

Ejercicio 2 De Métodos De Predicción

(“Mario García Berenguer”, “David Hernando González”, “Eder Tarifa Fernández”)

2023-11-28

Antes de comenzar con el ejercicio vamos a instalar todas las librerías que vamos a usar para su realización.

```
library(FactoMineR)
library(factoextra)
library(FactoClass)
library(RColorBrewer)
library("corrplot")

library(readxl)
```

Una vez tenemos las librerías instaladas, procedemos a leer los datos y a echarles un primer vistazo.

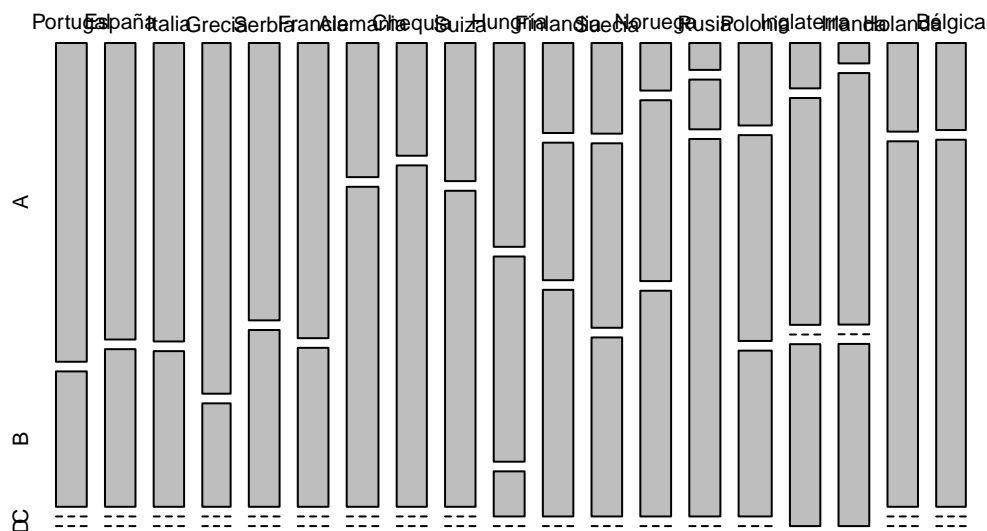
```
cochesExcel <- read_excel("coches.xls", col_names = TRUE)
tablaCoches <- data.frame(A = cochesExcel$A, B = cochesExcel$B, C = cochesExcel$C, D = cochesExcel$D, r
#View(tablaCoches)
summary(tablaCoches)
```

##	A	B	C	D
## Min.	: 45.0	Min. :111.0	Min. : 0.0	Min. : 0.00
## 1st Qu.:	189.5	1st Qu.:342.5	1st Qu.: 0.0	1st Qu.: 0.00
## Median :	244.0	Median :405.0	Median : 0.0	Median : 0.00
## Mean :	345.4	Mean :467.6	Mean :143.3	Mean : 42.05
## 3rd Qu.:	630.5	3rd Qu.:623.0	3rd Qu.:246.5	3rd Qu.: 0.00
## Max.	:712.0	Max. :800.0	Max. :842.0	Max. :402.00

Si queremos visualizar la tabla de contingencia podemos hacer lo siguiente:

```
mosaicplot(tablaCoches, main="", space=c(40, 40))
```

```
## Warning: In mosaicplot.default(tablaCoches, main = "", space = c(40, 40)) :
## extra argument 'space' will be disregarded
```



Para comenzar el análisis por correspondencias, debemos primero ver si existe alguna relación o dependencia entre las variables. Para ello, realizaremos el test de la chi-cuadrado.

```
chisq <- chisq.test(tablaCoches)
chisq
```

```
##
## Pearson's Chi-squared test
##
## data:  tablaCoches
## X-squared = 19222, df = 54, p-value < 2.2e-16
```

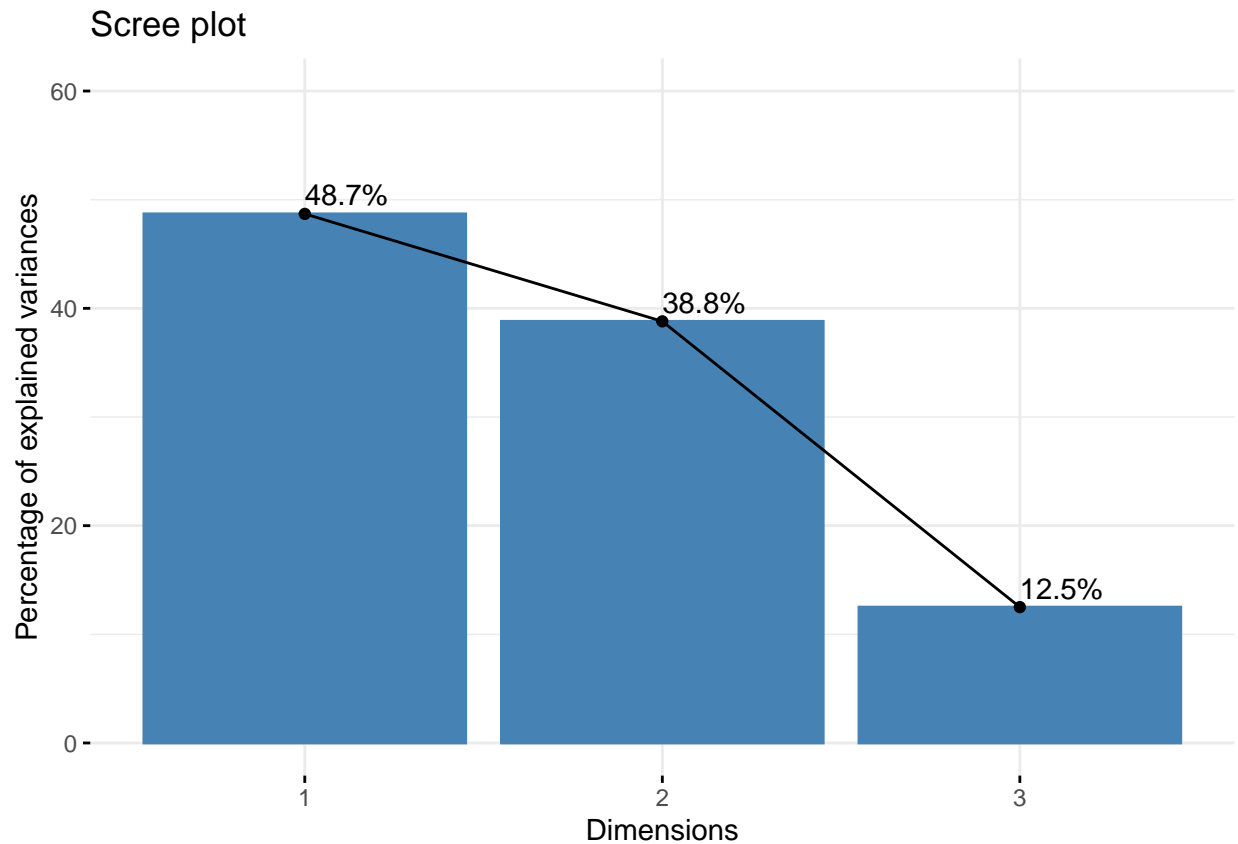
Vemos claramente que el p-valor que obtenemos al realizar el test es minúsculo, mucho menor que 0.05, por lo que podemos descartar que las variables sean independientes.

Por tanto, podemos continuar con nuestro análisis por correspondencia.

```
analisisCoches <- CA(tablaCoches, graph = FALSE)
analisisCoches$eig
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.4934210           48.69323           48.69323
## dim 2  0.3931978           38.80271           87.49595
## dim 3  0.1267068           12.50405          100.00000
```

```
fviz_screepplot(analisisCoches, addlabels = TRUE, ylim = c(0, 60))
```



Vemos que el 87% de la variabilidad es explicada por las dos primeras componentes, por lo que vamos a poder representar los datos en 2 dimensiones sin problemas.

```
analisisCoches$row
```

```
## $coord
##           Dim 1      Dim 2      Dim 3
## Portugal -0.33442575 -0.5755171  0.41304571
## España   -0.33053279 -0.5314122  0.31257499
## Italia    -0.33087780 -0.5353210  0.32147918
## Grecia    -0.34001679 -0.6388601  0.55734064
## Serbia    -0.32718332 -0.4934649  0.22613119
## Francia   -0.33030088 -0.5287849  0.30659004
## Alemania  -0.30211253 -0.2094285 -0.42090225
## Chequia   -0.29836500 -0.1669713 -0.51761942
## Suiza      -0.30280899 -0.2173190 -0.40292772
## Hungría    -0.05627043 -0.2874395  0.03376016
## Finlandia  1.00224152  0.1806389  0.08218236
## Suecia     0.72977827  0.1158189 -0.06246899
## Noruega    1.00449967  0.2635415 -0.11221364
## Rusia      1.87650583  0.5071992  0.26476145
## Polonia    0.65577184  0.1143809 -0.14015473
## Inglaterra -0.92262508  1.5641429  0.27173597
## Irlanda    -0.91919441  1.6160233  0.16054860
```

```
## Holanda -0.29414837 -0.1191996 -0.62644321
## Bélgica -0.29387326 -0.1160827 -0.63354345
```

```
##
```

```
## $contrib
```

```
##          Dim 1      Dim 2      Dim 3
## Portugal  1.19850071  4.4541196  7.11955569
## España    1.16025488  3.7635172  4.04064378
## Italia     1.16033892  3.8114014  4.26553118
## Grecia     1.13885802  5.0452986 11.91593382
## Serbia     1.14600885  3.2713326  2.13178979
## Francia    1.16562112  3.7488891  3.91085586
## Alemania   1.01026465  0.6092222  7.63621201
## Chequia    0.93494745  0.3674360 10.95797691
## Suiza      0.97280259  0.6287660  6.70747818
## Hungría    0.03406652  1.1154931  0.04775229
## Finlandia 10.66765836  0.4348639  0.27931828
## Suecia     5.67302830  0.1793074  0.16187503
## Noruega    10.80202693  0.9330613  0.52494615
## Rusia      38.11072524  3.4939049  2.95443662
## Polonia    4.95752486  0.1892662  0.88184594
## Inglaterra 9.01284905 32.5065572  3.04455369
## Irlanda    9.04524643 35.0838629  1.07457443
## Holanda    0.91887665  0.1893565 16.22951439
## Bélgica    0.89040047  0.1743439 16.11520595
```

```
##
```

```
## $cos2
```

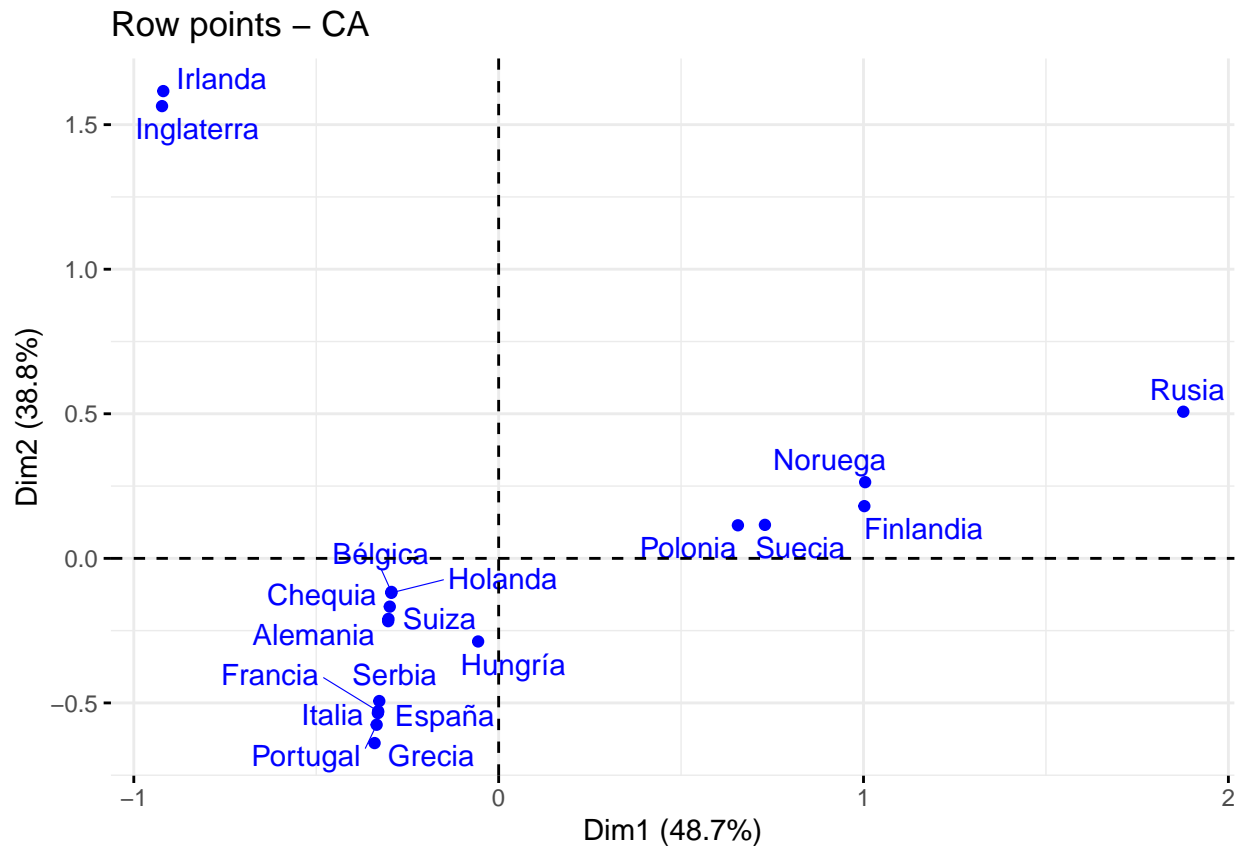
```
##          Dim 1      Dim 2      Dim 3
## Portugal  0.18224954  0.53973865  0.278011808
## España    0.22325744  0.57708521  0.199657351
## Italia     0.21922437  0.57382856  0.206947073
## Grecia     0.13855930  0.48915497  0.372285725
## Serbia     0.26649514  0.60620501  0.127299853
## Francia    0.22601304  0.57925819  0.194728764
## Alemania   0.29226583  0.14044687  0.567287299
## Chequia    0.23132669  0.07244590  0.696227415
## Suiza      0.30435426  0.15676070  0.538885042
## Hungría    0.03642527  0.95046326  0.013111470
## Finlandia  0.96227092  0.03125900  0.006470082
## Suecia     0.96850952  0.02439387  0.007096611
## Noruega    0.92480194  0.06365714  0.011540920
## Rusia      0.91494370  0.06684235  0.018213946
## Polonia    0.92928056  0.02827147  0.042447965
## Inglaterra 0.25247118  0.72562825  0.021900576
## Irlanda    0.24263746  0.74996042  0.007402120
## Holanda    0.17544561  0.02881105  0.795743342
## Bélgica    0.17230463  0.02688511  0.800810262
```

```
##
```

```
## $inertia
```

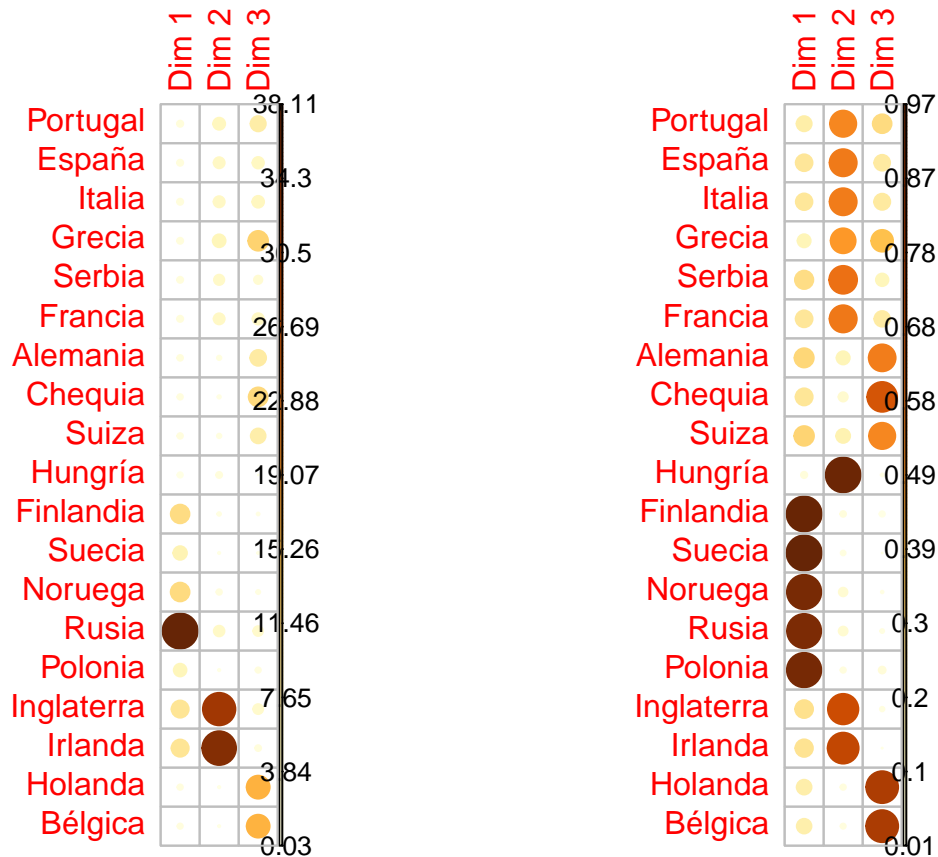
```
## [1] 0.032448115 0.025642779 0.026116421 0.040555665 0.021218579 0.025447290
## [7] 0.017055904 0.019942477 0.015771136 0.004614692 0.054700257 0.028902052
## [13] 0.057633389 0.205527752 0.026323018 0.176144027 0.183941696 0.025842370
## [19] 0.025497997
```

```
fviz_ca_row(analisisCoches, repel = TRUE)
```



En este gráfico podemos ver la variable País representada en dos dimensiones, con un 87.5% de variabilidad explicada. Vemos que los Países más mediterráneos y más occidentales de europa parecen estar muy pegados, y tener comportamientos muy similares. Por otro lado tenemos los países nórdicos, como polonia o suecia, y a rusia, aunque está bastante alejado. Por último, como era de esperar, están Inglaterra y Suecia, donde se conduce por la izquierda.

```
par(mfrow = c(1,2))
corrplot(analisisCoches$row$contrib, is.corr=FALSE,main="")
corrplot(analisisCoches$row$cos2, is.corr=FALSE,main="")
```



Cuando nos fijamos en la contribución absoluta, nos damos cuenta de que Rusia es quien más peso tiene en la primera dimensión con diferencia. En la segunda dimensión este papel parecen tomarlo los países angloparlantes Inglaterra y Irlanda. En la tercera dimensión, aunque con una variabilidad bastante menor encontramos a países como Holanda, Bélgica o Grecia. Este comentario lógicamente coincide con el gráfico mostrado, ya que Rusia es el país más alejado en el eje X (1ª dimensión), y Inglaterra y Irlanda son los países más alejados en el eje Y (2ª dimensión).

Si hablamos de la contribución relativa, debemos de mencionar que esta mide el peso que tiene cada dimensión en la explicación de los países. Podemos ver que los países nórdicos están explicados prácticamente en la primera dimensión, mientras que los países mediterráneos como España e Italia, los países angloparlantes, y Portugal se explican mejor en la segunda dimensión. La tercera dimensión ya no la podemos mostrar, pero sería la que explicaría a países más del interior de Europa, como Alemania o Suiza.

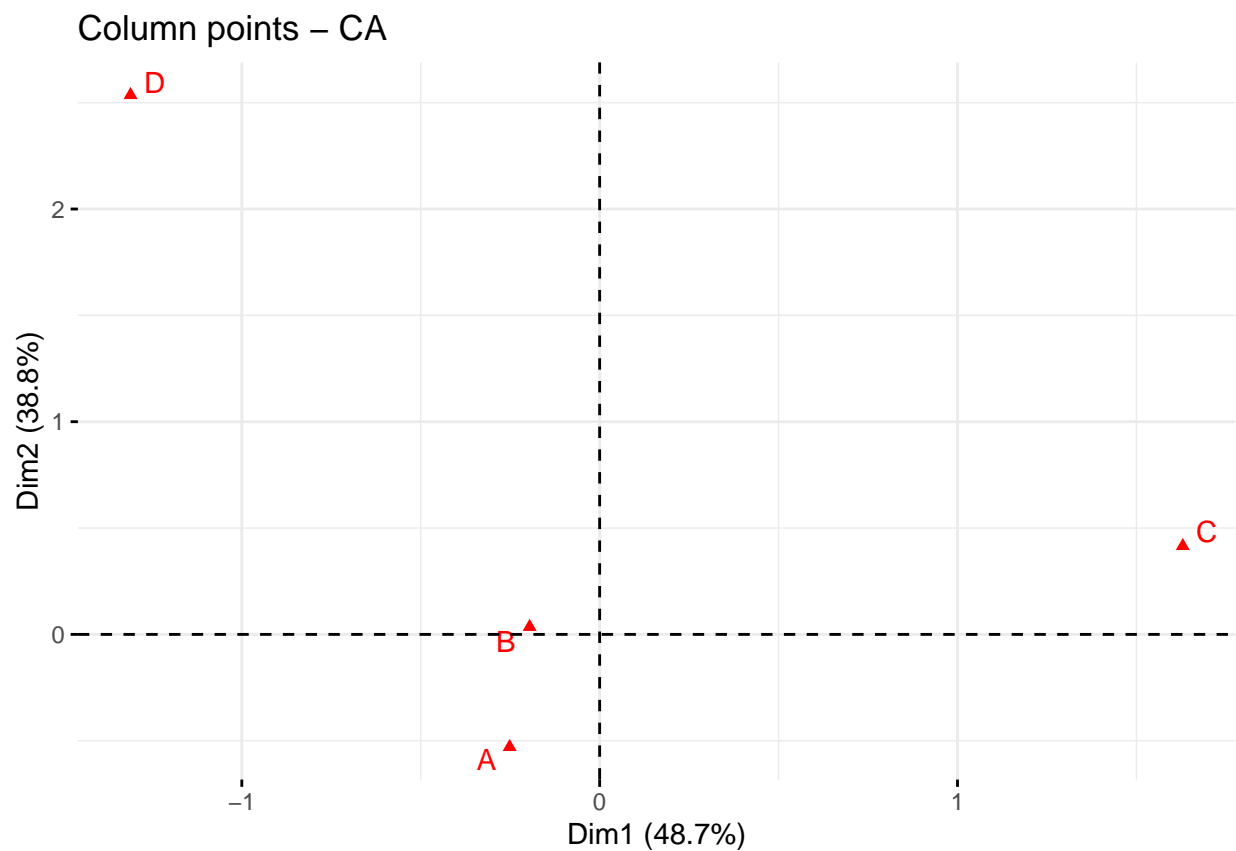
Una vez hemos acabado de analizar los perfiles de las filas, vamos a realizar un análisis de los perfiles de las columnas.

```
analisisCoches$col
```

```
## $coord
##      Dim 1      Dim 2      Dim 3
## A -0.2515582 -0.52921494  0.3647076
## B -0.1957242  0.03546436 -0.3655043
## C  1.6295559  0.41566825  0.1355795
## D -1.3110015  2.53605462  0.6062339
##
## $contrib
##      Dim 1      Dim 2      Dim 3
```

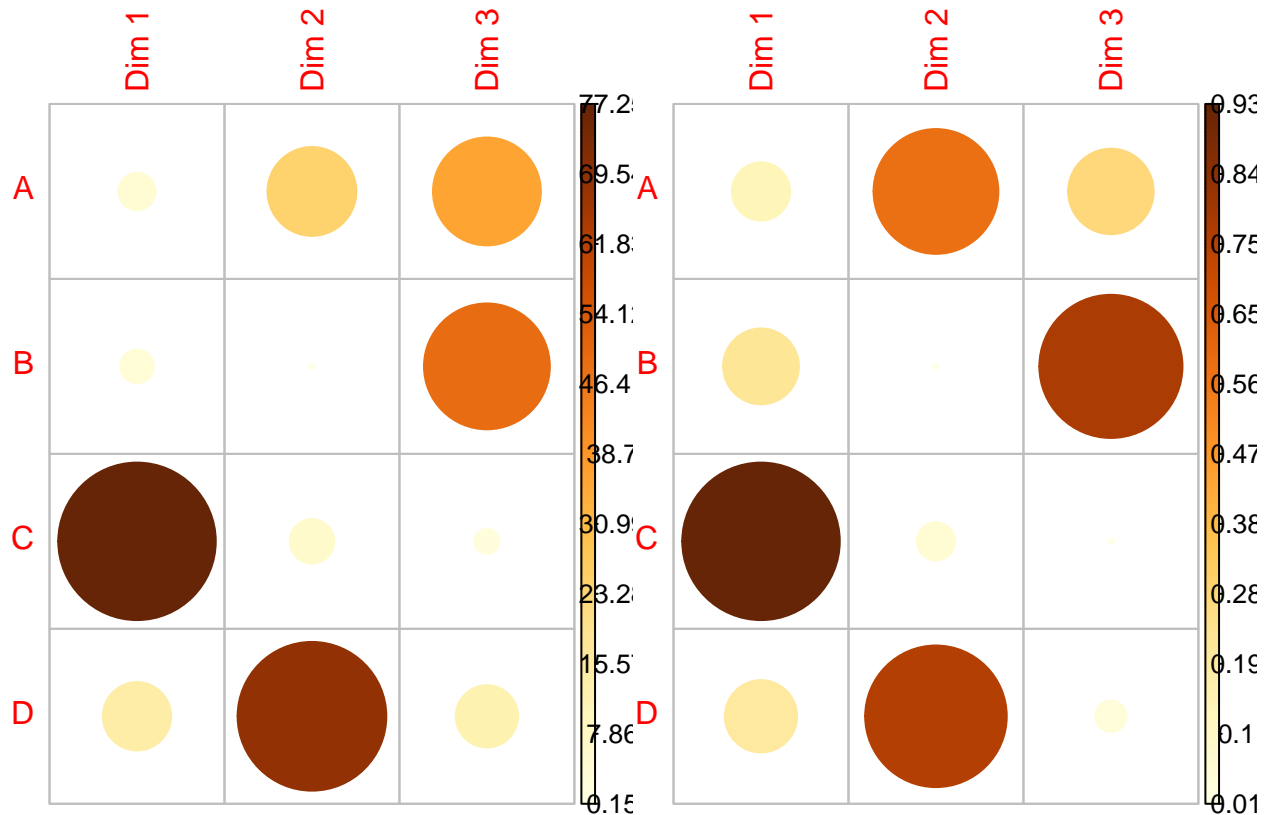
```
## A  4.437286 24.6439903 36.320168
## B  3.636097  0.1498087 49.379785
## C 77.254556  6.3079143  2.082529
## D 14.672060 68.8982867 12.217518
##
## $cos2
##      Dim 1      Dim 2      Dim 3
## A 0.1328435 0.587932389 0.279224105
## B 0.2212299 0.007263384 0.771506718
## C 0.9328459 0.060696699 0.006457421
## D 0.2017800 0.755072881 0.043147152
##
## $inertia
## [1] 0.16481425 0.08109785 0.40863149 0.35878202
```

```
fviz_ca_col(analisisCoches, repel=TRUE)
```



Vemos que las marcas D y C están muy alejadas del resto, mientras que A y B si que parecen más similares en dos dimensiones, ya que se encuentran más cerca. Veremos que nos dice la contribución absoluta y la contribución relativa.

```
par(mfrow = c(1,2))
corrplot(analisisCoches$col$contrib, is.corr=FALSE,main="")
corrplot(analisisCoches$col$cos2, is.corr=FALSE,main="")
```

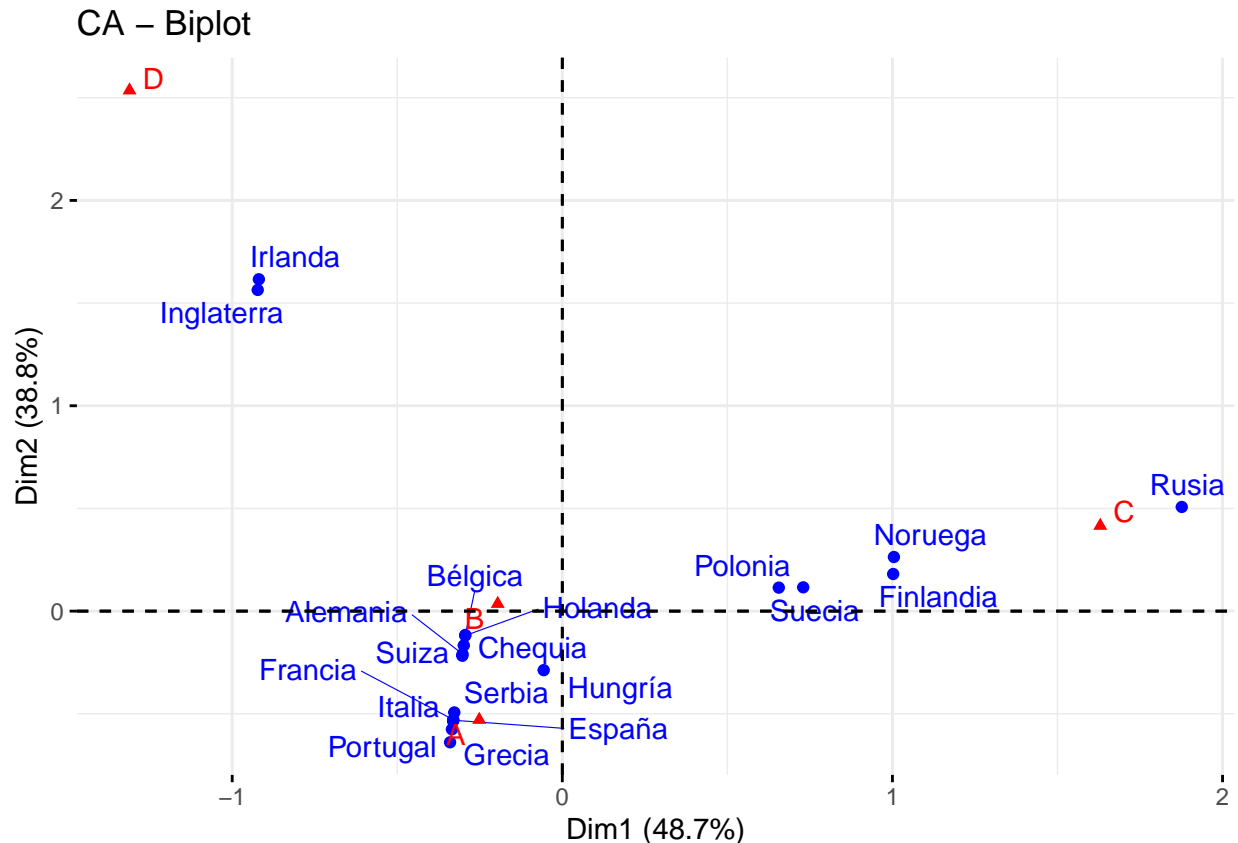


Cuando analizamos las contribuciones absolutas nos damos cuenta de que C es la marca de coche que más peso tiene en la primera dimensión (eje X). En la segunda dimensión debemos destacar la marca de coches D, aunque también la marca A tiene un poco de peso. La tercera dimensión ya no es representable, pero se la reparten mayoritariamente entre las marcas A y B.

En cuanto a las contribuciones relativas, C parece explicable casi al completo por la primera dimensión (eje X). En la segunda dimensión vemos explicadas las marcas A y D, aunque realmente A lo está en un 59%, ya que también se explica en la primera (13%) y tercera (28%) dimensión. La marca B esta explicada en la tercera dimensión en un 77%.

Después de analizar tanto los países como las marcas de coches, podemos intentar sacar conclusiones más generales.

```
fviz_ca_biplot(analisisCoches, repel = TRUE)
```

Cuando mostramos el gráfico del análisis de componentes completo, nos damos cuenta de que marcas están más relacionadas con cada país.

Por un lado, vemos que la marca de coches A está bastante cerca de los países occidentales de Europa, pero sobretodo parece acercarse a los mediterráneos, como España, Portugal, Grecia, Italia, ... Podemos suponer que esta marca de coches ofrece un buen servicio en cuanto a coches descapotables o con un buen sistema de aire acondicionado, ya que estas zonas son las más cálidas de Europa.

En cuanto a la marca de coches B, esta también está muy cercana a los países occidentales europeos, aunque parece ser más del estilo de países como Alemania, Holanda, Suiza, ... Esta claro que esta marca de coches, aunque es similar a la marca A, no comparte todas sus características, y parece enfocarse más a ofrecer un buen coche en zonas más lluviosas y frías.

La marca de coches C se diferencia bastante del resto, posicionandose en uno de los extremos del eje X. Los países más cercanos a esta marca son los países escandinavos, junto con Rusia y Polonia. La principal hipótesis por la que creemos que la marca C se diferencia tanto de las demás es que se trata de una marca histórica desarrollada por la industria de la URSS, ya que parece muy popular en países donde este Estado tuvo gran influencia, y es muy distante con las marcas más alineadas con los países occidentales. Lógicamente, parece una marca de coches que vende vehículos preparados para funcionar a temperaturas extremadamente frías, y sobre terrenos como nieve o incluso hielo.

Por último, la marca de coches D es una marca claramente popular en los países angloparlantes, tanto en Inglaterra como en Irlanda. La principal razón por la que vemos esta clara diferencia con las otras tres marcas es que en estos países se conduce por la izquierda, por lo que las marcas de coches que vendan vehículos allí deben de tener las modificaciones necesarias como para que se pueda conducir según sus reglas.

En conclusión, podemos decir que la marca A es la más popular en los países mediterráneos, la marca B lo es en zonas como Alemania o Países Bajos, la marca C está pensada más para usarse en países fríos como Rusia o países escandinavos, y la marca D probablemente sea una marca de coches para conducir por la izquierda.