HW_7

GA

2022-10-31

Exercise 1

Extract all matches from the England Premier League and calculate the mean team difference (average of home team goals minus away team goals) each day in the "2010/2011" season. Plot this proportion against time. (hint: you'll need separate date and time. You'll also need to use before you plot parse_date()).

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(RSQLite)
```

```
con <- dbConnect(drv = SQLite(), dbname = "/Users/fam/Desktop/Desktop - FENTAW's MacBook Air/American_U
```

```
dbListTables(con)
```

```
## [1] "Country"           "League"            "Match"
## [4] "Player"            "Player_Attributes" "Team"
## [7] "Team_Attributes"   "sqlite_sequence"
```

```
 Match_db   <- tbl(con, "Match")
```

```
Match_db %>%
  select(league_id, season, date, home_team_goal, away_team_goal) %>%
  filter(league_id == 1729, season == "2010/2011") %>%
  collect() ->
      subMatch

    subMatch %>%
      separate(col = "date", into = c("date", "time"), sep = " ") %>%
      select(-time) %>%
      group_by(date) %>%
      summarize(mean_diff = mean(home_team_goal - away_team_goal)) %>%
      mutate(date = parse_date(x = date, format = "%Y-%m-%d")) ->
      ave_diff

    ggplot(data = ave_diff, mapping = aes(x = date, y = mean_diff)) +
      geom_line() +
```

```
    xlab("Date") +
    ylab("Mean Difference") +
    theme_bw()
```