University of BRISTOL

Visual Analytics: Data Abstraction

Ian Nabney

ian.nabney@bristol.ac.uk

bristol.ac.uk

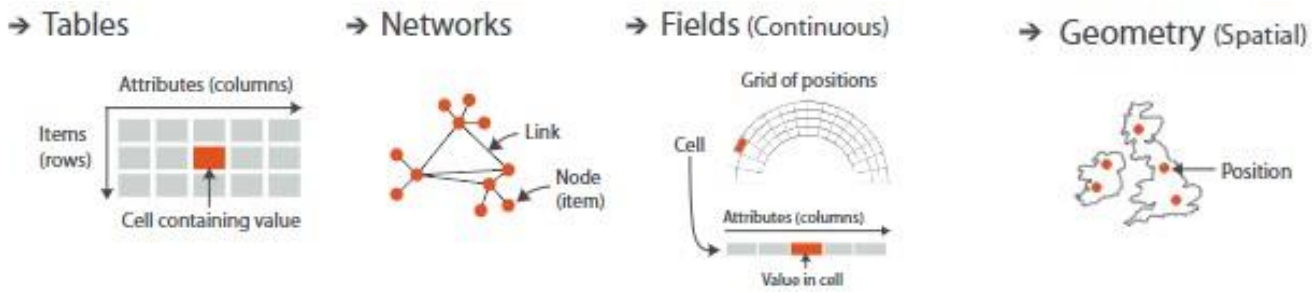
University of BRISTOL

Overview

- Reading: Chapter 2 of Munzner (e-book that can be downloaded from UoB library website)
- Understand what we are presenting: the categorisation of data
- Able to describe the type and semantics of variables and datasets

The Big Picture

- Three primary dataset types: tables, networks, spatial
- Four data types: items, attributes, links, positions
- Datasets may be available all at once as a static file or dynamically processed in the form of a stream
- Attributes (also known as variables or dimensions) may be nominal, ordinal, or continuous



Data semantics and types

- Measurement theory states that you should treat variables/attributes in a way that respects the meaning of the measurement ('semantics')
- **Nominal** attributes are distinct symbols (reasonably small number). The value serve as labels. The only comparison that makes sense is equality (two values are the same or different): e.g. colour
- **Ordinal** attributes are symbols with a meaningful order but no distance measure: e.g. 'hot', 'mild', 'cold'. Ordered comparison makes sense 'temperature > mild'
- **Interval** attributes are numeric and the size of the difference in values is meaningful: e.g. temperature on Celsius scale
- **Ratio** attributes are numeric and there is a well-defined zero point (so comparisons involving ratios make sense): e.g. weight

Dataset types

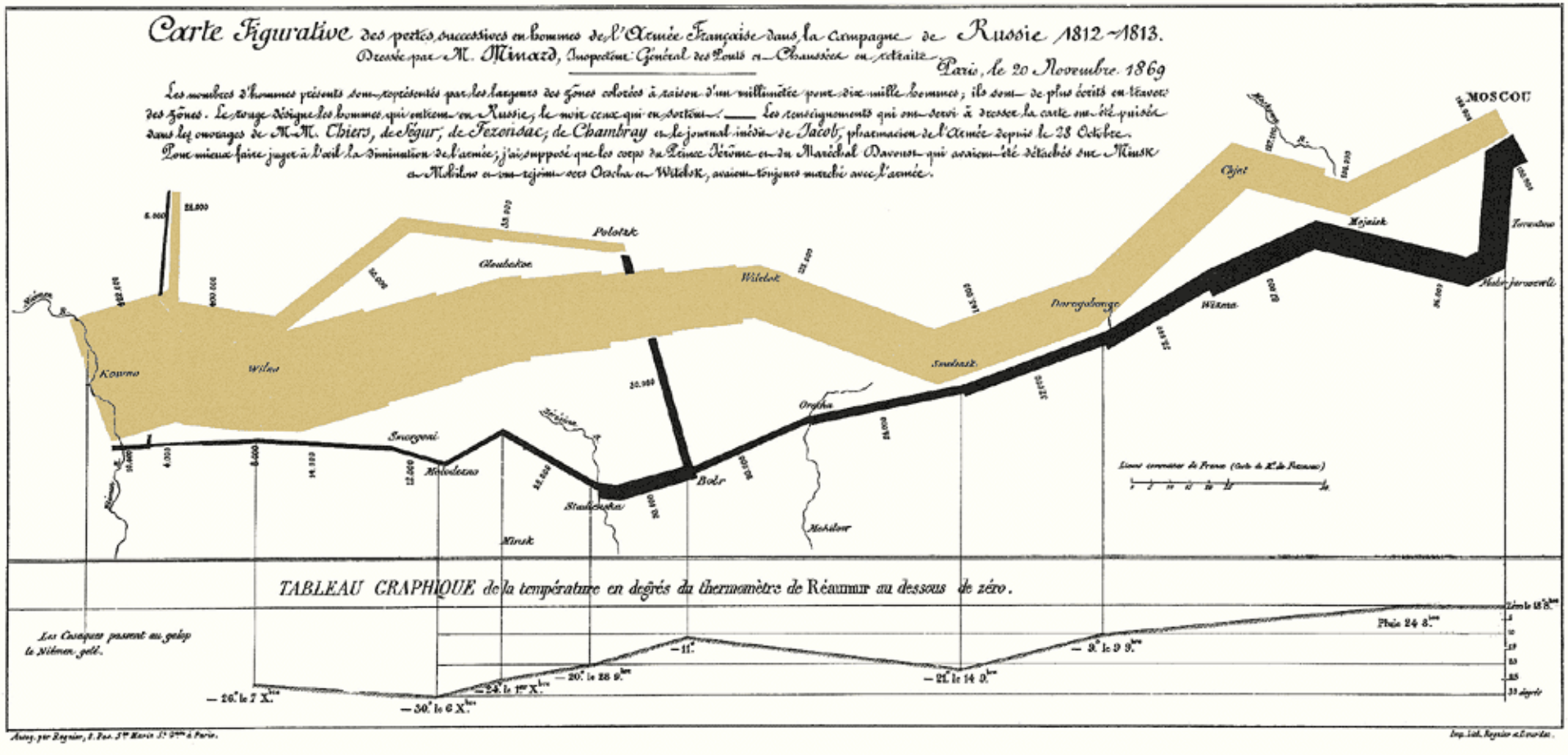
- For a simple **flat table**, each row represents an **item** of data, and each column is an **attribute** of the dataset.
- **Networks** are well suited for modelling relationships between two or more items. An item in a network is often called a **node**. A **link** is a relation between two items. Nodes may also have associated attributes. A **tree** has no cycles/loops.
- **Spatial** datasets store attribute values together with a location where they were measured. In regular grids, this location may be implicit.

- More complex datasets may well combine these types with other measurements (such as geometry, time series, etc.).

Dataset availability

- The default approach to vis assumes that the entire dataset is available all at once, as a **static** file.
- However, some datasets are instead **dynamic streams**, where the dataset information trickles in over the course of the vis session.
  - One kind of dynamic change is to add new items or delete previous items.
  - Another is to change the values of existing items.
- A dataset may be static in principle, but so large that it has to be accessed in a streamed form.

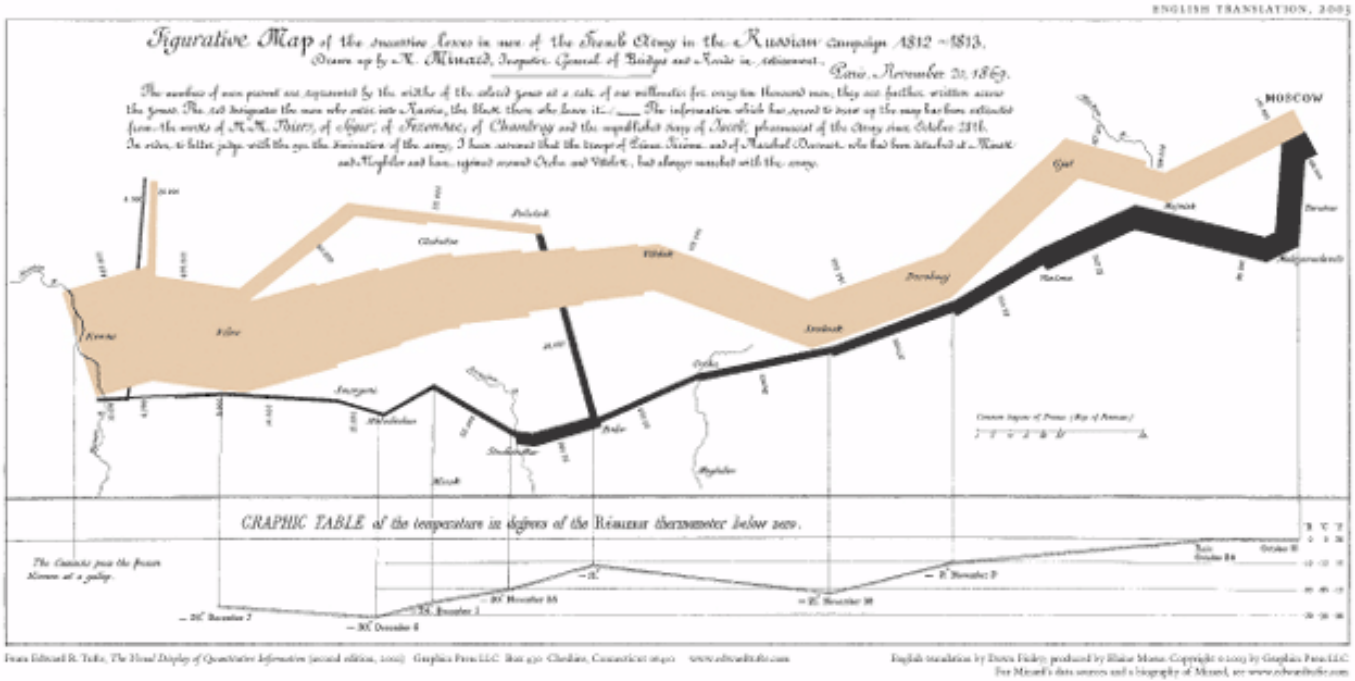
Minard the map maker



Map in English

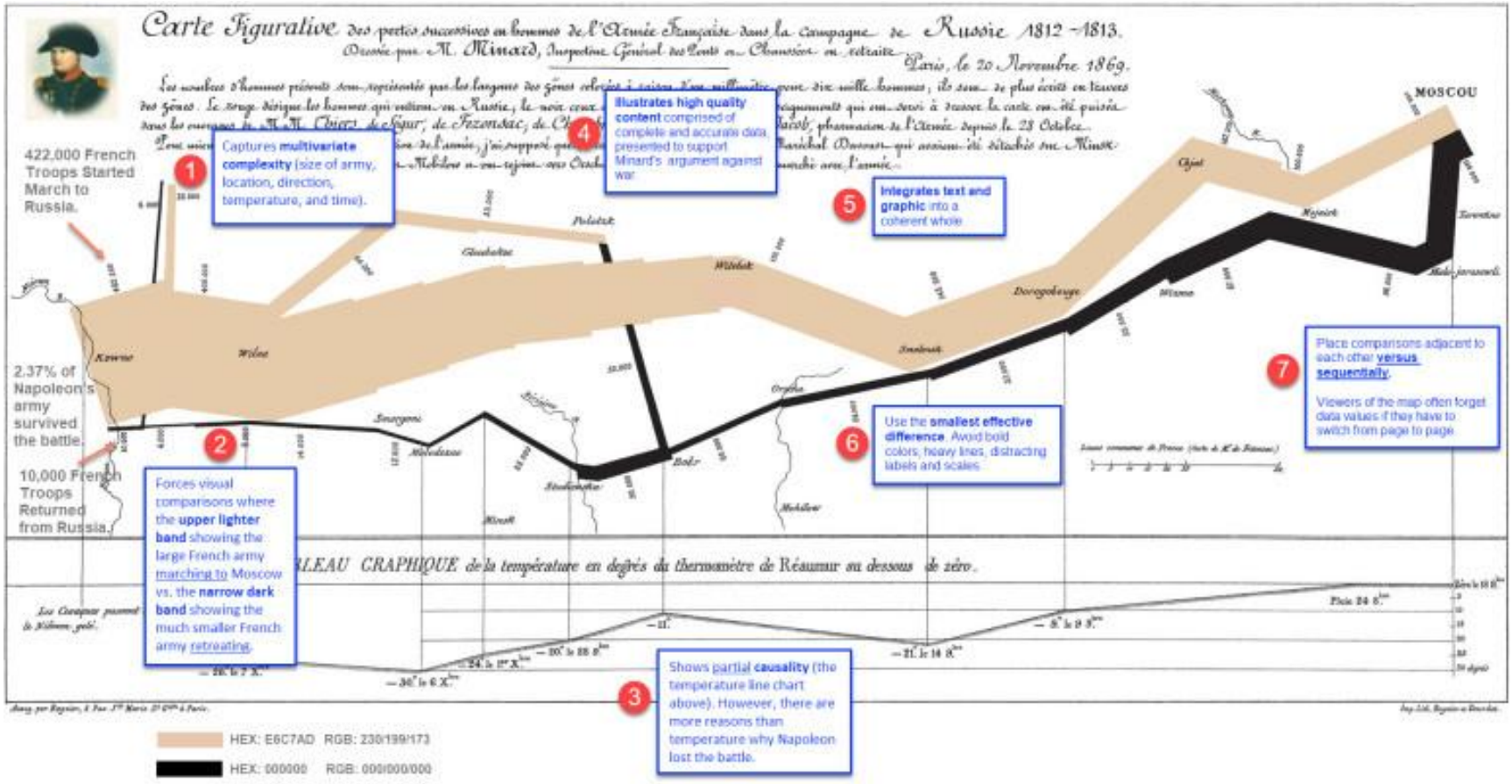
This classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally struggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. Minard does not mention Napoleon, the point of the graphic is to memorialize the deaths of the soldiers. It may well be the best statistical graphic ever drawn.

ENGLISH TRANSLATION, 2003



- Minard was a French civil engineer
- Map drawn in 1861
- Six types of variables shown in a single plot

Annotated map



https://datavizblog.com/2017/07/11/dataviz-as-history-annotating-minards-napoleon-map/