CÓDIGO R

knitr::opts\_chunk$set(echo = T)  
knitr::opts\_chunk$set(echo = T)

library(readxl)  
library(ggplot2)  
library(dplyr)

library(fBasics)

library(MASS)

library(corrgram)  
library(gclus)

library(descr)  
library(GGally)

library(agricolae)

library(plotly)

library(fdth)

library(caTools)  
library(ROCR)  
library(pROC)

library(car)

library(foreign)  
library(apaTables)  
library(PerformanceAnalytics)

library(psych)

library(corrr)  
library(igraph)

library(corrgram)  
library(gclus)  
library(GGally)  
library(gvlma)

library(readxl)  
BDa2censo <- read\_excel("BD inversionistas depurada V6 anonimizada.xlsx")

# ANÁLISIS ESTADÍSTICO BASE DE DATOS INVERSIONISTAS DE A2CENSO

## Resumen univariado de variables cualitativas

### Variable 1: SectorCampaña

SectorCampaña\_= BDa2censo$SectorCampaña  
SCtabla=data.frame(table(SectorCampaña\_))  
porcentaje=prop.table(SCtabla[,2])  
SCtabla2= cbind(SCtabla, porcentaje)  
cum\_frequencia=cumsum(SCtabla2[,2])  
SCtabla3= cbind(SCtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(SCtabla3[,3])  
SCtabla4= cbind(SCtabla3, cum\_porcentaje)  
knitr::kable(  
 (SCtabla4)  
)

barplot(table(BDa2censo$SectorCampaña), col = "black", main="Diagrama de barras variable Sector Campaña",ylim=c(0,4500),las=2, cex.names = 0.32, lwd = 1.5, cex.axis=0.9 , ylab = "Frecuencias Absolutas")

t <- list(  
family = "Ariel",  
size = 7,  
color = 'black')  
  
#plot\_ly(SCtabla4, labels = SCtabla4[,1], values = SCtabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por Sector Campaña",autosize = T, font = t)

### 

### Variable 2: Campaña

Campaña= BDa2censo$Campaña  
CPtabla=data.frame(table(Campaña))  
porcentaje=prop.table(CPtabla[,2])  
CPtabla2= cbind(CPtabla, porcentaje)  
cum\_frequencia=cumsum(CPtabla2[,2])  
CPtabla3= cbind(CPtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(CPtabla3[,3])  
CPtabla4= cbind(CPtabla3, cum\_porcentaje)  
CPtabla4o <- CPtabla4[order(-CPtabla4$Freq),]  
knitr::kable(  
 (CPtabla4o)  
)

ggplot(CPtabla4o, aes(x=Campaña, y=Freq)) +   
 geom\_bar(stat = "identity") + coord\_flip()

barplot(table(CPtabla4o$Freq), col = "blue", main="Diagrama de barras",ylim=c(0,2000),las=2, cex.names = 0.5, lwd = 1.5, cex.axis=0.9, xlab="Campaña", ylab = "Conteos")

t <- list(  
family = "Ariel",  
size = 7,  
color = 'black')  
  
#plot\_ly(CPtabla4, labels = CPtabla4o[,1], values = CPtabla4o[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por Campaña", font = t)

### Variable 3: Departamento

Departamento= BDa2censo$Departamento  
Dotabla=data.frame(table(Departamento))  
porcentaje=prop.table(Dotabla[,2])  
Dotabla2= cbind(Dotabla, porcentaje)  
cum\_frequencia=cumsum(Dotabla2[,2])  
Dotabla3= cbind(Dotabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Dotabla3[,3])  
Dotabla4= cbind(Dotabla3, cum\_porcentaje)  
Dotabla4o <- Dotabla4[order(-Dotabla4$Freq),]  
knitr::kable(  
 (Dotabla4o)  
)

barplot(table(BDa2censo$Departamento), col = "black", main="Cantidad de inversionistas por departamento",ylim=c(0,16000),las=2, cex.names = 0.38, lwd = 1, cex.axis=0.6, xlab="Departamento", ylab = "Conteos")

t <- list(  
family = "Ariel",  
size = 7,  
color = 'black')  
  
#plot\_ly(Dotabla4, labels = Dotabla4[,1], values = Dotabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por Departamento", font = t)

### 

### Variable 4: Sector Económico del inversionista

SectorEconomico= BDa2censo$SectorEconomico  
SEtabla=data.frame(table(SectorEconomico))  
porcentaje=prop.table(SEtabla[,2])  
SEtabla2= cbind(SEtabla, porcentaje)  
cum\_frequencia=cumsum(SEtabla2[,2])  
SEtabla3= cbind(SEtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(SEtabla3[,3])  
SEtabla4= cbind(SEtabla3, cum\_porcentaje)  
SEtabla4o <- SEtabla4[order(-SEtabla4$Freq),]  
knitr::kable(  
 (SEtabla4o)  
)

ggplot(SEtabla4o, aes(x=SectorEconomico, y=Freq)) +   
 geom\_bar(stat = "identity") + ggtitle("Numero de inversionistas por Sector Económico") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) +   
 xlab("Sector Económico")+ ylab("Cantidad de inversionistas") + coord\_flip()+theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 5,  
color = 'black')  
  
#plot\_ly(SEtabla4, labels = SEtabla4[,1], values = SEtabla4[,3], type = "pie",autosize = T) %>%  
# layout(title = "Diagrama de sectores por Sector Económico del Inversionista",autosize = T, font=t)#, #width = 450, height = 450, margin = m)

### Variable 5: Tipo Inversionista

TipoInversionista= BDa2censo$TipoInversionista  
TItabla=data.frame(table(TipoInversionista))  
porcentaje=prop.table(TItabla[,2])  
TItabla2= cbind(TItabla, porcentaje)  
cum\_frequencia=cumsum(TItabla2[,2])  
TItabla3= cbind(TItabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(TItabla3[,3])  
TItabla4= cbind(TItabla3, cum\_porcentaje)  
TItabla4 <- TItabla4[order(-TItabla4$Freq),]  
knitr::kable(  
 (TItabla4)  
)

ggplot(TItabla4, aes(x=TipoInversionista, y=Freq)) +   
 geom\_bar(stat = "identity") + ggtitle("Numero de inversionistas por tipo de inversionistas") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) +   
 xlab("Tipo de inversionista")+ ylab("Conteo") +theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 10,  
color = 'black')  
  
#plot\_ly(TItabla4, labels = TItabla4[,1], values = TItabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por Tipo Inversionista", font = t)

### 

### Variable 6: Propósito del inversionista en A2censo

Proposito = BDa2censo$Propósito  
PTtabla=data.frame(table(Proposito))  
porcentaje=prop.table(PTtabla[,2])  
PTtabla2= cbind(PTtabla, porcentaje)  
cum\_frequencia=cumsum(PTtabla2[,2])  
PTtabla3= cbind(PTtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(PTtabla3[,3])  
PTtabla4= cbind(PTtabla3, cum\_porcentaje)  
PTtabla4 <- PTtabla4[order(-PTtabla4$Freq),]  
knitr::kable(  
 (PTtabla4)  
)

ggplot(PTtabla4, aes(x=Proposito, y=Freq)) +   
 geom\_bar(stat = "identity") + ggtitle("Numero de inversionistas por Propósito") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) + coord\_flip() +  
 xlab("Propósito del inversionista")+ ylab("Conteo") +theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 10,  
color = 'black')  
  
  
#plot\_ly(PTtabla4, labels = PTtabla4[,1], values = PTtabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por propósito del inversionista en A2censo", font = t)

### 

### Variable 7: Grupo de Edad

GrupoEdad = BDa2censo$GrupoEdad  
GEtabla=data.frame(table(GrupoEdad))  
porcentaje=prop.table(GEtabla[,2])  
GEtabla2= cbind(GEtabla, porcentaje)  
cum\_frequencia=cumsum(GEtabla2[,2])  
GEtabla3= cbind(GEtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(GEtabla3[,3])  
GEtabla4= cbind(GEtabla3, cum\_porcentaje)  
GEtabla4 <- GEtabla4[order(-GEtabla4$Freq),]  
knitr::kable(  
 (GEtabla4)  
)

ggplot(GEtabla4, aes(x=GrupoEdad, y=Freq)) +   
 geom\_bar(stat = "identity") + coord\_flip() + ggtitle("Titulo grafico") + ggtitle("Numero de inversionistas por Rango de edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) +  
 xlab("Rango de edad")+ ylab("Conteo") +theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 10,  
color = 'black')  
  
#plot\_ly(GEtabla4, labels = GEtabla4[,1], values = GEtabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por rango de Edad del inversionista", font = t)

### Variable 8: Grupo de Ingresos

GrupoIngresos = BDa2censo$GrupoIngresos  
GItabla=data.frame(table(GrupoIngresos))  
porcentaje=prop.table(GItabla[,2])  
GItabla2= cbind(GItabla, porcentaje)  
cum\_frequencia=cumsum(GItabla2[,2])  
GItabla3= cbind(GItabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(GItabla3[,3])  
GItabla4= cbind(GItabla3, cum\_porcentaje)  
GItabla4 <- GItabla4[order(-GItabla4$Freq),]  
knitr::kable(  
 (GItabla4)  
)

ggplot(GItabla4, aes(x=GrupoIngresos, y=Freq)) +   
 geom\_bar(stat = "identity") + coord\_flip() + ggtitle("Numero de inversionistas por rango de ingresos") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) +  
 xlab("Rango de ingresos")+ ylab("Conteo") +theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 10,  
color = 'black')  
  
  
#plot\_ly(GItabla4, labels = GItabla4[,1], values = GItabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores por rango de Ingresos", font = t)

### Variable 9: Inversionistas recurrentes

InvRecurr = BDa2censo$CodEsInversionistaRecurrente  
Rtabla=data.frame(table(InvRecurr))  
porcentaje=prop.table(Rtabla[,2])  
Rtabla2= cbind(Rtabla, porcentaje)  
cum\_frequencia=cumsum(Rtabla2[,2])  
Rtabla3= cbind(Rtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Rtabla3[,3])  
Rtabla4= cbind(Rtabla3, cum\_porcentaje)  
Rtabla4 <- Rtabla4[order(-Rtabla4$Freq),]  
knitr::kable(  
 (Rtabla4)  
)

ggplot(Rtabla4, aes(x=InvRecurr, y=Freq)) +   
 geom\_bar(stat = "identity") + coord\_flip() + ggtitle("Numero de inversionistas por recurrencia de inversión") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic")) +  
 xlab("Recurrencia del inversionista")+ ylab("Conteo") +theme(axis.text=element\_text(size=5,face="bold"),  
 axis.title=element\_text(size=6,face="bold"))

t <- list(  
family = "Ariel",  
size = 10,  
color = 'black')  
  
  
  
#plot\_ly(Rtabla4, labels = Rtabla4[,1], values = Rtabla4[,3], type = "pie") %>%  
# layout(title = "Diagrama de sectores ¿Es inversionista recurrente", font = t)

## 

## Resumen univariado de variables cuantitativas

### Variable 10: Ingresos Mes

basicStats(BDa2censo$IngresosMes)

IMtabla\_Sturges <- fdt(BDa2censo$IngresosMes,breaks="Sturges",right=F, na.rm=TRUE)  
  
  
knitr::kable(  
 (IMtabla\_Sturges)  
)

plot(IMtabla\_Sturges,type='fh',main="Histograma de frecuencias absolutas Ingresos Mes", cex.main = 1 ,las=2, cex.names = 0.35, lwd = 1, cex.axis=0.45,  
 xlab="Ingresos mes",ylab="Conteos",col="black", las=2)

plot(IMtabla\_Sturges,type='fp',main="Poligono de frecuencias absolutas Ingresos Mes",cex.main = 1, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab="Ingresos mes",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=IngresosMes)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("Ingresos Mes") + coord\_flip() + ggtitle("Boxplot de ingresos mes") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=IngresosMes)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Ingresos Mes")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Diagrama de ojiva de Ingresos mes") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

### Variable 11: Patrimonio

basicStats(BDa2censo$Patrimonio)

POtabla\_St <- fdt(BDa2censo$Patrimonio,breaks="Sturges",right=F)  
  
knitr::kable(  
 (POtabla\_St)  
)

plot(POtabla\_St,type='fh',main="Histograma de frecuencias absolutas variable Patrimonio",cex.main = 1 ,las=2, cex.names = 0.35, lwd = 1, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="blue", las=2)

plot(POtabla\_St,type='fp',main="Poligono de frecuencias absolutas de Patrimonio", cex.main = 1, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=Patrimonio)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("Patrimonio") + coord\_flip() + ggtitle("Boxplot de Patrimonio") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=Patrimonio)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Patrimonio")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Boxplot de Patrimonio") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

### Variable 14: Monto Inversión

basicStats(BDa2censo$MontoInversion)

MItabla\_st <- fdt(BDa2censo$MontoInversion,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (MItabla\_st)  
)

plot(MItabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Monto Inversión",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="blue", las=2)

plot(MItabla\_st,type='fp',main="Poligono de frecuencias absolutas",cex.main = 1, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=MontoInversion)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("Monto invertido") + coord\_flip() + ggtitle("Boxplot de Monto invertido") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=MontoInversion)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Monto invertido")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Ojiva de Monto invertido") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

### Variable: Edad

basicStats(BDa2censo$Edad)

Edtabla\_st <- fdt(BDa2censo$Edad,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (Edtabla\_st)  
)

plot(Edtabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Edad",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black", las=2)

plot(Edtabla\_st,type='fp',main="Poligono de frecuencias absolutas Variable Edad",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=Edad)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("Edad") + coord\_flip() + ggtitle("Boxplot de Edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=Edad)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Edad")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Ojiva de Edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

### 

### Variable 15 : Cantidad Inversiones

basicStats(BDa2censo$CantidadInversiones)

CItabla\_st <- fdt(BDa2censo$CantidadInversiones,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (CItabla\_st)  
)

plot(CItabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Cantidad de inversiones",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="blue", las=2)

plot(CItabla\_st,type='fp',main="Poligono de frecuencias absolutas Variable Cantidad de inversiones",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=CantidadInversiones)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("Cantidad inversiones") + coord\_flip() + ggtitle("Boxplot de cantidad de inversiones") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=CantidadInversiones)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Cantidad inversiones")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Ojiva de Cantidad de inversiones") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

### Variable : Tasa

basicStats(BDa2censo$Tasa)

CItabla\_st <- fdt(BDa2censo$Tasa,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (CItabla\_st)  
)

plot(CItabla\_st,type='fh',main="Histograma de frecuencias absolutas variable tasa",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="blue", las=2)

plot(CItabla\_st,type='fp',main="Poligono de frecuencias absolutas Variable tasa",cex.main = 0.5, las=2, cex.names = 0.35, cex.axis=0.45,  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

ggplot(BDa2censo, aes( y=Tasa)) +   
 geom\_boxplot(fill="black", alpha=0.2) +   
 ylab("tasa") + coord\_flip() + ggtitle("Boxplot de tasa de adjudicacion") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo, aes( y=Tasa)) +   
 geom\_freqpoly(fill="black", alpha=0.2) +   
 ylab("Tasa de inversion adjudicada")+xlab("Frecuencia absoluta") + coord\_flip() + ggtitle("Ojiva de tasa adjudicada") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

## Resumen de medidas de localización

# Media Geométrica  
#Función  
geometric<-function(x) exp(sum(log(x))/length(x))  
  
#-------------------------------------------------------------------------------  
  
# Media Armónica  
#Función  
armonic<-function(x) 1/mean(1/x)  
  
  
#-------------------------------------------------------------------------------  
  
#Media Recortada en 100α% :  
#Función:  
recortada<-function(x) mean(x, trim=0.1)  
  
  
#-------------------------------------------------------------------------------  
  
#Media Winsorizada:  
library(psych)#Paquete útil para calcular la media Winsorizada:  
#Funcion  
winsorizada<-function(x) winsor.mean(x, trim= 0.1, na.rm = TRUE)  
  
#--------------------------------------------------------------------------------  
#Cálculo Resumen de medidas de localización  
  
options(scipen = 999)  
ResumenVaribles<- c("Edad","Tasa","Ingresos Mes", "Patrimonio","Monto Inversion","Cantidad Inversiones" )  
ResumenPromedio<- c(mean(BDa2censo$Edad),mean(BDa2censo$Tasa),mean(BDa2censo$IngresosMes),mean(BDa2censo$Patrimonio), mean(BDa2censo$MontoInversion),mean(BDa2censo$CantidadInversiones))  
ResumenMediana<- c(median(BDa2censo$Edad),median(BDa2censo$Tasa),median(BDa2censo$IngresosMes),median(BDa2censo$Patrimonio), median(BDa2censo$MontoInversion),median(BDa2censo$CantidadInversiones))  
ResumenGeometrica<- c(geometric(BDa2censo$Edad),geometric(BDa2censo$Tasa),geometric(BDa2censo$IngresosMes),geometric(BDa2censo$Patrimonio), geometric(BDa2censo$MontoInversion),geometric(BDa2censo$CantidadInversiones))  
ResumenArmonica<- c(armonic(BDa2censo$Edad),armonic(BDa2censo$Tasa),armonic(BDa2censo$IngresosMes),armonic(BDa2censo$Patrimonio), armonic(BDa2censo$MontoInversion),armonic(BDa2censo$CantidadInversiones))  
ResumenRecortada<- c(recortada(BDa2censo$Edad),recortada(BDa2censo$Tasa),recortada(BDa2censo$IngresosMes),recortada(BDa2censo$Patrimonio), recortada(BDa2censo$MontoInversion),recortada(BDa2censo$CantidadInversiones))  
ResumenWinsorizada<- c(winsorizada(BDa2censo$Edad),winsorizada(BDa2censo$Tasa),winsorizada(BDa2censo$IngresosMes),winsorizada(BDa2censo$Patrimonio), winsorizada(BDa2censo$MontoInversion),winsorizada(BDa2censo$CantidadInversiones))  
  
resumenMedia <- data.frame(ResumenVaribles ,ResumenPromedio,ResumenMediana,ResumenGeometrica,ResumenArmonica,ResumenRecortada,ResumenWinsorizada)

## Análisis Bivariado Cualitativo

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por grupo de ingresos vs sector económico") + ylab("Frecuencia") + xlab("Sector Campaña") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Grupo ingresos") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por grupo de ingresos vs departamento") + ylab("Frecuencia") + xlab("Departamento") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Grupo ingresos") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$CodEsInversionistaRecurrente,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por grupo de ingresos vs Recurrencia de inversion") + ylab("Frecuencia") + xlab("Recurrencia de inversion") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Grupo ingresos") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$GrupoEdad,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por rango de ingresos vs rango de edad") + ylab("Frecuencia") + xlab("Rango de edad") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Rango ingresos") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por rango de edad vs sector económico de la campaña") + ylab("Frecuencia") + xlab("Sector Campaña") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Rango de edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic() + theme\_classic() + ggtitle("Frecuencias por rango de edad vs departamento") + ylab("Frecuencia") + xlab("Departamento") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Rango de edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$CodEsInversionistaRecurrente,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por rango de edad vs recurrencia de inversión") + ylab("Frecuencia") + xlab("Recurrencia de inversión") + theme(legend.title = element\_text(colour="red", size=7,face="bold")) + labs(fill = "Rango de edad") + theme(plot.title = element\_text(color="red", size=7, face="bold.italic"))

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por recurrencia de inversion vs sector campaña") + ylab("Frecuencia") + xlab("Sector Campaña") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + labs(fill = "Recurrencia de inversión") + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por recurrencia de inversion vs departamento") + ylab("Frecuencia") + xlab("Departamento") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + labs(fill = "Recurrencia de inversión") + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(BDa2censo,aes(x=BDa2censo$GrupoEdad,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic() + ggtitle("Frecuencias por recurrencia de inversion vs rango de edad") + ylab("Frecuencia") + xlab("Rango de edad") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + labs(fill = "Recurrencia de inversión") + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

## 2. Resumen Bivariado entre variables Cualitativas y cuantitativas

# 1. Analisis de las variables desde el monto invertido

ggplot(data = BDa2censo, mapping = aes(x = MontoInversion , y = GrupoIngresos)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Monto invertido vs grupo de ingresos") + ylab("Rango de ingresos") + xlab("Monto invertido") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = IngresosMes , y = GrupoIngresos)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa ingresos mes vs grupo de ingresos") + ylab("Rango de ingresos") + xlab("ingresos mes") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = Patrimonio , y = GrupoIngresos)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa patrimonio vs grupo de ingresos") + ylab("Rango de ingresos") + xlab("patrimonio") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = GrupoIngresos)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa edad vs grupo de ingresos") + ylab("Rango de ingresos") + xlab("edad") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = CantidadInversiones , y = GrupoIngresos)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa cantidad de inversiones vs grupo de ingresos") + ylab("Rango de ingresos") + xlab("cantidad de inversiones") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

# 2. Analisis desde el tipo de inversionista (Recurrente o no Recurrente)

ggplot(data = BDa2censo, mapping = aes(x = MontoInversion , y = CodEsInversionistaRecurrente)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Monto invertido vs Recurrencia del inversionista") + ylab("Recurrencia del inversionista") + xlab("Monto invertido") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = IngresosMes , y = CodEsInversionistaRecurrente)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Ingresos mensuales vs Recurrencia del inversionista") + ylab("Recurrencia del inversionista") + xlab("Ingresos mensuales") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = Patrimonio , y = CodEsInversionistaRecurrente)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Patrimonio vs Recurrencia del inversionista") + ylab("Recurrencia del inversionista") + xlab("Patrimonio") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = CodEsInversionistaRecurrente)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Edad vs Recurrencia del inversionista") + ylab("Recurrencia del inversionista") + xlab("Edad") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = Propósito)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Edad vs Propósito del inversionista") + ylab("Propósito del inversionista") + xlab("Edad") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = CantidadInversiones , y = Propósito)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa cantidad de inversiones vs Propósito del inversionista") + ylab("Propósito del inversionista") + xlab("cantidad de inversiones") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

ggplot(data = BDa2censo, mapping = aes(x = MontoInversion , y = Propósito)) + geom\_boxplot(color="red", fill="orange", alpha=0.2) + theme\_classic() + ggtitle("Comparativa Monto invertido vs Propósito del inversionista") + ylab("Propósito del inversionista") + xlab("Monto invertido") + theme(legend.title = element\_text(colour="blue", size=7,face="bold")) + theme(plot.title = element\_text(color="blue", size=8, face="bold"))

## 3. Resumen Multivariado entre variables Cualitativas y cuantitativas

# Analisis del propósito de inversión a la luz de las variables de interés cualitativas y cuantitativas

library(ggplot2)  
library(dplyr)  
library(tidyr)

df1 <- BDa2censo %>%  
 dplyr::group\_by(Propósito, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(MontoInversion, na.rm=TRUE)))

ggplot(df1, aes(Propósito, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=50, hjust=1)) + ggtitle("Comparativa Rangos de edad vs Propósito vs monto invertido") + ylab("Rango de edad") + xlab("Propósito del inversionista") + labs(fill = "promedio de Monto invertido")+  
 geom\_text(aes(label = Mean), color = "black", size = 2)

# Analisis de la cantidad de inversiones realizadas respecto al grupo de edad y el propósito de inversión

df2 <- BDa2censo %>%  
 dplyr::group\_by(Propósito, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(CantidadInversiones, na.rm=TRUE)))

ggplot(df2, aes(Propósito, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=50, hjust=1))+ ggtitle("Comparativa Rangos de edad vs Propósito vs Cantidad de inversiones") + ylab("Rango de edad") + xlab("Propósito del inversionista") + labs(fill = "Cantidad de inversiones") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

# Analisis de la cantidad de inversiones realizadas respecto al sector económico de la campaña y el propósito de inversión

df3 <- BDa2censo %>%  
 dplyr::group\_by(Propósito, GrupoIngresos)%>%  
 dplyr::summarize(Mean = round(mean(MontoInversion, na.rm=TRUE)))

ggplot(df3, aes(Propósito, GrupoIngresos , fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=50, hjust=1))+ ggtitle("Comparativa Proposito vs ingresos vs Monto invertido") + ylab("Rango de ingresos") + xlab("Propósito del inversionista") + labs(fill = "Promedio invertido") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

df4 <- BDa2censo %>%  
 dplyr::group\_by(Propósito, GrupoIngresos)%>%  
 dplyr::summarize(Mean = round(mean(CantidadInversiones, na.rm=TRUE)))

ggplot(df4, aes(Propósito, GrupoIngresos , fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=50, hjust=1))+ ggtitle("Comparativa Proposito vs ingresos vs Cantidad inversiones") + ylab("Rango de ingresos") + xlab("Propósito del inversionista") + labs(fill = "Promedio de inversiones") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

df5 <- BDa2censo %>%  
 dplyr::group\_by(CodEsInversionistaRecurrente, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(MontoInversion, na.rm=TRUE)))

ggplot(df5, aes(CodEsInversionistaRecurrente, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1)) + ggtitle("Comparativa Rango de edad vs Recurrencia vs Monto invertido") + ylab("Rango de edad") + xlab("Recurrencia del inversionista") + labs(fill = "Promedio invertido") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

df6 <- BDa2censo %>%  
 dplyr::group\_by(CodEsInversionistaRecurrente, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(CantidadInversiones, na.rm=TRUE)))

ggplot(df6, aes(CodEsInversionistaRecurrente, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1)) + ggtitle("Comparativa Rango de edad vs Recurrencia vs cantidad de inversiones") + ylab("Rango de edad") + xlab("Recurrencia del inversionista") + labs(fill = "Promedio de inversiones") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

df7 <- BDa2censo %>%  
 dplyr::group\_by(GrupoIngresos, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(MontoInversion, na.rm=TRUE)))

ggplot(df7, aes(GrupoIngresos, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1)) + ggtitle("Comparativa Rango de edad vs rango ingresos vs monto invertido") + ylab("Rango de edad") + xlab("Rango de ingresos") + labs(fill = "Promedio de inversiones") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

df8 <- BDa2censo %>%  
 dplyr::group\_by(GrupoIngresos, GrupoEdad)%>%  
 dplyr::summarize(Mean = round(mean(CantidadInversiones, na.rm=TRUE)))

ggplot(df8, aes(GrupoIngresos, GrupoEdad, fill= Mean)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1)) + ggtitle("Comparativa Rango de edad vs rango ingresos vs cantidad de inversiones") + ylab("Rango de edad") + xlab("Rango de ingresos") + labs(fill = "Promedio de inversiones") +  
 geom\_text(aes(label = Mean), color = "black", size = 2)

# 4. Resumen Bivariado Cuantitativas - Regresiones Lineales simples, multiples y regresión logistica multiple

## Correlaciones Lineales simples entre las variables cuantitativas y su influencia en el monto invertido

cor(x = BDa2censo$Edad, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.144479

cor(x = BDa2censo$IngresosMes, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.09333871

cor(x = BDa2censo$Patrimonio, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.2583145

cor(x = BDa2censo$CantidadInversiones, y = BDa2censo$MontoInversion, method = "pearson")

## [1] -0.02984937

#cor(x = BDa2censo$CodProposito, y = BDa2censo$MontoInversion, method = "pearson")  
  
#cor(x = BDa2censo$CodProposito, y = BDa2censo$CantidadInversiones, method = "pearson")  
  
#cor(x = BDa2censo$PotencialP2, y = BDa2censo$Edad, method = "pearson")

## Regresion Lineales simples entre las variables cuantitativas y su influencia en la cantidad de inversiones

cor(x = BDa2censo$Edad, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.1549182

cor(x = BDa2censo$IngresosMes, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.09968228

cor(x = BDa2censo$Patrimonio, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.2861183

cor(x = BDa2censo$MontoInversion, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] -0.02984937

**Grafico General de Correlaciones lineales**

corrgram(BDa2censo, order=TRUE, lower.panel=panel.shade,  
 upper.panel=panel.pie, text.panel=panel.txt,  
 main="Matriz de Correlaciones General")

### Analisis de correlaciones múltiples

modelotest <- lm(MontoInversion ~ CantidadInversiones + Edad + IngresosMes + Patrimonio, data = BDa2censo )  
  
summary(modelotest)

Predicciones1<-modelotest$fitted.values  
  
library(ggplot2)  
ggplot(data = BDa2censo, aes(x = predict(modelotest),   
 y = abs(rstudent(modelotest))))+  
 geom\_hline(yintercept = 3, color = "grey", linetype = "dashed")+  
 geom\_point(aes(color = ifelse(abs(rstudent(modelotest)) > 2, "red", "black")))+  
 scale\_color\_identity()+  
 labs(title = "DistribuciÃ³n de los residuos estudentizados",   
 x = "PredicciÃ³n modelo",   
 y = "Residuos estudentizados")+  
 theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

modelotest2 <- lm(CantidadInversiones ~ Patrimonio + MontoInversion + Edad + IngresosMes, data = BDa2censo )  
  
summary(modelotest2)

Predicciones1<-modelotest2$fitted.values

## Analisis de modelo de regresión incluyendo solo las variables que demostraron aportar información al modelo general

step(object = modelotest, direction = "both", trace = 1) ## calculo de criterio de perdida de info

names(BDa2censo)

modeloDef <- lm(MontoInversion ~ CantidadInversiones ,data = BDa2censo )  
  
summary(modeloDef)

# —————-Inicio nuevas regresiones Guillermo———————-

BDa2censo$IngresosMes <- BDa2censo$IngresosMes / 1000000  
BDa2censo$Patrimonio <- BDa2censo$Patrimonio / 1000000  
BDa2censo$MontoInversion <- BDa2censo$MontoInversion / 1000000

# Paso 1 :

pairs.panels(BDa2censo[,15:20],pch = 20,stars = TRUE, main="Correlaciones entre variables")

# Paso 2:

Modelo de Regresion Lineal Simple

modeloSimple <- lm(MontoInversion ~ Patrimonio, data = BDa2censo)  
summary(modeloSimple)

modeloSimple$coefficients

## (Intercept) Patrimonio   
## 0.725892957 0.001537974

# paso 3:

Modelo de Regresión Lienal Multiple

modeloMultiple1 <- lm(MontoInversion ~ Edad +Tasa+IngresosMes+Patrimonio+CantidadInversiones, data = BDa2censo )  
summary(modeloMultiple1)

modeloMultiple1$coefficients

step(object = modeloMultiple1, direction = "both", trace = 1)

## 1. Multicolinealidad

vif(modeloMultiple1)

## 2. Relacion Lineal entre predictores y variable respuesta

crPlots(modeloMultiple1)

3. **Distribuciónnormaldelosresiduos**

mean(modeloMultiple1$residuals)

# Si la relación es lineal,los residuos se distribuyen de forma aleatoria entorno a cero

library(ggfortify)  
autoplot(modeloMultiple1,2) ## Qqplot

**4.Homocedasticidad**

library(lmtest)  
bptest(modeloMultiple1) #Breusch-Pagan Test For Homoscedasticity

par(mfrow=c(1,1))  
spreadLevelPlot(modeloMultiple1)

autoplot(modeloMultiple1,1)

## 5.Independencia

#independencia ()  
durbinWatsonTest(modeloMultiple1) # no cumple test de independencia

plot(modeloMultiple1$resid) # existe un patron , significa q no pasa la prueba , no son independintes

acf(modeloMultiple1$residuals) # Si usamos una gracica de autocorrelacion envidenciamos q los residuos no se establizand de manera alterna al interio del intervalo de confianza , por lo tanto se peude inferir que existe autocorrelacion

### Regresión logistica

BDa2censo$CodEsInversionistaRecurrente<-as.factor(BDa2censo$CodEsInversionistaRecurrente)  
BDa2censo$SectorCampaña<-as.factor(BDa2censo$SectorCampaña)  
#BDa2censo$CalidadTributaria<-as.factor(BDa2censo$CalidadTributaria)  
BDa2censo$Propósito<-as.factor(BDa2censo$Propósito)  
#BDa2censo$CodProposito<-as.factor(BDa2censo$CodProposito)  
#BDa2censo$PotencialP2<-as.factor(BDa2censo$PotencialP2)

set.seed(88)  
split <- sample.split(BDa2censo$CodEsInversionistaRecurrente, SplitRatio = 0.75)  
training <- subset(BDa2censo,split =="TRUE")  
testing <- subset(BDa2censo,split=="FALSE")

logMultiple=glm(data=training,CodEsInversionistaRecurrente~MontoInversion+Edad+IngresosMes+Patrimonio+SectorCampaña+Propósito, family = binomial)  
summary(logMultiple)

str(data)

logMultiple4=glm(data=training,CodEsInversionistaRecurrente~MontoInversion+Edad+IngresosMes+Patrimonio+SectorCampaña+Propósito, family = binomial)  
summary(logMultiple4)

**PREDICCIONES**

res <- predict(logMultiple4, training,type = 'response')

# matriz de confusion

Matriz\_Conf<-table(Predicción=res>0.48, Real=training$CodEsInversionistaRecurrente )  
Matriz\_Conf

# Precisión

(Matriz\_Conf[[1,1]]+Matriz\_Conf[[2,2]])/sum(Matriz\_Conf)

# Curva ROC

test\_prob = predict(logMultiple4, newdata = training, type = "response")  
test\_roc = roc(training$CodEsInversionistaRecurrente ~ test\_prob, plot = TRUE, print.auc = TRUE)

# Muestreo

N=nrow(BDa2censo)  
n=500  
set.seed(123)  
masBD <- BDa2censo [ sample (N, size = n ),]  
  
  
MediaEdad=round(mean(masBD$Edad),2)  
MediaEdad

VarEdad=var(masBD$Edad)  
VarEdad

VarEstim=round(((1-n/N)\*VarEdad/n),2)  
VarEstim

EE=round(sqrt(VarEstim),2)  
EE

CV=round((EE/MediaEdad)\*100,2)  
CV

valort=qt(c(0.025),df=(n-1),lower.tail = FALSE)# probabilidad de cola 0.025 equivle a nivel de confianza del 95%, ya q la dist t es simetrica  
  
Lsup=round(MediaEdad+(valort\*EE),2)   
Linf=round(MediaEdad-(valort\*EE),2)  
  
  
round(c(VarEdad,Linf),digit=2)

resumenMediaEdad <- data.frame(n ,MediaEdad,VarEstim,EE,Linf,Lsup,CV)  
resumenMediaEdad

mean(BDa2censo$Edad)

MediaInversion=mean(masBD$MontoInversion)  
VarMuestraInv =var(masBD$MontoInversion)  
VarEstimadaMediaInv=((N-n)/N)\*(VarMuestraInv/n)  
  
  
VarEstimInv=round(((1-n/N)\*VarEstimadaMediaInv/n),2)  
VarEstimInv

EEmediaInv=round(sqrt(VarEstimInv),2)  
EEmediaInv

CVmediaInv=round((EEmediaInv/MediaInversion)\*100,2)  
CVmediaInv

valort=qt(c(0.025),df=(n-1),lower.tail = FALSE)# probabilidad de cola 0.025 equivle a nivel de confianza del 95%, ya q la dist t es simetrica  
  
LsupMediaInv=round(MediaInversion+(valort\*EEmediaInv),2)   
LinfMediaInv=round(MediaInversion-(valort\*EEmediaInv),2)  
  
  
  
resumenMediaInversion <- data.frame(n ,MediaInversion,VarEstimInv,EEmediaInv,LinfMediaInv,LsupMediaInv,CVmediaInv)  
resumenMediaInversion

mean(BDa2censo$MontoInversion)

MediaInversion=mean(masBD$MontoInversion)  
TotalInversion=N\*MediaInversion  
TotalInversion

VarMuestra =var(masBD$MontoInversion)  
VarEstimadaMedia=((N-n)/N)\*(VarMuestra/n)  
  
VarEstimadaTotal=N\*N\*VarEstimadaMedia  
EEtotal=sqrt(VarEstimadaTotal)  
  
  
CVtotal=round((EEtotal/TotalInversion)\*100,2)  
CVtotal

LsupTotal=round(TotalInversion+(valort\*EEtotal),2)   
LinfTotal=round(TotalInversion-(valort\*EEtotal),2)  
  
sum(BDa2censo$MontoInversion)

resumenTotalInversion <- data.frame(n ,TotalInversion,VarEstimadaTotal,EEtotal,LinfTotal,LsupTotal,CVtotal)  
resumenTotalInversion

library(rgdal)

library(sf)

sp\_df <- readOGR(dsn = "~/GitHub/ProyectoEstadistica/ProyectoEstadistica/Sergio/MGN2021\_DPTO\_POLITICO", layer = "MGN\_DPTO\_POLITICO")

#head(sp\_df)  
#fix(sp\_df)  
#as.data.frame(sp\_df)  
  
DPTO\_SH="~/GitHub/ProyectoEstadistica/ProyectoEstadistica/Sergio/MGN2021\_DPTO\_POLITICO/MGN\_DPTO\_POLITICO.shp"  
  
DPTO\_SH2 <- st\_read(DPTO\_SH)

library(raster)

library(leaflet)

library(dplyr)  
library(sf)  
library(tmap)  
library(tigris)

library(dplyr)  
InverDepto <- masBD %>%  
 group\_by(Departamento) %>%  
 summarise(promedioEdad = mean(Edad),promedioInver = mean(MontoInversion),Cod\_Dpto)

# Para visualizar la base resumida  
ResumenInv=as.data.frame(InverDepto)  
#ResumenInv  
  
  
library(tidyr)  
Etiquetas=unite(ResumenInv, Etiqueta,c(1,2,3), sep = ": ", remove = TRUE)  
Etiquetas=Etiquetas[,1]  
#Etiquetas  
  
  
Resumen3=cbind(ResumenInv, Etiquetas )  
#Resumen3  
  
  
#DPTO\_SH2  
  
DPTO\_JOIN <- geo\_join(DPTO\_SH2, Resumen3,"DPTO\_CCDGO", "Cod\_Dpto")

##DPTO\_JOIN = as.data.frame(DPTO\_JOIN)  
#DPTO\_JOIN  
  
  
pal <- colorNumeric( palette = "RdYlBu", domain=DPTO\_JOIN$promedioInver) #palette = "YlGnBu" "RdBu" "RdYlBu" "Spectral" "Paired" "PuRd" "RdYlGn"  
  
popup\_sb <- paste0("Promedio de Monto de Inversión: ", as.character(DPTO\_JOIN$promedioInver))  
  
  
#leaflet(sp\_df) %>%  
# addProviderTiles("CartoDB.Positron") %>%  
# #setView(-98.483330, 38.712046, zoom = 4) %>%   
# addPolygons(data = DPTO\_JOIN ,   
# fillColor = ~pal(DPTO\_JOIN$promedioInver),  
# opacity = 1,  
# color = "black",  
# dashArray = "3",fillOpacity = 0.9,  
# highlight = highlightOptions(  
# weight = 1,  
# color = "#666",  
# dashArray = "",  
# fillOpacity = 1,  
# bringToFront = TRUE),  
# label = DPTO\_JOIN$Etiquetas,  
# labelOptions = labelOptions(  
# style = list("font-weight" = "normal", padding = "3px 8px"),  
# textsize = "15px",  
# direction = "auto"))%>%  
# addLegend(pal = pal, values =DPTO\_JOIN$promedioInver, opacity = 0.7, title = NULL,  
# position = "bottomright")