

Análisis Descriptivo

```
df <- read.csv("./BDD_DICIEMBRE_2022.csv")
head(df)
```

```
##      X ID ANIO      SINIESTROS LESIONADOS FALLECIDOS ENTE_DE_CONTROL LATITUD_Y
## 1 1 1 2017 DMQ00001012017      1      0      DMQ -0.083501
## 2 2 2 2017 ATM00002012017      1      0      ATM -2.246682
## 3 3 3 2017 PNE00003012017      1      0      PNE -0.253881
## 4 4 4 2017 DMQ00004012017      0      0      DMQ -0.116059
## 5 5 5 2017 DMQ00005012017      0      0      DMQ -0.239721
## 6 6 6 2017 DMQ00006012017      0      0      DMQ -0.116354
##      LONGITUD_X DPA_1      PROVINCIA DPA_2      CANTON DPA_3
## 1 -78.41774      17      PICHINCHA 1701      QUITO 170155
## 2 -79.89775      9      GUAYAS 901      GUAYAQUIL 90150
## 3 -79.21740      23 SANTO DOMINGO DE LOS TSACHILAS 2301 SANTO DOMINGO 230150
## 4 -78.46419      17      PICHINCHA 1701      QUITO 170150
## 5 -78.51206      17      PICHINCHA 1701      QUITO 170150
## 6 -78.46504      17      PICHINCHA 1701      QUITO 170150
##      PARROQUIA      DIRECCION
## 1      CALDERON (CARAPUNGO)      GIOVANNI CALLES & RAFAEL CALVACHE
## 2      GUAYAQUIL CALLE PUYO Y CALLE BAEZA, CDLA. LOS ESTEROS
## 3 SANTO DOMINGO DE LOS COLORADOS      COOP. LUZ DEL DIA
## 4      QUITO      EUGENIO DELGADO Y CARLOS FORTINEZ
## 5      QUITO      GUAYABAMBA Y CHAMBO
## 6      QUITO      CALDAS Y RIOS
##      ZONA_PLANIFICACION      ZONA ID_DE_LA_VIA NOMBRE_DE_LA_VIA UBICACION_DE_LA_VIA
## 1      ZONA 9 RURAL      ND      ND      ND
## 2      ZONA 8 URBANA      ND      ND      ND
## 3      ZONA 4 URBANA      ND      ND      ND
## 4      ZONA 9 RURAL      ND      ND      ND
## 5      ZONA 9 URBANA      ND      ND      ND
## 6      ZONA 9 URBANA      ND      ND      ND
##      JERARQUIA_DE_LA_VIA      FECHA      HORA      PERIODO_1 PERIODO_2      DIA_1
## 1      ND 01/01/2017 00:15:00 DE 00H00 A 00H59      0 DOMINGO
## 2      ND 01/01/2017 00:25:00 DE 00H00 A 00H59      0 DOMINGO
## 3      ND 01/01/2017 00:25:00 DE 00H00 A 00H59      0 DOMINGO
## 4      ND 01/01/2017 00:25:00 DE 00H00 A 00H59      0 DOMINGO
## 5      ND 01/01/2017 00:30:00 DE 00H00 A 00H59      0 DOMINGO
## 6      ND 01/01/2017 00:30:00 DE 00H00 A 00H59      0 DOMINGO
##      DIA_2 MES_1 MES_2 FERIADO CODIGO_CAUSA
## 1      7 ENERO      1      SI      C16
## 2      7 ENERO      1      SI      C16
## 3      7 ENERO      1      SI      C23
## 4      7 ENERO      1      SI      C03
## 5      7 ENERO      1      SI      C14
## 6      7 ENERO      1      SI      C06
##
```

```

## 1 NO TRANSITAR POR LAS ACERAS O ZONAS D
## 2 NO TRANSITAR POR LAS ACERAS O ZONAS D
## 3 NO RESPETAR LAS SENIALES REGLAMENTARIAS DE TRANSITO. (PARE, CEDA
## 4 CONDUCIR EN ESTADO DE SOMNOLENCIA O MALAS CONDICIONES
## 5 CONDUCIR DESATENTO A LAS CONDICIONES DE TRANSITO (CELULAR, PANTALLAS DE VIDEO, COMIDA, MAQUILLAJE (
## 6 CONDUCE BAJO LA INFLUENCIA DE ALCOHOL, SUSTANCIAS ESTUPEFACIENTES
## TIPO_DE_SINIESTRO TIPO_DE_VEHICULO_1 SERVICIO_1 AUTOMOVIL BICICLETA BUS
## 1 ATROPELLOS NO IDENTIFICADO PARTICULAR 0 0 0
## 2 ATROPELLOS AUTOMOVIL PARTICULAR 1 0 0
## 3 ATROPELLOS MOTOCICLETA PARTICULAR 0 0 0
## 4 ESTRELLAMIENTOS AUTOMOVIL PARTICULAR 1 0 0
## 5 CAIDA DE PASAJERO NO IDENTIFICADO PARTICULAR 0 0 0
## 6 ESTRELLAMIENTOS BUS PUBLICO 1 0 1
## CAMION CAMIONETA EMERGENCIAS ESPECIAL FURGONETA MOTOCICLETA NO_IDENTIFICADO
## 1 0 0 0 0 0 0 1
## 2 0 0 0 0 0 0 0
## 3 0 0 0 0 0 1 0
## 4 0 0 0 0 0 0 0
## 5 0 0 0 0 0 0 1
## 6 0 0 0 0 0 0 0
## SCOOTER_ELECTRICO TRICIMOTO VEHICULO_DEPORTIVO_UTILITARIO SUMA_DE_VEHICULOS
## 1 0 0 0 1
## 2 0 0 0 1
## 3 0 0 0 1
## 4 0 0 0 1
## 5 0 0 0 1
## 6 0 0 1 3
## TIPO_ID_1 EDAD_1 SEXO_1 CONDICION_1 PARTICIPANTE_1 CASCO_1
## 1 CEDULA 36 MUJER LESIONADO PEATON NO
## 2 CEDULA 12 HOMBRE LESIONADO PEATON NO
## 3 CEDULA 26 HOMBRE LESIONADO PEATON NO
## 4 NO IDENTIFICADO -1 NO IDENTIFICADO NO IDENTIFICADO CONDUCTOR NO
## 5 CEDULA 11 HOMBRE ILESO PASAJERO NO
## 6 CEDULA 36 HOMBRE ILESO CONDUCTOR NO
## CINTURON_1
## 1 NO
## 2 NO
## 3 NO
## 4 NO
## 5 NO
## 6 NO

```

```
table(sapply(df, class))
```

```

##
## character integer numeric
##      29      26      2

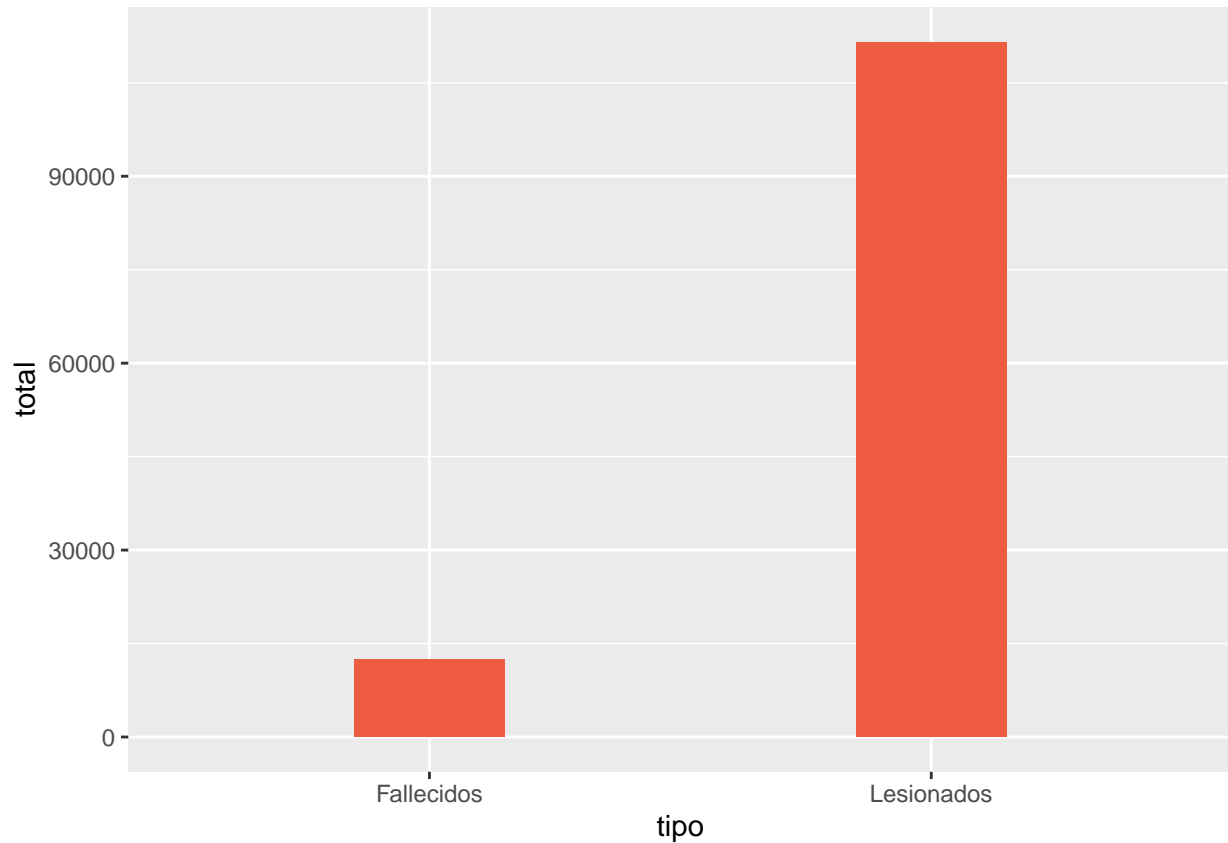
```

La base de datos cuenta con un total de 56 columnas, de las cuales 29 son identificadas del tipo “character”, 25 “integer” y 2 “numeric”.

```
library(ggplot2)
```

```
datos_les_fal <- data.frame(tipo=c("Lesionados", "Fallecidos"),total=c(sum(df$LESIONADOS),sum(df$FALLECIDOS)))

ggplot(data = datos_les_fal)+
  geom_bar(aes(x=tipo, y=total),
           stat = "identity",
           width = 0.3,
           fill = "tomato2")
```

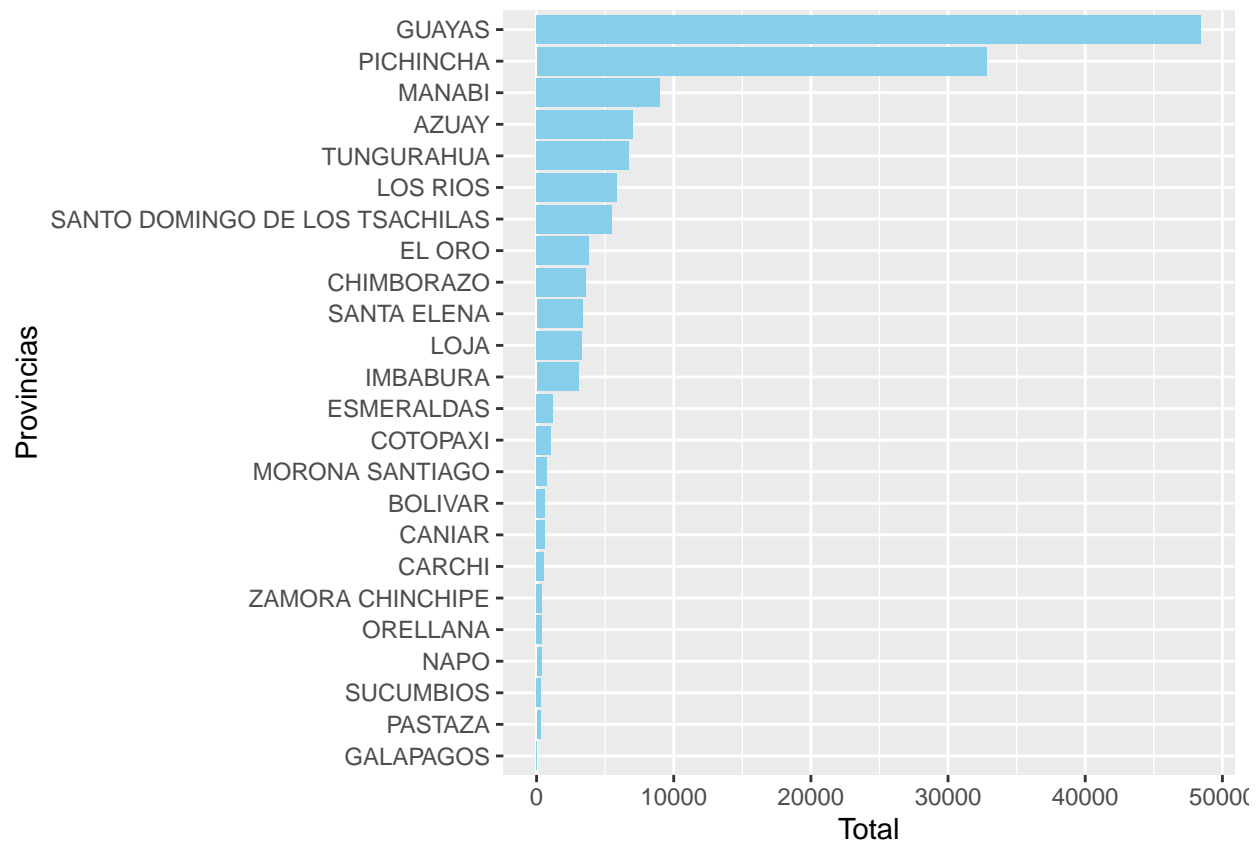


Hay un total de 111512 (104918) lesionados y 12408(11676) fallecidos

```
freq_prov <- data.frame(table(df$PROVINCIA))
colnames(freq_prov) <- c("Provincias", "Total")

freq_prov$Provincias <- factor(freq_prov$Provincias, levels=freq_prov$Provincias[order(freq_prov$Total)])

ggplot(data = freq_prov, aes(x=Total,y=Provincias))+
  geom_bar(
    stat = "identity",
    fill = "skyblue")
```

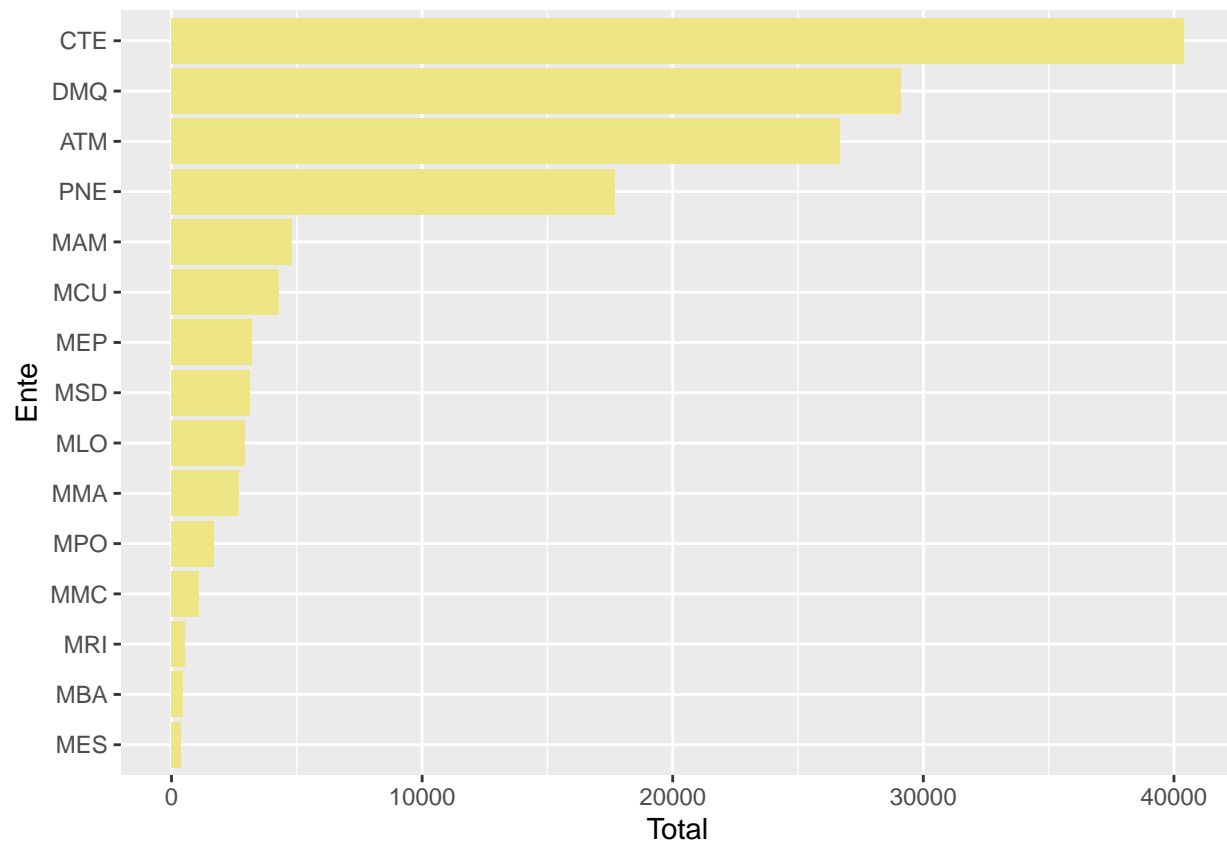


Se puede observar que las provincias que más registran siniestros son Guayas y Pichincha. Por el otro lado; Napo, Sucumbios, Pastaza y Galápagos entre otras tienen pocos accidentes.

```
freq_ente_cont <- data.frame(table(df$ENTE_DE_CONTROL))
colnames(freq_ente_cont) <- c("Ente", "Total")

freq_ente_cont$Ente <- factor(freq_ente_cont$Ente, levels = freq_ente_cont$Ente[order(freq_ente_cont$Total)])

ggplot(data = freq_ente_cont, aes(y=Ente, x=Total))+
  geom_bar(stat = "identity",
    fill = "khaki2")
```

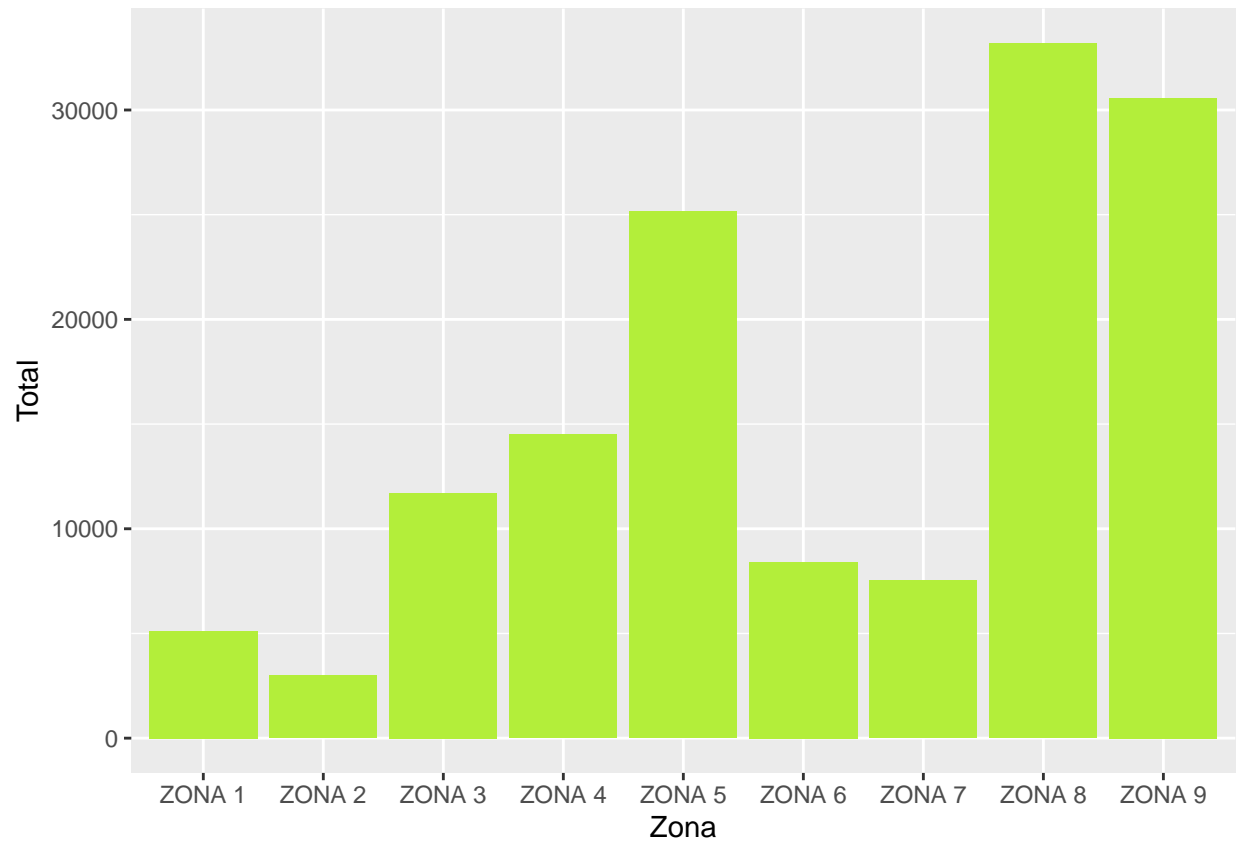


Del gráfico se observa que los entes que registran un mayor número de siniestro son CTE, DMQ, ATM y PNE

```
freq_zona <- data.frame(table(df$ZONA_PLANIFICACION))
colnames(freq_zona) <- c("Zona", "Total")

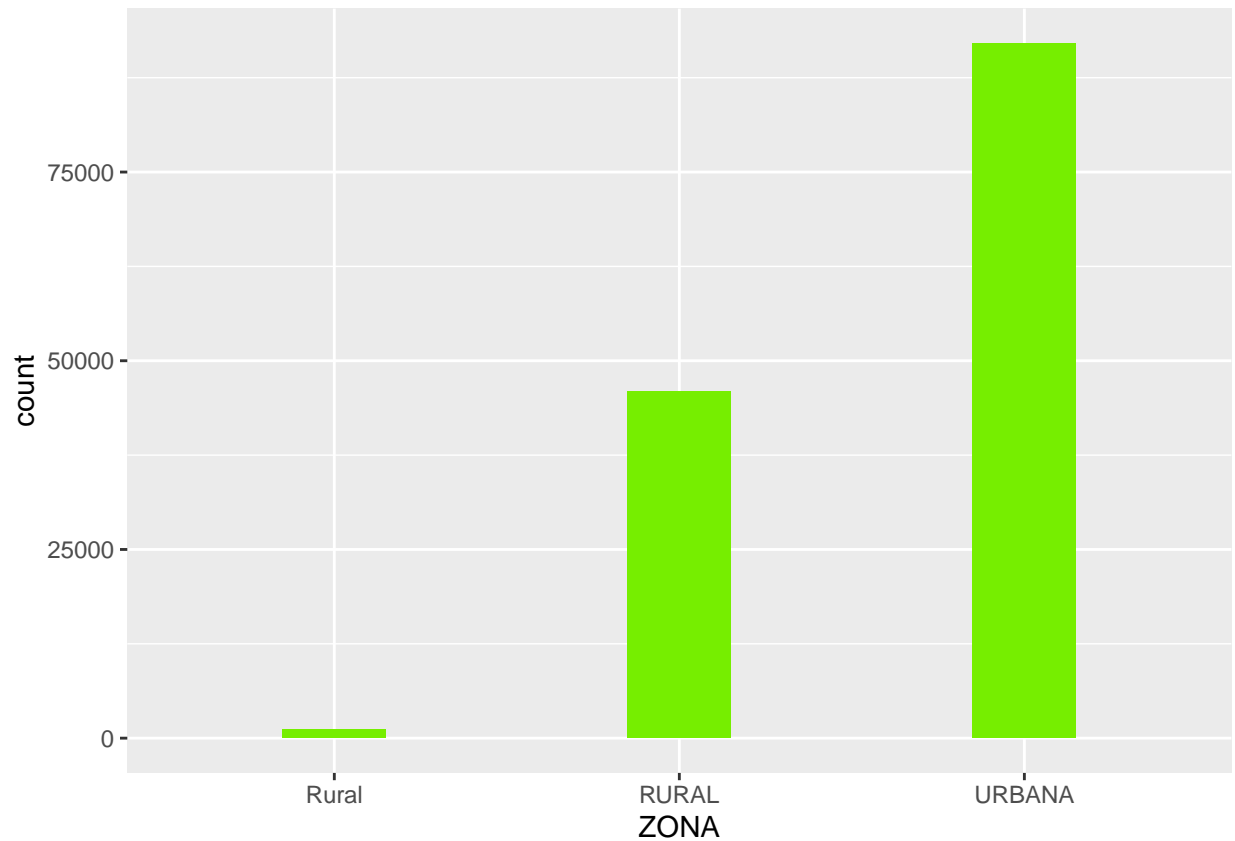
freq_zona$Zona <- factor(freq_zona$Zona, levels = freq_zona$Zona[order(freq_zona$Zona)])

ggplot(data = freq_zona, aes(y=Total, x=Zona))+
  geom_bar(stat = "identity",
           fill = "olivedrab2")
```



Zona 8, Zona 9 y Zona 5 con mayor número de siniestros

```
ggplot(data = df, aes(x=ZONA))+  
  geom_bar(  
    fill = "chartreuse2",  
    width = 0.3)
```

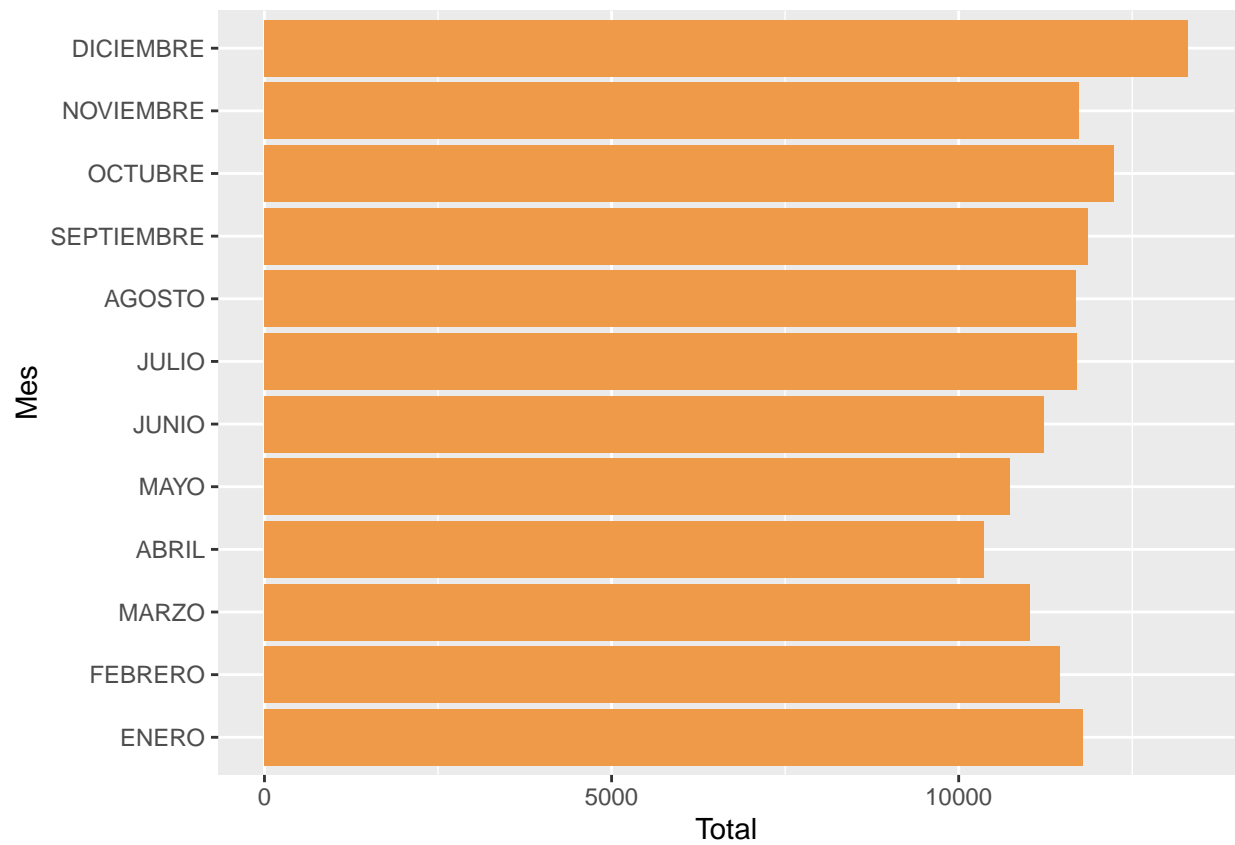


Predominan los siniestros en las zonas urbanas del país

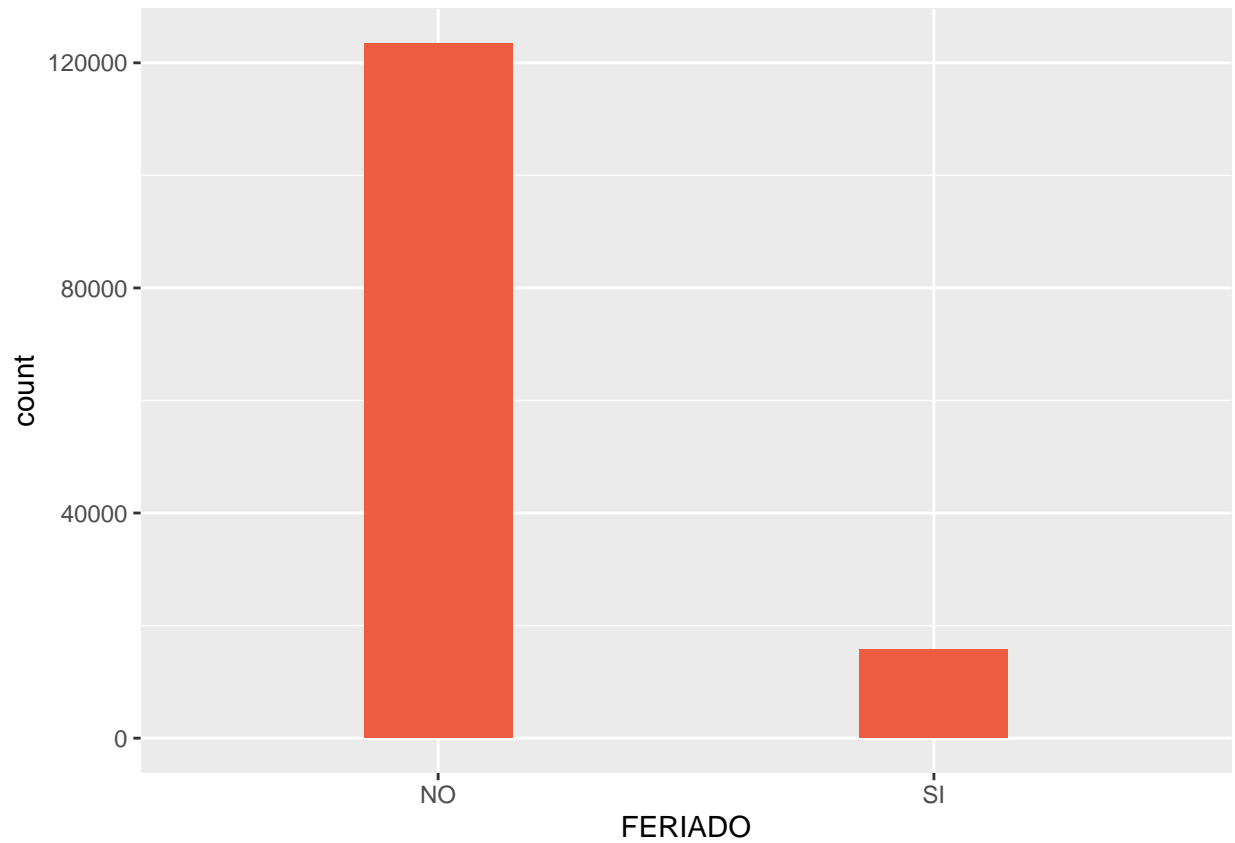
```
freq_mes <- data.frame(table(df$MES_1))
colnames(freq_mes) <- c("Mes", "Total")

freq_mes$Mes <- factor(freq_mes$Mes, levels = c("ENERO", "FEBRERO", "MARZO",
"ABRIL", "MAYO", "JUNIO", "JULIO",
"AGOSTO", "SEPTIEMBRE", "OCTUBRE",
"NOVIEMBRE", "DICIEMBRE"))

ggplot(data = freq_mes, aes(y=Mes, x=Total))+
  geom_bar(
    stat = "identity",
    fill = "tan2"
  )
```



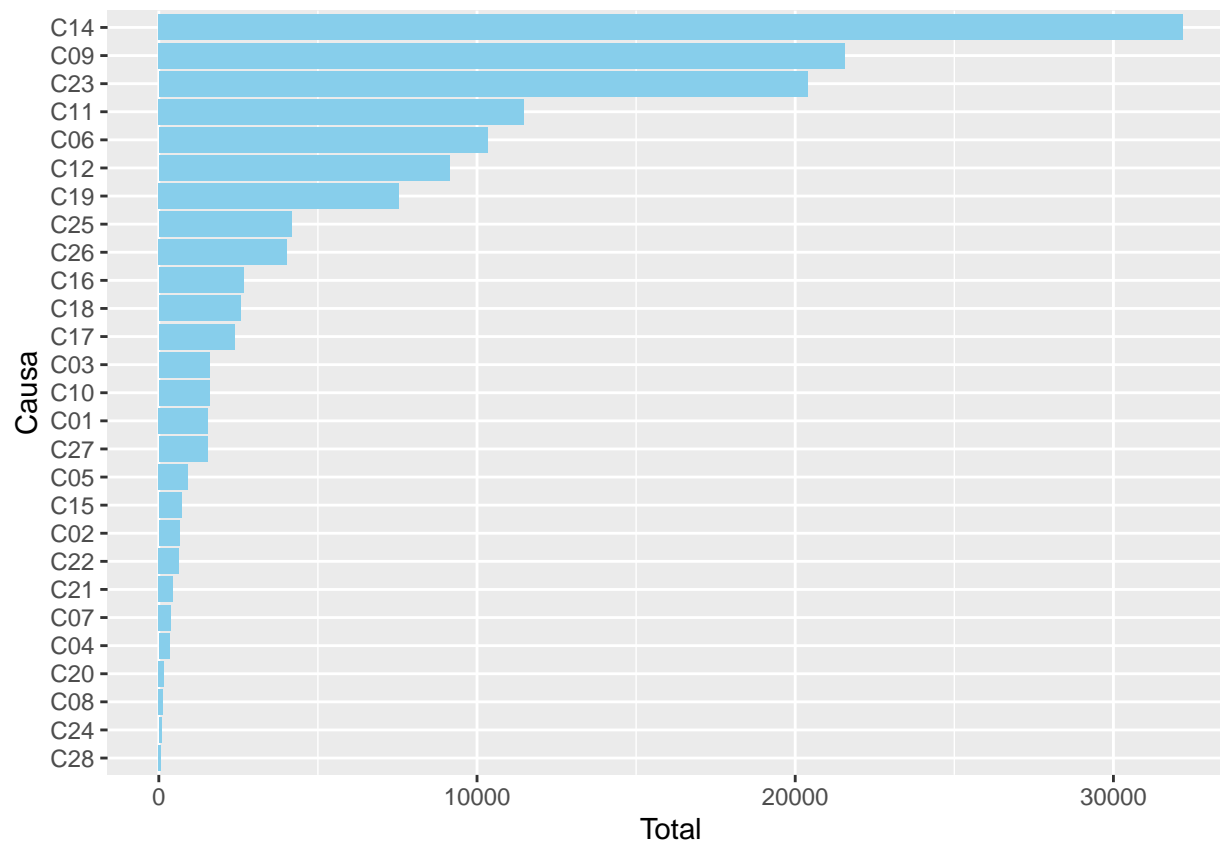
```
ggplot(data = df, aes(x=FERIADO))+  
  geom_bar(  
    width = 0.3,  
    fill = "tomato2"  
  )
```

```
freq_causa <- data.frame(table(df$CODIGO_CAUSA))
colnames(freq_causa) <- c("Causa", "Total")

freq_causa$Causa <- factor(freq_causa$Causa, levels = freq_causa$Causa[order(freq_causa$Total)])

ggplot(data = freq_causa, aes(y=Causa,x=Total))+
  geom_bar(
    stat = "identity",
    fill = "skyblue")
```

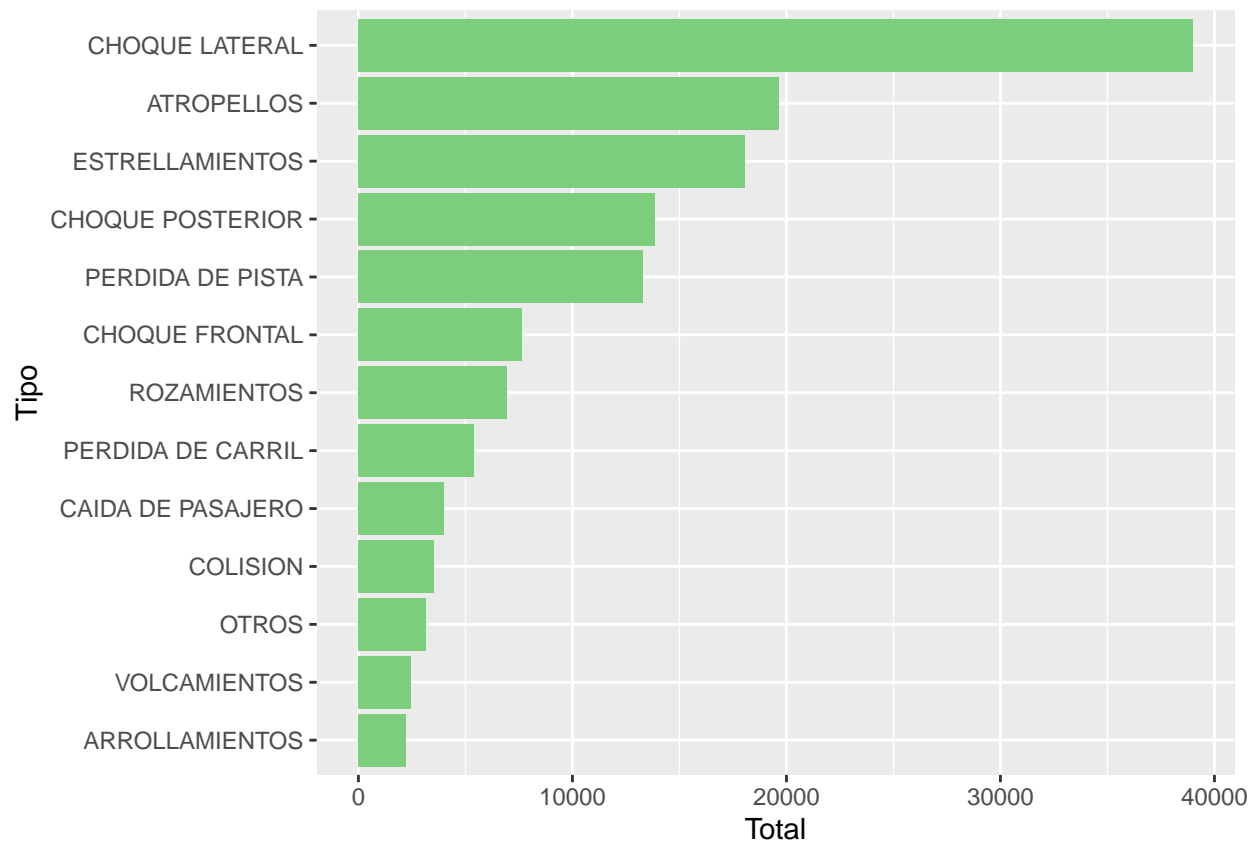


Los códigos de causa

```
freq_tipo <- data.frame(table(df$TIPO_DE_SINIESTRO))
colnames(freq_tipo) <- c("Tipo", "Total")

freq_tipo$Tipo <- factor(freq_tipo$Tipo, levels = freq_tipo$Tipo[order(freq_tipo$Total)])

ggplot(data = freq_tipo, aes(y=Tipo,x=Total))+
  geom_bar(
    stat = "identity",
    fill = "palegreen3")
```

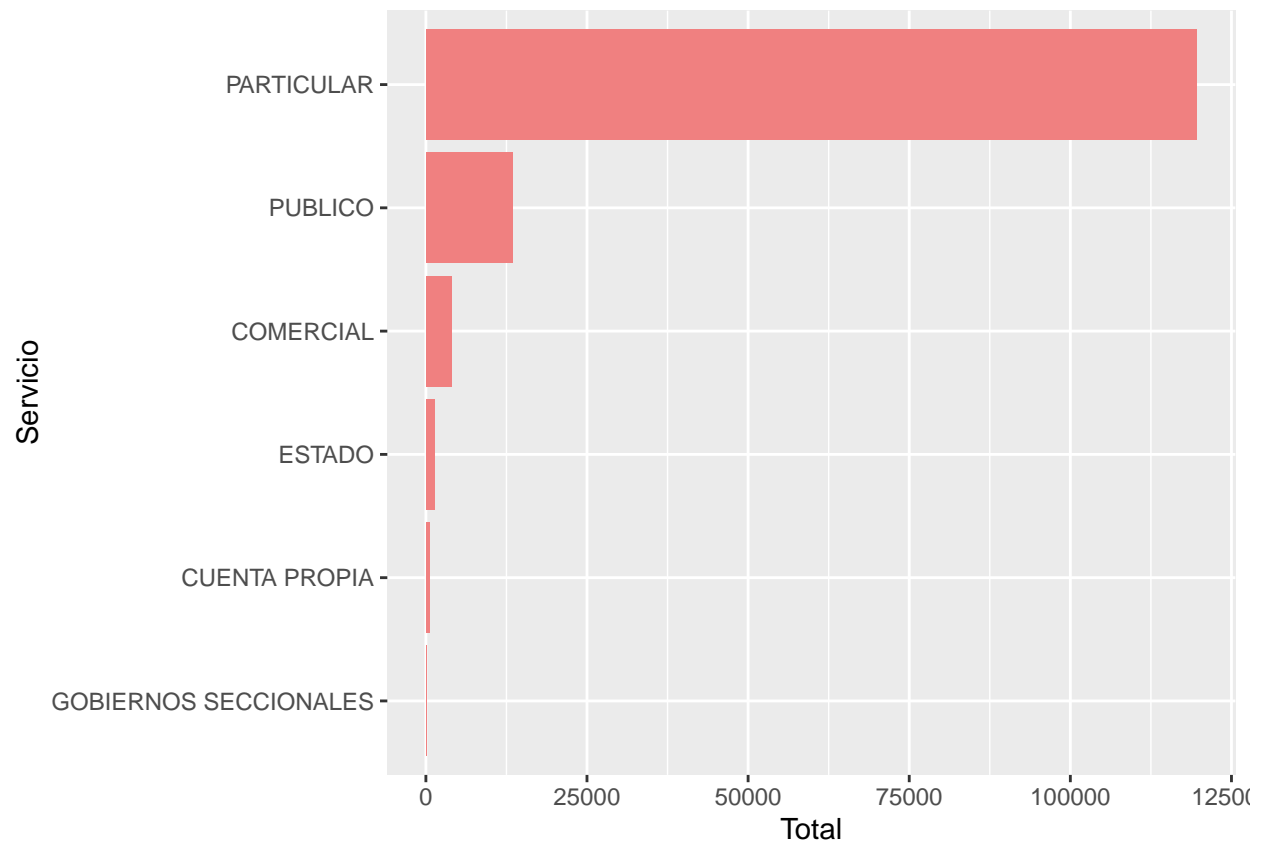


El choque lateral es el tipo de siniestro más común

```
freq_servicio <- data.frame(table(df$SERVICIO_1))
colnames(freq_servicio) <- c("Servicio", "Total")

freq_servicio$Servicio <- factor(freq_servicio$Servicio, levels = freq_servicio$Servicio[order(freq_ser

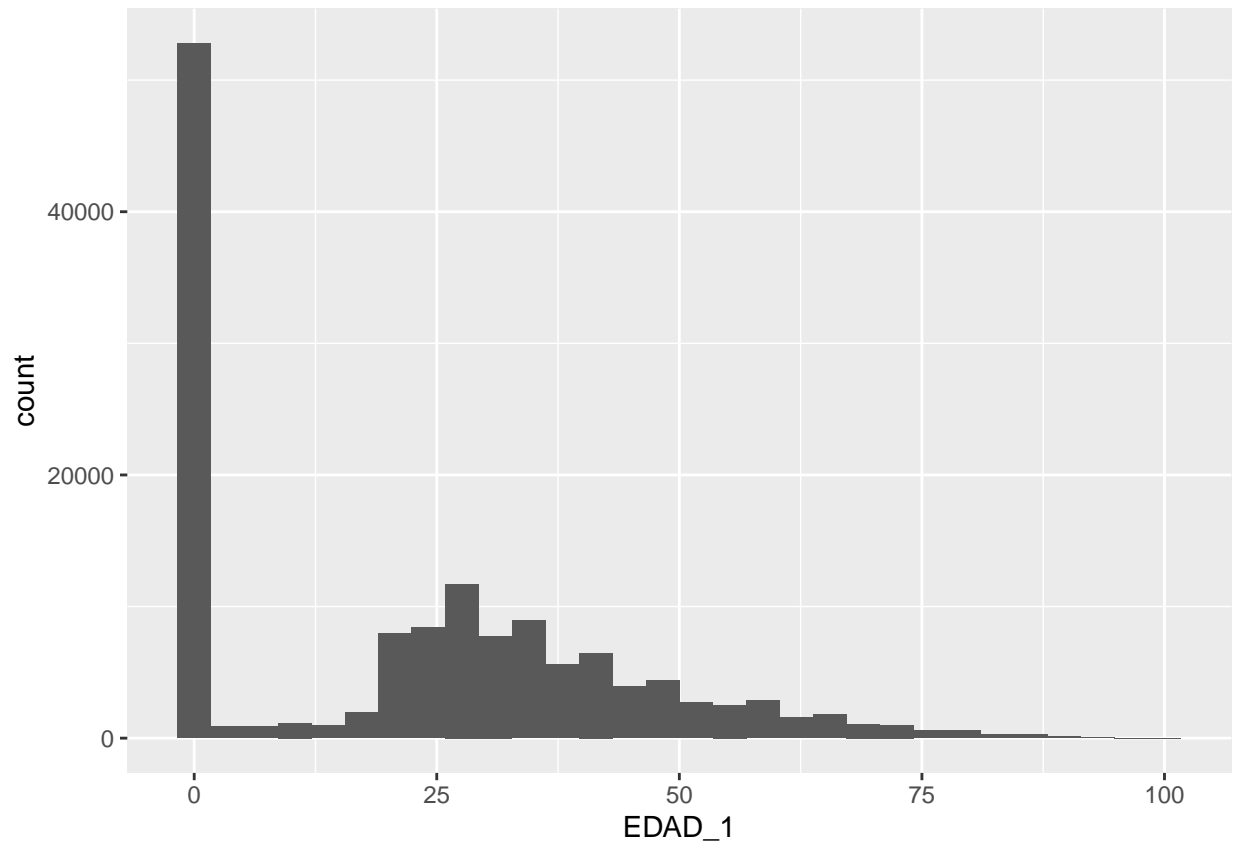
ggplot(data = freq_servicio, aes(y=Servicio,x=Total)))+
  geom_bar(
    stat = "identity",
    fill = "lightcoral")
```



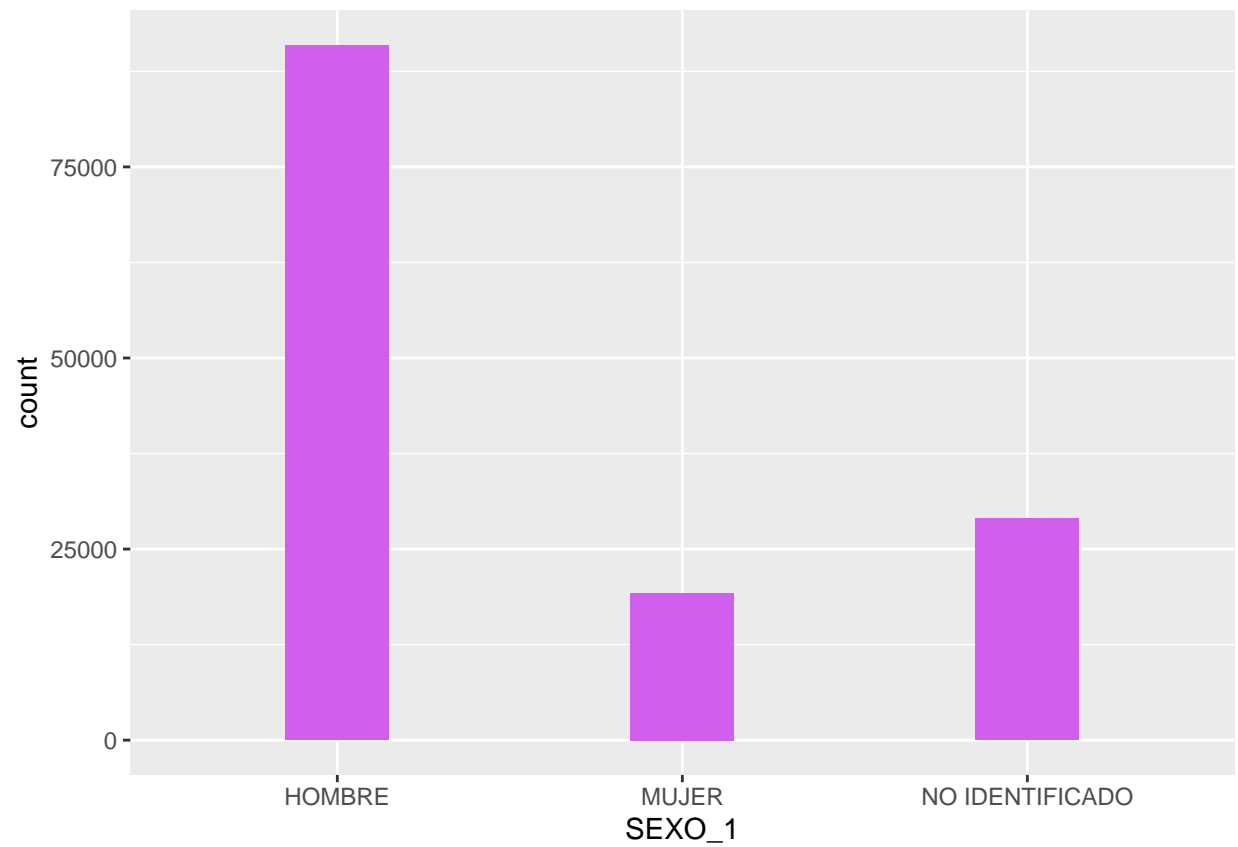
La gran mayoría ocurre en servicios particulares

```
ggplot(data = df, aes(x=EDAD_1))+  
  geom_histogram()
```

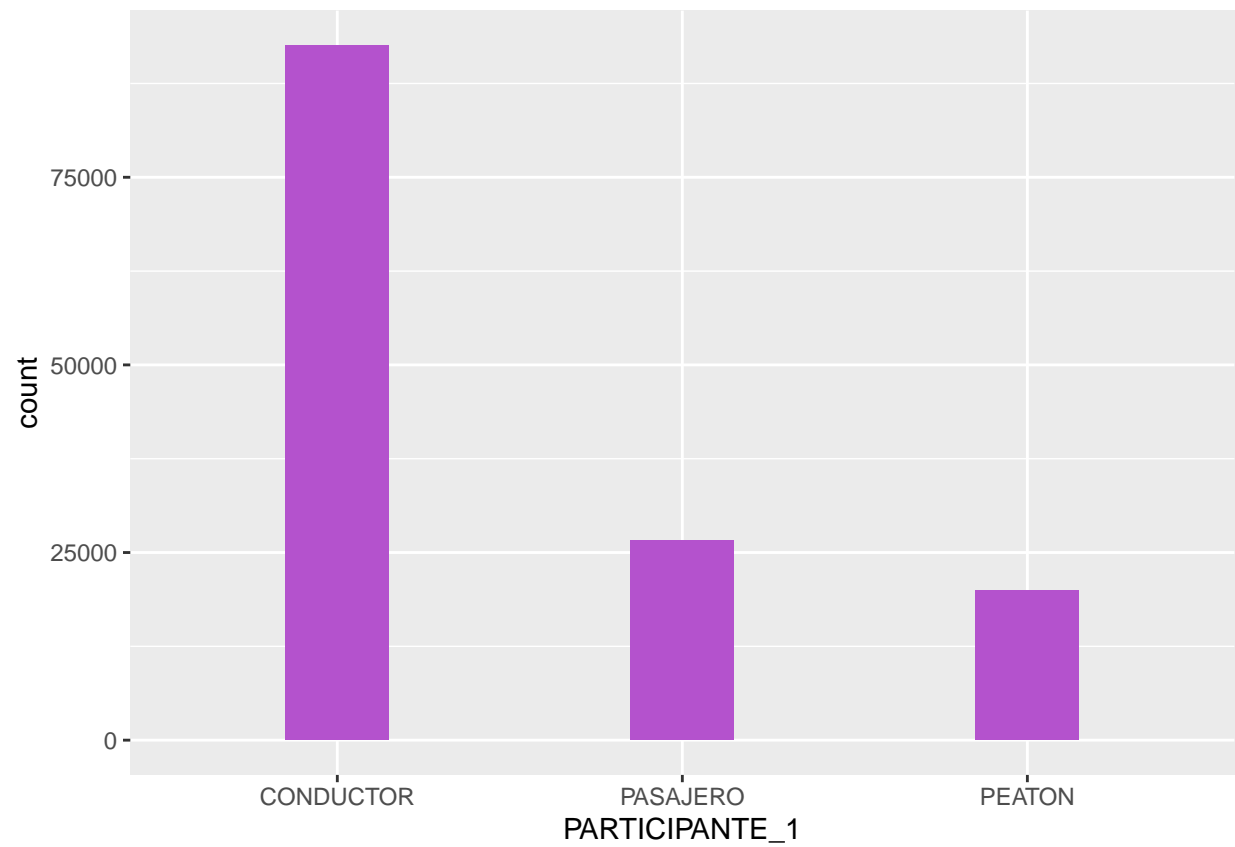
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



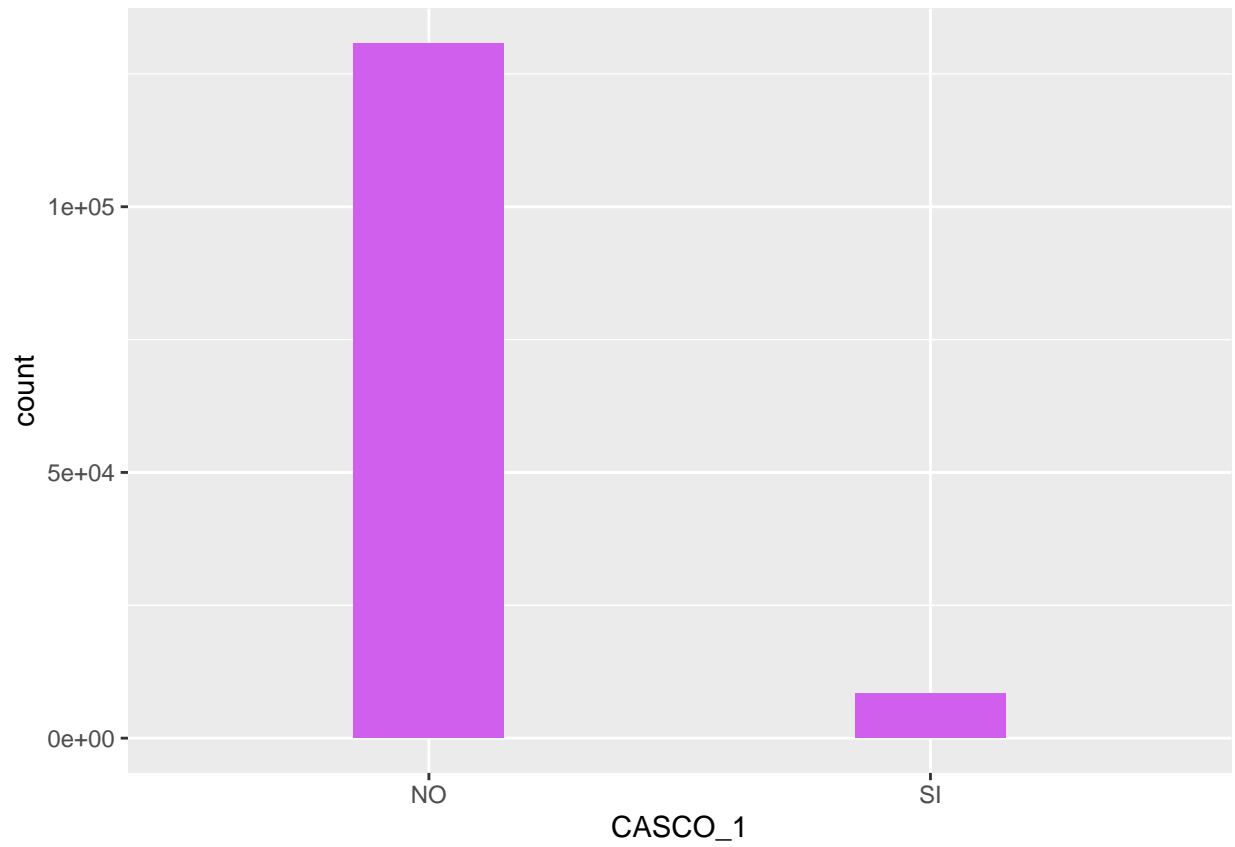
```
ggplot(data = df, aes(x=SEXO_1))+  
  geom_bar(  
    width = 0.3,  
    fill = "mediumorchid2"  
  )
```



```
ggplot(data = df, aes(x=PARTICIPANTE_1))+  
  geom_bar(  
    width = 0.3,  
    fill = "mediumorchid3"  
  )
```



```
ggplot(data = df, aes(x=CASCO_1))+  
  geom_bar(  
    width = 0.3,  
    fill = "mediumorchid2"  
  )
```



```
ggplot(data = df, aes(x=CINTURON_1))+  
  geom_bar(  
    width = 0.3,  
    fill = "mediumorchid2"  
  )
```