

Seminar on Innovative Approaches to Turn Statistics into Knowledge
15-16 July 2009, Washington DC

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:



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

Find patterns in data
Distinguish between different patterns

Control for disturbing factors
Cause and effect
Control for confounding variables

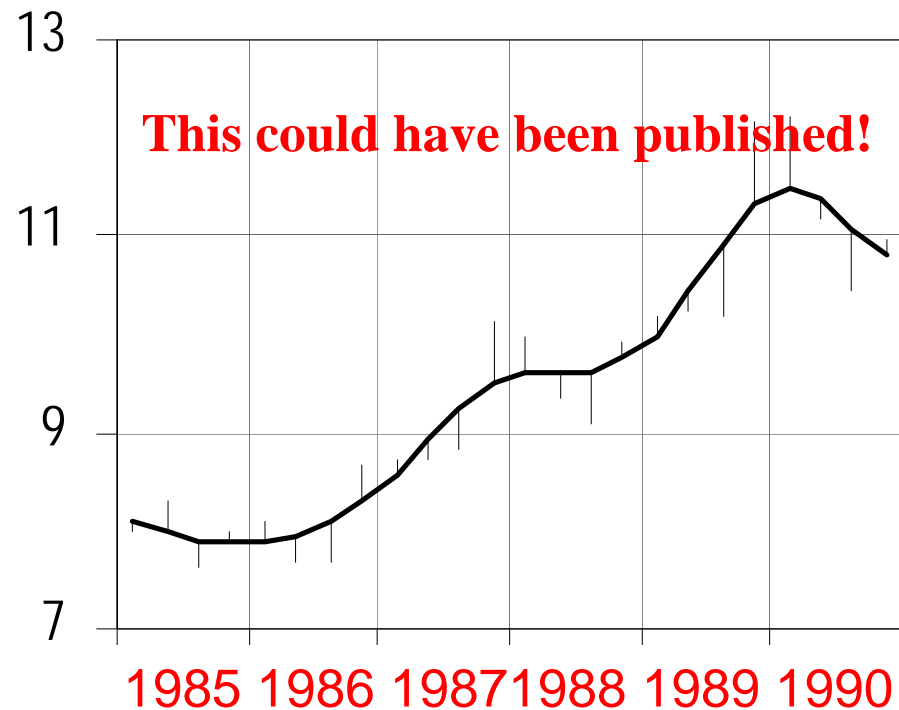
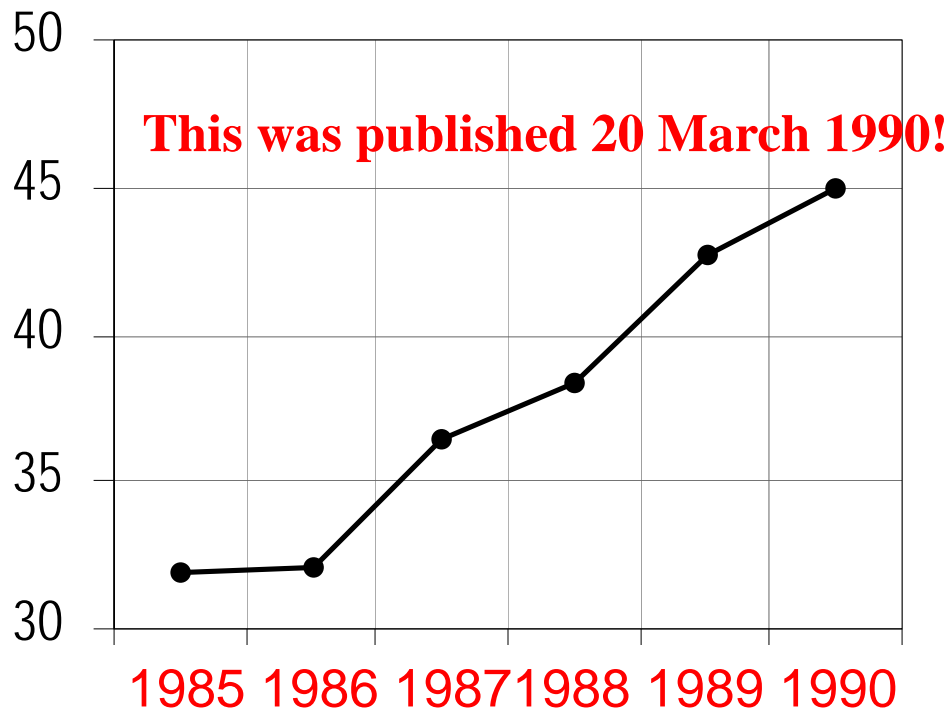
Is it true what you say?
Have you forgotten something important?
Is it understandable?
Is it interesting?

Quality of Data

Example 1:

Yearly data don't describe changes within years

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



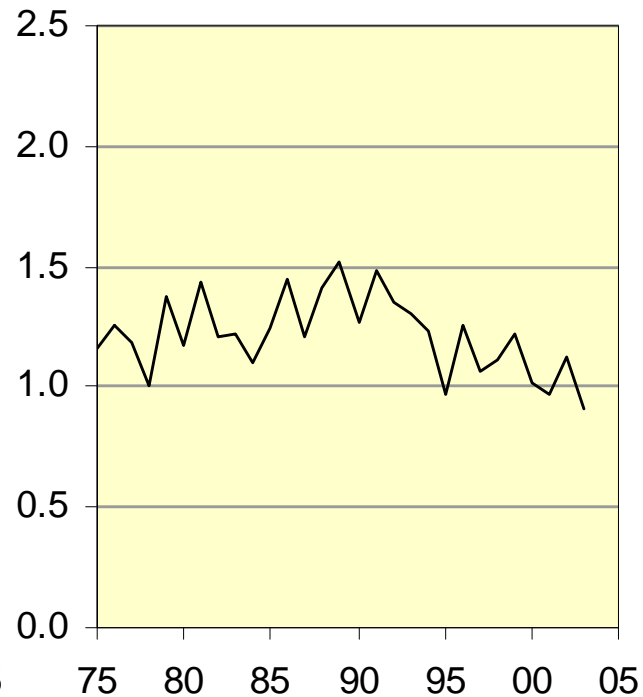
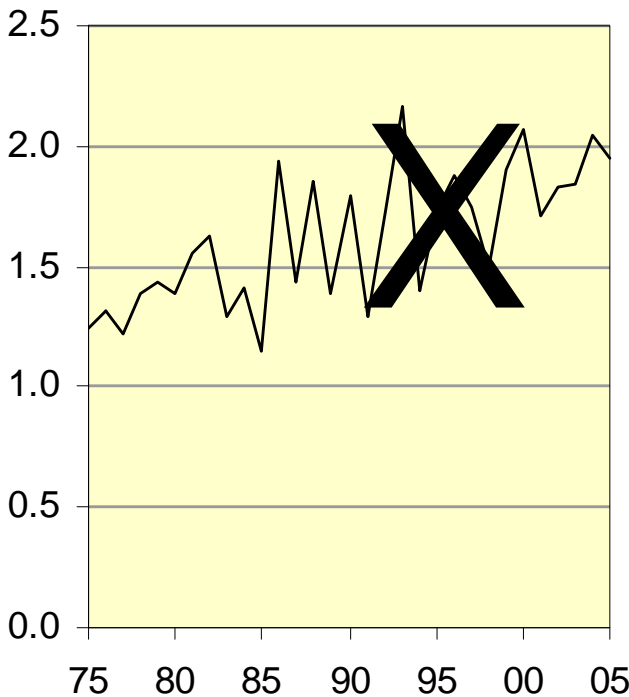
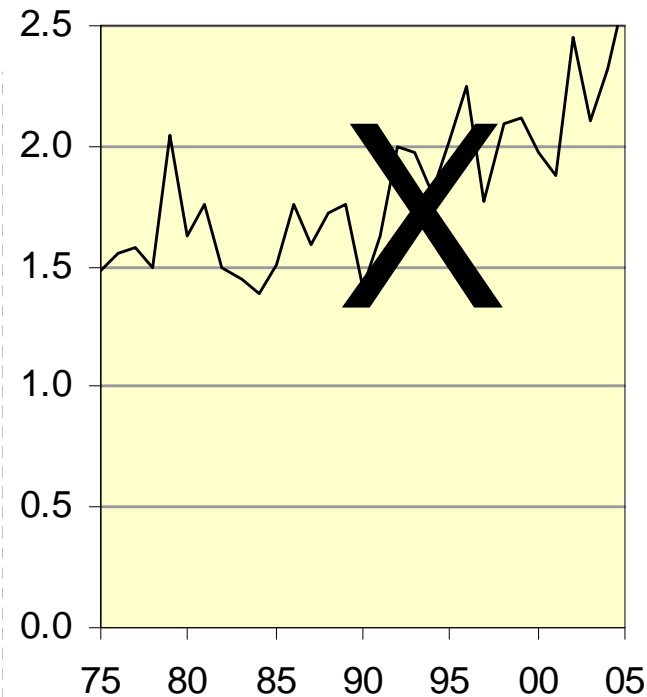
Quality of Data

Example 2: Quality of **different sources**

Murder/manslaughter in Sweden 1975-2005

Number of persons/cases per 100 000 inhabitants

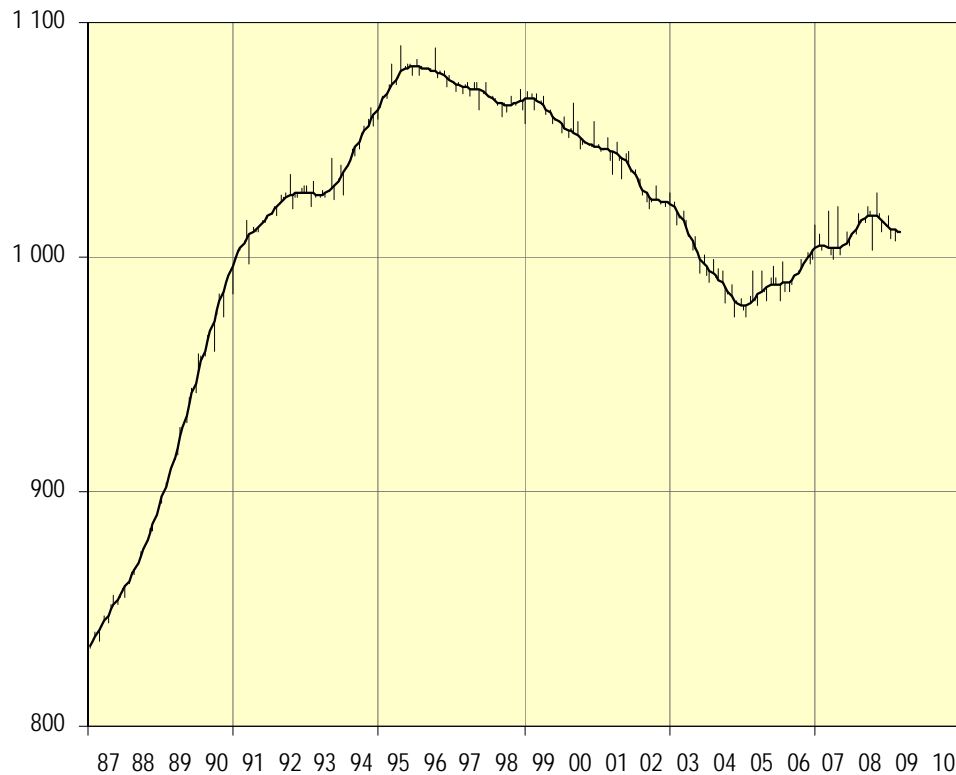
1. Reported 2. Persons found guilty 3. Persons killed



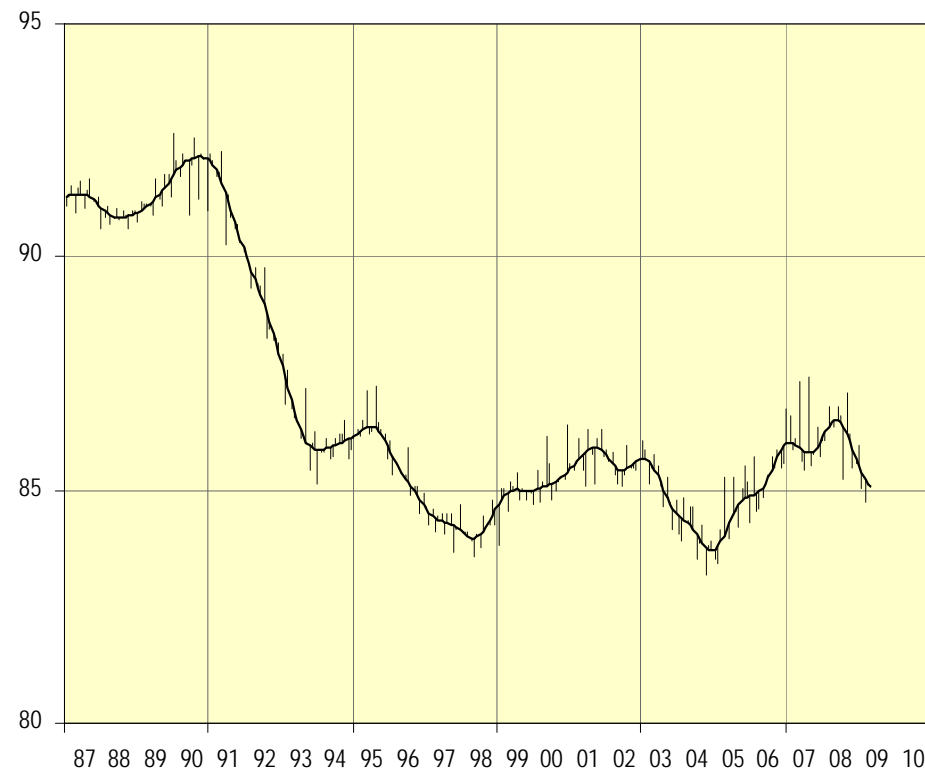
Quality of Analysis

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, %



Quality of Analysis

Example 4: Is segregation in Gothenburg increasing?

13 parts of the city of Gothenburg are compared:

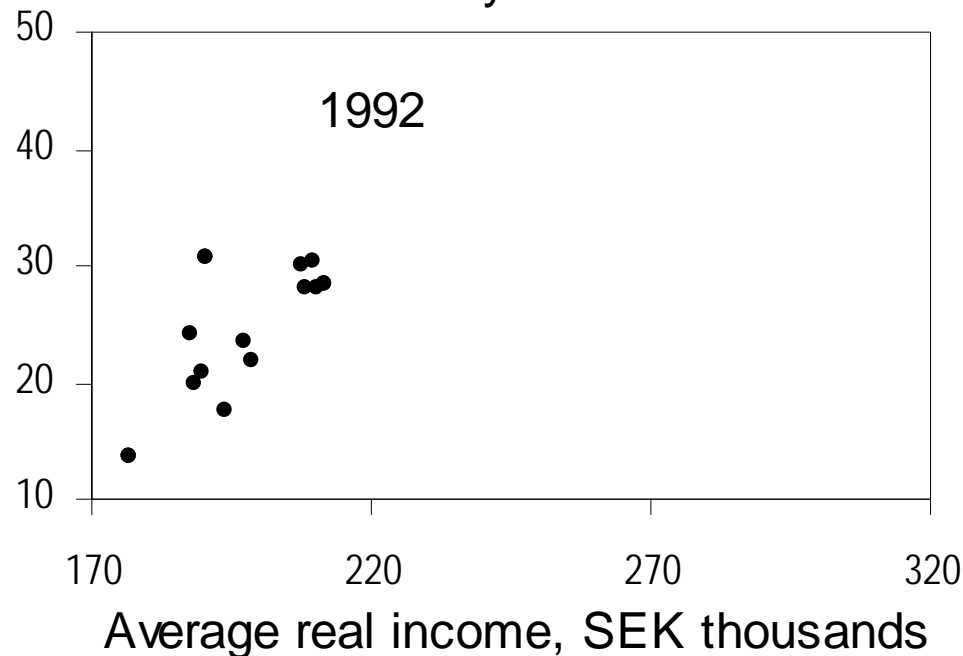
Analysis:

Disturbing factors must be eliminated/controlled!

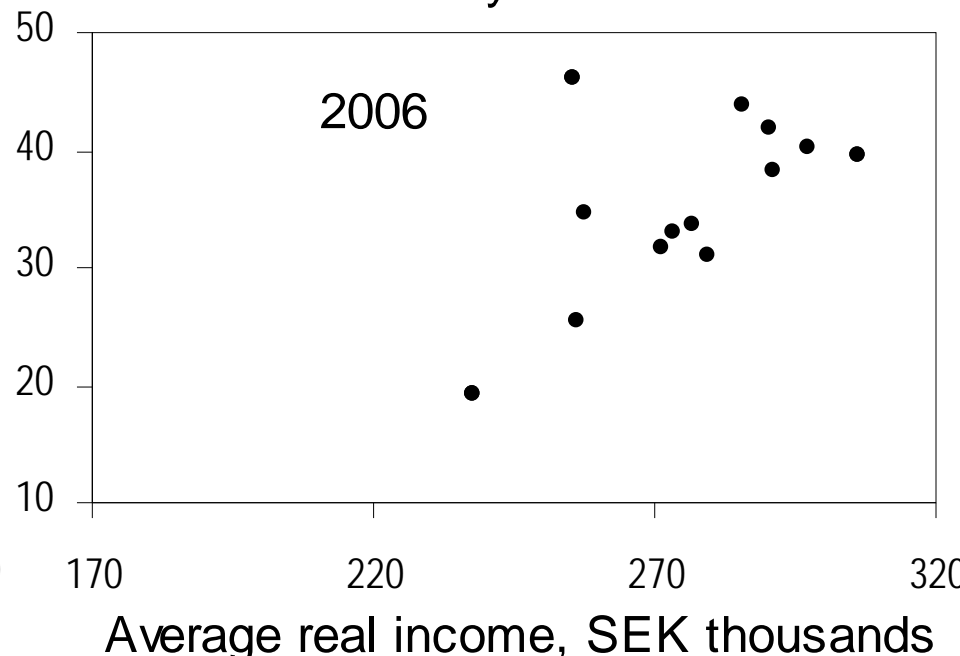
Remove the trends in both variables! More variables than two!

Control for differences in age distribution!

Percent with university education



Percent with university education

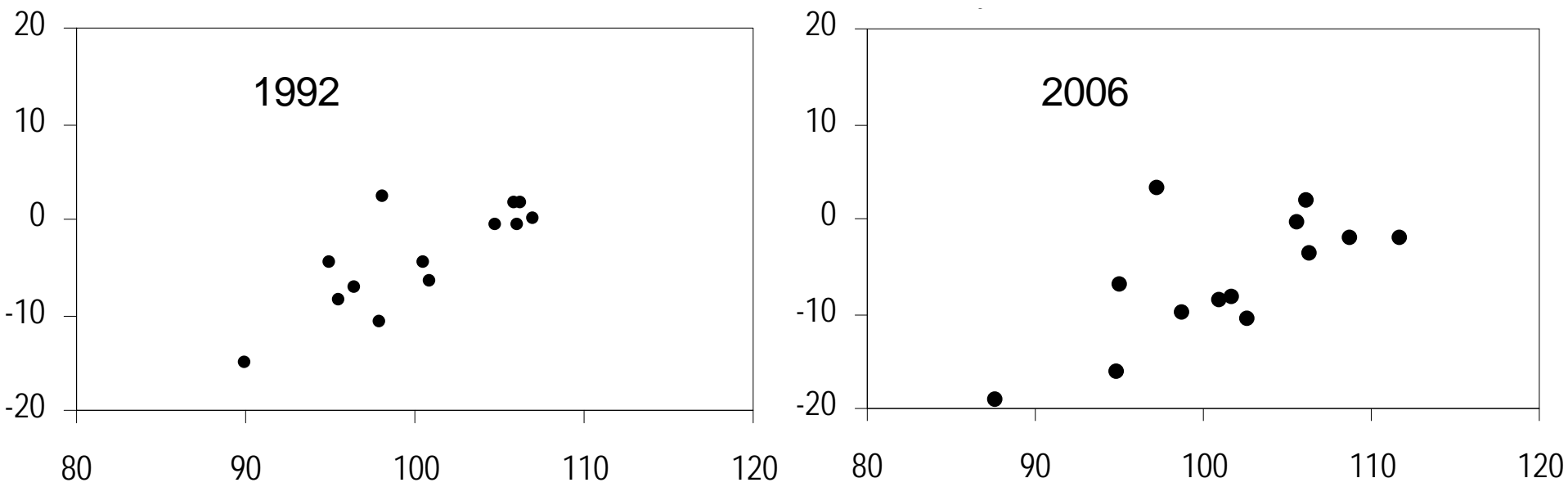


Quality of Analysis

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



Quality of
Presentation

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

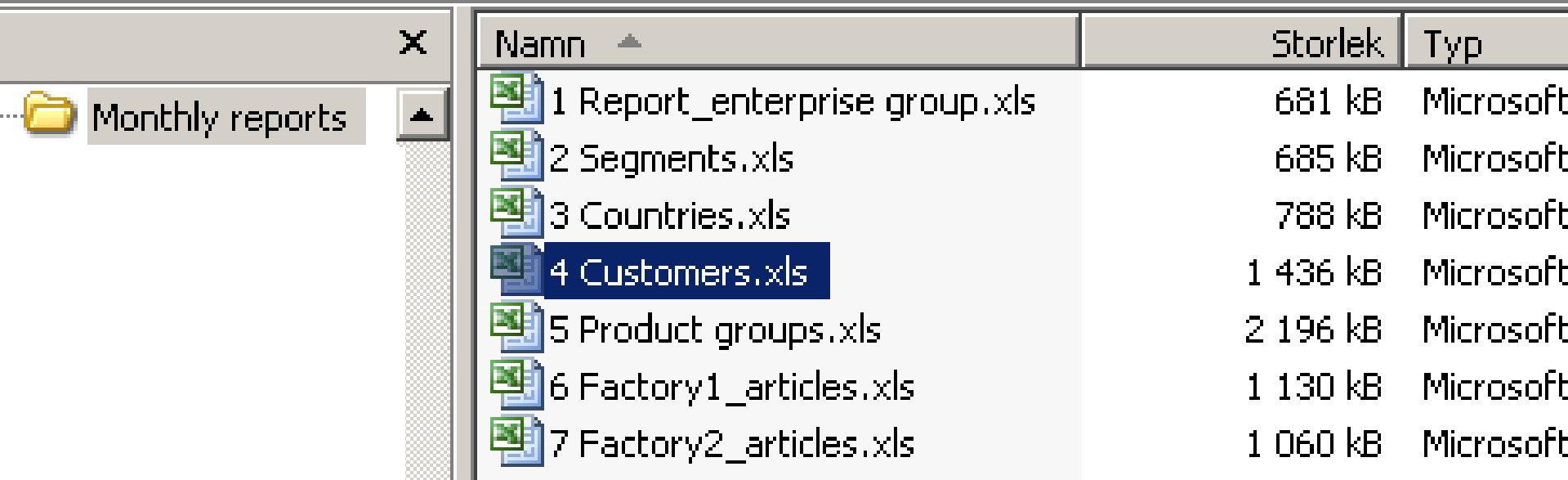
What is typical here:

- 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
1 600 monthly time-series describing sales:



Namn	Storlek	Typ
1 Report_enterprise group.xls	681 kB	Microsoft
2 Segments.xls	685 kB	Microsoft
3 Countries.xls	788 kB	Microsoft
4 Customers.xls	1 436 kB	Microsoft
5 Product groups.xls	2 196 kB	Microsoft
6 Factory1_articles.xls	1 130 kB	Microsoft
7 Factory2_articles.xls	1 060 kB	Microsoft

Customers **Data including 2009 period 1**

Price index

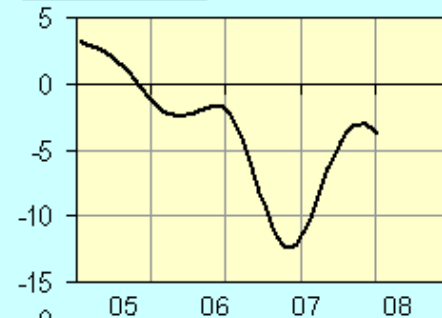
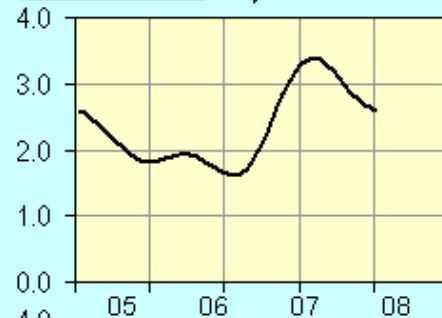
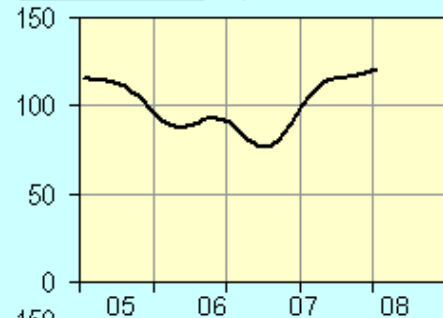
Sales, fixed prices, \$ millions

Profit margin %

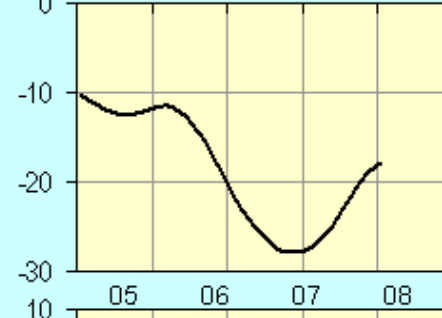
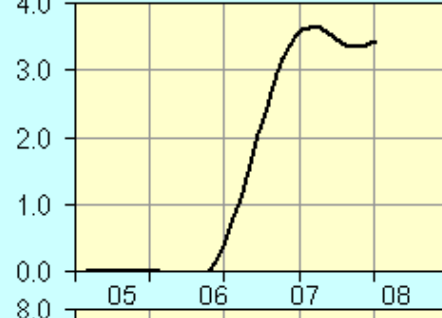
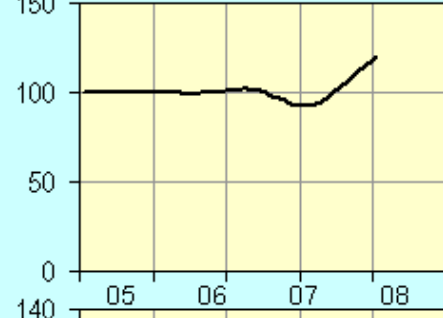
Sort by level
Sort by rate of change

Sort by level
Sort by rate of change

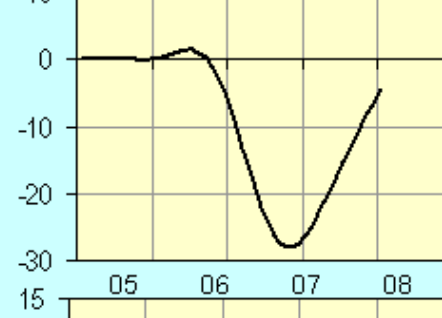
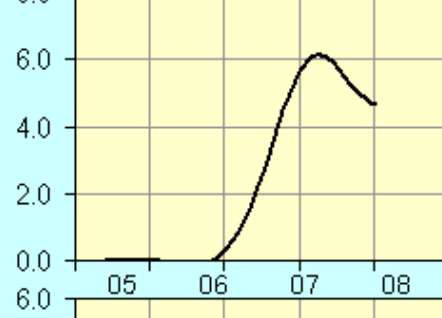
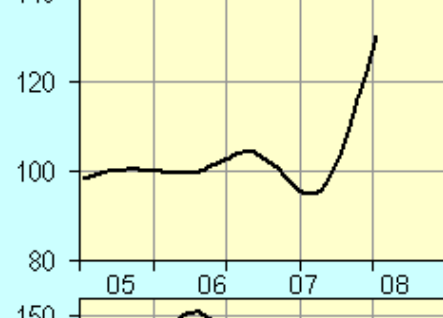
Sort by level
Sort by rate of change



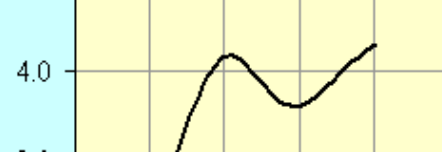
Customer 1



Customer 2



Customer 3



Customer 4

Customer 1

Customer 2

Customer 3

Customer 4

Customer 5

Customer 6

Customer 7

⋮

Customer 19

Customer gr. A

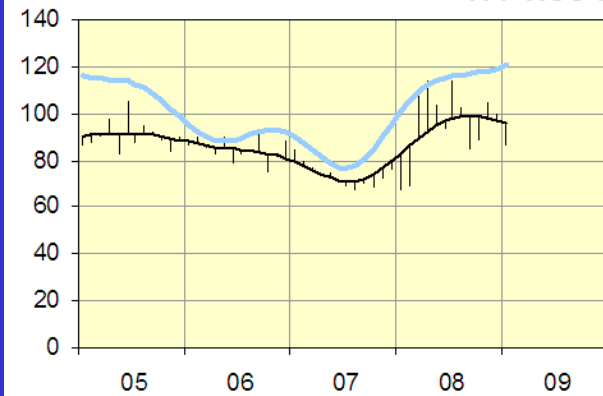
Customer gr. B

Other customers

Customer 1

Data including 2009 period 1

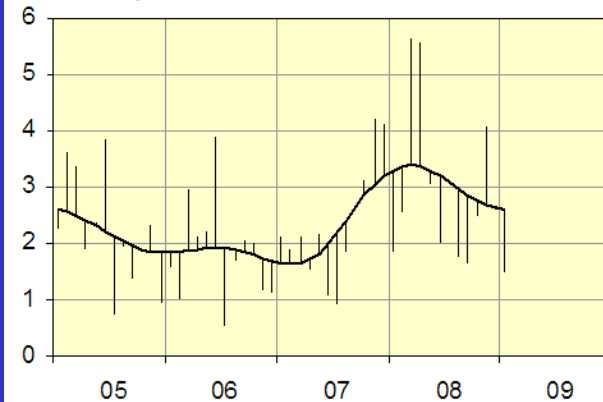
1. Price index, 2008 = 100



	Prices before discount	Prices after discount	Prices - raw material cost
2005	91.6	94.5	97.6
2006	84.8	88.3	80.7
2007	73.7	76.6	70.7
2008	100.0	100.0	100.0
Up to now:			
2009	86.3	86.4	77.5
Forecast			
2009	86.3		

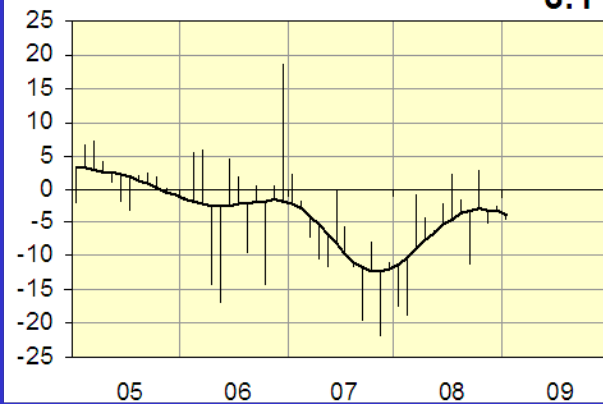
2. Volume of sales, 2008 years prices

\$ millions per month



	\$ millions per year
2005	26.4
2006	22.2
2007	27.8
2008	36.3
ARIMA-extrapolation	
2009	31.5
Forecast	
2009	31.5

3. Profit margin



	Mkr per år			Procent
	Sales-bonus	Profit	Costs	Profit margin
2005	24.1	0.5	23.6	2.0
2006	18.7	-0.4	19.1	-2.1
2007	20.5	-2.1	22.6	-10.3
2008	36.3	-1.4	37.7	-3.9
Up to now :				
2009	1.3	-0.1	1.3	-4.5
Forecast:				
2009				-4.5

Storytelling:

Relevant charts and tables together in a meaningful combination

Two "clicks" to get this detailed information based on some of the 1 600 series