

# Global river runoff and climate change

## Introduction

*Your team is a famous research group specializing in environmental data analytics. As the concerns about the impact of global warming to river runoff, the United Nations committee for global water resources comes to your aid. You are asked to present a report about the global changes in river runoff.*

## Objective

Your main objective is to present the global changes in river runoff for the years after 1990 CE. It is up to you to design the steps of your exploratory data analysis and answer the questions that will help you address this objective. Some examples of these questions can be:

- What is the difference of river runoff between 1961-1990 and 1991-2020?
- Which are the slopes in river runoff of the last 100 years?
- Are their differences between summer and winter?
- Are their differences per continent, latitude, or other geographic characterization?
- Does the elevation play a role? Does catchment size? Does something else?

Feel free to add your own questions!

## Data

To achieve this, you are provided with two datasets from the [Global Runoff Data Centre](#). The first consists of raw daily and monthly data, while the second of pre-processed average annual data (all units are in  $\text{m}^3/\text{s}$ ). You can use any of them you like or a combination of them. The datasets come in these files:

**grdc\_selection.csv** A ready-to-use curated dataset by [Markonis et al.](#) in tidy format. All records have at least 80 years of data until 2015, but some of them end earlier. Annual values of minimum (LQ), average (MQ), and maximum runoff (HQ) are provided, as well as a class for data quality (V.Qual; with A being the best and C being the worst quality class).

**grdc\_raw.tgz** The compressed dataset of the raw data at monthly and daily scale. Each file corresponds to a station for monthly or daily time step. Here, not all the records are needed. The stations that are crucial for the analysis are described in the next file.

**stations\_info.xlsx** A spreadsheet with all the stations that have at least 60 years of data, with records that end after 2010, catchment size above 10 thousand  $\text{km}^2$ , and missing values below 10%.

## Presentation and other information

Your code should be uploaded to github by 27<sup>th</sup> of May 2022, along with a markdown report. Please use a private repository up to that day. Your results will be also presented live in a 10-minute formal presentation in the week 30/5 to 3/6, which will be open to the audience. The exact date remains to be decided. After the presentation there will be time for questions by the evaluation committee, the other teams, and the audience.

The size of the team must be 3-4 persons, and it is not obligatory for all of them to participate in the presentation. Additionally, you can have two 30-minute consultations, one from Rajani and one from me, to discuss and get feedback for your work. Use it wisely to gain as much information as possible.

## Evaluation

The evaluation consists of two parts. Evaluation of your code/report and evaluation of your presentation. Among the evaluation criteria for the code/report will be the quality of the results, the correct methodology, cleanness of code, figure representativeness, and the ability to answer the main research question. This evaluation will be performed by Rajani and me. The evaluation of the presentations will be performed by a three-member committee consisting of a faculty member, a research associate, and a Ph.D. candidate. Their criteria will involve the quality of the presentation, the clarity of the results, the ability to answer the questions during the discussion.

**Good luck!**

