Stats 101A Extra Credit Lecture 1B

Charles Liu (304804942)

3/6/2020

Loading Necessary Packages:

```r
library(readr)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------ tidyverse 1.3.0 --
```

```
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.2     v stringr 1.4.0
## v purrr   0.3.3     v forcats 0.4.0
```

```
## -- Conflicts --------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:dplyr':
##
##     nasa
```

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.3
```

```
## Loading required package: magrittr
##
## Attaching package: 'magrittr'
## The following object is masked from 'package:purrr':
##
##     set_names
## The following object is masked from 'package:tidyr':
##
##     extract
```

Loading the Data:

```
Bordeaux <- read_csv("C:/Users/cliuk/Documents/UCLA Works/UCLA Winter 2020/Stats 101A/Extra Credit/Bord
```

```
## Parsed with column specification:
## cols(
##   Wine = col_character(),
##   Price = col_double(),
##   ParkerPoints = col_double(),
##   CoatesPoints = col_double(),
##   P95andAbove = col_double(),
##   FirstGrowth = col_double(),
##   CultWine = col_double(),
##   Pomerol = col_double(),
##   VintageSuperstar = col_double()
## )
```

```
attach(Bordeaux)
```

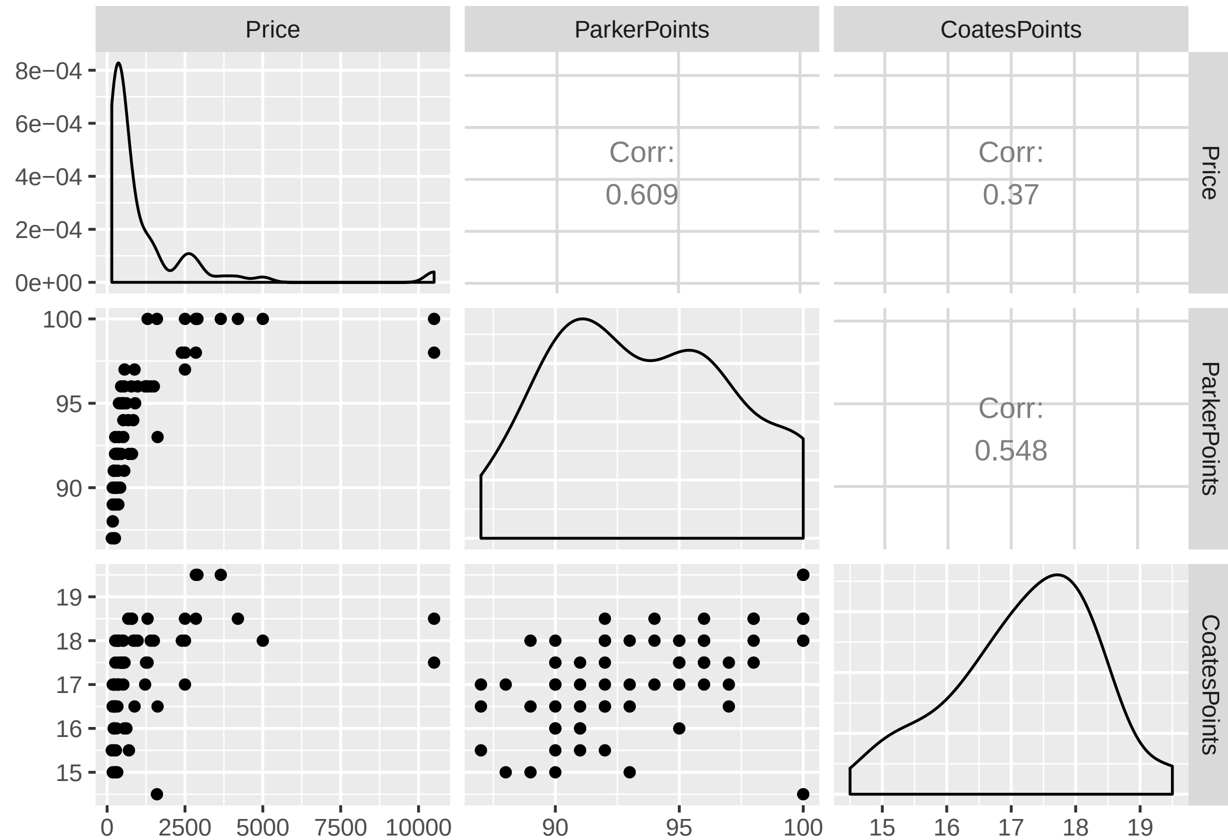
Question (1):

1A) First, using ifelse function, convert x3, x4, x5, x6 and x7 variables into categorical with "Yes" instead of 1 and "No" instead of 0. Name the new variables as P95andAboveNew, FirstGrowthNew, CultWineNew, PomerolNew and VintageSuperstarNew respectively.

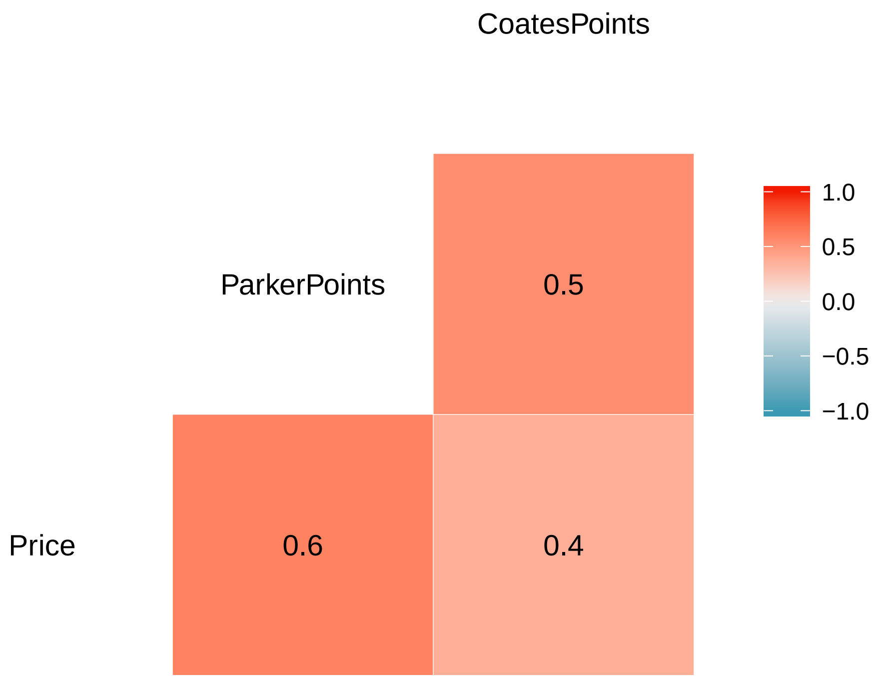
```
## x3
P95andAboveNew <- ifelse(P95andAbove == 1, "Yes", "No")

## x4
FirstGrowthNew <- ifelse(FirstGrowth == 1, "Yes", "No")

## x5
CultWineNew <- ifelse(CultWine == 1, "Yes", "No")

## x6
PomerolNew <- ifelse(Pomerol == 1, "Yes", "No")

## x7
VintageSuperstarNew <- ifelse(VintageSuperstar == 1, "Yes", "No")
```

1B) Create ggpairs plot for the response and the other two numerical predictors. What did you notice?

```r
df_bordeaux <- Bordeaux[, c(2, 3, 4)]
ggpairs(df_bordeaux)
```




```r
# Using ggcorr() to get a better look at the correlation between these variables
ggcorr(df_bordeaux, palette = "RdBu", label = TRUE)
```



What I noticed is that there is that the correlation between the variables are not related as much to each other. The highest correlation is between Price and ParkerPoints at 0.609. This is good because we don't want our predictors in a regression model to be highly correlated with each other. We see that our predictors are not completely independent of each other. We want them to be more correlated with the response variable. I can see that the correlation between Price and CoatesPoints are not very correlated, which could result in it as a not very good predictor.

1C) Create a MLR (call it m0) using the two numerical predictors only. Study the summary, anova, vif and diagnostics of the model.

```
m0 <- lm(Price ~ ParkerPoints + CoatesPoints, data = Bordeaux)
summary(m0)
```

```
##
## Call:
## lm(formula = Price ~ ParkerPoints + CoatesPoints, data = Bordeaux)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1839.9  -626.9  -207.3   293.2  8029.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -27629.03    4508.03  -6.129 4.84e-08 ***
## ParkerPoints    291.74      57.32   5.089 2.97e-06 ***
## CoatesPoints     86.20     190.98   0.451    0.653
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1515 on 69 degrees of freedom
## Multiple R-squared:  0.3722, Adjusted R-squared:  0.354
## F-statistic: 20.45 on 2 and 69 DF,  p-value: 1.058e-07
```

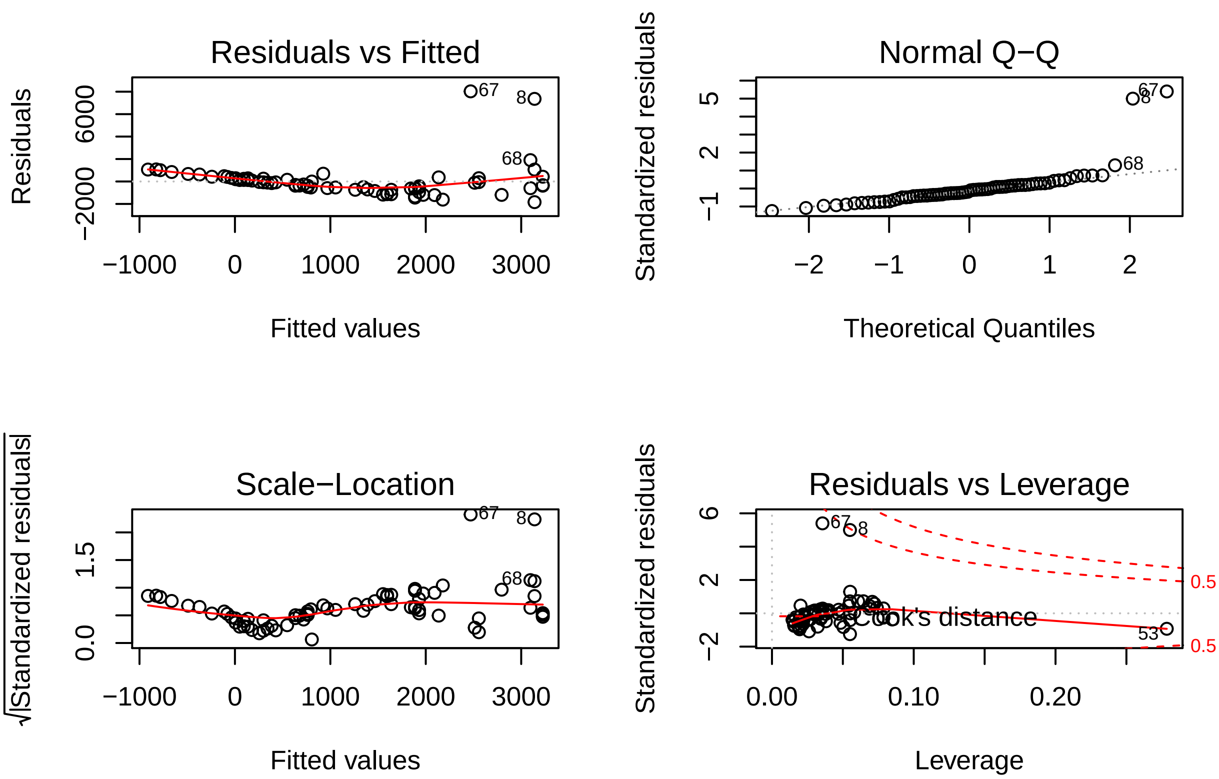
```r
anova(m0)
```

```
## Analysis of Variance Table
##
## Response: Price
##              Df    Sum Sq  Mean Sq F value    Pr(>F)
## ParkerPoints  1  93400998 93400998 40.7058 1.733e-08 ***
## CoatesPoints  1    467440   467440  0.2037    0.6532
## Residuals    69 158323083  2294537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
vif(m0)
```

```
## ParkerPoints CoatesPoints
##      1.42927      1.42927
```

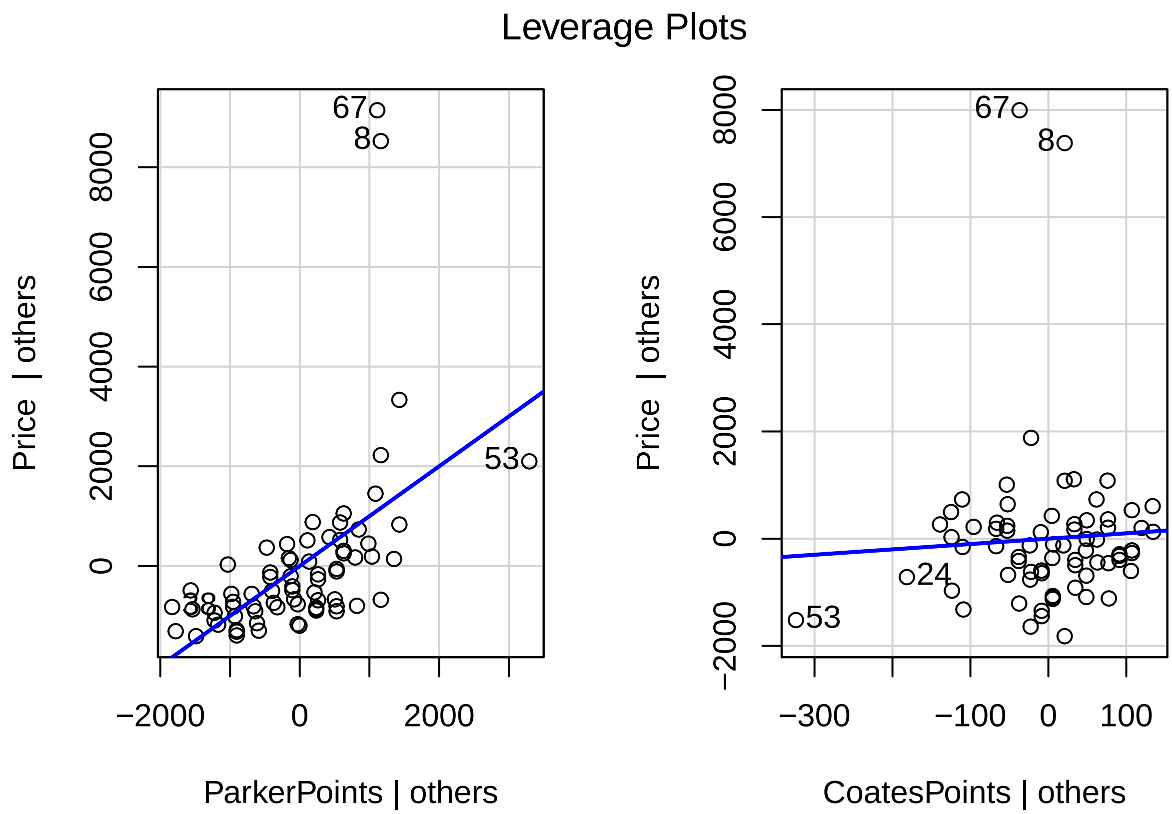
```r
par(mfrow = c(2, 2))
plot(m0)
```



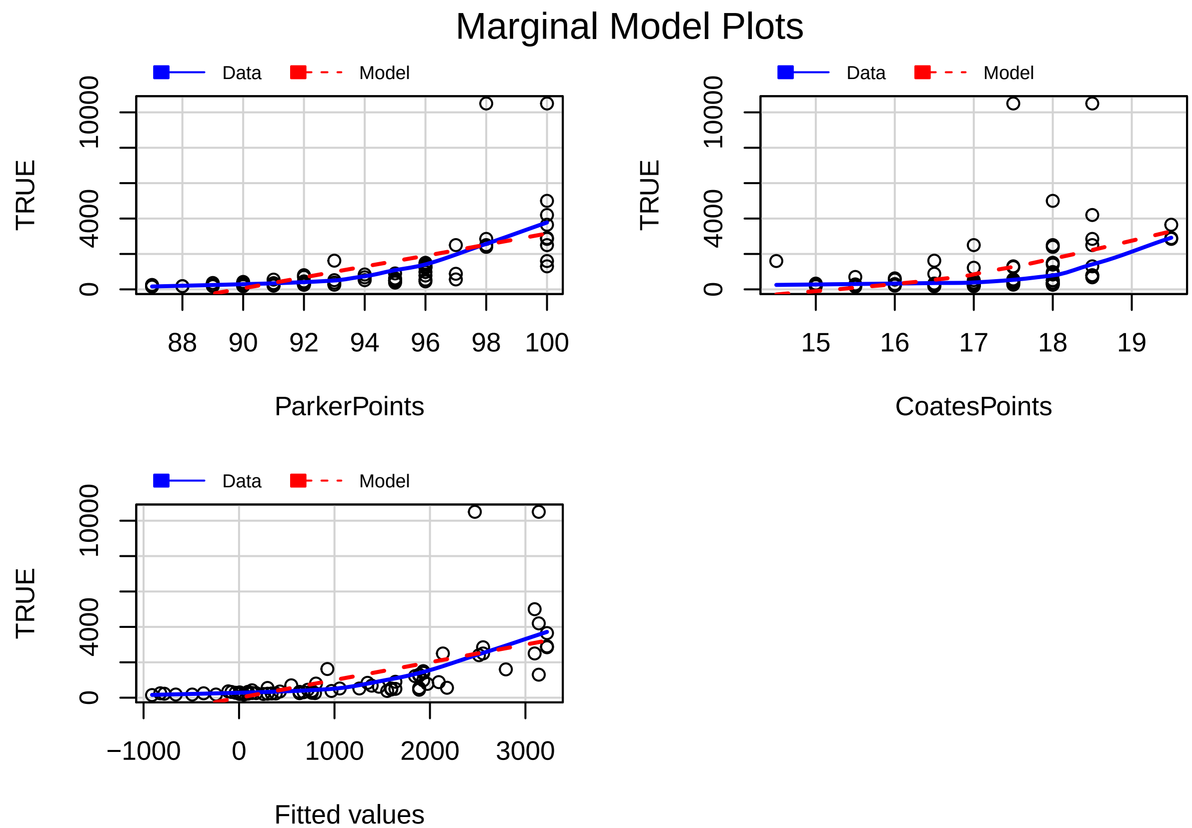
From our summary, we can see that the intercept and ParkerPoints are statistically significant, but our R-squared is not very high (0.3722). This tells us that maybe our model isn't the best fitted model. We can also see that the variable CoatesPoints is not statisticallty significant. Our anova test tells us that only ParkerPoints is significant. Since CoatesPoints is not statistically significant, we REJECT the null hypothesis. The ViF is good since none of the predictors have a value greater than 5. For the diagnostic plots, we can see that the Scale-Location and Residuals vs. Fitted plots have their linearity assumptions not satisfied. As for the normality assumption, it is overall satisfied with the exception of a few small points. We can see the Leverage and Outliers have a couple of bad leverages. We would need to conduct some transformations to make the plots look better.

1D) Use leveragePlots and mmps on m0. What did you notice? Run powerTransform and inversResponsePlot functions on m0. What do you need to do to make m0 a better model? Do it.

```
leveragePlots(m0)
```




```
mmps(m0)
```

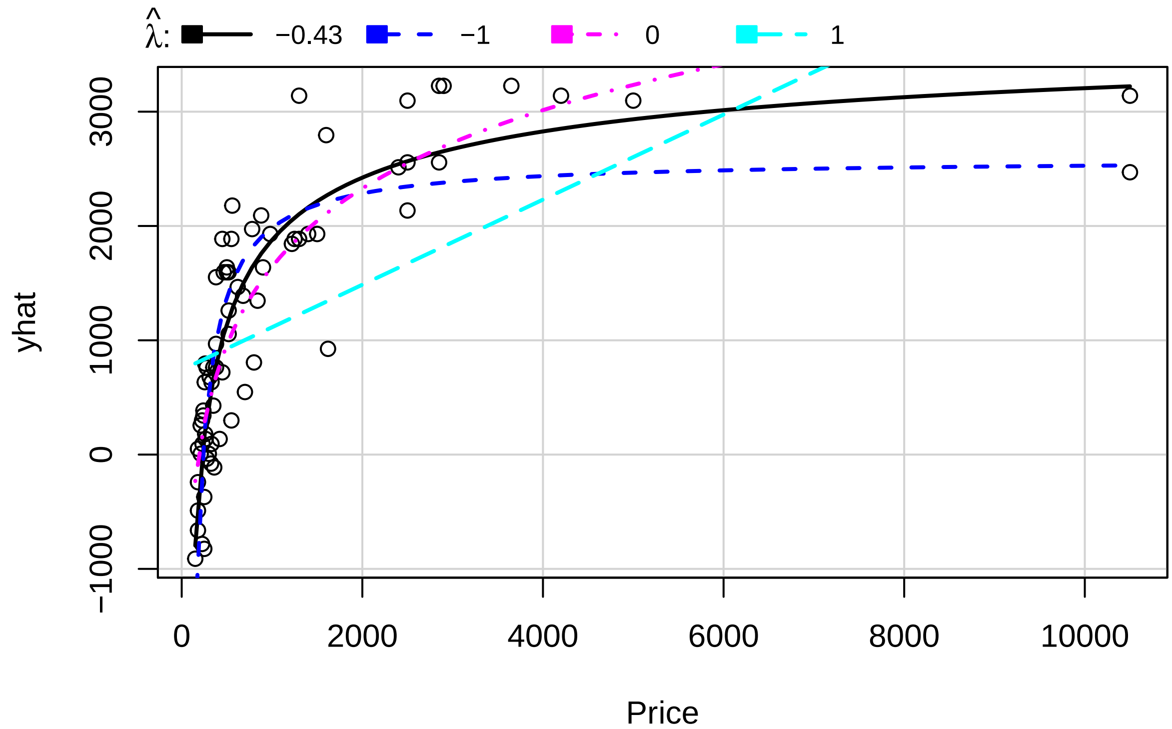


From our leveragePlots(), we can see that ParkerPoints variable has some increasing pattern, and this mean it is significant. As for CoatesPoints, the line is horizontal, mean this is not significant. From our mmps(), we observe that our model fits the data relatively well for ParkerPoints and Fitted Values. Meanwhile, the CoatesPoints data is not very well aligned with our model, telling us this is not a significant predictor to our model.

```
summary(powerTransform(cbind(Price, ParkerPoints, CoatesPoints) ~ 1))
```

```
## bcPower Transformations to Multinormality
##              Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Price          -0.4408        -0.5      -0.6218      -0.2599
## ParkerPoints    0.4785         1.0      -3.9053       4.8623
## CoatesPoints    4.1948         2.0       1.3738       7.0159
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                 LRT df       pval
## LR test, lambda = (0 0 0) 38.11429  3 2.6733e-08
##
## Likelihood ratio test that no transformations are needed
##                                 LRT df       pval
## LR test, lambda = (1 1 1) 260.7559  3 < 2.22e-16
```

```
par(mfrow = c(1, 1))
inverseResponsePlot(m0)
```




```
##        lambda      RSS
## 1 -0.4316934 17810925
## 2 -1.0000000 22176953
## 3  0.0000000 21876238
## 4  1.0000000 58929581
```

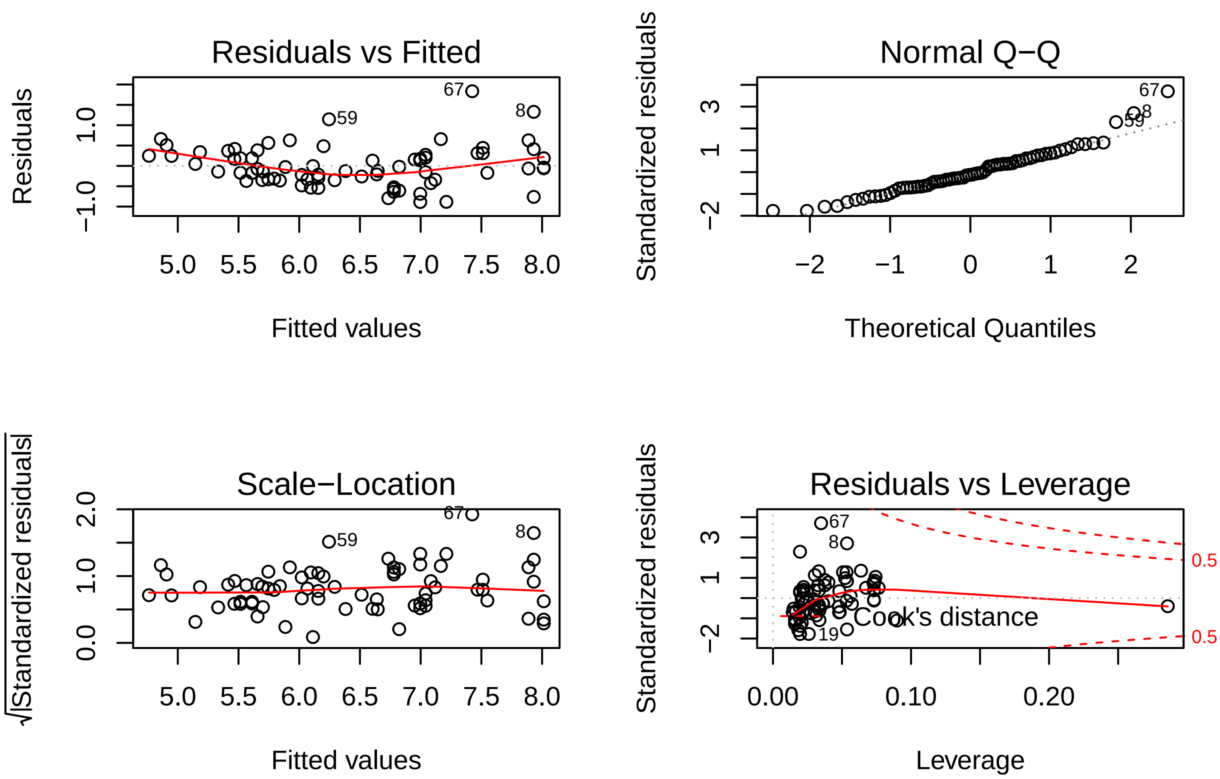
Inverse response plot is suggesting power of 0 as the best transformation of the response variable. We would take the log() of x and y.

```
m0_log <- lm(log(Price) ~ log(ParkerPoints) +
               log(CoatesPoints), data = Bordeaux)

summary(m0_log)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints),
##     data = Bordeaux)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88735 -0.33310 -0.06152  0.28237  1.83445
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -92.255      6.923 -13.327   <2e-16 ***
## log(ParkerPoints)    20.761      1.774  11.700   <2e-16 ***
```

```
## log(CoatesPoints)    1.569      1.067   1.471    0.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5044 on 69 degrees of freedom
## Multiple R-squared:  0.7618, Adjusted R-squared:  0.7549
## F-statistic: 110.3 on 2 and 69 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(m0_log)
```



The Normal Q-Q plot looks better and the R-squared has gone up to 0.7618. Not only that, but now both our predictors are significant.

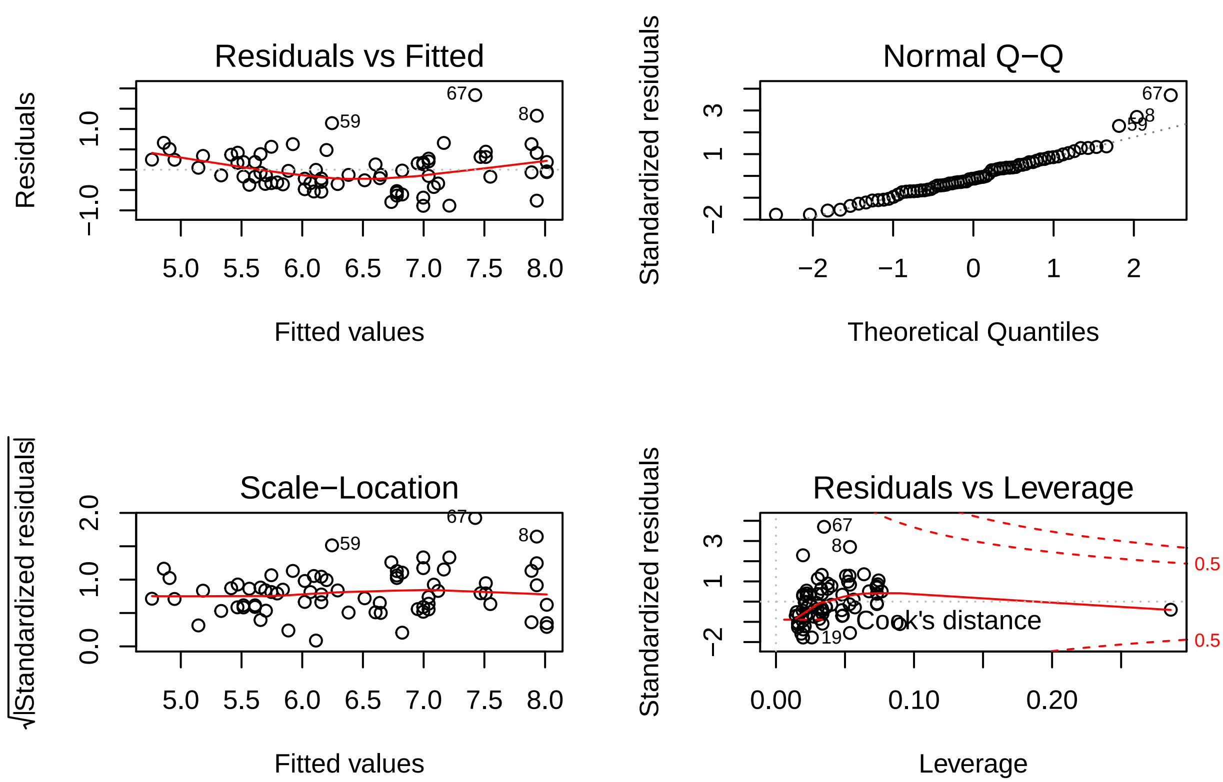
1E) A statistician suggested to use log transformation on the response variable and the two numerical predictors. Do you agree or disagree with that suggestion? Try it first then compare to m0.

```r
m0_log <- lm(log(Price) ~ log(ParkerPoints) +
               log(CoatesPoints), data = Bordeaux)

summary(m0_log)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints),
##     data = Bordeaux)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88735 -0.33310 -0.06152  0.28237  1.83445
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -92.255      6.923 -13.327   <2e-16 ***
## log(ParkerPoints)   20.761      1.774  11.700   <2e-16 ***
## log(CoatesPoints)    1.569      1.067   1.471    0.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5044 on 69 degrees of freedom
## Multiple R-squared:  0.7618, Adjusted R-squared:  0.7549
## F-statistic: 110.3 on 2 and 69 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(m0_log)
```

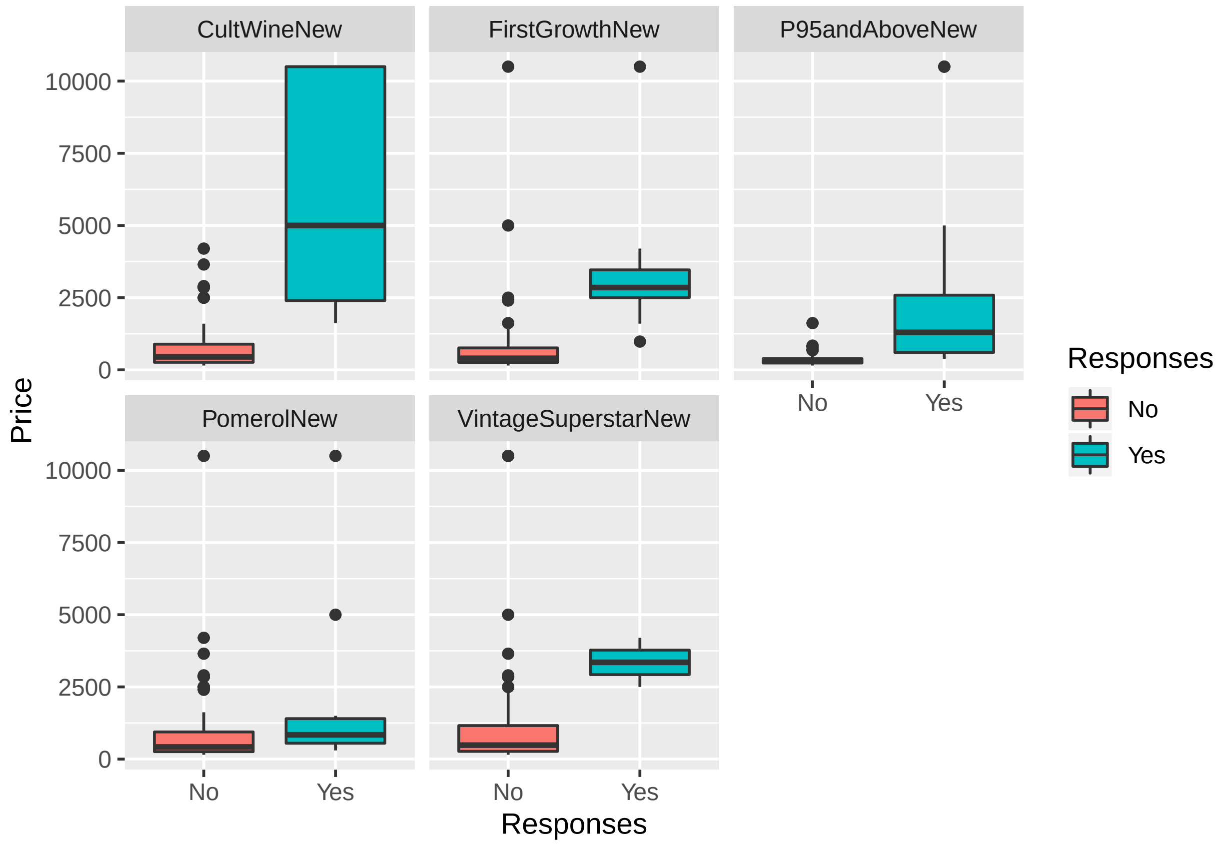


The Normal Q-Q plot looks better and the R-squared has gone up to 0.7618. Not only that, but now both our predictors are significant. The diagnostic plot for Scale-Location has its assumption satisfied. The Residuals vs. Fitted and Residuals vs. Leverage are better as well, but they are not satisfied. I agree with the statistian suggestion.

Question (2):

2A) Create a side by side box plots of the variable Price for each of the categorical variables created in Question 1 Part A. Which of these predictors you think are good predictor for the variable "Price"? Why?

```
total_Bordeaux <- add_column(Bordeaux, P95andAboveNew, FirstGrowthNew, CultWineNew, PomerolNew, VintageS


total_Bordeaux %>%
  select(Price, P95andAboveNew, FirstGrowthNew, CultWineNew, PomerolNew, VintageSuperstarNew) %>%
  pivot_longer(
    cols = "P95andAboveNew":"VintageSuperstarNew",
    names_to = "Types",
    values_to = "Responses") %>%
  ggplot(aes(x = Responses, Price, fill = Responses)) +
  geom_boxplot() +
  facet_wrap(~ Types, nrow = 2)
```



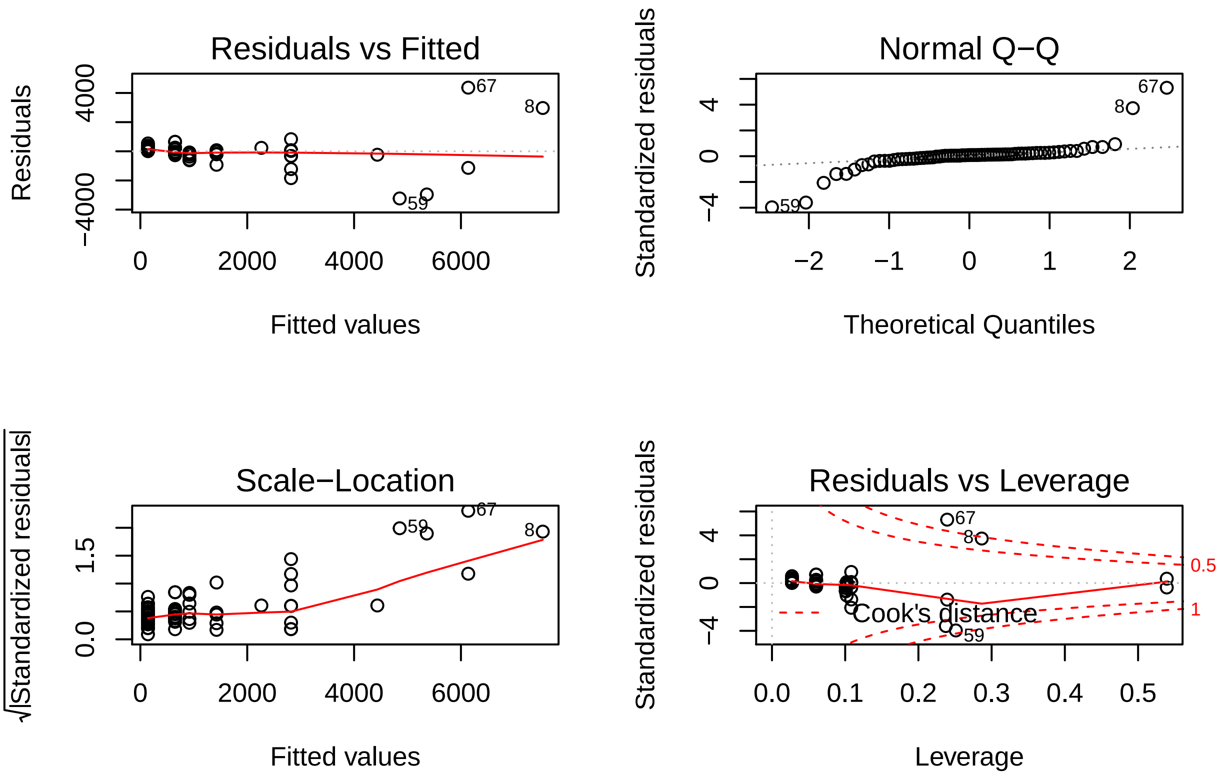
I would choose the predictor, CultWineNew, because both of the responses cover a wider range for the Price response. We can get a better prediction based of the variability of the responses. You can clearly see that the "Yes" boxplot covers a wider range of the Price.

2B) Create a MLR (call it mcat) using all categorical predictor to predict "Price". Check summary, anova, vif and diagnostics of the model. Does the MLR summary agrees with your answer in part A?

```r
mcat <- lm(Price ~ P95andAboveNew + FirstGrowthNew + CultWineNew +
           PomerolNew + VintageSuperstarNew)

summary(mcat)
```

```
##
## Call:
## lm(formula = Price ~ P95andAboveNew + FirstGrowthNew + CultWineNew +
##     PomerolNew + VintageSuperstarNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3233.6  -156.3    77.0   187.6  4363.2
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)               142.4      155.7   0.914   0.3638
## P95andAboveNewYes         507.5      261.5   1.941   0.0566 .
## FirstGrowthNewYes        2170.6      372.9   5.821 1.88e-07 ***
## CultWineNewYes           4711.3      450.9  10.448 1.26e-15 ***
## PomerolNewYes             775.7      303.1   2.559   0.0128 *
## VintageSuperstarNewYes   1614.9      695.3   2.323   0.0233 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 941.7 on 66 degrees of freedom
## Multiple R-squared:  0.7679, Adjusted R-squared:  0.7503
## F-statistic: 43.67 on 5 and 66 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(mcat)
```




```r
anova(mcat)
```

```
## Analysis of Variance Table
##
## Response: Price
##                     Df    Sum Sq   Mean Sq  F value    Pr(>F)
## P95andAboveNew       1  58316213  58316213  65.7561 1.689e-11 ***
## FirstGrowthNew       1  23288652  23288652  26.2598 2.806e-06 ***
## CultWineNew          1 102095811 102095811 115.1211 4.143e-16 ***
## PomerolNew           1   5174276   5174276   5.8344   0.01849 *
## VintageSuperstarNew  1   4784088   4784088   5.3944   0.02329 *
## Residuals           66  58532482    886856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
vif(mcat)
```

```
##      P95andAboveNew      FirstGrowthNew         CultWineNew          PomerolNew
##            1.370678            1.350000            1.066822            1.103529
## VintageSuperstarNew
##            1.059948
```

Based off of the summary(), we can see that our predictor for CultWineNew is significant, which is a good sign. Our diagnostic plots are the following: (Residuals vs. Fitted) assumption satisfied, (Normal Q-Q) assumption of normality not satisfied, (Scale-Location) not satisfied since there is an increasing line, (Residuals vs Leverage) not satisfied. Our anova() tells us that CultWineNew is definitely significant at a value for our F-statistics of Pr(>F) = 4.143e-16. Therefore, we REJECT our null hypothesis. The vif() is satisfied for all as the predictors all have values less than 5. Overall, the summary agrees with my answer from Q2a.

2C) Interpret the y-intercept and all the partial slopes in your mcat MLR.

Our intercept is the average "No" responses from all our predictors. Below is the partial slopes for mcat MLR. For every 1 unit of increase of Price, the slope for P95andAboveNewYes is 507.5, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for FirstGrowthNewYes is 2170.6, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for CultWineNewYes is 4711.3, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for PomerolNewYes is 775.7, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for VintageSuperstarNewYes is 1614.9, which is the average "Yes" responses.

```
summary(mcat)
```

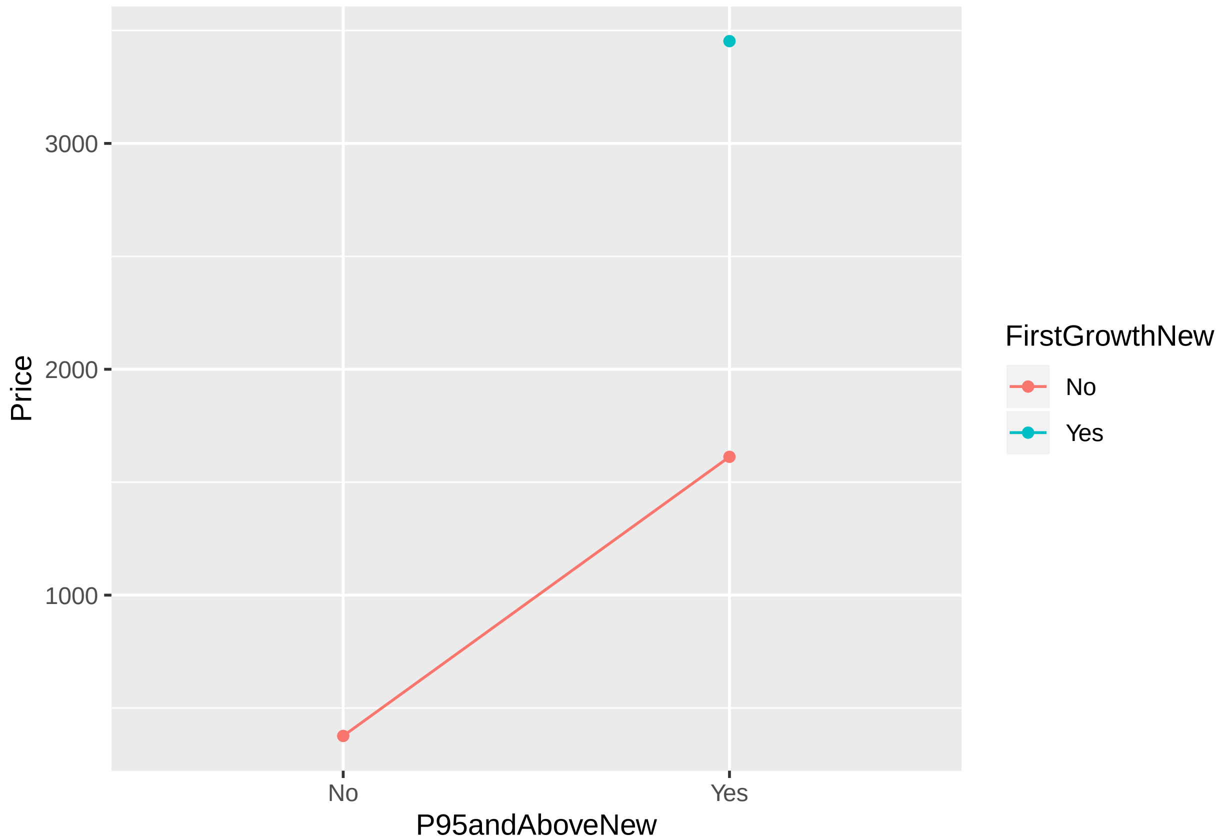
```
##
## Call:
## lm(formula = Price ~ P95andAboveNew + FirstGrowthNew + CultWineNew +
##     PomerolNew + VintageSuperstarNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3233.6  -156.3    77.0   187.6  4363.2
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)               142.4      155.7   0.914   0.3638
## P95andAboveNewYes         507.5      261.5   1.941   0.0566 .
## FirstGrowthNewYes        2170.6      372.9   5.821 1.88e-07 ***
## CultWineNewYes           4711.3      450.9  10.448 1.26e-15 ***
## PomerolNewYes             775.7      303.1   2.559   0.0128 *
## VintageSuperstarNewYes   1614.9      695.3   2.323   0.0233 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 941.7 on 66 degrees of freedom
## Multiple R-squared:  0.7679, Adjusted R-squared:  0.7503
## F-statistic: 43.67 on 5 and 66 DF,  p-value: < 2.2e-16
```

```
# Y = B0 + B1*x1 + B2*x2 + B3*x3 + B4*x4 + B5*x5 -->

# Price = Intercept + P95andAboveNewYes*(x1) + FirstGrowthNewYes*(x2) + CultWineNewYes*(x3) + PomerolNe

# Price = 142.4  + 507.5*(x1) + 2170.6*(x2) + 4711.3*(x3) + 775.7*(x4) + 1614.9*(x5)

### The slopes are as follows:
### P95andAboveNewYes --> 507.5
### FirstGrowthNewYes --> 2170.6
### CultWineNewYes --> 4711.3
### PomerolNewYes --> 775.7
### VintageSuperstarNewYes --> 1614.9
```

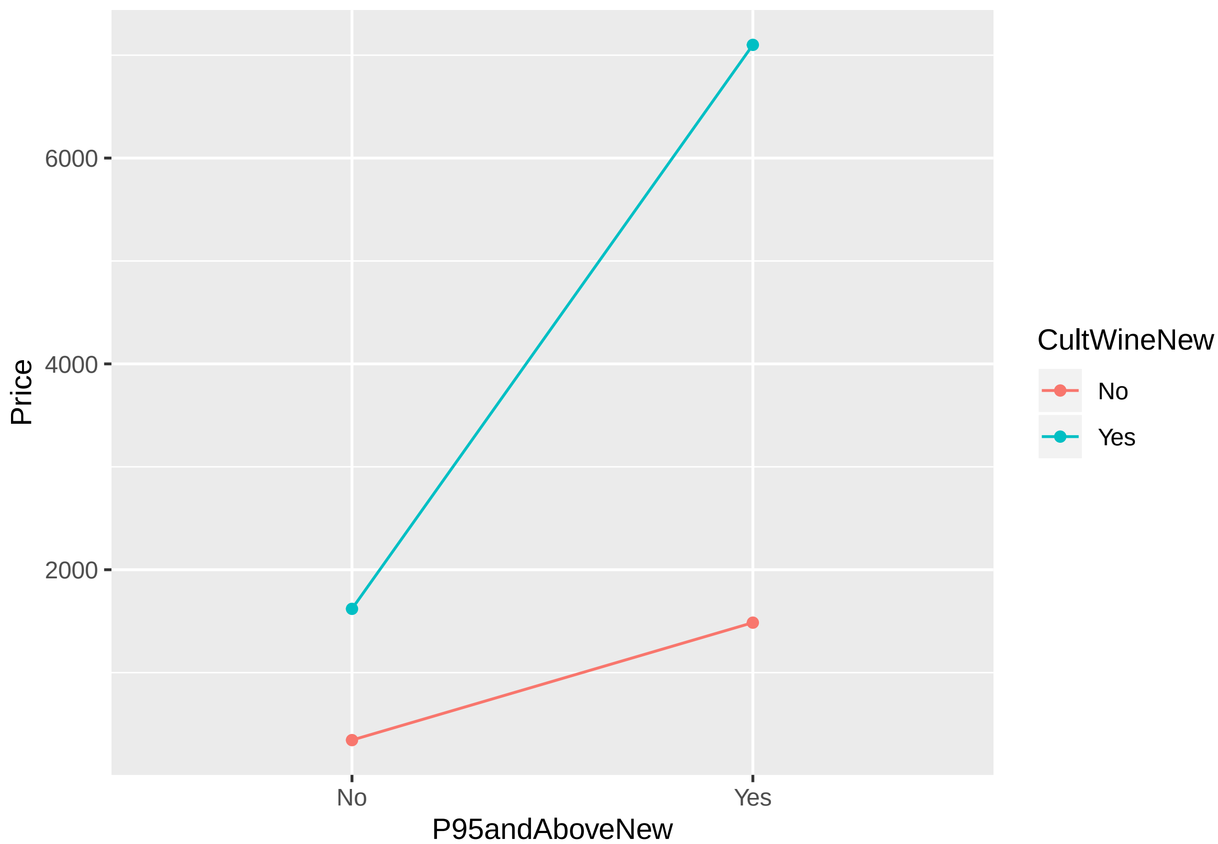
2D) Create 5C2 = 10 pairwise interaction plots. Which ones you think should be added to your model mcat?

```
## P95andAboveNew & FirstGrowthNew
p1 <- ggplot() + aes(x = P95andAboveNew, y = Price, color = FirstGrowthNew,
               group = FirstGrowthNew) +
```

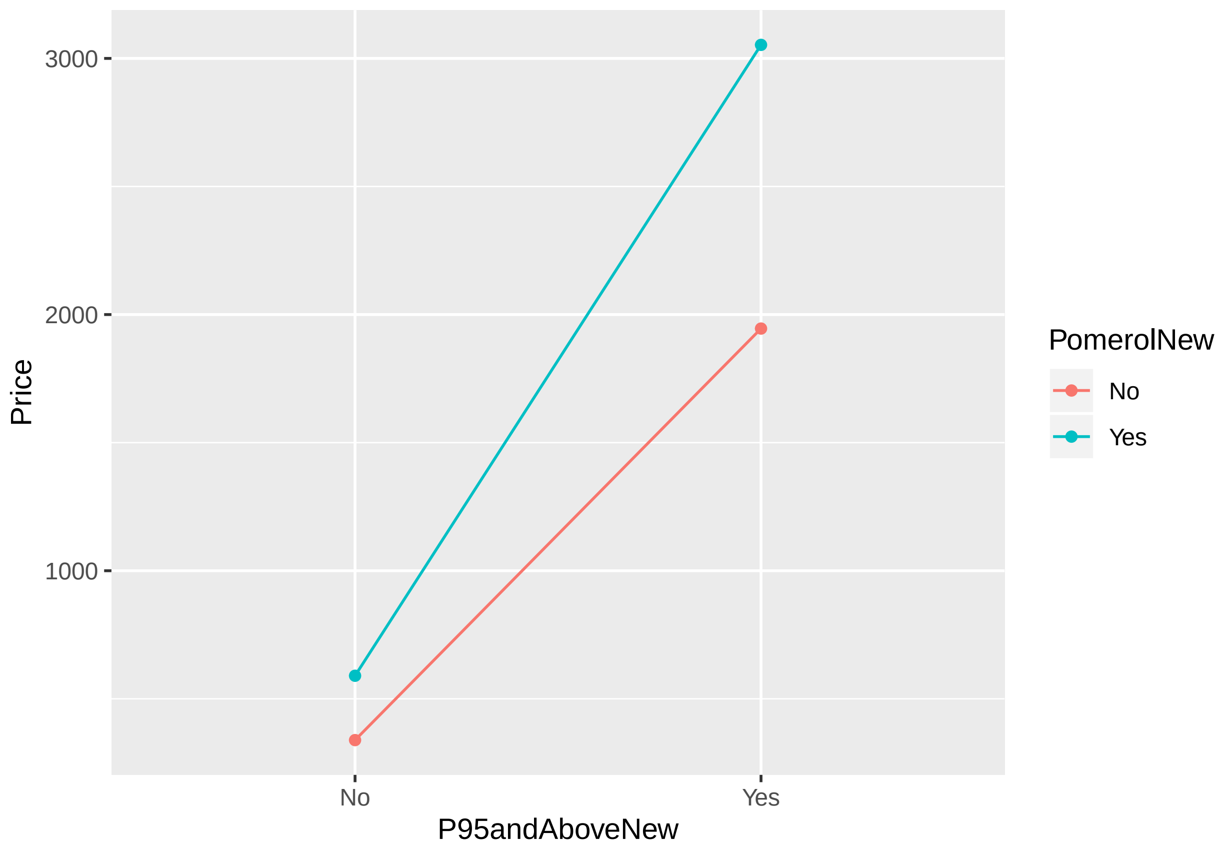
```r
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p1 # Do NOT add
```



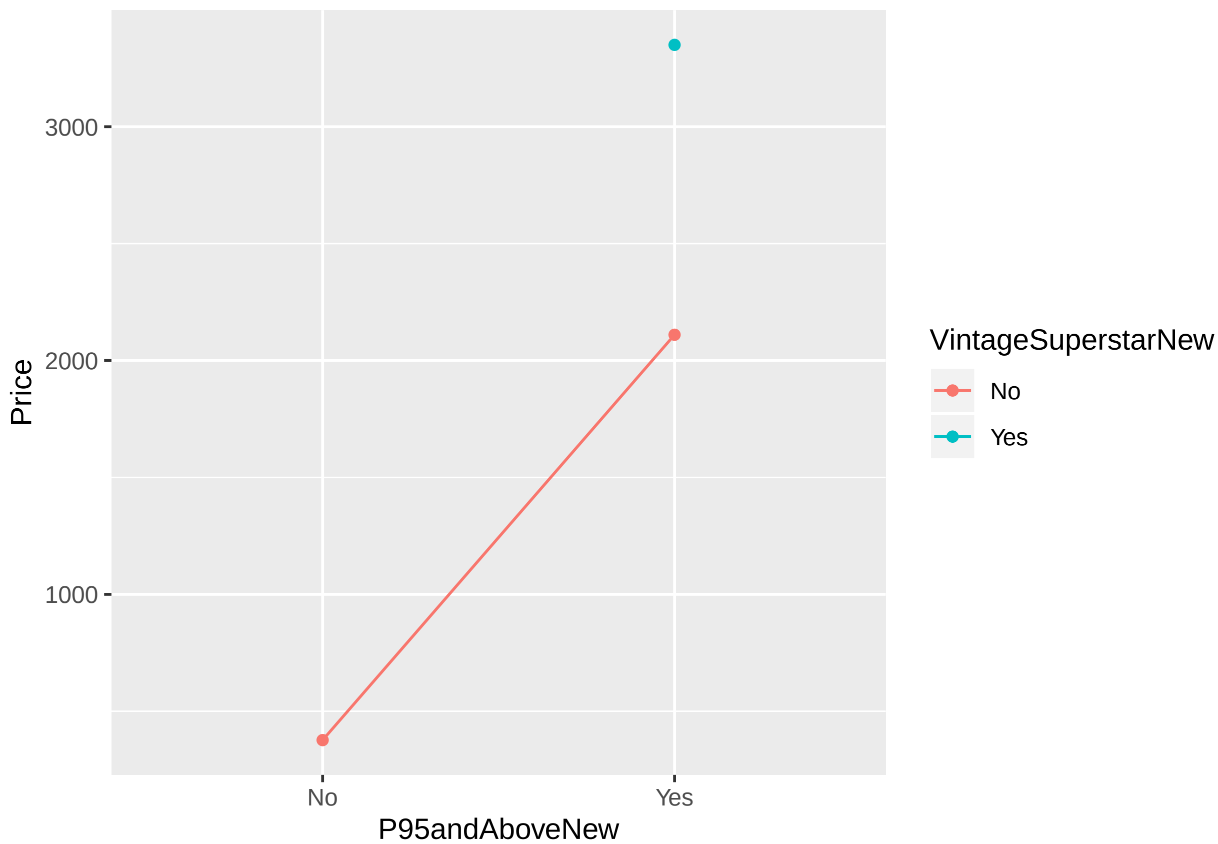

```r
## P95andAboveNew & CultWineNew
p2 <- ggplot() + aes(x = P95andAboveNew, y = Price, color = CultWineNew,
               group = CultWineNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p2 # Do add
```



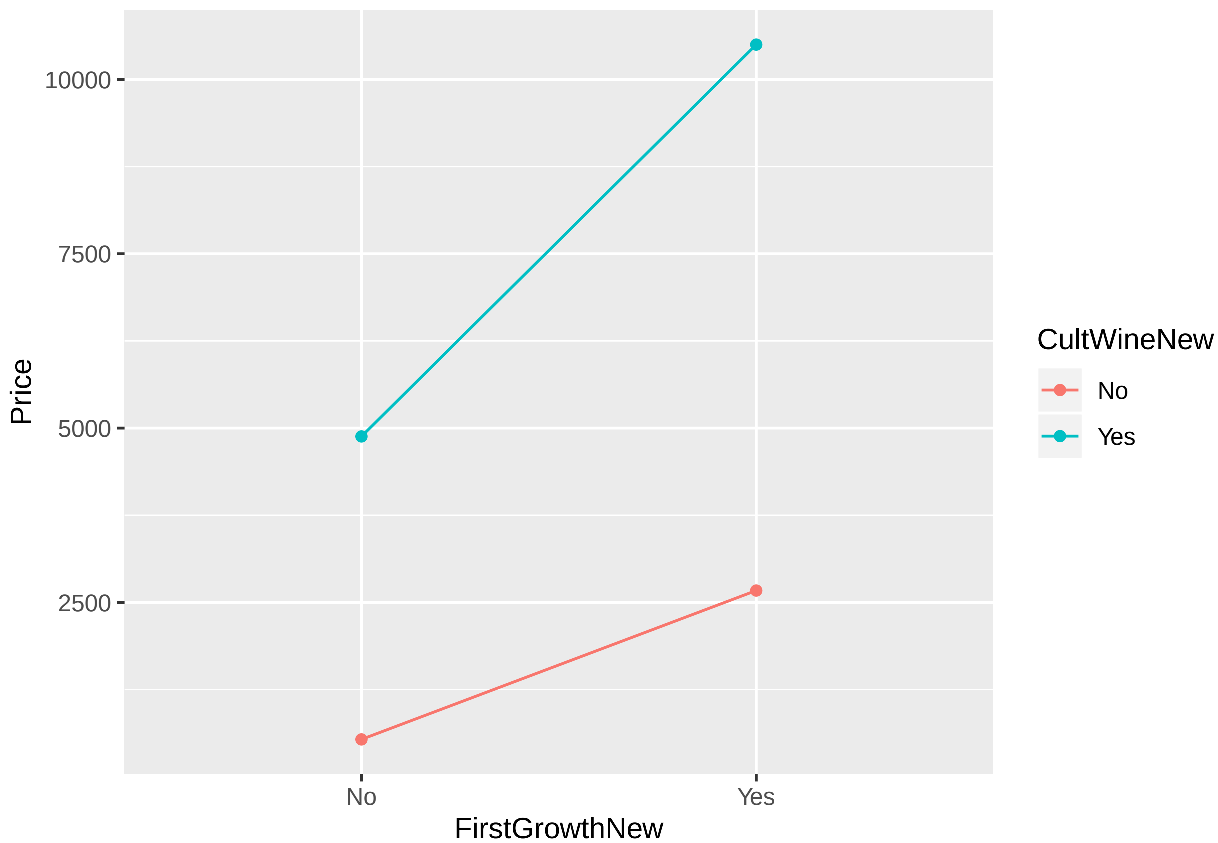

```
## P95andAboveNew & PomerolNew
p3 <- ggplot() + aes(x = P95andAboveNew, y = Price, color = PomerolNew,
               group = PomerolNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p3 # Do add
```



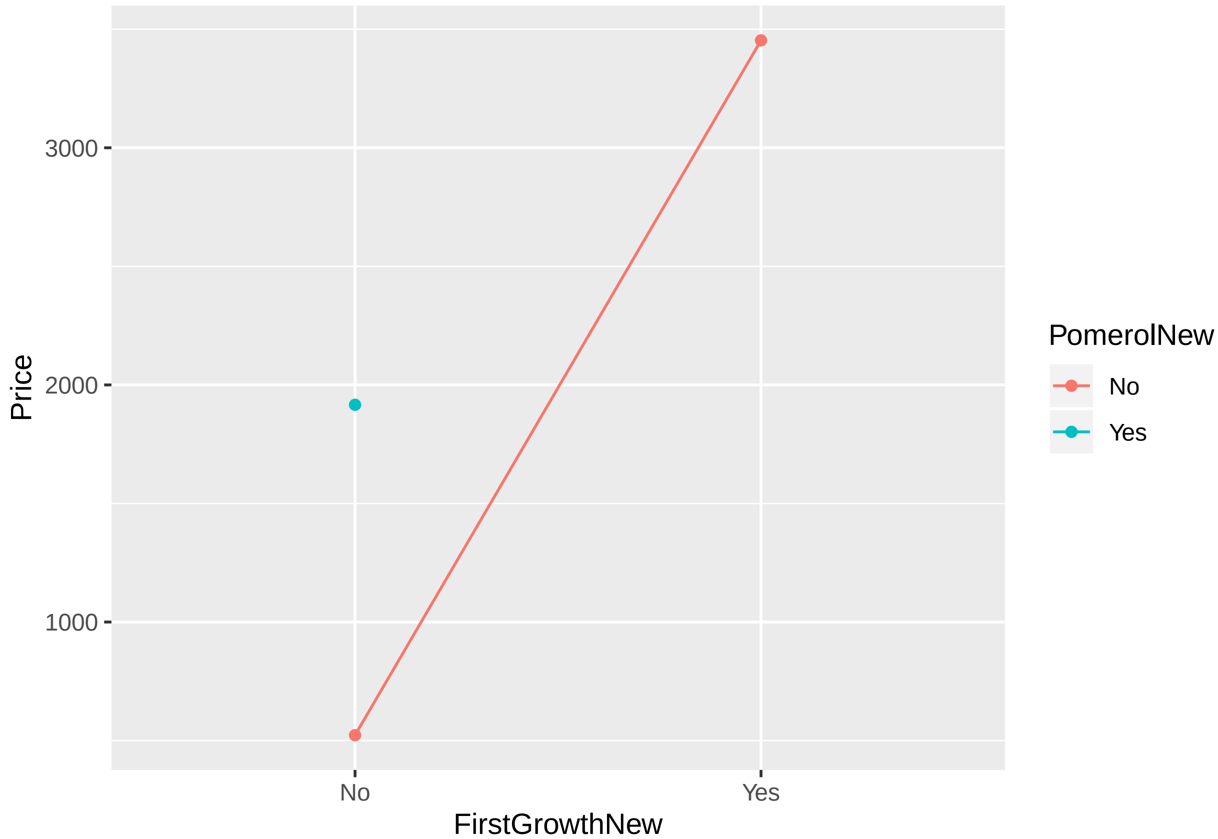

```r
## P95andAboveNew & VintageSuperstarNew
p4 <- ggplot() + aes(x = P95andAboveNew, y = Price, color = VintageSuperstarNew,
                     group = VintageSuperstarNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p4 # Do NOT add
```



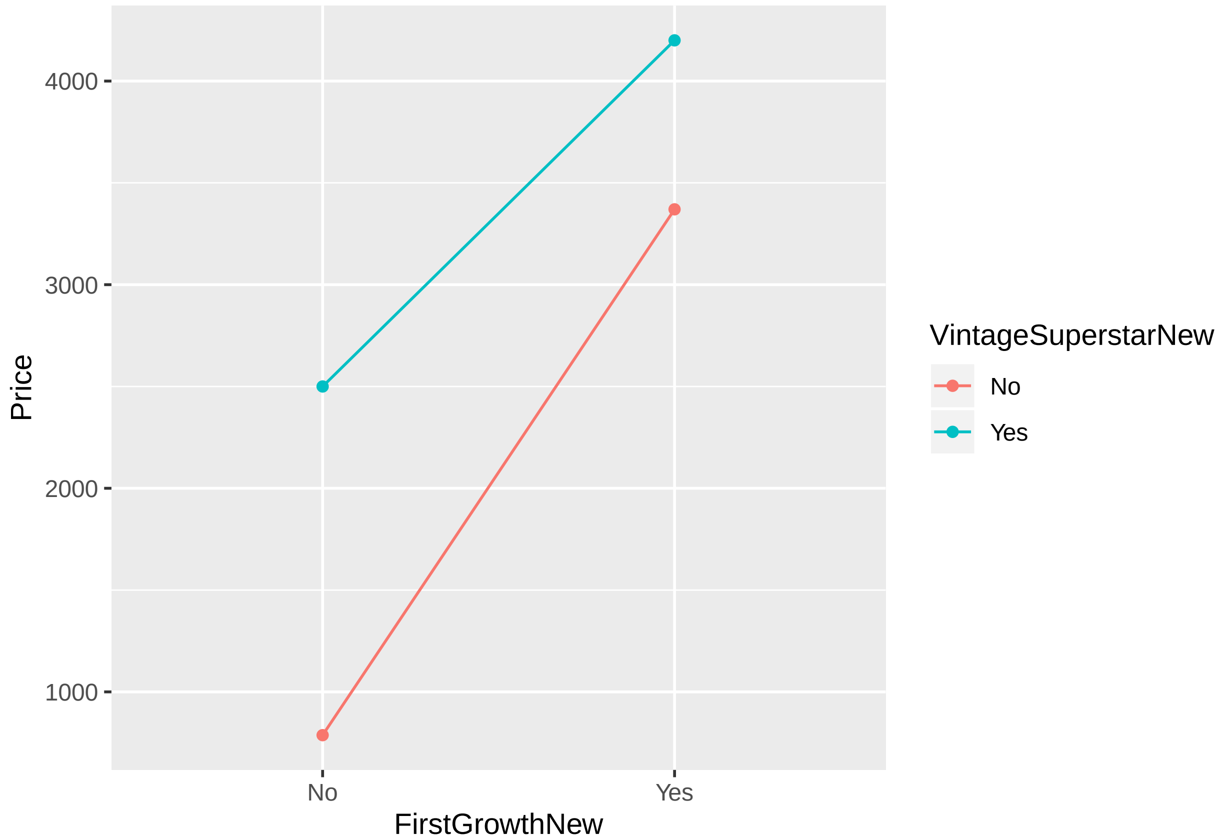

```r
## FirstGrowthNew & CultWineNew
p5 <- ggplot() + aes(x = FirstGrowthNew, y = Price, color = CultWineNew,
               group = CultWineNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p5 # Do add
```



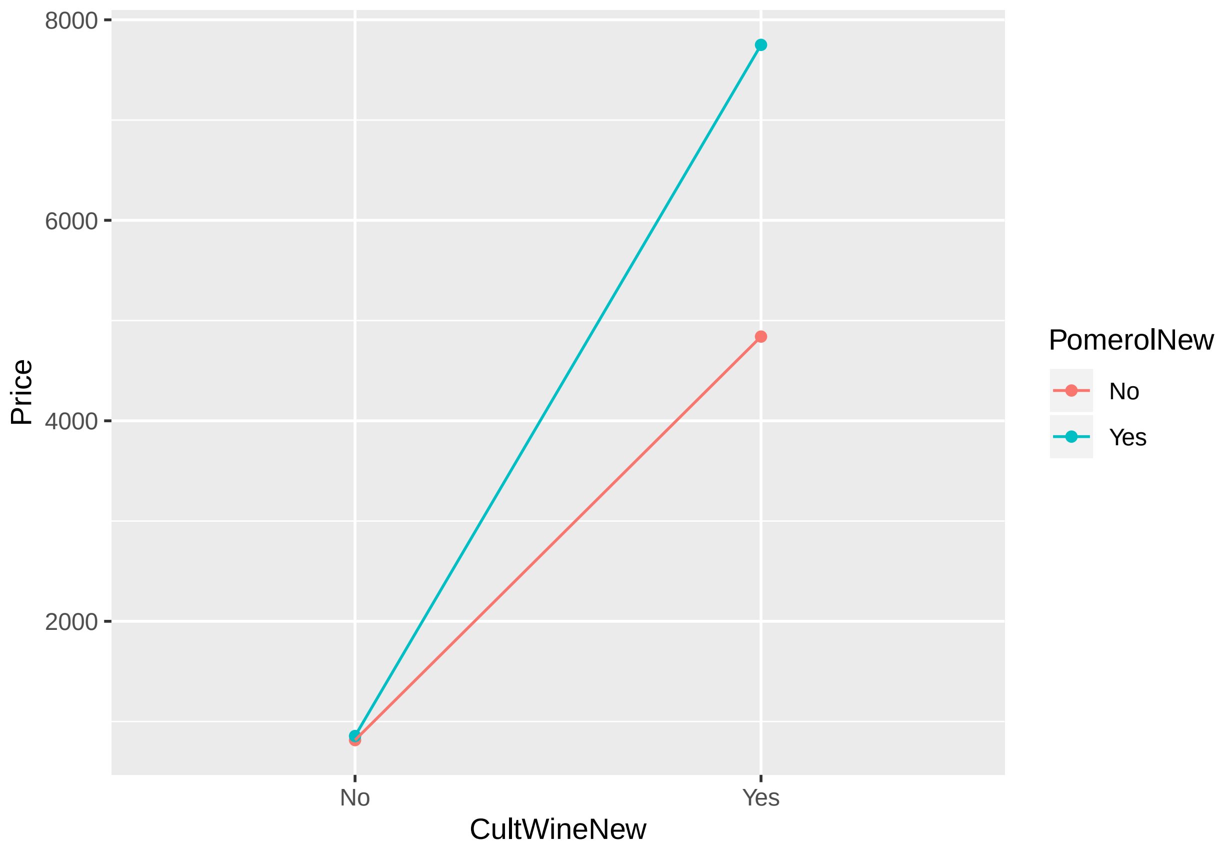

```
## FirstGrowthNew & PomerolNew
p6 <- ggplot() + aes(x = FirstGrowthNew, y = Price, color = PomerolNew,
               group = PomerolNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p6 # Do NOT add
```



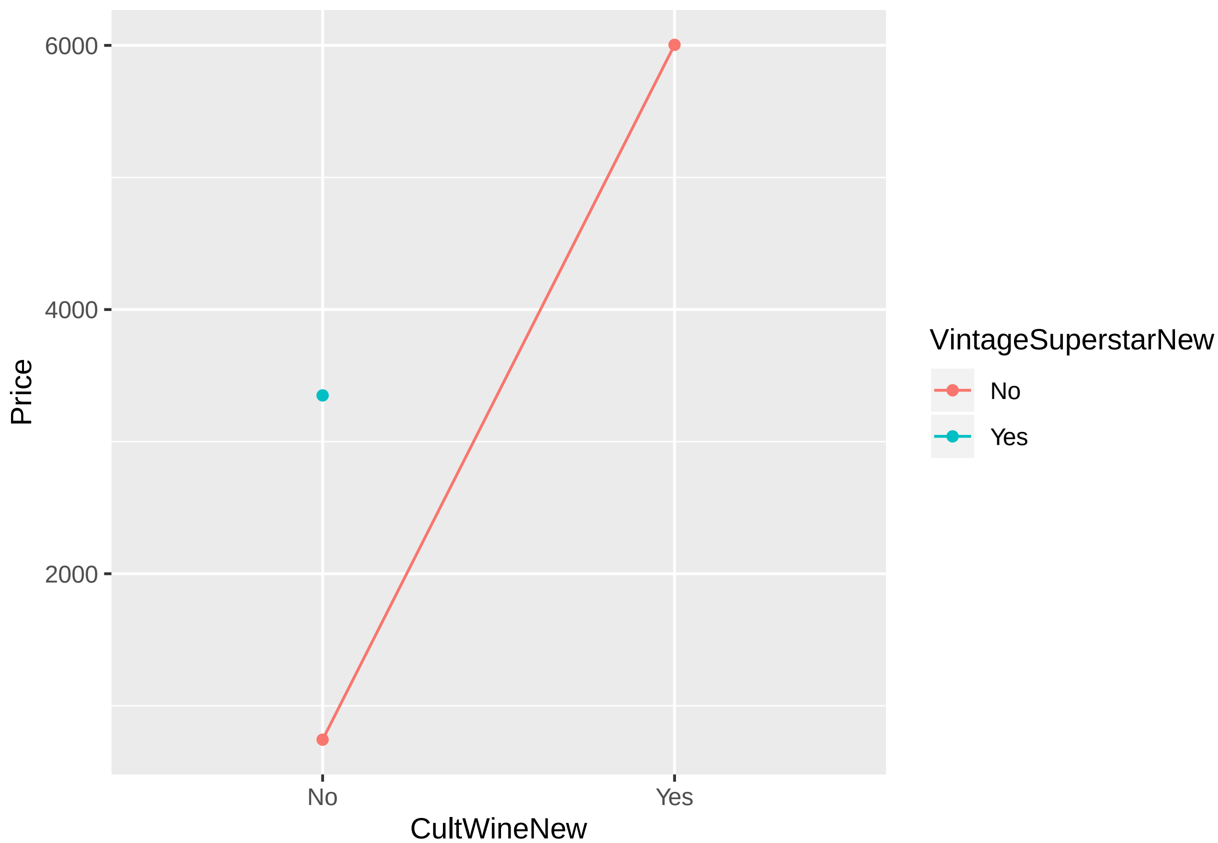

```
## FirstGrowthNew & VintageSuperstarNew
p7 <- ggplot() + aes(x = FirstGrowthNew, y = Price, color = VintageSuperstarNew,
               group = VintageSuperstarNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p7 # Do add
```



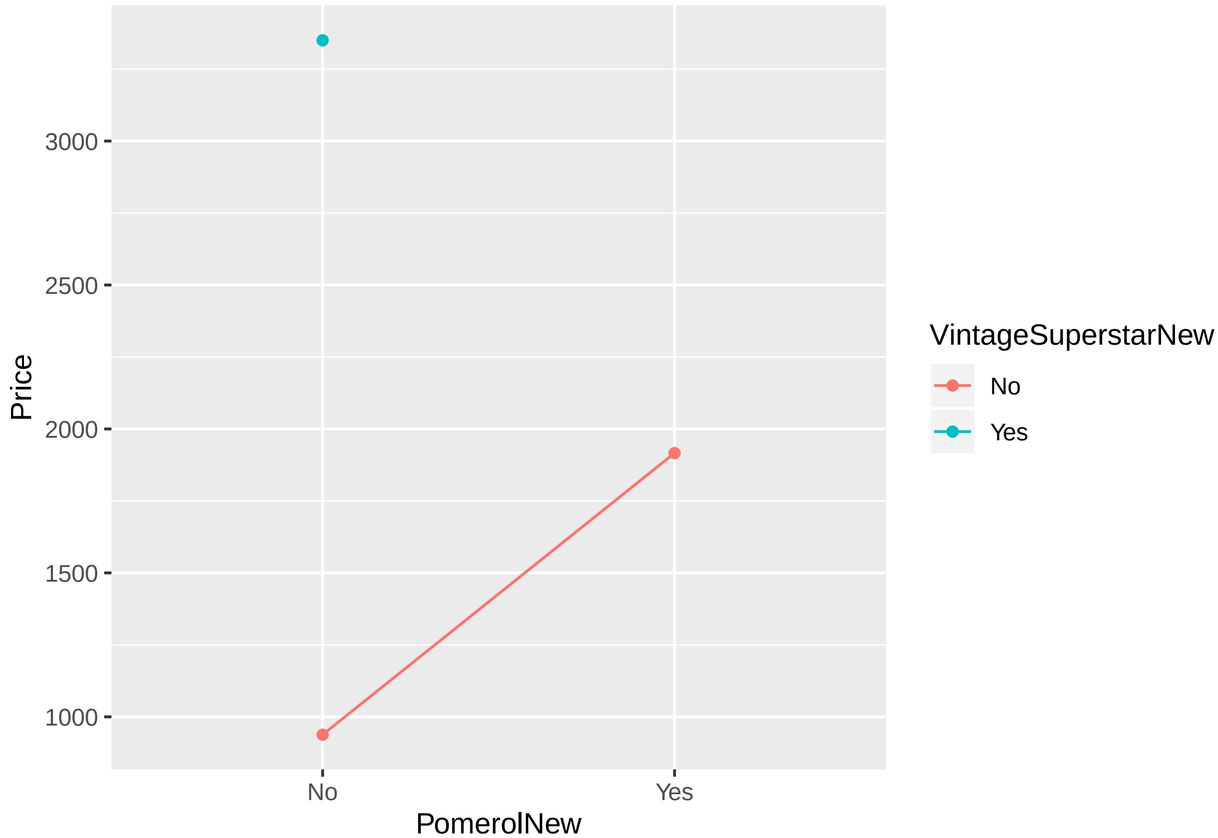

```
## CultWineNew & PomerolNew
p8 <- ggplot() + aes(x = CultWineNew, y = Price, color = PomerolNew,
               group = PomerolNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p8 # Do add
```




```r
## CultWineNew & VintageSuperstarNew
p9 <- ggplot() + aes(x = CultWineNew, y = Price, color = VintageSuperstarNew,
                     group = VintageSuperstarNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p9 # Do NOT add
```




```
## PomerolNew & VintageSuperstarNew
p10 <- ggplot() + aes(x = PomerolNew, y = Price, color = VintageSuperstarNew,
               group = VintageSuperstarNew) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line")
p10 # Do NOT add
```



I would add p5 and p8 to my model. They are FirstGrowthNew & CultWineNew (p5) and CultWineNew & PomerolNew (p8).

2E) Create a new MLR using the categorical predictors and the significant pairwise interactions (call it mcat2). Check summary, anova, and diagnostics of the model.

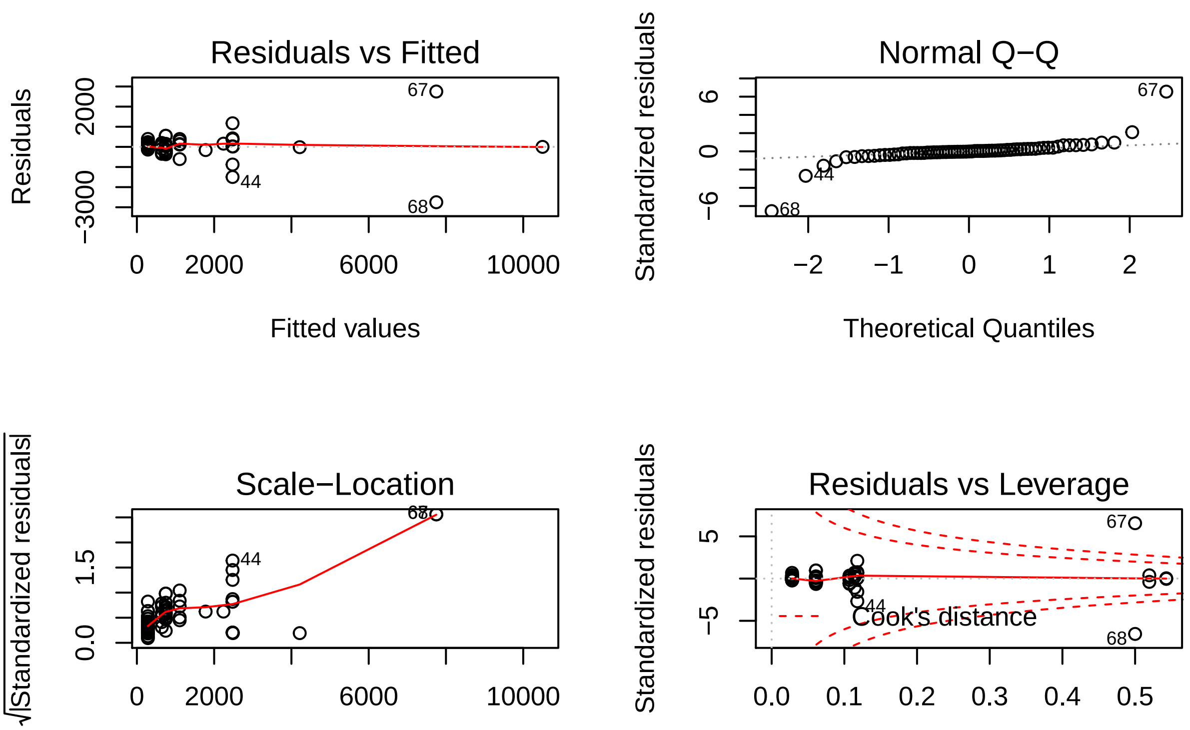
```r
mcat2 <- lm(Price ~ P95andAboveNew + FirstGrowthNew + CultWineNew +
              PomerolNew + VintageSuperstarNew +
              FirstGrowthNew:CultWineNew +
              CultWineNew:PomerolNew)

summary(mcat2)
```

```
##
## Call:
## lm(formula = Price ~ P95andAboveNew + FirstGrowthNew + CultWineNew +
##     PomerolNew + VintageSuperstarNew + FirstGrowthNew:CultWineNew +
##     CultWineNew:PomerolNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2750.0  -105.1    -7.6   136.8  2750.0
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                        285.10      99.46   2.866 0.005612 **
## P95andAboveNewYes                  461.96     165.92   2.784 0.007047 **
## FirstGrowthNewYes                 1729.83     245.86   7.036 1.61e-09 ***
## CultWineNewYes                    1493.92     429.99   3.474 0.000924 ***
## PomerolNewYes                      360.38     200.85   1.794 0.077503 .
## VintageSuperstarNewYes            1738.03     438.71   3.962 0.000190 ***
## FirstGrowthNewYes:CultWineNewYes  6529.19     759.75   8.594 2.91e-12 ***
## CultWineNewYes:PomerolNewYes      5148.64     628.40   8.193 1.47e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 593.5 on 64 degrees of freedom
## Multiple R-squared:  0.9106, Adjusted R-squared:  0.9008
## F-statistic: 93.15 on 7 and 64 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(mcat2)
```

```
## Warning: not plotting observations with leverage one:
##   8

## Warning: not plotting observations with leverage one:
##   8
```




```r
anova(mcat2)
```

```
## Analysis of Variance Table
##
```

```
## Response: Price
##                           Df    Sum Sq   Mean Sq F value    Pr(>F)    
## P95andAboveNew             1  58316213  58316213 165.582 < 2.2e-16 ***
## FirstGrowthNew             1  23288652  23288652  66.125 1.889e-11 ***
## CultWineNew                1 102095811 102095811 289.888 < 2.2e-16 ***
## PomerolNew                 1   5174276   5174276  14.692 0.0002914 ***
## VintageSuperstarNew        1   4784088   4784088  13.584 0.0004718 ***
## FirstGrowthNew:CultWineNew  1  12349678  12349678  35.065 1.379e-07 ***
## CultWineNew:PomerolNew     1  23642620  23642620  67.130 1.471e-11 ***
## Residuals                 64  22540183    352190                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see the interaction effects are significant in our model, along with all our other predictors still. It also has a much higher R-squared of 0.9106. From the diagnostic plot, The Residuals vs. Fitted and Residuals vs. Leverage have both their assumptions satisfied for assumption of linearity. The Normal Q-Q and Scale-Location plots do NOT have their assumptions satisfied though. From our anova F-test, we can see they all significant and REJECT our Null Hypothesis.

2F) Conduct a Partial F-test between mcat and mcat2. What do you conclude?

anova(mcat, mcat2)

We can conclude that our mcat2 model is better than our mcat model. It is significant when conducting the partial F-test between mcat and mcat2. We would REJECT the Null Hypothesis. Therefore, I would choose the mcat2 model.

Question (3):

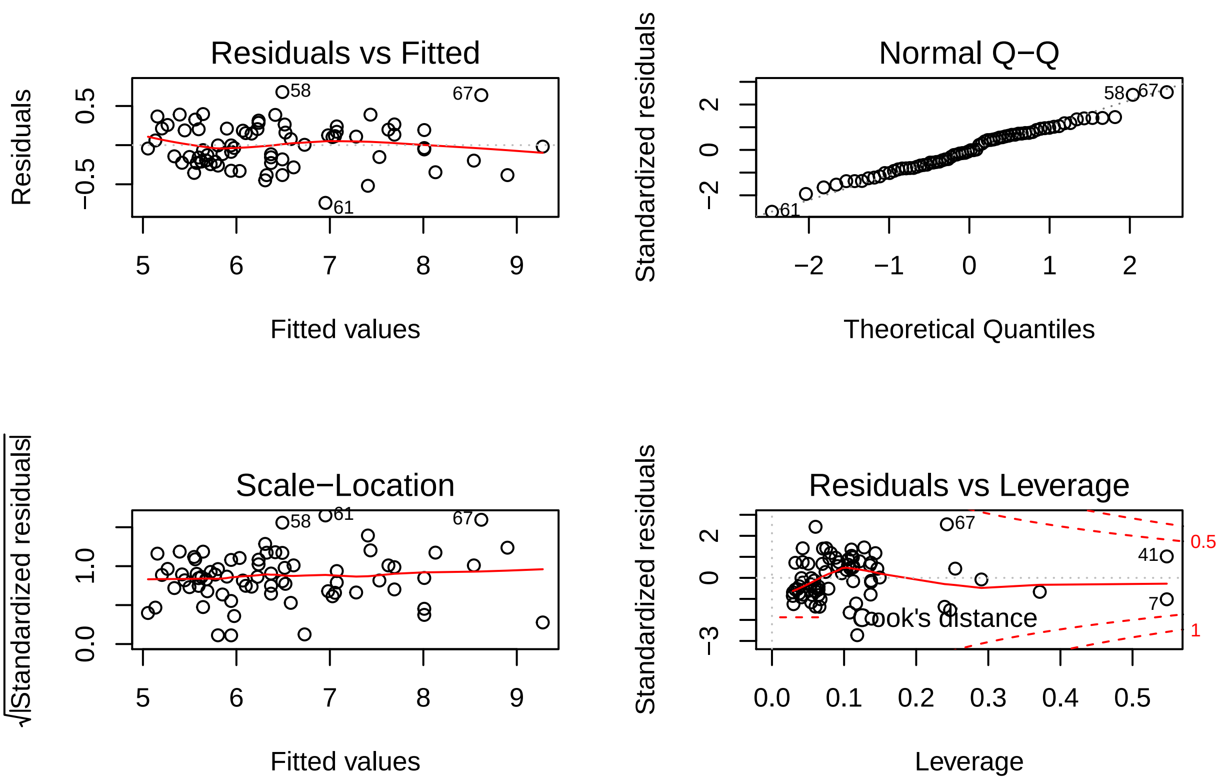
3A) Create a MLR (call it m1) using the suggested transformation on the numerical variables in the data along with the categorical predictors listed in your MLR mcat (No interaction terms). A total of 7 predictors. Check summary, anova, vif and diagnostics of the model.

```r
m1 <- lm(log(Price) ~ log(ParkerPoints) +
               log(CoatesPoints) + P95andAboveNew + FirstGrowthNew +
               CultWineNew + PomerolNew + VintageSuperstarNew, data = Bordeaux)

summary(m1)
```

```
## 
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) + 
##     P95andAboveNew + FirstGrowthNew + CultWineNew + PomerolNew + 
##     VintageSuperstarNew, data = Bordeaux)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -0.7379 -0.2008 -0.0111  0.1989  0.6784 
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           -51.14156    8.98557  -5.692 3.39e-07 ***
## log(ParkerPoints)      11.58862    2.06763   5.605 4.74e-07 ***
```

```
## log(CoatesPoints)        1.62053    0.61154    2.650  0.01013 *
## P95andAboveNewYes        0.10055    0.13697    0.734  0.46556
## FirstGrowthNewYes        0.86970    0.12524    6.944 2.33e-09 ***
## CultWineNewYes           1.35317    0.14569    9.288 1.78e-13 ***
## PomerolNewYes            0.53644    0.09366    5.727 2.95e-07 ***
## VintageSuperstarNewYes   0.61590    0.22067    2.791  0.00692 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2883 on 64 degrees of freedom
## Multiple R-squared:  0.9278, Adjusted R-squared:  0.9199
## F-statistic: 117.5 on 7 and 64 DF,  p-value: < 2.2e-16
```

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




```r
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                        Df Sum Sq Mean Sq  F value    Pr(>F)
## log(ParkerPoints)       1 55.603  55.603 668.9717 < 2.2e-16 ***
## log(CoatesPoints)       1  0.550   0.550   6.6202 0.0124118 *
## P95andAboveNew          1  0.129   0.129   1.5557 0.2168397
## FirstGrowthNew          1  1.383   1.383  16.6392 0.0001276 ***
## CultWineNew             1  7.578   7.578  91.1718 6.308e-14 ***
```

```
## PomerolNew           1  2.500   2.500  30.0807 7.541e-07 ***
## VintageSuperstarNew  1  0.647   0.647   7.7897 0.0069176 **
## Residuals           64  5.319   0.083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From our newest model (m1), we can see from the summary that all the predictors are significant, except for P95andAboveNew. Our model has a R-squared of 0.9278 that is really good for fitting our model. From our diagnostic plots, all of our assumptions (linearity, normality, equal variance, leverages) are satisfied. From our anova test, we can see that all the F-tests on our predictors are significant, except for P95andAboveNew. Therefore, we would REJECT the Null Hypothesis, except for P95andAboveNew.

3B) Create another MLR (call it mfull) using the suggested transformation on the numerical variables in the data along with the categorical predictors listed in your MLR mcat with the significant interaction terms). A total of 7 predictors. Check summary, anova, and diagnostics of the model.

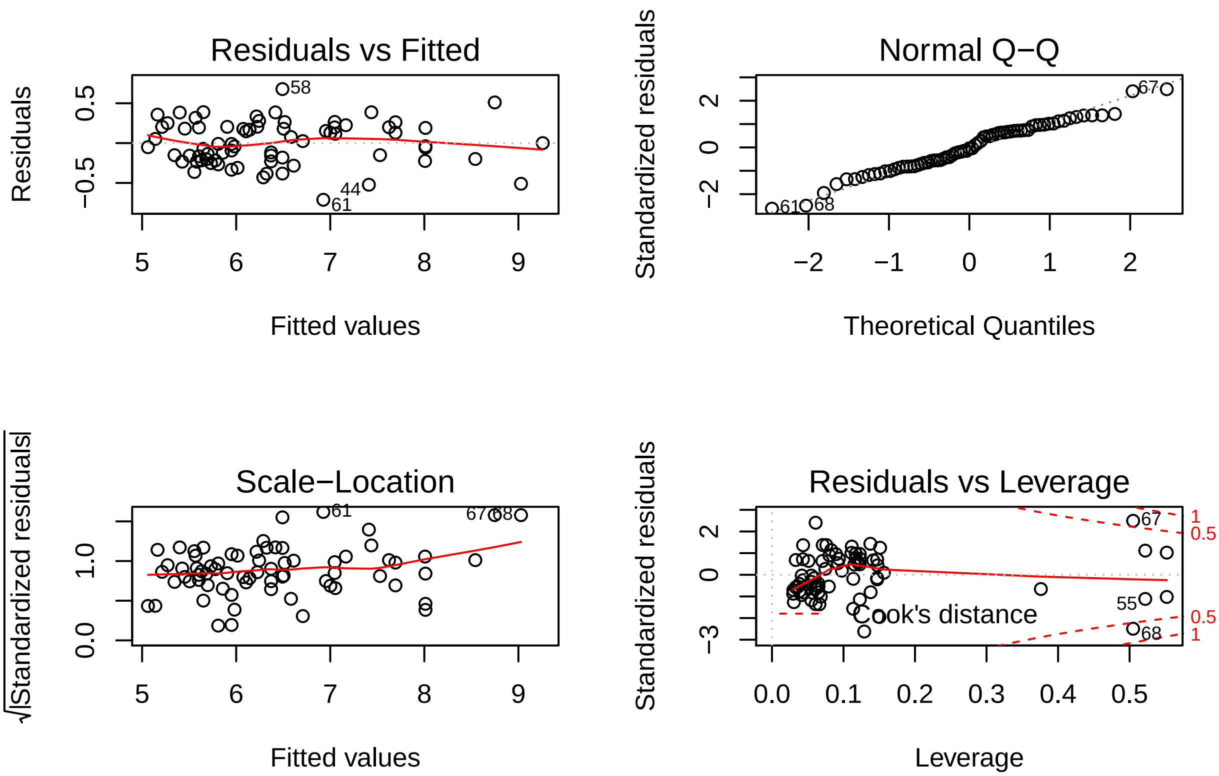
```r
mfull <- lm(log(Price) ~ log(ParkerPoints) +
              log(CoatesPoints) + P95andAboveNew + FirstGrowthNew +
              CultWineNew + PomerolNew + VintageSuperstarNew +
              FirstGrowthNew:CultWineNew +
              CultWineNew:PomerolNew,
            data = Bordeaux)

summary(mfull)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) +
##     P95andAboveNew + FirstGrowthNew + CultWineNew + PomerolNew +
##     VintageSuperstarNew + FirstGrowthNew:CultWineNew + CultWineNew:PomerolNew,
##     data = Bordeaux)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71175 -0.20578 -0.01028  0.20030  0.67822
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -50.94429    9.08776  -5.606 5.10e-07 ***
## log(ParkerPoints)            11.53989    2.09214   5.516 7.20e-07 ***
## log(CoatesPoints)             1.63100    0.61811   2.639  0.01051 *
## P95andAboveNewYes             0.09584    0.13836   0.693  0.49110
## FirstGrowthNewYes             0.87269    0.13205   6.609 1.02e-08 ***
## CultWineNewYes                1.23195    0.21643   5.692 3.66e-07 ***
## PomerolNewYes                 0.50935    0.09924   5.132 3.06e-06 ***
## VintageSuperstarNewYes        0.61583    0.22287   2.763  0.00753 **
## FirstGrowthNewYes:CultWineNewYes   0.10088    0.37322   0.270  0.78784
## CultWineNewYes:PomerolNewYes       0.27747    0.30821   0.900  0.37146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.291 on 62 degrees of freedom
## Multiple R-squared:  0.9288, Adjusted R-squared:  0.9184
```

```
## F-statistic: 89.83 on 9 and 62 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(mfull)
```

```
## Warning: not plotting observations with leverage one:
##   8

## Warning: not plotting observations with leverage one:
##   8
```




```r
anova(mfull)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                           Df Sum Sq Mean Sq  F value    Pr(>F)
## log(ParkerPoints)          1 55.603  55.603 656.5983 < 2.2e-16 ***
## log(CoatesPoints)          1  0.550   0.550   6.4977 0.0132900 *
## P95andAboveNew             1  0.129   0.129   1.5269 0.2212347
## FirstGrowthNew             1  1.383   1.383  16.3314 0.0001492 ***
## CultWineNew                1  7.578   7.578  89.4855 1.220e-13 ***
## PomerolNew                 1  2.500   2.500  29.5243 9.838e-07 ***
## VintageSuperstarNew        1  0.647   0.647   7.6456 0.0074867 **
## FirstGrowthNew:CultWineNew 1  0.000   0.000   0.0058 0.9396323
## CultWineNew:PomerolNew     1  0.069   0.069   0.8105 0.3714645
## Residuals                 62  5.250   0.085
## ---
```

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

From our newest model (mfull), we can see from the summary that all the predictors are significant, except for P95andAboveNew and our interaction effects. Our model has a R-squared of 0.9288 that is really good for fitting our model. From our diagnostic plots, the Residual vs. Fitted plot and Residuals vs. Leverage plot satisfied the assumptions. Although, the Normal Q-Q is less normal, and the Scale-Location plot is less horizontal. From our anova test, we can see that all the F-tests on our predictors are significant, except for P95andAboveNew and our interaction effects. Therefore, we would REJECT the Null Hypothesis, except for P95andAboveNew and our interaction effects.

3C) Interpret the y-intercept and all the partial slopes in your mfull MLR.

m1 model: Our intercept is the average "No" responses from all our predictors. For every 1 unit of increase of Price, we see, on average, an increase by 11.58862 units for log(ParkerPoints). For every 1 unit of increase of Price, we see, on average, an increase by 1.62053 units for log(CoatesPoints). For every 1 unit of increase of Price, the slope for P95andAboveNewYes is 0.10055, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for FirstGrowthNewYes is 0.86970, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for CultWineNewYes is 1.35317, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for PomerolNewYes is 0.53644, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for VintageSuperstarNewYes is 0.61590, which is the average "Yes" responses.

```r
summary(m1)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) +
##     P95andAboveNew + FirstGrowthNew + CultWineNew + PomerolNew +
##     VintageSuperstarNew, data = Bordeaux)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7379 -0.2008 -0.0111  0.1989  0.6784
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -51.14156    8.98557  -5.692 3.39e-07 ***
## log(ParkerPoints)       11.58862    2.06763   5.605 4.74e-07 ***
## log(CoatesPoints)        1.62053    0.61154   2.650  0.01013 *
## P95andAboveNewYes        0.10055    0.13697   0.734  0.46556
## FirstGrowthNewYes        0.86970    0.12524   6.944 2.33e-09 ***
## CultWineNewYes           1.35317    0.14569   9.288 1.78e-13 ***
## PomerolNewYes            0.53644    0.09366   5.727 2.95e-07 ***
## VintageSuperstarNewYes   0.61590    0.22067   2.791  0.00692 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2883 on 64 degrees of freedom
## Multiple R-squared:  0.9278, Adjusted R-squared:  0.9199
## F-statistic: 117.5 on 7 and 64 DF,  p-value: < 2.2e-16
```

```r
### log(Price) = Intercept + log(ParkerPoints)*(x1) + log(CoatesPoints)*(x2) +
###         P95andAboveNewYes*(x3) + FirstGrowthNewYes*(x4) +
```

```
###  CultWineNewYes*(x5) + PomerolNewYes*(x6) + VintageSuperstarNewYes*(x7) -->

### log(Price) = -51.14156  + 11.58862*(x1) + 1.62053 *(x2) + 0.10055*(x3) +
###         0.86970*(x4) + 1.35317*(x5) + 0.53644*(x6) + 0.61590*(x7)

### The slopes are as follows:
### log(ParkerPoints) --> 11.58862
### log(CoatesPoints) --> 1.62053
### P95andAboveNewYes --> 0.10055
### FirstGrowthNewYes --> 0.86970
### CultWineNewYes --> 1.35317
### PomerolNewYes --> 0.53644
### VintageSuperstarNewYes --> 0.61590
```

mfull model: Our intercept is the average "No" responses from all our predictors. For every 1 unit of increase of Price, we see, on average, an increase by 11.53989 units for log(ParkerPoints). For every 1 unit of increase of Price, we see, on average, an increase by 1.63100 units for log(CoatesPoints). For every 1 unit of increase of Price, the slope for P95andAboveNewYes is 0.09584, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for FirstGrowthNewYes is 0.87269, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for CultWineNewYes is 1.23195, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for PomerolNewYes is 0.50935, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for VintageSuperstarNewYes is 0.61583, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for our interaction effect of FirstGrowthNewYes:CultWineNewYes is 0.10088, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for our interaction effect of CultWineNewYes:PomerolNewYes is 0.27747, which is the average "Yes" responses.

```
summary(mfull)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) +
##     P95andAboveNew + FirstGrowthNew + CultWineNew + PomerolNew +
##     VintageSuperstarNew + FirstGrowthNew:CultWineNew + CultWineNew:PomerolNew,
##     data = Bordeaux)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71175 -0.20578 -0.01028  0.20030  0.67822
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -50.94429    9.08776  -5.606 5.10e-07 ***
## log(ParkerPoints)                 11.53989    2.09214   5.516 7.20e-07 ***
## log(CoatesPoints)                  1.63100    0.61811   2.639  0.01051 *
## P95andAboveNewYes                  0.09584    0.13836   0.693  0.49110
## FirstGrowthNewYes                  0.87269    0.13205   6.609 1.02e-08 ***
## CultWineNewYes                     1.23195    0.21643   5.692 3.66e-07 ***
## PomerolNewYes                      0.50935    0.09924   5.132 3.06e-06 ***
## VintageSuperstarNewYes             0.61583    0.22287   2.763  0.00753 **
## FirstGrowthNewYes:CultWineNewYes   0.10088    0.37322   0.270  0.78784
```

```
## CultWineNewYes:PomerolNewYes       0.27747    0.30821   0.900  0.37146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.291 on 62 degrees of freedom
## Multiple R-squared:  0.9288, Adjusted R-squared:  0.9184
## F-statistic: 89.83 on 9 and 62 DF,  p-value: < 2.2e-16
```

```
### log(Price) = Intercept + log(ParkerPoints)*(x1) + log(CoatesPoints)*(x2) +
###         P95andAboveNewYes*(x3) + FirstGrowthNewYes*(x4) +
###  CultWineNewYes*(x5) + PomerolNewYes*(x6) + VintageSuperstarNewYes*(x7)
###  FirstGrowthNewYes:CultWineNewYes*(x8) + CultWineNewYes:PomerolNewYes*(x9)
###                          -->

### log(Price) = -50.94429  + 11.53989*(x1) + 1.63100 *(x2) + 0.09584*(x3) +
###         0.87269*(x4) + 1.23195*(x5) + 0.50935*(x6) + 0.61583*(x7) +
###         0.10088*(x8) + 0.27747*(x9)

### The slopes are as follows:
### log(ParkerPoints) --> 11.53989
### log(CoatesPoints) --> 1.63100
### P95andAboveNewYes --> 0.09584
### FirstGrowthNewYes --> 0.87269
### CultWineNewYes --> 1.23195
### PomerolNewYes --> 0.50935
### VintageSuperstarNewYes --> 0.61583
### FirstGrowthNewYes:CultWineNewYes --> 0.10088
### CultWineNewYes:PomerolNewYes --> 0.27747
```

3D) Which of the predictors need to be dropped from mfull?

```
anova(m1, mfull)
```

```
## Analysis of Variance Table
##
## Model 1: log(Price) ~ log(ParkerPoints) + log(CoatesPoints) + P95andAboveNew +
##     FirstGrowthNew + CultWineNew + PomerolNew + VintageSuperstarNew
## Model 2: log(Price) ~ log(ParkerPoints) + log(CoatesPoints) + P95andAboveNew +
##     FirstGrowthNew + CultWineNew + PomerolNew + VintageSuperstarNew +
##     FirstGrowthNew:CultWineNew + CultWineNew:PomerolNew
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     64 5.3195
## 2     62 5.2504  2  0.069123 0.4081 0.6667
```

We should drop P95andAboveNewYes, FirstGrowthNewYes:CultWineNewYes, and CultWineNewYes:PomerolNewYes variables from our model (mfull).

3E) Create a MLR (mred). Conduct partial F-test.

```
mred <- lm(log(Price) ~ log(ParkerPoints) +
             log(CoatesPoints) + FirstGrowthNew +
             CultWineNew + PomerolNew + VintageSuperstarNew,
           data = Bordeaux)
```

```r
anova(mred, mfull)
```

```
## Analysis of Variance Table
##
## Model 1: log(Price) ~ log(ParkerPoints) + log(CoatesPoints) + FirstGrowthNew +
##     CultWineNew + PomerolNew + VintageSuperstarNew
## Model 2: log(Price) ~ log(ParkerPoints) + log(CoatesPoints) + P95andAboveNew +
##     FirstGrowthNew + CultWineNew + PomerolNew + VintageSuperstarNew +
##     FirstGrowthNew:CultWineNew + CultWineNew:PomerolNew
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     65 5.3643
## 2     62 5.2504  3   0.11392 0.4484 0.7193
```

We cannot conclude that our mred model is better than our mfull model. It is NOT significant when conducting the partial F-test between mred and mfull. We would NOT REJECT the Null Hypothesis. Therefore, I cannot say that the full model is better than the reduced model.

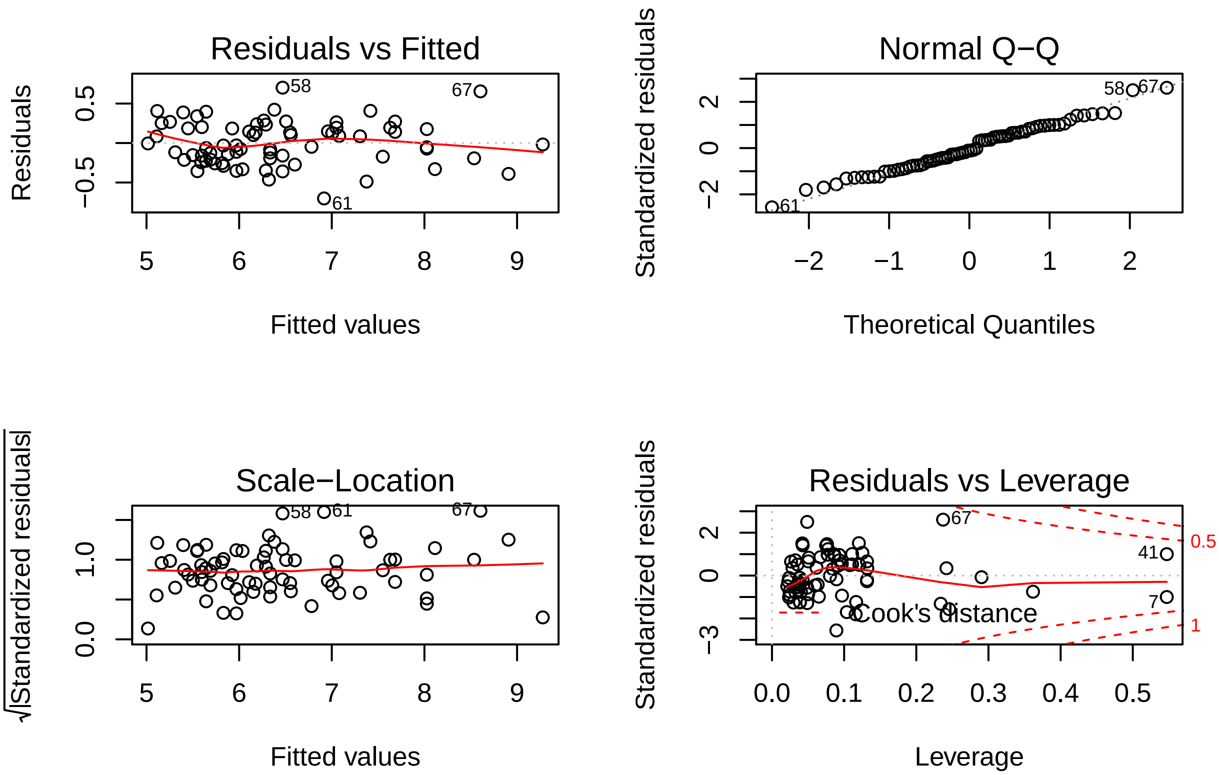
Question (4)

4A) State your final MLR based on your answers to the previous three questions.

```r
summary(mred)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) +
##     FirstGrowthNew + CultWineNew + PomerolNew + VintageSuperstarNew,
##     data = Bordeaux)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70185 -0.19752 -0.03061  0.19347  0.70118
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -56.47547    5.26798 -10.721 5.20e-16 ***
## log(ParkerPoints)       12.78432    1.26915  10.073 6.66e-15 ***
## log(CoatesPoints)        1.60447    0.60898   2.635  0.01052 *
## FirstGrowthNewYes        0.86149    0.12430   6.931 2.30e-09 ***
## CultWineNewYes           1.33601    0.14330   9.323 1.34e-13 ***
## PomerolNewYes            0.53619    0.09333   5.745 2.64e-07 ***
## VintageSuperstarNewYes   0.59470    0.21800   2.728  0.00819 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2873 on 65 degrees of freedom
## Multiple R-squared:  0.9272, Adjusted R-squared:  0.9205
## F-statistic:   138 on 6 and 65 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2))
plot(mred)
```




```r
anova(mred)
```

```
## Analysis of Variance Table
##
## Response: log(Price)
##                     Df Sum Sq Mean Sq  F value    Pr(>F)
## log(ParkerPoints)    1 55.603  55.603 673.7508 < 2.2e-16 ***
## log(CoatesPoints)    1  0.550   0.550   6.6675  0.012081 *
## FirstGrowthNew       1  1.436   1.436  17.4013 9.157e-05 ***
## CultWineNew          1  7.639   7.639  92.5675 4.045e-14 ***
## PomerolNew           1  2.504   2.504  30.3354 6.659e-07 ***
## VintageSuperstarNew  1  0.614   0.614   7.4420  0.008186 **
## Residuals           65  5.364   0.083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I would choose the reduce model (mred) as my final MLR. Not only does (mred) have diagnostic plots that satisfy all the assumptions, but all the predictors being used have significance. We have the highest R-squared of 92.72% of our model explained. It is also shown not to be overfitting nor is it underfitting for our model. Our F-test has a smaller p-value than 0.05, so we would REJECt our NUll Hypothesis. Therefore, our reduced model (mred) is my choice of MLR to use.

4B) Interpret the y-intercept and all the partial slopes in your final MLR.

The intercept tells us the average "No" responses from all our predictors. For every 1 unit of increase of Price, we see, on average, an increase by 12.78432 units for log(ParkerPoints). For every 1 unit of increase of Price, we see, on average, an increase by 1.60447 units for log(CoatesPoints). For every 1 unit of increase of

Price, the slope for FirstGrowthNewYes is 0.86149, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for CultWineNewYes is 1.33601, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for PomerolNewYes is 0.53619, which is the average "Yes" responses. For every 1 unit of increase of Price, the slope for VintageSuperstarNewYes is 0.59470, which is the average "Yes" responses.

```r
summary(mred)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(ParkerPoints) + log(CoatesPoints) +
##     FirstGrowthNew + CultWineNew + PomerolNew + VintageSuperstarNew,
##     data = Bordeaux)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70185 -0.19752 -0.03061  0.19347  0.70118
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -56.47547    5.26798 -10.721 5.20e-16 ***
## log(ParkerPoints)        12.78432    1.26915  10.073 6.66e-15 ***
## log(CoatesPoints)         1.60447    0.60898   2.635  0.01052 *
## FirstGrowthNewYes         0.86149    0.12430   6.931 2.30e-09 ***
## CultWineNewYes            1.33601    0.14330   9.323 1.34e-13 ***
## PomerolNewYes             0.53619    0.09333   5.745 2.64e-07 ***
## VintageSuperstarNewYes   0.59470    0.21800   2.728  0.00819 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2873 on 65 degrees of freedom
## Multiple R-squared:  0.9272, Adjusted R-squared:  0.9205
## F-statistic:   138 on 6 and 65 DF,  p-value: < 2.2e-16
```

```r
### log(Price) = Intercept + log(ParkerPoints)*(x1) + log(CoatesPoints)*(x2) +
###          FirstGrowthNewYes*(x3) + CultWineNewYes*(x4) + PomerolNewYes*(x5) +
###          VintageSuperstarNewYes*(x6) -->

### log(Price) = -56.47547  + 12.78432*(x1) + 1.60447 *(x2) + 0.86149*(x3) +
###         1.33601*(x4) + 0.53619*(x5) + 0.59470*(x6)

### The slopes are as follows:
### log(ParkerPoints) --> 12.78432
### log(CoatesPoints) --> 1.60447
### FirstGrowthNewYes --> 0.86149
### CultWineNewYes --> 1.33601
### PomerolNewYes --> 0.53619
### VintageSuperstarNewYes --> 0.59470
```

4C) Identify the Unusually highly priced wines and the Unusually lowly priced wines based on your final model.

```r
leverages <- hatvalues(mred)
```

```r
## Detects leverages:
which(leverages >= 2 * mean(leverages))
```

```
##  7  8 41 53 55 59 67 68
##  7  8 41 53 55 59 67 68
```

```r
### Leverages are: 7, 8, 41, 53, 55, 59, 67, 68

## Detects Outliers:
which(abs(rstandard(mred)) >= 2)
```

```
## 58 61 67
## 58 61 67
```

```r
### Outliers are: 58, 61, 67
```

The leverage points are 7, 8, 41, 53, 55, 59, 67, 68. The outlier points are 58, 61, 67. These are our unusually high/low priced wines.

4D) Identify the wines that can be considered good leverage points in your final MLR.

```r
## Detect Bad leverages:
nrow(Bordeaux) # n = 72, p = 6
```

```
## [1] 72
```

```r
which(leverages >= 2 * mean(leverages) & abs(rstandard(mred)) >= 2)
```

```
## 67
## 67
```

```r
which(leverages >= 2 * (6 + 1)/72 & abs(rstandard(mred)) >= 2)
```

```
## 67
## 67
```

```r
### Bad Leverages is: 67

### Good Leverages are: 7, 8, 41, 53, 55, 59, 68
```

The good leverage points are 7, 8, 41, 53, 55, 59, 68.