

Proyecto Bioestadística

Jorge Iván Aranda Caraballo

Utiliza los primeros 100,000 individuos de la base de datos de casos positivos por COVID-19 para realizar cada uno de los siguientes puntos:

```
library(dplyr)

library(readr)

library(ggplot2)

covid_19 <-
read_delim("C:/Users/jorge/Downloads/datos_abiertos_covid19/230228COVID1
9MEXICO.csv")

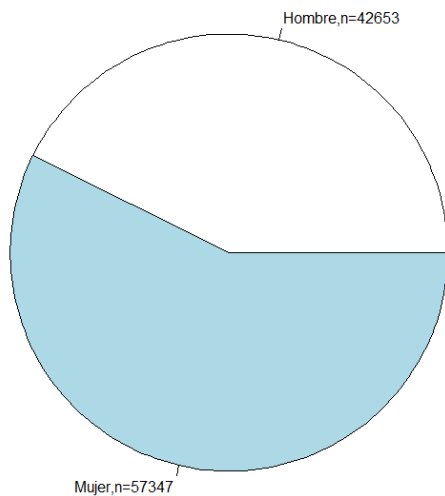
covid_19 <- covid_19 %>% filter(CLASIFICACION_FINAL %in% c(1,2,3))
covid_19 <- head(covid_19,n=100000)
```

- 1) Dibuja una gráfica de pastel correspondiente a la variable "SEXO". Reporta el tamaño de cada grupo en esta variable.

```
sex <- covid_19 %>% mutate(SEXO= case_when(
  .$SEXO==1 ~ "Mujer",
  .$SEXO==2 ~ "Hombre",
  .$SEXO==99 ~ "No Especificado")) %>% group_by(SEXO) %>%
summarise(Total=n())

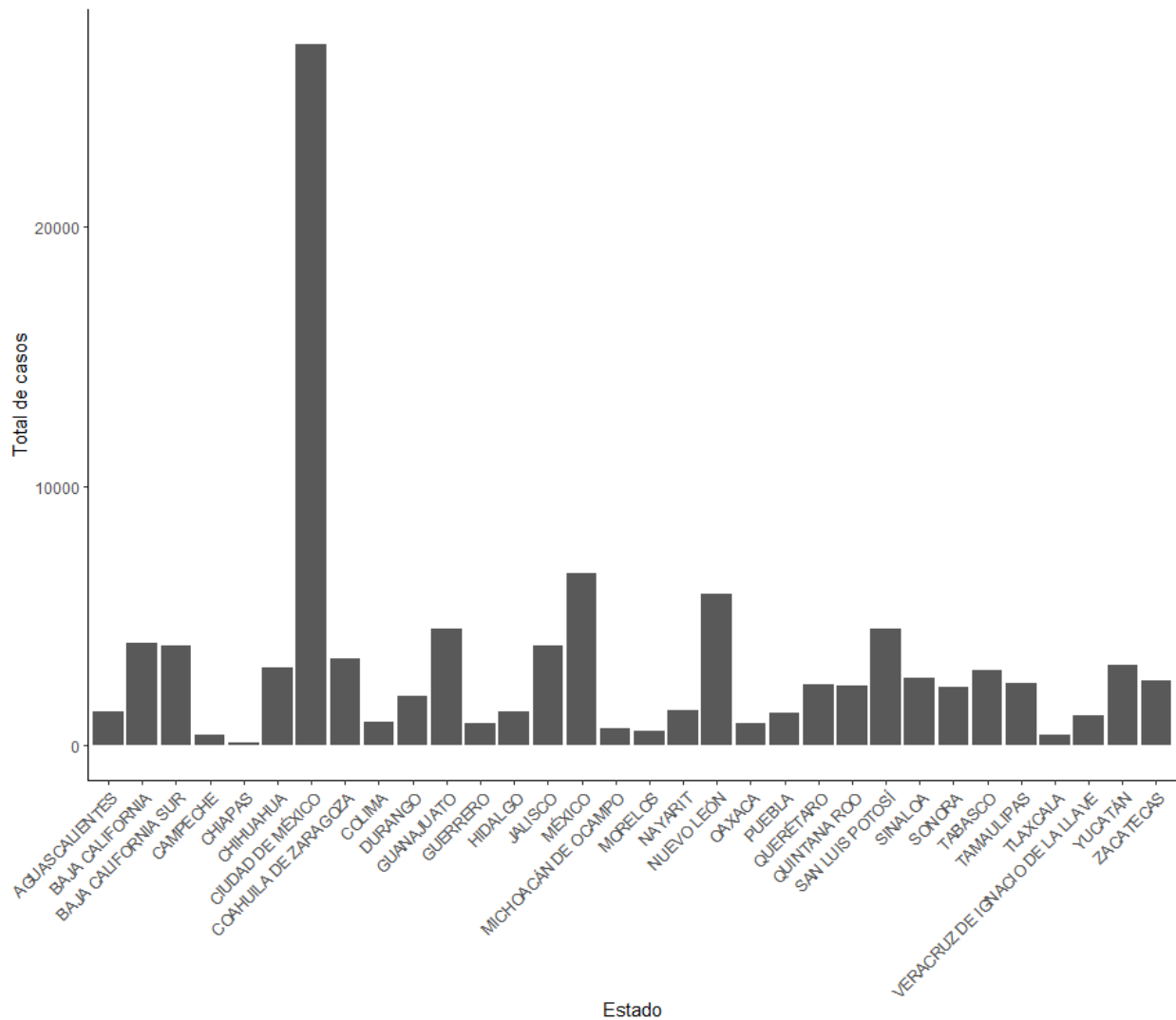
pie_labels=paste0(sex$SEXO, ",", "n=", sex$Total)

pie(sex$Total, labels = pie_labels)
```



- 2) En una gráfica de barras muestra el número de casos correspondiente a cada entidad (la variable "ENTIDAD_RES").

```
library(XLConnect)
Catalogos<-
loadWorkbook("C:/Users/jorge/Downloads/Catalogos_071020.xlsx")
Entidad_RED <- readWorksheet(Catalogos,sheet=9)
Entidad_RES <- as_tibble(Entidad_RED)
Entidades <- covid_19 %>%
  left_join(Entidad_RES,by=c('ENTIDAD_RES'='CLAVE_ENTIDAD'))
Entidades %>% ggplot(aes(ENTIDAD_FEDERATIVA)) +
  geom_bar() +
  theme_classic() +
  theme(axis.text.x = element_text(angle=90,hjust=1)) + ylab("Total de casos")
```

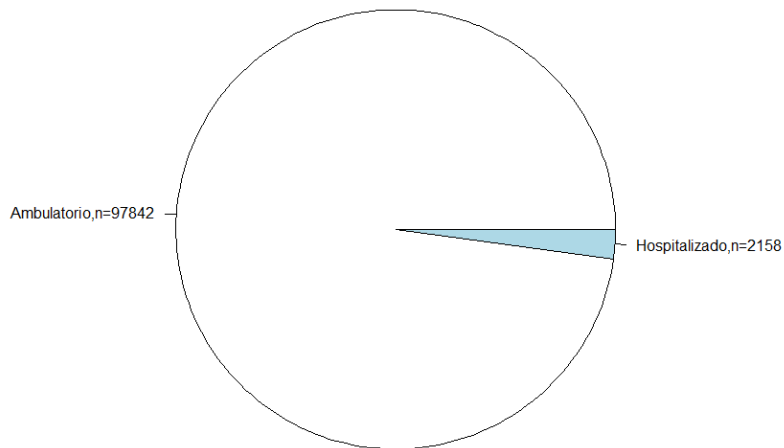


- 3) Dibuja otra gráfica de barras con el número de casos por entidad dividido por la población de la entidad correspondiente (investiga estas poblaciones).

Estado	Casos por cada 100 mil habitantes
ASACIENTES	2.2
BAJA CALIFORNIA	0.8
BAJA CALIFORNIA SUR	4.0
CAMPECHE	3.4
CHAPAS	0.6
CHIHUAHUA	0.8
CIUDAD DE MÉXICO	0.4
COAHUILA DE ZARAGOZA	1.0
COLIMA	4.4
DURANGO	1.7
GUANAJUATO	0.5
GUERRERO	0.9
HIDALGO	1.0
JALISCO	0.4
MÉXICO	0.2
MICHOACÁN DE OCAMPO	0.7
MORELOS	1.6
NAYARIT	2.6
NUEVO LEÓN	0.6
OAXACA	0.8
PUEBLA	0.5
QUERÉTARO	1.3
QUINTANA ROO	1.7
SAN LUIS POTOSÍ	1.1
SINALOA	1.0
SONORA	1.0
TABASCO	1.3
TAMAULIPAS	0.9
TLAXCALA	2.4
VERACRUZ DE LA Llave	0.4
YUCATÁN	1.4
ZACATECAS	2.0

```
paciente <- covid_19 %>% mutate(TIPO_PACIENTE=case_when(
  .$TIPO_PACIENTE==1 ~ "Ambulatorio",
  .$TIPO_PACIENTE==2 ~ "Hospitalizado",
  .$TIPO_PACIENTE==99 ~ "No especificado"
)) %>% group_by(TIPO_PACIENTE) %>% summarise(Total=n())

pie_labels<-paste0(paciente$TIPO_PACIENTE, ",", "", "", "n=", paciente$Total)
pie(paciente$Total, labels = pie_labels)
```



- 5) Dibuja una gráfica de pastel donde muestres los números de positivos por prueba de laboratorio y los positivos por prueba de antígeno.

```
library(tidyr)

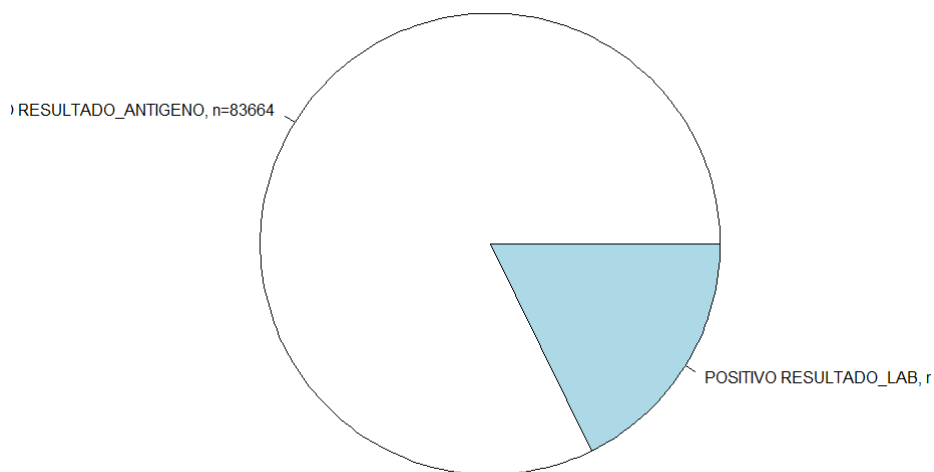
Resultados<- covid_19 %>%

mutate(RESULTADO_LAB= case_when(
  .$RESULTADO_LAB==1 ~ "Positivo",
  .$RESULTADO_LAB==2 ~ "Negativo",
  .$RESULTADO_LAB==3 ~ "Pendiente",
  .$RESULTADO_LAB==4 ~ "No adecuado",
  .$RESULTADO_LAB==97 ~ "Sin muestra"
), RESULTADO_ANTIGENO= case_when(
  .$RESULTADO_ANTIGENO==1 ~ "Positivo",
  .$RESULTADO_ANTIGENO==2 ~ "Negativo",
  .$RESULTADO_ANTIGENO==3 ~ "Pendiente",
  .$RESULTADO_ANTIGENO==4 ~ "No adecuado",
```

```

        .$RESULTADO_ANTIGENO==97 ~ "Sin muestra"
    ))
Resultados1 <- Resultados %>%
  select (RESULTADO_LAB, RESULTADO_ANTIGENO) %>%
  gather()
colnames(Resultados1) <- c('Prueba', 'Resultado')
Resultados1 <- Resultados1 %>%
  filter (Resultado=="Positivo") %>%
  group_by (Prueba) %>%
  summarize (Total=n())
pie_labels <- paste0("POSITIVO", " ",
                     Resultados1$Prueba, ", ", " ",
                     "n=", Resultados1$Total)
pie (Resultados1$Total, labels = pie_labels)

```

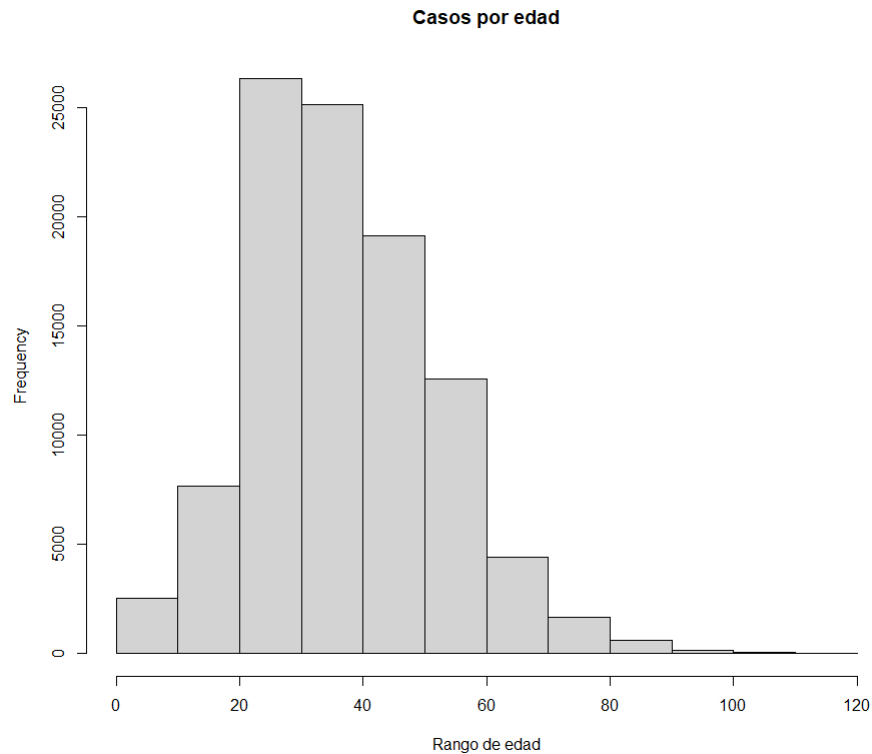


- 6) Dibuja un histograma de la variable "EDAD", con los extremos de los intervalos en 0, 10, 20, etc, años.

```

hist(covid_19$EDAD, breaks=11,
     main=c("Casos por edad"),
     xlab=c("Rango de edad"))

```



- 7) Para cada grupo de edad del ejercicio (5), divide su frecuencia (calculada en (5)) entre el tamaño de la población en ese grupo de edad (investiga estas poblaciones). Dibuja una gráfica de barras con los valores encontrados.

```
Covid_19_edades <- covid_19 %>% mutate(EDAD= case_when(
  .$EDAD<10 ~ "0 a 9 años",
  .$EDAD>=10 & .$EDAD<20 ~ "10 a 19 años",
  .$EDAD>=20 & .$EDAD<30 ~ "20 a 29 años",
  .$EDAD>=30 & .$EDAD<40 ~ "30 a 39 años",
  .$EDAD>=40 & .$EDAD<50 ~ "40 a 49 años",
  .$EDAD>=50 & .$EDAD<60 ~ "50 a 59 años",
  .$EDAD>=60 & .$EDAD<70 ~ "60 a 69 años",
  .$EDAD>=70 & .$EDAD<80 ~ "70 a 79 años",
  .$EDAD>=80 & .$EDAD<90 ~ "80 a 89 años",
  .$EDAD>=90 & .$EDAD<100 ~ "90 a 99 años",
  .$EDAD>100 ~ "mas de 100 años",
  !is.na(.$EDAD) ~ "No Especificado"
)) %>% mutate(EDAD=as.factor(EDAD))

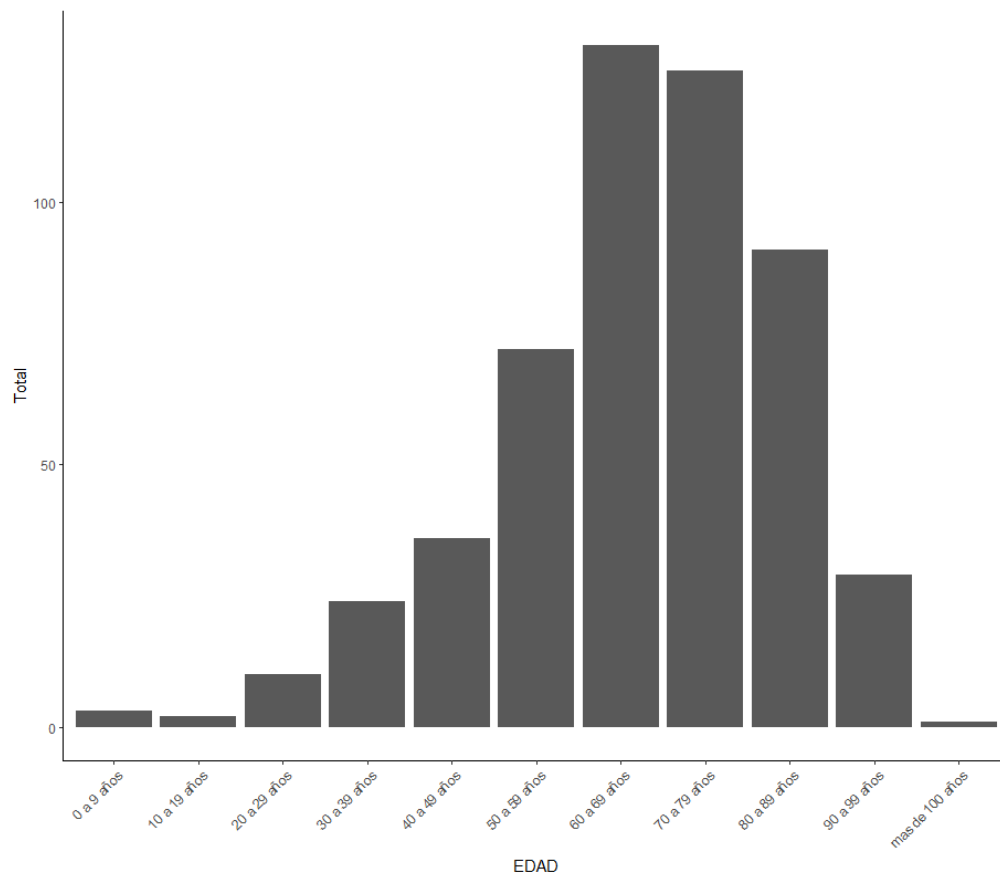
Pob_edades<- loadWorkbook("C:/Users/jorge/Downloads/Total_edades2.xlsx")
Pob_edades <- readWorksheet(Pob_edades,sheet=1)
casos_por_edad <- Covid_19_edades %>% group_by(EDAD) %>%
summarize(Total=n()) %>%
  inner_join(Pob_edades,by='EDAD',suffix=c("_casos","_poblacion")) %>%
  mutate(casos_x10000=Total_casos*10000/Total_poblacion)
```

A bar chart titled 'Caso de COVID-19 por edad' showing the incidence of COVID-19 cases per 10,000 inhabitants across different age groups in Mexico. The x-axis is labeled 'EDAD' and lists age ranges from '0 a 9 años' to 'mas de 100 años'. The y-axis is labeled 'Casos por cada 10mil habitantes' and ranges from 0 to 15. The bars show a peak in the 30-39 age group, with a secondary peak in the 20-29 group. The incidence decreases for older age groups, with a slight increase for those aged 100 and over.

EDAD	Casos por cada 10mil habitantes
0 a 9 años	1.0
10 a 19 años	3.0
20 a 29 años	13.0
30 a 39 años	14.5
40 a 49 años	12.5
50 a 59 años	10.5
60 a 69 años	5.5
70 a 79 años	4.0
80 a 89 años	3.5
90 a 99 años	3.2
mas de 100 años	6.8

- ```
muerteres_por_grupo<- Covid_19_edades %>% count(EDAD,FECHA_DEF) %>%
 filter(!is.na(FECHA_DEF)) %>%
 group by(EDAD) %>% summarise(Total=sum(n))
```

```
muerteres_por_grupo %>% ggplot(aes(EDAD, Total)) +
geom_bar(stat='identity') +
theme_classic() + theme(axis.text.x = element_text(hjust=1,
angle=45))
```



- 9) Divide las frecuencias encontradas en el ejercicio (7) entre las frecuencias correspondientes encontradas en el ejercicio (5). Con esto encontrarás las tasas de letalidad de la enfermedad. Dibuja una gráfica de barras con tus resultados.

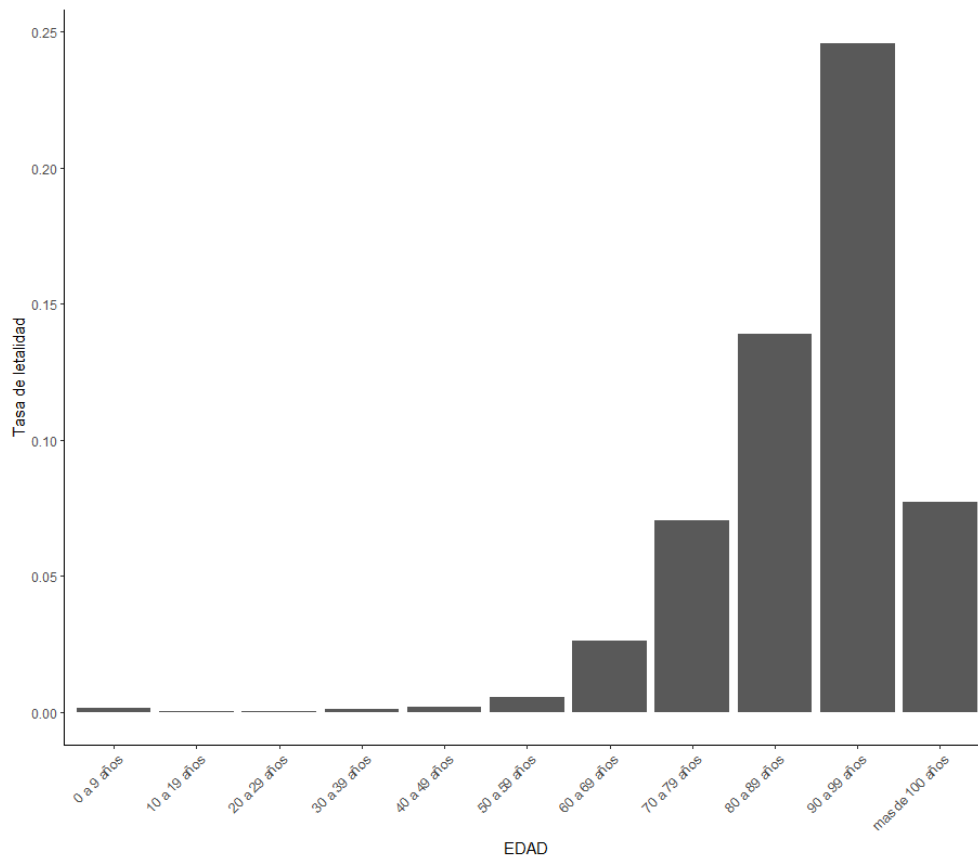
```

Letalidad <- casos_por_edad %>% select(EDAD, Total_casos) %>%
 inner_join(muertes_por_grupo, by='EDAD') %>% transmute(
 EDAD, Total_casos, Total_muertes=Total, indice_letalidad=
 Total/Total_casos
)

Letalidad %>% ggplot(aes(EDAD, indice_letalidad)) +
 geom_col() + theme_classic() +
 theme(axis.text.x = element_text(hjust=1, angle=45)) +
 ylab("Tasa de letalidad")

```





- 10) Divide las frecuencias encontradas en el ejercicio (7) entre el tamaño de la población en el grupo de edad correspondiente (como en el ejercicio (6)). Con esto encontrarás las tasas de mortalidad de la enfermedad. Dibuja una gráfica de barras con tus resultados.

```
Mortalidad <- muertes_por_grupo %>%
 inner_join(Pob_edades, by='EDAD', suffix=c(
 "_muertes", "_poblacion")) %>% mutate(Tasa_mortalidad=
 Total_muertes*10000/Total_poblacion)

Mortalidad %>% ggplot(aes(EDAD, Tasa_mortalidad)) + geom_col() +
 theme_classic() + theme(axis.text.x = element_text(hjust=1, angle=45))
+
 ylab("Muertes por cada 10mil habitantes")
```

