Pokemon

```
library(readr)
Pokemon <- read.csv("Pokemon.csv")
View(Pokemon)
names(Pokemon)
```

```
##  [1] "X."         "Name"       "Type.1"     "Type.2"     "Total"
##  [6] "HP"         "Attack"     "Defense"    "Sp..Atk"    "Sp..Def"
## [11] "Speed"      "Generation" "Legendary"
```

```
Pokemon_data <- as.matrix(Pokemon[,c(6:11)])
```

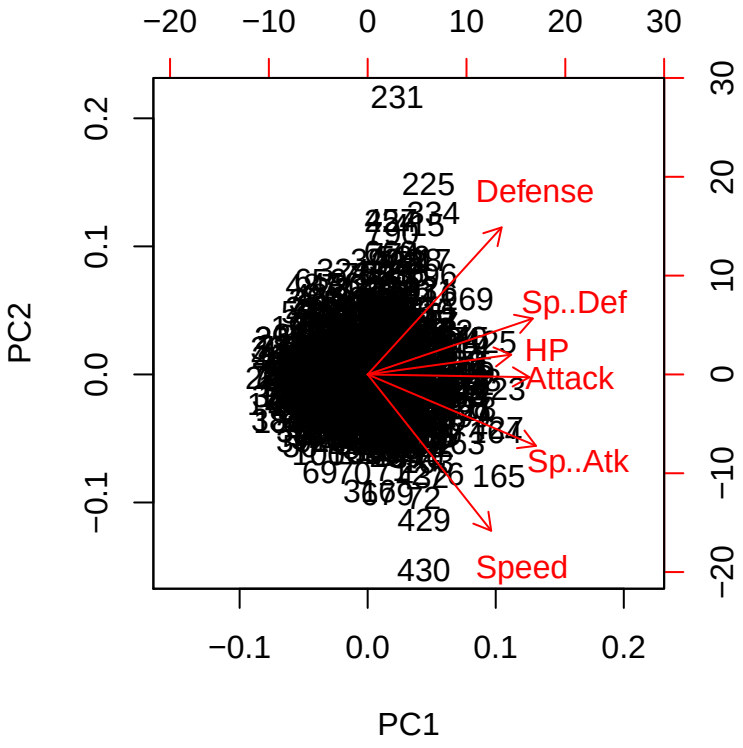
```
##Peform PCA
colMeans(Pokemon_data)
```

```
##       HP   Attack  Defense  Sp..Atk  Sp..Def    Speed
## 69.25875 79.00125 73.84250 72.82000 71.90250 68.27750
```

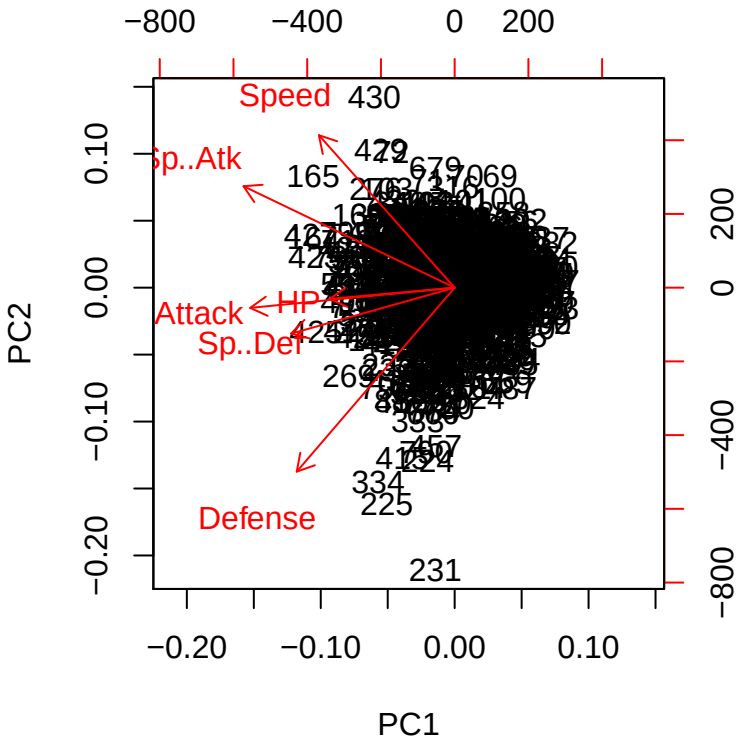
```
pok.pr <- prcomp(Pokemon_data,scale=TRUE)

summary(pok.pr)
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4     PC5     PC6
## Standard deviation     1.6466 1.0457 0.8825 0.8489 0.65463 0.51681
## Proportion of Variance 0.4519 0.1822 0.1298 0.1201 0.07142 0.04451
## Cumulative Proportion  0.4519 0.6342 0.7640 0.8841 0.95549 1.00000
```

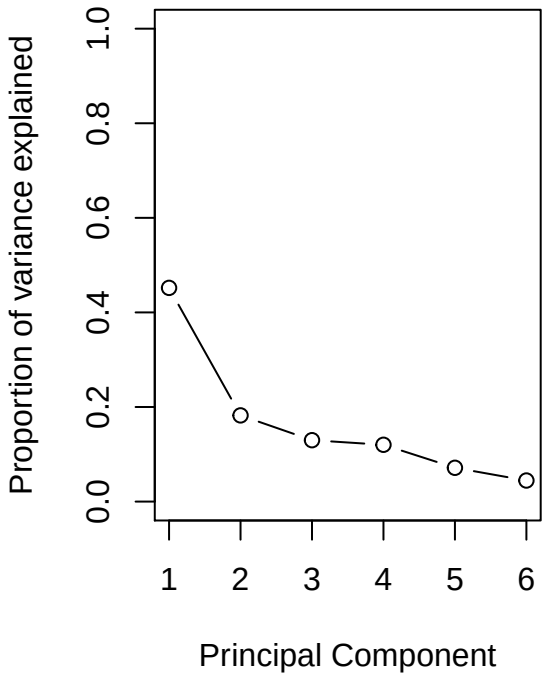
```
# Interpreting PCA results with scale=TRUE
biplot(pok.pr)
```



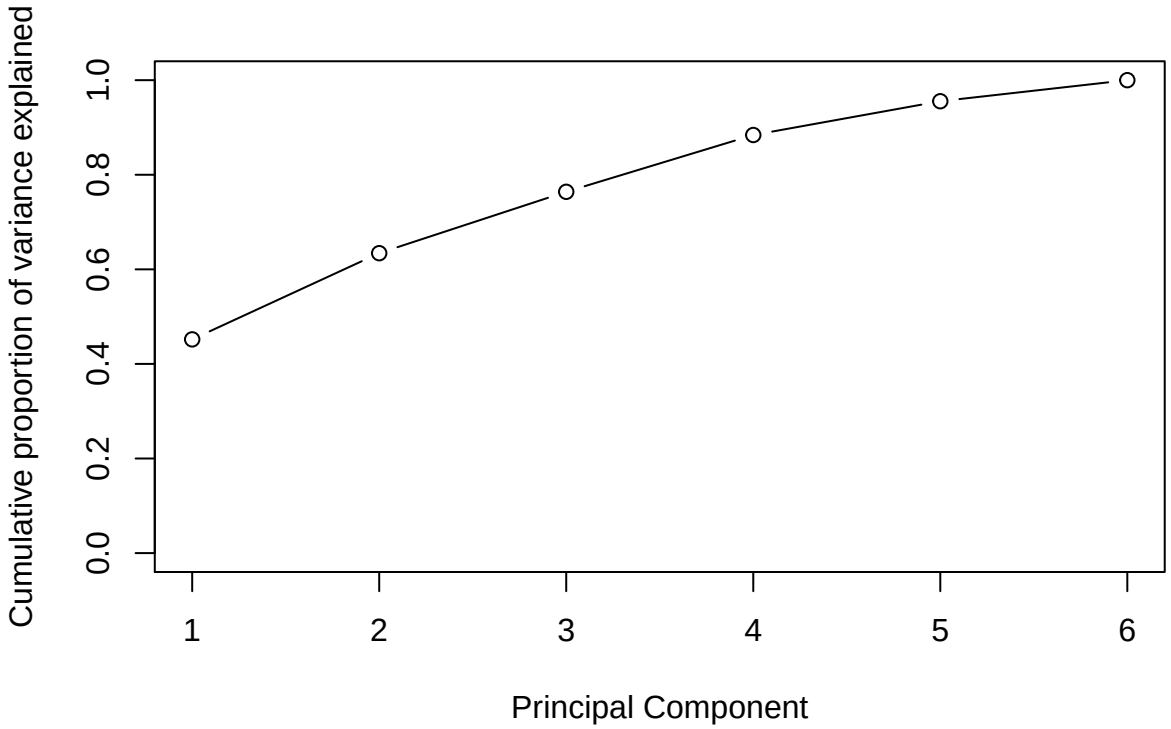
```r
#Interpreting PCA results with scale=FALSE
pok.pr.ns <- prcomp(Pokemon_data,scale=FALSE)
biplot(pok.pr.ns)
```



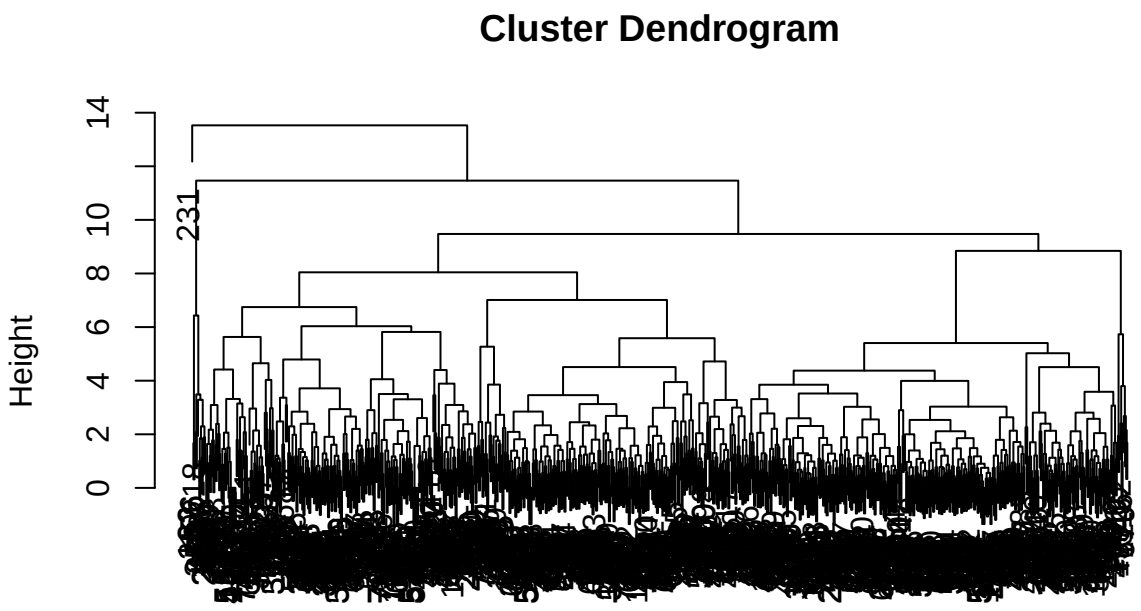

```r
##Calculating variability of each component
pr.var <- pok.pr$sdev ^2

#Variance explained by each PC
pve <- pr.var/sum(pr.var)

# Plotting variance explained by each component
par(mfrow=c(1,2))
plot(pve,xlab="Principal Component",ylab="Proportion of variance explained",ylim=c(0,1),type="b")
```




```
plot(cumsum(pve),xlab="Principal Component" ,ylab="Cumulative proportion of variance explained",ylim=c(
```

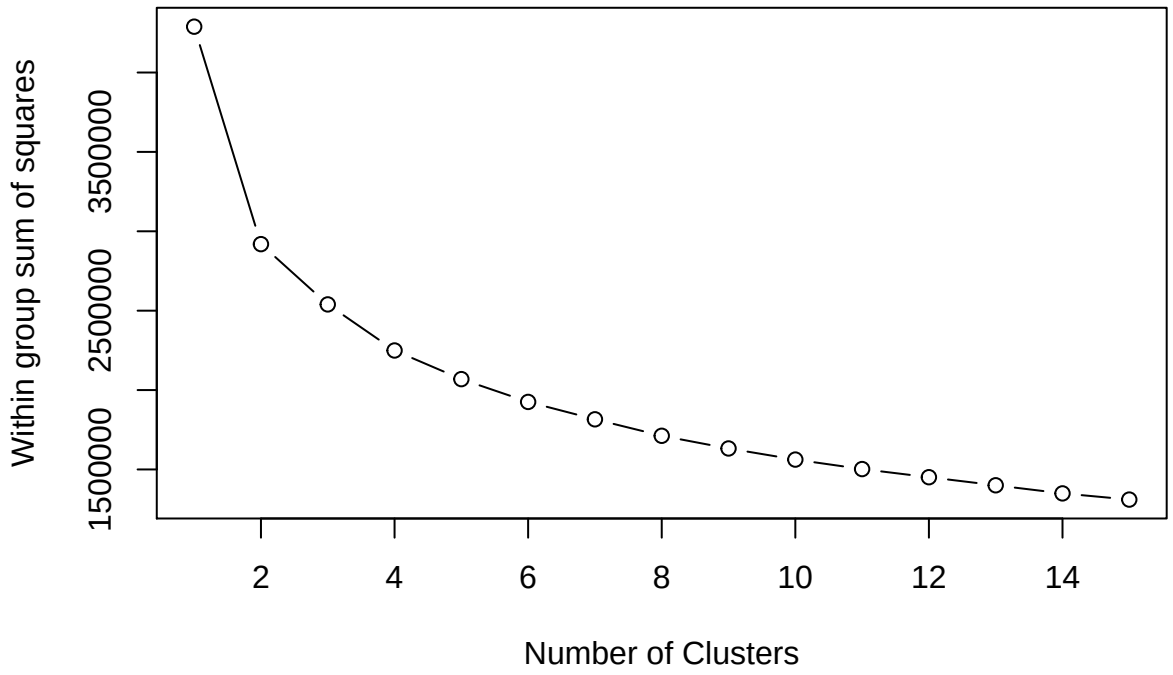



```
#Performing hierarchical clustering using euclidean distance
scaled_data <- scale(Pokemon_data)

Pokemon_hclust<- hclust(dist(scaled_data),method="complete")

plot(Pokemon_hclust)
```




```
# selecting number of clusters
Pokemon_hclust_clusters <- cutree(Pokemon_hclust,k=4)
```

```
#intialize wss
wss <- 0
for(i in 1:15){

  km.out<- kmeans(Pokemon_data,centers = i,nstart = 20,iter.max=50)
  wss[i] <- km.out$tot.withinss
}
plot(1:15,wss,type="b",xlab="Number of Clusters",ylab="Within group sum of squares")
```

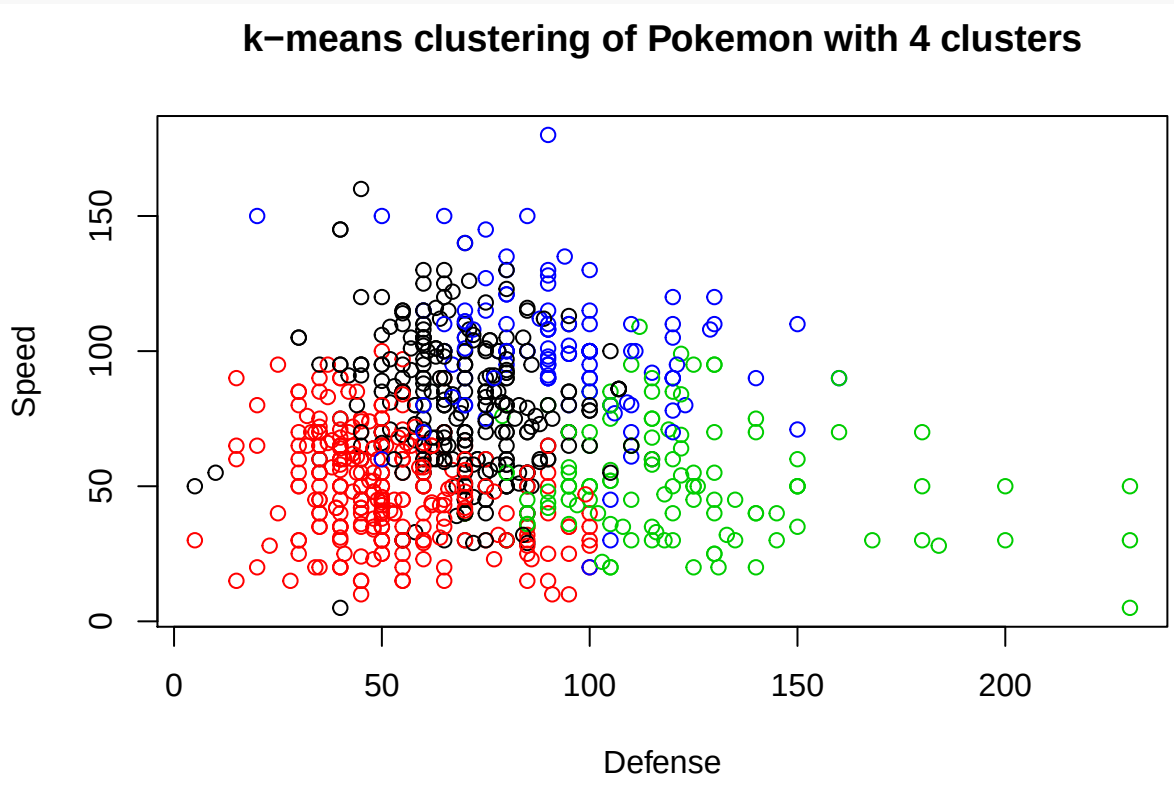



```
k <- 4

# Build model with k clusters: km.out
km.out <- kmeans(Pokemon_data, centers = k, nstart = 20, iter.max = 50)
km.out
```

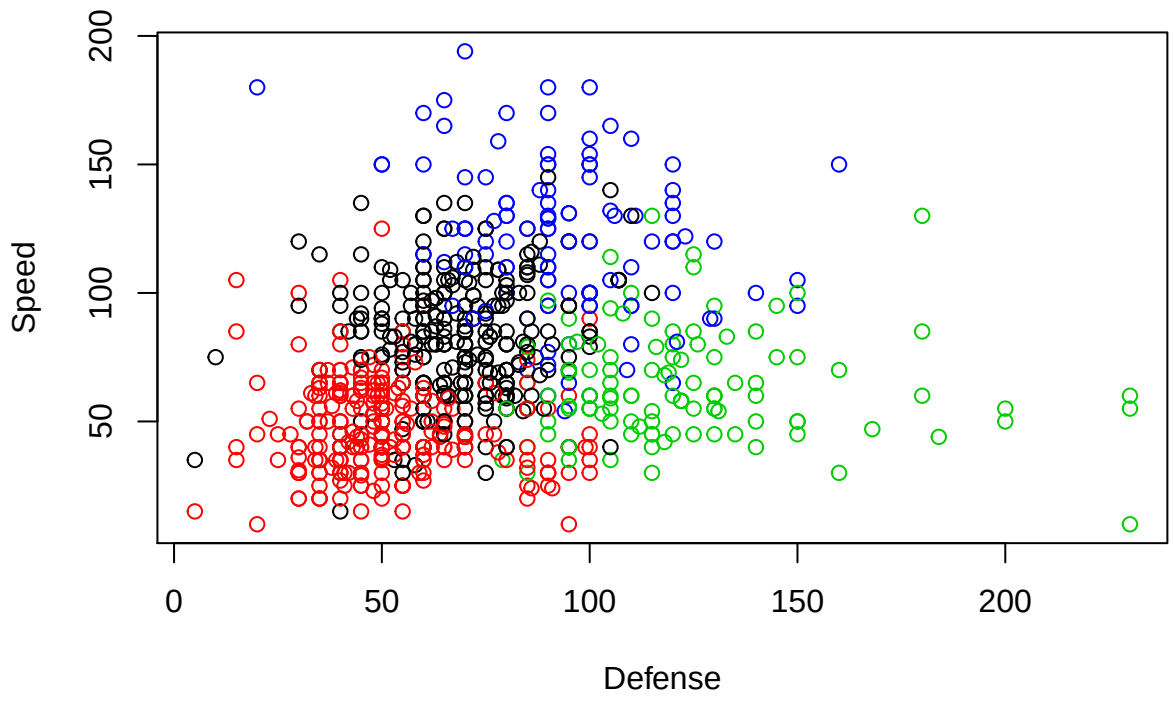
```
## K-means clustering with 4 clusters of sizes 288, 283, 115, 114
##
## Cluster means:
##         HP    Attack   Defense   Sp..Atk  Sp..Def     Speed
## 1 79.18056  81.31944  69.19097  82.01042 77.53125  80.10417
## 2 50.29682  54.03180  51.62898  47.90459 49.15548  49.74912
## 3 71.30435  92.91304 121.42609  63.89565 88.23478  52.36522
## 4 89.20175 121.09649  92.73684 120.45614 97.67544 100.44737
##
## Clustering vector:
##   [1] 2 1 1 4 2 1 1 4 4 2 1 1 4 2 2 1 2 2 1 1 2 2 1 4 2 1 2 1 2 1 2 1 2 3 2
##  [36] 2 1 2 2 1 2 1 2 1 2 1 2 1 2 1 1 2 3 2 1 2 1 2 1 2 1 2 1 2 4 2 2 1 2 1
##  [71] 1 4 2 1 3 2 1 1 2 1 2 3 3 1 1 2 3 3 2 1 2 2 1 2 1 2 1 2 3 2 1 1 4 3 2
## [106] 1 2 3 2 1 2 1 2 3 1 3 2 2 3 2 3 1 1 1 4 2 1 2 1 2 1 1 1 1 1 1 3 4 1 2
## [141] 1 4 1 2 2 1 1 4 1 2 3 2 3 1 4 1 4 4 4 2 1 4 4 4 4 4 2 1 1 2 1 1 2 1 1
## [176] 2 1 2 1 2 1 2 2 1 2 1 2 2 2 2 1 2 1 2 2 1 4 3 2 1 3 1 2 2 1 2 2 1 1 2
## [211] 3 1 3 1 1 1 2 1 1 2 3 1 3 3 3 2 1 1 3 3 3 1 3 1 2 1 2 3 2 1 2 2 1 2 1
## [246] 3 2 1 4 1 2 3 1 1 2 2 3 2 2 2 1 1 4 4 3 2 1 4 4 4 4 4 2 1 1 4 2 1 4 4
## [281] 2 1 1 4 2 1 2 1 2 2 1 2 2 2 2 1 2 2 1 2 1 2 1 2 2 1 4 2 1 2 1 2 1 4 2
## [316] 1 2 2 2 1 2 1 2 3 2 2 2 3 2 3 2 3 3 3 2 1 1 2 1 4 1 1 1 1 1 2 1 2 1 4
## [351] 1 1 2 1 4 3 2 1 2 2 2 1 2 1 2 1 4 1 1 1 1 2 1 2 1 2 3 2 3 2 3 2 1 1 3
## [386] 2 1 4 2 3 1 1 1 4 2 2 1 4 2 1 1 2 3 3 3 2 2 3 4 4 2 3 3 4 3 3 3 4 4 4
## [421] 4 4 4 4 4 4 4 4 4 4 3 4 2 3 3 2 1 1 2 1 1 2 2 1 2 1 2 2 2 2 1 2 1 2 1
## [456] 3 3 2 3 3 3 1 2 3 1 2 1 2 1 2 1 1 2 1 2 1 4 1 1 2 1 2 2 1 2 3 2 2 2 1
## [491] 3 2 1 4 4 1 2 4 4 2 3 2 3 2 1 1 2 1 2 2 1 4 1 3 3 3 3 4 4 1 1 3 1 3 1
## [526] 1 1 4 3 3 1 1 1 1 1 1 1 3 4 4 4 4 4 4 4 4 3 1 4 4 4 4 4 4 2 1 1 2 1 1
## [561] 2 1 1 2 1 2 2 3 2 1 2 1 2 1 2 1 2 1 2 2 1 2 1 2 3 3 2 1 2 1 1 3 2 3 3
```

```
## [596] 2 1 1 3 1 2 3 1 2 2 1 2 1 2 1 1 2 2 1 2 1 1 1 2 3 2 3 1 2 3 3 3 1 4 2
## [631] 1 2 1 2 1 2 1 1 2 2 1 2 1 2 1 1 2 1 1 2 3 2 1 2 1 1 2 1 2 3 2 3 3 2 1
## [666] 1 2 1 2 2 1 2 1 4 2 1 1 2 1 1 2 1 3 2 3 2 3 3 2 1 2 3 1 3 2 1 4 2 1 4
## [701] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 3 3 2 1 1 2 1 1 2 1 2 2 1 2 2 1
## [736] 2 1 2 2 1 2 1 2 1 1 2 1 1 2 3 4 3 2 1 2 1 2 1 2 3 2 3 2 1 2 1 2 3 2 1
## [771] 1 1 1 3 2 1 4 1 2 1 2 2 2 2 3 3 3 3 2 3 2 1 4 4 4 3 4 4 4 4
##
## Within cluster sum of squares by cluster:
## [1] 871531.3 513476.6 455965.7 408271.9
##  (between_SS / total_SS =  47.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```r
# Plot of Defense vs. Speed by cluster membership
plot(Pokemon_data[, c("Defense", "Speed")],
     col = km.out$cluster,
     main = paste("k-means clustering of Pokemon with", k, "clusters"),
     xlab = "Defense", ylab = "Speed")
```




```r
# Plot of Defense vs. Sp. Atk by cluster membership
plot(Pokemon_data[,3], Pokemon_data[,4],
     col = km.out$cluster,
     main = paste("k-means clustering of Pokemon with", k, "clusters"),
     xlab = "Defense", ylab = "Speed")
```




```
table(km.out$cluster,Pokemon_hclust_clusters)
```

```
##    Pokemon_hclust_clusters
##       1   2   3   4
##   1  32 245  11   0
##   2 277   6   0   0
##   3  10 104   0   1
##   4   5 109   0   0
```

```
# Create a hierarchical clustering model: wisc.pr.hclust
pokemon.pr.hclust <- hclust(dist(pok.pr$x[, 1:6]), method = "complete")
```

```
# Cut model into 4 clusters: wisc.pr.hclust.clusters
pokemon.pr.hclust.clusters <- cutree(pokemon.pr.hclust, k = 4)
```