

ggplot2

willandru

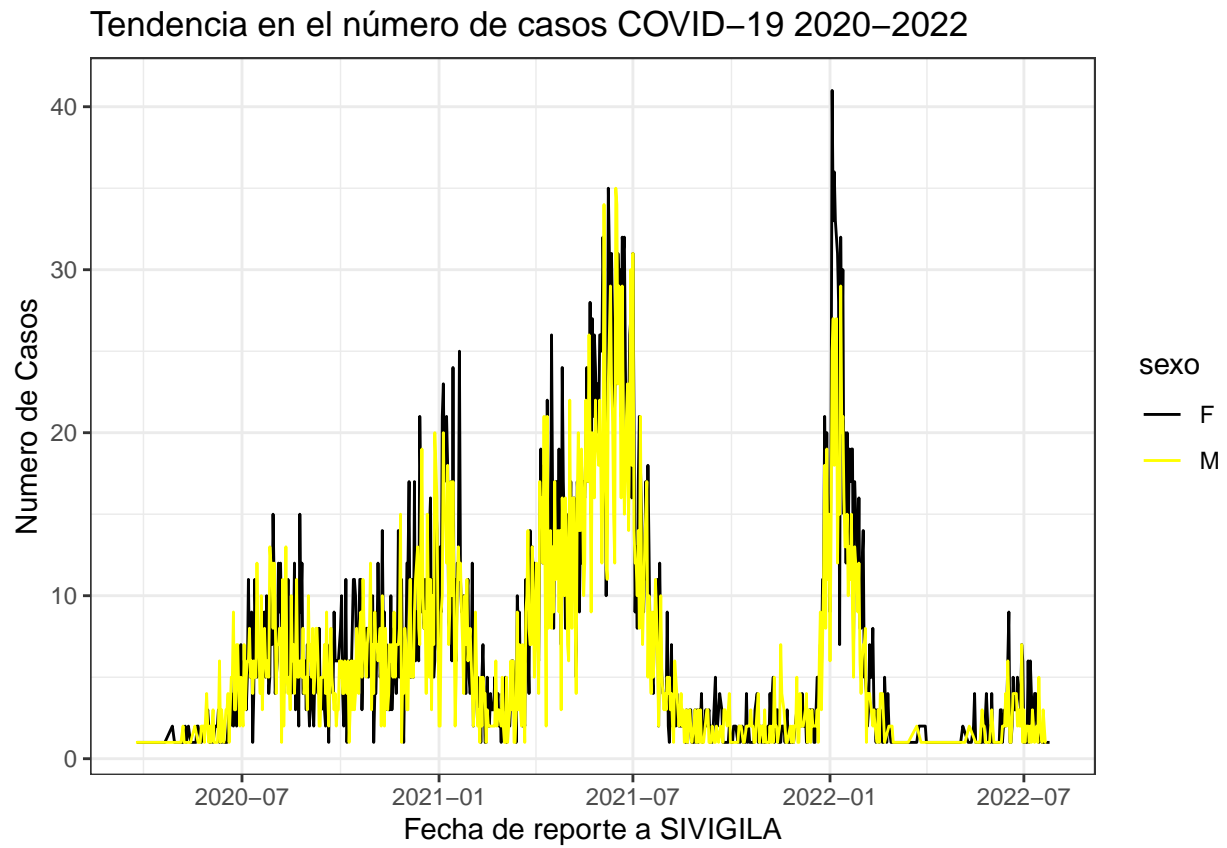
2023-12-03

## Carga de Datos

```
url <- "https://github.com/TRACE-LAC/TRACE-LAC-data/blob/main/otros/muestra_covid.RDS?raw=true"
datos <- readr::read_rds(url)
```

## Ejercicio Recrear Gráfico

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



# Análisis Exploratorio de Datos (EDA)

```
str(datos)
```

```
## tibble [10,000 x 23] (S3: tbl_df/tbl/data.frame)
## $ fecha_reporte_web      : POSIXct[1:10000], format: "2021-06-29" "2021-04-23" ...
## $ id_de_caso             : num [1:10000] 4208058 2737390 3576919 2944674 2300590 ...
## $ fecha_de_notificacion  : POSIXct[1:10000], format: "2021-06-15" "2021-04-19" ...
## $ codigo_divipola_departamento: num [1:10000] 68 5 68 11 13001 ...
## $ nombre_departamento   : chr [1:10000] "SANTANDER" "ANTIOQUIA" "SANTANDER" "BOGOTA" ...
## $ codigo_divipola_municipio : num [1:10000] 68001 5001 68001 11001 13001 ...
## $ nombre_municipio       : chr [1:10000] "BUCARAMANGA" "MEDELLIN" "BUCARAMANGA" "BOGOTA" ...
## $ edad                   : num [1:10000] 39 73 15 45 22 20 45 48 71 24 ...
## $ sexo                   : chr [1:10000] "F" "F" "M" "F" ...
## $ tipo_de_contagio       : chr [1:10000] "Relacionado" "Comunitaria" "Relacionado" "Comunitaria" ...
## $ ubicacion_del_caso     : chr [1:10000] "Casa" "Casa" "Casa" "Casa" ...
## $ estado                 : chr [1:10000] "Leve" "Leve" "Leve" "Leve" ...
## $ recuperado             : chr [1:10000] "Recuperado" "Recuperado" "Recuperado" "Recuperado" ...
## $ fecha_de_inicio_de_sintomas : POSIXct[1:10000], format: "2021-06-10" "2021-04-16" ...
## $ fecha_de_muerte        : Date[1:10000], format: NA NA ...
## $ fecha_de_diagnostico    : POSIXct[1:10000], format: "2021-06-26" "2021-04-20" ...
## $ fecha_recuperacion     : POSIXct[1:10000], format: "2021-06-30" "2021-04-30" ...
## $ tipo_de_recuperacion    : chr [1:10000] "Tiempo" "Tiempo" "Tiempo" "Tiempo" ...
## $ pertenencia_etnica     : num [1:10000] 6 6 6 6 6 6 6 6 6 6 ...
## $ nombre_del_grupo_etnico : chr [1:10000] NA NA NA NA ...
## $ fecha_nacimiento       : num [1:10000] 1983 1949 2007 1977 2000 ...
## $ retraso_notificacion    : num [1:10000] 5 3 5 12 7 4 1 4 4 1 ...
## $ tiempo_recuperacion    : num [1:10000] 20 14 20 19 14 14 14 14 18 213 ...
```

```
glimpse(datos)
```

```
## Rows: 10,000
## Columns: 23
## $ fecha_reporte_web      <dtm> 2021-06-29, 2021-04-23, 2021-06-07, 2021-
## $ id_de_caso             <dbl> 4208058, 2737390, 3576919, 2944674, 23005~
## $ fecha_de_notificacion  <dtm> 2021-06-15, 2021-04-19, 2021-05-24, 2021-
## $ codigo_divipola_departamento <dbl> 68, 5, 68, 11, 13001, 54, 66, 5, 81, 25, ~
## $ nombre_departamento   <chr> "SANTANDER", "ANTIOQUIA", "SANTANDER", "B~
## $ codigo_divipola_municipio <dbl> 68001, 5001, 68001, 11001, 13001, 54001, ~
## $ nombre_municipio       <chr> "BUCARAMANGA", "MEDELLIN", "BUCARAMANGA",~
## $ edad                   <dbl> 39, 73, 15, 45, 22, 20, 45, 48, 71, 24, 3~
## $ sexo                   <chr> "F", "F", "M", "F", "M", "M", "M", "F", "~
## $ tipo_de_contagio       <chr> "Relacionado", "Comunitaria", "Relacionad~
## $ ubicacion_del_caso     <chr> "Casa", "Casa", "Casa", "Casa", "Casa", "~
## $ estado                 <chr> "Leve", "Leve", "Leve", "Leve", "Leve", "~
## $ recuperado             <chr> "Recuperado", "Recuperado", "Recuperado",~
## $ fecha_de_inicio_de_sintomas <dtm> 2021-06-10, 2021-04-16, 2021-05-19, 2021-
## $ fecha_de_muerte        <date> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ fecha_de_diagnostico    <dtm> 2021-06-26, 2021-04-20, 2021-06-04, 2021-
## $ fecha_recuperacion     <dtm> 2021-06-30, 2021-04-30, 2021-06-08, 2021-
## $ tipo_de_recuperacion    <chr> "Tiempo", "Tiempo", "Tiempo", "Tiempo", "~
```

```
## $ pertenencia_etnica      <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ~
## $ nombre_del_grupo_etnico <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ fecha_nacimiento       <dbl> 1983, 1949, 2007, 1977, 2000, 2002, 1977, ~
## $ retraso_notificacion   <dbl> 5, 3, 5, 12, 7, 4, 1, 4, 4, 1, 0, NA, 1, ~
## $ tiempo_recuperacion    <dbl> 20, 14, 20, 19, 14, 14, 14, 14, 18, 213, ~
```

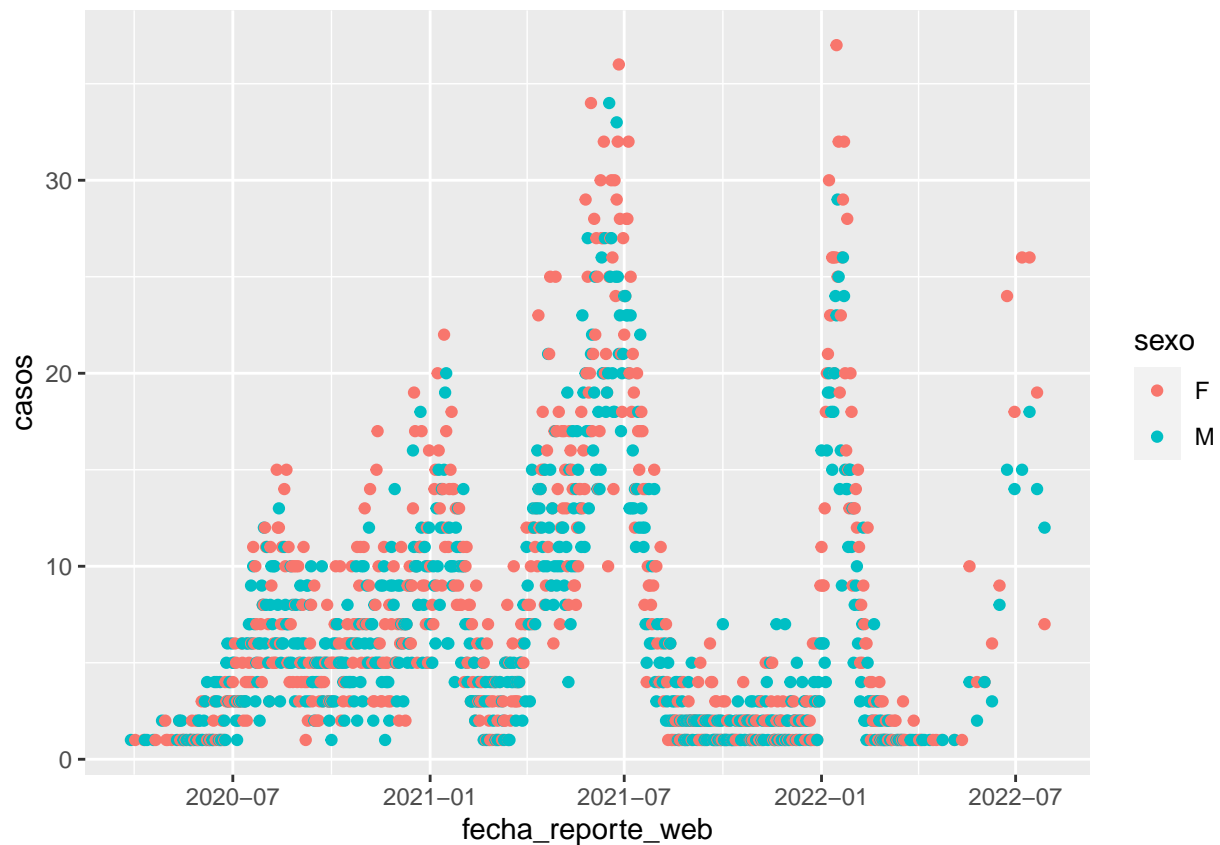
## Gráficos Exploratorios

### Gráfico de Puntos por Fecha y Sexo

```
COVID <- datos %>% group_by(fecha_reporte_web, sexo) %>%
  summarise(casos = n())
```

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

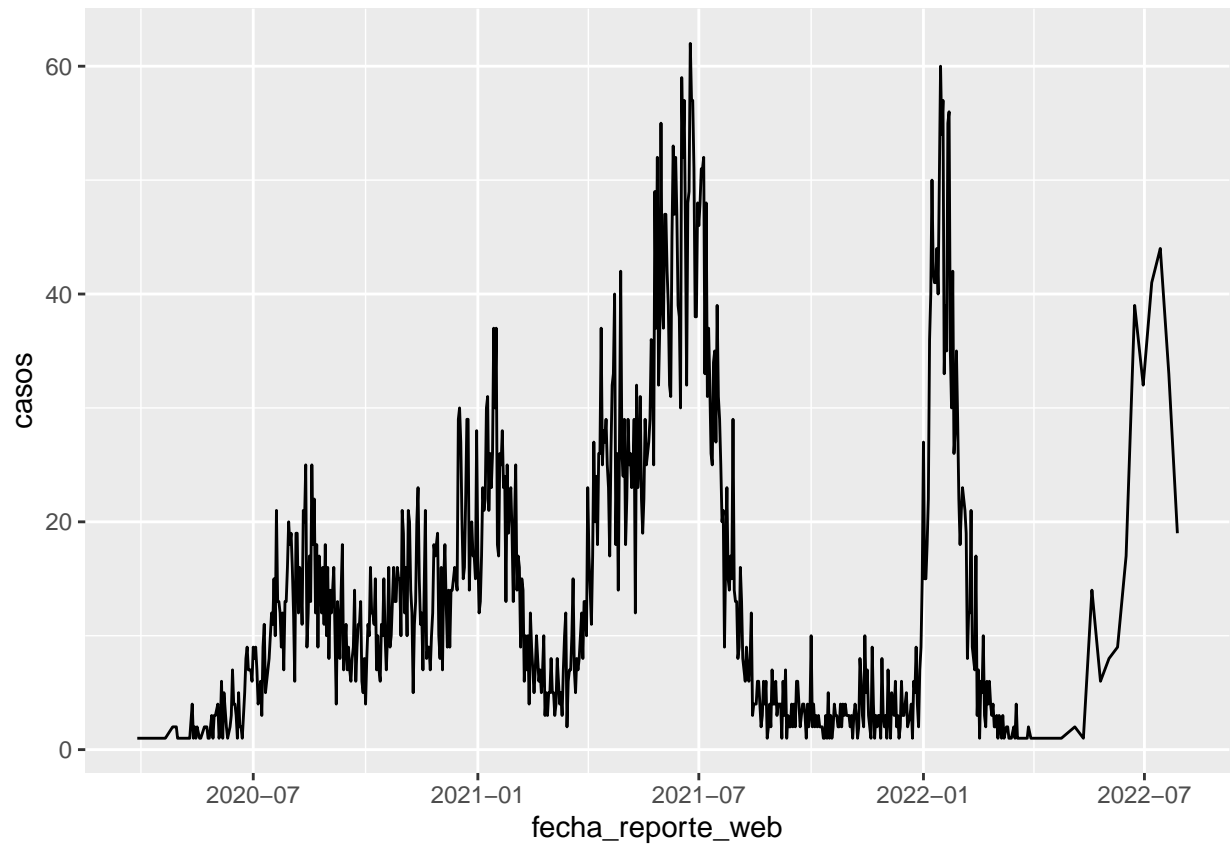
```
ggplot(data = COVID, aes(x = fecha_reporte_web, y = casos, colour = sexo)) +
  geom_point()
```



### Gráfico de Líneas por Fecha

```
COVID_FECHA <- datos %>% group_by(fecha_reporte_web) %>%
  summarise(casos = n())

ggplot(data = COVID_FECHA, aes(x = fecha_reporte_web, y = casos)) +
  geom_line()
```



### Gráfico de Barras por Sexo

```
ggplot(data = datos) +
  geom_bar(aes(x = sexo))
```

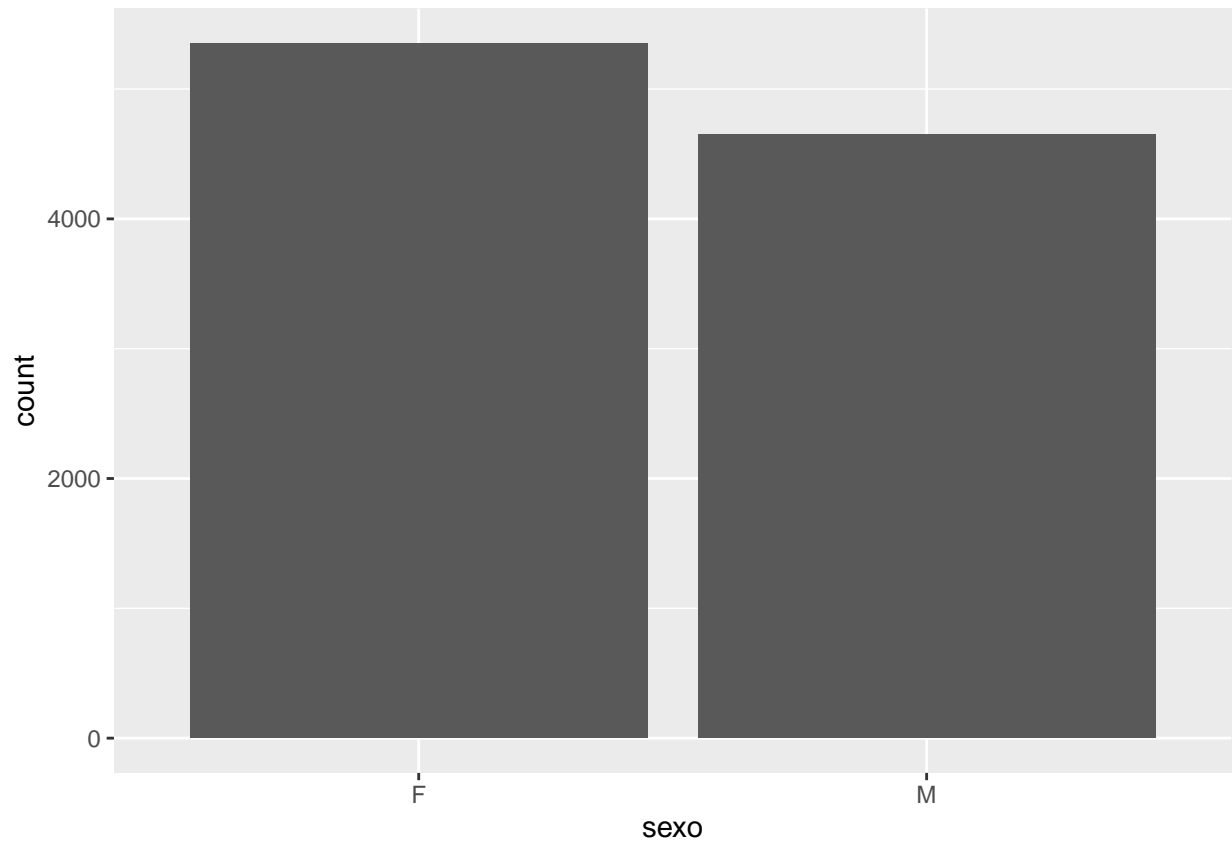


Gráfico de Barras por Departamento

```
COVID_DEPAS <- datos %>% group_by(nombre_departamento) %>% summarise(casos = n())  
  
ggplot(data = COVID_DEPAS, aes(x = reorder(nombre_departamento, -casos), y = casos)) +  
  geom_bar(stat = "identity") + coord_flip()
```

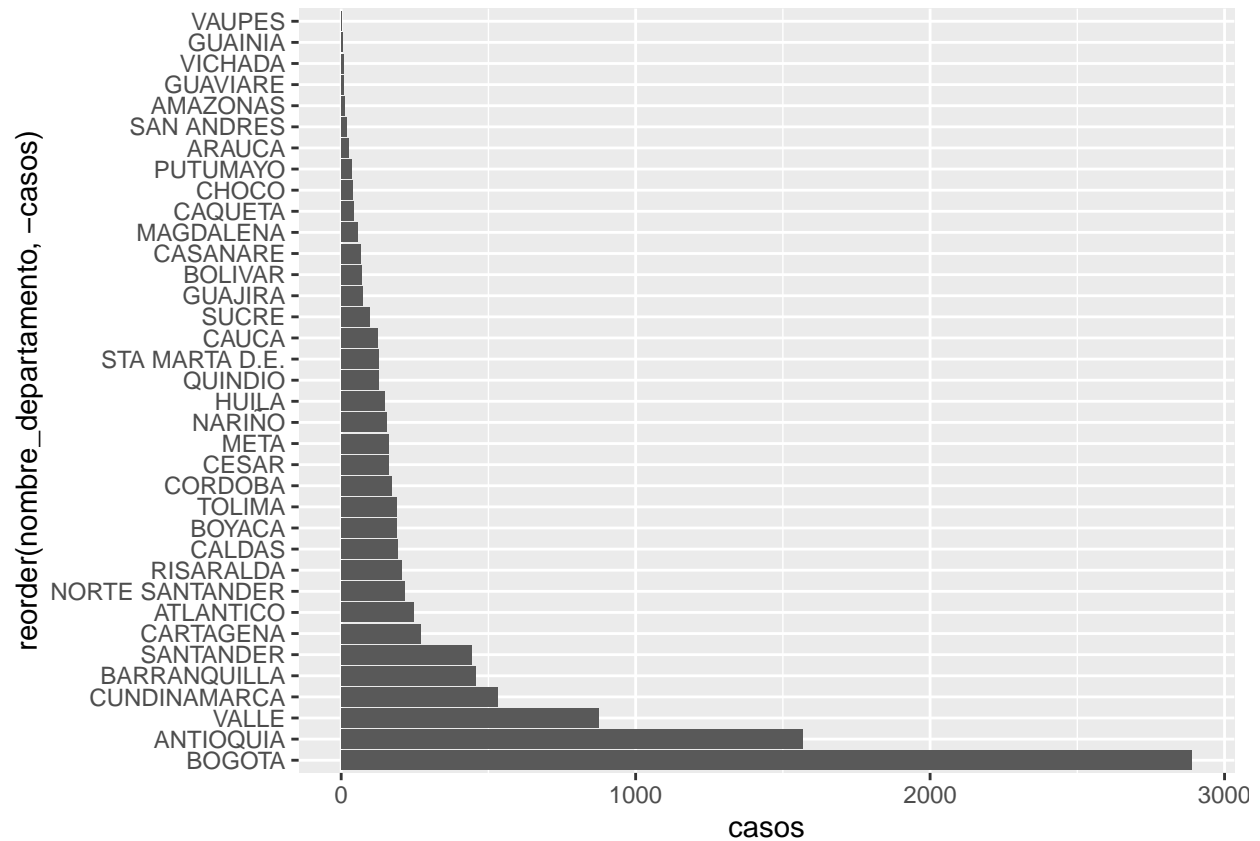
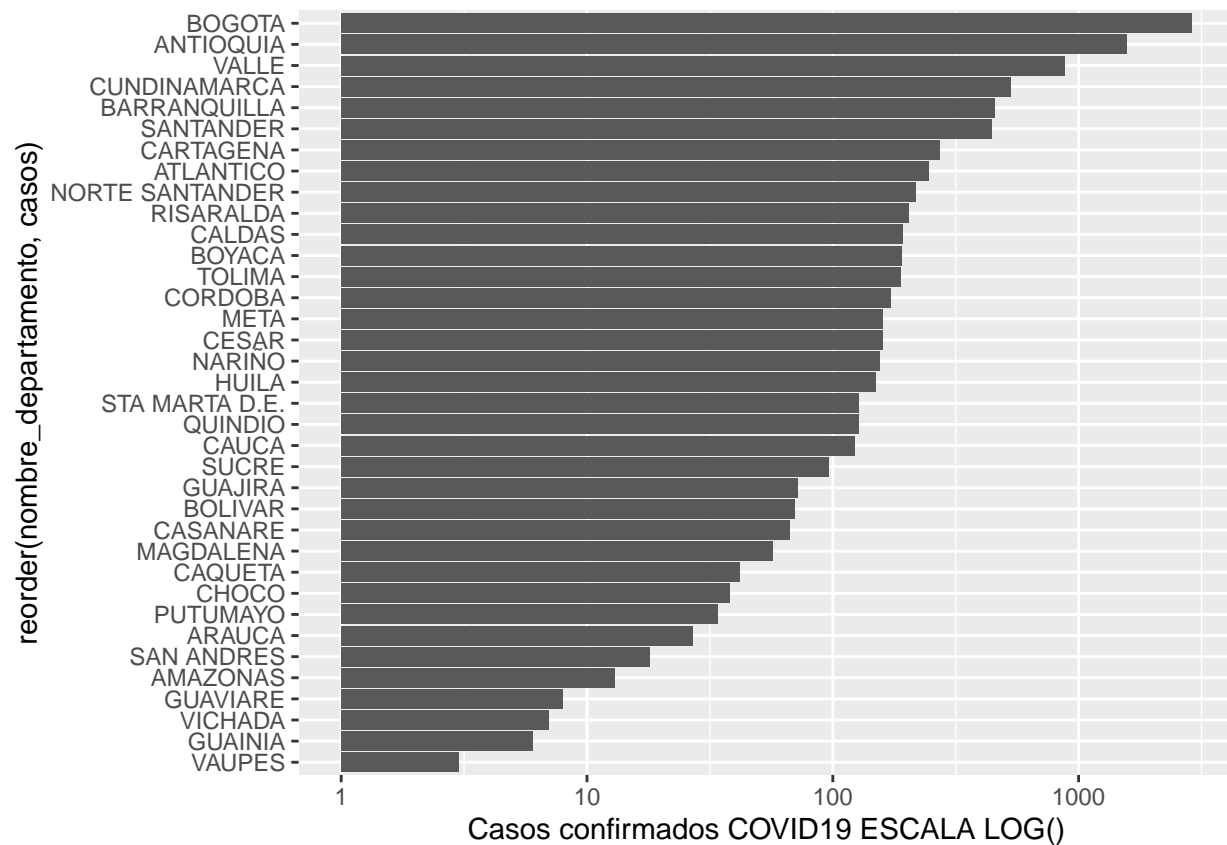


Gráfico de Barras por Departamento con Escala Logarítmica

```
ggplot(data = COVID_DEPAS, aes(x = reorder(nombre_departamento, casos), y = casos)) +
  geom_bar(stat = "identity") + coord_flip() + scale_y_log10(name = 'Casos confirmados COVID19 ESCALA L
```



## Gráficos con Facetas y Temas

### Gráfico de Puntos por Edad y Sexo con Facetas

```
COVID_SEXO <- datos %>% group_by(edad, sexo) %>%
  summarise(casos = n())
```

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

```
ggplot(COVID_SEXO, aes(x = edad, y = casos, colour = casos)) +
  geom_point() +
  facet_wrap(~sexo) +
  scale_color_gradient2()
```

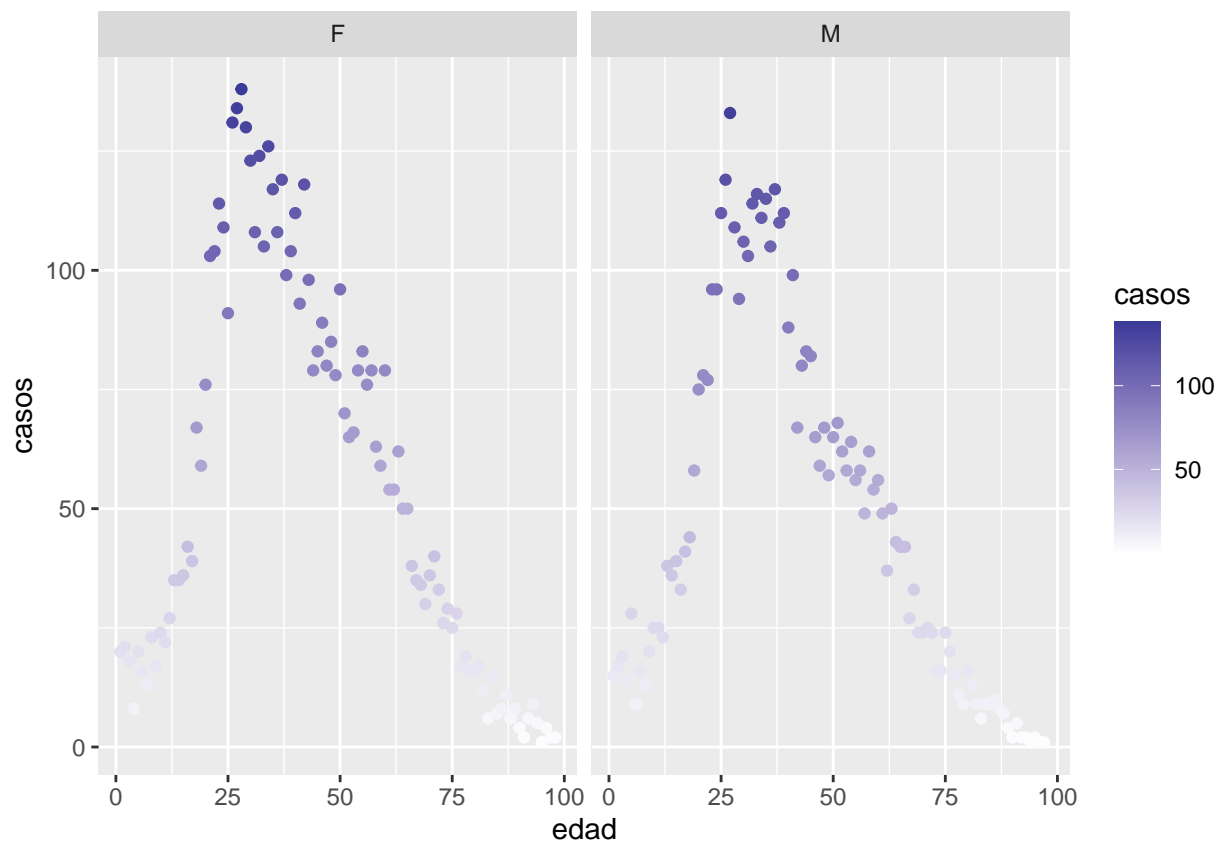


Gráfico de Puntos por Edad y Sexo con Tema Clásico

```
ggplot(data = COVID_SEX0, aes(x = edad, y = casos)) +  
  geom_point() +  
  facet_wrap(~sexo) +  
  theme_classic()
```



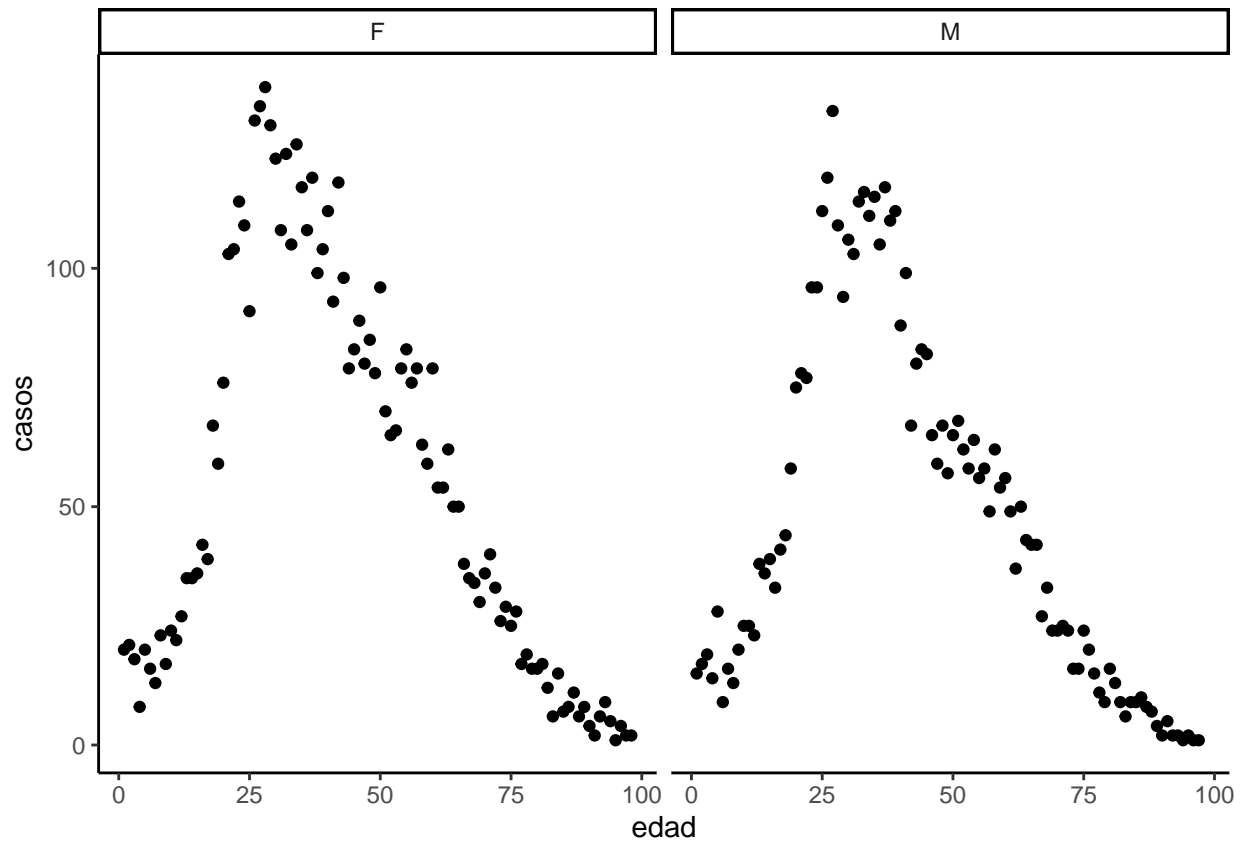
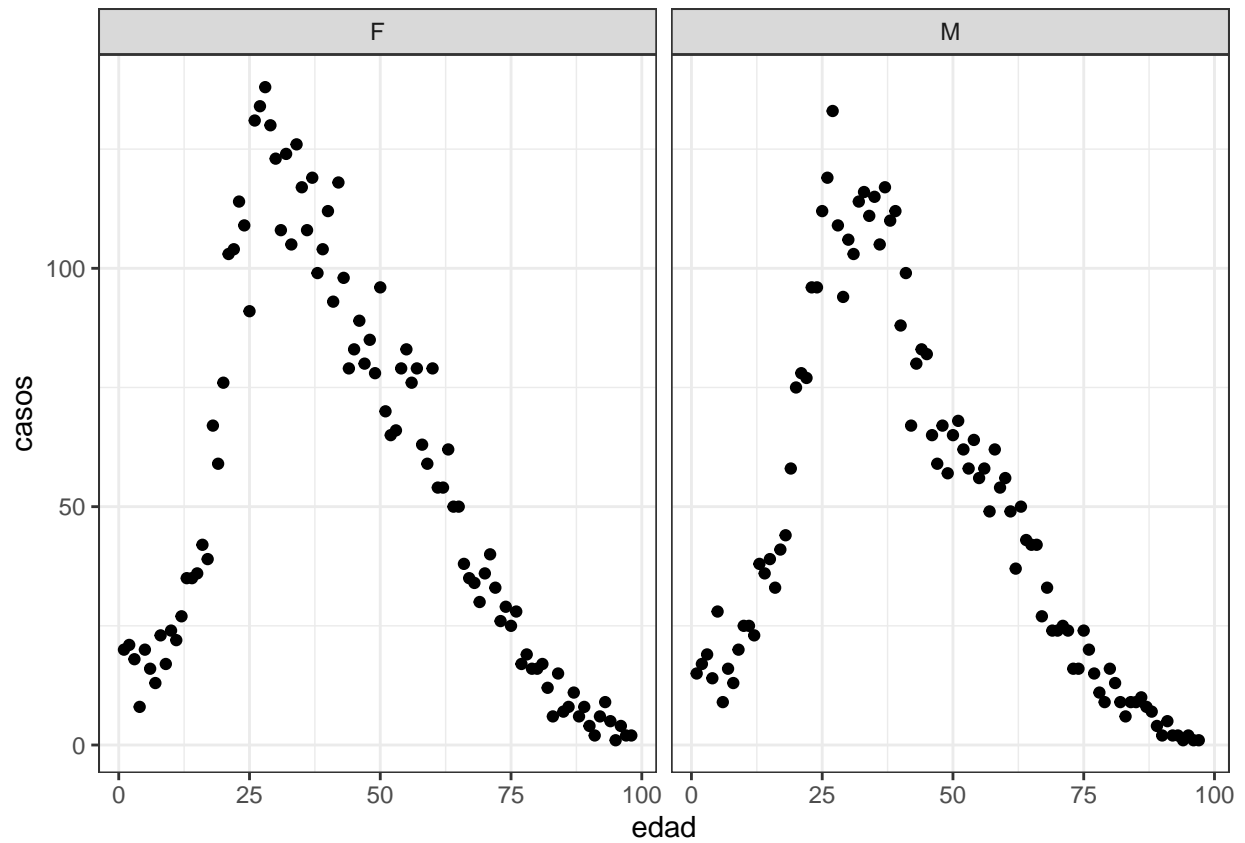


Gráfico de Puntos por Edad y Sexo con Tema Blanco y Negro

```
ggplot(data = COVID_SEX0, aes(x = edad, y = casos)) +  
  geom_point() +  
  facet_wrap(~sexo) +  
  theme_bw()
```



## Personalización de Títulos y Ejes

```
ggplot(data = COVID_SEX0, aes(x = edad, y = casos, colour = sexo)) +  
  geom_point() +  
  facet_wrap(~sexo) +  
  labs(  
    y = "Casos Diarios",  
    x = "Edad en años",  
    title = "Distribucion de casos de COVID19 en COLOMBIA"  
  )
```

Distribucion de casos de COVID19 en COLOMBIA

