Rucha Damle

23060641010

Div: A

1. Write R code for a principle component analysis.

Code:

```
install.packages("corrr")
library('corrr')
install.packages("ggcorrplot")
library(ggcorrplot)
install.packages("FactoMineR")
library("FactoMineR")
install.packages("factoextra")
library("factoextra")
wine_quality <- read.csv("C:/Users/rucha/OneDrive/Desktop/Ruchi/wine_quality.csv")
str(wine_quality)
colSums(is.na(wine_quality))
numerical_data <- wine_quality[,2:10]
head(numerical_data)
data_normalized <- scale(numerical_data)
head(data_normalized)
corr_matrix <- cor(data_normalized)
ggcorrplot(corr_matrix)
data.pca <- princomp(corr_matrix)
summary(data.pca)
data.pca$loadings[, 1:2]
fviz_eig(data.pca, addlabels = TRUE)
# Graph of the variables
fviz_pca_var(data.pca, col.var = "black")
fviz_cos2(data.pca, choice = "var", axes = 1:2)
fviz_pca_var(data.pca, col.var = "cos2",
```

```
gradient.cols = c("black", "orange", "green"),
repel = TRUE)
```

Output:

> wine_quality <- read.csv("C:/Users/rucha/OneDrive/Desktop/Ruchi/wine_quality.csv")

> str(wine_quality)

```
'data.frame':   1143 obs. of  13 variables:
 $ fixed.acidity       : num  7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 6.7 ...
 $ volatile.acidity    : num  0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.58 ...
 $ citric.acid         : num  0 0 0.04 0.56 0 0 0.06 0 0.02 0.08 ...
 $ residual.sugar      : num  1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 1.8 ...
 $ chlorides           : num  0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.097 ...
 $ free.sulfur.dioxide : num  11 25 15 17 11 13 15 15 9 15 ...
 $ total.sulfur.dioxide: num  34 67 54 60 34 40 59 21 18 65 ...
 $ density             : num  0.998 0.997 0.997 0.998 0.998 ...
 $ pH                  : num  3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.28 ...
 $ sulphates           : num  0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.54 ...
 $ alcohol             : num  9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 9.2 ...
 $ quality             : int  5 5 5 6 5 5 5 7 7 5 ...
 $ Id                  : int  0 1 2 3 4 5 6 7 8 10 ...
```

> colSums(is.na(wine_quality))

```
       fixed.acidity     volatile.acidity          citric.acid 
                   0                    0                    0 
      residual.sugar            chlorides  free.sulfur.dioxide 
                   0                    0                    0 
total.sulfur.dioxide              density                   pH 
                   0                    0                    0 
           sulphates              alcohol              quality 
                   0                    0                    0 
                  Id 
                   0 
```

> numerical_data <- wine_quality[,2:10]

> head(numerical_data)

```
  volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1             0.70        0.00            1.9     0.076                  11
2             0.88        0.00            2.6     0.098                  25
```

```
3           0.76      0.04          2.3     0.092                 15
4           0.28      0.56          1.9     0.075                 17
5           0.70      0.00          1.9     0.076                 11
6           0.66      0.00          1.8     0.075                 13
  total.sulfur.dioxide density   pH sulphates
1                   34  0.9978 3.51      0.56
2                   67  0.9968 3.20      0.68
3                   54  0.9970 3.26      0.65
4                   60  0.9980 3.16      0.58
5                   34  0.9978 3.51      0.56
6                   40  0.9978 3.51      0.56
```

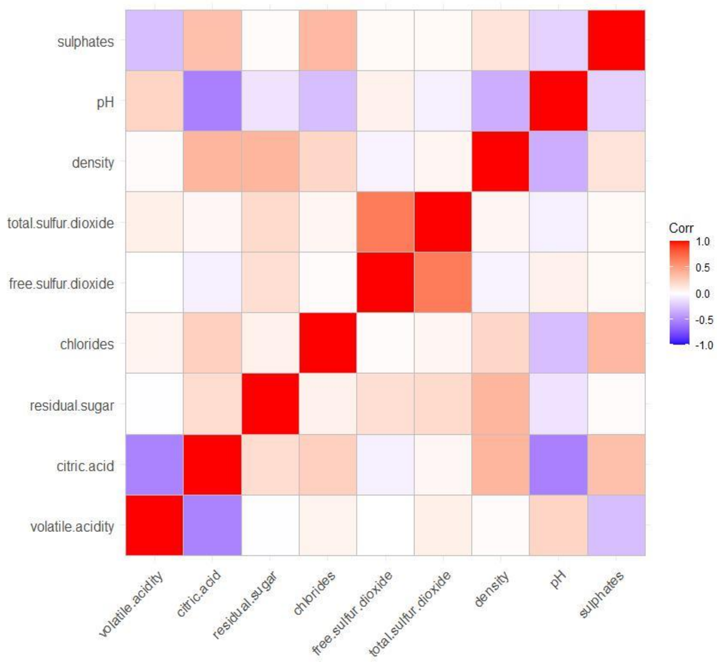
> data_normalized <- scale(numerical_data)

> head(data_normalized)

```
     volatile.acidity citric.acid residual.sugar  chlorides free.sulfur.dioxide
[1,]        0.9389212   -1.364429    -0.46621734 -0.2312936         -0.45026992
[2,]        1.9409632   -1.364429     0.05003827  0.2341441          0.91551896
[3,]        1.2729352   -1.161059    -0.17121413  0.1072065         -0.06004452
[4,]       -1.3991767    1.482750    -0.46621734 -0.2524499          0.13506817
[5,]        0.9389212   -1.364429    -0.46621734 -0.2312936         -0.45026992
[6,]        0.7162452   -1.364429    -0.53996814 -0.2524499         -0.25515722
     total.sulfur.dioxide    density         pH   sulphates
[1,]           -0.3634510 0.55561117  1.2701390 -0.57340683
[2,]            0.6431950 0.03614877 -0.7086174  0.13082384
[3,]            0.2466375 0.14004125 -0.3256323 -0.04523383
[4,]            0.4296640 0.65950365 -0.9639408 -0.45603505
[5,]           -0.3634510 0.55561117  1.2701390 -0.57340683
[6,]           -0.1804245 0.55561117  1.2701390 -0.57340683
```

> corr_matrix <- cor(data_normalized)

> ggcorrplot(corr_matrix)



> data.pca <- princomp(corr_matrix)

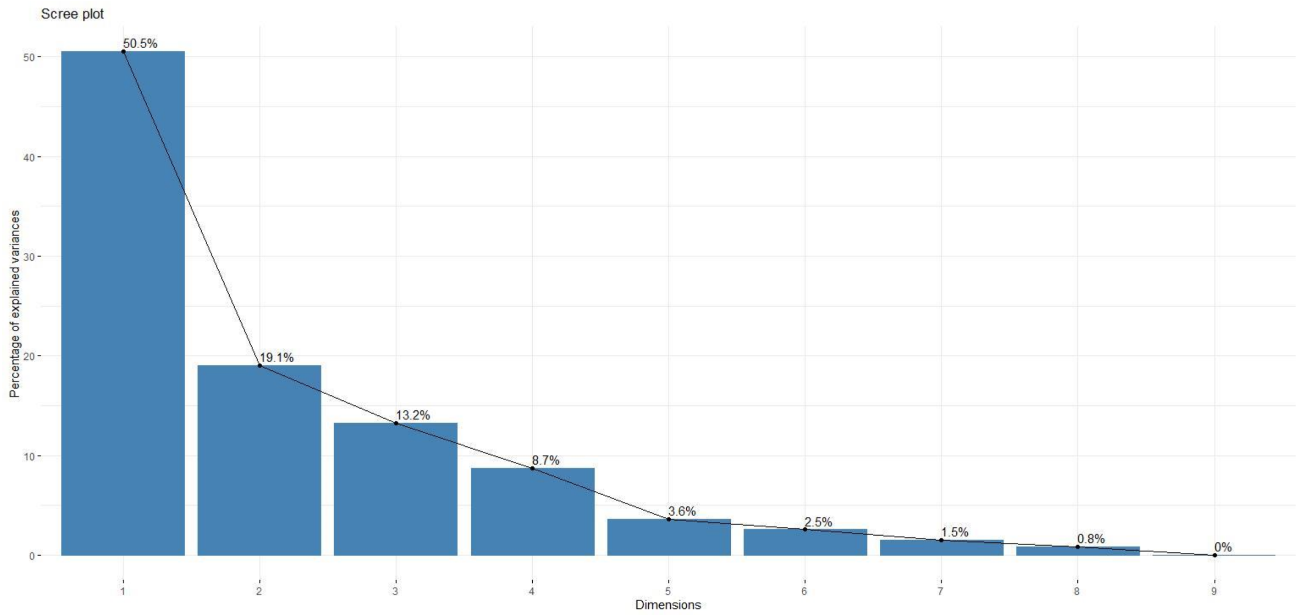
> summary(data.pca)

```
Importance of components:
                         Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     0.7891957 0.4846892 0.4036058 0.32805920 0.21118795
Proportion of Variance 0.5051674 0.1905428 0.1321237 0.08729116 0.03617463
Cumulative Proportion  0.5051674 0.6957101 0.8278338 0.91512498 0.95129961
                         Comp.6     Comp.7     Comp.8 Comp.9
Standard deviation     0.1772605 0.13565085 0.10109970      0
Proportion of Variance 0.0254853 0.01492488 0.00829021      0
Cumulative Proportion  0.9767849 0.99170979 1.00000000      1
```

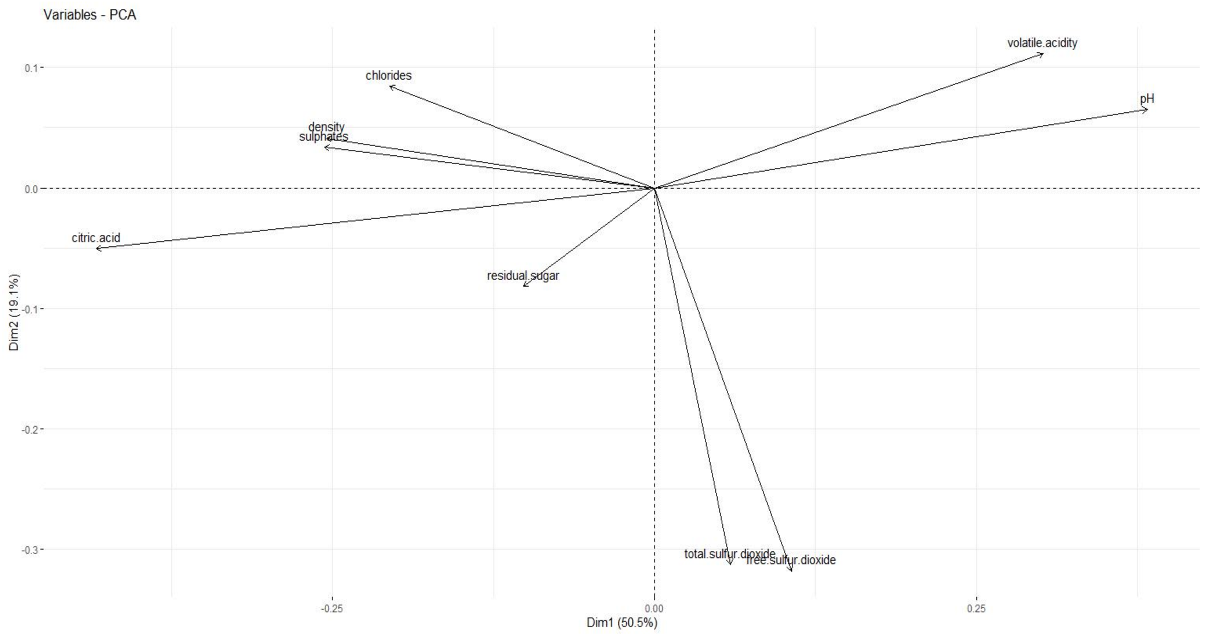
> data.pca$loadings[, 1:2]

```
                          Comp.1      Comp.2
volatile.acidity      0.38225034  0.22988464
citric.acid          -0.54879930 -0.10423407
residual.sugar       -0.12890286 -0.16890443
chlorides            -0.26027239  0.17370873
free.sulfur.dioxide   0.13458440 -0.65639029
total.sulfur.dioxide  0.07510444 -0.64550076
density              -0.32220371  0.08501984
pH                    0.48504728  0.13424557
sulphates            -0.32481781  0.06939157
```

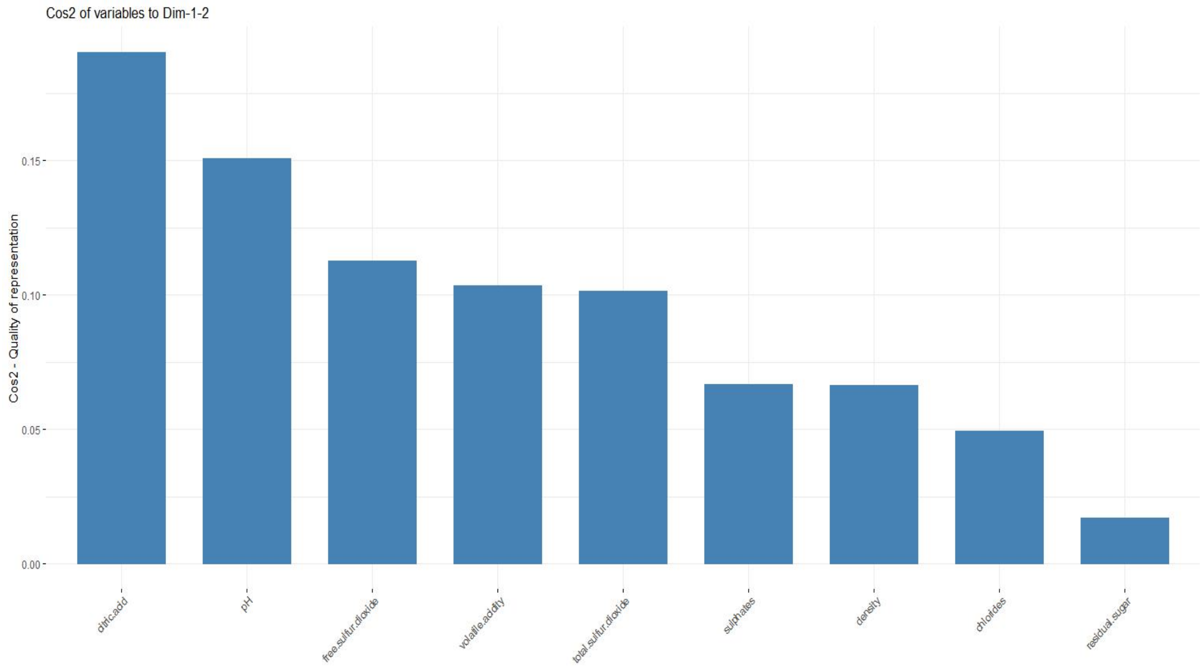
> fviz_eig(data.pca, addlabels = TRUE)



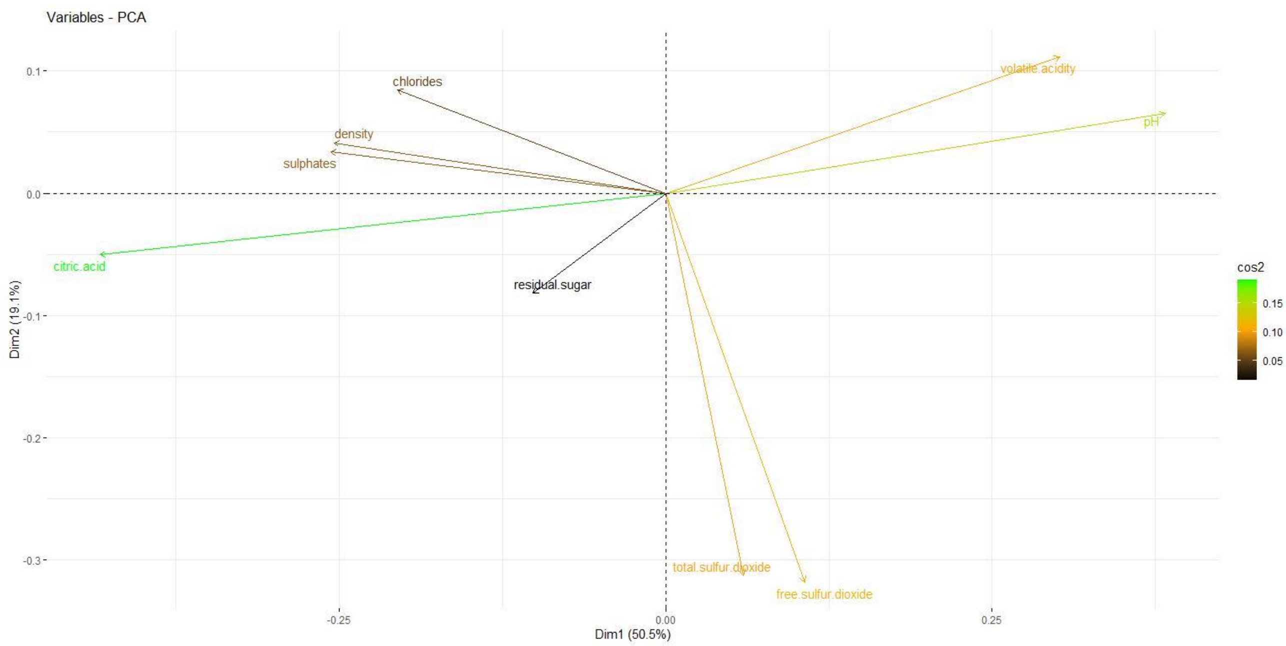
> fviz_pca_var(data.pca, col.var = "black")



> fviz_cos2(data.pca, choice = "var", axes = 1:2)




```
> fviz_pca_var(data.pca, col.var = "cos2",
+           gradient.cols = c("black", "orange", "green"),
+           repel = TRUE)
```



2. Write R code for factor analysis

Code:

```
> factor_analysis <- factanal(numerical_data, factors = 3)
> print(factor_analysis)
```

Output:

```
Call:
factanal(x = numerical_data, factors = 3)

Uniquenesses:
    volatile.acidity          citric.acid       residual.sugar
               0.005                0.275                0.872
           chlorides  free.sulfur.dioxide total.sulfur.dioxide
               0.802                0.020                0.518
             density                   pH            sulphates
               0.644                0.561                0.836

Loadings:
                     Factor1 Factor2 Factor3
volatile.acidity      0.994
citric.acid          -0.595           0.608
residual.sugar                0.190   0.301
chlorides                             0.441
free.sulfur.dioxide           0.986
total.sulfur.dioxide          0.681   0.127
density                               0.596
pH                    0.270          -0.604
sulphates            -0.301           0.268

               Factor1 Factor2 Factor3
SS loadings      1.510   1.481   1.474
Proportion Var   0.168   0.165   0.164
Cumulative Var   0.168   0.332   0.496

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 263.05 on 12 degrees of freedom.
The p-value is 2.58e-49
```

3.Write R code for canonical correleation analysis.

Code:

```
attach(wine_quality)

# Making two vectors X and Y

X<-wine_quality[1:6]

Y<-wine_quality[7:13]

print(X)

print(Y)

#Perform Canonical Correlation Analysis

cca_result <- cancor(X, Y)

# Summary of the results

summary(cca_result)

# Canonical Correlation Coefficients

cca_result$cor

# Canonical Loadings for X

cca_result$xcoef

# Canonical Loadings for Y

cca_result$ycoef
```

Output:

```
> attach(wine_quality)
The following objects are masked from wine_quality (pos = 3):

    alcohol, chlorides, citric.acid, density, fixed.acidity, free.sulfur.dioxide,
    Id, pH, quality, residual.sugar, sulphates, total.sulfur.dioxide,
    volatile.acidity

> # Making two vectors X and Y
> X<-wine_quality[1:6]
> Y<-wine_quality[7:13]

> print(X)

    fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1          7.4          0.700        0.00         1.90     0.076              11
2          7.8          0.880        0.00         2.60     0.098              25
3          7.8          0.760        0.04         2.30     0.092              15
4         11.2          0.280        0.56         1.90     0.075              17
5          7.4          0.700        0.00         1.90     0.076              11
6          7.4          0.660        0.00         1.80     0.075              13
7          7.9          0.600        0.06         1.60     0.069              15
8          7.3          0.650        0.00         1.20     0.065              15
9          7.8          0.580        0.02         2.00     0.073               9
10         6.7          0.580        0.08         1.80     0.097              15
11         5.6          0.615        0.00         1.60     0.089              16
12         7.8          0.610        0.29         1.60     0.114               9
13         8.5          0.280        0.56         1.80     0.092              35
14         7.9          0.320        0.51         1.80     0.341              17
15         7.6          0.390        0.31         2.30     0.082
7
```

```
 [ reached 'max' / getOption("max.print") -- omitted 977 rows ]

> print(Y)

total.sulfur.dioxide density   pH sulphates alcohol quality  Id
1                   34  0.9978 3.51      0.56     9.4       5   0
2                   67  0.9968 3.20      0.68     9.8       5   1
3                   54  0.9970 3.26      0.65     9.8       5   2
4                   60  0.9980 3.16      0.58     9.8       6   3
5                   34  0.9978 3.51      0.56     9.4       5   4
6                   40  0.9978 3.51      0.56     9.4       5   5
7                   59  0.9964 3.30      0.46     9.4       5   6
8                   21  0.9946 3.39      0.47    10.0       7   7
9                   18  0.9968 3.36      0.57     9.5       7   8
10                  65  0.9959 3.28      0.54     9.2       5  10
11                  59  0.9943 3.58      0.52     9.9       5  12
12                  29  0.9974 3.26      1.56     9.1       5  13
13                 103  0.9969 3.30      0.75    10.5       7  16
14                  56  0.9969 3.04      1.08     9.2       6  19
15                  71  0.9982 3.52      0.65     9.7       5  21
[ reached 'max' / getOption("max.print") -- omitted 1001 rows ]

> #Perform Canonical Correlation Analysis
> cca_result <- cancor(X, Y)
> # Summary of the results

> summary(cca_result)
        Length Class  Mode
cor      6     -none- numeric
xcoef   36     -none- numeric
ycoef   49     -none- numeric
xcenter  6     -none- numeric
ycenter  7     -none- numeric

> # Canonical Correlation Coefficients
> cca_result$cor
[1] 0.9378924 0.7214574 0.5450951 0.4424002 0.2072509 0.1717869

> # Canonical Loadings for X
> cca_result$xcoef
                              [,1]         [,2]          [,3]         [,4]         [,5]
fixed.acidity         0.0151129983 -0.002747175 -0.0024220413 -0.007071534  0.011754488
volatile.acidity      0.0039791213  0.044548960 -0.0734630741 -0.104203458  0.004131750
citric.acid           0.0020930050  0.068395672  0.0230216427  0.072300854 -0.039338579
residual.sugar        0.0055607180  0.005371927 -0.0095220968  0.002226067 -0.017719366
chlorides             0.0554710020 -0.092511713  0.4795052581 -0.342570077 -0.227954644
free.sulfur.dioxide -0.0002366815  0.002529444  0.0005919518 -0.000337976  0.001311562
                            [,6]
fixed.acidity       -0.011907887
volatile.acidity     0.157520340
citric.acid          0.230105875
residual.sugar      -0.007033329
chlorides           -0.194556123
free.sulfur.dioxide -0.000567475

> # Canonical Loadings for Y
> cca_result$ycoef
                             [,1]          [,2]          [,3]          [,4]          [,5]
total.sulfur.dioxide -8.666323e-05  9.094777e-04  4.160168e-05  5.367763e-05  1.445482e-04
density               1.236426e+01  4.900994e+00 -9.090534e+00 -1.294681e+00 -9.135633e+00
pH                   -1.026796e-01  1.342842e-02 -6.870228e-02 -1.411117e-02 -1.162437e-01
sulphates            -1.071846e-02 -5.580117e-03  1.475149e-01 -1.826556e-02 -1.092587e-01
alcohol               1.198400e-02  7.751857e-03 -1.328509e-02  1.192299e-02 -1.412388e-02
quality               3.928203e-04  9.806246e-04  1.970071e-03  2.611354e-02  1.374521e-02
Id                   -3.242203e-06  1.476356e-05  1.738818e-06  9.615776e-07 -1.850081e-05
```

	[,6]	[,7]
total.sulfur.dioxide	-1.854650e-05	-9.626301e-05
density	-5.761793e+00	1.533869e+00
pH	-1.008529e-01	-6.016749e-02
sulphates	3.975081e-03	-2.172450e-02
alcohol	2.440422e-02	-2.358211e-03
quality	-3.258403e-02	-2.963884e-03
Id	-1.473282e-05	6.308493e-05