Visualization
Prof. Bernhard Schmitzer, Uni Göttingen, summer term 2024

Problem sheet 2

- *Submission by* ***2024-04-29*** *18:00 via StudIP* ***as a single PDF/ZIP****. Please combine all results into one PDF or archive. If you work in another format (markdown, jupyter notebooks), add a PDF converted version to your submission.*
- *Use Python 3 for the programming tasks as shown in the lecture. If you cannot install Python on your system, the GWDG jupyter server at* https://jupyter-cloud.gwdg.de/ *might help. Your submission should contain the final images as well as the code that was used to generate them.*
- *Work in groups of up to three. Clearly indicate names and enrollment numbers of all group members at the beginning of the submission.*

Exercise 2.1: algorithm runtimes.
The dataset stored in `runtimes.csv` contains information on the runtime of two algorithms (a serial version and a distributed version) on test problems of different size (measured in pixels), and for various numbers of worker threads (for the distributed version).

1. Create a chart that examines how fast runtime increases with problem size, for the single and distributed versions, and for different numbers of threads for the latter.
2. Create a chart that examines how the runtime depends on the number of threads for the distributed version. What would be the ideal dependency? The chart should allow to compare the ideal case with the actual observations.

Exercise 2.2: precipitation data.
In this exercise we process precipitation data of the Deutscher Wetterdienst (DWD). The original data is available at https://www.dwd.de/DE/leistungen/cdc/cdc_ueberblick-klimadaten.html, but all data required for the exercise is already provided in the zip file of the problem sheet.

1. The file `zehn_min_rr_Beschreibung_Stationen.txt` (as available on the DWD website) contains basic information about the weather measurement stations. Its format should be self-explanatory. Convert it into reasonable CSV format. As a warmup, create a scatter plot of the geographical position of all weather stations and their elevation.
2. The file `10min_processed.csv` contains condensed precipitation data for the day 2024-04-20 in intervals of 10 minutes for (a subset of) the weather stations listed above. The column `stationid` corresponds to the column `Stations_id` in the other table. The column `date` indicates the beginning of the 10 minute interval in the format YYYYMMDDHHMM. The column `rain` encodes precipitation in this interval in millimeters, missing values are encoded as -999.

   Compute the total precipitation at each weather station within each hour of the observed day. For one (or multiple) hours, show the precipitation of all weather stations in a scatter plot. Truncate missing values to zero.

3. The file `griddata.npz` contains the arrays `geolat`, `geolong` and `ind`. The two former arrays contain latitude and longitude positions of a regular rectangular Cartesian grid covering the measurement area, the third array contains an approximate binary indicator of the region covered by the measurements (with `True` indicating coverage). Interpolate the above precipitation data from the weather stations to this grid (only to the region indicated by `ind`) and display the data as an image. This gives a smoother visual impression of the geographical distribution of precipitation. Do this for all hours of the day as a small multiple.

   The resulting figure could look approximately like this: