

IS5126 Guided Project

IS5126 Hands-on with Applied Analytics

Team 1: Neal Wong, Nicholas Leong, Sean Xu Shunjie, Ong Yi Long, Nigel Ngan Zisheng

Due by: Mar 07, 2022 5:00 PM

```
# load required packages
library(dplyr)
library(wooldridge)
library(rcompanion)
library(readxl)
library(tidyr)
library(ggplot2)
library(car)
library(sandwich)
library(lmtest)
library(caret)
library(pca3d)
library(rms)
library(leaps)
library(psych)
library(animation)
library(factoextra)
```

Part C: Data Analytics (50 points)

- Dataset required: `data('out')`

In this section, you need to perform data analytics techniques on the prepared data set, with either R or Python, and justify/interpret the model. You should have a basic understanding of basketball such as how the game is played, the role of each position, and basic information on how players are recruited, drafted, for each team. Answer the following questions.

```
nba = read.csv(file = './data/output_final.csv', header= TRUE)
```

Details 1. Clustering. Which players are similar? (10 points)

- a. Apply `k-means` algorithm.

```
#nba_salary = subset(nba, select = c(salary,tm,age,fg))

# nba_rmnull = nba[-which(sapply(nba, is.null))]
# nba_unlist <- as.numeric(unlist(nba_rmnull))
```

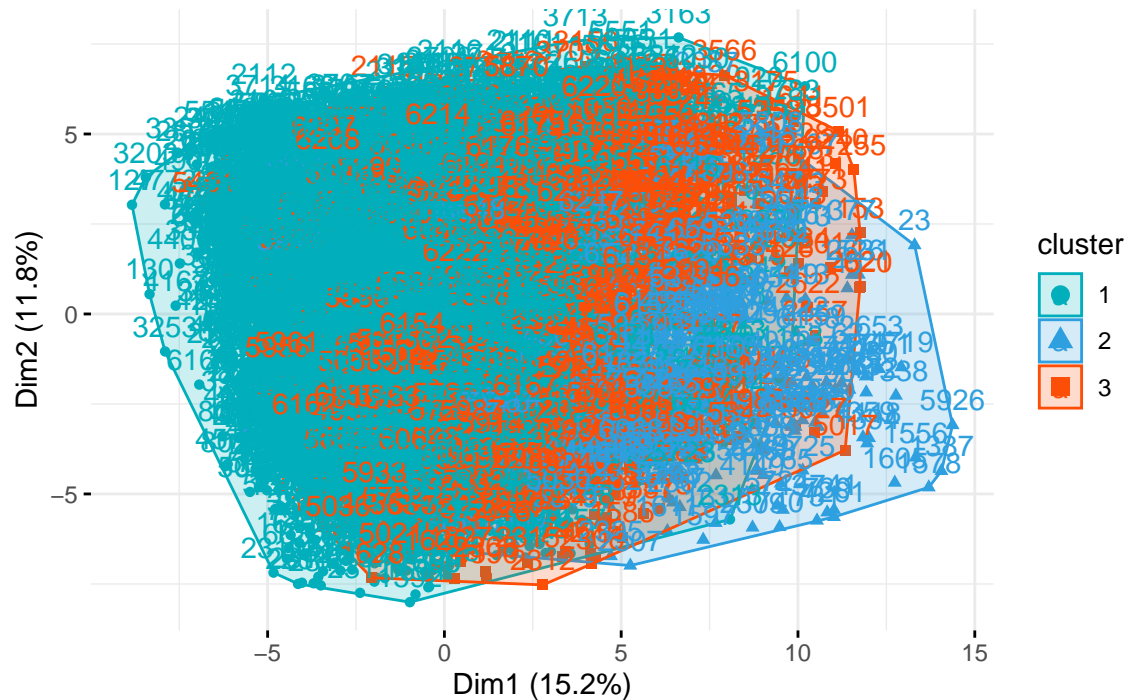
```

km_nba = kmeans(nba, center = 3, nstart = 10)

fviz_cluster(km_nba, data = nba,
  palette = c("#00AFBB", "#2E9FDF", "#FC4E07"),
  ggtheme = theme_minimal(),
  main = "Three clusters on the plane of first two PCs of 'nba'.")

```

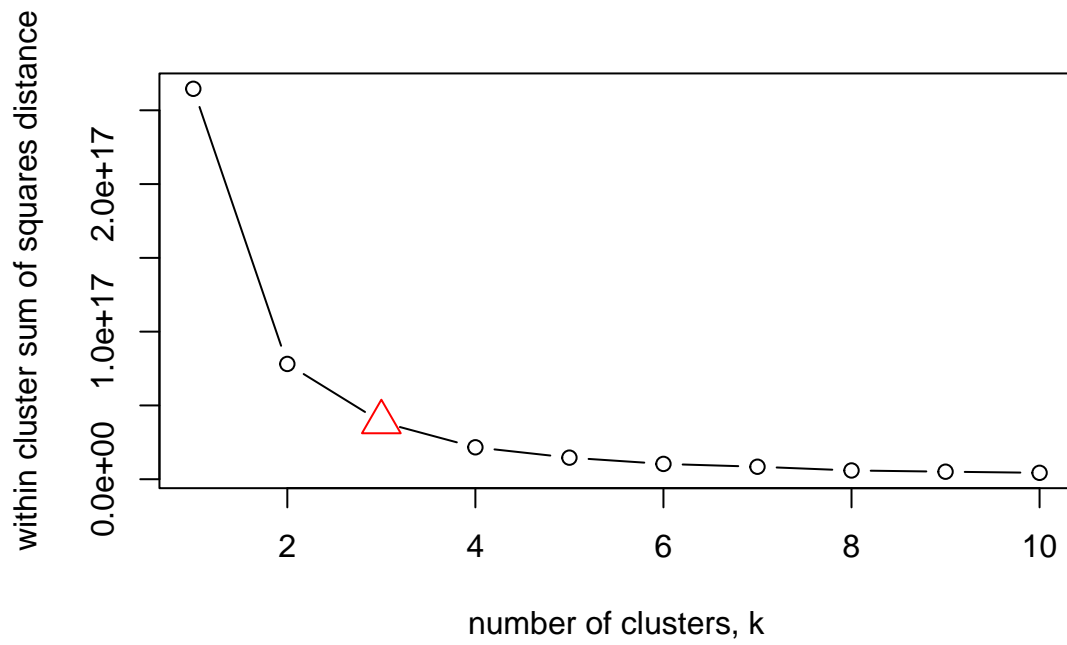
Three clusters on the plane of first two PCs of 'nba'.



```

# plot the ``within-cluster sum of squares distance'' as a function of the number of clusters
wss = rep(NA, 10)
for (k in c(1:10)){
  wss[k] = kmeans(nba, k, nstart = 10)$tot.withinss
}
cexs = rep(1, length(wss))
cexs[3] = 2
pchs = rep(1, length(wss))
pchs[3] = 2
cols = rep('black', length(wss))
cols[3] = 'red'
plot(wss, type = 'b', xlab = 'number of clusters, k', ylab = 'within cluster sum of squares distance',

```



when choosing the ``best'' number of clusters, look for the ``elbow''! In this case, either $k = 3$ or $k = 4$