Student Substance Use Survey: Understanding Drug and Alcohol Consumption Patterns

Betis Mohamed Dhia

2024-04-23

Description:

This project investigates drug and alcohol consumption patterns among students using survey data analysis techniques. It includes exploratory data analysis, visualization, and statistical modeling methods to comprehend the prevalence, perceptions, and behaviors associated with substance use. Techniques employed include pie charts, bar plots, correlation matrices, principal component analysis (PCA), multiple correspondence analysis (MCA), and clustering algorithms such as hierarchical clustering and k-means clustering. The findings provide insights into the prevalence of substance use, perceptions of harm, and clustering patterns within the student population, contributing to a better understanding of substance use behaviors and informing potential interventions.

Importing Data

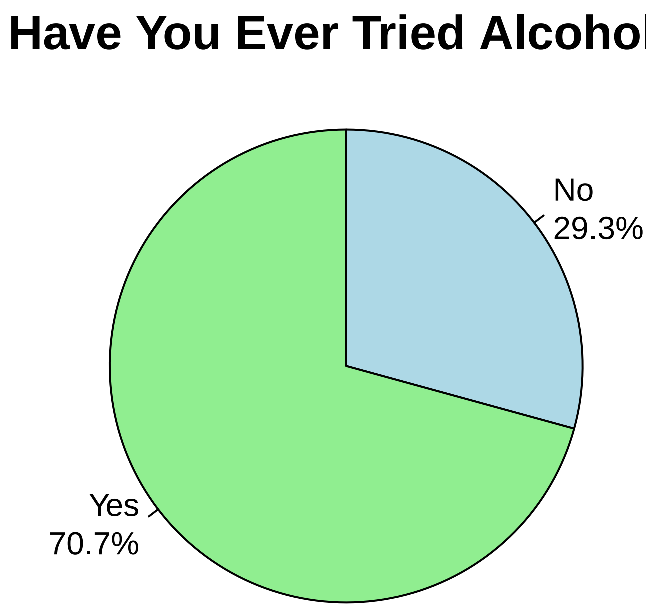
```r
data<-read.csv("C:\\ESSAI\\1ere\\Analyse de donnes\\projet\\Student Substance Use Survey Understanding
```

```r
# Remove rows with any empty values
data <- data[rowSums(data == "" | is.na(data)) == 0, ]
```

Data Visualisation

```r
alcohol_counts <- table(data$Have.you.ever.tried.alcohol.)

# Calculate percentages
percentages <- round(100 * alcohol_counts / sum(alcohol_counts), 1)

# Concatenate labels with percentages
labels_with_percentages <- paste(c("No", "Yes"), "\n", percentages, "%", sep="")

# Create a pie chart with percentages
pie(alcohol_counts,
    main = "Have You Ever Tried Alcohol?",
```

```
    labels = labels_with_percentages,
    col = c("lightblue", "lightgreen"),
    border = "black",
    clockwise = TRUE,
    radius = 1,
    init.angle = 90,
    cex.main = 1.5)
```



The majority of respondents (70.7%) have tried alcohol at least once, indicating that alcohol consumption is prevalent among the surveyed population.

```
# Create a table of counts for the variable "How do you prefer to drink?"
drink_counts <- table(data$How.do.you.prefer.to.drink.)

# Define colors for each option
colors <- c("lightblue", "lightgreen", "lightcoral", "lightyellow", "lightpink")

# Calculate percentages
drink_percentages <- round(100 * drink_counts / sum(drink_counts), 1)

# Create labels with percentages for each category
labels_with_percentages <- paste(names(drink_counts), "\n", drink_percentages, "%", sep="")

# Create a pie chart
pie(drink_counts,
    main = "Preference for Drinking",
```

```
    labels = labels_with_percentages,
    col = colors,
    border = "black",
    clockwise = TRUE,
    radius = 1,
    init.angle = 90,
    cex.main = 1.5)
```

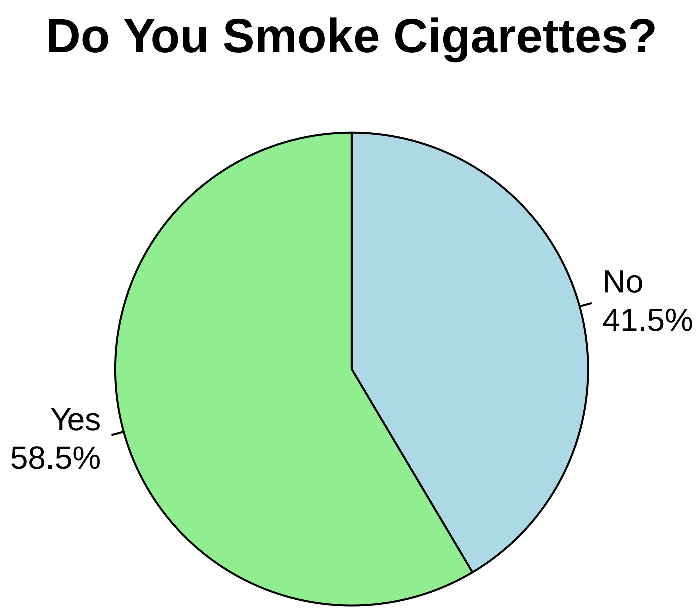
The majority of respondents (34.1%) reported that they don't drink alcohol.

Among those who do drink, a significant portion prefer to drink enough to feel a lot (24.4%), followed by those who prefer enough to feel a little (19.5%) and just a glass or two (14.6%).

A smaller proportion (7.3%) indicated a preference for drinking until they get drunk.

```
# Create a table of counts for the variable "Do you smoke cigarettes?"
smoking_counts <- table(data$Do.you.smoke.cigarettes.)

# Define colors for each option
colors <- c("lightblue", "lightgreen")

# Calculate percentages
smoking_percentages <- round(100 * smoking_counts / sum(smoking_counts), 1)

# Concatenate labels with percentages
labels_with_percentages <- paste(c("No", "Yes"), "\n", smoking_percentages, "%", sep="")
```

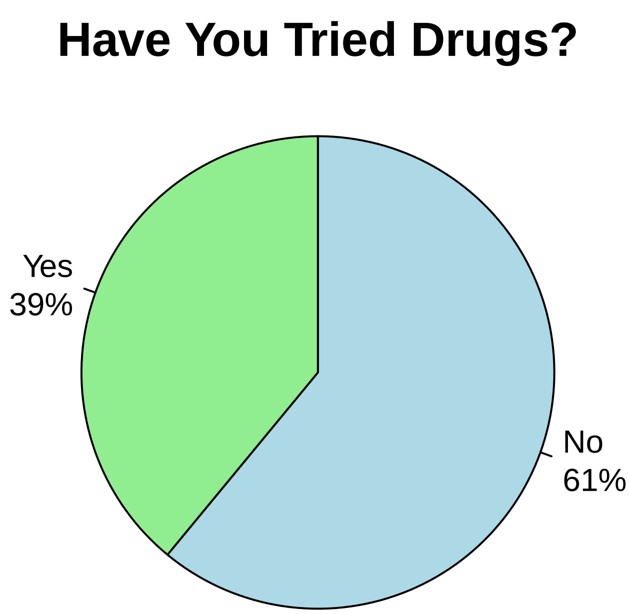
```
# Create a pie chart
pie(smoking_counts,
    main = "Do You Smoke Cigarettes?",
    labels = labels_with_percentages,
    col = colors,
    border = "black",
    clockwise = TRUE,
    radius = 1,
    init.angle = 90,
    cex.main = 1.5)
```



The majority of respondents (58.5%) reported that they smoke cigarettes, indicating a significant prevalence of smoking behavior among the surveyed population.

```
# Create a table of counts for the variable "Have you tried any of the following drugs?"
drug_counts <- table(data$Have.you.tried.any.of.the.following.drugs..Weed..Cocaine..LSD..Heroin..Ecstasy

# Define colors for each option
colors <- c("lightblue", "lightgreen")

# Calculate percentages
drug_percentages <- round(100 * drug_counts / sum(drug_counts), 1)

# Concatenate labels with percentages
labels_with_percentages <- paste(c("No", "Yes"), "\n", drug_percentages, "%", sep="")
```

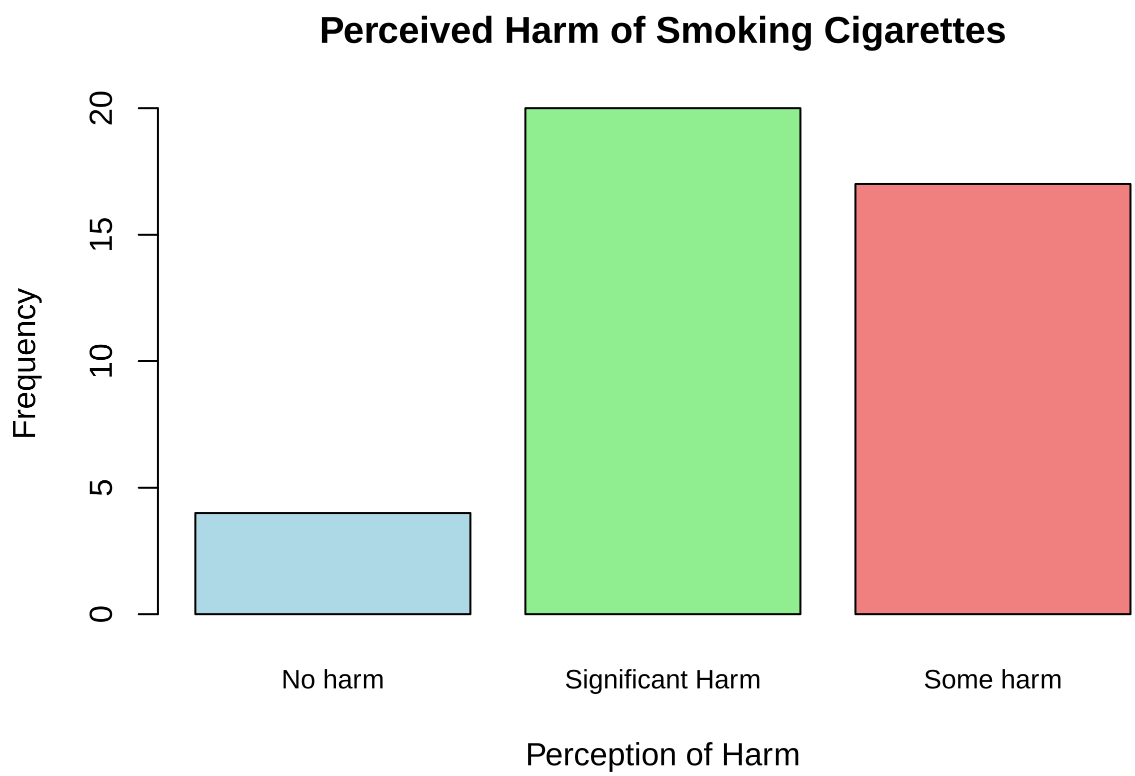
```r
# Create a pie chart
pie(drug_counts,
    main = "Have You Tried Drugs?",
    labels = labels_with_percentages,
    col = colors,
    border = "black",
    clockwise = TRUE,
    radius = 1,
    init.angle = 90,
    cex.main = 1.5)
```



A significant proportion (39%) of respondents admitted to having tried one or more of the listed drugs. However, the majority (61%) stated that they have not experimented with any of the specified drugs.

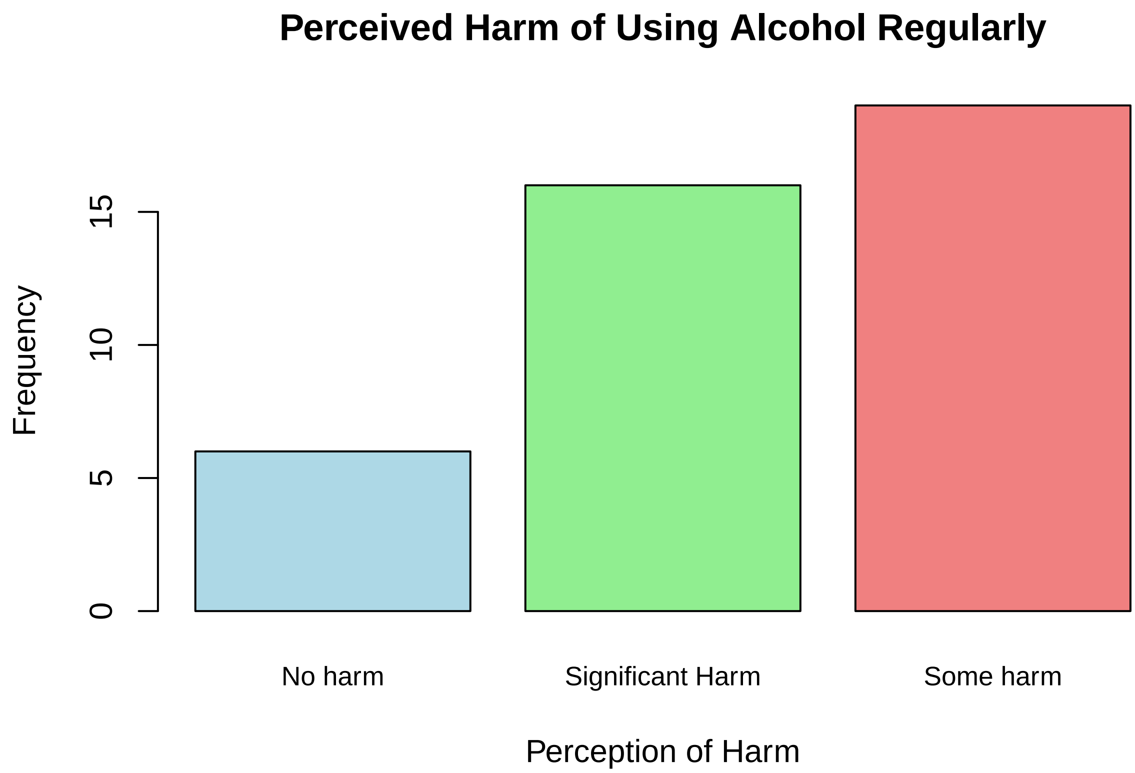
```r
# Create a table of counts for the variable
variable_counts <- table(data$How.harmful.do.you.perceive.regular.use.of.the.following.substances.to.be

# Define colors for each option
colors <- c("lightblue", "lightgreen", "lightcoral")

# Create a bar plot
barplot(variable_counts,
        main = "Perceived Harm of Smoking Cigarettes",
        xlab = "Perception of Harm",
        ylab = "Frequency",
```

```
        col = colors,
        border = "black",
        font.axis = 1,
        cex.names = 0.8)
```



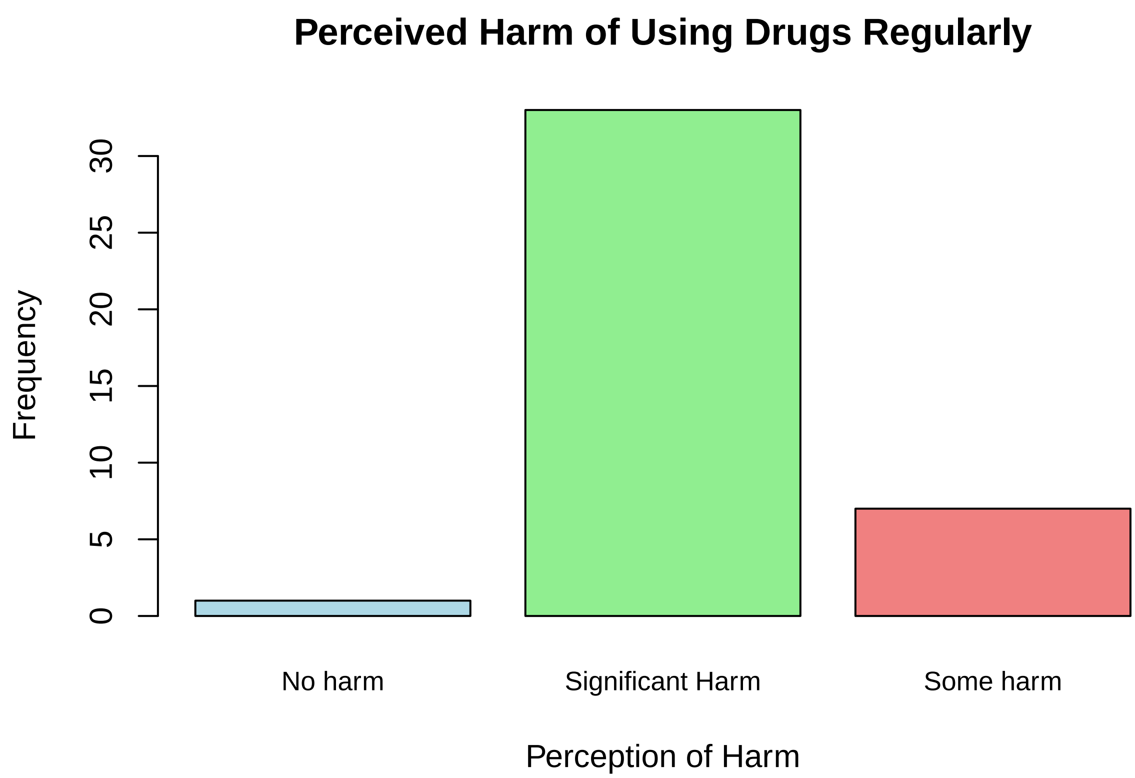
The bar plot shows perceptions of the harm associated with regular alcohol use. The majority perceive "some harm," followed by "significant harm," and a smaller group see "no harm."

```
# Create a table of counts for the variable
variable_counts <- table(data$How.harmful.do.you.perceive.regular.use.of.the.following.substances.to.be

# Create a bar plot
barplot(variable_counts,
        main = "Perceived Harm of Using Alcohol Regularly",
        xlab = "Perception of Harm",
        ylab = "Frequency",
        col = colors,
        border = "black",
        font.axis = 1,
        cex.names = 0.8) #
```



The bar plot illustrates how individuals perceive the harm associated with regular alcohol use. The majority of respondents perceive "some harm," followed by "significant harm," with a smaller group indicating "no harm".

```
# Create a table of counts for the variable
variable_counts <- table(data$How.harmful.do.you.perceive.regular.use.of.the.following.substances.to.be

# Create a bar plot
barplot(variable_counts,
        main = "Perceived Harm of Using Drugs Regularly",
        xlab = "Perception of Harm",
        ylab = "Frequency",
        col = colors,
        border = "black",
        font.axis = 1,
        cex.names = 0.8)
```



The bar plot depicts respondents' perceptions of the harm associated with regular drug use. It shows that the majority of individuals perceive "significant harm," followed by "some harm," with a smaller proportion perceiving "no harm."

Corrolation Matrix

```
subset_data <- data[, 17:28]
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.3.2
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

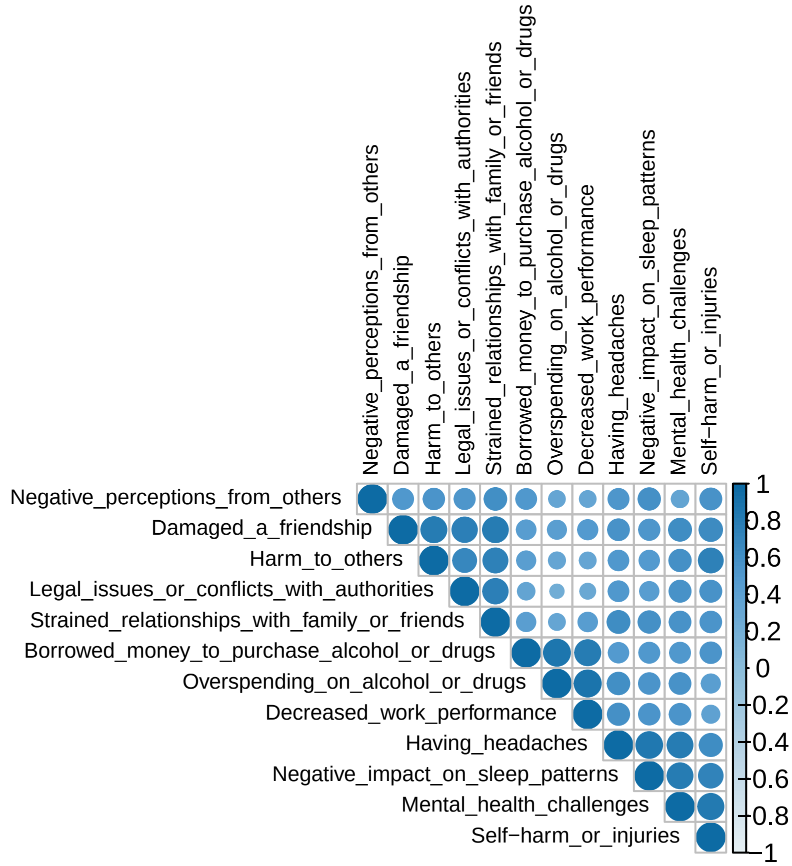
```
colnames(subset_data) <-
  c(
    "Mental_health_challenges",
    "Overspending_on_alcohol_or_drugs",
    "Damaged_a_friendship",
```

```
    "Legal_issues_or_conflicts_with_authorities",
    "Decreased_work_performance",
    "Having_headaches",
    "Strained_relationships_with_family_or_friends",
    "Harm_to_others",
    "Negative_impact_on_sleep_patterns",
    "Self-harm_or_injuries",
    "Negative_perceptions_from_others",
    "Borrowed_money_to_purchase_alcohol_or_drugs"
  )

numeric_data <- subset_data
numeric_data[] <- lapply(subset_data, function(x) {
  factor(
    x,
    levels = c("Never", "Rarely", "Sometimes", "Often", "Always"),
    ordered = TRUE
  )
})
numeric_data[] <- lapply(numeric_data, as.integer)

library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.2
```

```
## corrplot 0.92 loaded
```

```
M <- cor(numeric_data)
library(RColorBrewer)
col <- colorRampPalette(c("#E6F0F3", "#BDD7E7", "#8CBEDB", "#539ACF", "#0D6BA4"))(100)
corrplot(M, type = "upper", order = "hclust", col = col, tl.col = "black", tl.cex = 0.7)
```



PCA

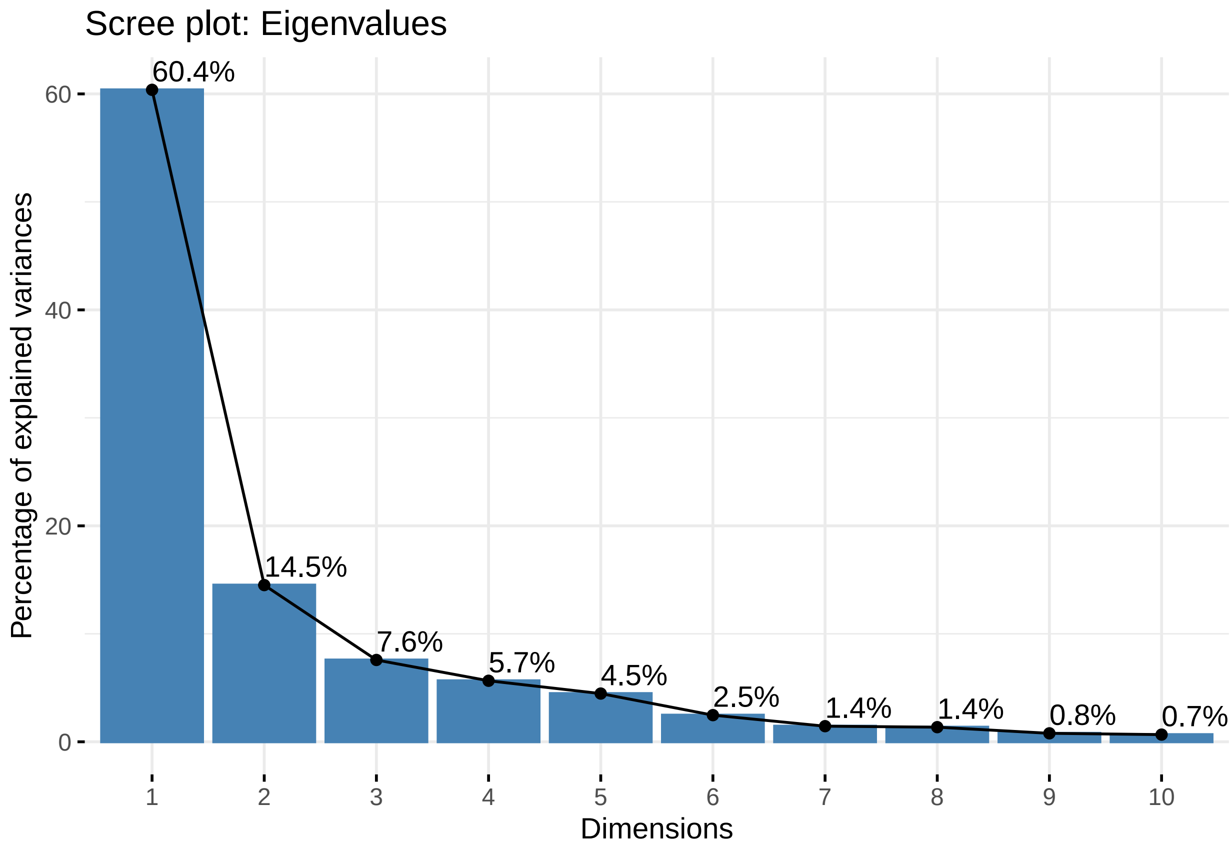
The goal of PCA (Principal Component Analysis) in the alcohol and drug survey is to reduce the dimensionality of the data and uncover underlying patterns by transforming variables into a set of linearly uncorrelated components, facilitating interpretation and visualization of the data's structure.

```
pca_result <- PCA(numeric_data, graph = FALSE)
head(pca_result$eig)
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1  7.2450173              60.375144                          60.37514
## comp 2  1.7414853              14.512377                          74.88752
## comp 3  0.9093015               7.577512                          82.46503
## comp 4  0.6784208               5.653506                          88.11854
## comp 5  0.5360308               4.466923                          92.58546
## comp 6  0.2962347               2.468623                          95.05409
```

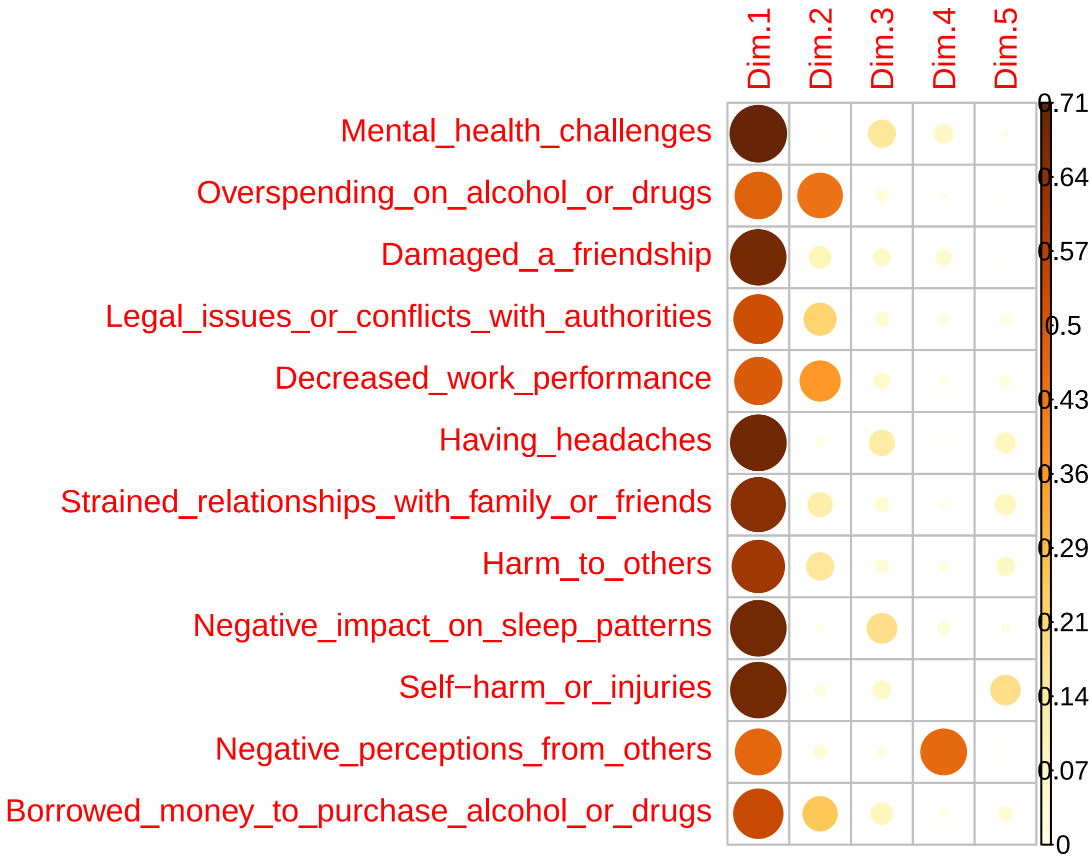
Interpretation: The first two principal components explain approximately 75% of the total variance in the data, while the cumulative percentage of variance explained by these two components is around 75%

```
fviz_eig(pca_result, addlabels = TRUE) +
 ggtitle("Scree plot: Eigenvalues")
```



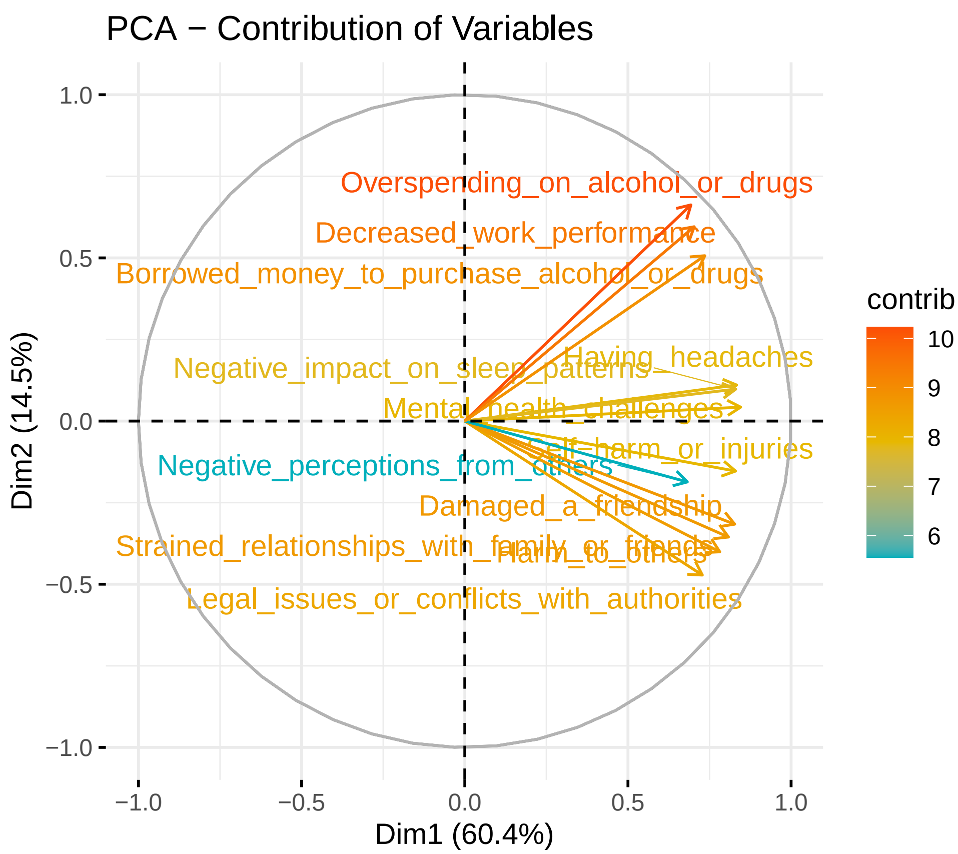
Interpretation: 1- Kaiser Criterion: We would retain principal components with eigenvalues greater than 1. In our case, only the first two components have eigenvalues exceeding 1, indicating their significance. 2- Cumulative Inertia Criterion: According to the cumulative inertia criterion, we would retain the first two principal components, explaining approximately 74.89% of the total inertia. 3- Elbow Criterion: When using the elbow method, we observe a clear bend after the second principal component, suggesting that we can suffice with only two components. => Thus, we opt to retain only the first two principal components, which collectively explain around 74.89% of the total inertia. This decision is supported by both the Kaiser criterion and the elbow method, indicating the significance of these two axes in capturing the variance within the data.

```r
var<-get_pca_var(pca_result)
corrplot(var$cos2,is.corr=FALSE)
```



The higher the cos2, the better the variable is represented.

```
fviz_pca_var(
  pca_result,
  col.var = "contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  title = "PCA - Contribution of Variables"
)
```



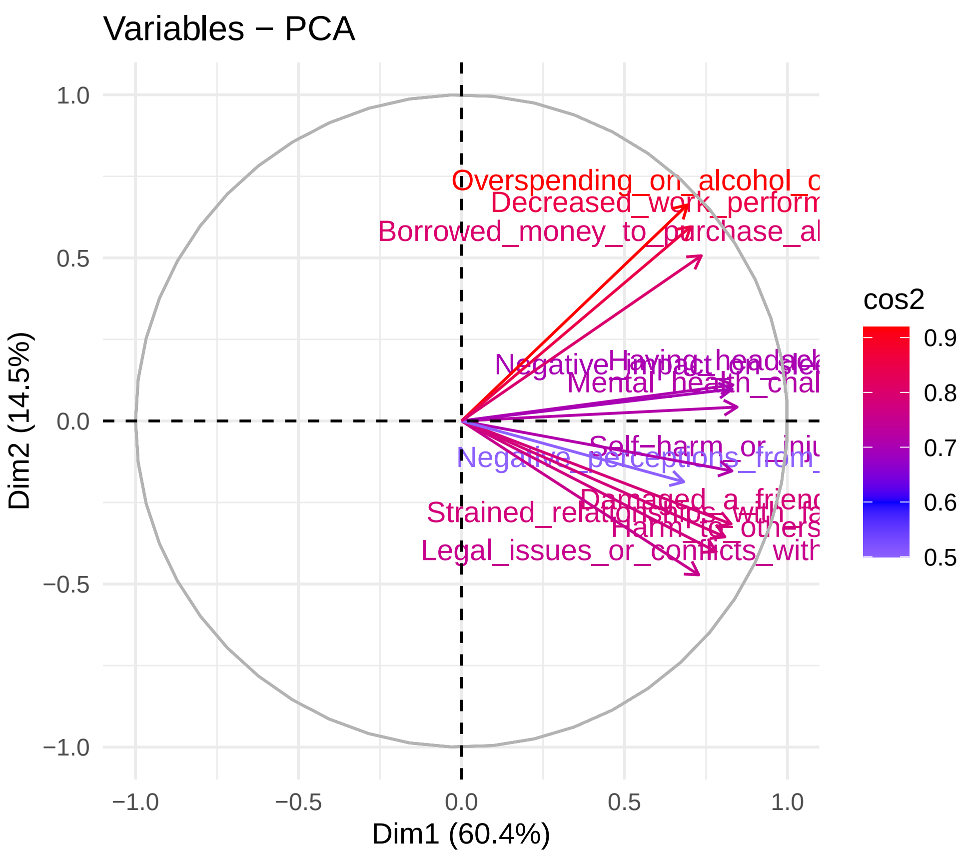
Interpretation: The PCA plot reveals distinct clusters of variables. Health-related factors, like mental health challenges and negative impact on sleep patterns, are closely grouped, suggesting a strong correlation. Similarly, variables related to behavioral and social aspects, such as overspending on alcohol, decreased work performance, and strained relationships, show tight associations. This implies that the first principal component likely represents health issues, given the close correlation among variables related to mental health and well-being.

```
pca_result$var$cos2
```

```
##                                                  Dim.1       Dim.2      Dim.3
## Mental_health_challenges                     0.7147113 0.001814660 0.15824880
## Overspending_on_alcohol_or_drugs             0.4801540 0.438831875 0.02536641
## Damaged_a_friendship                         0.6841489 0.099859106 0.05808064
## Legal_issues_or_conflicts_with_authorities   0.5299194 0.222695255 0.04047049
## Decreased_work_performance                   0.4943428 0.354554294 0.05049311
## Having_headaches                             0.6939159 0.012200001 0.13700163
## Strained_relationships_with_family_or_friends 0.6530318 0.126327601 0.04127343
## Harm_to_others                               0.6105945 0.160574843 0.02987066
## Negative_impact_on_sleep_patterns            0.6894138 0.009582687 0.19197708
## Self-harm_or_injuries                        0.6885361 0.023506301 0.06653067
## Negative_perceptions_from_others             0.4650981 0.034813096 0.01429935
## Borrowed_money_to_purchase_alcohol_or_drugs  0.5411507 0.256725548 0.09568921
##                                                     Dim.4        Dim.5
## Mental_health_challenges                     7.174065e-02 0.0059260855
## Overspending_on_alcohol_or_drugs             5.436582e-03 0.0009614676
## Damaged_a_friendship                         4.835237e-02 0.0008396351
```

```
## Legal_issues_or_conflicts_with_authorities    1.715165e-02 0.0211030208
## Decreased_work_performance                    1.347488e-02 0.0231505460
## Having_headaches                              9.051860e-04 0.0825199711
## Strained_relationships_with_family_or_friends 9.664917e-06 0.0882025819
## Harm_to_others                                1.342724e-02 0.0682295798
## Negative_impact_on_sleep_patterns             3.001802e-02 0.0142317385
## Self-harm_or_injuries                         1.916534e-05 0.1896403978
## Negative_perceptions_from_others              4.637658e-01 0.0004501193
## Borrowed_money_to_purchase_alcohol_or_drugs   1.411960e-02 0.0407756491
```

```
fviz_pca_var(pca_result, col.var = "cos2")+
 scale_color_gradient2(low="white" ,mid="blue" ,
 high="red", midpoint = 0.6)+
 theme_minimal()
```



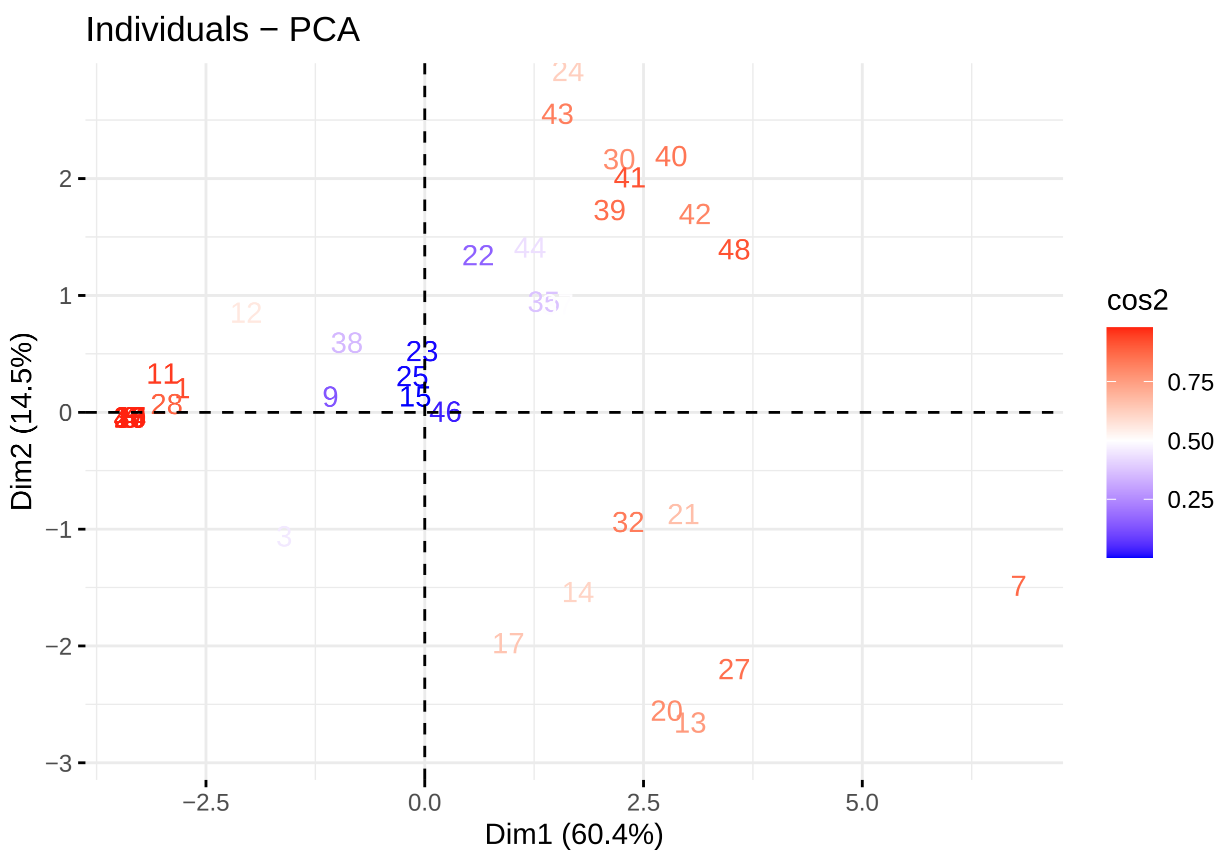
In the PCA analysis, the squared cosines of variables on the first two principal components (Dim.1 and Dim.2) suggest their relationships with these axes:

1. "Mental_health_challenges" exhibits a strong association with Dim.1, while "Having_headaches" and "Negative_impact_on_sleep_patterns" also contribute significantly to this component.
2. Dim.2 is primarily influenced by "Overspending_on_alcohol_or_drugs," "Damaged_a_friendship," and "Strained_relationships_with_family_or_friends," indicating their importance in explaining the variability along this axis.

These observations suggest that Dim.1 might represent factors related to mental health and well-being, while

Dim.2 may capture aspects such as social interactions and interpersonal relationships, including challenges like overspending on alcohol or drugs and strained relationships.

```
fviz_pca_ind(pca_result, geom = "text", col.ind = "cos2") + scale_color_gradient2(
  low = "blue",
  mid = "white",
  high = "red",
  midpoint = 0.5
)
```



MCA

The goal of MCA (Multiple Correspondence Analysis) in the alcohol and drug survey is to analyze the relationships between categorical variables, such as different types of substance use and demographic characteristics, in order to identify patterns and associations among these variables within the dataset.

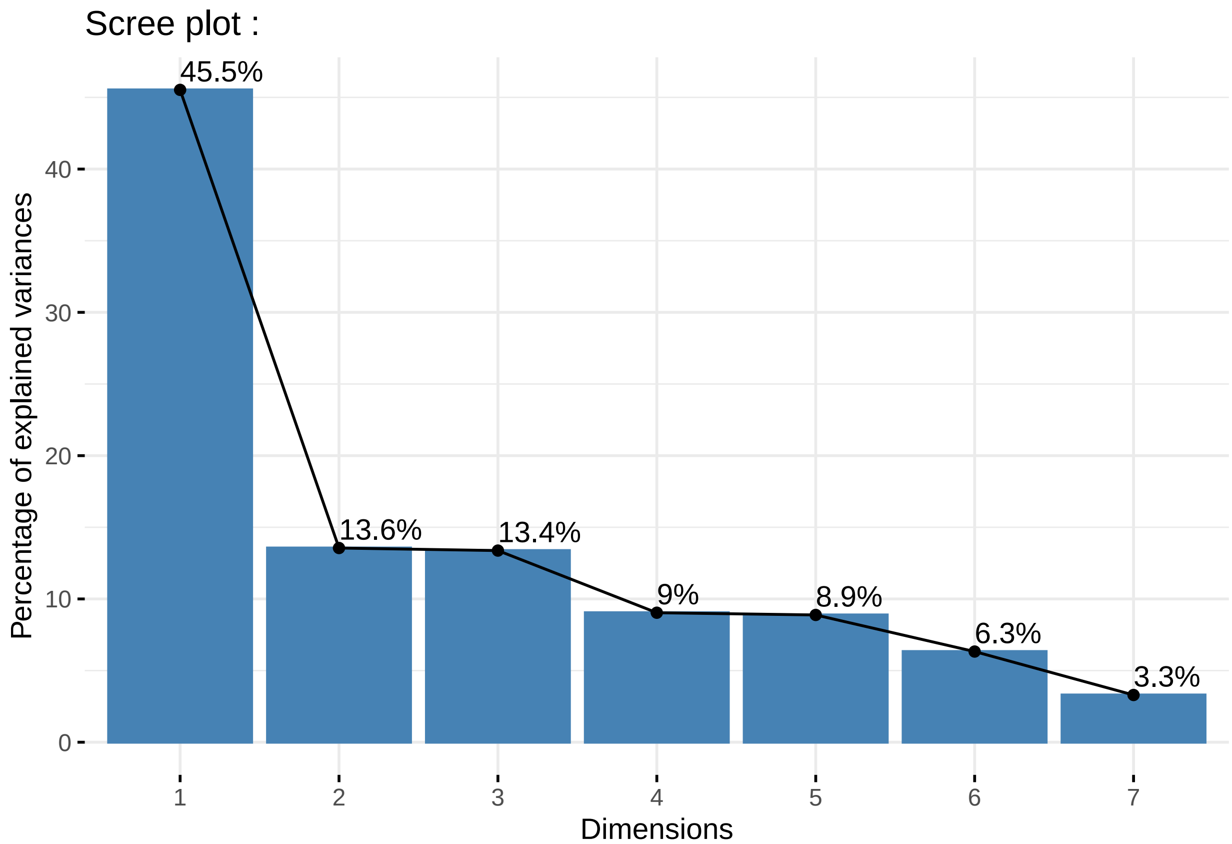
```
library(FactoMineR)
library(factoextra)
mca_subset <- data[, c(6, 8:10, 14:16)]
# Current column names
current_colnames <- colnames(mca_subset)

# Shorter column names
short_colnames <- c(
  "Tried_Alcohol",
  "Smoke_Cigarettes",
```

```r
  "Tried_Drugs",
  "Accessible_Drugs",
  "Addicted",
  "Attended_Educational_Programs",
  "Seeking_Help"
)

# Change column names
colnames(mca_subset) <- short_colnames

mca_result <- MCA (mca_subset, graph = FALSE)
```

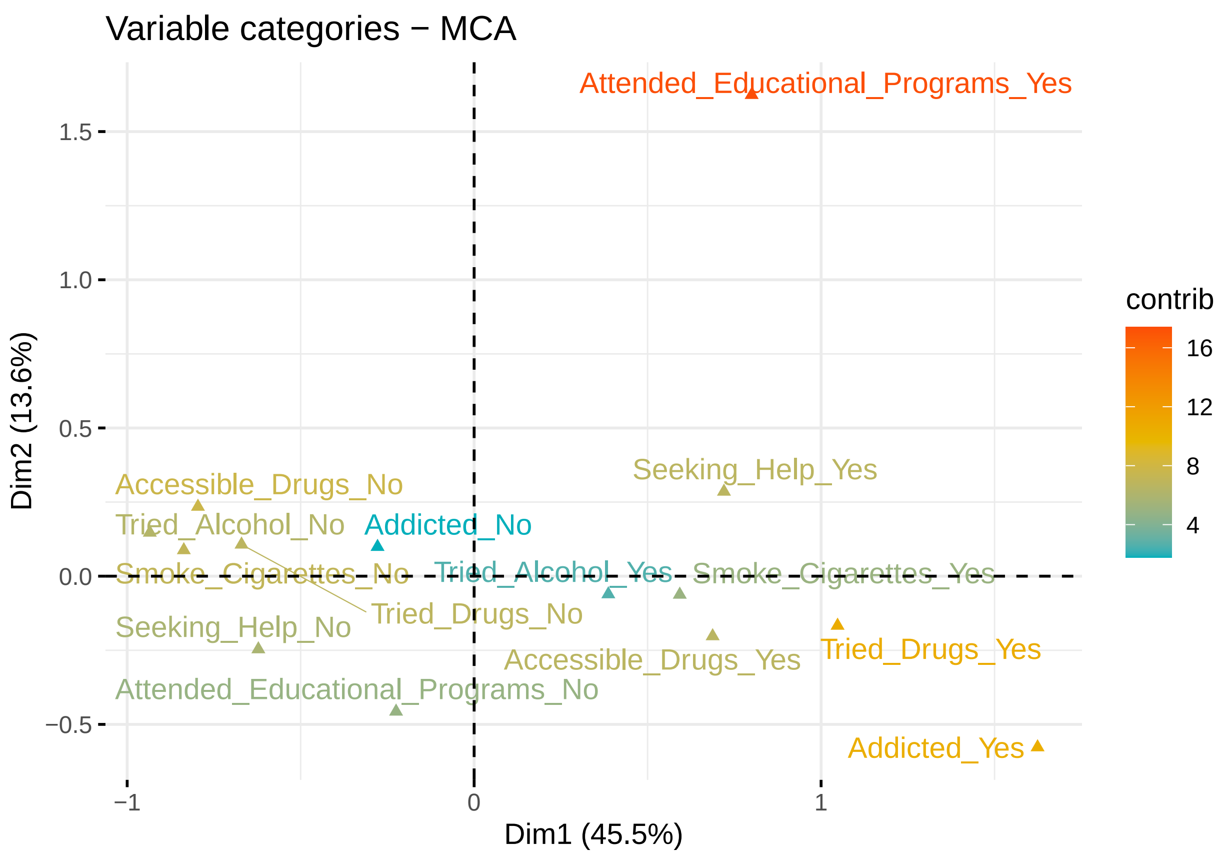
```r
mca_result$eig
```

```
##       eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.45524329              45.524329                          45.52433
## dim 2 0.13552889              13.552889                          59.07722
## dim 3 0.13375477              13.375477                          72.45269
## dim 4 0.09037448               9.037448                          81.49014
## dim 5 0.08883304               8.883304                          90.37345
## dim 6 0.06328552               6.328552                          96.70200
## dim 7 0.03298002               3.298002                         100.00000
```

```r
library(ggplot2)
library(factoextra)
fviz_eig(mca_result, addlabels = TRUE) +
 ggtitle("Scree plot :")
```




```r
fviz_mca_var(mca_result,
 col.var = "contrib", # Couleur en fonction de la contribution
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), # Choix de couleurs
 repel = TRUE, # Évite le chevauchement des labels
 ggtheme = theme_minimal() # Style du graphique
 )
```

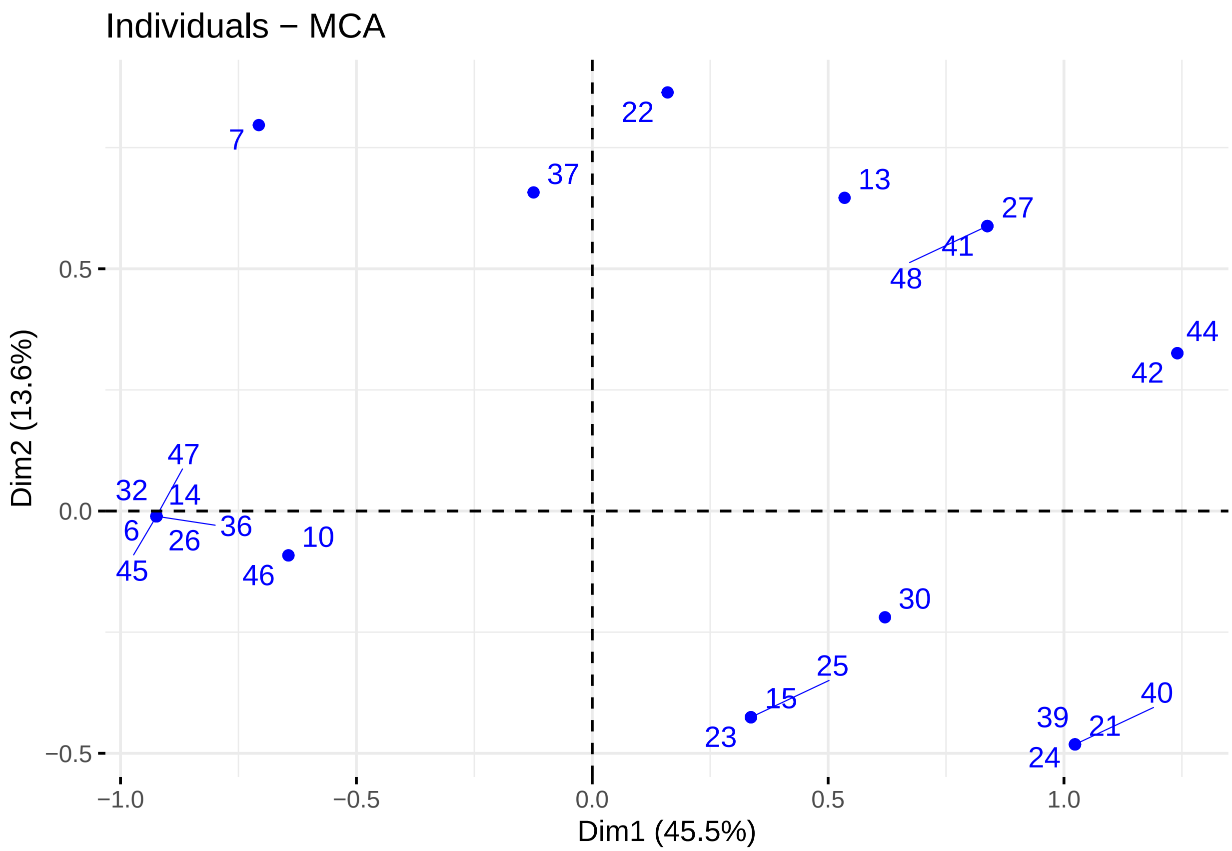



```r
dimdesc(mca_result, axes=1:2, proba=0.05)$`Dim 1`
```

```
##
## Link between the variable and the categorical variable (1-way anova)
## =============================================
##                                      R2      p.value
## Tried_Drugs                   0.7022349 8.251524e-12
## Accessible_Drugs              0.5472190 3.280818e-08
## Smoke_Cigarettes              0.4959335 2.781179e-07
## Addicted                      0.4520997 1.474500e-06
## Seeking_Help                  0.4478054 1.724393e-06
## Tried_Alcohol                 0.3614806 3.223871e-05
## Attended_Educational_Programs 0.1799298 5.710860e-03
##
## Link between variable and the categories of the categorical variables
## ================================================================
##                                                            Estimate
## Tried_Drugs=Tried_Drugs_Yes                               0.5795446
## Accessible_Drugs=Accessible_Drugs_Yes                     0.5004585
## Smoke_Cigarettes=Smoke_Cigarettes_Yes                     0.4822334
## Addicted=Addicted_Yes                                     0.6417755
## Seeking_Help=Seeking_Help_Yes                             0.4527226
## Tried_Alcohol=Tried_Alcohol_Yes                           0.4457881
## Attended_Educational_Programs=Attended_Educational_Programs_Yes  0.3457251
## Attended_Educational_Programs=Attended_Educational_Programs_No  -0.3457251
```

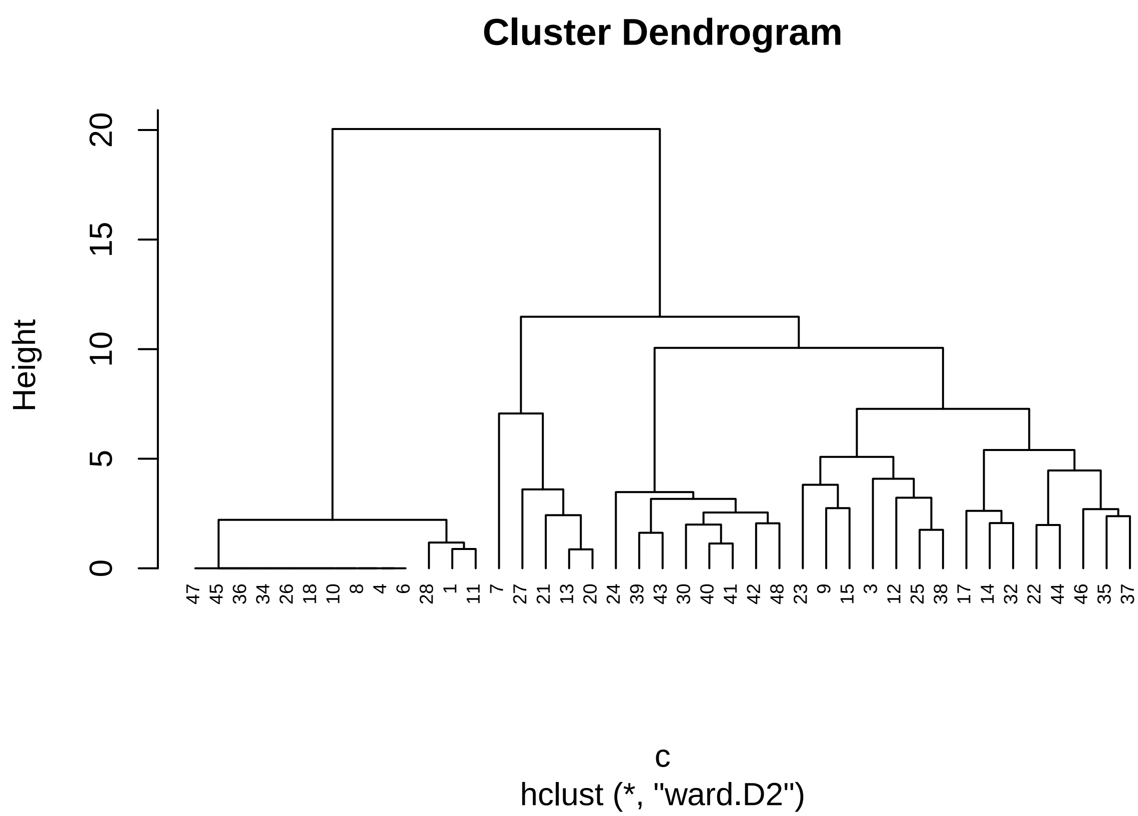
```
## Tried_Alcohol=Tried_Alcohol_No                                   -0.4457881
## Seeking_Help=Seeking_Help_No                                     -0.4527226
## Addicted=Addicted_No                                             -0.6417755
## Smoke_Cigarettes=Smoke_Cigarettes_No                             -0.4822334
## Accessible_Drugs=Accessible_Drugs_No                             -0.5004585
## Tried_Drugs=Tried_Drugs_No                                       -0.5795446
##                                                                       p.value
## Tried_Drugs=Tried_Drugs_Yes                                      8.251524e-12
## Accessible_Drugs=Accessible_Drugs_Yes                            3.280818e-08
## Smoke_Cigarettes=Smoke_Cigarettes_Yes                            2.781179e-07
## Addicted=Addicted_Yes                                            1.474500e-06
## Seeking_Help=Seeking_Help_Yes                                    1.724393e-06
## Tried_Alcohol=Tried_Alcohol_Yes                                  3.223871e-05
## Attended_Educational_Programs=Attended_Educational_Programs_Yes 5.710860e-03
## Attended_Educational_Programs=Attended_Educational_Programs_No  5.710860e-03
## Tried_Alcohol=Tried_Alcohol_No                                   3.223871e-05
## Seeking_Help=Seeking_Help_No                                     1.724393e-06
## Addicted=Addicted_No                                             1.474500e-06
## Smoke_Cigarettes=Smoke_Cigarettes_No                             2.781179e-07
## Accessible_Drugs=Accessible_Drugs_No                             3.280818e-08
## Tried_Drugs=Tried_Drugs_No                                       8.251524e-12
```

```
fviz_mca_ind (mca_result,select.ind = list(cos2 = 0.4),
 repel = TRUE,
 ggtheme = theme_minimal ())
```



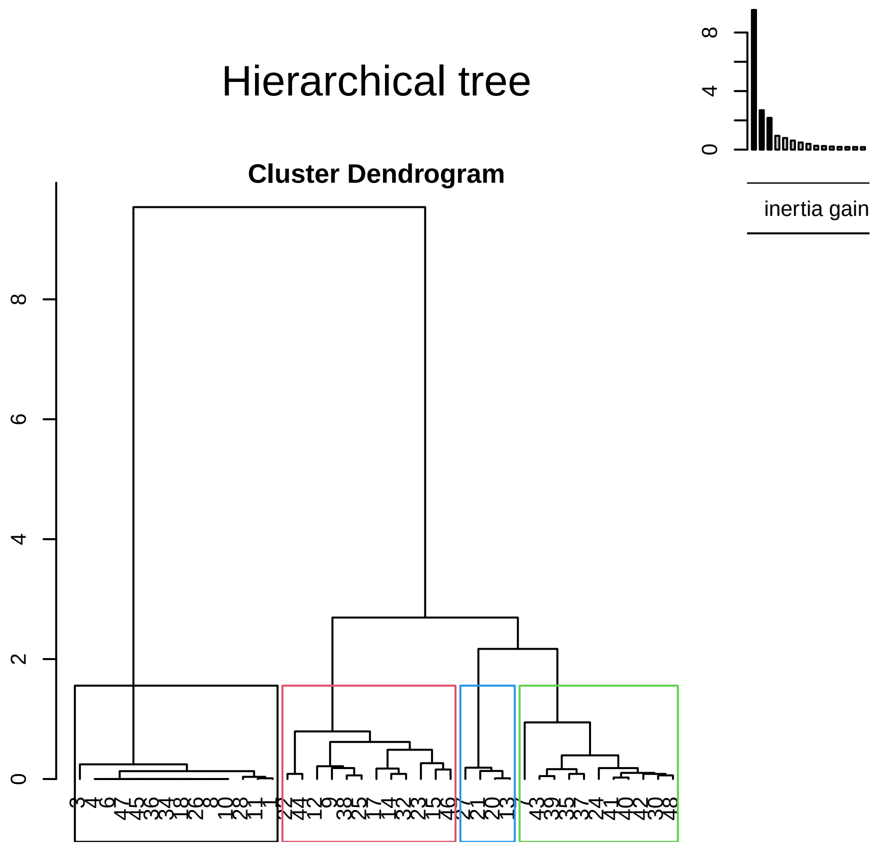
Clustering

```
# Load necessary libraries
library(FactoMineR)

# Calculate distance matrix
c <- dist(scale(numeric_data), method = "euclidean")

# Perform hierarchical clustering
h <- hclust(c, method = "ward.D2")

# Plot hierarchical clustering dendrogram
plot(h, hang = -1, cex = 0.6)
```



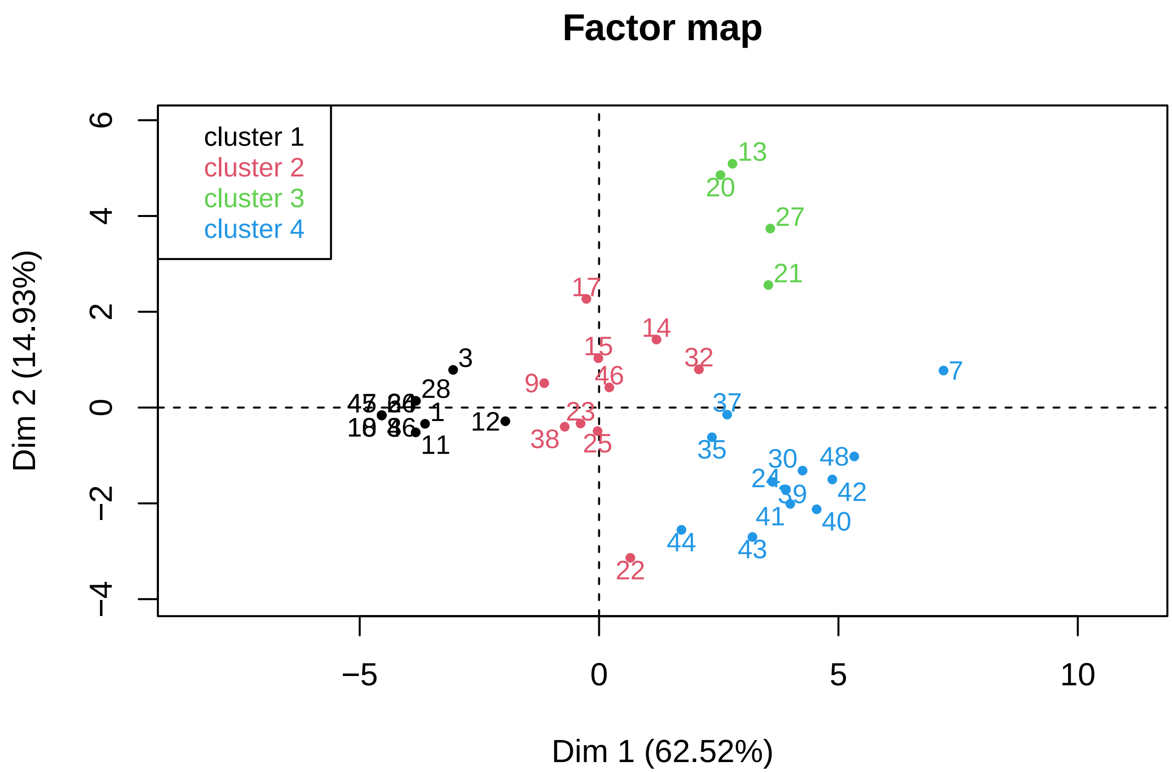
The hierarchical clustering dendrogram illustrates the relationship and grouping of observations based on their similarity in the scaled numeric data. Observations that are clustered closer together in the dendrogram are more similar to each other in terms of their numerical attributes, while those in separate branches are more dissimilar. The height of the dendrogram represents the distance between clusters; shorter branches indicate higher similarity.

```
# Perform HCPC
res.HCPC <- HCPC(numeric_data, consol = TRUE, graph = FALSE)

# Plot hierarchical tree
plot(res.HCPC, choice = 'tree', title = 'Hierarchical tree')
```



The HCPC analysis resulted in the formation of four distinct clusters, visually represented by the hierarchical tree.

```
plot.HCPC(res.HCPC, choice = 'map', draw.tree=FALSE, title ='Factor map')
```



We can clearly see that "cluster 1" individuals are more correlated to dim1 which represent health problems while "cluster 2" are slightly correlated to dim 2 which is financial and social problems.

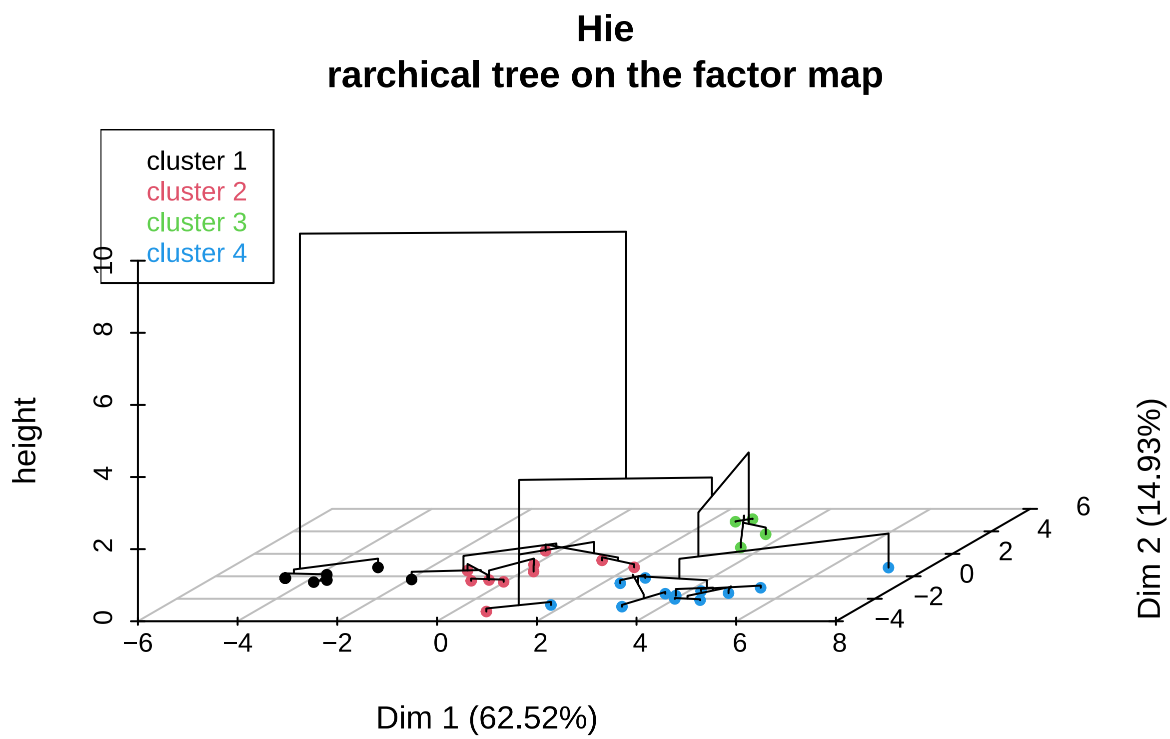
```
plot.HCPC(res.HCPC,choice = '3D.map', ind.names=FALSE, centres.plot=FALSE, angles=60,title='Hie
rarchical tree on the factor map')
```

```
## Warning in title(main, sub, ...): "centres.plot" is not a graphical parameter
```

```
## Warning in title(main, sub, ...): "angles" is not a graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "centres.plot" is not a
## graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "angles" is not a
## graphical parameter
```




```r
res.HCPC$desc.var
```

```
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##                                                   Eta2      P-value
## Decreased_work_performance                   0.8307834 2.414790e-14
## Mental_health_challenges                     0.8218210 6.241904e-14
## Overspending_on_alcohol_or_drugs             0.7997014 5.369801e-13
## Negative_impact_on_sleep_patterns            0.7926866 1.010949e-12
## Borrowed_money_to_purchase_alcohol_or_drugs  0.7696254 7.017220e-12
## Having_headaches                             0.7594846 1.547305e-11
## Self-harm_or_injuries                        0.6626858 7.577887e-09
## Strained_relationships_with_family_or_friends 0.5950758 2.117306e-07
## Negative_perceptions_from_others             0.5138442 5.829579e-06
## Damaged_a_friendship                         0.4851810 1.639194e-05
## Harm_to_others                               0.4648447 3.292528e-05
## Legal_issues_or_conflicts_with_authorities   0.3284483 1.884217e-03
##
## Description of each cluster by quantitative variables
## =====================================================
## $`1`
##                                                 v.test Mean in category
## Legal_issues_or_conflicts_with_authorities    -3.107905        1.133333
## Damaged_a_friendship                          -3.686395        1.066667
## Harm_to_others                                -3.882582        1.000000
```

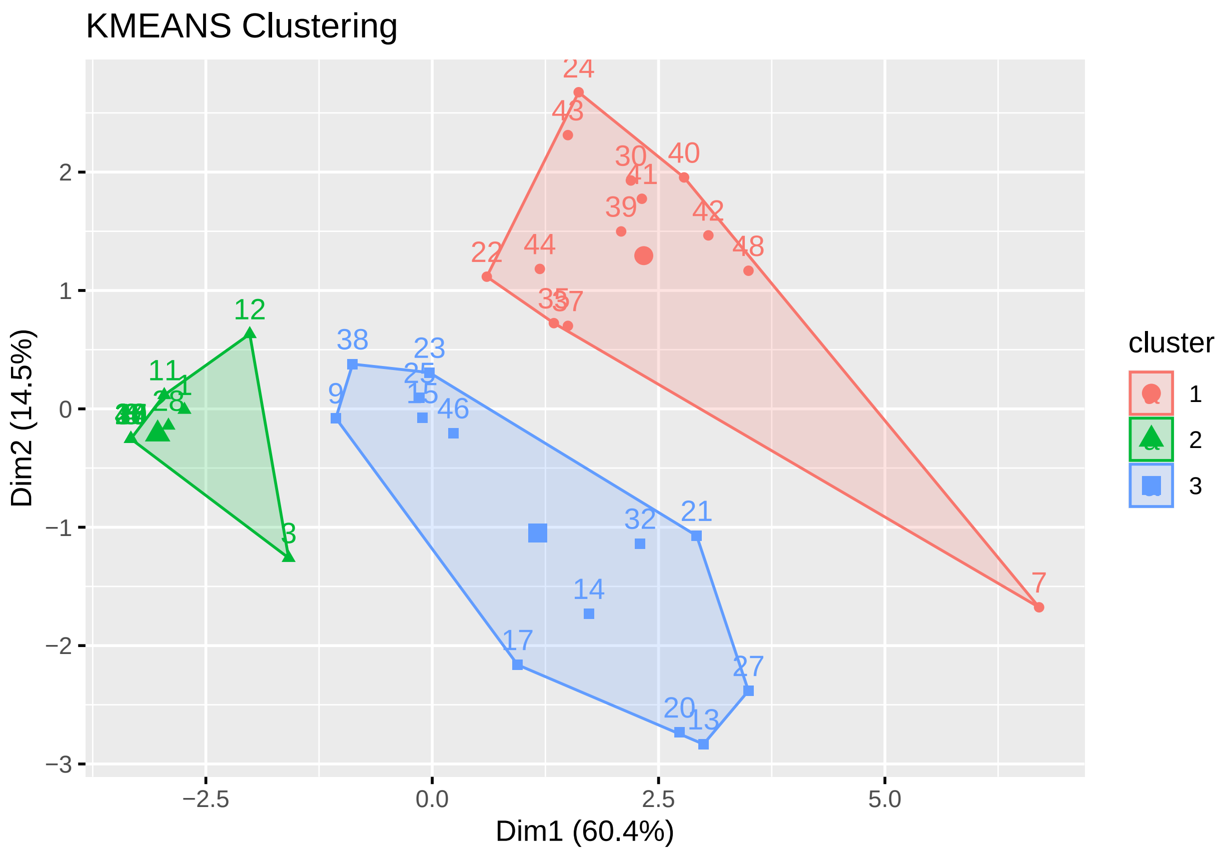
```
## Overspending_on_alcohol_or_drugs             -3.940135         1.266667
## Mental_health_challenges                     -4.226990         1.000000
## Self-harm_or_injuries                        -4.330850         1.066667
## Decreased_work_performance                   -4.398708         1.000000
## Borrowed_money_to_purchase_alcohol_or_drugs  -4.444827         1.066667
## Negative_perceptions_from_others             -4.471778         1.200000
## Strained_relationships_with_family_or_friends -4.637186         1.000000
## Having_headaches                             -4.920602         1.466667
## Negative_impact_on_sleep_patterns            -5.062222         1.133333
##                                              Overall mean sd in category
## Legal_issues_or_conflicts_with_authorities       1.780488      0.4988877
## Damaged_a_friendship                             1.853659      0.2494438
## Harm_to_others                                   1.658537      0.0000000
## Overspending_on_alcohol_or_drugs                 2.560976      0.5734884
## Mental_health_challenges                         2.292683      0.0000000
## Self-harm_or_injuries                            2.097561      0.2494438
## Decreased_work_performance                       2.487805      0.0000000
## Borrowed_money_to_purchase_alcohol_or_drugs      2.317073      0.2494438
## Negative_perceptions_from_others                 2.243902      0.5416026
## Strained_relationships_with_family_or_friends    2.000000      0.0000000
## Having_headaches                                 2.975610      0.8844333
## Negative_impact_on_sleep_patterns                2.878049      0.4988877
##                                              Overall sd      p.value
## Legal_issues_or_conflicts_with_authorities     1.0002974 1.884185e-03
## Damaged_a_friendship                           1.0255510 2.274528e-04
## Harm_to_others                                 0.8147948 1.033532e-04
## Overspending_on_alcohol_or_drugs               1.5780318 8.143563e-05
## Mental_health_challenges                       1.4690947 2.368386e-05
## Self-harm_or_injuries                          1.1434837 1.485351e-05
## Decreased_work_performance                     1.6248370 1.088972e-05
## Borrowed_money_to_purchase_alcohol_or_drugs    1.3514044 8.796270e-06
## Negative_perceptions_from_others               1.1214209 7.757196e-06
## Strained_relationships_with_family_or_friends  1.0359395 3.531841e-06
## Having_headaches                               1.4731384 8.627835e-07
## Negative_impact_on_sleep_patterns              1.6556647 4.143983e-07
##
## $`2`
##                                                v.test Mean in category
## Negative_perceptions_from_others             2.421912              3.0
## Strained_relationships_with_family_or_friends 2.080492              2.6
##                                              Overall mean sd in category
## Negative_perceptions_from_others                 2.243902            1.0
## Strained_relationships_with_family_or_friends    2.000000            0.8
##                                              Overall sd    p.value
## Negative_perceptions_from_others               1.121421 0.01543909
## Strained_relationships_with_family_or_friends  1.035940 0.03748046
##
## $`3`
##                                                v.test Mean in category
## Mental_health_challenges                     3.478328             4.75
## Self-harm_or_injuries                        3.459712             4.00
## Having_headaches                             2.857655             5.00
## Damaged_a_friendship                         2.831350             3.25
## Harm_to_others                               2.785608             2.75
```

```
## Negative_impact_on_sleep_patterns            2.665153            5.00
## Legal_issues_or_conflicts_with_authorities    2.535223            3.00
## Strained_relationships_with_family_or_friends 2.509197            3.25
##                                              Overall mean sd in category
## Mental_health_challenges                         2.292683      0.4330127
## Self-harm_or_injuries                            2.097561      0.7071068
## Having_headaches                                 2.975610      0.0000000
## Damaged_a_friendship                             1.853659      0.4330127
## Harm_to_others                                   1.658537      0.4330127
## Negative_impact_on_sleep_patterns                2.878049      0.0000000
## Legal_issues_or_conflicts_with_authorities       1.780488      0.7071068
## Strained_relationships_with_family_or_friends    2.000000      0.4330127
##                                              Overall sd      p.value
## Mental_health_challenges                      1.4690947 0.0005045519
## Self-harm_or_injuries                         1.1434837 0.0005407544
## Having_headaches                              1.4731384 0.0042678404
## Damaged_a_friendship                          1.0255510 0.0046352034
## Harm_to_others                                0.8147948 0.0053427364
## Negative_impact_on_sleep_patterns             1.6556647 0.0076953302
## Legal_issues_or_conflicts_with_authorities    1.0002974 0.0112375817
## Strained_relationships_with_family_or_friends  1.0359395 0.0121006033
##
## $‘4‘
##                                              v.test Mean in category
## Overspending_on_alcohol_or_drugs            5.428758         4.666667
## Decreased_work_performance                  5.246931         4.583333
## Borrowed_money_to_purchase_alcohol_or_drugs 4.815550         3.916667
## Negative_impact_on_sleep_patterns           3.575997         4.333333
## Mental_health_challenges                    3.574213         3.583333
## Having_headaches                            3.289354         4.166667
## Self-harm_or_injuries                       2.321302         2.750000
## Damaged_a_friendship                        2.233464         2.416667
##                                             Overall mean sd in category
## Overspending_on_alcohol_or_drugs                2.560976      0.6236096
## Decreased_work_performance                      2.487805      0.4930066
## Borrowed_money_to_purchase_alcohol_or_drugs     2.317073      0.7592028
## Negative_impact_on_sleep_patterns               2.878049      0.8498366
## Mental_health_challenges                        2.292683      0.9537936
## Having_headaches                                2.975610      0.5527708
## Self-harm_or_injuries                           2.097561      0.8291562
## Damaged_a_friendship                            1.853659      1.0374916
##                                             Overall sd      p.value
## Overspending_on_alcohol_or_drugs              1.578032 5.674743e-08
## Decreased_work_performance                    1.624837 1.546534e-07
## Borrowed_money_to_purchase_alcohol_or_drugs   1.351404 1.467948e-06
## Negative_impact_on_sleep_patterns             1.655665 3.488955e-04
## Mental_health_challenges                      1.469095 3.512826e-04
## Having_headaches                              1.473138 1.004177e-03
## Self-harm_or_injuries                         1.143484 2.027057e-02
## Damaged_a_friendship                          1.025551 2.551834e-02
```

The output provides information on the relationship between cluster variables and quantitative variables, showing their strength of association, significance, and descriptive statistics within each cluster.

For instance, Cluster 1 exhibits higher levels of various issues related to alcohol, drugs, and mental health, while Cluster 4 is characterized by elevated levels of overspending on alcohol or drugs, decreased work performance, and borrowing money for alcohol or drugs, indicating potential addiction and behavioral concerns.

```
res.km <- eclust(numeric_data, "kmeans", nstart = 25,k=3)
```



This allows for the segmentation of the data into three distinct groups based on their similarities in the numeric features.

1. First cluster: Financial problems
2. Second cluster: Social problems
3. Third cluster: Health problems

Conclusion:

In conclusion, this study utilized various data analysis techniques to investigate patterns of drug and alcohol consumption among students. Through exploratory data analysis, visualization, and statistical modeling, we gained valuable insights into the prevalence, perceptions, and behaviors associated with substance use within the surveyed population.

The findings revealed that alcohol consumption is prevalent among the students, with a significant portion reporting experimentation with various drugs. Moreover, perceptions of harm associated with substance use varied among individuals, with distinct clusters emerging based on factors such as financial, social, and health-related problems.

Overall, this study contributes to a better understanding of substance use behaviors among students and highlights the need for targeted interventions to address the diverse challenges associated with drug and alcohol consumption.