Training for compilers of statistical releases

Session 14
2 June 2016
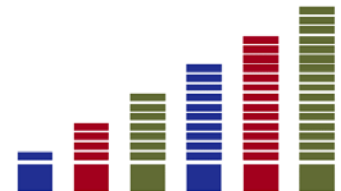
Kim Huuhko, Erja Seppänen
Statistics Finland
kim.huuhko@stat.fi,
erja.seppanen@stat.fi

Madrid, May 31 - June 3

Abstract

- At Statistics Finland, statistical experts produce statistical releases themselves and publish them online
- There are around 160 active statistics which produced around 550 releases in 2015
- During 2014 and 2015, Statistics Finland organised an extensive training round for statistical experts that covered all statistics
- The authors of statistical releases were invited to the training sessions as small groups and examples were used to review the recommendations for compiling statistics
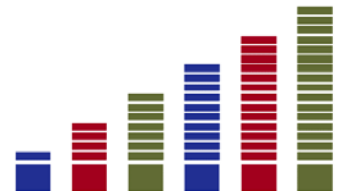
Background for the training sessions

Many changes in the data environment related to releases in recent years:

- New intranet
- The website's new graphic design -> narrower text column, new colours
- New colour palettes and graph templates to Excel and SAS
- New version of PX-Web -> richer metadata required

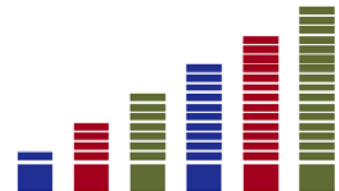-> New instructions and training sessions were needed!

Different release elements are answers for different needs!

- Graphs are a way to disseminate quick mental images of phenomena behind numbers
- Html tables are for further use of those exact numbers
- Database tables are for further processing of more extensive data

Correct graph types and correct presentation methods are essential for understandability and usability of statistical data!
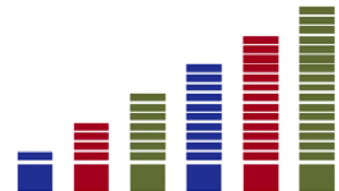
- Each graph type conveys a different view to data and answers to a different need
- Wrong graph types and presentation methods convey the wrong knowledge
    -> A bad graph is far worse than no graph at all!
- Positioning of variables and values in presentation tables radically affects the focus of the table

Role of the statistical experts

Experts were inspired to consider their role in conveying information:

- To summarise the most essential information from vast data masses
- Their insight can make statistics significantly more understandable to our non-professional customers
- Their expertise is needed to highlight the most interesting characteristics of the data
- Experts were provoked to consider critically the contents of their own releases
- All statistics have their own specific features -> global rules and recommendations cannot always be applied!
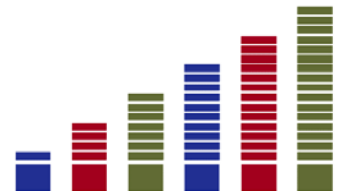
Changing data environment

- Decreasing resources
    -> automatic reprocessing methods are needed
- API solutions and machine processing are becoming more frequent in our statistical production
- Also, our customers use API solutions more often next to the direct use of the PX-Web user interface

-> Data have to be in a machine-readable format, which means highly standardised structures and metadata
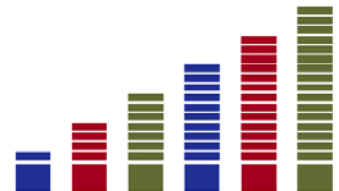   - Connectivity!
    - Stability!

Conclusions

Discoveries:
- Too many html tables in the releases
- Too much information in individual tables and graphs
    -> difficult to perceive the most relevant information

-> Focus of the releases should be slightly shifted from the static html tables to multi-useful database tables

Results:
- All the recent releases and their elements were reviewed and evaluated
- Useful for compilers as their own tables and graphs were used as examples
- Compilers consult "technical support" more often than before
- Collaboration in general has evolved significantly
- Reasoning behind recommendations was clarified to all
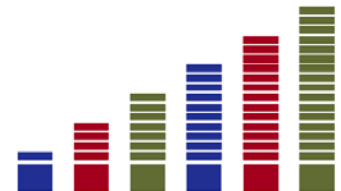- Many unnecessary misconceptions were removed

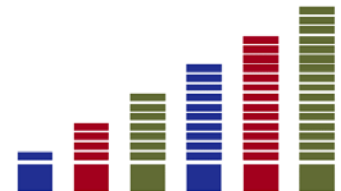# Text structure of a statistical release: heading

- The heading is the main element of the text: it makes the reader interested or uninterested in reading
- Search engines rank headings high
- A good heading of a statistical release:
  - Summarises the main topic of the release
  - Highlights the most important, interesting, topical issue: the biggest, smallest, new, continuous trend, etc.
  - Does not try to say everything
  - Includes an active verb

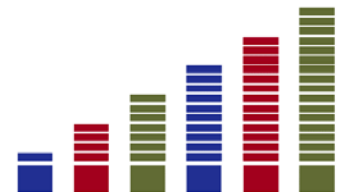Nearly every fifth woman aged 45 to 49 has not given birth

Structure of statistical releases: first paragraph

- Summarises the main issue of the news
- Complements the heading
- What is new, important, surprising, interesting
- Only one or two main points of the release
- Short, at most five lines

Good statistical text

- Things are presented in the order of importance, from the most important to the least important
- Use short, simple sentences: one sentence = one idea
- Active language: turn substantives into verbs, passive forms into active forms
- Make abstract expressions concrete
- Explain statistical terms
- Recommendation for web texts: one paragraph is at most five lines
- Various paragraph lengths give the text rhythm
- Add sub-headings every two to three paragraphs

Good statistical text

- Terms that describe figures and graphs are easily transferred to the text even if it was more descriptive to use other expressions
- Avoid:
  - "Elevator statistics": this went up, this went down
  - Listing figures shown in tables
  - Abbreviations