Euro 2024 Players Analysis

Juma

2024-08-17

Load Packages

```
library(flexdashboard)
library(tidyverse)
library(cluster)
library(plotly)
library(data.table)
library(DT)
library(htmltools)
library(webshot2)
```

Set working directory and load data

```
setwd("C:/Users/hp/Desktop/Dataset")
euro_24 <- read_csv("euro2024_players.csv")
```

```
## Rows: 623 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): Name, Position, Club, Foot, Country
## dbl (5): Age, Height, Caps, Goals, MarketValue
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
euro_24
```

```
## # A tibble: 623 x 10
##    Name        Position   Age Club  Height Foot   Caps Goals MarketValue Country
##    <chr>       <chr>    <dbl> <chr>  <dbl> <chr> <dbl> <dbl>       <dbl> <chr>
##  1 Marc-André~ Goalkee~    32 FC B~    187 right    40     0    28000000 Germany
##  2 Manuel Neu~ Goalkee~    38 Baye~    193 right   119     0     4000000 Germany
##  3 Oliver Bau~ Goalkee~    34 TSG ~    187 right     0     0     3000000 Germany
##  4 Nico Schlo~ Centre-~    24 Boru~    191 left     12     0    40000000 Germany
##  5 Jonathan T~ Centre-~    28 Baye~    195 right    25     0    30000000 Germany
##  6 Antonio Rü~ Centre-~    31 Real~    190 right    69     3    25000000 Germany
```

```
##  7 Waldemar A~ Centre-~    27 VfB ~    189 right     2     0    20000000 Germany
##  8 Robin Koch  Centre-~    27 Eint~    191 right     9     0    18000000 Germany
##  9 David Raum  Left-Ba~    26 RB L~    180 left     21     0    20000000 Germany
## 10 Maximilian~ Left-Ba~    27 VfB ~    180 left      4     1    17000000 Germany
## # i 613 more rows
```

View Data

```
view(euro_24)
```

Data Exploration

```
glimpse(euro_24)
```

```
## Rows: 623
## Columns: 10
## $ Name        <chr> "Marc-André ter Stegen", "Manuel Neuer", "Oliver Baumann",~
## $ Position    <chr> "Goalkeeper", "Goalkeeper", "Goalkeeper", "Centre-Back", "~
## $ Age         <dbl> 32, 38, 34, 24, 28, 31, 27, 27, 26, 27, 29, 27, 20, 29, 33~
## $ Club        <chr> "FC Barcelona", "Bayern Munich", "TSG 1899 Hoffenheim", "B~
## $ Height      <dbl> 187, 193, 187, 191, 195, 190, 189, 191, 180, 180, 177, 185~
## $ Foot        <chr> "right", "right", "right", "left", "right", "right", "righ~
## $ Caps        <dbl> 40, 119, 0, 12, 25, 69, 2, 9, 21, 4, 86, 15, 1, 5, 77, 109~
## $ Goals       <dbl> 0, 0, 0, 0, 0, 3, 0, 0, 0, 1, 6, 0, 0, 0, 18, 17, 1, 1, 2,~
## $ MarketValue <dbl> 2.8e+07, 4.0e+06, 3.0e+06, 4.0e+07, 3.0e+07, 2.5e+07, 2.0e~
## $ Country     <chr> "Germany", "Germany", "Germany", "Germany", "Germany", "Ge~
```

```
class(euro_24)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

```
summary(euro_24)
```

```
##      Name             Position              Age            Club
##  Length:623         Length:623         Min.   :16.00   Length:623
##  Class :character   Class :character   1st Qu.:24.00   Class :character
##  Mode  :character   Mode  :character   Median :27.00   Mode  :character
##                                        Mean   :27.04
##                                        3rd Qu.:30.00
##                                        Max.   :41.00
##      Height          Foot                Caps            Goals
##  Min.   :167.0   Length:623         Min.   :  0.00   Min.   :  0.000
##  1st Qu.:180.0   Class :character   1st Qu.:  7.00   1st Qu.:  0.000
##  Median :185.0   Mode  :character   Median : 21.00   Median :  1.000
##  Mean   :184.2                      Mean   : 30.34   Mean   :  4.152
##  3rd Qu.:189.0                      3rd Qu.: 42.00   3rd Qu.:  4.000
```

```
##  Max.   :202.0                      Max.   :206.00   Max.   :128.000
##   MarketValue          Country
##  Min.   :    50000   Length:623
##  1st Qu.:  2900000   Class :character
##  Median :  9000000   Mode  :character
##  Mean   : 18409029
##  3rd Qu.: 25000000
##  Max.   :180000000
```

```
head(euro_24,10)
```

```
## # A tibble: 10 x 10
##    Name        Position   Age Club  Height Foot   Caps Goals MarketValue Country
##    <chr>       <chr>    <dbl> <chr>  <dbl> <chr> <dbl> <dbl>       <dbl> <chr>
##  1 Marc-André~ Goalkee~    32 FC B~    187 right    40     0    28000000 Germany
##  2 Manuel Neu~ Goalkee~    38 Baye~    193 right   119     0     4000000 Germany
##  3 Oliver Bau~ Goalkee~    34 TSG ~    187 right     0     0     3000000 Germany
##  4 Nico Schlo~ Centre-~    24 Boru~    191 left     12     0    40000000 Germany
##  5 Jonathan T~ Centre-~    28 Baye~    195 right    25     0    30000000 Germany
##  6 Antonio Rü~ Centre-~    31 Real~    190 right    69     3    25000000 Germany
##  7 Waldemar A~ Centre-~    27 VfB ~    189 right     2     0    20000000 Germany
##  8 Robin Koch  Centre-~    27 Eint~    191 right     9     0    18000000 Germany
##  9 David Raum  Left-Ba~    26 RB L~    180 left     21     0    20000000 Germany
## 10 Maximilian~ Left-Ba~    27 VfB ~    180 left      4     1    17000000 Germany
```

```
tail(euro_24,10)
```

```
## # A tibble: 10 x 10
##    Name        Position   Age Club  Height Foot   Caps Goals MarketValue Country
##    <chr>       <chr>    <dbl> <chr>  <dbl> <chr> <dbl> <dbl>       <dbl> <chr>
##  1 Pavel Sulc  Attacki~    23 FC V~    177 right     1     0     6000000 Czech ~
##  2 Antonín Ba~ Attacki~    29 ACF ~    190 left     39     8     4500000 Czech ~
##  3 Ondrej Lin~ Attacki~    25 Feye~    175 right    13     0     3500000 Czech ~
##  4 Matej Jura~ Right W~    20 SK S~    181 -         1     0     8000000 Czech ~
##  5 Vaclav Cer~ Right W~    26 VfL ~    182 left     15     5     7000000 Czech ~
##  6 Adam Hlozek Second ~    21 Baye~    188 right    31     2    12000000 Czech ~
##  7 Patrik Sch~ Centre-~    28 Baye~    191 left     37    18    22000000 Czech ~
##  8 Mojmír Chy~ Centre-~    25 SK S~    187 -        12     4     6500000 Czech ~
##  9 Jan Kuchta  Centre-~    27 AC S~    185 right    20     3     5000000 Czech ~
## 10 Tomas Chory Centre-~    29 FC V~    199 right     3     2     3200000 Czech ~
```

Data Wrangling

Data type conversion

```
euro_24$Age <- as.integer(euro_24$Age)
class(euro_24$Age)
```

```
## [1] "integer"
```

```r
euro_24$Goals <- as.integer(euro_24$Goals)
class(euro_24$Goals)
```

```
## [1] "integer"
```

```r
euro_24 <- euro_24%>%
  mutate(across(c(Country, Club, Position), as.factor))
glimpse(euro_24)
```

```
## Rows: 623
## Columns: 10
## $ Name        <chr> "Marc-André ter Stegen", "Manuel Neuer", "Oliver Baumann",~
## $ Position    <fct> Goalkeeper, Goalkeeper, Goalkeeper, Centre-Back, Centre-Ba~
## $ Age         <int> 32, 38, 34, 24, 28, 31, 27, 27, 26, 27, 29, 27, 20, 29, 33~
## $ Club        <fct> FC Barcelona, Bayern Munich, TSG 1899 Hoffenheim, Borussia~
## $ Height      <dbl> 187, 193, 187, 191, 195, 190, 189, 191, 180, 180, 177, 185~
## $ Foot        <chr> "right", "right", "right", "left", "right", "right", "righ~
## $ Caps        <dbl> 40, 119, 0, 12, 25, 69, 2, 9, 21, 4, 86, 15, 1, 5, 77, 109~
## $ Goals       <int> 0, 0, 0, 0, 0, 3, 0, 0, 0, 1, 6, 0, 0, 0, 18, 17, 1, 1, 2,~
## $ MarketValue <dbl> 2.8e+07, 4.0e+06, 3.0e+06, 4.0e+07, 3.0e+07, 2.5e+07, 2.0e~
## $ Country     <fct> Germany, Germany, Germany, Germany, Germany, Germany, Germ~
```

Handling missing values

```r
unique(euro_24$Foot)
```

```
## [1] "right" "left"  "both"  NA      "-"
```

```r
missing_foot <- euro_24[is.na(euro_24$Foot),]
missing_foot
```

```
## # A tibble: 3 x 10
##   Name         Position   Age Club  Height Foot   Caps Goals MarketValue Country
##   <chr>        <fct>    <int> <fct>  <dbl> <chr> <dbl> <int>       <dbl> <fct>
## 1 Maximilian ~ Centre-~    26 TSV ~    186 <NA>      2     1     2000000 Austria
## 2 Thomas Kami~ Goalkee~    31 Luto~    190 <NA>      1     0     3000000 Belgium
## 3 Bogdan Raco~ Centre-~    24 Rakó~    187 <NA>      2     0     1800000 Romania
```

Replace missing values with the mode

```r
mode_foot <- as.character(names(sort(table(euro_24$Foot), decreasing = TRUE))[1])
mode_foot
```

```
## [1] "right"
```

```
(euro_24$Foot[is.na(euro_24$Foot)] <- mode_foot)
```

```
## [1] "right"
```

```
euro_24$Foot <- as.factor(euro_24$Foot)
```

Descriptive Statistics

Summary statistics for numerical columns

```
num_summary <- euro_24 %>%
    select_if(is.numeric) %>%
    summary()
num_summary
```

```
##       Age             Height           Caps             Goals
##  Min.   :16.00   Min.   :167.0   Min.   :  0.00   Min.   :  0.000
##  1st Qu.:24.00   1st Qu.:180.0   1st Qu.:  7.00   1st Qu.:  0.000
##  Median :27.00   Median :185.0   Median : 21.00   Median :  1.000
##  Mean   :27.04   Mean   :184.2   Mean   : 30.34   Mean   :  4.152
##  3rd Qu.:30.00   3rd Qu.:189.0   3rd Qu.: 42.00   3rd Qu.:  4.000
##  Max.   :41.00   Max.   :202.0   Max.   :206.00   Max.   :128.000
##   MarketValue
##  Min.   :    50000
##  1st Qu.:  2900000
##  Median :  9000000
##  Mean   : 18409029
##  3rd Qu.: 25000000
##  Max.   :180000000
```

Distribution of categorical variables

```
cat_summary <- euro_24 %>%
    select_if(is.factor) %>%
    summary()
cat_summary
```

```
##                Position                     Club        Foot
##  Centre-Back       :125   Inter Milan        : 13   -    :  3
##  Central Midfield  : 81   Manchester City    : 13   both : 31
##  Centre-Forward    : 77   FC Barcelona       : 12   left :150
##  Goalkeeper        : 72   Paris Saint-Germain: 12   right:439
##  Defensive Midfield: 50   Real Madrid        : 12
##  Attacking Midfield: 48   Bayern Munich      : 11
##  (Other)           :170   (Other)            :550
##           Country
##  Poland        : 27
```

```
##  Albania       : 26
##  Austria       : 26
##  Croatia       : 26
##  Czech Republic: 26
##  Denmark       : 26
##  (Other)       :466
```

Summary statistics for character variable

```
char_summary <- euro_24 %>%
    select_if(is.character) %>%
    summary()
char_summary
```

```
##      Name
##  Length:623
##  Class :character
##  Mode  :character
```

Demographics

There were a total of 623 Players

The average age of Players was 27

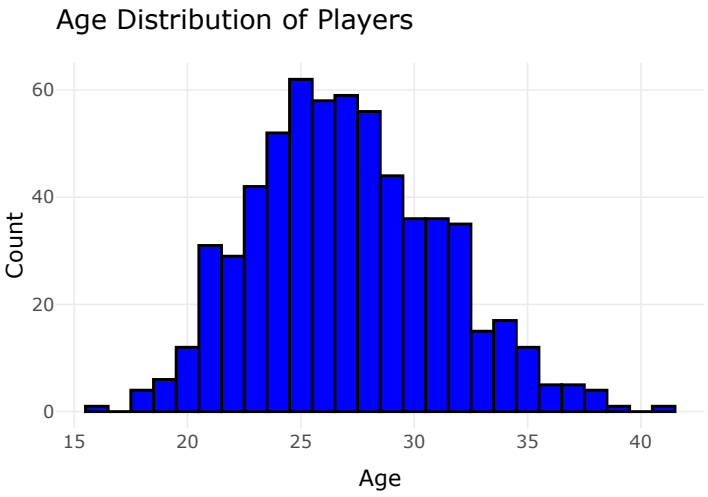
184 was the average height of Players

The total number of countries that participated in the Euro24 was 24

Age distribution

The histogram shows that the majority of players are in their mid-20s, with fewer players at both the younger (15-20 years) and older (30-40 years) ends of the spectrum.

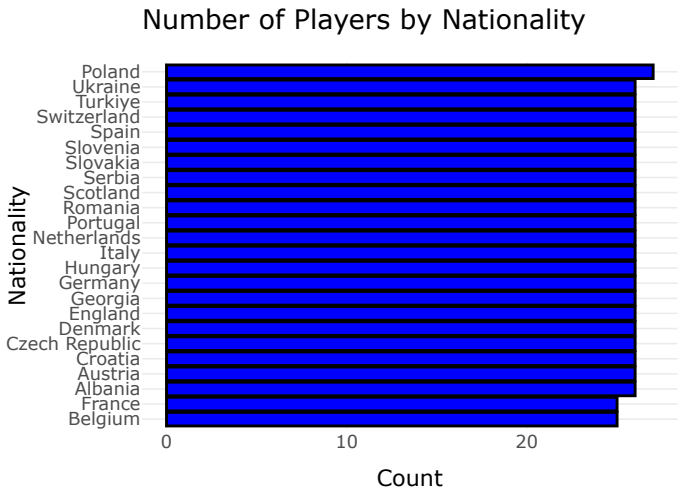
```
p <- ggplot(euro_24, aes(x = Age)) +
  geom_histogram(binwidth = 1, fill = 'blue', color = 'black') +
  theme_minimal() +
  labs(title = "Age Distribution of Players", x = "Age", y = "Count")
ggplotly(p)
```


Age Distribution of Players
Count
60
40
20
0
15
20
25
30
35
40
Age

Distribution by nationality

Poland had the highest number of players while countries like France and Belgium had the lowest.

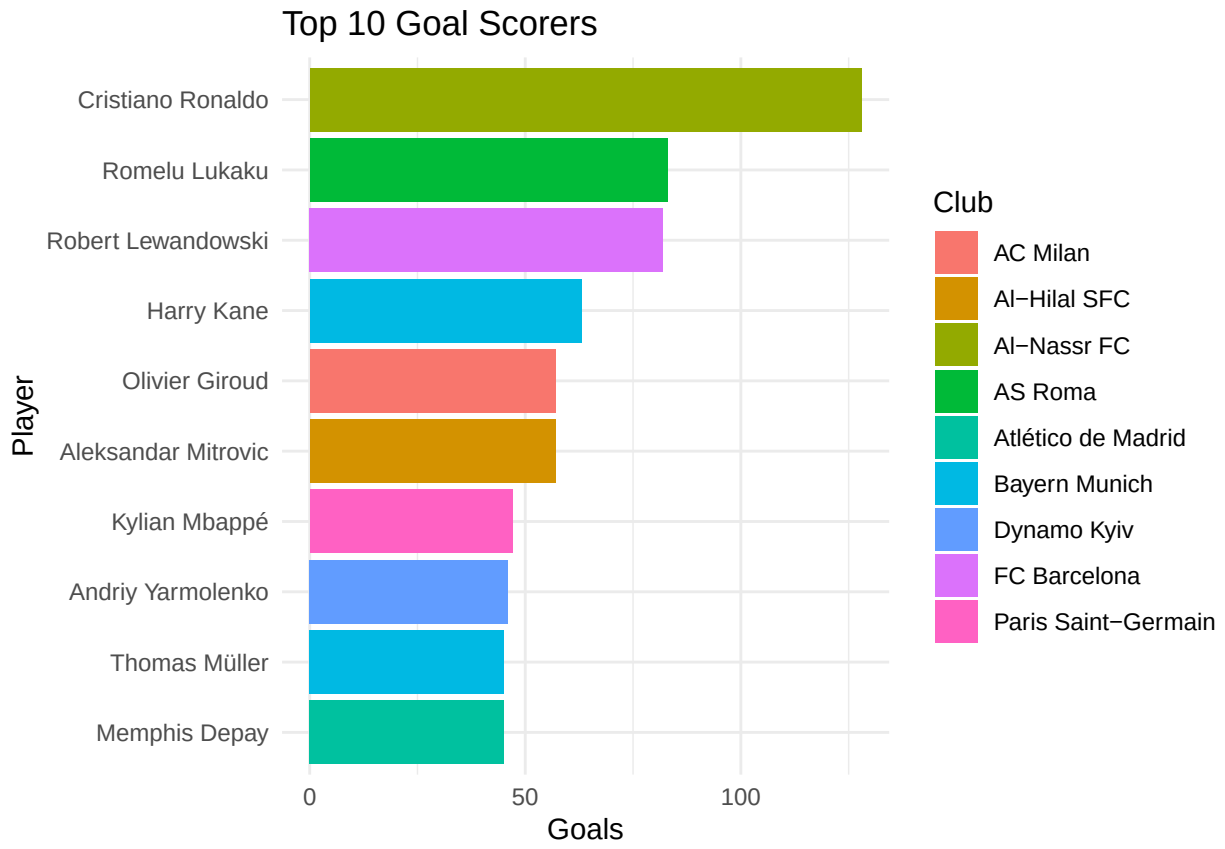
```
nationality_dist <-euro_24 %>%
    select(Country) %>%
    count(Country) %>%
    arrange(desc(n))
k <- ggplot(nationality_dist, aes(x = reorder(Country, n), y = n)) +
  geom_bar(stat = 'identity', fill = 'blue', color = 'black') +
  theme_minimal() +
  coord_flip() +
  labs(title = "Number of Players by Nationality", x = "Nationality", y = "Count")
ggplotly(k)
```


Number of Players by Nationality
Nationality
Poland
Ukraine
Turkiye
Switzerland
Spain
Slovenia
Slovakia
Serbia
Scotland
Romania
Portugal
Netherlands
Italy
Hungary
Germany
Georgia
England
Denmark
Czech Republic
Croatia
Austria
Albania
France
Belgium
0
10
20
Count

Top scorers

The bar chart shows the top 10 goal scorers. Cristiano Ronaldo has the highest score, as indicated by the longest bar while Kylian Mbappé, Andriy Yarmolenko, Thomas Müller, and Memphis Depay have shorter bars, indicating lower scores compared.

```
top_scorers <- euro_24 %>%
  arrange(desc(Goals)) %>%
  select(Name, Club, Country, Goals) %>%
  head(10)
ggplot(top_scorers, aes(x = reorder(Name, Goals), y = Goals, fill = Club)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Top 10 Goal Scorers",
       x = "Player",
       y = "Goals",
       fill = "Club") +
  theme_minimal()
```



Filter the relevant demographic columns

```
demographics_data <- euro_24 %>% select(Name, Age, Country, Club, Position, Height, Foot, Caps, Goals, 
datatable(demographics_data,
```

```
          options = list(pageLength = 10,
                         searchHighlight = TRUE),
          filter = 'top')
```

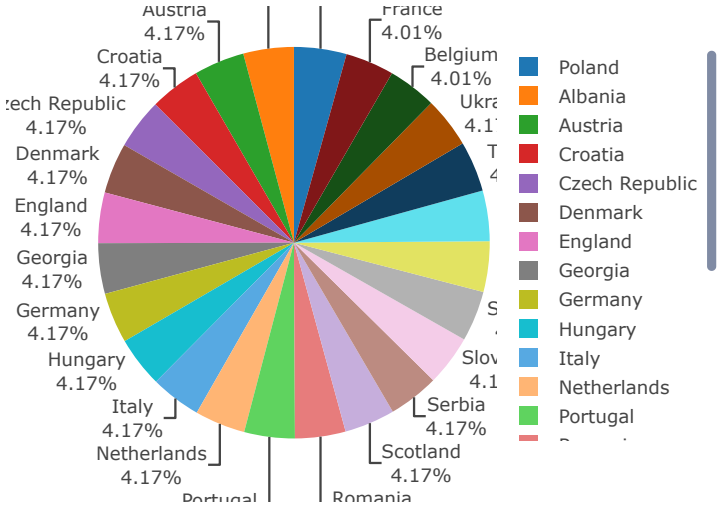
Show 10 entries Search:

	Name	Age	Country	Club	Position	Height	Foot	Caps	Goals	MarketValue
	All	All	All	All	All	All	All	All	All	All
1	Marc-André ter Stegen	32	Germany	FC Barcelona	Goalkeeper	187	right	40	0	28000000
2	Manuel Neuer	38	Germany	Bayern Munich	Goalkeeper	193	right	119	0	4000000
3	Oliver Baumann	34	Germany	TSG 1899 Hoffenheim	Goalkeeper	187	right	0	0	3000000
4	Nico Schlotterbeck	24	Germany	Borussia Dortmund	Centre-Back	191	left	12	0	40000000
5	Jonathan Tah	28	Germany	Bayer 04 Leverkusen	Centre-Back	195	right	25	0	30000000
6	Antonio Rüdiger	31	Germany	Real Madrid	Centre-Back	190	right	69	3	25000000
7	Waldemar Anton	27	Germany	VfB Stuttgart	Centre-Back	189	right	2	0	20000000
8	Robin Koch	27	Germany	Eintracht Frankfurt	Centre-Back	191	right	9	0	18000000
9	David Raum	26	Germany	RB Leipzig	Left-Back	180	left	21	0	20000000
10	Maximilian Mittelstädt	27	Germany	VfB Stuttgart	Left-Back	180	left	4	1	17000000

Showing 1 to 10 of 623 entries Previous 1 2 3 4 5 … 63 Next

Players National Distribution in Pie Chart

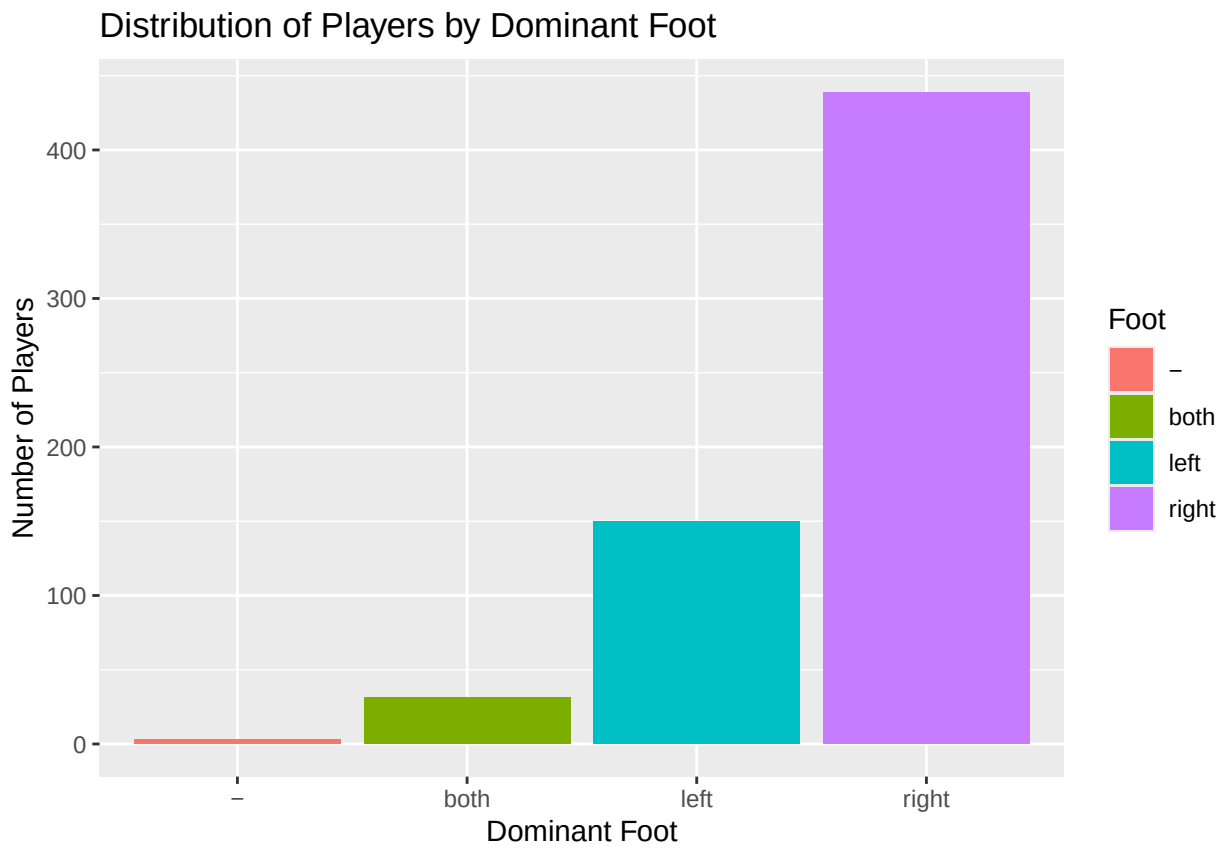
```
pie_chart <- plot_ly(nationality_dist, labels = ~Country, values = ~n, type = 'pie',
                     textinfo = 'label+percent', insidetextorientation = 'radial') %>%
  layout(xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),
         yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))
pie_chart
```


Austria
4.17%
France
4.01%
Croatia
4.17%
Belgium
4.01%
Ukra
4.1
ech Republic
4.17%
Denmark
4.17%
England
4.17%
Georgia
4.17%
Germany
4.17%
Hungary
4.17%
Italy
4.17%
Netherlands
4.17%
Portugal
Romania
Scotland
4.17%
Serbia
4.17%
Slov
4.1
Poland
Albania
Austria
Croatia
Czech Republic
Denmark
England
Georgia
Germany
Hungary
Italy
Netherlands
Portugal

Distribution of players according to their dominant foot

Right-footed players dominate the distribution, with nearly 400 players while "-" category has the least number of players, close to zero.

```
foot_distribution <- euro_24 %>%
  filter(!is.na(Foot)) %>%
  group_by(Foot) %>%
  summarise(Count = n())
ggplot(foot_distribution, aes(x = Foot, y = Count, fill = Foot)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Players by Dominant Foot",
       x = "Dominant Foot",
       y = "Number of Players")
```



Comparative Analysis

Club with the highest average age

```
team_age <- euro_24 %>%
  group_by(Club) %>%
  summarize(Average_Age = mean(Age, na.rm = TRUE)) %>%
```

```
  arrange(desc(Average_Age))
top_club <- team_age %>% slice(1)
top_club
```

```
## # A tibble: 1 x 2
##   Club           Average_Age
##   <fct>                <dbl>
## 1 Dinamo Tbilisi          38
```

Club with the highest average goals

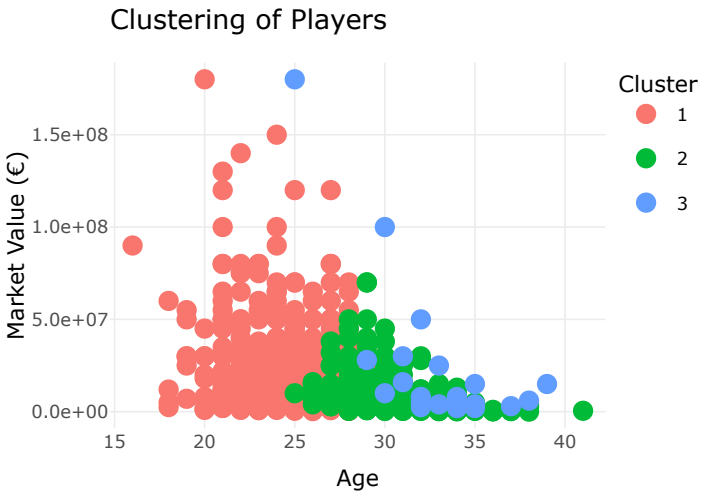
```
team_performance <- euro_24 %>%
  group_by(Club) %>%
  summarize(Average_Goals = mean(Goals, na.rm = TRUE)) %>%
  arrange(desc(Average_Goals))
# Display the top team
top_team <- team_performance[1, ]
top_team
```

```
## # A tibble: 1 x 2
##   Club        Average_Goals
##   <fct>               <dbl>
## 1 Al-Nassr FC          45.3
```

Visualization of Clusters

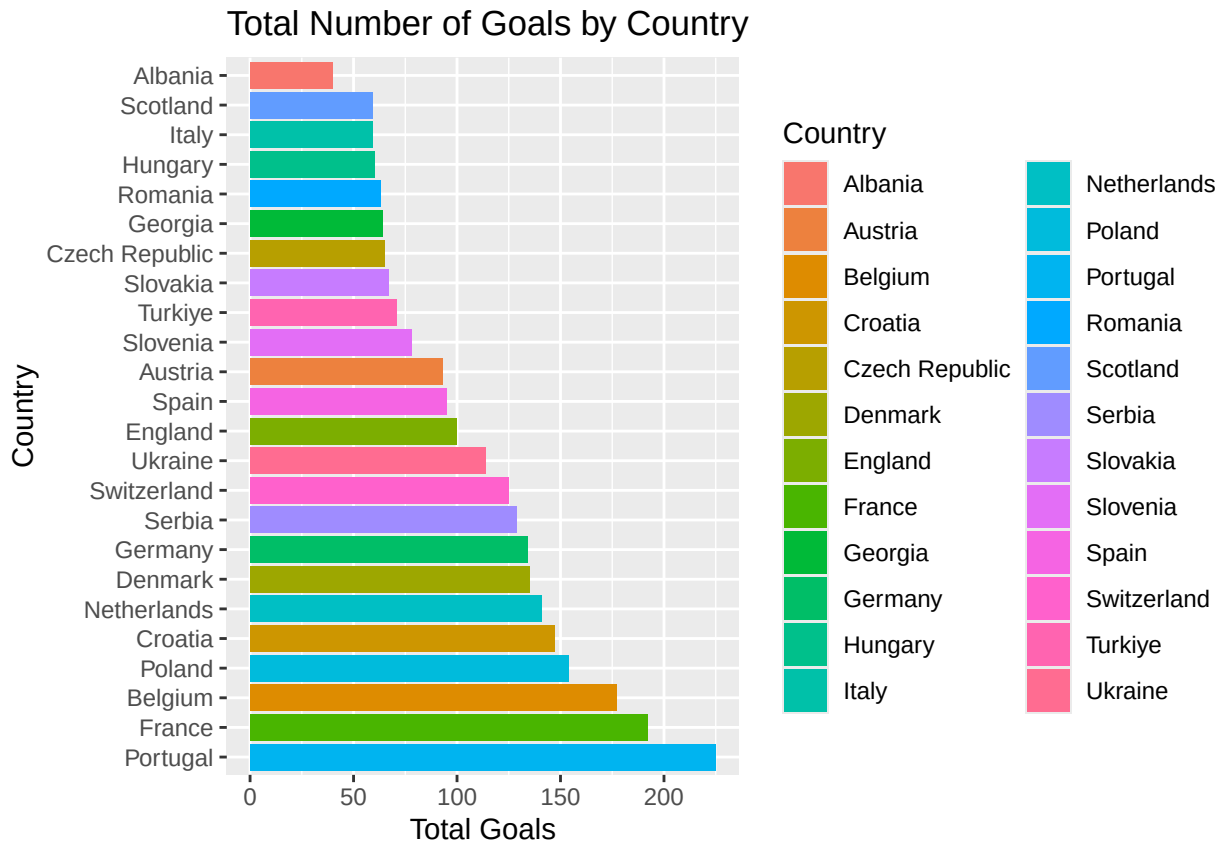
```
# Scaling Data and Clustering
num_data <- euro_24 %>%
  select_if(is.numeric) %>%
  na.omit()
num_data_scaled <- scale(num_data)
# K-means clustering
set.seed(123)
kmeans_result <- kmeans(num_data_scaled, centers = 3, nstart = 25)
# Add cluster results to the dataset
euro_24$Cluster <- as.factor(kmeans_result$cluster)
# Visualize clusters
k <- ggplot(euro_24, aes(x = Age, y = MarketValue, color = Cluster)) +
  geom_point(size = 3) +
  theme_minimal() +
  labs(title = "Clustering of Players", x = "Age", y = "Market Value (€)")

  ggplotly(k)
```


Clustering of Players
Market Value (€)
1.5e+08
1.0e+08
5.0e+07
0.0e+00
15
20
25
30
35
40
Age
Cluster
1
2
3

Total number of goals scored by players of each country

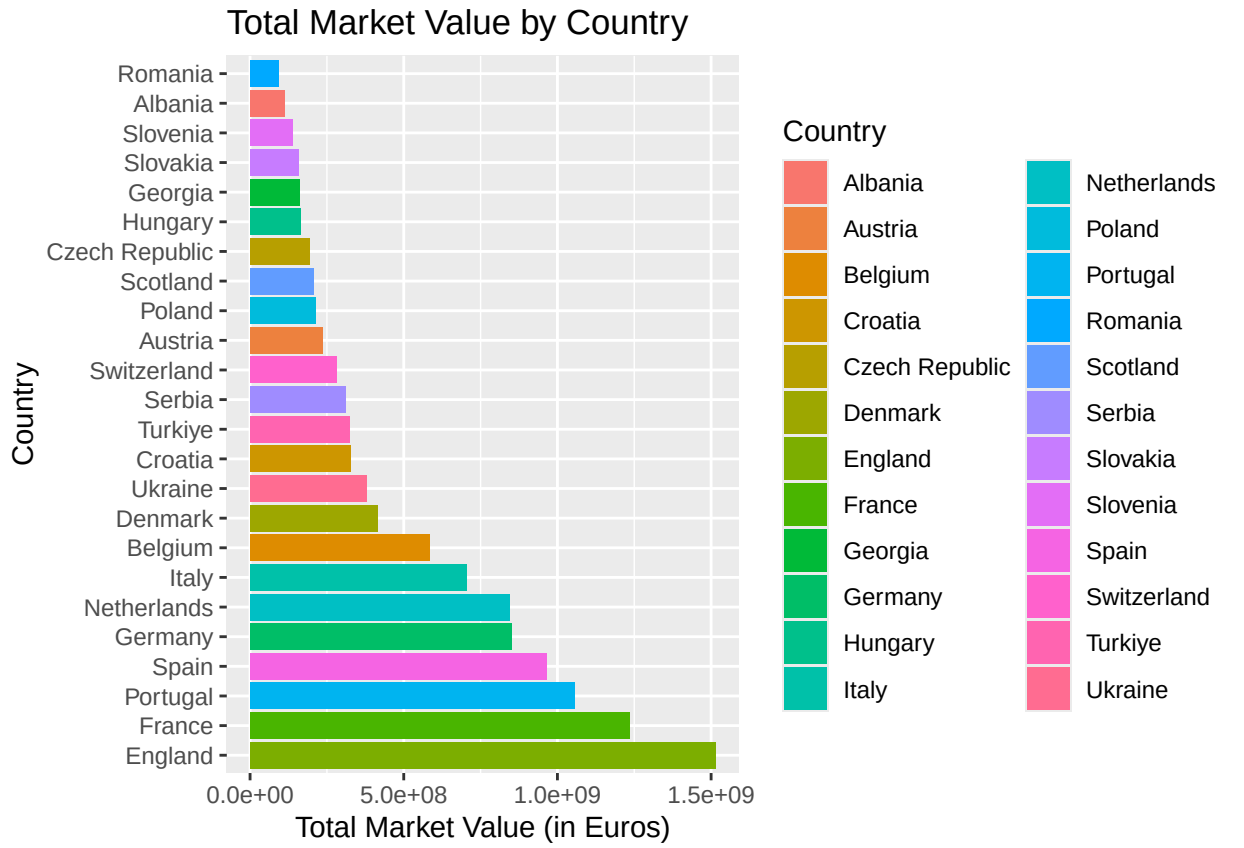
```
goals_by_country <- euro_24 %>%
  group_by(Country) %>%
  summarise(Total_Goals = sum(Goals))
ggplot(goals_by_country, aes(x = reorder(Country, -Total_Goals), y = Total_Goals, fill = Country)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Total Number of Goals by Country",
       x = "Country",
       y = "Total Goals")
```



Total market value of players of each country

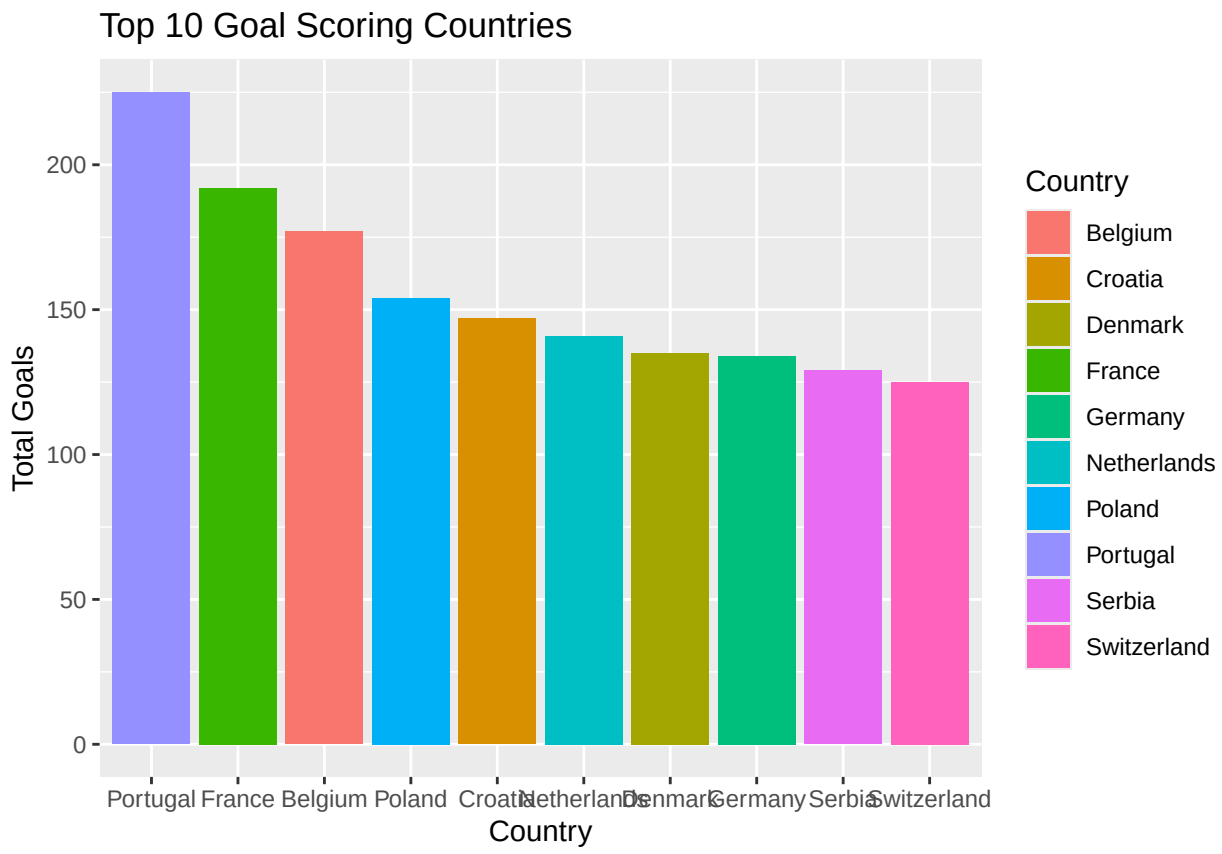
```
market_value_by_country <- euro_24 %>%
  group_by(Country) %>%
  summarise(Total_MarketValue = sum(MarketValue))
ggplot(market_value_by_country, aes(x = reorder(Country, -Total_MarketValue), y = Total_MarketValue, fi
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Total Market Value by Country",
```

```
      x = "Country",
      y = "Total Market Value (in Euros)")
```



Top Goal Scoring Countries

```
top_goal_scoring_countries <- goals_by_country %>%
  top_n(10, Total_Goals)
ggplot(top_goal_scoring_countries, aes(x = reorder(Country, -Total_Goals), y = Total_Goals, fill = Coun
  geom_bar(stat = "identity") +
  labs(title = "Top 10 Goal Scoring Countries",
       x = "Country",
       y = "Total Goals")
```


Top 10 Goal Scoring Countries
Total Goals
200
150
100
50
0
Portugal
France
Belgium
Poland
Croatia
Netherlands
Denmark
Germany
Serbia
Switzerland
Country
Country
Belgium
Croatia
Denmark
France
Germany
Netherlands
Poland
Portugal
Serbia
Switzerland