TOY_1

Paula Moreno Blazquez

Enero 2022

```r
library(tidyr)
library(dplyr)
library(chron)                    # Para CHR_to_Time
library(stringr)                  # Para str_pad
library(lubridate)
library(reshape)
library(tidyselect)

library(corrplot)
library(RColorBrewer)
library("colorspace")
library(graphics)
library(rpart)
```

PARTE 1: Crear Stints

DATA

Se crea una base de datos de juguete random para poder trabajar con un df más pequeño.

```r
equipo1 <- c("Juan","Diego","Maria","Andrea","Carla")
equipo2 <- c("Ignasi","Anna","Gerard","Jose","Paula")
equipo3 <- c("Bella","Gus","Alba","Erik","Kevin")
equipo4 <- c("Emma","Mauri","Berta","Judith","Roger")

toy <- data.frame(
  "id_play" = 1:15,
  "season" = c(rep("S2017",15)),
  "game_code_code" = c(rep("G1",10), rep("G2",5)),
  "quarter" = c(rep(1,3), rep(2,2),rep(3,2),rep(4,3),rep(1,4),rep(2,1)),
  "time" = c('00:00:00','00:03:20','00:05:12','00:11:20','00:15:36','00:21:45','00:27:20','00:31:45','0
  "points_home" =    c(0,0,0,2,2,2,5,5,7,7,0,3,3,5,6),
  "points_away" = c(0,2,2,2,4,4,7,7,9,9,0,0,2,2,2),
  "homeplayer1" = c(rep(equipo1[1],4), rep(equipo1[2],6), rep(equipo2[1],5)),
  "homeplayer2" = c(rep(equipo1[3],10),                   rep(equipo2[2],5)),
  "homeplayer3" = c(rep(equipo1[4],7), rep(equipo1[5],3), rep(equipo2[3],5)),
  "awayplayer1" = c(rep(equipo3[2],10),                   rep(equipo4[2],5)),
  "awayplayer2" = c(rep(equipo3[1],8), rep(equipo3[3],2), rep(equipo4[6],5)),
  "awayplayer3" = c(rep(equipo3[4],10),                   rep(equipo4[4],5))
  )

toy2 <- data.frame(                               #DF2 para comprovar que funciona
```

```r
  "id_play"    = 1:20,
  "season"     = c(rep("S2017",20)),
  "game_code"  = c(rep(1,10), rep(2,10)),
  "quarter"    = c(rep(1,3), rep(2,2),rep(3,2),rep(4,3),rep(1,4),rep(2,6)),
  "time"       = c('00:00:00','00:03:20','00:05:12','00:11:20','00:15:36','00:21:45','00:27:20','00:31:4
                   '00:33:00','00:36:17','00:00:00','00:03:48','00:04:45','00:06:46','00:11:21','00:12:3
                   '00:14:20','00:16:12','00:17:09','00:19:54'),
  "points_home" = c(0,0,0,2,2,2,5,5,7,7,0,2,2,5,5,7,7,7,9,10),
  "points_away" = c(0,2,2,2,4,4,7,7,9,9,0,0,2,2,4,4,6,6,6,6),
  "homeplayer1" = c(rep(equipo1[1],4),  rep(equipo1[2],6),    rep(equipo2[1],7), rep(equipo2[2],3)),
  "homeplayer2" = c(rep(equipo1[3],10),                       rep(equipo2[3],10)),
  "homeplayer3" = c(rep(equipo1[4],7),  rep(equipo1[5],3),    rep(equipo2[4],8), rep(equipo2[5],2)),
  "awayplayer1" = c(rep(equipo2[1],10),                       rep(equipo3[1],10)),
  "awayplayer2" = c(rep(equipo2[2],8),  rep(equipo2[4],2),    rep(equipo3[2],8), rep(equipo3[5],2)),
  "awayplayer3" = c(rep(equipo2[3],10),                       rep(equipo3[3],7), rep(equipo3[4],3))
  )

toy_backup <- toy2

df <- toy2
```

Modificar variable tiempo

Para facilitar los calculos con el tiempo, se va a pasar los mm:ss a segundos.

- CHR_to_Time: Para pasar la variable tiempo que nos llega como chracter a segundos.
- Print_MS: Que nos devolverá los segundos a formato mm:SS (se hará servir más adelante)

```r
CHR_to_Time <- function(x){
  a <- as.POSIXct(x, tz = '', format = "%H:%M:%S", usetz = FALSE)
  tms <- times(format(a, "%H:%M:%S"))
  s <- period_to_seconds(hms(tms))
  return(s)
}

Print_MS <- function(x){
  t <- seconds_to_period(x)
  sprintf('%02d:%02d:%02d', t@hour, minute(t), second(t))
}

df$time <- CHR_to_Time(df$time)
```

Se crea variable 'lineups' que recoge los quintetos de ambos equipos en pista. También se modifica variable 'game_code' para que quede categorizada con el mismo numero de caracteres

```r
#names(df)
df <- df %>%
        unite("lineups_H", matches("homeplayer[1-5]"), remove = TRUE) %>%
        unite("lineups_A", matches("awayplayer[1-5]"), remove = TRUE) %>%
        unite("lineups", matches("lineups_"), remove = FALSE) %>%
        mutate(game_code = paste0("G", str_pad(game_code, 6, pad = "0")),
               quarter   = paste0("Q", quarter))

df
```

```
##     id_play season game_code quarter time points_home points_away
```

```
## 1        1  S2017   G000001    Q1    0            0            0
## 2        2  S2017   G000001    Q1  200            0            2
## 3        3  S2017   G000001    Q1  312            0            2
## 4        4  S2017   G000001    Q2  680            2            2
## 5        5  S2017   G000001    Q2  936            2            4
## 6        6  S2017   G000001    Q3 1305            2            4
## 7        7  S2017   G000001    Q3 1640            5            7
## 8        8  S2017   G000001    Q4 1905            5            7
## 9        9  S2017   G000001    Q4 1980            7            9
## 10      10  S2017   G000001    Q4 2177            7            9
## 11      11  S2017   G000002    Q1    0            0            0
## 12      12  S2017   G000002    Q1  228            2            0
## 13      13  S2017   G000002    Q1  285            2            2
## 14      14  S2017   G000002    Q1  406            5            2
## 15      15  S2017   G000002    Q2  681            5            4
## 16      16  S2017   G000002    Q2  750            7            4
## 17      17  S2017   G000002    Q2  860            7            6
## 18      18  S2017   G000002    Q2  972            7            6
## 19      19  S2017   G000002    Q2 1029            9            6
## 20      20  S2017   G000002    Q2 1194           10            6
##                                 lineups          lineups_H           lineups_A
## 1      Juan_Maria_Andrea_Ignasi_Anna_Gerard   Juan_Maria_Andrea Ignasi_Anna_Gerard
## 2      Juan_Maria_Andrea_Ignasi_Anna_Gerard   Juan_Maria_Andrea Ignasi_Anna_Gerard
## 3      Juan_Maria_Andrea_Ignasi_Anna_Gerard   Juan_Maria_Andrea Ignasi_Anna_Gerard
## 4      Juan_Maria_Andrea_Ignasi_Anna_Gerard   Juan_Maria_Andrea Ignasi_Anna_Gerard
## 5     Diego_Maria_Andrea_Ignasi_Anna_Gerard  Diego_Maria_Andrea Ignasi_Anna_Gerard
## 6     Diego_Maria_Andrea_Ignasi_Anna_Gerard  Diego_Maria_Andrea Ignasi_Anna_Gerard
## 7     Diego_Maria_Andrea_Ignasi_Anna_Gerard  Diego_Maria_Andrea Ignasi_Anna_Gerard
## 8      Diego_Maria_Carla_Ignasi_Anna_Gerard   Diego_Maria_Carla Ignasi_Anna_Gerard
## 9      Diego_Maria_Carla_Ignasi_Jose_Gerard   Diego_Maria_Carla Ignasi_Jose_Gerard
## 10     Diego_Maria_Carla_Ignasi_Jose_Gerard   Diego_Maria_Carla Ignasi_Jose_Gerard
## 11       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 12       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 13       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 14       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 15       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 16       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 17       Ignasi_Gerard_Jose_Bella_Gus_Alba  Ignasi_Gerard_Jose     Bella_Gus_Alba
## 18         Anna_Gerard_Jose_Bella_Gus_Erik    Anna_Gerard_Jose     Bella_Gus_Erik
## 19     Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
## 20     Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
```

MERGE Temporada+game_code

Unimos 'Season' y 'Game_Code' para tener una variable identificativa del partido.

```
df_merged <- df %>%  unite("SeasonGame", c("season", "game_code"))
df_merged
```

```
##    id_play    SeasonGame quarter time points_home points_away
## 1        1 S2017_G000001      Q1    0           0           0
## 2        2 S2017_G000001      Q1  200           0           2
## 3        3 S2017_G000001      Q1  312           0           2
## 4        4 S2017_G000001      Q2  680           2           2
## 5        5 S2017_G000001      Q2  936           2           4
```

```
## 6            6 S2017_G000001    Q3 1305              2              4
## 7            7 S2017_G000001    Q3 1640              5              7
## 8            8 S2017_G000001    Q4 1905              5              7
## 9            9 S2017_G000001    Q4 1980              7              9
## 10          10 S2017_G000001    Q4 2177              7              9
## 11          11 S2017_G000002    Q1    0              0              0
## 12          12 S2017_G000002    Q1  228              2              0
## 13          13 S2017_G000002    Q1  285              2              2
## 14          14 S2017_G000002    Q1  406              5              2
## 15          15 S2017_G000002    Q2  681              5              4
## 16          16 S2017_G000002    Q2  750              7              4
## 17          17 S2017_G000002    Q2  860              7              6
## 18          18 S2017_G000002    Q2  972              7              6
## 19          19 S2017_G000002    Q2 1029              9              6
## 20          20 S2017_G000002    Q2 1194             10              6
##                                 lineups          lineups_H          lineups_A
## 1    Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 2    Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 3    Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 4    Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 5   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 6   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 7   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 8    Diego_Maria_Carla_Ignasi_Anna_Gerard  Diego_Maria_Carla Ignasi_Anna_Gerard
## 9    Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 10   Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 11      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 12      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 13      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 14      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 15      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 16      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 17      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 18       Anna_Gerard_Jose_Bella_Gus_Erik   Anna_Gerard_Jose      Bella_Gus_Erik
## 19    Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
## 20    Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
```

ORDENAR DF -> Ya no hace falta

UNIQUE quintetos

Queremos obtener un df con los quintetos sin duplicar, con la información agupada por

```r
# Columnas TRUE o FALSE si son iguales a su lag row
df_merged <- df_merged %>%
  mutate(players_SAME = (lineups == lag(lineups)),
         SeasonGame_SAME = (SeasonGame == lag(SeasonGame)))

## Determinamos los primeros elementos de estas dos variables como FALSE para que no
## nos aparezca NA, ya que no puede evaluar con anterior
df_merged$players_SAME[1]    <- FALSE
df_merged$SeasonGame_SAME[1] <- FALSE

df_merged <- df_merged %>%
  mutate(NEW_STINT = ifelse(((players_SAME == FALSE)|(SeasonGame_SAME == FALSE )),TRUE, FALSE),
```

```
        ### 1ra condicions para cambio de jugadores y 2nda para cambio de partido.
        stint = ifelse((lead(NEW_STINT) == TRUE)|(SeasonGame_SAME == FALSE ), NEW_STINT, NA))
        ### La primera condicion nos mantiene las que son iguales anterior a la fila en la que estamos
        ### ha habiado cambio. La segunda condicion nos mantiene la primera entrada de cada partido.
```

```
df_merged
```

```
##    id_play    SeasonGame quarter time points_home points_away
## 1        1 S2017_G000001      Q1    0           0           0
## 2        2 S2017_G000001      Q1  200           0           2
## 3        3 S2017_G000001      Q1  312           0           2
## 4        4 S2017_G000001      Q2  680           2           2
## 5        5 S2017_G000001      Q2  936           2           4
## 6        6 S2017_G000001      Q3 1305           2           4
## 7        7 S2017_G000001      Q3 1640           5           7
## 8        8 S2017_G000001      Q4 1905           5           7
## 9        9 S2017_G000001      Q4 1980           7           9
## 10      10 S2017_G000001      Q4 2177           7           9
## 11      11 S2017_G000002      Q1    0           0           0
## 12      12 S2017_G000002      Q1  228           2           0
## 13      13 S2017_G000002      Q1  285           2           2
## 14      14 S2017_G000002      Q1  406           5           2
## 15      15 S2017_G000002      Q2  681           5           4
## 16      16 S2017_G000002      Q2  750           7           4
## 17      17 S2017_G000002      Q2  860           7           6
## 18      18 S2017_G000002      Q2  972           7           6
## 19      19 S2017_G000002      Q2 1029           9           6
## 20      20 S2017_G000002      Q2 1194          10           6
##                                  lineups          lineups_H        lineups_A
## 1     Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 2     Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 3     Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 4     Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 5   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 6   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 7   Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 8    Diego_Maria_Carla_Ignasi_Anna_Gerard  Diego_Maria_Carla Ignasi_Anna_Gerard
## 9    Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 10   Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 11      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 12      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 13      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 14      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 15      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 16      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 17      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose     Bella_Gus_Alba
## 18        Anna_Gerard_Jose_Bella_Gus_Erik   Anna_Gerard_Jose     Bella_Gus_Erik
## 19    Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
## 20    Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula   Bella_Kevin_Erik
##    players_SAME SeasonGame_SAME NEW_STINT stint
## 1         FALSE           FALSE      TRUE  TRUE
## 2          TRUE            TRUE     FALSE    NA
## 3          TRUE            TRUE     FALSE    NA
```

```
## 4          TRUE           TRUE      FALSE FALSE
## 5         FALSE           TRUE       TRUE    NA
## 6          TRUE           TRUE      FALSE    NA
## 7          TRUE           TRUE      FALSE FALSE
## 8         FALSE           TRUE       TRUE  TRUE
## 9         FALSE           TRUE       TRUE    NA
## 10         TRUE           TRUE      FALSE FALSE
## 11        FALSE          FALSE       TRUE  TRUE
## 12         TRUE           TRUE      FALSE    NA
## 13         TRUE           TRUE      FALSE    NA
## 14         TRUE           TRUE      FALSE    NA
## 15         TRUE           TRUE      FALSE    NA
## 16         TRUE           TRUE      FALSE    NA
## 17         TRUE           TRUE      FALSE FALSE
## 18        FALSE           TRUE       TRUE  TRUE
## 19        FALSE           TRUE       TRUE    NA
## 20         TRUE           TRUE      FALSE    NA
```

Hasta aquí tenemos detectados los stints que mantendremos.

STINTS

Para que nos aparezcan tambien la primera y la ultima entrada:

```
n_df <- dim(df_merged)[1]
df_merged$stint[1] <- df_merged$NEW_STINT[1]
df_merged$stint[n_df] <- df_merged$NEW_STINT[n_df]

df_merged
```

```
##    id_play    SeasonGame quarter time points_home points_away
## 1        1 S2017_G000001      Q1    0           0           0
## 2        2 S2017_G000001      Q1  200           0           2
## 3        3 S2017_G000001      Q1  312           0           2
## 4        4 S2017_G000001      Q2  680           2           2
## 5        5 S2017_G000001      Q2  936           2           4
## 6        6 S2017_G000001      Q3 1305           2           4
## 7        7 S2017_G000001      Q3 1640           5           7
## 8        8 S2017_G000001      Q4 1905           5           7
## 9        9 S2017_G000001      Q4 1980           7           9
## 10      10 S2017_G000001      Q4 2177           7           9
## 11      11 S2017_G000002      Q1    0           0           0
## 12      12 S2017_G000002      Q1  228           2           0
## 13      13 S2017_G000002      Q1  285           2           2
## 14      14 S2017_G000002      Q1  406           5           2
## 15      15 S2017_G000002      Q2  681           5           4
## 16      16 S2017_G000002      Q2  750           7           4
## 17      17 S2017_G000002      Q2  860           7           6
## 18      18 S2017_G000002      Q2  972           7           6
## 19      19 S2017_G000002      Q2 1029           9           6
## 20      20 S2017_G000002      Q2 1194          10           6
##                              lineups         lineups_H        lineups_A
## 1  Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 2  Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
## 3  Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea Ignasi_Anna_Gerard
```

```
## 4      Juan_Maria_Andrea_Ignasi_Anna_Gerard   Juan_Maria_Andrea Ignasi_Anna_Gerard
## 5  Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 6  Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 7  Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea Ignasi_Anna_Gerard
## 8   Diego_Maria_Carla_Ignasi_Anna_Gerard  Diego_Maria_Carla Ignasi_Anna_Gerard
## 9   Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 10  Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla Ignasi_Jose_Gerard
## 11      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 12      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 13      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 14      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 15      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 16      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 17      Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 18       Anna_Gerard_Jose_Bella_Gus_Erik   Anna_Gerard_Jose      Bella_Gus_Erik
## 19   Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula  Bella_Kevin_Erik
## 20   Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula  Bella_Kevin_Erik
##    players_SAME SeasonGame_SAME NEW_STINT stint
## 1         FALSE           FALSE      TRUE  TRUE
## 2          TRUE            TRUE     FALSE    NA
## 3          TRUE            TRUE     FALSE    NA
## 4          TRUE            TRUE     FALSE FALSE
## 5         FALSE            TRUE      TRUE    NA
## 6          TRUE            TRUE     FALSE    NA
## 7          TRUE            TRUE     FALSE FALSE
## 8         FALSE            TRUE      TRUE  TRUE
## 9         FALSE            TRUE      TRUE    NA
## 10         TRUE            TRUE     FALSE FALSE
## 11        FALSE           FALSE      TRUE  TRUE
## 12         TRUE            TRUE     FALSE    NA
## 13         TRUE            TRUE     FALSE    NA
## 14         TRUE            TRUE     FALSE    NA
## 15         TRUE            TRUE     FALSE    NA
## 16         TRUE            TRUE     FALSE    NA
## 17         TRUE            TRUE     FALSE FALSE
## 18        FALSE            TRUE      TRUE  TRUE
## 19        FALSE            TRUE      TRUE    NA
## 20         TRUE            TRUE     FALSE FALSE
```

```r
STINTS <- df_merged %>% drop_na(stint)          #Eliminar NA rows

PM_function <- function(h,a){
  (h-lag(h))-(a-lag(a))
}

STINTS <- STINTS %>%
  mutate(ID_Stint = 1:nrow(STINTS),
         stint_times = times(ifelse(SeasonGame_SAME == TRUE, time-lag(time), 0)),
         PM = ifelse(SeasonGame_SAME == TRUE, PM_function(points_home, points_away), 0)) #Home como ref

STINTS
```

```
##   id_play    SeasonGame quarter time points_home points_away
## 1       1 S2017_G000001      Q1    0           0           0
## 2       4 S2017_G000001      Q2  680           2           2
```

```
## 3         7 S2017_G000001      Q3 1640              5              7
## 4         8 S2017_G000001      Q4 1905              5              7
## 5        10 S2017_G000001      Q4 2177              7              9
## 6        11 S2017_G000002      Q1    0              0              0
## 7        17 S2017_G000002      Q2  860              7              6
## 8        18 S2017_G000002      Q2  972              7              6
## 9        20 S2017_G000002      Q2 1194             10              6
##                                lineups            lineups_H           lineups_A
## 1  Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea  Ignasi_Anna_Gerard
## 2  Juan_Maria_Andrea_Ignasi_Anna_Gerard  Juan_Maria_Andrea  Ignasi_Anna_Gerard
## 3 Diego_Maria_Andrea_Ignasi_Anna_Gerard Diego_Maria_Andrea  Ignasi_Anna_Gerard
## 4  Diego_Maria_Carla_Ignasi_Anna_Gerard  Diego_Maria_Carla  Ignasi_Anna_Gerard
## 5  Diego_Maria_Carla_Ignasi_Jose_Gerard  Diego_Maria_Carla  Ignasi_Jose_Gerard
## 6     Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 7     Ignasi_Gerard_Jose_Bella_Gus_Alba Ignasi_Gerard_Jose      Bella_Gus_Alba
## 8       Anna_Gerard_Jose_Bella_Gus_Erik   Anna_Gerard_Jose      Bella_Gus_Erik
## 9   Anna_Gerard_Paula_Bella_Kevin_Erik   Anna_Gerard_Paula    Bella_Kevin_Erik
##   players_SAME SeasonGame_SAME NEW_STINT  stint ID_Stint stint_times PM
## 1        FALSE           FALSE      TRUE   TRUE        1           0  0
## 2         TRUE            TRUE     FALSE  FALSE        2         680  0
## 3         TRUE            TRUE     FALSE  FALSE        3         960 -2
## 4        FALSE            TRUE      TRUE   TRUE        4         265  0
## 5         TRUE            TRUE     FALSE  FALSE        5         272  0
## 6        FALSE           FALSE      TRUE   TRUE        6           0  0
## 7         TRUE            TRUE     FALSE  FALSE        7         860  1
## 8        FALSE            TRUE      TRUE   TRUE        8         112  0
## 9         TRUE            TRUE     FALSE  FALSE        9         222  3
```

```
names(STINTS)
```

```
##  [1] "id_play"         "SeasonGame"      "quarter"         "time"
##  [5] "points_home"     "points_away"     "lineups"         "lineups_H"
##  [9] "lineups_A"       "players_SAME"    "SeasonGame_SAME" "NEW_STINT"
## [13] "stint"           "ID_Stint"        "stint_times"     "PM"
```

```
STINTS_Reduced <- STINTS %>% dplyr::select(c(SeasonGame, quarter, lineups, lineups_H, lineups_A, stint_
STINTS_Reduced
```

```
##      SeasonGame quarter                                lineups
## 1 S2017_G000001      Q1  Juan_Maria_Andrea_Ignasi_Anna_Gerard
## 2 S2017_G000001      Q2  Juan_Maria_Andrea_Ignasi_Anna_Gerard
## 3 S2017_G000001      Q3 Diego_Maria_Andrea_Ignasi_Anna_Gerard
## 4 S2017_G000001      Q4  Diego_Maria_Carla_Ignasi_Anna_Gerard
## 5 S2017_G000001      Q4  Diego_Maria_Carla_Ignasi_Jose_Gerard
## 6 S2017_G000002      Q1     Ignasi_Gerard_Jose_Bella_Gus_Alba
## 7 S2017_G000002      Q2     Ignasi_Gerard_Jose_Bella_Gus_Alba
## 8 S2017_G000002      Q2       Anna_Gerard_Jose_Bella_Gus_Erik
## 9 S2017_G000002      Q2     Anna_Gerard_Paula_Bella_Kevin_Erik
##           lineups_H           lineups_A stint_times PM
## 1  Juan_Maria_Andrea Ignasi_Anna_Gerard           0  0
## 2  Juan_Maria_Andrea Ignasi_Anna_Gerard         680  0
## 3 Diego_Maria_Andrea Ignasi_Anna_Gerard         960 -2
## 4  Diego_Maria_Carla Ignasi_Anna_Gerard         265  0
## 5  Diego_Maria_Carla Ignasi_Jose_Gerard         272  0
## 6 Ignasi_Gerard_Jose     Bella_Gus_Alba           0  0
## 7 Ignasi_Gerard_Jose     Bella_Gus_Alba         860  1
```

```
## 8   Anna_Gerard_Jose     Bella_Gus_Erik       112  0
## 9   Anna_Gerard_Paula   Bella_Kevin_Erik      222  3
```

PARTE 2: Dummys Jugadores

```r
#PRUEBA2 - fastDummies

# Vector con todos los nombres de los jugadores:
players <- c(equipo1, equipo2, equipo3, equipo4)
length(players)
```

```
## [1] 20
```

```r
df_dummys_H <- fastDummies::dummy_cols(STINTS_Reduced, select_columns = "lineups_H", split = "_")

df_dummys_A <- fastDummies::dummy_cols(STINTS_Reduced, select_columns = "lineups_A", split = "_")
df_dummys_A <- df_dummys_A %>%
                  mutate(across(starts_with("lineups_A_"), function(x) -x))

Remove_first_char <- function(data, char){
  num_char <- nchar(char)+1
  substring(names(data), num_char)
}

COL_From <- function(data, first_col){
  last_col = ncol(data)
  colnames(data[first_col:last_col])
}

COL_to <- function(data, first_col, char){
  last_col = ncol(data)
  Remove_first_char(data[first_col:last_col], char)
}

#Primera columna con el nombre de un jugador
match <- match(paste0("lineups_A_", players), names(df_dummys_A)) %>%
          na.omit() %>%
          min()

#nombre columnas sin primera palabras:
col_to_A <- COL_to(df_dummys_A, match, "lineups_A_")
col_to_H <- COL_to(df_dummys_H, match, "lineups_H_")

df_dummys_A <- df_dummys_A %>% rename_at(vars(COL_From(df_dummys_A,match)), function(x) col_to_A)
df_dummys_H <- df_dummys_H %>% rename_at(vars(COL_From(df_dummys_H,match)), function(x) col_to_H)

#Juntar los dos DF:
df_dummys_A[df_dummys_A == 0] <- NA
df_dummys_H[df_dummys_H == 0] <- NA

df_dummys <- coalesce(df_dummys_H,df_dummys_A)
df_dummys[is.na(df_dummys)] <- 0

df_dummys <- df_dummys %>% select(-c(starts_with("lineups_")))
```

```
df_dummys
```

```
##        SeasonGame quarter                              lineups stint_times PM
## 1 S2017_G000001      Q1  Juan_Maria_Andrea_Ignasi_Anna_Gerard           0  0
## 2 S2017_G000001      Q2  Juan_Maria_Andrea_Ignasi_Anna_Gerard         680  0
## 3 S2017_G000001      Q3 Diego_Maria_Andrea_Ignasi_Anna_Gerard         960 -2
## 4 S2017_G000001      Q4  Diego_Maria_Carla_Ignasi_Anna_Gerard         265  0
## 5 S2017_G000001      Q4  Diego_Maria_Carla_Ignasi_Jose_Gerard         272  0
## 6 S2017_G000002      Q1     Ignasi_Gerard_Jose_Bella_Gus_Alba           0  0
## 7 S2017_G000002      Q2     Ignasi_Gerard_Jose_Bella_Gus_Alba         860  1
## 8 S2017_G000002      Q2      Anna_Gerard_Jose_Bella_Gus_Erik         112  0
## 9 S2017_G000002      Q2     Anna_Gerard_Paula_Bella_Kevin_Erik        222  3
##   Anna Gerard Jose Paula Diego Maria Andrea Carla Ignasi Juan Bella Gus Alba
## 1   -1     -1    0     0     0     1      1     0     -1    1     0   0    0
## 2   -1     -1    0     0     0     1      1     0     -1    1     0   0    0
## 3   -1     -1    0     0     1     1      1     0     -1    0     0   0    0
## 4   -1     -1    0     0     1     1      0     1     -1    0     0   0    0
## 5    0     -1   -1     0     1     1      0     1     -1    0     0   0    0
## 6    0      1    1     0     0     0      0     0      1    0    -1  -1   -1
## 7    0      1    1     0     0     0      0     0      1    0    -1  -1   -1
## 8    1      1    1     0     0     0      0     0      0    0    -1  -1    0
## 9    1      1    0     1     0     0      0     0      0    0    -1   0    0
##   Erik Kevin
## 1    0     0
## 2    0     0
## 3    0     0
## 4    0     0
## 5    0     0
## 6    0     0
## 7    0     0
## 8   -1     0
## 9   -1    -1
```

Ahora mismo tenemos un DF con el PM con HOME como referencia (si es positivo, ganaban HOME. Si es Negativo ganaban AWAY). Luego tenemos variables "dummys" con 1 si estaban jugando como HOME, -1 si estaban jugando como AWAY y 0 si no estaban en pista.

PARTE 3: PM por stint (PM CLASSIC)

```
df_dummys_PM <- df_dummys %>%
                mutate(across(matches(players), function(x) x*PM)) %>%
                select(-c(PM))
```

```
### PM de los mismos lineups (sin tener en cuenta SeasonGame o Quarter):
```

```
df_dummys_PM_2 <- df_dummys_PM %>% select(-c(SeasonGame, quarter))
```

```
PM_Lineups <-  aggregate(. ~ lineups, df_dummys_PM_2, sum, na.rm = TRUE) %>%
                mutate(stint_times = Print_MS(stint_times))
PM_Lineups
```

```
##                             lineups stint_times Anna Gerard Jose Paula
## 1        Anna_Gerard_Jose_Bella_Gus_Erik    00:01:52    0      0    0     0
```
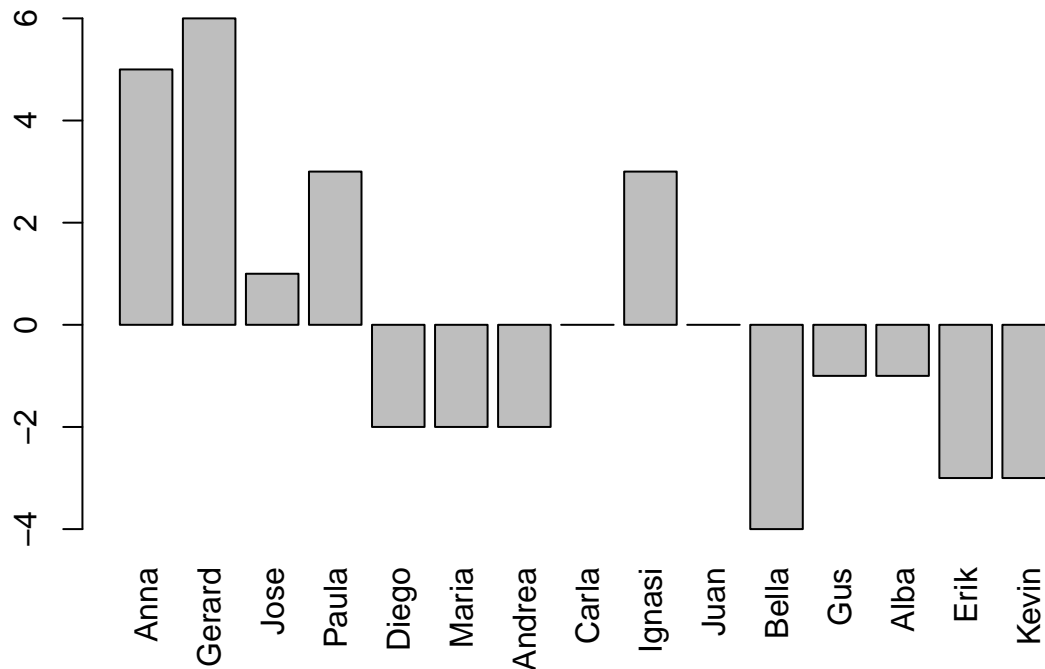
```
## 2      Anna_Gerard_Paula_Bella_Kevin_Erik      00:03:42     3        3     0      3
## 3 Diego_Maria_Andrea_Ignasi_Anna_Gerard        00:16:00     2        2     0      0
## 4  Diego_Maria_Carla_Ignasi_Anna_Gerard        00:04:25     0        0     0      0
## 5  Diego_Maria_Carla_Ignasi_Jose_Gerard        00:04:32     0        0     0      0
## 6      Ignasi_Gerard_Jose_Bella_Gus_Alba       00:14:20     0        1     1      0
## 7 Juan_Maria_Andrea_Ignasi_Anna_Gerard         00:11:20     0        0     0      0
##   Diego Maria Andrea Carla Ignasi Juan Bella Gus Alba Erik Kevin
## 1     0     0      0     0      0    0     0   0    0    0     0
## 2     0     0      0     0      0    0    -3   0    0   -3    -3
## 3    -2    -2     -2     0      2    0     0   0    0    0     0
## 4     0     0      0     0      0    0     0   0    0    0     0
## 5     0     0      0     0      0    0     0   0    0    0     0
## 6     0     0      0     0      1    0    -1  -1   -1    0     0
## 7     0     0      0     0      0    0     0   0    0    0     0
```

```
PM_C <- colSums(PM_Lineups[3:ncol(PM_Lineups)])
PM_C
```

```
##   Anna Gerard   Jose  Paula  Diego  Maria Andrea  Carla Ignasi   Juan  Bella
##      5      6      1      3     -2     -2     -2      0      3      0     -4
##    Gus   Alba   Erik  Kevin
##     -1     -1     -3     -3
```

```
barplot(PM_C, las=3)
```



```
table(PM_C)
```

```
## PM_C
```

```
## -4 -3 -2 -1  0  1  3  5  6
##  1  2  3  2  2  1  2  1  1
```

PARTE 4: Modelar

Variable outcome: PM

```
names(df_dummys)
```

```
##  [1] "SeasonGame"  "quarter"     "lineups"     "stint_times" "PM"
##  [6] "Anna"        "Gerard"      "Jose"        "Paula"       "Diego"
## [11] "Maria"       "Andrea"      "Carla"       "Ignasi"      "Juan"
## [16] "Bella"       "Gus"         "Alba"        "Erik"        "Kevin"
```

```
mod1 <- lm(PM ~ . -SeasonGame -quarter -lineups -stint_times, data=df_dummys)
summary(mod1)
```

```
##
## Call:
## lm(formula = PM ~ . - SeasonGame - quarter - lineups - stint_times,
##     data = df_dummys)
##
## Residuals:
##          1          2          3          4          5          6          7
##  4.237e-16 -3.285e-16 -8.842e-17 -3.784e-17  5.729e-18 -5.000e-01  5.000e-01
##          8          9
##  9.922e-18  1.320e-17
##
## Coefficients: (9 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.2500     0.4330   2.887   0.1020
## Anna         -0.5000     0.6124  -0.816   0.5000
## Gerard       -0.2500     0.9682  -0.258   0.8204
## Jose         -0.5000     0.9354  -0.535   0.6464
## Paula         2.5000     0.9354   2.673   0.1161
## Diego        -2.0000     0.6124  -3.266   0.0823 .
## Maria             NA         NA      NA       NA
## Andrea       -2.0000     0.7071  -2.828   0.1056
## Carla             NA         NA      NA       NA
## Ignasi            NA         NA      NA       NA
## Juan              NA         NA      NA       NA
## Bella             NA         NA      NA       NA
## Gus               NA         NA      NA       NA
## Alba              NA         NA      NA       NA
## Erik              NA         NA      NA       NA
## Kevin             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5 on 2 degrees of freedom
## Multiple R-squared:  0.9631, Adjusted R-squared:  0.8525
## F-statistic: 8.704 on 6 and 2 DF,  p-value: 0.1066
```

```
mod2 <- lm(PM ~ Anna * Diego, data=df_dummys)
summary(mod2)
```

```
## 
## Call:
## lm(formula = PM ~ Anna * Diego, data = df_dummys)
## 
## Residuals:
##          1          2          3          4          5          6          7
##  8.333e-02  8.333e-02 -1.000e+00  1.000e+00 -1.110e-16 -6.667e-01  3.333e-01
##          8          9
## -1.417e+00  1.583e+00
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6667     0.4859   1.372    0.228
## Anna          0.7500     0.5951   1.260    0.263
## Diego        -0.6667     1.2856  -0.519    0.626
## Anna:Diego    0.2500     1.5745   0.159    0.880
## 
## Residual standard error: 1.19 on 5 degrees of freedom
## Multiple R-squared:  0.4775, Adjusted R-squared:  0.1639
## F-statistic: 1.523 on 3 and 5 DF,  p-value: 0.3171
```