

Data Wrangling for 2019/20 and 2020/21

```
df_2020 <- read.csv('https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/master/data/2019-20')
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.0      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Separating the first column

```
df_2020 <- df_2020 %>%
  separate(name, into = c("First_Name", "Last_Name", "ID"), sep = "_")
head(df_2020)
```

```
##   First_Name Last_Name ID assists bonus bps clean_sheets creativity element
## 1   Aaron    Cresswell 376      0     0  7          0          1.5      376
## 2   Aaron      Lennon 430      0     0  3          0          0.0      430
## 3   Aaron      Mooy  516      0     0  0          0          0.0      516
## 4   Aaron    Ramsdale 494      0     0 11          0          0.0      494
## 5   Aaron Wan-Bissaka 122      0     2 34          1         16.1      122
## 6 Abdoulaye Doucouré 368      0     0  7          0         23.1      368
##   fixture goals_conceded goals_scored ict_index influence      kickoff_time
## 1      8              5              0      1.1      9.0 2019-08-10T11:30:00Z
## 2      3              0              0      0.2      2.0 2019-08-10T14:00:00Z
## 3      7              0              0      0.0      0.0 2019-08-10T14:00:00Z
## 4      2              1              0      1.0      9.8 2019-08-10T14:00:00Z
## 5      9              0              0      4.9     30.4 2019-08-11T15:30:00Z
## 6      7              3              0      3.0      4.4 2019-08-10T14:00:00Z
##   minutes opponent_team own_goals penalties_missed penalties_saved red_cards
## 1      90          11         0              0              0          0
## 2       6          16         0              0              0          0
## 3       0          18         0              0              0          0
## 4      90          15         0              0              0          0
## 5      90           6         0              0              0          0
## 6      90           4         1              0              0          0
##   round saves selected team_a_score team_h_score threat total_points
## 1     1     0    23399           5           0     0          0
## 2     1     0     8105           0           3     0          1
```

```
## 3      1      0    16261          3          0          0          0
## 4      1      2     3091          1          1          0          2
## 5      1      0   1879259          0          4          2          8
## 6      1      0    200815          3          0          2          0
##   transfers_balance transfers_in transfers_out value was_home yellow_cards GW
## 1              0              0              0   50      True           0  1
## 2              0              0              0   50      True           0  1
## 3              0              0              0   50     False           0  1
## 4              0              0              0   45      True           0  1
## 5              0              0              0   55      True           0  1
## 6              0              0              0   60      True           0  1
```

Creating a new first column

```
df_2020 <- df_2020 %>%
  select(-ID) %>%
  unite("player_name", c("First_Name", "Last_Name"), sep = " ")
head(df_2020)
```

```
##           player_name assists bonus bps clean_sheets creativity element fixture
## 1      Aaron Cresswell      0     0  7              0          1.5    376      8
## 2      Aaron Lennon      0     0  3              0          0.0    430      3
## 3      Aaron Mooy      0     0  0              0          0.0    516      7
## 4      Aaron Ramsdale      0     0 11              0          0.0    494      2
## 5      Aaron Wan-Bissaka      0     2 34              1         16.1    122      9
## 6      Abdoulaye Doucouré      0     0  7              0         23.1    368      7
##   goals_conceded goals_scored ict_index influence      kickoff_time minutes
## 1              5              0        1.1        9.0 2019-08-10T11:30:00Z    90
## 2              0              0        0.2        2.0 2019-08-10T14:00:00Z     6
## 3              0              0        0.0        0.0 2019-08-10T14:00:00Z     0
## 4              1              0        1.0        9.8 2019-08-10T14:00:00Z    90
## 5              0              0        4.9       30.4 2019-08-11T15:30:00Z    90
## 6              3              0        3.0        4.4 2019-08-10T14:00:00Z    90
##   opponent_team own_goals penalties_missed penalties_saved red_cards round
## 1             11          0                0                0          0      1
## 2             16          0                0                0          0      1
## 3             18          0                0                0          0      1
## 4             15          0                0                0          0      1
## 5              6          0                0                0          0      1
## 6              4          1                0                0          0      1
##   saves selected team_a_score team_h_score threat total_points
## 1      0    23399              5              0      0            0
## 2      0     8105              0              3      0            1
## 3      0    16261              3              0      0            0
## 4      2     3091              1              1      0            2
## 5      0   1879259              0              4      2            8
## 6      0    200815              3              0      2            0
##   transfers_balance transfers_in transfers_out value was_home yellow_cards GW
## 1              0              0              0   50      True           0  1
## 2              0              0              0   50      True           0  1
## 3              0              0              0   50     False           0  1
## 4              0              0              0   45      True           0  1
## 5              0              0              0   55      True           0  1
## 6              0              0              0   60      True           0  1
```

2021 dataset

```
df_2021 <- read.csv('https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/master/data/2020-21')
head(df_2021)
```

```
##           name position      team xP assists bonus bps clean_sheets
## 1   Aaron Connolly    FWD   Brighton 0.5      0    0  -3           0
## 2   Aaron Cresswell   DEF   West Ham 2.1      0    0  11           0
## 3   Aaron Mooy        MID   Brighton 0.0      0    0   0           0
## 4   Aaron Ramsdale    GK Sheffield Utd 2.5      0    0  12           0
## 5 Abdoulaye Doucouré  MID   Everton 1.3      0    0  20           1
## 6  Aboubakar Kamara   MID   Fulham 0.4      0    0  -2           0
##  creativity element fixture goals_conceded goals_scored ict_index influence
## 1      0.3      78      7           2           0      3.4      1.2
## 2     11.2     435      6           2           0      2.2     10.4
## 3      0.0      60      7           0           0      0.0      0.0
## 4      0.0     483      8           2           0      1.8     18.2
## 5     44.6     512      4           0           0      6.9     20.4
## 6      1.8     190      2           3           0      0.2      0.0
##  kickoff_time minutes opponent_team own_goals penalties_missed
## 1 2020-09-14T19:15:00Z      45           5           0           0
## 2 2020-09-12T19:00:00Z      90          14           0           0
## 3 2020-09-14T19:15:00Z       0           5           0           0
## 4 2020-09-14T17:00:00Z      90          20           0           0
## 5 2020-09-13T15:30:00Z      90          17           0           0
## 6 2020-09-12T11:30:00Z      62           1           0           0
##  penalties_saved red_cards round saves selected team_a_score team_h_score
## 1           0           0      1      0      32205           3           1
## 2           0           0      1      0       70232           2           0
## 3           0           0      1      0        2010           3           1
## 4           0           0      1      2     462939           2           0
## 5           0           0      1      0     143040           1           0
## 6           0           0      1      0        6711           3           0
##  threat total_points transfers_balance transfers_in transfers_out value
## 1     32           1           0           0           0      55
## 2      0           1           0           0           0      50
## 3      0           0           0           0           0      50
## 4      0           1           0           0           0      50
## 5      4           3           0           0           0      55
## 6      5           2           0           0           0      50
##  was_home yellow_cards GW
## 1     True           0      1
## 2     True           0      1
## 3     True           0      1
## 4     True           0      1
## 5    False           0      1
## 6     True           0      1
```

Season info for 2019/20

```
df_2020s <- read.csv('https://raw.githubusercontent.com/ewenme/squads/master/data/2019/premier_league.csv')
head(df_2020s)
```

```
##      player_name    position date_of_birth age nationality    club_name
## 1      Ederson    Goalkeeper   1993-08-17  25      Brazil Manchester City
## 2 Arijanet Muric    Goalkeeper   1998-11-07  20      Kosovo Manchester City
## 3  Claudio Bravo    Goalkeeper   1983-04-13  36      Chile Manchester City
## 4   Scott Carson    Goalkeeper   1985-09-03  33      England Manchester City
## 5 Aymeric Laporte Centre-Back   1994-05-27  25      France Manchester City
## 6   John Stones    Centre-Back   1994-05-28  25      England Manchester City
##      league_name    season year
## 1 Premier League 2019/2020 2019
## 2 Premier League 2019/2020 2019
## 3 Premier League 2019/2020 2019
## 4 Premier League 2019/2020 2019
## 5 Premier League 2019/2020 2019
## 6 Premier League 2019/2020 2019
```

```
write.csv(df_2020s, "~/DSI-SRP1/season2020.csv")
```

FPL data for 2019/20

```
write.csv(df_2020, "FPL_data_2019_20.csv")
```

Combine 2019/20 datasets

```
df_2020combined <- inner_join(df_2020, df_2020s, by = 'player_name')
head(df_2020combined)
```

```
##      player_name assists bonus bps clean_sheets creativity element fixture
## 1   Aaron Cresswell      0    0   7           0         1.5      376      8
## 2   Aaron Lennon      0    0   3           0         0.0      430      3
## 3   Aaron Mooy      0    0   0           0         0.0      516      7
## 4   Aaron Ramsdale      0    0  11           0         0.0      494      2
## 5   Aaron Wan-Bissaka      0    2  34           1        16.1      122      9
## 6 Abdoulaye Doucouré      0    0   7           0        23.1      368      7
##      goals_conceded goals_scored ict_index influence      kickoff_time minutes
## 1           5           0         1.1         9.0 2019-08-10T11:30:00Z      90
## 2           0           0         0.2         2.0 2019-08-10T14:00:00Z       6
## 3           0           0         0.0         0.0 2019-08-10T14:00:00Z       0
## 4           1           0         1.0         9.8 2019-08-10T14:00:00Z      90
## 5           0           0         4.9        30.4 2019-08-11T15:30:00Z      90
## 6           3           0         3.0         4.4 2019-08-10T14:00:00Z      90
##      opponent_team own_goals penalties_missed penalties_saved red_cards round
## 1           11           0           0           0           0           1
## 2           16           0           0           0           0           1
## 3           18           0           0           0           0           1
## 4           15           0           0           0           0           1
## 5            6           0           0           0           0           1
## 6            4           1           0           0           0           1
##      saves selected team_a_score team_h_score threat total_points
## 1      0    23399           5           0      0           0
## 2      0     8105           0           3      0           1
## 3      0    16261           3           0      0           0
## 4      2     3091           1           1      0           2
```

```
## 5      0 1879259      0      4      2      8
## 6      0  200815      3      0      2      0
##   transfers_balance transfers_in transfers_out value was_home yellow_cards GW
## 1              0              0              0   50      True           0  1
## 2              0              0              0   50      True           0  1
## 3              0              0              0   50     False           0  1
## 4              0              0              0   45      True           0  1
## 5              0              0              0   55      True           0  1
## 6              0              0              0   60      True           0  1
##           position date_of_birth age nationality      club_name
## 1      Left-Back   1989-12-15  29    England      West Ham United
## 2      Right Winger 1987-04-16  32    England      Burnley FC
## 3 Central Midfield 1990-09-15  28    Australia Brighton & Hove Albion
## 4      Goalkeeper  1998-05-14  21    England      AFC Bournemouth
## 5      Right-Back  1997-11-26  21    England      Manchester United
## 6 Central Midfield 1993-01-01  26    France       Watford FC
##   league_name      season year
## 1 Premier League 2019/2020 2019
## 2 Premier League 2019/2020 2019
## 3 Premier League 2019/2020 2019
## 4 Premier League 2019/2020 2019
## 5 Premier League 2019/2020 2019
## 6 Premier League 2019/2020 2019
```

Position Indexer function for 2019/20.

```
position_indexer <- function(df) {
  df %>%
    mutate(position_index = if_else(position %in% c("Goalkeeper"),
1,if_else(position %in% c("Centre-Back", "Left-Back", "Right-Back"),
2,if_else(position %in% c("Defensive Midfield", "Central Midfield", "Left Midfield",
"Right Midfield", "Midfielder"),
3,if_else(position %in% c("Left Winger","Right Winger","Attacking Midfield"),
4,if_else(position %in% c("Forward", "Second Striker", "Centre-Forward"),5,0))))))
}
df_2020c <- position_indexer(df_2020combined)
head(df_2020c)
```

```
##           player_name assists bonus bps clean_sheets creativity element fixture
## 1      Aaron Cresswell      0      0   7           0          1.5       376       8
## 2      Aaron Lennon      0      0   3           0          0.0       430       3
## 3      Aaron Mooy      0      0   0           0          0.0       516       7
## 4      Aaron Ramsdale      0      0  11           0          0.0       494       2
## 5      Aaron Wan-Bissaka      0      2  34           1         16.1       122       9
## 6 Abdoulaye Doucouré      0      0   7           0         23.1       368       7
##   goals_conceded goals_scored ict_index influence      kickoff_time minutes
## 1              5              0      1.1        9.0 2019-08-10T11:30:00Z      90
## 2              0              0      0.2        2.0 2019-08-10T14:00:00Z       6
## 3              0              0      0.0        0.0 2019-08-10T14:00:00Z       0
## 4              1              0      1.0        9.8 2019-08-10T14:00:00Z      90
## 5              0              0      4.9       30.4 2019-08-11T15:30:00Z      90
## 6              3              0      3.0        4.4 2019-08-10T14:00:00Z      90
##   opponent_team own_goals penalties_missed penalties_saved red_cards round
```

```
## 1      11      0      0      0      0      1
## 2      16      0      0      0      0      1
## 3      18      0      0      0      0      1
## 4      15      0      0      0      0      1
## 5       6      0      0      0      0      1
## 6       4      1      0      0      0      1
##      saves selected team_a_score team_h_score threat total_points
## 1      0    23399           5      0      0      0
## 2      0     8105           0      3      0      1
## 3      0    16261           3      0      0      0
## 4      2     3091           1      1      0      2
## 5      0   1879259          0      4      2      8
## 6      0    200815          3      0      2      0
##      transfers_balance transfers_in transfers_out value was_home yellow_cards GW
## 1           0           0           0    50     True           0 1
## 2           0           0           0    50     True           0 1
## 3           0           0           0    50    False           0 1
## 4           0           0           0    45     True           0 1
## 5           0           0           0    55     True           0 1
## 6           0           0           0    60     True           0 1
##      position date_of_birth age nationality      club_name
## 1    Left-Back   1989-12-15  29    England    West Ham United
## 2    Right Winger 1987-04-16  32    England    Burnley FC
## 3 Central Midfield 1990-09-15  28  Australia Brighton & Hove Albion
## 4    Goalkeeper   1998-05-14  21    England    AFC Bournemouth
## 5    Right-Back   1997-11-26  21    England    Manchester United
## 6 Central Midfield 1993-01-01  26    France      Watford FC
##      league_name  season year position_index
## 1 Premier League 2019/2020 2019           2
## 2 Premier League 2019/2020 2019           4
## 3 Premier League 2019/2020 2019           3
## 4 Premier League 2019/2020 2019           1
## 5 Premier League 2019/2020 2019           2
## 6 Premier League 2019/2020 2019           3
```

Position indexer for 2020/21

```
position_indexer <- function(df) {
  df %>%
    mutate(position_index = if_else(position %in% c("GK"),
  1,if_else(position %in% c("DEF"),
  2,if_else(position %in% c("MID"),
  3,if_else(position %in% c("FWD"),
  4,0))))))
}
df_2021c <- position_indexer(df_2021)
head(df_2021c)
```

```
##      name position      team xP assists bonus bps clean_sheets
## 1  Aaron Connolly    FWD  Brighton 0.5      0      0  -3           0
## 2  Aaron Cresswell    DEF  West Ham 2.1      0      0  11           0
## 3   Aaron Mooy       MID  Brighton 0.0      0      0   0           0
## 4  Aaron Ramsdale    GK Sheffield Utd 2.5      0      0  12           0
```

```
## 5 Abdoulaye Doucouré      MID      Everton 1.3      0      0      20      1
## 6 Aboubakar Kamara        MID      Fulham 0.4      0      0      -2      0
##      creativity element fixture goals_conceded goals_scored ict_index influence
## 1      0.3      78      7      2      0      3.4      1.2
## 2      11.2     435     6      2      0      2.2     10.4
## 3      0.0      60      7      0      0      0.0      0.0
## 4      0.0     483     8      2      0      1.8     18.2
## 5     44.6     512     4      0      0      6.9     20.4
## 6      1.8     190     2      3      0      0.2      0.0
##      kickoff_time minutes opponent_team own_goals penalties_missed
## 1 2020-09-14T19:15:00Z      45      5      0      0
## 2 2020-09-12T19:00:00Z      90     14      0      0
## 3 2020-09-14T19:15:00Z      0      5      0      0
## 4 2020-09-14T17:00:00Z      90     20      0      0
## 5 2020-09-13T15:30:00Z      90     17      0      0
## 6 2020-09-12T11:30:00Z      62      1      0      0
##      penalties_saved red_cards round saves selected team_a_score team_h_score
## 1      0      0      1      0      32205      3      1
## 2      0      0      1      0      70232      2      0
## 3      0      0      1      0      2010      3      1
## 4      0      0      1      2     462939      2      0
## 5      0      0      1      0     143040      1      0
## 6      0      0      1      0      6711      3      0
##      threat total_points transfers_balance transfers_in transfers_out value
## 1     32      1      0      0      0      55
## 2      0      1      0      0      0      50
## 3      0      0      0      0      0      50
## 4      0      1      0      0      0      50
## 5      4      3      0      0      0      55
## 6      5      2      0      0      0      50
##      was_home yellow_cards GW position_index
## 1      True      0      1      4
## 2      True      0      1      2
## 3      True      0      1      3
## 4      True      0      1      1
## 5     False      0      1      3
## 6      True      0      1      3
```

Cumulative function

```
accumulated <- function(df) {
  df %>%
    mutate(n = 1) %>%
    group_by(player_name, club_name) %>%
    mutate(goals_scored = cumsum(goals_scored) - goals_scored,
           assists = cumsum(assists) - assists,
           ict_index = cumsum(ict_index) - ict_index ,
           goals_conceded = cumsum(goals_conceded) - goals_conceded,
           minutes = cumsum(minutes) - minutes,
           own_goals = cumsum(own_goals) - own_goals,
           total_points = cumsum(total_points), bps = cumsum(bps)-bps,
           clean_sheets = cumsum(clean_sheets)-clean_sheets,
           bonus = cumsum(bonus)- bonus) %>%
    select(player_name, club_name, position, position_index, goals_scored,
```

```

      assists, ict_index, goals_conceded, minutes, own_goals,
      total_points, bps, bonus, clean_sheets, GW, season, kickoff_time) %>%
    filter(ict_index > 0)
  }
df_2020f <- accumulated(df_2020c)
head(df_2020f)

```

```

## # A tibble: 6 x 17
## # Groups:   player_name, club_name [6]
##   player_name club_name position position_index goals_scored assists ict_index
##   <chr>        <fct>    <fct>          <dbl>         <int>   <int>   <dbl>
## 1 Aaron Cress~ West Ham ~ Left-Ba~         2           0       0       1.1
## 2 Aaron Lennon Burnley FC Right W~         4           0       0       0.2
## 3 Aaron Ramsd~ AFC Bourn~ Goalkee~         1           0       0        1
## 4 Aaron Wan-B~ Mancheste~ Right-B~         2           0       0       4.9
## 5 Abdoulaye D~ Watford FC Central~         3           0       0      3.00
## 6 Adam Smith   AFC Bourn~ Right-B~         2           0       0      1.40
## # ... with 10 more variables: goals_conceded <int>, minutes <int>,
## #   own_goals <int>, total_points <int>, bps <int>, bonus <int>,
## #   clean_sheets <int>, GW <int>, season <fct>, kickoff_time <fct>

```

Cumulative for 2019/20

```

accumulated2 <- function(df) {
  df %>%
    mutate(n = 1) %>%
    group_by(name, team) %>%
    mutate(goals_scored = cumsum(goals_scored) - goals_scored,
           assists = cumsum(assists) - assists,
           ict_index = cumsum(ict_index) - ict_index ,
           goals_conceded = cumsum(goals_conceded) - goals_conceded,
           minutes = cumsum(minutes) - minutes,
           own_goals = cumsum(own_goals) - own_goals,
           total_points = cumsum(total_points), bps = cumsum(bps)-bps,
           clean_sheets = cumsum(clean_sheets)-clean_sheets,
           bonus = cumsum(bonus)- bonus) %>%
    select(name, team, position, position_index, goals_scored,
           assists, ict_index, goals_conceded, minutes, own_goals,
           total_points, bps, bonus, clean_sheets, GW, kickoff_time) %>%
    filter(ict_index > 0)
}
df_2021f <- accumulated2(df_2021c)
head(df_2021f)

```

```

## # A tibble: 6 x 16
## # Groups:   name, team [6]
##   name      team      position position_index goals_scored assists ict_index
##   <fct>    <fct>    <fct>          <dbl>         <int>   <int>   <dbl>
## 1 Aaron Connol~ Brighton FWD         4           0       0       3.4
## 2 Aaron Cressw~ West Ham DEF         2           0       0       2.2
## 3 Aaron Ramsda~ Sheffield~ GK         1           0       0       1.8
## 4 Abdoulaye Do~ Everton  MID         3           0       0       6.9

```



```
## 5 Aboubakar Ka~ Fulham    MID           3           0           0       0.200
## 6 Adama Traoré Wolves    MID           3           0           0       3.2
## # ... with 9 more variables: goals_conceded <int>, minutes <int>,
## #   own_goals <int>, total_points <int>, bps <int>, bonus <int>,
## #   clean_sheets <int>, GW <int>, kickoff_time <fct>
```

Write datasets to csv files

```
write.csv(df_2020c, "~/DSI-SRP1/FPL_2019_20.csv")
write.csv(df_2021c, "~/DSI-SRP1/FPL_2020_21.csv")
```