

Stats 101A Final Project

Suoyi Yang

12/8/2017

Starting Off

```
library(car)

## Loading required package: carData
library(MASS)

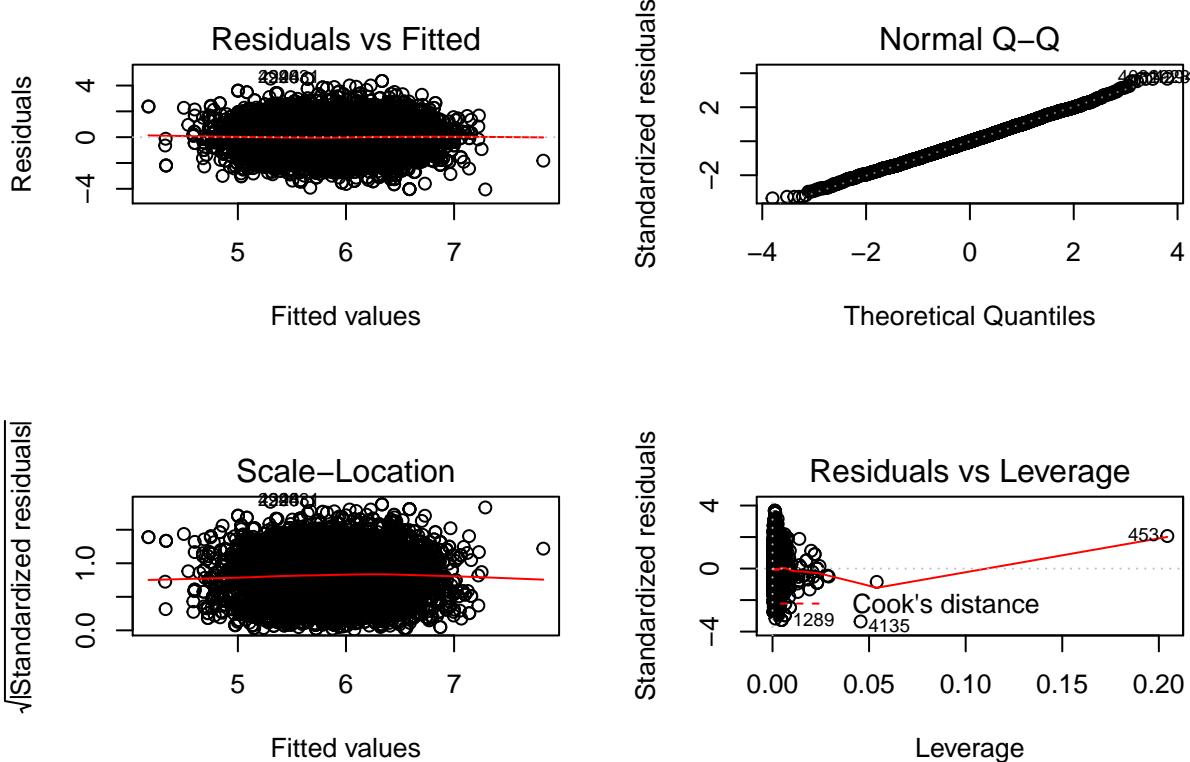
## Warning: package 'MASS' was built under R version 3.5.2
firstWineTraining <-read.csv("Wine2017Training2.csv", stringsAsFactors = TRUE)

fwineColor <- firstWineTraining$Wine.Color
ffixedAcidity <- firstWineTraining$fixed.acidity
fvolatileAcidity <- firstWineTraining$volatile.acidity
fcitricAcid <- firstWineTraining$citric.acid
fresidualSugar <-firstWineTraining$residual.sugar
fchlorides <- firstWineTraining$chlorides
ffreeSulfurDioxide <- firstWineTraining$free.sulfur.dioxide
ftotalSulfurDioxide <- firstWineTraining$total.sulfur.dioxide
fwineDensity <- firstWineTraining$density
fph <- firstWineTraining$pH
fsulphates <- firstWineTraining$sulphates
falcohol <- firstWineTraining$alcohol
fquality <- firstWineTraining$Quality

m1 <- lm(fquality~ fwineColor + ffixedAcidity + fvolatileAcidity + fcitricAcid + fresidualSugar + fchlorides + ffreeSulfurDioxide + ftotalSulfurDioxide + fwineDensity + fph + fsulphates + falcohol)
```

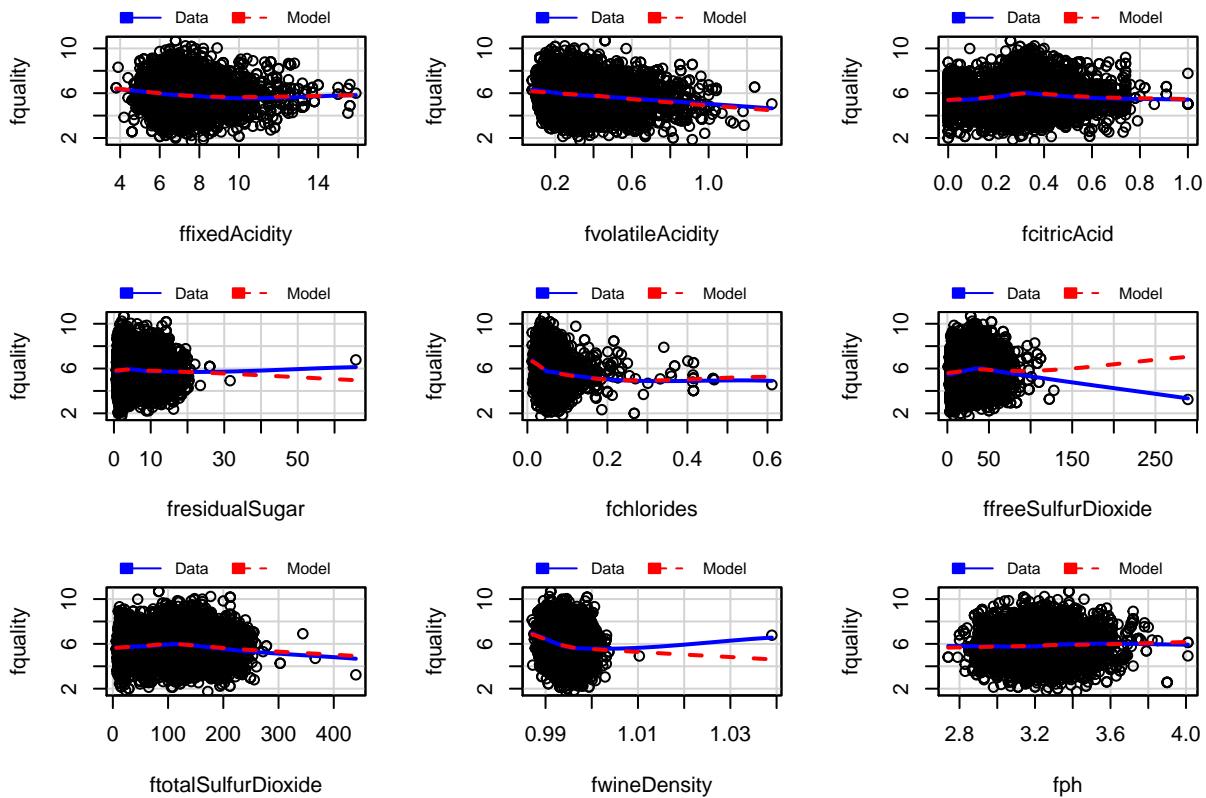
Diagnostics

```
par(mfrow=c(2,2))
plot(m1)
```



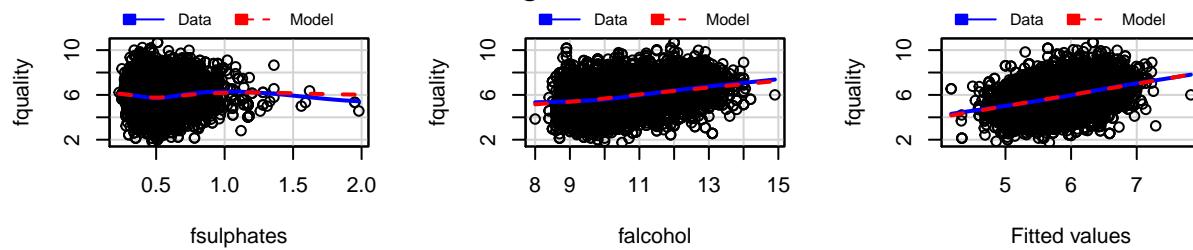
#Point 453 is a leverage point...not bad (yet) but suspicious
1289, 4125 also look out

mmps(m1)



```
## Warning in mmpps(m1): Interactions and/or factors skipped
```

Marginal Model Plots



```
# some points in residual sugar vs. quality, free sulphur dioxide vs quality, and quality vs. wine dens
```

```
#check them
```

```
plot(fresidualSugar, fquality)
```

```
plot(fwineDensity, fquality)
```

```
plot(ffreeSulfurDioxide, fquality)
```

```
firstWineTraining[fresidualSugar > 60,]$Case
```

```
## [1] 453
```

```
firstWineTraining[fwineDensity > 1.03,]$Case
```

```
## [1] 453
```

```
firstWineTraining[ffreeSulfurDioxide > 250,]$Case
```

```
## [1] 4135
```

```
#remove 453, 4135
```

```
newWineTraining <- firstWineTraining[-c(453,4135),]
```

```
wineColor <- newWineTraining$Wine.Color
```

```
fixedAcidity <- newWineTraining$fixed.acidity
```

```
volatileAcidity <- newWineTraining$volatile.acidity
```

```
citricAcid <- newWineTraining$citric.acid
```

```
residualSugar <- newWineTraining$residual.sugar
```

```
chlorides <- newWineTraining$chlorides
```

```
freeSulfurDioxide <- newWineTraining$free.sulfur.dioxide
```

```
totalSulfurDioxide <- newWineTraining$total.sulfur.dioxide
```

```
wineDensity <- newWineTraining$density
```

```
ph <- newWineTraining$pH
```

```
sulphates <- newWineTraining$sulphates
```

```
alcohol <- newWineTraining$alcohol
```

```
quality <- newWineTraining$Quality
```

```
m2 <- lm(quality ~ wineColor + fixedAcidity + volatileAcidity + citricAcid + residualSugar + chlorides + freeSulfurDioxide + totalSulfurDioxide + wineDensity + ph + sulphates + alcohol)
```

```
summary(m2)
```

```
##
```

```
## Call:
```

```
## lm(formula = quality ~ wineColor + fixedAcidity + volatileAcidity +
```

```
##      citricAcid + residualSugar + chlorides + freeSulfurDioxide +
```

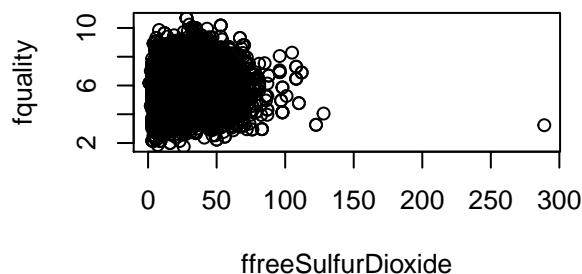
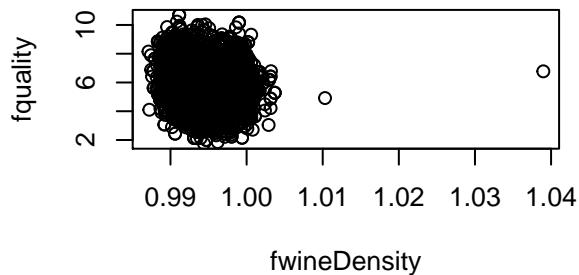
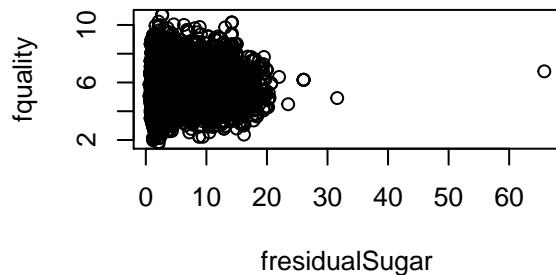
```
##      totalSulfurDioxide + wineDensity + ph + sulphates + alcohol)
```

```

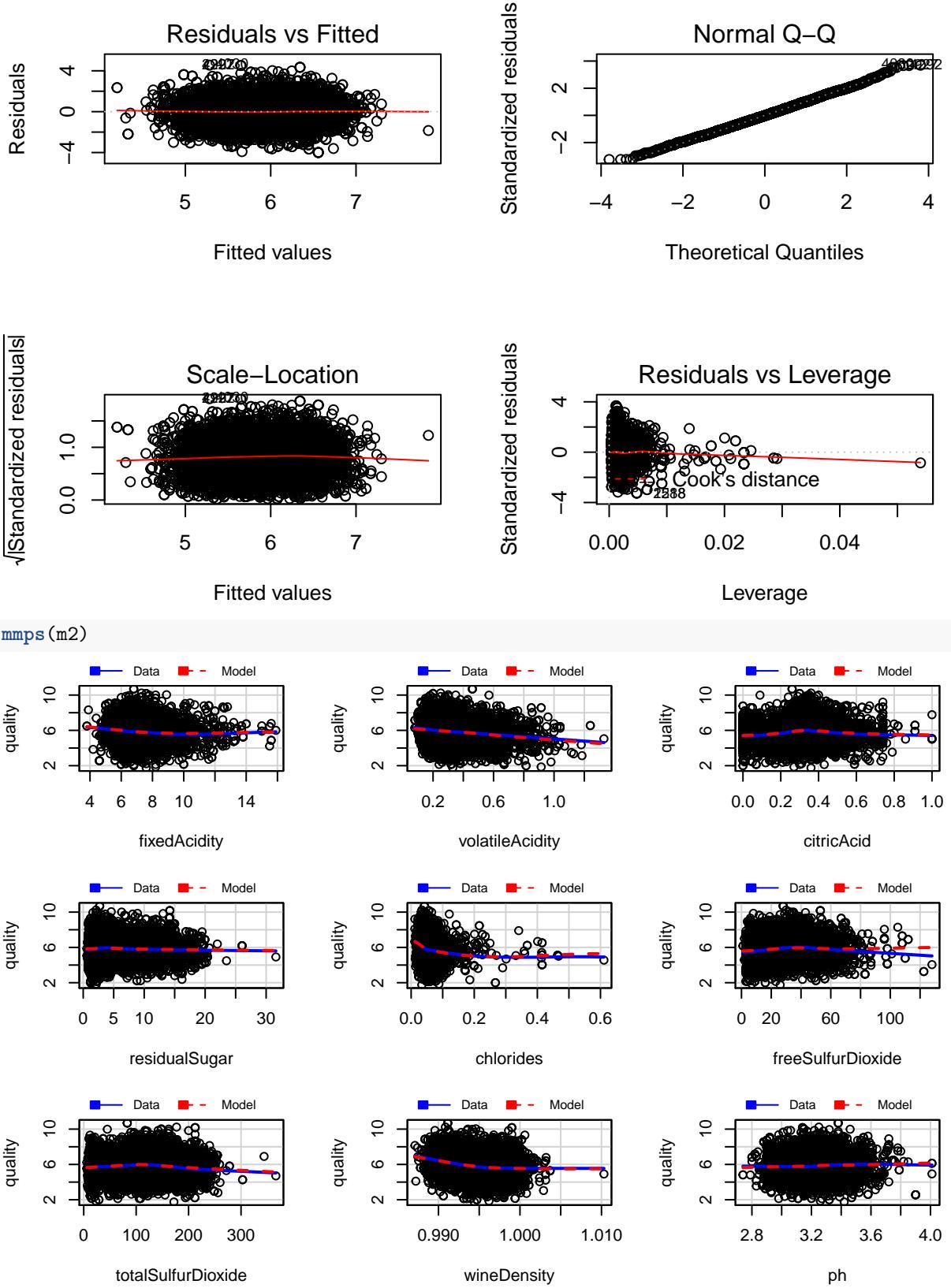
##
## Residuals:
##      Min       1Q   Median      3Q      Max
## -3.9956 -0.8467 -0.0254  0.8392  4.5512
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.377e+02  2.534e+01  5.432 5.77e-08 ***
## wineColorW          -2.623e-01  9.614e-02 -2.728 0.006393 **
## fixedAcidity        1.420e-01  2.718e-02  5.225 1.79e-07 ***
## volatileAcidity     -1.294e+00  1.325e-01 -9.768 < 2e-16 ***
## citricAcid         -1.100e-01  1.320e-01 -0.833 0.404916
## residualSugar       7.808e-02  1.016e-02  7.685 1.74e-14 ***
## chlorides          -1.452e-01  5.724e-01 -0.254 0.799701
## freeSulfurDioxide  7.798e-03  1.284e-03  6.075 1.31e-09 ***
## totalSulfurDioxide -1.881e-03  5.333e-04 -3.528 0.000422 ***
## wineDensity         -1.393e+02  2.568e+01 -5.424 6.04e-08 ***
## ph                  1.081e+00  1.520e-01  7.113 1.25e-12 ***
## sulphates          9.280e-01  1.243e-01  7.463 9.45e-14 ***
## alcohol             1.886e-01  3.219e-02  5.861 4.81e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.233 on 6985 degrees of freedom
## Multiple R-squared:  0.1354, Adjusted R-squared:  0.1339
## F-statistic: 91.16 on 12 and 6985 DF,  p-value: < 2.2e-16

```

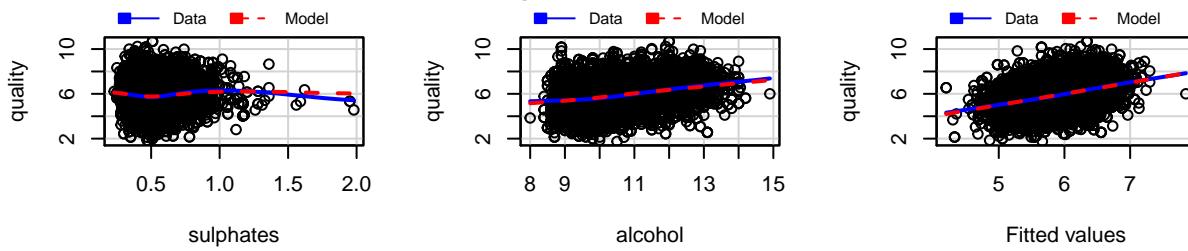
```
par(mfrow=c(2,2))
```



```
plot(m2)
```



Marginal Model Plots



Some Transforms

```
summary(powerTransform(cbind(quality, fixedAcidity, volatileAcidity, residualSugar, chlorides, freeSulfurDioxide, totalSulfurDioxide, wineDensity, ph, sulphates, alcohol)~1))

## bcPower Transformations to Multinormality
##           Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## quality      0.8185      0.82      0.7384      0.8985
## fixedAcidity -0.3130     -0.33     -0.3843     -0.2416
## volatileAcidity -0.0497    -0.05     -0.0929     -0.0065
## residualSugar  0.3113      0.33      0.2869      0.3357
## chlorides     -0.4008     -0.40     -0.4318     -0.3698
## freeSulfurDioxide  0.3480      0.33      0.3215      0.3744
## totalSulfurDioxide  0.6533      0.65      0.6254      0.6812
## wineDensity    -37.7457   -37.75    -41.3717   -34.1197
## ph            0.5654      0.50      0.2429      0.8878
## sulphates     -0.3420     -0.33     -0.4076     -0.2764
## alcohol       -0.5302     -0.50     -0.6916     -0.3687
##
## Likelihood ratio test that transformation parameters are equal to 0
## (all log transformations)
##                         LRT df      pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0 0) 5478.381 11 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                         LRT df      pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1 1) 21126.24 11 < 2.22e-16

caWine <- newWineTraining
caWine[citricAcid==0]$citric.acid <- 0.00000001
summary(powerTransform(cbind(caWine$citric.acid)~1))

## bcPower Transformation to Normality
##           Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1      0.6277      0.63      0.6103      0.6452
##
## Likelihood ratio test that transformation parameter is equal to 0
## (log transformation)
##                         LRT df      pval
## LR test, lambda = (0) 18868.29  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                         LRT df      pval
## LR test, lambda = (1) 1254.714  1 < 2.22e-16
```

```

#suggested Transforms (might not use all of them)
fa <- fixedAcidity^(1/3)
va <- log(volatileAcidity)
rs <- residualSugar^(1/3)
chl <- chlorides^(1/2) #yeah
fsd <- freeSulfurDioxide^(1/3) #yeah
tsd <- totalSulfurDioxide^(2/3)
newph <- ph^(1/2)
slpht <- sulphates^(1/3) #yeah
d <- wineDensity^(40)
a <- alcohol^(1/2)
ca <- citricAcid^(2/3)

#using all transformation
m3 <- lm(quality ~ wineColor + fa + va + ca + rs + chl +
          + fsd + tsd + newph + slpht + a + d)
summary(m3)

##
## Call:
## lm(formula = quality ~ wineColor + fa + va + ca + rs + chl +
##      + fsd + tsd + newph + slpht + a + d)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -3.8381 -0.8438 -0.0209  0.8248  4.5634
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.715146  1.543025  1.112  0.26637
## wineColorW -0.273358  0.096009 -2.847  0.00442 **
## fa          -6.219734  1.116858 -5.569 2.66e-08 ***
## va          -0.474662  0.047213 -10.054 < 2e-16 ***
## ca           0.118761  0.121282  0.979  0.32751
## rs           0.804264  0.085802  9.374 < 2e-16 ***
## chl          0.000234  0.025189  0.009  0.99259
## fsd          0.296206  0.037722  7.852 4.69e-15 ***
## tsd          -0.020946  0.003982 -5.260 1.48e-07 ***
## newph         3.690826  0.523376  7.052 1.93e-12 ***
## slpht        -1.303201  0.170554 -7.641 2.44e-14 ***
## a            -11.225336  2.155315 -5.208 1.96e-07 ***
## d             3.009461  0.456303  6.595 4.55e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.229 on 6985 degrees of freedom
## Multiple R-squared:  0.1415, Adjusted R-squared:  0.14
## F-statistic: 95.92 on 12 and 6985 DF,  p-value: < 2.2e-16

#using some transformation that mmps suggested
m4 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + residualSugar + chl +
          + fsd + totalSulfurDioxide + wineDensity + ph + slpht + alcohol)
summary(m4)

```

```

## 
## Call:
## lm(formula = quality ~ wineColor + fixedAcidity + va + citricAcid +
##      residualSugar + chlr + fsd + totalSulfurDioxide + wineDensity +
##      ph + slpht + alcohol)
## 
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -3.8782 -0.8516 -0.0279  0.8238  4.5471 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.240e+02 2.603e+01  4.764 1.94e-06 ***
## wineColorW -2.547e-01 9.664e-02 -2.636  0.00841 **  
## fixedAcidity 1.328e-01 2.732e-02  4.862 1.19e-06 *** 
## va          -4.705e-01 4.678e-02 -10.057 < 2e-16 ***
## citricAcid  -1.231e-02 1.273e-01 -0.097  0.92301  
## residualSugar 7.359e-02 1.035e-02  7.113 1.25e-12 *** 
## chlr         8.711e-04 2.534e-02  0.034  0.97258  
## fsd          2.967e-01 3.665e-02  8.097 6.62e-16 *** 
## totalSulfurDioxide -2.503e-03 5.459e-04 -4.585 4.63e-06 *** 
## wineDensity  -1.248e+02 2.626e+01 -4.755 2.03e-06 *** 
## ph           9.424e-01 1.517e-01  6.211 5.57e-10 *** 
## slpht        -1.227e+00 1.712e-01 -7.164 8.62e-13 *** 
## alcohol       2.115e-01 3.199e-02  6.612 4.07e-11 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 1.23 on 6985 degrees of freedom
## Multiple R-squared:  0.14, Adjusted R-squared:  0.1385 
## F-statistic: 94.76 on 12 and 6985 DF, p-value: < 2.2e-16

#maximizing R2 using transforms
m5 <-lm(quality~ wineColor + fixedAcidity + va + citricAcid + rs + chlr
         + fsd + tsd + wineDensity + ph + sulphates + alcohol)
summary(m5)

```

```

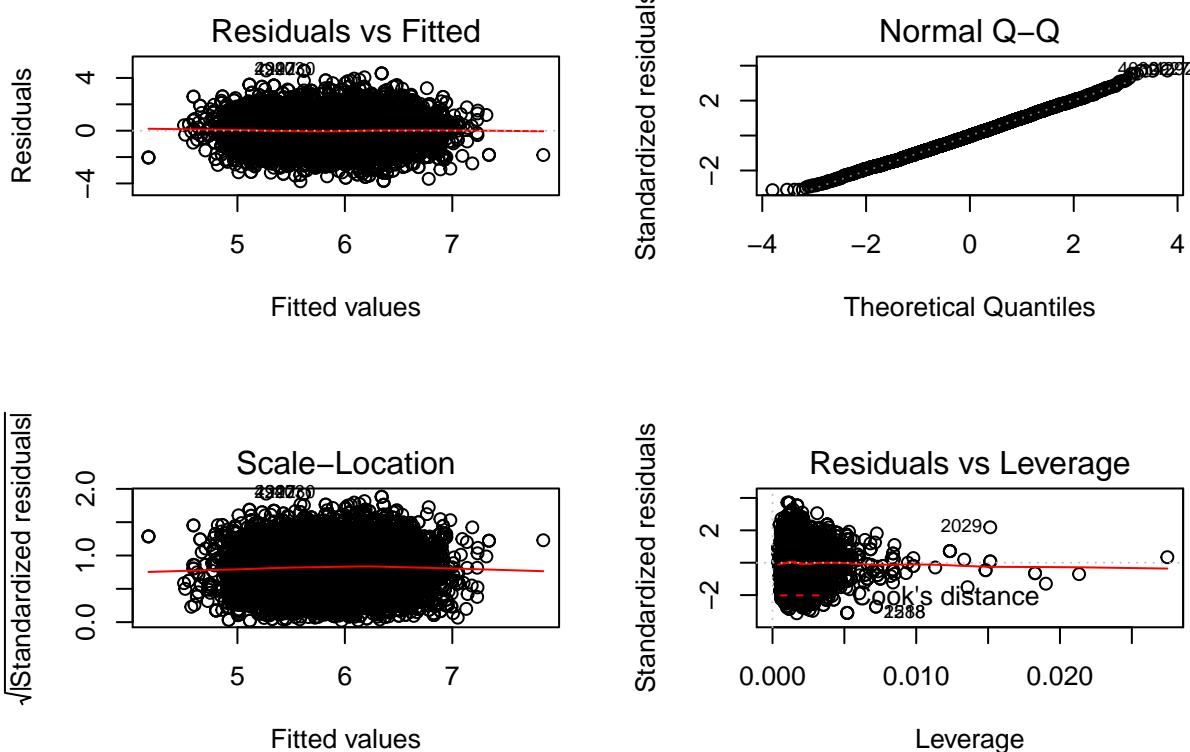
## 
## Call:
## lm(formula = quality ~ wineColor + fixedAcidity + va + citricAcid +
##      rs + chlr + fsd + tsd + wineDensity + ph + sulphates + alcohol)
## 
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -3.8347 -0.8418 -0.0244  0.8277  4.5665 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.078e+02 2.243e+01  4.805 1.58e-06 *** 
## wineColorW -1.751e-01 9.553e-02 -1.833  0.0668 .  
## fixedAcidity 1.149e-01 2.463e-02  4.664 3.16e-06 *** 
## va          -4.884e-01 4.669e-02 -10.460 < 2e-16 *** 
## citricAcid -1.449e-02 1.274e-01 -0.114  0.9095  
## rs           6.696e-01 8.721e-02  7.679 1.82e-14 *** 
## chlr        6.949e-03 2.511e-02  0.277  0.7820 

```

```

## fsd          3.083e-01  3.754e-02   8.214 2.53e-16 ***
## tsd         -2.066e-02  3.968e-03  -5.208 1.97e-07 ***
## wineDensity -1.110e+02  2.275e+01  -4.877 1.10e-06 ***
## ph           8.982e-01  1.413e-01   6.359 2.16e-10 ***
## sulphates    9.164e-01  1.210e-01   7.574 4.09e-14 ***
## alcohol      2.104e-01  3.028e-02   6.947 4.06e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.229 on 6985 degrees of freedom
## Multiple R-squared:  0.142, Adjusted R-squared:  0.1405
## F-statistic:  96.3 on 12 and 6985 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(m5)

```



```
anova(m4)
```

```

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq  F value    Pr(>F)
## wineColor                  1    77.0   76.99  50.8770 1.082e-12 ***
## fixedAcidity                1     6.4    6.37   4.2107  0.0402081 *
## va                          1   273.3  273.33 180.6164 < 2.2e-16 ***
## citricAcid                 1     1.3    1.34   0.8884  0.3459496
## residualSugar               1    18.7   18.74  12.3821 0.0004362 ***
## chlr                        1   268.8  268.81 177.6258 < 2.2e-16 ***
## fsd                         1    49.6   49.61  32.7807 1.074e-08 ***
## totalSulfurDioxide          1  145.4  145.44  96.1022 < 2.2e-16 ***

```

```

## wineDensity      1   419.3  419.27 277.0485 < 2.2e-16 ***
## ph              1   262.8  262.78 173.6438 < 2.2e-16 ***
## slpht           1   132.1  132.05  87.2600 < 2.2e-16 ***
## alcohol          1    66.2   66.17  43.7212 4.066e-11 ***
## Residuals       6985 10570.7    1.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m5)

## Analysis of Variance Table
##
## Response: quality
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## wineColor        1    77.0   76.99  50.9925 1.021e-12 ***
## fixedAcidity     1     6.4    6.37   4.2202  0.03998 *
## va               1   273.3  273.33 181.0267 < 2.2e-16 ***
## citricAcid       1     1.3    1.34   0.8904  0.34540
## rs                1     6.2    6.18   4.0929  0.04310 *
## chlr              1   280.1  280.14 185.5313 < 2.2e-16 ***
## fsd               1    43.0   43.02  28.4888 9.720e-08 ***
## tsd               1   160.2  160.15 106.0664 < 2.2e-16 ***
## wineDensity       1   451.4  451.38 298.9414 < 2.2e-16 ***
## ph                1   229.1  229.13 151.7494 < 2.2e-16 ***
## sulphates         1   144.0  143.96  95.3424 < 2.2e-16 ***
## alcohol            1    72.9   72.87  48.2637 4.062e-12 ***
## Residuals         6985 10546.7    1.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Adding In Some Interactions and Using BIC/AIC

```

forwardBIC <- step(lm(quality~1), scope=list(lower = ~1, upper = ~ wineColor + fixedAcidity + va + citr)

## Start:  AIC=3950.77
## quality ~ 1
##
##                 Df  Sum of Sq    RSS     AIC
## + alcohol        1   1067.01 11225 3324.1
## + wineDensity   1    493.41 11798 3672.9
## + chlr           1    373.06 11919 3743.9
## + va              1    324.07 11968 3772.6
## + fsd             1     84.64 12207 3911.3
## + wineColor       1     76.99 12215 3915.7
## + fixedAcidity    1     42.40 12249 3935.4
## + ph              1     27.52 12264 3943.9
## + citricAcid      1     27.50 12264 3944.0
## + sulphates        1     19.81 12272 3948.3
## <none>                  12292 3950.8
## + rs              1      1.63 12290 3958.7
## + tsd              1      1.63 12290 3958.7
##
## Step:  AIC=3324.14
## quality ~ alcohol

```

```

##                                     Df Sum of Sq   RSS   AIC
## + va                           1  302.957 10922 3141.5
## + fsd                          1  207.237 11017 3202.6
## + rs                           1  113.213 11111 3262.1
## + wineColor                     1   62.143 11162 3294.1
## + tsd                          1   49.098 11176 3302.3
## + chlr                         1   46.911 11178 3303.7
## + citricAcid                   1   35.404 11189 3310.9
## + sulphates                     1   17.053 11208 3322.4
## <none>                         11225 3324.1
## + fixedAcidity                 1     9.442 11215 3327.1
## + ph                            1     1.055 11224 3332.3
## + wineDensity                   1     0.869 11224 3332.5
##
## Step:  AIC=3141.52
## quality ~ alcohol + va
##
##                                     Df Sum of Sq   RSS   AIC
## + sulphates                     1    78.299 10843 3100.0
## + fsd                           1    68.532 10853 3106.3
## + wineDensity                   1    66.064 10856 3107.9
## + rs                            1    64.974 10857 3108.6
## + ph                            1    29.419 10892 3131.5
## + wineColor                      1    17.345 10904 3139.2
## <none>                         10922 3141.5
## + chlr                          1     2.179 10920 3149.0
## + fixedAcidity                  1     0.962 10921 3149.8
## + tsd                           1     0.212 10921 3150.2
## + citricAcid                    1     0.014 10922 3150.4
##
## Step:  AIC=3100.02
## quality ~ alcohol + va + sulphates
##
##                                     Df Sum of Sq   RSS   AIC
## + fsd                           1   97.110 10746 3045.9
## + rs                            1   91.949 10751 3049.3
## + wineDensity                   1   30.974 10812 3088.9
## + ph                            1   17.742 10826 3097.4
## <none>                         10843 3100.0
## + tsd                           1     3.051 10840 3106.9
## + chlr                          1     3.037 10840 3106.9
## + fixedAcidity                  1     2.251 10841 3107.4
## + citricAcid                    1     1.935 10841 3107.6
## + wineColor                      1     0.045 10843 3108.8
##
## Step:  AIC=3045.92
## quality ~ alcohol + va + sulphates + fsd
##
##                                     Df Sum of Sq   RSS   AIC
## + tsd                           1   55.627 10691 3018.5
## + rs                            1   42.105 10704 3027.3
## + wineDensity                   1   31.958 10714 3033.9
## + wineColor                      1   22.237 10724 3040.3

```

```

## + ph          1   21.531 10725 3040.7
## <none>           10746 3045.9
## + citricAcid  1    2.195 10744 3053.3
## + fixedAcidity 1    1.266 10745 3053.9
## + chlr         1    0.006 10746 3054.8
##
## Step: AIC=3018.46
## quality ~ alcohol + va + sulphates + fsd + tsd
##
##             Df Sum of Sq   RSS   AIC
## + rs          1   74.607 10616 2978.3
## + wineDensity 1   25.952 10665 3010.3
## + ph          1   13.820 10677 3018.3
## <none>           10691 3018.5
## + chlr         1    3.079 10688 3025.3
## + citricAcid  1    0.320 10690 3027.1
## + fixedAcidity 1    0.041 10691 3027.3
## + wineColor    1    0.007 10691 3027.3
##
## Step: AIC=2978.3
## quality ~ alcohol + va + sulphates + fsd + tsd + rs
##
##             Df Sum of Sq   RSS   AIC
## + ph          1   27.1713 10589 2969.2
## <none>           10616 2978.3
## + wineDensity 1   2.1635 10614 2985.7
## + citricAcid  1   1.9126 10614 2985.9
## + chlr         1   1.7743 10614 2986.0
## + fixedAcidity 1   0.6130 10615 2986.8
## + wineColor    1   0.2015 10616 2987.0
##
## Step: AIC=2969.22
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph
##
##             Df Sum of Sq   RSS   AIC
## <none>           10589 2969.2
## + wineDensity  1   5.3360 10584 2974.6
## + fixedAcidity 1   2.3779 10586 2976.5
## + chlr         1   2.2126 10587 2976.6
## + wineColor    1   0.2013 10589 2977.9
## + citricAcid   1   0.0209 10589 2978.1

forwardAIC <- step(lm(quality~1), scope=list(lower = ~1, upper = ~ wineColor + fixedAcidity + va + citr))

## Start: AIC=3943.92
## quality ~ 1
##
##             Df Sum of Sq   RSS   AIC
## + alcohol      1   1067.01 11225 3310.4
## + wineDensity  1   493.41 11798 3659.2
## + chlr         1   373.06 11919 3730.2
## + va           1   324.07 11968 3758.9
## + fsd          1    84.64 12207 3897.6
## + wineColor    1    76.99 12215 3901.9
## + fixedAcidity 1    42.40 12249 3921.7

```

```

## + ph          1    27.52 12264 3930.2
## + citricAcid 1    27.50 12264 3930.2
## + sulphates   1    19.81 12272 3934.6
## <none>           12292 3943.9
## + rs          1    1.63 12290 3945.0
## + tsd         1    1.63 12290 3945.0
##
## Step: AIC=3310.44
## quality ~ alcohol
##
##             Df Sum of Sq   RSS   AIC
## + va        1  302.957 10922 3121.0
## + fsd       1  207.237 11017 3182.0
## + rs        1  113.213 11111 3241.5
## + wineColor 1   62.143 11162 3273.6
## + tsd        1   49.098 11176 3281.8
## + chlr      1   46.911 11178 3283.1
## + citricAcid 1   35.404 11189 3290.3
## + sulphates 1   17.053 11208 3301.8
## + fixedAcidity 1   9.442 11215 3306.5
## <none>           11225 3310.4
## + ph          1   1.055 11224 3311.8
## + wineDensity 1   0.869 11224 3311.9
##
## Step: AIC=3120.96
## quality ~ alcohol + va
##
##             Df Sum of Sq   RSS   AIC
## + sulphates 1   78.299 10843 3072.6
## + fsd        1   68.532 10853 3078.9
## + wineDensity 1   66.064 10856 3080.5
## + rs         1   64.974 10857 3081.2
## + ph          1   29.419 10892 3104.1
## + wineColor   1   17.345 10904 3111.8
## <none>           10922 3121.0
## + chlr        1   2.179 10920 3121.6
## + fixedAcidity 1   0.962 10921 3122.3
## + tsd         1   0.212 10921 3122.8
## + citricAcid  1   0.014 10922 3122.9
##
## Step: AIC=3072.61
## quality ~ alcohol + va + sulphates
##
##             Df Sum of Sq   RSS   AIC
## + fsd        1   97.110 10746 3011.7
## + rs         1   91.949 10751 3015.0
## + wineDensity 1   30.974 10812 3054.6
## + ph          1   17.742 10826 3063.2
## <none>           10843 3072.6
## + tsd        1   3.051 10840 3072.6
## + chlr        1   3.037 10840 3072.7
## + fixedAcidity 1   2.251 10841 3073.2
## + citricAcid  1   1.935 10841 3073.4
## + wineColor   1   0.045 10843 3074.6

```

```

##
## Step: AIC=3011.66
## quality ~ alcohol + va + sulphates + fsd
##
##          Df Sum of Sq   RSS   AIC
## + tsd      1   55.627 10691 2977.3
## + rs       1   42.105 10704 2986.2
## + wineDensity 1   31.958 10714 2992.8
## + wineColor   1   22.237 10724 2999.2
## + ph        1   21.531 10725 2999.6
## <none>           10746 3011.7
## + citricAcid  1    2.195 10744 3012.2
## + fixedAcidity 1    1.266 10745 3012.8
## + chlr       1    0.006 10746 3013.7
##
## Step: AIC=2977.34
## quality ~ alcohol + va + sulphates + fsd + tsd
##
##          Df Sum of Sq   RSS   AIC
## + rs       1   74.607 10616 2930.3
## + wineDensity 1   25.952 10665 2962.3
## + ph        1   13.820 10677 2970.3
## + chlr     1    3.079 10688 2977.3
## <none>           10691 2977.3
## + citricAcid  1    0.320 10690 2979.1
## + fixedAcidity 1    0.041 10691 2979.3
## + wineColor   1    0.007 10691 2979.3
##
## Step: AIC=2930.33
## quality ~ alcohol + va + sulphates + fsd + tsd + rs
##
##          Df Sum of Sq   RSS   AIC
## + ph       1   27.1713 10589 2914.4
## <none>           10616 2930.3
## + wineDensity 1    2.1635 10614 2930.9
## + citricAcid  1    1.9126 10614 2931.1
## + chlr      1    1.7743 10614 2931.2
## + fixedAcidity 1    0.6130 10615 2931.9
## + wineColor   1    0.2015 10616 2932.2
##
## Step: AIC=2914.39
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph
##
##          Df Sum of Sq   RSS   AIC
## + wineDensity 1    5.3360 10584 2912.9
## <none>           10589 2914.4
## + fixedAcidity 1    2.3779 10586 2914.8
## + chlr       1    2.2126 10587 2914.9
## + wineColor   1    0.2013 10589 2916.3
## + citricAcid  1    0.0209 10589 2916.4
##
## Step: AIC=2912.87
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity
##

```

```

##          Df Sum of Sq   RSS   AIC
## + fixedAcidity  1  31.4642 10552 2894.0
## <none>           10584 2912.9
## + wineColor     1   1.4161 10582 2913.9
## + citricAcid    1   0.9041 10583 2914.3
## + chlr          1   0.4713 10583 2914.6
##
## Step:  AIC=2894.03
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity
##
##          Df Sum of Sq   RSS   AIC
## + wineColor     1   5.1371 10547 2892.6
## <none>           10552 2894.0
## + citricAcid    1   0.1393 10552 2895.9
## + chlr          1   0.0581 10552 2896.0
##
## Step:  AIC=2892.62
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor
##
##          Df Sum of Sq   RSS   AIC
## + wineColor:ph    1   30.9600 10516 2874.1
## + wineColor:fixedAcidity  1   17.9544 10529 2882.7
## + wineColor:tsd    1   13.3326 10534 2885.8
## + wineColor:rs     1   6.2092 10541 2890.5
## + wineColor:fsd    1   5.1264 10542 2891.2
## + wineColor:wineDensity  1   3.5121 10543 2892.3
## <none>             10547 2892.6
## + wineColor:va     1   1.3205 10546 2893.8
## + wineColor:sulphates  1   1.1936 10546 2893.8
## + chlr            1   0.1234 10547 2894.5
## + wineColor:alcohol   1   0.1144 10547 2894.6
## + citricAcid      1   0.0274 10547 2894.6
##
## Step:  AIC=2874.05
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor
##
##          Df Sum of Sq   RSS   AIC
## + wineColor:tsd    1   15.2267 10501 2865.9
## + wineColor:rs     1   7.6499 10508 2871.0
## + wineColor:va     1   3.4579 10512 2873.8
## + wineColor:fsd    1   3.4381 10512 2873.8
## <none>             10516 2874.1
## + wineColor:fixedAcidity  1   1.4726 10514 2875.1
## + citricAcid      1   0.8900 10515 2875.5
## + wineColor:wineDensity  1   0.8730 10515 2875.5
## + wineColor:sulphates  1   0.4869 10515 2875.7
## + wineColor:alcohol   1   0.4723 10516 2875.7
## + chlr            1   0.3972 10516 2875.8
##
## Step:  AIC=2865.91
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +

```

```

##      fixedAcidity + wineColor + ph:wineColor + tsd:wineColor
##
##                                Df Sum of Sq   RSS     AIC
## + wineColor:va             1   5.4743 10495 2864.3
## + wineColor:rs             1   3.5943 10497 2865.5
## + wineColor:wineDensity    1   3.4535 10497 2865.6
## <none>                      10501 2865.9
## + wineColor:fsd            1   1.4178 10499 2867.0
## + chlr                      1   1.0781 10500 2867.2
## + wineColor:sulphates      1   0.9740 10500 2867.3
## + citricAcid                1   0.7775 10500 2867.4
## + wineColor:fixedAcidity    1   0.7673 10500 2867.4
## + wineColor:alcohol          1   0.0488 10501 2867.9
##
## Step:  AIC=2864.26
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor
##
##                                Df Sum of Sq   RSS     AIC
## + wineColor:rs             1   3.8727 10491 2863.7
## <none>                      10495 2864.3
## + wineColor:wineDensity    1   2.6731 10493 2864.5
## + wineColor:fsd            1   2.2074 10493 2864.8
## + wineColor:sulphates      1   1.7778 10493 2865.1
## + wineColor:fixedAcidity    1   1.2497 10494 2865.4
## + chlr                      1   1.1175 10494 2865.5
## + citricAcid                1   0.0957 10495 2866.2
## + wineColor:alcohol          1   0.0240 10495 2866.2
##
## Step:  AIC=2863.68
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor + rs:wineColor
##
##                                Df Sum of Sq   RSS     AIC
## + wineColor:wineDensity    1   7.0171 10484 2861.0
## <none>                      10491 2863.7
## + wineColor:fsd            1   2.1558 10489 2864.2
## + wineColor:fixedAcidity    1   1.8016 10490 2864.5
## + wineColor:sulphates      1   1.6186 10490 2864.6
## + chlr                      1   0.8057 10490 2865.1
## + wineColor:alcohol          1   0.1087 10491 2865.6
## + citricAcid                1   0.0362 10491 2865.7
##
## Step:  AIC=2861
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor + rs:wineColor + wineDensity:wineColor
##
##                                Df Sum of Sq   RSS     AIC
## + wineColor:alcohol          1   8.0598 10476 2857.6
## + wineColor:fsd              1   3.1745 10481 2860.9
## <none>                      10484 2861.0

```

```

## + wineColor:sulphates      1    0.9325 10483 2862.4
## + chlr                      1    0.6323 10484 2862.6
## + wineColor:fixedAcidity   1    0.1159 10484 2862.9
## + citricAcid                1    0.0325 10484 2863.0
##
## Step: AIC=2857.61
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor + rs:wineColor + wineDensity:wineColor + alcohol:wineColor
##
##                         Df Sum of Sq   RSS   AIC
## + wineColor:fixedAcidity   1    7.0341 10469 2854.9
## + wineColor:fsd            1    3.1779 10473 2857.5
## <none>                   10476 2857.6
## + chlr                     1    1.0159 10475 2858.9
## + wineColor:sulphates     1    0.1013 10476 2859.6
## + citricAcid               1    0.0920 10476 2859.6
##
## Step: AIC=2854.91
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor + rs:wineColor + wineDensity:wineColor + alcohol:wineColor +
##           fixedAcidity:wineColor
##
##                         Df Sum of Sq   RSS   AIC
## + wineColor:fsd            1    3.7789 10466 2854.4
## <none>                   10469 2854.9
## + chlr                     1    0.9915 10468 2856.2
## + wineColor:sulphates     1    0.0664 10469 2856.9
## + citricAcid               1    0.0267 10469 2856.9
##
## Step: AIC=2854.39
## quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
##           fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
##           va:wineColor + rs:wineColor + wineDensity:wineColor + alcohol:wineColor +
##           fixedAcidity:wineColor + fsd:wineColor
##
##                         Df Sum of Sq   RSS   AIC
## <none>                   10466 2854.4
## + chlr                     1    1.06059 10464 2855.7
## + wineColor:sulphates     1    0.08463 10465 2856.3
## + citricAcid               1    0.00031 10466 2856.4
backBIC <- step(lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + wineD

## Start: AIC=3018.56
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + wineDensity + ph + sulphates + alcohol + wineDensity +
##           wineColor:wineDensity + wineColor:fixedAcidity + wineColor:va +
##           wineColor:chlr + wineColor:tsd + wineColor:ph + wineColor:sulphates +
##           wineColor:citricAcid + wineColor:rs + wineColor:fsd + wineColor:alcohol
##
##                         Df Sum of Sq   RSS   AIC
## - wineColor:sulphates     1    0.132 10450 3009.8
## - wineColor:va             1    2.312 10452 3011.2

```

```

## - wineColor:citricAcid    1    2.765 10453 3011.6
## - wineColor:fsd           1    3.409 10454 3012.0
## - wineColor:fixedAcidity  1    7.318 10457 3014.6
## - wineColor:chlر          1    9.587 10460 3016.1
## <none>                      10450 3018.6
## - wineColor:tsd           1    14.059 10464 3019.1
## - wineColor:alcohol        1    15.033 10465 3019.8
## - wineColor:rs             1    21.765 10472 3024.3
## - wineColor:wineDensity   1    26.174 10476 3027.2
## - wineColor:ph             1    45.778 10496 3040.3
##
## Step: AIC=3009.79
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##         fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##         wineColor:fixedAcidity + wineColor:va + wineColor:chlر +
##         wineColor:tsd + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##         wineColor:fsd + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:va               1    2.222 10452 3002.4
## - wineColor:citricAcid      1    2.692 10453 3002.7
## - wineColor:fsd             1    3.443 10454 3003.2
## - wineColor:fixedAcidity   1    8.067 10458 3006.3
## - wineColor:chlر           1    9.455 10460 3007.3
## <none>                      10450 3009.8
## - wineColor:tsd             1   13.953 10464 3010.3
## - wineColor:alcohol         1   17.475 10468 3012.6
## - wineColor:rs              1   24.115 10474 3017.1
## - wineColor:wineDensity    1   29.491 10480 3020.7
## - wineColor:ph              1   47.662 10498 3032.8
## - sulphates                 1   73.455 10524 3050.0
##
## Step: AIC=3002.43
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##         fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##         wineColor:fixedAcidity + wineColor:chlر + wineColor:tsd +
##         wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:fsd +
##         wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:fsd             1    2.648 10455 2995.3
## - wineColor:citricAcid     1    7.348 10460 2998.5
## - wineColor:fixedAcidity   1    7.846 10460 2998.8
## - wineColor:chlر           1    8.116 10460 2999.0
## - wineColor:tsd             1   12.260 10465 3001.8
## <none>                      10452 3002.4
## - wineColor:alcohol         1   18.185 10471 3005.7
## - wineColor:rs              1   23.957 10476 3009.6
## - wineColor:wineDensity    1   30.665 10483 3014.1
## - wineColor:ph              1   48.044 10500 3025.7
## - sulphates                 1   72.116 10524 3041.7
## - va                        1   165.397 10618 3103.4
##
## Step: AIC=2995.35

```

```

## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##      fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##      wineColor:fixedAcidity + wineColor:chlr + wineColor:tsd +
##      wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:fixedAcidity    1     7.076 10462 2991.2
## - wineColor:citricAcid     1     7.845 10463 2991.7
## - wineColor:chlr            1     7.881 10463 2991.8
## - wineColor:tsd             1    10.767 10466 2993.7
## <none>                      10455 2995.3
## - wineColor:alcohol         1    17.798 10473 2998.4
## - wineColor:rs              1    23.006 10478 3001.9
## - wineColor:wineDensity     1    29.019 10484 3005.9
## - wineColor:ph              1    45.919 10501 3017.2
## - sulphates                 1    72.809 10528 3035.1
## - fsd                        1    91.991 10547 3047.8
## - va                         1   163.910 10619 3095.3
##
## Step:  AIC=2991.23
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##      fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##      wineColor:chlr + wineColor:tsd + wineColor:ph + wineColor:citricAcid +
##      wineColor:rs + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:chlr            1     5.360 10468 2986.0
## - wineColor:tsd             1     8.364 10470 2988.0
## - fixedAcidity              1     9.669 10472 2988.8
## - wineColor:alcohol          1    10.875 10473 2989.6
## - wineColor:citricAcid       1    12.729 10475 2990.9
## <none>                      10462 2991.2
## - wineColor:rs              1    16.122 10478 2993.2
## - wineColor:wineDensity      1    26.659 10489 3000.2
## - wineColor:ph              1    45.561 10508 3012.8
## - sulphates                 1    75.939 10538 3033.0
## - fsd                        1    89.775 10552 3042.2
## - va                         1   167.076 10629 3093.2
##
## Step:  AIC=2985.96
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##      fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##      wineColor:tsd + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##      wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - chlr                       1     0.661 10468 2977.6
## - wineColor:tsd               1     8.331 10476 2982.7
## - fixedAcidity                1    10.616 10478 2984.2
## <none>                      10468 2986.0
## - wineColor:citricAcid        1    13.401 10481 2986.1
## - wineColor:alcohol            1    13.712 10481 2986.3
## - wineColor:rs                 1    16.250 10484 2988.0
## - wineColor:wineDensity        1    23.693 10491 2992.9

```

```

## - wineColor:ph      1   43.262 10511 3006.0
## - sulphates        1   72.519 10540 3025.4
## - fsd               1   88.572 10556 3036.1
## - va                1   171.439 10639 3090.8
##
## Step: AIC=2977.55
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + fsd +
##          tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##          wineColor:tsd + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##          wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:tsd           1    7.947 10476 2974.0
## - fixedAcidity            1   11.586 10480 2976.4
## <none>                      10468 2977.6
## - wineColor:alcohol       1   13.422 10482 2977.7
## - wineColor:citricAcid    1   13.564 10482 2977.8
## - wineColor:rs             1   16.653 10485 2979.8
## - wineColor:wineDensity   1   23.625 10492 2984.5
## - wineColor:ph             1   42.903 10511 2997.3
## - sulphates                1   72.084 10540 3016.7
## - fsd                      1   88.506 10557 3027.6
## - va                       1   174.188 10642 3084.2
##
## Step: AIC=2974
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + fsd +
##          tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##          wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## <none>                      10476 2974.0
## - fixedAcidity            1   13.275 10489 2974.0
## - wineColor:citricAcid    1   15.166 10491 2975.3
## - wineColor:alcohol       1   19.210 10495 2978.0
## - wineColor:rs             1   23.753 10500 2981.0
## - wineColor:wineDensity   1   25.394 10502 2982.1
## - tsd                      1   27.823 10504 2983.7
## - wineColor:ph             1   44.467 10521 2994.8
## - sulphates                1   71.758 10548 3012.9
## - fsd                      1   87.089 10563 3023.1
## - va                       1   172.231 10648 3079.3
backAIC <- step(lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + wineD
## Start: AIC=2854.08
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##          fsd + tsd + wineDensity + ph + sulphates + alcohol + wineDensity +
##          wineColor:wineDensity + wineColor:fixedAcidity + wineColor:va +
##          wineColor:chlr + wineColor:tsd + wineColor:ph + wineColor:sulphates +
##          wineColor:citricAcid + wineColor:rs + wineColor:fsd + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:sulphates     1    0.132 10450 2852.2
## - wineColor:va             1    2.312 10452 2853.6
## - wineColor:citricAcid    1    2.765 10453 2853.9

```

```

## <none>                               10450 2854.1
## - wineColor:fsd                     1      3.409 10454 2854.4
## - wineColor:fixedAcidity            1      7.318 10457 2857.0
## - wineColor:chl                         1      9.587 10460 2858.5
## - wineColor:tsd                     1     14.059 10464 2861.5
## - wineColor:alcohol                  1     15.033 10465 2862.1
## - wineColor:rs                      1     21.765 10472 2866.6
## - wineColor:wineDensity              1     26.174 10476 2869.6
## - wineColor:ph                      1     45.778 10496 2882.7
##
## Step: AIC=2852.16
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##          fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##          wineColor:fixedAcidity + wineColor:va + wineColor:chl +
##          wineColor:tsd + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##          wineColor:fsd + wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:va                   1      2.222 10452 2851.7
## - wineColor:citricAcid           1      2.692 10453 2852.0
## <none>                           10450 2852.2
## - wineColor:fsd                 1      3.443 10454 2852.5
## - wineColor:fixedAcidity         1      8.067 10458 2855.6
## - wineColor:chl                 1      9.455 10460 2856.5
## - wineColor:tsd                 1     13.953 10464 2859.5
## - wineColor:alcohol              1     17.475 10468 2861.9
## - wineColor:rs                  1     24.115 10474 2866.3
## - wineColor:wineDensity          1     29.491 10480 2869.9
## - wineColor:ph                  1     47.662 10498 2882.0
## - sulphates                      1    73.455 10524 2899.2
##
## Step: AIC=2851.65
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##          fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##          wineColor:fixedAcidity + wineColor:chl + wineColor:tsd +
##          wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:fsd +
##          wineColor:alcohol
##
##                                     Df Sum of Sq   RSS   AIC
## - wineColor:fsd                 1      2.648 10455 2851.4
## <none>                           10452 2851.7
## - wineColor:citricAcid           1      7.348 10460 2854.6
## - wineColor:fixedAcidity          1      7.846 10460 2854.9
## - wineColor:chl                 1      8.116 10460 2855.1
## - wineColor:tsd                 1     12.260 10465 2857.9
## - wineColor:alcohol              1     18.185 10471 2861.8
## - wineColor:rs                  1     23.957 10476 2865.7
## - wineColor:wineDensity          1     30.665 10483 2870.2
## - wineColor:ph                  1     48.044 10500 2881.7
## - sulphates                      1     72.116 10524 2897.8
## - va                             1    165.397 10618 2959.5
##
## Step: AIC=2851.42
## quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +

```

```

##   fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
##   wineColor:fixedAcidity + wineColor:chlر + wineColor:tsd +
##   wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol
##
##                                Df Sum of Sq    RSS     AIC
## <none>                            10455 2851.4
## - wineColor:fixedAcidity    1      7.076 10462 2854.2
## - wineColor:citricAcid      1      7.845 10463 2854.7
## - wineColor:chlر           1      7.881 10463 2854.7
## - wineColor:tsd             1     10.767 10466 2856.6
## - wineColor:alcohol          1     17.798 10473 2861.3
## - wineColor:rs               1     23.006 10478 2864.8
## - wineColor:wineDensity       1     29.019 10484 2868.8
## - wineColor:ph               1     45.919 10501 2880.1
## - sulphates                  1     72.809 10528 2898.0
## - fsd                         1     91.991 10547 2910.7
## - va                          1    163.910 10619 2958.3

```

Testing Summary, VIF, and ANOVA

```

forwardBICmodel <- lm(quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph)

forwardAICmodel <- lm(quality ~ alcohol + va + sulphates + fsd + tsd + rs + ph + wineDensity +
                      fixedAcidity + wineColor + ph:wineColor + tsd:wineColor +
                      va:wineColor + rs:wineColor + wineDensity:wineColor + alcohol:wineColor +
                      fixedAcidity:wineColor + fsd:wineColor)

backBICmodel <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + fsd +
                     tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
                     wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol)

backAICmodel <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
                     fsd + tsd + wineDensity + ph + sulphates + alcohol + wineColor:wineDensity +
                     wineColor:fixedAcidity + wineColor:chlر + wineColor:tsd +
                     wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol)

options(scipen = 999)
vif(forwardBICmodel) #no violation

##   alcohol         va sulphates        fsd        tsd        rs        ph
## 1.174185  1.303292  1.152847  2.538543  3.000252  1.501067  1.154939

vif(forwardAICmodel) #violations abound

##          alcohol            va        sulphates
## 15.590233 7.993363 1.562280
##          fsd            tsd            rs
## 13.122080 21.037679 49.884857
##          ph        wineDensity        fixedAcidity
## 13.373374 64.474750 15.295422
##          wineColor    ph:wineColor    tsd:wineColor
## 1881723.725275 1219.844692 62.206309
##          va:wineColor    rs:wineColor wineDensity:wineColor
## 22.213367 192.818702 1930556.330435

```

```

##      alcohol:wineColor fixedAcidity:wineColor          fsd:wineColor
##      370.942439           124.989436            85.256936

vif(backBICmodel) #violations abound

##      wineColor      fixedAcidity          va
##  849744.616331       5.395425   2.124842
##      citricAcid          rs          fsd
##      4.540807       41.536225  2.777616
##      tsd      wineDensity          ph
##      5.292205       36.687342  9.130371
##      sulphates      alcohol wineColor:wineDensity
##      1.558738       11.945058 841674.869144
##      wineColor:ph wineColor:citricAcid wineColor:rs
##      685.473720        9.726427 142.246755
##      wineColor:alcohol
##      243.116218

vif(backAICmodel) #violations abound

##      wineColor      fixedAcidity          va
##  1976221.119193       16.980758  2.151678
##      citricAcid          rs          chlr
##      4.908854       50.158021 15.759406
##      fsd          tsd      wineDensity
##      2.788612       13.091493 67.882660
##      ph      sulphates      alcohol
##      14.028061       1.585357 15.859562
##      wineColor:wineDensity wineColor:fixedAcidity wineColor:chlr
##      2026807.941783       135.401410 109.760017
##      wineColor:tsd      wineColor:ph wineColor:citricAcid
##      33.338567       1271.875006 10.433227
##      wineColor:rs      wineColor:alcohol
##      194.636741        378.090119

summary(forwardBICmodel)

## 
## Call:
## lm(formula = quality ~ alcohol + va + sulphates + fsd + tsd +
##     rs + ph)
## 
## Residuals:
##    Min      1Q  Median      3Q      Max
## -3.9551 -0.8414 -0.0192  0.8362  4.5393
## 
## Coefficients:
##             Estimate Std. Error t value     Pr(>|t|)    
## (Intercept) -1.052684   0.375325 -2.805     0.00505 **  
## alcohol      0.344827   0.013399 25.735 < 0.0000000000000002 *** 
## va          -0.513293   0.037603 -13.650 < 0.0000000000000002 *** 
## sulphates    0.749865   0.105655   7.097  0.0000000000013973 *** 
## fsd          0.323886   0.036482   8.878 < 0.0000000000000002 *** 
## tsd         -0.022226   0.003068  -7.244  0.0000000000004834 *** 
## rs           0.288473   0.037858   7.620  0.0000000000000287 *** 
## ph          0.420406   0.099266   4.235  0.0000231314367622 *** 

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.231 on 6990 degrees of freedom
## Multiple R-squared: 0.1385, Adjusted R-squared: 0.1377
## F-statistic: 160.6 on 7 and 6990 DF, p-value: < 0.0000000000000022
summary(forwardAICmodel)

##
## Call:
## lm(formula = quality ~ alcohol + va + sulphates + fsd + tsd +
##     rs + ph + wineDensity + fixedAcidity + wineColor + ph:wineColor +
##     tsd:wineColor + va:wineColor + rs:wineColor + wineDensity:wineColor +
##     alcohol:wineColor + fixedAcidity:wineColor + fsd:wineColor)
##
## Residuals:
##      Min    1Q   Median    3Q   Max 
## -3.8341 -0.8543 -0.0214  0.8132  4.5691
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -13.226027  38.945734 -0.340  0.734167    
## alcohol       0.353251   0.048577  7.272 0.00000000000393
## va           -0.342726   0.092653 -3.699  0.000218
## sulphates    0.814863   0.122371  6.659 0.0000000000029679
## fsd          0.416746   0.082524  5.050 0.000000453081112
## tsd          -0.040877   0.008084 -5.057 0.000000437885919
## rs            -0.047497   0.217137 -0.219  0.826859
## ph            -0.659624   0.336075 -1.963  0.049718
## wineDensity  16.366516  39.783621  0.411  0.680800
## fixedAcidity -0.015092   0.044090 -0.342  0.732136
## wineColorW   172.649003  46.476108  3.715  0.000205
## ph:wineColorW 1.857151   0.369561  5.025 0.000000515165060
## tsd:wineColorW 0.029972   0.009338  3.210  0.001335
## va:wineColorW -0.195715   0.106269 -1.842  0.065562
## rs:wineColorW  0.909015   0.236122  3.850  0.000119
## wineDensity:wineColorW -180.244968  47.358248 -3.806  0.000142
## alcohol:wineColorW   -0.194341   0.060412 -3.217  0.001302
## fixedAcidity:wineColorW  0.121372   0.053789  2.256  0.024074
## fsd:wineColorW   -0.147462   0.092893 -1.587  0.112458
##
## (Intercept)
## alcohol      ***
## va           ***
## sulphates   ***
## fsd          ***
## tsd          ***
## rs            *
## ph            ***
## wineDensity  ***
## fixedAcidity ***
## wineColorW   ***
## ph:wineColorW ***
## tsd:wineColorW **

```

```

## va:wineColorW .
## rs:wineColorW ***
## wineDensity:wineColorW ***
## alcohol:wineColorW **
## fixedAcidity:wineColorW *
## fsd:wineColorW
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.225 on 6979 degrees of freedom
## Multiple R-squared: 0.1486, Adjusted R-squared: 0.1464
## F-statistic: 67.65 on 18 and 6979 DF, p-value: < 0.0000000000000022
summary(backBICmodel)

##
## Call:
## lm(formula = quality ~ wineColor + fixedAcidity + va + citricAcid +
##      rs + fsd + tsd + wineDensity + ph + sulphates + alcohol +
##      wineColor:wineDensity + wineColor:ph + wineColor:citricAcid +
##      wineColor:rs + wineColor:alcohol)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -3.8013 -0.8488 -0.0173  0.8198  4.5926 
##
## Coefficients:
##             Estimate Std. Error t value
## (Intercept) 20.314474 29.664028  0.685
## wineColorW 123.378848 31.243119  3.949
## fixedAcidity 0.077912  0.026196  2.974
## va          -0.511954  0.047788 -10.713
## citricAcid  -0.655600  0.216487 -3.028
## rs           0.012214  0.198208  0.062
## fsd          0.289345  0.037982  7.618
## tsd          -0.017465  0.004056 -4.306
## wineDensity -18.987209 30.021055 -0.632
## ph           -0.362008  0.277791 -1.303
## sulphates    0.845554  0.122277  6.915
## alcohol       0.345189  0.042536  8.115
## wineColorW:wineDensity -128.680525 31.281309 -4.114
## wineColorW:ph      1.508568  0.277132  5.443
## wineColorW:citricAcid 0.824761  0.259436  3.179
## wineColorW:rs       0.807157  0.202881  3.978
## wineColorW:alcohol  -0.175048  0.048925 -3.578
##                  Pr(>|t|) 
## (Intercept) 0.493481
## wineColorW 0.000079252734245164 ***
## fixedAcidity 0.002947 ** 
## va         < 0.000000000000002 ***
## citricAcid 0.002468 ** 
## rs           0.950867
## fsd          0.00000000000029168 ***
## tsd          0.000016861269129638 ***
## wineDensity 0.527105

```

```

## ph                      0.192562
## sulphates              0.0000000000005091181 ***
## alcohol                 0.000000000000000568 ***
## wineColorW:wineDensity 0.000039391761742950 ***
## wineColorW:ph            0.000000054023020416 ***
## wineColorW:citricAcid   0.001484 **
## wineColorW:rs            0.000070059310268688 ***
## wineColorW:alcohol       0.000349 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.225 on 6981 degrees of freedom
## Multiple R-squared:  0.1477, Adjusted R-squared:  0.1457
## F-statistic: 75.61 on 16 and 6981 DF,  p-value: < 0.0000000000000022
summary(backAICmodel)

##
## Call:
## lm(formula = quality ~ wineColor + fixedAcidity + va + citricAcid +
##      rs + chlr + fsd + tsd + wineDensity + ph + sulphates + alcohol +
##      wineColor:wineDensity + wineColor:fixedAcidity + wineColor:chlr +
##      wineColor:tsd + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##      wineColor:alcohol)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -3.8242 -0.8431 -0.0145  0.8211  4.5898 
##
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)             -42.431053  39.960625 -1.062
## wineColorW                205.485713  47.611925  4.316
## fixedAcidity             -0.015278  0.046440 -0.329
## va                       -0.502579  0.048054 -10.459
## citricAcid               -0.466670  0.224928 -2.075
## rs                        -0.059810  0.217654 -0.275
## chlr                      0.150731  0.063475  2.375
## fsd                       0.297965  0.038030  7.835
## tsd                      -0.031823  0.006375 -4.992
## wineDensity               45.886027  40.807036  1.124
## ph                        -0.867468  0.344081 -2.521
## sulphates                 0.858964  0.123228  6.971
## alcohol                   0.363920  0.048977  7.430
## wineColorW:wineDensity   -213.459821  48.507271 -4.401
## wineColorW:fixedAcidity   0.121616  0.055965  2.173
## wineColorW:chlr           -0.158523  0.069126 -2.293
## wineColorW:tsd             0.018318  0.006834  2.681
## wineColorW:ph              2.088186  0.377226  5.536
## wineColorW:citricAcid     0.614349  0.268503  2.288
## wineColorW:rs              0.929211  0.237148  3.918
## wineColorW:alcohol         -0.210123  0.060970 -3.446
##                               Pr(>|t|) 
## (Intercept)                  0.288353
## wineColorW 0.00001611892517029 ***

```

```

## fixedAcidity          0.742170
## va                   < 0.0000000000000002 ***
## citricAcid           0.038046 *
## rs                    0.783483
## chlr                 0.017592 *
## fsd                  0.0000000000000538 ***
## tsd                  0.00000061201242871 ***
## wineDensity           0.260855
## ph                   0.011721 *
## sulphates            0.0000000000344713 ***
## alcohol               0.0000000000012115 ***
## wineColorW:wineDensity 0.00001095718586840 ***
## wineColorW:fixedAcidity 0.029808 *
## wineColorW:chlr      0.021864 *
## wineColorW:tsd       0.007367 **
## wineColorW:ph        0.00000003213809872 ***
## wineColorW:citricAcid 0.022165 *
## wineColorW:rs         0.00009003853527512 ***
## wineColorW:alcohol    0.000572 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.224 on 6977 degrees of freedom
## Multiple R-squared:  0.1494, Adjusted R-squared:  0.147
## F-statistic: 61.28 on 20 and 6977 DF,  p-value: < 0.0000000000000022
anova(forwardBICmodel)

## Analysis of Variance Table
##
## Response: quality
##             Df  Sum Sq Mean Sq F value          Pr(>F)
## alcohol      1 1067.0 1067.01 704.368 < 0.0000000000000022 ***
## va          1  303.0  302.96 199.991 < 0.0000000000000022 ***
## sulphates   1   78.3   78.30  51.687 0.00000000000718367 ***
## fsd          1   97.1   97.11  64.105 0.0000000000001371 ***
## tsd          1   55.6   55.63  36.721 0.000000001433685011 ***
## rs           1   74.6   74.61  49.250 0.00000000002464892 ***
## ph           1   27.2   27.17  17.937 0.000023131436762205 ***
## Residuals 6990 10588.8     1.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(forwardAICmodel)

## Analysis of Variance Table
##
## Response: quality
##             Df  Sum Sq Mean Sq F value          Pr(>F)
## alcohol      1 1067.0 1067.01 711.5491 < 0.0000000000000022
## va          1  303.0  302.96 202.0298 < 0.0000000000000022
## sulphates   1   78.3   78.30  52.2142 0.0000000000005504553
## fsd          1   97.1   97.11  64.7588 0.0000000000009873
## tsd          1   55.6   55.63  37.0956 0.0000000011844995871
## rs           1   74.6   74.61  49.7523 0.00000000000019120001
## ph           1   27.2   27.17  18.1195 0.0000210185442058805

```

```

## wineDensity           1     5.3    5.34   3.5584      0.059288
## fixedAcidity         1    31.5   31.46  20.9822  0.0000047161103012807
## wineColor            1     5.1    5.14   3.4257      0.064229
## ph:wineColor          1    31.0   30.96  20.6460  0.0000056182609895881
## tsd:wineColor          1    15.2   15.23  10.1541      0.001446
## va:wineColor          1     5.5    5.47   3.6506      0.056090
## rs:wineColor          1     3.9    3.87   2.5826      0.108091
## wineDensity:wineColor  1     7.0    7.02   4.6794      0.030560
## alcohol:wineColor     1     8.1    8.06   5.3748      0.020459
## fixedAcidity:wineColor 1     7.0    7.03   4.6908      0.030359
## fsd:wineColor          1     3.8    3.78   2.5200      0.112458
## Residuals             6979 10465.5   1.50

##
## alcohol                 ***
## va                      ***
## sulphates               ***
## fsd                     ***
## tsd                     ***
## rs                      ***
## ph                      ***
## wineDensity              .
## fixedAcidity             ***
## wineColor                .
## ph:wineColor              ***
## tsd:wineColor             **
## va:wineColor              .
## rs:wineColor              .
## wineDensity:wineColor    *
## alcohol:wineColor         *
## fixedAcidity:wineColor   *
## fsd:wineColor
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(backBICmodel)

```

```

## Analysis of Variance Table
##
## Response: quality
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## wineColor                  1   77.0   76.99  51.3070 0.00000000000008708
## fixedAcidity                1    6.4    6.37   4.2462  0.0393743
## va                          1  273.3  273.33 182.1430 < 0.00000000000000022
## citricAcid                 1    1.3    1.34   0.8959  0.3439195
## rs                           1    6.2    6.18   4.1181  0.0424642
## fsd                         1   25.8   25.75  17.1620 0.0000347293852099
## tsd                         1  264.3  264.27 176.1029 < 0.00000000000000022
## wineDensity                 1  634.2  634.21 422.6207 < 0.00000000000000022
## ph                           1  239.3  239.28 159.4478 < 0.00000000000000022
## sulphates                   1  143.4  143.43  95.5805 < 0.00000000000000022
## alcohol                      1   73.6   73.58  49.0316 0.0000000000027536
## wineColor:wineDensity        1    3.5    3.53   2.3511  0.1252385
## wineColor:ph                 1   29.2   29.17  19.4356 0.0000105607860207
## wineColor:citricAcid         1    7.4    7.36   4.9067  0.0267851

```

```

## wineColor:rs           1    11.5   11.49   7.6584      0.0056657
## wineColor:alcohol      1    19.2   19.21  12.8011      0.0003488
## Residuals            6981 10476.1    1.50

##
## wineColor          ***
## fixedAcidity        *
## va                  ***
## citricAcid
## rs                 *
## fsd                ***
## tsd                ***
## wineDensity        ***
## ph                 ***
## sulphates          ***
## alcohol             ***
## wineColor:wineDensity
## wineColor:ph         ***
## wineColor:citricAcid *
## wineColor:rs          **
## wineColor:alcohol      ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(backAICmodel)

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq F value      Pr(>F)
## wineColor                   1    77.0   76.99  51.3808 0.00000000000008389
## fixedAcidity                 1     6.4    6.37   4.2523 0.0392330
## va                          1   273.3   273.33 182.4051 < 0.00000000000000022
## citricAcid                  1     1.3    1.34   0.8972 0.3435727
## rs                           1     6.2    6.18   4.1240 0.0423159
## chlr                         1   280.1   280.14 186.9441 < 0.00000000000000022
## fsd                          1    43.0   43.02  28.7058 0.0000000869455742
## tsd                          1   160.2   160.15 106.8741 < 0.00000000000000022
## wineDensity                  1   451.4   451.38 301.2177 < 0.00000000000000022
## ph                           1   229.1   229.13 152.9049 < 0.00000000000000022
## sulphates                    1   144.0   143.96  96.0684 < 0.00000000000000022
## alcohol                       1    72.9    72.87  48.6312 0.0000000000033726
## wineColor:wineDensity        1     3.5    3.51   2.3404 0.1261025
## wineColor:fixedAcidity       1    15.5   15.51  10.3523 0.0012991
## wineColor:chlr                1     2.6    2.59   1.7310 0.1883305
## wineColor:tsd                 1    12.3   12.33   8.2306 0.0041315
## wineColor:ph                  1    24.8   24.77  16.5330 0.0000483369236362
## wineColor:citricAcid         1     6.8    6.77   4.5206 0.0335237
## wineColor:rs                  1     8.4    8.40   5.6024 0.0179629
## wineColor:alcohol              1    17.8   17.80  11.8774 0.0005715
## Residuals                     6977 10455.1    1.50

##
## wineColor          ***
## fixedAcidity        *
## va                  ***

```

```

## citricAcid
## rs *
## chlr ***
## fsd ***
## tsd ***
## wineDensity ***
## ph ***
## sulphates ***
## alcohol ***
## wineColor:wineDensity
## wineColor:fixedAcidity **
## wineColor:chlr
## wineColor:tsd **
## wineColor:ph ***
## wineColor:citricAcid *
## wineColor:rs *
## wineColor:alcohol ***
## Residuals
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- forwardBICmodel has no violations but it seems much too simplistic so far. But all the other models have violations so far. Thus I am going to start with the model that has the largest amount of predictors and slowly reduce the number of predictors (refering to other models when deciding which ones to drop) so that VIF is not violated and SSE is the most diminished.

Look at some correlations with quality first

```

cor(quality, fixedAcidity)

## [1] -0.05873244
cor(quality, va)

## [1] -0.1623726
cor(quality, citricAcid)

## [1] 0.04729954
cor(quality, rs)

## [1] -0.01152722
cor(quality, chlr)

## [1] 0.1742143
cor(quality, fsd)

## [1] 0.08298183
cor(quality, tsd)

## [1] -0.01152238
cor(quality, ph )

```

```

## [1] 0.0473173
cor(quality, sulphates)

## [1] 0.04014197
cor(quality, alcohol)

## [1] 0.2946325
cor(quality, wineDensity)

## [1] -0.2003556
#alcohol and winedensity has highest correlations with quality (alcohol is the highest)

cor(alcohol, wineDensity)

## [1] -0.7003862
#however, they are also highly correlated with one another, I suspect that one will need to be dropped

```

Using backAICmodel and dropping predictors using other models as reference

```

BeginningModel <- backAICmodel
vif(BeginningModel)

##          wineColor      fixedAcidity           va
## 1976221.119193       16.980758      2.151678
##         citricAcid            rs          chlr
##        4.908854       50.158021     15.759406
##          fsd            tsd      wineDensity
##        2.788612       13.091493     67.882660
##          ph            sulphates      alcohol
##        14.028061       1.585357     15.859562
## wineColor:wineDensity wineColor:fixedAcidity      wineColor:chlr
##        2026807.941783      135.401410    109.760017
## wineColor:tsd          wineColor:ph      wineColor:citricAcid
##        33.338567       1271.875006     10.433227
## wineColor:rs          wineColor:alcohol
##        194.636741       378.090119

# we can start easy and first check in the model if any of the numeric predictors are too highly correlated

vif(lm(quality~fixedAcidity+fixedAcidity:wineColor))

##          fixedAcidity fixedAcidity:wineColor
##             1.138931          1.138931

vif(lm(quality~fsd + fsd:wineColor))

##          fsd fsd:wineColor
##        2.317025      2.317025

vif(lm(quality~tsd + tsd:wineColor)) #violation

##          tsd tsd:wineColor

```

```

##      5.736625      5.736625
vif(lm(quality~sulphates + sulphates:wineColor))

##          sulphates sulphates:wineColor
##          1.02998           1.02998
vif(lm(quality~citricAcid + citricAcid:wineColor))

##          citricAcid citricAcid:wineColor
##          1.49181           1.49181
vif(lm(quality~rs + rs:wineColor))

##          rs rs:wineColor
##      2.368335      2.368335
vif(lm(quality~ph + ph:wineColor))

##          ph ph:wineColor
##      1.07374       1.07374
vif(lm(quality~alcohol + alcohol:wineColor))

##          alcohol alcohol:wineColor
##      1.056701       1.056701
vif(lm(quality~chlr + chlr:wineColor))

##          chlr chlr:wineColor
##      3.085519       3.085519
#remove totalSulfurDioxide:wineColor interaction term
newmodel1 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
  fsd + tsd + wineDensity + ph + sulphates +
  alcohol + wineColor:wineDensity + wineColor:fixedAcidity +
  wineColor:chlr + wineColor:ph +
  wineColor:citricAcid + wineColor:rs + wineColor:alcohol)

vif(newmodel1)

##          wineColor      fixedAcidity          va
##      1963785.845729      16.403747      2.151317
##          citricAcid            rs          chlr
##          4.845711      49.292852      15.677427
##          fsd            tsd      wineDensity
##          2.785047      5.401496      67.616197
##          ph            sulphates        alcohol
##          13.876222      1.585356      15.704646
##  wineColor:wineDensity wineColor:fixedAcidity      wineColor:chlr
##          2012464.766568      131.917287      109.625504
##          wineColor:ph    wineColor:citricAcid      wineColor:rs
##          1263.237419      10.306745      191.335934
##          wineColor:alcohol
##          375.495728

anova(newmodel1) #SSE: 10465.8

## Analysis of Variance Table
##

```

```

## Response: quality
##                               Df  Sum Sq Mean Sq F value      Pr(>F)
## wineColor                  1    77.0   76.99  51.3353 0.0000000000008584
## fixedAcidity                1     6.4    6.37   4.2486 0.0393200
## va                          1   273.3  273.33 182.2436 < 0.0000000000000022
## citricAcid                 1     1.3    1.34   0.8964 0.3437864
## rs                           1     6.2    6.18   4.1204 0.0424072
## chlr                         1   280.1  280.14 186.7785 < 0.0000000000000022
## fsd                          1    43.0   43.02  28.6804 0.0000000880893795
## tsd                          1   160.2  160.15 106.7794 < 0.0000000000000022
## wineDensity                  1   451.4  451.38 300.9509 < 0.0000000000000022
## ph                           1   229.1  229.13 152.7695 < 0.0000000000000022
## sulphates                    1   144.0  143.96  95.9833 < 0.0000000000000022
## alcohol                       1    72.9   72.87  48.5882 0.000000000034469
## wineColor:wineDensity        1     3.5    3.51   2.3383 0.1262704
## wineColor:fixedAcidity       1    15.5   15.51   10.3431 0.0013055
## wineColor:chlr                1     2.6    2.59   1.7294 0.1885263
## wineColor:ph                  1    18.8   18.80   12.5361 0.0004018
## wineColor:citricAcid         1     9.4    9.42   6.2811 0.0122258
## wineColor:rs                  1    10.8   10.78   7.1880 0.0073564
## wineColor:alcohol              1    20.3   20.31  13.5383 0.0002355
## Residuals                     6978 10465.8   1.50
##
## wineColor                   ***
## fixedAcidity                  *
## va                          ***
## citricAcid
## rs                           *
## chlr                         ***
## fsd                          ***
## tsd                          ***
## wineDensity                  ***
## ph                           ***
## sulphates                    ***
## alcohol                      ***
## wineColor:wineDensity
## wineColor:fixedAcidity  **
## wineColor:chlr
## wineColor:ph                  ***
## wineColor:citricAcid        *
## wineColor:rs                  **
## wineColor:alcohol             ***
## Residuals
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#before looking at interaction terms, see if there are any violations just among numerical predictors
vif(lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + wineDensity + ph

```

```

##   wineColor fixedAcidity          va   citricAcid      rs
## 7.896544    4.738833    2.016203    1.562970 7.991085
##   chlr      fsd          tsd wineDensity      ph
## 2.448134    2.696151    5.034294   20.943674 2.346481
## sulphates   alcohol
## 1.516848    6.017532

```

```

#wineDensity has highest VIF so try removing that
vif(lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates))

##      wineColor fixedAcidity          va    citricAcid          rs
## 6.281208     2.173436    2.012546    1.558674    1.508763
##      chlr        fsd        tsd         ph    sulphates
## 2.365970     2.655217    4.973080    1.586751    1.419091
##      alcohol
## 1.641370

# add in the interactions in newmodel1 but without wineDensity and wineDensity:wineColor
newmodel2 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + wineColor:fixedAcidity + wineColor:va + wineColor:citricAcid + wineColor:rs + wineColor:chlr + wineColor:ph + wineColor:alcohol + wineColor:wineColor:fixedAcidity + wineColor:wineColor:va + wineColor:wineColor:citricAcid + wineColor:wineColor:rs + wineColor:wineColor:chlr + wineColor:wineColor:ph + wineColor:wineColor:alcohol + wineColor:wineColor:wineColor:fixedAcidity + wineColor:wineColor:wineColor:va + wineColor:wineColor:wineColor:citricAcid + wineColor:wineColor:wineColor:rs + wineColor:wineColor:wineColor:chlr + wineColor:wineColor:wineColor:ph + wineColor:wineColor:wineColor:alcohol)

vif(newmodel2)

##      wineColor      fixedAcidity          va
## 1220.637986     5.561935    2.147128
##      citricAcid          rs        chlr
## 4.835013     29.276015   14.572420
##      fsd        tsd         ph
## 2.717568     5.277454    9.198201
##      sulphates      alcohol wineColor:fixedAcidity
## 1.473413     6.629972   55.813868
##      wineColor:chlr      wineColor:ph wineColor:citricAcid
## 102.249019    852.246043   10.206791
##      wineColor:rs      wineColor:alcohol
## 99.054284    132.823777

anova(newmodel2) #SSE: 10519.2

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq F value    Pr(>F)
## wineColor                      1    77.0   76.99  51.0895 0.0000000000000972
## fixedAcidity                   1     6.4    6.37   4.2282 0.0397937
## va                            1   273.3  273.33 181.3710 < 0.00000000000000022
## citricAcid                    1     1.3    1.34   0.8921 0.3449442
## rs                            1     6.2    6.18   4.1007 0.0429045
## chlr                          1   280.1  280.14 185.8842 < 0.00000000000000022
## fsd                           1    43.0   43.02  28.5430 0.000000094535043
## tsd                           1   160.2  160.15 106.2682 < 0.00000000000000022
## ph                            1    66.1   66.11  43.8698 0.000000000037707
## sulphates                     1   112.5  112.55  74.6802 < 0.00000000000000022
## alcohol                        1   682.8  682.77 453.0503 < 0.00000000000000022
## wineColor:fixedAcidity        1    20.4   20.45  13.5672 0.0002319
## wineColor:chlr                 1     2.5    2.50   1.6622 0.1973523
## wineColor:ph                   1    26.7   26.73  17.7379 0.000025670819792
## wineColor:citricAcid          1     7.9    7.95   5.2732 0.0216861
## wineColor:rs                   1     5.8    5.75   3.8179 0.0507462
## wineColor:alcohol              1     0.1    0.08   0.0533 0.8174932
## Residuals                      6980 10519.2   1.51
##
## wineColor ***  

## fixedAcidity *

```

```

## va          ***
## citricAcid
## rs          *
## chlr        ***
## fsd         ***
## tsd         ***
## ph          ***
## sulphates   ***
## alcohol      ***
## wineColor:fixedAcidity ***
## wineColor:chlr
## wineColor:ph    ***
## wineColor:citricAcid  *
## wineColor:rs    .
## wineColor:alcohol
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#there is a lot of violation for residual sugar so since it's interaction does not contribute much to the model

newmodel3 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity + wineColor:chlr + wineColor:ph + wineColor:citricAcid + wineColor:rs + wineColor:alcohol)

vif(newmodel3) # that improved a lot :)
```

	wineColor	fixedAcidity	va
##	1186.713795	5.543772	2.146996
##	citricAcid	rs	chlr
##	4.805261	1.613318	14.356604
##	fsd	tsd	ph
##	2.715975	5.245377	9.178340
##	sulphates	alcohol	wineColor:fixedAcidity
##	1.468270	6.552991	55.717488
##	wineColor:chlr	wineColor:ph	wineColor:citricAcid
##	100.981665	851.510316	10.158052
##	wineColor:alcohol		
##	132.572648		

```

anova(newmodel2, newmodel3) #we can probaby drop predictors(p val is 0.05217)

## Analysis of Variance Table
## Model 1: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:chlr + wineColor:ph + wineColor:citricAcid + wineColor:rs +
##           wineColor:alcohol
## Model 2: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:chlr + wineColor:ph + wineColor:citricAcid + wineColor:alcohol
##   Res.Df   RSS Df Sum of Sq    F Pr(>F)
## 1   6980 10519
## 2   6981 10525 -1   -5.6841 3.7717 0.05217 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(newmodel3) #SSE: 10524.9
```

```
## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq F value      Pr(>F)
## wineColor                  1    77.0   76.99  51.0692 0.0000000000000982
## fixedAcidity                1     6.4    6.37   4.2266  0.0398331
## va                          1   273.3  273.33 181.2990 < 0.0000000000000022
## citricAcid                 1     1.3    1.34   0.8917  0.3450399
## rs                           1     6.2    6.18   4.0990  0.0429458
## chlr                        1   280.1  280.14 185.8104 < 0.0000000000000022
## fsd                          1    43.0   43.02  28.5317 0.000000095087045
## tsd                          1   160.2  160.15 106.2260 < 0.0000000000000022
## ph                           1    66.1   66.11  43.8524 0.000000000038042
## sulphates                   1   112.5  112.55  74.6506 < 0.0000000000000022
## alcohol                      1   682.8  682.77 452.8705 < 0.0000000000000022
## wineColor:fixedAcidity      1    20.4   20.45  13.5618  0.0002326
## wineColor:chlr               1     2.5    2.50   1.6615  0.1974412
## wineColor:ph                 1    26.7   26.73  17.7309 0.000025765701316
## wineColor:citricAcid        1     7.9    7.95   5.2711  0.0217122
## wineColor:alcohol            1     0.2    0.15   0.0995  0.7524216
## Residuals                    6981 10524.9   1.51
##
## wineColor                  ***
## fixedAcidity                *
## va                          ***
## citricAcid                 *
## rs                           *
## chlr                        ***
## fsd                         ***
## tsd                         ***
## ph                           ***
## sulphates                  ***
## alcohol                      ***
## wineColor:fixedAcidity      ***
## wineColor:chlr               ***
## wineColor:ph                 ***
## wineColor:citricAcid        *
## wineColor:alcohol            *
## Residuals
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#chlr has large vif violation and its interaction with color does not contribute much to SSR
#get rid of wineColor:chlr
newmodel4 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + alcohol)
```

```
vif(newmodel4) #still violation in fixedAcidity ph , and alcohol
```

```
##          wineColor      fixedAcidity           va
## 1181.965633          5.543563 2.133237
##          citricAcid             rs          chlr
##          4.753273          1.608083 2.412914
```

```

##          fsd                  tsd                  ph
##      2.711301            5.183898        9.144246
##      sulphates           alcohol wineColor:fixedAcidity
##      1.446264            5.969191        55.715554
##      wineColor:ph    wineColor:citricAcid    wineColor:alcohol
##      848.617046           10.063709       116.428857

anova(newmodel4) #SSE: 10529.4

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq F value      Pr(>F)
## wineColor                   1    77.0   76.99  51.0545 0.0000000000009894
## fixedAcidity                1     6.4    6.37   4.2253  0.0398617
## va                          1   273.3  273.33 181.2467 < 0.0000000000000022
## citricAcid                 1     1.3    1.34   0.8915  0.3451095
## rs                          1     6.2    6.18   4.0978  0.0429758
## chlr                        1   280.1  280.14 185.7568 < 0.0000000000000022
## fsd                         1    43.0   43.02  28.5235 0.0000000954899240
## tsd                         1   160.2  160.15 106.1954 < 0.0000000000000022
## ph                          1    66.1   66.11  43.8397 0.00000000000382862
## sulphates                   1   112.5  112.55  74.6290 < 0.0000000000000022
## alcohol                      1   682.8  682.77 452.7399 < 0.0000000000000022
## wineColor:fixedAcidity      1    20.4   20.45  13.5579  0.0002331
## wineColor:ph                 1    24.1   24.12  15.9951 0.0000641610591596
## wineColor:citricAcid        1     8.5    8.52   5.6514  0.0174679
## wineColor:alcohol            1     0.1    0.14   0.0960  0.7567101
## Residuals                   6982 10529.4   1.51
##
## wineColor                   ***
## fixedAcidity                 *
## va                          ***
## citricAcid
## rs                          *
## chlr                        ***
## fsd                         ***
## tsd                         ***
## ph                          ***
## sulphates                   ***
## alcohol                      ***
## wineColor:fixedAcidity      ***
## wineColor:ph                 ***
## wineColor:citricAcid        *
## wineColor:alcohol
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## citric acid does not provide and SSR but its interaction does. So instead, since wineColor:alcohol has

newmodel5 <- lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates)

vif(newmodel5) #still violations in fixedAcidity and ph

```

```

##          wineColor      fixedAcidity           va
## 1161.992085          5.537478        2.097156
##          citricAcid            rs           chlr
##          4.646950          1.535762        2.371846
##          fsd              tsd             ph
##          2.708569          5.180994       8.709002
##          sulphates      alcohol wineColor:fixedAcidity
##          1.442918          1.680410      55.606461
## wineColor:ph  wineColor:citricAcid
##          813.399662          9.893368

```

`anova(newmodel4,newmodel5) #it's okay to remove :) and fixed alcohol vif!`

```

## Analysis of Variance Table
##
## Model 1: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid + wineColor:alcohol
## Model 2: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid
##   Res.Df   RSS Df Sum of Sq    F Pr(>F)
## 1    6982 10529
## 2    6983 10530 -1  -0.14475 0.096 0.7567

```

`anova(newmodel5) #SSE: 10529.4`

```

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq  F value    Pr(>F)
## wineColor                  1    77.0   76.99  51.0611 0.0000000000000986
## fixedAcidity                1     6.4    6.37   4.2259  0.0398488
## va                          1   273.3  273.33 181.2702 < 0.0000000000000022
## citricAcid                 1     1.3    1.34   0.8916  0.3450783
## rs                           1     6.2    6.18   4.0984  0.0429623
## chlr                         1   280.1  280.14 185.7809 < 0.0000000000000022
## fsd                          1    43.0   43.02  28.5272 0.000000095308280
## tsd                          1   160.2  160.15 106.2091 < 0.0000000000000022
## ph                           1    66.1   66.11  43.8454 0.000000000038176
## sulphates                   1   112.5  112.55  74.6387 < 0.0000000000000022
## alcohol                      1   682.8  682.77 452.7985 < 0.0000000000000022
## wineColor:fixedAcidity       1    20.4   20.45  13.5596  0.0002328
## wineColor:ph                 1    24.1   24.12  15.9972 0.000064090978147
## wineColor:citricAcid         1     8.5    8.52   5.6522  0.0174606
## Residuals                   6983 10529.6    1.51
##
## wineColor                  ***
## fixedAcidity                *
## va                          ***
## citricAcid
## rs                           *
## chlr                         ***
## fsd                          ***
## tsd                          ***

```

```

## ph          ***
## sulphates ***
## alcohol    ***
## wineColor:fixedAcidity ***
## wineColor:ph ***
## wineColor:citricAcid *
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#the vif ph does not seem to be reducing that much, despite all the terms we have eliminated

vif(lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates +)

##          wineColor      fixedAcidity           va
##        41.134311       4.289122      2.096131
##          citricAcid        rs            chlr
##        4.562786       1.511573      2.370806
##          fsd             tsd            ph
##        2.697544       5.120984      1.602106
##          sulphates     alcohol wineColor:fixedAcidity
##        1.435813       1.673122      43.146205
## wineColor:citricAcid
##        9.732586

anova(newmodel5, lm(quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity + wineColor:ph + wineColor:citricAcid))

## Analysis of Variance Table
##
## Model 1: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid
## Model 2: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:citricAcid
##   Res.Df   RSS Df Sum of Sq    F    Pr(>F)
## 1   6983 10530
## 2   6984 10557 -1   -27.494 18.234 0.0000198 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#but the interaction terms is SUPER significant (0.0000198) to get rid of...lets see if any other less

#try wineColor
vif(lm(quality ~ fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + alcohol +)

##          fixedAcidity           va          citricAcid
##        3.722690       2.097156      4.615055
##          rs            chlr            fsd
##        1.512803       2.371831      2.698028
##          tsd            ph            sulphates
##        5.138765       1.802214      1.433149
##          alcohol fixedAcidity:wineColor ph:wineColor
##        1.672926       34.900179      28.794202
## citricAcid:wineColor
##        9.820022

```

```

anova(newmodel5, lm(quality ~ fixedAcidity + va + citricAcid + rs + chlr + fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity + wineColor:ph + wineColor:citricAcid))

## Analysis of Variance Table
##
## Model 1: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid
## Model 2: quality ~ fixedAcidity + va + citricAcid + rs + chlr + fsd +
##           tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     6983 10530
## 2     6984 10548 -1   -18.148 12.036 0.0005251 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#but the interaction terms is SUPER significant (0.0005251) to get rid of...lets see if any other less

#eliminated fixedAcidity, wineColor:fixedAcidity, citricAcid, wineColor:citricAcid
vif(lm(quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates + alcohol + wineColor:ph + chl

##      wineColor          va            rs            fsd            tsd
## 494.111500    1.846242    1.514455    2.630085    4.928064
##      ph        sulphates       alcohol         chlr wineColor:ph
## 4.515683     1.414815    1.618286    2.347519   464.739302

anova(newmodel5, lm(quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates + alcohol + wineCo

## Analysis of Variance Table
##
## Model 1: quality ~ wineColor + fixedAcidity + va + citricAcid + rs + chlr +
##           fsd + tsd + ph + sulphates + alcohol + wineColor:fixedAcidity +
##           wineColor:ph + wineColor:citricAcid
## Model 2: quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates +
##           alcohol + wineColor:ph + chlorides
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     6983 10530
## 2     6987 10539 -4   -9.8252 1.629 0.1639

#p-value: 0.1156 so this is a good option! and now all extreme VIF issues are gone

newmodel6 <- lm(quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates + alcohol + wineColor:ph + chl

vif(newmodel6)

##      wineColor          va            rs            fsd            tsd
## 494.111500    1.846242    1.514455    2.630085    4.928064
##      ph        sulphates       alcohol         chlr wineColor:ph
## 4.515683     1.414815    1.618286    2.347519   464.739302

anova(newmodel6)

## Analysis of Variance Table
##
## Response: quality
##              Df  Sum Sq Mean Sq  F value          Pr(>F)
## wineColor     1    77.0   76.99  51.0362 0.0000000000009985 ***

```

```

## va          1  259.4  259.36 171.9168 < 0.00000000000000022 ***
## rs          1     7.9    7.87   5.2163           0.02241 *
## fsd         1    31.5   31.49  20.8723   0.0000049936734306 ***
## tsd         1   264.6  264.60 175.3901 < 0.00000000000000022 ***
## ph          1    74.9   74.89  49.6401   0.0000000000020234 ***
## sulphates   1   104.7  104.74  69.4292 < 0.00000000000000022 ***
## alcohol      1   883.0  883.05 585.3342 < 0.00000000000000022 ***
## chlr         1     2.1    2.05   1.3622           0.24320
## wineColor:ph 1    45.8   45.82  30.3740   0.0000000368988618 ***
## Residuals    6987 10540.7   1.51

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

#SSE: 10540.7

```

There is a good model now. Try to see if it can be improved more!

```

#more general: chlorides instead of chlr
anova(lm(quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates + alcohol + wineColor:ph + chl

## Analysis of Variance Table
##
## Response: quality
##                         Df  Sum Sq Mean Sq F value            Pr(>F)
## wineColor          1    77.0   76.99  51.0427   0.0000000000009952 ***
## va                 1  259.4  259.36 171.9390 < 0.00000000000000022 ***
## rs                 1     7.9    7.87   5.2169           0.0224 *
## fsd                1    31.5   31.49  20.8750   0.0000049866838418 ***
## tsd                1   264.6  264.60 175.4127 < 0.00000000000000022 ***
## ph                 1    74.9   74.89  49.6465   0.0000000000020169 ***
## sulphates         1   104.7  104.74  69.4382 < 0.00000000000000022 ***
## alcohol             1   883.0  883.05 585.4097 < 0.00000000000000022 ***
## chlorides           1     1.4    1.35   0.8960           0.3439
## wineColor:ph       1    47.9   47.89  31.7450   0.0000000182646430 ***
## Residuals          6987 10539.4   1.51

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

anova(lm(quality ~ wineColor + va + rs + fsd + tsd + ph + sulphates + alcohol + wineColor:ph + chl

## Analysis of Variance Table
##
## Response: quality
##                         Df  Sum Sq Mean Sq F value            Pr(>F)
## wineColor          1    77.0   76.99  51.0459   0.0000000000009936
## va                 1  259.4  259.36 171.9495 < 0.00000000000000022
## rs                 1     7.9    7.87   5.2173           0.02239
## fsd                1    31.5   31.49  20.8763   0.0000049833688280
## tsd                1   264.6  264.60 175.4235 < 0.00000000000000022
## ph                 1    74.9   74.89  49.6496   0.0000000000020138
## sulphates         1   104.7  104.74  69.4424 < 0.00000000000000022
## alcohol             1   883.0  883.05 585.4456 < 0.00000000000000022
## chlorides           1     1.4    1.35   0.8961           0.34387
## wineColor:ph       1    47.9   47.89  31.7469   0.0000000182464970

```

```

## wineColor:chlorides      1      2.2      2.16    1.4292          0.23194
## Residuals                 6986  10537.2     1.51
##
## wineColor                  ***
## va                         ***
## rs                          *
## fsd                         ***
## tsd                         ***
## ph                          ***
## sulphates                   ***
## alcohol                      ***
## chlorides
## wineColor:ph                ***
## wineColor:chlorides
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
vif(lm(quality ~  wineColor +  va + rs  +  fsd + tsd +  ph + sulphates + alcohol + wineColor:ph + chlori

##           wineColor            va            rs
## 532.439682        1.864159        1.523191
##           fsd              tsd            ph
## 2.629895         4.913168       4.773458
##           sulphates        alcohol        chlorides
## 1.472267         1.472727       2.499634
## wineColor:ph wineColor:chlorides
## 484.570240        3.976886

```

FINAL MODEL

```

#va, rs, fsd, tsd, ph all all tranformed variables while the others are not
finalModel <- lm(quality ~  wineColor +  va + rs  +  fsd + tsd +  ph + sulphates + alcohol + wineColor:ph + chlori

anova(finalModel)

## Analysis of Variance Table
##
## Response: quality
##                               Df  Sum Sq Mean Sq   F value      Pr(>F)
## wineColor                  1    77.0   76.99  51.0459  0.00000000000009936
## va                        1   259.4   259.36 171.9495 < 0.00000000000000022
## rs                         1     7.9     7.87   5.2173      0.02239
## fsd                        1    31.5    31.49  20.8763  0.0000049833688280
## tsd                        1   264.6   264.60 175.4235 < 0.00000000000000022
## ph                          1    74.9    74.89  49.6496  0.00000000000020138
## sulphates                  1   104.7   104.74  69.4424 < 0.00000000000000022
## alcohol                     1   883.0   883.05 585.4456 < 0.00000000000000022
## chlorides                   1     1.4     1.35   0.8961      0.34387
## wineColor:ph                1    47.9    47.89  31.7469  0.000000182464970
## wineColor:chlorides         1     2.2     2.16   1.4292          0.23194
## Residuals                 6986  10537.2     1.51
##
```

```

## wineColor      ***
## va            ***
## rs             *
## fsd            ***
## tsd            ***
## ph             ***
## sulphates     ***
## alcohol         ***
## chlorides      ***
## wineColor:ph   ***
## wineColor:chlorides
## Residuals
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(finalModel)

##
## Call:
## lm(formula = quality ~ wineColor + va + rs + fsd + tsd + ph +
##      sulphates + alcohol + wineColor:ph + chlorides + chlorides:wineColor)
##
## Residuals:
##    Min      1Q  Median      3Q      Max
## -3.8928 -0.8308 -0.0163  0.8408  4.5648
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) 2.439228  0.707791  3.446      0.000572 ***
## wineColorW -4.483850  0.784067 -5.719  0.00000001117708295 ***
## va          -0.467420  0.044875 -10.416 < 0.0000000000000002 ***
## rs           0.302457  0.038053  7.948  0.0000000000000219 ***
## fsd          0.343986  0.037052  9.284 < 0.0000000000000002 ***
## tsd          -0.024764  0.003918 -6.320  0.00000000027683248 ***
## ph           -0.588179  0.201372 -2.921      0.003502 **
## sulphates    0.727423  0.119141  6.106  0.00000000107915420 ***
## alcohol       0.342590  0.014974 22.880 < 0.0000000000000002 ***
## chlorides    -1.395056  0.687580 -2.029      0.042503 *
## wineColorW:ph 1.344183  0.233603  5.754  0.00000000907703568 ***
## wineColorW:chlorides 1.354136  1.132710  1.195      0.231939
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.228 on 6986 degrees of freedom
## Multiple R-squared:  0.1427, Adjusted R-squared:  0.1414
## F-statistic: 105.7 on 11 and 6986 DF,  p-value: < 0.0000000000000022
vif(finalModel)

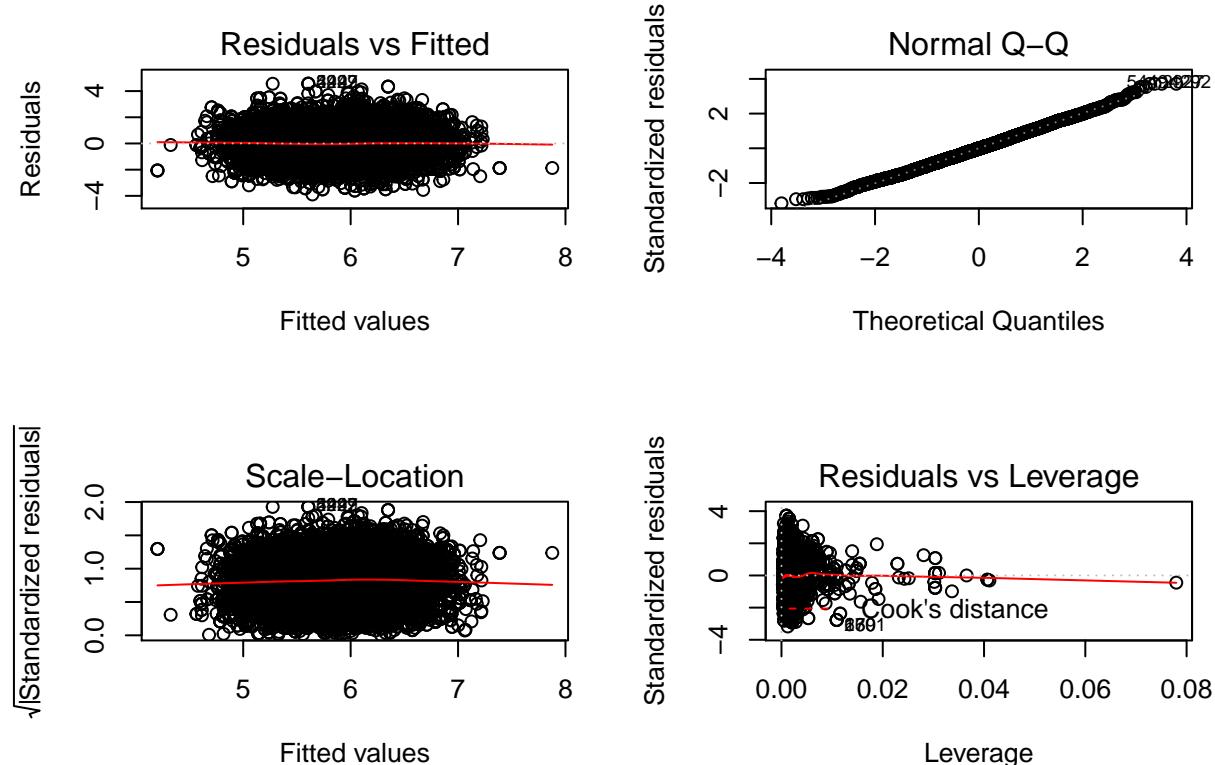
##          wineColor          va            rs
## 532.439682  1.864159  1.523191
##          fsd            tsd            ph
## 2.629895   4.913168  4.773458
##          sulphates        alcohol        chlorides
## 1.472267   1.472727  2.499634

```

```

##          wineColor:ph wineColor:chlorides
##          484.570240           3.976886
par(mfrow=c(2,2))
plot(finalModel)

```



Creating the csv file to turn in

```

wineTest <- read.csv("Wine2017Testing.csv", stringsAsFactors = TRUE)
wineColor <- ifelse(wineTest$Wine.Color=="W", 1, 0)
wineTest$Quality <- 2.439228 + (-4.483850*wineColor) + (-0.467420*log(wineTest$volatile.acidity)) + (0.000000*wineTest$alcohol)

write.csv(wineTest, file = "f304449872.csv")

```