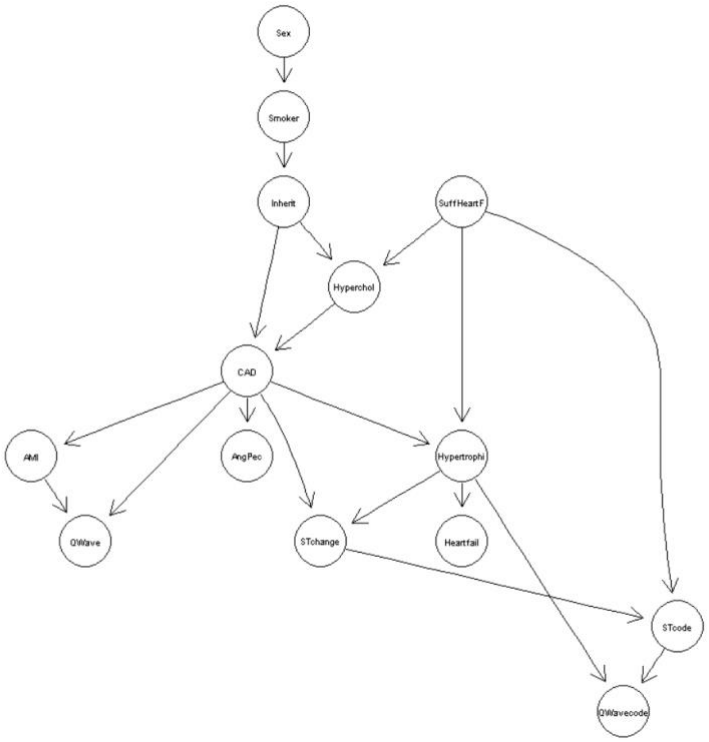
SDM Final Homework

1.

(B)

After filling the illegal edges and fitting the network with the new constraints we get the following graph.



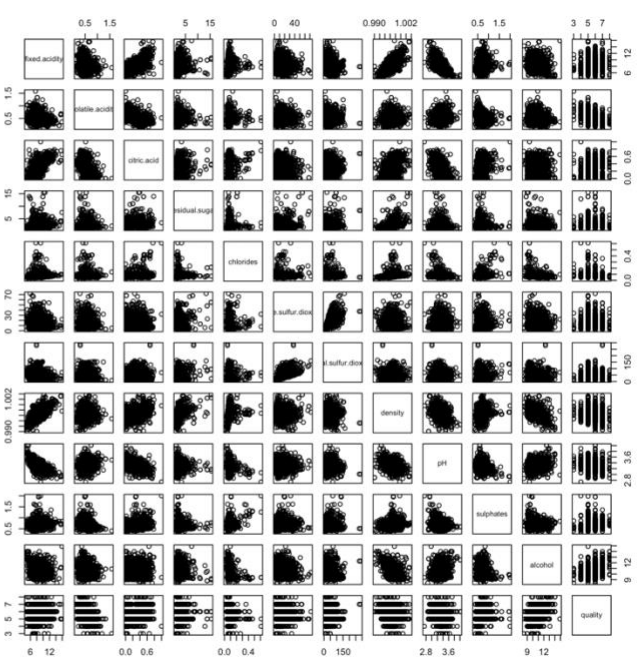
(D) output - 0.838983050847458

2.

(a)

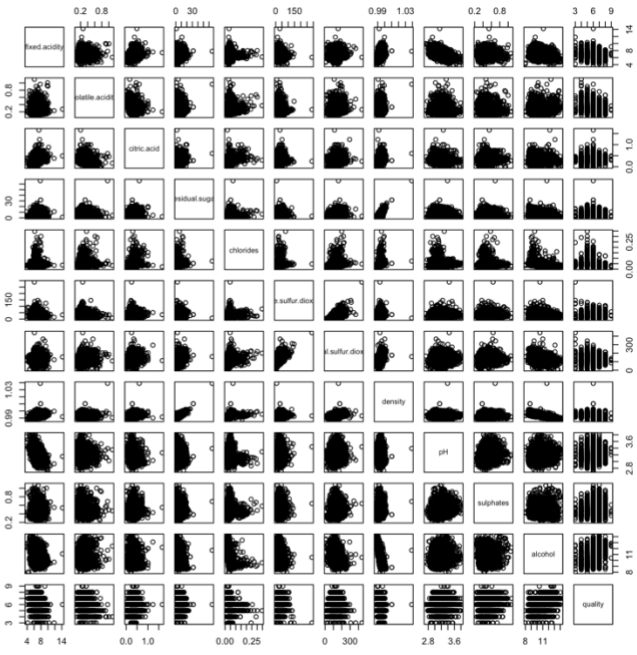
RED data

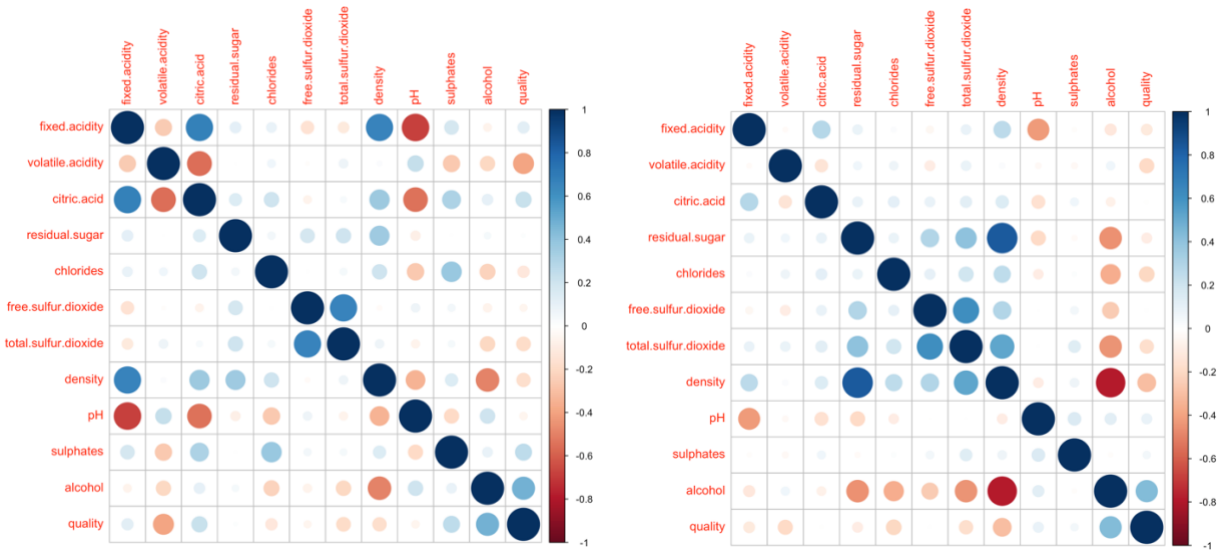
```
>
> print(dim(red))
[1] 1599   12
> print(dim(white))
[1] 4898   12
>
> head(red)
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
1           7.4             0.70        0.00            1.9     0.076
2           7.8             0.88        0.00            2.6     0.098
3           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 quality
1                  11                   34  0.9978 3.51      0.56     9.4       5
2                  25                   67  0.9968 3.20      0.68     9.8       5
3                  15                   54  0.9970 3.26      0.65     9.8       5
4                  17                   60  0.9980 3.16      0.58     9.8       6
5                  11                   34  0.9978 3.51      0.56     9.4       5
6                  13                   40  0.9978 3.51      0.56     9.4       5
> summary(red)
 fixed.acidity   volatile.acidity  citric.acid    residual.sugar
 Min.   : 4.60   Min.   :0.1200   Min.   :0.000   Min.   : 0.900
 1st Qu.: 7.10   1st Qu.:0.3900   1st Qu.:0.090   1st Qu.: 1.900
 Median : 7.90   Median :0.5200   Median :0.260   Median : 2.200
 Mean   : 8.32   Mean   :0.5278   Mean   :0.271   Mean   : 2.539
 3rd Qu.: 9.20   3rd Qu.:0.6400   3rd Qu.:0.420   3rd Qu.: 2.600
 Max.   :15.90   Max.   :1.5800   Max.   :1.000   Max.   :15.500
   chlorides       free.sulfur.dioxide total.sulfur.dioxide    density
 Min.   :0.01200   Min.   : 1.00       Min.   :  6.00       Min.   :0.9901
 1st Qu.:0.07000   1st Qu.: 7.00       1st Qu.: 22.00       1st Qu.:0.9956
 Median :0.07900   Median :14.00       Median : 38.00       Median :0.9968
 Mean   :0.08747   Mean   :15.87       Mean   : 46.47       Mean   :0.9967
 3rd Qu.:0.09000   3rd Qu.:21.00       3rd Qu.: 62.00       3rd Qu.:0.9978
 Max.   :0.61100   Max.   :72.00       Max.   :289.00       Max.   :1.0037
       pH          sulphates         alcohol         quality
 Min.   :2.740   Min.   :0.3300   Min.   : 8.40   Min.   :3.000
 1st Qu.:3.210   1st Qu.:0.5500   1st Qu.: 9.50   1st Qu.:5.000
 Median :3.310   Median :0.6200   Median :10.20   Median :6.000
 Mean   :3.311   Mean   :0.6581   Mean   :10.42   Mean   :5.636
 3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   3rd Qu.:6.000
 Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :8.000
> plot(red)
```



WHITE DATA

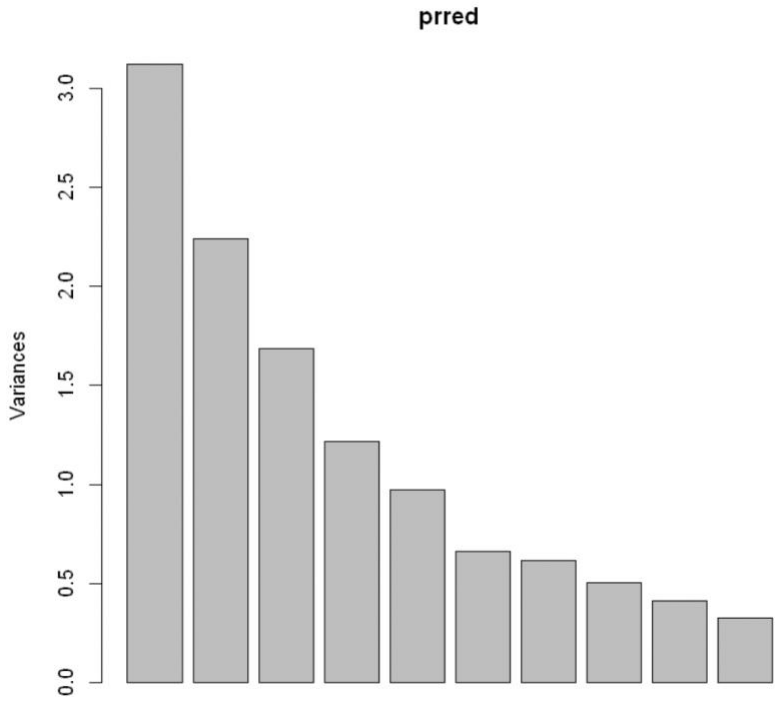
```
> plot(white)
> head(white)
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
1           7.0             0.27        0.36           20.7     0.045
2           6.3             0.30        0.34            1.6     0.049
3           8.1             0.28        0.40            6.9     0.050
4           7.2             0.23        0.32            8.5     0.058
5           7.2             0.23        0.32            8.5     0.058
6           8.1             0.28        0.40            6.9     0.050
  free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol quality
1                  45                  170  1.0010 3.00      0.45     8.8       6
2                  14                  132  0.9940 3.30      0.49     9.5       6
3                  30                   97  0.9951 3.26      0.44    10.1       6
4                  47                  186  0.9956 3.19      0.40     9.9       6
5                  47                  186  0.9956 3.19      0.40     9.9       6
6                  30                   97  0.9951 3.26      0.44    10.1       6
> summary(white)
 fixed.acidity    volatile.acidity  citric.acid     residual.sugar
 Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600
 1st Qu.: 6.300   1st Qu.:0.2100   1st Qu.:0.2700   1st Qu.: 1.700
 Median : 6.800   Median :0.2600   Median :0.3200   Median : 5.200
 Mean   : 6.855   Mean   :0.2782   Mean   :0.3342   Mean   : 6.391
 3rd Qu.: 7.300   3rd Qu.:0.3200   3rd Qu.:0.3900   3rd Qu.: 9.900
 Max.   :14.200   Max.   :1.1000   Max.   :1.6600   Max.   :65.800
   chlorides       free.sulfur.dioxide total.sulfur.dioxide    density
 Min.   :0.00900   Min.   :  2.00      Min.   :  9.0        Min.   :0.9871
 1st Qu.:0.03600   1st Qu.: 23.00      1st Qu.:108.0        1st Qu.:0.9917
 Median :0.04300   Median : 34.00      Median :134.0        Median :0.9937
 Mean   :0.04577   Mean   : 35.31      Mean   :138.4        Mean   :0.9940
 3rd Qu.:0.05000   3rd Qu.: 46.00      3rd Qu.:167.0        3rd Qu.:0.9961
 Max.   :0.34600   Max.   :289.00      Max.   :440.0        Max.   :1.0390
       pH          sulphates         alcohol         quality
 Min.   :2.720   Min.   :0.2200   Min.   : 8.00   Min.   :3.000
 1st Qu.:3.090   1st Qu.:0.4100   1st Qu.: 9.50   1st Qu.:5.000
 Median :3.180   Median :0.4700   Median :10.40   Median :6.000
 Mean   :3.188   Mean   :0.4898   Mean   :10.51   Mean   :5.878
 3rd Qu.:3.280   3rd Qu.:0.5500   3rd Qu.:11.40   3rd Qu.:6.000
 Max.   :3.820   Max.   :1.0800   Max.   :14.20   Max.   :9.000
> plot(white)
>
>
>
```

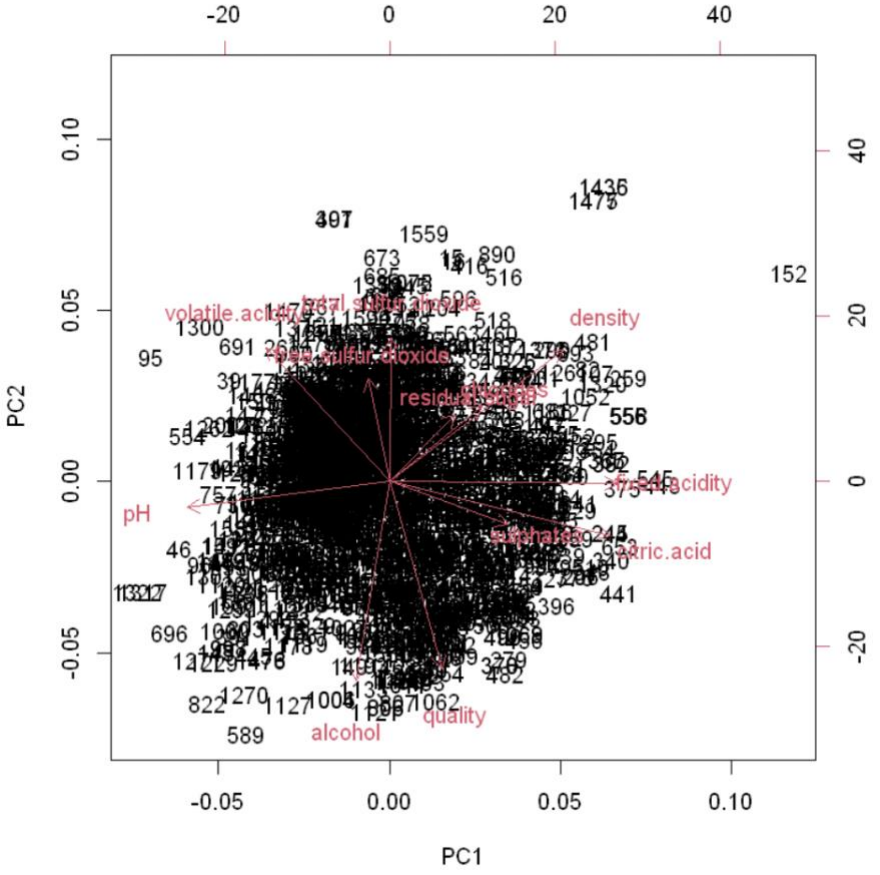




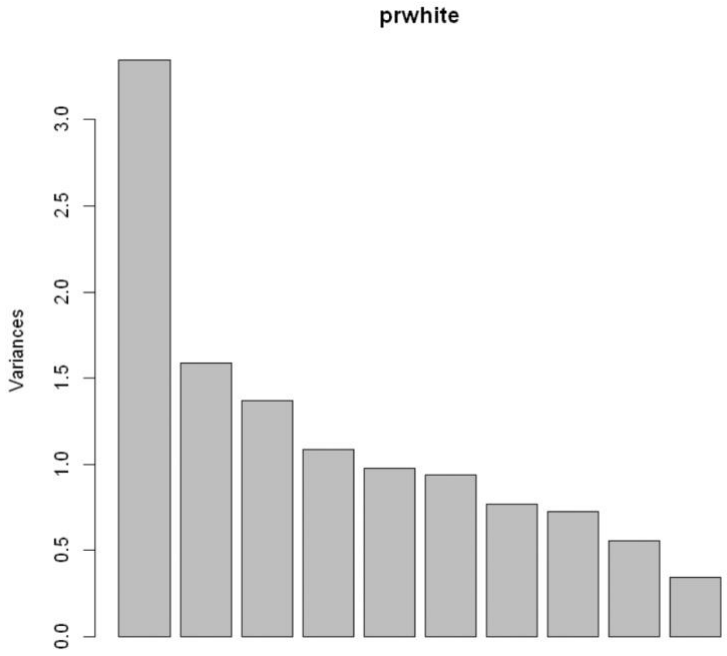
(b)

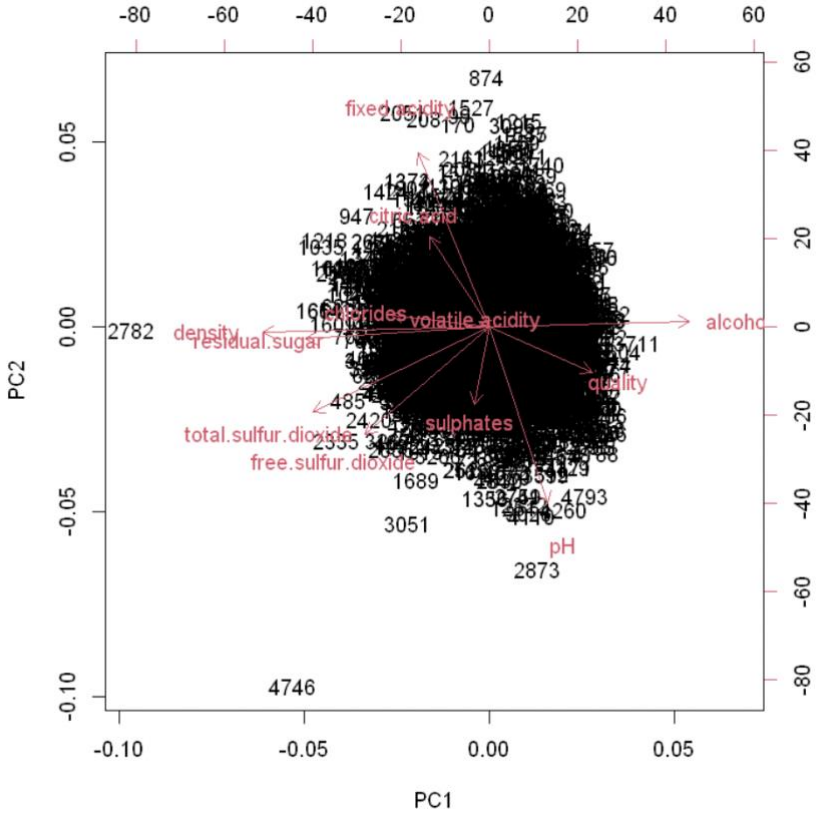
Predictions



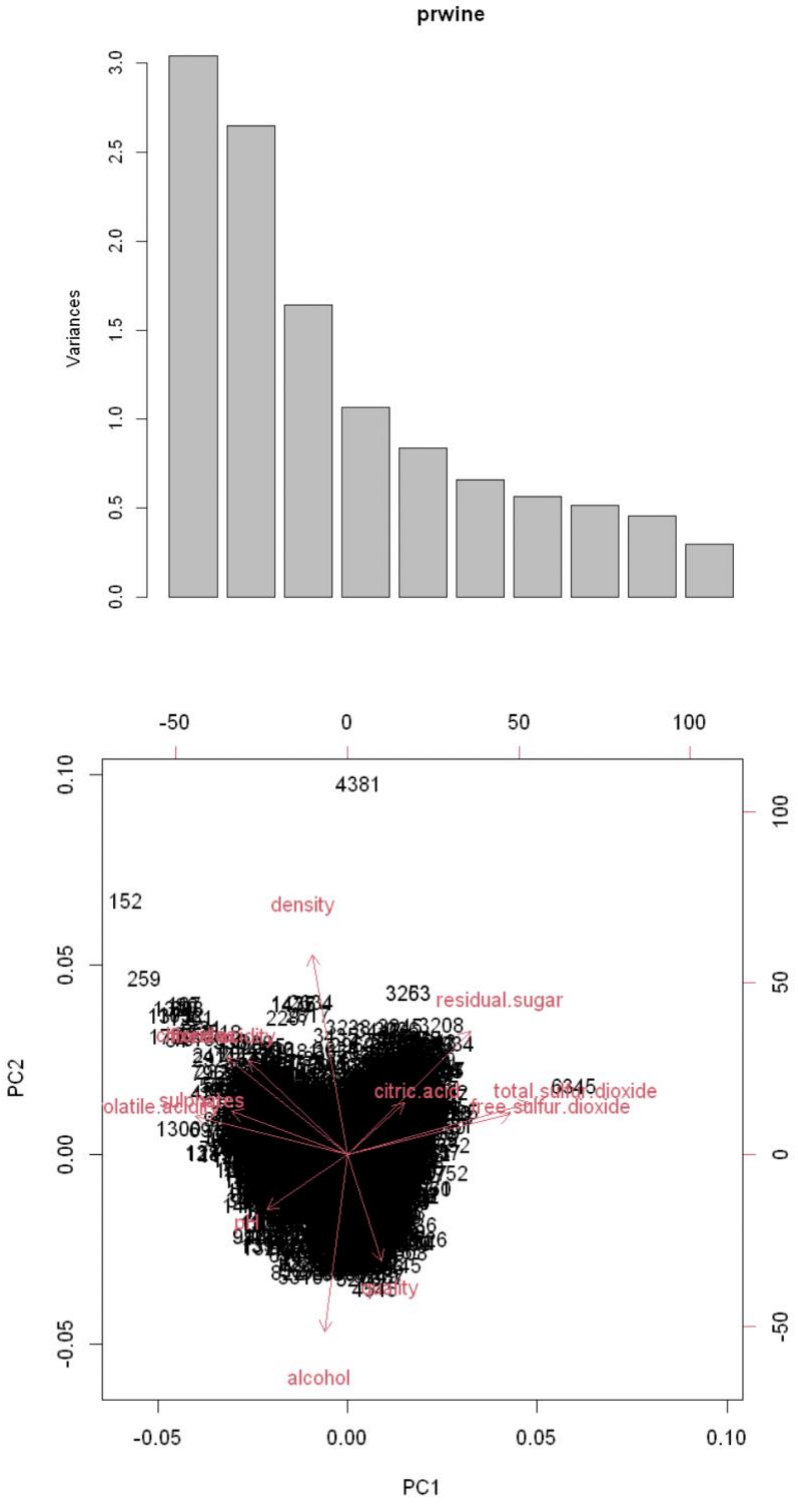

-20
0
20
40
0.10
0.05
0.00
-0.05
PC2
40
20
0
-20
density
pH
alcohol
quality
-0.05
0.00
0.05
0.10
PC1

(C)


prwhite
Variances
0.0
0.5
1.0
1.5
2.0
2.5
3.0

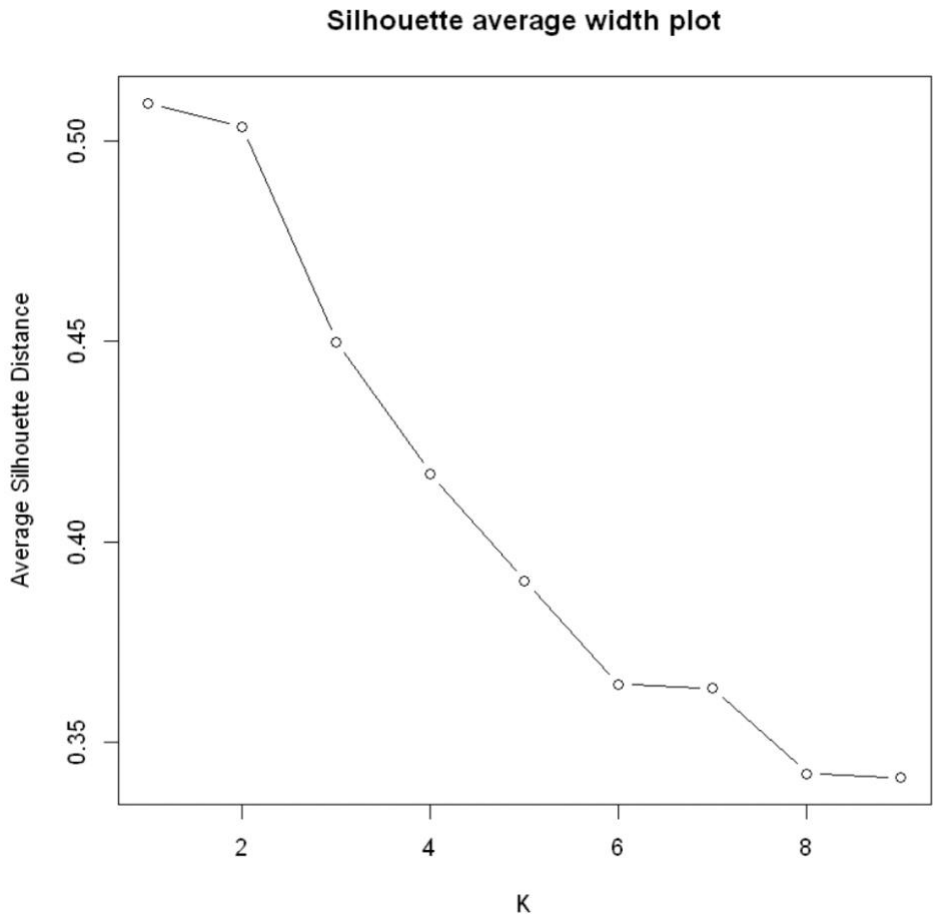

-80
-60
-40
-20
0
20
40
60
874
fixed.acidity
1527
947
citric.acid
chlorides
volatile.acidity
alcohol
2782
density
residual.sugar
quality
485
sulphates
total.sulfur.dioxide
free.sulfur.dioxide
1689
4793
3051
pH
2873
4746
0.05
0.00
-0.05
-0.10
PC2
PC1

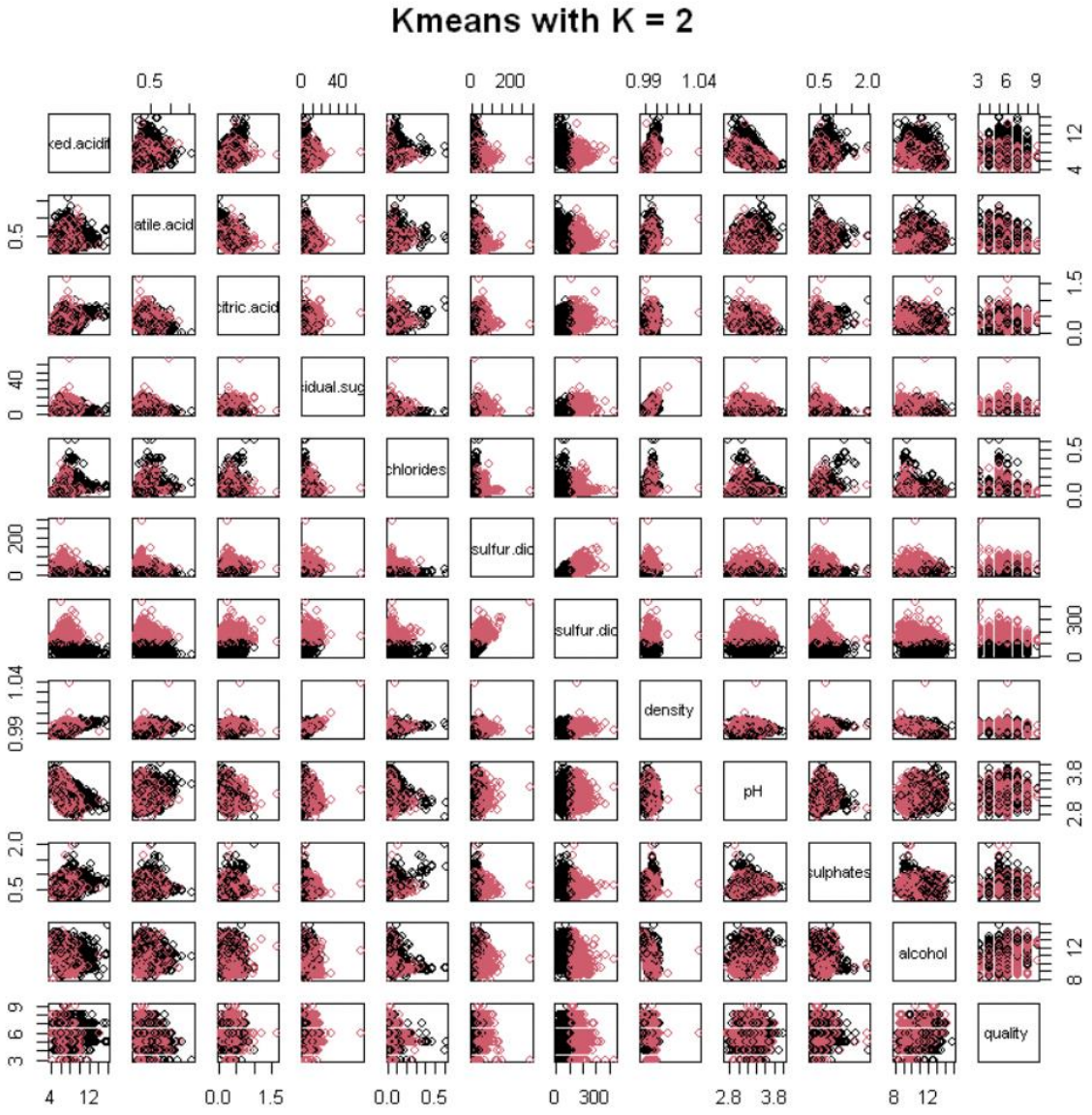
(d)


prwine
Variances
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-50
0
50
100
4381
152
density
259
3263
residual.sugar
6345
citric.acid
total.sulfur.dioxide
free.sulfur.dioxide
alcohol
PC2
PC1
0.10
0.05
0.00
-0.05
-0.05
0.00
0.05
0.10
100
50
0
-50

(e)

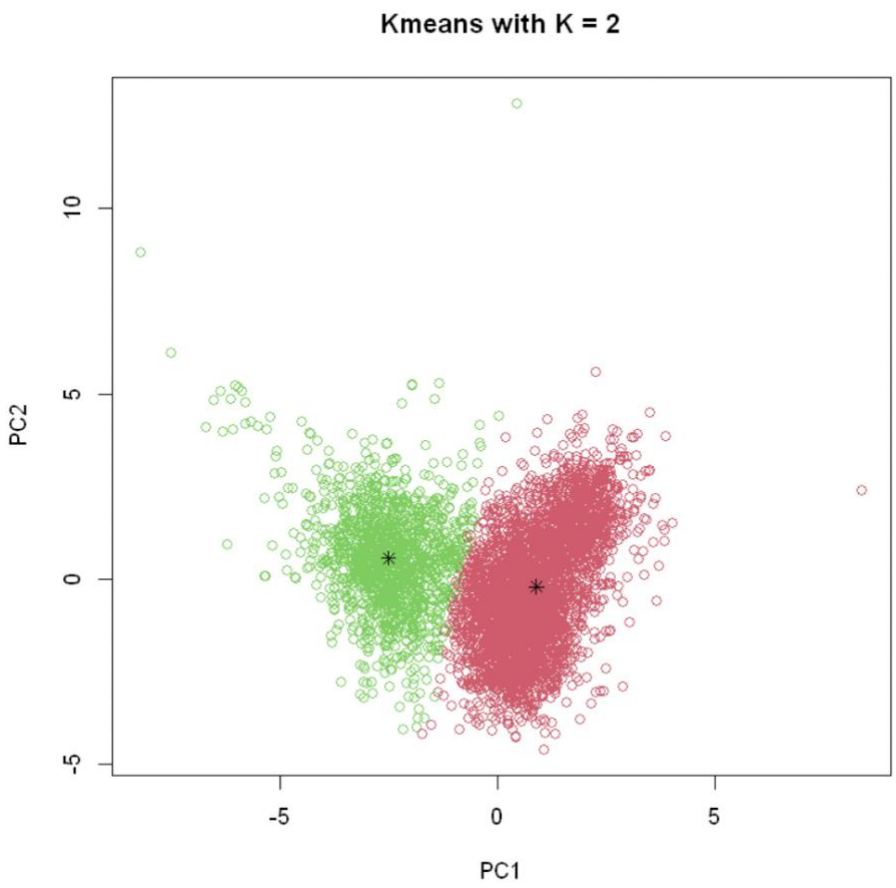
The silhouette plot reveals that the highest average distance occurs at 2, which aligns with the presence of two distinct types of wines in the dataset. Consequently, we can proceed with clustering the dataset using the k-means algorithm, specifying k as 2.





(F)

Based on the scree plot, it is evident that the first four principal components account for approximately 80% of the variation in the data. Therefore, we can utilize these four principal components to assess the performance of the k-means algorithm.

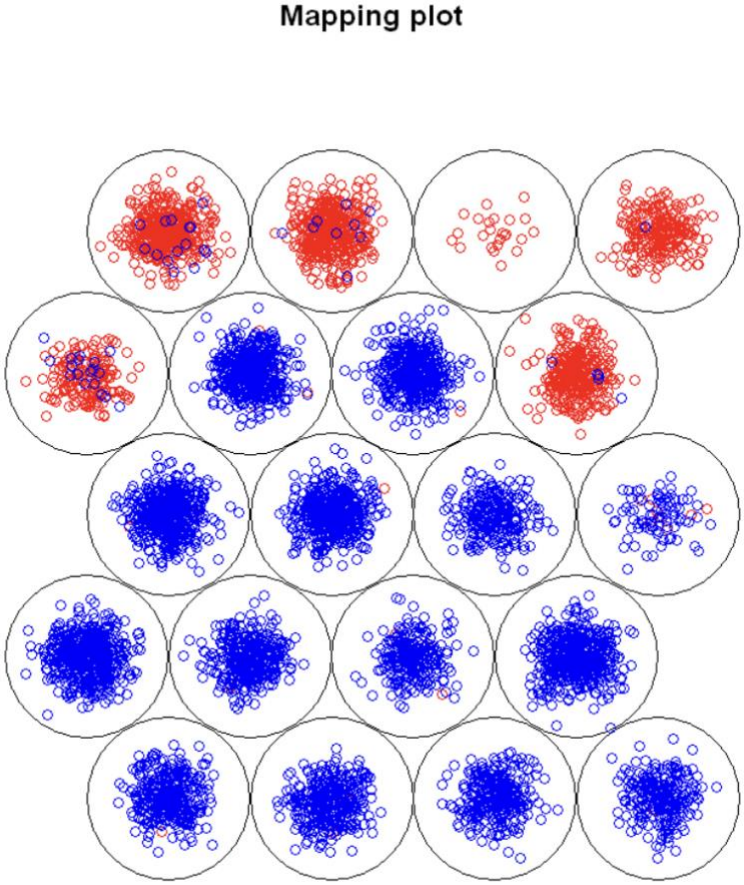


(g)

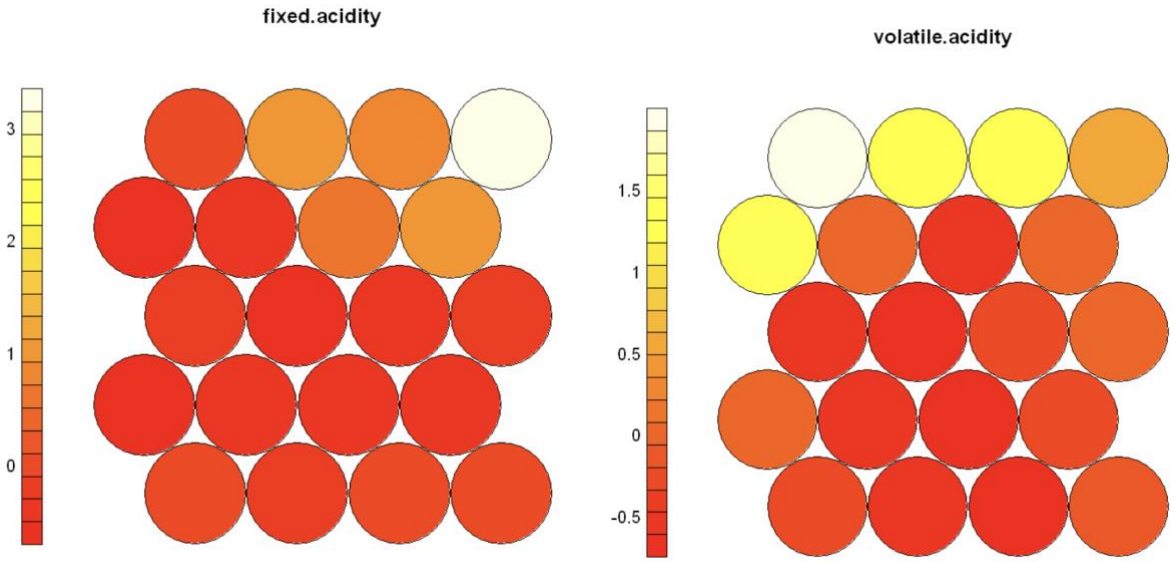
The observations from both E and F indicate that there is minimal disparity in the data. In both cases, the clusters appear almost identical, making it challenging to distinguish between them. This similarity arises because PCA retains sufficient information that enables clearer differentiation between the clusters.

3.

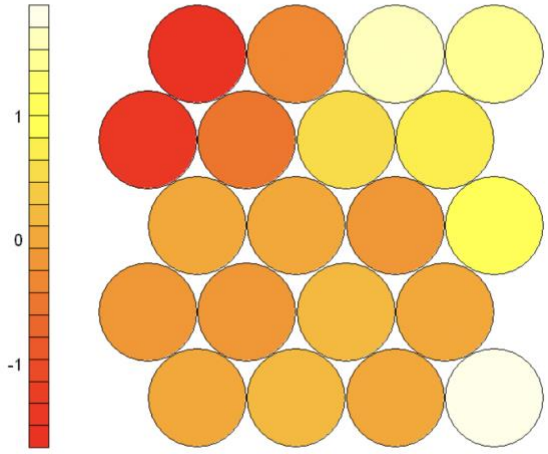
(A)


Mapping plot

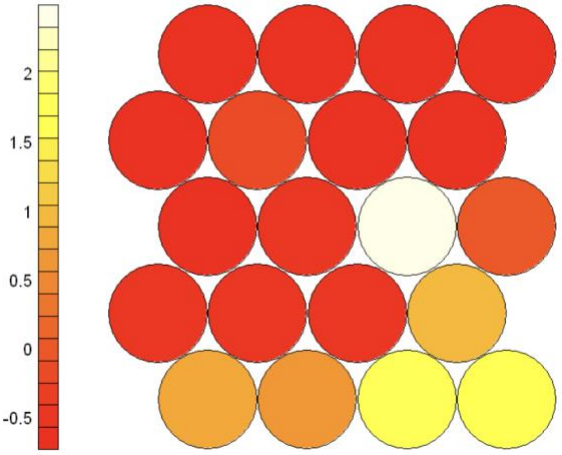
(B)


fixed.acidity
3
2
1
0
volatile.acidity
1.5
1
0.5
0
-0.5

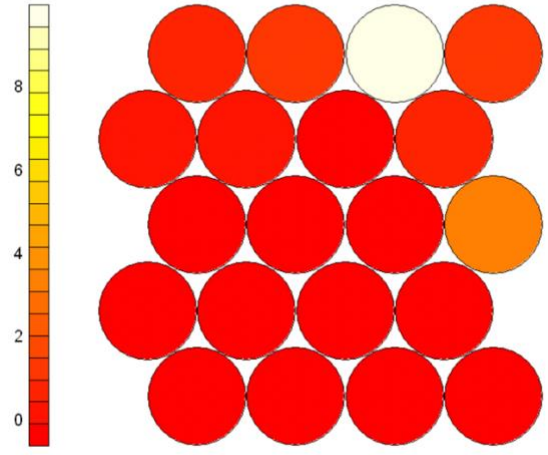
citric.acid



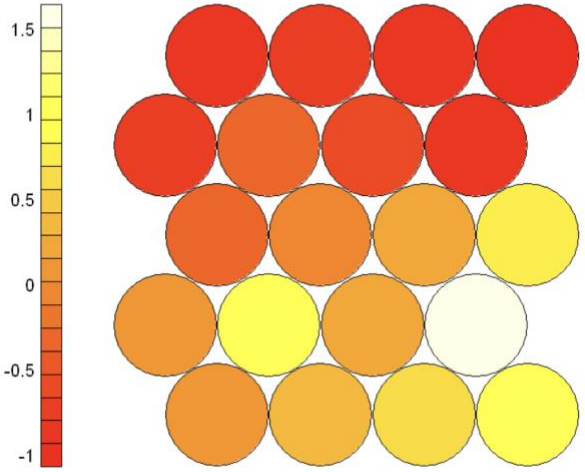
residual.sugar



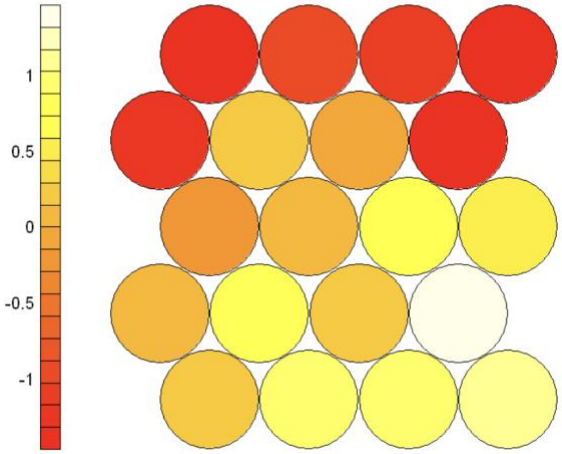
chlorides



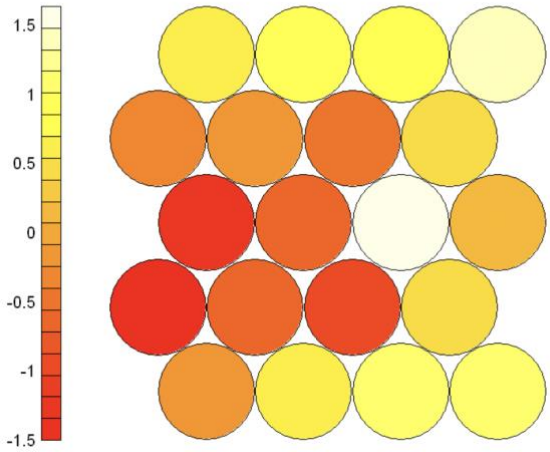
free.sulfur.dioxide



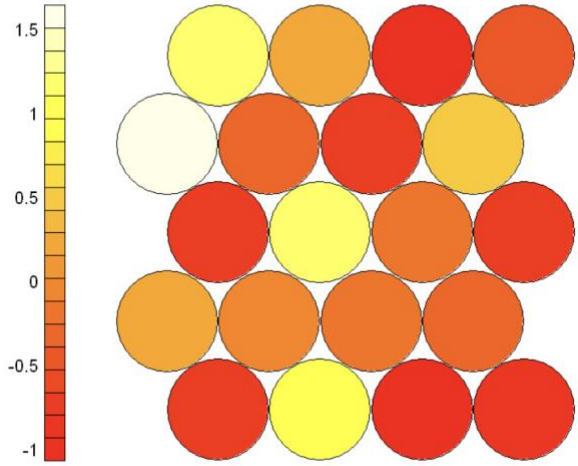
total.sulfur.dioxide



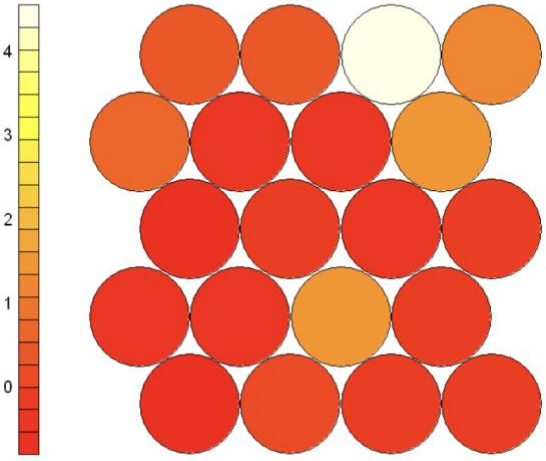
density

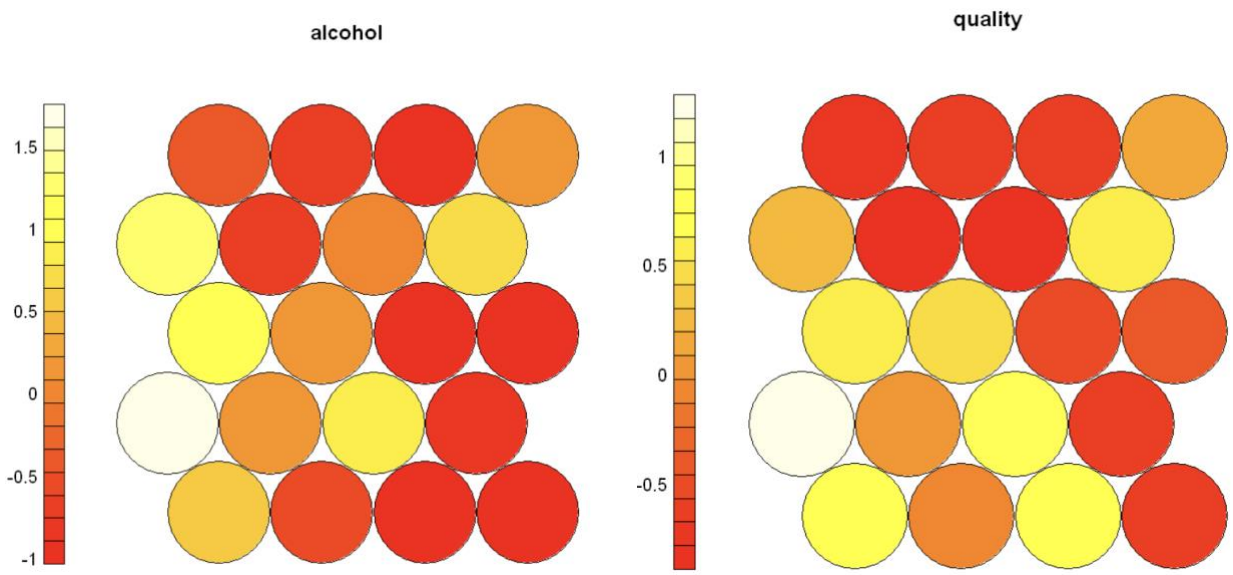


pH

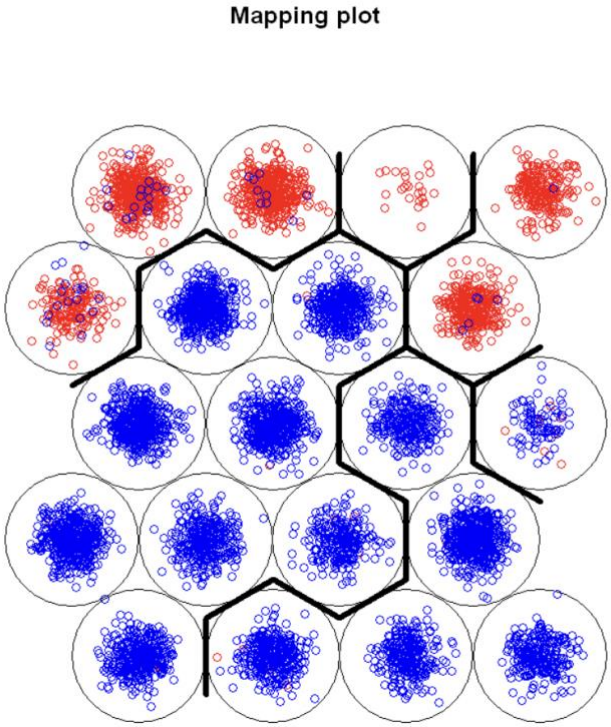


sulphates





(C)



4.

```
> head(titanic)
  Survived Pclass                                                Name    Sex Age Siblings.Spouses.Aboard Parents.Children.Aboard    Fare
1        0      3                             Mr. Owen Harris Braund   male  22                       1                       0  7.2500
2        1      1 Mrs. John Bradley (Florence Briggs Thayer) Cumings female  38                       1                       0 71.2833
3        1      3                              Miss. Laina Heikkinen female  26                       0                       0  7.9250
4        1      1        Mrs. Jacques Heath (Lily May Peel) Futrelle female  35                       1                       0 53.1000
5        0      3                            Mr. William Henry Allen   male  35                       0                       0  8.0500
6        0      3                                    Mr. James Moran   male  27                       0                       0  8.4583
```

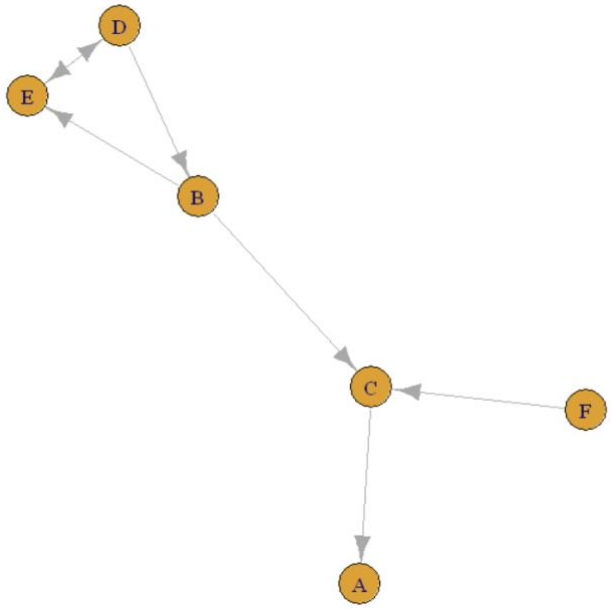
```
> prop.table(table(titanic$Sex, titanic$Age, titanic$Survived), margin = c(1, 2))
, ,  = 0

             0.42       0.67       0.75       0.83       0.92          1          2          3          4          5          6          7
  female                    0.00000000                    0.00000000 0.57142857 0.66666667 0.00000000 0.00000000 0.50000000 0.50000000
  male   0.00000000 0.00000000 0.00000000 0.00000000 0.40000000 0.75000000 0.00000000 0.50000000 1.00000000 0.00000000 0.66666667

               8          9         10         11         12         13         14       14.5         15         16         17         18
  female 0.75000000 1.00000000 1.00000000 1.00000000 1.00000000 0.00000000 0.40000000 1.00000000 0.00000000 0.14285714 0.16666667 0.35294118
  male   0.50000000 0.50000000 1.00000000 0.66666667 0.00000000 1.00000000            1.00000000 0.92307692 0.90000000 0.89473684

              19         20       20.5         21         22         23       23.5         24       24.5         25         26         27
  female 0.11111111 0.66666667 0.58333333 0.12500000 0.16666667 0.11111111 0.50000000 0.33333333 0.14285714
  male   0.87500000 0.85000000 1.00000000 0.95454545 0.95652174 0.89473684 1.00000000 0.87500000 1.00000000 0.77777778 0.80000000 0.63157895

              28       28.5         29         30       30.5         31         32       32.5         33         34       34.5         35
  female 0.27272727 0.28571429 0.18181818 1.00000000 0.22222222 0.50000000 0.00000000 0.00000000 0.00000000 0.00000000
  male   0.92307692 1.00000000 0.66666667 0.95454545 1.00000000 0.70000000 0.52941176 1.00000000 1.00000000 0.83333333 1.00000000 0.75000000

              36       36.5         37         38         39         40       40.5         41         42         43         44         45
  female 0.00000000 1.00000000 0.16666667 0.33333333 0.33333333 0.40000000 0.00000000 0.50000000 0.33333333 0.50000000
  male   0.75000000 1.00000000 0.72727273 0.83333333 0.91666667 0.88888889 1.00000000 1.00000000 0.75000000 1.00000000 0.83333333 0.62500000

            45.5         46         47         48         49         50         51         52         53         54         55       55.5
  female 0.00000000 0.50000000 0.14285714 0.00000000 0.20000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
  male   1.00000000 0.80000000 0.87500000 0.40000000 0.66666667 0.80000000 0.83333333 0.75000000 1.00000000 1.00000000 1.00000000

              56         57         58         59         60         61         62         63         64         65         66         69
  female 0.00000000 1.00000000 0.00000000 0.00000000 0.50000000 0.00000000
  male   0.66666667 1.00000000 1.00000000 1.00000000 0.75000000 1.00000000 0.66666667 1.00000000 1.00000000 1.00000000 1.00000000

              70       70.5         71         74         80
  female
  male   1.00000000 1.00000000 1.00000000 1.00000000 0.00000000

, ,  = 1
```

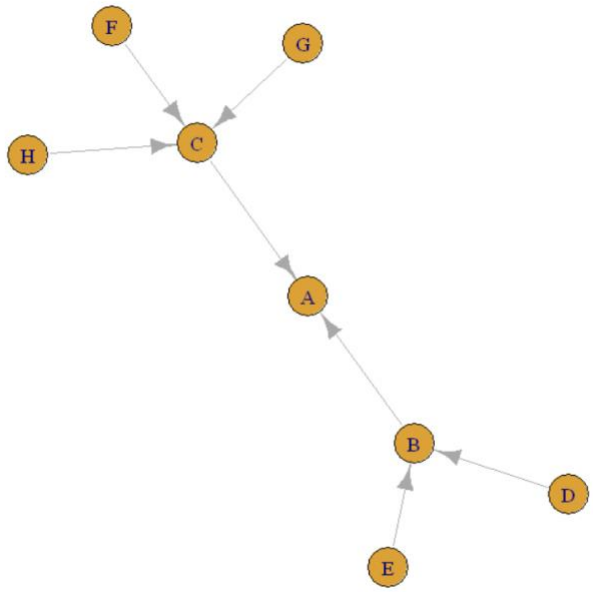
```
            0.42       0.67       0.75       0.83       0.92          1          2          3          4          5          6          7
female                        1.00000000                       1.00000000 0.42857143 0.33333333 1.00000000 1.00000000 0.50000000 0.50000000
male   1.00000000 1.00000000            1.00000000 1.00000000 0.60000000 0.25000000 1.00000000 0.50000000 0.00000000 1.00000000 0.33333333

               8          9         10         11         12         13         14       14.5         15         16         17         18
female 0.25000000 0.00000000 0.00000000 0.00000000 0.00000000 1.00000000 0.60000000 0.00000000 1.00000000 0.85714286 0.83333333 0.64705882
male   0.50000000 0.50000000 0.00000000 0.33333333 1.00000000            0.00000000            0.00000000 0.07692308 0.10000000 0.10526316

              19         20       20.5         21         22         23       23.5         24       24.5         25         26         27
female 0.88888889 0.33333333            0.41666667 0.87500000 0.83333333            0.88888889            0.50000000 0.66666667 0.85714286
male   0.12500000 0.15000000 0.00000000 0.04545455 0.04347826 0.10526316 0.00000000 0.12500000 0.00000000 0.22222222 0.20000000 0.36842105

              28       28.5         29         30       30.5         31         32       32.5         33         34       34.5         35
female 0.72727273            0.71428571 0.81818182 0.00000000 0.77777778 0.50000000 1.00000000 1.00000000 1.00000000            1.00000000
male   0.07692308 0.00000000 0.33333333 0.04545455 0.00000000 0.30000000 0.47058824 0.00000000 0.00000000 0.16666667 0.00000000 0.25000000

              36       36.5         37         38         39         40       40.5         41         42         43         44         45
female 1.00000000            0.00000000 0.83333333 0.66666667 0.66666667            0.60000000 1.00000000 0.50000000 0.66666667 0.50000000
male   0.25000000 0.00000000 0.27272727 0.16666667 0.08333333 0.11111111 0.00000000 0.00000000 0.25000000 0.00000000 0.16666667 0.37500000

            45.5         46         47         48         49         50         51         52         53         54         55       55.5
female             1.00000000 0.50000000 0.85714286 1.00000000 0.80000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
male   0.00000000 0.20000000 0.12500000 0.60000000 0.33333333 0.20000000 0.16666667 0.25000000            0.00000000 0.00000000 0.00000000

              56         57         58         59         60         61         62         63         64         65         66         69
female 1.00000000 0.00000000 1.00000000            1.00000000            0.50000000 1.00000000
male   0.33333333 0.00000000 0.00000000 0.00000000 0.25000000 0.00000000 0.33333333            0.00000000 0.00000000 0.00000000 0.00000000

              70       70.5         71         74         80
female
male   0.00000000 0.00000000 0.00000000 0.00000000 1.00000000

> titanic %>% filter(age_group == 1) %>% summarize(Child_Prob_Survived = mean(Survived))
  Child_Prob_Survived
1           0.5443038
> titanic %>% filter(age_group == 0) %>% summarize(Adult_Prob_Survived = mean(Survived))
  Adult_Prob_Survived
1           0.3700495
> titanic %>% filter(age_group == 1, Sex == "female") %>% summarize(Female_Child_Prob_Survived = mean(Survived))
  Female_Child_Prob_Survived
1                  0.5263158
> titanic %>% filter(age_group == 1, Sex == "male") %>% summarize(Male_Child_Prob_Survived = mean(Survived))
  Male_Child_Prob_Survived
1                0.5609756
```

```
> prop.table(g, margin =1)

       female      male
  0 0.1486239 0.8513761
  1 0.6812865 0.3187135
> prop.table(age, margin =1)

             0          1
  0 0.93394495 0.06605505
  1 0.87426901 0.12573099
> class <- table(titanic$Survived, titanic$Pclass)
> prop.table(class, margin =1)

            1         2         3
  0 0.1467890 0.1779817 0.6752294
  1 0.3976608 0.2543860 0.3479532
```

```
> titanic %>% filter(Pclass == 1, Sex == "female") %>% summarize(Class1_F_Prob_Survived = mean(Survived))
  Class1_F_Prob_Survived
1              0.9680851
> titanic %>% filter(Pclass == 3, Sex == "male") %>% summarize(Class3_M_Prob_Survived = mean(Survived))
  Class3_M_Prob_Survived
1              0.1370262
```

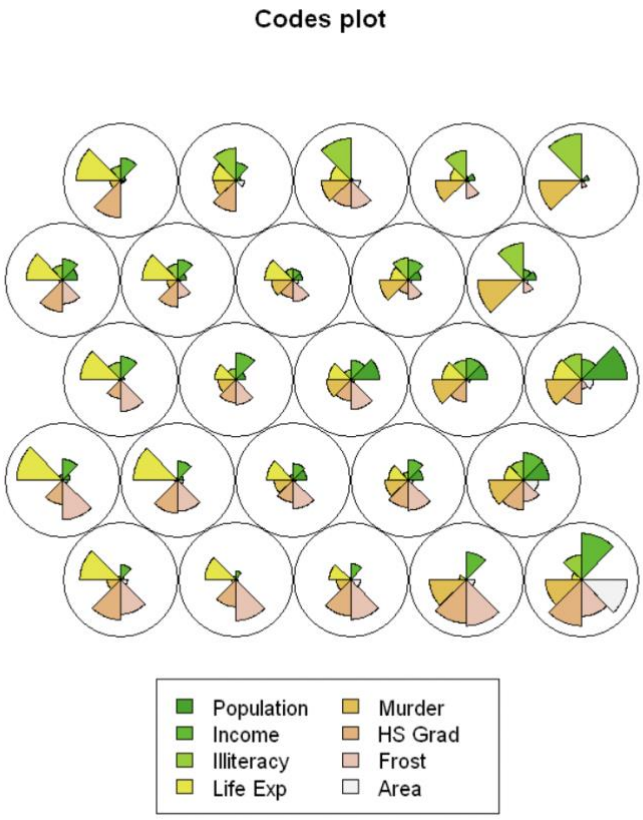
5.

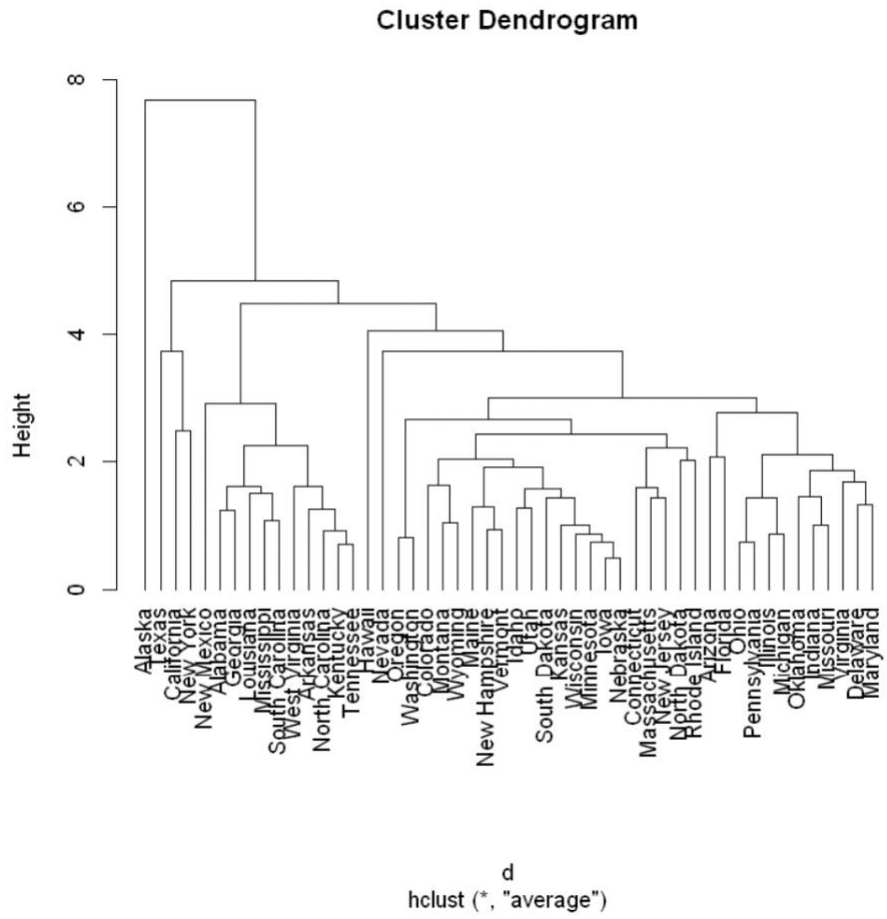




6.

```
> dim(statedata)
[1] 50  8
>
> summary(statedata)
   Population          Income          Illiteracy        Life Exp          Murder          HS Grad           Frost             Area
 Min.   :-0.8694   Min.   :-2.1772   Min.   :-1.0992   Min.   :-2.1742   Min.   :-1.6194   Min.   :-1.89526   Min.   :-2.0096   Min.   :-0.8167
 1st Qu.:-0.7094   1st Qu.:-0.7210   1st Qu.:-0.8941   1st Qu.:-0.5670   1st Qu.:-0.8203   1st Qu.:-0.62622   1st Qu.:-0.7351   1st Qu.:-0.3955
 Median :-0.3154   Median : 0.1354   Median :-0.3609   Median :-0.1517   Median :-0.1430   Median : 0.01758   Median : 0.1931   Median :-0.1929
 Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.1617   3rd Qu.: 0.6147   3rd Qu.: 0.6644   3rd Qu.: 0.7553   3rd Qu.: 0.8931   3rd Qu.: 0.74805   3rd Qu.: 0.6789   3rd Qu.: 0.1222
 Max.   : 3.7970   Max.   : 3.0582   Max.   : 2.6742   Max.   : 2.0273   Max.   : 2.0918   Max.   : 1.75709   Max.   : 1.6071   Max.   : 5.8094
>
> statedata = scale(statedata)
> summary(statedata)
   Population          Income          Illiteracy        Life Exp          Murder          HS Grad           Frost             Area
 Min.   :-0.8694   Min.   :-2.1772   Min.   :-1.0992   Min.   :-2.1742   Min.   :-1.6194   Min.   :-1.89526   Min.   :-2.0096   Min.   :-0.8167
 1st Qu.:-0.7094   1st Qu.:-0.7210   1st Qu.:-0.8941   1st Qu.:-0.5670   1st Qu.:-0.8203   1st Qu.:-0.62622   1st Qu.:-0.7351   1st Qu.:-0.3955
 Median :-0.3154   Median : 0.1354   Median :-0.3609   Median :-0.1517   Median :-0.1430   Median : 0.01758   Median : 0.1931   Median :-0.1929
 Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.1617   3rd Qu.: 0.6147   3rd Qu.: 0.6644   3rd Qu.: 0.7553   3rd Qu.: 0.8931   3rd Qu.: 0.74805   3rd Qu.: 0.6789   3rd Qu.: 0.1222
 Max.   : 3.7970   Max.   : 3.0582   Max.   : 2.6742   Max.   : 2.0273   Max.   : 2.0918   Max.   : 1.75709   Max.   : 1.6071   Max.   : 5.8094
```





1. 0.530222559638633 2. 0.291425879300336 3. 0.37803771005104 4. 0.321126308804489
5. 0.272373298240413 6. 0.251923480012512 7. 0.256732418449407 8. 0.237350663799007
9. 0.233627900939559

Among the given clustering methods, hierarchical clustering demonstrates the highest average distance, reaching 0.53 when the number of clusters (k) is set to 2. Consequently, we will cut the hierarchical tree at 2 and employ this clustering solution to visualize and draw clusters on the self-organizing map (SOM).

Mapping plot