ggplot2

willandru

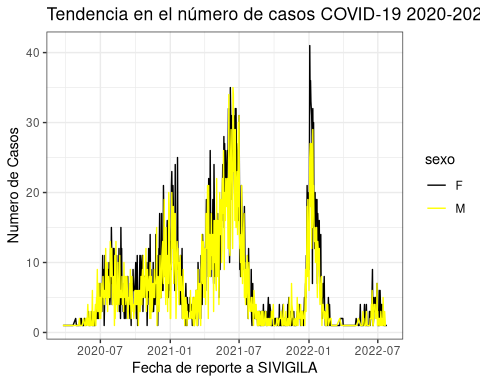
2023-12-03

## 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 × 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 <dttm> 2021-06-29, 2021-04-23, 2021-06-07, 2021…  
## $ id\_de\_caso <dbl> 4208058, 2737390, 3576919, 2944674, 23005…  
## $ fecha\_de\_notificacion <dttm> 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 <dttm> 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 <dttm> 2021-06-26, 2021-04-20, 2021-06-04, 2021…  
## $ fecha\_recuperacion <dttm> 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, 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, …

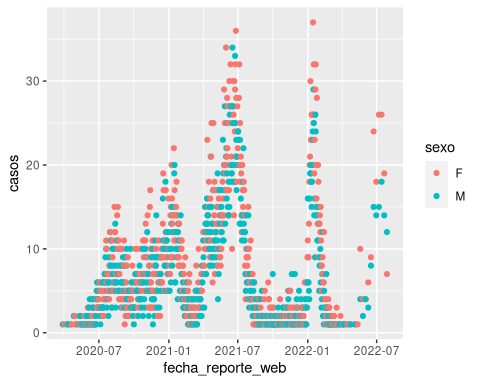
## Gráficos Exploratorios

### Gráfico de Puntos por Fecha y Sexo

COVID <- datos %>% group\_by(fecha\_reporte\_web, sexo) %>%  
 summarise(casos = n())

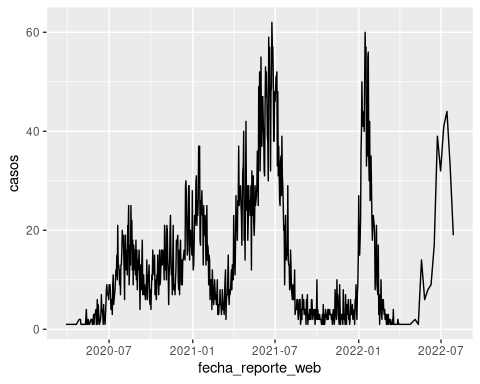
## `summarise()` has grouped output by 'fecha\_reporte\_web'. You can override using  
## the `.groups` argument.

ggplot(data = COVID, aes(x = fecha\_reporte\_web, y = casos, colour = sexo)) +  
 geom\_point()



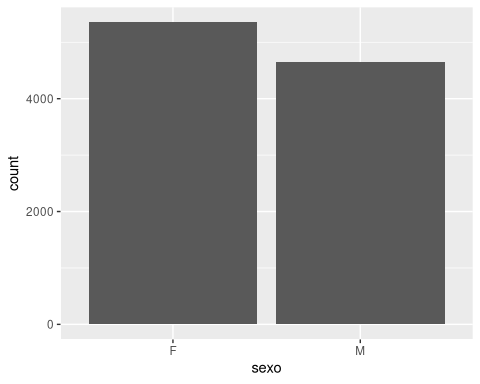
### Gráfico de Líneas por Fecha

COVID\_FECHA <- datos %>% group\_by(fecha\_reporte\_web) %>%   
 summarise(casos = n())  
  
ggplot(data = COVID\_FECHA, aes(x = fecha\_reporte\_web, y = casos)) +  
 geom\_line()



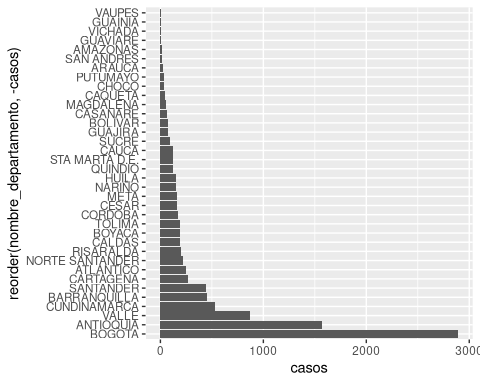
### Gráfico de Barras por Sexo

ggplot(data = datos) +   
 geom\_bar(aes(x = sexo))



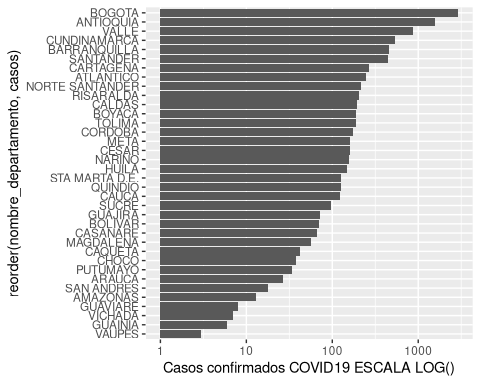
### Gráfico de Barras por Departamento

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 LOG()')



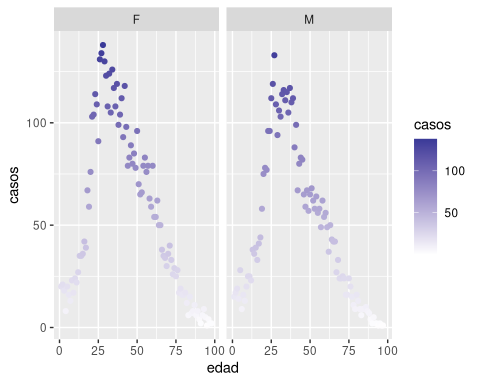
## Gráficos con Facetas y Temas

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

COVID\_SEXO <- datos %>% group\_by(edad, sexo) %>%  
 summarise(casos = n())

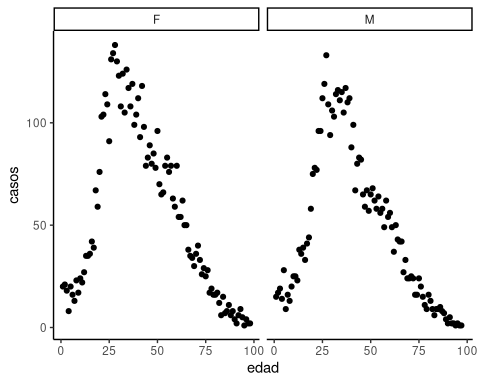
## `summarise()` has grouped output by 'edad'. You can override using the  
## `.groups` argument.

ggplot(COVID\_SEXO, aes(x = edad, y = casos, colour = casos)) +  
 geom\_point() +  
 facet\_wrap(~sexo) +  
 scale\_color\_gradient2()



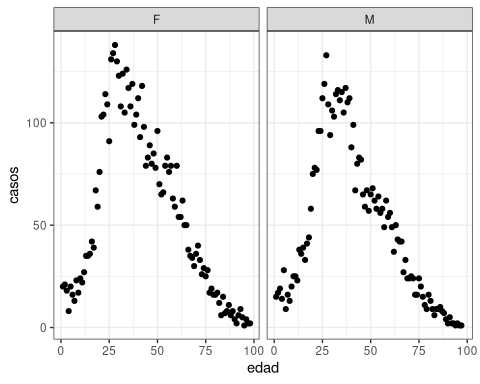
### Gráfico de Puntos por Edad y Sexo con Tema Clásico

ggplot(data = COVID\_SEXO, 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\_SEXO, aes(x = edad, y = casos)) +  
 geom\_point() +  
 facet\_wrap(~sexo) +  
 theme\_bw()



## Personalización de Títulos y Ejes

ggplot(data = COVID\_SEXO, 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"  
 )

