INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Lab Report

Course: Data Analytics in R (CS6E23L)

Course Instructor: Dr. Kavi Mahesh

Lab Instructor: Pragya Verma

By:

Paka Sravan Kumar Yadav

6th Semester

3rd Year

16CS11

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Lab – 02

Cricket Data set - Batting

```
odibatting2007 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2007odibattingrating.csv")
odibatting2008 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2008odibattingrating.csv")
odibatting2009 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2009odibattingrating.csv")
odibatting2010 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\Player Ratings\\2010odibattingrating.csv")
odibatting2011 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2011odibattingrating.csv")
odibatting2012 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2012odibattingrating.csv")
odibatting2013 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2013odibattingrating.csv")
odibatting2014 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2014odibattingrating.csv")
odibatting2015 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2015odibattingrating.csv")
odibatting2016 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2016odibattingrating.csv")

testbatting2007 <- read.csv G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cric
ket Data Set\\Player Ratings\\2007testbattingrating.csv")
testbatting2008 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2008testbattingrating.csv")
testbatting2009 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2009testbattingrating.csv")
testbatting2010 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2010testbattingrating.csv")
testbatting2011 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2011testbattingrating.csv")
testbatting2012 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2012testbattingrating.csv")
testbatting2013 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2013testbattingrating.csv")
testbatting2014 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2014testbattingrating.csv")
testbatting2015 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
cket Data Set\\Player Ratings\\2015testbattingrating.csv")
testbatting2016 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\\Cri
```

```
cket Data Set\\Player Ratings\\2016testbattingrating.csv")

twentybatting2007 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2007twenty20battingrating.csv")
twentybatting2008 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2008twenty20battingrating.csv")
twentybatting2009 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2009twenty20battingrating.csv")
twentybatting2010 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2010twenty20battingrating.csv")
twentybatting2011 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2011twenty20battingrating.csv")
twentybatting2012 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2012twenty20battingrating.csv")
twentybatting2013 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2013twenty20battingrating.csv")
twentybatting2014 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2014twenty20battingrating.csv")
twentybatting2015 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2015twenty20battingrating.csv")
twentybatting2016 <- read.csv("G:\\Required\\6th Sem\\DA\\Lab\\Lab2\
\Cricket Data Set\\Player Ratings\\2016twenty20battingrating.csv")


library(dplyr)

##
## 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

dataBatting <- bind_rows(odibatting2007, odibatting2008, odibatting2009, o
dibatting2010,
                         odibatting2011, odibatting2012, odibatting2013, odibatting
2014,
                         odibatting2015, odibatting2016,
                         testbatting2007, testbatting2008, testbatting2009, testbatting
2010,
                         testbatting2011, testbatting2012, testbatting2013, testbatting
2014,
```

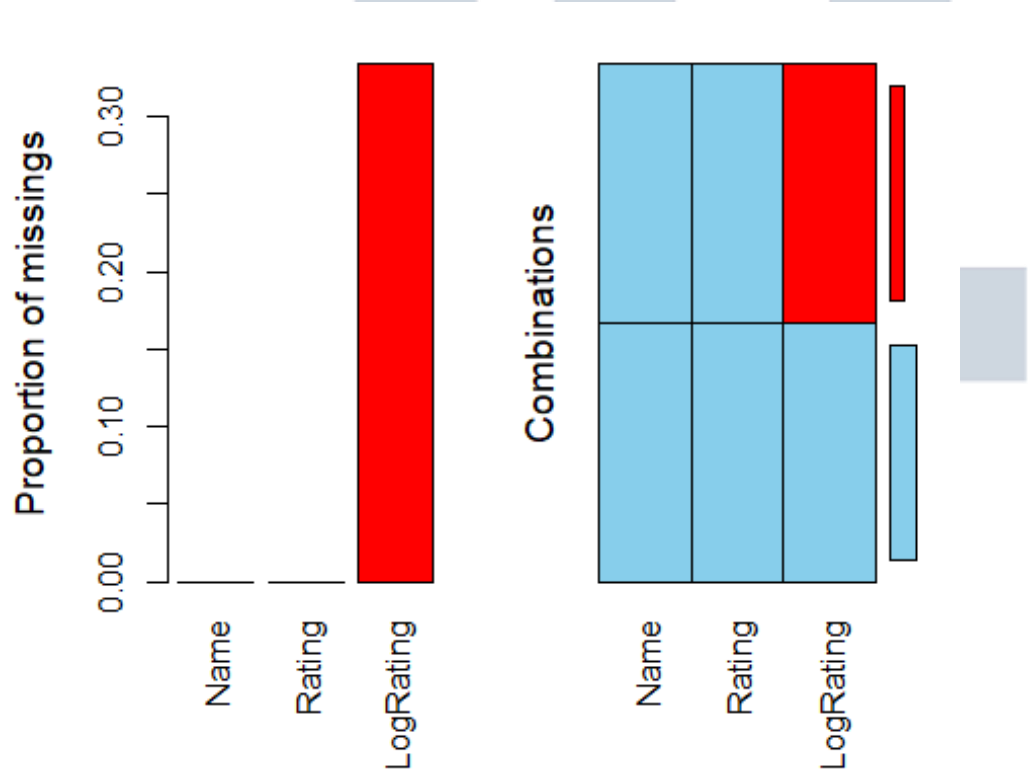
```
                    testbatting2015, testbatting2016,
                  twentybatting2007, twentybatting2008, twentybatting2009,tw
entybatting2010,
                  twentybatting2011,twentybatting2012,twentybatting2013,twe
ntybatting2014,
                  twentybatting2015,twentybatting2016)

summary(dataBatting)

##     Name              Rating       LogRating
##  Length:3000        Min.   : 10.0   Min.   :2.233
##  Class :character   1st Qu.:323.0   1st Qu.:2.608
##  Mode  :character   Median :460.0   Median :2.699
##                     Mean   :458.1   Mean   :2.688
##                     3rd Qu.:586.2   3rd Qu.:2.795
##                     Max.   :936.0   Max.   :2.971
##                                     NA's   :1001

library(VIM)

aggr(dataBatting)
```




```
dataBatting <- dataBatting %>%
  group_by(Name) %>%
  summarise(avg = mean(Rating))
```

```r
set.seed(20)

batcluster <- kmeans(dataBatting[, 2], 5)

batcluster$cluster <- as.factor(batcluster$cluster)

str(batcluster)
```

```
## List of 9
##  $ cluster     : Factor w/ 5 levels "1","2","3","4",..: 3 5 3 3 5 3 5 3 2 3 ...
##  $ centers     : num [1:5, 1] 628.3 247 81.3 468.8 364.6
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5] "1" "2" "3" "4" ...
##   .. ..$ : chr "avg"
##  $ totss       : num 10751533
##  $ withinss    : num [1:5] 142920 127651 111123 141721 129637
##  $ tot.withinss: num 653052
##  $ betweenss   : num 10098481
##  $ size        : int [1:5] 60 96 42 118 136
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```r
library(ggplot2)
ggplot(dataBatting, aes(dataBatting$Name, avg, color = batcluster$cluster)) +
  geom_point(size = 2) +
  scale_color_hue(labels = c("Good", "Avarage", "Useless", "Best", "Better")) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  ggtitle(" Batting Ratings(2007-2016)")
```