

# Multivariate Analysis of Player Performance in VNL 2023 using R

– Xuân Hương Nguyễn –

2025-30-6

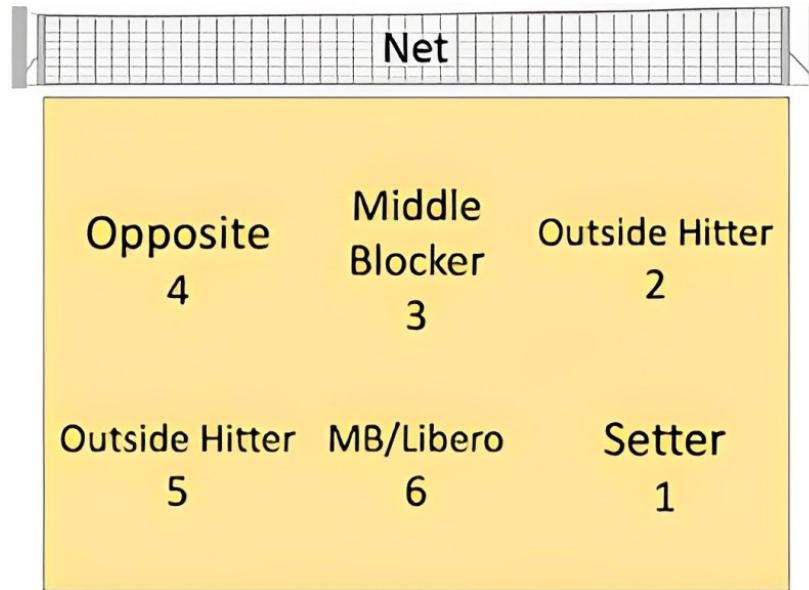
## Dataset: Volleyball National League in 2023

Bộ dữ liệu về các cầu thủ bóng chuyền trong Giải bóng chuyền quốc gia năm 2023 (Volleyball National League in 2023). Bộ dữ liệu gồm 131 quan trắc và 10 biến:

1. **Player:** Tên của vận động viên.
2. **Country:** Quốc tịch của vận động viên.
3. **Age:** Tuổi của vận động viên.
4. **Attack:** Trung bình tổng thể của vận động viên trong mỗi trận đấu khi tham gia khâu tấn công.
5. **Block (Chắn bóng):** Trung bình tổng thể của vận động viên trong mỗi trận đấu khi tham gia phòng thủ trên lưới.
6. **Serve (Phát bóng):** Trung bình khả năng phát bóng của từng vận động viên trong trận đấu.
7. **Set (Chuyền bóng):** Trung bình số lần chuyền thành công, lỗi chuyền và số lần chuyền thử trong mỗi pha bóng.
8. **Dig (Phòng thủ):** Trung bình số lần cứu bóng, lỗi cứu bóng và đỡ bóng từ đối phương tấn công.
9. **Receive (Đỡ bóng):** Trung bình của các lần chạm được bóng khi phòng thủ thành công và pha bóng vẫn còn tiếp tục.
10. **Position (Vị trí):** Vị trí đại diện cho vị trí chơi của vận động viên và bao gồm 5 vị trí cơ bản sau đây:
  - OH (Outside Hitter): Chủ công ngoài

- OP (Opposite): Đối chuyền
- MB (Middle Blocker): Trung phong chắn bóng
- S (Setter): Chuyên hai
- L (Libero): Cầu thủ chuyên phòng thủ

Sơ đồ cơ bản của vận động viên trong sân:



*Lưu ý: Trong mỗi yếu tố, bộ dữ liệu sử dụng điểm trung bình của vận động viên trong một trận đấu, là tác nhân gây ra lỗi, điểm và tỷ lệ phần trăm hiệu suất tích cực của vận động viên đó trong mỗi yếu tố.*

(Nguồn dữ liệu: Kaggle - <https://www.kaggle.com/datasets/yeganehbavafa/vnl-men-2023> ).

## Load dữ liệu của bài toán

```
vnl_players <- read.csv(file="datasets/VNL2023.csv") |>
  janitor::clean_names()
head(vnl_players)
```

```
##           player country age attack block serve  set  dig receive
## 1  Ichikawa Yuki   Japan  28  15.80  1.13  1.40 0.07 4.80    5.60
## 2   Romano Yuri    Italy  26  12.33  1.07  1.47 0.00 3.87    0.00
## 3 Abdel-Aziz Nimir Nederland 31  15.33  0.67  2.08 0.00 3.17    0.25
## 4 Herrera Jaime Jesus     Cuba  28  15.00  0.92  1.75 0.00 3.33    0.17
## 5  Takahashi Ran     Japan  22  11.53  0.67  1.00 0.07 6.40    5.07
## 6  Luburic Drazen    Serbia  30  14.33  0.58  1.08 0.00 4.25    0.00
##      position
```

```
## 1      OH
## 2      OP
## 3      OP
## 4      OP
## 5      OH
## 6      OP
```

```
dim(vnl_players)
```

```
## [1] 131 10
```

## Khám phá dữ liệu

Kiểm tra giá trị khuyết:

```
colSums(is.na(vnl_players))
```

```
##  player country      age  attack  block  serve  set  dig
##      0        0        0        0      0      0    0    0
## receive position
##      0        0
```

```
colSums(is.na(vnl_players))/nrow(vnl_players)*100
```

```
##  player country      age  attack  block  serve  set  dig
##      0        0        0        0      0      0    0    0
## receive position
##      0        0
```

**Nhận xét:** Bộ dữ liệu được ghi lại đầy đủ các quan sát, không xuất hiện giá trị khuyết nào.

Kiểm tra giá trị lặp:

```
sum(duplicated(vnl_players))
```

```
## [1] 0
```

**Nhận xét:** Không có quan sát nào bị trùng lặp.

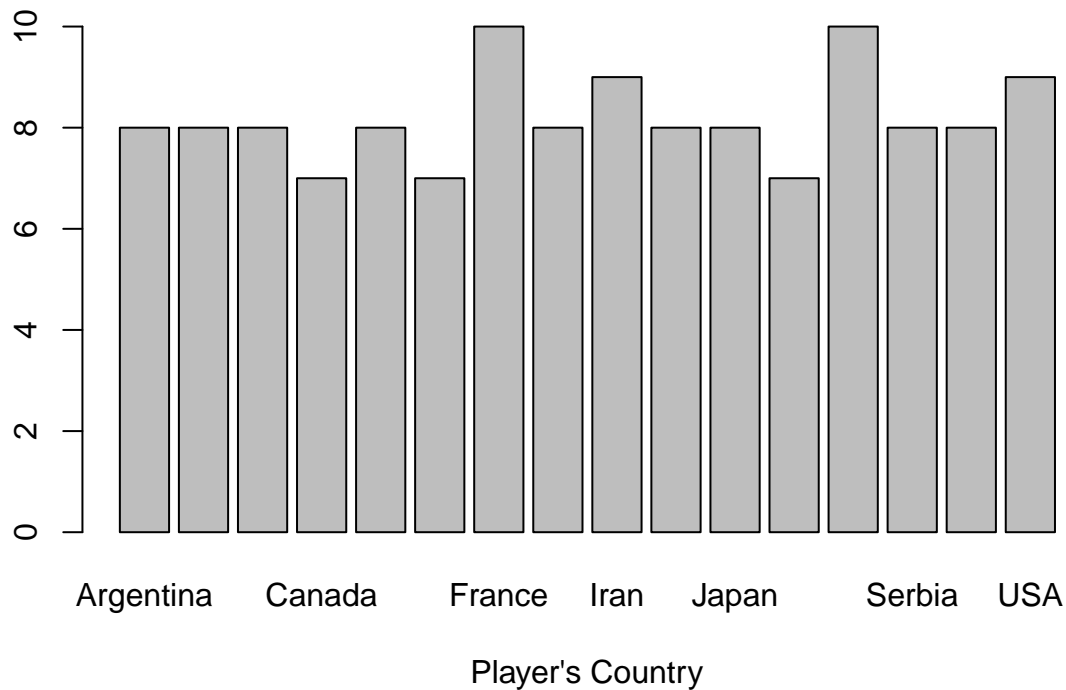
**Phân tích các biến định tính:**

Thống kê số lượng mỗi nhóm

```
table(vnl_players$country)
```

```
##
## Argentina      Brazil      Bulgaria      Canada      China      Cuba      France      Germany
##           8           8           8           7           8           7           10           8
##      Iran      Italy      Japan      Nederland      Poland      Serbia      Slovenia      USA
##           9           8           8           7           10           8           8           9
```

Countplot cho biến `country`:

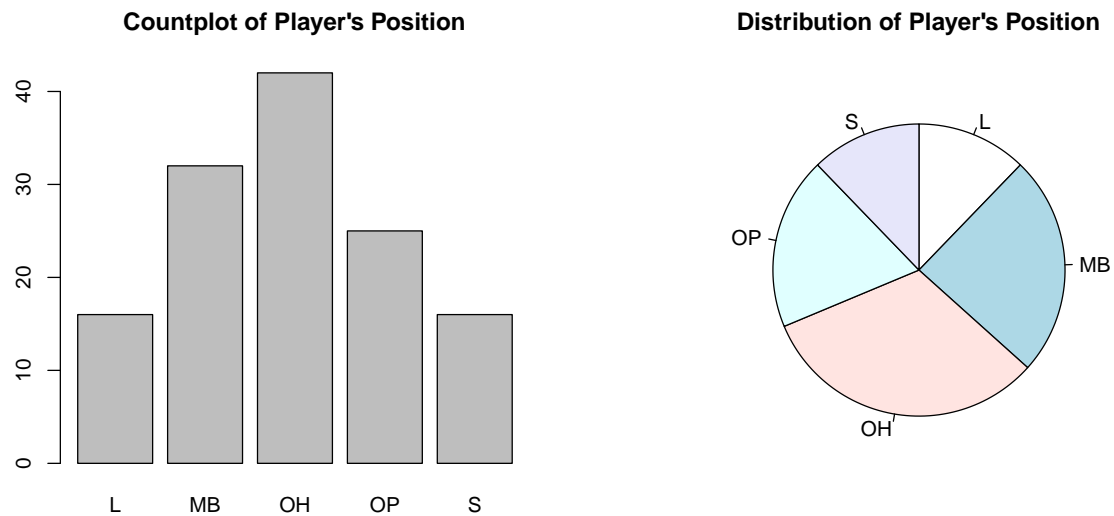


**Nhận xét:** Cầu thủ tham gia Giải bóng chuyền quốc gia năm 2023 đến từ 16 quốc gia khác nhau, số lượng cầu thủ từ mỗi nước dao động từ 7 đến 10 người. Phân phối số lượng cầu thủ ở mỗi quốc gia khá đồng đều, không có quốc gia nào áp đảo, có thể so sánh công bằng.

```
table(vnl_players$position)
```

```
##
## L MB OH OP S
## 16 32 42 25 16
```

Countplot và Pie Chart của biến `position`:



**Nhận xét:**

- **Outside Hitter (OH)** là vị trí phổ biến nhất (chiếm gần 1/3 toàn bộ cầu thủ). Điều này phản ánh vai trò đa năng của chủ công ngoài trong bóng chuyền hiện đại: tấn công biên, phòng thủ, và cả nhận bóng.
- **Setter (S)** và **Libero (L)** chiếm số lượng ít hơn, tương ứng với tính chất chuyên môn hóa cao (chuyên chuyền hai, chuyên phòng thủ) và không thay thế nhau tùy tiện trong thi đấu.

**Phân tích các biến định lượng:**

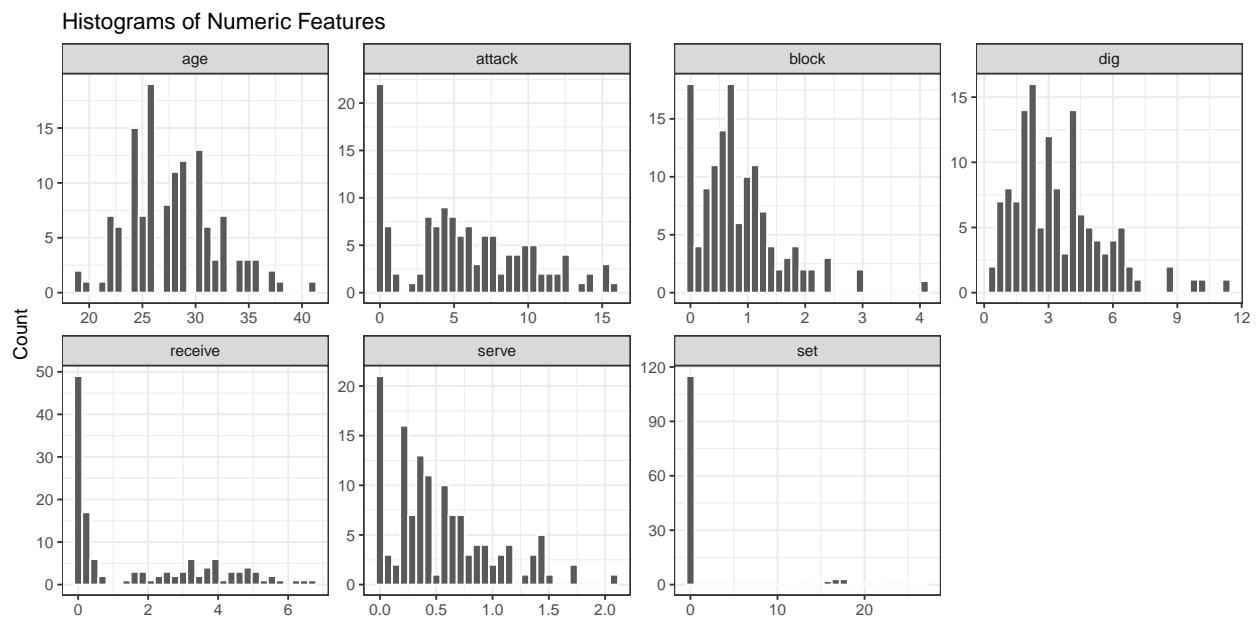
Thống kê mô tả:

```
vnl_players |>
  select(age, attack, block, serve, set, dig, receive) |>
  summary()
```

##	age	attack	block	serve
##	Min. :19.00	Min. : 0.000	Min. :0.0000	Min. :0.0000
##	1st Qu.:25.00	1st Qu.: 2.800	1st Qu.:0.3700	1st Qu.:0.2400
##	Median :27.00	Median : 5.170	Median :0.6900	Median :0.4200
##	Mean :27.81	Mean : 5.643	Mean :0.8456	Mean :0.5358

```
## 3rd Qu.:30.00  3rd Qu.: 8.600  3rd Qu.:1.1400  3rd Qu.:0.7600
## Max.   :41.00  Max.   :15.800  Max.   :4.0800  Max.   :2.0800
##      set      dig      receive
## Min.   : 0.000  Min.   : 0.530  Min.   :0.000
## 1st Qu.: 0.000  1st Qu.: 1.920  1st Qu.:0.000
## Median : 0.000  Median : 3.000  Median :0.330
## Mean   : 2.193  Mean   : 3.428  Mean   :1.684
## 3rd Qu.: 0.000  3rd Qu.: 4.510  3rd Qu.:3.385
## Max.   :26.890  Max.   :11.440  Max.   :6.690
```

Biểu đồ histogram của các biến `age`, `attack`, `block`, `serve`, `set`, `dig`, `receive`:

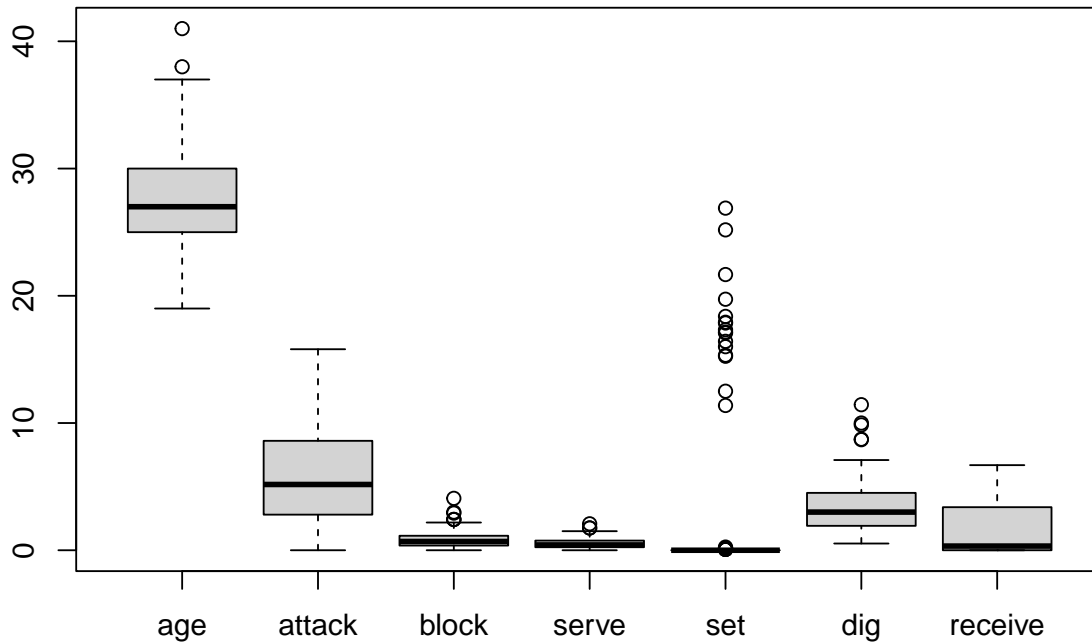


**Nhận xét:**

- Biến `set` có trung vị bằng 0, nhưng giá trị lớn nhất lên tới gần 27, khả năng phần lớn cầu thủ không thực hiện chuyền hai, trừ các Setter chuyên biệt.
- Nhiều biến có phân phối lệch phải, đặc biệt là `attack`, `set`, `receive` và `serve`.
- Một số biến có giá trị bằng 0 nhiều, do đó cần chú ý cân nhắc khi phân tích nhiều chiều.

Boxplot của các biến:

**Boxplot of Numeric Features**



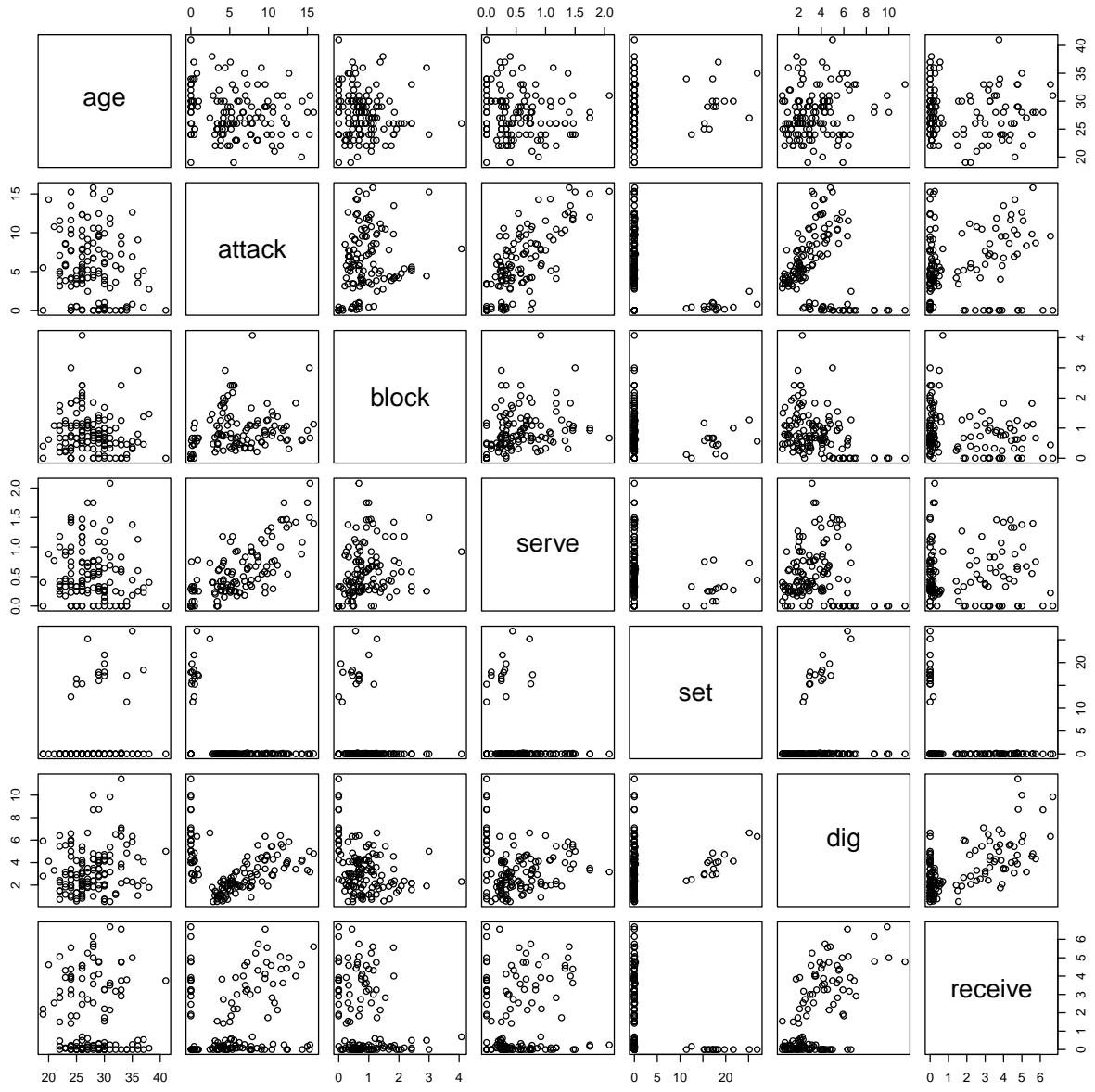
**Nhận xét:** Các biến có thang đo chênh lệch nhau nhiều, do đó cần thực hiện chuẩn hoá dữ liệu trước khi áp dụng các phương pháp phân tích nhiều chiều.

Ma trận tương quan của các biến:

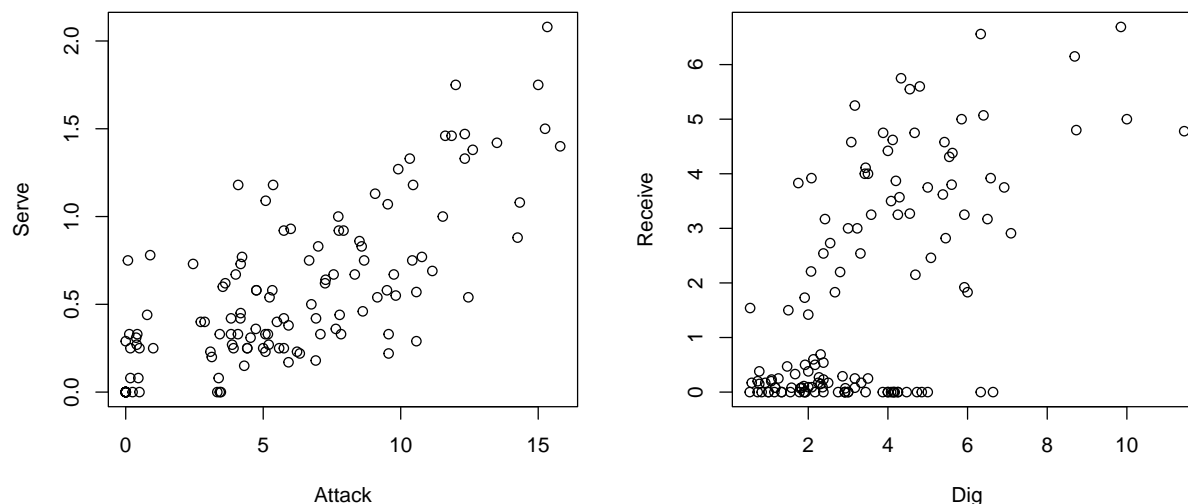
```
cor_matrix <- cor(vnl_players[, c("age", "attack", "block", "serve",
                                   "set", "dig", "receive")])
round(cor_matrix, 3)
```

```
##          age attack  block  serve   set   dig receive
## age      1.000 -0.178 -0.101 -0.108  0.178  0.167 -0.011
## attack  -0.178  1.000  0.338  0.769 -0.431 -0.099  0.170
## block   -0.101  0.338  1.000  0.336 -0.132 -0.348 -0.265
## serve   -0.108  0.769  0.336  1.000 -0.155 -0.053  0.040
## set      0.178 -0.431 -0.132 -0.155  1.000  0.132 -0.306
## dig      0.167 -0.099 -0.348 -0.053  0.132  1.000  0.625
## receive -0.011  0.170 -0.265  0.040 -0.306  0.625  1.000
```

Biểu đồ scatter cho từng cặp biến định lượng:







**Nhận xét:** Dữ liệu tồn tại những cặp biến có tương quan mạnh như “**attack** (tấn công) và **serve** (phát bóng)” (0.769), “**dig** (phòng thủ) và **receive** (đỡ bóng)” (0.625) cho thấy các biến kỹ năng này có xu hướng thay đổi cùng nhau, tức là không hoàn toàn độc lập. Ngoài ra, biến **age** có tương quan thấp với tất cả các biến kỹ năng cho thấy tuổi không đóng vai trò quyết định kỹ năng chuyên môn.

Việc xuất hiện nhiều cặp biến kỹ năng có tương quan dương mạnh rõ rệt phản ánh bản chất **đa chiều** và **có cấu trúc lặp lại** trong không gian kỹ năng của các vận động viên. Đây chính là điều kiện thích hợp để áp dụng phân tích nhiều chiều nhằm phát hiện các chiều kỹ năng tổng hợp tiềm ẩn trong cấu trúc dữ liệu thực, rút gọn dữ liệu nhưng vẫn bảo toàn thông tin và làm rõ sự phân bố kỹ năng giữa các vị trí thi đấu.

## Đề xuất phương pháp

Bộ dữ liệu ghi nhận nhiều chỉ số kỹ thuật chuyên môn như **attack** (tấn công), **block** (chặn bóng), **serve** (phát bóng), **dig** (phòng thủ), **receive** (đỡ bóng), **set** (chuyền bóng), cùng với thông tin tuổi và vị trí thi đấu. Các chỉ số này có thể liên quan chặt chẽ đến vai trò thi đấu, song không phải lúc nào cũng rõ ràng với nghi vấn có thể tồn tại những vận động viên đa năng hoặc ngoại lệ chưa được phân tích kỹ.

Vì vậy, cần áp dụng các phương pháp phân tích nhiều chiều như **Phân tích thành phần chính PCA** để khám phá cấu trúc nhóm, và **Phân tích nhân tố FA** để phân tích các nhóm kỹ năng tiềm ẩn chi phối hiệu suất thi đấu.

## Vấn đề nghiên cứu

Khám phá mối liên hệ giữa các biến và phân nhóm đặc điểm thi đấu của các cầu thủ bóng chuyền dựa trên các chỉ số kỹ thuật, nhằm:

- Hỗ trợ nhận diện vai trò thi đấu.
- Phát hiện cầu thủ nổi bật và phục vụ mục tiêu huấn luyện chuyên sâu.

## Triển khai phương pháp

### Phương pháp 1: Sử dụng phương pháp phân tích thành phần chính - PCA:

Với mục tiêu khám phá cấu trúc **kỹ năng kỹ thuật chuyên môn**, ta loại biến age ra khỏi bộ dữ liệu dùng thực hiện PCA:

```
dat_vnl_players <- vnl_players |> select(-player, -country, -position, -age)
names(dat_vnl_players)
```

```
## [1] "attack" "block" "serve" "set" "dig" "receive"
```

So sánh kết quả thực hiện trên ma trận hiệp phương sai S và ma trận tương quan R

Thực hiện tính toán trên phần mềm R thông qua hàm princomp() dựa trên ma trận hiệp phương sai S:

```
vnl_players.pca.cov <- princomp(dat_vnl_players, cor=F)
summary(vnl_players.pca.cov, loadings=TRUE)
```

```
## Importance of components:
##
##              Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## Standard deviation  6.4643763 3.6002481 2.5853947 1.05473904 0.605134969
## Proportion of Variance 0.6634525 0.2057887 0.1061232 0.01766228 0.005813814
## Cumulative Proportion 0.6634525 0.8692412 0.9753644 0.99302669 0.998840500
##
##              Comp.6
## Standard deviation  0.2702446
## Proportion of Variance 0.0011595
## Cumulative Proportion 1.0000000
##
## Loadings:
##              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## attack      0.424  0.898
## block                                -0.990
## serve                                -0.993
## set        -0.899  0.425
## dig              0.745  0.663
## receive              0.657 -0.740
```

```
round(vnl_players.pca$cdev^2, 5)
```

```
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6  
## 41.78816 12.96179  6.68427  1.11247  0.36619  0.07303
```

Thực hiện tính toán trên phần mềm R thông qua hàm `princomp()` dựa trên ma trận tương quan mẫu **R**:

```
vnl_players.pca <- princomp(dat_vnl_players, cor=T)  
summary(vnl_players.pca, loadings=TRUE)
```

```
## Importance of components:  
##  
##               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5  
## Standard deviation    1.4916239 1.3454863 0.9663580 0.7807868 0.49689369  
## Proportion of Variance 0.3708236 0.3017222 0.1556413 0.1016047 0.04115056  
## Cumulative Proportion 0.3708236 0.6725459 0.8281871 0.9297918 0.97094236  
##  
##               Comp.6  
## Standard deviation    0.41754740  
## Proportion of Variance 0.02905764  
## Cumulative Proportion 1.00000000  
##  
## Loadings:  
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## attack    0.592  0.213  0.117  0.227          0.730  
## block     0.425 -0.285          -0.856  
## serve     0.540  0.129  0.470  0.268          -0.626  
## set       -0.340 -0.256  0.785          0.410  0.184  
## dig       -0.245  0.572  0.364 -0.304 -0.609  0.131  
## receive          0.682 -0.107 -0.227  0.669 -0.154
```

```
round(vnl_players.pca$cdev^2, 5)
```

```
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6  
## 2.22494 1.81033 0.93385 0.60963 0.24690 0.17435
```

### Nhận xét:

Có sự khác nhau rõ rệt trong kết quả phân tích sử dụng ma trận hiệp phương sai **S** và ma trận tương quan mẫu **R**:

- Với ma trận hiệp phương sai: thành phần chính đầu tiên (PC1) chiếm tới 66.3% tổng phương sai, chủ yếu bị chi phối bởi các biến có giá trị lớn và độ biến động cao như **attack** và **set**.

- Với ma trận tương quan: PC1 chỉ chiếm 37.1%, và phương sai được phân bố đều hơn giữa các thành phần.

**Lý do của sự khác nhau:** Sự chênh lệch đến từ việc ma trận hiệp phương sai giữ nguyên đơn vị gốc của các biến, khiến những biến có độ lớn và phương sai cao (ví dụ: **attack**, **set**) lấn át kết quả. Ngược lại, ma trận tương quan chuẩn hóa dữ liệu, đưa các biến về cùng thang đo (trung bình 0, độ lệch chuẩn 1), do đó đảm bảo mọi biến có cơ hội đóng góp công bằng vào mô hình phân tích.

**Kết luận:** Do dữ liệu gồm nhiều biến khác nhau về đơn vị và thang đo, ta lựa chọn phân tích thành phần chính dựa trên ma trận tương quan. Cách tiếp cận này giúp khai thác toàn diện các chiều thông tin, tránh thiên lệch bởi quy mô biến, và phản ánh cấu trúc đa chiều của dữ liệu một cách công bằng, khách quan và chính xác hơn.

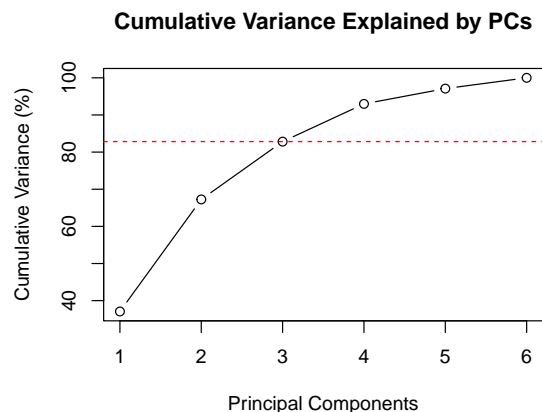
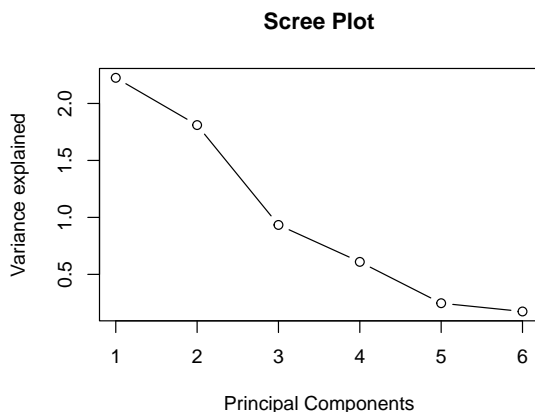
## Phân tích kết quả

Tiếp tục, thực hiện phân tích kết quả của PCA trên ma trận tương quan **R** trên 2 tiêu chí.

### 1. Tiêu chí 1: Phương sai giải thích và phương sai cộng dồn của từng PC

```
summary(vnl_players.pca, loadings=FALSE)
```

```
## Importance of components:
##                               Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## Standard deviation          1.4916239 1.3454863 0.9663580 0.7807868 0.49689369
## Proportion of Variance      0.3708236 0.3017222 0.1556413 0.1016047 0.04115056
## Cumulative Proportion      0.3708236 0.6725459 0.8281871 0.9297918 0.97094236
##                               Comp.6
## Standard deviation          0.41754740
## Proportion of Variance      0.02905764
## Cumulative Proportion      1.00000000
```



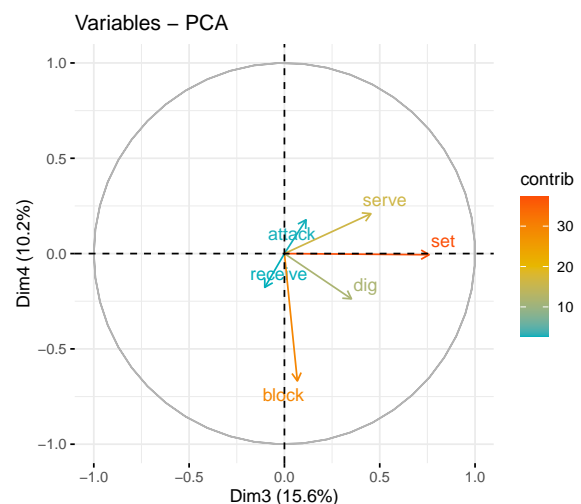
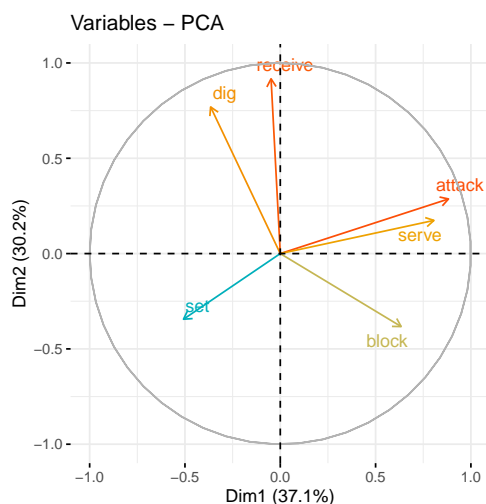
### Nhận xét:

Thành phần chính đầu tiên giải thích khoảng 37.1% tổng phương sai giải thích. Với chỉ ba thành phần chính đầu tiên, dữ liệu cầu thủ đã được thể hiện tốt mà không mất nhiều thông tin với 82.8% tổng phương sai giải thích. So sánh các tỷ lệ này với các tỷ lệ thu được khi sử dụng các dữ liệu không chuẩn hóa, phân tích này yêu cầu nhiều thành phần hơn để giải thích cùng một khoảng lượng biến thiên như phân tích ban đầu khi sử dụng ma trận hiệp phương sai. Scree-plot cũng thể hiện “elbow-point” tại PC4. Do đó, ta có thể cân nhắc giảm chiều xuống 3 chiều.

## 2. Tiêu chí 2: Loadings của từng PC

```
loadings(vnl_players.pca)
```

```
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## attack  0.592  0.213  0.117  0.227          0.730
## block   0.425 -0.285          -0.856
## serve   0.540  0.129  0.470  0.268          -0.626
## set     -0.340 -0.256  0.785          0.410  0.184
## dig     -0.245  0.572  0.364 -0.304 -0.609  0.131
## receive          0.682 -0.107 -0.227  0.669 -0.154
##
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var 0.167  0.167  0.167  0.167  0.167  0.167
## Cumulative Var 0.167  0.333  0.500  0.667  0.833  1.000
```



**Nhận xét:** Dựa vào kết quả loadings cùng với biểu đồ trực quan, em rút ra nhận định về các thành phần chính như sau:

**Thành phần chính thứ nhất:**

PC1 phản ánh mạnh mẽ các kỹ năng tấn công trực tiếp trong bóng chuyền. Các biến có tải số lớn bao gồm **attack** (0.592), **serve** (0.540), và **block** (0.425) đều là những chỉ số liên quan đến khả năng ghi điểm, gây áp lực lên đối thủ, và thực hiện các pha chặn bóng trên lưới. Những vận động viên có giá trị PC1 càng cao thì càng có khả năng tấn công tốt. Thành phần này giúp phân biệt rõ ràng giữa những cầu thủ thiên về ghi điểm như **Outside Hitter (OH)** và **Opposite (OP)** với các cầu thủ đóng vai trò hỗ trợ như **Setter (S)**.

Như vậy: PC1 có thể được hiểu là trực thể hiện **năng lực tấn công tổng hợp**, và đóng vai trò chính trong việc nhận diện những “chủ công” của đội.

**Thành phần chính thứ hai:**

Ngược lại, PC2 tập trung thể hiện năng lực phòng thủ từ tuyến sau, với hai biến có tải số cao là **receive** (0.682) và **dig** (0.572), đại diện cho kỹ năng đỡ bước một và cứu bóng. Đây là những chỉ số tiêu biểu cho những vận động viên có khả năng phòng ngự chắc chắn, thường gặp ở những người chơi vị trí **Libero (L)** hoặc những **OH có thiên hướng phòng thủ**. Những cầu thủ có giá trị PC2 cao là những người thường xuyên tham gia vào các pha cứu bóng và đỡ bóng đầu tiên sau phát bóng hoặc tấn công của đối thủ. Thành phần này giúp tách biệt rõ nhóm cầu thủ chuyên về phòng thủ bền bỉ và ổn định trong hệ thống chiến thuật của đội, từ đó có thể đánh giá hiệu quả hoạt động của tuyến sau trong mỗi đội bóng.

Như vậy: PC2 có thể được hiểu là trực thể hiện **năng lực phòng thủ**.

**Thành phần chính thứ ba:**

PC3 cho thấy mức độ điều phối và tổ chức trận đấu của vận động viên, chủ yếu qua biến **set** (0.785) và một phần nhỏ biến **serve** (0.470). Đây là những kỹ năng đặc trưng của vị trí chuyên hai **Setter (S)**, người có vai trò phân phối bóng, điều tiết nhịp độ trận đấu và tạo điều kiện cho các đợt tấn công. Vận động viên có điểm số cao trên PC3 thường không trực tiếp ghi điểm nhưng lại rất quan trọng trong việc xây dựng lối chơi chiến thuật.

Như vậy: PC3 đại diện cho trực thể hiện **năng lực tổ chức chiến thuật trận đấu**.

**Thành phần chính thứ tư:**

PC4 chủ yếu được định hình bởi biến **block** với hệ số âm mạnh ( $-0.856$ ), phản ánh khả năng chắn bóng vượt trội, một đặc trưng rất rõ của các trung phong chắn bóng **Middle Blocker (MB)**. Mức độ ảnh hưởng của các biến khác ở trục này là nhỏ, nên có thể xem đây là một thành phần đơn trục, giúp tách bạch những cầu thủ có khả năng ngăn chặn tấn công đối phương ngay từ trên lưới (phòng thủ sớm).

Như vậy: PC4 thể hiện trực **năng lực chắn bóng độc lập**.

**Thành phần chính thứ năm:**

PC5 là sự kết hợp giữa **receive** (0.669), **set** (0.410), và **dig** (−0.609). Điều này dẫn đến sự kết hợp giữa đỡ bóng tuyến sau, chuyền bóng điều phối, nhưng lại phản ánh tiêu cực ở khả năng cứu bóng. PC5 khó diễn giải vì kết hợp kỹ năng đối lập.

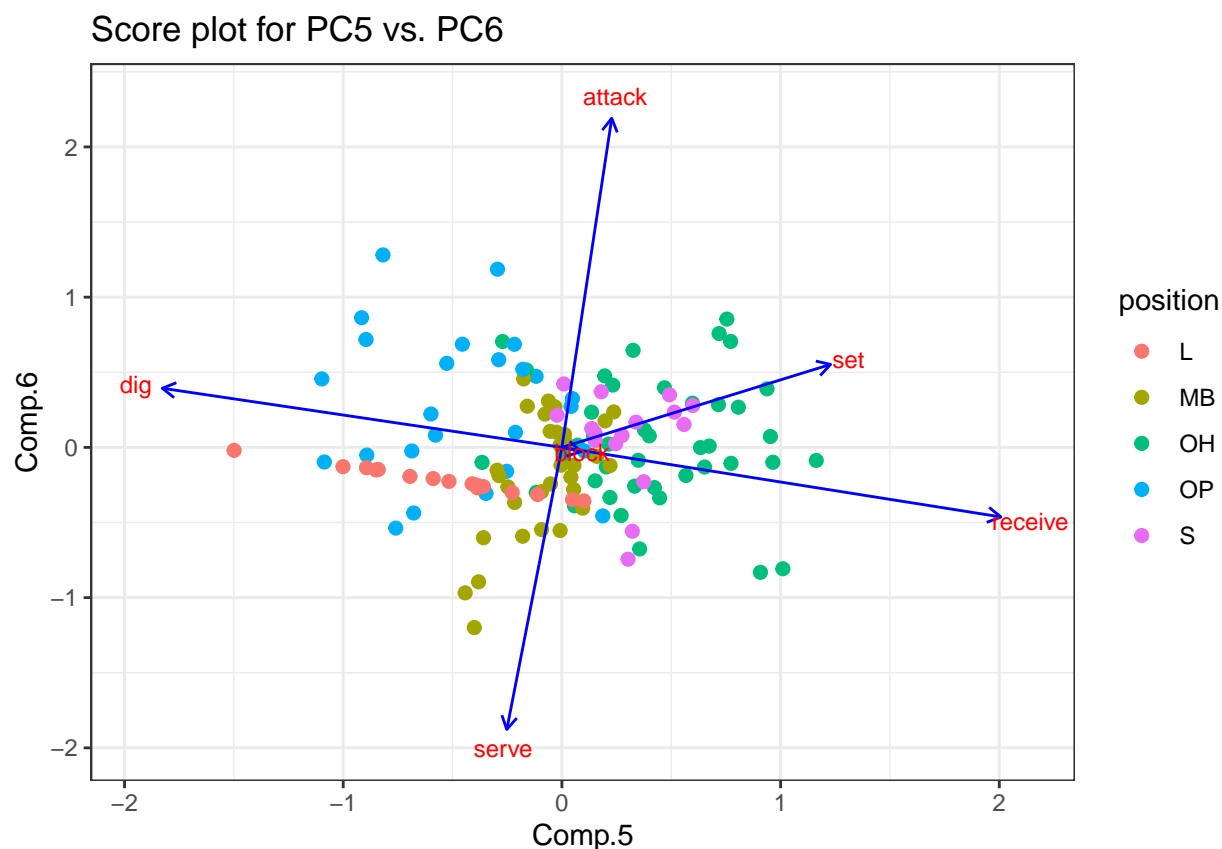
#### Thành phần chính thứ sáu:

PC6 thể hiện sự phân tách giữa kỹ năng tấn công và khả năng phát bóng chiến thuật thông qua hai biến đối lập: **attack** (0.730) và **serve** (−0.626). Thành phần này đại diện cho một cầu thủ tấn công giỏi nhưng lại kém ở khả năng phát bóng. Điều này không phù hợp với thực tế thi đấu.

### Chọn số thành phần chính giữ lại

Nhìn vào kết quả phân tích loadings của các thành phần chính ta thấy:

- PC4 gần như chỉ phản ánh một chiều kỹ năng duy nhất là khả năng chắn bóng. Đây là một yếu tố quan trọng nhưng không đủ để tạo nên một nhóm kỹ năng hay vai trò thi đấu hoàn chỉnh. Ngoài ra, biến **block** cũng đã có đóng góp đáng kể trong PC1 (0.425), vì vậy PC4 chỉ lặp lại một phần thông tin đã có, không cung cấp thêm insight mới, đồng thời không giúp phân nhóm cầu thủ rõ ràng.
- PC5 là một sự pha trộn thiếu nhất quán, ví dụ, vị trí **Libero (L)** thường giỏi cả phòng thủ **receive** lẫn cứu bóng **dig**, còn **Setter (S)** thì thường giỏi chuyền bóng **set** nhưng không tham gia phòng thủ nhiều, PC5 lại là sự đối lập giữa (**receive**, **set**) và (**dig**). Do đó, PC5 phản ánh một tổ hợp kỹ năng thiếu logic thực tế, gây khó khăn trong việc phân loại hoặc gán vai trò cụ thể cho nhóm cầu thủ có điểm số cao ở trục này.
- PC6 thể hiện sự đối lập không hợp lý: một cầu thủ tấn công giỏi nhưng lại kém ở khả năng phát bóng. Điều này không phù hợp với thực tế thi đấu, đặc biệt là ở các vị trí như **Outside Hitter (OH)** hay **Opposite (OP)**, vốn thường mạnh cả hai. Thành phần này vì thế không phản ánh được một xu hướng chiến thuật cụ thể nào.



Như vậy, ba thành phần chính cuối cùng không tách biệt được tốt các nhóm vai trò cầu thủ trong thực tế. Biểu đồ trực quan trên hai trục chính PC5-PC6 cho thấy sự trộn lẫn các nhóm vị trí, thể hiện sự phân tách yếu. Song, ba thành phần này cũng đóng góp phần nhỏ phương sai vào tổng phương sai giải thích. Do đó, nên được loại bỏ khỏi mô hình khi rút gọn dữ liệu để phục vụ cho phân tích hoặc phân nhóm cầu thủ tốt hơn.

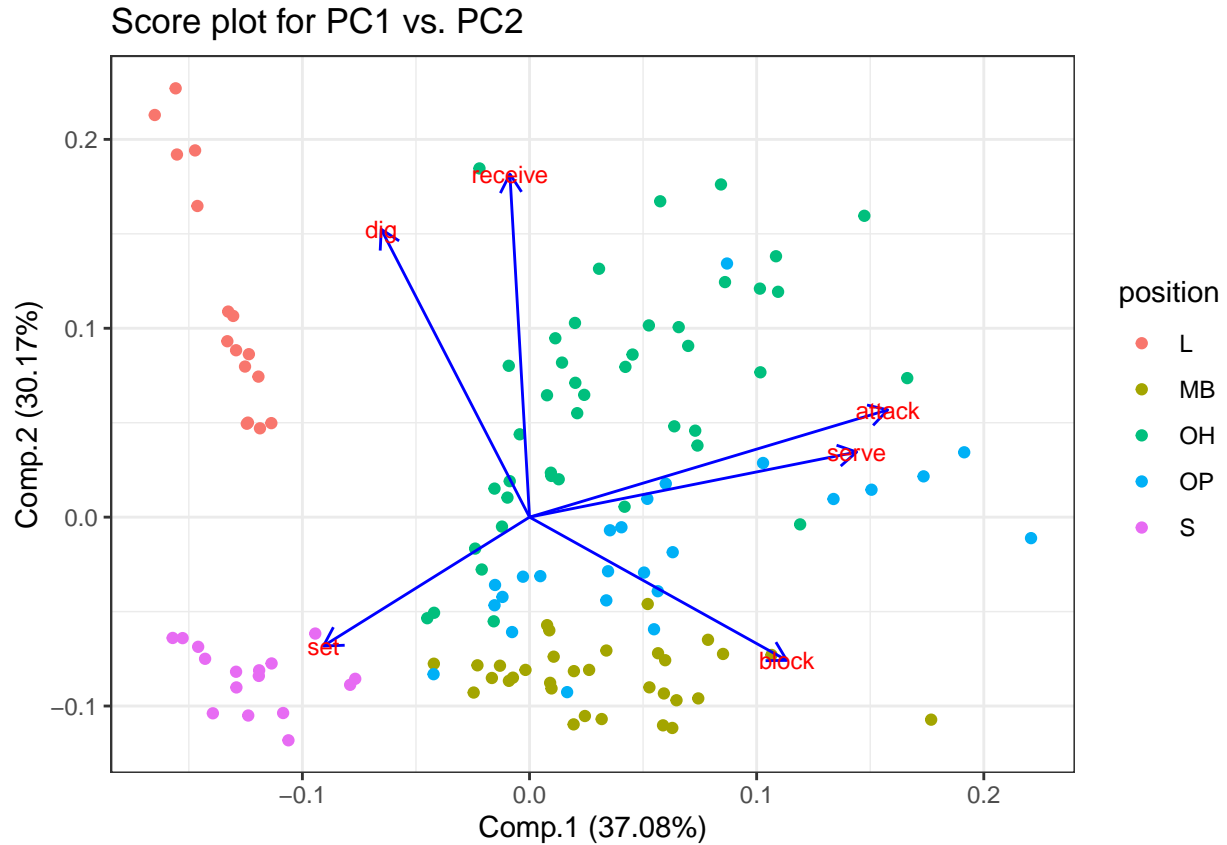
**Kết luận:** Quyết định giữ lại ba thành phần chính đầu tiên (PC1–PC3). Chúng đủ thông tin, có ý nghĩa thực tiễn rõ ràng, và giải thích được 82.8% tổng phương sai.

Thành phần chính	Vai trò kỹ năng đại diện
– PC1	Tấn công tổng hợp
– PC2	Phòng thủ tuyến sau
– PC3	Tổ chức chiến thuật

### Phân biệt các nhóm cầu thủ dựa trên các thành phần chính

Ở phần này, ta tiến hành trực quan các điểm dữ liệu trên kết quả phân tích. Trước hết, phân tích kết quả trên hai trục thành phần chính đầu tiên PC1 và PC2:

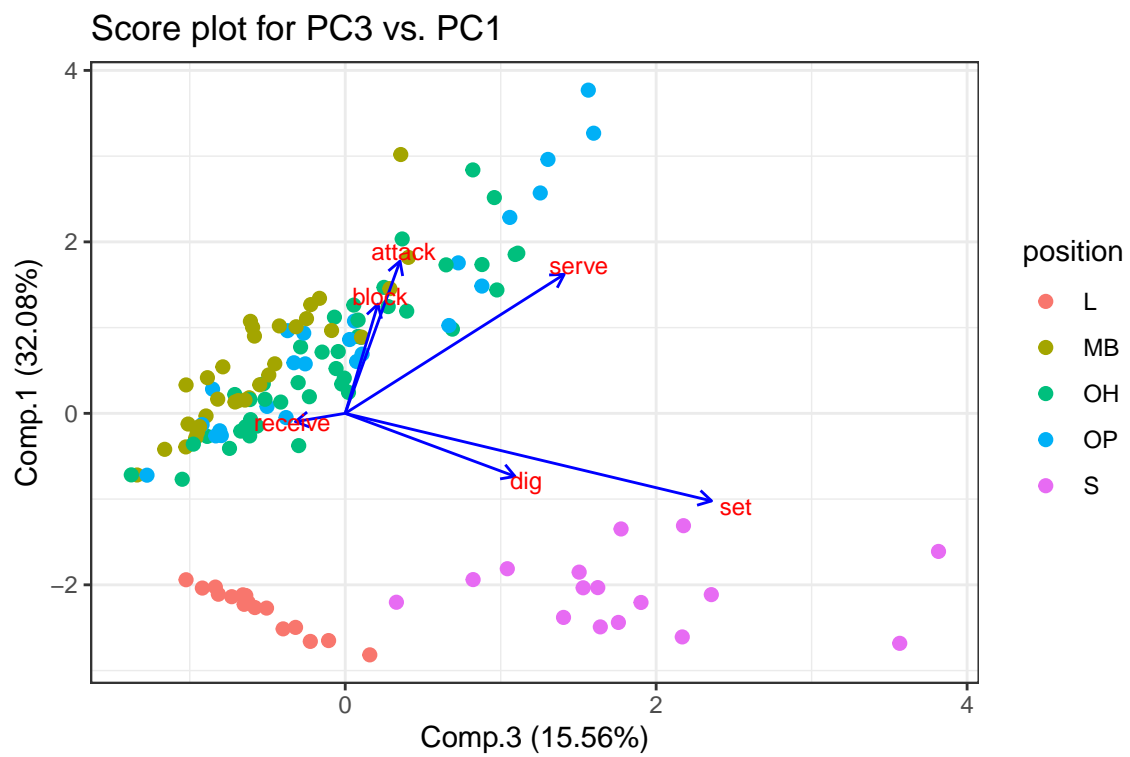
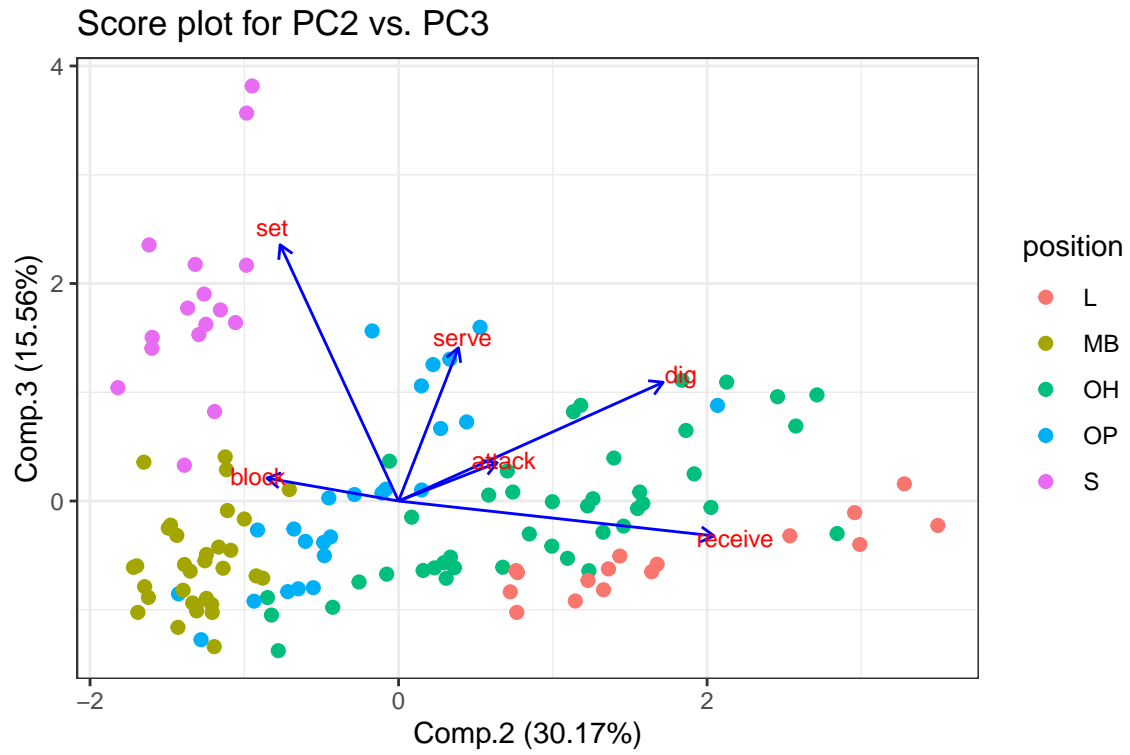




**Nhận xét:** Dựa trên biểu đồ điểm biểu diễn hai thành phần chính đầu tiên (PC1 và PC2), có thể thấy rằng mỗi trục có khả năng phân biệt các nhóm vị trí ở mức độ nhất định.

- Trục PC1 giúp phân tách rõ ràng nhóm tấn công ra khỏi các nhóm còn lại. Các vận động viên có giá trị PC1 cao nằm bên phải đồ thị thiên về tấn công OH, OP, và MB. Trong khi đó, những vận động viên có giá trị PC1 thấp hơn như L và S không thường xuyên tham gia tấn công nên bị đẩy về bên trái đồ thị.
- Trục PC2 giúp phân biệt tốt hơn giữa hai nhóm chuyên biệt L và S. Các L có giá trị PC2 cao, phản ánh sự vượt trội về phòng thủ, trong khi Setter có giá trị PC2 thấp nhất do ít tham gia vào các tình huống cứu bóng.

Khi kết hợp cả hai trục, PCA thể hiện khá rõ cấu trúc phân nhóm của các vị trí trên sân, đặc biệt ở những vai trò chuyên biệt về phòng thủ và tấn công. Các nhóm có xu hướng phân bố ở bốn góc phần tư I, II, III, IV lần lượt là nhóm vị trí (OH), (L), (S) và (OP, MB).



**Nhận xét:** Hai trục PC2-PC3 tiếp tục giúp phân tách tốt nhóm S so với các nhóm còn lại.

## Phát hiện các cầu thủ “đa năng” hoặc có phong cách thi đấu khác biệt so với vị trí truyền thống

Ở phần này, ta thực hiện tìm kiếm những cầu thủ được xem như là một giá trị ngoại lai từ kết quả PCA. Bằng cách tính trung bình và độ lệch chuẩn của từng nhóm vị trí theo hệ trục scale theo PCA (Comp.1, Comp.2, Comp.3), sau đó lọc ra những cầu thủ có giá trị cao hơn hẳn hoặc thấp hơn hẳn so với trung bình và độ lệch chuẩn trong nhóm vị trí cầu thủ đó thuộc về. Từ đó, phân tích xem những cầu thủ này là những cầu thủ đa năng, những cầu thủ có phong cách thi đấu khác biệt thực sự nổi trội hay những cầu thủ chơi sai vị trí so với năng lực hoặc thể hiện vai trò của mình chưa tốt. Kết quả cuối cùng này cung cấp cho các huấn luyện viên thêm thông tin để tận dụng “quân bài” và điều chỉnh chiến lược thi đấu trong các mùa giải tiếp theo.

Tạo dataframe theo hệ trục PCA:

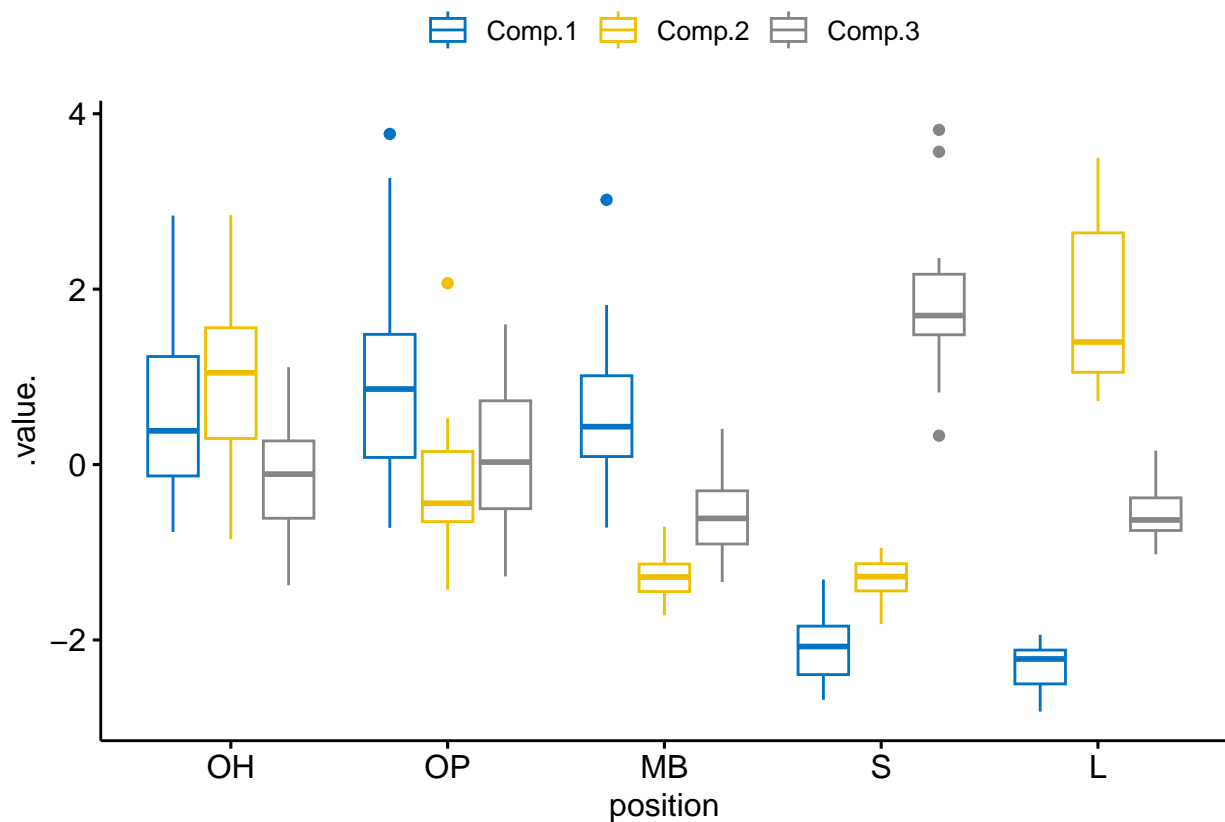
```
scores.pca <- as.data.frame(vnl_players.pca$scores[, 1:3])
scores.pca$position <- vnl_players$position
scores.pca$player <- vnl_players$player
head(scores.pca)
```

```
##      Comp.1   Comp.2   Comp.3 position      player
## 1 2.5166578 2.4570085 0.9591354      OH      Ichikawa Yuki
## 2 2.2850424 0.1478982 1.0587751      OP      Romano Yuri
## 3 3.2674064 0.5285791 1.5979310      OP  Abdel-Aziz Nimir
## 4 2.9618549 0.3324903 1.3040177      OP Herrera Jaime Jesus
## 5 0.9822707 2.5753504 0.6895210      OH      Takahashi Ran
## 6 1.7551340 0.4415720 0.7271021      OP      Luburic Drazen
```

Tính toán trung bình và độ lệch chuẩn theo vị trí:

```
(stats_by_position <- scores.pca |>
  group_by(position) |>
  summarise(across(starts_with("Comp"), list(mean = mean, sd = sd))))
```

```
## # A tibble: 5 x 7
##   position Comp.1_mean Comp.1_sd Comp.2_mean Comp.2_sd Comp.3_mean Comp.3_sd
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 L        -2.29     0.260     1.76     0.967    -0.555    0.312
## 2 MB         0.561     0.766    -1.30     0.250    -0.564    0.442
## 3 OH         0.635     0.884     0.962     0.955    -0.103    0.629
## 4 OP         1.00      1.18    -0.244     0.707     0.0732    0.842
## 5 S        -2.07     0.413    -1.30     0.255     1.84     0.888
```



Lọc các cầu thủ có giá trị lệch hẳn so với nhóm vị trí:

```
scores.outliers <- scores.pca |>
  left_join(stats_by_position, by = "position")
```

```
scores.outliers$outlier_flag <- with(scores.outliers,
  abs(Comp.1 - Comp.1_mean) > 2 * Comp.1_sd |
  abs(Comp.2 - Comp.2_mean) > 2 * Comp.2_sd |
  abs(Comp.3 - Comp.3_mean) > 2 * Comp.3_sd
)
head(scores.outliers |> select(player, position, outlier_flag))
```

```
##           player position outlier_flag
## 1    Ichikawa Yuki      OH          TRUE
## 2      Romano Yuri      OP         FALSE
## 3 Abdel-Aziz Nimir      OP         FALSE
## 4 Herrera Jaime Jesus      OP         FALSE
## 5    Takahashi Ran      OH         FALSE
## 6    Luburic Drazen      OP         FALSE
```

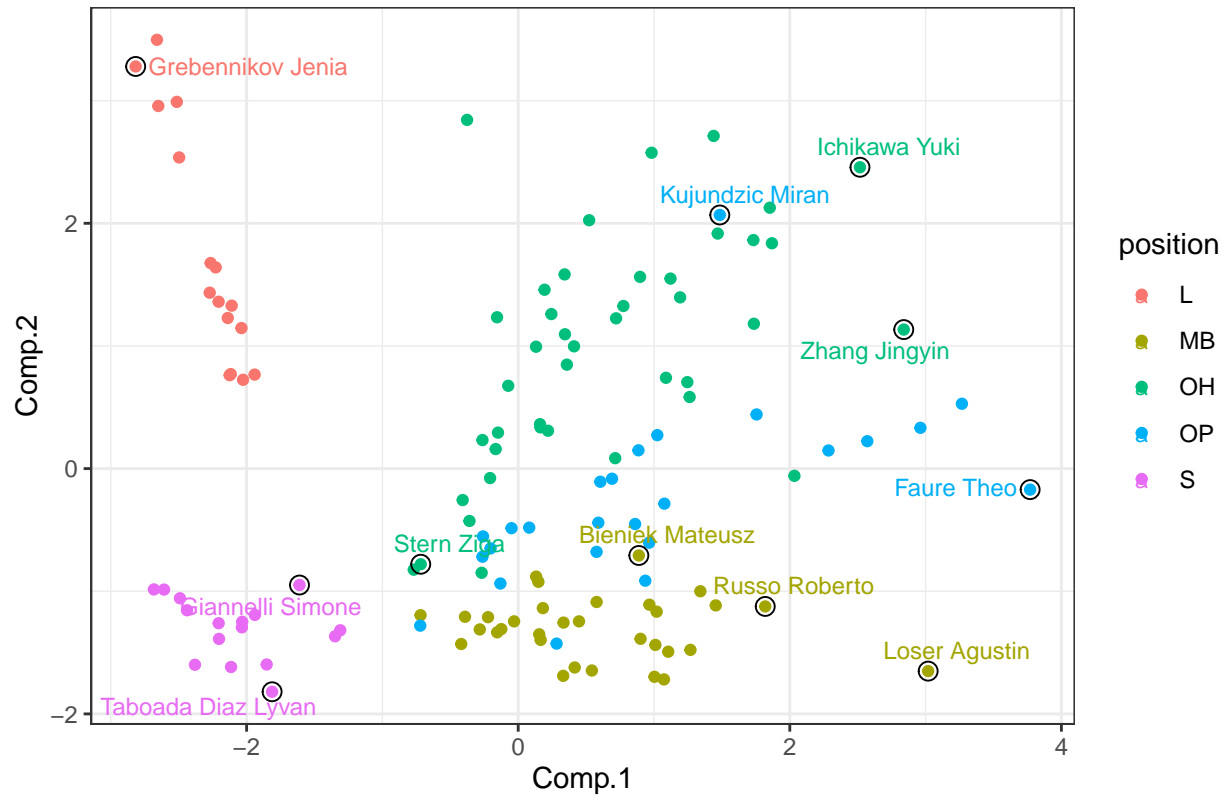
```

outliers <- scores.outliers |> rstatix::filter(outlier_flag == TRUE)
outliers <- outliers |>
  left_join(vnl_players |> select(-position), by = "player")
outliers[, c("player", "position",
             "Comp.1", "Comp.2", "Comp.3")]

```

##	player	position	Comp.1	Comp.2	Comp.3
## 1	Ichikawa Yuki	OH	2.5166578	2.4570085	0.9591354
## 2	Zhang Jingyin	OH	2.8392051	1.1335944	0.8210198
## 3	Kujundzic Miran	OP	1.4845232	2.0679013	0.8791899
## 4	Loser Agustin	MB	3.0187513	-1.6512164	0.3572480
## 5	Faure Theo	OP	3.7696947	-0.1712018	1.5635477
## 6	Russo Roberto	MB	1.8190521	-1.1237026	0.4061402
## 7	Bieniek Mateusz	MB	0.8884874	-0.7081993	0.1052268
## 8	Stern Ziga	OH	-0.7176721	-0.7797259	-1.3761095
## 9	Giannelli Simone	S	-1.6104773	-0.9490658	3.8161457
## 10	Taboada Diaz Lyvan	S	-1.8122342	-1.8194719	1.0425933
## 11	Grebennikov Jenia	L	-2.8168382	3.2787328	0.1580007

Players with different playing styles (PC1 vs. PC2)



**Nhận xét:** Như vậy, có 11 cầu thủ được xét vào diện nổi bật hơn so với các cầu thủ còn lại trong mùa giải.

### 1. Cầu thủ đa năng (Chỉ số vượt trội ở nhiều khía cạnh):

- **Ichikawa Yuki (OH):** toàn diện cả tấn công (PC1: 2.52), phòng thủ (PC2: 2.46) và tương đối ổn cả tổ chức chuyên bóng (PC3: 0.96). Là một OH đa năng cực kỳ xuất sắc, nằm xa trung tâm nhóm OH ở cả ba chiều, đóng vai trò chiến thuật toàn diện.
- **Zhang Jingyin (OH):** PC1 = 2.84, PC2 = 1.13, PC3 = 0.82. Giống Ichikawa, thiên công nhưng vẫn giữ được phòng thủ ổn. Là một chủ công có khả năng công thủ toàn diện.
- **Kujundzic Miran (OP):** PC2 = 2.07 cực kỳ cao so với OP trung bình (-0.24) trong khi đây là vai trò của Libero chứ không phải thiên về tấn công như OP. PC1 = 1.48 và PC3 = 0.87 cũng cao so với trung bình. Là một OP đa năng hiếm thấy.

Đây là những cầu thủ có thể thích nghi tốt nhiều vị trí và đóng vai trò chiến lược trong thi đấu.

### 2. Cầu thủ vượt trội thiên về một mặt, xuất sắc cá biệt:

- **Giannelli Simone (S):** PC3 = 3.82 > trung bình S = 1.83 cho thấy kỹ năng chuyền cực xuất sắc. PC2 = -0.95, PC1 = -1.61 vẫn nằm đúng vùng Setter. Là một Setter mẫu mực, vượt trội về kỹ năng đặc trưng, đóng vai trò thủ lĩnh điều phối lối chơi.
- **Faure Theo (OP):** PC1 = 3.77 rất cao, cho thấy khả năng tấn công toàn diện. Chuyên công mạnh, dứt điểm cực tốt, phù hợp với OP kiểu cổ điển. Tuy nhiên có khả năng chuyền bóng bất ngờ, PC3 = 1.56, có thể thử đổi chuyên chiến thuật hỗ trợ điều phối bóng.
- **Grebennikov Jenia (L):** PC2 = 3.28, phòng thủ siêu đỉnh. Đây là một Libero đẳng cấp, rất đúng vai trò.

Những cầu thủ này “không toàn diện”, nhưng lại là “mũi nhọn chuyên biệt” cực kỳ hiệu quả.

### 3. Cầu thủ có phong độ yếu hoặc có dấu hiệu “thi đấu lệch vị trí”:

- **Bieniek Mateusz (MB):** PC1 = 0.89 so với trung bình cho thấy tạm ổn về tấn công. Tuy nhiên, PC2 = -0.71 lại cho thấy cầu thủ này phòng thủ yếu với vị trí MB và PC3 = 0.10 không đặc trưng chuyền bóng. MB có block kém, phòng thủ không tốt, có thể thử vai trò OP, phù hợp thiên công.
- **Stern Ziga (OH):** Chỉ số cực thấp ở cả ba trục, lệch hoàn toàn khỏi nhóm OH, không nổi bật ở kỹ năng nào. Do đó, nên xem lại vai trò hoặc đang thi đấu dưới phong độ.

Những trường hợp này cần được xem xét lại về chiến thuật sử dụng hoặc phong độ.

Các cầu thủ còn lại **Loser Agustin (MB)**, **Russo Roberto (MB)**, **Taboada Diaz Lyvan (S)** đều là những cầu thủ chơi khá tốt, ổn định, đúng vị trí nhưng chưa thật sự nổi bật vượt trội như các cầu thủ trên.

**Kết luận:** Như vậy, các huấn luyện viên có thể căn cứ trên kết quả phân tích này để đưa ra các quyết định chiến lược.

## Mở rộng: Phân nhóm cầu thủ dựa vào kết quả PCA bằng Clustering

Sau khi áp dụng Phân tích thành phần chính (PCA) để rút gọn dữ liệu xuống ba thành phần chính đầu tiên (PC1, PC2, PC3), ta tiếp tục thực hiện triển khai thuật toán phân cụm K-Means để phân nhóm các cầu thủ dựa trên đặc điểm kỹ thuật đã được rút trích.

Việc sử dụng PCA trước khi phân cụm giúp giảm nhiễu và loại bỏ sự phụ thuộc tuyến tính giữa các biến gốc, từ đó làm tăng hiệu quả và tính ổn định cho thuật toán K-Means.

Cụ thể, thuật toán K-Means được triển khai như sau:

```
set.seed(123)
km <- kmeans(vnl_players.pca$scores[, 1:3], centers = 5)
```

Đánh giá độ trùng khớp giữa phân nhóm theo thuật toán và vai trò thực tế của cầu thủ:

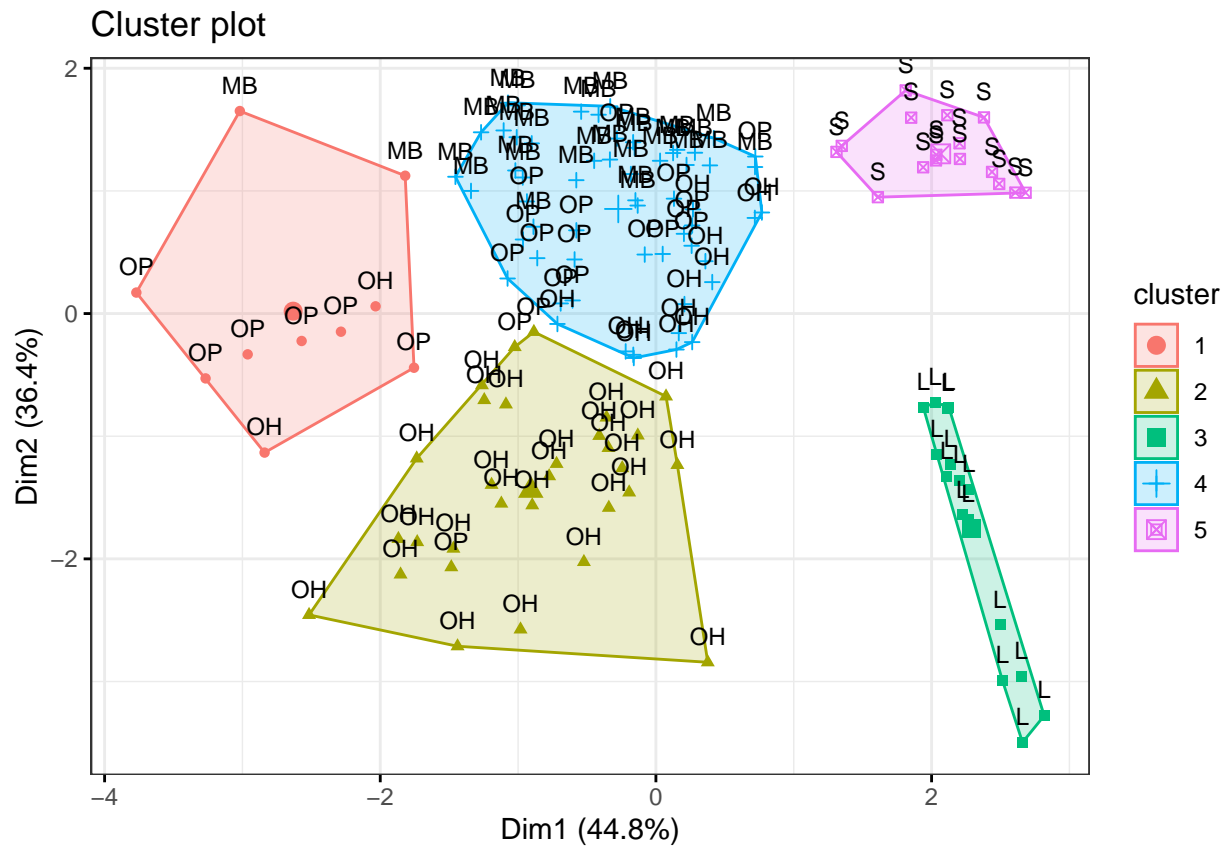
```
table(km$cluster, vnl_players$position)
```

```
##
##      L MB OH OP  S
##  1  0  2  2  6  0
##  2  0  0 27  3  0
##  3 16  0  0  0  0
##  4  0 30 13 16  0
##  5  0  0  0  0 16
```

**Nhận xét:** Đọc kết quả bảng số liệu so sánh:

- Cụm 1 gồm chủ yếu là OP (6), một số MB (2) và OH (2): đây là một cụm thiên về vị trí tấn công, nhưng có phần trộn lẫn giữa các vai trò, có thể là các cầu thủ có kỹ năng giao thoa.
- Cụm 2 gần như thuần OH (27), chỉ có 3 OP, cho thấy OH là nhóm có phong cách kỹ thuật rất đặc trưng và được phân biệt rõ ràng trong không gian PCA.
- Cụm 3 hoàn toàn là L (16). Đây là cụm phân biệt rất tốt, chứng tỏ đặc điểm kỹ thuật phòng thủ của Libero là rất riêng biệt.
- Cụm 4 là sự kết hợp giữa MB (30), OH (13) và OP (16) là một cụm tấn công hỗn hợp, đại diện cho các cầu thủ chơi thiên về tấn công nhưng khó phân tách hoàn toàn bằng kỹ năng.

- Cụm 5 hoàn toàn là S (16) – giống như Libero, Setter cũng có phong cách chuyên biệt rõ ràng về chuyền bóng và điều phối.



### Kết luận:

- Hai vị trí S và L được phân cụm rất rõ ràng lần lượt là cụm 5 và cụm 3, nằm trọn vẹn trong một cụm riêng, cho thấy hai vị trí này có kỹ năng đặc thù, dễ nhận diện bằng PCA.
- OH và OP có một phần trùng lặp, cho thấy khó phân biệt tuyệt đối giữa hai vị trí này chỉ qua các chỉ số kỹ thuật. Điều này phản ánh đúng thực tế khi hai vai trò này cùng chơi ở biên và đều tham gia tấn công nhiều.
- MB khá đa dạng, một phần vào cụm 1, phần lớn vào cụm 4. Lý do là vì MB vừa tham gia chắn bóng vừa hỗ trợ tấn công nhanh, dẫn đến đặc điểm kỹ thuật có sự pha trộn.

Tổng quát, việc kết hợp PCA + Clustering phân biệt tốt những vị trí có đặc điểm chuyên biệt (L, S), nhưng khó hơn với những vị trí mang tính lai và thiên về tấn công (OH, OP, MB).

### Xem xét các cầu thủ thuộc Cụm 1 (OH, OP, MB):



```
scores.pca$cluster <- km$cluster
cluster_1 <- scores.pca |> rstatix::filter(scores.pca$cluster == 1)
cluster_1$player
```

```
## [1] "Romano Yuri"          "Abdel-Aziz Nimir"    "Herrera Jaime Jesus"
## [4] "Luburic Drazen"       "Zhang Jingyin"        "Leon Venero Wilfredo"
## [7] "Loser Agustin"        "Esmaeilnezhad Amin"  "Faure Theo"
## [10] "Russo Roberto"
```

**Nhận xét:** Cụm 1 xuất hiện các cầu thủ với tên quen thuộc “Zhang Jingyin”, “Faure Theo”, “Russo Roberto” và “Loser Agustin”. Đây là những cái tên xuất hiện trong danh sách các cầu thủ nổi bật ở phần trước với cùng một điểm chung là có chỉ số năng lực khác vị trí đảm nhiệm cao bất ngờ hoặc chơi ổn định đều ở các năng lực. Do đó có thể nói rằng, cụm 1 chính là đại diện cho các cầu thủ có nhiều tiềm năng phát triển về sau ở nhiều vị trí các nhau, những cái tên còn lại trong nhóm cần được xem xét kỹ lưỡng, để ý và quan tâm nhiều hơn, hứa hẹn có một sự bùng nổ nếu khai thác được tiềm năng từ các cầu thủ này.

## Phương pháp 2: Sử dụng phương pháp phân tích nhân tố - FA:

Ngoài ra, tiếp tục thực hiện phân tích nhân tố FA bên cạnh phương pháp PCA để so sánh kết quả từ hai phương pháp. FA cung cấp cái nhìn sâu về mối quan hệ ẩn giữa các biến gốc dựa trên giả định mô hình. Do đó, em tiếp tục thực hiện FA với phép xoay nhân tố **varimax** để hiểu biết sâu sắc hơn về cấu trúc dữ liệu đang nghiên cứu.

### Chọn số nhân tố phù hợp

Sử dụng hàm `factanal()` trong phần mềm R (dùng phương pháp hợp lý cực đại), thực hiện phân tích nhân tố trên ma trận hệ số tương quan **R**:

```
round(cor_matrix, 4)
```

```
##          age  attack  block  serve    set    dig receive
## age      1.0000 -0.1778 -0.1010 -0.1084  0.1778  0.1671 -0.0111
## attack  -0.1778  1.0000  0.3384  0.7689 -0.4308 -0.0990  0.1699
## block   -0.1010  0.3384  1.0000  0.3360 -0.1320 -0.3483 -0.2652
## serve   -0.1084  0.7689  0.3360  1.0000 -0.1548 -0.0525  0.0396
## set      0.1778 -0.4308 -0.1320 -0.1548  1.0000  0.1317 -0.3059
## dig      0.1671 -0.0990 -0.3483 -0.0525  0.1317  1.0000  0.6247
## receive -0.0111  0.1699 -0.2652  0.0396 -0.3059  0.6247  1.0000
```

**Mô hình hai nhân tố:**

```

vnl_players.ft <- factanal(covmat = cor_matrix, factors = 2,
                           rotation = "varimax", n.obs = 131)
vnl_players.ft

##
## Call:
## factanal(factors = 2, covmat = cor_matrix, n.obs = 131, rotation = "varimax")
##
## Uniquenesses:
##      age  attack  block  serve   set   dig receive
##  0.968   0.005   0.776   0.398   0.759   0.564   0.005
##
## Loadings:
##           Factor1 Factor2
## age       -0.178
## attack    0.997
## block     0.335  -0.334
## serve     0.769  -0.104
## set       -0.434  -0.229
## dig              0.654
## receive   0.184   0.980
##
##           Factor1 Factor2
## SS loadings      1.961   1.564
## Proportion Var   0.280   0.223
## Cumulative Var   0.280   0.504
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 39.57 on 8 degrees of freedom.
## The p-value is 3.86e-06

```

**Nhận xét:** Qua kết quả R, ta thấy mô hình với hai nhân tố để phân tích cho bộ dữ liệu này là chưa đủ vì p-value rất nhỏ ( $3.86e-06$ ) trong kiểm định mô hình, tức bác bỏ “ $H_0$ : mô hình với 2 nhân tố là đủ”, với mức ý nghĩa  $\alpha = 0.05$ .

**Mô hình ba nhân tố:**

```

vnl_players.ft <- factanal(covmat = cor_matrix, factors = 3,
                           rotation = "varimax", n.obs = 131)
vnl_players.ft

##
## Call:

```

```
## factanal(factors = 3, covmat = cor_matrix, n.obs = 131, rotation = "varimax")
##
## Uniquenesses:
##      age  attack  block  serve    set    dig receive
##    0.921  0.160   0.715  0.169   0.344  0.252  0.204
##
## Loadings:
##           Factor1 Factor2 Factor3
## age                0.246
## attack    0.820         -0.409
## block     0.351  -0.373  -0.151
## serve     0.909
## set      -0.123  -0.101   0.794
## dig                0.822   0.268
## receive         0.849  -0.271
##
##           Factor1 Factor2 Factor3
## SS loadings    1.649   1.556   1.029
## Proportion Var  0.236   0.222   0.147
## Cumulative Var  0.236   0.458   0.605
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 0.82 on 3 degrees of freedom.
## The p-value is 0.845
```

**Nhận xét:** Mô hình ba nhân tố thỏa kiểm định ba nhân tố là đủ với  $p\text{-value} = 0.845 > 0.05$ . Do đó, chọn số nhân tố cho mô hình là 3.

## Phân tích kết quả mô hình ba nhân tố

Kết quả mô hình ba nhân tố:

```
vnl_players.ft
```

```
##
## Call:
## factanal(factors = 3, covmat = cor_matrix, n.obs = 131, rotation = "varimax")
##
## Uniquenesses:
##      age  attack  block  serve    set    dig receive
##    0.921  0.160   0.715  0.169   0.344  0.252  0.204
##
## Loadings:
```

```
##          Factor1 Factor2 Factor3
## age                0.246
## attack    0.820          -0.409
## block     0.351   -0.373   -0.151
## serve     0.909
## set       -0.123   -0.101    0.794
## dig                0.822    0.268
## receive                0.849   -0.271
##
##          Factor1 Factor2 Factor3
## SS loadings      1.649    1.556    1.029
## Proportion Var   0.236    0.222    0.147
## Cumulative Var   0.236    0.458    0.605
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 0.82 on 3 degrees of freedom.
## The p-value is 0.845
```

### Nhận xét:

1. Tổng phương sai được giải thích: Mô hình năm nhân tố giải thích khoảng 60.5% phương sai chung trong dữ liệu, đây là mức khá tốt, cho thấy mô hình có khả năng tóm tắt dữ liệu hiệu quả mà vẫn giữ lại phần lớn thông tin quan trọng.
2. Tác động tổng của các nhân tố chung lên từng biến:

```
# communalities
1-vnl_players.ft$uniquenesses
```

```
##          age      attack      block      serve      set      dig      receive
## 0.07879296 0.84009865 0.28512049 0.83060010 0.65560021 0.74817564 0.79593265
```

- Nhóm 6 biến kỹ năng, ngoại trừ **block** thì tất cả các biến còn lại đều có communality cao ( $> 0.6$ ).
- Biến **age** hầu như không liên quan đến cấu trúc nhân tố.

3. Factor loadings:

```
loadings(vnl_players.ft)
```

```
##
## Loadings:
##          Factor1 Factor2 Factor3
## age                0.246
```

```

## attack    0.820          -0.409
## block     0.351   -0.373   -0.151
## serve     0.909
## set       -0.123   -0.101    0.794
## dig                0.822    0.268
## receive                0.849   -0.271
##
##                               Factor1 Factor2 Factor3
## SS loadings                1.649    1.556    1.029
## Proportion Var             0.236    0.222    0.147
## Cumulative Var             0.236    0.458    0.605

```

**Factor 1** – Nhân tố về khả năng tấn công chủ động:

- Tải cao ở: **attack** (0.820), **serve** (0.909).
- Nhân tố này đại diện cho nhóm kỹ năng thiên về tấn công, đặc biệt là kỹ năng tấn công chủ động và phát bóng. Những cầu thủ có điểm cao ở nhân tố này thường là Outside Hitter (OH), Opposite (OP), hoặc Middle Blocker (MB).

**Factor 2** – Nhân tố về khả năng phòng thủ và phản xạ:

- Tải cao ở: **dig** (0.822), **receive** (0.849).
- Nhân tố này thể hiện khả năng phòng thủ, đặc biệt là kỹ năng cứu bóng và đỡ bước một. Đây là nhóm kỹ năng quan trọng với Libero (L) và các vị trí phòng thủ.

**Factor 3** – Nhân tố về khả năng tổ chức chuyền bóng:

- Tải cao ở: **set** (0.794).
- Nhân tố này đại diện cho kỹ năng tổ chức lối chơi, đặc biệt là chuyền hai. Rất đặc trưng cho vị trí Setter (S).
- Ở nhân tố này đặc biệt có sự góp mặt của biến **age** với hệ số tải dương 0.246. Điều này thể hiện rõ mối liên hệ tương đối của tuổi và kỹ năng tổ chức chiến lược. Khác với các vị trí thiên về thể lực như OH hay OP, Setter, vị trí tổ chức chuyền bóng không cần quá nhanh nhạy về thể chất, mà cần kinh nghiệm thi đấu, sự điềm tĩnh, khả năng quan sát và phán đoán, vốn là các kỹ năng được tích lũy theo thời gian. Do đó, dễ hiểu khi **age** chỉ đóng góp vào nhân tố này mặc dù tác động không đủ lớn.

## Sử dụng kết quả FA để đánh giá năng lực cầu thủ

Bảng điểm đánh giá các cầu thủ ở từng năng lực theo nhân tố:

```

dat_vnl_players$age <- vnl_players$age
vnl_players.ft <- factanal(dat_vnl_players, factors = 3,
                           scores = "regression", rotation = "varimax")
scores.ft <- as.data.frame(vnl_players.ft$scores)
scores.ft$player <- vnl_players$player
scores.ft$position <- vnl_players$position
scores.ft <- scores.ft|> select(player, position, Factor1, Factor2, Factor3)
head(scores.ft)

```

```

##           player position  Factor1    Factor2    Factor3
## 1      Ichikawa Yuki      OH 2.058547  1.3624895 -0.65206037
## 2        Romano Yuri      OP 1.920333 -0.3742869  0.20451081
## 3   Abdel-Aziz Nimir      OP 2.974381 -0.3909057  0.34869500
## 4 Herrera Jaime Jesus      OP 2.521398 -0.4174520  0.13413185
## 5      Takahashi Ran      OH 1.152418  1.5835043 -0.37299347
## 6      Luburic Drazen      OP 1.591571 -0.2222093 -0.05046336

```

*Top 5 cầu thủ chuyên tấn công:*

```

scores.ft |> arrange(-Factor1) |> head() |> select(player, position, Factor1)

```

```

##           player position  Factor1
## 1   Abdel-Aziz Nimir      OP 2.974381
## 2 Herrera Jaime Jesus      OP 2.521398
## 3         Faure Theo      OP 2.396137
## 4 Esmaeilnezhad Amin      OP 2.236297
## 5      Ichikawa Yuki      OH 2.058547
## 6        Romano Yuri      OP 1.920333

```

**Nhận xét:** Các vận động viên có điểm số cao nhất ở nhân tố tấn công đều là những OP (đối chuyền), là những tay ghi điểm chính trong đội hình. Các cầu thủ này sở hữu khả năng tấn công toàn diện, nổi bật cả về tấn công, phát bóng và chắn bóng. Đặc biệt, **Zhang Jingyin (OH)** là chủ công hiếm hoi góp mặt trong nhóm này, thể hiện anh là một cầu thủ nặng ký.

*Top 5 cầu thủ chuyên phòng thủ:*

```

scores.ft |> arrange(-Factor2) |> head() |> select(player, position, Factor2)

```

```

##           player position  Factor2
## 1      Kovacic Jani      L 2.865590
## 2 Grebennikov Jenia      L 2.704158

```

## 3	Danani Santiago	L 2.450737
## 4	Balaso Fabio	L 2.426202
## 5	Yamamoto Tomohiro	L 2.096260
## 6	Tillie Kevin	OH 2.035975

**Nhận xét:** Libero luôn dẫn đầu về năng lực phòng thủ. Tuy nhiên, **Tillie Kevin (OH)** xuất hiện cho thấy một chủ công cũng có khả năng phòng ngự cực tốt, một trường hợp cầu thủ đa năng.

*Top 5 cầu thủ chuyên chuyên bóng tổ chức lối chơi:*

```
scores.ft |> arrange(-Factor3) |> head() |> select(player, position, Factor3)
```

##	player	position	Factor3
## 1	De Cecco Luciano	S	3.424620
## 2	Giannelli Simone	S	3.236210
## 3	Sekita Masahiro	S	2.408187
## 4	Christenson Micah	S	2.403926
## 5	Rezende Bruno Mossa	S	2.238613
## 6	Brizard Antoine	S	2.182889

**Nhận xét:** Các Setter hàng đầu có điểm cực cao ở khả năng tổ chức, chuyên bóng và điều phối lối chơi, đúng với vai trò chiến thuật của họ.

**Kết luận:** Qua một số phân tích và thống kê bên trên, ta nhận thấy FA là một công cụ tốt để xây dựng bảng điểm năng lực của từng cầu thủ. Do đó, huấn luyện viên có thể theo dõi bảng điểm này để đánh giá năng lực cầu thủ và xây dựng hệ thống rèn luyện phù hợp cho từng cá nhân. Đồng thời, có thể sử dụng FA cho từng giai đoạn huấn luyện để có thể đánh giá tiến bộ của cầu thủ khi so sánh các bảng điểm năng lực với nhau.

**Bổ sung: Kiểm định MANOVA dựa vào kết quả FA**

**Mục tiêu:** Liệu quốc tịch (**country**) và/hoặc độ tuổi (**age**) có ảnh hưởng đến các năng lực tiềm ẩn của các vận động viên?

**Mô hình kiểm định:** Two-way MANOVA.

- Hai nhân tố: **country**, **age**.
- MANOVA với các biến phụ thuộc: **F1**, **F2**, **F3**.

```
scores.ft$country <- vnl_players$country
scores.ft$age <- vnl_players$age
```

Bỏ qua các giả định phức tạp của phương pháp MANOVA, em chọn thực hiện Permutation MANOVA cho bài toán này:

```
library(vegan)
X <- scores.ft[, c("Factor1", "Factor2", "Factor3")]
perm_result <- adonis2(X ~ country + age, data = scores.ft, permutations = 999)
print(perm_result)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = X ~ country + age, data = scores.ft, permutations = 999)
##          Df SumOfSqs      R2      F Pr(>F)
## Model      16   493172 0.74117 20.403  0.208
## Residual  114   172221 0.25883
## Total     130   665393 1.00000
```

**Nhận xét:** Giá trị  $p = 0.208$  ( $> 0.05$ ) cho thấy mô hình không có ý nghĩa thống kê với mức ý nghĩa 5%. Tức là:

Không có bằng chứng đủ mạnh để khẳng định rằng tuổi và quốc tịch có ảnh hưởng đồng thời đến bộ ba năng lực tiềm ẩn của cầu thủ.

## So sánh hai phương pháp

Cả phân tích thành phần chính (PCA) và phân tích nhân tố (FA) đều chỉ ra rằng dữ liệu cầu thủ có thể được mô tả hiệu quả bằng ba năng lực chính: **tấn công**, **phòng thủ** và **chuyên bóng tổ chức lối chơi**. Sự tương đồng giữa hai phương pháp này cho thấy cấu trúc kỹ năng trong bộ dữ liệu là ổn định, rõ ràng, nhất quán và có tính giải thích cao.

Trong nghiên cứu này, cả Phân tích thành phần chính (PCA) và Phân tích nhân tố (FA) đều được khai thác để hiểu sâu về đặc điểm thi đấu của các vận động viên. Mỗi phương pháp đều có ưu thế riêng, phục vụ các mục tiêu khác nhau:

- PCA đặc biệt hữu ích trong giai đoạn khám phá dữ liệu ban đầu, khi mục tiêu là giảm chiều, tìm ra các mẫu phân nhóm và nhận diện các cầu thủ có phong cách thi đấu tương đồng hoặc ngoại lệ. Nhờ đó, PCA giúp xây dựng cái nhìn tổng thể, trực quan hóa cấu trúc dữ liệu và hỗ trợ các bước như phân cụm.



- FA lại phù hợp hơn khi mục tiêu là phân tích sâu các năng lực tiềm ẩn chi phối thành tích thi đấu. Thông qua việc trích xuất các nhân tố ẩn như khả năng tấn công, phòng thủ hay tổ chức lối chơi, FA cung cấp nền tảng để xây dựng bảng điểm năng lực cá nhân hóa và thiết kế chiến lược huấn luyện phù hợp cho từng vận động viên.

Do đó, việc kết hợp PCA để khám phá và FA để đào sâu là một hướng tiếp cận toàn diện, vừa giúp hiểu được cấu trúc dữ liệu, vừa phục vụ tốt mục tiêu ra quyết định huấn luyện mang tính chiến lược.

## Kết luận

Thông qua việc kết hợp các phương pháp phân tích dữ liệu đa chiều như Phân tích thành phần chính (PCA) và Phân tích nhân tố (FA), đề án đã đạt được mục tiêu khám phá đặc điểm thi đấu của các vận động viên bóng chuyền quốc tế dựa trên các chỉ số kỹ thuật chuyên môn. Kết quả từ PCA cho thấy rõ cấu trúc phân nhóm vị trí thi đấu, hỗ trợ hiệu quả việc phát hiện các cầu thủ có phong cách thi đấu khác biệt, đa năng hoặc lệch chuẩn. Tiếp theo, phân tích FA đã trích xuất ba năng lực tiềm ẩn chính (tấn công chủ động, phòng thủ phản xạ, tổ chức chiến thuật), giúp xây dựng bảng điểm năng lực hỗ trợ huấn luyện chuyên sâu.

Điểm đặc biệt của bộ dữ liệu được thể hiện khi cả PCA và FA đều cho ra cùng một kết quả phân tích giống nhau. Song, phương pháp PCA phù hợp cho bước khám phá tổng quan và phân nhóm vận động viên, trong khi FA có ưu thế trong việc đánh giá năng lực tiềm ẩn và xây dựng chiến lược huấn luyện cá nhân hóa. Việc kết hợp cả hai phương pháp mang lại một cách tiếp cận toàn diện, vừa khám phá được cấu trúc dữ liệu, vừa đưa ra định hướng hành động cụ thể. Đề án mở ra tiềm năng ứng dụng thực tiễn trong công tác tuyển chọn, huấn luyện và phân tích chiến thuật cho các đội tuyển thể thao chuyên nghiệp.