Final Soccer

STAT 184 Course Project

Evin karimbaniyil, William Carrera

06 30 2025

Table of contents

Introduction

Our project is focused on the best soccer players this world has seen. Every year, football fans around the world are always asking: Who is the best soccer player? Why are they the best soccer player? What makes them the best soccer player? Through the years, these questions have had different answers. It has gotten to the point where the answers in the past would be unacceptable answers to football fanatics and analysts. Our goal with this project is to demonstrate who truly deserved to win the most coveted trophy in soccer history: the Ballon d'Or. The Ballon d'Or is awarded yearly to the single best player throughout a season. We, and many other fans, think that the trophy has lost some of its significance in recent years, as an ever-expanding list of criteria may eclipse true on-field performance.

Background information

Football's top individual honor, the Ballon d'Or, was first presented by France Football magazine in 1956. Its eligibility was initially restricted to European players, but in 1995 it was broadened to allow any player at a European club. Since 2007, nominations from all around the world have been accepted. Coaches, captains, and a few journalists from the national team each submit ranked ballots, with votes being cast based on individual performance, team performance, and overall match influence.

Although a player's season-long consistency is reflected in the Ballon d'Or, other awards highlight various aspects of exceptional performance:

- European Golden Boot:

- UEFA Champions League: Europe's greatest club tournament frequently recognizes its finest player through media and peer honors, highlighting clutch efforts on the largest stage.
- FIFA World Cup: As the pinnacle of international football, the "Best Player" award honors individuals who thrive under the most intense worldwide pressure.

By comparing Ballon d'Or winners to Golden Shoe winners and exceptional Champions League and World Cup performers, we want to determine if the Ballon d'Or actually belongs to the season's most important player or if prominent tournament heroes are being neglected.

This table shows the player who has won the most Ballon d'Or awards, highlighting their incredible achievements and dominance in the world of football over the years.

```
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
```

```
# A tibble: 6 x 4
  Player            Winner                          `Second place` `Third place`
  <chr>             <chr>                           <chr>          <chr>
1 Lionel Messi      8 (2009, 2010, 2011, 2012, 201~ "5 (2008, 201~ "1 (2007)"
2 Cristiano Ronaldo 5 (2008, 2013, 2014, 2016, 201~ "6 (2007, 200~ "1 (2019)"
3 Michel Platini    3 (1983, 1984, 1985)            ""             "2 (1977, 19~
4 Johan Cruyff      3 (1971, 1973, 1974)            ""             "1 (1975)"
5 Marco van Basten  3 (1988, 1989, 1992)            ""             ""
6 Franz Beckenbauer 2 (1972, 1976)                  "2 (1974, 197~ "1 (1966)"
```

This table displays the player with the most Golden Boot awards, showcasing their consistent goal-scoring excellence and remarkable impact in top-level competitions.

```
# A tibble: 6 x 3
  Player             Wins Seasons
  <chr>             <int> <chr>
1 Lionel Messi          6 2009-10, 2011-12, 2012-13, 2016-17, 2017-18, 2018-
19
2 Cristiano Ronaldo     4 2007-08, 2010-11, 2013-14 (shared), 2014-15
3 Eusébio               2 1967-68, 1972-73
4 Gerd Müller           2 1969-70, 1971-72
5 Dudu Georgescu        2 1974-75, 1976-77
6 Fernando Gomes        2 1982-83, 1984-85
```

This graphs provides information about the winners of the Ballon d'Or, Champions League, and World Cup. It includes the best player of each competition, as well as the teams that won the Champions League and the World Cup. This summary allows us to see the top performers and most successful teams across the biggest football tournaments in the world.

```
Attaching package: 'readr'


The following object is masked from 'package:rvest':

    guess_encoding


Rows: 10 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): WinnerBD, Nationality, Team, pts
dbl (1): Year

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.
```
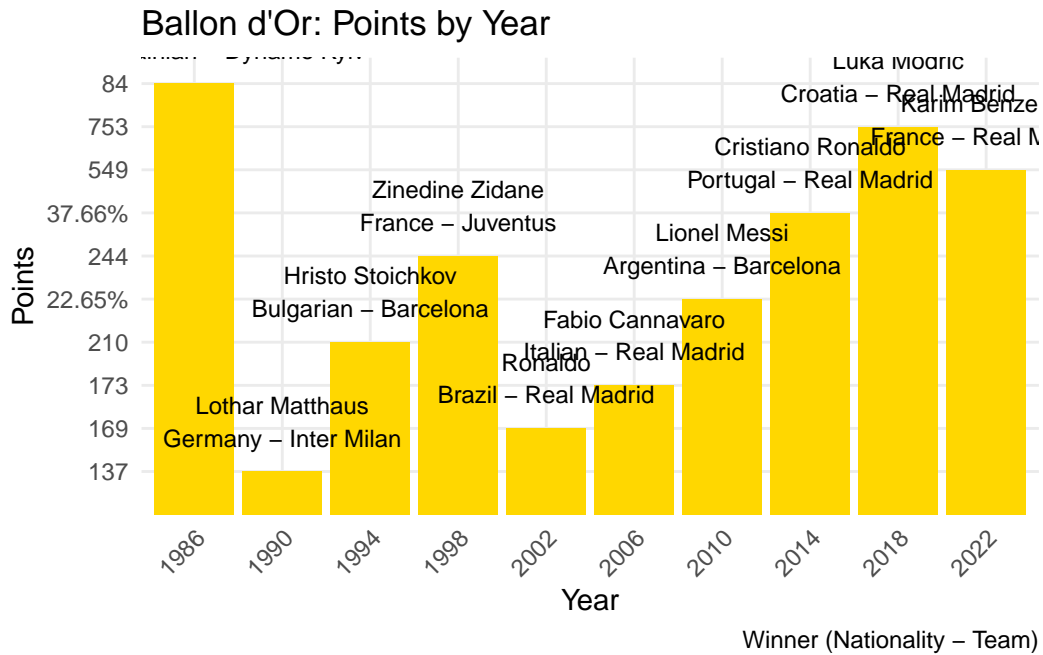
Ballon d'Or: Points by Year

Milan – Dynamo Kyiv

Luka Modric
Croatia – Real Madrid
Karim Benze
Cristiano Ronaldo France – Real M
Portugal – Real Madrid

Zinedine Zidane
France – Juventus

Lionel Messi
Argentina – Barcelona

Hristo Stoichkov
Bulgarian – Barcelona

Fabio Cannavaro
Ronaldo Italian – Real Madrid

Brazil – Real Madrid

Lothar Matthaus
Germany – Inter Milan

Points (y-axis): 84, 753, 549, 37.66%, 244, 22.65%, 210, 173, 169, 137

Year (x-axis): 1986, 1990, 1994, 1998, 2002, 2006, 2010, 2014, 2018, 2022

Winner (Nationality – Team)

```
Rows: 10 Columns: 5
-- Column specification -------------------------------------------------------
Delimiter: ","
chr (4): WinnerCL, SecondCL, BestplayerCL, Nationality
dbl (1): Year

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.
```
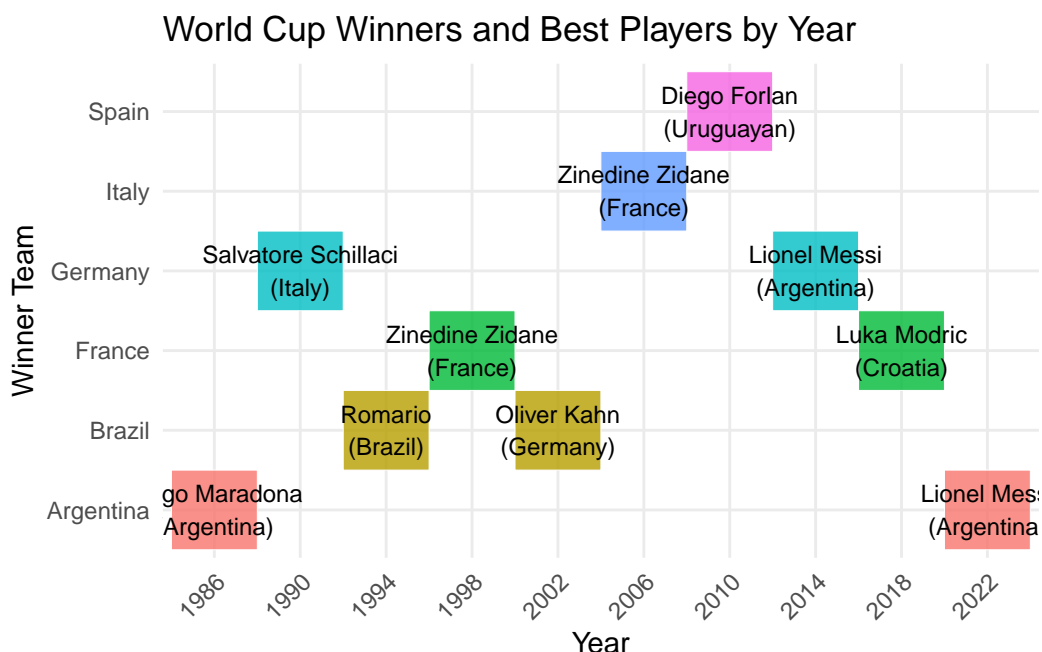
Champions League Points Over Years


```
Rows: 10 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): WinnerWC, SecondWC, BestplayerWC, Nationality
dbl (1): Year

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.
```

World Cup Winners and Best Players by Year

Conclusion

Our cross-competition study demonstrates that the Ballon d'Or usually goes to the team that wins the most league games and the most Champions League games. Icons such as Lionel Messi and Cristiano Ronaldo, whose continuous goal-scoring and match-winning performances helped them win both club and individual awards, demonstrate this confluence. However, there have been other instances where players made tournament-defining World Cup or Champions League performances while falling short in the Ballon d'Or voting, indicating that the award's criteria may undervalue peak effect in favor of season-long measurements.

To ensure the Ballon d'Or retains its prestige, we recommend adopting a more balanced voting system that explicitly factors in major tournament achievements alongside domestic and continental league performances. This approach would honor both the season-long consistency of top scorers and the defining moments delivered on football's grandest stages, reaffirming the trophy's status as the unquestioned marker of the world's best player each year. Furthermore, expanding the voting pool beyond just coaches, players, and journalists to include fans from around the globe would democratize the process and reflect soccer's truly universal appeal. While any voting system will inevitably carry biases and personal preferences, a broader electorate helps to balance those influences and strengthens the award's legitimacy. By embracing these reforms, the Ballon d'Or can evolve into an even more inclusive and representative celebration of footballing excellence.

```
Rows: 10 Columns: 5
-- Column specification ---------------------------------------------------------
```

```
Delimiter: ","
chr (4): WinnerBD, Nationality, Team, pts
dbl (1): Year

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.
Rows: 10 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): WinnerCL, SecondCL, BestplayerCL, Nationality
dbl (1): Year

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.
Rows: 10 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): WinnerWC, SecondWC, BestplayerWC, Nationality
dbl (1): Year

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.
```

Code Appendix

```r
library(dplyr)
library(rvest)

Dor <- "https://en.wikipedia.org/wiki/Ballon_d%27Or"
page <- read_html(Dor)
tables <- page %>%
  html_nodes("table")
tables[[5]] %>%
  html_table(fill = TRUE)%>%
  head()
```

```
# A tibble: 6 x 4
  Player           Winner                          `Second place` `Third place`
  <chr>            <chr>                           <chr>          <chr>
1 Lionel Messi     8 (2009, 2010, 2011, 2012, 201~ "5 (2008, 201~ "1 (2007)"
2 Cristiano Ronaldo 5 (2008, 2013, 2014, 2016, 201~ "6 (2007, 200~ "1 (2019)"
3 Michel Platini   3 (1983, 1984, 1985)            ""             "2 (1977, 19~
```

```
4 Johan Cruyff      3 (1971, 1973, 1974)           ""              "1 (1975)"
5 Marco van Basten  3 (1988, 1989, 1992)           ""              ""
6 Franz Beckenbauer 2 (1972, 1976)                 "2 (1974, 197~  "1 (1966)"
```

```r
library(dplyr)
library(rvest)

Gool <- "https://en.wikipedia.org/wiki/European_Golden_Shoe"
page <- read_html(Gool)
tables <- page %>%
  html_nodes("table")
tables[[4]] %>%
  html_table(fill = TRUE)%>%
  head()
```

```
# A tibble: 6 x 3
  Player             Wins Seasons
  <chr>             <int> <chr>
1 Lionel Messi          6 2009-10, 2011-12, 2012-13, 2016-17, 2017-18, 2018-
19
2 Cristiano Ronaldo     4 2007-08, 2010-11, 2013-14 (shared), 2014-15
3 Eusébio               2 1967-68, 1972-73
4 Gerd Müller           2 1969-70, 1971-72
5 Dudu Georgescu        2 1974-75, 1976-77
6 Fernando Gomes        2 1982-83, 1984-85
```

```r
library(readr)
library(ggplot2)
library(dplyr)

Ballon_d_Or_Sheet1_ <- read_csv("Ballon dOr(Sheet1).csv")
```
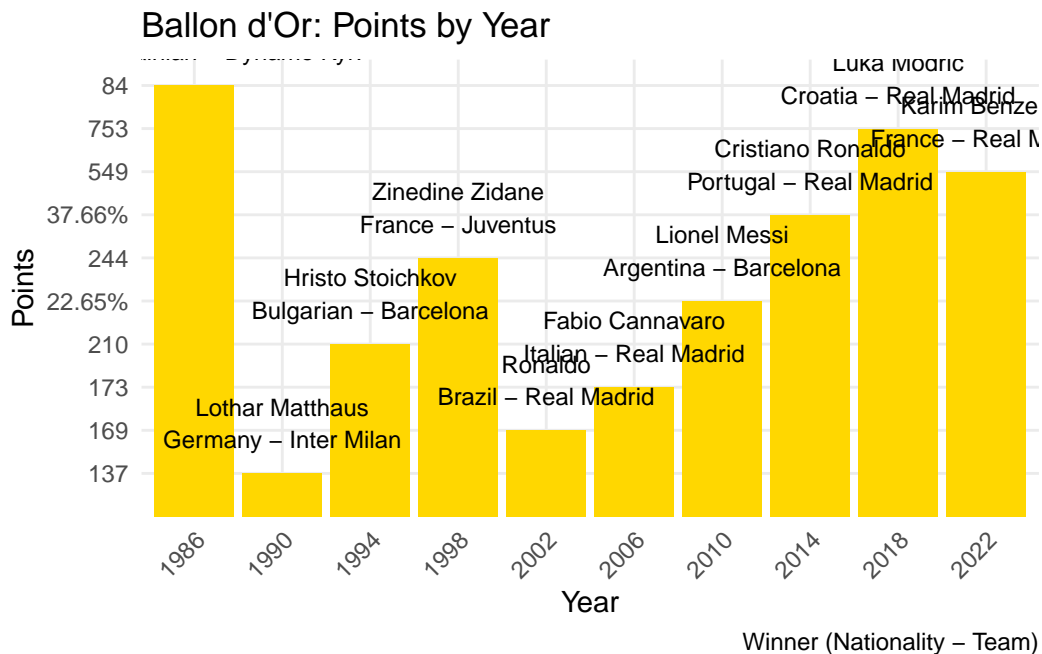
```
Rows: 10 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): WinnerBD, Nationality, Team, pts
dbl (1): Year

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.
```

```
ggplot(Ballon_d_Or_Sheet1_, aes(x = as.factor(Year), y = pts)) +
  geom_col(fill = "gold") +
  geom_text(aes(label = paste0(WinnerBD, "\n", Nationality, " - ", Team)),
            vjust = -0.5, size = 3) +
  labs(title = "Ballon d'Or: Points by Year",
       x = "Year",
       y = "Points",
       caption = "Winner (Nationality - Team)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Ballon d'Or: Points by Year

Winner (Nationality – Team)

```
FIFA_Champions_Sheet1_ <- read_csv("FIFA Champions(Sheet1).csv")
```

```
Rows: 10 Columns: 5
-- Column specification ------------------------------------------------------
Delimiter: ","
chr (4): WinnerCL, SecondCL, BestplayerCL, Nationality
dbl (1): Year

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.
```

```
ggplot(FIFA_Champions_Sheet1_, aes(x = Year, y = WinnerCL)) +
  geom_line(color = "blue") +
  geom_point(color = "red", size = 3) +
  geom_text(aes(label = paste0("Winner: ", WinnerCL, "\nBest: ", BestplayerCL)),
            vjust = -1, size = 3, check_overlap = TRUE) +
  labs(title = "Champions League Points Over Years",
       x = "Year",
       y = "Points") +
  theme_minimal()
```

Champions League Points Over Years


```
Worldcup_Sheet1_ <- read_csv("Worldcup(Sheet1).csv")
```

```
Rows: 10 Columns: 5
-- Column specification -----------------------------------------------------
Delimiter: ","
chr (4): WinnerWC, SecondWC, BestplayerWC, Nationality
dbl (1): Year

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.
```
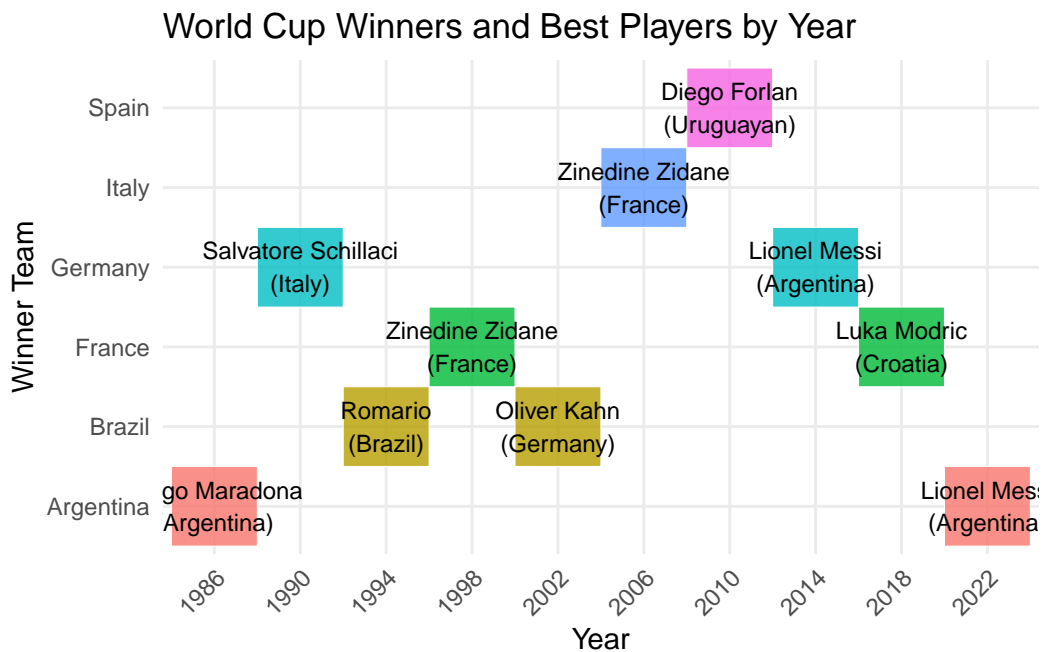
```r
ggplot(Worldcup_Sheet1_, aes(x = as.factor(Year), y = WinnerWC, fill = WinnerWC)) +
  geom_tile(color = "white", alpha = 0.8) +
  geom_text(aes(label = paste0(BestplayerWC, "\n(", Nationality, ")")),
            color = "black", size = 3) +
  labs(title = "World Cup Winners and Best Players by Year",
       x = "Year",
       y = "Winner Team") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1))
```


World Cup Winners and Best Players by Year

```r
library(dplyr)
library(readr)
library(ggplot2)


Ballon_d_Or_Sheet1_ <- read_csv("Ballon dOr(Sheet1).csv")
```

```
Rows: 10 Columns: 5
-- Column specification ------------------------------------------------------
Delimiter: ","
chr (4): WinnerBD, Nationality, Team, pts
dbl (1): Year
```

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.

```
FIFA_Champions_Sheet1_ <- read_csv("FIFA Champions(Sheet1).csv")
```

Rows: 10 Columns: 5
-- Column specification --
Delimiter: ","
chr (4): WinnerCL, SecondCL, BestplayerCL, Nationality
dbl (1): Year

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.

```
Worldcup_Sheet1_ <- read_csv("Worldcup(Sheet1).csv")
```

Rows: 10 Columns: 5
-- Column specification --
Delimiter: ","
chr (4): WinnerWC, SecondWC, BestplayerWC, Nationality
dbl (1): Year

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.

```
worldcup <- Worldcup_Sheet1_ %>%
  rename(NationallyWC = Nationality)
championsleague <- FIFA_Champions_Sheet1_ %>%
  rename(nationalityCL = Nationality)
joined_data <- worldcup %>%
  left_join(championsleague, by = "Year")%>%
  left_join(Ballon_d_Or_Sheet1_, by = "Year")
```