HOME ADVANTAGE

Narek Sahakyan

October 11, 2019

```r
calculate_points <- function(df,df1) {
    temp <- data.frame(df[,1], df$HW*3 + df$HD, df$GFH, df$GAH, df1$AW*3 + df1$AD, df1$GFA, df1$GAA)
    rownames(temp) <- NULL
    colnames(temp) <- c("TEAM","HOME.POINTS","GF.HOME","GA.HOME","AWAY.POINTS","GF.AWAY","GA.AWAY")
    temp$P <- temp$HOME.POINTS + temp$AWAY.POINTS
    temp$PD <- temp$HOME.POINTS - temp$AWAY.POINTS
    temp$GFHAD <- temp$GF.HOME - temp$GF.AWAY #The difference of goals scored at home and and away
    temp$GAHAD <- temp$GA.HOME - temp$GA.AWAY ##The difference of goals allowed at home and and away
    return(temp <- temp %>%
      arrange(desc(P)))
}

teams_ha_points <- function(data = f_data_sm){
    df <- data %>%
    group_by(HOMETEAM) %>% summarise(HW = sum(FTR == "H"),
                                     HD = sum(FTR == "D"),
                                     GFH = sum(FTHG),
                                     GAH = sum(FTAG))

    df1 <- data %>%
    group_by(AWAYTEAM) %>% summarise(AW = sum(FTR == "A"),
                                     AD = sum(FTR == "D"),
                                     GFA = sum(FTAG),
                                     GAA = sum(FTHG))
    result <- calculate_points(df, df1)
    return(result)
}

team_ha_adv_filtered <- function(data = f_data_sm, country, season) {
  result <- data %>%
    filter(COUNTRY == country, SEASON == season) %>%
    teams_ha_points()
  result$SEASON = season
  result$COUNTRY = country
  cols <- colnames(result)
  len <- length(cols)
  return(result[c("SEASON","COUNTRY",colnames(result)[1:(len-2)])])
}


home_advantage <- function(data){

  data_home_wins <- data %>%
    group_by(SEASON) %>%
    filter(FTR == "H") %>%
    summarise(POINTS = 3*n())
```

```r
  data_home_draws <- data %>%
    group_by(SEASON) %>%
    filter(FTR == "D") %>%
    summarise(POINTS = n())

  data_home_points <- data.frame(SEASON = data_home_wins$SEASON,
                                 HOME_POINTS = data_home_wins$POINTS + data_home_draws$POINTS)

  data_away_wins <- data %>%
    group_by(SEASON) %>%
    filter(FTR == "A") %>%
    summarise(POINTS = 3*n())

  #The draw points for home and away teams are the same

  data_away_points <- data.frame(SEASON = data_away_wins$SEASON,
                                 AWAY_POINTS = data_away_wins$POINTS + data_home_draws$POINTS)

  data_points <- data_home_points
  data_points$AWAY_POINTS <- data_away_points$AWAY_POINTS

  return(data_points <- data_points %>%
    group_by(SEASON) %>%
    mutate(DIFF = (HOME_POINTS - AWAY_POINTS),
           MATCHES = (HOME_POINTS - AWAY_POINTS)/3 ) %>%
    arrange(desc(DIFF)))
}


plot_ha <- function(data) {
  data %>%
  ggplot(aes(x = reorder(factor(SEASON), DIFF))) +
  geom_bar(aes(y = DIFF),
           stat = "identity",
           fill = "red") +
  geom_bar(aes(y = MATCHES),
         stat = "identity",
         fill = "blue") +
  coord_flip() +
  labs(x = "SEASON", y = "The difference in points gained by home teams and away teams")
  ggtitle("The average number of how many more games won by home teams")
}

plot_ha_line <- function(data){
  data %>%
  ggplot(aes(x = SEASON, y = DIFF)) + geom_line(color="red")
}

all_leagues_detailed <- f_data_sm %>%
  group_by(SEASON, COUNTRY) %>%
  group_modify(~ teams_ha_points(.x)) %>%
  summarise(MEAN.PD = mean(PD),
            MEAN.GFD = mean(GFHAD),
```

```
              MEAN.GAD = mean(GAHAD)) %>%
    mutate(MEAN.GWH = round(MEAN.PD / 3), #On average how many more games were won at home
           MEAN.GDH = MEAN.GFD - MEAN.GAD,# The average difference of goals in home games
           MEAN.HGWG = MEAN.GDH / MEAN.GWH ) %>% # The average goal difference for each won home game
    arrange(desc(MEAN.PD), desc(MEAN.HGWG))

eredivisie_detailed_ha <- all_leagues_detailed %>%
  filter(COUNTRY == "Netherlands")

teams_ha_2002 <- team_ha_adv_filtered(country = "Netherlands",
                                       season = 2002)
teams_ha_2016 <- team_ha_adv_filtered(country = "Netherlands",
                                       season = 2016)

all_leagues <- f_data_sm %>%
  group_by(COUNTRY) %>%
  group_modify(~home_advantage(.x)) %>%
  arrange(desc(SEASON), desc(DIFF))
```

Eredivisie Home advantage over time in terms of points

```
ered_points <- home_advantage(eredivisie)

ered_points %>%
  ggplot(aes(x = reorder(factor(SEASON), DIFF))) +
  geom_bar(aes(y = DIFF),
           stat = "identity",
           fill = "red") +
  geom_bar(aes(y = MATCHES),
           stat = "identity",
           fill = "blue") +
  coord_flip() +
  labs(x = "SEASON", y = "The difference in points gained by home teams and away teams") +
  ggtitle("The average number of how many more games won by home teams")
```
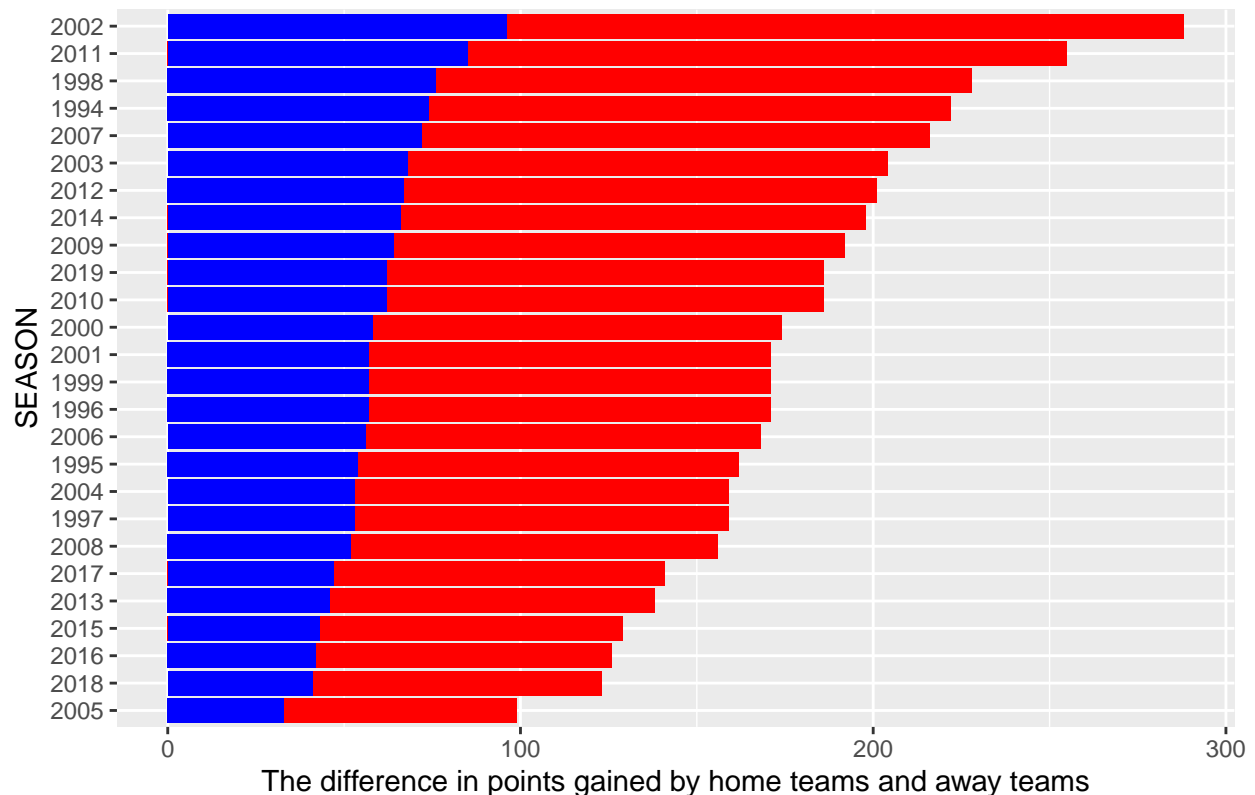
The average number of how many more games won by home teams

The difference in points gained by home teams and away teams

As we can se the biggest difference in the amount of points gained by teams at home and away was in 2002 and the smallest difference was in 2005 If we divide the largest difference in points gained in 2002(288) by 3, we get that on average home team won the game in 96 more matches and if we go further and divide $96/18 = 5.33333$ we get that in that season a team won on average 5 more games at home

Eredivisie Home advantage over time in terms of goals

```r
eredivisie_detailed_ha %>%
  ggplot(aes(x = SEASON)) +
    geom_point(aes(y = MEAN.GFD), col = "red") +
    geom_line(aes(y = MEAN.GFD), col="red") +
      geom_point(aes(y = MEAN.GWH), col = "blue") +
      geom_line(aes(y = MEAN.GWH), col="blue")+
        geom_point(aes(y = MEAN.GAD), col = "orange") +
        geom_line(aes(y = MEAN.GAD), col="orange") +
      labs(x = "SEASON", y = "HOME ADVANTAGE") +
      ggtitle("HOME ADVANTAGE OVER TIME IN TERMS OF POINTS, GF,GA")
```

4

HOME ADVANTAGE OVER TIME IN TERMS OF POINTS, GF,GA
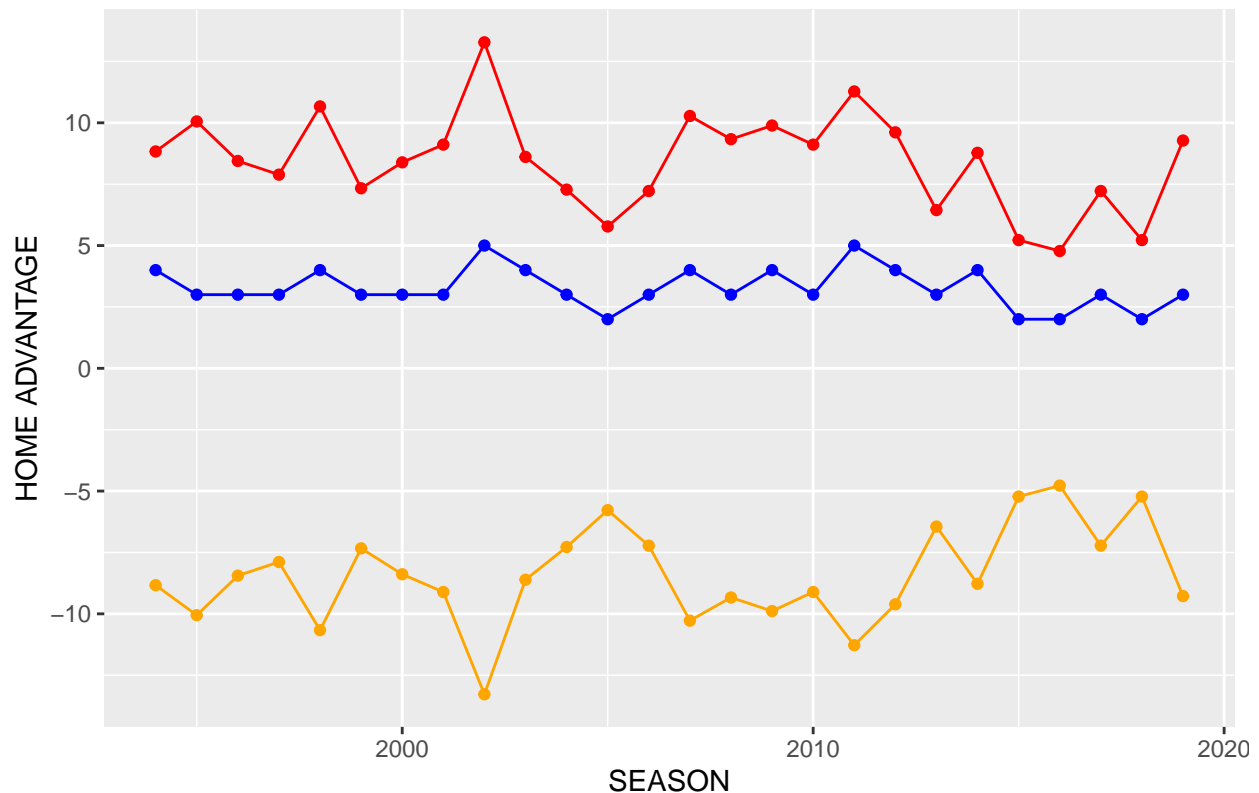
As you can see, the factor home advantage had the highest effect in season 2002. In this season, we can see that the difference in the average number of goals scored at home and away was the highest, leading to the opposite effect in the same difference for goals allowed at home and away. As the goals are the only key factors deciding the game results, we can see that the highest differences in the mentioned fields led to the highest difference in the amount of points gained in home matches and away matches, which divided by 3 can give us the average number of games. As the average number of home advantaged goals is the highest and the allowed goals difference is the lowest, then we can conclude that in season 2002, the home teams won their games in their stadiums, by the highest average goal difference.

The other side of the factor is visible in the season 2015/2016. All of the factors described above had the opposite results. So it means that decrease(increase) in the difference of average goals scored(allowed) at home and away lead to lower effect of home advantage.

Also, the essence of the home advantage is proved by the observations, which show that the difference of average goals scored(allowed) at home and away was always positive(negative). So in general teams score more in home games than in away games and allow more in away games rather than in home games.
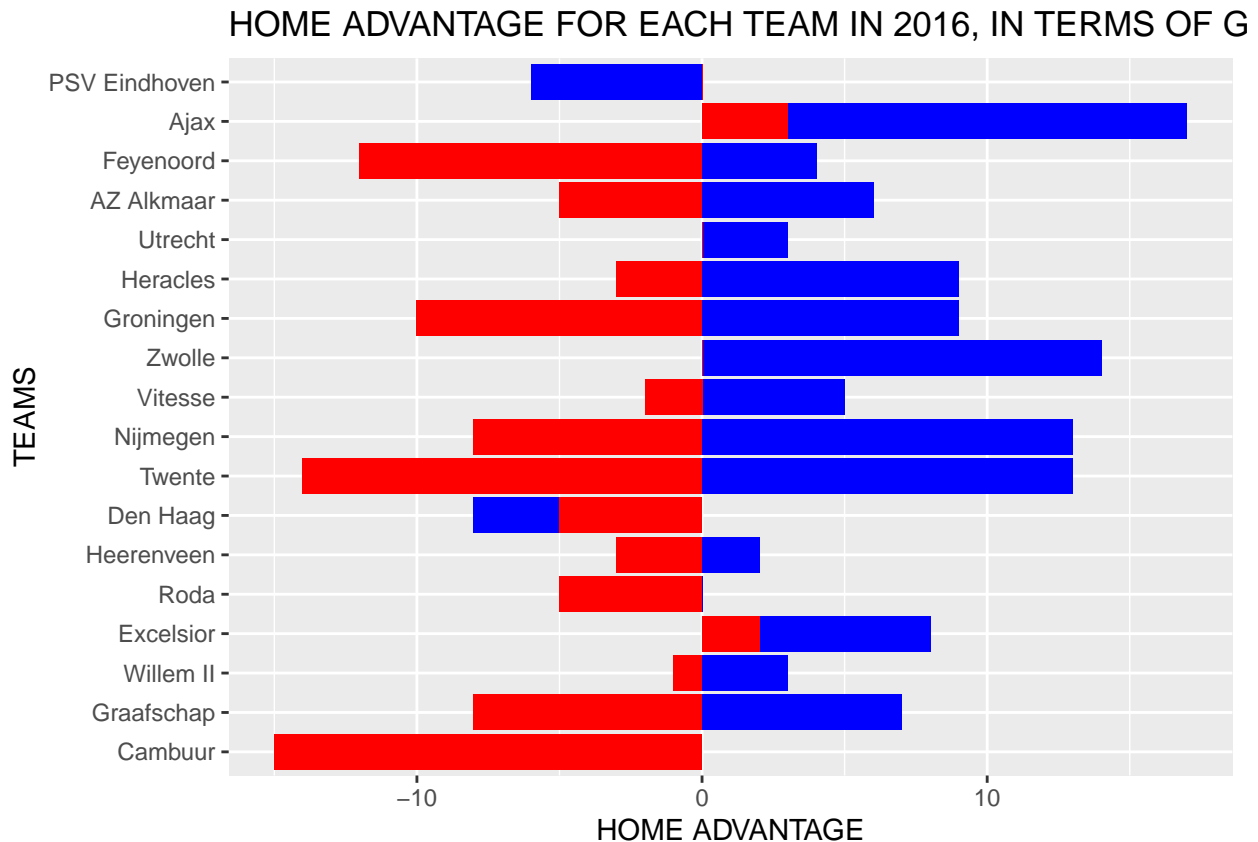
Eredivisie Home advantage for season 2018/19 in terms of teams

```
teams_ha_2016 %>%
   ggplot(aes(x = reorder(TEAM, P))) +
     geom_bar(aes(y = GFHAD),
              stat = "identity",
              fill = "blue") +
       geom_bar(aes(y = GAHAD),
```

```
          stat = "identity",
          fill = "red") +
    coord_flip() +
    labs(x = "TEAMS", y = "HOME ADVANTAGE") +
  ggtitle("HOME ADVANTAGE FOR EACH TEAM IN 2016, IN TERMS OF GF,GA")
```


We already know that the home advantage factor had the lowest effects in the season 2016. It is clearly observable from this plot, as the champion PSV managed to score more goals in away games than in home games. DEN Haag also scored more away goals, but unlike PSV they allowed more goals away, whereas PSV allowed equal number of goals in home and away goals. It is visible that many teams managed to allow the same number of goals in home and away goals. Cambuur was the first in terms of using home advantage for not allowing goals, and Feyenord was the best, overcoming Excelsior only a little bit

```
teams_ha_2002 %>%
    ggplot(aes(x = reorder(TEAM, P))) +
    geom_bar(aes(y = GFHAD),
          stat = "identity",
          fill = "blue") +
      geom_bar(aes(y = GAHAD),
          stat = "identity",
          fill = "red") +
      coord_flip() +
      labs(x = "TEAMS", y = "HOME ADVANTAGE") +
    ggtitle("HOME ADVANTAGE FOR EACH TEAM IN 2002, IN TERMS OF GF,GA")
```

HOME ADVANTAGE FOR EACH TEAM IN 2002, IN TERMS OF G
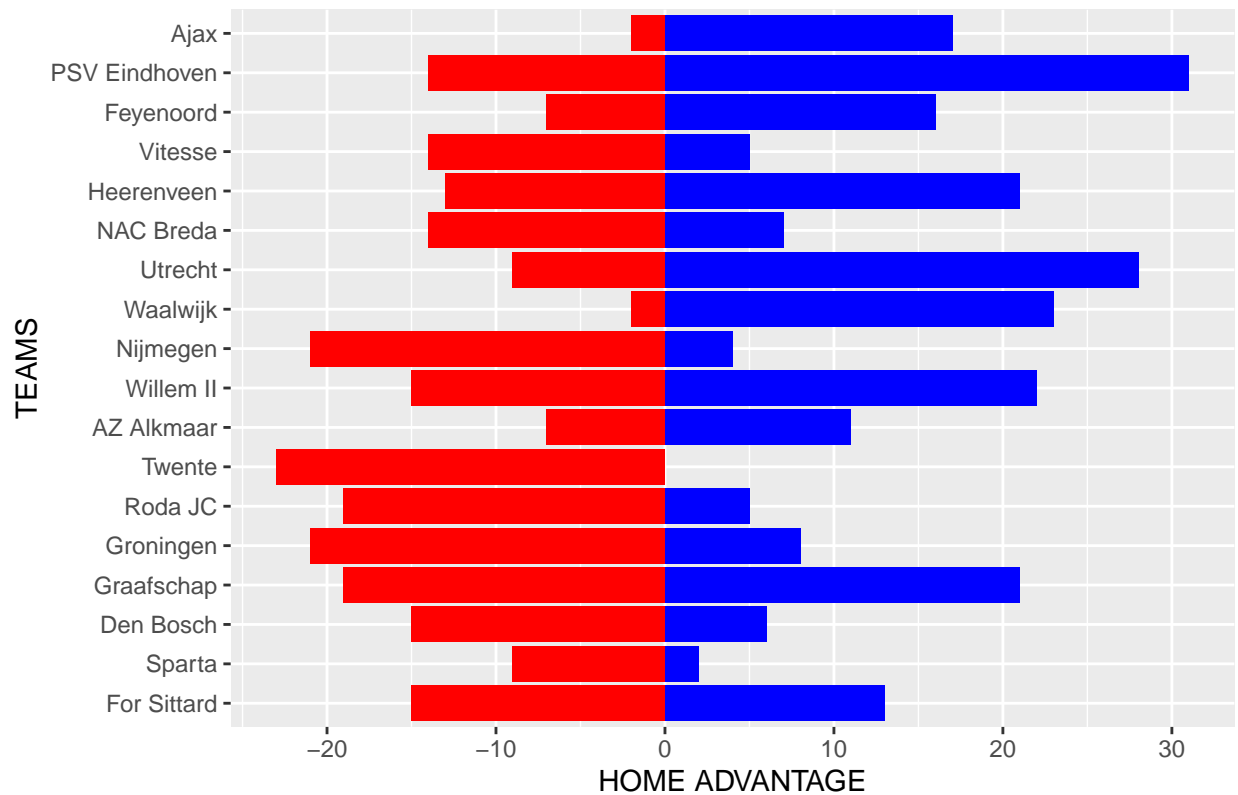
For the season 2001/2002, we see almost completely opposite picture of the 2016 season. Only Twente was not able to use its home advantage in terms of scoring goals, whereas even the relegated teams have positive values in the differences of goals scored at home and away. Ajax, the champion got lower difference in home goals, but due to lower difference in allowed home goals compared to Feyenord, they became a champion. It is interesting that the team Waalwijk was even better in these factors compared to Ajax, but they were only the 8th team of the leageue.

```
plot_ha_line(all_leagues) + facet_grid(rows = vars(COUNTRY),
                                       scales = "free",
                                       space = "free")
```
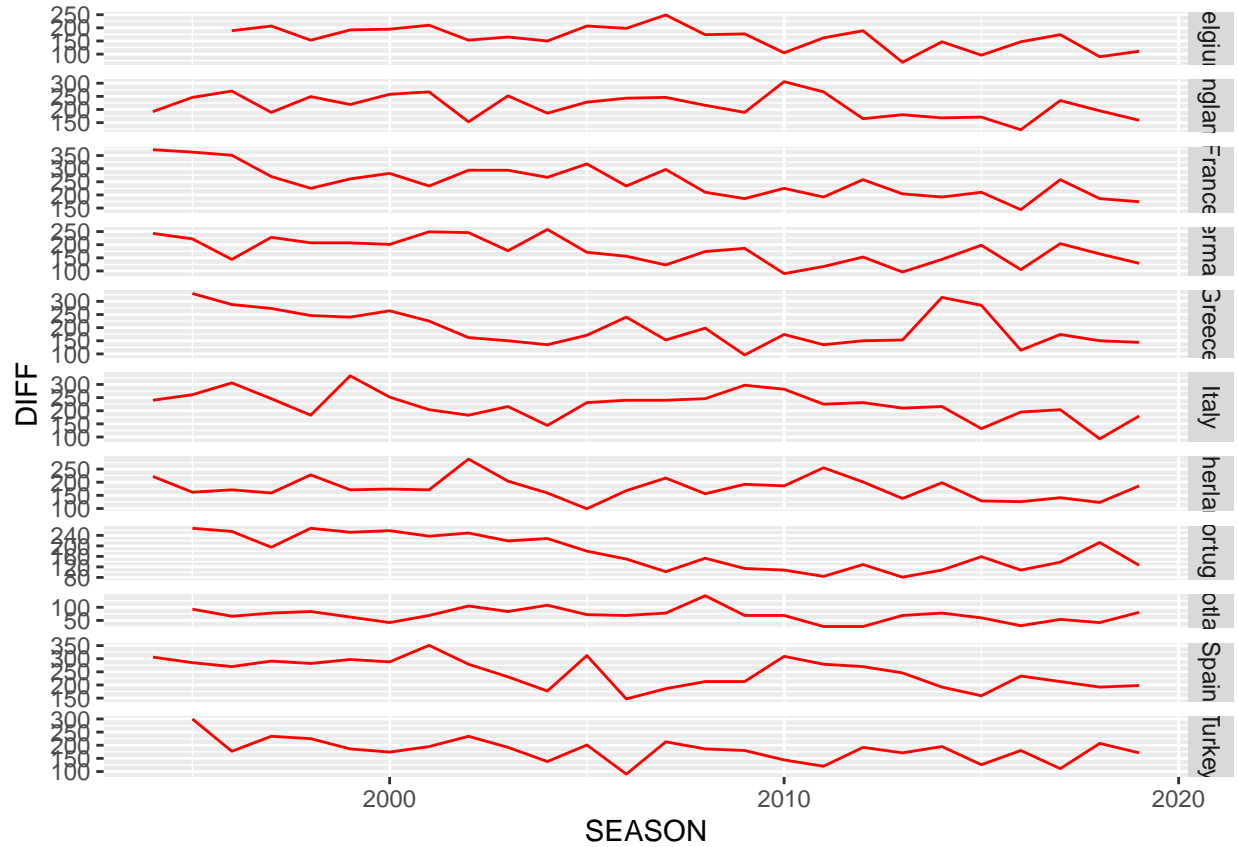
As we can see from the visualization, almost all the leages had a period of signifcant decrease and then increase or increase and then decrease in the factor of home advantage. Portugal league was more or less stable in this factor for a long period of time, but it dropped to its lowest value in around 2008 and started to rise but not significinatly later. The difference in points seems to approach 200 over time. Which divided by 3 gives us around 66 matches on average. For example, Taking as an average sample 34 tours for a league, meaning that contains 612 games. 66/612 =∼ 10%, meaning that the home teams on average win 10% more games in their home stadium.