

FinalDraft

Kevin Tran

December 13, 2017

```
library(readr)

## Warning: package 'readr' was built under R version 3.3.3
wine <- read_csv("~/Stats101A/wine.csv")

## Parsed with column specification:
## cols(
##   Case = col_integer(),
##   Wine.Color = col_character(),
##   fixed.acidity = col_double(),
##   volatile.acidity = col_double(),
##   citric.acid = col_double(),
##   residual.sugar = col_double(),
##   chlorides = col_double(),
##   free.sulfur.dioxide = col_double(),
##   total.sulfur.dioxide = col_double(),
##   density = col_double(),
##   pH = col_double(),
##   sulphates = col_double(),
##   alcohol = col_double(),
##   Quality = col_double()
## )

WineTesting <- read_csv("~/Stats101A/WineTesting.csv")

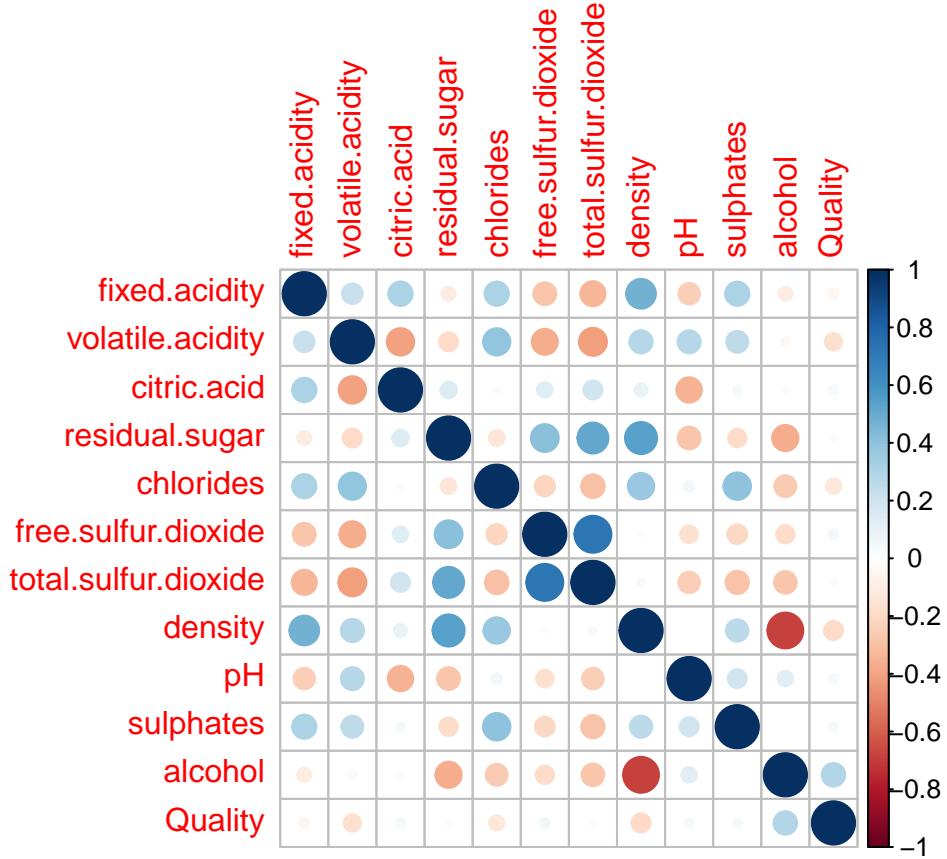
## Parsed with column specification:
## cols(
##   Case = col_integer(),
##   Wine.Color = col_character(),
##   fixed.acidity = col_double(),
##   volatile.acidity = col_double(),
##   citric.acid = col_double(),
##   residual.sugar = col_double(),
##   chlorides = col_double(),
##   free.sulfur.dioxide = col_double(),
##   total.sulfur.dioxide = col_double(),
##   density = col_double(),
##   pH = col_double(),
##   sulphates = col_double(),
##   alcohol = col_double()
## )

attach(wine)

#check for correlation or our numeric variables
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.3.3
```

```
## corrplot 0.84 loaded
corrgraph <- cor(wine[,3:14])
corrplot(corrgraph)
```



#we see that alcohol has the highest r with quality and should definitely include that in our model.
#Meanwhile alcohol and density are related pretty highly and may be redundant.
#Total sulfur and free sulfur also have high r and we should only include one if we want to.

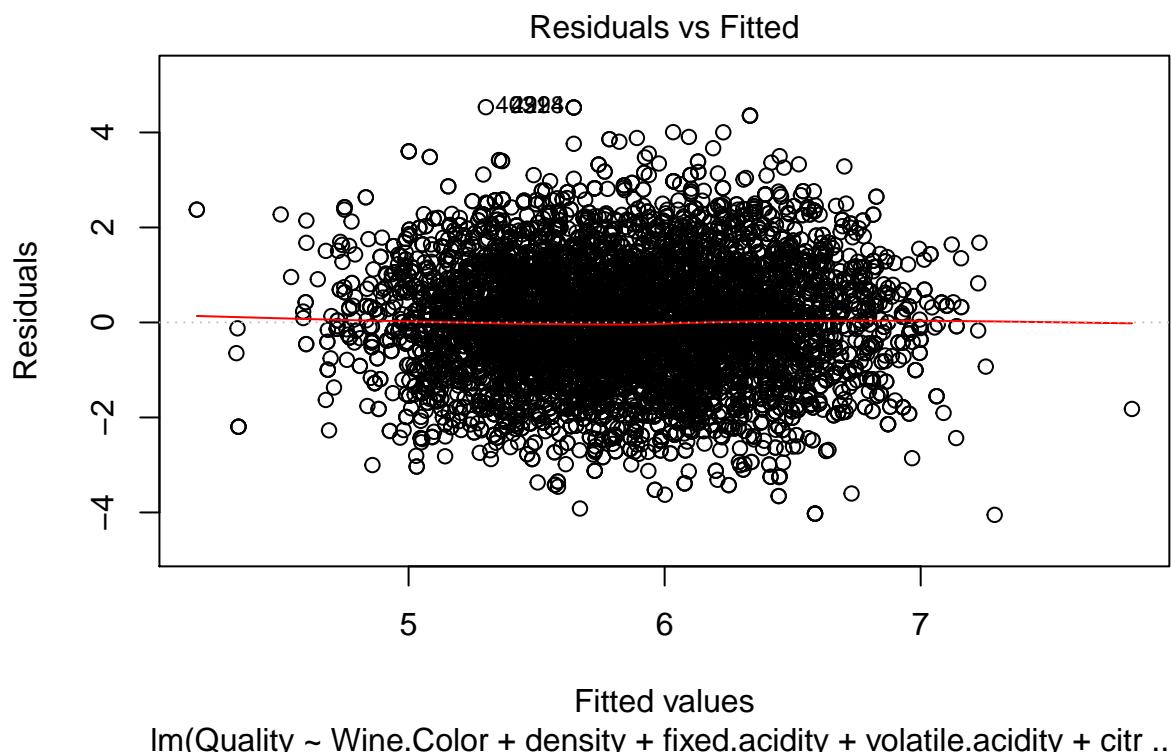
```
#Our first model
m1 <- lm(Quality ~ Wine.Color + density + fixed.acidity + volatile.acidity + citric.acid + residual.sugar
summary(m1)
```

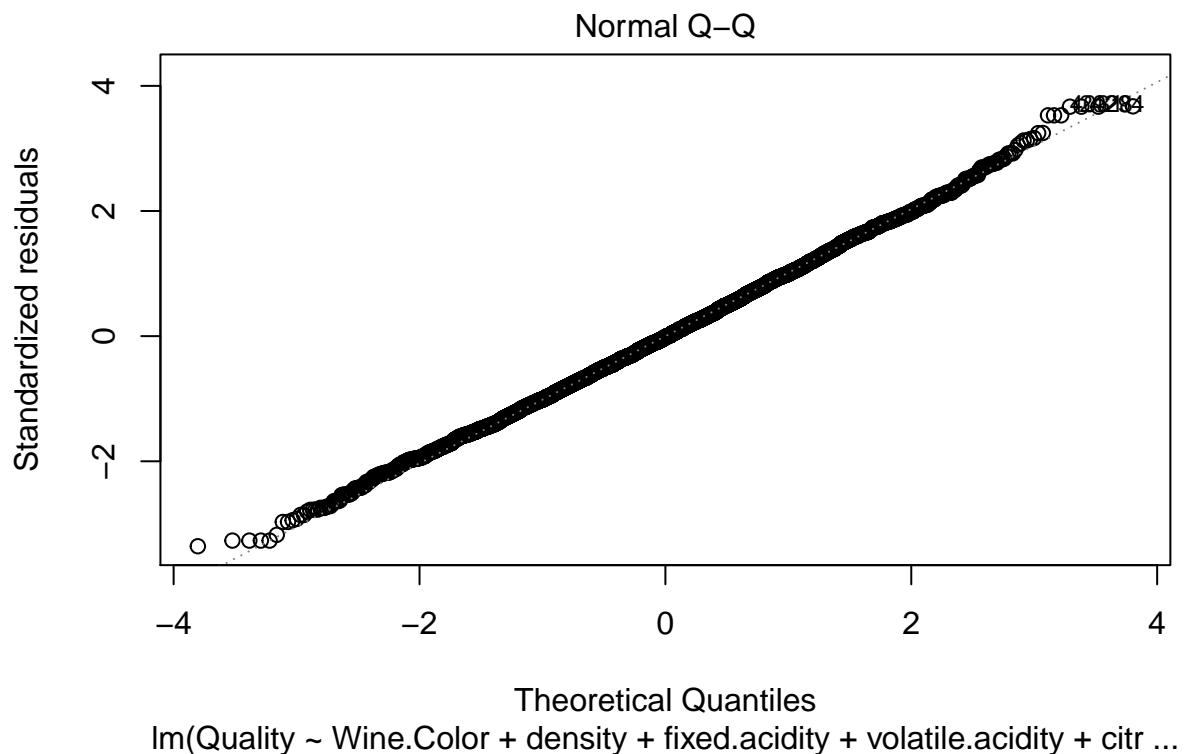
```
##
## Call:
## lm(formula = Quality ~ Wine.Color + density + fixed.acidity +
##     volatile.acidity + citric.acid + residual.sugar + chlorides +
##     free.sulfur.dioxide + total.sulfur.dioxide + pH + sulphates +
##     alcohol)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.0502 -0.8456 -0.0244  0.8431  4.5320 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.169e+02  2.313e+01   5.055 4.41e-07 ***
##
```

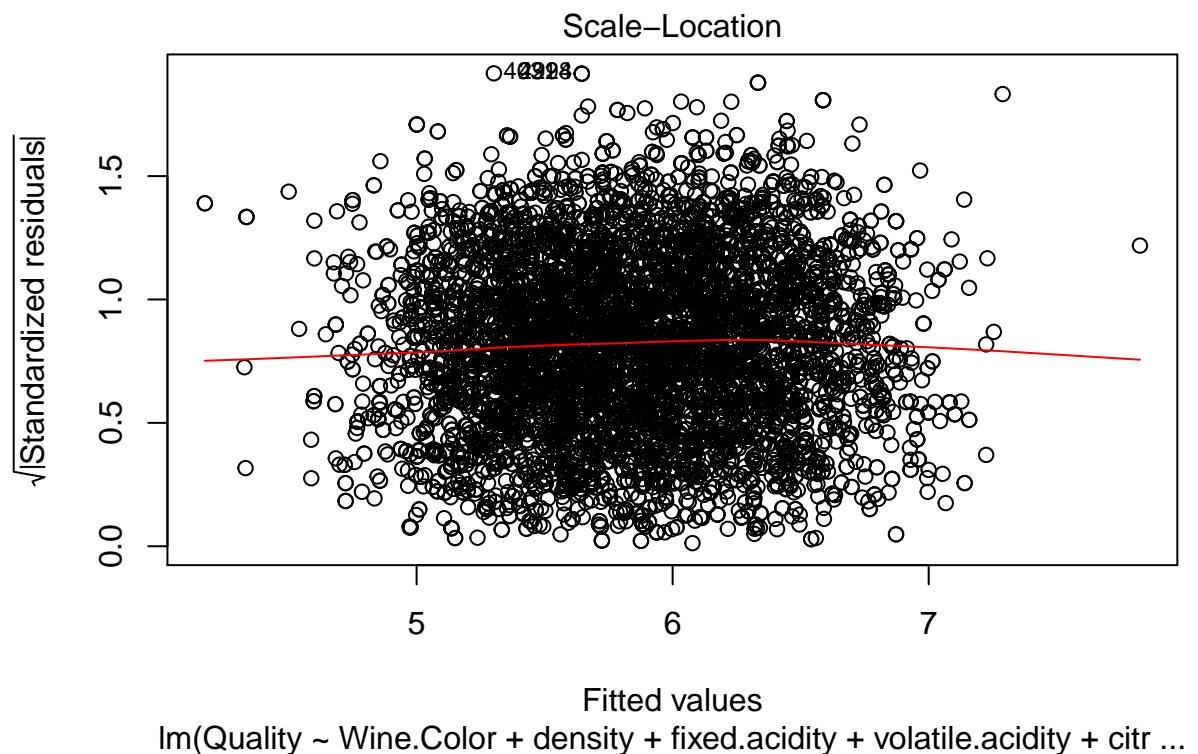
```

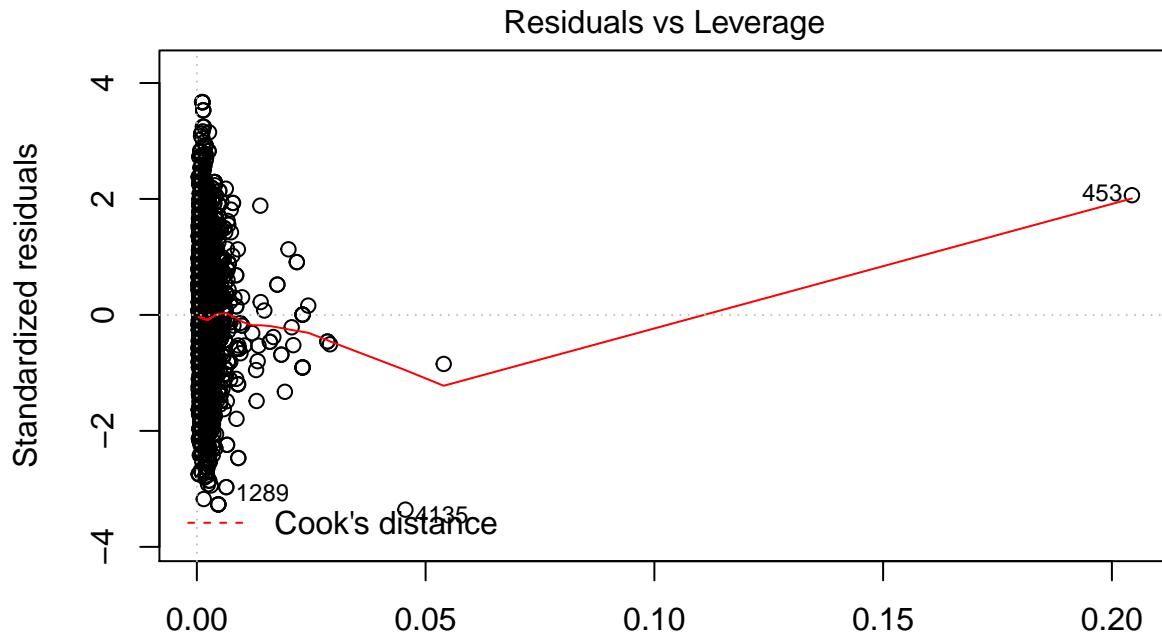
## Wine.ColorW      -2.075e-01 9.278e-02 -2.237 0.025319 *
## density         -1.183e+02 2.345e+01 -5.044 4.68e-07 ***
## fixed.acidity   1.238e-01 2.584e-02  4.791 1.69e-06 ***
## volatile.acidity -1.298e+00 1.326e-01 -9.788 < 2e-16 ***
## citric.acid    -1.039e-01 1.321e-01 -0.786 0.431605
## residual.sugar  7.259e-02 9.665e-03  7.510 6.63e-14 ***
## chlorides        -1.582e-01 5.729e-01 -0.276 0.782431
## free.sulfur.dioxide 7.052e-03 1.263e-03  5.585 2.43e-08 ***
## total.sulfur.dioxide -1.944e-03 5.313e-04 -3.660 0.000254 ***
## pH                1.009e+00 1.483e-01  6.801 1.12e-11 ***
## sulphates        9.015e-01 1.239e-01  7.278 3.77e-13 ***
## alcohol          2.131e-01 2.967e-02  7.180 7.67e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.235 on 6987 degrees of freedom
## Multiple R-squared:  0.134, Adjusted R-squared:  0.1325
## F-statistic: 90.09 on 12 and 6987 DF, p-value: < 2.2e-16
plot(m1)

```









Leverage
lm(Quality ~ Wine.Color + density + fixed.acidity + volatile.acidity + citr ...

```
#remove outliers
wine <- wine[-c(270,4135,5523),]
#diagnostic shows transformation is not needed
detach(wine)
attach(wine)

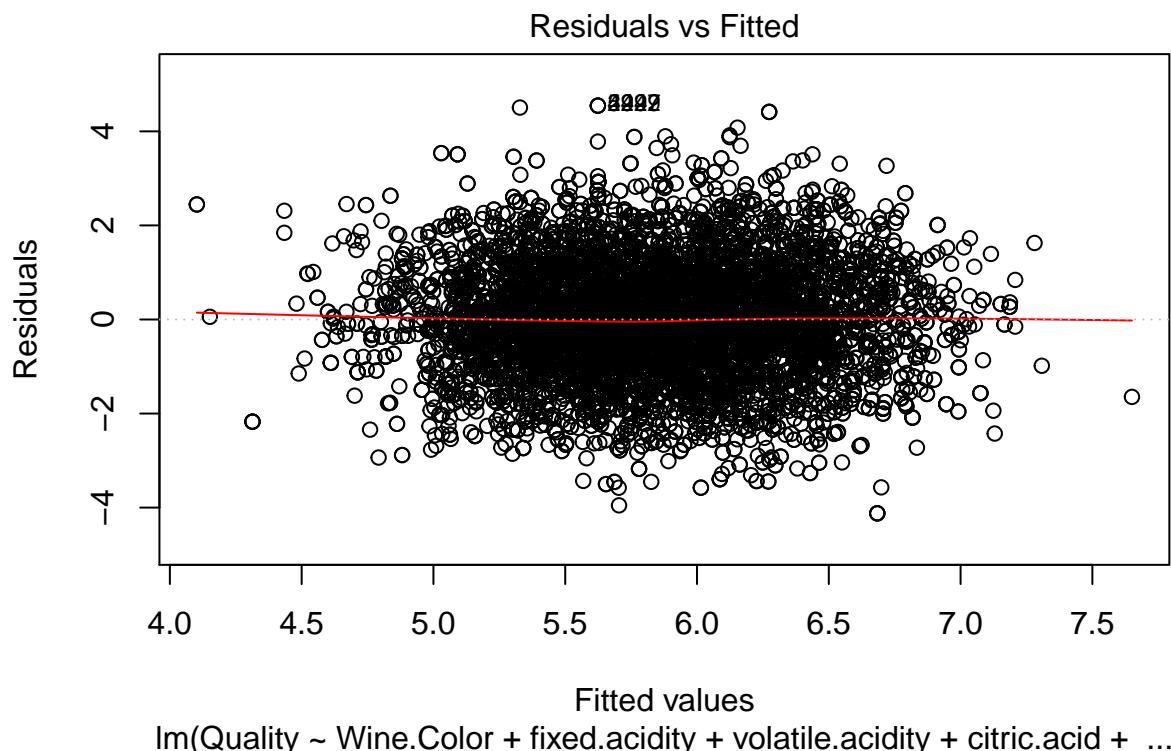
#see changes
m1a <- lm(Quality ~ Wine.Color + fixed.acidity + volatile.acidity + citric.acid + residual.sugar + chlor
summary(m1a)

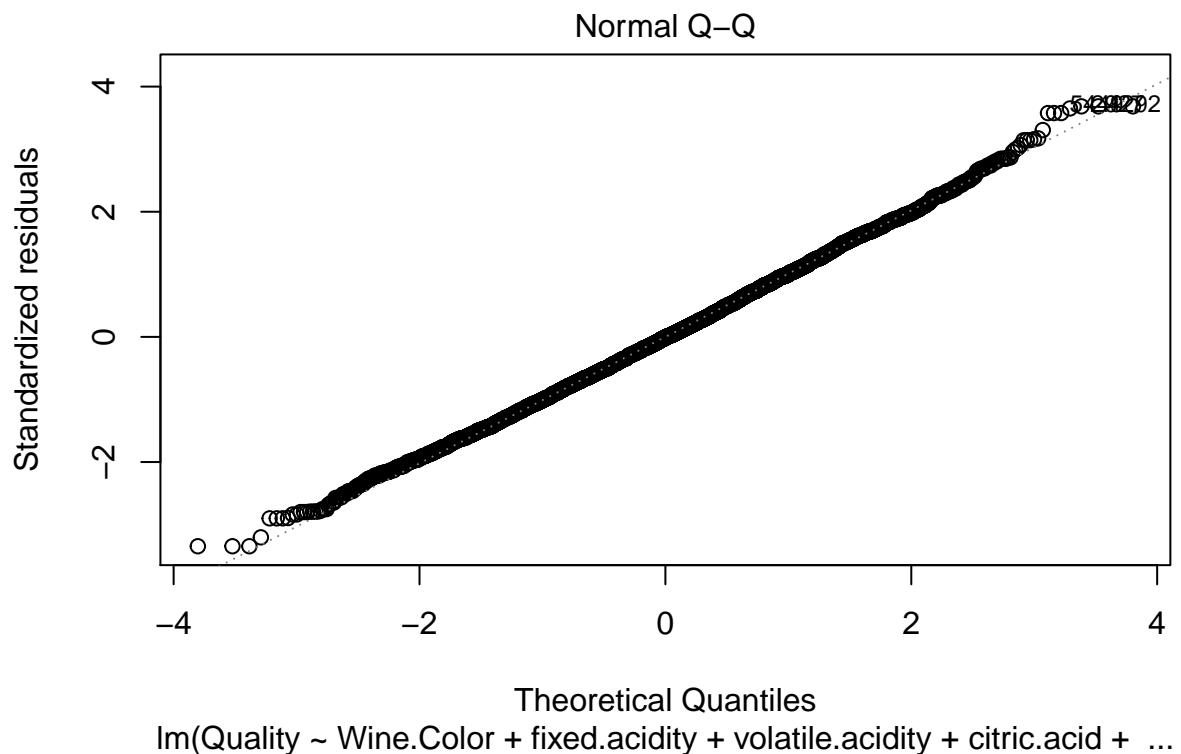
##
## Call:
## lm(formula = Quality ~ Wine.Color + fixed.acidity + volatile.acidity +
##     citric.acid + residual.sugar + chlorides + free.sulfur.dioxide +
##     total.sulfur.dioxide + pH + sulphates + alcohol)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -4.1215 -0.8408 -0.0126  0.8409  4.5464
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.1943534  0.5055229  0.384   0.701
## Wine.ColorW           0.0284839  0.0787290  0.362   0.718
## fixed.acidity          0.0256902  0.0168472  1.525   0.127
## volatile.acidity      -1.3707389  0.1316066 -10.415 < 2e-16 ***
## citric.acid           0.0000000  0.0000000  0.000   1.000
## residual.sugar         0.0000000  0.0000000  0.000   1.000
## chlorides              0.0000000  0.0000000  0.000   1.000
## free.sulfur.dioxide   0.0000000  0.0000000  0.000   1.000
## total.sulfur.dioxide  0.0000000  0.0000000  0.000   1.000
## pH                      0.0000000  0.0000000  0.000   1.000
## sulphates              0.0000000  0.0000000  0.000   1.000
## alcohol                 0.0000000  0.0000000  0.000   1.000
```

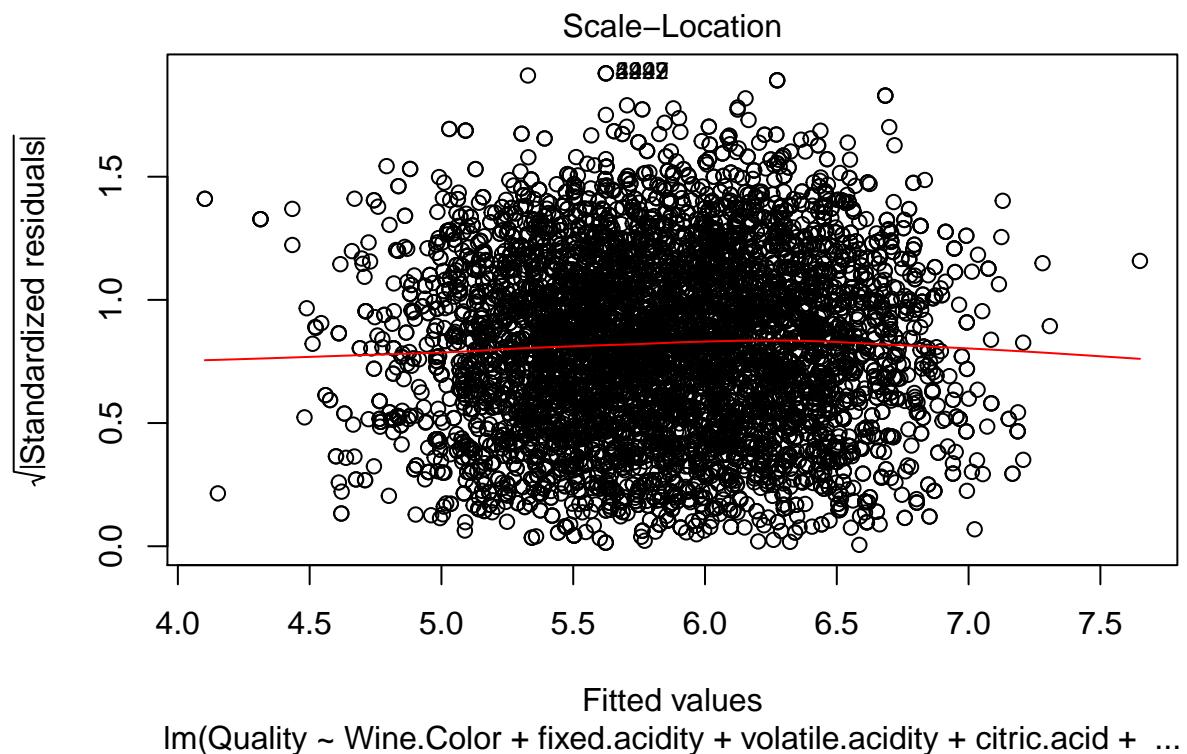
```

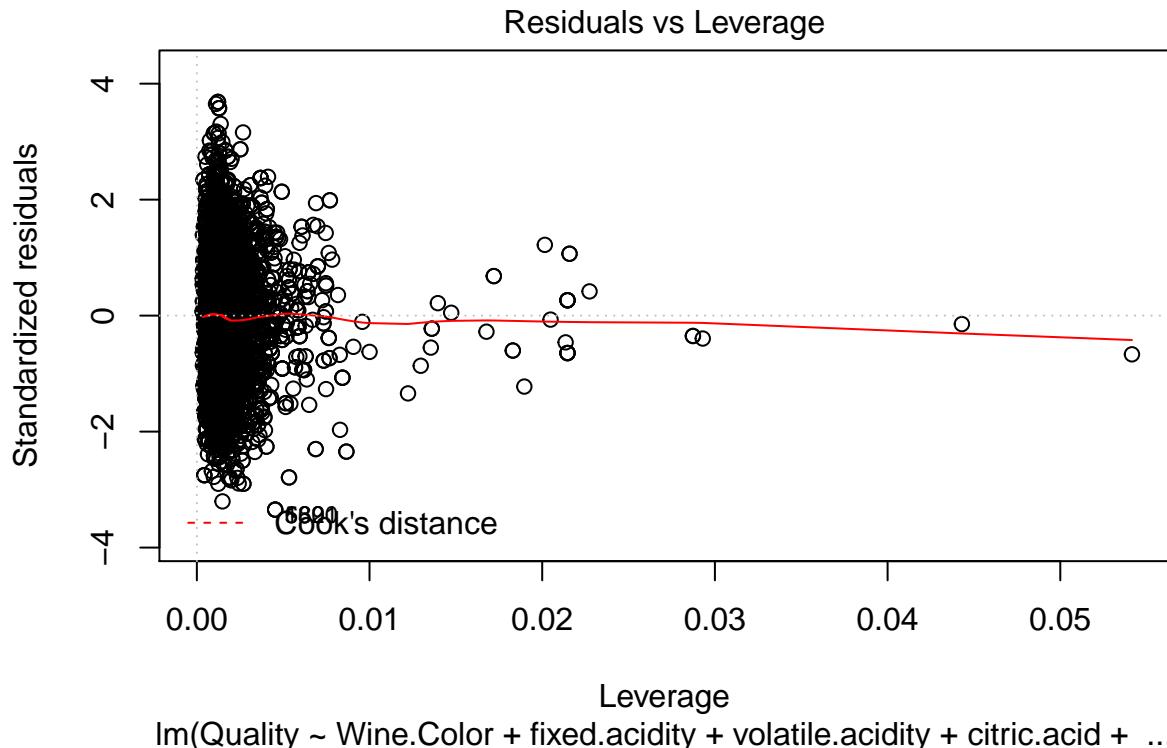
## citric.acid      -0.1441730  0.1319216 -1.093    0.274
## residual.sugar   0.0273795  0.0038838  7.050 1.96e-12 ***
## chlorides        -0.4740883  0.5742848 -0.826    0.409
## free.sulfur.dioxide  0.0086364  0.0012736  6.781 1.29e-11 ***
## total.sulfur.dioxide -0.0023540  0.0005254 -4.481 7.56e-06 ***
## pH                 0.5757279  0.1179280  4.882 1.07e-06 ***
## sulphates         0.7319246  0.1196235  6.119 9.95e-10 ***
## alcohol            0.3437154  0.0150552 22.830 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.235 on 6985 degrees of freedom
## Multiple R-squared:  0.1326, Adjusted R-squared:  0.1312
## F-statistic: 97.07 on 11 and 6985 DF,  p-value: < 2.2e-16
plot(m1a)

```









```
#r2 is .1326 r2a

#horrendous r2
#let us start variable selection using forward selection
X <- lm(Quality~1)
forwardAIC <- step(X,scope=list(lower=~1, upper=~Wine.Color + fixed.acidity + volatile.acidity + citric.acid + residual.sugar))

## Start: AIC=3936.3
## Quality ~ 1
##
##                               Df  Sum of Sq    RSS     AIC
## + alcohol                  1   1075.88 11202 3296.6
## + volatile.acidity         1    317.33 11960 3755.1
## + chlorides                1    204.74 12073 3820.6
## + Wine.Color                1     76.72 12201 3894.4
## + fixed.acidity             1     43.57 12234 3913.4
## + free.sulfur.dioxide      1     38.67 12239 3916.2
## + pH                        1     29.82 12248 3921.3
## + citric.acid               1     26.91 12251 3922.9
## + sulphates                 1     19.22 12258 3927.3
## + total.sulfur.dioxide     1      5.83 12272 3935.0
## + residual.sugar            1      5.13 12272 3935.4
## <none>                      12278 3936.3
##
## Step: AIC=3296.6
## Quality ~ alcohol
```

```

##                                     Df Sum of Sq   RSS   AIC
## + volatile.acidity      1   278.369 10923 3122.5
## + free.sulfur.dioxide   1   156.692 11045 3200.0
## + residual.sugar        1   110.457 11091 3229.3
## + Wine.Color             1    61.643 11140 3260.0
## + total.sulfur.dioxide  1    46.576 11155 3269.5
## + chlorides              1    36.786 11165 3275.6
## + citric.acid            1    34.439 11167 3277.1
## + sulphates              1    16.315 11185 3288.4
## + fixed.acidity           1    10.055 11192 3292.3
## <none>                      11202 3296.6
## + pH                         1    1.596 11200 3297.6
##
## Step:  AIC=3122.53
## Quality ~ alcohol + volatile.acidity
##
##                                     Df Sum of Sq   RSS   AIC
## + sulphates                  1    74.376 10849 3076.7
## + residual.sugar              1    49.156 10874 3093.0
## + free.sulfur.dioxide         1    44.930 10878 3095.7
## + pH                           1    40.843 10882 3098.3
## + Wine.Color                  1    19.257 10904 3112.2
## <none>                        10923 3122.5
## + citric.acid                 1     0.842 10922 3124.0
## + chlorides                   1     0.557 10923 3124.2
## + total.sulfur.dioxide        1     0.475 10923 3124.2
## + fixed.acidity                1     0.337 10923 3124.3
##
## Step:  AIC=3076.72
## Quality ~ alcohol + volatile.acidity + sulphates
##
##                                     Df Sum of Sq   RSS   AIC
## + residual.sugar              1    69.671 10779 3033.6
## + free.sulfur.dioxide          1    62.589 10786 3038.2
## + pH                           1    28.175 10821 3060.5
## + chlorides                   1     6.194 10843 3074.7
## + citric.acid                 1     5.520 10843 3075.2
## + fixed.acidity                1     3.478 10845 3076.5
## <none>                        10849 3076.7
## + total.sulfur.dioxide         1     1.392 10847 3077.8
## + Wine.Color                  1     0.310 10848 3078.5
##
## Step:  AIC=3033.64
## Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar
##
##                                     Df Sum of Sq   RSS   AIC
## + pH                           1    48.507 10731 3004.1
## + free.sulfur.dioxide          1    30.211 10749 3016.0
## + Wine.Color                  1    10.104 10769 3029.1
## + citric.acid                 1    10.046 10769 3029.1
## + total.sulfur.dioxide         1     5.580 10774 3032.0
## <none>                        10779 3033.6
## + chlorides                   1     1.697 10778 3034.5

```

```

## + fixed.acidity      1     1.473 10778 3034.7
##
## Step: AIC=3004.08
## Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar +
##          pH
##
##                               Df Sum of Sq   RSS   AIC
## + free.sulfur.dioxide  1    28.0187 10703 2987.8
## + Wine.Color           1     5.3881 10725 3002.6
## + total.sulfur.dioxide 1     5.0836 10726 3002.8
## <none>                      10731 3004.1
## + fixed.acidity        1     2.8588 10728 3004.2
## + citric.acid          1     1.6905 10729 3005.0
## + chlorides             1     0.3497 10730 3005.9
##
## Step: AIC=2987.79
## Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar +
##          pH + free.sulfur.dioxide
##
##                               Df Sum of Sq   RSS   AIC
## + total.sulfur.dioxide 1    47.439 10655 2958.7
## + Wine.Color            1    15.778 10687 2979.5
## + fixed.acidity         1     8.372 10694 2984.3
## <none>                      10703 2987.8
## + citric.acid          1     1.323 10701 2988.9
## + chlorides              1     0.042 10703 2989.8
##
## Step: AIC=2958.71
## Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar +
##          pH + free.sulfur.dioxide + total.sulfur.dioxide
##
##                               Df Sum of Sq   RSS   AIC
## <none>                      10655 2958.7
## + fixed.acidity  1    1.88530 10653 2959.5
## + chlorides       1    1.68518 10654 2959.6
## + citric.acid    1    0.70318 10654 2960.2
## + Wine.Color      1    0.00006 10655 2960.7
#result: Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar + pH + free.sulfur.dioxide

m2 <- lm(Quality ~ alcohol + volatile.acidity + sulphates + residual.sugar +
          pH + free.sulfur.dioxide + total.sulfur.dioxide)
summary(m2) #.1321
```

```

## 
## Call:
## lm(formula = Quality ~ alcohol + volatile.acidity + sulphates +
##      residual.sugar + pH + free.sulfur.dioxide + total.sulfur.dioxide)
##
## Residuals:
##      Min      1Q Median      3Q      Max
## -4.1642 -0.8403 -0.0160  0.8408  4.5703
##
## Coefficients:
## (Intercept)  alcohol  volatile.acidity sulphates
##             10655.00000  1.8853000  1.6851800  0.7031800
## residual.sugar  pH  free.sulfur.dioxide total.sulfur.dioxide
##             0.0000600  1.0000000  1.0000000  1.0000000
## Estimate Std. Error t value Pr(>|t|)
```

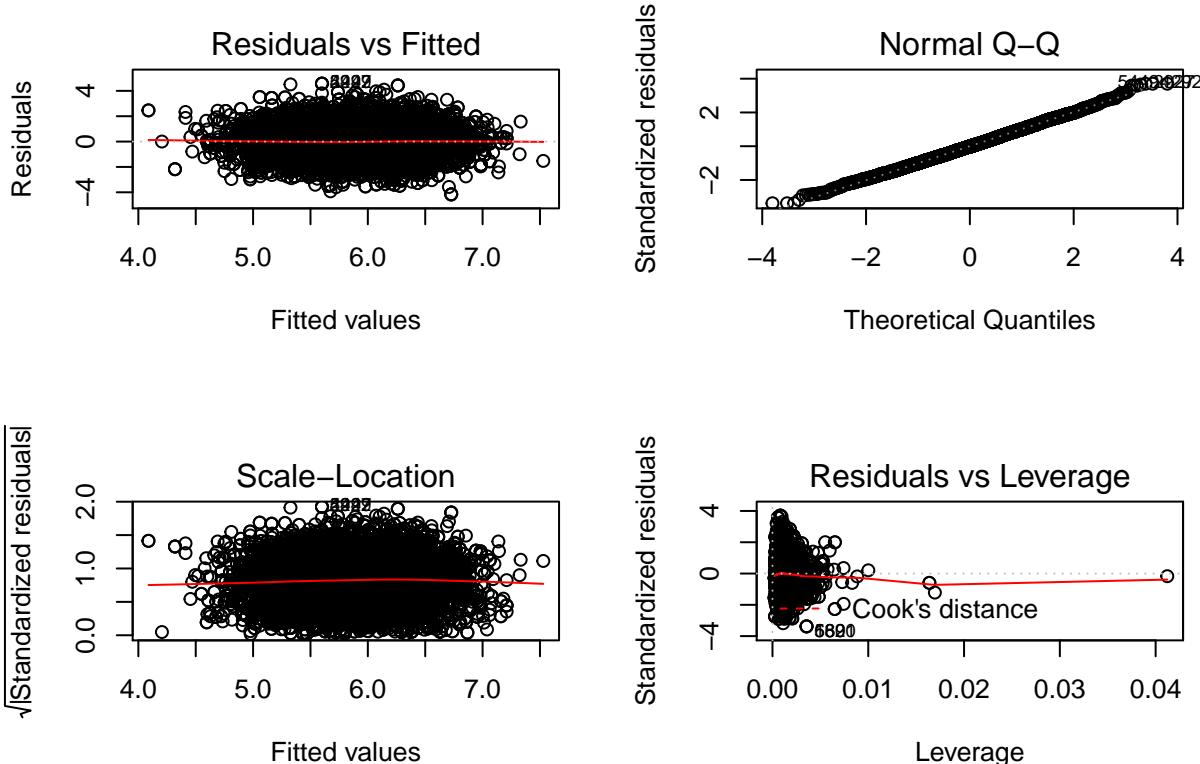
```

## (Intercept)      0.474241   0.356748   1.329    0.184
## alcohol         0.345089   0.013672  25.241   < 2e-16 ***
## volatile.acidity -1.353642  0.101935 -13.279   < 2e-16 ***
## sulphates        0.711488   0.105424   6.749  1.61e-11 ***
## residual.sugar   0.027349   0.003858   7.089  1.49e-12 ***
## pH               0.529782   0.100481   5.272  1.39e-07 ***
## free.sulfur.dioxide 0.008511   0.001253   6.794  1.18e-11 ***
## total.sulfur.dioxide -0.002376  0.000426 -5.578  2.52e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.235 on 6989 degrees of freedom
## Multiple R-squared:  0.1321, Adjusted R-squared:  0.1313
## F-statistic:  152 on 7 and 6989 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(m2)

```

#transformations of our best model
`library(car)`

`## Warning: package 'car' was built under R version 3.3.3`



```

summary(powerTransform(cbind(alcohol, volatile.acidity , sulphates , residual.sugar , pH , free.sulfur.

## bcPower Transformations to Multinormality
##          Est.Power Std.Err. Wald Lower Bound Wald Upper Bound

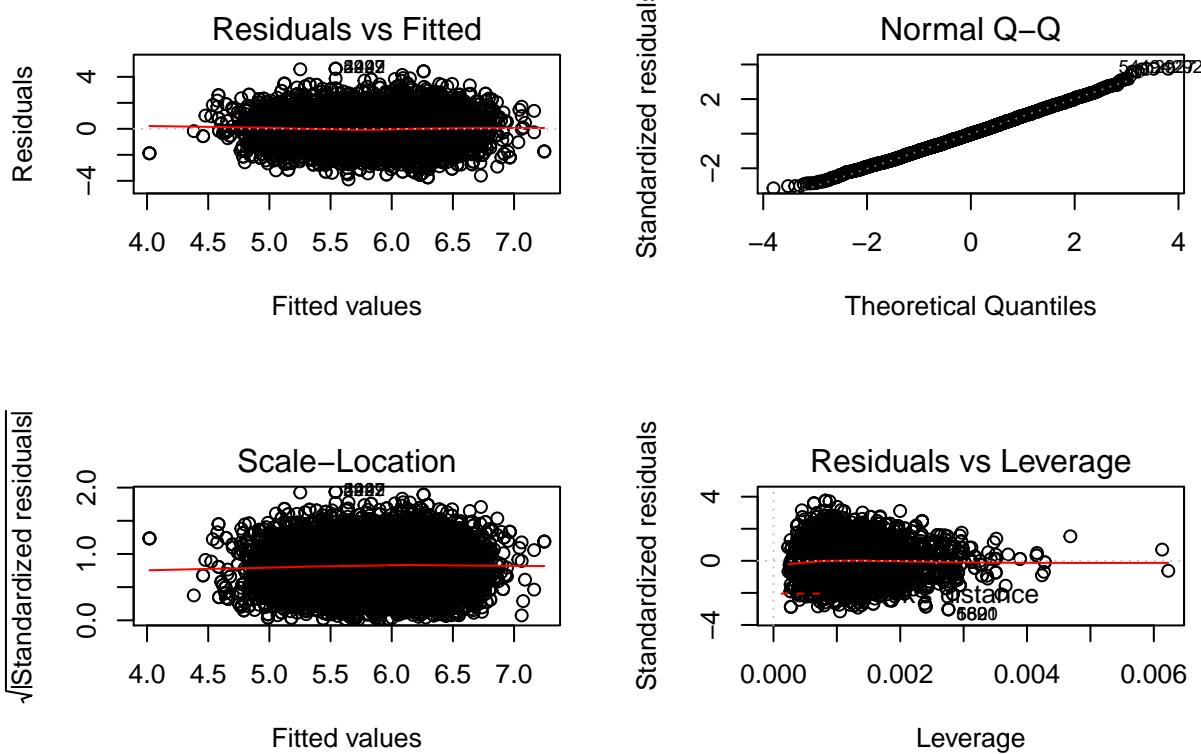
```

```

## alcohol           -2.1061  0.1103   -2.3222  -1.8900
## volatile.acidity -0.0908  0.0226   -0.1351  -0.0466
## sulphates        -0.4409  0.0357   -0.5109  -0.3708
## residual.sugar   -0.0467  0.0154   -0.0768  -0.0166
## pH                -1.3804  0.1953   -1.7631  -0.9977
## free.sulfur.dioxide 0.3686  0.0135    0.3420  0.3951
## total.sulfur.dioxide 0.7436  0.0150    0.7141  0.7730
## Quality          0.8029  0.0409    0.7228  0.8830
##
## Likelihood ratio tests about transformation parameters
##                                         LRT df
## LR test, lambda = (0 0 0 0 0 0 0)             4238.959803  8
## LR test, lambda = (1 1 1 1 1 1 1)             12162.291204  8
## LR test, lambda = (-2 -0.09 -0.5 -0.05 -1 0.37 0.74 0.8) 7.455386  8
##                                         pval
## LR test, lambda = (0 0 0 0 0 0 0)            0.0000000
## LR test, lambda = (1 1 1 1 1 1 1)            0.0000000
## LR test, lambda = (-2 -0.09 -0.5 -0.05 -1 0.37 0.74 0.8) 0.4883884
#apply suggest transformations
m3 <- lm(Quality ~ I(alcohol^-2) + log(volatile.acidity) + I(sulphates^-0.5) + log(residual.sugar) +
  I(pH^-1) + I(free.sulfur.dioxide^(1/3)) + I(total.sulfur.dioxide^(3/4)))
summary(m3) #.1326

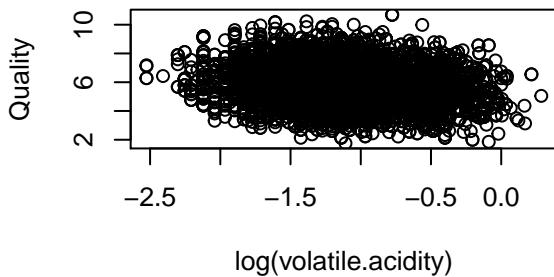
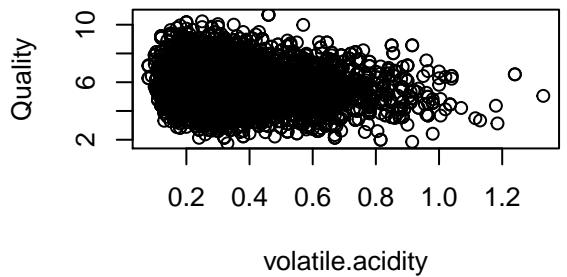
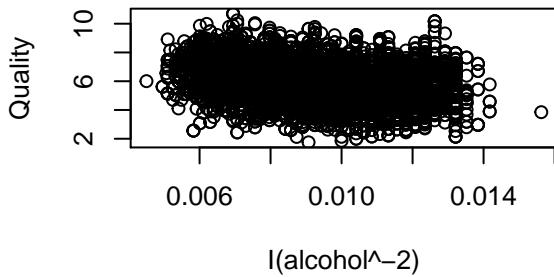
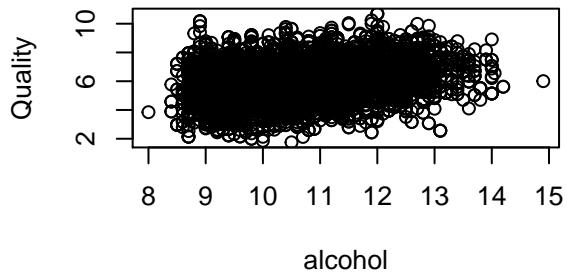
##
## Call:
## lm(formula = Quality ~ I(alcohol^-2) + log(volatile.acidity) +
##     I(sulphates^-0.5) + log(residual.sugar) + I(pH^-1) + I(free.sulfur.dioxide^(1/3)) +
##     I(total.sulfur.dioxide^(3/4)))
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -3.8949 -0.8412 -0.0140  0.8408  4.6291 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               8.409e+00  3.307e-01  25.423 < 2e-16 ***
## I(alcohol^-2)            -2.011e+02  8.097e+00 -24.837 < 2e-16 ***
## log(volatile.acidity)   -4.977e-01  3.784e-02 -13.153 < 2e-16 ***
## I(sulphates^-0.5)        -5.565e-01  8.890e-02 -6.260 4.07e-10 ***
## log(residual.sugar)      1.738e-01  2.075e-02  8.377 < 2e-16 ***
## I(pH^-1)                  -3.851e+00  1.048e+00 -3.675  0.00024 ***
## I(free.sulfur.dioxide^(1/3)) 3.239e-01  3.641e-02  8.896 < 2e-16 ***
## I(total.sulfur.dioxide^(3/4)) -1.366e-02  1.869e-03 -7.310 2.97e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.234 on 6989 degrees of freedom
## Multiple R-squared:  0.1326, Adjusted R-squared:  0.1317 
## F-statistic: 152.6 on 7 and 6989 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(m3)

```

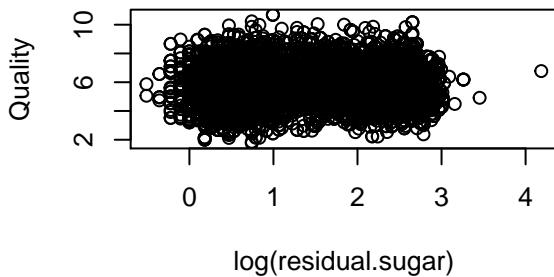
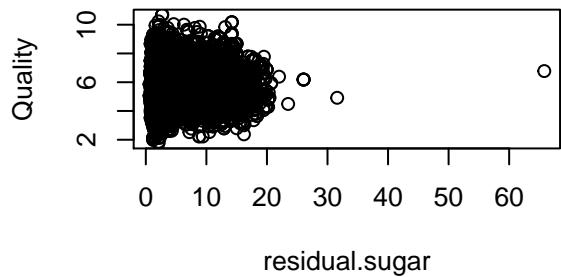
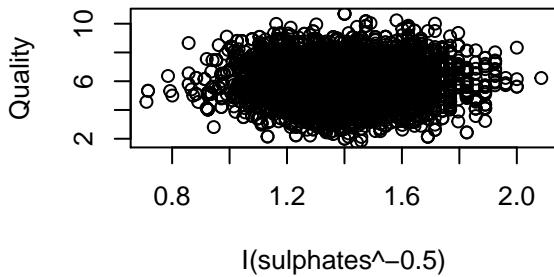
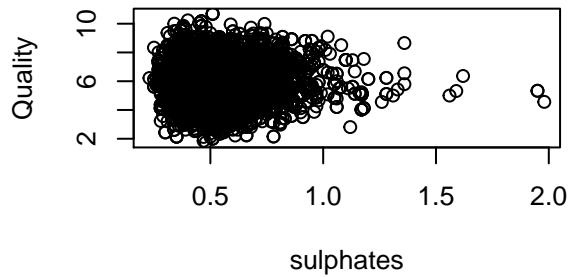


```
vif(m3)
```

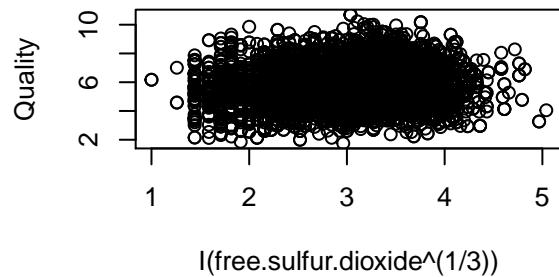
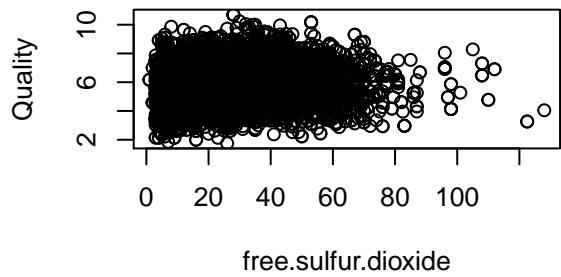
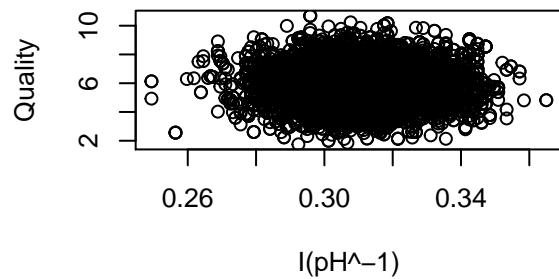
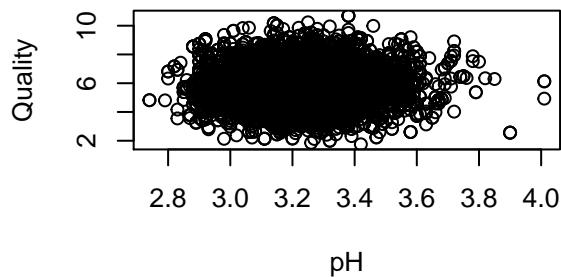
```
##          I(alcohol^-2)      log(volatile.acidity)
##                1.209807            1.313082
##          I(sulphates^-0.5)    log(residual.sugar)
##                1.157490            1.467661
##          I(pH^-1)  I(free.sulfur.dioxide^(1/3))
##                1.167374            2.513894
## I(total.sulfur.dioxide^(3/4))
##                2.955645
#before and afters of transformations
par(mfrow=c(2,2))
plot(Quality ~ alcohol)
plot(Quality ~ I(alcohol^-2))
plot(Quality ~ volatile.acidity)
plot(Quality ~ log(volatile.acidity))
```



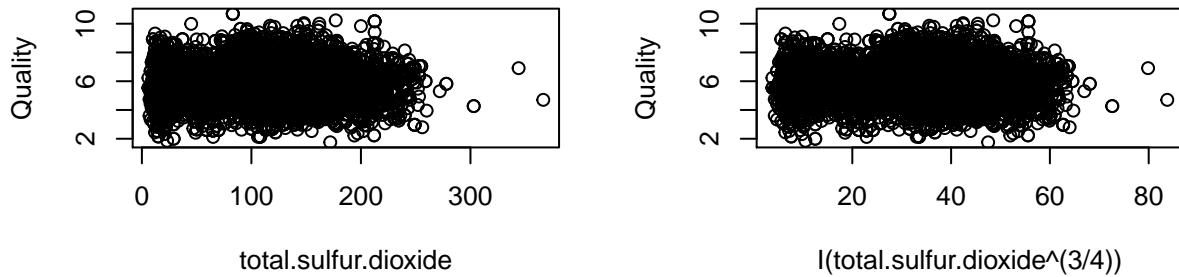
```
par(mfrow=c(2,2))
plot(Quality ~ sulphates)
plot(Quality ~ I(sulphates^-.5) )
plot(Quality ~ residual.sugar)
plot(Quality ~ log(residual.sugar))
```



```
par(mfrow=c(2,2))
plot(Quality ~ pH)
plot(Quality ~ I(pH^-1))
plot(Quality ~ free.sulfur.dioxide)
plot(Quality ~ I(free.sulfur.dioxide^(1/3)))
```

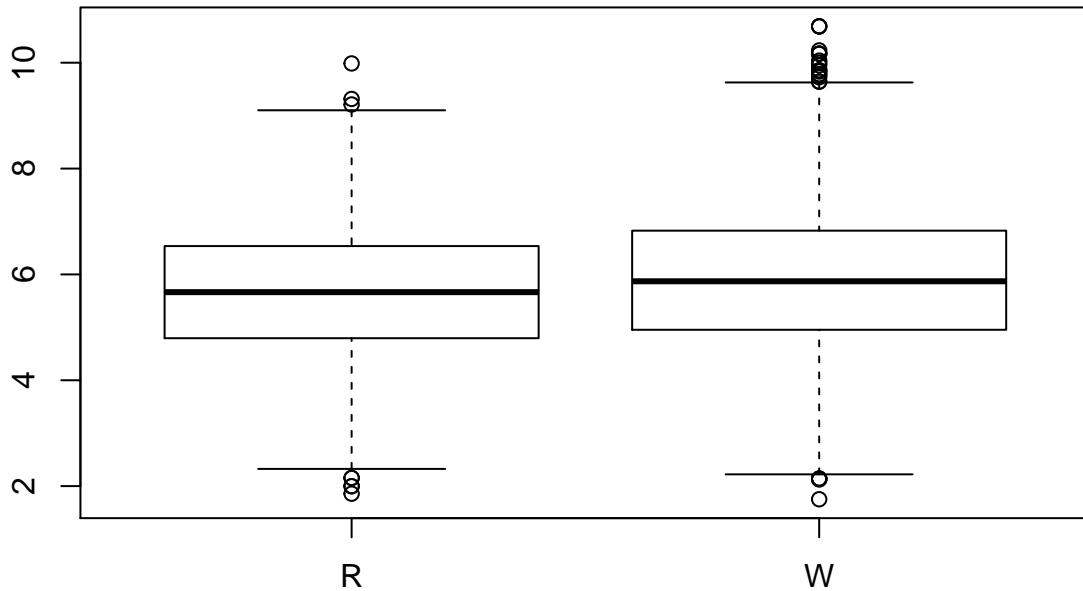


```
plot(Quality ~ total.sulfur.dioxide)
plot(Quality ~ I(total.sulfur.dioxide^(3/4)))
```



```
#now lets start checking interactions between wine color and our current predictors
#Quality ~ I(alcohol^-2) + log(volatible.acidity) + I(sulphates^-5) + log(residual.sugar) + I(pH^-1) +
summary(lm(Quality ~ Wine.Color)) #significant

##
## Call:
## lm(formula = Quality ~ Wine.Color)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.1569 -0.9314 -0.0319  0.8991  4.7787 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.66626   0.03170 178.761 < 2e-16 ***
## Wine.ColorW 0.24243   0.03655   6.632 3.56e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.321 on 6995 degrees of freedom
## Multiple R-squared:  0.006249,    Adjusted R-squared:  0.006107 
## F-statistic: 43.98 on 1 and 6995 DF,  p-value: 3.557e-11
boxplot(Quality~Wine.Color)
```



```

int1a <- lm(Quality ~ Wine.Color*I(alcohol^-2))
int1b <- lm(Quality ~ I(alcohol^-2) + Wine.Color)
anova(int1b,int1a) #signif .002

## Analysis of Variance Table
##
## Model 1: Quality ~ I(alcohol^-2) + Wine.Color
## Model 2: Quality ~ Wine.Color * I(alcohol^-2)
##   Res.Df   RSS Df Sum of Sq    F    Pr(>F)
## 1    6994 11204
## 2    6993 11190  1     14.466 9.04 0.002651 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(int1a)

##
## Call:
## lm(formula = Quality ~ Wine.Color * I(alcohol^-2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.2161 -0.8823 -0.0326  0.8533  4.8329 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 7.8886    0.1672  47.169 < 2e-16 ***

```

```

## Wine.ColorW          -0.3121      0.1857   -1.681   0.09274 .
## I(alcohol^2)        -235.4616     17.4252  -13.513  < 2e-16 ***
## Wine.ColorW:I(alcohol^2) 58.1209     19.3307    3.007   0.00265 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 6993 degrees of freedom
## Multiple R-squared:  0.08857,   Adjusted R-squared:  0.08818
## F-statistic: 226.5 on 3 and 6993 DF,  p-value: < 2.2e-16
int1a <- lm(Quality ~ log(volatile.acidity) + Wine.Color + Wine.Color:log(volatile.acidity))
int1b <- lm(Quality ~ log(volatile.acidity) + Wine.Color)
anova(int1b,int1a) #signif .01

## Analysis of Variance Table
##
## Model 1: Quality ~ log(volatile.acidity) + Wine.Color
## Model 2: Quality ~ log(volatile.acidity) + Wine.Color + Wine.Color:log(volatile.acidity)
##   Res.Df   RSS Df Sum of Sq   F Pr(>F)
## 1   6994 11944
## 2   6993 11934  1    10.374 6.0791 0.0137 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(int1a)

##
## Call:
## lm(formula = Quality ~ log(volatile.acidity) + Wine.Color + Wine.Color:log(volatile.acidity))
##
## Residuals:
##       Min     1Q     Median      3Q     Max 
## -4.0454 -0.9088 -0.0437  0.8772  5.0530 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 5.16087   0.06831 75.552 <2e-16 ***
## log(volatile.acidity)      -0.74628   0.08962 -8.328 <2e-16 ***
## Wine.ColorW                  0.09289   0.09961  0.933  0.3511  
## log(volatile.acidity):Wine.ColorW 0.25614   0.10389  2.466  0.0137 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.306 on 6993 degrees of freedom
## Multiple R-squared:  0.02798,   Adjusted R-squared:  0.02756
## F-statistic: 67.09 on 3 and 6993 DF,  p-value: < 2.2e-16
int1a <- lm(Quality ~ log(residual.sugar) + Wine.Color + Wine.Color:log(residual.sugar))
int1b <- lm(Quality ~ log(residual.sugar) + Wine.Color)
anova(int1b,int1a) #insig

## Analysis of Variance Table
##
## Model 1: Quality ~ log(residual.sugar) + Wine.Color
## Model 2: Quality ~ log(residual.sugar) + Wine.Color + Wine.Color:log(residual.sugar)
##   Res.Df   RSS Df Sum of Sq   F Pr(>F)

```

```

## 1 6994 12189
## 2 6993 12186 1 2.9292 1.6809 0.1948
int1a <- lm(Quality ~ I(free.sulfur.dioxide^(1/3)) + Wine.Color + Wine.Color:I(free.sulfur.dioxide^(1/3))
int1b <- lm(Quality ~ I(free.sulfur.dioxide^(1/3)) + Wine.Color)
anova(int1b,int1a) #insignif

## Analysis of Variance Table
##
## Model 1: Quality ~ I(free.sulfur.dioxide^(1/3)) + Wine.Color
## Model 2: Quality ~ I(free.sulfur.dioxide^(1/3)) + Wine.Color + Wine.Color:I(free.sulfur.dioxide^(1/3))
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    6994 12174
## 2    6993 12169  1     4.965 2.8532 0.09124 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
int1a <- lm(Quality ~ I(total.sulfur.dioxide^(3/4)) + Wine.Color + Wine.Color:I(total.sulfur.dioxide^(3/4))
int1b <- lm(Quality ~ I(total.sulfur.dioxide^(3/4)) + Wine.Color)
anova(int1b,int1a) #insignif

## Analysis of Variance Table
##
## Model 1: Quality ~ I(total.sulfur.dioxide^(3/4)) + Wine.Color
## Model 2: Quality ~ I(total.sulfur.dioxide^(3/4)) + Wine.Color + Wine.Color:I(total.sulfur.dioxide^(3/4))
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    6994 12062
## 2    6993 12062  1    0.52446 0.3041 0.5814
int1a <- lm(Quality ~ I(pH^-1) + Wine.Color + Wine.Color:I(pH^-1))
int1b <- lm(Quality ~ I(pH^-1) + Wine.Color)
anova(int1b,int1a) #signif 3e-7

## Analysis of Variance Table
##
## Model 1: Quality ~ I(pH^-1) + Wine.Color
## Model 2: Quality ~ I(pH^-1) + Wine.Color + Wine.Color:I(pH^-1)
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    6994 12122
## 2    6993 12077  1    45.014 26.065 3.388e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(int1a)

##
## Call:
## lm(formula = Quality ~ I(pH^-1) + Wine.Color + Wine.Color:I(pH^-1))
##
## Residuals:
##       Min     1Q     Median      3Q     Max 
## -4.3879 -0.9300 -0.0211  0.8874  4.5841 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 4.8724     0.6814   7.150 9.54e-13 ***
## I(pH^-1)    2.6243     2.2502   1.166   0.244    
## 
```

```

## Wine.ColorW          4.3465    0.7876   5.518 3.55e-08 ***
## I(pH^-1):Wine.ColorW -13.1550   2.5767  -5.105 3.39e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.314 on 6993 degrees of freedom
## Multiple R-squared:  0.01634,   Adjusted R-squared:  0.01592
## F-statistic: 38.71 on 3 and 6993 DF,  p-value: < 2.2e-16
int1a <- lm(Quality ~ I(sulphates^-0.5) + Wine.Color + Wine.Color:I(sulphates^-0.5))
int1b <- lm(Quality ~ I(sulphates^-0.5) + Wine.Color)
anova(int1b,int1a) #signif 2e-10

## Analysis of Variance Table
##
## Model 1: Quality ~ I(sulphates^-0.5) + Wine.Color
## Model 2: Quality ~ I(sulphates^-0.5) + Wine.Color + Wine.Color:I(sulphates^-0.5)
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   6994 12134
## 2   6993 12064  1     69.48 40.274 2.345e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(int1a)

##
## Call:
## lm(formula = Quality ~ I(sulphates^-0.5) + Wine.Color + Wine.Color:I(sulphates^-0.5))
##
## Residuals:
##       Min     1Q     Median     3Q     Max 
## -4.1614 -0.9308 -0.0305  0.9046  4.7604 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  8.1069    0.2910  27.854 < 2e-16 ***
## I(sulphates^-0.5) -1.9443    0.2305  -8.436 < 2e-16 ***
## Wine.ColorW -1.7351    0.3343  -5.191 2.15e-07 ***
## I(sulphates^-0.5):Wine.ColorW  1.6266    0.2563   6.346 2.34e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.313 on 6993 degrees of freedom
## Multiple R-squared:  0.01738,   Adjusted R-squared:  0.01695
## F-statistic: 41.22 on 3 and 6993 DF,  p-value: < 2.2e-16

par(mfrow=c(2,2))
plot(c(5,15), c(0,12), type='n', ylab='Quality', xlab='Alcohol')
lines(alcohol[Wine.Color=="W"], Quality[Wine.Color=="W"], type='p', col='blue', pch = 1, cex=1.2)
lines(alcohol[Wine.Color=="R"], Quality[Wine.Color=="R"], type='p', col='red', pch = 20, cex=1.2)
legend('topleft', c('White Wine 1','Red Wine 2'), pch=c(1,20,8), col=c('blue','red'), cex=0.8)
reg1<-lm(Quality[Wine.Color=="W"] ~ alcohol[Wine.Color=="W"])
abline(reg1,col="blue")
reg2<-lm(Quality[Wine.Color=="R"] ~ alcohol[Wine.Color=="R"])
abline(reg2,col="red")

```

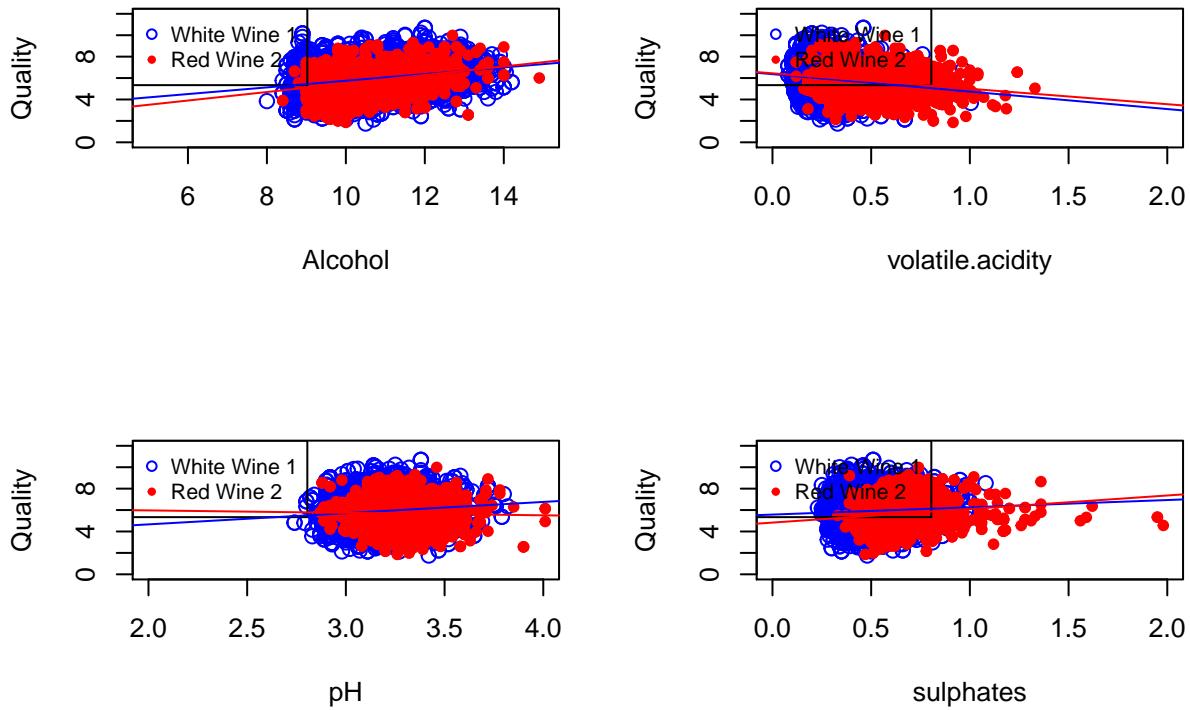
```

plot(c(0,2), c(0,12), type='n', ylab='Quality', xlab='volatile.acidity')
lines(volatile.acidity[Wine.Color=="W"], Quality[Wine.Color=="W"], type='p', col='blue', pch = 1, cex=1)
lines(volatile.acidity[Wine.Color=="R"], Quality[Wine.Color=="R"], type='p', col='red', pch = 20, cex=1)
legend('topleft', c('White Wine 1','Red Wine 2'), pch=c(1,20,8), col=c('blue','red')), cex=0.8)
reg1<-lm(Quality[Wine.Color=="W"] ~ volatile.acidity[Wine.Color=="W"])
abline(reg1,col="blue")
reg2<-lm(Quality[Wine.Color=="R"] ~ volatile.acidity[Wine.Color=="R"])
abline(reg2,col="red")

plot(c(2,4), c(0,12), type='n', ylab='Quality', xlab='pH')
lines(pH[Wine.Color=="W"], Quality[Wine.Color=="W"], type='p', col='blue', pch = 1, cex=1.2)
lines(pH[Wine.Color=="R"], Quality[Wine.Color=="R"], type='p', col='red', pch = 20, cex=1.2)
legend('topleft', c('White Wine 1','Red Wine 2'), pch=c(1,20,8), col=c('blue','red')), cex=0.8)
reg1<-lm(Quality[Wine.Color=="W"] ~ pH[Wine.Color=="W"])
abline(reg1,col="blue")
reg2<-lm(Quality[Wine.Color=="R"] ~ pH[Wine.Color=="R"])
abline(reg2,col="red")

plot(c(0,2), c(0,12), type='n', ylab='Quality', xlab='sulphates')
lines(sulphates[Wine.Color=="W"], Quality[Wine.Color=="W"], type='p', col='blue', pch = 1, cex=1.2)
lines(sulphates[Wine.Color=="R"], Quality[Wine.Color=="R"], type='p', col='red', pch = 20, cex=1.2)
legend('topleft', c('White Wine 1','Red Wine 2'), pch=c(1,20,8), col=c('blue','red')), cex=0.8)
reg1<-lm(Quality[Wine.Color=="W"] ~ sulphates[Wine.Color=="W"])
abline(reg1,col="blue")
reg2<-lm(Quality[Wine.Color=="R"] ~ sulphates[Wine.Color=="R"])
abline(reg2,col="red")

```



```
#now we add interactions and remove 1 by 1 based on anova
m4a <- lm(Quality ~ I(alcohol^2) + log(volatile.acidity) + I(sulphates^-0.5) + log(residual.sugar) +
  I(pH^-1) + I(free.sulfur.dioxide^(1/3)) + I(total.sulfur.dioxide^(3/4))+ Wine.Color:I(sulphates^-0.5)
anova(m4a)
```

```
## Analysis of Variance Table
##
## Response: Quality
##                                     Df  Sum Sq Mean Sq F value    Pr(>F)
## I(alcohol^2)                      1 1000.3 1000.26 659.2773 < 2.2e-16
## log(volatile.acidity)              1   294.1   294.06 193.8137 < 2.2e-16
## I(sulphates^-0.5)                 1    61.5    61.52 40.5468 2.041e-10
## log(residual.sugar)               1   115.7   115.66 76.2327 < 2.2e-16
## I(pH^-1)                          1    29.8    29.77 19.6230 9.577e-06
## I(free.sulfur.dioxide^(1/3))      1    44.7    44.70 29.4610 5.897e-08
## I(total.sulfur.dioxide^(3/4))     1    81.4    81.43 53.6702 2.637e-13
## Wine.Color                         1     0.5     0.47  0.3103 0.5774895
## I(sulphates^-0.5):Wine.Color      1    16.4    16.45 10.8409 0.0009977
## I(pH^-1):Wine.Color                1    34.7    34.75 22.9013 1.740e-06
## log(volatile.acidity):Wine.Color   1     2.2     2.19  1.4406 0.2300870
## I(alcohol^2):Wine.Color            1     0.0     0.00  0.0001 0.9927301
## Residuals                           6984 10596.2   1.52
##
##                                     ***
## I(alcohol^2)                      ***
## log(volatile.acidity)              ***
## I(sulphates^-0.5)                 ***
## log(residual.sugar)               ***
```

```

## I(pH^-1) ***  

## I(free.sulfur.dioxide^(1/3)) ***  

## I(total.sulfur.dioxide^(3/4)) ***  

## Wine.Color  

## I(sulphates^-0.5):Wine.Color ***  

## I(pH^-1):Wine.Color ***  

## log(volatile.acidity):Wine.Color  

## I(alcohol^-2):Wine.Color  

## Residuals  

## ---  

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

#remove color:alc, contribution is 0 to SSR  

m4b <-lm(Quality ~ I(alcohol^-2) + log(volatile.acidity) + I(sulphates^-0.5) + log(residual.sugar) +  

    I(pH^-1) + I(free.sulfur.dioxide^(1/3)) + I(total.sulfur.dioxide^(3/4))+ Wine.Color:I(sulphates^-0.5)  

anova(m4b)

## Analysis of Variance Table  

##  

## Response: Quality  

##  

## Df Sum Sq Mean Sq F value Pr(>F)  

## I(alcohol^-2) 1 1000.3 1000.26 659.3717 < 2.2e-16  

## log(volatile.acidity) 1 294.1 294.06 193.8415 < 2.2e-16  

## I(sulphates^-0.5) 1 61.5 61.52 40.5526 2.035e-10  

## log(residual.sugar) 1 115.7 115.66 76.2436 < 2.2e-16  

## I(pH^-1) 1 29.8 29.77 19.6258 9.563e-06  

## I(free.sulfur.dioxide^(1/3)) 1 44.7 44.70 29.4652 5.884e-08  

## I(total.sulfur.dioxide^(3/4)) 1 81.4 81.43 53.6779 2.627e-13  

## Wine.Color 1 0.5 0.47 0.3104 0.5774622  

## I(sulphates^-0.5):Wine.Color 1 16.4 16.45 10.8425 0.0009969  

## I(pH^-1):Wine.Color 1 34.7 34.75 22.9046 1.737e-06  

## log(volatile.acidity):Wine.Color 1 2.2 2.19 1.4408 0.2300536  

## Residuals 6985 10596.2 1.52  

##  

## I(alcohol^-2) ***  

## log(volatile.acidity) ***  

## I(sulphates^-0.5) ***  

## log(residual.sugar) ***  

## I(pH^-1) ***  

## I(free.sulfur.dioxide^(1/3)) ***  

## I(total.sulfur.dioxide^(3/4)) ***  

## Wine.Color  

## I(sulphates^-0.5):Wine.Color ***  

## I(pH^-1):Wine.Color ***  

## log(volatile.acidity):Wine.Color  

## Residuals  

## ---  

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

summary(m4b)

##  

## Call:  

## lm(formula = Quality ~ I(alcohol^-2) + log(volatile.acidity) +  

##     I(sulphates^-0.5) + log(residual.sugar) + I(pH^-1) + I(free.sulfur.dioxide^(1/3)) +  

##     I(total.sulfur.dioxide^(3/4)) + Wine.Color:I(sulphates^-0.5) +

```

```

##      Wine.Color:I(pH^-1) + Wine.Color:log(volatile.acidity) +
##      Wine.Color)
##
## Residuals:
##      Min       1Q   Median      3Q      Max
## -3.9540 -0.8413 -0.0123  0.8287  4.6426
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               6.514e+00 7.474e-01  8.715 < 2e-16
## I(alcohol^-2)            -1.991e+02 8.631e+00 -23.069 < 2e-16
## log(volatile.acidity)    -3.523e-01 9.011e-02 -3.909 9.35e-05
## I(sulphates^-0.5)        -1.164e+00 2.256e-01 -5.160 2.54e-07
## log(residual.sugar)      1.800e-01 2.088e-02  8.620 < 2e-16
## I(pH^-1)                  5.120e+00 2.202e+00  2.325 0.02009
## I(free.sulfur.dioxide^(1/3)) 3.375e-01 3.705e-02  9.111 < 2e-16
## I(total.sulfur.dioxide^(3/4)) -1.496e-02 2.375e-03 -6.301 3.14e-10
## Wine.ColorW                2.703e+00 8.430e-01  3.207 0.00135
## I(sulphates^-0.5):Wine.ColorW 7.752e-01 2.498e-01  3.103 0.00192
## I(pH^-1):Wine.ColorW      -1.229e+01 2.501e+00 -4.914 9.11e-07
## log(volatile.acidity):Wine.ColorW -1.247e-01 1.039e-01 -1.200 0.23005
##
## (Intercept)                 ***
## I(alcohol^-2)                ***
## log(volatile.acidity)        ***
## I(sulphates^-0.5)             ***
## log(residual.sugar)          ***
## I(pH^-1)                      *
## I(free.sulfur.dioxide^(1/3)) ***
## I(total.sulfur.dioxide^(3/4)) ***
## Wine.ColorW                   **
## I(sulphates^-0.5):Wine.ColorW **
## I(pH^-1):Wine.ColorW          ***
## log(volatile.acidity):Wine.ColorW
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.232 on 6985 degrees of freedom
## Multiple R-squared:  0.1369, Adjusted R-squared:  0.1356
## F-statistic: 100.8 on 11 and 6985 DF,  p-value: < 2.2e-16
vif(m4b)

##                               I(alcohol^-2)           log(volatile.acidity)
##                               1.380682              7.479800
##                               I(sulphates^-0.5)       log(residual.sugar)
##                               7.485827              1.493497
##                               I(pH^-1)             I(free.sulfur.dioxide^(1/3))
##                               5.176032              2.614084
##                               I(total.sulfur.dioxide^(3/4)) Wine.Color
##                               4.795692              611.462972
##                               I(sulphates^-0.5):Wine.Color I(pH^-1):Wine.Color
##                               119.735014              536.165061
## log(volatile.acidity):Wine.Color
##                               20.993167

```

```

#remove wine color
m4c <- lm(Quality ~ I(alcohol^-2) + log(volatile.acidity) + I(sulphates^-0.5) + log(residual.sugar) +
  I(pH^-1) + I(free.sulfur.dioxide^(1/3)) + I(total.sulfur.dioxide^(3/4))+ Wine.Color:I(sulphates^-0.5)
anova(m4c)

## Analysis of Variance Table
##
## Response: Quality
##                                     Df  Sum Sq Mean Sq F value    Pr(>F)
## I(alcohol^-2)                      1 1000.3 1000.26 658.4966 < 2.2e-16
## log(volatile.acidity)               1  294.1  294.06 193.5842 < 2.2e-16
## I(sulphates^-0.5)                  1   61.5   61.52  40.4988 2.092e-10
## log(residual.sugar)                1   115.7   115.66  76.1424 < 2.2e-16
## I(pH^-1)                          1    29.8    29.77  19.5998 9.693e-06
## I(free.sulfur.dioxide^(1/3))       1    44.7    44.70  29.4261 6.004e-08
## I(total.sulfur.dioxide^(3/4))      1    81.4    81.43  53.6067 2.723e-13
## I(sulphates^-0.5):Wine.Color      1     2.4     2.44   1.6095  0.2046
## I(pH^-1):Wine.Color              1    33.2    33.18  21.8447 3.012e-06
## log(volatile.acidity):Wine.Color  1     2.6     2.62   1.7268  0.1889
## Residuals                         6986 10611.8    1.52
##
##                                     ***
## I(alcohol^-2)                      ***
## log(volatile.acidity)               ***
## I(sulphates^-0.5)                  ***
## log(residual.sugar)                ***
## I(pH^-1)                          ***
## I(free.sulfur.dioxide^(1/3))       ***
## I(total.sulfur.dioxide^(3/4))      ***
## I(sulphates^-0.5):Wine.Color      ***
## I(pH^-1):Wine.Color              ***
## log(volatile.acidity):Wine.Color
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(m4c)

##
## Call:
## lm(formula = Quality ~ I(alcohol^-2) + log(volatile.acidity) +
##     I(sulphates^-0.5) + log(residual.sugar) + I(pH^-1) + I(free.sulfur.dioxide^(1/3)) +
##     I(total.sulfur.dioxide^(3/4)) + Wine.Color:I(sulphates^-0.5) +
##     Wine.Color:I(pH^-1) + Wine.Color:log(volatile.acidity))
##
## Residuals:
##      Min      1Q Median      3Q      Max 
## -3.9257 -0.8520 -0.0101  0.8391  4.6262 
##
## Coefficients:
## (Intercept)          Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          8.610e+00 3.626e-01 23.746 < 2e-16  
## I(alcohol^-2)        -1.984e+02 8.634e+00 -22.981 < 2e-16  
## log(volatile.acidity) -3.589e-01 9.015e-02 -3.981 6.94e-05 
## I(sulphates^-0.5)     -1.466e+00 2.050e-01 -7.151 9.50e-13 
## log(residual.sugar)   1.751e-01 2.084e-02  8.400 < 2e-16 

```

```

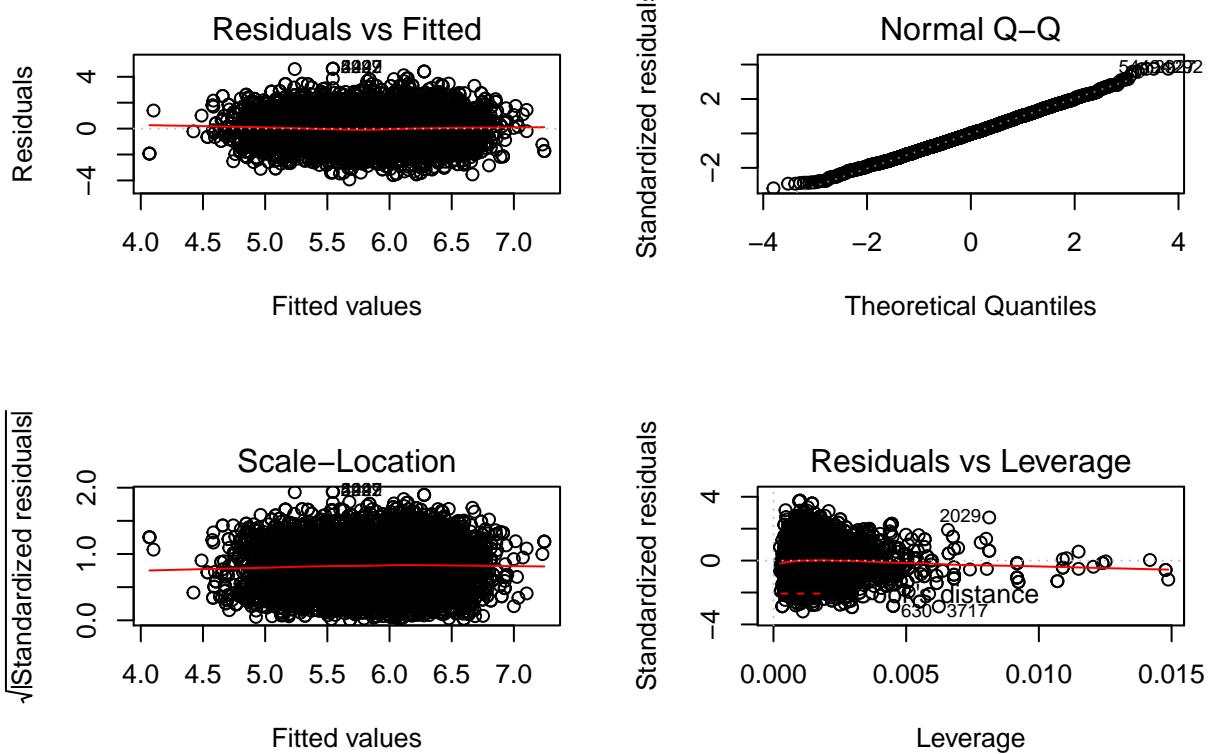
## I(pH^-1)           -5.261e-01  1.323e+00  -0.398    0.691
## I(free.sulfur.dioxide^(1/3))   3.257e-01  3.689e-02   8.829  < 2e-16
## I(total.sulfur.dioxide^(3/4)) -1.400e-02  2.358e-03  -5.940  2.98e-09
## I(sulphates^-0.5):Wine.ColorW 1.125e+00  2.249e-01   5.002  5.82e-07
## I(pH^-1):Wine.ColorW        -5.012e+00  1.051e+00  -4.768  1.90e-06
## log(volatile.acidity):Wine.ColorW -1.366e-01  1.039e-01  -1.314    0.189
##
## (Intercept)      ***
## I(alcohol^-2)     ***
## log(volatile.acidity) ***
## I(sulphates^-0.5) ***
## log(residual.sugar) ***
## I(pH^-1)          ***
## I(free.sulfur.dioxide^(1/3)) ***
## I(total.sulfur.dioxide^(3/4)) ***
## I(sulphates^-0.5):Wine.ColorW ***
## I(pH^-1):Wine.ColorW ***
## log(volatile.acidity):Wine.ColorW
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.232 on 6986 degrees of freedom
## Multiple R-squared:  0.1357, Adjusted R-squared:  0.1344
## F-statistic: 109.7 on 10 and 6986 DF,  p-value: < 2.2e-16

vif(m4c)

##                   I(alcohol^-2)          log(volatile.acidity)
##                   1.379820            7.475893
##                   I(sulphates^-0.5)      log(residual.sugar)
##                   6.177539            1.485346
##                   I(pH^-1)            I(free.sulfur.dioxide^(1/3))
##                   1.866697            2.588098
##       I(total.sulfur.dioxide^(3/4)) I(sulphates^-0.5):Wine.Color
##                   4.719543            96.919252
##                   I(pH^-1):Wine.Color log(volatile.acidity):Wine.Color
##                   94.600368            20.966731

plot(m4c)

```



```
wine <- wine[-c(630,3717,2029),]
detach(wine)
attach(wine)

#we should compare our current model if all transformations were simply logs to see if we are overfitting
m4d <- lm(Quality ~ log(alcohol) + log(volatile.acidity) + log(sulphates) + log(residual.sugar) +
           log(pH) + log(free.sulfur.dioxide) + log(total.sulfur.dioxide)+ Wine.Color:log(sulphates) + Wine.Color)

anova(m4d)

## Analysis of Variance Table
##
## Response: Quality
##                         Df  Sum Sq Mean Sq  F value    Pr(>F)
## log(alcohol)            1 1064.6 1064.63 707.1926 < 2.2e-16
## log(volatile.acidity)   1  296.5  296.51 196.9583 < 2.2e-16
## log(sulphates)          1   74.7   74.69  49.6136 2.051e-12
## log(residual.sugar)     1  107.1  107.13  71.1639 < 2.2e-16
## log(pH)                 1   36.5   36.45  24.2130 8.822e-07
## log(free.sulfur.dioxide) 1   50.3   50.33  33.4291 7.710e-09
## log(total.sulfur.dioxide) 1   94.4   94.36  62.6772 2.813e-15
## log(sulphates):Wine.Color 1     9.8     9.79   6.5005  0.01081
## log(pH):Wine.Color       1     1.4     1.43   0.9494  0.32991
## log(volatile.acidity):Wine.Color 1     0.2     0.18   0.1166  0.73278
## Residuals                6983 10512.5      1.51
## log(alcohol)               ***
```

```

## log(volatile.acidity)      ***
## log(sulphates)            ***
## log(residual.sugar)       ***
## log(pH)                   ***
## log(free.sulfur.dioxide)  ***
## log(total.sulfur.dioxide) ***
## log(sulphates):Wine.Color *
## log(pH):Wine.Color
## log(volatile.acidity):Wine.Color
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(m4d)

##
## Call:
## lm(formula = Quality ~ log(alcohol) + log(volatile.acidity) +
##     log(sulphates) + log(residual.sugar) + log(pH) + log(free.sulfur.dioxide) +
##     log(total.sulfur.dioxide) + Wine.Color:log(sulphates) + Wine.Color:log(pH) +
##     Wine.Color:log(volatile.acidity))
##
## Residuals:
##    Min      1Q  Median      3Q      Max
## -3.9066 -0.8481 -0.0143  0.8310  4.5365
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -4.53989  0.55829 -8.132 4.97e-16 ***
## log(alcohol)                 3.66538  0.15018 24.406 < 2e-16 ***
## log(volatile.acidity)      -0.45887  0.08714 -5.266 1.43e-07 ***
## log(sulphates)               0.72369  0.13611  5.317 1.09e-07 ***
## log(residual.sugar)         0.15241  0.02023  7.535 5.48e-14 ***
## log(pH)                      1.49689  0.33669  4.446 8.89e-06 ***
## log(free.sulfur.dioxide)     0.34206  0.03539  9.666 < 2e-16 ***
## log(total.sulfur.dioxide)   -0.32130  0.04694 -6.844 8.33e-12 ***
## log(sulphates):Wine.ColorW -0.39225  0.15584 -2.517  0.0119 *
## log(pH):Wine.ColorW        -0.11228  0.11835 -0.949  0.3428
## log(volatile.acidity):Wine.ColorW -0.03415  0.10001 -0.341  0.7328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.227 on 6983 degrees of freedom
## Multiple R-squared:  0.1417, Adjusted R-squared:  0.1405
## F-statistic: 115.3 on 10 and 6983 DF,  p-value: < 2.2e-16
#we actually see a smaller SSE and suspect our power transforms were overfitting. get rid of interaction
m5a <- lm(Quality ~ log(alcohol) + log(volatile.acidity) + log(sulphates) + log(residual.sugar) +
           log(pH) + log(free.sulfur.dioxide) + log(total.sulfur.dioxide)+ Wine.Color:log(sulphates) + Wine.Co
anova(m5a)

## Analysis of Variance Table
##
## Response: Quality
##                               Df  Sum Sq Mean Sq  F value    Pr(>F)
## log(alcohol)                  1  1064.6 1064.63 707.2821 < 2.2e-16 ***

```

```

## log(volatile.acidity)      1   296.5  296.51 196.9832 < 2.2e-16 ***
## log(sulphates)            1    74.7   74.69  49.6199 2.044e-12 ***
## log(residual.sugar)       1   107.1  107.13 71.1729 < 2.2e-16 ***
## log(pH)                   1    36.5   36.45  24.2161 8.808e-07 ***
## log(free.sulfur.dioxide)  1    50.3   50.33  33.4333 7.693e-09 ***
## log(total.sulfur.dioxide) 1    94.4   94.36  62.6851 2.802e-15 ***
## log(sulphates):Wine.Color 1     9.8    9.79   6.5013   0.0108 *
## log(pH):Wine.Color        1     1.4    1.43   0.9495   0.3299
## Residuals                 6984 10512.6   1.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove interaction between ph and color
m5b <- lm(Quality ~ log(alcohol) + log(volatile.acidity) + log(sulphates) + log(residual.sugar) +
           log(pH) + log(free.sulfur.dioxide) + log(total.sulfur.dioxide)+ Wine.Color:log(sulphates))
anova(m5b) #significant contributions to SS

## Analysis of Variance Table
##
## Response: Quality
##                               Df  Sum Sq Mean Sq F value    Pr(>F)
## log(alcohol)                1 1064.6 1064.63 707.2872 < 2.2e-16 ***
## log(volatile.acidity)       1   296.5  296.51 196.9846 < 2.2e-16 ***
## log(sulphates)              1    74.7   74.69  49.6203 2.044e-12 ***
## log(residual.sugar)         1   107.1  107.13 71.1734 < 2.2e-16 ***
## log(pH)                     1    36.5   36.45  24.2162 8.808e-07 ***
## log(free.sulfur.dioxide)    1    50.3   50.33  33.4335 7.692e-09 ***
## log(total.sulfur.dioxide)   1    94.4   94.36  62.6856 2.801e-15 ***
## log(sulphates):Wine.Color  1     9.8    9.79   6.5014   0.0108 *
## Residuals                  6985 10514.1   1.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(m5b)

##
## Call:
## lm(formula = Quality ~ log(alcohol) + log(volatile.acidity) +
##     log(sulphates) + log(residual.sugar) + log(pH) + log(free.sulfur.dioxide) +
##     log(total.sulfur.dioxide) + Wine.Color:log(sulphates))
##
## Residuals:
##      Min      1Q Median      3Q      Max 
## -3.9117 -0.8493 -0.0138  0.8294  4.5463 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -4.38950   0.53718  -8.171 3.59e-16 ***
## log(alcohol)              3.64177   0.14805   24.598 < 2e-16 ***
## log(volatile.acidity)   -0.47146   0.04244  -11.108 < 2e-16 ***
## log(sulphates)            0.65119   0.11650   5.590 2.36e-08 ***
## log(residual.sugar)       0.15106   0.02012   7.509 6.71e-14 ***
## log(pH)                   1.42587   0.32955   4.327 1.53e-05 ***
## log(free.sulfur.dioxide)  0.34779   0.03494   9.955 < 2e-16 ***
## log(total.sulfur.dioxide) -0.34142   0.04225  -8.082 7.47e-16 ***

```

```

## log(sulphates):Wine.ColorW -0.27801      0.10903   -2.550    0.0108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.227 on 6985 degrees of freedom
## Multiple R-squared:  0.1416, Adjusted R-squared:  0.1406
## F-statistic:  144 on 8 and 6985 DF,  p-value: < 2.2e-16
#m5b is our final model
plot(m5b)

```

