Mini-Project2

Tenzin Gyaltsen, Shen Rothermel

Introduction

In this project, we explored soccer statistics from multiple professional football leagues using data from FBref (https://fbref.com/en/comps/22/Major-League-Soccer-Stats), a trusted site for advanced football analytics. While we initially focused on Major League Soccer (MLS), we extended our analysis to include other major international competitions such as the Premier League, La Liga, Bundesliga, and Serie A.

Our goal was to collect and organize standardized squad-level statistics across leagues to support comparative analysis. Specifically, we targeted the "Squad Standard Stats" tables on each competition's main stats page. These tables contain information on team performance metrics such as matches played, goals, assists, average age, possession %, and more.

Motivation

We chose this dataset primarily out of personal interest: one of us enjoys following global football news, while the other is an avid FC25 player. Beyond our curiosity, we recognized that this data offers a rich opportunity for cross-league comparisons.

By scraping the same type of statistics from each league, we aimed to answer questions such as:

- Do older squads tend to score more or less?
- Is there a relationship between average age and possession percentage?
- How does team performance (e.g., goals, assists) vary across leagues?

These questions open the door for future data visualizations (like scatterplots or heatmaps) and statistical modeling (e.g., regression of goals on age or possession).

To acquire the data, we used a custom scraping function along with an iteration technique (pmap) to systematically collect comparable squad stats from each league's respective webpage. This ensures consistency while handling slight variations in webpage structure — such as differing table positions.

Scraping the “Squad Standard Stats” table:

To begin, we manually scrape the Major League Soccer (MLS) stats page using rvest. This allows us to locate and inspect the structure of all tables on the page, which helps identify the correct table containing squad-level statistics.

Once we confirm the correct table is loaded (in this case, table 5), we clean it by promoting the first row to column headers, standardizing names, and parsing numeric columns. This results in a tidy dataset ready for analysis.

```r
#| results: hide

library(rvest)
library(janitor)
```

```
Warning: package 'janitor' was built under R version 4.4.3


Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(dplyr)
```

```

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
```

```r
library(purrr)
library(stringr)
library(readr)
```

```
Attaching package: 'readr'


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

    guess_encoding
```

```r
# Check permissions for the specific stats page
robotstxt::paths_allowed("https://fbref.com/en/comps/22/Major-League-Soccer-Stats")
```

```
fbref.com




[1] TRUE
```

```r
# Step 1: Read the page with rvest
MLS_table <- read_html("https://fbref.com/en/comps/22/Major-League-Soccer-Stats")

# Step 2: Extract tables from the page
Squad <- html_nodes(MLS_table, "table")
html_table(Squad, header = TRUE, fill = TRUE)  # find right table
```

```
[[1]]
# A tibble: 15 x 20
      Rk Squad       MP     W     D     L    GF    GA    GD   Pts `Pts/MP`    xG
   <int> <chr>    <int> <int> <int> <int> <int> <int> <int> <int>    <dbl> <dbl>
 1     1 Columbu~     7     4     3     0    10     5     5    15     2.14   8.7
 2     2 Inter M~     6     4     2     0    12     6     6    14     2.33  11.3
 3     3 Philade~     7     4     1     2    13     8     5    13     1.86  13.2
 4     4 Charlot~     7     4     1     2    12     7     5    13     1.86   9.3
 5     5 FC Cinc~     7     4     1     2     9     9     0    13     1.86  10  
 6     6 Orlando~     7     3     2     2    15    12     3    11     1.57  12.4
 7     7 Chicago~     7     3     2     2    14    12     2    11     1.57  11.7
 8     8 NY Red ~     7     3     2     2     9     7     2    11     1.57  11  
 9     9 Nashvil~     7     3     1     3    10     7     3    10     1.43  11.2
10    10 Atlanta~     7     2     3     2    11    12    -1     9     1.29  10  
11    11 NYCFC        7     2     2     3    10    11    -1     8     1.14   9.9
12    12 D.C. Un~     7     1     3     3     9    17    -8     6     0.86  11.3
13    13 NE Revo~     6     1     1     4     3     7    -4     4     0.67   4.3
```

```
14    14 Toronto~      7     0     3     4     7    13    -6     3     0.43   5.7
15    15 CF Mont~      7     0     2     5     4    12    -8     2     0.29   7.1
# i 8 more variables: xGA <dbl>, xGD <dbl>, `xGD/90` <dbl>, `Last 5` <chr>,
#   Attendance <chr>, `Top Team Scorer` <chr>, Goalkeeper <chr>, Notes <lgl>

[[2]]
# A tibble: 16 x 28
   ``    ``    Home  Home  Home  Home  Home  Home  Home  Home  Home  Home  Home
   <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
 1 Rk    Squad MP    W     D     L     GF    GA    GD    Pts   "Pts~ xG    xGA
 2 1     Colu~ 4     2     2     0     6     3     +3    8     "2.0~ 6.1   2.8
 3 2     Inte~ 4     2     2     0     6     4     +2    8     "2.0~ 7.8   5.0
 4 3     Phil~ 4     2     1     1     6     4     +2    7     "1.7~ 8.0   3.5
 5 4     Char~ 4     4     0     0     10    2     +8    12    "3.0~ 7.5   5.6
 6 5     FC C~ 4     3     1     0     6     2     +4    10    "2.5~ 6.4   3.8
 7 6     Orla~ 3     2     0     1     10    7     +3    6     "2.0~ 7.0   3.5
 8 7     Chic~ 2     0     2     0     3     3     0     2     "1.0~ 3.1   4.5
 9 8     NY R~ 4     3     1     0     8     4     +4    10    "2.5~ 8.1   4.8
10 9     Nash~ 4     2     1     1     6     2     +4    7     "1.7~ 7.3   3.2
11 10    Atla~ 5     2     2     1     9     8     +1    8     "1.6~ 7.3   7.3
12 11    NYCFC 3     2     0     1     5     4     +1    6     "2.0~ 6.3   3.9
13 12    D.C.~ 4     1     2     1     5     5     0     5     "1.2~ 5.3   4.4
14 13    NE R~ 3     1     0     2     2     4     -2    3     "1.0~ 2.8   3.1
15 14    Toro~ 2     0     1     1     1     2     -1    1     "0.5~ 1.3   2.4
16 15    CF M~ 0     0     0     0     0     0     0     0     ""    0.0   0.0
# i 15 more variables: Home <chr>, Home <chr>, Away <chr>, Away <chr>,
#   Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>,
#   Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>

[[3]]
# A tibble: 15 x 20
      Rk Squad        MP     W     D     L    GF    GA    GD   Pts `Pts/MP`    xG
   <int> <chr>     <int> <int> <int> <int> <int> <int> <int> <int>    <dbl> <dbl>
 1     1 Vancouv~      7     5     1     1    12     5     7    16     2.29  11.7
 2     2 San Die~      7     4     2     1    13     6     7    14     2     12.7
 3     3 Minneso~      7     4     2     1    11     7     4    14     2     13.8
 4     4 Austin        7     4     1     2     5     3     2    13     1.86   8.2
 5     5 Portlan~      7     3     2     2     9     8     1    11     1.57   7.4
 6     6 FC Dall~      7     3     2     2    10    10     0    11     1.57  10.1
 7     7 Colorad~      7     3     2     2     8     9    -1    11     1.57   7.8
 8     8 SJ Eart~      7     3     1     3    15    10     5    10     1.43  14.6
 9     9 LAFC          7     3     0     4     8    10    -2     9     1.29   7.6
10    10 Real Sa~      7     3     0     4     7    11    -4     9     1.29   9.7
```

```
11    11 St. Lou~     7     2     2     3     4     4     0     8     1.14   7.2
12    12 Seattle~     7     1     3     3     8    11    -3     6     0.86  10
13    13 Houston~     7     1     2     4     5    11    -6     5     0.71   5.6
14    14 Sportin~     7     1     1     5     8    12    -4     4     0.57   6.4
15    15 LA Gala~     7     0     2     5     5    14    -9     2     0.29   6
# i 8 more variables: xGA <dbl>, xGD <dbl>, `xGD/90` <dbl>, `Last 5` <chr>,
#   Attendance <chr>, `Top Team Scorer` <chr>, Goalkeeper <chr>, Notes <lgl>

[[4]]
# A tibble: 16 x 28
   ``    ``    Home  Home  Home  Home  Home  Home  Home  Home  Home  Home  Home
   <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
 1 Rk    Squad MP    W     D     L     GF    GA    GD    Pts   Pts/~ xG    xGA
 2 1     Vanc~ 4     3     0     1     7     4     +3    9     2.25  6.6   4.4
 3 2     San ~ 4     2     2     0     7     3     +4    8     2.00  7.0   3.8
 4 3     Minn~ 3     2     1     0     5     2     +3    7     2.33  6.9   2.4
 5 4     Aust~ 4     2     1     1     3     2     +1    7     1.75  5.5   2.8
 6 5     Port~ 4     2     1     1     6     6     0     7     1.75  4.5   5.0
 7 6     FC D~ 3     1     0     2     3     5     -2    3     1.00  3.2   3.6
 8 7     Colo~ 3     1     1     1     5     6     -1    4     1.33  4.9   4.5
 9 8     SJ E~ 5     2     1     2     12    5     +7    7     1.40  10.3  10.1
10 9     LAFC  3     2     0     1     2     1     +1    6     2.00  2.5   2.2
11 10    Real~ 4     2     0     2     5     4     +1    6     1.50  5.5   5.9
12 11    St. ~ 3     1     1     1     1     1     0     4     1.33  4.1   1.7
13 12    Seat~ 3     1     2     0     7     4     +3    5     1.67  4.4   2.0
14 13    Hous~ 4     1     0     3     4     8     -4    3     0.75  4.3   5.4
15 14    Spor~ 4     1     1     2     6     7     -1    4     1.00  4.4   6.4
16 15    LA G~ 3     0     0     3     1     7     -6    0     0.00  3.7   5.7
# i 15 more variables: Home <chr>, Home <chr>, Away <chr>, Away <chr>,
#   Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>,
#   Away <chr>, Away <chr>, Away <chr>, Away <chr>, Away <chr>

[[5]]
# A tibble: 31 x 32
   ``             ``    ``    ``    `Playing Time` `Playing Time` `Playing Time`
   <chr>          <chr> <chr> <chr> <chr>          <chr>          <chr>
 1 Squad          # Pl  Age   Poss  MP             Starts         Min
 2 Atlanta Utd    23    29.2  49.1  7              77             630
 3 Austin         19    28.2  42.4  7              77             630
 4 CF Montréal    23    24.2  51.9  7              77             630
 5 Charlotte      18    29.3  50.1  7              77             630
 6 Chicago Fire   22    25.9  47.7  7              77             630
 7 Colorado Rapi~ 24    26.5  45.7  7              77             630
 8 Columbus Crew  20    26.6  56.6  7              77             630
 9 D.C. United    20    26.1  52.4  7              77             630
10 FC Cincinnati  22    27.5  52.3  7              77             630
# i 21 more rows
# i 25 more variables: `Playing Time` <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Expected <chr>,
#   Expected <chr>, Expected <chr>, Expected <chr>, Progression <chr>,
#   Progression <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>,
#   `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, ...

[[6]]
# A tibble: 31 x 32
   ``             ``    ``    ``    `Playing Time` `Playing Time` `Playing Time`
   <chr>          <chr> <chr> <chr> <chr>          <chr>          <chr>
 1 Squad          # Pl  Age   Poss  MP             Starts         Min
 2 vs Atlanta Utd 23    27.3  50.9  7              77             630
 3 vs Austin      19    26.7  57.6  7              77             630
 4 vs CF Montréal 23    27.1  48.1  7              77             630
 5 vs Charlotte   18    28.1  49.9  7              77             630
 6 vs Chicago Fi~ 22    26.3  52.3  7              77             630
 7 vs Colorado R~ 24    28.0  54.3  7              77             630
 8 vs Columbus C~ 20    26.1  43.4  7              77             630
 9 vs D.C. United 20    26.9  47.6  7              77             630
10 vs FC Cincinn~ 22    27.6  47.7  7              77             630
# i 21 more rows
# i 25 more variables: `Playing Time` <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Expected <chr>,
#   Expected <chr>, Expected <chr>, Expected <chr>, Progression <chr>,
#   Progression <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>,
#   `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, ...

[[7]]
# A tibble: 31 x 21
   ``          ``    `Playing Time` `Playing Time` `Playing Time` `Playing Time`
   <chr>       <chr> <chr>          <chr>          <chr>          <chr>
 1 Squad       # Pl  MP             Starts         Min            90s
 2 Atlanta Utd 1     7              7              630            7.0
 3 Austin      1     7              7              630            7.0
 4 CF Montréal 1     7              7              630            7.0
 5 Charlotte   1     7              7              630            7.0
 6 Chicago Fi~ 1     7              7              630            7.0
 7 Colorado R~ 2     7              7              630            7.0
 8 Columbus C~ 2     7              7              630            7.0
 9 D.C. United 1     7              7              630            7.0
10 FC Cincinn~ 1     7              7              630            7.0
# i 21 more rows
# i 15 more variables: Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>,
#   `Penalty Kicks` <chr>, `Penalty Kicks` <chr>, `Penalty Kicks` <chr>,
#   `Penalty Kicks` <chr>, `Penalty Kicks` <chr>

[[8]]
# A tibble: 31 x 21
   ``          ``    `Playing Time` `Playing Time` `Playing Time` `Playing Time`
   <chr>       <chr> <chr>          <chr>          <chr>          <chr>
 1 Squad       # Pl  MP             Starts         Min            90s
 2 vs Atlanta~ 1     7              7              630            7.0
 3 vs Austin   1     7              7              630            7.0
 4 vs CF Mont~ 1     7              7              630            7.0
 5 vs Charlot~ 1     7              7              627            7.0
 6 vs Chicago~ 1     7              7              630            7.0
 7 vs Colorad~ 2     7              7              630            7.0
 8 vs Columbu~ 2     7              7              630            7.0
 9 vs D.C. Un~ 1     7              7              630            7.0
10 vs FC Cinc~ 1     7              7              630            7.0
# i 21 more rows
# i 15 more variables: Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>,
#   `Penalty Kicks` <chr>, `Penalty Kicks` <chr>, `Penalty Kicks` <chr>,
#   `Penalty Kicks` <chr>, `Penalty Kicks` <chr>

[[9]]
# A tibble: 31 x 28
   ``       ``    ``    Goals Goals Goals Goals Goals Expected Expected Expected
   <chr>    <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>    <chr>    <chr>
 1 Squad    # Pl  90s   GA    PKA   FK    CK    OG    PSxG     PSxG/SoT PSxG+/-
 2 Atlanta~ 1     7.0   12    1     1     2     0     10.6     0.29     -1.4
 3 Austin   1     7.0   3     0     1     0     0     3.9      0.19     +0.9
 4 CF Mont~ 1     7.0   12    0     0     3     0     13.2     0.33     +1.2
 5 Charlot~ 1     7.0   7     2     0     0     0     9.5      0.21     +2.5
 6 Chicago~ 1     7.0   12    1     0     1     1     12.4     0.33     +1.4
 7 Colorad~ 2     7.0   9     0     0     1     1     11.6     0.33     +3.6
 8 Columbu~ 2     7.0   5     0     0     1     0     6.5      0.35     +1.5
 9 D.C. Un~ 1     7.0   17    2     0     3     0     15.6     0.38     -1.4
10 FC Cinc~ 1     7.0   9     0     0     0     1     9.0      0.29     +1.0
# i 21 more rows
# i 17 more variables: Expected <chr>, Launched <chr>, Launched <chr>,
#   Launched <chr>, Passes <chr>, Passes <chr>, Passes <chr>, Passes <chr>,
#   `Goal Kicks` <chr>, `Goal Kicks` <chr>, `Goal Kicks` <chr>, Crosses <chr>,
#   Crosses <chr>, Crosses <chr>, Sweeper <chr>, Sweeper <chr>, Sweeper <chr>

[[10]]
# A tibble: 31 x 28
   ``       ``    ``    Goals Goals Goals Goals Goals Expected Expected Expected
   <chr>    <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>    <chr>    <chr>
 1 Squad    # Pl  90s   GA    PKA   FK    CK    OG    PSxG     PSxG/SoT PSxG+/-
 2 vs Atla~ 1     7.0   11    1     0     1     2     11.2     0.32     +2.2
 3 vs Aust~ 1     7.0   5     0     0     2     0     6.7      0.32     +1.7
 4 vs CF M~ 1     7.0   4     0     0     1     0     6.9      0.37     +2.9
 5 vs Char~ 1     7.0   12    1     0     1     1     10.4     0.40     -0.6
 6 vs Chic~ 1     7.0   14    1     0     2     0     15.0     0.42     +1.0
 7 vs Colo~ 2     7.0   8     1     0     0     0     5.9      0.27     -2.1
 8 vs Colu~ 2     7.0   10    0     0     0     1     9.5      0.30     +0.5
 9 vs D.C.~ 1     7.0   9     1     0     1     0     12.6     0.33     +3.6
10 vs FC C~ 1     7.0   9     2     2     0     0     9.2      0.24     +0.2
# i 21 more rows
# i 17 more variables: Expected <chr>, Launched <chr>, Launched <chr>,
#   Launched <chr>, Passes <chr>, Passes <chr>, Passes <chr>, Passes <chr>,
#   `Goal Kicks` <chr>, `Goal Kicks` <chr>, `Goal Kicks` <chr>, Crosses <chr>,
#   Crosses <chr>, Crosses <chr>, Sweeper <chr>, Sweeper <chr>, Sweeper <chr>

[[11]]
# A tibble: 31 x 20
   ``          ``    ``    Standard Standard Standard Standard Standard Standard
   <chr>       <chr> <chr> <chr>    <chr>    <chr>    <chr>    <chr>    <chr>
 1 Squad       # Pl  90s   Gls      Sh       SoT      SoT%     Sh/90    SoT/90
 2 Atlanta Utd 23    7.0   9        87       32       36.8     12.43    4.57
 3 Austin      19    7.0   5        85       21       24.7     12.14    3.00
 4 CF Montréal 23    7.0   4        65       20       30.8     9.29     2.86
 5 Charlotte   18    7.0   11       69       24       34.8     9.86     3.43
 6 Chicago Fi~ 22    7.0   14       75       33       44.0     10.71    4.71
 7 Colorado R~ 24    7.0   8        68       18       26.5     9.71     2.57
 8 Columbus C~ 20    7.0   9        84       32       38.1     12.00    4.57
 9 D.C. United 20    7.0   9        87       36       41.4     12.43    5.14
10 FC Cincinn~ 22    7.0   9        94       32       34.0     13.43    4.57
```

```
# i 21 more rows
# i 11 more variables: Standard <chr>, Standard <chr>, Standard <chr>,
#   Standard <chr>, Standard <chr>, Standard <chr>, Expected <chr>,
#   Expected <chr>, Expected <chr>, Expected <chr>, Expected <chr>

[[12]]
# A tibble: 31 x 20
   ``          ``    ``    Standard Standard Standard Standard Standard Standard
   <chr>       <chr> <chr> <chr>    <chr>    <chr>    <chr>    <chr>    <chr>
 1 Squad       # Pl  90s   Gls      Sh       SoT      SoT%     Sh/90    SoT/90
 2 vs Atlanta~ 23    7.0   12       73       33       45.2     10.43    4.71
 3 vs Austin   19    7.0   3        79       21       26.6     11.29    3.00
 4 vs CF Mont~ 23    7.0   12       90       41       45.6     12.86    5.86
 5 vs Charlot~ 18    7.0   7        112      36       32.1     16.00    5.14
 6 vs Chicago~ 22    7.0   11       85       30       35.3     12.14    4.29
 7 vs Colorad~ 24    7.0   8        113      38       33.6     16.14    5.43
 8 vs Columbu~ 20    7.0   5        54       19       35.2     7.71     2.71
 9 vs D.C. Un~ 20    7.0   17       87       37       42.5     12.43    5.29
10 vs FC Cinc~ 22    7.0   8        90       27       30.0     12.86    3.86
# i 21 more rows
# i 11 more variables: Standard <chr>, Standard <chr>, Standard <chr>,
#   Standard <chr>, Standard <chr>, Standard <chr>, Expected <chr>,
#   Expected <chr>, Expected <chr>, Expected <chr>, Expected <chr>

[[13]]
# A tibble: 31 x 26
   ``         ``    ``    Total Total Total Total Total Short Short Short Medium
   <chr>      <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
 1 Squad      # Pl  90s   Cmp   Att   Cmp%  TotD~ PrgD~ Cmp   Att   Cmp%  Cmp
 2 Atlanta U~ 23    7.0   2958  3627  81.6  53208 19100 1257  1441  87.2  1359
 3 Austin     19    7.0   2390  2969  80.5  43653 16477 1069  1190  89.8  972
 4 CF Montré~ 23    7.0   2848  3558  80.0  51391 18004 1204  1370  87.9  1268
 5 Charlotte  18    7.0   2840  3457  82.2  50448 17196 1235  1354  91.2  1309
 6 Chicago F~ 22    7.0   2826  3426  82.5  46504 17309 1383  1531  90.3  1110
 7 Colorado ~ 24    7.0   2309  3063  75.4  39847 15706 1088  1246  87.3  947
 8 Columbus ~ 20    7.0   3650  4246  86.0  56171 19991 1986  2142  92.7  1332
 9 D.C. Unit~ 20    7.0   2718  3488  77.9  50305 19605 1210  1401  86.4  1135
10 FC Cincin~ 22    7.0   3109  3814  81.5  53759 18705 1318  1510  87.3  1473
# i 21 more rows
# i 14 more variables: Medium <chr>, Medium <chr>, Long <chr>, Long <chr>,
#   Long <chr>, `` <chr>, `` <chr>, Expected <chr>, Expected <chr>, `` <chr>,
#   `` <chr>, `` <chr>, `` <chr>, `` <chr>
```

```
[[14]]
# A tibble: 31 x 26
   ``         ``    ``    Total Total Total Total Total Short Short Short Medium
   <chr>      <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
 1 Squad      # Pl  90s   Cmp   Att   Cmp%  TotD~ PrgD~ Cmp   Att   Cmp%  Cmp
 2 vs Atlant~ 23    7.0   3153  3788  83.2  54639 18143 1406  1552  90.6  1425
 3 vs Austin  19    7.0   3403  4068  83.7  61991 21857 1419  1595  89.0  1620
 4 vs CF Mon~ 23    7.0   2628  3295  79.8  46749 18752 1176  1342  87.6  1129
 5 vs Charlo~ 18    7.0   2888  3509  82.3  49946 18592 1332  1482  89.9  1230
 6 vs Chicag~ 22    7.0   3063  3745  81.8  53027 17421 1366  1522  89.8  1340
 7 vs Colora~ 24    7.0   2920  3618  80.7  52084 19208 1295  1435  90.2  1261
 8 vs Columb~ 20    7.0   2616  3254  80.4  44517 15942 1251  1418  88.2  1008
 9 vs D.C. U~ 20    7.0   2512  3156  79.6  43818 17259 1169  1302  89.8  1026
10 vs FC Cin~ 22    7.0   2774  3489  79.5  49237 18218 1218  1373  88.7  1239
# i 21 more rows
# i 14 more variables: Medium <chr>, Medium <chr>, Long <chr>, Long <chr>,
#   Long <chr>, `` <chr>, `` <chr>, Expected <chr>, Expected <chr>, `` <chr>,
#   `` <chr>, `` <chr>, `` <chr>, `` <chr>

[[15]]
# A tibble: 31 x 18
   ``       ``    ``    ``    `Pass Types` `Pass Types` `Pass Types` `Pass Types`
   <chr>    <chr> <chr> <chr> <chr>        <chr>        <chr>        <chr>
 1 Squad    # Pl  90s   Att   Live         Dead         FK           TB
 2 Atlant~ 23    7.0   3627  3299         316          93           3
 3 Austin   19    7.0   2969  2676         277          81           8
 4 CF Mon~ 23    7.0   3558  3230         310          110          5
 5 Charlo~ 18    7.0   3457  3122         326          87           5
 6 Chicag~ 22    7.0   3426  3090         318          89           7
 7 Colora~ 24    7.0   3063  2738         313          70           11
 8 Columb~ 20    7.0   4246  3937         299          118          11
 9 D.C. U~ 20    7.0   3488  3127         347          80           6
10 FC Cin~ 22    7.0   3814  3465         329          75           9
# i 21 more rows
# i 10 more variables: `Pass Types` <chr>, `Pass Types` <chr>,
#   `Pass Types` <chr>, `Pass Types` <chr>, `Corner Kicks` <chr>,
#   `Corner Kicks` <chr>, `Corner Kicks` <chr>, Outcomes <chr>, Outcomes <chr>,
#   Outcomes <chr>

[[16]]
# A tibble: 31 x 18
   ``       ``    ``    ``    `Pass Types` `Pass Types` `Pass Types` `Pass Types`
   <chr>    <chr> <chr> <chr> <chr>        <chr>        <chr>        <chr>
 1 Squad   # Pl  90s   Att   Live         Dead         FK           TB
 2 vs Atl~ 23    7.0   3788  3494         282          98           5
 3 vs Aus~ 19    7.0   4068  3722         343          106          2
 4 vs CF ~ 23    7.0   3295  2968         320          96           9
 5 vs Cha~ 18    7.0   3509  3184         312          79           11
 6 vs Chi~ 22    7.0   3745  3384         338          110          7
 7 vs Col~ 24    7.0   3618  3275         332          87           10
 8 vs Col~ 20    7.0   3254  2960         269          75           10
 9 vs D.C~ 20    7.0   3156  2819         322          127          10
10 vs FC ~ 22    7.0   3489  3171         306          86           5
# i 21 more rows
# i 10 more variables: `Pass Types` <chr>, `Pass Types` <chr>,
#   `Pass Types` <chr>, `Pass Types` <chr>, `Corner Kicks` <chr>,
#   `Corner Kicks` <chr>, `Corner Kicks` <chr>, Outcomes <chr>, Outcomes <chr>,
#   Outcomes <chr>

[[17]]
# A tibble: 31 x 19
   ``    ``    ``    SCA   SCA   `SCA Types` `SCA Types` `SCA Types` `SCA Types`
   <chr> <chr> <chr> <chr> <chr> <chr>       <chr>       <chr>       <chr>
 1 Squad # Pl  90s   SCA   SCA90 PassLive    PassDead    TO          Sh
 2 Atla~ 23    7.0   162   23.14 123         15          10          6
 3 Aust~ 19    7.0   153   21.86 117         12          10          6
 4 CF M~ 23    7.0   114   16.29 85          9           6           8
 5 Char~ 18    7.0   120   17.14 90          7           7           6
 6 Chic~ 22    7.0   136   19.43 105         10          6           9
 7 Colo~ 24    7.0   122   17.43 94          10          4           5
 8 Colu~ 20    7.0   142   20.29 104         6           13          6
 9 D.C.~ 20    7.0   149   21.29 113         14          6           10
10 FC C~ 22    7.0   174   24.86 131         11          10          11
# i 21 more rows
# i 10 more variables: `SCA Types` <chr>, `SCA Types` <chr>, GCA <chr>,
#   GCA <chr>, `GCA Types` <chr>, `GCA Types` <chr>, `GCA Types` <chr>,
#   `GCA Types` <chr>, `GCA Types` <chr>, `GCA Types` <chr>

[[18]]
# A tibble: 31 x 19
   ``    ``    ``    SCA   SCA   `SCA Types` `SCA Types` `SCA Types` `SCA Types`
   <chr> <chr> <chr> <chr> <chr> <chr>       <chr>       <chr>       <chr>
 1 Squad # Pl  90s   SCA   SCA90 PassLive    PassDead    TO          Sh
 2 vs A~ 23    7.0   123   17.57 87          7           7           10
 3 vs A~ 19    7.0   145   20.71 109         14          5           11
 4 vs C~ 23    7.0   151   21.57 111         12          8           11
 5 vs C~ 18    7.0   211   30.14 169         12          7           10
 6 vs C~ 22    7.0   159   22.71 116         13          10          8
 7 vs C~ 24    7.0   204   29.14 155         22          4           14
 8 vs C~ 20    7.0   88    12.57 64          12          2           5
 9 vs D~ 20    7.0   159   22.71 133         7           2           7
10 vs F~ 22    7.0   159   22.71 120         12          7           11
# i 21 more rows
# i 10 more variables: `SCA Types` <chr>, `SCA Types` <chr>, GCA <chr>,
#   GCA <chr>, `GCA Types` <chr>, `GCA Types` <chr>, `GCA Types` <chr>,
#   `GCA Types` <chr>, `GCA Types` <chr>, `GCA Types` <chr>

[[19]]
# A tibble: 31 x 19
   ``             ``    ``    Tackles Tackles Tackles Tackles Tackles Challenges
   <chr>          <chr> <chr> <chr>   <chr>   <chr>   <chr>   <chr>   <chr>
 1 Squad          # Pl  90s   Tkl     TklW    Def 3rd Mid 3rd Att 3rd Tkl
 2 Atlanta Utd    23    7.0   101     62      47      47      7       40
 3 Austin         19    7.0   107     53      68      31      8       48
 4 CF Montréal    23    7.0   114     72      41      56      17      47
 5 Charlotte      18    7.0   93      64      38      42      13      47
 6 Chicago Fire   22    7.0   92      51      44      33      15      42
 7 Colorado Rapi~ 24    7.0   101     51      41      46      14      47
 8 Columbus Crew  20    7.0   87      51      33      40      14      36
 9 D.C. United    20    7.0   136     84      63      49      24      87
10 FC Cincinnati  22    7.0   124     78      59      48      17      63
# i 21 more rows
# i 10 more variables: Challenges <chr>, Challenges <chr>, Challenges <chr>,
#   Blocks <chr>, Blocks <chr>, Blocks <chr>, `` <chr>, `` <chr>, `` <chr>,
#   `` <chr>

[[20]]
# A tibble: 31 x 19
   ``             ``    ``    Tackles Tackles Tackles Tackles Tackles Challenges
   <chr>          <chr> <chr> <chr>   <chr>   <chr>   <chr>   <chr>   <chr>
 1 Squad          # Pl  90s   Tkl     TklW    Def 3rd Mid 3rd Att 3rd Tkl
 2 vs Atlanta Utd 23    7.0   92      54      30      54      8       50
 3 vs Austin      19    7.0   90      57      38      30      22      46
 4 vs CF Montréal 23    7.0   95      55      46      33      16      33
 5 vs Charlotte   18    7.0   133     85      59      52      22      70
 6 vs Chicago Fi~ 22    7.0   144     88      59      56      29      67
 7 vs Colorado R~ 24    7.0   112     66      56      40      16      39
 8 vs Columbus C~ 20    7.0   126     69      69      42      15      65
 9 vs D.C. United 20    7.0   129     74      68      40      21      52
```

```
10 vs FC Cincinn~ 22    7.0   146     82      69      55      22      67
# i 21 more rows
# i 10 more variables: Challenges <chr>, Challenges <chr>, Challenges <chr>,
#   Blocks <chr>, Blocks <chr>, Blocks <chr>, `` <chr>, `` <chr>, `` <chr>,
#   `` <chr>

[[21]]
# A tibble: 31 x 26
   ``           ``    ``    ``    Touches Touches Touches Touches Touches Touches
   <chr>        <chr> <chr> <chr> <chr>   <chr>   <chr>   <chr>   <chr>   <chr>
 1 Squad        # Pl  Poss  90s   Touches Def Pen Def 3rd Mid 3rd Att 3rd Att Pen
 2 Atlanta Utd  23    49.1  7.0   4307    408     1401    1960    973     178
 3 Austin       19    42.4  7.0   3803    507     1431    1575    832     135
 4 CF Montréal  23    51.9  7.0   4322    501     1653    1841    860     122
 5 Charlotte    18    50.1  7.0   4207    483     1485    1824    934     136
 6 Chicago Fi~  22    47.7  7.0   4241    574     1677    1788    806     126
 7 Colorado R~  24    45.7  7.0   3886    416     1158    1683    1084    160
 8 Columbus C~  20    56.6  7.0   4943    409     1507    2232    1241    167
 9 D.C. United  20    52.4  7.0   4257    385     1308    1830    1158    174
10 FC Cincinn~  22    52.3  7.0   4702    442     1534    2150    1060    167
# i 21 more rows
# i 16 more variables: Touches <chr>, `Take-Ons` <chr>, `Take-Ons` <chr>,
#   `Take-Ons` <chr>, `Take-Ons` <chr>, `Take-Ons` <chr>, Carries <chr>,
#   Carries <chr>, Carries <chr>, Carries <chr>, Carries <chr>, Carries <chr>,
#   Carries <chr>, Carries <chr>, Receiving <chr>, Receiving <chr>

[[22]]
# A tibble: 31 x 26
   ``           ``    ``    ``    Touches Touches Touches Touches Touches Touches
   <chr>        <chr> <chr> <chr> <chr>   <chr>   <chr>   <chr>   <chr>   <chr>
 1 Squad        # Pl  Poss  90s   Touches Def Pen Def 3rd Mid 3rd Att 3rd Att Pen
 2 vs Atlanta~  23    50.9  7.0   4488    445     1465    2125    938     147
 3 vs Austin    19    57.6  7.0   4772    353     1258    2207    1341    173
 4 vs CF Mont~  23    48.1  7.0   4082    447     1440    1697    971     156
 5 vs Charlot~  18    49.9  7.0   4288    415     1419    1839    1071    187
 6 vs Chicago~  22    52.3  7.0   4498    393     1427    2062    1046    183
 7 vs Colorad~  24    54.3  7.0   4513    435     1588    1985    971     197
 8 vs Columbu~  20    43.4  7.0   3978    476     1541    1658    806     126
 9 vs D.C. Un~  20    47.6  7.0   4084    520     1589    1686    843     146
10 vs FC Cinc~  22    47.7  7.0   4377    490     1608    1875    945     137
# i 21 more rows
# i 16 more variables: Touches <chr>, `Take-Ons` <chr>, `Take-Ons` <chr>,
#   `Take-Ons` <chr>, `Take-Ons` <chr>, `Take-Ons` <chr>, Carries <chr>,
#   Carries <chr>, Carries <chr>, Carries <chr>, Carries <chr>, Carries <chr>,
#   Carries <chr>, Carries <chr>, Receiving <chr>, Receiving <chr>

[[23]]
# A tibble: 31 x 23
   ``    ``    ``    `Playing Time` `Playing Time` `Playing Time` `Playing Time`
   <chr> <chr> <chr> <chr>          <chr>          <chr>          <chr>
 1 Squad # Pl  Age   MP             Min            Mn/MP          Min%
 2 Atla~ 23    29.2  7              630            90             100
 3 Aust~ 19    28.2  7              630            90             100
 4 CF M~ 23    24.2  7              630            90             100
 5 Char~ 18    29.3  7              630            90             100
 6 Chic~ 22    25.9  7              630            90             100
 7 Colo~ 24    26.5  7              630            90             100
 8 Colu~ 20    26.6  7              630            90             100
 9 D.C.~ 20    26.1  7              630            90             100
10 FC C~ 22    27.5  7              630            90             100
# i 21 more rows
# i 16 more variables: `Playing Time` <chr>, Starts <chr>, Starts <chr>,
#   Starts <chr>, Subs <chr>, Subs <chr>, Subs <chr>, `Team Success` <chr>,
#   `Team Success` <chr>, `Team Success` <chr>, `Team Success` <chr>,
#   `Team Success` <chr>, `Team Success (xG)` <chr>, `Team Success (xG)` <chr>,
#   `Team Success (xG)` <chr>, `Team Success (xG)` <chr>

[[24]]
# A tibble: 31 x 23
   ``    ``    ``    `Playing Time` `Playing Time` `Playing Time` `Playing Time`
   <chr> <chr> <chr> <chr>          <chr>          <chr>          <chr>
 1 Squad # Pl  Age   MP             Min            Mn/MP          Min%
 2 vs A~ 23    27.3  7              630            90             100
 3 vs A~ 19    26.7  7              630            90             100
 4 vs C~ 23    27.1  7              630            90             100
 5 vs C~ 18    28.1  7              630            90             100
 6 vs C~ 22    26.3  7              630            90             100
 7 vs C~ 24    28.0  7              630            90             100
 8 vs C~ 20    26.1  7              630            90             100
 9 vs D~ 20    26.9  7              630            90             100
10 vs F~ 22    27.6  7              630            90             100
# i 21 more rows
# i 16 more variables: `Playing Time` <chr>, Starts <chr>, Starts <chr>,
#   Starts <chr>, Subs <chr>, Subs <chr>, Subs <chr>, `Team Success` <chr>,
#   `Team Success` <chr>, `Team Success` <chr>, `Team Success` <chr>,
#   `Team Success` <chr>, `Team Success (xG)` <chr>, `Team Success (xG)` <chr>,
#   `Team Success (xG)` <chr>, `Team Success (xG)` <chr>

[[25]]
# A tibble: 31 x 19
   ``    ``    ``    Performance Performance Performance Performance Performance
   <chr> <chr> <chr> <chr>       <chr>       <chr>       <chr>       <chr>
 1 Squad # Pl  90s   CrdY        CrdR        2CrdY       Fls         Fld
 2 Atla~ 23    7.0   15          0           0           92          82
 3 Aust~ 19    7.0   19          0           0           91          80
 4 CF M~ 23    7.0   18          0           0           88          102
 5 Char~ 18    7.0   11          1           0           78          74
 6 Chic~ 22    7.0   11          0           0           99          68
 7 Colo~ 24    7.0   11          0           0           78          61
 8 Colu~ 20    7.0   11          1           0           72          95
 9 D.C.~ 20    7.0   20          0           0           119         66
10 FC C~ 22    7.0   13          1           1           69          66
# i 21 more rows
# i 11 more variables: Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, `Aerial Duels` <chr>, `Aerial Duels` <chr>,
#   `Aerial Duels` <chr>

[[26]]
# A tibble: 31 x 19
   ``    ``    ``    Performance Performance Performance Performance Performance
   <chr> <chr> <chr> <chr>       <chr>       <chr>       <chr>       <chr>
 1 Squad # Pl  90s   CrdY        CrdR        2CrdY       Fls         Fld
 2 vs A~ 23    7.0   10          0           0           86          87
 3 vs A~ 19    7.0   19          0           0           84          87
 4 vs C~ 23    7.0   11          0           0           107         82
 5 vs C~ 18    7.0   13          2           1           79          74
 6 vs C~ 22    7.0   5           0           0           72          94
 7 vs C~ 24    7.0   9           0           0           66          74
 8 vs C~ 20    7.0   19          0           0           102         67
 9 vs D~ 20    7.0   18          0           0           70          116
10 vs F~ 22    7.0   8           0           0           74          66
# i 21 more rows
# i 11 more variables: Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, `Aerial Duels` <chr>, `Aerial Duels` <chr>,
#   `Aerial Duels` <chr>
# Step 3: Extract the correct table (the fifth table on the page)
Squad2 <- html_table(Squad, header = TRUE, fill = TRUE)[[5]]
Squad2
```

```
# A tibble: 31 x 32
   ``              ``    ``    ``    `Playing Time` `Playing Time` `Playing Time`
   <chr>           <chr> <chr> <chr> <chr>          <chr>          <chr>
 1 Squad           # Pl  Age   Poss  MP             Starts         Min
 2 Atlanta Utd     23    29.2  49.1  7              77             630
 3 Austin          19    28.2  42.4  7              77             630
 4 CF Montréal     23    24.2  51.9  7              77             630
 5 Charlotte       18    29.3  50.1  7              77             630
 6 Chicago Fire    22    25.9  47.7  7              77             630
 7 Colorado Rapi~  24    26.5  45.7  7              77             630
 8 Columbus Crew   20    26.6  56.6  7              77             630
 9 D.C. United     20    26.1  52.4  7              77             630
10 FC Cincinnati   22    27.5  52.3  7              77             630
# i 21 more rows
# i 25 more variables: `Playing Time` <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Performance <chr>,
#   Performance <chr>, Performance <chr>, Performance <chr>, Expected <chr>,
#   Expected <chr>, Expected <chr>, Expected <chr>, Progression <chr>,
#   Progression <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>,
#   `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, `Per 90 Minutes` <chr>, ...
```

```
# Step 4: Keep only relevant columns and clean the data
Squad2_cleaned <- Squad2 |>
  row_to_names(row_number = 1) |>   # promotes row 1 to column names
  clean_names() |>                  # make the column names snake_case
  select(1:16) |>                   # keep only the first 16 columns
  filter(squad != "Squad") |>       # remove header repeats if any
  mutate(across(2:16, parse_number))# apply parse_number to cols 2-16
```

```
Warning: Row 1 does not provide unique names. Consider running clean_names()
after row_to_names().
```

```
Squad2_cleaned
```

```
# A tibble: 30 x 16
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Atlan~        23  29.2  49.1     7     77   630     7     9     6    15     8
 2 Austin        19  28.2  42.4     7     77   630     7     5     5    10     5
 3 CF Mo~        23  24.2  51.9     7     77   630     7     4     3     7     4
 4 Charl~        18  29.3  50.1     7     77   630     7    11     6    17    10
 5 Chica~        22  25.9  47.7     7     77   630     7    14     9    23    13
 6 Color~        24  26.5  45.7     7     77   630     7     8     5    13     7
 7 Colum~        20  26.6  56.6     7     77   630     7     9     8    17     9
 8 D.C. ~        20  26.1  52.4     7     77   630     7     9     7    16     8
 9 FC Ci~        22  27.5  52.3     7     77   630     7     9     4    13     7
10 FC Da~        20  27.9  44.4     7     77   630     7    10     6    16    10
# i 20 more rows
# i 4 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>
```

Creating a Custom Web Scraping Function:

Next, we generalize this scraping process by writing a custom function called scrape_fbref_table(). This function takes in a URL and table number and performs all the cleaning steps automatically. We use it to easily scrape multiple pages later on.

```
# Custom Function
scrape_fbref_table <- function(url, table_number = 5, n_cols = 16) {
  page <- read_html(url)
  tables <- html_nodes(page, "table")
  raw_table <- html_table(tables, fill = TRUE)[[table_number]]

  cleaned_table <- raw_table |>
    row_to_names(row_number = 1) |>
    clean_names() |>
    select(1:n_cols) |>
    filter(squad != "Squad") |>
    mutate(across(all_of(2:n_cols), parse_number))

  return(cleaned_table)
}

Squad2_cleaned <- scrape_fbref_table("https://fbref.com/en/comps/22/Major-League-Soccer-Stats
```

```
Warning: Row 1 does not provide unique names. Consider running clean_names()
after row_to_names().
```

```
Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
i Please use `all_of()` or `any_of()` instead.
  # Was:
  data %>% select(n_cols)

  # Now:
  data %>% select(all_of(n_cols))

See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```

```
Squad2_cleaned
```

```
# A tibble: 30 x 16
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Atlan~        23  29.2  49.1     7     77   630     7     9     6    15     8
 2 Austin        19  28.2  42.4     7     77   630     7     5     5    10     5
 3 CF Mo~        23  24.2  51.9     7     77   630     7     4     3     7     4
 4 Charl~        18  29.3  50.1     7     77   630     7    11     6    17    10
 5 Chica~        22  25.9  47.7     7     77   630     7    14     9    23    13
 6 Color~        24  26.5  45.7     7     77   630     7     8     5    13     7
 7 Colum~        20  26.6  56.6     7     77   630     7     9     8    17     9
 8 D.C. ~        20  26.1  52.4     7     77   630     7     9     7    16     8
 9 FC Ci~        22  27.5  52.3     7     77   630     7     9     4    13     7
10 FC Da~        20  27.9  44.4     7     77   630     7    10     6    16    10
# i 20 more rows
# i 4 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>
```

Iterating Over Multiple Competitions

We used purrr::pmap() to iterate over multiple variables — specifically, league URLs, the table numbers containing the “Squad Standard Stats” table for each competition, and the league names. This allowed us to apply our custom scraping function across multiple soccer leagues, each with its own unique webpage and table structure. This approach demonstrates how iteration over multiple inputs can automate the data collection process across structured but inconsistent sources.

```
# Step 1: Define league names, URLs, and their specific table numbers
leagues <- tibble::tibble(
  league = c("MLS", "Premier_League", "La_Liga", "Bundesliga", "Serie_A"),
  url = c(
    "https://fbref.com/en/comps/22/Major-League-Soccer-Stats",
    "https://fbref.com/en/comps/9/Premier-League-Stats",
    "https://fbref.com/en/comps/12/La-Liga-Stats",
    "https://fbref.com/en/comps/20/Bundesliga-Stats",
    "https://fbref.com/en/comps/11/Serie-A-Stats"
  ),
  table_number = c(5, 3, 3, 3, 3)  # Specify table index for each league
)

# Step 2: Scrape each league using map3 to pass 3 arguments
league_tables <- pmap(
  list(leagues$url, leagues$table_number, leagues$league),
  function(url, table_num, league_name) {
    scrape_fbref_table(url, table_number = table_num) |>
      mutate(league = league_name)  # Optionally tag league in each table
  }
)
```

```
Warning: Row 1 does not provide unique names. Consider running clean_names() after row_to_nam
Row 1 does not provide unique names. Consider running clean_names() after row_to_names().
Row 1 does not provide unique names. Consider running clean_names() after row_to_names().
Row 1 does not provide unique names. Consider running clean_names() after row_to_names().
Row 1 does not provide unique names. Consider running clean_names() after row_to_names().
```

```
# Step 3: Name each list entry by league
names(league_tables) <- leagues$league

# Now each league table is separate and named:
league_tables$MLS
```

```
# A tibble: 30 x 17
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Atlan~        23  29.2  49.1     7     77   630     7     9     6    15     8
 2 Austin        19  28.2  42.4     7     77   630     7     5     5    10     5
 3 CF Mo~        23  24.2  51.9     7     77   630     7     4     3     7     4
 4 Charl~        18  29.3  50.1     7     77   630     7    11     6    17    10
 5 Chica~        22  25.9  47.7     7     77   630     7    14     9    23    13
 6 Color~        24  26.5  45.7     7     77   630     7     8     5    13     7
 7 Colum~        20  26.6  56.6     7     77   630     7     9     8    17     9
 8 D.C. ~        20  26.1  52.4     7     77   630     7     9     7    16     8
 9 FC Ci~         22  27.5  52.3     7     77   630     7     9     4    13     7
10 FC Da~         20  27.9  44.4     7     77   630     7    10     6    16    10
# i 20 more rows
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```

```
league_tables$Premier_League
```

```
# A tibble: 20 x 17
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Arsen~        24  26.5  56.1    31    341  2790    31    54    43    97    52
 2 Aston~        28  27.8  51.2    31    341  2790    31    45    36    81    43
 3 Bourn~        28  25.8  47.5    31    341  2790    31    50    37    87    44
 4 Brent~        27  26.6  48.1    31    341  2790    31    51    32    83    46
 5 Brigh~        31  25.6  52.3    31    341  2790    31    47    31    78    44
 6 Chels~        29  24.4  57.9    31    341  2790    31    53    41    94    50
 7 Cryst~        29  26.9  44.2    30    330  2700    30    37    29    66    35
 8 Evert~        26  28.8  40        31    341  2790    31    30    19    49    28
 9 Fulham        26  28.7  52        31    341  2790    31    46    39    85    43
10 Ipswi~        32  26.5  41.4    31    341  2790    31    30    22    52    28
11 Leice~        29  27.3  45.7    31    341  2790    31    25    20    45    23
12 Liver~        24  27.9  57.8    31    341  2790    31    72    54   126    63
13 Manch~        29  27.5  61        31    341  2790    31    56    41    97    54
14 Manch~        29  26.3  52.8    31    341  2790    31    35    23    58    32
15 Newca~        24  27.9  50        30    330  2700    30    51    38    89    48
16 Nott'~        23  26.9  40        31    341  2790    31    50    36    86    47
17 South~        34  26    50.1    31    341  2790    31    22    14    36    22
18 Totte~        31  25.9  56.4    31    341  2790    31    55    43    98    53
19 West ~        27  28.8  47.6    31    341  2790    31    33    20    53    30
20 Wolves        29  27.6  47.8    31    341  2790    31    43    35    78    43
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```

```
league_tables$La_Liga
```

```
# A tibble: 20 x 17
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Alavés        29  27.2  45.3    30    330  2700    30    33    18    51    27
 2 Athle~        30  27.6  49.2    30    330  2700    30    46    36    82    44
 3 Atlét~        24  29.1  51.8    30    330  2700    30    49    38    87    45
 4 Barce~        27  25.4  67.3    30    330  2700    30    81    59   140    76
 5 Betis         35  28.1  51.5    30    330  2700    30    39    24    63    33
 6 Celta~        30  27.7  54.1    30    330  2700    30    43    30    73    37
 7 Espan~        26  26    38.5    29    319  2610    29    29    21    50    25
 8 Getafe        30  28.4  43      30    330  2700    30    30    17    47    25
 9 Girona        29  28    56.6    30    330  2700    30    36    27    63    32
10 Las P~        30  27.7  50.1    30    330  2700    30    32    23    55    30
11 Legan~        26  28.6  42.3    30    330  2700    30    29    21    50    24
12 Mallo~        28  29.4  46.9    30    330  2700    30    28    19    47    24
13 Osasu~        24  28.2  46      30    330  2700    30    32    17    49    25
14 Rayo ~        26  29.8  51.3    30    330  2700    30    31    24    55    31
15 Real ~        27  27.6  60.9    30    330  2700    30    63    44   107    54
16 Real ~        31  26    54.1    30    330  2700    30    29    20    49    27
17 Sevil~        34  26.8  51.5    30    330  2700    30    31    26    57    30
18 Valen~        32  25.3  46.9    30    330  2700    30    34    23    57    31
19 Valla~        35  26.2  43.3    30    330  2700    30    19    12    31    16
20 Villa~        28  27.6  49.2    29    319  2610    29    49    32    81    44
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```

```
league_tables$Bundesliga
```

```
# A tibble: 18 x 17
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Augsb~        29  27.5  43      28    308  2520    28    31    23    54    30
 2 Bayer~        28  28.4  68.9    28    308  2520    28    78    50   128    69
 3 Bochum        26  28.8  44      28    308  2520    28    27    20    47    26
 4 Dortm~        28  27.3  59.9    28    308  2520    28    50    39    89    46
 5 Eint ~        26  25.7  49.9    28    308  2520    28    55    35    90    52
 6 Freib~        26  28    49      28    308  2520    28    36    27    63    36
 7 Gladb~        26  27.2  50      28    308  2520    28    44    33    77    41
 8 Heide~        26  27.6  43.3    28    308  2520    28    32    21    53    27
 9 Hoffe~        34  27.1  49.5    28    308  2520    28    34    20    54    31
10 Holst~        27  26.3  44.3    28    308  2520    28    39    23    62    36
11 Lever~        23  27.8  58.8    28    308  2520    28    61    46   107    59
12 Mainz~        24  28.2  49.5    28    308  2520    28    44    32    76    41
13 RB Le~        29  26.3  52.7    28    308  2520    28    42    29    71    40
14 St. P~        27  27.7  44.5    28    308  2520    28    22    20    42    21
15 Stutt~        28  25.5  55.7    28    308  2520    28    48    35    83    46
16 Union~        27  27.9  41      28    308  2520    28    27    18    45    24
```

```
17 Werde~        23  28.3  49.5    28    308  2520    28    44    32    76    42
18 Wolfs~        26  25.9  45.9    28    308  2520    28    47    32    79    43
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```

```
league_tables$Serie_A
```

```
# A tibble: 20 x 17
   squad  number_pl   age  poss    mp starts   min  x90s   gls   ast   g_a  g_pk
   <chr>      <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Atala~        32  27.8  55.8    31    341  2790    31    61    42   103    57
 2 Bolog~        30  27    58.1    31    341  2790    31    50    38    88    44
 3 Cagli~        26  27.4  45.6    31    341  2790    31    29    23    52    26
 4 Como          38  27    54.2    31    341  2790    31    34    29    63    34
 5 Empoli        34  26.2  40.9    31    341  2790    31    23    14    37    21
 6 Fiore~        34  26.9  49.3    31    341  2790    31    47    31    78    41
 7 Genoa         34  27.1  45.8    31    341  2790    31    28    21    49    28
 8 Hella~        32  25.8  38      31    341  2790    31    27    17    44    25
 9 Inter         25  30.1  59.5    31    341  2790    31    66    50   116    60
10 Juven~        29  25.4  58.4    31    341  2790    31    45    30    75    40
11 Lazio         27  27.9  54.2    31    341  2790    31    51    37    88    46
12 Lecce         30  26.8  44.9    31    341  2790    31    22    17    39    20
13 Milan         34  26.4  54.6    31    341  2790    31    46    32    78    42
14 Monza         35  27.6  47.7    31    341  2790    31    24    16    40    21
15 Napoli        27  29.3  53      31    341  2790    31    45    32    77    41
16 Parma         32  24.6  45.2    31    341  2790    31    36    26    62    30
17 Roma          28  27.3  55.4    31    341  2790    31    45    27    72    38
18 Torino        28  27.4  47.6    31    341  2790    31    33    21    54    32
19 Udine~        29  27.1  47.2    31    341  2790    31    36    24    60    34
20 Venez~        36  26.1  44.6    31    341  2790    31    23    12    35    19
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```