Docente: Jaime Lincovil



**CATEDRA N°1 PROGRAMACION EN R**

**Alumno**

**Carola Soto Oberreuter**

[**carolasotooberreuter@gmail.com**](mailto:carolasotooberreuter@gmail.com)

**github: https://github.com/ayayena/**

Introducción

Una de las tareas habituales a las que se enfrenta todo analista o científico de datos es la de entender las características de las variables con las que se está trabajando.

Para este trabajo se pretende hacer un análisis exploratorio de datos, que permita identificar posibles errores, datos incorrectamente introducidos ó la ausencia de estos , revelar la presencia de valores atípicos (outliers).

Además de poder visualizar la relación existente entre las variables, poder ver los aspectos más significativos de las variables , mediante resúmenes o representaciones gráficas

Definición del problema/objetivo de investigación

Se requiere analizar el comportamiento de los datos que tienen relación a las atenciones de urgencia en los establecimientos de atención de salud pública durante el año 2023, en el país de Chile.

Dado que diariamente acuden a los establecimientos de salud personas de diversas edades y consultas por diversas causas, el analizar posibles tendencias en relación a la a edad o al establecimiento donde se producen más o menos consultas, permitirán deducir conclusiones que pueden transformarse en la base para la implementación de nuevas acciones ó nuevos análisis

1. Introducción al análisis

Para el análisis requerido se han obtenido los datos que contiene la información de las atenciones de urgencias realizadas en el año 2023 de los establecimientos de salud pública.

Este conjunto de datos se encuentra en los datos abiertos en la página oficial del Departamento de Estadísticas e Información (DEIS), dependiente de la Subsecretaría de Salud Pública en el Ministerio de Salud de chile.

En el conjunto de datos de encuentran los datos que corresponde a la identificación de los establecimientos que realizaron la atención, la categoría del establecimiento, la comuna a la que pertenece, la fecha de la realización, la causa de atención

|  |  |
| --- | --- |
| Nombre | Descripción |
| IdEstablecimiento | Codigo del Establecimiento |
| NEstablecimiento | Nombre del establecimiento |
| IdCausa | Código de categoría de la causa de atencion |
| GlosaCausa | Descripcion de la causa de atencion |
| Total | Numero total de atenciones |
| Menores\_1 | Numero total de atenciones de pacientes menores de 1 año |
| De\_1\_a\_4 | Numero total de atenciones de pacientes menores que están entre 1 y 4 años |
| De\_5\_a\_14 | Numero total de atenciones de pacientes menores entre 5 a 14 años |
| De\_15\_a\_64 | Numero total de atenciones de pacientes entre 15 a 64 años |
| De\_65\_y\_mas | Numero total de atenciones de pacientes mayores de 65 años |
| fecha | Fecha de la atención |
| semana | Corresponde a número de semana de la atención de urgencia |
| GLOSATIPOESTABLECIMIENTO | Cear:Centro Exclusivo de Atencion Respiratoria  Sapu:Servicio de atención primaria de urgencia  SAR: Servicio de atención primaria de urgencia de alta resolutividad  SUR:Servicio de urgencia rural |
| GLOSATIPOATENCION | Categoría de atención adulto, indiferenciado , infantil |
| GlosaTipoCampana | Corresponde a la identificación si corresponde a una campaña |
| CodigoRegion | Corresponde al código de la región a la que pertenece |
| NombreRegion | Correponde a la región a la que pertenece |
| CodigoDependencia | Corresponde al de la dependencia administrativa a la que pertenece el establecimiento, es decir el servicio de salud |
| NombreDependencia | Corresponde a la descripción de la dependencia administrativa a la que pertenece el establecimiento, es decir el servicio de salud |
| CodigoComuna | Código de la comuna a la que pertenece el establecimiento |
| NombreComuna | Descripción de la comuna a la que pertenece el establecimiento |

Los objetivos del análisis es poder ver el comportamiento de los datos y poder dar respuesta las siguientes situaciones

* Poder identificar en que establecimientos de la región de los lagos se producen más atenciones durante el año
* Identificar en que periodos del año se producen consultas por atenciones de urgencia en los establecimientos de la región de los lagos
* Identificar cual es el grupo de personas que consultan las urgencias, niños, adultos o mayores de edad

1. Preprocesamiento de Datos

Para el manejo de las observaciones y sus variables se transformó todo el conjunto de datos en un tipo de datos dataframe.

Se evaluó si el conjunto de datos contenía en su interior el tipo de datos perdidos na o la ausencia de estos comprobando la existencia de datos null

Dado que en los registros numéricos el ser humano tiende a confundir el número cero con la letra “o”, algunos registros pueden presentar este carácter en una variable numérica se decidió verificar la presencia del carácter “o” y remplazar el valor en la situación de encontrarlo.

Al igual que los datos perdidos a la existencia de estos, se decidió remplazar estos datos por el valor cero en las variables cuantitativas encontradas en el dataset.

Se trabajaron las variables cualitativas con caracteres especiales

Dado los objetivos del análisis se necesita trabajar con una las observaciones requeridas que corresponden a la región, se obtiene un resumen de las observaciones de la región.

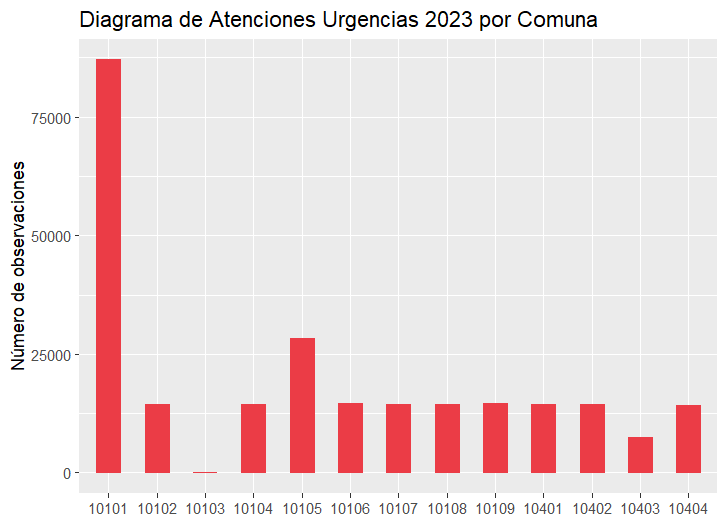
1. Análisis Exploratorio de Datos

Se analizaron los datos delas variables categóricas por comuna, establecimiento y fecha

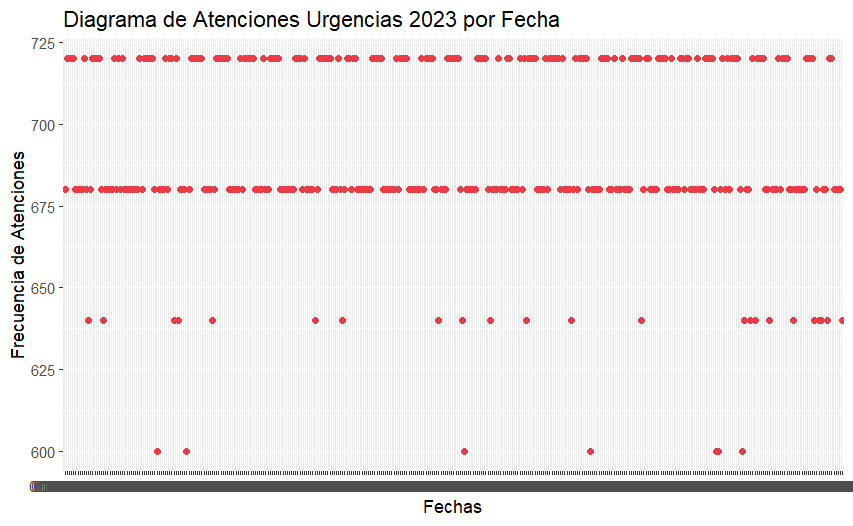
La variable categórica Comuna nos nuestra que la comuna de Puerto Montt es la que presentó más atenciones de urgencia durente el año 2023.

Los rangos de las atenciones de urgencia van entre las 40 y 87040 durante el año

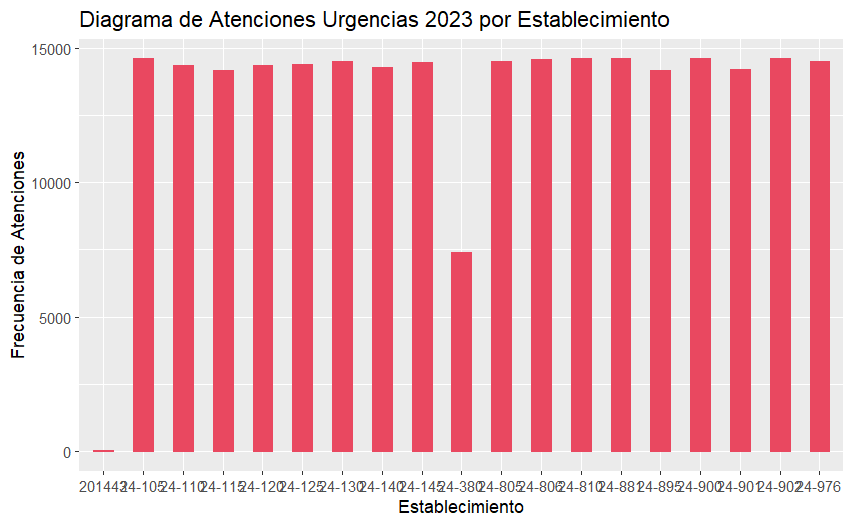
|  |  |
| --- | --- |
| Comuna | Fecuencia |
| Calbuco | 14520 |
| Chaiten | 14520 |
| Cochamo | 40 |
| Fresia | 14360 |
| Frutillar | 28360 |
| Futaleufu | 14480 |
| Hualaihue | 7400 |
| Llanquihue | 14360 |
| Los Muermos | 14600 |
| Maullin | 14400 |
| Palena | 14280 |
| Puerto Montt | 87040 |
| Puerto Varas | 14600 |



Al analizar el conjunto de datos por fecha se puede observar que las atenciones son uniformes en cuanto a los días. Se visualiza que las atenciones están entre las 600 y 725 atenciones diarias en los establecimientos dependientes de la red del servicio de salud Reloncaví.

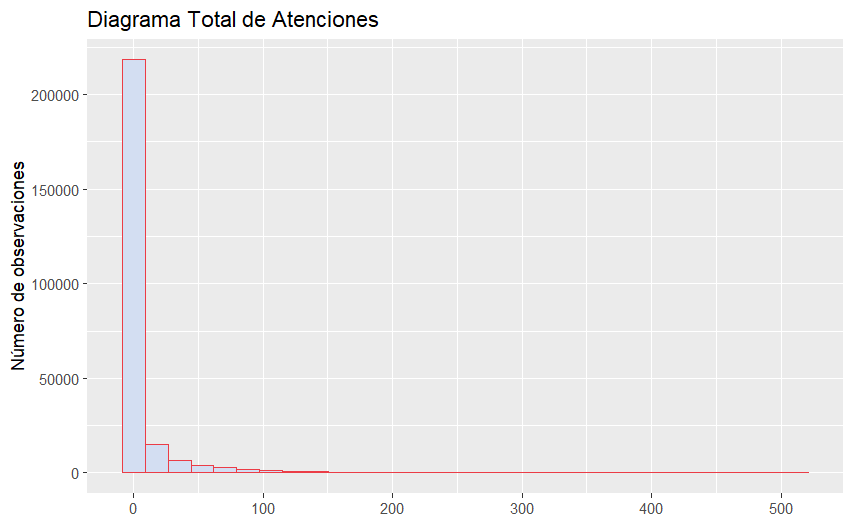


Analizando el conjunto de datos considerando la variable establecimiento se observa que las atenciones en frecuencia son similares en la gran mayoría de los establecimientos, con excepción de dos de estos, que muestran un número menor de atenciones



Las variables cuantitativas analizadas corresponden a el total de urgencias y las atenciones que fueron realizadas a los menores de 1 año, estas variables corresponden a valores discretos y se observa una clara tendencia que se registran mayor cantidad de observaciones con poca frecuencia de atenciones de urgencia.

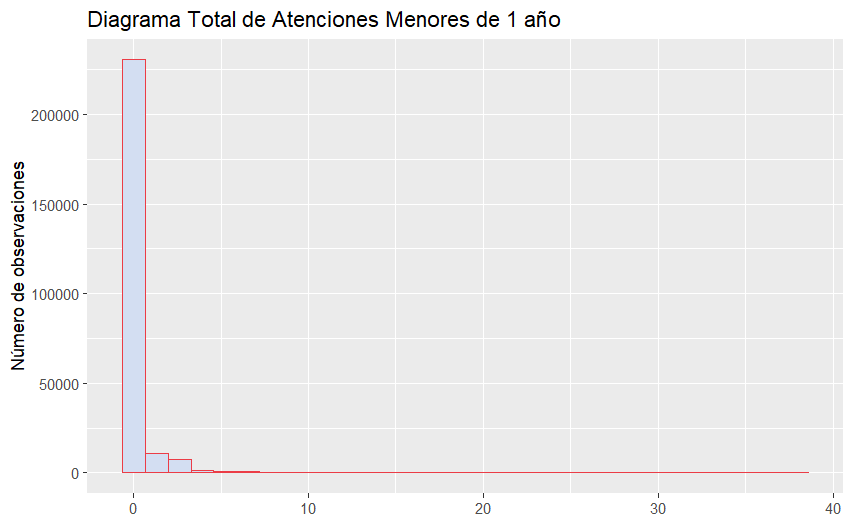
Para la variable Total de atenciones se registran mayoritariamente valores ceros, el rango de atenciones va desde 0 a 513 , siendo la media 7.701 atenciones . La moda presenta valor cero y desviación estándar es de 26.71666



**Tabla de medidas de tendencia central y medidas de posición variable cuantitativa Totales atenciones Urgencia**

|  |
| --- |
| Min. 1st Qu. Median Mean 3rd Qu. Max. |
| 0.000 0.000 0.000 7.701 2.000 513.000 |

Para la variable Menores de 1 año de atenciones se registran mayoritariamente valores ceros, el rango de atenciones va desde 0 a 38, siendo la media 0.2374 atenciones. La moda presenta valor cero y desviación estándar es de 1.206614



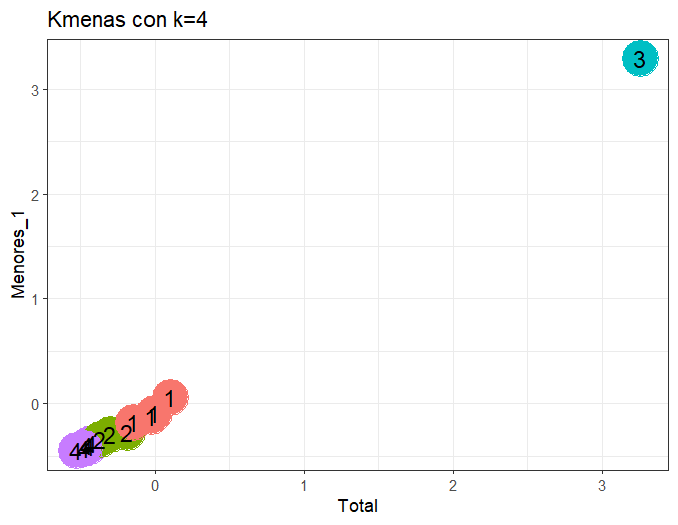
**Tabla de medidas de tendencia central y medidas de posición variable cuantitativa Menores de 1 año atenciones Urgencia**

|  |
| --- |
| Min. 1st Qu. Median Mean 3rd Qu. Max. |
| 0.0000 0.0000 0.0000 0.2374 0.0000 38.0000 |

1. Modelado de Datos

Se utilizó el algoritmo de k-means para analizar el comportamiento de grupos de la variable total de atenciones de urgencia en relación a las atenciones de menores de 1 año, en los establecimientos de la administración dependiente del servicio de salud del Reloncavi.

Se consideraron 4 grupos en el algoritmo y se observó que los datos de atenciones totales y menores de 1 año se agrupan mayoritariamente en frecuencias menores tendientes a cero.



1. Interpretación de Resultados

Del análisis del conjunto de datos de atenciones de urgencia año 2023, se observa cómo se mencionó anteriormente a bajas frecuencias, donde se encuentra mayoritariamente el valor cero.

Y al observar el comportamiento de los datos por fecha de atención , se observa que los datos son uniforme en algunos días, hay variabilidad pero se visualiza una tendencia lineal, se ven claramente 4 grupos de frecuencias en las atenciones.

En relación a la cantidad de atenciones por comunas dependientes del servicio de salud del Reloncaví, se observa un número de atenciones de urgencia mayor en el establecimiento de la comuna de Puerto Monnt, superior a los otros establecimientos de la administración.

Y en relación a las atenciones por establecimiento se visualiza que hay una frecuencia similar en los establecimientos.

Cuando se analizó el comportamiento por grupo tambien se observe una tendencia a frecuencia de bajo valor y presencia mayoritamiente del valor cero.

1. Código

library(RPostgreSQL)

library(readr)

library(dplyr)

library(ggplot2)

############### DIRECTORIO ###############################################

setwd("D:/Magister Ciencias Datos UDLA/Clases R/Practicos R/MagisterCsDatos/")

source("D:/Magister Ciencias Datos UDLA/Clases R/Practicos R/MagisterCsDatos/funcionesEvaluacion.R")

############COMIENZO SCRIPT ##############################################

repositorio="local"

atenUrgencia=data.frame()

if(repositorio == "DB")

{

conex=recolecion\_Datos("localhost","5432","Urgencias","postgres","rootpostgres")

#dbListTables(conex)

atenUrgencia=dbGetQuery(conex, 'select \* from AtencionesUrgencia2023')

}else

{

AtenUrgencia <- read\_delim("AtencionesUrgencia2023.csv",

delim = ";", escape\_double = FALSE, trim\_ws = TRUE)

}

############# DESCRIPCION DEL CONJUNTO DE DATOS#################################

# Se obtiene el nombre de la variables

names(AtenUrgencia)

# Se visualiza el data set que contien los datos de las atenciones

View(AtenUrgencia)

# Se obtiene un resumen de las variables que tiene el data set

glimpse(AtenUrgencia)

#se obtiene los tipos de datos que contiene el dataset de atenciones

str(AtenUrgencia)

############### PREPROCESAMIENTO DE DATOS######################################

#se tranforma el data set en tipo de datos dataframe

dataAten<-data.frame(AtenUrgencia)

#veirificar datos nulos

dnulos=verificaNulos(dataAten)

dnulos

if (dnulos == FALSE)

{

#busca registros nulos y remplaza con cero en variable numericas

dataAten$Total<-replace(dataAten$Total,is.null(dataAten$Total),0)

dataAten$Menores\_1<-replace(dataAten$Menores\_1,is.null(dataAten$Menores\_1),0)

dataAten$De\_1\_a\_4-replace(dataAten$De\_1\_a\_4,is.null(dataAten$De\_1\_a\_4),0)

dataAten$De\_5\_a\_14<-replace(dataAten$De\_5\_a\_14,is.null(dataAten$De\_5\_a\_14),0)

dataAten$De\_15\_a\_64<-replace(dataAten$De\_15\_a\_64,is.null(dataAten$De\_15\_a\_64),0)

dataAten$De\_65\_y\_mas<-replace(dataAten$De\_65\_y\_mas,is.null(dataAten$De\_65\_y\_mas),0)

}

#verifica na

dna = verificaNA(dataAten)

dna

if(dna == FALSE)

{

#busca registros na

which(is.na(dataAten))

#busca registros nulos y remplaza con cero en variable numericas

dataAten$Total<-replace(dataAten$Total,is.null(dataAten$Total),0)

dataAten$Menores\_1<-replace(dataAten$Menores\_1,is.null(dataAten$Menores\_1),0)

dataAten$De\_1\_a\_4-replace(dataAten$De\_1\_a\_4,is.null(dataAten$De\_1\_a\_4),0)

dataAten$De\_5\_a\_14<-replace(dataAten$De\_5\_a\_14,is.null(dataAten$De\_5\_a\_14),0)

dataAten$De\_15\_a\_64<-replace(dataAten$De\_15\_a\_64,is.null(dataAten$De\_15\_a\_64),0)

dataAten$De\_65\_y\_mas<-replace(dataAten$De\_65\_y\_mas,is.null(dataAten$De\_65\_y\_mas),0)

}

#verificar caracter 0 en variables numericas

total=dataAten$Total

total

class(total)

vtotal=verificaCaracteres(dataAten$Total)

vtotal

if(vtotal == FALSE)

{

dataAten$Total<-replace(dataAten$Total,is.nul(dataAten$Total),0)

}

vm1=verificaCaracteres(dataAten$Menores\_1)

vm1

if(vm1 == FALSE)

{

dataAten$Menores\_1<-replace(dataAten$Menores\_1,is.null(dataAten$Menores\_1),0)

}

vm1y4=verificaCaracteres(dataAten$De\_1\_a\_4)

vm1y4

if(vm1y4 == FALSE)

{

dataAten$De\_1\_a\_4-replace(dataAten$De\_1\_a\_4,is.null(dataAten$De\_1\_a\_4),0)

}

vm5y14=verificaCaracteres(dataAten$De\_5\_a\_14)

vm5y14

if(vm5y14 == FALSE)

{

dataAten$De\_5\_a\_14<-replace(dataAten$De\_5\_a\_14,is.null(dataAten$De\_5\_a\_14),0)

}

vm15y64=verificaCaracteres(dataAten$De\_15\_a\_64)

vm15y64

if(vm15y64 == FALSE)

{

dataAten$De\_15\_a\_64<-replace(dataAten$De\_15\_a\_64,is.null(dataAten$De\_15\_a\_64),0)

}

vm65ymas=verificaCaracteres(dataAten$De\_65\_y\_mas)

vm65ymas

if(vm65ymas== FALSE)

{

dataAten$De\_65\_y\_mas<-replace(dataAten$De\_65\_y\_mas,is.null(dataAten$De\_65\_y\_mas),0)

}

# Quitar caracteres especiales de la variable nombre de communa

dataAten$NombreComuna<-replace(dataAten$NombreComuna,dataAten$NombreComuna == "Maull\xedn","Maullin")

dataAten$NombreComuna<-replace(dataAten$NombreComuna,dataAten$NombreComuna == "Chait\xe9n","Chaiten")

dataAten$NombreComuna<-replace(dataAten$NombreComuna,dataAten$NombreComuna == "Futaleuf\xfa","Futaleufu")

dataAten$NombreComuna<-replace(dataAten$NombreComuna,dataAten$NombreComuna == "Cocham\xf3","Cochamo")

dataAten$NombreComuna<-replace(dataAten$NombreComuna,dataAten$NombreComuna == "Hualaihu\xe9","Hualaihue")

# Quitar caracteres especiales de la variable nestablecimiento

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "Hospital de Maull\xedn","Hospital de Maullin")

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "Hospital de Futaleuf\xfa","Hospital de Futaleufu")

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "Hospital de Chait\xe9n","Hospital de Chaiten")

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "Centro de Salud Familiar R\xedo Negro Hornopir\xe9n","Centro de Salud Familiar Rio Negro Hornopiren")

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "SUR Cocham\xf3 - R\xedo Puelo","Sur Cochamo - Rio Puelo")

dataAten$NEstablecimiento<-replace(dataAten$NEstablecimiento,dataAten$NEstablecimiento == "SAPU Angelm\xf3","Sapu Angelmo")

#################VARIABLES CATEGORICAS###########################

#Definicion Varibales Categoricas

#obnter el conjunto de observacionde de las comunas con dependencia del servico salud del reloncavi

dataRegion=dataAten%>%filter(CodigoDependencia == 24)

dataRegion

dataRegion$CodigoComuna=factor(dataRegion$CodigoComuna)

tregion<-table(dataRegion$CodigoComuna)

tregion

dfregion=data.frame(tregion)

dfregion

glimpse(dataRegion)

ggplot(dataRegion, aes(x = CodigoComuna)) +

geom\_bar(width=0.5, colour="#eb3c46", fill="#eb3c46")+

ggtitle("Diagrama de Atenciones Urgencias 2023 por Comuna") +

labs(x = NULL, y = "Número de observaciones")

#obtener variable categorica pr fecha

dataRegion$fecha=factor(dataRegion$fecha)

tfecha=table(dataRegion$fecha)

tfecha

dfecha=data.frame(tfecha)

dfecha

ggplot(dfecha, aes(x = Var1, y=Freq)) +

geom\_line()+

geom\_point(colour="#eb3c46")+

labs(x = "Fechas", y = "Frecuencia de Atenciones",

title = "Diagrama de Atenciones Urgencias 2023 por Fecha")

#geom\_line(width=0.5, colour="#e94860", fill="#e94860")+

#ggtitle("Diagrama de Atenciones Urgencias 2023 por Comuna")

#obtener el conjunto de datos de la establecmientos de la dependencia del reloncavi

dataRegion$IdEstablecimiento=factor(dataRegion$IdEstablecimiento)

testablecmiento<-table(dataRegion$IdEstablecimiento)

testablecmiento

ggplot(dataRegion, aes(x = IdEstablecimiento)) +

geom\_bar(width=0.5, colour="#e94860", fill="#e94860")+

ggtitle("Diagrama de barras") +

labs(x = "Establecimiento", y = "Frecuencia de Atenciones",

title = "Diagrama de Atenciones Urgencias 2023 por Establecimiento")

#################VARIABLES CUANTITATIVAS###########################

#Definicion Variables Cuantitativas

#variable de total

ttotal=table(dataRegion$Total)

ttotal

ggplot(dataRegion, aes(x=Total)) +

geom\_histogram(color="#eb3c46", fill="#d3def2")+

labs(x = NULL, y = "Número de observaciones",title = "Diagrama Total de Atenciones")

#variable de menores de 1 año

ggplot(dataRegion, aes(x=Menores\_1)) +

geom\_histogram(color="#eb3c46", fill="#d3def2")+

labs(x = NULL, y = "Número de observaciones",title = "Diagrama Total de Atenciones Menores de 1 año")

#variable de menores de 1 año

ggplot(dataRegion, aes(x=De\_1\_a\_4)) +

geom\_histogram(color="#eb3c46", fill="#d3def2")

#########################MEDIDAS DE TENDENCIA CENTRAL Y POSICION##############

#medidas de tendencia central media, moda, mediana

glimpse(dataRegion$Total)

#variable total

summary(dataRegion$Total)

modatotal<-Moda(dataRegion$Total)

modatotal

desvtotal=sd(dataRegion$Total)

desvtotal

#variable menores de 1 año

summary(dataRegion$Menores\_1)

modam1<-Moda(dataRegion$Menores\_1)

modam1

desvm1=sd(dataRegion$Menores\_1)

desvm1

###############ALGORITMO DE AGRUPACION######################################3

# Algoritmo de cluster de K-means

#crear dataset para algoritmo

class(dataRegion)

dataRegion

gruposcomunast=aggregate(Total ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunast)

gruposcomunasm1=aggregate(Menores\_1 ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunasm1)

gruposcomunasm1a4=aggregate(De\_1\_a\_4 ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunasm1a4)

gruposcomunasm5a14=aggregate(De\_5\_a\_14 ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunasm5a14)

gruposcomunasm15a64=aggregate(De\_15\_a\_64 ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunasm15a64)

gruposcomunasm65ymas=aggregate(De\_65\_y\_mas ~ NombreComuna, data=dataRegion, FUN = sum)

class(gruposcomunasm65ymas)

grupo1=merge(x = gruposcomunas, y = gruposcomunasm1)

grupo1

grupo2=merge(x = grupo1, y = gruposcomunasm1a4)

grupo2

grupo3=merge(x = grupo2, y = gruposcomunasm5a14)

grupo3

grupo4=merge(x = grupo3, y = gruposcomunasm15a64)

grupo4

grupokm=merge(x = grupo4, y = gruposcomunasm65ymas)

grupokm

ciudades=factor(grupokm$NombreComuna)

ciudades

rownames(grupokm)=ciudades

dfkm=grupokm %>% select(Total, Menores\_1, De\_65\_y\_mas)

totalidad = scale(dfkm, center = TRUE, scale = TRUE)

summary(inseguridad)

totalidad= as.data.frame(totalidad)

#creamos 4 cluster en funcion a su grado de inseguridad

kmcluster = kmeans(totalidad,centers=4,nstart = 50)

kmcluster

#graficamos los cluster en funcion del %total y %menores de 1 año

totalidad = totalidad %>% mutate(cluster = kmcluster$cluster)

(g1=ggplot(totalidad, aes(x = Total, y = Menores\_1)) +

geom\_point(aes(color=as.factor(cluster)), size=10)+

geom\_text(aes(label = cluster), size = 5) +

theme\_bw() +

theme(legend.position = "none")+

labs(title = "Kmenas con k=4")

)

(g1=ggplot(totalidad, aes(x = Total, y = De\_65\_y\_mas)) +

geom\_point(aes(color=as.factor(cluster)), size=10)+

geom\_text(aes(label = cluster), size = 5) +

theme\_bw() +

theme(legend.position = "none")+

labs(title = "Kmenas con k=4")

)

######################FUNCIONES################################

#establecerconexion a la base de datos

recolecion\_Datos<-function(host,port,dbname,user, passd)

{

dvr <- PostgreSQL()

db <- dbname ##Nombre de la BBDD

#host\_db <- '10.6.43.108'

host\_db<-host

db\_port <- port

db\_user <- user ##Tu usuario

db\_password <- passd ##Tu contraseña

con <- dbConnect(dvr, dbname = db, host=host\_db, port=db\_port,

user=db\_user, password=db\_password)

return(con)

}

#obtener list de nombres dataframe

#listaNombres<-function(df)

#{

# nombres=names(df)

# return(nombres)

#}

#verificar nulos

verificaNulos<-function(df)

{

#cond <- c(names(df))

#cond

dnulos=any(is.na(df))

return(dnulos)

}

#verificar na

verificaNA<-function(df)

{

dna=anyNA(df)

return(dna)

}

verificaCaracteres<-function(varia)

{

if (is.numeric(varia))

{

respuesta=TRUE

}

else

{

respuesta=FALSE

}

return(respuesta)

}

Moda = function(x) { #G) Otra función para hallar la moda

q <- unique(x)

q[which.max(tabulate(match(x, q)))]

}