



Procesamiento y Visualización de Datos Espaciales en R

Profesor: José Luis Texcalac Sangrador

Laboratorio — 05

El procesamiento de datos es una tarea fundamental que antecede al análisis de datos, hasta ahora usted ha revisado en clase cómo diversos paquetes nos apoyan en la “limpieza de datos” y su visualización gráfica.

¿Qué se espera de ti en este laboratorio?

Esta actividad pretende que usted desarrolle habilidades para el procesamiento de datos utilizando los diversos paquetes hasta ahora vistos en clase, mismos que le permitirán procesar y visualizar información, por ejemplo: `{tidyverse}`, `{janitor}`, `{DataExplorer}`, `{fauxnaif}`, `{RColorBrewer}`, `{Viridis}`.

Se requiere que usted aplique los conocimientos adquiridos en clase y que explore mediante búsquedas en internet la información disponible y de utilidad para la realización del laboratorio.

Indicaciones por considerar:

- Suba su laboratorio a la plataforma [Google Classroom](#) a más tardar el martes 19 de noviembre a las 12:00 h.
- Lo que debe usted entregar para evaluar su laboratorio es:
 - Archivo (script) con extensión **.R**
 - Imágenes con los gráficos que se le solicitan.
 - Archivos (CSV, RDS) que usted exporte en 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**
- Cualquier duda puede publicarla en [Google Classroom](#) y su profesor o compañeros le pueden auxiliar.
- Revise las dispositivas de este tema, ahí encontrará código que le apoye a la realización de este laboratorio.

Instrucciones

Genere un nuevo script y en él programe lo que se le solicita.

- El script debe ir con encabezado y secciones.
 - Comente cada uno de sus ejercicios del script.
 - Comente los resultados de sus gráficas.
 - Puede tomar como referencia el estilo de comentar que se muestra en los distintos scripts de apoyo disponibles en [Google Classroom](#).
 - Utilice el formato `snake_case` para sus columnas.
1. Ingrese a la página [Datos Abiertos Dirección General de Epidemiología](#) y revise la información relativa a [Influenza, COVID-19 y otros virus respiratorios](#).
 - Descargar [Diccionario de Datos](#).
 - Ingrese a [Bases Históricas](#).
 - Descargar [Cierre Datos Abiertos Históricos 2020](#).
 - Mueva los archivos descargados a su carpeta data dentro de su proyecto `clase_r` y revise los archivos descargados.
 2. Importe la malla de datos de COVID del año 2020 que usted descargó, llame al objeto como `covid`.
 - ¿qué sucede al importar?
 - ¿qué representa cada fila?
 3. Sobreescriba la malla `covid`, la malla deberá contener sólo los casos registrados como defunciones.
 - ¿cuántas defunciones se registraron en el 2020?
 4. Sobreescriba la malla `covid`, en ella conserve sólo los casos que cumplieron alguna de las siguientes condiciones:
 - Caso de COVID-19 confirmado por asociación clínica epidemiológica
 - Caso de COVID-19 confirmado por comité de dictaminación
 - Caso de SARS-COV-2 confirmado
 - ¿Cuántas defunciones ocurrieron cuando se cumplió alguna de las anteriores causas?
 5. Revise la entidad y municipio de residencia de cada caso

- Conserve en la malla **covid** sólo a aquellos que residieron en estados y municipios del país plenamente identificados.
- ¿cuántos registros se excluyeron y por qué?
- Genere columna **ent_mun** dónde se combine la entidad y municipio de residencia, guarde los cambios (o sea sobreescriba la malla).

6. Mantenga en la malla **covid** sólo a las siguientes columnas, guarde los cambios.

- | | |
|-----------------|------------------|
| - entidad_res | - asma |
| - municipio_res | - inmusupr |
| - edad | - hipertension |
| - sexo | - hipertension |
| - intubado | - cardiovascular |
| - indigena | - obesidad |
| - diabetes | - obesidad |
| - epoc | - tabaquismo |

7. Revise la malla de datos **covid**

- Genere estadísticas descriptivas.
- ¿qué modificaciones debe hacer a sus variables y por qué?
- Realice las modificaciones
- Genere nuevamente estadísticas descriptivas de la malla **covid** y comente el resultado.
- Guarde los cambios.

8. Exporte la malla **covid** y guarde el nombre del archivo como **covid_2020** con extensión CSV y RDS.

- Importe ambas mallas, nombre a sus objetos como **covid_csv** y **covid_rds**, comente la diferencia de ambos objetos y las implicaciones de importar/exportar en ambos formatos.

9. Genere código que responda las siguientes preguntas, comente los resultados en su script, **NO guarde los cambios**.

- ¿Cuáles fueron los 5 estados con mayor número de defunciones?
- ¿cuál fue el municipio con mayor número de defunciones y de qué estado fue?
- ¿cuál fue el día con mayor número de defunciones?
- ¿cuál fue el mes con mayor número de defunciones?

- Genere tres distintos gráficos con los cuales podamos comunicar algún resultado de los datos.
- Comente en su script cada gráfico.
- Exporte como imagen sus gráficos, elija el diseño y use alguna paleta de colores.

10. Continúe trabajando con la malla **covid**.

- Genere una nueva malla de nombre **cov_ent** con los datos de la entidad que se le asigna:
 - Areli: Ciudad de México.
 - Mayra: Nuevo León.
 - Indra: Tlaxcala.
 - Mónica: Estado de México.
 - Adolfo: Jalisco.
 - Karen: Morelos.
- Reduzca la nueva malla **cov_ent** a tres columnas que deberán contener la siguiente información:
 - entidad: Nombre del estado que se le asignó.
 - mes: nombre o número del mes del año.
 - tot_def: total de defunciones de cada mes del año
- Genere y exporte una gráfica que muestre la tendencia de defunciones mensuales para el estado que se le asignó, comente el resultado en su script.

11. Continúe trabajando con la malla **covid**.

- Descargue la malla de datos del **SUN 2018** (Base_SUN_2018.csv).
- Importe la malla del SUN, nombre al objeto como **sun**.
- Conserve sólo las columnas **cve_mun**, **cve_sun** y **nom_sun** de la malla **sun**, guarde los cambios.
- Conserve en la malla **sun** sólo los datos de zonas metropolitanas, puede indentificarlas por la letra **M** en la columna **cve_sun**. (M: zona metropolitana, C: conurbación, P: población). Guarde los cambios.
- Transforme las variables de la malla **covid** al formato que mejor ajuste a su naturaleza, guarde los cambios.

- Genere una nueva malla de nombre **covid_sun** en dónde usted mantenga sólo los datos de los municipios que pertenecen a alguna zona metropolitana.
- Exporte la malla de datos **covid_sun** con el formato que usted considere.
- Genere y exporte una gráfica con datos de la malla **covid_sun**, comente el resultado en su script. La gráfica debe ser distinta a las incluidas anteriormente.

Material de apoyo

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

- Diapositivas de las clases.
- [Estadísticas descriptivas en R](#)
- [Ideas y sugerencias de gráficas.](#)
- Identificar top de valores más altos en una columna ([how to use slice_max\(\)](#)).
- Paleta de colores de estudios [Ghibli](#).
- Paleta de colores inspirada en pinturas del [Museo de Arte Moderno de Nueva York](#).