

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)

Import Text Data

File/URL:
~/curso_R_humanitarios/data/decisiones_asilo_peru.csv Browse...

Data Preview:

Anio (double)	Codigo Pais Origen (character)	Codigo Pais Asilo (character)	Nombre Pais de Origen (character)	Nombre Pais Asilo (character)	Tipo de procedimiento (character)	Nombre del Procedimiento (character)	Codigo de Tipo de Decision (character)	Tipo de Datos de Decision (character)	Datos de Decision (character)	Prom de Perso de decisi por C
2000	COL	PER	Colombia	Peru	U	UNHCR	FI	P	Persons	
2000	CUB	PER	Cuba	Peru	U	UNHCR	FI	P	Persons	
2001	RUS	PER	Russian Federation	Peru	G	Government	FI	P	Persons	
2002	COL	PER	Colombia	Peru	G	Government	FI	P	Persons	
2002	CUB	PER	Cuba	Peru	G	Government	FI	P	Persons	
2003	ARG	PER	Argentina	Peru	G	Government	FI	P	Persons	
2003	COL	PER	Colombia	Peru	G	Government	FI	P	Persons	
2003	CUB	PER	Cuba	Peru	G	Government	FI	P	Persons	
2003	PSE	PER	State of Palestine	Peru	G	Government	FI	P	Persons	
2003	IRQ	PER	Iraq	Peru	G	Government	FI	P	Persons	
2004	COL	PER	Colombia	Peru	G	Government	FI	P	Persons	

Previewing first 50 entries.

Import Options:

Name: ☒ First Row as Names Delimiter: Escape:
Skip: ☒ Trim Spaces Quotes: Comment:
☒ Open Data Viewer Locale: NA:

Code Preview:

```
library(readr)
decisiones_asilo_peru <- read_csv("data/decisiones_asilo_peru.csv")
view(decisiones_asilo_peru)
```

? Reading rectangular data using readr

Import Cancel

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 saltar (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

Import Excel Data

File/URL:

Data Preview:

SSID (character)	Commune (character)	CampName (character)	Latitude (character)	Longitude (character)	Households (character)	Individuals (character)	#Shelters (character)	Percent_TShelter (double)	CMA Presence (character)
#meta+ssid	#adm3 +name	#loc +name	#geo+lat	#geo+lon	#affected +idps +hh	#affected+idps+ind	#need+shelter		
121_03_326	LEOGANE	MOPAL	18.530571999999999	-72.568282999999994	6	28	6		NA
116_03_001	GRESSIER	Teren de la Kolin	18.532571999999998	-72.551085999999998	139	604	139		0 No
114_05_353	PETION-VILLE	Tabarre ISA	18.536300000000000	18.532571999999998	546	2894	553		100 No
118_03_027	TABARRE	Centre Refugies Hatien	18.562207999999998	-72.271899000000005	260	1237	224		100 No
131_05_431	CROIX-DES-BOUQUETS	New Life Village	18.587793999999999	-72.173890999999998	11	45	28		100 No
121_03_357	LEOGANE	MODSOL	18.501783	-72.604838000000001	458	1720	584		0 No
113_10_450	CARREFOUR	CHEGER/ Des gens sans abri	18.523854	-72.387383999999997	328	1502	328		0 No
117_02_304	CITE SOLEIL	RADIO COMMERCE	18.584156	-72.306999000000005	417	1843	445		70 No
131_09_406	CROIX-DES-BOUQUETS	Corail Sector3	18.662151999999999	-72.242909999999995	900	3939	898		100 No
118_03_427	TABARRE	Village Eden	18.566165	-72.280704999999998	67	334	69		100 No
113_10_301	CARREFOUR	Centre d'hebergement Ti Kajou	18.524480000000001	-72.398989999999998	425	1275	425		0 No

Previewing first 50 entries.

Import Options:

Name: Max Rows:
☒ First Row as Names
Sheet: Skip:
Range: NA:
☒ Open Data Viewer

Code Preview:

```
library(readxl)
url <- "https://data.humdata.org/dataset/27baf423-2d13-49a4-bd72-cb65d8f924da/resource/9f71ab56-8a9a-4df2-be5d-f26d3600a285/download/dtm-haiti-site-assessment-round-33.xlsx"
destfile <- "dtm_haiti_site_assessment_round33.xlsx"
curl::curl_download(url, destfile)
dtm_haiti_site_assessment_round33 <- read_excel(destfile)
```

[Reading Excel files using readxl](#)

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 mayoría de las posibilidades que nos brinda un paquete
📁

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