



Exercises for Chapter 1: Introduction

1 EXERCISE 1

Consider a data-intensive application domain of your choice and from your daily experience. Possible examples are the medical/healthcare domain, following an education, doing online shopping, or browsing the internet. Consider next a data-related problem for this domain. For your dataset and problem, give three examples of quantitative questions, and 3 examples of qualitative questions, that visualization could address. Argue why the quantitative ones comply with the definition thereof given in Chapter 1, and why the qualitative ones cannot be (easily) made quantitative.

2 EXERCISE 2

Consider the domains and questions from Exercise 1. Give three examples of quantitative questions, different from the ones you may have given as answers for Exercise 1, which could be answered *without* using data visualization techniques, *e.g.* by using simple query techniques. Argue, for each of your example questions, why it can be answered without recurring to visualization techniques. What makes such types of questions different from the ones where data visualization seems to be indispensable?

3 EXERCISE 3

Give three examples of data visualization scenarios from your current experience. Consider, for instance, the domains listed in Exercise 1. For each scenario, outline a few typical ques-



tions that users would like to answer, and argue which relates to the “confirm the known” and which to the “discover the unknown” high-level goals of data visualization. Separately, describe, in your words, what would qualify as ‘insight’ for each scenario.

4 EXERCISE 4

One of the stated goals of data visualization is to enable users to find unknown insights, *e.g.* in terms of data patterns or correlations, present in a dataset. However, one could argue that this quest is strongly limited by the inherent design of visualization tools and techniques, which can (by construction) only show data patterns for which they have been designed. As such, many types of insight would be undiscovered unless we created *specific* visualization tools that are able to capture them. In turn, this would imply that we knew what we want to find in the data, and hence that there is little to discover as new insights. Explain whether and why you agree (or disagree) with the above reasoning.

5 EXERCISE 5

Data size is one of the main challenges that are posed to visualization tools and techniques. Consider two types of very large datasets: The entire internet traffic over a large geographical area (a country, or even the entire world) over one hour, and the entire set of transactions performed over all stock exchanges over one hour over a geographical area of the same size. Let us assume that, in terms of data bytes, the two datasets are of comparable size.

- Argue which of the two datasets would be more challenging to visualize, and why. Support your answer by thinking about typical questions which you would like to answer for such a dataset.
- Assume, next, that you have only a very small screen (smartphone-size) on which you are allowed to show your visualization. How would you simplify the data so that its depiction fits on such a screen, without being allowed to use zoom or pan techniques? Explain how much of the questions addressed at the previous sub-point can be still answered with this constraint, and what would get lost.



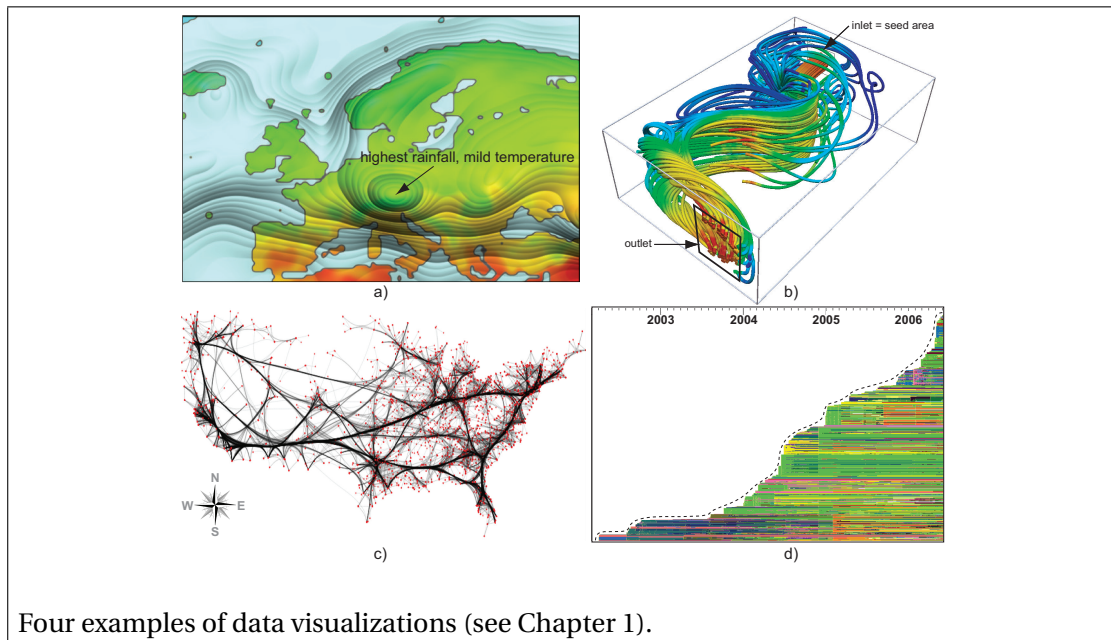
6 EXERCISE 6

One of the main challenges in data visualization is the size of datasets to be explored. Give an example of visualization application (from your current experience and/or daily life) where very large datasets are present. Describe the kind of data present in this application. Next, describe, in your words, why data size is a problem in this context (*Hints*: consider potential problems such as computational efficiency, or the limited size of display screens). Next, consider that we would like to address this data-size problem by reducing the amount of data to half. For your chosen example, describe a strategy to reduce the dataset size. In particular, answer the following questions

- How would you reduce the data size? Consider, for example, strategies such as throwing away data values, or averaging data values.
- What kinds of data-patterns does your data-size-reducing proposal keep? What patterns may we lose in this process?

7 EXERCISE 7

Figure 1.2 in Chapter 1 (also shown below) presents four examples of data visualizations taken from four different domains: Weather forecast, computational fluid dynamics, geographical systems, and software maintenance. Study the examples based on the material presented in Chapter 1– do not look for further information in the following chapters or elsewhere. After this, rank these four visualizations in terms of your difficulty to understand them, from easiest to hardest. For each visualization, describe briefly the reasons why understanding it has proven challenging.



Hints: Reasons for the understanding challenge may be e.g. difficulty of understanding the type of data involved in the problem; difficulty of understanding typical questions that the visualization should answer; difficulty to see how the visualization answers these questions; or difficulty in perceiving the elements of the displayed image (e.g. small-scale details, different colors or shades).

End of Exercises for Chapter 1: Introduction
