

# Taller de R: Estadística y Programación

## Taller 2

28/10/2020

En este taller se evalúan los módulos 1 y 2 del curso. Se presentan 2 tipos de taller (A y B), pero usted solo debe desarrollar 1 de ellos. Para realizar este trabajo, podrán trabajar de manera individual o hacer grupos de hasta dos personas. Sea creativo en su código (no hay una respuesta única, todos los métodos que permitan obtener la misma respuesta, son válidos). Cuando encuentre una ayuda en línea que le permita solucionar algún problema, no olvide dar los créditos. Por último lea atentamente las instrucciones del taller.

## Instrucciones

- No seguir las instrucciones tiene una penalización del **10%** de la nota total.
- Este taller representa el 30% de la nota del curso.
- El taller estará disponible en SICUA desde las 20:00 horas del jueves 29 de octubre de 2020.
- Debe elegir y desarrollar solo 1 taller (A o B).
- Por favor sea lo más organizado posible y comente paso a paso cada línea de código, pero recuerden **NO** usar ningún carácter especial dentro del código para evitar problemas al abrir los scripts en los diferentes sistemas operativos.
- El taller debe ser entregado antes de las 23:59 horas del martes 20 de noviembre de 2020.
- Si usted decide hacer el taller A, debe enviar el script con la solución al correo [ef.martinezg@uniandes.edu.co](mailto:ef.martinezg@uniandes.edu.co) con el siguiente asunto: *Taller 1 de R, 2020-2*. Al comienzo del script debe estar el nombre(s) y el código de la(s) persona(s) que trabajaron en el código. Además el script debe ser guardado con el primer apellido de cada uno de los integrantes más el código estudiantil separado por \_\_, tal como se muestra a continuación: *martinez\_201725842\_arevalo\_201522054.R*.
- Si usted decide hacer el taller B, debe invitar al usuario [eduard-martinez](#) como colaborador de su repositorio. Este [vídeo](#) le puede ayudar a entender mejor el funcionamiento de los repositorios de GitHub.

## Sobre la GEIH

La recolección de la Gran Encuesta Integrada de Hogares -GEIH- empezó el 7 de agosto de 2006 en su módulo central de mercado laboral e ingresos y, a partir del 11 de septiembre, con su módulo de gastos de los hogares. A partir del 10 de julio de 2006 se amplió la cobertura de la ECH a once ciudades adicionales, a las trece principales ciudades y áreas metropolitanas, al resto de cabeceras y al resto rural; cobertura que en la actualidad mantiene la GEIH. Actualmente la encuesta se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Esta tiene una muestra total anual de 240.000 hogares aproximadamente, lo que hace que sea la de mayor cobertura a nivel nacional.

La GEIH recoge información a tres niveles geográficos: Áreas metropolitanas, Cabeceras y Restos. En **Áreas** se recoge información para las 13 principales áreas metropolitanas del país. Por su parte, **cabecera** lo hace para todas las cabeceras municipales (o zonas urbanas del país, inclusive las áreas metropolitanas). Finalmente, **resto** recoge información para las zonas rurales del país. Para cada nivel geográfico se puede acceder de manera libre a los siguientes módulos:

- 1. Características generales personas: se recoge información de algunas características observables de las personas como la edad, sexo, ...
- 2. Desocupados: información de las personas que reportaron estar desocupadas pero que se encontraban buscando empleo.
- 3. Fuerza de trabajo: información de las personas que pertenecen a la fuerza de trabajo.
- 4. Inactivos: información de las personas que reportaron no estar trabajando pero que tampoco están buscando empleo.
- 5. Ocupados: información de las personas que reportaron estar ocupadas al momento de la encuesta.
- 6. Otras actividades y ayudas en la semana: información sobre ingresos.
- 7. Otros ingresos: información sobre ingresos.
- 8. Vivienda y hogares: características de la vivienda y el hogar de la persona encuestada.

Todos los módulos poseen estas 3 variables (`secuencia_p`, `orden` y `directorio`) que permiten cruzar información de todos los módulos para un mismo individuo. Puede obtener una descripción detallada de todas las variables [aquí](#). Sin embargo, para los propósitos de este taller puede ser suficiente con la información que le será suministrada en los siguientes incisos.

## Taller A

### 1 Organizar la GEIH

#### 1.1. Descargar bases de datos

Dirijase a la pagina de <http://microdatos.dane.gov.co> y descargue los microdatos anonimizados de la GEIH para todos los meses del año 2019.

#### 1.2. Importar datos

Usando loops o funciones, importe a R los módulos de *Características generales*, *ocupados*, *inactivos*, *desocupados* y *fuerza de trabajo* para las cabeceras municipales y el resto. Asegúrese de crear un objeto para cada modulo en el que apile todos los meses del mismo modulo. **Hint:** Asegúrese de crear una variable que le permita identificar las observaciones de cada mes y una variable dicótoma que identifique las observaciones de *cabecera* y *resto*.

**Bonus:** puede intentar guardarlos en listas y usar la función `rbindlist()` para apilarlos.

#### 1.3. Unir datos

Use la funcion `merge` para crear una base de datos que contenga todos los modulos. **Hint:** use las variables `secuencia_p`, `orden` y `directorio`.

## 1.4 Descriptivas

Use las funciones `subset()`, `ggplot()`, `group_by()` y `summarize()` entre otras, para generar algunas estadísticas descriptivas (gráficos y tablas) a nivel departamental de la tasa de analfabetismo, la tasa de graduados de educación superior y los ingresos laborales para las personas entre 18 y 28 años. Tenga en cuenta algunas dimensiones como mes, sexo y urbano/rural. Las tablas las puede plotear sobre la consola, pero los gráficos los debe exportar en formato `.jpeg` a la carpeta *results*.

## 2. Mapas en R

De la carpeta *data/original* importe el shapefile `\textit{MGN_DPTO_POLITICO}` y pinte la tasa de analfabetismo y la tasa de graduados de educación superior departamental que calculo en el punto anterior.

### Taller B

Intente hacer el **Taller B** pero incorporando dentro de la función/loop los archivos del año 2018. Puede descargarlos [aquí](#).