

NBA 2015

Maribel Crespí valero y Carlos Bosh Pino

4/25/2020

Contents

1	Introducción:	2
1.1	Carga de datos	3
2	Estadísticos Descriptivos básicos globales: Medias, Medianas, Desviaciones típicas y Varianzas	5
2.1	Medias y medianas:	5
2.2	Desviaciones típicas:	8
2.3	Varianzas:	10
3	Estadísticas descriptivas básicas por agrupaciones:	13
3.1	Representaciones gráficas:	13
3.2	Altura relacionada con otros datos:	16
4	Aplicaciones de estadística de inferencia a los datos	26
4.1	Bondad de ajuste con χ^2 de Pearson	26
4.2	Anova 1	27
4.3	Test KS	28
4.4	Boxplot de Puntos por equipo	31
4.5	Segunda ANOVA	32
4.6	Comparación de medias dos a dos con ajuste de Bonferroni	32
4.7	Clustering	33
5	Conclusiones:	39



1 Introducción:

En baloncesto abundan estadísticas de todo tipo, así que hemos pensado que era un tema bastante idóneo a tratar en este trabajo. La elección de la NBA es debido a que es la liga más atractiva a la hora de sacar datos, ya que en ésta es donde juegan los mejores jugadores de todo el mundo y por lo tanto es un foco de talento inmenso. Un ejemplo de la inmensa cantidad de estadísticas son actuaciones históricas donde jugadores han llegado a marcar 100 puntos, dar 30 asistencias o coger 42 rebotes. De jugadores que promedian triples-dobles por temporada o incluso aquellos que han conseguido llegar al cuádruple-doble en algún partido.

NOTA: Un triple-doble es conseguir doble dígito en al menos tres de los datos estadísticos cuantificables tradicionales. Estos son los puntos, los rebotes, las asistencias, los robos y los tapones. Un cuádruple-doble es lo mismo pero en cuatro datos estadísticos.

Nosotros nos vamos a encargar de estudiar todas estas estadísticas, ver que cómo se relacionan unas con otras y sacar las conclusiones pertinentes. Investigaremos qué es lo que hace a un equipo el dueño de su conferencia, comparar a los mejores jugadores para ver quién es realmente eficiente para su equipo, ver la evolución de la liga y de los equipos, entre otras muchas cosas.

Hemos escogido 449 jugadores con un total de 34 variables (“Nombre”, “Partidos jugados”, “Minutos”, “Puntos”, “Canastas_ anotadas”, “Canastas intentadas”, “Porcentaje canastas”, “3P anotados”, “3P intentados”, “Porcentaje triples”, “Tiros libres anotados”, “Tiros libres intentados”, “Porcentaje tiros libres”, “Rebotes ofensivos”, “Rebotes defensivos”, “Rebotes”, “Asistencias”, “Robos”, “Tapones”, “Perdidas”, “Faltas en ataque”, “Faltas en defensa”, “Asistencias/Perdida”, “Robo/Pérdida”, “Edad”, “Nacimiento”, “Fecha de nacimiento”, “Instituto”, “Experiencia”, “Peso”, “Posición”, “Equipo”, “Altura”, “BMI”).

Los objetivos del proyecto y los apartados que se tratan son:

1. Estadísticos Descriptivos básicos globales.
2. Estadísticos Descriptivos básicos por agrupaciones.
3. Análisis multidimensional: Relación entre las variables.

4. Presentación gráfica de los resultados Anteriores.
5. Conclusiones: Interpretaciones básicas del análisis exploratorio.

EL link con los datos es: NBA 2015

1.1 Carga de datos

```
data=read.csv("players_stats.csv.xls",header=TRUE)

colnames(data)<-c("Nombre","Partidos jugados","Minutos","Puntos",
  "Canastas anotadas","Canastas intentadas",
  "Porcentaje canastas","3P anotados",
  "3P intentados","Porcentaje triples",
  "Tiros libres anotados","Tiros libres intentados",
  "Porcentaje tiros libres ","Rebotes ofensivos",
  "Rebotes defensivos","Rebotes",
  "Asistencias","Robos","Tapones","Perdidas","Faltas en ataque",
  "Faltas en defensa","Asistencias/Perdida","Robo/Pérdida",
  "Edad","Nacimiento","Fecha de nacimiento","Instituto",
  "Experiencia","Peso","Posición","Equipo","Altura","BMI")

str(data)
```

```
## 'data.frame':   490 obs. of  34 variables:
##  $ Nombre           : Factor w/ 490 levels "Aaron Brooks",...: 4 1 2 3 5 6 8 9 10 11 ...
##  $ Partidos jugados  : int   26 82 47 32 76 65 74 27 5 69 ...
##  $ Minutos           : int  324 1885 797 740 2318 1992 1744 899 14 1518 ...
##  $ Puntos            : int  133 954 243 213 1156 1082 545 374 4 432 ...
##  $ Canastas anotadas : int   51 344 93 91 519 486 195 121 1 179 ...
##  $ Canastas intentadas : int  137 817 208 220 965 1010 440 300 4 353 ...
##  $ Porcentaje canastas : num  37.2 42.1 44.7 41.4 53.8 48.1 44.3 40.3 25 50.7 ...
##  $ 3P anotados       : int   15 121 13 1 11 2 73 26 0 1 ...
##  $ 3P intentados      : int   57 313 48 9 36 5 210 68 0 3 ...
##  $ Porcentaje triples : num  26.3 38.7 27.1 11.1 30.6 40 34.8 38.2 0 33.3 ...
##  $ Tiros libres anotados : int   16 145 44 30 107 108 82 106 2 73 ...
##  $ Tiros libres intentados : int   24 174 61 46 141 165 101 129 2 104 ...
##  $ Porcentaje tiros libres : num  66.7 83.3 72.1 65.2 75.9 65.5 81.2 82.2 100 70.2 ...
##  $ Rebotes ofensivos   : int    6 32 46 48 131 99 31 19 1 142 ...
##  $ Rebotes defensivos  : int   26 134 123 114 413 449 173 95 0 312 ...
##  $ Rebotes            : int   32 166 169 162 544 548 204 114 1 454 ...
##  $ Asistencias         : int   46 261 33 30 244 113 83 82 1 32 ...
##  $ Robos              : int    7 54 21 19 68 47 56 17 0 34 ...
##  $ Tapones            : int    0 15 22 9 98 84 5 5 0 105 ...
##  $ Perdidas           : int   14 157 38 44 100 68 60 52 0 74 ...
##  $ Faltas en ataque    : int   15 189 83 88 121 139 148 64 1 213 ...
##  $ Faltas en defensa   : int  110 791 318 244 1530 1225 569 338 3 778 ...
##  $ Asistencias/Perdida : num   3.29 1.66 0.87 0.68 2.44 1.66 1.38 1.58 0 0.43 ...
##  $ Robo/Pérdida       : num   0.5 0.34 0.55 0.43 0.68 0.69 0.93 0.33 0 0.46 ...
##  $ Edad              : int   29 30 20 24 29 30 33 24 24 22 ...
##  $ Nacimiento         : Factor w/ 42 levels "", "ar", "au", "ba",...: 40 40 40 40 13 40 40 40 40 39
##  $ Fecha de nacimiento : Factor w/ 409 levels "", "April 1, 1988",...: 370 138 380 111 226 154 347
##  $ Instituto          : Factor w/ 113 levels "", "Arizona State University",...: 66 90 60 33 69 1
```

```
## $ Experiencia      : Factor w/ 21 levels "", "1", "10", "11", ...: 16 17 21 21 18 3 16 14 21 2 ..
## $ Peso             : num  185 180 202 205 205 ...
## $ Posición         : Factor w/ 6 levels "", "C", "PF", "PG", ...: 4 4 3 3 2 2 6 6 2 2 ...
## $ Equipo           : Factor w/ 31 levels "", "ATL", "BOS", ...: 25 5 23 2 2 4 19 30 6 25 ...
## $ Altura           : num  81.5 72.5 99 106.7 110.2 ...
## $ BMI              : num  23.8 22.4 24.1 25.4 26.2 ...
```

```
head(data)
```

```
##      Nombre Partidos jugados Minutos Puntos Canastas anotadas
## 1      AJ Price           26      324      133              51
## 2    Aaron Brooks          82     1885      954             344
## 3    Aaron Gordon          47      797      243             93
## 4 Adreian Payne           32      740      213             91
## 5      Al Horford          76     2318     1156            519
## 6    Al Jefferson          65     1992     1082            486
##      Canastas intentadas Porcentaje canastas 3P anotados 3P intentados
## 1              137              37.2              15              57
## 2              817              42.1              121             313
## 3              208              44.7               13             48
## 4              220              41.4               1              9
## 5              965              53.8               11             36
## 6             1010              48.1               2              5
##      Porcentaje triples Tiros libres anotados Tiros libres intentados
## 1              26.3              16              24
## 2              38.7             145             174
## 3              27.1              44              61
## 4              11.1              30              46
## 5              30.6             107             141
## 6              40.0             108             165
##      Porcentaje tiros libres Rebotes ofensivos Rebotes defensivos Rebotes
## 1              66.7              6              26              32
## 2              83.3              32             134             166
## 3              72.1              46             123             169
## 4              65.2              48             114             162
## 5              75.9             131             413             544
## 6              65.5              99             449             548
##      Asistencias Robos Tapones Perdidas Faltas en ataque Faltas en defensa
## 1              46           7           0           14              15             110
## 2             261          54          15          157             189             791
## 3              33          21          22           38              83             318
## 4              30          19           9           44              88             244
## 5             244          68          98          100             121            1530
## 6             113          47          84           68             139            1225
##      Asistencias/Perdida Robo/Pérdida Edad Nacimiento Fecha de nacimiento
## 1              3.29              0.50   29          us   October 7, 1986
## 2              1.66              0.34   30          us   January 14, 1985
## 3              0.87              0.55   20          us   September 16, 1995
## 4              0.68              0.43   24          us   February 19, 1991
## 5              2.44              0.68   29          do    June 3, 1986
## 6              1.66              0.69   30          us   January 4, 1985
##      Instituto Experiencia  Peso Posición Equipo Altura      BMI
## 1 University of Connecticut      5 185.0      PG   PHO  81.45 23.79839
## 2      University of Oregon      6 180.0      PG   CHI  72.45 22.36111
```

## 3	University of Arizona	R 202.5	PF	ORL 99.00 24.14266
## 4	Michigan State University	R 205.0	PF	ATL 106.65 25.37775
## 5	University of Florida	7 205.0	C	ATL 110.25 26.23438
## 6		10 205.0	C	CHA 130.05 30.94587

2 Estadísticos Descriptivos básicos globales: Medias, Medianas, Desviaciones típicas y Varianzas

En este apartado, calculamos las estadísticas globales de todos los jugadores.

2.1 Medias y medianas:

Media, Mediana y maximos y minimos de minutos jugados:

```
summary(data$Minutos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.0   492.2  1193.0  1214.7  1905.8  2981.0
```

Media, Mediana y maximos y minimos de puntos:

```
summary(data$Puntos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   145.2   423.0   502.1   774.0  2217.0
```

Media, Mediana y maximos y minimos de canastas anotadas:

```
summary(data$`Canastas anotadas`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0    55.5   156.0   188.3   286.0   659.0
```

Media, Mediana y maximos y minimos de canastas intentadas:

```
summary(data$`Canastas intentadas`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   139.0   357.5   419.5   642.8  1471.0
```

Media, Mediana y maximos y minimos de los porcentajes de anotación:

```
summary(data$`Porcentaje canastas`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   39.60   42.90   43.10   47.58  100.00
```

Media, Mediana y maximos y minimos de triples anotados:

```
summary(data$`3P anotados`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   1.00   18.00   39.39   66.00   286.00
```

Media, Mediana y maximos y minimos de triples intentados:

```
summary(data$`3P intentados`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0.0     6.0   58.0   112.5   192.0   646.0
```

Media, Mediana y maximos y minimos de porcentajes de triple:

```
summary(data$`Porcentaje triples`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   15.50   31.30   25.52   36.40   100.00
```

Media, Mediana y maximos y minimos de tiros libres anotados:

```
summary(data$`Tiros libres anotados`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   18.50   58.00   86.04  126.75   715.00
```

Media, Mediana y maximos y minimos de tiros libres intentados:

```
summary(data$`Tiros libres intentados`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   26.25   80.00  114.69  166.75   824.00
```

Media, Mediana y maximos y minimos del porcentaje de acierto de tiros libres:

```
summary(data$`Porcentaje tiros libres`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   64.35   75.00   70.45   82.00   100.00
```

Media, Mediana y maximos y minimos de rebotes ofensivos:

```
summary(data$`Rebotes ofensivos`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   13.00   31.50   54.66   75.75   437.00
```

Media, Mediana y maximos y minimos de rebotes defensivos:

```
summary(data$`Rebotes ofensivos`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   13.00   31.50   54.66   75.75   437.00
```

Media, Mediana y maximos y minimos de rebotes en general:

```
summary(data$Rebotes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   70.25  176.00  217.35  314.75  1226.00
```

Media, Mediana y maximos y minimos de asistencias:

```
summary(data$Asistencias)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0    24.0    70.0   110.6   150.8   838.0
```

Media, Mediana y maximos y minimos de robos de balón:

```
summary(data$Robos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   12.25   32.00   38.84   55.50   163.00
```

Media, Mediana y maximos y minimos de tapones:

```
summary(data$Tapones)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00    5.00   14.00   24.08   29.75   200.00
```

Media, Mediana y maximos y minimos de pérdidas de balón:

```
summary(data$`Perdidas`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   23.00   56.50   68.83  100.00   321.00
```

Media, Mediana y maximos y minimos de faltas en ataque:

```
summary(data$`Faltas en ataque`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0    45.5   103.0   101.5   148.8   285.0
```

Media, Mediana y maximos y minimos de faltas en defensa:

```
summary(data$`Faltas en defensa`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      -3.0   165.0   490.5   564.3   837.0  2202.0
```

Media, Mediana y maximos y minimos de la edad de los jugadores:

```
summary(data$Edad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      20.00   24.00   27.00   27.51   30.00   39.00     68
```

2.2 Desviaciones típicas:

Desviación típica de minutos jugados:

```
sd(data$Minutos)
```

```
## [1] 820.5701
```

Desviación típica de puntos:

```
sd(data$Puntos)
```

```
## [1] 422.0842
```

Desviación típica de canastas anotadas:

```
sd(data$`Canastas anotadas`)
```

```
## [1] 156.2658
```

Desviación típica de canastas intentadas:

```
sd(data$`Canastas intentadas`)
```

```
## [1] 337.3671
```

Desviación típica de los porcentajes de anotación:

```
sd(data$`Porcentaje canastas`)
```

```
## [1] 9.625231
```

Desviación típica de triples anotados:


```
sd(data$`3P anotados`)
```

```
## [1] 47.88091
```

Desviación típica de triples intentados:

```
sd(data$`3P intentados`)
```

```
## [1] 127.3857
```

Desviación típica de porcentajes de triple:

```
sd(data$`Porcentaje triples`)
```

```
## [1] 15.7966
```

Desviación típica de tiros libres anotados:

```
sd(data$`Tiros libres anotados`)
```

```
## [1] 91.31532
```

Desviación típica de tiros libres intentados:

```
sd(data$`Tiros libres intentados`)
```

```
## [1] 115.1392
```

Desviación típica del porcentaje de acierto de tiros libres:

```
sd(data$`Porcentaje tiros libres`)
```

```
## [1] 19.08782
```

Desviación típica de rebotes ofensivos:

```
sd(data$`Rebotes ofensivos`)
```

```
## [1] 61.06604
```

Desviación típica de rebotes defensivos:

```
sd(data$`Rebotes defensivos`)
```

```
## [1] 61.06604
```

Desviación típica de rebotes en general:

```
sd(data$Rebotes)
```

```
## [1] 191.9839
```

Desviación típica de asistencias:

```
sd(data$Asistencias)
```

```
## [1] 125.9821
```

Desviación típica de robos de balón:

```
sd(data$Robos)
```

```
## [1] 33.38951
```

Desviación típica de tapones:

```
sd(data$Tapones)
```

```
## [1] 31.66285
```

Desviación típica de pérdidas de balón:

```
sd(data$`Perdidas`)
```

```
## [1] 58.4807
```

Desviación típica de faltas en ataque:

```
sd(data$`Faltas en ataque`)
```

```
## [1] 65.32681
```

Desviación típica de faltas en defensa:

```
sd(data$`Faltas en defensa`)
```

```
## [1] 464.428
```

2.3 Varianzas:

Varianza de minutos jugados:

```
var(data$Minutos)
```

```
## [1] 673335.3
```

Varianza de puntos:

```
var(data$Puntos)
```

```
## [1] 178155.1
```

Varianza de canastas anotadas:

```
var(data$`Canastas anotadas`)
```

```
## [1] 24418.99
```

Varianza de canastas intentadas:

```
var(data$`Canastas intentadas`)
```

```
## [1] 113816.6
```

Varianza de los porcentajes de anotación:

```
var(data$`Porcentaje canastas`)
```

```
## [1] 92.64507
```

Varianza de triples anotados:

```
var(data$`3P anotados`)
```

```
## [1] 2292.581
```

Varianza de triples intentados:

```
var(data$`3P intentados`)
```

```
## [1] 16227.13
```

Varianza de porcentajes de triple:

```
var(data$`Porcentaje triples`)
```

```
## [1] 249.5327
```

Varianza de tiros libres anotados:

```
var(data$`Tiros libres anotados`)
```

```
## [1] 8338.487
```

Varianza de tiros libres intentados:

```
var(data$`Tiros libres intentados`)
```

```
## [1] 13257.04
```

Varianza del porcentaje de acierto de tiros libres:

```
var(data$`Porcentaje tiros libres`)
```

```
## [1] 364.3448
```

Varianza de rebotes ofensivos:

```
var(data$`Rebotes ofensivos`)
```

```
## [1] 3729.061
```

Varianza de rebotes defensivos:

```
var(data$`Rebotes defensivos`)
```

```
## [1] 3729.061
```

Varianza de rebotes en general:

```
var(data$Rebotes)
```

```
## [1] 36857.82
```

Varianza de asistencias:

```
var(data$Asistencias)
```

```
## [1] 15871.48
```

Varianza de robos de balón:

```
var(data$Robos)
```

```
## [1] 1114.859
```

Varianza de tapones:

```
var(data$Tapones)
```

```
## [1] 1002.536
```

Varianza de pérdidas de balón:

```
var(data$`Perdidas`)
```

```
## [1] 3419.992
```

Varianza de faltas en ataque:

```
var(data$`Faltas en ataque`)
```

```
## [1] 4267.592
```

Varianza de faltas en defensa:

```
var(data$`Faltas en defensa`)
```

```
## [1] 215693.4
```

3 Estadísticas descriptivas básicas por agrupaciones:

Hemos decidido escoger los cuatro mejores equipos de la liga. Los dos mejores de la conferencia Oeste(Golden State Warriors y Houston Rockets) y los dos mejores de la conferencia Este(Atlanta Hawks y Cleveland Cavaliers).

```
Mejores_equipos=c("GSW","CLE","HOU","ATL")
data2=data[as.character(data$Equipo)%in% Mejores_equipos,c("Equipo","Puntos",
                                                             "Asistencias","Rebotes")]
data2=na.omit(data2)
summary(data2$Puntos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         4      182     390     584     855    2217
```

```
sd(data2$`Asistencias`)
```

```
## [1] 158.2523
```

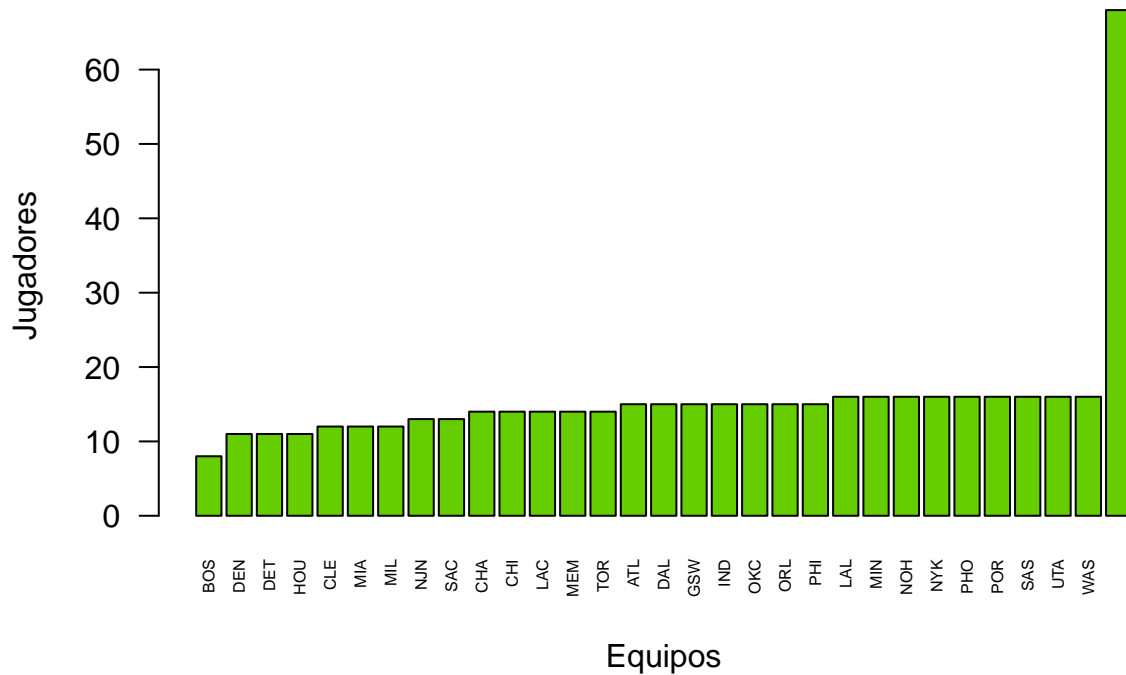
```
var(data2$Rebotes)
```

```
## [1] 35635.78
```

3.1 Representaciones gráficas:

```
barplot(sort(table(data$Equipo)),las=2,cex.names = 0.5,col="chartreuse3",
        main="Número de jugadores de cada equipo",xlab="Equipos",ylab="Jugadores")
```

Número de jugadores de cada equipo

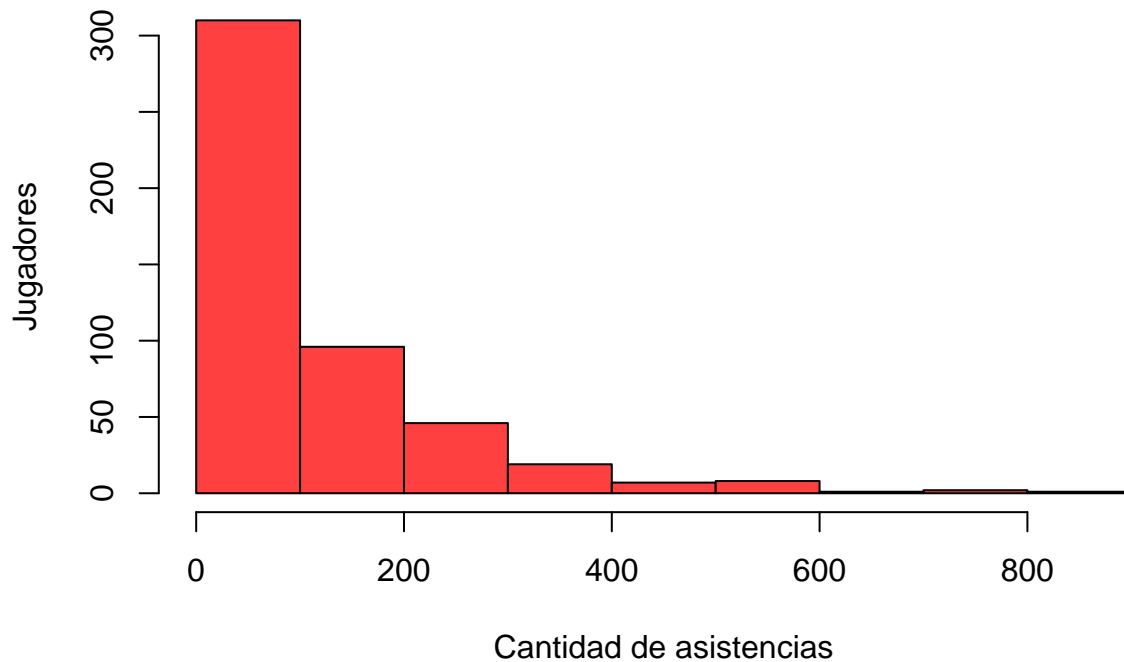


En el gráfico podemos ver como en general el número de jugadores por equipo está bastante balanceado. Algunos equipos optan por una plantilla más corta pero con jugadores estrella con salarios más elevados. Otros equipos optan por más jugadores de nivel medio pero de salarios más reducidos. En general los equipos con las super estrellas mejor pagadas, son con los que menor plantel cuentan. Un ejemplo podría ser Cleveland Cavaliers, cuyo plantel cuenta con LeBron James y Kyrie Irving, dos jugadores de contrato máximo. Por lo tanto dejan poco espacio salarial para los otros jugadores.

NOTA: Todos los equipos de la NBA tienen un límite salarial.

```
hist(data$`Asistencias`,main = "Frecuencia absoluta de asistencias",
      xlab = "Cantidad de asistencias", ylab = "Jugadores",col="brown1")
```

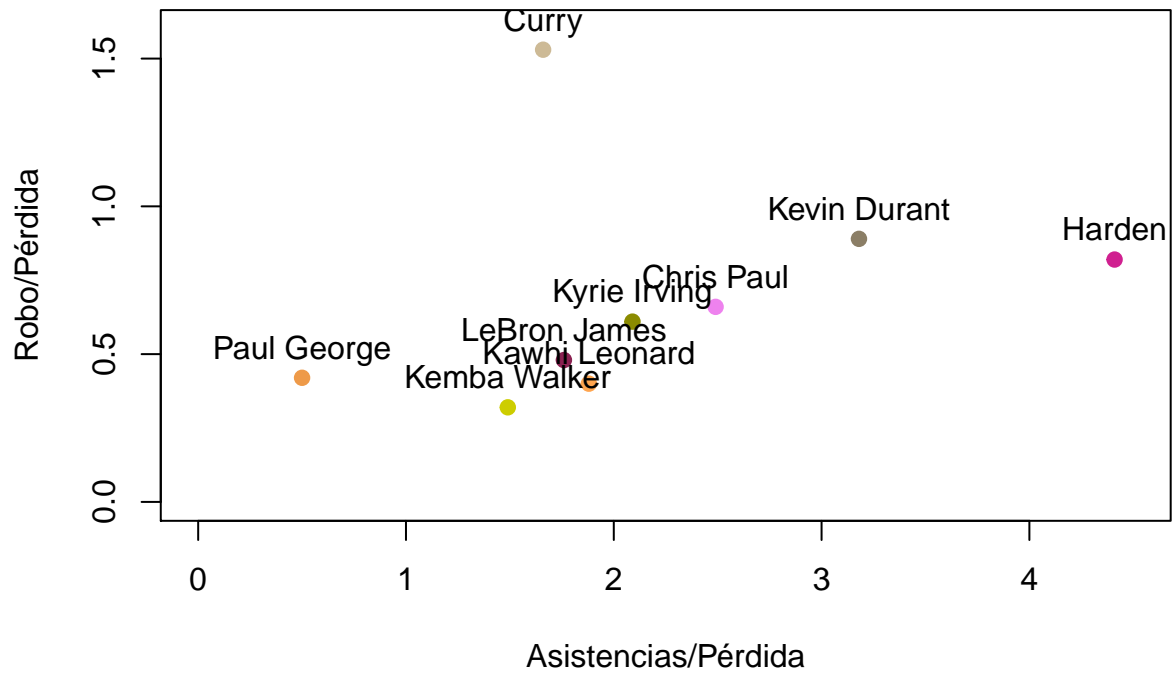
Frecuencia absoluta de asistencias



Cantidad de asistencias por jugador.

```
Jugadores_top=c("James Harden","LeBron James","Stephen Curry","Kevin Durant",
                "Kemba Walker","Kyrie Irving","Kawhi Leonard","Paul George",
                "Chris Paul")
data3=data[as.character(data$Nombre)%in% Jugadores_top,23:24]
data3=na.omit(data3)
plot(data3,col=c("violetred","violetred4","wheat3","wheat4",
                "yellow3","yellow4","tan1","tan2","violet"),
     main="Gráfico Multidimensional",xlab="Asistencias/Pérdida",
     pch=19,xlim=c(0,4.5),ylim=c(0,1.6))
text(data3,labels=c("Harden","LeBron James","Curry","Kevin Durant",
                    "Kemba Walker","Kyrie Irving","Kawhi Leonard",
                    "Paul George","Chris Paul"),pos=c(rep(3,times=9)))
```

Gráfico Multidimensional

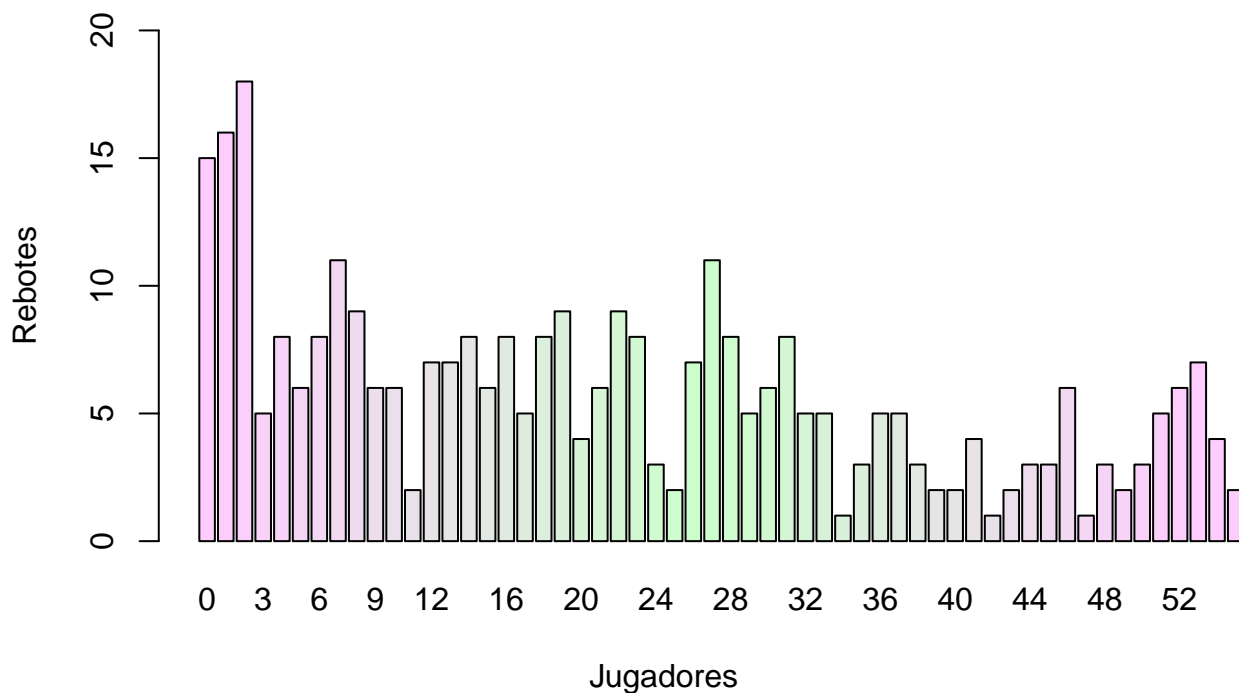


Robos/Pérdidas y Asistencias/Pérdida de los mejores jugadores de la liga. Como vemos, Harden es el jugador con más asistencias por pérdidas, lo que quiere decir que es un jugador bastante eficiente en este aspecto. Por el lado de los robos, el jugador que sobresale por encima de los demás es Stephen Curry.

3.2 Altura relacionada con otros datos:

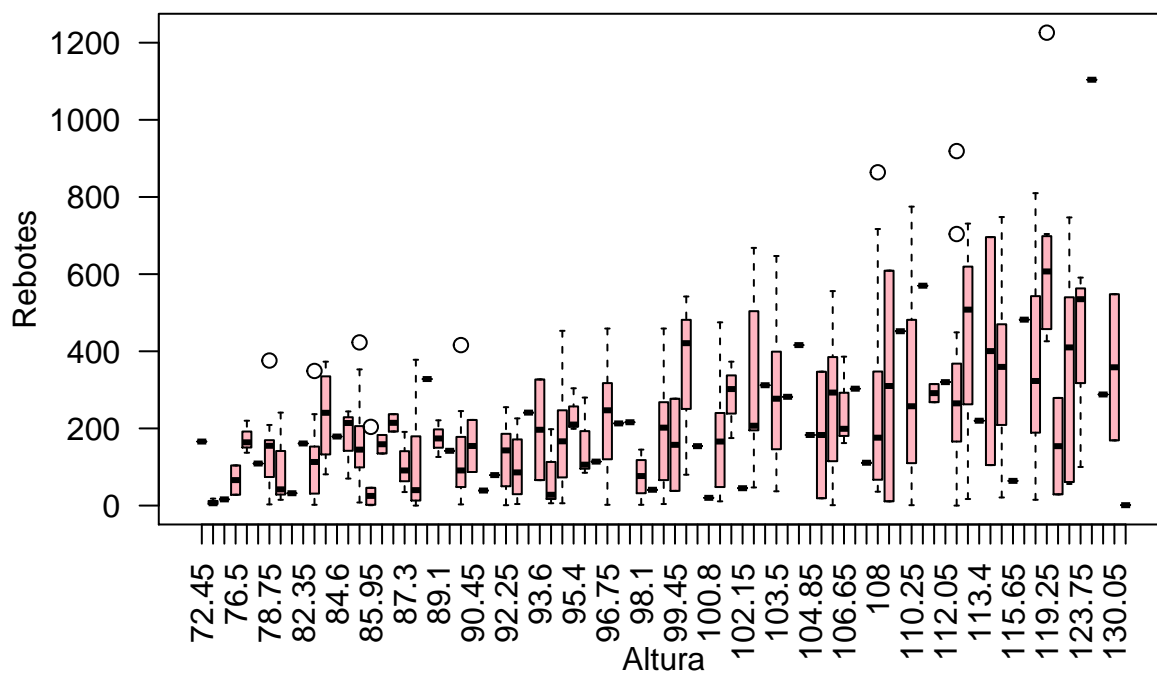
```
mis.colores.2 <- colorRampPalette(c("#ffccff", "#ccffcc", "#ffccff"))
barplot(table(data$`Rebotes ofensivos`), main = "Frecuencia de rebotes por jugador",
        xlab = "Jugadores", ylab = "Rebotes", col = mis.colores.2(54), xlim=c(0,60), ylim=c(0,20))
```


Frecuencia de rebotes por jugador



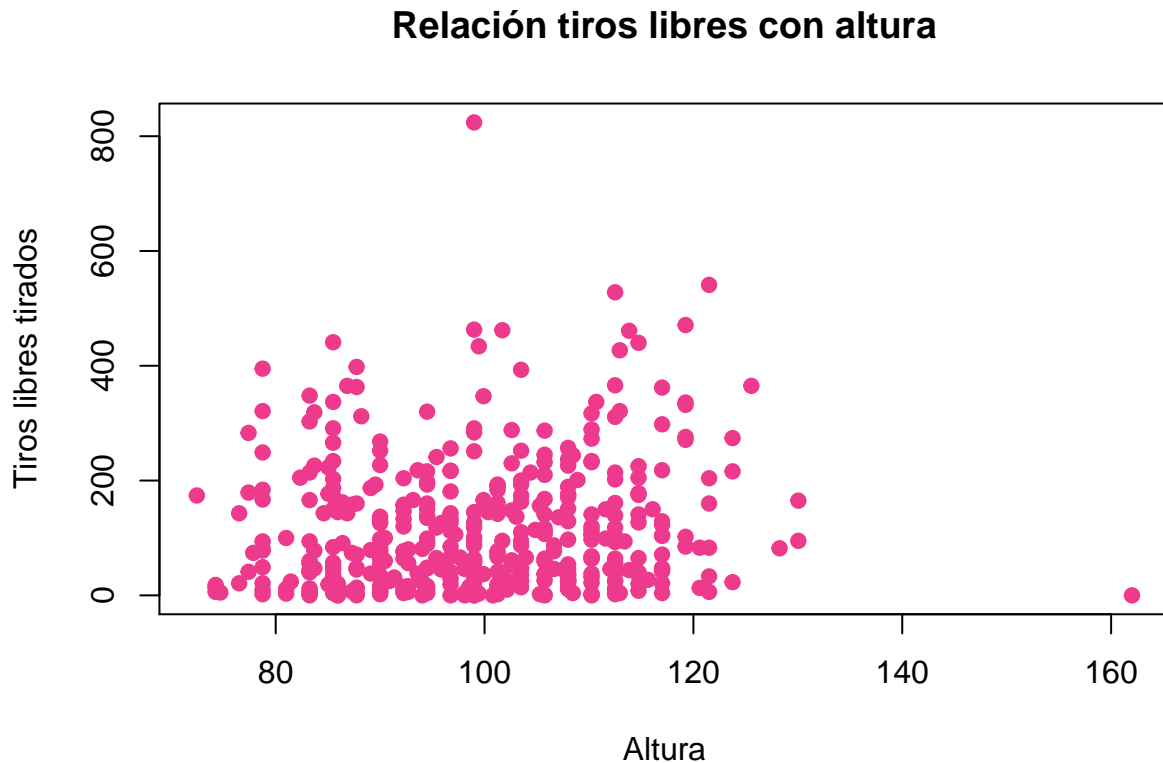
```
boxplot(data$Rebotes~data$Altura,col="lightpink", las=2, cex.label=2,xlab="Altura",
        ylab="Rebotes",main="Rebotes por altura")
```

Rebotes por altura



Este gráfico nos muestra la cantidad de rebotes totales que cogen los jugadores en relación a su altura. Como era de esperar, los jugadores más altos por lo general, son los que más rebotes capturan.

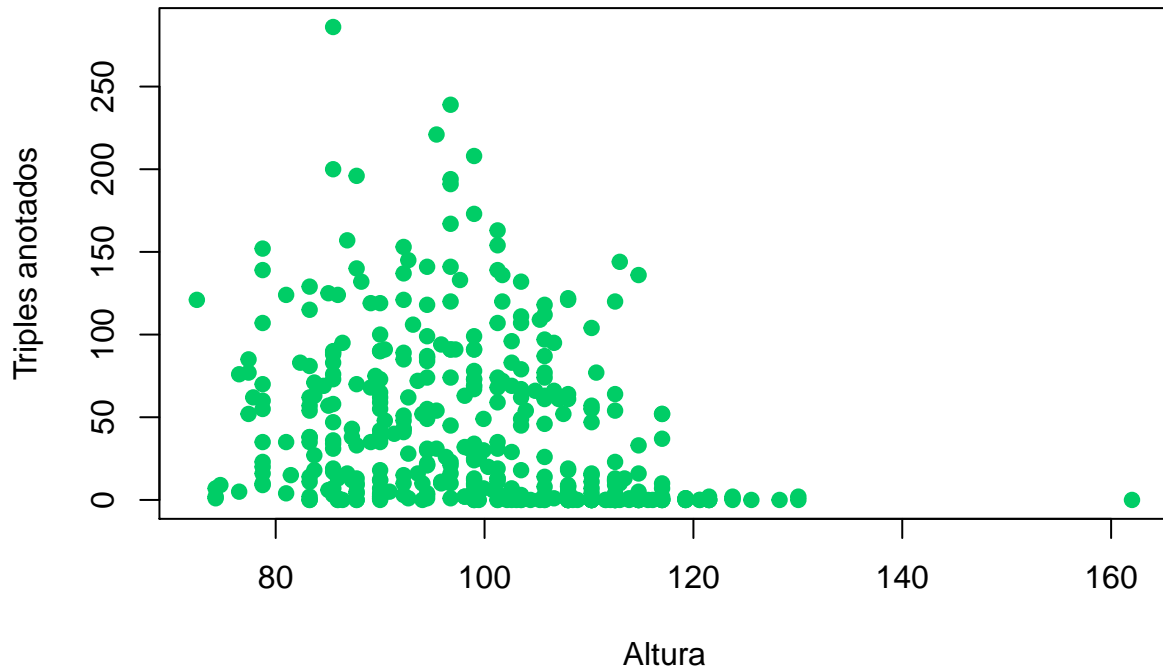
```
plot(data$`Tiros libres intentados`~data$Altura,
     xlab= "Altura", ylab="Tiros libres tirados",
     col="violetred2",main="Relación tiros libres con altura",
     pch=19)
```



De este gráfico sacamos la conclusión de que los jugadores más altos, como juegan cerca del aro, son los que suelen conseguir sacar más faltas y por lo tanto ir a la línea de tiros libres más frecuentemente.

```
plot(data$`3P anotados`~data$Altura,xlab= "Altura",
     ylab="Triples anotados",col="springgreen3",
     main="Relación triples anotados con la altura",pch=19)
```

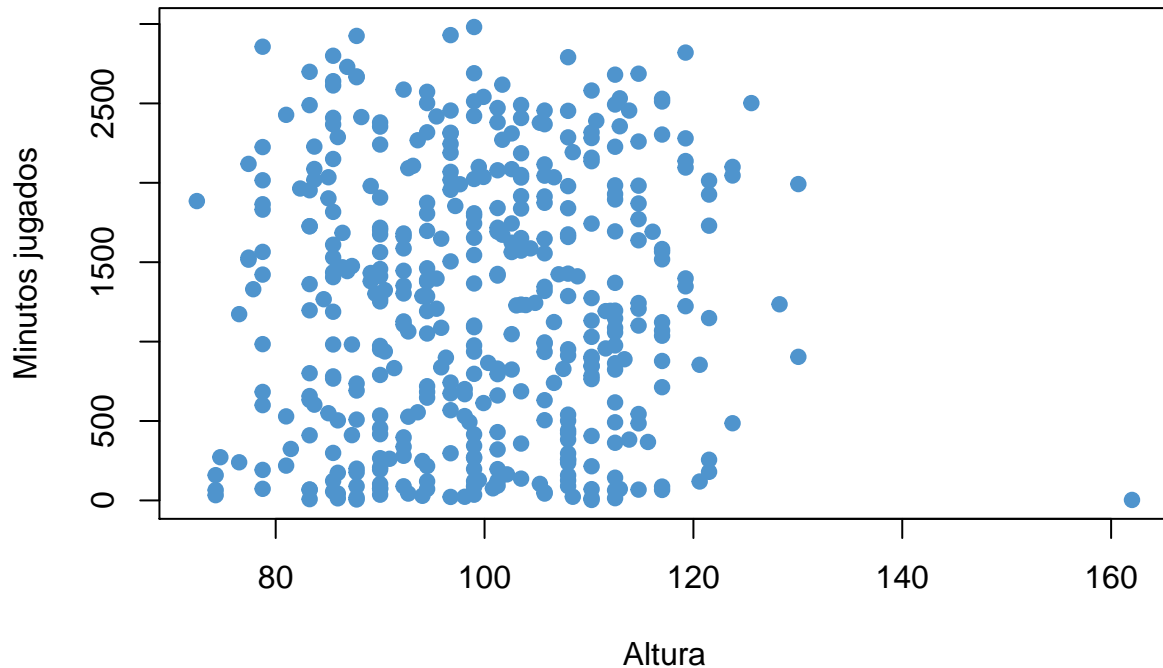
Relación triples anotados con la altura



Podemos comprobar como los mejores triplistas son jugadores de estatura “media” en la nba. De hecho, tiene bastante sentido, ya que cuando pensamos en jugadores de mayor estatura, los relacionamos inevitablemente cerca de la canasta, posteando o reboteando. Sería un “desperdicio” posicionarlos lejos de donde supuestamente tienen una ventaja. En cambio, los jugadores de menor estatura, suelen ser los que están más lejos del aro y por lo tanto es normal que estos sean los que más se atreven desde el perímetro. Además, los jugadores bajitos también suelen ser mejores en el triple ya que tienen el centro de gravedad más bajo y pueden efectuar una buena mecánica de tiro de forma mucho más rápida.

```
plot(data$Minutos~data$Altura,xlab= "Altura", ylab="Minutos jugados",
      col="steelblue3",main="Relación de minutos en cancha con la altura",
      pch=19)
```

Relación de minutos en cancha con la altura



Cabría esperar que los jugadores más altos son los que más juegan en baloncesto, pero este gráfico nos desmiente ese mito. A pesar de que la altura en la NBA es importante; no es imprescindible. Como podemos observar, los minutos están bastante repartidos entre jugadores de todos los tamaños. De jugadores bajitos jugando casi 3000 minutos, a jugadores enormes jugando no más de 1000, y viceversa. El talento también influye mucho.

```
NBA_CURRY=data %>% filter(Equipo=="GSW",Nombre=="Stephen Curry") %>% summarise(
  Nombre="Stephen Curry",
  Triples=max(`3P anotados`),
  Asistencias=max(Asistencias),
  Robos=max(Robos),
  Tapones=max(Tapones))

knitr::kable(NBA_CURRY)
```

Nombre	Triples	Asistencias	Robos	Tapones
Stephen Curry	286	619	163	16

```
NBA_LILLARD=data %>% filter(Equipo=="POR",Nombre=="Damian Lillard") %>% summarise(
  Nombre="Damian Lillard",
  Triples=max(`3P anotados`),
  Asistencias=max(Asistencias),
  Robos=max(Robos),
  Tapones=max(Tapones))

knitr::kable(NBA_LILLARD)
```

Nombre	Triples	Asistencias	Robos	Tapones
Damian Lillard	196	507	97	21

Datos de jugadores concretos. En este caso he seleccionado a Stephen Curry y Damian Lillard porque son dos de los bases más destacados de la liga.

```
Instituto= table(data$Instituto)
head(Instituto)
```

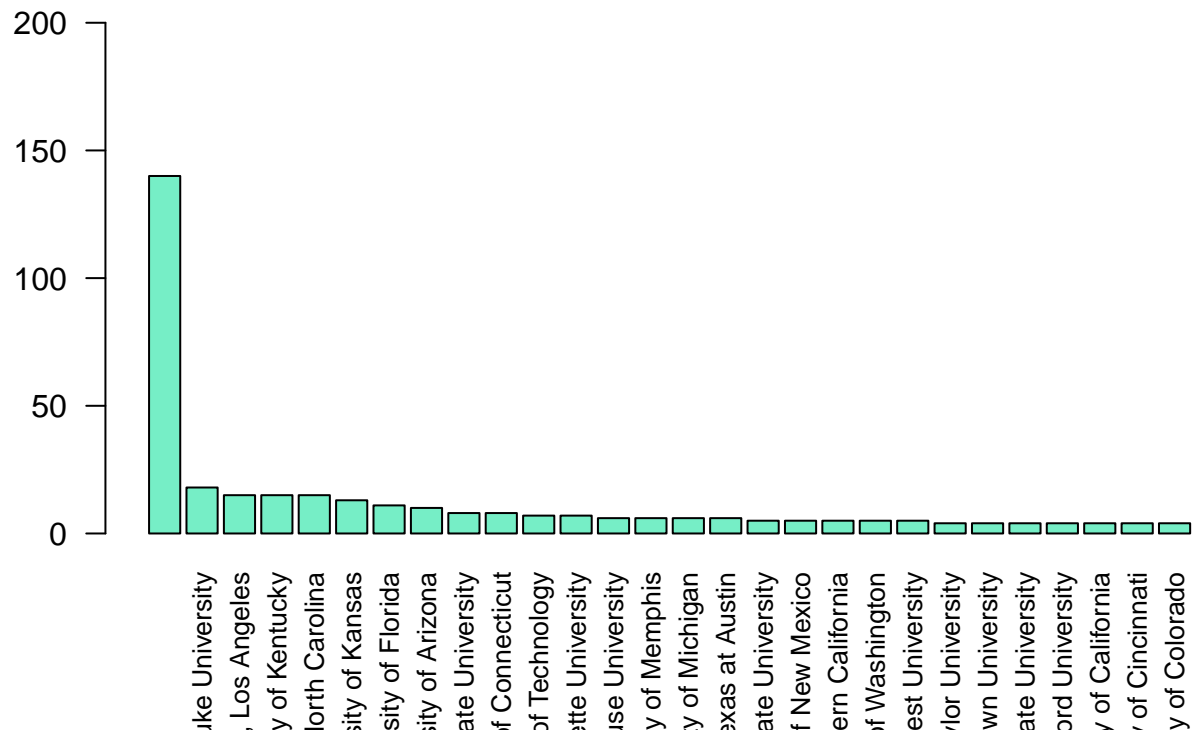
```
##
##              Arizona State University      Baylor University
##              140                      2                      4
##      Belmont University      Blinn College      Boston College
##              1                      1                      2
```

```
max(Instituto)
```

```
## [1] 140
```

```
barplot(sort(Instituto,decreasing=TRUE),
        las=2,cex.names = 0.8,main="Cantidad de jugadores por instituto
que han llegado a la NBA",
        xlab="", col="aquamarine2",ylim=c(0,200),xlim=c(0,30))
```

Cantidad de jugadores por instituto que han llegado a la NBA



Como podemos ver, en general los jugadores que llegaron a la NBA se han distribuido a lo largo de todas las universidades de los 50 estados. Sin embargo, algunas universidades tienen más caché que otras, son vistas con mejores ojos por los jugadores y los analistas, por lo que generalmente los mejores prospectos se marchan a las mejores universidades.

Se puede ver como una sobresale por encima de todas, pero en este caso son los jugadores que no han asistido a la universidad, ya sea porque han llegado del instituto a la NBA directamente, o por que estuvieron jugando en otras ligas.

De las universidades más escogidas encontramos Duke University. Esto no es de extrañar ya que es de las más prestigiosas. Aquí han nacido grandes jugadores como Kyrie Irving, Zion Williamson o Jayson Tatum, entre otros.

```
NBA_resumen_season= data %>% group_by(Equipo) %>% summarise(N=n(),
                                                             Mean_pts=mean(Puntos),
                                                             Sd_pts=sd(Puntos),
                                                             Max_pts=max(Puntos),
                                                             Min_pts=min(Puntos))

NBA_resumen_season
```

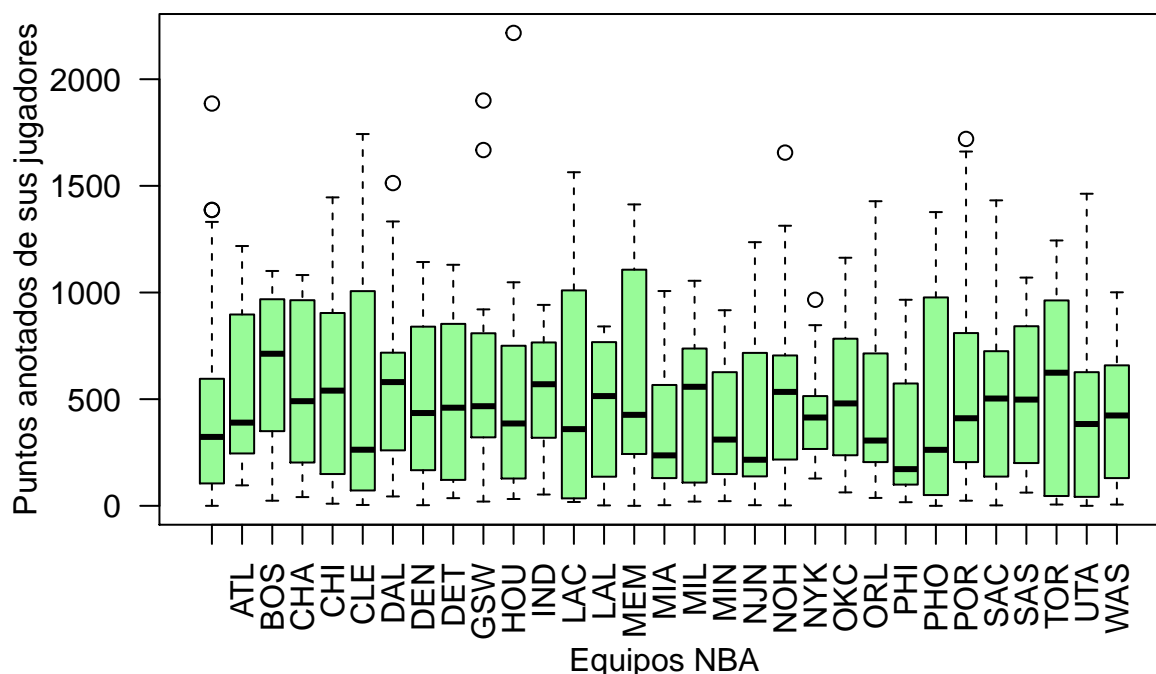
```
## # A tibble: 31 x 6
##   Equipo      N Mean_pts Sd_pts Max_pts Min_pts
##   <fct> <int>    <dbl> <dbl>    <int>    <int>
## 1 ""         68    417.   387.   1886      0
## 2 "ATL"       15    573.   404.   1218     96
## 3 "BOS"        8    648.   392.   1101     24
## 4 "CHA"       14    518.   375.   1082     41
## 5 "CHI"       14    590.   455.   1446     10
## 6 "CLE"       12    544.   642.   1743      4
## 7 "DAL"       15    588.   443.   1513     44
## 8 "DEN"       11    519.   408.   1143      3
## 9 "DET"       11    531.   420.   1130     36
## 10 "GSW"      15    640.   549.   1900     20
## # ... with 21 more rows
```

```
knitr::kable(NBA_resumen_season)
```

Equipo	N	Mean_pts	Sd_pts	Max_pts	Min_pts
	68	416.5735	387.2092	1886	0
ATL	15	572.7333	404.1010	1218	96
BOS	8	648.5000	391.5993	1101	24
CHA	14	518.4286	374.8462	1082	41
CHI	14	590.3571	454.9346	1446	10
CLE	12	544.0000	641.9919	1743	4
DAL	15	588.2667	442.8538	1513	44
DEN	11	518.8182	407.6299	1143	3
DET	11	530.9091	419.5992	1130	36
GSW	15	639.6000	548.6547	1900	20
HOU	11	567.0000	645.9474	2217	32
IND	15	544.0000	273.1182	942	53
LAC	14	574.7857	582.2333	1564	18
LAL	16	456.7500	331.0465	841	2
MEM	14	574.0714	466.9939	1413	0

Equipo	N	Mean_pts	Sd_pts	Max_pts	Min_pts
MIA	12	360.5000	352.4621	1007	3
MIL	12	503.5833	365.4592	1055	20
MIN	16	383.8125	294.0447	917	22
NJN	13	470.1538	440.9289	1236	3
NOH	16	542.8750	453.5050	1656	2
NYK	16	449.6875	239.1693	966	128
OKC	15	530.4667	357.5410	1163	63
ORL	15	523.1333	446.2682	1428	37
PHI	15	362.0000	331.6855	966	17
PHO	16	503.6250	519.7365	1377	0
POR	16	587.3125	518.8004	1720	24
SAC	13	526.4615	499.5402	1432	2
SAS	16	531.5000	353.8224	1070	62
TOR	14	562.7857	492.3390	1244	6
UTA	16	421.7500	454.5755	1463	0
WAS	16	434.6875	334.1101	1001	6

```
boxplot(data$Puntos~data$Equipo,las=2,col="palegreen",xlab="Equipos NBA",
        ylab="Puntos anotados de sus jugadores")
```

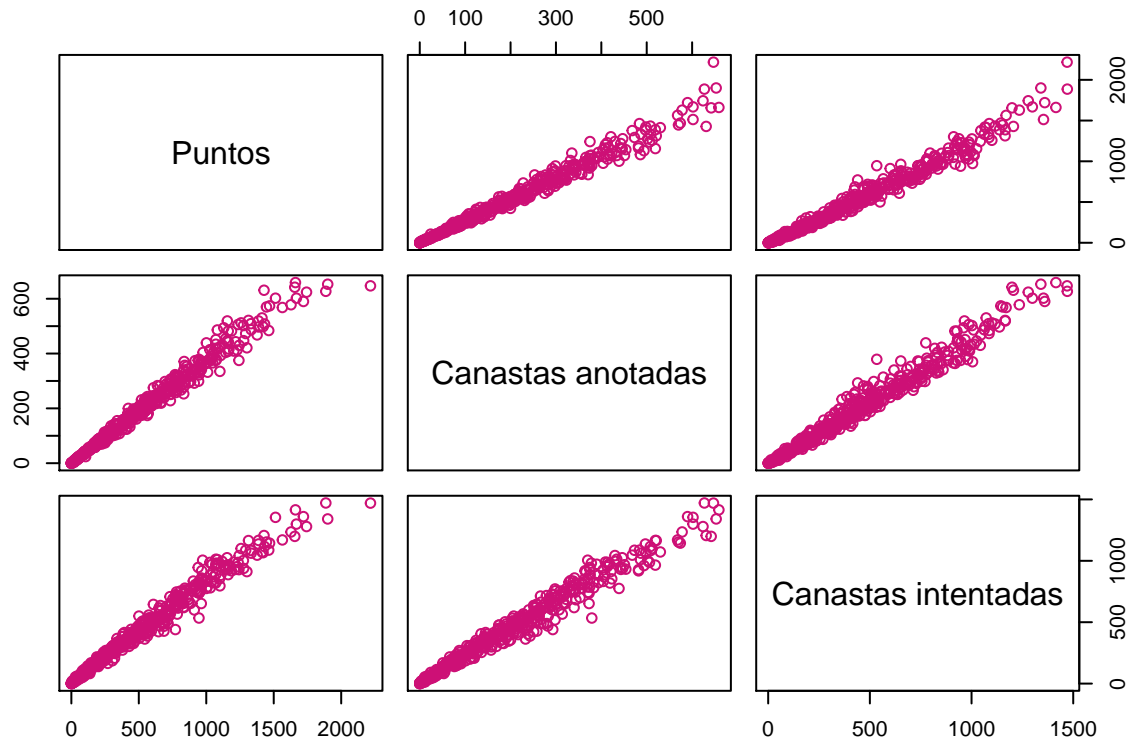


Este gráfico nos muestra la cantidad de puntos totales que ha encestado en todo su conjunto los equipos de la NBA de esta temporada. Mirando la imagen, se puede observar que los equipos más anotadores son los Memphis Grizzlies, seguido de Los Ángeles Clippers, y finalmente la medalla de bronce es para los Phoenix Suns. Todos pertenecientes a la conferencia Oeste. Pero si miramos sus clasificaciones en liga, los Grizzlies quedaron quintos, los Clippers terceros y los Suns décimos. Podemos deducir que a pesar de que tuviesen un muy buen ataque, no eran equipos tan balanceados en defensa.

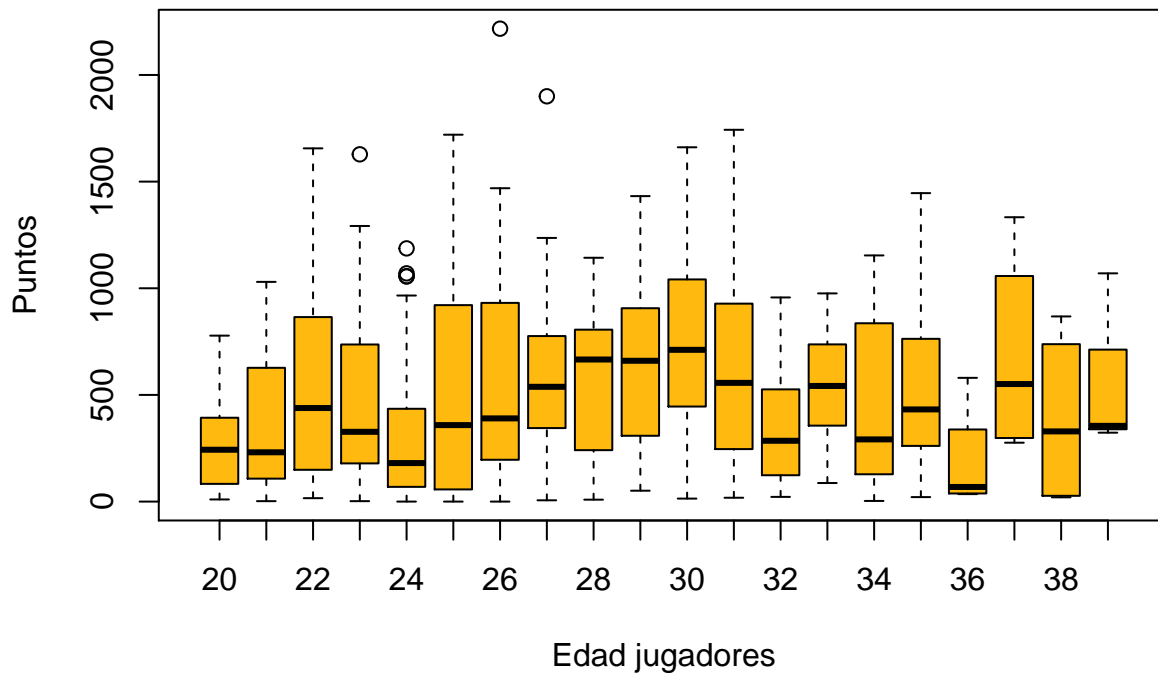
En la otra cara de la moneda, vemos a los equipos menos anotadores, los New York knicks, Miami Heat y Philadelphia 76ers. De hecho, si miramos sus clasificaciones(todos de la conferencia Este),quedaron últimos, décimos y penúltimos respectivamente. Podemos entender que sus bajas anotaciones se han visto reflejadas en una temporada bastante pobre.

Como conclusión final, la anotación en la NBA es muy importante, ya que los equipos con bajo porcentaje de anotación se han visto mermados por este aspecto hasta el punto de caer al fondo de la clasificación. Sin embargo, si un equipo NBA quiere aspirar a lo más alto, deberá tener en cuenta tanto ataque como defensa, ya que sin este último se quedarán en un buen equipo y nada más.

```
plot(data[,4:6],col="deeppink3")
```



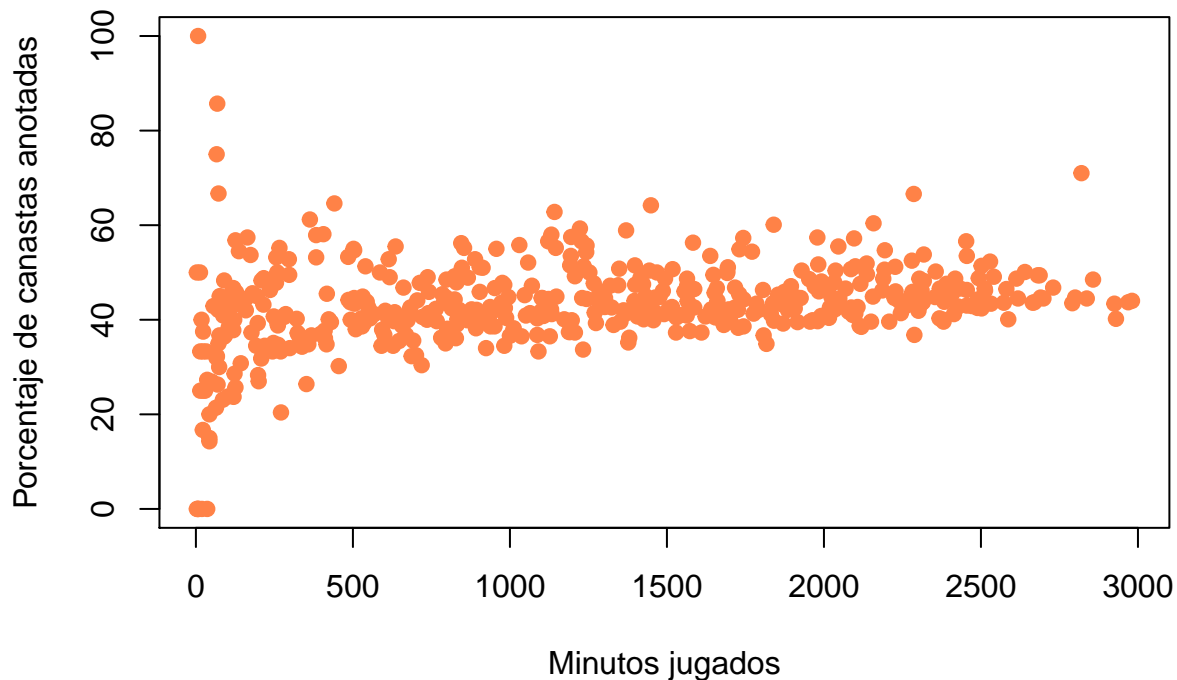
```
boxplot(data$Puntos~data$Edad,xlab="Edad jugadores",
        ylab="Puntos",col="darkgoldenrod1")
```

En este histograma se muestran la cantidad de puntos metidos en esta temporada en relación a la edad. Podemos observar que los jugadores que más anotan son en general los que oscilan entre 25 y 31 años. Esto también es lógico puesto que son las edades de madurez donde el cuerpo humano alcanza las mayores capacidades físicas. Los jóvenes por lo general anotan menos por la inexperiencia. Los más veteranos anotan menos por el desgaste físico de los años.

```
plot(data$`Porcentaje canastas`~data$Minutos,xlab= "Minutos jugados",
      ylab="Porcentaje de canastas anotadas",col="sienna1",
      main="Relación de las canastas anotadas con los minutos jugados",
      pch=19)
```

Relación de las canastas anotadas con los minutos jugados



Podemos observar como la muestra se distribuye de forma bastante uniforme a lo largo del gráfico. Esto quiere decir que en general los porcentajes, tanto de jugadores con un rol de suplente, como de los que son titulares indiscutibles, son bastante similares.

Lo que uno podría pensar es que los jugadores que juegan más minutos, deberían tener los porcentajes más elevados, puesto que son los más talentosos y sin embargo no es así. Esto se debe a que los jugadores que juegan más, generalmente asumen más responsabilidades y por tanto tiran más. En cambio, los jugadores cuya función es más de rol, suelen seleccionar pocos tiros, pero que suelen ser más efectivos ya que buscan la máxima eficiencia para lograr entrar en el quinteto titular. Además, como pequeño apunte, muchos de las estrellas de la nba buscan inflar estadísticas, así que no es de extrañar que tiren de forma desmedida solo para su beneficio propio. Por lo que aunque la situación de tiro no sea la idónea (por lo tanto las posibilidades de anotar sean menores), lo asuman igual.

4 Aplicaciones de estadística de inferencia a los datos

4.1 Bondad de ajuste con χ^2 de Pearson

```
FreEquipo=table(data$Equipo)
chisq.test(FreEquipo)

##
##  Chi-squared test for given probabilities
##
## data:  FreEquipo
## X-squared = 185.67, df = 30, p-value < 2.2e-16
```

```
x=data$Robos
ks.test(x,"pnorm",3,1)
```

```
## Warning in ks.test(x, "pnorm", 3, 1): ties should not be present for the
## Kolmogorov-Smirnov test
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: x
## D = 0.85375, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

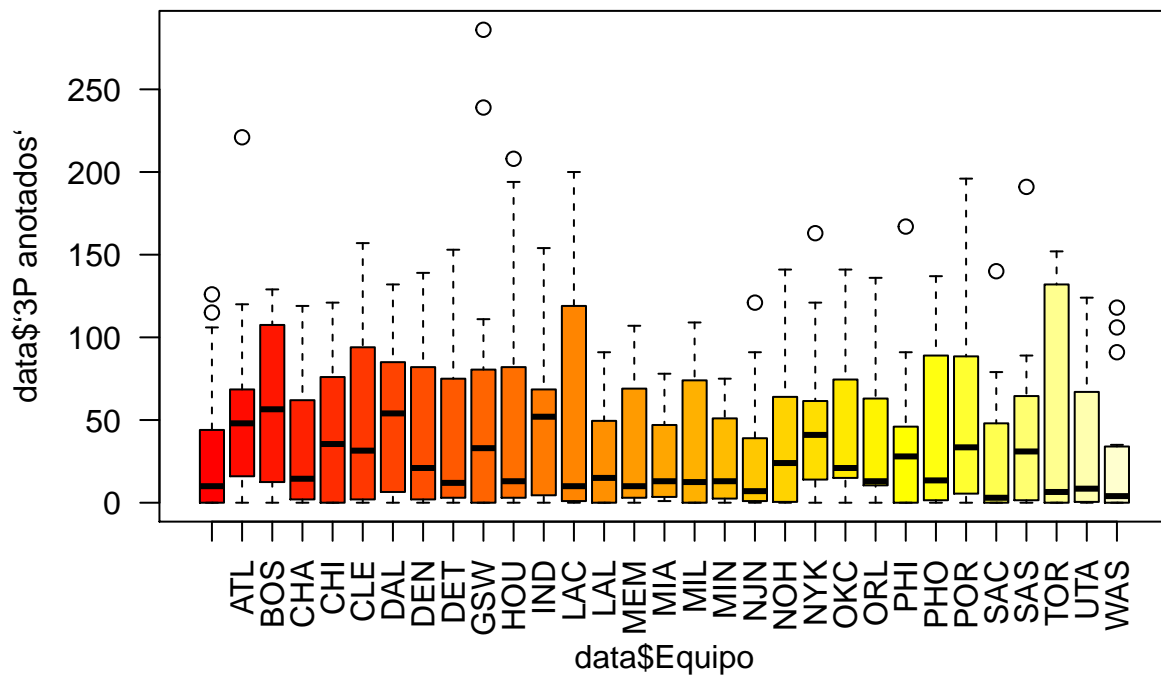
4.2 Anova 1

A continuación vamos a realizar un análisis de la varianza(*ANOVA*) para comparar entre las siguientes hipótesis:

$$\begin{cases} H_0 : \text{Media Equipo 1} = \text{Media Equipo 2} \\ H_1 : \text{Media Equipo 1} \neq \text{Media Equipo 2} \end{cases}$$

Primero, se puede ver un boxplot, donde se pueden apreciar qué variables hemos escogido para realizar el *ANOVA*.

```
boxplot(data$`3P anotados`~data$Equipo,col=heat.colors(32),las=2)
```



Este gráfico muestra los triples anotados por los equipos estudiados. Como vemos, el equipo con más cantidad de triples anotados son Los Ángeles Clippers. El equipo que más triples ha anotado en un partido son los Golden State Warriors, ya que en el gráfico aparecen dos valores atípicos asignados a ellos; valores bastante superiores a los de sus rivales. De hecho esto se puede justificar ya que Golden State cuentan en su plantilla

con los “*Splash Brothers*” (Stephen Curry y Klay Thompson), considerados de los mejores triplistas de la historia.

A continuación realizamos el anova comparando los triples anotados por equipo.

```
data=na.omit(data)
sol_anova=aoa(data$`3P anotados`~data$Equipo)
summary(sol_anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## data$Equipo  29  42521    1466   0.582  0.961
## Residuals   392 987571    2519
```

Comprobamos que el p-valor es muy grande. Concluimos, por tanto, que no tenemos evidencias suficientes para rechazar la hipótesis nula, y por lo tanto la media de los equipos sí que es distinta.

4.3 Test KS

A continuación vamos a realizar el test de Kolmogorov-Smirnov que utilizaremos para verificar si nuestro conjunto de datos se adapta a una distribución normal o no. Por consiguiente nuestro contraste será:

$$\begin{cases} H_0 : \text{Los datos proceden de una distribución normal} \\ H_1 : \text{Los datos no proceden de una distribución normal} \end{cases}$$

Para la realización de este apartado hemos cogido 4 equipos y hemos realizado 4 test ks, uno para cada muestra. La elección del número de equipos es simplemente por comodidad y porque hemos pensado que con esta cantidad ya sería suficiente para llegar a conclusiones más generales. Los equipos seleccionados son los que se encuentran en la parte alta de la clasificación.

```
Mejores_equipos=c("GSW","CLE","HOU","ATL")
data4=data[as.character(data$Equipo)%in% Mejores_equipos,c("Equipo","Puntos",
                                                             "Asistencias","Rebotes")]
data4=na.omit(data4)
data4
```

```
##      Equipo Puntos Asistencias Rebotes
## 4      ATL    213         30      162
## 5      ATL   1156        244      544
## 9      CLE      4          1         1
## 21     GSW    604        228      257
## 26     GSW    422        180      543
## 54     GSW     30         12        41
## 55     CLE     35          2        29
## 83     HOU     32          2        36
## 103    GSW    388         85      257
## 111    ATL    883        118      372
## 112    ATL    768        317      164
## 123    HOU    855        130      421
## 126    GSW    921        291      647
## 130    HOU    646         50      431
## 140    ATL     96         22       100
## 148    GSW    779        449      417
```

## 150	HOU	45	15	17
## 173	GSW	827	116	453
## 195	HOU	2217	565	459
## 197	CLE	250	24	62
## 198	GSW	62	2	37
## 216	ATL	1162	513	184
## 233	CLE	136	26	41
## 238	HOU	184	25	279
## 241	ATL	135	13	39
## 262	GSW	254	48	73
## 271	ATL	390	78	222
## 276	CLE	1228	168	731
## 282	GSW	1668	222	247
## 285	HOU	182	85	118
## 288	ATL	911	196	304
## 292	CLE	1628	389	237
## 301	GSW	467	99	91
## 302	CLE	1743	511	416
## 324	GSW	791	72	325
## 330	CLE	319	204	126
## 338	CLE	109	46	91
## 339	ATL	197	25	119
## 340	ATL	529	75	199
## 354	HOU	74	11	39
## 364	GSW	20	6	17
## 369	HOU	568	192	237
## 375	ATL	1218	223	570
## 377	ATL	356	49	189
## 424	GSW	461	259	183
## 426	CLE	276	51	202
## 429	ATL	299	155	79
## 434	GSW	1900	619	341
## 441	HOU	386	36	220
## 443	ATL	278	75	223
## 450	CLE	785	55	591
## 460	HOU	1048	209	459
## 483	CLE	15	8	5

```
X=data4$Puntos
F=as.factor(as.character(data4$Equipo))
levels(F)
```

```
## [1] "ATL" "CLE" "GSW" "HOU"
```

```
muestra1=X[F=="GSW"]
ks.test(muestra1,"pnorm",mean(muestra1),sd(muestra1))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: muestra1
## D = 0.17068, p-value = 0.7134
## alternative hypothesis: two-sided
```

```
muestra2=X[F=="CLE"]
ks.test(muestra2,"pnorm",mean(muestra2),sd(muestra2))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: muestra2
## D = 0.30367, p-value = 0.177
## alternative hypothesis: two-sided
```

```
muestra3=X[F=="HOU"]
ks.test(muestra3,"pnorm",mean(muestra3),sd(muestra3))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: muestra3
## D = 0.20377, p-value = 0.6797
## alternative hypothesis: two-sided
```

```
muestra4=X[F=="ATL"]
ks.test(muestra4,"pnorm",mean(muestra4),sd(muestra4))
```

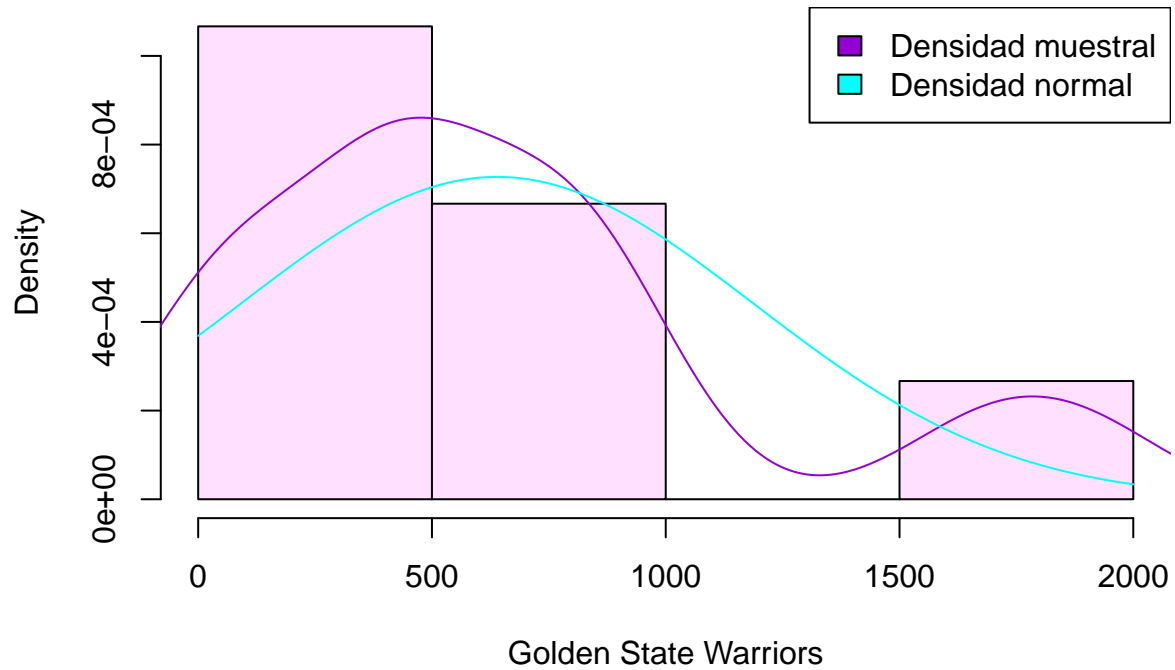
```
##
## One-sample Kolmogorov-Smirnov test
##
## data: muestra4
## D = 0.20777, p-value = 0.4743
## alternative hypothesis: two-sided
```

Hemos decidido crear una tabla más pequeña solo con los datos que nos interesan y con los equipos seleccionados. Y previamente hemos realizado los test KS. Como vemos, el p-valor de todos los test es bastante alto, por lo que no podemos rechazar la hipótesis nula, por lo que concluimos que los datos proceden de una distribución normal.

A continuación realizamos el histograma de puntos de los Golden State Warriors:

```
hist(muestra1,freq=FALSE,main="Histograma de canastas",xlab="Golden State Warriors",
     col="thistle1")
lines(density(muestra1),col="darkviolet")
curve(dnorm(x,mean(muestra1),sd(muestra1)),add=TRUE,col="cyan")
legend("topright",c("Densidad muestral","Densidad normal"),fill=c("darkviolet","cyan"))
```

Histograma de canastas

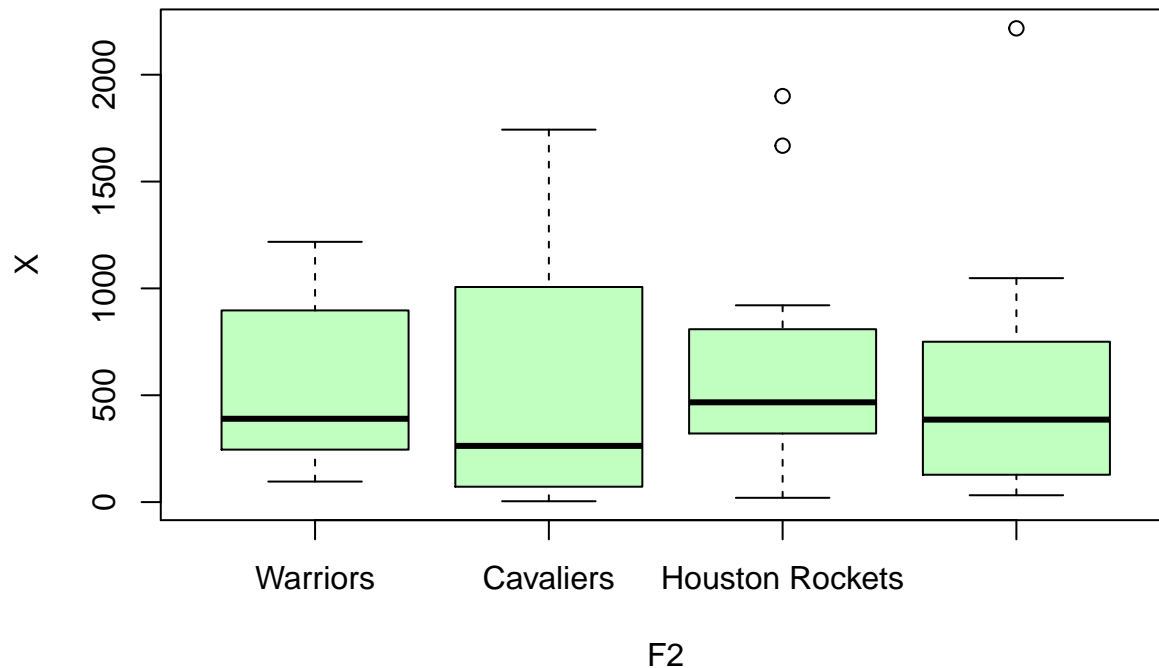


En el gráfico podemos ver la comparación entre la densidad muestral y la normal. Vemos como la empírica oscila de forma más bruca que la normal.

4.4 Boxplot de Puntos por equipo

```
F2=F
levels(F2)=paste0(c("Warriors","Cavaliers","Houston Rockets","Atlanta Hawks"))
boxplot(X~F2,main="Boxplot de puntos por equipo",col="darkseagreen1")
```

Boxplot de puntos por equipo



En este boxplot vemos las comparaciones entre los mejores equipos de la liga.

4.5 Segunda ANOVA

A continuación vamos a realizar un segundo anova entre los puntos anotados por cada equipo.

```
sola_aov=aov(X~F)
summary(sola_aov)
```

```
##           Df    Sum Sq Mean Sq F value Pr(>F)
## F           3    70653   23551   0.076  0.973
## Residuals  49 15206645   310340
```

De nuevo obtenemos un p-valor muy muy grande, y llegamos a la misma conclusión que en el primer anova; la media de puntos entre los cuatro equipos es distinta.

4.6 Comparación de medias dos a dos con ajuste de Bonferroni

```
pairwise.t.test(X,F,p.adjust.method = "bonferroni")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  X and F
##
```

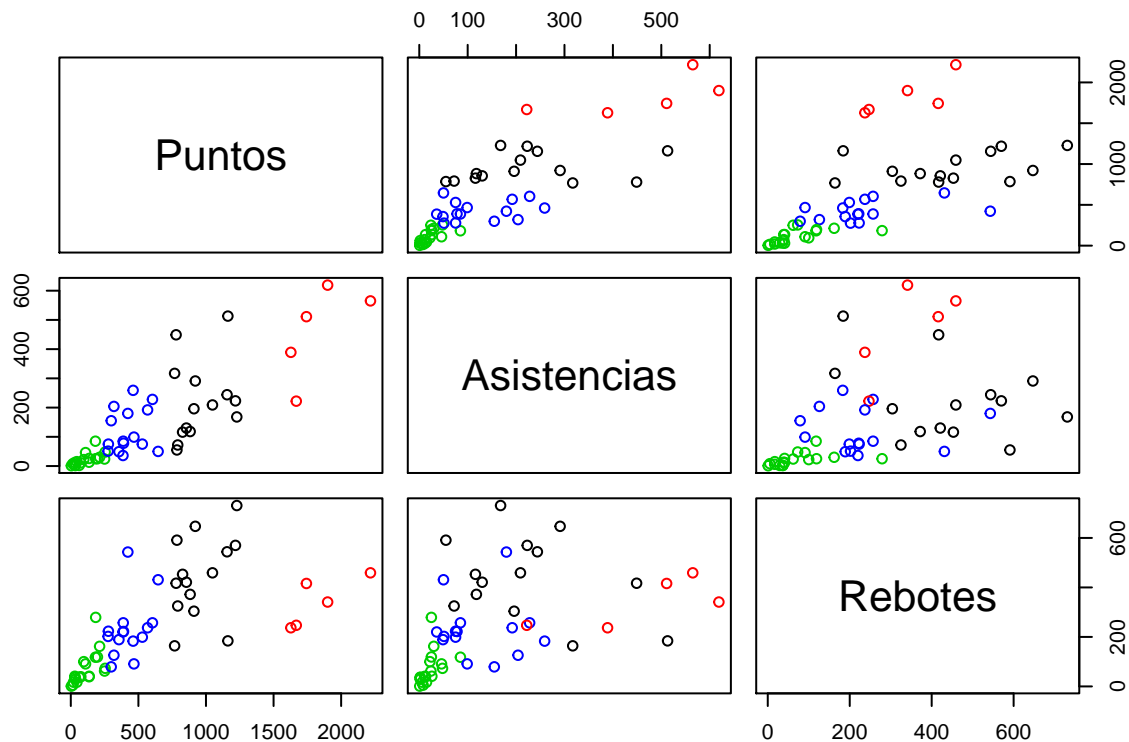


```
##      ATL CLE GSW
## CLE 1  -  -
## GSW 1   1  -
## HOU 1   1  1
##
## P value adjustment method: bonferroni
```

```
cluster4=kmeans(data4[,2:4],centers = 4)
table(cluster4$cluster,F)
```

```
##      F
##      ATL CLE GSW HOU
## 1     6   2   4   2
## 2     0   2   2   1
## 3     4   6   4   5
## 4     5   2   5   3
```

```
plot(data2[,2:4],col=cluster4$cluster)
```



4.7 Clustering

```
X1=data$`Canastas intentadas`
X2=data$`3P intentados`
X3=data$`Tiros libres intentados`
Datos=data.frame(X1,X2,X3)

kmeans(X1,4,algorithm = "Lloyd",iter.max = 40)
```

```

## K-means clustering with 4 clusters of sizes 155, 85, 53, 129
##
## Cluster means:
##      [,1]
## 1  96.23226
## 2 705.91765
## 3 1056.60377
## 4  391.79070
##
## Clustering vector:
##  [1] 1 2 1 1 3 3 4 4 1 4 4 4 4 1 1 4 1 3 4 4 1 4 1 4 1 4 3 2 1 1 4 3 1 4 3 4 2
## [38] 1 1 3 2 2 2 2 1 4 3 1 1 4 3 1 1 4 1 4 2 2 4 1 2 4 4 1 2 4 1 1 2 3 1 4 2 1
## [75] 1 4 2 1 4 2 2 1 3 4 2 1 2 4 4 1 1 4 2 4 1 2 4 1 3 3 2 2 3 2 4 1 1 3 4 2 1
## [112] 2 4 1 1 4 1 1 1 4 2 1 3 3 2 1 2 1 2 2 1 1 1 1 1 4 4 2 2 2 4 1 3 3 4 1 2 1
## [149] 1 2 2 1 1 4 4 2 4 1 1 4 1 2 4 4 3 4 1 1 1 1 1 4 2 2 1 1 1 1 1 3 3 1 4 1 1
## [186] 4 2 1 1 4 1 3 4 3 4 4 1 4 3 1 1 1 1 4 1 1 1 4 4 2 1 2 1 1 2 1 4 1 2 4 1 1
## [223] 1 1 2 4 3 1 1 2 4 3 1 4 4 3 2 4 2 4 4 3 2 4 1 4 1 2 3 1 4 3 3 2 1 4 1 4 4
## [260] 3 1 1 4 3 2 1 2 1 1 2 1 2 3 2 4 2 4 2 1 3 2 4 4 2 1 4 1 4 1 3 4 3 4 1 1 4
## [297] 4 1 2 3 1 2 2 4 4 1 2 4 2 3 1 2 2 1 4 4 4 1 4 1 4 3 1 3 2 4 1 1 4 4 1 4 2
## [334] 4 4 4 4 1 1 1 3 4 1 1 2 4 4 2 1 2 3 4 1 2 1 4 1 1 2 1 1 4 4 1 4 1 4 4 4 1
## [371] 2 1 4 3 4 1 4 4 4 4 4 2 4 4 2 1 2 2 3 1 4 2 4 1 3 4 3 1 3 1 1 1 3 4 1 1 1
## [408] 3 4 2 2 2 4 1 1 4 3 1 2 3 4 1
##
## Within cluster sum of squares by cluster:
## [1] 834487.6 688372.4 1177806.7 965821.3
## (between_SS / total_SS = 92.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"

```

```

km=kmeans(X1,4,algorithm = "Lloyd",iter.max = 40)
km$cluster

```

```

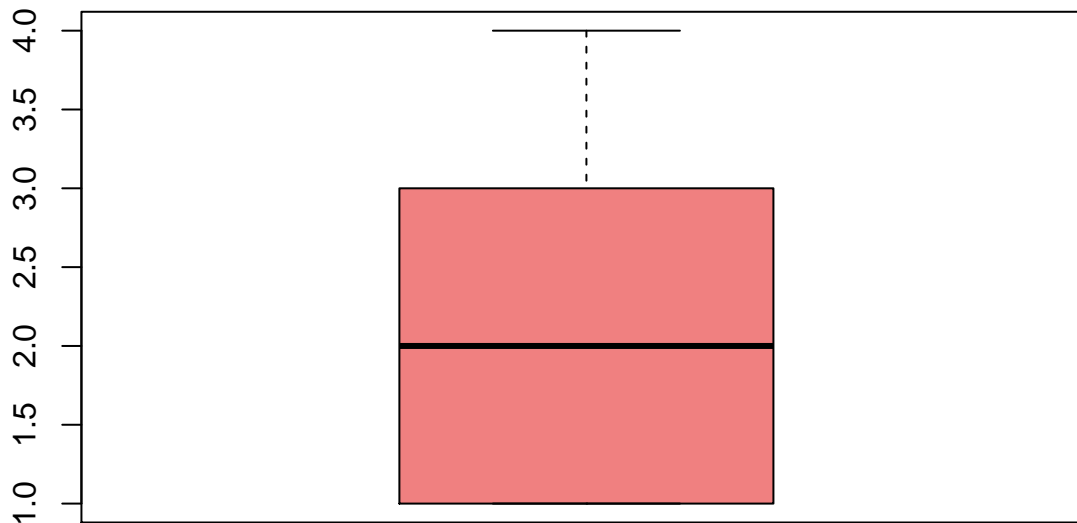
##  [1] 1 2 1 1 4 4 3 3 1 3 3 3 3 1 1 3 1 4 3 3 1 3 1 3 1 3 4 2 1 1 3 4 1 3 4 3 2
## [38] 1 1 4 2 2 2 2 1 3 4 1 1 3 4 1 1 3 1 3 2 2 3 1 2 3 3 1 2 3 1 1 2 4 1 3 2 1
## [75] 1 3 2 1 3 2 2 1 4 3 2 1 2 3 3 1 1 3 2 3 1 2 3 1 4 4 2 2 4 2 3 1 1 4 3 2 1
## [112] 2 3 1 1 3 1 1 1 3 2 1 4 4 2 1 2 1 2 2 1 1 1 1 1 3 3 2 2 2 3 1 4 4 3 1 2 1
## [149] 1 2 2 1 1 3 3 2 3 1 1 3 1 2 3 3 4 3 1 1 1 1 1 3 2 2 1 1 1 1 1 4 4 1 3 1 1
## [186] 3 2 1 1 3 1 4 3 4 3 3 1 3 4 1 1 1 1 3 1 1 1 3 3 2 1 2 1 1 2 1 3 1 2 3 1 1
## [223] 1 1 2 3 4 1 1 2 3 4 1 3 3 4 2 3 2 3 3 4 2 3 1 3 1 2 4 1 3 4 4 2 1 3 1 3 3
## [260] 4 1 1 3 4 2 1 2 1 1 2 1 2 4 2 3 2 3 2 1 4 2 3 3 2 1 3 1 3 1 4 3 4 3 1 1 3
## [297] 3 1 2 4 1 2 2 3 3 1 2 3 2 4 1 2 2 1 3 3 3 1 3 1 3 4 1 4 2 3 1 1 3 3 1 3 2
## [334] 3 3 3 3 1 1 1 4 3 1 1 2 3 3 2 1 2 4 3 1 2 1 3 1 1 2 1 1 3 3 1 3 1 3 3 3 1
## [371] 2 1 3 4 3 1 3 3 3 3 3 2 3 3 2 1 2 2 4 1 3 2 3 1 4 3 4 1 4 1 1 1 4 3 1 1 1
## [408] 4 3 2 2 2 3 1 1 3 4 1 2 4 3 1

```

```

boxplot(km$cluster,col="lightcoral")

```



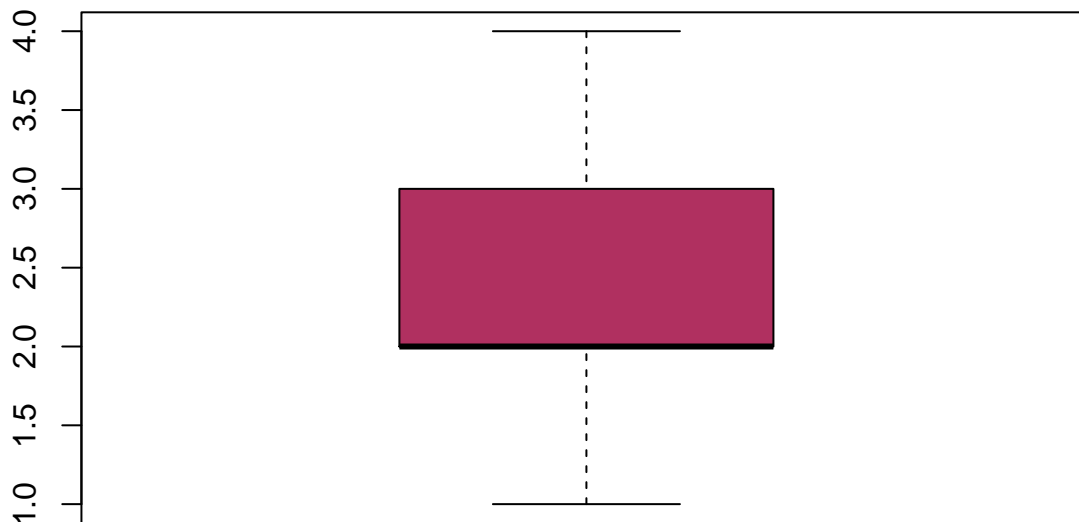
```
kmeans(X2,4,algorithm = "Lloyd",iter.max = 40)
```

```
## K-means clustering with 4 clusters of sizes 220, 89, 75, 38
##
## Cluster means:
##      [,1]
## 1  17.01818
## 2 132.98876
## 3 245.34667
## 4 409.89474
##
## Clustering vector:
##  [1] 1 3 1 1 1 1 3 1 1 1 2 1 2 2 1 1 1 1 3 1 2 1 1 1 1 1 1 3 1 1 1 4 1 2 4 2 4
## [38] 1 1 1 3 2 3 1 1 3 3 1 1 3 1 1 1 2 1 1 1 2 3 1 4 4 2 1 2 2 1 2 1 4 1 2 4 2
## [75] 1 1 3 1 1 3 3 1 4 2 3 2 4 1 3 1 1 2 2 1 1 1 1 1 2 1 3 2 1 3 3 1 1 3 2 2 1
## [112] 4 1 1 1 1 1 1 1 1 1 1 1 1 3 3 1 3 1 3 2 1 1 1 2 2 2 3 4 2 1 1 1 3 4 1 1 1 1
## [149] 1 4 3 1 1 3 3 4 2 1 1 3 1 4 3 2 4 1 2 1 1 1 1 3 2 2 1 2 1 1 1 3 3 1 2 1 1
## [186] 2 2 1 1 2 1 3 1 4 1 3 2 3 4 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 2 1 4 2 1 1
## [223] 2 1 3 2 3 1 1 1 2 4 1 2 1 4 3 1 3 2 2 4 2 1 2 1 1 4 4 1 3 4 2 2 1 2 1 1 2
## [260] 4 1 1 1 4 3 1 1 2 1 3 1 3 1 1 3 3 3 3 1 2 1 3 1 4 2 2 1 2 1 2 1 3 3 2 1 3
## [297] 1 1 4 3 1 1 1 1 1 1 1 3 2 3 1 1 2 3 1 1 2 2 2 3 1 2 1 1 3 3 2 1 1 2 1 1 3 2
## [334] 1 3 3 2 1 1 1 3 2 1 1 4 1 2 2 2 1 3 1 1 4 1 2 1 1 3 1 1 1 2 1 1 1 1 2 2 1
## [371] 3 1 2 4 3 1 1 1 1 1 1 4 2 1 1 1 4 1 3 1 1 2 3 1 4 2 4 2 3 1 1 1 3 1 1 1 1
## [408] 3 3 3 3 4 2 1 1 2 4 1 2 1 1 1
##
## Within cluster sum of squares by cluster:
## [1] 87171.93 93558.99 113522.99 223101.58
## (between_SS / total_SS = 92.8 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
km1=kmeans(X2,4,algorithm = "Lloyd",iter.max = 40)
km1$cluster
```

```
## [1] 2 3 2 2 2 2 3 2 2 2 1 2 1 1 2 2 2 2 3 2 1 2 2 2 2 2 2 3 2 2 2 4 2 1 4 1 4
## [38] 2 2 2 3 1 3 2 2 3 3 2 2 3 2 2 2 1 2 2 2 1 3 2 4 4 1 2 1 1 2 1 2 4 2 1 4 1
## [75] 2 2 3 2 2 3 3 2 4 1 3 1 4 2 3 2 2 1 1 2 2 2 2 2 1 2 3 1 2 3 3 2 2 3 1 1 2
## [112] 4 2 2 2 2 2 2 2 2 2 2 2 3 3 2 3 2 3 1 2 2 2 1 1 1 3 4 1 2 2 2 3 4 2 2 2 2
## [149] 2 4 3 2 2 3 3 4 1 2 2 3 2 4 3 1 4 2 1 2 2 2 2 3 1 1 2 1 2 2 2 3 3 2 1 2 2
## [186] 1 1 2 2 1 2 3 2 4 2 3 1 3 4 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 1 2 4 1 2 2
## [223] 1 2 3 1 3 2 2 2 1 4 2 1 2 4 3 2 3 1 1 4 1 2 1 2 2 4 4 2 3 4 1 1 2 1 2 2 1
## [260] 4 2 2 2 4 3 2 2 1 2 3 2 3 2 2 3 3 3 3 2 1 2 3 2 4 1 1 2 1 2 1 2 3 3 1 2 3
## [297] 2 2 4 3 2 2 2 2 2 2 3 1 3 2 2 1 3 2 2 1 1 1 3 2 1 2 2 3 3 1 2 2 1 2 2 3 1
## [334] 2 3 3 1 2 2 2 3 1 2 2 4 2 1 1 1 2 3 2 2 4 2 1 2 2 3 2 2 2 1 2 2 2 2 1 1 2
## [371] 3 2 1 4 3 2 2 2 2 2 2 4 1 2 2 2 4 2 3 2 2 1 3 2 4 1 4 1 3 2 2 2 3 2 2 2 2
## [408] 3 3 3 3 4 1 2 2 1 4 2 1 2 2 2
```

```
boxplot(km1$cluster,col="maroon")
```



```
kmeans(X3,4,algorithm = "Lloyd",iter.max = 40)
```

```
## K-means clustering with 4 clusters of sizes 30, 196, 77, 119
```

```
##
```

```
## Cluster means:
```

```
## [1,]
```

```
## 1 404.46667
```

```
## 2 28.31633
```

```
## 3 226.76623
```

```
## 4 118.42857
```

```
##
```

```
## Clustering vector:
```

```
## [1] 2 3 2 2 4 4 4 4 2 4 4 4 4 2 4 4 2 1 4 4 2 4 2 2 2 1 4 4 2 4 3 2 4 4 2 4
```

```
## [38] 2 3 1 4 4 4 3 2 4 3 2 2 4 3 2 2 4 2 4 4 3 2 2 3 2 4 4 3 2 2 2 4 1 2 2 4 2
```

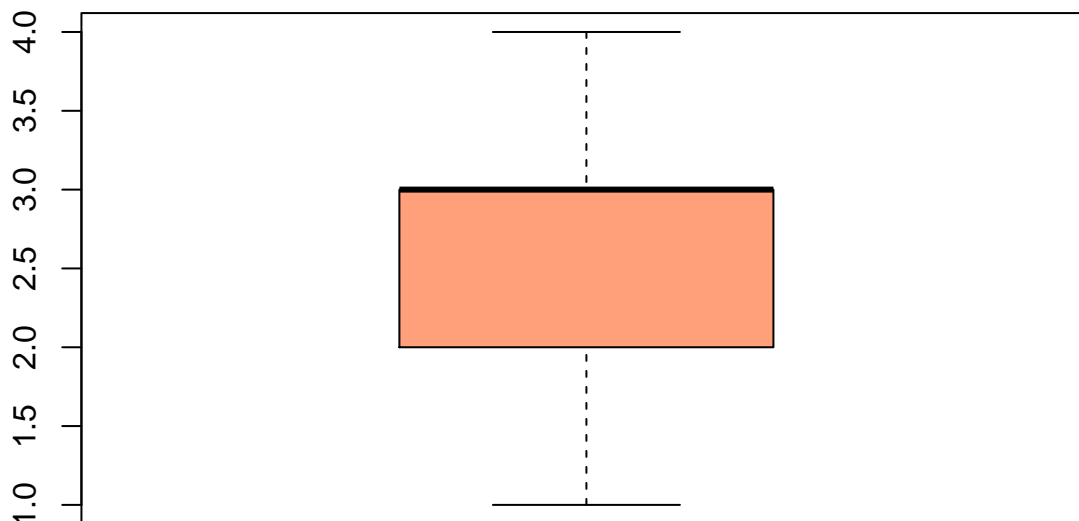
```
## [75] 2 2 3 2 4 4 3 2 1 2 3 2 4 2 2 2 2 2 3 4 2 4 1 2 1 1 3 3 1 3 4 2 2 3 4 4 2
```

```
## [112] 3 2 2 2 3 2 2 2 3 3 2 3 1 4 2 4 2 4 4 2 2 2 2 2 4 4 4 3 1 4 2 3 1 3 2 1 2
```

```
## [149] 2 2 4 2 4 2 2 1 4 4 2 4 2 3 2 4 1 4 2 2 2 2 2 2 4 3 2 2 2 2 2 3 1 2 3 2 2
```

```
## [186] 2 3 2 2 4 2 1 3 3 3 4 2 2 3 2 2 4 2 4 2 2 2 2 2 3 2 4 2 2 4 2 2 2 2 4 2 2
## [223] 2 2 3 4 3 2 4 3 4 4 2 4 2 1 3 4 4 2 4 3 3 4 2 4 2 4 3 2 2 1 1 4 2 4 2 2 4
## [260] 1 2 2 4 1 4 2 3 2 2 3 2 3 1 3 4 4 4 3 2 3 3 4 1 4 2 2 2 2 2 3 3 3 4 2 2 4
## [297] 2 2 3 3 2 3 3 2 2 2 4 2 3 3 2 4 4 2 3 3 4 2 2 2 2 1 2 1 3 4 2 2 2 4 2 4 4
## [334] 3 2 2 4 2 2 4 3 4 2 2 3 4 4 3 2 3 1 3 2 4 2 4 2 2 4 2 2 4 4 2 4 2 2 2 2
## [371] 3 2 2 1 2 2 3 3 4 2 4 2 2 4 3 2 4 3 3 2 4 4 2 2 4 4 4 2 1 2 4 2 3 3 2 2 2
## [408] 1 2 2 4 4 4 2 2 2 4 2 3 3 4 2
##
## Within cluster sum of squares by cluster:
## [1] 298183.47 91476.39 116133.79 98721.14
## (between_SS / total_SS = 89.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

```
km2=kmeans(X3,4,algorithm = "Lloyd",iter.max = 40)
boxplot(km2$cluster,col="lightsalmon")
```



```
km2$cluster
```

```
## [1] 3 4 3 3 2 2 2 2 3 2 2 2 2 3 2 2 3 1 2 2 3 2 3 3 3 3 1 2 2 3 2 4 3 2 2 3 2
## [38] 3 4 1 2 2 2 4 3 2 4 3 3 2 4 3 3 2 3 2 2 4 3 3 4 3 2 2 4 3 3 3 2 1 3 3 2 3
## [75] 3 3 4 3 2 2 4 3 1 3 4 3 2 3 3 3 3 3 4 2 3 2 1 3 1 1 4 4 1 4 2 3 3 4 2 2 3
## [112] 4 3 3 3 4 3 3 3 4 4 3 4 1 2 3 2 3 2 2 3 3 3 3 3 2 2 2 4 1 2 3 4 1 4 3 1 3
## [149] 3 3 2 3 2 3 3 1 2 2 3 2 3 4 3 2 1 2 3 3 3 3 3 3 2 4 3 3 3 3 3 4 1 3 4 3 3
## [186] 3 4 3 3 2 3 1 4 4 4 2 3 3 4 3 3 2 3 2 3 3 3 3 3 4 3 2 3 3 2 3 3 3 3 2 3 3
## [223] 3 3 4 2 4 3 2 4 2 2 3 2 3 1 4 2 2 3 2 4 4 2 3 2 3 2 4 3 3 1 1 2 3 2 3 3 2
## [260] 1 3 3 2 1 2 3 4 3 3 4 3 4 1 4 2 2 2 4 3 4 4 2 1 2 3 3 3 3 3 4 4 4 2 3 3 2
## [297] 3 3 4 4 3 4 4 3 3 3 2 3 4 4 3 2 2 3 4 4 2 3 3 3 3 1 3 1 4 2 3 3 3 2 3 2 2
## [334] 4 3 3 2 3 3 2 4 2 3 3 4 2 2 4 3 4 1 4 3 2 3 2 3 3 2 3 3 2 2 3 2 3 3 3 3 3
## [371] 4 3 3 1 3 3 4 4 2 3 2 3 3 2 4 3 2 4 4 3 2 2 3 3 2 2 2 3 1 3 2 3 4 4 3 3 3
## [408] 1 3 3 2 2 2 3 3 3 2 3 4 4 2 3
```

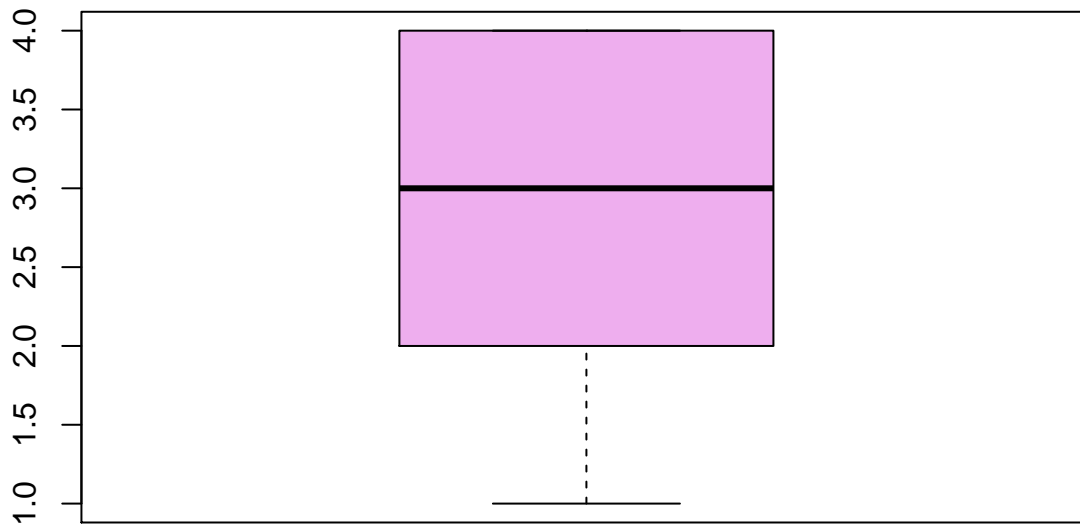
```
kmeans(Datos,4,algorithm = "Lloyd",iter.max = 40)
```

```
## K-means clustering with 4 clusters of sizes 90, 124, 52, 156
##
## Cluster means:
##           X1           X2           X3
## 1  699.65556 214.94444 179.67778
## 2  388.59677 105.15323 101.79032
## 3 1059.92308 247.76923 315.86538
## 4   97.36538  27.49359  25.76923
##
## Clustering vector:
##  [1] 4 1 4 4 3 3 2 2 4 2 2 2 2 4 4 2 4 3 2 2 4 2 4 2 4 2 3 1 4 4 2 1 4 2 3 2 1
## [38] 4 4 3 1 1 1 1 4 1 3 4 4 2 3 4 4 2 4 2 1 1 2 4 1 2 2 4 1 2 4 4 2 3 4 2 1 4
## [75] 4 2 1 4 2 1 1 4 3 2 1 4 1 2 2 4 4 2 1 2 4 1 1 4 3 3 1 1 3 1 1 4 4 3 2 1 4
## [112] 1 2 4 4 2 4 4 4 2 1 4 3 3 1 4 1 4 1 1 4 4 4 4 4 2 1 1 1 1 2 4 3 3 2 4 1 4
## [149] 4 1 1 4 4 2 2 1 2 4 4 1 4 1 2 2 3 2 4 4 4 4 4 2 1 1 4 4 4 4 4 3 3 4 2 4 4
## [186] 2 1 4 4 2 4 3 2 3 2 1 4 2 3 4 4 4 4 2 4 4 4 2 2 1 4 1 4 4 1 4 2 4 1 2 4 4
## [223] 4 4 1 2 3 4 4 1 2 3 4 2 2 3 1 2 1 2 2 3 1 2 4 2 4 1 3 4 2 3 3 2 4 2 4 2 2
## [260] 3 4 4 2 3 1 4 1 4 4 1 4 1 3 1 2 1 2 1 4 3 1 2 2 1 4 2 4 2 4 3 2 3 2 4 4 2
## [297] 4 4 1 3 4 1 1 4 2 4 1 2 1 3 4 1 1 4 2 2 2 2 1 4 2 3 4 3 1 2 4 4 2 2 4 2 1
## [334] 2 2 2 2 4 4 4 3 2 4 4 1 2 2 1 4 1 3 2 4 1 4 2 4 4 1 4 4 2 2 4 2 4 2 2 2 4
## [371] 1 4 2 3 2 4 2 2 2 2 2 1 2 2 1 4 1 2 3 4 2 1 2 4 3 2 3 4 3 4 4 4 3 2 4 4 4
## [408] 3 2 1 1 1 2 4 4 2 3 4 1 3 2 4
##
## Within cluster sum of squares by cluster:
## [1] 2811181 2250364 3555607 1164979
## (between_SS / total_SS =  84.0 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

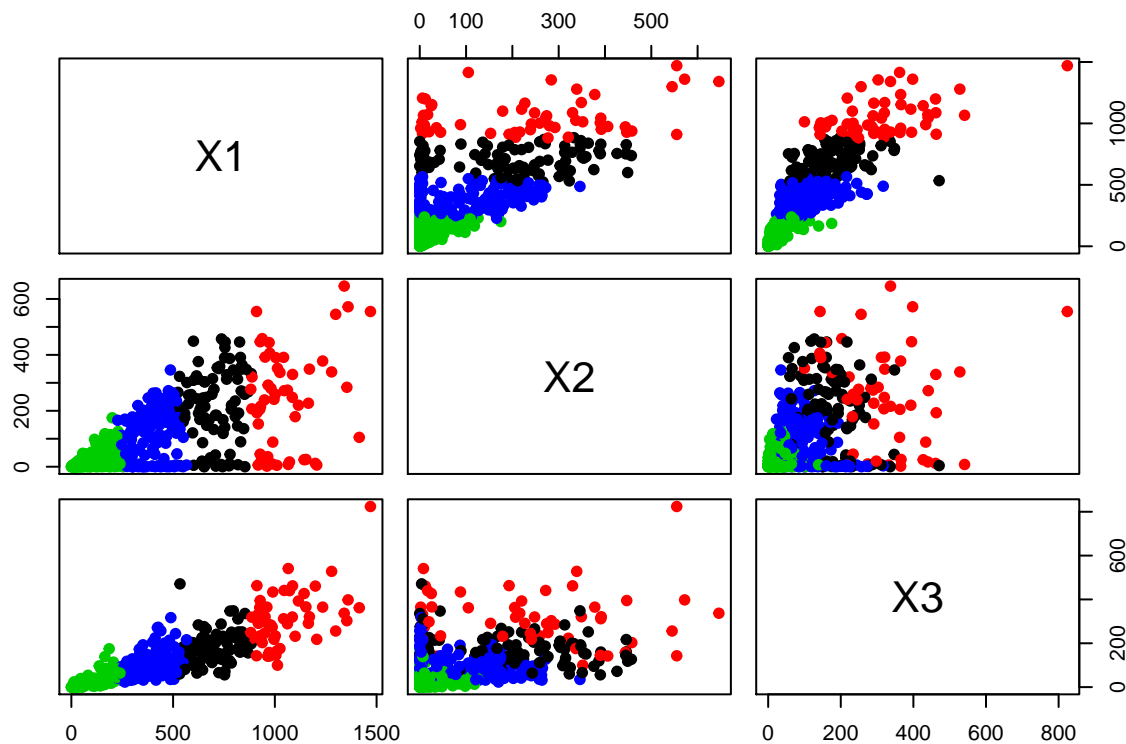
```
km3=kmeans(Datos,4,algorithm = "Lloyd",iter.max = 40)
km3$cluster
```

```
##  [1] 3 1 3 3 2 2 4 4 3 4 4 4 4 3 3 4 3 2 4 4 3 4 3 4 3 4 2 1 3 3 4 1 3 4 2 4 1
## [38] 3 3 2 1 1 1 1 3 1 2 3 3 4 2 3 3 4 3 4 1 1 4 3 1 4 4 3 1 4 3 3 4 2 3 4 1 3
## [75] 3 4 1 3 4 1 1 3 2 4 1 3 1 4 4 3 3 4 1 4 3 1 1 3 2 2 1 1 2 1 1 3 3 2 4 1 3
## [112] 1 4 3 3 4 3 3 3 4 1 3 2 2 1 3 1 3 1 1 3 3 3 3 3 4 1 1 1 1 4 3 2 2 4 3 1 3
## [149] 3 1 1 3 3 4 4 1 4 3 3 1 3 1 4 4 2 4 3 3 3 3 3 4 1 1 3 3 3 3 3 2 2 3 4 3 3
## [186] 4 1 3 3 4 3 2 4 2 4 1 3 4 2 3 3 3 3 4 3 3 3 4 1 3 1 3 3 1 3 4 3 1 4 3 3
## [223] 3 3 1 4 2 3 3 1 4 2 3 4 4 2 1 4 1 4 4 2 1 4 3 4 3 1 2 3 4 2 2 4 3 4 3 4 4
## [260] 2 3 3 4 2 1 3 1 3 3 1 3 1 2 1 4 1 4 1 3 2 1 4 4 1 3 4 3 4 3 2 4 2 4 3 3 4
## [297] 3 3 1 2 3 1 1 3 4 3 1 4 1 2 3 1 1 3 4 4 4 4 1 3 4 2 3 2 1 4 3 3 4 4 3 4 1
## [334] 4 4 4 4 3 3 3 2 4 3 3 1 4 4 1 3 1 2 4 3 1 3 4 3 3 1 3 3 4 4 3 4 3 4 4 4 3
## [371] 1 3 4 2 4 3 4 4 4 4 4 1 4 4 1 3 1 4 2 3 4 1 4 3 2 4 2 3 2 3 3 3 2 4 3 3 3
## [408] 2 4 1 1 1 4 3 3 4 2 3 1 2 4 3
```

```
boxplot(km3$cluster,col="plum2")
```



```
plot(Datos,col=km3$cluster,pch=19)
```



5 Conclusiones:

En el transcurso de la realización de este trabajo se han llevado a cabo varios estudios estadísticos muy interesantes, aprovechando que en el mundo NBA se pueden hacer gran cantidad de análisis y comparaciones. Aprovechando esto, hemos cotejado a los mejores jugadores de la liga, ver en qué aspecto destacaban cada uno de ellos, comprobando así que tanto los altos como los bajos pueden destacar, aunque de forma totalmente distinta. También hemos explorado los datos de los equipos de la liga, analizando cuáles de ellos podían ser serios candidatos al anillo, y por contra, cuáles eran los que necesitaban un cambio de dinámica. Hemos analizado estadísticas cruciales como la altura y las hemos relacionado con otras como el triple, los tiros libres

anotados, los robos, etc. También hemos hecho una parte de estadísticas de inferencia a los datos. Así que con los últimos temas dados en el temario, decidimos realizar varios test ANOVA, un test de Kolmogorov-Smirnov, una comparación de medias dos a dos con ajuste de Bonferroni y un Clustering.

Podemos decir que gracias a este proyecto hemos aprendido mucho sobre las herramientas de R, pues hemos podido averiguar de primera mano cómo se hacen los análisis estadísticos, aplicando las técnicas y conceptos dados en clase. Este trabajo nos ha ayudado a comprender mejor la asignatura en general.

En nuestra opinión, R es una herramienta muy útil con un gran abanico de posibilidades; hemos podido personalizar a nuestro antojo la gran cantidad de funcionalidades que se nos ofrecían ya que cuentan con un número muy extenso de variables con las cuales ir jugando. También hemos podido realizar fácilmente operaciones que a mano hubiesen sido de lo más tediosas. Por último, cabe mencionar que a pesar de las circunstancias, hemos podido interactuar más con nuestro profesor, el cuál ha supuesto una gran ayuda a la hora de realizar el proyecto.