Data Management with R

M2 EIE, M2 I3E

Project: Climate change and natural disasters

Due date: 31 December 2025

Laurent R. Bergé*

This project is optional.

Objective. The objective of this project is to study the link between climate change and the frequence, and consequences, of natural disasters.

Groups. This project should be done in groups of at least 3 persons and maximum 5 persons.

AI. You are allowed to use AI. However: you must document at the beginning of the script how you did use it. Just write bullet points.

1. Expected output

This project has two components, each graded differently:

1. A data management task. The output of this part is an R document containing the code to create the data set for the analysis. **This script must be returned on Moodle.**
2. One web site which will display the results of the analysis. You will have a file **`website.txt`** which contains the link to your final website; **this file must be returned on Moodle**.

2. Descritpion of the data sets

You will use two data sets:

1. Meteorological data from Meteo France.[1] It reports data from 62 meteorological stations in France, with infra-day frequency.
2. Natural disaster data from the EM-DAT data set.[2]

*BSE, UMR CNRS 6060, University of Bordeaux, e-mail: laurent.berge@u-bordeaux.fr.

[1] See : https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32.

[2] See https://www.emdat.be/.

2.1. Meteo France

This data set is a subset from the Synop data base. It contains the following variables:

station_id	Identifier of the meteorological station
pressure_Pa	Pressure in Pascal
wind_strength_m_per_s	Wind strength in m/s
temperature_K	Temperature in Kelvin
humidity_pct	Humidity in %
year	Year of recording
month	Month of recording
day	Day of recording
hour	Hour of recording

2.2. EM-DAT data set

This is a subset from the main EM-DAT data set. It only contains information on natural disasters from Western Europe. You can find the documentation here. Only the names of the variables have been changed. Here are the modifications to the names:

- converted to lowercase,
- spaces have been turned into underscores (_),
- **'000US$** has been turned into **thousand_usd**,
- the commas have been deleted.

Only the names have been modified, not the data.

2.3. Access to the data

You can access these two data sets through moodle: https://moodle.u-bordeaux.fr/course/view.php?id=180385. You can load the data sets into **R** with the function **fread** from **data.table** (without any extra manipulation).

3. Data management: 12 points

The objective of this step is to create a data set linking meteorological data to natural disasters.

You need to end up with a data set containing **monthly** data for the following variables:

- year
- month
- average temperature of the month in Celsius degrees
- gap between the month's temperature (variable above) and the overall average temperatures for this month (across all years)
- total number of natural disasters in France for this month
- total number of damages, in dollars, in France for this month
- total number of deaths in France for this month

As you can see, you will need to aggregate the data at a monthly level and only consider natural disasters occurring in France (concerns the EM-DAT data set).

NOTA: in the EM-DAT data set, only use the date of start of the event.

Output. You should complete all these steps in a single **R** file whose name is **`NAME1_NAME2_etc.R`**. It should contain at the top of the document, in comments, the full names of all the authors.

Grading. Reaching this final data set requires to perform many small tasks. The validity of each of these tasks will be graded. Be sure the document is **clear** and easy to read, be sure to comment all the small steps you take.

4. Results: 8 points

You will present an analysis of the data. You will show evidence on the link between climate change and natural disasters. You will try to answer the questions: *Does climate change increase the occurence of natural disasters? Does climate change increase the damages from natural disasters?*

You will showcase your analysis on a website.

Important. In this part, the R code will not be evaluated: Do not place it in the same script as the data management. Only the content of the web site will be evaluated.

Output. You will have a file named **`website.txt`** which contains the link to your website, and you will upload this file on Moodle.

4.1. Creating a website

4 pts Q1. Create a web site hosted on Github, which will contain the results of your analysis.

4.1.1. Details

In this part, you need to create a webpage hosted on Github, with the following characteristics:

- it must contain at least 3 pages:
  1. one welcome page
  2. one page for the descriptive statistics
  3. one page for the data analysis
- the welcome page will contain, at least:
  - one welcome message
  - the first and last names of all the persons of the group
  - the links to the other two pages
  - of course it's better if it looks nice!

4.2. Descriptive statistics

2 pts Q2. On your website, create a page containing some descriptive statitics for the main variables.

This page should include:

- tables
- graphics (in particular displaying a temporal evolution)

4.3. Results

2 pts Q3. On your website, create a page containing the analysis of the link between climate change and natural disasters.

This page should include:

- tables
- graphics (eg bivariate graphs / related temporal evolutions, etc)
- econometrics

5. Resources

Thème	Ressource
data manipulation	• introduction to data.table by Laurent Bergé
create a web site with RStudio	• create web sites with Quarto, by Posit
Git (for Github)	• introduction to Git, by Peter Cottle • Git and GitHub Starter Guide by Dom Skinnion
publish a website on Github	• official documentation • Professional Website Starter Guide by IQSS • tons of online videos
graphics	• what's a nice graph? by Laurent Bergé • introduction to ggplot2 by Garrick Aden-Buie

5.1. Examples (last year projects)

- https://0joanne0.github.io/projectR/
- https://lambda-ck.github.io/R-project
- https://israel-ngarti.github.io/Site-web/
- https://ajaxrl.github.io/RSHA.ProjetR/

Those websites are linked to Github projects, so it's easy to see the underlying code that generated it.

Type "https://github.com/username" to get the Github page of **`username`**, then you can browse his/her projects – and you'll find the code that backed the web page. Example: "https://github.com/0joanne0".