly initiatives tend to have higher EV Sales. Moreover, a 1-percent increase in Light Trucks purchased in a state leads to a 7% decrease in EV sales. This suggests that light vehicles, like SUVs and pickup trucks, and EV are substitutes for each other. This can be attributed to the fact that Light-duty vehicles are currently responsible for 58% of US transportation sector emissions which has motivated state and local governments to push for replacing light trucks with EVs (Lewis). Therefore, the sales of EVs and light trucks vary in opposite directions.

6.2. Interpretation and Analysis of Qualitative Predictors (all interpretations are "on average and holding everything else constant"): One of the most astounding results of this study is that states that have adopted the ZEV mandate have EV sales 135% lower than states that have not adopted the ZEV mandate. This negative impact of the ZEV mandate is robust because it is consistent with almost all our models. This is because the ZEV mandate, despite the portrayals found in the press where its impact on carbon emission has been lauded, does not directly stimulate customer purchasing behavior; it merely drives the production of EVs. The ZEV mandate is, essentially a state regulation, adopted by 12 states, that requires manufacturers to reach a certain number of "ZEV credits", and not actual ZEV sales or production. Those credits can be reached by producing fewer ZEV cars that are more efficient which does not necessarily increase EV production or sales (UCSUSA). Therefore, while the ZEV mandate may have an (indirect) impact on production of EVs, our model shows that the ZEV mandate does not have a positive impact on sales of EVs. Finally, the Republican Majority states have EV sales 377% lower than Democratic majority states. This was expected because generally speaking the democratic party has more aggressive pro-EV policies.

7. Conclusions and Lessons Learned

7.1. Conclusions from the Analysis: This project aimed at determining the factors that contributed to the sales of EVs. Particularly, we were interested in producing interpretable results and testing whether *ZEV, Tax, Gas, Solar, Models, PoliticalParty20,* and *Income* have an impact on the EV sales. We found that the number of available EV models has a significant and positive impact on EV Sales which suggests that when customers have a greater range of models to choose from, they are more likely to purchase EVs. The other major contributor is Solar; as people consume more solar technology, they become more inclined towards buying EV cars. The third major factor is Green Score. The states that have more environment friendly initiatives promote more EV sales than states that don't have environmentally friendly initiatives. Additionally, our findings regarding the ZEV mandates suggests that states that are interested in increasing EV sales need to offer incentivization policies that directly impact the EV sale by incentivizing consumers instead of incentivizing producers only. Finally, our analysis shows that variables such as median household income, oil price per gallon, tax rebate and the ratio of charging ports per 100,000 people do not have a significant impact on the sales of EV because these variables were dropped during the Stepwise Selection process. In terms of hypothesis mentioned in section 2.3, we can confirm that we reject the null hypothesis for *ZEV, Solar* and *Political Party* and we fail to reject the null-hypothesis for *Tax, Gas* and *Income.*

7.2. Project Issues, Challenges and Lessons Learned

There were three major issues that we faced in this project. There was an issue of dimensionality and low degrees of freedom because of limited observations which meant that our model was at risk of being overfit and highly unstable. We tried to avoid this problem by using the Stepwise Selection method to drop the variables that were insignificant. This also helped us reduce multicollinearity which we found, in the OLS testing, was a result of high dimensionality. We then shrunk our coefficients using the Ridge regression model. This enabled us to reduce dimensionality while also retaining variables that are important for our analysis.

Another issue we struggled with was explaining the counterintuitive effect of the ZEV mandate. We expected a positive impact of the mandate on the sales of EVS because of our initial literature review. However, our models showed that it has a negative and robust impact on EV sales. We explained this impact by digging deeper into the requirements of the ZEV mandate and recognised that it does not directly impact the sales of EVS, rather the production of EVs.

Finally, we also found it difficult to conduct cross validation testing on our models because we have very limited observations (i.e, 50 observations) with a few extremely high outliers like Florida and California. This made it challenging to conduct cross validation using the random splitting method because our results varied significantly when we changed the split. We addressed this issue using the 10-Fold Cross validation method which repeated the splitting and testing 10 times, each time using a different split.

Appendices

A. Data Information

Figure 1: Predictor Variables and Rationale

Predictor Variables	Detail	Rationale for Selection
Oil Price per Gallon	Continuous variable in USD	Given the scope of this project, this predictor seems like a reasonable choice because EV are direct substitutes of gasoline powered cars. It is safe to assume that as gasoline prices rise, people might start preferring EV cars to save money. This is also worth exploring given the status of global political turmoil and its resulting impact on the oil prices.
Charging Ports per 100,000 People, by State	Count/Discrete Variable	This is critical when comparing the EV sales by State because it is safe to assume that with a better charging infrastructure EV users would feel more comfortable with driving EV cars as they would have the assurity of finding a charging port nearby if they run out of power
EV Tax Credit	Binary variable that takes the value of 1 if the state gives tax credit for owning an EV, 0 otherwise	This variable determines the impact of incentivization policies on a State/Federal level and how it might impact the EV sales

ZEV Mandate	Categorical variable that assumes one of the following levels: **-** 'ZEV' if the state has adopted California's Low Emissions Vehicle (LEV) standards, Zero Emissions Vehicle Standard (ZEV) or both); -'No' if the state has adopted neither LEV nor ZEV standards.	This variable also represents the impact of government incentivization policies. https://www.ucsusa.org/resources/what-zev
Median Household Income .	Continuous variable in USD	This is essential to study the socioeconomic factors linked with the purchase of an EV. This is also important from a policymaking perspective
Political Affiliation of the State based on 2020 Elections	Categorical variable that assumes two of the following levels: Democrat or Republican	Given the politicized nature of the topic of climate change in the United States, it seems timely and justified to explore the political affiliation by State and whether it has any impact on the EV sales or not;
Green Score of the State	Continuous variable developed by WalletHub to assess the eco-friendliness of a state, graded on a 100-point scale, with a score of 100 representing the highest level of eco-friendliness	This predictor is selected to explore any potential links between EV sales and the existing initiatives for climate change in each State;
Percentage of Solar installation per household	Continuous variable in percentage	This predictor will help determine whether a conscious and climate-friendly choice of individual households has any impact on EV sales

EV Models available per state	Continuous Variable	This variable will help gauge if the availability of a high number of EV models is a factor in EV sales.
Percentage of Light Trucks Sold (Pickups, SUVs, CUVs, and Vans – combined)	Continuous Variable	Sales of Light Trucks form the majority of US auto sales. It would be interesting to see if it influences EV Sales.
Percentage of Renewable Electricity from Grid	Continuous Variable	This predictor is selected to explore any potential links between EV sales and the existing energy infrastructure for climate change in each State.

Data Source

EVAdoption is a website which focuses on monitoring and analyzing the Electric Vehicle Industry trends and data. EVAdoption compiles and collects the data from various sources, such as, the Alliance of Auto Manufacturers, CFA Institute, US Census Bureau among others.

Figure 2: Summary Statistics

Variable	N	Mean	Median	SD of X	Min	Max
EVS	50	42818.74	14787.5	130980.311	656	930811
ZEV	50					
... No	38	76%				
... ZEV	12	24%				
Tax	50					
... 0	38	76%				
... 1	12	24%				
Gas	50	2.867	2.795	0.336	2.46	4.05
Truck	50	67.318	65.9	6.923	50.7	82.4
Income	50	61205.22	59827	9050.913	43441	81084
GreenScore	50	56.4	57.665	12.604	23.96	76.35
Solar	50	0.486	0.045	1.004	0	5
Electricity	50	22.72	12	24.46	1	100
Models	50	31.015	31.08	4.162	24.83	41.58
PoliticalParty20	50					
... Democratic	28	56%				
... Republican	22	44%				
portstopop	50	29.172	21.578	21.733	7.331	128.723

Figure 3: Summary Statistics, by Tax Rebate

Summary Statistics of States, by Tax Rebate

Characteristic	**0**, N = 38[1]	**1**, N = 12[1]
EVS	9,847 (4,267, 28,839)	43,390 (18,698, 58,522)
ZEV		
No	34 (89%)	4 (33%)
ZEV	4 (11%)	8 (67%)
Gas	2.77 (2.64, 2.90)	2.86 (2.75, 3.00)
Truck	68 (63, 74)	65 (62, 66)
Income	58,372 (54,980, 63,724)	65,500 (62,415, 72,892)
GreenScore	54 (50, 63)	68 (59, 73)
Solar	0.02 (0.01, 0.12)	0.41 (0.16, 1.58)
Electricity	12 (7, 32)	10 (5, 22)
Models	29.2 (28.1, 31.1)	34.8 (31.8, 39.6)
Charging1	2.54 (2.32, 2.71)	2.68 (2.41, 2.94)
BEVS19	765 (413, 2,821)	4,754 (1,932, 6,016)
PHEVS19	364 (174, 887)	1,790 (764, 2,239)
Port1	5 (1, 15)	20 (14, 35)
Port2	626 (265, 1,407)	2,129 (798, 3,826)
DCPorts	154 (81, 374)	437 (237, 569)
TotalPorts	830 (358, 1,729)	2,626 (1,042, 4,319)
EVtoCharger	13.2 (9.7, 17.3)	16.0 (13.9, 17.9)
PoliticalParty20		
Democratic	17 (45%)	11 (92%)
Republican	21 (55%)	1 (8.3%)

[1] Median (IQR); n (%)

Figure 4: Summary Statistics by Political Affiliation

Summary Statistics of States, by Political Affiliation

Characteristic	**Democratic**, N = 28[1]	**Republican**, N = 22[1]
EVS	27,553 (6,917, 48,161)	5,964 (1,916, 14,839)
ZEV		
No	16 (57%)	22 (100%)
ZEV	12 (43%)	0 (0%)
Tax	11 (39%)	1 (4.5%)
Gas	2.85 (2.74, 3.00)	2.66 (2.55, 2.86)
Truck	65 (61, 70)	68 (64, 76)
Income	64,207 (57,529, 72,834)	57,854 (54,980, 59,731)
GreenScore	65 (59, 69)	50 (43, 54)
Solar	0.19 (0.02, 0.81)	0.01 (0.01, 0.05)
Electricity	10 (6, 26)	12 (6, 32)
Models	31.1 (30.7, 35.6)	28.1 (27.8, 31.1)
Charging1	2.53 (2.35, 2.74)	2.59 (2.36, 2.72)
BEVS19	2,820 (642, 5,654)	604 (192, 1,768)
PHEVS19	910 (342, 1,744)	206 (92, 456)
Port1	18 (8, 54)	2 (0, 6)
Port2	1,022 (450, 2,048)	366 (184, 1,388)
DCPorts	335 (116, 502)	106 (71, 236)
TotalPorts	1,418 (534, 2,584)	624 (262, 1,664)
EVtoCharger	16.0 (13.5, 19.7)	11.1 (7.0, 14.3)

[1] Median (IQR); n (%)

B. Correlation Plots

Figure 5: Correlation Matrix

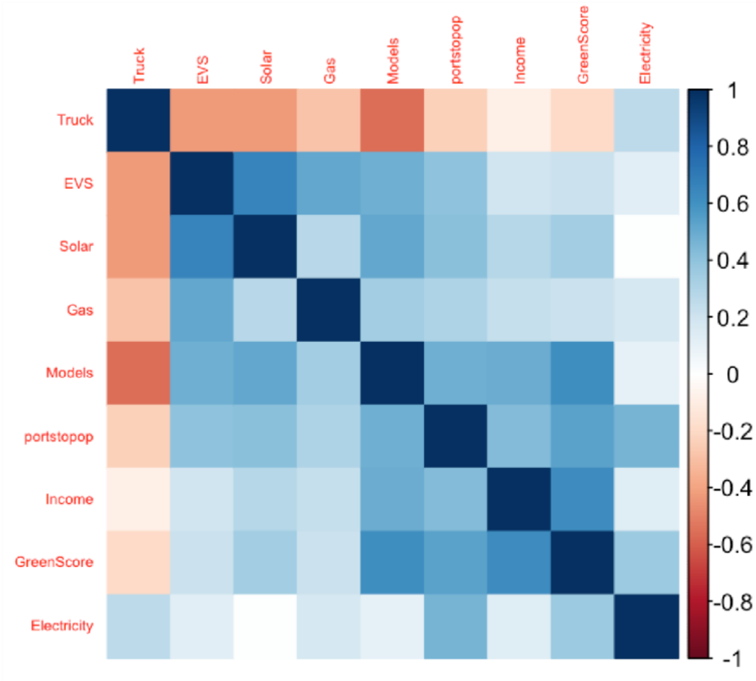
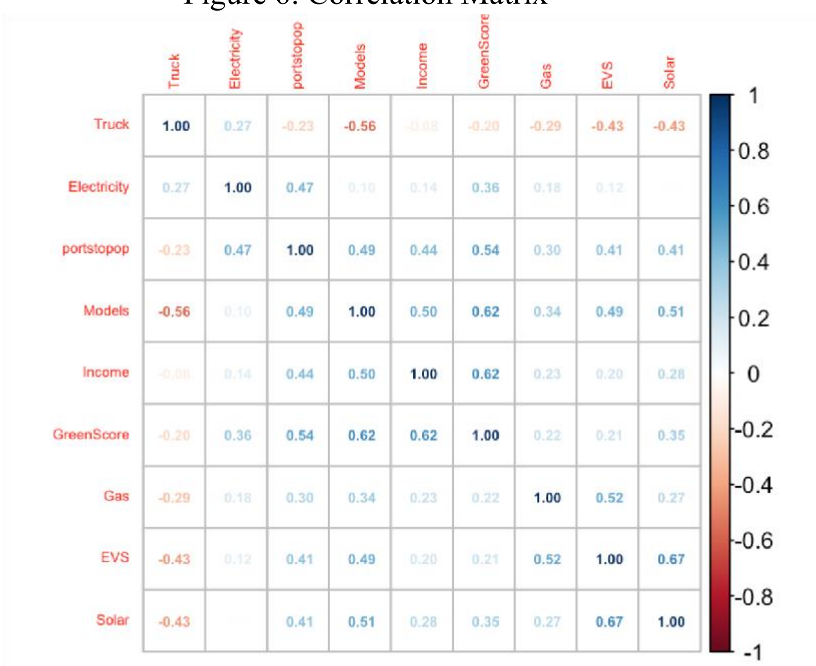


Figure 6: Correlation Matrix



Scatter Plots

Figure 7: EV Sales vs Gas price per Gallon

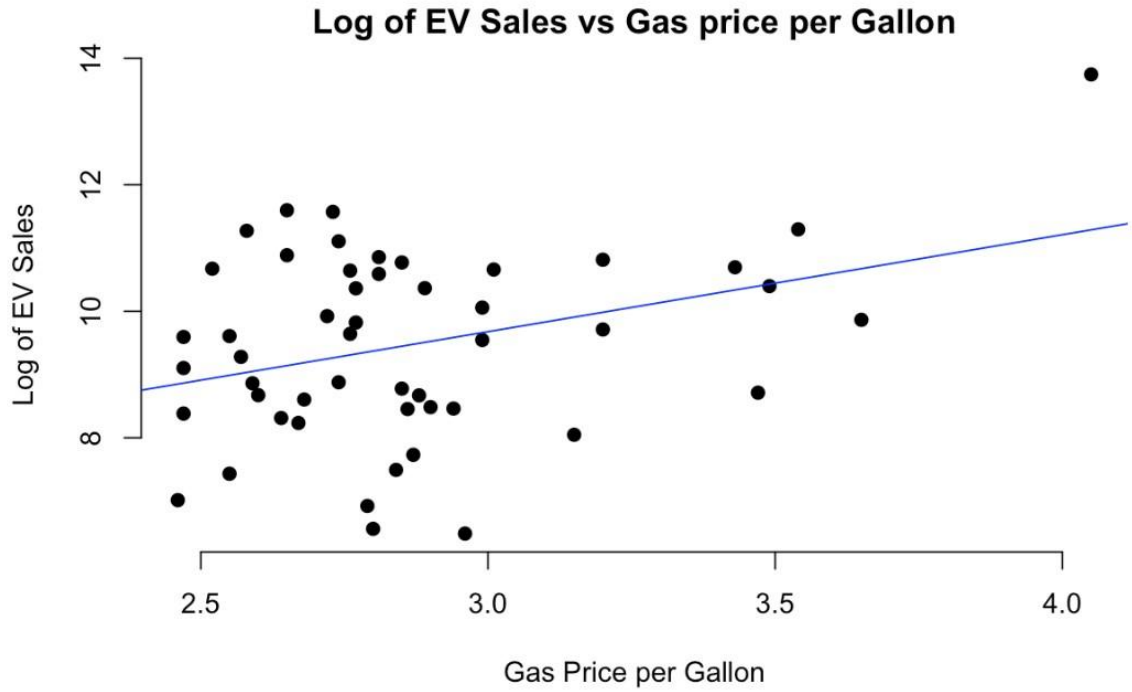


Figure 8: Sales vs Median Household Income

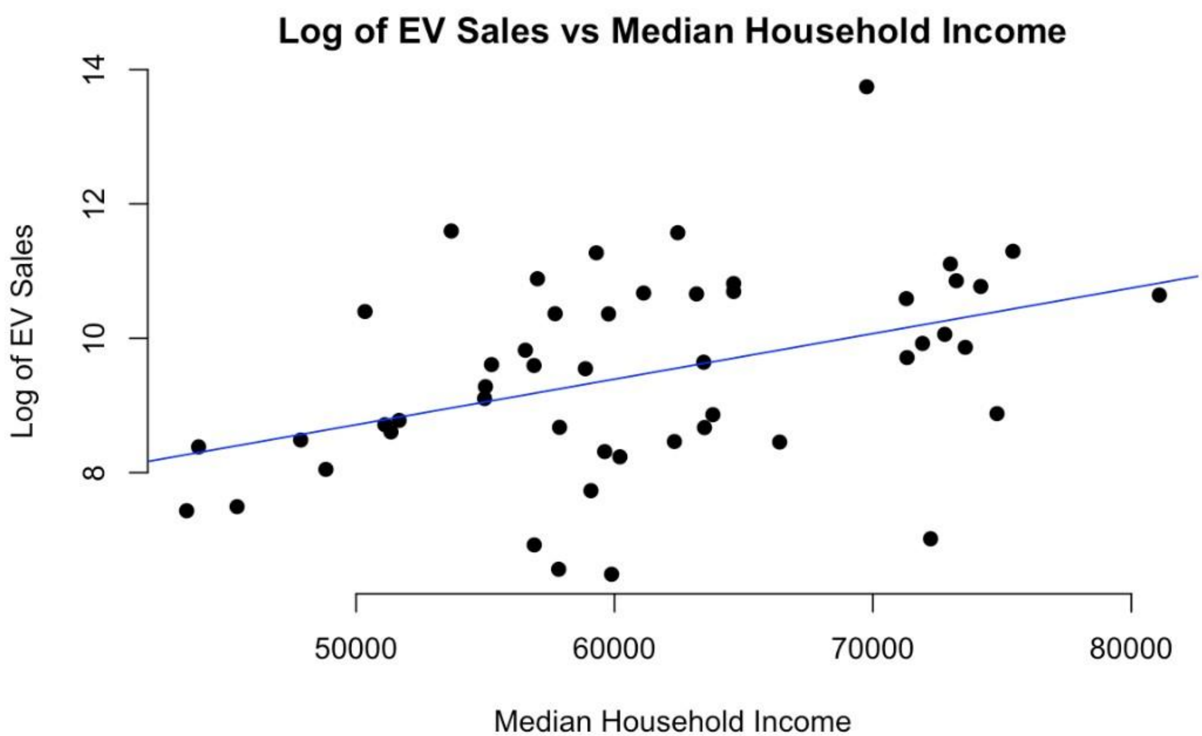


Figure 9: EV Sales vs Green Score

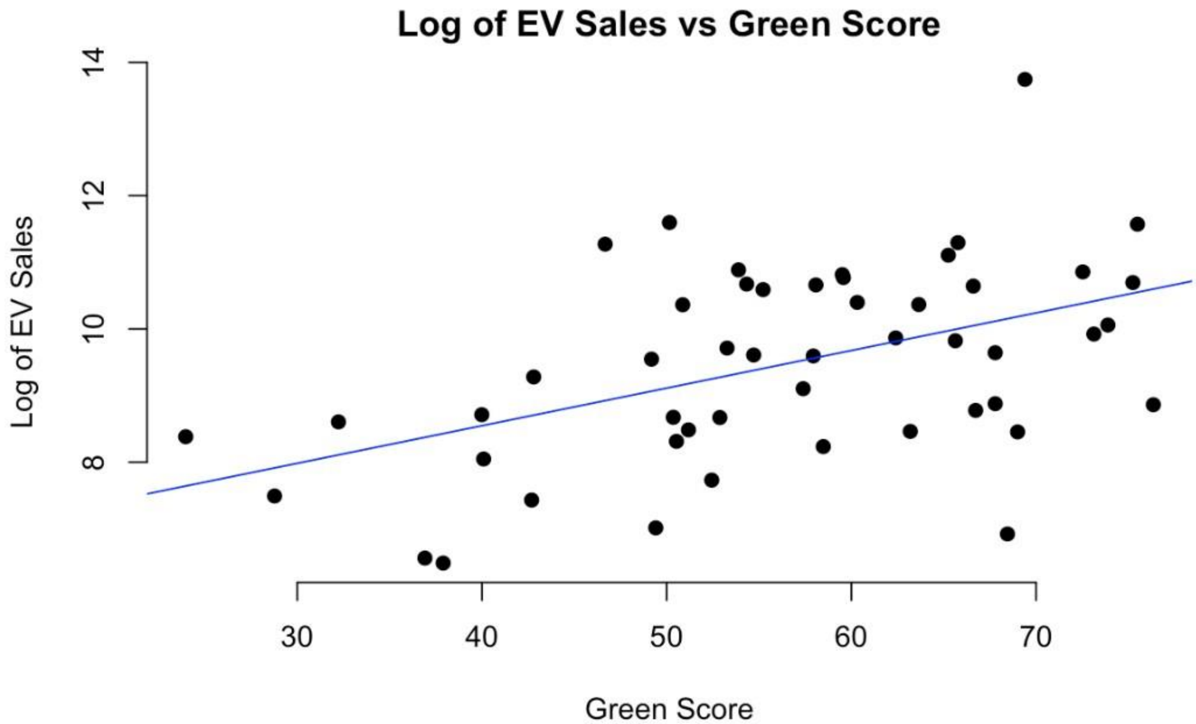


Figure 10: EV Sales vs EV Models available per state

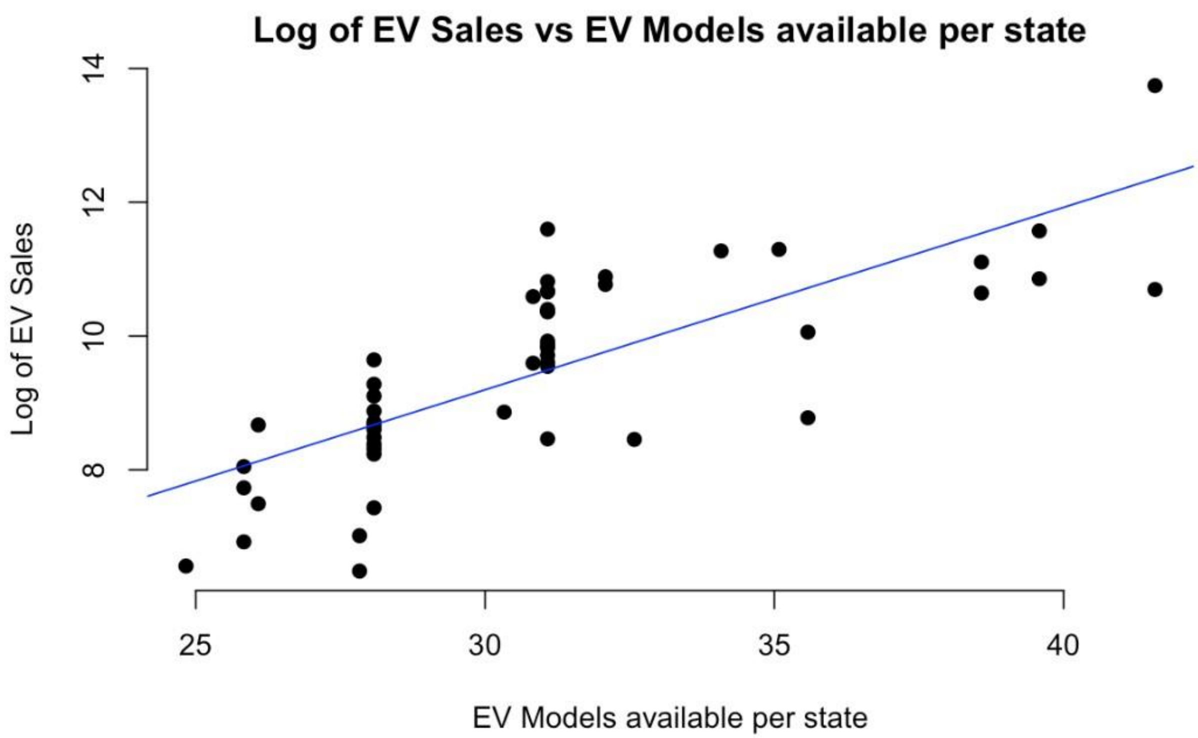


Figure 11: EV Sales vs %Renewable Electricty from grid

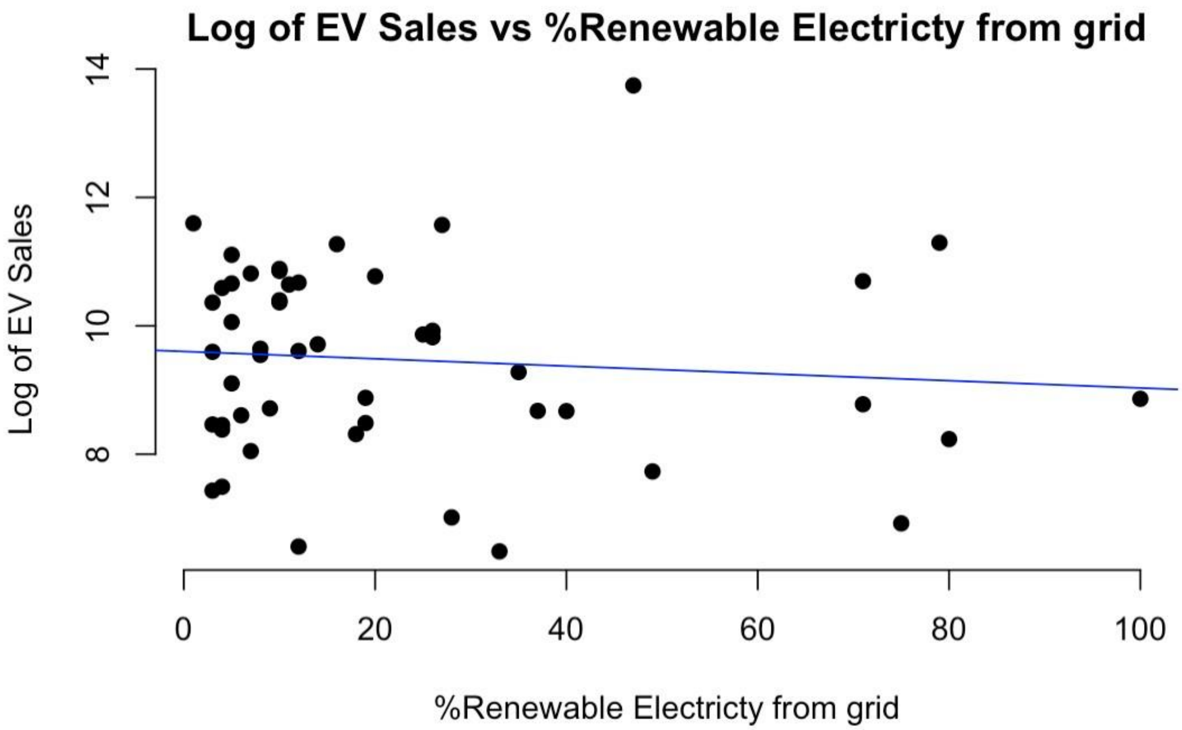


Figure 12: EV Sales vs ports per 100,000 people

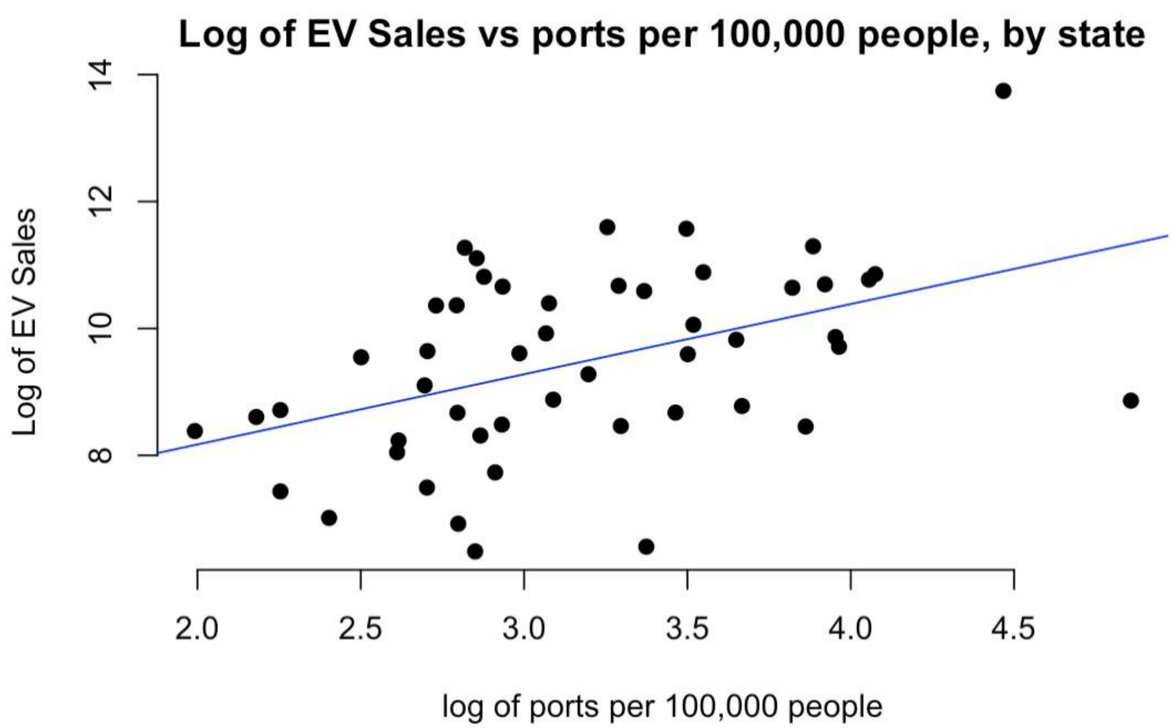


Figure 13: EV Sales vs Solar Installation per household

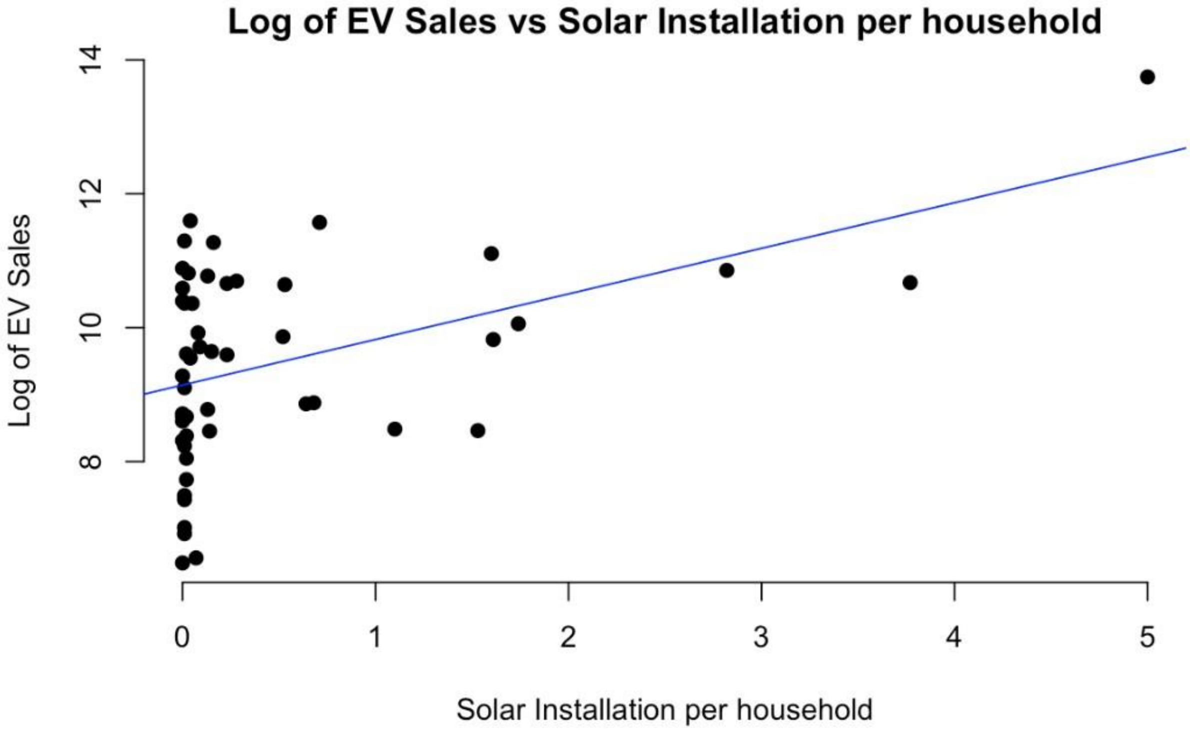
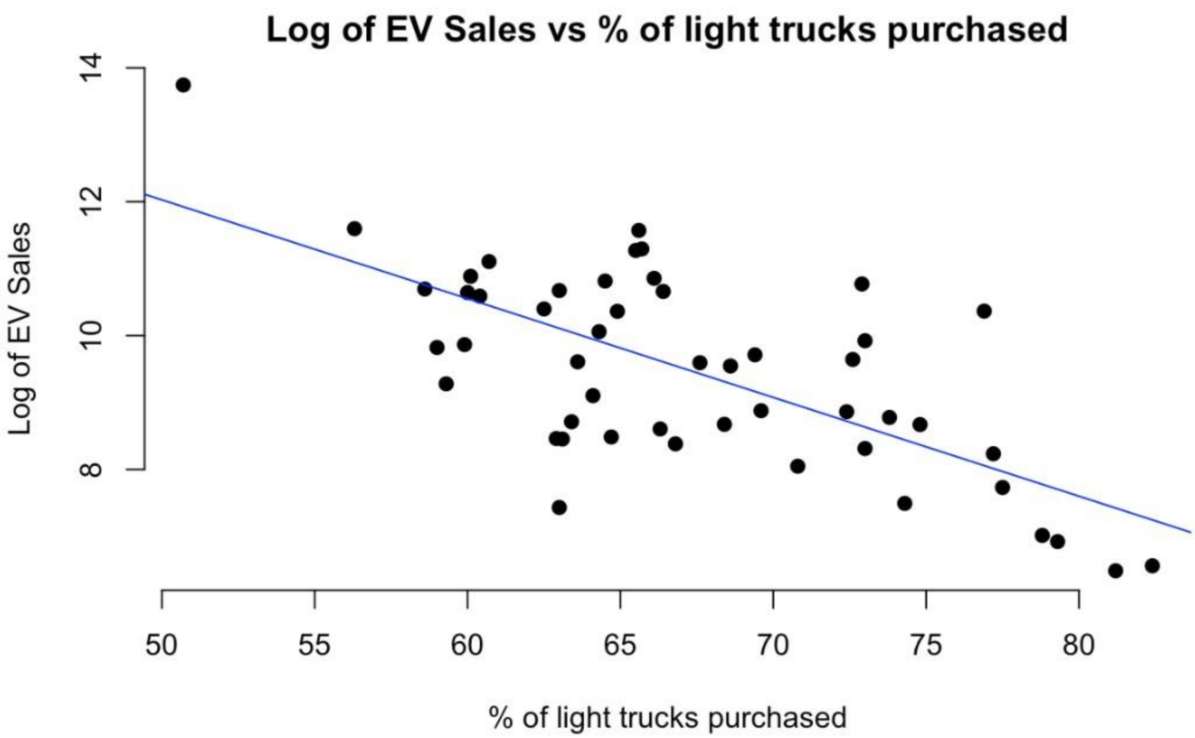
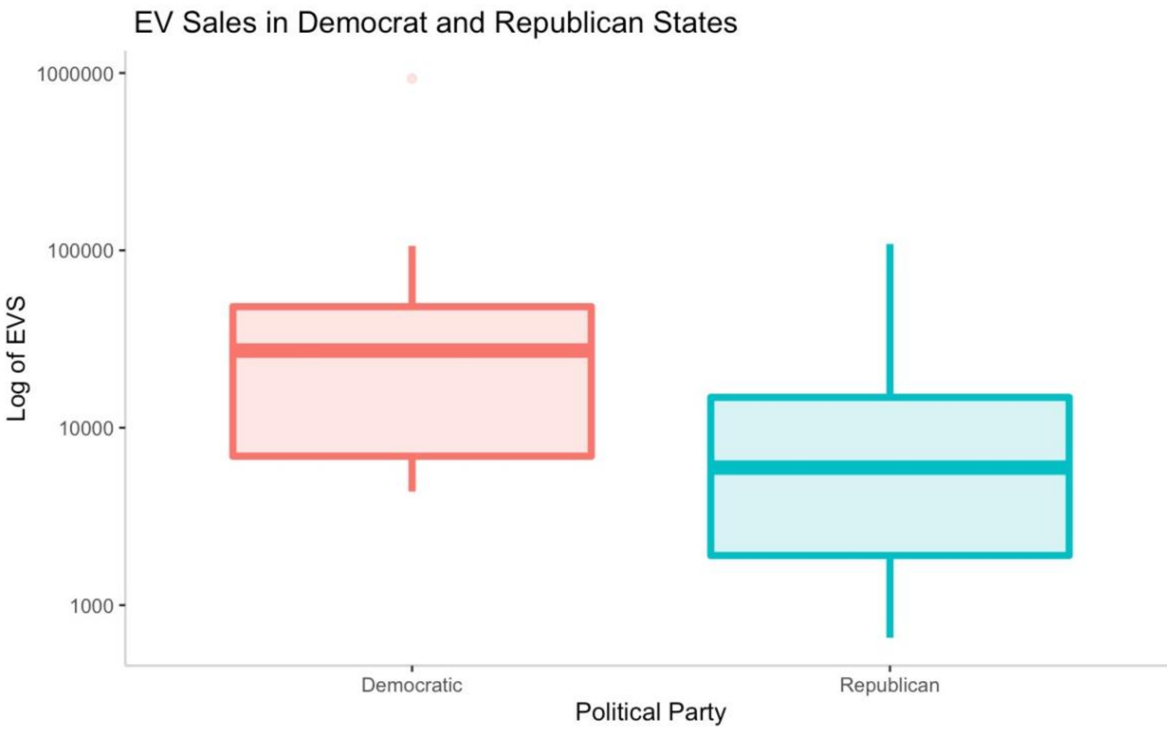


Figure 14: EV Sales vs % of light trucks purchased



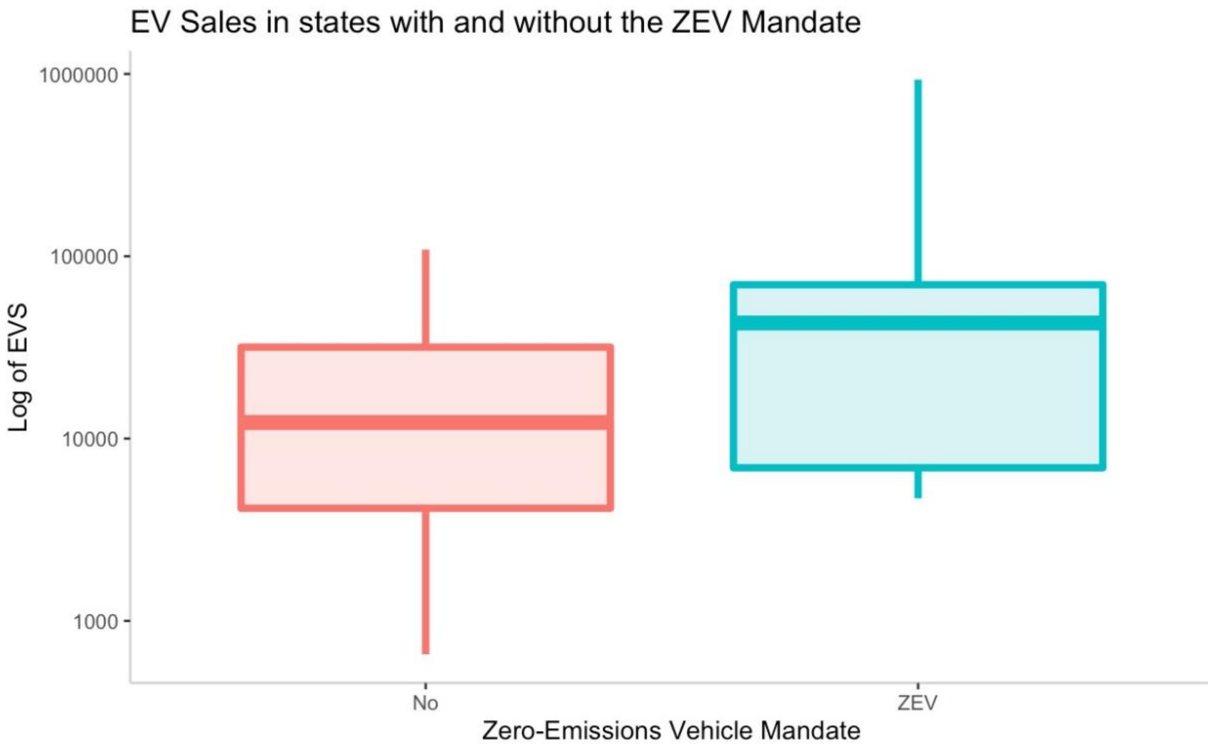
C. ANOVA and Boxplots

Figure 15: EV Sales in Democrat and Republican States



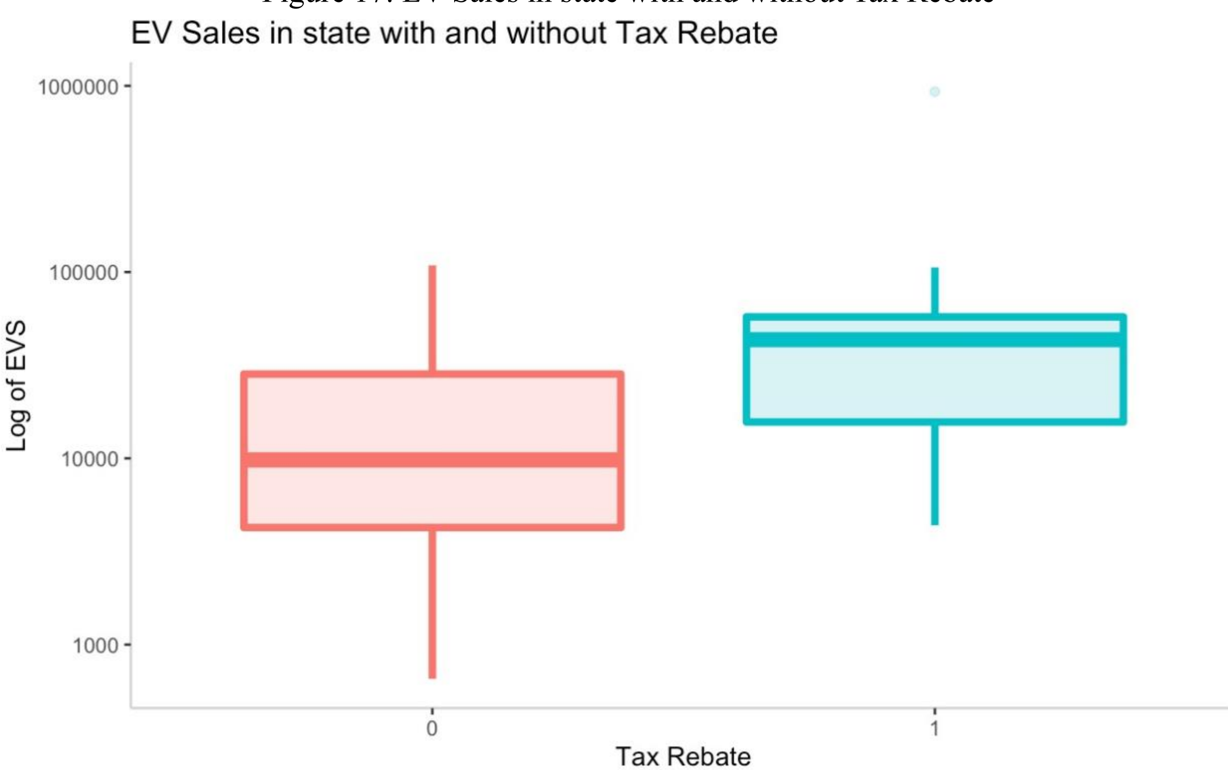
ANOVA P-Value=0.00143**

Figure 16: EV Sales in states with and without the ZEV Mandate



ANOVA P-Value= 0.0141*

Figure 17: EV Sales in state with and without Tax Rebate



ANOVA P-Value = 0.00674**

Figure 18: ANOVA of log of EVS vs Categorical Variables

```
            Df Sum Sq Mean Sq F value  Pr(>F)
Tax          1  15.51  15.511   8.022 0.00674 **
Residuals   48  92.81   1.934
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
            Df Sum Sq Mean Sq F value Pr(>F)
ZEV          1  12.90  12.903   6.491 0.0141 *
Residuals   48  95.42   1.988
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
                 Df Sum Sq Mean Sq F value  Pr(>F)
PoliticalParty20  1  20.88  20.877   11.46 0.00143 **
Residuals        48  87.44   1.822
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

D. Regression Models

Chart 1: Modeling Method and Specifications

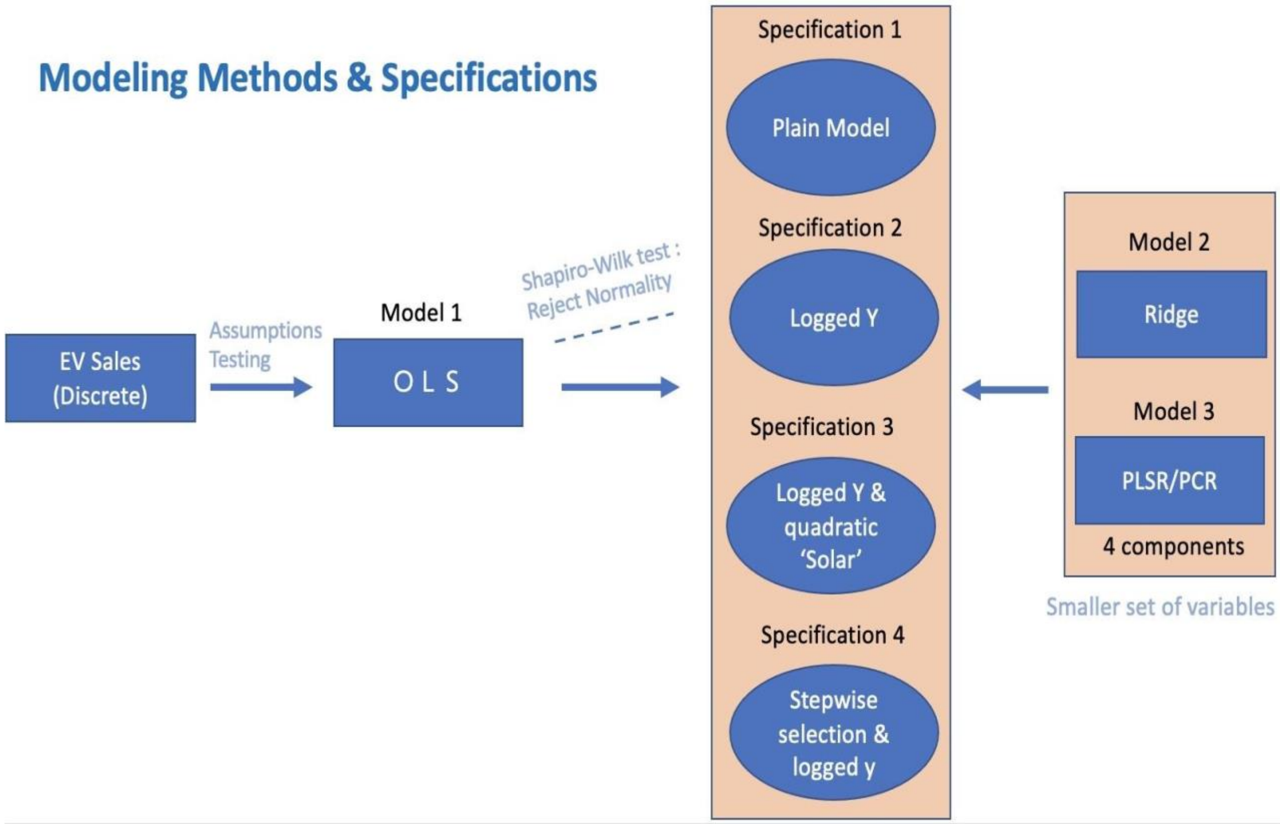


Chart 2: Modeling Method and Specifications Cross Validation

Model	Plain EVS with initial set of predictors	Logged EVS with initial set of predictors	Logged EVS with initial set of predictors and a quadratic term for Solar	Logged EVS with Stepwise Selected predictors
OLS	155,000	0.85	0.86	0.756
Ridge	132,000	0.825	0.86	0.728
PCR/PLSR	139,000 *(4 components PCR)*	0.9725 *(4 component PLSR)*	0.9035 *(4 component PLSR)*	0.7738 *(3 component PLSR)*

Chart 3: Results of three models with Specification 4 (Final Specification)

Predictor	OLS	Ridge	PLSR
(Intercept)	4.156	6.423	-
ZEV	-1.865	-1.353	-0.765
GreenScore	0.023	0.021	0.438
Solar	-0.520	-0.188	-0.200
I(Solar^2)	0.153	0.082	0.231
Models	0.291	0.226	1.031
PoliticalPartyRepublican	-0.443	-0.377	-0.189
CV. RMSE	0.7558	**0.7281**	0.7738

Figure 19: Plain OLS Model

```
Call:
glm(formula = EVS ~ ZEV + Tax + Gas + Truck + Income + GreenScore +
    Solar + Electricity + Models + portstopop + PoliticalParty20,
    family = gaussian(link = "identity"), data = EV)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-201989   -24400     4890    34289   328011

Coefficients:
                              Estimate   Std. Error t value Pr(>|t|)
(Intercept)               -466816.5508 323022.9180  -1.445 0.156612
ZEVZEV                     -90491.0155  54448.2886  -1.662 0.104749
Tax1                        30731.1872  40646.6970   0.756 0.454277
Gas                        120335.1070  43075.2205   2.794 0.008121 **
Truck                       -1317.9378   2704.4092  -0.487 0.628825
Income                         -0.5542      1.9610  -0.283 0.779025
GreenScore                  -1692.3444   1712.5509  -0.988 0.329305
Solar                       67230.5014  15738.7097   4.272 0.000125 ***
Electricity                   871.0579    747.8533   1.165 0.251383
Models                      10143.4232   6599.7224   1.537 0.132592
portstopop                    758.5862    865.9453   0.876 0.386523
PoliticalParty20Republican  18061.0795  33318.2490   0.542 0.590930
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 7842141414)

    Null deviance: 840636252992  on 49  degrees of freedom
Residual deviance: 298001373725  on 38  degrees of freedom
AIC: 1293.3

Number of Fisher Scoring iterations: 2

[1] 0.645505
[1] "10-Fold Cross Validation RMSE is 155855.8  and the R-squared value is 0.65"
```

Figure 20: OLS Model with log-transformed *EVS*

```
Call:
glm(formula = log(EVS) ~ ZEV + Tax + Gas + Truck + Income + GreenScore +
    Solar + Electricity + Models + portstopop + PoliticalParty20,
    family = gaussian(link = "identity"), data = EV)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.41358  -0.29566   0.00559   0.43738   1.52602

Coefficients:
                             Estimate   Std. Error t value    Pr(>|t|)
(Intercept)               2.723082288 2.578236452   1.056     0.29755
ZEVZEV                   -2.088799275 0.434583909  -4.806 0.000024259 ***
Tax1                     -0.136504010 0.324425266  -0.421     0.67630
Gas                       0.123830855 0.343808744   0.360     0.72071
Truck                    -0.059903401 0.021585485  -2.775     0.00851 **
Income                    0.000009994 0.000015652   0.639     0.52697
GreenScore                0.011654171 0.013668879   0.853     0.39922
Solar                     0.062174709 0.125619926   0.495     0.62349
Electricity              -0.001905889 0.005969058  -0.319     0.75125
Models                    0.310353423 0.052676277   5.892 0.000000801 ***
portstopop                0.007692172 0.006911620   1.113     0.27273
PoliticalParty20Republican -0.327717015 0.265932599  -1.232     0.22540
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.4995897)

    Null deviance: 108.321  on 49  degrees of freedom
Residual deviance:  18.984  on 38  degrees of freedom
AIC: 119.47

Number of Fisher Scoring iterations: 2

[1] "10-Fold Cross Validation RMSE is 0.8473  and the R-squared value is 0.82"
```

Figure 21: OLS Model with log-transformed *EVS* and quadratic *Solar*

```
Call:
glm(formula = log(EVS) ~ ZEV + Tax + Gas + Truck + Income + GreenScore +
    Solar + I(Solar^2) + Electricity + Models + portstopop +
    PoliticalParty20, family = gaussian(link = "identity"), data = EV)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.32600  -0.30770  -0.03155   0.41229   1.48470

Coefficients:
                              Estimate  Std. Error t value   Pr(>|t|)
(Intercept)                 3.96294607  2.55468691   1.551   0.129355
ZEVZEV                     -1.78910694  0.44356021  -4.034   0.000264 ***
Tax1                       -0.17238257  0.31254328  -0.552   0.584574
Gas                        -0.13754004  0.35511429  -0.387   0.700744
Truck                      -0.06132171  0.02077317  -2.952   0.005455 **
Income                      0.00001153  0.00001507   0.765   0.449103
GreenScore                  0.01840903  0.01356592   1.357   0.183001
Solar                      -0.62493024  0.36110615  -1.731   0.091857 .
I(Solar^2)                  0.17229535  0.08533036   2.019   0.050760 .
Electricity                -0.00467662  0.00590286  -0.792   0.433257
Models                      0.28931761  0.05172498   5.593 0.00000223 ***
portstopop                  0.00612232  0.00669302   0.915   0.366256
PoliticalParty20Republican -0.43999329  0.26175298  -1.681   0.101197
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.4621665)

    Null deviance: 108.32  on 49  degrees of freedom
Residual deviance:  17.10  on 37  degrees of freedom
AIC: 116.25

Number of Fisher Scoring iterations: 2

[1] 0.8421344
[1] "10-Fold Cross Validation RMSE is 0.8477  and the R-squared value is 0.84"
```

Figure 22: OLS Model with log-transformed *EVS* and Stepwise Selected Predictors

```
Call:
glm(formula = log(EVS) ~ ZEV + Truck + GreenScore + Solar + I(Solar^2) +
    Models + PoliticalParty20, family = gaussian(link = "identity"),
    data = EV)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5443  -0.2771  -0.0001   0.4188   1.4845

Coefficients:
                           Estimate Std. Error t value    Pr(>|t|)
(Intercept)                 4.15557    2.11482   1.965    0.056054 .
ZEVZEV                     -1.86535    0.37641  -4.956 0.000012294 ***
Truck                      -0.06363    0.01789  -3.557    0.000945 ***
GreenScore                  0.02294    0.01085   2.115    0.040401 *
Solar                      -0.52039    0.30884  -1.685    0.099407 .
I(Solar^2)                  0.15295    0.07077   2.161    0.036437 *
Models                      0.29068    0.04562   6.372 0.000000116 ***
PoliticalParty20Republican -0.44268    0.23850  -1.856    0.070462 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.433867)

    Null deviance: 108.321  on 49  degrees of freedom
Residual deviance:  18.222  on 42  degrees of freedom
AIC: 109.43

Number of Fisher Scoring iterations: 2

[1] "10-Fold Cross Validation RMSE is 0.7558"
```

Figure 23:Plain Ridge Model

```
     Best Lambda Best Log Lambda  Best 10FCV BEST RMSE 10FCV
[1,]     1904824           14.46 17446101563        132083.7
12 x 2 sparse Matrix of class "dgCMatrix"
                           Best Ridge 0-Lambda Ridge
(Intercept)                 -6190.772    -367861.591
ZEVZEV                       4277.832     -63792.138
Tax1                         4852.393      28683.908
Gas                         11670.872     116531.004
Truck                        -450.227      -1802.416
Income                          0.125         -0.422
GreenScore                     89.038      -1347.061
Solar                        5049.914      61737.482
Electricity                    34.656        716.764
Models                        799.676       7498.992
portstopop                    129.456        710.199
PoliticalParty20Republican  -1805.230      19596.227
```

Figure 24: Ridge Model with log-transformed *EVS*

```
     Best Lambda Best Log Lambda Best 10FCV BEST RMSE 10FCV
[1,]       0.112          -2.186      0.681           0.825
12 x 2 sparse Matrix of class "dgCMatrix"
                           Best Ridge 0-Lambda Ridge
(Intercept)                     5.193          5.193
ZEVZEV                         -1.418         -1.418
Tax1                           -0.058         -0.058
Gas                             0.206          0.206
Truck                          -0.068         -0.068
Income                          0.000          0.000
GreenScore                      0.014          0.014
Solar                           0.076          0.076
Electricity                    -0.003         -0.003
Models                          0.225          0.225
portstopop                      0.005          0.005
PoliticalParty20Republican     -0.264         -0.264
```

Figure 25: Ridge Model with log-transformed *EVS* and quadratic *Solar*

```
     Best Lambda Best 10FCV BEST RMSE 10FCV
[1,]       0.112      0.618           0.786
13 x 2 sparse Matrix of class "dgCMatrix"
                           Best Ridge 0-Lambda Ridge
(Intercept)                     5.637          5.637
ZEVZEV                         -1.323         -1.323
Tax1                           -0.076         -0.076
Gas                             0.082          0.082
Truck                          -0.067         -0.067
Income                          0.000          0.000
GreenScore                      0.016          0.016
I(Solar^2)                      0.080          0.080
Solar                          -0.210         -0.210
Electricity                    -0.004         -0.004
Models                          0.218          0.218
portstopop                      0.004          0.004
PoliticalParty20Republican     -0.323         -0.323
```

Figure 26a: Ridge Model with Stepwise Selected Predictors

```
     Best Lambda Best Log Lambda Best 10FCV BEST RMSE 10FCV
[1,]       0.112          -2.186      0.529           0.728
8 x 2 sparse Matrix of class "dgCMatrix"
                           Best Ridge 0-Lambda Ridge
(Intercept)                     6.423          6.423
ZEVZEV                         -1.353         -1.353
Truck                          -0.069         -0.069
GreenScore                      0.021          0.021
Solar                          -0.188         -0.188
I(Solar^2)                      0.082          0.082
Models                          0.226          0.226
PoliticalParty20Republican     -0.377         -0.377
```

Figure 26b: Final Model without Solar ^2

```
     Best Lambda Best Log Lambda Best 10FCV BEST RMSE 10FCV
[1,]       0.112          -2.186      0.595           0.771
>
> ridge.coef <- coef(ridge.count, s = ridge.best.lambda)
> ridge.coef.0 <- coef(ridge.count, s = 0)
> all.coefs <- round(cbind(ridge.coef), digits = 3)
> colnames(all.coefs) <- c("Best Ridge")
> all.coefs
7 x 1 sparse Matrix of class "dgCMatrix"
                           Best Ridge
(Intercept)                     6.262
ZEVZEV                         -1.411
Truck                          -0.069
GreenScore                      0.019
Solar                           0.113
Models                          0.234
PoliticalParty20Republican     -0.328
```

Figure 27a: Plain PCR Models

```
> summary(pcr.fit)
Data:   X dimension: 50 11
        Y dimension: 50 1
Fit method: svdpc
Number of components considered: 11

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV          132310   136945   137208   139318   139721   143198   148424   152141
adjCV       132310   135675   135838   136422   136524   140385   144928   148381
       8 comps  9 comps  10 comps  11 comps
CV      159252   158676    157089    155126
adjCV   155109   154379    152945    150963

TRAINING: % variance explained
     1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X      43.00    57.68    66.69    74.15    80.58    86.08    90.52    94.36
EVS    27.19    30.77    52.12    54.02    55.16    59.49    60.61    60.68
     9 comps  10 comps  11 comps
X      96.76     98.72    100.00
EVS    62.19     62.67     64.55
> validationplot(pcr.fit, val.type = "RMSEP", legendpos= "bottomright")
> pcr.fit$coefficients[,,c(1,2,4)]
                              1 comps    2 comps    4 comps
ZEVZEV                      12150.271  10929.625   9205.855
Tax1                         8952.478  14354.329  -2378.796
Gas                          6613.550   6788.784  36838.612
Truck                       -7070.406 -16922.790 -36186.936
Income                       9308.961   6016.573 -19598.179
GreenScore                  11191.435   6380.825  -8032.604
Solar                        8841.624  13005.597  36643.700
Electricity                  3822.726  -8941.461  10435.416
Models                      12653.241  15779.932  14421.378
portstopop                  10335.391   4993.253  21875.204
PoliticalParty20Republican -9640.593 -10446.177  10797.707
> #4 Component 139721
```

Figure 27b: Plain PLSR Model

```
> summary(plsr.fit)
Data:   X dimension: 50 11
        Y dimension: 50 1
Fit method: kernelpls
Number of components considered: 11

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
CV          132310   140740   149497   154403   157563   156519
adjCV       132310   138575   145840   150401   153237   152251
       6 comps  7 comps  8 comps  9 comps  10 comps  11 comps
CV      155253   155040   155140   155138    155131    155126
adjCV   151082   150882   150976   150974    150967    150963

TRAINING: % variance explained
     1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
X      41.25    52.71    64.24    70.52    74.46    79.64    83.55
EVS    40.78    60.79    62.83    63.97    64.45    64.52    64.55
     8 comps  9 comps  10 comps  11 comps
X      88.19    93.38     96.24    100.00
EVS    64.55    64.55     64.55     64.55
> validationplot(plsr.fit, val.type = "RMSEP", legendpos= "topright")
> plsr.fit$coefficients[,,c(1,2,4)]
                              1 comps    2 comps    4 comps
ZEVZEV                      10181.347  -6636.085 -27367.466
Tax1                        10332.746   7365.367  17116.758
Gas                         17016.190  43598.479  41789.340
Truck                      -14177.059 -27082.529 -10514.251
Income                       6499.722  -8123.695  -1096.764
GreenScore                   6969.575 -14201.064 -18007.037
Solar                       21986.456  56054.931  70694.840
Electricity                  3985.249   4823.443  17473.901
Models                      16057.600  16416.637  25241.302
portstopop                  13422.284  15709.021  14067.963
PoliticalParty20Republican  -5824.087  13889.306  13514.243
```

Figure 28a: PCR Models with log-transformed *EVS*

```
> summary(pcr.fit)
Data:   X dimension: 50 11
        Y dimension: 50 1
Fit method: svdpc
Number of components considered: 11

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502    1.160    1.054    1.059    1.048     1.03    1.020   0.9974
adjCV        1.502    1.156    1.048    1.053    1.048     1.03    1.011   0.9812
       8 comps  9 comps  10 comps  11 comps
CV       1.100    1.081     1.117    0.9164
adjCV    1.089    1.062     1.102    0.9002

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X           43.00    57.68    66.69    74.15    80.58    86.08    90.52    94.36
log(EVS)    44.03    56.31    57.40    58.90    62.09    64.58    68.21    68.30
          9 comps  10 comps  11 comps
X           96.76     98.72    100.00
log(EVS)    71.87     72.12     82.47
> validationplot(pcr.fit, val.type = "RMSEP", legendpos= "bottomright")
> pcr.fit$coefficients[,,c(1,2,4)]
                                1 comps     2 comps     4 comps
ZEVZEV                       0.17552286  0.14988143  0.08259819
Tax1                         0.12932754  0.24280125  0.14960753
Gas                          0.09553937  0.09922041  0.30462434
Truck                       -0.10213911 -0.30910272 -0.35307839
Income                       0.13447729  0.06531591  0.08795206
GreenScore                   0.16167152  0.06061767  0.05851254
Solar                        0.12772613  0.21519644  0.18140492
Electricity                  0.05522312 -0.21290713 -0.22056410
Models                       0.18278878  0.24846945  0.22027915
portstopop                   0.14930510  0.03708575  0.01977401
PoliticalParty20Republican -0.13926804 -0.15619050 -0.22454258
> #4 Component  1.048
```

Figure 28b: PLSR Models with log-transformed *EVS*

```
> summary(plsr.fit)
Data:   X dimension: 50 11
        Y dimension: 50 1
Fit method: kernelpls
Number of components considered: 11

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502    1.065   0.9678   0.9664   0.9725   0.9406   0.9294   0.9214
adjCV        1.502    1.061   0.9609   0.9514   0.9515   0.9223   0.9122   0.9047
       8 comps  9 comps  10 comps  11 comps
CV      0.9177   0.9168    0.9165    0.9164
adjCV   0.9013   0.9005    0.9002    0.9002

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X           41.90    55.61    62.70    67.33    73.07    78.40    82.81    88.34
log(EVS)    55.57    70.24    77.64    81.27    82.22    82.42    82.47    82.47
          9 comps  10 comps  11 comps
X           93.77     96.37    100.00
log(EVS)    82.47     82.47     82.47
> validationplot(plsr.fit, val.type = "RMSEP", legendpos= "topright")
> plsr.fit$coefficients[,,c(1,2,4)]
                               1 comps     2 comps      4 comps
ZEVZEV                      0.11960248 -0.14600921 -0.788462222
Tax1                        0.13113285  0.05255473  0.017960804
Gas                         0.12006374  0.14588982 -0.028340237
Truck                      -0.23821801 -0.57402666 -0.574469699
Income                      0.14313067  0.13126835  0.154822035
GreenScore                  0.16576405  0.13807217  0.234886795
Solar                       0.15923195  0.17413752 -0.009363282
Electricity                -0.03259202 -0.26072944 -0.006537071
Models                      0.26453522  0.44757212  1.101287912
portstopop                  0.12704952  0.02129077  0.104395381
PoliticalParty20Republican -0.15213400 -0.12608831 -0.084376328
> #4 component 0.9725
```

Figure 29a: PCR Models with log-transformed *EVS* and quadratic *Solar*

```
> summary(pcr.fit)
Data:   X dimension: 50 12
        Y dimension: 50 1
Fit method: svdpc
Number of components considered: 12

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502    1.129    1.081    1.073    1.062    1.043    1.034   1.0134
adjCV        1.502    1.126    1.073    1.069    1.055    1.043    1.022   0.9994
       8 comps  9 comps  10 comps  11 comps  12 comps
CV       1.093    1.055     1.086    0.8763    0.9393
adjCV    1.083    1.035     1.072    0.8609    0.9195

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X           42.04    56.65    67.57    74.99    81.28    86.50    90.90    94.46
log(EVS)    46.02    53.54    56.01    58.60    60.19    65.82    68.18    68.20
          9 comps  10 comps  11 comps  12 comps
X           96.68     98.51     99.70    100.00
log(EVS)    72.49     72.83     83.37     84.21
> validationplot(pcr.fit, val.type = "RMSEP", legendpos= "bottomright")
> pcr.fit$coefficients[,,c(1, 2,4)]
                               1 comps     2 comps     4 comps
ZEVZEV                      0.16363328  0.11921034  0.10445460
Tax1                        0.12234965  0.16323464  0.17195139
Gas                         0.09469084  0.11021374  0.27061118
Truck                      -0.10271567 -0.23342582 -0.34156605
Income                      0.12356290  0.05017250  0.05091529
GreenScore                  0.14811797  0.05187936  0.04981318
Solar                       0.13959997  0.24997738  0.11528789
I(Solar^2)                  0.12233627  0.24206717  0.10639984
Electricity                 0.04876877 -0.11324080 -0.19945814
Models                      0.17392131  0.19152861  0.22172526
portstopop                  0.14223931  0.07262187  0.01493539
PoliticalParty20Republican -0.12602045 -0.08935093 -0.23079046
>
> #4component 1.062
```

Figure 29b: PLSR Models with log-transformed *EVS* and quadratic *Solar*

```
> summary(plsr.fit)
Data: 	X dimension: 50 12 
	Y dimension: 50 1
Fit method: kernelpls
Number of components considered: 12

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502    1.048   0.9569   0.9474   0.9035   0.8841   0.8985   0.9043
adjCV        1.502    1.045   0.9498   0.9332   0.8876   0.8682   0.8821   0.8873
       8 comps  9 comps  10 comps  11 comps  12 comps
CV      0.9280   0.9349    0.9327    0.9393    0.9393
adjCV   0.9095   0.9154    0.9133    0.9194    0.9195

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X           41.21    53.58    61.52    68.04    74.63    80.40    83.84    88.45
log(EVS)    55.94    70.86    78.07    81.71    83.34    83.74    83.95    84.02
          9 comps  10 comps  11 comps  12 comps
X           90.92     94.59     96.85    100.00
log(EVS)    84.16     84.21     84.21     84.21
> validationplot(plsr.fit, val.type = "RMSEP", legendpos= "topright")
> plsr.fit$coefficients[,,c(1,2,4)]
                                1 comps       2 comps      4 comps
ZEVZEV                       0.11058680 -0.1566940555 -0.727481199
Tax1                         0.12124800  0.0439820098 -0.088371761
Gas                          0.11101328  0.1232474425 -0.002111136
Truck                       -0.22026104 -0.5625787022 -0.607542256
Income                       0.13234142  0.1340913981  0.168978372
GreenScore                   0.15326869  0.1430023357  0.261124483
Solar                        0.14722898  0.0843620878 -0.155964967
I(Solar^2)                   0.14744822  0.1675388846  0.317758092
Electricity                 -0.03013522 -0.2680696783 -0.004953512
Models                       0.24459445  0.4390523998  0.981403555
portstopop                   0.11747247 -0.0006534936  0.106503494
PoliticalParty20Republican  -0.14066607 -0.1406830082 -0.055971046
> #4 component 0.9035
```

Figure 30a: PCR Models with log-transformed *EVS* and Stepwise Selected Predictors

```
> summary(pcr.fit)
Data: 	X dimension: 50 7 
	Y dimension: 50 1
Fit method: svdpc
Number of components considered: 7

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502    1.071    1.092   0.9481    1.020   0.9365   0.7476    0.793
adjCV        1.502    1.068    1.089   0.9424    1.016   0.9311   0.7405    0.782

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
X           52.88    72.87    83.83    91.55    96.86    99.30   100.00
log(EVS)    50.73    50.83    63.08    63.15    69.11    82.31    83.18
> validationplot(pcr.fit, val.type = "RMSEP", legendpos= "bottomright")
> pcr.fit$coefficients[,,c(1,3,4)]
                              1 comps      3 comps      4 comps
ZEVZEV                      0.2270846  0.118102765  0.094845868
Truck                      -0.1740189 -0.684429002 -0.682889491
GreenScore                  0.2012923 -0.008072197 -0.008508523
Solar                       0.2210887  0.094013915  0.102800446
I(Solar^2)                  0.1929478  0.077371681  0.084887456
Models                      0.2517637  0.334292648  0.314017077
PoliticalParty20Republican -0.1762665 -0.219880646 -0.261572708
> #3 component model is best  0.9481
```

Figure 30b: PLSR Models with log-transformed *EVS* and Stepwise Selected Predictors

```
> summary(plsr.fit)
Data:   X dimension: 50 7
        Y dimension: 50 1
Fit method: kernelpls
Number of components considered: 7

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
CV           1.502   1.0004   0.8659   0.7738    0.771   0.7388   0.7692    0.793
adjCV        1.502   0.9987   0.8568   0.7584    0.764   0.7324   0.7597    0.782

TRAINING: % variance explained
          1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
X           52.36    62.76    70.00    86.28    91.55    95.31   100.00
log(EVS)    57.91    74.99    81.11    82.22    82.92    83.09    83.18
> validationplot(plsr.fit, val.type = "RMSEP", legendpos= "topright")
> plsr.fit$coefficients[,,c(1,2,3)]
                             1 comps     2 comps    3 comps
ZEVZEV                     0.1471942 -0.36503303 -0.7646421
Truck                     -0.2931737 -0.78698612 -0.5461856
GreenScore                 0.2040050  0.17355095  0.4384220
Solar                      0.1959659 -0.02720948 -0.2003257
I(Solar^2)                 0.1962578  0.14632391  0.2310254
Models                     0.3255621  0.63324006  1.0313130
PoliticalParty20Republican -0.1872305 -0.17921722 -0.1894096
> #3 component with 0.7738 error
```

E. OLS Assumption Testing

Figure 31: OLS Testing Summaries

Assumption	Testing

Assumption 1: is Y Continuous?	No, it is discrete. Outcome variable is in units of EVs sold. It is not in decimals, and it cannot be negative. The distribution of EVS is not normal. This is shown by the histogram in fig. 19 in which the data is highly skewed to the right. The QQ-plot in Fig 20, however, shows that the data is aligned with the qq-line. The Shapiro-Wilk test is significant in Fig 21. The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05.
Assumption 2: Are Errors Normally Distributed?	Yes. This is shown in the QQ-plot in Fig 22, in which most of the data aligns with the qq-line. The residual plot in Figure 23 also shows an even, cloud like distribution. Fig 24 also shows that the Shapiro-wilks test is insignificant. Thus, we fail to reject the hypothesis of normality.
Assumption 3: Are the X's independent?	No. The last CI Value is 89, in fig 25, which is way higher than threshold of 50. Thus, there is severe overall multicollinearity in the model. The VIFs for all variables, in Fig 26, are less than 10 which means multicollinearity for individual predictors is not an issue.

Assumption 4: Do Y and X's have a linear relationship?	This has been evaluated using the plots created to study the direction of correlation between Y and Xs in which the EVS variable has been logged and a trend line has been added. Figure 6, 7,8,9, 10, 11 and 13 show that there is a linear relationship between log of EVS and the predictor variables Oil Price per Gallon, Median Household Income, Green Score, EV Models per states, Total Charging ports, Charging Connections per Location and % of light truck purchased, respectively. Figure 12 shows that EVS has a polynomial relationship with the percentage of solar installation per household (solar), where the log of EVS dramatically increases at lower values of 'solar' and then increases at a lower rate for higher values of 'solar'.
Assumption 5 and 6: Are error independence and observation independent?	We conducted a Durbin-Watson test on our model for serial correlation and the DW value is 1.6436, close to 2. Therefore, there is no serial correlation in the model.
	Moreover, there is no time variable in our model which make testing for serial correlation redundant.
Assumption 7: Is the Error Average zero?	The average of the errors of the regression model is 3.363887e-13 which is very close to zero.

Assumption 8: Is the Error variance Constant i.e., Homoscedastic?	No. The residual plot in Fig 28, shows that the errors are not spread in a cloud like form. We also conducted a Breusch-Pagan test which was significant at p=0.009 level. Therefore, we reject the null hypotheses of no heteroskedasticity. Thus, the outcomes of the residual plot and the BP test show that there is a problem with heteroskedasticity in this model. Based on the BP test, the p-value of the residual regression is significant, and the error is correlated with the predicted values.

Figure 32: Histogram of EV Sales

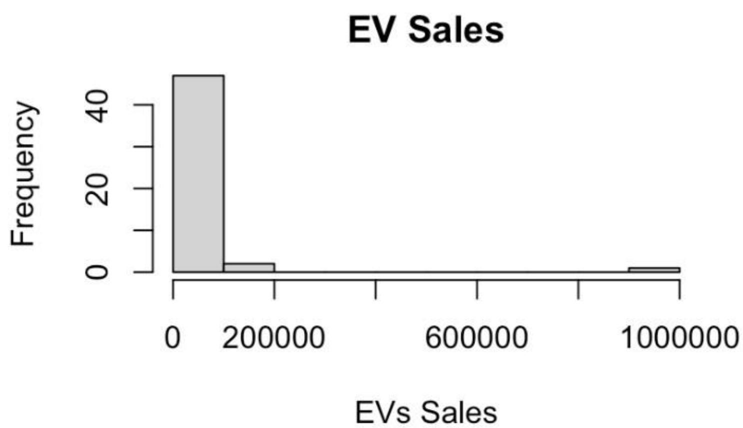


Figure 33: QQ-plot of EV Sales

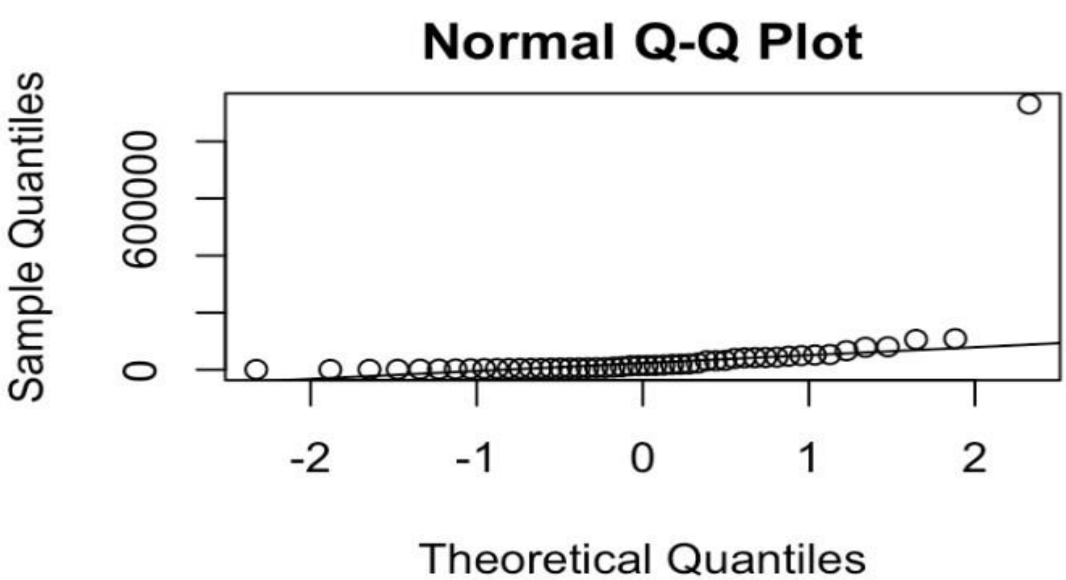


Figure 34: Normality test of EVS

```
        Shapiro-Wilk normality test

data:  EV$EVS
W = 0.26628, p-value = 2.144e-14
```

Figure 35: QQ Plot of Residuals

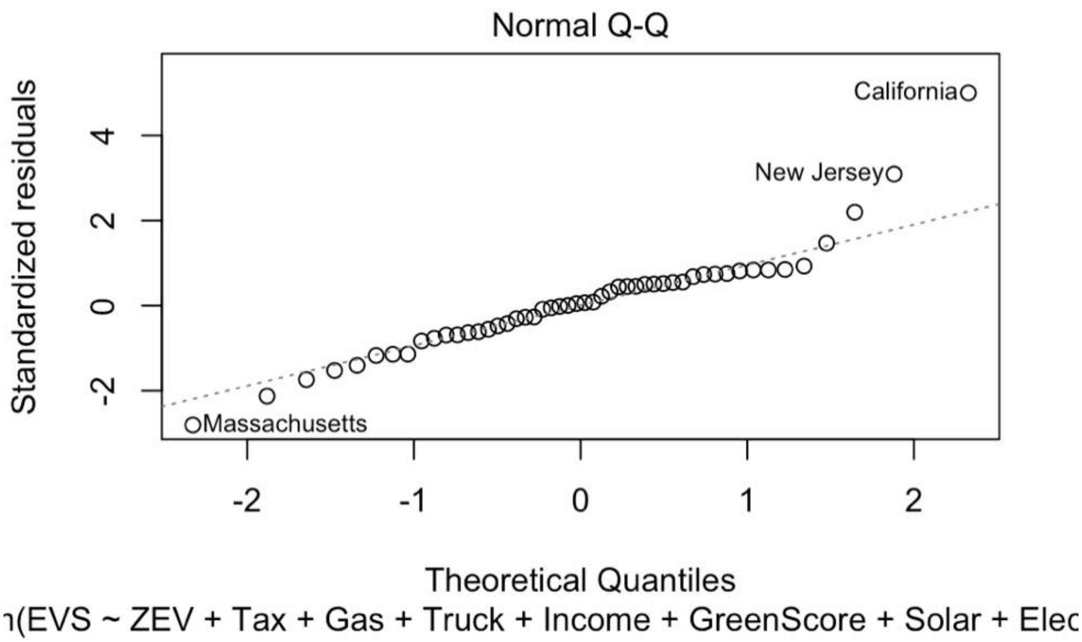


Figure 36: Residual Plot

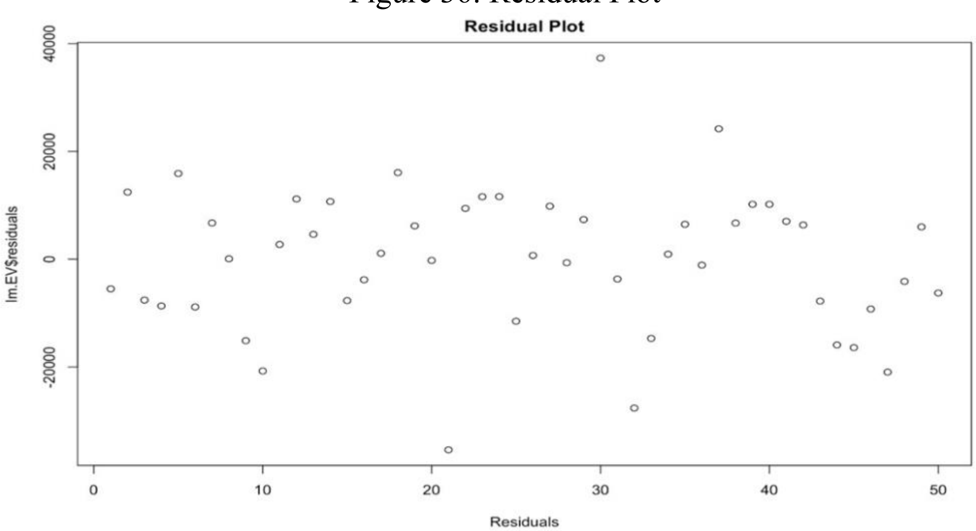


Figure 37: Multicollinearity Tests (CI and VIF)

```
[1]  1.000000  2.325404  3.597464  3.943357  5.179089  6.439931  7.741913
[8] 20.481797 27.165726 31.615168 39.534645 90.921552
         ZEV             Tax              Gas           Truck
    3.447693        1.921369         1.309154        2.190242
      Income      GreenScore            Solar     Electricity
    1.968410        2.911304         1.558720        2.090759
      Models     portstopop PoliticalParty20
    4.713248        2.213003         1.743975
```

Figure 38: DW Test

```
	Durbin-Watson test

data:  lm.EV
DW = 2.2079, p-value = 0.7133
alternative hypothesis: true autocorrelation is greater than 0
```

Figure 39: Residuals vs Fitted Plot

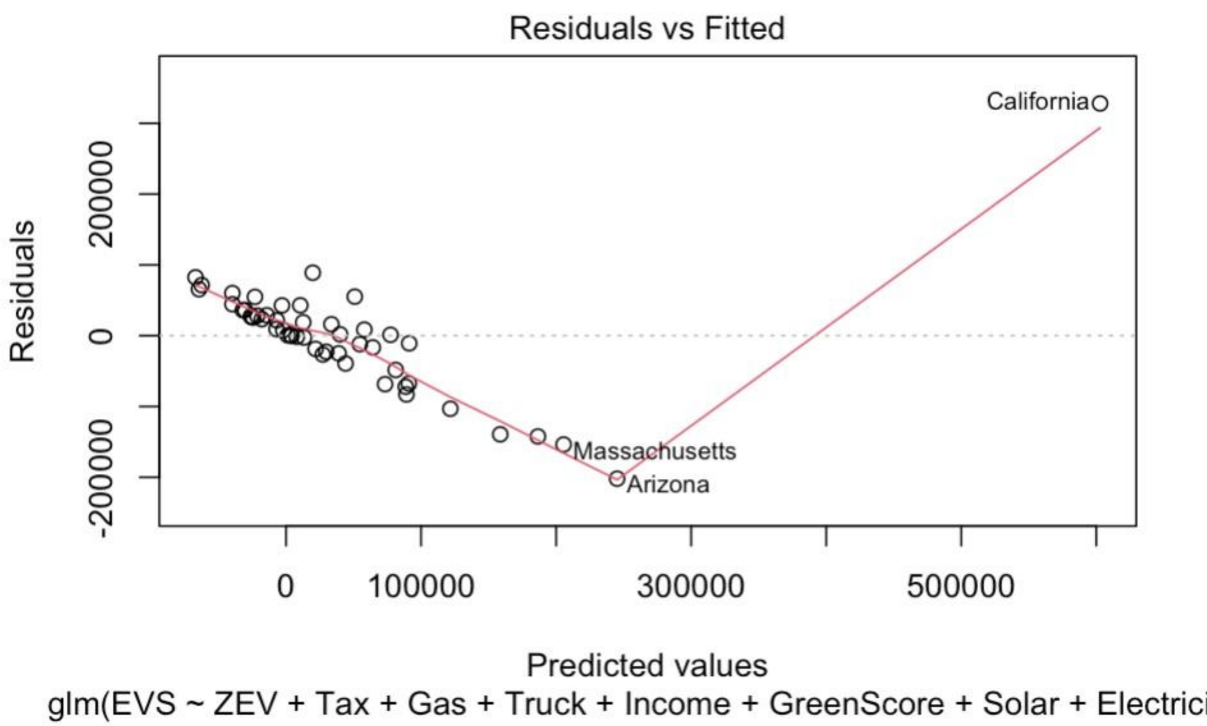


Figure 40: BP Test

```
	studentized Breusch-Pagan test

data:  lm.EV
BP = 42.501, df = 11, p-value = 0.00001326
```

F. References

ICCT. “A Global Comparison of the Life-Cycle Greenhouse Gas Emissions of Combustion Engine and Electric Passenger Cars.” *International Council on Clean Transportation*, 25 Jan. 2022, https://theicct.org/publication/a-global-comparison-of-the-life-cycle-greenhouse-gas-emissi ons-of-combustion-engine-and-electric-passenger-cars/.

EPA. “Sources of Greenhouse Gas Emissions.” *EPA*, Environmental Protection Agency, https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions.

United Nations. “Goal 7 | Department of Economic and Social Affairs.” *United Nations*, United Nations, https://sdgs.un.org/goals/goal7.

DeSilver, Drew. “Today's Electric Vehicle Market: Slow Growth in U.S., Faster in China, Europe.” *Pew Research Center*, Pew Research Center, 10 June 2021, https://www.pewresearch.org/fact-tank/2021/06/07/todays-electric-vehicle-market-slow-growth-in-u-s-faster-in-china-europe/.

Lewis, Michelle. “Light-Duty EVS Have 64% Lower Life-Cycle Emissions than ICE Vehicles – Ford Study.” *Electrek*, 4 Mar. 2022, https://electrek.co/2022/03/04/light-duty-evs-have-64-lower-life-cycle-emissions-than-icevehicles-ford-study/.

UCSUSA. “What Is Zev?” *Union of Concerned Scientists*, https://www.ucsusa.org/resources/what-zev.