

Lista de ejercicios 2

Q-lab CC.SS.

Objetivo:

El objetivo de esta lista de ejercicios es practicar las habilidades de manipulación de datos en R utilizando el paquete `dplyr`, aplicándolas a una base de datos que contiene información socioeconómica y política de Estados Unidos. A través de estos ejercicios, se busca que los estudiantes se familiaricen con técnicas clave como la filtración, agrupación, creación de nuevas variables, y ordenamiento de datos, sin necesidad de realizar un análisis profundo de los resultados. La práctica se enfoca en el manejo eficiente y correcto de los datos para prepararlos para posibles análisis futuros.

Indicaciones:

- Desarrollar esta la tarea en un archivo R Markdown.
- Colocar sus archivos de la siguiente manera: Nombres_Apellidos_Lista2
- Fecha de entrega: Hasta las 11 pm del 24 de agosto al correo: i.vargasv@pucp.edu.pe o subirlo a la carpeta de drive

Preparativos:

a. Instalación y Carga de Librerías: Asegúrense de tener instaladas y cargadas las librerías `tidyverse`, `rio` y `dplyr`.

b. Importar la base de datos “usa.sav” (<https://github.com/qlabpucp/R-Studio-basics/blob/main/Datasets/usa.sav>) Pueden usar el siguiente código:

```
library(foreign)
usa <- read.spss("usa.sav", #en formato spss
  use.value.labels = TRUE, #cargar la base de datos con las etiquetas de los datos
  to.data.frame=TRUE) #que la base de datos se cree en un data frame
```

Ejercicios:

1. Exploración y Selección de Datos

1.1. Filtra los datos para mostrar únicamente las observaciones donde la edad (`edad`) es mayor o igual a 50 años. Usa la función `filter`.

1.2. Filtra los datos para mostrar únicamente a los hombres mayores de 50 años que se consideran políticamente muy liberales y que opinan que se debe permitir el aborto. Usa la función `filter` con múltiples condiciones.

1.3. Filtra las observaciones para mostrar solo a mujeres (sexo) que sean casadas (marital) y con ingresos mayores de 30,000. Usa la función filter con múltiples condiciones.

2. Creación de Nuevas Variables y Categorización:

2.1. Analizar qué variables deben ser configuradas. Usar 'str' para ver el tipo de variables. Particularmente, convertir la variable gorepre y gorepst a un tipo numérico.

2.2. Crear una nueva variable difgore. Luego, crear una nueva variable "difgore" que calcule la diferencia entre la percepción de Gore antes (gorepre) y después (gorepst) de las elecciones. Esto te permitirá analizar cómo cambió la percepción de Gore en los entrevistados. Posteriormente, ordena los datos en orden descendente según la percepción posterior a las elecciones (gorepst). Esto te permitirá ver a aquellos que tuvieron la percepción más alta de Gore después de las elecciones.

2.3. Crea una nueva variable llamada grupo_edad que clasifique a las personas en tres grupos: "Joven" (menos de 30 años), "Adulto" (30 a 60 años) y "Mayor" (más de 60 años). Usa la función mutate y case_when.

3. Agrupación y Sumarización

3.1. Agrupa los datos por region y calcula la media y la desviación estándar de los ingresos (ingresos). Luego, ordena los resultados de mayor a menor media de ingresos. Para lo último usa arrange(desc(variable)).

3.2. Agrupa los datos por aborto (opinión sobre el aborto) y calcula la media y la mediana de la edad (edad) para cada grupo.

3.3. Agrupa los datos por region y calcula la diferencia promedio en la percepción de Clinton entre 1996 (clintpre) y 2000 (clintpst).

3.4. Agrupa los datos por marital (estado civil) y voto00 para calcular la media de los ingresos (ingresos) según si votaron o no en el año 2000. Luego, analiza las diferencias por estado civil.