|  |  |
| --- | --- |
|  | Master of Science in Business Analytics and Information Systems  Course: Statistical Datamining  Instructor: PhD. Daniel Zantedeschi |

**FINAL PROJECT :**

**Boat Trader.com**

By: Pedro Jose Ortiz Rodriguez

USF Student ID: U93789013

November 16th , 2019

TABLE OF CONTENTS

1. **Data Extraction Description………………………………………………………………………………..3**
2. **Visualization and Data Exploration …………………………………………………………………….6**
3. **Statistical Analyses ……………………………………………………………………………………………14**
4. **Summary of results …………………………………………………………………………………………..28**
5. **Comments and recommendations for future iterations……………………………………..29**
6. **Appendix……………………………………………………………………………………………………………30**

**1) Data Extraction Description**

For this project, data was extracted from the Boat Trader website (<https://www.boattrader.com/>) using the provided Python script in the course with Spyder (Python 3.6).

When trying to use this code at it is, the results generally came out empty because some elements within the code where deprecated possibly to changes in the website structure.

For the script to run, the following URL was used in line 13:

search\_url = 'http://www.boattrader.com/search-results/NewOrUsed-any/Type-all/Category-all/Zip-77552/Radius-400/Sort-Length:DESC'

Additionally, in the loop section of the pages, a limit was set to 357 pages, since after this number of pages, an exception seemed to be triggered in the output, as seen below:

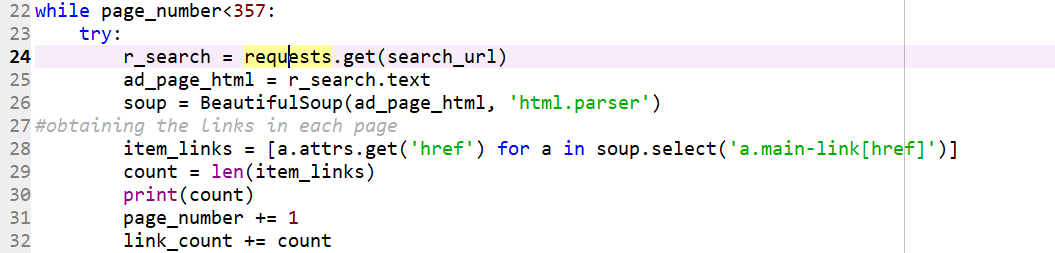


Figure Python Script for Web Scrapping

Note also that the process of getting the links was update in line 28 within the soup.select to use “a.main-link[href]” , since the initial version was generating non-working links. The following section of the code which focuses on extracting listing URL, year of the boat, make, length, engine type, listing price, was used as the original version of the script:

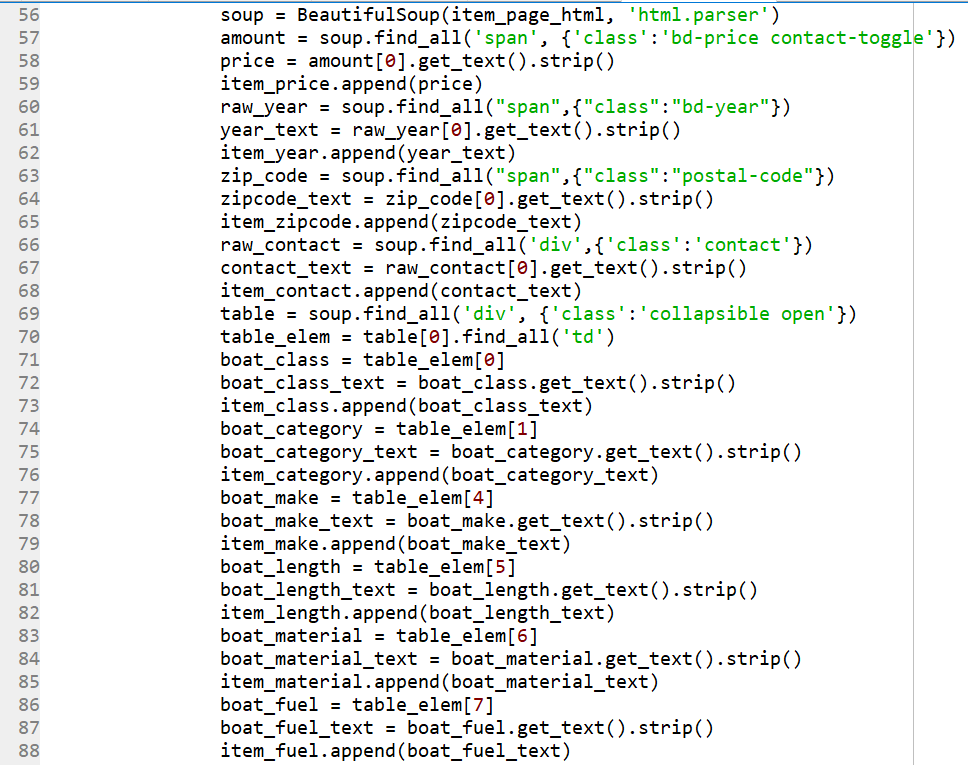


Figure Python Script for Web Scrapping (Cont.)

When logging into the site, and making a search for all the available boat listings we get around 109,560 listing results as shown below:

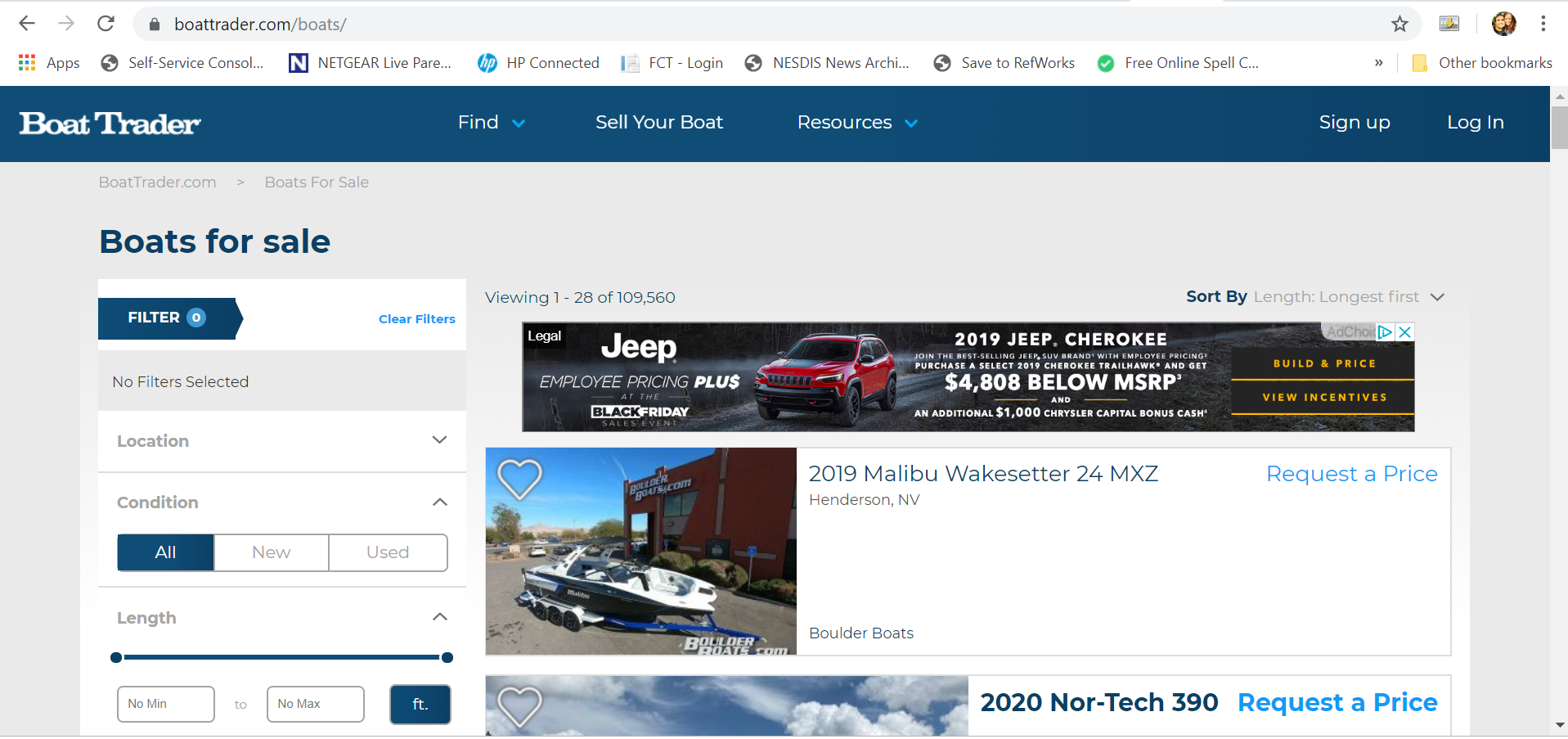


Figure Boat Trader Web Site

Due to the limitations mentioned in the code, a smaller portion of this data was extracted for later use in R for statistical analysis. The strategy that I implemented consisted of getting data from 5 geographically randomly picked areas from the continental United States based on one zip code, which was later used for the before mentioned search URL in a range of 400 miles. The regions from where the data was extracted where the following:

* Northeast (Zip Code: 10020, New York, NY)
* Southeast (Zip Code: 33620, Tampa, FL)
* GreatLakes(Zip Code: 60064, North Chicago, IL)
* Pacific (Zip Code: 90009, Los Angeles, CA)
* Gulfcoast (Zip Code: 77552, Galveston TX)

This was done in parallel using 5 concurrent Python sessions running virtually the same script, only changing the zip code in the URL and having different output files. The results where stores in the following files:

* Output\_File\_Pacific.csv
* Output\_File\_New\_Great\_LAkes.csv
* Output\_File\_New\_SouthEast.csv
* Output\_File\_New\_NorthEast.csv
* Output\_File\_New\_Gulfcoast.csv

This files where later imported in R, merged into a single dataframe and later went through the following data cleaning process:

* Eliminated duplicate listings based on its posting link
* Excluded observations with price values of “Request a Price”
* Clean price variable, by taking out “$” and “,” to transform it into a numeric variable.
* Clean length by eliminating “’” in order to transform it into a numeric variable.
* Calculated the boat’s age using the listing’s listed year.

After this step, the completed data contained 27169 observations with the following column structure as detailed below:



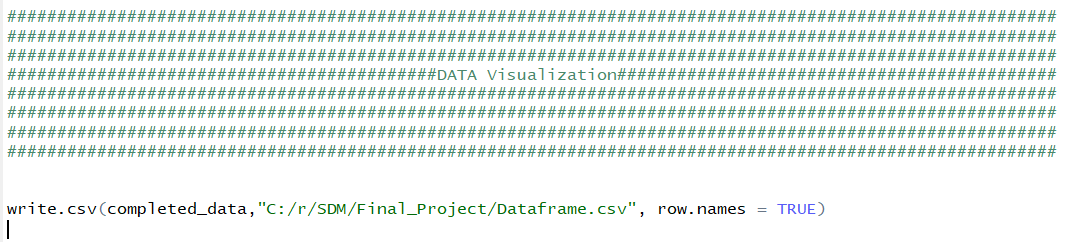
Figure Loaded Dataframe with complete data

**2) Visualization and Data Exploration**

Before starting to do a definition of the data gathered from the Boat Trader website, first we are going to briefly describe the structure of the data. The data is comprised of:

* Dependent Variables
  + Price (Numerical data)
* Independent Variables
  + Posting Link
  + Year (Numerical data)
  + Contact
  + Zipcode
  + Class
  + Category
  + Length (Numerical data)
  + Maker
  + Material
  + Fuel
  + Region
  + Age (Numerical data)

After cleaning the data in R, I exported the data out to an CSV file in order to do some visualization and analysis of the data using Tableau software as seen below:



After getting the data into the file, I made an initial exploration of the Boat Trader data to see how prices, ages, lengths and quantities of listings are distributed among the United states.

When analyzing the average age of the boat listings within the sample data, we can see that the age can go as much as 45.18 years. The older models seem to be in the Southeastern US, the Northwest as well as southern California:

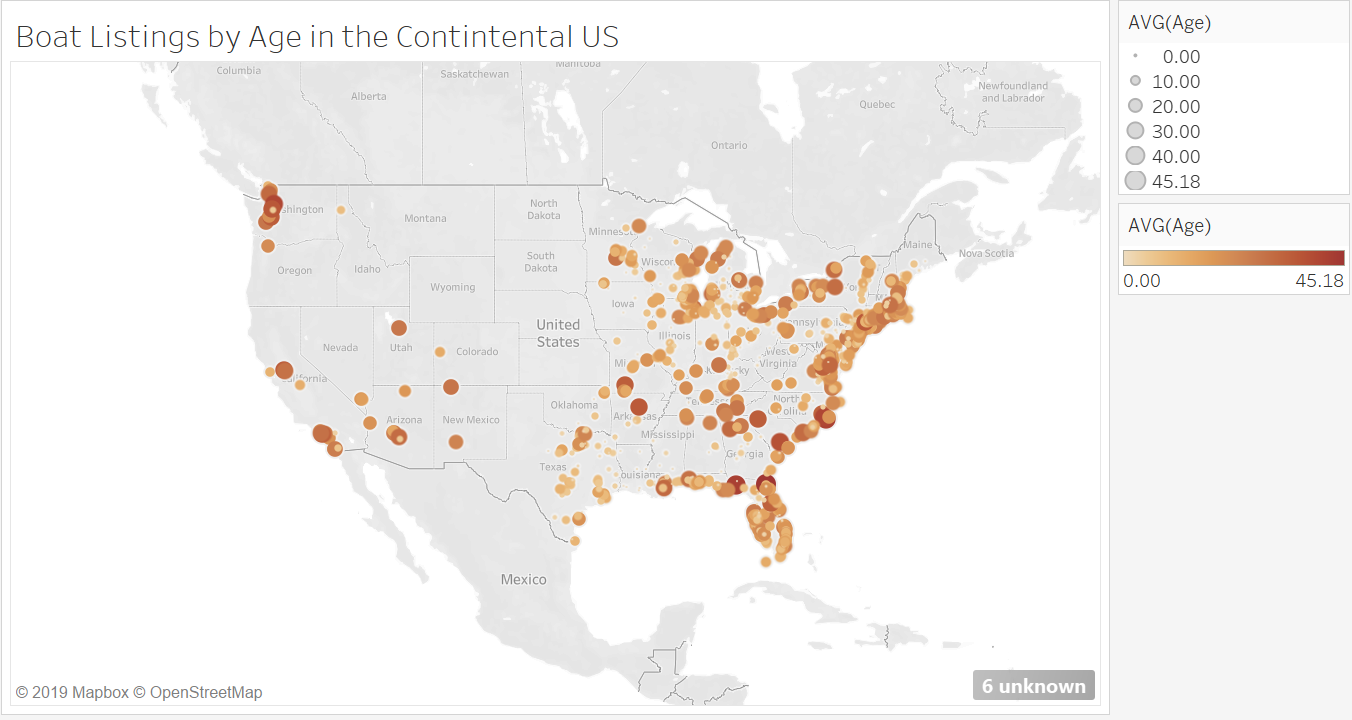


Figure Visualization of Boat Listings by Average Age in the US

In Terms of the price, from a general point of view, there seem to be some relative pattern of prices of 7000$ or below across the US.

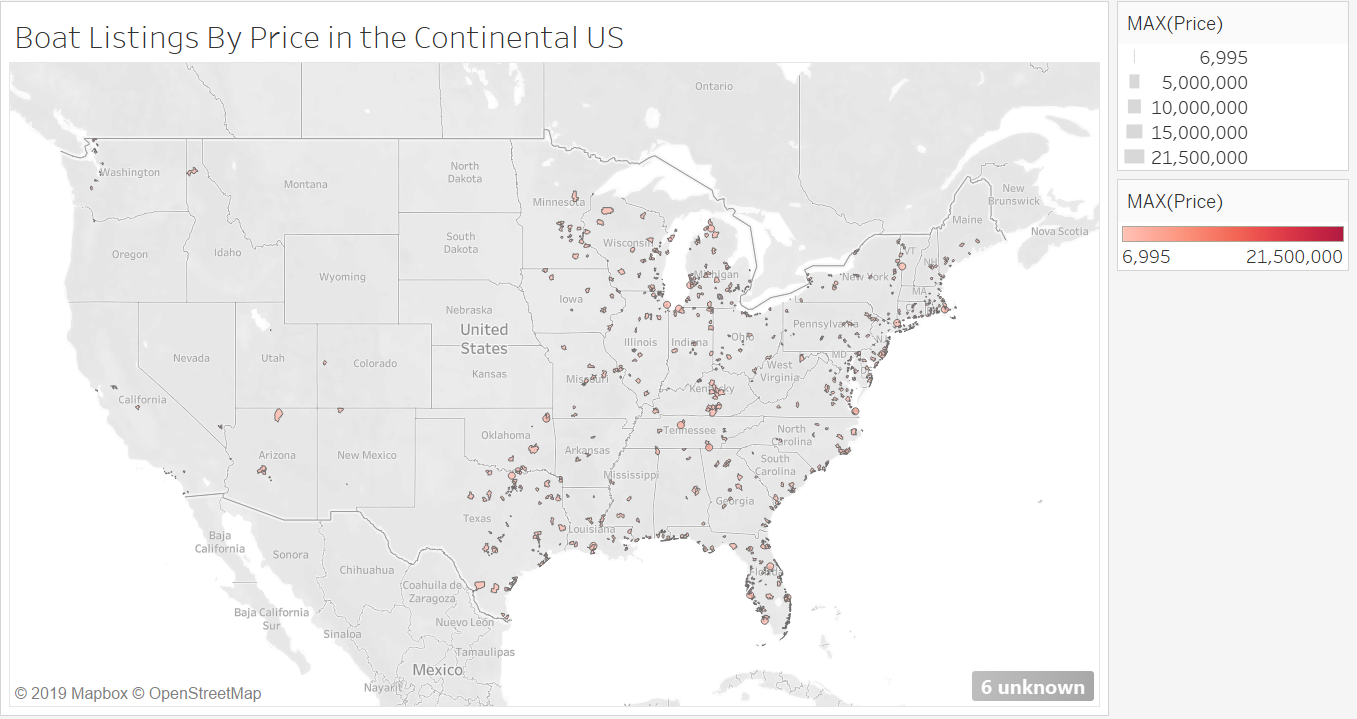


Figure Boat Listing by Max Price in the US

As the legend suggests, there are states that have some multimillion-dollar boat listings. An example of this can be found in south Florida as shown:

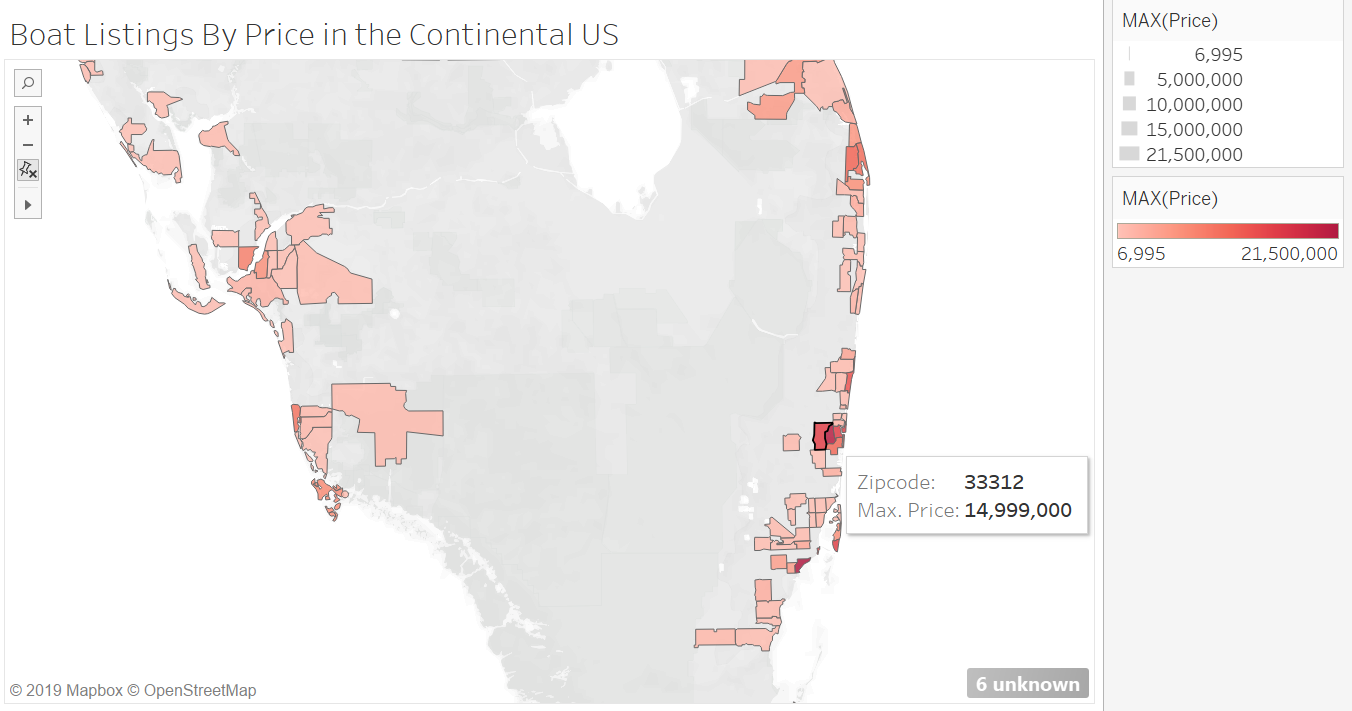


Figure Boat Listings by Price in the US (Focus on South Florida)

In terms of length, we can see that the regions of the US that the highest average length tend to be in southeastern portion (Florida specifically) and some states such as Oklahoma, Mississippi, Ohio, among others.

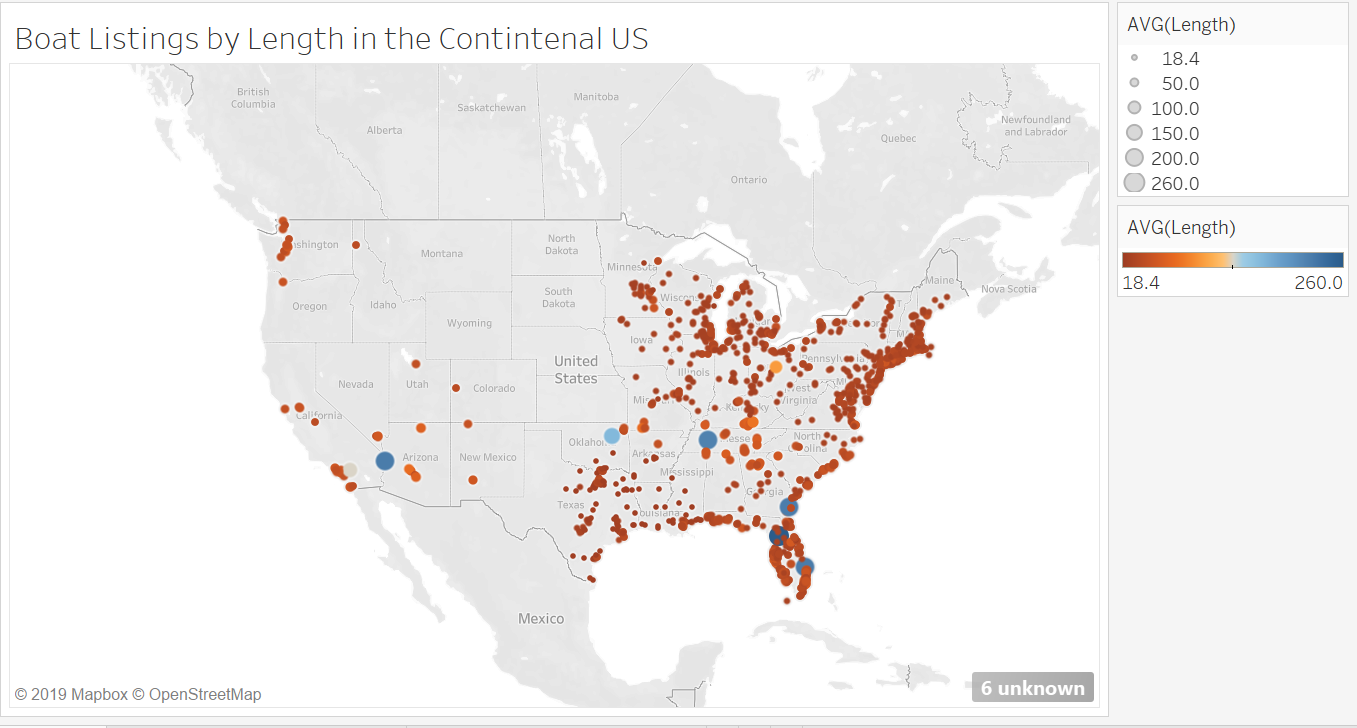


Figure Boat Listings by Average Length in the US

Finally, in terms of available listings in the dataset, we can see that locations (zip codes) with the most listings tend to be in the east and particularly in the southeastern as well as the western gulf coast.

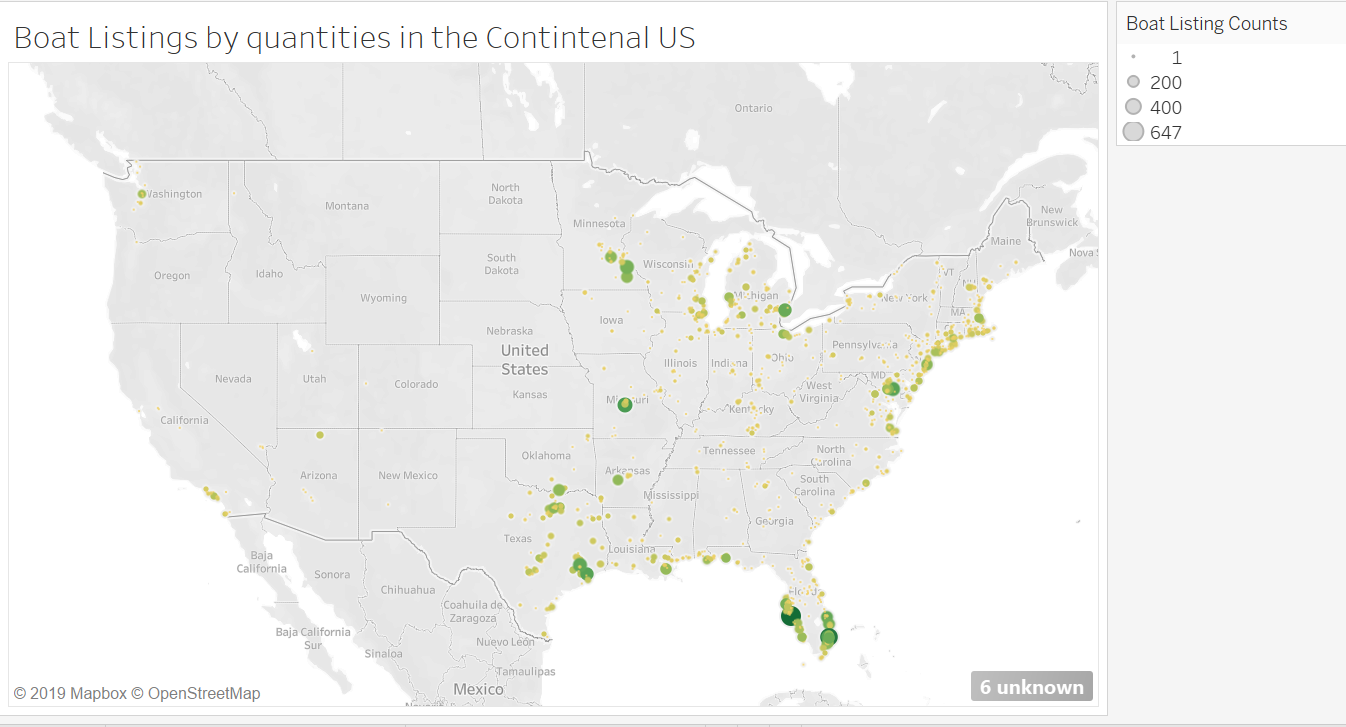


Figure Boat Listings present in Data Throughout the US

In order to complement the data in the dataset in terms of the geographical location of the listings, I provided some basic exploration of the numeric variables of the model

In R Studio, we can briefly do a briefly description on the independent variables and how they relate to the price variable:

In terms of the length, it seems that the majority of the listing’s length fall bellow 50, having right skewedness.

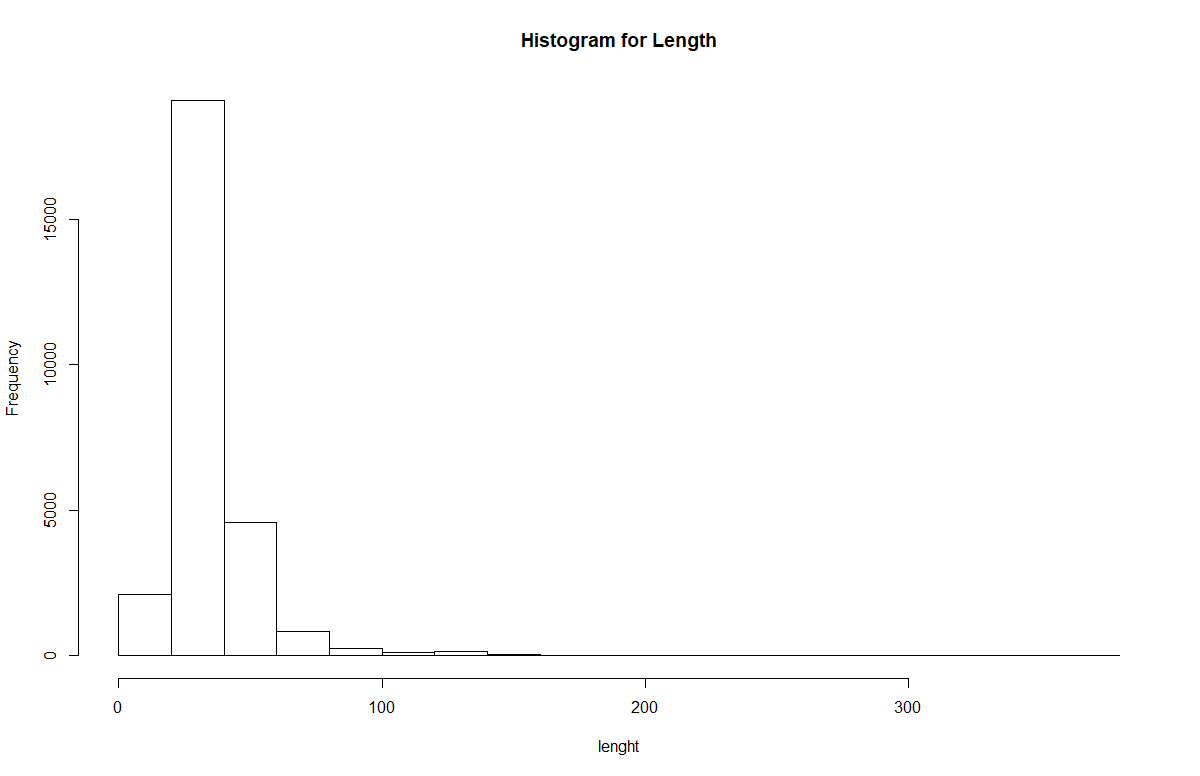


Figure Length Histogram

Going over the age of the listings, we can see that there also right skewedness, and we can see a large majority of listing with ages close to 0 and son on

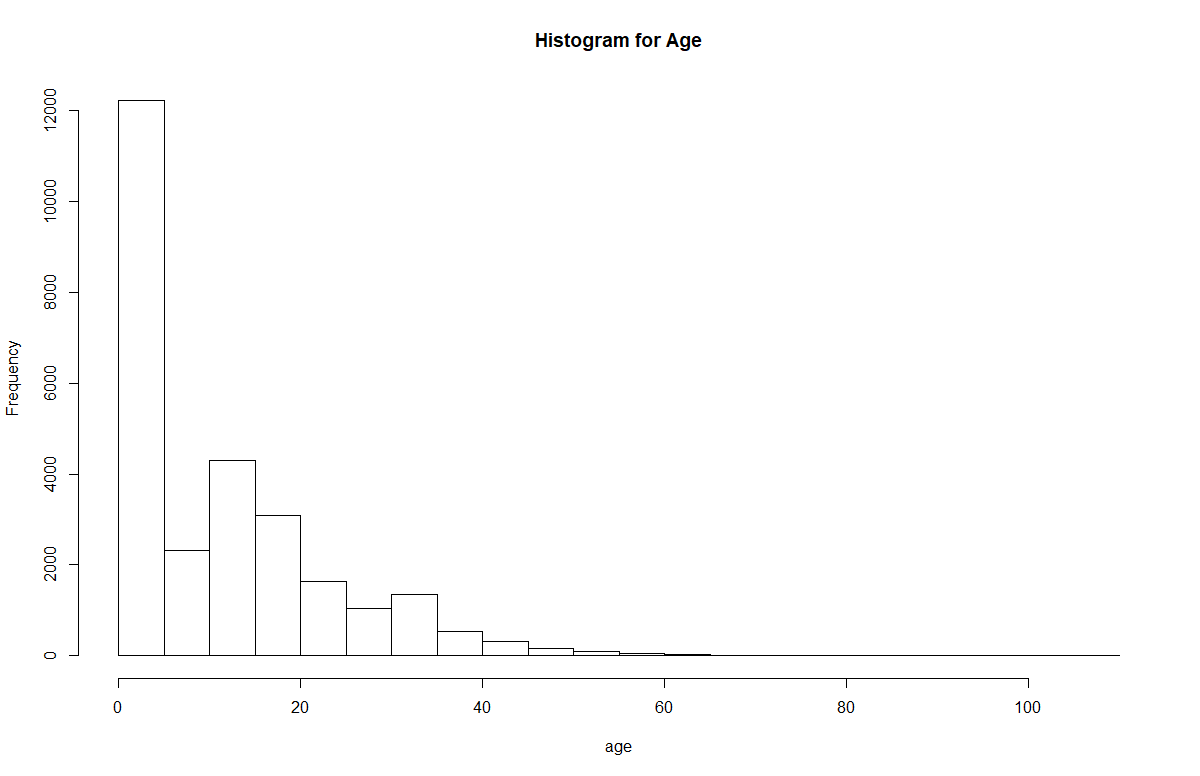


Figure Age Histogram

Complementing the information from the previous histogram, we can see that the Age of the boat listings is left skewed, since most of the boats seem to have a year of manufacture of 1960 and above (most nearing 2020).

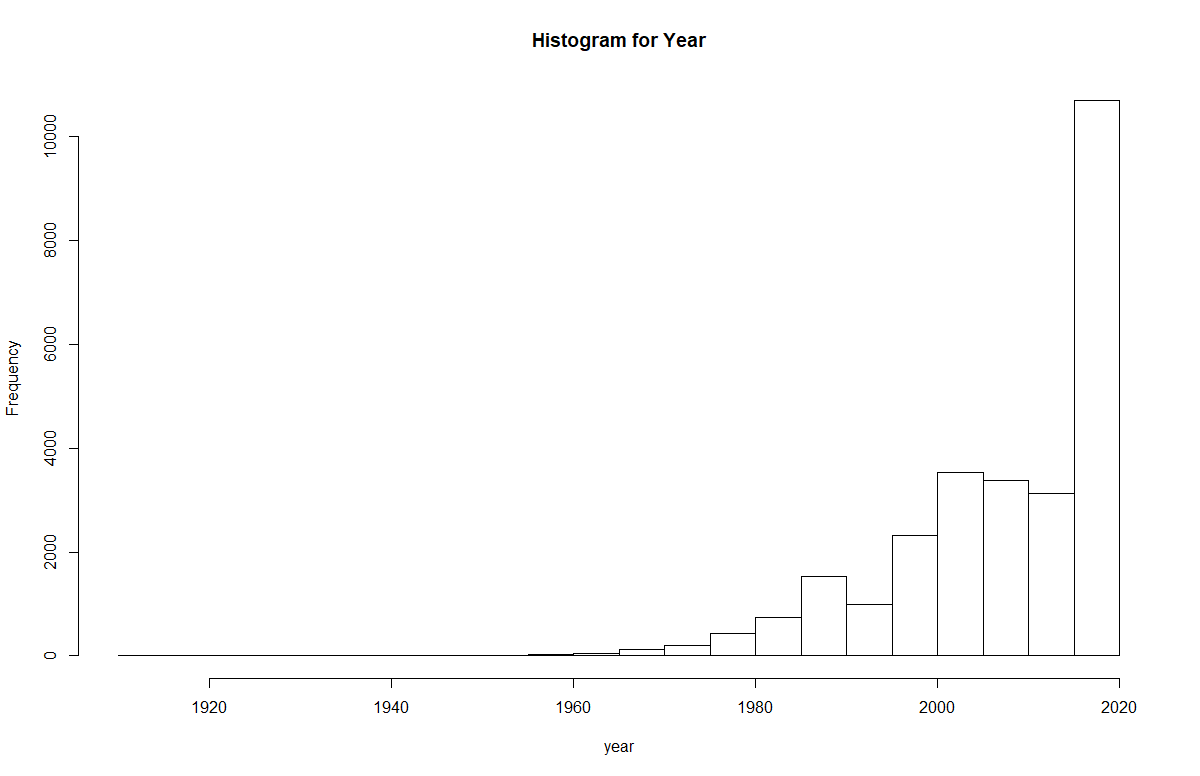


Figure Year Histogram

When comparing how the dependent price behaves with changes length, there is logical trend that as the length increases, the price increases. There are some listings with a higher length (for example 170 and onwards) that have low prices that are due to other methods of payments

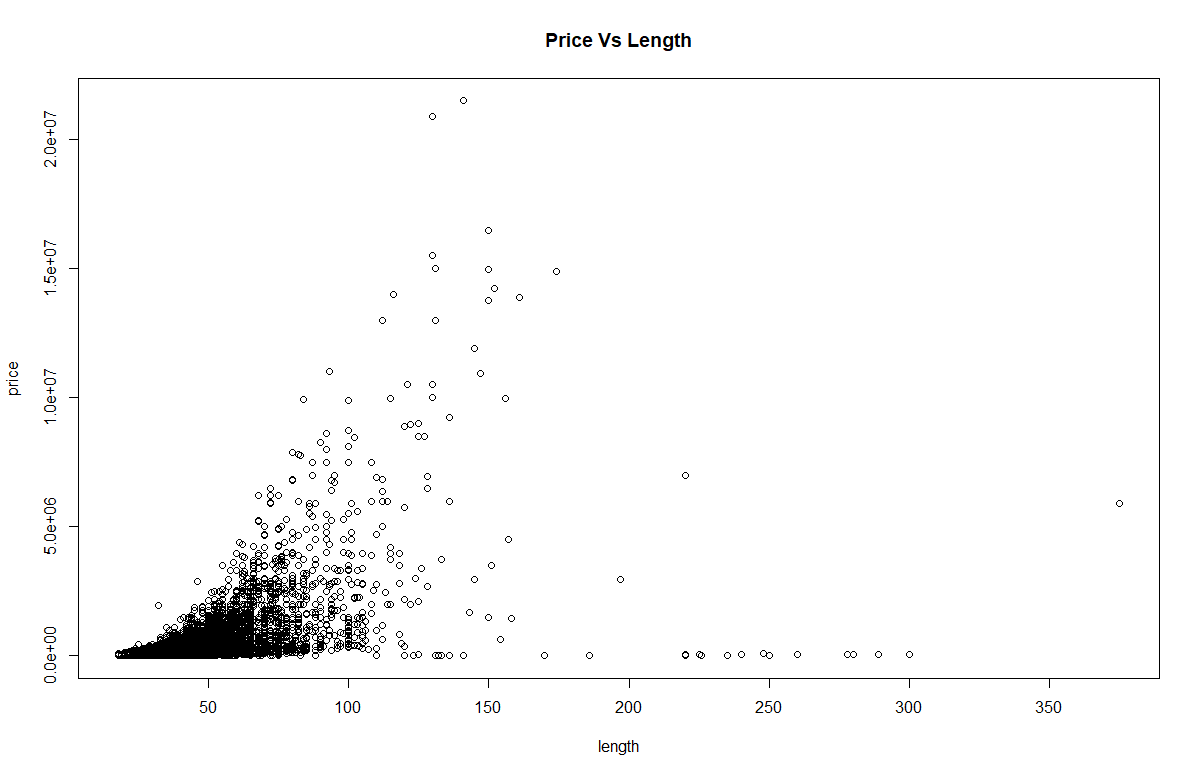


Figure Price Vs Length Plot

In the following plot, we can see another expected behavior : that the price increases as the age of the boat decreases.

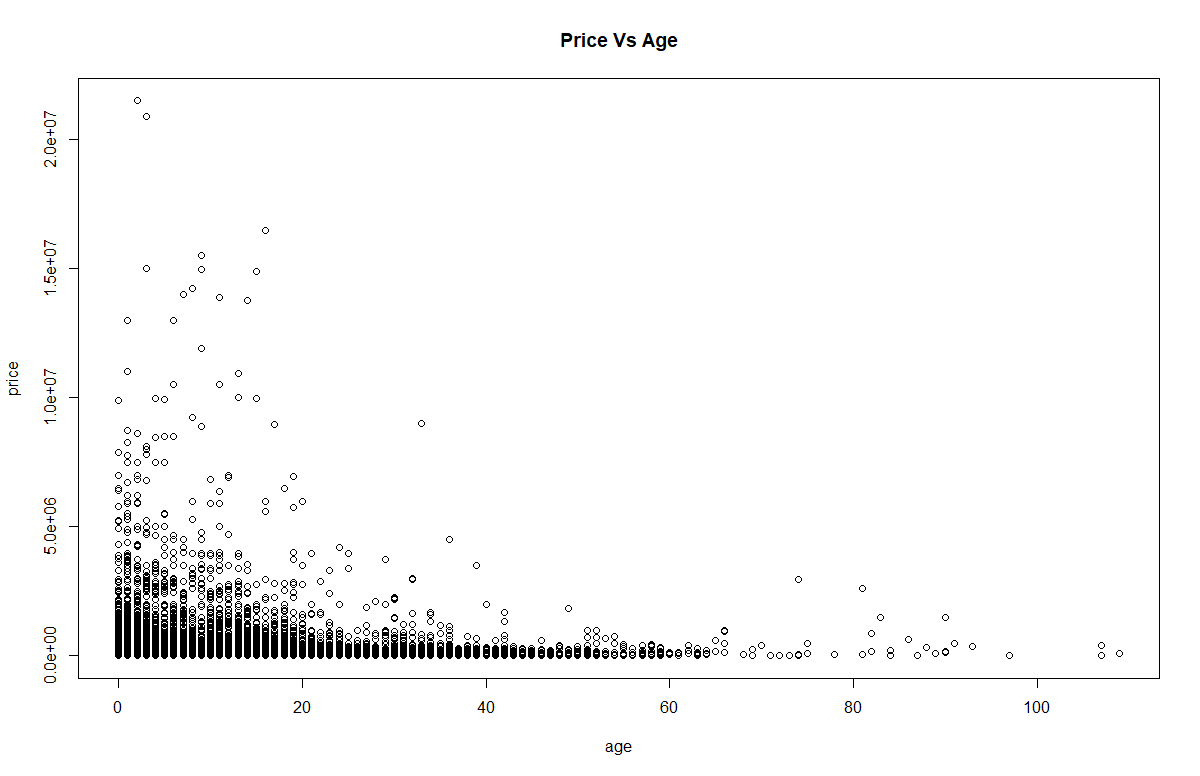


Figure Price Vs Age Plot

**3) Statistical Analyses (using, in the proper contexts, all the tools we are exploring in the class)**

For the purpose of analyzing the data from the Boat Trader website, a linear regression model will be defined with the intention of identifying relationships within price and each of the other mentioned independent variables

For the purpose of this project, the following variables are not being considered into the model:

* Posting Link
* Contact

The following variables will be handled as categorical variables for having a limited amount of possible values:

* Region
* Zipcode
* Material
* Year
* Category
* Make
* Fuel
* Region

Additionally, the following variables will be handled as continuous variables:

* Length
* Age
* Price (Dependent variable)

Including only the continuous variables in the decision model (age+ price):

Call:

lm(formula = price ~ age + length, data = completed\_data)

Residuals:

Min 1Q Median 3Q Max

-5984790 -83777 -29899 35273 18893300

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -398385.6 7651.8 -52.06 <2e-16 \*\*\*

age -10423.8 282.3 -36.93 <2e-16 \*\*\*

length 21460.5 203.1 105.65 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 548300 on 27166 degrees of freedom

Multiple R-squared: 0.2937, Adjusted R-squared: 0.2937

F-statistic: 5649 on 2 and 27166 DF, p-value: < 2.2e-16

From the regression results, we can see that both age and length have significant values at 5% and they play a role in this model that has a Multiple R Square of 0.2937.

*Checking LINE Assumptions:*

To check how this model complies with normality, I’m going to make a residual plot, QQ plot and density residual plots as follows:

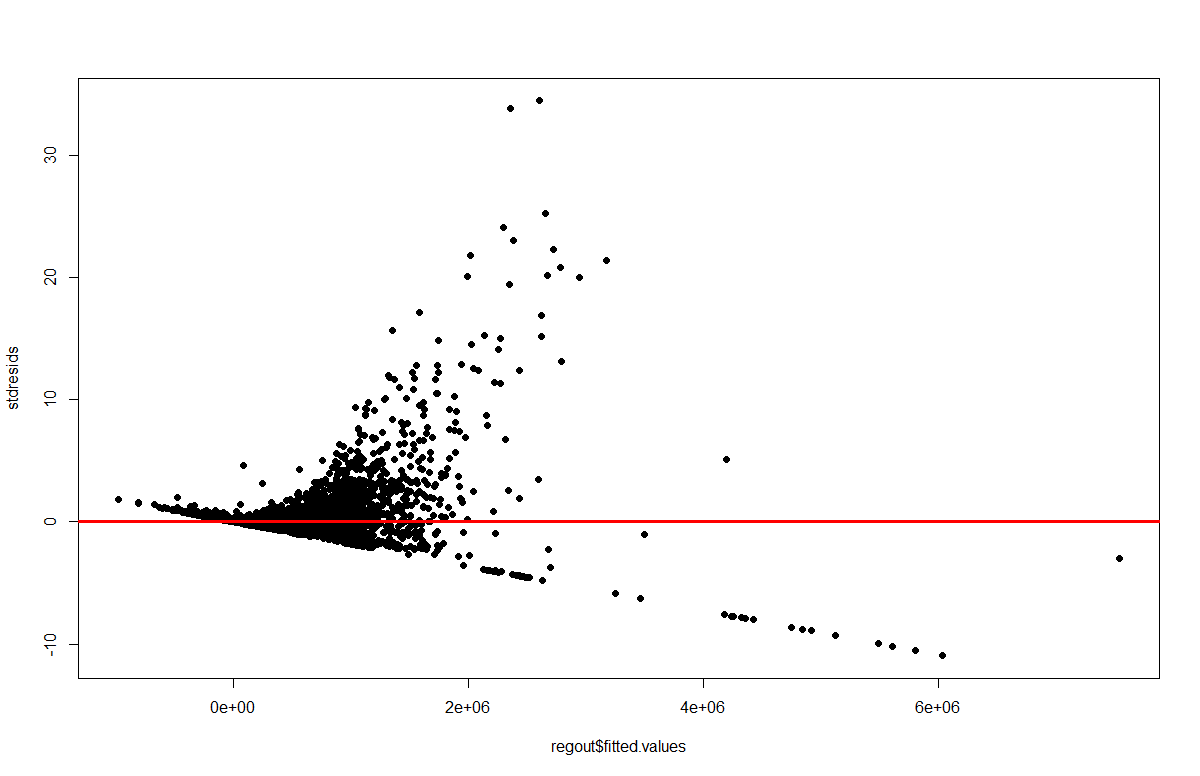


Figure Residual Plot for Regression Model with only continuous variables

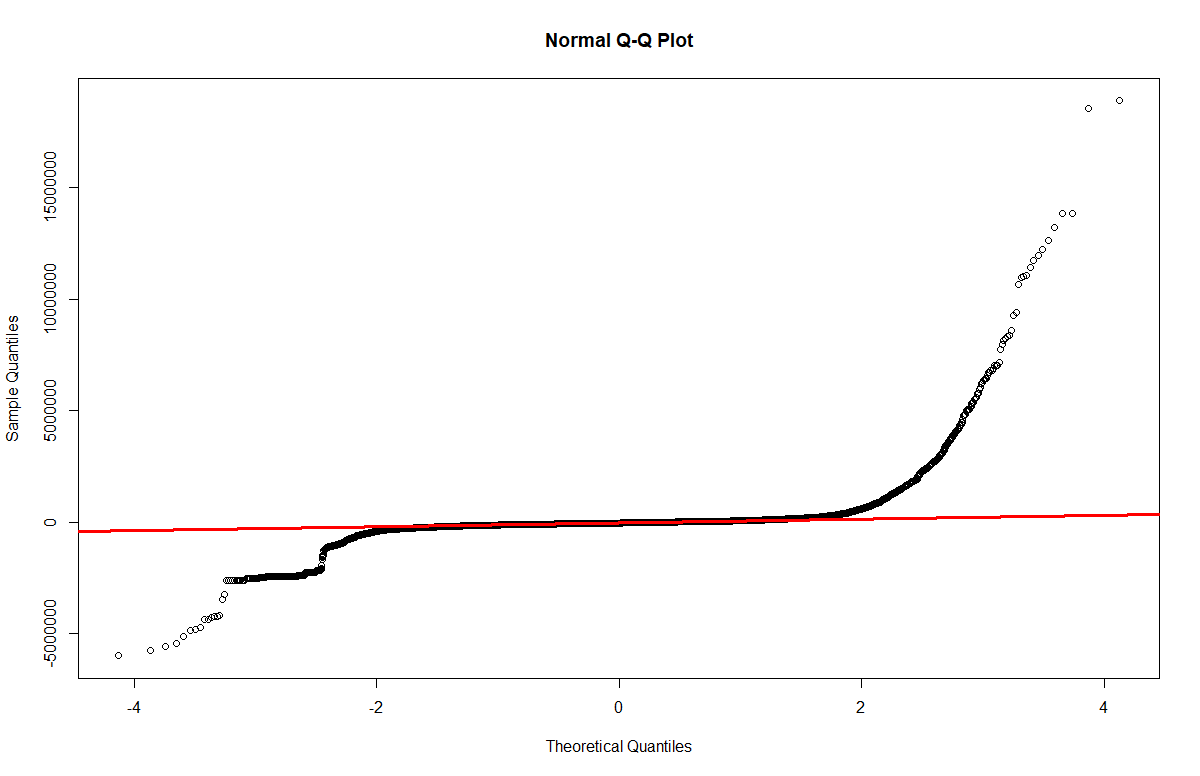


Figure QQ Plot for Regression Model with only continuous variables

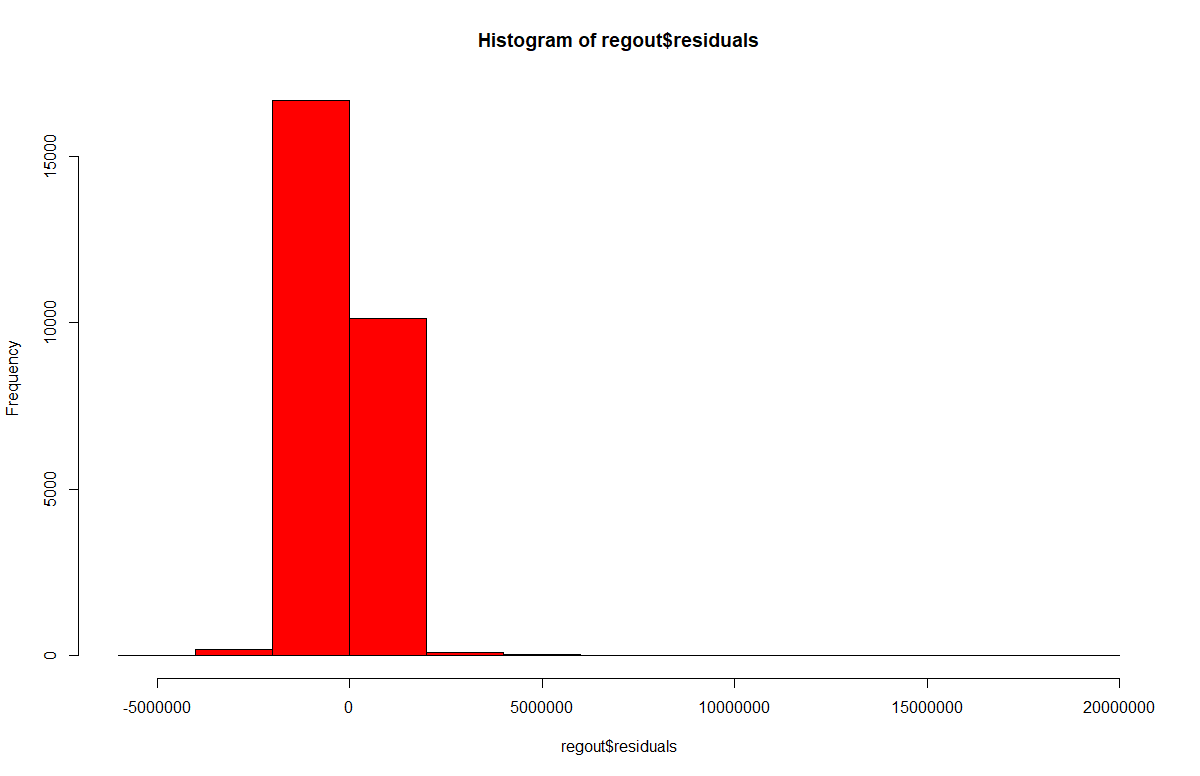


Figure Histogram of Residuals for Regression Model with only continuous variables

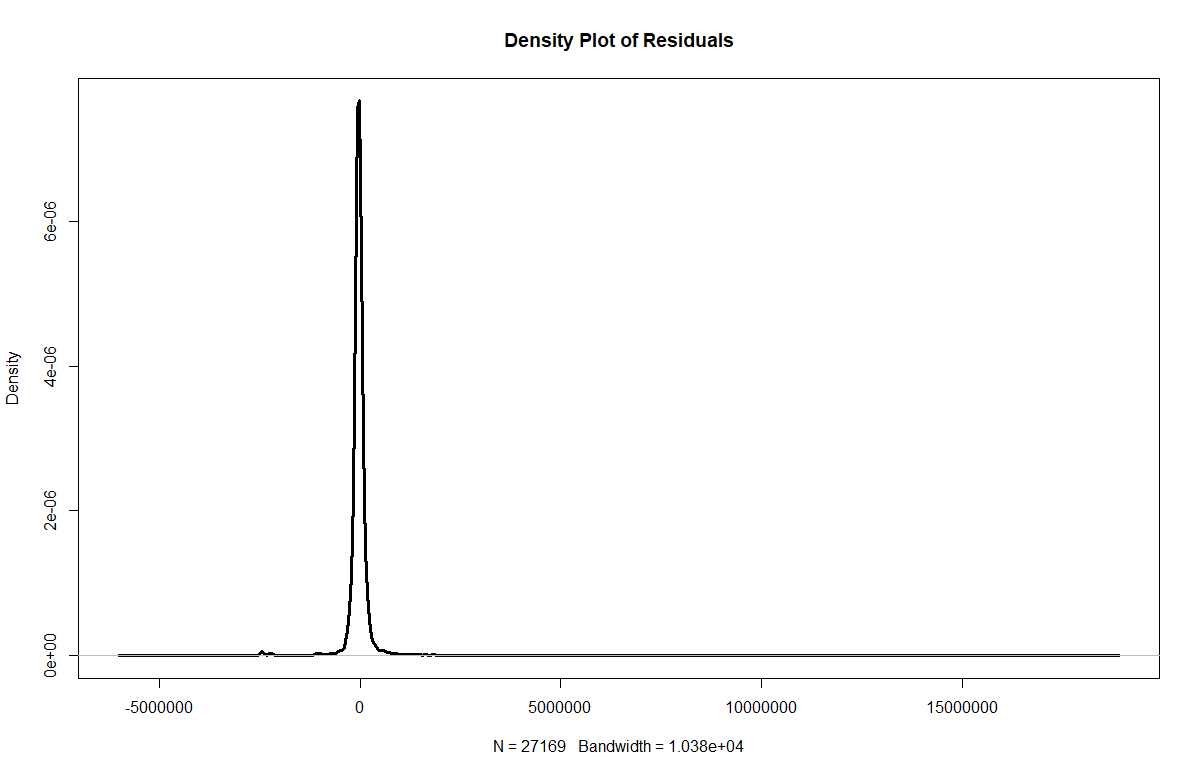


Figure Density Plot of Residuals for Regression Model with only continuous variables

To see if there is correlation between the independent variables:

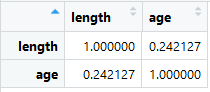


Figure Correlation of independent variables for Regression Model with only continuous variables

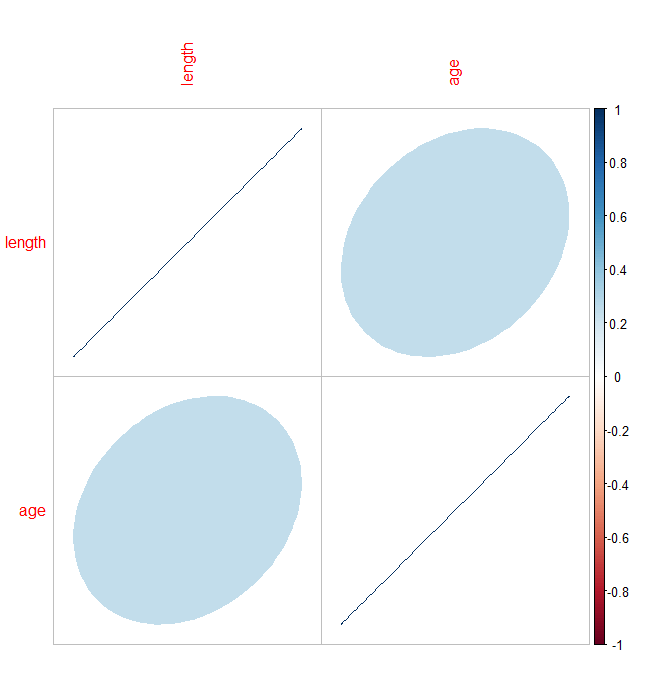


Figure Correlation Plot of Residuals for Regression Model with only continuous variables

As we can see in figures 19 and 20, there isn’t a strong relationship between these two variables as the correlation plot suggests.

The next step would be to include the categorial variables present

As we can see in figures 15 and 16, there seems to be violation of heteroskedasticity as well of linearity. In figures 17 and 18 Some slight normal behavior can be observed both in the histogram and the density plot of the residuals.

Including the continuous variables+ categorial variables in the decision model:

Due to the size in levels of the “category” and zip code variables, they were not included in this version of the linear regression model.

The linear regression model would be:

regout=lm(price ~ age+

length+

as.factor(region)+

as.factor(make)+

as.factor(material)+

as.factor(year)+

as.factor(region)+

as.factor(fuel),data = completed\_data)

From R, we get a better Multiple R Squared (0.3495) and a slightly lower Adjusted R Squared which suggest that there might be some overfitting in the model.

lm(formula = price ~ age + length + as.factor(region) + as.factor(make) +

as.factor(material) + as.factor(year) +

as.factor(fuel), data = completed\_data)

Residuals:

Min 1Q Median 3Q Max

-5231181 -122939 -10875 85488 18795065

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -366524.8 18970.4 -19.321 < 2e-16 \*\*\*

age -6393.7 4984.1 -1.283 0.199566

length 18527.9 241.0 76.873 < 2e-16 \*\*\*

as.factor(region)Gulf Coast 80817.4 10199.0 7.924 2.39e-15 \*\*\*

as.factor(region)NorthEast 13468.1 9691.1 1.390 0.164619

as.factor(region)Pacific -106155.8 14690.4 -7.226 5.10e-13 \*\*\*

as.factor(region)SorthEast 94754.6 10653.6 8.894 < 2e-16 \*\*\*

as.factor(make)Single Inboard -108736.2 14263.3 -7.623 2.55e-14 \*\*\*

as.factor(make)Single Outboard -39935.2 12149.4 -3.287 0.001014 \*\*

as.factor(make)Triple Outboard -44732.8 36198.2 -1.236 0.216552

as.factor(make)Twin Inboard 76386.0 11081.8 6.893 5.58e-12 \*\*\*

as.factor(make)Twin Outboard -38963.0 20458.9 -1.904 0.056862 .

as.factor(material)Composite 320318.6 28057.4 11.417 < 2e-16 \*\*\*

as.factor(material)Ferro cement 18062.9 167497.5 0.108 0.914123

as.factor(material)Fiberglass 59117.4 11515.9 5.134 2.86e-07 \*\*\*

as.factor(material)Hypalon 27350.9 215689.5 0.127 0.899094

as.factor(material)Other -66159.8 16533.7 -4.002 6.31e-05 \*\*\*

as.factor(material)Pvc 276695.2 527970.1 0.524 0.600232

as.factor(material)Steel 29197.8 51347.7 0.569 0.569613

as.factor(material)Wood 314636.0 52536.7 5.989 2.14e-09 \*\*\*

as.factor(year)1912 173725.3 647758.5 0.268 0.788552

as.factor(year)1922 278899.4 713981.9 0.391 0.696077

as.factor(year)1926 -1237699.6 700769.1 -1.766 0.077373 .

as.factor(year)1928 -403857.7 694223.4 -0.582 0.560746

as.factor(year)1929 -404532.9 540524.3 -0.748 0.454220

as.factor(year)1930 256157.9 687841.6 0.372 0.709592

as.factor(year)1931 -352641.4 684688.4 -0.515 0.606530

as.factor(year)1932 293097.7 679515.4 0.431 0.666229

as.factor(year)1933 525033.0 678454.0 0.774 0.439016

as.factor(year)1935 -655609.2 559375.2 -1.172 0.241192

as.factor(year)1936 -413835.4 669538.1 -0.618 0.536521

as.factor(year)1937 -321302.7 550403.7 -0.584 0.559388

as.factor(year)1938 784907.6 548275.7 1.432 0.152272

as.factor(year)1941 13737.1 654295.3 0.021 0.983250

as.factor(year)1944 -634523.6 525900.8 -1.207 0.227618

as.factor(year)1945 415766.9 477350.2 0.871 0.383768

as.factor(year)1946 -61724.8 513606.2 -0.120 0.904342

as.factor(year)1947 168479.1 514711.0 0.327 0.743423

as.factor(year)1948 -77844.4 634378.0 -0.123 0.902338

as.factor(year)1949 -174770.9 631649.0 -0.277 0.782020

as.factor(year)1950 -125344.3 506517.2 -0.247 0.804552

as.factor(year)1951 -392941.5 626371.7 -0.627 0.530448

as.factor(year)1953 -338841.7 402236.0 -0.842 0.399574

as.factor(year)1954 -310356.1 493412.3 -0.629 0.529355

as.factor(year)1955 -195896.4 438889.4 -0.446 0.655351

as.factor(year)1956 -467410.3 378101.2 -1.236 0.216393

as.factor(year)1957 -360298.5 405761.5 -0.888 0.374572

as.factor(year)1958 -324733.7 382153.5 -0.850 0.395473

as.factor(year)1959 -341573.0 396980.5 -0.860 0.389561

as.factor(year)1960 -134930.3 327766.9 -0.412 0.680588

as.factor(year)1961 -213407.7 340242.9 -0.627 0.530519

as.factor(year)1962 -148552.7 336152.1 -0.442 0.658550

as.factor(year)1963 -130599.2 321104.5 -0.407 0.684218

as.factor(year)1964 -232422.7 307410.7 -0.756 0.449616

as.factor(year)1965 -254946.7 305588.8 -0.834 0.404130

as.factor(year)1966 -188252.0 280139.0 -0.672 0.501593

as.factor(year)1967 -222979.2 280680.8 -0.794 0.426956

as.factor(year)1968 -231887.1 276108.8 -0.840 0.401006

as.factor(year)1969 -190349.4 261263.5 -0.729 0.466270

as.factor(year)1970 -173571.2 265844.1 -0.653 0.513822

as.factor(year)1971 -275462.9 262526.5 -1.049 0.294060

as.factor(year)1972 -150227.0 242659.3 -0.619 0.535865

as.factor(year)1973 -239459.7 239585.5 -0.999 0.317574

as.factor(year)1974 -231096.2 235271.7 -0.982 0.325984

as.factor(year)1975 -220599.4 226988.5 -0.972 0.331132

as.factor(year)1976 -240175.9 221417.6 -1.085 0.278056

as.factor(year)1977 -261144.6 212253.5 -1.230 0.218579

as.factor(year)1978 -209748.1 207273.5 -1.012 0.311576

as.factor(year)1979 -232328.1 201323.4 -1.154 0.248509

as.factor(year)1980 -228145.9 198014.4 -1.152 0.249262

as.factor(year)1981 -242411.3 192187.1 -1.261 0.207201

as.factor(year)1982 -243765.0 187168.0 -1.302 0.192796

as.factor(year)1983 -227411.2 180206.1 -1.262 0.206978

as.factor(year)1984 -221848.4 174285.4 -1.273 0.203063

as.factor(year)1985 -244454.4 168628.9 -1.450 0.147165

as.factor(year)1986 -235955.4 162944.9 -1.448 0.147609

as.factor(year)1987 -245801.3 157797.6 -1.558 0.119316

as.factor(year)1988 -248412.9 152524.1 -1.629 0.103393

as.factor(year)1989 -248001.8 147954.0 -1.676 0.093709 .

as.factor(year)1990 -267406.6 143950.9 -1.858 0.063233 .

as.factor(year)1991 -272652.7 142003.2 -1.920 0.054863 .

as.factor(year)1992 -241459.2 137037.3 -1.762 0.078081 .

as.factor(year)1993 -235361.7 131539.7 -1.789 0.073580 .

as.factor(year)1994 -231415.1 125081.6 -1.850 0.064308 .

as.factor(year)1995 -228514.6 119215.0 -1.917 0.055270 .

as.factor(year)1996 -221241.9 113960.1 -1.941 0.052220 .

as.factor(year)1997 -246258.7 108936.2 -2.261 0.023793 \*

as.factor(year)1998 -252017.5 103610.3 -2.432 0.015007 \*

as.factor(year)1999 -240618.3 98057.1 -2.454 0.014139 \*

as.factor(year)2000 -209371.1 92876.3 -2.254 0.024185 \*

as.factor(year)2001 -232994.3 88052.8 -2.646 0.008148 \*\*

as.factor(year)2002 -202160.0 83678.4 -2.416 0.015702 \*

as.factor(year)2003 -164476.6 78395.0 -2.098 0.035910 \*

as.factor(year)2004 -172087.0 73424.9 -2.344 0.019100 \*

as.factor(year)2005 -176718.3 68211.1 -2.591 0.009582 \*\*

as.factor(year)2006 -166044.3 63233.6 -2.626 0.008647 \*\*

as.factor(year)2007 -161879.9 58958.0 -2.746 0.006043 \*\*

as.factor(year)2008 -86484.3 54668.6 -1.582 0.113667

as.factor(year)2009 -86358.2 53211.5 -1.623 0.104617

as.factor(year)2010 100580.5 51486.9 1.954 0.050769 .

as.factor(year)2011 -26655.9 44894.6 -0.594 0.552688

as.factor(year)2012 -53311.3 39489.8 -1.350 0.177027

as.factor(year)2013 -1308.4 34620.3 -0.038 0.969853

as.factor(year)2014 -19842.6 30441.8 -0.652 0.514522

as.factor(year)2015 -925.6 26863.0 -0.034 0.972513

as.factor(year)2016 66715.1 23831.2 2.799 0.005122 \*\*

as.factor(year)2017 88126.2 21595.4 4.081 4.50e-05 \*\*\*

as.factor(year)2018 78642.0 18304.7 4.296 1.74e-05 \*\*\*

as.factor(year)2019 -59558.5 14296.3 -4.166 3.11e-05 \*\*\*

as.factor(year)2020 NA NA NA NA

as.factor(fuel)Diesel 212545.2 15292.8 13.898 < 2e-16 \*\*\*

as.factor(fuel)Electric 392726.5 107164.8 3.665 0.000248 \*\*\*

as.factor(fuel)Gas -7835.4 11064.2 -0.708 0.478843

as.factor(fuel)Other -35675.9 15812.6 -2.256 0.024068 \*

as.factor(fuel)Propane 317135.1 528085.2 0.601 0.548153

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 527300 on 27055 degrees of freedom

Multiple R-squared: 0.3495, Adjusted R-squared: 0.3468

F-statistic: 128.6 on 113 and 27055 DF, p-value: < 2.2e-16

In a similar fashion, revising the LINE requirements of normality, I still see elements which suggest that the model is violating heteroskedasticity and linearity as shown in Figure 21, and barely shows some normal appearance in Figures 22 to 24.

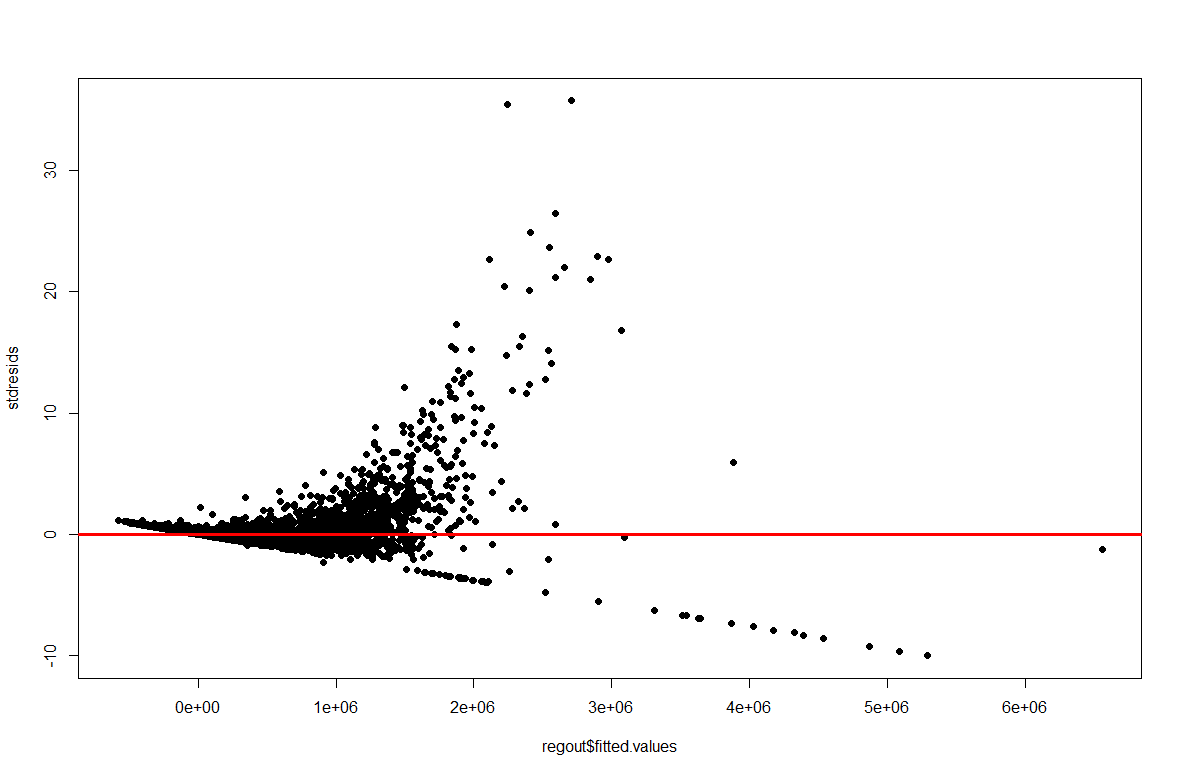


Figure Residual Plot for Regression Model with only continuous variables+ categorical variables

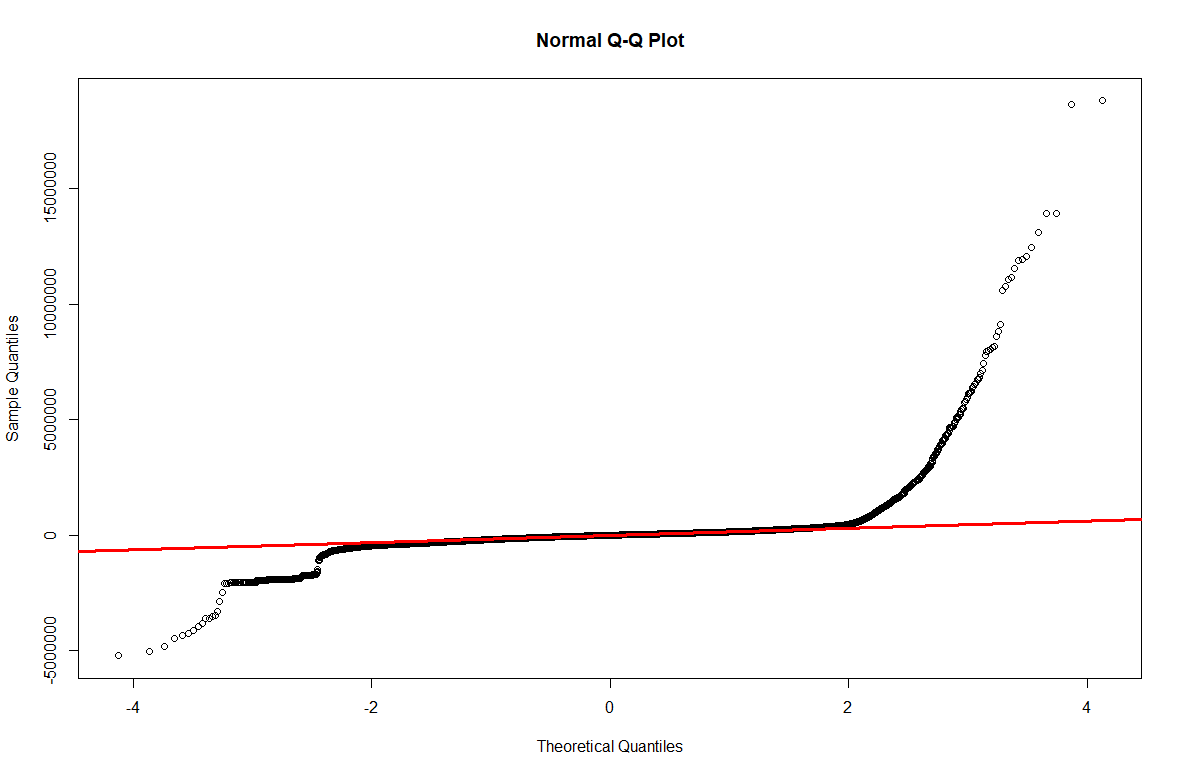


Figure QQ Plot for Regression Model with only continuous variables+ categorical variables

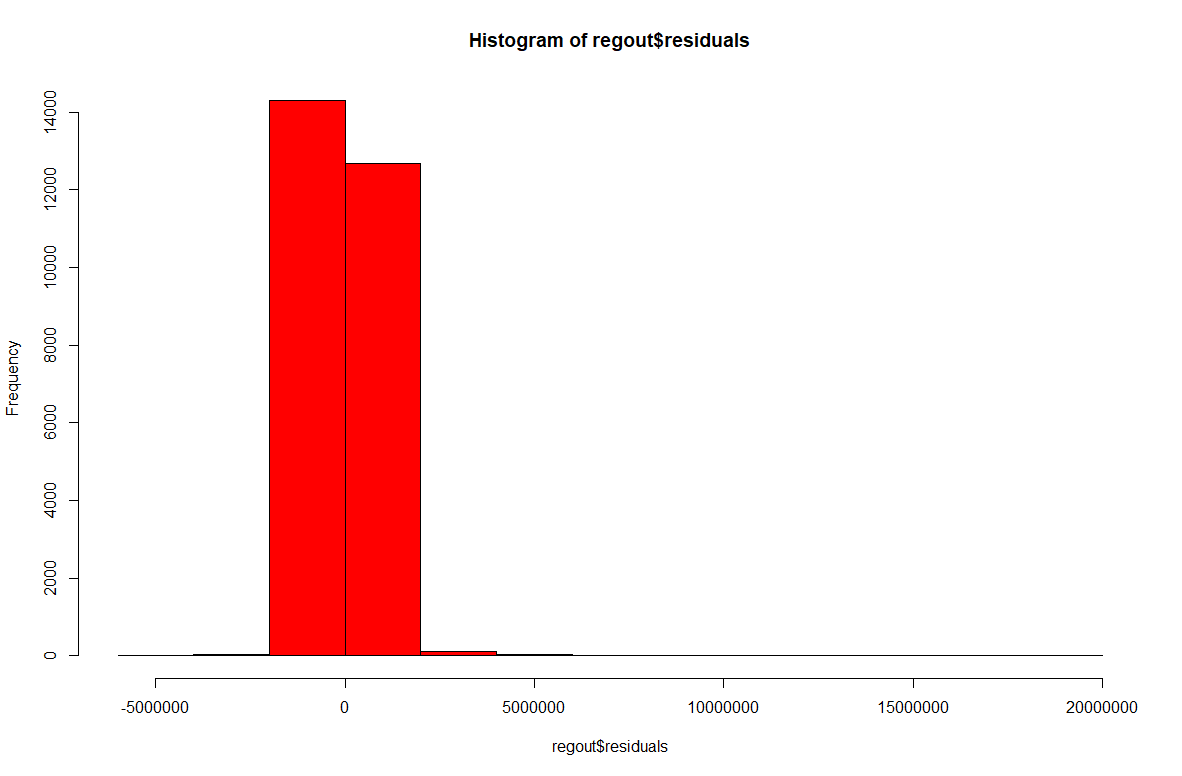


Figure Histogram of Residuals for Regression Model with only continuous variables+ categorical variables

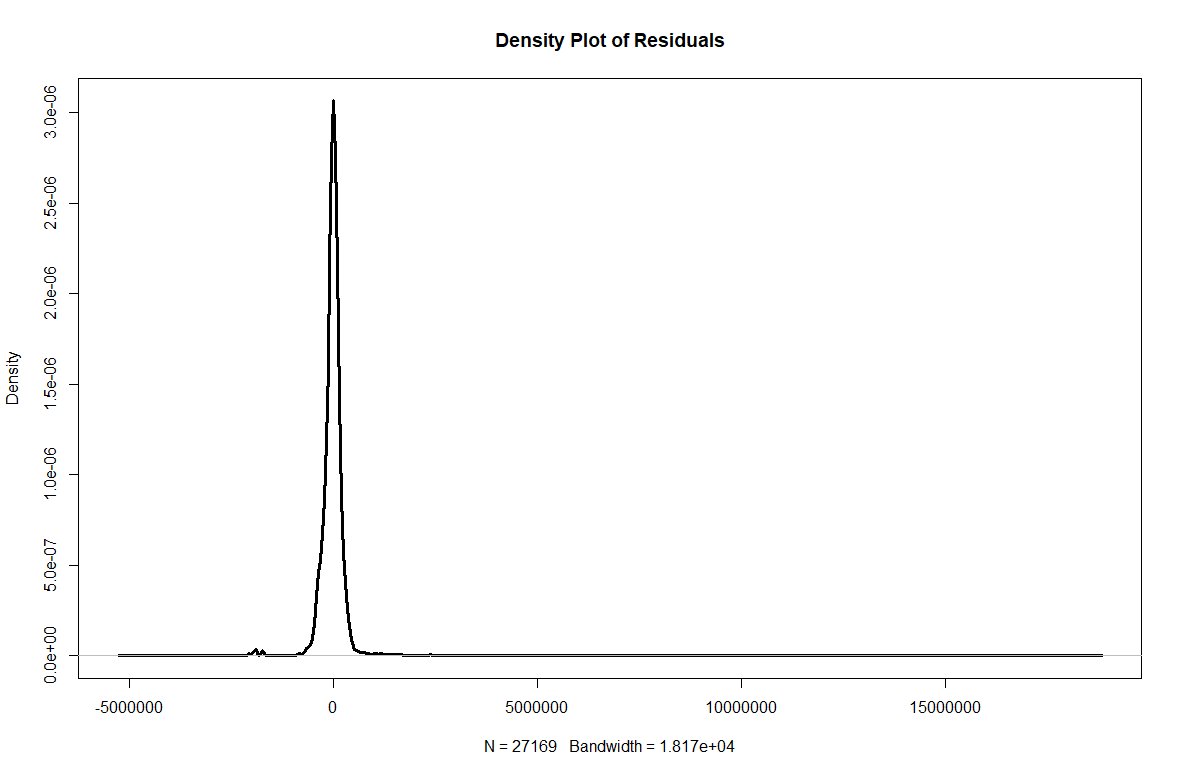


Figure Density of Residuals for Regression Model with only continuous variables+ categorical variables

In appendix I, a linear regression model running all continuous and categorical was appended. Notice how the R Square improved:

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 441300 on 24371 degrees of freedom

Multiple R-squared: 0.5896, Adjusted R-squared: 0.5425

F-statistic: 12.52 on 2797 and 24371 DF, p-value: < 2.2e-16

Since we have data that seems to not follow linear patterns, one approach that will be tested next is the Generalized Additive Model, which has the flexibility to be able o work around complex situations.

Generalized Additive Model:

The model was designed as the last iteration of the linear model that ran previously (all continuous+categorial variables):

regout2=gam(completed\_data$price ~ completed\_data$age+

completed\_data$length+

as.factor(completed\_data$region)+

as.factor(completed\_data$make)+

as.factor(completed\_data$material)+

as.factor(completed\_data$year)+

as.factor(completed\_data$fuel)+

as.factor(completed\_data$zipcode)+

as.factor(completed\_data$category))

summary(regout2)

The entire output of the GAM model has been put in section II of the Appendix, for now we can se below that the R Square is almost like the one obtained in the last linear regression model with all the continuous and categorial variables:

Rank: 2798/2809

R-sq.(adj) = 0.542 Deviance explained = 59%

GCV = 2.1708e+11 Scale est. = 1.9472e+11 n = 27169

Interactions

From the data visualizations shown before, the following variables where chosen as candidates for interaction due to the relationship they had geographically and their effect on the price:

* Length
* Age
* Region

Several runs where made with each of these combinations of interactions, and the end the implementation o the interactions between Length-Region, Age-Region and Age-Length was included in the model

#Interaction Between Length/Region, Age/Region and Age/Length

regout\_interaction4=lm(price ~ age+

length+

as.factor(make)+

as.factor(material)+

as.factor(year)+

as.factor(region)+

as.factor(fuel)+

as.factor(category)+

as.factor(zipcode)+age\*

as.factor(region)+

age\*length+

as.factor(region)\*length,data = completed\_data)

This generated a much better R-Squared in the output of the execution of the model.

Residual standard error: 392100 on 24362 degrees of freedom

Multiple R-squared: 0.6762, Adjusted R-squared: 0.6389

F-statistic: 18.13 on 2806 and 24362 DF, p-value: < 2.2e-16

Revising the LINE requirements of normality, even though this model seems to have a good R squared compared to other model to explain all the variability of price in term of the independent variables, I still see elements which suggest that the model is violating heteroskedasticity and linearity as shown in Figure 25, and barely shows some normal appearance in Figures 26 to 28.

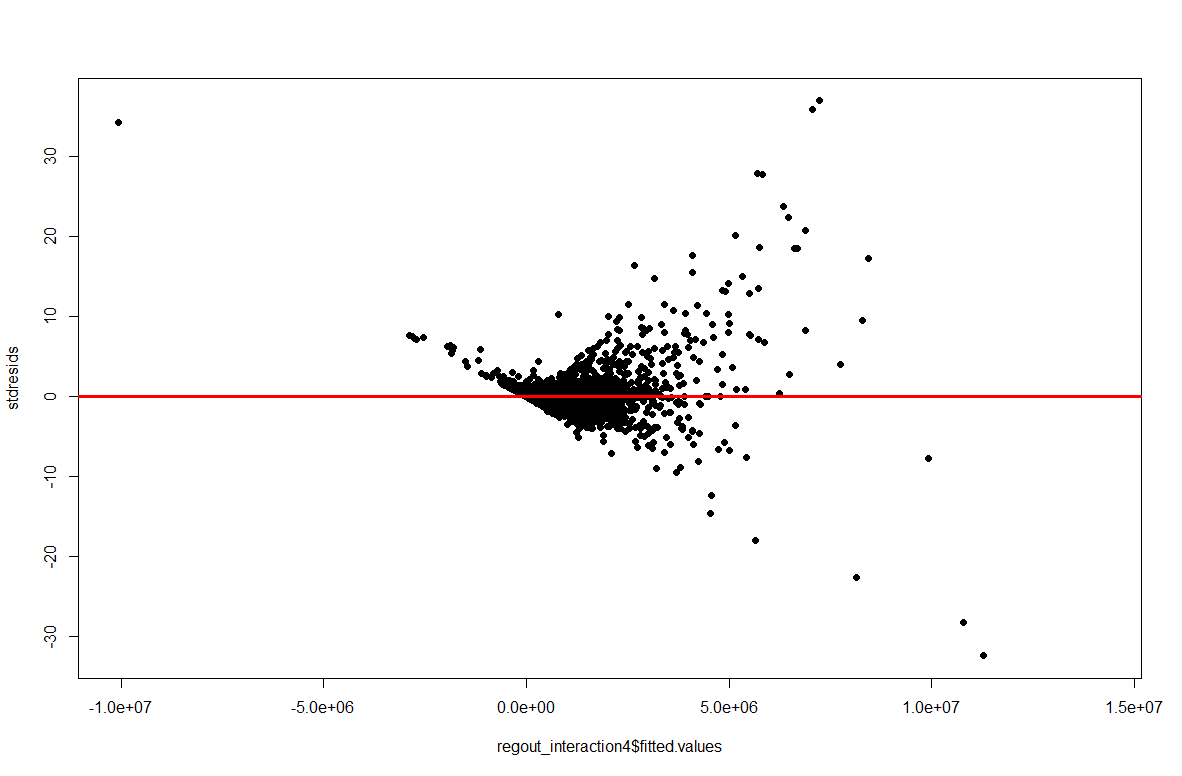


Figure Residual Plot for Regression Model with Interaction variables

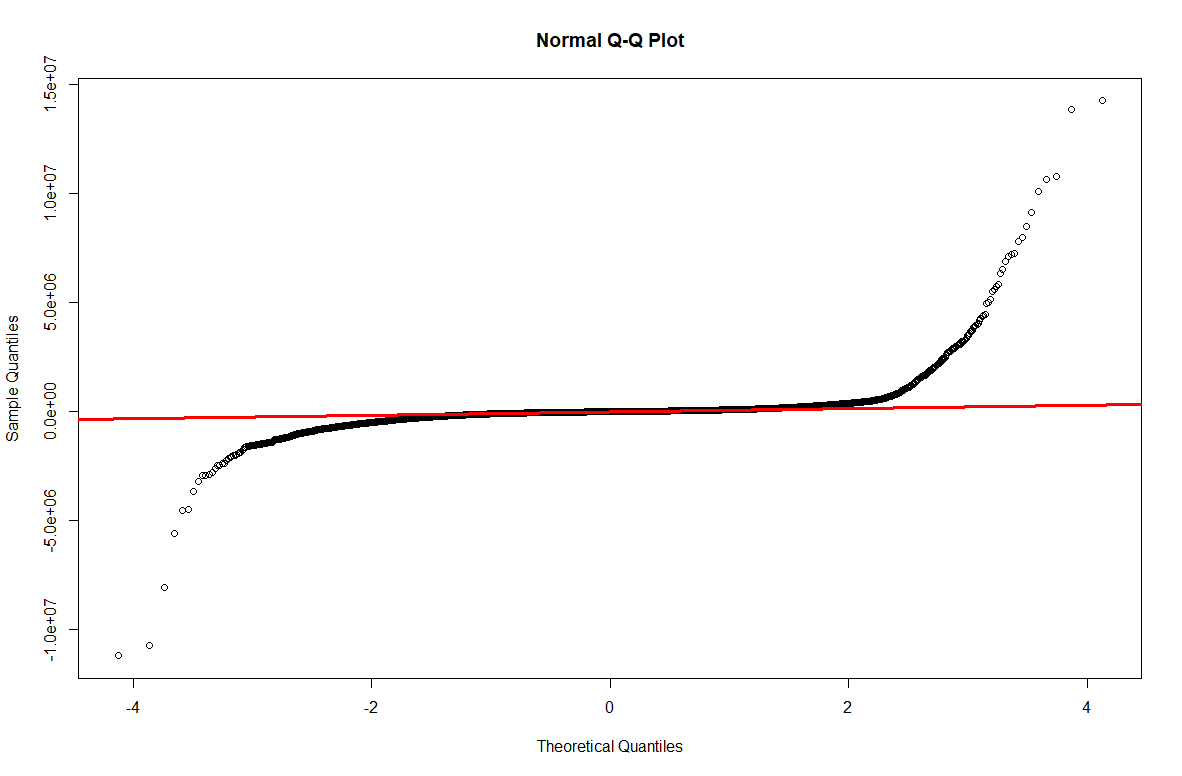


Figure QQ Plot for Regression Model with Interaction variables

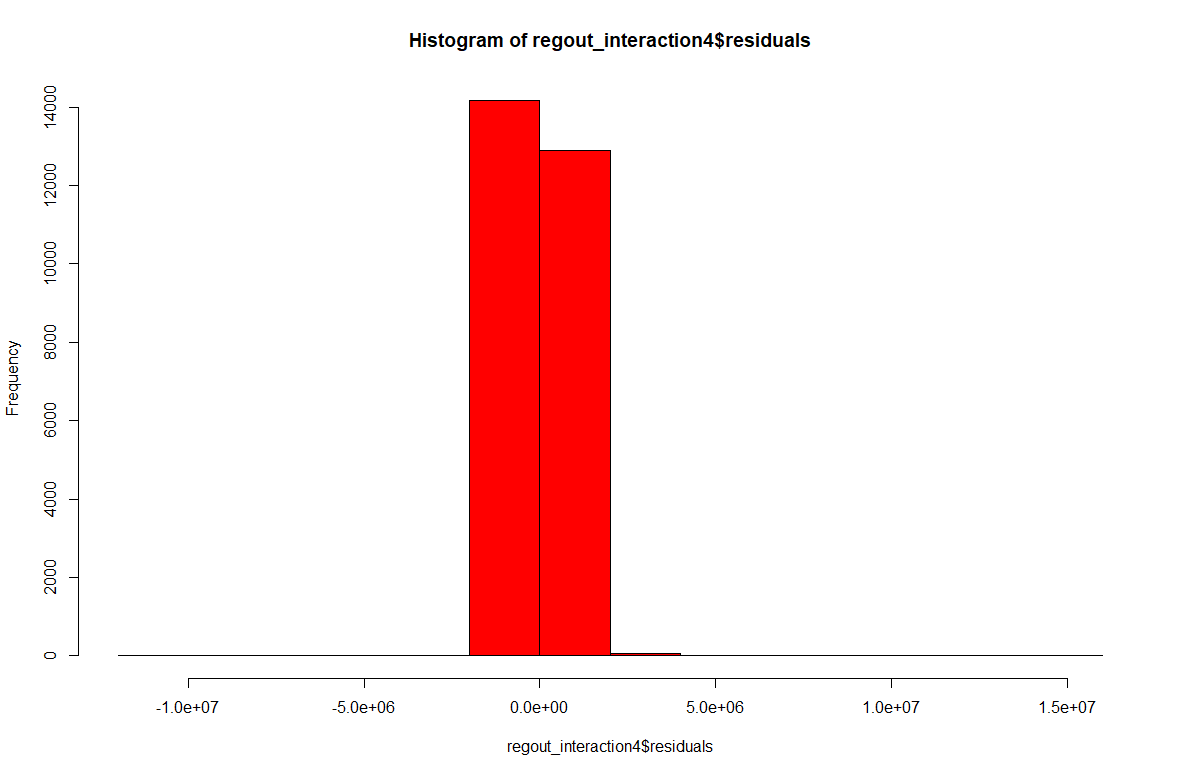


Figure Histogram of Residuals for Regression Model with Interaction variables

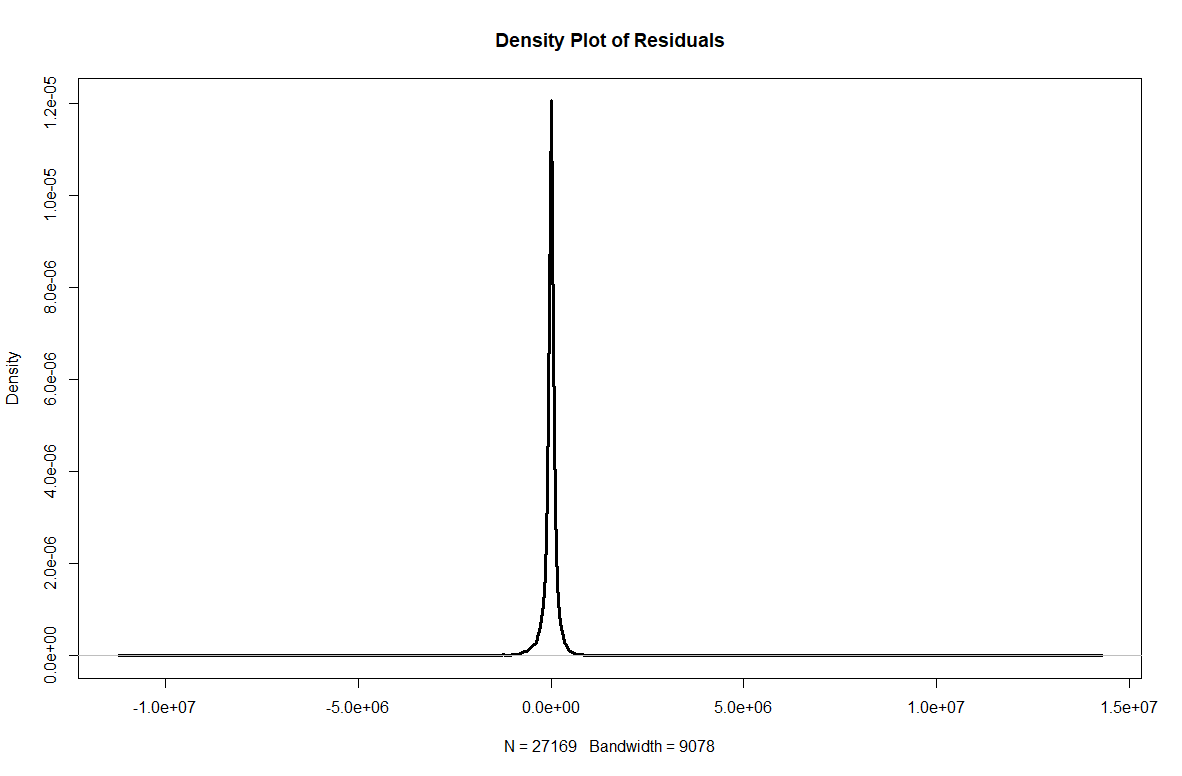


Figure Density of Residuals for Regression Model with Interaction Variables

**4) Summary of results**

From all the stages of this process, the following information could be discovered about the Boat Trader web site.

* From Figure 5, we can see that in terms of the data that was web scrapped for this project, the places with the lowest mean age tend to be in the Easter US. The places where the mean age appeared to be the highest where in the Pacific region and some portion of the South East US
* From Figures 6 and 7, we learned that the costliest listings of boats where located in the Eastern US. The max value found was $21.500.000
* In terms of boat length, many boats listed tend to have a size of 100 or less average length. There where some places like Central US and South E where average lengths of 260 or near where found. This can bee see in Figure 8.
* Much of the listings for this dataset seemed to be centered around the South East as well as the North Western Gulf Coast, as seen in Figure 9.
* Many of the listings seemed to be of boating of length less than 100, as seen in Figure 10.
* Most boats seem to have an age of 20 years or less, as seen in Figure 11.
* When checking the age of the listed boat, most of the boats seem to be from the year 2000 and onwards, which makes sense, people would normally like to purchase new boats, as seen in Figure 12
* Length and price seem to have a natural relationship that the longer the boat, the price increases as seen in Figure 13
* Similarly, for age, the shorter the age, the more the value of the boat as seen in Figure 14. For certain vintage boats, there seems to be some increased value for some for some historic/ memorabilia value.
* When building regression model with only the continuous variables age and length, they both seem to play a significance in prince, although that model can be improved as shown in later adaptions.
* Including the categorical variables in the model increased the performance of the regression model in identifying the change of price in terms of the other variables.
* One common element throughout the project is that there seems to violations in LINE assumptions of normality.
* The General Additive Model was implemented based on the previous idea. Similar results where achieved in terms of model performance.
* From all the information gathered from the visualizations of the listings throughout the US, it was identified that length, age and region must have some level of interaction to affect price. Because of this, a lineal regression model was made that included the interactions between Length-Region, Age-Region and Age-Length. This model at the end up in the end having the best performance of the project, compared to the ones developed in earlier stages.
* Because of this, it can be noted that region of the listen, the age of the boat and its length play a key role in its price.

**5) Comments and recommendations for future iterations.**

This project was a great hands opportunity to put together diverse tools that I have learned throughout the MS BAIS program. In this case, the reuse of the Python script for web scrapping was a key take way during the development of this project. The application of this can be endless in search of generating business opportunity from using untapped data sources hidden on websites. The use of Tableau complements the information obtained in R using the tools obtained during our statistical datamining course., helping see certain interaction among variables.

Future iteration of this project can be performed with this same website, to apply how the regression model behaves when logs are implemented on the dependent/and or independent variables, specially since in this case it was determined that data did not comply with normality. This would be an opportunity to develop better models that relate the change of price in terms of the studied variables.

Also, it would be a great addition to further projects to implement cluster analysis, which could prove to be useful in business settings to identify opportunities.

**Appendix**

I. Result of Linear Regression taking into consideration all continuous as well as categorical variables

|  |
| --- |
| > summary(regout)  Call:  lm(formula = price ~ age + length + as.factor(region) + as.factor(make) +  as.factor(material) + as.factor(year) + as.factor(region) +  as.factor(fuel) + as.factor(category) + as.factor(zipcode),  data = completed\_data)  Residuals:  Min 1Q Median 3Q Max  -8453958 -91624 0 72013 16689148  Coefficients: (11 not defined because of singularities)  Estimate Std. Error t value Pr(>|t|)  (Intercept) -8.915e+05 4.458e+05 -2.000 0.045532  age -5.074e+03 4.661e+03 -1.089 0.276316  length 3.520e+04 3.520e+02 99.988 < 2e-16  as.factor(region)Gulf Coast 2.189e+05 3.725e+04 5.877 4.24e-09  as.factor(region)NorthEast -3.190e+04 3.172e+04 -1.006 0.314638  as.factor(region)Pacific -1.049e+05 2.972e+04 -3.529 0.000418  as.factor(region)SorthEast 5.814e+04 2.906e+04 2.001 0.045427  as.factor(make)Single Inboard 7.399e+03 1.554e+04 0.476 0.634063  as.factor(make)Single Outboard -1.788e+04 1.265e+04 -1.413 0.157581  as.factor(make)Triple Outboard -1.759e+05 3.418e+04 -5.145 2.70e-07  as.factor(make)Twin Inboard -1.337e+04 1.159e+04 -1.154 0.248706  as.factor(make)Twin Outboard -5.189e+04 2.016e+04 -2.574 0.010073  as.factor(material)Composite 2.006e+05 3.046e+04 6.585 4.64e-11  as.factor(material)Ferro cement -4.376e+04 1.607e+05 -0.272 0.785386  as.factor(material)Fiberglass 2.524e+04 1.652e+04 1.528 0.126553  as.factor(material)Hypalon 2.623e+04 1.890e+05 0.139 0.889607  as.factor(material)Other 1.405e+04 2.014e+04 0.697 0.485627  as.factor(material)Pvc 2.366e+05 4.482e+05 0.528 0.597673  as.factor(material)Steel -2.641e+04 5.201e+04 -0.508 0.611562  as.factor(material)Wood 2.143e+05 5.428e+04 3.947 7.93e-05  as.factor(year)1912 2.429e+05 7.368e+05 0.330 0.741692  as.factor(year)1922 3.307e+03 6.305e+05 0.005 0.995815  as.factor(year)1926 -1.970e+06 6.329e+05 -3.113 0.001854  as.factor(year)1928 -2.098e+06 6.605e+05 -3.177 0.001491  as.factor(year)1929 -7.513e+05 6.590e+05 -1.140 0.254229  as.factor(year)1930 2.002e+05 6.025e+05 0.332 0.739624  as.factor(year)1931 -7.245e+05 6.029e+05 -1.202 0.229488  as.factor(year)1932 1.051e+04 6.742e+05 0.016 0.987560  as.factor(year)1933 6.116e+05 5.998e+05 1.020 0.307872  as.factor(year)1935 -1.703e+06 5.928e+05 -2.873 0.004070  as.factor(year)1936 -2.440e+06 7.485e+05 -3.259 0.001118  as.factor(year)1937 -7.697e+05 5.851e+05 -1.316 0.188340  as.factor(year)1938 5.424e+05 5.168e+05 1.049 0.293972  as.factor(year)1941 3.302e+04 5.735e+05 0.058 0.954089  as.factor(year)1944 -1.123e+06 4.790e+05 -2.345 0.019036  as.factor(year)1945 -1.170e+05 5.655e+05 -0.207 0.836094  as.factor(year)1946 -4.165e+04 4.897e+05 -0.085 0.932223  as.factor(year)1947 5.202e+04 4.547e+05 0.114 0.908929  as.factor(year)1948 -3.183e+05 5.692e+05 -0.559 0.576048  as.factor(year)1949 8.305e+04 5.691e+05 0.146 0.883976  as.factor(year)1950 -2.610e+05 5.479e+05 -0.476 0.633793  as.factor(year)1951 -2.842e+05 5.500e+05 -0.517 0.605407  as.factor(year)1953 -1.271e+06 4.194e+05 -3.030 0.002449  as.factor(year)1954 -8.442e+05 7.544e+05 -1.119 0.263125  as.factor(year)1955 2.224e+04 4.339e+05 0.051 0.959118  as.factor(year)1956 -1.028e+06 3.711e+05 -2.770 0.005608  as.factor(year)1957 -5.159e+05 4.014e+05 -1.285 0.198709  as.factor(year)1958 -4.606e+05 3.836e+05 -1.201 0.229883  as.factor(year)1959 -2.641e+05 3.853e+05 -0.685 0.493043  as.factor(year)1960 -1.682e+05 3.138e+05 -0.536 0.592037  as.factor(year)1961 -4.897e+05 3.139e+05 -1.560 0.118795  as.factor(year)1962 -2.523e+05 3.117e+05 -0.809 0.418288  as.factor(year)1963 -3.193e+05 3.021e+05 -1.057 0.290498  as.factor(year)1964 -2.032e+05 2.846e+05 -0.714 0.475155  as.factor(year)1965 -4.460e+05 2.909e+05 -1.533 0.125282  as.factor(year)1966 -2.897e+05 2.601e+05 -1.114 0.265281  as.factor(year)1967 -2.534e+05 2.618e+05 -0.968 0.333185  as.factor(year)1968 -5.511e+05 2.580e+05 -2.136 0.032687  as.factor(year)1969 -1.380e+05 2.429e+05 -0.568 0.569905  as.factor(year)1970 -1.699e+05 2.465e+05 -0.689 0.490728  as.factor(year)1971 -2.442e+05 2.453e+05 -0.995 0.319591  as.factor(year)1972 -1.485e+05 2.260e+05 -0.657 0.511146  as.factor(year)1973 -3.146e+05 2.207e+05 -1.426 0.154016  as.factor(year)1974 -2.676e+05 2.198e+05 -1.217 0.223553  as.factor(year)1975 -2.123e+05 2.129e+05 -0.998 0.318521  as.factor(year)1976 -2.619e+05 2.062e+05 -1.270 0.203974  as.factor(year)1977 -3.539e+05 1.980e+05 -1.787 0.073909  as.factor(year)1978 -2.471e+05 1.933e+05 -1.278 0.201189  as.factor(year)1979 -2.748e+05 1.881e+05 -1.460 0.144174  as.factor(year)1980 -2.835e+05 1.847e+05 -1.535 0.124763  as.factor(year)1981 -2.743e+05 1.792e+05 -1.530 0.125968  as.factor(year)1982 -2.631e+05 1.749e+05 -1.504 0.132644  as.factor(year)1983 -2.646e+05 1.679e+05 -1.576 0.115092  as.factor(year)1984 -2.528e+05 1.626e+05 -1.554 0.120160  as.factor(year)1985 -3.145e+05 1.571e+05 -2.002 0.045243  as.factor(year)1986 -2.633e+05 1.519e+05 -1.733 0.083062  as.factor(year)1987 -3.055e+05 1.473e+05 -2.074 0.038052  as.factor(year)1988 -3.031e+05 1.426e+05 -2.126 0.033527  as.factor(year)1989 -3.346e+05 1.381e+05 -2.423 0.015418  as.factor(year)1990 -3.261e+05 1.342e+05 -2.429 0.015127  as.factor(year)1991 -3.166e+05 1.324e+05 -2.391 0.016809  as.factor(year)1992 -3.762e+05 1.274e+05 -2.952 0.003159  as.factor(year)1993 -2.685e+05 1.226e+05 -2.190 0.028538  as.factor(year)1994 -2.761e+05 1.167e+05 -2.367 0.017962  as.factor(year)1995 -2.799e+05 1.113e+05 -2.516 0.011876  as.factor(year)1996 -2.765e+05 1.063e+05 -2.602 0.009274  as.factor(year)1997 -2.956e+05 1.016e+05 -2.910 0.003613  as.factor(year)1998 -2.989e+05 9.665e+04 -3.093 0.001985  as.factor(year)1999 -2.906e+05 9.152e+04 -3.175 0.001501  as.factor(year)2000 -2.674e+05 8.670e+04 -3.084 0.002042  as.factor(year)2001 -2.821e+05 8.224e+04 -3.430 0.000604  as.factor(year)2002 -2.382e+05 7.798e+04 -3.055 0.002253  as.factor(year)2003 -1.991e+05 7.322e+04 -2.719 0.006545  as.factor(year)2004 -2.132e+05 6.864e+04 -3.106 0.001896  as.factor(year)2005 -2.234e+05 6.377e+04 -3.504 0.000460  as.factor(year)2006 -2.116e+05 5.918e+04 -3.575 0.000351  as.factor(year)2007 -2.092e+05 5.514e+04 -3.795 0.000148  as.factor(year)2008 -1.459e+05 5.109e+04 -2.856 0.004287  as.factor(year)2009 -1.559e+05 4.945e+04 -3.152 0.001624  as.factor(year)2010 -1.828e+04 4.738e+04 -0.386 0.699616  as.factor(year)2011 -1.166e+05 4.162e+04 -2.801 0.005102  as.factor(year)2012 -1.371e+05 3.666e+04 -3.740 0.000185  as.factor(year)2013 -4.745e+04 3.228e+04 -1.470 0.141594  as.factor(year)2014 -4.551e+04 2.849e+04 -1.597 0.110174  as.factor(year)2015 -3.273e+04 2.508e+04 -1.305 0.191904  as.factor(year)2016 2.327e+04 2.227e+04 1.045 0.295932  as.factor(year)2017 6.759e+04 2.027e+04 3.334 0.000856  as.factor(year)2018 5.408e+04 1.727e+04 3.131 0.001744  as.factor(year)2019 -1.116e+04 1.341e+04 -0.832 0.405334  as.factor(year)2020 NA NA NA NA  as.factor(fuel)Diesel -9.068e+04 1.646e+04 -5.510 3.62e-08  as.factor(fuel)Electric 8.705e+04 1.208e+05 0.721 0.471153  as.factor(fuel)Gas -6.744e+03 1.207e+04 -0.559 0.576337  as.factor(fuel)Other -2.726e+04 1.950e+04 -1.398 0.162073  as.factor(fuel)Propane 1.931e+05 4.943e+05 0.391 0.696105  as.factor(category)Aft Cabin, Cruisers 1.086e+05 1.746e+05 0.622 0.534162  as.factor(category)Aft Cabin, Cruisers, Motor Yachts 1.635e+05 4.477e+05 0.365 0.714931  as.factor(category)Aft Cabin, Express Cruiser 6.168e+04 3.212e+05 0.192 0.847702  as.factor(category)Aft Cabin, Flybridge, Motor Yachts -1.992e+03 1.486e+05 -0.013 0.989306  as.factor(category)Aft Cabin, Motor Yachts -2.893e+04 7.345e+04 -0.394 0.693699  as.factor(category)Aft Cabin, Motor Yachts, Express Cruiser -4.112e+04 4.444e+05 -0.093 0.926277  as.factor(category)Aft Cabin, Motor Yachts, Flybridge -1.807e+05 1.753e+05 -1.031 0.302691  as.factor(category)Aft Cabin, Motor Yachts, Trawlers -9.065e+04 1.683e+05 -0.539 0.590173  as.factor(category)Aft Cabin, Pilothouse, Flybridge 1.649e+05 4.476e+05 0.368 0.712548  as.factor(category)Aft Cabin, Sports Fishing Boats, Trawlers 3.766e+04 5.447e+05 0.069 0.944884  as.factor(category)Aft Cabin, Trawlers, Motor Yachts -5.079e+03 1.199e+05 -0.042 0.966205  as.factor(category)Aluminum Fish Boats 7.568e+04 5.494e+04 1.378 0.168344  as.factor(category)Antique and Classics -1.345e+04 1.383e+05 -0.097 0.922490  as.factor(category)Antique and Classics, Runabouts 1.361e+05 2.242e+05 0.607 0.544027  as.factor(category)Antique and Classics, Sloop -2.721e+04 3.193e+05 -0.085 0.932096  as.factor(category)Barge -4.316e+05 3.255e+05 -1.326 0.184799  as.factor(category)Bass Boats 6.591e+04 4.804e+04 1.372 0.170068  as.factor(category)Bass Boats, Downeast 1.826e+05 4.703e+05 0.388 0.697771  as.factor(category)Bay Boats -4.679e+04 5.439e+04 -0.860 0.389607  as.factor(category)Bay Boats, Center Consoles -1.315e+04 6.426e+04 -0.205 0.837838  as.factor(category)Bay Boats, Saltwater Fishing -3.488e+04 1.301e+05 -0.268 0.788530  as.factor(category)Bay Boats, Saltwater Fishing, Sports Fishing Boats -5.141e+05 4.708e+05 -1.092 0.274802  as.factor(category)Bowrider -1.935e+04 4.473e+04 -0.433 0.665332  as.factor(category)Bowrider, Bay Boats 1.664e+05 3.169e+05 0.525 0.599592  as.factor(category)Bowrider, Bowrider 2.998e+04 3.172e+05 0.095 0.924706  as.factor(category)Bowrider, Bowrider, Runabouts -4.614e+04 1.759e+05 -0.262 0.793082  as.factor(category)Bowrider, Center Consoles 6.040e+04 2.761e+05 0.219 0.826863  as.factor(category)Bowrider, Cruisers -1.873e+04 1.334e+05 -0.140 0.888368  as.factor(category)Bowrider, Cruisers, Sports Cruiser 4.025e+04 3.181e+05 0.127 0.899306  as.factor(category)Bowrider, Cuddy Cabin -1.744e+05 1.055e+05 -1.654 0.098219  as.factor(category)Bowrider, Cuddy Cabin, Dual Console -4.693e+04 4.451e+05 -0.105 0.916022  as.factor(category)Bowrider, Deck Boats -6.796e+03 9.241e+04 -0.074 0.941376  as.factor(category)Bowrider, Deck Boats, Bowrider -1.105e+05 4.801e+05 -0.230 0.817989  as.factor(category)Bowrider, Deck Boats, Runabouts 1.070e+05 1.182e+05 0.905 0.365590  as.factor(category)Bowrider, Dual Console -4.110e+04 1.200e+05 -0.342 0.732064  as.factor(category)Bowrider, Dual Console, Cruisers -2.892e+05 4.541e+05 -0.637 0.524201  as.factor(category)Bowrider, Dual Console, Cuddy Cabin -4.737e+04 4.452e+05 -0.106 0.915259  as.factor(category)Bowrider, Dual Console, Runabouts 4.063e+04 4.453e+05 0.091 0.927291  as.factor(category)Bowrider, Express Cruiser, Cuddy Cabin 2.580e+05 4.455e+05 0.579 0.562551  as.factor(category)Bowrider, High Performance Boats, Sports Cruiser -1.851e+04 4.453e+05 -0.042 0.966836  as.factor(category)Bowrider, Jet Boats -4.097e+05 1.509e+05 -2.716 0.006614  as.factor(category)Bowrider, Jet Boats, Runabouts -4.769e+04 4.627e+05 -0.103 0.917917  as.factor(category)Bowrider, Other -4.858e+04 3.426e+05 -0.142 0.887248  as.factor(category)Bowrider, Runabouts -1.059e+04 5.543e+04 -0.191 0.848546  as.factor(category)Bowrider, Runabouts, Bowrider 3.267e+03 1.877e+05 0.017 0.986112  as.factor(category)Bowrider, Runabouts, Cruisers 3.251e+04 2.605e+05 0.125 0.900671  as.factor(category)Bowrider, Runabouts, Jet Boats -3.925e+04 4.457e+05 -0.088 0.929833  as.factor(category)Bowrider, Runabouts, Ski and Fish 9.136e+03 1.598e+05 0.057 0.954407  as.factor(category)Bowrider, Runabouts, Ski and Wakeboard Boats 9.461e+04 1.551e+05 0.610 0.541970  as.factor(category)Bowrider, Saltwater Fishing -6.029e+04 3.178e+05 -0.190 0.849554  as.factor(category)Bowrider, Saltwater Fishing, Sports Fishing Boats -2.569e+05 4.457e+05 -0.576 0.564368  as.factor(category)Bowrider, Ski and Fish, Runabouts -4.306e+03 1.679e+05 -0.026 0.979541  as.factor(category)Bowrider, Ski and Wakeboard Boats 1.062e+04 9.103e+04 0.117 0.907100  as.factor(category)Bowrider, Ski and Wakeboard Boats, Runabouts 3.597e+04 1.220e+05 0.295 0.768104  as.factor(category)Bowrider, Sports Cruiser -8.621e+04 2.039e+05 -0.423 0.672473  as.factor(category)Bowrider, Sports Cruiser, Cruisers 5.665e+04 3.202e+05 0.177 0.859587  as.factor(category)Catamaran 5.167e+04 8.733e+04 0.592 0.554121  as.factor(category)Catamaran, Catamaran -1.368e+05 4.459e+05 -0.307 0.758933  as.factor(category)Catamaran, Commercial Boats -2.027e+05 4.460e+05 -0.455 0.649429  as.factor(category)Catamaran, Cruisers -3.546e+04 1.861e+05 -0.191 0.848862  as.factor(category)Catamaran, Multi-Hulls -6.880e+04 1.066e+05 -0.645 0.518665  as.factor(category)Catamaran, Multi-Hulls, Catamaran -2.147e+05 2.605e+05 -0.824 0.409768  as.factor(category)Center Cockpit -1.938e+05 1.106e+05 -1.752 0.079773  as.factor(category)Center Cockpit, Cruisers -1.224e+05 1.791e+05 -0.683 0.494427  as.factor(category)Center Cockpit, Cruisers, Sloop -4.047e+04 2.620e+05 -0.154 0.877234  as.factor(category)Center Cockpit, Cutter 9.168e+04 3.174e+05 0.289 0.772728  as.factor(category)Center Cockpit, Ketch -9.162e+04 2.612e+05 -0.351 0.725754  as.factor(category)Center Cockpit, Racers and Cruisers 9.570e+04 3.210e+05 0.298 0.765601  as.factor(category)Center Cockpit, River Cruiser -7.287e+05 6.275e+05 -1.161 0.245498  as.factor(category)Center Cockpit, Sloop -1.153e+05 2.283e+05 -0.505 0.613707  as.factor(category)Center Consoles -2.595e+04 4.465e+04 -0.581 0.561131  as.factor(category)Center Consoles, Aluminum Fish Boats 3.190e+04 2.372e+05 0.134 0.893033  as.factor(category)Center Consoles, Antique and Classics, Saltwater Fishing 3.410e+05 4.528e+05 0.753 0.451454  as.factor(category)Center Consoles, Bay Boats -3.583e+04 6.627e+04 -0.541 0.588694  as.factor(category)Center Consoles, Bay Boats, Saltwater Fishing -8.474e+04 2.631e+05 -0.322 0.747386  as.factor(category)Center Consoles, Bay Boats, Ski and Fish 5.472e+04 1.463e+05 0.374 0.708382  as.factor(category)Center Consoles, Bowrider, Saltwater Fishing -8.356e+04 2.651e+05 -0.315 0.752600  as.factor(category)Center Consoles, Center Consoles -7.868e+04 3.166e+05 -0.248 0.803753  as.factor(category)Center Consoles, Center Consoles, Saltwater Fishing -7.327e+04 2.034e+05 -0.360 0.718626  as.factor(category)Center Consoles, Commercial Boats 2.107e+05 3.218e+05 0.655 0.512624  as.factor(category)Center Consoles, Cuddy Cabin -4.425e+04 9.117e+04 -0.485 0.627383  as.factor(category)Center Consoles, Cuddy Cabin, High Performance Boats 1.987e+05 4.472e+05 0.444 0.656846  as.factor(category)Center Consoles, Cuddy Cabin, Sports Fishing Boats -1.410e+05 2.073e+05 -0.680 0.496260  as.factor(category)Center Consoles, Deck Boats 7.352e+04 2.594e+05 0.283 0.776880  as.factor(category)Center Consoles, Dive Boat -1.071e+05 3.232e+05 -0.331 0.740331    (Intercept) \*  age  length \*\*\*  as.factor(region)Gulf Coast \*\*\*  as.factor(region)NorthEast  as.factor(region)Pacific \*\*\*  as.factor(region)SorthEast \*  as.factor(make)Single Inboard  as.factor(make)Single Outboard  as.factor(make)Triple Outboard \*\*\*  as.factor(make)Twin Inboard  as.factor(make)Twin Outboard \*  as.factor(material)Composite \*\*\*  as.factor(material)Ferro cement  as.factor(material)Fiberglass  as.factor(material)Hypalon  as.factor(material)Other  as.factor(material)Pvc  as.factor(material)Steel  as.factor(material)Wood \*\*\*  as.factor(year)1912  as.factor(year)1922  as.factor(year)1926 \*\*  as.factor(year)1928 \*\*  as.factor(year)1929  as.factor(year)1930  as.factor(year)1931  as.factor(year)1932  as.factor(year)1933  as.factor(year)1935 \*\*  as.factor(year)1936 \*\*  as.factor(year)1937  as.factor(year)1938  as.factor(year)1941  as.factor(year)1944 \*  as.factor(year)1945  as.factor(year)1946  as.factor(year)1947  as.factor(year)1948  as.factor(year)1949  as.factor(year)1950  as.factor(year)1951  as.factor(year)1953 \*\*  as.factor(year)1954  as.factor(year)1955  as.factor(year)1956 \*\*  as.factor(year)1957  as.factor(year)1958  as.factor(year)1959  as.factor(year)1960  as.factor(year)1961  as.factor(year)1962  as.factor(year)1963  as.factor(year)1964  as.factor(year)1965  as.factor(year)1966  as.factor(year)1967  as.factor(year)1968 \*  as.factor(year)1969  as.factor(year)1970  as.factor(year)1971  as.factor(year)1972  as.factor(year)1973  as.factor(year)1974  as.factor(year)1975  as.factor(year)1976  as.factor(year)1977 .  as.factor(year)1978  as.factor(year)1979  as.factor(year)1980  as.factor(year)1981  as.factor(year)1982  as.factor(year)1983  as.factor(year)1984  as.factor(year)1985 \*  as.factor(year)1986 .  as.factor(year)1987 \*  as.factor(year)1988 \*  as.factor(year)1989 \*  as.factor(year)1990 \*  as.factor(year)1991 \*  as.factor(year)1992 \*\*  as.factor(year)1993 \*  as.factor(year)1994 \*  as.factor(year)1995 \*  as.factor(year)1996 \*\*  as.factor(year)1997 \*\*  as.factor(year)1998 \*\*  as.factor(year)1999 \*\*  as.factor(year)2000 \*\*  as.factor(year)2001 \*\*\*  as.factor(year)2002 \*\*  as.factor(year)2003 \*\*  as.factor(year)2004 \*\*  as.factor(year)2005 \*\*\*  as.factor(year)2006 \*\*\*  as.factor(year)2007 \*\*\*  as.factor(year)2008 \*\*  as.factor(year)2009 \*\*  as.factor(year)2010  as.factor(year)2011 \*\*  as.factor(year)2012 \*\*\*  as.factor(year)2013  as.factor(year)2014  as.factor(year)2015  as.factor(year)2016  as.factor(year)2017 \*\*\*  as.factor(year)2018 \*\*  as.factor(year)2019  as.factor(year)2020  as.factor(fuel)Diesel \*\*\*  as.factor(fuel)Electric  as.factor(fuel)Gas  as.factor(fuel)Other  as.factor(fuel)Propane  as.factor(category)Aft Cabin, Cruisers  as.factor(category)Aft Cabin, Cruisers, Motor Yachts  as.factor(category)Aft Cabin, Express Cruiser  as.factor(category)Aft Cabin, Flybridge, Motor Yachts  as.factor(category)Aft Cabin, Motor Yachts  as.factor(category)Aft Cabin, Motor Yachts, Express Cruiser  as.factor(category)Aft Cabin, Motor Yachts, Flybridge  as.factor(category)Aft Cabin, Motor Yachts, Trawlers  as.factor(category)Aft Cabin, Pilothouse, Flybridge  as.factor(category)Aft Cabin, Sports Fishing Boats, Trawlers  as.factor(category)Aft Cabin, Trawlers, Motor Yachts  as.factor(category)Aluminum Fish Boats  as.factor(category)Antique and Classics  as.factor(category)Antique and Classics, Runabouts  as.factor(category)Antique and Classics, Sloop  as.factor(category)Barge  as.factor(category)Bass Boats  as.factor(category)Bass Boats, Downeast  as.factor(category)Bay Boats  as.factor(category)Bay Boats, Center Consoles  as.factor(category)Bay Boats, Saltwater Fishing  as.factor(category)Bay Boats, Saltwater Fishing, Sports Fishing Boats  as.factor(category)Bowrider  as.factor(category)Bowrider, Bay Boats  as.factor(category)Bowrider, Bowrider  as.factor(category)Bowrider, Bowrider, Runabouts  as.factor(category)Bowrider, Center Consoles  as.factor(category)Bowrider, Cruisers  as.factor(category)Bowrider, Cruisers, Sports Cruiser  as.factor(category)Bowrider, Cuddy Cabin .  as.factor(category)Bowrider, Cuddy Cabin, Dual Console  as.factor(category)Bowrider, Deck Boats  as.factor(category)Bowrider, Deck Boats, Bowrider  as.factor(category)Bowrider, Deck Boats, Runabouts  as.factor(category)Bowrider, Dual Console  as.factor(category)Bowrider, Dual Console, Cruisers  as.factor(category)Bowrider, Dual Console, Cuddy Cabin  as.factor(category)Bowrider, Dual Console, Runabouts  as.factor(category)Bowrider, Express Cruiser, Cuddy Cabin  as.factor(category)Bowrider, High Performance Boats, Sports Cruiser  as.factor(category)Bowrider, Jet Boats \*\*  as.factor(category)Bowrider, Jet Boats, Runabouts  as.factor(category)Bowrider, Other  as.factor(category)Bowrider, Runabouts  as.factor(category)Bowrider, Runabouts, Bowrider  as.factor(category)Bowrider, Runabouts, Cruisers  as.factor(category)Bowrider, Runabouts, Jet Boats  as.factor(category)Bowrider, Runabouts, Ski and Fish  as.factor(category)Bowrider, Runabouts, Ski and Wakeboard Boats  as.factor(category)Bowrider, Saltwater Fishing  as.factor(category)Bowrider, Saltwater Fishing, Sports Fishing Boats  as.factor(category)Bowrider, Ski and Fish, Runabouts  as.factor(category)Bowrider, Ski and Wakeboard Boats  as.factor(category)Bowrider, Ski and Wakeboard Boats, Runabouts  as.factor(category)Bowrider, Sports Cruiser  as.factor(category)Bowrider, Sports Cruiser, Cruisers  as.factor(category)Catamaran  as.factor(category)Catamaran, Catamaran  as.factor(category)Catamaran, Commercial Boats  as.factor(category)Catamaran, Cruisers  as.factor(category)Catamaran, Multi-Hulls  as.factor(category)Catamaran, Multi-Hulls, Catamaran  as.factor(category)Center Cockpit .  as.factor(category)Center Cockpit, Cruisers  as.factor(category)Center Cockpit, Cruisers, Sloop  as.factor(category)Center Cockpit, Cutter  as.factor(category)Center Cockpit, Ketch  as.factor(category)Center Cockpit, Racers and Cruisers  as.factor(category)Center Cockpit, River Cruiser  as.factor(category)Center Cockpit, Sloop  as.factor(category)Center Consoles  as.factor(category)Center Consoles, Aluminum Fish Boats  as.factor(category)Center Consoles, Antique and Classics, Saltwater Fishing  as.factor(category)Center Consoles, Bay Boats  as.factor(category)Center Consoles, Bay Boats, Saltwater Fishing  as.factor(category)Center Consoles, Bay Boats, Ski and Fish  as.factor(category)Center Consoles, Bowrider, Saltwater Fishing  as.factor(category)Center Consoles, Center Consoles  as.factor(category)Center Consoles, Center Consoles, Saltwater Fishing  as.factor(category)Center Consoles, Commercial Boats  as.factor(category)Center Consoles, Cuddy Cabin  as.factor(category)Center Consoles, Cuddy Cabin, High Performance Boats  as.factor(category)Center Consoles, Cuddy Cabin, Sports Fishing Boats  as.factor(category)Center Consoles, Deck Boats  as.factor(category)Center Consoles, Dive Boat  [ reached getOption("max.print") -- omitted 2609 rows ]  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 441300 on 24371 degrees of freedom  Multiple R-squared: 0.5896, Adjusted R-squared: 0.5425  F-statistic: 12.52 on 2797 and 24371 DF, p-value: < 2.2e-16 |
|  |
| |  | | --- | |  | |

II. Result of General Additive Model taking into consideration all continuous as well as categorical variables

Family: gaussian

Link function: identity

Formula:

completed\_data$price ~ completed\_data$age + completed\_data$length +

as.factor(completed\_data$region) + as.factor(completed\_data$make) +

as.factor(completed\_data$material) + as.factor(completed\_data$year) +

as.factor(completed\_data$fuel) +

as.factor(completed\_data$zipcode) + as.factor(completed\_data$category)

Parametric coefficients:

Estimate Std. Error

(Intercept) -1.229e+06 4.799e+05

completed\_data$age -1.979e+03 6.496e+03

completed\_data$length 3.520e+04 3.520e+02

as.factor(completed\_data$region)Gulf Coast 2.189e+05 3.725e+04

as.factor(completed\_data$region)NorthEast -3.190e+04 3.172e+04

as.factor(completed\_data$region)Pacific -1.049e+05 2.972e+04

as.factor(completed\_data$region)SorthEast 5.814e+04 2.906e+04

as.factor(completed\_data$make)Single Inboard 7.399e+03 1.554e+04

as.factor(completed\_data$make)Single Outboard -1.788e+04 1.265e+04

as.factor(completed\_data$make)Triple Outboard -1.759e+05 3.418e+04

as.factor(completed\_data$make)Twin Inboard -1.337e+04 1.159e+04

as.factor(completed\_data$make)Twin Outboard -5.189e+04 2.016e+04

as.factor(completed\_data$material)Composite 2.006e+05 3.046e+04

as.factor(completed\_data$material)Ferro cement -4.376e+04 1.607e+05

as.factor(completed\_data$material)Fiberglass 2.524e+04 1.652e+04

as.factor(completed\_data$material)Hypalon 2.623e+04 1.890e+05

as.factor(completed\_data$material)Other 1.405e+04 2.014e+04

as.factor(completed\_data$material)Pvc 2.366e+05 4.482e+05

as.factor(completed\_data$material)Steel -2.641e+04 5.201e+04

as.factor(completed\_data$material)Wood 2.143e+05 5.428e+04

as.factor(completed\_data$year)1912 2.491e+05 7.344e+05

as.factor(completed\_data$year)1922 4.045e+04 6.157e+05

as.factor(completed\_data$year)1926 -1.921e+06 6.132e+05

as.factor(completed\_data$year)1928 -2.042e+06 6.396e+05

as.factor(completed\_data$year)1929 -6.925e+05 6.372e+05

as.factor(completed\_data$year)1930 2.622e+05 5.791e+05

as.factor(completed\_data$year)1931 -6.595e+05 5.787e+05

as.factor(completed\_data$year)1932 7.861e+04 6.506e+05

as.factor(completed\_data$year)1933 3.414e+05 2.862e+05

as.factor(completed\_data$year)1935 -1.626e+06 5.633e+05

as.factor(completed\_data$year)1936 -2.359e+06 7.244e+05

as.factor(completed\_data$year)1937 -6.861e+05 5.537e+05

as.factor(completed\_data$year)1938 6.290e+05 4.806e+05

as.factor(completed\_data$year)1941 6.448e+04 2.695e+05

as.factor(completed\_data$year)1944 -1.018e+06 4.355e+05

as.factor(completed\_data$year)1945 -8.660e+03 5.283e+05

as.factor(completed\_data$year)1946 6.978e+04 4.491e+05

as.factor(completed\_data$year)1947 1.665e+05 4.090e+05

as.factor(completed\_data$year)1948 -2.007e+05 5.306e+05

as.factor(completed\_data$year)1949 1.019e+05 2.655e+05

as.factor(completed\_data$year)1950 -1.372e+05 5.104e+05

as.factor(completed\_data$year)1951 -1.573e+05 5.124e+05

as.factor(completed\_data$year)1953 -1.138e+06 3.660e+05

as.factor(completed\_data$year)1954 -7.080e+05 7.235e+05

as.factor(completed\_data$year)1955 1.615e+05 3.837e+05

as.factor(completed\_data$year)1956 -8.857e+05 3.139e+05

as.factor(completed\_data$year)1957 -3.704e+05 3.441e+05

as.factor(completed\_data$year)1958 -3.120e+05 3.255e+05

as.factor(completed\_data$year)1959 -1.125e+05 3.357e+05

as.factor(completed\_data$year)1960 -1.340e+04 2.393e+05

as.factor(completed\_data$year)1961 -3.319e+05 2.397e+05

as.factor(completed\_data$year)1962 -9.138e+04 2.371e+05

as.factor(completed\_data$year)1963 -1.553e+05 2.258e+05

as.factor(completed\_data$year)1964 -3.610e+04 2.034e+05

as.factor(completed\_data$year)1965 -2.757e+05 2.122e+05

as.factor(completed\_data$year)1966 -1.164e+05 1.692e+05

as.factor(completed\_data$year)1967 -7.697e+04 1.726e+05

as.factor(completed\_data$year)1968 -3.716e+05 1.677e+05

as.factor(completed\_data$year)1969 4.460e+04 1.462e+05

as.factor(completed\_data$year)1970 1.581e+04 1.532e+05

as.factor(completed\_data$year)1971 -5.538e+04 1.541e+05

as.factor(completed\_data$year)1972 4.341e+04 1.235e+05

as.factor(completed\_data$year)1973 -1.196e+05 1.189e+05

as.factor(completed\_data$year)1974 -6.948e+04 1.182e+05

as.factor(completed\_data$year)1975 -1.116e+04 1.086e+05

as.factor(completed\_data$year)1976 -5.765e+04 1.005e+05

as.factor(completed\_data$year)1977 -1.465e+05 8.614e+04

as.factor(completed\_data$year)1978 -3.666e+04 8.080e+04

as.factor(completed\_data$year)1979 -6.121e+04 7.420e+04

as.factor(completed\_data$year)1980 -6.686e+04 7.286e+04

as.factor(completed\_data$year)1981 -5.451e+04 6.541e+04

as.factor(completed\_data$year)1982 -4.021e+04 6.178e+04

as.factor(completed\_data$year)1983 -3.866e+04 5.196e+04

as.factor(completed\_data$year)1984 -2.372e+04 4.500e+04

as.factor(completed\_data$year)1985 -8.241e+04 3.987e+04

as.factor(completed\_data$year)1986 -2.807e+04 3.423e+04

as.factor(completed\_data$year)1987 -6.718e+04 3.232e+04

as.factor(completed\_data$year)1988 -6.170e+04 3.095e+04

as.factor(completed\_data$year)1989 -9.013e+04 3.332e+04

as.factor(completed\_data$year)1990 -7.850e+04 3.818e+04

as.factor(completed\_data$year)1991 -6.593e+04 4.889e+04

as.factor(completed\_data$year)1992 -1.224e+05 5.104e+04

as.factor(completed\_data$year)1993 -1.156e+04 5.390e+04

as.factor(completed\_data$year)1994 -1.613e+04 5.564e+04

as.factor(completed\_data$year)1995 -1.685e+04 5.907e+04

as.factor(completed\_data$year)1996 -1.031e+04 6.377e+04

as.factor(completed\_data$year)1997 -2.636e+04 6.920e+04

as.factor(completed\_data$year)1998 -2.654e+04 7.448e+04

as.factor(completed\_data$year)1999 -1.510e+04 7.985e+04

as.factor(completed\_data$year)2000 1.116e+04 8.566e+04

as.factor(completed\_data$year)2001 -4.255e+02 9.183e+04

as.factor(completed\_data$year)2002 4.652e+04 9.855e+04

as.factor(completed\_data$year)2003 8.874e+04 1.044e+05

as.factor(completed\_data$year)2004 7.771e+04 1.106e+05

as.factor(completed\_data$year)2005 7.061e+04 1.167e+05

as.factor(completed\_data$year)2006 8.555e+04 1.229e+05

as.factor(completed\_data$year)2007 9.101e+04 1.295e+05

as.factor(completed\_data$year)2008 1.574e+05 1.360e+05

as.factor(completed\_data$year)2009 1.506e+05 1.433e+05

as.factor(completed\_data$year)2010 2.912e+05 1.501e+05

as.factor(completed\_data$year)2011 1.961e+05 1.559e+05

as.factor(completed\_data$year)2012 1.786e+05 1.621e+05

as.factor(completed\_data$year)2013 2.714e+05 1.682e+05

as.factor(completed\_data$year)2014 2.764e+05 1.746e+05

as.factor(completed\_data$year)2015 2.923e+05 1.810e+05

as.factor(completed\_data$year)2016 3.514e+05 1.874e+05

as.factor(completed\_data$year)2017 3.988e+05 1.937e+05

as.factor(completed\_data$year)2018 3.884e+05 2.000e+05

as.factor(completed\_data$year)2019 3.262e+05 2.063e+05

as.factor(completed\_data$year)2020 3.343e+05 2.000e+05

as.factor(completed\_data$fuel)Diesel -9.068e+04 1.646e+04

as.factor(completed\_data$fuel)Electric 8.705e+04 1.208e+05

as.factor(completed\_data$fuel)Gas -6.744e+03 1.207e+04

as.factor(completed\_data$fuel)Other -2.726e+04 1.950e+04

as.factor(completed\_data$fuel)Propane 1.931e+05 4.943e+05

as.factor(completed\_data$zipcode)02035 1.359e+05 6.256e+05

as.factor(completed\_data$zipcode)02043 2.697e+05 6.300e+05

as.factor(completed\_data$zipcode)02050 3.648e+04 5.466e+05

as.factor(completed\_data$zipcode)02171 -1.214e+03 3.150e+05

as.factor(completed\_data$zipcode)02191 -1.140e+05 5.153e+05

as.factor(completed\_data$zipcode)02540 1.541e+05 6.316e+05

as.factor(completed\_data$zipcode)02649 -2.981e+04 6.258e+05

as.factor(completed\_data$zipcode)02655 1.097e+05 7.666e+05

as.factor(completed\_data$zipcode)02748 5.879e+04 6.258e+05

as.factor(completed\_data$zipcode)02840 -2.395e+05 4.661e+05

as.factor(completed\_data$zipcode)02896 -2.291e+05 4.796e+05

as.factor(completed\_data$zipcode)03840 -4.137e+04 4.990e+05

as.factor(completed\_data$zipcode)04101 9.109e+04 5.473e+05

as.factor(completed\_data$zipcode)06340 1.376e+05 6.371e+05

as.factor(completed\_data$zipcode)06410 7.528e+04 4.738e+05

as.factor(completed\_data$zipcode)06426 3.616e+05 6.872e+05

as.factor(completed\_data$zipcode)06475 2.147e+04 5.137e+05

as.factor(completed\_data$zipcode)06480 1.204e+05 6.292e+05

as.factor(completed\_data$zipcode)06615 -3.878e+04 6.273e+05

as.factor(completed\_data$zipcode)06854 4.341e+05 6.259e+05

as.factor(completed\_data$zipcode)07760 1.305e+05 5.460e+05

as.factor(completed\_data$zipcode)08204 -1.020e+04 4.824e+05

as.factor(completed\_data$zipcode)08215 -2.353e+05 5.145e+05

as.factor(completed\_data$zipcode)08224 -2.437e+04 5.417e+05

as.factor(completed\_data$zipcode)08244 -1.585e+05 6.278e+05

as.factor(completed\_data$zipcode)08611 -1.090e+05 6.253e+05

as.factor(completed\_data$zipcode)08721 1.321e+05 5.436e+05

as.factor(completed\_data$zipcode)08724 1.371e+04 4.555e+05

as.factor(completed\_data$zipcode)08731 -1.171e+04 6.326e+05

as.factor(completed\_data$zipcode)08742 5.379e+04 4.598e+05

as.factor(completed\_data$zipcode)08753 1.285e+05 5.344e+05

as.factor(completed\_data$zipcode)10011 -2.465e+05 6.256e+05

as.factor(completed\_data$zipcode)10017 2.211e+05 6.267e+05

as.factor(completed\_data$zipcode)10805 2.876e+05 4.791e+05

as.factor(completed\_data$zipcode)10980 2.256e+05 4.612e+05

as.factor(completed\_data$zipcode)11050 3.181e+05 5.019e+05

as.factor(completed\_data$zipcode)11520 1.596e+05 4.450e+05

as.factor(completed\_data$zipcode)11542 1.536e+05 4.526e+05

as.factor(completed\_data$zipcode)11558 1.606e+05 5.123e+05

as.factor(completed\_data$zipcode)11572 2.446e+05 4.566e+05

as.factor(completed\_data$zipcode)11701 1.086e+05 4.854e+05

as.factor(completed\_data$zipcode)11702 1.458e+05 4.556e+05

as.factor(completed\_data$zipcode)11726 -3.393e+04 6.253e+05

as.factor(completed\_data$zipcode)11733 3.595e+05 4.702e+05

as.factor(completed\_data$zipcode)11743 9.657e+04 4.492e+05

as.factor(completed\_data$zipcode)11752 1.623e+05 4.539e+05

as.factor(completed\_data$zipcode)11757 1.351e+05 4.450e+05

as.factor(completed\_data$zipcode)11768 9.331e+04 4.533e+05

as.factor(completed\_data$zipcode)11769 7.418e+04 4.496e+05

as.factor(completed\_data$zipcode)11772 1.240e+05 4.503e+05

as.factor(completed\_data$zipcode)11782 1.396e+05 4.575e+05

as.factor(completed\_data$zipcode)11783 1.936e+05 4.504e+05

as.factor(completed\_data$zipcode)11933 1.432e+05 5.128e+05

as.factor(completed\_data$zipcode)11940 3.110e+05 6.289e+05

as.factor(completed\_data$zipcode)11941 1.327e+05 4.798e+05

as.factor(completed\_data$zipcode)11944 1.430e+05 4.486e+05

as.factor(completed\_data$zipcode)11946 2.021e+05 4.527e+05

as.factor(completed\_data$zipcode)11952 1.288e+05 4.470e+05

as.factor(completed\_data$zipcode)11963 4.456e+04 4.514e+05

as.factor(completed\_data$zipcode)11971 1.807e+05 4.570e+05

as.factor(completed\_data$zipcode)11978 1.093e+05 4.543e+05

as.factor(completed\_data$zipcode)12192 1.989e+05 4.955e+05

as.factor(completed\_data$zipcode)12303 2.729e+05 4.703e+05

as.factor(completed\_data$zipcode)12553 2.584e+05 4.490e+05

as.factor(completed\_data$zipcode)12804 7.845e+04 6.261e+05

as.factor(completed\_data$zipcode)12845 1.753e+05 4.503e+05

as.factor(completed\_data$zipcode)12866 2.418e+05 5.113e+05

as.factor(completed\_data$zipcode)12883 1.940e+05 4.951e+05

as.factor(completed\_data$zipcode)12901 1.820e+05 4.578e+05

as.factor(completed\_data$zipcode)13029 3.476e+05 6.265e+05

as.factor(completed\_data$zipcode)13030 2.057e+05 4.704e+05

as.factor(completed\_data$zipcode)13502 1.620e+05 4.735e+05

as.factor(completed\_data$zipcode)13607 1.174e+05 4.784e+05

as.factor(completed\_data$zipcode)13624 2.440e+05 3.135e+05

as.factor(completed\_data$zipcode)13664 1.835e+05 5.420e+05

as.factor(completed\_data$zipcode)14047 2.875e+05 6.263e+05

as.factor(completed\_data$zipcode)14072 2.048e+05 4.518e+05

as.factor(completed\_data$zipcode)14120 1.031e+05 4.551e+05

as.factor(completed\_data$zipcode)14150 1.527e+05 4.568e+05

as.factor(completed\_data$zipcode)14174 2.453e+05 5.437e+05

as.factor(completed\_data$zipcode)14203 2.051e+05 4.541e+05

as.factor(completed\_data$zipcode)14207 1.863e+05 5.432e+05

as.factor(completed\_data$zipcode)14218 2.174e+05 4.583e+05

as.factor(completed\_data$zipcode)14225 1.732e+05 4.950e+05

as.factor(completed\_data$zipcode)14424 1.862e+05 4.851e+05

t value Pr(>|t|)

(Intercept) -2.561 0.010445 \*

completed\_data$age -0.305 0.760643

completed\_data$length 99.988 < 2e-16 \*\*\*

as.factor(completed\_data$region)Gulf Coast 5.877 4.24e-09 \*\*\*

as.factor(completed\_data$region)NorthEast -1.006 0.314638

as.factor(completed\_data$region)Pacific -3.529 0.000418 \*\*\*

as.factor(completed\_data$region)SorthEast 2.001 0.045427 \*

as.factor(completed\_data$make)Single Inboard 0.476 0.634063

as.factor(completed\_data$make)Single Outboard -1.413 0.157581

as.factor(completed\_data$make)Triple Outboard -5.145 2.70e-07 \*\*\*

as.factor(completed\_data$make)Twin Inboard -1.154 0.248706

as.factor(completed\_data$make)Twin Outboard -2.574 0.010073 \*

as.factor(completed\_data$material)Composite 6.585 4.64e-11 \*\*\*

as.factor(completed\_data$material)Ferro cement -0.272 0.785386

as.factor(completed\_data$material)Fiberglass 1.528 0.126553

as.factor(completed\_data$material)Hypalon 0.139 0.889607

as.factor(completed\_data$material)Other 0.697 0.485627

as.factor(completed\_data$material)Pvc 0.528 0.597673

as.factor(completed\_data$material)Steel -0.508 0.611562

as.factor(completed\_data$material)Wood 3.947 7.93e-05 \*\*\*

as.factor(completed\_data$year)1912 0.339 0.734500

as.factor(completed\_data$year)1922 0.066 0.947617

as.factor(completed\_data$year)1926 -3.132 0.001736 \*\*

as.factor(completed\_data$year)1928 -3.193 0.001409 \*\*

as.factor(completed\_data$year)1929 -1.087 0.277102

as.factor(completed\_data$year)1930 0.453 0.650761

as.factor(completed\_data$year)1931 -1.140 0.254397

as.factor(completed\_data$year)1932 0.121 0.903831

as.factor(completed\_data$year)1933 1.193 0.232961

as.factor(completed\_data$year)1935 -2.886 0.003905 \*\*

as.factor(completed\_data$year)1936 -3.257 0.001128 \*\*

as.factor(completed\_data$year)1937 -1.239 0.215281

as.factor(completed\_data$year)1938 1.309 0.190547

as.factor(completed\_data$year)1941 0.239 0.810923

as.factor(completed\_data$year)1944 -2.337 0.019432 \*

as.factor(completed\_data$year)1945 -0.016 0.986922

as.factor(completed\_data$year)1946 0.155 0.876515

as.factor(completed\_data$year)1947 0.407 0.683859

as.factor(completed\_data$year)1948 -0.378 0.705319

as.factor(completed\_data$year)1949 0.384 0.701194

as.factor(completed\_data$year)1950 -0.269 0.788062

as.factor(completed\_data$year)1951 -0.307 0.758947

as.factor(completed\_data$year)1953 -3.109 0.001881 \*\*

as.factor(completed\_data$year)1954 -0.979 0.327792

as.factor(completed\_data$year)1955 0.421 0.673772

as.factor(completed\_data$year)1956 -2.822 0.004782 \*\*

as.factor(completed\_data$year)1957 -1.076 0.281811

as.factor(completed\_data$year)1958 -0.959 0.337661

as.factor(completed\_data$year)1959 -0.335 0.737639

as.factor(completed\_data$year)1960 -0.056 0.955335

as.factor(completed\_data$year)1961 -1.385 0.166185

as.factor(completed\_data$year)1962 -0.385 0.699894

as.factor(completed\_data$year)1963 -0.688 0.491737

as.factor(completed\_data$year)1964 -0.177 0.859143

as.factor(completed\_data$year)1965 -1.299 0.193843

as.factor(completed\_data$year)1966 -0.688 0.491385

as.factor(completed\_data$year)1967 -0.446 0.655660

as.factor(completed\_data$year)1968 -2.216 0.026686 \*

as.factor(completed\_data$year)1969 0.305 0.760323

as.factor(completed\_data$year)1970 0.103 0.917795

as.factor(completed\_data$year)1971 -0.359 0.719347

as.factor(completed\_data$year)1972 0.351 0.725224

as.factor(completed\_data$year)1973 -1.006 0.314367

as.factor(completed\_data$year)1974 -0.588 0.556828

as.factor(completed\_data$year)1975 -0.103 0.918176

as.factor(completed\_data$year)1976 -0.574 0.566157

as.factor(completed\_data$year)1977 -1.701 0.089001 .

as.factor(completed\_data$year)1978 -0.454 0.650065

as.factor(completed\_data$year)1979 -0.825 0.409419

as.factor(completed\_data$year)1980 -0.918 0.358802

as.factor(completed\_data$year)1981 -0.833 0.404671

as.factor(completed\_data$year)1982 -0.651 0.515175

as.factor(completed\_data$year)1983 -0.744 0.456839

as.factor(completed\_data$year)1984 -0.527 0.598087

as.factor(completed\_data$year)1985 -2.067 0.038765 \*

as.factor(completed\_data$year)1986 -0.820 0.412107

as.factor(completed\_data$year)1987 -2.078 0.037691 \*

as.factor(completed\_data$year)1988 -1.993 0.046247 \*

as.factor(completed\_data$year)1989 -2.705 0.006829 \*\*

as.factor(completed\_data$year)1990 -2.056 0.039783 \*

as.factor(completed\_data$year)1991 -1.349 0.177464

as.factor(completed\_data$year)1992 -2.398 0.016493 \*

as.factor(completed\_data$year)1993 -0.214 0.830195

as.factor(completed\_data$year)1994 -0.290 0.771885

as.factor(completed\_data$year)1995 -0.285 0.775444

as.factor(completed\_data$year)1996 -0.162 0.871622

as.factor(completed\_data$year)1997 -0.381 0.703209

as.factor(completed\_data$year)1998 -0.356 0.721526

as.factor(completed\_data$year)1999 -0.189 0.849978

as.factor(completed\_data$year)2000 0.130 0.896338

as.factor(completed\_data$year)2001 -0.005 0.996303

as.factor(completed\_data$year)2002 0.472 0.636867

as.factor(completed\_data$year)2003 0.850 0.395365

as.factor(completed\_data$year)2004 0.703 0.482242

as.factor(completed\_data$year)2005 0.605 0.545054

as.factor(completed\_data$year)2006 0.696 0.486489

as.factor(completed\_data$year)2007 0.703 0.482238

as.factor(completed\_data$year)2008 1.158 0.247074

as.factor(completed\_data$year)2009 1.050 0.293522

as.factor(completed\_data$year)2010 1.940 0.052405 .

as.factor(completed\_data$year)2011 1.257 0.208589

as.factor(completed\_data$year)2012 1.102 0.270548

as.factor(completed\_data$year)2013 1.613 0.106797

as.factor(completed\_data$year)2014 1.583 0.113532

as.factor(completed\_data$year)2015 1.615 0.106291

as.factor(completed\_data$year)2016 1.875 0.060785 .

as.factor(completed\_data$year)2017 2.059 0.039553 \*

as.factor(completed\_data$year)2018 1.942 0.052168 .

as.factor(completed\_data$year)2019 1.581 0.113853

as.factor(completed\_data$year)2020 1.671 0.094717 .

as.factor(completed\_data$fuel)Diesel -5.510 3.62e-08 \*\*\*

as.factor(completed\_data$fuel)Electric 0.721 0.471153

as.factor(completed\_data$fuel)Gas -0.559 0.576337

as.factor(completed\_data$fuel)Other -1.398 0.162073

as.factor(completed\_data$fuel)Propane 0.391 0.696105

as.factor(completed\_data$zipcode)02035 0.217 0.827985

as.factor(completed\_data$zipcode)02043 0.428 0.668599

as.factor(completed\_data$zipcode)02050 0.067 0.946789

as.factor(completed\_data$zipcode)02171 -0.004 0.996924

as.factor(completed\_data$zipcode)02191 -0.221 0.824923

as.factor(completed\_data$zipcode)02540 0.244 0.807181

as.factor(completed\_data$zipcode)02649 -0.048 0.962011

as.factor(completed\_data$zipcode)02655 0.143 0.886181

as.factor(completed\_data$zipcode)02748 0.094 0.925149

as.factor(completed\_data$zipcode)02840 -0.514 0.607297

as.factor(completed\_data$zipcode)02896 -0.478 0.632890

as.factor(completed\_data$zipcode)03840 -0.083 0.933930

as.factor(completed\_data$zipcode)04101 0.166 0.867816

as.factor(completed\_data$zipcode)06340 0.216 0.829022

as.factor(completed\_data$zipcode)06410 0.159 0.873757

as.factor(completed\_data$zipcode)06426 0.526 0.598765

as.factor(completed\_data$zipcode)06475 0.042 0.966662

as.factor(completed\_data$zipcode)06480 0.191 0.848294

as.factor(completed\_data$zipcode)06615 -0.062 0.950710

as.factor(completed\_data$zipcode)06854 0.694 0.487935

as.factor(completed\_data$zipcode)07760 0.239 0.811070

as.factor(completed\_data$zipcode)08204 -0.021 0.983134

as.factor(completed\_data$zipcode)08215 -0.457 0.647420

as.factor(completed\_data$zipcode)08224 -0.045 0.964121

as.factor(completed\_data$zipcode)08244 -0.253 0.800637

as.factor(completed\_data$zipcode)08611 -0.174 0.861573

as.factor(completed\_data$zipcode)08721 0.243 0.807931

as.factor(completed\_data$zipcode)08724 0.030 0.975996

as.factor(completed\_data$zipcode)08731 -0.019 0.985229

as.factor(completed\_data$zipcode)08742 0.117 0.906881

as.factor(completed\_data$zipcode)08753 0.241 0.809930

as.factor(completed\_data$zipcode)10011 -0.394 0.693503

as.factor(completed\_data$zipcode)10017 0.353 0.724223

as.factor(completed\_data$zipcode)10805 0.600 0.548383

as.factor(completed\_data$zipcode)10980 0.489 0.624727

as.factor(completed\_data$zipcode)11050 0.634 0.526193

as.factor(completed\_data$zipcode)11520 0.359 0.719918

as.factor(completed\_data$zipcode)11542 0.339 0.734362

as.factor(completed\_data$zipcode)11558 0.313 0.753917

as.factor(completed\_data$zipcode)11572 0.536 0.592207

as.factor(completed\_data$zipcode)11701 0.224 0.822993

as.factor(completed\_data$zipcode)11702 0.320 0.748968

as.factor(completed\_data$zipcode)11726 -0.054 0.956724

as.factor(completed\_data$zipcode)11733 0.765 0.444571

as.factor(completed\_data$zipcode)11743 0.215 0.829782

as.factor(completed\_data$zipcode)11752 0.358 0.720617

as.factor(completed\_data$zipcode)11757 0.304 0.761408

as.factor(completed\_data$zipcode)11768 0.206 0.836924

as.factor(completed\_data$zipcode)11769 0.165 0.868938

as.factor(completed\_data$zipcode)11772 0.275 0.782971

as.factor(completed\_data$zipcode)11782 0.305 0.760182

as.factor(completed\_data$zipcode)11783 0.430 0.667315

as.factor(completed\_data$zipcode)11933 0.279 0.780044

as.factor(completed\_data$zipcode)11940 0.494 0.620986

as.factor(completed\_data$zipcode)11941 0.277 0.782075

as.factor(completed\_data$zipcode)11944 0.319 0.749930

as.factor(completed\_data$zipcode)11946 0.446 0.655320

as.factor(completed\_data$zipcode)11952 0.288 0.773271

as.factor(completed\_data$zipcode)11963 0.099 0.921365

as.factor(completed\_data$zipcode)11971 0.395 0.692506

as.factor(completed\_data$zipcode)11978 0.241 0.809913

as.factor(completed\_data$zipcode)12192 0.401 0.688189

as.factor(completed\_data$zipcode)12303 0.580 0.561757

as.factor(completed\_data$zipcode)12553 0.576 0.564934

as.factor(completed\_data$zipcode)12804 0.125 0.900284

as.factor(completed\_data$zipcode)12845 0.389 0.696966

as.factor(completed\_data$zipcode)12866 0.473 0.636288

as.factor(completed\_data$zipcode)12883 0.392 0.695138

as.factor(completed\_data$zipcode)12901 0.397 0.691017

as.factor(completed\_data$zipcode)13029 0.555 0.579066

as.factor(completed\_data$zipcode)13030 0.437 0.661922

as.factor(completed\_data$zipcode)13502 0.342 0.732187

as.factor(completed\_data$zipcode)13607 0.245 0.806142

as.factor(completed\_data$zipcode)13624 0.778 0.436337

as.factor(completed\_data$zipcode)13664 0.339 0.734934

as.factor(completed\_data$zipcode)14047 0.459 0.646166

as.factor(completed\_data$zipcode)14072 0.453 0.650406

as.factor(completed\_data$zipcode)14120 0.227 0.820717

as.factor(completed\_data$zipcode)14150 0.334 0.738090

as.factor(completed\_data$zipcode)14174 0.451 0.651847

as.factor(completed\_data$zipcode)14203 0.452 0.651537

as.factor(completed\_data$zipcode)14207 0.343 0.731588

as.factor(completed\_data$zipcode)14218 0.474 0.635310

as.factor(completed\_data$zipcode)14225 0.350 0.726458

as.factor(completed\_data$zipcode)14424 0.384 0.701067

[ reached getOption("max.print") -- omitted 2609 rows ]

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Rank: 2798/2809

R-sq.(adj) = 0.542 Deviance explained = 59%

GCV = 2.1708e+11 Scale est. = 1.9472e+11 n = 27169