R para contextos humanitarios de emergencia

Lectura de datos

Violeta Roizman

Lectura de datos

En general, el primer paso de nuestro flujo de trabajo es importar los datos a R

Una vez importados los datos, podremos:

- Modificarlos
- Graficarlos
- Modelarlos
- Comunicar nuestros resultados

Los tipos de archivos de almacenamiento de datos más comunes son:

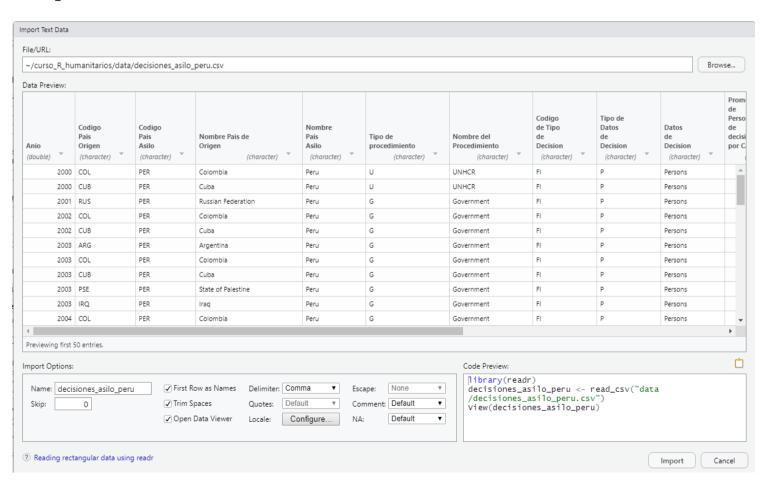
- Excel
- CSV (Comma Separated Values)

Tambien existen

- json
- XML, etc.

Lectura de datos CSV

File > Import Dataset > From Text...(readr)



Lectura de datos CSV - Opciones

- 1. Seleccionamos el archivo deseado
- 2. Si es CSV, debe decir Delimiter: Comma
- 3. También prestar atención donde dice First Row as Names
- 4. Si hay filas que hay que saltear (que no sean los nombres), podemos usar Skip
- 5. RStudio adivina el tipo de la columna en base a sus primeros valores, podemos cambiar el tipo si está equivocado
- 6. Importante!: Cuando terminamos de elegir las opciones, copiamos el código que se encuentra en el rectángulo Code Preview
- 7. Hacer click en importar

Una vez que importamos datos vamos a tener disponible un dataset del tipo tibble en la mesada

Lectura de datos

La última línea de código que copiamos es

```
View(nombre_dataset)
```

Ese comando sirve para visualizar las filas y columnas del dataset. También si corremos

```
nombre_dataset
```

veremos en la consola (o justo debajo del bloque de código si es un RMarkdown) algunas pocas filas del dataset.

El comando str (nombre_dataset) devuelve una descripción general del dataset. Algunos campos que nos devuelve el comando son:

- Tipo de dato (tibble)
- Tamaño ([filas x columnas])
- Nombre y tipo de las columnas, etc.

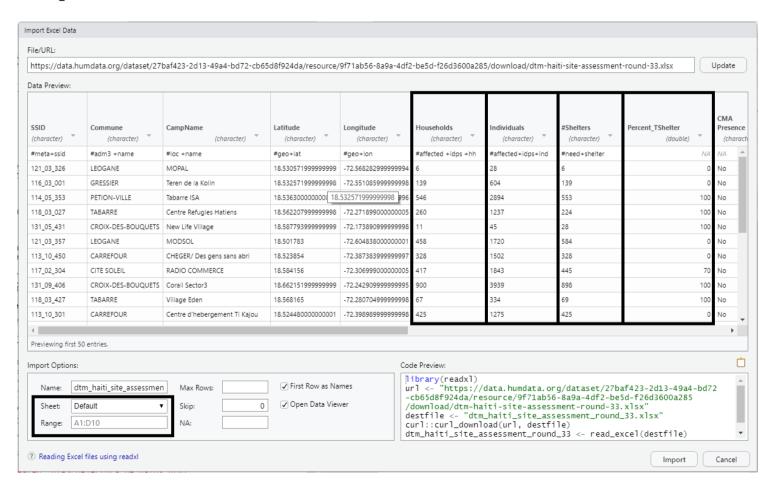
Posibles tipos de columnas: texto (character), números (double, numeric, integer), fechas (Date, DateTime), lógicos (logical), categóricas (factor), etc.

Tu turno: Lectura de datos CSV

- 1. importar el archivo "decisiones_asilo_peru.csv" (Fuente: UNHCR) con información sobre las decisiones tomadas en Perú frente a los pedidos de asilo recibidos provenientes de distintos países
- 2. Indicar que cantidad de filas y columnas tiene

Lectura de datos de Excel

File > Import Dataset > From Excel



Lectura de datos de Excel

• El paquete utilizado es readxl

Importante!

- Número de página donde se encuentra el dataset
- Rango de celdas donde se encuentra

Desafío 1: Lectura de datos

DESAFÍO 1: Lectura de datos

El objetivo de este desafío es que armes un proyecto y leas una archivo de datos para aplicar lo que aprendas en el resto del curso.

- 1. Crea un proyecto de RStudio llamado datos-haiti
- 2. Crea un archivo de RMarkdown (por las dudas, ésta es la sección asociada).
- 3. Arma un chunk para leer los datos. Importa el archivo disponible en https://data.humdata.org/dataset/27baf423-2d13-49a4-bd72-cb65d8f924da/resource/9f71ab56-8a9a-4df2-be5d-f26d3600a285/download/dtm-haiti-site-assessment-round-33.xlsx con información sobre distintos campamentos en Haití (Fuente: OIM).
- 4. Además de poner un título informativo, describe los datos con tus palabras. ¿Cuál es la fuente? ¿Qué variables incluyen? ¿Cuántas observaciones tiene? ¿Qué tipo de datos tiene cada columna?

Observaciones sobre el desafío:

- Notar que se puede importar directo usando el url
- Observa que la segunda fila no debería aparecer como una observación del dataset. En el futuro veremos cómo quitarla
- Esta fila arruina la detección del tipo de todas las columnas numéricas
- Arreglar el tipo de las columnas

Escritura de datos

Les presento las cheatsheets o guías rápidas (en español) 🕏

https://rstudio.com/resources/cheatsheets/

Nos describen brevemente la mayoria de las posibilidades que nos brinda un paquete $\stackrel{\hbox{\scriptsize th}}{\boxplus}$

Escritura de datos

Con readr puedo escribir archivos csv o de texto

Tu turno: Con la ayuda de la guías rápida de readr guardar el dataset de haiti con un nombre distinto en la carpeta data

Para escribir archivos excel puedes usar el paquete xls

Escritura de datos

En este modulo aprendimos a

- leer datos en R
 - en formato csv
 - en formato .xls o .xlsx
- escribir datos en R

Licencia y material usado

Licencia: Creative Commons Attribution-ShareAlike 4.0 International License.

Este material está inspirado y utiliza explicaciones de:

- R para Clima de Paola Corrales y Elio Campitelli
- Master the Tidyverse de Garrett Grolemund