Project 3

MACT423301- MACT423302- Applied Multivariate Analysis (Spring 2023)

Group members:

Maya Hany Elshweiky 900204233

Omar Ahmed 900201077

Dr. Ali Hadi

Table of contents:

1. **Statement of the problem**
    - **1.1. What question(s) can be answered by analysis of the data**
    - **1.2. Background information**
2. **Data description**
    - **2.1. Variables' Description**
3. **Data visualization and analysis**
    - **3.1. The Relationship Between Variables**
    - **3.2. Testing Normality of the Data**
        - **3.2.1. Data Standardization**
4. **Principal Components Analysis (PCA)**
    - **4.1. Classical PCA**
    - **4.2. Robust PCA**
        - **4.2.1 MV BACON**
        - **4.2.2 Robust PCA Implementation**
5. **Conclusion**
6. **Appendix**

Red Wine Quality Data Analysis

1. Statement of the problem:

1.1 What question(s) can be answered by the analysis of the data?

The following questions can be answered by analyzing the data:

- How good/accurate is our analysis and method of PCA
- Which variables are most important for explaining the variation in the data?
- Are there any patterns or groupings in the data?
- Are there any outliers or extreme values that are affecting the results?
- Can the data be represented in a lower-dimensional space while still maintaining most of the variation?

1.2 Background Information

We will analyze the "Wine Quality" data set to assess the questions above, the Data set contains the following variables: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH level, sulfates, alcohol level and quality of wine; the units of measurements for the listed variables are mentioned below; all of the variables are quantitative variables except for the "quality of wine" variable. First, we will scale the data by standardizing our data the variables will have the same variance (equal to 1). Scaling is important because PCA is a variance-based method, and each variable in the dataset should be of equal importance. Principal Components Analysis is the linear combination of x-variables, and the first principal component has maximum variance (among all linear combinations). Then, we will apply Classical PCA using the princomp() function. The summary() function is then used to view the results of the analysis, including the standard deviation, proportion of variance, and cumulative proportion explained by each principal component. Afterward, we will plot the results using plot(pc), which displays a scree plot, and we will also analyze the results of pc$scores, pc$sd, and pc$loadings. Finally, we will apply Robust PCA using the mvBACON() function and compare it with the Classical PCA approach.

2. Data Description

Who: 1599 observations of wine quality based on physicochemical tests

When: The data was collected in 2009

Where: The data was obtained from kaggle.com and the Citation for the data set is found below:

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

Citation: P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

What: There are 12 variables in the data, for each, we have the wine quality and eleven chemical attributes (quantitative), and their description is shown below in the table.

Variable	Type	Unit of measurement	Description
fixed acidity	Quantitative (Numeric)	tartaric acid - g / dm^3	Fixed acids, numeric from 3.8 to 15.9
volatile acidity	Quantitative (Numeric)	acetic acid - g / dm^3	Volatile acids, numeric from 0.1 to 1.6
citric acid	Quantitative (Numeric)	g / dm^3	Citric acids, numeric from 0.0 to 1.7
residual sugar	Quantitative (Numeric)	g / dm^3	residual sugar, numeric from 0.6 to 65.8

chlorides	Quantitative (Numeric)	sodium chloride - g / dm^3	numeric from 0.01 to 0.61
free sulfur dioxide	Quantitative (Numeric)	mg / dm^3	Free sulfur dioxide, numeric: from 1 to 289
total sulfur dioxide	Quantitative (Numeric)	mg / dm^3	Total sulfur dioxide, numeric: from 6 to 440
density	Quantitative (Numeric)	g / dm^3	Density, numeric: from 0.987 to 1.039
pH	Quantitative (Numeric)	0-14 scale (0 being most acidic, 14 most alkaline)	pH, numeric: from 2.7 to 4.0
sulphates	Quantitative (Numeric)	(potassium sulphate - g / dm3	Sulfates, numeric: from 0.2 to 2.0
alcohol	Quantitative (Numeric)	% by volume	the percent alcohol content of the wine,

			numeric: from 8.0 to 14.9
quality	Categorical	0-10 scale	Wine quality rating between 0 (very bad) and 10 (very excellent)

3. Data Visualization and Analysis

3.1 The Relationship Between Variables

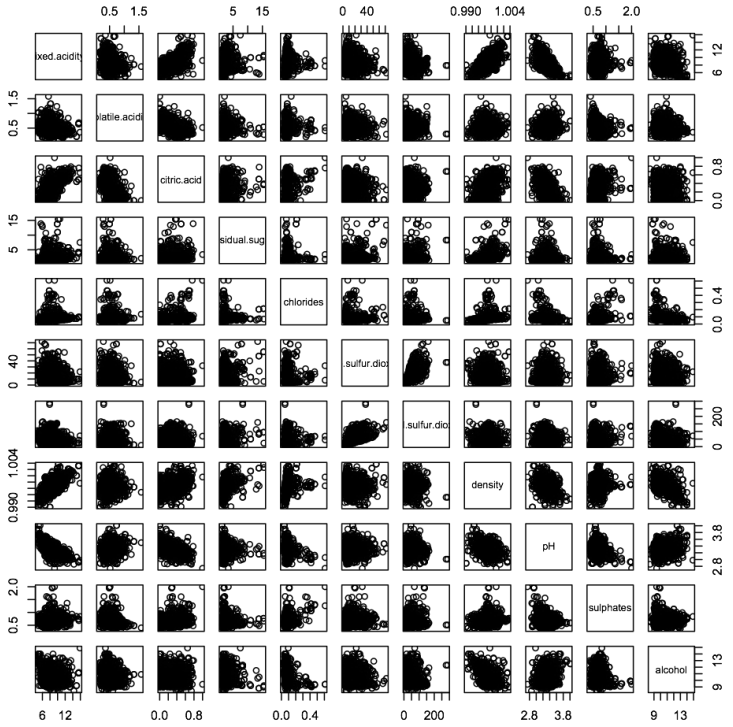


Figure 1: pairs of variables before scaling the data

As shown in Figure 1, we can see that many of the variables have linear relationships with each other, for example, fixed acidity and density have a positive linear relationship, while fixed

acidity and pH have a negative linear relationship. There are also several variables that do not have linear relationships with each other, such as fixed acidity and sulfur dioxide, residual sugar and sulfur dioxide, residual sugar, and fixed acidity, and so much more. Lastly, we can see that there are some outliers in the data set, as we can see points on the graph that are on the edges of the plot and are far away from the rest of the data.

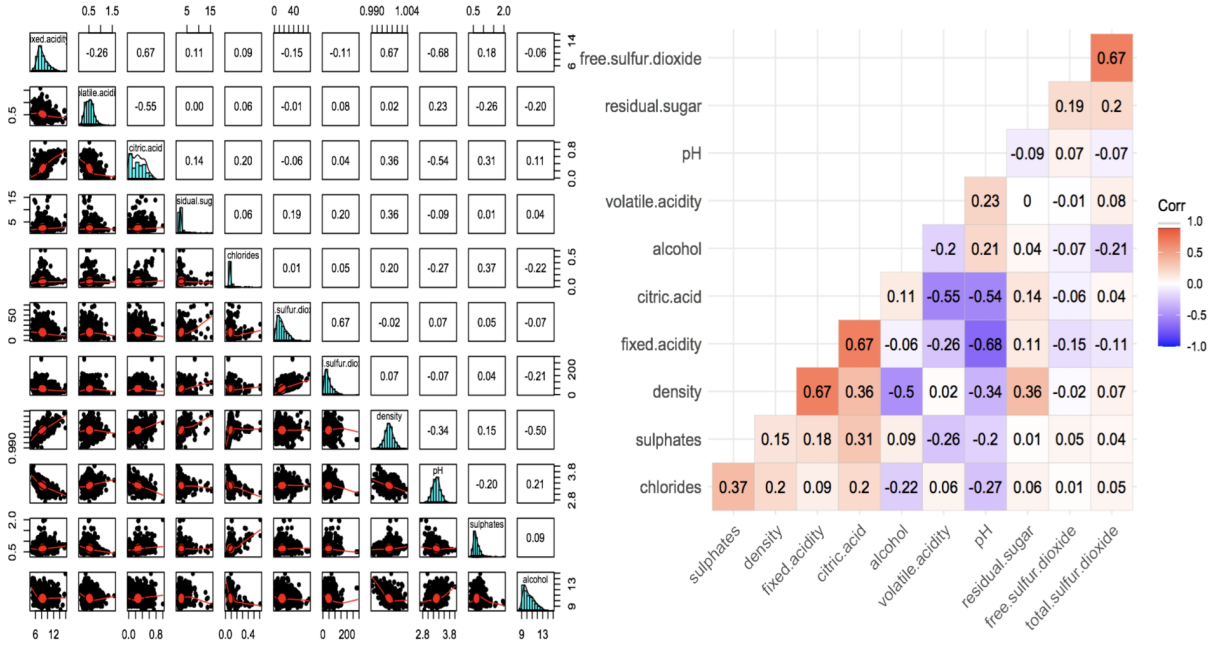


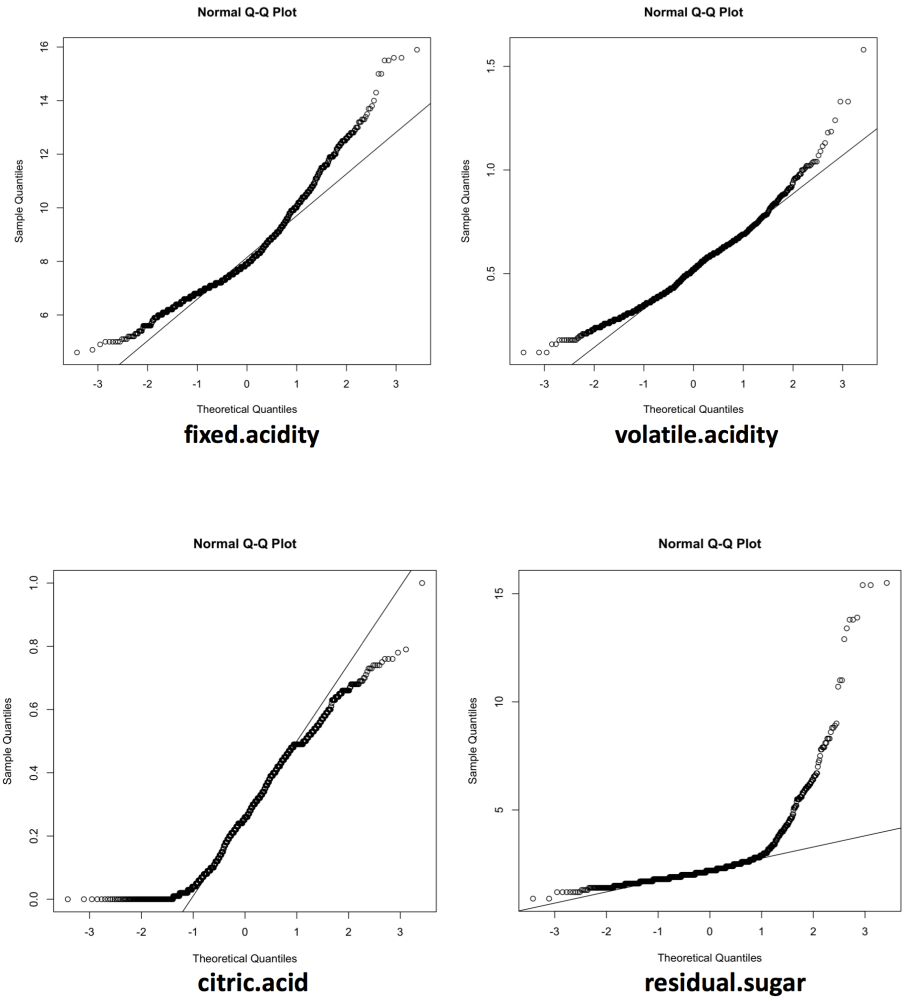
Figure 2&3: scatter plot of matrices, correlation, and histogram and correlation matrix

This is a very important graph for our analysis; firstly, it shows the correlation between the variables and shows their relationship, secondly, it plots the histograms of the variables on the diagonals, which will help us analyze which distributions the variables follow, for example, we can see that the density variable and the pH variable follows a normal distribution; Sulfur dioxide follows a gamma distribution. The plot and correlation matrix above also shows the correlation between the variables, the correlation numbers are not that high, which indicates that not many variables have strong correlations, the strongest positive correlation value we obtained was 0.67, which is the correlation between fixed acidity and density, the weakest positive correlation was 0.01, which is the correlation between chlorides and sulfur dioxide. And finally, the strongest negative correlation was between fixed acidity and pH level, with a value of -0.68. These values indicate that the relationship or dependency between the variables is not that strong and that they are to some extent independent.

3.2 Testing Normality of the Data

As shown in the previous analysis, not all variables have the same distribution; therefore, we will be using the Q-Q plot to test the normality of the variables.

Figure 3: Q-Q plots of four different variables are shown below



As shown in the QQ plots above, not all variables follow a normal distribution as the data is not normally distributed, and the points deviate from the QQ line such as but not limited to fixed.acidity, volatile.acidity, citric.acid, and residual.sugar.

3.2.1. Data Standardization

After standardizing the data using scale(x) function, our data is prepared for implementing PCA as the method assumes that the data is scaled. Now as shown in the output below, the mean of all variables is equal to zero, and the standard deviation is 0.999, which is approximately 1.

```
> # scaling the data
> z=scale(x, center = TRUE, scale = TRUE)
> # using only numeric values so we can procced with PCA
> z=z[,1:11]
> z=as.matrix(z)
> summary(z) # mean of all variables is equal to zero
 fixed.acidity     volatile.acidity    citric.acid       residual.sugar
 Min.   :-2.1364   Min.   :-2.27757   Min.   :-1.39104   Min.   :-1.1623
 1st Qu.:-0.7005   1st Qu.:-0.76969   1st Qu.:-0.92903   1st Qu.:-0.4531
 Median :-0.2410   Median :-0.04367   Median :-0.05634   Median :-0.2403
 Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
 3rd Qu.: 0.5056   3rd Qu.: 0.62649   3rd Qu.: 0.76501   3rd Qu.: 0.0434
 Max.   : 4.3538   Max.   : 5.87614   Max.   : 3.74240   Max.   : 9.1928
   chlorides        free.sulfur.dioxide total.sulfur.dioxide    density
 Min.   :-1.60344   Min.   :-1.4221     Min.   :-1.2302      Min.   :-3.53762
 1st Qu.:-0.37111   1st Qu.:-0.8485     1st Qu.:-0.7438      1st Qu.:-0.60757
 Median :-0.17989   Median :-0.1792     Median :-0.2574      Median : 0.00176
 Mean   : 0.00000   Mean   : 0.0000     Mean   : 0.0000      Mean   : 0.00000
 3rd Qu.: 0.05383   3rd Qu.: 0.4900     3rd Qu.: 0.4722      3rd Qu.: 0.57664
 Max.   :11.12355   Max.   : 5.3656     Max.   : 7.3728      Max.   : 3.67890
       pH             sulphates          alcohol
 Min.   :-3.69924   Min.   :-1.9359   Min.   :-1.8983
 1st Qu.:-0.65494   1st Qu.:-0.6380   1st Qu.:-0.8661
 Median :-0.00721   Median :-0.2251   Median :-0.2092
 Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.57574   3rd Qu.: 0.4239   3rd Qu.: 0.6353
 Max.   : 4.52687   Max.   : 7.9162   Max.   : 4.2011
> sd(z) # standard deviation of z is 0.999 approximately 1
[1] 0.9997157
```

Figure 4: R output of scaled data

4. PCA

Principal Components Analysis (PCA) is a statistical method that is used to transform a set of correlated variables into a smaller set of uncorrelated variables (called principal components) that capture most of the variation in the original data. This can be useful for reducing the dimensionality of the data, simplifying analysis, identifying important variables, and visualizing patterns.

4.1 Classical PCA

Classical PCA is a common method for performing PCA. It assumes that the data is standardized. The method calculates the covariance matrix of the dataset, then uses

eigendecomposition or singular value decomposition to find the principal components. The principal components are sorted by descending order of their eigenvalues, which represent the amount of variation explained by each component. The PCA's goal is to reduce our data's dimension without loss of data.

```
> pc=princomp(z, cor=T);
> summary(pc,loadings=T)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4     Comp.5     Comp.6
Standard deviation     1.7604353 1.3877715 1.2452082 1.1014684 0.97943457 0.81216271
Proportion of Variance 0.2817393 0.1750827 0.1409585 0.1102939 0.08720837 0.05996439
Cumulative Proportion  0.2817393 0.4568220 0.5977805 0.7080744 0.79528275 0.85524714
                           Comp.7     Comp.8    Comp.9    Comp.10     Comp.11
Standard deviation     0.76406231 0.65035121 0.5870623 0.42583232 0.244045717
Proportion of Variance 0.05307193 0.03845061 0.0313311 0.01648483 0.005414392
Cumulative Proportion  0.90831906 0.94676967 0.9781008 0.99458561 1.000000000
```

Figure 4: R output of PCA and Summary

After applying PCA to our data, we will analyze it using three rules. First, we will look at the Eigenvalues, which is the variance of our components, for scaled data, the rule tells us that we should choose the components with the Eigenvalues above 1; therefore, if we look at the importance of components, we will find the first row of the output gives us the Standard Deviation, which is the square root of the variance (eigenvalues). The eigenvalues of the first five components, respectively, to the nearest three decimal places are as follows: 3.099, 1.926, 1.551, 1.212, and 0.959. We will find that components 1,2,3 and 4 with values above 1, which means that they represent our original data, while component 5 is less than 1. Therefore, from the first rule, we get 4 components. Now, using the second rule, which is examining the cumulative proportion, if we look at the third row in the output, we will find that the cumulative proportion reaches 1 or close to 1 at component 9, component 10 adds only 0.01648, which is very small, so we can neglect it, same with component 11, so we see here that our data can be reduced to 9 components instead of 11 variables. In our data set, the first 2 rules give us different values and contradict each other, so we will look at the scree plot, which is the last rule for examining the number of components.

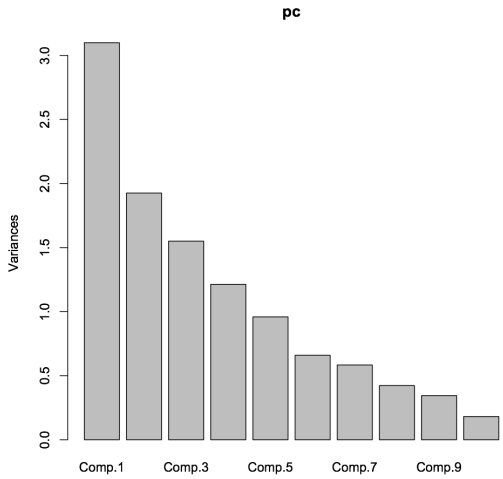


Figure 5: Scree plot of Non-robust PCA with the variances on the y-axis

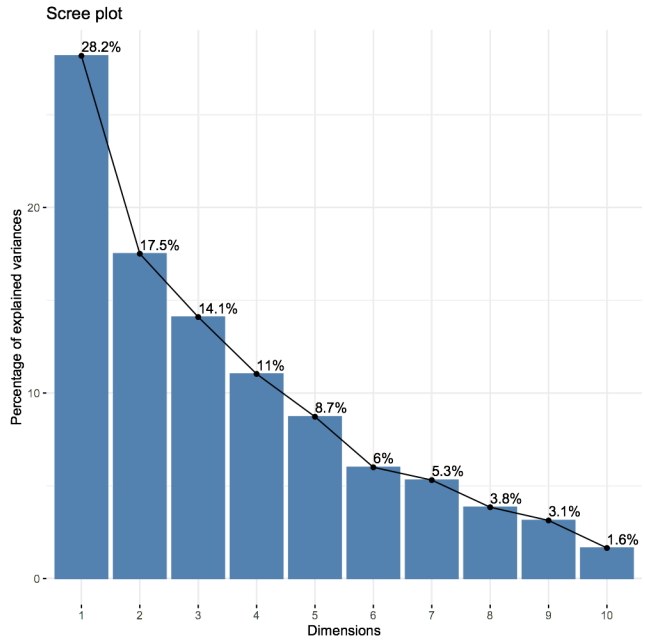


Figure 6: Scree plot of Non-robust PCA with the percentage of variances on the y-axis

As seen in the scree plots, the number of components can be determined by looking at the bars when they reach 0 or close to 0, so at component 9, the variances are very small, so we can determine the number of components to be 9, which is the same number of components we got when we examined the cumulative proportion. Overall, the non-robust PCA reduced our data to 9 components, which was not very effective, as it did not reduce our data's dimension significantly.

```
Loadings:
                      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
fixed.acidity          0.489  0.111  0.123  0.230         0.101  0.350  0.178  0.194
volatile.acidity      -0.239 -0.275  0.450        -0.219  0.411  0.534        -0.129
citric.acid            0.464  0.152 -0.238                      -0.105  0.378 -0.381
residual.sugar         0.146 -0.272 -0.101  0.373 -0.732        -0.291 -0.300
chlorides              0.212 -0.148        -0.666 -0.247  0.304 -0.370  0.357  0.111
free.sulfur.dioxide          -0.514 -0.429         0.159         0.117  0.205  0.635
total.sulfur.dioxide         -0.569 -0.322         0.222  0.136                -0.592
density                0.395 -0.234  0.339  0.174 -0.157 -0.391  0.170  0.239
pH                    -0.439                      -0.268 -0.522         0.561 -0.168
sulphates              0.243        -0.280 -0.551 -0.226 -0.381  0.447 -0.375
alcohol               -0.113  0.386 -0.472  0.122 -0.351  0.362  0.328  0.218
                      Comp.10 Comp.11
fixed.acidity          0.250   0.640
volatile.acidity      -0.366
citric.acid           -0.622
residual.sugar                 0.184
chlorides              0.218
free.sulfur.dioxide   -0.248
total.sulfur.dioxide   0.371
density                0.240  -0.567
pH                             0.341
sulphates             -0.112
alcohol                0.303  -0.315
```

Figure 6: R output pc$loadings

As shown above, the loadings output in Figure 6, in component , the variables with the highest variability were as follows: fixed.acidity - 0.489, citric.acid - 0.464, and pH - 0.439, while the following variables: free.sulfur.dioxide and total.sulfur.dioxide had no loadings, which indicates that they have no variability with component 1. In component 2, the variables with the highest variability were as follows: total.sulfur.dioxide - 0.569, free.sulfur.dioxide - 0.514, while the following variables: pH and sulphates have no variability with component 2. In component 3, the variables with the highest variability were as follows: alcohol - 0.472, volatile.acidity - 0.450 and freesulfur.dioxide - 0.429, while the following variables: chlorides and pH had no variability with component 3. Noting that for component 4 it had high variability with chlorides - 0.666, and for component 5 it had high variability with residual.sugar - 0.732. The rest of the components had small variability with the variables which means that they do not affect the components significantly.

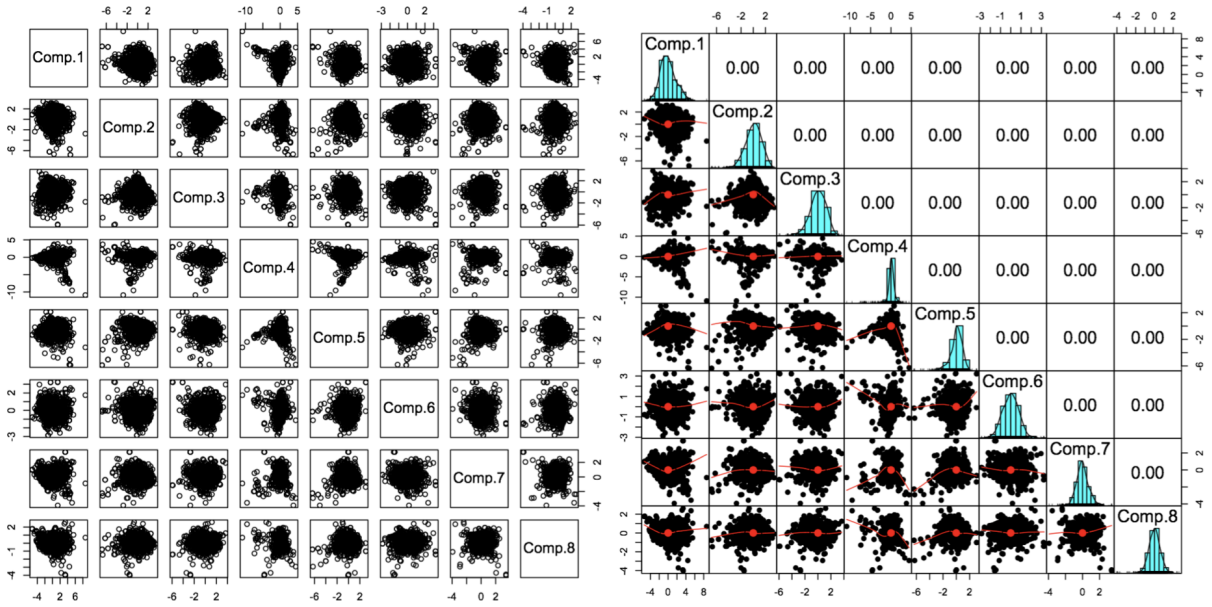


Figure 7: Pairwise graph of components

The pairwise function of the components shows the correlation between the components; since, PCA reduces our data's number of variables to fewer components that are orthogonal to each other, which means that the correlation between them all is zero. As illustrated in the figure above, there is no correlation between any of the components, which suggests that the PCA algorithm was efficient and resulted in components that are orthogonal to each other.

4.2 Robust PCA

4.2.1 MV BACON

```
> b=mvBACON(z)
rank(x.ord[1:m,] >= p  ==> chosen m =  44
MV-BACON (subset no. 1): 44 of 1599 (2.75 %)
MV-BACON (subset no. 2): 1254 of 1599 (78.42 %)
MV-BACON (subset no. 3): 1397 of 1599 (87.37 %)
MV-BACON (subset no. 4): 1422 of 1599 (88.93 %)
MV-BACON (subset no. 5): 1426 of 1599 (89.18 %)
MV-BACON (subset no. 6): 1428 of 1599 (89.31 %)
MV-BACON (subset no. 7): 1429 of 1599 (89.37 %)
MV-BACON (subset no. 8): 1429 of 1599 (89.37 %)
```

Figure 8: MV-BACON

We applied the BACON algorithm to get rid of outliers from our data set so we can apply the **Robust PCA.** Bacon took 7 iterations to reach the basic subset that is free from outliers, this basic subset contained 1429 out of 1599 observations, and the non-basic subset that contained the outliers had 170 observations. So we can now take the basic subset and use it in our PCA to get the number of components after removing the outliers from our data set.

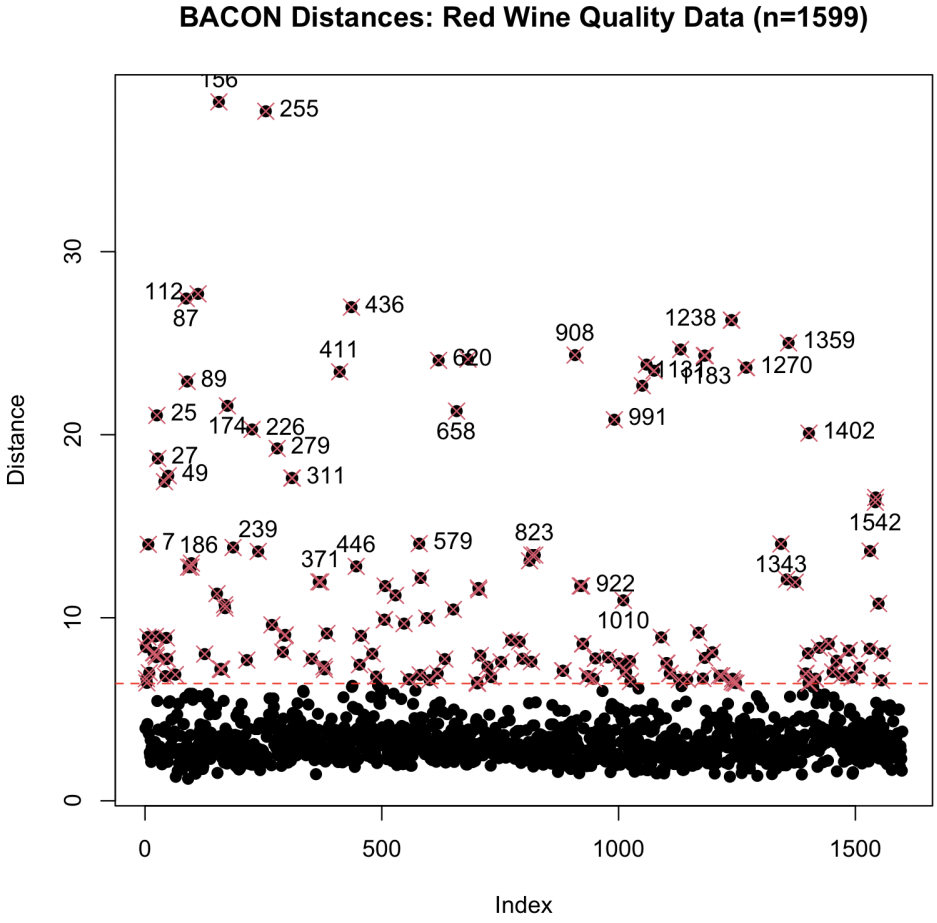


Figure 9: BACON scatter plot

As seen in the figure above, the observations above the red line are the outliers in our data set, and the observations below the red line are our basic subset that contains no outliers; some observations above the red-dotted line are very close to the line, which may be misclassified as outliers. Overall, this plot gives us an image of the basic and non-basic subsets of our data after applying the BACON algorithm to determine outliers.

4.2.2 Robust PCA Implementation

Robust PCA is a variant of PCA that is more resistant to outliers and heavy-tailed distributions. It uses a modified covariance matrix (such as a robust estimator) to calculate the principal components. This can help to prevent outliers from dominating the results and allow for a more accurate representation of the underlying patterns in the data.

```
> summary(pcr,loadings=T)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4     Comp.5     Comp.6     Comp.7
Standard deviation     1.8047324 1.4173208 1.2660380 1.0678341 0.90534090 0.82200378 0.75537066
Proportion of Variance 0.2960963 0.1826180 0.1457138 0.1036609 0.07451292 0.06142638 0.05187135
Cumulative Proportion  0.2960963 0.4787143 0.6244282 0.7280890 0.80260196 0.86402834 0.91589969
                           Comp.8     Comp.9    Comp.10     Comp.11
Standard deviation     0.60824880 0.57535613 0.41506458 0.227647783
Proportion of Variance 0.03363333 0.03009406 0.01566169 0.004711228
Cumulative Proportion  0.94953302 0.97962708 0.99528877 1.000000000

Loadings:
                  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
fixed.acidity      0.507                              0.225  0.227  0.282  0.177  0.342
volatile.acidity  -0.216 -0.403  0.319 -0.205         0.272  0.606  0.237 -0.158 -0.352
citric.acid        0.443  0.225 -0.188         0.185        -0.151  0.375 -0.304 -0.647
residual.sugar     0.223 -0.128        -0.742  0.201  0.208 -0.271 -0.429        -0.140
chlorides          0.244 -0.326  0.103 -0.227 -0.114 -0.856  0.106
free.sulfur.dioxide        -0.264 -0.646                              0.110  0.654 -0.221
total.sulfur.dioxide       -0.412 -0.551         0.166                      -0.621  0.330
density            0.420 -0.298  0.114        -0.313  0.291 -0.203  0.273         0.246
pH                -0.413                -0.308 -0.352        -0.450  0.542
sulphates          0.162  0.300 -0.309 -0.163 -0.741         0.345 -0.244 -0.150
alcohol                   0.494 -0.133 -0.469  0.320         0.315  0.301         0.308
                  Comp.11
fixed.acidity      0.631
volatile.acidity
citric.acid
residual.sugar     0.148
chlorides
free.sulfur.dioxide
total.sulfur.dioxide
density           -0.591
pH                 0.327
sulphates
alcohol           -0.338
```

Figure 10: R Output of Robust PCA

As shown in Figure 10, the robust PCA is computed using the data set after applying the BACON algorithm to identify the removing outliers, now, using rule 1, which is observing the standard deviations, we stop considering components that have a standard deviation < 1; components 1,2,3 and 4 have eigenvalues above 1, so the number of components extracted using rule 1 is four components. Now using rule 2, the cumulative proportion illustrates that the number of components is 9 components, as at comp. 9, we have a value of 0.979627, which is almost all of our data, and we only neglect about 2% of our data, which is negligible. So, now we will analyze the scree plot, which is rule 3, to find the number of components given by PCA.

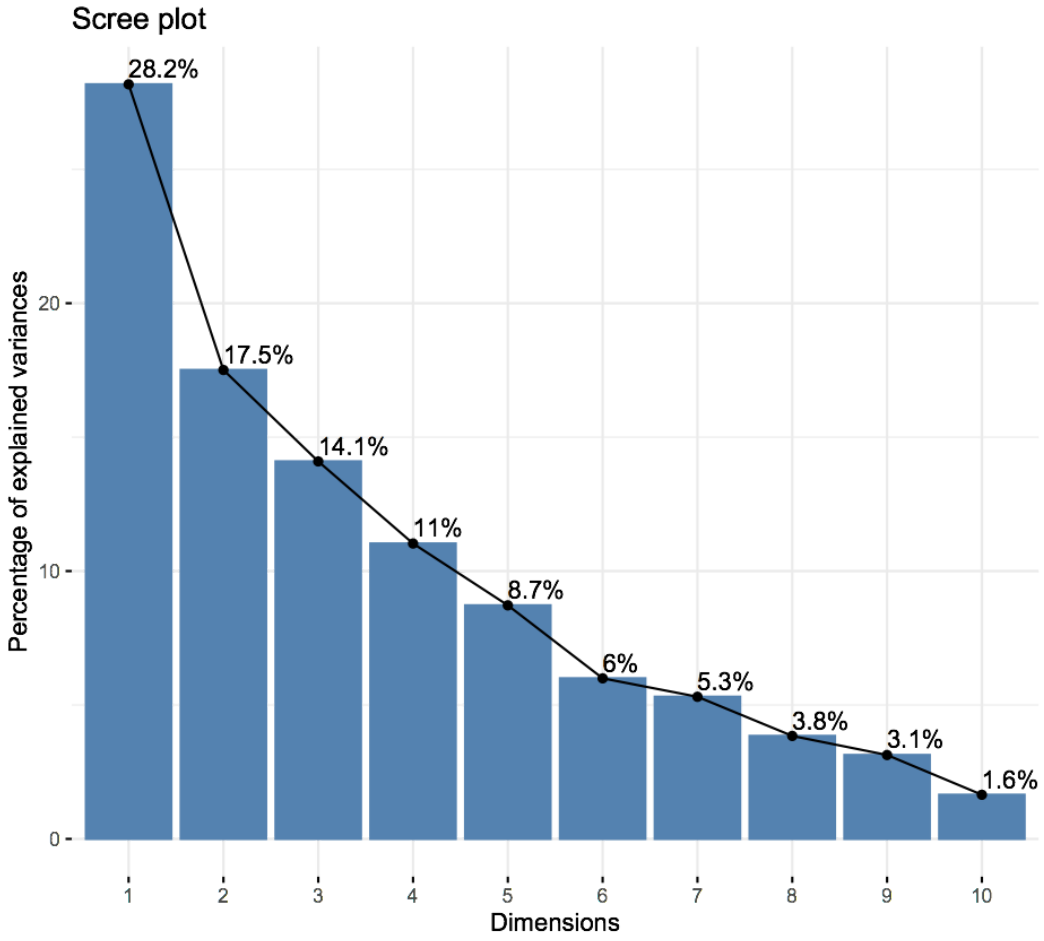


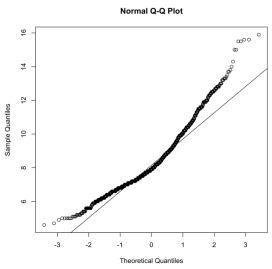
Figure 11: Scree plot of robust PCA

Lastly, analyzing PCA using the Scree plot, we can see that it is very similar to the classical PCA scree plot, as it also gives us 9 components; so, the cumulative proportion rule and the scree plot gave us a value of 9 for the number of components for the robust PCA method.

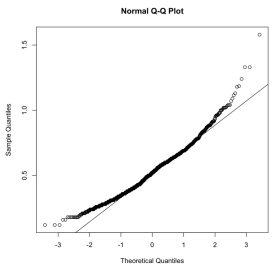
6. Conclusion

In conclusion, our analysis has shown that PCA can be a useful tool for exploring patterns and reducing the dimensionality of a dataset, but it is important to consider the assumptions and limitations of the method, as well as the potential need for robust alternatives in the presence of outliers or other deviations from normality. We conducted both the classical PCA and the robust PCA, and we found out that the first rule in both methods gave us a value of 4 components, however, when we observed the cumulative proportion and the scree plots, we found that the number of components was 9 for both methods; which tells us that our data's dimension was reduced to 9 components; it also tells us, that the outliers in our data do not affect the dimension of our data when we reduced it, as it gave us the same number of components. Lastly, we can conclude that the PCA did indeed reduce our data's dimension, but not by many variables, as our original data contained 11 variables, and PCA reduced it to 9 components; which is because the correlation between the variables is very low and almost non-existent, so every variable is important and neglecting any more variables would result in significant loss of data.

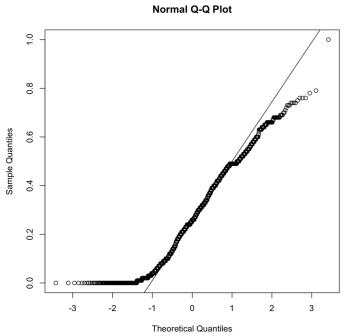
7. Appendix



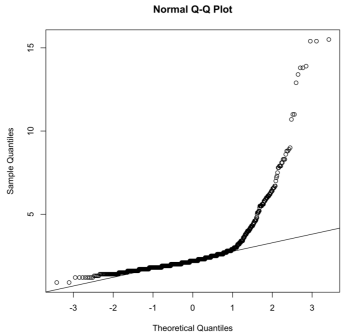
fixed.acidity



volatile.acidity

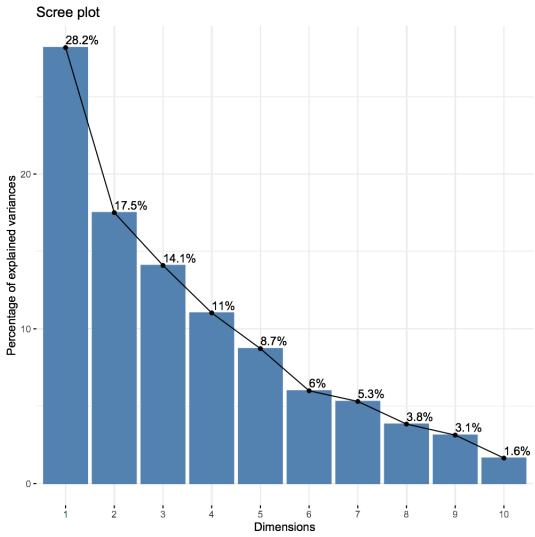


citric.acid



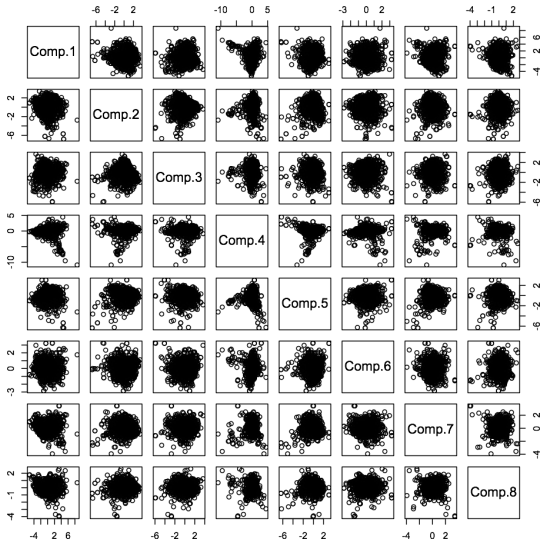
residual.sugar

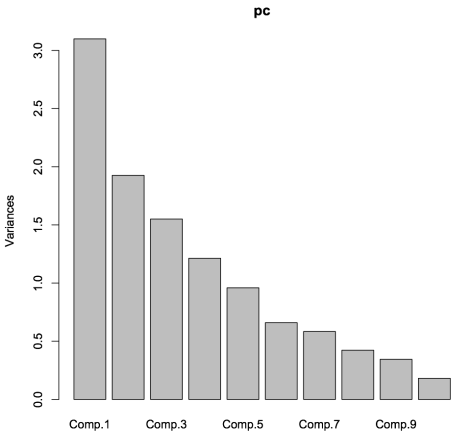
```
> # scaling the data
> z=scale(x, center = TRUE, scale = TRUE)
> # using only numeric values so we can procced with PCA
> z=z[,1:11]
> z=as.matrix(z)
> summary(z) # mean of all variables is equal to zero
 fixed.acidity     volatile.acidity    citric.acid       residual.sugar
 Min.   :-2.1364   Min.   :-2.27757   Min.   :-1.39104   Min.   :-1.1623
 1st Qu.:-0.7005   1st Qu.:-0.76969   1st Qu.:-0.92903   1st Qu.:-0.4531
 Median :-0.2410   Median :-0.04367   Median :-0.05634   Median :-0.2403
 Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
 3rd Qu.: 0.5056   3rd Qu.: 0.62649   3rd Qu.: 0.76501   3rd Qu.: 0.0434
 Max.   : 4.3538   Max.   : 5.87614   Max.   : 3.74240   Max.   : 9.1928
   chlorides        free.sulfur.dioxide total.sulfur.dioxide    density
 Min.   :-1.60344   Min.   :-1.4221     Min.   :-1.2302      Min.   :-3.53762
 1st Qu.:-0.37111   1st Qu.:-0.8485     1st Qu.:-0.7438      1st Qu.:-0.60757
 Median :-0.17989   Median :-0.1792     Median :-0.2574      Median : 0.00176
 Mean   : 0.00000   Mean   : 0.0000     Mean   : 0.0000      Mean   : 0.00000
 3rd Qu.: 0.05383   3rd Qu.: 0.4900     3rd Qu.: 0.4722      3rd Qu.: 0.57664
 Max.   :11.12355   Max.   : 5.3656     Max.   : 7.3728      Max.   : 3.67890
       pH            sulphates          alcohol
 Min.   :-3.69924   Min.   :-1.9359   Min.   :-1.8983
 1st Qu.:-0.65494   1st Qu.:-0.6380   1st Qu.:-0.8661
 Median :-0.00721   Median :-0.2251   Median :-0.2092
 Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.57574   3rd Qu.: 0.4239   3rd Qu.: 0.6353
 Max.   : 4.52687   Max.   : 7.9162   Max.   : 4.2011
> sd(z) # standard deviation of z is 0.999 approximately 1
[1] 0.9997157
```




```
> pc=princomp(z, cor=T);
> summary(pc,loadings=T)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4     Comp.5     Comp.6
Standard deviation     1.7604353 1.3877715 1.2452082 1.1014684 0.97943457 0.81216271
Proportion of Variance 0.2817393 0.1750827 0.1409585 0.1102939 0.08720837 0.05996439
Cumulative Proportion  0.2817393 0.4568220 0.5977805 0.7080744 0.79528275 0.85524714
                           Comp.7     Comp.8     Comp.9     Comp.10     Comp.11
Standard deviation     0.76406231 0.65035121 0.5870623 0.42583232 0.244045717
Proportion of Variance 0.05307193 0.03845061 0.0313311 0.01648483 0.005414392
Cumulative Proportion  0.90831906 0.94676967 0.9781008 0.99458561 1.000000000
```

```
Loadings:
                      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
fixed.acidity          0.489  0.111  0.123  0.230         0.101  0.350  0.178  0.194
volatile.acidity      -0.239 -0.275  0.450         -0.219  0.411  0.534         -0.129
citric.acid            0.464  0.152 -0.238                       -0.105  0.378 -0.381
residual.sugar         0.146 -0.272 -0.101  0.373 -0.732         -0.291 -0.300
chlorides              0.212 -0.148         -0.666 -0.247  0.304 -0.370  0.357  0.111
free.sulfur.dioxide          -0.514 -0.429         0.159         0.117  0.205  0.635
total.sulfur.dioxide         -0.569 -0.322         0.222  0.136                -0.592
density                0.395 -0.234  0.339  0.174 -0.157 -0.391  0.170  0.239
pH                    -0.439                      -0.268 -0.522         0.561 -0.168
sulphates              0.243        -0.280 -0.551 -0.226 -0.381  0.447 -0.375
alcohol               -0.113  0.386 -0.472  0.122 -0.351  0.362  0.328  0.218
                      Comp.10 Comp.11
fixed.acidity          0.250   0.640
volatile.acidity      -0.366
citric.acid           -0.622
residual.sugar                 0.184
chlorides              0.218
free.sulfur.dioxide   -0.248
total.sulfur.dioxide   0.371
density                0.240  -0.567
pH                             0.341
sulphates             -0.112
alcohol                0.303  -0.315
```

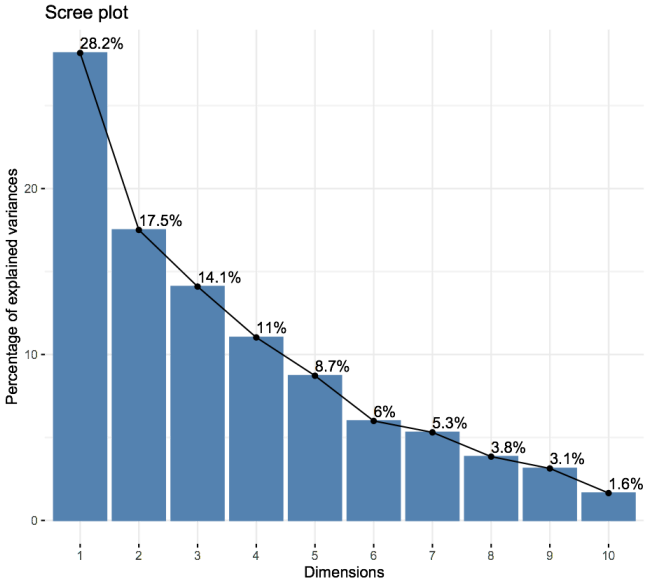


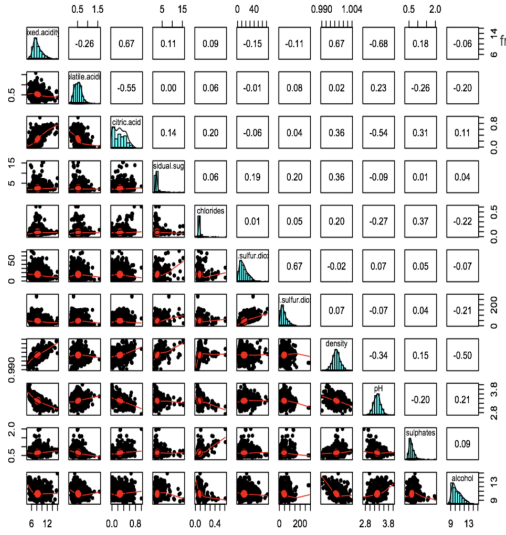


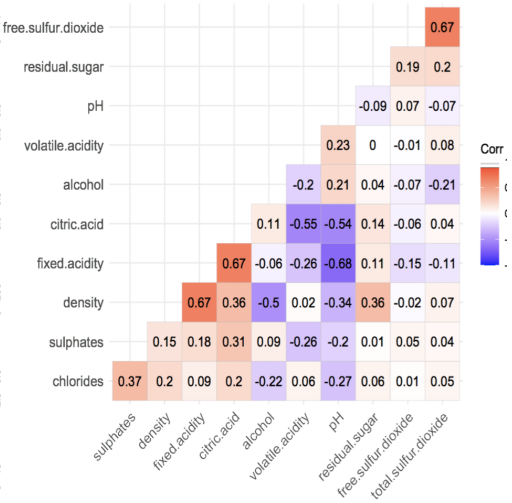

```
> b=mvBACON(z)
rank(x.ord[1:m,] >= p  ==> chosen m =  44
MV-BACON (subset no. 1): 44 of 1599 (2.75 %)
MV-BACON (subset no. 2): 1254 of 1599 (78.42 %)
MV-BACON (subset no. 3): 1397 of 1599 (87.37 %)
MV-BACON (subset no. 4): 1422 of 1599 (88.93 %)
MV-BACON (subset no. 5): 1426 of 1599 (89.18 %)
MV-BACON (subset no. 6): 1428 of 1599 (89.31 %)
MV-BACON (subset no. 7): 1429 of 1599 (89.37 %)
MV-BACON (subset no. 8): 1429 of 1599 (89.37 %)
```

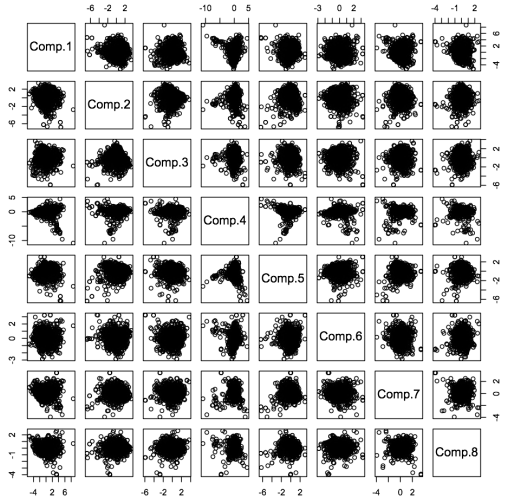
```
> summary(pcr,loadings=T)
Importance of components:
                            Comp.1    Comp.2    Comp.3    Comp.4     Comp.5     Comp.6     Comp.7
Standard deviation     1.8047324 1.4173208 1.2660380 1.0678341 0.90534090 0.82200378 0.75537066
Proportion of Variance 0.2960963 0.1826180 0.1457138 0.1036609 0.07451292 0.06142638 0.05187135
Cumulative Proportion  0.2960963 0.4787143 0.6244282 0.7280890 0.80260196 0.86402834 0.91589969
                           Comp.8     Comp.9    Comp.10     Comp.11
Standard deviation     0.60824880 0.57535613 0.41506458 0.227647783
Proportion of Variance 0.03363333 0.03009406 0.01566169 0.004711228
Cumulative Proportion  0.94953302 0.97962708 0.99528877 1.000000000
```

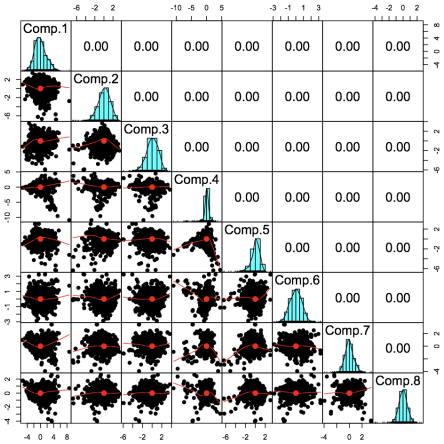
```
Loadings:
                      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
fixed.acidity          0.507                                0.225  0.227  0.282  0.177  0.342
volatile.acidity      -0.216 -0.403  0.319 -0.205         0.272  0.606  0.237 -0.158 -0.352
citric.acid            0.443  0.225 -0.188         0.185        -0.151  0.375 -0.304 -0.647
residual.sugar         0.223 -0.128        -0.742  0.201  0.208 -0.271 -0.429        -0.140
chlorides              0.244 -0.326  0.103 -0.227 -0.114 -0.856  0.106
free.sulfur.dioxide           -0.264 -0.646                                0.110  0.654 -0.221
total.sulfur.dioxide          -0.412 -0.551         0.166                       -0.621  0.330
density                0.420 -0.298  0.114        -0.313  0.291 -0.203  0.273         0.246
pH                    -0.413               -0.308 -0.352        -0.450  0.542
sulphates              0.162  0.300 -0.309 -0.163 -0.741         0.345 -0.244 -0.150
alcohol                       0.494 -0.133 -0.469  0.320         0.315  0.301         0.308
                      Comp.11
fixed.acidity          0.631
volatile.acidity
citric.acid
residual.sugar         0.148
chlorides
free.sulfur.dioxide
total.sulfur.dioxide
density               -0.591
pH                     0.327
sulphates
alcohol               -0.338
```








Comp.1
Comp.2
Comp.3
Comp.4
Comp.5
Comp.6
Comp.7
Comp.8


Comp.1
Comp.2
Comp.3
Comp.4
Comp.5
Comp.6
Comp.7
Comp.8
0.00