Class_Activity#6

Sunny Lee

2021-02-22

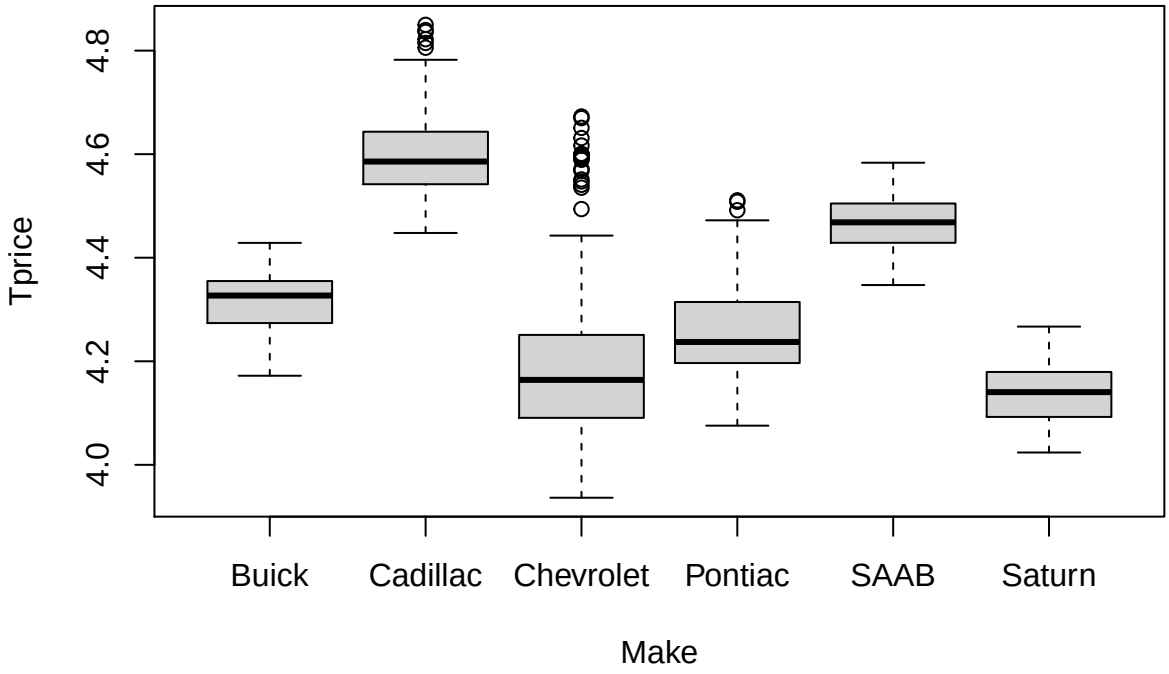
Categorical Explanatory Variables:

```
cars <- read.csv("C3 Cars.csv")
attach(cars)
table(Make)
```

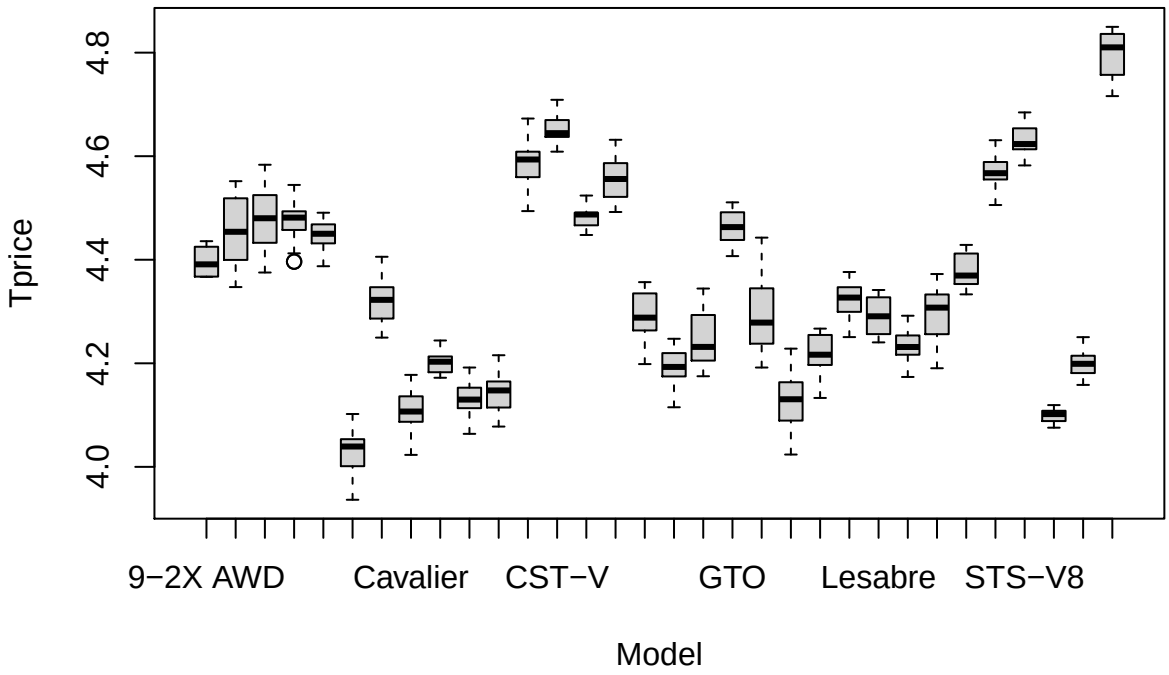
```
## Make
##     Buick  Cadillac Chevrolet   Pontiac      SAAB    Saturn
##        80        80       320       150       114        60
```

1) Create boxplots or individual value plots of the response variable TPrice versus the categorical variables Make, Model, Trim, and Type. Describe any pattern you see.

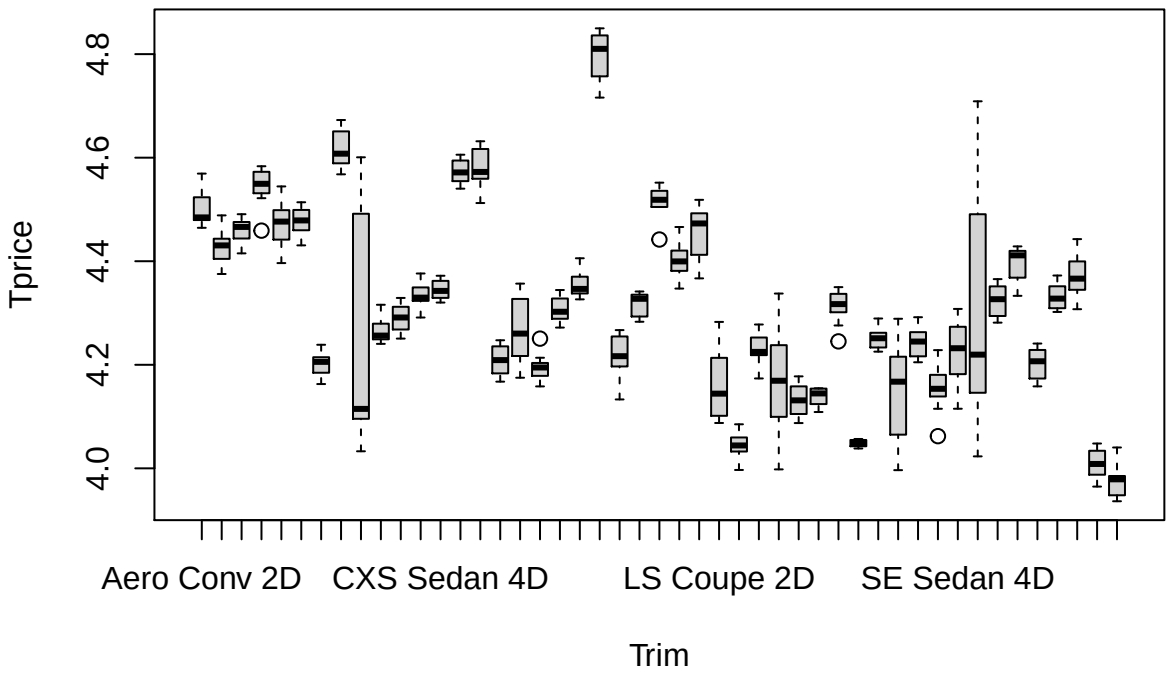
```
Tprice <- log10(Price)
boxplot(Tprice~Make)
```



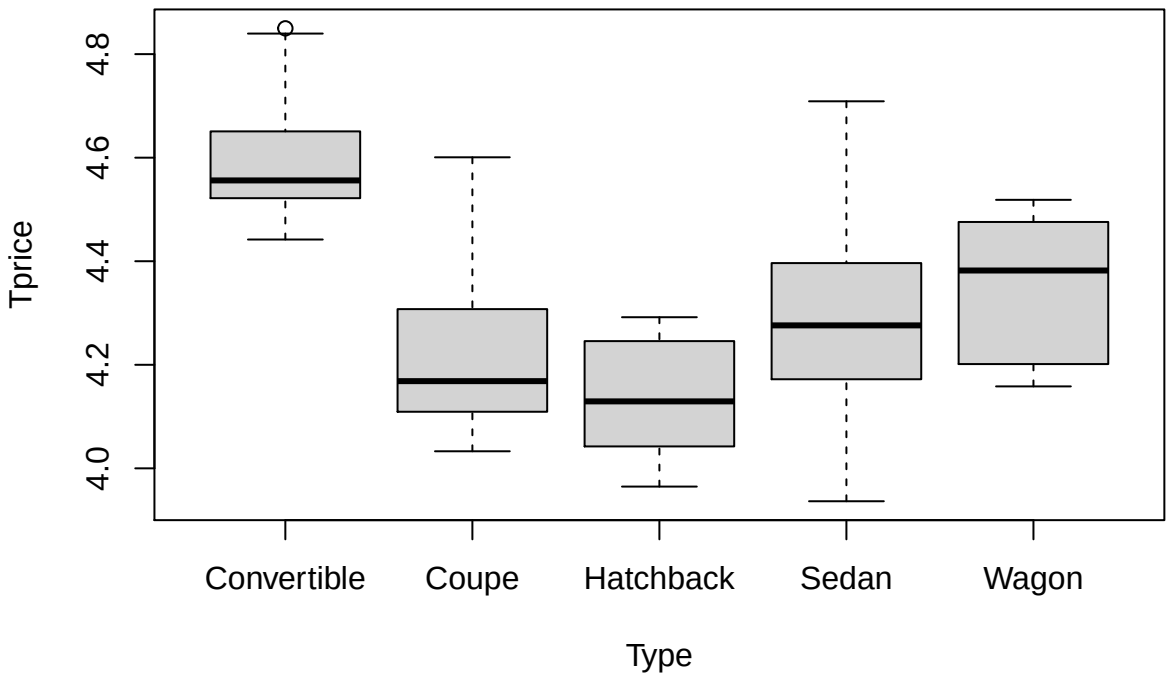
```
boxplot(Tprice~Model)
```




```
boxplot(Tprice~Trim)
```




```
boxplot(Tprice~Type)
```



Throughout every plot, we find each of the categorical variables seem to have very different means and variances.

2) Create an indicator variables for Make. Name the columns, in order, Buick, Cadillac, Chevrolet, Pontiac, SAAB, and Saturn. Look at the new data columns and describe how the indicator variables are defined. For example, list all possible outcomes for the Cadillac indicator variable and explain what each outcome represents.

```
Buick <- (Make == "Buick")*1
Cadillac <- (Make == "Cadillac")*1
Chevrolet<-(Make == "Chevrolet")*1
Pontiac<-(Make == "Pontiac")*1
SAAB<-(Make == "SAAB")*1
Saturn<-(Make == "Saturn")*1
cars[Buick, ]
```

```
##      Mileage   Price  Make   Model     Trim  Type Cyl Liter Doors Cruise Sound
## 1       8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.1     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.2     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.3     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.4     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.5     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.6     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.7     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.8     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.9     8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
## 1.10    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4      1     1
```

```
## 1.11    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.12    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.13    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.14    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.15    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.16    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.17    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.18    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.19    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.20    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.21    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.22    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.23    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.24    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.25    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.26    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.27    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.28    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.29    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.30    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.31    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.32    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.33    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.34    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.35    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.36    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.37    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.38    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.39    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.40    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.41    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.42    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.43    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.44    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.45    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.46    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.47    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.48    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.49    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.50    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.51    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.52    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.53    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.54    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.55    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.56    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.57    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.58    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.59    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.60    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.61    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.62    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.63    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.64    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
```

```
## 1.65    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.66    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.67    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.68    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.69    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.70    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.71    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.72    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.73    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.74    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.75    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.76    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.77    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.78    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
## 1.79    8221 17314.1 Buick Century Sedan 4D Sedan   6   3.1     4       1     1
##      Leather  X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10 X.11
## 1          1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.1        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.2        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.3        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.4        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.5        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.6        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.7        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.8        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.9        1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.10       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.11       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.12       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.13       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.14       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.15       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.16       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.17       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.18       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.19       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.20       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.21       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.22       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.23       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.24       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.25       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.26       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.27       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.28       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.29       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.30       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.31       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.32       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.33       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.34       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.35       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.36       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.37       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
```

```
## 1.38       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.39       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.40       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.41       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.42       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.43       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.44       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.45       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.46       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.47       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.48       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.49       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.50       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.51       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.52       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.53       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.54       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.55       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.56       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.57       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.58       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.59       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.60       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.61       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.62       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.63       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.64       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.65       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.66       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.67       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.68       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.69       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.70       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.71       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.72       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.73       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.74       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.75       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.76       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.77       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.78       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
## 1.79       1 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA
```

We define the indicator variables by replacing the make with a 1 if the indicator variable is that make or 0 if it is not.

3) Build a new regression model using Tprice as the response and Mileage, Liter, Saturn, Cadillac, Chevrolet, Pontiac, and SAAB as the explanatory variables. Explain why you expect the R^2 value to increase when you add terms for make?

```
model <- lm(Tprice~Mileage+Liter+Saturn+Cadillac+Chevrolet+Pontiac+SAAB)
model1 <- lm(Tprice~Mileage+Liter+Make)
summary(model)
```

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

```
## lm(formula = Tprice ~ Mileage + Liter + Saturn + Cadillac + Chevrolet +
##     Pontiac + SAAB)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.132503 -0.033074 -0.004553  0.028345  0.181074
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.020e+00  1.033e-02 389.178  < 2e-16 ***
## Mileage     -3.476e-06  2.227e-07 -15.609  < 2e-16 ***
## Liter        9.972e-02  2.000e-03  49.867  < 2e-16 ***
## Saturn      -3.997e-02  9.200e-03  -4.344 1.58e-05 ***
## Cadillac     2.093e-01  8.290e-03  25.250  < 2e-16 ***
## Chevrolet   -4.934e-02  6.641e-03  -7.430 2.81e-13 ***
## Pontiac     -2.636e-02  7.181e-03  -3.670 0.000259 ***
## SAAB         3.053e-01  8.110e-03  37.651  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05158 on 796 degrees of freedom
## Multiple R-squared:  0.9169, Adjusted R-squared:  0.9161
## F-statistic:  1254 on 7 and 796 DF,  p-value: < 2.2e-16
```

When we add more varaibles to our model, the more we train our model to predict different makes. For example, if the model only had Saturn as one of the exaplanatory variables, our model would do a poor job of predicting other makes such as Cadillac. Thus, when we add more variables to our model, we are able to predict a much larger set of makes.

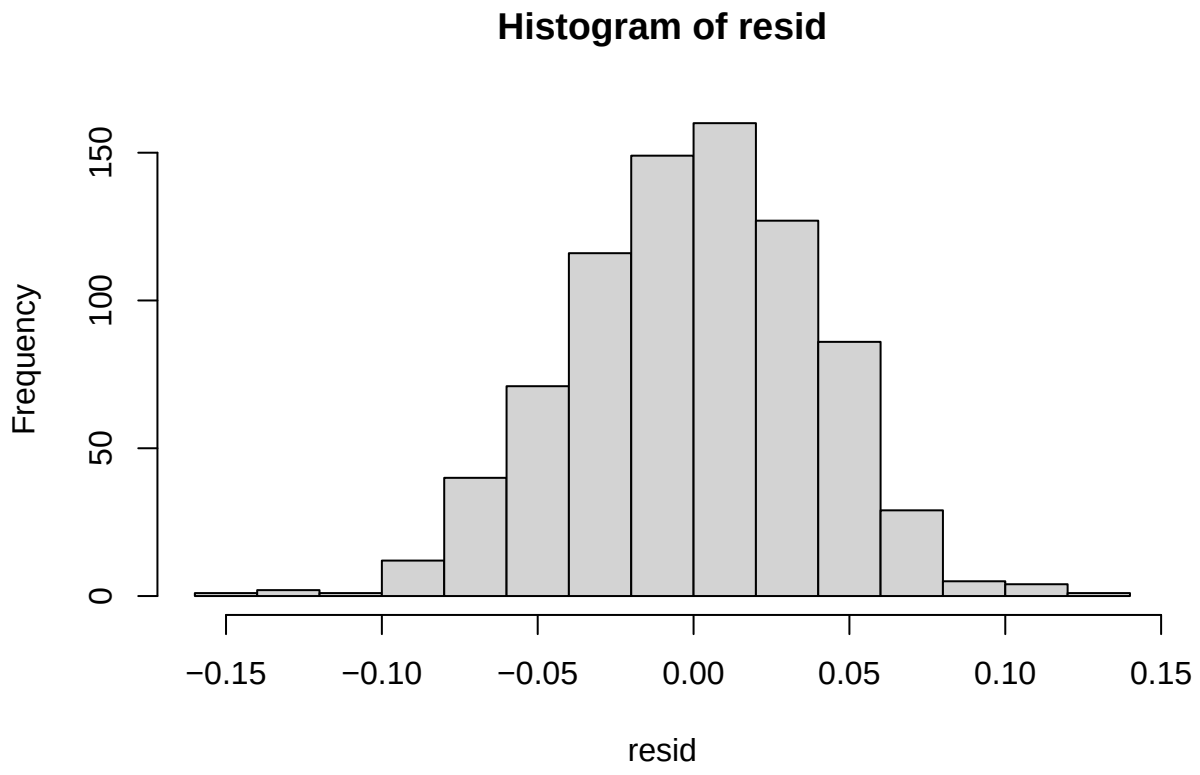
4) Include the Make and Type indicator variables, plus the variables Liter, Doors, Cruise, Sound, Leather, and Mileage, in a model to predict Tprice. Does the normality plot suggest that the residuals could follow a normal distribution? Describe whether the residuals versus fit, the residuals versus order, and the residuals versus each explanatory variables.

```
model2 <- lm(Tprice~Liter+Doors+Cruise+Sound+Leather+Mileage+Make+Type)
summary(model2)
```

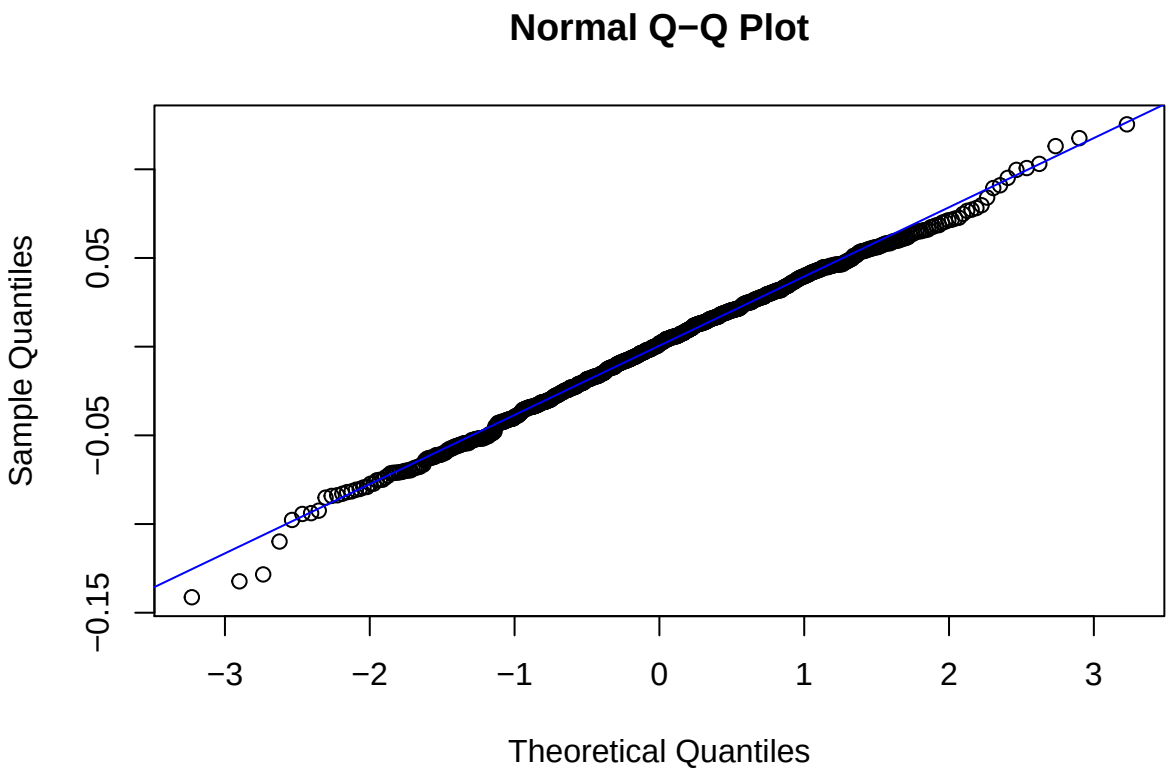
```
##
## Call:
## lm(formula = Tprice ~ Liter + Doors + Cruise + Sound + Leather +
##     Mileage + Make + Type)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.141272 -0.025786  0.001806  0.026857  0.125339
##
## Coefficients: (1 not defined because of singularities)
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.235e+00  1.692e-02 250.290  < 2e-16 ***
## Liter          9.576e-02  1.721e-03  55.642  < 2e-16 ***
## Doors         -3.538e-02  3.985e-03  -8.879  < 2e-16 ***
## Cruise         7.517e-03  4.009e-03   1.875   0.0611 .
## Sound          5.223e-03  3.170e-03   1.647   0.0999 .
## Leather        6.260e-03  3.397e-03   1.843   0.0657 .
## Mileage       -3.576e-06  1.702e-07 -21.018  < 2e-16 ***
## MakeCadillac   1.914e-01  6.650e-03  28.779  < 2e-16 ***
```

```
## MakeChevrolet -5.497e-02  5.605e-03  -9.808  < 2e-16 ***
## MakePontiac   -4.207e-02  5.743e-03  -7.326 5.88e-13 ***
## MakeSAAB       2.394e-01  6.961e-03  34.397  < 2e-16 ***
## MakeSaturn    -4.165e-02  7.441e-03  -5.598 3.00e-08 ***
## TypeCoupe     -1.378e-01  7.306e-03 -18.865  < 2e-16 ***
## TypeHatchback -8.899e-02  8.164e-03 -10.900  < 2e-16 ***
## TypeSedan     -7.115e-02  6.019e-03 -11.821  < 2e-16 ***
## TypeWagon             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03937 on 789 degrees of freedom
## Multiple R-squared:  0.952,  Adjusted R-squared:  0.9512
## F-statistic:  1118 on 14 and 789 DF,  p-value: < 2.2e-16
```

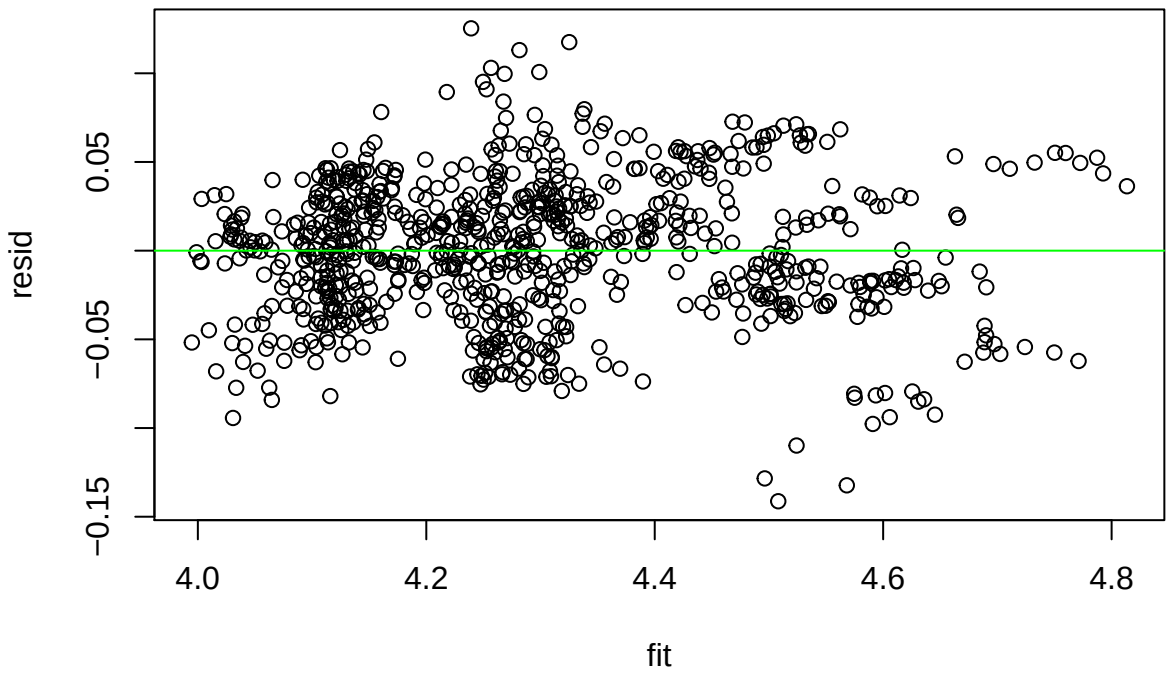
```
resid <- model2$residuals
fit <- model2$fitted.values

hist(resid)
```



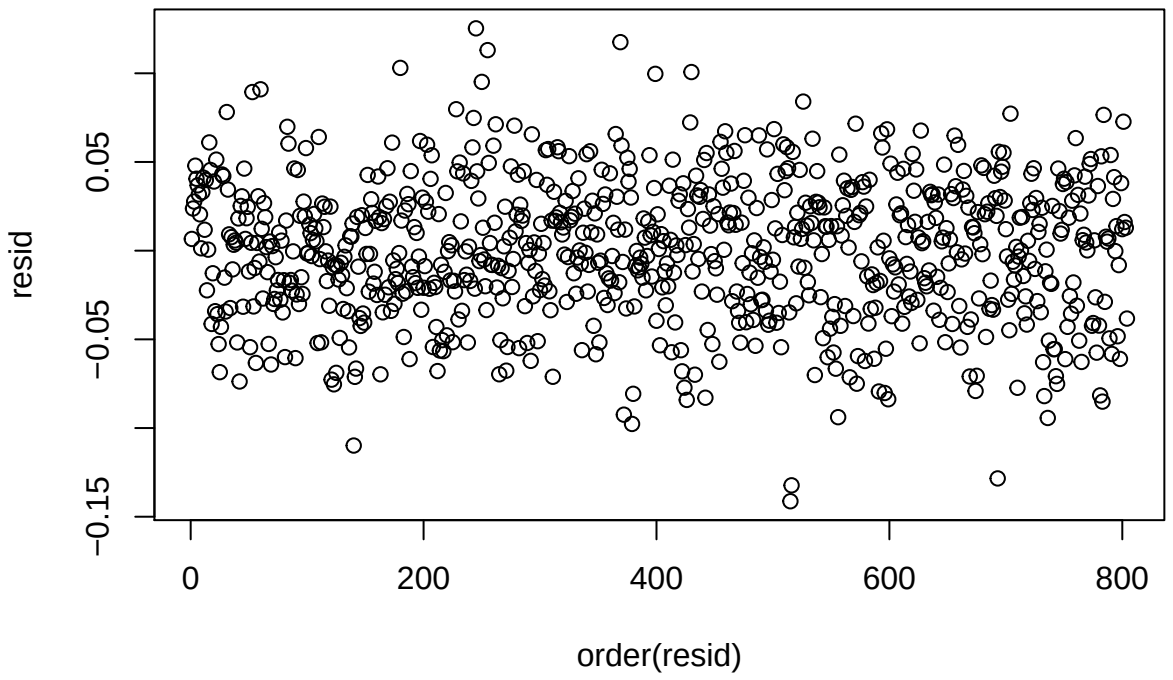

```
qqnorm(resid)
qqline(resid, col = "blue")
```



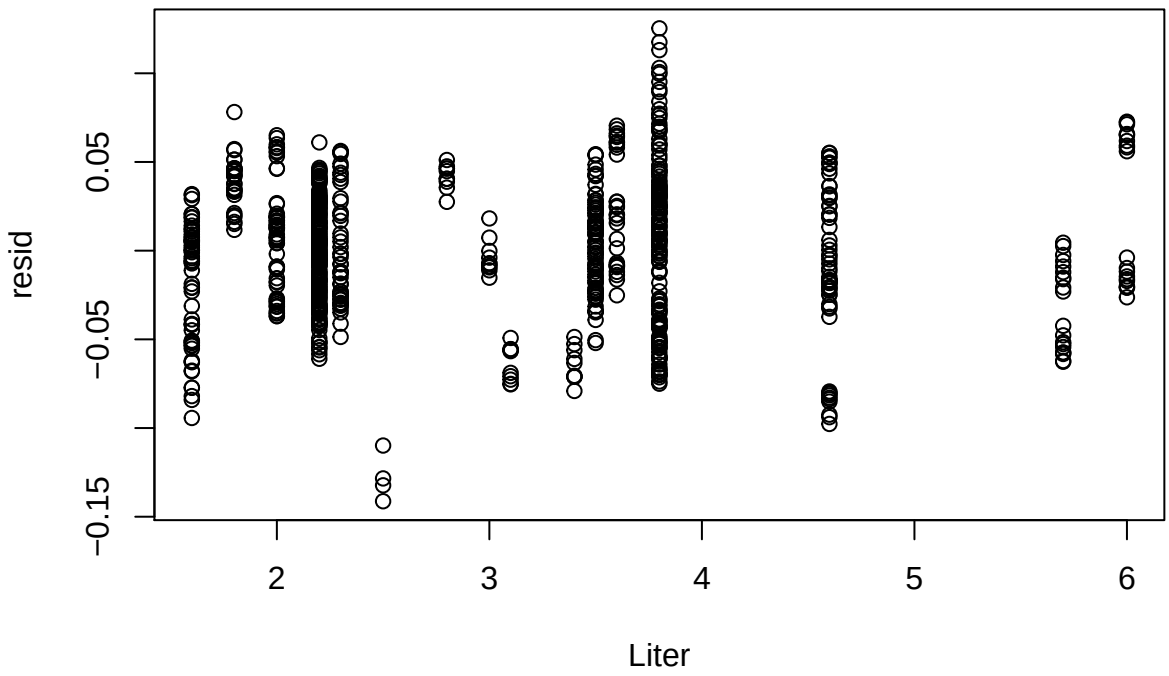

```
plot(resid~fit)
abline(h=0, col = "green")
```

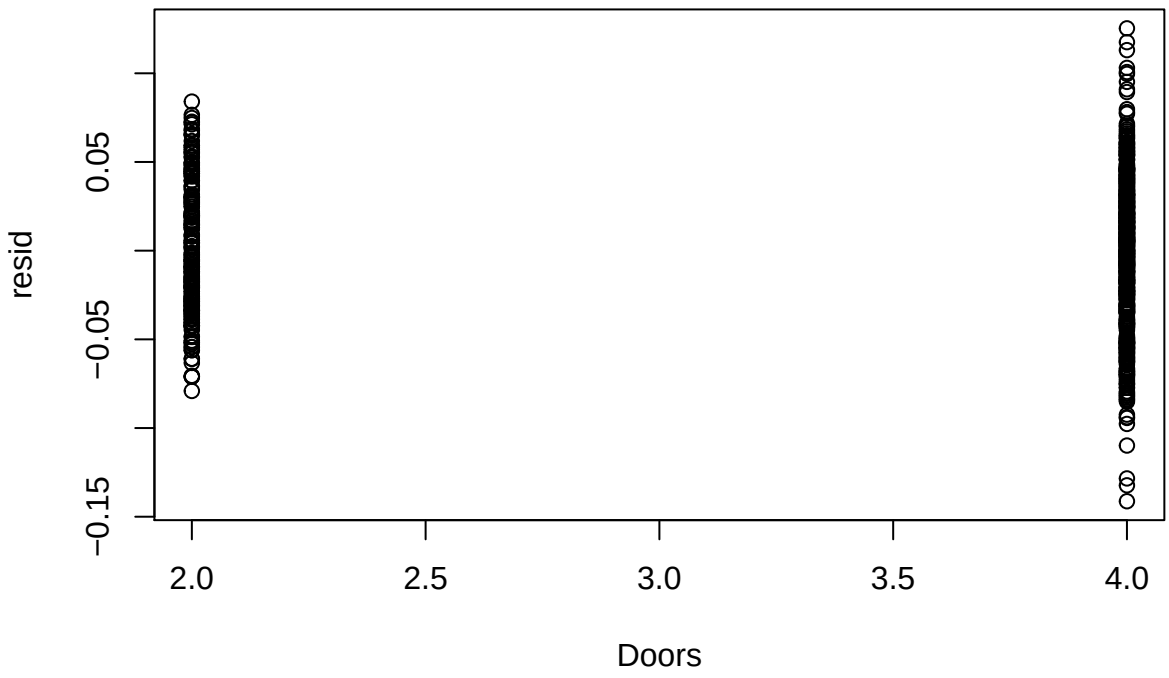



```
plot(resid~order(resid))
```

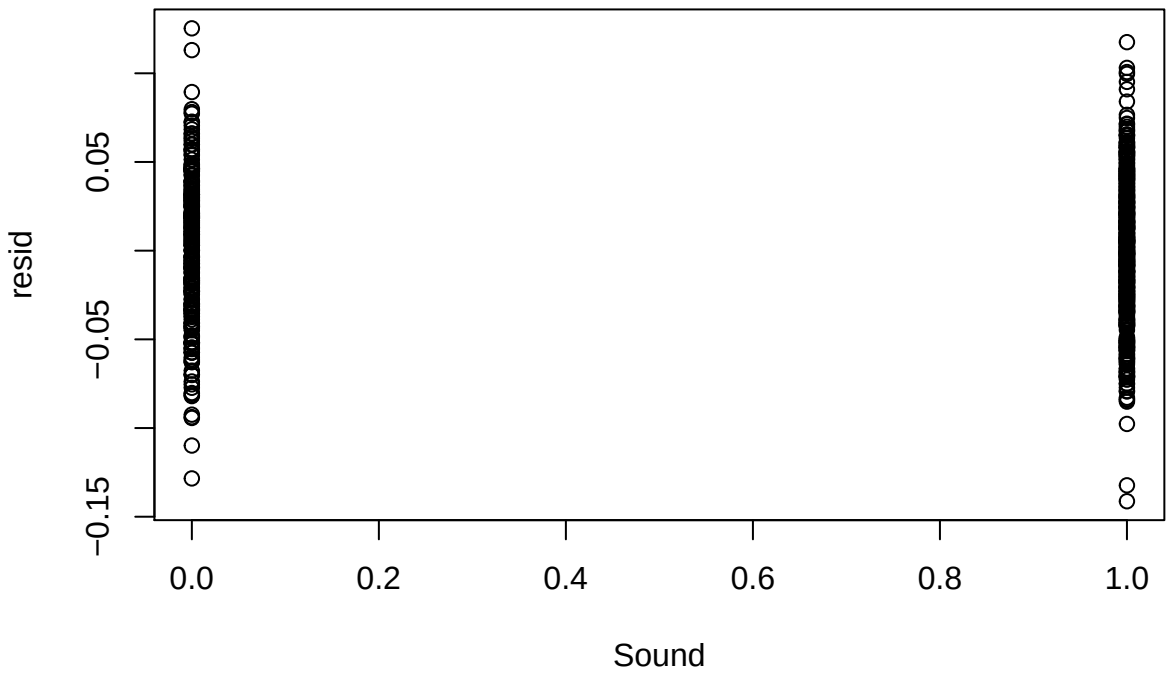


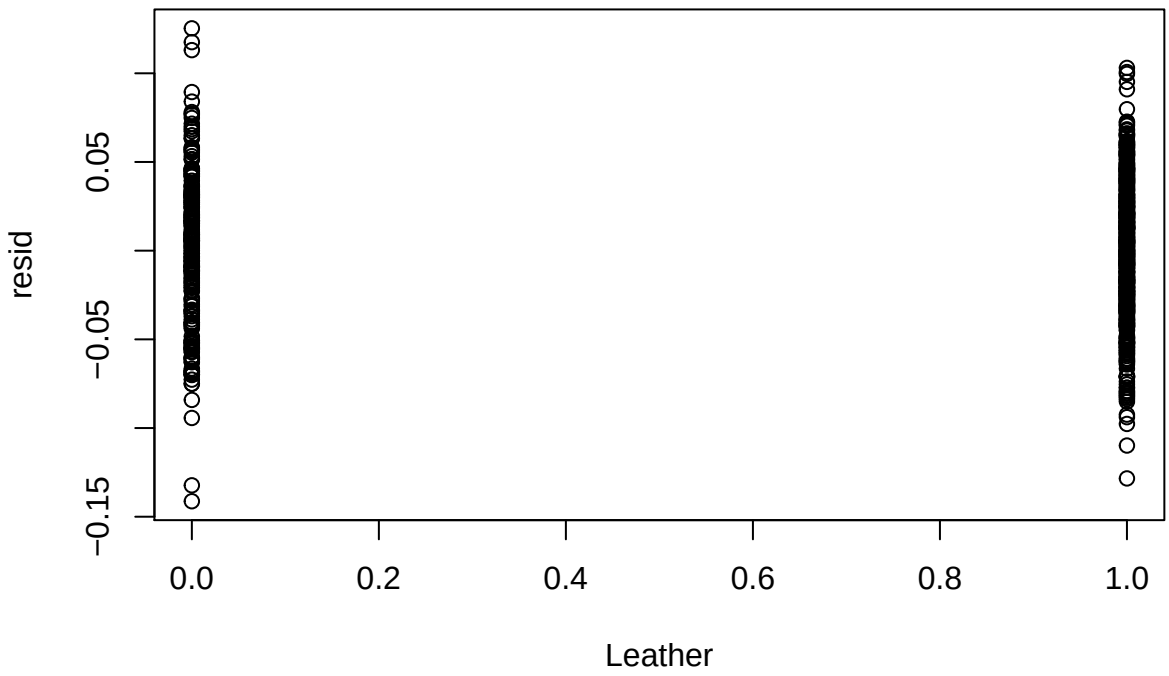

```
plot(resid~Liter+Doors+Cruise+Sound+Leather+Mileage)
```

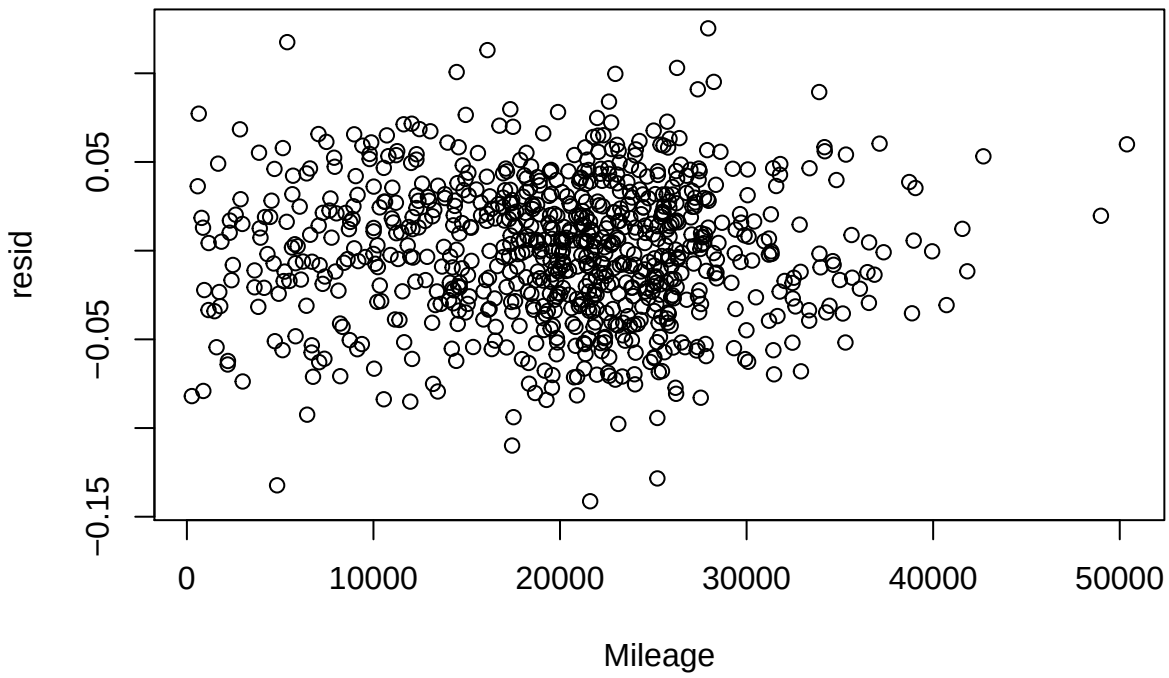

resid
0.05
−0.05
−0.15
2
3
4
5
6
Liter


resid
0.05
−0.05
−0.15
2.0
2.5
3.0
3.5
4.0
Doors

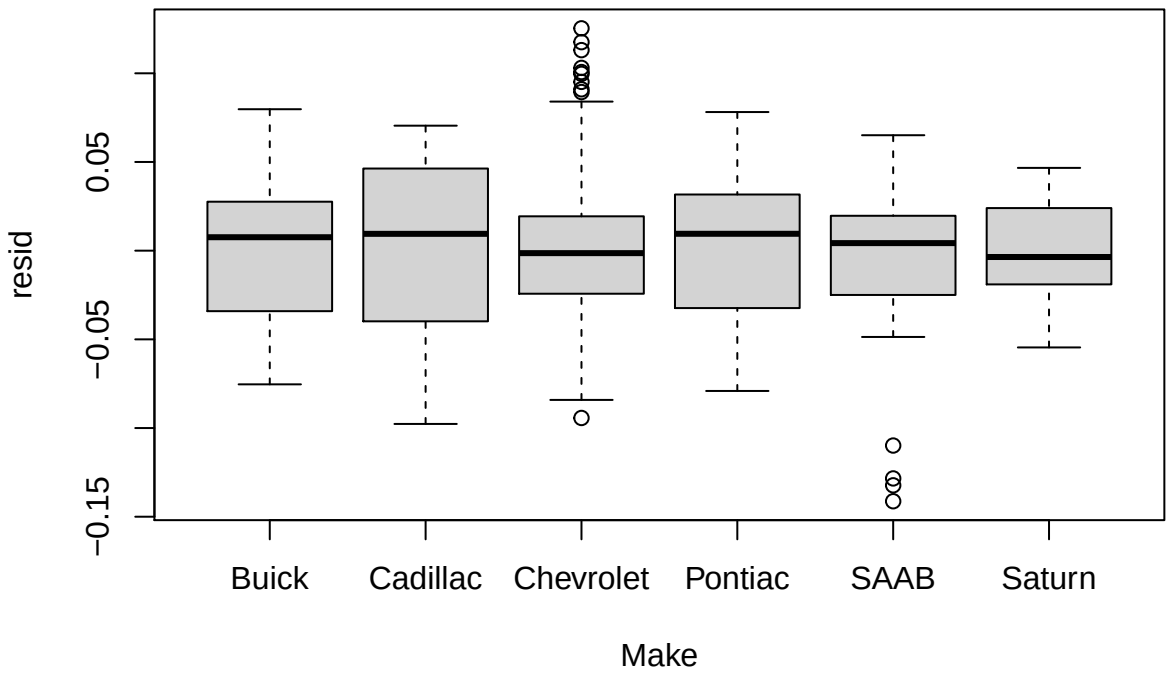

resid
0.05
−0.05
−0.15
0.0
0.2
0.4
0.6
0.8
1.0
Cruise


resid
0.05
−0.05
−0.15
0.0
0.2
0.4
0.6
0.8
1.0
Sound

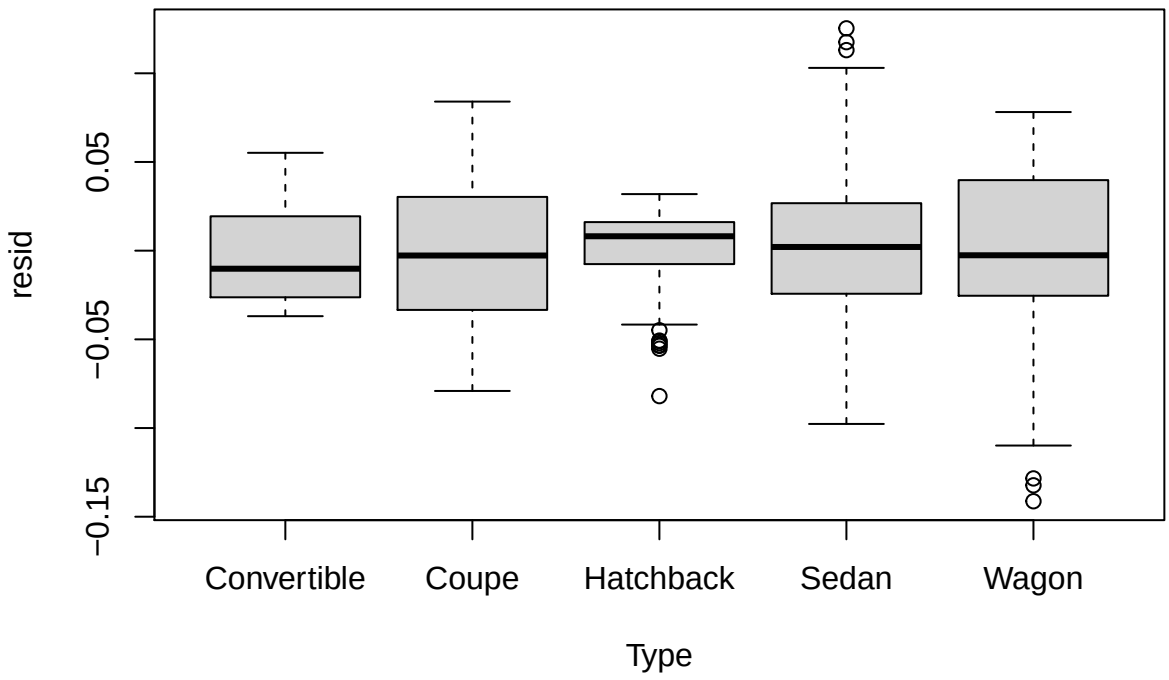

resid
0.05
−0.05
−0.15
0.0
0.2
0.4
0.6
0.8
1.0
Leather




```
boxplot(resid~Make)
```




```
boxplot(resid~Type)
```



Looking at the histogram and the qqplot, we see the residuals are very normally distrubted, except for a couple of outliers. The residuals vs fit and order both seem to have a constant variance and zero mean. For the explanatory variables, residuals against Sound and Leather seem to have mean zero and a constant variance, however every other exaplanatory variable seems to have nonconstant variance.

5) Create a regression model too predict price from Mileage for the Cavalier data. Calculate the total sum of squares (SST), residual sum of squares (SSE), and regression sum of squares (SSR). Verify that SST=SSE+SSR.

```
cav <- read.csv("C3 Cavalier.csv")
model3 <- lm(cav$Price~cav$Mileage)
price_avg <- mean(cav$Price)
yhat <- model3$fitted.values
SSR <- sum((yhat - price_avg)^2)
SSR
```

```
## [1] 29255352
```

```
SSE <- sum((model3$residuals)^2)
SSE
```

```
## [1] 12933122
```

```
SST <- sum((cav$Price - price_avg)^2)
SST
```

```
## [1] 42188474
```

```
SSE+SSR
```

```
## [1] 42188474
```

We calculate SSE by taking the residuals of the model and squaring each value. To calculate the SSR, we take the difference between our predictions and the overall average price and sum the square of those values. SST we calculate by summing over the square difference between the price and the average. Now that we have SSE, SSR and SST, we can verify SST=SSE+SSR, which we see is correct.

6) show that $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$ for the model in the previous question.

```
err <- model3$residuals
reg <- model3$fitted.values - mean(cav$Price)
round(sum(err*reg))
```

```
## [1] 0
```

Since $(y_i - \hat{y}_i)$ is just our residuals, we can use the residuals which are calculated with the lm() function. $(\hat{y}_i - \bar{y})$ is just our fitted values minus the mean, thus we can easily calculate this by using the fitted.values from the lm() funciton and taking the mean of the Price. Then, by summing over err*reg, we find the sum is zero.

7) Consider the Cavalier data set and the regression model

$$y = \beta_0 + \beta_1(Mileage) + \beta_2(Cruise) + \epsilon.$$

Submit the ANOVA table, F-statistic, and p-value to test the hypothesis: $H_o : \beta_1 = \beta_2 = 0$ versus H_a : at least one of the coefficients is not 0.

```
#just talk about the results
model7 <- lm(cav$Price~cav$Mileage + cav$Cruise)
summary(model7)
```

```
##
## Call:
## lm(formula = cav$Price ~ cav$Mileage + cav$Cruise)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1484.27  -429.78    18.81   447.32  1074.22
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15191.6065   313.0957  48.521  < 2e-16 ***
## cav$Mileage    -0.1173     0.0148  -7.927 1.61e-08 ***
## cav$Cruise    317.1039   265.4794   1.194    0.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 674.5 on 27 degrees of freedom
## Multiple R-squared:  0.7088, Adjusted R-squared:  0.6873
## F-statistic: 32.86 on 2 and 27 DF,  p-value: 5.834e-08
```

```
summary(aov(model7))
```

```
##             Df   Sum Sq  Mean Sq F value   Pr(>F)
## cav$Mileage  1 29255352 29255352  64.303 1.29e-08 ***
## cav$Cruise   1   649109   649109   1.427    0.243
## Residuals   27 12284013   454963
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For our multiple linear regression, we obtain an F-value of 32.86 and a p-value of 5.834e-08. Since our p-value is quite low, we can reject the null hypothesis. Thus, we can conclude at least one of the coefficients are not zero.

8) Conduct an extra sum of squares test to determine if Trim is useful. More specifically, use the reduced model in the previous question and the full model

$$y = \beta_0 + \beta_1(Mileage) + \beta_2(Cruise) + \beta_3(LsSportSedan4D) + \beta_4(Sedan4D) + \epsilon.$$

to test the hypothesis $H_0 : \beta_3 = \beta_4 = 0$ versus H_a : at least one of the coefficients is not 0.

```
M8 <- lm(cav$Price~cav$Mileage+cav$Cruise)
summary(M8)
```

```
##
## Call:
## lm(formula = cav$Price ~ cav$Mileage + cav$Cruise)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1484.27  -429.78    18.81   447.32  1074.22
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15191.6065   313.0957  48.521  < 2e-16 ***
## cav$Mileage    -0.1173     0.0148  -7.927 1.61e-08 ***
## cav$Cruise    317.1039   265.4794   1.194    0.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 674.5 on 27 degrees of freedom
## Multiple R-squared:  0.7088, Adjusted R-squared:  0.6873
## F-statistic: 32.86 on 2 and 27 DF,  p-value: 5.834e-08
```

```
M81 <- lm(cav$Price~cav$Mileage+cav$Cruise+cav$Trim)
summary(M81)
```

```
##
## Call:
## lm(formula = cav$Price ~ cav$Mileage + cav$Cruise + cav$Trim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -726.41 -226.61  -66.98  306.69  816.31
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.514e+04  2.077e+02  72.911  < 2e-16 ***
## cav$Mileage                -1.054e-01  8.437e-03 -12.496 3.00e-12 ***
## cav$Cruise                  3.173e+02  1.483e+02   2.139   0.0424 *
## cav$TrimLS Sport Sedan 4D  3.605e+02  1.705e+02   2.114   0.0447 *
## cav$TrimSedan 4D          -9.439e+02  1.689e+02  -5.590 8.17e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 376.7 on 25 degrees of freedom
## Multiple R-squared:  0.9159, Adjusted R-squared:  0.9025
```

```
## F-statistic: 68.08 on 4 and 25 DF,  p-value: 4.509e-13
```

```
anova(M8, M81)
```

```
## Analysis of Variance Table
##
## Model 1: cav$Price ~ cav$Mileage + cav$Cruise
## Model 2: cav$Price ~ cav$Mileage + cav$Cruise + cav$Trim
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1     27 12284013
## 2     25  3547400  2   8736613 30.785 1.808e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From our anova table, we find that the two models are significantly different from one another. We conclude this by looking at our p-value which is quite small. Since the p-value is quite small, we can reject the null hypothesis, indicating that our Trim variable is significant in predicting the price. We also see a big jump in R^2 when we look at the two models side by side.