



Procesamiento y Visualización de Datos Espaciales en R

Profesor: José Luis Texcalac Sangrador

Laboratorio — 07

Los factores se utilizan para trabajar con variables categóricas, es decir, variables que tienen un conjunto fijo y conocido de valores posibles. Los valores que un factor puede contener están delimitados por los niveles, es decir, las diferentes categorías presentes en una variable categórica. Los niveles pueden estar codificados tanto en valores numéricos como caracteres. En el entorno de [Tidyverse](#) los factores se trabajan con el paquete [forcats](#) que nos facilita varias de las tareas básicas y primordiales para el procesamiento de este tipo de información.

¿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 y en el procesamiento de datos de tipo factor. 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](#) (opcional)
- 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: [L07_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](#).
- 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.
- Recuerde que sus tablas a incluir en su archivo R Notebook deben contener sólo un máximo de 100 registros (revise el ejemplo del laboratorio 06 para mayor detalle).

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. Para esta parte del laboratorio usted utilizará el archivo [covid_clase.csv](#) que usted descargó del sitio de la clase en Google Classroom para el laboratorio pasado.
 - i. Ejecute el siguiente código:

```
covid <-
  read_csv("../data/covid_clase.csv") %>%
  clean_names() %>%
  print()

cov_def <-
  covid %>%
  select(entidad_res, municipio_res, fecha_ingreso:fecha_def, edad, sexo,
         intubado, embarazo, indigena:hipertension, cardiovascular:tabaquismo,
         clasificacion_final) %>%
  mutate(across(c(entidad_res, municipio_res, sexo:tabaquismo),
                na_if_in, 36, 97, 98, 99, 997, 998, 999)) %>%
  filter(clasificacion_final < 4) %>%
  arrange(fecha_def) %>%
  print()
```

- ii. Convierta a factor las variables: sexo, intubado, indígena, diabetes, epoc y asma, para la variable sexo utilice las etiquetas del catálogo de variables. Guarde su objeto como [cov_def](#) y ejecute el [summary](#)

- iii. Seleccione las columnas: `entidad_res`, `municipio_res`, `fecha_ingreso`, `fecha_sintomas`, `fecha_def`, `edad`, `sexo`, `intubado`, `indigena`, `diabetes`, `epoc`, `asma` y guarde su objeto como `cov_def`.
- iv. Genere las columnas: "day", "month", "year" y "wek" que contengan el día, mes, año y número de semana respectivamente. Guarde su objeto como `cov_def`.
- v. Usando el comando `group_by`, genere los procesamientos para resolver los siguientes ejercicios

- Tabla muestre las muertes por entidad en orden descendente.
- Compare el resultado anterior con el siguiente código

```
cov_def %>%  
  count(entidad_res, name = "tot_def") %>%  
  arrange(desc(tot_def))
```

- Tabla que muestre los 20 municipios con mayor número de muertes y su entidad.
 - Tabla que muestre las 10 entidades con mayor número de muertes del sexo femenino (también intente la tabla usando el comando `count`).
 - Tabla que muestre las 10 entidades con mayor número de muertes del sexo masculino
 - Revise las páginas que se le sugieren al final de este archivo y revise los contenidos que le ayudarán a realizar los siguientes ejercicios:
 - Genere una tabla que contenga las 10 entidades con mayor número de defunciones y el porcentaje de cada una (use el comando `tabyl`).
 - Genere una tabla de contingencia que nos muestre, para la Ciudad de México ("09"), las frecuencias por sexo según su condición de intubado (use el comando `tabyl`).
2. Importe a su R Notebook el archivo `Base_SUN_2018.csv`, guarde el objeto como `sun` y realice las siguientes tareas.
- i. Filtre los registros que inicien con la letra "M" en la columna `cve_sun` de la malla `sun`.

Código de ejemplo para filtrar con el comando `str_detect(...)`

```
[1] malla <-
```

```
[2] malla %>%  
[3] filter(str_detect(categ, "^z")) %>%  
[4] print()
```

El código anterior indica lo siguiente:

[1] Se guarda el resultado como malla.

[2] Se trabaja con el objeto malla.

[3] El comando `str_detect` identifica patrones en los valores de una columna, el operador `^` puede interpretarse como "inicia con_", en conjunción con la letra `z` podemos interpretar el resultado como: "detecta todos los valores de la columna `categ` que inicien con la letra `z`". El comando `filter` nos va a devolver como resultado, todas aquellas filas que cumplan el patrón definido dentro de `str_detect`.

- ii. Seleccione a las columnas; `nom_ent`, `nom_mun`, `cve_mun`, `cve_sun`, `nom_sun`, `pob_2018`.
- iii. Genere una tabla que muestre las 20 Zonas Metropolitanas (`nom_sun`) con mayor número de habitantes.

Material de apoyo

Material de apoyo para resolver los ejercicios del punto 1.

- [The janitor package](#)
- [tabyls: a tidy, fully-featured approach to counting things](#)