Trabajo Colaborativo 4

Leonidas Fiquitiva Castro

13 de septiembre de 2017

## Proyecto colaborativo cuarta etapa

Para finalizar nuestro proyecto de curso, a continuación encontrará las actividades de la última etapa, las cuales debe desarrollar junto con su grupo de trabajo.

Tenga en cuenta que para esta entrega se espera que usted y su grupo utilicen las diferentes herramientas vistas en el curso, como filtros, agregaciones, recodificaciones, integración de bases de datos (merge y bind), además de la elaboración de visualizaciones avanzadas.

Usando las bases de datos con las que han venido trabajando, realicen un detallado análisis exploratorio de datos, que les permita dar cumplimiento a su objetivo de investigación.

Se revisarán los siguientes aspectos:

1. Uso conveniente de las funciones de recodificación de variables para obtener salidas y gráficos con resultados más legibles.

{r Leonidas}

library(readr)

Base\_Hospitales <- read.csv("BaseHospitales.csv", header = T, sep = ";") dato<-(Base\_Hospitales$Número.de.personas.afiliadas.al.régimen.subsidiado) table(dato)

library(vegan)

attach(dato) dato[dato %in% 1:30000] <- "Bajo" dato[dato %in% 300000:1000000] <- "Medio" dato[dato %in% 1000000:2000000] <- "Alto" dato[is.na(dato)] <- "No sabe / No responde" detach(dato)

library(car) estR <- Recode(Base\_Hospitales$Número.de.personas.afiliadas.al.régimen.subsidiado, "1:200000='A'; 200000:600000='B';600000:6000000='C', NA='D")

table(estR)

b. Integración de las diferentes tablas de su proyecto, utilizando (merge, bind, etc.).  
  
  
  
  
datos1<-c(Base\_Hospitales$Código.del.municipio,Base\_Hospitales$Municipio)  
datos2<-c(Base\_Hospitales$regionandina)  
datos2  
df\_left <- merge(x = datos1, y = datos2, by =  
"IdCliente",  
+ all.x = TRUE)  
df\_left

1. Desarrollo de al menos 3 consultas que involucren filtros (utilizando los operadores lógicos & |). Para ello, usen filtros relevantes para la resolución de su problema de investigación.

Abejorral <- subset(Base\_Hospitales, Municipio == "Abejorral") dim(Abejorral) Abejorral

departamento<- subset(Base\_Hospitales, Departamento == "Boyacá") dim(departamento) departamento

departamento<- subset(Base\_Hospitales, Departamento == "Boyacá") dim(departamento) departamento

Clinicas<- subset(Base\_Hospitales, regionandina == 1) dim(Clinicas) Clinicas

``` d. Desarrollo de al menos 5 agregaciones por diferentes variables categóricas (utilizando medidas como los percentiles, promedio, desviación, coeficiente de variación, simetría, curtosis, entre otras) que les permitan realizar un análisis desagregado para cumplir los objetivos de su proyecto. Por ejemplo, si están realizando un análisis de ingresos empresariales, hagan agregaciones por sedes.

1. Desarrollo de análisis multivariados (cálculo de correlaciones, tablas de contingencias, visualizaciones de diagramas de dispersión con variables agrupadas, entre otras).
2. Uso de información externa de otra base u otro módulo, que les permita complementar sus análisis integrando esa información con las bases de datos que han venido trabajando. Por ejemplo, si están haciendo un estudio de ingresos empresariales por municipios, busquen información de la población de municipios y péguenla a la base de datos, utilizando funciones como merge.
3. Desarrollo de diferentes visualizaciones que involucren gráficos realizados en ggplot2 o el paquete básico (pueden realizar gráficos de densidad para diferentes subpoblaciones, boxplots, histogramas, etc.).
4. Creatividad en las visualizaciones y estética de las mismas.
5. Explicaciones claras de los resultados obtenidos.
6. Calidad de las conclusiones, en relación con el objetivo del proyecto.