Assignment 3 – PCA

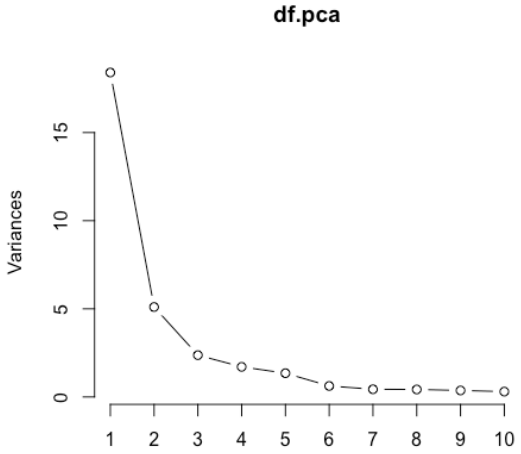
Code:

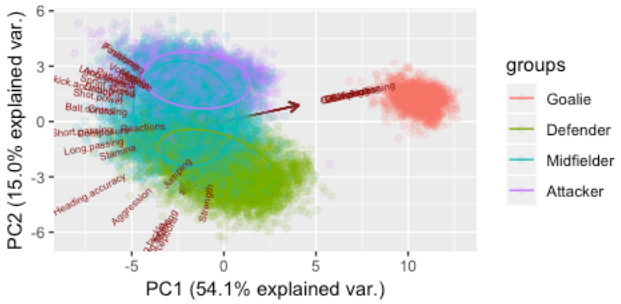
```
library(tidyverse)
library(ggbiplot)

df <- read.csv('FIFA18.csv', header = TRUE, sep = ',')

# Cleaning up a data a little
df$Interceptions <- as.numeric(df$Interceptions)
df$Positioning <- as.numeric(df$Positioning)

# Assigning every player a primary position by removing additional positions, then making them
factors
df$Preferred.Positions <- sub("\\s.*", "", df$Preferred.Positions)
df$Preferred.Positions <- as.factor(df$Preferred.Positions)

# Assigning a role to each player
pos <- as.factor(df$Preferred.Positions)
levels(pos) <- list(Goalie  = c("GK"),
                    Defender = c("CB","RB","LB","RWB","LWB"),
                    Midfielder = c("CDM","CM","RM","LM","CAM"),
                    Attacker = c("CF","RW","LW","ST"))
df <- mutate(df, Role = pos)

# Performing PCA on all the skills
df.pca <- df %>%
  select(Acceleration:Volleys) %>%
  prcomp(center = TRUE, scale. = TRUE)

# Screeplot of skill ratings
screeplot(df.pca, type="lines")

# Making a biplot
ggbiplot(df.pca, obs.scale = 1, var.scale = 1, alpha = 0.1,
         groups=df$Role, varname.size = 2, varname.adjust = 2,
         ellipse = TRUE, circle = FALSE)

# Performing PCA on only goalie related skills
goalie.pca <- df %>%
  select(GK.diving:GK.reflexes) %>%
  prcomp(center = TRUE, scale. = TRUE)

# Screeplot of goalie ratings
screeplot(goalie.pca, type="lines")
```

```
summary(goalie.pca)

# Taking only non-goalies

nongoalies <- subset(df, Role != "Goalie")

# Performing PCA on non-goalie related skills
nongoalie.pca <- nongoalies %>%
  dplyr::select(Acceleration:Free.kick.accuracy, Heading.accuracy:Volleys) %>%
  prcomp(center = TRUE, scale. = TRUE)

# Screeplot of non-goalie ratings
screeplot(nongoalie.pca, type="lines")
summary(nongoalie.pca)

# Making a biplot of non-goalie ratings
ggbiplot(nongoalie.pca, obs.scale = 1, var.scale = 1, alpha = 0.1,
        groups=nongoalies$Role, varname.size = 2, varname.adjust = 2,
        ellipse = TRUE, circle = FALSE)
```

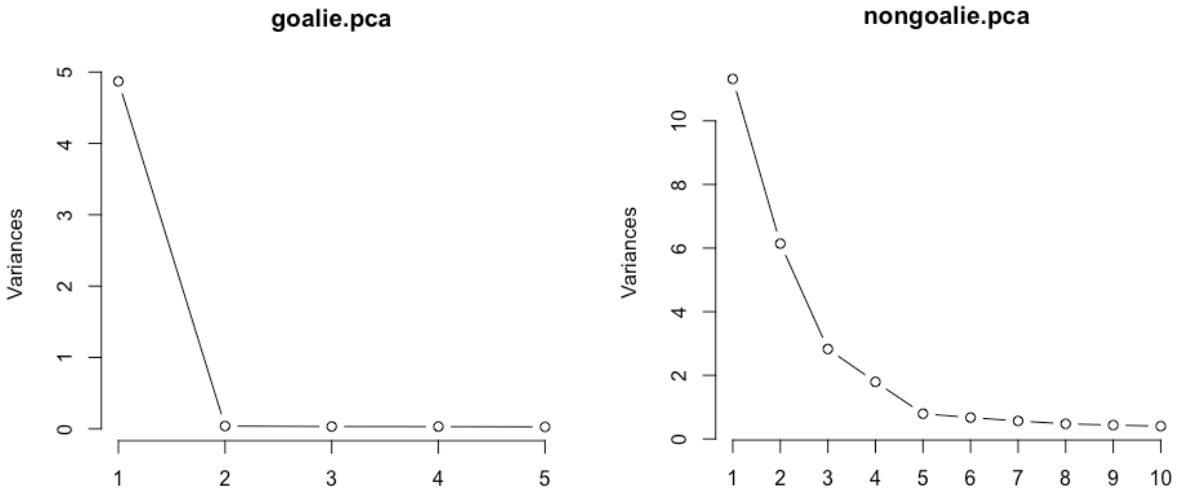
Analysis:

Looking at the screeplot for all the skills, we see that 3 skills account for 75% of the variance. Looking at the biplot, we can see that goalkeeper skills were grouped together separately, so I broke them down to goalkeepers and others to see if we can glean more information.





Looking just goalkeeper related skills, we can see just 1 of them already accounts for 97% of the variance.



Meanwhile the other players' skills are more spread out. Attackers and defenders have quite separate skills, but midfielders have a combination of attacking and defending skills and help bridge the gap so the 3 roles form 1 big group.