

# Proyecto 1

Fernando Stein Vallarta 165455      Pablo López Landeros 178863  
Manuel García Garduño 162136

## Introducción:

Para el proyecto utilizamos la base de datos de Servicio Línea de Mujeres que se encuentra disponible en la página de datos abiertos de la CDMX. En esta base de datos podemos encontrar la información de las llamadas realizadas al servicio Línea de Mujeres de LOCATEL desde noviembre de 2016. Sin embargo, para este proyecto únicamente consideramos los datos del año 2020.

La captura de la información se realiza a través de lo que el gobierno de la CDMX ha denominado Sistema de Registro de Información de Locatel (SIRILO). Cabe resaltar que la información obtenida es otorgada de forma opcional por las mujeres que llaman y requiere del consentimiento de las usuarias para ser difundida.

Para obtener la información, las operadoras realizan una entrevista a las mujeres que se comunican al servicio de asesorías. La atención que brindan las operadoras depende del motivo de la llamada ya que este permite a la operadora elaborar un plan de acción. Cabe resaltar que la información obtenida es otorgada de forma voluntaria por las mujeres que llaman y requiere del consentimiento de las usuarias para ser difundida.

En esta base podrás encontrar toda la información de las llamadas realizadas al servicio Línea de Mujeres de LOCATEL desde noviembre de 2016. Esta información se actualizará semanalmente a partir del 6 de abril de 2020 La Línea mujeres brinda atención enfocada a las mujeres que sufren de violencia, inequidad y desigualdad de género en los siguientes rubros:

1. Información sobre programas y acciones del gobierno de la CDMX
2. Asesoría psicológica
3. Asesoría Jurídica
4. Orientación médica

```
#Paquetes requeridos
library(tidyverse)
library(cowplot)
library(kableExtra)
library(knitr)
library(lubridate)
library(dplyr)
library(moments)
library(readr)
library(rgdal)
library(broom)
library(scales)
library(zoo)
library(knitr)
library(meme)
library(extrafont)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)#Con esto el código
#no debe salirse del pdf.
```

### *#Lectura de Datos y Diccionario*

```
linea.mujeres <- read_csv(paste("https://raw.githubusercontent.com/pablolopez2733/",
                                "Aplicada1/master/Bases%20de%20Datos/linea-mujeres.csv",
                                sep=""))

diccionario.linea <- read_csv(paste("https://github.com/pablolopez2733/Aplicada1/raw/master/",
                                    "Bases%20de%20Datos/DiccionarioLineaMujeres.csv",
                                    sep=""))
```

## Descripción de los datos

Para realizar una breve descripción de los datos, empleamos el diccionario proporcionado por la página de datos abiertos de la CDMX. Observamos que la base de datos se compone de la siguiente forma:

### *#Descripción de la base de datos usando el Diccionario*

```
glimpse(linea.mujeres)
```

```
## Rows: 66,315
## Columns: 27
## $ FOLIO <dbl> 656650, 656651, 656654, 656659, 656663, 656666, 6...
## $ FECHA_HORA_ALTA <dtm> 2020-02-11 14:01:00, 2020-02-11 14:01:00, 2020-0...
## $ DÍA_ALTA <dbl> 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 1...
## $ MES_ALTA <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ AÑO_ALTA <dbl> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2...
## $ SEXO <chr> "FEMENINO", "FEMENINO", "FEMENINO", "FEMENINO", "...
## $ EDAD <dbl> 83, 29, 13, 67, 39, 56, 60, 40, 68, 20, 69, 45, 6...
## $ ESTADO_CIVIL <chr> "SOLTERO (A)", "SOLTERO (A)", "SOLTERO (A)", "SOL...
## $ OCUPACION <chr> "HOGAR", "EMPLEADO (A)", "HOGAR", "EMPLEADO (A)",...
## $ ESCOLARIDAD <chr> "BACHILLERATO", "BACHILLERATO", "PRIMARIA", "SECU...
## $ ESTADO_USUARIA <chr> "CIUDAD DE MÉXICO", "ESTADO DE MÉXICO", "CIUDAD D...
## $ MUNICIPIO_USUARIA <chr> "CUAUHTÉMOC", "NEZAHUALCÓYOTL", "GUSTAVO A. MADER...
## $ COLONIA_USUARIA <chr> "ROMA NORTE", "EL SOL", "GUADALUPE PROLETARIA", "...
## $ CP_USUARIA <dbl> 6700, 57200, 7670, 1470, 9660, 7320, 9030, 57410,...
## $ ESTADO_HECHOS <chr> NA, NA, NA, NA, NA, "CIUDAD DE MÉXICO", NA, NA, N...
## $ MUNICIPIO_HECHOS <chr> NA, NA, NA, NA, NA, "GUSTAVO A. MADERO", NA, NA, ...
## $ COLONIA_HECHOS <chr> NA, NA, NA, NA, NA, "FOVISSSTE", NA, NA, NA, NA, ...
## $ CP_HECHOS <dbl> NA, NA, NA, NA, NA, 7320, NA, NA, NA, NA, NA, NA,...
## $ ORIGEN <chr> "ASESORÍA", "ASESORÍA", "ASESORÍA", "ASESORÍA", "...
## $ SERVICIO <chr> "PSICOLÓGICO", "PSICOLÓGICO", "PSICOLÓGICO", "JUR...
## $ TEMATICA_1 <chr> "VIOLENCIA", "CONFLICTOS EMOCIONALES", "SERVICIOS...
## $ TEMATICA_2 <chr> "OTRA", "ESTADOS DE ANIMO", "INFORMACION", NA, "D...
## $ TEMATICA_3 <chr> "SOLICITUD DE SERVICIOS DE APOYO", "ENFRENTAR SIT...
## $ TEMATICA_4 <chr> NA, NA, NA, NA, NA, "RELACIÓN PADRES E HIJOS", NA...
## $ TEMATICA_5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
## $ TEMATICA_6 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
## $ TEMATICA_7 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
```

```
variables <- diccionario.linea %>% select(`Nombre de la variable`, Definición)
kable(variables, booktabs = T) %>% kable_styling(latex_options = "striped")
```

Nombre de la variable	Definición
FOLIO	Identificador único de la llamada
FECHA_ALTA	Fecha en la que se generó la llamada de la usuaria (día, mes y año)
AÑO_ALTA	Año en que se generó la llamada de la usuaria.
MES_ALTA	Mes en que se generó la llamada de la usuaria.
DIAS_ALTA	Día en que se generó la llamada de la usuaria.
HORA_ALTA	Año en que se generó la llamada de la usuaria.
SEXO	Sexo de la usuaria que realizó la llamada.
EDAD	Edad de la usuaria que realizó la llamada.
ESTADO_CIVIL	Estado civil de la usuaria que realizó la llamada.
OCUPACION	Ocupación de la usuaria que realizó la llamada.
ESCOLARIDAD	Escolaridad de la usuaria que realizó la llamada.
ESTADO_USUARIA	Entidad Federativa de la usuaria que realizó la llamada.
MUNICIPIO_USUARIA	Municipio de la usuaria que realizó la llamada.
COLONIA_USUARIA	Colonia de la usuaria que realizó la llamada.
CP_USUARIA	Código Postal de la usuaria que realizó la llamada.
ESTADO_HECHOS	Estado en el que se registraron los hechos.
MUNICIPIO_HECHOS	Municipio en el que se registraron los hechos.
COLONIA_HECHOS	Colonia en la que se registraron los hechos.
CP_HECHOS	Código Postal en el que se registraron los hechos.
ORIGEN	Plataforma desde la que se inicio la llamada.
SERVICIO	Servicio que se le proporcionó a la usuaria.
TEMATICA_1	Primera temática en la que se categoriza la llamada.
TEMATICA_2	Segunda temática en la que se categoriza la llamada.
TEMATICA_3	Tercera temática en la que se categoriza la llamada.
TEMATICA_4	Cuarta temática en la que se categoriza la llamada.
TEMATICA_5	Quinta temática en la que se categoriza la llamada.
TEMATICA_6	Sexta temática en la que se categoriza la llamada.
TEMATICA_7	Séptima temática en la que se categoriza la llamada.

La base de datos no se limita a la CDMX, existe un porcentaje de llamadas provenientes del interior de la república y algunos estados del extranjero. La base de datos contiene 51,557 llamadas de la CDMX y 66,315 llamadas en total. Es decir, 14,758 llamadas vienen del interior de la república o el extranjero (Estados Unidos). Haremos un análisis general de la base de datos pero posteriormente nos enfocaremos en la Ciudad de México.

```
#Denuncias en CDMX solamente
denuncias.cdmx <- linea.mujeres %>% filter(ESTADO_USUARIA=="CIUDAD DE MÉXICO")
num_cdmx <- nrow(denuncias.cdmx)
num_tot <- nrow(linea.mujeres)
llam_int <- num_tot-num_cdmx

origen <- c('Llamadas totales','Llamadas CDMX','Llamadas Interior de la República')
denuncias <- c(num_tot, num_cdmx, llam_int)
porc_total=100
porc_cdmx <- (num_cdmx/num_tot)*100
porc_int <- 100-porc_cdmx

porcentaje <- c(porc_total,porc_cdmx,porc_int)
calls_df <- data.frame(origen, denuncias,porcentaje)
```

```
kable(calls_df, booktabs = T) %>% kable_styling(latex_options = "striped")
```

origen	denuncias	porcentaje
Llamadas totales	66315	100.00000
Llamadas CDMX	51557	77.74561
Llamadas Interior de la República	14758	22.25439

Ahora bien, en el presente proyecto buscamos estudiar múltiples situaciones que rodean a la realidad de discriminación por género basado en los datos obtenidos.

En primer lugar, nos gustaría analizar si desde la cuarentena ha existido un incremento en las llamadas realizadas a la Línea Mujeres. La Jornada Nacional de Sana Distancia (JNSD), o cuarentena comenzó el 23 de marzo del presente año. Luego entonces, construimos un histograma en donde observamos el número de llamadas por mes, contemplando que los meses de cuarentena son abril, mayo y algunos días de marzo y junio:

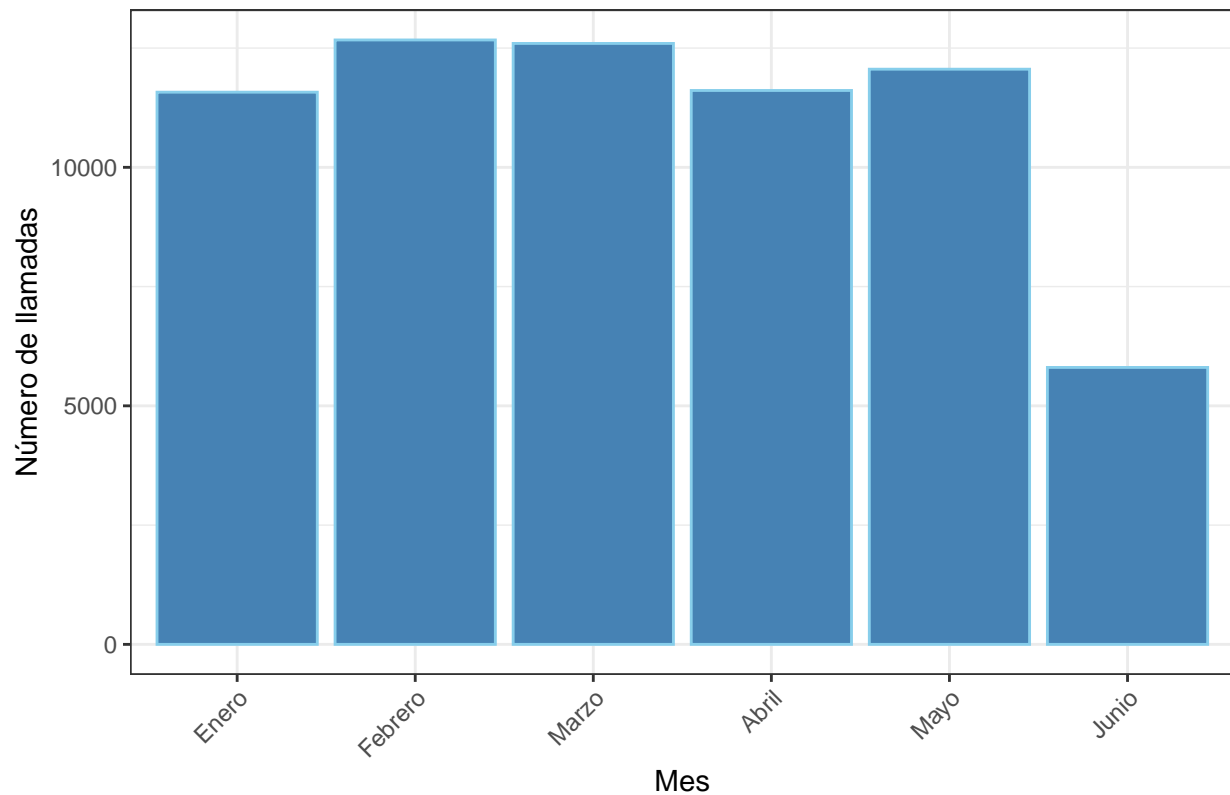
#### *#ANALISIS ESTADISTICO-DESCRIPTIVO*

*#En primer lugar, realizaremos un análisis de casos por mes*

```
llamadas.mes <- linea.mujeres %>%
  group_by(MES_ALTA) %>%
  count()
llamadas.mes$mes<-c("Enero", "Febrero", "Marzo", "Abril", "Mayo", "Junio")

ggplot(llamadas.mes) +
  geom_col(aes(x = reorder(mes, MES_ALTA), y = n),
           color='skyblue',
           fill='steelblue') +
  theme_bw() +
  labs(
    title = "Llamadas realizadas por mes a la Línea Mujeres",
    x = "Mes",
    y = "Número de llamadas"
  ) +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```

Llamadas realizadas por mes a la Línea Mujeres



En la anterior gráfica notamos que para los meses de abril, mayo y junio no se nota un aumento en el número de llamadas, de hecho, las llamadas han disminuido. Sin embargo, el mes de junio aún no ha concluido, de ahí que la gráfica anterior no parece ser contundente.

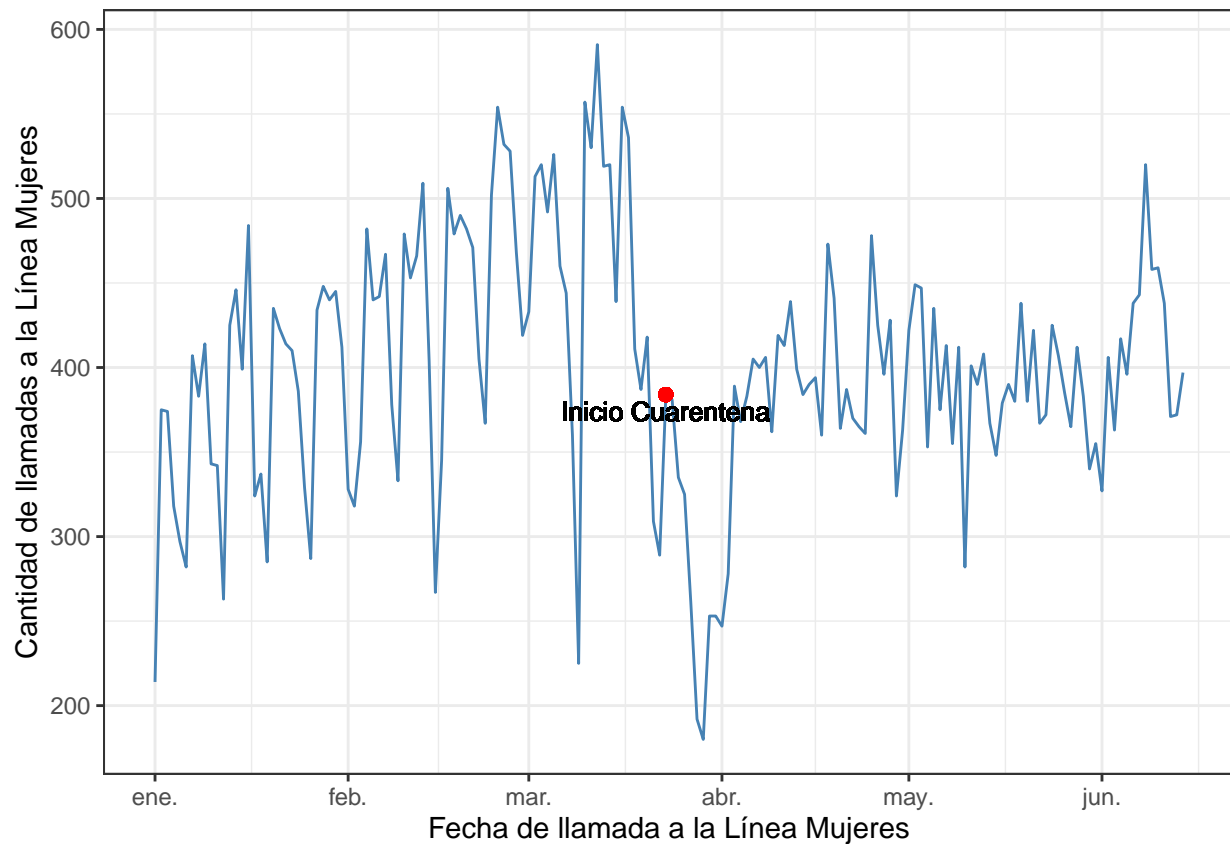
Ahora bien, en busca de un resultado más claro sobre si existe un aumento en las llamadas a la línea mujeres a partir del cuarentena realizamos otra gráfica, ahora de líneas. En esta gráfica es más clara la tendencia a la baja a raíz del comienzo de la cuarentena.

```
#PreCovid - PostCovid

#Inicia 23 - Marzo - 2020
linea.mujeres <- linea.mujeres %>%
  mutate(FECHA_HORA_ALTA = ymd_hms(FECHA_HORA_ALTA))
linea.mujeres <- linea.mujeres %>%
  mutate(fecha = date(FECHA_HORA_ALTA))
conteo_llamadas<-linea.mujeres %>%
  group_by(fecha) %>%
  tally() #llamadas por día

#Gráfica de Líneas por fecha
ggplot(conteo_llamadas) +
  geom_line(aes(x = fecha, y = n), colour = "steelblue") +
  theme_bw() +
  labs(
    x = "Fecha de llamada a la Línea Mujeres",
    y = "Cantidad de llamadas a la Línea Mujeres"
  ) +
```

```
geom_point(aes(x = dmy("23/03/2020"), y = 384), color = "red", size = 2) +
geom_text(aes(x = dmy("23/03/2020"), y = 384), label = "Inicio Cuarentena",
          size= 3.5, nudge_y= -10, colour = "black")
```



*#OJO : No se nota ninguna tendencia a partir del inicio de la cuarentena*

De igual forma, nos gustaría realizar un análisis por edad. En primera instancia, calculamos algunas estadísticas de resumen de la variable edad como su: desviación media absoluta, media, mediana, rango intercuartílico y varianza.

*#Análisis por edades*

*#Tabla de resumen*

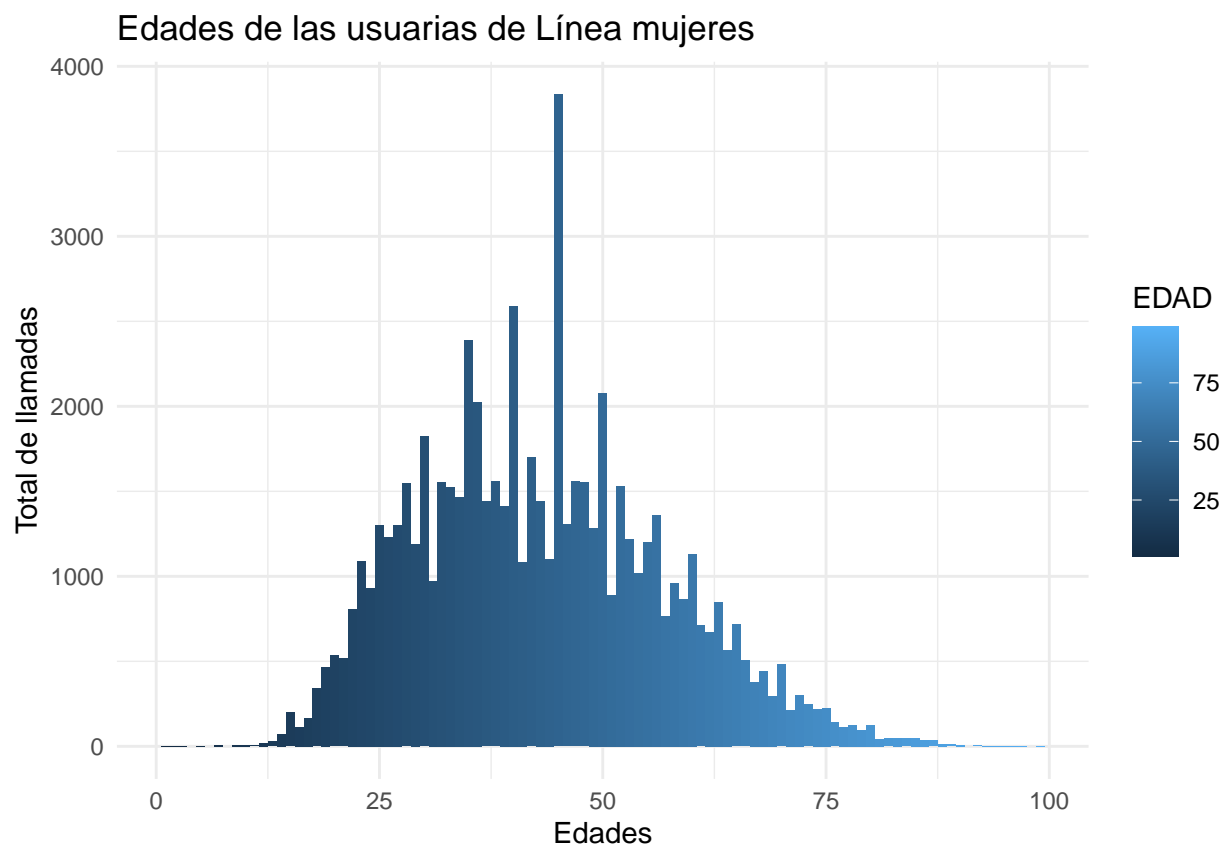
```
llamadas.tabla.edad <- linea.mujeres %>%
  summarise(MAD = mad(EDAD),
            Promedio = mean(EDAD),
            Mediana = median(EDAD),
            IQR = IQR(EDAD),
            Varianza = var(EDAD))

kable(llamadas.tabla.edad, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

MAD	Promedio	Mediana	IQR	Varianza
14.826	43.11089	42	19	192.6623

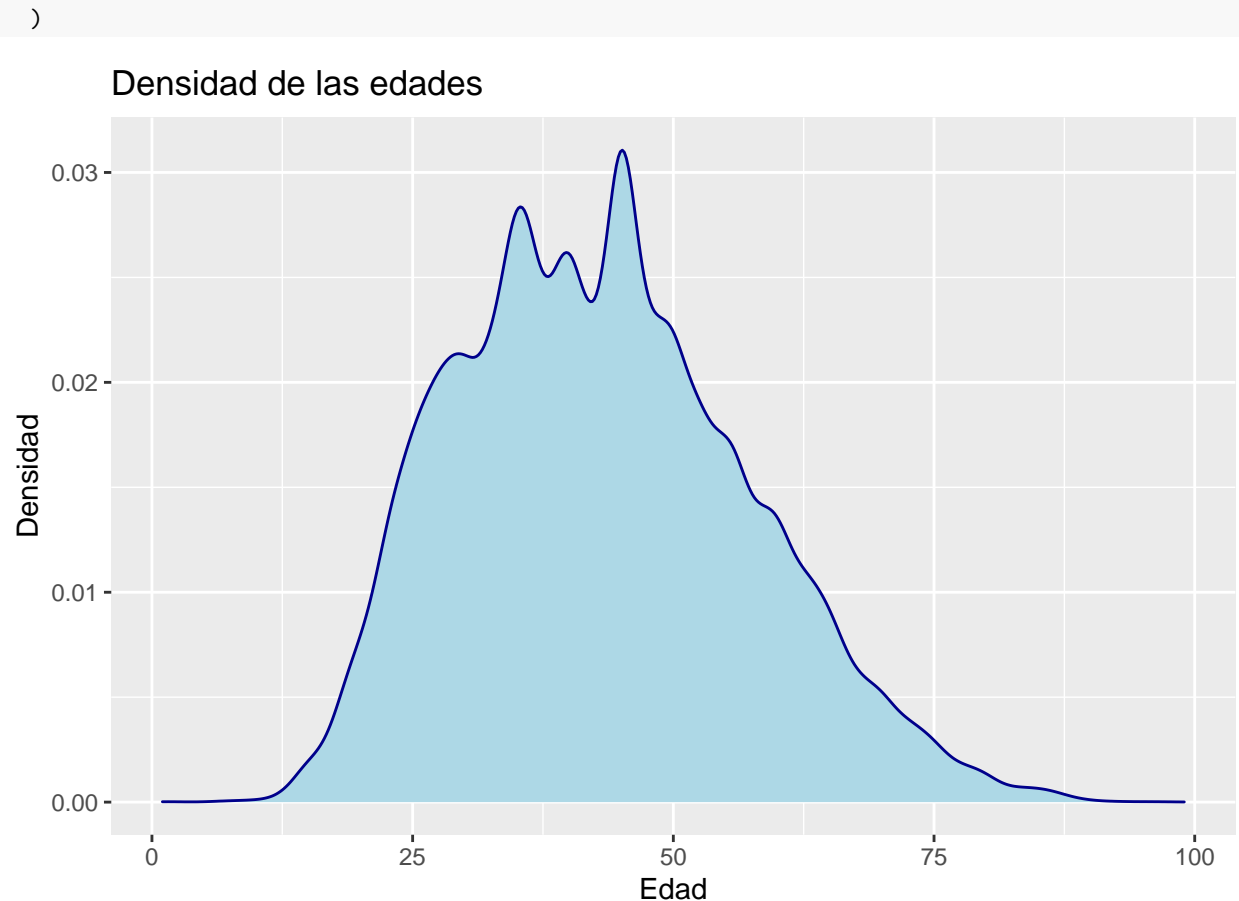
Con los anteriores resultados observamos que el promedio de edad esta alrededor de los 43 años, que el 50% de la población es menor a los 42 años, que, en promedio, el resto de las personas están a 14 años de 42 años, y que el 50% de la población centrada se encuentra en el intervalo [23, 61] años, y finalmente la variabilidad de las edades es mayor a 192. Si quisiéramos representar lo anterior en una gráfica, podemos usar una gráfica de barras.

```
llamadas.edad <- linea.mujeres %>% group_by(EDAD) %>%
  count()
ggplot(llamadas.edad)+
  geom_col(aes(x = EDAD, y = n, fill = EDAD)) +
  ggtitle("Edades de las usuarias de Línea mujeres") +
  theme_minimal() +
  labs(
    x = "Edades",
    y = "Total de llamadas"
  )
```



En la anterior gráfica, observamos que la distribución de las edades pareciera ser normal. De ahí que decidimos realizar un diagrama de densidad que valida nuestra hipótesis de normalidad, a pesar de tener algunos picos.

```
ggplot(linea.mujeres) +
  geom_density(aes(x=EDAD), color="darkblue", fill="lightblue")+
  labs(
    title = "Densidad de las edades",
    x = "Edad",
    y = "Densidad"
```



## Análisis de temática

Otras variables que resultan de interés para el análisis son las siguientes: estado civil, escolaridad, la temática principal de la llamada, el servicio brindado y el origen de la llamada. En particular, nos interesan las relaciones que existen entre las anteriores variables y su dinámica. En primera instancia, nos gustaría entender la dinámica entre el estado civil de las mujeres que utilizan la línea mujeres, y la ocupación. La siguiente gráfica presenta la anterior relación.

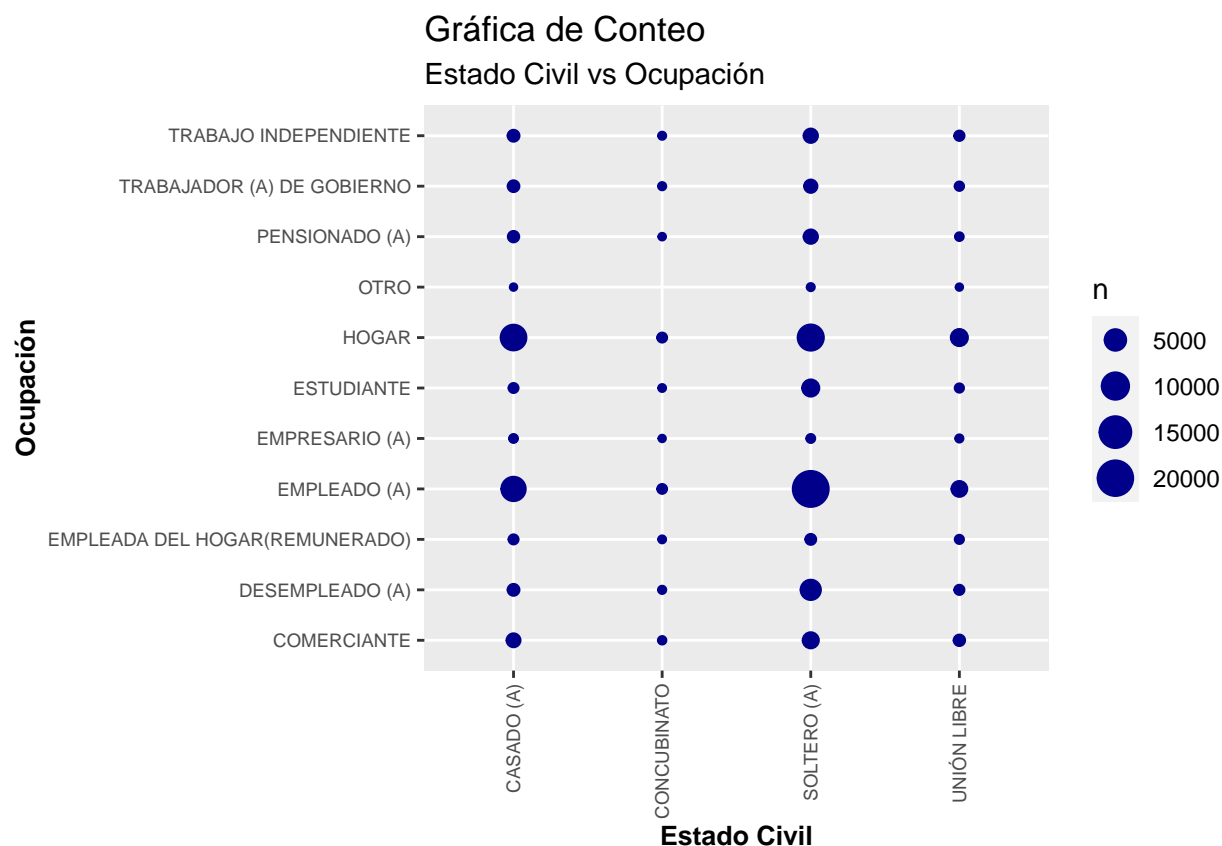
```
#Nos deshacemos de valores missing en variables de interés

linea.mujeres.filtered<-filter(linea.mujeres, !is.na(TEMATICA_1),
                                !is.na(ORIGEN), !is.na(SERVICIO),
                                !is.na(ESCOLARIDAD), !is.na(ESTADO_CIVIL),
                                !is.na(OCUPACION))

#Realizamos un Count Plot entre el Estado Civil y la ocupación
ggplot(linea.mujeres.filtered)+
  geom_count(aes(x =ESTADO_CIVIL , y = OCUPACION),
             color = "darkblue", show.legend=TRUE) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) +
  labs(subtitle="Estado Civil vs Ocupación",
       y="Ocupación",
       x="Estado Civil",
```



```
title="Gráfica de Conteo") +
theme(axis.text=element_text(size=7),
axis.title=element_text(size=10,face="bold"))
```



De la anterior gráfica, podemos concluir lo siguiente:

- La mayoría de las mujeres casadas suele dedicarse al hogar, a diferencia de las mujeres solteras, cuya mayoría son las empleadas.
- No se nota una tendencia importante en el estado civil de las usuarias de acuerdo con la ocupación, sin embargo, parece que el estado civil tiene una clara influencia sobre la ocupación de las mujeres.

Ahora bien, la atención que brinda la Línea Mujeres de la CDMX se clasifica en 3 tipos: médica, jurídica y psicológica dependiendo de la problemática que resalta la usuaria. De igual forma, Línea Mujeres puede ser contactada de distintas formas: llamar directamente al número de LOCATEL y solicitar una asesoría o atención inmediata; utilizar la línea de LOCATEL o los medios electrónicos de CDMX Contigo; marcar el número de emergencias 911 para ser redirigido a la Línea Mujeres, o utilizar los botones de apoyo localizados en las 16 delegaciones de Instituto Nacional de Mujeres.

La siguiente gráfica, busca analizar si la Línea Mujeres de LOCATEL ha logrado su objetivo de focalizar las llamadas relacionadas con respecto a la atención a las mujeres, y, de igual forma, estudia que tipo de servicio es el más solicitado.

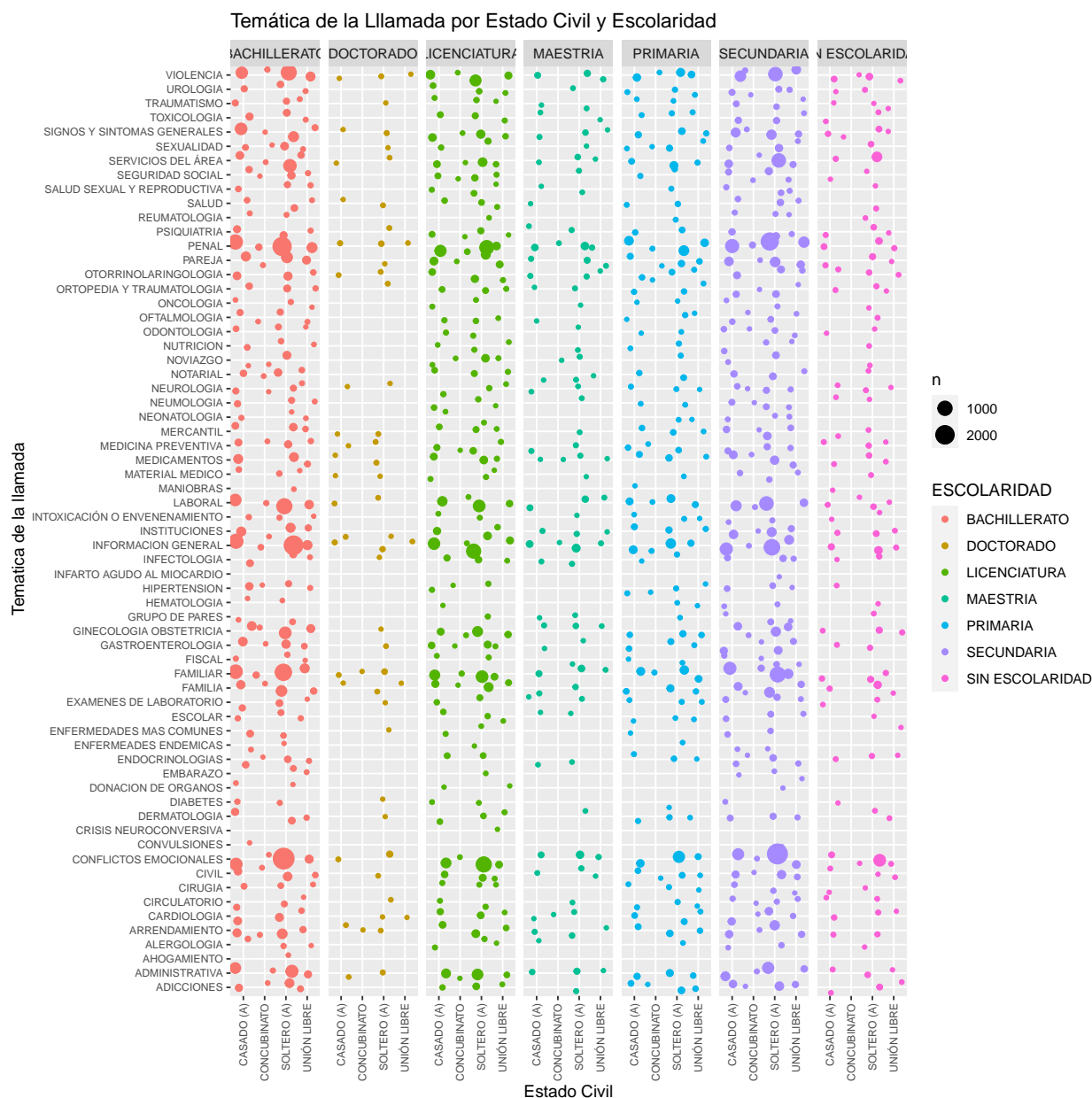
```
tematica.escolaridad <- linea.mujeres.filtered %>%
  group_by(TEMATICA_1, ESCOLARIDAD, ESTADO_CIVIL) %>%
  count()

ggplot(tematica.escolaridad)+
  labs(x = "Estado Civil",
```

```

y = " Temática de la llamada",
title="Temática de la Llamada por Estado Civil y Escolaridad") +
geom_jitter(aes(ESTADO_CIVIL, TEMATICA_1, size=n, color = ESCOLARIDAD)) +
theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
      axis.text=element_text(size=7)) +
facet_grid(cols = vars(ESCOLARIDAD))

```

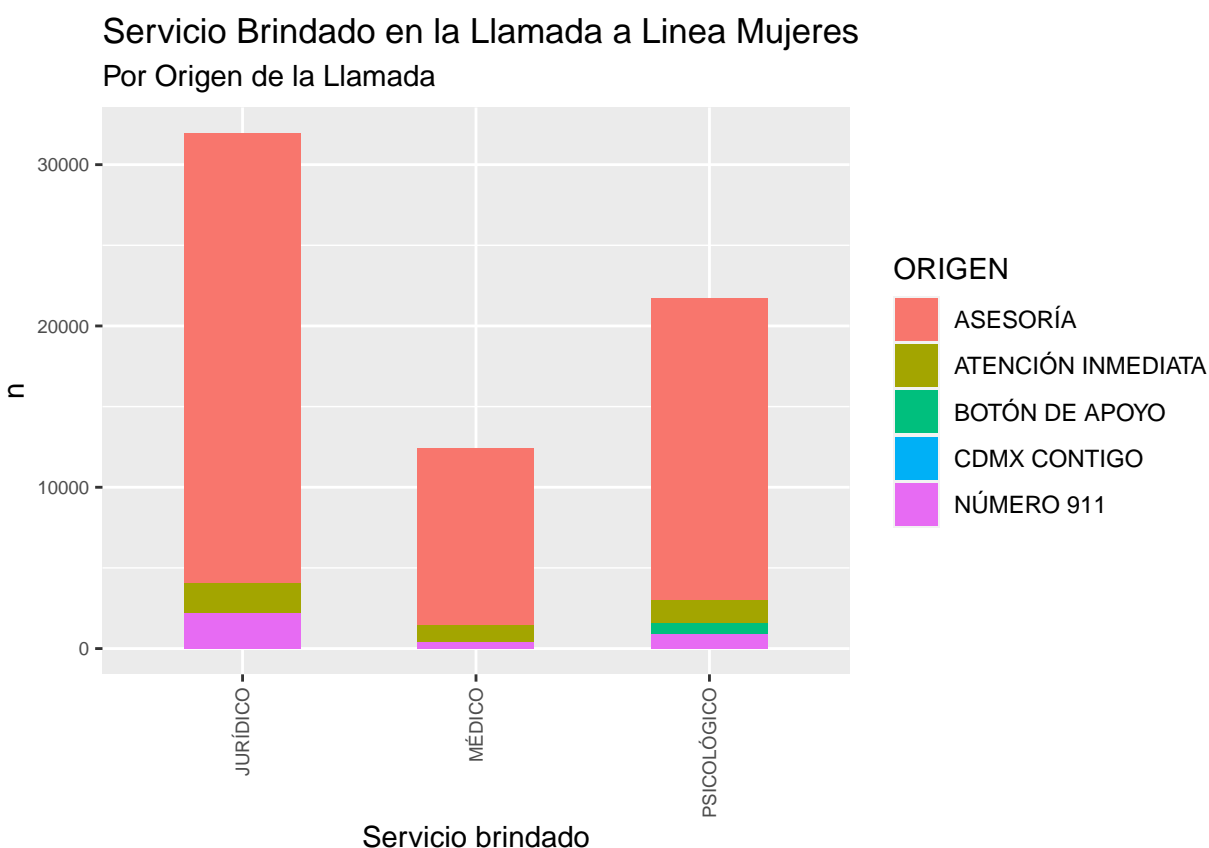


En la anterior gráfica observamos lo siguiente:

- El servicio mayormente brindado es jurídico, seguido de atención psicológica, y por último atención médica.
- Notemos que la mayoría de las llamadas proviene de la Línea Mujeres, ya sea en su modalidad de asesoría o de atención inmediata. Sin embargo, el número 911 sigue siendo utilizado.

- El servicio de CDMX Contigo y los botones de apoyo no tienen un gran volumen de llamadas. Asimismo, las asesoras de la Línea Mujeres enlistan la temática principal y las temáticas secundarias de la llamada realizada. Para este estudio, nos enfocaremos en la temática principal, y en particular estudiaremos la relación que existe entre la temática, el estado civil y el nivel de escolaridad de las usuarias.

```
#Realizamos un Bubble Plot
ggplot(linea.mujeres.filtered) +
  geom_bar(aes(SERVICIO, fill = ORIGEN), width = 0.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
        axis.text=element_text(size=7)) +
  labs(
    title = "Servicio Brindado en la Llamada a Linea Mujeres",
    subtitle = "Por Origen de la Llamada",
    x = "Servicio brindado",
    y = "n"
  )
```



Podemos obtener las siguientes conclusiones del anterior estudio:

- El nivel de escolaridad tiene un impacto sobre el número de denuncias. Podríamos establecer que las personas sin escolaridad podrían tener dificultades de acceso a la Línea Mujeres, de donde si nos olvidamos de las personas sin escolaridad notamos que a mayor escolaridad menor es el número de llamadas.
- Las temáticas de violencia, penal, información general y familia son comunes en todos los niveles de escolaridad. Mientras que otras como salud sexual y embarazo desaparecen en los niveles de escolaridad más altos.
- Las temáticas de conflictos emocionales, violencia, familiar, pareja y penal son las más comunes, sin

importar el nivel de escolaridad.

- Si ahora consideramos las temáticas de acuerdo al estado civil, podemos observar que la temática de violencia es más común en mujeres solteras que en los demás estados civiles, y en general, sin importar el nivel de escolaridad, las llamadas a la Línea Mujeres son realizadas en su mayoría por mujeres solteras.
- Resulta interesante que la temática de pareja es más recurrente en usuarias solteras que en usuarias en una relación. Debemos de resaltar que para el presente estudio no se realizó un enfoque de regresión lineal o no lineal para evidenciar las correlaciones que podrían existir entre las variables. Sin embargo, consideramos pertinente que se realice para poder “explicar” de mejor forma las correlaciones que parecen existir entre las variables.

## Ciudad de México

Nos parece importante enfocar una parte del análisis en la Ciudad de México ya que es el entorno dónde nos desenvolvemos la mayoría de los estudiantes de este curso. Como primer análisis quisimos explorar el número de llamadas a la línea mujeres por alcaldía:

```
#Recuento por delegaciones:
recuento_delegaciones <- denuncias.cdmx %>%
  group_by(MUNICIPIO_USUARIA) %>%
  tally()
#Imprimimos tabla:
kable(recuento_delegaciones, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

MUNICIPIO_USUARIA	n
ÁLVARO OBREGÓN	3420
AZCAPOTZALCO	4726
BENITO JUÁREZ	3733
COYOACÁN	4386
CUAJIMALPA DE MORELOS	1169
CUAUHTÉMOC	4422
GUSTAVO A. MADERO	5632
IZTACALCO	3199
IZTAPALAPA	8506
LA MAGDALENA CONTRERAS	1012
MIGUEL HIDALGO	2012
MILPA ALTA	396
TLÁHUAC	1492
TLALPAN	3011
VENUSTIANO CARRANZA	2935
XOCHIMILCO	1506

Vemos que la delegación Iztapalapa es la que más denuncias presenta. Mientras que en Milpa Alta es de dónde menos llamadas se han recibido.

A continuación, se muestra un mapa de la CDMX con la cantidad de denuncias por delegación:

```
#-----
#Denuncias por delegación en CDMX
#código base tomado de:
```

```

#https://rpubs.com/huanfaChen/ggplotShapefile
#-----

#Leemos el shapefile
setwd("C:/Users/pablo/Desktop/GithubRepos/Aplicada1/Proyecto1")
shape_cdmx <- readOGR(dsn = "./alcaldias.shp", layer = "alcaldias")

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\pablo\Desktop\GithubRepos\Aplicada1\Proyecto1\alcaldias.shp", layer: "alcaldias"
## with 16 features
## It has 5 fields

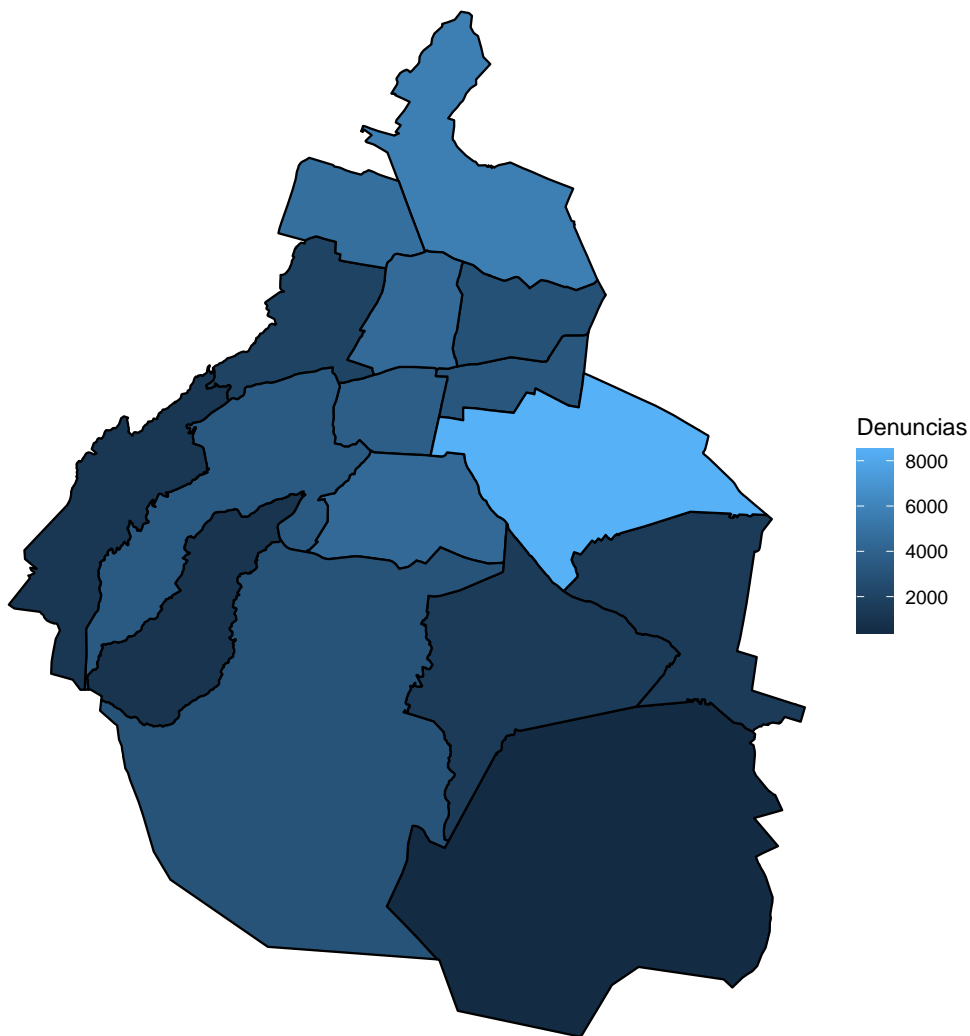
shp_df <- broom::tidy(shape_cdmx)

## Regions defined for each Polygons
#intentemos ponerle nombre a cada delegacion
#matcheamos el id de cada delegacion con su nombre
id=as.character(c(9,7,1,3,13,14,2,8,15,5,4,0,6,12,11,10))
recuento_delegaciones=cbind(recuento_delegaciones, id)
shp_df <- left_join(shp_df, recuento_delegaciones, by="id")
names(shp_df)[names(shp_df) == "n"] <- "Denuncias"
#merge the two datasets #standardise the values in the Total
#column to ensure we get an even spread of colours across the map rather
#than very light or dark colours for outliers
map <- ggplot() +
  geom_polygon(data = shp_df,
    aes(x = long, y = lat, group = group, fill=Denuncias),
    colour = "black")+
  labs(title="Llamadas a línea mujeres por alcaldía en la CDMX",
    subtitle = "Año:2020",
    caption = "Fuente: Gobierno de la CDMX")+
  theme(legend.title = "Llamadas recibidas")
map + theme_void()

```

## Llamadas a línea mujeres por alcaldía en la CDMX

Año:2020



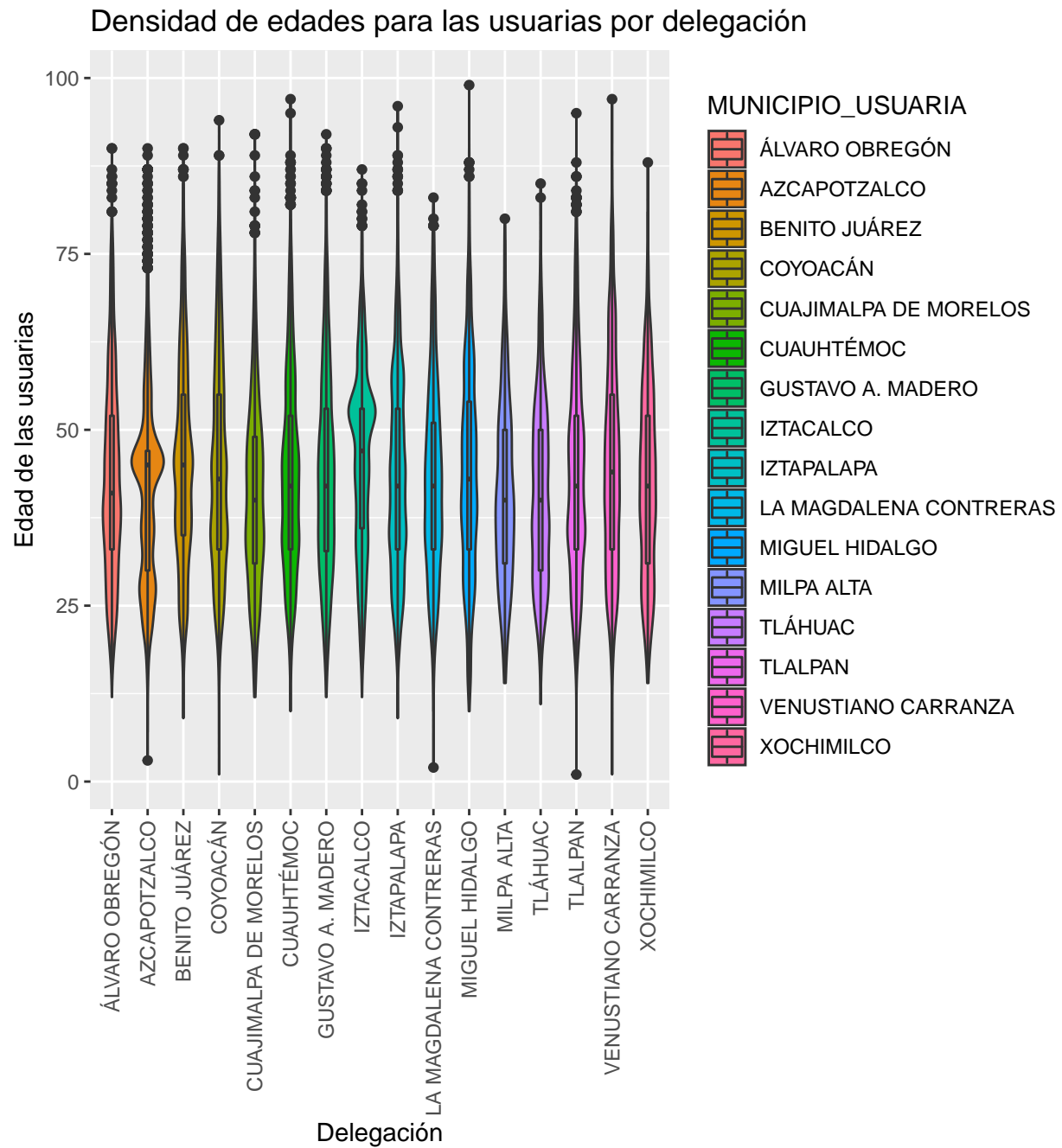
Fuente: Gobierno de la CDMX

Aquí confirmamos lo que ya se había mencionado y más aun, podemos percibir una fuerte diferencia de tonalidad en esta delegación con respecto al resto del mapa. Esto nos indica que sí existe un fuerte problema de violencia hacia el género femenino en esta delegación.

Continuando el análisis por delegación, interesa la edad de las usuarias al presentar las denuncias. Por lo tanto, a continuación se presenta un diagrama de violín para la edad de las usuarias por municipio:

```
#Gráfica de violín: edad por delegación-----
violin <- ggplot(denuncias.cdmx,
  aes(x=MUNICIPIO_USUARIA, y=EDAD , fill=MUNICIPIO_USUARIA)) +
  labs(x="Delegación",
    y="Edad de las usuarias",
    title="Densidad de edades para las usuarias por delegación")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  geom_violin()
```

```
violin + geom_boxplot(width=0.1)
```



En la gráfica anterior podemos ver:

1. La media para cada delegación está entre los 30 y los 40 años.
2. Si se puede rescatar algo “positivo”, es que parece haber pocas llamadas de menores de edad.
3. Todas las delegaciones parecen tener una distribución normal salvo Azcapotzalco e Iztacalco, para las cuales parece haber dos picos. Esto podría ser indicativo de otro tipo de distribución.

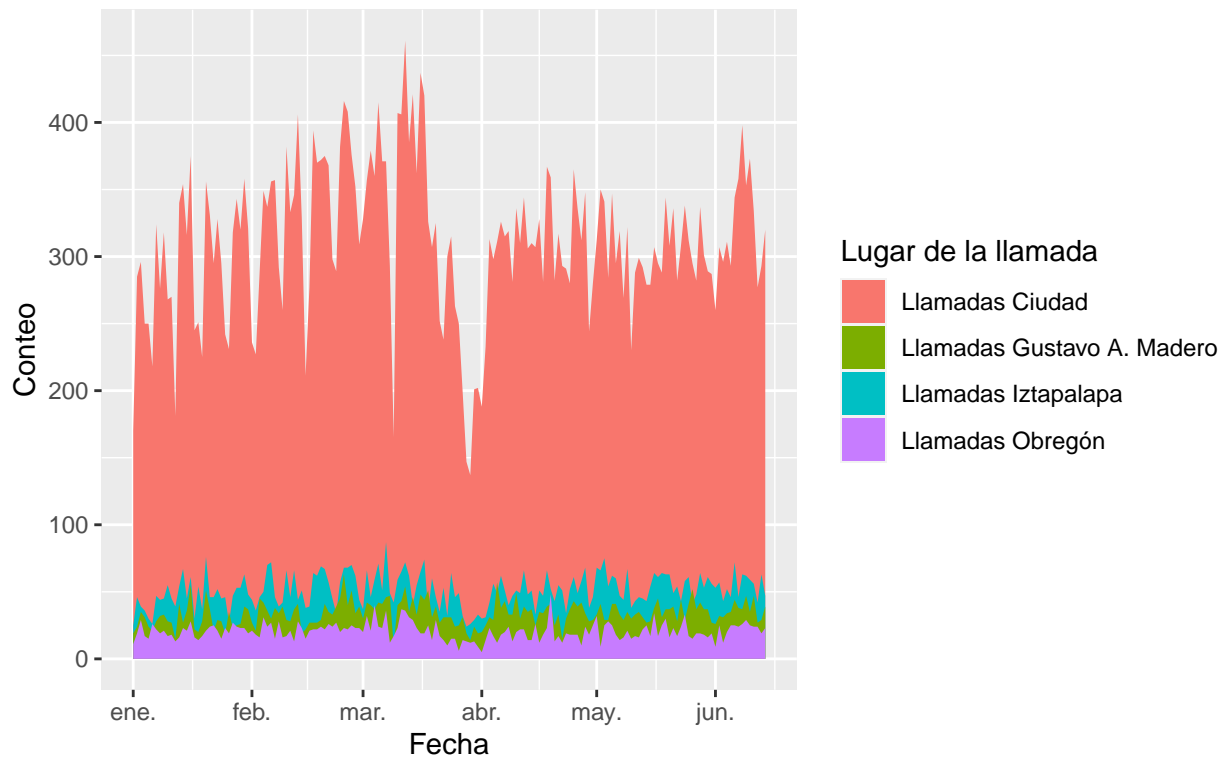
## Algunas delegaciones importantes

**Álvaro Obregón** En virtud de que como parte de la comunidad estudiantil del ITAM nos corresponde estar informados acerca del bienestar de nuestra comunidad y nuestro entorno, se optó por incluir en el análisis exploratorio el conteo de las llamadas realizadas desde la delegación Álvaro Obregón y compararlo con otras delegaciones de la Ciudad de México.

```
#Creamos una columna nueva que contenga la fecha de las denuncias en formato Date
linea.mujeres <- linea.mujeres %>%
  mutate(FECHA_HORA_ALTA = ymd_hms(FECHA_HORA_ALTA))
linea.mujeres <- linea.mujeres %>%
  mutate(fecha = date(FECHA_HORA_ALTA))
#####
#####
#GRAFICA 1
#Creamos una base de datos con el número de llamadas en la delegación Alvaro Obregón en relación a las
conteo_ciudad <- linea.mujeres%>%
  filter(str_detect(ESTADO_USUARIA,"CIUDAD DE"))%>%
  group_by(fecha)%>%
  tally()
conteo_obregon <- linea.mujeres%>%
  filter(str_detect(ESTADO_USUARIA,"CIUDAD DE") & str_detect(MUNICIPIO_USUARIA,"OBREG"))%>%
  group_by(fecha)%>%
  tally()
conteo_iztapalapa <- linea.mujeres%>%
  filter(str_detect(MUNICIPIO_USUARIA,"IZTAPALAPA"))%>%
  group_by(fecha)%>%
  tally()
conteo_madero <- linea.mujeres%>%
  filter(str_detect(MUNICIPIO_USUARIA,"MADERO"))%>%
  group_by(fecha)%>%
  tally()
obregon_vs_ciudad <- conteo_ciudad%>%
  mutate(O = conteo_obregon$n)%>%
  mutate(I = conteo_iztapalapa$n)%>%
  mutate(G = conteo_madero$n)
#Una vez construida la nueva base de datos, graficamos:
ggplot(obregon_vs_ciudad,aes(x=fecha))+
  geom_area(aes(y=n, fill = "Llamadas Ciudad"))+
  geom_area(aes(y=I, fill="Llamadas Iztapalapa"))+
  geom_area(aes(y=G, fill="Llamadas Gustavo A. Madero"))+
  geom_area(aes(y=O, fill="Llamadas Obregón"))+
  labs(title="Grafica de Area de Conteo de Llamadas",
       caption = "Fuente: Base de datos de la CDMX",
       fill="Lugar de la llamada",
       x="Fecha",
       y="Conteo")
```



## Grafica de Area de Conteo de Llamadas



Fuente: Base de datos de la CDMX

Podemos ver que, en lo que va del año 2020, el día donde hubo el conteo máximo de llamadas fue el 19 de abril con casi 50 llamadas, mientras que fue a principios de ese mismo mes en donde se registró el conteo mínimo. Además, tras una minuciosa observación, se puede notar que a finales del mes de marzo hay una notable disminución en el número de llamadas provenientes de la ciudad, así como en las delegaciones que se muestran en la gráfica.

**Iztapalapa** Anteriormente, se estableció que en la delegación Iztapalapa ocurre el mayor número de llamadas. Resulta entonces de interés analizar la distribución de las llamadas en esa delegación a lo largo del año. A continuación, se muestra un mapa de calor en forma de calendario. Colores rojos indican un bajo conteo de llamadas mientras que colores verdes indican un alto conteo.

```
#GRAFICA 2
#FUNCIONES QUE SERAN NECESARIAS PARA ESTA GRAFICA:
#FUNCION 1: recibe el numero de mes y que regrese la abreviacion del mes
month.abb <- function(x){
  indices <- seq(1,length(x))
  y <- indices
  for(i in indices){
    y[i] <- switch(x[i],"Ene","Feb","Mar",
                  "Abr","May","Jun","Jul",
                  "Ago","Sep","Oct","Nov","Dic")
  }
  return(y)
}

#FUNCION 2: recibe un vector de fechas y devuelve el número de la semana del mes
semana_del_mes <- function(x){
```

```

indices <- seq(1, length(x))
y <- indices
for(i in indices){
  dia <- day(x[i])
  semana <- ceiling(dia/7)
  y[i] <- semana
}
return(y)
}

#Creamos una base de datos alternativa para crear el calendario:
linea.mujeres.copia <- linea.mujeres%>%
  filter(str_detect(MUNICIPIO_USUARIA,"IZTAPALAPA"))%>%
  group_by(fecha)%>%
  tally()

linea.mujeres.copia$year <- year(linea.mujeres.copia$fecha)
linea.mujeres.copia$yearmonthf <- factor(as.yearmon(linea.mujeres.copia$fecha))
linea.mujeres.copia$monthf <- month.abb(month(linea.mujeres.copia$fecha))
linea.mujeres.copia$monthf <- factor(linea.mujeres.copia$monthf,
  levels = c("Ene","Feb","Mar",
             "Abr","May","Jun"))

linea.mujeres.copia$week <- week(linea.mujeres.copia$fecha)
linea.mujeres.copia$monthweek <- semana_del_mes(linea.mujeres.copia$fecha)
linea.mujeres.copia$weekdayf <- weekdays(linea.mujeres.copia$fecha)
levels(linea.mujeres.copia$weekdayf)

## NULL

linea.mujeres.copia$weekdayf <- factor(linea.mujeres.copia$weekdayf,
  levels = c("domingo","sábado",
             "viernes","jueves",
             "miércoles","martes",
             "lunes"), order= TRUE)

levels(linea.mujeres.copia$weekdayf)

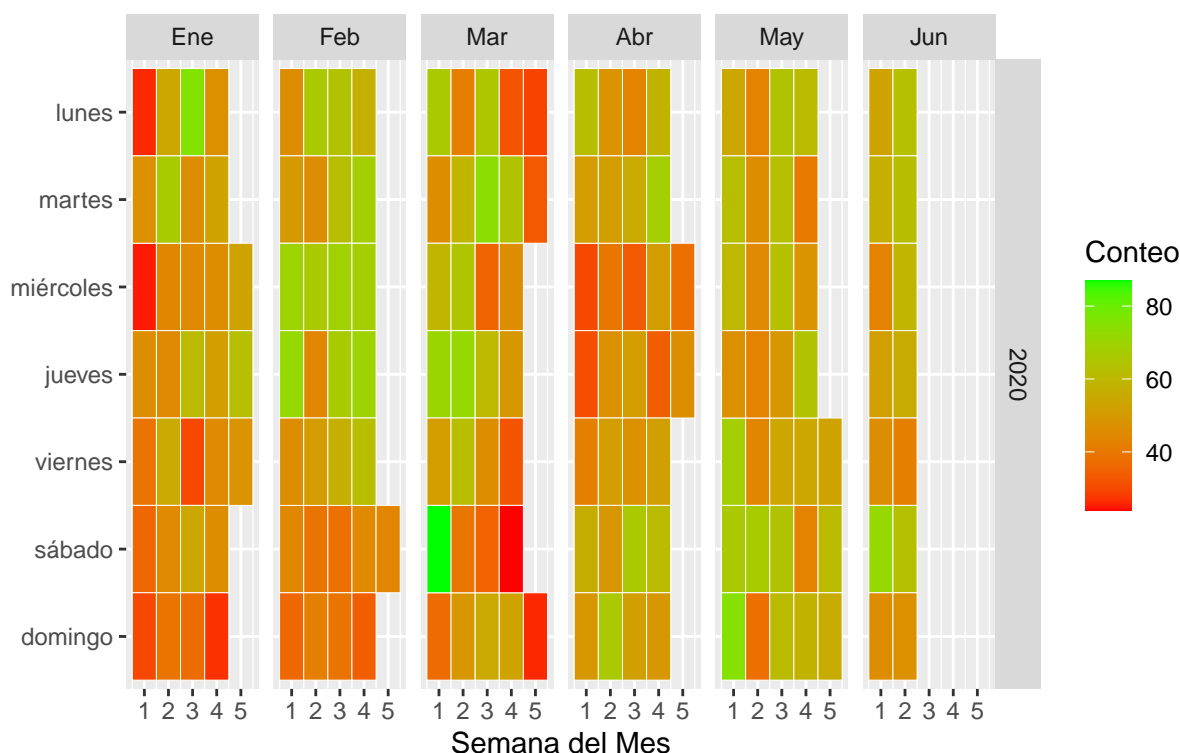
## [1] "domingo" "sábado" "viernes" "jueves" "miércoles" "martes"
## [7] "lunes"

#Una vez construida la nueva base de datos graficamos
ggplot(linea.mujeres.copia, aes(monthweek, weekdayf, fill = n)) +
  geom_tile(colour = "white") +
  facet_grid(year~monthf) +
  scale_fill_gradient(low="red", high="green") +
  labs(x="Semana del Mes",
       y="",
       title = "Calendario de Mapa de Calor",
       subtitle="Conteo de llamadas",
       fill="Conteo")

```

## Calendario de Mapa de Calor

### Conteo de Llamadas



Tras una minuciosa observación se puede notar que existe una ligera tendencia a que el número de llamadas disminuya en los días domingo. La cuarta semana de febrero fue una de las que presentó mayor volumen de llamadas, mientras que la primera semana de enero fue la que recibió menos llamadas en lo que va del año.

## Conclusión

Como conclusiones generales, debemos resaltar que la Línea Mujeres parece tener un impacto positivo. A pesar de que solamente lleva alrededor de 3 años en operación registra más de 10,000 llamadas mensuales para los meses del año en curso. Lo que representa más de 300 llamadas diarias a la Línea. Además, parece que tiene un impacto significativo sobre las mujeres con niveles de escolaridad bajos que son las que con mayor frecuencia utilizan el servicio. Sin embargo, debemos de notar que a partir del aislamiento por la pandemia se han disminuido el número de llamada, lo que nos lleva a pensar que el hecho de que las mujeres se encuentren en casa con sus parejas podría limitar las posibilidades de hacer la llamada. Luego entonces, existen algunas limitantes al uso del servicio.

Ahora bien, el número de temáticas principales que son discutidas en las llamadas es muy alto. Existen algunas temáticas que no tienen que ver con el objetivo fundamental de la Línea Mujeres, de ahí que sugerimos limitar el espectro de temáticas posibles para poder realizar un análisis más preciso. De igual forma, la Línea Mujeres es un servicio que se propone para la CDMX, sin embargo, en la base de datos existen entradas que son de otros estados de la República e incluso de otros países. De ahí que nos parece pertinente recomendar el uso a nivel nacional.

#Código tomado de:  
[#https://cran.r-project.org/web/packages/meme/vignettes/meme.html](https://cran.r-project.org/web/packages/meme/vignettes/meme.html)

```

#Ajuste si la computadora es Windows
if (.Platform$OS.type == "windows") {
  windowsFonts(
    Impact = windowsFont("Impact"),
    Courier = windowsFont("Courier")
  )
}
#Lectura de imagen
img<-"http://i0.kym-cdn.com/entries/icons/mobile/000/000/745/success.jpg"
#Creacion del Meme
x <- meme(img)
plot(x, size = 2, "Proyecto", "Terminado", color = "white", font = "Courier")

```

Un bonito meme para cerrar:

