

Tarea 3

Dalia Camacho

September 1, 2018

En la carpeta de arriba encontrarás un archivo de excel (m_013.xls), este archivo contiene información de causas de mortalidad en México entre 2000 y 2008. Contesta las siguientes preguntas:

```
library(readxl)
library(tidyr)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(ggplot2)
library(gridExtra)

##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##   combine
library(grid)

Mortalidad <- read_xls("m_013.xls")

#View(Mortalidad)
```

¿Cuáles son las variables en esta base de datos?

Las variables son: entidad federativa, sexo, tasa de mortalidad por 100,000 habitantes para enfermedades transmisibles, nutricionales y de la reproducción, enfermedades no transmisibles y lesiones.

¿La tabla de datos cumple con los principios de datos limpios? ¿Qué problemas presenta?

La tabla no cumple con los principios de los datos limpios, porque distintas columnas tienen la misma variable divididas por sexo. Mientras que la variable sexo no se encuentra definida en una columna.

La información del archivo de excel se ha guardado también en archivos de texto (csv) 2001-2008, lee y limpia los datos para que cumplan los principios de datos limpios. Recuerda que las modificaciones deben de ser reproducibles, para esto guarda tu trabajo en un script.

```
#Leer bases
files <- list.files( pattern = "\\\\.csv")
Bases <- read.csv(files[1], stringsAsFactors = FALSE)
Bases$year <- as.numeric( substring(files[1], 1,4))
for(i in 2:length(files)){
  aux      <- read.csv(files[i])
  aux$year <- as.numeric( substring(files[i], 1,4))
  Bases    <- rbind(Bases, aux)
}

Trans  <- Bases %>% gather( "sex", "trans", starts_with("trans")) %>%
  select(sex, year, trans, edo) %>%
  mutate(sex=substring(sex, 7, 100)) %>%
  filter(sex!="Total")

noTrans <- Bases %>% gather( "sex", "notrans", contains("noTrans")) %>%
  select(sex, year, notrans, edo)%>%
  mutate(sex=substring(sex, 9, 100)) %>%
  filter(sex!="Total")

Lesiones <- Bases %>% gather( "sex", "lesiones", contains("lesiones")) %>%
  select(sex, year, lesiones, edo)%>%
  mutate(sex=substring(sex, 10, 100)) %>%
  filter(sex!="Total")

NewBase <- Trans %>% inner_join(noTrans) %>%
  inner_join(Lesiones, by = c("sex", "year", "edo"))

## Joining, by = c("sex", "year", "edo")
head(NewBase)
```

```
##      sex year trans      edo notrans lesiones
## 1 Mujeres 2000  54.8    Aguascalientes  416.8    25.0
## 2 Mujeres 2000  63.9    Baja California  503.8    29.8
## 3 Mujeres 2000  61.5 Baja California Sur  409.9    23.7
## 4 Mujeres 2000  44.9      Campeche    391.5    19.5
## 5 Mujeres 2000  51.7      Coahuila    500.6    18.9
## 6 Mujeres 2000  51.0      Colima     436.4    21.9
```

El archivo de excel indice_marginacion.xlsx contiene el índice por entidad para los años 2000 y 2010. Realiza una gráfica donde compares la marginación por entidad con las tasas de mortalidad correspondientes al 2000. Deberás unir las dos fuentes de información.

```
Marginacion <- read_excel("indice_marginacion.xlsx", range = "A4:F36",
  col_names = TRUE) %>%
  mutate(edo=`Entidad federativa`) %>%
  mutate(ind_marg = `Índice absoluto de marginación 2000`) %>%
  mutate(lugar_marg=`Lugar que ocupa en el contexto nacional`) %>%
  select(c(edo,ind_marg, lugar_marg))
```

```

Marg_mort <- NewBase %>% filter(year==2000) %>%
  left_join(Marginacion, by=c("edo"))

g1 <- ggplot(Marg_mort)+theme_bw()+
  geom_point(aes(ind_marg, lesiones, col=sex))+
  geom_smooth(aes(ind_marg, lesiones, group=(sex), col=sex), se=FALSE, method = "loess")+
  xlab("Índice de marginación")+
  ylab("")+
  ggtitle("Lesiones")+
  theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5),
        legend.position = "none")

g2 <- ggplot(Marg_mort)+theme_bw()+
  geom_point(aes(ind_marg, trans, col=sex))+
  geom_smooth(aes(ind_marg, trans, group=(sex), col=sex), se=FALSE, method = "loess")+
  ylab("")+
  xlab("Índice de marginación")+
  ggtitle("Enfermedades transmisibles")+
  theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5),
        legend.position = "none")

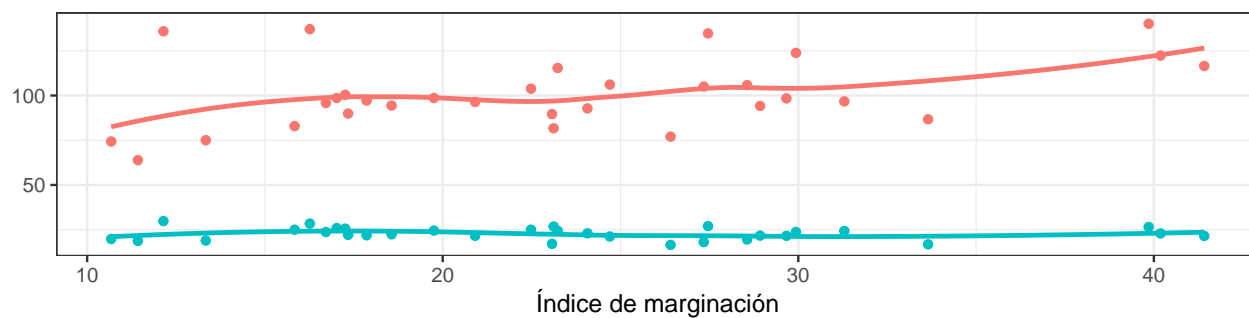
g3 <- ggplot(Marg_mort)+theme_bw()+
  geom_point(aes(ind_marg, notrans, col=sex))+
  geom_smooth(aes(ind_marg, notrans, group=(sex), col=sex), se=FALSE, method = "loess")+
  xlab("Índice de marginación")+
  ylab("")+
  ggtitle("Enfermedades no transmisibles")+
  #ggtitle("Tasa de muertes por sexo y nivel de marginación")+
  theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5),
        legend.position = "bottom")

grid.arrange(g1,g2,g3, top= textGrob("Tasa de muertes por sexo y nivel de marginación", gp=gpar(fontsize=12,
  heights=c(12,12,17))

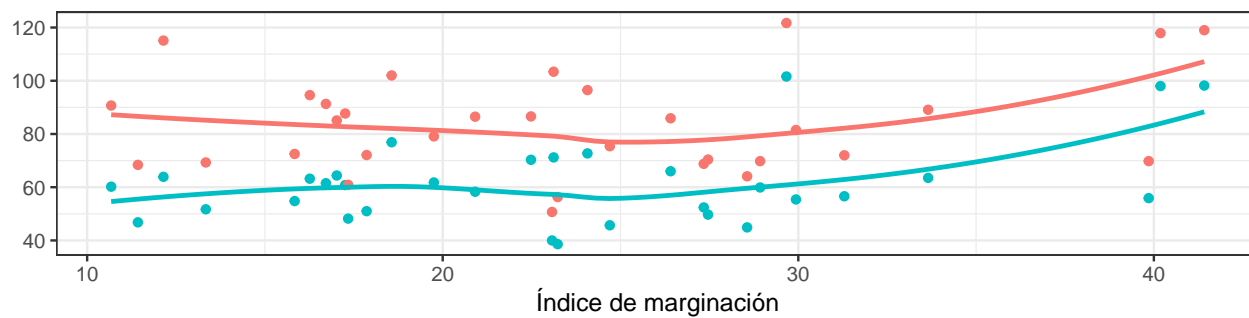
```

Tasa de muertes por sexo y nivel de marginación

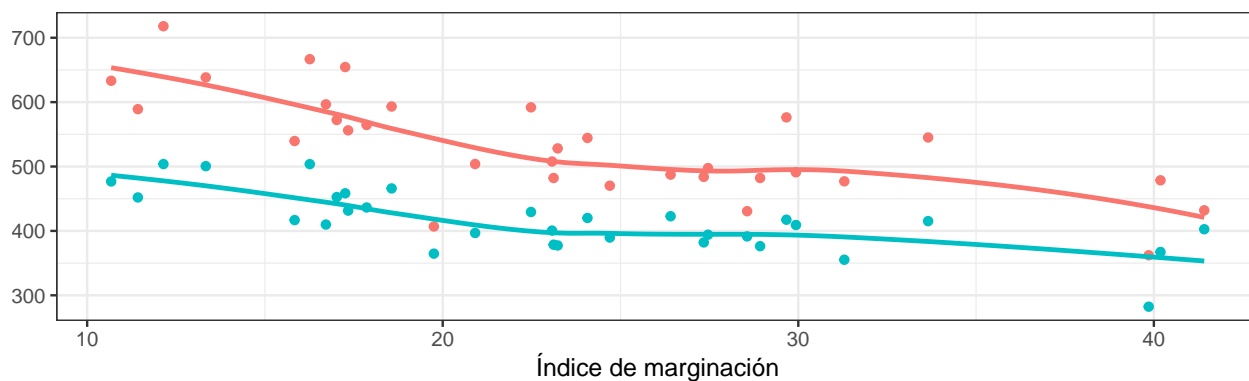
Lesiones



Enfermedades transmisibles



Enfermedades no transmisibles



—●— Hombres —●— Mujeres