Assignment 4

Drew Murray

2022-12-19

```
wine = read.csv("C:/Users/dgmur/Downloads/wineFl2022.csv")
str(wine)
```

```
## 'data.frame':    350 obs. of  13 variables:
##  $ fixed_acidity       : num  5.9 7 6.9 9.9 6.7 7.6 8 6.5 6.8 12.2 ...
##  $ volatile_acidity    : num  0.32 0.23 0.24 0.49 0.16 0.78 0.18 0.22 0.28 0.45 ...
##  $ citric_acid         : num  0.2 0.42 0.49 0.23 0.28 0 0.37 0.27 0.37 0.49 ...
##  $ residual_sugar      : num  14.4 1.1 1.3 2.4 2.5 1.7 0.9 1.6 7 1.4 ...
##  $ chlorides           : num  0.05 0.062 0.032 0.087 0.046 0.076 0.049 0.039 0.057 0.075 ...
##  $ free_sulfur_dioxide : num  29 35 35 19 40 33 36 36 35 3 ...
##  $ total_sulfur_dioxide: num  144 100 148 115 153 45 109 116 208 6 ...
##  $ density             : num  0.997 0.993 0.993 0.995 0.992 ...
##  $ pH                  : num  3.24 3.04 3.45 2.77 3.38 3.31 2.89 3.38 3.57 3.13 ...
##  $ sulphates           : num  0.41 0.4 0.57 0.44 0.51 0.62 0.44 0.57 0.55 0.63 ...
##  $ alcohol             : num  10.3 9.2 10.7 9.4 11.4 10.7 12.7 11 10.2 10.4 ...
##  $ quality             : int  6 5 7 6 7 6 6 7 5 5 ...
##  $ type                : int  0 0 0 0 0 1 1 0 0 1 ...
```

```
wine$type=as.factor(wine$type)
wine$quality=as.numeric(wine$quality)
winec = wine[1:11]
View(winec)
#1

#a
cor(winec)
```

```
##                       fixed_acidity volatile_acidity citric_acid residual_sugar
## fixed_acidity           1.00000000      0.225020854  0.38344924    -0.09864763
## volatile_acidity        0.22502085      1.000000000 -0.32286866    -0.17139713
## citric_acid             0.38344924     -0.322868655  1.00000000     0.16225541
## residual_sugar         -0.09864763     -0.171397128  0.16225541     1.00000000
## chlorides               0.32756202      0.341374902  0.08681857    -0.12546264
## free_sulfur_dioxide    -0.28237859     -0.287109795  0.04245716     0.43035646
## total_sulfur_dioxide   -0.36919701     -0.358818093  0.10598165     0.47333598
## density                 0.53262405      0.244588159  0.19580930     0.54020970
## pH                     -0.23369013      0.228314657 -0.27738639    -0.22678197
## sulphates               0.30713463      0.186344437  0.10589346    -0.19561141
## alcohol                -0.20003736      0.004018921 -0.09567631    -0.34337904
##                        chlorides free_sulfur_dioxide total_sulfur_dioxide
## fixed_acidity         0.32756202        -0.282378587           -0.3691970
## volatile_acidity      0.34137490        -0.287109795           -0.3588181
## citric_acid           0.08681857         0.042457162            0.1059817
## residual_sugar       -0.12546264         0.430356464            0.4733360
## chlorides             1.00000000        -0.248831028           -0.3270541
## free_sulfur_dioxide  -0.24883103         1.000000000            0.7626948
## total_sulfur_dioxide -0.32705409         0.762694847            1.0000000
## density               0.37847916         0.005382083           -0.0229437
## pH                    0.01866647        -0.128416387           -0.1705793
## sulphates             0.39260836        -0.147085345           -0.2853768
## alcohol              -0.27335277        -0.101607695           -0.1933318
##                           density           pH    sulphates      alcohol
## fixed_acidity         0.532624054 -0.233690126  0.307134630 -0.200037357
## volatile_acidity      0.244588159  0.228314657  0.186344437  0.004018921
## citric_acid           0.195809304 -0.277386386  0.105893456 -0.095676312
## residual_sugar        0.540209698 -0.226781965 -0.195611413 -0.343379037
## chlorides             0.378479157  0.018666467  0.392608361 -0.273352774
## free_sulfur_dioxide   0.005382083 -0.128416387 -0.147085345 -0.101607695
## total_sulfur_dioxide -0.022943702 -0.170579290 -0.285376799 -0.193331790
## density               1.000000000 -0.005440951  0.228176692 -0.717527825
## pH                   -0.005440951  1.000000000  0.169653803  0.129038264
## sulphates             0.228176692  0.169653803  1.000000000  0.008272161
## alcohol              -0.717527825  0.129038264  0.008272161  1.000000000
```

```
cov(winec)
```

```
##                       fixed_acidity volatile_acidity   citric_acid
## fixed_acidity          1.890202538     0.0509847073  7.466255e-02
## volatile_acidity       0.050984707     0.0271597634 -7.535804e-03
## citric_acid            0.074662546    -0.0075358043  2.005772e-02
## residual_sugar        -0.642850594    -0.1338861850  1.089204e-01
## chlorides              0.014041393     0.0017541110  3.833680e-04
## free_sulfur_dioxide   -7.137701187    -0.8699282440  1.105514e-01
## total_sulfur_dioxide -29.064296357    -3.3859831355  8.594466e-01
## density                0.002214716     0.0001219106  8.387203e-05
## pH                    -0.047560426     0.0055699075 -5.815366e-03
## sulphates              0.061132820     0.0044460147  2.171208e-03
## alcohol               -0.326777023     0.0007869709 -1.610020e-02
##                       residual_sugar     chlorides free_sulfur_dioxide
## fixed_acidity           -0.642850594  1.404139e-02       -7.137701e+00
## volatile_acidity        -0.133886185  1.754111e-03       -8.699282e-01
## citric_acid              0.108920385  3.833680e-04        1.105514e-01
## residual_sugar          22.466657798 -1.854156e-02        3.750334e+01
## chlorides               -0.018541555  9.721314e-04       -1.426394e-01
## free_sulfur_dioxide     37.503344249 -1.426394e-01        3.380219e+02
## total_sulfur_dioxide   128.465413426 -5.838880e-01        8.029171e+02
## density                  0.007744167  3.569008e-05        2.992719e-04
## pH                      -0.159121531  8.615407e-05       -3.494975e-01
## sulphates               -0.134231641  1.772203e-03       -3.915014e-01
## alcohol                 -1.933878769 -1.012680e-02       -2.219654e+00
##                       total_sulfur_dioxide       density            pH
## fixed_acidity                -2.906430e+01  2.214716e-03 -4.756043e-02
## volatile_acidity             -3.385983e+00  1.219106e-04  5.569907e-03
## citric_acid                   8.594466e-01  8.387203e-05 -5.815366e-03
## residual_sugar                1.284654e+02  7.744167e-03 -1.591215e-01
## chlorides                    -5.838880e-01  3.569008e-05  8.615407e-05
## free_sulfur_dioxide           8.029171e+02  2.992719e-04 -3.494975e-01
## total_sulfur_dioxide          3.278649e+03 -3.973325e-03 -1.445856e+00
## density                      -3.973325e-03  9.147159e-06 -2.435954e-06
## pH                           -1.445856e+00 -2.435954e-06  2.191303e-02
## sulphates                    -2.365689e+00  9.990929e-05  3.635853e-03
## alcohol                      -1.315337e+01 -2.578506e-03  2.269635e-02
##                          sulphates       alcohol
## fixed_acidity         6.113282e-02 -3.267770e-01
## volatile_acidity      4.446015e-03  7.869709e-04
## citric_acid           2.171208e-03 -1.610020e-02
## residual_sugar       -1.342316e-01 -1.933879e+00
## chlorides             1.772203e-03 -1.012680e-02
## free_sulfur_dioxide  -3.915014e-01 -2.219654e+00
## total_sulfur_dioxide -2.365689e+00 -1.315337e+01
## density               9.990929e-05 -2.578506e-03
## pH                    3.635853e-03  2.269635e-02
## sulphates             2.095960e-02  1.422973e-03
## alcohol               1.422973e-03  1.411798e+00
```

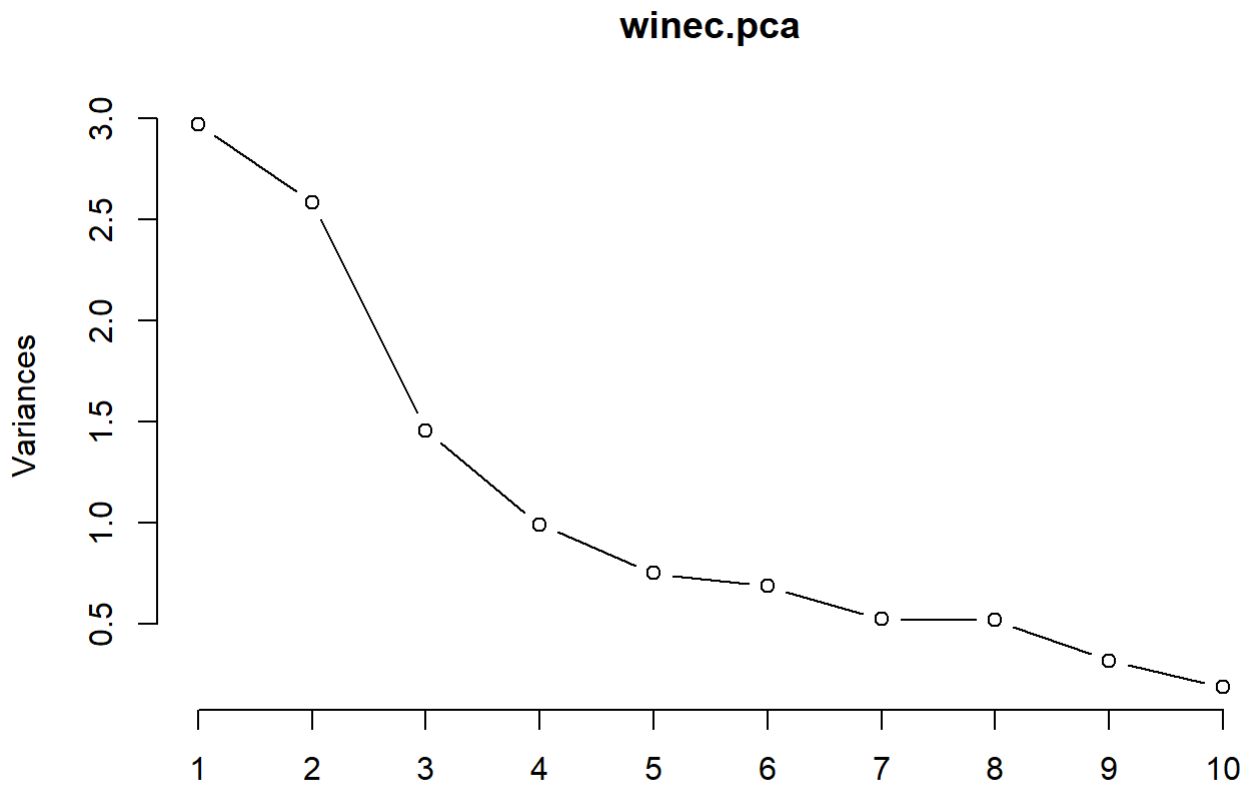
```
#b

#I chose correlation because the variables are on a different scale

winec.pca = prcomp(winec, scale. = T)

winec.pca
```

```
## Standard deviations (1, .., p=11):
##  [1] 1.7232494 1.6076239 1.2068558 0.9946672 0.8652731 0.8279042 0.7221999
##  [8] 0.7184746 0.5619289 0.4299401 0.1660712
##
## Rotation (n x k) = (11 x 11):
##                              PC1         PC2         PC3         PC4
## fixed_acidity         0.348489546 -0.29235631  0.32295067 -0.09387512
## volatile_acidity      0.343846756  0.03276812 -0.40392763 -0.17954875
## citric_acid          -0.009760994 -0.28502639  0.57385827  0.24915131
## residual_sugar       -0.267480445 -0.40830325 -0.18326601 -0.09808649
## chlorides             0.369862176 -0.19645234 -0.08349223  0.17270436
## free_sulfur_dioxide  -0.418729255 -0.17289180 -0.14492089  0.36002435
## total_sulfur_dioxide -0.472907822 -0.18332739 -0.11478151  0.21541693
## density               0.185040901 -0.53684289 -0.23345742 -0.05441968
## pH                    0.127515499  0.22004395 -0.46478704  0.39420304
## sulphates             0.318732272 -0.07236050  0.02559981  0.70297463
## alcohol              -0.028613894  0.47069391  0.23940149  0.16991977
##                              PC5         PC6         PC7         PC8
## fixed_acidity         0.04334508 -0.35893889  0.08376151 -0.42807410
## volatile_acidity      0.39351946 -0.44374763  0.34170306  0.03221217
## citric_acid          -0.26312144 -0.08990380  0.48010365  0.20787468
## residual_sugar       -0.12195280 -0.30847793 -0.25710155  0.59045465
## chlorides             0.35390879  0.51829635  0.30728852  0.42378092
## free_sulfur_dioxide   0.35796425 -0.13909626  0.15758820 -0.22149241
## total_sulfur_dioxide  0.19238345  0.01155103  0.23717318 -0.17675022
## density              -0.20539242 -0.15702906 -0.08509948 -0.01480819
## pH                   -0.61576992 -0.08600367  0.32724636 -0.05467107
## sulphates             0.14303805 -0.07636788 -0.53667765 -0.04357396
## alcohol               0.16412954 -0.49498534  0.04190356  0.39858458
##                              PC9        PC10         PC11
## fixed_acidity         0.40906345  0.27322565  0.342065272
## volatile_acidity     -0.43991846 -0.12679298  0.078278568
## citric_acid          -0.35278986 -0.22642878 -0.022744994
## residual_sugar        0.07573828  0.05299393  0.430373805
## chlorides             0.31160684  0.14293522  0.048809367
## free_sulfur_dioxide   0.36690764 -0.53679357 -0.002643888
## total_sulfur_dioxide -0.24778764  0.70221161 -0.063510466
## density               0.11286958  0.02068696 -0.729810143
## pH                    0.17856128  0.07100579  0.174135418
## sulphates            -0.27646554  0.02552777  0.071039626
## alcohol               0.30784128  0.21769828 -0.341627469
```

```
#c
summary(winec.pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     1.723 1.6076 1.2069 0.99467 0.86527 0.82790 0.72220
## Proportion of Variance 0.270 0.2349 0.1324 0.08994 0.06806 0.06231 0.04742
## Cumulative Proportion  0.270 0.5049 0.6373 0.72726 0.79533 0.85764 0.90505
##                            PC8     PC9   PC10    PC11
## Standard deviation     0.71847 0.56193 0.4299 0.16607
## Proportion of Variance 0.04693 0.02871 0.0168 0.00251
## Cumulative Proportion  0.95198 0.98069 0.9975 1.00000
```

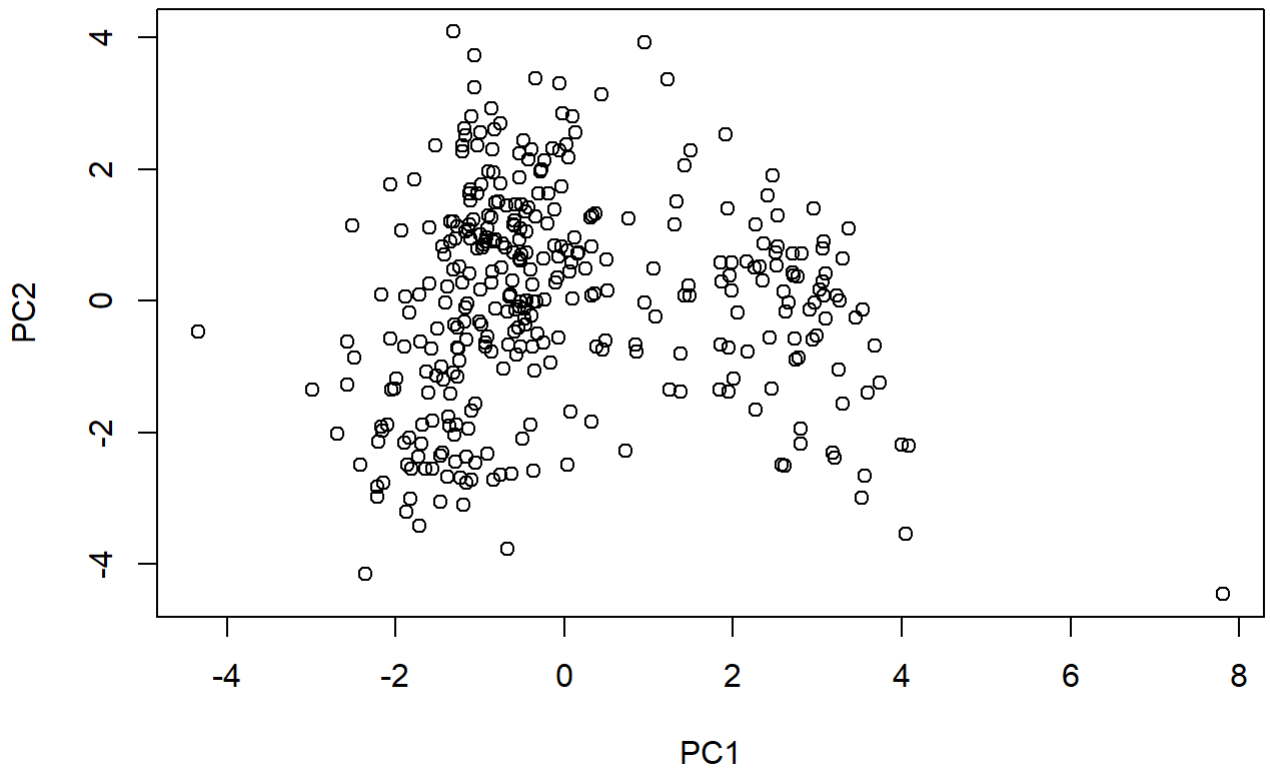
```
screeplot(winec.pca, type = "lines")
```



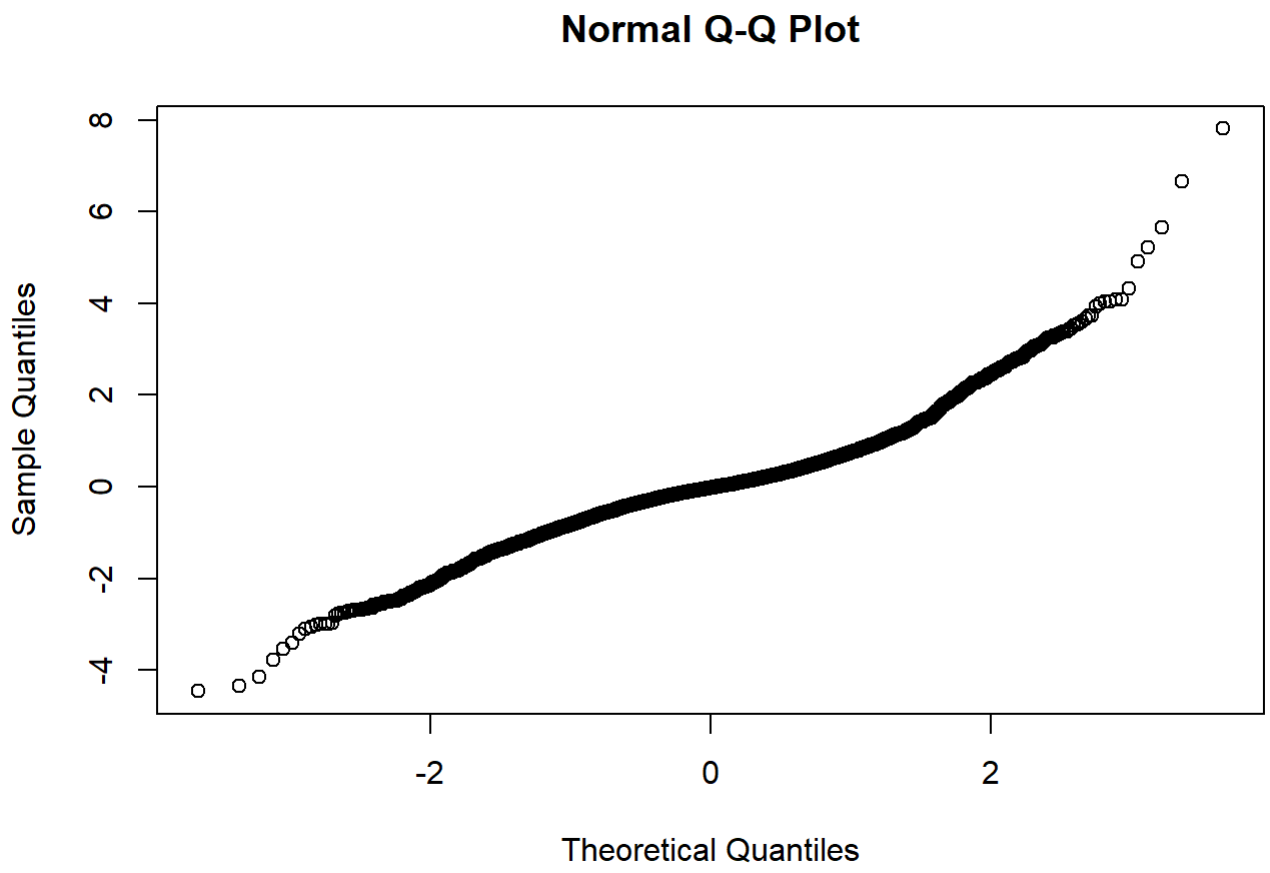

```
#10 PCs

#d

c=cor(winec, winec.pca$x)
round(c,2)
```

```
##                         PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9
## fixed_acidity          0.60 -0.47  0.39 -0.09  0.04 -0.30  0.06 -0.31  0.23
## volatile_acidity       0.59  0.05 -0.49 -0.18  0.34 -0.37  0.25  0.02 -0.25
## citric_acid           -0.02 -0.46  0.69  0.25 -0.23 -0.07  0.35  0.15 -0.20
## residual_sugar        -0.46 -0.66 -0.22 -0.10 -0.11 -0.26 -0.19  0.42  0.04
## chlorides              0.64 -0.32 -0.10  0.17  0.31  0.43  0.22  0.30  0.18
## free_sulfur_dioxide   -0.72 -0.28 -0.17  0.36  0.31 -0.12  0.11 -0.16  0.21
## total_sulfur_dioxide -0.81 -0.29 -0.14  0.21  0.17  0.01  0.17 -0.13 -0.14
## density                0.32 -0.86 -0.28 -0.05 -0.18 -0.13 -0.06 -0.01  0.06
## pH                     0.22  0.35 -0.56  0.39 -0.53 -0.07  0.24 -0.04  0.10
## sulphates              0.55 -0.12  0.03  0.70  0.12 -0.06 -0.39 -0.03 -0.16
## alcohol               -0.05  0.76  0.29  0.17  0.14 -0.41  0.03  0.29  0.17
##                         PC10  PC11
## fixed_acidity          0.12  0.06
## volatile_acidity      -0.05  0.01
## citric_acid           -0.10  0.00
## residual_sugar         0.02  0.07
## chlorides              0.06  0.01
## free_sulfur_dioxide   -0.23  0.00
## total_sulfur_dioxide   0.30 -0.01
## density                0.01 -0.12
## pH                     0.03  0.03
## sulphates              0.01  0.01
## alcohol                0.09 -0.06
```

```
#PC1: Sulfur Dioxide

#PC2: Density of Alcohol by amount of sugar

#PC3: PH level of acidity

#PC4: PH level of sulfur containing bases

#PC5: Ph level of volatile acidity and choloride

#PC6: Chlorides residual sugar and acidity based on alcohol percantage

#PC7: PH level of volatile, citric acid and sulphate

#PC8: fixed acidity and residual sugar based on percentage of alcohol

#PCA9: acidity, chlorides and sulfur dioxide

#PCA10: Sulfur dioxide
#e

#No, the proportion of variance is too low, and too close to PC2. PC1 has a 27% proportion of va
riance,while PC has 23.4%

#F

plot(winec.pca$x[,1:2])
```




```
#There are no distingiuishing groups, however, there is an outlier that is skewed to the higher end of PC1

#bonus

qqnorm(winec.pca$x)
```




```
#The plot shows that the distribution is not normal, this shows evidence of high variation.


#2

#a
wined = wine[-13]

cor(wined)
```

```
##                       fixed_acidity volatile_acidity citric_acid residual_sugar
## fixed_acidity           1.00000000      0.225020854  0.38344924    -0.09864763
## volatile_acidity        0.22502085      1.000000000 -0.32286866    -0.17139713
## citric_acid             0.38344924     -0.322868655  1.00000000     0.16225541
## residual_sugar         -0.09864763     -0.171397128  0.16225541     1.00000000
## chlorides               0.32756202      0.341374902  0.08681857    -0.12546264
## free_sulfur_dioxide    -0.28237859     -0.287109795  0.04245716     0.43035646
## total_sulfur_dioxide   -0.36919701     -0.358818093  0.10598165     0.47333598
## density                 0.53262405      0.244588159  0.19580930     0.54020970
## pH                     -0.23369013      0.228314657 -0.27738639    -0.22678197
## sulphates               0.30713463      0.186344437  0.10589346    -0.19561141
## alcohol                -0.20003736      0.004018921 -0.09567631    -0.34337904
## quality                -0.11163094     -0.234230528  0.02093266    -0.01517474
##                        chlorides free_sulfur_dioxide total_sulfur_dioxide
## fixed_acidity         0.32756202        -0.282378587          -0.36919701
## volatile_acidity      0.34137490        -0.287109795          -0.35881809
## citric_acid           0.08681857         0.042457162           0.10598165
## residual_sugar       -0.12546264         0.430356464           0.47333598
## chlorides             1.00000000        -0.248831028          -0.32705409
## free_sulfur_dioxide  -0.24883103         1.000000000           0.76269485
## total_sulfur_dioxide -0.32705409         0.762694847           1.00000000
## density               0.37847916         0.005382083          -0.02294370
## pH                    0.01866647        -0.128416387          -0.17057929
## sulphates             0.39260836        -0.147085345          -0.28537680
## alcohol              -0.27335277        -0.101607695          -0.19333179
## quality              -0.21776159         0.020542135          -0.01026426
##                           density           pH    sulphates      alcohol
## fixed_acidity         0.532624054 -0.233690126  0.307134630 -0.200037357
## volatile_acidity      0.244588159  0.228314657  0.186344437  0.004018921
## citric_acid           0.195809304 -0.277386386  0.105893456 -0.095676312
## residual_sugar        0.540209698 -0.226781965 -0.195611413 -0.343379037
## chlorides             0.378479157  0.018666467  0.392608361 -0.273352774
## free_sulfur_dioxide   0.005382083 -0.128416387 -0.147085345 -0.101607695
## total_sulfur_dioxide -0.022943702 -0.170579290 -0.285376799 -0.193331790
## density               1.000000000 -0.005440951  0.228176692 -0.717527825
## pH                   -0.005440951  1.000000000  0.169653803  0.129038264
## sulphates             0.228176692  0.169653803  1.000000000  0.008272161
## alcohol              -0.717527825  0.129038264  0.008272161  1.000000000
## quality              -0.278927117  0.019394617  0.124581349  0.396693357
##                          quality
## fixed_acidity        -0.11163094
## volatile_acidity     -0.23423053
## citric_acid           0.02093266
## residual_sugar       -0.01517474
## chlorides            -0.21776159
## free_sulfur_dioxide   0.02054213
## total_sulfur_dioxide -0.01026426
## density              -0.27892712
## pH                    0.01939462
## sulphates             0.12458135
```

```
## alcohol              0.39669336
## quality              1.00000000
```

```
cov(wined)
```

```
##                       fixed_acidity volatile_acidity   citric_acid
## fixed_acidity          1.890202538     0.0509847073  7.466255e-02
## volatile_acidity       0.050984707     0.0271597634 -7.535804e-03
## citric_acid            0.074662546    -0.0075358043  2.005772e-02
## residual_sugar        -0.642850594    -0.1338861850  1.089204e-01
## chlorides              0.014041393     0.0017541110  3.833680e-04
## free_sulfur_dioxide   -7.137701187    -0.8699282440  1.105514e-01
## total_sulfur_dioxide -29.064296357    -3.3859831355  8.594466e-01
## density                0.002214716     0.0001219106  8.387203e-05
## pH                    -0.047560426     0.0055699075 -5.815366e-03
## sulphates              0.061132820     0.0044460147  2.171208e-03
## alcohol               -0.326777023     0.0007869709 -1.610020e-02
## quality               -0.132358576    -0.0332904625  2.556693e-03
##                       residual_sugar     chlorides free_sulfur_dioxide
## fixed_acidity          -0.642850594  1.404139e-02       -7.137701e+00
## volatile_acidity       -0.133886185  1.754111e-03       -8.699282e-01
## citric_acid             0.108920385  3.833680e-04        1.105514e-01
## residual_sugar         22.466657798 -1.854156e-02        3.750334e+01
## chlorides              -0.018541555  9.721314e-04       -1.426394e-01
## free_sulfur_dioxide    37.503344249 -1.426394e-01        3.380219e+02
## total_sulfur_dioxide  128.465413426 -5.838880e-01        8.029171e+02
## density                 0.007744167  3.569008e-05        2.992719e-04
## pH                     -0.159121531  8.615407e-05       -3.494975e-01
## sulphates              -0.134231641  1.772203e-03       -3.915014e-01
## alcohol                -1.933878769 -1.012680e-02       -2.219654e+00
## quality                -0.062030291 -5.855407e-03        3.257102e-01
##                       total_sulfur_dioxide       density            pH
## fixed_acidity               -2.906430e+01  2.214716e-03 -4.756043e-02
## volatile_acidity            -3.385983e+00  1.219106e-04  5.569907e-03
## citric_acid                  8.594466e-01  8.387203e-05 -5.815366e-03
## residual_sugar               1.284654e+02  7.744167e-03 -1.591215e-01
## chlorides                   -5.838880e-01  3.569008e-05  8.615407e-05
## free_sulfur_dioxide          8.029171e+02  2.992719e-04 -3.494975e-01
## total_sulfur_dioxide         3.278649e+03 -3.973325e-03 -1.445856e+00
## density                     -3.973325e-03  9.147159e-06 -2.435954e-06
## pH                          -1.445856e+00 -2.435954e-06  2.191303e-02
## sulphates                   -2.365689e+00  9.990929e-05  3.635853e-03
## alcohol                     -1.315337e+01 -2.578506e-03  2.269635e-02
## quality                     -5.068604e-01 -7.275238e-04  2.475972e-03
##                          sulphates       alcohol       quality
## fixed_acidity         6.113282e-02 -3.267770e-01 -0.1323585755
## volatile_acidity      4.446015e-03  7.869709e-04 -0.0332904625
## citric_acid           2.171208e-03 -1.610020e-02  0.0025566926
## residual_sugar       -1.342316e-01 -1.933879e+00 -0.0620302906
## chlorides             1.772203e-03 -1.012680e-02 -0.0058554073
## free_sulfur_dioxide  -3.915014e-01 -2.219654e+00  0.3257101924
## total_sulfur_dioxide -2.365689e+00 -1.315337e+01 -0.5068604175
## density               9.990929e-05 -2.578506e-03 -0.0007275238
## pH                    3.635853e-03  2.269635e-02  0.0024759722
## sulphates             2.095960e-02  1.422973e-03  0.0155545641
```

```
## alcohol              1.422973e-03  1.411798e+00  0.4064943375
## quality              1.555456e-02  4.064943e-01  0.7437494883
```

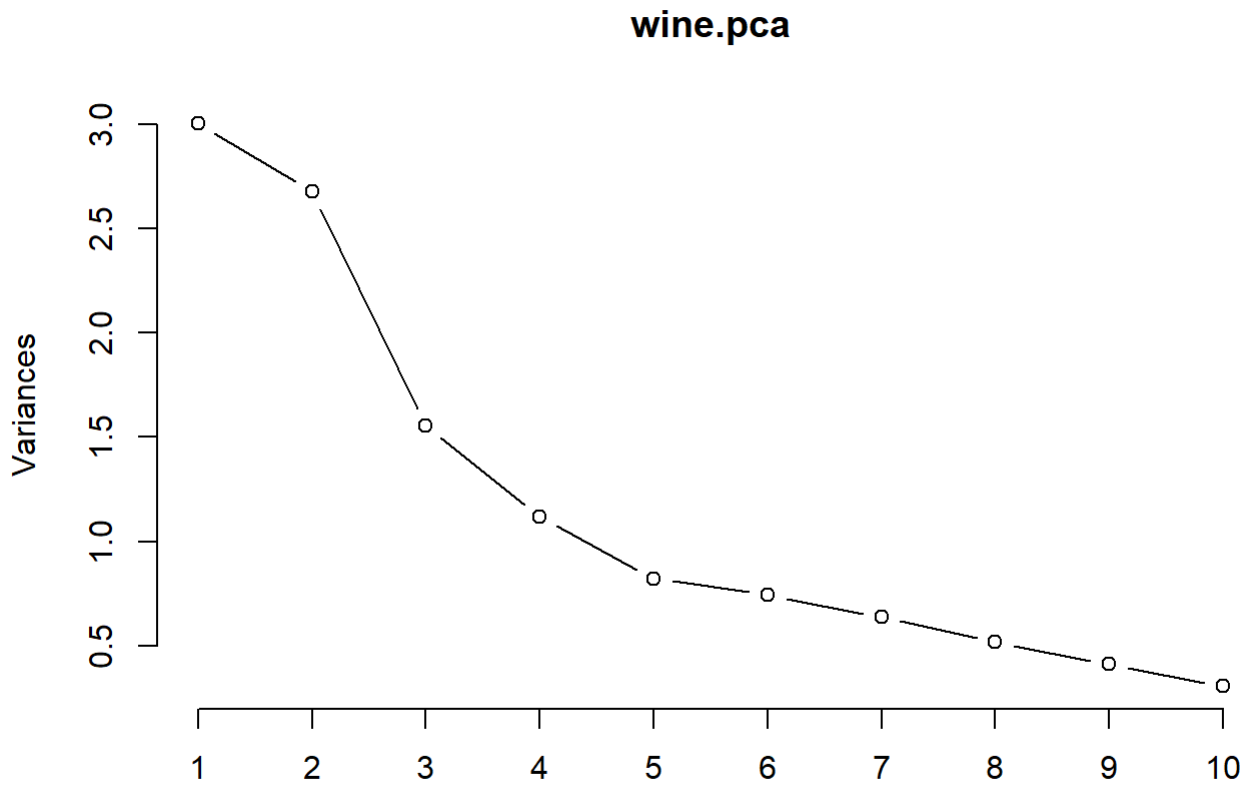
```
#b I chose correlation maxtrix because the variables are not one the same scale.

wine.pca = prcomp(wined,scale. = T)

wine.pca
```

```
## Standard deviations (1, .., p=12):
##  [1] 1.7324119 1.6357004 1.2467205 1.0580945 0.9067586 0.8621958 0.7981909
##  [8] 0.7194678 0.6417402 0.5537722 0.4277998 0.1658754
##
## Rotation (n x k) = (12 x 12):
##                              PC1         PC2         PC3          PC4
## fixed_acidity         0.37384753 -0.20920790  0.34944193  0.101481689
## volatile_acidity      0.34605826  0.06241715 -0.38344695 -0.006200858
## citric_acid           0.01550334 -0.24883481  0.57469924  0.007293069
## residual_sugar       -0.21073912 -0.43076891 -0.08045098 -0.163775376
## chlorides             0.39301292 -0.13510339 -0.03755022 -0.114788780
## free_sulfur_dioxide  -0.38691015 -0.23592001 -0.09789778 -0.287638032
## total_sulfur_dioxide -0.43615175 -0.25936172 -0.09887064 -0.145936568
## density               0.25687987 -0.49174628 -0.11498773 -0.167868006
## pH                    0.10042460  0.22184714 -0.39264147 -0.448034762
## sulphates             0.30735851  0.01054705  0.17042324 -0.613979534
## alcohol              -0.10540151  0.47361058  0.21246419 -0.082411454
## quality              -0.14068511  0.21960665  0.36299199 -0.482597698
##                              PC5         PC6         PC7         PC8
## fixed_acidity         0.13882760 -0.12231102  0.37554679 -0.32337939
## volatile_acidity      0.17854081 -0.49137727  0.39081629  0.20269632
## citric_acid          -0.21716706  0.30102439  0.34520552  0.42157398
## residual_sugar        0.45985782 -0.04057706 -0.03621810  0.36955879
## chlorides            -0.38370342 -0.19562045 -0.44540196  0.52910531
## free_sulfur_dioxide  -0.29102109 -0.32440834  0.22851724 -0.10881479
## total_sulfur_dioxide -0.25210986 -0.13491619  0.13311385 -0.02804903
## density               0.29496975  0.11660006  0.04826955 -0.05610048
## pH                   -0.02773031  0.61105097  0.31137342  0.12066866
## sulphates            -0.27804516 -0.13216386 -0.04388815 -0.32078443
## alcohol               0.05413015 -0.26854548  0.36266713  0.35593643
## quality               0.47615444 -0.11424921 -0.29101433  0.01265819
##                              PC9        PC10        PC11         PC12
## fixed_acidity         0.29643825  0.3518858  0.27428184 -0.342475359
## volatile_acidity      0.11151836 -0.4705296 -0.15021984 -0.075668433
## citric_acid           0.02304073 -0.3421192 -0.23390324  0.023508793
## residual_sugar       -0.41382212  0.1504179  0.07075169 -0.432262885
## chlorides             0.26011021  0.2458745  0.13589718 -0.047862076
## free_sulfur_dioxide   0.20186465  0.3555843 -0.53021442  0.003308181
## total_sulfur_dioxide  0.20364509 -0.3097345  0.68535808  0.063560810
## density               0.01043673  0.1040176  0.02101411  0.730305663
## pH                    0.20849284  0.1460274  0.07056923 -0.174230911
## sulphates            -0.51367297 -0.1692788  0.04347695 -0.073442865
## alcohol              -0.24893466  0.3759888  0.24899992  0.337711305
## quality               0.45675359 -0.1735244 -0.06215342  0.009685102
```

```
#c
summary(wine.pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4     PC5     PC6     PC7
## Standard deviation     1.7324 1.6357 1.2467 1.0581 0.90676 0.86220 0.79819
## Proportion of Variance 0.2501 0.2230 0.1295 0.0933 0.06852 0.06195 0.05309
## Cumulative Proportion  0.2501 0.4731 0.6026 0.6959 0.76440 0.82635 0.87945
##                            PC8     PC9    PC10    PC11    PC12
## Standard deviation     0.71947 0.64174 0.55377 0.42780 0.16588
## Proportion of Variance 0.04314 0.03432 0.02556 0.01525 0.00229
## Cumulative Proportion  0.92258 0.95690 0.98246 0.99771 1.00000
```

```
screeplot(wine.pca, type="lines")
```



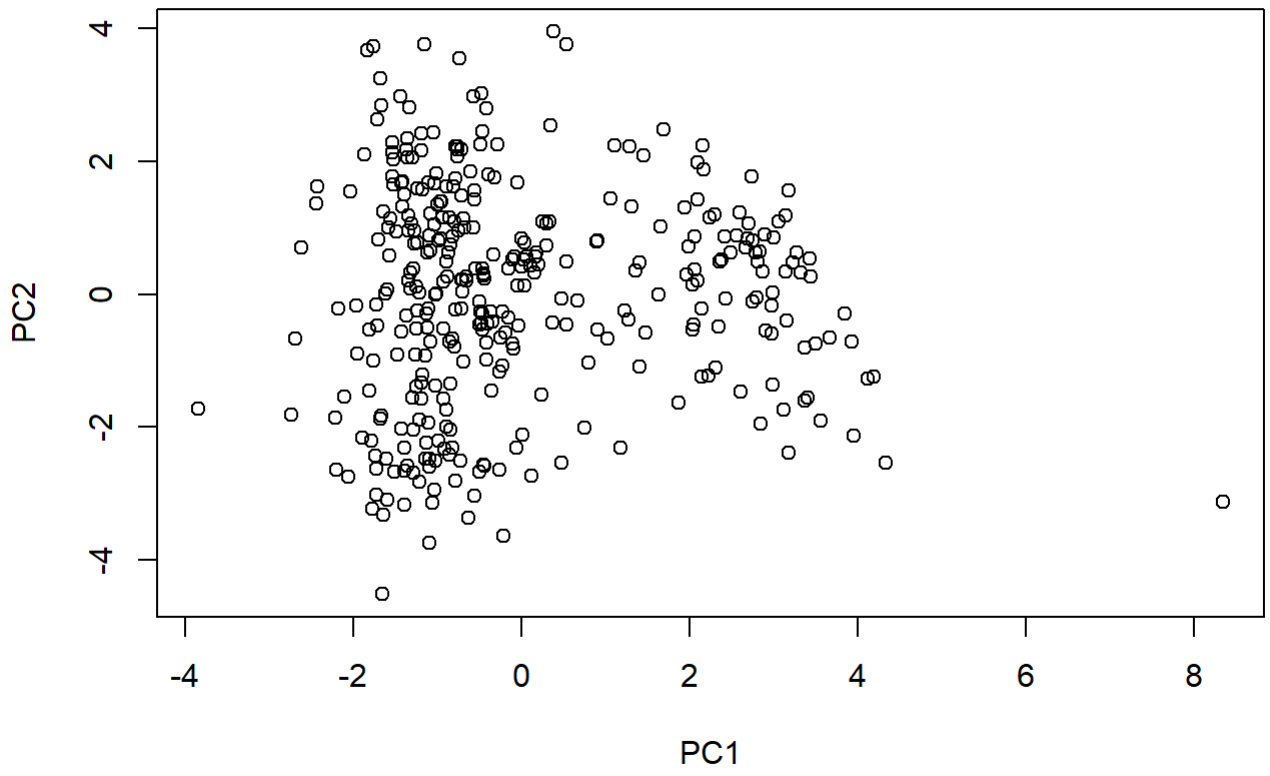

```
#d

#10 PCs, however, I could get away with 9 PCs in comparison to question 1 due to PC9 in PC2 has a lower variance explained

#e
d=cor(wined, wine.pca$x)

round(d,2)
```

```
##                         PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9
## fixed_acidity          0.65 -0.34  0.44  0.11  0.13 -0.11  0.30 -0.23  0.19
## volatile_acidity       0.60  0.10 -0.48 -0.01  0.16 -0.42  0.31  0.15  0.07
## citric_acid            0.03 -0.41  0.72  0.01 -0.20  0.26  0.28  0.30  0.01
## residual_sugar        -0.37 -0.70 -0.10 -0.17  0.42 -0.03 -0.03  0.27 -0.27
## chlorides              0.68 -0.22 -0.05 -0.12 -0.35 -0.17 -0.36  0.38  0.17
## free_sulfur_dioxide   -0.67 -0.39 -0.12 -0.30 -0.26 -0.28  0.18 -0.08  0.13
## total_sulfur_dioxide -0.76 -0.42 -0.12 -0.15 -0.23 -0.12  0.11 -0.02  0.13
## density                0.45 -0.80 -0.14 -0.18  0.27  0.10  0.04 -0.04  0.01
## pH                     0.17  0.36 -0.49 -0.47 -0.03  0.53  0.25  0.09  0.13
## sulphates              0.53  0.02  0.21 -0.65 -0.25 -0.11 -0.04 -0.23 -0.33
## alcohol               -0.18  0.77  0.26 -0.09  0.05 -0.23  0.29  0.26 -0.16
## quality               -0.24  0.36  0.45 -0.51  0.43 -0.10 -0.23  0.01  0.29
##                         PC10  PC11  PC12
## fixed_acidity          0.19  0.12 -0.06
## volatile_acidity      -0.26 -0.06 -0.01
## citric_acid           -0.19 -0.10  0.00
## residual_sugar         0.08  0.03 -0.07
## chlorides              0.14  0.06 -0.01
## free_sulfur_dioxide    0.20 -0.23  0.00
## total_sulfur_dioxide -0.17  0.29  0.01
## density                0.06  0.01  0.12
## pH                     0.08  0.03 -0.03
## sulphates             -0.09  0.02 -0.01
## alcohol                0.21  0.11  0.06
## quality               -0.10 -0.03  0.00
```

```
#PC1: Wine type based of acidity, chlorides and sulfur dioxide

#PC2: Amount of density based on percentage of alcohol and residual sugar

#PC3: Acidity based on PH level and quality

#PC4: Amount of sulphates based on quality and PH level

#PC5: Quality based on Residual sugar and chlorides

#PC6: Average Ph level based on percentage of  volatile acidity and free sulfur dioxide

#PC7: PH level of acidity

#PC8: Total Amount of residual sugars, chlorides and citric acid based on percentage of alcohol

#PC9: Quality based Sulphates and residual sugar

#PC10: Amount of Volatile Acidity based on quality of alchohol

#Yes, my intepretations do differ, categorical variable dominated the PCs

#F

summary(wine.pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4     PC5     PC6     PC7
## Standard deviation     1.7324 1.6357 1.2467 1.0581 0.90676 0.86220 0.79819
## Proportion of Variance 0.2501 0.2230 0.1295 0.0933 0.06852 0.06195 0.05309
## Cumulative Proportion  0.2501 0.4731 0.6026 0.6959 0.76440 0.82635 0.87945
##                           PC8     PC9    PC10    PC11    PC12
## Standard deviation     0.71947 0.64174 0.55377 0.42780 0.16588
## Proportion of Variance 0.04314 0.03432 0.02556 0.01525 0.00229
## Cumulative Proportion  0.92258 0.95690 0.98246 0.99771 1.00000
```

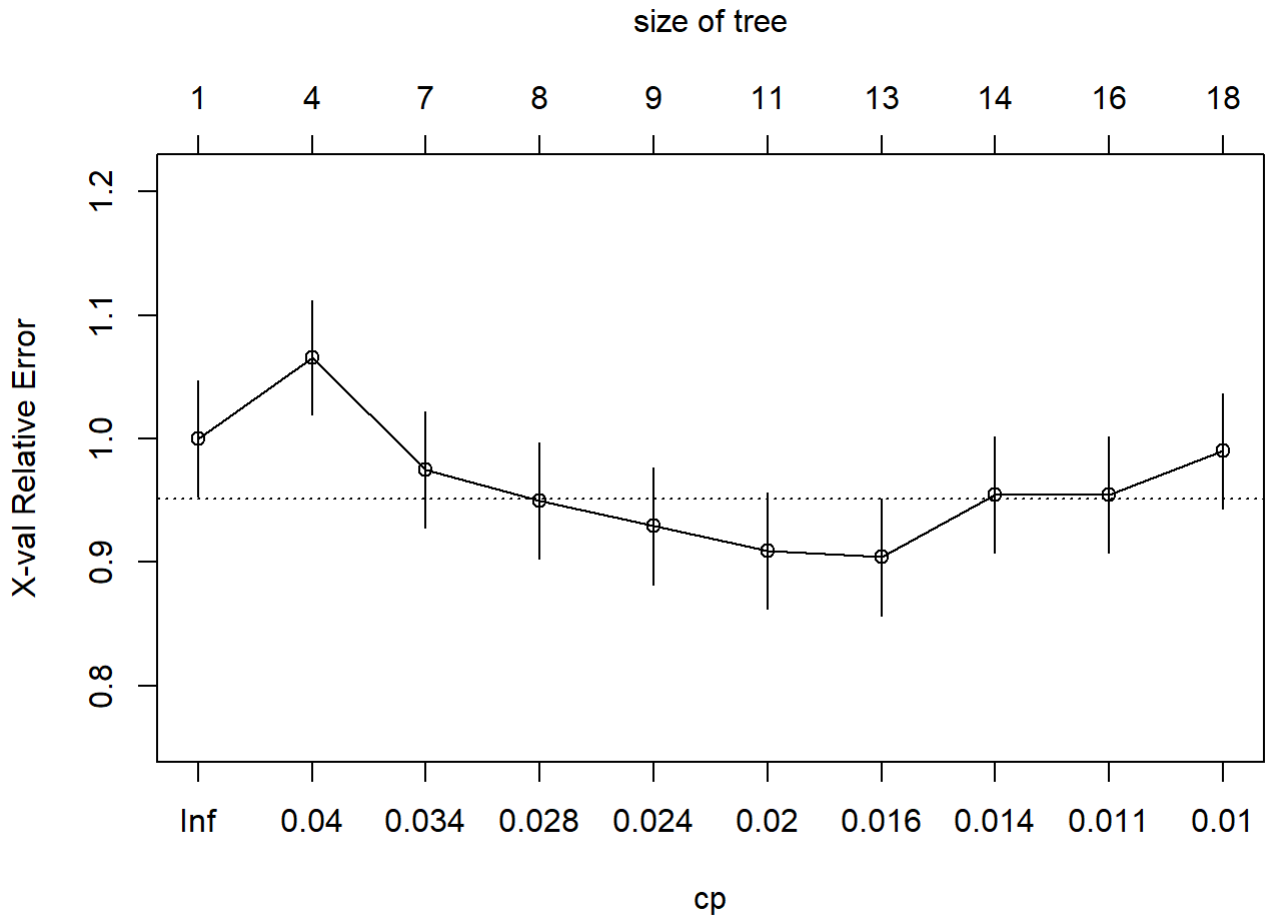
```
#No, the proportion of variance is too close in proportion of variance to PC2


#G

plot(wine.pca$x[,1:2])
```




```
#There are distinguishable groups, and there is one outlier on the PC1 axis. There is a differen
ce compared to question one because the groups are distinuishable


#Bonus
#I would prefer with categorical variables because it more easier to interpret relationships in
each PC. An alternative method would be creating subsets for the data based factors of the categ
orical variables, and then use PCA for each subset.


#3

library(rpart)
library(rpart.plot)



#a
wine$quality=as.factor(wine$quality)

fit= rpart(quality~., method = "class", data=wine)

printcp(fit)
```

```
##
## Classification tree:
## rpart(formula = quality ~ ., data = wine, method = "class")
##
## Variables actually used in tree construction:
## [1] alcohol              chlorides            density
## [4] fixed_acidity        pH                   sulphates
## [7] total_sulfur_dioxide volatile_acidity
##
## Root node error: 198/350 = 0.56571
##
## n= 350
##
##          CP nsplit rel error  xerror     xstd
## 1  0.043771      0   1.00000 1.00000 0.046833
## 2  0.037037      3   0.86869 1.06566 0.046233
## 3  0.030303      6   0.75758 0.97475 0.046993
## 4  0.025253      7   0.72727 0.94949 0.047113
## 5  0.022727      8   0.70202 0.92929 0.047181
## 6  0.017677     10   0.65657 0.90909 0.047224
## 7  0.015152     12   0.62121 0.90404 0.047231
## 8  0.012626     13   0.60606 0.95455 0.047092
## 9  0.010101     15   0.58081 0.95455 0.047092
## 10 0.010000     17   0.56061 0.98990 0.046902
```

```
plotcp(fit)
```


size of tree
1
4
7
8
9
11
13
14
16
18
X-val Relative Error
0.8
0.9
1.0
1.1
1.2
Inf
0.04
0.034
0.028
0.024
0.02
0.016
0.014
0.011
0.01
cp

```
summary(fit)
```

```
## Call:
## rpart(formula = quality ~ ., data = wine, method = "class")
##   n= 350
##
##            CP nsplit rel error    xerror       xstd
## 1  0.04377104      0 1.0000000 1.0000000 0.04683334
## 2  0.03703704      3 0.8686869 1.0656566 0.04623272
## 3  0.03030303      6 0.7575758 0.9747475 0.04699258
## 4  0.02525253      7 0.7272727 0.9494949 0.04711262
## 5  0.02272727      8 0.7020202 0.9292929 0.04718063
## 6  0.01767677     10 0.6565657 0.9090909 0.04722386
## 7  0.01515152     12 0.6212121 0.9040404 0.04723081
## 8  0.01262626     13 0.6060606 0.9545455 0.04709173
## 9  0.01010101     15 0.5808081 0.9545455 0.04709173
## 10 0.01000000     17 0.5606061 0.9898990 0.04690176
##
## Variable importance
##              alcohol              density total_sulfur_dioxide
##                   18                   13                   13
##            sulphates            chlorides        fixed_acidity
##                   12                   11                    8
##     volatile_acidity  free_sulfur_dioxide       residual_sugar
##                    6                    4                    4
##                 type                   pH          citric_acid
##                    4                    4                    2
##
## Node number 1: 350 observations,    complexity param=0.04377104
##   predicted class=6  expected loss=0.5657143  P(node) =1
##     class counts:    14   108   152    67     9
##    probabilities: 0.040 0.309 0.434 0.191 0.026
##   left son=2 (218 obs) right son=3 (132 obs)
##   Primary splits:
##       alcohol             < 10.75    to the left,  improve=20.929650, (0 missing)
##       density             < 0.99222  to the right, improve=10.996840, (0 missing)
##       chlorides           < 0.0445   to the right, improve= 7.709710, (0 missing)
##       free_sulfur_dioxide < 17.5     to the left,  improve= 3.759377, (0 missing)
##       volatile_acidity    < 0.455    to the right, improve= 3.660000, (0 missing)
##   Surrogate splits:
##       density        < 0.992935 to the right, agree=0.820, adj=0.523, (0 split)
##       chlorides      < 0.0365   to the right, agree=0.731, adj=0.288, (0 split)
##       fixed_acidity  < 6.05     to the right, agree=0.649, adj=0.068, (0 split)
##       residual_sugar < 1.25     to the right, agree=0.649, adj=0.068, (0 split)
##       sulphates      < 0.335    to the right, agree=0.646, adj=0.061, (0 split)
##
## Node number 2: 218 observations,    complexity param=0.04377104
##   predicted class=6  expected loss=0.559633  P(node) =0.6228571
##     class counts:    11    95    96    11     5
##    probabilities: 0.050 0.436 0.440 0.050 0.023
##   left son=4 (209 obs) right son=5 (9 obs)
##   Primary splits:
##       total_sulfur_dioxide < 220      to the left,  improve=4.775559, (0 missing)
##       volatile_acidity     < 0.2275   to the right, improve=4.748625, (0 missing)
```

```
##       density              < 0.99369  to the left,  improve=2.768169, (0 missing)
##       pH                   < 3.055    to the right, improve=2.679951, (0 missing)
##       sulphates            < 0.375    to the left,  improve=2.469719, (0 missing)
##
## Node number 3: 132 observations,    complexity param=0.03703704
##   predicted class=6  expected loss=0.5757576  P(node) =0.3771429
##     class counts:     3    13    56    56     4
##    probabilities: 0.023 0.098 0.424 0.424 0.030
##   left son=6 (125 obs) right son=7 (7 obs)
##   Primary splits:
##       chlorides            < 0.023    to the right, improve=3.863152, (0 missing)
##       total_sulfur_dioxide < 80.5     to the left,  improve=3.237374, (0 missing)
##       volatile_acidity     < 0.485    to the right, improve=3.213442, (0 missing)
##       sulphates            < 0.375    to the left,  improve=3.043466, (0 missing)
##       free_sulfur_dioxide  < 15.5     to the left,  improve=2.932734, (0 missing)
##
## Node number 4: 209 observations,    complexity param=0.04377104
##   predicted class=5  expected loss=0.5454545  P(node) =0.5971429
##     class counts:    11    95    87    11     5
##    probabilities: 0.053 0.455 0.416 0.053 0.024
##   left son=8 (154 obs) right son=9 (55 obs)
##   Primary splits:
##       volatile_acidity     < 0.2275   to the right, improve=5.501982, (0 missing)
##       sulphates            < 0.555    to the left,  improve=3.023857, (0 missing)
##       pH                   < 3.295    to the left,  improve=2.873411, (0 missing)
##       total_sulfur_dioxide < 22.5     to the left,  improve=2.214736, (0 missing)
##       density              < 0.99369  to the left,  improve=2.144072, (0 missing)
##   Surrogate splits:
##       residual_sugar      < 14       to the left,  agree=0.770, adj=0.127, (0 split)
##       alcohol             < 8.95     to the right, agree=0.756, adj=0.073, (0 split)
##       free_sulfur_dioxide < 75.5     to the left,  agree=0.742, adj=0.018, (0 split)
##       density             < 0.999405 to the left,  agree=0.742, adj=0.018, (0 split)
##
## Node number 5: 9 observations
##   predicted class=6  expected loss=0  P(node) =0.02571429
##     class counts:     0     0     9     0     0
##    probabilities: 0.000 0.000 1.000 0.000 0.000
##
## Node number 6: 125 observations,    complexity param=0.03703704
##   predicted class=6  expected loss=0.552  P(node) =0.3571429
##     class counts:     3    13    56    49     4
##    probabilities: 0.024 0.104 0.448 0.392 0.032
##   left son=12 (32 obs) right son=13 (93 obs)
##   Primary splits:
##       total_sulfur_dioxide < 80.5     to the left,  improve=2.975253, (0 missing)
##       chlorides            < 0.0485   to the right, improve=2.779764, (0 missing)
##       residual_sugar       < 1.55     to the left,  improve=2.767268, (0 missing)
##       volatile_acidity     < 0.485    to the right, improve=2.697455, (0 missing)
##       free_sulfur_dioxide  < 14       to the left,  improve=2.552000, (0 missing)
##   Surrogate splits:
##       type                splits as  RL,           agree=0.928, adj=0.719, (0 split)
##       chlorides           < 0.0585   to the right, agree=0.896, adj=0.594, (0 split)
```

```
##       free_sulfur_dioxide < 18.5     to the left,  agree=0.896, adj=0.594, (0 split)
##       density             < 0.99372  to the right, agree=0.864, adj=0.469, (0 split)
##       fixed_acidity       < 8.05     to the right, agree=0.840, adj=0.375, (0 split)
##
## Node number 7: 7 observations
##   predicted class=7  expected loss=0  P(node) =0.02
##     class counts:     0     0     0     7     0
##    probabilities: 0.000 0.000 0.000 1.000 0.000
##
## Node number 8: 154 observations,    complexity param=0.03030303
##   predicted class=5  expected loss=0.4675325  P(node) =0.44
##     class counts:     9    82    56     5     2
##    probabilities: 0.058 0.532 0.364 0.032 0.013
##   left son=16 (88 obs) right son=17 (66 obs)
##   Primary splits:
##       sulphates            < 0.555    to the left,  improve=3.987013, (0 missing)
##       alcohol              < 9.85     to the left,  improve=3.692264, (0 missing)
##       total_sulfur_dioxide < 174      to the right, improve=2.783800, (0 missing)
##       fixed_acidity        < 6.25     to the right, improve=2.257234, (0 missing)
##       free_sulfur_dioxide  < 36.5     to the right, improve=1.979978, (0 missing)
##   Surrogate splits:
##       type                 splits as  LR,           agree=0.721, adj=0.348, (0 split)
##       total_sulfur_dioxide < 92.5     to the right, agree=0.714, adj=0.333, (0 split)
##       chlorides            < 0.069    to the left,  agree=0.675, adj=0.242, (0 split)
##       fixed_acidity        < 8.05     to the left,  agree=0.662, adj=0.212, (0 split)
##       free_sulfur_dioxide  < 18.5     to the right, agree=0.630, adj=0.136, (0 split)
##
## Node number 9: 55 observations,    complexity param=0.01262626
##   predicted class=6  expected loss=0.4363636  P(node) =0.1571429
##     class counts:     2    13    31     6     3
##    probabilities: 0.036 0.236 0.564 0.109 0.055
##   left son=18 (37 obs) right son=19 (18 obs)
##   Primary splits:
##       pH                   < 3.275    to the left,  improve=3.896970, (0 missing)
##       fixed_acidity        < 7.05     to the left,  improve=3.797332, (0 missing)
##       sulphates            < 0.525    to the right, improve=3.431768, (0 missing)
##       total_sulfur_dioxide < 137.5    to the left,  improve=3.131276, (0 missing)
##       chlorides            < 0.039    to the right, improve=2.855303, (0 missing)
##   Surrogate splits:
##       sulphates            < 0.565    to the left,  agree=0.782, adj=0.333, (0 split)
##       volatile_acidity     < 0.1175   to the right, agree=0.709, adj=0.111, (0 split)
##       residual_sugar       < 1.7      to the right, agree=0.709, adj=0.111, (0 split)
##       total_sulfur_dioxide < 167.5    to the left,  agree=0.709, adj=0.111, (0 split)
##       density              < 0.99908  to the left,  agree=0.709, adj=0.111, (0 split)
##
## Node number 12: 32 observations,    complexity param=0.01010101
##   predicted class=6  expected loss=0.4375  P(node) =0.09142857
##     class counts:     1     7    18     6     0
##    probabilities: 0.031 0.219 0.562 0.188 0.000
##   left son=24 (14 obs) right son=25 (18 obs)
##   Primary splits:
##       sulphates           < 0.605    to the left,  improve=3.393849, (0 missing)
```

```
##       free_sulfur_dioxide < 14       to the left,  improve=2.988366, (0 missing)
##       density             < 0.9938   to the left,  improve=2.598128, (0 missing)
##       chlorides           < 0.0695   to the left,  improve=2.473214, (0 missing)
##       volatile_acidity    < 0.325    to the right, improve=1.687500, (0 missing)
##   Surrogate splits:
##       density              < 0.9938   to the left,  agree=0.844, adj=0.643, (0 split)
##       chlorides            < 0.0565   to the left,  agree=0.781, adj=0.500, (0 split)
##       type                 splits as  LR,           agree=0.781, adj=0.500, (0 split)
##       fixed_acidity        < 6.7      to the left,  agree=0.750, adj=0.429, (0 split)
##       total_sulfur_dioxide < 37.5     to the right, agree=0.719, adj=0.357, (0 split)
##
## Node number 13: 93 observations,    complexity param=0.03703704
##   predicted class=7  expected loss=0.5376344  P(node) =0.2657143
##     class counts:     2     6    38    43     4
##    probabilities: 0.022 0.065 0.409 0.462 0.043
##   left son=26 (70 obs) right son=27 (23 obs)
##   Primary splits:
##       sulphates            < 0.565    to the left,  improve=5.525893, (0 missing)
##       residual_sugar       < 1.55     to the left,  improve=3.730862, (0 missing)
##       density              < 0.990985 to the right, improve=2.280156, (0 missing)
##       total_sulfur_dioxide < 153.5    to the right, improve=2.126679, (0 missing)
##       pH                   < 3.055    to the left,  improve=1.542593, (0 missing)
##   Surrogate splits:
##       citric_acid         < 0.165    to the right, agree=0.785, adj=0.130, (0 split)
##       alcohol             < 13.7     to the left,  agree=0.785, adj=0.130, (0 split)
##       volatile_acidity    < 0.115    to the right, agree=0.774, adj=0.087, (0 split)
##       free_sulfur_dioxide < 60.5     to the left,  agree=0.763, adj=0.043, (0 split)
##
## Node number 16: 88 observations,    complexity param=0.01515152
##   predicted class=5  expected loss=0.3636364  P(node) =0.2514286
##     class counts:     6    56    24     2     0
##    probabilities: 0.068 0.636 0.273 0.023 0.000
##   left son=32 (73 obs) right son=33 (15 obs)
##   Primary splits:
##       fixed_acidity        < 6.25     to the right, improve=3.040349, (0 missing)
##       total_sulfur_dioxide < 156.5    to the right, improve=2.786830, (0 missing)
##       alcohol              < 10.55    to the left,  improve=1.503987, (0 missing)
##       volatile_acidity     < 0.69     to the right, improve=1.448292, (0 missing)
##       residual_sugar       < 5.65     to the left,  improve=1.415388, (0 missing)
##   Surrogate splits:
##       density        < 0.99234  to the right, agree=0.875, adj=0.267, (0 split)
##       pH             < 3.395    to the left,  agree=0.852, adj=0.133, (0 split)
##       residual_sugar < 16.15    to the left,  agree=0.841, adj=0.067, (0 split)
##
## Node number 17: 66 observations,    complexity param=0.02525253
##   predicted class=6  expected loss=0.5151515  P(node) =0.1885714
##     class counts:     3    26    32     3     2
##    probabilities: 0.045 0.394 0.485 0.045 0.030
##   left son=34 (8 obs) right son=35 (58 obs)
##   Primary splits:
##       total_sulfur_dioxide < 22.5     to the left,  improve=2.452194, (0 missing)
##       citric_acid          < 0.425    to the right, improve=2.201154, (0 missing)
```

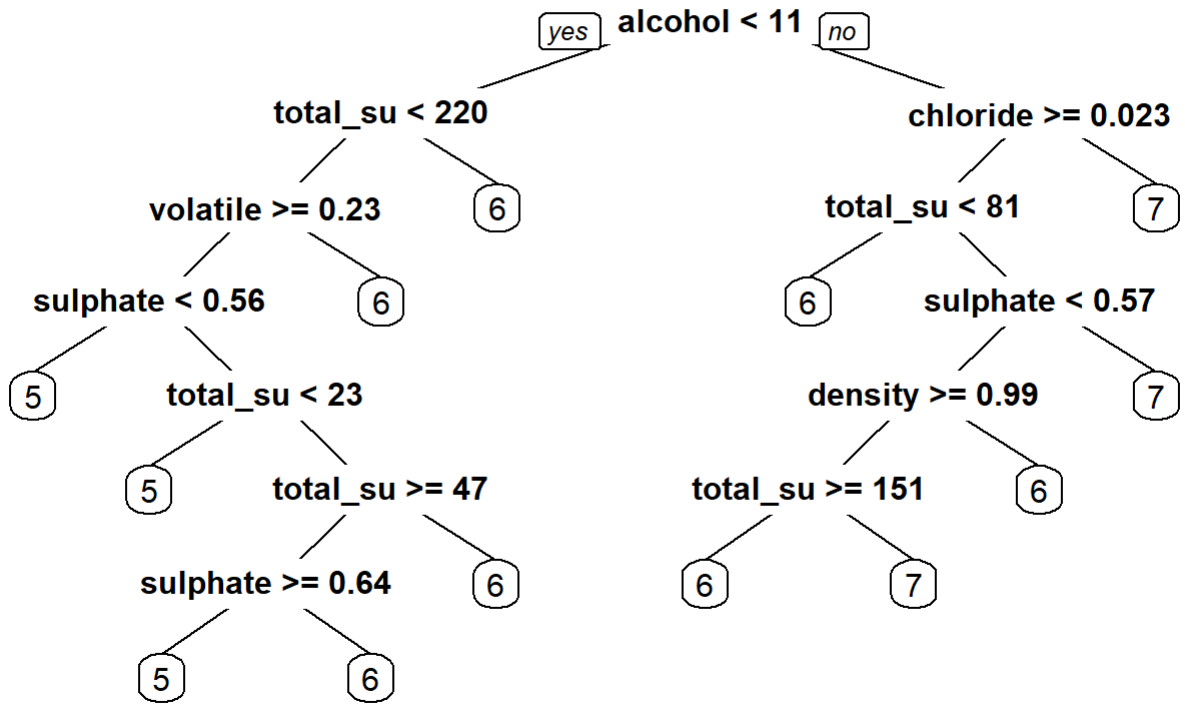
```
##       alcohol              < 9.85     to the left,  improve=1.836398, (0 missing)
##       sulphates            < 0.625    to the right, improve=1.601399, (0 missing)
##       chlorides            < 0.093    to the right, improve=1.566234, (0 missing)
##   Surrogate splits:
##       free_sulfur_dioxide < 4.5      to the left,  agree=0.909, adj=0.25, (0 split)
##
## Node number 18: 37 observations,    complexity param=0.01262626
##   predicted class=6  expected loss=0.5945946  P(node) =0.1057143
##     class counts:     2    12    15     5     3
##    probabilities: 0.054 0.324 0.405 0.135 0.081
##   left son=36 (15 obs) right son=37 (22 obs)
##   Primary splits:
##       fixed_acidity        < 6.85     to the left,  improve=2.872727, (0 missing)
##       sulphates            < 0.415    to the left,  improve=2.132867, (0 missing)
##       density              < 0.99495  to the right, improve=1.807692, (0 missing)
##       pH                   < 3.05     to the right, improve=1.767241, (0 missing)
##       total_sulfur_dioxide < 137.5    to the left,  improve=1.417647, (0 missing)
##   Surrogate splits:
##       sulphates   < 0.4      to the left,  agree=0.730, adj=0.333, (0 split)
##       density     < 0.992415 to the left,  agree=0.703, adj=0.267, (0 split)
##       citric_acid < 0.245    to the left,  agree=0.676, adj=0.200, (0 split)
##       pH          < 3.16     to the right, agree=0.676, adj=0.200, (0 split)
##       alcohol     < 10.45    to the right, agree=0.676, adj=0.200, (0 split)
##
## Node number 19: 18 observations
##   predicted class=6  expected loss=0.1111111  P(node) =0.05142857
##     class counts:     0     1    16     1     0
##    probabilities: 0.000 0.056 0.889 0.056 0.000
##
## Node number 24: 14 observations
##   predicted class=5  expected loss=0.5  P(node) =0.04
##     class counts:     1     7     5     1     0
##    probabilities: 0.071 0.500 0.357 0.071 0.000
##
## Node number 25: 18 observations
##   predicted class=6  expected loss=0.2777778  P(node) =0.05142857
##     class counts:     0     0    13     5     0
##    probabilities: 0.000 0.000 0.722 0.278 0.000
##
## Node number 26: 70 observations,    complexity param=0.02272727
##   predicted class=6  expected loss=0.5  P(node) =0.2
##     class counts:     2     5    35    25     3
##    probabilities: 0.029 0.071 0.500 0.357 0.043
##   left son=52 (43 obs) right son=53 (27 obs)
##   Primary splits:
##       density             < 0.990985 to the right, improve=2.926935, (0 missing)
##       alcohol             < 11.05    to the right, improve=2.385714, (0 missing)
##       chlorides           < 0.0345   to the left,  improve=2.035047, (0 missing)
##       residual_sugar      < 1.65     to the left,  improve=1.768627, (0 missing)
##       free_sulfur_dioxide < 22.5     to the right, improve=1.330939, (0 missing)
##   Surrogate splits:
##       alcohol              < 12.1     to the left,  agree=0.871, adj=0.667, (0 split)
```

```
##       residual_sugar       < 3.6      to the right, agree=0.771, adj=0.407, (0 split)
##       chlorides            < 0.0355   to the right, agree=0.729, adj=0.296, (0 split)
##       total_sulfur_dioxide < 112.5    to the right, agree=0.729, adj=0.296, (0 split)
##       pH                   < 3.03     to the right, agree=0.657, adj=0.111, (0 split)
##
## Node number 27: 23 observations
##   predicted class=7  expected loss=0.2173913  P(node) =0.06571429
##     class counts:     0     1     3    18     1
##    probabilities: 0.000 0.043 0.130 0.783 0.043
##
## Node number 32: 73 observations
##   predicted class=5  expected loss=0.3150685  P(node) =0.2085714
##     class counts:     6    50    15     2     0
##    probabilities: 0.082 0.685 0.205 0.027 0.000
##
## Node number 33: 15 observations
##   predicted class=6  expected loss=0.4  P(node) =0.04285714
##     class counts:     0     6     9     0     0
##    probabilities: 0.000 0.400 0.600 0.000 0.000
##
## Node number 34: 8 observations
##   predicted class=5  expected loss=0.25  P(node) =0.02285714
##     class counts:     0     6     1     0     1
##    probabilities: 0.000 0.750 0.125 0.000 0.125
##
## Node number 35: 58 observations,    complexity param=0.01767677
##   predicted class=6  expected loss=0.4655172  P(node) =0.1657143
##     class counts:     3    20    31     3     1
##    probabilities: 0.052 0.345 0.534 0.052 0.017
##   left son=70 (39 obs) right son=71 (19 obs)
##   Primary splits:
##       total_sulfur_dioxide < 47       to the right, improve=3.011215, (0 missing)
##       alcohol              < 9.85     to the left,  improve=2.359824, (0 missing)
##       free_sulfur_dioxide  < 8.5      to the right, improve=1.943591, (0 missing)
##       citric_acid          < 0.48     to the right, improve=1.685157, (0 missing)
##       sulphates            < 0.72     to the right, improve=1.605187, (0 missing)
##   Surrogate splits:
##       free_sulfur_dioxide < 11       to the right, agree=0.828, adj=0.474, (0 split)
##       chlorides           < 0.0715   to the left,  agree=0.793, adj=0.368, (0 split)
##       residual_sugar      < 2.3      to the right, agree=0.776, adj=0.316, (0 split)
##       volatile_acidity    < 0.525    to the left,  agree=0.759, adj=0.263, (0 split)
##       type                splits as  LR,           agree=0.759, adj=0.263, (0 split)
##
## Node number 36: 15 observations
##   predicted class=5  expected loss=0.4  P(node) =0.04285714
##     class counts:     1     9     4     0     1
##    probabilities: 0.067 0.600 0.267 0.000 0.067
##
## Node number 37: 22 observations
##   predicted class=6  expected loss=0.5  P(node) =0.06285714
##     class counts:     1     3    11     5     2
##    probabilities: 0.045 0.136 0.500 0.227 0.091
```

```
##
## Node number 52: 43 observations,    complexity param=0.02272727
##   predicted class=7  expected loss=0.5581395  P(node) =0.1228571
##     class counts:     2     5    16    19     1
##    probabilities: 0.047 0.116 0.372 0.442 0.023
##   left son=104 (16 obs) right son=105 (27 obs)
##   Primary splits:
##       total_sulfur_dioxide < 150.5    to the right, improve=2.620155, (0 missing)
##       fixed_acidity        < 5.95     to the left,  improve=1.924917, (0 missing)
##       density              < 0.99217  to the right, improve=1.820155, (0 missing)
##       alcohol              < 11.05    to the right, improve=1.820155, (0 missing)
##       pH                   < 3.135    to the right, improve=1.765610, (0 missing)
##   Surrogate splits:
##       fixed_acidity       < 7.05     to the right, agree=0.721, adj=0.25, (0 split)
##       volatile_acidity    < 0.42     to the right, agree=0.721, adj=0.25, (0 split)
##       residual_sugar      < 10.75    to the right, agree=0.721, adj=0.25, (0 split)
##       chlorides           < 0.049    to the right, agree=0.721, adj=0.25, (0 split)
##       free_sulfur_dioxide < 36.5     to the right, agree=0.721, adj=0.25, (0 split)
##
## Node number 53: 27 observations,    complexity param=0.01010101
##   predicted class=6  expected loss=0.2962963  P(node) =0.07714286
##     class counts:     0     0    19     6     2
##    probabilities: 0.000 0.000 0.704 0.222 0.074
##   left son=106 (18 obs) right son=107 (9 obs)
##   Primary splits:
##       alcohol              < 12.84667 to the left,  improve=3.370370, (0 missing)
##       free_sulfur_dioxide  < 37.5     to the left,  improve=1.919577, (0 missing)
##       density              < 0.98916  to the right, improve=1.411306, (0 missing)
##       total_sulfur_dioxide < 108      to the right, improve=1.348148, (0 missing)
##       residual_sugar       < 1.45     to the left,  improve=1.038258, (0 missing)
##   Surrogate splits:
##       volatile_acidity    < 0.36     to the left,  agree=0.815, adj=0.444, (0 split)
##       density             < 0.98916  to the right, agree=0.815, adj=0.444, (0 split)
##       fixed_acidity       < 5.55     to the right, agree=0.778, adj=0.333, (0 split)
##       residual_sugar      < 2.7      to the left,  agree=0.778, adj=0.333, (0 split)
##       free_sulfur_dioxide < 37.5     to the left,  agree=0.778, adj=0.333, (0 split)
##
## Node number 70: 39 observations,    complexity param=0.01767677
##   predicted class=5  expected loss=0.5641026  P(node) =0.1114286
##     class counts:     3    17    16     3     0
##    probabilities: 0.077 0.436 0.410 0.077 0.000
##   left son=140 (20 obs) right son=141 (19 obs)
##   Primary splits:
##       sulphates            < 0.635    to the right, improve=2.653576, (0 missing)
##       free_sulfur_dioxide  < 24.5     to the left,  improve=2.289817, (0 missing)
##       chlorides            < 0.0705   to the right, improve=2.289817, (0 missing)
##       total_sulfur_dioxide < 140.5    to the left,  improve=1.880769, (0 missing)
##       residual_sugar       < 5.55     to the left,  improve=1.788701, (0 missing)
##   Surrogate splits:
##       citric_acid         < 0.33     to the right, agree=0.718, adj=0.421, (0 split)
##       chlorides           < 0.043    to the right, agree=0.692, adj=0.368, (0 split)
##       pH                  < 3.355    to the left,  agree=0.641, adj=0.263, (0 split)
```

```
##       fixed_acidity       < 7.05     to the right, agree=0.615, adj=0.211, (0 split)
##       free_sulfur_dioxide < 21       to the left,  agree=0.615, adj=0.211, (0 split)
##
## Node number 71: 19 observations
##   predicted class=6  expected loss=0.2105263  P(node) =0.05428571
##     class counts:     0     3    15     0     1
##    probabilities: 0.000 0.158 0.789 0.000 0.053
##
## Node number 104: 16 observations
##   predicted class=6  expected loss=0.4375  P(node) =0.04571429
##     class counts:     1     2     9     3     1
##    probabilities: 0.062 0.125 0.562 0.188 0.062
##
## Node number 105: 27 observations
##   predicted class=7  expected loss=0.4074074  P(node) =0.07714286
##     class counts:     1     3     7    16     0
##    probabilities: 0.037 0.111 0.259 0.593 0.000
##
## Node number 106: 18 observations
##   predicted class=6  expected loss=0.1111111  P(node) =0.05142857
##     class counts:     0     0    16     1     1
##    probabilities: 0.000 0.000 0.889 0.056 0.056
##
## Node number 107: 9 observations
##   predicted class=7  expected loss=0.4444444  P(node) =0.02571429
##     class counts:     0     0     3     5     1
##    probabilities: 0.000 0.000 0.333 0.556 0.111
##
## Node number 140: 20 observations
##   predicted class=5  expected loss=0.35  P(node) =0.05714286
##     class counts:     1    13     6     0     0
##    probabilities: 0.050 0.650 0.300 0.000 0.000
##
## Node number 141: 19 observations
##   predicted class=6  expected loss=0.4736842  P(node) =0.05428571
##     class counts:     2     4    10     3     0
##    probabilities: 0.105 0.211 0.526 0.158 0.000
```

```
pfit<- prune(fit, cp=fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])

#b
prp(pfit,main = "Pruned Classification Tree for Wine Qualities")
```

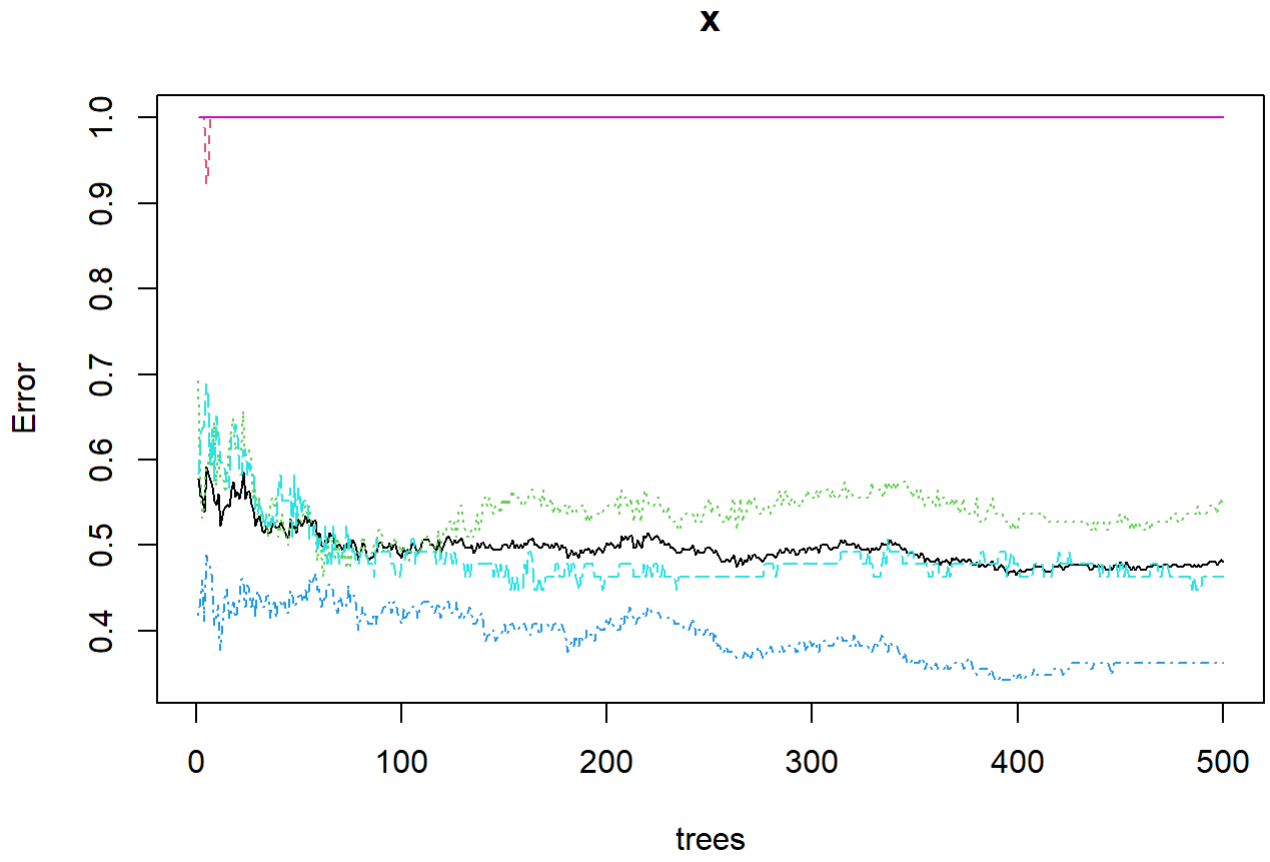



```
#The predicted the wine quality would be a 6

#c
library(randomForest)
```

```
## randomForest 4.7-1.1
```

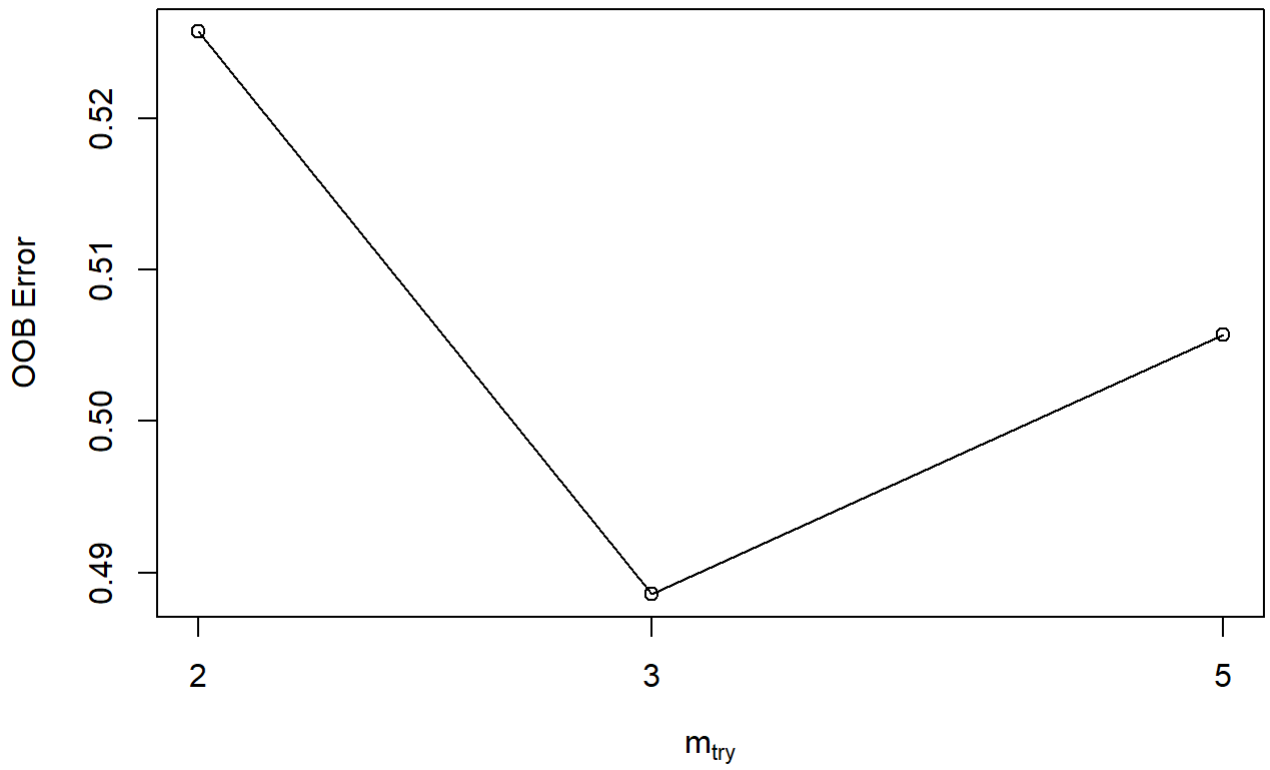
```
## Type rfNews() to see new features/changes/bug fixes.
```

```
set.seed(478)

wine$quality=as.factor(wine$quality)

x = randomForest(quality~., data=wine)

plot(x)
```




```
t = tuneRF(wine[,-12],wine[,12],
           stepFactor = 0.7, plot = TRUE,
           ntreeTry = 300,
           trace = TRUE,
           improve=1)
```

```
## mtry = 3  OOB error = 48.86%
## Searching left ...
## mtry = 5     OOB error = 50.57%
## -0.03508772 1
## Searching right ...
## mtry = 2     OOB error = 52.57%
## -0.07602339 1
```




```
x1 = randomForest(quality~., ntree=300,mtry=3,nodesize=2, data=wine)
print(x1)
```

```
##
## Call:
##  randomForest(formula = quality ~ ., data = wine, ntree = 300,      mtry = 3, nodesize = 2)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 50%
## Confusion matrix:
##   4  5  6  7 8 class.error
## 4 0  7  7  0 0   1.0000000
## 5 0 51 51  6 0   0.5277778
## 6 0 42 91 19 0   0.4013158
## 7 0  4 30 33 0   0.5074627
## 8 0  4  4  1 0   1.0000000
```

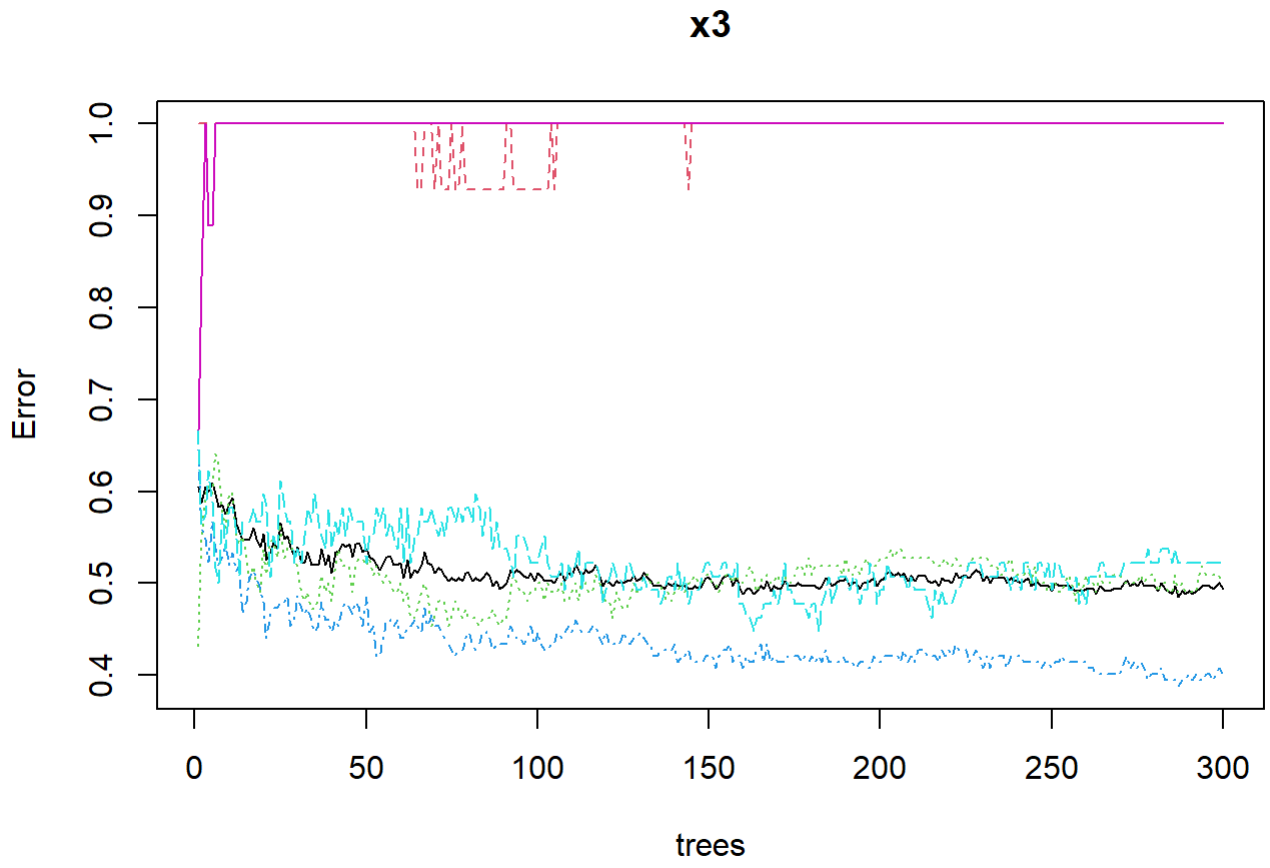
```
x2 = randomForest(quality~., ntree=300,mtry=2,nodesize=2,data=wine)
print(x2)
```

```
##
## Call:
##  randomForest(formula = quality ~ ., data = wine, ntree = 300,      mtry = 2, nodesize = 2)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 50.29%
## Confusion matrix:
##   4  5  6  7 8 class.error
## 4 0  7  6  1 0   1.0000000
## 5 0 49 55  4 0   0.5462963
## 6 0 38 93 21 0   0.3881579
## 7 0  4 31 32 0   0.5223881
## 8 0  4  3  2 0   1.0000000
```

```
x3 = randomForest(quality~., ntree=300,mtry=5,data=wine)
print(x3)
```

```
##
## Call:
##  randomForest(formula = quality ~ ., data = wine, ntree = 300,      mtry = 5)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 49.43%
## Confusion matrix:
##   4  5  6  7 8 class.error
## 4 0  7  7  0 0   1.0000000
## 5 0 54 48  6 0   0.5000000
## 6 0 36 91 25 0   0.4013158
## 7 0  3 32 32 0   0.5223881
## 8 0  4  4  1 0   1.0000000
```

```
plot(x3)
```




```
#x3 had the lowest oob error rate



#I chose x3

#d
a = data.frame(fixed_acidity = 8, volatile_acidity = 0.6, citric_acid = 0.6, residual_sugar = 1
5, chlorides = 0.3,
                free_sulfur_dioxide = 50, total_sulfur_dioxide = 200, density = 1, pH = 3.5, sul
phates = 0.75,
                alcohol = 10, type=factor("1",c(0:1)))

b=predict(x3, a, interval="predict", level=0.95)

b
```

```
## 1
## 5
## Levels: 4 5 6 7 8
```

```
#E

printcp(pfit)
```

```
##
## Classification tree:
## rpart(formula = quality ~ ., data = wine, method = "class")
##
## Variables actually used in tree construction:
## [1] alcohol              chlorides            density
## [4] sulphates            total_sulfur_dioxide volatile_acidity
##
## Root node error: 198/350 = 0.56571
##
## n= 350
##
##         CP nsplit rel error  xerror     xstd
## 1 0.043771      0   1.00000 1.00000 0.046833
## 2 0.037037      3   0.86869 1.06566 0.046233
## 3 0.030303      6   0.75758 0.97475 0.046993
## 4 0.025253      7   0.72727 0.94949 0.047113
## 5 0.022727      8   0.70202 0.92929 0.047181
## 6 0.017677     10   0.65657 0.90909 0.047224
## 7 0.015152     12   0.62121 0.90404 0.047231
```

```
print(x3)
```

```
##
## Call:
##  randomForest(formula = quality ~ ., data = wine, ntree = 300,      mtry = 5)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 49.43%
## Confusion matrix:
##   4  5  6  7 8 class.error
## 4 0  7  7  0 0   1.0000000
## 5 0 54 48  6 0   0.5000000
## 6 0 36 91 25 0   0.4013158
## 7 0  3 32 32 0   0.5223881
## 8 0  4  4  1 0   1.0000000
```

```
#The classification tree predicted 6 while the random forest predicted 5. The error rate for
# the CT was about 56.5% while, RF was about 48%


bikes = read.csv("C:/Users/dgmur/Downloads/SeoulBikes_Fl2022.csv")

dim(bikes)
```

```
## [1] 730  16
```

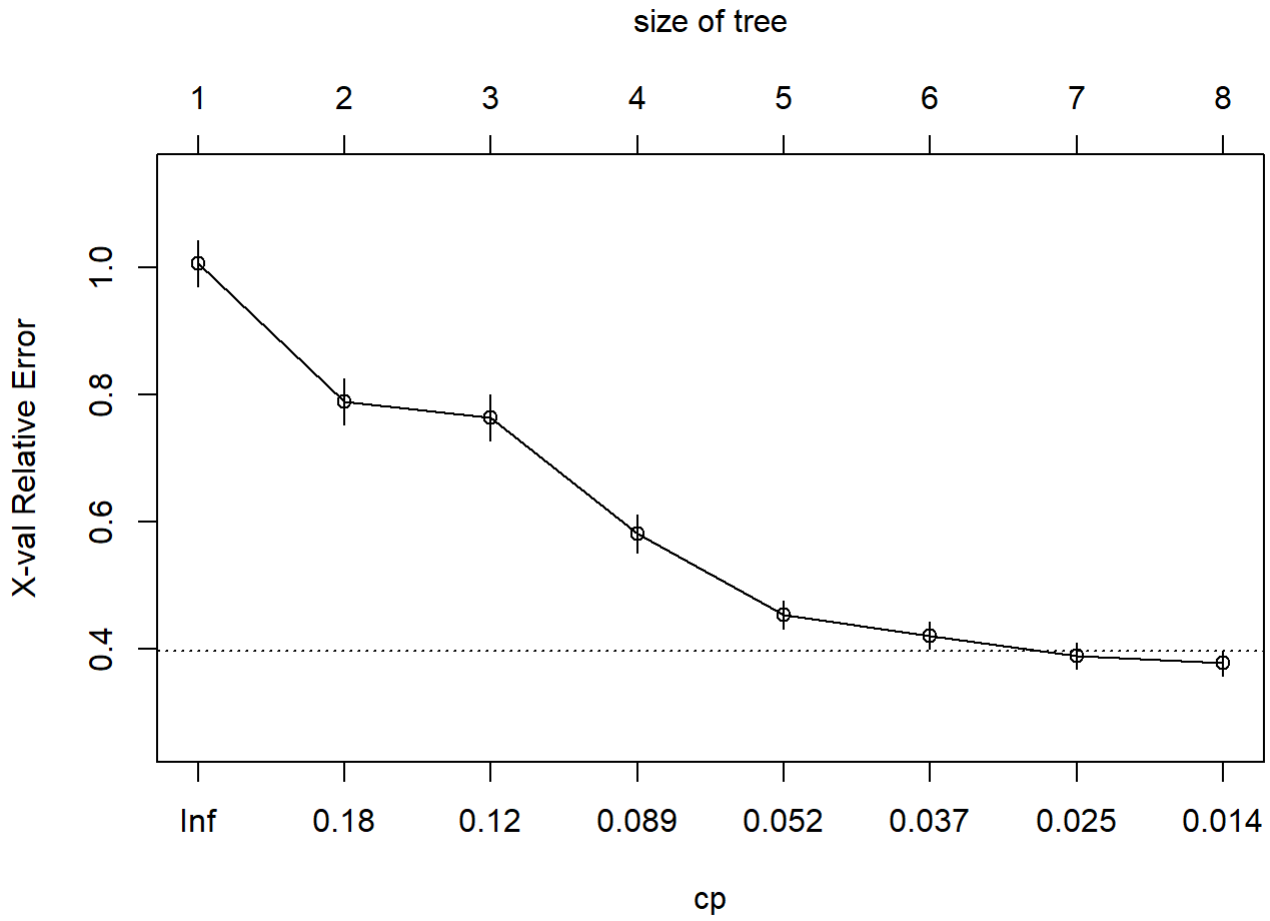
```
bikes = bikes[4:16]

x.new <- data.frame(Holiday=factor("Holiday", c("Holiday", "No Holiday")),
                    Seasons=factor("Spring", c("Autumn", "Spring", "Summer", "Winter")),
                    Functioning.Day=factor("Yes",c("No","Yes")),
                    Temperature=12.1, Humidity=29, WindSpeed=2.3,
                    Visibility=1734, DewPointTemp=-5.4, SolarRadiation=2.26,
                    Rainfall=0, Snowfall=0,
                    DayF=factor("21",c(1:31)), MonthF=factor("3",c(1:12)),
                    Time=factor("Morning",c("Morning","Evening")))

str(bikes)
```

```
## 'data.frame':    730 obs. of  13 variables:
##  $ RentedBikeCount: int  930 219 85 937 812 802 754 780 222 35 ...
##  $ Temperature    : num  -7.6 -4.2 3.2 -0.8 -8.1 -4.2 -0.2 -7 -5.4 -0.2 ...
##  $ Humidity       : int  37 79 92 69 36 78 96 59 56 90 ...
##  $ WindSpeed      : num  1.1 2.1 1.8 3.5 1.6 0 0.8 0.8 0.7 1.2 ...
##  $ Visibility     : int  2000 1436 244 954 2000 1518 173 1559 1940 378 ...
##  $ DewPointTemp   : num  -19.8 -7.3 2 -5.7 -20.5 -7.4 -0.7 -13.6 -12.8 -1.6 ...
##  $ SolarRadiation : num  0.01 0.01 0 0.01 0.01 0.01 0.01 0.01 0.01 0 ...
##  $ Rainfall       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Snowfall       : num  0 0 0 0 0 0.2 0.8 0 0 4.1 ...
##  $ Seasons        : chr  "Winter" "Winter" "Winter" "Winter" ...
##  $ Holiday        : chr  "No Holiday" "No Holiday" "No Holiday" "No Holiday" ...
##  $ Functioning.Day: chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Time           : chr  "Morning" "Morning" "Morning" "Morning" ...
```

```
bikes$Seasons= as.factor(bikes$Seasons)

bikes$Holiday= as.factor(bikes$Holiday)

bikes$Functioning.Day=as.factor(bikes$Functioning.Day)

bikes$Time=as.factor(bikes$Time)


#a

fit2= rpart(RentedBikeCount~., method = "anova", data=bikes)

printcp(fit2)
```

```
##
## Regression tree:
## rpart(formula = RentedBikeCount ~ ., data = bikes, method = "anova")
##
## Variables actually used in tree construction:
## [1] Functioning.Day Humidity        Seasons         SolarRadiation
## [5] Temperature     Time
##
## Root node error: 640097003/730 = 876845
##
## n= 730
##
##         CP nsplit rel error  xerror     xstd
## 1 0.269598      0   1.00000 1.00621 0.036343
## 2 0.116823      1   0.73040 0.78870 0.036018
## 3 0.114820      2   0.61358 0.76427 0.036189
## 4 0.068987      3   0.49876 0.58097 0.030018
## 5 0.039809      4   0.42977 0.45350 0.022044
## 6 0.034160      5   0.38996 0.42065 0.020941
## 7 0.018238      6   0.35580 0.38881 0.020175
## 8 0.010000      7   0.33756 0.37699 0.019396
```

```
plotcp(fit2)
```



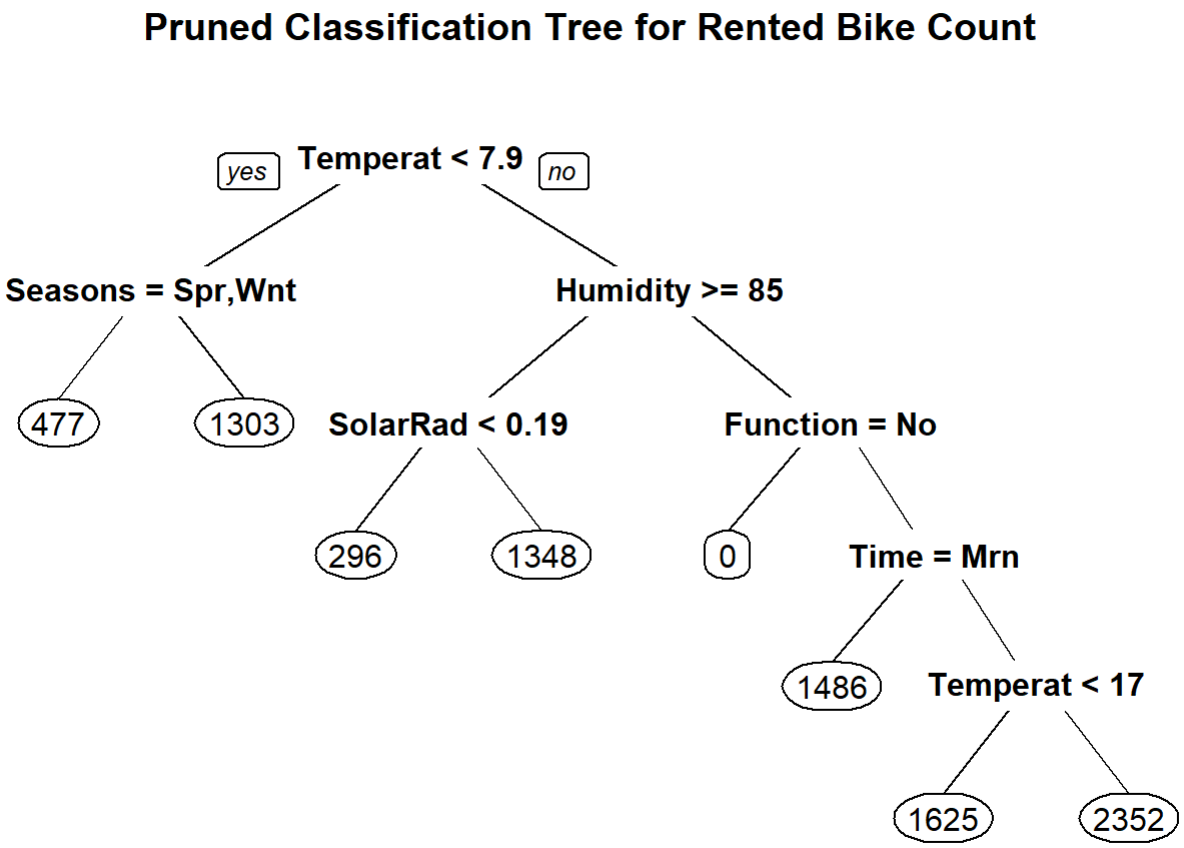

```
summary(fit2)
```

```
## Call:
## rpart(formula = RentedBikeCount ~ ., data = bikes, method = "anova")
##   n= 730
##
##           CP nsplit rel error    xerror       xstd
## 1 0.26959787      0 1.0000000 1.0062097 0.03634349
## 2 0.11682322      1 0.7304021 0.7887003 0.03601774
## 3 0.11481981      2 0.6135789 0.7642709 0.03618950
## 4 0.06898737      3 0.4987591 0.5809696 0.03001769
## 5 0.03980925      4 0.4297717 0.4534989 0.02204431
## 6 0.03415988      5 0.3899625 0.4206475 0.02094054
## 7 0.01823837      6 0.3558026 0.3888055 0.02017526
## 8 0.01000000      7 0.3375642 0.3769880 0.01939600
##
## Variable importance
##     Temperature         Seasons    DewPointTemp  SolarRadiation        Humidity
##              22              16              15              12              11
## Functioning.Day            Time        Rainfall        Snowfall       WindSpeed
##               8               5               4               3               3
##      Visibility
##               2
##
## Node number 1: 730 observations,    complexity param=0.2695979
##   mean=1259.314, MSE=876845.2
##   left son=2 (266 obs) right son=3 (464 obs)
##   Primary splits:
##       Temperature    < 7.85   to the left,  improve=0.26959790, (0 missing)
##       Seasons        splits as  RRRL,       improve=0.25654450, (0 missing)
##       SolarRadiation < 0.205  to the left,  improve=0.25609000, (0 missing)
##       DewPointTemp   < -2.25  to the left,  improve=0.17980060, (0 missing)
##       Humidity       < 84.5   to the right, improve=0.09142905, (0 missing)
##   Surrogate splits:
##       DewPointTemp   < 2.75   to the left,  agree=0.892, adj=0.703, (0 split)
##       Seasons        splits as  RRRL,       agree=0.882, adj=0.677, (0 split)
##       SolarRadiation < 0.075  to the left,  agree=0.818, adj=0.500, (0 split)
##       Snowfall       < 0.1    to the right, agree=0.688, adj=0.143, (0 split)
##       Humidity       < 34.5   to the left,  agree=0.653, adj=0.049, (0 split)
##
## Node number 2: 266 observations,    complexity param=0.03980925
##   mean=617.1617, MSE=239589
##   left son=4 (221 obs) right son=5 (45 obs)
##   Primary splits:
##       Seasons        splits as  RL-L,       improve=0.39983550, (0 missing)
##       Temperature    < 3.25   to the left,  improve=0.14399360, (0 missing)
##       DewPointTemp   < -10.25 to the left,  improve=0.07283931, (0 missing)
##       SolarRadiation < 0.115  to the left,  improve=0.07059487, (0 missing)
##       Holiday        splits as  LR,         improve=0.05290116, (0 missing)
##   Surrogate splits:
##       Temperature     < 6.75   to the left,  agree=0.842, adj=0.067, (0 split)
##       Functioning.Day splits as  RL,         agree=0.835, adj=0.022, (0 split)
##
## Node number 3: 464 observations,    complexity param=0.1168232
```

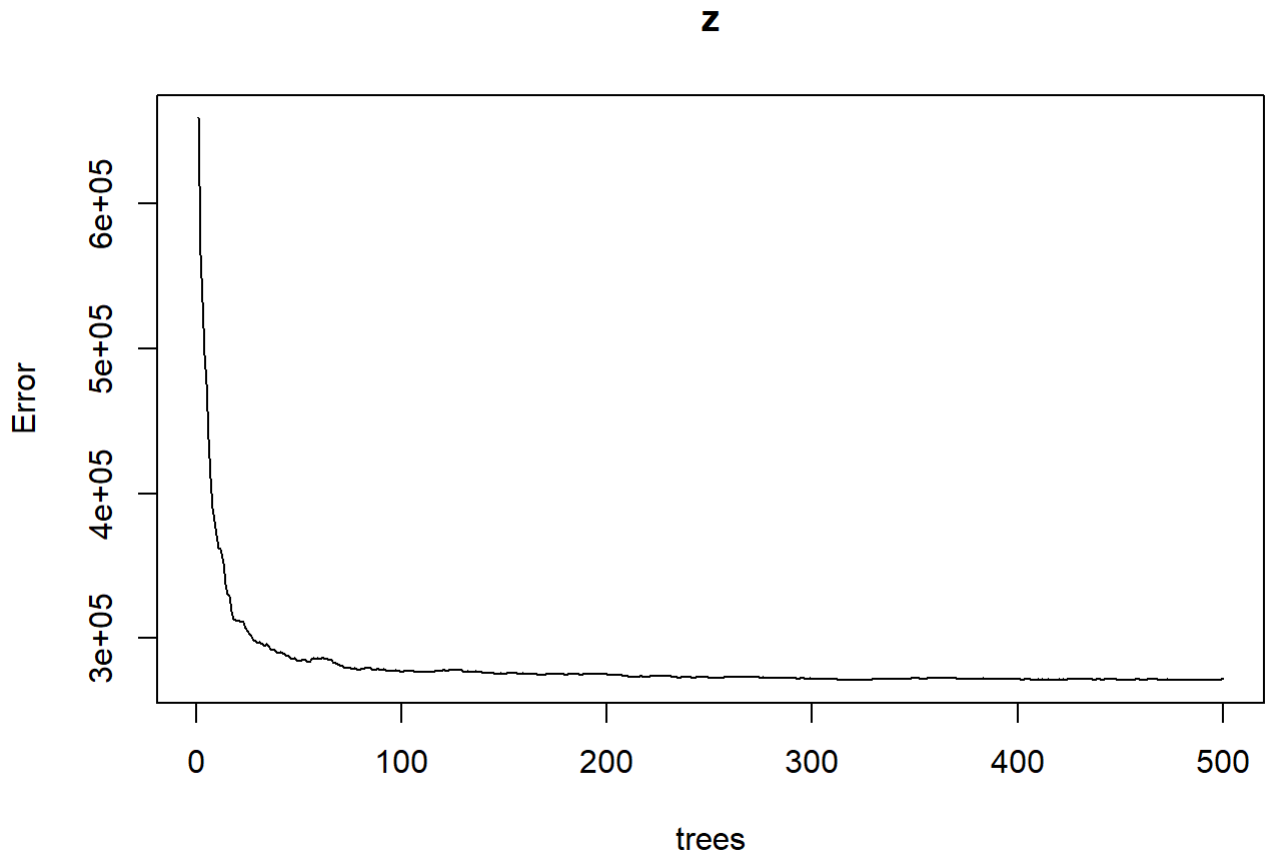
```
##   mean=1627.444, MSE=870253.3
##   left son=6 (57 obs) right son=7 (407 obs)
##   Primary splits:
##       Humidity        < 84.5   to the right, improve=0.1851873, (0 missing)
##       Rainfall        < 0.15   to the right, improve=0.1700205, (0 missing)
##       Functioning.Day splits as  LR,         improve=0.1587288, (0 missing)
##       Time            splits as  RL,         improve=0.1203948, (0 missing)
##       SolarRadiation  < 0.155  to the left,  improve=0.1168390, (0 missing)
##   Surrogate splits:
##       Rainfall     < 0.4    to the right, agree=0.933, adj=0.456, (0 split)
##       Visibility   < 283.5  to the left,  agree=0.897, adj=0.158, (0 split)
##       DewPointTemp < 25.45  to the right, agree=0.884, adj=0.053, (0 split)
##       WindSpeed    < 0.1    to the left,  agree=0.881, adj=0.035, (0 split)
##
## Node number 4: 221 observations
##   mean=477.4977, MSE=91569.86
##
## Node number 5: 45 observations
##   mean=1303.067, MSE=400265.6
##
## Node number 6: 57 observations,    complexity param=0.01823837
##   mean=554.7193, MSE=395396.7
##   left son=12 (43 obs) right son=13 (14 obs)
##   Primary splits:
##       SolarRadiation < 0.185  to the left,  improve=0.51799300, (0 missing)
##       Rainfall       < 0.75   to the right, improve=0.30090910, (0 missing)
##       Humidity       < 90.5   to the right, improve=0.14186150, (0 missing)
##       Temperature    < 19.3   to the left,  improve=0.07291409, (0 missing)
##       DewPointTemp   < 18.55  to the left,  improve=0.07291409, (0 missing)
##   Surrogate splits:
##       Temperature  < 24.5   to the left,  agree=0.825, adj=0.286, (0 split)
##       DewPointTemp < 23.8   to the left,  agree=0.807, adj=0.214, (0 split)
##
## Node number 7: 407 observations,    complexity param=0.1148198
##   mean=1777.678, MSE=753026.4
##   left son=14 (22 obs) right son=15 (385 obs)
##   Primary splits:
##       Functioning.Day splits as  LR,         improve=0.23980490, (0 missing)
##       Time            splits as  RL,         improve=0.14092870, (0 missing)
##       Temperature     < 20.45  to the left,  improve=0.11122760, (0 missing)
##       WindSpeed       < 2.25   to the left,  improve=0.07788779, (0 missing)
##       Humidity        < 57.5   to the right, improve=0.06370545, (0 missing)
##
## Node number 12: 43 observations
##   mean=296.4884, MSE=99407.6
##
## Node number 13: 14 observations
##   mean=1347.857, MSE=470625.8
##
## Node number 14: 22 observations
##   mean=0, MSE=0
##
```

```
## Node number 15: 385 observations,    complexity param=0.06898737
##   mean=1879.26, MSE=605158.2
##   left son=30 (164 obs) right son=31 (221 obs)
##   Primary splits:
##       Time        splits as  RL,         improve=0.18953340, (0 missing)
##       WindSpeed   < 1.35   to the left,  improve=0.10008420, (0 missing)
##       Temperature < 17.05  to the left,  improve=0.08926161, (0 missing)
##       Humidity    < 57.5   to the right, improve=0.06494347, (0 missing)
##       Rainfall    < 0.05   to the right, improve=0.04771395, (0 missing)
##   Surrogate splits:
##       WindSpeed   < 1.25   to the left,  agree=0.769, adj=0.457, (0 split)
##       Humidity    < 57.5   to the right, agree=0.735, adj=0.378, (0 split)
##       Visibility  < 1064   to the left,  agree=0.631, adj=0.134, (0 split)
##       Temperature < 13.85  to the left,  agree=0.590, adj=0.037, (0 split)
##       Rainfall    < 0.05   to the right, agree=0.577, adj=0.006, (0 split)
##
## Node number 30: 164 observations
##   mean=1486.116, MSE=463720.8
##
## Node number 31: 221 observations,    complexity param=0.03415988
##   mean=2171.005, MSE=510303.6
##   left son=62 (55 obs) right son=63 (166 obs)
##   Primary splits:
##       Temperature    < 16.55  to the left,  improve=0.19388370, (0 missing)
##       DewPointTemp   < 0.4    to the left,  improve=0.09247077, (0 missing)
##       SolarRadiation < 0.685  to the left,  improve=0.08959975, (0 missing)
##       Humidity       < 72.5   to the right, improve=0.08502754, (0 missing)
##       Visibility     < 642.5  to the left,  improve=0.04144956, (0 missing)
##   Surrogate splits:
##       SolarRadiation < 0.075  to the left,  agree=0.860, adj=0.436, (0 split)
##       DewPointTemp   < 2.85   to the left,  agree=0.855, adj=0.418, (0 split)
##       Visibility     < 505.5  to the left,  agree=0.769, adj=0.073, (0 split)
##
## Node number 62: 55 observations
##   mean=1624.545, MSE=444930.4
##
## Node number 63: 166 observations
##   mean=2352.06, MSE=400242.7
```

```
pfit2<- prune(fit2, cp=fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])

#b
prp(pfit2,main = "Pruned Classification Tree for Rented Bike Count")
```

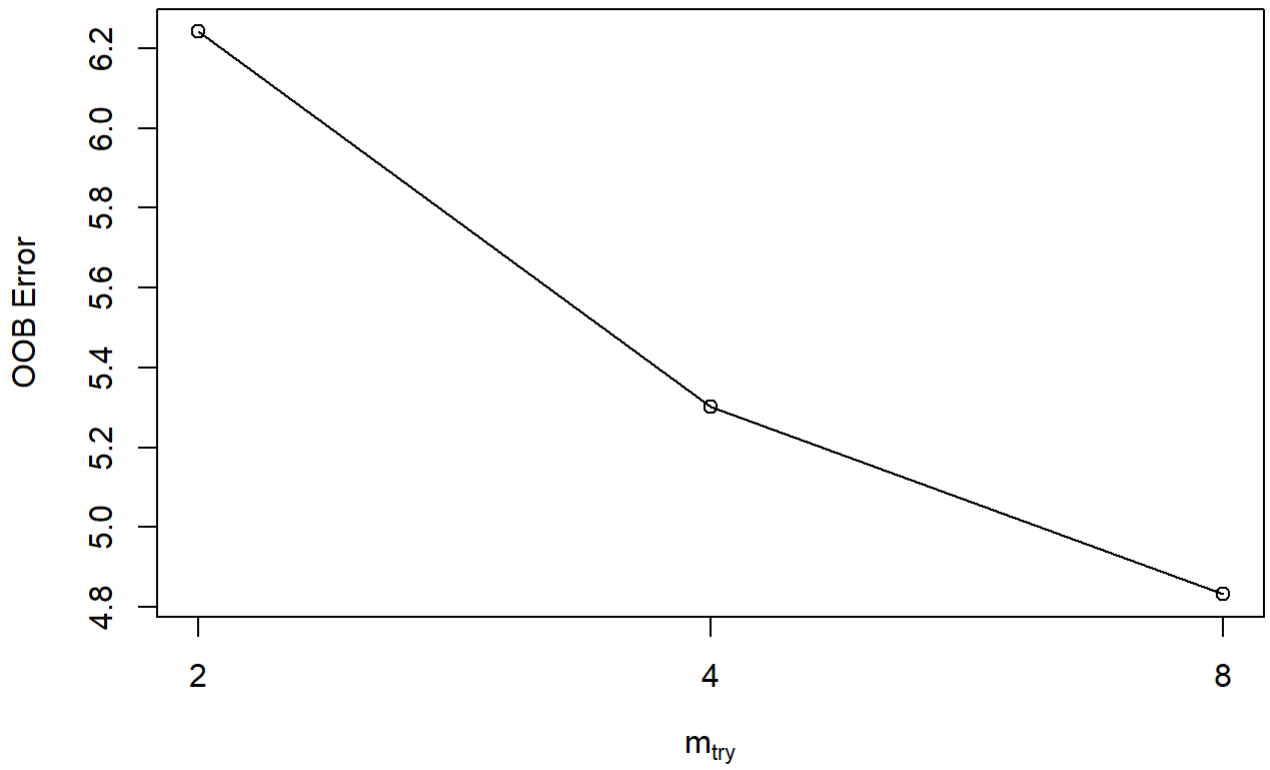



```
 #tree predicted 1486

#c
set.seed(55)


z = randomForest(RentedBikeCount~., data=bikes)
plot(z)
```




```
t2 = tuneRF(wine[,-4],wine[,4],
            stepFactor = .5, plot = TRUE,
            ntreeTry = 400,
            trace = TRUE,
            improve=1)
```

```
## mtry = 4  OOB error = 5.302175
## Searching left ...
## mtry = 8     OOB error = 4.832046
## 0.0886673 1
## Searching right ...
## mtry = 2     OOB error = 6.241474
## -0.1771534 1
```




```
z1 = randomForest(RentedBikeCount~., ntree=400,mtry=8,nodesize=2, data=bikes)
print(z1)
```

```
##
## Call:
##  randomForest(formula = RentedBikeCount ~ ., data = bikes, ntree = 400,      mtry = 8, nodesi
ze = 2)
##                Type of random forest: regression
##                      Number of trees: 400
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 278243
##                     % Var explained: 68.27
```

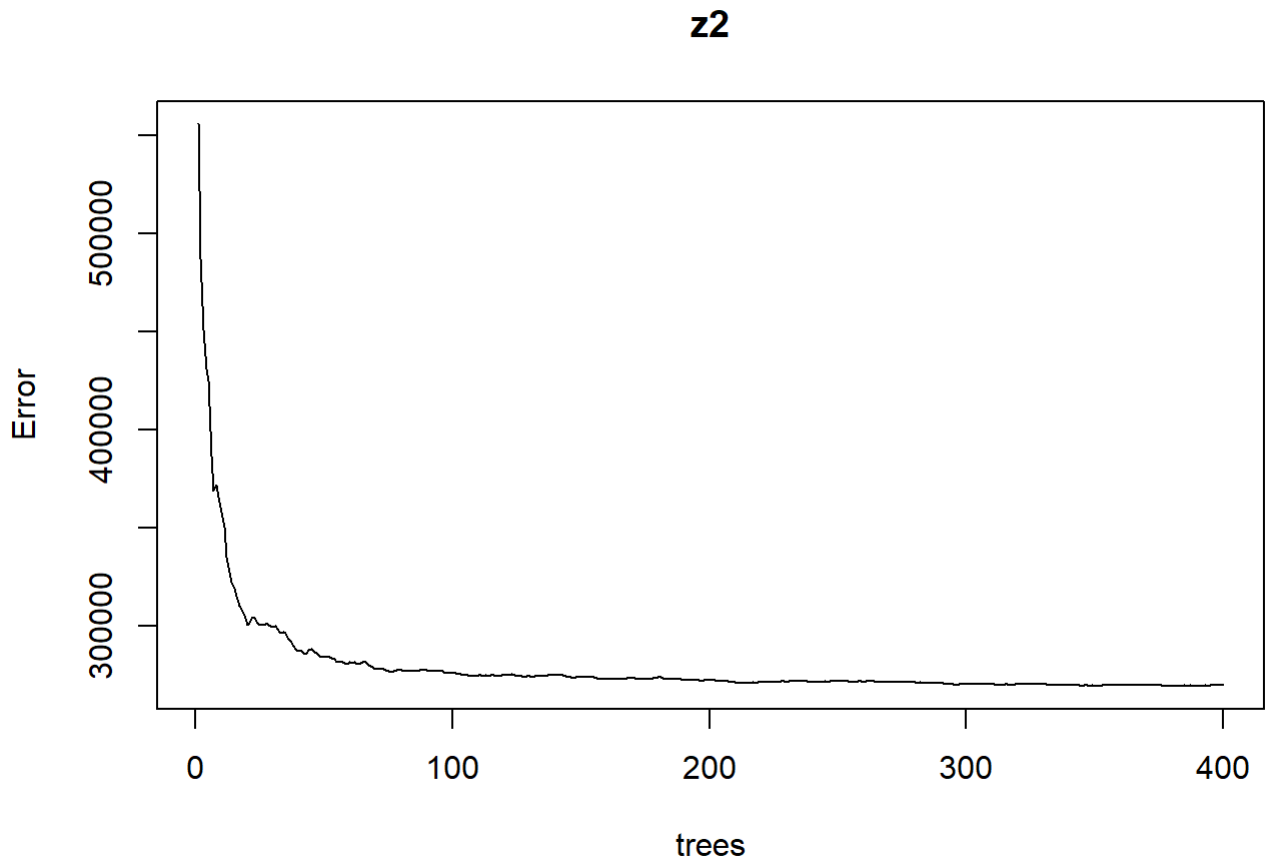
```
z2 = randomForest(RentedBikeCount~., ntree=400,mtry=8,nodesize=1, data=bikes)
print(z2)
```

```
##
## Call:
##  randomForest(formula = RentedBikeCount ~ ., data = bikes, ntree = 400,      mtry = 8, nodesi
ze = 1)
##                Type of random forest: regression
##                      Number of trees: 400
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 270018.4
##                     % Var explained: 69.21
```

```
z3 = randomForest(RentedBikeCount~., ntree=400,mtry=8,nodesize=5, data=bikes)
print(z3)
```

```
##
## Call:
##  randomForest(formula = RentedBikeCount ~ ., data = bikes, ntree = 400,      mtry = 8, nodesi
ze = 5)
##                Type of random forest: regression
##                      Number of trees: 400
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 271002.1
##                     % Var explained: 69.09
```

```
#I selected z2

plot(z2)
```




```
#d
zp=predict(z1, x.new, interval="predict", level=0.95)
zp
```

```
##        1
## 1192.925
```

```
#RF predicted 1147.564

#E
printcp(pfit2)
```

```
##
## Regression tree:
## rpart(formula = RentedBikeCount ~ ., data = bikes, method = "anova")
##
## Variables actually used in tree construction:
## [1] Functioning.Day Humidity        Seasons         SolarRadiation
## [5] Temperature     Time
##
## Root node error: 640097003/730 = 876845
##
## n= 730
##
##         CP nsplit rel error  xerror     xstd
## 1 0.269598      0   1.00000 1.00621 0.036343
## 2 0.116823      1   0.73040 0.78870 0.036018
## 3 0.114820      2   0.61358 0.76427 0.036189
## 4 0.068987      3   0.49876 0.58097 0.030018
## 5 0.039809      4   0.42977 0.45350 0.022044
## 6 0.034160      5   0.38996 0.42065 0.020941
## 7 0.018238      6   0.35580 0.38881 0.020175
## 8 0.010000      7   0.33756 0.37699 0.019396
```

```
print(z2)
```

```
##
## Call:
##  randomForest(formula = RentedBikeCount ~ ., data = bikes, ntree = 400,      mtry = 8, nodesi
ze = 1)
##                Type of random forest: regression
##                      Number of trees: 400
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 270018.4
##                     % Var explained: 69.21
```

```
#The regression tree had a rootnode error rate of 100% and rf had a 69% variance explained
zp
```

```
##        1
## 1192.925
```

```
# regression tree predicted 1486 while rf was 1147.564
```