

ORDENA TU LLUVIA: Una herramienta sencilla para ordenar los datos climáticos diarios descargados de AEMET (Agencia Española de Meteorología).

Ed. Nuria Naranjo Fernández y Héctor Aguilera

Editado en Málaga en octubre de 2022.

Objetivo

“Ordena tu lluvia” es una herramienta sencilla, construida bajo en código libre R, cuyo principal objetivo es facilitar la organización de los datos de precipitación descargados o solicitados a la Agencia Española de Meteorología (AEMET).

Motivación

Los datos de precipitación diaria solicitados a AEMET son entregados en un fichero/hoja de cálculo de manera que cada día del mes ocupa una columna diferente en una hoja de cálculo.

Los usuarios de estos datos, para su tratamiento o análisis, suelen organizar los datos con las fechas diarias de forma continua, y cada estación de AEMET ocupando una columna.

Mediante la herramienta “Ordena tu lluvia” se satisface la necesidad de realizar este cambio de formato con una inversión de tiempo mínima y un coste cero de recursos económicos.

Herramientas necesarias:

- Hoja de cálculo (tipo Excel o LibreCalc)
- R
- RStudio

¿Cómo son los datos de salida?

Los datos de salida habrán sufrido tres importantes transformaciones:

- Colocación de los datos en una columna para cada estación
- Cambio de unidades: los datos originales de AEMET vienen expresados en décimas de mm y los datos de salida ya están transformados a mm.
- Eliminación de los datos negativos: como la precipitación no puede ser negativa, se ha incluido en el script un bucle que transforma los datos negativos en cero.

Índice

0.	Prepara los datos de entrada.....	3
1.	Descarga e instala R y RStudio.....	3
2.	Descarga el script del repositorio.....	3
3.	Inicia RStudio y ejecuta el script.....	4
4.	Comprueba los datos de salida.....	4
5.	Transforma los datos de salida para poder tratarlos y explorarlos	4
	Referencias.....	5

0. Prepara los datos de entrada

- Crea un CSV en tu programa de Hojas de Cálculo (**leer * más abajo**) o descarga el archivo de muestra (ordenatulluvia_misdatos.csv)
- Copia y pega los datos originales según las siguientes columnas:

	A	B	C	D	E	F	G	H
1	INDICATIVO	ANO	MES	Dia1	Dia2	Dia3	Dia4	Dia5
2	Z_estacion1	1972	1	0	50	65	0	0
3	Z_estacion1	1972	2	100	0	15	110	150
4	Z_estacion1	1972	3	30	0	0	0	200
5	Z_estacion1	1972	4	0	0	0	0	0
6	Z_estacion1	1972	5	240	50	200	0	100
7	X_estacion2	1972	1	0	0	0	0	0
8	X_estacion2	1972	2	0	0	0	0	0
9	X_estacion2	1972	3	0	0	0	0	0
10	X_estacion2	1972	4	0	0	0	80	20
11	X_estacion2	1972	5	0	0	0	0	0

Columna INDICATIVO: el nombre de la estación no debe comenzar por un número. Es conveniente renombrar aquellas estaciones que comienzan su código por un número utilizando la herramienta de Excel “concatenar” (o la herramienta similar en el software que estes utilizando) y añadiendo alguna letra delante del nombre de la estación.

Columnas ANO y MES: se refiere al año y mes de medida.

Columnas de la D a la AH: contienen la información de la precipitación de AEMET de los días 1 a 31 del mes. (Nota: No modificar los datos originales, dejando los huecos en los días 29,30 y 31 para el mes de febrero; y día 31 para abril, junio, septiembre y noviembre.)

*Tener en cuenta que cuando utilizo R, el archivo CSV no debe contener ñ, espacios, (), nombres con acentos, ni cualquier otro símbolo que no sea compatible. Se recomienda limitarse a letras y números.

1. Descarga e instala R y RStudio

- Descarga e instala R: <https://cran.r-project.org/bin/windows/base/>
- Descarga e instala RStudio Desktop: <https://www.rstudio.com/products/rstudio/>

2. Descarga el script del repositorio

- Descarga en script del repositorio GitHub y guárdalo en un directorio conocido en tu equipo.
- Asegúrate que el script y los datos de entrada se encuentran en el mismo directorio.

3. Inicia RStudio y ejecuta el script

- Abre el script Ordenatulluvia: File → OpenFile y selecciona Ordenatulluvia.R
- Antes de ejecutar el script, asegúrate de cambiar en la **fila 19** el directorio del archivo de los datos de entrada: p.e. `setwd("C:/Users/miequipo/Documents/")`

* Para saber cuál es el directorio de la carpeta en la que están guardados los datos, búscala, haz clic con el botón derecho del ratón sobre la carpeta y selecciona Propiedades. Busca la Ubicación de la carpeta y copia y pega en la línea.

**Es importante comprobar que el separador de directorios es "/" y no "\" que es el separador por defecto.

- Ejecuta el script línea por línea, haciendo clic sobre Run



- Los datos de salida podrás encontrarlos en un archivo .csv en el directorio que has indicado en la **fila 12**.

4. Comprueba los datos de salida

Ten en cuenta que los datos de salida habrán sufrido tres importantes transformaciones:

- Colocación de los datos de filas a columnas
- Cambio de unidades: los datos originales de AEMET vienen expresados en décimas de mm y los datos de salida ya están transformados a mm.
- Eliminación de los datos negativos: como la precipitación no puede ser negativa, se ha incluido en el script un bucle que transforma los datos negativos en cero.

5. Transforma los datos de salida para poder tratarlos y explorarlos

- Si tu equipo está en castellano, para no tener problemas con el formato del separador decimal (punto o coma), ten en cuenta que puedes modificar el separador decimal en cualquier hoja de cálculo o procesador de textos (bloc de notas)

Referencias

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Wickham H (2007). “Reshaping Data with the reshape Package.” *Journal of Statistical Software*, 21(12), 1–20. <http://www.jstatsoft.org/v21/i12/>

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. DOI: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686)

Wickham H, François R, Henry L, Müller K (2022). dplyr: A Grammar of Data Manipulation. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>

Wickham H, Girlich M (2022). tidyr: Tidy Messy Data. <https://tidyr.tidyverse.org>, <https://github.com/tidyverse/tidyr>

Zeileis A, Grothendieck G (2005). “zoo: S3 Infrastructure for Regular and Irregular Time Series.” *Journal of Statistical Software*, 14(6), 1–27. DOI: [10.18637/jss.v014.i06](https://doi.org/10.18637/jss.v014.i06)