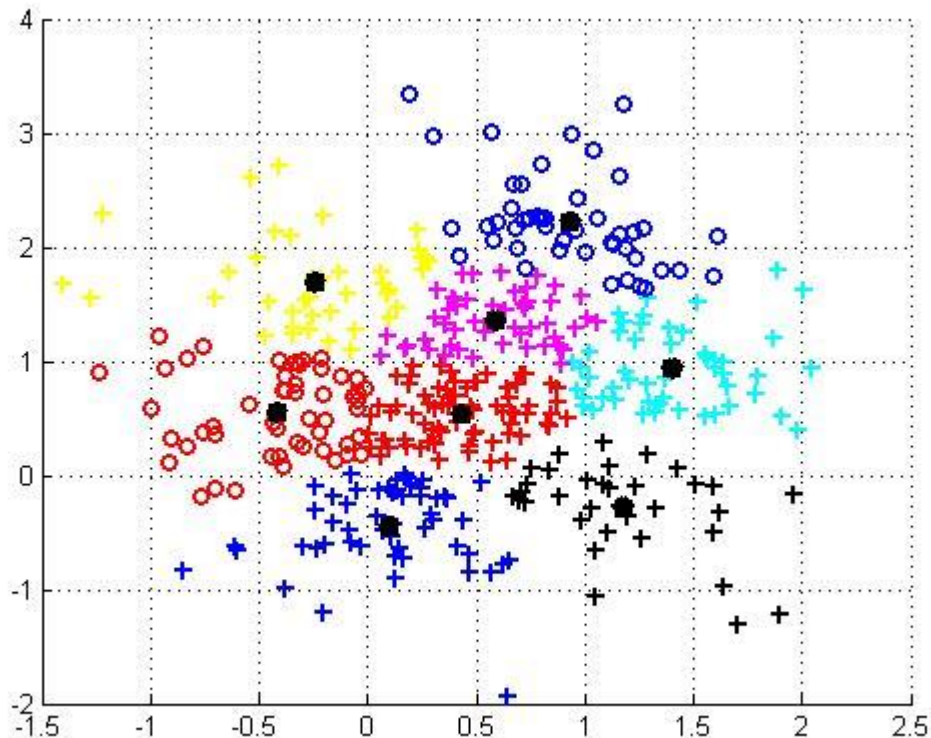# Exercise 3.1.3

Perform K-means clustering on each cipher individually for the training data from all the available datasets (disjunct). Represent the training data as a number of cluster centroids and compare performance, try multiple cluster sizes.

e.g.
Data of cypher '0'

Use the 7 centroids to represent 1000 data for training

Performance here means the KNN accuracy and time

Exercise 3.3.1

Plot the precision-recall curves for 1 to 13 "k" with "l" values up to the "k" value. Here, the results should be one plot containing "k" lines, and each one have "k" datapoints.

```r
for(k in 1:13) {
for(i in 1:k){
    id_test_pred <- knn(train = id_train, test = id_test, cl = id_train_labels, k=k, l=i)
# this is the task about: train 'l' up to 'k'
    # then you can calculate true positive and false positive
    res = truP_falP(id_test_pred, id_test_labels)
    # then you can calculate  precision and recall
    rec <-  res$truePositive/length(id_test_pred)
    pre <- res$truePositive/(res$truePositive + res$falsePositive)
    # then you can plot it
    …
  }
}
```

Exercise 4

When you work on raw data, you may need to generate a formula to link the 'cypher' to 324 pixels.

```
model.randomforest <- randomForest(V1 ~ . , data = id)
```