



Procesamiento y Visualización de Datos Espaciales en R

Profesor: José Luis Texcalac Sangrador

Laboratorio — 06

El paquete [dplyr](#) forma parte del núcleo de paquetes que integran [Tidyverse](#), facilita varias de las tareas básicas y primordiales para el procesamiento de información, a través de una serie de “verbos” mediante los cuales es posible programar la manipulación de datos.

¿Qué se espera de ti en este laboratorio?

Esta actividad pretende que fomente su habilidad en la redacción de código y sus resultados utilizando el lenguaje de marcado [Markdown](#); se busca que fortalezca el aprendizaje en el uso de “verbos” del paquete [dplyr](#) para la manipulación de datos. Se espera que la edición que programe en su archivo [R Notebook](#) contenga el formato, diseño y calidad que usted considere adecuada para la presentación de un reporte [html](#).

Indicaciones por considerar:

- Suba su laboratorio a la plataforma [Google Classroom](#) a más tardar antes del inicio de la próxima sesión (jueves 14 de octubre).
- No es necesario el envío de su laboratorio por correo electrónico.
- Lo que debe usted entregar para evaluar su laboratorio es:
 - Archivo con extensión [.Rmd](#)
 - Archivo con extensión [.html](#)
 - Archivo con extensión [.css](#)
- La cuenta de correo electrónico para cualquier duda respecto al curso es: [✉ jtexcalac@insp.edu.mx](mailto:jtexcalac@insp.edu.mx).
- Nombre todos sus archivos con el patrón: [L06_Nombre](#) (o puede usar su apellido en sustitución del nombre).
- Publique sus dudas en [Google Classroom](#), su profesor o compañeros le pueden auxiliar.
- ¡Muy importante! Comente su [R Notebook](#), explique los procedimientos que va a realizar en cada bloque de código, comente los resultados, la idea es que practique la redacción en lenguaje [Markdown](#).

- Para la realización de este laboratorio es necesario instalar el paquete [fauxnaif](#).

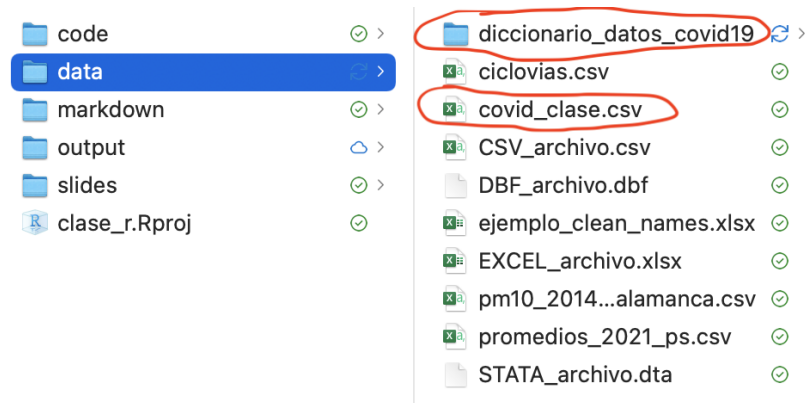
```
remotes::install_github("rossellhayes/fauxnaif")
```

- Es necesario ejecutar todos los chunk para que todos los resultados de su código (tablas, gráficos, etc.) sean visibles en su archivo [R Notebook](#).
- Los nombres de columna de sus mallas de datos deben estar en formato [snake_case](#).
- Revise los hipervínculos que se le comparten en el texto a lo largo del documento.

Instrucciones:

Para la realización de su laboratorio trabaje sobre el proyecto que generó para la clase ([clase_r](#)), dentro del proyecto genere un nuevo documento [R Notebook](#), guárdelo en la carpeta markdown y en el programe las siguientes tareas.

1. Descargue del sitio de la clase en Google Classroom el archivo [covid_clase.zip](#):
 - i. En la siguiente imagen se le sugiere cómo organizar sus archivos una vez que los descomprima, los archivos que importan son los resaltados en color rojo, el resto de los archivos pueden variar de usuario en usuario.



- ii. Cargue a su sesión la malla ["covid_clase.csv"](#) y nombre a su objeto como [covid](#).
- iii. Seleccione las columnas: SEXO, ENTIDAD_RES, MUNICIPIO_RES, FECHA_INGRESO, FECHA_SINTOMAS, FECHA_DEF, INTUBADO, EDAD, EMBARAZO, INDIGENA, DIABETES, EPOC, ASMA, INMUSUPR, HIPERTENSION, OTRA_COM, CARDIOVASCULAR, OBESIDAD, RENAL_CRONICA, TABAQUISMO, CLASIFICACION_FINAL.
- iv. Convierta a [NA](#) los valores de cada una de las columnas de acuerdo con lo especificado en los [catálogos](#), revise los [descriptores](#) para entender el contenido de cada una de las variables.

- Por ejemplo: La variable [SEXO](#) contiene valores [1](#), [2](#) y [99](#) que corresponden a Mujer, Hombre y NO ESPECIFICADO, entonces hay que convertir los valores [99](#) a [NA](#).
 - Identifique los distintos valores en cada una de las variables que deben convertirse a [NA](#).
 - Las imágenes al final de este archivo le apoyarán a realizar esta parte del laboratorio.
- v. Posicione las columnas en el siguiente orden: ENTIDAD_RES, MUNICIPIO_RES, FECHA_INGRESO, FECHA_SINTOMAS, FECHA_DEF, EDAD, SEXO, INTUBADO, EMBARAZO, INDIGENA, DIABETES, EPOC, ASMA, INMUSUPR, HIPERTENSION, CARDIOVASCULAR, OBESIDAD, RENAL_CRONICA, TABAQUISMO, CLASIFICACION_FINAL.
 - vi. Filtre los registros de la columna CLASIFICACION_FINAL que cumplan alguno de los siguientes criterios:
 - Caso de COVID-19 confirmado por asociación clínica epidemiológica.
 - Caso de COVID-19 confirmado por comité de dictaminación.
 - Caso de SARS-CoV-2 confirmado.
 - vii. Ordene los datos por fecha de defunción.
 - viii. Aplique el comando [summary\(\)](#) a la malla de datos
 - ¿Son correctos los resultados descriptivos de todas las variables?
 - ¿Por qué?
 - ix. Para evitar que sus tablas se alarguen en el html, utilice el comando [scroll_box\(\)](#) con las dimensiones que le parezcan apropiadas.
 - El comando y ejemplos de uso los puede usted encontrar en una de las ligas que se les sugirió para la edición de tablas en el laboratorio pasado.
 - El comando pertenece al paquete [KableExtra](#).
2. Visite la página [aire](#) de la Secretaría del Medio Ambiente de la Ciudad de México.
 - i. Clic a la pestaña [Datos](#)
 - ii. En el apartado [Promedio 24 horas](#) seleccione partículas suspendidas.
 - iii. Descargue los datos del año 2021 en formato [.csv](#) ([promedios_2021_ps](#)).

- iv. Importe la malla de datos "promedios_2021_ps.csv" a su sesión, nombre al objeto como `contam`.
- v. Deberá utilizar el argumento `skip` dentro del comando `read_csv` para leer adecuadamente la malla de datos (puede usar el comando `?read_csv` para buscar ayuda dentro de R).
- vi. Calcule el promedio por estación y contaminante.
- vii. Ordene los datos por estación.
- viii. Separe en una malla los resultados de PM_{10} y otra para $PM_{2.5}$, nombre a sus objetos como `pm10` y `pm25`.

Material de apoyo

Puede usted apoyarse de los siguientes recursos para la realización de su laboratorio

- [Introducción a dplyr](#)
- [Tutorial KableExtra](#)
- [Tablas en R con KableExtra](#)
- [Rstudio y Markdown](#)

A. Recodificar valores como NA (variable por variable).



Recodificar valores

Genero columna, si ya existe entonces la sobrescribe.

Activar paquete en la sesión

Comando para convertir valores a NA

Columna que contiene los valores NA

```
library(fauxnaif)
malla %>%
  mutate(sexo = na_if_in(sexo, 98, 99),
         edad = na_if_in(edad, -88, -99),
         peso = na_if_in(peso, -88, -99)) %>%
  print()
```

malla			
nombre	sexo	edad	peso
Kevin	1	17	-99
Brayan	1	-99	61.7
Kimberly	98	15	51.9
Britany	2	16	59.3
Brandon	99	17	-88
Melany	2	-88	61.6

nombre	sexo	edad	peso
Kevin	1	17	NA
Brayan	1	NA	61.7
Kimberly	NA	15	51.9
Britany	2	16	59.3
Brandon	NA	17	NA
Melany	2	NA	61.6

Valores a convertir a NA

B. Recodificar valores como NA (Varias variables a la vez).



Recodificar valores

Es posible hacer todo el proceso para varias columnas a la vez

```
library(fauxnaif)
malla %>%
  mutate(across(c(sexo:peso), na_if_in, -88, -99, 98, 99)) %>%
  print()
```

malla			
nombre	sexo	edad	peso
Kevin	1	17	-99
Brayan	1	-99	61.7
Kimberly	98	15	51.9
Britany	2	16	59.3
Brandon	99	17	-88
Melany	2	-88	61.6

nombre	sexo	edad	peso
Kevin	1	17	NA
Brayan	1	NA	61.7
Kimberly	NA	15	51.9
Britany	2	16	59.3
Brandon	NA	17	NA
Melany	2	NA	61.6