

Tarea individual 3

27/4/2019

Entrega

Esta tarea tiene que estar disponible en su repositorio de GitHub con el resto de las actividades y tareas del curso el 4 de Mayo al mediodía. Asegurate que tanto Lucía como yo seamos colaboradoras de tu proyecto Tareas_STAT_NT creado hace dos semanas. Recordar seleccionar en en opciones de proyecto, codificación de código UTF-8. La tarea debe ser realizada en RMarkdown, la tarea es individual por lo que cada uno tiene que escribir su propia versión de la misma. El repositorio debe contener unicamente el archivo .Rmd con la solución de la tarea.

Recordá que todas las Figuras deben ser autocontenidas, deben tener toda la información necesaria para que se entienda la información que se presenta. Todas las Figuras deben tener leyendas, títulos. El título (caption) debe contener el número de la Figura así como una breve explicación de la información en la misma. Adicionalmente en las Figuras los nombre de los ejes tienen que ser informativos. En el YAML en Tarea_2.Rmd verás `fig_caption: true` para que salgan los `caption` en el chunk de código debes incluir `fig.cap = "Poner el que tipo de gráfico es y algún comentario interesante de lo que ves"`.

Ejercicio 1

Los datos que vamos a utilizar en este ejercicio están disponibles en el catálogo de datos abiertos uruguay <https://catalogodatos.gub.uy>. Los datos que seleccioné son sobre las emisiones de dióxido de carbono (CO2) correspondientes a las actividades de quema de los combustibles en las industrias de la energía y los sectores de consumo. Se incluyen también emisiones de CO2 provenientes de la quema de biomasa y de bunkers internacionales, las cuales se presentan como partidas informativas ya que no se consideran en los totales. En el siguiente link se encuentran los datos y los meta datos con información que describe la base de datos <https://catalogodatos.gub.uy/dataset/emisiones-de-co2-por-sector>.

Deben bajar los datos y guardarlos en formato csv.

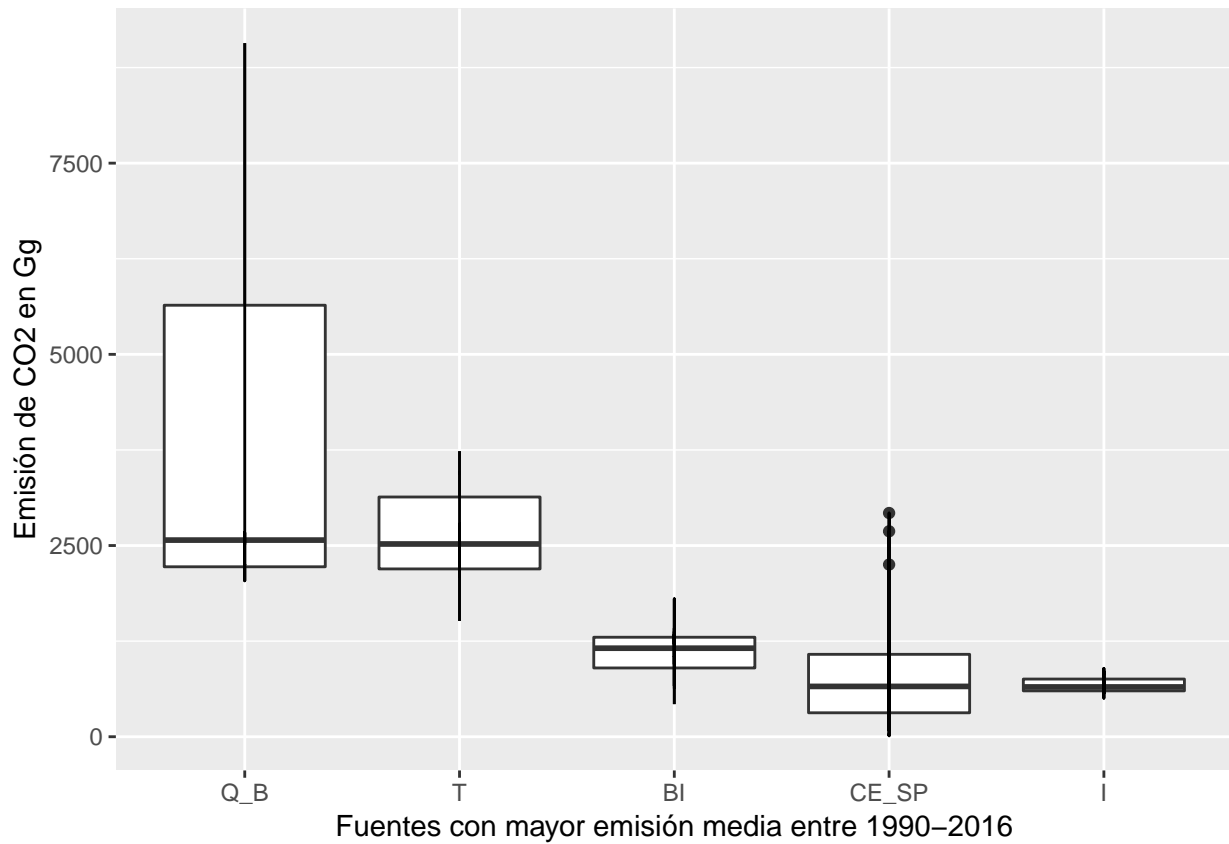
Debe leer con cuidado la información en los meta datos para entender los datos y responder correctamente cada pregunta.

1. Leer los datos usando el paquete `readr` y guardarlos en un objeto llamado `dat`.
2. Reestructure los datos con `gather` de forma de tener tres variables AÑO, una que contenga cada fuente y una que contenga el valor correspondiente a la emisión. Para esto debes identificar cuál es la `key` y el `value` para transformar los datos originales. Guardar los nuevos datos en un objeto llamado `dato_emision`.

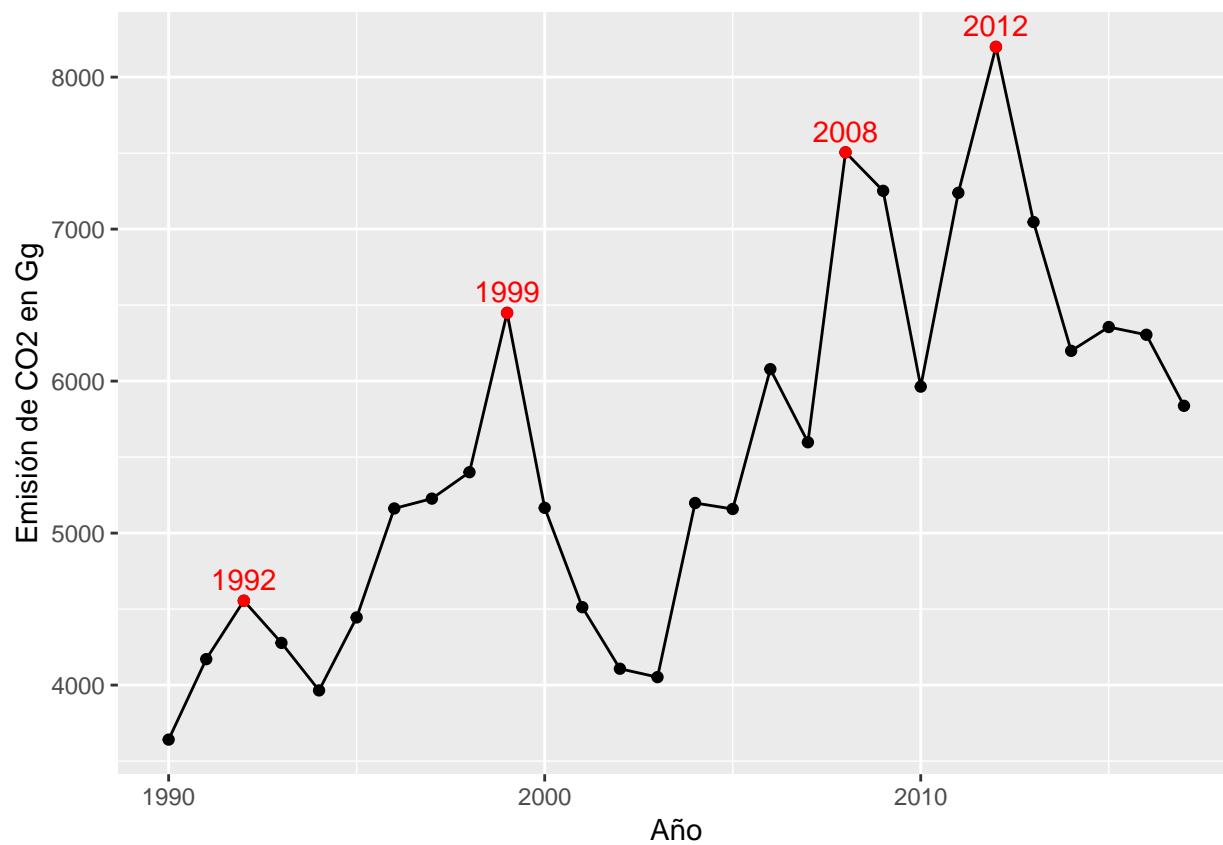
Con los datos `dato_emision` respondé las siguientes preguntas:

3. Usando las funciones de la librería `dplyr` obtenga qué fuentes tienen la emisión máxima. Recuerde que TOTAL debería ser excluido para esta respuesta así como los subtotales.
4. ¿En qué año se dió la emisión máxima para la fuente que respondió en la pregunta anterior?
5. Usando las funciones de la librería `dplyr` obtenga las 5 fuentes, sin incluir TOTAL ni subtotales, qué tienen un valor medio de emisión a lo largo de todos los años más grandes.
6. Usando `ggplot2` realice un gráfico de las emisiones a lo largo de los años para cada fuente. Utilice dos elementos geométricos, puntos y líneas. Selecciones para dibujar solamente las 5 fuentes que a lo largo de los años tienen una emisión media mayor que el resto (respuesta de la pregunta 5). Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un `caption` en la figura con algún comentario de interés que describa el gráfico.

7. Remplice el siguiente gráfico usando `ggplot2`. Incluir un `caption` en la figura con algún comentario de interés que describa el gráfico.



6. Usando la librería `ggplot2` y `ggpmisc` replique el siguiente gráfico de las emisiones totales entre 1990 y 2016. Los puntos rojos indican los máximos locales o picos de emisión de CO2 en Gg. Use `library(help = ggpmisc)` para ver todas las funciones de la librería `ggpmisc` e identificar cual o cuales necesita para replicar el gráfico. Incluir un `caption` en la figura con algún comentario de interés que describa el gráfico.



Ejercicio 2

Los datos que vamos a utilizar en este ejercicio están disponibles en el catálogo de datos abiertos uruguay <https://catalogodatos.gub.uy>.

Los datos que seleccioné son sobre los gastos realizados por actos médicos Los datos y los metadatos se encuentran disponibles en:

<https://catalogodatos.gub.uy/dataset/gasto-por-actos-medicos-fondo-nacional-de-recursos/resource/cae06492-d355-4ab0-9b42-79e0cca87def>

Por defecto los datos bajan en formato `xlsx`, usar el paquete `readxl` para leer los datos. Este ejercicio tiene como objetivo que realice un análisis exploratorio de datos utilizando todo lo aprendido en el curso. Debe contener al menos 5 preguntas orientadoras y visualizaciones apropiadas para responderlas. La exploración deberá contener las preguntas a responder sus respuestas con el correspondiente resumen de información o visualización. Incluya en su exploración el análisis de la variabilidad tanto de variables cuantitativas como cualitativas y covariaciones entre las mismas. Recuerde que en las visualizaciones, las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un `caption` en la figura con algún comentario de interés que describa el gráfico.