Ejercicios Computación

Venus Puertas Daniel Mesias Julieth Moreno

2022-08

Ejercicio English Premiere league

Para este ejercicio se importan los conjuntos de datos usando la librería readxl de R, por ello lo primero que hacemos es cargar la librería.

```
library(readxl)
```

Ahora vamos a cargar el conjunto de datos season-1415.xlsx, para ello se usa la función read_excel() y especificamos el nombre de las columnas que queremos que R lea con col_names, adicionalmente con col_type le esepcificamos cómo debe leerlas y cuáles debe omitir. Dado que le estamos especificando los nombres de las columnas, debemos usar el argumento skip=1 para que evite leer la primera línea donde se encuentran los títulos.

```
season1415<-read_excel("season-1415.xlsx",col_names=c("Season","Date","HomeTeam"
,"HomeAway","FTHG","FTAG","Referee","HS","AS","HST","AST","HF","AF","HC","AC",
"HY","AY","HR", "AR") ,col_types = c("text", "date", "text", "text", "numeric",
"numeric","skip", "skip", "skip", "skip", "text", "numeric","numeric","numeric",
"numeric","numeric","numeric","numeric","numeric","numeric","numeric","numeric",
"numeric"), skip=1)
str(season1415)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Season  : chr [1:380] "14-15" "14-15" "14-15" "14-15" ...
##  $ Date    : POSIXct[1:380], format: "2014-08-16" "2014-08-16" ...
##  $ HomeTeam: chr [1:380] "Arsenal" "Leicester" "Man United" "QPR" ...
##  $ HomeAway: chr [1:380] "Crystal Palace" "Everton" "Swansea" "Hull" ...
##  $ FTHG    : num [1:380] 2 2 1 0 0 2 0 2 0 1 ...
##  $ FTAG    : num [1:380] 1 2 2 1 1 2 1 1 2 3 ...
##  $ Referee : chr [1:380] "J Moss" "M Jones" "M Dean" "C Pawson" ...
##  $ HS      : num [1:380] 14 11 14 19 12 10 18 12 12 9 ...
##  $ AS      : num [1:380] 4 13 5 11 7 7 10 12 13 11 ...
##  $ HST     : num [1:380] 6 3 5 6 2 5 4 5 0 2 ...
##  $ AST     : num [1:380] 2 3 4 4 2 2 4 6 5 3 ...
##  $ HF      : num [1:380] 13 16 14 10 14 18 12 8 8 6 ...
##  $ AF      : num [1:380] 19 10 20 10 9 9 10 11 11 7 ...
##  $ HC      : num [1:380] 9 3 4 8 2 6 8 2 3 4 ...
##  $ AC      : num [1:380] 3 6 0 9 8 3 5 6 3 3 ...
##  $ HY      : num [1:380] 2 1 2 1 0 3 1 1 1 1 ...
##  $ AY      : num [1:380] 2 1 4 2 3 1 0 2 5 1 ...
##  $ HR      : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
##  $ AR      : num [1:380] 1 0 0 0 0 0 1 0 0 0 ...
```

A continuación vamos a personalizar el nombre y tipo de variables del archivo season-1415.xlsx

```
season1415.2<-read_excel("season-1415.xlsx",sheet = 1, skip = 1,  col_names =
c("Temp", "Fecha","E.local", "E.visitante", "GLTC", "GVTC", "Arbitro", "DL",
"DV", "DAL", "DAV", "FL", "FV", "FLL", "FLV", "TAL", "TAV", "TRL", "TRV"),
col_types = c("text", "date", "text", "text","numeric","numeric","skip", "skip",
"skip", "skip", "text", "numeric","numeric","numeric","numeric","numeric","numeric"
,"numeric","numeric","numeric","numeric","numeric","numeric"))
str(season1415.2)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Temp       : chr [1:380] "14-15" "14-15" "14-15" "14-15" ...
##  $ Fecha      : POSIXct[1:380], format: "2014-08-16" "2014-08-16" ...
##  $ E.local    : chr [1:380] "Arsenal" "Leicester" "Man United" "QPR" ...
##  $ E.visitante: chr [1:380] "Crystal Palace" "Everton" "Swansea" "Hull" ...
##  $ GLTC       : num [1:380] 2 2 1 0 0 2 0 2 0 1 ...
##  $ GVTC       : num [1:380] 1 2 2 1 1 2 1 1 2 3 ...
##  $ Arbitro    : chr [1:380] "J Moss" "M Jones" "M Dean" "C Pawson" ...
##  $ DL         : num [1:380] 14 11 14 19 12 10 18 12 12 9 ...
##  $ DV         : num [1:380] 4 13 5 11 7 7 10 12 13 11 ...
##  $ DAL        : num [1:380] 6 3 5 6 2 5 4 5 0 2 ...
##  $ DAV        : num [1:380] 2 3 4 4 2 2 4 6 5 3 ...
##  $ FL         : num [1:380] 13 16 14 10 14 18 12 8 8 6 ...
##  $ FV         : num [1:380] 19 10 20 10 9 9 10 11 11 7 ...
##  $ FLL        : num [1:380] 9 3 4 8 2 6 8 2 3 4 ...
##  $ FLV        : num [1:380] 3 6 0 9 8 3 5 6 3 3 ...
##  $ TAL        : num [1:380] 2 1 2 1 0 3 1 1 1 1 ...
##  $ TAV        : num [1:380] 2 1 4 2 3 1 0 2 5 1 ...
##  $ TRL        : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
##  $ TRV        : num [1:380] 1 0 0 0 0 0 1 0 0 0 ...
```

Para el archivo season-1516.xlsx

```
season1516<-read_excel("season-1415.xlsx",col_names=c("Season","Date","HomeTeam"
,"HomeAway","FTHG","FTAG","Referee","HS","AS","HST","AST","HF","AF","HC","AC",
"HY","AY","HR", "AR") ,col_types = c("text", "date", "text", "text", "numeric",
"numeric","skip", "skip", "skip", "skip", "text", "numeric","numeric","numeric",
"numeric","numeric","numeric","numeric","numeric","numeric","numeric","numeric",
"numeric"), skip=1)
str(season1516)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Season  : chr [1:380] "14-15" "14-15" "14-15" "14-15" ...
##  $ Date    : POSIXct[1:380], format: "2014-08-16" "2014-08-16" ...
##  $ HomeTeam: chr [1:380] "Arsenal" "Leicester" "Man United" "QPR" ...
##  $ HomeAway: chr [1:380] "Crystal Palace" "Everton" "Swansea" "Hull" ...
##  $ FTHG    : num [1:380] 2 2 1 0 0 2 0 2 0 1 ...
##  $ FTAG    : num [1:380] 1 2 2 1 1 2 1 1 2 3 ...
##  $ Referee : chr [1:380] "J Moss" "M Jones" "M Dean" "C Pawson" ...
##  $ HS      : num [1:380] 14 11 14 19 12 10 18 12 12 9 ...
##  $ AS      : num [1:380] 4 13 5 11 7 7 10 12 13 11 ...
##  $ HST     : num [1:380] 6 3 5 6 2 5 4 5 0 2 ...
##  $ AST     : num [1:380] 2 3 4 4 2 2 4 6 5 3 ...
##  $ HF      : num [1:380] 13 16 14 10 14 18 12 8 8 6 ...
##  $ AF      : num [1:380] 19 10 20 10 9 9 10 11 11 7 ...
##  $ HC      : num [1:380] 9 3 4 8 2 6 8 2 3 4 ...
##  $ AC      : num [1:380] 3 6 0 9 8 3 5 6 3 3 ...
```

```
##  $ HY      : num [1:380] 2 1 2 1 0 3 1 1 1 1 ...
##  $ AY      : num [1:380] 2 1 4 2 3 1 0 2 5 1 ...
##  $ HR      : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
##  $ AR      : num [1:380] 1 0 0 0 0 0 1 0 0 0 ...
```

A continuación vamos a personalizar el nombre y tipo de variables del archivo season-1516.xlsx

```
season1516.2<-read_excel("season-1516.xlsx",sheet = 1, skip = 1,  col_names =
c("Temp", "Fecha","E.local", "E.visitante", "GLTC", "GVTC", "Arbitro", "DL",
"DV", "DAL", "DAV", "FL", "FV", "FLL", "FLV", "TAL", "TAV", "TRL", "TRV"),
col_types = c("text", "date", "text", "text","numeric","numeric","skip", "skip",
"skip", "skip", "text", "numeric","numeric","numeric","numeric","numeric","numeric"
,"numeric","numeric","numeric","numeric","numeric","numeric"))
str(season1516.2)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Temp       : chr [1:380] "15-16" "15-16" "15-16" "15-16" ...
##  $ Fecha      : POSIXct[1:380], format: "2015-08-08" "2015-08-08" ...
##  $ E.local    : chr [1:380] "Bournemouth" "Chelsea" "Everton" "Leicester" ...
##  $ E.visitante: chr [1:380] "Aston Villa" "Swansea" "Watford" "Sunderland" ...
##  $ GLTC       : num [1:380] 0 2 2 4 1 1 0 2 0 0 ...
##  $ GVTC       : num [1:380] 1 2 2 2 0 3 2 2 1 3 ...
##  $ Arbitro    : chr [1:380] "M Clattenburg" "M Oliver" "M Jones" "L Mason" ...
##  $ DL         : num [1:380] 11 11 10 19 9 17 22 9 7 9 ...
##  $ DV         : num [1:380] 7 18 11 10 9 11 8 15 8 19 ...
##  $ DAL        : num [1:380] 2 3 5 8 1 6 6 4 1 2 ...
##  $ DAV        : num [1:380] 3 10 5 5 4 7 4 5 3 7 ...
##  $ FL         : num [1:380] 13 15 7 13 12 14 12 9 9 12 ...
##  $ FV         : num [1:380] 13 16 13 17 12 20 9 12 16 9 ...
##  $ FLL        : num [1:380] 6 4 8 6 1 1 5 6 3 6 ...
##  $ FLV        : num [1:380] 3 8 2 3 2 4 4 6 5 6 ...
##  $ TAL        : num [1:380] 3 1 1 2 2 1 1 2 2 4 ...
##  $ TAV        : num [1:380] 4 3 2 4 3 0 3 4 4 1 ...
##  $ TRL        : num [1:380] 0 1 0 0 0 0 0 0 0 0 ...
##  $ TRV        : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
```

Para el archivo season-1617.xlsx

```
season1617<-read_excel("season-1617.xlsx",col_names=c("Season","Date","HomeTeam"
,"HomeAway","FTHG","FTAG","Referee","HS","AS","HST","AST","HF","AF","HC","AC",
"HY","AY","HR", "AR") ,col_types = c("text", "date", "text", "text", "numeric",
"numeric","skip", "skip", "skip", "skip", "text", "numeric","numeric","numeric",
"numeric","numeric","numeric","numeric","numeric","numeric","numeric","numeric",
"numeric"), skip=1)
str(season1617)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Season  : chr [1:380] "16-17" "16-17" "16-17" "16-17" ...
##  $ Date    : POSIXct[1:380], format: "2016-08-13" "2016-08-13" ...
##  $ HomeTeam: chr [1:380] "Burnley" "Crystal Palace" "Everton" "Hull" ...
##  $ HomeAway: chr [1:380] "Swansea" "West Brom" "Tottenham" "Leicester" ...
##  $ FTHG    : num [1:380] 0 0 1 2 2 1 1 3 1 2 ...
##  $ FTAG    : num [1:380] 1 1 1 1 1 1 1 4 3 1 ...
##  $ Referee : chr [1:380] "J Moss" "C Pawson" "M Atkinson" "M Dean" ...
##  $ HS      : num [1:380] 10 14 12 14 16 12 24 9 9 16 ...
##  $ AS      : num [1:380] 17 13 13 18 7 12 5 16 11 7 ...
##  $ HST     : num [1:380] 3 4 6 5 4 2 6 5 3 6 ...
```

```
##  $ AST     : num [1:380] 9 3 4 5 3 1 1 7 7 3 ...
##  $ HF      : num [1:380] 10 12 10 8 11 18 8 13 7 16 ...
##  $ AF      : num [1:380] 14 15 14 17 14 14 12 17 10 16 ...
##  $ HC      : num [1:380] 7 3 5 5 9 9 6 5 4 7 ...
##  $ AC      : num [1:380] 4 6 6 3 6 6 2 4 2 1 ...
##  $ HY      : num [1:380] 3 2 0 2 1 3 1 3 0 5 ...
##  $ AY      : num [1:380] 2 2 0 2 2 5 2 3 1 2 ...
##  $ HR      : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
##  $ AR      : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
```

A continuación vamos a personalizar el nombre y tipo de variables del archivo season-1617.xlsx

```
season1617.2<-read_excel("season-1617.xlsx",sheet = 1, skip = 1,  col_names =
c("Temp", "Fecha","E.local", "E.visitante", "GLTC", "GVTC", "Arbitro", "DL",
"DV", "DAL", "DAV", "FL", "FV", "FLL", "FLV", "TAL", "TAV", "TRL", "TRV"),
col_types = c("text", "date", "text", "text","numeric","numeric","skip", "skip",
"skip", "skip", "text", "numeric","numeric","numeric","numeric","numeric","numeric"
,"numeric","numeric","numeric","numeric","numeric","numeric"))
str(season1617.2)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Temp       : chr [1:380] "16-17" "16-17" "16-17" "16-17" ...
##  $ Fecha      : POSIXct[1:380], format: "2016-08-13" "2016-08-13" ...
##  $ E.local    : chr [1:380] "Burnley" "Crystal Palace" "Everton" "Hull" ...
##  $ E.visitante: chr [1:380] "Swansea" "West Brom" "Tottenham" "Leicester" ...
##  $ GLTC       : num [1:380] 0 0 1 2 2 1 1 3 1 2 ...
##  $ GVTC       : num [1:380] 1 1 1 1 1 1 1 4 3 1 ...
##  $ Arbitro    : chr [1:380] "J Moss" "C Pawson" "M Atkinson" "M Dean" ...
##  $ DL         : num [1:380] 10 14 12 14 16 12 24 9 9 16 ...
##  $ DV         : num [1:380] 17 13 13 18 7 12 5 16 11 7 ...
##  $ DAL        : num [1:380] 3 4 6 5 4 2 6 5 3 6 ...
##  $ DAV        : num [1:380] 9 3 4 5 3 1 1 7 7 3 ...
##  $ FL         : num [1:380] 10 12 10 8 11 18 8 13 7 16 ...
##  $ FV         : num [1:380] 14 15 14 17 14 14 12 17 10 16 ...
##  $ FLL        : num [1:380] 7 3 5 5 9 9 6 5 4 7 ...
##  $ FLV        : num [1:380] 4 6 6 3 6 6 2 4 2 1 ...
##  $ TAL        : num [1:380] 3 2 0 2 1 3 1 3 0 5 ...
##  $ TAV        : num [1:380] 2 2 0 2 2 5 2 3 1 2 ...
##  $ TRL        : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
##  $ TRV        : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
```

Para el archivo season-1718.xlsx

```
season1718<-read_excel("season-1718.xlsx",col_names=c("Season","Date","HomeTeam"
,"HomeAway","FTHG","FTAG","Referee","HS","AS","HST","AST","HF","AF","HC","AC",
"HY","AY","HR", "AR") ,col_types = c("text", "date", "text", "text", "numeric",
"numeric","skip", "skip", "skip", "skip", "text", "numeric","numeric","numeric",
"numeric","numeric","numeric","numeric","numeric","numeric","numeric","numeric",
"numeric"), skip=1)
str(season1718)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Season  : chr [1:380] "17-18" "17-18" "17-18" "17-18" ...
##  $ Date    : POSIXct[1:380], format: "2017-08-11" "2017-08-12" ...
##  $ HomeTeam: chr [1:380] "Arsenal" "Brighton" "Chelsea" "Crystal Palace" ...
##  $ HomeAway: chr [1:380] "Leicester" "Man City" "Burnley" "Huddersfield" ...
##  $ FTHG    : num [1:380] 4 0 2 0 1 0 3 1 4 0 ...
```

```
##  $ FTAG    : num [1:380] 3 2 3 3 0 0 3 0 0 2 ...
##  $ Referee : chr [1:380] "M Dean" "M Oliver" "C Pawson" "J Moss" ...
##  $ HS      : num [1:380] 27 6 19 14 9 29 9 16 22 6 ...
##  $ AS      : num [1:380] 6 14 10 8 9 4 14 9 9 18 ...
##  $ HST     : num [1:380] 10 2 6 4 4 2 4 6 6 3 ...
##  $ AST     : num [1:380] 3 4 5 6 1 0 5 2 1 6 ...
##  $ HF      : num [1:380] 9 6 16 7 13 10 14 15 19 6 ...
##  $ AF      : num [1:380] 12 9 11 19 10 13 8 3 7 10 ...
##  $ HC      : num [1:380] 9 3 8 12 6 13 3 8 11 5 ...
##  $ AC      : num [1:380] 4 10 5 9 7 0 3 2 1 7 ...
##  $ HY      : num [1:380] 0 0 3 1 1 2 0 3 2 1 ...
##  $ AY      : num [1:380] 1 2 3 3 1 1 3 1 2 2 ...
##  $ HR      : num [1:380] 0 0 2 0 0 0 0 0 0 1 ...
##  $ AR      : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
```

A continuación vamos a personalizar el nombre y tipo de variables del archivo season-1718.xlsx

```
season1718.2<-read_excel("season-1718.xlsx",sheet = 1, skip = 1,  col_names =
c("Temp", "Fecha","E.local", "E.visitante", "GLTC", "GVTC", "Arbitro", "DL",
"DV", "DAL", "DAV", "FL", "FV", "FLL", "FLV", "TAL", "TAV", "TRL", "TRV"),
col_types = c("text", "date", "text", "text","numeric","numeric","skip", "skip",
"skip", "skip", "text", "numeric","numeric","numeric","numeric","numeric","numeric"
,"numeric","numeric","numeric","numeric","numeric","numeric"))
str(season1718.2)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Temp       : chr [1:380] "17-18" "17-18" "17-18" "17-18" ...
##  $ Fecha      : POSIXct[1:380], format: "2017-08-11" "2017-08-12" ...
##  $ E.local    : chr [1:380] "Arsenal" "Brighton" "Chelsea" "Crystal Palace" ...
##  $ E.visitante: chr [1:380] "Leicester" "Man City" "Burnley" "Huddersfield" ...
##  $ GLTC       : num [1:380] 4 0 2 0 1 0 3 1 4 0 ...
##  $ GVTC       : num [1:380] 3 2 3 3 0 0 3 0 0 2 ...
##  $ Arbitro    : chr [1:380] "M Dean" "M Oliver" "C Pawson" "J Moss" ...
##  $ DL         : num [1:380] 27 6 19 14 9 29 9 16 22 6 ...
##  $ DV         : num [1:380] 6 14 10 8 9 4 14 9 9 18 ...
##  $ DAL        : num [1:380] 10 2 6 4 4 2 4 6 6 3 ...
##  $ DAV        : num [1:380] 3 4 5 6 1 0 5 2 1 6 ...
##  $ FL         : num [1:380] 9 6 16 7 13 10 14 15 19 6 ...
##  $ FV         : num [1:380] 12 9 11 19 10 13 8 3 7 10 ...
##  $ FLL        : num [1:380] 9 3 8 12 6 13 3 8 11 5 ...
##  $ FLV        : num [1:380] 4 10 5 9 7 0 3 2 1 7 ...
##  $ TAL        : num [1:380] 0 0 3 1 1 2 0 3 2 1 ...
##  $ TAV        : num [1:380] 1 2 3 3 1 1 3 1 2 2 ...
##  $ TRL        : num [1:380] 0 0 2 0 0 0 0 0 0 1 ...
##  $ TRV        : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
```

Para el archivo season-1819.xlsx

```
season1819<-read_excel("season-1819.xlsx",col_names=c("Season","Date","HomeTeam"
,"HomeAway","FTHG","FTAG","Referee","HS","AS","HST","AST","HF","AF","HC","AC",
"HY","AY","HR", "AR") ,col_types = c("text", "date", "text", "text", "numeric",
"numeric","skip", "skip", "skip", "skip", "text", "numeric","numeric","numeric",
"numeric","numeric","numeric","numeric","numeric","numeric","numeric","numeric",
"numeric"), skip=1)
str(season1819)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
```

```
##  $ Season  : chr [1:380] "18-19" "18-19" "18-19" "18-19" ...
##  $ Date    : POSIXct[1:380], format: "2018-08-10" "2018-08-11" ...
##  $ HomeTeam: chr [1:380] "Man United" "Bournemouth" "Fulham" "Huddersfield" ...
##  $ HomeAway: chr [1:380] "Leicester" "Cardiff" "Crystal Palace" "Chelsea" ...
##  $ FTHG    : num [1:380] 2 2 0 0 1 2 2 0 4 0 ...
##  $ FTAG    : num [1:380] 1 0 2 3 2 0 2 2 0 0 ...
##  $ Referee : chr [1:380] "A Marriner" "K Friend" "M Dean" "C Kavanagh" ...
##  $ HS      : num [1:380] 8 12 15 6 15 19 11 9 18 18 ...
##  $ AS      : num [1:380] 13 10 10 13 15 6 6 17 5 16 ...
##  $ HST     : num [1:380] 6 4 6 1 2 5 4 3 8 3 ...
##  $ AST     : num [1:380] 4 1 9 4 5 0 5 8 2 6 ...
##  $ HF      : num [1:380] 11 11 9 9 11 10 8 11 14 10 ...
##  $ AF      : num [1:380] 8 9 11 8 12 16 7 14 9 9 ...
##  $ HC      : num [1:380] 2 7 5 2 3 8 3 2 5 8 ...
##  $ AC      : num [1:380] 5 4 5 5 5 2 6 9 4 5 ...
##  $ HY      : num [1:380] 2 1 1 2 2 2 0 2 1 0 ...
##  $ AY      : num [1:380] 1 1 2 1 2 2 1 2 2 1 ...
##  $ HR      : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
##  $ AR      : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
```

A continuación vamos a personalizar el nombre y tipo de variables del archivo season-1819.xlsx

```
season1819.2<-read_excel("season-1819.xlsx",sheet = 1, skip = 1,  col_names =
c("Temp", "Fecha","E.local", "E.visitante", "GLTC", "GVTC", "Arbitro", "DL",
"DV", "DAL", "DAV", "FL", "FV", "FLL", "FLV", "TAL", "TAV", "TRL", "TRV"),
col_types = c("text", "date", "text", "text","numeric","numeric","skip", "skip",
"skip", "skip", "text", "numeric","numeric","numeric","numeric","numeric","numeric"
,"numeric","numeric","numeric","numeric","numeric","numeric"))
str(season1819.2)
```

```
## tibble [380 x 19] (S3: tbl_df/tbl/data.frame)
##  $ Temp       : chr [1:380] "18-19" "18-19" "18-19" "18-19" ...
##  $ Fecha      : POSIXct[1:380], format: "2018-08-10" "2018-08-11" ...
##  $ E.local    : chr [1:380] "Man United" "Bournemouth" "Fulham" "Huddersfield" ...
##  $ E.visitante: chr [1:380] "Leicester" "Cardiff" "Crystal Palace" "Chelsea" ...
##  $ GLTC       : num [1:380] 2 2 0 0 1 2 2 0 4 0 ...
##  $ GVTC       : num [1:380] 1 0 2 3 2 0 2 2 0 0 ...
##  $ Arbitro    : chr [1:380] "A Marriner" "K Friend" "M Dean" "C Kavanagh" ...
##  $ DL         : num [1:380] 8 12 15 6 15 19 11 9 18 18 ...
##  $ DV         : num [1:380] 13 10 10 13 15 6 6 17 5 16 ...
##  $ DAL        : num [1:380] 6 4 6 1 2 5 4 3 8 3 ...
##  $ DAV        : num [1:380] 4 1 9 4 5 0 5 8 2 6 ...
##  $ FL         : num [1:380] 11 11 9 9 11 10 8 11 14 10 ...
##  $ FV         : num [1:380] 8 9 11 8 12 16 7 14 9 9 ...
##  $ FLL        : num [1:380] 2 7 5 2 3 8 3 2 5 8 ...
##  $ FLV        : num [1:380] 5 4 5 5 5 2 6 9 4 5 ...
##  $ TAL        : num [1:380] 2 1 1 2 2 2 0 2 1 0 ...
##  $ TAV        : num [1:380] 1 1 2 1 2 2 1 2 2 1 ...
##  $ TRL        : num [1:380] 0 0 0 0 0 0 0 0 0 0 ...
##  $ TRV        : num [1:380] 0 0 0 0 0 0 1 0 0 0 ...
```

De acuerdo a la información que nos brinda R con la función str() podemos ver que todos los conjuntos de datos cuentan con 380 registros cada uno.

Ejercicio IMDb

Para desarrollar este ejercicio usaremos la libreria "readr" por lo cual el primer paso es cargarla.

```
library(readr)
```

Para poder cargar los archivos sin descargarlos y descomprimirlos previamente usaremos la función read_delim() de la librería.

A continuación importamos los datos de title.basics, solamente las primeras 5000 observaciones y especificando las columnas como caracter, factor, caracter, caracter, lógico, fecha en año, fecha en año, tiempo en minutos y factor respectivamente.

```
tbasics<-read_delim('https://datasets.imdbws.com/title.basics.tsv.gz',n_max=5000,
                    col_types = cols('c','f','c','c','l',col_date('%Y'),col_date('%Y'),
                                     col_time('%M'),'f'))
```

```
## Warning: One or more parsing issues, see `problems()` for details
```

```
str(tbasics)
```

```
## spec_tbl_df [5,000 x 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ tconst        : chr [1:5000] "tt0000001" "tt0000002" "tt0000003" "tt0000004" ...
##  $ titleType     : Factor w/ 2 levels "short","movie": 1 1 1 1 1 1 1 1 2 1 ...
##  $ primaryTitle  : chr [1:5000] "Carmencita" "Le clown et ses chiens" "Pauvre Pierrot" "Un bon bock"
##  $ originalTitle : chr [1:5000] "Carmencita" "Le clown et ses chiens" "Pauvre Pierrot" "Un bon bock"
##  $ isAdult       : logi [1:5000] FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ startYear     : Date[1:5000], format: "1894-01-01" "1892-01-01" ...
##  $ endYear       : Date[1:5000], format: NA NA ...
##  $ runtimeMinutes: 'hms' num [1:5000] NA NA NA 00:12:00 ...
##   ..- attr(*, "units")= chr "secs"
##  $ genres        : Factor w/ 179 levels "Documentary,Short",..: 1 2 3 2 4 5 6 1 7 1 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   tconst = col_character(),
##   ..   titleType = col_factor(levels = NULL, ordered = FALSE, include_na = FALSE),
##   ..   primaryTitle = col_character(),
##   ..   originalTitle = col_character(),
##   ..   isAdult = col_logical(),
##   ..   startYear = col_date(format = "%Y"),
##   ..   endYear = col_date(format = "%Y"),
##   ..   runtimeMinutes = col_time(format = "%M"),
##   ..   genres = col_factor(levels = NULL, ordered = FALSE, include_na = FALSE)
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Este conjunto de datos cuenta con 9 variables y sus formatos son: Texto, Factor, Texto, Texto, Valor Lógico, Fecha, Fecha, Fecha y factor.

Ahora vamos a cargar las primeras 5000 observaciones de title.principals omitiendo la variable ordering y tomando las demás columnas como texto.

```
tprincipals<-read_delim('https://datasets.imdbws.com/title.principals.tsv.gz',n_max=5000,
                        col_types=cols('c','-','c','c','c','c'))

str(tprincipals)
```

```
## spec_tbl_df [5,000 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ tconst    : chr [1:5000] "tt0000001" "tt0000001" "tt0000001" "tt0000002" ...
```

```
##  $ nconst    : chr [1:5000] "nm1588970" "nm0005690" "nm0374658" "nm0721526" ...
##  $ category  : chr [1:5000] "self" "director" "cinematographer" "director" ...
##  $ job       : chr [1:5000] "\\N" "\\N" "director of photography" "\\N" ...
##  $ characters: chr [1:5000] "[\"Self\"]" "\\N" "\\N" "\\N" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   tconst = col_character(),
##   ..   ordering = col_skip(),
##   ..   nconst = col_character(),
##   ..   category = col_character(),
##   ..   job = col_character(),
##   ..   characters = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Este conjunto de datos tiene 5 variables todas en formato texto.

Ahora vamos a cargar las primeras 5000 observaciones de name.basics omitiendo la variable knownForTitles y tomando las demás columnas como caracter, caracter, fecha en año, fecha en año y caracter respectivamente.

```
nbasics<-read_delim('https://datasets.imdbws.com/name.basics.tsv.gz',n_max=5000,
                    col_types=cols('c','c',col_date('%Y'),col_date('%Y'),'c','-'))
```

```
## Warning: One or more parsing issues, see `problems()` for details
```

```
str(nbasics)
```

```
## spec_tbl_df [5,000 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ nconst           : chr [1:5000] "nm0000001" "nm0000002" "nm0000003" "nm0000004" ...
##  $ primaryName      : chr [1:5000] "Fred Astaire" "Lauren Bacall" "Brigitte Bardot" "John Belushi" .
##  $ birthYear        : Date[1:5000], format: "1899-01-01" "1924-01-01" ...
##  $ deathYear        : Date[1:5000], format: "1987-01-01" "2014-01-01" ...
##  $ primaryProfession: chr [1:5000] "soundtrack,actor,miscellaneous" "actress,soundtrack" "actress,sou
##  - attr(*, "spec")=
##   .. cols(
##   ..   nconst = col_character(),
##   ..   primaryName = col_character(),
##   ..   birthYear = col_date(format = "%Y"),
##   ..   deathYear = col_date(format = "%Y"),
##   ..   primaryProfession = col_character(),
##   ..   knownForTitles = col_skip()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Este conjunto de datos tiene 5 Variables en los formatos: Texto, Texto, Fecha, Fecha y Texto.

Ejercicio Shows en Netflix

Para este ejercicio nuevamente usamos la función read_delim() de la libreria "readr" por lo que la cargamos primero.

```
library(readr)
```

Posteriormente descargamos, descomprimimos y guardamos el archivo en el directorio de trabajo, para importarlo usando read_delim(), dandole títulos personalizados y especificando el tipo de columnas como caracter, factor, caracter, caracter, caracter, caracter, fecha en formato mes dia año, fecha en año, factor, caracter y caracter respectivamente.

```
Titles<-read_delim("netflix_titles.csv",skip = 1,col_names = c("IDShow","Tipo","Titulo","Director","Ele
                  col_types= cols("c","f","c","c","c","c",col_date("%B %d, %Y"),col_date("%Y"),"f","c",
str(Titles)
```

```
## spec_tbl_df [8,807 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ IDShow              : chr [1:8807] "s1" "s2" "s3" "s4" ...
##  $ Tipo                : Factor w/ 2 levels "Movie","TV Show": 1 2 2 2 2 2 1 1 2 1 ...
##  $ Titulo              : chr [1:8807] "Dick Johnson Is Dead" "Blood & Water" "Ganglands" "Jailbirds
##  $ Director            : chr [1:8807] "Kirsten Johnson" NA "Julien Leclercq" NA ...
##  $ Elenco              : chr [1:8807] NA "Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Di
##  $ País                : chr [1:8807] "United States" "South Africa" NA NA ...
##  $ Fecha de ingreso    : Date[1:8807], format: "2021-09-25" "2021-09-24" ...
##  $ Fecha de publicación: Date[1:8807], format: "2020-01-01" "2021-01-01" ...
##  $ Puntuación          : Factor w/ 18 levels "PG-13","TV-MA",..: 1 2 2 2 2 2 3 2 4 1 ...
##  $ Duración            : chr [1:8807] "90 min" "2 Seasons" "1 Season" "1 Season" ...
##  $ Listado en          : chr [1:8807] "Documentaries" "International TV Shows, TV Dramas, TV Mysteri
##  $ Sinópsis            : chr [1:8807] "As her father nears the end of his life, filmmaker Kirsten Jo
##  - attr(*, "spec")=
##   .. cols(
##   ..   IDShow = col_character(),
##   ..   Tipo = col_factor(levels = NULL, ordered = FALSE, include_na = FALSE),
##   ..   Titulo = col_character(),
##   ..   Director = col_character(),
##   ..   Elenco = col_character(),
##   ..   País = col_character(),
##   ..   `Fecha de ingreso` = col_date(format = "%B %d, %Y"),
##   ..   `Fecha de publicación` = col_date(format = "%Y"),
##   ..   Puntuación = col_factor(levels = NULL, ordered = FALSE, include_na = FALSE),
##   ..   Duración = col_character(),
##   ..   `Listado en` = col_character(),
##   ..   Sinópsis = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

El conjunto de datos cuenta con 8807 registros.