

Análisis Champions League 2021 2022

Isaias Arcos, Nicolás Bejarano, Mateo Erazo

12/12/2023



En la temporada 2021-2022 al igual que todos los años se disputó la competencia más importante a nivel clubes la *UEFA Champions League*, la competencia más exigente y esperada del año. El analizar estos datos de esta competencia es importante debido a que los mejores futbolistas participan en la disputa de este trofeo.

Tratamiento previo

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.3      v readr      2.1.4  
## v forcats    1.0.0      v stringr   1.5.0  
## v ggplot2    3.4.3      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.0  
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
```

```
library(kableExtra)
```

```

## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'

##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##      group_rows
library(corrplot)

## corrplot 0.92 loaded

attacking <- read.csv("data/attacking.csv")
attempts <- read.csv("data/attempts.csv")
defending <- read.csv("data/defending.csv")
goalkeeper <- read.csv("data/goalkeeping.csv")
midfielders <- read.csv("data/distributon.csv")
goals <- read.csv("data/goals.csv")

# cambio de nombres dataset portero
goalkeeper <- goalkeeper %>%
  rename(
    Serial = serial,
    Nombre_jugador = player_name,
    Equipo = club,
    Posicion = position,
    Paradas = saved,
    Goles_concedidos = conceded,
    Penales_parados = saved_penalties,
    Partidos_sin_goles = cleansheets,
    #Despejes = punches_made,
    Partidos_jugados = match_played
  )

# cambio de nombres dataset defensa
defending <- defending %>%
  rename(
    serial = serial,
    Nombre_jugador = player_name,
    Equipo = club,
    Posicion = position,
    Balones_recuperados = balls_recoverd,
    Entradas = tackles,
    Entradas_ganadas = t_won,
    Entradas_perdidas = t_lost,
    Despejes_intentados = clearance_attempted,
    Partidos_jugados = match_played
  )

# medio campo
midfielders <- midfielders %>%
  rename(

```

```

    serial = serial,
    Nombre_jugador = player_name,
    Equipo = club,
    Posicion = position,
    Precisión_pase = pass_accuracy,
    Pases_intentados = pass_attempted,
    Pases_completados = pass_completed,
    Precisión_centro = cross_accuracy,
    Centros_intentados = cross_attempted,
    Centros_completados = cross_complted
  )

# renombrar variables de goles
goals <- goals %>%
  rename(
    serial = serial,
    nombre_jugador = player_name,
    equipo = club,
    posicion = position,
    goles = goals,
    pie_derecho = right_foot,
    pie_izquierdo = left_foot,
    cabezazos = headers,
    otros = others,
    dentro_area = inside_area
  )

```

El fútbol es un deporte colectivo y de altísima complejidad por lo que es importante analizar todas las fasetas del juego:

- Arco
- Defensa
- Medio campo
- Ataque

Análisis de porteros

Análisis de datos

Las salvadas así como intervenciones importantes prácticamente son un “gol” para el equipo, por lo que si el arquero está en un gran nivel sin duda ayuda mucho en una competición corta como esta.

```

top_goalkeepers <- goalkeeper %>%
  arrange(desc(Paradas)) %>%
  head(10)

ggplot(top_goalkeepers, aes(x=Nombre_jugador, y=Paradas, fill=Equipo)) +
  geom_bar(stat="identity", position="dodge") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title="Paradas realizadas por Portero (Top 10)",
       x="Portero",
       y="Número de Paradas") +
  theme(legend.title=element_blank()) # Esto retira la leyenda

```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```



```

## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for
## <9f>

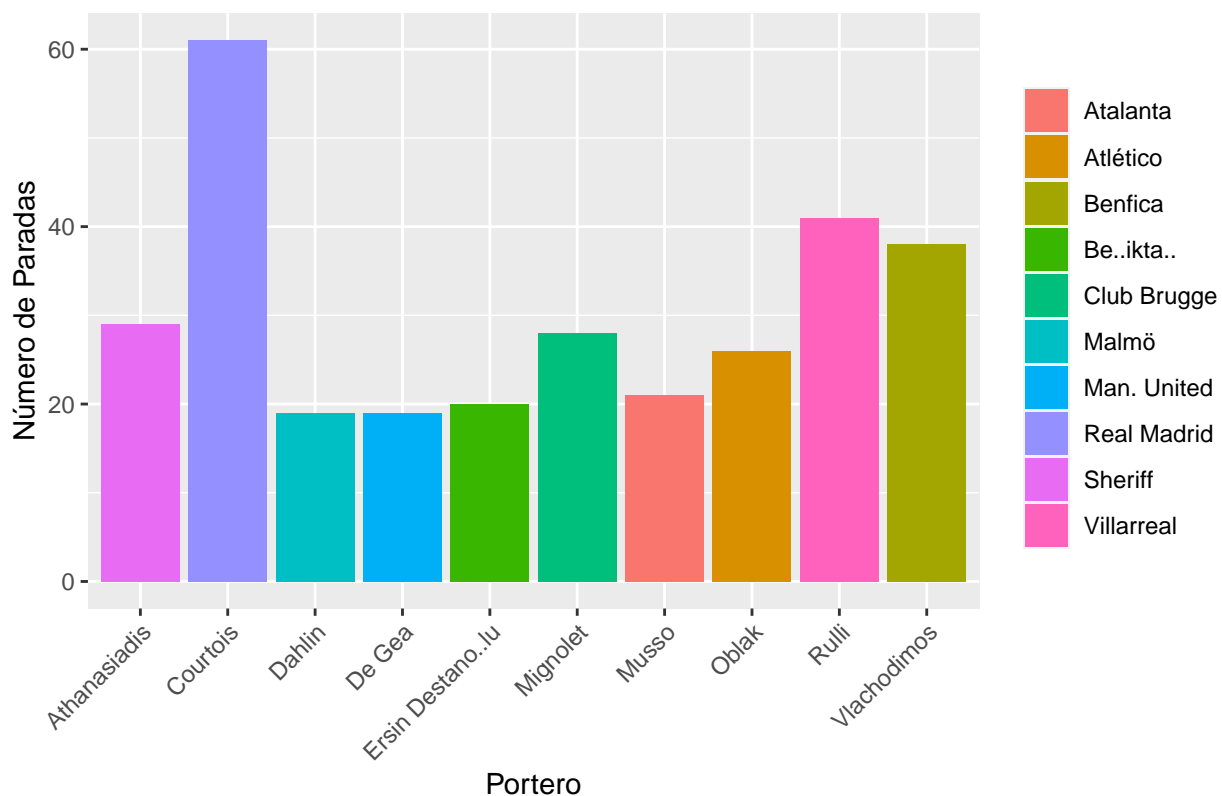
```


[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>
```

Paradas realizadas por Portero (Top 10)



Vemos claramente que Courtois es el claro ganador, pero esto se puede deber a que jugó más partidos, vamos a realizar un análisis más profundo, para medir el rendimiento de los porteros dependiendo de la cantidad de

participaciones.

```
# creamos una nueva métrica
goalkeeper <- goalkeeper %>%
  mutate(Paradas_por_partido = Paradas / Partidos_jugados) %>%
  arrange(desc(Paradas_por_partido)) %>%
  head(10)

ggplot(goalkeeper, aes(x=Nombre_jugador, y=Paradas_por_partido, fill=Equipo)) +
  geom_bar(stat="identity", position="dodge") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title="Paradas por Partido por Portero (Top 10)",
       x="Portero",
       y="Paradas por Partido") +
  theme(legend.title=element_blank())
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>
```

[illegible]

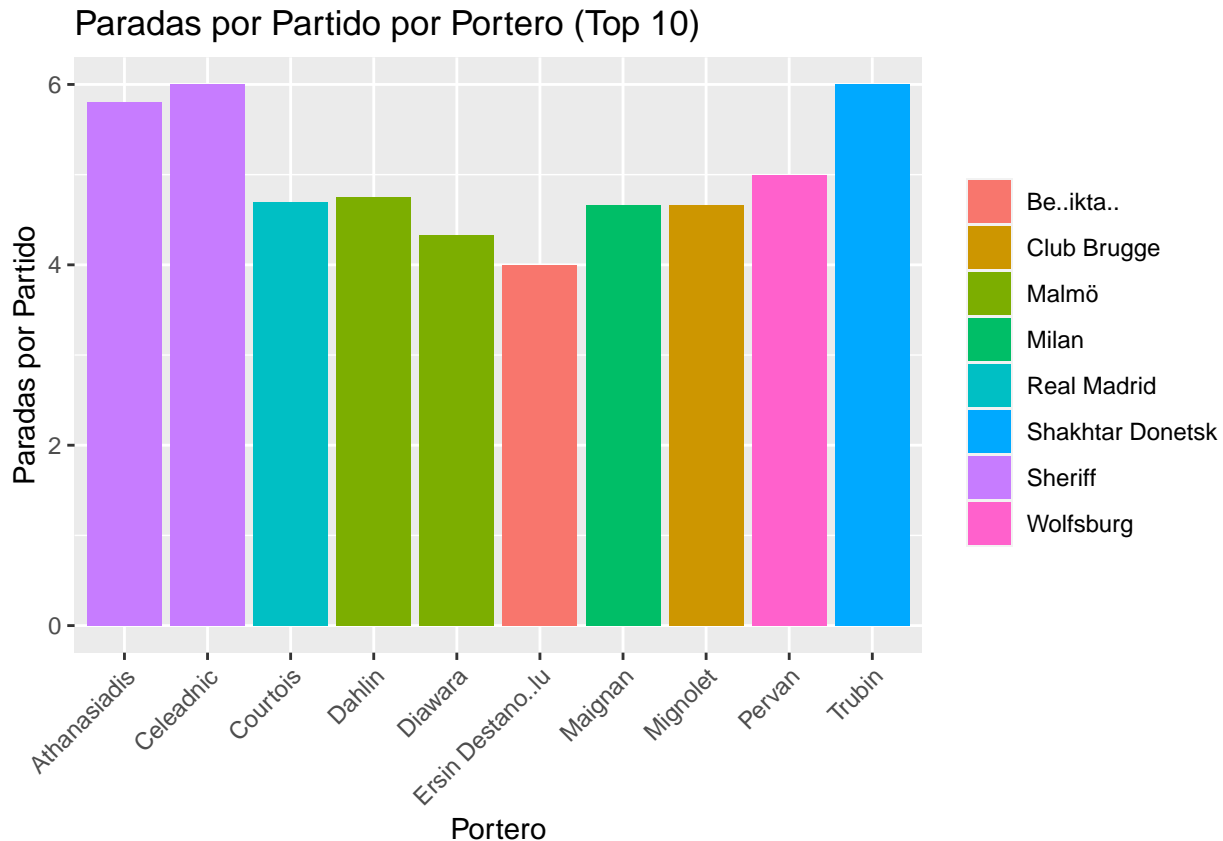
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <c4>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <c4>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <c4>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <c4>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Ersin Destanoğlu' in 'mbcsToSbcs': dot substituted for  
## <c4>
```


[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>
```



```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Beşiktaş' in 'mbcsToSbcs': dot substituted for <9f>
```



Realizando este análisis vemos que aún así Courtois sigue estando entre los mejores sin embargo los porteros *Trubin* del *Shakhtar Donetsk* y *Celeadnic* de *Sheriff* también han tenido un alto rendimiento.

Creación de modelo

Variable dependiente: Partidos sin goles

Nos parece importante predecir cuales serán los factores que influyen en que un partido termine sin goles.

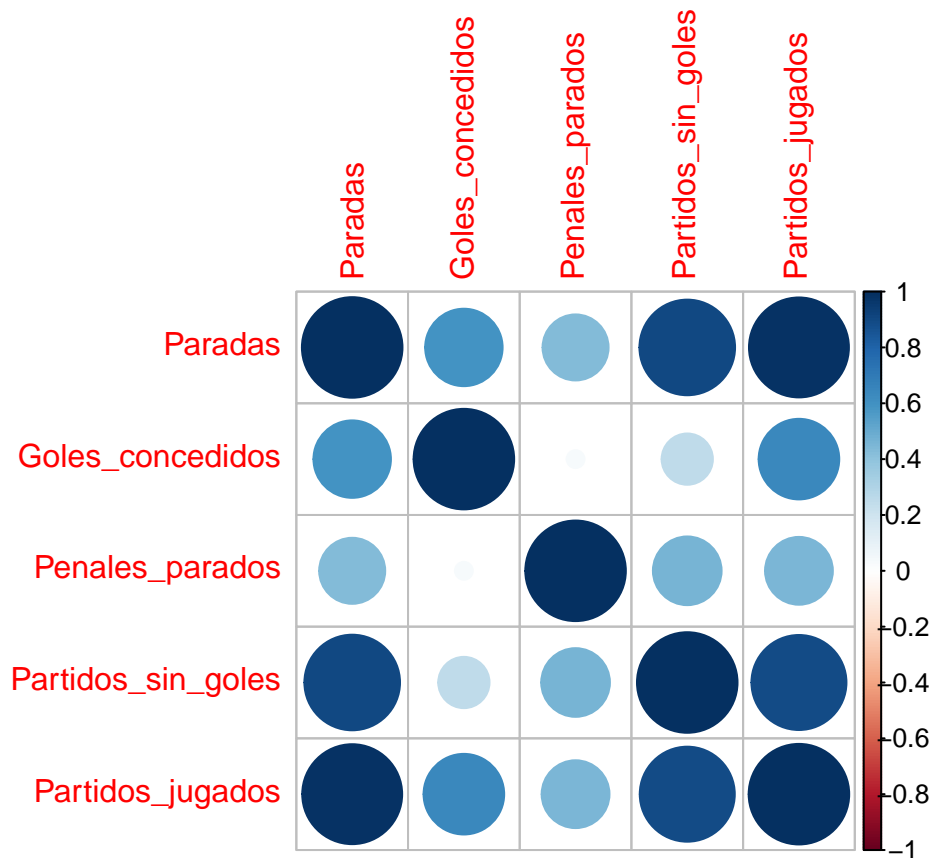
Para ello primero estudiaremos la correlación de variables

```
correlation_matrix <- cor(goalkeeper[, c("Paradas", "Goles_concedidos", "Penales_parados", "Partidos_sin_goles")])
print(correlation_matrix)
```

```
##               Paradas Goles_concedidos Penales_parados
## Paradas         1.0000000      0.5932259      0.4312246
## Goles_concedidos 0.5932259      1.0000000      0.0314357
## Penales_parados  0.4312246      0.0314357      1.0000000
## Partidos_sin_goles 0.9045057      0.2598798      0.4665695
## Partidos_jugados  0.9878259      0.6425767      0.4549131
##               Partidos_sin_goles Partidos_jugados
## Paradas         0.9045057      0.9878259
## Goles_concedidos 0.2598798      0.6425767
## Penales_parados  0.4665695      0.4549131
## Partidos_sin_goles 1.0000000      0.8930827
## Partidos_jugados  0.8930827      1.0000000
```

Visualización de la matriz de correlación

```
corrplot(correlation_matrix, method = "circle")
```



Podemos ver que existe alguna relación entre partidos jugados y paradas que nos podrían ayudar a entender la relación. Sin embargo vamos a utilizar psych para poder identificar mejor las relaciones mediante dispersión y distribuciones.

```
library(psych)
```

```
##
```

```
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
```

```
##
```

```
## %+%, alpha
```

```
# Subconjunto de datos
```

```
subset_data <- goalkeeper[, c("Paradas", "Goles_concedidos", "Penales_parados", "Partidos_sin_goles", "Partidos_jugados")]
```

```
# Crea el gráfico de dispersión
```

```
pairs.panels(subset_data,
```

```
  main = "Gráficos de Dispersión con Títulos",
```

```
  cex.main = 1.2, # Tamaño del título principal
```

```
  cex.labels = 1.2, # Tamaño de las etiquetas de variables
```

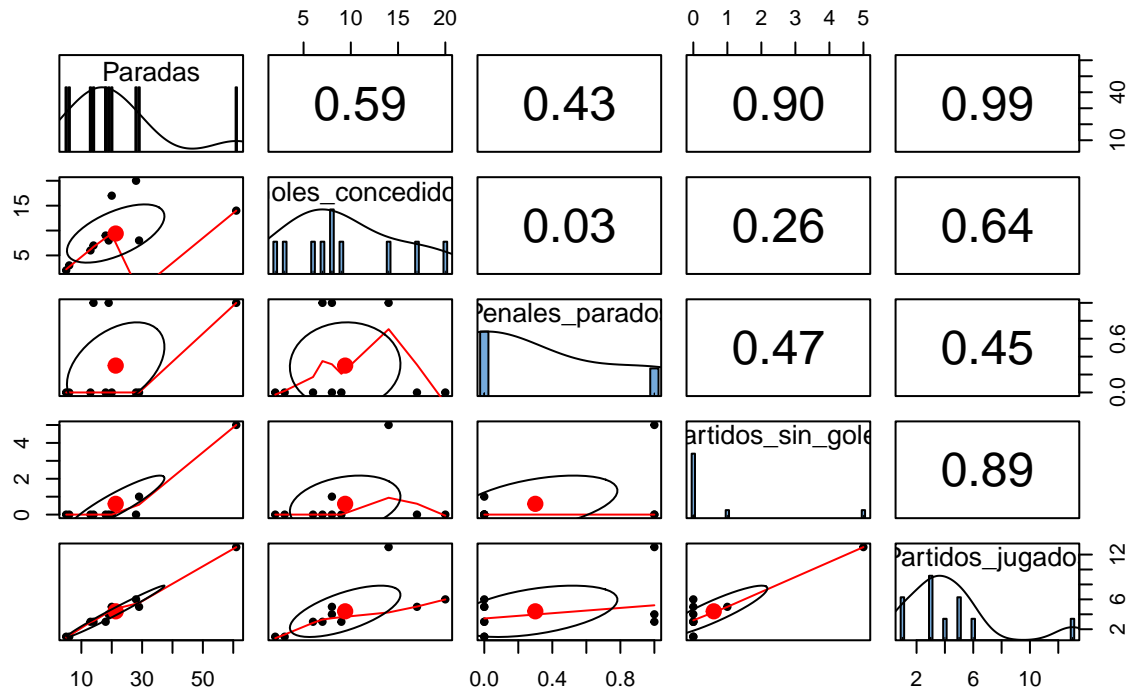
```
  method = "pearson", # Método de correlación
```

```
  hist.col = "#75AADB", # Color de los histogramas
```

```
  density = TRUE, # Mostrar densidades en los histogramas
```

```
  ellipses = TRUE) # Mostrar elipse de confianza
```

Gráficos de Dispersión con Títulos



Como podemos ver observar la librería incluso nos marca que existe la relación entre las variables Paradas, Goles concedidos y Cantidad de Partidos Jugados para predecir la Cantidad de Partidos sin goles. Observamos también que existe un distribución que no es normal y tiende a aproximarse a una poca cantidad de partidos sin goles.

```
vars <- c("Paradas", "Goles_concedidos", "Partidos_jugados", "Partidos_sin_goles")
subset_data <- goalkeeper[, vars]
```

```
# Crea el modelo de regresión lineal múltiple
```

```
modelo_arco <- lm(Partidos_sin_goles ~ Paradas + Goles_concedidos + Partidos_jugados, data = subset_data)
```

```
# Muestra un resumen del modelo
```

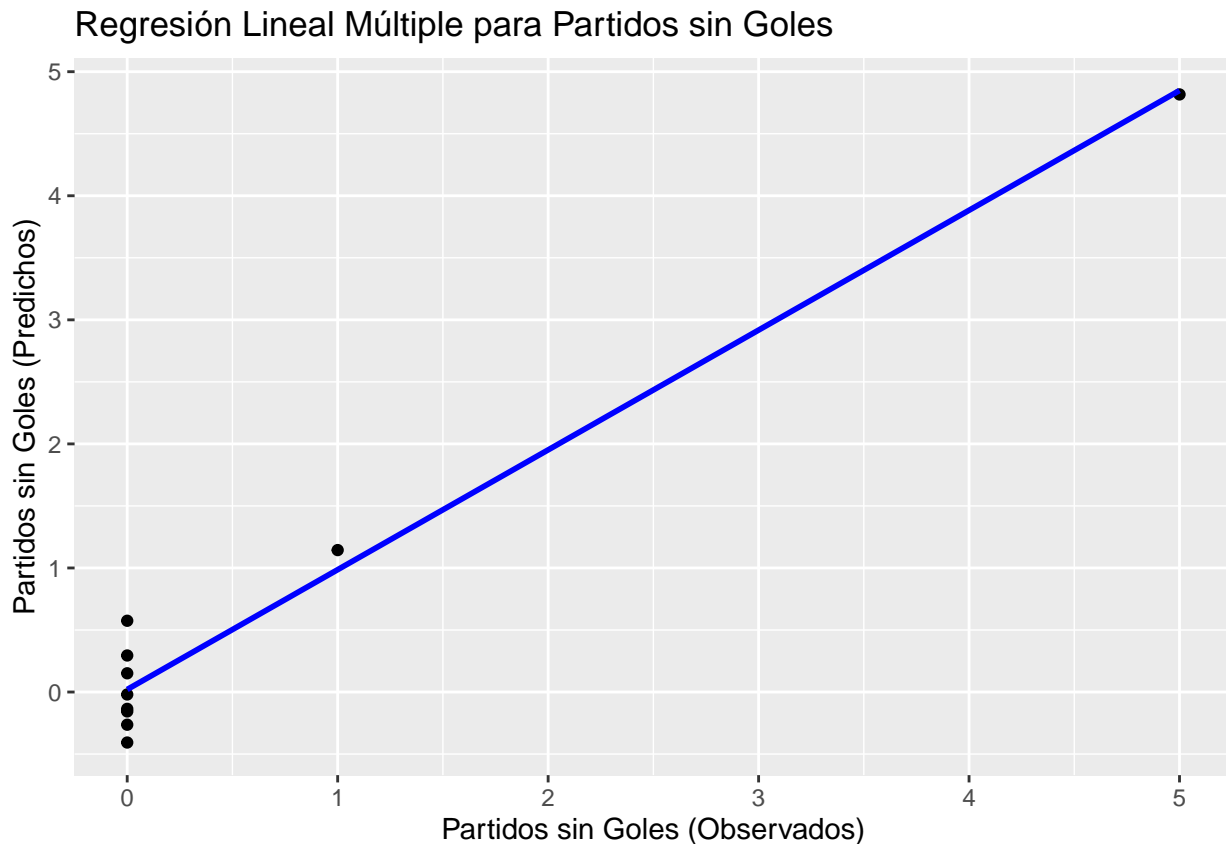
```
summary(modelo_arco)
```

```
##
## Call:
## lm(formula = Partidos_sin_goles ~ Paradas + Goles_concedidos +
##     Partidos_jugados, data = subset_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57456 -0.14933  0.07749  0.17644  0.40693
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.542667   0.246150  -2.205   0.0697 .
## Paradas       0.000394   0.051075   0.008   0.9941
## Goles_concedidos -0.144062   0.028442  -5.065   0.0023 **
## Partidos_jugados  0.565559   0.250584   2.257   0.0648 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.3587 on 6 degrees of freedom
## Multiple R-squared:  0.9655, Adjusted R-squared:  0.9483
## F-statistic: 56.02 on 3 and 6 DF,  p-value: 8.843e-05

# Graficar los valores observados vs. predichos
ggplot(subset_data, aes(x = Partidos_sin_goles, y = predict(modelo_arco))) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Regresión Lineal Múltiple para Partidos sin Goles",
        x = "Partidos sin Goles (Observados)",
        y = "Partidos sin Goles (Predichos)")

## `geom_smooth()` using formula = 'y ~ x'
```



Resultados de nuestro modelo de partidos sin goles

Las métricas de los estadísticos F y el p valor, en conjunto nos indican que con un F alto y un p valor bajo, nos permite rechazar la hipótesis nula, esto nos indica que al menos una de las variables que estamos utilizando para predicción de nuestro modelo es altamente significativa para predecir un partido sin goles.

Tenemos un R2 y R2 ajustada alto por lo que el modelo es bueno para explicar la variabilidad de partidos sin goles.

Hemos obtenido un modelo realmente bueno tal como muestran las métricas y el gráfico de las relaciones respectivas.

Análisis de defensa

Primero obtenemos las estadísticas generales de la competición en el apartado defensivo

Recuperaciones de balón

Las recuperaciones de balón son fundamentales pues permiten interrumpir un ataque rival y recuperar la posesión del balón o armar un contra ataque, aquí observamos el **Top 10 jugadores** con mayores recuperaciones de balón.

```
top_balls_recovered <- defending %>%  
  arrange(desc(Balones_recuperados)) %>%  
  head(10) %>%  
  select(Nombre_jugador, Equipo, Balones_recuperados)  
#print(top_balls_recovered)  
kable(top_balls_recovered, "html") %>%  
kable_styling()
```

Nombre_jugador	Equipo	Balones_recuperados
Casemiro	Real Madrid	76
Éder Militão	Real Madrid	76
Reinildo	Atlético	76
Fabinho	Liverpool	74
Albiol	Villarreal	73
Thiago Silva	Chelsea	65
Otamendi	Benfica	63
Capoue		

Villarreal

62

Van Dijk

Liverpool

62

João Cancelo

Man. City

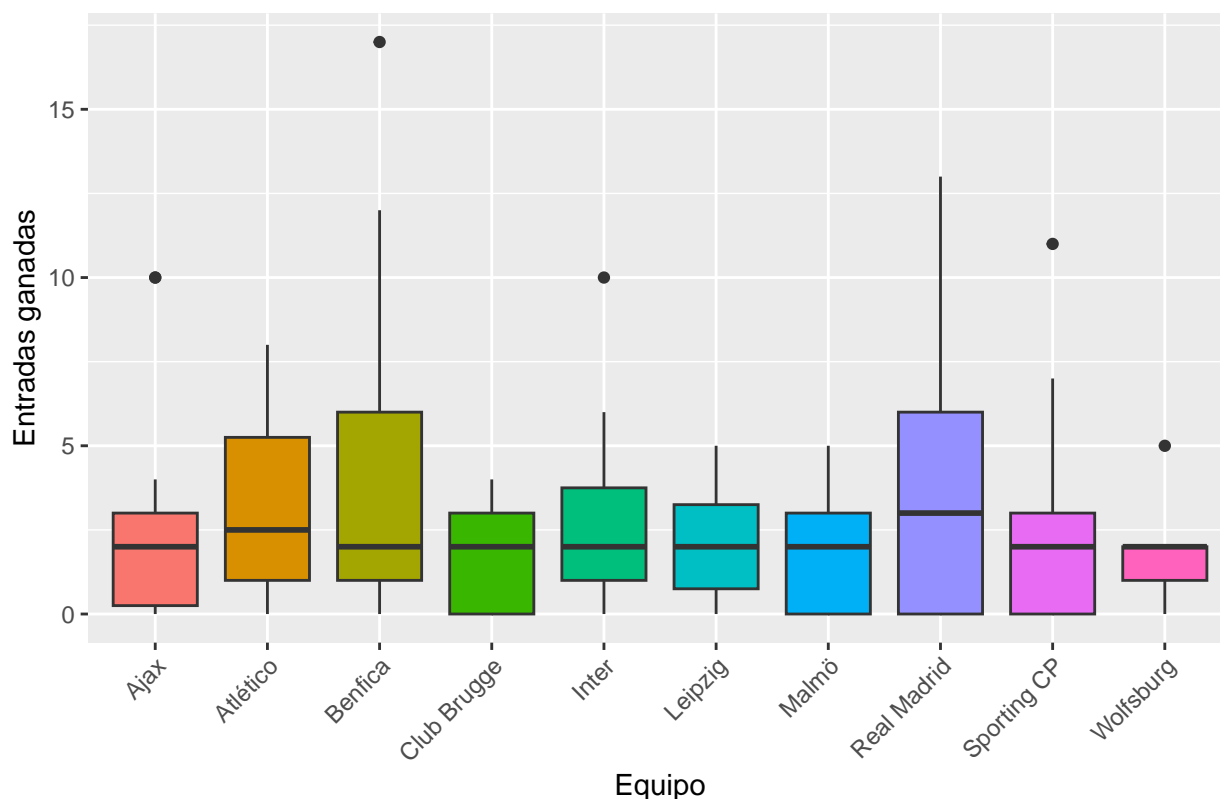
56

Distribución de entradas exitosas por equipo

```
# para evitar la carga cognitiva vamos a filtrar el top 10 equipos con más entradas exitosas
top_teams_success_tackles <- defending %>%
  group_by(Equipo) %>%
  summarize(mediana_entradas = median(Entradas_ganadas)) %>%
  arrange(desc(mediana_entradas)) %>%
  head(10) %>%
  pull(Equipo)

# Diagrama de cajas
defending %>%
  filter(Equipo %in% top_teams_success_tackles) %>%
  ggplot(aes(x=Equipo, y=Entradas_ganadas)) +
  geom_boxplot(aes(fill=Equipo)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title="Distribución de Entradas ganadas por Equipo (Top 10)",
        x="Equipo",
        y="Entradas ganadas") +
  theme(legend.position="none") #elimina leyenda
```

Distribución de Entradas ganadas por Equipo (Top 10)



Se puede observar que el Real Madrid es el claro ganador en cuanto a entradas exitosas, considerando que se está analizando el top 10 vemos que el resto de equipos no llegaron a instancias decisivas de la competición, lo que claramente indica que sin duda el **Real Madrid** tuvo una superioridad defensiva frente a sus rivales, su diagrama de caja muestra una simetría casi perfecta, lo que indica una distribución uniforme en el número de entradas exitosas por partido y nos indica su *consistencia y solidez defensiva*.

Construcción de modelo defensa

Variable dependiente: Entradas o duelos ganados Consideramos que es importante el estudio de esta variable porque permite cancelar opciones de gol y armar posibles contragolpes.

```
# Correlación
correlation_matrix_defensa <- cor(defending[, c("Balones_recuperados", "Entradas", "Entradas_ganadas",
print(correlation_matrix_defensa)
```

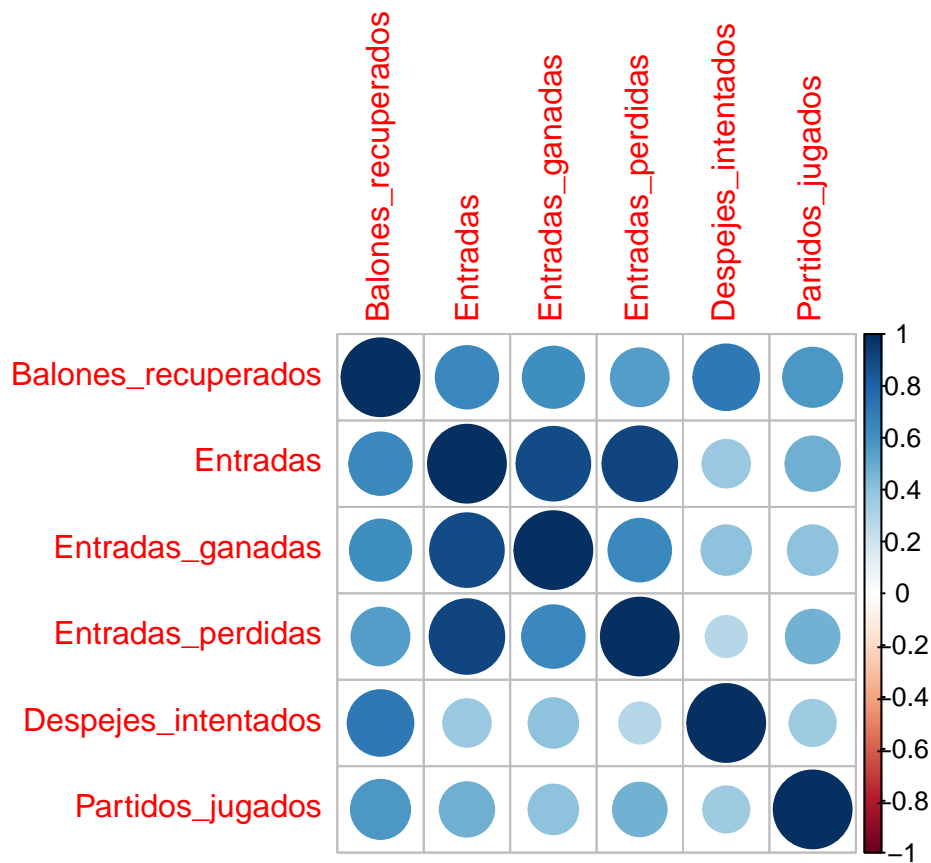
```
##          Balones_recuperados  Entradas  Entradas_ganadas
## Balones_recuperados          1.0000000  0.6436342          0.6189064
## Entradas                     0.6436342  1.0000000          0.8962542
## Entradas_ganadas             0.6189064  0.8962542          1.0000000
## Entradas_perdidas            0.5532103  0.9179407          0.6467478
## Despejes_intentados          0.7115820  0.3740017          0.4054458
## Partidos_jugados             0.5776482  0.4885425          0.4078533
##          Entradas_perdidas  Despejes_intentados  Partidos_jugados
## Balones_recuperados          0.5532103          0.7115820          0.5776482
## Entradas                     0.9179407          0.3740017          0.4885425
## Entradas_ganadas             0.6467478          0.4054458          0.4078533
## Entradas_perdidas            1.0000000          0.2804817          0.4752904
## Despejes_intentados          0.2804817          1.0000000          0.3528119
```



```
## Partidos_jugados          0.4752904          0.3528119          1.0000000
```

```
# Visualización de la matriz de correlación
```

```
corrplot(correlation_matrix_defensa, method = "circle")
```

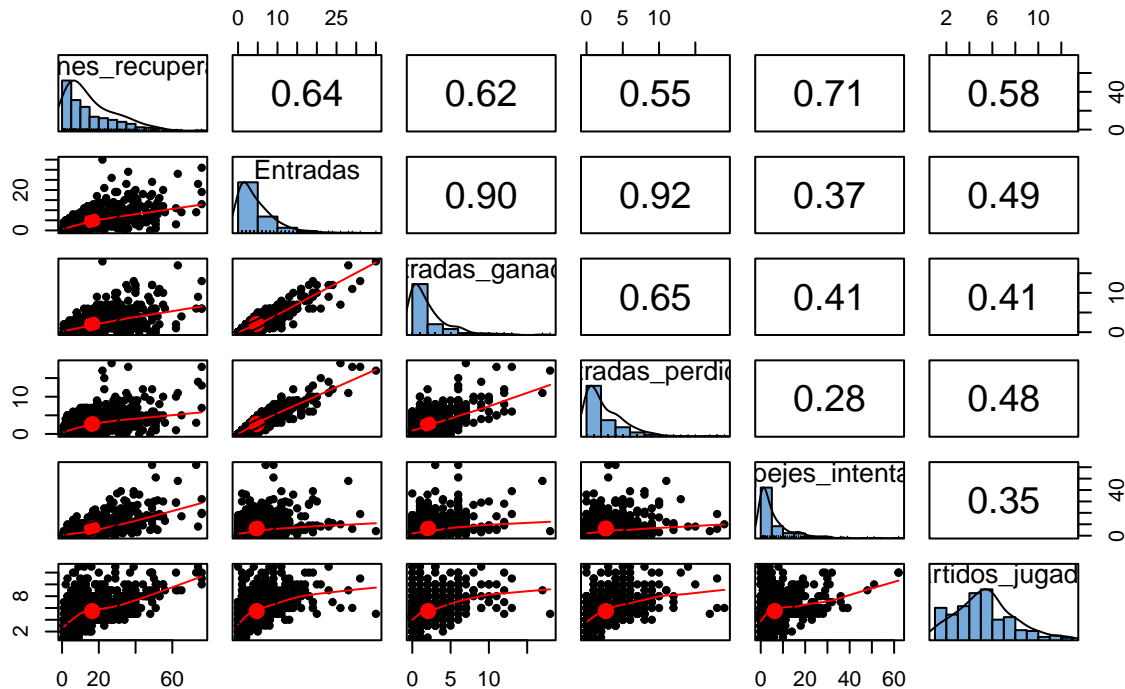


```
subset_data_defensa <- defending[, c("Balones_recuperados", "Entradas", "Entradas_ganadas", "Entradas_perdidas", "Despejes_intentados", "Partidos_jugados")]
```

```
# Gráfico de dispersión
```

```
pairs.panels(subset_data_defensa,
  main = "Gráficos de Dispersión para Variables de Defensa",
  cex.main = 1.2,
  cex.labels = 1.2,
  method = "pearson",
  hist.col = "#75AADB",
  density = TRUE,
  ellipses = TRUE)
```

Gráficos de Dispersión para Variables de Defensa



Con las observaciones estadísticas realizadas nos gustaría armar un modelo de regresión lineal simple que permita determinar que tan probable es ganar un entrada cuando se intenta una, tanto los coeficientes de correlación como la distribución gráfica nos indican que es posible.

Ajustar el modelo de regresión lineal simple

```
modelo_defensa <- lm(Entradas_ganadas ~ Entradas, data = defending)
```

Resumen del modelo

```
summary(modelo_defensa)
```

```
##
```

```
## Call:
```

```
## lm(formula = Entradas_ganadas ~ Entradas, data = defending)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -4.9833 -0.6597  0.1374  0.6712  4.2799
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.137425   0.063589  -2.161   0.0311 *
```

```
## Entradas     0.466183   0.009228  50.517  <2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

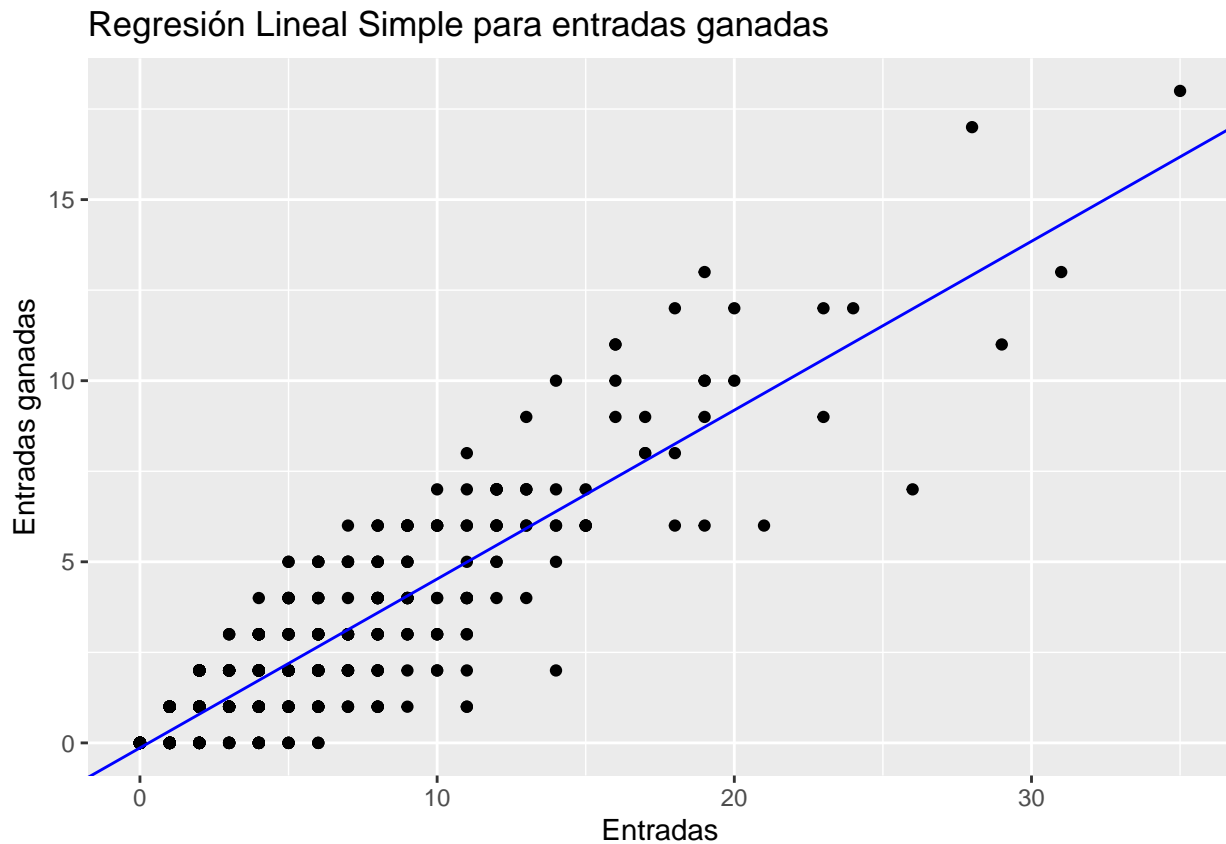
```
## Residual standard error: 1.157 on 625 degrees of freedom
```

```
## Multiple R-squared:  0.8033, Adjusted R-squared:  0.803
```

```
## F-statistic: 2552 on 1 and 625 DF, p-value: < 2.2e-16
```

Graficar los valores observados vs. predichos para el modelo de regresión lineal múltiple

```
ggplot(defending, aes(x = Entradas, y = Entradas_ganadas)) +
  geom_point() + # Puntos de datos
  geom_abline(intercept = coef(modelo_defensa)[1], slope = coef(modelo_defensa)[2], color = "blue") +
  labs(title = "Regresión Lineal Simple para entradas ganadas",
        x = "Entradas",
        y = "Entradas ganadas")
```



Resultados de nuestro modelo de partidos sin goles

Nuestro análisis visual y estadístico de las métricas F, p valor y R2 indican que nuestro modelo es adecuado para predecir la proporción de entradas ganadas que es aproximadamente (0.45), revisando la pendiente de la recta de nuestro modelo. Es decir hay una probabilidad del 45% de recuperar un valor frente a una entrada intentada.

Análisis de medio campo

```
# top de distribucion basados en nueva metrica
# pases completados e intentados
top_midfielders <- midfielders %>%
  mutate(Puntuacion_compuesta = Pases_intentados * Pases_completados) %>%
  arrange(desc(Puntuacion_compuesta)) %>%
  head(10)

# establecer top para leyenda de burbuja pases completados
tres_tamanos_maximos <- sort(unique(top_midfielders$Pases_completados), decreasing = TRUE)[1:1]

ggplot(top_midfielders, aes(x=reorder(Nombre_jugador, -Puntuacion_compuesta), y=Pases_intentados, size=
```

```

geom_point(alpha=0.7) +
scale_size_continuous(range = c(3, 30), breaks=tres_tamamos_maximos) +
labs(
  title="Mediocampo por Pases Intentados y Completados (Top 10)",
  x="Jugador",
  y="Pases Intentados",
  size="Pases Completados",
  color="Equipo"
) +
theme(axis.text.x = element_text(angle = 45, hjust = 1),
      legend.position = "right",
      legend.spacing.y = unit(0.5, "cm")) +
guides(size=guide_legend(order=1), color=guide_legend(order=2))

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

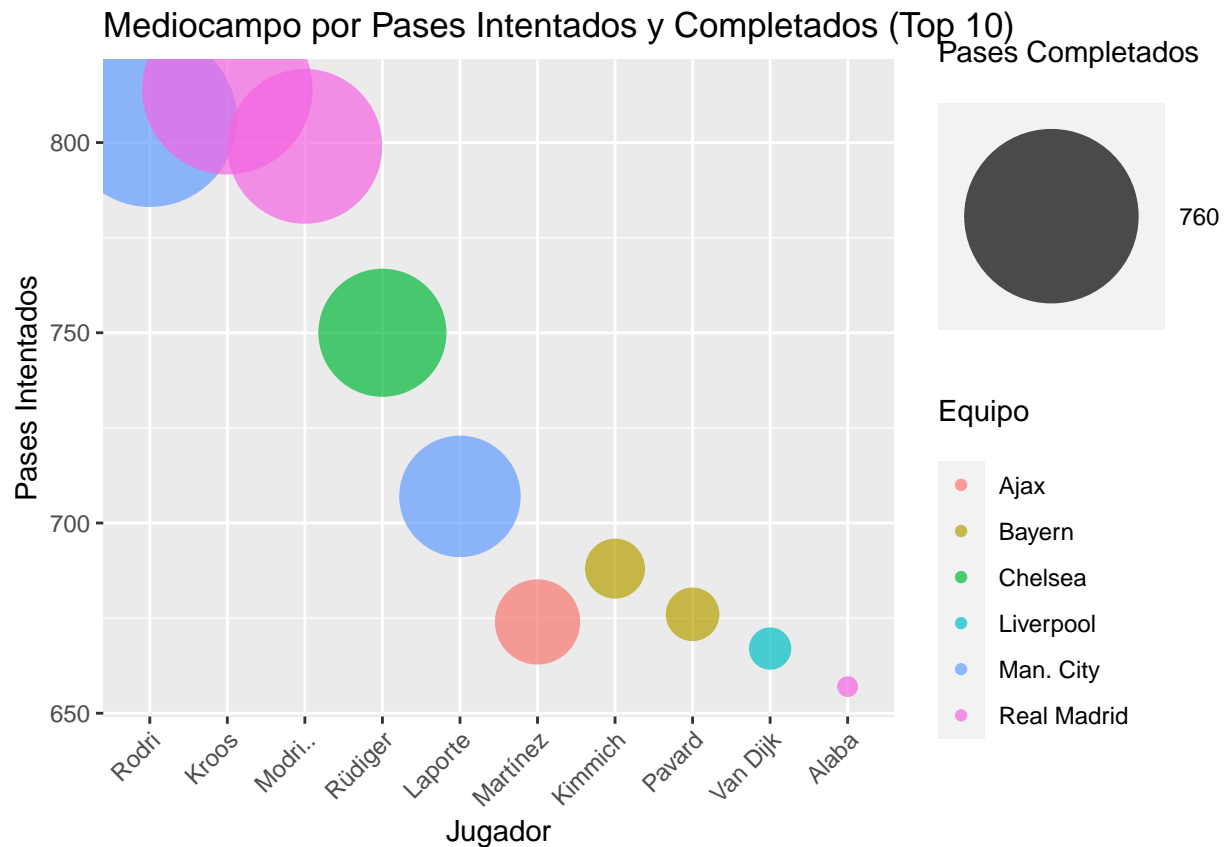
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Modrić' in 'mbcsToSbcs': dot substituted for <87>

```

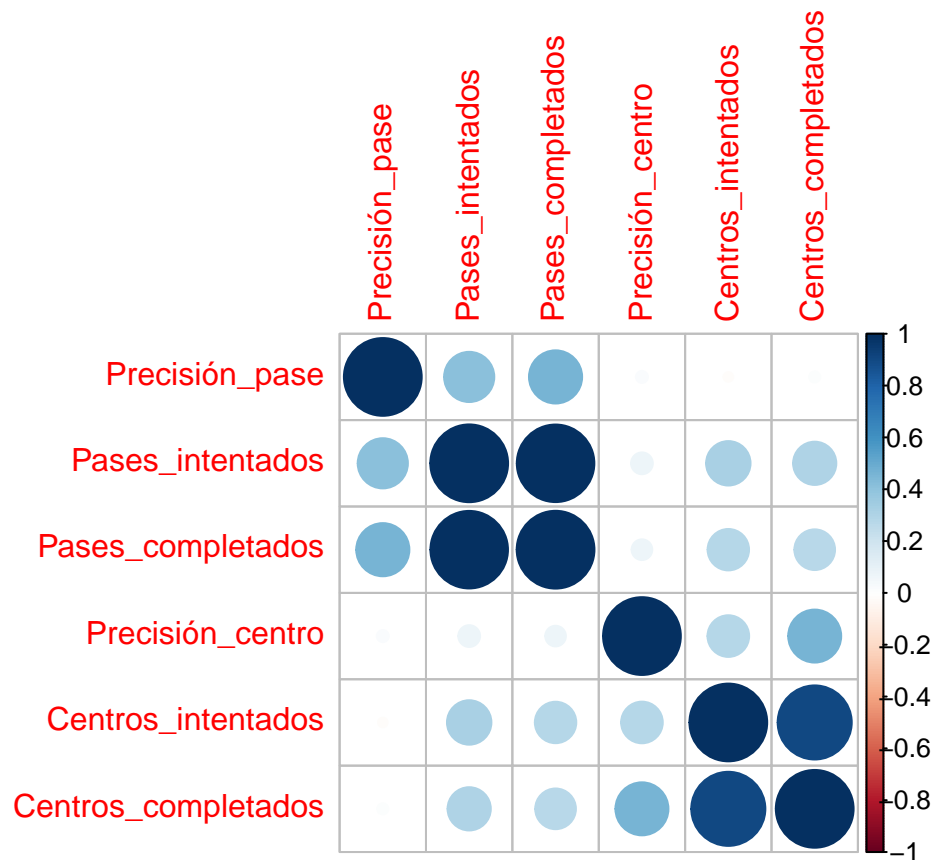


En este gráfico se puede apreciar el top de futbolistas con mayor cantidad de pases intentados y su relación con pases completados con éxito.

Creación de modelo para medio campo

```
# Matriz de correlación
correlation_matrix_midfielders <- cor(midfielders[, c("Precisión_pase", "Pases_intentados", "Pases_comp

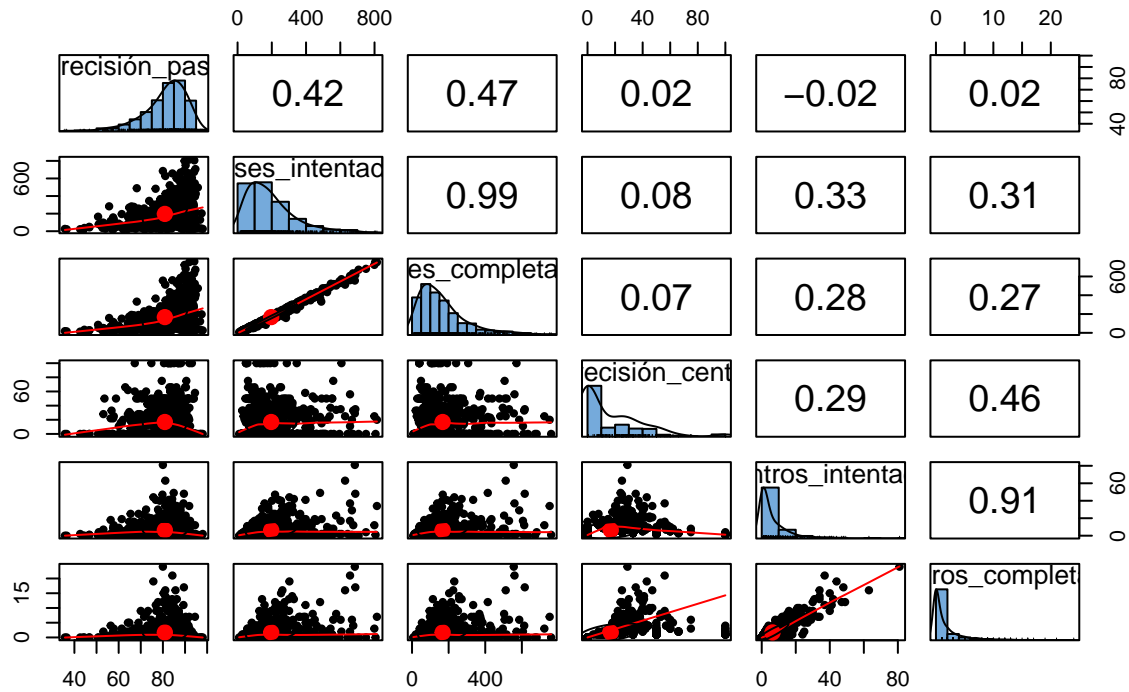
# Visualización de la matriz de correlación
corrplot(correlation_matrix_midfielders, method = "circle")
```



```
subset_data_midfielders <- midfielders[, c("Precisión_pase", "Pases_intentados", "Pases_completados", "Precisión_centro", "Centros_intentados", "Centros_completados")]

# Crea el gráfico de dispersión
pairs.panels(subset_data_midfielders,
  main = "Gráficos de Dispersión con Títulos",
  cex.main = 1.2, # Tamaño del título principal
  cex.labels = 1.2, # Tamaño de las etiquetas de variables
  method = "pearson", # Método de correlación
  hist.col = "#75AADB", # Color de los histogramas
  density = TRUE, # Mostrar densidades en los histogramas
  ellipses = TRUE) # Mostrar elipse de confianza
```

Gráficos de Dispersión con Títulos



Con el

análisis de correlación y de dispersión analizaremos la precisión en los pases

Creación de modelo para mediocampo

```
# Ajustar el modelo de regresión lineal simple
modelo_pases <- lm(Precisión_pase ~ Pases_intentados, data = midfielders)

# Obtener predicciones del modelo
predicciones <- predict(modelo_pases)

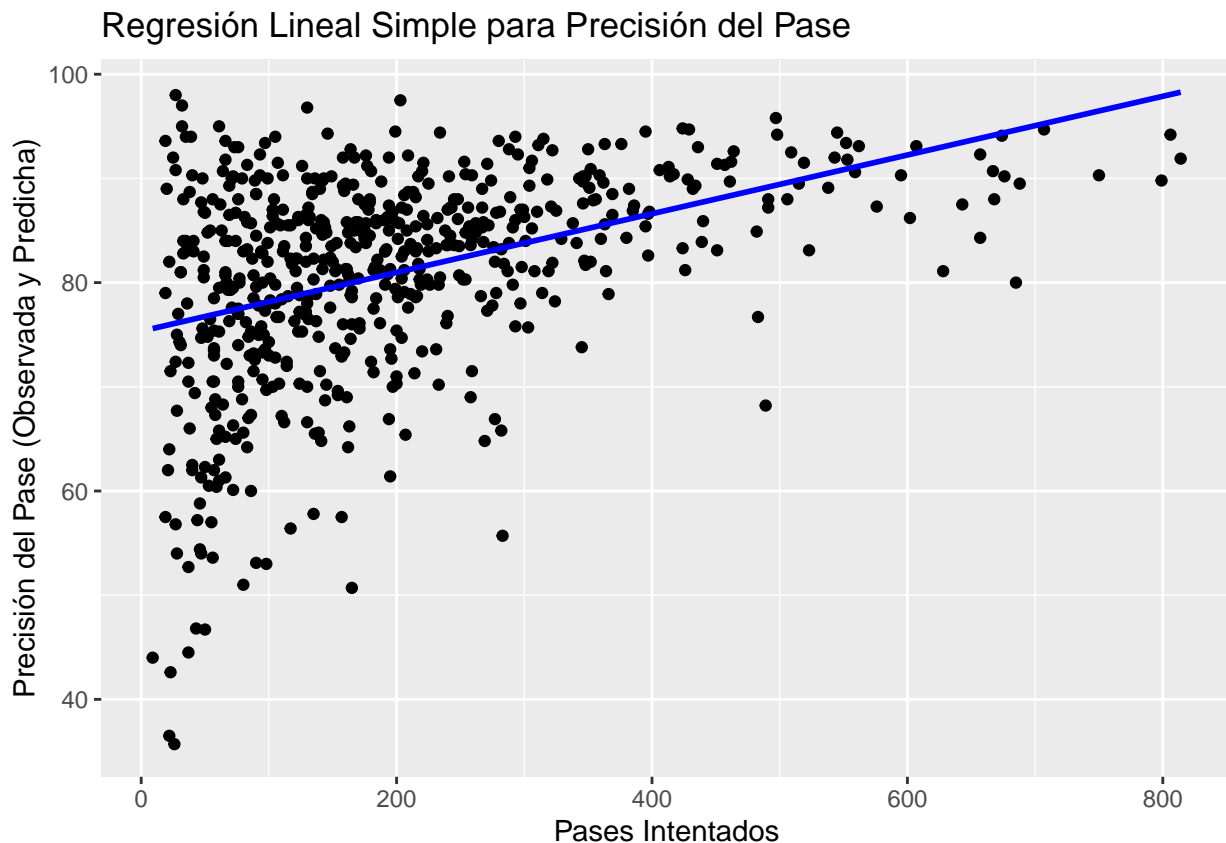
summary(modelo_pases)
```

```
##
## Call:
## lm(formula = Precisión_pase ~ Pases_intentados, data = midfielders)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40.373  -4.146   1.451   5.670  21.899
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   75.33998    0.61532  122.44  <2e-16 ***
## Pases_intentados 0.02818    0.00250   11.27  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.128 on 606 degrees of freedom
## Multiple R-squared:  0.1734, Adjusted R-squared:  0.172
## F-statistic: 127.1 on 1 and 606 DF, p-value: < 2.2e-16
```



```
ggplot(midfielders, aes(x = Pases_intentados, y = Precisión_pase)) +
  geom_point() + # Puntos de datos
  geom_line(aes(y = predicciones), color = "blue", size = 1) + # Línea de regresión
  labs(title = "Regresión Lineal Simple para Precisión del Pase",
        x = "Pases Intentados",
        y = "Precisión del Pase (Observada y Predicha)")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Como vemos de cierta manera se puede predecir la precisión de los pases, sin embargo nuestros estadísticos son malos, tenemos un R2 muy bajo, por lo que vamos a intentar con un modelo diferente.

```
# Ajustar un modelo de regresión polinómica
modelo_medio <- lm(Precisión_pase ~ poly(Pases_intentados, degree = 4), data = midfielders)

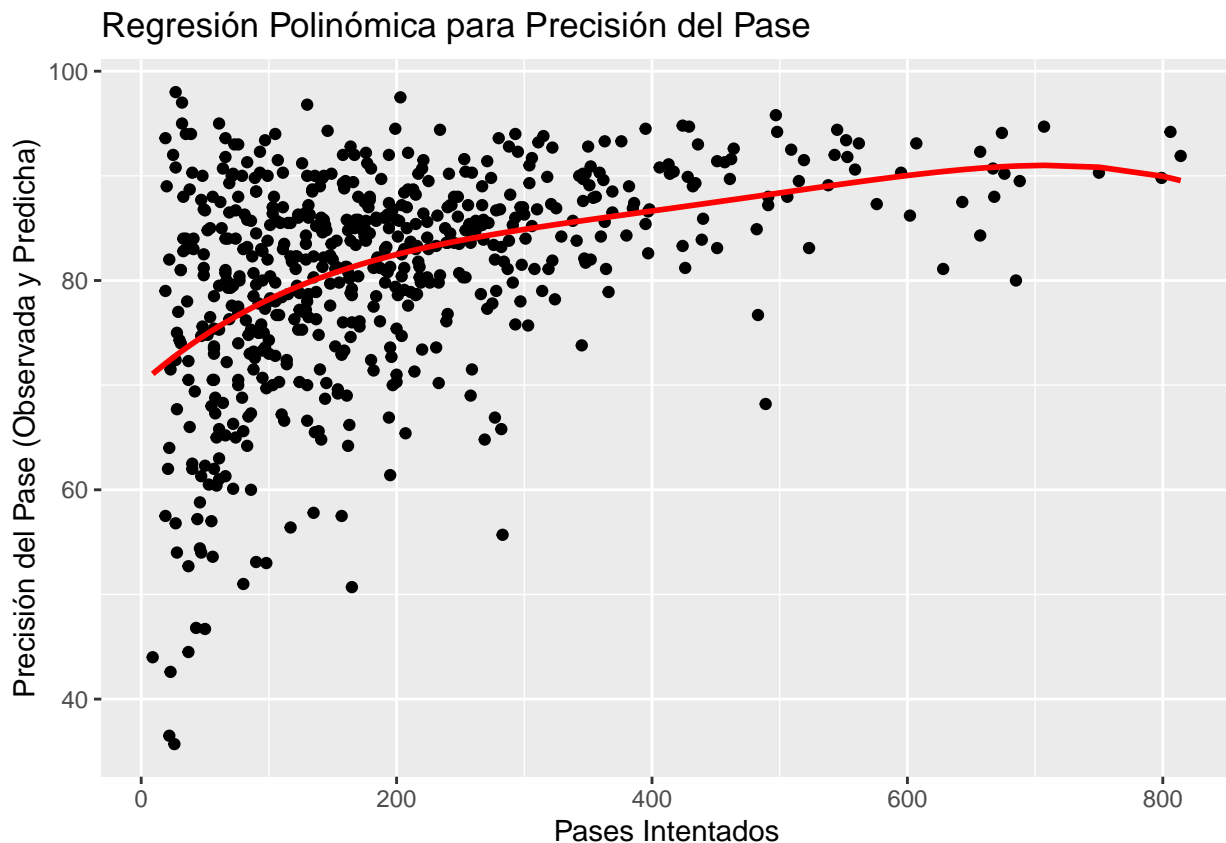
# Hacer predicciones con el modelo polinómico
predicciones_polinomicas <- predict(modelo_medio, newdata = data.frame(Pases_intentados = midfielders$Pases_intentados))

summary(modelo_medio)

##
## Call:
## lm(formula = Precisión_pase ~ poly(Pases_intentados, degree = 4),
##     data = midfielders)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.024  -4.076   0.947   5.182  25.185
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      80.8811     0.3653  221.420 < 2e-16 ***
## poly(Pases_intentados, degree = 4)1 102.9164     9.0070   11.426 < 2e-16 ***
## poly(Pases_intentados, degree = 4)2 -36.2777     9.0070   -4.028 6.35e-05 ***
## poly(Pases_intentados, degree = 4)3  13.2878     9.0070    1.475  0.141
## poly(Pases_intentados, degree = 4)4  -9.2453     9.0070   -1.026  0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.007 on 603 degrees of freedom
## Multiple R-squared:  0.1992, Adjusted R-squared:  0.1939
## F-statistic: 37.5 on 4 and 603 DF, p-value: < 2.2e-16
```

```
# Graficar los resultados
ggplot(midfielders, aes(x = Pases_intentados, y = Precisión_pase)) +
  geom_point() + # Puntos de datos
  geom_line(aes(y = predicciones_polinomicas), color = "red", size = 1) + # Línea de regresión polinómica
  labs(title = "Regresión Polinómica para Precisión del Pase",
       x = "Pases Intentados",
       y = "Precisión del Pase (Observada y Predicha)")
```



Como podemos ver no estamos logrando ajustar el modelo porque existe demasiada variabilidad en primeros

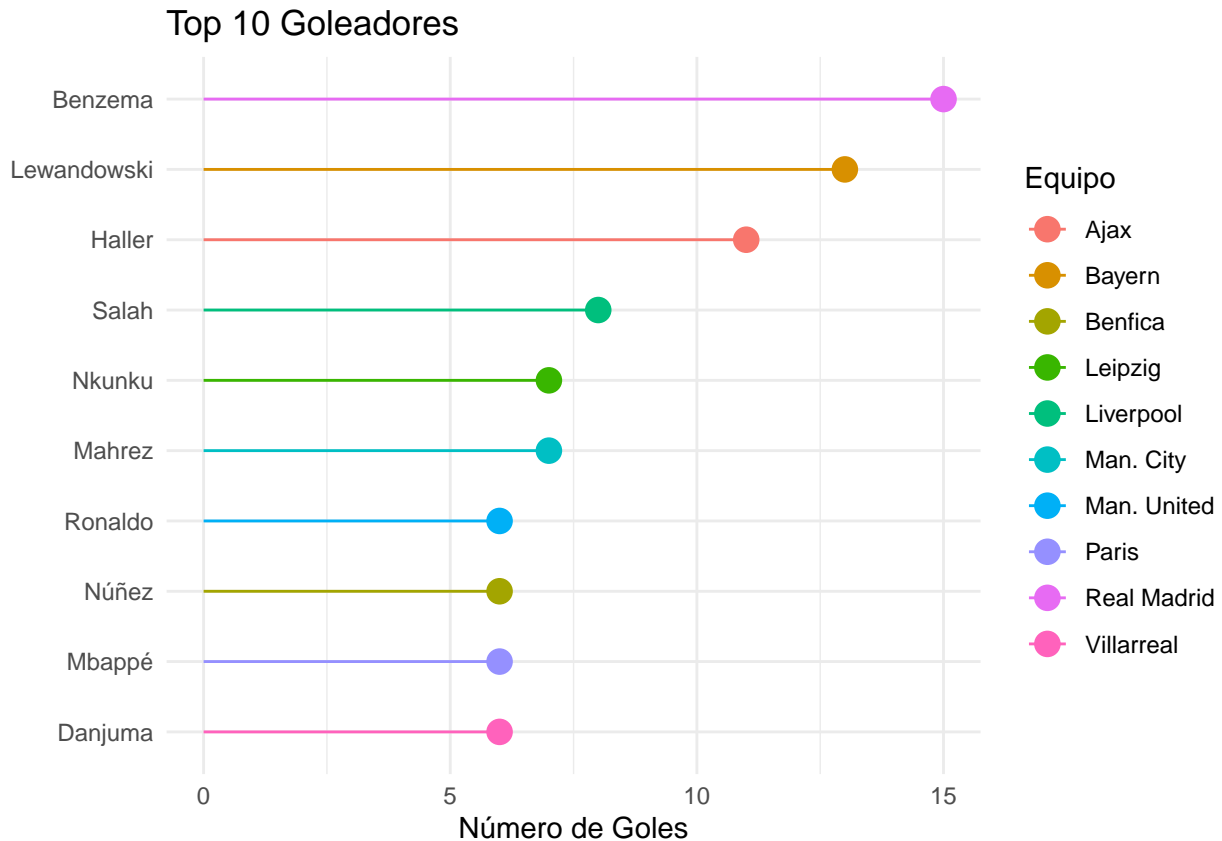
200 pases, no hemos logrado dar con un modelo que sea capaz de predecir, nos haría falta otro tipo de información, mucho depende del equipo que realice los pases y otros factores que nos hacen falta en este dataset para determinar un modelo.

Análisis de delanteros

Finalmente el aparatado defensivo es vital, por ello tenemos el top delanteros de la temporada.

```
top_goleadores <- goals %>%
  arrange(desc(goles)) %>%
  head(10)

ggplot(top_goleadores, aes(x=reorder(nombre_jugador, goles), y=goles, color=equipo)) +
  geom_segment(aes(xend=nombre_jugador, yend=0, color=equipo), size=0.5) +
  geom_point(size=4) +
  coord_flip() +
  labs(
    title="Top 10 Goleadores",
    y="Número de Goles",
    x="Jugador"
  ) +
  theme_minimal() +
  theme(axis.title.y=element_blank()) +
  scale_color_discrete(name="Equipo")
```



Como podemos observar el podio se lo llevan Benzema, Lewandowski y Haller.

Construcción del modelo

****Variable dependiente: Goles***

Realizamos el análisis de correlación y dispersión

```
correlation_matrix_goals <- cor(goals[, c("goles", "pie_derecho", "pie_izquierdo", "cabezazos", "otros", "dentro_area")])  
print(correlation_matrix_goals)
```

```
##           goles pie_derecho pie_izquierdo cabezazos  otros  
## goles      1.0000000  0.73983819   0.55053077 0.46813438 0.3534190  
## pie_derecho 0.7398382  1.00000000  -0.05685574 0.18980981 0.1902586  
## pie_izquierdo 0.5505308 -0.05685574   1.00000000 0.03427709 0.1716038  
## cabezazos   0.4681344  0.18980981   0.03427709 1.00000000 0.1644312  
## otros       0.3534190  0.19025858   0.17160379 0.16443116 1.0000000  
## dentro_area 0.9666028  0.72416318   0.49465289 0.49940796 0.3455990  
##  
##           dentro_area  
## goles      0.9666028  
## pie_derecho 0.7241632  
## pie_izquierdo 0.4946529  
## cabezazos   0.4994080  
## otros       0.3455990  
## dentro_area 1.0000000
```

Visualización de la matriz de correlación

```
corrplot(correlation_matrix_goals, method = "circle")
```

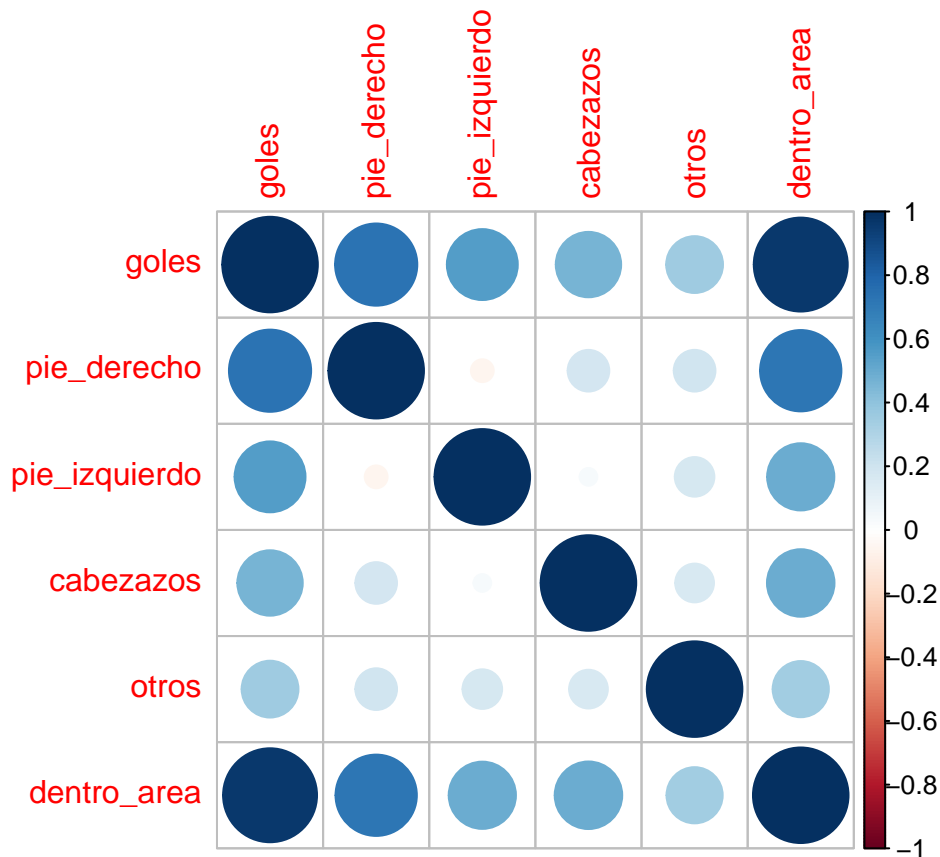
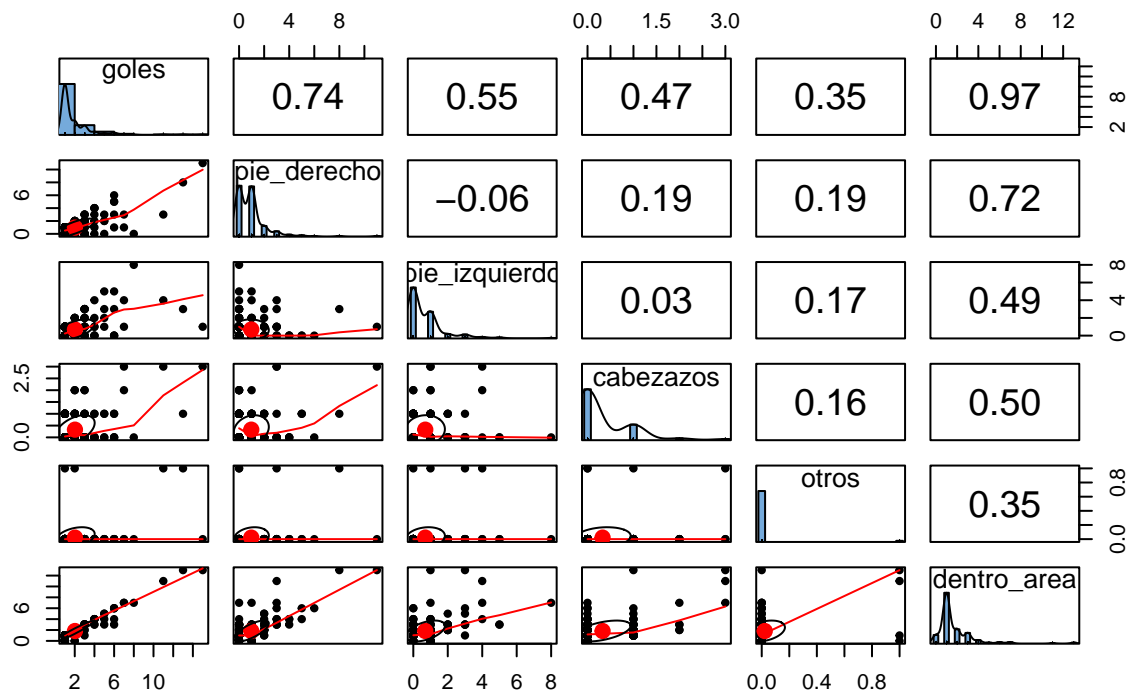


Gráfico de dispersión

```
subset_data_goals <- goals[, c("goles", "pie_derecho", "pie_izquierdo", "cabezazos", "otros", "dentro_area")]

pairs.panels(subset_data_goals,
  main = "Gráficos de Dispersión con Títulos",
  cex.main = 1.2,
  cex.labels = 1.2,
  method = "pearson",
  hist.col = "#75AADB",
  density = TRUE,
  ellipses = TRUE)
```

Gráficos de Dispersión con Títulos



Dado el análisis intentaremos construir un modelo para predecir la cantidad de goles dentro del area dado que el remata escogido sea con pie derecho.

```
modelo_delantero <- lm(goles ~ pie_derecho, data = goals)
```

```
# Resumen del modelo
```

```
summary(modelo_delantero)
```

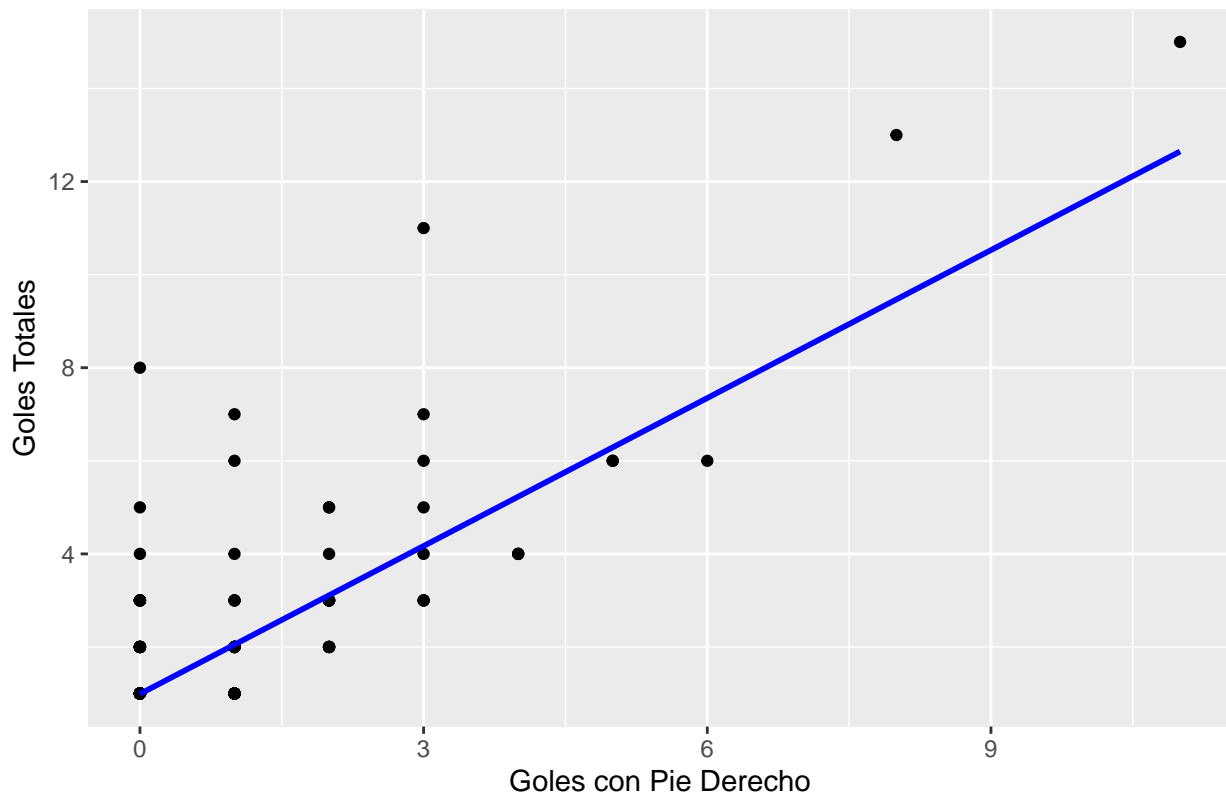
```
##
## Call:
## lm(formula = goles ~ pie_derecho, data = goals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3485 -1.0508 -0.0508  0.0087  7.0087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.99127    0.12188   8.133 6.45e-14 ***
## pie_derecho    1.05954    0.07162  14.794 < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.353 on 181 degrees of freedom
## Multiple R-squared:  0.5474, Adjusted R-squared:  0.5449
## F-statistic: 218.9 on 1 and 181 DF,  p-value: < 2.2e-16

ggplot(goals, aes(x = pie_derecho, y = goles)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Regresión Lineal Simple para Predecir Goles con Pie Derecho",
       x = "Goles con Pie Derecho",
       y = "Goles Totales")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Regresión Lineal Simple para Predecir Goles con Pie Derecho



Este modelo de regresión lineal simple nos permite explicar de una manera relativamente precisa que la cantidad de goles con pie derecho, es en una proporción de 0.75, los estadísticos indican que es un modelo que de alguna manera explica la variable goles.

Resultados finales del modelos

```
library(stargazer)
```

```
##
## Please cite as:
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
#summary(ModeloVenBen)
```

```
lista_modelos <- list(modelo_arco, modelo_defensa, modelo_medio, modelo_delantero)
```

```
stargazer(lista_modelos, title="Resumen de modelos usados", type = "text")
```

```
##
## Resumen de modelos usados
## =====
##                                     Dependent variable:
##                                     -----
##                                     Partidos_sin_goles      Entradas_ganadas      Precisión_p
##                                     (1)                    (2)                    (3)
## -----
## Paradas                                0.0004
##                                     (0.051)
##
## Goles_concedidos                     -0.144***
##                                     (0.028)
##
## Partidos_jugados                      0.566*
##                                     (0.251)
##
## Entradas                                0.466***
##                                     (0.009)
##
## poly(Pases_intentados, degree = 4)1                                102.916**
##                                     (9.007)
##
## poly(Pases_intentados, degree = 4)2                                -36.278**
##                                     (9.007)
##
## poly(Pases_intentados, degree = 4)3                                13.288
##                                     (9.007)
##
## poly(Pases_intentados, degree = 4)4                                -9.245
##                                     (9.007)
##
## pie_derecho
##
## Constant                             -0.543*
##                                     (0.246)
##                                     -0.137**
##                                     (0.064)
##                                     80.881**
##                                     (0.365)
## -----
## Observations                        10                        627                        608
## R2                                0.966                        0.803                        0.199
## Adjusted R2                        0.948                        0.803                        0.194
## Residual Std. Error                0.359 (df = 6)                1.157 (df = 625)                9.007 (df = 6)
## F Statistic                        56.021*** (df = 3; 6) 2,551.968*** (df = 1; 625) 37.503*** (df = 1; 6)
## =====
## Note:
```

Analziando nuestros modelos, sin duda nuestro mejor trabajo está en el modelo de predicción de partidos sin

goles (regresión múltiple), el modelo de entradas gaandas es también bastante bueno, los modelos de goles y medio campos no son tan adecuados.

Consideramos que esto se debe a que en los datasets de medio campo y goles, falta más información propia del fútbol para poder establecer mejores relaciones y crear modelos más completos.

En cambio los modelos de defensa tanto en arco como en entradas ganadas en defensa son exelentes para predecir sus respectivas variables independientes por lo que podemos usarlos para la toma de decisiones.

Conclusiones

1. En la temporada 2021-2022 el equipo ganador fue el Real Madrid, si realizamos un análisis detallado es un más que justo ganador, obteniendo podio en todas las fasetas del campo, se puede destacar el volumen de pases intentados/exitosos de sus medio campistas Kroos Y Modric, así como también una temporada que roza la gloria de Benzema y Courtois. Son jugadores importantes que deben subir su valor.
2. Como apartado adicional jugadores muy interesantes son Haller, debido a la cantidad de goles marcados a pesar de no haber tenido tantas participaciones. Kimmich a pesar de no haber llegado a la final tienen una gran cantidad de pases intentados y un gran volumen de acierto, además de ser un jugador que no solo da pases “defensivos”.
3. Una acción importante es ver que el número de paradas, goles concedidos y partidos jugados influyen directamente sobre partidos sin goles, destacando que con estos parámetros esta cantidad de partidos está entre 1 y 5 lo cual es bueno dado que al público le interesa partidos con la mayor cantidad de goles posibles. Esto indica que el formato actual de la competición resulta atractivo para el público, se está planteando un formato estilo liga, que consideramos inadecuado ya que se generarían más partidos sin goles, haciendo menos atractiva la competición y por ende generando menos ingresos.
4. Es importante mantener un buen margen de publicidad y centrar el atractivo de la competición en equipos que tengan un estilo ofensivo pues como muestra nuestro modelo de predicción de entradas exitosas, tan solo hay una probabilidad de 45% de tener una entrada exitosa, es decir que mientras un equipo se mantenga atacando es bastante probable que logre superar marcas y generar ocasiones. Esto es especialmente importante para mantener el atractivo de la competición, se debe premiar más a los equipos que ataquen, para que así los equipos que tienden a cerrarse mucho intenten jugar más equilibrados.

