



Procesamiento y Visualización de Datos Espaciales en R

Profesor: José Luis Texcalac Sangrador

Laboratorio — 05

[Tidyverse](#) es un conjunto de paquetes de **R** diseñados para la ciencia de datos. Los paquetes comparten una filosofía de diseño, gramática y estructuras de datos que los conforman. El paquete [dplyr](#) forma parte del núcleo que conforma [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 usted en este laboratorio?

Esta actividad pretende que consolide de manera práctica los contenidos compartidos en clase, que fomente su habilidad en la redacción de código para el procesamiento de datos y generación de resultados utilizando el lenguaje de marcado [Markdown](#) en un documento [Quarto](#). Busca fomentar su proactividad para ampliar sus capacidades en **R** a través de búsquedas de internet que le permitan identificar información disponible y de utilidad para la realización del laboratorio.

Indicaciones generales:

- Suba su laboratorio a la plataforma [Google Classroom](#) a más tardar antes del inicio de la sesión del jueves 19 de octubre).
- Lo que debe usted entregar para evaluar su laboratorio es:
 - Archivo comprimido **.zip**
 - El archivo comprimido contendrá la carpeta con las salidas de su laboratorio, el archivo con su código con extensión **.qmd** y el archivo **.html** añadir también el archivo **style.css** en caso de haberlo utilizado para su laboratorio.
- La cuenta de correo electrónico para el envío de laboratorios y cualquier asunto relacionado al curso es: ✉ jtexcalac@insp.edu.mx.
- Nombre a su archivo con el siguiente patrón: **L05_Nombre**.
- Publique sus dudas en Google Classroom y su profesor o compañeros le pueden auxiliar.
- Revise las dispositivas de este tema, ahí encontrará el código necesario para la realización de este laboratorio.
- Comente su archivo [Quarto](#), explique los procedimientos que va a realizar en cada bloque de código, comente los resultados, por ejemplo: qué hace el comando, el por qué los argumentos que utilizará, etc.. La idea es que practique la redacción en lenguaje [Markdown](#).

Instrucciones:

1. Para este laboratorio usted no requiere instalar ningún paquete adicional, el código requerido para resolverlo está en las diapositivas de este tema y fue ejecutado por usted en la sesión respectiva.
2. Cuando se requiera una función o paquete adicional para la resolución del laboratorio le será indicado en las especificaciones del ejercicio.
3. Importe el archivo `contaminantes_2019.csv` y guarde el objeto como `cdmx`
 - Recuerde no desplegar toda la tabla en su html, con las primeras 10 o 20 filas es suficiente.
4. Genere un objeto de nombre `pm25` el cual deberá contener únicamente las mediciones (concentraciones) de este contaminante.
5. Excluya de la malla `pm25` las columnas `id_parameter` y `unit`, guarde el resultado con el mismo nombre.
6. Continúe trabajando con la malla `pm25`, asigne el nombre `site` para la columna que contiene el nombre de las estaciones de monitoreo y `pm25` para la que contiene las concentraciones, sobrescriba el objeto con el mismo nombre.
7. A partir de la malla `pm25` genere una nueva malla de nombre `pm25_cdmx` que contenga sólo la información de las estaciones que pertenecen a la Ciudad de México.
 - Para verificar que estaciones pertenecen a la Ciudad de México revise la página que visitó en clase para la descarga de datos de ozono, clic en pestaña "Monitoreo" y finalmente seleccione "Estaciones de monitoreo".
8. Utilice la malla `pm25_cdmx` y redacte el código necesario para replicar la siguiente salida, su html deberá mostrar el código utilizado. No es necesario generar un objeto para este ejercicio. El sufijo `_mean` se refiere al promedio y `_sd` a la desviación estándar, la estación del ejercicio es CAM.

```
# A tibble: 1 × 2
  cam_mean cam_sd
  <dbl>    <dbl>
1    28.4    16.9
```

9. Utilice la malla `pm25_cdmx` y redacte el código necesario para replicar la siguiente salida, su html deberá mostrar el código utilizado. No es necesario generar un objeto para este ejercicio. El sufijo `_max` se refiere al valor máximo y `_min` al valor mínimo, la estación del ejercicio es MER.

```
# A tibble: 1 × 2
  mer_max mer_min
  <dbl>    <dbl>
1    165         1
```

10. Muestre los 15 valores más altos de la estación UAX, para se le pide utilice la función `slice_max()`. Revise el material de apoyo para la solución de este ejercicio. No es necesario generar un objeto para este ejercicio.
11. Utilice la malla `pm25_cdmx` y genere una columna de nombre ciudad y el valor que contendrá la nueva variable en cada una de las filas deberá decir "CDMX", guarde el resultado con el mismo nombre.
12. Ejercicio opcional, no evaluable: use la malla `pm25_cdmx`, agrupe los datos por la columna `site` y genere el promedio para cada estación. A continuación se le comparten un par de sitios de internet que le pueden ayudar a resolver este ejercicio.
 - [Agrupar y resumir](#)
 - [Agrupar y resumir](#)
13. Inserte una imagen, video, texto, meme, canción o algo que quiera compartir con su profesor.

Material de apoyo

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

- [Filtrar los 10 valores mayores de una columna de un DF en R](#) (`slice_max`)
- `dplyr` función `slice`
- [Manipulación de datos con dplyr](#) (`min`, `max`)
- [Summarise, min, max](#)