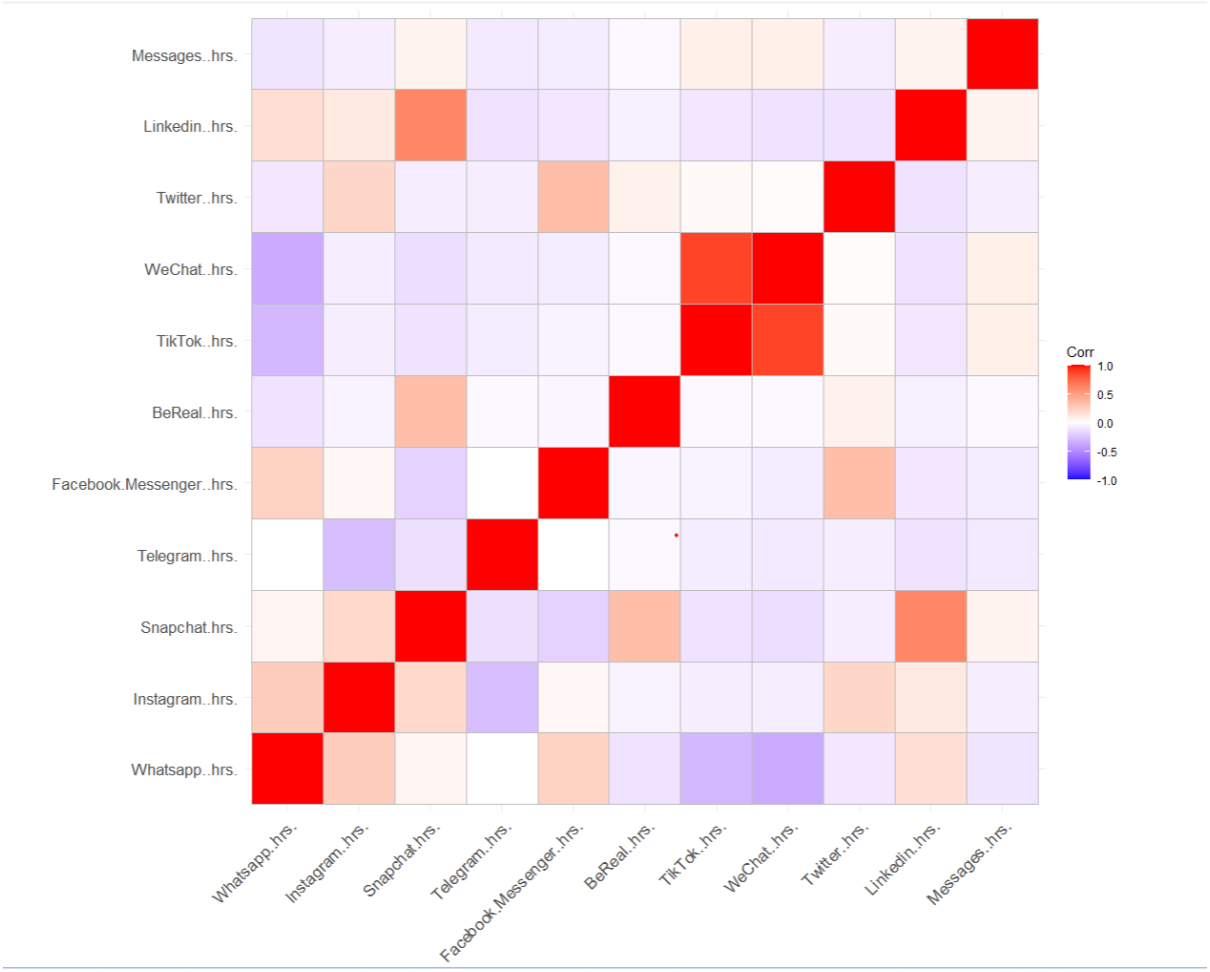
Social Media Usage of the Class

The dataset includes information on the amount of time (in hours) that individuals spend on different social media platforms weekly.

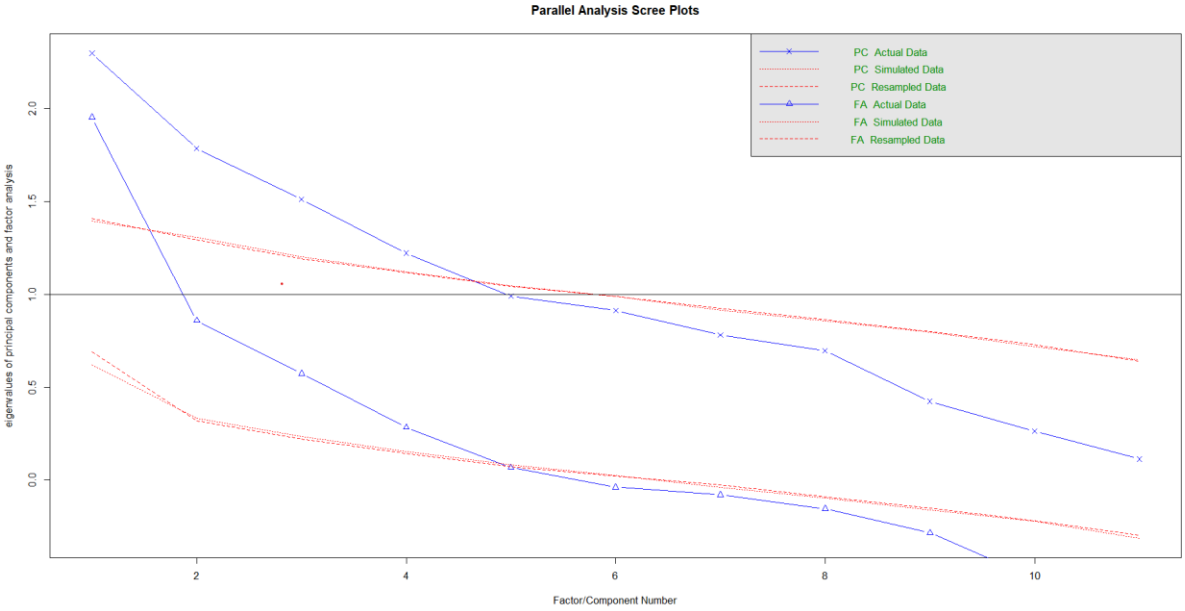
1. the relationship between the different social media apps in terms of usage time?
2. identifying clusters of social media that are similar?
3. Are there any underlying factors that drive social media usage?

Correlation Matrix

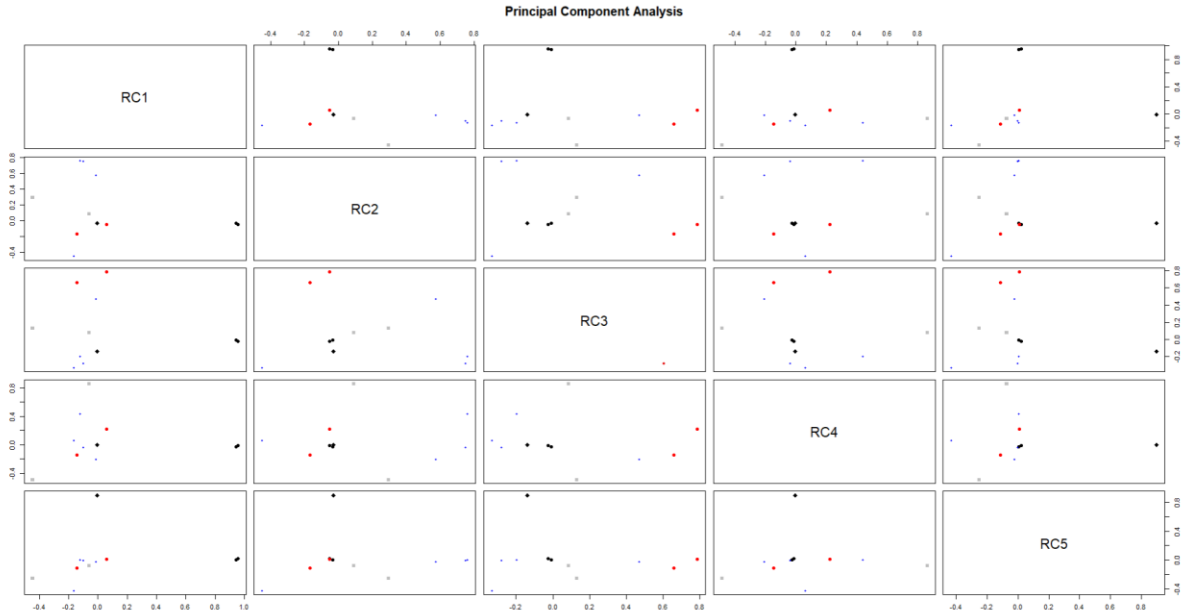


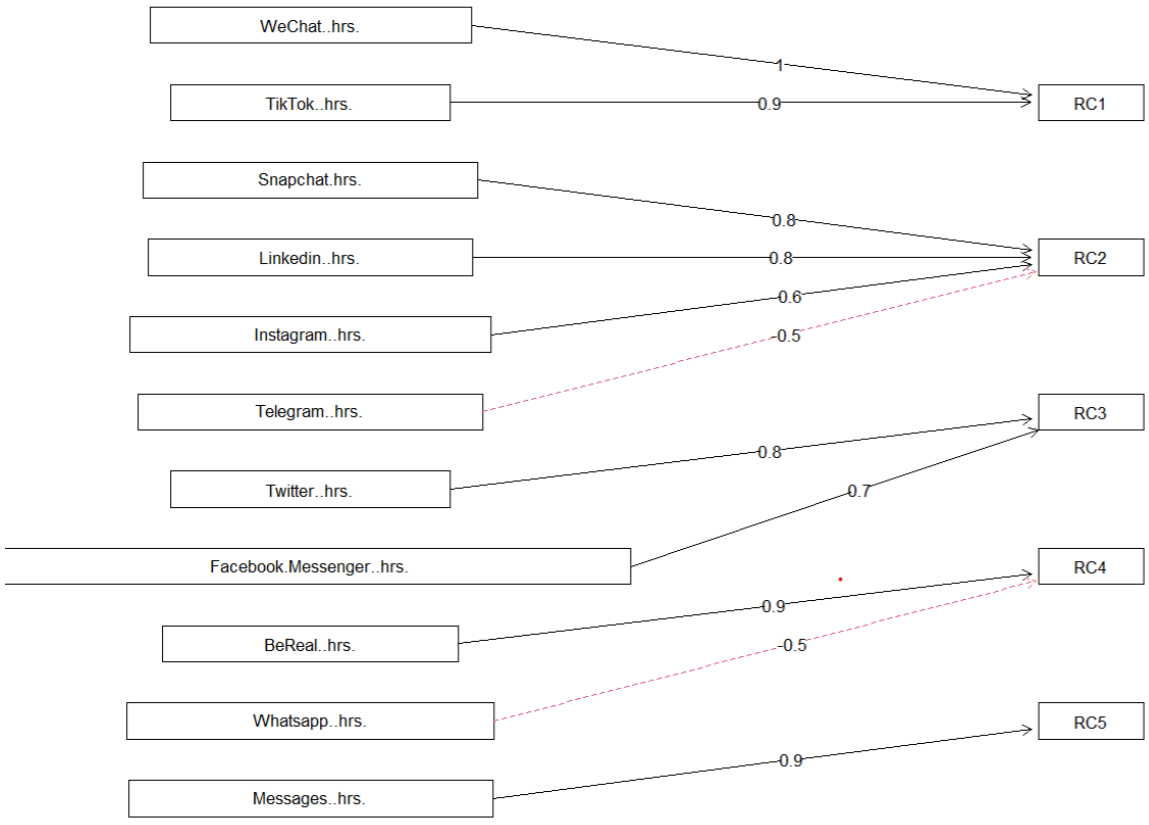
WeChat and Tiktok usage time(in hrs) are highly correlated which suggests that these variables may be measuring similar constructs or concepts.

Exploratory Factor Analysis




```
> fit.pc <- principal(class_data_analysis, nfactors=5, rotate="varimax")
> fit.pc
Principal Components Analysis
Call: principal(r = class_data_analysis, nfactors = 5, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
                          RC1   RC2   RC3   RC4   RC5   h2    u2 com
Whatsapp..hrs.          -0.45  0.30  0.13 -0.49 -0.25 0.61 0.392 3.4
Instagram..hrs.         -0.01  0.58  0.47 -0.21 -0.03 0.60 0.405 2.2
Snapchat.hrs.           -0.12  0.76 -0.20  0.44  0.00 0.83 0.173 1.8
Telegram..hrs.          -0.17 -0.45 -0.33  0.06 -0.43 0.53 0.468 3.2
Facebook.Messenger..hrs. -0.14 -0.17  0.66 -0.15 -0.12 0.52 0.483 1.4
BeReal..hrs.            -0.06  0.09  0.08  0.86 -0.08 0.77 0.234 1.1
TikTok..hrs.             0.94 -0.03 -0.01 -0.03  0.00 0.89 0.108 1.0
WeChat..hrs.             0.95 -0.05 -0.02 -0.01  0.02 0.91 0.085 1.0
Twitter..hrs.            0.06 -0.05  0.78  0.22  0.01 0.67 0.328 1.2
Linkedin..hrs.          -0.10  0.75 -0.28 -0.04  0.00 0.65 0.346 1.3
Messages..hrs.          -0.01 -0.03 -0.14 -0.01  0.90 0.83 0.169 1.0

                       RC1  RC2  RC3  RC4  RC5
SS loadings           2.08 1.81 1.54 1.29 1.08
Proportion Var        0.19 0.16 0.14 0.12 0.10
Cumulative Var        0.19 0.35 0.49 0.61 0.71
Proportion Explained  0.27 0.23 0.20 0.17 0.14
Cumulative Proportion 0.27 0.50 0.70 0.86 1.00

Mean item complexity =  1.7
Test of the hypothesis that 5 components are sufficient.

The root mean square of the residuals (RMSR) is  0.1
 with the empirical chi square  185.49  with prob <  1.7e-34

Fit based upon off diagonal values = 0.76
>
```


Principal Component Analysis
RC1
RC2
RC3
RC4
RC5


WeChat..hrs.
TikTok..hrs.
Snapchat.hrs.
Linkedin..hrs.
Instagram..hrs.
Telegram..hrs.
Twitter..hrs.
Facebook.Messenger..hrs.
BeReal..hrs.
Whatsapp..hrs.
Messages..hrs.
RC1
RC2
RC3
RC4
RC5
1
0.9
0.8
0.8
0.6
0.5
0.8
0.7
0.9
0.5
0.9

```
> fit.pc$loadings

Loadings:
                        RC1    RC2    RC3    RC4    RC5
Whatsapp..hrs.          -0.449  0.295  0.129 -0.488 -0.253
Instagram..hrs.                 0.575  0.470 -0.208
Snapchat.hrs.           -0.125  0.762 -0.198  0.438
Telegram..hrs.          -0.167 -0.451 -0.332        -0.433
Facebook.Messenger..hrs. -0.143 -0.167  0.659 -0.145 -0.115
BeReal..hrs.                                0.861
TikTok..hrs.             0.944
WeChat..hrs.             0.954
Twitter..hrs.                          0.785  0.223
Linkedin..hrs.          -0.102  0.750 -0.281
Messages..hrs.                        -0.138         0.901

                 RC1   RC2   RC3   RC4   RC5
SS loadings    2.084 1.808 1.543 1.292 1.083
Proportion Var 0.189 0.164 0.140 0.117 0.098
Cumulative Var 0.189 0.354 0.494 0.612 0.710
```

Factor 1 includes WeChat and Tiktok these two platforms share similar characteristics or are used for similar purposes. This could suggest that both WeChat and TikTok are social media platforms that are primarily used for communication and socialization purposes.

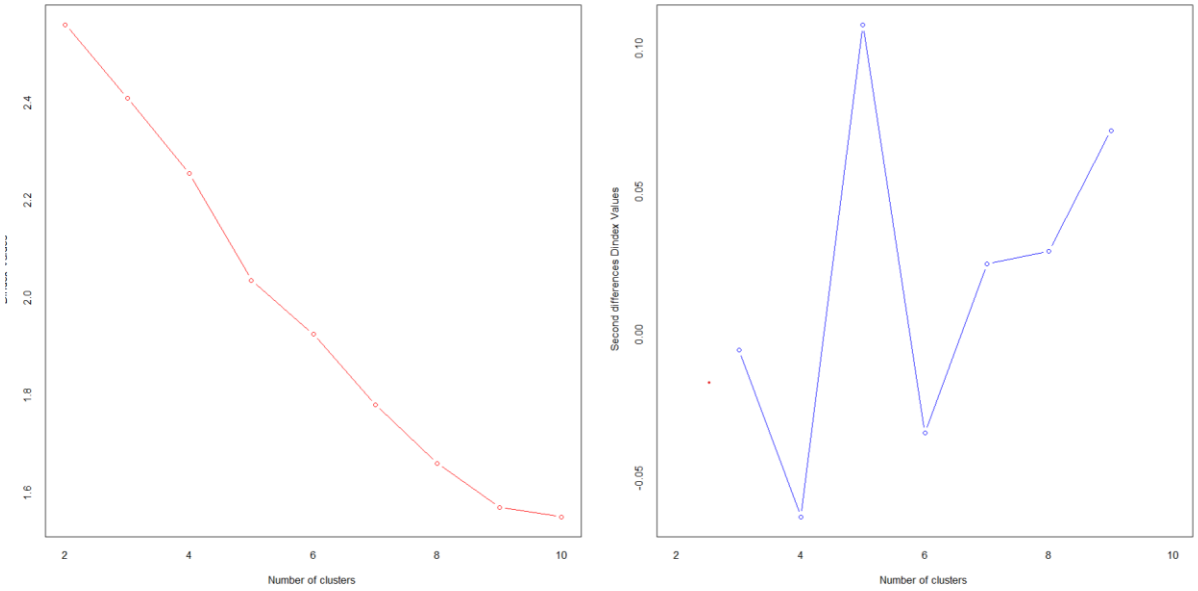
Factor 2 includes Snapchat, Linkedin, Instagram reflects a social networking factor, where these platforms are used primarily for building and maintaining social connections, sharing personal updates and experiences, and staying informed about others' lives.

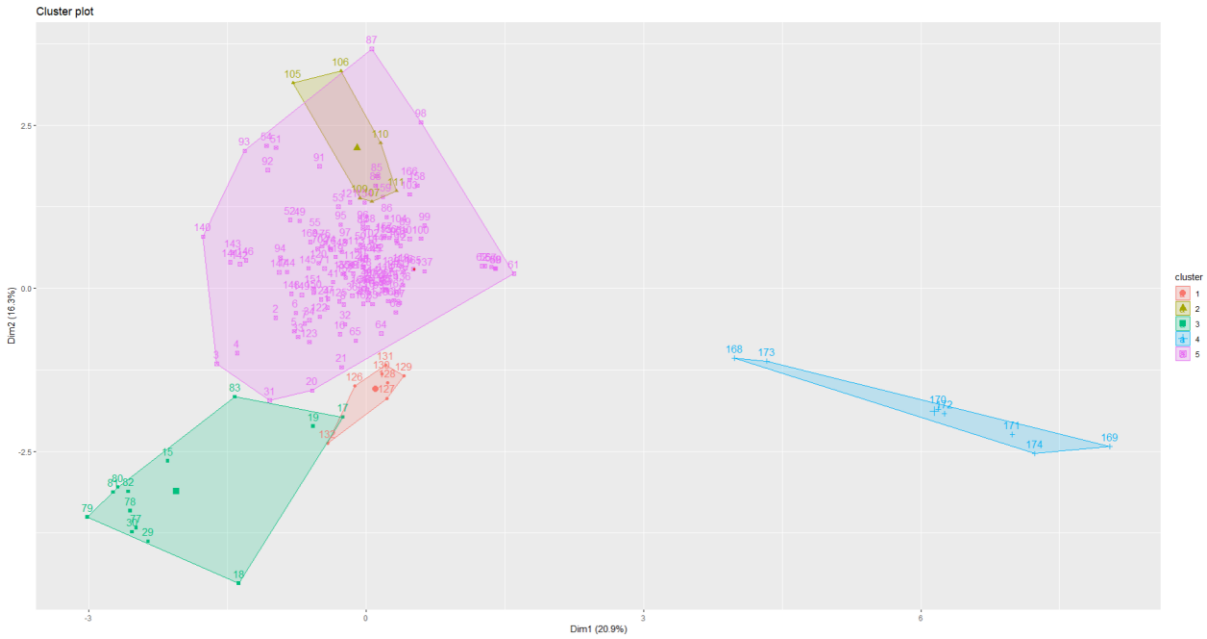
Factor 4 includes BeReal usage time, which is a social media platform

Factor 3 includes Twitter and Facebook, represent the public and real-time nature of information sharing on these platforms, where users can share their thoughts and opinions with a wider audience.

Cluster Analysis

```
> nb <- NbClust(norm_data, distance = "euclidean", min.nc = 2,
+               max.nc = 10, method = "kmeans")
*** : The Hubert index is a graphical method of determining the number of clusters.
               In the plot of Hubert index, we seek a significant knee that corresponds to a
               significant increase of the value of the measure i.e the significant peak in Hubert
               index second differences plot.

*** : The D index is a graphical method of determining the number of clusters.
               In the plot of D index, we seek a significant knee (the significant peak in Dindex
               second differences plot) that corresponds to a significant increase of the value of
               the measure.

*******************************************************************
* Among all indices:
* 4 proposed 2 as the best number of clusters
* 3 proposed 3 as the best number of clusters
* 5 proposed 5 as the best number of clusters
* 2 proposed 7 as the best number of clusters
* 5 proposed 8 as the best number of clusters
* 2 proposed 9 as the best number of clusters
* 2 proposed 10 as the best number of clusters

                   ***** Conclusion *****

* According to the majority rule, the best number of clusters is  5


*******************************************************************
```



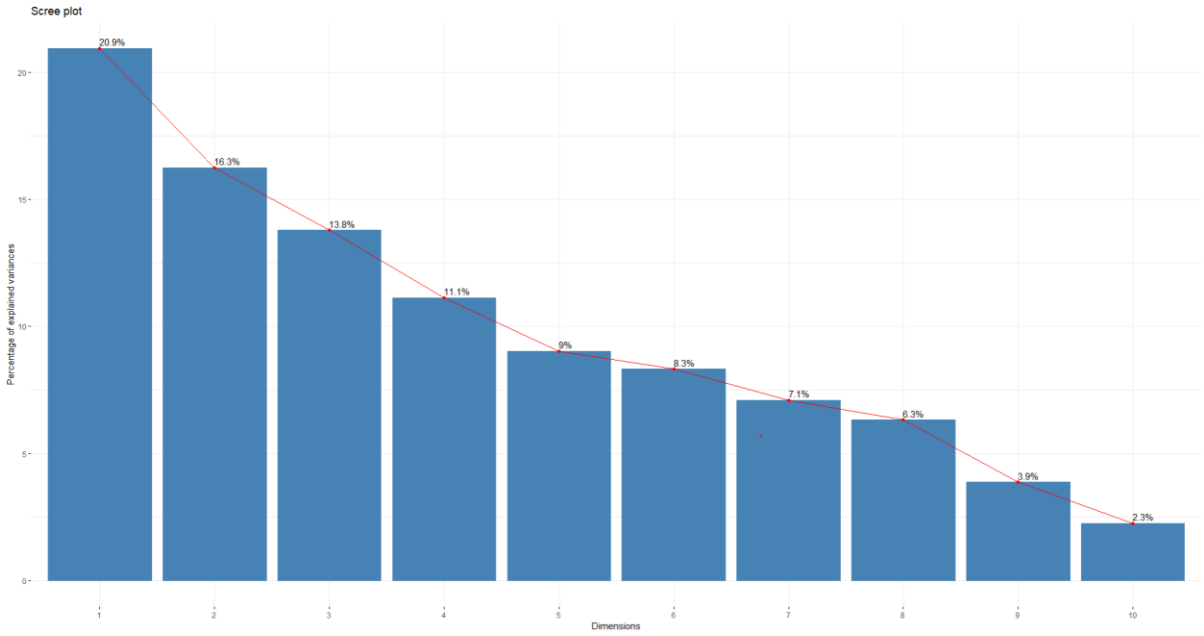


Number of clusters suggested by nbclust is 5

cluster analysis has identified groups of users with similar usage patterns or preferences for social media apps.

PCA

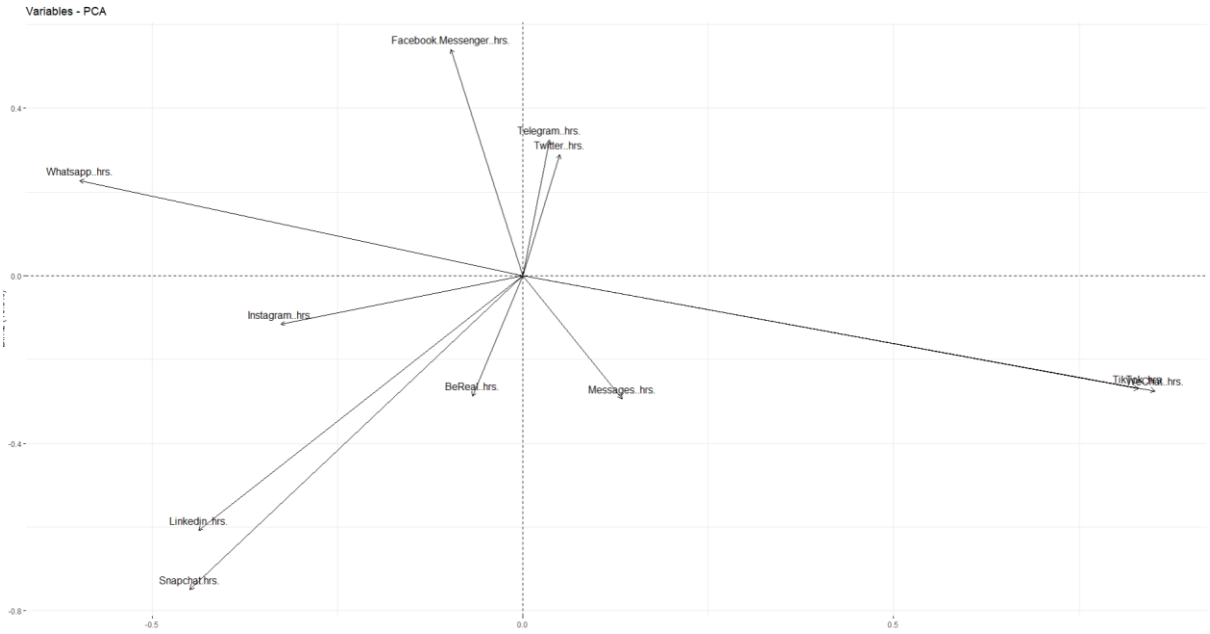
```
> summary(pca_data)
Importance of components:
                          PC1    PC2    PC3    PC4    PC5     PC6    PC7     PC8     PC9    PC10    PC11
Standard deviation     1.5196 1.3388 1.2329 1.1077 0.9974 0.95776 0.8843 0.83471 0.65311 0.49840 0.33764
Proportion of Variance 0.2094 0.1625 0.1378 0.1113 0.0902 0.08318 0.0709 0.06318 0.03868 0.02252 0.01034
Cumulative Proportion  0.2094 0.3719 0.5097 0.6210 0.7112 0.79438 0.8653 0.92846 0.96714 0.98966 1.00000
```

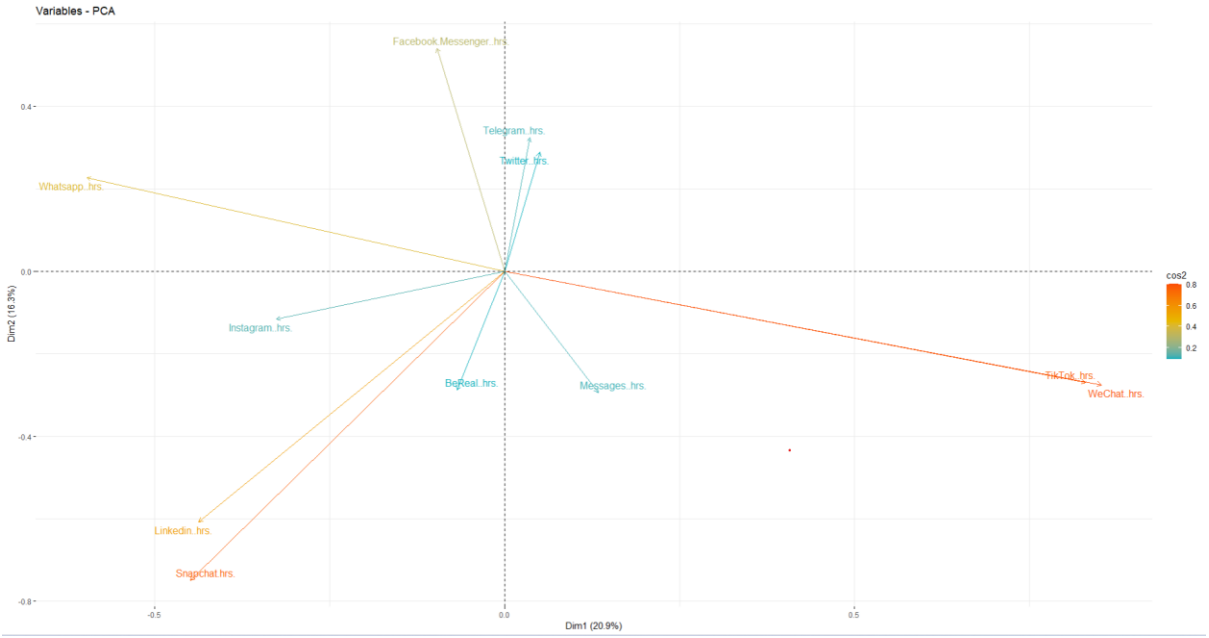


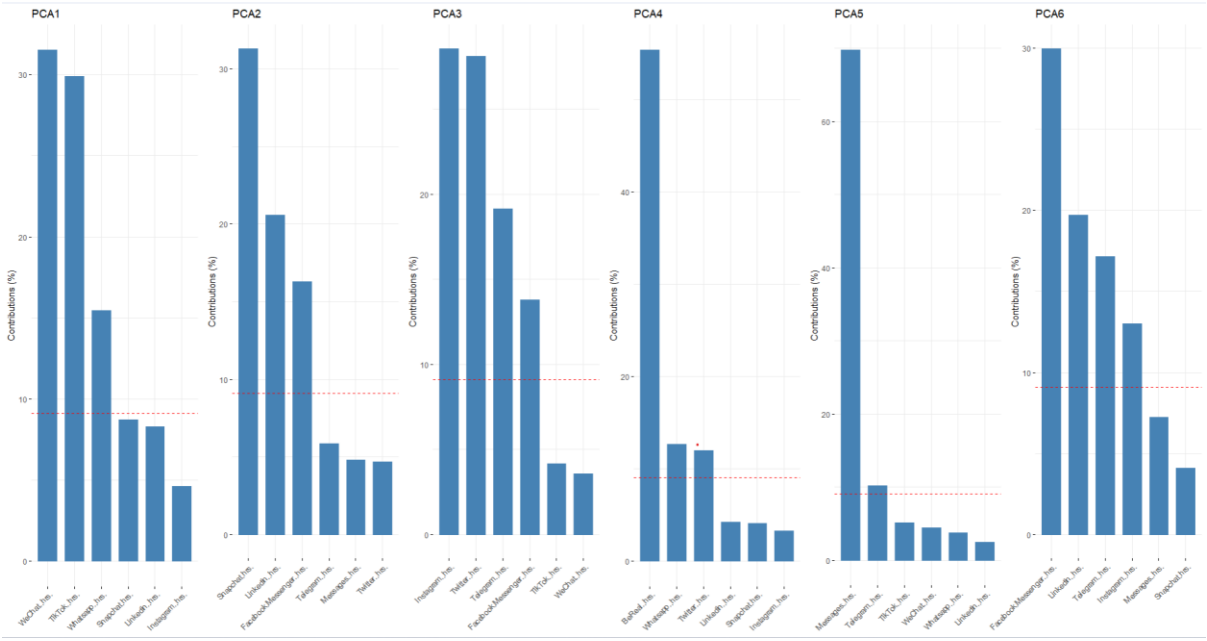
From the scree plot, it can be observed that the first six PCs explain over 80% of the total variance in the data.

```
> pca_data$rotation[,1:6]
                                PC1         PC2         PC3         PC4         PC5         PC6
Whatsapp..hrs.           -0.39305303  0.16963762 -0.10713227 -0.35617893  0.19346133 -0.11462970
Instagram..hrs.          -0.21463239 -0.08628024 -0.53410204 -0.18078722  0.07181896  0.36039675
Snapchat.hrs.            -0.29542184 -0.55938805 -0.07683288  0.20306890  0.12438300 -0.20259326
Telegram..hrs.            0.02362935  0.24183165  0.43726559  0.14915562  0.31875968 -0.41407176
Facebook.Messenger..hrs. -0.06361940  0.40373626 -0.37105087  0.06958413 -0.05459531 -0.54733433
BeReal..hrs.             -0.04485826 -0.21437668 -0.02852030  0.74403381  0.02495200  0.08112231
TikTok..hrs.              0.54662323 -0.20113734 -0.20373946 -0.13184153  0.22579244 -0.15729474
WeChat..hrs.              0.56128938 -0.20524489 -0.18928642 -0.12308446  0.21083294 -0.10878250
Twitter..hrs.             0.03308135  0.21604781 -0.53012448  0.34593567 -0.13928092 -0.17957740
Linkedin..hrs.           -0.28778484 -0.45390192 -0.04506509 -0.20487022  0.15679365 -0.44399781
Messages..hrs.            0.08827763 -0.21939832  0.08520039 -0.15609178 -0.83505477 -0.26900129
```

```
> eigenvalues
     PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9     PC10     PC11
0.20938 0.16252 0.13784 0.11126 0.09020 0.08318 0.07090 0.06318 0.03868 0.02252 0.01034
>
```







Based on the PCA analysis performed on the dataset, several conclusions can be made:

The first six principal components explain over 80% of the total variance in the data, with PC1 contributing the most at 31%.

PC1: The variables that have the highest positive correlation with PC1 are WeChat, Whatsapp, and TikTok, indicating that these platforms have a significant impact on the overall social media usage patterns.

PC2: The variables that contribute most to this component are Snapchat, Linkedin, and Facebook. This component can be interpreted as a measure of the usage of visually-oriented social media platforms.

PC3: The variables that contribute most to this component are Instagram, Twitter, Telegram, and Facebook. This component can be interpreted as a measure of the usage of social media applications to get to know others' social lives.

PC4: The variables that contribute most to this component are BeReal, Whatsapp, and Twitter. This component can be interpreted as a measure of the usage of social media platforms for interaction with others.

PC5: The variable that contributes most to this component is Messages. This component can be interpreted as a measure of the usage of texting others not using the internet.

PC6: The variables that contribute most to this component are Facebook, Linkedin, and Telegram. This component is much very similar to PC3.

Logistic Regression

```
> # Fit logistic regression model
> model <- glm(Social.Media.Addiction ~ ., data = train_data, family = "binomial")
>
> # Print model summary
> summary(model)

Call:
glm(formula = Social.Media.Addiction ~ ., family = "binomial",
    data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8232  -0.7382  -0.3615   0.7988   2.7758

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            1.07513    0.59742   1.800   0.0719 .
Whatsapp..hrs.         0.03415    0.05443   0.627   0.5303
Instagram..hrs.       -0.27820    0.06184  -4.499 6.84e-06 ***
Snapchat.hrs.          0.19167    0.20555   0.932   0.3511
Telegram..hrs.        -0.69611    0.64843  -1.074   0.2830
Facebook.Messenger..hrs.  0.38305    0.62056   0.617   0.5371
BeReal..hrs.           1.51161    1.96360   0.770   0.4414
TikTok..hrs.          -0.65693    1.00763  -0.652   0.5144
WeChat..hrs.           0.42097    0.31673   1.329   0.1838
Twitter..hrs.          0.27563    0.24044   1.146   0.2516
Linkedin..hrs.        -0.09825    0.07616  -1.290   0.1970
Messages..hrs.         0.33349    0.22652   1.472   0.1410
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 189.22  on 139  degrees of freedom
Residual deviance: 139.22  on 128  degrees of freedom
AIC: 163.22

Number of Fisher Scoring iterations: 6

>
```

```
> confusionMatrix(table(test_data$Social.Media.Addiction, test_data$predicted.Social.Media.Addiction))
Confusion Matrix and Statistics


    0  1
  0 15  1
  1  8 10

               Accuracy : 0.7353
                 95% CI : (0.5564, 0.8712)
    No Information Rate : 0.6765
    P-Value [Acc > NIR] : 0.2969

                  Kappa : 0.4814

 Mcnemar's Test P-Value : 0.0455

            Sensitivity : 0.6522
            Specificity : 0.9091
         Pos Pred Value : 0.9375
         Neg Pred Value : 0.5556
             Prevalence : 0.6765
         Detection Rate : 0.4412
   Detection Prevalence : 0.4706
      Balanced Accuracy : 0.7806

       'Positive' Class : 0

>
```

Accuracy is 73% and the AIC score is 163.22