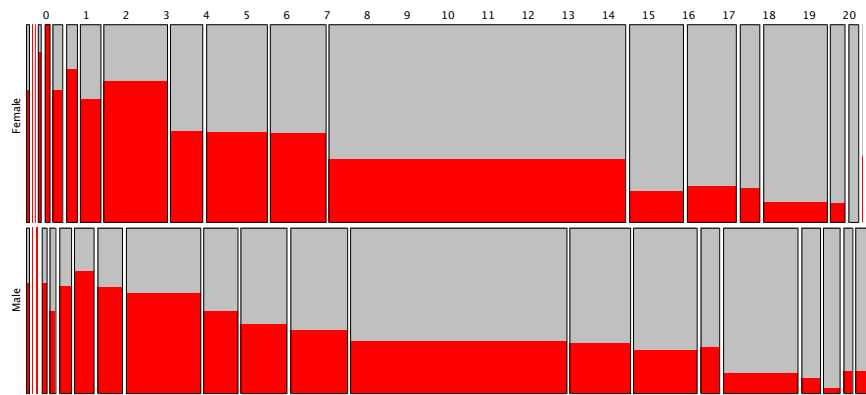
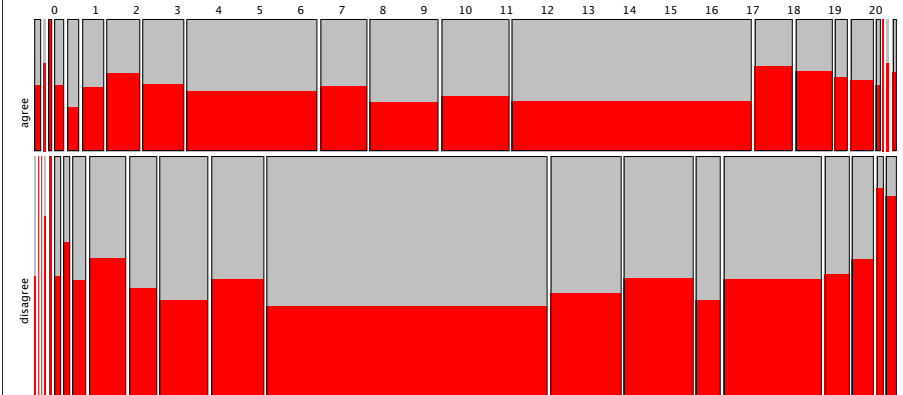


Womensrole — Agree by sex and education



Which probabilities are represented?

Womensrole — Sex by agree and education



Which probabilities are represented?

A Quickstart Guide (1)

Preliminaries

- Find out the background
 - what are the goals of the study?
 - what are the data? (numbers of variables and cases)
- Prepare the data
 - formatting, storing
 - merging and matching

A Quickstart Guide (2)

Variables

- What types of data?
 - nominal, ordinal, continuous, spatial,...
- What properties do the data have?
 - individual plots
 - range, categories, missings, outliers
- Assess the data quality and clean the data
- Check for consistency and plausibility

Data cleaning

- Never underestimate “the rawness of raw data” (Peter Huber)
- Header problems, lines too short/long
- Spaces, quotes, special symbols
- Errors, missing values, duplicates, special cases, outliers
- Numeric codes instead of readable text

Visualization helps for identifying and amending

A Quickstart Guide (3)

First analyses

- Use multiple basic plots together
 - barcharts, histograms, scatterplots,...
- Explore the plots using interaction
 - querying
 - linking
 - reformatting
- Summarise low dimensional properties of the data

Example: Crimean War

Florence Nightingale

- managed nursing the hospital in Scutari for British wounded from the Crimean War in the 1850s
- collected data on patient mortality
- attempted to influence British Government policy on treatment of the war wounded and on hospital management in general

Extract from Table B

Division and Corps		Average Strength	Total.				Remarks
			Admitted into Hospital	Died in the Crimea, &c.	Sent to Scutari, &c.	Died at Scutari, &c.	
2nd	30th Foot	522	934	108	308	93	99
Division	55th "	695	1,462	61	265	96	100
	62nd "	430	949	96	135	42	24
	95th "	417	1,250	199	345	155s	114
	41st "	684	1,323	104	320	94	81
	47th "	637	1,223	91	280	71	102
	49th "	655	1,071	66	274	90	89
3rd	1st "	771	1,048	229	354	118	63
Division	14th "	423	878	8	42	2	6 a)
	38th "	689	1,728	149	319	118	73

Scutari lies just across the Bosphorus from Istanbul



Case study: Web keywords

- 13765 records of 103 keywords by day
 - number of clicks
 - number of page impressions
 - sales
 - turnover (Umsatz)
- Data from January to August 2006
 - 243 days

A Quickstart Guide (4)

Multivariate Displays

- Choose subsets of variables relevant to the goals
- Draw multivariate plots
 - parallel coordinates, scatterplot matrices for continuous
 - mosaic plot variations for categorical variables
- Explore using interaction: sorting, zooming
- Identify any higher dimensional structure

A Quickstart Guide (5)

Grouping and selection

- Condition on grouping variables and compare subsets
 - boxplots by one grouping variable
 - small multiples, trellis displays
- Consider interesting selected subsets in isolation (drill-down)

A Quickstart Guide (6)

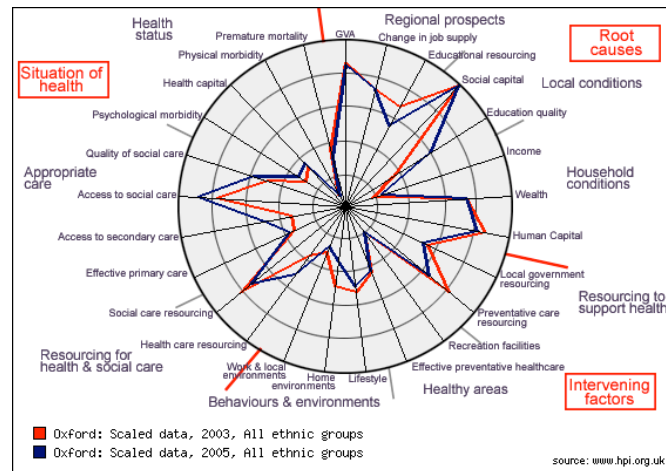
Special features

- Reduce dataset
 - aggregate data by groups or clusters (case reduction)
 - combine related variables (dimension reduction)
- Use any temporal structure (time series plots)
- Link to maps if there is spatial information

A Quickstart Guide (7)

Presenting results

- Choose graphics that convey the main conclusions
- Polish graphics
 - scales, order, size, aspect ratio, colours, layout
- Ensure that all graphics are fully explained
 - captions, legends, axes, annotations, text and headlines



Health Poverty Index: www.hpi.org.uk/