

Clustering

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

data(iris)

View(iris)

table(iris$Species)

##
##      setosa versicolor  virginica
##         50         50         50

iris_new = iris %>% select(-Species)
View(iris_new)

iris_model = kmeans(iris_new, centers = 3)
iris_model$cluster

##   [1] 1 2 2 2 1 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 1 1 1
##   [36] 1 1 1 2 1 1 2 2 1 1 2 1 2 1 1 3 3 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3
##   [71] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3
##  [106] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
##  [141] 3 3 3 3 3 3 3 3 3 3

length(iris_model$cluster)

## [1] 150

table(iris_model$cluster)
```

```

##
##  1  2  3
## 33 21 96

iris_model

## K-means clustering with 3 clusters of sizes 33, 21, 96
##
## Cluster means:
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1    5.175758    3.624242    1.472727    0.2727273
## 2    4.738095    2.904762    1.790476    0.3523810
## 3    6.314583    2.895833    4.973958    1.7031250
##
## Clustering vector:
##   [1] 1 2 2 2 1 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 1 1 1
##   2
##  [36] 1 1 1 2 1 1 2 2 1 1 2 1 2 1 1 3 3 3 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3
##   3
##  [71] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3
##   3
## [106] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
##   3
## [141] 3 3 3 3 3 3 3 3 3 3
##
## Within cluster sum of squares by cluster:
## [1]  6.432121 17.669524 118.651875
## (between_SS / total_SS =  79.0 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"

```