Final Project:
MNXB11 - Introduction to Programming for Scientists

Shilpa Rani Sahu, Yuxin Lin, Sebastian Wiederkehr

October 26, 2023

1 Introduction

For this project, there were a variety of datasets provided. All of them contain recordings of air temperatures from cities in Sweden, the earliest dating back to the 18th century. The data was gathered by the Swedish Meteorological and Hydrological Institute (SMHI). It is publicly available and can be accessed through their website.
Out of this great pool of available data, we chose to examine the dataset for the weather station in Lund, as it felt the most relevant as we are experiencing the temperatures here firsthand. Furthermore, it has a large range, even when compared to the other datasets available, with its first entry dating back to the first of January 1780.

2 Research and Goals

As the supposed extent of the project was set by the instructions, we first decided on the three research questions we wanted to answer with our dataset on air temperatures in Lund. We came up with the following:

1. How much did the mean air temperatures on April the 1st vary in Lund since 1780?
2. What were the ,on average, warmest and coldest days in Lund since 1780?
3. Can we observe long-term trends in air temperature in Lund?

Especially the third question, arguably the most interesting one, gives further reason why we chose to use the data from Lunds weather station. Due to its proximity to the sea, it is an area with oceanic climate. This ensures minimal temperature fluctuations between the seasons and thus making it easier to observe long term trends as well as making the results more general.

3 Method

The datasets we were provided with still include Metadata, such as the exact location of the weather station, the different time periods in which the data was gathered as well as information about the quality of the data. Additionally, the data is portrayed

in a rather unusual format, using semicolons instead of commas to separate each entry. Since we need clean comma-separated values, as a first step we have to clean the data and replace the separators. Therefore we wrote a shell (bash) script, as this is usually the most straightforward tool for this kind of job. It is called *cleandata.sh*. As a first step, this bash script skips over the first 13 lines of the original datafile, as those are filled with the metadata described above. Afterwards, it iterates through each line; first making sure that the contents are valid data, and then extracting the information for *date*, *time* and *temperature*. Whilst doing this, it gets rid of additionally metadata stored next to the actual data in some lines. The output is then saved into a new file called *processed-data.csv*. In this step, we also leave out the data describing the quality of each observation.

4 Results

5 Complications

As you might have noticed, we were not able to reach our goals and carry out our planned analyses. This came as a result of multiple complications and problems we faced during our process, which we will now describe.

6 Conclusion