31_question_10

Sudarshan Budhathoki

2024-05-31

```r
# Load the required library
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
library(factoextra)
```

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

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

```r
# Load the "iris" dataset
data(iris)

# Select the first four variables from the iris dataset
iris_data <- iris[, 1:4]

# a) Fit a K-means clustering model with k=2 and k=3
k2 <- kmeans(iris_data, centers = 2)
k3 <- kmeans(iris_data, centers = 3)

# b) Plot the clusters formed with k=3 in a single graph and interpret them
p1 <- fviz_cluster(k3, data = iris_data, geom = "point", stand = FALSE)
print(p1)
```
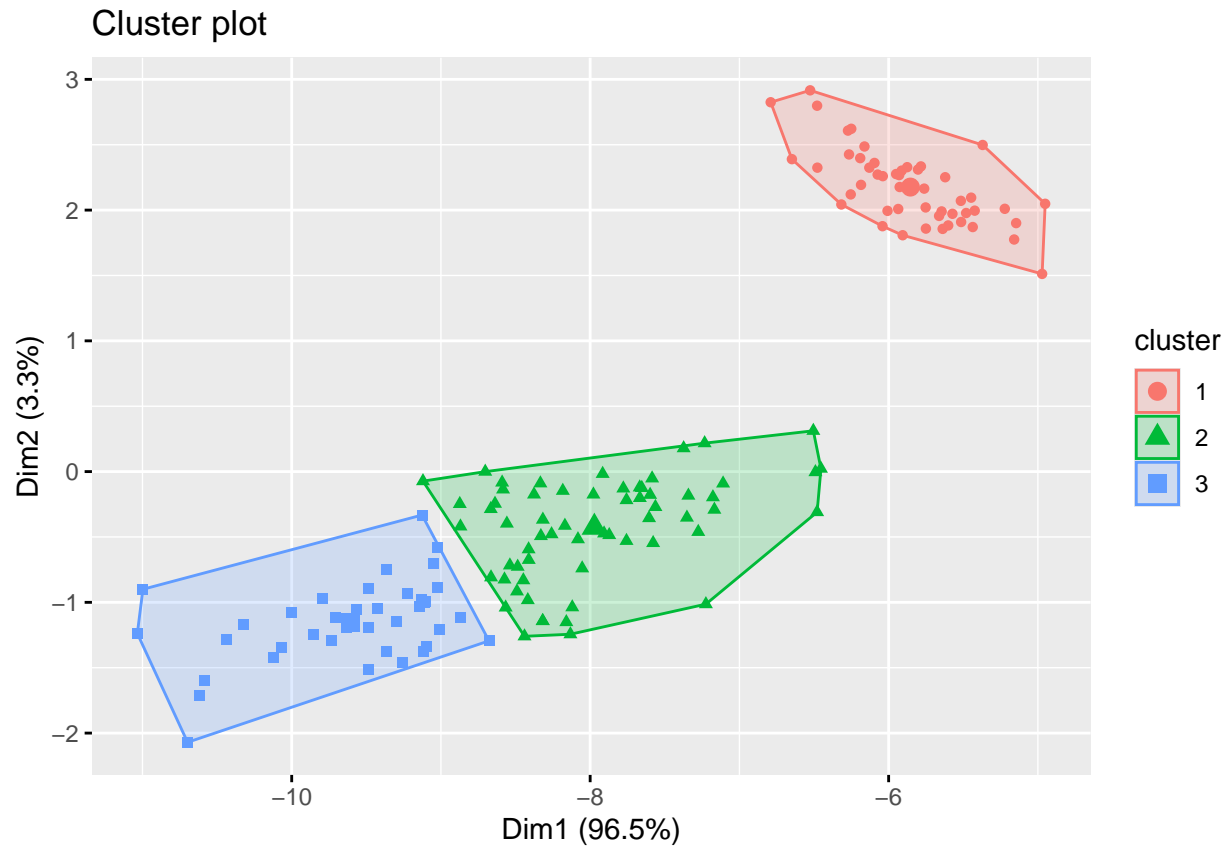
Cluster plot


```r
# c) Add cluster centers for the plot of clusters formed with k=3 and interpret them
p2 <- p1 + geom_point(data = as.data.frame(k3$centers), aes(x = Sepal.Length, y = Sepal.Width),
                      color = "red", size = 3, shape = 17)
print(p2)
```
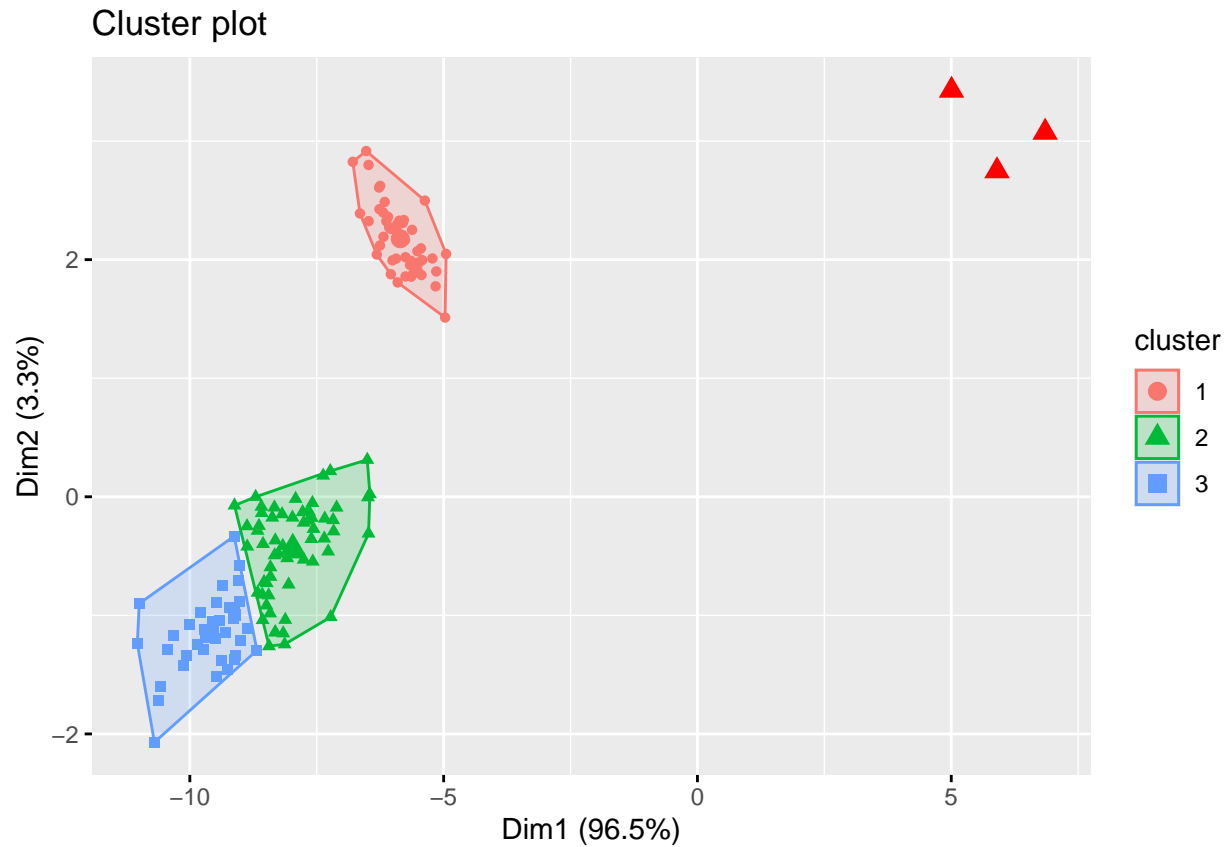
Cluster plot


```
# d) Compare the k=3 clusters with species variable using confusion matrix and interpret the result
species_map <- c(setosa = 3, versicolor = 2, virginica = 1)  # Mapping species to cluster number

iris_pred <- as.factor(k3$cluster)
iris_pred_mapped <- as.factor(species_map[as.character(iris_pred)])
iris_true <- as.factor(species_map[iris$Species])

conf_matrix <- table(iris_pred_mapped, iris_true)
conf_matrix
```

```
## < table of extent 0 x 3 >
```