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)

## X..BDa2censo.IngresosMes  
## nobs 2.572000e+04  
## NAs 0.000000e+00  
## Minimum 1.010000e+06  
## Maximum 1.540000e+08  
## 1. Quartile 3.000000e+06  
## 3. Quartile 8.000000e+06  
## Mean 7.590004e+06  
## Median 4.878000e+06  
## Sum 1.952149e+11  
## SE Mean 7.490181e+04  
## LCL Mean 7.443193e+06  
## UCL Mean 7.736816e+06  
## Variance 1.442964e+14  
## Stdev 1.201235e+07  
## Skewness 7.140657e+00  
## Kurtosis 6.439523e+01

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)

## X..BDa2censo.Patrimonio  
## nobs 2.572000e+04  
## NAs 0.000000e+00  
## Minimum 1.000000e+06  
## Maximum 8.064210e+08  
## 1. Quartile 2.000000e+07  
## 3. Quartile 2.158680e+08  
## Mean 1.570082e+08  
## Median 1.000000e+08  
## Sum 4.038252e+12  
## SE Mean 1.120823e+06  
## LCL Mean 1.548114e+08  
## UCL Mean 1.592051e+08  
## Variance 3.231059e+16  
## Stdev 1.797515e+08  
## Skewness 1.613582e+00  
## Kurtosis 2.098108e+00

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"))

### 14: Monto Inversión

basicStats(BDa2censo$MontoInversion)

## X..BDa2censo.MontoInversion  
## nobs 2.572000e+04  
## NAs 0.000000e+00  
## Minimum 2.000000e+05  
## Maximum 5.300000e+06  
## 1. Quartile 3.000000e+05  
## 3. Quartile 1.000000e+06  
## Mean 9.673676e+05  
## Median 5.000000e+05  
## Sum 2.488069e+10  
## SE Mean 6.673247e+03  
## LCL Mean 9.542876e+05  
## UCL Mean 9.804475e+05  
## Variance 1.145369e+12  
## Stdev 1.070219e+06  
## Skewness 2.150895e+00  
## Kurtosis 4.542472e+00

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)

## X..BDa2censo.Edad  
## nobs 2.572000e+04  
## NAs 0.000000e+00  
## Minimum 1.800000e+01  
## Maximum 8.600000e+01  
## 1. Quartile 3.000000e+01  
## 3. Quartile 4.200000e+01  
## Mean 3.718176e+01  
## Median 3.500000e+01  
## Sum 9.563150e+05  
## SE Mean 6.133800e-02  
## LCL Mean 3.706154e+01  
## UCL Mean 3.730199e+01  
## Variance 9.676859e+01  
## Stdev 9.837102e+00  
## Skewness 1.263190e+00  
## Kurtosis 2.000467e+00

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)

## X..BDa2censo.CantidadInversiones  
## nobs 2.572000e+04  
## NAs 0.000000e+00  
## Minimum 1.000000e+00  
## Maximum 9.100000e+01  
## 1. Quartile 3.000000e+00  
## 3. Quartile 1.700000e+01  
## Mean 1.219810e+01  
## Median 8.000000e+00  
## Sum 3.137350e+05  
## SE Mean 7.678000e-02  
## LCL Mean 1.204760e+01  
## UCL Mean 1.234859e+01  
## Variance 1.516242e+02  
## Stdev 1.231358e+01  
## Skewness 1.748056e+00  
## Kurtosis 3.334008e+00

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 15 : Tasa

basicStats(BDa2censo$Tasa)

## X..BDa2censo.Tasa  
## nobs 25720.000000  
## NAs 0.000000  
## Minimum 0.080000  
## Maximum 0.130000  
## 1. Quartile 0.100000  
## 3. Quartile 0.110000  
## Mean 0.104405  
## Median 0.100000  
## Sum 2685.293400  
## SE Mean 0.000064  
## LCL Mean 0.104280  
## UCL Mean 0.104530  
## Variance 0.000105  
## Stdev 0.010254  
## Skewness 0.436234  
## Kurtosis -0.255612

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)  
  
resumenMedia1 <- data.frame(ResumenVaribles ,ResumenArmonica,ResumenRecortada,ResumenWinsorizada)  
  
  
knitr::kable(  
 (resumenMedia)  
)

**Análisis Bivariado Cualitativo**

**Visualización de las variables cualitativas desde la perspectiva de grupo de ingresos**

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"))

**Visualización de las variables cualitativas desde la perspectiva de inversionista recurrente**

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"))

Los inversionistas Recurrentes llegan a tener un promedio de patrimonio superior, aun cuando hay personas consideradas inversionistas no recurrentes con un muy alto nivel de patrimonio

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)

##   
## Attaching package: 'tidyr'

## The following object is masked from 'package:igraph':  
##   
## crossing

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

## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.

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)))

## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.

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)))

## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.

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)))

## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.

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)

**Analisis desde la recurrencia de inversión a la luz de las demás variables de interés cualitativas y cuantitativas**

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

## `summarise()` has grouped output by 'CodEsInversionistaRecurrente'. You can  
## override using the `.groups` argument.

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)))

## `summarise()` has grouped output by 'CodEsInversionistaRecurrente'. You can  
## override using the `.groups` argument.

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)

**Analisis desde los rangos de edad y rangos de ingresos a la luz de las variables cuantitativas de interés**

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

## `summarise()` has grouped output by 'GrupoIngresos'. You can override using the  
## `.groups` argument.

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)))

## `summarise()` has grouped output by 'GrupoIngresos'. You can override using the  
## `.groups` argument.

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

## 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

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

# Paso Cero:

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)

##   
## Call:  
## lm(formula = MontoInversion ~ Patrimonio, data = BDa2censo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.7563 -0.5720 -0.3720 0.2126 4.5592   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.72589296 0.00856002 84.80 <0.0000000000000002 \*\*\*  
## Patrimonio 0.00153797 0.00003587 42.88 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.034 on 25718 degrees of freedom  
## Multiple R-squared: 0.06673, Adjusted R-squared: 0.06669   
## F-statistic: 1839 on 1 and 25718 DF, p-value: < 0.00000000000000022

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)

##   
## Call:  
## lm(formula = MontoInversion ~ Edad + Tasa + IngresosMes + Patrimonio +   
## CantidadInversiones, data = BDa2censo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.1659 -0.5821 -0.3297 0.2066 4.5504   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.08695033 0.07034580 1.236 0.216   
## Edad 0.00433078 0.00072268 5.993 0.00000000209 \*\*\*  
## Tasa 5.31433292 0.62444368 8.511 < 0.0000000000000002 \*\*\*  
## IngresosMes 0.00584545 0.00054045 10.816 < 0.0000000000000002 \*\*\*  
## Patrimonio 0.00157547 0.00004086 38.558 < 0.0000000000000002 \*\*\*  
## CantidadInversiones -0.01042623 0.00054229 -19.226 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.023 on 25714 degrees of freedom  
## Multiple R-squared: 0.08625, Adjusted R-squared: 0.08607   
## F-statistic: 485.4 on 5 and 25714 DF, p-value: < 0.00000000000000022

modeloMultiple1$coefficients

## (Intercept) Edad Tasa IngresosMes   
## 0.086950331 0.004330777 5.314332923 0.005845453   
## Patrimonio CantidadInversiones   
## 0.001575472 -0.010426229

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

## Start: AIC=1182.07  
## MontoInversion ~ Edad + Tasa + IngresosMes + Patrimonio + CantidadInversiones  
##   
## Df Sum of Sq RSS AIC  
## <none> 26917 1182.1  
## - Edad 1 37.59 26955 1216.0  
## - Tasa 1 75.82 26993 1252.4  
## - IngresosMes 1 122.46 27040 1296.8  
## - CantidadInversiones 1 386.95 27304 1547.2  
## - Patrimonio 1 1556.24 28473 2625.7

##   
## Call:  
## lm(formula = MontoInversion ~ Edad + Tasa + IngresosMes + Patrimonio +   
## CantidadInversiones, data = BDa2censo)  
##   
## Coefficients:  
## (Intercept) Edad Tasa   
## 0.086950 0.004331 5.314333   
## IngresosMes Patrimonio CantidadInversiones   
## 0.005845 0.001575 -0.010426

## Paso4 Validacion de Supuestos

## 1. Multicolinealidad

vif(modeloMultiple1)

## Edad Tasa IngresosMes Patrimonio   
## 1.241702 1.007372 1.035539 1.325390   
## CantidadInversiones   
## 1.095531

## 2. Relacion Lineal entre predictores y variable respuesta

crPlots(modeloMultiple1)

## 3. Distribución normal de los residuos

mean(modeloMultiple1$residuals)

## [1] -0.0000000000000001302145

# 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

##   
## studentized Breusch-Pagan test  
##   
## data: modeloMultiple1  
## BP = 1108.5, df = 5, p-value < 0.00000000000000022

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

##   
## Suggested power transformation: -0.2003072

autoplot(modeloMultiple1,1)

## 5.Independencia

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

## lag Autocorrelation D-W Statistic p-value  
## 1 0.9619126 0.07553598 0  
## Alternative hypothesis: rho != 0

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

library(plot3D)  
library(plot3Drgl)

## Loading required package: rgl

z<-BDa2censo$MontoInversion  
y<-BDa2censo$Patrimonio  
x<-BDa2censo$Edad  
  
scatter3D(x, y, z, theta = 15, phi = 20)

scatter3D(x, y, z, phi = 0, bty ="g")

scatter3D(x, y, z, pch = 18, theta = 20, phi = 20,  
 main = "Modelo Monto de Inversión", xlab = "Edad",  
 ylab ="Patrimonio", zlab = "Inversion")

scatter3D(x, y, z, phi = 0, bty = "g",  
 pch = 20, cex = 2, ticktype = "detailed",xlab = "Edad",  
 ylab ="Patrimonio", zlab = "Inversion")

objr<-lm(z ~ x+y)  
objr

##   
## Call:  
## lm(formula = z ~ x + y)  
##   
## Coefficients:  
## (Intercept) x y   
## 0.583237 0.004267 0.001436

grid.lines = 26  
x.pred <- seq(min(x), max(x), length.out = grid.lines)  
y.pred <- seq(min(y), max(y), length.out = grid.lines)  
xy <- expand.grid( x = x.pred, y = y.pred)  
z.pred <- matrix(predict(objr, newdata = xy),   
 nrow = grid.lines, ncol = grid.lines)  
  
# Marcamos las líneas de iteracción para que busquen la recta de regresión  
fitpoints <- predict(objr)  
#ploteamos la gráfica en 3d con recta de regresión  
scatter3D(x, y, z, pch = 18, cex = 2,   
 theta = 20, phi = 20, ticktype = "detailed",  
 xlab = "Edad", ylab = "Patrimonio", zlab = "Inversion",   
 surf = list(x = x.pred, y = y.pred, z = z.pred,   
 facets = NA, fit = fitpoints), main = "")

plotrgl()

## Regresión logistica

BDa2censo$Inv\_Recurrente<-BDa2censo$CodEsInversionistaRecurrente  
  
BDa2censo$Inv\_Recurrente[BDa2censo$Inv\_Recurrente=="Inversionista No Recurrente"]<-0  
BDa2censo$Inv\_Recurrente[BDa2censo$Inv\_Recurrente=="Inversionista Recurrente"]<-1  
BDa2censo$Inv\_Recurrente<-as.numeric(BDa2censo$Inv\_Recurrente)  
BDa2censo$SectorCampaña<-as.factor(BDa2censo$SectorCampaña)  
BDa2censo$Propósito<-as.factor(BDa2censo$Propósito)  
  
  
library(fdth)  
library(caTools)  
library(ROCR)  
library(pROC)  
library(magrittr)

##   
## Attaching package: 'magrittr'

## The following object is masked from 'package:tidyr':  
##   
## extract

library(InformationValue)  
  
set.seed(88)  
split <- sample.split(BDa2censo$Inv\_Recurrente, SplitRatio = 0.75)  
training <- subset(BDa2censo,split =="TRUE")  
testing <- subset(BDa2censo,split=="FALSE")  
  
  
logMultiple=glm(data=training,Inv\_Recurrente~MontoInversion+Edad+IngresosMes+Patrimonio+SectorCampaña+Propósito, family = binomial)  
summary(logMultiple)

##   
## Call:  
## glm(formula = Inv\_Recurrente ~ MontoInversion + Edad + IngresosMes +   
## Patrimonio + SectorCampaña + Propósito, family = binomial,   
## data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0912 -0.8618 -0.6968 1.2166 2.1939   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -1.9246331 0.2979440 -6.460  
## MontoInversion -0.2243309 0.0166163 -13.501  
## Edad 0.0173327 0.0017639 9.826  
## IngresosMes 0.0092231 0.0013452 6.856  
## Patrimonio 0.0022523 0.0001008 22.348  
## SectorCampañaAlcantarillado 0.2878625 0.3045023 0.945  
## SectorCampañaAlojamiento Y Servicios De Comida -0.0295437 0.2920597 -0.101  
## SectorCampañaAutomotriz 0.5099706 0.2968569 1.718  
## SectorCampañaComercio al por mayor 0.3723162 0.2912361 1.278  
## SectorCampañaComercio Al Por Menor 0.0446268 0.2923675 0.153  
## SectorCampañaEducación 1.0029118 0.3505771 2.861  
## SectorCampañaEntretenimiento 0.5914717 0.3454151 1.712  
## SectorCampañaIndustrías Creativas Y Culturales 0.6087700 0.3732196 1.631  
## SectorCampañaInformación y Comunicaciones 0.7096011 0.2925527 2.426  
## SectorCampañaInmobiliarias -0.0060257 0.2926296 -0.021  
## SectorCampañaInvestigación y Ciencias 0.1011636 0.2947185 0.343  
## SectorCampañaManufactura 0.4045251 0.2898151 1.396  
## SectorCampañaSalud 0.4026100 0.3058514 1.316  
## SectorCampañaSaneamiento Ambiental -0.2091516 0.3053564 -0.685  
## SectorCampañaServicios Administrativos -0.2422360 0.2926781 -0.828  
## SectorCampañaServicios Domésticos 0.8118238 0.3515981 2.309  
## SectorCampañaServicios Energéticos 0.2204440 0.2981431 0.739  
## SectorCampañaTransporte y Almacenamiento 0.1203064 0.3000769 0.401  
## PropósitoAprender de financiación 0.1616005 0.0884718 1.827  
## PropósitoDiversificar mi portafolio de inversión 0.3406143 0.0533117 6.389  
## PropósitoHacer parte de la comunidad a2censo - -0.4472824 0.1410949 -3.170  
## PropósitoRentabilizar portafolio de inversión -0.0382445 0.0525907 -0.727  
## Pr(>|z|)   
## (Intercept) 0.00000000010490 \*\*\*  
## MontoInversion < 0.0000000000000002 \*\*\*  
## Edad < 0.0000000000000002 \*\*\*  
## IngresosMes 0.00000000000707 \*\*\*  
## Patrimonio < 0.0000000000000002 \*\*\*  
## SectorCampañaAlcantarillado 0.34448   
## SectorCampañaAlojamiento Y Servicios De Comida 0.91943   
## SectorCampañaAutomotriz 0.08581 .   
## SectorCampañaComercio al por mayor 0.20111   
## SectorCampañaComercio Al Por Menor 0.87868   
## SectorCampañaEducación 0.00423 \*\*   
## SectorCampañaEntretenimiento 0.08683 .   
## SectorCampañaIndustrías Creativas Y Culturales 0.10286   
## SectorCampañaInformación y Comunicaciones 0.01529 \*   
## SectorCampañaInmobiliarias 0.98357   
## SectorCampañaInvestigación y Ciencias 0.73141   
## SectorCampañaManufactura 0.16277   
## SectorCampañaSalud 0.18805   
## SectorCampañaSaneamiento Ambiental 0.49338   
## SectorCampañaServicios Administrativos 0.40787   
## SectorCampañaServicios Domésticos 0.02095 \*   
## SectorCampañaServicios Energéticos 0.45967   
## SectorCampañaTransporte y Almacenamiento 0.68848   
## PropósitoAprender de financiación 0.06776 .   
## PropósitoDiversificar mi portafolio de inversión 0.00000000016686 \*\*\*  
## PropósitoHacer parte de la comunidad a2censo - 0.00152 \*\*   
## PropósitoRentabilizar portafolio de inversión 0.46710   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 24421 on 19289 degrees of freedom  
## Residual deviance: 22772 on 19263 degrees of freedom  
## AIC: 22826  
##   
## Number of Fisher Scoring iterations: 4

logMultiple1=glm(data=training,Inv\_Recurrente~MontoInversion+Edad+IngresosMes+Patrimonio, family = binomial)  
summary(logMultiple1)

##   
## Call:  
## glm(formula = Inv\_Recurrente ~ MontoInversion + Edad + IngresosMes +   
## Patrimonio, family = binomial, data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8052 -0.8464 -0.7460 1.2711 2.1235   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.57829952 0.06327048 -24.945 < 0.0000000000000002 \*\*\*  
## MontoInversion -0.21313284 0.01633425 -13.048 < 0.0000000000000002 \*\*\*  
## Edad 0.01583860 0.00172735 9.169 < 0.0000000000000002 \*\*\*  
## IngresosMes 0.00831926 0.00125744 6.616 0.0000000000369 \*\*\*  
## Patrimonio 0.00242873 0.00009816 24.744 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 24421 on 19289 degrees of freedom  
## Residual deviance: 23218 on 19285 degrees of freedom  
## AIC: 23228  
##   
## Number of Fisher Scoring iterations: 4

logMultiple2=glm(data=training,Inv\_Recurrente~MontoInversion+Edad+IngresosMes+Patrimonio, family = binomial)  
summary(logMultiple2)

##   
## Call:  
## glm(formula = Inv\_Recurrente ~ MontoInversion + Edad + IngresosMes +   
## Patrimonio, family = binomial, data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8052 -0.8464 -0.7460 1.2711 2.1235   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.57829952 0.06327048 -24.945 < 0.0000000000000002 \*\*\*  
## MontoInversion -0.21313284 0.01633425 -13.048 < 0.0000000000000002 \*\*\*  
## Edad 0.01583860 0.00172735 9.169 < 0.0000000000000002 \*\*\*  
## IngresosMes 0.00831926 0.00125744 6.616 0.0000000000369 \*\*\*  
## Patrimonio 0.00242873 0.00009816 24.744 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 24421 on 19289 degrees of freedom  
## Residual deviance: 23218 on 19285 degrees of freedom  
## AIC: 23228  
##   
## Number of Fisher Scoring iterations: 4

## Predicciones

# Elaboramos las respectivas predicciones de acaurdo al modelo escogido para evalaur el desempeño del mismo

predict <- predict(logMultiple, type = 'response')  
predict2 <- as.data.frame(predict(logMultiple, type = 'response'))

optCutOff <- optimalCutoff(training$Inv\_Recurrente, predict)[1]  
  
optCutOff

## [1] 0.4576869

## matriz de confusion

confusionMatrix(training$Inv\_Recurrente, predict, threshold = optCutOff)

## 0 1  
## 0 11540 4606  
## 1 1416 1728

## Sensibilidad y especificidad

sensitivity(training$Inv\_Recurrente, predict, threshold = optCutOff)

## [1] 0.2728134

specificity(training$Inv\_Recurrente,predict, threshold = optCutOff)

## [1] 0.890707

## Curva ROC

plotROC(training$Inv\_Recurrente, predict)

# Muestreo

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

## [1] 37.04

VarEdad=var(masBD$Edad)  
VarEdad

## [1] 92.59374

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

## [1] 0.18

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

## [1] 0.42

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

## [1] 1.13

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)

## [1] 92.59 36.21

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

## n MediaEdad VarEstim EE Linf Lsup CV  
## 1 500 37.04 0.18 0.42 36.21 37.87 1.13

mean(BDa2censo$Edad)

## [1] 37.18177

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

## [1] 0

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

## [1] 0

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

## [1] 0

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

## n MediaInversion VarEstimInv EEmediaInv LinfMediaInv LsupMediaInv  
## 1 500 0.9407 0 0 0.94 0.94  
## CVmediaInv  
## 1 0

mean(BDa2censo$MontoInversion)

## [1] 0.9673676

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

## [1] 24194.8

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

## [1] 4.91

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

## [1] 24880.69

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

## n TotalInversion VarEstimadaTotal EEtotal LinfTotal LsupTotal CVtotal  
## 1 500 24194.8 1413995 1189.115 21858.51 26531.09 4.91

####  
  
library(rgdal)

## Loading required package: sp

## Please note that rgdal will be retired by the end of 2023,  
## plan transition to sf/stars/terra functions using GDAL and PROJ  
## at your earliest convenience.  
##   
## rgdal: version: 1.5-32, (SVN revision 1176)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 3.3.2, released 2021/09/01  
## Path to GDAL shared files: C:/Program Files/R/R-4.2.0/library/rgdal/gdal  
## GDAL binary built with GEOS: TRUE   
## Loaded PROJ runtime: Rel. 7.2.1, January 1st, 2021, [PJ\_VERSION: 721]  
## Path to PROJ shared files: C:/Program Files/R/R-4.2.0/library/rgdal/proj  
## PROJ CDN enabled: FALSE  
## Linking to sp version:1.4-7  
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,  
## use options("rgdal\_show\_exportToProj4\_warnings"="none") before loading sp or rgdal.

##   
## Attaching package: 'rgdal'

## The following object is masked from 'package:fBasics':  
##   
## getDescription

library(sf)

## Linking to GEOS 3.9.1, GDAL 3.3.2, PROJ 7.2.1; sf\_use\_s2() is TRUE

sp\_df <- readOGR(dsn = "MGN2021\_DPTO\_POLITICO", layer = "MGN\_DPTO\_POLITICO")

## Warning in OGRSpatialRef(dsn, layer, morphFromESRI = morphFromESRI, dumpSRS =  
## dumpSRS, : Discarded datum Marco\_Geocentrico\_Nacional\_de\_Referencia in Proj4  
## definition: +proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no\_defs

## OGR data source with driver: ESRI Shapefile   
## Source: "D:\Scripts\_SQL\ProyectoEstadistica\ProyectoEstadistica\Sergio\MGN2021\_DPTO\_POLITICO", layer: "MGN\_DPTO\_POLITICO"  
## with 33 features  
## It has 9 fields

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

## Reading layer `MGN\_DPTO\_POLITICO' from data source   
## `D:\Scripts\_SQL\ProyectoEstadistica\ProyectoEstadistica\Sergio\MGN2021\_DPTO\_POLITICO\MGN\_DPTO\_POLITICO.shp'   
## using driver `ESRI Shapefile'  
## Simple feature collection with 33 features and 9 fields  
## Geometry type: MULTIPOLYGON  
## Dimension: XY  
## Bounding box: xmin: -81.73562 ymin: -4.229406 xmax: -66.84722 ymax: 13.39473  
## Geodetic CRS: MAGNA-SIRGAS

library(raster)

##   
## Attaching package: 'raster'

## The following object is masked from 'package:plotly':  
##   
## select

## The following object is masked from 'package:MASS':  
##   
## select

## The following object is masked from 'package:dplyr':  
##   
## select

library(leaflet)

##   
## Attaching package: 'leaflet'

## The following object is masked from 'package:xts':  
##   
## addLegend

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

## To enable caching of data, set `options(tigris\_use\_cache = TRUE)`  
## in your R script or .Rprofile.

##   
## Attaching package: 'tigris'

## The following object is masked from 'package:igraph':  
##   
## blocks

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

## `summarise()` has grouped output by 'Departamento'. You can override using the  
## `.groups` argument.

# 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")

## Warning: We recommend using the dplyr::\*\_join() family of functions instead.

## Warning: `group\_by\_()` was deprecated in dplyr 0.7.0.  
## Please use `group\_by()` instead.  
## See vignette('programming') for more help  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.

##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")