Crowfunding en el Mercado de Capitales Colombiano - Proyecto de Estadística para la Analítica de Datos

Sergio Castañeda, Guillermo Castillo, Alexander Guecha

18/05/2022

Justificación: el presente analisis estadístico propone estimar el monto a invertir por un inversionista, teniendo en cuenta las variables que componen la base con la información recopilada de su interacción con el producto.

Con la anterior información se puede ser mas especifico con el portafolio que se le ofrezca a un inversionistas actual y potencial, de cara a generar una diversificación del uso de las alternativas de inversión en personas que han tomado cercanía con la oferta de valor de a2censo.

knitr::opts\_chunk$set(echo = F)  
knitr::opts\_chunk$set(echo = F)  
library(readxl)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(fBasics)

## Loading required package: timeDate

## Loading required package: timeSeries

library(MASS)

##   
## Attaching package: 'MASS'

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

library(corrgram)  
library(gclus)

## Loading required package: cluster

library(descr)  
library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(agricolae)

##   
## Attaching package: 'agricolae'

## The following objects are masked from 'package:timeDate':  
##   
## kurtosis, skewness

library(plotly)

##   
## Attaching package: 'plotly'

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

## The following object is masked from 'package:timeSeries':  
##   
## filter

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(fdth)

##   
## Attaching package: 'fdth'

## The following objects are masked from 'package:stats':  
##   
## sd, var

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

## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following object is masked from 'package:fdth':  
##   
## var

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

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

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

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

## Loading required package: xts

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following object is masked from 'package:timeSeries':  
##   
## time<-

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

##   
## Attaching package: 'PerformanceAnalytics'

## The following objects are masked from 'package:agricolae':  
##   
## kurtosis, skewness

## The following objects are masked from 'package:timeDate':  
##   
## kurtosis, skewness

## The following object is masked from 'package:graphics':  
##   
## legend

library(psych)

##   
## Attaching package: 'psych'

## The following object is masked from 'package:car':  
##   
## logit

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

## The following object is masked from 'package:timeSeries':  
##   
## outlier

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(corrr)  
library(igraph)

##   
## Attaching package: 'igraph'

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

## The following object is masked from 'package:agricolae':  
##   
## similarity

## The following objects are masked from 'package:gclus':  
##   
## diameter, star

## The following objects are masked from 'package:dplyr':  
##   
## as\_data\_frame, groups, union

## The following objects are masked from 'package:stats':  
##   
## decompose, spectrum

## The following object is masked from 'package:base':  
##   
## union

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

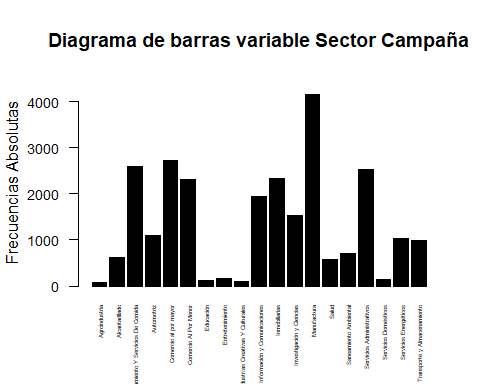
# ANÁLISIS ESTADÍSTICO BASE DE DATOS INVERSIONISTAS DE A2CENSO

## Resumen univariado de variables cualitativas

### Variable 1: SectorCampaña

Esta variable agrupa el conjunto de actividades productivas o comerciales que reúnen una serie de características similares en relación con una o más campañas, es decir, son negocios que cuentan con una naturaleza común. Mediante el siguiente análisis de frecuencias, se busca determinar cual sector de campaña cuenta con una mayor aceptación por parte de los inversionista

|  | SectorCampaña\_ | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 13 | Manufactura | 4166 | 0.1619751 | 4166 | 0.1619751 |
| 5 | Comercio al por mayor | 2723 | 0.1058709 | 6889 | 0.2678460 |
| 3 | Alojamiento Y Servicios De Comida | 2593 | 0.1008165 | 9482 | 0.3686625 |
| 16 | Servicios Administrativos | 2532 | 0.0984448 | 12014 | 0.4671073 |
| 11 | Inmobiliarias | 2324 | 0.0903577 | 14338 | 0.5574650 |
| 6 | Comercio Al Por Menor | 2307 | 0.0896967 | 16645 | 0.6471617 |
| 10 | Información y Comunicaciones | 1941 | 0.0754666 | 18586 | 0.7226283 |
| 12 | Investigación y Ciencias | 1531 | 0.0595257 | 20117 | 0.7821540 |
| 4 | Automotriz | 1092 | 0.0424572 | 21209 | 0.8246112 |
| 18 | Servicios Energéticos | 1033 | 0.0401633 | 22242 | 0.8647745 |
| 19 | Transporte y Almacenamiento | 982 | 0.0381804 | 23224 | 0.9029549 |
| 15 | Saneamiento Ambiental | 714 | 0.0277605 | 23938 | 0.9307154 |
| 2 | Alcantarillado | 610 | 0.0237170 | 24548 | 0.9544323 |
| 14 | Salud | 569 | 0.0221229 | 25117 | 0.9765552 |
| 8 | Entretenimiento | 155 | 0.0060264 | 25272 | 0.9825816 |
| 17 | Servicios Domésticos | 142 | 0.0055210 | 25414 | 0.9881026 |
| 7 | Educación | 125 | 0.0048600 | 25539 | 0.9929627 |
| 9 | Industrías Creativas Y Culturales | 104 | 0.0040435 | 25643 | 0.9970062 |
| 1 | Agroindustria | 77 | 0.0029938 | 25720 | 1.0000000 |

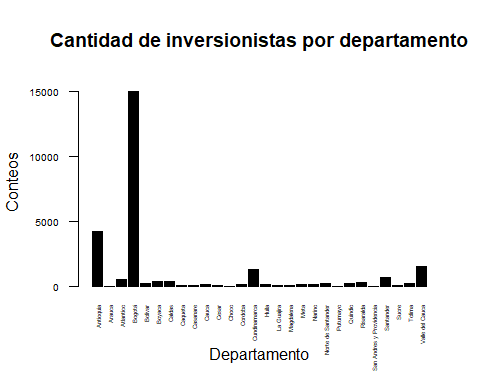


Podemos evidenciar que las campañas de inversión que mayor tracción de inversionistas tuvieron, se encontraban asociadas con los sectores económicos de Manufactura (n= 4166), Comercio al por mayor (n= 2724)y Alojamiento y servicios de comida (n = 2593), por otro lado las campañas asociadas a sectores económicos de:Agroindustria (n=77), industrias creativas (n=104) y educación (n=124)fueron los 3 sectores económicos que menor tracción de numero de inversionistas tuvieron.

### Variable 2: Departamento

Con esta variable se buscar conocer el grado de participación de los inversionistas así como de las campañas en el territorio colombiano.

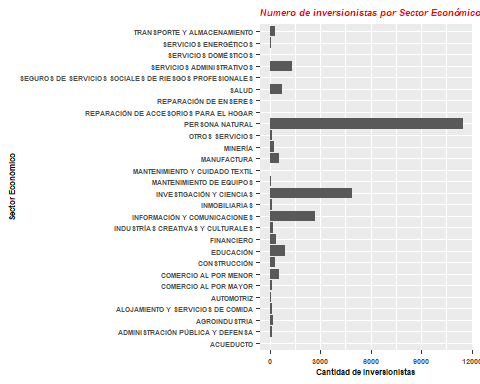
|  | Departamento | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 4 | Bogotá | 15026 | 0.5842146 | 15026 | 0.5842146 |
| 1 | Antioquia | 4207 | 0.1635692 | 19233 | 0.7477838 |
| 28 | Valle del Cauca | 1534 | 0.0596423 | 20767 | 0.8074261 |
| 14 | Cundinamarca | 1260 | 0.0489891 | 22027 | 0.8564152 |
| 25 | Santander | 688 | 0.0267496 | 22715 | 0.8831649 |
| 3 | Atlantico | 477 | 0.0185459 | 23192 | 0.9017107 |
| 7 | Caldas | 346 | 0.0134526 | 23538 | 0.9151633 |
| 6 | Boyaca | 323 | 0.0125583 | 23861 | 0.9277216 |
| 23 | Risaralda | 295 | 0.0114697 | 24156 | 0.9391913 |
| 22 | Quindio | 230 | 0.0089425 | 24386 | 0.9481337 |
| 5 | Bolivar | 205 | 0.0079705 | 24591 | 0.9561042 |
| 27 | Tolima | 182 | 0.0070762 | 24773 | 0.9631804 |
| 20 | Norte de Santander | 164 | 0.0063764 | 24937 | 0.9695568 |
| 19 | Narino | 114 | 0.0044323 | 25051 | 0.9739891 |
| 15 | Huila | 113 | 0.0043935 | 25164 | 0.9783826 |
| 10 | Cauca | 102 | 0.0039658 | 25266 | 0.9823484 |
| 18 | Meta | 100 | 0.0038880 | 25366 | 0.9862364 |
| 13 | Cordoba | 97 | 0.0037714 | 25463 | 0.9900078 |
| 11 | Cesar | 78 | 0.0030327 | 25541 | 0.9930404 |
| 9 | Casanare | 55 | 0.0021384 | 25596 | 0.9951788 |
| 26 | Sucre | 47 | 0.0018274 | 25643 | 0.9970062 |
| 17 | Magdalena | 30 | 0.0011664 | 25673 | 0.9981726 |
| 8 | Caqueta | 25 | 0.0009720 | 25698 | 0.9991446 |
| 16 | La Guajira | 8 | 0.0003110 | 25706 | 0.9994557 |
| 2 | Arauca | 5 | 0.0001944 | 25711 | 0.9996501 |
| 21 | Putumayo | 5 | 0.0001944 | 25716 | 0.9998445 |
| 24 | San Andres y Providencia | 3 | 0.0001166 | 25719 | 0.9999611 |
| 12 | Choco | 1 | 0.0000389 | 25720 | 1.0000000 |

 De otro lado, si bien es esperado que la mayor participación de inversionistas se encuentren ubicados en Bogotá (n= 15032), resulta llamativo observar las participaciones que se tienen en otros departamentos como Antioquia (n= 4207) , Valle del cauca (n= 1534), cundinamarca (n= 1260) santander (n= 688) y atlantico (n= 477).

### Variable 3: Sector Económico del inversionista

Esta variable categórica es muy relevante para nuestro análisis estadístico, toda vez que nos permite conocer de primera mano los sectores económicos de los inversionistas interesados en las campañas publicitarias que ofrece a2censo, así como la capacidad de inversión para financiar los distintos proyectos que ofrece la compañía.

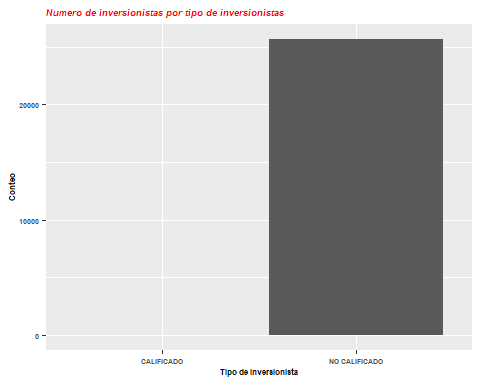
|  | SectorEconomico | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 20 | PERSONA NATURAL | 11467 | 0.4458398 | 11467 | 0.4458398 |
| 14 | INVESTIGACIÓN Y CIENCIAS | 4893 | 0.1902411 | 16360 | 0.6360809 |
| 12 | INFORMACIÓN Y COMUNICACIONES | 2672 | 0.1038880 | 19032 | 0.7399689 |
| 25 | SERVICIOS ADMINISTRATIVOS | 1337 | 0.0519829 | 20369 | 0.7919518 |
| 9 | EDUCACIÓN | 938 | 0.0364697 | 21307 | 0.8284215 |
| 23 | SALUD | 711 | 0.0276439 | 22018 | 0.8560653 |
| 7 | COMERCIO AL POR MENOR | 583 | 0.0226672 | 22601 | 0.8787325 |
| 17 | MANUFACTURA | 552 | 0.0214619 | 23153 | 0.9001944 |
| 10 | FINANCIERO | 403 | 0.0156687 | 23556 | 0.9158631 |
| 28 | TRANSPORTE Y ALMACENAMIENTO | 318 | 0.0123639 | 23874 | 0.9282271 |
| 8 | CONSTRUCCIÓN | 310 | 0.0120529 | 24184 | 0.9402799 |
| 18 | MINERÍA | 277 | 0.0107698 | 24461 | 0.9510498 |
| 3 | AGROINDUSTRIA | 170 | 0.0066096 | 24631 | 0.9576594 |
| 11 | INDUSTRÍAS CREATIVAS Y CULTURALES | 170 | 0.0066096 | 24801 | 0.9642691 |
| 13 | INMOBILIARIAS | 162 | 0.0062986 | 24963 | 0.9705677 |
| 6 | COMERCIO AL POR MAYOR | 157 | 0.0061042 | 25120 | 0.9766719 |
| 2 | ADMINISTRACIÓN PÚBLICA Y DEFENSA | 148 | 0.0057543 | 25268 | 0.9824261 |
| 19 | OTROS SERVICIOS | 119 | 0.0046267 | 25387 | 0.9870529 |
| 4 | ALOJAMIENTO Y SERVICIOS DE COMIDA | 114 | 0.0044323 | 25501 | 0.9914852 |
| 27 | SERVICIOS ENERGÉTICOS | 66 | 0.0025661 | 25567 | 0.9940513 |
| 5 | AUTOMOTRIZ | 57 | 0.0022162 | 25624 | 0.9962675 |
| 15 | MANTENIMIENTO DE EQUIPOS | 57 | 0.0022162 | 25681 | 0.9984837 |
| 26 | SERVICIOS DOMÉSTICOS | 22 | 0.0008554 | 25703 | 0.9993390 |
| 1 | ACUEDUCTO | 7 | 0.0002722 | 25710 | 0.9996112 |
| 16 | MANTENIMIENTO Y CUIDADO TEXTIL | 5 | 0.0001944 | 25715 | 0.9998056 |
| 21 | REPARACIÓN DE ACCESORIOS PARA EL HOGAR | 2 | 0.0000778 | 25717 | 0.9998834 |
| 22 | REPARACIÓN DE ENSERES | 2 | 0.0000778 | 25719 | 0.9999611 |
| 24 | SEGUROS DE SERVICIOS SOCIALES DE RIESGOS PROFESIONALES | 1 | 0.0000389 | 25720 | 1.0000000 |

 La mayor proporsion de inversionistas autodenominan su actividad económica como persona natural (n= 11471), siendo que es la opcion mas entendible para usuarios que no conocen estas clasificaciones económicas, no obtante resulta llamativo observar la participacion de personas que se dedican a actividades de investigación y ciencias (n= 4893), información y comunicaciones (n= 2672) y servicios administrativos (n= 1337) en donde podemos encontrar posiblemente un proporsion de inversionistas asalariados que usan sus excedentes de dinero en inversiones en a2censo.

### Variable 4: Tipo Inversionista

Esta variable categórica es usada por a2censo para segmentar la población que aplica a las distintas campañas de inversión en función del interés y nivel de experticia en los actuales ecosistemas de inversión y financiación.

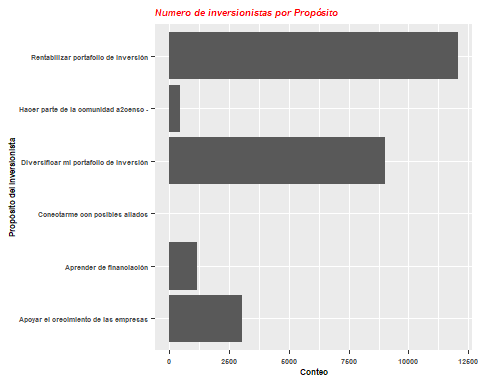
|  | TipoInversionista | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 2 | NO CALIFICADO | 25674 | 0.9982115 | 25674 | 0.9982115 |
| 1 | CALIFICADO | 46 | 0.0017885 | 25720 | 1.0000000 |

 Por otro lado, podemos encontrar que la proporsion de inversionistas que se autodenominan inversionistas calificados y que han invertido en a2censo es muy baja (n= 46), respecto a los inversionistas no calificados (n= 25680) representando una oportunidad de abordar más a este nicho de inversionistas calificados con este producto y hacer un mayor detalle a su comportamiento de inversión.

### Variable 5: Propósito del inversionista en A2censo

Mediante esta variable categórica, a2censo busca conocer la finalidad o el objetivo de inversión de cada uno de los inversionistas, así como la correlación que pueda existir entre estos objetivos y las campañas realizadas, para brindar mejores opciones por parte de la compañía y por consiguiente acrecentar la motivación y confianza de los clientes.

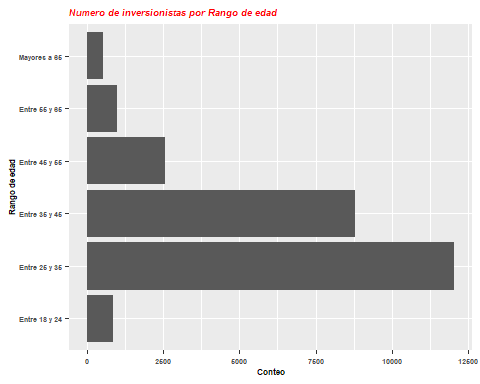
|  | Proposito | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 6 | Rentabilizar portafolio de inversión | 12063 | 0.4690124 | 12063 | 0.4690124 |
| 4 | Diversificar mi portafolio de inversión | 9032 | 0.3511664 | 21095 | 0.8201788 |
| 1 | Apoyar el crecimiento de las empresas | 3027 | 0.1176905 | 24122 | 0.9378694 |
| 2 | Aprender de financiación | 1138 | 0.0442457 | 25260 | 0.9821151 |
| 5 | Hacer parte de la comunidad a2censo - | 459 | 0.0178460 | 25719 | 0.9999611 |
| 3 | Conectarme con posibles aliados | 1 | 0.0000389 | 25720 | 1.0000000 |

 De acuerdo a los anteriores analisis podemos evidenciar como la mayoria de inversionistas actuales siguen propósitos de inversión asociados al crecimiento y rentabilidad del dinero y el portafolio de inversión actual (n = 12066); y por otro lado a la diversificación del portafolio de inversion ( n= 9034). No obstante podemos encontrar que hay un 11.8% de la muestra que tienen fines filantrópicos de apoyar el emprendimiento Colombiano (n = 3027).

### Variable 6: Grupo de Edad

Mediante esta variable categórica, a2censo realizad una segmentación de los clientes por grupo etario, el cual brinda información de gran importancia a la compañía al momento de lanzar un proyecto de inversión, toda vez que le permite predecir de antemano la población objetivo en función de la edad.

|  | GrupoEdad | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 2 | Entre 25 y 35 | 12048 | 0.4684292 | 12048 | 0.4684292 |
| 3 | Entre 35 y 45 | 8786 | 0.3416019 | 20834 | 0.8100311 |
| 4 | Entre 45 y 55 | 2556 | 0.0993779 | 23390 | 0.9094090 |
| 5 | Entre 55 y 65 | 970 | 0.0377138 | 24360 | 0.9471229 |
| 1 | Entre 18 y 24 | 838 | 0.0325816 | 25198 | 0.9797045 |
| 6 | Mayores a 65 | 522 | 0.0202955 | 25720 | 1.0000000 |

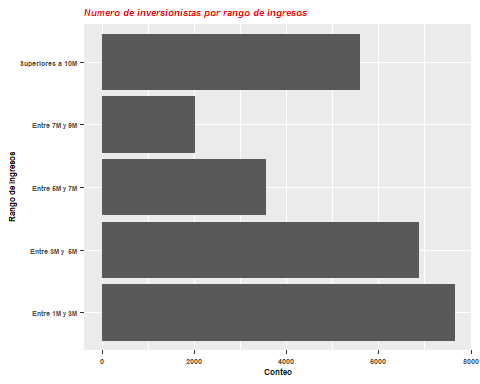


De acuerdo al anterior analisis, resulta interesante observar la importante participación del 47% de inversionistas jovenes entre 25 y 35 años (n = 12049) y tambien de inversionistas entre 35 y 45 años (n = 8787). No obstante la participacion de inversionistas entre 45 y 55 años (9,9%, n = 2558) no es una participación despreciable dentro del grupo total de inversionistas.

### Variable 7: Grupo de Ingresos

Mediante esta variable categórica se busca agrupar a los clientes inversionistas en razón al nivel de ingresos mensuales y de esta forma tener un proyección del capital disponible al momento de realizar una campaña en particular y, del grupo de interés al cual se puede brindar el lanzamiento del proyecto. De otra parte, para fines de este análisis se proyecta conocer si existe correlación de esta variable con la proyección del monto de inversión.

|  | GrupoIngresos | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 1 | Entre 1M y 3M | 7652 | 0.2975117 | 7652 | 0.2975117 |
| 2 | Entre 3M y 5M | 6873 | 0.2672240 | 14525 | 0.5647356 |
| 5 | Superiores a 10M | 5609 | 0.2180793 | 20134 | 0.7828149 |
| 3 | Entre 5M y 7M | 3563 | 0.1385303 | 23697 | 0.9213453 |
| 4 | Entre 7M y 9M | 2023 | 0.0786547 | 25720 | 1.0000000 |

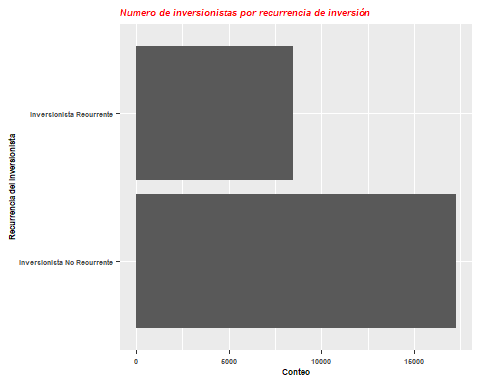


Dentro de los inversionistas de a2censo encontramos inversionistas de tres grupos de rangos salariales predominantes, dentro de los cuales se encuentran en primer lugar una mayor proporsión de inversionistas de ingresos bajos (menor a 1 millon mensual; n= 7652), Inversionistas de ingresos medios - bajos (entre 3 y 5 millones mensuales; n= 6873), Inversionistas de ingresos altos (Superiores a 10 Millones mensuales; n = 5615).No obstante si agruparamos, podemos observar que mas del 50% de inversionistas suelen ser personas con ingresos inferiores a 5 millones de pesos.

### Variable 8: Inversionistas recurrentes

Esta variable categórica está con la frecuencia de participación de cada uno de los inversionistas de a2censo en las campañas ofrecidas, el cual se categoriza a cada cliente entre frecuente o no, partiendo de un parámetro de medición previamente establecido por la compañía, ayudando en la toma de decisiones, toda vez que brinda información relevante para predecir el monto de inversión que puede llegar a tener una campaña en particular.

| InvRecurr | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Inversionista No Recurrente | 17274 | 0.6716174 | 17274 | 0.6716174 |
| Inversionista Recurrente | 8446 | 0.3283826 | 25720 | 1.0000000 |

 Podemos evidenciar que la mayoría (67%) de inversionistas no son recurrentes ( n = 17280), Sin embargo existe una alta proporsión de inversionistas recurrentes (33%; n= 8446) que pueden ser de interés en el analisis realizado

## Resumen univariado de variables cuantitativas

A continuación se presenta el analisis variable por variable en lo que respecta a información cuantitativa.

Vale la pena resaltar que dichos analisis se hicieron siguiendo la regla de establecimiento de clases de Sturges que incluye la funcion fdt del paquete fdth.

Este proceso se realizó de esta manera puesto que al realizar los analisis bajo un numero entre 3 y 7 intervalos de clase que fueron probados en cada una de las variables, la distribución de la información no permitía desagregarla en grupos que contribuyeran al analisis y que se representaran de mejor manera, en todo caso la distribución en clases bajo la regla de sturges permitió la mejor visualización de la información cuantitativa.

### Variable 9: Ingresos Mes

Mediante esta variable cuantitativa, a2cendo lleva el registro del ingreso mensual reportado por el inversionista en pesos colombianos. Para el presente análisis estadístico fue seleccionada esta variable, con el propósito de establecer el nivel de correlación que existe entre el nivel de ingresos mensuales de cada inversionista con la proyección del monto de inversión.

## 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [999900,10658656.2) | 21389 | 0.8316096 | 83.1609642 | 21389 | 83.16096 |
| [10658656.2,20317412.5) | 3345 | 0.1300544 | 13.0054432 | 24734 | 96.16641 |
| [20317412.5,29976168.8) | 383 | 0.0148911 | 1.4891135 | 25117 | 97.65552 |
| [29976168.8,39634925) | 155 | 0.0060264 | 0.6026439 | 25272 | 98.25816 |
| [39634925,49293681.2) | 115 | 0.0044712 | 0.4471229 | 25387 | 98.70529 |
| [49293681.2,58952437.5) | 48 | 0.0018663 | 0.1866252 | 25435 | 98.89191 |
| [58952437.5,68611193.8) | 39 | 0.0015163 | 0.1516330 | 25474 | 99.04355 |
| [68611193.8,78269950) | 13 | 0.0005054 | 0.0505443 | 25487 | 99.09409 |
| [78269950,87928706.2) | 37 | 0.0014386 | 0.1438569 | 25524 | 99.23795 |
| [87928706.2,97587462.5) | 29 | 0.0011275 | 0.1127527 | 25553 | 99.35070 |
| [97587462.5,107246219) | 68 | 0.0026439 | 0.2643857 | 25621 | 99.61509 |
| [107246219,116904975) | 4 | 0.0001555 | 0.0155521 | 25625 | 99.63064 |
| [116904975,126563731) | 34 | 0.0013219 | 0.1321928 | 25659 | 99.76283 |
| [126563731,136222488) | 15 | 0.0005832 | 0.0583204 | 25674 | 99.82115 |
| [136222488,145881244) | 12 | 0.0004666 | 0.0466563 | 25686 | 99.86781 |
| [145881244,155540000) | 34 | 0.0013219 | 0.1321928 | 25720 | 100.00000 |

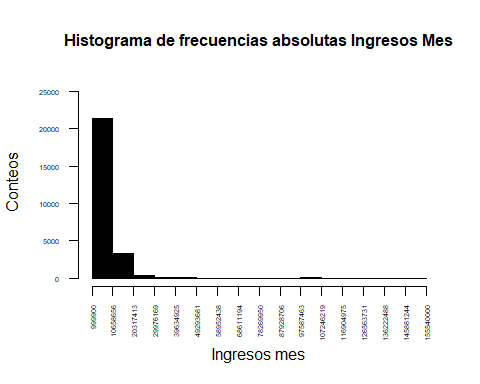
|  | x |
| --- | --- |
| start | 999900 |
| end | 155540000 |
| h | 9658756 |
| right | 0 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

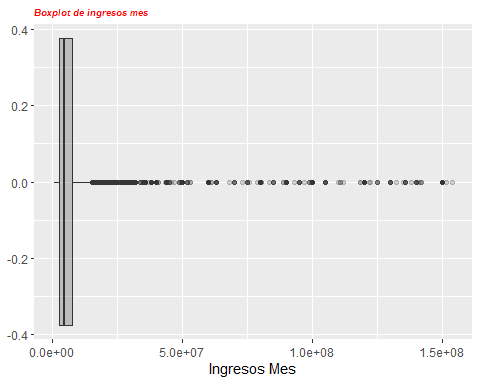
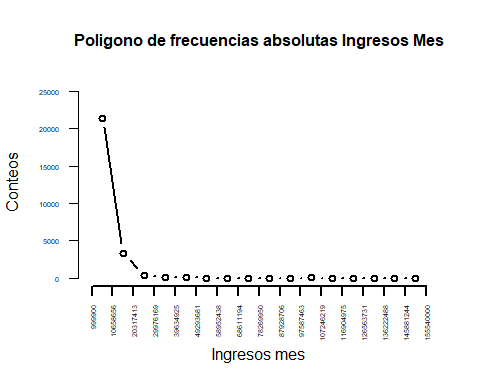
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

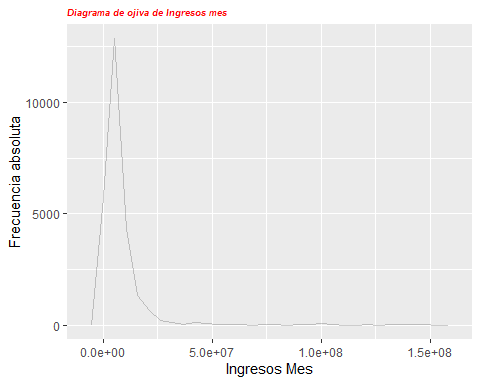
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 De acuerdo al analisis de frecuencia podemos encontrar que un total de 21.389 que corresponde al 83% de los inversionistas se agrupan en salarios entre $990. 000 y 10.658.656 pesos siendo en rango salarial predominante. No obstante el grupo de inversionistas con ingresos entre 10.658.656 y 20.317.412 de pesos representa el 13% (n = 3345) de la muestra de inversionsitas.

### Variable 10: Patrimonio

Mediante esta variable cuantitativa, a2cendo lleva el registro del valor en pesos colombianos de los activos que reporta tener el inversionista. Para el presente análisis estadístico fue seleccionada esta variable, con el propósito de establecer el nivel de correlación que existe entre el valor del patrimonio de cada inversionista con la proyección del monto de inversión.

## 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [990000,51833450.6) | 9512 | 0.3698289 | 36.9828927 | 9512 | 36.98289 |
| [51833450.6,102676901) | 5510 | 0.2142302 | 21.4230171 | 15022 | 58.40591 |
| [102676901,153520352) | 2476 | 0.0962675 | 9.6267496 | 17498 | 68.03266 |
| [153520352,204363803) | 1597 | 0.0620918 | 6.2091757 | 19095 | 74.24184 |
| [204363803,255207253) | 1136 | 0.0441680 | 4.4167963 | 20231 | 78.65863 |
| [255207253,306050704) | 1381 | 0.0536936 | 5.3693624 | 21612 | 84.02799 |
| [306050704,356894154) | 650 | 0.0252722 | 2.5272162 | 22262 | 86.55521 |
| [356894154,407737605) | 873 | 0.0339425 | 3.3942457 | 23135 | 89.94946 |
| [407737605,458581056) | 403 | 0.0156687 | 1.5668740 | 23538 | 91.51633 |
| [458581056,509424506) | 521 | 0.0202566 | 2.0256610 | 24059 | 93.54199 |
| [509424506,560267957) | 315 | 0.0122473 | 1.2247278 | 24374 | 94.76672 |
| [560267957,611111408) | 389 | 0.0151244 | 1.5124417 | 24763 | 96.27916 |
| [611111408,661954858) | 245 | 0.0095257 | 0.9525661 | 25008 | 97.23173 |
| [661954858,712798309) | 271 | 0.0105365 | 1.0536547 | 25279 | 98.28538 |
| [712798309,763641759) | 190 | 0.0073872 | 0.7387247 | 25469 | 99.02411 |
| [763641759,814485210) | 251 | 0.0097589 | 0.9758942 | 25720 | 100.00000 |

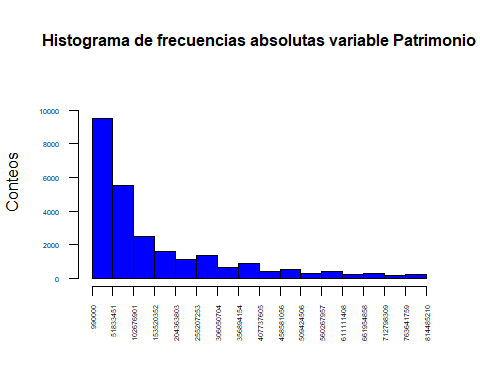
|  | x |
| --- | --- |
| start | 990000 |
| end | 814485210 |
| h | 50843451 |
| right | 0 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

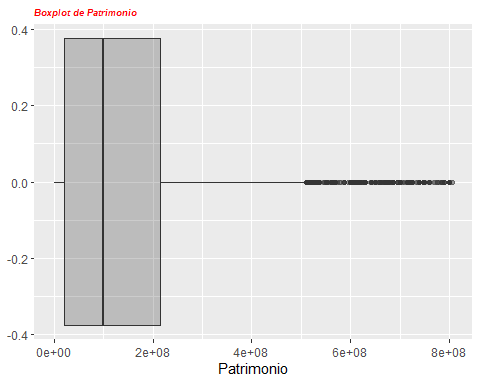
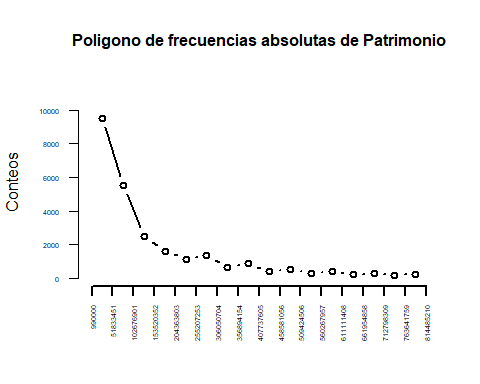
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

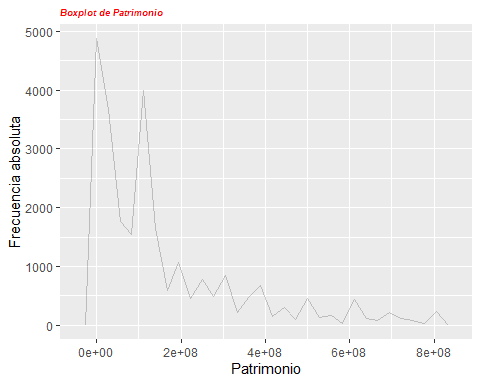
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Por otro lado en lo que respecta al patrimonio de los inversionistas podemos encontrar que el 36% de los inversionistas que representan un total de 9.514 individuos tienen patrimonios entre 990.000 y 51.833.450. Adicionalmente podemos encontrar que el 21% (n= 5510) de los inversionistas tienen patrimonios entre 51.833.450. y 102.676.901 y una proporsión del 9% de los inversionistas ( n= ) tienen patrimonios entre 102.676.901 y 153.520.352

### Variable 11: Monto inversion

Mediante esta variable cuantitativa, a2cendo lleva el registro del valor en pesos colombianos invertidos por cada inversionista en cada una de las campañas disponibles. Para el presente análisis estadístico, fue seleccionada como variable dependiente, con el propósito de desarrollar el mejor modelo que permita a la compañía hacer una proyección de la demanda con base en las variables explicativas seleccionadas para este estudio en particular.

## 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [198000,520187.5) | 13900 | 0.5404355 | 54.0435459 | 13900 | 54.04355 |
| [520187.5,842375) | 1842 | 0.0716174 | 7.1617418 | 15742 | 61.20529 |
| [842375,1164563) | 4175 | 0.1623250 | 16.2325039 | 19917 | 77.43779 |
| [1164563,1486750) | 568 | 0.0220840 | 2.2083981 | 20485 | 79.64619 |
| [1486750,1808938) | 791 | 0.0307543 | 3.0754277 | 21276 | 82.72162 |
| [1808938,2131125) | 1899 | 0.0738336 | 7.3833593 | 23175 | 90.10498 |
| [2131125,2453313) | 183 | 0.0071151 | 0.7115086 | 23358 | 90.81649 |
| [2453313,2775500) | 284 | 0.0110420 | 1.1041991 | 23642 | 91.92068 |
| [2775500,3097688) | 755 | 0.0293546 | 2.9354588 | 24397 | 94.85614 |
| [3097688,3419875) | 87 | 0.0033826 | 0.3382582 | 24484 | 95.19440 |
| [3419875,3742063) | 108 | 0.0041991 | 0.4199067 | 24592 | 95.61431 |
| [3742063,4064250) | 339 | 0.0131804 | 1.3180404 | 24931 | 96.93235 |
| [4064250,4386438) | 34 | 0.0013219 | 0.1321928 | 24965 | 97.06454 |
| [4386438,4708625) | 56 | 0.0021773 | 0.2177294 | 25021 | 97.28227 |
| [4708625,5030813) | 661 | 0.0256998 | 2.5699844 | 25682 | 99.85226 |
| [5030813,5353000) | 38 | 0.0014774 | 0.1477449 | 25720 | 100.00000 |

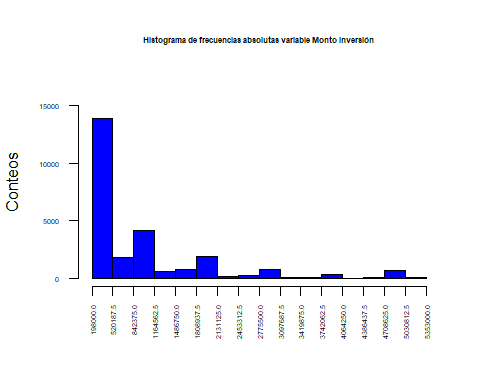
|  | x |
| --- | --- |
| start | 198000.0 |
| end | 5353000.0 |
| h | 322187.5 |
| right | 0.0 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

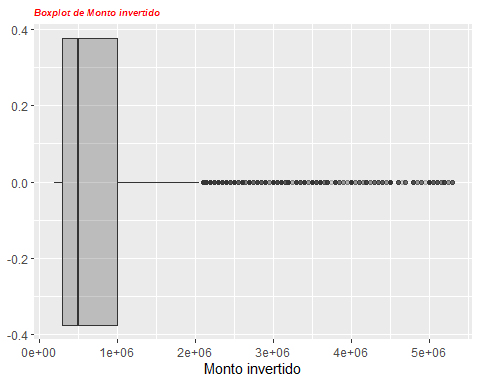
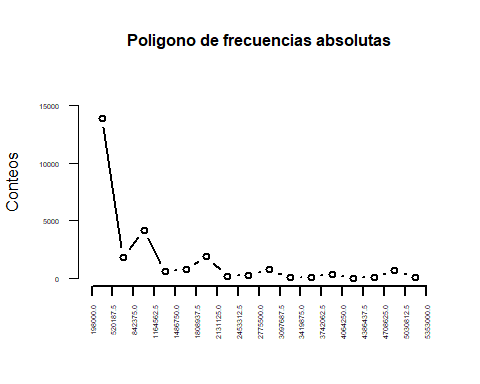
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

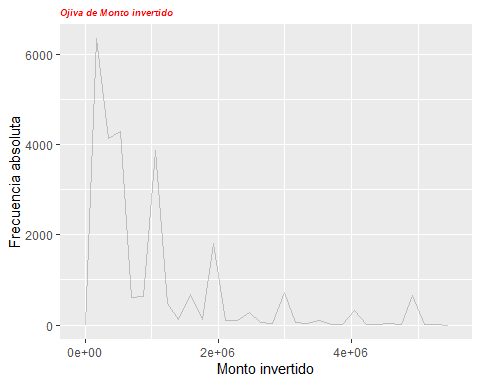
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Respecto al monto invertido podemos encontrar que un 54% de los inversionistas (n= 13.904) invierten en un rango entre 198.000 pesos y 520.187 pesos,encontrandose que este es rango ded ticket mayoritatio de inversión actual. No obstante se encuentra una proporsion del 16% de los inversionistas ( n= 4175) que invierten rangos entre 842.375 pesos y 1.164.563 pesos. Por lo tanto más del 60% de los inversionistas invierten entre 198.000 pesos y 842.375 pesos

### Variable 12: Edad

Mediante esta variable cuantitativa, a2cendo lleva el registro de la edad de cada inversionista, lo cual es un dato que se debe considerar para el presente análisis estadístico, toda vez que este dato puede afectar el cálculo de la proyección del monto de inversión.

## 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [17.82,22.135) | 216 | 0.0083981 | 0.8398134 | 216 | 0.8398134 |
| [22.135,26.45) | 2083 | 0.0809876 | 8.0987558 | 2299 | 8.9385692 |
| [26.45,30.765) | 4741 | 0.1843313 | 18.4331260 | 7040 | 27.3716952 |
| [30.765,35.08) | 5846 | 0.2272939 | 22.7293935 | 12886 | 50.1010886 |
| [35.08,39.395) | 4610 | 0.1792379 | 17.9237947 | 17496 | 68.0248834 |
| [39.395,43.71) | 3092 | 0.1202177 | 12.0217729 | 20588 | 80.0466563 |
| [43.71,48.025) | 2228 | 0.0866252 | 8.6625194 | 22816 | 88.7091757 |
| [48.025,52.34) | 898 | 0.0349145 | 3.4914463 | 23714 | 92.2006221 |
| [52.34,56.655) | 673 | 0.0261664 | 2.6166407 | 24387 | 94.8172628 |
| [56.655,60.97) | 366 | 0.0142302 | 1.4230171 | 24753 | 96.2402799 |
| [60.97,65.285) | 445 | 0.0173017 | 1.7301711 | 25198 | 97.9704510 |
| [65.285,69.6) | 283 | 0.0110031 | 1.1003110 | 25481 | 99.0707621 |
| [69.6,73.915) | 112 | 0.0043546 | 0.4354588 | 25593 | 99.5062208 |
| [73.915,78.23) | 91 | 0.0035381 | 0.3538103 | 25684 | 99.8600311 |
| [78.23,82.545) | 21 | 0.0008165 | 0.0816485 | 25705 | 99.9416796 |
| [82.545,86.86) | 15 | 0.0005832 | 0.0583204 | 25720 | 100.0000000 |

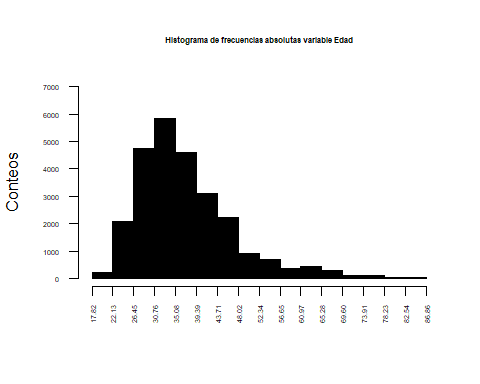
|  | x |
| --- | --- |
| start | 17.820 |
| end | 86.860 |
| h | 4.315 |
| right | 0.000 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

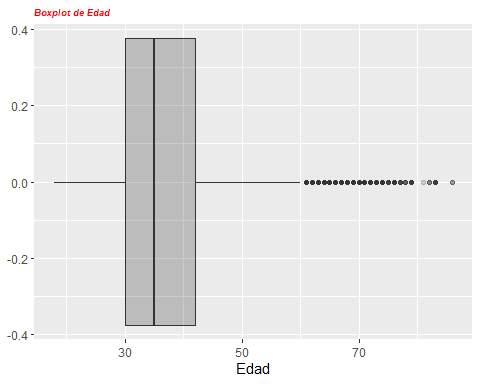
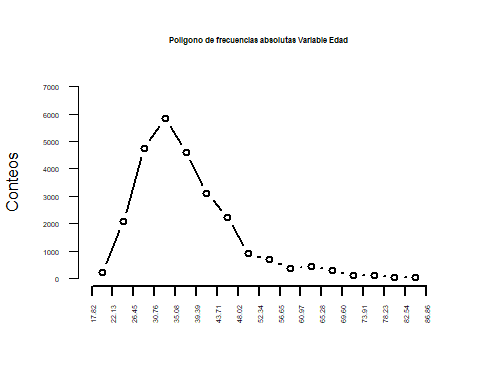
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

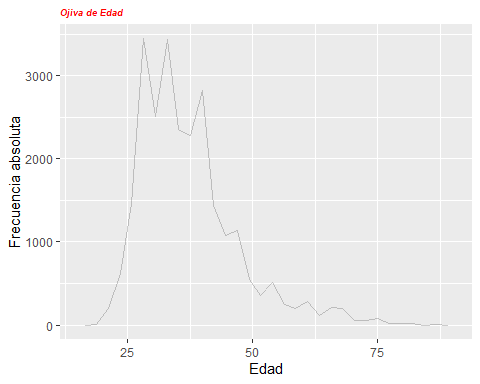
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 De acuerdo al analisis cuantitativo de la variable edad podemos encontrar que la mayor cantidad de inversionistas (22%, n = 5847) se encuentran en un rango de edad entre 31 años y 35 años. No obstante la participación de inversionistas entre los 26 años y 31 años tambien es relevante con un 18% de participación (n= 4741) y tambien el rango comprendido entre los 35 años y 39 años con un 18% de participación (n = 4611) y el rango entre 39 años y 44 años con un 12% de la participación total (n = 3092). Dicho lo anterior los inversionistas entre los 26 años y 45 años suman un total del 59% de participacion de los inversionistas siendo la edad media 35 años.

### Variable 14: Cantidad de inversiones

Mediante esta variable cuantitativa, a2censo lleva el registro de la cantidad de inversiones realizadas por cada inversionista, lo cual es un dato que se debe considerar para el presente análisis estadístico, toda vez que este dato puede afectar el cálculo de la proyección del monto de inversión.

## 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [0.99,6.6725) | 11094 | 0.4313375 | 43.1337481 | 11094 | 43.13375 |
| [6.6725,12.355) | 5411 | 0.2103810 | 21.0381026 | 16505 | 64.17185 |
| [12.355,18.037) | 3623 | 0.1408631 | 14.0863142 | 20128 | 78.25816 |
| [18.037,23.72) | 1649 | 0.0641135 | 6.4113530 | 21777 | 84.66952 |
| [23.72,29.402) | 1566 | 0.0608865 | 6.0886470 | 23343 | 90.75816 |
| [29.402,35.085) | 761 | 0.0295879 | 2.9587869 | 24104 | 93.71695 |
| [35.085,40.768) | 538 | 0.0209176 | 2.0917574 | 24642 | 95.80871 |
| [40.768,46.45) | 359 | 0.0139580 | 1.3958009 | 25001 | 97.20451 |
| [46.45,52.133) | 263 | 0.0102255 | 1.0225505 | 25264 | 98.22706 |
| [52.133,57.815) | 212 | 0.0082426 | 0.8242613 | 25476 | 99.05132 |
| [57.815,63.498) | 115 | 0.0044712 | 0.4471229 | 25591 | 99.49844 |
| [63.498,69.18) | 128 | 0.0049767 | 0.4976672 | 25719 | 99.99611 |
| [69.18,74.862) | 0 | 0.0000000 | 0.0000000 | 25719 | 99.99611 |
| [74.862,80.545) | 0 | 0.0000000 | 0.0000000 | 25719 | 99.99611 |
| [80.545,86.227) | 0 | 0.0000000 | 0.0000000 | 25719 | 99.99611 |
| [86.227,91.91) | 1 | 0.0000389 | 0.0038880 | 25720 | 100.00000 |

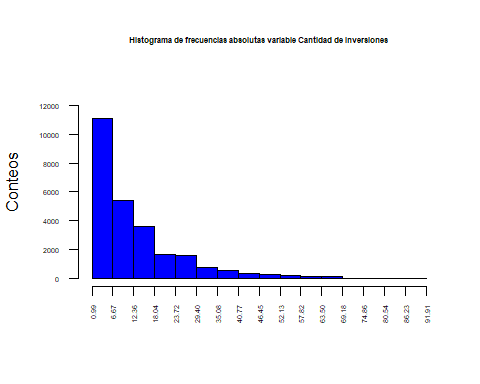
|  | x |
| --- | --- |
| start | 0.9900 |
| end | 91.9100 |
| h | 5.6825 |
| right | 0.0000 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

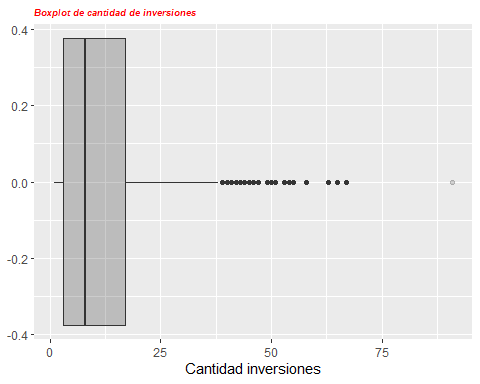
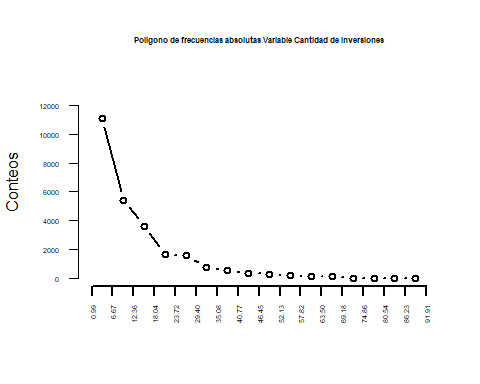
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

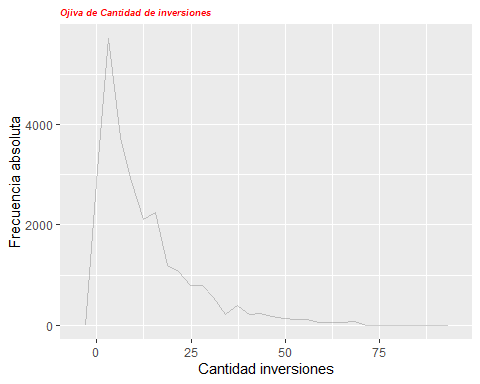
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Respecto a la variable Cantidad de inversiones podemos encontrar que el promedio de inversiones es 12 inversiones. No obstante un 43% (n = 11.099) de los inversionistas han realizado entre 1 y 7 inversiones. un 21% (n = 5.412) han realizado entre 7 y 12 inversiones, un 14% (n= 3.623) han realizado entre 12 y 18 inversiones y un 6% (n = 1.649) han realizado entre 18 y 24 inversiones en la plataforma de a2censo.

### Variable 15: 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

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [0.0792,0.08246) | 187 | 0.0072706 | 0.7270607 | 187 | 0.7270607 |
| [0.08246,0.08571) | 370 | 0.0143857 | 1.4385692 | 557 | 2.1656299 |
| [0.08571,0.08897) | 77 | 0.0029938 | 0.2993779 | 634 | 2.4650078 |
| [0.08897,0.09223) | 2636 | 0.1024883 | 10.2488336 | 3270 | 12.7138414 |
| [0.09223,0.09548) | 1581 | 0.0614697 | 6.1469673 | 4851 | 18.8608087 |
| [0.09548,0.09874) | 1289 | 0.0501166 | 5.0116641 | 6140 | 23.8724728 |
| [0.09874,0.102) | 7593 | 0.2952177 | 29.5217729 | 13733 | 53.3942457 |
| [0.102,0.1053) | 2749 | 0.1068818 | 10.6881804 | 16482 | 64.0824261 |
| [0.1053,0.1085) | 213 | 0.0082815 | 0.8281493 | 16695 | 64.9105754 |
| [0.1085,0.1118) | 3067 | 0.1192457 | 11.9245723 | 19762 | 76.8351477 |
| [0.1118,0.115) | 2284 | 0.0888025 | 8.8802488 | 22046 | 85.7153966 |
| [0.115,0.1183) | 51 | 0.0019829 | 0.1982893 | 22097 | 85.9136858 |
| [0.1183,0.1215) | 2765 | 0.1075039 | 10.7503888 | 24862 | 96.6640747 |
| [0.1215,0.1248) | 0 | 0.0000000 | 0.0000000 | 24862 | 96.6640747 |
| [0.1248,0.128) | 119 | 0.0046267 | 0.4626750 | 24981 | 97.1267496 |
| [0.128,0.1313) | 739 | 0.0287325 | 2.8732504 | 25720 | 100.0000000 |

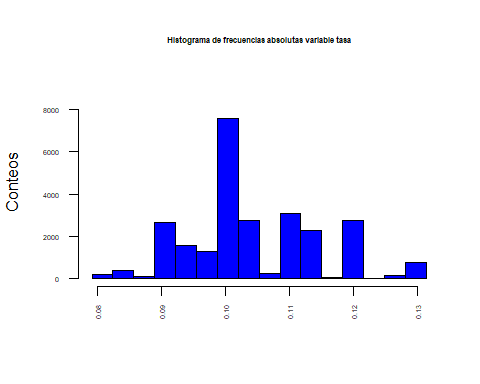
|  | x |
| --- | --- |
| start | 0.0792000 |
| end | 0.1313000 |
| h | 0.0032562 |
| right | 0.0000000 |

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "cex.names" is not  
## a graphical parameter

## Warning in axis(2, ...): "cex.names" is not a graphical parameter

## Warning in rect(brk[-length(brk)], 0, brk[-1], y, col = col, ...): "cex.names"  
## is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning in plot.window(...): "cex.names" is not a graphical parameter

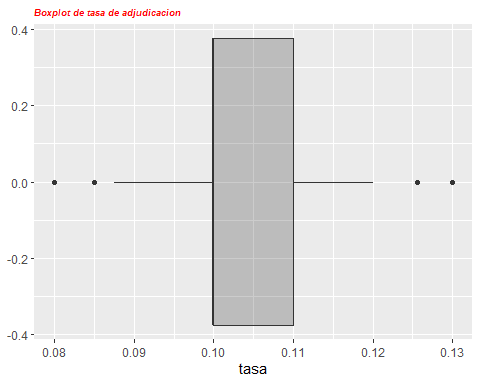
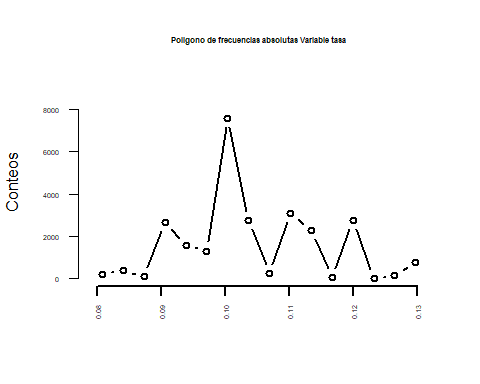
## Warning in plot.xy(xy, type, ...): "cex.names" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "cex.names" is not  
## a graphical parameter

## Warning in box(...): "cex.names" is not a graphical parameter

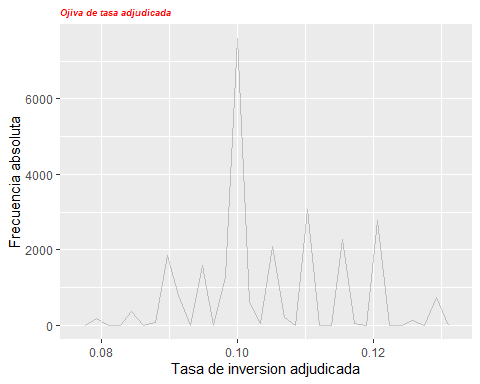
## Warning in title(...): "cex.names" is not a graphical parameter

## Warning in axis(1, at = round(brk, x.round), las = xlas, ...): "cex.names" is  
## not a graphical parameter



## Warning: Ignoring unknown parameters: fill

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 De acuerdo al analisis de la variable cuantitativa de tasa de adjudicación de la campaña podemos encontrar que hay dos rangos de tasas predominantes en cantidad de inversionistas, el 29.52% (n = 7595) de los inversionistas lograron tasas de adjudicacion de la inversion entre los 0.098 y el 0.102. En siguiente lugar podemos encontrar que el 12% de los inversionistas (n= 3068 )lograron tasas entre el 0.1085 y el 0.1118

## Resumen de medidas de localización

## ResumenVaribles ResumenPromedio ResumenMediana ResumenGeometrica  
## 1 Edad 37.1817652 35.0 36.0365311  
## 2 Tasa 0.1044049 0.1 0.1039105  
## 3 Ingresos Mes 7590004.2873250 4878000.0 5017014.3233656  
## 4 Patrimonio 157008240.3223561 100000000.0 64676964.5259231  
## 5 Monto Inversion 967367.5738725 500000.0 611385.6725858  
## 6 Cantidad Inversiones 12.1980949 8.0 7.1994345  
## ResumenArmonica ResumenRecortada ResumenWinsorizada  
## 1 35.0149957 35.9849339 36.4879471  
## 2 0.1034249 0.1040474 0.1044279  
## 3 3793266.2022598 5523128.3514288 6008502.6811431  
## 4 14267648.1084715 122361003.4177683 139643202.7342146  
## 5 430745.4551011 728999.5139969 813199.6111975  
## 6 3.9411647 10.0013122 11.0010498

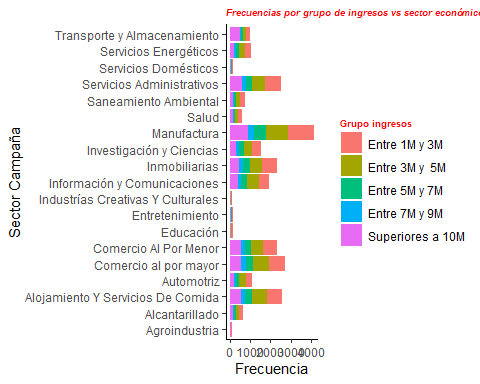
## Análisis Bivariado Cualitativo

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

A continuación el analisis bivariado cualitativo se realizará comparando las variables cualitativas contra la variable de grupo ingresos

## Warning: Use of `BDa2censo$GrupoIngresos` is discouraged. Use `GrupoIngresos`  
## instead.

## Warning: Use of `BDa2censo$SectorCampaña` is discouraged. Use `SectorCampaña`  
## instead.

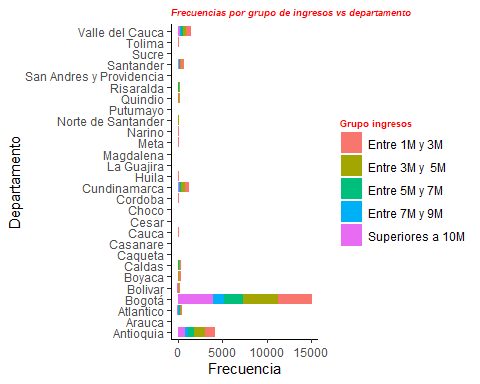


A partir del anterior grafico podemos observar la concentración de inversionistas con ingresos superiores a 10M de pesos en sectores económicos de manufactura y servicios administrativos, en los cuales la pacticipacion de personas con ingresos entre 3M y 5M también es relevante.

Resulta interesante observar que en general las empresas financiadas en a2censo que pertenecen al sector de manufactura Servicios administrativos e inmobiliarias, demuestran mayor traccion de inversionistas de alto capital y medio-bajo capital caraterizandose por tener inversionistas con ingresos superiories a 10 millones, y entre 3 y 5 MM

## Warning: Use of `BDa2censo$GrupoIngresos` is discouraged. Use `GrupoIngresos`  
## instead.

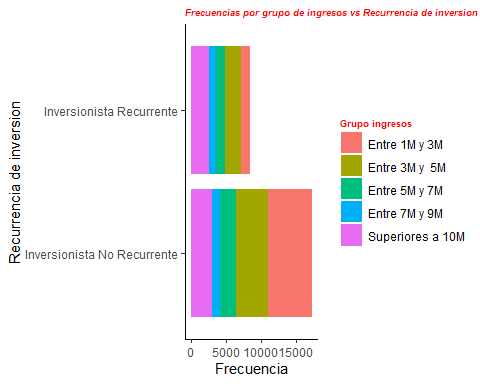
## Warning: Use of `BDa2censo$Departamento` is discouraged. Use `Departamento`  
## instead.



Lejos de encontrar relevante sacar alguna conclusión de acuerdo al rango de ingresos, de manera general podemos observar que departamentos como valle del cauca, Santander, Antioquia son regiones potenciales para atracción de inversionistas en donde a2censo no ha tenido una alta participacion a nivel de marketing o acercamiento con inversionistas. Para el caso de Bogotá podemos evidenciar como la proporsion de inversionistas con rangos de ingresos medios- bajos (3 a 5 millones) y medios altos (superior a 10 millones) es predominante

## Warning: Use of `BDa2censo$GrupoIngresos` is discouraged. Use `GrupoIngresos`  
## instead.

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

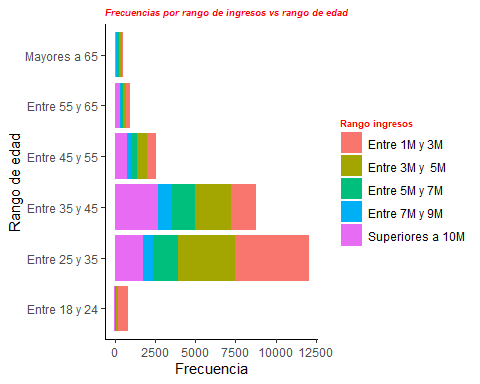


De acuerdo a la anterior grafica podemos observar como los inversionistas recurrentes son en mayor medida personas con ingresos superiores a 10 M, aun cuando hay una proporsion importante de inversionistas recurrentes con ingresos entre 3M y 5M que no son necesariamente de un alto rango de ingresos y aun asi estan viendo una oportunidad recurrente de invertir en el mercado

En el caso de los inversionistas no recurrentes podemos encontrar como una importante proporsion de inversionistas son personas con ingresos entre 1 M a 5 M

## Warning: Use of `BDa2censo$GrupoIngresos` is discouraged. Use `GrupoIngresos`  
## instead.

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.



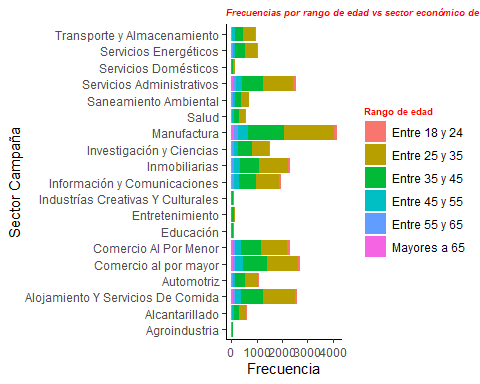
Realizando un analisis bivariado entre el grupo de edad y el rango de ingresos resulta interesante observar la participacion que tienen personas entre rangos de edad de 25 a 35 con ingresos entre 1 a 5 millones de pesos los cuales representan un volumen importante, no obstante volumenes similares se observan el personas con edades entre 35 y 45 años suelen ser en mayor medida personas con ingresos superiores a 10 millones de pesos.

#Visualización de las variables cualitativas desde la perspectiva de grupo de edad

A continuación el analisis bivariado cualitativo se realizará comparando las variables contra la variable de grupo de edad.

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

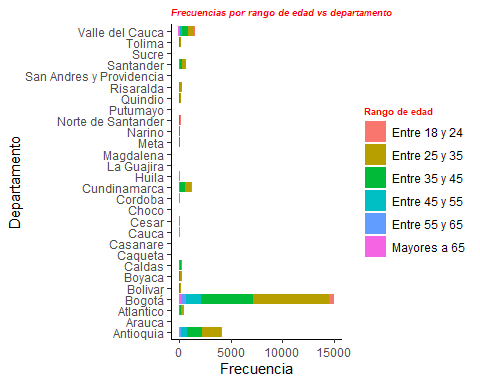
## Warning: Use of `BDa2censo$SectorCampaña` is discouraged. Use `SectorCampaña`  
## instead.



Podemos observar que de manera general independiente al sector economico de la empresa financiada, los grupos poblacionales que más invierten en a2censo se encuentran entre 25 y 45 años , no obstante se observa una interesante participación de personas mayores de 65 años en las campañas de tecnología , servicios alimenticios y salud

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

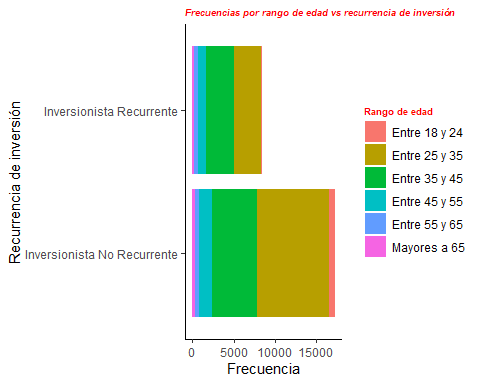
## Warning: Use of `BDa2censo$Departamento` is discouraged. Use `Departamento`  
## instead.



No es el grupo poblacional mas grande dentro de la comunidad de inversionistas, no obstante puede ser un grupo de atención las personas entre 45 y 55 años los cuales pueden ser un segmento a desarrollar en Bogotá, Antioquia y valle del cauca

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.



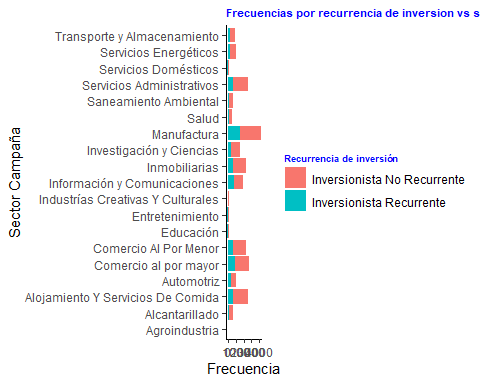
Del anterior grafico podemos concluir que para el caso de los inversionistas recurrentes hay una leve proporsion mayor de inversionistas en rangos de edades entre 35 y 45 años mientras que esta proporsion mayor de inversionistas no recurrentes de centra en personas entre 25 y 35 años, aun cuando ambos grupos de edades son predominantes estos dos grupos de edades

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

A continuación el analisis bivariado cualitativo se realizará comparando las variables contra la variable de categorización si es un inversionista recurrente o no lo es.

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

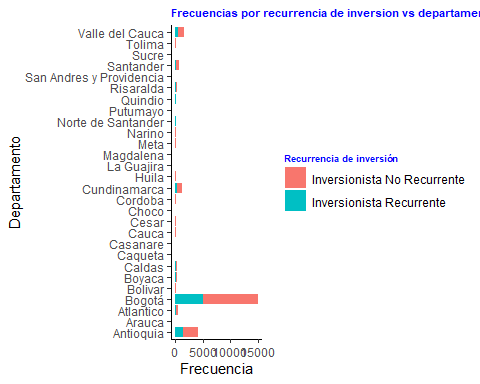
## Warning: Use of `BDa2censo$SectorCampaña` is discouraged. Use `SectorCampaña`  
## instead.



De manera general se puede observar que la cantidad de inversionistas NO recurrentes es mayor para todos los sectores economicos de las empresas financiadas, es una distribución muy acorde al numero de inversionistas sin considerar si es recurrente o no. No obstante llama la atención la proporsion grande que mantienen los inversionistas recurrentes en sector económico de manufactura

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

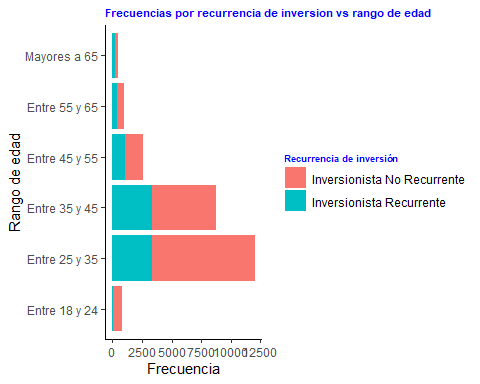
## Warning: Use of `BDa2censo$Departamento` is discouraged. Use `Departamento`  
## instead.



Se puede evidencia la alta proporsión de inversionistas recurrentes ubicados en Bogotá

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

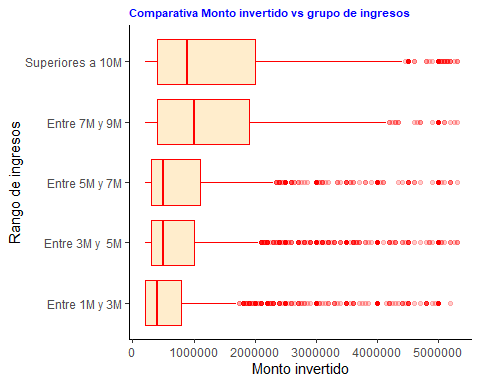


De la anterior grafica resulta interesante observar que actualmente la cantidad de inversionistas recurrentes son mas en personas entre 25 y 45 años, en donde podemos observar que para este rango de edad entre 25 y 35 años la mayor cantidad de inversionistas son no recurrentes

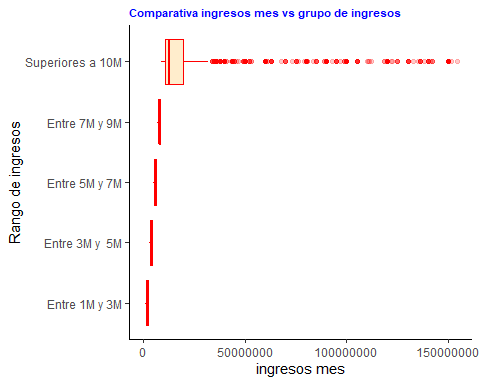
## 2. Resumen Bivariado entre variables Cualitativas y cuantitativas

A continuación se realizará un analisis Bivariado que contemple una de las variables cualitativas de mayor interés del estudio en cuestión, respecto a una de las variables cuantitativas evaluadas.

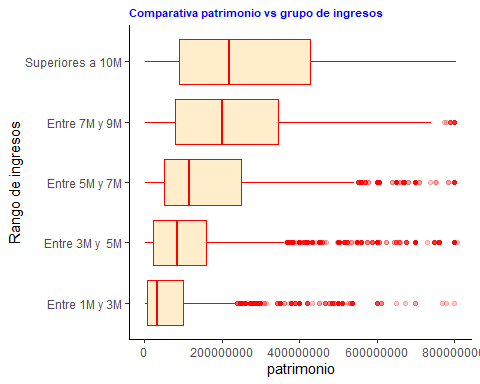
# 1. Analisis de las variables desde el monto invertido



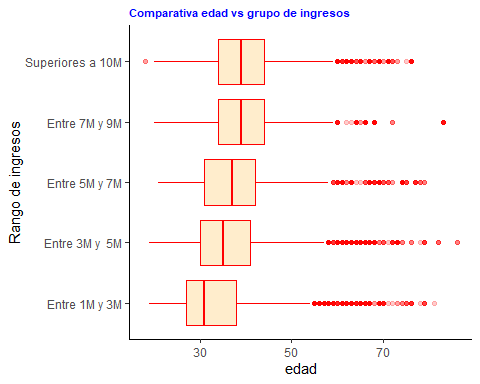
Del anterior diagrama de cajas y bigotes podemos concluir que aunque desde la correlacion lineal simple no se encuentra correlacion dada la alta dispersion de los datos que se tienen, a nivel del analisis de la media del monto invertido parece comportarse de manera acorde al rango de ingresos en donde podemos observar que a mayor rango de ingresos la media de monto invertido va incrementando proporsionalmente. El unico rango de ingresos que muestra un comportamiento atipico es el de los inversionistas con rangos de ingresos entre 7 y 9 millones mensuales, quienes muestran tener un promedio de inversión mayor incluso que las personas con ingresos mayores a 10 millones.



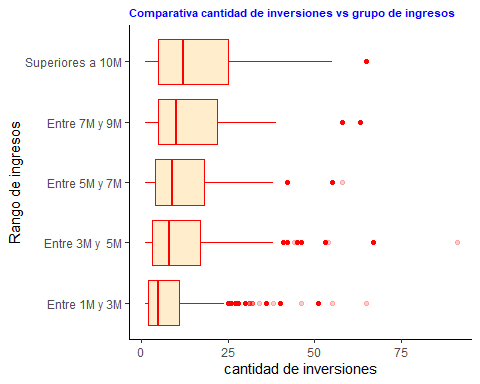
De manera similar al anterior grafico podemos encontrar que a) la mayor dispersión de datos respecto a ingresos mes se encuentra en las personas con ingresos superiores a 10 millones de pesos, no obstante desde el analisis de medidas de tendencia central y dispersión podemos observar menores desviaciones estandar en lo que respecto a ingresos menores a 10 millones de pesos



De manera similar, la medida de tendencia central es acorde a la conclusión de que un mayor promedio de patrimonio reportado es acorde a un mayor rango de ingresos, no obstante la dispersión no permitiría generalizar esta observación para todos los sujetos encontrandose sujetos con alto patrimonio en ingresos entre 1 y 7 millones, lo cual puede hablar de la calidad de la información que están diligenciando los inversionistas

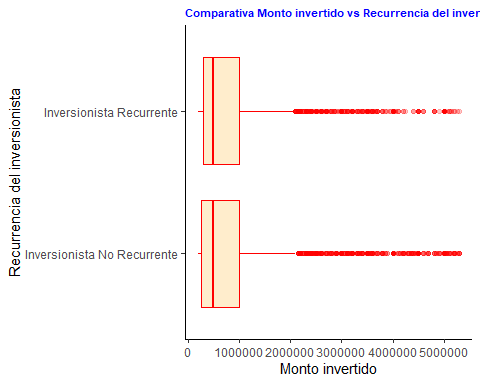


De acuerdo con el comportamiento que muchas veces se ha observado en las estadisticas Nacionales podemos observar que a mayor edad el promedio de ingresos es mayor encontrandose que el promedio de edad de nuestros inversionistas se encuentra en mayor medida entre los 30 y 40 ños

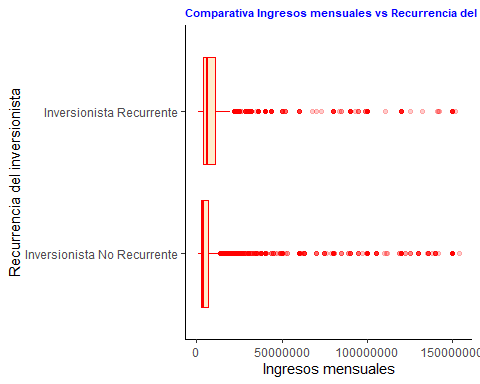


Si bien hay outliers en todos los rangos de ingresos, desde la tendencia central podemos ver que el mayor foco de inversionistas recurrentes se encuentra en personas con ingresos entre 7 a 10 millones de pesos

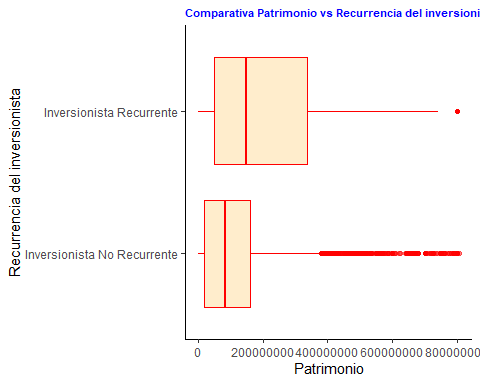
#2. Analisis desde el tipo de inversionista (Recurrente o no Recurrente)



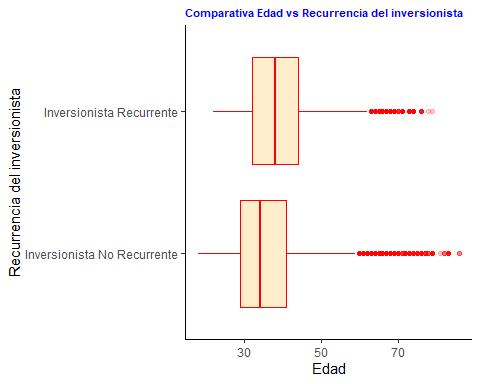
Desde el promedio podemos concluir que los inversionistas NO recurrentes invierten en promedio los mismos montos que inversionistas recurrentes, no obtante la dispersión de los datos no permite encontrar dicha relacion



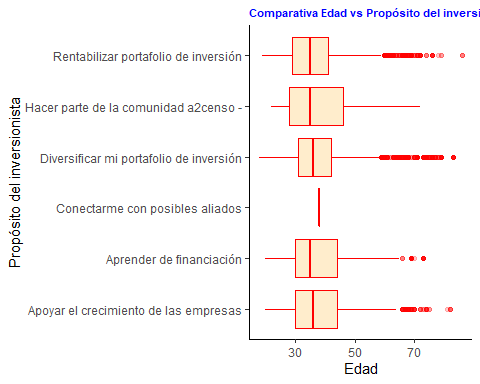
Los inversionistas recurrentes pueden llegar a tener un promedio de ingresos levemente mayor a los inversionistas no recurrentes



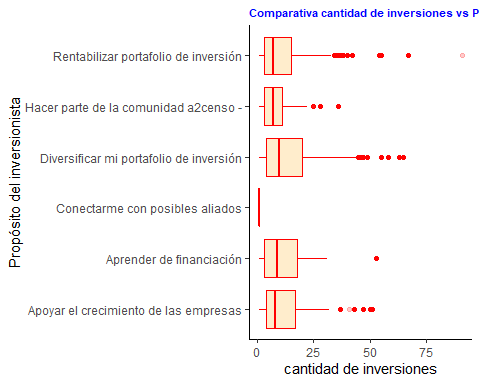
Los inversionistas Recurrentes llegan a tener un promedio de patrimonio superior, aun cuando hay personas consideradas inversionistas no recurrentes con un muy alto nivel de patrimonio



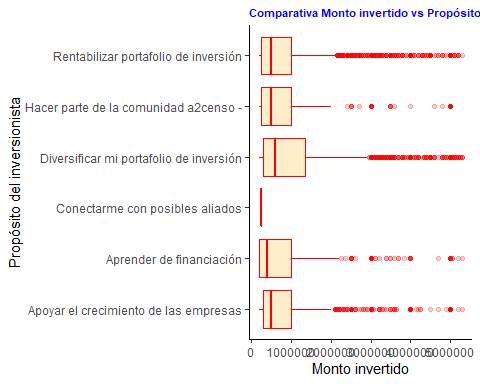
Los inversionistas recurrentes tiene un promedio de edad mayor, lo cual puede estar asociado con la siguiente grafica en donde observamos que en este rango de edad dichas inversiones recurrentes pueden estar asociadas con propositos de inversión como diversificar portafolio, aprender de financiación o sencillamente rentabilidad



Aqui podemos observar que no se evidencia alguna relación entre la edad y el propósito, es decir que tanto personas jovenes como viejas se distribuyen entre los diferentes propósitos de inversión.



Resulta interesante identificar que el promedio de numero de inversiones es superior respecto a propósitos de inversion asociados a la diversificación, por el contrario aquellos que han invertido con objetivos de conexión con otros posibles aliados son en promedio menos cantidades de inversiones

 Llama la atención observar que en general el promedio del monto invertido es mayor en personas con propósitos de diversificar el portafolio de inversión, en donde puede haber una mayor posibilidad de profesionalizar al inversionista.

## 3. Resumen Multivariado entre variables Cualitativas y cuantitativas

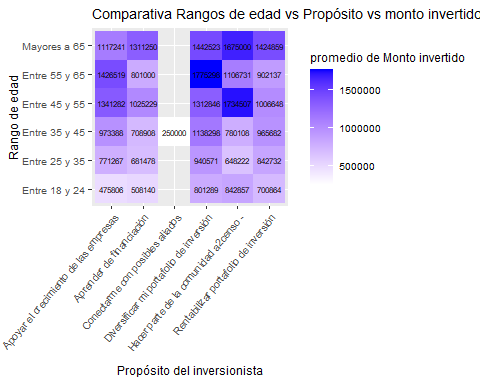
A continuación se realizará un analisis Multivariado que contemple dos de las variables cualitativas de mayor interés del estudio en cuestión, respecto a las variables cuantitativas evaluadas que permiten entender el comportamiento de inversión. Dichas variables son el monto invertido y la recurrencia de inversión

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

##   
## Attaching package: 'tidyr'

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

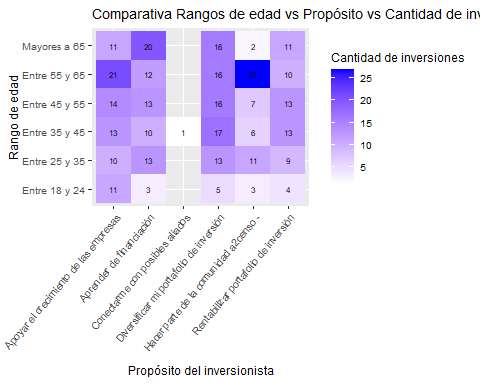
## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.



Del anterior grafico podemos observar como de manera general el promedio de monto invertido incrementa con el rango de edad, en donde podemos observar que los mayores montos promedios invertidos concuerdan con un interés de diversificar el portafolio de inversión en personas mayores de 55 años

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

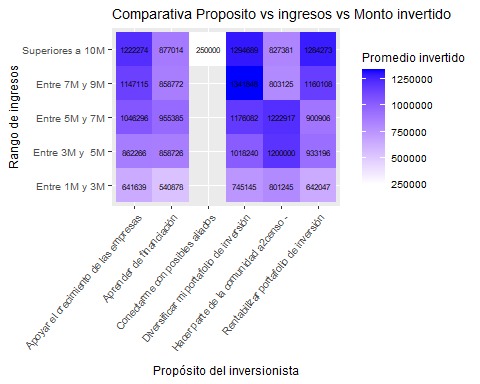
## `summarise()` has grouped output by 'Propósito'. You can override using the  
## `.groups` argument.



Respecto a la cantidad de inversiones realizadas resulta llamativo observar que personas entre 55 y 65 años tienen un proposito de afiliación a la comunidad de a2censo y en general son las personas con mayor recurrencia de inversiones desde los datos promedio de inversiones realizadas

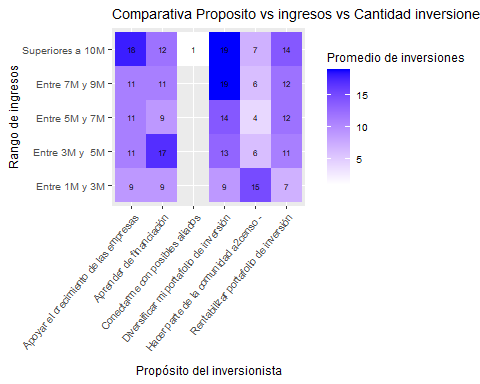
# Analisis de la cantidad de inversiones realizadas respecto al sector económico de la campaña y el propósito de inversión

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



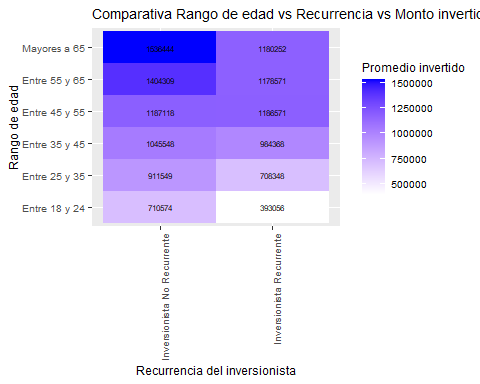
Desde los valores promedio de monto invertido resulta llamativo observar que los montos de inversión incrementan a medida que los rangos salariales tambien incrementan

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

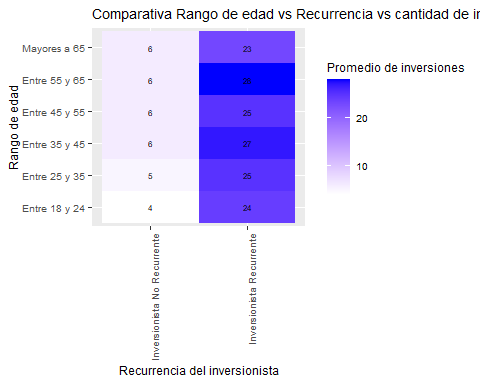
 Por otro lado si se analizan los rangos de ingresos y propósitos a la luz del promedio de inversiones realizadas podemos ver como personas con mayores ingresos y con interés de diversificar el portafolio de inversion son quienes mas inversiones recurrentes realizan. No obstante hay un grupo de personas de bajos ingresos entre 1 y3 millones mensualees con interéses de afiliación a la comunidad de a2censo que estan invirtiendo en promedio 15 veces, demostrando una recurrencia de inversiones de bajos montos

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

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

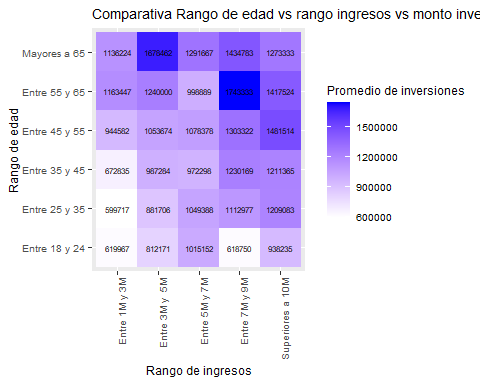
 Podemos encontrar como los montos promedios invertidos suelen ser mayores en inversionistas no recurrentes con rangos de edad superiores a los 55 años

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

 Podemos encontrar como en inversionistas recurrentes el promedio de inversiones es 6 inversiones respecto a mas o menos 20 inversiones que suelen realizar los inversionistas recurrentes

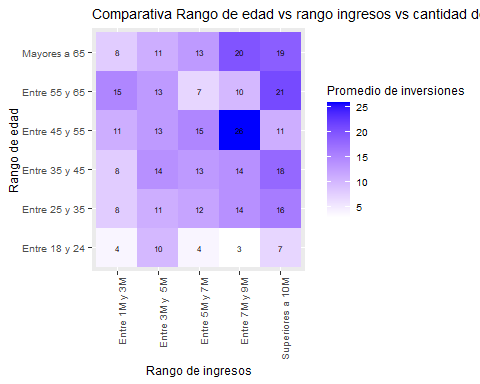
#Analisis desde los rangos de edad y rangos de ingresos a la luz de las variables cuantitativas de interés

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

 ahora bien, al realizar el analisis entre rangos de edad y de ingresos podemos observar como los mayores valores promedio de inversión se encuentran en personas con ingresos entre 3y 5 millones de mas de 65 años. personas con ingresos entre 7 y 9 millones con rangos de eedad entre 55 a 65 y las personas con ingresos de mas de 10 millones de pesos.

Analisis

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



Finalmente podemos observar que de acuerdo al valor promedio de inversiones realizads, los grupos con mayor recurrencia de inversión se encuentra en las edades entre 45 y 55 años con ingresos entre 7 y 9 millones. Por otro lado personas con ingresos superiores a 10 MM suelen invertir de manera más recurrente.

# 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

## [1] 0.144479

## [1] 0.09333871

## [1] 0.2583145

## [1] -0.02984937

Se encuentra que desde una correlacion lineal simple no hay correlación entre ninguna de las variables a la luz de la variable monto inversión.

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

## [1] 0.1549182

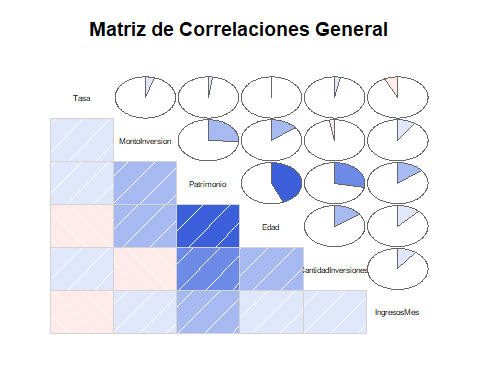
## [1] 0.09968228

## [1] 0.2861183

## [1] -0.02984937

Se encuentra que desde una correlacion lineal simple no hay correlación entre ninguna de las variables a la luz de la variable cantidad de inversiones.

#Grafico General de Correlaciones lineales

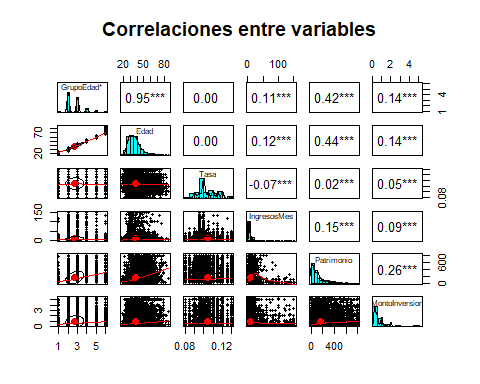
 En el anterior grafico podemos observar que aunque dichas correlaciones no son significativas la relación entre las variables Ingresos mes y patrimonio y la relación entre las variables patrimonio y edad son más fuertes respecto a las demás.

# Paso Cero:

Para facilidad de interpretación de las gráficas se pretende manejar una escala lo más similar posible por lo tanto se transformarán las variables de: ingreso mes , monto de inversión y patrimonio a millones dividiendo en un millón.

# Paso 1 :

Realizamos un diagrama de las variables cuantitativas con sus distribuciones y correlaciones, para identificar magnitud y sentido de la relación entre ellas



Luego de analizar las diferentes correlaciones entre variables cuantitativas identificamos que la mayor correlación se presenta entre el patrimonio y la edad (0.44) le sigue la correlación que se da entre la cantidad de inversiones con el patrimonio (0.29 ) y la siguiente es la correlación que se da entre monto de inversión y patrimonio con un 0.26

Dentro de los objetivos que se han planteado en este proyecto se pretende encontrar un modelo que pueda predecir el valor del monto de inversión dado las variables dependientes con las que se cuenta. Viendo que el monto de inversión presenta una correlación alrededor del 0.26 (aunque es una correlación moderada), plantearemos preliminarmente un modelo de regresión lineal simple que relacione precisamente estas dos variables patrimonio y monto de inversión.

# Paso 2:

Modelo de Regresion Lineal Simple

##   
## 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

## (Intercept) Patrimonio   
## 0.725892957 0.001537974

El modelos obtenido tiene la forma:

MontoInversion = 0.72 + Patrimonio\*0.0015 Evidenciamos que aunque tanto el coeficiente como el intercepto resultan ser significativos dentro del análisis, el R2 ajustado que nos arroja este modelo es muy bajo (6.6%). Es decir que solo el 6.6% de porcentaje de variación en la variable de respuesta ( monto inversion) es explicado por su relación con la variable predictora (Patrimonio)

Dado a este bajo desempeño del modelo de regresión lineal simple, buscaremos alternativas con un modelo de regresión múltiple incluyendo las otras variables o información disponible dentro de nuestra data. Intentaremos encontrar o predecir el monto de la inversión incluyendo la relación que puede existir con la edad del inversionista, con los ingresos mensuales, con la tasa a la cual se pacta el retorno de su inversión y otras variables que están bajo el análisis. Esperamos que con esto el desempeño del modelo sea mucho mejor.

# paso 3:

Modelo de Regresión Lienal Multiple

##   
## 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

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

El modelo de regresion multiple obtenido tiene la forma:

MontoInversion = 0.086 + Edad*0.004 + Tasa*5.31 + IngresosMes *0.0058 + Patrimonio*0.0015 - CantidadInversiones\*0.0104

Luego del análisis, podemos evidenciar que tanto el intercepto como los coeficientes de las variables predictoras: edad, tasa.ingresos mes, patrimonio, cantidad de inversiones son significativos ( a un nivel de significancia del 0.5%) la medida de desempeño R2ajustado sigue siendo muy bajo. Para este caso del modelo de regresión múltiple nos dio un 8.6% aproximadamente ( Solo el 8.6% de la variabilidad en los datos se puede explicar por el modelo). Es una leve mejoría con el modelo presentado anteriormente, pero va en contra del principio de parsimonia ya que hemos agregado y cuatro variables adicionales y la mejora fue prácticamente insignificante.

Propendiendo por el principio de parsimonia aplicaremos el algoritmo step para encontrar quizás una mejor combinación de variables que nos permitan un mejor índice de pérdida de información o AIC para un modelo de regresión múltiple

Luego de correr el algoritmo evidenciamos que el modelo que nos propone con menor pérdida de información AIC 1182.07 es el modelo que previamente hemos definido (modeloMultiple1), que incluye todas las variables predictoras cuantitativas

## 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 Es muy importante al momento de realizar un modelo de regresión lineal validar el cumplimiento de los supuestos ya que podemos ya que sin esta validación podemos caer en la estimación de variables no confiables

## 1. Multicolinealidad

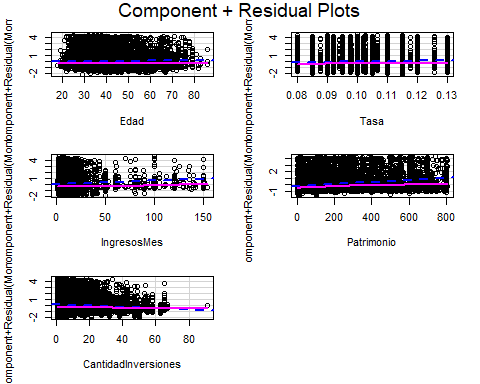
Los predictores deben ser independientes,no debe de haber colinialidad entre ellos , para validar esto calculamos FACTOR DE INFLACION DE LA VARIANZA(VIF)

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

Este resultado para todas las variables predciotres no esta dando ligeramente mayor a 1 , por lo tanto hay un indicio de una posible colinialidad. Es relvante seguir evaluando los demas supuestos, par atebner la certeza q no se cumple y q no se podria usar un modelo de regresion como herramienta predicora.

Para evaluar la influencia que tienen los predictores sobre la variable dependiente

## 2. Relacion Lineal entre predictores y variable respuesta

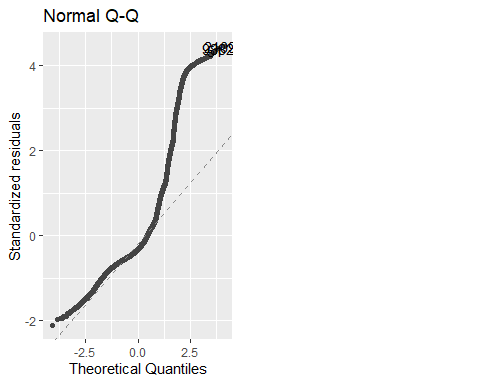
 Luego de graficar los residuos para cada variable predictora podemos identificar comportamientos aleatorios sobre el valor cero pero de alguna manera desiguales en los sentidos positivo y negativos. La gran mayoría de los residuos para cada variable predictora se encuentran por encima cero, No se cumple que se distribuyan en igual proporción alrededor del cero esto indicaría que este supuesto no se cumpliría

##3. Distribuciónnormaldelosresiduos

La media de los residuos nos da aproxidamente cero (-1.30e-16) , mas sin embargo la distribucion de los residuos estudentizados presentean una comportamiento anomalo ( gran parte de los residuos superan el limite de 2). Este comporamtiento lo podemos evidenciar en la grafica Normal QQplot donde es calro que los residus no se ajustan sobre la diagonal de los Quantiles teoricos , por lo tanto este supuesto tampoco se satisface.

## [1] -0.0000000000000001302145

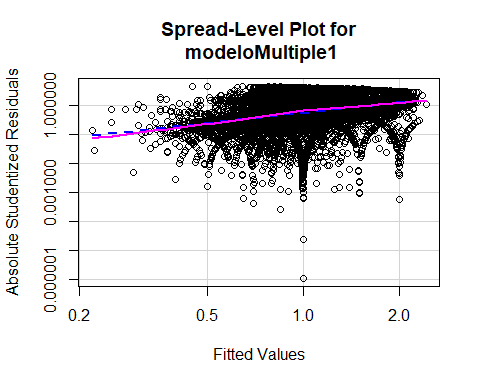
# Si la relación es lineal,los residuos se distribuyen de forma aleatoria entorno a cero



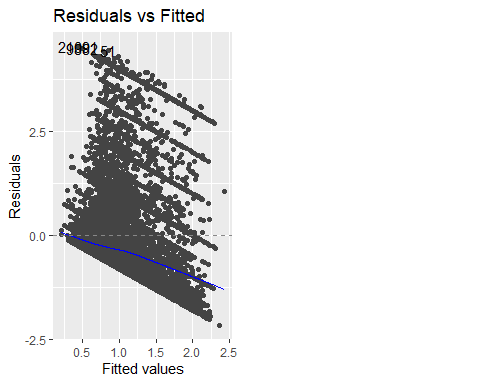
##4.Homocedasticidad

Para valdiar este supuesto acudiremos a realizar el test de Breusch-Pagan, el cual nos arrojo un p-value de 2.2e-16 , lo cual nso permite rechazr la hipotesis nula q existe varibilidad constante de los residuos. Es decir q no se presenta homcedasticidad, imcumpliendo asi el supuesto . Adcionalmente en la grafica podemos observar un incremnto en valor absoluto de los residuos estudentizados, confirmando la heterocedasticidad de los residuos.

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



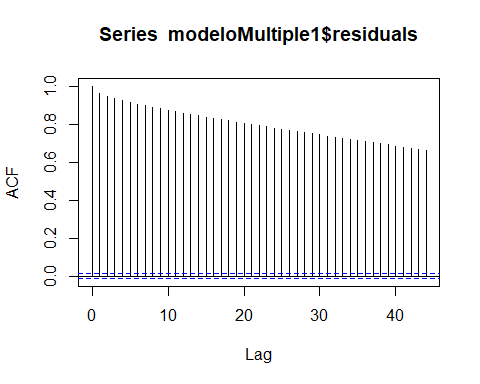
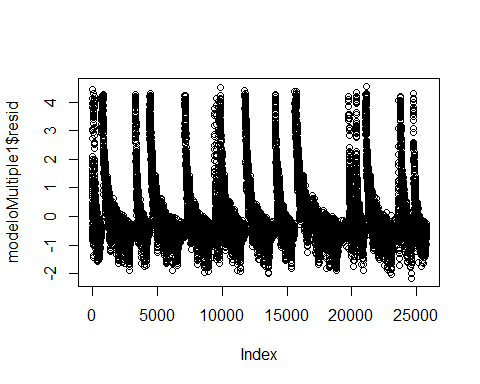
##   
## Suggested power transformation: -0.2003072



## 5.Independencia

Para validar esto supuesto , realizamos prueba Durbin Watson y evidecmoa q el p-value nos da ceor por lo tanto se rechaza la hipotesis nula , tambien podemos ver graficamente la que si existe autocorrelacion

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



Con animo ilustrativo y para confirmar se realiza una representación 3D de la regresión con solo dos variables predictoras ( edad y patrimonio)

## Regresión logistica

De acuerdo a los objetivos planteados para el presente estudio existe un alto interés por parte de la compañía en conocer si existe alguna variable que nos permita predecir la fidelización de los clientes con el producto, entendiendo que la mejor variable que nos puede hablar de fidelización es la recurrencia de inversión. Por esto se analizarán las variables para identificar si alguna de ellas o un conjunto de ellas permiten predecir el comportamiento de inversión referente a la recurrencia.Siendo asi plantearemos un modelo de regresion logisitica que no permita estimar la probabilidad de recurrencia de un inversionista.

##   
## Attaching package: 'magrittr'

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

##   
## 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

Desde el modelo de regresion logistica ajustado se encuentra que el sector económico de la campaña es una variable con un nivel de significancia en la variable respuesta de recurrencia del inversionista. De la misma forma en que muestra significancia respecto a los propósitos de inversión de diversificacion del portafolio y el motivador de hacer crecer el capital.

De otro lado, las variables, monto inversión, edad, patrimonio e ingresos mes tambien muestran significancia respecto a la variable asociada a si el individuo es inversionista recurrente o no.

Probaremos con otros modelos descartando algunas variables con el animo de comparar el criterio de perdida de informacion AIC y escoger el mejor modelo

##   
## 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

##   
## 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

De acuerdo a lo anterior encontramos que el modelo en donde menor pérdida de información es el designado con nombre: logMultiple , en donde el AIC es de 22826. Asociado a las variables MontoInversion + Edad + IngresosMes + Patrimonio + SectorCampaña + Propósito . No obstatne dentro de las variables cualitativas no todas sus categorías tienen un nivel de significancia relevante, por ejemplo para la variable sector de campaña, las categorías de variables con mayor nivel de significancia son: Educación, Información y Comunicación, y Servicios Domésticos. Para el caso del propósito se destacan: Diversificar Portafolio y Pertenecer a la comunidad de a2censo.

## Resultando un modelo de la forma:

Inv\_Recurrente = e (-1,925-0,0000002243*MontoInversion+0,01733*Edad+0,000000009223*IngresosMes+0,000000002252*Patrimonio+0,2879*SectorCampañaAlcantarillado-0,02954*SectorCampañaAlojamiento Y Servicios De Comida+0,51*SectorCampañaAutomotriz+0,3723*SectorCampañaComercio al por mayor+0,04463*SectorCampañaComercio Al Por Menor+1,003*SectorCampañaEducación+0,5915*SectorCampañaEntretenimiento+0,6088*SectorCampañaIndustrías Creativas Y Culturales+0,7096*SectorCampañaInformación y Comunicaciones-0,006026*SectorCampañaInmobiliarias+0,1012*SectorCampañaInvestigación y Ciencias+0,4045*SectorCampañaManufactura+0,4026*SectorCampañaSalud-0,2092*SectorCampañaSaneamiento Ambiental-0,2422*SectorCampañaServicios Administrativos+0,8118*SectorCampañaServicios Domésticos+0,2204*SectorCampañaServicios Energéticos+0,1203*SectorCampañaTransporte y Almacenamiento+0,1616*PropósitoAprender de financiación+0,3406*PropósitoDiversificar mi portafolio de inversión-0,4473*PropósitoHacer parte de la comunidad a2censo-0,03824*PropósitoRentabilizar portafolio de inversión) / (1+e(-1,925-0,0000002243*MontoInversion+0,01733*Edad+0,000000009223*IngresosMes+0,000000002252*Patrimonio+0,2879*SectorCampañaAlcantarillado-0,02954*SectorCampañaAlojamiento Y Servicios De Comida+0,51*SectorCampañaAutomotriz+0,3723*SectorCampañaComercio al por mayor+0,04463*SectorCampañaComercio Al Por Menor+1,003*SectorCampañaEducación+0,5915*SectorCampañaEntretenimiento+0,6088*SectorCampañaIndustrías Creativas Y Culturales+0,7096*SectorCampañaInformación y Comunicaciones-0,006026*SectorCampañaInmobiliarias+0,1012*SectorCampañaInvestigación y Ciencias+0,4045*SectorCampañaManufactura+0,4026*SectorCampañaSalud-0,2092*SectorCampañaSaneamiento Ambiental-0,2422*SectorCampañaServicios Administrativos+0,8118*SectorCampañaServicios Domésticos+0,2204*SectorCampañaServicios Energéticos+0,1203*SectorCampañaTransporte y Almacenamiento+0,1616*PropósitoAprender de financiación+0,3406*PropósitoDiversificar mi portafolio de inversión-0,4473*PropósitoHacer parte de la comunidad a2censo-0,03824*PropósitoRentabilizar portafolio de inversión))

## Predicciones

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

Tambien definimos el valor del punto de corte o treshold para determinar la clasificaicon de la variable respuesta, con la funcion optCutOff, q nos permite determinar de manera automatica el punto de corte para la clasificacio de las observaciones del modelo propuesto.

## [1] 0.4576869

## matriz de confusion

Por medio de la siguente matriz de confusion podemos evalaur el ajsute del modelo , tomado como criterio de corte el valor estimado 0.457 en el paso anterior.

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

## Sensibilidad y especificidad

## [1] 0.2728134

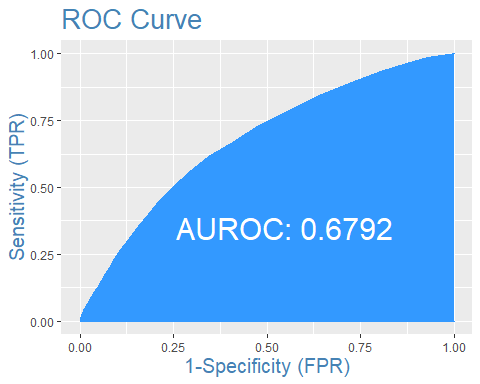
Se concluye que el modelo acertó en el 27% de las observaciones de verdaderos positivos ( Grado de Sensibilidad), es decir solo identifica un 27% de los positivos reales

## [1] 0.890707

La especificidad mide la tasa de verdadero negativos, es decir para este caso el modelo identifica el 89% de los negativos reales

## Curva ROC

La curva característica operativa o curva ROC permite identificar el rendimiento de un clasificador, en este caso el modelo de regresión logística. El área bajo la curva mide la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos (1-Especificidad) en varios valores de umbral



El criterio de precision del modelo concluye que el 68% de las observaciones son acertadas. Y asu vez son evidencia de un modelos q no presta un nivel adecudo para la toma de decisiones.

A pesar de haber escogido un modelos con el menor criterio de pérdida de informacion aún el nivel de precision no es suficiente para tomar decisiones fiables en el futuro.

# Muestreo

## [1] 37.04

## [1] 92.59374

## [1] 0.18

## [1] 0.42

## [1] 1.13

## [1] 92.59 36.21

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

## [1] 37.18177

## [1] 0

## [1] 0

## [1] 0

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

## [1] 0.9673676

## [1] 24194.8

## [1] 4.91

## [1] 24880.69

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

## 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: 1.5-31, (SVN revision 1171)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 3.4.1, released 2021/12/27  
## Path to GDAL shared files: C:/Users/alexg/Documents/R/win-library/4.1/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:/Users/alexg/Documents/R/win-library/4.1/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

## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1; sf\_use\_s2() is TRUE

## 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: "C:\Users\alexg\Documents\GitHub\ProyectoEstadistica\ProyectoEstadistica\Sergio\MGN2021\_DPTO\_POLITICO", layer: "MGN\_DPTO\_POLITICO"  
## with 33 features  
## It has 9 fields

## Reading layer `MGN\_DPTO\_POLITICO' from data source   
## `C:\Users\alexg\Documents\GitHub\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

##   
## 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

##   
## Attaching package: 'leaflet'

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

## 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

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

## 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.