

# trabajo-31enero16

Caio Moreno

January 30, 2016

Trabajo Complementos - 31 enero 2016 - Analise de la Correlacion entre las variables (Pobl, Natalidad, EsperanzaVida, Mortalidad).

```
# Author: Caio Fernandes Moreno <caiofern@ucm.es / caiomsouza@gmail.com>
```

```
setwd("/Users/caiomsouza/git/Bitbucket/ucm/COMPLEMENTOS_DE_FORMACION_EN_TECNICAS_DE_MINERIA_DE_DATOS/ta
```

```
países <- read.csv(file="DatosPaíses.csv",head=TRUE,sep=",")
head(países, 10)
```

##	Pais	Pobl	Natalidad	EsperanzaVida	Mortalidad
## 1	Afganist\x87n	27963	35.6	59.8	8.6
## 2	Albania	2902	13.1	77.5	7.2
## 3	Alemania	80435	8.3	80.7	10.8
## 4	Angola	21220	46.2	51.7	14.2
## 5	Arabia Saudita	28091	20.8	74.1	3.4
## 6	Argelia	36036	25.1	74.4	5.1
## 7	Argentina	41223	17.8	76.0	7.6
## 8	Armenia	2963	13.3	74.6	9.0
## 9	Australia	22163	13.5	82.1	6.7
## 10	Austria	8392	9.5	81.1	9.4
##	BalanzaComercial	PIB	ProdCereales		
## 1	-4766	566	157.13532		
## 2	-2861	3786	577.68526		
## 3	205408	41100	2659.28619		
## 4	29864	4221	19.54153		
## 5	144283	19327	10.28894		
## 6	17558	4350	113.11257		
## 7	12057	11508	310.53185		
## 8	-2771	3125	191.30778		
## 9	10724	57593	82.63341		
## 10	-5712	46377	1593.54430		

```
#Dejar solo POBL NATALIDA ESPERANZ MORTALID
```

```
países.valores <- países
```

```
# Remove la columna Países
```

```
países.valores$Pais <- NULL
```

```
# Remove la columna BalanzaComercial
```

```
países.valores$BalanzaComercial <- NULL
```

```
# Remove la columna PIB
```

```
países.valores$PIB <- NULL
```

```
# Remove la columna ProdCereales
```

```
países.valores$ProdCereales <- NULL
```

```
head(países.valores,10)
```

```
##      Pobl Natalidad EsperanzaVida Mortalidad
## 1  27963      35.6          59.8          8.6
## 2   2902      13.1          77.5          7.2
## 3  80435       8.3          80.7         10.8
## 4  21220      46.2          51.7         14.2
## 5  28091      20.8          74.1          3.4
## 6  36036      25.1          74.4          5.1
## 7  41223      17.8          76.0          7.6
## 8   2963      13.3          74.6          9.0
## 9  22163      13.5          82.1          6.7
## 10 8392       9.5          81.1          9.4
```

```
colnames(países.valores)
```

```
## [1] "Pobl"          "Natalidad"      "EsperanzaVida" "Mortalidad"
```

```
# Normaliza las variables
```

```
países.valores.normalizar <- scale(países.valores)
head(países.valores.normalizar, 10)
```

```
##      Pobl Natalidad EsperanzaVida Mortalidad
## [1,] -0.03332065  1.2692333    -1.2643593  0.07557593
## [2,] -0.64060228 -0.8127601     0.7680260 -0.41474593
## [3,]  1.23818813 -1.2569187     1.1354628  0.84608170
## [4,] -0.19671796  2.2500835    -2.1944339  2.03686334
## [5,] -0.03021893 -0.1002557     0.3776243 -1.74561954
## [6,]  0.16230541  0.2976364     0.4120715 -1.15022871
## [7,]  0.28799751 -0.3778548     0.5957899 -0.27465397
## [8,] -0.63912412 -0.7942535     0.4350363  0.21566788
## [9,] -0.17386705 -0.7757469     1.2962165 -0.58986088
## [10,] -0.50756783 -1.1458791     1.1813925  0.35575984
```

```
países.cor <- cor(países.valores.normalizar)
#View(países.cor)
países.cor
```

```
##      Pobl Natalidad EsperanzaVida Mortalidad
## Pobl      1.00000000 -0.03243038 -0.01445153 -0.03519014
## Natalidad -0.03243038  1.00000000 -0.87063840  0.06833273
## EsperanzaVida -0.01445153 -0.87063840  1.00000000 -0.38664292
## Mortalidad -0.03519014  0.06833273 -0.38664292  1.00000000
```

```
cat("Se puede ver una correlacion muy alta entre EsperanzaVida y Natalidad de -0.87063840")
```

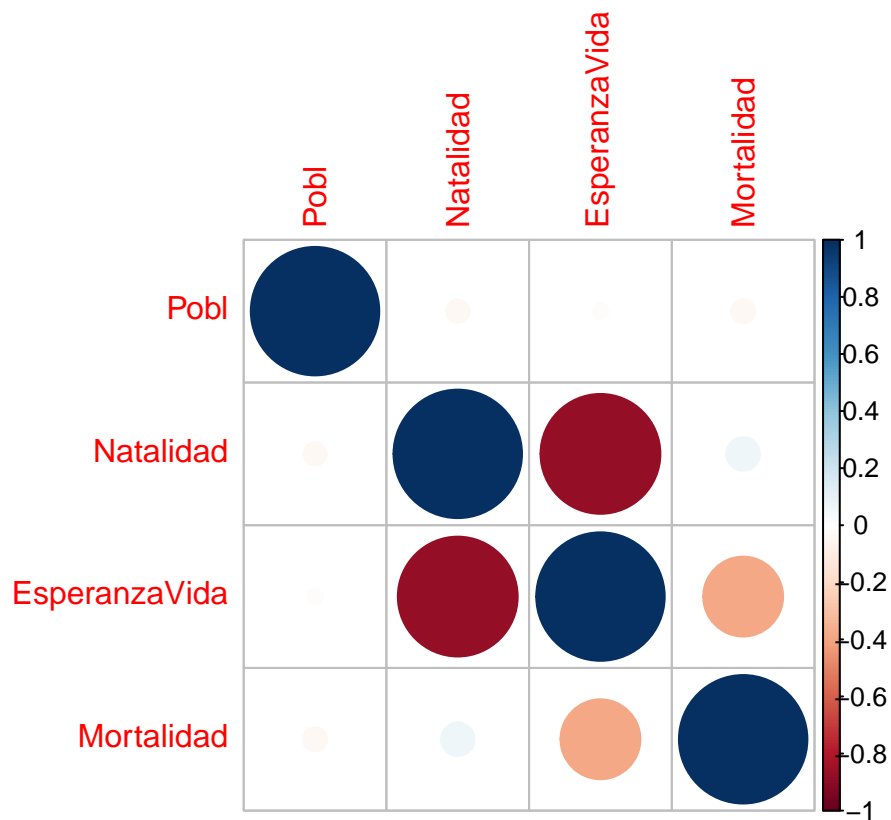
```
## Se puede ver una correlacion muy alta entre EsperanzaVida y Natalidad de -0.87063840
```

```
cat("Se percibe que cuanto mayor la Esperanza de Vida menos niños en un pais.")
```

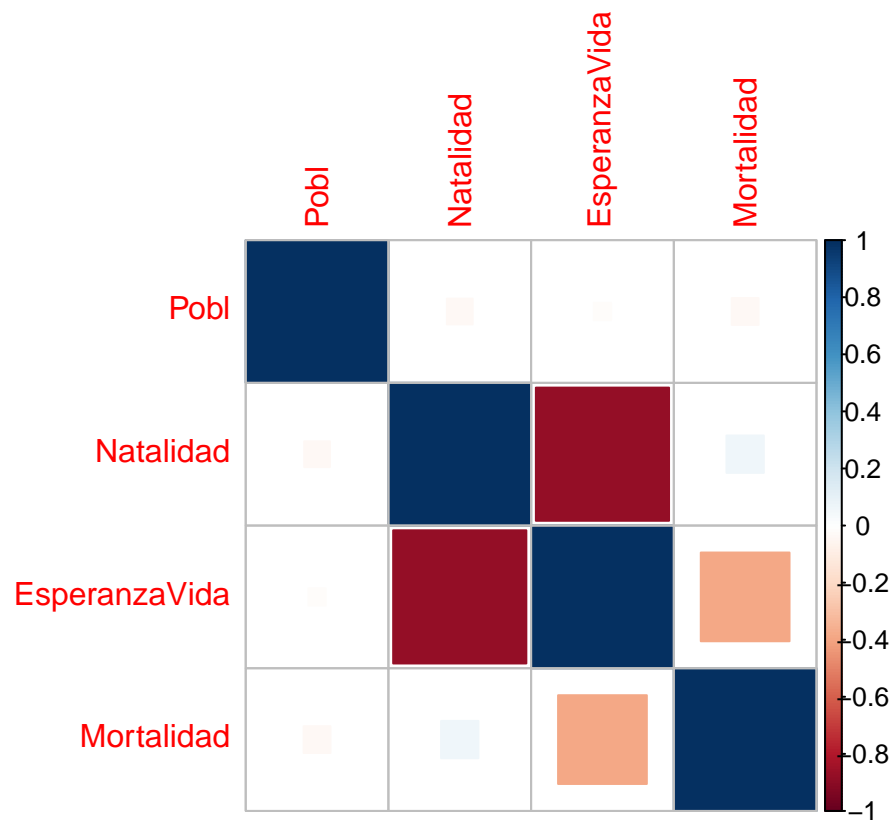
```
## Se percibe que cuanto mayor la Esperanza de Vida menos niños en un pais.
```

```
# Utilizando la libreria corrplot para visualizar mejor las correlaciones entre las variables.  
# https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html
```

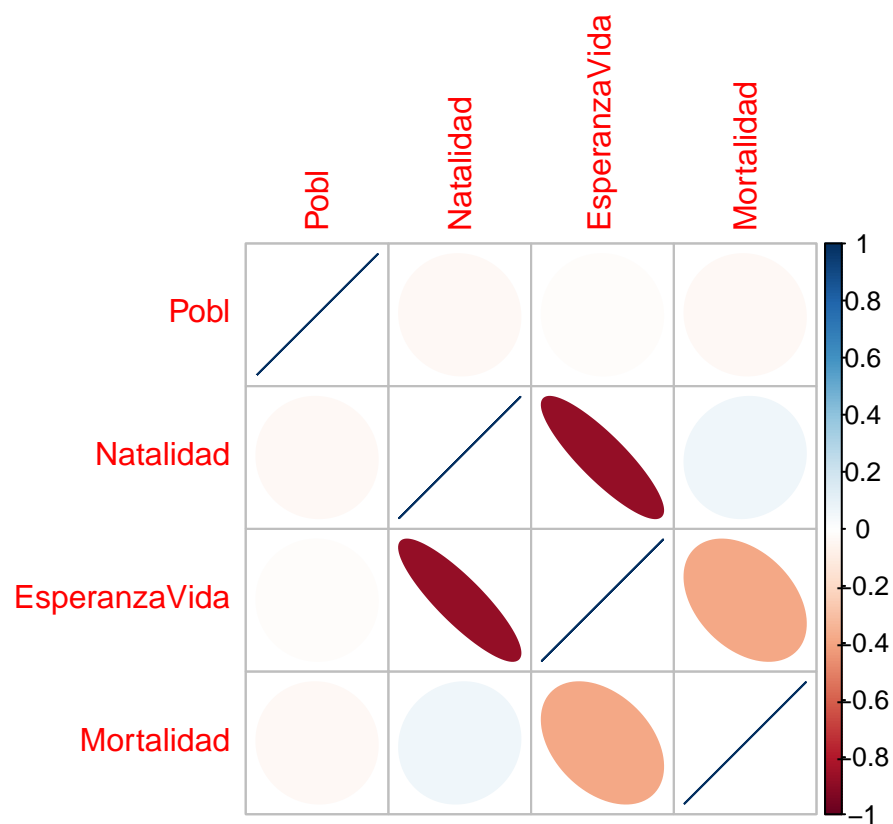
```
library(corrplot)  
M <- cor(países.valores.normalizar)  
corrplot(M, method = "circle")
```



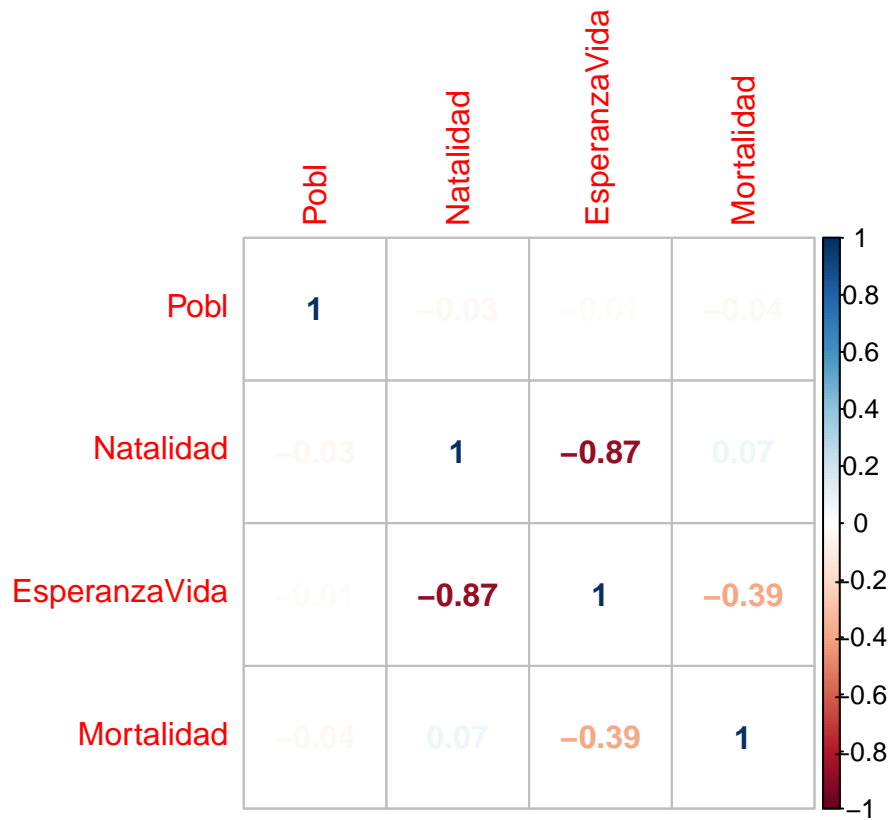
```
corrplot(M, method = "square")
```



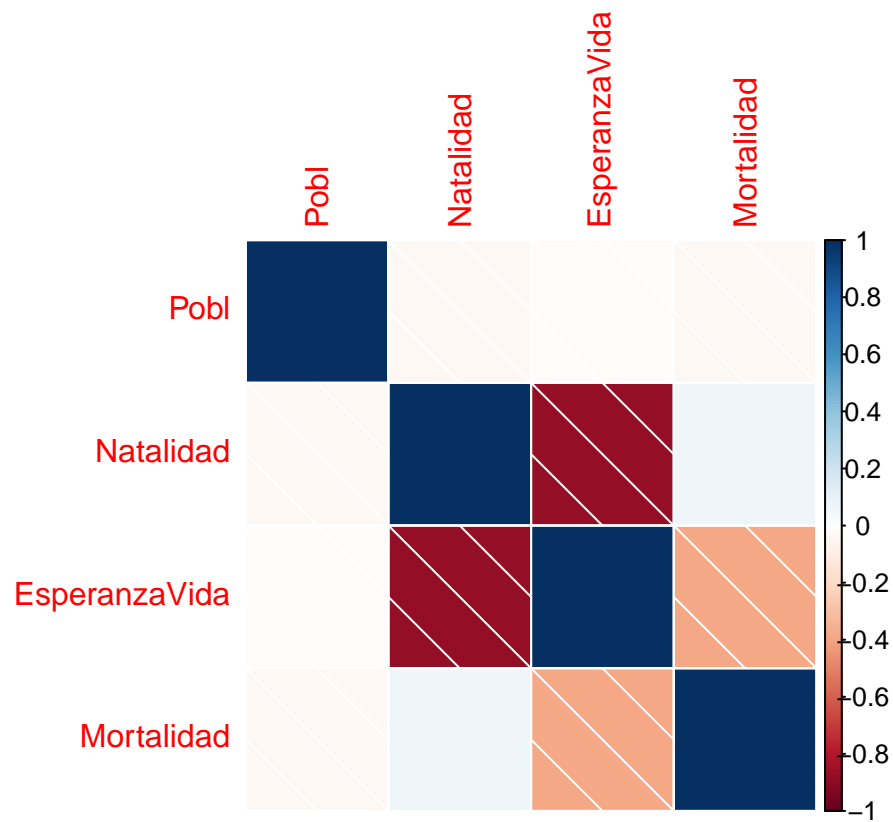
```
corrplot(M, method = "ellipse")
```



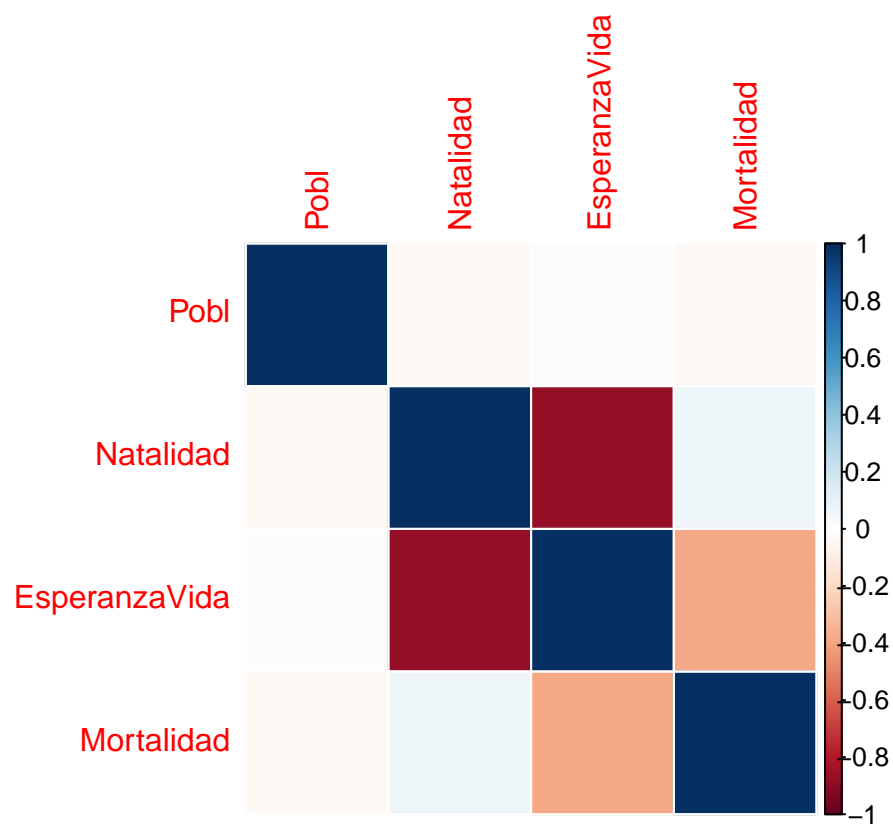
```
corrplot(M, method = "number")
```



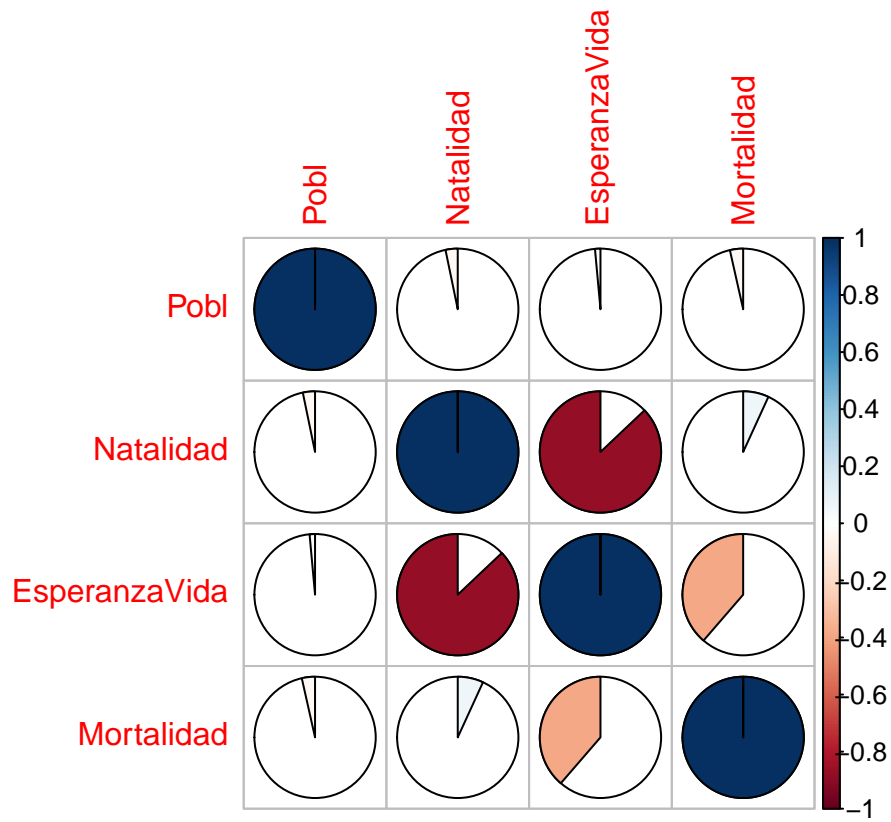
```
corrplot(M, method = "shade")
```



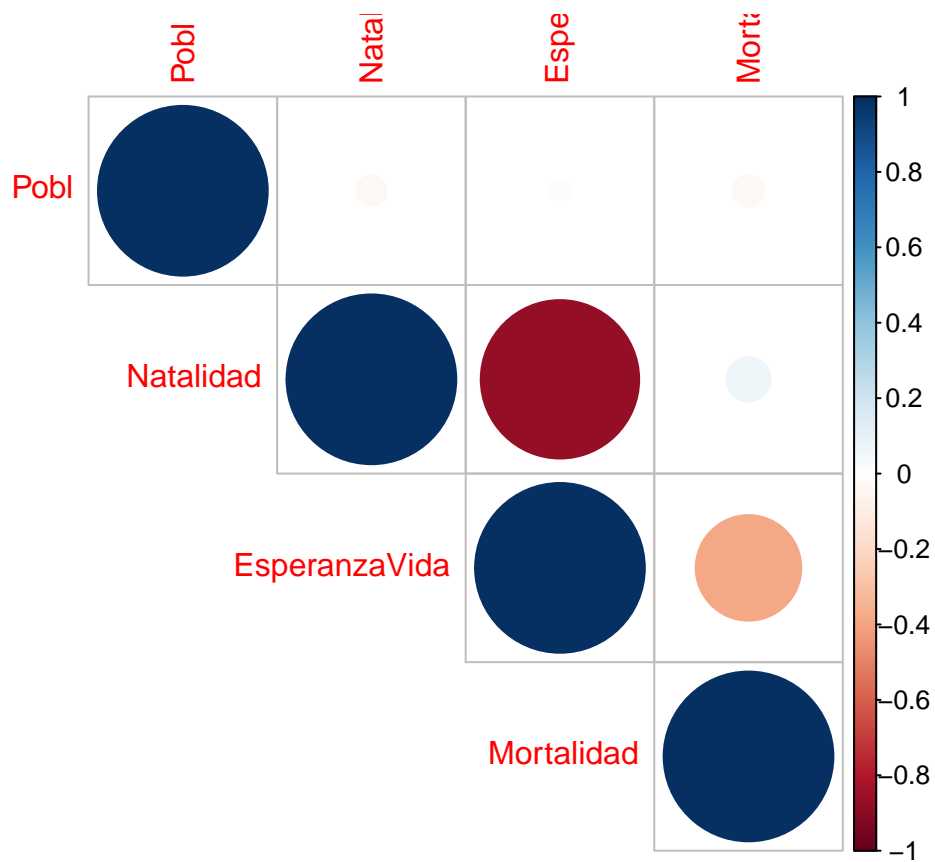
```
corrplot(M, method = "color")
```



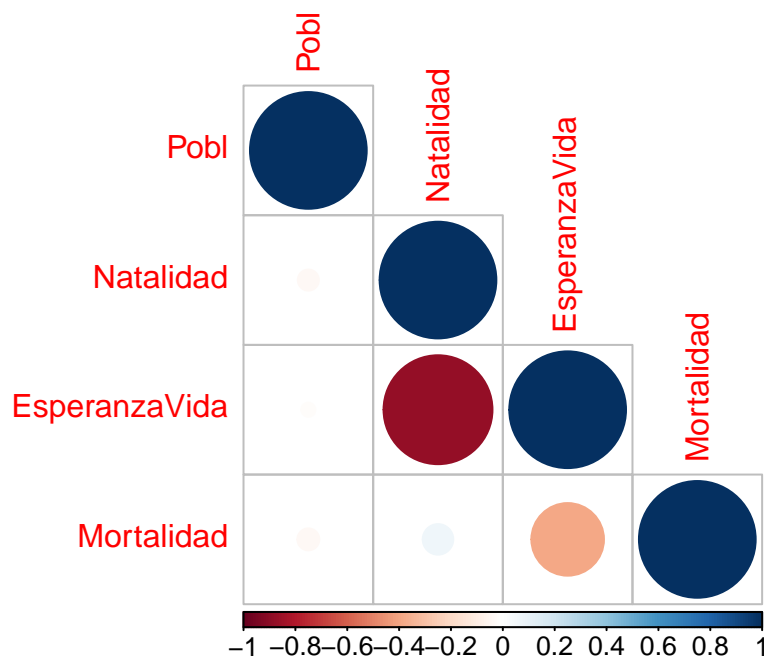
```
corrplot(M, method = "pie")
```



```
corrplot(M, type = "upper")
```

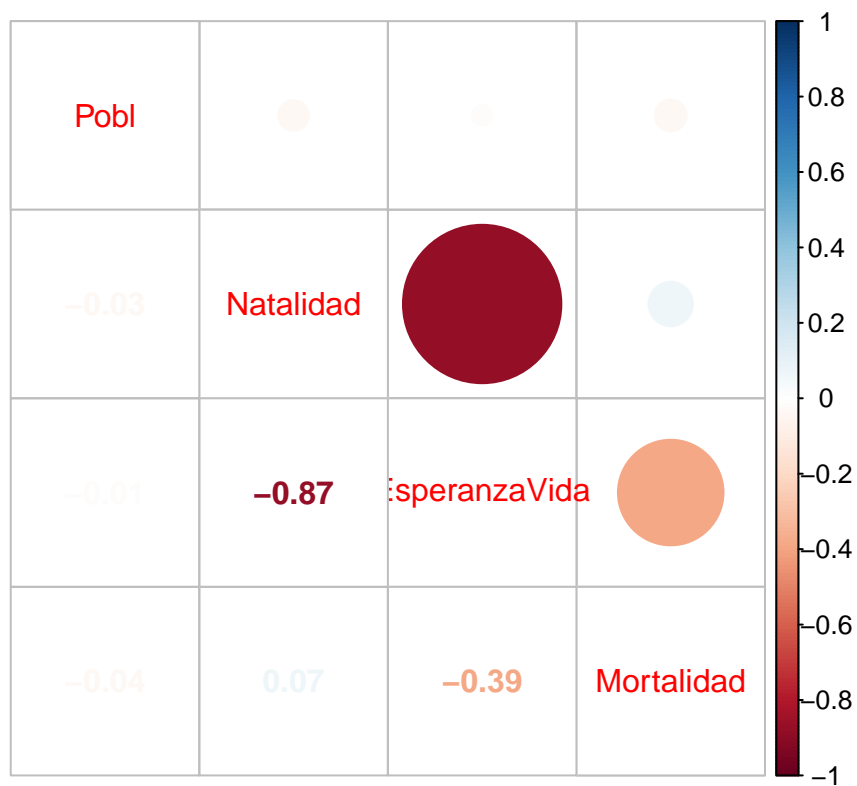


```
corrplot(M, type = "lower")
```

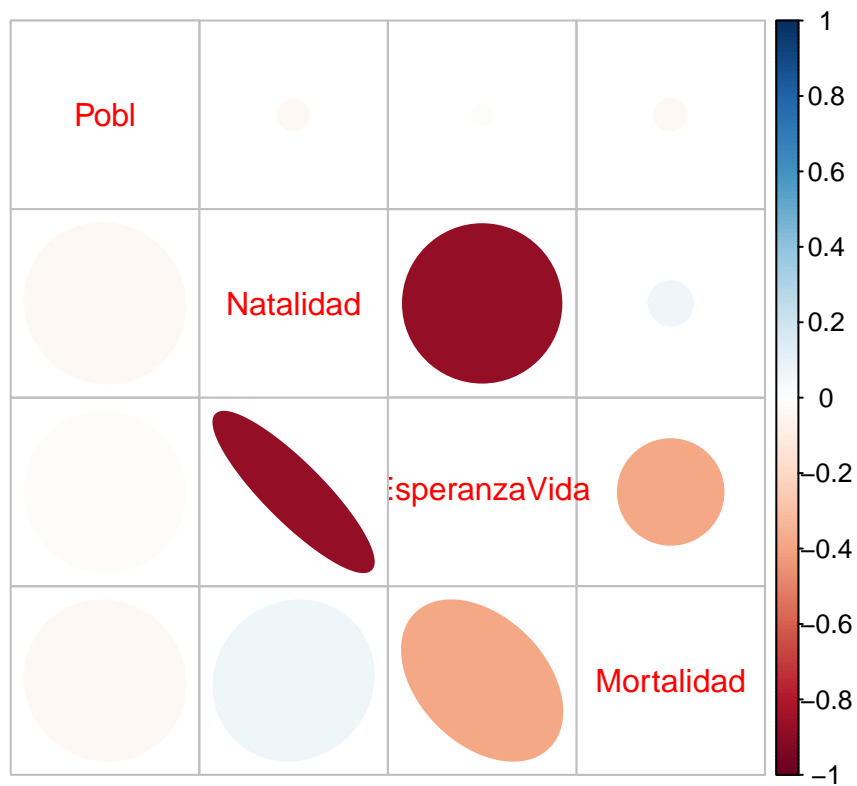


```
corrplot.mixed(M)
```

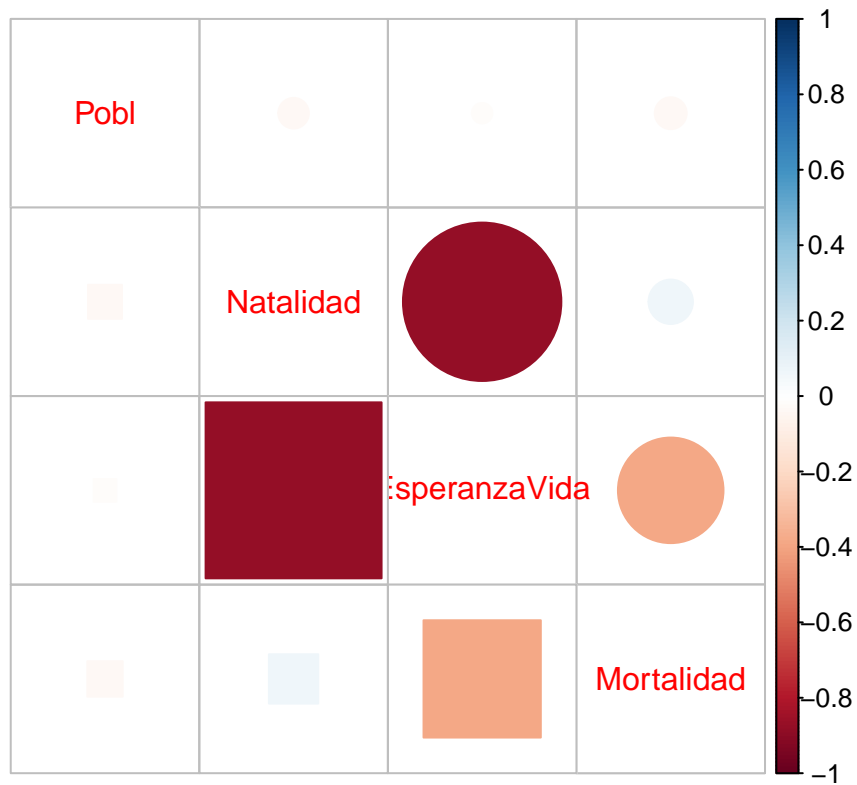




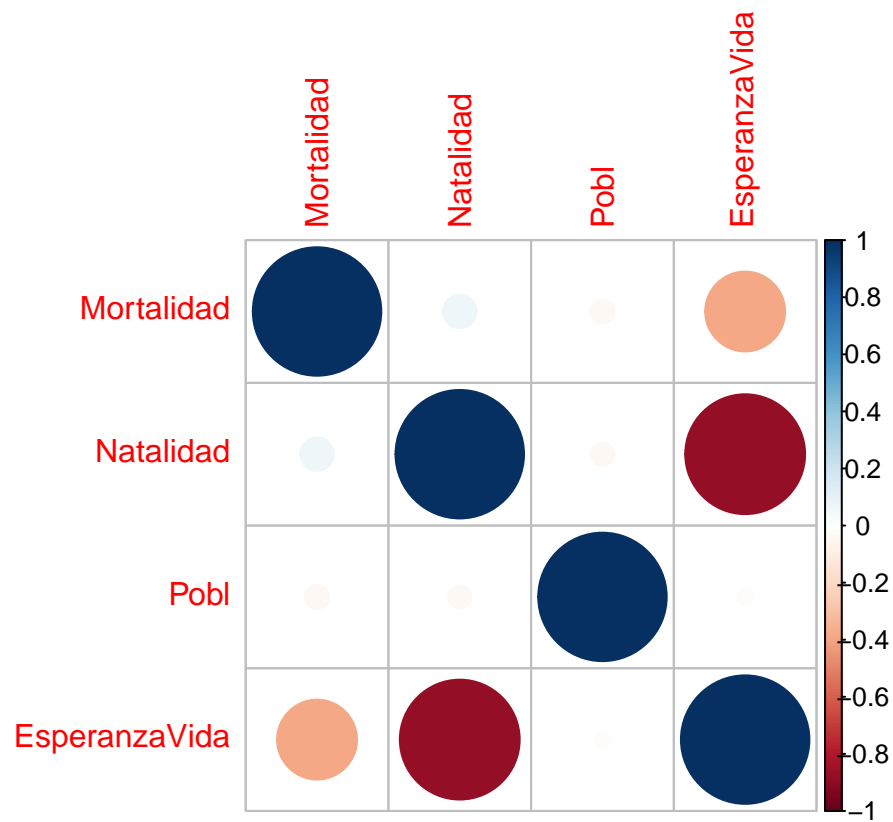
```
corrplot.mixed(M, lower = "ellipse", upper = "circle")
```



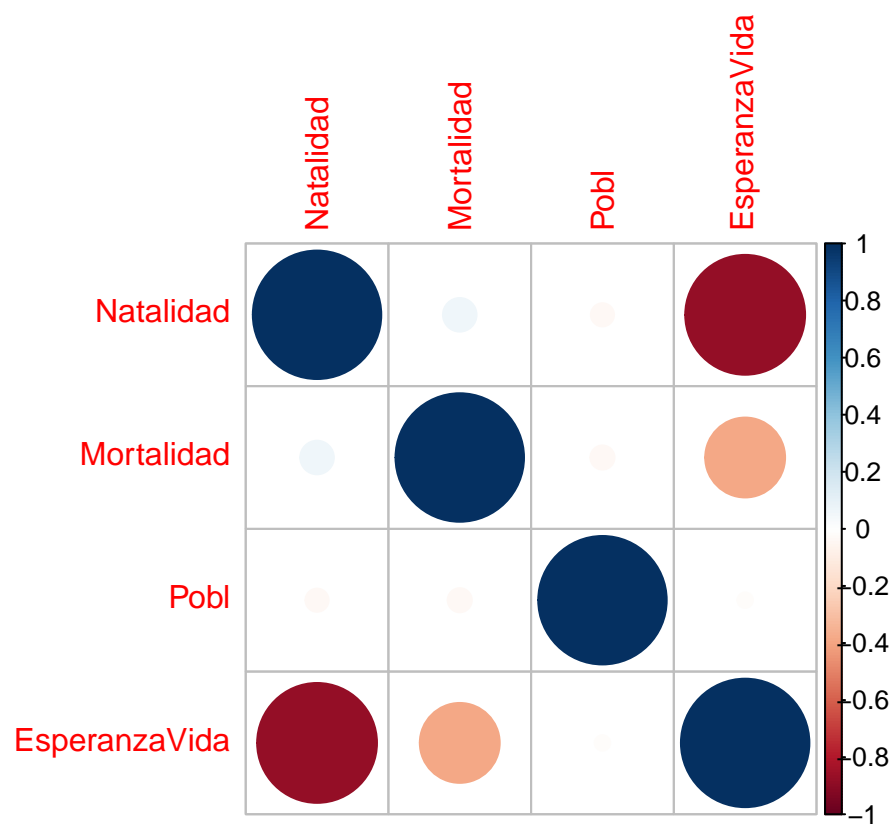
```
corrplot.mixed(M, lower = "square", upper = "circle")
```



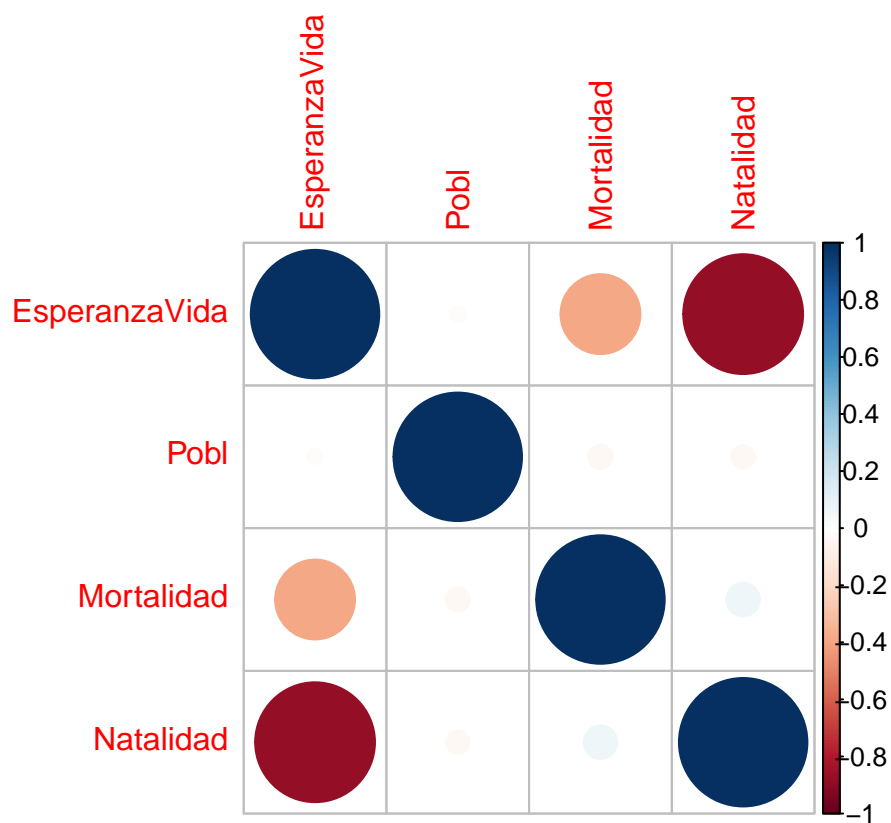
```
corrplot(M, order = "AOE")
```



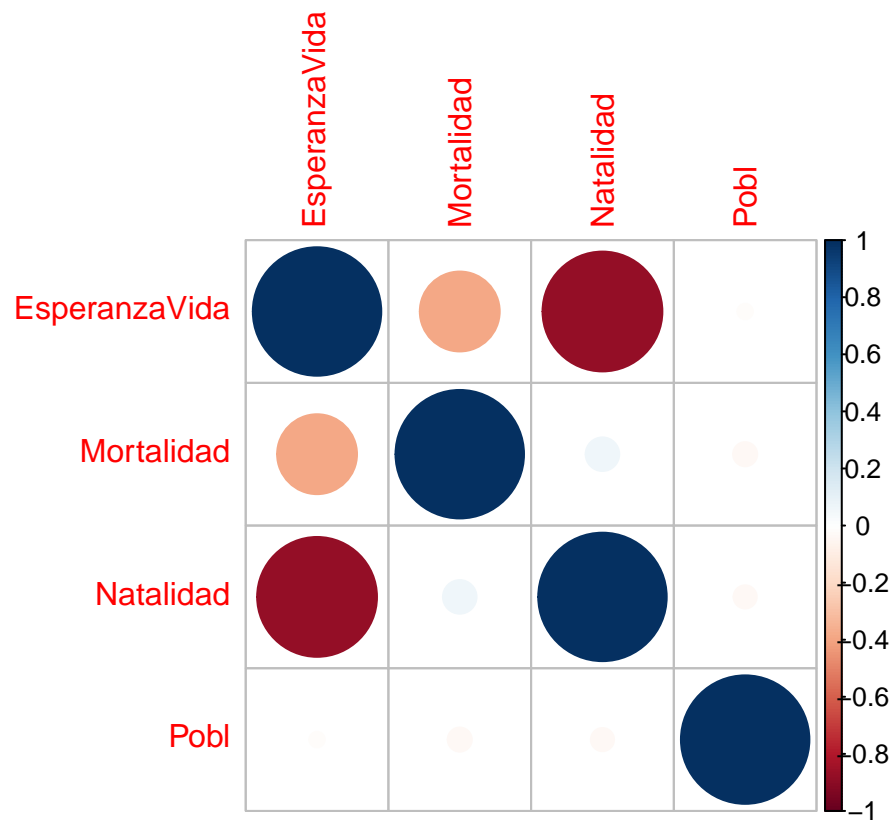
```
corrplot(M, order = "hclust")
```



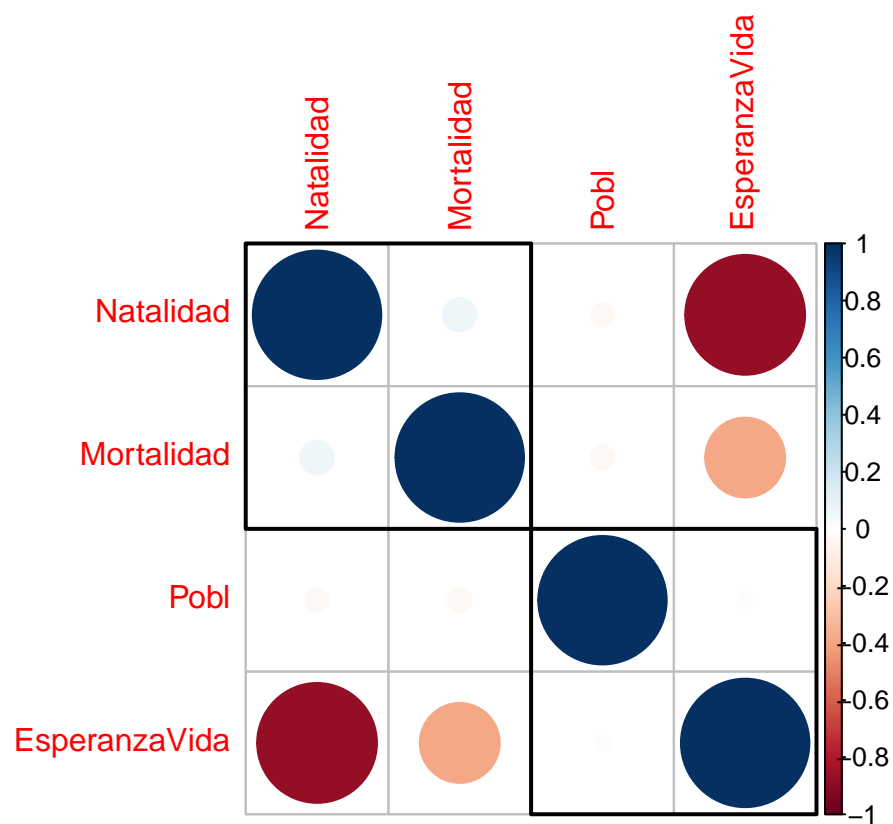
```
corrplot(M, order = "FPC")
```



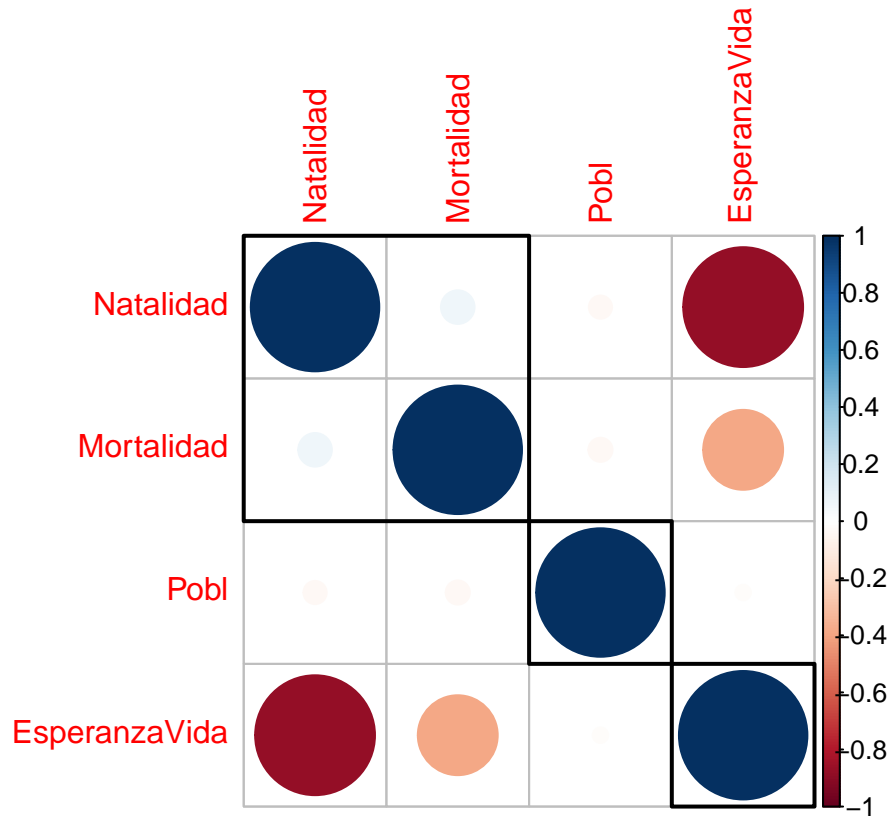
```
corrplot(M, order = "alphabet")
```



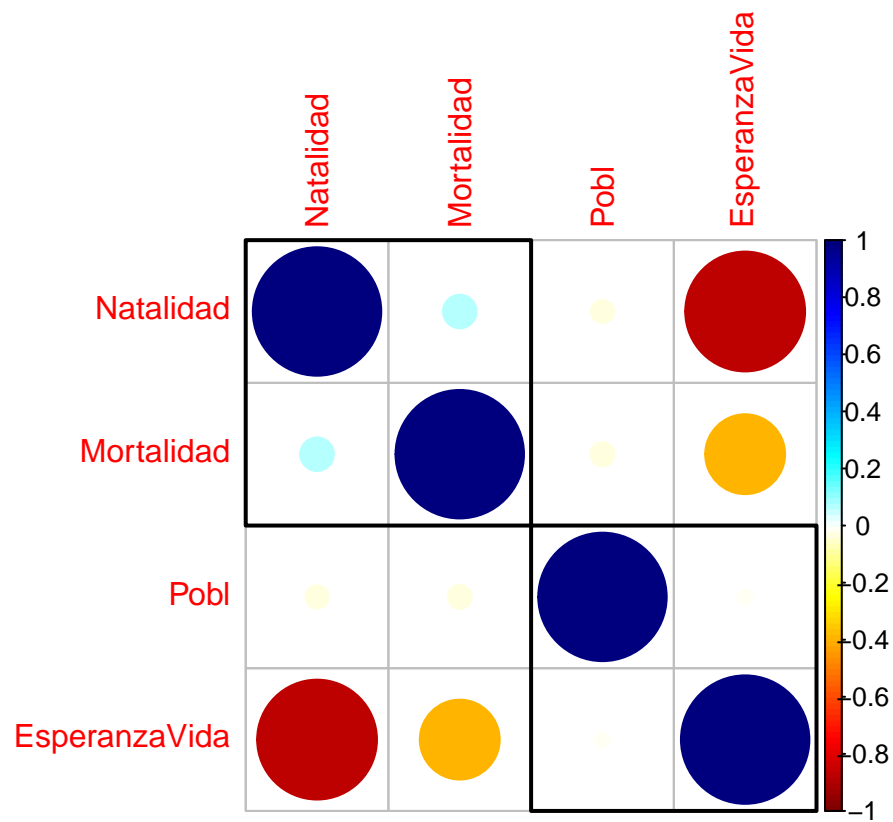
```
corrplot(M, order = "hclust", addrect = 2)
```



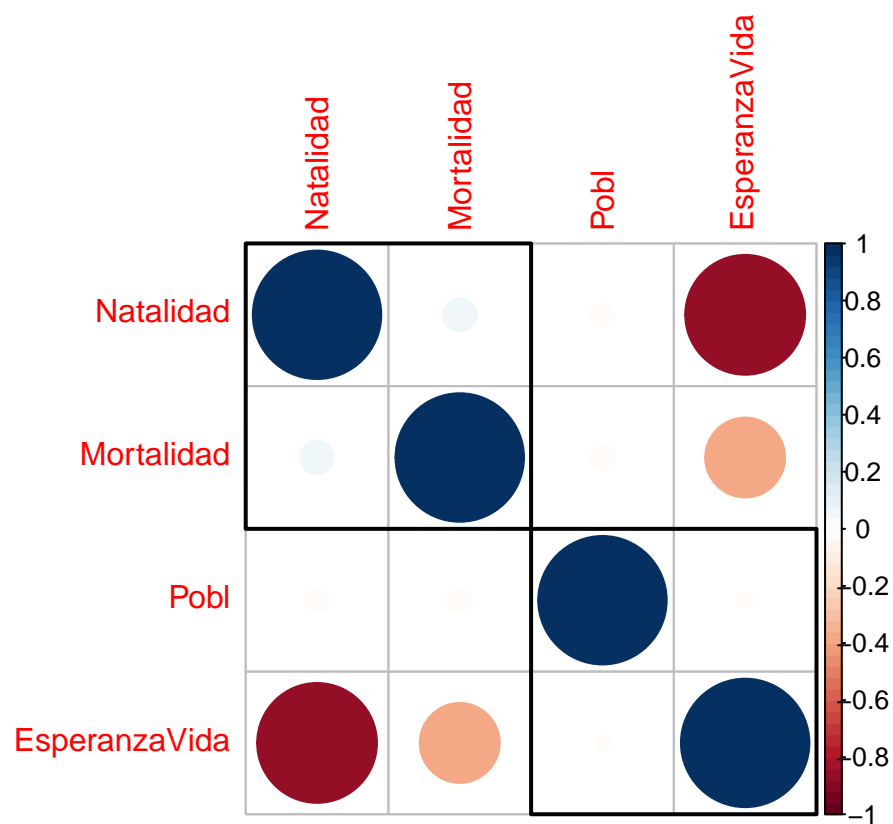
```
corrplot(M, order = "hclust", addrect = 3)
```



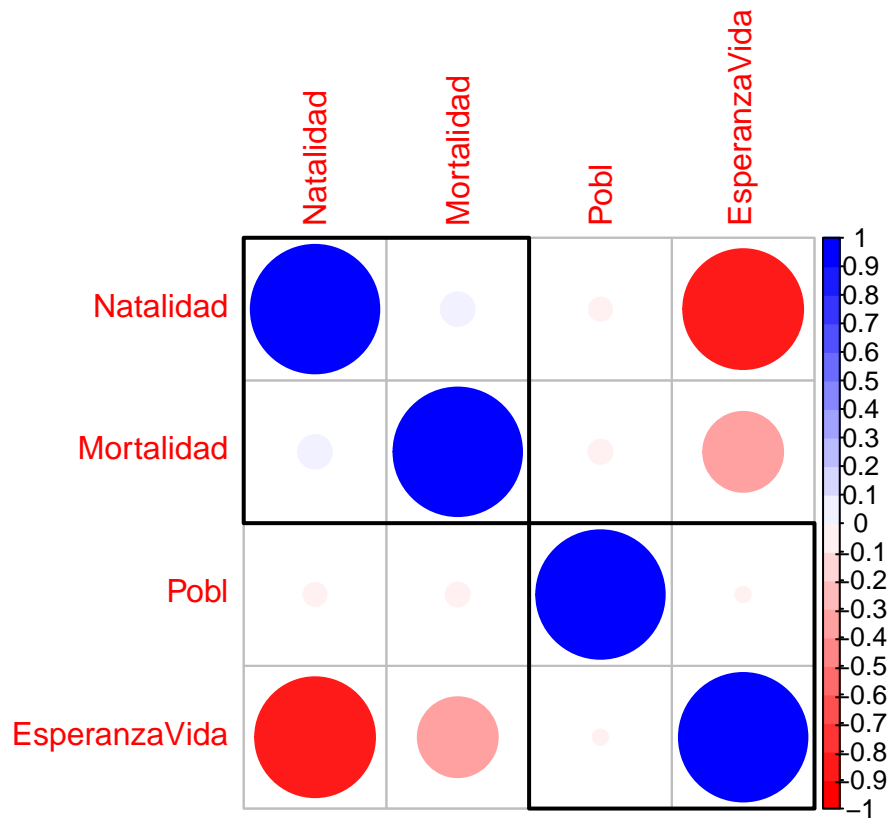
```
col1 <- colorRampPalette(c("#7F0000", "red", "#FF7F00", "yellow", "white", "cyan",
                           "#007FFF", "blue", "#00007F"))
col2 <- colorRampPalette(c("#67001F", "#B2182B", "#D6604D", "#F4A582", "#FDDBC7",
                           "#FFFFFF", "#D1E5F0", "#92C5DE", "#4393C3", "#2166AC", "#053061"))
col3 <- colorRampPalette(c("red", "white", "blue"))
col4 <- colorRampPalette(c("#7F0000", "red", "#FF7F00", "yellow", "#7FFF7F",
                           "cyan", "#007FFF", "blue", "#00007F"))
wb <- c("white", "black")
## using these color spectrums
corrplot(M, order = "hclust", addrect = 2, col = col1(100))
```



```
corrplot(M, order = "hclust", addrect = 2, col = col2(50))
```

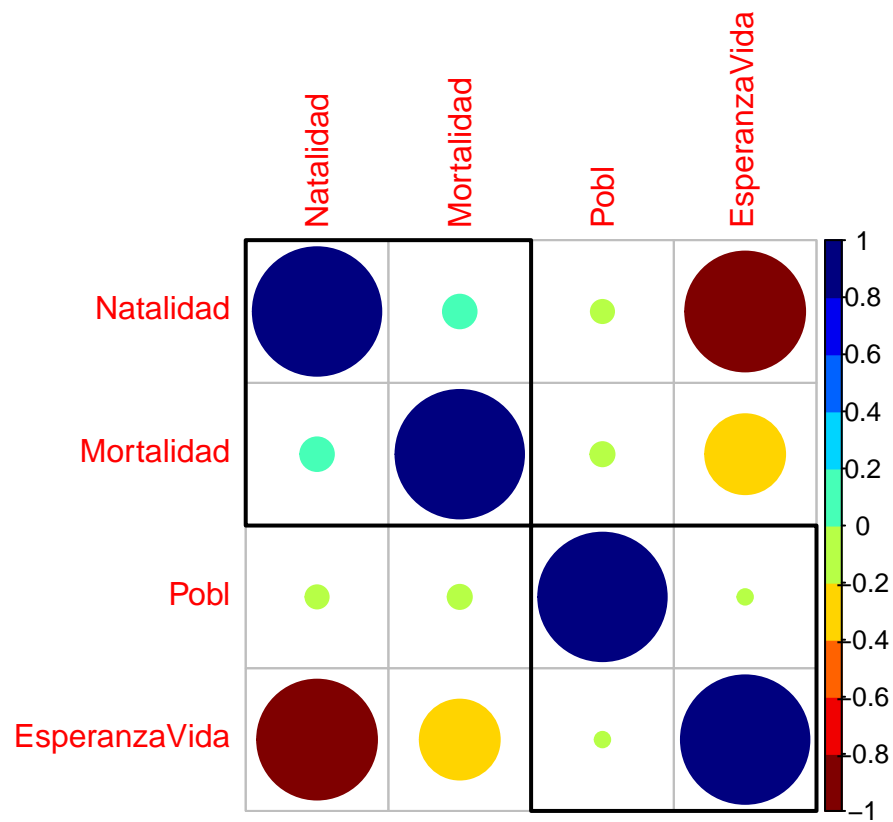


```
corrplot(M, order = "hclust", addrect = 2, col = col3(20))
```

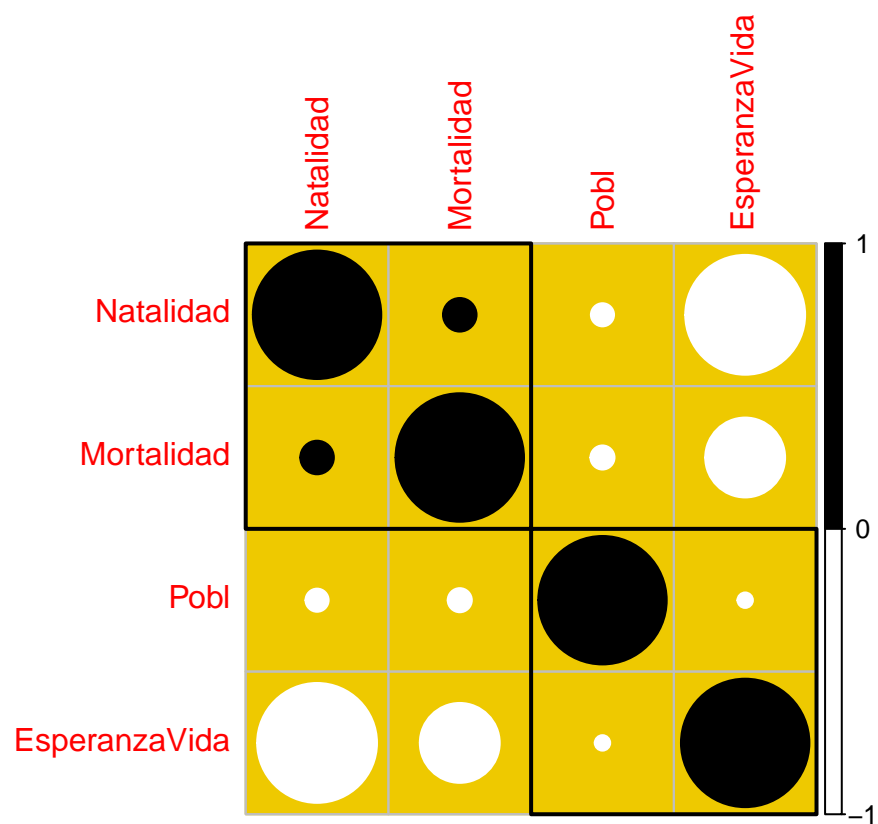


```
corrplot(M, order = "hclust", addrect = 2, col = col4(10))
```

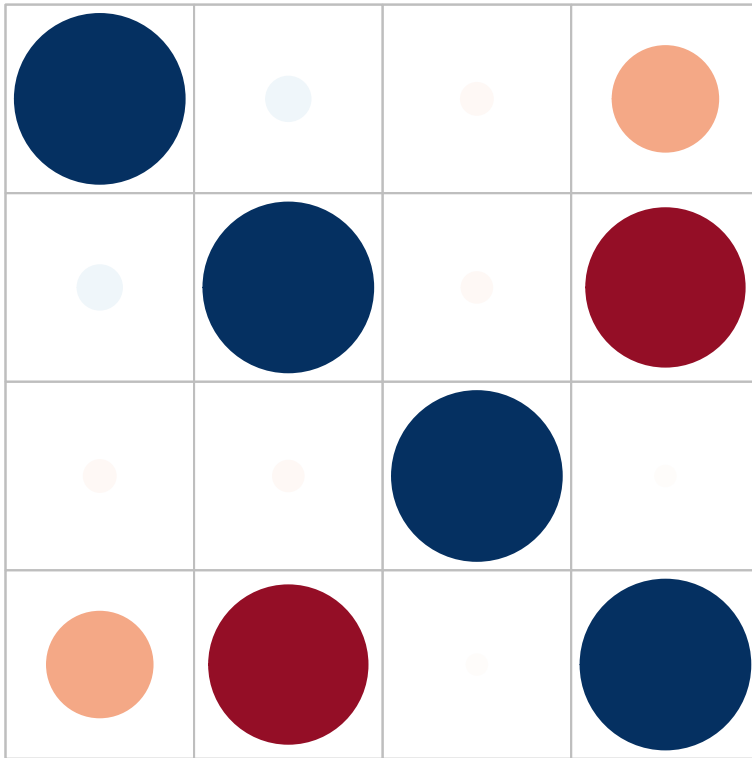




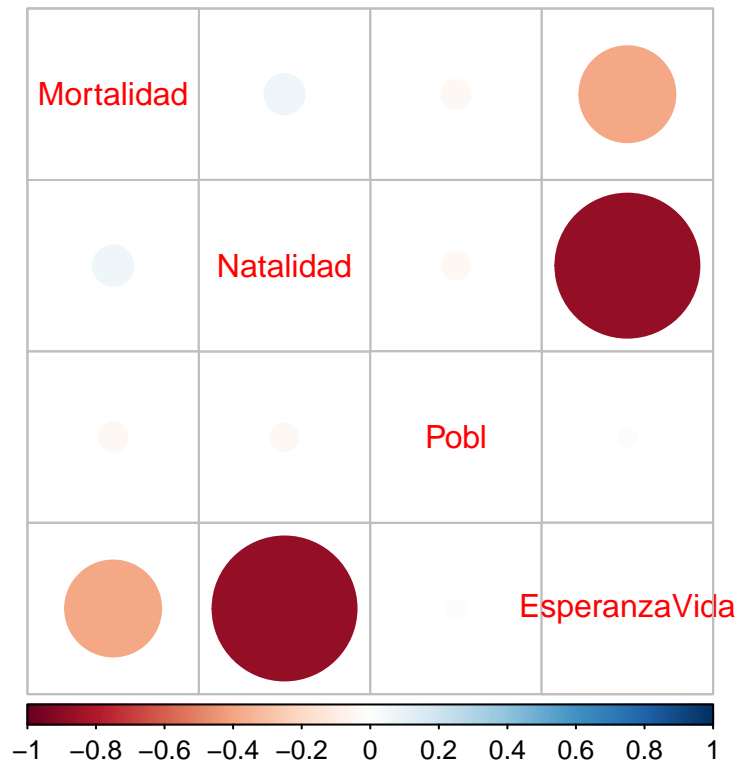
```
corrplot(M, order = "hclust", addrect = 2, col = "wb", bg = "gold2")
```



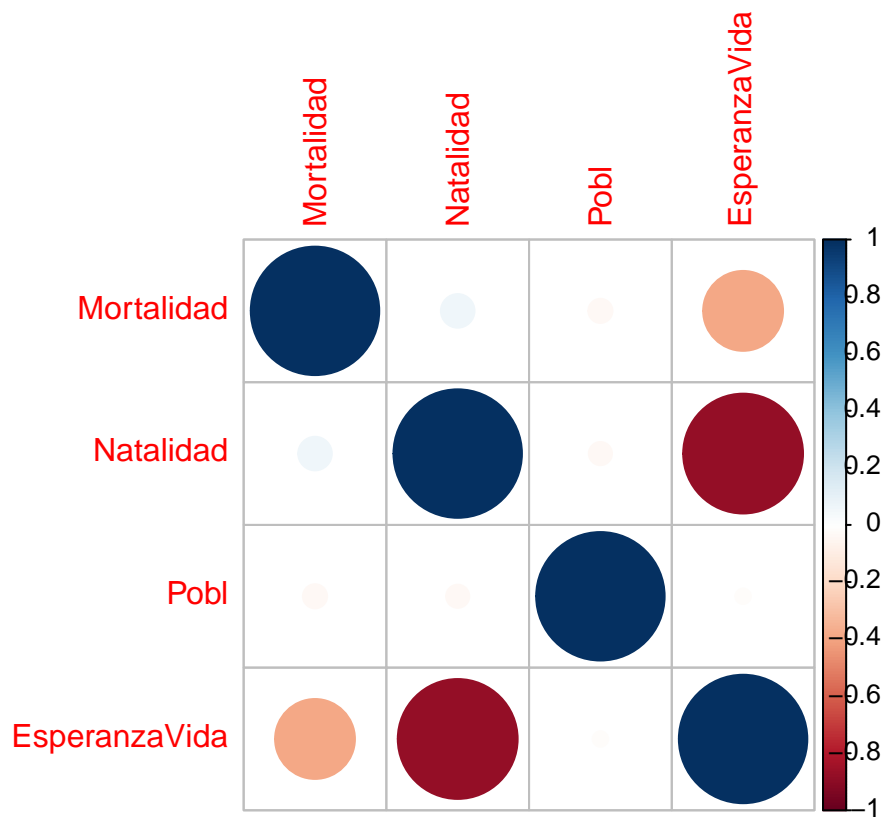
```
## remove color legend and text legend
corrplot(M, order = "AOE", cl.pos = "n", tl.pos = "n")
```



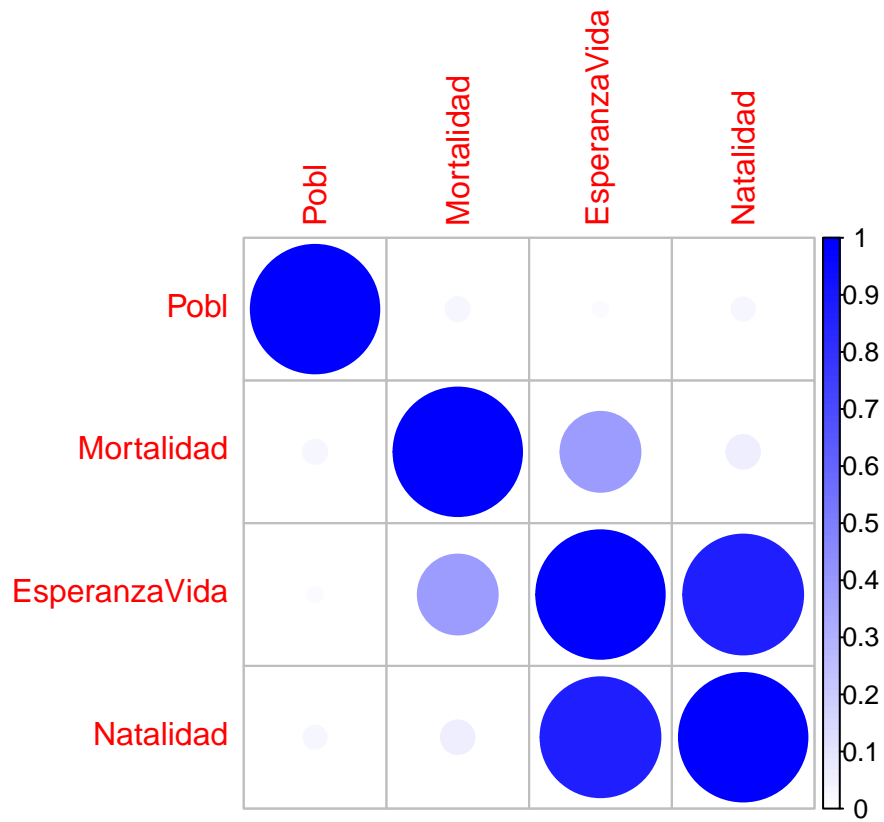
```
## bottom color legend, diagonal text legend, rotate text label
corrplot(M, order = "AOE", cl.pos = "b", tl.pos = "d", tl.srt = 60)
```



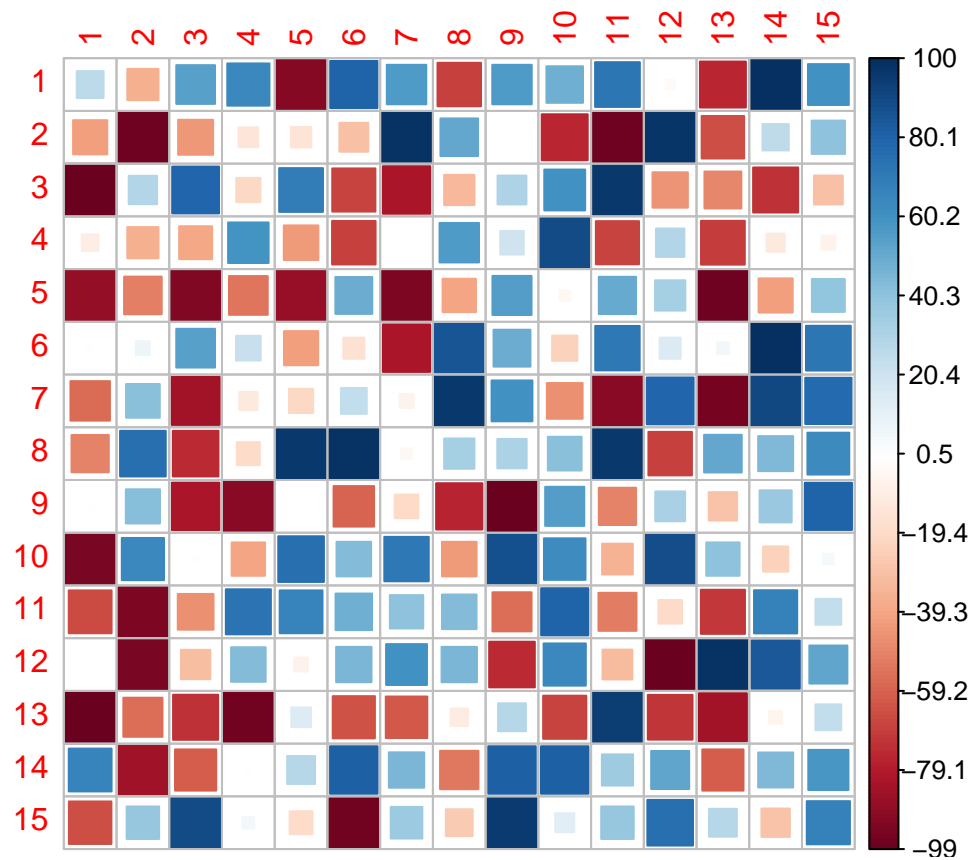
```
## a wider color legend with numbers right aligned
corrplot(M, order = "AOE", cl.ratio = 0.2, cl.align = "r")
```



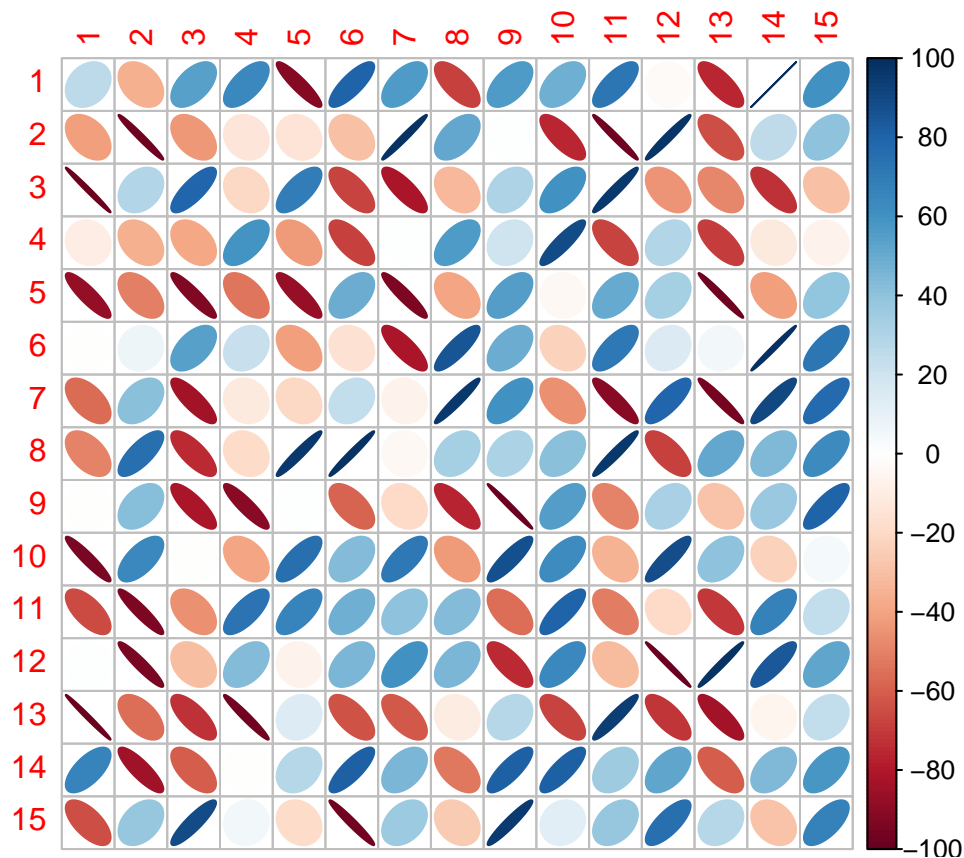
```
corrplot(abs(M), order = "AOE", col = col3(200), cl.lim = c(0, 1))
```



```
## visualize a matrix in [-100, 100]
ran <- round(matrix(runif(225, -100, 100), 15))
corrplot(ran, is.corr = FALSE, method = "square")
```

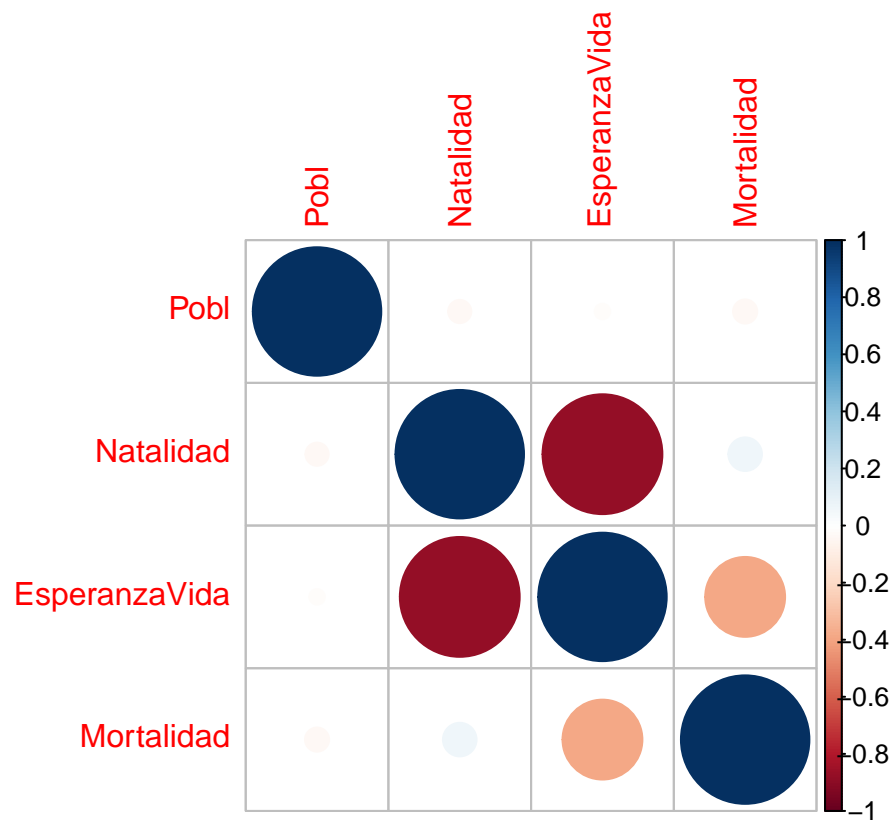


```
## a beautiful color legend
corrplot(ran, is.corr = FALSE, method = "ellipse", cl.lim = c(-100, 100))
```

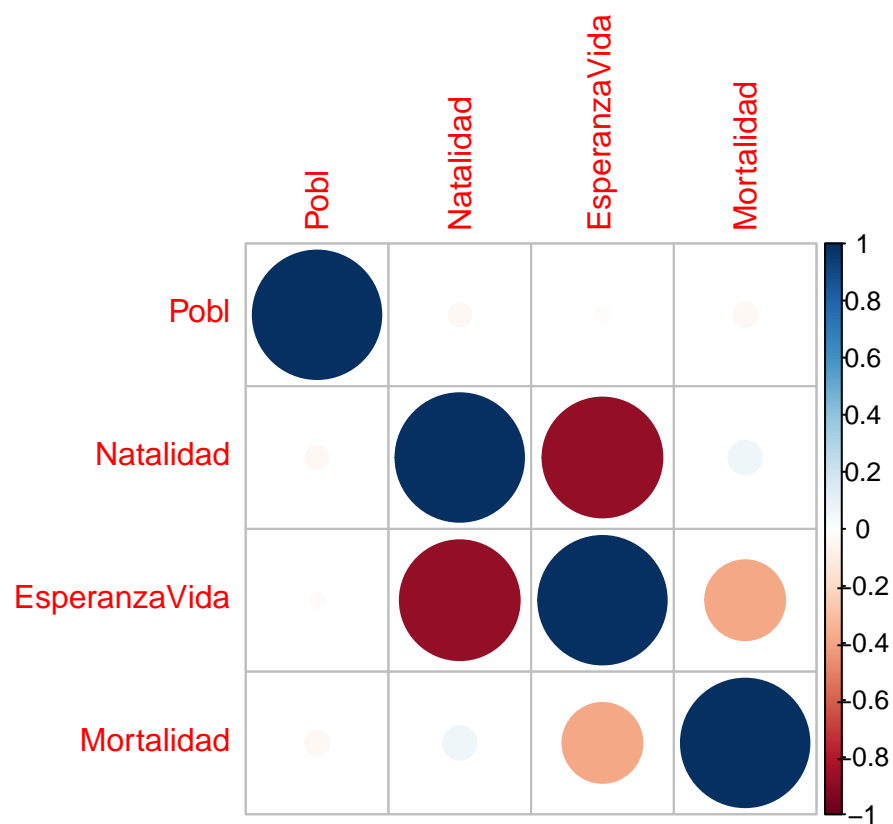


```
cor.mtest <- function(mat, conf.level = 0.95) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat <- lowCI.mat <- uppCI.mat <- matrix(NA, n, n)
  diag(p.mat) <- 0
  diag(lowCI.mat) <- diag(uppCI.mat) <- 1
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], conf.level = conf.level)
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
      lowCI.mat[i, j] <- lowCI.mat[j, i] <- tmp$conf.int[1]
      uppCI.mat[i, j] <- uppCI.mat[j, i] <- tmp$conf.int[2]
    }
  }
  return(list(p.mat, lowCI.mat, uppCI.mat))
}

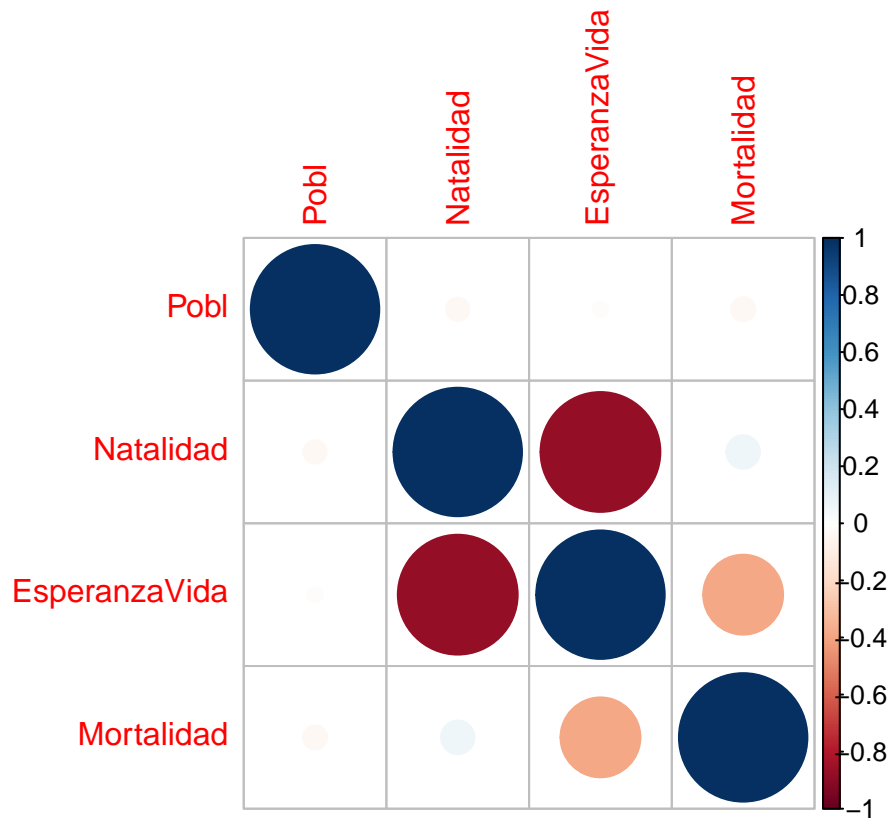
res1 <- cor.mtest(mtcars, 0.95)
res2 <- cor.mtest(mtcars, 0.99)
## specialized the insignificant value according to the significant level
corrplot(M, p.mat = res1[[1]], sig.level = 0.2)
```



```
corrplot(M, p.mat = res1[[1]], sig.level = 0.05)
```

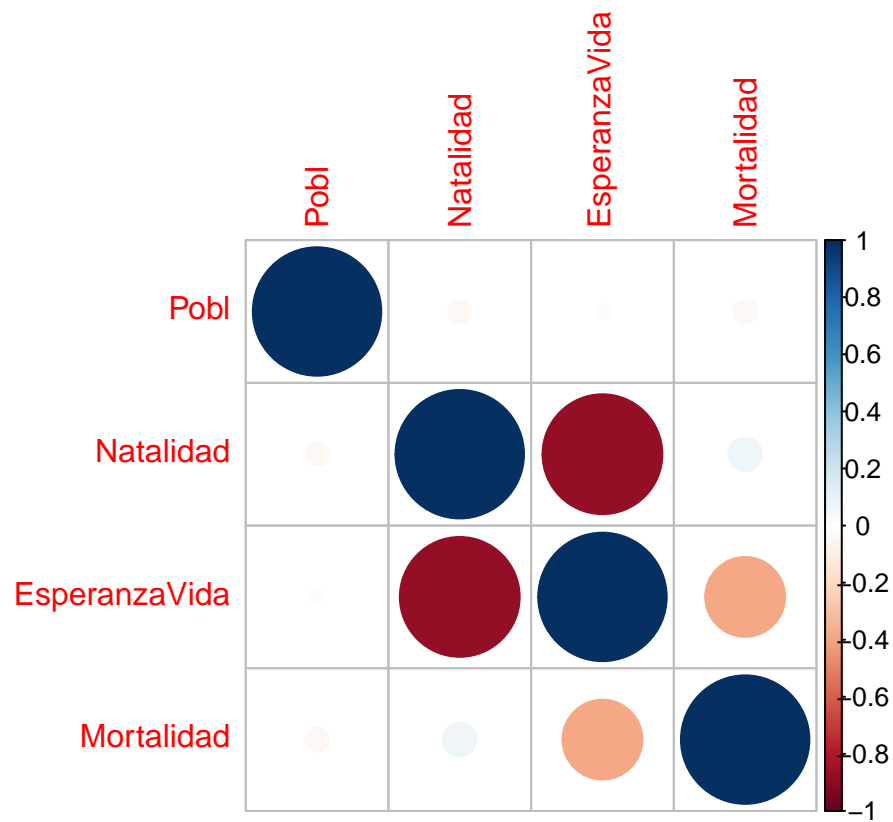


```
corrplot(M, p.mat = res1[[1]], sig.level = 0.01)
```

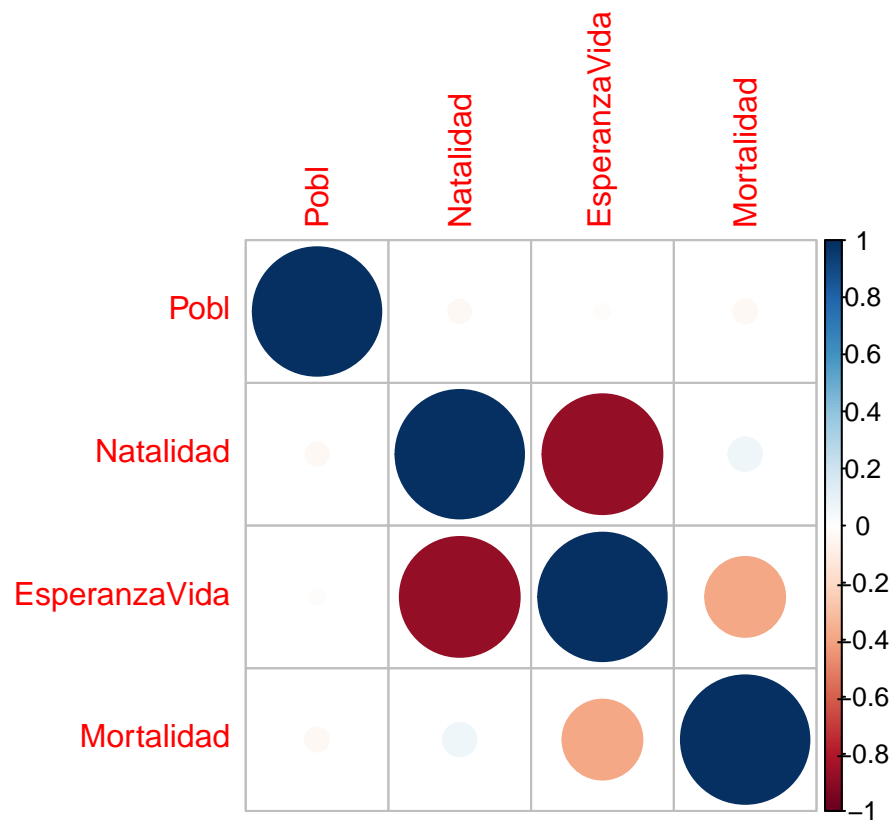


```
## leave blank on no significant coefficient
corrplot(M, p.mat = res1[[1]], insig = "blank")
```

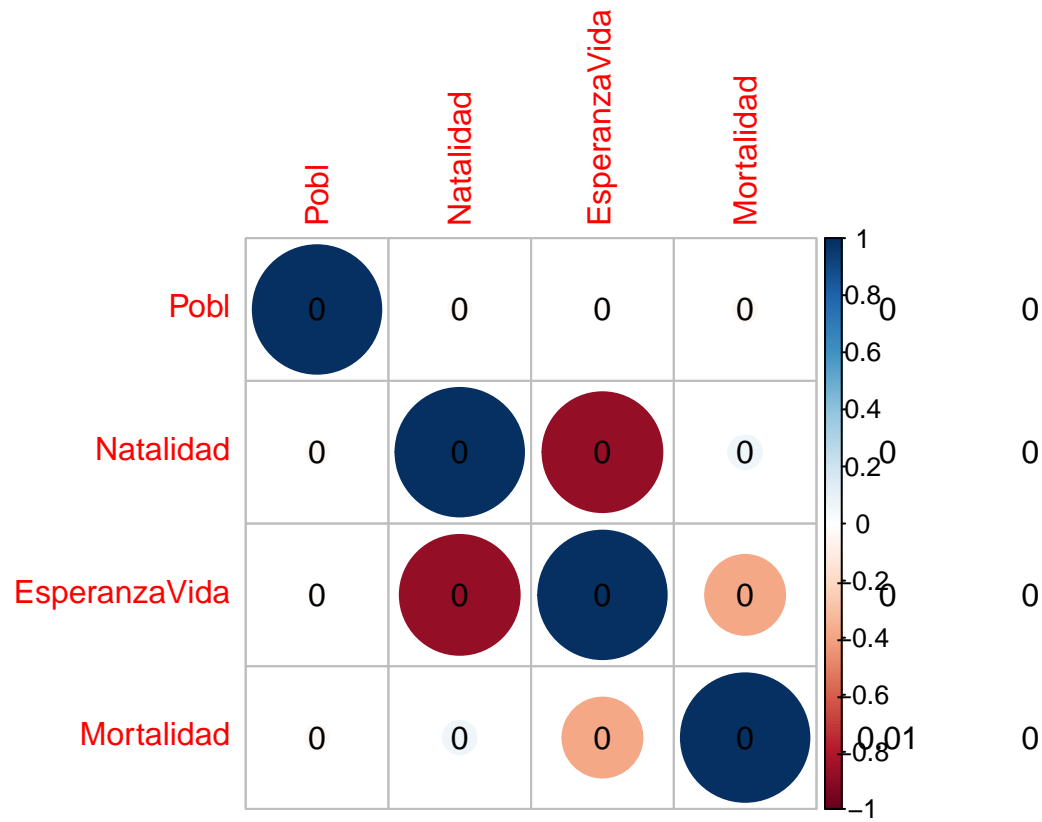




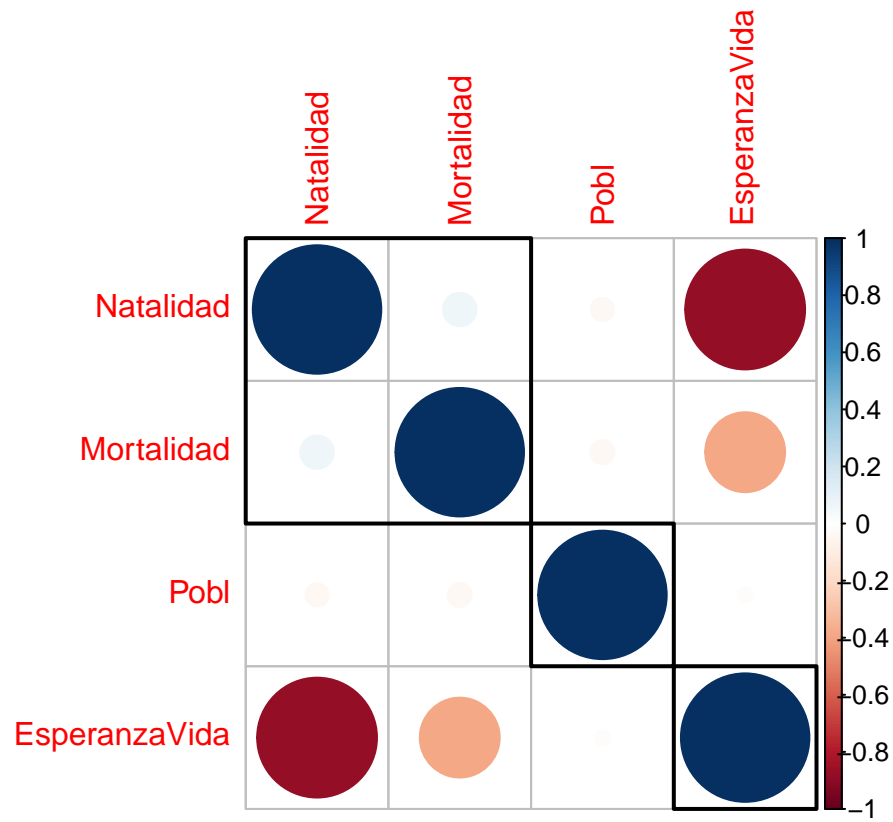
```
## add p-values on no significant coefficient
corrplot(M, p.mat = res1[[1]], insig = "p-value")
```



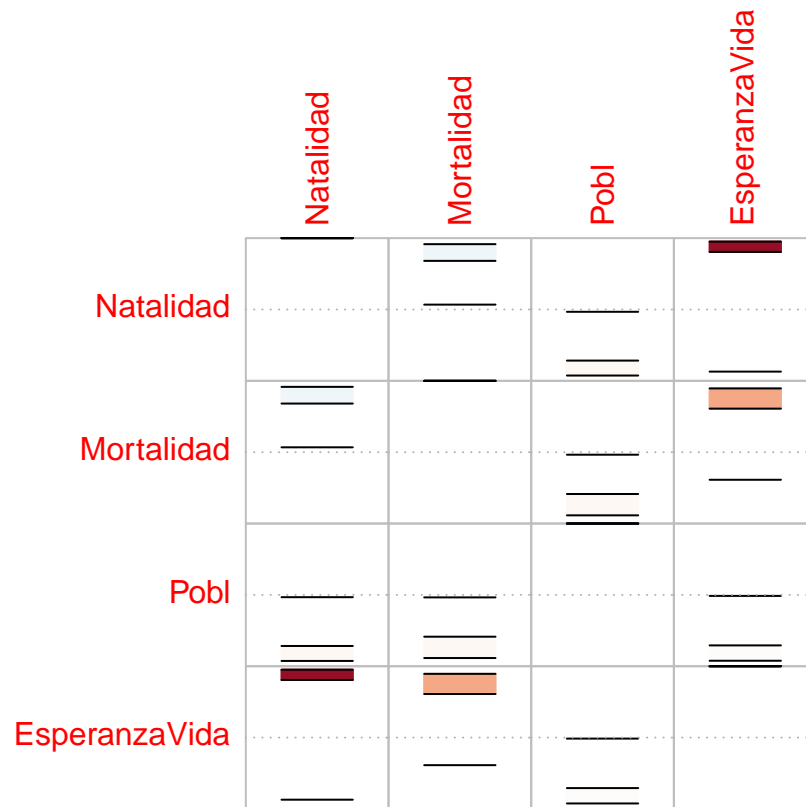
```
## add all p-values
corrplot(M, p.mat = res1[[1]], insig = "p-value", sig.level = -1)
```



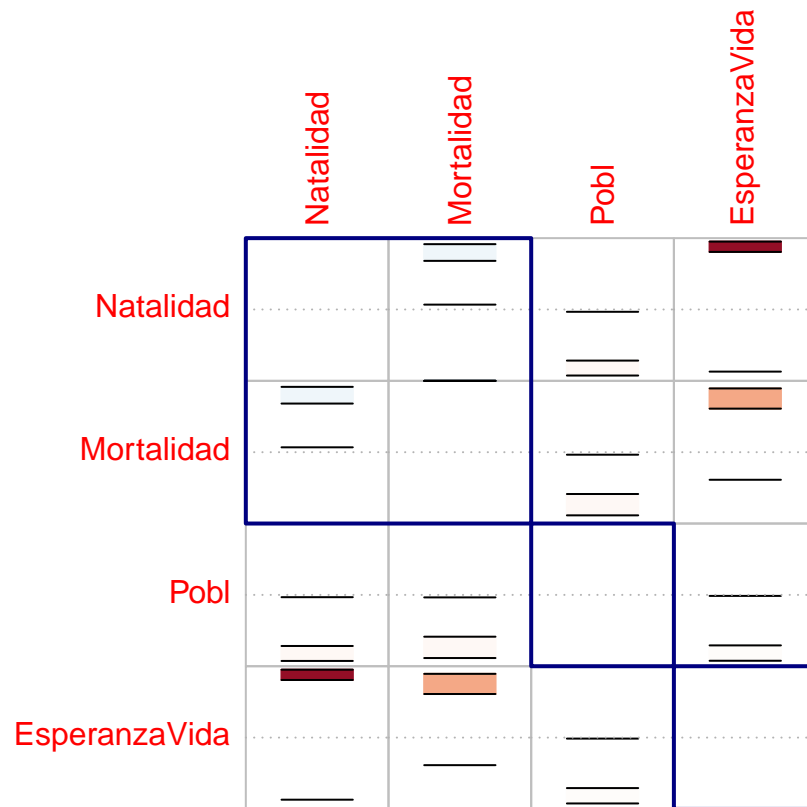
```
## add cross on no significant coefficient
corrplot(M, p.mat = res1[[1]], order = "hclust", insig = "pch", addrect = 3)
```



```
## plot confidence interval(0.95, 0.95, 0.99), 'rect' method
corrplot(M, low = res1[[2]], upp = res1[[3]], order = "hclust", rect.col = "navy",
  plotC = "rect", cl.pos = "n")
```

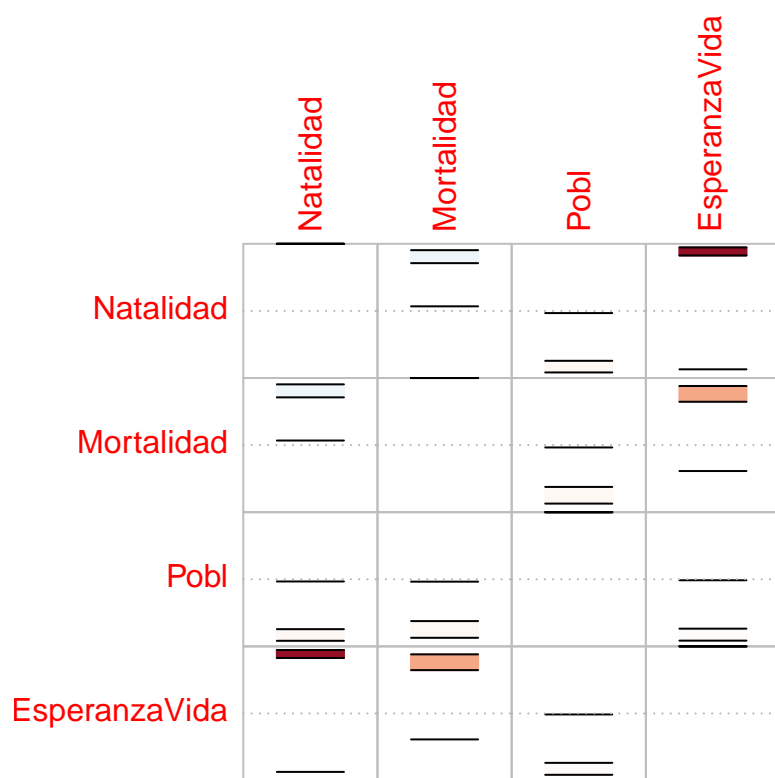


```
corrplot(M, p.mat = res1[[1]], low = res1[[2]], upp = res1[[3]], order = "hclust",
  pch.col = "red", sig.level = 0.01, addrect = 3, rect.col = "navy", plotC = "rect",
  cl.pos = "n")
```

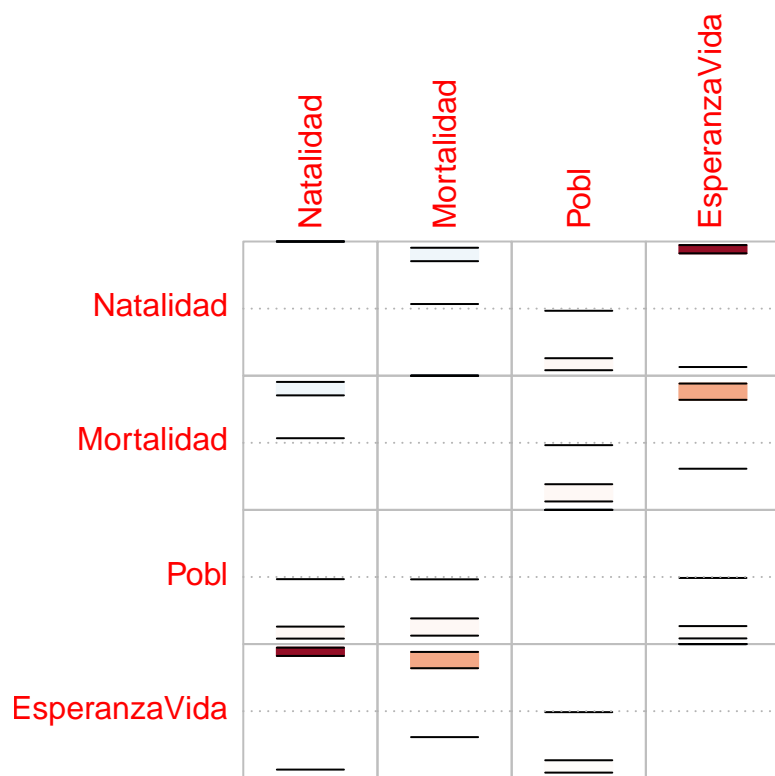


```
for (i in seq(0.1, 0, -0.005)) {
  tmp <- cor.mtest(mtcars, 1 - i)
  corrplot(M, p.mat = tmp[[1]], low = tmp[[2]], upp = tmp[[3]], order = "hclust",
    pch.col = "red", sig.level = i, plotC = "rect", cl.pos = "n", mar = c(0,
    0, 1, 0), title = subs
}
```

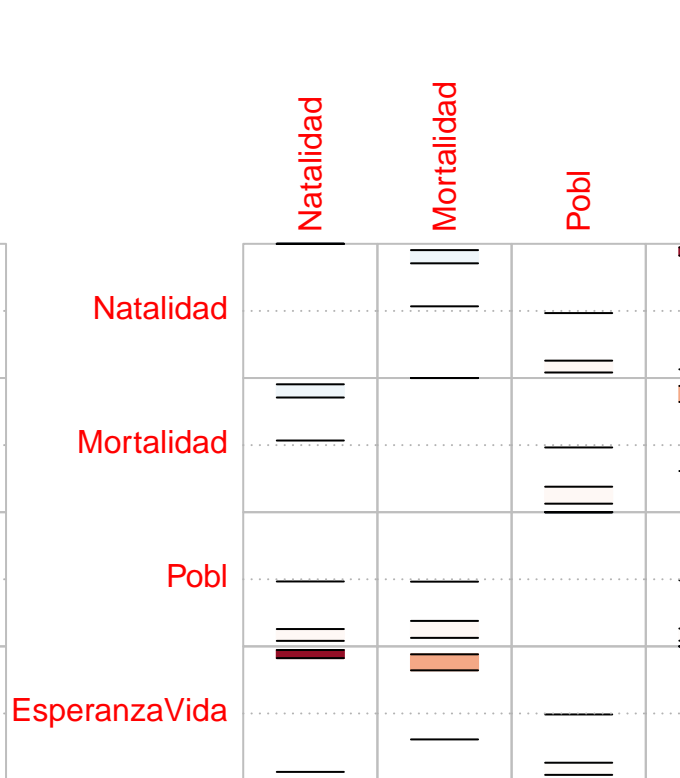
$\alpha = 0.100$



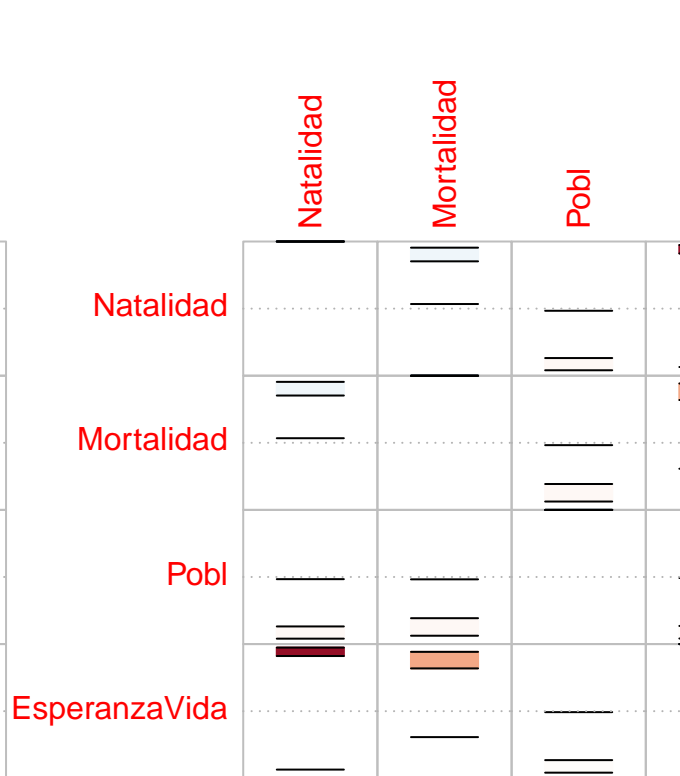
$\alpha = 0.090$



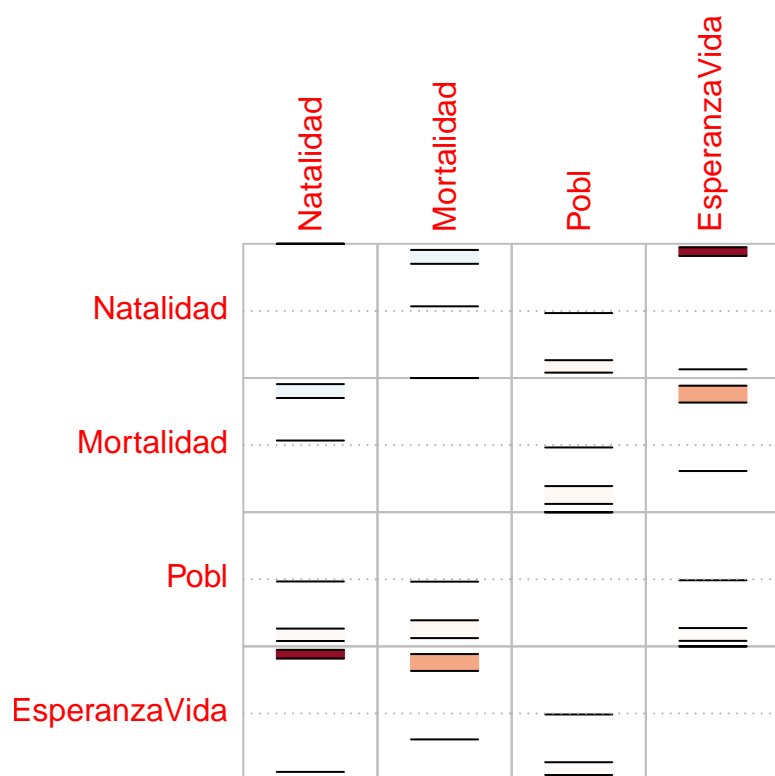
$\alpha = 0.095$



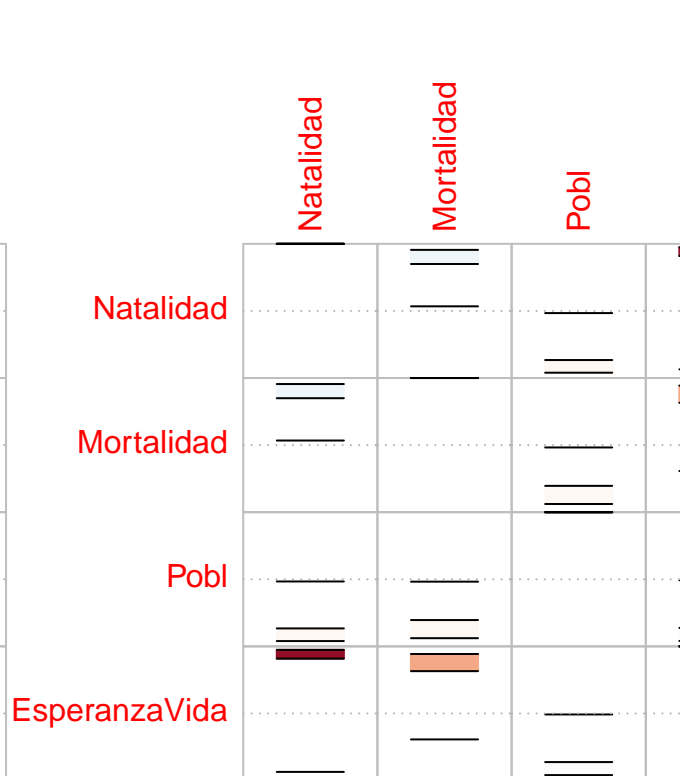
$\alpha = 0.085$



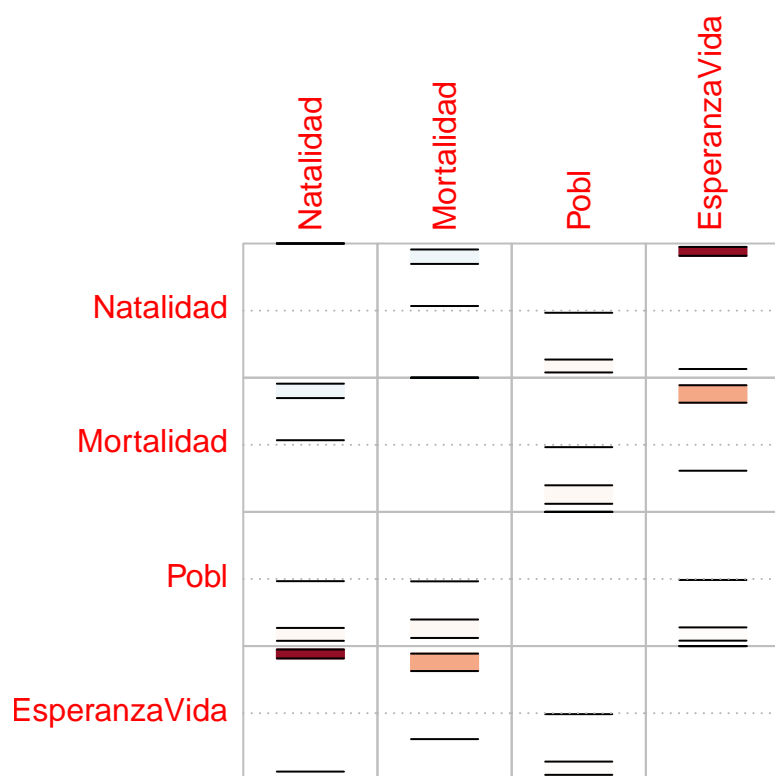
$\alpha = 0.080$



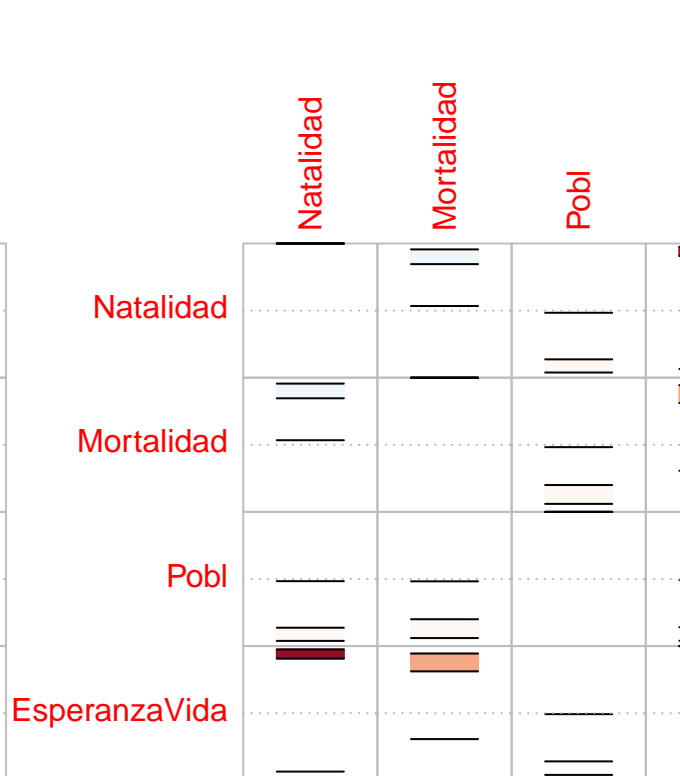
$\alpha = 0.075$



$\alpha = 0.070$

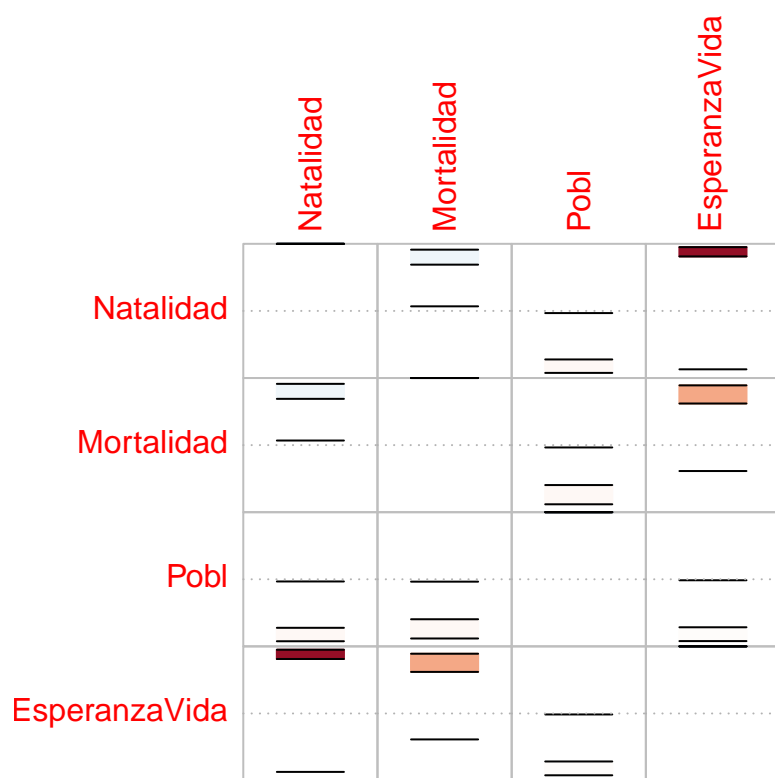


$\alpha = 0.065$

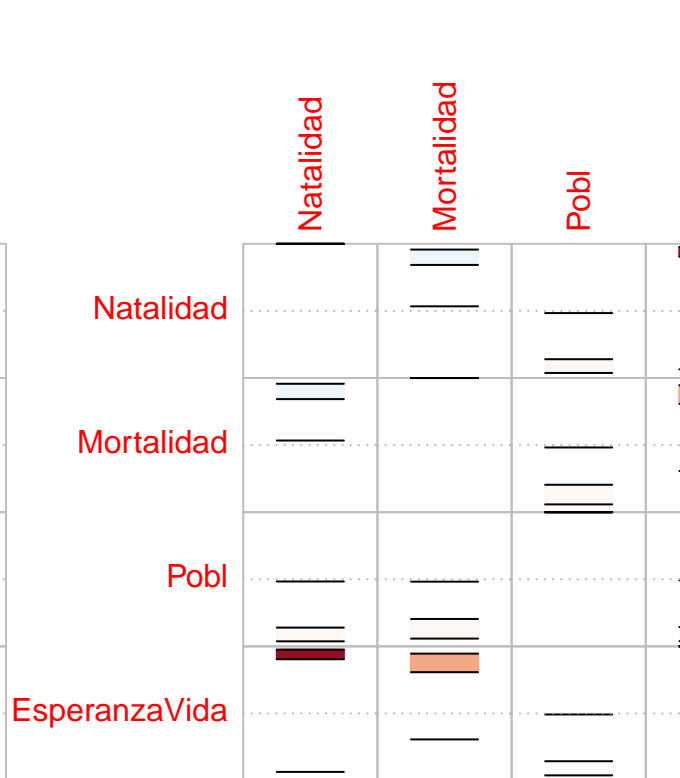




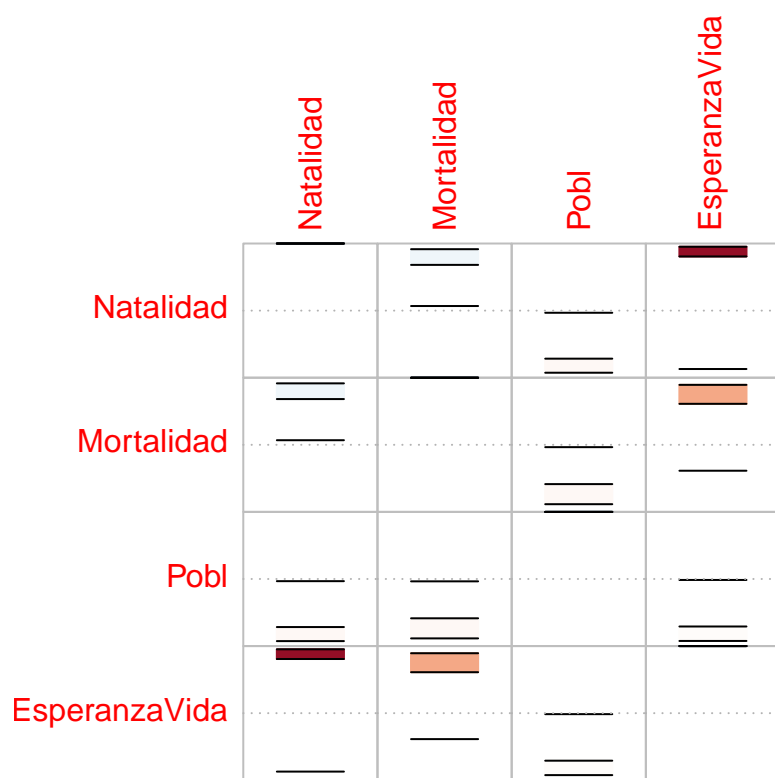
$\alpha = 0.060$



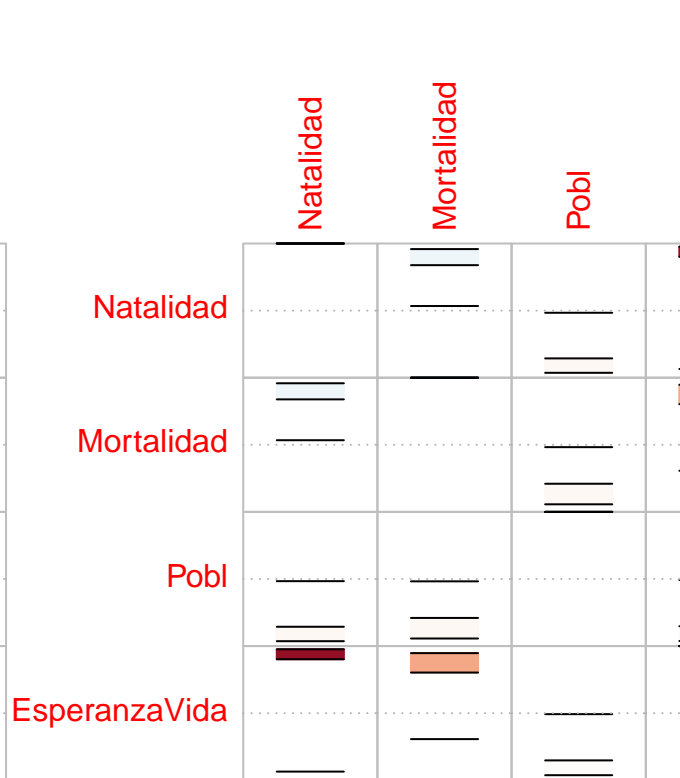
$\alpha = 0.055$



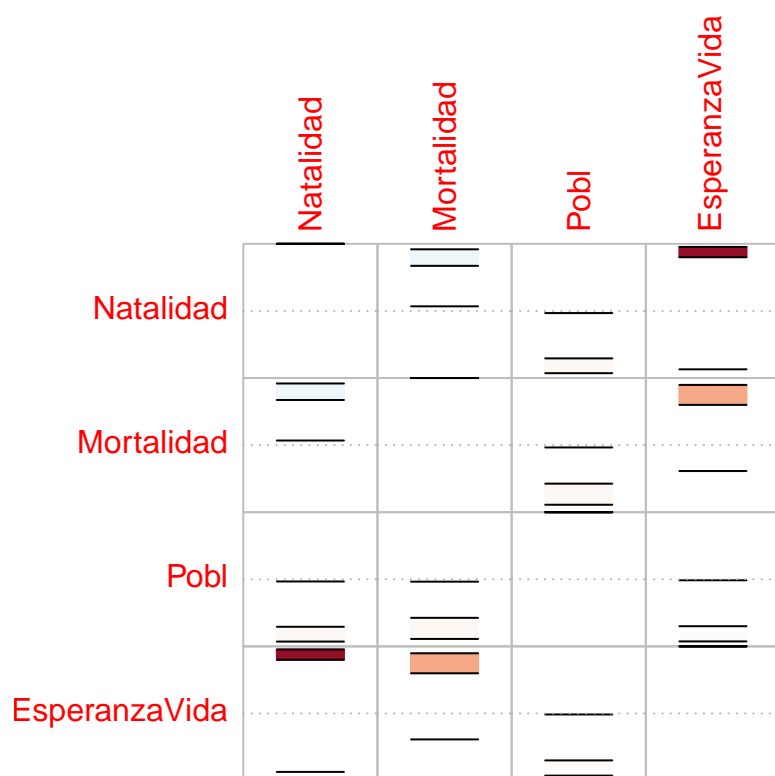
$\alpha = 0.050$



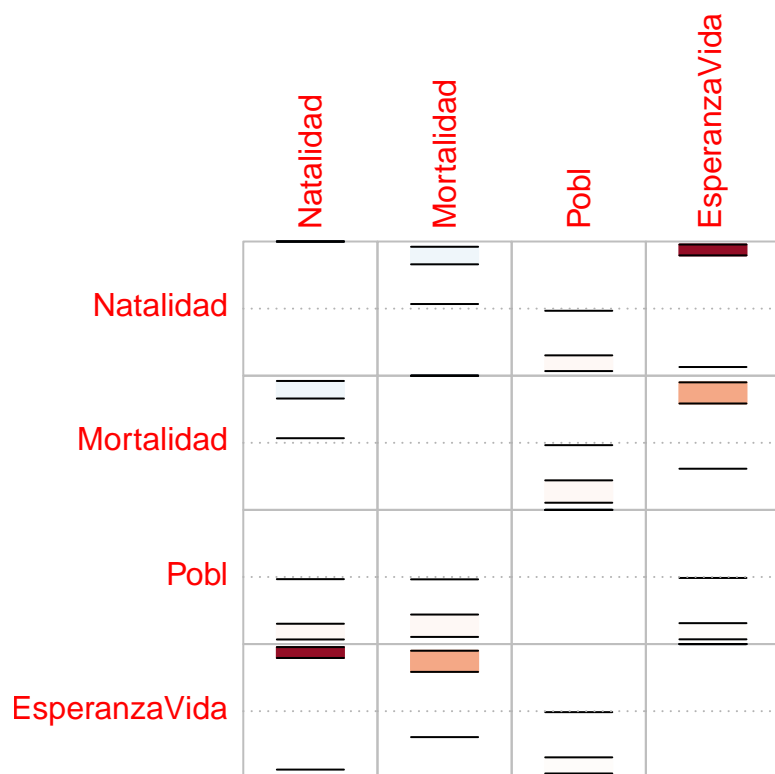
$\alpha = 0.045$



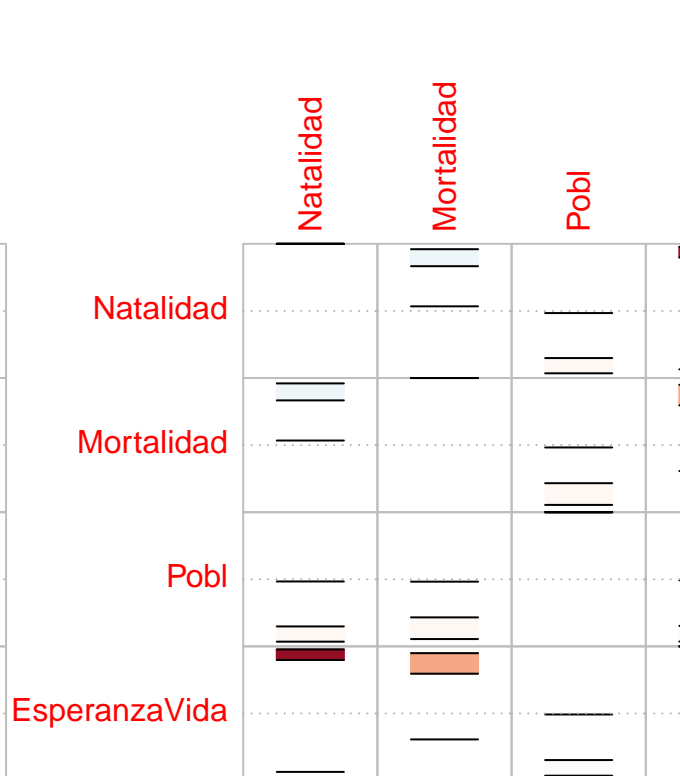
$\alpha = 0.040$



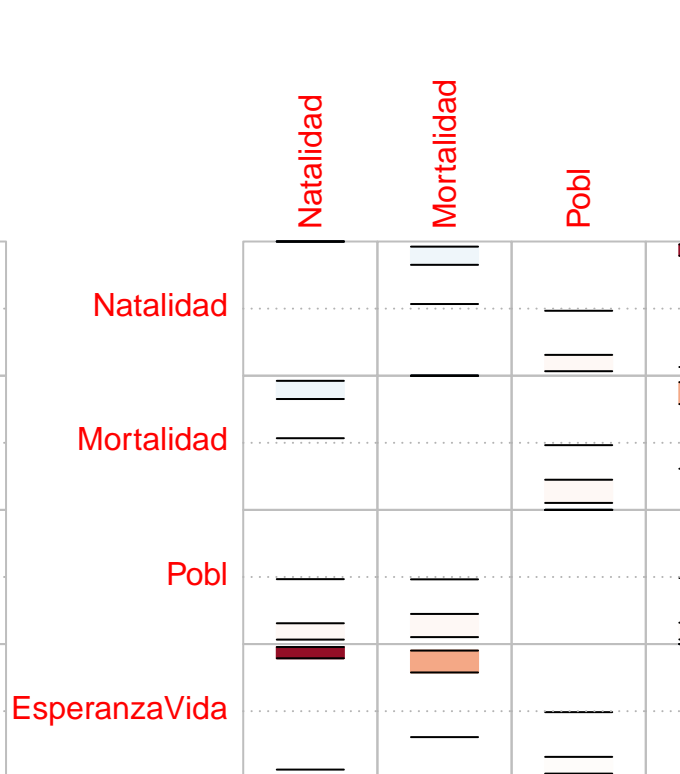
$\alpha = 0.030$



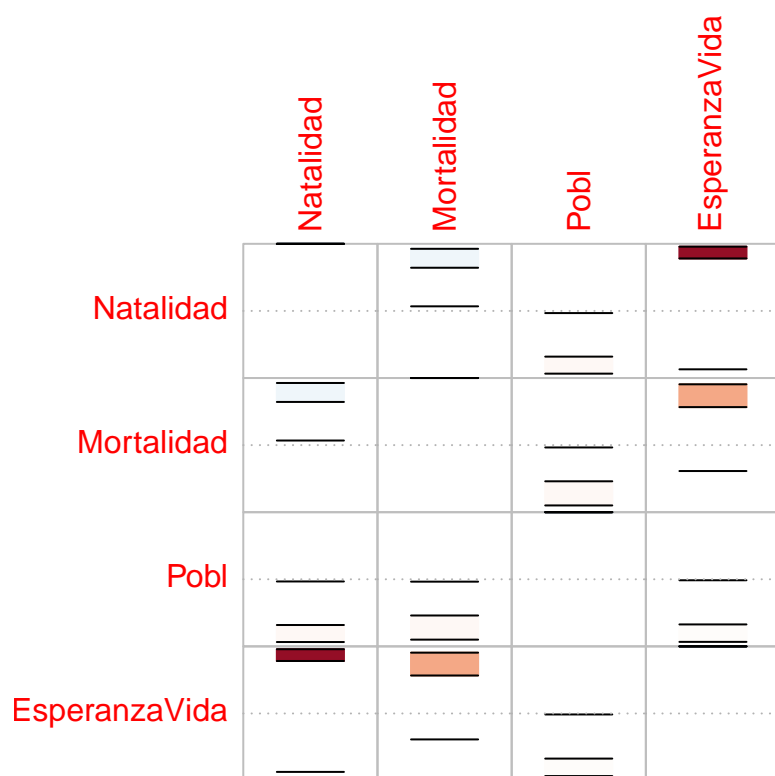
$\alpha = 0.035$



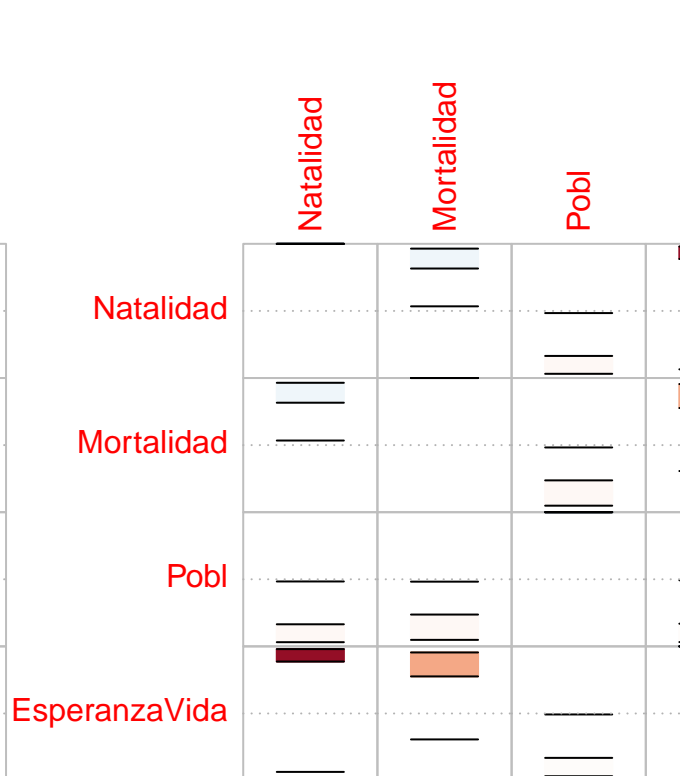
$\alpha = 0.025$



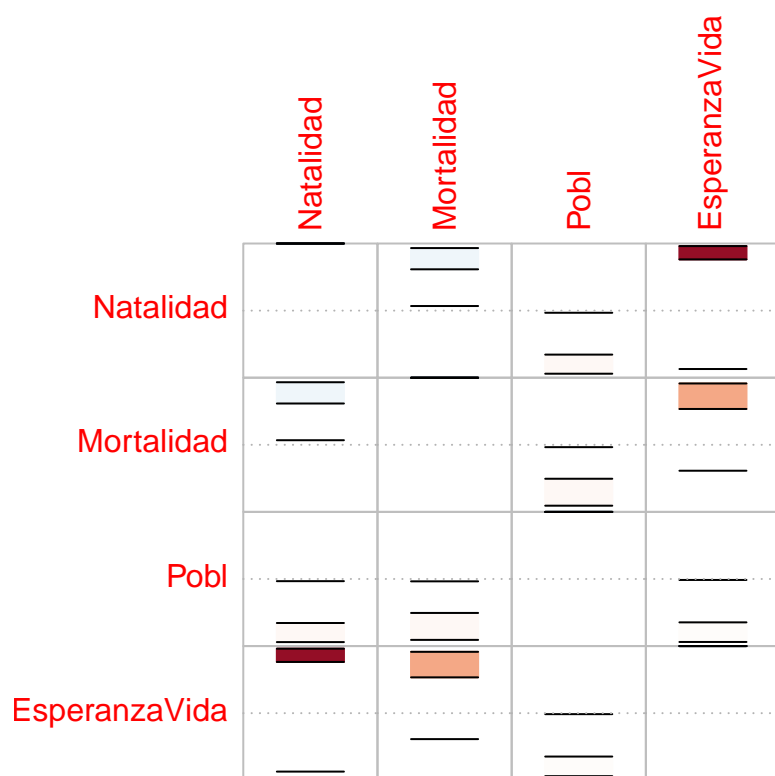
$\alpha = 0.020$



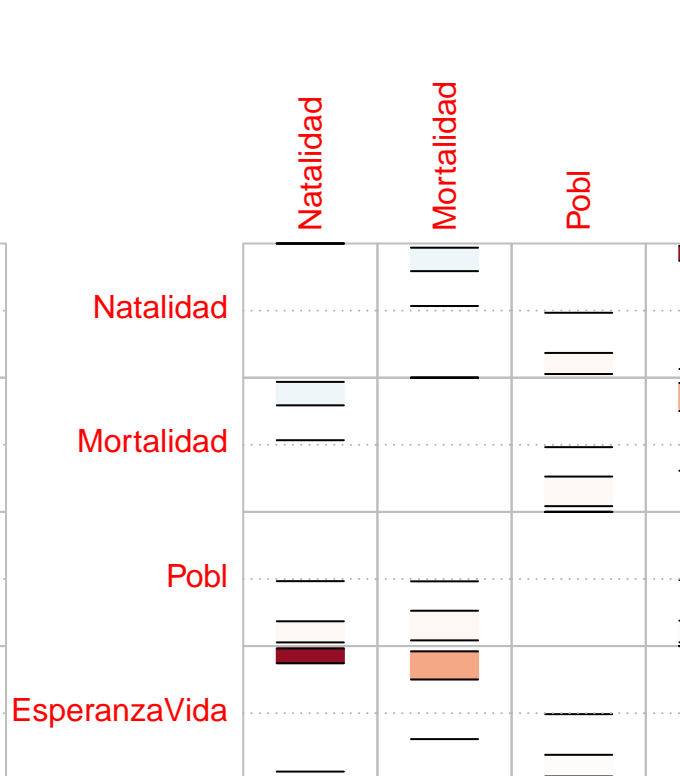
$\alpha = 0.015$



$\alpha = 0.010$



$\alpha = 0.005$



$$\alpha = 0.000$$

	Natalidad	Mortalidad	Pobl	EsperanzaVida
Natalidad		X	X	X
Mortalidad	X		X	X
Pobl	X	X		X
EsperanzaVida	X	X	X	