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.

```
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
```

```
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
```

```
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 Philade~     5     4     0     1    12     6     6    12      2.4   9.6
 2     2 Charlot~     5     3     1     1    10     4     6    10      2     7
 3     3 Inter M~     4     3     1     0     9     4     5    10      2.5   6.6
 4     4 Nashvil~     5     3     1     1     8     3     5    10      2     7.8
 5     5 Chicago~     5     3     1     1    12     9     3    10      2     9
 6     6 Columbu~     5     2     3     0     6     3     3     9      1.8   5.9
 7     7 NY Red ~     5     2     2     1     6     4     2     8      1.6   8.2
 8     8 NYCFC        5     2     2     1     6     5     1     8      1.6   5.3
 9     9 Orlando~     5     2     1     2    13    11     2     7      1.4  10
10    10 FC Cinc~     5     2     1     2     6     8    -2     7      1.4   6.1
11    11 D.C. Un~     5     1     3     1     7     9    -2     6      1.2   8.3
12    12 Atlanta~     5     1     2     2     6     8    -2     5      1     6.7
13    13 NE Revo~     4     0     1     3     1     5    -4     1      0.25  1.7
```

```
14    14 Toronto~     5     0     1     4     6    12    -6     1     0.2    3.8
15    15 CF Mont~     5     0     1     4     2     9    -7     1     0.2    5.1
# 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     Phil~ 3     2     0     1     6     4     +2    6     "2.0~ 6.5   2.9
 3 2     Char~ 3     3     0     0     8     1     +7    9     "3.0~ 5.8   4.3
 4 3     Inte~ 2     1     1     0     3     2     +1    4     "2.0~ 3.2   1.7
 5 4     Nash~ 3     2     1     0     5     0     +5    7     "2.3~ 5.3   1.4
 6 5     Chic~ 1     0     1     0     2     2     0     1     "1.0~ 2.3   3.0
 7 6     Colu~ 3     1     2     0     4     2     +2    5     "1.6~ 4.5   2.3
 8 7     NY R~ 3     2     1     0     6     3     +3    7     "2.3~ 6.7   3.0
 9 8     NYCFC 2     2     0     0     4     2     +2    6     "3.0~ 3.7   1.9
10 9     Orla~ 3     2     0     1     10    7     +3    6     "2.0~ 7.0   3.5
11 10    FC C~ 3     2     1     0     5     2     +3    7     "2.3~ 4.2   3.3
12 11    D.C.~ 3     1     2     0     4     3     +1    5     "1.6~ 4.0   3.2
13 12    Atla~ 3     1     1     1     4     4     0     4     "1.3~ 4.0   4.6
14 13    NE R~ 2     0     0     2     0     3     -3    0     "0.0~ 0.6   1.6
15 14    Toro~ 1     0     0     1     1     2     -1    0     "0.0~ 0.6   1.0
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~     5     4     0     1    10     5     5    12      2.4   8.4
 2     2 Austin       5     3     0     2     4     3     1     9      1.8   6.3
 3     3 LAFC         5     3     0     2     6     6     0     9      1.8   5.4
 4     4 San Die~     5     2     2     1     7     4     3     8      1.6   8.6
 5     5 St. Lou~     5     2     2     1     4     1     3     8      1.6   5.5
 6     6 Minneso~     5     2     2     1     7     6     1     8      1.6   9.5
 7     7 Colorad~     5     2     2     1     6     7    -1     8      1.6   4.7
 8     8 FC Dall~     5     2     1     2     7     8    -1     7      1.4   8.1
 9     9 Portlan~     5     2     1     2     6     7    -1     7      1.4   4.5
10    10 SJ Eart~     5     2     0     3     8     8     0     6      1.2  10.1
```

```
11    11 Real Sa~     5     2     0     3     5     9    -4     6      1.2   6.7
12    12 Seattle~     5     1     2     2     7     7     0     5      1     6
13    13 Houston~     5     0     2     3     3     8    -5     2      0.4   4.4
14    14 LA Gala~     5     0     2     3     4    10    -6     2      0.4   4.5
15    15 Sportin~     5     0     1     4     5    10    -5     1      0.2   4
# 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~ 3     2     0     1     5     4     +1    6     2.00  4.6   4.1
 3 2     Aust~ 3     2     0     1     3     2     +1    6     2.00  4.4   2.3
 4 3     LAFC  3     2     0     1     2     1     +1    6     2.00  2.5   2.2
 5 4     San ~ 2     0     2     0     1     1     0     2     1.00  3.0   1.1
 6 5     St. ~ 2     1     1     0     1     0     +1    4     2.00  3.3   1.0
 7 6     Minn~ 2     1     1     0     3     2     +1    4     2.00  4.6   0.9
 8 7     Colo~ 2     0     1     1     3     6     -3    1     0.50  2.1   3.9
 9 8     FC D~ 2     0     0     2     1     4     -3    0     0.00  1.9   2.6
10 9     Port~ 3     1     1     1     3     5     -2    4     1.33  2.0   4.4
11 10    SJ E~ 3     1     0     2     5     3     +2    3     1.00  5.9   5.6
12 11    Real~ 3     1     0     2     3     4     -1    3     1.00  4.1   5.6
13 12    Seat~ 3     1     2     0     7     4     +3    5     1.67  4.4   2.0
14 13    Hous~ 3     0     0     3     3     8     -5    0     0.00  3.7   4.7
15 14    LA G~ 2     0     0     2     0     5     -5    0     0.00  2.5   3.9
16 15    Spor~ 3     0     1     2     4     7     -3    1     0.33  3.0   5.4
# 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    29.3  48.2  5              55             450
 3 Austin         19    28.3  43.4  5              55             450
 4 CF Montréal    22    24.2  52.6  5              55             450
 5 Charlotte      18    29.1  48.4  5              55             450
 6 Chicago Fire   22    25.7  46.8  5              55             450
 7 Colorado Rapi~ 21    26.3  45.2  5              55             450
```

```
 8 Columbus Crew  18    26.8  57.6  5              55             450
 9 D.C. United    20    26.1  52.6  5              55             450
10 FC Cincinnati  22    27.5  51.8  5              55             450
# 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    27.5  51.8  5              55             450
 3 vs Austin      19    26.6  56.6  5              55             450
 4 vs CF Montréal 22    27.6  47.4  5              55             450
 5 vs Charlotte   18    28.0  51.6  5              55             450
 6 vs Chicago Fi~ 22    26.6  53.2  5              55             450
 7 vs Colorado R~ 21    28.1  54.8  5              55             450
 8 vs Columbus C~ 18    26.5  42.4  5              55             450
 9 vs D.C. United 20    26.6  47.4  5              55             450
10 vs FC Cincinn~ 22    27.7  48.2  5              55             450
# 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     5              5              450            5.0
 3 Austin       1     5              5              450            5.0
 4 CF Montréal  1     5              5              450            5.0
 5 Charlotte    1     5              5              450            5.0
 6 Chicago Fi~  1     5              5              450            5.0
```

```
 7 Colorado R~ 2     5              5              450            5.0
 8 Columbus C~ 2     5              5              450            5.0
 9 D.C. United 1     5              5              450            5.0
10 FC Cincinn~ 1     5              5              450            5.0
# 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     5              5              450            5.0
 3 vs Austin   1     5              5              450            5.0
 4 vs CF Mont~ 1     5              5              450            5.0
 5 vs Charlot~ 1     5              5              447            5.0
 6 vs Chicago~ 1     5              5              450            5.0
 7 vs Colorad~ 2     5              5              450            5.0
 8 vs Columbu~ 2     5              5              450            5.0
 9 vs D.C. Un~ 1     5              5              450            5.0
10 vs FC Cinc~ 1     5              5              450            5.0
# 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     5.0   8     0     1     2     0     7.5      0.31     -0.5
 3 Austin   1     5.0   3     0     1     0     0     2.5      0.18     -0.5
 4 CF Mont~ 1     5.0   9     0     0     3     0     10.6     0.32     +1.6
 5 Charlot~ 1     5.0   4     1     0     0     0     6.4      0.22     +2.4
 6 Chicago~ 1     5.0   9     0     0     1     1     8.2      0.39     +0.2
 7 Colorad~ 2     5.0   7     0     0     1     1     8.9      0.32     +2.9
```

```
 8 Columbu~ 2     5.0   3     0     0     1     0     4.6      0.29     +1.6
 9 D.C. Un~ 1     5.0   9     1     0     2     0     10.0     0.44     +1.0
10 FC Cinc~ 1     5.0   8     0     0     0     1     7.5      0.38     +0.5
# 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     5.0   6     0     0     1     1     8.4      0.34     +3.4
 3 vs Aust~ 1     5.0   4     0     0     2     0     5.0      0.28     +1.0
 4 vs CF M~ 1     5.0   2     0     0     1     0     3.8      0.36     +1.8
 5 vs Char~ 1     5.0   10    0     0     1     1     7.4      0.43     -1.6
 6 vs Chic~ 1     5.0   12    1     0     2     0     10.8     0.38     -1.2
 7 vs Colo~ 2     5.0   6     0     0     0     0     4.1      0.31     -1.9
 8 vs Colu~ 2     5.0   6     0     0     0     1     6.6      0.33     +1.6
 9 vs D.C.~ 1     5.0   7     1     0     1     0     8.9      0.34     +1.9
10 vs FC C~ 1     5.0   6     1     1     0     0     6.4      0.23     +0.4
# 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    5.0   5        63       25       39.7     12.60    5.00
 3 Austin       19    5.0   4        66       18       27.3     13.20    3.60
 4 CF Montréal  22    5.0   2        44       12       27.3     8.80     2.40
 5 Charlotte    18    5.0   9        55       18       32.7     11.00    3.60
 6 Chicago Fi~  22    5.0   12       56       25       44.6     11.20    5.00
 7 Colorado R~  21    5.0   6        41       13       31.7     8.20     2.60
 8 Columbus C~  18    5.0   5        64       20       31.3     12.80    4.00
 9 D.C. United  20    5.0   7        63       24       38.1     12.60    4.80
10 FC Cincinn~  22    5.0   6        59       23       39.0     11.80    4.60
```

```
# 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    5.0   8        56       25       44.6     11.20    5.00
 3 vs Austin   19    5.0   3        51       14       27.5     10.20    2.80
 4 vs CF Mont~ 22    5.0   9        73       34       46.6     14.60    6.80
 5 vs Charlot~ 18    5.0   4        79       24       30.4     15.80    4.80
 6 vs Chicago~ 22    5.0   8        62       18       29.0     12.40    3.60
 7 vs Colorad~ 21    5.0   6        87       30       34.5     17.40    6.00
 8 vs Columbu~ 18    5.0   3        42       16       38.1     8.40     3.20
 9 vs D.C. Un~ 20    5.0   9        59       21       35.6     11.80    4.20
10 vs FC Cinc~ 22    5.0   7        62       20       32.3     12.40    4.00
# 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    5.0   2128  2605  81.7  37977 13850 920   1060  86.8  971
 3 Austin      19    5.0   1743  2163  80.6  31799 12075 793   883   89.8  681
 4 CF Montré~  22    5.0   2035  2549  79.8  37790 13356 844   964   87.6  929
 5 Charlotte   18    5.0   1972  2431  81.1  35143 12010 852   936   91.0  899
 6 Chicago F~  22    5.0   2018  2430  83.0  33184 12070 1005  1096  91.7  787
 7 Colorado ~  21    5.0   1596  2175  73.4  27348 11105 764   883   86.5  645
 8 Columbus ~  18    5.0   2708  3126  86.6  40491 13993 1515  1617  93.7  947
 9 D.C. Unit~  20    5.0   1967  2511  78.3  35383 13666 918   1055  87.0  779
10 FC Cincin~  22    5.0   2199  2732  80.5  38631 13462 893   1039  85.9  1091
# 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    5.0   2355  2818  83.6  41341 13175 1029  1131  91.0  1109
 3 vs Austin  19    5.0   2409  2858  84.3  44020 15519 1002  1124  89.1  1139
 4 vs CF Mon~ 22    5.0   1810  2290  79.0  32990 13380 791   910   86.9  774
 5 vs Charlo~ 18    5.0   2182  2648  82.4  38012 13762 981   1103  88.9  956
 6 vs Chicag~ 22    5.0   2273  2758  82.4  39027 12708 1045  1157  90.3  957
 7 vs Colora~ 21    5.0   2094  2622  79.9  37469 13997 915   1020  89.7  916
 8 vs Columb~ 18    5.0   1869  2299  81.3  31145 11023 928   1042  89.1  694
 9 vs D.C. U~ 20    5.0   1763  2238  78.8  30407 11757 827   918   90.1  715
10 vs FC Cin~ 22    5.0   2020  2557  79.0  34817 12926 926   1034  89.6  892
# 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    5.0   2605  2380         215          61           2
 3 Austin   19    5.0   2163  1945         208          60           7
 4 CF Mon~  22    5.0   2549  2323         220          78           3
 5 Charlo~  18    5.0   2431  2197         230          56           4
 6 Chicag~  22    5.0   2430  2200         216          63           7
 7 Colora~  21    5.0   2175  1943         224          52           8
 8 Columb~  18    5.0   3126  2913         206          84           9
 9 D.C. U~  20    5.0   2511  2253         249          59           5
10 FC Cin~  22    5.0   2732  2478         238          50           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>

[[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    5.0   2818  2603         205          74           5
 3 vs Aus~ 19    5.0   2858  2618         239          76           2
 4 vs CF ~ 22    5.0   2290  2069         216          58           8
 5 vs Cha~ 18    5.0   2648  2416         225          55           8
 6 vs Chi~ 22    5.0   2758  2505         238          74           3
 7 vs Col~ 21    5.0   2622  2374         239          64           9
 8 vs Col~ 18    5.0   2299  2095         188          52           9
 9 vs D.C~ 20    5.0   2238  1989         238          98           3
10 vs FC ~ 22    5.0   2557  2321         225          67           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    5.0   116   23.20 91          10          5           5
 3 Aust~ 19    5.0   118   23.60 91          11          3           6
 4 CF M~ 22    5.0   79    15.80 59          6           4           5
 5 Char~ 18    5.0   91    18.20 69          5           5           6
 6 Chic~ 22    5.0   99    19.80 75          8           4           7
 7 Colo~ 21    5.0   71    14.20 59          7           1           1
 8 Colu~ 18    5.0   110   22.00 78          5           12          5
 9 D.C.~ 20    5.0   113   22.60 88          9           5           7
10 FC C~ 22    5.0   109   21.80 81          7           7           6
# 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    5.0   95    19.00 64          7           7           8
 3 vs A~ 19    5.0   93    18.60 70          9           4           6
 4 vs C~ 22    5.0   122   24.40 89          11          5           11
```

```
 5 vs C~ 18    5.0   150   30.00 121         9           5           6
 6 vs C~ 22    5.0   116   23.20 86          8           6           6
 7 vs C~ 21    5.0   153   30.60 112         20          2           14
 8 vs C~ 18    5.0   72    14.40 53          10          1           4
 9 vs D~ 20    5.0   109   21.80 90          5           1           5
10 vs F~ 22    5.0   111   22.20 82          9           7           6
# 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    5.0   76      49      34      36      6       27
 3 Austin          19    5.0   79      44      53      21      5       31
 4 CF Montréal     22    5.0   78      51      28      40      10      30
 5 Charlotte       18    5.0   69      42      31      28      10      36
 6 Chicago Fire    22    5.0   68      36      27      27      14      29
 7 Colorado Rapi~ 21    5.0   83      43      34      40      9       37
 8 Columbus Crew   18    5.0   62      35      22      29      11      27
 9 D.C. United     20    5.0   86      54      42      30      14      52
10 FC Cincinnati   22    5.0   82      54      37      35      10      41
# 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    5.0   68      41      23      39      6       37
 3 vs Austin       19    5.0   68      45      28      22      18      33
 4 vs CF Montréal  22    5.0   70      40      34      24      12      24
 5 vs Charlotte    18    5.0   107     76      50      40      17      54
 6 vs Chicago Fi~ 22    5.0   99      58      42      37      20      48
 7 vs Colorado R~ 21    5.0   81      45      43      26      12      29
 8 vs Columbus C~ 18    5.0   87      44      49      31      7       39
 9 vs D.C. United  20    5.0   103     59      53      31      19      36
```

```
10 vs FC Cincinn~ 22    5.0   104     60      45      42      17      48
# 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    48.2  5.0   3099    300     1016    1420    680     129
 3 Austin      19    43.4  5.0   2729    317     971     1157    624     98
 4 CF Montréal 22    52.6  5.0   3087    371     1152    1335    624     85
 5 Charlotte   18    48.4  5.0   2971    352     1053    1231    715     103
 6 Chicago Fi~ 22    46.8  5.0   3040    415     1169    1321    573     97
 7 Colorado R~ 21    45.2  5.0   2765    324     847     1167    774     97
 8 Columbus C~ 18    57.6  5.0   3597    271     1013    1670    945     130
 9 D.C. United 20    52.6  5.0   3044    270     922     1310    838     123
10 FC Cincinn~ 22    51.8  5.0   3367    296     1112    1593    695     101
# 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    51.8  5.0   3317    304     1065    1585    694     109
 3 vs Austin   19    56.6  5.0   3360    243     868     1569    946     123
 4 vs CF Mont~ 22    47.4  5.0   2872    299     960     1228    703     130
 5 vs Charlot~ 18    51.6  5.0   3214    307     1061    1401    779     122
 6 vs Chicago~ 22    53.2  5.0   3287    284     1001    1539    775     144
 7 vs Colorad~ 21    54.8  5.0   3282    298     1172    1430    703     145
 8 vs Columbu~ 18    42.4  5.0   2796    338     1095    1182    533     82
 9 vs D.C. Un~ 20    47.4  5.0   2901    382     1121    1217    590     96
10 vs FC Cinc~ 22    48.2  5.0   3172    322     1094    1434    685     96
# 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    29.3  5              450            90             100
 3 Aust~ 19    28.3  5              450            90             100
 4 CF M~ 22    24.2  5              450            90             100
 5 Char~ 18    29.1  5              450            90             100
 6 Chic~ 22    25.7  5              450            90             100
 7 Colo~ 21    26.3  5              450            90             100
 8 Colu~ 18    26.8  5              450            90             100
 9 D.C.~ 20    26.1  5              450            90             100
10 FC C~ 22    27.5  5              450            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    27.5  5              450            90             100
 3 vs A~ 19    26.6  5              450            90             100
 4 vs C~ 22    27.6  5              450            90             100
 5 vs C~ 18    28.0  5              450            90             100
 6 vs C~ 22    26.6  5              450            90             100
 7 vs C~ 21    28.1  5              450            90             100
 8 vs C~ 18    26.5  5              450            90             100
 9 vs D~ 20    26.6  5              450            90             100
10 vs F~ 22    27.7  5              450            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    5.0   13          0           0           68          52
 3 Aust~ 19    5.0   12          0           0           67          60
 4 CF M~ 22    5.0   12          0           0           59          72
 5 Char~ 18    5.0   6           1           0           56          50
 6 Chic~ 22    5.0   9           0           0           64          48
 7 Colo~ 21    5.0   6           0           0           58          44
 8 Colu~ 18    5.0   9           1           0           50          68
 9 D.C.~ 20    5.0   14          0           0           93          49
10 FC C~ 22    5.0   11          1           1           53          43
# 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    5.0   10          0           0           54          65
 3 vs A~ 19    5.0   15          0           0           62          65
 4 vs C~ 22    5.0   8           0           0           75          54
 5 vs C~ 18    5.0   10          2           1           51          54
 6 vs C~ 22    5.0   4           0           0           52          60
 7 vs C~ 21    5.0   5           0           0           46          55
 8 vs C~ 18    5.0   10          0           0           74          47
 9 vs D~ 20    5.0   15          0           0           52          90
10 vs F~ 22    5.0   5           0           0           46          52
# 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>
```

```r
# 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    29.3  48.2  5              55             450
 3 Austin         19    28.3  43.4  5              55             450
 4 CF Montréal    22    24.2  52.6  5              55             450
 5 Charlotte      18    29.1  48.4  5              55             450
 6 Chicago Fire   22    25.7  46.8  5              55             450
 7 Colorado Rapi~ 21    26.3  45.2  5              55             450
 8 Columbus Crew  18    26.8  57.6  5              55             450
 9 D.C. United    20    26.1  52.6  5              55             450
10 FC Cincinnati  22    27.5  51.8  5              55             450
# 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>, ...
```

```r
# 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().
```

```r
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  29.3  48.2     5     55   450     5     5     4     9     5
 2 Austin        19  28.3  43.4     5     55   450     5     4     4     8     4
 3 CF Mo~        22  24.2  52.6     5     55   450     5     2     2     4     2
 4 Charl~        18  29.1  48.4     5     55   450     5     9     5    14     9
 5 Chica~        22  25.7  46.8     5     55   450     5    12     7    19    11
 6 Color~        21  26.3  45.2     5     55   450     5     6     4    10     6
 7 Colum~        18  26.8  57.6     5     55   450     5     5     4     9     5
 8 D.C. ~        20  26.1  52.6     5     55   450     5     7     5    12     6
 9 FC Ci~        22  27.5  51.8     5     55   450     5     6     3     9     5
10 FC Da~        18  28.1  46.4     5     55   450     5     7     4    11     7
# 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.

```r
# 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  29.3  48.2     5     55   450     5     5     4     9     5
 2 Austin        19  28.3  43.4     5     55   450     5     4     4     8     4
 3 CF Mo~        22  24.2  52.6     5     55   450     5     2     2     4     2
 4 Charl~        18  29.1  48.4     5     55   450     5     9     5    14     9
 5 Chica~        22  25.7  46.8     5     55   450     5    12     7    19    11
 6 Color~        21  26.3  45.2     5     55   450     5     6     4    10     6
 7 Colum~        18  26.8  57.6     5     55   450     5     5     4     9     5
 8 D.C. ~        20  26.1  52.6     5     55   450     5     7     5    12     6
 9 FC Ci~        22  27.5  51.8     5     55   450     5     6     3     9     5
10 FC Da~        18  28.1  46.4     5     55   450     5     7     4    11     7
# 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_nar
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  29.3  48.2     5     55   450     5     5     4     9     5
 2 Austin        19  28.3  43.4     5     55   450     5     4     4     8     4
 3 CF Mo~        22  24.2  52.6     5     55   450     5     2     2     4     2
 4 Charl~        18  29.1  48.4     5     55   450     5     9     5    14     9
 5 Chica~        22  25.7  46.8     5     55   450     5    12     7    19    11
 6 Color~        21  26.3  45.2     5     55   450     5     6     4    10     6
 7 Colum~        18  26.8  57.6     5     55   450     5     5     4     9     5
 8 D.C. ~        20  26.1  52.6     5     55   450     5     7     5    12     6
```

```
 9 FC Ci~        22  27.5  51.8     5     55   450     5     6     3     9     5
10 FC Da~        18  28.1  46.4     5     55   450     5     7     4    11     7
# 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  55.8    29    319  2610    29    51    40    91    49
 2 Aston~        28  27.7  51.4    29    319  2610    29    40    31    71    38
 3 Bourn~        28  25.8  47      29    319  2610    29    47    36    83    41
 4 Brent~        27  26.6  48.2    29    319  2610    29    50    32    82    46
 5 Brigh~        30  25.6  51.8    29    319  2610    29    46    30    76    43
 6 Chels~        29  24.3  58.1    29    319  2610    29    52    40    92    49
 7 Cryst~        27  26.9  43.9    28    308  2520    28    34    26    60    32
 8 Evert~        26  28.9  40.8    29    319  2610    29    29    19    48    28
 9 Fulham        26  28.7  52.7    29    319  2610    29    42    37    79    39
10 Ipswi~        32  26.4  41.4    29    319  2610    29    27    19    46    25
11 Leice~        28  27.3  45.9    29    319  2610    29    25    20    45    23
12 Liver~        24  27.9  57.1    29    319  2610    29    69    51   120    60
13 Manch~        27  27.5  60.7    29    319  2610    29    54    40    94    52
14 Manch~        29  26.4  52.7    29    319  2610    29    35    23    58    32
15 Newca~        24  27.8  50.3    28    308  2520    28    46    35    81    43
16 Nott'~        23  26.9  40      29    319  2610    29    48    34    82    45
17 South~        34  26    50.4    29    319  2610    29    20    12    32    20
18 Totte~        31  25.8  56.8    29    319  2610    29    52    41    93    51
19 West ~        27  28.9  47      29    319  2610    29    31    18    49    28
20 Wolves        28  27.6  47.8    29    319  2610    29    40    34    74    40
# 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    28    308  2520    28    32    17    49    26
 2 Athle~        30  27.6  48.9    28    308  2520    28    46    36    82    44
```

```
 3 Atlét~        24  29.1  51.3    28    308  2520    28    46    37    83    43
 4 Barce~        27  25.3  67.1    28    308  2520    28    77    55   132    72
 5 Betis         35  28    52.3    28    308  2520    28    36    21    57    30
 6 Celta~        30  27.7  53.8    28    308  2520    28    40    27    67    34
 7 Espan~        26  26    38.7    27    297  2430    27    24    18    42    22
 8 Getafe        30  28.3  42.4    28    308  2520    28    25    13    38    20
 9 Girona        29  28    57.3    28    308  2520    28    35    26    61    31
10 Las P~        30  27.8  50.4    28    308  2520    28    30    21    51    28
11 Legan~        25  28.6  42.2    28    308  2520    28    26    20    46    22
12 Mallo~        27  29.4  46.7    28    308  2520    28    27    18    45    23
13 Osasu~        23  28.2  46.3    28    308  2520    28    31    16    47    24
14 Rayo ~        24  29.8  50.9    28    308  2520    28    29    23    52    29
15 Real ~        26  27.6  60.4    28    308  2520    28    59    42   101    51
16 Real ~        30  25.9  54.3    28    308  2520    28    24    18    42    22
17 Sevil~        34  26.8  52      28    308  2520    28    29    25    54    28
18 Valen~        32  25.3  47.4    28    308  2520    28    31    20    51    28
19 Valla~        32  26.2  43      28    308  2520    28    18    11    29    15
20 Villa~        28  27.6  49.1    27    297  2430    27    47    30    77    42
# 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~        28  27.5  43.4    26    286  2340    26    29    21    50    28
 2 Bayer~        27  28.5  69      26    286  2340    26    73    46   119    64
 3 Bochum        26  28.8  44      26    286  2340    26    26    19    45    25
 4 Dortm~        28  27.3  61.1    26    286  2340    26    43    32    75    39
 5 Eint ~        26  25.7  49.6    26    286  2340    26    54    35    89    51
 6 Freib~        26  27.9  47.9    26    286  2340    26    34    25    59    34
 7 Gladb~        26  27.3  50.9    26    286  2340    26    42    32    74    39
 8 Heide~        26  27.6  43.2    26    286  2340    26    31    21    52    27
 9 Hoffe~        34  27.1  49.8    26    286  2340    26    32    19    51    30
10 Holst~        27  26.2  44      26    286  2340    26    38    22    60    35
11 Lever~        23  27.7  58.8    26    286  2340    26    57    42    99    55
12 Mainz~        24  28.2  49      26    286  2340    26    42    30    72    39
13 RB Le~        29  26.4  51.9    26    286  2340    26    39    27    66    37
14 St. P~        27  27.7  44.2    26    286  2340    26    19    17    36    18
15 Stutt~        27  25.5  56.2    26    286  2340    26    44    31    75    42
16 Union~        27  27.8  41.4    26    286  2340    26    24    16    40    21
```

```
17 Werde~        23  28.3  50.1    26    286  2340    26    39    29    68    37
18 Wolfs~        25  25.9  45      26    286  2340    26    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  27.7  55.8    29    319  2610    29    61    42   103    57
 2 Bolog~        29  27    58      29    319  2610    29    48    36    84    42
 3 Cagli~        26  27.3  45.4    29    319  2610    29    26    21    47    23
 4 Como          38  27    53.4    29    319  2610    29    30    26    56    30
 5 Empoli        33  26.2  41.2    29    319  2610    29    22    13    35    20
 6 Fiore~        34  26.8  49.5    29    319  2610    29    45    30    75    39
 7 Genoa         34  27.1  45.3    29    319  2610    29    27    21    48    27
 8 Hella~        32  25.8  37.8    29    319  2610    29    26    17    43    24
 9 Inter         25  30.1  59.3    29    319  2610    29    62    47   109    56
10 Juven~        29  25.4  58.7    29    319  2610    29    43    29    72    38
11 Lazio         26  27.9  54.4    29    319  2610    29    49    35    84    44
12 Lecce         30  26.7  44.8    29    319  2610    29    21    16    37    19
13 Milan         34  26.3  54.9    29    319  2610    29    43    29    72    39
14 Monza         35  27.4  47.6    29    319  2610    29    23    15    38    20
15 Napoli        26  29.3  53.6    29    319  2610    29    42    30    72    38
16 Parma         32  24.5  45      29    319  2610    29    34    24    58    28
17 Roma          28  27.2  55.9    29    319  2610    29    43    26    69    36
18 Torino        28  27.4  47.4    29    319  2610    29    31    20    51    30
19 Udine~        29  27.2  47.4    29    319  2610    29    35    24    59    33
20 Venez~        36  26.1  44.6    29    319  2610    29    23    12    35    19
# i 5 more variables: pk <dbl>, p_katt <dbl>, crd_y <dbl>, crd_r <dbl>,
#   league <chr>
```