

Solución ejercicios - Sesión 5

Fundación Paz Ciudadana

Ignacio Agloni, Verónica Canales, Klaus Lehmann

Ejercicio 1

Problema

Utilizando la base de datos de la **ENUSC 2017**, grafica la victimización personal por tramos de edad. Utiliza como referencia los siguientes tramos:

- 15 a 19 años
- 20 a 35 años
- 36 a 45 años
- 46 a 55 años
- 56 a 65 años
- 66 años y más

Utiliza el tipo de gráfico que creas sea el más adecuado para mostrar estos datos.

- Considera solo los datos a nivel muestral y descriptivo.
- Recuerda que debes filtrar tus datos por el informante Kish.

Solución

Carguemos primero los paquetes y datos a nuestra sesión.

```
library(haven)
library(tidyverse)

# Puedo utilizar el mismo código que se usó en el taller
enusc <- read_sav("data/Base de Datos - XIV ENUSC 2017.sav") %>%
  filter(Kish == 1) %>%
  select(enc_region, rph_sexo, VP_DC, VA_DC, rph_edad, P13_1_1) %>%
  mutate(sexo = haven::as_factor(rph_sexo),
         region = haven::as_factor(enc_region),
         vic_hog = haven::as_factor(VA_DC),
         edad = as.numeric(rph_edad))
```

Ahora es necesario crear la variable tramos de edad, la cual debiese ser una variable de tipo factor. Hay varias maneras de hacer esto. Una de estas es:

```
# Primero agrego la variable a la base
# Ahora no es más que la variable edad con otro nombre
enusc$edad_tramos <- enusc$edad

# El código especial "no responde" de esta variable es 999
sum(if_else(enusc$edad == 999, 1, 0))
```

```
## [1] 0
```

```
# Reviso que no existan NA
sum(is.na(enusc$edad))
```

```
## [1] 0
```

```

# Asigno un nuevo valor dependiendo de los valores de edad
# Sé que no hay valores perdidos
enusc$edad_tramos[enusc$edad >= 15 & enusc$edad <= 19] <- 1
enusc$edad_tramos[enusc$edad >= 20 & enusc$edad <= 35] <- 2
enusc$edad_tramos[enusc$edad >= 36 & enusc$edad <= 45] <- 3
enusc$edad_tramos[enusc$edad >= 46 & enusc$edad <= 55] <- 4
enusc$edad_tramos[enusc$edad >= 56 & enusc$edad <= 65] <- 5
enusc$edad_tramos[enusc$edad >= 66 & enusc$edad <= max(enusc$edad)] <- 6

# Ahora cambio la variable a tipo factor
enusc$edad_tramos <- factor(enusc$edad_tramos, labels = c("15 a 19 años",
  "20 a 35 años", "36 a 45 años", "46 a 55 años",
  "56 a 65 años", "66 años y más"))

table(enusc$edad_tramos)

##
## 15 a 19 años 20 a 35 años 36 a 45 años 46 a 55 años 56 a 65 años
##      1914      7520      4520      4552      4190
## 66 años y más
##      4920

# ¿Se les ocurre como hacer esto con if_else?

```

Listo, tenemos nuestra variable de edad en tramos. La variable victimización personal actualmente es `haven_labelled`. Dejémosla como tipo factor.

```

enusc$vic_per <- haven::as_factor(enusc$VP_DC)
table(enusc$vic_per)

##
## NO SI
## 24951 2665

```

No me gustan los nombres con mayúsculas y quiero que la categorías “Sí” esté antes que “No”. Haré estos cambios con el paquete `forcats`.

```

enusc$vic_per <- enusc$vic_per %>%
  fct_recode("Sí" = "SI", "No" = "NO") %>%
  fct_relevel("Sí", "No")
table(enusc$vic_per)

##
## Sí No
## 2665 24951

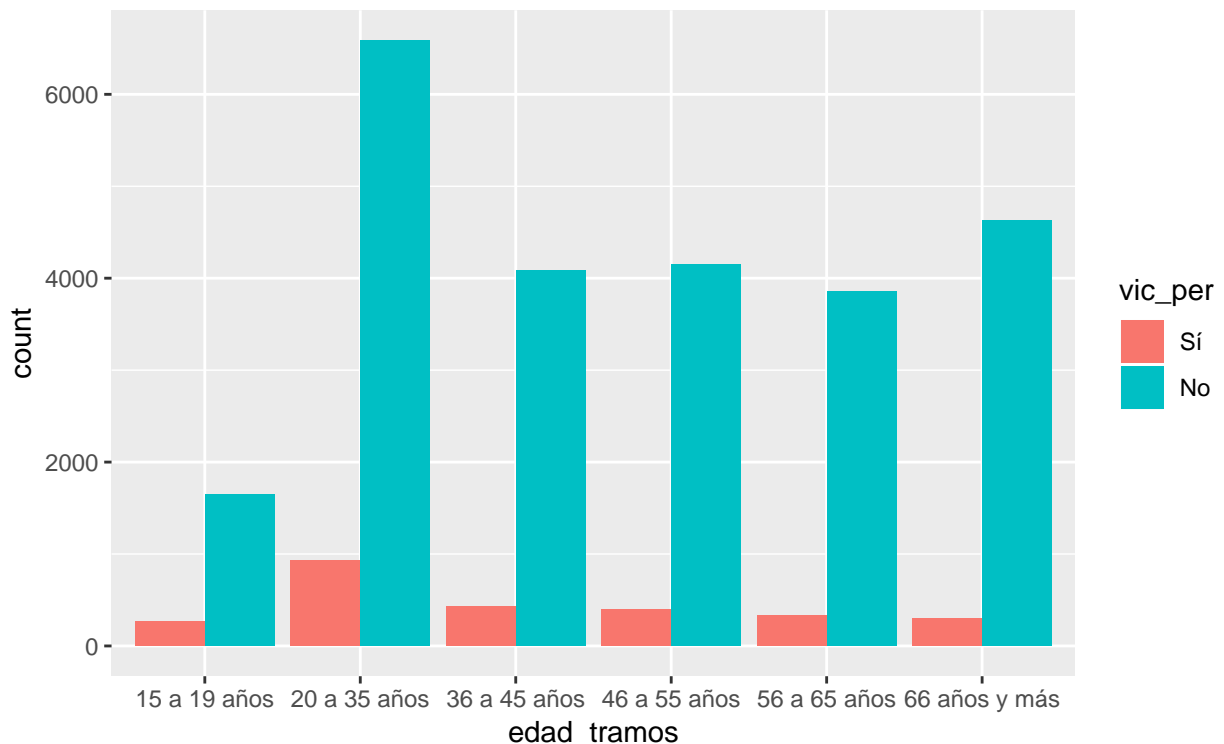
```

Ahora veamos como usar estas variable para visualizar nuestros datos. Probemos con un gráfico de barras.

```

ggplot(data = enusc, aes(x = edad_tramos, fill = vic_per)) +
  geom_bar(position = "dodge")

```



Lo anterior resuelve el ejercicio, pero mejor hagamos una tabla para graficar la frecuencia relativa de la victimización personal por tramos de edad.

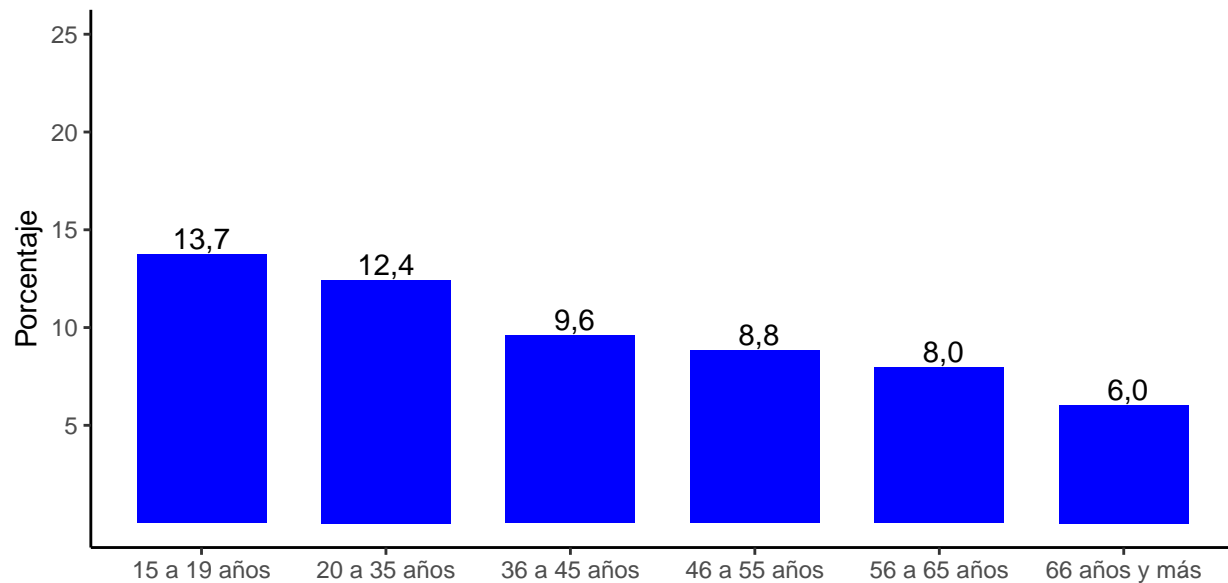
```
# Hagamos una tabla que calcule el % de victimizados
# por tramo de edad
tabla_vic <- enusc %>%
  group_by(edad_tramos, vic_per) %>%
  summarise(total_edad = n()) %>% # calcula el denominador para c/grupo
  mutate(vic_tramo = 100 * total_edad / sum(total_edad)) # calcula el %
```

Ahora usemos esta información para hacer nuestro gráfico de barras. Además, como la victimización es mucho más baja que la no victimización, grafiquemos solo esa categoría para poder apreciar las diferencias entre los tramos de edad.

```
tabla_vic %>% filter(vic_per == "Sí") %>%
ggplot(aes(x = edad_tramos, y = vic_tramo)) +
  geom_bar(stat = "identity", fill = "blue", width = 0.7) +
  scale_y_continuous(name = "Porcentaje", breaks = seq(5,25,5), limits = c(0,25)) +
  scale_x_discrete(name = NULL) +
  # en la siguiente línea se agrega el valor sobre las barras
  geom_text(aes(label = format(round(vic_tramo, 1), decimal.mark = ",")), hjust = 0.5, vjust = -0.25) +
  ggtitle("Victimización personal por tramos de edad",
    subtitle = "Delitos de robo con violencia, robo por sorpresa, hurto
    y lesiones (valores muestrales)") +
  labs(caption = "Fuente: Elaboración propia en base a ENUSC 2017 - INE") +
  theme_classic() +
  # En las siguientes líneas se ajustan tamaño y posición título, caption, etc
  theme(plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 11, hjust = 0.5),
    plot.caption = element_text(hjust = 0.5, size = 10))
```

Victimización personal por tramos de edad

Delitos de robo con violencia, robo por sorpresa, hurto y lesiones (valores muestrales)



Fuente: Elaboración propia en base a ENUSC 2017 – INE

Ejercicio 2

Problema

Compara en un solo gráfico el porcentaje de personas víctimas de un delito (victimización personal) por región y el porcentaje de personas que creen serán nuevamente víctimas de un delito en los próximos 12 meses.

- Considera solo los datos a nivel muestral y descriptivo.
- Recuerda que debes filtrar tus datos por el informante Kish.
- La pregunta P13_1_1 tiene valores NA.

Puedes usar si quieres el siguiente código para partir

```
# Convertir p13 en factor y VP_DC en factor
enusc$p13 <- as_factor(enusc$P13_1_1)
enusc$vic_per <- as_factor(enusc$VP_DC)
# Dejar como NA categorías distintas a Si y No
enusc$p13[enusc$p13 != "Si" & enusc$p13 != "No"] <- NA
# Borrar niveles que ahora son NA con fct_drop de forcats
enusc$p13 <- fct_drop(enusc$p13, c("No sabe", "Sin dato",
                                   "No responde"))

# Calcular percepción P13 por región
enusc %>% filter(!is.na(p13)) %>%
  group_by(region, p13) %>%
  summarise(total = n()) %>%
  mutate(perc_vic = 100 * (total / sum(total)))
```

Solución

Al igual que en el ejercicio anterior construiremos la tabla desde la cual haremos nuestro gráfico. Partiremos con el código que dejamos en el enunciado.

```
# Convertir p13 en factor y VP_DC en factor
enusc$p13 <- as_factor(enusc$P13_1_1)
enusc$vic_per <- as_factor(enusc$VP_DC)
# Reacomodamos las categorías de victimización personal
enusc$vic_per <- enusc$vic_per %>%
  fct_recode("Sí" = "SI", "No" = "NO") %>%
  fct_relevel("Sí", "No")
# Dejar como NA categorías distintas a Si y No en p13
enusc$p13[enusc$p13 != "Si" & enusc$p13 != "No"] <- NA
# Borrar niveles que ahora son NA con fct_drop de forcats
enusc$p13 <- fct_drop(enusc$p13, c("No sabe", "Sin dato",
                                   "No responde")) %>%
  fct_recode("Sí" = "Si")

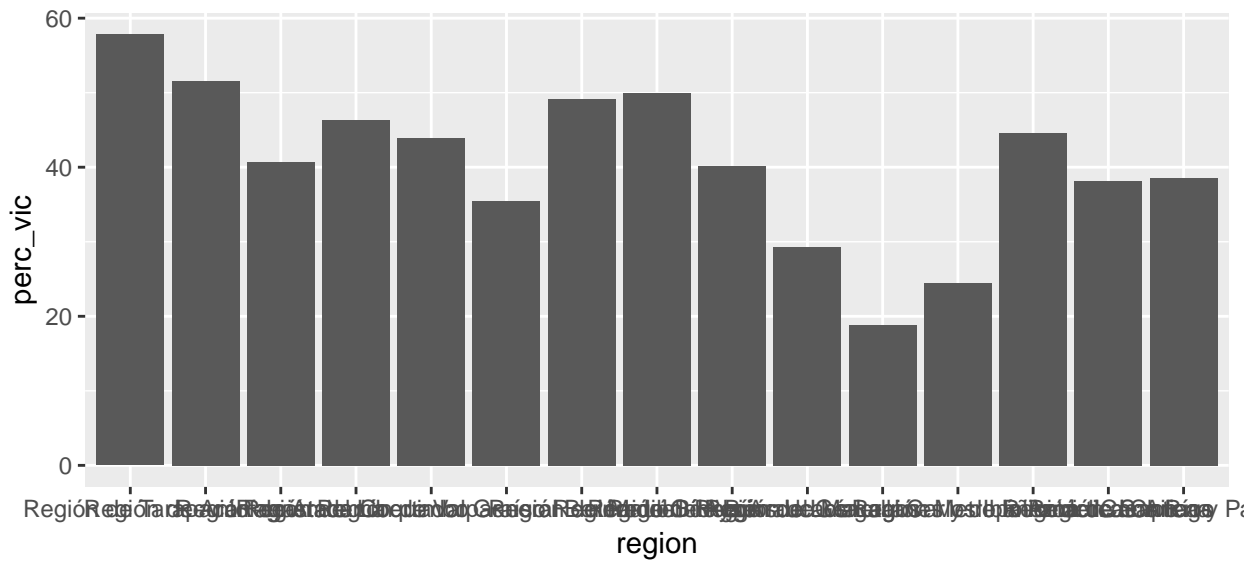
# Calcular percepción P13 por región
tabla_p13 <- enusc %>% filter(!is.na(p13)) %>% # deja fuera los valores NA
  group_by(region, p13) %>%
  summarise(total = n()) %>%
  mutate(perc_vic = 100 * (total/ sum(total))) %>%
  select(-total) # la borramos porque no la necesitamos

# Calcular victimizacion por región
tabla_vicper <- enusc %>% filter(!is.na(vic_per)) %>%
  group_by(region, vic_per) %>%
  summarise(total = n()) %>%
  mutate(vic = 100 * (total/ sum(total))) %>%
  select(-total)

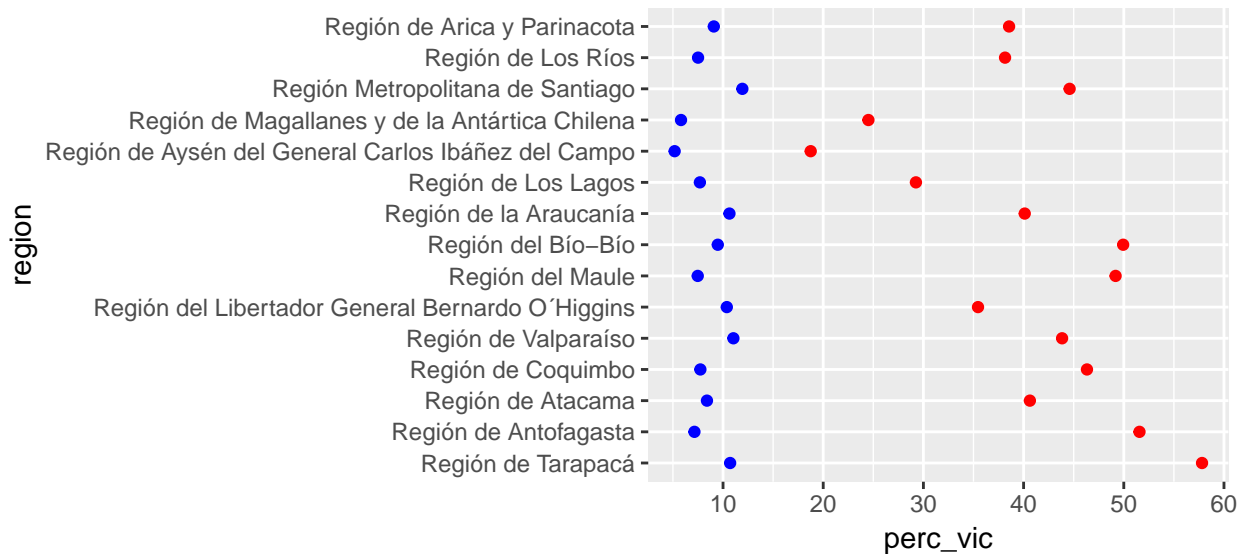
# Ahora juntamos ambas tablas
tabla_ej2 <- bind_cols(tabla_p13, tabla_vicper) %>%
  select(-region1)
```

Veamos si esta tabla nos sirve para crear un gráfico que contenga ambos datos. Usemos `geom_bar` y `geom_point`.

```
tabla_ej2 %>% filter(p13 == "Sí") %>%
  ggplot() +
  geom_bar(aes(x = region, y = perc_vic), stat = "identity") +
  geom_bar(aes(x = region, y = vic), stat = "identity")
```



```
tabla_ej2 %>% filter(p13 == "Sí") %>%
  ggplot() +
  geom_point(aes(y = region, x = perc_vic), colour = "red") +
  geom_point(aes(y = region, x = vic), colour = "blue")
```

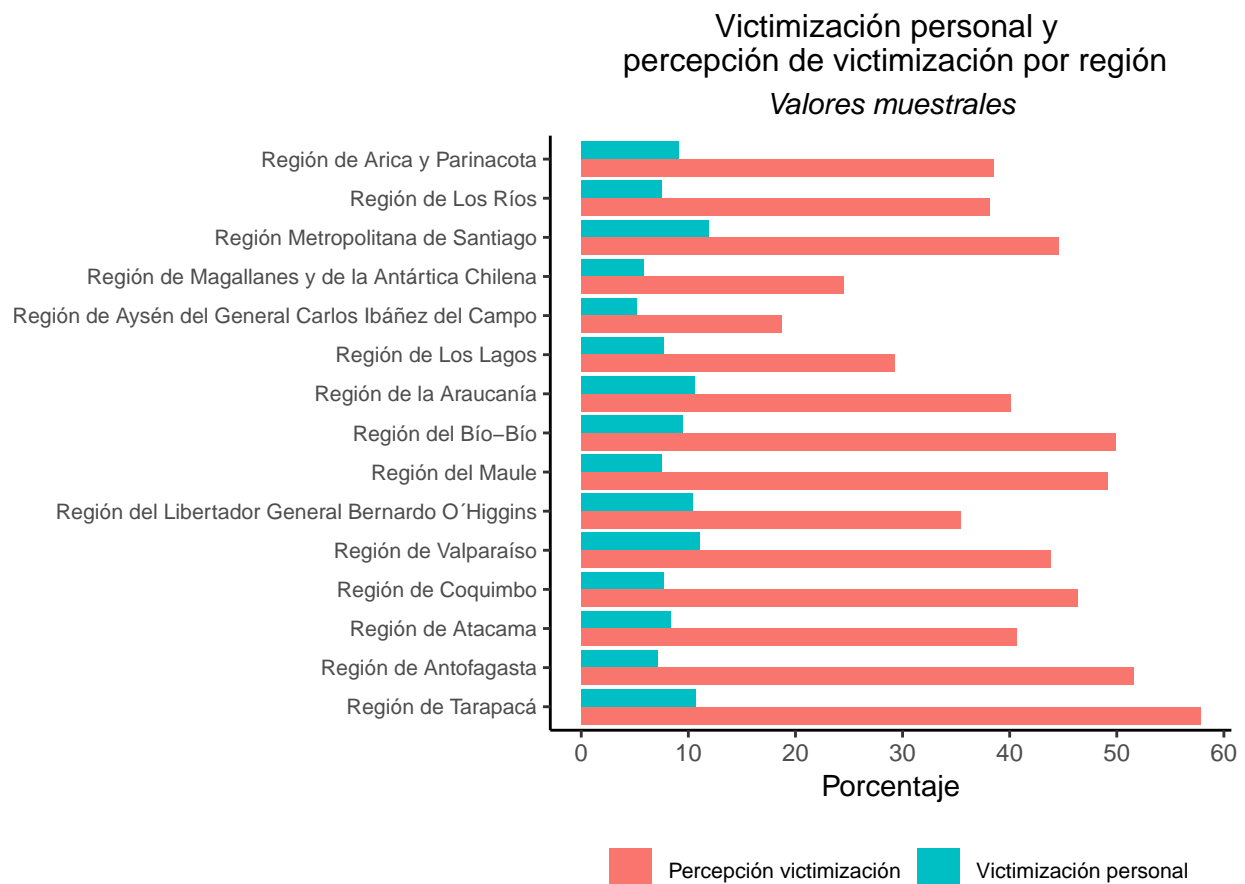


Parece que esto no salió tan bien. Por esta vez quiero hacer un gráfico de barra. La estructura actual de la tabla no me permite usarla para hacer un gráfico de barras. Si en vez de dos columnas separadas, la información del porcentaje de victimización y de percepción de ser víctima estuviese en una sola columna, sería más sencillo utilizar un `aes` como `fill` para diferenciar entre el indicador de victimización personal y la pregunta p13. Para reunir esta información en una sola columna podríamos usar `gather`.

```
tabla_ej2_gather <- tabla_ej2 %>%
  filter(p13 == "Sí") %>%
  select(-p13, -vic_per) %>%
  gather(perc_vic, vic, key = "indicador", value = "porcentaje")
```

Ahora sí podemos usar esta tabla para construir el gráfico de barras.

```
ggplot(data = tabla_ej2_gather) +
  geom_bar(aes(x = region, y = porcentaje, fill = indicador), stat = "identity",
    position = "dodge") +
  # \n permite el salto de línea. Vean que ocurre si lo quitan
  ggtitle("Victimización personal y \n percepción de victimización por región",
    subtitle = "Valores muestrales") +
  labs(caption = "Fuente: Elaboración propia en base a ENUSC 2017 - INE") +
  theme_classic() +
  scale_y_continuous(name = "Porcentaje", breaks = seq(0,60,10)) +
  scale_x_discrete(name = NULL) +
  scale_fill_discrete(name = NULL, labels = c("Percepción victimización", "Victimización personal")) +
  theme(plot.title = element_text(size = 12, hjust = 0.5),
    plot.subtitle = element_text(size = 11, hjust = 0.5, face = "italic"),
    plot.caption = element_text(hjust = 0.5, size = 8),
    legend.position = "bottom",
    legend.text = element_text(size = 8),
    axis.text.y = element_text(size = 8)) +
  coord_flip() # con esto giramos el gráfico
```



Fuente: Elaboración propia en base a ENUSC 2017 – INE

Si bien hay otros arreglos que podríamos hacer en este gráfico (como reordenar las regiones, acortar sus nombres porque algunos son muy largos, agregar el porcentaje a cada barra, etc.), dejaré este gráfico hasta aquí por si Uds. quieren seguir desarrollándolo.