UNIVERSITY OF CALCUTTA

LADY BRABOURNE COLLEGE

B.SC SEMESTER 6 EXAMINATION 2021

SUBJECT : **STATISTICS**

PAPER : **DSE B - PROJECT WORK**

PROJECT TITLE : **STATISTICAL STUDY OF FACTORS AFFECTING SUSCEPTIBILITY OF COAL TO SPONTANEOUS COMBUSTION**

SUBMITTED BY:

CU ROLL NUMBER: 183031-11-0190

CU REGISTRATION NUMBER: 031-1211-0290-18

EXECUTIVE SUMMARY

Spontaneous combustion is an inherent problem to the coal mining industry. The paper describes an analysis of factors which are important for determining the susceptibility of coal to spontaneous combustion. A bar diagram is plotted for each intrinsic variable and a scatter plot is drawn to find their relation to Crossing Point Temperature and their correlation is measured. A Principal Component Analysis is conducted to summarize the intrinsic variables of coal by a smaller number of components and reduce the multicollinearity of the dataset before conducting the regression analysis of Crossing Point Temperature with the principal components as the predictor variables. A paired sample sign test has been carried out to find out if there is significant difference between the effects of different condition on the Wet Oxidation Potential Difference of coal samples. The results of Wet Oxidation Potential Difference have been correlated with Crossing Point Temperature to find the relationship between them.

INTRODUCTION

Coal is the prevailing energy supply source in India and meets 56% of country's primary commercial energy supply. Spontaneous heating and fire in coal mines is a major problem worldwide and has been a great concern both for the industry and researchers in this field. Majority of fires existing today in different coalfields are mainly due to spontaneous combustion of coal. The auto oxidation of coal ultimately leads to spontaneous combustion which is the major root cause for the disastrous of coal mine in leading and coal producing countries like USA, China, Australia, India and Germany. It is a slow process and the heat evolved is carried away by air. This process of self-heating of coal or other carbonaceous material resulting eventually in its ignition is termed as "spontaneous heating" or "auto oxidation". Spontaneous heating or spontaneous combustion of coal can lead to loss of desirable coal properties and products, creates environmental pollution, agricultural land degradation, and raise concerns about safety and economic aspects of mining especially in coal stockpiles, transportation over long distances, and in underground mining, etc. Indian mines have a historical record of extensive fire activity for over eighty years. Self-heating potential of coal is investigated by a number of methods, specifically crossing point temperature (CPT)- one of the most widely used methods in many countries of the world, wet oxidation potential difference(WOPD) method, etc. Lower is the crossing point temperature of coal, the higher is its susceptibility to spontaneous combustion. All the experimental methods for finding out the liability of coal to spontaneous combustion mainly depend on the intrinsic properties of coal. The main objective of the paper is to study the factors affecting susceptibility of coal, explain the variability in coal by reducing the intrinsic variables of coal to a smaller number of principal components and see a comparison of the different methods of measuring susceptibility of coal to spontaneous heating.

METHODOLOGY

The dataset consists of data for 78 different coal samples from India on
1.Intrinsic variables content(%) in coal
Proximate Variables-Moisture, Volatile Matter Yield , Ash Yield, and Fixed Carbon
Ultimate Variables-Carbon, Hydrogen, Nitrogen, Sulfur, Oxygen
Petrographic Variables-Vitrinite, Intertinite , Liptinite
2.Crossing Point Temperature(degree Celsius)
3.Wet Oxidation Potential Difference(WOPD)

- WOPD at PD1 –WOPD measured at $KMnO_4$=0.05N, KOH=1N, Temp=45^0 C
- WOPD at PD2 -WOPD measured at $KMnO_4$=0.1N, KOH=1N, Temp=45^0 C
- WOPD at PD3 -WOPD measured at $KMnO_4$=0.15N, KOH=1N, Temp=45^0 C

At first the data for each intrinsic variable is visualized using a bar diagram and a scatter plot is drawn to understand its relation with crossing point temperature and then their correlation coefficient is found out.

A Principal Component Analysis is done on these variable. It is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its objectives are data reduction and interpretations because it reveals relationship that were not previously detected. Much of the total system variability can be accounted by only k(say) of the principal components, k≤n, the total number of variables. So the original dataset on n variables can be reduced to a dataset on k principal components.

Algebraically, principal components are particular uncorrelated linear combinations of n random variables whose variances are as large as possible. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with the random variables as the coordinate axes. These new axes represent the directions with maximum variability.

Let the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ have the covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Sigma}$ have the eigen value-eigen vector pairs as $(\lambda_1, \boldsymbol{a_1})$, $(\lambda_2, \boldsymbol{a_2})$,..., $(\lambda_n, \boldsymbol{a_n})$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Then the ith principal component is given by:

$$Y_i = \boldsymbol{a_i}'\boldsymbol{X} = a_{i1}X_1 + \cdots + a_{in}X_n \ , \qquad i = 1,2, \ldots, n$$

$$Var(Y_i) = \boldsymbol{a_i}'\boldsymbol{\Sigma}\boldsymbol{a_i} = \lambda_i \ , \qquad i = 1,2, . . , n$$

$$Cov(Y_i, Y_j) = \boldsymbol{a_i}'\boldsymbol{\Sigma}\boldsymbol{a_j} = 0\,, \qquad i \neq j$$

A paired sample sign test is a nonparametric test for location parameter(median) θ of a population. Here, no assumption of the normality of parent population is made .Suppose a random sample of n pairs (x_i,y_i) , i=1,2,...,n. Let $d_i=x_i-y_i$, i=1,2,...,n. It is assumed that the distribution of differences is continuous in the vicinity of the median θ.

The null hypothesis is $H_o : \theta = 0$ and the alternative is $H_1 : \theta \neq 0$.

Let r be the number of plus signs i.e., number of pairs where $d_i > 0$ and s be the number of minus signs i.e., number of pairs where $d_i < 0$. The critical region is given by:

$r \geq r_{\alpha/2}$ and $r \leq r'_{\alpha/2}$ where $r_{\alpha/2}$ and $r'_{\alpha/2}$ are such that,

$\sum_{r_{\alpha/2}}^{n} \binom{n}{r} \left(\frac{1}{2}\right)^n \leq \frac{\alpha}{2}$ and $\sum_{0}^{r'_{\alpha/2}} \binom{n}{r} \left(\frac{1}{2}\right)^n \leq \frac{\alpha}{2}$

Under H_0, $r \sim Binomial\ (n, \frac{1}{2})$

$\therefore E(r) = \frac{n}{2} \quad , \quad Var(r) = \frac{n}{4}$

For large n(>25), normal approximation may be used.

$$Z = \frac{r - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,1)$$

and the test can be conducted by τ_α values.

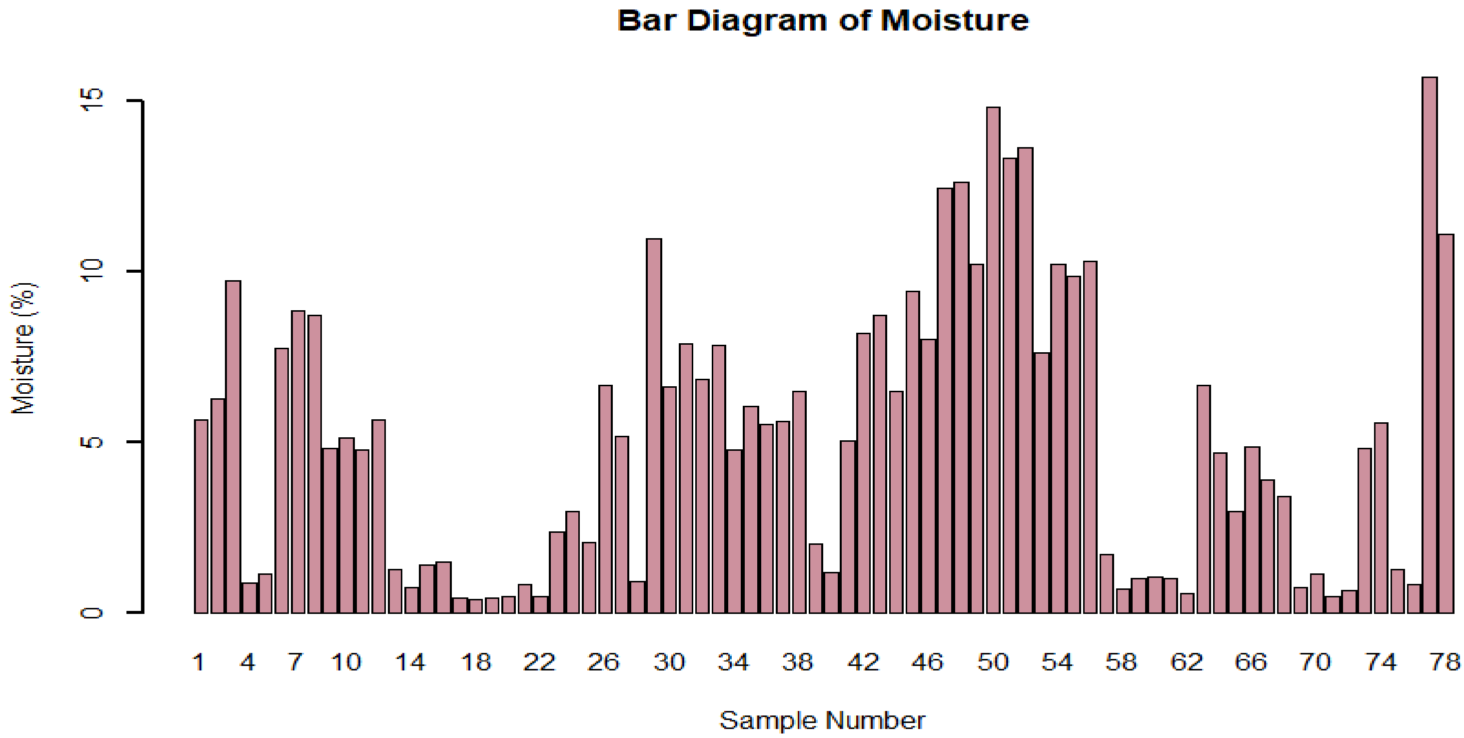
❖ RStudio Version 1.4.1717 is used to carry out the entire analysis.

RESULTS

Moisture:

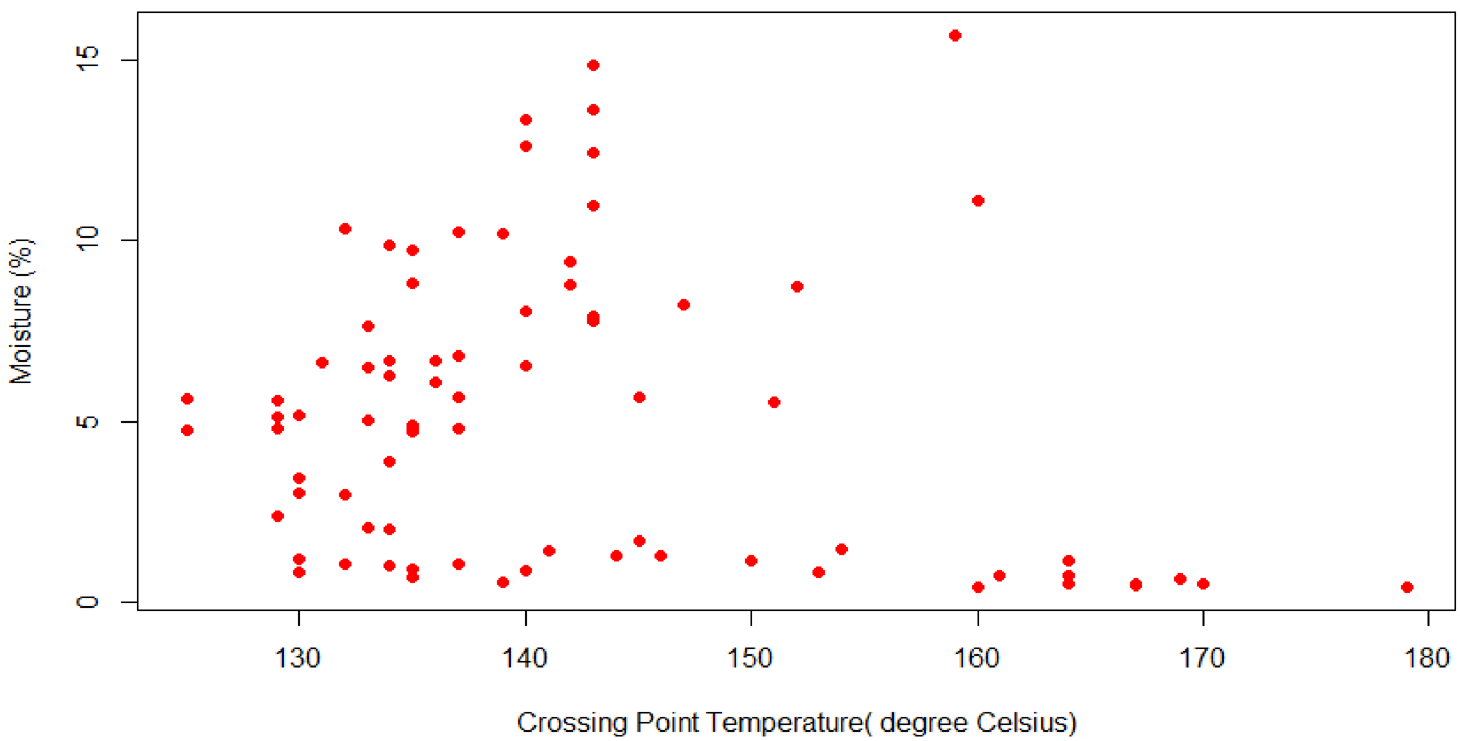
The data is visualized with a bar diagram:

>barplot(a.1,names=s,xlab="Sample Number",ylab="Moisture (%)",main="Bar Diagram of Moisture",col="pink3",lwd=2)



The scatter plot between moisture and crossing point temperature is:

> plot(a.1$cpt,a.1$moisture,xlab="Crossing Point Temperature(degree Celsius)",ylab="Moisture (%)",col="red",pch=19,cex=1)



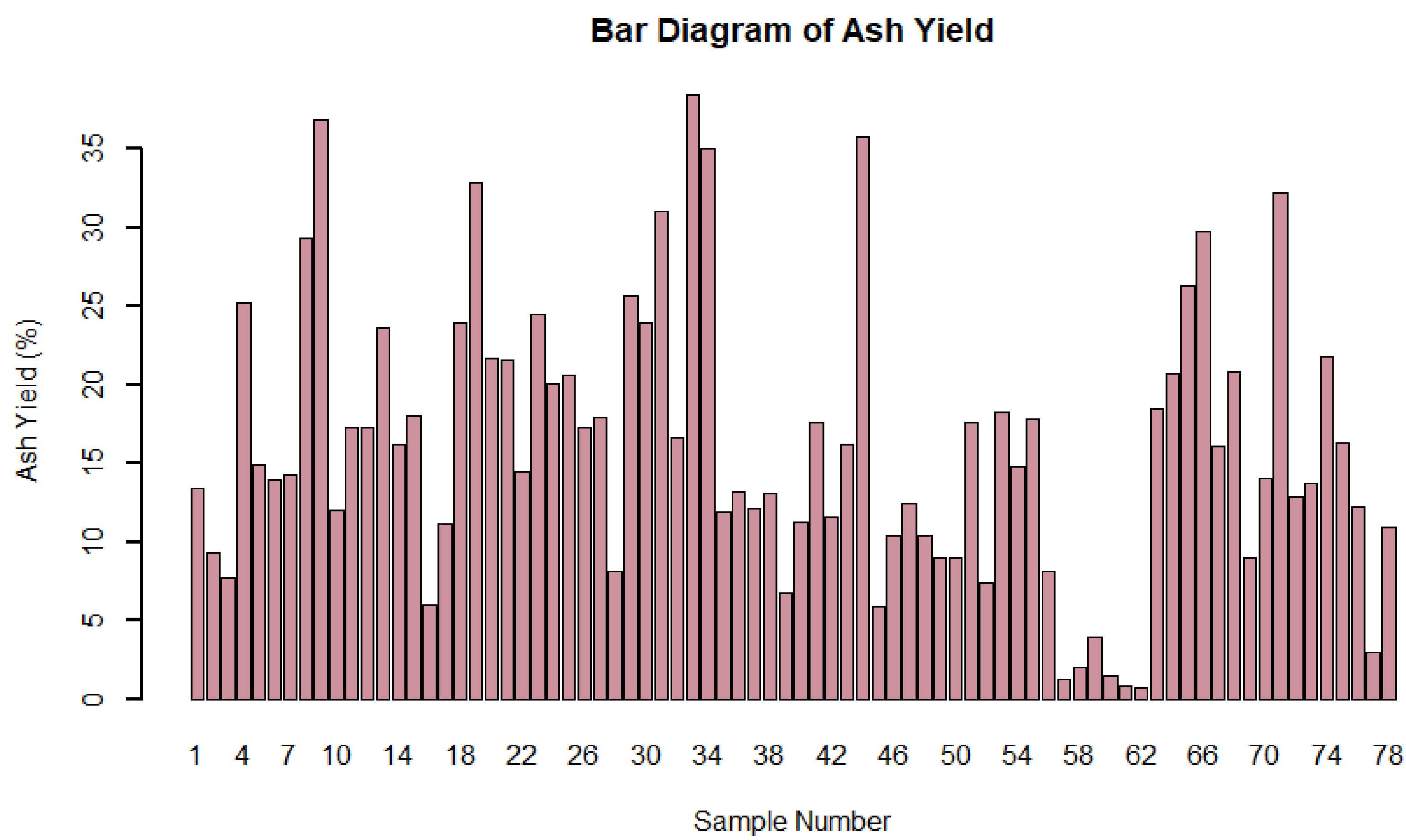
The correlation coefficient between moisture and crossing point temperature is found to be -0.202

```
> cor(a.1,method="pearson")
            moisture        cpt
moisture     1.000         -0.202
cpt         -0.202          1.00
```

This shows that moisture and crossing point temperature have weak negative correlation.

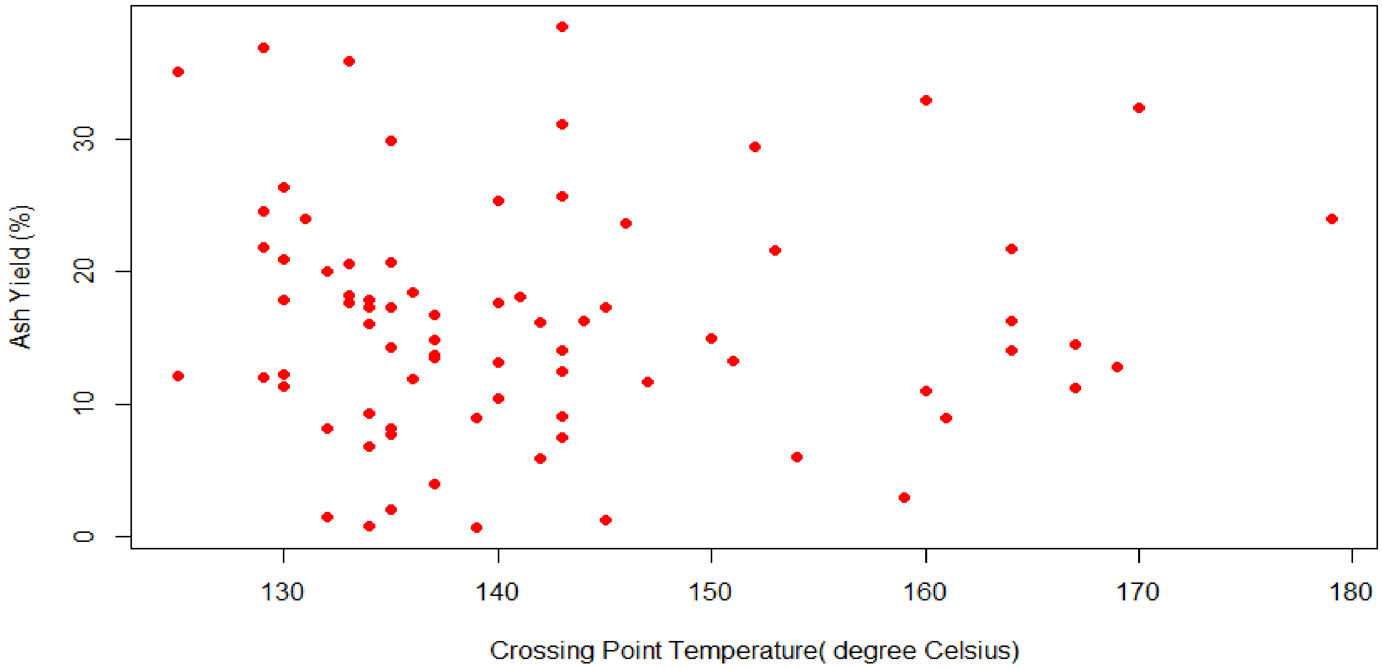
<u>Ash Yield:</u>

The data is visualized with a bar diagram:



The scatter plot between ash yield and crossing point temperature is:

```
> plot(a.1$cpt,a.2$ash.yld,xlab="Crossing Point Temperature( degree Celsius)",ylab="Ash Yield (%)",col="red",pch=19,cex=1)
```



The correlation coefficient between ash yield and crossing point temperature is found to be 0.005, which is almost equal to 0.

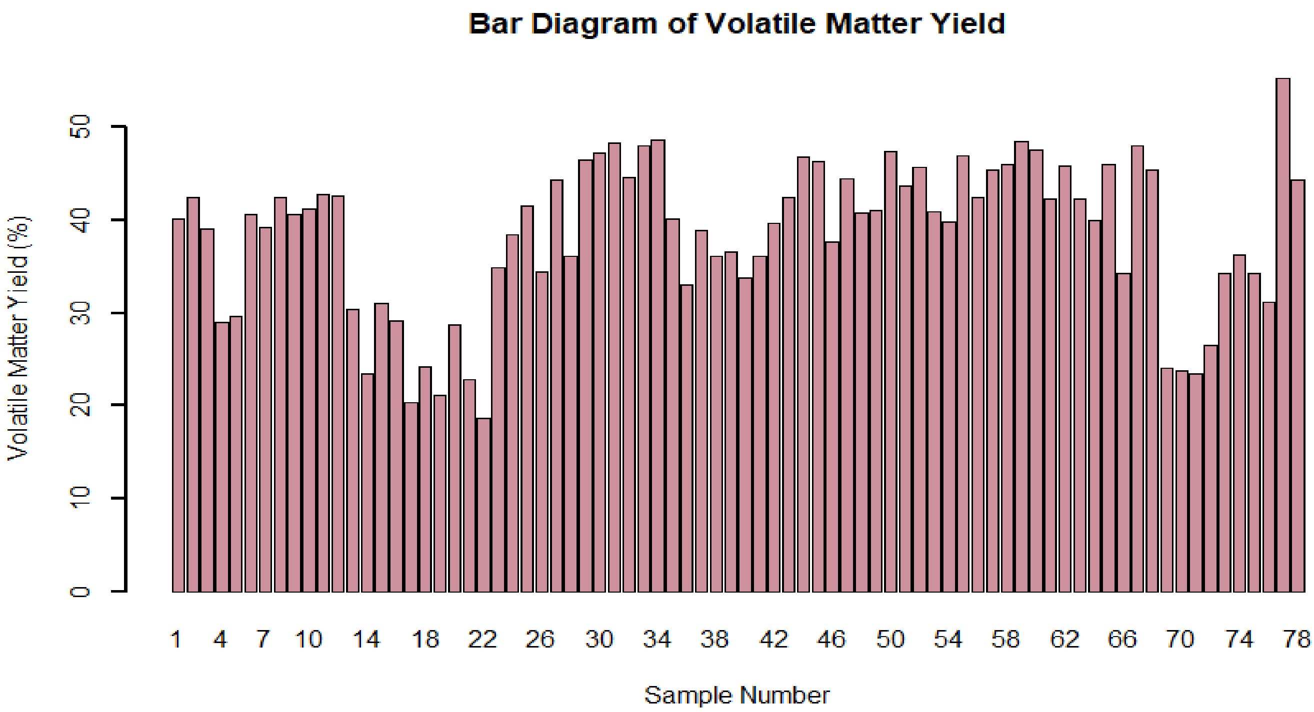
```
> cor(a.2,method="pearson")
          ash.yld     cpt
ash.yld    1.000     0.005
cpt        0.005     1.000
```

This shows that ash yield and crossing point temperature have negligible correlation.

<u>Volatile Matter Yield:</u>

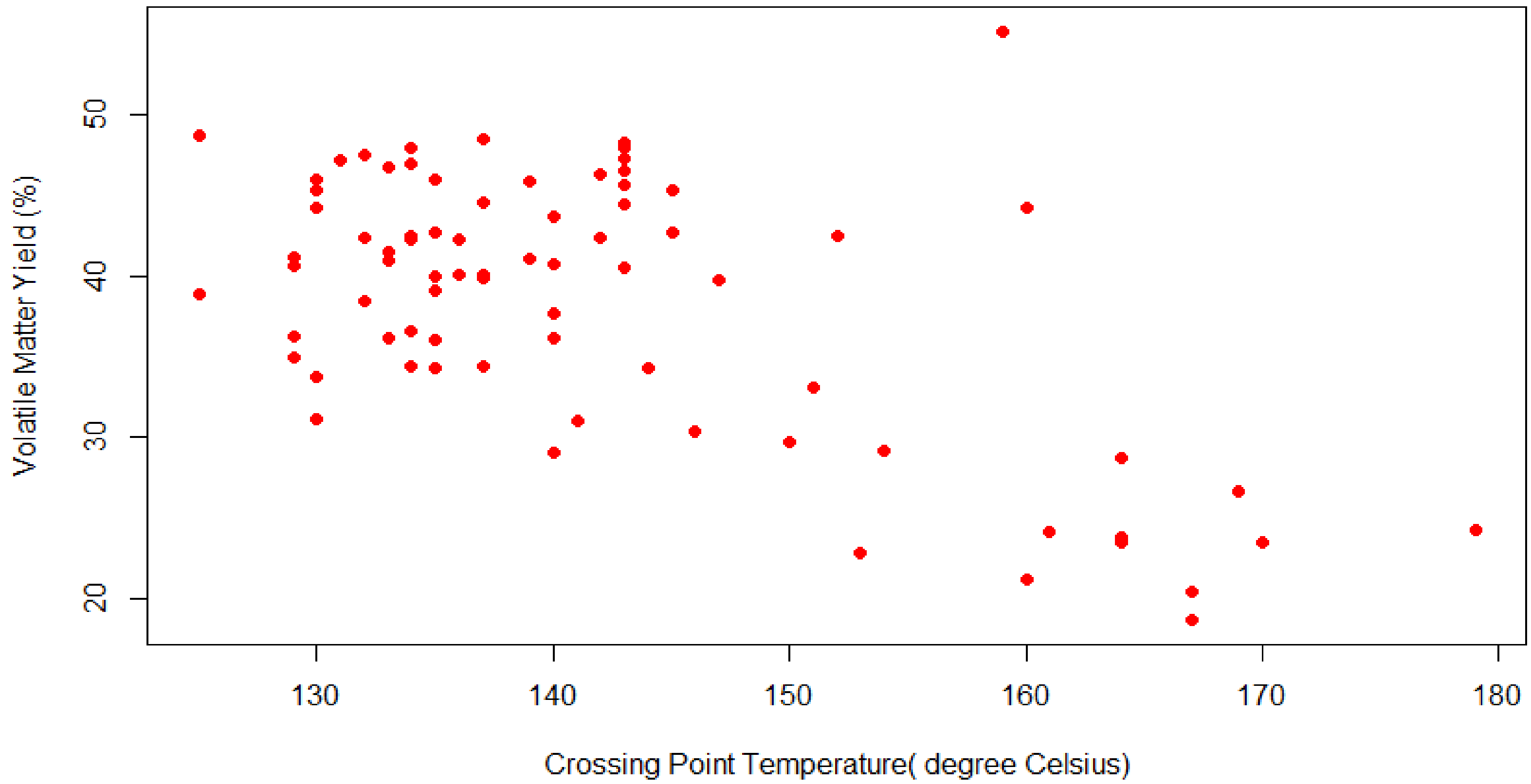
The data is visualized with a bar diagram:

```
> barplot(a.3,names=s,xlab="Sample Number",ylab="Volatile Matter Yield (%)",main="Bar Diagram of Volatile Matter Yield",col="pink3",lwd=2)
```



The scatter plot between volatile matter yield and crossing point temperature is :

> plot(a.3$cpt,a.3$vm.yld,xlab="Crossing Point Temperature(degree Celsius)",ylab="Volatile Matter Yield (%)",col="red",pch=19,cex=1)



The correlation coefficient between volatile matter yield and crossing point temperature is found to be -0.6007487.

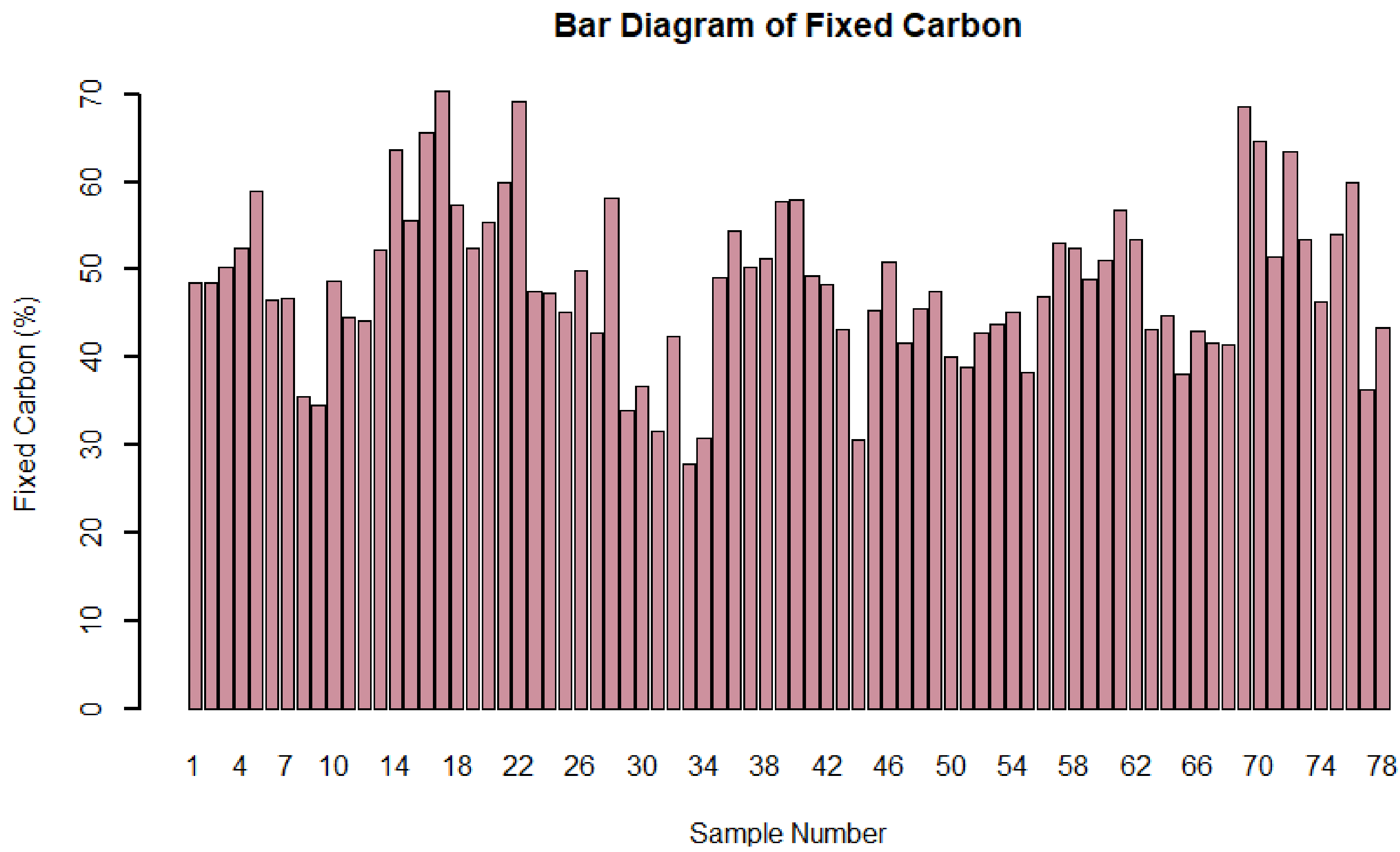
```
>cor(a.3,method="pearson")
              vm.yld          cpt
vm.yld     1.0000000   -0.6007487
cpt       -0.6007487    1.0000000
```

This shows that volatile matter yield and crossing point temperature have strong negative correlation.

<u>Fixed Carbon:</u>

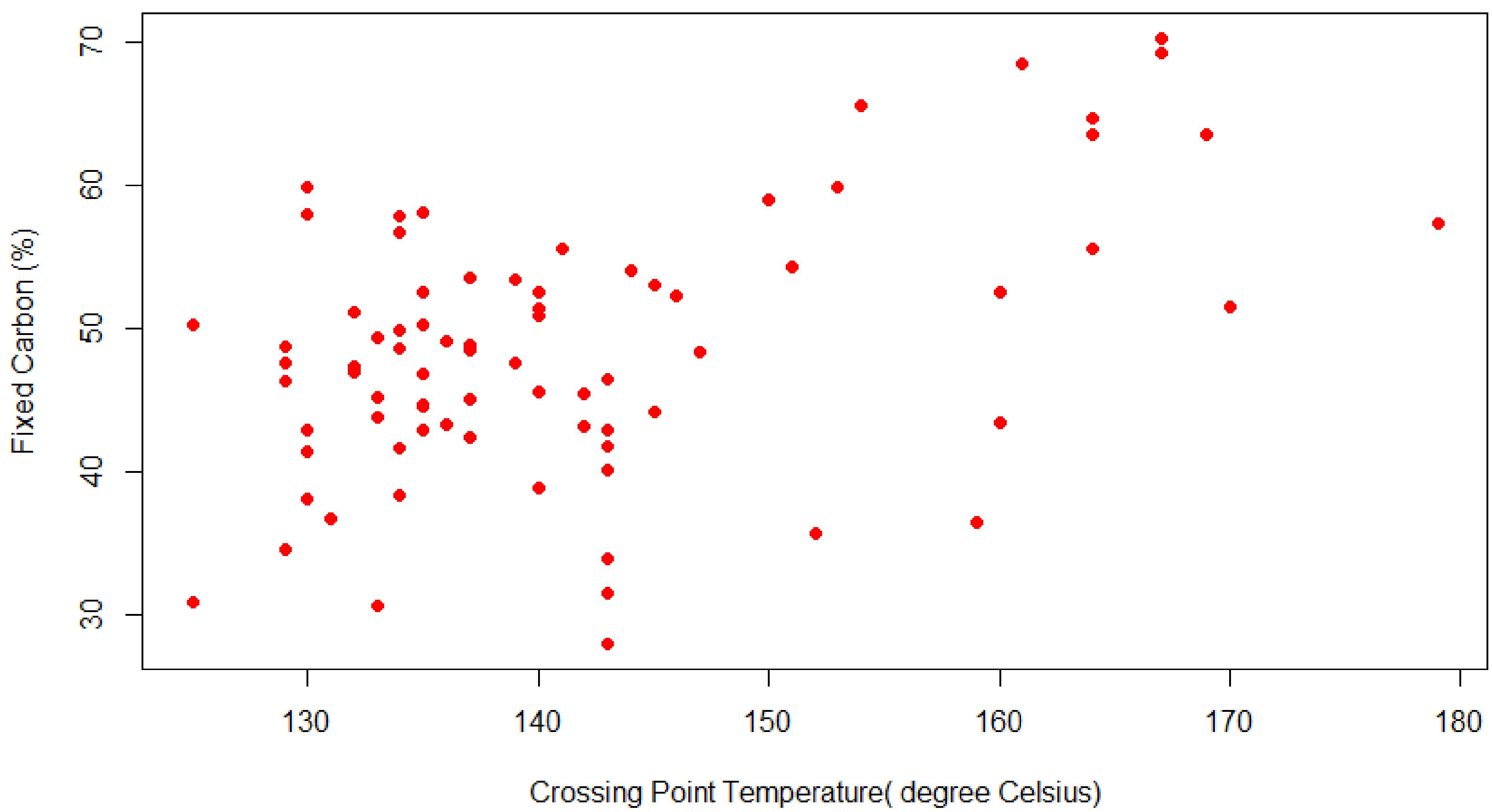
The data is visualized with a bar diagram:

> barplot(a.4,names=s,xlab="Sample Number",ylab="Fixed Carbon (%)",main="Bar Diagram of Fixed Carbon",col="pink3",lwd=2



The scatter plot between fixed carbon and crossing point temperature is

> plot(a.4$cpt,a.4$fxd.crbn,xlab="Crossing Point Temperature(degree Celsius)",ylab="Fixed Carbon (%)",col="red",pch=19,cex=1)



The correlation coefficient between fixed carbon and crossing point temperature is found to be 0.4767388.

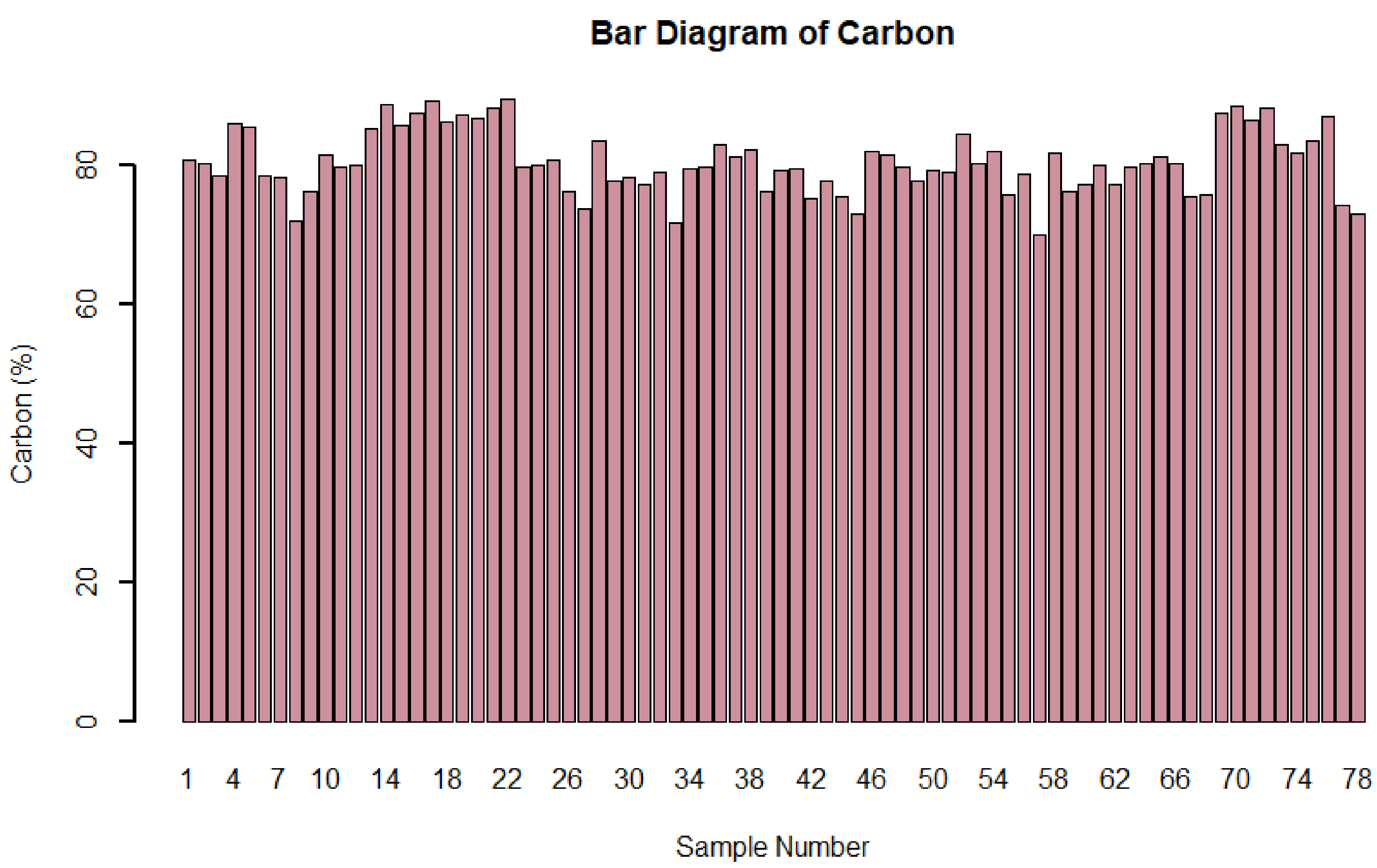
```
> cor(a.4,method="pearson")
          fxd.crbn       cpt
fxd.crbn  1.0000000  0.4767388
cpt       0.4767388  1.0000000
```

This shows that fixed carbon and crossing point temperature have moderate positive correlation.

Carbon:

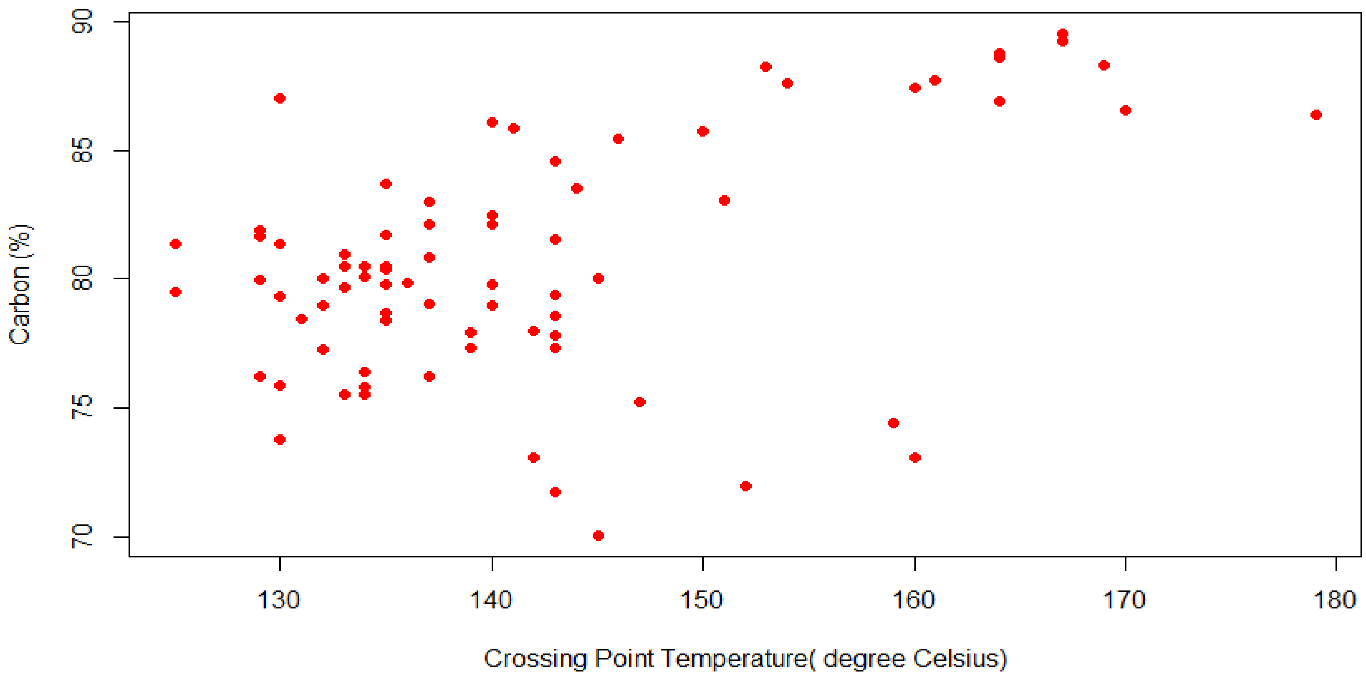
The data is visualized with a bar diagram:

```
> barplot(a.5,names=s,xlab="Sample Number",ylab="Carbon (%)",main="Bar Diagram of Carbon "
, col="pink3",lwd=2)
```



The scatter plot between carbon and crossing point temperature is

```
> plot(a.5$cpt,a.5$carbon,xlab="Crossing Point Temperature( degree Celsius)", ylab="Carbon
(%)",col="red",pch=19,cex=1)
```



The Correlation coefficient between carbon and crossing point temperature is found to be 0.4838925

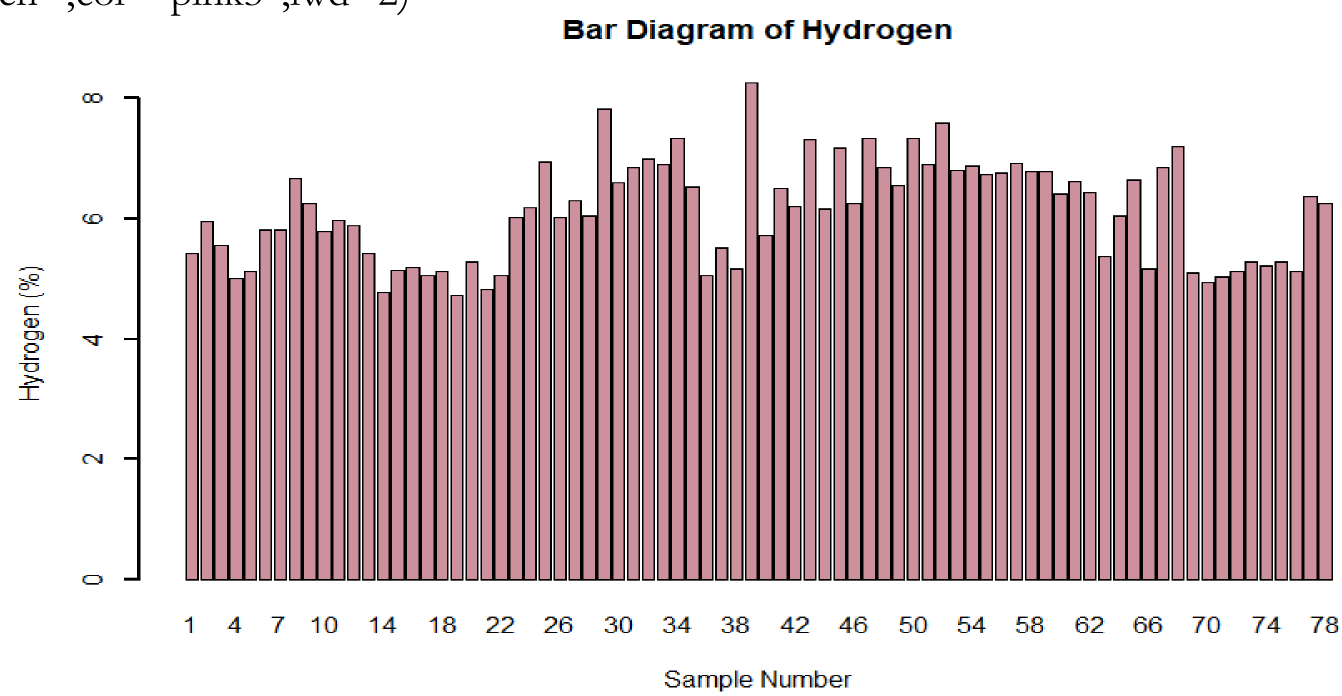
```
> cor(a.5,method="pearson")
           carbon       cpt
carbon  1.0000000  0.4838925
cpt     0.4838925  1.0000000
```

This shows that carbon and crossing point temperature have moderate positive correlation.

<u>Hydrogen:</u>

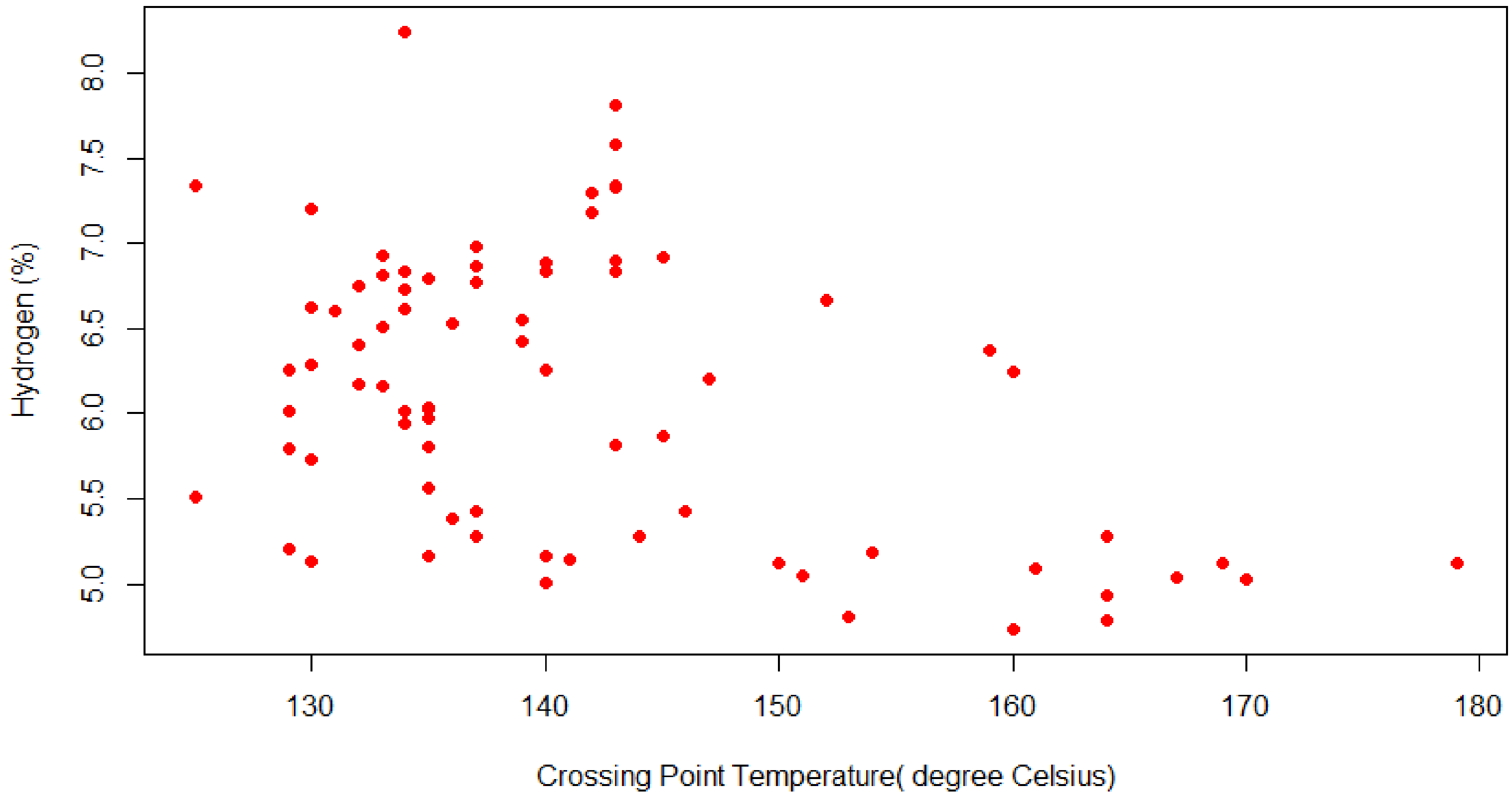
The data is visualized with a bar diagram:

```
> barplot(a.6,names=s,xlab="Sample Number",ylab="Hydrogen (%)",main="Bar Diagram of Hydrogen" ,col="pink3",lwd=2)
```



The scatter plot between hydrogen and crossing point temperature is

```
> plot(a.6$cpt,a.6$hydrogen,xlab="Crossing Point Temperature( degree Celsius)",ylab="Hydrogen (%)",col="red",pch=19,cex=1)
```



The correlation coefficient between hydrogen and crossing point temperature is found to be -0.4466731,

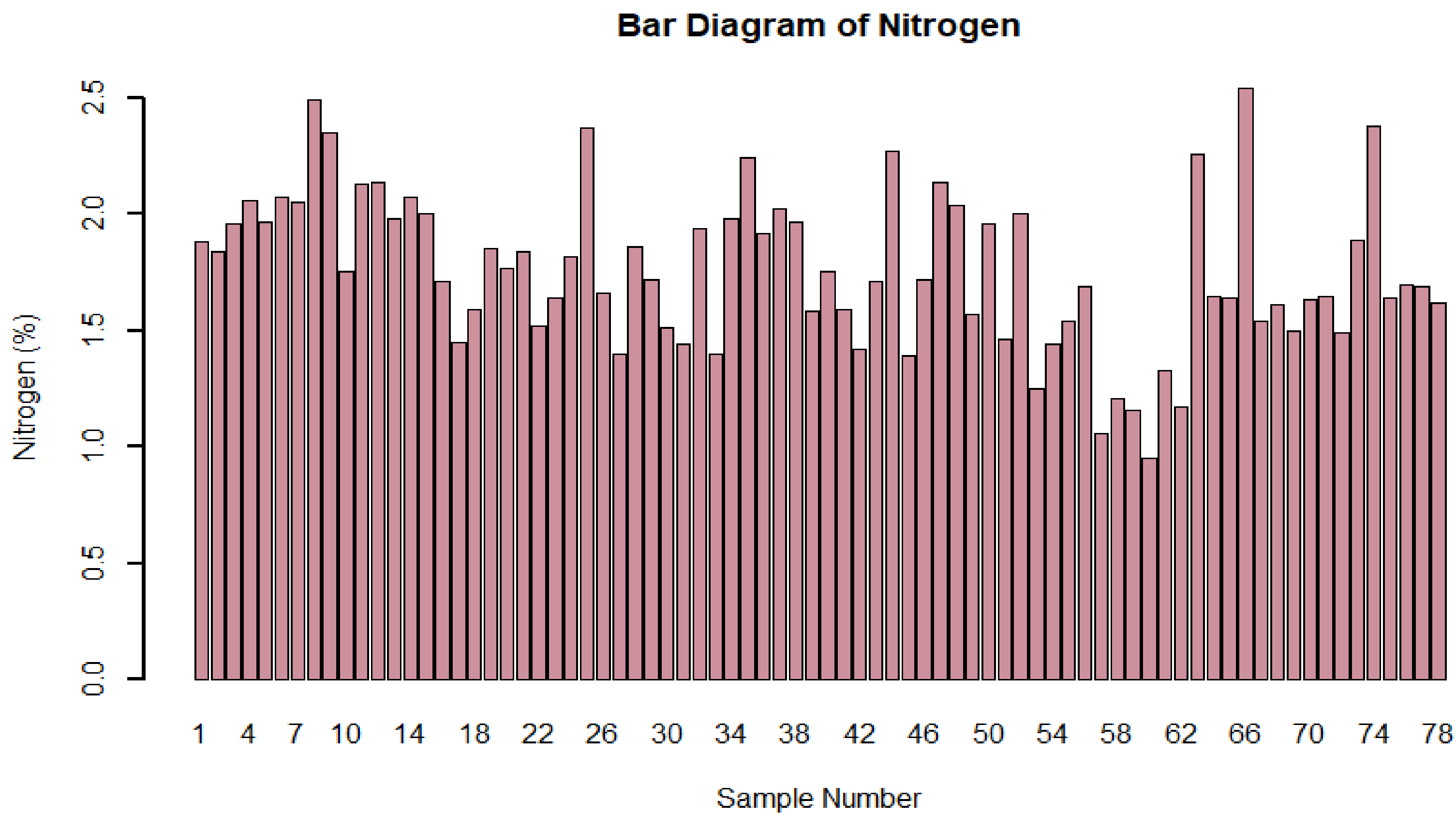
```
> cor(a.6,method="pearson")
           hydrogen        cpt
hydrogen  1.0000000 -0.4466731
cpt      -0.4466731  1.0000000
```

This shows that hydrogen and crossing point temperature have moderate negative correlation.

<u>Nitrogen:</u>

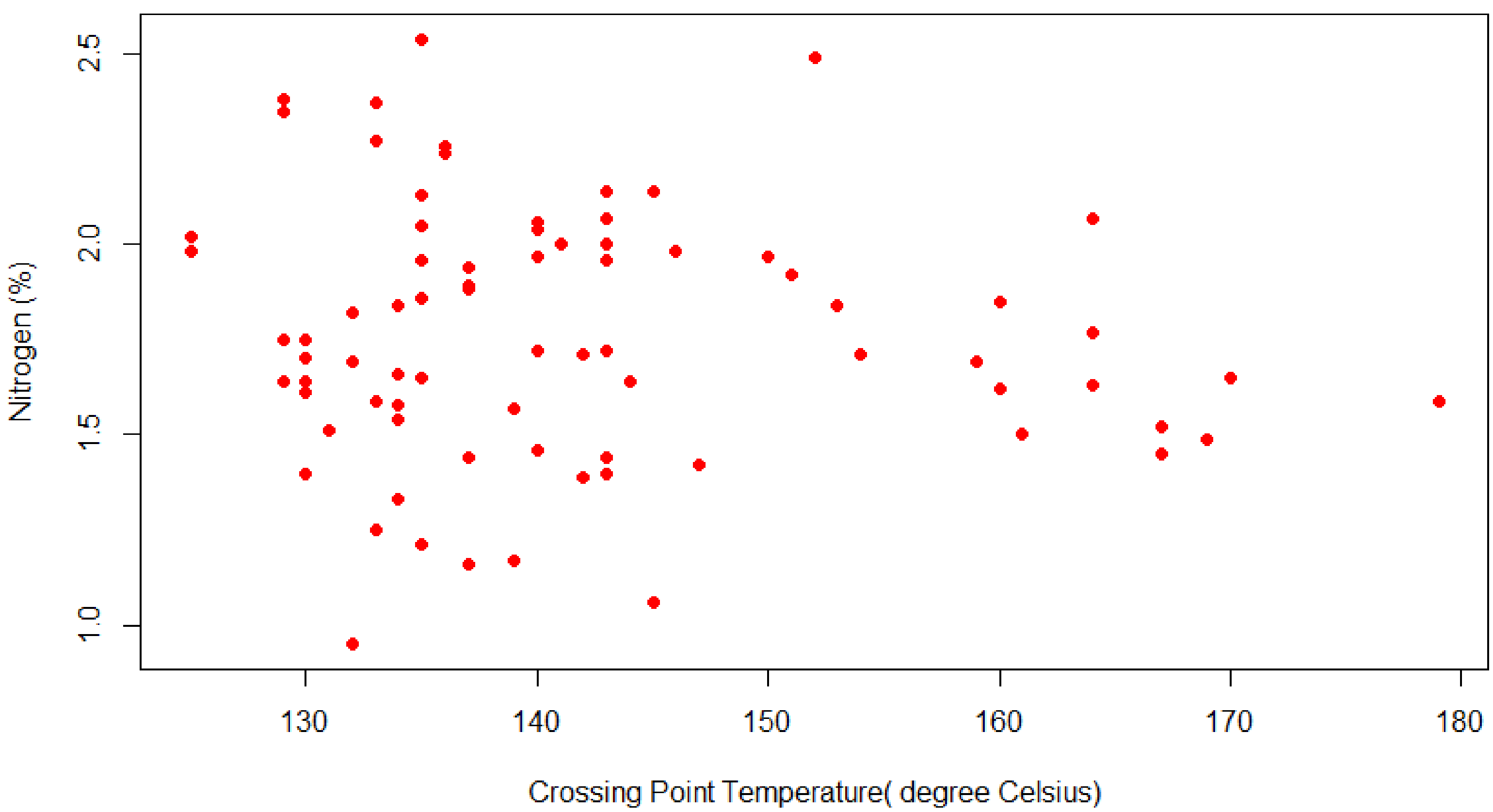
The data is visualized with a bar diagram:

```
> barplot(a.7,names=s,xlab="Sample Number",ylab="Nitrogen (%)",main="Bar Diagram of Nitrogen",col="pink3",lwd=2)
```



The scatter plot between nitrogen and crossing point temperature is

> plot(a.7$cpt,a.7$nitrogen,xlab="Crossing Point Temperature(degree Celsius)",ylab="Nitrogen (%)",col="red",pch=19,cex=1)



The correlation coefficient between nitrogen and crossing point temperature is found to be -0.1042976,

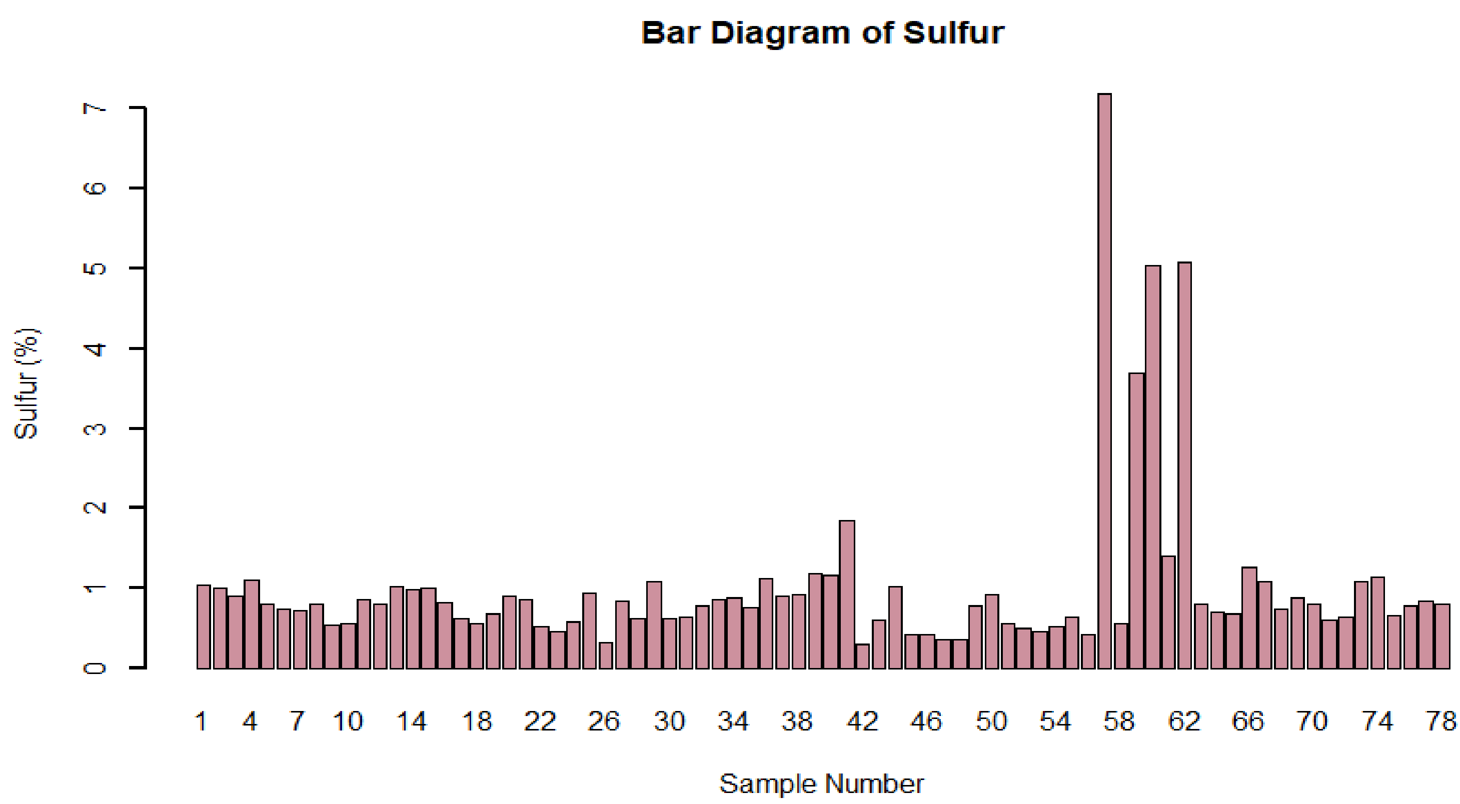
```
> cor(a.7,method="pearson")
            nitrogen        cpt
nitrogen   1.0000000  -0.1042976
cpt       -0.1042976   1.0000000
```

This shows that nitrogen and crossing point temperature have weak negative correlation.

<u>Sulfur:</u>

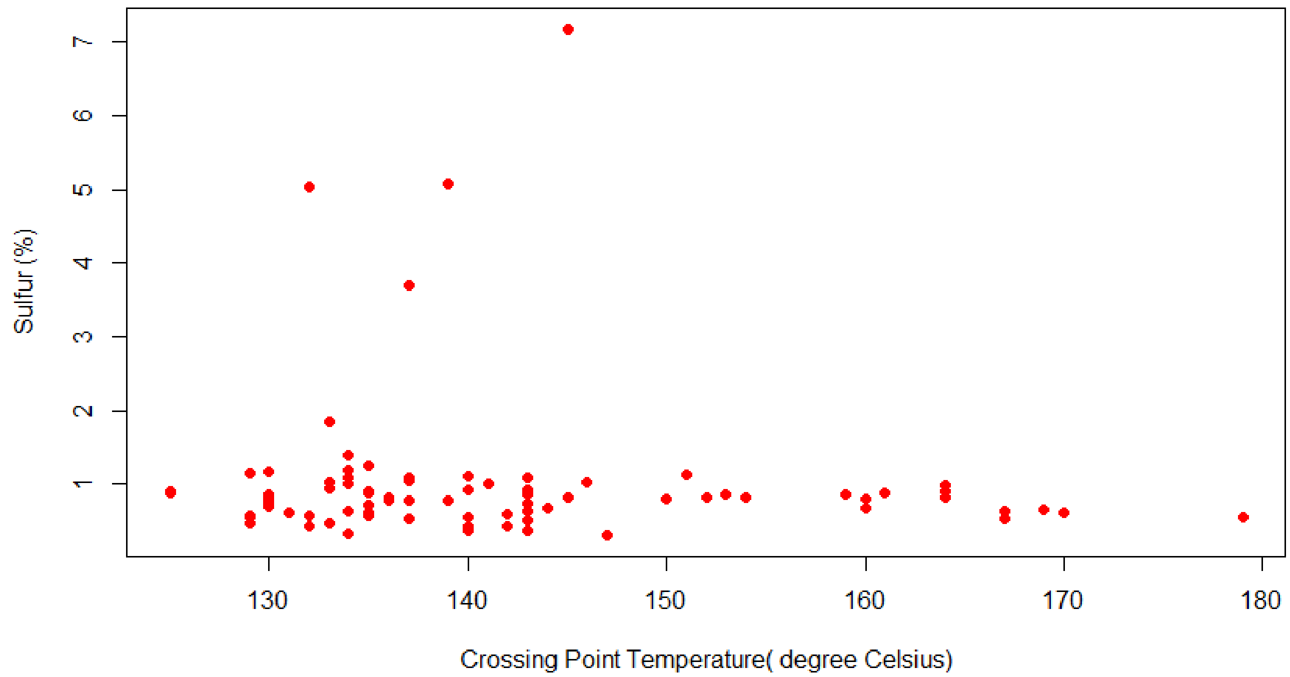
The data is visualized with a bar diagram:

```
> barplot(a.8,names=s,xlab="Sample Number",ylab="Sulfur (%)",main="Bar Diagram of Sulfur", col="pink3",lwd=2)
```



The scatter plot between sulfur and crossing point temperature is

```
> plot(a.8$cpt,a.8$sulfur,xlab="Crossing Point Temperature( degree Celsius)",ylab="Sulfur (%)", col="red",pch=19,cex=1)
```



The correlation coefficient between sulfur and crossing point temperature is found to be -0.07689693,

> cor(a.8,method="pearson")

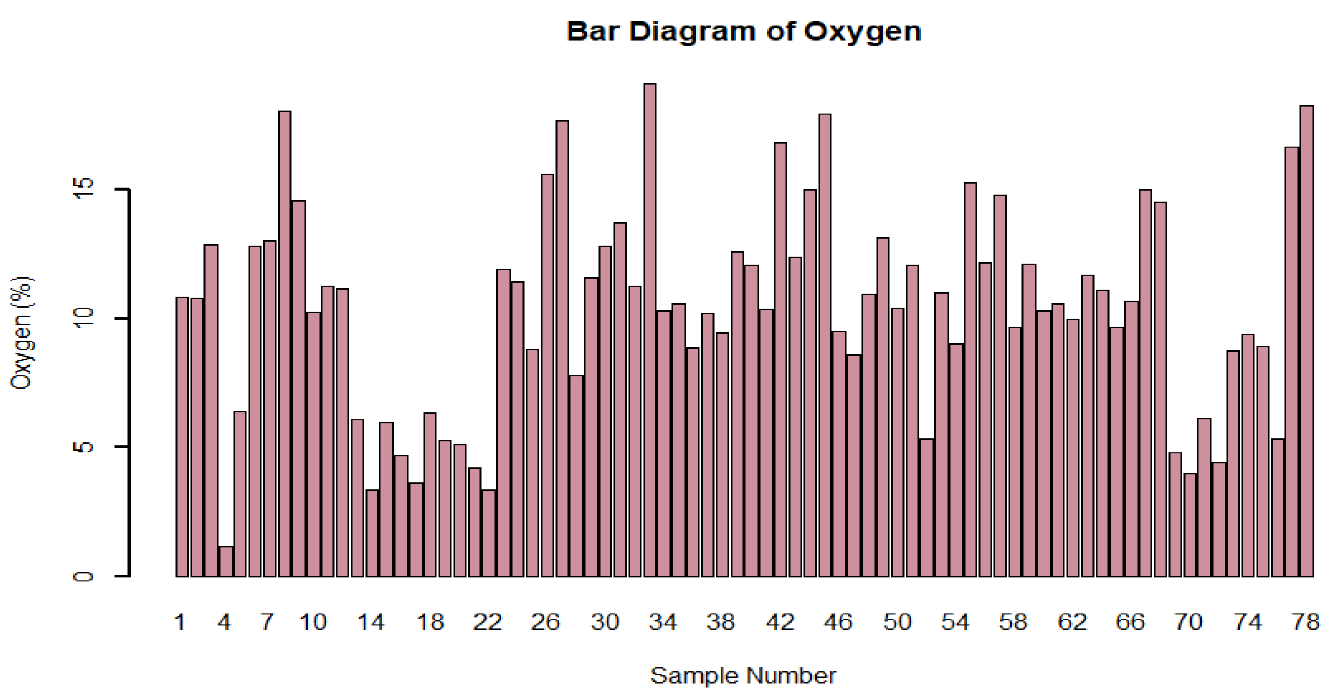
	sulfur	cpt
sulfur	1.00000000	-0.07689693
cpt	-0.07689693	1.00000000

This shows that sulfur and crossing point temperature have negligible correlation.

<u>Oxygen:</u>

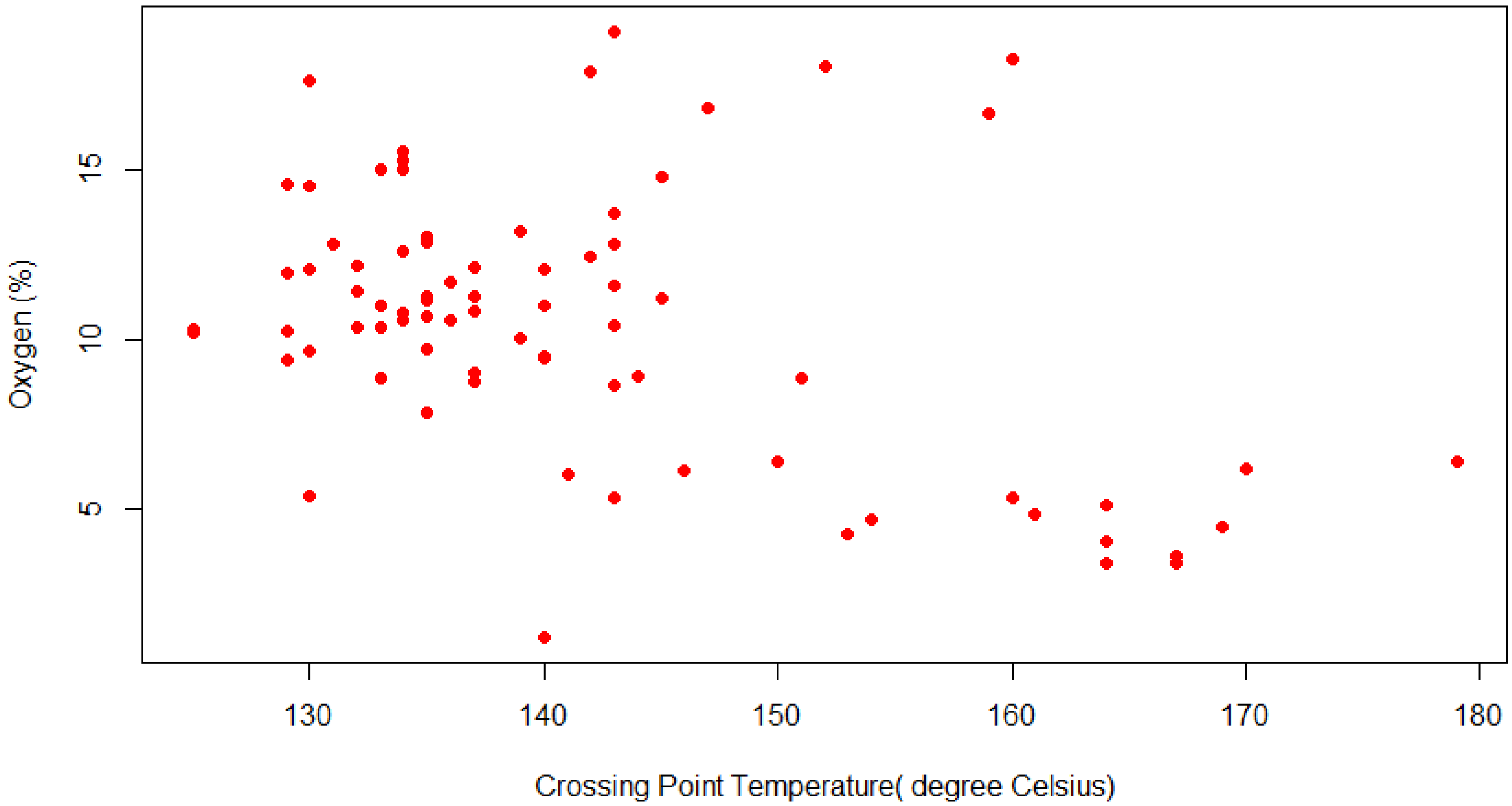
The data is visualized with a bar diagram:

> barplot(a.9,names=s,xlab="Sample Number",ylab="Oxygen (%)",main="Barplot of Oxygen", col="pink3",lwd=2)



The scatter plot between oxygen and crossing point temperature is

> plot(a.9$cpt,a.9$oxygen,xlab="Crossing Point Temperature(degree Celsius)",ylab="Oxygen (%)",col="red",pch=19,cex=1)



The correlation coefficient between oxygen and crossing point temperature is found to be -0.4287,

> cor(a.9,method="pearson")

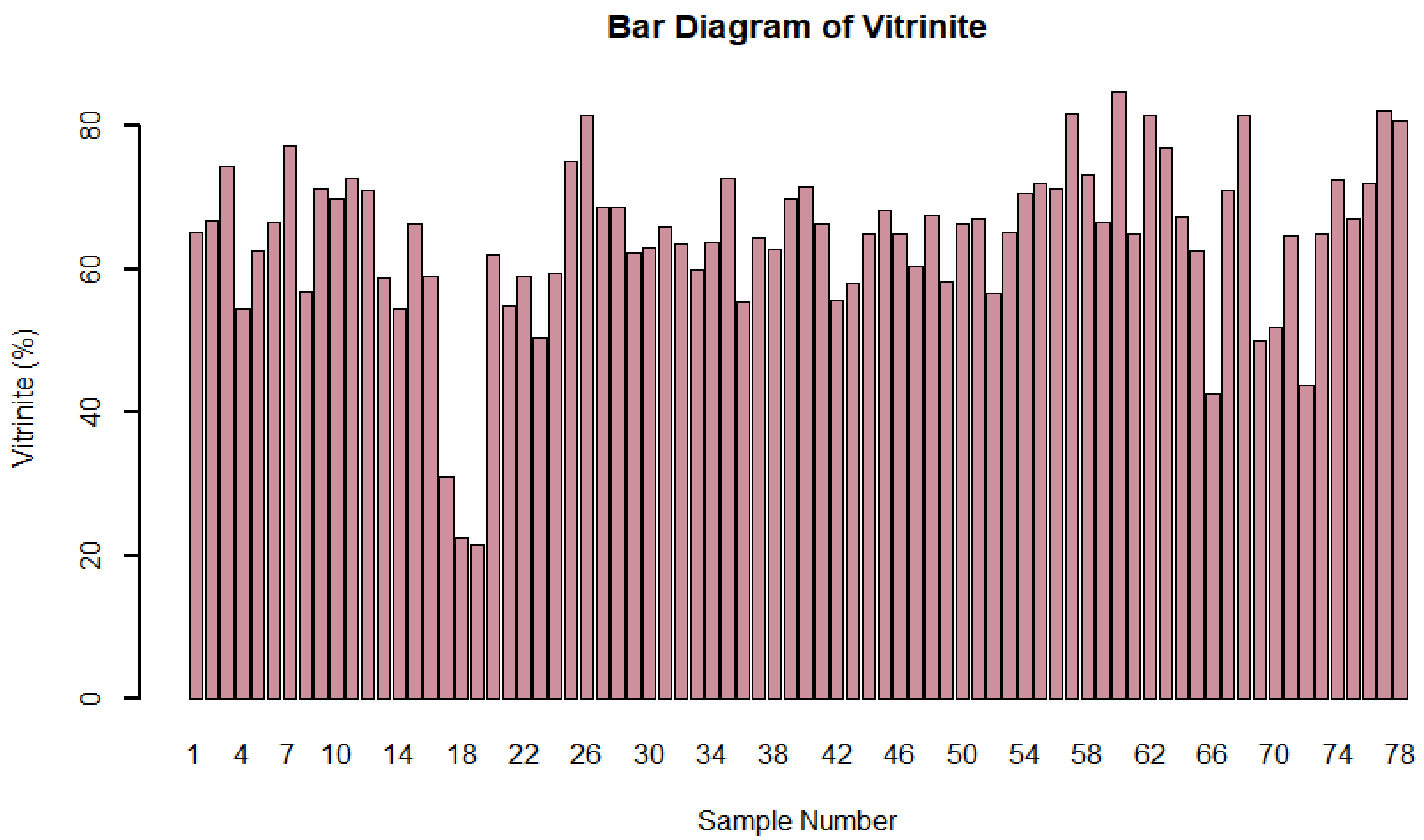
	oxygen	cpt
oxygen	1.0000	-0.4287
cpt	-0.4287	1.0000

This shows that oxygen and crossing point temperature have moderate negative correlation.

<u>Vitrinite:</u>

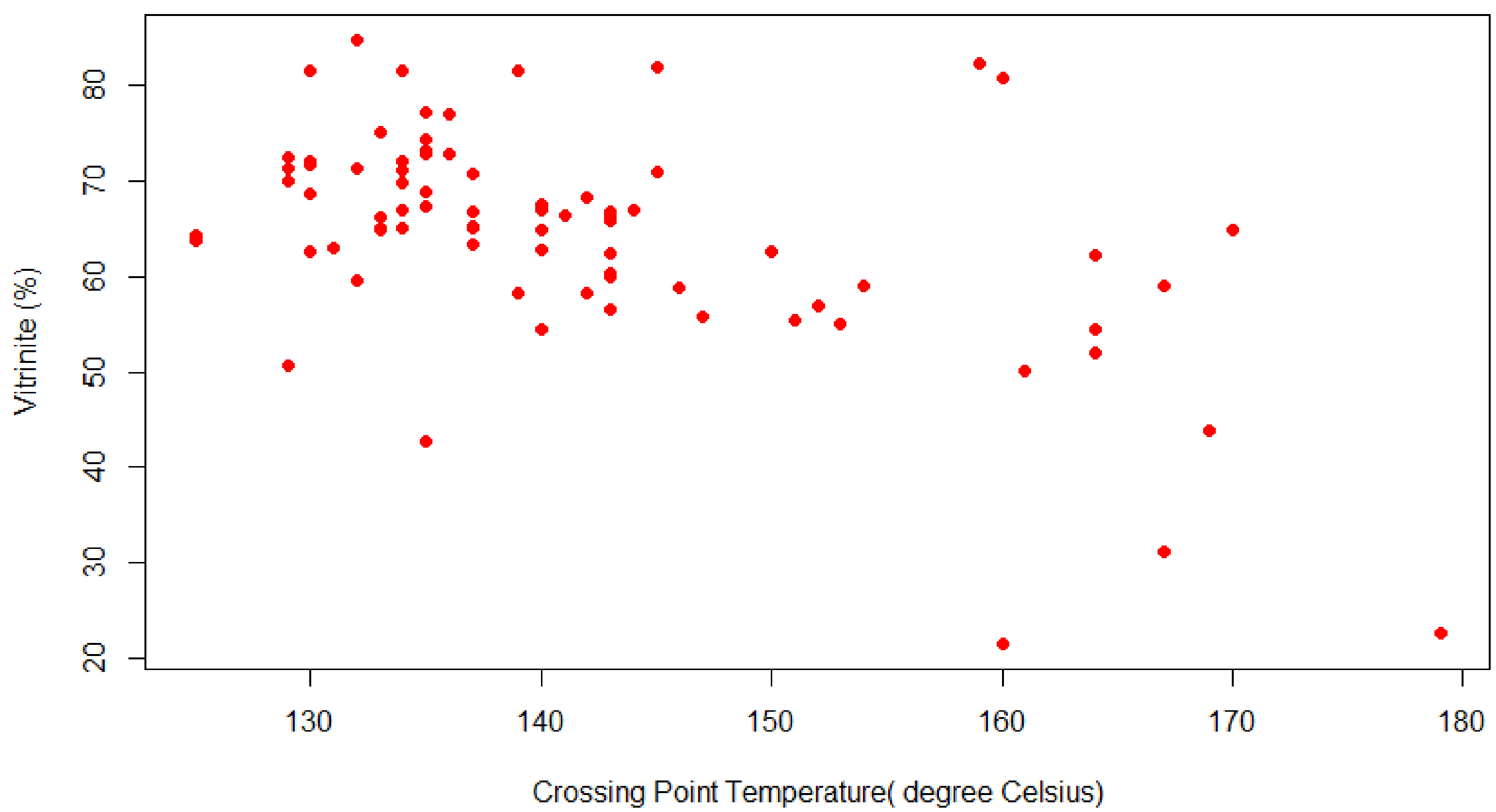
The data is visualized with a bar diagram:

> barplot(a.10,names=s,xlab="Sample Number",ylab="Vitrinite (%)",main="Bar Diagram of Vitrinite", col="pink3",lwd=2)



The scatter plot between vitrinite and crossing point temperature is

> plot(a.10$cpt,a.10$vitrinite,xlab="Crossing Point Temperature(degree Celsius)",ylab="Vitrinite (%)",col="red",pch=19,cex=1)



The correlation coefficient between vitrinite and crossing point temperature is found to be -0.5454505,

> cor(a.10,method="pearson")

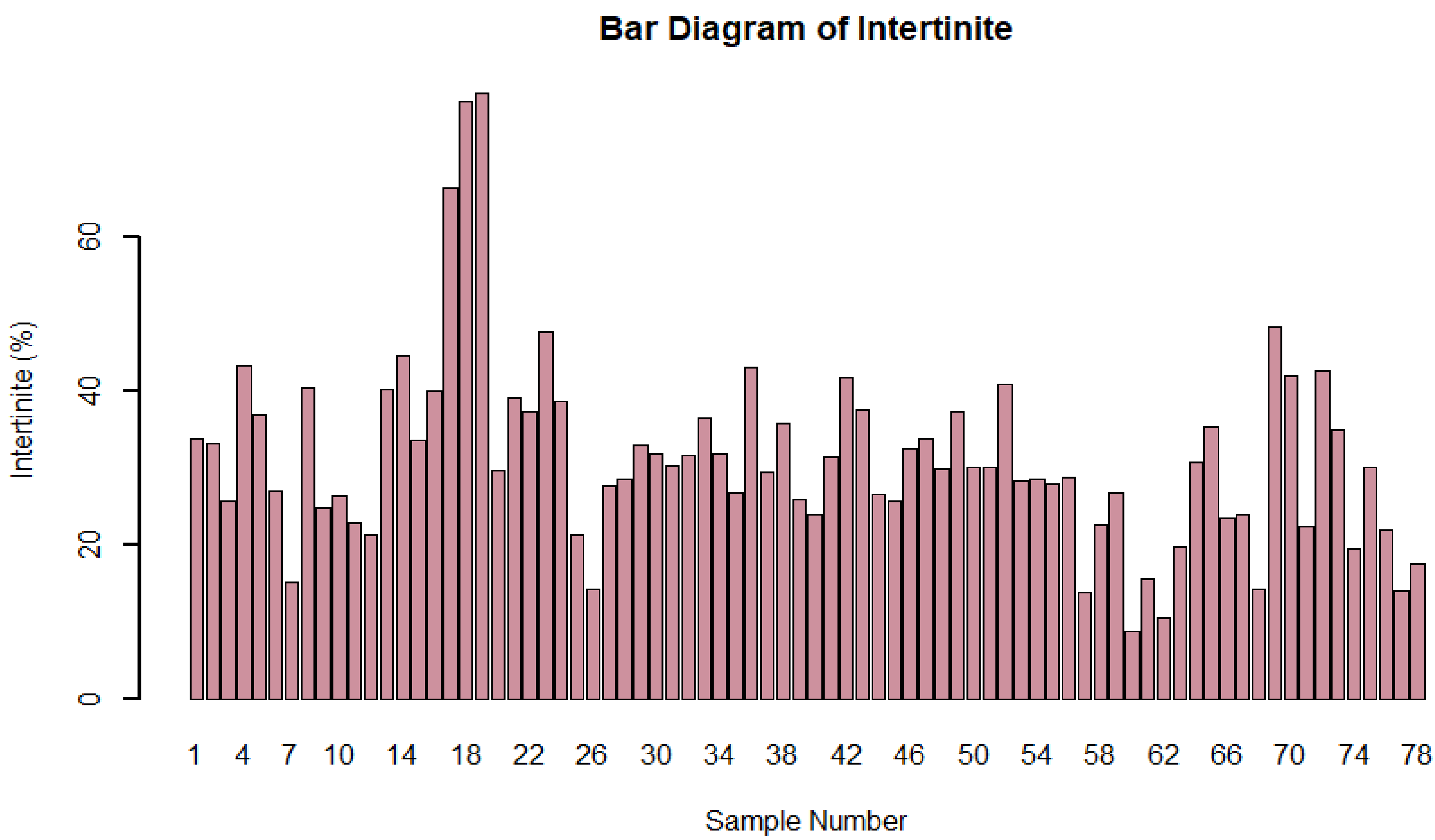
	vitrinite	cpt
vitrinite	1.0000000	-0.5454505
cpt	-0.5454505	1.0000000

This shows that vitrinite and crossing point temperature have moderate negative correlation.

<u>Intertinite:</u>

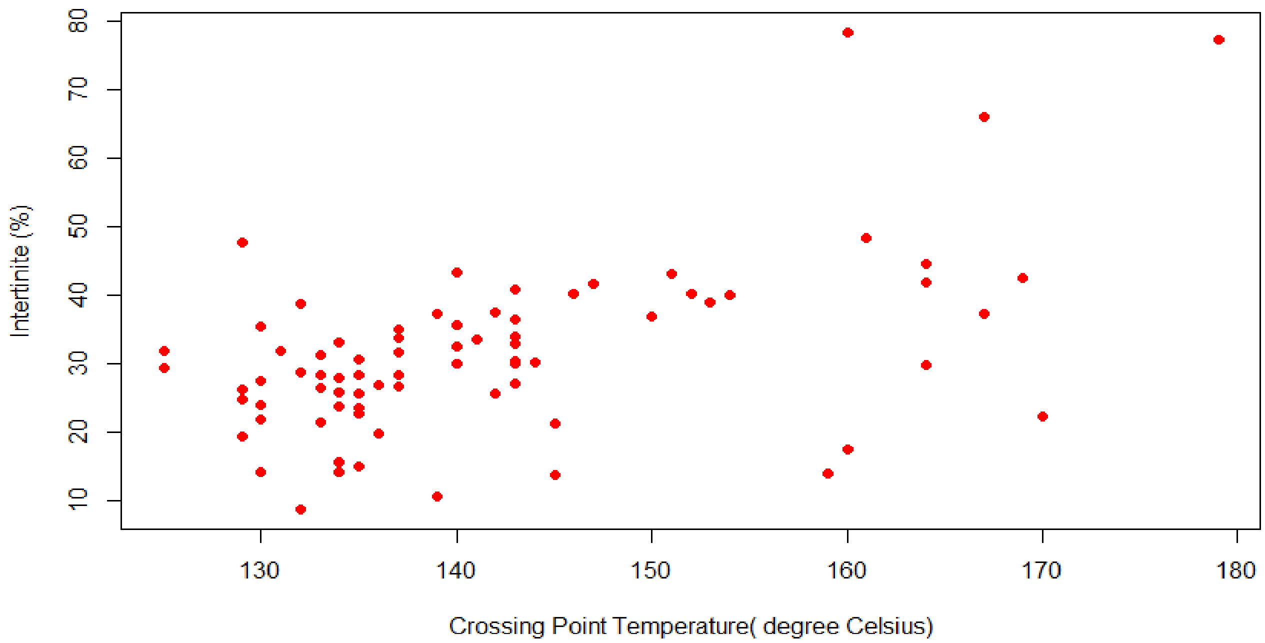
The data is visualized with a bar diagram:

> barplot(a.11,names=s,xlab="Sample Number",ylab="Intertinite (%)",main="Bar Diagram of Intertinite", col="pink3",lwd=2)



The scatter plot between intertinite and crossing point temperature is

> plot(a.11$cpt,a.11$intertinite,xlab="Crossing Point Temperature(degree Celsius)",ylab="Intertinite (%)",col="red",pch=19,cex=1)



The correlation coefficient between intertinite and crossing point temperature is found to be 0.5282726,

> cor(a.11,method="pearson")

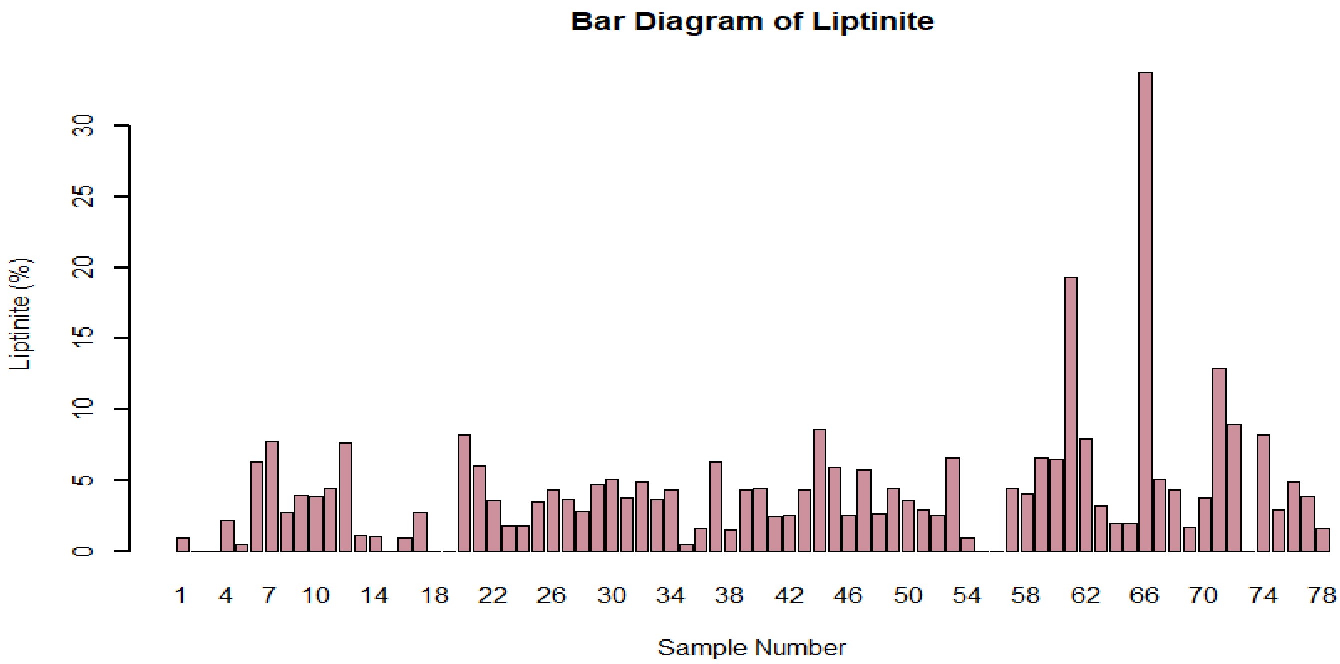
	intertinite	cpt
intertinite	1.0000000	0.5282726
cpt	0.5282726	1.0000000

This shows that intertinite and crossing point temperature have moderate positive correlation.

<u>Liptinite:</u>

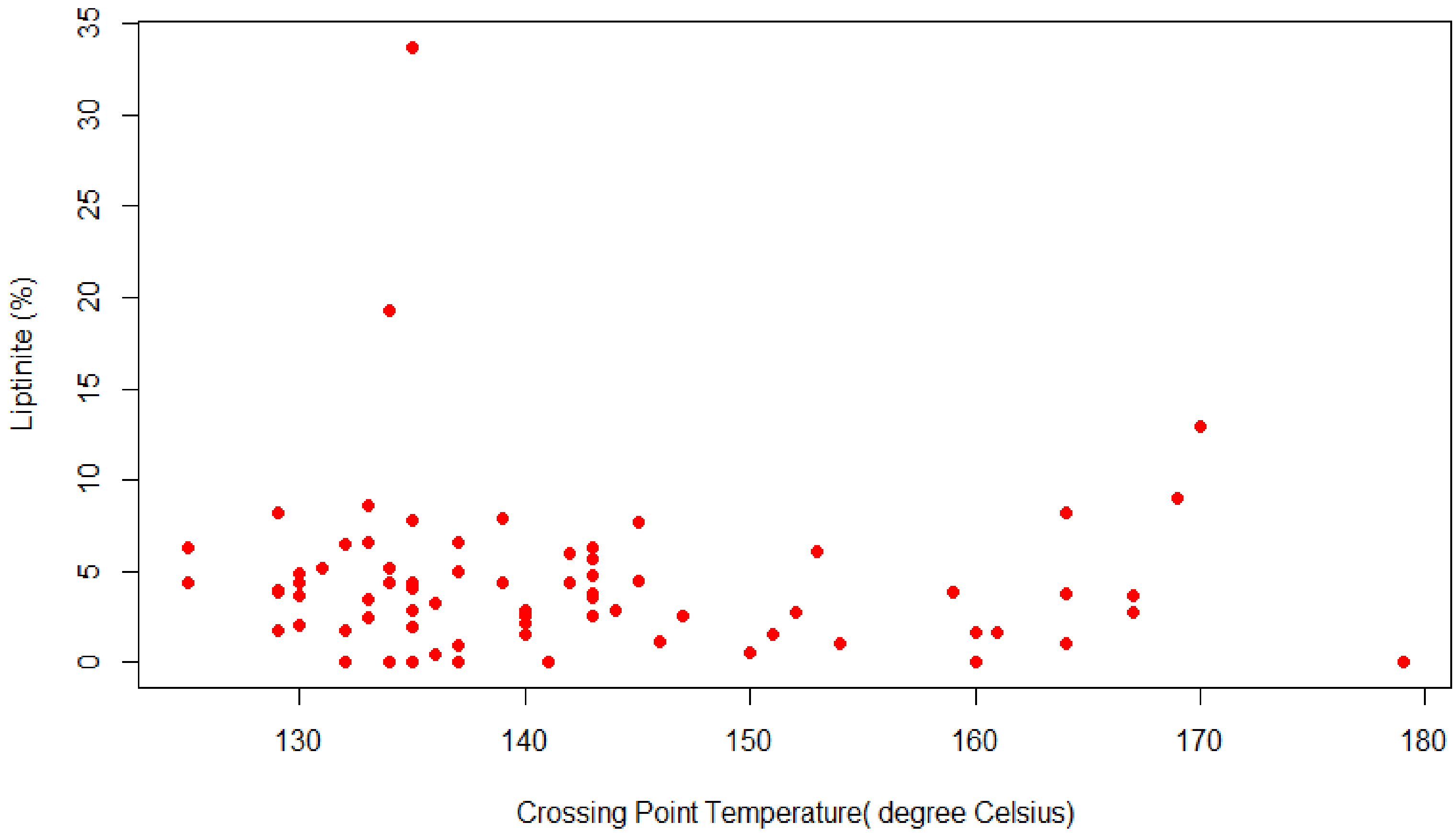
The data is visualized with a bar diagram:

> barplot(a.12,names=s,xlab="Sample Number",ylab="Liptinite (%)",main="Bar Diagram of Liptinite", col="pink3",lwd=2)



The scatter plot between liptinite and crossing point temperature is

> plot(a.12$cpt,a.12$liptinite,xlab="Crossing Point Temperature(degree Celsius)",ylab="Liptinite (%)",col="red",pch=19,cex=1)



The correlation coefficient between liptinite and crossing point temperature is found to be 0.005,

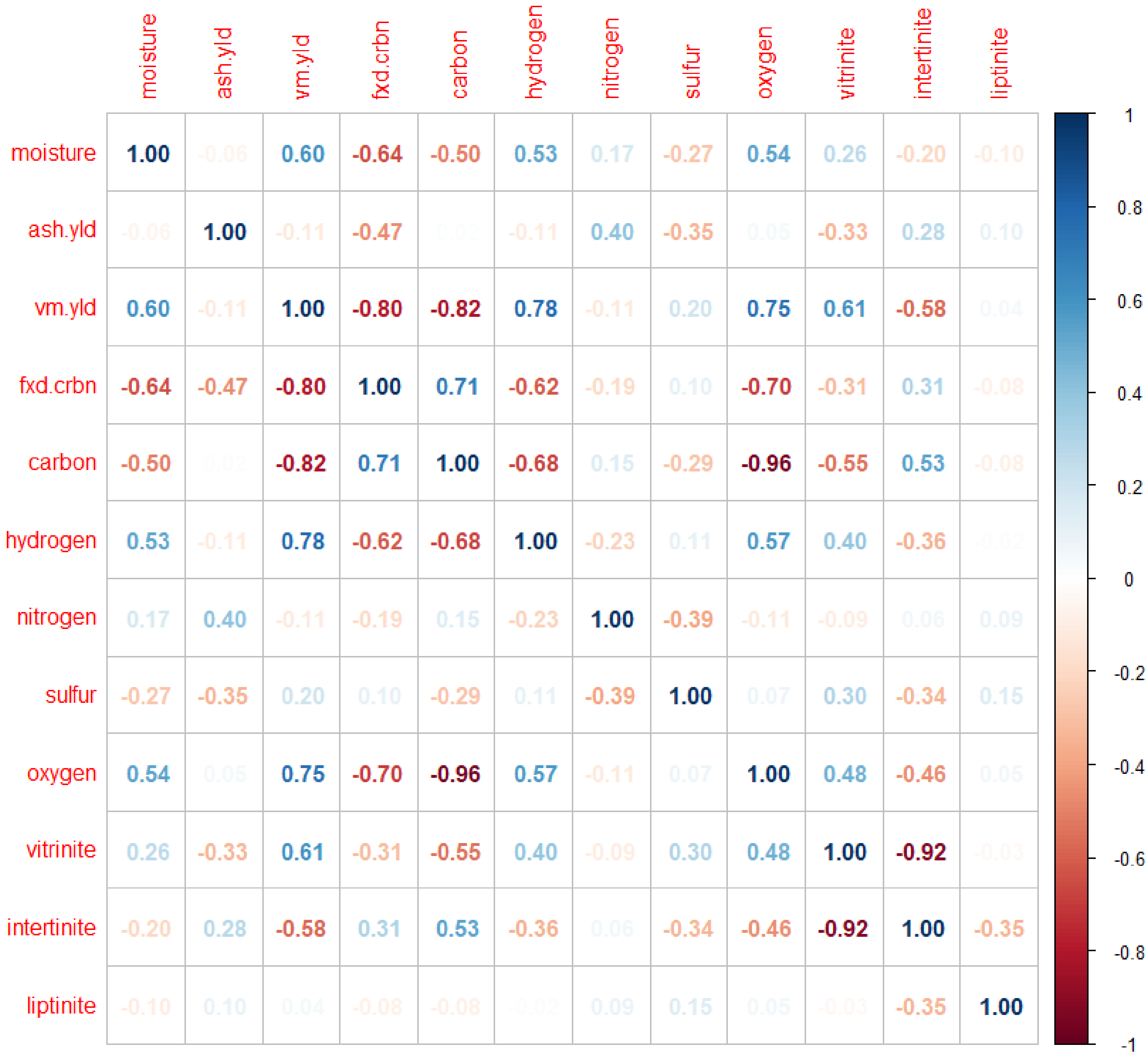
> cor(a.12,method="pearson")

	liptinite	cpt
liptinite	1.00000000	-0.06949183
cpt	-0.06949183	1.00000000

This shows that liptinite and crossing point temperature are have negligible correlation.

Now we will find the regression equation for Crossing Point Temperature with the intrinsic variables as the predictor variables, but before that we will check whether the variables are independent or not by finding their correlations

```
>a.c=cor(a,method="pearson")
>corrplot::corrplot(a.c,method="number")
```



We can see there is a high degree of multicollinearity in this dataset. So we will do a principal component analysis to reduce the multicollinearity of the dataset and also reduce the dimensionality of dataset.

ANALYSIS OF RESULTS USING PRINCIPAL COMPONENT ANALYSIS(PCA)

The principal component analysis is done below, which includes standardization of dataset.

```
> a.p=prcomp(a,center=TRUE,scale.=TRUE)
```

The coefficients for the 12 principal components (PC) are obtained as follows:

```
> -a.p$rotation
```

	PC1	PC2	PC3	PC4	PC5	PC6
moisture	0.2804191	-0.2762068	0.2371360	-0.2920490	0.3984487	0.1774998
ash.yld	-0.0400913	-0.49639126	-0.2823918	0.3435947	-0.4260219	-0.1309934
vm.yld	0.4137618	-0.01718778	0.0623166	0.0312052	0.0430865	0.1905205
fxd.crbn	-0.3438672	0.3574824	0.0694049	-0.1512798	0.1075028	-0.1231588
carbon	-0.4055941	0.0068222	-0.0166622	-0.2077098	0.1211280	0.0710722
hydrogen	0.3437625	-0.0275011	0.2364014	0.1496877	0.2424906	0.1982621
nitrogen	-0.0501105	-0.4049520	-0.3744769	-0.4558831	-0.1239246	0.4932266
sulfur	0.0930942	0.4746070	-0.1239468	0.3529375	-0.2588623	0.6416403
oxygen	0.3812613	-0.0901231	0.0398475	0.1519203	-0.0918033	-0.3616698
vitrinite	0.3117444	0.2684600	-0.1422301	-0.4321896	-0.3133949	-0.1650068
intertinite	-0.3039198	-0.2764367	0.3913193	0.3098716	0.0577924	0.1915941
liptinite	0.0393115	0.0506507	-0.6845538	0.2625998	0.6168516	-0.0761174

	PC7	PC8	PC9	PC10	PC11	PC12
moisture	0.2099727	-0.5325536	0.3849697	-0.0300191	-0.1873820	0.0369389
ash.yld	-0.2599529	-0.1833539	0.2900127	-0.0749277	-0.3984625	0.0764463
vm.yld	-0.1532497	-0.0468212	-0.7017270	-0.0871603	-0.5002970	0.0974766
fxd.crbn	0.1853001	0.2924178	0.2346930	-0.1050226	-0.7032242	0.1380398
carbon	-0.3835402	-0.2353290	-0.1194526	-0.7350843	0.1231359	0.0062820
hydrogen	-0.5254516	0.5228373	0.3716070	-0.1266879	0.0323487	-0.0048142
nitrogen	0.2285524	0.4130291	-0.0224980	-0.0583175	0.0064944	-0.0056610
sulfur	0.1816821	-0.2060184	0.1956138	-0.1833324	0.0106678	0.0008085
oxygen	0.5016882	0.2006052	0.0150774	-0.6166516	0.1009801	-0.0003536
vitrinite	-0.1949643	-0.0773064	0.1262105	0.0210068	0.1269122	0.6494763
intertinite	0.1779826	0.0834637	-0.1132778	0.0233285	0.1306356	0.6876993
liptinite	0.0262169	-0.0103668	-0.0083372	0.0142425	0.0394974	0.2634191

The i^{th} principal component PCi, denoted by Y_i can be written as:

$Y_i = \sum_{j=1}^{12} a_{ij} X_j \quad , \quad i, j = 1,2, \dots ,12$,where, X_j is the j[th] intrinsic variable of coal and a_{ij} is the coeffcient of the j[th] intrinsic variable in the i[th] principal component.

The table below shows the eigen values, variances, standard deviation, proportion of variance explained and cumulative proportion of variance explained by the principal components(PC).

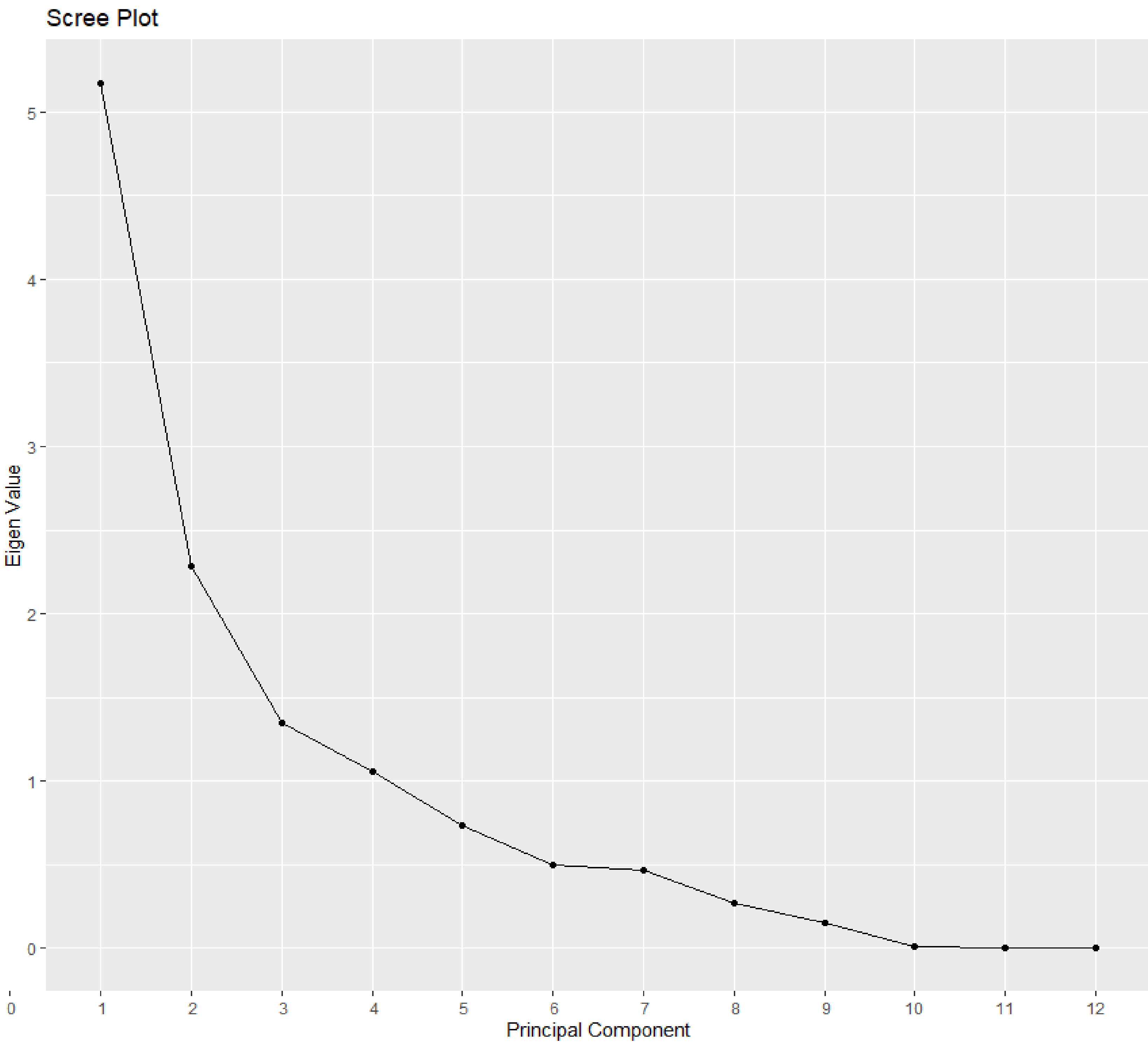
```
> a.p$sdev
> eigs<-a.p$sdev^2
>rbind(Eigen.Value=eigs,Variance=eigs,Standard.Deviation=sqrt(eigs),Proportion=eigs/sum(eigs),Cumulative=cumsum(eigs)/sum(eigs))
```

	Eigen value	Variance	Standard deviation	Proportion of variance explained	Cumulative proportion of variance explained	Cumulative Percentage of Variance Explained (%)
PC1	5.1719784	5.1719784	2.2741984	0.4309982	0.4309982	43.09982
PC2	2.2871203	2.2871203	1.5123228	0.1905934	0.6215916	62.15916
PC3	1.3458978	1.3458978	1.1601284	0.1121581	0.7337497	73.37497
PC4	1.05609804	1.05609804	1.02766631	0.08800817	0.82175787	82.175787
PC5	0.73398046	0.73398046	0.85672660	0.06116504	0.88292291	88.292291
PC6	0.4989792	0.4989792	0.7063846	0.0415816	0.9245045	92.45045
PC7	0.46971284	0.46971284	0.68535599	0.03914274	0.96364725	96.364725
PC8	0.27219014	0.27219014	0.52171845	0.02268251	0.98632976	98.632976
PC9	0.15375944	0.15375944	0.39212172	0.01281329	0.99914305	99.914305
PC10	0.0063520156	0.0063520156	0.0796995330	0.0005293346	0.9996723849	99.96723849
PC11	0.0030311819	0.0030311819	0.0550561700	0.0002525985	0.9999249834	99.99249834
PC12	9.001989×10^{-4}	9.001989×10^{-4}	3.000332×10^{-2}	7.501658×10^{-5}	1.0000000	100.0000

DETERMINING THE NUMBER OF PRINCIPAL COMPONENTS

To determine the number of principal components, we will first draw the screee plot.

```
> a.s <- qplot(c(1:12),eigs)+geom_line()+xlab("Principal Component")+ylab("Eigen Value")
+ggtitle("Scree Plot")
> a.s+scale_x_discrete(breaks=c(0:12),limits=c("0":"12"))
```

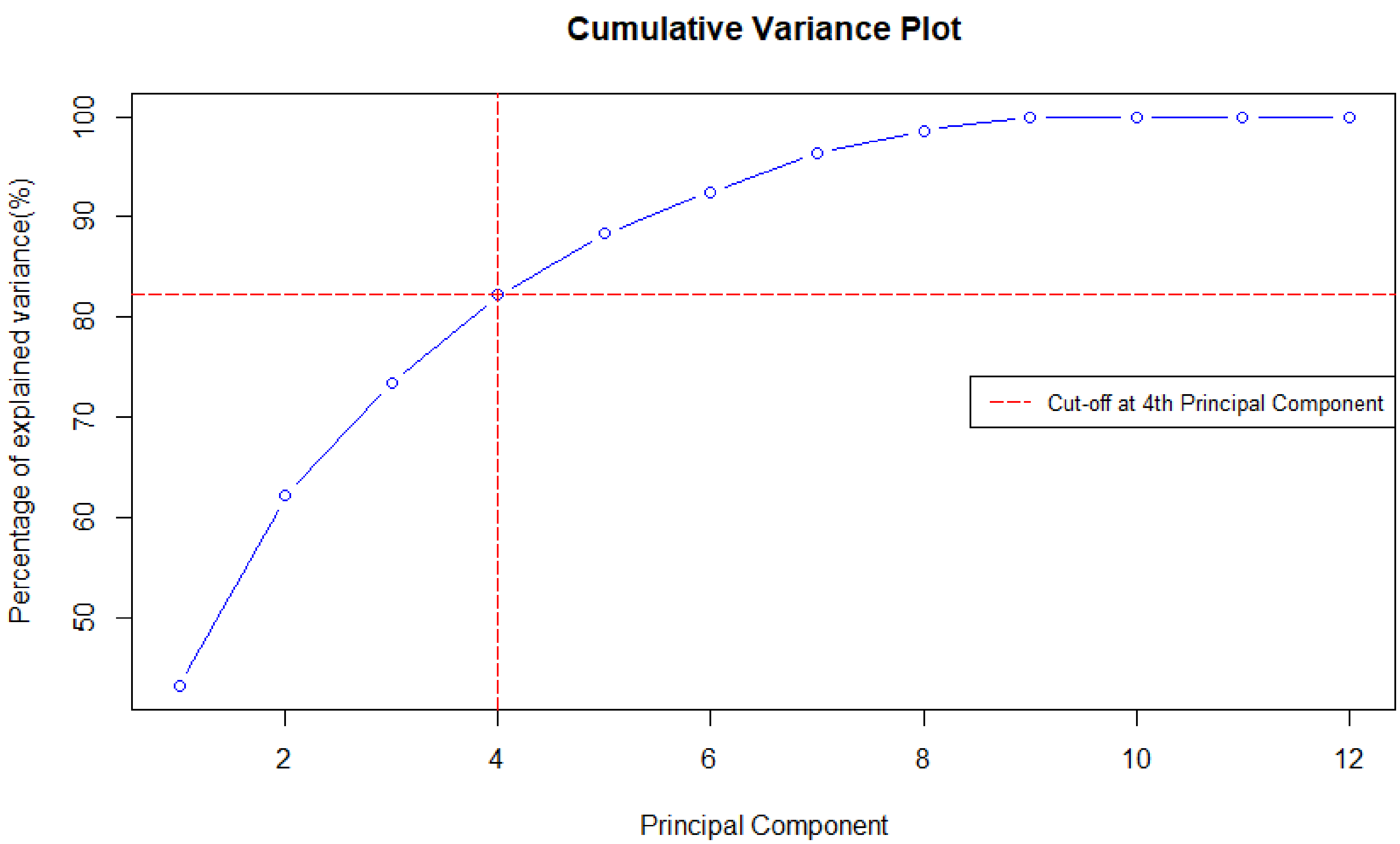


From the above scree plot we see that there is an elbow bend at the 6th principal component. So we retain 6 principal components which are capable of explaining about 95.45% variance.

A rule of thumb also suggests retaining only those components whose variances or eigen values are greater than unity .
From the table, we can see that only the first 4 principal components satisfy the above criteria and explain around 82.18% variance.

The cumulative variance plot is shown below:

```
> Cumulative=cumsum(eigs)/sum(eigs)
> plot(Cumulative[0:10], xlab = "Principal Component", ylab = “Percentage of explained variance (%)" , main = "Cumulative Variance Plot",type="b",col="blue")
> abline(v = 4, col="red", lty=5)
> abline(h = 82.175787, col="red", lty=5)
> legend( "right",legend=c("Cut-off at 4th Principal Component"),col=c("red"), lty=5, cex=0.8)
```



Hence a reduction in the data 78 observations on 12 variables to 78 observations on 4 principal components in reasonable.
Therefore, we conclude that four principal components effectively summarize the total variance present in the dataset.

INTERPRETATION OF PRNCIPAL COMPONENTS

For the interpretation of principal components, both the coeffcients and the correlation of the variables are taken into consideration. This is because variables with relatively large coefficients (absolute value) tend to have relatively large correlations, so these two measures frequently give the same results.

At first we will consider the coefficients given above and examine the magnitude and direction of the coefficients for the original variables. The larger the absolute value of the coefficient, the more important the corresponding variable is in calculating the component.

Here, first principal component has large positive association with carbon, and intertinite and large negative association with volatile matter yield. The second principal component has large positive association with ash yield and large negative association with sulfur. The third principal component has large positive association with liptinite. The fourth principal component has large positive association with nitrogen and vitrinite.

Next we will consider the correlations between the principal components and original variables .

```
> var <- get_pca_var(a.p)
> -var$cor
```

	PC1	PC2	PC3	PC4
moisture	0.63772886	-0.41771378	0.27510824	-0.30012897
ash.yld	-0.09117575	-0.75070382	-0.32761075	0.35310066
vm.yld	0.94097652	-0.02599348	0.07229521	0.03206859
fxd.crbn	-0.78202223	0.54062882	0.08051858	-0.15546519
carbon	-0.92240151	0.01031735	- 0.01933033	-0.21345633
hydrogen	0.78178409	-0.04159058	0.27425592	0.15382904
nitrogen	-0.11396129	-0.61241823	-0.43444134	-0.46849570
sulfur	0.21171478	0.71775905	-0.14379425	0.36270196
oxygen	0.86706388	-0.13629522	0.04622823	0.15612342
vitrinite	0.70896852	0.40599796	-0.16500523	-0.44414667
intertinite	-0.69117405	-0.41806149	0.45398059	0.31844460
liptinite	0.08940208	0.07660019	-0.79417027	0.26986493

The first principal component are most strongly correlated with volatile matter yield and carbon. The second principal component are most strongly correlated with ash yield and sulfur. The third principal component are most strongly correlated with liptinite. The fourth principal component are most strongly correlated with nitrogen and vitrinite. Combining both the results, we conclude that the first principal component is primarily a measure of volatile matter yield and carbon, the second principal component is primarily a measure of ash yield and sulfur, the third principal component is primarily a measure of liptinite, the fourth principal component is primarily a measure of nitrogen and vitrinite.

We can also get the principal component analysis scores for the 78 coal samples.

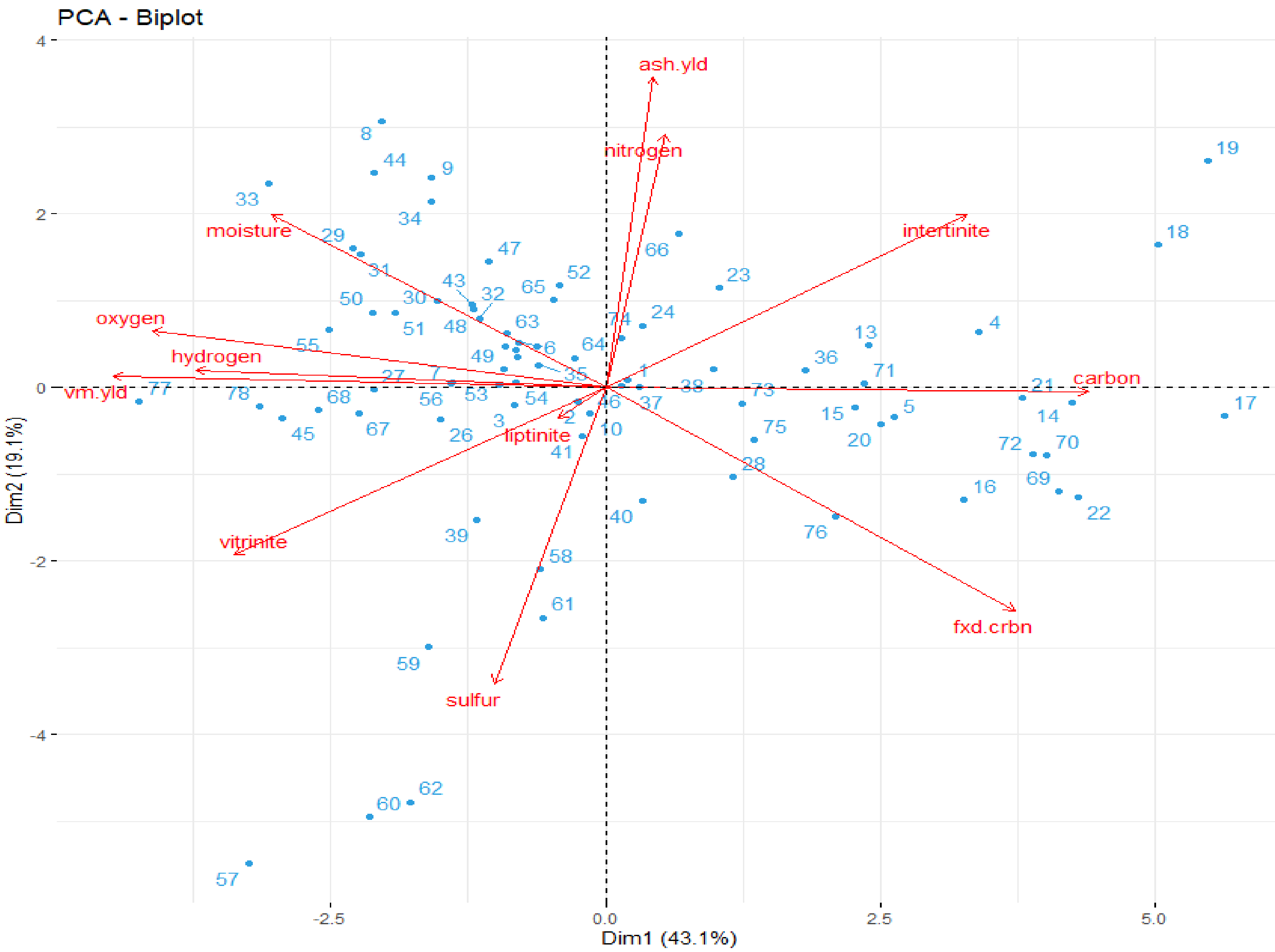
```
> a.p$x
            PC1          PC2          PC3          PC4
 [1,]  -0.2046202  -0.077126403   0.352756058  -0.563438560
 [2,] 0.2604738    0.167563362   0.837493297  -0.745400564
 [3,] 0.8264372    0.211243300   0.548133679  -1.661673219
 [4,]  -3.3948758  -0.632196041  -0.547141820  -0.065901461
 [5,]  -2.6189827   0.342674971   0.027730334  -0.896042480
 [6,] 0.7818535   -0.512736181  -0.611693885  -0.689255576
 [7,] 1.3922387   -0.049130002  -1.271911929  -1.333664811
 [8,] 2.0317427   -3.060213748  -0.219703723   0.616944842
 [9,] 1.5792257   -2.410551475  -1.532548921   0.050705603
[10,] 0.1387846    0.303636802  -0.061231961  -0.747300856
[11,] 0.8044096   -0.339517277  -0.879265513  -0.877367654
[12,] 0.8163753   -0.428580468  -1.380156008  -0.775962425
[13,]  -2.3946587  -0.483091402  -0.186184899  -0.076250670
[14,]  -4.2377815   0.177038797  -0.066818714  -0.778394198
[15,]  -2.2628917   0.228958790  -0.210595296  -0.964736296
[16,]  -3.2584040   1.299321817   0.723749257  -0.919172516
[17,]  -5.6329992   0.328612571   1.640354344   1.188696562
[18,]  -5.0199497  -1.642972405   1.932094971   2.390948287
[19,]  -5.4713356  -2.601545310   1.208464688   2.400166689
[20,]  -2.5045783   0.425958119  -1.392519666  -0.020264459
[21,]  -3.7906160   0.124257555  -0.903481864  -0.056172405
[22,]  -4.3014557   1.271116384   0.056097138  -0.536092278
[23,]  -1.0301942  -1.148987952   0.782911398   1.352354062
[24,]  -0.3316849  -0.709188491   0.410322668   0.396165404
[25,] 0.6303030   -0.472597378  -1.116044185  -1.021038992
[26,] 1.4990069    0.368388664  -0.499960990  -0.886105422
[27,] 2.1068591    0.033181186   0.429286727   0.946797761
[28,]  -1.1650481   1.033726263  -0.002919786  -1.010006340
[29,] 2.2911716   -1.593321634   0.527324633   0.919869873
[30,] 1.5323967   -0.997106789   0.211850945   1.027335343
[31,] 2.2334015   -1.535855265   0.296943290   1.311155831
[32,] 1.1513791   -0.787505064   0.083421771   0.040150778
[33,] 3.0673941   -2.346031141   0.419410653   2.623297345
```

[34,]	1.5860687	-2.131492623	-0.555525434	1.052683014
[35,]	0.6048375	-0.255250228	0.132296417	-1.497425216
[36,]	-1.8176563	0.190003949	0.479089691	-0.320503777
[37,]	-0.3067695	-0.003120689	-0.656753346	-0.689086010
[38,]	-0.9738512	-0.202455489	0.234827446	-0.870375871
[39,]	1.1690781	1.527153582	0.740441466	0.320600915
[40,]	-0.3321531	1.307020695	-0.479555730	-0.402837246
[41,]	0.2099410	0.559050015	0.384783231	0.489677700
[42,]	0.9118487	-0.472927035	1.584711771	0.826918755
[43,]	1.2143723	-0.952175983	0.923588067	0.573494731
[44,]	2.1011037	-2.464678560	-1.929664167	0.816525418
[45,]	2.9420085	0.362025166	1.014849716	0.313982274
[46,]	-0.1373300	-0.009469649	0.777972590	-0.768081379
[47,]	1.0632370	-1.443037940	0.419007326	-0.858887357
[48,]	1.1988665	-0.890418876	0.739419596	-1.349319951
[49,]	0.9238144	-0.207767995	1.177043835	0.256802040
[50,]	2.1202946	-0.859711565	0.958689489	-0.966318833
[51,]	1.9060860	-0.853620273	1.158540660	-0.023110257
[52,]	0.4230841	-1.174028881	1.573147372	-1.001158989
[53,]	0.9638777	-0.022903504	0.440201538	0.687331971
[54,]	0.8145745	-0.053850872	1.259797341	-0.564695007
[55,]	2.5204028	-0.659026218	1.167320509	0.002995482
[56,]	1.3988410	-0.043140809	1.389354502	-1.049550888
[57,]	3.2461184	5.490185967	-0.064060735	2.339916668
[58,]	0.5928329	2.099316656	0.771991107	-0.179174911
[59,]	1.6090968	2.997726859	0.281366625	1.874755337
[60,]	2.1508093	4.953733375	-0.456199105	1.164193729
[61,]	0.5642811	2.655829294	-1.827678650	0.857445440
[62,]	1.7752003	4.781603268	-0.808155595	1.069411788
[63,]	0.8941864	-0.614245250	-1.090338915	-1.527906150
[64,]	0.2850195	-0.334876345	0.241791563	0.089771396
[65,]	0.4753729	-1.001465109	0.320160028	0.875520286
[66,]	-0.6587526	-1.764416668	-6.072729205	1.784578193
[67,]	2.2481892	0.301079996	0.003550499	0.711590314
[68,]	2.6150184	0.261775520	-0.452518922	0.072360088
[69,]	-4.1236971	1.198676187	1.047247164	0.056331697
[70,]	-4.0061072	0.784636853	0.146535087	-0.126214678
[71,]	-2.3581415	-0.039292258	-2.667729341	0.500405395
[72,]	-3.8884951	0.773651661	-0.260070219	0.708874686
[73,]	-1.2390723	0.189621431	0.388068663	-0.818705096
[74,]	-0.1505664	-0.571870331	-2.232175675	-1.217899933
[75,]	-1.3515857	0.607539027	-0.172665751	-0.314251216
[76,]	-2.0966746	1.491007960	-0.844626743	-1.187426477
[77,]	-4.2381220	0.167442146	0.612625130	-1.516125087
[78,]	3.1508915	0.218747281	0.595862416	-0.837460189

BIPLOT

A biplot is a very popular way for visualizing results from PCA as it combines both the principal component scores and loading or coordinate vectors into a single biplot display. The projection values or the length of the vectors on each principal component show how much weight they have on that principal component. The cosine of the angle between a vector and an aixs indicates the importance of contribution of the corresponding variable to the principal component. The cosine of the angle between pairs of vectors indicate correlation between the correponding variables in this space. When two vactors are close, forming a small angle on the same side of the quadrant the two variables they represent are positively correlated.If they meet each other at right angle i.e., they are perpendicular , they are uncorrelated. When they diverge and form a large angle i.e., on opposite sides of plot origin they are negatively correlated. Points that are close to each other in the biplot represent observations with similar values. The position of the points in the biplot represent the corresponding score of that observation in the principal component.
A biplot of individuals and variables for the components that explain maximum variability i.e., PC1 and PC2 is shown below:

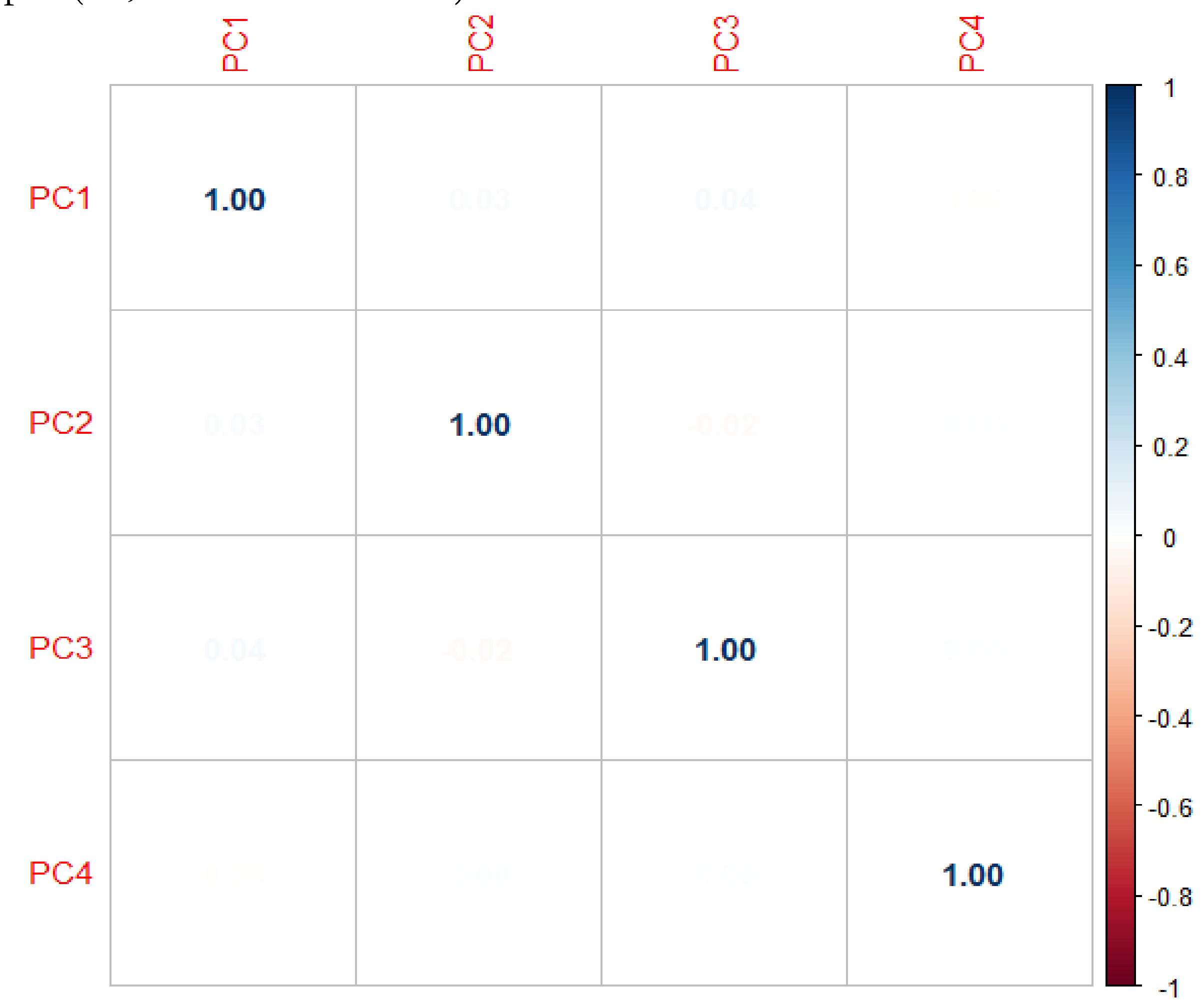
> fviz_pca_biplot(a.p, repel = TRUE,col.var = "red",col.ind = "#2E9FDF",axes=1:2)



We can see that length of vectors for vm.yld(volatile matter yield) and carbon are the largest on the x axis and the angles they form with x axis is very small. So they have the maximum contribution to Dim1 (1st principal component). The length of vectors for ash.yld(ash yield) and sulfur are the largest on the y axis and the angles they form with y axis is the minimum. So they have the maximum contribution to Dim2 (2nd principal component). Variables like vm.yld(volatile matter yield), hydrogen, oxygen are grouped together indicating they are positively correlated. Variables like hydrogen and ash yield, moisture and intertinite are uncorrelated. Variables like vm.yld and fxd.crbn; oxygen and carbon are negatively correlated. The maximum and minium score for the 1st principal component is of 17th and 77th observation respectively. Observations like 46th , 10th have scores close to zero. The maximum and minium score for the 2nd principal component is that of 8th and 57th observation respectively. Observations like 46th , 37th ,53th ,27th have scores almost close to zero.

Now we will find the correlation between the first four principal components.

```
> a.c=cor(-a.p$rotation[,c(1:4)],method="pearson")
> corrplot::corrplot(a.c,method="number")
```



So, the principal components are uncorrelated , indicating absence of multicollinearity.

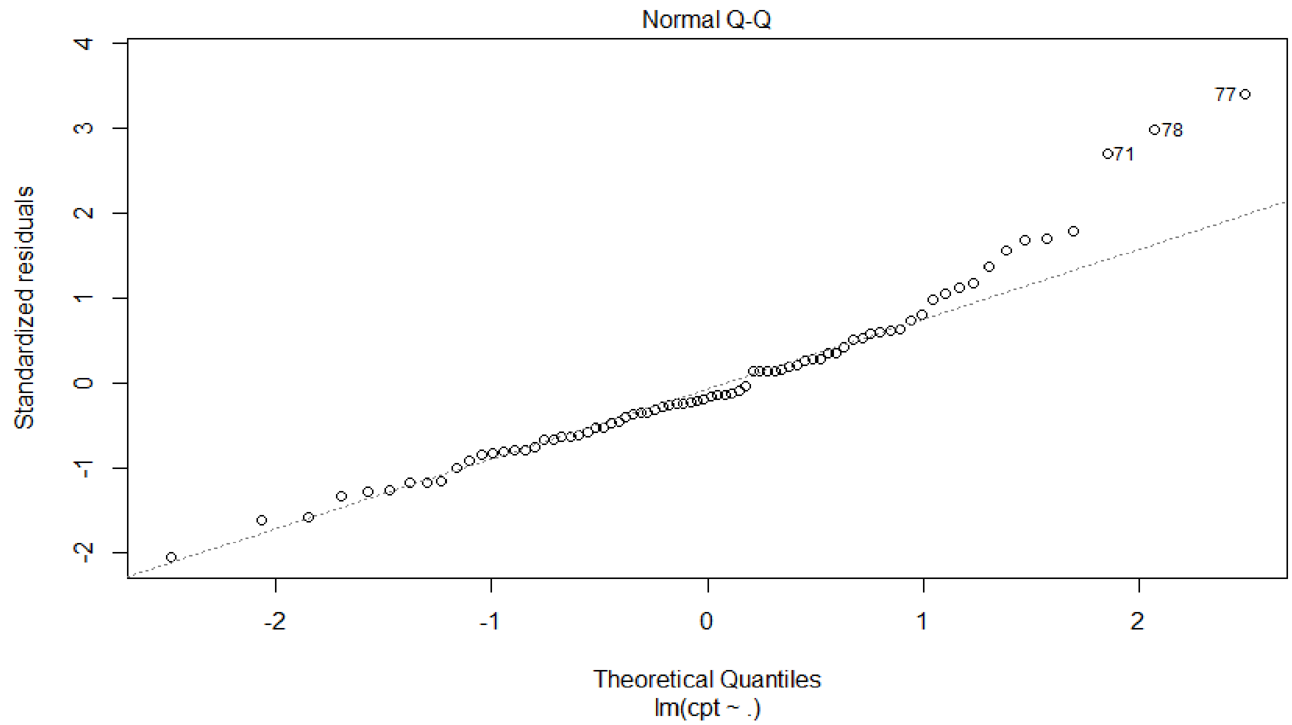
Now we will use these four principal components as the predictor variables for CPT and fnd out the regresson equation.

```
> b=as.data.frame(-a.p$x[,c(1:4)]
> c=read.table("india_coal.txt",header=TRUE,sep='\t')
> cpt=scale(c$cpt)
> e=as.data.frame(cbind(b,cpt))
> e.lm=lm(cpt~PC1+PC2+PC3+PC4,data=e)
> summary(e.lm)
```

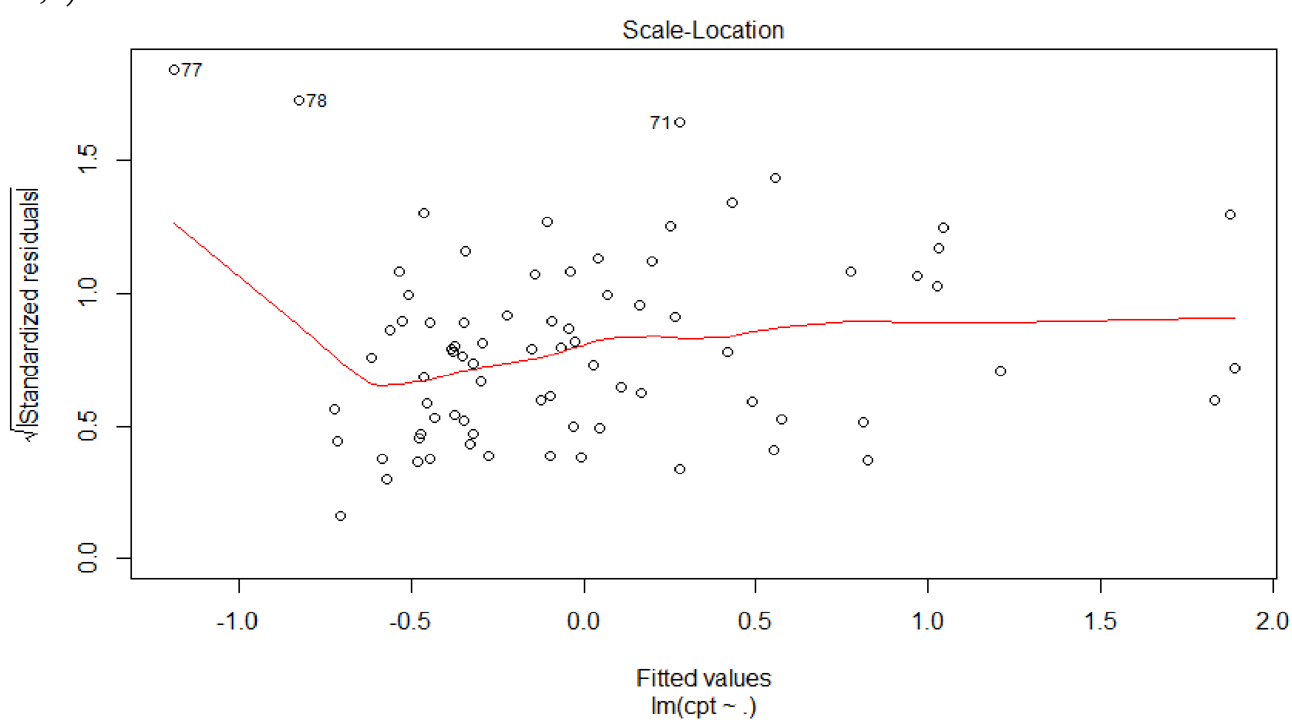
```
Call:
lm(formula = cpt ~ PC1 + PC2 + PC3 + PC4, data = e)

Residuals:
   Min     1Q   Median    3Q     Max
-1.6151 -0.4960 -0.1326  0.3759  2.6250

Coefficients:
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) -7.742e-16  9.137e-02    0.000    1.0000
PC1         -2.548e-01  4.044e-02   -6.301     2e-08 ***
PC2         -9.403e-04  6.081e-02   -0.015    0.9877
PC3          1.457e-01  7.927e-02    1.837    0.0702 .
PC4          1.316e-01  8.949e-02    1.470    0.1458
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.807 on 73 degrees of freedom
Multiple R-squared:  0.3826,	Adjusted R-squared:  0.3488
F-statistic: 11.31 on 4 and 73 DF,  p-value: 3.391e-07
```

So, we have found the estimates of the independent variables, but before writing the final regression equation, we have to check whether the assumptions of normality and homoscedasticity are followed here.

```
> e.lm=lm(cpt~.,data=e)
> plot(e.lm,2)
```



> plot(e.lm,3)



We can see that in the 1st plot, all the points do not fall on the line, specially near the end. So, we conclude that the residuals do not follow a normal distribution. In the 2nd plot, the red line is not horizontal, implying that the condition of homoscedasticity is not satisfied here. So , the regression assumptions are not satisfied and and hence, the estimates are not reliable and the regression equation will be incorrect for prediction and forecasting purposes.

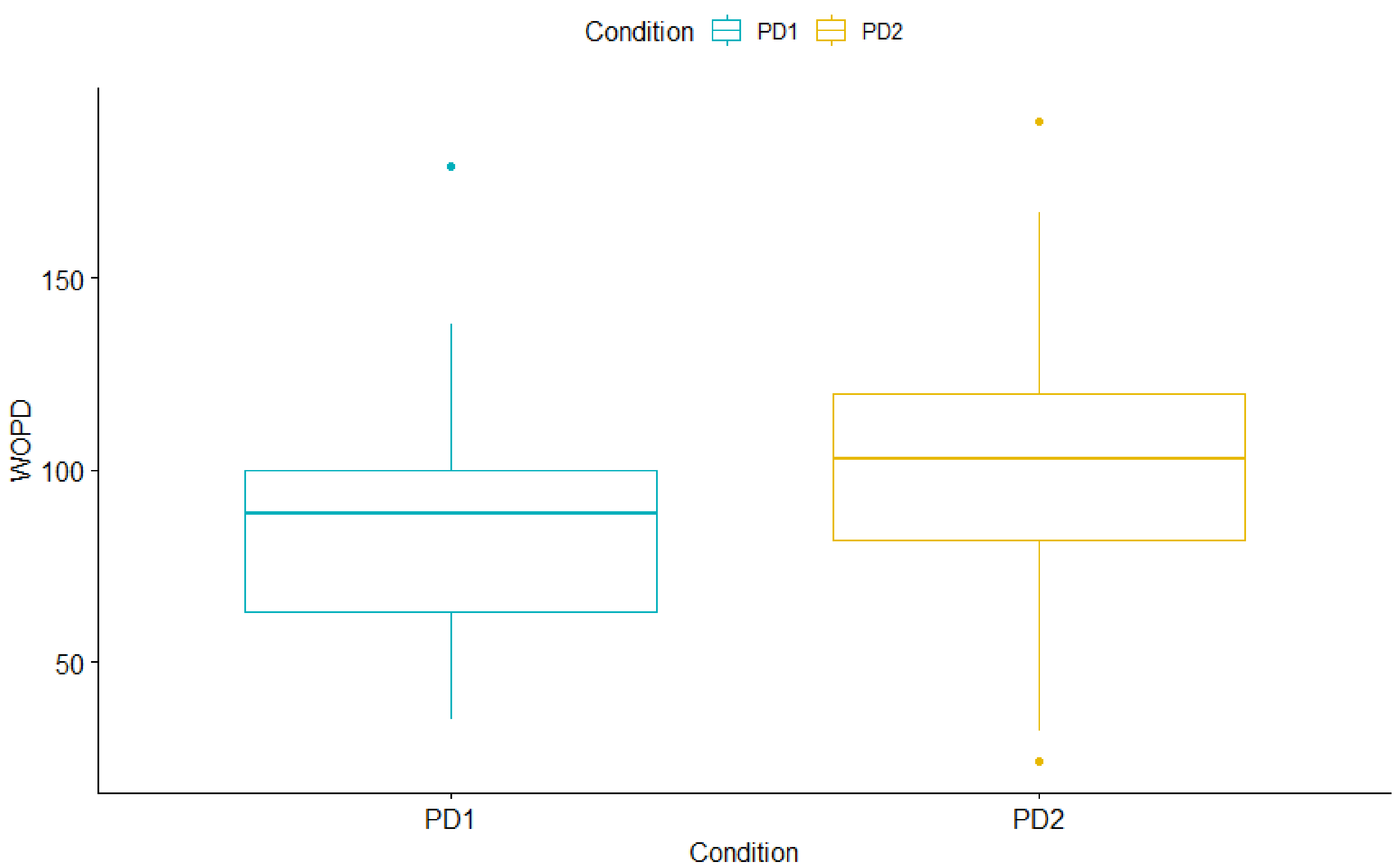
Now we will test whether there is any difference in effect of different conditions PD1, PD2 and PD3 on average WOPD by nonparametric tests because the basic assumptions for parametric tests like homoscedasticity and normality do not hold here.
The null hypothesis is H_0 = no difference in effect of different conditions on WOPD
The alternative hypothesis is H_1 = there is a difference in the effect

To test the above hypothesis, we can use either paired sample sign test or paired sample Wilcoxon Signed Rank test. Paired sample Wilcoxon Signed Rank test requires an additional assumption that the distribution of differences between the two samples is symmetric.

1.PD1 and PD2

The boxplot can be drawn as follows:

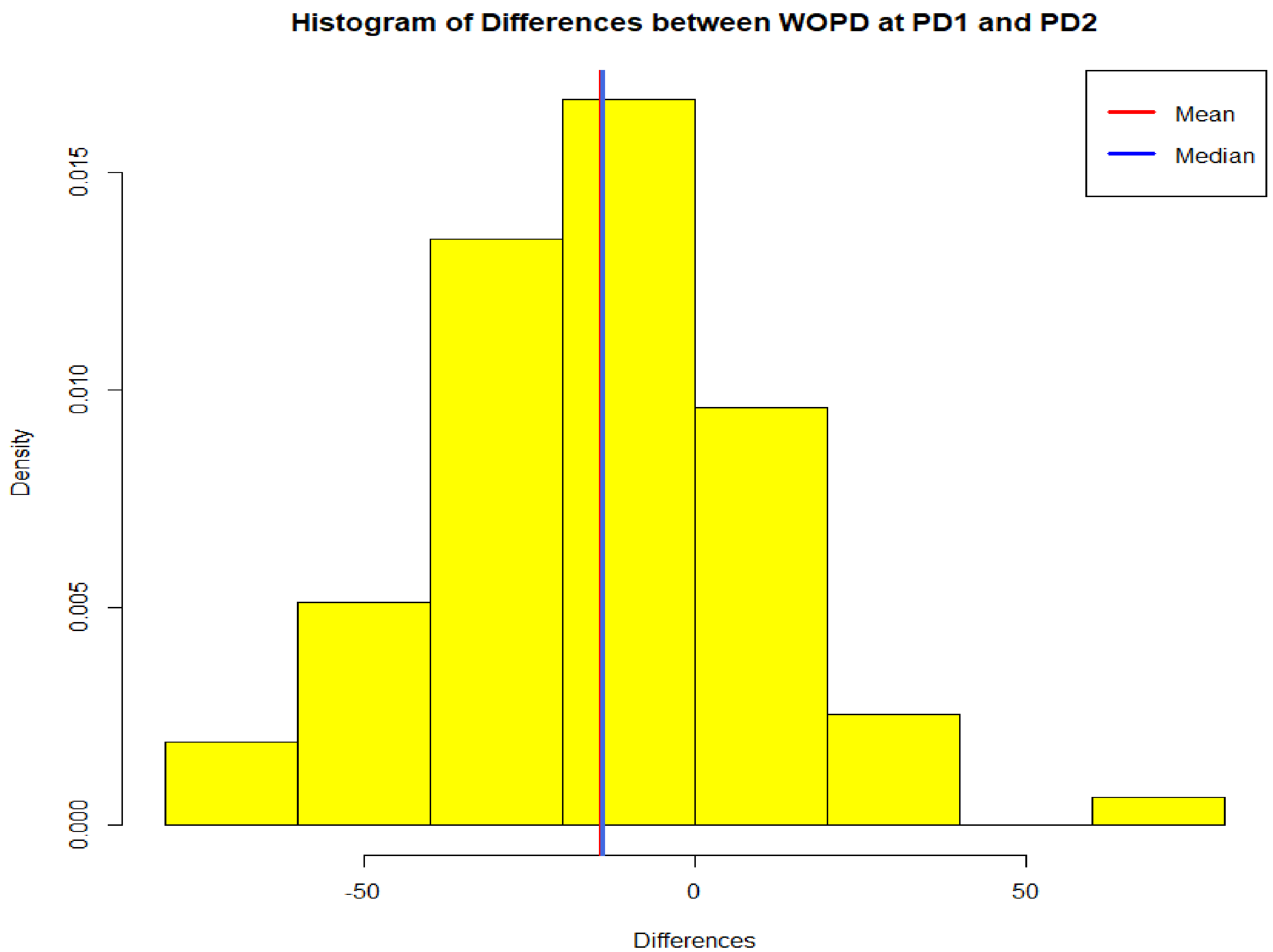
```
> ggboxplot(b,x="Condition",y="WOPD",color="Condition",palette = c("#00AFBB", "#E7B800"))
```



The histogram of difference between WOPD at the above conditions is:

```
> d1=b$PD1-b$PD2
```

```
> hist(d.1,col="yellow",freq=FALSE,xlab="Differences",main="Histogram of Differences between WOPD at pd1 and pd2")
> abline(v=mean(d.1),col="red",lwd=3)
> abline(v=median(d.1),col="royal blue",lwd=3)
> legend(x="topright",c("Mean","Median"),col=c("red","blue"),lwd=c(3,3))
```



We can see that the differences are not symmetrically distributed and the skewness is found to be 0.4374516 , which implies that they are positively skewed.

```
> library(moments)
> skewness(d1)
[1] 0.4374516
```

So we are using paired sample sign test instead of Wilcoxon Signed rank test.

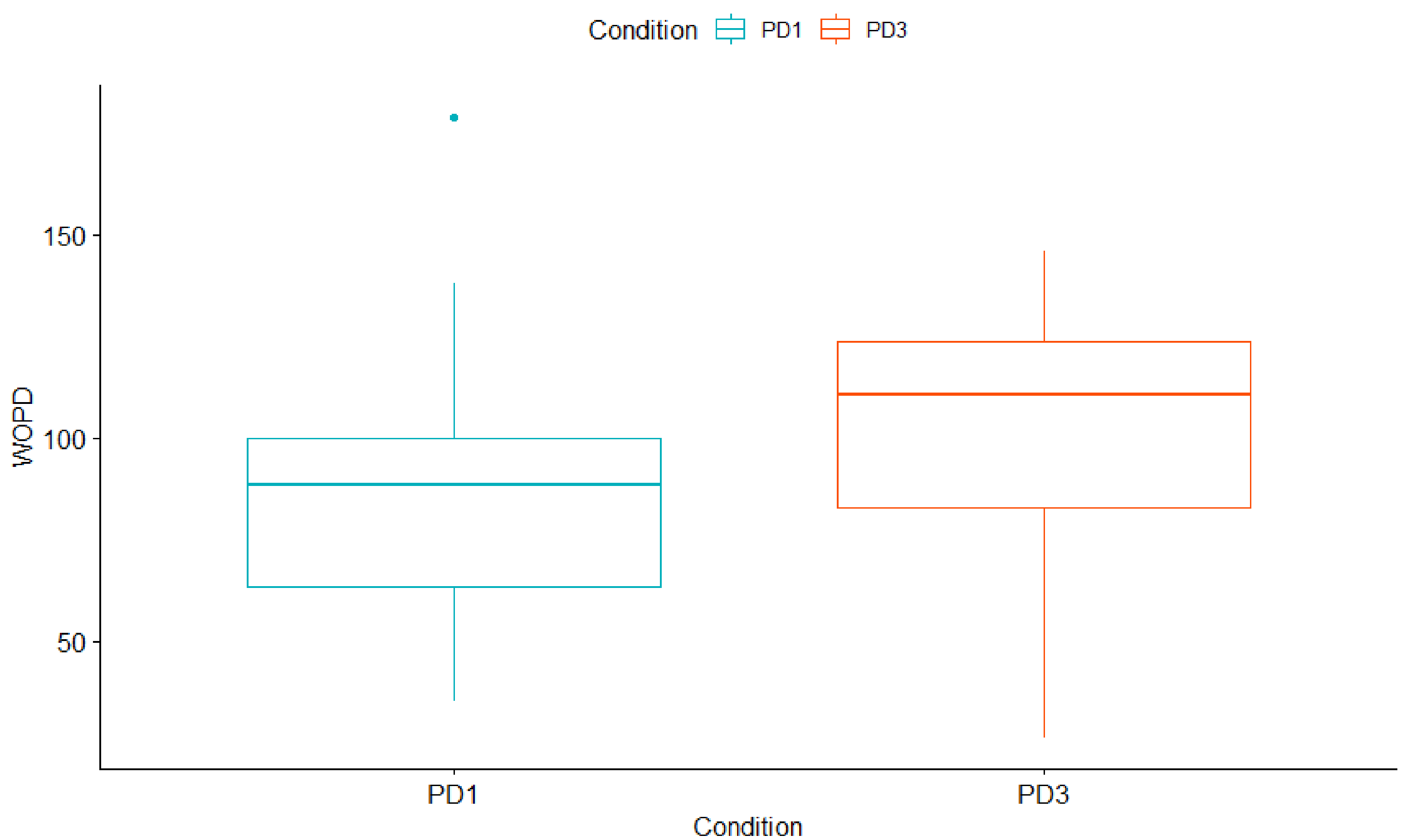
```
> sign_test(b.1,WOPD~PD,mu=0,alternative="two.sided",p.adjust.method="none")
# A tibble: 1 x 8
    .y.   group1 group2   n1    n2   statistic   df      p
* <chr>  <chr>  <chr>  <int> <int>   <dbl>  <dbl>   <dbl>
1 WOPD    PD1    PD2     78    78      20      77   0.0000293
```

Since the p value 0.0000293 is less than the significance level 0.05 we reject the null hypothesis.

2.PD1 and PD3

The boxplot can be drawn as follows:

> ggboxplot(c, x="Condition", y "WOPD",color="Condition",palette = c("#00AFBB", "#FC4E07"))



The histogram of difference between WOPD at the above conditions is:

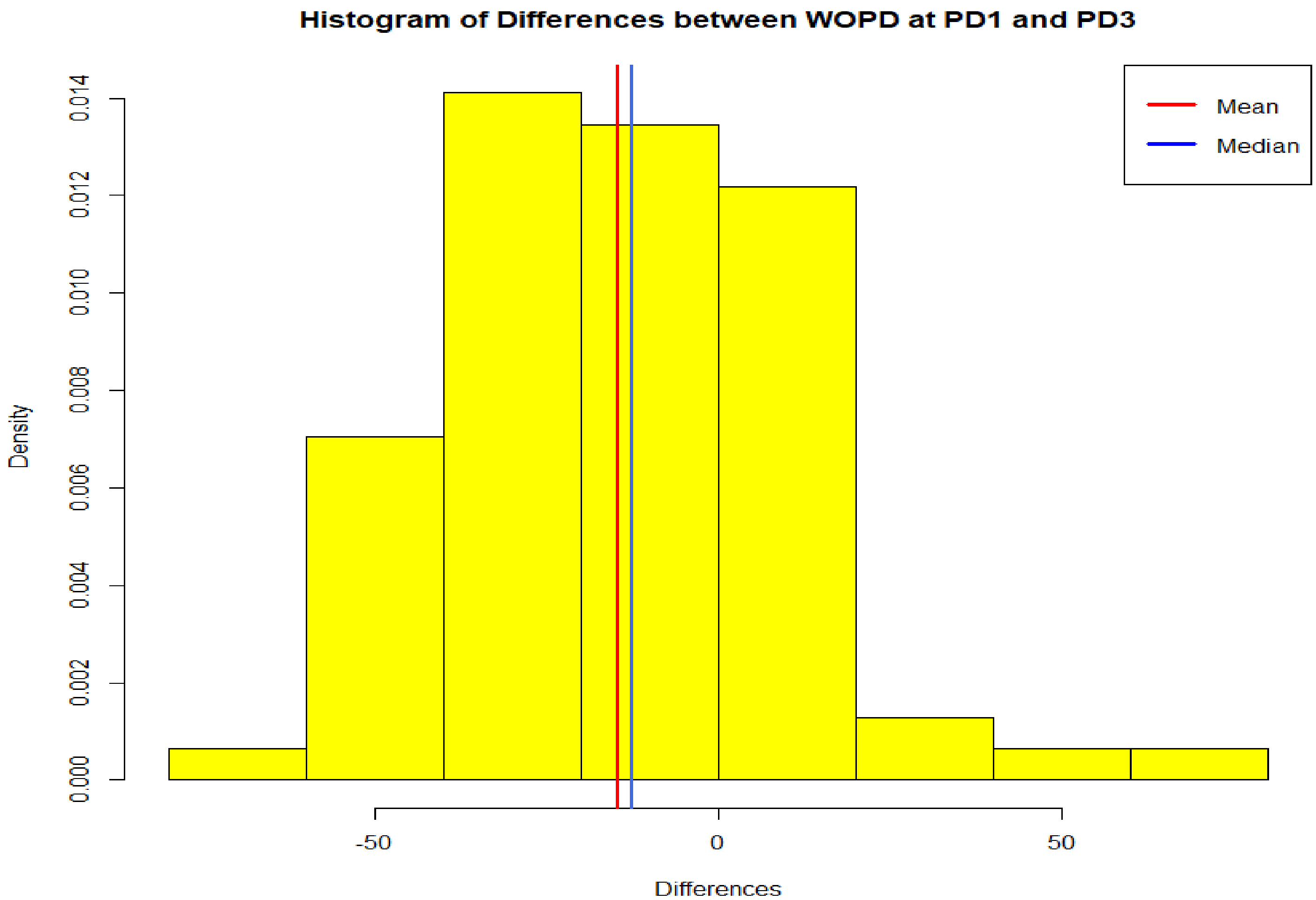
> d.2=b$PD1-b$PD3

> hist(d.2,col="yellow",freq=FALSE,xlab="Differences",main="Histogram of Differences between WOPD at pd1 and pd3")

> abline(v=mean(d.2),col="red",lwd=3)

> abline(v=median(d.2),col="royal blue",lwd=3)

> legend(x="topright",c("Mean","Median"),col=c("red","blue"),lwd=c(3,3))



We can see that the differences are not symmetrically distributed and the skewness is found to be 0.38775, which implies that they are positively skewed.

```
> library(moments)
> skewness(d.2)
[1] 0.38775
```

So we are using paired sample sign test instead of Wilcoxon Signed rank test.

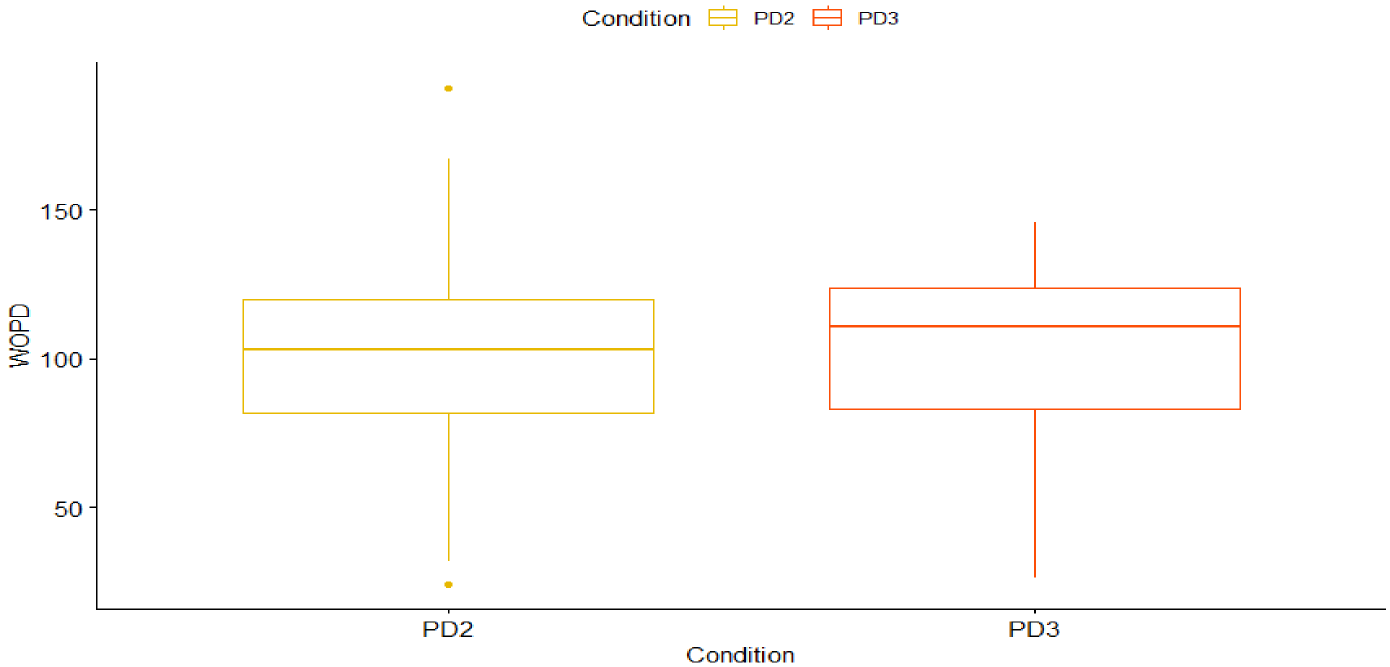
```
> sign_test(a,WOPD~PD,mu=0,alternative="two.sided",p.adjust.method="none")
# A tibble: 1 x 8
     .y.   group1 group2   n1    n2    statistic   df      p
*  <chr> <chr>  <chr>  <int> <int>    <dbl> <dbl>    <dbl>
1 WOPD   PD1    PD3      78    78        23     78  0.000378
```

Since the p value 0.000378 is less than the significance level 0.05, we reject the null hypothesis.

3.PD2 and PD3

The boxplot can be drawn as follows:

> ggboxplot(d, x="Condition", y="WOPD",color ="Condition",palette=c("#E7B800", "#FC4E07"))



The histogram of difference between WOPD at the above conditions is:

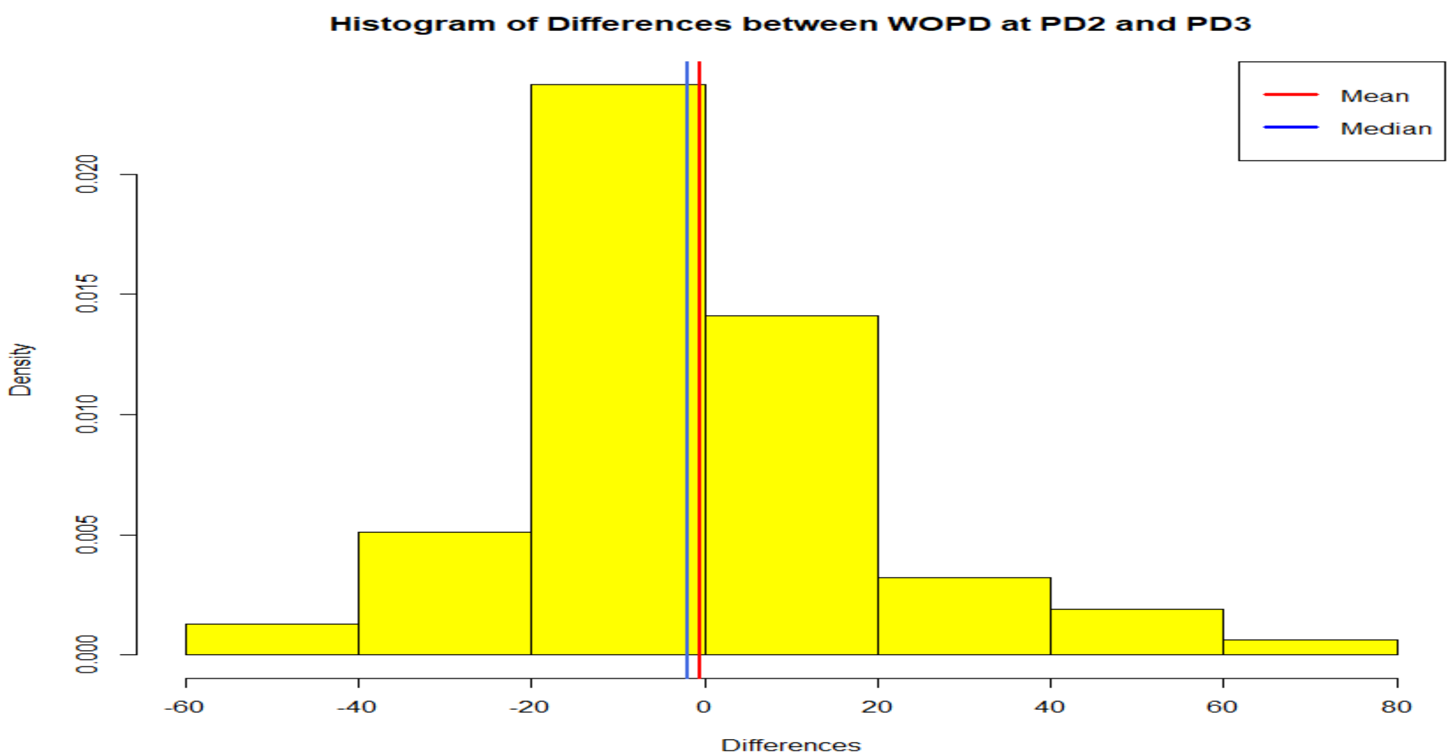
> d.3=b$PD2-b$PD3

> hist(d.6,col="yellow",freq=FALSE,xlab="Differences",main="Histogram of Differences between WOPD at pd2 and pd3")

> abline(v=mean(d.6),col="red",lwd=3)

> abline(v=median(d.6),col="royal blue",lwd=3)

> legend(x="topright",c("Mean","Median"),col=c("red","blue"),lwd=c(3,3))



We can see that the differences are not symmetrically distributed and the skewness is found to be 1.016788, which implies that they are positively skewed.

```
> library(moments)
> skewness(d1)
[1] 1.016788
```

So we are using paired sample sign test instead of Wilcoxon Signed rank test.

```
> sign_test(a,WOPD~PD,mu=0,alternative="two.sided",p.adjust.method="none")
# A tibble: 1 x 8
    .y.   group1 group2   n1    n2   statistic  df     p
*  <chr> <chr>  <chr>  <int> <int>  <dbl>   <dbl>  <dbl>
1 WOPD  PD2     PD3     78    78     31      75    0.165
```

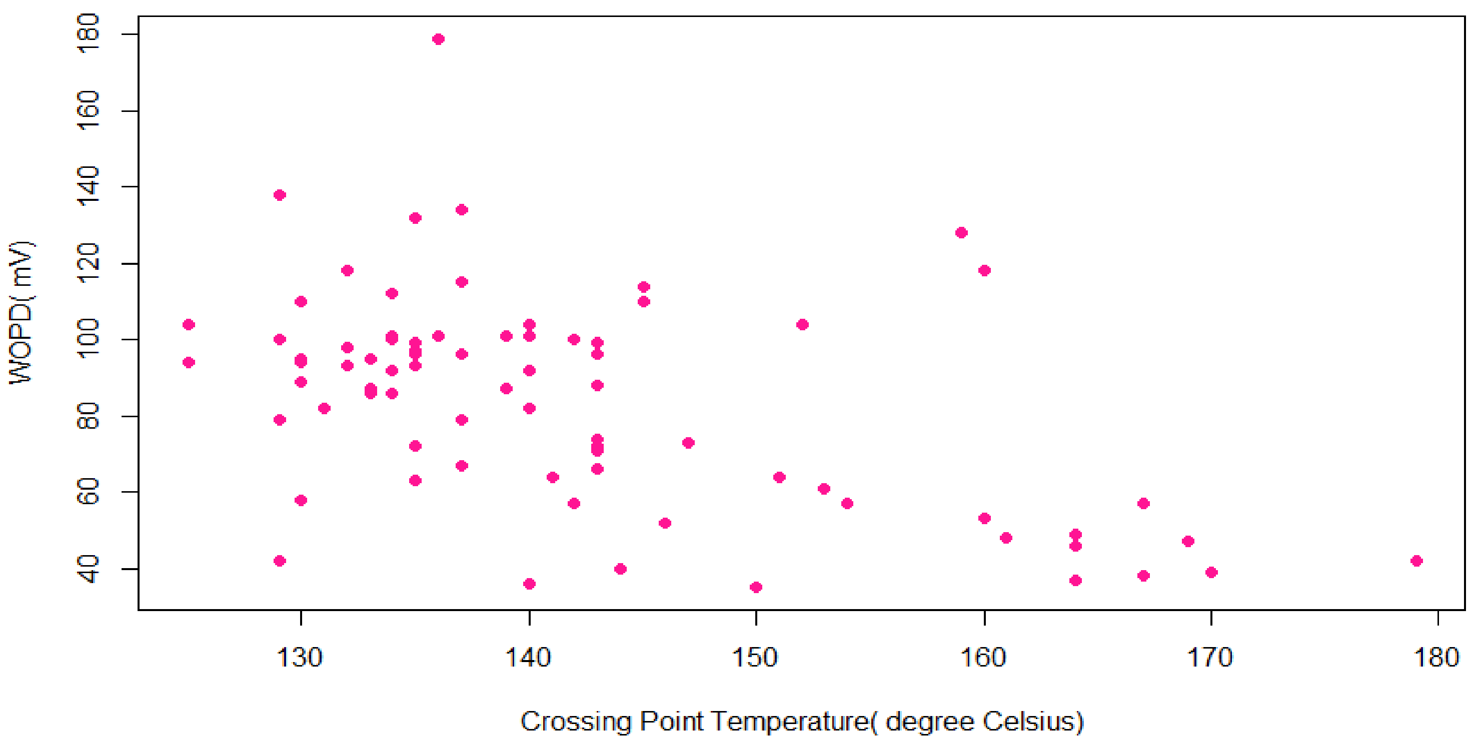
Since the p value 0.165 is more than the significance level 0.05, we accept the null hypothesis.

Therefore, we see that the effect of PD1 on WOPD is significantly different from that of PD2 and PD3 but there is no difference between PD2 and PD3.

Now , we find the relation between the crossing point temperature(CPT) and Wet Oxidation Potential Difference(WOPD) by drawing the scatter plot and finding the correlation between them.

The scater plot between CPT and WOPD at PD1 with the equatiois:

```
> plot(b$cpt,b$pd3,xlab="Crossing Point Temperature( degree Celsius)",ylab="WOPD ( mV)",
col="deep pink",pch=19,cex=1)
```



The correlation between them is found to be -0.5107481

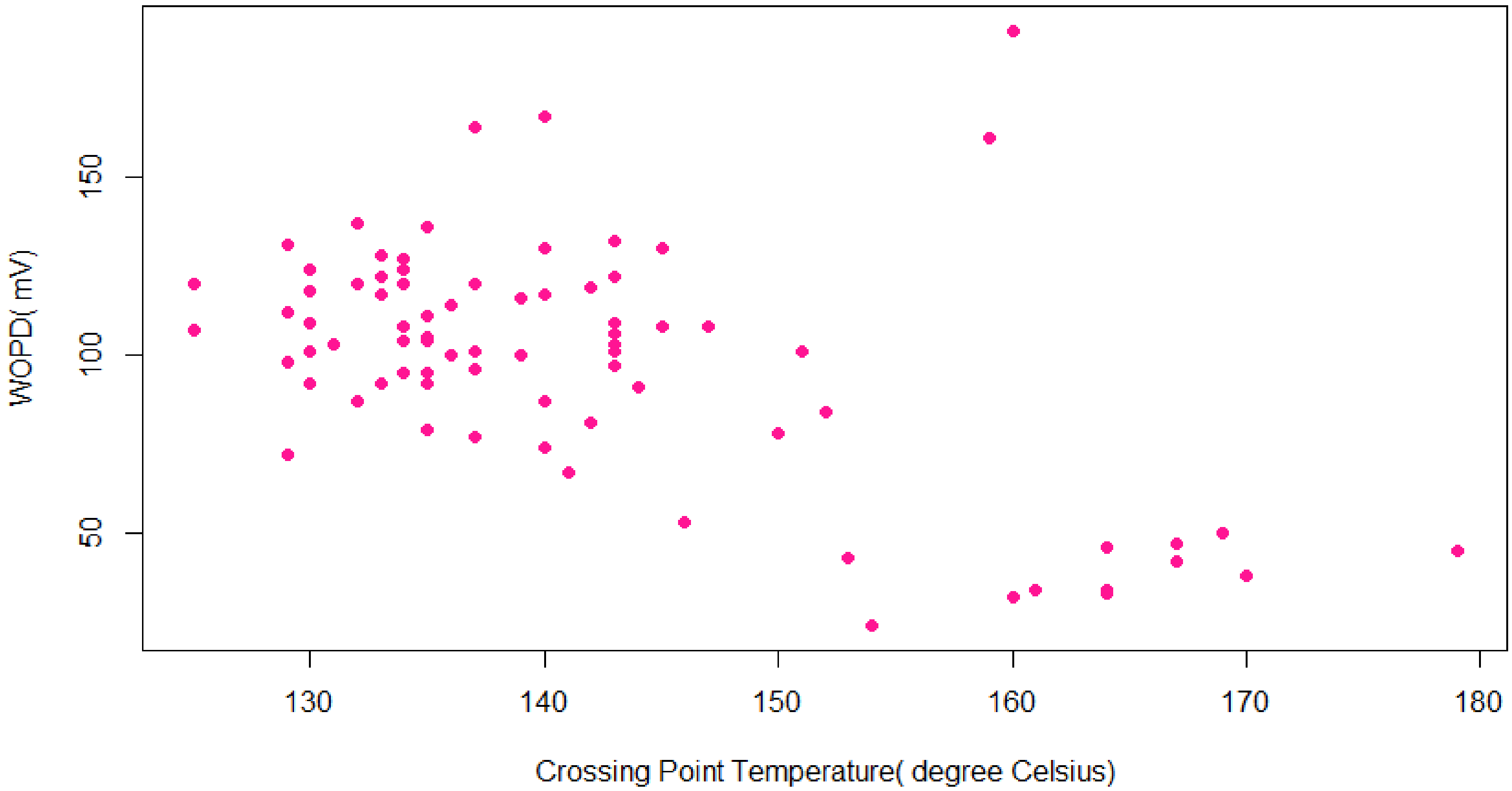
> cor(b,method="pearson")

	CPT	WOPD
CPT	1.0000000	-0.5107481
WOPD	-0.5107481	1.0000000

So, this implies that CPT and WOPD at PD1 have a moderate negative correlation.

The scatter plot between CPT and WOPD at PD2 is:

> plot(bcpt,bpd7,xlab="Crossing Point Temperature(degree Celsius)",ylab="WOPD (mV)", col="deep pink",pch=19,cex=1)



The correlation between them is found to be -0.550537.

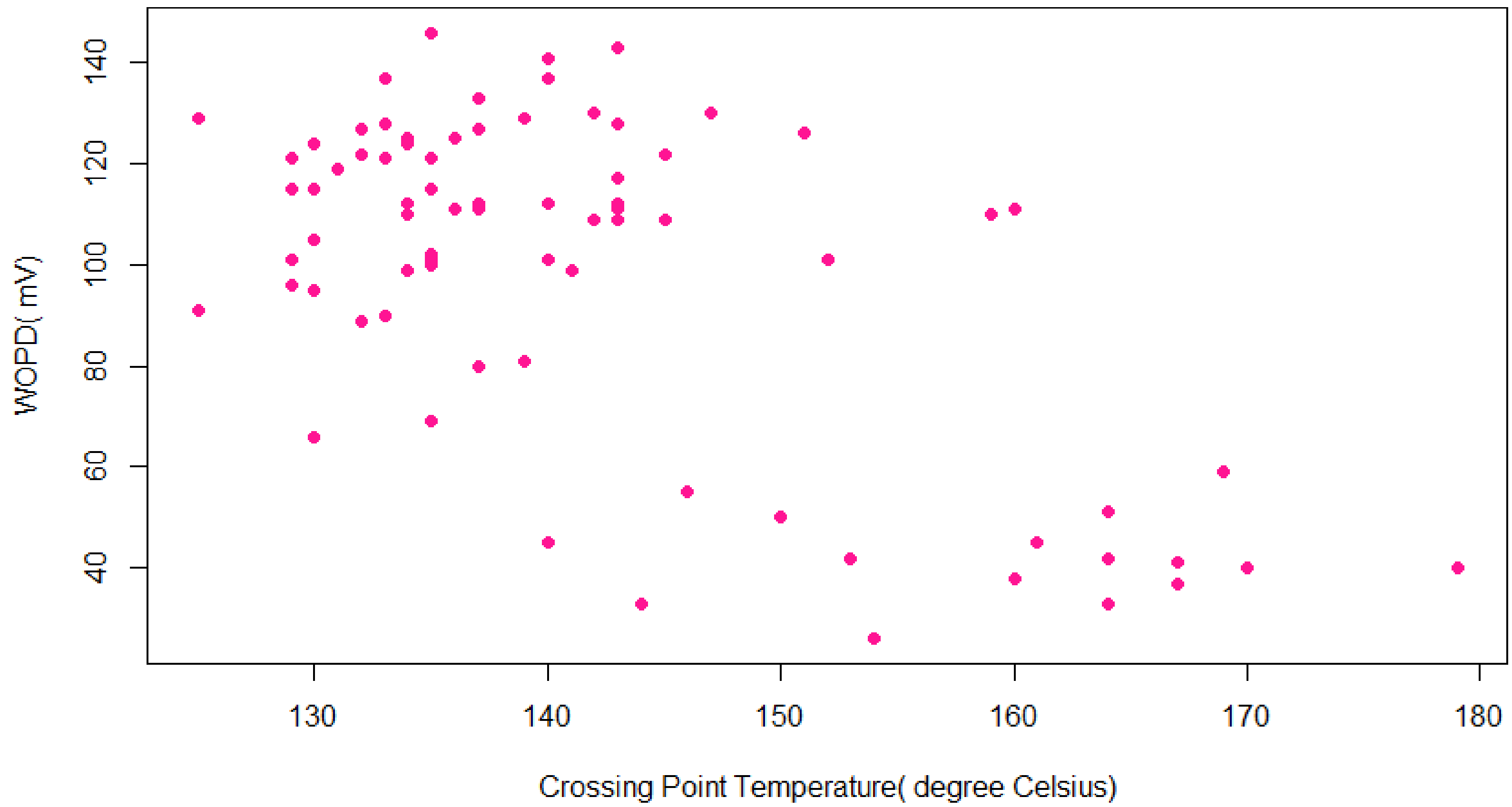
> cor(c,method="pearson")

	CPT	WOPD
CPT	1.000000	**-0.550537**
WOPD	**-0.550537**	1.000000

So, this implies that CPT and WOPD at PD2 have a moderate negative correlation, but stronger than the previous one.

The scatter plot between CPT and WOPD at PD3 is:

```
> plot(d$cpt,d$pd3,xlab="Crossing Point Temperature( degree Celsius)",ylab="WOPD( mV)",
col="deep pink",pch=19,cex=1)
```



Here also we see a stronger negative relation between CPT and WOPD.

The correlation between them is found to be -0.6336818.

```
> cor(d,method="pearson")
            CPT         WOPD
CPT     1.0000000   -0.6336818
WOPD   -0.6336818    1.0000000
```

So, this implies that CPT and WOPD at PD3 have a strong negative correlation, but stronger than both the previous ones.

Hence for all the cases, we conclude that as the crossing point temperature increases, Wet Oxidation Potential Difference decreases implying a negative correlation between them. Crossing Point Temperature(CPT) is most strongly correlated with WOPD at PD3.

CONCLUSION

From the above results we conclude the following:

- Fixed carbon, Carbon and Intertinite have moderate positive correlation with crossing point temperature (CPT) . So, lower the fixed carbon, carbon and intertinite content, lower is the CPT and higher is its susceptibility to spontaneous combustion.
- Ash Yield, Sulfur and Liptinite are found to be uncorrelated with CPT.
- Moisture and Nitrogen has a weak negative correlation with CPT.
- Hydrogen, Oxygen, and Vitrinite have moderate negative correlation with CPT. Volatile Matter Yield have strong negative correlation with CPT. So, higher their content ,lower is the CPT and higher is its susceptibility to spontaneous combustion.
- By applying PCA, the dataset of 12 variables is reduced to a dataset of 4 principal components which is capable of explaining of about 82.18% variability of the dataset. The first principal component can be called as a measure of volatile matter yield and carbon, the second principal component as a measure of ash yield and sulfur, the third principal component as a measure of liptinite, the fourth principal component is as a measure of nitrogen and vitrinite.
- The regression equation of CPT as the dependent variable and the first 4 principal components as the predictor variables, will be incorrect for prediction and forecasting purposes because of the violation of assumptions like normality and homoscedasticity.
- There is a significant difference between the effect of PD1 and PD2 and between PD1 and PD3 on the Wet Oxidation Potential Difference. But, there is no difference between the effects of PD2 and PD3 on Wet Oxidation Potential Difference.
- Crossing Point Temperature is most strongly correlated with WOPD measured at 3rd condition PD3. There exists a simple relationship between WOPD and CPT. As the WOPD increases, CPT decreases which increases its susceptibility to spontaneous combustion.

So, to decrease the susceptibility of coal to spontaneous combustion, crossing point temperature should be high, and for that, its Wet Oxidation Potential Difference should be low, Fixed Carbon, Carbon and Intertinite content should be high and Volatile matter yield, Hydrogen, Oxygen, Vitrinite content should be low.

REFERENCE

1.Prediction of spontaneous heating susceptibility of Indian coals using fuzzy logic and artificial neural network models – H.B.Sahu , S.Padhee, S.S.Mahapatra

2.Applied Multivariate Statistical Analysis – Johnson and Wichern

3.Fundamentals of Statistics (Volume One) – A.M.Gun , M.K.Gupta , B.Dasgupta

4. Practical Guide to Principal Component Methods in R - Alboukadel KASSAMBARA

5. https://rpubs.com/esobolewska/pcr-step-by-step

6. https://www.datacamp.com/community/tutorials/pca-analysis-r

8. https://towardsdatascience.com/principal-component-analysis-pca-101-using-r-361f4c53a9ff

9. https://www.analyticsvidhya.com/blog/2016/03/pca-practical-guide-principal-component-analysis-python/

ACKNOWLEDGEMENT

I would like to express my thanks of gratitude to Professor Ipshita Samanta , Dr . Moutushi Chatterjee and Dr. Bratati Chakraborty of the Department of Statistics of Lady Brabourne College for helping me and guiding me throughout the project work. I would also like to thank my friends and family for their constant support and encouragement.

APPENDIX

Dataset:

sample	moisture (%)	ash.yld (%)	vm.yld (%)	fxd.crbn (%)	carbon (%)	hydrogen (%)	nitrogen (%)	sulfur (%)	oxygen (%)	vitrinite (%)	intertinite (%)	liptinite (%)	CPT (degree Celsius)	WOPD at PD1 (mV)	WOPD at PD2 (mV)	WOPD at PD3 (mV)
1	5.66	13.47	40.08	48.46	80.81	5.43	1.88	1.04	10.84	65.27	33.76	0.97	137	115	120	127
2	6.27	9.34	42.47	48.55	80.46	5.94	1.84	1	10.77	66.88	33.12	0	134	112	127	125
3	9.74	7.72	39.11	50.26	78.7	5.56	1.96	0.91	12.87	74.31	25.69	0	135	93	136	101
4	0.86	25.29	28.98	52.46	86.08	5.01	2.06	1.11	1.16	54.52	43.32	2.17	140	36	87	45
5	1.15	14.93	29.68	59.01	85.72	5.12	1.97	0.8	6.39	62.59	36.93	0.48	150	35	78	50
6	7.79	13.99	40.55	46.5	78.53	5.82	2.07	0.74	12.84	66.67	27.05	6.28	143	71	122	109
7	8.85	14.26	39.15	46.79	78.36	5.81	2.05	0.72	13.06	77.16	15.07	7.77	135	72	92	100
8	8.73	29.36	42.45	35.63	71.98	6.67	2.49	0.81	18.06	56.92	40.31	2.78	152	104	84	101
9	4.82	36.89	40.66	34.59	76.24	6.26	2.35	0.55	14.6	71.32	24.72	3.96	129	42	112	101
10	5.13	12.07	41.14	48.74	81.64	5.79	1.75	0.57	10.25	69.88	26.22	3.9	129	138	98	96
11	4.8	17.28	42.76	44.6	79.76	5.97	2.13	0.87	11.27	72.83	22.77	4.4	135	132	95	146
12	5.68	17.35	42.67	44.13	79.99	5.87	2.14	0.81	11.19	70.99	21.33	7.68	145	114	130	122
13	1.27	23.61	30.4	52.28	85.46	5.43	1.98	1.03	6.11	58.76	40.13	1.11	146	52	53	55
14	0.73	16.28	23.41	63.56	88.78	4.78	2.07	0.98	3.39	54.44	44.52	1.04	164	37	33	33
15	1.4	18.1	31.04	55.51	85.84	5.14	2	1.01	6.01	66.36	33.64	0	141	64	67	99
16	1.48	5.99	29.14	65.57	87.61	5.18	1.71	0.82	4.68	59.05	39.96	0.99	154	57	24	26
17	0.45	11.21	20.44	70.28	89.27	5.04	1.45	0.63	3.61	31.1	66.2	2.71	167	57	42	37
18	0.4	23.99	24.18	57.33	86.38	5.12	1.59	0.56	6.36	22.63	77.37	0	179	42	45	40
19	0.43	32.92	21.17	52.54	87.44	4.73	1.85	0.68	5.31	21.51	78.49	0	160	53	32	38
20	0.49	21.69	28.69	55.49	86.92	5.28	1.77	0.91	5.11	62.12	29.7	8.19	164	49	34	42
21	0.82	21.64	22.8	59.86	88.25	4.81	1.84	0.86	4.23	54.99	38.96	6.05	153	61	43	42
22	0.5	14.5	18.62	69.17	89.56	5.04	1.52	0.52	3.36	59.03	37.35	3.62	167	38	47	41
23	2.38	24.49	34.9	47.61	79.94	6.02	1.64	0.46	11.94	50.59	47.65	1.76	129	100	131	121
24	2.99	20.05	38.45	47.37	80.02	6.17	1.82	0.58	11.41	59.53	38.69	1.78	132	93	120	122
25	2.05	20.59	41.53	45.23	80.93	6.93	2.37	0.94	8.83	75.16	21.38	3.45	133	86	128	128
26	6.66	17.29	34.41	49.88	76.42	6.02	1.66	0.32	15.58	81.48	14.2	4.32	134	92	124	124
27	5.18	17.9	44.23	42.9	73.79	6.29	1.4	0.85	17.67	68.71	27.61	3.68	130	94	109	124
28	0.92	8.21	36.05	58.11	83.69	6.03	1.86	0.62	7.8	68.75	28.42	2.84	135	99	104	102
29	10.98	25.64	46.53	33.89	77.82	7.81	1.72	1.09	11.57	62.32	32.96	4.72	143	99	106	112
30	6.64	23.95	47.18	36.66	78.46	6.6	1.51	0.62	12.81	62.97	31.9	5.13	131	82	103	119
31	7.91	31.11	48.25	31.56	77.35	6.84	1.44	0.64	13.73	65.86	30.34	3.8	143	72	97	128
32	6.83	16.71	44.56	42.39	79.05	6.98	1.94	0.78	11.25	63.43	31.65	4.92	137	67	96	111
33	7.88	38.46	47.93	27.94	71.73	6.9	1.4	0.86	19.12	59.88	36.41	3.71	143	74	109	117
34	4.77	35.04	48.68	30.89	79.5	7.34	1.98	0.88	10.3	63.77	31.88	4.35	125	94	107	91
35	6.06	11.89	40.1	49.15	79.87	6.53	2.24	0.77	10.59	72.75	26.78	0.47	136	179	100	111
36	5.53	13.26	33.09	54.34	83.04	5.05	1.92	1.13	8.85	55.36	43.07	1.57	151	64	101	126
37	5.61	12.12	38.92	50.25	81.37	5.51	2.02	0.91	10.2	64.37	29.31	6.32	125	104	120	129
38	6.52	13.1	36.12	51.35	82.48	5.16	1.97	0.93	9.46	62.84	35.67	1.49	140	104	117	141
39	2	6.77	36.62	57.82	76.37	8.25	1.58	1.18	12.62	69.86	25.81	4.32	134	100	104	99
40	1.2	11.32	33.78	57.93	79.29	5.73	1.75	1.17	12.07	71.62	23.98	4.4	130	110	101	95
41	5.04	17.6	36.14	49.4	79.68	6.51	1.59	1.85	10.37	66.24	31.3	2.46	133	87	122	121
42	8.22	11.63	39.73	48.31	75.25	6.2	1.42	0.3	16.83	55.71	41.71	2.58	147	73	108	130
43	8.76	16.21	42.42	43.2	77.97	7.3	1.71	0.6	12.42	58.17	37.5	4.33	142	100	119	130
44	6.51	35.83	46.79	30.68	75.53	6.16	2.27	1.02	15.02	64.95	26.44	8.61	133	87	117	137
45	9.44	5.87	46.35	45.44	73.07	7.18	1.39	0.43	17.94	68.32	25.7	5.98	142	57	81	109
46	8.04	10.46	37.64	50.82	82.1	6.26	1.72	0.43	9.5	64.88	32.54	2.59	140	101	167	137
47	12.45	12.48	44.45	41.7	81.51	7.34	2.14	0.37	8.63	60.41	33.88	5.71	143	96	132	143
48	12.63	10.47	40.78	45.54	79.78	6.84	2.04	0.36	10.98	67.43	29.91	2.66	140	82	130	112
49	10.21	9.01	41.07	47.6	77.93	6.55	1.57	0.78	13.17	58.25	37.35	4.4	139	101	116	129
50	14.85	9.07	47.34	40.06	79.35	7.33	1.96	0.93	10.42	66.36	30.06	3.59	143	66	103	111
51	13.36	17.6	43.74	38.84	78.98	6.89	1.46	0.56	12.09	67.01	30.1	2.89	140	92	74	101
52	13.64	7.45	45.62	42.91	84.59	7.58	2	0.51	5.32	56.55	40.91	2.54	143	88	101	112
53	7.62	18.23	40.93	43.8	80.46	6.81	1.25	0.46	11.02	65.13	28.26	6.61	133	95	92	90
54	10.24	14.82	39.83	45.09	82.13	6.87	1.44	0.52	9.03	70.67	28.42	0.91	137	79	77	80
55	9.88	17.87	46.93	38.34	75.83	6.73	1.54	0.64	15.27	72.06	27.94	0	134	92	95	112
56	10.33	8.17	42.42	46.93	78.96	6.75	1.69	0.42	12.18	71.27	28.73	0	132	98	87	89
57	1.7	1.34	45.34	53	70.02	6.92	1.06	7.18	14.82	81.86	13.72	4.42	145	110	108	109
58	0.7	2.03	46	52.53	81.74	6.79	1.21	0.57	9.69	73.27	22.67	4.06	135	96	79	69
59	1.03	4.01	48.52	48.89	76.25	6.77	1.16	3.69	12.13	66.71	26.7	6.59	137	96	101	112
60	1.07	1.52	47.53	51.11	77.28	6.4	0.95	5.04	10.33	84.79	8.69	6.52	132	118	137	127
61	1.02	0.88	42.22	56.68	80.1	6.61	1.33	1.4	10.57	65.02	15.68	19.3	134	101	108	110
62	0.56	0.71	45.85	53.46	77.31	6.43	1.17	5.07	10.01	81.57	10.53	7.9	139	87	100	81
63	6.69	18.45	42.27	43.22	79.84	5.38	2.26	0.81	11.7	76.92	19.85	3.23	136	101	114	125
64	4.71	20.75	39.99	44.73	80.47	6.04	1.65	0.71	11.13	67.36	30.7	1.94	135	97	105	121
65	3	26.38	46.02	38.12	81.38	6.63	1.64	0.69	9.66	62.6	35.4	1.99	130	95	124	105
66	4.88	29.8	34.28	42.93	80.37	5.16	2.54	1.26	10.67	42.68	23.56	33.76	135	63	111	115
67	3.9	16.08	47.96	41.64	75.54	6.84	1.54	1.09	15	71.03	23.83	5.14	134	86	120	125
68	3.42	20.89	45.37	41.35	75.89	7.2	1.61	0.74	14.56	81.48	14.2	4.32	130	89	118	115
69	0.75	8.99	24.13	68.48	87.72	5.09	1.5	0.88	4.82	50.02	48.32	1.67	161	48	34	45
70	1.15	14.03	23.83	64.61	88.6	4.93	1.63	0.81	4.03	51.97	41.93	3.76	164	46	46	51

71	0.49	32.3	23.4	51.48	86.56	5.03	1.65	0.61	6.14	64.77	22.28	12.95	170	39	38	40
72	0.66	12.85	26.62	63.47	88.31	5.12	1.49	0.65	4.43	43.84	42.55	9.01	169	47	50	59
73	4.82	13.74	34.34	53.47	82.99	5.28	1.89	1.08	8.75	65.03	34.97	0	137	134	164	133
74	5.57	21.83	36.21	46.31	81.89	5.21	2.38	1.14	9.38	72.41	19.42	8.17	129	79	72	115
75	1.29	16.35	34.32	54.09	83.5	5.28	1.64	0.67	8.91	66.96	30.16	2.88	144	40	91	33
76	0.84	12.26	31.15	59.83	87.02	5.13	1.7	0.79	5.35	71.99	21.95	4.91	130	58	92	66
77	15.71	2.98	55.22	36.4	74.42	6.37	1.69	0.85	16.68	82.18	13.95	3.87	159	128	161	110
78	11.1	10.97	44.26	43.44	73.04	6.25	1.62	0.8	18.3	80.79	17.58	1.63	160	118	191	111

Abbreviations :

- ash.yld – Ash yield
- vm.yld – Volatile matter yield
- fxd.crbn – Fixed Carbon
- CPT – Crossing Point Temperature
- WOPD – Wet Oxidation Potential Difference
- WOPD at PD1 – Wet Oxidation Potential Difference measured at $KMn0_4$=0.05N, KOH=1N, Temp=45^0 C
- WOPD at PD2 - Wet Oxidation Potential Difference measured at $KMn0_4$=0.1N, KOH=1N, Temp=45^0 C
- WOPD at PD3 - Wet Oxidation Potential Difference measured at $KMn0_4$=0.15N, KOH=1N, Temp=45^0 C
- PC1 = Dim1 = 1st principal component
- PC2 = Dim2 = 2nd principal component
- PC3 = Dim3 = 3rd principal component
- PC4 = Dim4 = 4th principal component

Source of dataset : https://openmv.net/