Statistically sound methods to turn time-series data into knowledge

Anders Wallgren & Britt Wallgren

Örebro University and Statistics Sweden
How do we turn statistics into knowledge?

Are there quality aspects here? Yes, there are three quality aspects here:

- Quality of Data
  - At NSI:s we think of quality of estimates one by one
    - not on patterns in data
    - not on the knowledge we give rise to

- Quality of Analysis
  - Find patterns in data
  - Distinguish between different patterns
  - Control for disturbing factors
  - Cause and effect
  - Control for confounding variables

- Quality of Presentation
  - Is it true what you say?
  - Have you forgotten something important?
  - Is it understandable?
  - Is it interesting?
Example 1: 
**Yearly data don’t describe changes within years**

Real case: Planned investments in manufacturing 89/90 + 5 %!

This was published 20 March 1990!

This could have been published!
Example 2: Quality of different sources

Murder/manslaughter in Sweden 1975-2005
Number of persons/cases per 100 000 inhabitants
1. Reported 2. Persons found guilty 3. Persons killed
Example 3: Number of persons depends on the size of the population – *a disturbing factor here!*

Employed persons in Sweden
45-54 years old, thousands

Employment ratio among persons
45-54 years old, %
Example 4: Is segregation in Gothenburg increasing?  

13 parts of the city of Gothenburg are compared:

Analysis: Disturbing factors must be eliminated/controlled!  
Remove the trends in both variables! More variables than two!  
Control for differences in age distribution!
Example 4: Is segregation in Gothenburg increasing?

Both variables are corrected for differences in age distributions.

Percent with university education compared with the entire Gothenburg:

Average income, index with the entire Gothenburg = 100.
Edward Tufte: *The Visual Display of Quantitative Information*
"Above all else show the data!"

William Cleveland: *The Elements of Graphing Data*
"Make the data stand out!"
Perception, Trellis Displays

Wallgren et al: *Graphing Statistics & Data*
How do we turn statistics into knowledge?

Conclusions:

Storytelling – is my message true?

- Do investments go up or down?
- Do murders go up or down?
- Does employment go up or down?
- Does segregation increase or not?

Misleading charts are dangerous – the effect of visualization can be very strong!
How do we turn statistics into knowledge?

We think that the Analysis part of the chain is the weakest!

There are plenty of good data
There are many good IT-tools for presentation
But we lack good analysis!
How do we turn *time-series data* into knowledge?

1. Let the users do it themselves
   - Difficult for most users
   - Risk for misleading charts

2. We do the analysis ourselves
   - “We” = a small team of experts that develop the methods for analysis and design the reports
   - Every month/quarter/year a time-series database is updated
   - Every month/... a large number of standardized reports are updated by the regular staff
Every month/quarter/year a time-series database is updated

Every month/... a large number of standardized reports are updated by the regular staff

What is typical here:

- Often a very large number of time-series
- Many similar time-series for a large number of groups/domains of study
- Same kind of analysis can be used at a large scale
Three examples (the first will be shown):

1. One manufacturing enterprise with 200 employed
   1 600 monthly time-series describing sales

2. Accident rates:
   300 municipalities, 3 age classes, 2 sexes and
   3 kinds of accidents
   5 400 yearly time-series

3. Labor Force Survey: about 15 000 monthly series
   Labor force participation for a large number of
   subpopulations by age, sex, region, education, ...
1. One manufacturing enterprise with 200 employed
1600 monthly time-series describing sales:
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customers</strong></td>
<td><strong>Data including 2009 period 1</strong></td>
<td></td>
</tr>
<tr>
<td>Price index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sort by level</td>
<td>Sort by level</td>
<td>Sort by level</td>
</tr>
<tr>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales, fixed prices, $ millions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sort by level</td>
<td>Sort by level</td>
<td>Sort by level</td>
</tr>
<tr>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profit margin %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sort by level</td>
<td>Sort by level</td>
<td>Sort by level</td>
</tr>
<tr>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
<td>Sort by rate of change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer 1</td>
<td>Customer 2</td>
<td>Customer 3</td>
</tr>
<tr>
<td>Customer 4</td>
<td>Customer 5</td>
<td>Customer 6</td>
</tr>
<tr>
<td>Customer 7</td>
<td>Customer 8</td>
<td>Customer 9</td>
</tr>
<tr>
<td>Customer 10</td>
<td>Customer 11</td>
<td>Customer 12</td>
</tr>
<tr>
<td>Customer 13</td>
<td>Customer 14</td>
<td>Customer 15</td>
</tr>
<tr>
<td>Customer 16</td>
<td>Customer 17</td>
<td>Customer 18</td>
</tr>
<tr>
<td>Other customers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Storytelling:
Relevant charts and tables together in a meaningful combination

Two “clics” to get this detailed information based on some of the 1 600 series