Visualization
Prof. Bernhard Schmitzer, Uni Göttingen, summer term 2022

Problem sheet 1

Submission by 2022-05-09 18:00 via StudIP as a single PDF. Work in groups of up to three. Clearly indicate names and matrikelnr of all group members at the beginning of the submission.

Exercise 1.1: Tufte and data-ink.
Apply the principle of data-ink by Edward Tufte to a statistical graphic of your choosing from a peer-reviewed scientific publication. Identify data and non-data ink. Try to be strict in the sense of Tufte.
Remark: It is not necessary to re-create the graphic with plotting software. A simple 'dissection' as shown in the lecture, using e.g. Paint or Gimp is fully sufficient.

Exercise 1.2: Visualizing the geography and economy of Countrystan.
In the attached file `data.csv` you can find some basic information about cities in the absurdly named fictional state of Countrystan. The first line is a comment, each subsequent line contains information about a city in Countrystan, giving its longitude, latitude, the fraction of the city's GDP associated with the agriculture sector, and finally its 'tier', with regular cities being tier 0, the three province capitals being tier 1, and the capital being tier 2. The fields are separated by a space. The capital is called 'Apolis', the three province capitals are called 'Bplace', 'Ctown' and 'Dingen' (in the order that they appear in the file).

1. Plot the locations of the cities, encode information about the agricultural GDP fraction, and their tier status.

2. If done properly, you should see a strong relation between the longitude and the agricultural GDP fraction. Try to visualize this relation more clearly in a second plot. Do the capital and province capitals play a special role in this relation? In one or two sentences, describe the structures that you observed.

Remark: The plots do not need to be visually perfect. Use your favorite software (or draw them by hand if you feel strongly that visualization is not tied to the computer). The exercise is about discovering and corroborating a story by showing appropriate data. There is no clear right answer, just give it your best shot.