

Visualization

Prof. Bernhard Schmitzer, Uni Göttingen, summer term 2025

Problem sheet 7

- *Submission by 2025-07-02 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 7.1: Conditional distributions in census data.

The file `census.csv` contains a simplified dataset on German census data of 2022 (the column separation character is `;`). Each row represents one administrative unit at the level of *Gemeinde*. It contains four columns: `Name`, `Total`, `rel_foreign`, and `rel_women`, which indicate the name, total number of inhabitants, relative fraction of non-German national inhabitants, and relative fraction of women inhabitants for the administrative unit.

1. Import the dataset into Python.
2. Show the distribution of the number of inhabitants throughout all Gemeinden.
3. Show the distribution of the fraction of women, dependent on the number of inhabitants. For small Gemeinden this should be quite spread out, and it should stabilize for larger Gemeinden.
4. Show the distribution of the fraction of foreigners, dependent on the number of inhabitants.

Exercise 7.2: Visualizing a relational database as decorated graph.

Consider a simple relational database that represents an online newspaper. *Journalists* can author *articles* (for simplicity, each article will be written by precisely one author) and articles can be *assigned* to (multiple) *categories*. *Readers* can create accounts and write *comments* on articles, they can express *reactions* to other readers' comments (such as 'agree' or 'disagree'), and they can *follow* certain authors (to be automatically informed, when they publish a new article). Assume that each of the concepts above that were highlighted in *italics* is represented by a separate table, and that relations between the concepts is encoded by simple key/foreign key references.

1. For each of the above tables, except for the *follow* and *assignment* tables, list at least two examples of columns that these tables should have (beyond keys and foreign keys).
2. Draw a graph that represents the above database, in particular the table columns and reference relations. Possibly you can choose a separate visual representation for the relations encoded by the auxiliary tables *follow* and *assigned*. You can do this in any software you want, or with a simple hand drawing (scanned or on a tablet).