### Autores:

Yulian Esteban Vargas Pardo, David Alejandro Mosquera Moreno, Kevin Stiven Moreno Martinez, Lina Dayana Rojas Angarita, Camila Alejandra Velasco Ruiz, yeferson David Aparicio Gómez

### Año:

2022

### Curso:

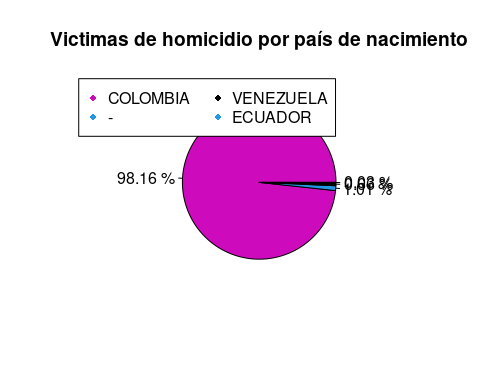
FUNDAMENTOS DE ESTADÍSTICA PARA ANALÍTICA DE DATOS

# Homicidios en la ciudad de Cali

## 1. Introducción

En la base de datos de “Homicidios” se observa que durante el año 2017 hubo 11965 homicidios, al realizar un análisis detallado el 98.16% personas afectadas tenían nacionalidad colombiana, el 1.01% no fue reportada esta información y el 0.66% era de nacionalidad venezolana. Por lo anterior observado, se pretende analizar en la base de datos el país Colombia, donde se encuentra que, de los 32 departamentos observados, el departamento del Valle es el que tiene más homicidios registrados con el 20% de los casos seguido por los departamentos de Antioquia 16% y Cundinamarca el 13%, por tal razón para analizar los diferentes factores del homicidio en los ciudadanos colombianos nos enfocaremos en el municipio de Cali.

library(readxl)  
Paisdenacimiento <- read\_excel("Homicidios (1).xlsx",   
 sheet = "Pais de nacimiento", col\_types = c("text"))  
##View(Paisdenacimiento)  
  
#nrow(Paisdenacimiento) # numero de filas de Refresco  
#ncol(Paisdenacimiento)  
# tabla de frecuencias. El resto se hace de la misma forma que  
# en el caso 1  
frec.abs1 <- table(Paisdenacimiento);#frec.abs  
frec.abs1 <- sort(frec.abs1,decreasing = T)[1:4]  
n <- dim(Paisdenacimiento)[1];#n  
frec.relat1 <- frec.abs1/n;#frec.relat1  
prctg <- frec.relat1\*100  
##windows(50,30) # abriendo ventana grafica externa  
  
  
#?barplot  
#barplot(frec.abs1,xlab="Pais de nacimiento",  
 # ylab="Frecuencia absoluta",  
 # main="Homicidios",col=c(8,2,3,4),ylim=c(0,12000))  
  
n1 <- rownames(frec.abs1)  
#n1  
  
pie(frec.abs1,labels=paste(round((frec.abs1/n)\*100,2),"%"), main="Victimas de homicidio por país de nacimiento",  
 col=c(6,4,9,12))  
legend("topleft",legend =n1,ncol=2,col=c(6,4,9,12),pch=18)



Por ello, el siguiente informe tiene como objetivo presentar el índice de homicidios de la ciudad de Cali en el año 2017, para ello se utilizó una base de datos de Homicidios ocurridos en Colombia en dicho año.

Esta problemática se ha presentado desde años atrás, según el informe de UNDOC (Oficina de Naciones Unidas contra la Droga y el Delito), la ciudad de Cali registra entre las ciudades con mayor tasa de hasta 80 homicidios. Comparado con otras ciudades de Colombia en el año 2017, Medellín registró 573 homicidios, Bogotá registró 1130, Barranquilla 346, Bucaramanga 91 y finalmente Cali registró 1234 homicidios siendo así la ciudad con la mayor cantidad de víctimas en el país.

## 2. Identificación de los tipos de variables para la base de datos Homicidios

| Variable cualitativa | Cualitativa | Cuantitativa |
| --- | --- | --- |
| Fecha |  | Continua |
| Departamento | Nominal |  |
| Municipio | Nominal |  |
| Día |  | Continua |
| Hora | Discreta |  |
| Barrio | Nominal |  |
| Zona | Nominal |  |
| Clase de sitio | Nominal |  |
| Arma empleada | Nominal |  |
| Móvil Agresor | Nominal |  |
| Móvil Victima | Nominal |  |
| Edad |  | Discreta |
| Sexo | Nominal |  |
| Estado civil | Nominal |  |
| País de nacimiento | Nominal |  |
| Clase de empleado | Nominal |  |
| Profesión | Nominal |  |
| Escolaridad | Ordinal |  |
| Código DANE | Ordinal |  |
| Cantidad Departamento |  | Continua |

## 3. Análisis Gráfico

### 3.1 Comparativa de homicidios de las principales ciudades del país

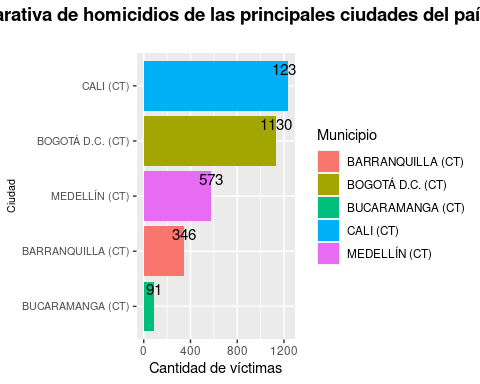
library(readr)  
Homicidios <- read\_csv("Homicidios.csv")

## Rows: 11965 Columns: 20  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): Fecha, Departamento, Municipio, Día, Barrio, Zona, Clase de sitio...  
## dbl (2): Edad, Cantidad  
## time (1): Hora  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#View(Homicidios)  
library(sqldf)

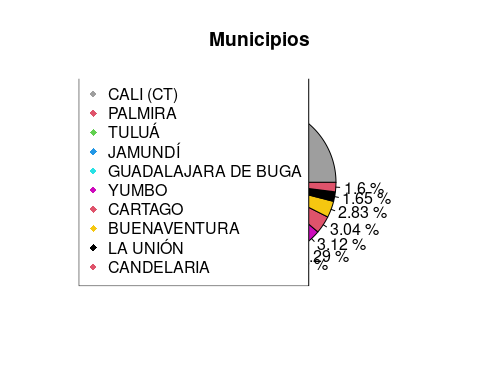
## Loading required package: gsubfn  
## Loading required package: proto  
## Loading required package: RSQLite

library(ggplot2)  
  
cities <- sqldf("WITH A AS(SELECT COUNT(\*) as counter, Municipio FROM Homicidios WHERE Departamento = 'VALLE' AND Municipio = 'CALI (CT)'  
 UNION ALL  
 SELECT COUNT(\*) as counter, Municipio FROM Homicidios WHERE Departamento = 'ANTIOQUIA' AND Municipio = 'MEDELLÍN (CT)'  
 UNION ALL  
 SELECT COUNT(\*) as counter, Municipio FROM Homicidios WHERE Departamento = 'CUNDINAMARCA' AND Municipio = 'BOGOTÁ D.C. (CT)'  
 UNION ALL  
 SELECT COUNT(\*) as counter, Municipio FROM Homicidios WHERE Departamento = 'ATLÁNTICO' AND Municipio = 'BARRANQUILLA (CT)'  
 UNION ALL  
 SELECT COUNT(\*) as counter, Municipio FROM Homicidios WHERE Departamento = 'SANTANDER' AND Municipio = 'BUCARAMANGA (CT)')  
 SELECT \* FROM A ORDER BY 1 DESC   
 ")  
  
ggplot(cities, aes(x= reorder(Municipio, counter), y=counter, fill=Municipio))+geom\_col()+ coord\_flip() +  
   
 labs( title = "Comparativa de homicidios de las principales ciudades del país\n", x= "Ciudad" , y ="Cantidad de víctimas" )+  
 geom\_text(aes(y = counter, label = counter),   
 position = position\_dodge(width = 0.9), size=4, vjust=-1, hjust=0.5 ,col="black")+  
 theme(  
 axis.text.y = element\_text(size = 8), axis.title.y = element\_text(size = 8),  
 plot.title = element\_text(size = 14, face = "bold", color = "Black", hjust= .5)  
 )



### 3.2 Comparativa de municipios del departamento del Valle

library(readxl)  
Municipios <- read\_excel("Homicidios (1).xlsx",   
 sheet = "municipio", col\_types = c("text"))  
##View(Municipios)  
  
n <- dim(Municipios)[1];#n  
frec.abs <- sort(table(Municipios),decreasing = T)[1:10];#frec.abs  
frec.relat = frec.abs/n  
prctg <- frec.relat\*100;#prctg  
n1 <- rownames(frec.abs);#n1  
##windows(50,30) # abriendo ventana grafica externa  
  
  
  
  
#?barplot  
#barplot(frec.abs,xlab="Municipios",  
 # ylab="Frecuencia absoluta",  
 # main="Homicidios",col=c(8,2,3,4,5,6,7,15,9,10,11,12,13,14),ylim=c(0,1300))  
  
  
pie(frec.abs,labels=paste(round((frec.abs/n\*100),2),"%"), main="Municipios",  
 col=c(8,2,3,4,5,6,18,15,9,10 ))  
  
legend("bottomleft",legend =n1,ncol=1,col=c(8,2,3,4,5,6,18,15,9,10 ),pch=18)

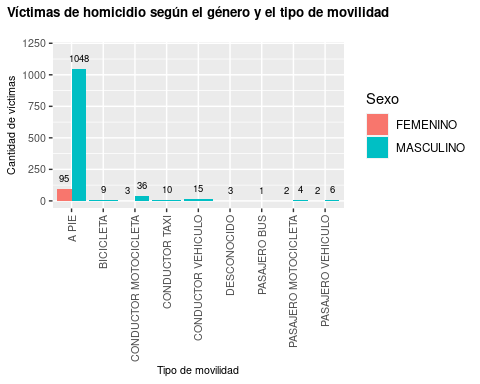


### 3.3 Móvil víctimas de homicidios por sexo

library(readr)  
Homicidios <- read\_csv("Homicidios.csv")

## Rows: 11965 Columns: 20  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): Fecha, Departamento, Municipio, Día, Barrio, Zona, Clase de sitio...  
## dbl (2): Edad, Cantidad  
## time (1): Hora  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#View(Homicidios)  
library(sqldf)  
library(ggplot2)  
  
  
# Movil Victima con mayores homicidios haciendo distincion de genero  
Datos <- sqldf("SELECT CASE WHEN [Móvil Victima] !='-' THEN [Móvil Victima] ELSE 'DESCONOCIDO' END AS [Móvil Victima] , sexo, count(sexo) AS Cantidad FROM Homicidios WHERE Departamento = 'VALLE' AND Municipio = 'CALI (CT)' group by sexo, [Móvil Victima]")  
#Visualizar Consulta#  
#View(Datos)  
  
#Graficar  
ggplot(Datos, aes(x = `Móvil Victima`, y=Cantidad, fill =Sexo))+  
 #Rango de datos  
 scale\_y\_continuous(limit = c(0,1200))+  
 #Separa la columna de acuerdo al género  
 geom\_col(position='dodge' ) +  
 labs( title = "Víctimas de homicidio según el género y el tipo de movilidad\n", x= "Tipo de movilidad" , y ="Cantidad de víctimas" )+  
 #Valores encima de las columnas  
 geom\_text(aes(y = Cantidad, label = Cantidad),   
 position = position\_dodge(width = 0.9), size=2.5, vjust=-1, hjust=0.5 ,col="black")+  
 #Eje X vertical, y personaliza el título  
 theme(  
 axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1, size = 8), axis.title.x = element\_text(size = 8),  
 axis.text.y = element\_text(size = 8), axis.title.y = element\_text(size = 8),  
 plot.title = element\_text(size = 10, face = "bold", color = "black",hjust = 0.5)  
 )



### 3.4 Los cinco Barrios con mayor cantidad de homicidios en la ciudad de Cali

library(readr)  
Homicidios <- read\_csv("Homicidios.csv")

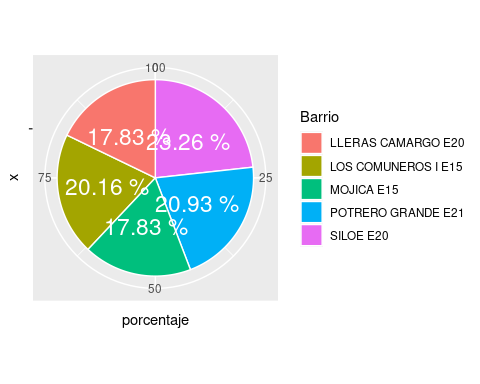
## Rows: 11965 Columns: 20  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): Fecha, Departamento, Municipio, Día, Barrio, Zona, Clase de sitio...  
## dbl (2): Edad, Cantidad  
## time (1): Hora  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#View(Homicidios)  
library(sqldf)  
library(ggplot2)  
library(knitr)  
#Top 5 Barrios con mayor cantidad homicidios en la ciudad de Cali  
  
test3 <- sqldf("SELECT Barrio, Count(Barrio) AS victimas,  
 ROUND(Count(Barrio) / ( SELECT (SUM(victimas) \*1.0)  
 FROM  
 (SELECT Barrio, Count(Barrio) AS victimas  
 FROM Homicidios  
 WHERE Departamento = 'VALLE' AND Municipio = 'CALI (CT)'  
 Group By Barrio  
 Order By 2 DESC limit 5)A)\*100 ,2) AS porcentaje  
FROM Homicidios  
WHERE Departamento = 'VALLE' AND Municipio = 'CALI (CT)'  
Group By Barrio  
Order By 2 DESC  
limit 5")  
  
  
kable(test3, caption = "Tabla de datos")

Tabla de datos

| Barrio | victimas | porcentaje |
| --- | --- | --- |
| SILOE E20 | 30 | 23.26 |
| POTRERO GRANDE E21 | 27 | 20.93 |
| LOS COMUNEROS I E15 | 26 | 20.16 |
| MOJICA E15 | 23 | 17.83 |
| LLERAS CAMARGO E20 | 23 | 17.83 |

ggplot(test3,aes(x="",y=porcentaje, fill=Barrio))+  
 geom\_bar(stat = "identity",  
 color="white")+  
 geom\_text(aes(label=paste(porcentaje, "%")),  
 position=position\_stack(vjust=0.5),color="white",size=6)+  
 coord\_polar(theta = "y")



#--------------------------------------

## 4. Referencias

- Tejeda Puentes, D. S. (2017). Producción del poder mafioso en Santiago de Cali y su relación con la salud urbana: territorios y redes. repositorio.unal.edu.co. <https://repositorio.unal.edu.co/bitstream/handle/unal/59802/31841192.2017.pdf?sequence=1&isAllowed=y>