

Curso básico de R para ciencia de datos

Stephanie Hereira Pacheco

Proyecto 2

Parte 1

1. Escoger un dataset para trabajar: En este primer paso podemos escoger algunos de los datasets vistos o trabajar con uno de nuestro trabajo/carrera/investigación (importarlo a R).
2. Crear un proyecto nuevo que se llame “**Proyecto_2**” y crea un script nuevo también que se llame “**Proyecto_2.R**” y en este script escribe todo lo que vayamos a hacer en este primer proyecto. Describe cada paso que des usando el “#” para los comentarios y explicando cada paso en detalle.
3. Explora este dataset, usar las funciones del *tidyverse* o R base para hacer subconjuntos de tus datos (al menos dos subconjuntos). **Nota: Poner todos estos pasos en el script**
4. Reliza 3 gráficas (1 de R base y 2 de ggplot2). Sé creativo, trata de cambiar formato, colores, escalas, etc.

Parte 2

Para esta parte trabajaremos con el dataset en este link: <https://raw.githubusercontent.com/ndphillips/ThePiratesGuideToR/master/data/cafeinestudy.txt>

Este dataset presenta las variables:

- “drink”ó tipo de bebida y las opciones son: “coffee” (café) y “greentea” (té verde)
- cups (tazas) : 1 ó 5 tazas
- gender (género): female (femenino) y male (masculino).
- age (edad)
- score (puntuación)

Con estos datos, realiza:

1. Carga este dataset a R (descárgalo, cárgalo o súbelo).
2. Calcula el promedio de edad para cada género
3. Calcula el promedio de edad para cada nivel combinado de género y bebida
4. Sólo para los hombres, calcula el valor máximo de la variable “score” para cada edad
5. Crea un dataframe mostrando de cada bebida, su promedio, mediana, máximo y mínimo de la variable “score”
6. Sólo para mujeres encima de los 20 años, crea una tabla mostrando para cada nivel combinado de bebida (drink) y tazas (cups), el promedio, mediana, máximo y mínimo de la variable “score”. También incluye una columna mostrando cuántas personas hay en cada grupo (n)