Project

Bob Ding, Lynn Fan, Alice Jiang

11/18/2018

Data Exploration

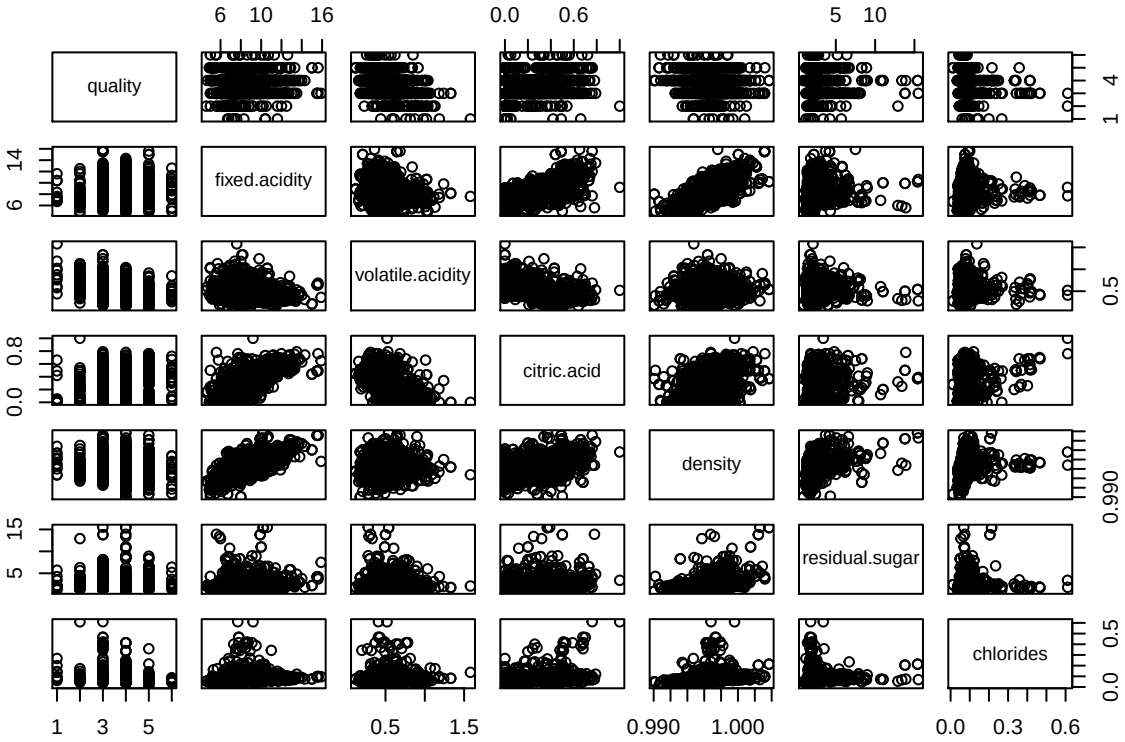
```
# input data source
data <- read.csv("./redwine_quality.csv")
```

```
# convert quality into factor variable
data <- data %>% mutate(quality = as.factor(quality))
```

```
# overview of the dataset
glimpse(data)
```

```
## Observations: 1,599
## Variables: 12
## $ fixed.acidity        <dbl> 7.4, 7.8, 7.8, 11.2, 7.4, 7.4, 7.9, 7.3, ...
## $ volatile.acidity     <dbl> 0.700, 0.880, 0.760, 0.280, 0.700, 0.660,...
## $ citric.acid          <dbl> 0.00, 0.00, 0.04, 0.56, 0.00, 0.00, 0.06,...
## $ residual.sugar       <dbl> 1.9, 2.6, 2.3, 1.9, 1.9, 1.8, 1.6, 1.2, 2...
## $ chlorides            <dbl> 0.076, 0.098, 0.092, 0.075, 0.076, 0.075,...
## $ free.sulfur.dioxide  <dbl> 11, 25, 15, 17, 11, 13, 15, 15, 9, 17, 15...
## $ total.sulfur.dioxide <dbl> 34, 67, 54, 60, 34, 40, 59, 21, 18, 102, ...
## $ density              <dbl> 0.9978, 0.9968, 0.9970, 0.9980, 0.9978, 0...
## $ pH                   <dbl> 3.51, 3.20, 3.26, 3.16, 3.51, 3.51, 3.30,...
## $ sulphates            <dbl> 0.56, 0.68, 0.65, 0.58, 0.56, 0.56, 0.46,...
## $ alcohol              <dbl> 9.4, 9.8, 9.8, 9.8, 9.4, 9.4, 9.4, 10.0, ...
## $ quality              <fct> 5, 5, 5, 6, 5, 5, 5, 7, 7, 5, 5, 5, 5, 5,...
```

```
# scatter plot matrix one
pairs(quality ~ fixed.acidity + volatile.acidity + citric.acid + density + residual.sugar + chlorides, d
```



The scatter plot matrix suggests a fairly strong positive linear relationship between fixed.acidity and citric.acid; fixed.acidity and density; and density and residual.sugar. There also appears to be a strong negative linear relationship between volatile.acidity and citric.acid.

```
# correlation coefficients for potential multicollinearity
cor(data$fixed.acidity, data$citric.acid)
```

```
## [1] 0.6717034
```

```
cor(data$fixed.acidity, data$density)
```

```
## [1] 0.6680473
```

```
cor(data$volatile.acidity, data$citric.acid)
```

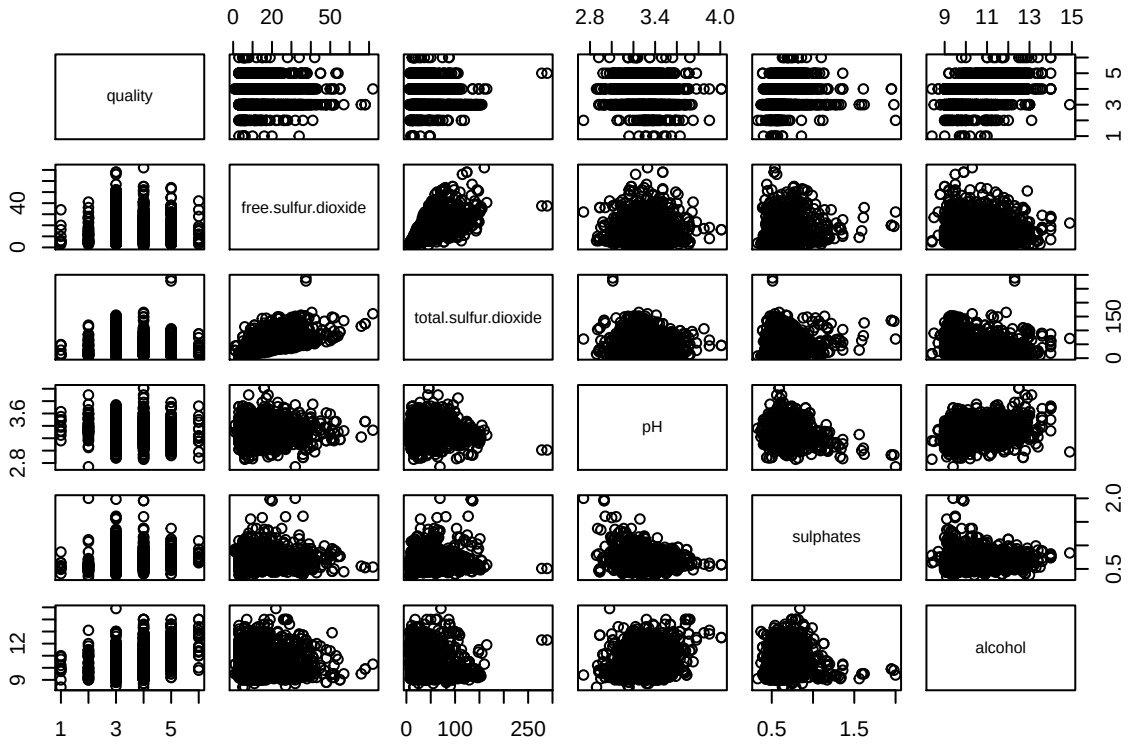
```
## [1] -0.5524957
```

```
cor(data$density, data$residual.sugar)
```

```
## [1] 0.3552834
```

The correlation coefficent for fixed.acidity and citric.acid (0.6717034) and the correlation coefficient for fixed.acidity and density (0.6680473) show moderate to strong linear relationship between the variables. To remove potential multicollinearity impact, we **will not include fixed.acidity** in the model.

```
# scatter plot matrix two
pairs(quality ~ free.sulfur.dioxide + total.sulfur.dioxide + pH + sulphates + alcohol, data=data)
```



The scatter plot shows strong positive linear relationship between free.sulfur.dioxide and total.sulfur.dioxide.

```
# correlation coefficients for potential multicollinearity
cor(data$free.sulfur.dioxide, data$total.sulfur.dioxide)
```

```
## [1] 0.6676665
```

The correlation coefficient of 0.6676665 between free.sulfur.dioxide and total.sulfur.dioxide indicates strong positive linear relationship. We also know from the data description that total.sulfur dioxide includes free.sulfur.dioxide - whose concentration level could impact the smell and taste of the wine, thus impacting quality. So it is sufficient to include only one of these two variables in the model, and we **will include free.sulphur.dioxide**.

```
# distribution of each variable
p1  <- ggplot(data = data, aes(x = quality) ) + geom_histogram(stat="count",fill= "lightblue")
```

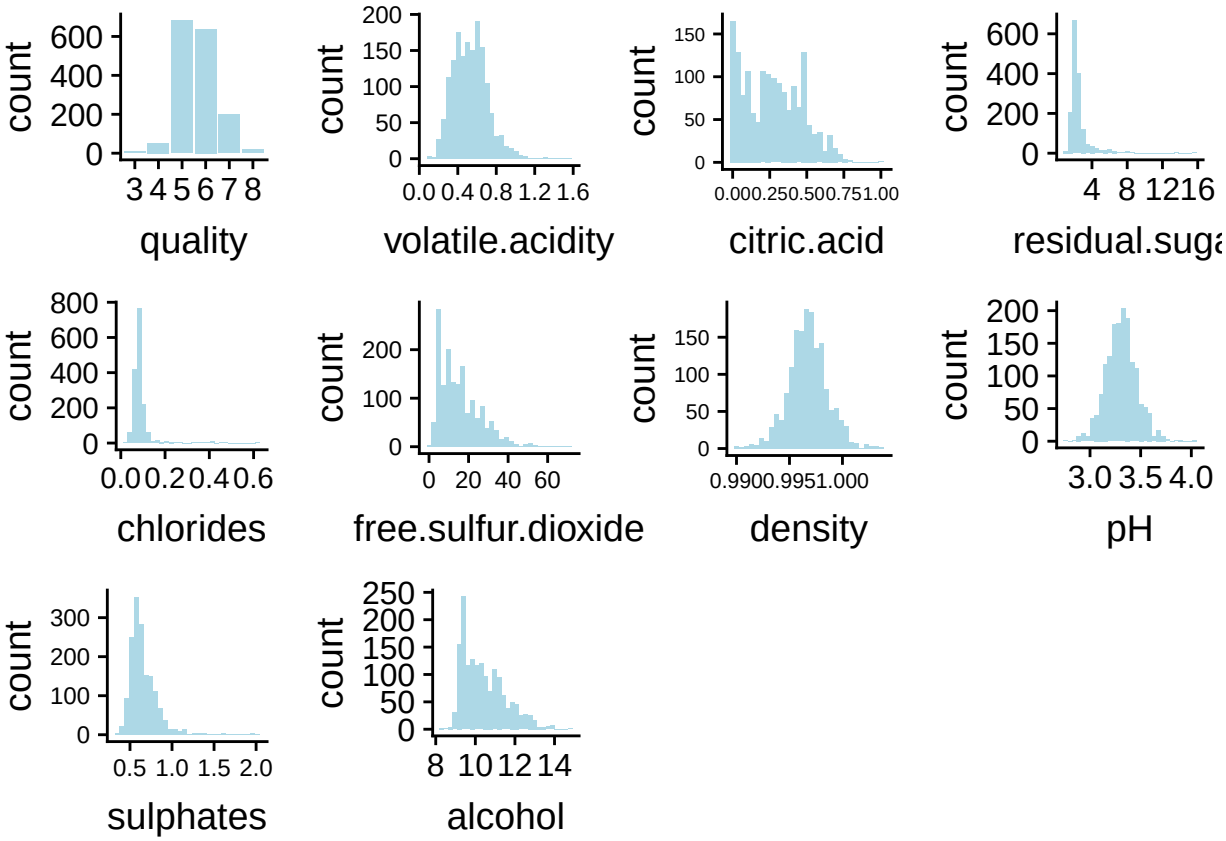
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
p3  <- ggplot(data = data, aes(x = volatile.acidity) ) + theme(axis.text=element_text(size=9)) + geom_hi
p4  <- ggplot(data = data, aes(x = citric.acid) ) + theme(axis.text=element_text(size=7)) + geom_histogr
p5  <- ggplot(data = data, aes(x = residual.sugar) ) + geom_histogram(fill= "lightblue")
p6  <- ggplot(data = data, aes(x = chlorides) ) + theme(axis.text=element_text(size=11)) + geom_histogra
p7  <- ggplot(data = data, aes(x = free.sulfur.dioxide) ) + theme(axis.text=element_text(size=9)) + geom
p9  <- ggplot(data = data, aes(x = density) ) + theme(axis.text=element_text(size=7.5)) + geom_histogram
p10 <- ggplot(data = data, aes(x = pH) ) + geom_histogram(fill= "lightblue")
p11 <- ggplot(data = data, aes(x = sulphates) ) + theme(axis.text=element_text(size=9)) + geom_histogram
p12 <- ggplot(data = data, aes(x = alcohol) ) + geom_histogram(fill= "lightblue")

plot_grid(p1,p3,p4,p5,p6,p7,p9,p10,p11,p12,ncol = 4,nrow = 3)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



We can see that free.sulfur.dioxide and alcohol have an obvious rightly skewed distribution. residual.sugar, cholorides, and sulphates also have a slightly rightward skewedness. citric.acid at first appears to have a bimodal distribution, but this is because there are some wines with zero citric acid, so we see a spike at 0 in the histogram. Based on the data definition, we know it is possible for wines to have citric acid of 0. ***** The distribution of citric acid is overall fairly symmetric. ***** All other variables: quality, volatile.acidity, density, and pH are unimodal and fairly symmetric.

******sort quality********

```
# summary statistics of each variable
summary(data)
```

```
##  fixed.acidity   volatile.acidity  citric.acid    residual.sugar
##  Min.   : 4.60   Min.   :0.1200   Min.   :0.000   Min.   : 0.900
##  1st Qu.: 7.10   1st Qu.:0.3900   1st Qu.:0.090   1st Qu.: 1.900
##  Median : 7.90   Median :0.5200   Median :0.260   Median : 2.200
##  Mean   : 8.32   Mean   :0.5278   Mean   :0.271   Mean   : 2.539
##  3rd Qu.: 9.20   3rd Qu.:0.6400   3rd Qu.:0.420   3rd Qu.: 2.600
##  Max.   :15.90   Max.   :1.5800   Max.   :1.000   Max.   :15.500
##    chlorides       free.sulfur.dioxide total.sulfur.dioxide
##  Min.   :0.01200   Min.   : 1.00       Min.   :  6.00
##  1st Qu.:0.07000   1st Qu.: 7.00       1st Qu.: 22.00
```

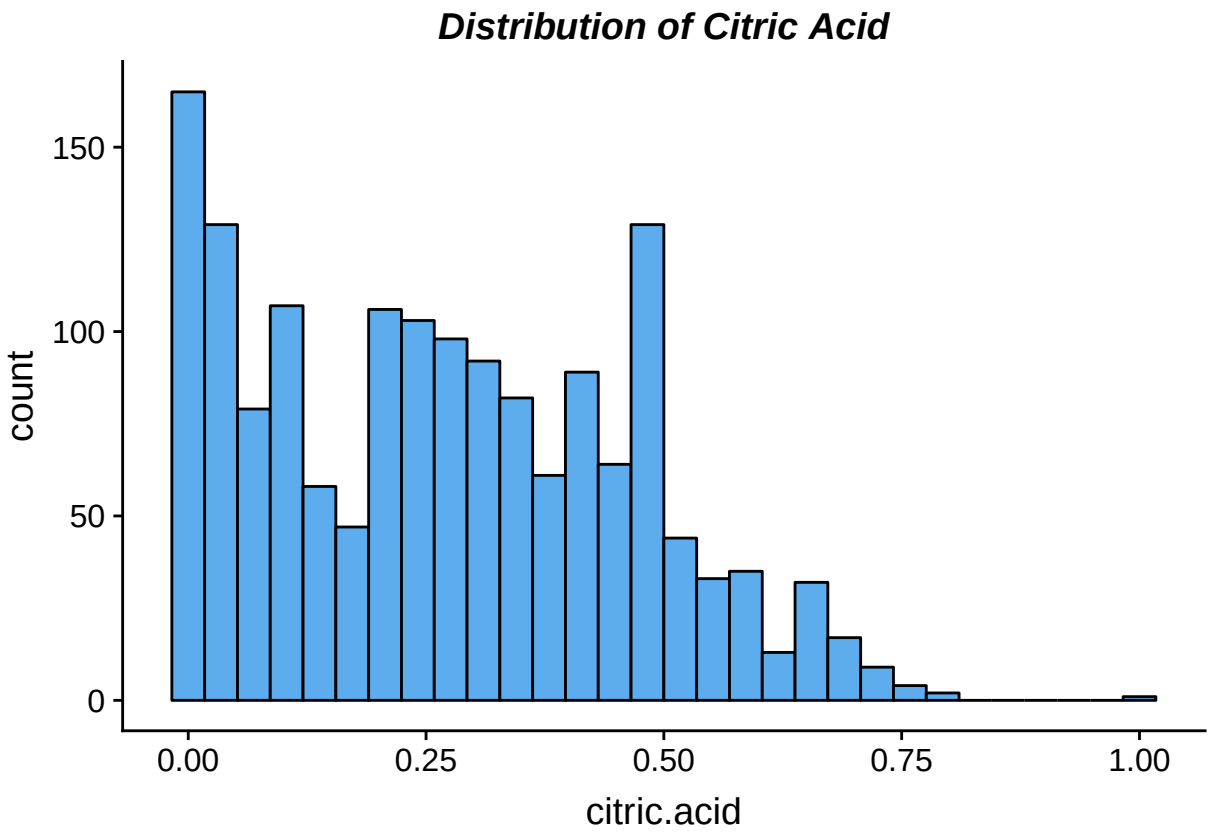
```
##  Median :0.07900   Median :14.00       Median : 38.00
##  Mean   :0.08747   Mean   :15.87       Mean   : 46.47
##  3rd Qu.:0.09000   3rd Qu.:21.00       3rd Qu.: 62.00
##  Max.   :0.61100   Max.   :72.00       Max.   :289.00
##     density             pH          sulphates         alcohol      quality
##  Min.   :0.9901   Min.   :2.740   Min.   :0.3300   Min.   : 8.40   3: 10
##  1st Qu.:0.9956   1st Qu.:3.210   1st Qu.:0.5500   1st Qu.: 9.50   4: 53
##  Median :0.9968   Median :3.310   Median :0.6200   Median :10.20   5:681
##  Mean   :0.9967   Mean   :3.311   Mean   :0.6581   Mean   :10.42   6:638
##  3rd Qu.:0.9978   3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   7:199
##  Max.   :1.0037   Max.   :4.010   Max.   :2.0000   Max.   :14.90   8: 18
```

From the previous histogram, we can see that quality has a fairly normal distribution. The summary statistic also shows that quality rating of 5 has the most number of observations (it is also the middle rating on the scale of 0-10). So quality rating of 5 is the most appropriate **reference level** for the model.

```
# makes quality==5 the reference level
data <- data %>% mutate(quality = relevel(quality, ref=5))
```

```
# distribution of citric acid
ggplot(aes(x=citric.acid), data=data) +
  geom_histogram(fill="steelblue2", color="black") +
  ggtitle("Distribution of Citric Acid") +
  theme(plot.title=element_text(color="black", size=14, face="bold.italic", hjust=0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# add mean-centered numerical variables besides citric.acid
data <- data %>%
  mutate(volatile.acidityCent = volatile.acidity-mean(volatile.acidity),
```

```
          residual.sugarCent = residual.sugar-mean(residual.sugar),
          chloridesCent = chlorides-mean(chlorides),
          free.sulfur.dioxideCent = free.sulfur.dioxide-mean(free.sulfur.dioxide),
          sulphatesCent = sulphates-mean(sulphates),
          densityCent = density-mean(density), pHCent = pH-mean(pH),
          alcoholCent = alcohol-mean(alcohol))
```

Multiple Regression Analysis

```
# change quality into a numerical variable
data <- data %>% mutate(quality = as.numeric(quality))
```

```
# primary model
full.model <- lm(quality ~ volatile.acidityCent + residual.sugarCent + citric.acid + chloridesCent + fr
```

```
# view the primary model
kable(tidy(full.model), format="markdown")
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.2078379	0.0770411	54.6181017	0.0000000
volatile.acidityCent	-0.1346401	0.2266206	-0.5941213	0.5525155
residual.sugarCent	-0.0658173	0.0256206	-2.5689180	0.0102922
citric.acid	-0.7577667	0.2602039	-2.9122039	0.0036390
chloridesCent	1.4116625	0.7778250	1.8148844	0.0697301
free.sulfur.dioxideCent	0.0070710	0.0031087	2.2745911	0.0230632
sulphatesCent	-0.5609678	0.2159487	-2.5976900	0.0094720
densityCent	52.5804142	25.2708430	2.0806751	0.0376235
pHCent	0.0995892	0.2550857	0.3904146	0.6962823
alcoholCent	-0.1430496	0.0411678	-3.4747890	0.0005250

```
# check model fit
glance(full.model)
```

```
## # A tibble: 1 x 11
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC   BIC
## *     <dbl>         <dbl> <dbl>     <dbl>    <dbl> <int>  <dbl> <dbl> <dbl>
## 1    0.0559        0.0506  1.24      10.5 7.54e-16    10 -2610. 5241. 5301.
## # ... with 2 more variables: deviance <dbl>, df.residual <int>
```

This is the primary model we constructed based on the exploratory analysis. We can see that volatile.acidityCent, chloridesCent, and pHCent have p-values greater than 0.05. The model's R^2 value 0.0559 is noticeably higher than the Adjusted R^2 of 0.0506, penalizing for unnecessary variables in the model.

We will then use model selection to build a better model, first including only the main effects.

Model Selections

```
# backward selection process
backward <- ols_step_backward_aic(full.model, details=TRUE)
```

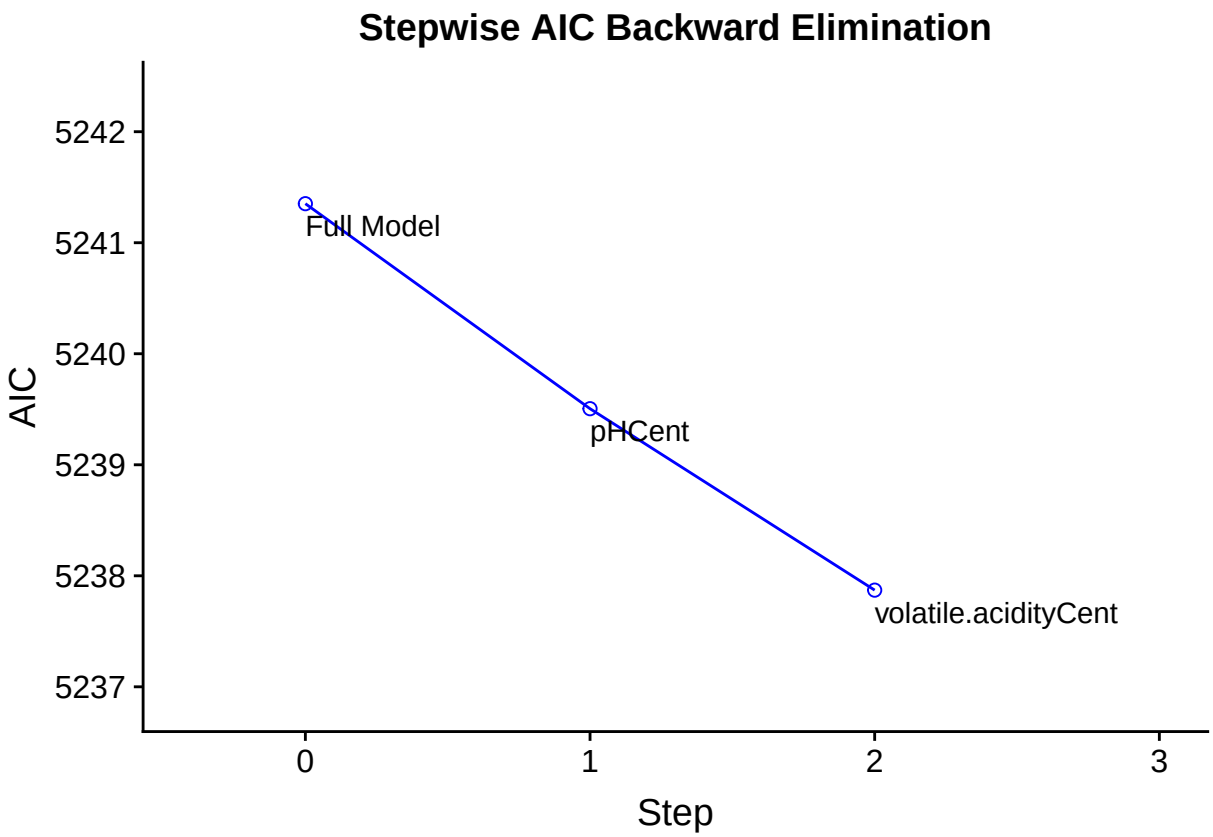
```
## Backward Elimination Method
```

```
## ---------------------------
##
## Candidate Terms:
##
## 1 . volatile.acidityCent
## 2 . residual.sugarCent
## 3 . citric.acid
## 4 . chloridesCent
## 5 . free.sulfur.dioxideCent
## 6 . sulphatesCent
## 7 . densityCent
## 8 . pHCent
## 9 . alcoholCent
##
##  Step 0: AIC = 5241.352
##  quality ~ volatile.acidityCent + residual.sugarCent + citric.acid + chloridesCent + free.sulfur.dio
##
## ------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------
## pHCent                      1     5239.505     0.235    2449.128    0.056        0.051
## volatile.acidityCent        1     5239.707     0.544    2449.437    0.056        0.051
## chloridesCent               1     5242.663     5.076    2453.970    0.054        0.049
## densityCent                 1     5243.702     6.672    2455.565    0.053        0.049
## free.sulfur.dioxideCent     1     5244.550     7.974    2456.867    0.053        0.048
## residual.sugarCent          1     5245.979    10.171    2459.064    0.052        0.047
## sulphatesCent               1     5246.128    10.400    2459.293    0.052        0.047
## citric.acid                 1     5247.863    13.070    2461.964    0.051        0.046
## alcoholCent                 1     5251.456    18.608    2467.502    0.049        0.044
## ------------------------------------------------------------------------------------
##
##
## - pHCent
##
##
##   Step 1 : AIC = 5239.505
##  quality ~ volatile.acidityCent + residual.sugarCent + citric.acid + chloridesCent + free.sulfur.dio
##
## ------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------
## volatile.acidityCent        1     5237.871     0.560    2449.689    0.056        0.051
## chloridesCent               1     5240.715     4.921    2454.049    0.054        0.050
## densityCent                 1     5241.900     6.741    2455.869    0.053        0.049
## free.sulfur.dioxideCent     1     5242.873     8.235    2457.363    0.053        0.049
## residual.sugarCent          1     5244.217    10.302    2459.430    0.052        0.048
## sulphatesCent               1     5244.351    10.507    2459.636    0.052        0.048
## citric.acid                 1     5249.254    18.062    2467.190    0.049        0.045
## alcoholCent                 1     5249.666    18.698    2467.826    0.049        0.044
## ------------------------------------------------------------------------------------
##
## - volatile.acidityCent
##
##
```

```
##   Step 2 : AIC = 5237.871
##  quality ~ residual.sugarCent + citric.acid + chloridesCent + free.sulfur.dioxideCent + sulphatesCen
##
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## chloridesCent               1     5238.740     4.400    2454.089    0.054        0.050
## densityCent                 1     5239.909     6.195    2455.883    0.053        0.050
## free.sulfur.dioxideCent     1     5241.338     8.391    2458.079    0.052        0.049
## sulphatesCent               1     5242.354     9.952    2459.641    0.052        0.048
## residual.sugarCent          1     5242.528    10.219    2459.908    0.052        0.048
## alcoholCent                 1     5248.410    19.286    2468.974    0.048        0.045
## citric.acid                 1     5250.116    21.922    2471.610    0.047        0.044
## ----------------------------------------------------------------------------------------
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - pHCent
## - volatile.acidityCent
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ------------------------------------------------------------------
## R                         0.236       RMSE                 1.241
## R-Squared                 0.056       Coef. Var           31.002
## Adj. R-Squared            0.051       MSE                  1.540
## Pred R-Squared            0.045       MAE                  0.925
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares          DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     144.301           7         20.614    13.389    0.0000
## Residual      2449.689        1591          1.540
## Total         2593.990        1598
## ---------------------------------------------------------------------
##
##                                     Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                    model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##              (Intercept)     4.198         0.060                69.597    0.000     4.079     4.316
##      residual.sugarCent     -0.066         0.026       -0.073   -2.576    0.010    -0.116    -0.016
```

```
##             citric.acid    -0.720          0.191       -0.110    -3.773    0.000    -1.095    -0.346
##           chloridesCent     1.267          0.750        0.047     1.691    0.091    -0.203     2.738
## free.sulfur.dioxideCent     0.007          0.003        0.059     2.334    0.020     0.001     0.013
##           sulphatesCent    -0.536          0.211       -0.071    -2.542    0.011    -0.950    -0.123
##             densityCent    49.122         24.490        0.073     2.006    0.045     1.087    97.158
##             alcoholCent    -0.141          0.040       -0.118    -3.539    0.000    -0.219    -0.063
## ----------------------------------------------------------------------------------------------------
```

```
# view stepwise AIC backward elimination
plot(backward)
```



The backward selection removes pHCent and volatile.acidityCent from the model.

```
# forward selection process
forward <- ols_step_forward_aic(full.model, details=TRUE)
```

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . volatile.acidityCent
## 2 . residual.sugarCent
## 3 . citric.acid
## 4 . chloridesCent
## 5 . free.sulfur.dioxideCent
## 6 . sulphatesCent
## 7 . densityCent
## 8 . pHCent
```

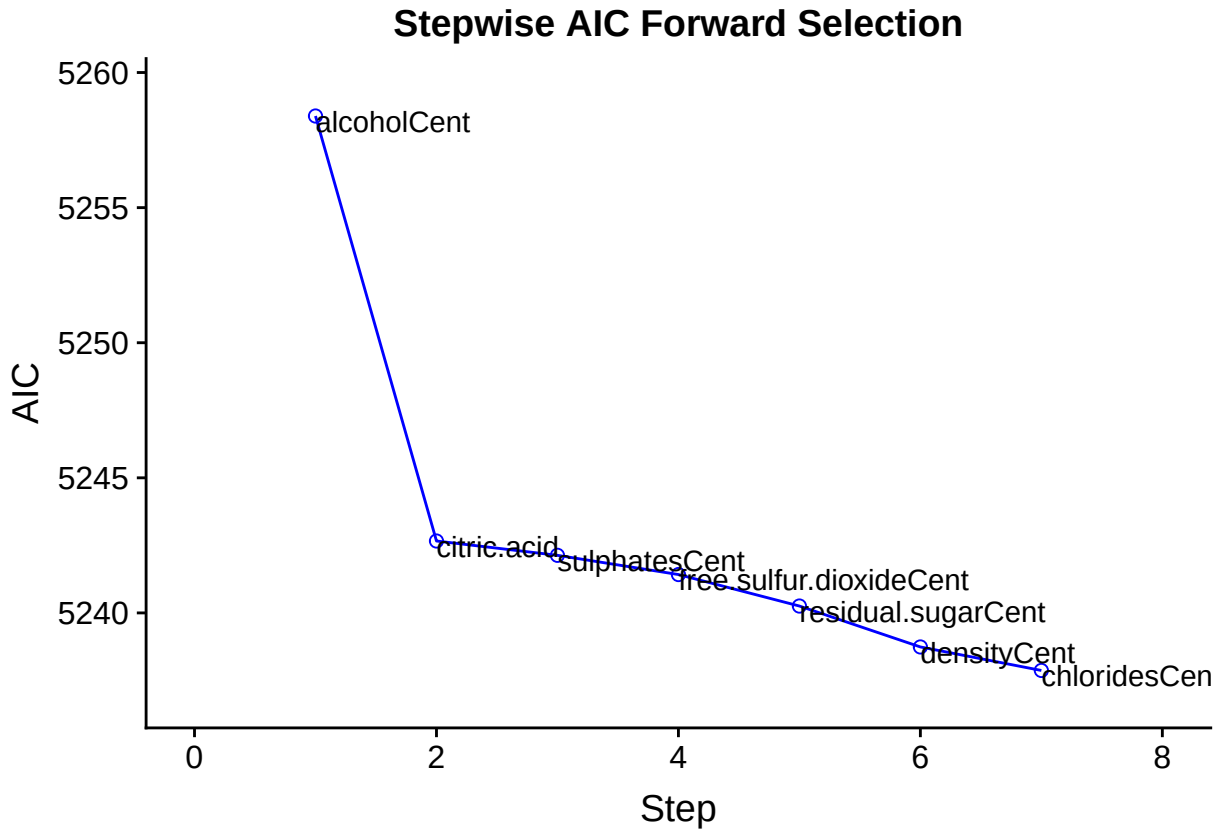
```
## 9 . alcoholCent
##
##  Step 0: AIC = 5315.392
##  quality ~ 1
##
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## alcoholCent                  1    5258.390    93.971    2500.019    0.036        0.036
## citric.acid                  1    5292.863    39.488    2554.502    0.015        0.015
## volatile.acidityCent         1    5302.658    23.791    2570.199    0.009        0.009
## sulphatesCent                1    5305.527    19.176    2574.814    0.007        0.007
## densityCent                  1    5311.101    10.185    2583.805    0.004        0.003
## free.sulfur.dioxideCent      1    5312.446     8.011    2585.979    0.003        0.002
## residual.sugarCent           1    5312.626     7.720    2586.270    0.003        0.002
## chloridesCent                1    5315.451     3.146    2590.844    0.001        0.001
## pHCent                       1    5315.916     2.392    2591.598    0.001        0.000
## ----------------------------------------------------------------------------------------
##
##
## - alcoholCent
##
##
##  Step 1 : AIC = 5258.39
##  quality ~ alcoholCent
##
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## citric.acid                  1    5242.662    27.566    2472.453    0.047        0.046
## pHCent                       1    5251.999    13.085    2486.934    0.041        0.040
## sulphatesCent                1    5252.594    12.160    2487.860    0.041        0.040
## volatile.acidityCent         1    5254.707     8.870    2491.149    0.040        0.038
## residual.sugarCent           1    5256.785     5.630    2494.389    0.038        0.037
## free.sulfur.dioxideCent      1    5257.396     4.677    2495.342    0.038        0.037
## densityCent                  1    5258.166     3.475    2496.544    0.038        0.036
## chloridesCent                1    5260.298     0.144    2499.875    0.036        0.035
## ----------------------------------------------------------------------------------------
##
## - citric.acid
##
##
##  Step 2 : AIC = 5242.662
##  quality ~ alcoholCent + citric.acid
##
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## sulphatesCent                1    5242.132     3.908    2468.545    0.048        0.047
## free.sulfur.dioxideCent      1    5242.367     3.546    2468.908    0.048        0.046
## residual.sugarCent           1    5242.890     2.737    2469.716    0.048        0.046
## chloridesCent                1    5244.159     0.776    2471.677    0.047        0.045
## densityCent                  1    5244.261     0.620    2471.833    0.047        0.045
## pHCent                       1    5244.344     0.491    2471.963    0.047        0.045
```

```
## volatile.acidityCent        1    5244.649     0.020    2472.433    0.047        0.045
## ---------------------------------------------------------------------------------------
##
## - sulphatesCent
##
##
##  Step 3 : AIC = 5242.132
##  quality ~ alcoholCent + citric.acid + sulphatesCent
##
## ---------------------------------------------------------------------------------------
## Variable                   DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------
## free.sulfur.dioxideCent     1    5241.420     4.183    2464.362    0.050        0.048
## residual.sugarCent          1    5242.164     3.036    2465.509    0.050        0.047
## chloridesCent               1    5242.253     2.900    2465.646    0.049        0.047
## densityCent                 1    5243.517     0.949    2467.596    0.049        0.046
## pHCent                      1    5243.908     0.347    2468.199    0.048        0.046
## volatile.acidityCent        1    5244.130     0.004    2468.542    0.048        0.046
## ---------------------------------------------------------------------------------------
##
## - free.sulfur.dioxideCent
##
##
##  Step 4 : AIC = 5241.42
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent
##
## ---------------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------
## residual.sugarCent        1    5240.255     4.874    2459.488    0.052        0.049
## chloridesCent             1    5241.410     3.097    2461.266    0.051        0.048
## densityCent               1    5242.659     1.172    2463.190    0.050        0.047
## pHCent                    1    5243.294     0.194    2464.168    0.050        0.047
## volatile.acidityCent      1    5243.418     0.003    2464.359    0.050        0.047
## ---------------------------------------------------------------------------------------
##
## - residual.sugarCent
##
##
##  Step 5 : AIC = 5240.255
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent
##
## ---------------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------
## densityCent               1    5238.740     5.399    2454.089    0.054        0.050
## chloridesCent             1    5239.909     3.605    2455.883    0.053        0.050
## pHCent                    1    5242.171     0.128    2459.360    0.052        0.048
## volatile.acidityCent      1    5242.191     0.097    2459.391    0.052        0.048
## ---------------------------------------------------------------------------------------
##
## - densityCent
##
##
```

```
##  Step 6 : AIC = 5238.74
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent +
##
## ------------------------------------------------------------------------------------
## Variable                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------
## chloridesCent            1    5237.871     4.400    2449.689    0.056        0.051
## pHCent                   1    5240.685     0.085    2454.004    0.054        0.050
## volatile.acidityCent     1    5240.715     0.040    2454.049    0.054        0.050
## ------------------------------------------------------------------------------------
##
## - chloridesCent
##
##
##  Step 7 : AIC = 5237.871
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent +
##
## ------------------------------------------------------------------------------------
## Variable                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------
## volatile.acidityCent     1    5239.505     0.560    2449.128    0.056        0.051
## pHCent                   1    5239.707     0.251    2449.437    0.056        0.051
## ------------------------------------------------------------------------------------
##
##
## No more variables to be added.
##
## Variables Entered:
##
## - alcoholCent
## - citric.acid
## - sulphatesCent
## - free.sulfur.dioxideCent
## - residual.sugarCent
## - densityCent
## - chloridesCent
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.236       RMSE                1.241
## R-Squared               0.056       Coef. Var          31.002
## Adj. R-Squared          0.051       MSE                 1.540
## Pred R-Squared          0.045       MAE                 0.925
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------
```

```
##                 Sum of
##                Squares          DF     Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression     144.301           7         20.614     13.389     0.0000
## Residual      2449.689        1591          1.540
## Total         2593.990        1598
## ----------------------------------------------------------------------
##
##                                      Parameter Estimates
## ---------------------------------------------------------------------------------------------------
##                    model      Beta     Std. Error    Std. Beta      t        Sig      lower     upper
## ---------------------------------------------------------------------------------------------------
##              (Intercept)     4.198          0.060                69.597    0.000     4.079     4.316
##              alcoholCent    -0.141          0.040       -0.118   -3.539    0.000    -0.219    -0.063
##              citric.acid    -0.720          0.191       -0.110   -3.773    0.000    -1.095    -0.346
##            sulphatesCent    -0.536          0.211       -0.071   -2.542    0.011    -0.950    -0.123
## free.sulfur.dioxideCent     0.007          0.003        0.059    2.334    0.020     0.001     0.013
##      residual.sugarCent    -0.066          0.026       -0.073   -2.576    0.010    -0.116    -0.016
##              densityCent    49.122         24.490        0.073    2.006    0.045     1.087    97.158
##            chloridesCent     1.267          0.750        0.047    1.691    0.091    -0.203     2.738
## ---------------------------------------------------------------------------------------------------
```

```r
# view stepwise AIC forward selection
plot(forward)
```



We can see the foward selection process includes the same explanatory variables as the model from the backward selection - both models do not include volatile.acidityCent and pHCent from the full.model.

```r
# stepwise selection process
stepwise <- ols_step_both_aic(full.model, details=TRUE)
```

```
## Stepwise Selection Method
## -------------------------
##
## Candidate Terms:
##
## 1 . volatile.acidityCent
## 2 . residual.sugarCent
## 3 . citric.acid
## 4 . chloridesCent
## 5 . free.sulfur.dioxideCent
## 6 . sulphatesCent
## 7 . densityCent
## 8 . pHCent
## 9 . alcoholCent
##
##  Step 0: AIC = 5315.392
##  quality ~ 1
##
##
##                                 Enter New Variables
## ----------------------------------------------------------------------------------------
## Variable                   DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## alcoholCent                 1    5258.390    93.971    2500.019    0.036         0.036
## citric.acid                 1    5292.863    39.488    2554.502    0.015         0.015
## volatile.acidityCent        1    5302.658    23.791    2570.199    0.009         0.009
## sulphatesCent               1    5305.527    19.176    2574.814    0.007         0.007
## densityCent                 1    5311.101    10.185    2583.805    0.004         0.003
## free.sulfur.dioxideCent     1    5312.446     8.011    2585.979    0.003         0.002
## residual.sugarCent          1    5312.626     7.720    2586.270    0.003         0.002
## chloridesCent               1    5315.451     3.146    2590.844    0.001         0.001
## pHCent                      1    5315.916     2.392    2591.598    0.001         0.000
## ----------------------------------------------------------------------------------------
##
## - alcoholCent added
##
##
##  Step 1 : AIC = 5258.39
##  quality ~ alcoholCent
##
##                                 Enter New Variables
## ----------------------------------------------------------------------------------------
## Variable                   DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## citric.acid                 1    5242.662    121.537    2472.453    0.047         0.046
## pHCent                      1    5251.999    107.056    2486.934    0.041         0.040
## sulphatesCent               1    5252.594    106.130    2487.860    0.041         0.040
## volatile.acidityCent        1    5254.707    102.841    2491.149    0.040         0.038
## residual.sugarCent          1    5256.785     99.601    2494.389    0.038         0.037
## free.sulfur.dioxideCent     1    5257.396     98.648    2495.342    0.038         0.037
## densityCent                 1    5258.166     97.446    2496.544    0.038         0.036
```

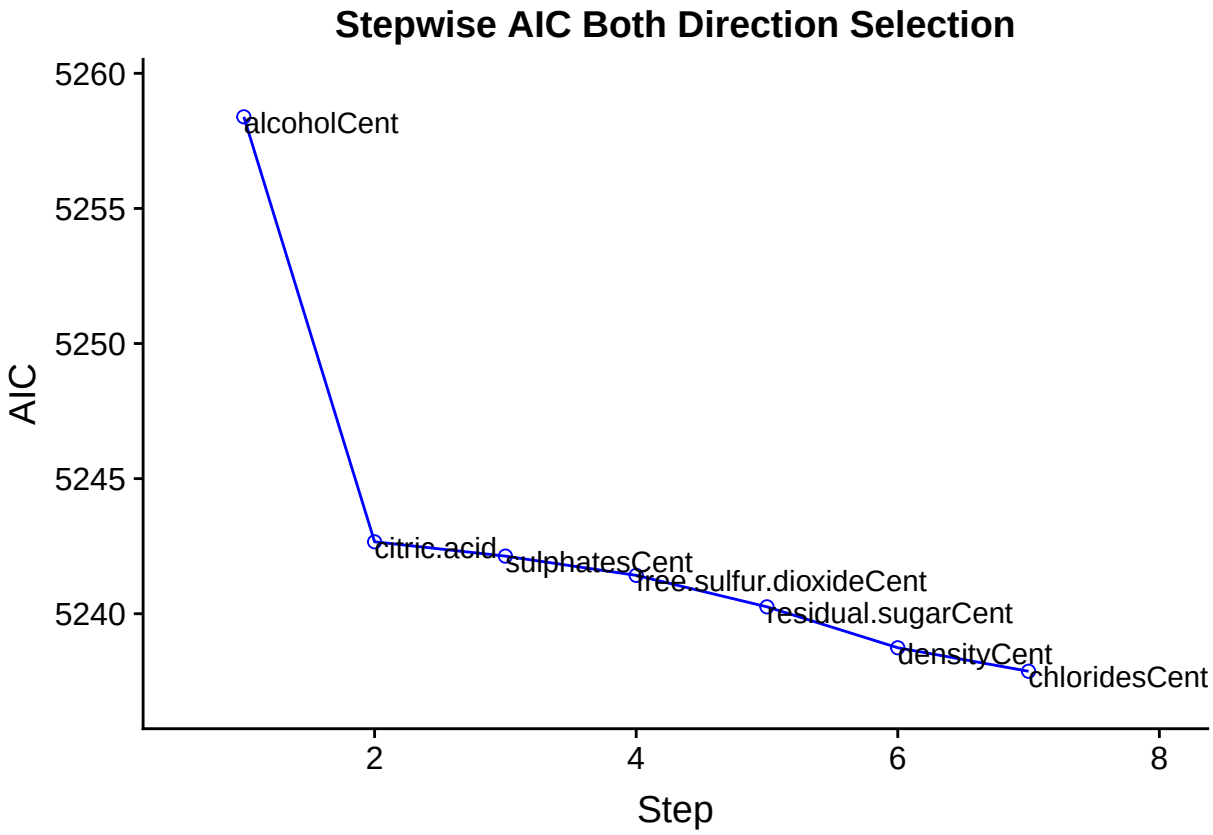
```
## chloridesCent               1    5260.298     94.115    2499.875    0.036        0.035
## ---------------------------------------------------------------------------------------
##
## - citric.acid added
##
##
##  Step 2 : AIC = 5242.662
##  quality ~ alcoholCent + citric.acid
##
##                          Remove Existing Variables
## ------------------------------------------------------------------------
## Variable       DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## citric.acid     1    5258.390    93.971    2500.019    0.036        0.036
## alcoholCent     1    5292.863    39.488    2554.502    0.015        0.015
## ------------------------------------------------------------------------
##
##                                  Enter New Variables
## ---------------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------
## sulphatesCent                1    5242.132    125.445    2468.545    0.048        0.047
## free.sulfur.dioxideCent      1    5242.367    125.082    2468.908    0.048        0.046
## residual.sugarCent           1    5242.890    124.274    2469.716    0.048        0.046
## chloridesCent                1    5244.159    122.313    2471.677    0.047        0.045
## densityCent                  1    5244.261    122.157    2471.833    0.047        0.045
## pHCent                       1    5244.344    122.027    2471.963    0.047        0.045
## volatile.acidityCent         1    5244.649    121.557    2472.433    0.047        0.045
## ---------------------------------------------------------------------------------------------
##
## - sulphatesCent added
##
##
##  Step 3 : AIC = 5242.132
##  quality ~ alcoholCent + citric.acid + sulphatesCent
##
##                           Remove Existing Variables
## ------------------------------------------------------------------------------
## Variable          DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------
## sulphatesCent      1    5242.662    121.537    2472.453    0.047        0.046
## citric.acid        1    5252.594    106.130    2487.860    0.041        0.040
## alcoholCent        1    5290.816     45.945    2548.045    0.018        0.016
## ------------------------------------------------------------------------------
##
##                                  Enter New Variables
## ---------------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent      1    5241.420    129.628    2464.362    0.050        0.048
## residual.sugarCent           1    5242.164    128.481    2465.509    0.050        0.047
## chloridesCent                1    5242.253    128.344    2465.646    0.049        0.047
## densityCent                  1    5243.517    126.394    2467.596    0.049        0.046
## pHCent                       1    5243.908    125.791    2468.199    0.048        0.046
```

```
## volatile.acidityCent        1    5244.130    125.448    2468.542    0.048        0.046
## ----------------------------------------------------------------------------------------
##
## - free.sulfur.dioxideCent added
##
##
##  Step 4 : AIC = 5241.42
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent
##
##                                Remove Existing Variables
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## free.sulfur.dioxideCent      1    5242.132    125.445    2468.545    0.048        0.047
## sulphatesCent                1    5242.367    125.082    2468.908    0.048        0.046
## citric.acid                  1    5250.977    111.752    2482.238    0.043        0.041
## alcoholCent                  1    5288.408     52.960    2541.030    0.020        0.019
## ----------------------------------------------------------------------------------------
##
##                                  Enter New Variables
## ---------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------
## residual.sugarCent        1    5240.255    134.502    2459.488    0.052        0.049
## chloridesCent             1    5241.410    132.724    2461.266    0.051        0.048
## densityCent               1    5242.659    130.800    2463.190    0.050        0.047
## pHCent                    1    5243.294    129.822    2464.168    0.050        0.047
## volatile.acidityCent      1    5243.418    129.631    2464.359    0.050        0.047
## ---------------------------------------------------------------------------------
##
## - residual.sugarCent added
##
##
##  Step 5 : AIC = 5240.255
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent
##
##                                Remove Existing Variables
## ----------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------
## residual.sugarCent           1    5241.420    129.628    2464.362    0.050        0.048
## sulphatesCent                1    5241.579    129.383    2464.607    0.050        0.047
## free.sulfur.dioxideCent      1    5242.164    128.481    2465.509    0.050        0.047
## citric.acid                  1    5247.653    120.003    2473.987    0.046        0.044
## alcoholCent                  1    5286.171     59.684    2534.306    0.023        0.021
## ----------------------------------------------------------------------------------------
##
##                                  Enter New Variables
## ---------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------
## densityCent               1    5238.740    139.901    2454.089    0.054        0.050
## chloridesCent             1    5239.909    138.107    2455.883    0.053        0.050
## pHCent                    1    5242.171    134.630    2459.360    0.052        0.048
```

```
## volatile.acidityCent     1    5242.191    134.599    2459.391    0.052        0.048
## ----------------------------------------------------------------------------------
##
## - densityCent added
##
##
##  Step 6 : AIC = 5238.74
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent +
##
##                                Remove Existing Variables
## ------------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------------
## densityCent                  1    5240.255    134.502    2459.488    0.052        0.049
## sulphatesCent                1    5241.007    133.344    2460.646    0.051        0.048
## free.sulfur.dioxideCent      1    5241.808    132.112    2461.878    0.051        0.048
## residual.sugarCent           1    5242.659    130.800    2463.190    0.050        0.047
## citric.acid                  1    5249.498    120.243    2473.747    0.046        0.043
## alcoholCent                  1    5254.096    113.119    2480.871    0.044        0.041
## ------------------------------------------------------------------------------------------
##
##                                  Enter New Variables
## ----------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------
## chloridesCent             1    5237.871    144.301    2449.689    0.056        0.051
## pHCent                    1    5240.685    139.986    2454.004    0.054        0.050
## volatile.acidityCent      1    5240.715    139.941    2454.049    0.054        0.050
## ----------------------------------------------------------------------------------
##
## - chloridesCent added
##
##
##  Step 7 : AIC = 5237.871
##  quality ~ alcoholCent + citric.acid + sulphatesCent + free.sulfur.dioxideCent + residual.sugarCent +
##
##                                Remove Existing Variables
## ------------------------------------------------------------------------------------------
## Variable                    DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------------
## chloridesCent                1    5238.740    139.901    2454.089    0.054        0.050
## densityCent                  1    5239.909    138.107    2455.883    0.053        0.050
## free.sulfur.dioxideCent      1    5241.338    135.911    2458.079    0.052        0.049
## sulphatesCent                1    5242.354    134.349    2459.641    0.052        0.048
## residual.sugarCent           1    5242.528    134.082    2459.908    0.052        0.048
## alcoholCent                  1    5248.410    125.016    2468.974    0.048        0.045
## citric.acid                  1    5250.116    122.380    2471.610    0.047        0.044
## ------------------------------------------------------------------------------------------
##
##                                  Enter New Variables
## ----------------------------------------------------------------------------------
## Variable                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------
## volatile.acidityCent      1    5239.505    144.862    2449.128    0.056        0.051
```

```
## pHCent                    1    5239.707    144.553    2449.437    0.056        0.051
## ----------------------------------------------------------------------------------
##
##
## No more variables to be added or removed.
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.236       RMSE                 1.241
## R-Squared                0.056       Coef. Var           31.002
## Adj. R-Squared           0.051       MSE                  1.540
## Pred R-Squared           0.045       MAE                  0.925
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares          DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     144.301           7         20.614    13.389    0.0000
## Residual      2449.689        1591          1.540
## Total         2593.990        1598
## ---------------------------------------------------------------------
##
##                                     Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                    model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##              (Intercept)     4.198         0.060                69.597    0.000     4.079     4.316
##              alcoholCent    -0.141         0.040       -0.118   -3.539    0.000    -0.219    -0.063
##              citric.acid    -0.720         0.191       -0.110   -3.773    0.000    -1.095    -0.346
##            sulphatesCent    -0.536         0.211       -0.071   -2.542    0.011    -0.950    -0.123
## free.sulfur.dioxideCent     0.007         0.003        0.059    2.334    0.020     0.001     0.013
##      residual.sugarCent    -0.066         0.026       -0.073   -2.576    0.010    -0.116    -0.016
##              densityCent   49.122        24.490        0.073    2.006    0.045     1.087    97.158
##            chloridesCent    1.267         0.750        0.047    1.691    0.091    -0.203     2.738
## ----------------------------------------------------------------------------------------------------
```

```r
# view stepwise AIC two-direction selection
plot(stepwise)
```



The plot for stepwise selection looks the same as the plot for forward selection. Therefore, for all three model selections, we obtain the same result of including all variables from the full.model, except for volatile.acidityCent and pHCent.

```
# new model from model selection
selected <- lm(quality ~ residual.sugarCent + citric.acid + chloridesCent + free.sulfur.dioxideCent + su
```

```
# view coefficients of the new model
kable(tidy(selected), format="markdown")
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.1976490	0.0603138	69.596873	0.0000000
residual.sugarCent	-0.0659177	0.0255864	-2.576285	0.0100763
citric.acid	-0.7201660	0.1908611	-3.773246	0.0001670
chloridesCent	1.2672316	0.7495997	1.690544	0.0911197
free.sulfur.dioxideCent	0.0072279	0.0030962	2.334443	0.0196964
sulphatesCent	-0.5363162	0.2109507	-2.542377	0.0111042
densityCent	49.1224937	24.4896196	2.005850	0.0450413
alcoholCent	-0.1409656	0.0398305	-3.539138	0.0004130

We can see that chloridesCent has a p-value much greater than 0.05, so we consider chloridesCent to be a statistically insignificant predictor and remove it from the model.

```
# model after removing chloridesCent
model <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + densi
```

```
# view the model
kable(tidy(model), format="markdown")
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.1855088	0.0599196	69.852023	0.0000000
residual.sugarCent	-0.0619426	0.0254929	-2.429795	0.0152173
citric.acid	-0.6753641	0.1891225	-3.571040	0.0003661
free.sulfur.dioxideCent	0.0069545	0.0030938	2.247889	0.0247197
sulphatesCent	-0.4044692	0.1961185	-2.062372	0.0393342
densityCent	45.7026073	24.4201616	1.871511	0.0614575
alcoholCent	-0.1596179	0.0382941	-4.168207	0.0000324

We can see that the p-values of most of the explanatory variables are smaller than the threshold 0.05, which means there is sufficient evidence that these varaibles are significant predictors of wine quality. There is one exception, which is densityCent, whose p-value of 0.06 is slightly greater than 0.05. We remember that the p-value of densityCent is the selected model that included chloridesCent was 0.45 (already very close to 0.05). We believe 0.06 p-value is a tolerable level and will still keep densityCent in the model, since it helps with fitting the model.

```
# multicollinearity check
kable(tidy(vif(model)))
```

```
## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")
```

names	x
residual.sugarCent	1.339256
citric.acid	1.407016
free.sulfur.dioxideCent	1.085642
sulphatesCent	1.145623
densityCent	2.202034
alcoholCent	1.726386

A quick examination of the model using the VIF() function shows no noticeable detection of multicollinearity effects - all vif values are quite small, ranging around 1 to 2.

We will then examine possible interaction effects between the selected variables.

Nested F Tests on Interactions

```
# interaction effects with residual.sugarCent
model_rs <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + d
```

```
# Nested F Test for interaction with residual.sugarCent
ano1 <- anova(model_rs, model)
kable(tidy(ano1))
```

```
## Warning: Unknown or uninitialised column: 'term'.
```

res.df	rss	df	sumsq	statistic	p.value
1587	2447.619	NA	NA	NA	NA

res.df	rss	df	sumsq	statistic	p.value
1592	2454.089	-5	-6.470218	0.8390388	0.5218901

p-value is much higher than 0.05, so there is no significant interactions with residual.sugarCent that impact wine quality.

```
# Nested F Test for interaction effect between alcohol and density
model_da <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + d
ano2 <- anova(model_da, model)
kable(tidy(ano2))
```

```
## Warning: Unknown or uninitialised column: 'term'.
```

res.df	rss	df	sumsq	statistic	p.value
1591	2452.915	NA	NA	NA	NA
1592	2454.089	-1	-1.173729	0.7612995	0.3830538

p-value is much larger than 0.05, so there is no significant interaction effect between alcoholCent an densityCent.

```
# Nested F Test for all other two-way interactions
model1 <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + den
ano3 <- anova(model1, model)
kable(tidy(ano3))
```

```
## Warning: Unknown or uninitialised column: 'term'.
```

res.df	rss	df	sumsq	statistic	p.value
1583	2355.529	NA	NA	NA	NA
1592	2454.089	-9	-98.55976	7.359521	0

The p-value is extremely small, so there is at least one significant interaction effects. We will then investigate which specific interactions are significant using backward model selection.

```
# backward selection
backward_new <- ols_step_backward_aic(model1, details=TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . residual.sugarCent
## 2 . citric.acid
## 3 . free.sulfur.dioxideCent
## 4 . sulphatesCent
## 5 . densityCent
## 6 . alcoholCent
## 7 . citric.acid:free.sulfur.dioxideCent
## 8 . citric.acid:sulphatesCent
## 9 . citric.acid:densityCent
## 10 . citric.acid:alcoholCent
## 11 . free.sulfur.dioxideCent:sulphatesCent
```

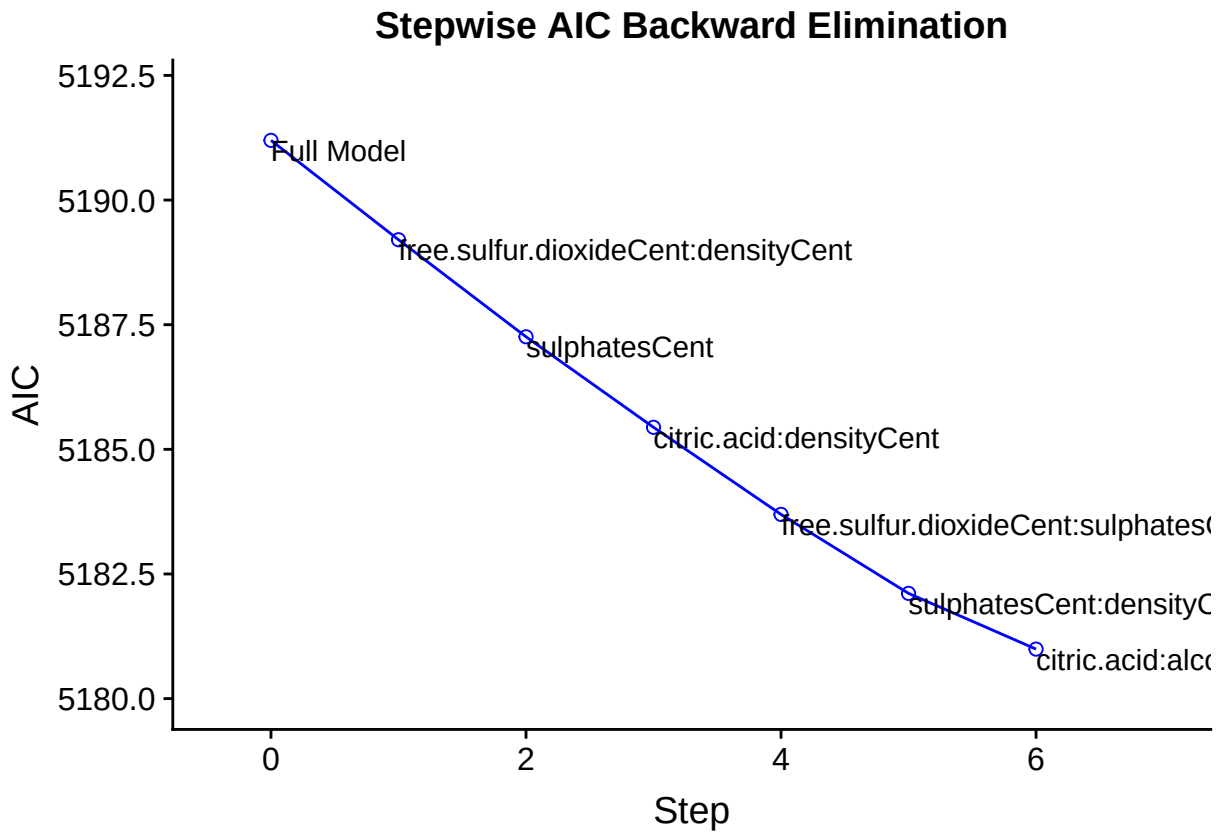
```
## 12 . free.sulfur.dioxideCent:densityCent
## 13 . free.sulfur.dioxideCent:alcoholCent
## 14 . sulphatesCent:densityCent
## 15 . sulphatesCent:alcoholCent
##
##  Step 0: AIC = 5191.197
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + densityCent +
##
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent:densityCent      1     5189.204     0.010    2355.539    0.092         0.084
## sulphatesCent                            1     5189.249     0.076    2355.605    0.092         0.084
## citric.acid:densityCent                  1     5189.348     0.222    2355.751    0.092         0.084
## sulphatesCent:densityCent                1     5189.368     0.251    2355.780    0.092         0.084
## free.sulfur.dioxideCent:sulphatesCent    1     5189.401     0.301    2355.830    0.092         0.084
## citric.acid:alcoholCent                  1     5189.595     0.586    2356.115    0.092         0.084
## densityCent                              1     5190.154     1.409    2356.939    0.091         0.083
## alcoholCent                              1     5191.894     3.976    2359.505    0.090         0.082
## residual.sugarCent                       1     5192.655     5.099    2360.628    0.090         0.082
## citric.acid:sulphatesCent                1     5195.135     8.763    2364.292    0.089         0.080
## citric.acid:free.sulfur.dioxideCent      1     5195.927     9.935    2365.464    0.088         0.080
## free.sulfur.dioxideCent:alcoholCent      1     5196.262    10.431    2365.960    0.088         0.080
## citric.acid                              1     5200.598    16.855    2372.384    0.085         0.077
## free.sulfur.dioxideCent                  1     5202.826    20.163    2375.692    0.084         0.076
## sulphatesCent:alcoholCent                1     5216.367    40.366    2395.895    0.076         0.068
## ---------------------------------------------------------------------------------------------------
##
##
## - free.sulfur.dioxideCent:densityCent
##
##
##   Step 1 : AIC = 5189.204
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + densityCent +
##
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## sulphatesCent                            1     5187.257     0.079    2355.618    0.092         0.084
## citric.acid:densityCent                  1     5187.364     0.235    2355.774    0.092         0.084
## sulphatesCent:densityCent                1     5187.372     0.248    2355.787    0.092         0.084
## free.sulfur.dioxideCent:sulphatesCent    1     5187.403     0.293    2355.832    0.092         0.084
## citric.acid:alcoholCent                  1     5187.595     0.576    2356.115    0.092         0.084
## densityCent                              1     5188.154     1.400    2356.939    0.091         0.084
## alcoholCent                              1     5189.925     4.011    2359.550    0.090         0.083
## residual.sugarCent                       1     5190.813     5.323    2360.862    0.090         0.082
## citric.acid:sulphatesCent                1     5193.136     8.755    2364.294    0.089         0.081
## citric.acid:free.sulfur.dioxideCent      1     5194.661    11.011    2366.550    0.088         0.080
## free.sulfur.dioxideCent:alcoholCent      1     5198.642    16.911    2372.450    0.085         0.078
## citric.acid                              1     5198.661    16.938    2372.477    0.085         0.078
## free.sulfur.dioxideCent                  1     5201.890    21.734    2377.273    0.084         0.076
## sulphatesCent:alcoholCent                1     5214.658    40.793    2396.332    0.076         0.069
## ---------------------------------------------------------------------------------------------------
##
```

```
## - sulphatesCent
##
##
##   Step 2 : AIC = 5187.257
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + densityCent + alcoholCent +
##
## ----------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------------------
## citric.acid:densityCent                  1     5185.442     0.272    2355.890    0.092        0.085
## free.sulfur.dioxideCent:sulphatesCent    1     5185.471     0.314    2355.932    0.092        0.085
## sulphatesCent:densityCent                1     5185.510     0.372    2355.990    0.092        0.085
## citric.acid:alcoholCent                  1     5185.641     0.565    2356.183    0.092        0.085
## densityCent                              1     5186.156     1.324    2356.942    0.091        0.085
## alcoholCent                              1     5188.164     4.285    2359.903    0.090        0.083
## residual.sugarCent                       1     5188.822     5.258    2360.876    0.090        0.083
## citric.acid:free.sulfur.dioxideCent      1     5192.670    10.946    2366.564    0.088        0.081
## free.sulfur.dioxideCent:alcoholCent      1     5196.651    16.845    2372.463    0.085        0.078
## citric.acid                              1     5196.856    17.150    2372.767    0.085        0.078
## free.sulfur.dioxideCent                  1     5199.914    21.691    2377.309    0.084        0.077
## citric.acid:sulphatesCent                1     5208.876    35.053    2390.671    0.078        0.071
## sulphatesCent:alcoholCent                1     5213.431    41.872    2397.490    0.076        0.069
## ----------------------------------------------------------------------------------------------------
##
## - citric.acid:densityCent
##
##
##   Step 3 : AIC = 5185.442
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + densityCent + alcoholCent +
##
## ----------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent:sulphatesCent    1     5183.695     0.373    2356.263    0.092        0.085
## sulphatesCent:densityCent                1     5183.915     0.697    2356.587    0.092        0.085
## citric.acid:alcoholCent                  1     5184.457     1.497    2357.386    0.091        0.085
## alcoholCent                              1     5186.494     4.501    2360.391    0.090        0.084
## residual.sugarCent                       1     5187.117     5.420    2361.310    0.090        0.083
## densityCent                              1     5188.648     7.683    2363.573    0.089        0.083
## citric.acid:free.sulfur.dioxideCent      1     5191.299    11.605    2367.495    0.087        0.081
## free.sulfur.dioxideCent:alcoholCent      1     5194.825    16.832    2372.721    0.085        0.079
## citric.acid                              1     5194.908    16.954    2372.844    0.085        0.079
## free.sulfur.dioxideCent                  1     5198.364    22.089    2377.979    0.083        0.077
## citric.acid:sulphatesCent                1     5207.093    35.105    2390.995    0.078        0.072
## sulphatesCent:alcoholCent                1     5212.354    42.984    2398.874    0.075        0.069
## ----------------------------------------------------------------------------------------------------
##
## - free.sulfur.dioxideCent:sulphatesCent
##
##
##   Step 4 : AIC = 5183.695
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + densityCent + alcoholCent +
##
## ----------------------------------------------------------------------------------------------------
```

```
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## sulphatesCent:densityCent               1     5182.111     0.613    2356.876    0.091         0.086
## citric.acid:alcoholCent                 1     5182.719     1.510    2357.772    0.091         0.085
## alcoholCent                             1     5184.611     4.300    2360.563    0.090         0.084
## residual.sugarCent                      1     5185.492     5.602    2361.865    0.089         0.084
## densityCent                             1     5187.246     8.194    2364.457    0.088         0.083
## citric.acid:free.sulfur.dioxideCent     1     5189.493    11.518    2367.781    0.087         0.081
## citric.acid                             1     5193.349    17.235    2373.498    0.085         0.079
## free.sulfur.dioxideCent:alcoholCent     1     5193.425    17.349    2373.612    0.085         0.079
## free.sulfur.dioxideCent                 1     5196.607    22.076    2378.339    0.083         0.077
## citric.acid:sulphatesCent               1     5205.119    34.771    2391.034    0.078         0.072
## sulphatesCent:alcoholCent               1     5212.030    45.128    2401.390    0.074         0.068
## ---------------------------------------------------------------------------------------------------
##
## - sulphatesCent:densityCent
##
##
##   Step 5 : AIC = 5182.111
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + densityCent + alcoholCent +
##
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## citric.acid:alcoholCent                 1     5180.993     1.300    2358.176    0.091         0.086
## alcoholCent                             1     5183.207     4.567    2361.443    0.090         0.084
## residual.sugarCent                      1     5183.769     5.398    2362.274    0.089         0.084
## densityCent                             1     5185.686     8.232    2365.108    0.088         0.083
## citric.acid:free.sulfur.dioxideCent     1     5187.644    11.129    2368.005    0.087         0.082
## free.sulfur.dioxideCent:alcoholCent     1     5191.709    17.157    2374.032    0.085         0.080
## citric.acid                             1     5191.925    17.478    2374.354    0.085         0.079
## free.sulfur.dioxideCent                 1     5194.802    21.753    2378.629    0.083         0.078
## citric.acid:sulphatesCent               1     5203.267    34.379    2391.255    0.078         0.073
## sulphatesCent:alcoholCent               1     5221.454    61.733    2418.609    0.068         0.062
## ---------------------------------------------------------------------------------------------------
##
## - citric.acid:alcoholCent
##
##
##   Step 6 : AIC = 5180.993
##  quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + densityCent + alcoholCent +
##
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq      RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## residual.sugarCent                      1     5182.880     5.740    2363.916    0.089         0.084
## densityCent                             1     5184.351     7.915    2366.091    0.088         0.083
## citric.acid:free.sulfur.dioxideCent     1     5186.574    11.207    2369.383    0.087         0.082
## free.sulfur.dioxideCent:alcoholCent     1     5191.230    18.117    2376.293    0.084         0.079
## citric.acid                             1     5191.425    18.406    2376.582    0.084         0.079
## alcoholCent                             1     5192.379    19.825    2378.001    0.083         0.079
## free.sulfur.dioxideCent                 1     5194.265    22.632    2380.807    0.082         0.078
## citric.acid:sulphatesCent               1     5202.118    34.353    2392.529    0.078         0.073
## sulphatesCent:alcoholCent               1     5226.572    71.223    2429.399    0.063         0.059
```

```
## ---------------------------------------------------------------------------------------------
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - free.sulfur.dioxideCent:densityCent
## - sulphatesCent
## - citric.acid:densityCent
## - free.sulfur.dioxideCent:sulphatesCent
## - sulphatesCent:densityCent
## - citric.acid:alcoholCent
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.302       RMSE                 1.218
## R-Squared               0.091       Coef. Var           30.436
## Adj. R-Squared          0.086       MSE                  1.484
## Pred R-Squared          0.076       MAE                  0.880
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     235.814           9         26.202    17.655    0.0000
## Residual      2358.176        1589          1.484
## Total         2593.990        1598
## --------------------------------------------------------------------
##
##                                                Parameter Estimates
## ----------------------------------------------------------------------------------------------
##                                model      Beta    Std. Error    Std. Beta      t        Sig      lowe
## ----------------------------------------------------------------------------------------------
##                           (Intercept)     4.230         0.058                72.489    0.000     4.11
##                  residual.sugarCent    -0.050         0.026       -0.056    -1.967    0.049    -0.10
##                           citric.acid    -0.656         0.186       -0.100    -3.522    0.000    -1.02
##              free.sulfur.dioxideCent     0.020         0.005        0.165     3.905    0.000     0.01
##                           densityCent    55.633        24.090        0.082     2.309    0.021     8.38
##                           alcoholCent    -0.139         0.038       -0.116    -3.655    0.000    -0.21
## citric.acid:free.sulfur.dioxideCent    -0.043         0.016       -0.118    -2.748    0.006    -0.07
##           citric.acid:sulphatesCent    -2.357         0.490       -0.130    -4.811    0.000    -3.31
## free.sulfur.dioxideCent:alcoholCent     0.010         0.003        0.087     3.494    0.000     0.00
##           alcoholCent:sulphatesCent    -1.339         0.193       -0.180    -6.928    0.000    -1.71
## ----------------------------------------------------------------------------------------------
```

```
# view the stepwise AIC backward elimination
plot(backward_new)
```



Backward model selection removes 5 interaction effects and sulphatesCent. However, since the model still includes some interaction effects with sulphatesCent, we will keep this main effect in the model.

```
# new model after backward selection
model2 <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent + dens
kable(tidy(model2), format="markdown")
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.2245320	0.0592509	71.2990038	0.0000000
residual.sugarCent	-0.0512202	0.0256054	-2.0003694	0.0456304
citric.acid	-0.6497755	0.1867955	-3.4785390	0.0005178
free.sulfur.dioxideCent	0.0204311	0.0051881	3.9380530	0.0000857
sulphatesCent	-0.1836422	0.3561499	-0.5156318	0.6061833
densityCent	57.4648324	24.3560389	2.3593669	0.0184266
alcoholCent	-0.1355440	0.0386503	-3.5069347	0.0004659
citric.acid:free.sulfur.dioxideCent	-0.0437867	0.0157370	-2.7824077	0.0054596
citric.acid:sulphatesCent	-1.9995045	0.8489262	-2.3553337	0.0186271
free.sulfur.dioxideCent:alcoholCent	0.0103289	0.0029290	3.5264524	0.0004331
sulphatesCent:alcoholCent	-1.3476692	0.1940602	-6.9445928	0.0000000

We can see that p-value of all the coefficients are less than the threshold of 0.05, except for sulphatesCent. Therefore, these main and interaction effects are significant predictors of wine quality (besides sulphatesCent). So we will make model2 our final model.

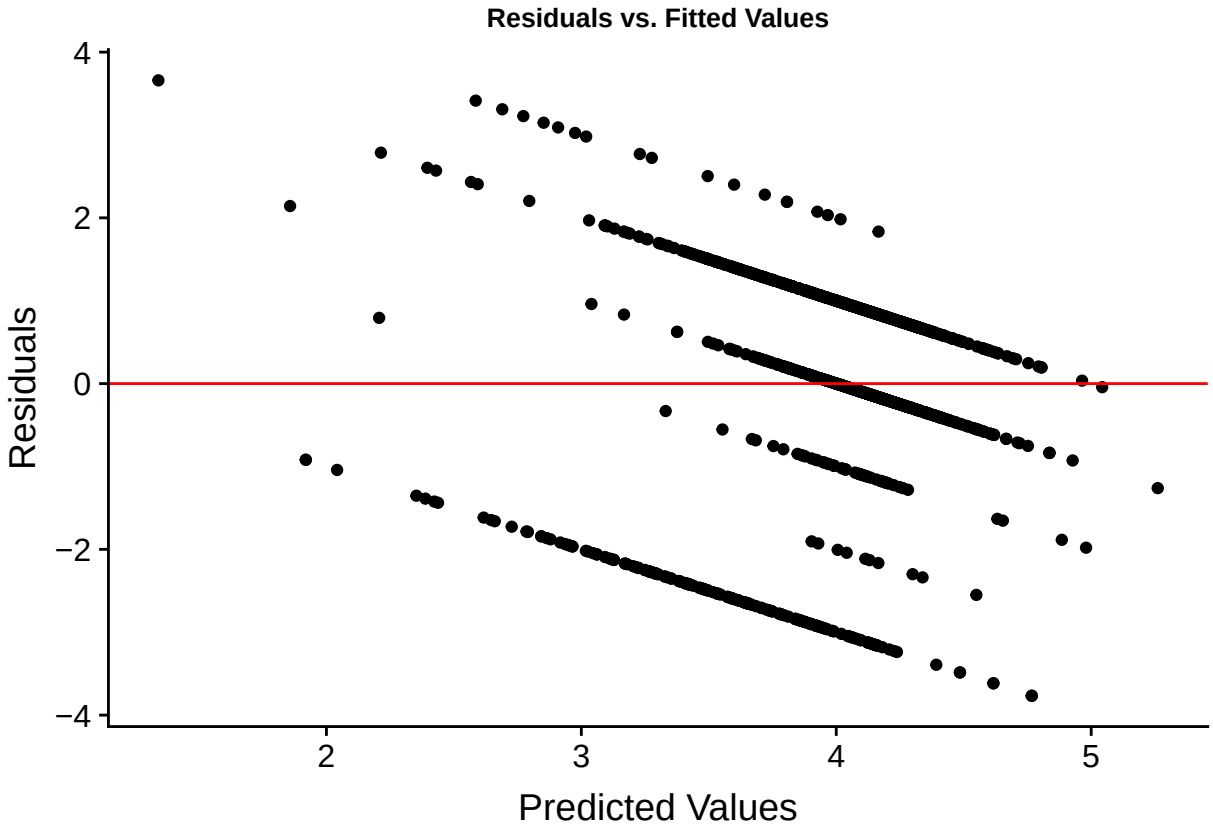
```
# make model2 the final multiple regression model
final.model <- lm(quality ~ residual.sugarCent + citric.acid + free.sulfur.dioxideCent + sulphatesCent +
```

Our final multiple reression model:
$quality = 4.2245 - 0.0512 * residual.sugarCent - 0.6498 * citric.acid + 0.0204 * free.sulfur.dioxideCent - 0.1836 * sulphatesCent + 57.4648 * densityCent - 0.1355 * alcoholCent - 0.0438 * citric.acid * free.sulfur.dioxideCent - 1.9995 * citric.acid * sulphatesCent + 0.0103 * free.sulfur.dioxideCent * alcoholCent - 1.3477 * sulphatesCent * alcoholCent$

Assumptions

Residual Plots

```
data <- data %>% mutate(predicted = predict.lm(final.model))
data <- data %>% mutate(resid=resid(final.model))
ggplot(data, aes(predicted, resid)) + geom_point() +
  geom_hline(yintercept=0, color="red") + labs(x="Predicted Values", y="Residuals",
  title="Residuals vs. Fitted Values") + theme(plot.title=element_text(hjust=0.5,
  size=10))
```

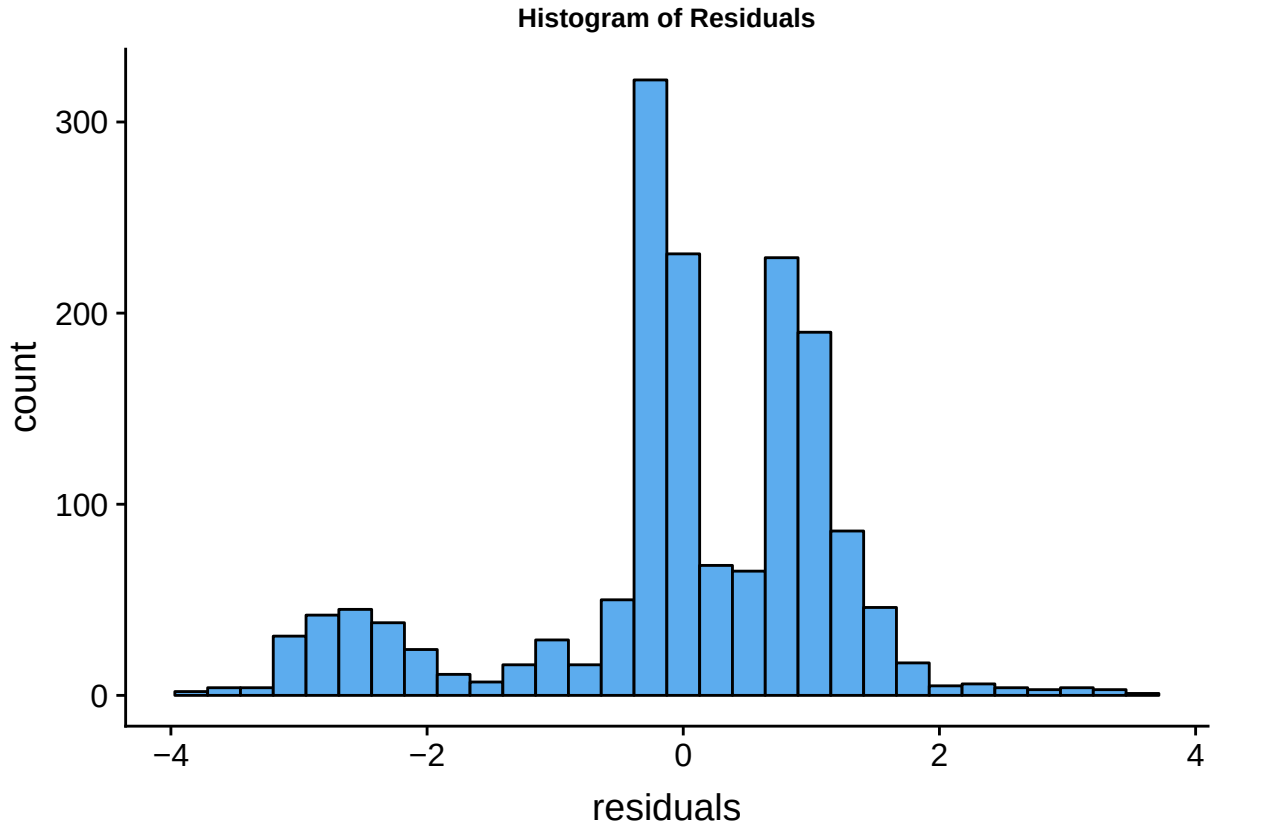


Normaltiy

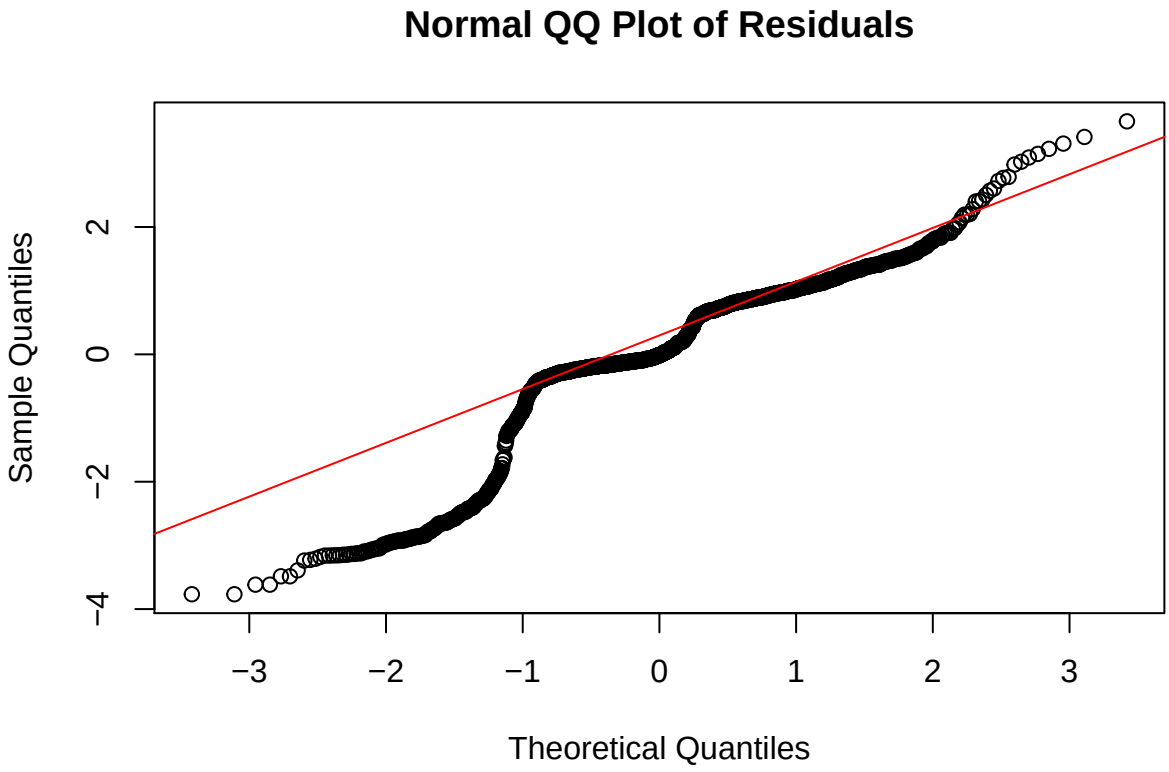
```
# distribution of residuals
ggplot(data, aes(resid)) + geom_histogram(color="black",
```

```
  fill="steelblue2") + labs(x="residuals", y="count", title="Histogram of Residuals") +
  theme(plot.title=element_text(hjust=0.5, size=10))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# QQ Plot
qqnorm(data$resid, main="Normal QQ Plot of Residuals")
qqline(data$resid, col="red")
```



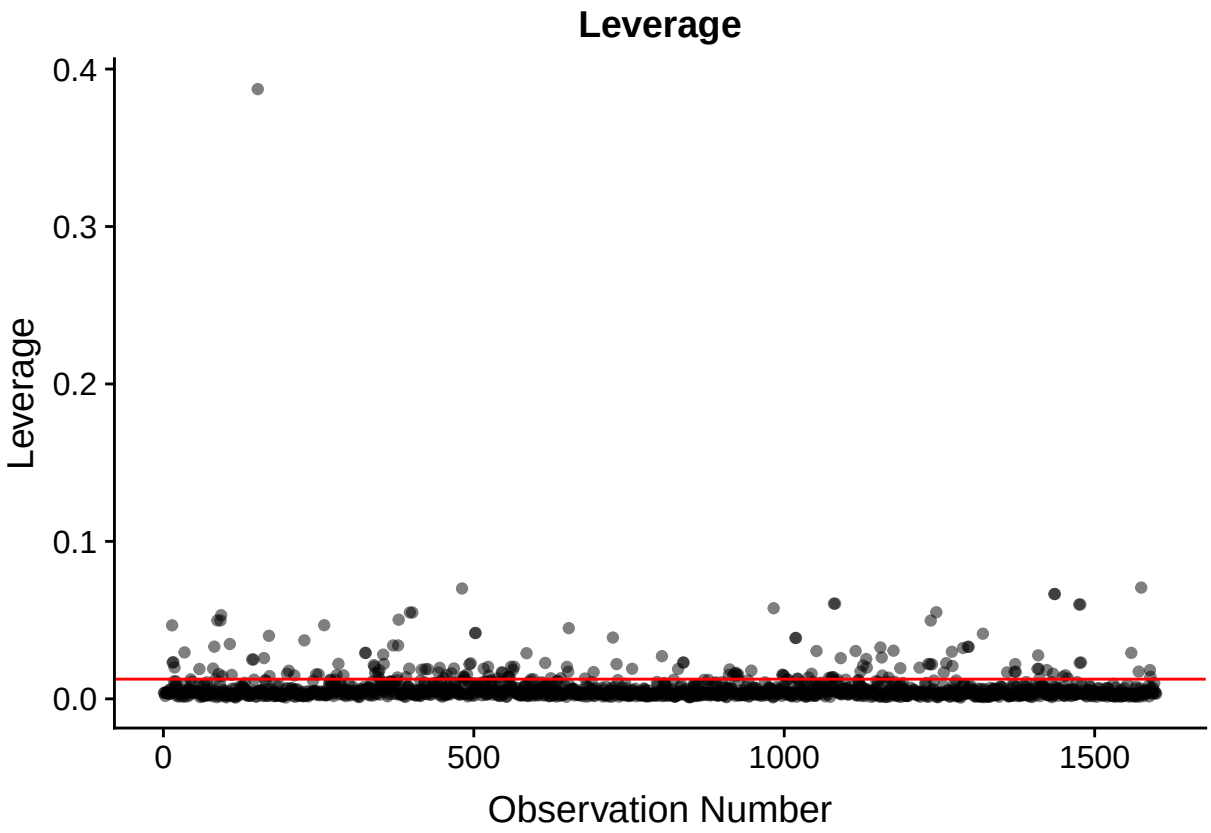
The histogram is unimodal and fairly symmetric. In the QQ plot, we can see most of the points lie right on top of the red diagonal line, with some deviation on the sides, especially the left side. Overall, the Normaltiy Assumption is well satisfied.

Influence Points

```
data <- data %>% mutate(leverage = hatvalues(final.model), cooks =
  cooks.distance(final.model), stand.resid = rstandard(final.model), obs.num =
  row_number())
```

```
# calculate leverage threshold
t <- 2*10/1599
```

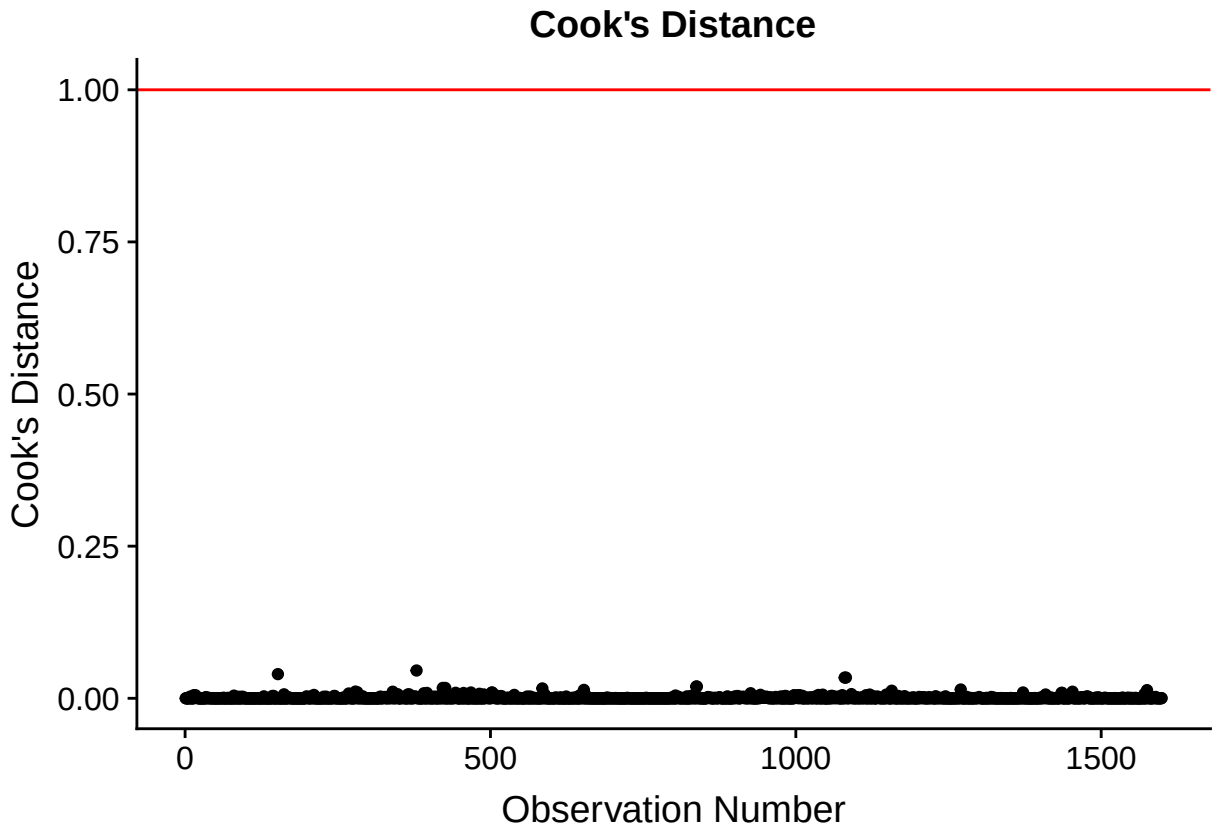
```
# leverage
ggplot(data=data, aes(x=obs.num, y=leverage)) +
  geom_point(alpha=0.5) +
  geom_hline(yintercept=t, color="red") +
  labs(x="Observation Number", y="Leverage", title="Leverage")
```



We can see that there is one point with a significantly high leverage around 0.4, comparing with other observations. This could be an outlier and might be an influence point.

```
#data %>% filter(leverage > 0.3) %>%
  #select(citric.acid, residual.sugarCent, free.sulfur.dioxideCent, sulphatesCent, densityCent, alcohol
```

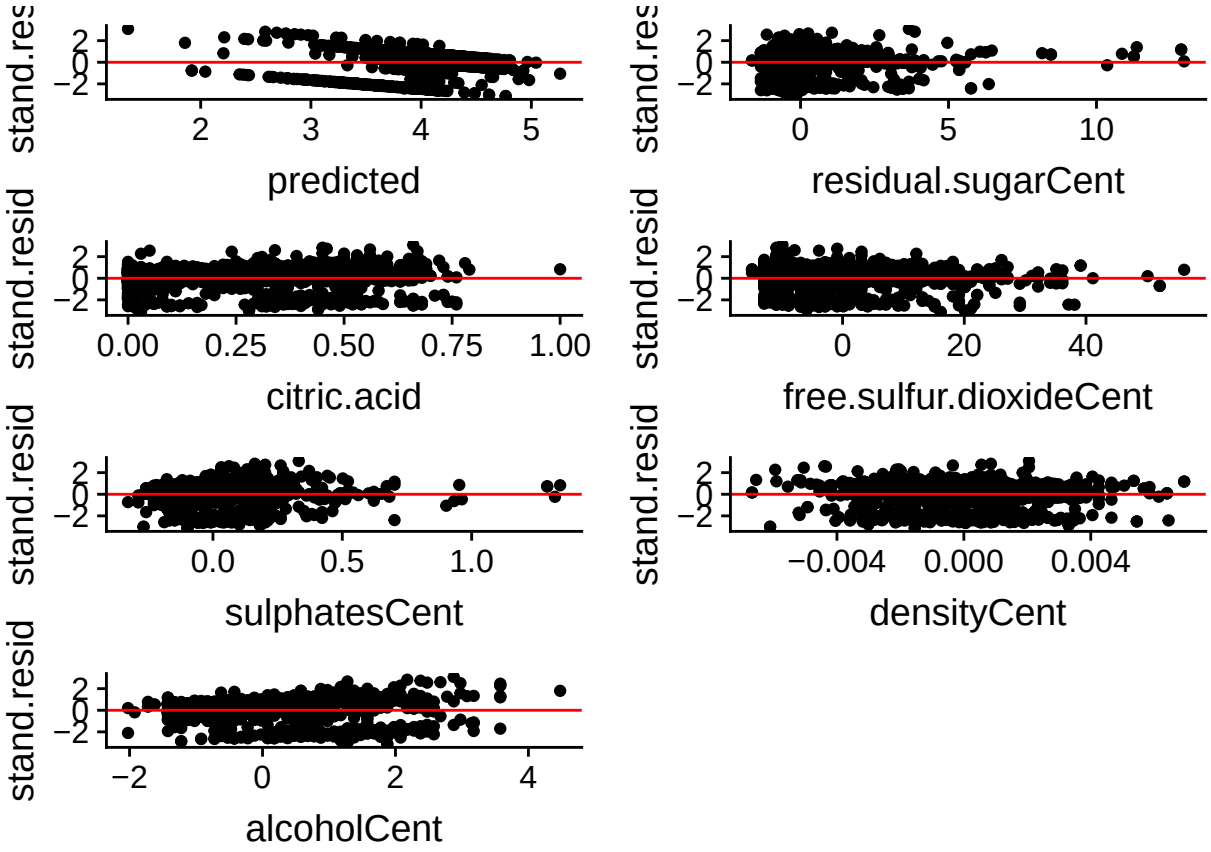
```
# Cook's Distance
ggplot(data=data, aes(x=obs.num, y=cooks)) +
  geom_point() +
  geom_hline(yintercept=1, color="red")+
  labs(x="Observation Number", y="Cook's Distance", title="Cook's Distance")
```



The Cook's Distance for all observations are far below the threshold of 1.

```
# standardized residuals
p1 <- ggplot(data=data, aes(x=predicted, y=stand.resid)) + geom_point()+  geom_hline(yintercept=0, colo
p2 <- ggplot(data=data, aes(x=residual.sugarCent, y=stand.resid)) + geom_point()+  geom_hline(yintercep
p3 <- ggplot(data=data, aes(x=citric.acid, y=stand.resid)) + geom_point()+  geom_hline(yintercept=0, co
p4 <- ggplot(data=data, aes(x=free.sulfur.dioxideCent, y=stand.resid)) + geom_point()+  geom_hline(yint
p5 <- ggplot(data=data, aes(x=sulphatesCent, y=stand.resid)) + geom_point()+  geom_hline(yintercept=0,
p6 <- ggplot(data=data, aes(x=densityCent, y=stand.resid)) + geom_point()+  geom_hline(yintercept=0, co
p7 <- ggplot(data=data, aes(x=alcoholCent, y=stand.resid)) + geom_point()+  geom_hline(yintercept=0, co

plot_grid(p1, p2, p3, p4, p5, p6, p7, ncol=2, nrow=4)
```



The standardized residuals show some points with magnitude greater than 2, but overall, in combination with our observation from Cook's Distance and just one data point with high leverage away from other points, we can conclude there isn't any obvious influential points in this model.

Multicollinearity

```
kable(tidy(vif(final.model)))
```

```
## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")
```

names	x
residual.sugarCent	1.402748
citric.acid	1.425081
free.sulfur.dioxideCent	3.169724
sulphatesCent	3.922514
densityCent	2.274231
alcoholCent	1.825881
citric.acid:free.sulfur.dioxideCent	3.249458
citric.acid:sulphatesCent	3.832337
free.sulfur.dioxideCent:alcoholCent	1.097922
sulphatesCent:alcoholCent	1.187895

Additional Work

```
# compare model from stepwise selection with final model from backward selection
stepwise_new <- ols_step_both_aic(model1, details=TRUE)
```

```
## Stepwise Selection Method
## -------------------------
##
## Candidate Terms:
##
## 1 . residual.sugarCent
## 2 . citric.acid
## 3 . free.sulfur.dioxideCent
## 4 . sulphatesCent
## 5 . densityCent
## 6 . alcoholCent
## 7 . citric.acid:free.sulfur.dioxideCent
## 8 . citric.acid:sulphatesCent
## 9 . citric.acid:densityCent
## 10 . citric.acid:alcoholCent
## 11 . free.sulfur.dioxideCent:sulphatesCent
## 12 . free.sulfur.dioxideCent:densityCent
## 13 . free.sulfur.dioxideCent:alcoholCent
## 14 . sulphatesCent:densityCent
## 15 . sulphatesCent:alcoholCent
##
##  Step 0: AIC = 5315.392
##  quality ~ 1
##
##
##                                        Enter New Variables
## ---------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------
## citric.acid:alcoholCent                  1    5252.279    103.508    2490.482    0.040         0.039
## alcoholCent                              1    5258.390     93.971    2500.019    0.036         0.036
## sulphatesCent:alcoholCent                1    5291.643     41.437    2552.553    0.016         0.015
## citric.acid                              1    5292.863     39.488    2554.502    0.015         0.015
## citric.acid:sulphatesCent                1    5303.179     22.954    2571.036    0.009         0.008
## sulphatesCent                            1    5305.527     19.176    2574.814    0.007         0.007
## free.sulfur.dioxideCent:alcoholCent      1    5307.604     15.829    2578.161    0.006         0.005
## densityCent                              1    5311.101     10.185    2583.805    0.004         0.003
## free.sulfur.dioxideCent                  1    5312.446      8.011    2585.979    0.003         0.002
## residual.sugarCent                       1    5312.626      7.720    2586.270    0.003         0.002
## citric.acid:densityCent                  1    5313.982      5.526    2588.464    0.002         0.002
## free.sulfur.dioxideCent:densityCent      1    5314.283      5.037    2588.953    0.002         0.001
## sulphatesCent:densityCent                1    5315.451      3.146    2590.844    0.001         0.001
## citric.acid:free.sulfur.dioxideCent      1    5316.525      1.406    2592.584    0.001         0.000
## free.sulfur.dioxideCent:sulphatesCent    1    5317.389      0.004    2593.986    0.000        -0.001
## ---------------------------------------------------------------------------------------------
##
## - citric.acid:alcoholCent added
##
##
```

```
##  Step 1 : AIC = 5252.279
##  quality ~ citric.acid:alcoholCent
##
##                                          Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## citric.acid:sulphatesCent                1    5239.063    127.094    2466.896    0.049         0.048
## citric.acid                              1    5240.154    125.411    2468.579    0.048         0.047
## sulphatesCent:alcoholCent                1    5242.098    122.408    2471.582    0.047         0.046
## sulphatesCent                            1    5244.293    119.012    2474.978    0.046         0.045
## free.sulfur.dioxideCent:alcoholCent      1    5247.491    114.058    2479.932    0.044         0.043
## free.sulfur.dioxideCent:densityCent      1    5248.462    112.552    2481.438    0.043         0.042
## alcoholCent                              1    5250.417    109.515    2484.475    0.042         0.041
## residual.sugarCent                       1    5252.168    106.793    2487.197    0.041         0.040
## free.sulfur.dioxideCent                  1    5252.306    106.579    2487.411    0.041         0.040
## sulphatesCent:densityCent                1    5252.722    105.932    2488.058    0.041         0.040
## citric.acid:densityCent                  1    5253.692    104.422    2489.568    0.040         0.039
## citric.acid:free.sulfur.dioxideCent      1    5254.226    103.591    2490.399    0.040         0.039
## free.sulfur.dioxideCent:sulphatesCent    1    5254.267    103.527    2490.463    0.040         0.039
## densityCent                              1    5254.275    103.513    2490.477    0.040         0.039
## ---------------------------------------------------------------------------------------------------
##
## - citric.acid:sulphatesCent added
##
##
##  Step 2 : AIC = 5239.063
##  quality ~ citric.acid:alcoholCent + citric.acid:sulphatesCent
##
##                                Remove Existing Variables
## ------------------------------------------------------------------------------------
## Variable                      DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------
## citric.acid:sulphatesCent      1    5252.279    103.508    2490.482    0.040         0.039
## citric.acid:alcoholCent        1    5303.179     22.954    2571.036    0.009         0.008
## ------------------------------------------------------------------------------------
##
##                                          Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## sulphatesCent:alcoholCent                1    5213.741    168.889    2425.101    0.065         0.063
## sulphatesCent:densityCent                1    5233.111    139.333    2454.657    0.054         0.052
## citric.acid                              1    5233.373    138.930    2455.060    0.054         0.052
## free.sulfur.dioxideCent:alcoholCent      1    5234.315    137.483    2456.507    0.053         0.051
## free.sulfur.dioxideCent:densityCent      1    5236.242    134.521    2459.469    0.052         0.050
## alcoholCent                              1    5238.203    131.502    2462.487    0.051         0.049
## free.sulfur.dioxideCent                  1    5238.630    130.845    2463.145    0.050         0.049
## residual.sugarCent                       1    5239.014    130.254    2463.736    0.050         0.048
## free.sulfur.dioxideCent:sulphatesCent    1    5240.551    127.884    2466.106    0.049         0.048
## densityCent                              1    5240.854    127.417    2466.573    0.049         0.047
## sulphatesCent                            1    5240.969    127.240    2466.750    0.049         0.047
## citric.acid:densityCent                  1    5241.053    127.110    2466.880    0.049         0.047
## citric.acid:free.sulfur.dioxideCent      1    5241.057    127.105    2466.885    0.049         0.047
```

```
## ---------------------------------------------------------------------------------------------------
##
## - sulphatesCent:alcoholCent added
##
##
##  Step 3 : AIC = 5213.741
##  quality ~ citric.acid:alcoholCent + citric.acid:sulphatesCent + sulphatesCent:alcoholCent
##
##                                Remove Existing Variables
## ---------------------------------------------------------------------------------
## Variable                     DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------
## sulphatesCent:alcoholCent     1    5239.063    127.094    2466.896    0.049         0.048
## citric.acid:sulphatesCent     1    5242.098    122.408    2471.582    0.047         0.046
## citric.acid:alcoholCent       1    5258.357     97.148    2496.842    0.037         0.036
## ---------------------------------------------------------------------------------
##
##                                       Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent:alcoholCent       1    5205.419    184.492    2409.498    0.071         0.069
## citric.acid                               1    5208.193    180.309    2413.681    0.070         0.067
## alcoholCent                               1    5209.383    178.512    2415.478    0.069         0.066
## free.sulfur.dioxideCent:densityCent       1    5209.851    177.805    2416.185    0.069         0.066
## free.sulfur.dioxideCent                   1    5211.959    174.617    2419.373    0.067         0.065
## residual.sugarCent                        1    5213.343    172.522    2421.468    0.067         0.064
## densityCent                               1    5214.581    170.647    2423.343    0.066         0.063
## sulphatesCent:densityCent                 1    5215.364    169.460    2424.530    0.065         0.063
## citric.acid:densityCent                   1    5215.500    169.254    2424.736    0.065         0.063
## citric.acid:free.sulfur.dioxideCent       1    5215.632    169.054    2424.936    0.065         0.063
## free.sulfur.dioxideCent:sulphatesCent     1    5215.714    168.930    2425.060    0.065         0.063
## sulphatesCent                             1    5215.741    168.889    2425.101    0.065         0.063
## ---------------------------------------------------------------------------------------------------
##
## - free.sulfur.dioxideCent:alcoholCent added
##
##
##  Step 4 : AIC = 5205.419
##  quality ~ citric.acid:alcoholCent + citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sul
##
##                                    Remove Existing Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent:alcoholCent       1    5213.741    168.889    2425.101    0.065         0.063
## sulphatesCent:alcoholCent                 1    5234.315    137.483    2456.507    0.053         0.051
## citric.acid:sulphatesCent                 1    5235.125    136.238    2457.752    0.053         0.051
## citric.acid:alcoholCent                   1    5245.882    119.649    2474.341    0.046         0.044
## ---------------------------------------------------------------------------------------------------
##
##                                       Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
```

```
## ---------------------------------------------------------------------------------------------------
## alcoholCent                                1    5199.748    196.024    2397.966    0.076        0.073
## citric.acid                                1    5201.365    193.598    2400.392    0.075        0.072
## free.sulfur.dioxideCent                    1    5202.608    191.731    2402.259    0.074        0.071
## densityCent                                1    5204.982    188.163    2405.827    0.073        0.070
## citric.acid:densityCent                    1    5206.663    185.631    2408.359    0.072        0.069
## residual.sugarCent                         1    5206.676    185.612    2408.378    0.072        0.069
## free.sulfur.dioxideCent:densityCent        1    5206.740    185.515    2408.475    0.072        0.069
## sulphatesCent:densityCent                  1    5206.840    185.365    2408.625    0.071        0.069
## citric.acid:free.sulfur.dioxideCent        1    5207.327    184.632    2409.358    0.071        0.068
## sulphatesCent                              1    5207.362    184.578    2409.412    0.071        0.068
## free.sulfur.dioxideCent:sulphatesCent      1    5207.366    184.572    2409.418    0.071        0.068
## ---------------------------------------------------------------------------------------------------
##
## - alcoholCent added
##
##
##  Step 5 : AIC = 5199.748
##  quality ~ citric.acid:alcoholCent + citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.su
##
##                                    Remove Existing Variables
## ------------------------------------------------------------------------------------------------
## Variable                                DF       AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------------------
## citric.acid:alcoholCent                  1    5200.092    192.507    2401.483    0.074        0.072
## alcoholCent                              1    5205.419    184.492    2409.498    0.071        0.069
## free.sulfur.dioxideCent:alcoholCent      1    5209.383    178.512    2415.478    0.069        0.066
## citric.acid:sulphatesCent                1    5229.161    148.450    2445.540    0.057        0.055
## sulphatesCent:alcoholCent                1    5232.931    142.676    2451.314    0.055        0.053
## ------------------------------------------------------------------------------------------------
##
##                                      Enter New Variables
## ------------------------------------------------------------------------------------------------
## Variable                                DF       AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------------------------
## citric.acid                              1    5194.836    206.368    2387.622    0.080        0.076
## free.sulfur.dioxideCent                  1    5196.351    204.104    2389.886    0.079        0.075
## residual.sugarCent                       1    5200.731    197.548    2396.442    0.076        0.073
## free.sulfur.dioxideCent:densityCent      1    5201.007    197.135    2396.855    0.076        0.073
## citric.acid:densityCent                  1    5201.292    196.708    2397.282    0.076        0.072
## sulphatesCent:densityCent                1    5201.488    196.414    2397.576    0.076        0.072
## densityCent                              1    5201.571    196.290    2397.700    0.076        0.072
## citric.acid:free.sulfur.dioxideCent      1    5201.644    196.180    2397.810    0.076        0.072
## free.sulfur.dioxideCent:sulphatesCent    1    5201.729    196.052    2397.938    0.076        0.072
## sulphatesCent                            1    5201.742    196.034    2397.956    0.076        0.072
## ------------------------------------------------------------------------------------------------
##
## - citric.acid added
##
##
##  Step 6 : AIC = 5194.836
##  quality ~ citric.acid:alcoholCent + citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.su
##
##                                    Remove Existing Variables
```

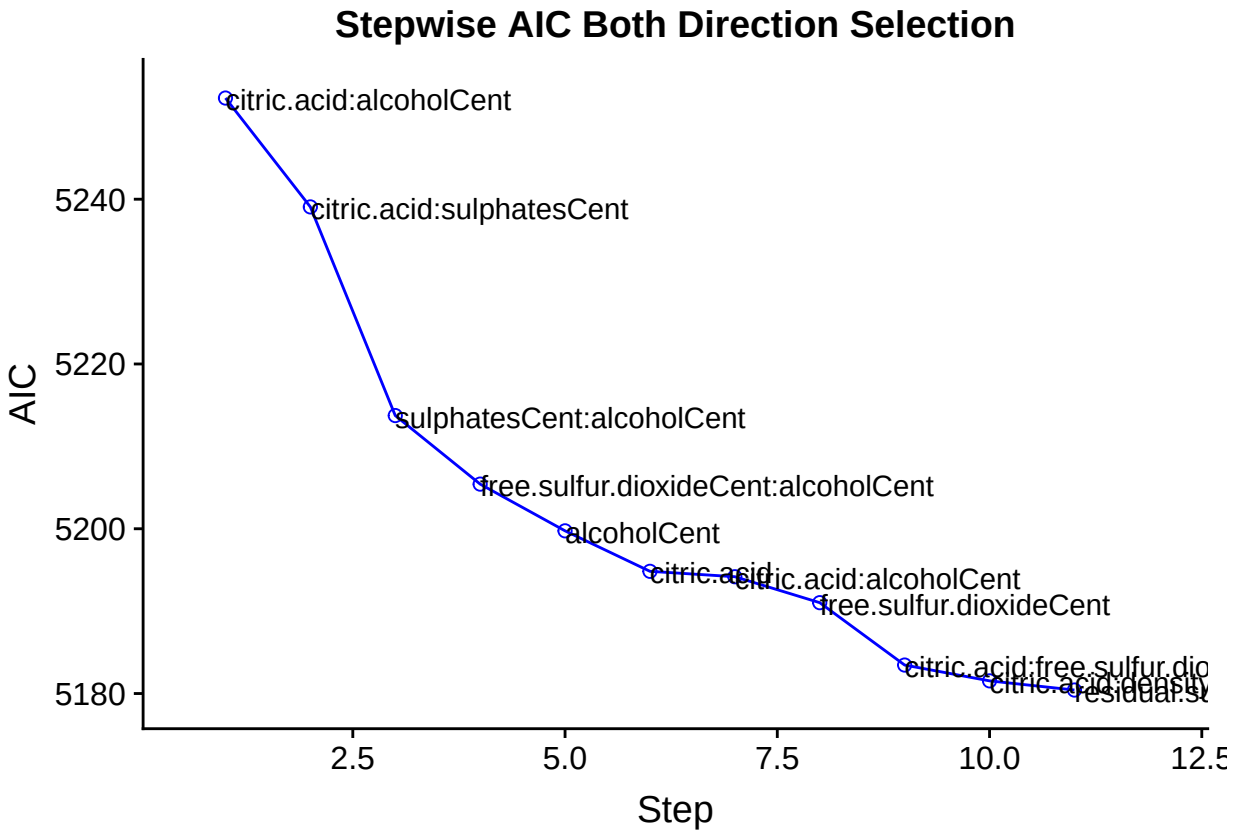
```
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## citric.acid:alcoholCent                  1    5194.178    204.363    2389.627    0.079         0.076
## citric.acid                              1    5199.748    196.024    2397.966    0.076         0.073
## alcoholCent                              1    5201.365    193.598    2400.392    0.075         0.072
## free.sulfur.dioxideCent:alcoholCent      1    5202.861    191.351    2402.639    0.074         0.071
## citric.acid:sulphatesCent                1    5214.775    173.382    2420.608    0.067         0.064
## sulphatesCent:alcoholCent                1    5227.838    153.526    2440.464    0.059         0.056
## ---------------------------------------------------------------------------------------------------
##
## - citric.acid:alcoholCent removed
##
##
##  Step 7 : AIC = 5194.178
##  quality ~ citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sulfur.dioxideCent:alcoholCen
##
##                                          Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent                  1    5191.029    212.046    2381.944    0.082         0.078
## citric.acid:densityCent                  1    5191.652    211.117    2382.873    0.081         0.078
## densityCent                              1    5193.418    208.483    2385.507    0.080         0.077
## free.sulfur.dioxideCent:densityCent      1    5194.921    206.241    2387.749    0.080         0.076
## residual.sugarCent                       1    5195.544    205.310    2388.680    0.079         0.076
## sulphatesCent                            1    5196.079    204.511    2389.479    0.079         0.075
## sulphatesCent:densityCent                1    5196.112    204.462    2389.528    0.079         0.075
## free.sulfur.dioxideCent:sulphatesCent    1    5196.131    204.432    2389.558    0.079         0.075
## citric.acid:free.sulfur.dioxideCent      1    5196.169    204.376    2389.614    0.079         0.075
## ---------------------------------------------------------------------------------------------------
##
## - free.sulfur.dioxideCent added
##
##
##  Step 8 : AIC = 5191.029
##  quality ~ citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sulfur.dioxideCent:alcoholCen
##
##                                       Remove Existing Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## free.sulfur.dioxideCent                  1    5194.178    204.363    2389.627    0.079         0.076
## citric.acid                              1    5195.968    201.687    2392.303    0.078         0.075
## free.sulfur.dioxideCent:alcoholCent      1    5200.974    194.185    2399.805    0.075         0.072
## citric.acid:sulphatesCent                1    5212.469    176.871    2417.119    0.068         0.065
## alcoholCent                              1    5230.898    148.852    2445.138    0.057         0.054
## sulphatesCent:alcoholCent                1    5232.007    147.155    2446.835    0.057         0.054
## ---------------------------------------------------------------------------------------------------
##
##                                          Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
```

```
## citric.acid:free.sulfur.dioxideCent       1    5183.459    226.259    2367.731    0.087        0.083
## citric.acid:densityCent                   1    5187.880    219.703    2374.287    0.085        0.081
## densityCent                               1    5189.775    216.888    2377.102    0.084        0.080
## residual.sugarCent                        1    5191.441    214.410    2379.580    0.083        0.079
## free.sulfur.dioxideCent:densityCent       1    5191.500    214.322    2379.668    0.083        0.079
## free.sulfur.dioxideCent:sulphatesCent     1    5192.962    212.145    2381.845    0.082        0.078
## sulphatesCent                             1    5193.000    212.088    2381.902    0.082        0.078
## sulphatesCent:densityCent                 1    5193.002    212.086    2381.904    0.082        0.078
## ----------------------------------------------------------------------------------------------------
##
## - citric.acid:free.sulfur.dioxideCent added
##
##
##  Step 9 : AIC = 5183.459
##  quality ~ citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sulfur.dioxideCent:alcoholCen
##
##                                    Remove Existing Variables
## ----------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------------------
## citric.acid                              1    5190.406    212.972    2381.018    0.082        0.079
## citric.acid:free.sulfur.dioxideCent      1    5191.029    212.046    2381.944    0.082        0.078
## free.sulfur.dioxideCent:alcoholCent      1    5195.788    204.945    2389.045    0.079        0.076
## free.sulfur.dioxideCent                  1    5196.169    204.376    2389.614    0.079        0.075
## citric.acid:sulphatesCent                1    5201.881    195.825    2398.165    0.075        0.072
## sulphatesCent:alcoholCent                1    5226.052    159.298    2434.692    0.061        0.058
## alcoholCent                              1    5226.707    158.300    2435.690    0.061        0.057
## ----------------------------------------------------------------------------------------------------
##
##                                      Enter New Variables
## ----------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------------------
## citric.acid:densityCent                  1    5181.540    232.055    2361.935    0.089        0.085
## densityCent                              1    5182.880    230.074    2363.916    0.089        0.084
## residual.sugarCent                       1    5184.351    227.899    2366.091    0.088        0.083
## free.sulfur.dioxideCent:sulphatesCent    1    5184.915    227.064    2366.926    0.088        0.083
## sulphatesCent:densityCent                1    5185.241    226.582    2367.408    0.087        0.083
## sulphatesCent                            1    5185.433    226.297    2367.693    0.087        0.083
## free.sulfur.dioxideCent:densityCent      1    5185.452    226.269    2367.721    0.087        0.083
## ----------------------------------------------------------------------------------------------------
##
## - citric.acid:densityCent added
##
##
##  Step 10 : AIC = 5181.54
##  quality ~ citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sulfur.dioxideCent:alcoholCen
##
##                                    Remove Existing Variables
## ----------------------------------------------------------------------------------------------------
## Variable                                DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------------------------------
## citric.acid:densityCent                  1    5183.459    226.259    2367.731    0.087        0.083
## citric.acid:free.sulfur.dioxideCent      1    5187.880    219.703    2374.287    0.085        0.081
```

```
## citric.acid                              1    5191.838    213.819    2380.171    0.082        0.078
## free.sulfur.dioxideCent                  1    5193.551    211.267    2382.723    0.081        0.077
## free.sulfur.dioxideCent:alcoholCent      1    5194.612    209.686    2384.304    0.081        0.077
## citric.acid:sulphatesCent                1    5201.770    198.988    2395.002    0.077        0.073
## alcoholCent                              1    5209.379    187.565    2406.425    0.072        0.068
## sulphatesCent:alcoholCent                1    5225.648    162.955    2431.035    0.063        0.059
## ---------------------------------------------------------------------------------------------------
##
##                                        Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## residual.sugarCent                        1    5180.426    236.650    2357.340    0.091        0.086
## free.sulfur.dioxideCent:sulphatesCent     1    5183.275    232.447    2361.543    0.090        0.084
## free.sulfur.dioxideCent:densityCent       1    5183.508    232.102    2361.888    0.089        0.084
## sulphatesCent                             1    5183.523    232.079    2361.911    0.089        0.084
## sulphatesCent:densityCent                 1    5183.524    232.078    2361.912    0.089        0.084
## densityCent                               1    5183.534    232.064    2361.926    0.089        0.084
## ---------------------------------------------------------------------------------------------------
##
## - residual.sugarCent added
##
##
##  Step 11 : AIC = 5180.426
##  quality ~ citric.acid:sulphatesCent + sulphatesCent:alcoholCent + free.sulfur.dioxideCent:alcoholCen
##
##                                     Remove Existing Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## residual.sugarCent                        1    5181.540    232.055    2361.935    0.089        0.085
## citric.acid:densityCent                   1    5184.351    227.899    2366.091    0.088        0.083
## citric.acid:free.sulfur.dioxideCent       1    5185.670    225.947    2368.043    0.087        0.083
## free.sulfur.dioxideCent:alcoholCent       1    5190.246    219.160    2374.830    0.084        0.080
## citric.acid                               1    5190.406    218.923    2375.067    0.084        0.080
## free.sulfur.dioxideCent                   1    5193.017    215.041    2378.949    0.083        0.078
## citric.acid:sulphatesCent                 1    5201.961    201.697    2392.293    0.078        0.073
## alcoholCent                               1    5204.448    197.973    2396.017    0.076        0.072
## sulphatesCent:alcoholCent                 1    5224.961    167.038    2426.952    0.064        0.060
## ---------------------------------------------------------------------------------------------------
##
##                                        Enter New Variables
## ---------------------------------------------------------------------------------------------------
## Variable                                 DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------------------------------
## densityCent                               1    5181.920    237.396    2356.594    0.092        0.086
## free.sulfur.dioxideCent:sulphatesCent     1    5182.248    236.913    2357.077    0.091        0.086
## sulphatesCent:densityCent                 1    5182.394    236.697    2357.293    0.091        0.086
## sulphatesCent                             1    5182.405    236.681    2357.309    0.091        0.086
## free.sulfur.dioxideCent:densityCent       1    5182.410    236.673    2357.317    0.091        0.086
## ---------------------------------------------------------------------------------------------------
##
##
## No more variables to be added or removed.
```

```
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                        0.302       RMSE                 1.218
## R-Squared                0.091       Coef. Var           30.431
## Adj. R-Squared           0.086       MSE                  1.484
## Pred R-Squared           0.076       MAE                  0.879
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression     236.650           9         26.294    17.724    0.0000
## Residual      2357.340        1589          1.484
## Total         2593.990        1598
## --------------------------------------------------------------------------
##
##                                            Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                              model       Beta    Std. Error    Std. Beta      t        Sig      lowe
## ----------------------------------------------------------------------------------------------------
##                         (Intercept)      4.202         0.055                76.846    0.000     4.09
##                         alcoholCent     -0.163         0.032       -0.137   -5.106    0.000    -0.22
##                         citric.acid     -0.618         0.179       -0.094   -3.457    0.001    -0.96
##             free.sulfur.dioxideCent      0.020         0.005        0.161    3.817    0.000     0.01
##                  residual.sugarCent     -0.042         0.024       -0.047   -1.760    0.079    -0.09
##           citric.acid:sulphatesCent     -2.382         0.491       -0.131   -4.854    0.000    -3.34
##           sulphatesCent:alcoholCent     -1.318         0.192       -0.177   -6.850    0.000    -1.69
## alcoholCent:free.sulfur.dioxideCent      0.010         0.003        0.086    3.434    0.001     0.00
## citric.acid:free.sulfur.dioxideCent     -0.042         0.016       -0.116   -2.686    0.007    -0.07
##             citric.acid:densityCent    128.418        52.874        0.071    2.429    0.015    24.70
## ----------------------------------------------------------------------------------------------------
```

```
# view the stepwise AIC backward elimination
plot(stepwise_new)
```



Logistic Regression

```
rawdata <- read.csv("./redwine_quality.csv")
data$quality <- rawdata$quality
data$quality[data$quality < 6] <- 0
data$quality[data$quality >= 6] <- 1

# Model Reduced
modellg1.null <- glm(quality ~ 1, data=data , family=binomial)

# Model Full
modellg1.full <- glm(quality ~ volatile.acidityCent  + residual.sugarCent + citric.acid + chloridesCent
```

```
# forward Selection
step(modellg1.null,scope=list(upper=modellg1.full),direction="forward")
```

```
## Start:  AIC=2210.97
## quality ~ 1
##
##                          Df Deviance    AIC
## + alcoholCent             1   1865.0 1869.0
## + volatile.acidityCent    1   2033.4 2037.4
## + sulphatesCent           1   2122.8 2126.8
## + densityCent             1   2167.8 2171.8
## + citric.acid             1   2168.0 2172.0
```

```
## + chloridesCent            1   2188.6 2192.6
## + free.sulfur.dioxideCent  1   2202.9 2206.9
## <none>                         2209.0 2211.0
## + pHCent                   1   2209.0 2213.0
## + residual.sugarCent       1   2209.0 2213.0
##
## Step:  AIC=1869.03
## quality ~ alcoholCent
##
##                           Df Deviance    AIC
## + volatile.acidityCent     1   1752.5 1758.5
## + sulphatesCent            1   1804.7 1810.7
## + citric.acid              1   1841.4 1847.4
## + pHCent                   1   1846.8 1852.8
## + densityCent              1   1857.9 1863.9
## <none>                         1865.0 1869.0
## + free.sulfur.dioxideCent  1   1863.3 1869.3
## + residual.sugarCent       1   1864.2 1870.2
## + chloridesCent            1   1864.8 1870.8
##
## Step:  AIC=1758.5
## quality ~ alcoholCent + volatile.acidityCent
##
##                           Df Deviance    AIC
## + sulphatesCent            1   1724.5 1732.5
## + free.sulfur.dioxideCent  1   1749.7 1757.7
## + densityCent              1   1750.1 1758.1
## <none>                         1752.5 1758.5
## + pHCent                   1   1750.7 1758.7
## + citric.acid              1   1750.8 1758.8
## + residual.sugarCent       1   1751.8 1759.8
## + chloridesCent            1   1752.4 1760.4
##
## Step:  AIC=1732.49
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent
##
##                           Df Deviance    AIC
## + chloridesCent            1   1715.0 1725.0
## + citric.acid              1   1717.5 1727.5
## + free.sulfur.dioxideCent  1   1720.7 1730.7
## <none>                         1724.5 1732.5
## + residual.sugarCent       1   1723.8 1733.8
## + densityCent              1   1724.2 1734.2
## + pHCent                   1   1724.4 1734.4
##
## Step:  AIC=1725.01
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent +
##     chloridesCent
##
##                           Df Deviance    AIC
## + free.sulfur.dioxideCent  1   1710.7 1722.7
## + citric.acid              1   1710.8 1722.8
## <none>                         1715.0 1725.0
## + pHCent                   1   1714.0 1726.0
```

```
## + residual.sugarCent       1   1714.5 1726.5
## + densityCent              1   1714.8 1726.8
##
## Step:  AIC=1722.71
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent +
##     chloridesCent + free.sulfur.dioxideCent
##
##                       Df Deviance    AIC
## + citric.acid          1   1705.7 1719.7
## <none>                     1710.7 1722.7
## + pHCent               1   1710.2 1724.2
## + residual.sugarCent   1   1710.6 1724.6
## + densityCent          1   1710.6 1724.6
##
## Step:  AIC=1719.68
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent +
##     chloridesCent + free.sulfur.dioxideCent + citric.acid
##
##                       Df Deviance    AIC
## + pHCent               1   1701.7 1717.7
## + densityCent          1   1702.7 1718.7
## <none>                     1705.7 1719.7
## + residual.sugarCent   1   1705.7 1721.7
##
## Step:  AIC=1717.69
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent +
##     chloridesCent + free.sulfur.dioxideCent + citric.acid + pHCent
##
##                       Df Deviance    AIC
## + densityCent          1   1698.9 1716.9
## <none>                     1701.7 1717.7
## + residual.sugarCent   1   1701.7 1719.7
##
## Step:  AIC=1716.88
## quality ~ alcoholCent + volatile.acidityCent + sulphatesCent +
##     chloridesCent + free.sulfur.dioxideCent + citric.acid + pHCent +
##     densityCent
##
##                       Df Deviance    AIC
## <none>                     1698.9 1716.9
## + residual.sugarCent   1   1698.3 1718.3

##
## Call:  glm(formula = quality ~ alcoholCent + volatile.acidityCent +
##     sulphatesCent + chloridesCent + free.sulfur.dioxideCent +
##     citric.acid + pHCent + densityCent, family = binomial, data = data)
##
## Coefficients:
##             (Intercept)              alcoholCent     volatile.acidityCent
##                 0.71114                  1.05568                  -3.80216
##           sulphatesCent            chloridesCent  free.sulfur.dioxideCent
##                 2.60505                 -3.79466                  -0.01193
##             citric.acid                  pHCent              densityCent
##                -1.68434                 -0.95434                  75.68124
```

```
##
## Degrees of Freedom: 1598 Total (i.e. Null);  1590 Residual
## Null Deviance:       2209
## Residual Deviance: 1699  AIC: 1717
```

```
modellg1.selected <- glm(quality ~ alcoholCent  + volatile.acidityCent + sulphatesCent + chloridesCent +
summary(modellg1.selected)
```

```
##
## Call:
## glm(formula = quality ~ alcoholCent + volatile.acidityCent +
##     sulphatesCent + chloridesCent + free.sulfur.dioxideCent +
##     citric.acid + pHCent + densityCent, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3861  -0.8677   0.3184   0.8572   2.4028
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               0.711145   0.148087   4.802 1.57e-06 ***
## alcoholCent               1.055683   0.084814  12.447  < 2e-16 ***
## volatile.acidityCent     -3.802157   0.482445  -7.881 3.25e-15 ***
## sulphatesCent             2.605051   0.454235   5.735 9.75e-09 ***
## chloridesCent            -3.794657   1.516658  -2.502 0.012350 *
## free.sulfur.dioxideCent -0.011926   0.005849  -2.039 0.041462 *
## citric.acid              -1.684338   0.509051  -3.309 0.000937 ***
## pHCent                   -0.954338   0.488481  -1.954 0.050739 .
## densityCent              75.681238  45.291656   1.671 0.094727 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2209.0  on 1598  degrees of freedom
## Residual deviance: 1698.9  on 1590  degrees of freedom
## AIC: 1716.9
##
## Number of Fisher Scoring iterations: 4
```

```
kable(tidy(modellg1.selected))
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.7111446	0.1480871	4.802204	0.0000016
alcoholCent	1.0556830	0.0848137	12.447077	0.0000000
volatile.acidityCent	-3.8021566	0.4824455	-7.881008	0.0000000
sulphatesCent	2.6050510	0.4542355	5.735023	0.0000000
chloridesCent	-3.7946575	1.5166583	-2.501986	0.0123499
free.sulfur.dioxideCent	-0.0119260	0.0058493	-2.038875	0.0414625
citric.acid	-1.6843382	0.5090512	-3.308780	0.0009370
pHCent	-0.9543379	0.4884814	-1.953683	0.0507387
densityCent	75.6812385	45.2916564	1.670975	0.0947266

```
modellg1.final <- glm(quality ~ alcoholCent  + volatile.acidityCent + sulphatesCent + chloridesCent + f
summary(modellg1.final)
```

```
##
## Call:
## glm(formula = quality ~ alcoholCent + volatile.acidityCent +
##     sulphatesCent + chloridesCent + free.sulfur.dioxideCent +
##     citric.acid, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2135  -0.8735   0.3138   0.8834   2.3526
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.495163   0.120787   4.099 4.14e-05 ***
## alcoholCent                0.951016   0.070958  13.402  < 2e-16 ***
## volatile.acidityCent      -3.548641   0.451559  -7.859 3.88e-15 ***
## sulphatesCent              2.703043   0.446393   6.055 1.40e-09 ***
## chloridesCent             -3.770068   1.466806  -2.570   0.0102 *
## free.sulfur.dioxideCent -0.013254   0.005858  -2.262   0.0237 *
## citric.acid               -0.875529   0.391892  -2.234   0.0255 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2209.0  on 1598  degrees of freedom
## Residual deviance: 1705.7  on 1592  degrees of freedom
## AIC: 1719.7
##
## Number of Fisher Scoring iterations: 4
```

```
kable(tidy(modellg1.final))
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.4951634	0.1207867	4.099486	0.0000414
alcoholCent	0.9510158	0.0709582	13.402476	0.0000000
volatile.acidityCent	-3.5486411	0.4515594	-7.858636	0.0000000
sulphatesCent	2.7030432	0.4463930	6.055299	0.0000000
chloridesCent	-3.7700684	1.4668058	-2.570257	0.0101623
free.sulfur.dioxideCent	-0.0132536	0.0058580	-2.262476	0.0236680
citric.acid	-0.8755285	0.3918921	-2.234106	0.0254761

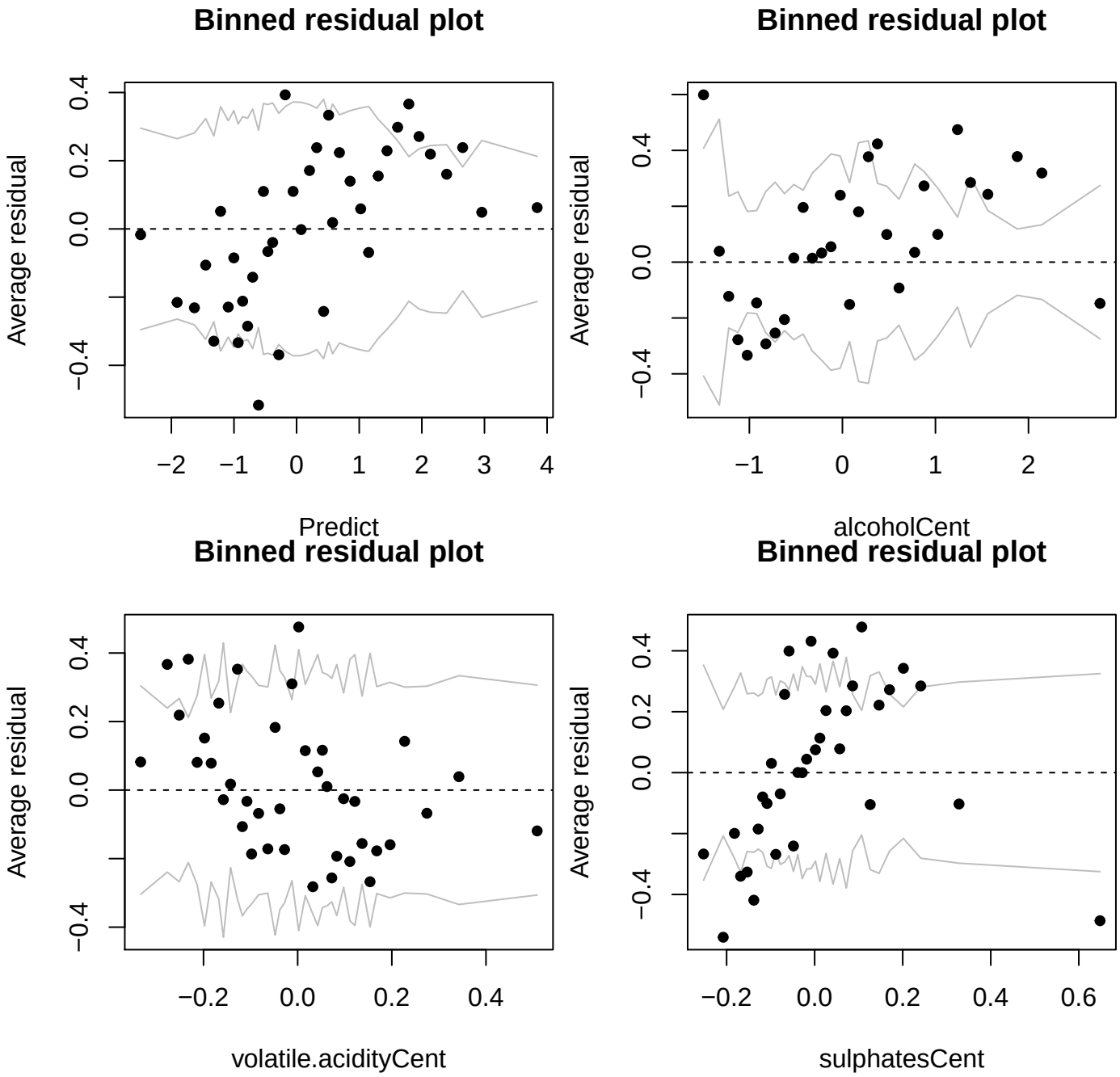
```
kable(tidy(vif(modellg1.final)))
```

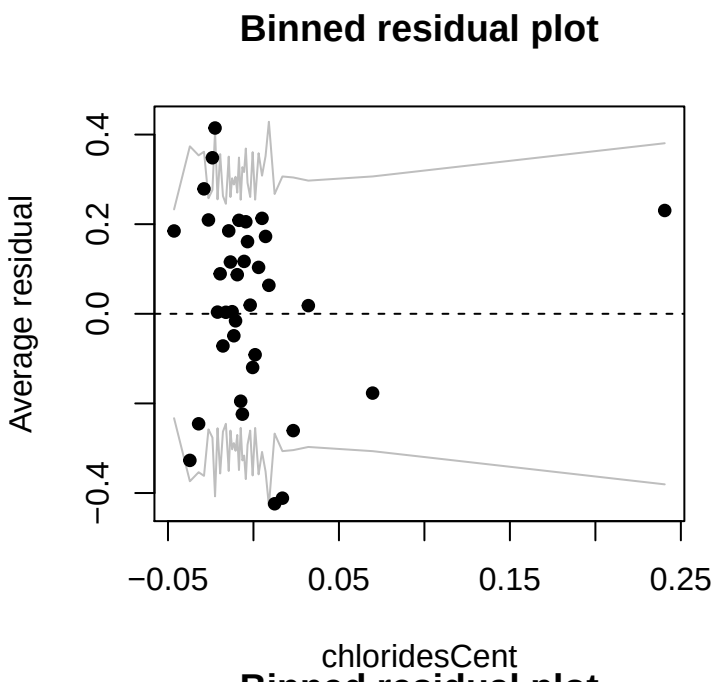
```
## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")
```

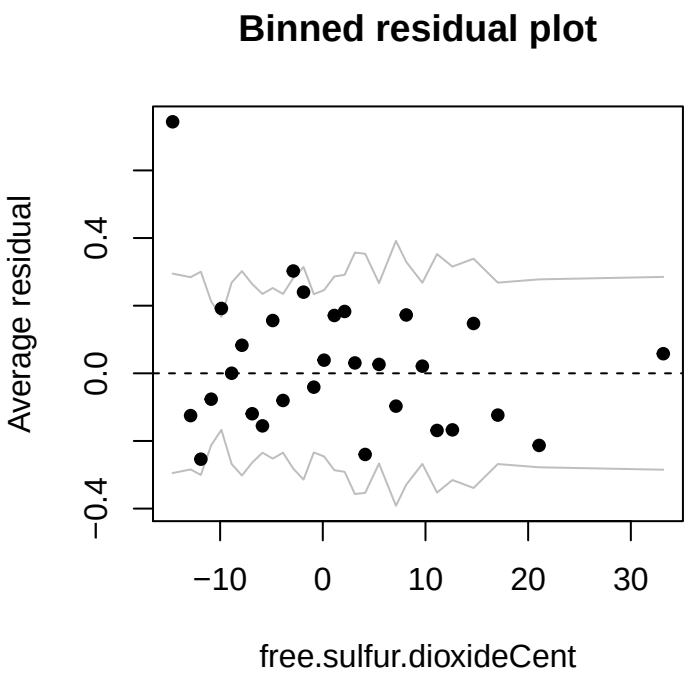
names	x
alcoholCent	1.050002
volatile.acidityCent	1.542159

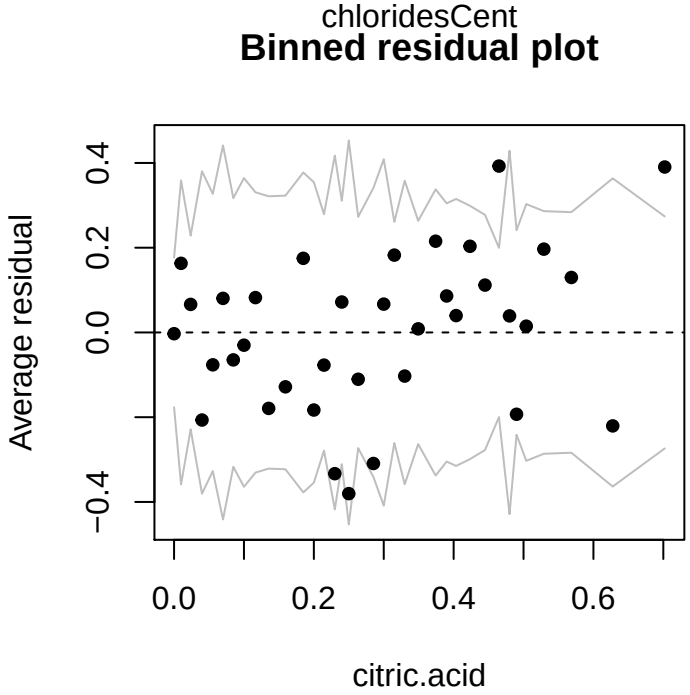
names	x
sulphatesCent	1.416972
chloridesCent	1.433665
free.sulfur.dioxideCent	1.018900
citric.acid	1.650265

```
data$pred <- predict.glm(modellg1.final)
data$res <- residuals.glm(modellg1.final)
```



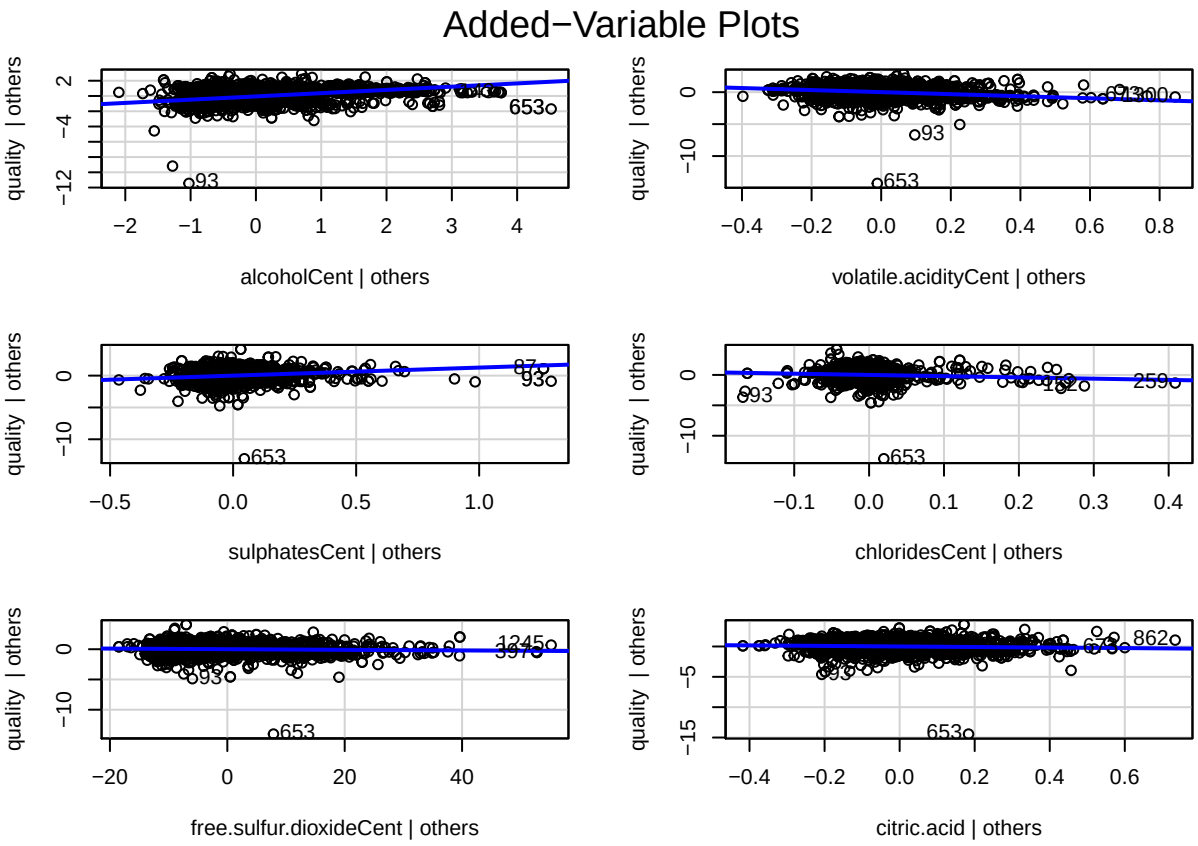




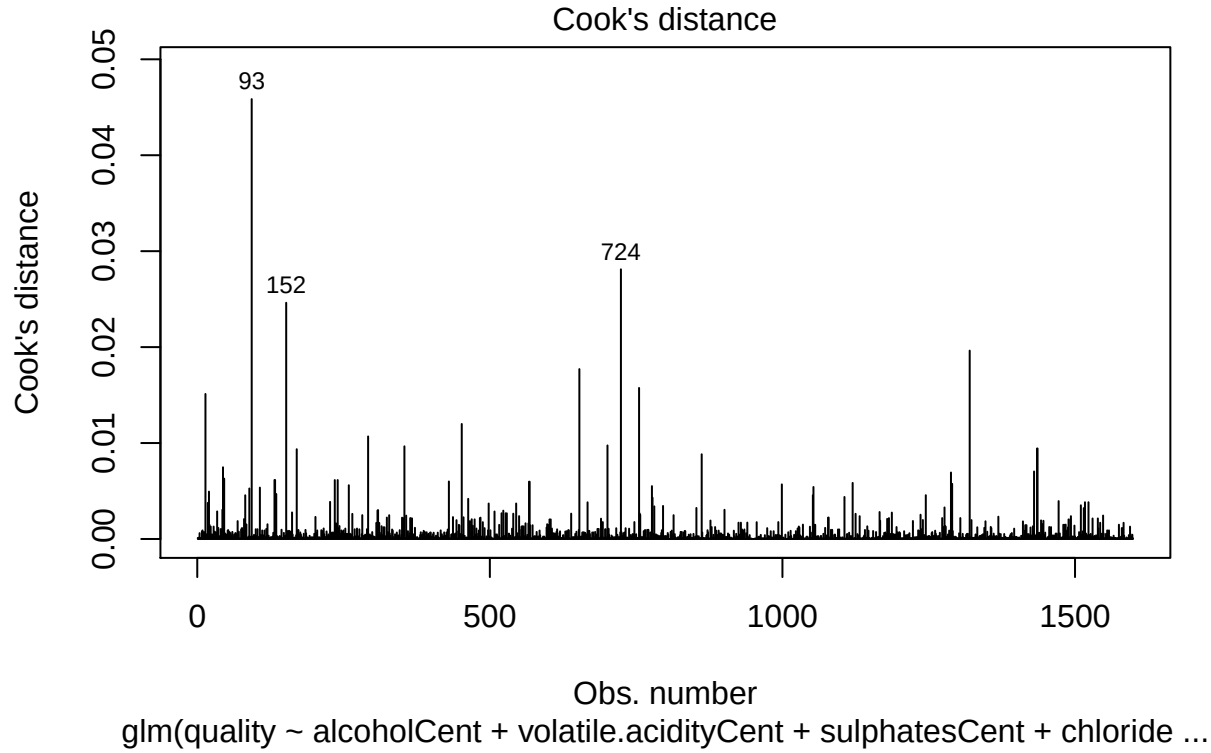


sulfates cent need transformation volatile.acidityCent also needs tranformation alcoholCent also needs transformation

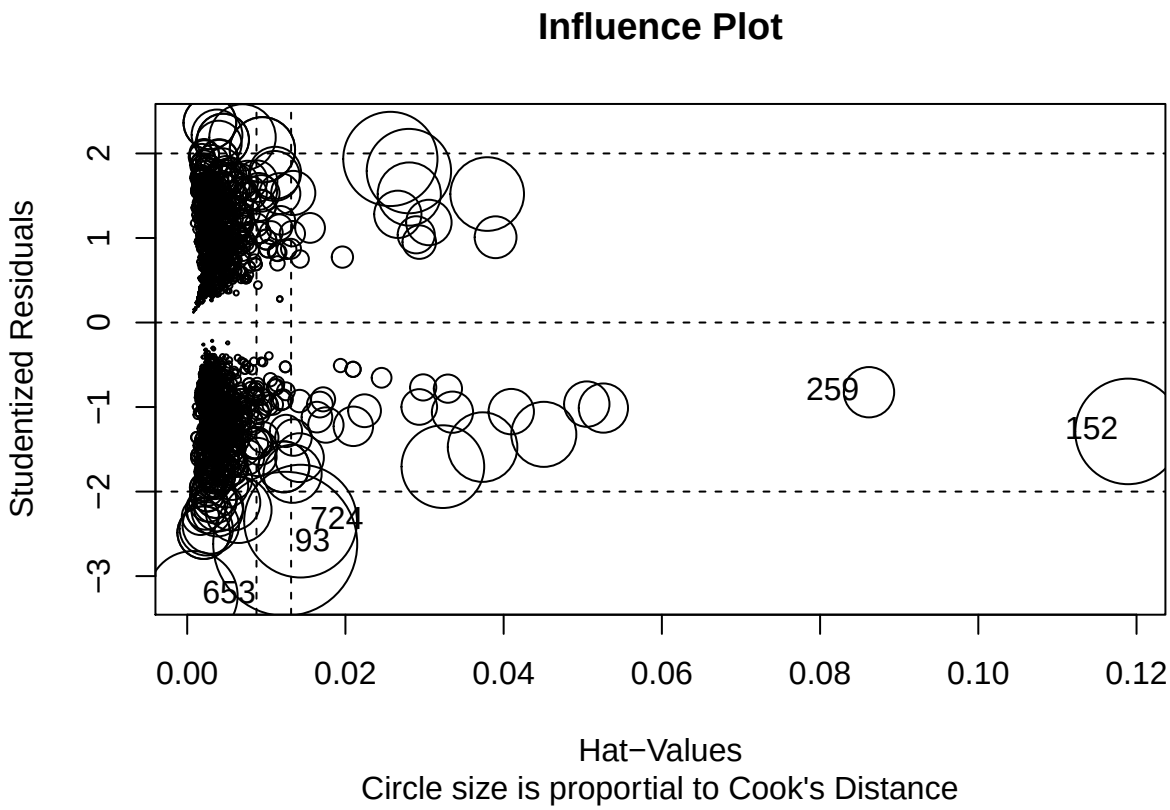
```
avPlots(modellg1.final)
```




```
cutoff <- cooks.distance(modellg1.final)
plot(modellg1.final, which=4, cook.levels=cutoff)
```



```
influencePlot(modellg1.final,    main="Influence Plot", sub="Circle size is proportial to Cook's Distanc
```

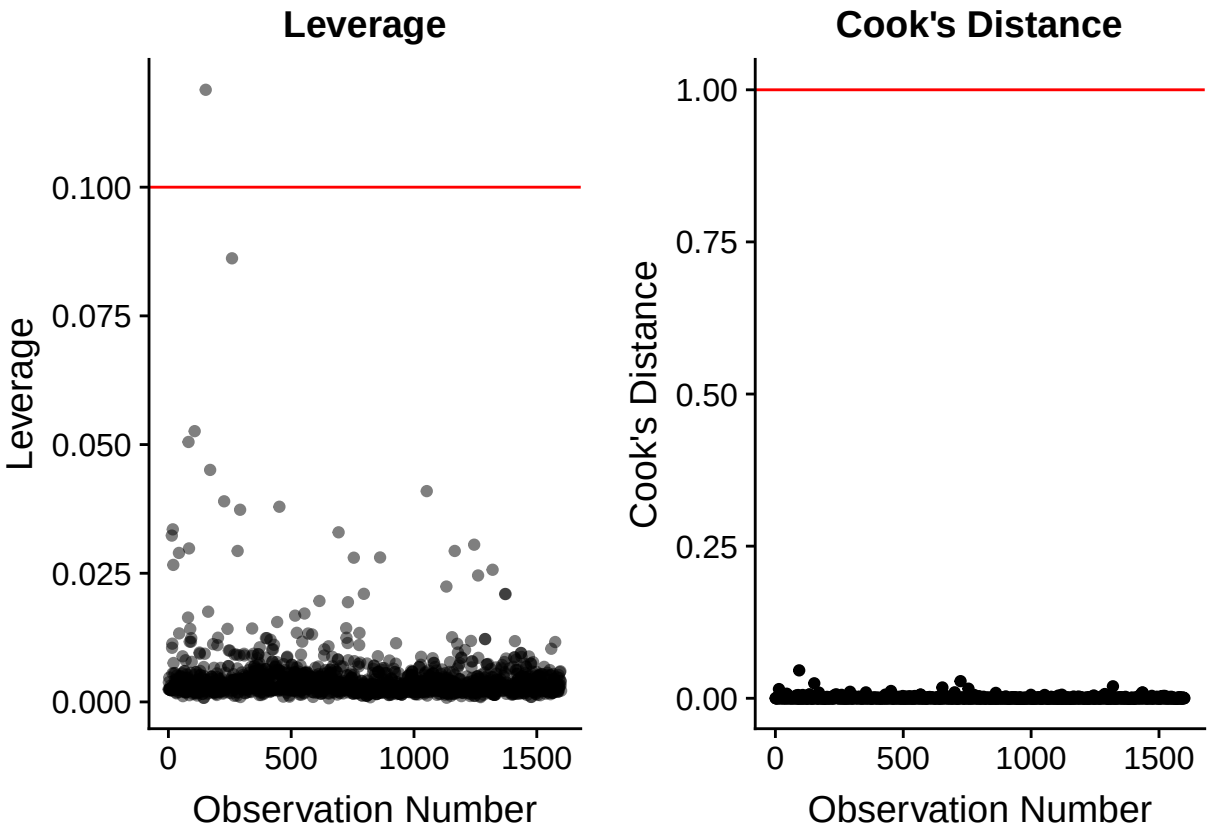



```
##        StudRes          Hat       CookD
## 93  -2.6180001 0.0123886268 0.045853660
## 152 -1.2882242 0.1189215890 0.024622603
## 259 -0.8246309 0.0861821330 0.005602629
## 653 -3.2327072 0.0007125379 0.017709691
## 724 -2.3495059 0.0143290897 0.028108832
```

```
data <- data %>%
  mutate(leverage = hatvalues(modellg1.final),
         cooks = cooks.distance(modellg1.final),
         stand.resid = rstandard(modellg1.final),
         obs.num = row_number())
p3 <- ggplot(data=data, aes(x=obs.num,y=leverage)) +
  geom_point(alpha=0.5) +
  geom_hline(yintercept=0.1,color="red")+
  labs(x="Observation Number",y="Leverage",title="Leverage")

p4 <- ggplot(data=data, aes(x=obs.num,y=cooks)) +
  geom_point() +
  geom_hline(yintercept=1,color="red")+
  labs(x="Observation Number",y="Cook's Distance",title="Cook's Distance")

plot_grid(p3,p4,ncol = 2)
```



no significant outlier or influential plot.