```r
library(car)
```

```
## Loading required package: carData
```

```r
library(reshape2)
library(ggplot2)
library(MASS)
library(interactions)

source("clean_data.R")


df <- remove_cols(df, c("Color", "Model"))

# Remove columns with only one observation and affected rows
res <- convert_categorical(df, categorical)
design <- as.data.frame(res$dummy)

singles <- c()
bad_idx <- c()
for (col in colnames(design)) {
  if (sum(design[, col] != 0) <= 1) {
    singles <- c(singles, col)
    bad_idx <- c(bad_idx, which(design[, col] != 0))
  }
}
singles
```

```
##  [1] "MakeFiat"            "MakeLexus"           "Fuel.TypePetrol + CNG"
##  [4] "LocationDak. Kannada" "LocationFaizabad"    "LocationGorakhpur"
##  [7] "LocationPurnea"      "LocationRohtak"      "LocationRudrapur"
## [10] "LocationSamastipur"  "LocationValsad"
```
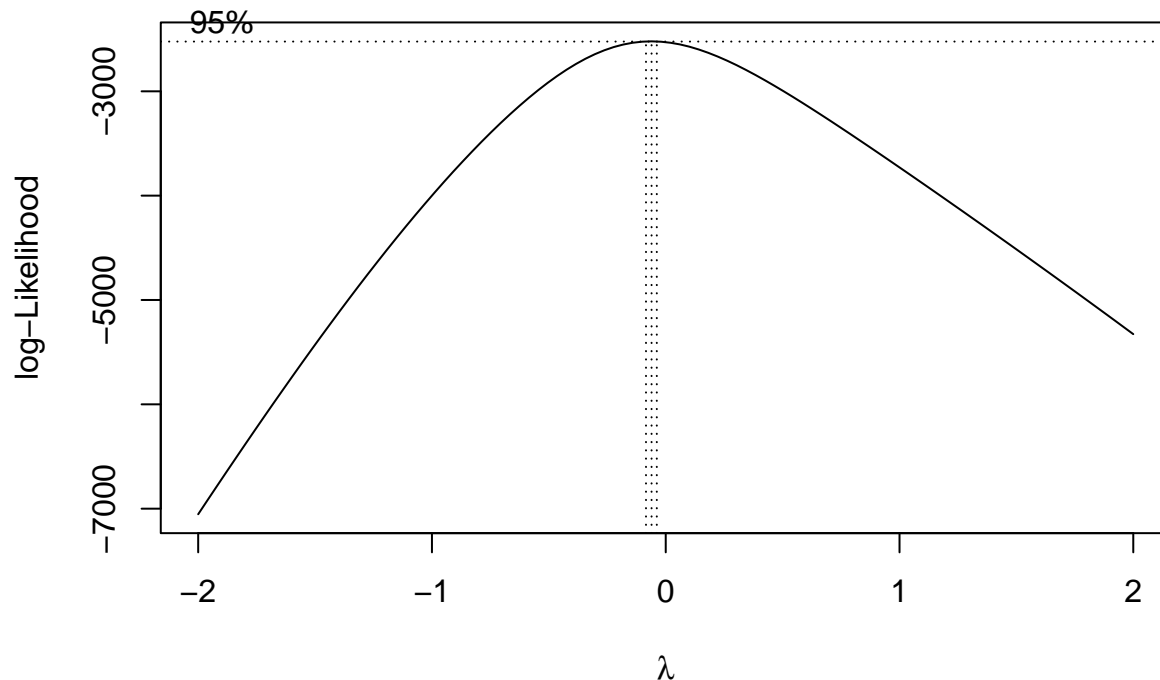
```r
bad_idx
```

```
##  [1]  662 1009 1169 1077  779  569  728  245  510  264  162
```

```r
df <- df[-bad_idx, ]
```

Box cox

```r
bc <- boxcox(Price ~., data = df)
```

1

```r
bc$x[which.max(bc$y)]
```

```
## [1] -0.06060606
```

For the sake of interpretability, we will use a log transformation, since $\lambda \approx 0$.

```r
x <- df
x$Price <- log(df$Price)
names(x)[names(x) == "Price"] <- "log(Price)"

colnames(x)
```
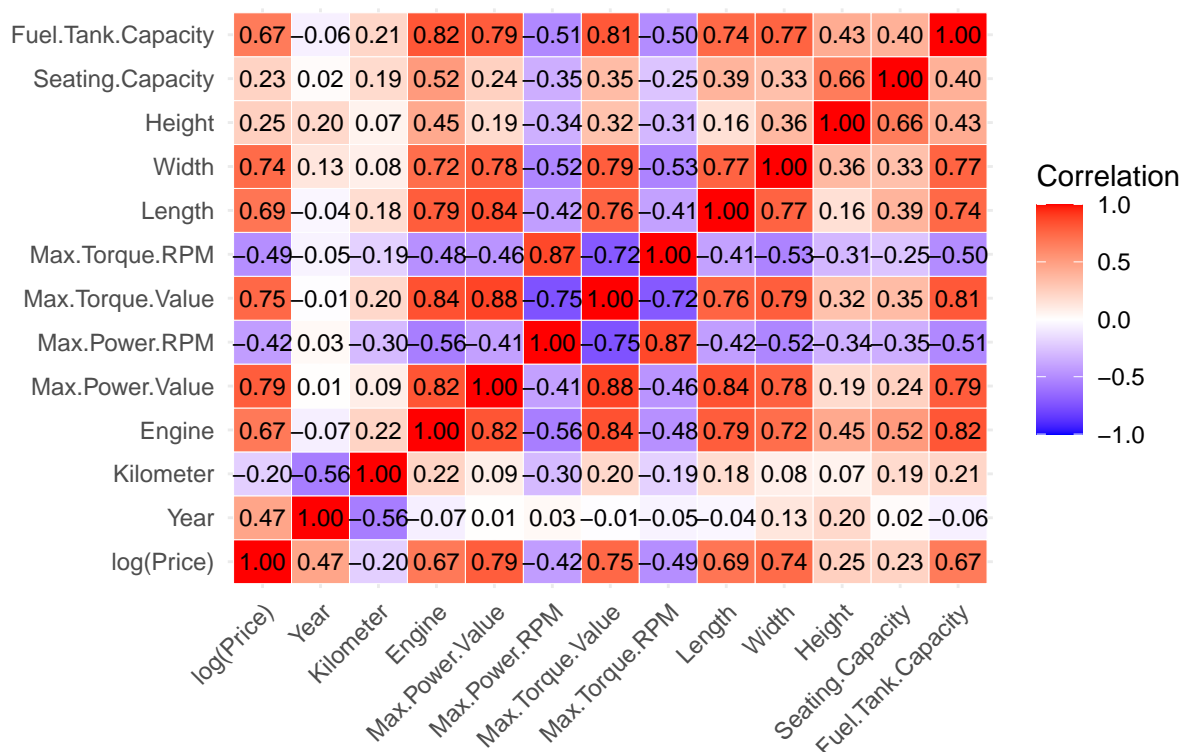
```
##  [1] "Make"              "log(Price)"        "Year"
##  [4] "Kilometer"         "Fuel.Type"         "Transmission"
##  [7] "Location"          "Owner"             "Seller.Type"
## [10] "Engine"            "Max.Power.Value"   "Max.Power.RPM"
## [13] "Max.Torque.Value"  "Max.Torque.RPM"    "Drivetrain"
## [16] "Length"            "Width"             "Height"
## [19] "Seating.Capacity"  "Fuel.Tank.Capacity"
```

Inspect correlation

```r
cor_matrix <- cor(x[, sapply(x, is.numeric)])
cor_melted <- melt(cor_matrix)
ggplot(data = cor_melted, aes(x = Var1, y = Var2, fill = value)) +
    geom_tile(color = "white") +
    scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                         midpoint = 0, limit = c(-1, 1), space = "Lab",
                         name="Correlation") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
    labs(x = "", y = "", title = "Correlation Matrix") +
    geom_text(aes(label = sprintf("%.2f", value)), size = 3)
```

Correlation Matrix

	log(Price)	Year	Kilometer	Engine	Max.Power.Value	Max.Power.RPM	Max.Torque.Value	Max.Torque.RPM	Length	Width	Height	Seating.Capacity	Fuel.Tank.Capacity
Fuel.Tank.Capacity	0.67	−0.06	0.21	0.82	0.79	−0.51	0.81	−0.50	0.74	0.77	0.43	0.40	1.00
Seating.Capacity	0.23	0.02	0.19	0.52	0.24	−0.35	0.35	−0.25	0.39	0.33	0.66	1.00	0.40
Height	0.25	0.20	0.07	0.45	0.19	−0.34	0.32	−0.31	0.16	0.36	1.00	0.66	0.43
Width	0.74	0.13	0.08	0.72	0.78	−0.52	0.79	−0.53	0.77	1.00	0.36	0.33	0.77
Length	0.69	−0.04	0.18	0.79	0.84	−0.42	0.76	−0.41	1.00	0.77	0.16	0.39	0.74
Max.Torque.RPM	−0.49	−0.05	−0.19	−0.48	−0.46	0.87	−0.72	1.00	−0.41	−0.53	−0.31	−0.25	−0.50
Max.Torque.Value	0.75	−0.01	0.20	0.84	0.88	−0.75	1.00	−0.72	0.76	0.79	0.32	0.35	0.81
Max.Power.RPM	−0.42	0.03	−0.30	−0.56	−0.41	1.00	−0.75	0.87	−0.42	−0.52	−0.34	−0.35	−0.51
Max.Power.Value	0.79	0.01	0.09	0.82	1.00	−0.41	0.88	−0.46	0.84	0.78	0.19	0.24	0.79
Engine	0.67	−0.07	0.22	1.00	0.82	−0.56	0.84	−0.48	0.79	0.72	0.45	0.52	0.82
Kilometer	−0.20	−0.56	1.00	0.22	0.09	−0.30	0.20	−0.19	0.18	0.08	0.07	0.19	0.21
Year	0.47	1.00	−0.56	−0.07	0.01	0.03	−0.01	−0.05	−0.04	0.13	0.20	0.02	−0.06
log(Price)	1.00	0.47	−0.20	0.67	0.79	−0.42	0.75	−0.49	0.69	0.74	0.25	0.23	0.67

Correlation

1.0
0.5
0.0
−0.5
−1.0

Trim Regressors Based on Intuition

Engine has high correlation with other regressors, and is highly related to more relevant statistics like torque and horsepower, so it is dropped.

We need really only one dimension of the car. Height and length are not as correlated to price as width, and width gives more information about engine capacity,

All the torque and horsepower regressors are highly related. A typical car driver will prioritize maximum horsepower over the others, so only Max.Power.Value was kept.

Fuel tank capacity information seems to be included in many other stats due its high correlation with many other regressors, so it was dropped.

We do not think that color will be a useful predictor, and there are too many models so these categories are removed.

```r
x_trim <- remove_cols(x, c("Engine", "Length", "Max.Torque.RPM", "Max.Torque.Value", "Max.Power.RPM", "
model <- lm(`log(Price)` ~., data = x_trim)
colnames(x_trim)
```

```
## [1] "Make"            "log(Price)"       "Year"            "Kilometer"
## [5] "Fuel.Type"       "Transmission"     "Location"        "Owner"
## [9] "Seller.Type"     "Max.Power.Value"  "Drivetrain"      "Width"
## [13] "Seating.Capacity"
```

Attempt to add interactions

We consider only numerical interactions due to data sparsity; including categorical values produces NA's for most interactions. We do not expect kilometers driven to interact with max power or width, so those interactions are not considered.

```
interactions <- lm(
  `log(Price)` ~ . + Year:Kilometer + Year:Max.Power.Value
                 + Year:Width + Max.Power.Value:Width,
  data = x_trim
)

anova(interactions)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                        Df Sum Sq Mean Sq   F value     Pr(>F)
## Make                   25 354.44  14.178  524.3350 < 2.2e-16 ***
## Year                    1 178.52 178.525 6602.4314 < 2.2e-16 ***
## Kilometer               1   0.98   0.982   36.2998 2.185e-09 ***
## Fuel.Type               2  19.23   9.615  355.6012 < 2.2e-16 ***
## Transmission            1  20.31  20.313  751.2240 < 2.2e-16 ***
## Location               63  16.76   0.266    9.8370 < 2.2e-16 ***
## Owner                   3   0.46   0.153    5.6705 0.0007375 ***
## Seller.Type             2   0.33   0.165    6.1157 0.0022703 **
## Max.Power.Value         1  60.04  60.045 2220.6537 < 2.2e-16 ***
## Drivetrain              2   0.01   0.007    0.2758 0.7589795
## Width                   1   6.16   6.161  227.8650 < 2.2e-16 ***
## Seating.Capacity        1   0.96   0.959   35.4778 3.296e-09 ***
## Year:Kilometer          1   1.11   1.105   40.8782 2.236e-10 ***
## Year:Max.Power.Value    1   2.44   2.445   90.4110 < 2.2e-16 ***
## Year:Width              1   0.09   0.092    3.3988 0.0654663 .
## Max.Power.Value:Width   1   1.01   1.008   37.2728 1.345e-09 ***
## Residuals            1333  36.04   0.027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remove drivetrain and year x width interaction as they are not significant. All the others are significant even
with Bonferroni adjustment.

```
x_final <- remove_cols(x_trim, c("Drivetrain"))

final <- lm(
  `log(Price)` ~ . + Year:Kilometer + Year:Max.Power.Value + Max.Power.Value:Width,
  data = x_final
)

anova(final)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                 Df Sum Sq Mean Sq   F value     Pr(>F)
## Make            25 354.44  14.178  517.3215 < 2.2e-16 ***
## Year             1 178.52 178.525 6514.1167 < 2.2e-16 ***
## Kilometer        1   0.98   0.982   35.8143 2.784e-09 ***
## Fuel.Type        2  19.23   9.615  350.8447 < 2.2e-16 ***
## Transmission     1  20.31  20.313  741.1756 < 2.2e-16 ***
## Location        63  16.76   0.266    9.7055 < 2.2e-16 ***
## Owner            3   0.46   0.153    5.5946 0.0008202 ***
```

```
## Seller.Type                2    0.33    0.165      6.0339 0.0024618 **
## Max.Power.Value             1   60.04   60.045   2190.9500 < 2.2e-16 ***
## Width                       1    6.08    6.076    221.7070 < 2.2e-16 ***
## Seating.Capacity            1    0.81    0.805     29.3898 7.016e-08 ***
## Year:Kilometer              1    1.01    1.009     36.8257 1.680e-09 ***
## Year:Max.Power.Value        1    2.43    2.433     88.7645 < 2.2e-16 ***
## Max.Power.Value:Width       1    0.89    0.891     32.5080 1.459e-08 ***
## Residuals                1336   36.61    0.027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inspect Multicollinearity and Leverage

```
vif(final, type = "predictor")
```

```
## GVIFs computed for predictors

##                          GVIF Df GVIF^(1/(2*Df))              Interacts With
## Make             5.512981e+01 25        1.083497                          --
## Year             2.165977e+03  5        2.155586 Kilometer, Max.Power.Value
## Kilometer        1.979648e+01  3        1.644743                        Year
## Fuel.Type        2.689951e+00  2        1.280667                          --
## Transmission     1.915462e+00  1        1.384002                          --
## Location         1.353010e+01 63        1.020889                          --
## Owner            1.632720e+00  3        1.085139                          --
## Seller.Type      1.328754e+00  2        1.073646                          --
## Max.Power.Value  6.523716e+01  5        1.518621                 Year, Width
## Width            1.689522e+06  3       10.913415             Max.Power.Value
## Seating.Capacity 2.040449e+00  1        1.428443                          --
##
## Make             Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Year                                        Make, Fuel.Type, Transmission, Location, Owner, Seller.T
## Kilometer                  Make, Fuel.Type, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Fuel.Type           Make, Year, Kilometer, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Transmission          Make, Year, Kilometer, Fuel.Type, Location, Owner, Seller.Type, Max.Power.Va
## Location          Make, Year, Kilometer, Fuel.Type, Transmission, Owner, Seller.Type, Max.Power.Va
## Owner         Make, Year, Kilometer, Fuel.Type, Transmission, Location, Seller.Type, Max.Power.Va
## Seller.Type       Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Max.Power.Va
## Max.Power.Value              Make, Kilometer, Fuel.Type, Transmission, Location, Owner, S
## Width                     Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, S
## Seating.Capacity        Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller
```

The multicollinearity from width is surprising. We will remove that interaction term.

```
final <- lm(
  `log(Price)` ~ . + Year:Kilometer + Year:Max.Power.Value,
  data = x_final
)
```

```
vif(final, type = "predictor")
```

```
## GVIFs computed for predictors

##                   GVIF Df GVIF^(1/(2*Df))              Interacts With
## Make         40.192834 25        1.076671                          --
## Year         16.007788  5        1.319572 Kilometer, Max.Power.Value
```

```
## Kilometer       19.689673  3       1.643261                          Year
## Fuel.Type         2.503787  2       1.257909                          --
## Transmission      1.903712  1       1.379751                          --
## Location         12.911019 63       1.020510                          --
## Owner             1.632165  3       1.085077                          --
## Seller.Type       1.327301  2       1.073352                          --
## Max.Power.Value  39.472819  3       1.845227                          Year
## Width             4.240769  1       2.059313                          --
## Seating.Capacity  1.982470  1       1.408002                          --
##
## Make              Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Year                                          Make, Fuel.Type, Transmission, Location, Owner, Seller.T
## Kilometer                   Make, Fuel.Type, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Fuel.Type           Make, Year, Kilometer, Transmission, Location, Owner, Seller.Type, Max.Power.Va
## Transmission        Make, Year, Kilometer, Fuel.Type, Location, Owner, Seller.Type, Max.Power.Va
## Location            Make, Year, Kilometer, Fuel.Type, Transmission, Owner, Seller.Type, Max.Power.Va
## Owner               Make, Year, Kilometer, Fuel.Type, Transmission, Location, Seller.Type, Max.Power.Va
## Seller.Type         Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Max.Power.Va
## Max.Power.Value             Make, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller.T
## Width               Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller.Type, Max.
## Seating.Capacity            Make, Year, Kilometer, Fuel.Type, Transmission, Location, Owner, Seller
```
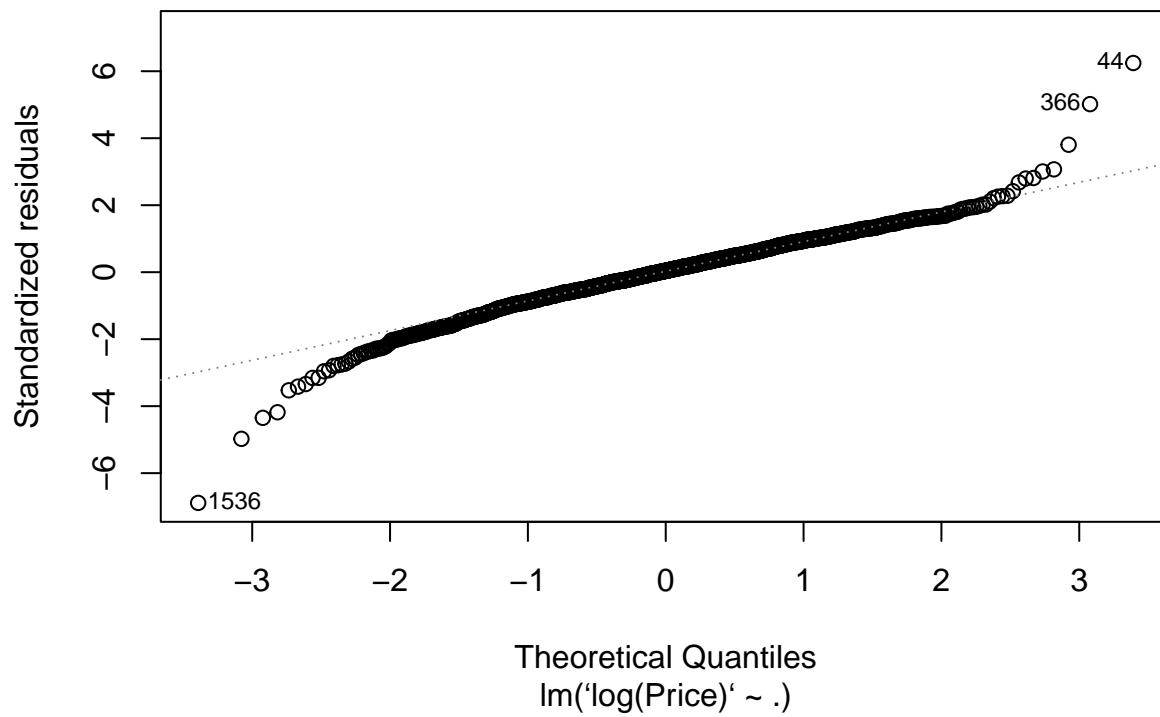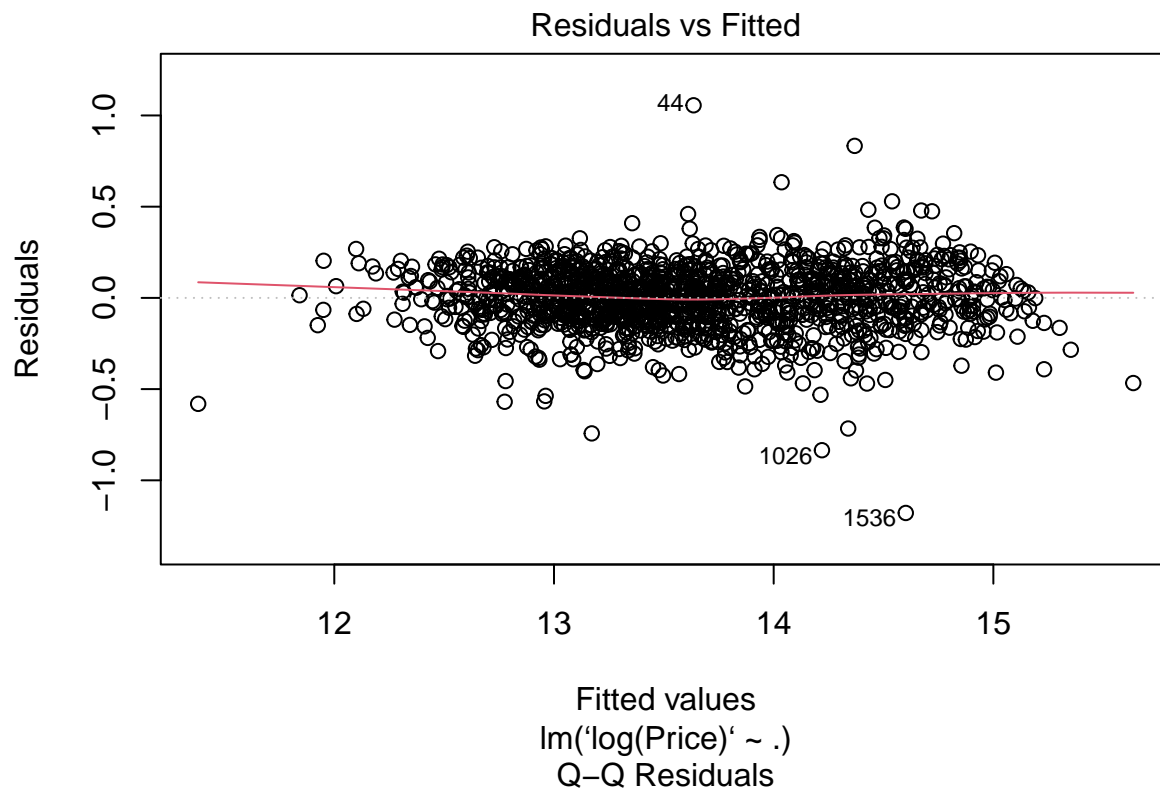
We get that GVIF^(1/(2*Df)) $< 2.236068 \approx \sqrt{5}$, for all the regressors, so multicollinearity is low.
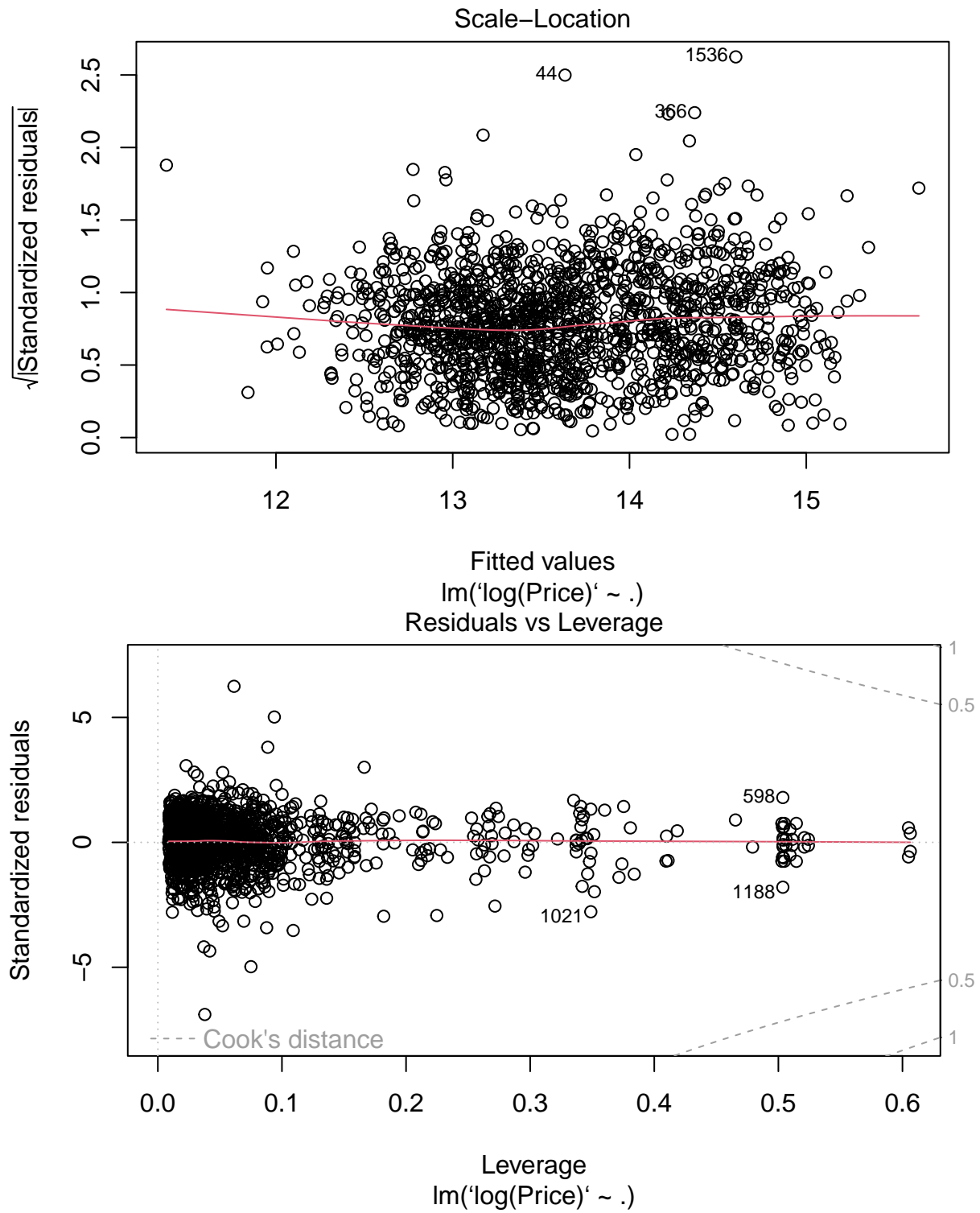
```
which(abs(rstudent(final)) > 4)
```

```
##   44  366  492 1026 1536 1681
##   39  268  345  721 1071 1171
```

After inspecting these data points, we do not find a good reason to remove them (i.e., they are not clerical errors).

```
plot(model)
```

Residuals vs Fitted

Fitted values
lm('log(Price)' ~ .)

Q−Q Residuals

Theoretical Quantiles
lm('log(Price)' ~ .)

Scale–Location

Fitted values
lm(`log(Price)` ~ .)

Residuals vs Leverage

Leverage
lm(`log(Price)` ~ .)

Inspect Model Coefficients

```
summary(final)
```

```
##
## Call:
```

```
## lm(formula = `log(Price)` ~ . + Year:Kilometer + Year:Max.Power.Value,
##     data = x_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99321 -0.09676  0.00175  0.10187  1.06780
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -8.847e+01  1.155e+01  -7.662 3.49e-14 ***
## MakeBMW                -8.977e-02  3.691e-02  -2.432 0.015152 *
## MakeChevrolet          -7.444e-01  8.863e-02  -8.399  < 2e-16 ***
## MakeDatsun             -8.348e-01  6.813e-02 -12.252  < 2e-16 ***
## MakeFord               -5.387e-01  3.793e-02 -14.202  < 2e-16 ***
## MakeHonda              -3.762e-01  2.908e-02 -12.935  < 2e-16 ***
## MakeHyundai            -4.086e-01  2.700e-02 -15.135  < 2e-16 ***
## MakeIsuzu              -4.525e-01  1.240e-01  -3.650 0.000272 ***
## MakeJaguar             -9.940e-02  7.269e-02  -1.367 0.171729
## MakeJeep               -4.595e-01  4.957e-02  -9.270  < 2e-16 ***
## MakeKia                -3.375e-01  4.323e-02  -7.806 1.18e-14 ***
## MakeLand Rover          6.612e-02  9.940e-02   0.665 0.506083
## MakeMahindra           -5.776e-01  3.392e-02 -17.030  < 2e-16 ***
## MakeMaruti Suzuki      -3.958e-01  2.926e-02 -13.527  < 2e-16 ***
## MakeMercedes-Benz       1.436e-01  2.971e-02   4.835 1.49e-06 ***
## MakeMG                 -4.922e-01  5.347e-02  -9.206  < 2e-16 ***
## MakeMINI                5.300e-01  7.278e-02   7.282 5.60e-13 ***
## MakeMitsubishi         -2.703e-01  1.096e-01  -2.466 0.013779 *
## MakeNissan             -5.133e-01  5.247e-02  -9.783  < 2e-16 ***
## MakeRenault            -6.070e-01  3.894e-02 -15.589  < 2e-16 ***
## MakeSkoda              -3.688e-01  3.549e-02 -10.394  < 2e-16 ***
## MakeSsangyong          -8.007e-01  1.100e-01  -7.279 5.70e-13 ***
## MakeTata               -6.823e-01  3.650e-02 -18.694  < 2e-16 ***
## MakeToyota             -1.627e-01  3.009e-02  -5.407 7.58e-08 ***
## MakeVolkswagen         -3.384e-01  3.742e-02  -9.043  < 2e-16 ***
## MakeVolvo              -6.313e-02  5.482e-02  -1.151 0.249741
## Year                    4.904e-02  5.734e-03   8.552  < 2e-16 ***
## Kilometer              -5.198e-04  1.159e-04  -4.486 7.88e-06 ***
## Fuel.TypeDiesel        -6.911e-02  3.504e-02  -1.972 0.048760 *
## Fuel.TypePetrol        -2.103e-01  3.333e-02  -6.309 3.81e-10 ***
## TransmissionManual     -1.134e-01  1.253e-02  -9.049  < 2e-16 ***
## LocationAhmedabad       7.455e-02  4.725e-02   1.578 0.114896
## LocationAllahabad      -3.741e-02  1.055e-01  -0.355 0.722856
## LocationAmbala Cantt   -1.004e-01  6.791e-02  -1.479 0.139476
## LocationAmritsar        5.070e-02  9.516e-02   0.533 0.594248
## LocationAurangabad      2.610e-01  1.055e-01   2.475 0.013455 *
## LocationBangalore       2.410e-01  4.468e-02   5.393 8.17e-08 ***
## LocationBhopal          9.776e-02  1.101e-01   0.888 0.374532
## LocationBhubaneswar     3.593e-02  1.099e-01   0.327 0.743705
## LocationBulandshahar   -2.556e-01  1.281e-01  -1.996 0.046156 *
## LocationChandigarh      4.818e-02  6.026e-02   0.800 0.424108
## LocationChennai         2.248e-01  4.962e-02   4.531 6.41e-06 ***
## LocationCoimbatore      2.984e-01  5.759e-02   5.181 2.55e-07 ***
## LocationDehradun        9.870e-02  5.679e-02   1.738 0.082458 .
## LocationDelhi           6.062e-02  4.284e-02   1.415 0.157270
```

```
## LocationDharwad          1.888e-01  1.259e-01   1.500 0.133965
## LocationErnakulam        2.328e-01  1.262e-01   1.844 0.065391 .
## LocationFaridabad       -1.734e-02  5.493e-02  -0.316 0.752239
## LocationGhaziabad        7.271e-03  1.265e-01   0.057 0.954187
## LocationGoa              2.701e-01  9.542e-02   2.831 0.004711 **
## LocationGurgaon          1.683e-02  5.471e-02   0.308 0.758440
## LocationGuwahati         1.865e-01  8.689e-02   2.147 0.031994 *
## LocationHaldwani        -1.268e-01  1.258e-01  -1.008 0.313534
## LocationHyderabad        1.846e-01  4.531e-02   4.073 4.91e-05 ***
## LocationIndore           1.894e-01  9.412e-02   2.012 0.044409 *
## LocationJaipur           1.408e-01  5.340e-02   2.636 0.008476 **
## LocationJalandhar        1.340e-01  5.624e-02   2.383 0.017296 *
## LocationJamshedpur       2.981e-02  8.730e-02   0.341 0.732783
## LocationKanpur          -2.542e-02  5.045e-02  -0.504 0.614491
## LocationKarnal           1.299e-01  7.246e-02   1.792 0.073297 .
## LocationKharar          -4.311e-02  1.258e-01  -0.343 0.731955
## LocationKheda           -6.993e-02  1.259e-01  -0.555 0.578785
## LocationKolkata         -6.892e-02  4.737e-02  -1.455 0.145924
## LocationKollam           1.623e-01  1.262e-01   1.285 0.198869
## LocationKota             2.447e-02  1.373e-01   0.178 0.858577
## LocationLucknow          5.244e-02  4.742e-02   1.106 0.269012
## LocationLudhiana         6.154e-02  5.008e-02   1.229 0.219386
## LocationMangalore        1.583e-02  9.416e-02   0.168 0.866496
## LocationMeerut           3.421e-02  7.269e-02   0.471 0.638025
## LocationMirzapur         6.818e-02  1.370e-01   0.498 0.618784
## LocationMohali           1.272e-01  5.414e-02   2.349 0.018979 *
## LocationMumbai           1.145e-01  4.274e-02   2.679 0.007480 **
## LocationMuzaffurpur     -7.048e-02  1.260e-01  -0.560 0.575904
## LocationMysore           3.626e-01  7.618e-02   4.759 2.15e-06 ***
## LocationNagpur           1.879e-01  1.056e-01   1.780 0.075313 .
## LocationNashik           9.824e-02  7.597e-02   1.293 0.196158
## LocationNavi Mumbai      1.165e-01  6.121e-02   1.903 0.057199 .
## LocationNoida            1.260e-01  6.247e-02   2.017 0.043921 *
## LocationPanchkula       -3.072e-02  1.054e-01  -0.292 0.770688
## LocationPatna            1.288e-01  5.073e-02   2.539 0.011234 *
## LocationPune             1.574e-01  4.487e-02   3.508 0.000466 ***
## LocationRaipur           1.267e-01  5.896e-02   2.149 0.031786 *
## LocationRanchi          -1.564e-02  5.910e-02  -0.265 0.791367
## LocationRanga Reddy     -6.941e-02  1.270e-01  -0.546 0.584828
## LocationRoorkee          1.445e-01  1.070e-01   1.351 0.176837
## LocationSalem            4.298e-01  1.260e-01   3.411 0.000666 ***
## LocationSurat            6.901e-02  8.588e-02   0.804 0.421771
## LocationThane            1.154e-01  6.071e-02   1.901 0.057546 .
## LocationUdupi            1.787e-01  7.234e-02   2.471 0.013606 *
## LocationVadodara        -9.279e-02  1.057e-01  -0.878 0.379977
## LocationVaranasi        -6.602e-02  5.898e-02  -1.119 0.263210
## LocationWarangal        -7.010e-02  1.259e-01  -0.557 0.577784
## LocationYamunanagar      1.718e-01  1.062e-01   1.619 0.105785
## LocationZirakpur         2.381e-02  6.059e-02   0.393 0.694424
## OwnerSecond             -1.501e-02  1.306e-02  -1.149 0.250624
## OwnerThird              -5.843e-02  4.731e-02  -1.235 0.217049
## OwnerUnRegistered Car    1.085e-01  8.949e-02   1.213 0.225407
## Seller.TypeCorporate     1.462e-02  9.712e-02   0.151 0.880383
## Seller.TypeIndividual    2.271e-02  9.219e-02   0.246 0.805477
```

```
## Max.Power.Value      -8.091e-01  8.766e-02  -9.229  < 2e-16 ***
## Width                 1.543e-03  1.009e-04  15.293  < 2e-16 ***
## Seating.Capacity      4.081e-02  8.549e-03   4.774 2.01e-06 ***
## Year:Kilometer        2.571e-07  5.745e-08   4.475 8.30e-06 ***
## Year:Max.Power.Value  4.048e-04  4.346e-05   9.312  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1675 on 1337 degrees of freedom
## Multiple R-squared:  0.9463, Adjusted R-squared:  0.9422
## F-statistic: 228.9 on 103 and 1337 DF,  p-value: < 2.2e-16
```

```r
anova(final)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                     Df Sum Sq Mean Sq   F value     Pr(>F)
## Make                25 354.44  14.178  505.4109 < 2.2e-16 ***
## Year                 1 178.52 178.525 6364.1385 < 2.2e-16 ***
## Kilometer            1   0.98   0.982   34.9897 4.204e-09 ***
## Fuel.Type            2  19.23   9.615  342.7670 < 2.2e-16 ***
## Transmission         1  20.31  20.313  724.1111 < 2.2e-16 ***
## Location            63  16.76   0.266    9.4820 < 2.2e-16 ***
## Owner                3   0.46   0.153    5.4658 0.0009824 ***
## Seller.Type          2   0.33   0.165    5.8950 0.0028252 **
## Max.Power.Value      1  60.04  60.045 2140.5065 < 2.2e-16 ***
## Width                1   6.08   6.076  216.6025 < 2.2e-16 ***
## Seating.Capacity     1   0.81   0.805   28.7132 9.874e-08 ***
## Year:Kilometer       1   1.01   1.009   35.9779 2.565e-09 ***
## Year:Max.Power.Value 1   2.43   2.433   86.7208 < 2.2e-16 ***
## Residuals         1337  37.51   0.028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Normalized Coefficients

```r
normalized <- x_final
for (col in colnames(normalized)) {
  if (col != "log(Price)" && !(col %in% categorical)) {
    normalized[, col] <- (normalized[, col] - mean(normalized[, col])) / sd(normalized[, col])
  }
}

normalized_model <- lm(`log(Price)` ~ . + Year:Kilometer + Year:Max.Power.Value, data = normalized)

sort(abs(normalized_model$coefficients), decreasing = T)
```

```
##     (Intercept)          MakeDatsun        MakeSsangyong
##    14.083398105         0.834753631          0.800670555
##    MakeChevrolet            MakeTata          MakeRenault
##     0.744382773         0.682276167          0.607023106
##     MakeMahindra            MakeFord              MakeMINI
##     0.577601300         0.538687730          0.529950013
##      MakeNissan              MakeMG              MakeJeep
```

```
##        0.513258876            0.492244337            0.459497184
##          MakeIsuzu          LocationSalem            MakeHyundai
##        0.452503248            0.429808202            0.408611709
##    MakeMaruti Suzuki             MakeHonda              MakeSkoda
##        0.395791191            0.376204847            0.368848089
##       LocationMysore        MakeVolkswagen                MakeKia
##        0.362571923            0.338409217            0.337455723
##               Year      LocationCoimbatore        Max.Power.Value
##        0.309169876            0.298383889            0.289135842
##       MakeMitsubishi           LocationGoa       LocationAurangabad
##        0.270254156            0.270137564            0.261025878
## LocationBulandshahar      LocationBangalore       LocationErnakulam
##        0.255630170            0.240990910            0.232786160
##      LocationChennai        Fuel.TypePetrol          LocationIndore
##        0.224792375            0.210301849            0.189379678
##      LocationDharwad          LocationNagpur       LocationGuwahati
##        0.188751967            0.187929972            0.186538760
##    LocationHyderabad           LocationUdupi     LocationYamunanagar
##        0.184568584            0.178736786            0.171839738
##           MakeToyota          LocationKollam            LocationPune
##        0.162680016            0.162280107            0.157409319
##       LocationRoorkee       MakeMercedes-Benz         LocationJaipur
##        0.144528570            0.143629101            0.140775223
##               Width        LocationJalandhar          LocationKarnal
##        0.138997947            0.134032148            0.129880430
##        LocationPatna          LocationMohali       LocationHaldwani
##        0.128805119            0.127161838            0.126834461
##        LocationRaipur          LocationNoida      LocationNavi Mumbai
##        0.126723339            0.125986284            0.116505798
##        LocationThane          LocationMumbai      TransmissionManual
##        0.115394024            0.114492074            0.113399679
## OwnerUnRegistered Car  LocationAmbala Cantt             MakeJaguar
##        0.108542121            0.100415170            0.099397093
##      LocationDehradun          LocationNashik          LocationBhopal
##        0.098700492            0.098242282            0.097758921
##      LocationVadodara                MakeBMW        LocationAhmedabad
##        0.092791719            0.089768746            0.074545756
##    LocationMuzaffurpur       LocationWarangal            LocationKheda
##        0.070477682            0.070096294            0.069933069
##    LocationRanga Reddy        Fuel.TypeDiesel           LocationSurat
##        0.069409159            0.069109060            0.069009846
##      LocationKolkata         LocationMirzapur         MakeLand Rover
##        0.068923627            0.068177156            0.066116951
##      LocationVaranasi              MakeVolvo          LocationLudhiana
##        0.066015022            0.063129319            0.061538227
##        LocationDelhi             OwnerThird          LocationLucknow
##        0.060618355            0.058433339            0.052441008
##      LocationAmritsar       LocationChandigarh    Year:Max.Power.Value
##        0.050702042            0.048184226            0.046211274
##        LocationKharar         LocationAllahabad               Kilometer
##        0.043111915            0.037408412            0.036719114
##    LocationBhubaneswar           LocationMeerut        LocationPanchkula
##        0.035928949            0.034206532            0.030722943
##      LocationJamshedpur       Seating.Capacity           LocationKanpur
```

12

```
##        0.029812782              0.029667255           0.025417328
##           LocationKota       LocationZirakpur Seller.TypeIndividual
##        0.024474683              0.023809079           0.022707801
##         Year:Kilometer       LocationFaridabad         LocationGurgaon
##        0.020100327              0.017344872           0.016828717
##     LocationMangalore          LocationRanchi            OwnerSecond
##        0.015832281              0.015636096           0.015008599
##  Seller.TypeCorporate       LocationGhaziabad
##        0.014617277              0.007270709
```

Confidence Intervals

```
confint(final, level = 0.95)
```

```
##                             2.5 %        97.5 %
## (Intercept)           -1.111237e+02 -6.582168e+01
## MakeBMW               -1.621837e-01 -1.735379e-02
## MakeChevrolet         -9.182476e-01 -5.705180e-01
## MakeDatsun            -9.684109e-01 -7.010964e-01
## MakeFord              -6.130975e-01 -4.642780e-01
## MakeHonda             -4.332615e-01 -3.191482e-01
## MakeHyundai           -4.615744e-01 -3.556490e-01
## MakeIsuzu             -6.957042e-01 -2.093023e-01
## MakeJaguar            -2.419968e-01  4.320265e-02
## MakeJeep              -5.567326e-01 -3.622618e-01
## MakeKia               -4.222594e-01 -2.526520e-01
## MakeLand Rover        -1.288894e-01  2.611233e-01
## MakeMahindra          -6.441384e-01 -5.110642e-01
## MakeMaruti Suzuki     -4.531920e-01 -3.383904e-01
## MakeMercedes-Benz      8.534960e-02  2.019086e-01
## MakeMG                -5.971360e-01 -3.873526e-01
## MakeMINI               3.871803e-01  6.727197e-01
## MakeMitsubishi        -4.852245e-01 -5.528379e-02
## MakeNissan            -6.161828e-01 -4.103349e-01
## MakeRenault           -6.834112e-01 -5.306350e-01
## MakeSkoda             -4.384638e-01 -2.992324e-01
## MakeSsangyong         -1.016446e+00 -5.848947e-01
## MakeTata              -7.538747e-01 -6.106776e-01
## MakeToyota            -2.217030e-01 -1.036571e-01
## MakeVolkswagen        -4.118245e-01 -2.649939e-01
## MakeVolvo             -1.706809e-01  4.442223e-02
## Year                   3.778860e-02  6.028441e-02
## Kilometer             -7.470509e-04 -2.924697e-04
## Fuel.TypeDiesel       -1.378415e-01 -3.766142e-04
## Fuel.TypePetrol       -2.756955e-01 -1.449082e-01
## TransmissionManual    -1.379826e-01 -8.881678e-02
## LocationAhmedabad     -1.815207e-02  1.672436e-01
## LocationAllahabad     -2.442918e-01  1.694749e-01
## LocationAmbala Cantt  -2.336391e-01  3.280876e-02
## LocationAmritsar      -1.359736e-01  2.373777e-01
## LocationAurangabad     5.411343e-02  4.679383e-01
## LocationBangalore      1.533346e-01  3.286473e-01
## LocationBhopal        -1.181306e-01  3.136485e-01
## LocationBhubaneswar   -1.796040e-01  2.514619e-01
```

```
## LocationBulandshahar   -5.068953e-01 -4.365059e-03
## LocationChandigarh     -7.003699e-02  1.664054e-01
## LocationChennai         1.274570e-01  3.221277e-01
## LocationCoimbatore      1.854031e-01  4.113647e-01
## LocationDehradun       -1.271244e-02  2.101134e-01
## LocationDelhi          -2.341583e-02  1.446525e-01
## LocationDharwad        -5.817554e-02  4.356795e-01
## LocationErnakulam      -1.485169e-02  4.804240e-01
## LocationFaridabad      -1.251065e-01  9.041674e-02
## LocationGhaziabad      -2.409582e-01  2.554996e-01
## LocationGoa             8.293995e-02  4.573352e-01
## LocationGurgaon        -9.049991e-02  1.241573e-01
## LocationGuwahati        1.607444e-02  3.570031e-01
## LocationHaldwani       -3.736233e-01  1.199544e-01
## LocationHyderabad       9.567332e-02  2.734639e-01
## LocationIndore          4.740455e-03  3.740189e-01
## LocationJaipur          3.602532e-02  2.455251e-01
## LocationJalandhar       2.370898e-02  2.443553e-01
## LocationJamshedpur     -1.414476e-01  2.010732e-01
## LocationKanpur         -1.243911e-01  7.355647e-02
## LocationKarnal         -1.227173e-02  2.720326e-01
## LocationKharar         -2.899742e-01  2.037504e-01
## LocationKheda          -3.169911e-01  1.771249e-01
## LocationKolkata        -1.618561e-01  2.400889e-02
## LocationKollam         -8.538357e-02  4.099438e-01
## LocationKota           -2.449270e-01  2.938764e-01
## LocationLucknow        -4.059170e-02  1.454737e-01
## LocationLudhiana       -3.671050e-02  1.597870e-01
## LocationMangalore      -1.688845e-01  2.005491e-01
## LocationMeerut         -1.083959e-01  1.768090e-01
## LocationMirzapur       -2.005555e-01  3.369098e-01
## LocationMohali          2.095495e-02  2.333687e-01
## LocationMumbai          3.064645e-02  1.983377e-01
## LocationMuzaffurpur    -3.175823e-01  1.766269e-01
## LocationMysore          2.131244e-01  5.120195e-01
## LocationNagpur         -1.919501e-02  3.950550e-01
## LocationNashik         -5.078535e-02  2.472699e-01
## LocationNavi Mumbai    -3.569190e-03  2.365808e-01
## LocationNoida           3.436347e-03  2.485362e-01
## LocationPanchkula      -2.374581e-01  1.760122e-01
## LocationPatna           2.927935e-02  2.283309e-01
## LocationPune            6.938280e-02  2.454358e-01
## LocationRaipur          1.106101e-02  2.423857e-01
## LocationRanchi         -1.315666e-01  1.002945e-01
## LocationRanga Reddy    -3.185731e-01  1.797548e-01
## LocationRoorkee        -6.529401e-02  3.543512e-01
## LocationSalem           1.826441e-01  6.769723e-01
## LocationSurat          -9.945643e-02  2.374761e-01
## LocationThane          -3.700910e-03  2.344890e-01
## LocationUdupi           3.682441e-02  3.206492e-01
## LocationVadodara       -3.000647e-01  1.144813e-01
## LocationVaranasi       -1.817154e-01  4.968536e-02
## LocationWarangal       -3.170796e-01  1.768870e-01
## LocationYamunanagar    -3.643904e-02  3.801185e-01
```

```
## LocationZirakpur       -9.505607e-02  1.426742e-01
## OwnerSecond            -4.062604e-02  1.060885e-02
## OwnerThird             -1.512523e-01  3.438567e-02
## OwnerUnRegistered Car  -6.702301e-02  2.841073e-01
## Seller.TypeCorporate   -1.759010e-01  2.051356e-01
## Seller.TypeIndividual  -1.581458e-01  2.035614e-01
## Max.Power.Value        -9.810376e-01 -6.370908e-01
## Width                   1.345327e-03  1.741270e-03
## Seating.Capacity        2.404111e-02  5.758251e-02
## Year:Kilometer          1.443624e-07  3.697602e-07
## Year:Max.Power.Value    3.194923e-04  4.900242e-04
```

Inspection of Seller Type

```r
comp <- aov(`log(Price)`~Seller.Type, data = x)
summary(comp)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## Seller.Type    2    4.8  2.4087    4.99 0.00692 **
## Residuals   1438  694.1  0.4827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(comp, conf.level=.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = `log(Price)` ~ Seller.Type, data = x)
##
## $Seller.Type
##                                       diff         lwr         upr      p adj
## Corporate-Commercial Registration   0.7704811 -0.08982006  1.63078234 0.0899342
## Individual-Commercial Registration  0.4253785 -0.39077327  1.24153018 0.4398371
## Individual-Corporate               -0.3451027 -0.62403742 -0.06616795 0.0104896
```

Corporate sellers are able to sell their cars at a statistically significantly higher price compared to individual sellers.