Cluster Analysis

Yosune Miquelajauregui/Planeacion colaborativa,LANCIS

25 de abril de 2018

****The objective of this script is to perform a cluster analysis to obtain homogenous groups as a function of biophysical and socioeconmical variables. The script provides the code to graph the results as well as to calculate and graph Gowers residuals/El objetivo de este ejercicio es clasificar las AGEBs en funcion devariables biofisicas y socio-economicas.Obtener gráficas de agrupamientos, calcular los residuales de Gower y graficarlos.**

Cluster Analysis/Análisis de grupos

Load libraries

```
library(cluster)
library(ggplot2)
library(scales)
library(reshape2)
library(ggrepel)
```

Load Data/cargar datos

```
Agebs<-
read.csv("C:\\Users\\Usuario\\Documents\\ClustersAGEB\\matriz.agebs.csv",
header=TRUE, sep=",")

#es necesario pasar datos de factor a numerico
Agebs$POBTOT <- as.numeric(Agebs$POBTOT)
Agebs$OCUP_VIV <- as.numeric(Agebs$OCUP_VIV)
Agebs$precip <- as.numeric(Agebs$precip)
#es necesario reescalar las variables para comparaci?n
Agebs.stand <- scale(Agebs[-1])
Agebs.stand <- na.omit(Agebs.stand)
```

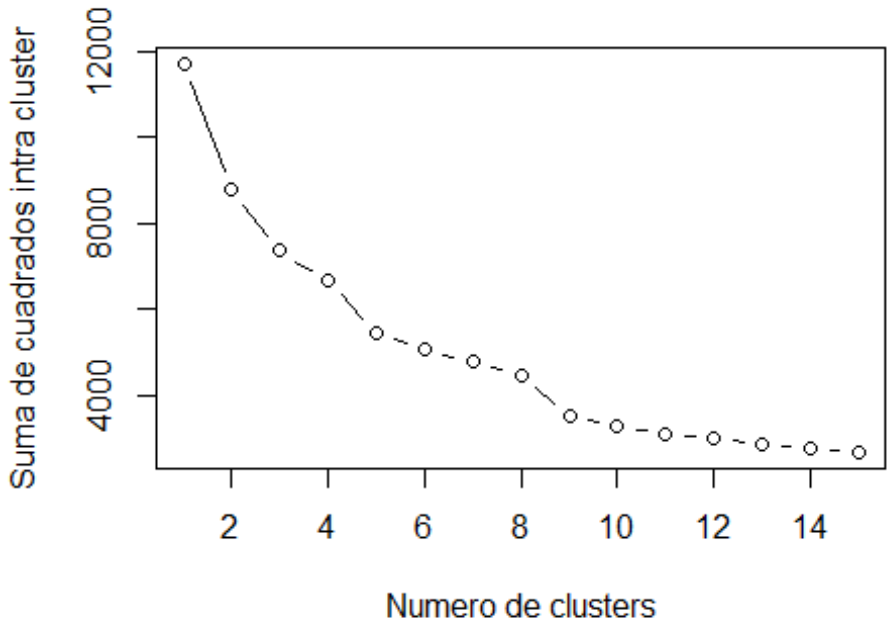
Utilizar el método K-means para generar grupos homogeneos. Se requiere especificar el numero de grupos a crear utilizando la suma de cuadrados intra cluster.

```
wssplot <- function(data, nc=15, seed=1234){
  wss <- (nrow(data)-1)*sum(apply(data,2,var))
  for (i in 2:nc){
    set.seed(seed)
    wss[i] <- sum(kmeans(data, centers=i)$withinss)}
```

```
  plot(1:nc, wss, type="b", xlab="Numero de clusters",
       ylab="Suma de cuadrados intra cluster")}
```

Graficar numero de clusters en funcion de la suma de cuadraddos intracluster

```
wssplot(Agebs.stand, nc=15)
```

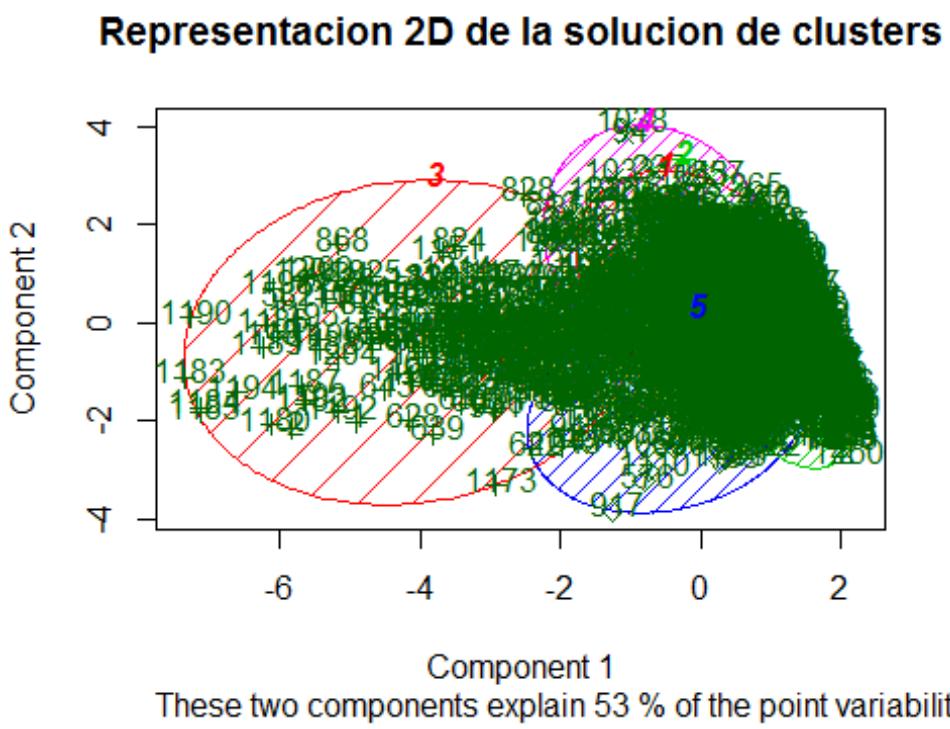


Crear grupos

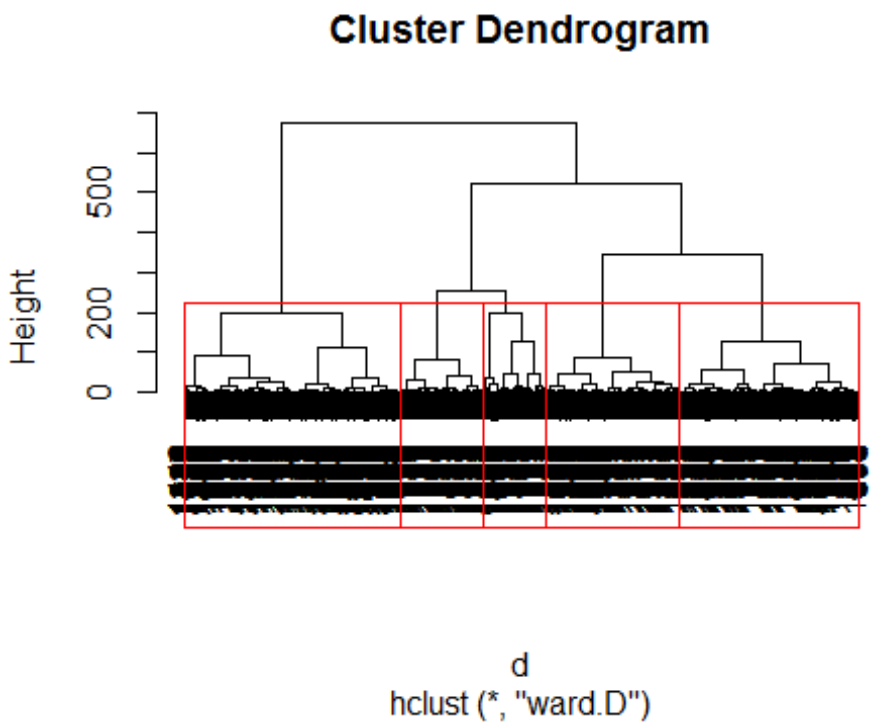
```
k.means.fit5 <- kmeans(na.omit(Agebs.stand), 5)
```

Graficar grupos

```
clusplot(Agebs.stand, k.means.fit5$cluster, main='Representacion 2D de la
solucion de clusters',
         color=TRUE, shade=TRUE,
         labels=2, lines=0)
```




```r
d <- dist(Agebs.stand, method = "euclidean") # matriz de distancia
fit <- hclust(d, method="ward.D")
plot(fit) # dendograma
groups <- cutree(fit, k=5) # crear los grupos
# graficar dendograma con bordes rojos para delimitar los grupos
rect.hclust(fit, k=5, border="red")
```



Gowers residulas/ Residuales de Gower

```
residuales_ageb5 <- read.csv(header=T, text='
Atributo,Grupo,Tipo,Valor
Urbano,1,Promedio,0.9786149163
Urbano,2,Promedio,0.1145258621
Urbano,3,Promedio,0.9669020501
Urbano,4,Promedio,0.8229090909
Urbano,5,Promedio,0.9943092105
Precip,1,Promedio,0.2581278539
Precip,2,Promedio,0.3881465517
Precip,3,Promedio,0.1746241458
Precip,4,Promedio,0.2436363636
Precip,5,Promedio,0.2875657895
Elev,1,Promedio,0.1246270928
Elev,2,Promedio,0.3107758621
Elev,3,Promedio,0.0219134396
Elev,4,Promedio,0.0638181818
Elev,5,Promedio,0.0761184211
Hund,1,Promedio,0.0214003044
Hund,2,Promedio,0.0289224138
Hund,3,Promedio,0.9540205011
Hund,4,Promedio,0.4127272727
Hund,5,Promedio,0.0195394737
POBTOT,1,Promedio,0.3008523592
POBTOT,2,Promedio,0.1743103448
POBTOT,3,Promedio,0.2138724374
```

```
POBTOT,4,Promedio,0.0138181818
POBTOT,5,Promedio,0.1332401316
V_S_AGUA,1,Promedio,0.0247488584
V_S_AGUA,2,Promedio,0.1993965517
V_S_AGUA,3,Promedio,0.009453303
V_S_AGUA,4,Promedio,0.0267272727
V_S_AGUA,5,Promedio,0.0168914474
OCUP_VIV,1,Promedio,0.6521613394
OCUP_VIV,2,Promedio,0.6825862069
OCUP_VIV,3,Promedio,0.5757403189
OCUP_VIV,4,Promedio,0.0201818182
OCUP_VIV,5,Promedio,0.5513322368
INGRESO,1,Promedio,0.7036225266
INGRESO,2,Promedio,0.6021551724
INGRESO,3,Promedio,0.7737129841
INGRESO,4,Promedio,0
INGRESO,5,Promedio,0.8720065789
Urbano,1,Residual,0.1653941556
Urbano,2,Residual,-0.6282778629
Urbano,3,Residual,0.0754207983
Urbano,4,Residual,0.1922304639
Urbano,5,Residual,0.1952324451
Precip,1,Residual,-0.0500608217
Precip,2,Residual,0.1503749118
Precip,3,Residual,-0.2118250209
Precip,4,Residual,0.1179898217
Precip,5,Residual,-0.0064788909
Elev,1,Residual,-0.0325920413
Elev,2,Residual,0.2239737636
Elev,3,Residual,-0.2135661856
Elev,4,Residual,0.0891411813
Elev,5,Residual,-0.0669567179
Hund,1,Residual,-0.3036902234
Hund,2,Residual,-0.2257510784
Hund,3,Residual,0.5506694822
Hund,4,Residual,0.2701788785
Hund,5,Residual,-0.291407059
POBTOT,1,Residual,0.0958651336
POBTOT,2,Residual,0.0397401549
POBTOT,3,Residual,-0.0693752794
POBTOT,4,Residual,-0.0086269102
POBTOT,5,Residual,-0.0576030989
V_S_AGUA,1,Residual,-0.0684631629
V_S_AGUA,2,Residual,0.1766015661
V_S_AGUA,3,Residual,-0.1620192095
V_S_AGUA,4,Residual,0.1160573851
V_S_AGUA,5,Residual,-0.0621765788
OCUP_VIV,1,Residual,0.1179924207
OCUP_VIV,2,Residual,0.2188343239
OCUP_VIV,3,Residual,-0.0366890909
```

```
OCUP_VIV,4,Residual,-0.3314449669
OCUP_VIV,5,Residual,0.0313073133
INGRESO,1,Residual,0.0755545395
INGRESO,2,Residual,0.044504221
INGRESO,3,Residual,0.0673845058
INGRESO,4,Residual,-0.4455258534
INGRESO,5,Residual,0.2580825871
')

residuales_ageb5$Grupo <- factor(residuales_ageb5$Grupo)
```

Generate point graphs/Generar una gráfica de puntos

Final graph showing Gowers residuals for 5 groups

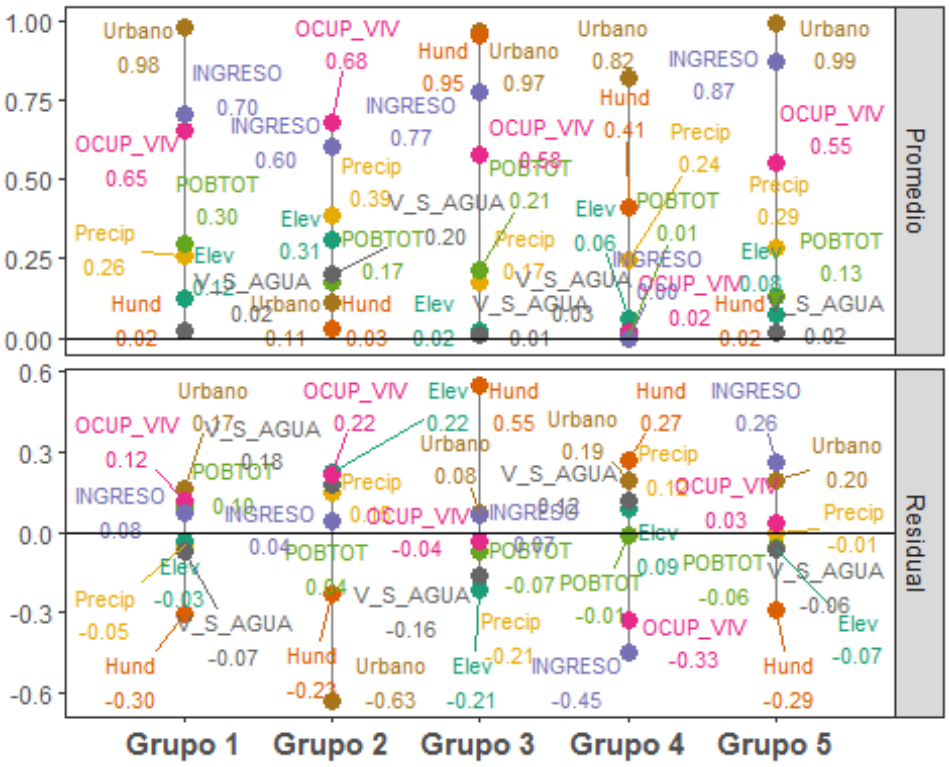
```
ggplot(data=residuales_ageb5, aes(x=Grupo, y=Valor, group=Atributo,
color=Atributo)) +
  geom_segment(aes(xend=Grupo),yend=0,colour="grey50",linetype="solid") +
  geom_point(size=3,show.legend=FALSE) +
  geom_text_repel(aes(label = paste(Atributo, sprintf("%.2f",
Valor),sep="\n")), size=3, family="Calibri", show.legend=FALSE)+
  geom_hline(aes(yintercept=0), linetype="solid") +
  scale_color_brewer(palette="Dark2") +
  scale_x_discrete(labels=c(
    "1"="Grupo 1",
    "2"="Grupo 2",
    "3"="Grupo 3",
    "4"="Grupo 4",
    "5"="Grupo 5"), expand=c(0.13,0.8)) +
  scale_y_continuous() +
  theme_bw() +
  theme(
    text=element_text(family="Calibri"),
    panel.grid.minor=element_blank(),
    panel.grid.major=element_blank(),
    axis.title.x=element_blank(),
    axis.title.y=element_blank(),
    axis.text.x=element_text(size=11,face="bold", hjust=0.5))
```

Separate mean and residualsSepara promedios de residuales

```
graf + facet_grid(Tipo ~ ., scales="free")
```

```
## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database
```

```
## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x,
x
## $y, : font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x,
x
## $y, : font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database
```



Name each part of the graphs as mean and residual/nombrar cada una de las gráficas

```
tipo_etiquetas <- c("Promedio" = "PROMEDIO",
                    "Residual" = "RESIDUAL")
```

Apply to y axis/Se aplica formato al eje Y

```
graf + facet_grid(Tipo ~ ., scales="free", labeller =
as_labeller(tipo_etiquetas)) +
  theme(strip.text.y=element_text(size=14, face="bold"))
```

```
## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database
```

```
## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x,
x
## $y, : font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x,
x
## $y, : font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database

## Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x,
x$y, :
## font family not found in Windows font database
```