Introduction to Tables and Graphs

Alex Shpenev

June 7, 2017

Frequency Distribution

A table reporting the number of observations falling into each category of the variable

Example: US Population by Marital Status, mln.

	1990	2000	2005	2010
Total	181.8	201.8	217.2	229.1
Never married	40.4	48.2	53.9	61.5
Married	112.6	120.1	127.4	129.5
Widowed	13.8	13.7	13.8	14.3
Divorced	15.1	19.8	22.1	23.7

Source: U.S. Census Bureau, 2012 Statistical Abstract of the United States

Frequency Distribution

Why did the population of married individuals change?

- Likelihood of marriage and divorce changed
- Selection (more people survive, life expectancy longer)
- Total population increased

Proportions

To eliminate the effect of population size, we use proportions

$$p = \frac{f}{N}$$

where

- f - frequency in the category
- N - total number of cases

In popular literature, proportions are often multiplied by 100 to get percentages.

Percentage distribution

We can now construct a table showing the percentage of observations falling into each category

Age	Frequency	Proportion	Percentage
0	1	$1/25 = 0.04$	4 %
1	1	$1/25 = 0.04$	4 %
2	3	$3/25 = 0.12$	12 %
3	5	$5/25 = 0.2$	20 %
4	15	$15/25 = 0.6$	60 %

Cumulative frequency distribution

The distribution showing frequency at or below each categori of the variable

Age	Frequency	Proportion	Cumulative frequency
0	1	0.04	0.04
1	1	0.04	0.08
2	3	0.12	0.2
3	5	0.2	0.4
4	15	0.6	1

You can also calculate cumulative proportions and percentages

Frequency distributions and scales

The tables we just saw are best suited for nominal and ordinal data.

When our variable is interval-ratio, we might want to create groups (e.g. age groups 0 - 10, 11 - 20 etc.)

Rate

Frequency distributions are meant for observations taken at a given moment. If we are interested in events that happen over time, we might calculate rates.

Rates can be defined as frequencies of events over a given period of time.

Don't forget that "time" flies separately for each individual, so if 2 people live a year, they will together accumulate 2 years of time.

$$Rate = \frac{f}{Exposure}$$

$$Exposure = Population * Time$$

When population changes (e.g. due to births, deaths, migration etc., calculate the average population)

Graphs

Here's an example of a graph that shows 6 variables in 2 dimensions:

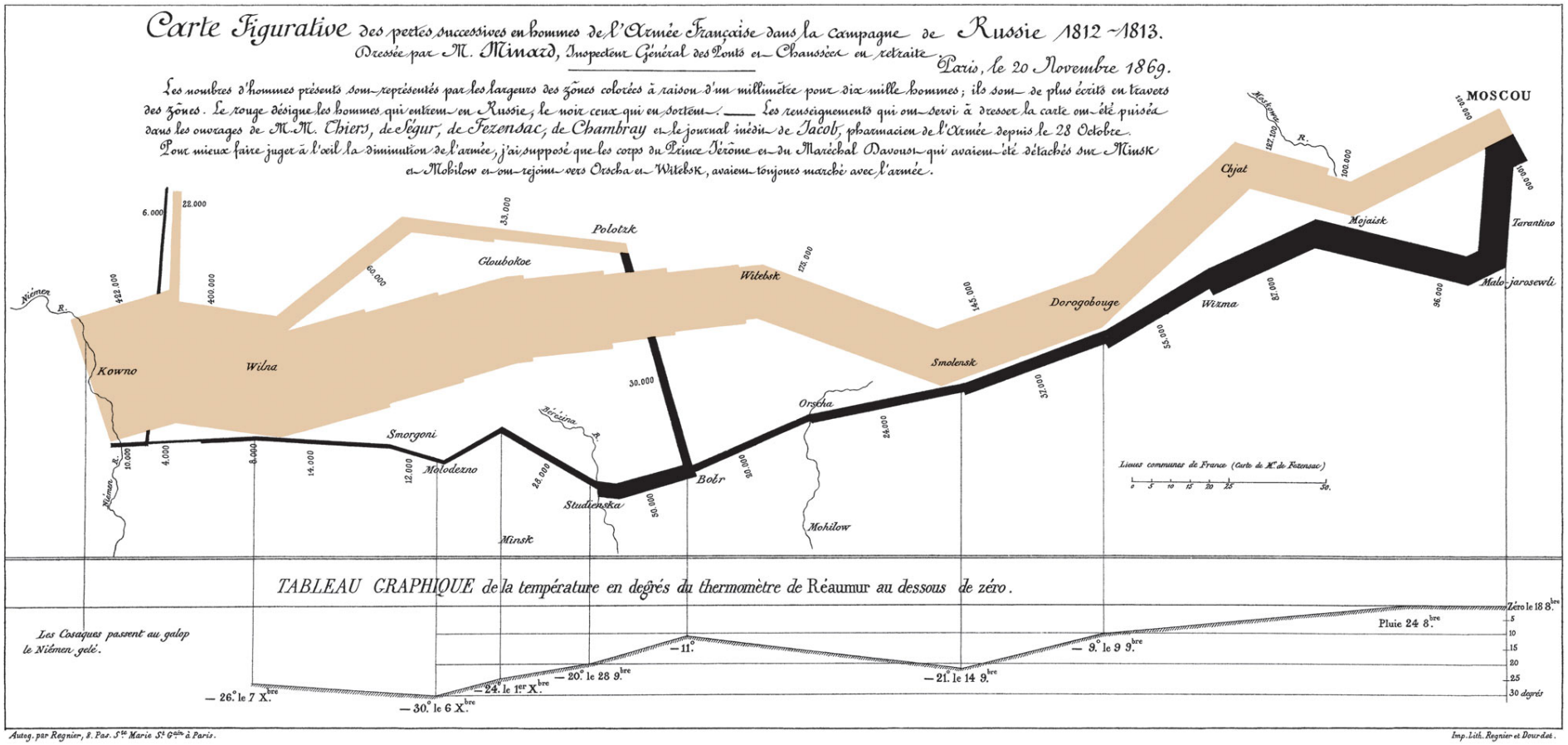

Figure 1: Famous graph by Charles Joseph Minard

Types of graphs

- Pie charts
- Bar charts
- Line charts
- Histograms
- Box plots
- Time series