

PantaleonData: Tabulación versus Graficación

Juan C. Correa

3/6/2021

Introducción

En ocasiones previas hemos visto algunas recomendaciones sobre cómo documentar nuestros análisis de resultados estadísticos. Ahora vamos a comprender comparativamente el rol de la tabulación y la graficación. Como de costumbre, empecemos por abrir nuestros datos de ejemplo, siguiendo las recomendaciones que hemos aprendido.

```
setwd("~/Documents/GitHub/Pantaleon")
library(readxl)
PantaleonData <- read_excel("data 19-20.xlsx")
```

```
## New names:
## * TOTAL -> TOTAL...23
## * CONTAR -> CONTAR...24
## * edad -> edad...33
## * edad -> edad...39
## * TOTAL -> TOTAL...88
## * ...
```

Con esta función, por ejemplo, estamos haciendo una descripción gruesa de la variable TCH.

```
summary(PantaleonData$TCH)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  50.97   88.69  102.50  102.91  116.83  181.81
```

Para resumir o describir el comportamiento estadístico de una variable, a veces incluso uno puede prescindir de la tabulación y la graficación porque la prosa cumple de manera adecuada el propósito de la descripción. La tabulación y la graficación se hacen mucho más interesantes cuando uno desea explorar relaciones entre variables.

La tabulación y La graficación

Ahora vamos a concentrarnos en ver cómo puede “automatizarse” la descripción estadística de TCH, pero en función de otra variable, por ejemplo, estación. Para ello, vamos a usar una nueva librería llamada apaTables.

```
library(apaTables)
apa.1way.table(iv = Estacion, dv = TCH, data = PantaleonData, show.conf.interval = TRUE)
```

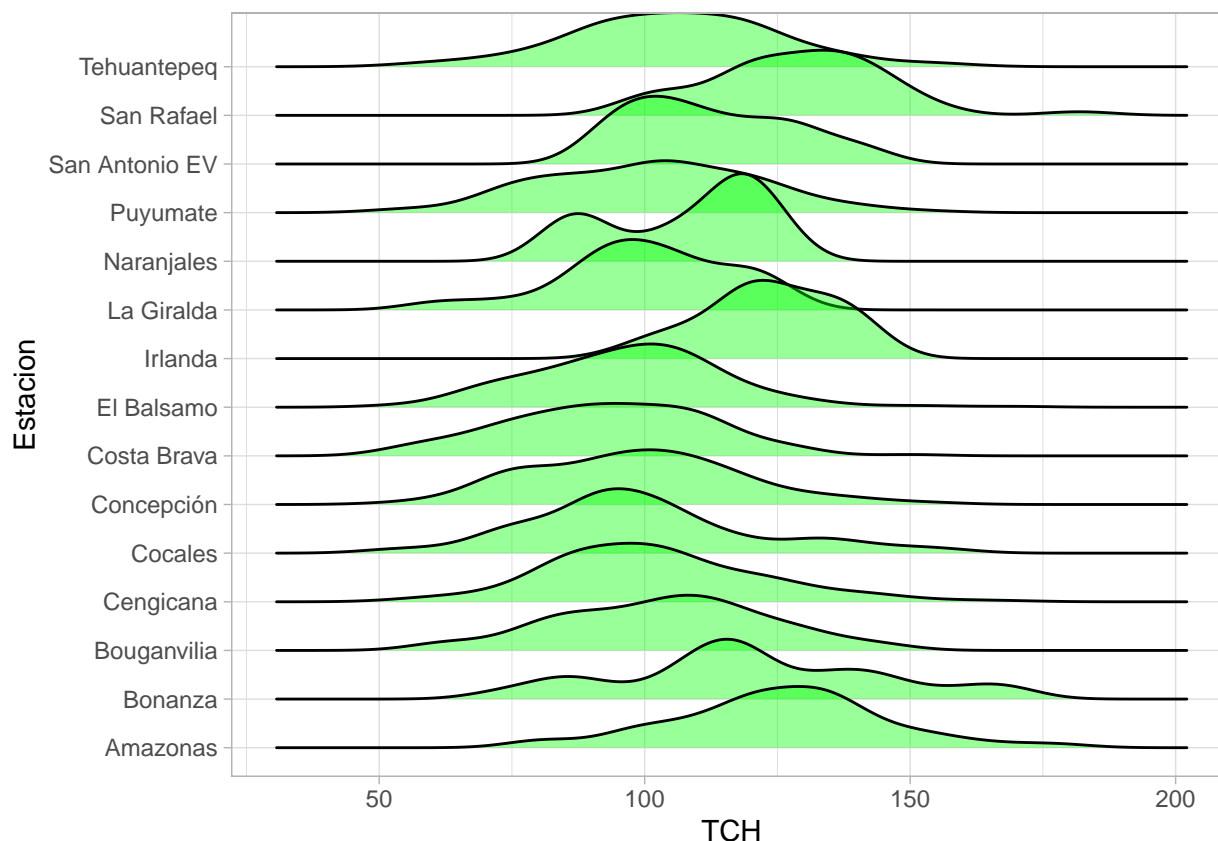
```
##
##
## Descriptive statistics for TCH as a function of Estacion.
##
##      Estacion      M      M_95%_CI      SD
##      Amazonas 125.30 [119.22, 131.39] 20.02
##      Bonanza 119.98 [109.97, 129.99] 24.78
##      Bouganvilia 103.59 [99.33, 107.86] 19.41
##      Cengicana 101.85 [99.17, 104.52] 19.89
##      Cocales 99.23 [93.52, 104.93] 21.50
##      Concepción 98.29 [95.38, 101.21] 19.59
##      Costa Brava 93.07 [89.79, 96.35] 18.67
##      El Balsamo 97.05 [94.60, 99.51] 18.11
##      Irlanda 123.51 [117.43, 129.60] 12.24
##      La Giralda 100.41 [93.35, 107.47] 15.92
##      Naranjales 107.68 [93.57, 121.78] 15.25
##      Puyumate 100.97 [97.97, 103.96] 19.99
##      San Antonio EV 112.55 [101.54, 123.56] 15.39
##      San Rafael 128.72 [123.81, 133.63] 16.54
##      Tehuantepeq 105.93 [103.77, 108.08] 19.45
##
## Note. M and SD represent mean and standard deviation, respectively.
## LL and UL indicate the lower and upper limits of the 95% confidence interval
## for the mean, respectively.
## The confidence interval is a plausible range of population means that could
## have caused a sample mean (Cumming, 2014).
```

La tabla anterior, nos presenta el promedio, la desviación estándar y los intervalos de confianza de la variable TCH para cada una de las estaciones de producción.

Ahora, vamos a mostrar un ejemplo que ilustre cómo podríamos presentar los resultados anteriores pero haciendo uso de gráficos. Para ello vamos a usar las librerías ggplot2 y ggridges.

```
library(ggplot2)
library(ggbridges)
ggplot(PantaleonData,
  aes(x=TCH, y=Estacion)) +
  geom_density_ridges(fill="green", alpha = 0.4) +
  ylab("Estacion") + xlab("TCH") +
  theme(axis.text.y = element_text(family="Arial", face="bold", colour="black", size=rel(1))) +
  theme(axis.text.x = element_text(family="Arial", face="bold", colour="black", size=rel(1))) +
  theme_light()
```

```
## Picking joint bandwidth of 6.77
```



Tabulación versus Graficación: Algunas mejores prácticas

- La tabulación enfatiza precisiones numéricas de los estadísticos reportados en la tabla que no son evidentes en las gráficas.
- La graficación suele ser más “amigable” a la vista de la audiencia en comparación con una fría tabla que muestra números.
- Si se quiere describir una variable, a veces es suficiente con la prosa (siendo innecesarias las tabulaciones y las gráficas). Con dos o más variables, hay que usar tabulación y graficación.
- La audiencia a la cual se dirige un informe debe ser el criterio más importante para decidir cómo reportar los resultados.
- La mejor práctica, sin embargo, es aquella que hace un uso inteligente de tabulaciones y graficaciones.

Veamos un ejemplo en el siguiente artículo:

Correa, J.C., Garzón, W., Brooker, P., Sakarkar, G., Carranza, S., Yunado, L. & Rincón, A. (2019). Evaluation of Collaborative Consumption of Online Food Delivery Services through Web Mining Techniques. *Journal of Retailing and Consumer Services*, 46, 45-50. doi: <https://doi.org/10.1016/j.jretconser.2018.05.002> cuyo repositorio está disponible en este enlace <https://github.com/jcorrean/WebMining-OFD>