Weekly FPL Points Dataset Creator

This document is used to create new players' datasets that describe weekly performance in the Premier league from 2016-2019

```
library(tidyverse)
```

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

## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidyr)
```

Loading the datasets

```
s17_1 <- read.csv("~/DSI-SRP1/season2017.csv", encoding = "UTF-8")
s17_2 <- read.csv("~/DSI-SRP1/FPL_2016.csv")
s18_1 <- read.csv("~/DSI-SRP1/season2018.csv", encoding = "UTF-8")
s18_2 <- read.csv("~/DSI-SRP1/FPL_2017.csv")
s19_1 <- read.csv("~/DSI-SRP1/season2019.csv", encoding = "UTF-8")
s19_2 <- read.csv("https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/master/data/2018-19/
```

Function for transforming and merging datsets

```
fpl_merge <- function(df, df2) {
  df2 <- df2 %>%
    rename(c(player_name = name))
  df1 <- df %>%
    select(player_name, club_name, position_index, position, season)
  df1$player_name <- as.character(df1$player_name)
  df2$player_name <- as.character(df2$player_name)
  df4 <- inner_join(df2, df1, by = "player_name")
  return(df4)
}
```

Premier League 2016/17

```
s17_2 <- s17_2 %>%
  separate(name, into = c("first", "last"), "_")
s17_2 <- s17_2 %>%
  unite(name, first,last, sep = " ")
s17_3 <- fpl_merge(s17_1,s17_2)
nrow(s17_3)
```

```
## [1] 22065
```

```
write_csv(s17_3, "FPL_2016_17.csv")
```

Premier League 2017/18

```
s18_2 <- s18_2 %>%
  separate(name, into = c("first", "last"), "_")
s18_2 <- s18_2 %>%
  unite(name, first,last, sep = " ")
s18_3 <- fpl_merge(s18_1,s18_2)
nrow(s18_3)
```

```
## [1] 20186
```

```
write_csv(s18_3, "FPL_2017_18.csv")
```

Premier League 2018/19

```
s19_2 <- s19_2 %>%
  separate(name, into = c("first", "last", "id"), "_")
s19_2 <- s19_2 %>%
 unite(name, first:last, sep = " ")
s19_2 <- s19_2 %>%
  select(-id)
s19_3 <- fpl_merge(s19_1,s19_2)
nrow(s19_3)
```

```
## [1] 16979
```

```
write_csv(s19_3, "FPL_2018_19.csv")
```

Loading newly made datasets

```
data17 <- read.csv("~/DSI-SRP1/FPL_2016_17.csv", encoding = "UTF-8")
data18 <- read.csv("~/DSI-SRP1/FPL_2017_18.csv", encoding = "UTF-8")
data19 <- read.csv("~/DSI-SRP1/FPL_2018_19.csv", encoding = "UTF-8")
```

Function for cumulatively adding specific columns

```
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,
           GW = cumsum(n)) %>%
    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) %>%
    filter(ict_index > 0)
}
```

Applying the functions on the datasets

```
year17 <- accumulated(data17)
year18 <- accumulated(data18)
year19 <- accumulated(data19)
```

Cleaning datasets

function for removing duplicate players' record

```
rem_dup <- function(df) {
  df %>%
    group_by(GW) %>%
    distinct(player_name, .keep_all = TRUE)
}
```

```
year19 <- year19 %>%
  filter(player_name != "Danny Ward")
year17 <- rem_dup(year17)
year18 <- rem_dup(year18)
year19 <- rem_dup(year19)
```

Number of rows in each dataset

```
nrow(year17)
```

```
## [1] 15348
```

```
nrow(year18)
```

```
## [1] 14738
```

```
nrow(year19)
```

```
## [1] 13413
```

Saving the newly datasets to csv files

```
write.csv(year17, "FPL_2016_17_new.csv")
write.csv(year18, "FPL_2017_18_new.csv")
write.csv(year19, "FPL_2018_19_new.csv")
```

Variable selector

```
var_sel <- function(df) {
  df %>%
    select(player_name, club_name, position, position_index, goals_scored,
           assists, ict_index, goals_conceded, minutes, own_goals,
           total_points, bps, bonus, clean_sheets, season)
}
```

Creating a dataset that combines the three seasons and cumulatively adds up certain rows

Creating a new accumulated function

```
accumulated2 <- function(df) {
  df %>%
    mutate(n = 1) %>%
    group_by(player_name,club_name) %>%
    mutate(goals_scored =cumsum(goals_scored),
           assists =cumsum(assists),
           ict_index =cumsum(ict_index),
           goals_conceded =cumsum(goals_conceded),
           minutes =cumsum(minutes),
           own_goals =cumsum(own_goals),
```

```
            total_points =cumsum(total_points), bps =cumsum(bps),
            clean_sheets =cumsum(clean_sheets),
            bonus = cumsum(bonus),
            GW = cumsum(n)) %>%
    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) %>%
    filter(ict_index > 0)
}
```

Selecting important rows

```
data17 <- var_sel(data17)
data18 <- var_sel(data18)
data19 <- var_sel(data19)
year3 <- rbind(data17, data18,data19)
year3 <- accumulated2(year3)
```

Cleaning the cumulated season dataset

```
year3 <- year3 %>%
  filter(player_name != "Danny Ward")
year3 <- rem_dup(year3)
nrow(year3)
```

```
## [1] 46659
```

```
write.csv(year3, "~/DSI-SRP1/FPL_ULTIMATE.csv")
```

2019/20 sample dataset

```
library(readxl)
data <- read_excel("~/DSI-SRP1/2019_20_sample.xlsx")
head(data)
```

```
## # A tibble: 6 x 8
##   player_name goals_scored assists minutes position_index clean_sheets
##   <chr>              <dbl>   <dbl>   <dbl>          <dbl>        <dbl>
## 1 Alisson                0       1    2363              1           13
## 2 Bravo                  0       0     347              1            1
## 3 de Gea                 0       0    3240              1           12
## 4 Caballero              0       0     360              1            0
## 5 Gunn                   0       0     900              1            5
## 6 Pope                   0       0    3330              1           15
## # ... with 2 more variables: total_points <dbl>, fpl_to_game <dbl>
```

```r
write.csv(data, "~/DSI-SRP1/2019_20_sample.csv")
```