

Exercise 11.2

Andrew Jordan

2022-06-02

Question 1: Introduction to Machine Learning

```
binary_df <- read.csv("data/binary-classifier-data.csv")
trinary_df <- read.csv("data/trinary-classifier-data.csv")
head(binary_df)
```

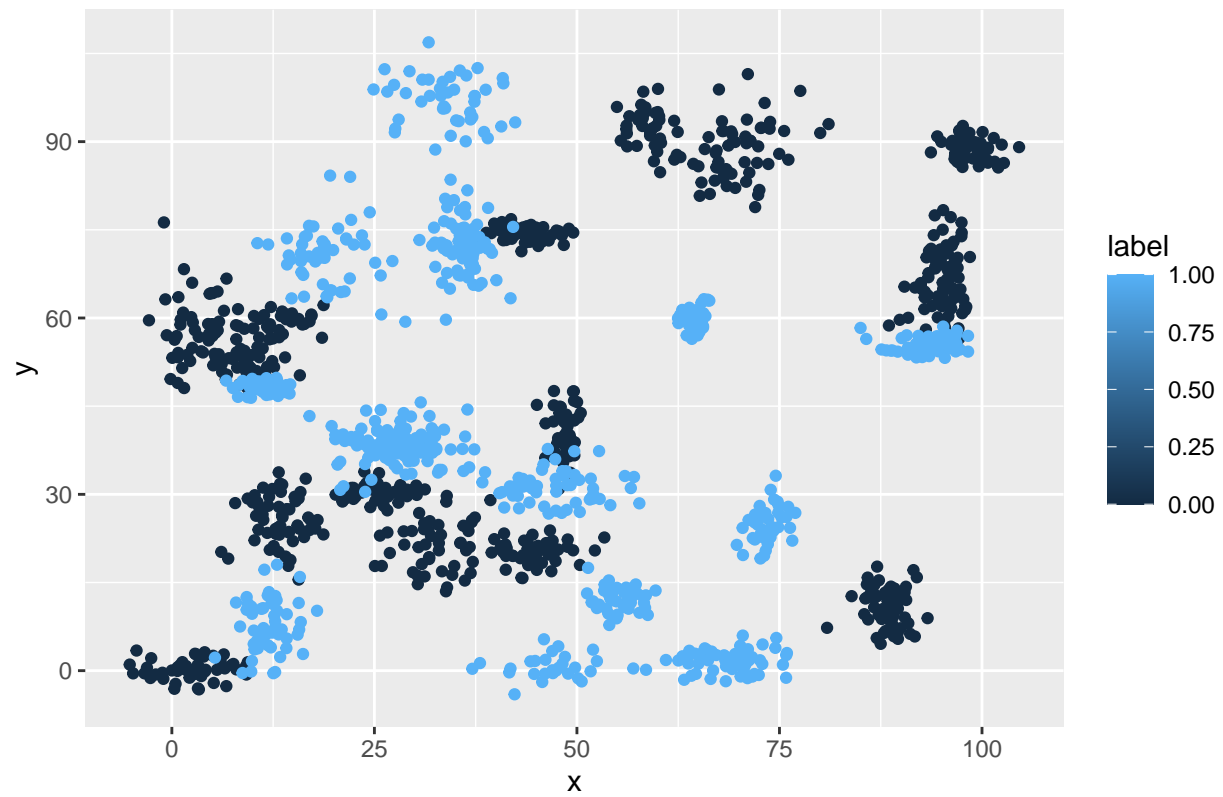
```
##   label      x      y
## 1     0 70.88469 83.17702
## 2     0 74.97176 87.92922
## 3     0 73.78333 92.20325
## 4     0 66.40747 81.10617
## 5     0 69.07399 84.53739
## 6     0 72.23616 86.38403
```

```
head(trinary_df)
```

```
##   label      x      y
## 1     0 30.08387 39.63094
## 2     0 31.27613 51.77511
## 3     0 34.12138 49.27575
## 4     0 32.58222 41.23300
## 5     0 34.65069 45.47956
## 6     0 33.80513 44.24656
```

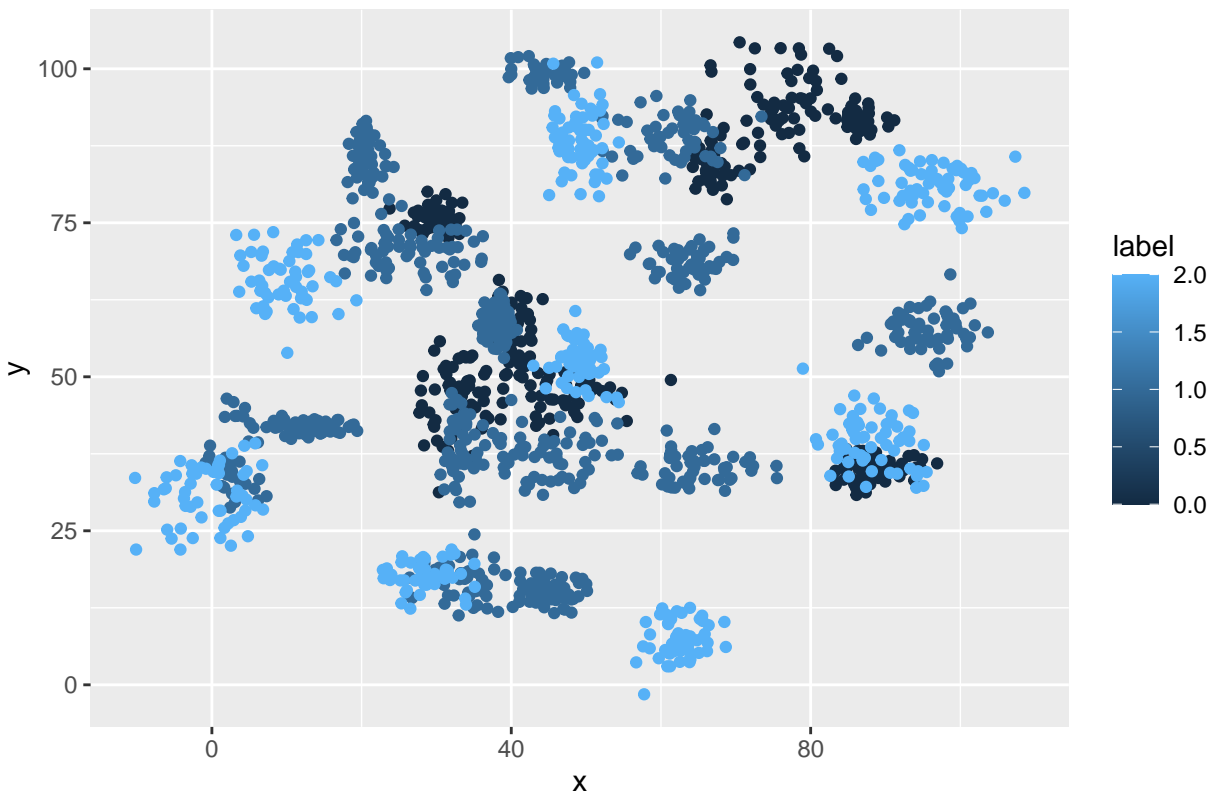
```
library(ggplot2)
ggplot(binary_df, aes(x = x, y = y, color = label)) + geom_point() + ggtitle("Binary Classifier Data")
```

Binary Classifier Data



```
ggplot(trinary_df, aes(x = x, y = y, color = label)) + geom_point() + ggtitle("Trinary Classifier Data")
```

Trinary Classifier Data



```
binary_dm <- as.matrix(binary_df)
binary_k3 <- kmeans(binary_dm, 3)
binary_k5 <- kmeans(binary_dm, 5)
binary_k10 <- kmeans(binary_dm, 10)
binary_k15 <- kmeans(binary_dm, 15)
binary_k20 <- kmeans(binary_dm, 20)
binary_k25 <- kmeans(binary_dm, 25)
```

```
trinary_dm <- as.matrix(trinary_df)
trinary_k3 <- kmeans(trinary_dm, 3)
trinary_k5 <- kmeans(trinary_dm, 5)
trinary_k10 <- kmeans(trinary_dm, 10)
trinary_k15 <- kmeans(trinary_dm, 15)
trinary_k20 <- kmeans(trinary_dm, 20)
trinary_k25 <- kmeans(trinary_dm, 25)
```

At this point professor, I am unsure of how to compute the accuracy of the resulting models for each value of k . I've gone through the textbook and additional materials, but I believe I am either overlooking or not understanding something essential to my comprehending checking the KNN models' accuracy.

Question 2: Clustering

```
cluster_df <- read.csv("data/clustering-data.csv")
head(cluster_df)
```

```
##      x    y
## 1  46 236
## 2  69 236
## 3 144 236
## 4 171 236
## 5 194 236
## 6 195 236
```

```
ggplot(cluster_df, aes(x = x, y = y, color = "red")) + geom_point() + ggtitle("Clustering Data")
```



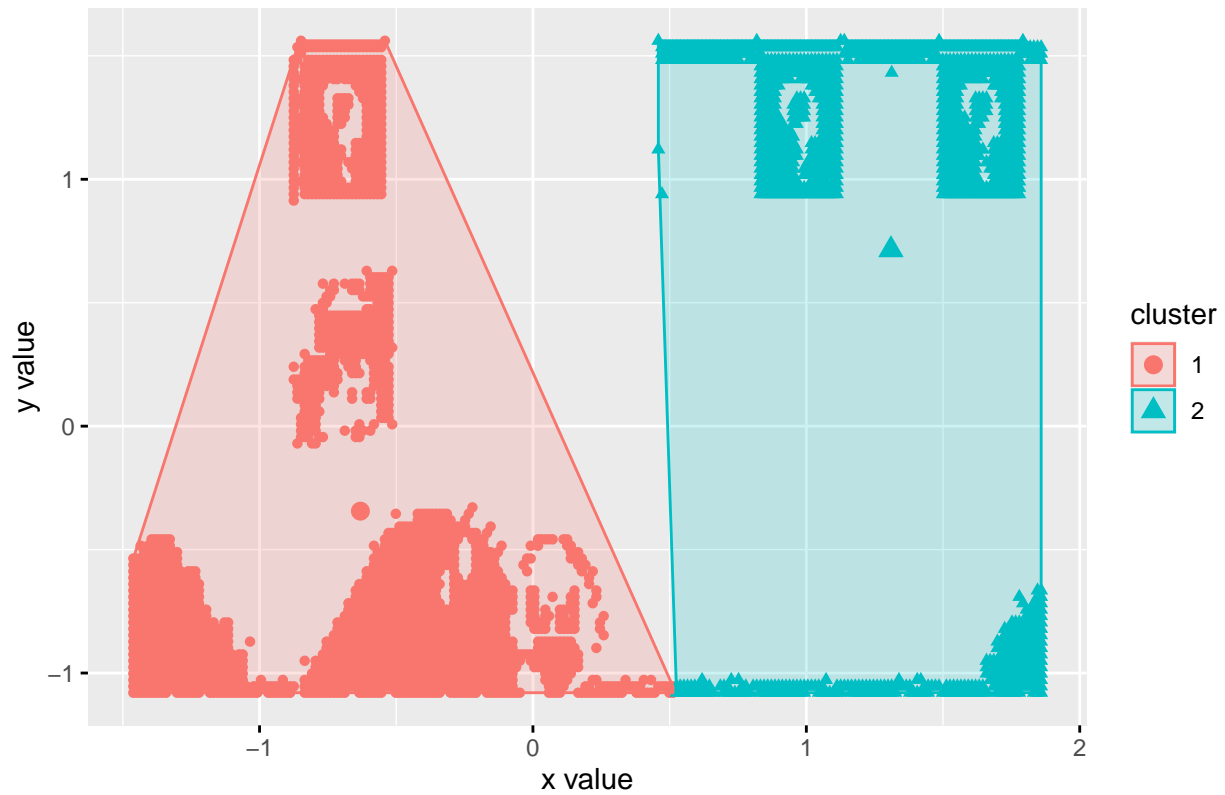
```
cluster_dm <- as.matrix(cluster_df)
cluster_k2 <- kmeans(cluster_dm, 2)
cluster_k3 <- kmeans(cluster_dm, 3)
cluster_k4 <- kmeans(cluster_dm, 4)
cluster_k5 <- kmeans(cluster_dm, 5)
cluster_k6 <- kmeans(cluster_dm, 6)
cluster_k7 <- kmeans(cluster_dm, 7)
cluster_k8 <- kmeans(cluster_dm, 8)
cluster_k9 <- kmeans(cluster_dm, 9)
cluster_k10 <- kmeans(cluster_dm, 10)
```

```
cluster_k11 <- kmeans(cluster_dm, 11)
cluster_k12 <- kmeans(cluster_dm, 12)
```

```
library(factoextra)
```

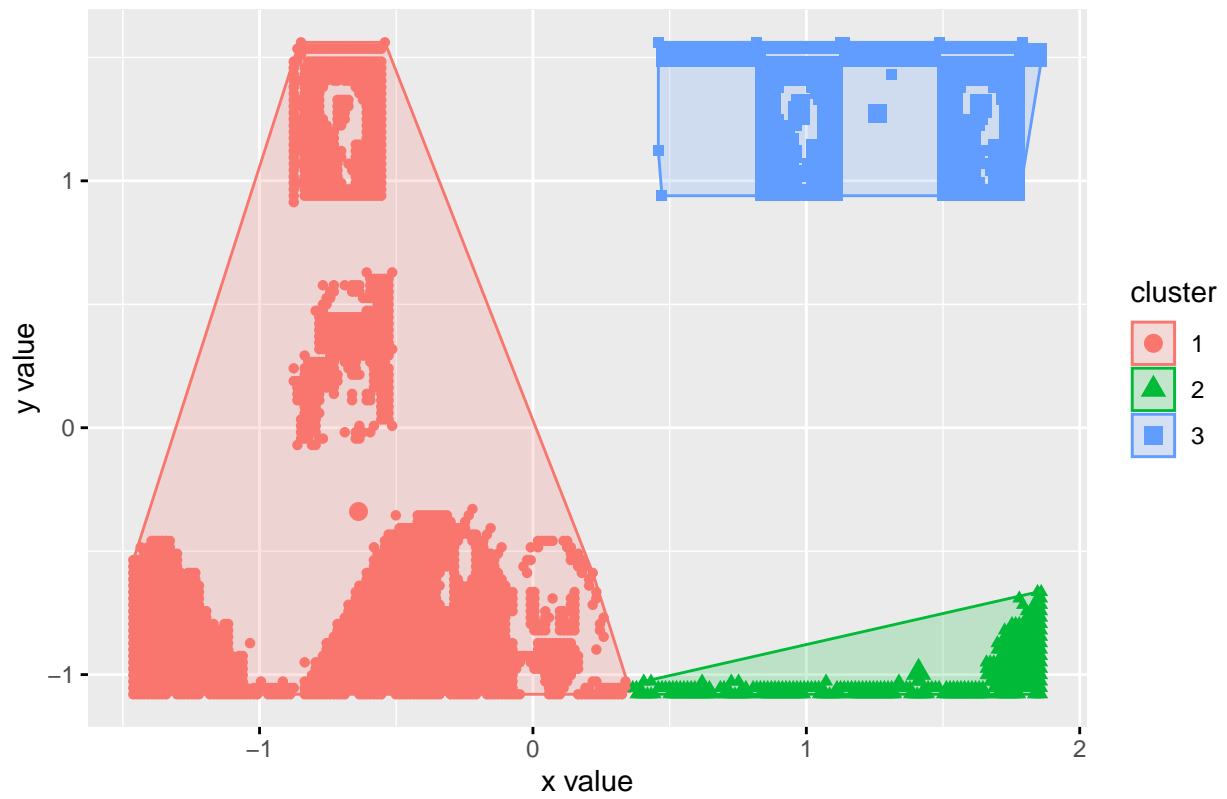
```
fviz_cluster(cluster_k2, cluster_df, geom = "point")
```

Cluster plot



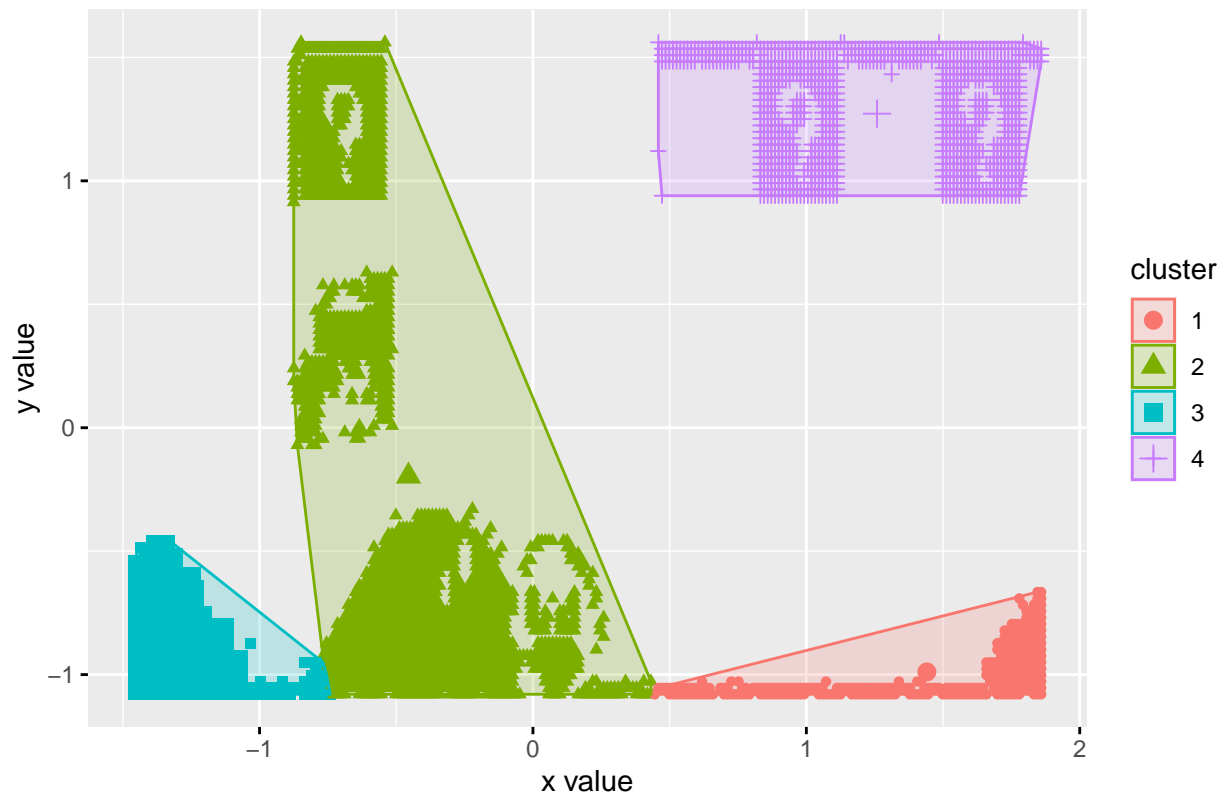
```
fviz_cluster(cluster_k3, cluster_df, geom = "point")
```

Cluster plot



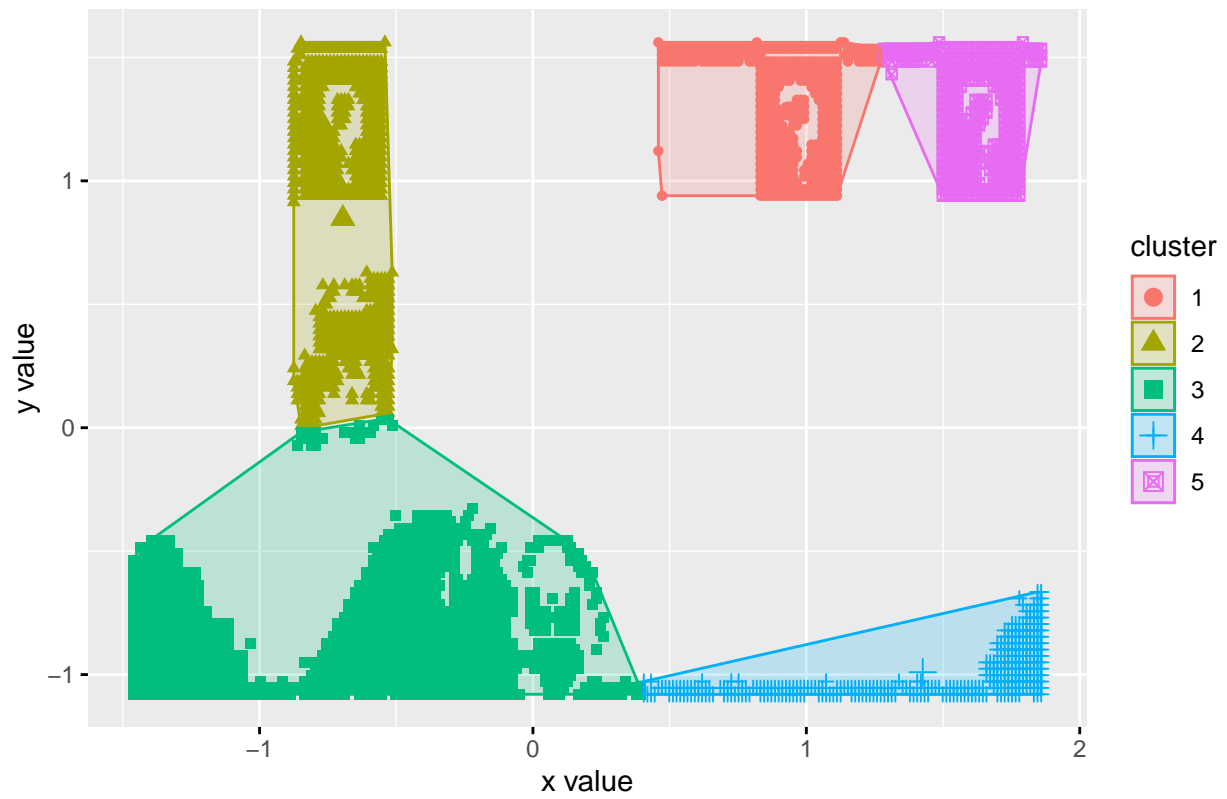
```
fviz_cluster(cluster_k4, cluster_df, geom = "point")
```

Cluster plot



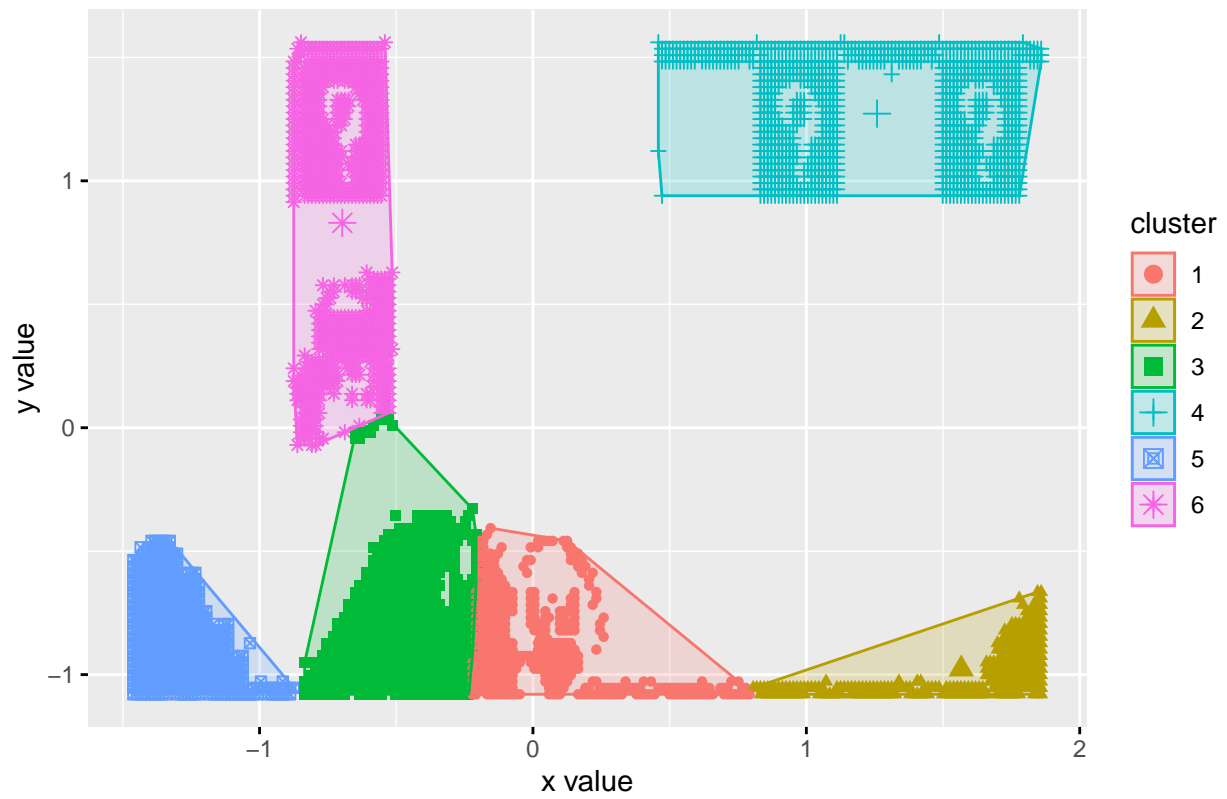
```
fviz_cluster(cluster_k5, cluster_df, geom = "point")
```

Cluster plot



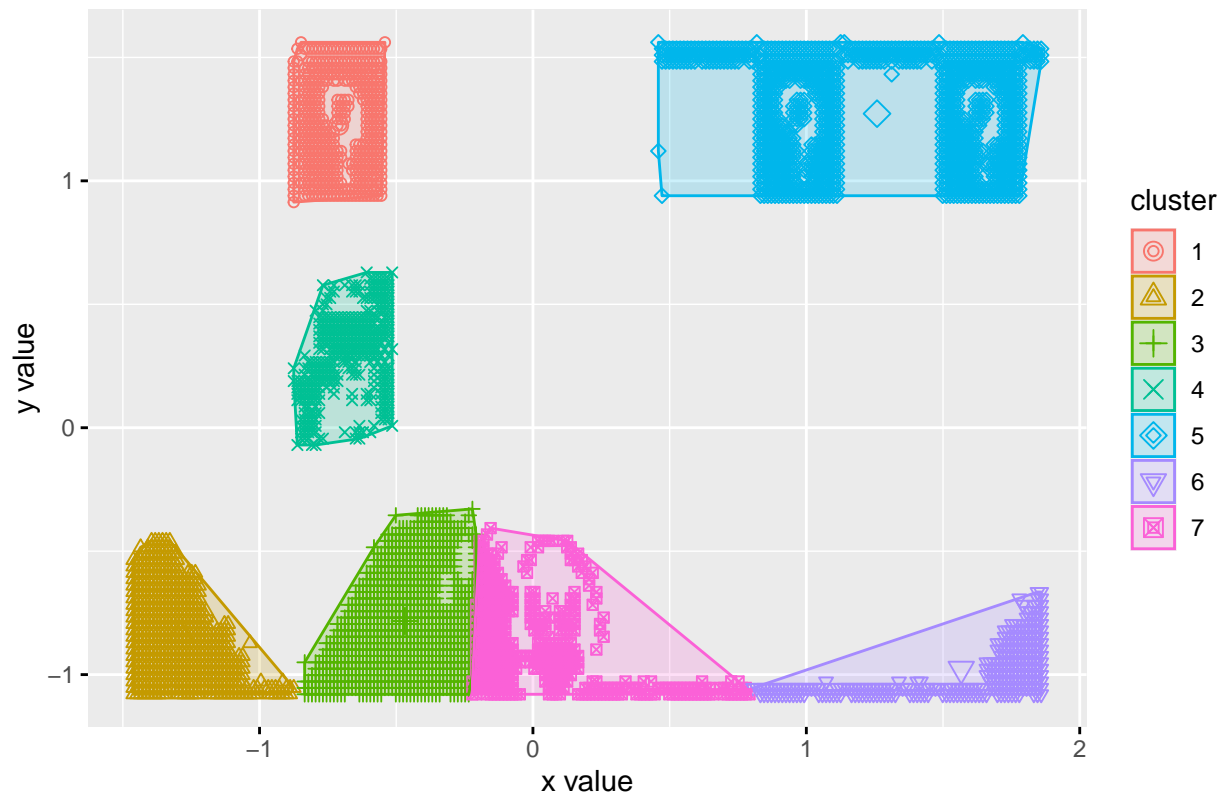
```
fviz_cluster(cluster_k6, cluster_df, geom = "point")
```


Cluster plot



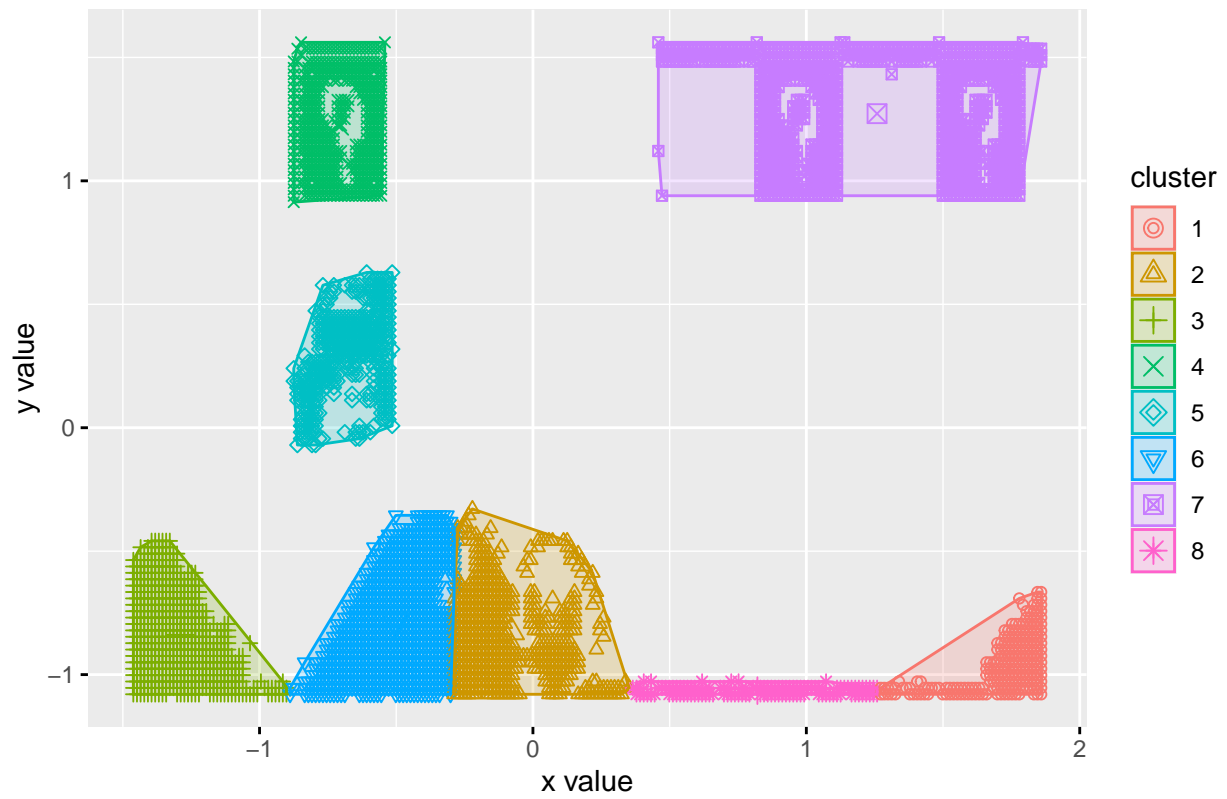
```
fviz_cluster(cluster_k7, cluster_df, geom = "point")
```

Cluster plot



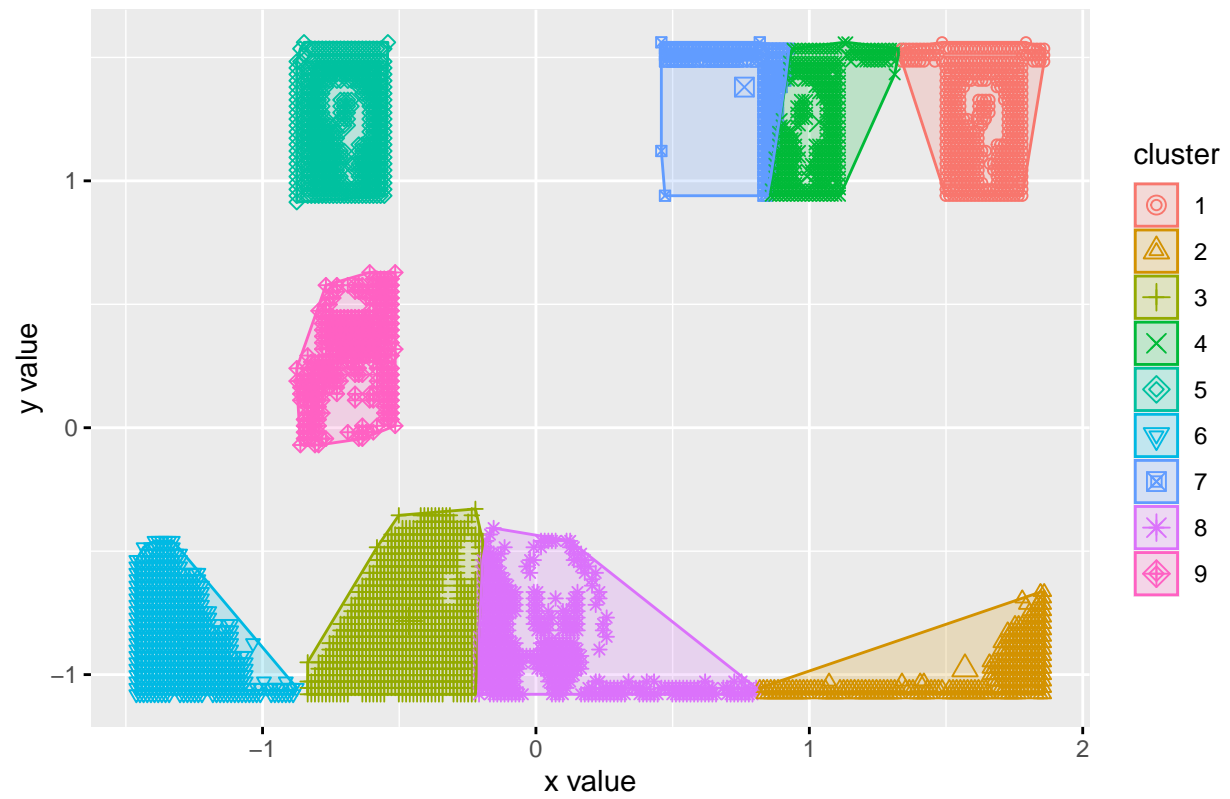
```
fviz_cluster(cluster_k8, cluster_df, geom = "point")
```

Cluster plot



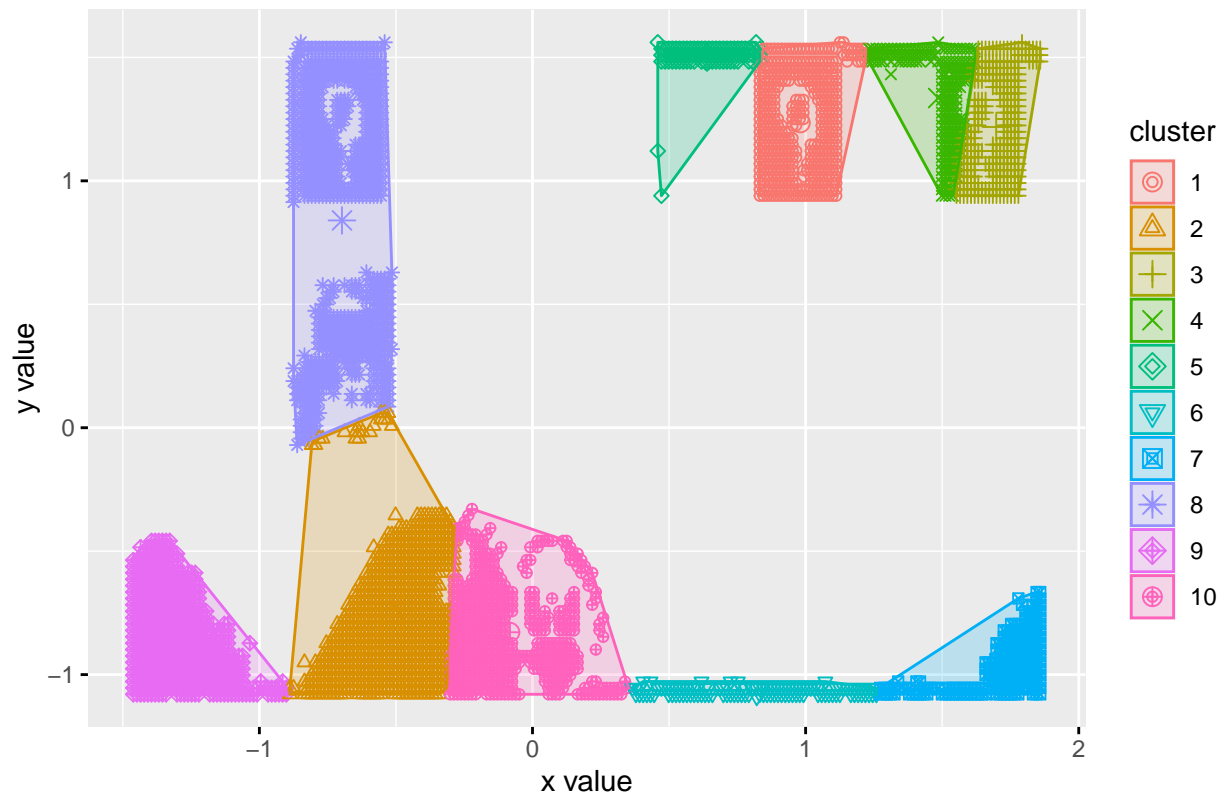
```
fviz_cluster(cluster_k9, cluster_df, geom = "point")
```

Cluster plot



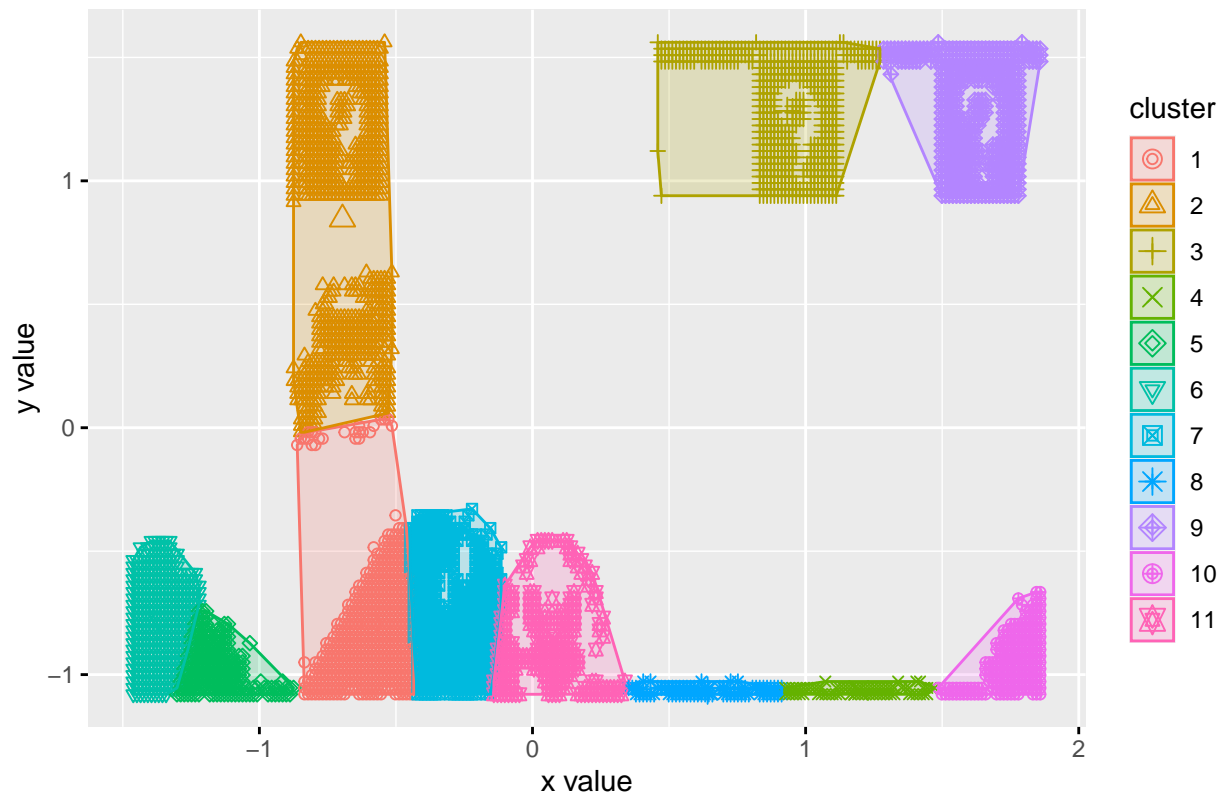
```
fviz_cluster(cluster_k10, cluster_df, geom = "point")
```

Cluster plot



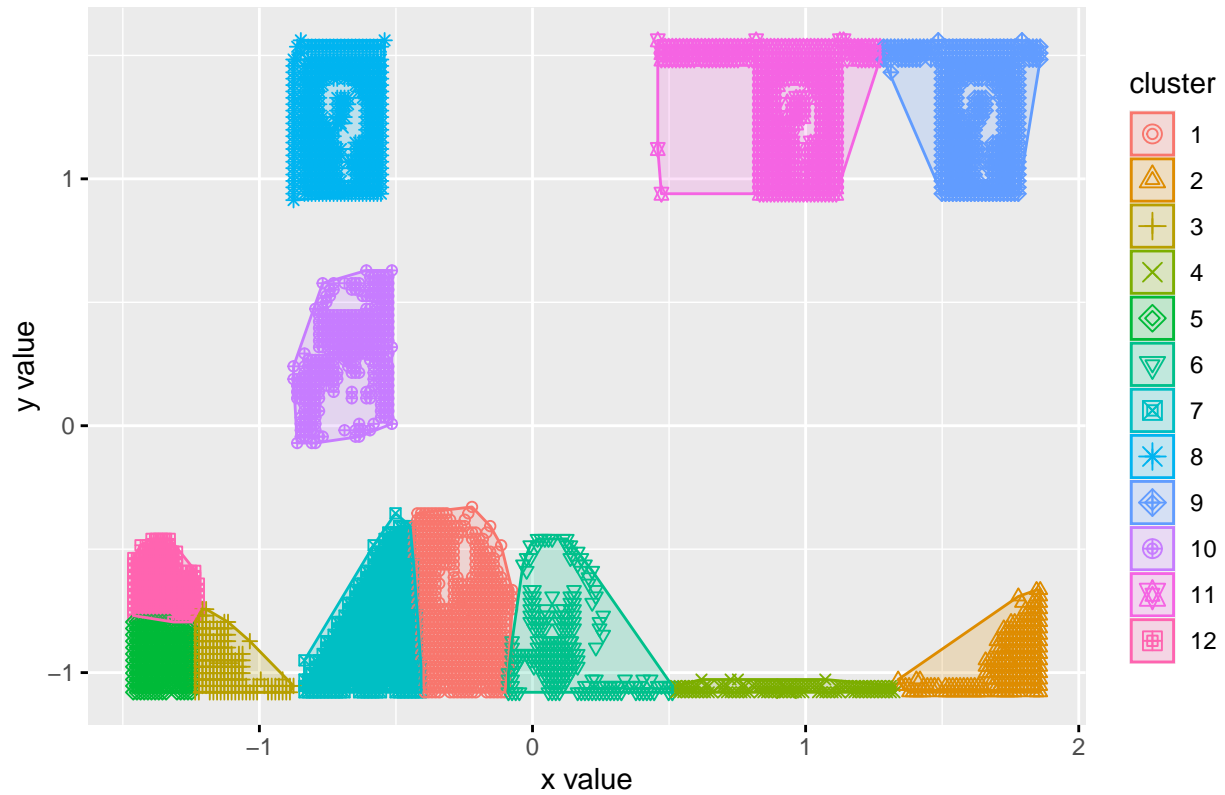
```
fviz_cluster(cluster_k11, cluster_df, geom = "point")
```

Cluster plot



```
fviz_cluster(cluster_k12, cluster_df, geom = "point")
```

Cluster plot



Professor, this again appears to be the point of my limitation in this assignment. After plotting the kmeans cluster results, I am at a loss as to how to proceed through the next step. I am continuing to search the internet for a solution, but at this time I again must be either overlooking or misunderstanding something from the textbook and additional readings.