Project2

Suraj Bhattarai
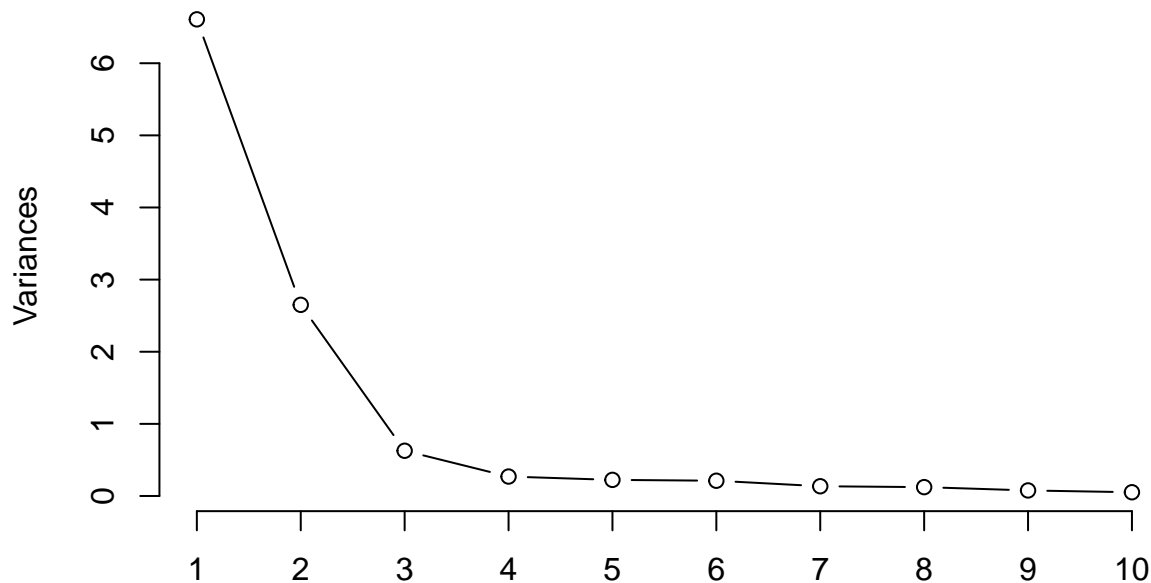
2023-07-04

1. Perform the principal component analysis in the data and exact the dimensions based on components with eigenvalues >1, check it with screeplot as well and interpret the result carefully.

```r
# Load the mtcars dataset
data(mtcars)

# 1. Perform Principal Component Analysis (PCA)
pca <- prcomp(mtcars, scale. = TRUE) # Perform PCA
summary(pca) # View summary of PCA results
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6    PC7
## Standard deviation     2.5707 1.6280 0.79196 0.51923 0.47271 0.46000 0.3678
## Proportion of Variance 0.6008 0.2409 0.05702 0.02451 0.02031 0.01924 0.0123
## Cumulative Proportion  0.6008 0.8417 0.89873 0.92324 0.94356 0.96279 0.9751
##                           PC8    PC9    PC10   PC11
## Standard deviation     0.35057 0.2776 0.22811 0.1485
## Proportion of Variance 0.01117 0.0070 0.00473 0.0020
## Cumulative Proportion  0.98626 0.9933 0.99800 1.0000
```

Screeplot of PCA

Interpretation: # The screeplot shows the eigenvalues for each principal component. # Look for the components with eigenvalues > 1, as they explain more variance than a single original variable. # These components can be considered significant in the analysis.

##2. Perform the principal component analysis with varimax rotation in the data and exact the dimensions based on eigenvalue >1 and check it with Screeplot as well and interpret the result carefully

```r
library(psych)  # Load the psych package for varimax rotation
```

```
## Warning: package 'psych' was built under R version 4.2.3
```

```r
# Check and remove rows with missing or infinite values again
mtcars_rot <- mtcars[complete.cases(mtcars), ]

pca_rot <- principal(mtcars_rot, nfactors = ncol(mtcars_rot), rotate = "varimax")  # Perform PCA with va
summary(pca_rot)  # View summary of rotated PCA results
```

```
##
## Factor analysis with Call: principal(r = mtcars_rot, nfactors = ncol(mtcars_rot), rotate = "varimax")
##
## Test of the hypothesis that 11 factors are sufficient.
## The degrees of freedom for the model is -11  and the objective function was  0
## The number of observations was  32  with Chi Square =  0  with prob <  NA
##
## The root mean square of the residuals (RMSA) is  0
```
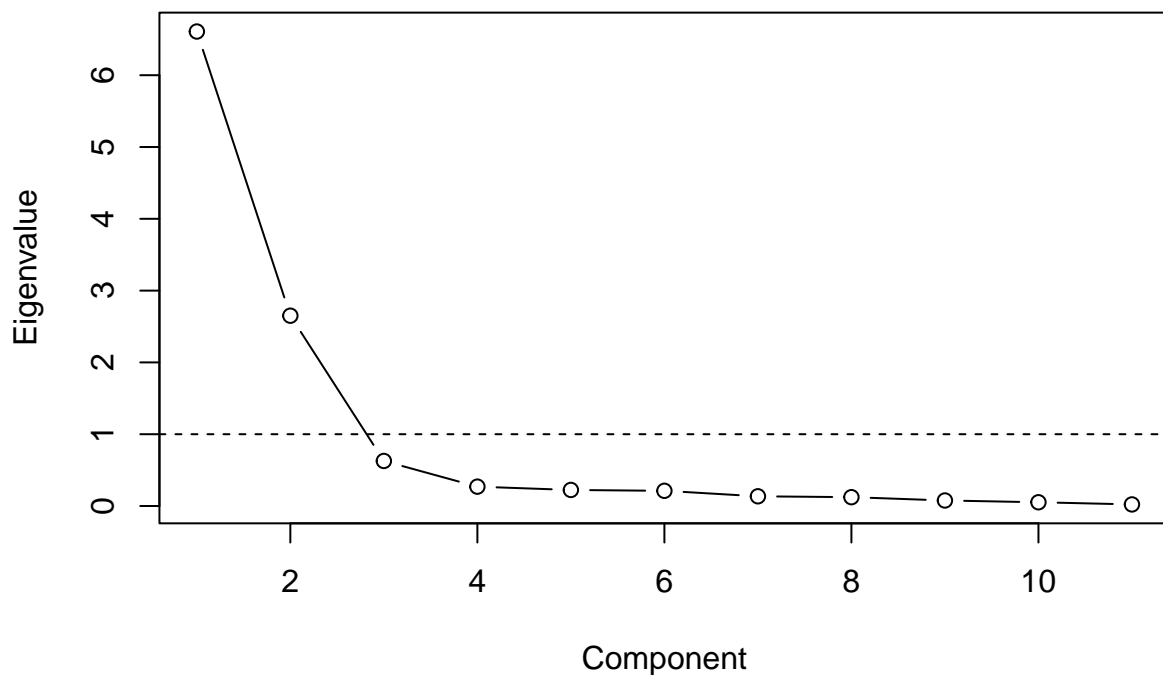
```
# Extract eigenvalues
eigenvalues <- pca_rot$values

# Screeplot
plot(1:length(eigenvalues), eigenvalues, type = "b", xlab = "Component", ylab = "Eigenvalue",
     main = "Screeplot of PCA with Varimax Rotation")

# Add a horizontal line at eigenvalue = 1
abline(h = 1, lty = 2)
```

Screeplot of PCA with Varimax Rotation


```
# Interpretation:
# The screeplot of the rotated PCA shows the eigenvalues for each rotated component.
# Look for components with eigenvalues > 1, which indicate significant dimensions.
```

##3. Perform the classical multidimensional scaling in the data, revise the results using stress values and interpret the result carefully
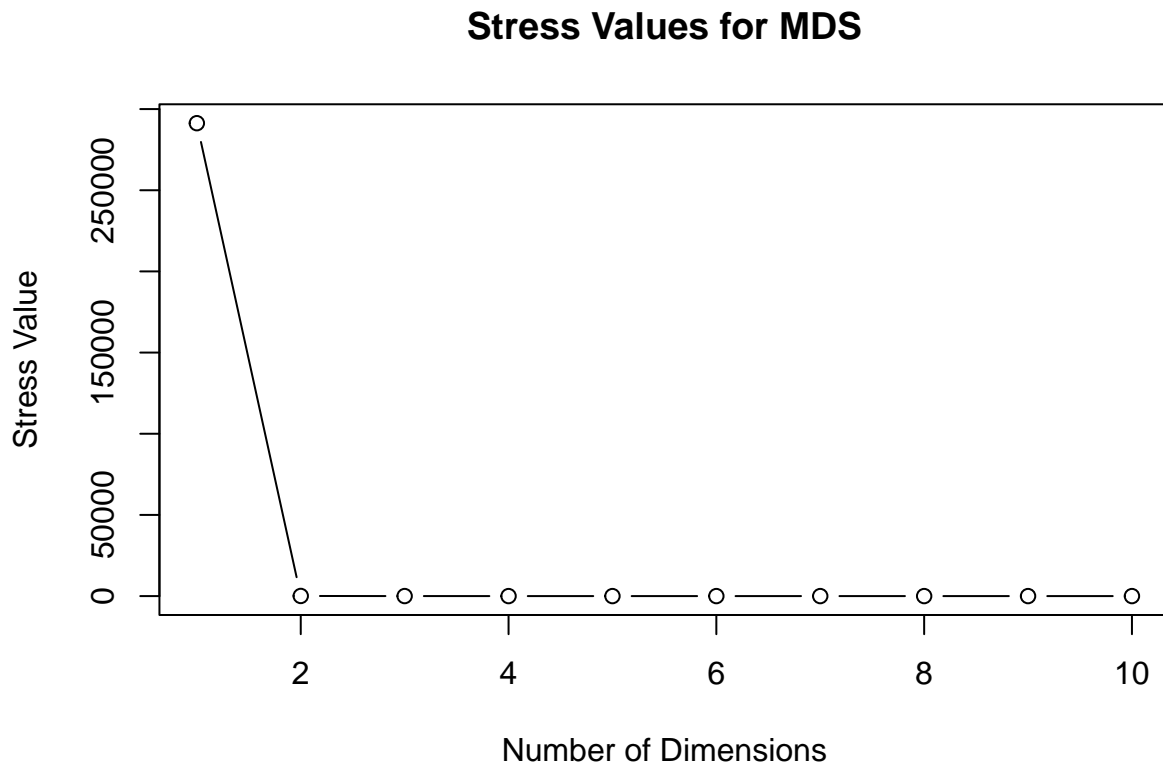
```
# Calculate stress values using cmdscale
stress_values <- c()
for (k in 1:10) {
  mds_k <- cmdscale(dist(mtcars), k = k)
  stress_values[k] <- sum((dist(mtcars) - dist(mds_k))^2)
}

# Plot stress values
```

```
plot(1:10, stress_values, type = "b",
     xlab = "Number of Dimensions", ylab = "Stress Value",
     main = "Stress Values for MDS")
```

Stress Values for MDS

Interpretation: # MDS represents the dissimilarities between cars in a low-dimensional space. # The stress values quantify the goodness of fit between the original distances and the distances in the reduced space. # Lower stress values indicate a better fit, and the interpretation of the result should consider the stress values obtained.

##4. Perform the hierarchical cluster analysis in the data and determine the number of clusters to exact using the dendogram and cut at the various distances with justification

```
# Perform Hierarchical Cluster Analysis (HCA)
hc <- hclust(dist(mtcars))
plot(hc, main = "Dendrogram for Hierarchical Clustering")
```

Dendrogram for Hierarchical Clustering

dist(mtcars)
hclust (*, "complete")
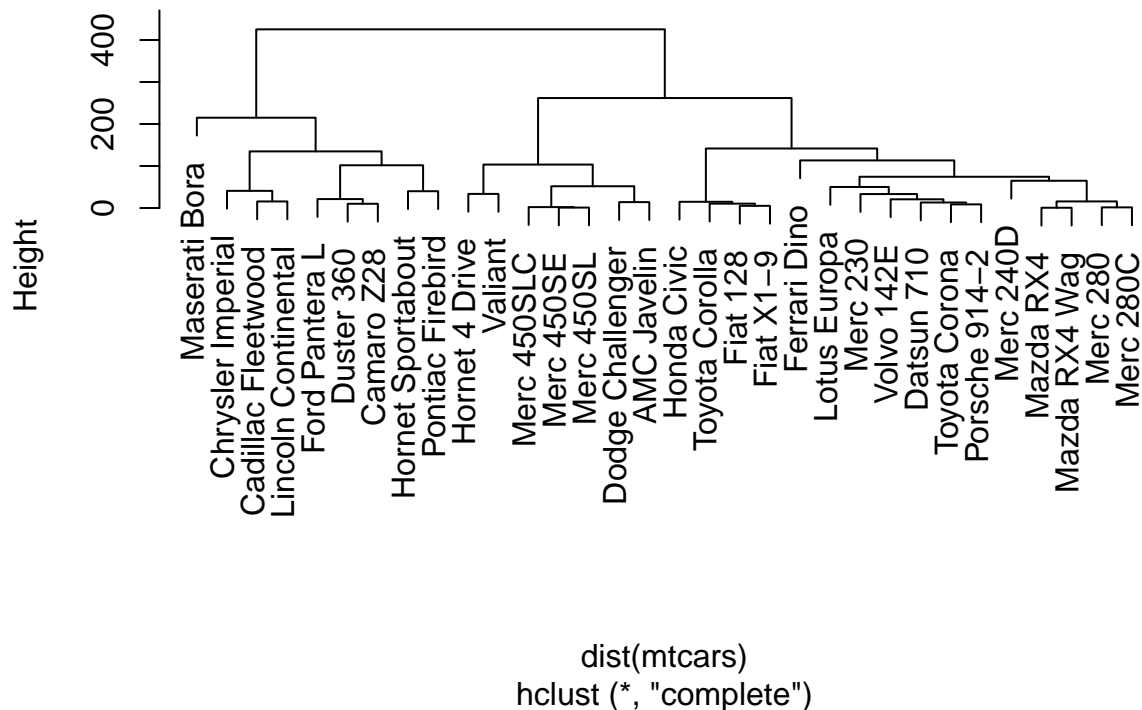
```
# Cut the dendrogram at various distances
cut_dendrogram <- cutree(hc, k = 3)
```

Justification:

The optimal number of clusters can be determined by examining the dendrogram and observing where the branches merge.

Cutting the dendrogram at various distances allows for exploring different cluster solutions.

##5. Perform the k-means cluster analysis in the data based on the number of clusters identified using dendogram and interpret the result carefully

```
# Perform K-means Cluster Analysis
kmeans_clusters <- kmeans(mtcars, centers = 3)  # Adjust centers as desired
kmeans_clusters$cluster
```

```
##           Mazda RX4      Mazda RX4 Wag          Datsun 710       Hornet 4 Drive
##                   2                  2                   2                    1
##    Hornet Sportabout            Valiant          Duster 360            Merc 240D
```

5

```
##                 3                 1                 3                 2
##          Merc 230          Merc 280         Merc 280C         Merc 450SE
##                 2                 2                 2                 1
##         Merc 450SL       Merc 450SLC Cadillac Fleetwood Lincoln Continental
##                 1                 1                 3                 3
##   Chrysler Imperial          Fiat 128        Honda Civic     Toyota Corolla
##                 3                 2                 2                 2
##       Toyota Corona  Dodge Challenger        AMC Javelin         Camaro Z28
##                 2                 1                 1                 3
##    Pontiac Firebird          Fiat X1-9       Porsche 914-2       Lotus Europa
##                 3                 2                 2                 2
##      Ford Pantera L       Ferrari Dino      Maserati Bora          Volvo 142E
##                 3                 2                 3                 2
```

Interpretation:

K-means clustering groups the observations into k clusters based on similarity.

The interpretation of the clustering result should consider the characteristics of each cluster and their differences.