

# Homework 2

2023-03-20

Francesco Tedesco 1635101

Oliver Einard Tarragó Boardman 1527541

First of all, we import the libraries that we will need.

```
library(MASS)
library(leaps)
library(olsrr)
```

Next, import the dataset which we are going to study

```
winequality = read.csv("winequality-red.csv")
head(winequality)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70         0.00             1.9      0.076
## 2           7.8             0.88         0.00             2.6      0.098
## 3           7.8             0.76         0.04             2.3      0.092
## 4          11.2             0.28         0.56             1.9      0.075
## 5           7.4             0.70         0.00             1.9      0.076
## 6           7.4             0.66         0.00             1.8      0.075
##   free.sulfur.dioxide total.sulfur.dioxide density    pH sulphates alcohol
## 1                  11                   34 0.9978 3.51      0.56      9.4
## 2                  25                   67 0.9968 3.20      0.68      9.8
## 3                  15                   54 0.9970 3.26      0.65      9.8
## 4                  17                   60 0.9980 3.16      0.58      9.8
## 5                  11                   34 0.9978 3.51      0.56      9.4
## 6                  13                   40 0.9978 3.51      0.56      9.4
##   quality
## 1        5
## 2        5
## 3        5
## 4        6
## 5        5
## 6        5
```

Now, we define one function to visualize data

```
basic_info_of_variable <- function(variable) {

  # This function returns two plots
  # Plot 1, histogram with:
  # - density (green line)
  # - mean (red line)
```

```

# - median (black line)

# Plot 2, boxplot with:
# - mean (red line)
# - median (black line)

par(mfrow = c(1, 2), cex.axis= 0.75, cex.lab= 0.75)

hist(variable, xlab = "", main = "", breaks = 20, probability = TRUE)
lines(density(variable), col = 'green', lwd = 3)
grid(nx = NULL, ny = NULL, col = "gray", lty = "dotted")
abline(v = mean(variable), col='red', lwd = 2.2)
abline(v = median(variable), col='black', lwd = 2.2)

boxplot(variable)
grid(nx = NULL, ny = NULL, col = "gray", lty = "dotted")
segments(x0 = 0.8, y0 = mean(variable),
          x1 = 1.2, y1 = mean(variable),
          col = "red", lwd = 2)
}

```

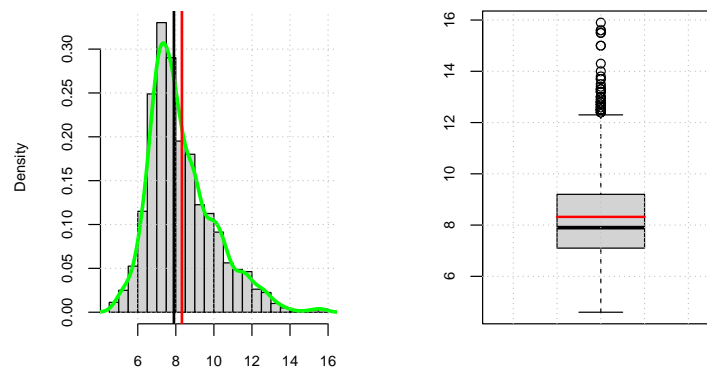
## Exercise 1

Provide details of the chosen dataset. Design models to be analysed for this dataset.

We choose the winequality-red dataset from kaggle, which contains information about several red wines, for each of them we have some measurements which describe the characteristics of the wine, our goal is to predict the wine quality from the given data. So, each characteristic will be a variable, we don't know yet if all variables are needed to make our regression model. The variables are:

- **fixed.acidity:** Indicates the amount of non-volatile acids presents in the wine, these acids doesn't evaporate easily when the wine is heated. Some examples are: tartaric, malic and citric acid...

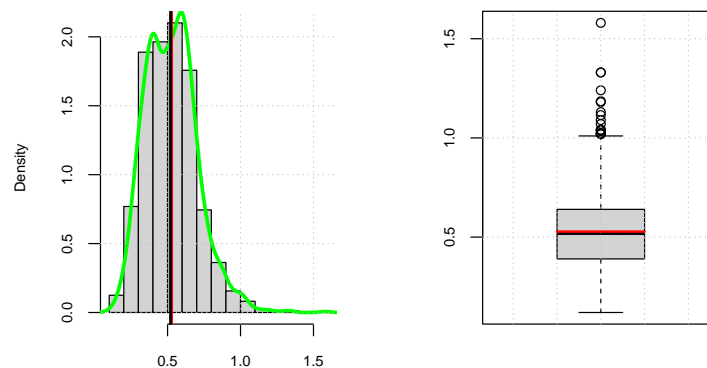
```
basic_info_of_variable(winequality$fixed.acidity)
```



As we can see by the plots, in a considerable number of wines we have around 8 % of fixed acidity, we can also say that find a wine with a fixed acidity higher than 12.1 approximately.

- **volatile.acidity:** The variable refers to the amounts of volatile acids, which unlike the previous acids, these can evaporate easily, one example can be acetic acid. When the amount is too high, the taste of the wine can be unpleasant.

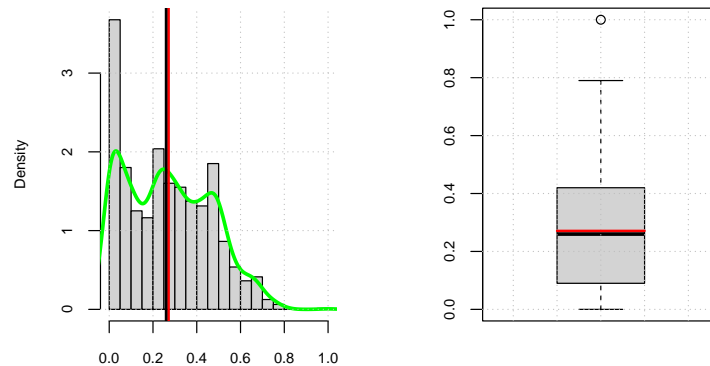
```
basic_info_of_variable(winequality$volatile.acidity)
```



Since higher amounts of volatile acidity are unpleasant, we can see that the majority of wine have volatility acidity around 0.5, which compared to the previous acidity is lower. We can also say that the distribution in this case is more symmetric since the median are and the mean are closer and the shape is also symmetric

- **citric.acid:** The quantity of citric acid in the wine, this acid can be found in some natural fruits, in small portions can help to add 'freshness' and flavor to wines

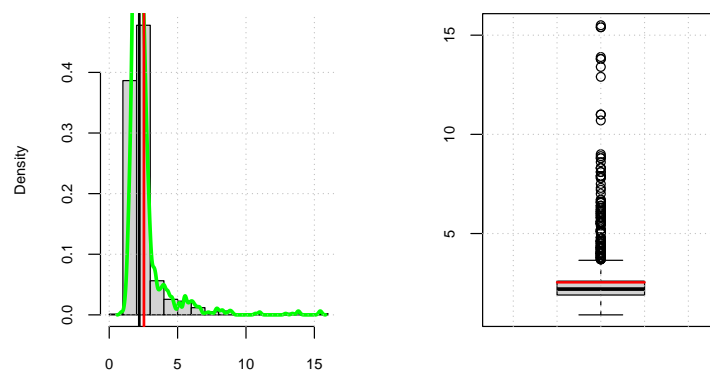
```
basic_info_of_variable(winequality$citric.acid)
```



The citric acid of the wines is a little more dispersed, since more wines have a similar amount of citric acid and the middle quartile of the box plot is wider

- **residual.sugar:** The amount of sugar that the wine has after the fermentation process.

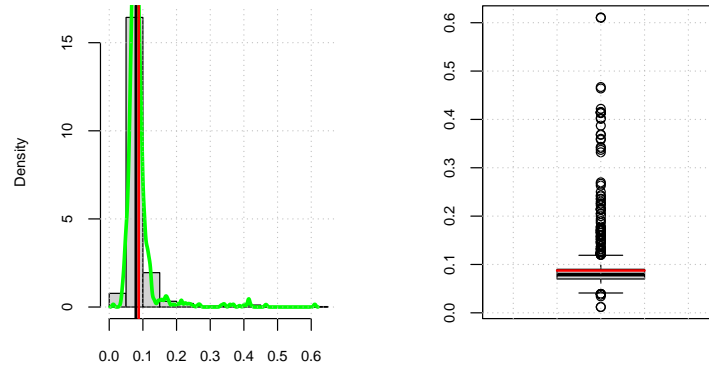
```
basic_info_of_variable(winequality$residual.sugar)
```



Unlike the previous variable, we can see less dispersion since the values of residual sugar are more concentrated at approximately 2.5.

- **chlorides:** Indicates the quantity of salt of the wine

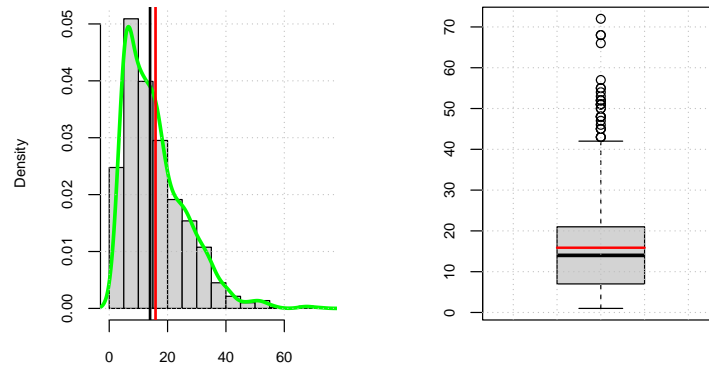
```
basic_info_of_variable(winequality$chlorides)
```



Like the previous variable, we the amount of chlorides of wines isn't so dispersed

- **free.sulfur.dioxide:** It refers to the amount of free sulfur dioxide (SO<sub>2</sub>) of the wine, this type of sulfur dioxide exists in equilibrium between dissolved gas and in its free form (bisulfite ion). In small quantities can help to prevent microbial growth and the oxidation of the wine. (parts per million ppm)

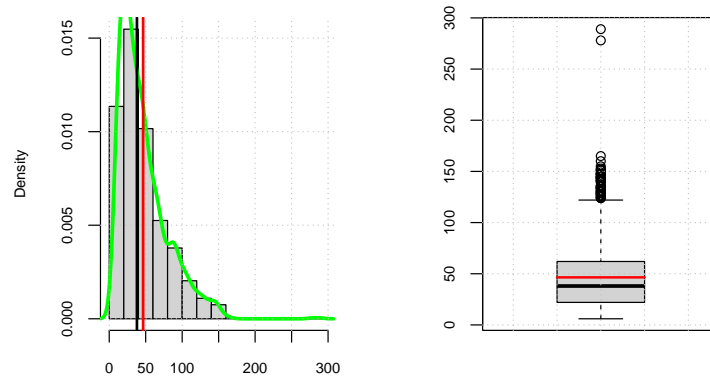
```
basic_info_of_variable(winequality$free.sulfur.dioxide)
```



Here the median is lower than the mean, so in majority of wines the free.sulfur.dioxide portion is below the mean, so there are a minority of higher values for free.sulfur.dioxide that increase the mean.

- **total.sulfur.dioxide:** In this case, we have the total amount of sulfur dioxide which can be found in the wine. It is measured in parts per million (ppm), in small quantities is almost undetectable, can become easy to detect when the concentration is over 50 ppm.

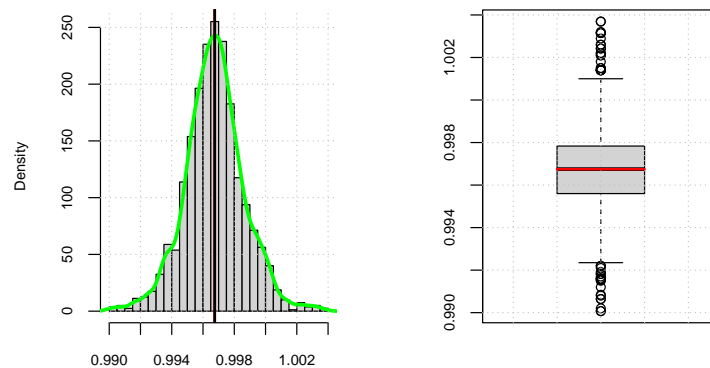
```
basic_info_of_variable(winequality$total.sulfur.dioxide)
```



Thanks to the median, we can say that in majority of wines we can notice the presence of the total sulfur dioxide in the wine, since the median is lower than 50.

- **density:** Indicates the ratio between liquid mass and its volume. In this case, mostly depends on the percent alcohol and sugar presents in the wine.

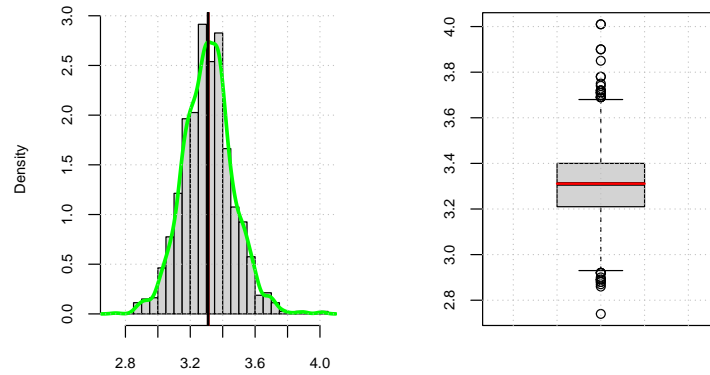
```
basic_info_of_variable(winequality$density)
```



Here, the distribution is very symmetric since the median and mean are almost the same and the shape of the distribution is symmetrical respect to the mean and median line.

- **pH:** This variable indicates the acidity (alkalinity) of the wine. Can take values from 0 to 14, and wines often have a 3-4 pH value.

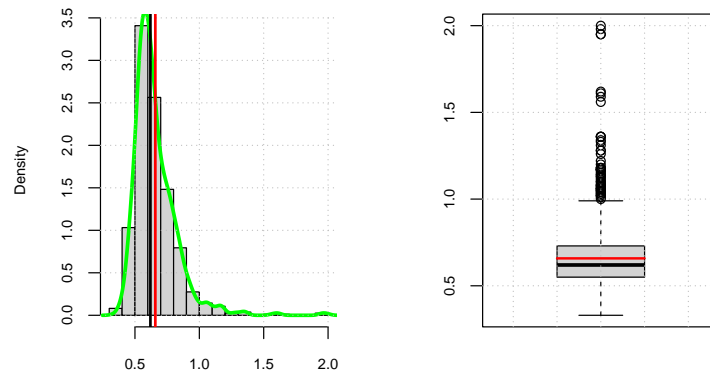
```
basic_info_of_variable(winequality$pH)
```



In this case, the distribution is also symmetrical, for the same reasons we mentioned in the previous variable.

- **sulphates:** It refers to the amount of sulfites present in the wine, this additive can contribute to sulfure dioxide levels, which helps to prevent oxidation and microbial growth.

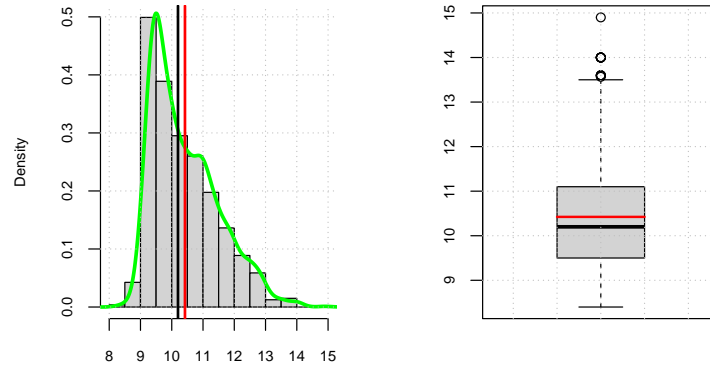
```
basic_info_of_variable(winequality$sulphates)
```



We can see, that we probably have some outliers due to the presence of values above the upper quartile.

- **alcohol:** The alcohol percentage of the wine

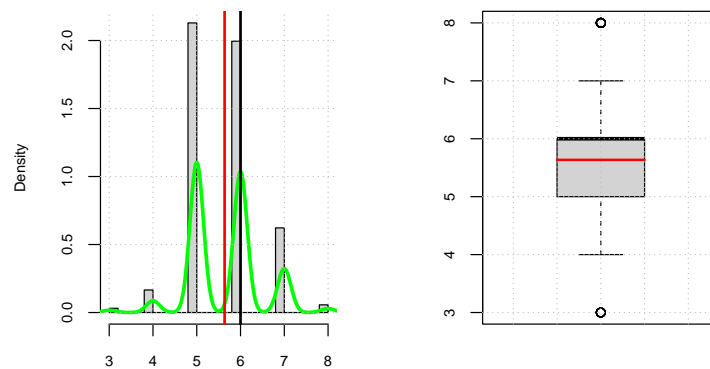
```
basic_info_of_variable(winequality$alcohol)
```



We can see, that all wines have a alcohol and the mean percentage is approximately around 10.4 %.

- **quality:** This is the variable that we want to estimate, consist of a score that can be between 0 and 10, we consider good quality if it's higher than 6.5.

```
basic_info_of_variable(winequality$quality)
```



Since the median is below 6.5, we can say that the majority of wines have a quality which wouldn't be considered "good".

For more details, you can check the following links:

<https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>

<https://archive.ics.uci.edu/ml/datasets/wine+quality>



## Exercise 2

Apply backward selection to find the best fit model using p-value and AIC criteria. Compare the results found by both methods. Do the same for forward selection. Comment on the results.

### backward selection (p-value)

The backward selection algorithm for p-value follow this steps:

1. Fit the model with all the variables
2. Update the model by removing the variable with higher p-value
3. Fit the updated model
4. Repeat steps 2 and 3 until the highest p-value of the updated model is below 0.05.

The following code applies the backward selection method for p-value

```
model_full <- lm(quality~.,data = winequality)
summary(model_full)

##
## Call:
## lm(formula = quality ~ ., data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68911 -0.36652 -0.04699  0.45202  2.02498
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.197e+01  2.119e+01   1.036   0.3002
## fixed.acidity    2.499e-02  2.595e-02   0.963   0.3357
## volatile.acidity -1.084e+00  1.211e-01  -8.948 < 2e-16 ***
## citric.acid     -1.826e-01  1.472e-01  -1.240   0.2150
## residual.sugar   1.633e-02  1.500e-02   1.089   0.2765
## chlorides       -1.874e+00  4.193e-01  -4.470 8.37e-06 ***
## free.sulfur.dioxide 4.361e-03  2.171e-03   2.009   0.0447 *
## total.sulfur.dioxide -3.265e-03  7.287e-04  -4.480 8.00e-06 ***
## density        -1.788e+01  2.163e+01  -0.827   0.4086
## pH             -4.137e-01  1.916e-01  -2.159   0.0310 *
## sulphates       9.163e-01  1.143e-01   8.014 2.13e-15 ***
## alcohol         2.762e-01  2.648e-02  10.429 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.648 on 1587 degrees of freedom
## Multiple R-squared:  0.3606, Adjusted R-squared:  0.3561
## F-statistic: 81.35 on 11 and 1587 DF, p-value: < 2.2e-16
model <- update(model_full, ~.-residual.sugar)
summary(model)

##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##      chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##      density + pH + sulphates + alcohol, data = winequality)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.69219 -0.36820 -0.04617  0.45516  2.02213
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.2233446  17.0262330   0.483  0.62918
## fixed.acidity      0.0128171   0.0234166   0.547  0.58422
## volatile.acidity  -1.0867572   0.1210734  -8.976 < 2e-16 ***
## citric.acid       -0.1740055   0.1469746  -1.184  0.23662
## chlorides        -1.8791757   0.4192830  -4.482 7.93e-06 ***
## free.sulfur.dioxide  0.0046603   0.0021540   2.164  0.03064 *
## total.sulfur.dioxide -0.0032358   0.0007283  -4.443 9.49e-06 ***
## density          -3.8644412  17.3849576  -0.222  0.82412
## pH                -0.4822839   0.1809384  -2.665  0.00777 **
## sulphates         0.8913499   0.1120168   7.957 3.32e-15 ***
## alcohol           0.2906726   0.0229049  12.690 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.648 on 1588 degrees of freedom
## Multiple R-squared:  0.3601, Adjusted R-squared:  0.356
## F-statistic: 89.35 on 10 and 1588 DF,  p-value: < 2.2e-16
```

```
model <- update(model, ~.-density)
summary(model)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + citric.acid +
##      chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##      pH + sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68601 -0.36723 -0.04516  0.45629  2.02723
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.4410923   0.6122514   7.254 6.31e-13 ***
## fixed.acidity      0.0090201   0.0160129   0.563  0.57331
## volatile.acidity  -1.0905804   0.1198096  -9.103 < 2e-16 ***
## citric.acid       -0.1756500   0.1467444  -1.197  0.23149
## chlorides        -1.8893071   0.4166737  -4.534 6.21e-06 ***
## free.sulfur.dioxide  0.0046664   0.0021532   2.167  0.03036 *
## total.sulfur.dioxide -0.0032560   0.0007224  -4.507 7.04e-06 ***
## pH                -0.5022333   0.1570654  -3.198  0.00141 **
## sulphates         0.8872849   0.1104810   8.031 1.86e-15 ***
## alcohol           0.2940206   0.0172514  17.043 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6479 on 1589 degrees of freedom
## Multiple R-squared:  0.3601, Adjusted R-squared:  0.3564
## F-statistic: 99.34 on 9 and 1589 DF,  p-value: < 2.2e-16
```

```
model <- update(model, .~-fixed.acidity)
summary(model)
```

```
##
## Call:
## lm(formula = quality ~ volatile.acidity + citric.acid + chlorides +
##     free.sulfur.dioxide + total.sulfur.dioxide + pH + sulphates +
##     alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.66890 -0.37044 -0.04474  0.45697  2.02363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.6680876   0.4608410  10.129 < 2e-16 ***
## volatile.acidity  -1.0736123   0.1159362  -9.260 < 2e-16 ***
## citric.acid       -0.1295444   0.1217717  -1.064  0.2876
## chlorides        -1.9494185   0.4026906  -4.841 1.42e-06 ***
## free.sulfur.dioxide  0.0047601   0.0021463   2.218  0.0267 *
## total.sulfur.dioxide -0.0033658   0.0006954  -4.840 1.42e-06 ***
## pH               -0.5491501   0.1331350  -4.125 3.90e-05 ***
## sulphates         0.8914283   0.1102122   8.088 1.19e-15 ***
## alcohol           0.2928780   0.0171280  17.099 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6477 on 1590 degrees of freedom
## Multiple R-squared:  0.3599, Adjusted R-squared:  0.3567
## F-statistic: 111.8 on 8 and 1590 DF,  p-value: < 2.2e-16
```

```
model <- update(model, .~-citric.acid)
summary(model)
```

```
##
## Call:
## lm(formula = quality ~ volatile.acidity + chlorides + free.sulfur.dioxide +
##     total.sulfur.dioxide + pH + sulphates + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68918 -0.36757 -0.04653  0.46081  2.02954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.4300987   0.4029168  10.995 < 2e-16 ***
## volatile.acidity  -1.0127527   0.1008429 -10.043 < 2e-16 ***
## chlorides        -2.0178138   0.3975417  -5.076 4.31e-07 ***
## free.sulfur.dioxide  0.0050774   0.0021255   2.389  0.017 *
## total.sulfur.dioxide -0.0034822   0.0006868  -5.070 4.43e-07 ***
## pH               -0.4826614   0.1175581  -4.106 4.23e-05 ***
## sulphates         0.8826651   0.1099084   8.031 1.86e-15 ***
## alcohol           0.2893028   0.0167958  17.225 < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6477 on 1591 degrees of freedom
## Multiple R-squared:  0.3595, Adjusted R-squared:  0.3567
## F-statistic: 127.6 on 7 and 1591 DF,  p-value: < 2.2e-16
```

### backward selection (AIC)

Now, we apply the same selection method, but with AIC measure:

$$AIC = 2k - 2\text{LogLikelihood}$$

Since the k is the number of variables and the log-likelihood measures how well the model fit the data, we can say that we want a lower AIC value, which implies a simpler model with better fit having in count its complexity.

In order to do the backward selection method with AIC, we will follow these steps:

1. Fit the model with all the variables
2. Calculate the AIC for the initial fit
3. For each variable, we get the AIC value for the fitted model without the variable
4. Since we want the AIC of the model lower, we remove the variable which produces the lowest increase at the model
5. Fit the updated model
6. Repeat steps 3, 4 and 5 until the variable which produces the lowest increase at the model reduces the AIC of the last updated model

Here are the steps of the backward selection method based on AIC using the r function stepAIC

```
model_full <- lm(quality~.,data = winequality)
model_backward <- stepAIC(model_full, trace = TRUE, direction = "backward")

## Start:  AIC=-1375.49
## quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
##          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
##          density + pH + sulphates + alcohol
##
##           Df Sum of Sq  RSS   AIC
## - density      1      0.287 666.70 -1376.8
## - fixed.acidity  1      0.389 666.80 -1376.5
## - residual.sugar  1      0.498 666.91 -1376.3
## - citric.acid     1      0.646 667.06 -1375.9
## <none>                        666.41 -1375.5
## - free.sulfur.dioxide  1      1.694 668.10 -1373.4
## - pH                  1      1.957 668.37 -1372.8
## - chlorides           1      8.391 674.80 -1357.5
## - total.sulfur.dioxide  1      8.427 674.84 -1357.4
## - sulphates           1     26.971 693.38 -1314.0
## - volatile.acidity     1     33.620 700.03 -1298.8
## - alcohol              1     45.672 712.08 -1271.5
##
## Step:  AIC=-1376.8
## quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
##          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
```

```

##      pH + sulphates + alcohol
##
##      Df Sum of Sq    RSS      AIC
## - fixed.acidity      1      0.108 666.81 -1378.5
## - residual.sugar     1      0.231 666.93 -1378.2
## - citric.acid        1      0.654 667.35 -1377.2
## <none>                                666.70 -1376.8
## - free.sulfur.dioxide 1      1.829 668.53 -1374.4
## - pH                 1      4.325 671.02 -1368.5
## - total.sulfur.dioxide 1      8.728 675.43 -1358.0
## - chlorides          1      8.761 675.46 -1357.9
## - sulphates          1     27.287 693.98 -1314.7
## - volatile.acidity    1     35.000 701.70 -1297.0
## - alcohol            1    119.669 786.37 -1114.8
##
## Step:  AIC=-1378.54
## quality ~ volatile.acidity + citric.acid + residual.sugar + chlorides +
##      free.sulfur.dioxide + total.sulfur.dioxide + pH + sulphates +
##      alcohol
##
##      Df Sum of Sq    RSS      AIC
## - residual.sugar      1      0.257 667.06 -1379.9
## - citric.acid         1      0.565 667.37 -1379.2
## <none>                                666.81 -1378.5
## - free.sulfur.dioxide 1      1.901 668.71 -1376.0
## - pH                 1      7.065 673.87 -1363.7
## - chlorides          1      9.940 676.75 -1356.9
## - total.sulfur.dioxide 1     10.031 676.84 -1356.7
## - sulphates          1     27.673 694.48 -1315.5
## - volatile.acidity    1     36.234 703.04 -1295.9
## - alcohol            1    120.633 787.44 -1114.7
##
## Step:  AIC=-1379.93
## quality ~ volatile.acidity + citric.acid + chlorides + free.sulfur.dioxide +
##      total.sulfur.dioxide + pH + sulphates + alcohol
##
##      Df Sum of Sq    RSS      AIC
## - citric.acid        1      0.475 667.54 -1380.8
## <none>                                667.06 -1379.9
## - free.sulfur.dioxide 1      2.064 669.13 -1377.0
## - pH                 1      7.138 674.20 -1364.9
## - total.sulfur.dioxide 1      9.828 676.89 -1358.5
## - chlorides          1      9.832 676.89 -1358.5
## - sulphates          1     27.446 694.51 -1317.5
## - volatile.acidity    1     35.977 703.04 -1297.9
## - alcohol            1    122.667 789.73 -1112.0
##
## Step:  AIC=-1380.79
## quality ~ volatile.acidity + chlorides + free.sulfur.dioxide +
##      total.sulfur.dioxide + pH + sulphates + alcohol
##
##      Df Sum of Sq    RSS      AIC
## <none>                                667.54 -1380.8
## - free.sulfur.dioxide 1      2.394 669.93 -1377.1

```

```
## - pH 1 7.073 674.61 -1365.9
## - total.sulfur.dioxide 1 10.787 678.32 -1357.2
## - chlorides 1 10.809 678.35 -1357.1
## - sulphates 1 27.060 694.60 -1319.2
## - volatile.acidity 1 42.318 709.85 -1284.5
## - alcohol 1 124.483 792.02 -1109.4

summary(model_backward)

##
## Call:
## lm(formula = quality ~ volatile.acidity + chlorides + free.sulfur.dioxide +
## total.sulfur.dioxide + pH + sulphates + alcohol, data = winequality)
##
## Residuals:
## Min 1Q Median 3Q Max
## -2.68918 -0.36757 -0.04653 0.46081 2.02954
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.4300987 0.4029168 10.995 < 2e-16 ***
## volatile.acidity -1.0127527 0.1008429 -10.043 < 2e-16 ***
## chlorides -2.0178138 0.3975417 -5.076 4.31e-07 ***
## free.sulfur.dioxide 0.0050774 0.0021255 2.389 0.017 *
## total.sulfur.dioxide -0.0034822 0.0006868 -5.070 4.43e-07 ***
## pH -0.4826614 0.1175581 -4.106 4.23e-05 ***
## sulphates 0.8826651 0.1099084 8.031 1.86e-15 ***
## alcohol 0.2893028 0.0167958 17.225 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6477 on 1591 degrees of freedom
## Multiple R-squared: 0.3595, Adjusted R-squared: 0.3567
## F-statistic: 127.6 on 7 and 1591 DF, p-value: < 2.2e-16
```

In this case, the sets of variable selected by the previous methods are the same, so we can say that the selected variables have a significant effect on the response variable. Then, the model fitted with this set of variables can be considered a model simple enough to prevent over-fitting and predict the quality of wines with a high accuracy.

### forward selection (p-value)

The forward selection method using p-value, follows these steps:

1. We fit a model with only one variable for all the predictor variables
2. We select the model with the variable added which has the lowest p-value
3. With the variable selected, we fit multiple models adding only one variable for model, for all the remaining predictor variables
4. Repeat 2 and 3 until the sum of p-values of all values is lower than 0.05

The following code shows the steps for this method:

```
# To check the results, you can find them in the r markdown that we sent via CV
model_forward_p_value <- lm(quality~fixed.acidity, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
```

```
## lm(formula = quality ~ fixed.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8248 -0.6061  0.1925  0.4341  2.5550
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.15732    0.09789  52.684 < 2e-16 ***
## fixed.acidity  0.05754    0.01152   4.996 6.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8016 on 1597 degrees of freedom
## Multiple R-squared:  0.01539, Adjusted R-squared:  0.01477
## F-statistic: 24.96 on 1 and 1597 DF, p-value: 6.496e-07

model_forward_p_value <- lm(quality~volatile.acidity, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79071 -0.54411 -0.00687  0.47350  2.93148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.56575    0.05791  113.39 <2e-16 ***
## volatile.acidity -1.76144    0.10389  -16.95 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7437 on 1597 degrees of freedom
## Multiple R-squared:  0.1525, Adjusted R-squared:  0.152
## F-statistic: 287.4 on 1 and 1597 DF, p-value: < 2.2e-16

model_forward_p_value <- lm(quality~citric.acid, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ citric.acid, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0011 -0.5976  0.1021  0.5057  2.5901
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.38172    0.03372 159.610 <2e-16 ***
## citric.acid   0.93845    0.10104   9.288 <2e-16 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7869 on 1597 degrees of freedom
## Multiple R-squared:  0.05124,    Adjusted R-squared:  0.05065
## F-statistic: 86.26 on 1 and 1597 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~residual.sugar, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6609 -0.6334  0.3580  0.3690  2.3729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.616055   0.041616 134.950  <2e-16 ***
## residual.sugar 0.007865   0.014331   0.549    0.583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8077 on 1597 degrees of freedom
## Multiple R-squared:  0.0001886, Adjusted R-squared: -0.0004375
## F-statistic: 0.3012 on 1 and 1597 DF,  p-value: 0.5832
model_forward_p_value <- lm(quality~free.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ free.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6864 -0.6394  0.3215  0.3762  2.4661
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.698107   0.036678 155.357  <2e-16 ***
## free.sulfur.dioxide -0.003911   0.001929  -2.027   0.0428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8068 on 1597 degrees of freedom
## Multiple R-squared:  0.002566, Adjusted R-squared:  0.001941
## F-statistic: 4.109 on 1 and 1597 DF,  p-value: 0.04283
model_forward_p_value <- lm(quality~total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ total.sulfur.dioxide, data = winequality)

```



```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8063 -0.6336  0.2164  0.3800  2.5527
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.8471792   0.0343670  170.140 < 2e-16 ***
## total.sulfur.dioxide -0.0045442   0.0006037  -7.527 8.62e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7939 on 1597 degrees of freedom
## Multiple R-squared:  0.03426, Adjusted R-squared:  0.03366
## F-statistic: 56.66 on 1 and 1597 DF, p-value: 8.622e-14
model_forward_p_value <- lm(quality~density, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ density, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7885 -0.6216  0.1554  0.4271  2.5177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    80.24      10.51    7.636 3.83e-14 ***
## density       -74.85      10.54   -7.100 1.87e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7954 on 1597 degrees of freedom
## Multiple R-squared:  0.0306, Adjusted R-squared:  0.02999
## F-statistic: 50.41 on 1 and 1597 DF, p-value: 1.875e-12
model_forward_p_value <- lm(quality~pH, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6817 -0.6394  0.3032  0.3878  2.4874
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.6359     0.4332  15.320 <2e-16 ***
## pH            -0.3020     0.1307  -2.311  0.021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.8065 on 1597 degrees of freedom
## Multiple R-squared:  0.003333, Adjusted R-squared:  0.002709
## F-statistic: 5.34 on 1 and 1597 DF, p-value: 0.02096
model_forward_p_value <- lm(quality~alcohol, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8442 -0.4112 -0.1690  0.5166  2.5888
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.87497    0.17471   10.73  <2e-16 ***
## alcohol      0.36084    0.01668   21.64  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7104 on 1597 degrees of freedom
## Multiple R-squared:  0.2267, Adjusted R-squared:  0.2263
## F-statistic: 468.3 on 1 and 1597 DF, p-value: < 2.2e-16
model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.81291 -0.54009 -0.00973  0.46502  2.95433
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.45088    0.12121  53.221  <2e-16 ***
## fixed.acidity  0.01192    0.01105   1.079   0.281
## volatile.acidity -1.73175    0.10747 -16.113  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7436 on 1596 degrees of freedom
## Multiple R-squared:  0.1532, Adjusted R-squared:  0.1521
## F-statistic: 144.3 on 2 and 1596 DF, p-value: < 2.2e-16
model_forward_p_value <- lm(quality~citric.acid + volatile.acidity, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ citric.acid + volatile.acidity, data = winequality)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79679 -0.53718 -0.01263  0.46847  2.93320
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.52857    0.08887   73.459  <2e-16 ***
## citric.acid       0.06320    0.11460    0.551    0.581
## volatile.acidity -1.72345    0.12467  -13.824  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7438 on 1596 degrees of freedom
## Multiple R-squared:  0.1527, Adjusted R-squared:  0.1516
## F-statistic: 143.8 on 2 and 1596 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78211 -0.53964 -0.00728  0.47217  2.94096
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.544754    0.066856   97.893  <2e-16 ***
## residual.sugar    0.008294    0.013197    0.628    0.53
## volatile.acidity -1.761563    0.103914  -16.952  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7438 on 1596 degrees of freedom
## Multiple R-squared:  0.1527, Adjusted R-squared:  0.1517
## F-statistic: 143.9 on 2 and 1596 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~free.sulfur.dioxide + volatile.acidity, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ free.sulfur.dioxide + volatile.acidity,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71431 -0.54272 -0.01128  0.48469  2.91593
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.634235    0.064582  102.725  <2e-16 ***
```

```
## free.sulfur.dioxide -0.004228 0.001776 -2.381 0.0174 *
## volatile.acidity -1.764032 0.103748 -17.003 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7426 on 1596 degrees of freedom
## Multiple R-squared: 0.1555, Adjusted R-squared: 0.1545
## F-statistic: 147 on 2 and 1596 DF, p-value: < 2.2e-16
model_forward_p_value <- lm(quality~total.sulfur.dioxide + volatile.acidity, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ total.sulfur.dioxide + volatile.acidity,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78011 -0.51369 -0.01658  0.46470  3.07333
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.7154491   0.0611231 109.868 < 2e-16 ***
## total.sulfur.dioxide -0.0038334  0.0005592  -6.855 1.01e-11 ***
## volatile.acidity    -1.7075850   0.1027303 -16.622 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7332 on 1596 degrees of freedom
## Multiple R-squared: 0.1768, Adjusted R-squared: 0.1757
## F-statistic: 171.4 on 2 and 1596 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~density + volatile.acidity, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ density + volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.67724 -0.54639 -0.02654  0.47391  2.67102
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      77.5250     9.6674   8.019 2.04e-15 ***
## density          -71.1996     9.7000  -7.340 3.38e-13 ***
## volatile.acidity  -1.7449     0.1022 -17.067 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7316 on 1596 degrees of freedom
## Multiple R-squared: 0.1802, Adjusted R-squared: 0.1792
## F-statistic: 175.4 on 2 and 1596 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~pH + volatile.acidity, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ pH + volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80704 -0.54411 -0.00415  0.46803  2.89689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.9621     0.4013   14.86 <2e-16 ***
## pH              0.1884     0.1239    1.52  0.129
## volatile.acidity -1.7996     0.1068  -16.84 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7434 on 1596 degrees of freedom
## Multiple R-squared:  0.1538, Adjusted R-squared:  0.1527
## F-statistic: 145 on 2 and 1596 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~alcohol + volatile.acidity, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ alcohol + volatile.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59342 -0.40416 -0.07426  0.46539  2.25809
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.09547     0.18450   16.78 <2e-16 ***
## alcohol          0.31381     0.01601   19.60 <2e-16 ***
## volatile.acidity -1.38364     0.09527  -14.52 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6678 on 1596 degrees of freedom
## Multiple R-squared:  0.317, Adjusted R-squared:  0.3161
## F-statistic: 370.4 on 2 and 1596 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol,
##      data = winequality)
##
## Residuals:
```

```

##      Min      1Q   Median      3Q      Max
## -2.65565 -0.39856 -0.08143  0.46879  2.28524
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.674469   0.218033   12.27 < 2e-16 ***
## fixed.acidity    0.035763   0.009961    3.59 0.00034 ***
## volatile.acidity -1.286244   0.098716  -13.03 < 2e-16 ***
## alcohol          0.320726   0.016064   19.96 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6653 on 1595 degrees of freedom
## Multiple R-squared:  0.3225, Adjusted R-squared:  0.3212
## F-statistic: 253.1 on 3 and 1595 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~citric.acid + volatile.acidity + alcohol, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ citric.acid + volatile.acidity + alcohol,
##     data = winequality)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.59992 -0.40354 -0.07282  0.47165  2.23655
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.05533   0.19433   15.722 <2e-16 ***
## citric.acid     0.06779   0.10291    0.659   0.51
## volatile.acidity -1.34286   0.11362  -11.818 <2e-16 ***
## alcohol         0.31384   0.01601   19.602 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6679 on 1595 degrees of freedom
## Multiple R-squared:  0.3172, Adjusted R-squared:  0.3159
## F-statistic: 247 on 3 and 1595 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity + alcohol, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol,
##     data = winequality)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.59520 -0.40432 -0.07387  0.46972  2.25759
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

```

```
## (Intercept)      3.098825   0.185901   16.67   <2e-16 ***
## residual.sugar  -0.001781   0.011864   -0.15    0.881
## volatile.acidity -1.383483   0.095303  -14.52   <2e-16 ***
## alcohol          0.313917   0.016027   19.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.668 on 1595 degrees of freedom
## Multiple R-squared:  0.317, Adjusted R-squared:  0.3157
## F-statistic: 246.8 on 3 and 1595 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~free.sulfur.dioxide + volatile.acidity + alcohol,
                             data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ free.sulfur.dioxide + volatile.acidity +
##     alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.55905 -0.40378 -0.07351  0.46641  2.23730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.142876   0.188527   16.671   <2e-16 ***
## free.sulfur.dioxide -0.001951   0.001601   -1.219    0.223
## volatile.acidity   -1.386553   0.095284  -14.552   <2e-16 ***
## alcohol           0.312384   0.016048   19.466   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6677 on 1595 degrees of freedom
## Multiple R-squared:  0.3176, Adjusted R-squared:  0.3164
## F-statistic: 247.5 on 3 and 1595 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~total.sulfur.dioxide + volatile.acidity + alcohol,
                             data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ total.sulfur.dioxide + volatile.acidity +
##     alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59554 -0.39683 -0.06971  0.45436  2.21917
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.3050091   0.1918403   17.228 < 2e-16 ***
## total.sulfur.dioxide -0.0019630   0.0005171  -3.796 0.000153 ***
## volatile.acidity    -1.3705246   0.0949333  -14.437 < 2e-16 ***
```

```
## alcohol          0.3017966  0.0162519  18.570  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.665 on 1595 degrees of freedom
## Multiple R-squared:  0.3231, Adjusted R-squared:  0.3218
## F-statistic: 253.8 on 3 and 1595 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~density + volatile.acidity + alcohol, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61526 -0.39782 -0.07035  0.46700  2.25079
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -18.40654    10.29779  -1.787   0.0741 .
## density        21.35964    10.22796   2.088   0.0369 *
## volatile.acidity -1.36524     0.09557 -14.285  <2e-16 ***
## alcohol         0.33321     0.01849  18.019  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6671 on 1595 degrees of freedom
## Multiple R-squared:  0.3189, Adjusted R-squared:  0.3176
## F-statistic: 248.9 on 3 and 1595 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~pH + volatile.acidity + alcohol, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ pH + volatile.acidity + alcohol, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.54671 -0.40614 -0.07843  0.46216  2.21636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.26881    0.36898  11.569  < 2e-16 ***
## pH             -0.42187    0.11504  -3.667  0.000253 ***
## volatile.acidity -1.27876    0.09911 -12.902  < 2e-16 ***
## alcohol         0.32994    0.01654  19.947  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6652 on 1595 degrees of freedom
## Multiple R-squared:  0.3227, Adjusted R-squared:  0.3214
```



```
## F-statistic: 253.3 on 3 and 1595 DF, p-value: < 2.2e-16
model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol
                             + total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65026 -0.38507 -0.06865  0.47237  2.21680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.9117067   0.2283290   12.752 < 2e-16 ***
## fixed.acidity      0.0315709   0.0100056    3.155 0.001633 **
## volatile.acidity   -1.2859043   0.0983932  -13.069 < 2e-16 ***
## alcohol           0.3091414   0.0163728   18.881 < 2e-16 ***
## total.sulfur.dioxide -0.0017602  0.0005197   -3.387 0.000724 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6632 on 1594 degrees of freedom
## Multiple R-squared:  0.3273, Adjusted R-squared:  0.3256
## F-statistic: 193.9 on 4 and 1594 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~citric.acid + volatile.acidity + alcohol
                             + total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ citric.acid + volatile.acidity + alcohol +
##     total.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.60574 -0.39986 -0.06932  0.46019  2.18458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.2478356   0.1997541   16.259 < 2e-16 ***
## citric.acid       0.1056835   0.1029229    1.027  0.30466
## volatile.acidity   -1.3066184   0.1135141  -11.511 < 2e-16 ***
## alcohol           0.3015240   0.0162538   18.551 < 2e-16 ***
## total.sulfur.dioxide -0.0020137  0.0005195   -3.876 0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.665 on 1594 degrees of freedom
## Multiple R-squared:  0.3236, Adjusted R-squared:  0.3219
## F-statistic: 190.6 on 4 and 1594 DF, p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity + alcohol
+ total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol +
##     total.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.58732 -0.39696 -0.07468  0.45673  2.21997
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.2977665   0.1921622  17.161 < 2e-16 ***
## residual.sugar      0.0083030   0.0120988   0.686 0.492644
## volatile.acidity    -1.3707104   0.0949494 -14.436 < 2e-16 ***
## alcohol              0.3008290   0.0163157  18.438 < 2e-16 ***
## total.sulfur.dioxide -0.0020416   0.0005297  -3.854 0.000121 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6651 on 1594 degrees of freedom
## Multiple R-squared:  0.3233, Adjusted R-squared:  0.3216
## F-statistic: 190.4 on 4 and 1594 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~free.sulfur.dioxide + volatile.acidity
+ alcohol + total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ free.sulfur.dioxide + volatile.acidity +
##     alcohol + total.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.66401 -0.39126 -0.07046  0.44828  2.24351
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.3008844   0.1917228  17.217 < 2e-16 ***
## free.sulfur.dioxide  0.0038366   0.0021490   1.785  0.0744 .
## volatile.acidity    -1.3592149   0.0950796 -14.296 < 2e-16 ***
## alcohol              0.2994969   0.0162918  18.383 < 2e-16 ***
## total.sulfur.dioxide -0.0027976   0.0006968  -4.015 6.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6646 on 1594 degrees of freedom
## Multiple R-squared:  0.3245, Adjusted R-squared:  0.3228
## F-statistic: 191.4 on 4 and 1594 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~density + volatile.acidity + alcohol
                             + total.sulfur.dioxide, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol +
##     total.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61607 -0.39286 -0.06512  0.46152  2.19572
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.694e+01  1.026e+01  -1.650  0.099130 .
## density        2.010e+01  1.019e+01   1.972  0.048737 *
## volatile.acidity -1.353e+00  9.524e-02 -14.211 < 2e-16 ***
## alcohol         3.203e-01  1.874e-02  17.088 < 2e-16 ***
## total.sulfur.dioxide -1.929e-03  5.169e-04  -3.732  0.000196 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6644 on 1594 degrees of freedom
## Multiple R-squared:  0.3248, Adjusted R-squared:  0.3231
## F-statistic: 191.7 on 4 and 1594 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~pH + volatile.acidity + alcohol + total.sulfur.dioxide
                             , data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ pH + volatile.acidity + alcohol + total.sulfur.dioxide,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.54704 -0.39786 -0.06261  0.46225  2.17603
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.5333226  0.3733444  12.142 < 2e-16 ***
## pH             -0.4387993  0.1145924  -3.829  0.000134 ***
## volatile.acidity -1.2609452  0.0987659 -12.767 < 2e-16 ***
## alcohol         0.3181250  0.0167351  19.009 < 2e-16 ***
## total.sulfur.dioxide -0.0020368  0.0005153  -3.953  8.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6622 on 1594 degrees of freedom
## Multiple R-squared:  0.3293, Adjusted R-squared:  0.3276
## F-statistic: 195.6 on 4 and 1594 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol
+ total.sulfur.dioxide + pH, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.57491 -0.39398 -0.06458  0.46291  2.18007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.1667587   0.5899471    7.063 2.43e-12 ***
## fixed.acidity    0.0107960   0.0134521    0.803 0.422356
## volatile.acidity -1.2527119   0.0993082  -12.614 < 2e-16 ***
## alcohol         0.3175516   0.0167523   18.956 < 2e-16 ***
## total.sulfur.dioxide -0.0019535  0.0005257   -3.716 0.000209 ***
## pH             -0.3558951   0.1542903   -2.307 0.021202 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6623 on 1593 degrees of freedom
## Multiple R-squared:  0.3296, Adjusted R-squared:  0.3275
## F-statistic: 156.6 on 5 and 1593 DF,  p-value: < 2.2e-16
```

```
model_forward_p_value <- lm(quality~citric.acid + volatile.acidity + alcohol
+ total.sulfur.dioxide+ pH, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ citric.acid + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.52451 -0.39926 -0.06223  0.44793  2.17555
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.8384193   0.4570365   10.587 < 2e-16 ***
## citric.acid    -0.1394403   0.1205103   -1.157 0.247412
## volatile.acidity -1.3247757   0.1131185  -11.711 < 2e-16 ***
## alcohol         0.3215376   0.0169913   18.924 < 2e-16 ***
## total.sulfur.dioxide -0.0019837  0.0005173   -3.835 0.000130 ***
## pH            -0.5208429   0.1347450   -3.865 0.000115 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6621 on 1593 degrees of freedom
## Multiple R-squared:  0.3299, Adjusted R-squared:  0.3277
```

```
## F-statistic: 156.8 on 5 and 1593 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.54366 -0.39603 -0.06248  0.46271  2.18269
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.5200377   0.3758315   12.027 < 2e-16 ***
## residual.sugar      0.0038077   0.0121073    0.314 0.753187
## volatile.acidity    -1.2619193   0.0988423  -12.767 < 2e-16 ***
## alcohol              0.3175488   0.0168398   18.857 < 2e-16 ***
## total.sulfur.dioxide -0.0020722   0.0005276   -3.928 8.95e-05 ***
## pH                 -0.4352400   0.1151822   -3.779 0.000163 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6624 on 1593 degrees of freedom
## Multiple R-squared:  0.3293, Adjusted R-squared:  0.3272
## F-statistic: 156.4 on 5 and 1593 DF,  p-value: < 2.2e-16
model_forward_p_value <- lm(quality~free.sulfur.dioxide + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ free.sulfur.dioxide + volatile.acidity +
##     alcohol + total.sulfur.dioxide + pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.63722 -0.39655 -0.05734  0.45409  2.25189
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.6608045   0.3763161   12.385 < 2e-16 ***
## free.sulfur.dioxide  0.0053481   0.0021681    2.467  0.0137 *
## volatile.acidity    -1.2332940   0.0992439  -12.427 < 2e-16 ***
## alcohol              0.3166903   0.0167186   18.942 < 2e-16 ***
## total.sulfur.dioxide -0.0032081   0.0007001   -4.582 4.96e-06 ***
## pH                 -0.4863946   0.1160257   -4.192 2.92e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6612 on 1593 degrees of freedom
```

```
## Multiple R-squared:  0.3318, Adjusted R-squared:  0.3297
## F-statistic: 158.2 on 5 and 1593 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~density + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.56111 -0.39861 -0.06655  0.45930  2.19134
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.9980861  10.7171656  -0.560  0.575783
## density        10.3721348  10.5486756   0.983  0.325627
## volatile.acidity -1.2600068  0.0987715 -12.757 < 2e-16 ***
## alcohol         0.3264758  0.0187670  17.396 < 2e-16 ***
## total.sulfur.dioxide -0.0020141  0.0005158  -3.905  9.82e-05 ***
## pH            -0.4072606  0.1189980  -3.422  0.000636 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6622 on 1593 degrees of freedom
## Multiple R-squared:  0.3297, Adjusted R-squared:  0.3276
## F-statistic: 156.7 on 5 and 1593 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide,
                             data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.66670 -0.39198 -0.06302  0.45827  2.25646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2792639  0.5907453   7.244 6.77e-13 ***
## fixed.acidity  0.0112546  0.0134317   0.838  0.40220
## volatile.acidity -1.2245814  0.0997963 -12.271 < 2e-16 ***
## alcohol        0.3160859  0.0167358  18.887 < 2e-16 ***
## total.sulfur.dioxide -0.0031268  0.0007069  -4.423 1.04e-05 ***
## pH            -0.4001910  0.1550759  -2.581  0.00995 **
## free.sulfur.dioxide  0.0053731  0.0021685   2.478  0.01332 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6612 on 1592 degrees of freedom
## Multiple R-squared:  0.3321, Adjusted R-squared:  0.3296
## F-statistic:   132 on 6 and 1592 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~citric.acid + volatile.acidity + alcohol
                             + total.sulfur.dioxide+ pH + free.sulfur.dioxide,
                             data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ citric.acid + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61656 -0.39728 -0.05674  0.45369  2.24815
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.8804162   0.4567516  10.685 < 2e-16 ***
## citric.acid      -0.1029700   0.1213496  -0.849  0.3963
## volatile.acidity  -1.2816634   0.1144571 -11.198 < 2e-16 ***
## alcohol           0.3192744   0.0169952  18.786 < 2e-16 ***
## total.sulfur.dioxide -0.0031167  0.0007084  -4.399 1.16e-05 ***
## pH               -0.5448564   0.1349486  -4.038 5.66e-05 ***
## free.sulfur.dioxide  0.0051095   0.0021865   2.337  0.0196 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6612 on 1592 degrees of freedom
## Multiple R-squared:  0.3321, Adjusted R-squared:  0.3296
## F-statistic:   132 on 6 and 1592 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide,
                             data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.63564 -0.39639 -0.05692  0.45424  2.25405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.6554467   0.3792972  12.274 < 2e-16 ***
## residual.sugar     0.0013966   0.0121284   0.115  0.9083
## volatile.acidity  -1.2337565   0.0993558 -12.418 < 2e-16 ***
```

```
## alcohol          0.3164844  0.0168191  18.817 < 2e-16 ***
## total.sulfur.dioxide -0.0032167  0.0007043  -4.567 5.32e-06 ***
## pH              -0.4849080  0.1167774  -4.152 3.46e-05 ***
## free.sulfur.dioxide  0.0053277  0.0021760   2.448  0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6614 on 1592 degrees of freedom
## Multiple R-squared:  0.3318, Adjusted R-squared:  0.3293
## F-statistic: 131.8 on 6 and 1592 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~density + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide,
                             data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65258 -0.39449 -0.06478  0.45261  2.26843
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.3356847  10.7006722  -0.592 0.553879
## density        10.8310996  10.5332124   1.028 0.303974
## volatile.acidity -1.2321119   0.0992487 -12.414 < 2e-16 ***
## alcohol         0.3254002   0.0187416  17.362 < 2e-16 ***
## total.sulfur.dioxide -0.0031930  0.0007003  -4.560 5.51e-06 ***
## pH             -0.4538083   0.1202736  -3.773 0.000167 ***
## free.sulfur.dioxide  0.0053872  0.0021684   2.484 0.013079 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6611 on 1592 degrees of freedom
## Multiple R-squared:  0.3323, Adjusted R-squared:  0.3298
## F-statistic: 132 on 6 and 1592 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide
                             + citric.acid, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65627 -0.39321 -0.06135  0.45404  2.25370
```



```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.3140011   0.5909495   7.300 4.52e-13 ***
## fixed.acidity      0.0238248   0.0157814   1.510 0.131324
## volatile.acidity  -1.3163583   0.1166965 -11.280 < 2e-16 ***
## alcohol           0.3208337   0.0170197  18.851 < 2e-16 ***
## total.sulfur.dioxide -0.0028440   0.0007308  -3.892 0.000104 ***
## pH                -0.4265986   0.1559890  -2.735 0.006311 **
## free.sulfur.dioxide  0.0049004   0.0021900   2.238 0.025383 *
## citric.acid       -0.2160929   0.1425787  -1.516 0.129818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6609 on 1591 degrees of freedom
## Multiple R-squared:  0.3331, Adjusted R-squared:  0.3302
## F-statistic: 113.5 on 7 and 1591 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide
                             + citric.acid, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61253 -0.39644 -0.05643  0.45495  2.25243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.8778209   0.4570221  10.673 < 2e-16 ***
## residual.sugar      0.0028628   0.0122452   0.234  0.8152
## volatile.acidity  -1.2844587   0.1151137 -11.158 < 2e-16 ***
## alcohol           0.3189511   0.0170564  18.700 < 2e-16 ***
## total.sulfur.dioxide -0.0031307   0.0007112  -4.402 1.14e-05 ***
## pH                -0.5440418   0.1350336  -4.029 5.87e-05 ***
## free.sulfur.dioxide  0.0050587   0.0021979   2.302  0.0215 *
## citric.acid       -0.1069025   0.1225455  -0.872  0.3831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6614 on 1591 degrees of freedom
## Multiple R-squared:  0.3322, Adjusted R-squared:  0.3292
## F-statistic: 113 on 7 and 1591 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~density + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide
                             + citric.acid, data = winequality)
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62286 -0.38775 -0.05943  0.45638  2.27393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -14.593406   12.003288  -1.216   0.2242
## density        19.403070   11.951025   1.624   0.1047
## volatile.acidity -1.329219    0.118089 -11.256 < 2e-16 ***
## alcohol         0.337531    0.020371  16.569 < 2e-16 ***
## total.sulfur.dioxide -0.002996  0.000712  -4.208 2.72e-05 ***
## pH             -0.546518    0.134883  -4.052 5.33e-05 ***
## free.sulfur.dioxide  0.004934   0.002188   2.255  0.0243 *
## citric.acid     -0.208715    0.137669  -1.516  0.1297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6609 on 1591 degrees of freedom
## Multiple R-squared:  0.3332, Adjusted R-squared:  0.3303
## F-statistic: 113.6 on 7 and 1591 DF,  p-value: < 2.2e-16
# here p-value > 0.05
model_forward_p_value <- lm(quality~residual.sugar + volatile.acidity
+ alcohol + total.sulfur.dioxide + pH
+ free.sulfur.dioxide + citric.acid
+ fixed.acidity , data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ residual.sugar + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid +
##     fixed.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65363 -0.39339 -0.06241  0.45531  2.25622
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.3157444   0.5912641   7.299 4.55e-13 ***
## residual.sugar  0.0017087   0.0122647   0.139 0.889216
## volatile.acidity -1.3178250   0.1172062 -11.244 < 2e-16 ***
## alcohol         0.3206316   0.0170866  18.765 < 2e-16 ***
## total.sulfur.dioxide -0.0028540  0.0007345  -3.886 0.000106 ***
## pH             -0.4267998   0.1560437  -2.735 0.006305 **
## free.sulfur.dioxide  0.0048713   0.0022006   2.214 0.026997 *
## citric.acid     -0.2177825   0.1431373  -1.521 0.128335
## fixed.acidity    0.0236864   0.0158175   1.497 0.134468
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6611 on 1590 degrees of freedom
## Multiple R-squared:  0.3331, Adjusted R-squared:  0.3297
## F-statistic: 99.27 on 8 and 1590 DF,  p-value: < 2.2e-16

model_forward_p_value <- lm(quality~density
                             + volatile.acidity + alcohol + total.sulfur.dioxide
                             + pH + free.sulfur.dioxide + citric.acid
                             + fixed.acidity , data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ density + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid +
##     fixed.acidity, data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.63952 -0.39293 -0.06271  0.45601  2.26844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -8.7444172  16.8878728  -0.518  0.60467
## density         13.3103516  17.2031421   0.774  0.43921
## volatile.acidity  -1.3305761   0.1181490 -11.262 < 2e-16 ***
## alcohol           0.3325305   0.0227660  14.606 < 2e-16 ***
## total.sulfur.dioxide -0.0029057  0.0007352  -3.952 8.09e-05 ***
## pH               -0.4904711   0.1765040  -2.779  0.00552 **
## free.sulfur.dioxide  0.0048913   0.0021903   2.233  0.02568 *
## citric.acid       -0.2286244   0.1435135  -1.593  0.11135
## fixed.acidity       0.0111864   0.0227143   0.492  0.62245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.661 on 1590 degrees of freedom
## Multiple R-squared:  0.3333, Adjusted R-squared:  0.33
## F-statistic: 99.38 on 8 and 1590 DF,  p-value: < 2.2e-16

# so, the model:
model_forward_p_value <- lm(quality~fixed.acidity + volatile.acidity + alcohol
                             + total.sulfur.dioxide + pH + free.sulfur.dioxide
                             + citric.acid, data = winequality)
summary(model_forward_p_value)

##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -2.65627 -0.39321 -0.06135  0.45404  2.25370
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.3140011   0.5909495    7.300 4.52e-13 ***
## fixed.acidity      0.0238248   0.0157814    1.510 0.131324
## volatile.acidity  -1.3163583   0.1166965   -11.280 < 2e-16 ***
## alcohol           0.3208337   0.0170197   18.851 < 2e-16 ***
## total.sulfur.dioxide -0.0028440  0.0007308   -3.892 0.000104 ***
## pH               -0.4265986   0.1559890   -2.735 0.006311 **
## free.sulfur.dioxide  0.0049004  0.0021900    2.238 0.025383 *
## citric.acid       -0.2160929  0.1425787   -1.516 0.129818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6609 on 1591 degrees of freedom
## Multiple R-squared:  0.3331, Adjusted R-squared:  0.3302
## F-statistic: 113.5 on 7 and 1591 DF,  p-value: < 2.2e-16
```

```
summary(model_forward_p_value)
```

```
##
## Call:
## lm(formula = quality ~ fixed.acidity + volatile.acidity + alcohol +
##     total.sulfur.dioxide + pH + free.sulfur.dioxide + citric.acid,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65627 -0.39321 -0.06135  0.45404  2.25370
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.3140011   0.5909495    7.300 4.52e-13 ***
## fixed.acidity      0.0238248   0.0157814    1.510 0.131324
## volatile.acidity  -1.3163583   0.1166965   -11.280 < 2e-16 ***
## alcohol           0.3208337   0.0170197   18.851 < 2e-16 ***
## total.sulfur.dioxide -0.0028440  0.0007308   -3.892 0.000104 ***
## pH               -0.4265986   0.1559890   -2.735 0.006311 **
## free.sulfur.dioxide  0.0049004  0.0021900    2.238 0.025383 *
## citric.acid       -0.2160929  0.1425787   -1.516 0.129818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6609 on 1591 degrees of freedom
## Multiple R-squared:  0.3331, Adjusted R-squared:  0.3302
## F-statistic: 113.5 on 7 and 1591 DF,  p-value: < 2.2e-16
```

### forward selection (AIC)

Having in mind the interpretation of AIC and the steps of backward selection for p-value, here we will add to the model the variable that decrease the most the AIC of the model, we stop when we can't get a lower AIC

The following code gets a set of variables following the forward selection method for AIC criteria

```

model_null <- lm(quality~1,data = winequality)
model_forward <- stepAIC(model_null, trace = TRUE, direction = "forward",
                        scope=list(lower=model_null, upper=~residual.sugar
                                + volatile.acidity + alcohol + total.sulfur.dioxide
                                + pH + free.sulfur.dioxide + citric.acid
                                + fixed.acidity + density + sulphates + chlorides))

```

```

## Start:  AIC=-682.5
## quality ~ 1
##
##              Df Sum of Sq    RSS    AIC
## + alcohol      1   236.295  805.87 -1091.65
## + volatile.acidity  1   158.967  883.20 -945.14
## + sulphates     1    65.865  976.30 -784.89
## + citric.acid   1    53.405  988.76 -764.61
## + total.sulfur.dioxide 1    35.707 1006.46 -736.24
## + density       1    31.887 1010.28 -730.19
## + chlorides     1    17.318 1024.85 -707.29
## + fixed.acidity  1    16.038 1026.13 -705.29
## + pH            1     3.473 1038.69 -685.84
## + free.sulfur.dioxide 1     2.674 1039.49 -684.61
## <none>                      1042.17 -682.50
## + residual.sugar      1     0.197 1041.97 -680.80
##
## Step:  AIC=-1091.65
## quality ~ alcohol
##
##              Df Sum of Sq    RSS    AIC
## + volatile.acidity  1    94.074  711.80 -1288.1
## + sulphates         1    44.977  760.89 -1181.5
## + citric.acid       1    31.953  773.92 -1154.3
## + pH                1    26.362  779.51 -1142.8
## + fixed.acidity     1    24.623  781.25 -1139.3
## + total.sulfur.dioxide 1     8.270  797.60 -1106.2
## + density           1     5.203  800.67 -1100.0
## <none>                      805.87 -1091.7
## + chlorides         1     0.611  805.26 -1090.9
## + free.sulfur.dioxide 1     0.325  805.55 -1090.3
## + residual.sugar    1     0.041  805.83 -1089.7
##
## Step:  AIC=-1288.14
## quality ~ alcohol + volatile.acidity
##
##              Df Sum of Sq    RSS    AIC
## + sulphates       1   19.6916  692.10 -1331.0
## + total.sulfur.dioxide 1    6.3730  705.42 -1300.5
## + pH              1    5.9515  705.84 -1299.6
## + fixed.acidity   1    5.7061  706.09 -1299.0
## + density         1    1.9410  709.86 -1290.5
## <none>                      711.80 -1288.1
## + free.sulfur.dioxide 1    0.6621  711.13 -1287.6
## + chlorides       1    0.3762  711.42 -1287.0
## + citric.acid     1    0.1936  711.60 -1286.6
## + residual.sugar  1    0.0101  711.79 -1286.2

```

```

##
## Step: AIC=-1331
## quality ~ alcohol + volatile.acidity + sulphates
##
##           Df Sum of Sq    RSS    AIC
## + total.sulfur.dioxide 1    8.2176 683.89 -1348.1
## + chlorides             1    7.4925 684.61 -1346.4
## + fixed.acidity         1    3.3282 688.78 -1336.7
## + pH                   1    3.0454 689.06 -1336.0
## + free.sulfur.dioxide   1    1.1129 690.99 -1331.6
## <none>                  692.10 -1331.0
## + citric.acid           1    0.2522 691.85 -1329.6
## + density               1    0.2222 691.88 -1329.5
## + residual.sugar        1    0.0143 692.09 -1329.0
##
## Step: AIC=-1348.1
## quality ~ alcohol + volatile.acidity + sulphates + total.sulfur.dioxide
##
##           Df Sum of Sq    RSS    AIC
## + chlorides             1    8.0370 675.85 -1365.0
## + pH                   1    3.3094 680.58 -1353.8
## + fixed.acidity         1    2.1037 681.78 -1351.0
## + free.sulfur.dioxide   1    1.3557 682.53 -1349.3
## <none>                  683.89 -1348.1
## + residual.sugar        1    0.2634 683.62 -1346.7
## + density               1    0.1077 683.78 -1346.3
## + citric.acid           1    0.0730 683.81 -1346.3
##
## Step: AIC=-1365
## quality ~ alcohol + volatile.acidity + sulphates + total.sulfur.dioxide +
## chlorides
##
##           Df Sum of Sq    RSS    AIC
## + pH                   1    5.9189 669.93 -1377.1
## + fixed.acidity         1    2.4065 673.44 -1368.7
## + free.sulfur.dioxide   1    1.2403 674.61 -1365.9
## <none>                  675.85 -1365.0
## + residual.sugar        1    0.5531 675.30 -1364.3
## + citric.acid           1    0.1615 675.69 -1363.4
## + density               1    0.1526 675.70 -1363.4
##
## Step: AIC=-1377.06
## quality ~ alcohol + volatile.acidity + sulphates + total.sulfur.dioxide +
## chlorides + pH
##
##           Df Sum of Sq    RSS    AIC
## + free.sulfur.dioxide   1    2.39413 667.54 -1380.8
## <none>                  669.93 -1377.1
## + citric.acid           1    0.80525 669.13 -1377.0
## + residual.sugar        1    0.28390 669.65 -1375.7
## + density               1    0.04468 669.89 -1375.2
## + fixed.acidity         1    0.01040 669.92 -1375.1
##
## Step: AIC=-1380.79

```

```
## quality ~ alcohol + volatile.acidity + sulphates + total.sulfur.dioxide +
## chlorides + pH + free.sulfur.dioxide
##
##           Df Sum of Sq    RSS    AIC
## <none>                667.54 -1380.8
## + citric.acid      1    0.47480 667.06 -1379.9
## + residual.sugar   1    0.16673 667.37 -1379.2
## + density          1    0.03079 667.51 -1378.9
## + fixed.acidity    1    0.00663 667.53 -1378.8

summary(model_forward)

##
## Call:
## lm(formula = quality ~ alcohol + volatile.acidity + sulphates +
## total.sulfur.dioxide + chlorides + pH + free.sulfur.dioxide,
## data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68918 -0.36757 -0.04653  0.46081  2.02954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.4300987   0.4029168   10.995 < 2e-16 ***
## alcohol         0.2893028   0.0167958   17.225 < 2e-16 ***
## volatile.acidity -1.0127527   0.1008429  -10.043 < 2e-16 ***
## sulphates       0.8826651   0.1099084    8.031 1.86e-15 ***
## total.sulfur.dioxide -0.0034822  0.0006868   -5.070 4.43e-07 ***
## chlorides      -2.0178138   0.3975417   -5.076 4.31e-07 ***
## pH             -0.4826614   0.1175581   -4.106 4.23e-05 ***
## free.sulfur.dioxide  0.0050774  0.0021255    2.389  0.017 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6477 on 1591 degrees of freedom
## Multiple R-squared:  0.3595, Adjusted R-squared:  0.3567
## F-statistic: 127.6 on 7 and 1591 DF,  p-value: < 2.2e-16

# https://stackoverflow.com/questions/33715860/problems-with-forward-selection-with-stepAIC-r
```

In this case, the set of variables selected by the forward selection using AIC is the same set obtained with backward selection, this fact can also explain that all these variables have a significant effect on the response variable. Since the model obtained with forward selection using p-value have higher p-values with a different set of sizes, knowing the alternative we should prioritize it.

### Exercise 3

**Find the best possible subset of variables to select the best fit model. Compare the results with the final models obtained in the previous point.**

To obtain the best subsets of variables of all possible subsets of variables, we can use the following code:

```
model_subsets_best <- lm(quality~., data = winequality)
ols_step_best_subset(model_subsets_best)
```

Output resized to be filled in the pdf

```
##                               Best Subsets Regression
## -----
## Model Index    Predictors
## -----
## 1  alcohol
## 2  volatile.acidity alcohol
## 3  volatile.acidity sulphates alcohol
## 4  volatile.acidity total.sulfur.dioxide sulphates alcohol
## 5  volatile.acidity chlorides total.sulfur.dioxide pH sulphates alcohol
## 6  volatile.acidity chlorides free.sulfur.dioxide total.sulfur.dioxide pH sulphates alcohol
## 7  volatile.acidity citric.acid chlorides free.sulfur.dioxide total.sulfur.dioxide pH sulphates alcohol
## 8  volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide pH sulphates alcohol
## 9  volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide pH sulphates alcohol
## 10 fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide pH sulphates alcohol
## 11 fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide density pH sulphates alcohol
## -----
##
##                               Subsets Regression Summary
## -----
## Model    R-Square R-2 Adj R-Pred    C(p) AIC      SBIC      SBC      MSEP      FPE      HSP      APC
## -----
## 1 0.2267  0.2263  0.2246 324.1115 3448.1135 -1090.3748 3464.2449 806.8797 0.5052 3e-04 0.7752
## 2 0.3170  0.3161  0.3139 102.0818 3251.6275 -1286.4844 3273.1360 713.1345 0.4468 3e-04 0.6856
## 3 0.3359  0.3346  0.3316 57.1879 3208.7683 -1329.2376 3235.6540 693.8409 0.4350 3e-04 0.6674
## 4 0.3438  0.3421  0.3386 39.6184 3191.6693 -1346.2787 3223.9321 686.0331 0.4304 3e-04 0.6603
## 5 0.3515  0.3495  0.3454 22.4791 3174.7667 -1363.0770 3212.4066 678.3967 0.4259 3e-04 0.6534
## 6 0.3572  0.3548  0.3501 10.3837 3162.7015 -1375.0322 3205.7185 672.8782 0.4227 3e-04 0.6485
## 7 0.3595  0.3567  0.3515 6.6823 3158.9769 -1378.6948 3207.3711 670.8952 0.4217 3e-04 0.6470
## 8 0.3599  0.3567  0.351 7.5516 3159.8391 -1377.8080 3213.6105 670.8399 0.4219 3e-04 0.6473
## 9 0.3602  0.3565  0.3501 8.9404 3161.2237 -1376.4025 3220.3722 671.0041 0.4223 3e-04 0.6479
## 10 0.3603 0.3562  0.3491 10.6832 3162.9648 -1374.6439 3227.4904 671.3182 0.4227 3e-04 0.6486
## 11 0.3606 0.3561  0.3483 12.0000 3164.2766 -1373.3075 3234.1793 671.4524 0.4231 3e-04 0.6491
## -----
## AIC: Akaike Information Criteria
## SBIC: Sawa's Bayesian Information Criteria
## SBC: Schwarz Bayesian Criteria
## MSEP: Estimated error of prediction, assuming multivariate normality
## FPE: Final Prediction Error
## HSP: Hocking's Sp
## APC: Amemiya Prediction Criteria
```

Since there are many possible best subsets for different criteria, in order to select a good model, we can get the best models based on a reduced number of criteria, like the  $R^2_{Adj}$  or BIC (Bayesian information criterion), the next code helps us to do that.

```
model_subsets <- regsubsets(quality-residual.sugar + volatile.acidity + alcohol
                           + total.sulfur.dioxide + pH + free.sulfur.dioxide
                           + citric.acid + fixed.acidity + density + sulphates
                           + chlorides, data = winequality, nbest = 2)
summary(model_subsets)$which
```

```
## (Intercept) residual.sugar volatile.acidity alcohol total.sulfur.dioxide
## 1          TRUE          FALSE          FALSE          TRUE          FALSE
## 1          TRUE          FALSE          TRUE          FALSE          FALSE
## 2          TRUE          FALSE          TRUE          TRUE          FALSE
## 2          TRUE          FALSE          FALSE          TRUE          FALSE
## 3          TRUE          FALSE          TRUE          TRUE          FALSE
```



```

## 3      TRUE      FALSE      TRUE      TRUE      TRUE
## 4      TRUE      FALSE      TRUE      TRUE      TRUE
## 4      TRUE      FALSE      TRUE      TRUE      FALSE
## 5      TRUE      FALSE      TRUE      TRUE      TRUE
## 5      TRUE      FALSE      TRUE      TRUE      FALSE
## 6      TRUE      FALSE      TRUE      TRUE      TRUE
## 6      TRUE      FALSE      TRUE      TRUE      TRUE
## 7      TRUE      FALSE      TRUE      TRUE      TRUE
## 7      TRUE      FALSE      TRUE      TRUE      TRUE
## 8      TRUE      FALSE      TRUE      TRUE      TRUE
## 8      TRUE      TRUE      TRUE      TRUE      TRUE
##      pH free.sulfur.dioxide citric.acid fixed.acidity density sulphates
## 1 FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## 1 FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## 2 FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## 2 FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
## 3 FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
## 3 FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## 4 FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
## 4 FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
## 5 FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
## 5 TRUE       FALSE      FALSE      FALSE      FALSE      TRUE
## 6 TRUE       FALSE      FALSE      FALSE      FALSE      TRUE
## 6 FALSE      FALSE      FALSE      TRUE      FALSE      TRUE
## 7 TRUE       TRUE      FALSE      FALSE      FALSE      TRUE
## 7 TRUE       FALSE      TRUE      FALSE      FALSE      TRUE
## 8 TRUE       TRUE      TRUE      FALSE      FALSE      TRUE
## 8 TRUE       TRUE      FALSE      FALSE      FALSE      TRUE
##      chlorides
## 1      FALSE
## 1      FALSE
## 2      FALSE
## 2      FALSE
## 3      FALSE
## 3      FALSE
## 4      FALSE
## 4      TRUE
## 5      TRUE
## 5      TRUE
## 6      TRUE
## 6      TRUE
## 7      TRUE
## 7      TRUE
## 8      TRUE
## 8      TRUE

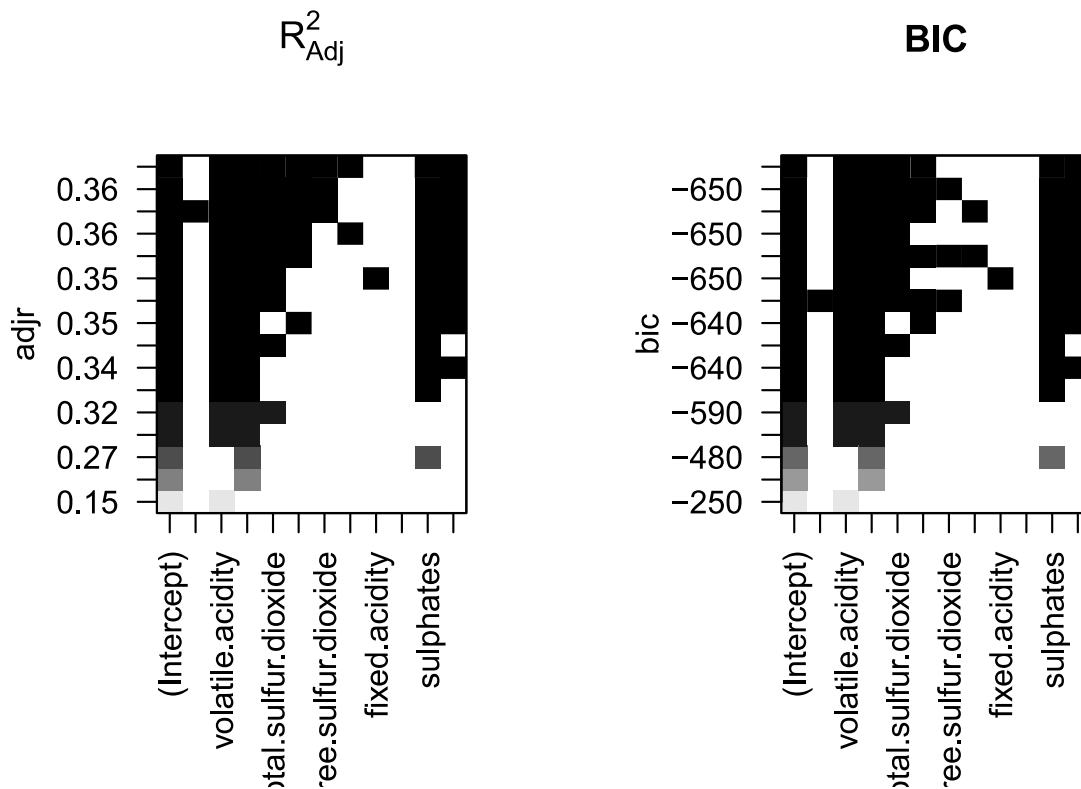
```

In order to understand easier the results, we can plot them:

```

par(mfrow = c(1,2))
plot(model_subsets, scale = "adjr", main=expression(R[Adj]^2))
plot(model_subsets, scale = "bic", main = "BIC")

```



In these plots, each line represents a model with a different set of variables, if we want the best subsets, we need to select the first row:

In order to do that, first of all we need to select get indices where these two measures get the maximum or minimum value

```
# for r2 adjusted we need to get the index of the maximum value
which(summary(model_subsets)$adjr2 ==max(summary(model_subsets)$adjr2))
```

```
## [1] 15
```

```
# for BIC criteria we want the minimum
which(summary(model_subsets)$bic ==min(summary(model_subsets)$bic))
```

```
## [1] 11
```

Once we have the indices, we can finally get the models

- Best model for  $R^2_{Adj}$

```
summary(model_subsets)$which[15,]
```

```
##      (Intercept)      residual.sugar      volatile.acidity
##             TRUE             FALSE             TRUE
##      alcohol total.sulfur.dioxide             pH
##             TRUE             TRUE             TRUE
## free.sulfur.dioxide      citric.acid      fixed.acidity
##             TRUE             TRUE             FALSE
##      density      sulphates      chlorides
##             FALSE             TRUE             TRUE
```

```
model_adj2 <- lm(quality~volatile.acidity + alcohol + total.sulfur.dioxide + pH
                + free.sulfur.dioxide + citric.acid + sulphates + chlorides,
                data = winequality)
summary(model_adj2)
```

```
##
## Call:
## lm(formula = quality ~ volatile.acidity + alcohol + total.sulfur.dioxide +
##     pH + free.sulfur.dioxide + citric.acid + sulphates + chlorides,
##     data = winequality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.66890 -0.37044 -0.04474  0.45697  2.02363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.6680876   0.4608410   10.129 < 2e-16 ***
## volatile.acidity -1.0736123   0.1159362   -9.260 < 2e-16 ***
## alcohol         0.2928780   0.0171280   17.099 < 2e-16 ***
## total.sulfur.dioxide -0.0033658   0.0006954   -4.840 1.42e-06 ***
## pH             -0.5491501   0.1331350   -4.125 3.90e-05 ***
## free.sulfur.dioxide  0.0047601   0.0021463    2.218  0.0267 *
## citric.acid      -0.1295444   0.1217717   -1.064  0.2876
## sulphates        0.8914283   0.1102122    8.088 1.19e-15 ***
## chlorides       -1.9494185   0.4026906   -4.841 1.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6477 on 1590 degrees of freedom
## Multiple R-squared:  0.3599, Adjusted R-squared:  0.3567
## F-statistic: 111.8 on 8 and 1590 DF, p-value: < 2.2e-16
```

The model we just obtained, has the same Adjusted  $R^2$ , but one variable more (**citric acid**), since this variable is more dispersed similar with less variance, and to prevent over-fitting and make the model less complex we should not consider this model in comparison to the model we found before

- Best model for BIC

```
summary(model_subsets)$which[11, ]
```

```
##      (Intercept)      residual.sugar      volatile.acidity
##              TRUE              FALSE              TRUE
##      alcohol total.sulfur.dioxide              pH
##              TRUE              TRUE              TRUE
## free.sulfur.dioxide      citric.acid      fixed.acidity
##              FALSE              FALSE              FALSE
##      density      sulphates      chlorides
##              FALSE              TRUE              TRUE
```

```
model_BIC <- lm(quality~volatile.acidity + alcohol + total.sulfur.dioxide + pH + sulphates +
                chlorides, data = winequality)
summary(model_BIC)
```

```
##
## Call:
## lm(formula = quality ~ volatile.acidity + alcohol + total.sulfur.dioxide +
##     pH + sulphates + chlorides, data = winequality)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.60575 -0.35883 -0.04806  0.46079  1.95643
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.2957316   0.3995603   10.751 < 2e-16 ***
## volatile.acidity -1.0381945   0.1004270  -10.338 < 2e-16 ***
## alcohol         0.2906738   0.0168108   17.291 < 2e-16 ***
## total.sulfur.dioxide -0.0023721  0.0005064  -4.684 3.05e-06 ***
## pH             -0.4351830   0.1160368   -3.750 0.000183 ***
## sulphates       0.8886802   0.1100419    8.076 1.31e-15 ***
## chlorides      -2.0022839   0.3980757   -5.030 5.46e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6487 on 1592 degrees of freedom
## Multiple R-squared:  0.3572, Adjusted R-squared:  0.3548
## F-statistic: 147.4 on 6 and 1592 DF,  p-value: < 2.2e-16
```

Now, the model hasn't the **fixed sulfur dioxide** variable, and in general has even lower p-values, since the **fixed sulfur dioxide** variable in the other model has the higher p-value, we can consider remove it, but, even that, removing this variable can help to reduce the complexity of the model and prevent over fitting, as we saw in exercise 2, with different selection method with different criteria we get the same set of variables, which probably indicates that all the variables of this set are significant to the prediction. So, at the moment, we can't arrive on a conclusion on which model select, a good idea might be test the two model with a validation data-set.

## References

[https://r-coder.com/boxplot-r/?utm\\_content=cmp-true](https://r-coder.com/boxplot-r/?utm_content=cmp-true)  
<https://www.geeksforgeeks.org/how-to-use-par-function-in-r/>  
<https://www.datacamp.com/tutorial/make-histogram-basic-r>