

01-Discharge of River Elbe: Data Management with LibreOffice

Thomas Petzoldt

2024-10-16

1 Introduction

Planning and maintenance of waterways and rivers needs adequate measurements and data. However, raw time series can often be long and confusing, so a first step is aggregation and visualization.

Scientific aim: plot an average year and calculate monthly averages, minima and maxima of the discharge.

2 Methods

The approach demonstrates the use of pivot tables for aggregating data according to certain criteria. This feature is available in spreadsheet software, e.g. Excel or LibreOffice. We will also use date and time computation to derive aggregation criteria from a single date column.

The data set consists of daily measurements for discharge of the Elbe River in Dresden (daily discharge sum in m^3d^{-1}). The data were kindly provided by the German Federal Institute for Hydrology (BfG)¹.

In order to simplify access, we will distribute an aggregated file during the course.

3 Exercises

3.1 Download the data set and inspect the data

1. Download the data set (**elbe_data.ods**) from the course home page and save it to a personal folder or your USB pendrive.

¹Data Source: Federal Waterways and Shipping Administration (WSV), provided by the Federal Institute for Hydrology (BfG).

2. Open it with LibreOffice **Calc**. Excel has a similar functionality, but details differ.
3. Make sure that you have set Calc to English language.
4. Inspect the data. You see that the date format has the form YYYY-MM-DD, that is the so-called “ISO 8601” date format, the international standard that makes data exchange between different software systems easier².

3.2 Creation of categories

In the following, we intend to aggregate the discharge data according to certain criteria, e.g. year and month. This can be done with the “pivot table” tool, so before we can do this, we need to create additional columns with the categories.

3.2.1 Date computations

Create the following categorical columns using formulas for date computation:

- year:= YEAR(A2)
- month: = MONTH(A2)
- day: = DAY(A2)
- weekday: = WEEKDAY(A2)
- doy: = A2 - DATE(YEAR(A2), 1, 1) + 1

The last formula computes the “day of year” (doy), also called Julian day. Here DATE(YEAR(A2), 1, 1) creates a date for the first January of the respective year and then the difference (+1) between the respective day and the corresponding 1st January. The formula respects the different length of months automatically, including 29th February in leap years.

Then fill the formulas down the column until the end of the data column.

Note: The formulas above assume that you use LibreOffice with English user interface. If you use another language (e.g. German) or other program (e.g. Excel), then the keywords and delimiters (semi-colon instead of comma) of the formulas may be different and you have to look up for them in the function library.

3.2.2 Pivot tables

In LibreOffice pivot tables are created like follows:

1. Select the data range for which the pivot table is to be created (including header!),
2. In the menu select **Insert, Pivot Table**
3. Select Source – Current selection – ok
4. Now drag the appropriate items to the fields, e.g. “year” to the Column field, “month” to Row field and “discharge” to Data field.

²For details, see http://en.wikipedia.org/wiki/ISO_dates

5. Double click on “discharge” and change Function to “Average”.
6. OK

That’s it, and you get the monthly average discharge values.

Task: Repeat the same for individual years, and for all years to extract minimum and maximum discharge and find a way to show the results graphically.

3.2.3 Average year and seasonality

Create a pivot table with “year” as Column fields, “doy” as Row fields and “Mean discharge” (i.e. Average) as Data fields.

Plot all years as function of the day of year.

Then create the some of following plots (you may need additional pivot tables):

1. a bar chart for annual discharge sums (y=annual discharge, x=year)
2. a bar chart with average discharge for the 12 months.
3. XY (Scatter) chart for all years like example before, and in addition average discharge for all observed years as thick line. Note: here it may be necessary to copy the numbers only from the pivot table to a separate sheet (Paste special – numbers) to remove the pivot table automatism for the graphics.
4. XY (Scatter) chart with confidence band. Calculate maximum, average and minimum per doyear over all years 3 lines: y = discharge min / average / max, x = doyear.

Now interpret the results. What was 2002? Google for “Elbe river 2002”.

3.2.4 Cumulative sum plot³

Cumulative sum plots of rainfall, discharge or temperature are useful for reservoir managers, or to classify years whether they are dry, wet, cold or warm.

Create a cumulative sum plot for each year by adding the discharge data (Q) as follows:

$$\begin{aligned}
 Q_{sum,1} &= Q_1 \\
 Q_{sum,2} &= Q_1 + Q_2 \\
 Q_{sum,3} &= Q_1 + Q_2 + Q_3 \\
 Q_{sum,n} &= Q_1 + Q_2 + Q_3 + \dots + Q_n
 \end{aligned}$$

Answer the following questions. Which year was:

³optional topic with higher difficulty

- the wettest,
- the driest,
- had a wet winter and a dry summer?

3.2.5 Additional ideas

The following ideas are intended as a stimulus for own explorations of the data and creative work. Feel free just to play around with downloaded data or develop your own project.

- Repeat the analysis with an additional [elbe.csv](#)⁴ file with more years.
- Download data from other measurement stations, e.g.
 - from <http://www.fgg-elbe.de/elbe-datenportal.html> or
 - U.S. Geological Survey <http://waterdata.usgs.gov/>

and analyse the data.

4 Discussion

Think about the hydrological situation in middle Europe in general, and Dresden in particular:

- Why do high discharge and floods occur in winter and spring?
- When did the highest flood occur?
- Do research on the internet about this.

Select your favorite tasks from above and create a two-page document to describe your results.

⁴<https://github.com/tpetzoldt/datasets/blob/main/data/>