

TAREA MODULO 3.

CASO DE ESTUDIO : ANÁLISIS DE FACTORES DE ATAQUE ENTRE LOS CENTRO CAMPISTAS CON MÁS DE 1000 ´EN EL CAMPEONATO DE LIGA ESPAÑOLA.

1. LECTURA Y FILTRADO.

Filtramos el data set de **FBREF_players.csv**. Para obtener los datos exclusivamente de los centrocampistas que han jugado más de 1000 minutos en La Liga.

Player	Squad	Pos	Age	MP	Min	Gls	G.PK	Ast	xG	xA	Gls.90				
1	Martin Agirregabiria				Alavés	DF,MF	24	23	1374	0	0	1	0.2	0.9	0.00
2	Rubén Alcaraz				Valladolid	MF	30	27	2020	1	1	0	1.5	0.6	0.04
3	Carles Aleñá				Getafe	MF,FW	23	18	1111	1	1	2	1.7	1.3	0.08
4	Sergio Álvarez				Eibar	MF,DF	29	26	1539	0	0	0	0.3	0.9	0.00
5	Mauro Arambarri				Getafe	MF	25	32	2803	3	3	0	2.2	2.1	0.10
6	Marco Asensio				Real Madrid	FW,MF	25	31	1728	4	4	2	4.6	3.0	0.21

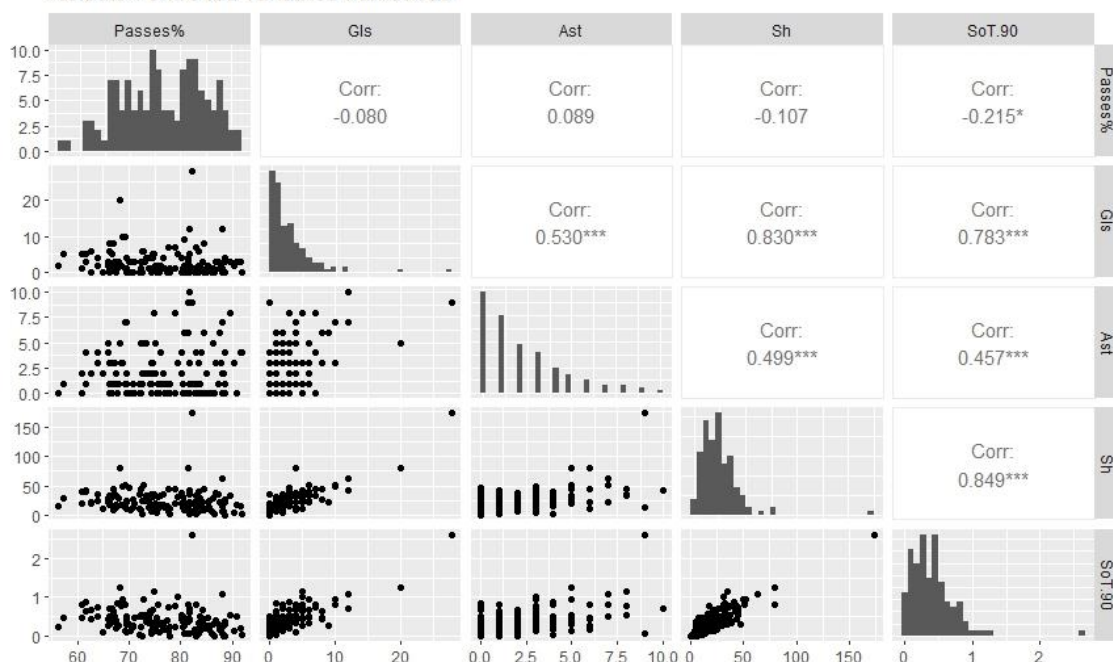
Seleccionamos solo las columnas que nos interesan : `metrics <- c("Player", "Squad", "MP", "Min", "Passes.", "Gls", "Ast", "Sh", "Sh.90", "SoT.90", "PassesCompleted.90", "LongPasses.", "LongPassesCompleted.90", "ShortPasses.", "MediumPasses.", "LongPasses.", "PassesProgressive.90", "PassesAttempted.90", "ShortPassesCompleted.90", "MediumPassesCompleted.90", "TotDistPasses.90", "FinalThirdPasses.90")`

Hay 301 jugadores con más de 1000 ´.

Y centro campistas exactamente : 135

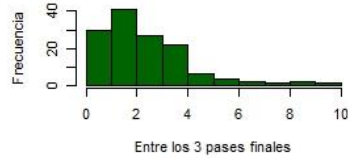
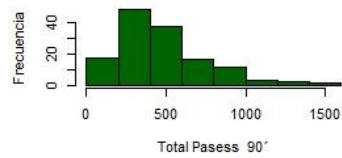
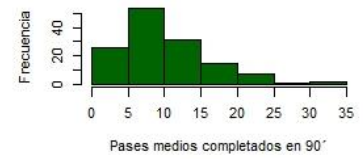
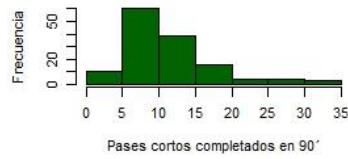
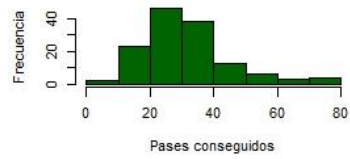
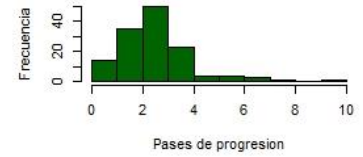
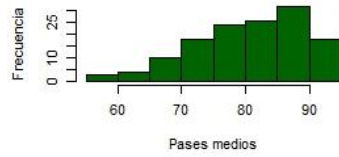
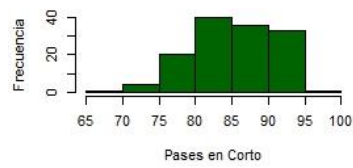
2. Análisis Exploratorio de datos del data set.

Relación entre las variables numéricas

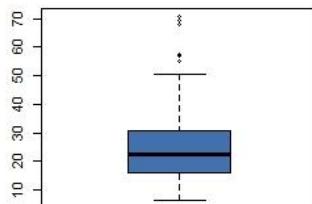


Vemos que no hay nulos .

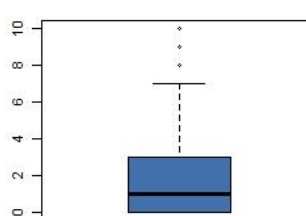
Histograma de las variables de analisis



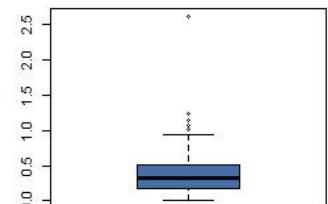
Pases completados por 90'



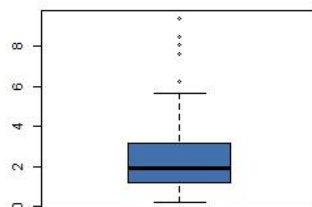
Identificación de outliers



Tiros a portería

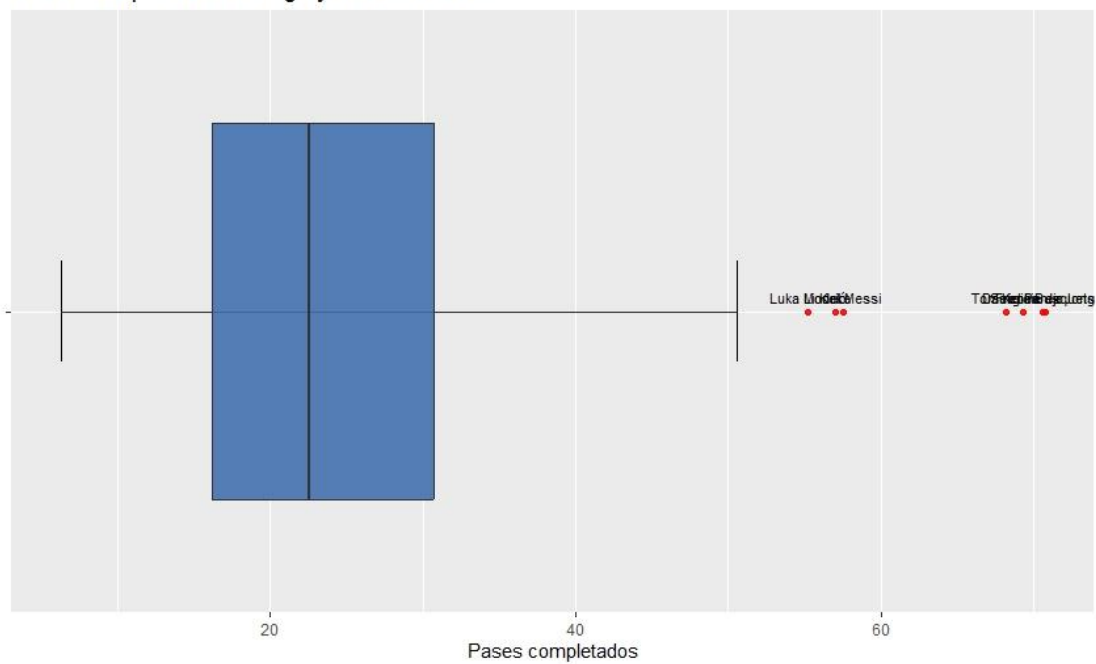


Últimos 3 pases

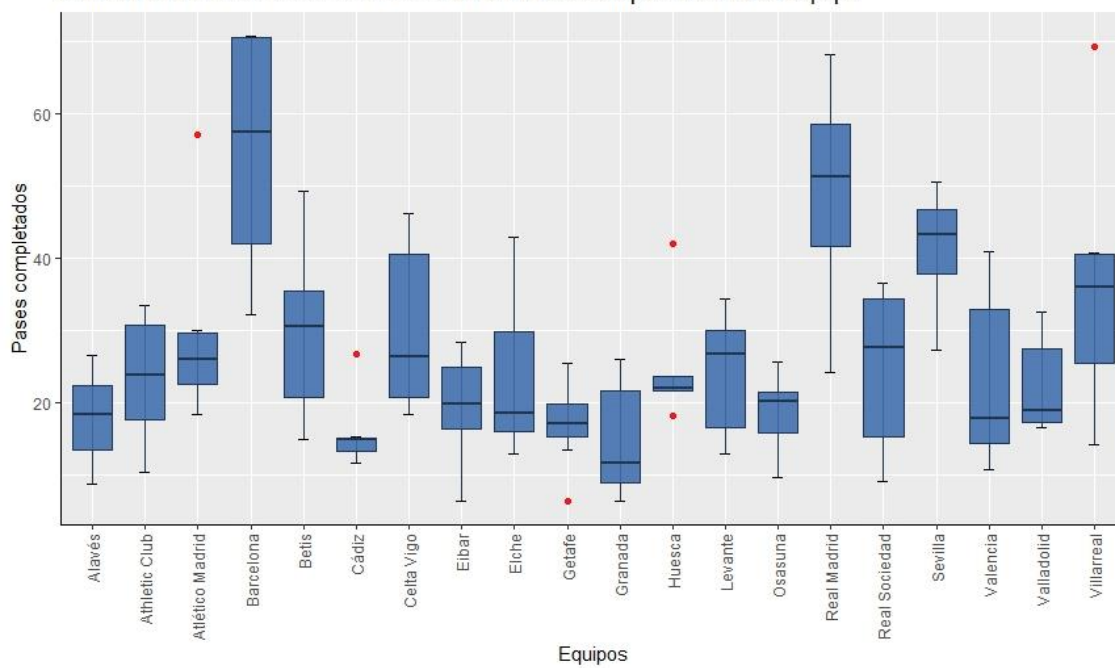


Analizamos la variable Pases completados en 90'

Centro campistas de La Liga y sus asistencias



Análisis de PASES COMPLETADOS de los centrocampistas de cada equipo



Describimos algunas variables numéricas:

Passes%	Gls	Ast	Sh	Sh.90
Min. :56.40	Min. : 0.000	Min. : 0.000	Min. : 1.00	Min. :0.070
1st Qu.:69.60	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 13.00	1st Qu.:0.665
Median :76.20	Median : 2.000	Median : 1.000	Median : 22.00	Median :1.100
Mean :76.46	Mean : 2.741	Mean : 2.178	Mean : 24.19	Mean :1.199
3rd Qu.:82.90	3rd Qu.: 4.000	3rd Qu.: 3.000	3rd Qu.: 32.00	3rd Qu.:1.620
Max. :91.80	Max. :28.000	Max. :10.000	Max. :173.00	Max. :5.660

ESTANDARIZAMOS LAS VARIABLES:

```
[1] -4.244258e-16 -4.231480e-17 1.142291e-16 -1.012115e-16 -4.051864e-17 1.427613e-17 -9.405564e-17 -6.152807e-17
```

```
[9] 7.214965e-17 2.597524e-16 2.173568e-16 -5.044100e-17 1.376284e-17 -1.835755e-17 -6.410446e-17 -5.383957e-17
```

```
[17] -9.057646e-17 8.330528e-17
```

```
> desv.estandar # vector igual a 1
```

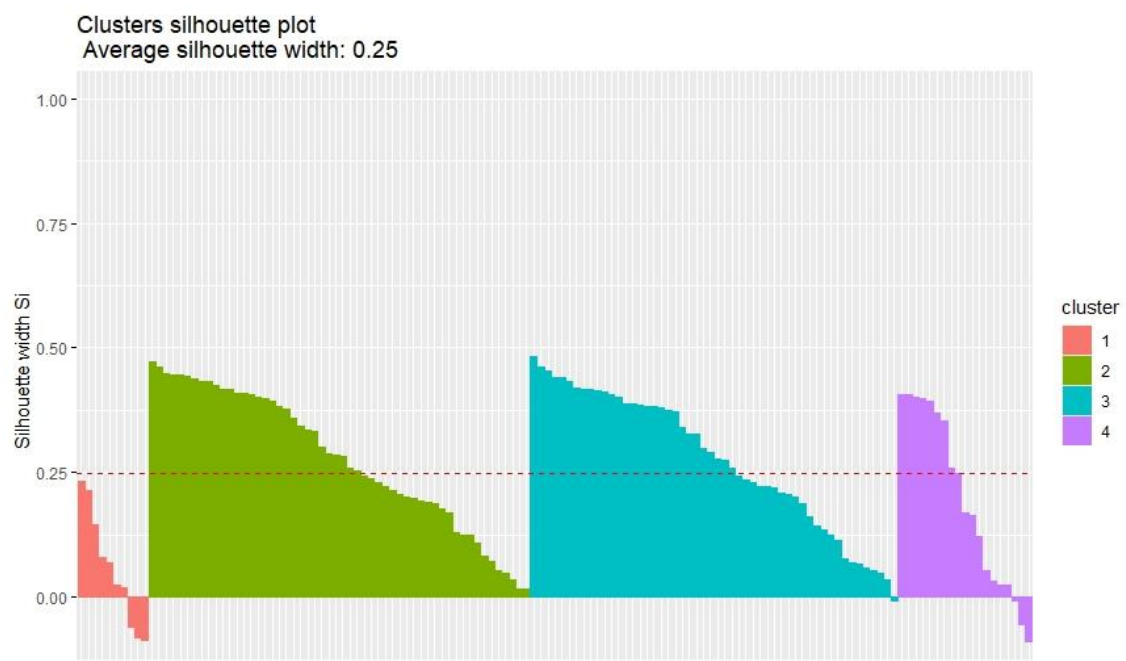
```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

VEMOS VALORES ATÍPICOS:

```
$hopkins_stat
```

```
[1] 0.7877879
```

PROBAMOS LA TENDENCIA DE AGRUPACIÓN:



3. CLUSTERIZACIÓN.

```
dist_eucl <- dist(df_teams_norm, method="euclidean")
```

4.1 Cálculo de conglomerados

```
hc_ward_eucl <- hclust(dist_eucl, method = "ward.D2")
```

```
hc_single_eucl <- hclust(dist_eucl, method = "single")
```

```
hc_complete_eucl <- hclust(dist_eucl, method = "complete")
```

```
hc_average_eucl <- hclust(dist_eucl, method = "average")
```

4.2 Evaluación del método de clusterización

```
ew <- cor(dist_eucl, cophenetic(hc_ward_eucl))
```

```
es <- cor(dist_eucl, cophenetic(hc_single_eucl))
```

```
ec <- cor(dist_eucl, cophenetic(hc_complete_eucl))
```

```
ea <- cor(dist_eucl, cophenetic(hc_average_eucl))
```

```
resumen <- rbind(ew, es, ec, ea)
```

```
row.names(resumen) <- c("Ward Euclidean", "Single Euclidean",  
                        "Complete Euclidean", "Average Euclidean")
```

```
kable(resumen, align='c',
```

```
      col.names = c("Distancias cophenetic"))
```

```
      | Distancias cophenetic |  
|:-----|:-----:|  
|Ward Euclidean | 0.4682520 |  
|Single Euclidean | 0.7113363 |  
|Complete Euclidean | 0.5875913 |  
|Average Euclidean | 0.8331728 |
```

¿Cuáles son las primeras uniones? ¿A qué distancias?

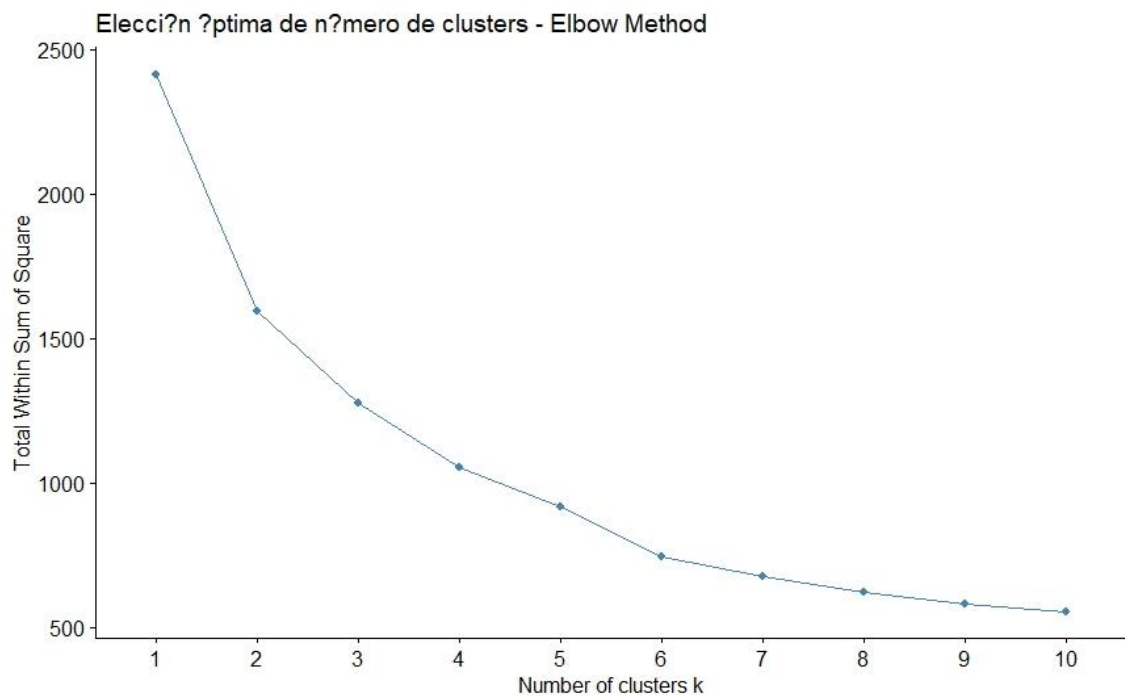
```
[,1][,2] [,3]  
[1,] -56 -66 0.7364932
```

[2,] -108 -120 0.9965799
[3,] -68 -86 0.9980248
[4,] -18 -43 1.1292399
[5,] -106 -135 1.1327232
[6,] -10 -46 1.1360244
[7,] -13 -88 1.1899602
[8,] -67 4 1.1905213
[9,] -118 1 1.2300554
[10,] -92 -119 1.2471191
[11,] -7 -112 1.2523886
[12,] -34 3 1.2730950
[13,] -115 -125 1.2811622
[14,] -49 7 1.3173255
[15,] -27 -70 1.3319337
[16,] -130 2 1.3452528
[17,] -37 -132 1.4084042
[18,] -55 -97 1.4257378
[19,] -81 -133 1.4285655
[20,] -41 -100 1.4345784
[21,] -4 9 1.4642636
[22,] -69 16 1.4844800
[23,] -21 -116 1.4873683
[24,] -9 -78 1.4894198
[25,] -1 -98 1.5189330
[26,] -61 -90 1.5289429
[27,] -25 -127 1.5290942
[28,] -2 8 1.5320807
[29,] -99 -123 1.5869794
[30,] -20 -23 1.5999430
[31,] -35 -104 1.6173478
[32,] -45 -111 1.6369255
[33,] -75 -77 1.6392931
[34,] -96 -131 1.6605099
[35,] -110 24 1.6646982
[36,] -83 -89 1.6676974
[37,] 15 21 1.6729125
[38,] -8 35 1.7141719
[39,] 6 28 1.7156059

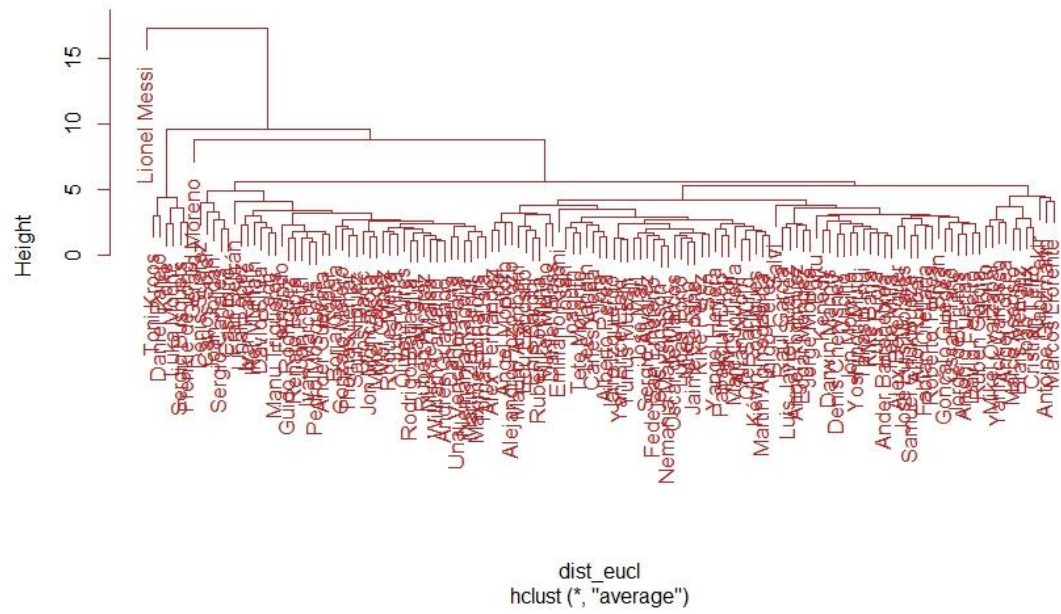
[40,] 22 29 1.7365843
[41,] -51 -85 1.7370723
[42,] -32 -39 1.7440807
[43,] -105 19 1.7543850
[44,] -113 25 1.7985139
[45,] 12 32 1.8119715
[46,] -40 33 1.8192224
[47,] 13 45 1.8467639
[48,] 5 20 1.8471020
[49,] -3 18 1.8494750
[50,] -58 -65 1.8689216
[51,] 10 44 1.8838757
[52,] -26 11 1.8844727
[53,] -73 -80 1.9402001
[54,] -30 -33 1.9446542
[55,] -128 39 1.9599512
[56,] -16 -19 1.9648006
[57,] -47 27 1.9736291
[58,] 37 40 1.9801947
[59,] -44 -101 1.9909274
[60,] -71 26 2.0295225
[61,] -48 38 2.0395964
[62,] -6 -72 2.0468933
[63,] -38 -50 2.0595694
[64,] 46 55 2.0633684
[65,] -121 34 2.1053579
[66,] -29 -126 2.1160421
[67,] -54 41 2.1172196
[68,] 17 48 2.1528947
[69,] -36 60 2.1544006
[70,] -12 -64 2.1751449
[71,] 36 61 2.1981411
[72,] -93 -129 2.1991828
[73,] -5 52 2.2218591
[74,] -24 51 2.2258701
[75,] -91 -94 2.2388706
[76,] -57 49 2.2912315
[77,] -114 53 2.3097654

[78,] -31 47 2.3180247
[79,] -63 -107 2.3249530
[80,] 23 72 2.3400385
[81,] -59 -79 2.3546277
[82,] 30 71 2.3615131
[83,] 64 68 2.3710199
[84,] -74 -134 2.4070474
[85,] 14 58 2.4172374
[86,] 63 77 2.4334830
[87,] 57 74 2.4349748
[88,] 31 42 2.4409268
[89,] -15 -52 2.4647266
[90,] -82 -103 2.4695062
[91,] 54 90 2.4863964
[92,] -87 65 2.4911451
[93,] 67 76 2.5082635
[94,] -17 -22 2.5129077
[95,] 59 88 2.5364169
[96,] 43 83 2.5734201
[97,] 50 82 2.5736683
[98,] 75 94 2.6851266
[99,] 69 96 2.6869299
[100,] 85 87 2.7011909
[101,] -53 56 2.7145896
[102,] 70 80 2.7704486
[103,] 62 66 2.8178895
[104,] 93 100 2.9233157
[105,] -14 95 2.9276139
[106,] -102 97 2.9700497
[107,] 79 92 3.0315239
[108,] 102 105 3.0369869
[109,] -60 -95 3.0418427
[110,] 106 108 3.0768138
[111,] 73 84 3.1042066
[112,] 91 111 3.2119735
[113,] 78 99 3.2335325
[114,] -122 101 3.2779894
[115,] 107 113 3.4082672

[116,] -109 104 3.5100809
 [117,] 86 110 3.5994160
 [118,] 81 89 3.6201294
 [119,] 98 103 3.6834557
 [120,] 112 116 3.7870426
 [121,] -117 117 3.8447314
 [122,] -124 114 3.9827803
 [123,] -11 115 4.0866646
 [124,] -42 -62 4.1119315
 [125,] 120 121 4.2218871
 [126,] 109 118 4.3954947
 [127,] -28 124 4.4021291
 [128,] 119 127 4.5355406
 [129,] 122 123 4.9425674
 [130,] 125 128 5.3226842
 [131,] 129 130 5.5842979
 [132,] -84 131 8.7892072
 [133,] 126 132 9.5744855
 [134,] -76 133 17.3359114



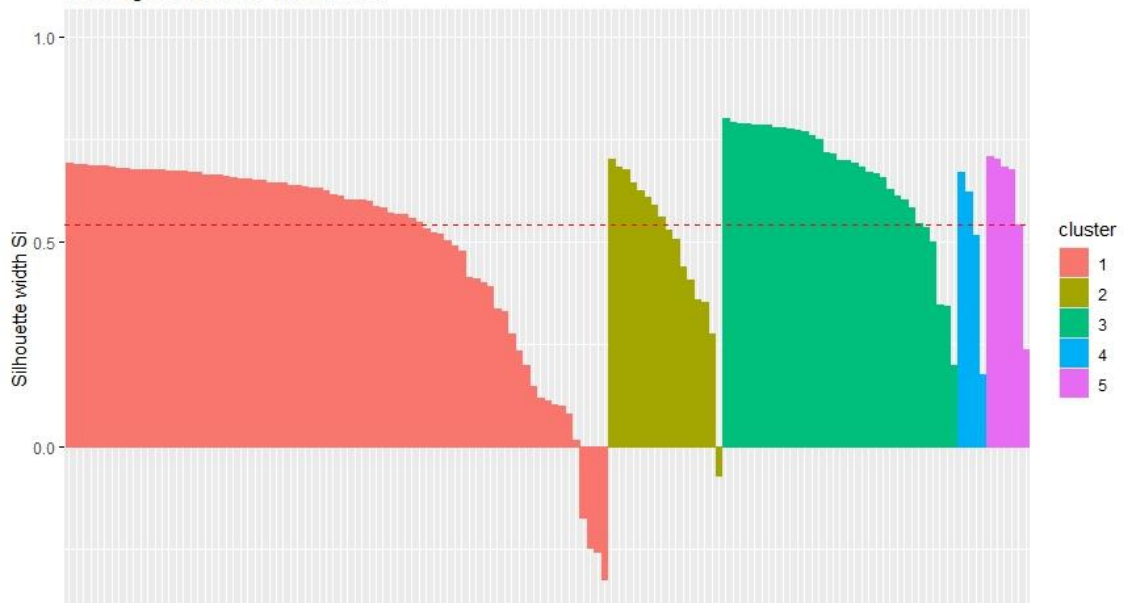
Dendograma Centrocampistas



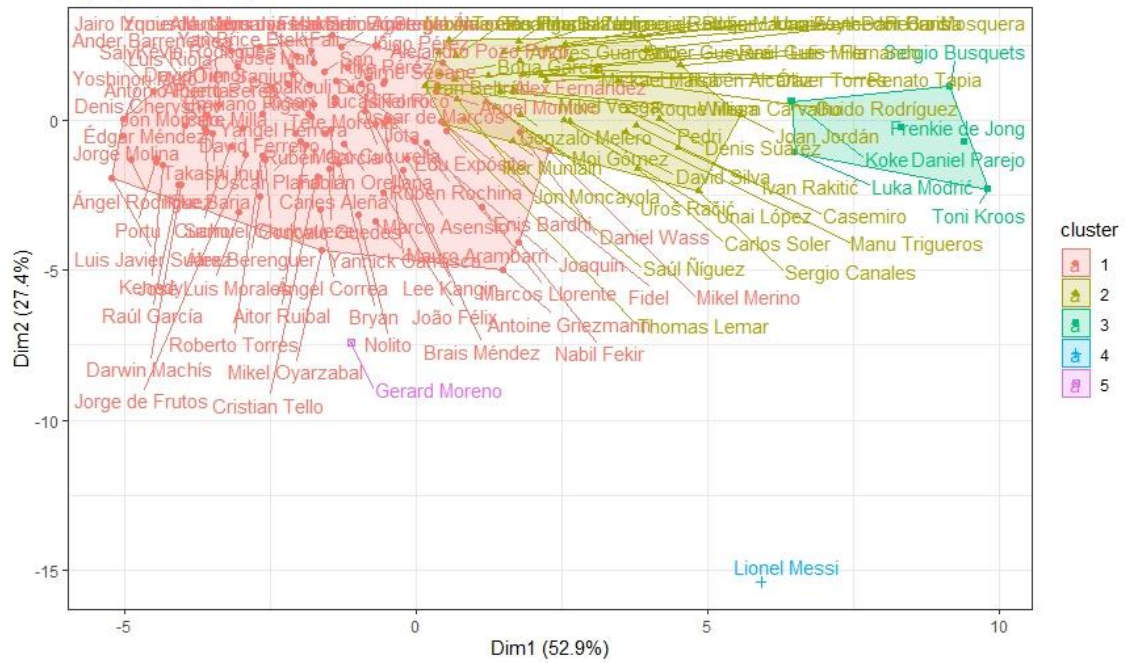
Clusters silhouette plot
Average silhouette width: 0.58



Clusters silhouette plot
Average silhouette width: 0.54



Agrupaci?n jer?quica - 5 clusters



Cluster Dendrogram

