

# Taller de R: Estadística y Programación

## Taller 3

Abril 26 de 2022

En este taller se evalúan las clases 9 a 15 del curso. Se presentan 2 tipos de taller (A y B), pero usted solo debe desarrollar 1 de ellos. Para realizar este trabajo, podrá hacer grupo de hasta tres personas. Sea creativo en su código (no hay una respuesta única, todos los métodos que permitan obtener la misma respuesta son válidos). Cuando encuentre material de apoyo en línea que le permita solucionar algún problema, no olvide citar la fuente. Por último, lea atentamente las instrucciones del taller.

### Instrucciones

- Este taller pesa el **45%** de la nota total del curso.
- No seguir las instrucciones tiene una penalización del **20%** de la nota final.
- El taller debe ser terminado antes de las 23:59 horas del 5 de junio de 2022.
- En el repositorio asignado, debe incluir tres carpetas: input (datos originales), output (datos procesados) y code (script con la respuesta del taller).
- Debe escribir en el siguiente Excel group-task.xlsx su código, su usuario de GitHub (por favor respetar mayúsculas y minúsculas) y en la columna grupo debe escribir un número que indique el grupo al cual pertenece (si va a desarrollar el task en grupo).
- Por favor sea lo más organizado posible y comente paso a paso cada línea de código, pero recuerden **NO** usar ningún acento o carácter especial dentro del código para evitar problemas al abrir los scripts en los diferentes sistemas operativos.
- En las primeras líneas del script debe escribir su nombre, código y la versión de R sobre la que está trabajando.
- Para este taller debe usar (al menos) las librerías `pacman`, `rio`, `tidyverse`, `sf`, `leaflet`, `rvest`, `xml2`, `osmdata`, `ggsn` y `XML`. Adicionalmente puede requerir algunas librerías para las regresiones.

## Taller A

### 1. Datos espaciales (50%)

- **1.1. Descargar datos:**
  - Seleccione una ciudad de Colombia.
  - Usando OpenStreetMaps **descargue los shapefiles de:**
    - \* Estaciones de autobús (puntos).
    - \* Vías principales y de transporte masivo (líneas). Por ejemplo, si usted selecciona Cali debe dejar las rutas avenidas y las rutas del MIO, si selecciona Bogotá-Transmilenio, Barranquilla-Transmetro...
    - \* Limites de los barrios o manzanas si usted lo prefiere (polígonos).
- **1.2. Visualizar información:**

Use la función `leaflet` para visualizar la información descargada en el punto anterior.

- **1.3. Estimar distancias:**

- Estime la distancia de cada barrio/manzana a las estaciones de autobús en la ciudad y encuentre la distancia promedio a una estación. Debe agregar esta información al objeto que contiene barrio/manzana. Llame a esta variable `mean_bs`.
- Estime la distancia de cada barrio/manzana a las vías principales de la ciudad y encuentre la distancia promedio a una vía principal. Debe agregar esta información al objeto que contiene barrio/manzana. Llame a esta variable `mean_via`.
- Genere una tabla con las estadísticas descriptivas de las variables de distancias.

- **1.4. Plot Mapping:**

Usando `ggplot` pinte dos mapas con los polígonos de barrio/manzana. Asigne a cada polígono un color de acuerdo con los valores de los cuartiles de distribución de las distancias a las estaciones de buses y las vías principales. Exporte estos mapas a la carpeta **output**.

- **1.5. Exportar datos:**

Extraiga la información de barrios y exporte la base de datos en un objeto **.rds**.

## 2. Regresiones (35%)

La Oficina del Alto Comisionado para la Paz (OACP) tiene el registro oficial de víctimas por minas antipersona (MAP) y municiones sin explotar (MUSE). Estos registros pueden obtenerse en la página oficial de la OACP. Para este ejercicio usted cuenta con una base de datos que contiene los registros de las víctimas de MAP-MUSE en el departamento de Norte de Santander. La variable **fallecido** toma el valor de 1 si la persona fallece en el accidente y 0 si resulta herida.

- **2.1. Importar:**

Lea el archivo **data/output/df\_mapmuse.rds** y estime un modelo de probabilidad lineal en el que **fallecido** es la variable dependiente. Y use las demás variables como variables explicativas. Almacene los resultados de la estimación en un objeto llamado **ols**.

- **2.2. Coefplot:**

Exporte a la carpeta **output** los gráficos con los coeficientes (`coef-plot`) de las estimaciones.

- **2.3. Estimación:**

Estime la ecuación del punto **2.1.** usando un modelo **logit** y un modelo **probit**, almacene los resultados de las estimaciones en dos objetos llamados **logit** y **probit** respectivamente.

- **2.4. Exportar resultados:**

Exporte los resultados de los tres modelos en una misma tabla usando la función **stargazer**.

- **2.5. Presentar resultados:** De los objetos **logit** y **probit** exporte a la carpeta **views** dos gráficos con el efecto marginal de la distancia a un centro médico sobre la probabilidad de fallecer.

## 3. Web-scraping (15%)

- **3.1.** Desde la consola de Rstudio lea la siguiente url <https://es.wikipedia.org/wiki> y cree un objeto que contenga el HTML de la página como un objeto **xml\_document**.
- **3.2.** Use el `xpath` para extraer el título de la página (Departamentos de Colombia).
- **3.3.** Extraiga la tabla que contiene los departamentos de Colombia.

## Taller B

Las personas que realicen este taller podrán participar por uno de los 2 mug personalizados de la clase y promocionaré sus resultados (páginas web) en mis redes sociales. Además, podrán contar con ayuda adicional para realizar el punto 2 de este taller.

### 1. Datos espaciales (50%)

Deberá realizar los mismos puntos del **Taller A**. Si desea puede adicionar otros amenities, bar, restaurantes, ...

### 2. Página WEB (50%)

Deberá consolidar los resultados en un Rmarkdown y colgarlo como una página web.