***ALY-6015 Week 6 Project***

***Week 6: Regression and Datamining***

**Assignment Title: ALY-6015 Week 6 Project**

***Regression and Data Mining***

**ALY6015 80877 Intermediate Analytics SEC 07 Spring 2019 CPS [VTL-A-OL] ALY6015.80877.201935**

**Submission Date: 05/19/2019**

**Student: Srinivasu Narahari**

**Professor: Steward Huang**

**Introduction**:

**Regression**: Regression analysis is set of the process used for determining the relation ship between two variables.

**Association**: By thus we find pattern of similar objects in a in a single transaction

**Classification**: In this method classify each item of data into a pre-defined class.

**Clustering**: The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

**Prediction:** Prediction is one pf the data mining technique in this process we discover the relating ships between independent variable and dependent and independent variable.

**Sequential Pattern:** By using this data mining technique we can find similar patterns or trends or similar events in a transaction data over a business period.

**Decision Trees :** In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it

Collaborative\_Group\_Project.R

narah

2019-05-19

car\_info <- read.csv(url("http://mlr.cs.umass.edu/ml/machine-learning-databases/auto-mpg/auto-mpg.names"), header=FALSE)  
car\_Info\_dataset = read.csv(url("http://mlr.cs.umass.edu/ml/machine-learning-databases/autos/imports-85.data"), header=FALSE)  
  
library(party)

library(dbplyr)  
library(dplyr)

car\_Info\_dataset\_scrubbed = select(car\_Info\_dataset, -contains("?"))  
  
#View(car\_Info\_dataset\_scrubbed)  
car\_Info\_dataset\_scrubbed = na.omit(car\_Info\_dataset\_scrubbed)  
plot(car\_Info\_dataset\_scrubbed$V1, car\_Info\_dataset\_scrubbed$V2)# + car\_Info\_dataset\_scrubbed$V3 + car\_Info\_dataset\_scrubbed$V4 + car\_Info\_dataset\_scrubbed$V5 + car\_Info\_dataset\_scrubbed$V6 + car\_Info\_dataset\_scrubbed$V7)  
  
model\_car\_Info\_dataset = lm(car\_Info\_dataset\_scrubbed$V1 ~ car\_Info\_dataset\_scrubbed$V2 + car\_Info\_dataset\_scrubbed$V3 + car\_Info\_dataset\_scrubbed$V4 + car\_Info\_dataset\_scrubbed$V5 + car\_Info\_dataset\_scrubbed$V6 + car\_Info\_dataset\_scrubbed$V7)

By using linear mode predicted the MPG miles per gallon of different cars based on the car size, make. model etc.

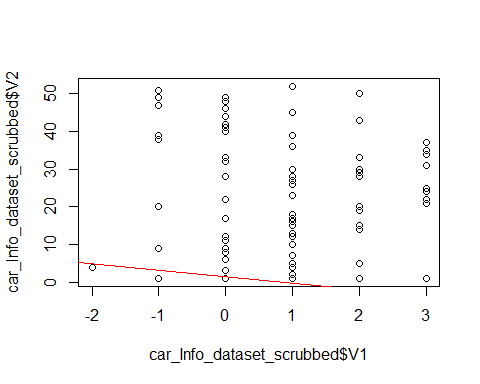
In the following summary we can see intercept as 1.44785 slope as 0.57539 and error value 2.516.

summary(model\_car\_Info\_dataset)

##   
## Call:  
## lm(formula = car\_Info\_dataset\_scrubbed$V1 ~ car\_Info\_dataset\_scrubbed$V2 +   
## car\_Info\_dataset\_scrubbed$V3 + car\_Info\_dataset\_scrubbed$V4 +   
## car\_Info\_dataset\_scrubbed$V5 + car\_Info\_dataset\_scrubbed$V6 +   
## car\_Info\_dataset\_scrubbed$V7)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.51775 -0.06648 0.00000 0.11599 1.28351   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 1.44785 0.57539 2.516  
## car\_Info\_dataset\_scrubbed$V2101 -1.71806 0.48936 -3.511  
## car\_Info\_dataset\_scrubbed$V2102 -1.95466 0.73304 -2.666  
## car\_Info\_dataset\_scrubbed$V2103 -0.78592 0.40967 -1.918  
## car\_Info\_dataset\_scrubbed$V2104 0.13587 0.46335 0.293  
## car\_Info\_dataset\_scrubbed$V2106 -2.23546 0.40716 -5.490  
## car\_Info\_dataset\_scrubbed$V2107 -1.65706 0.65189 -2.542  
## car\_Info\_dataset\_scrubbed$V2108 -1.55718 0.48977 -3.179  
## car\_Info\_dataset\_scrubbed$V2110 -1.52033 0.52901 -2.874  
## car\_Info\_dataset\_scrubbed$V2113 0.66249 0.51347 1.290  
## car\_Info\_dataset\_scrubbed$V2115 -0.35784 0.47474 -0.754  
## car\_Info\_dataset\_scrubbed$V2118 -0.70674 0.45465 -1.554  
## car\_Info\_dataset\_scrubbed$V2119 -0.69645 0.52364 -1.330  
## car\_Info\_dataset\_scrubbed$V2121 0.82449 0.92740 0.889  
## car\_Info\_dataset\_scrubbed$V2122 -0.14488 0.36751 -0.394  
## car\_Info\_dataset\_scrubbed$V2125 -0.35320 0.40202 -0.879  
## car\_Info\_dataset\_scrubbed$V2128 -1.73762 0.45483 -3.820  
## car\_Info\_dataset\_scrubbed$V2129 -0.41926 0.53476 -0.784  
## car\_Info\_dataset\_scrubbed$V2134 0.10073 0.45642 0.221  
## car\_Info\_dataset\_scrubbed$V2137 -1.25866 0.42010 -2.996  
## car\_Info\_dataset\_scrubbed$V2142 1.31469 0.62502 2.103  
## car\_Info\_dataset\_scrubbed$V2145 0.89023 0.46359 1.920  
## car\_Info\_dataset\_scrubbed$V2148 0.27991 0.54463 0.514  
## car\_Info\_dataset\_scrubbed$V2150 1.16439 0.47919 2.430  
## car\_Info\_dataset\_scrubbed$V2153 0.55837 0.40710 1.372  
## car\_Info\_dataset\_scrubbed$V2154 0.38500 0.50147 0.768  
## car\_Info\_dataset\_scrubbed$V2158 0.58403 0.45821 1.275  
## car\_Info\_dataset\_scrubbed$V2161 -0.54816 0.25579 -2.143  
## car\_Info\_dataset\_scrubbed$V2164 1.60407 0.46014 3.486  
## car\_Info\_dataset\_scrubbed$V2168 -0.62489 0.42279 -1.478  
## car\_Info\_dataset\_scrubbed$V2186 0.98538 0.56109 1.756  
## car\_Info\_dataset\_scrubbed$V2188 -0.50519 0.41756 -1.210  
## car\_Info\_dataset\_scrubbed$V2192 0.49481 0.41756 1.185  
## car\_Info\_dataset\_scrubbed$V2194 0.17586 0.51780 0.340  
## car\_Info\_dataset\_scrubbed$V2197 1.49413 0.51205 2.918  
## car\_Info\_dataset\_scrubbed$V2231 -1.80411 0.62339 -2.894  
## car\_Info\_dataset\_scrubbed$V2256 1.33503 0.55282 2.415  
## car\_Info\_dataset\_scrubbed$V265 -1.44191 0.42188 -3.418  
## car\_Info\_dataset\_scrubbed$V274 0.01921 0.40791 0.047  
## car\_Info\_dataset\_scrubbed$V277 -0.09376 0.61223 -0.153  
## car\_Info\_dataset\_scrubbed$V278 -1.30595 0.66510 -1.964  
## car\_Info\_dataset\_scrubbed$V281 -0.09376 0.61223 -0.153  
## car\_Info\_dataset\_scrubbed$V283 -1.02839 0.70942 -1.450  
## car\_Info\_dataset\_scrubbed$V285 -1.63631 0.54375 -3.009  
## car\_Info\_dataset\_scrubbed$V287 -0.50587 0.51205 -0.988  
## car\_Info\_dataset\_scrubbed$V289 -1.61628 0.73265 -2.206  
## car\_Info\_dataset\_scrubbed$V290 -1.42412 0.60634 -2.349  
## car\_Info\_dataset\_scrubbed$V291 -0.38742 0.40215 -0.963  
## car\_Info\_dataset\_scrubbed$V293 -0.50929 0.40897 -1.245  
## car\_Info\_dataset\_scrubbed$V294 1.43802 0.33064 4.349  
## car\_Info\_dataset\_scrubbed$V295 -0.04459 0.46785 -0.095  
## car\_Info\_dataset\_scrubbed$V298 -0.17551 0.92740 -0.189  
## car\_Info\_dataset\_scrubbed$V3audi -0.11981 0.44232 -0.271  
## car\_Info\_dataset\_scrubbed$V3bmw -0.52092 0.43059 -1.210  
## car\_Info\_dataset\_scrubbed$V3chevrolet -0.42198 0.85929 -0.491  
## car\_Info\_dataset\_scrubbed$V3dodge 0.21897 0.55940 0.391  
## car\_Info\_dataset\_scrubbed$V3honda 1.12058 0.52146 2.149  
## car\_Info\_dataset\_scrubbed$V3isuzu -0.29136 0.42068 -0.693  
## car\_Info\_dataset\_scrubbed$V3jaguar -1.15273 0.47163 -2.444  
## car\_Info\_dataset\_scrubbed$V3mazda -0.17823 0.51999 -0.343  
## car\_Info\_dataset\_scrubbed$V3mercedes-benz -1.01595 0.44657 -2.275  
## car\_Info\_dataset\_scrubbed$V3mercury -0.63756 0.59937 -1.064  
## car\_Info\_dataset\_scrubbed$V3mitsubishi 0.82411 0.41144 2.003  
## car\_Info\_dataset\_scrubbed$V3nissan 1.20662 0.49255 2.450  
## car\_Info\_dataset\_scrubbed$V3peugot -0.03539 0.43862 -0.081  
## car\_Info\_dataset\_scrubbed$V3plymouth 0.07893 0.52259 0.151  
## car\_Info\_dataset\_scrubbed$V3porsche 0.41713 0.39332 1.061  
## car\_Info\_dataset\_scrubbed$V3renault 0.10857 0.48390 0.224  
## car\_Info\_dataset\_scrubbed$V3saab 0.77990 0.60435 1.290  
## car\_Info\_dataset\_scrubbed$V3subaru 1.43090 0.72908 1.963  
## car\_Info\_dataset\_scrubbed$V3toyota -0.09162 0.50238 -0.182  
## car\_Info\_dataset\_scrubbed$V3volkswagen 0.06748 0.39107 0.173  
## car\_Info\_dataset\_scrubbed$V3volvo -1.48056 0.54302 -2.727  
## car\_Info\_dataset\_scrubbed$V4gas 0.07313 0.16819 0.435  
## car\_Info\_dataset\_scrubbed$V5turbo 0.04007 0.12844 0.312  
## car\_Info\_dataset\_scrubbed$V6four 0.15953 0.42150 0.378  
## car\_Info\_dataset\_scrubbed$V6two 1.18027 0.45842 2.575  
## car\_Info\_dataset\_scrubbed$V7hardtop -0.68489 0.31781 -2.155  
## car\_Info\_dataset\_scrubbed$V7hatchback -1.10377 0.29530 -3.738  
## car\_Info\_dataset\_scrubbed$V7sedan -1.16477 0.31857 -3.656  
## car\_Info\_dataset\_scrubbed$V7wagon -1.49514 0.34338 -4.354  
## Pr(>|t|)   
## (Intercept) 0.013136 \*   
## car\_Info\_dataset\_scrubbed$V2101 0.000624 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2102 0.008687 \*\*   
## car\_Info\_dataset\_scrubbed$V2103 0.057359 .   
## car\_Info\_dataset\_scrubbed$V2104 0.769825   
## car\_Info\_dataset\_scrubbed$V2106 2.17e-07 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2107 0.012256 \*   
## car\_Info\_dataset\_scrubbed$V2108 0.001864 \*\*   
## car\_Info\_dataset\_scrubbed$V2110 0.004772 \*\*   
## car\_Info\_dataset\_scrubbed$V2113 0.199378   
## car\_Info\_dataset\_scrubbed$V2115 0.452412   
## car\_Info\_dataset\_scrubbed$V2118 0.122627   
## car\_Info\_dataset\_scrubbed$V2119 0.185949   
## car\_Info\_dataset\_scrubbed$V2121 0.375707   
## car\_Info\_dataset\_scrubbed$V2122 0.694101   
## car\_Info\_dataset\_scrubbed$V2125 0.381338   
## car\_Info\_dataset\_scrubbed$V2128 0.000210 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2129 0.434531   
## car\_Info\_dataset\_scrubbed$V2134 0.825692   
## car\_Info\_dataset\_scrubbed$V2137 0.003303 \*\*   
## car\_Info\_dataset\_scrubbed$V2142 0.037447 \*   
## car\_Info\_dataset\_scrubbed$V2145 0.057116 .   
## car\_Info\_dataset\_scrubbed$V2148 0.608209   
## car\_Info\_dataset\_scrubbed$V2150 0.016536 \*   
## car\_Info\_dataset\_scrubbed$V2153 0.172670   
## car\_Info\_dataset\_scrubbed$V2154 0.444099   
## car\_Info\_dataset\_scrubbed$V2158 0.204831   
## car\_Info\_dataset\_scrubbed$V2161 0.034065 \*   
## car\_Info\_dataset\_scrubbed$V2164 0.000679 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2168 0.141937   
## car\_Info\_dataset\_scrubbed$V2186 0.081526 .   
## car\_Info\_dataset\_scrubbed$V2188 0.228633   
## car\_Info\_dataset\_scrubbed$V2192 0.238275   
## car\_Info\_dataset\_scrubbed$V2194 0.734710   
## car\_Info\_dataset\_scrubbed$V2197 0.004185 \*\*   
## car\_Info\_dataset\_scrubbed$V2231 0.004495 \*\*   
## car\_Info\_dataset\_scrubbed$V2256 0.017196 \*   
## car\_Info\_dataset\_scrubbed$V265 0.000855 \*\*\*  
## car\_Info\_dataset\_scrubbed$V274 0.962515   
## car\_Info\_dataset\_scrubbed$V277 0.878535   
## car\_Info\_dataset\_scrubbed$V278 0.051821 .   
## car\_Info\_dataset\_scrubbed$V281 0.878535   
## car\_Info\_dataset\_scrubbed$V283 0.149689   
## car\_Info\_dataset\_scrubbed$V285 0.003172 \*\*   
## car\_Info\_dataset\_scrubbed$V287 0.325110   
## car\_Info\_dataset\_scrubbed$V289 0.029222 \*   
## car\_Info\_dataset\_scrubbed$V290 0.020420 \*   
## car\_Info\_dataset\_scrubbed$V291 0.337235   
## car\_Info\_dataset\_scrubbed$V293 0.215360   
## car\_Info\_dataset\_scrubbed$V294 2.82e-05 \*\*\*  
## car\_Info\_dataset\_scrubbed$V295 0.924216   
## car\_Info\_dataset\_scrubbed$V298 0.850209   
## car\_Info\_dataset\_scrubbed$V3audi 0.786952   
## car\_Info\_dataset\_scrubbed$V3bmw 0.228657   
## car\_Info\_dataset\_scrubbed$V3chevrolet 0.624238   
## car\_Info\_dataset\_scrubbed$V3dodge 0.696151   
## car\_Info\_dataset\_scrubbed$V3honda 0.033584 \*   
## car\_Info\_dataset\_scrubbed$V3isuzu 0.489860   
## car\_Info\_dataset\_scrubbed$V3jaguar 0.015926 \*   
## car\_Info\_dataset\_scrubbed$V3mazda 0.732362   
## car\_Info\_dataset\_scrubbed$V3mercedes-benz 0.024624 \*   
## car\_Info\_dataset\_scrubbed$V3mercury 0.289530   
## car\_Info\_dataset\_scrubbed$V3mitsubishi 0.047362 \*   
## car\_Info\_dataset\_scrubbed$V3nissan 0.015693 \*   
## car\_Info\_dataset\_scrubbed$V3peugot 0.935820   
## car\_Info\_dataset\_scrubbed$V3plymouth 0.880194   
## car\_Info\_dataset\_scrubbed$V3porsche 0.290957   
## car\_Info\_dataset\_scrubbed$V3renault 0.822846   
## car\_Info\_dataset\_scrubbed$V3saab 0.199287   
## car\_Info\_dataset\_scrubbed$V3subaru 0.051929 .   
## car\_Info\_dataset\_scrubbed$V3toyota 0.855592   
## car\_Info\_dataset\_scrubbed$V3volkswagen 0.863287   
## car\_Info\_dataset\_scrubbed$V3volvo 0.007328 \*\*   
## car\_Info\_dataset\_scrubbed$V4gas 0.664451   
## car\_Info\_dataset\_scrubbed$V5turbo 0.755604   
## car\_Info\_dataset\_scrubbed$V6four 0.705722   
## car\_Info\_dataset\_scrubbed$V6two 0.011208 \*   
## car\_Info\_dataset\_scrubbed$V7hardtop 0.033093 \*   
## car\_Info\_dataset\_scrubbed$V7hatchback 0.000282 \*\*\*  
## car\_Info\_dataset\_scrubbed$V7sedan 0.000377 \*\*\*  
## car\_Info\_dataset\_scrubbed$V7wagon 2.77e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4793 on 124 degrees of freedom  
## Multiple R-squared: 0.9099, Adjusted R-squared: 0.8518   
## F-statistic: 15.66 on 80 and 124 DF, p-value: < 2.2e-16

abline(model\_car\_Info\_dataset,col = "red")

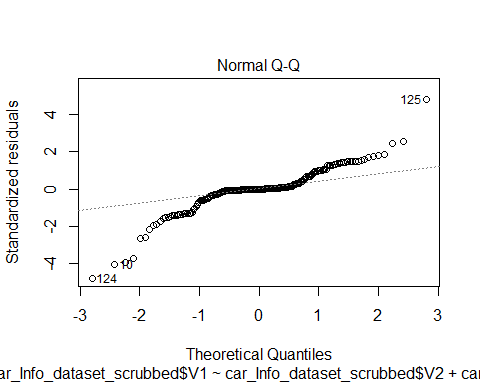
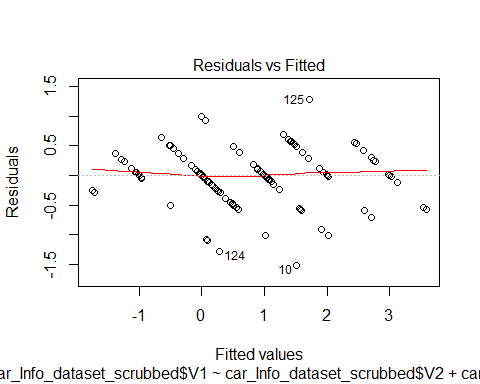
## Warning in abline(model\_car\_Info\_dataset, col = "red"): only using the  
## first two of 81 regression coefficients



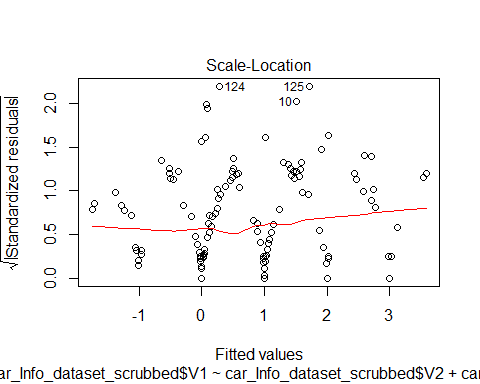
In the above plot redline is the prediction line. The fitted line is covering only few points. We can say linear model is not enough for this data. The curve is under fitting. Polynomial curve can fit all the points.

plot(model\_car\_Info\_dataset)

## Warning: not plotting observations with leverage one:  
## 37, 43, 76, 107, 126, 154, 155, 181, 191

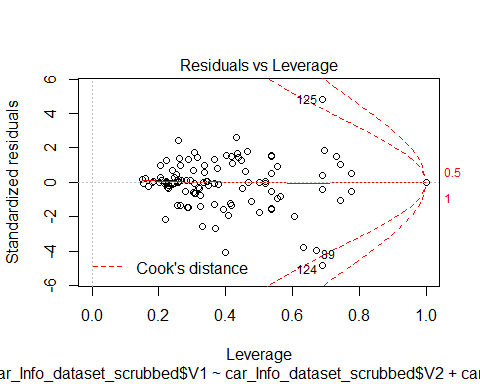


## Warning: not plotting observations with leverage one:  
## 37, 43, 76, 107, 126, 154, 155, 181, 191



## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

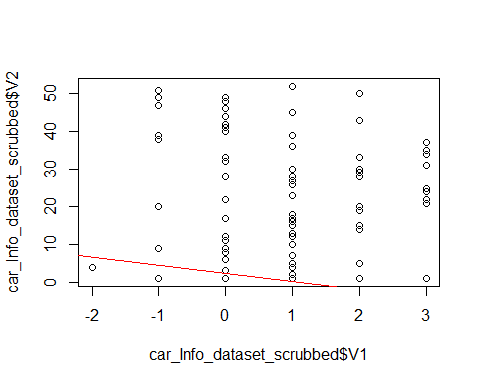


cis = car\_Info\_dataset\_scrubbed  
plot(car\_Info\_dataset\_scrubbed$V1,car\_Info\_dataset\_scrubbed$V2)  
#meanCarMPG = mean(car\_Info\_dataset\_scrubbed$V1)  
#meanCarMP  
#abline(h=meanCarMPG)  
model\_car\_Info\_dataset\_2 = lm(car\_Info\_dataset\_scrubbed$V1 ~ car\_Info\_dataset\_scrubbed$V2 + car\_Info\_dataset\_scrubbed$V3)  
summary(model\_car\_Info\_dataset\_2)

##   
## Call:  
## lm(formula = car\_Info\_dataset\_scrubbed$V1 ~ car\_Info\_dataset\_scrubbed$V2 +   
## car\_Info\_dataset\_scrubbed$V3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.7347 0.0000 0.0000 0.1136 1.8393   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 2.333e+00 3.483e-01 6.699  
## car\_Info\_dataset\_scrubbed$V2101 -2.131e+00 6.006e-01 -3.548  
## car\_Info\_dataset\_scrubbed$V2102 -3.131e+00 7.845e-01 -3.991  
## car\_Info\_dataset\_scrubbed$V2103 -1.115e+00 5.056e-01 -2.206  
## car\_Info\_dataset\_scrubbed$V2104 1.009e+00 4.921e-01 2.050  
## car\_Info\_dataset\_scrubbed$V2106 -2.467e+00 4.916e-01 -5.019  
## car\_Info\_dataset\_scrubbed$V2107 -2.131e+00 7.768e-01 -2.743  
## car\_Info\_dataset\_scrubbed$V2108 -1.804e+00 5.995e-01 -3.008  
## car\_Info\_dataset\_scrubbed$V2110 -3.038e+00 5.944e-01 -5.110  
## car\_Info\_dataset\_scrubbed$V2113 7.424e-01 5.790e-01 1.282  
## car\_Info\_dataset\_scrubbed$V2115 -2.576e-01 5.240e-01 -0.492  
## car\_Info\_dataset\_scrubbed$V2118 -7.729e-01 5.357e-01 -1.443  
## car\_Info\_dataset\_scrubbed$V2119 -2.653e-01 6.511e-01 -0.407  
## car\_Info\_dataset\_scrubbed$V2121 1.581e+00 1.135e+00 1.393  
## car\_Info\_dataset\_scrubbed$V2122 1.786e-02 4.160e-01 0.043  
## car\_Info\_dataset\_scrubbed$V2125 -1.261e+00 4.415e-01 -2.857  
## car\_Info\_dataset\_scrubbed$V2128 -9.702e-01 4.879e-01 -1.989  
## car\_Info\_dataset\_scrubbed$V2129 7.424e-01 5.790e-01 1.282  
## car\_Info\_dataset\_scrubbed$V2134 1.581e+00 5.064e-01 3.122  
## car\_Info\_dataset\_scrubbed$V2137 -1.841e+00 4.995e-01 -3.686  
## car\_Info\_dataset\_scrubbed$V2142 3.000e+00 6.966e-01 4.306  
## car\_Info\_dataset\_scrubbed$V2145 6.332e-01 5.651e-01 1.121  
## car\_Info\_dataset\_scrubbed$V2148 -9.446e-01 6.095e-01 -1.550  
## car\_Info\_dataset\_scrubbed$V2150 2.542e+00 4.754e-01 5.348  
## car\_Info\_dataset\_scrubbed$V2153 7.386e-01 5.055e-01 1.461  
## car\_Info\_dataset\_scrubbed$V2154 -2.653e-01 6.027e-01 -0.440  
## car\_Info\_dataset\_scrubbed$V2158 -8.742e-15 5.507e-01 0.000  
## car\_Info\_dataset\_scrubbed$V2161 -3.125e-01 2.883e-01 -1.084  
## car\_Info\_dataset\_scrubbed$V2164 1.000e+00 5.507e-01 1.816  
## car\_Info\_dataset\_scrubbed$V2168 5.040e-01 4.827e-01 1.044  
## car\_Info\_dataset\_scrubbed$V2186 5.000e-01 6.745e-01 0.741  
## car\_Info\_dataset\_scrubbed$V2188 -2.500e-01 5.225e-01 -0.478  
## car\_Info\_dataset\_scrubbed$V2192 7.500e-01 5.225e-01 1.435  
## car\_Info\_dataset\_scrubbed$V2194 1.196e+00 5.995e-01 1.996  
## car\_Info\_dataset\_scrubbed$V2197 2.581e+00 6.146e-01 4.199  
## car\_Info\_dataset\_scrubbed$V2231 -8.036e-01 7.358e-01 -1.092  
## car\_Info\_dataset\_scrubbed$V2256 1.839e+00 6.662e-01 2.761  
## car\_Info\_dataset\_scrubbed$V265 -1.419e+00 5.182e-01 -2.738  
## car\_Info\_dataset\_scrubbed$V274 -5.306e-01 4.898e-01 -1.083  
## car\_Info\_dataset\_scrubbed$V277 -4.191e-01 7.481e-01 -0.560  
## car\_Info\_dataset\_scrubbed$V278 -3.131e+00 7.768e-01 -4.031  
## car\_Info\_dataset\_scrubbed$V281 -4.191e-01 7.481e-01 -0.560  
## car\_Info\_dataset\_scrubbed$V283 -1.131e+00 8.149e-01 -1.388  
## car\_Info\_dataset\_scrubbed$V285 -3.131e+00 6.006e-01 -5.213  
## car\_Info\_dataset\_scrubbed$V287 5.809e-01 6.146e-01 0.945  
## car\_Info\_dataset\_scrubbed$V289 -3.131e+00 8.513e-01 -3.678  
## car\_Info\_dataset\_scrubbed$V290 -1.419e+00 7.481e-01 -1.897  
## car\_Info\_dataset\_scrubbed$V291 -4.191e-01 4.912e-01 -0.853  
## car\_Info\_dataset\_scrubbed$V293 -7.500e-01 4.608e-01 -1.628  
## car\_Info\_dataset\_scrubbed$V294 8.393e-01 3.907e-01 2.148  
## car\_Info\_dataset\_scrubbed$V295 -3.229e-01 5.811e-01 -0.556  
## car\_Info\_dataset\_scrubbed$V298 5.809e-01 1.135e+00 0.512  
## car\_Info\_dataset\_scrubbed$V3audi -1.333e+00 4.926e-01 -2.707  
## car\_Info\_dataset\_scrubbed$V3bmw -2.083e+00 4.608e-01 -4.521  
## car\_Info\_dataset\_scrubbed$V3chevrolet -1.914e+00 1.022e+00 -1.873  
## car\_Info\_dataset\_scrubbed$V3dodge -3.887e-01 6.339e-01 -0.613  
## car\_Info\_dataset\_scrubbed$V3honda 7.975e-01 6.006e-01 1.328  
## car\_Info\_dataset\_scrubbed$V3isuzu -1.583e+00 4.608e-01 -3.436  
## car\_Info\_dataset\_scrubbed$V3jaguar -2.544e+00 5.274e-01 -4.825  
## car\_Info\_dataset\_scrubbed$V3mazda -2.076e+00 5.240e-01 -3.962  
## car\_Info\_dataset\_scrubbed$V3mercedes-benz -2.333e+00 4.926e-01 -4.737  
## car\_Info\_dataset\_scrubbed$V3mercury -1.333e+00 6.966e-01 -1.914  
## car\_Info\_dataset\_scrubbed$V3mitsubishi -7.195e-02 4.415e-01 -0.163  
## car\_Info\_dataset\_scrubbed$V3nissan -5.298e-01 5.466e-01 -0.969  
## car\_Info\_dataset\_scrubbed$V3peugot -2.134e+00 4.337e-01 -4.922  
## car\_Info\_dataset\_scrubbed$V3plymouth -1.068e+00 6.027e-01 -1.772  
## car\_Info\_dataset\_scrubbed$V3porsche 1.667e-01 4.608e-01 0.362  
## car\_Info\_dataset\_scrubbed$V3renault -1.333e+00 5.507e-01 -2.421  
## car\_Info\_dataset\_scrubbed$V3saab -1.609e+00 6.227e-01 -2.584  
## car\_Info\_dataset\_scrubbed$V3subaru 7.975e-01 8.149e-01 0.979  
## car\_Info\_dataset\_scrubbed$V3toyota -1.914e+00 5.631e-01 -3.399  
## car\_Info\_dataset\_scrubbed$V3volkswagen -1.173e+00 4.486e-01 -2.614  
## car\_Info\_dataset\_scrubbed$V3volvo -3.010e+00 6.215e-01 -4.844  
## Pr(>|t|)   
## (Intercept) 5.51e-10 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2101 0.000538 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2102 0.000109 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2103 0.029144 \*   
## car\_Info\_dataset\_scrubbed$V2104 0.042304 \*   
## car\_Info\_dataset\_scrubbed$V2106 1.65e-06 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2107 0.006930 \*\*   
## car\_Info\_dataset\_scrubbed$V2108 0.003145 \*\*   
## car\_Info\_dataset\_scrubbed$V2110 1.10e-06 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2113 0.202000   
## car\_Info\_dataset\_scrubbed$V2115 0.623754   
## car\_Info\_dataset\_scrubbed$V2118 0.151479   
## car\_Info\_dataset\_scrubbed$V2119 0.684324   
## car\_Info\_dataset\_scrubbed$V2121 0.165909   
## car\_Info\_dataset\_scrubbed$V2122 0.965828   
## car\_Info\_dataset\_scrubbed$V2125 0.004969 \*\*   
## car\_Info\_dataset\_scrubbed$V2128 0.048824 \*   
## car\_Info\_dataset\_scrubbed$V2129 0.202000   
## car\_Info\_dataset\_scrubbed$V2134 0.002208 \*\*   
## car\_Info\_dataset\_scrubbed$V2137 0.000332 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2142 3.21e-05 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2145 0.264515   
## car\_Info\_dataset\_scrubbed$V2148 0.123547   
## car\_Info\_dataset\_scrubbed$V2150 3.81e-07 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2153 0.146387   
## car\_Info\_dataset\_scrubbed$V2154 0.660530   
## car\_Info\_dataset\_scrubbed$V2158 1.000000   
## car\_Info\_dataset\_scrubbed$V2161 0.280275   
## car\_Info\_dataset\_scrubbed$V2164 0.071676 .   
## car\_Info\_dataset\_scrubbed$V2168 0.298310   
## car\_Info\_dataset\_scrubbed$V2186 0.459838   
## car\_Info\_dataset\_scrubbed$V2188 0.633088   
## car\_Info\_dataset\_scrubbed$V2192 0.153513   
## car\_Info\_dataset\_scrubbed$V2194 0.048021 \*   
## car\_Info\_dataset\_scrubbed$V2197 4.89e-05 \*\*\*  
## car\_Info\_dataset\_scrubbed$V2231 0.276763   
## car\_Info\_dataset\_scrubbed$V2256 0.006588 \*\*   
## car\_Info\_dataset\_scrubbed$V265 0.007027 \*\*   
## car\_Info\_dataset\_scrubbed$V274 0.280647   
## car\_Info\_dataset\_scrubbed$V277 0.576307   
## car\_Info\_dataset\_scrubbed$V278 9.35e-05 \*\*\*  
## car\_Info\_dataset\_scrubbed$V281 0.576307   
## car\_Info\_dataset\_scrubbed$V283 0.167556   
## car\_Info\_dataset\_scrubbed$V285 6.99e-07 \*\*\*  
## car\_Info\_dataset\_scrubbed$V287 0.346304   
## car\_Info\_dataset\_scrubbed$V289 0.000341 \*\*\*  
## car\_Info\_dataset\_scrubbed$V290 0.060038 .   
## car\_Info\_dataset\_scrubbed$V291 0.395067   
## car\_Info\_dataset\_scrubbed$V293 0.105975   
## car\_Info\_dataset\_scrubbed$V294 0.033548 \*   
## car\_Info\_dataset\_scrubbed$V295 0.579414   
## car\_Info\_dataset\_scrubbed$V298 0.609561   
## car\_Info\_dataset\_scrubbed$V3audi 0.007692 \*\*   
## car\_Info\_dataset\_scrubbed$V3bmw 1.35e-05 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3chevrolet 0.063342 .   
## car\_Info\_dataset\_scrubbed$V3dodge 0.540771   
## car\_Info\_dataset\_scrubbed$V3honda 0.186522   
## car\_Info\_dataset\_scrubbed$V3isuzu 0.000789 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3jaguar 3.81e-06 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3mazda 0.000121 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3mercedes-benz 5.53e-06 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3mercury 0.057786 .   
## car\_Info\_dataset\_scrubbed$V3mitsubishi 0.870792   
## car\_Info\_dataset\_scrubbed$V3nissan 0.334189   
## car\_Info\_dataset\_scrubbed$V3peugot 2.51e-06 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3plymouth 0.078705 .   
## car\_Info\_dataset\_scrubbed$V3porsche 0.718147   
## car\_Info\_dataset\_scrubbed$V3renault 0.016836 \*   
## car\_Info\_dataset\_scrubbed$V3saab 0.010855 \*   
## car\_Info\_dataset\_scrubbed$V3subaru 0.329530   
## car\_Info\_dataset\_scrubbed$V3toyota 0.000893 \*\*\*  
## car\_Info\_dataset\_scrubbed$V3volkswagen 0.009984 \*\*   
## car\_Info\_dataset\_scrubbed$V3volvo 3.51e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6033 on 132 degrees of freedom  
## Multiple R-squared: 0.8481, Adjusted R-squared: 0.7653   
## F-statistic: 10.24 on 72 and 132 DF, p-value: < 2.2e-16

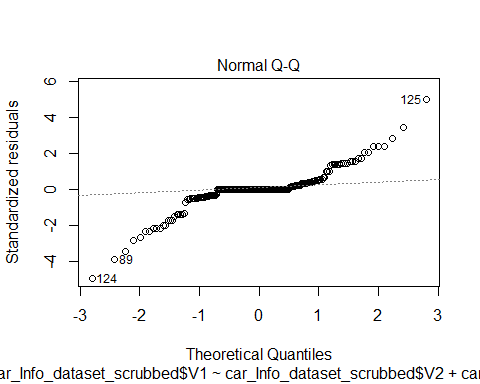
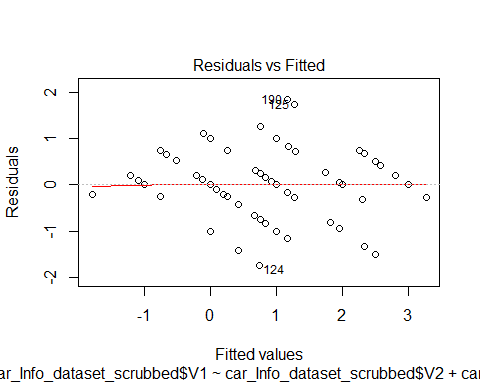
abline(model\_car\_Info\_dataset\_2,col = "red")

## Warning in abline(model\_car\_Info\_dataset\_2, col = "red"): only using the  
## first two of 73 regression coefficients

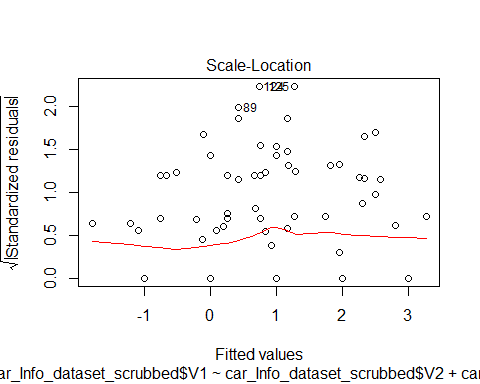


plot(model\_car\_Info\_dataset\_2)

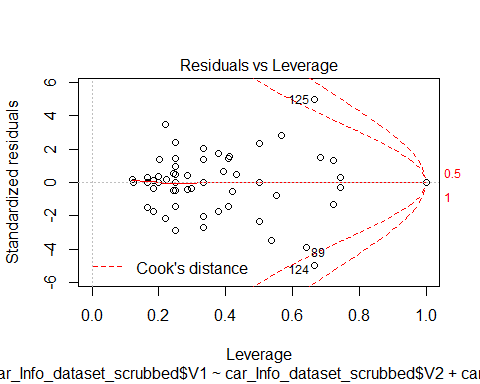
## Warning: not plotting observations with leverage one:  
## 37, 43, 76, 107, 126, 154, 155, 181, 191



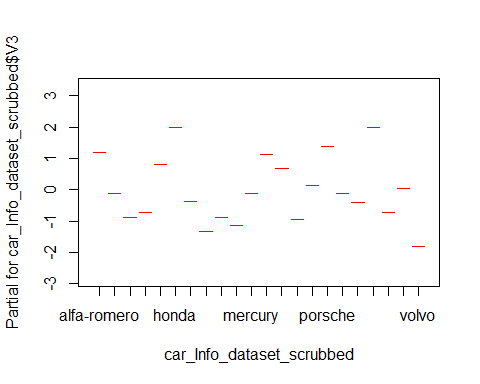
## Warning: not plotting observations with leverage one:  
## 37, 43, 76, 107, 126, 154, 155, 181, 191



## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced  
  
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



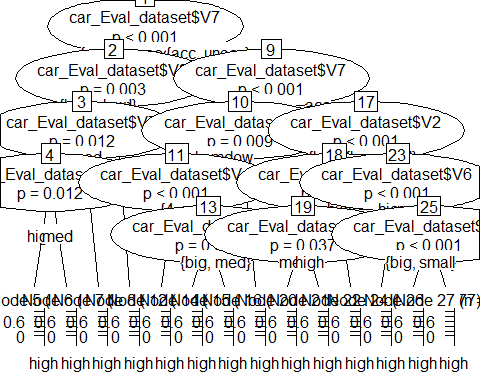
termplot(model\_car\_Info\_dataset\_2)



###http://mlr.cs.umass.edu/ml/machine-learning-databases/car/car.data  
car\_Eval\_dataset = read.csv(url("http://mlr.cs.umass.edu/ml/machine-learning-databases/car/car.data"), header=FALSE)  
  
set.seed(1728)  
ind <- sample(2, nrow(car\_Eval\_dataset), replace=T, prob=c(0.7, 0.3))  
car\_Eval\_dataset.train <- car\_Eval\_dataset[ind==1, ]  
car\_Eval\_dataset.test <- car\_Eval\_dataset[ind==2, ]  
  
  
car\_Eval\_dataset.formula <- car\_Eval\_dataset$V1 ~ car\_Eval\_dataset$V2 + car\_Eval\_dataset$V3 + car\_Eval\_dataset$V4 + car\_Eval\_dataset$V5 + car\_Eval\_dataset$V6 + car\_Eval\_dataset$V7  
car\_Eval\_dataset.ctree <- ctree(car\_Eval\_dataset.formula, data=car\_Eval\_dataset.train)  
car\_Eval\_dataset.ctree

##   
## Conditional inference tree with 14 terminal nodes  
##   
## Response: car\_Eval\_dataset$V1   
## Inputs: car\_Eval\_dataset$V2, car\_Eval\_dataset$V3, car\_Eval\_dataset$V4, car\_Eval\_dataset$V5, car\_Eval\_dataset$V6, car\_Eval\_dataset$V7   
## Number of observations: 1728   
##   
## 1) car\_Eval\_dataset$V7 == {good, vgood}; criterion = 1, statistic = 189.133  
## 2) car\_Eval\_dataset$V2 == {high, med}; criterion = 0.997, statistic = 15.109  
## 3) car\_Eval\_dataset$V7 == {vgood}; criterion = 0.988, statistic = 9.544  
## 4) car\_Eval\_dataset$V2 == {high}; criterion = 0.988, statistic = 9.5  
## 5)\* weights = 13   
## 4) car\_Eval\_dataset$V2 == {med}  
## 6)\* weights = 26   
## 3) car\_Eval\_dataset$V7 == {good}  
## 7)\* weights = 23   
## 2) car\_Eval\_dataset$V2 == {low}  
## 8)\* weights = 72   
## 1) car\_Eval\_dataset$V7 == {acc, unacc}  
## 9) car\_Eval\_dataset$V7 == {unacc}; criterion = 0.999, statistic = 21.041  
## 10) car\_Eval\_dataset$V6 == {high, med}; criterion = 0.991, statistic = 21.514  
## 11) car\_Eval\_dataset$V4 == {2}; criterion = 1, statistic = 69.991  
## 12)\* weights = 384   
## 11) car\_Eval\_dataset$V4 == {4, more}  
## 13) car\_Eval\_dataset$V5 == {small}; criterion = 0.996, statistic = 23.715  
## 14)\* weights = 130   
## 13) car\_Eval\_dataset$V5 == {big, med}  
## 15)\* weights = 120   
## 10) car\_Eval\_dataset$V6 == {low}  
## 16)\* weights = 576   
## 9) car\_Eval\_dataset$V7 == {acc}  
## 17) car\_Eval\_dataset$V2 == {high, vhigh}; criterion = 1, statistic = 152.301  
## 18) car\_Eval\_dataset$V2 == {high}; criterion = 1, statistic = 30.882  
## 19) car\_Eval\_dataset$V6 == {med}; criterion = 0.963, statistic = 10.158  
## 20)\* weights = 49   
## 19) car\_Eval\_dataset$V6 == {high}  
## 21)\* weights = 56   
## 18) car\_Eval\_dataset$V2 == {vhigh}  
## 22)\* weights = 72   
## 17) car\_Eval\_dataset$V2 == {low, med}  
## 23) car\_Eval\_dataset$V6 == {high}; criterion = 1, statistic = 43.17  
## 24)\* weights = 102   
## 23) car\_Eval\_dataset$V6 == {med}  
## 25) car\_Eval\_dataset$V5 == {big, med}; criterion = 1, statistic = 47.69  
## 26)\* weights = 77   
## 25) car\_Eval\_dataset$V5 == {small}  
## 27)\* weights = 28

plot(car\_Eval\_dataset.ctree)



pred <- predict(car\_Eval\_dataset.ctree, newdata = car\_Eval\_dataset.test)  
  
table(pred, car\_Eval\_dataset$V2,dnn = c("pred", "V1"))

## V1  
## pred high low med vhigh  
## high 296 318 341 240  
## low 62 86 63 72  
## med 0 0 0 0  
## vhigh 74 28 28 120

In the confusion matrix we can find the efficiency of the algorithm. In the diagonal values represents total number of points in each class available for high , low, med vhigh, High class has 296 points. But total (high,vhigh) + (high,low) +( high,med) = 240 +318 +341 = 899. It means we are losing few points in class high. The algorithm efficiency is very low. We need to choose different algorithm to classify.

**Conclusion**: The points in data set used for regression is complex due to that linear model could not fit all points, I considered both cases of considering single variable and 7 variable cases. In both cases we are getting almost similar values. So, we need to use different algorithm for the regression. In the classification efficiency is coming very low because each class has very less point then actual points. Due to this we need to use another algorithm for the classification.

**References**:

Zentut. (n.d.). Data Mining Techniques. *Data Mining Techniques*. Retrieved from <http://www.zentut.com/data-mining/data-mining-techniques/>

Zhao, Y. (2015). *Introduction to Data Mining with R1.* Retrieved from <https://docs.google.com/viewerng/viewer?url=http://www.rdatamining.com/docs/RDataMining-introduction-slides.pdf>

assambara, A. (2018). Determining The Optimal Number Of Clusters: 3 Must Know Methods . *DataNovia*. Retrieved from <https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/>

NEU. (2019). MPSA - ALY6015 Intermediate Analytics - V1. Retrieved from https://northeastern.acrobatiq.com/courseware/glossary/NEU\_MPSA\_INTERMEDIATE\_ANALYTICS\_ALY6015\_V1\_23

Maindonald, J. H. (2008, January 19). Using R for Data Analysis and Graphics . Introduction, Code and Commentary . Retrieved from https://cran.r-project.org/doc/contrib/usingR.pdf