Player_Classify

Chenjie Li

11/11/2019

```
# firts, let's connect R with our Postgres databse:
library(RPostgreSQL)
```

```
## Loading required package: DBI
```

```
# create connection
con <- dbConnect(PostgreSQL(), user= "lchenjie", dbname="csp571")

# query to fetch players' stats

Q = "select a.*,s.*,r.*,i.*
from player_assists a, player_scoring s, player_rebounds r, player_info i
where a.player_name = s.player_name and a.season_name=s.season_name and a.team_name = s.team_name and
s.player_name = r.player_name and s.season_name=r.season_name and s.team_name = r.team_name and
r.player_name = i.player_name; "

# return results
player_raw <- dbGetQuery(con,Q)
```

```
# remove duplicate cols
players <- player_raw[, !duplicated(colnames(player_raw))]
```

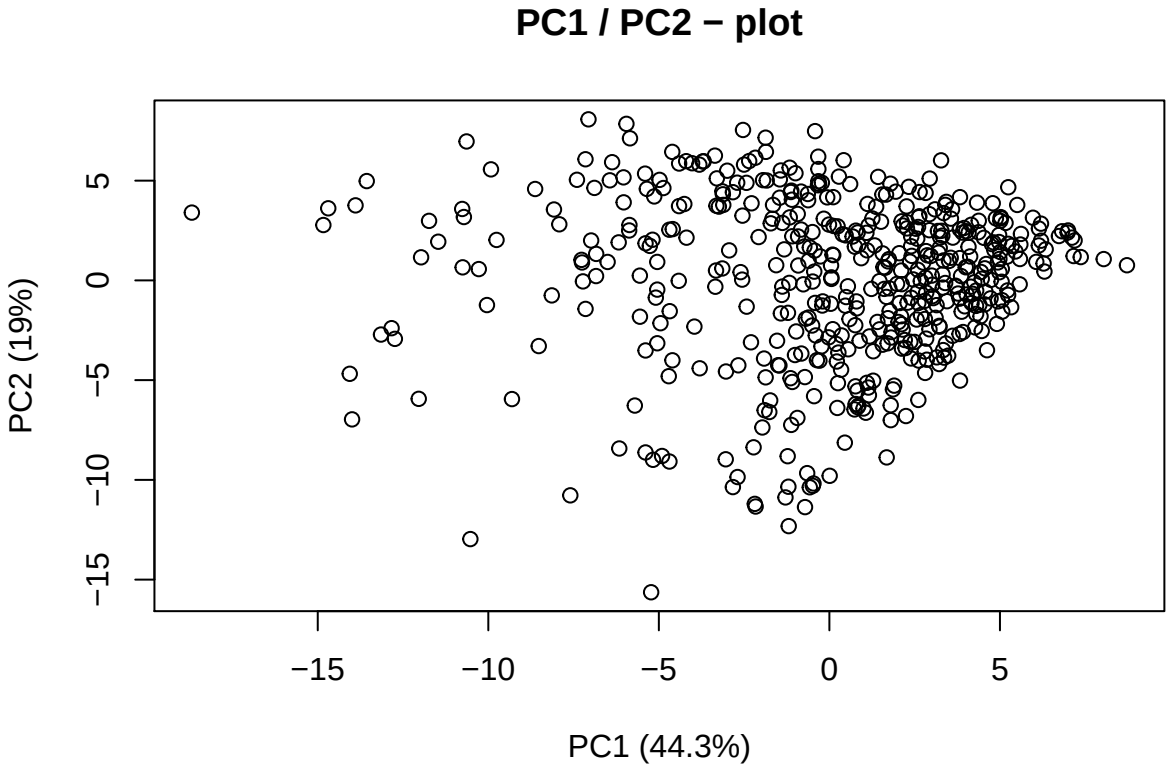
```
players_18 = players[players$season_name=='2018-19',]
```

```
#PCA for the data from season 2018-19
# players_18$position <- as.factor(players_18$position)
num_players_18 <- players_18[, sapply(players_18, class) != "character"]

for(i in 1:ncol(num_players_18)){
  num_players_18[is.na(num_players_18[,i]), i] <- mean(num_players_18[,i], na.rm = TRUE)
}

players_18.pr <-prcomp(num_players_18, center = TRUE, scale = TRUE)

summary(players_18.pr)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5    PC6
## Standard deviation     4.4363 4.0583 2.15870 1.78512 1.47881 1.3739
## Proportion of Variance 0.2982 0.2495 0.07061 0.04828 0.03313 0.0286
## Cumulative Proportion  0.2982 0.5477 0.61835 0.66663 0.69976 0.7284
##                           PC7     PC8     PC9    PC10    PC11   PC12
## Standard deviation     1.31367 1.27097 1.18799 1.14527 1.09021 0.9851
## Proportion of Variance 0.02615 0.02448 0.02138 0.01987 0.01801 0.0147
## Cumulative Proportion  0.75451 0.77899 0.80037 0.82024 0.83825 0.8530
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     0.94873 0.88716 0.84745 0.81679 0.78617 0.77201
## Proportion of Variance 0.01364 0.01193 0.01088 0.01011 0.00936 0.00903
## Cumulative Proportion  0.86659 0.87852 0.88940 0.89951 0.90887 0.91790
```

```
##                            PC19   PC20    PC21    PC22    PC23    PC24
## Standard deviation     0.69719 0.6700 0.66178 0.63485 0.61153 0.58065
## Proportion of Variance 0.00736 0.0068 0.00664 0.00611 0.00567 0.00511
## Cumulative Proportion  0.92527 0.9321 0.93870 0.94481 0.95048 0.95559
##                            PC25    PC26    PC27   PC28    PC29    PC30
## Standard deviation     0.55316 0.53590 0.48125 0.4597 0.45332 0.44995
## Proportion of Variance 0.00464 0.00435 0.00351 0.0032 0.00311 0.00307
## Cumulative Proportion  0.96022 0.96457 0.96808 0.9713 0.97440 0.97747
##                            PC31    PC32    PC33    PC34    PC35    PC36
## Standard deviation     0.42124 0.40970 0.39691 0.38353 0.35571 0.34239
## Proportion of Variance 0.00269 0.00254 0.00239 0.00223 0.00192 0.00178
## Cumulative Proportion  0.98015 0.98270 0.98508 0.98731 0.98923 0.99101
##                            PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation     0.32674 0.30542 0.27089 0.25237 0.23723 0.21035
## Proportion of Variance 0.00162 0.00141 0.00111 0.00097 0.00085 0.00067
## Cumulative Proportion  0.99262 0.99404 0.99515 0.99611 0.99697 0.99764
##                           PC43    PC44    PC45    PC46    PC47    PC48
## Standard deviation     0.1997 0.17531 0.13485 0.12297 0.11260 0.10478
## Proportion of Variance 0.0006 0.00047 0.00028 0.00023 0.00019 0.00017
## Cumulative Proportion  0.9982 0.99871 0.99898 0.99921 0.99940 0.99957
##                            PC49    PC50    PC51    PC52    PC53    PC54
## Standard deviation     0.08828 0.07574 0.06992 0.06068 0.05428 0.04120
## Proportion of Variance 0.00012 0.00009 0.00007 0.00006 0.00004 0.00003
## Cumulative Proportion  0.99969 0.99978 0.99985 0.99991 0.99995 0.99998
##                            PC55    PC56      PC57      PC58      PC59
## Standard deviation     0.03163 0.02416 2.442e-15 2.244e-15 2.165e-15
## Proportion of Variance 0.00002 0.00001 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  0.99999 1.00000 1.000e+00 1.000e+00 1.000e+00
##                             PC60      PC61     PC62      PC63      PC64
## Standard deviation     1.97e-15 1.912e-15 1.85e-15 1.609e-15 1.567e-15
## Proportion of Variance 0.00e+00 0.000e+00 0.00e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.00e+00 1.000e+00 1.00e+00 1.000e+00 1.000e+00
##                              PC65      PC66
## Standard deviation     1.476e-15 1.236e-15
## Proportion of Variance 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00
```

```
plot(players_18.pr$x[,1],players_18.pr$x[,2], xlab="PC1 (44.3%)", ylab = "PC2 (19%)", main = "PC1 / PC2
```


PC1 / PC2 − plot
PC2 (19%)
PC1 (44.3%)
−15
−10
−5
0
5