

# Displays of indicators and of their accuracy

**Beat Hulliger**  
**Statistical Methods Unit**  
**Swiss Federal Statistical Office**

**Conference on “Visualising and Presenting Indicator Systems”**  
**14 – 16 March 2005**

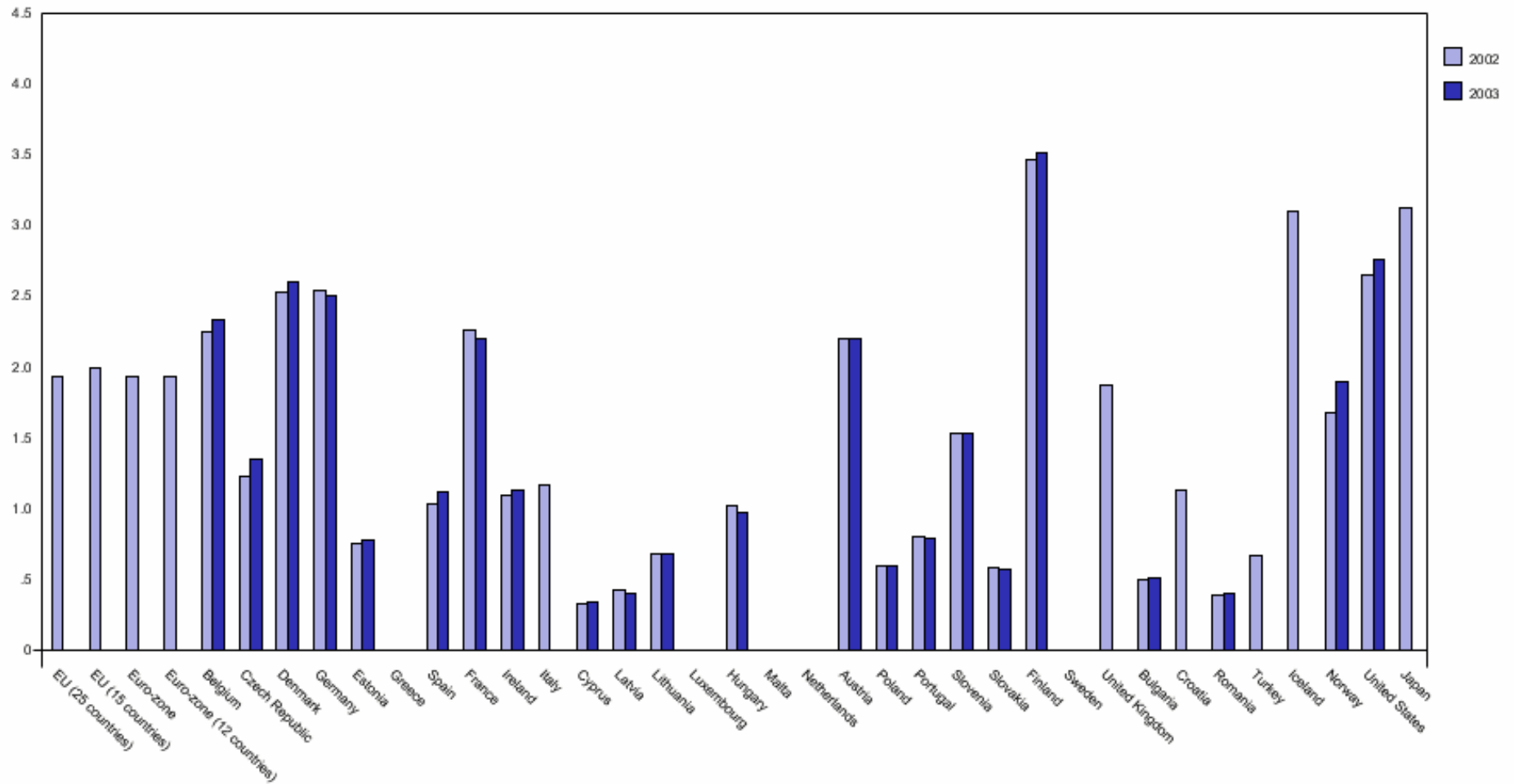
Copyright © 2005 Swiss Federal Statistical Office (SFSO) – All rights reserved. Use for non-commercial and non-licensing purposes on condition of acknowledgement of the source permitted

# Content

- Introduction
- Challenge of displaying indicators
- Displaying accuracy
- Examples of displays with accuracy
  - Categories
  - Time
  - Categories and time
  - League charts
- Some conclusions

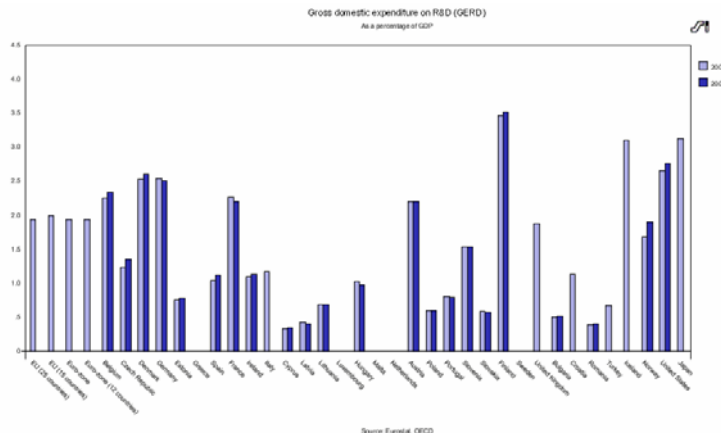
## Gross domestic expenditure on R&D (GERD)

As a percentage of GDP



Source: Eurostat, OECD

## Eurostat Structural Indicators on R+D



- Many countries across x-axis
- RD as % of GDP on y-axis
- Two years as paired categories: Change practically invisible
- Missing values
- No footnotes
- No variability

## TV-News

- An indicator in the main TV-news issue may take 10 seconds
- It has to compete with background visual material
- It may have to be animated and colored

## Public addressed and purpose of displays of indicators

- Public addressed: politicians and managers with limited statistical knowledge and less time.
- Purpose (of displays) is not to analyse a problem but to give a quick overview and highlight salient features
- 10 seconds on TV, 1 minute (?) in a report
- Displays of indicators must compete with other visual information
- Displays of indicators must be fancy!

## Displays of indicators

- 1 dimension for indicator (y-axis)
- 1 dimension for covariate (x-axis)
- Covariate categories: Indicator as bar
- Covariate time: indicator as line
- Covariate category and time: usually bars

## Challenge of displaying indicators

- No more than
  - 1 dimension for indicator (y-axis: usually quantitative)
  - 1 dimension for covariate (x-axis: categoric or time)
  - 1 display only!
- No time or space for footnotes
- Display must be self-explaining to a large part
- Accuracy and warnings must be integrated in the display
- The display must look fancy

## Displaying accuracy

- Bias
  - Difficult to display because often bias cannot be quantified
  - possible way out: scenarios and/or sensitivity analysis
- Variance
  - Confidence intervals for point estimates are good but ...
  - Testing is the real problem but ...
  - Users don't understand tests!

## Variance estimation

- Publication of variances is still the exception
- Publication of variances of change is practically inexistent
- If variances are published then often only sampling variances
  
- Here: Assume that a variance estimate is published or known to the statistician that plots the graph.

## Working horse: Confidence intervals to test equality of two normal means

Confidence interval for normal mean  $\bar{X}$   
 with estimated standard deviation  $\hat{\sigma}(\bar{X})$

$$CI(\bar{X}) = \bar{X} \pm z(\beta) \hat{\sigma}_{\bar{X}}$$

where  $z$  is the standard normal quantile  $z(\beta) = \Phi^{-1}\left(\frac{1+\beta}{2}\right)$

## Confidence interval for difference of two means

$$CI(\bar{X} - \bar{Y}) = (\bar{X} - \bar{Y}) \pm z(\beta) \sqrt{\hat{\sigma}_{\bar{X}}^2 + \hat{\sigma}_{\bar{Y}}^2 - 2\rho \hat{\sigma}_{\bar{X}} \hat{\sigma}_{\bar{Y}}}$$

where  $\rho$  is the correlation between X and Y.

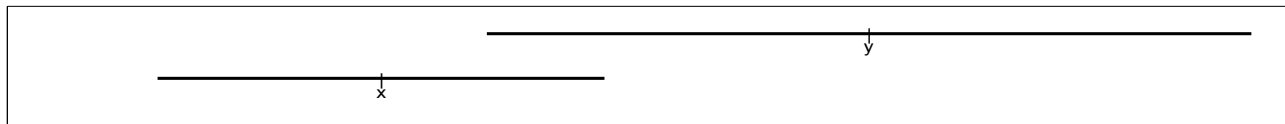
Usually  $\beta=95\%$

Correct test for equality at level  $1-\beta$ :  $0 \notin CI(\bar{X} - \bar{Y})$

is replaced by "overlap test":  $CI(\bar{X}) \cap CI(\bar{Y}) = \{ \}$

which corresponds to:  $0 \notin \bar{X} - \bar{Y} \pm z(\beta) (\hat{\sigma}(\bar{X}) + \hat{\sigma}(\bar{Y}))$

Group



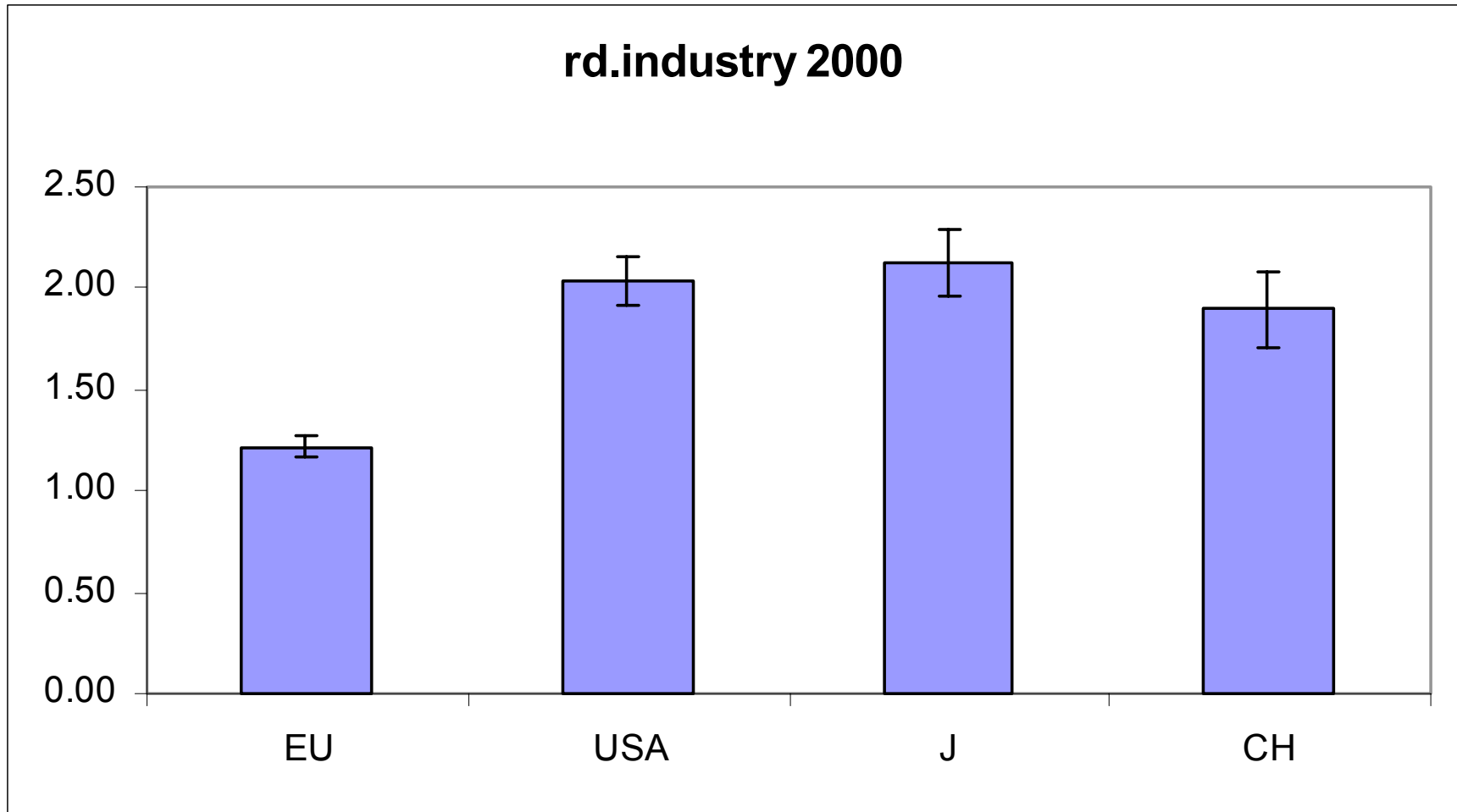
Indicator

## Overlap test is

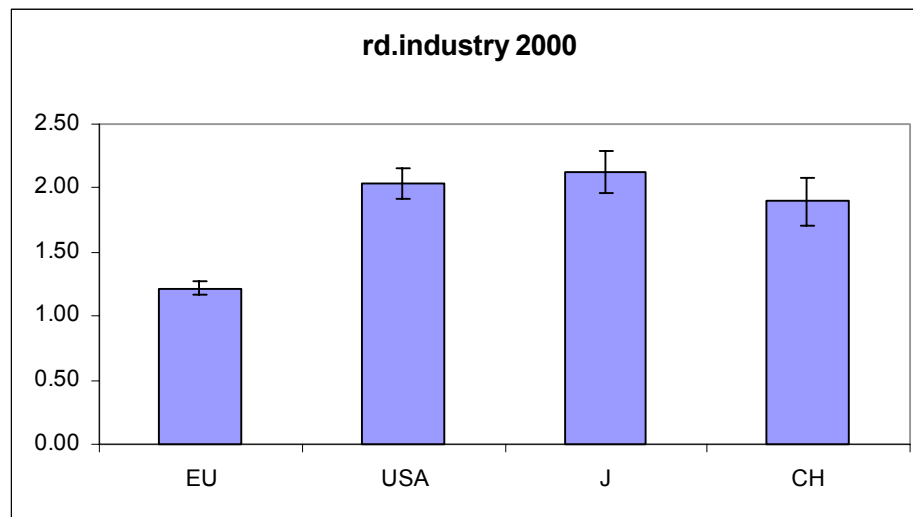
- Approximate because the correlation  $\rho$  is neglected
- Correct under perfect negative correlation ( $\rho=-1$ )
- Mildly conservative under negative correlation ( $\rho<0$ )
- Conservative under independence ( $\rho=0$ ) (At nominal level 95% true confidence level  $\approx 99.4\%$  for equal variances)
- Too conservative under positive correlation ( $\rho>0$ ) (E.g. at  $\rho=0.5$  and nominal level 95% true conf. level  $\approx 99.991\%$ )

## Examples

- Based on indicator for R+D expenditures
- Variances and correlations are fictive except for the coefficient of variation of RD expenditures of industry in Switzerland in 2000: 5% (not published).
- Assume independence between countries
- Examples with Excel and with R

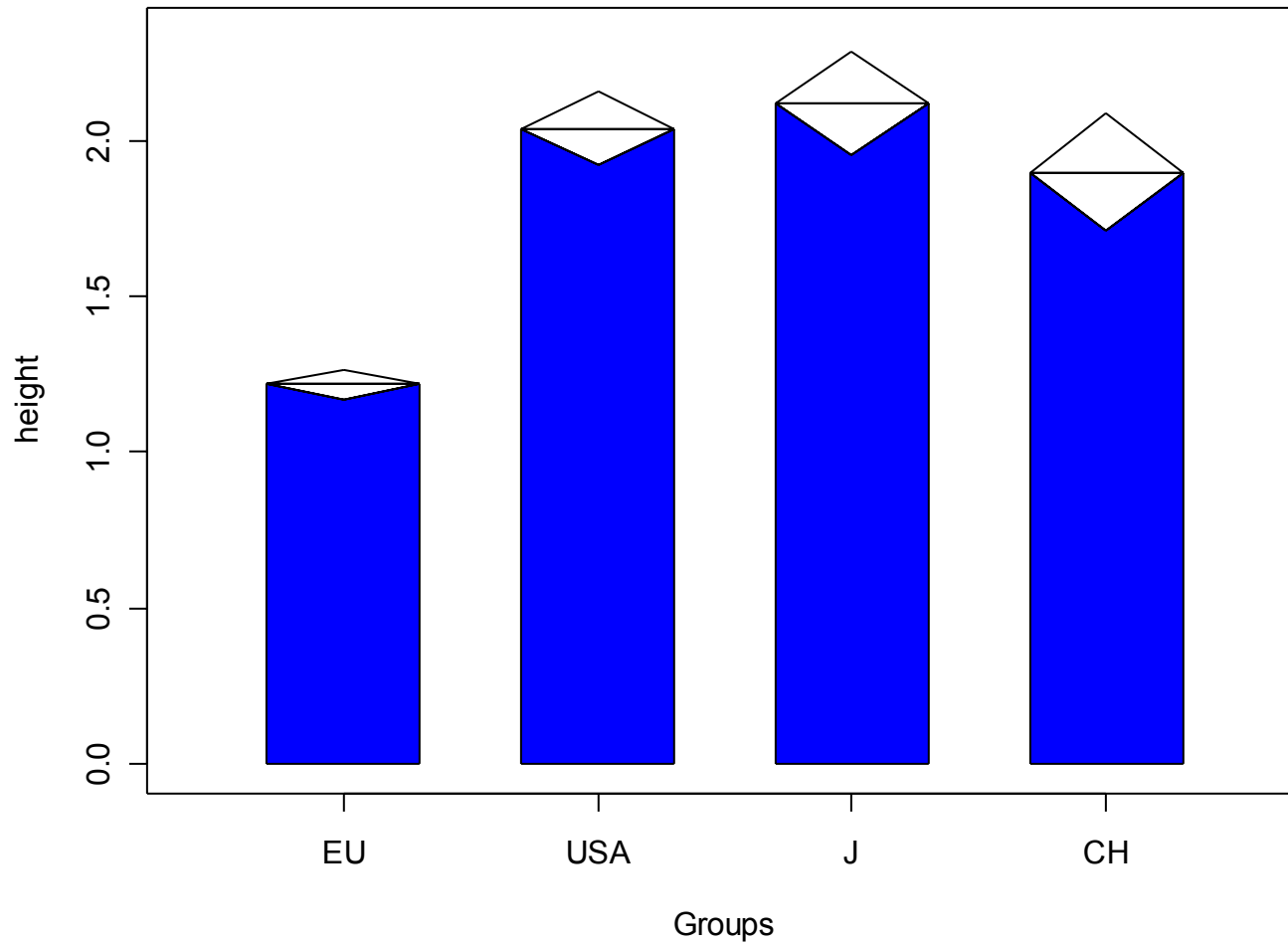


## Excel with error bars

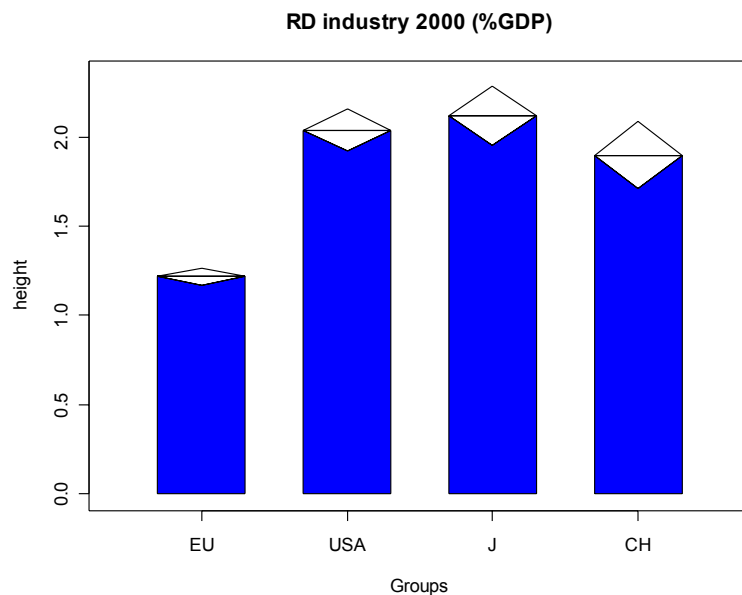


- Only EU is significantly lower
- Multiple comparison? (No correction! (Goldstein and Healy 1995))
- Error bars (whiskers) are not prominent enough
- Not fancy (in my opinion)

### RD industry 2000 (%GDP)



## Candle plot

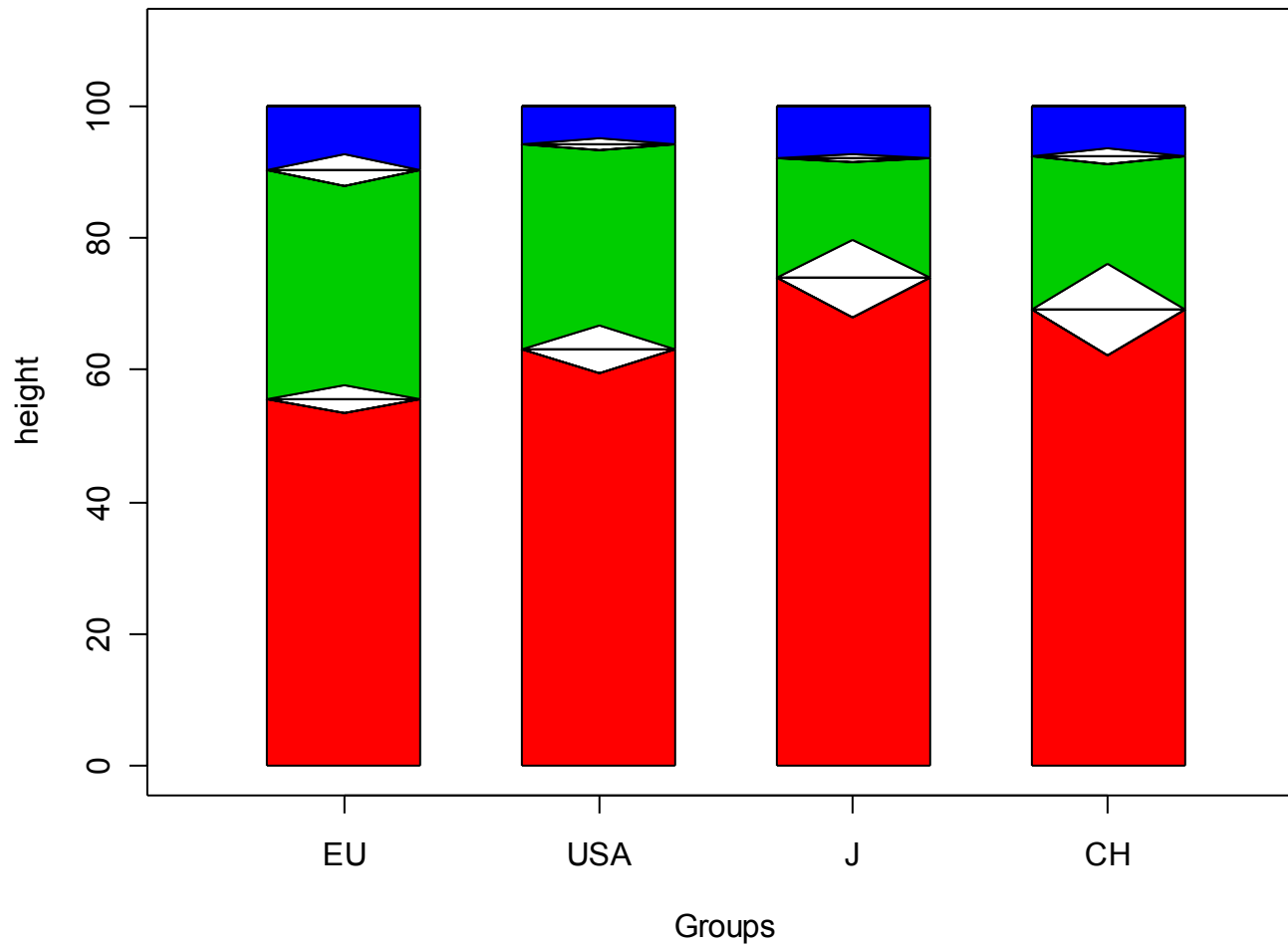


- "Flame" indicates CI enough prominently
- Colored surface diminished by half of lower confidence limit
- Suggests triangular distribution around mean
- Symmetrical
- Fancy enough?

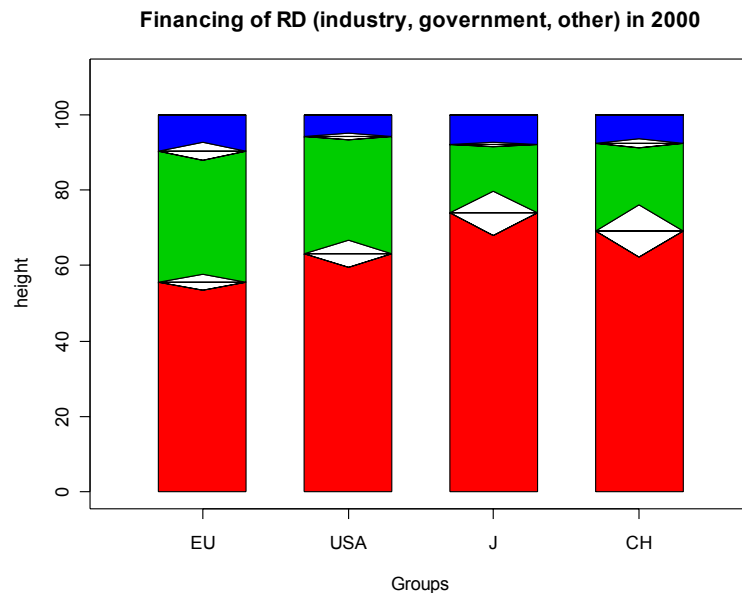
## Compositions

- Proportions adding up to 1
- Pie chart now seldom used because of its severe perceptual problems
- Stacked bar charts
- Usually two or more groups to compare

### Financing of RD (industry, government, other) in 2000

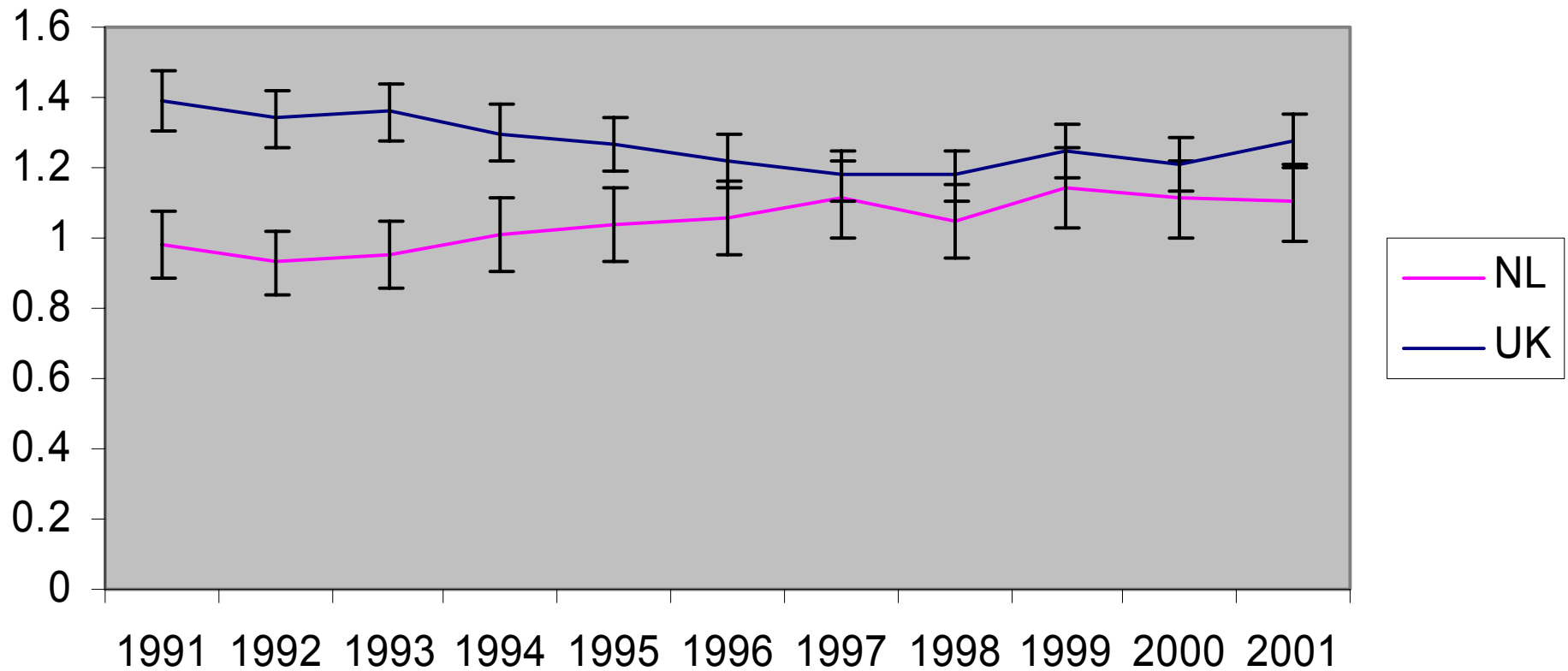


## Stacked candle plot

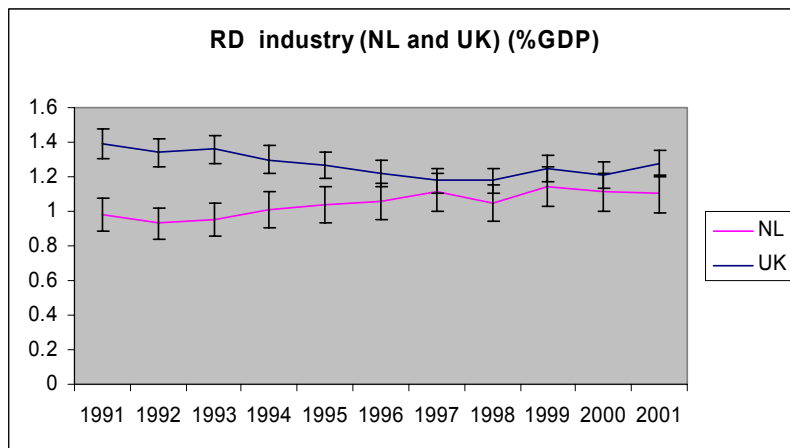


- Problem: Green surface is also diminished from below
- Proportions are negatively correlated: Joint confidence region for proportions needs  $d-1$  dimensions
- Only CI of cumulative proportions make sense:  $CI(p_1+p_2)$
- Intuitive? Clear?

### RD industry (NL and UK) (%GDP)

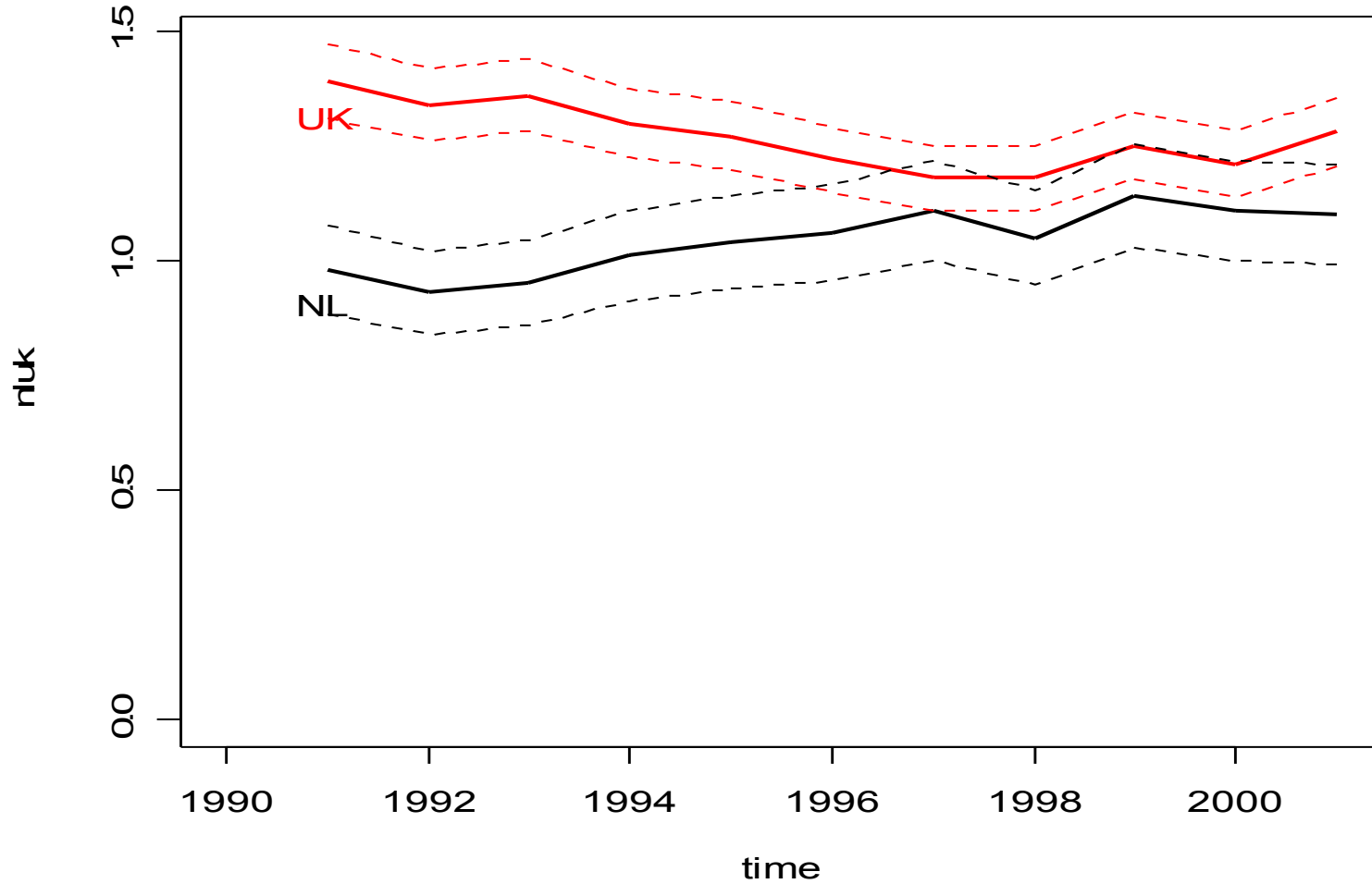


## Multiple time series: Several categories over time



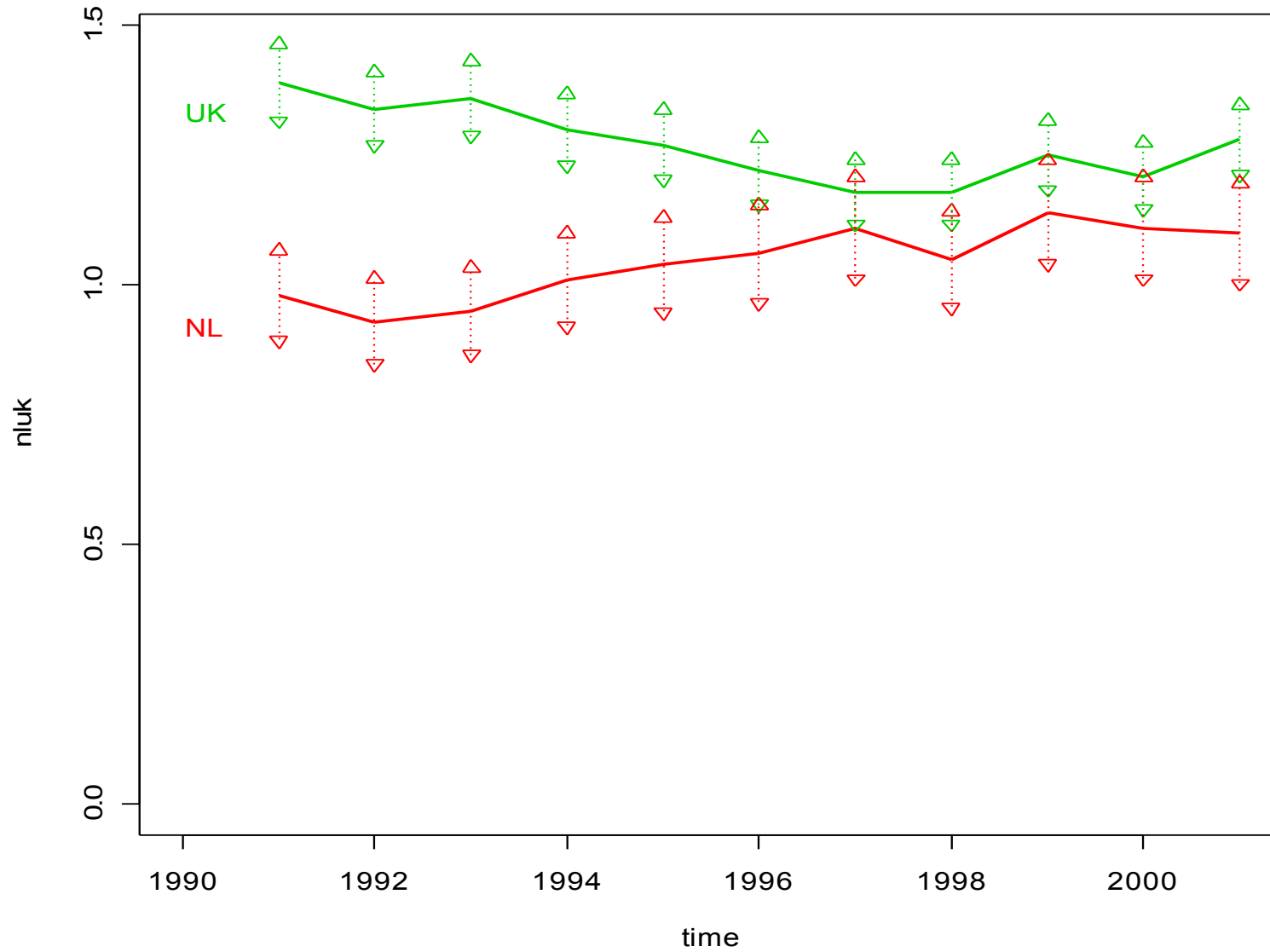
- Line plots adequate (not bars!)
- Interest: Comparison of time series
- Test: Difference between time series at specific time points
- Overlapping intervals are not well visible.
- Not fancy
- First guess: Draw confidence lines like for regression

### RD expenditures of industry (%GDP)

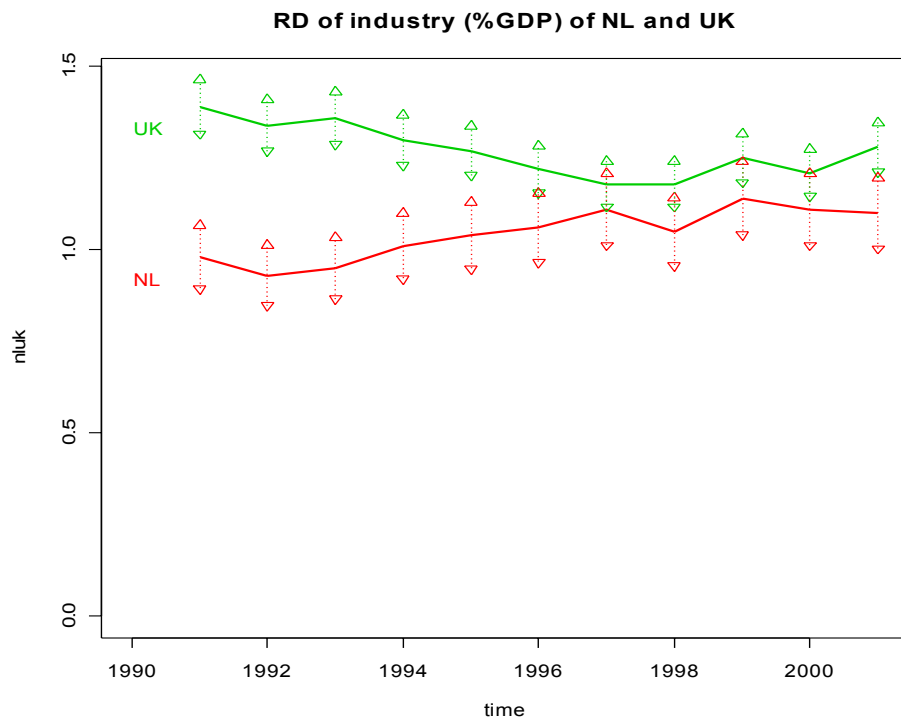


OECD

RD of industry (%GDP) of NL and UK

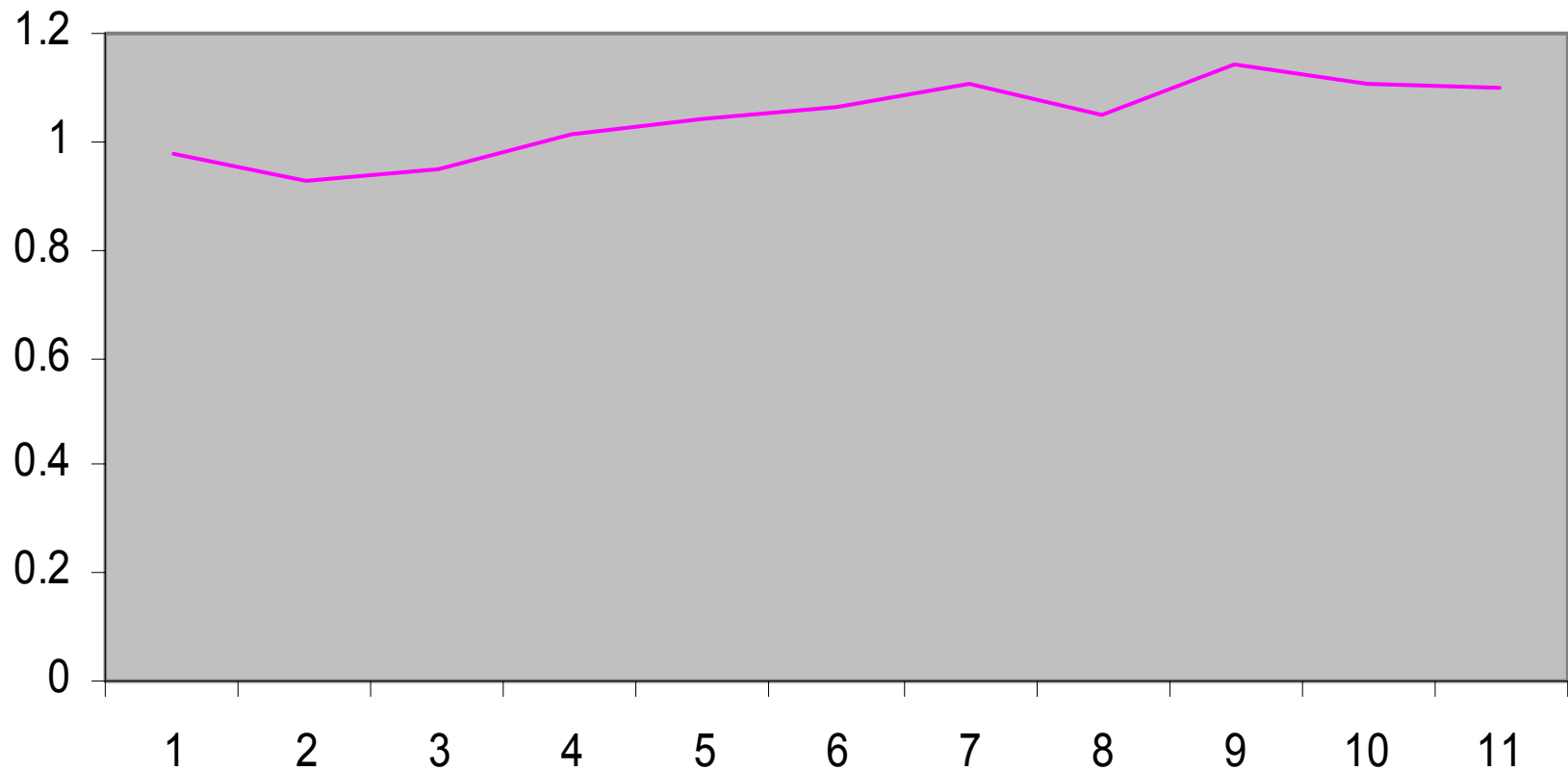


## Multiple line plot with CI at time points

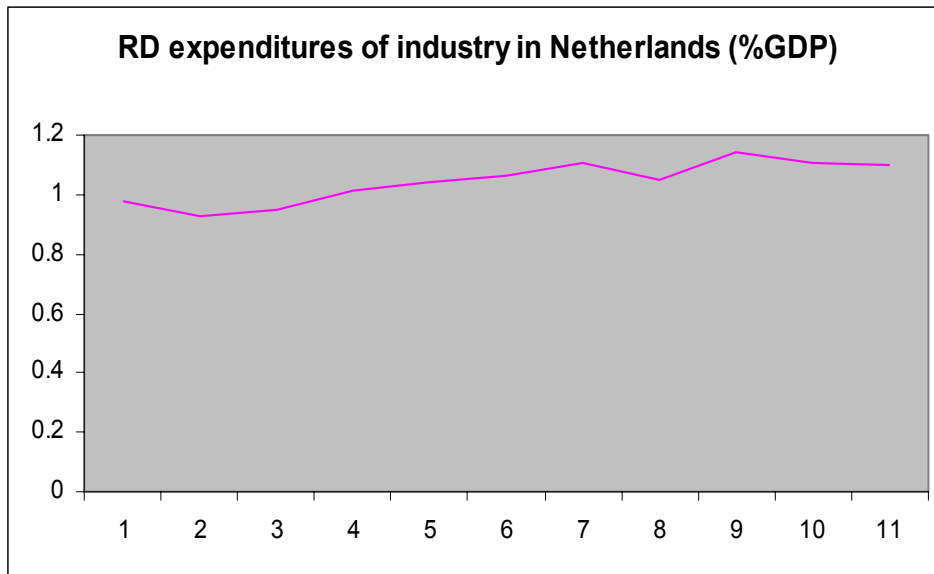


- The CI at the time points are drawn individually.
- To make the overlap visible use small triangles (2-dim!)
- The statistically interesting question, whether  $NL < UK$  consistently is not addressed!

## RD expenditures of industry in Netherlands (%GDP)

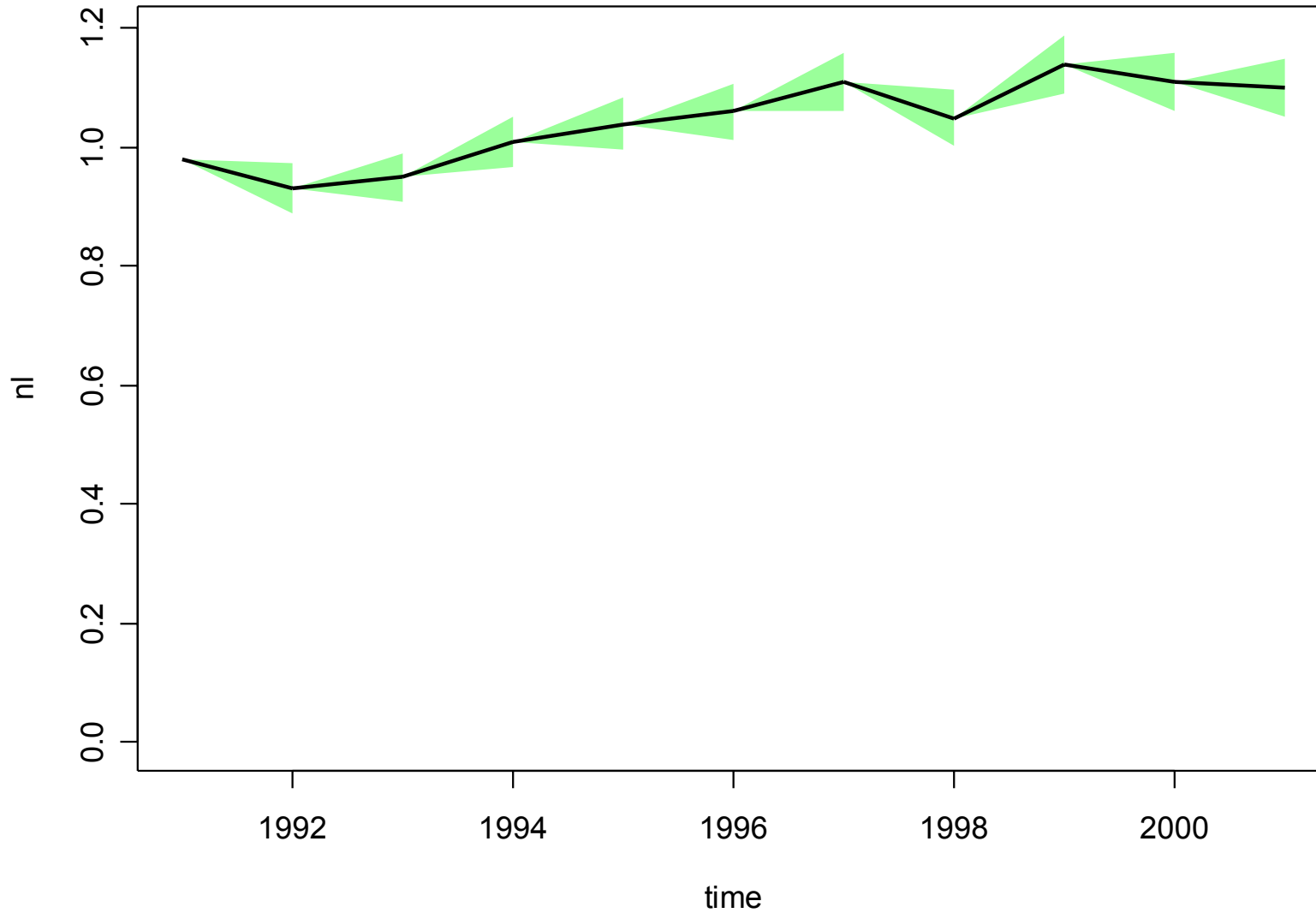


## Single time series

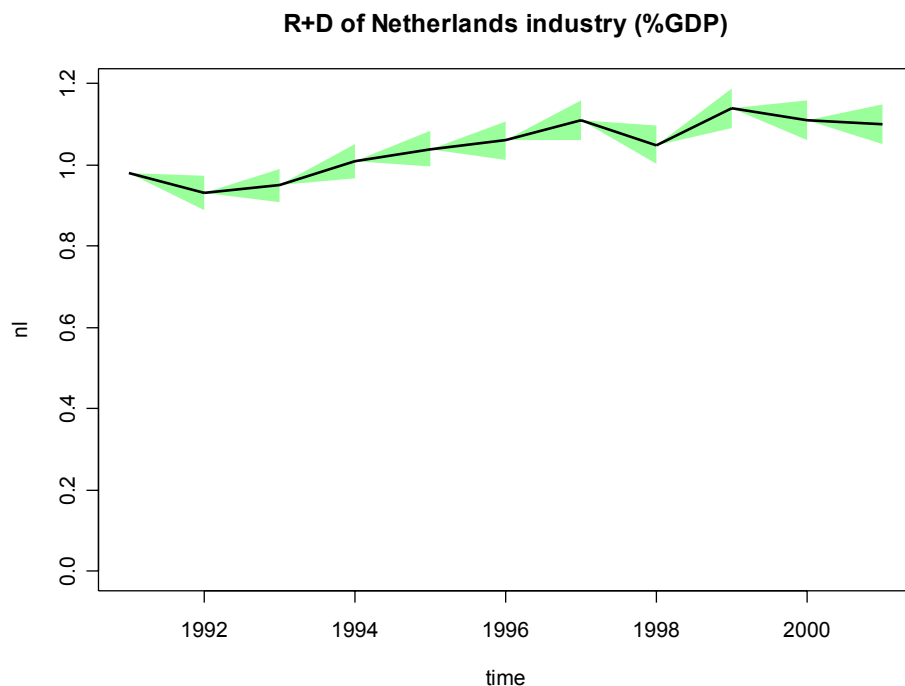


- Line plot shows evolution
- Interest: change between two specific time points
- How to make clear which time points to compare?
- Usually positive correlation (panels)
- Individual CIs for the time points will not convey the correct message

### R+D of Netherlands industry (%GDP)

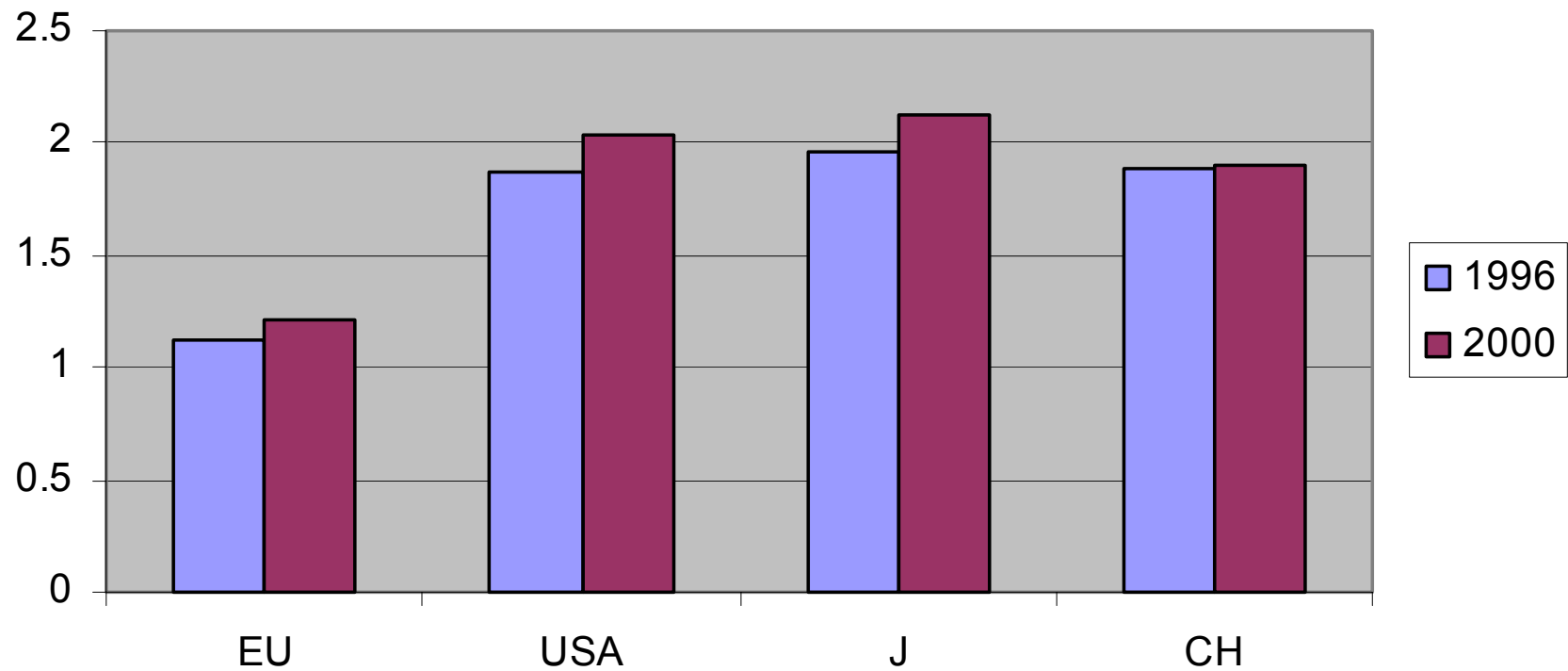


# Funnelplot

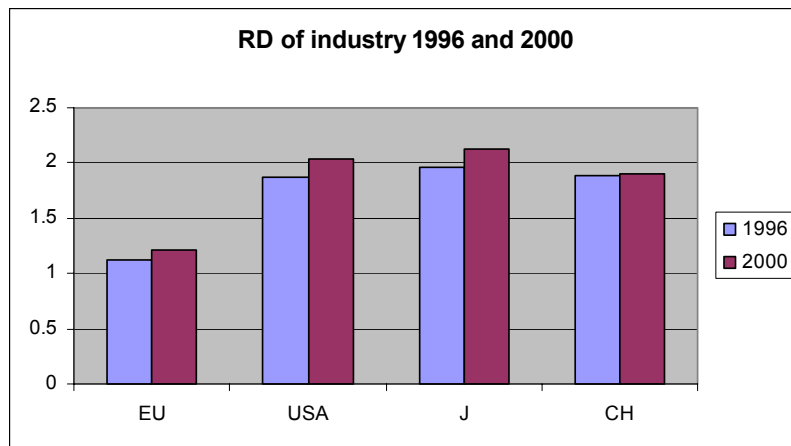


- Test for change from  $x(t)$  to  $x(t+1)$  indicated by a funnel
- Funnel opening is CI with length adjusted for correlation
- If funnel does not include horizontal line then change is significant
- Not indicated: change from start etc.!

### RD of industry 1996 and 2000

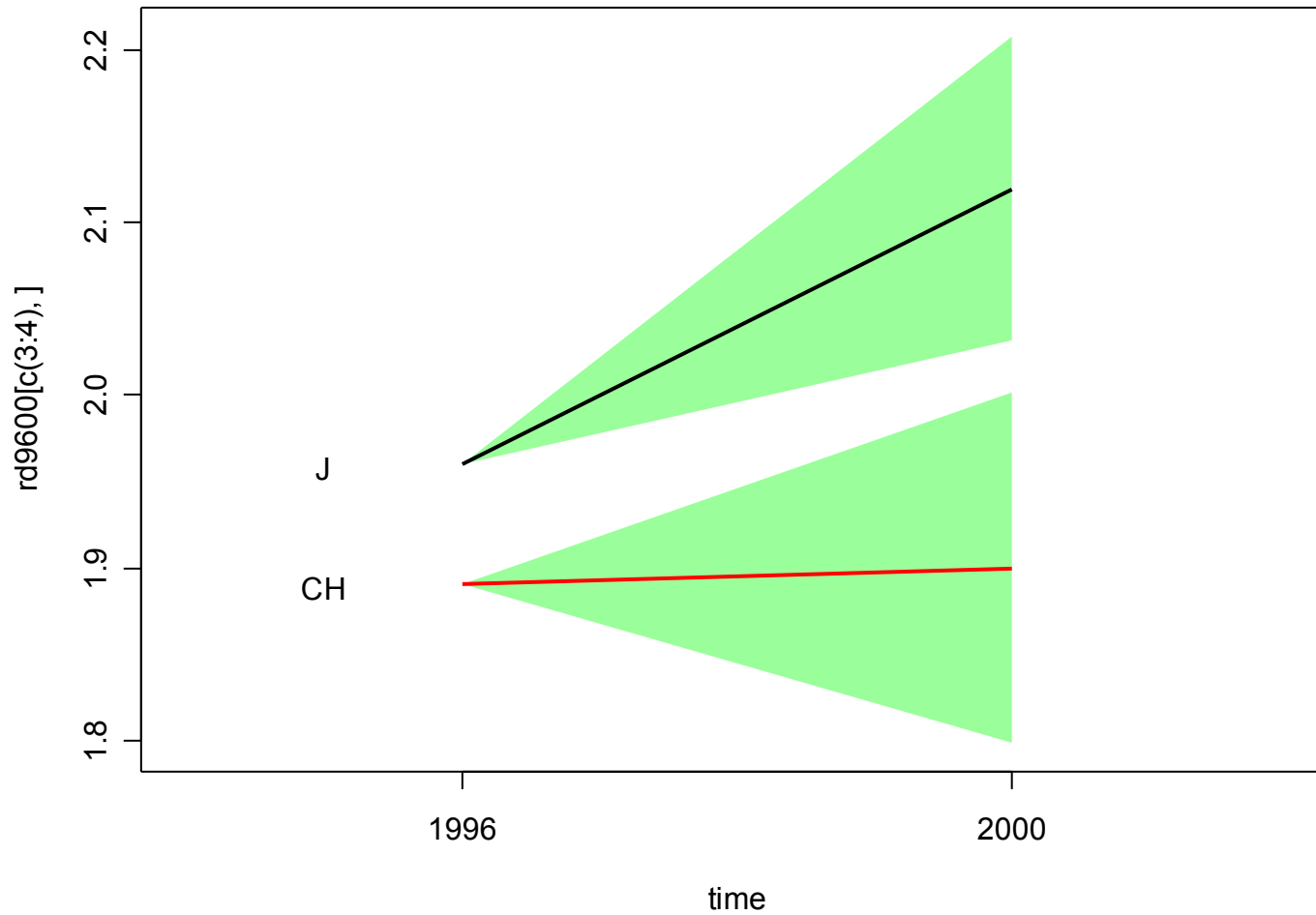


## Change and several categories

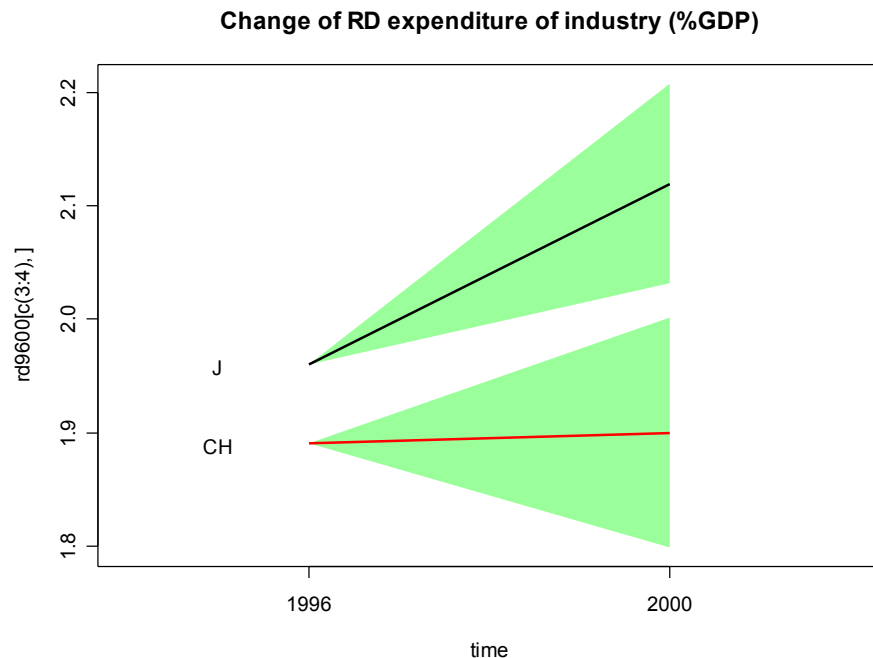


- A few categories are compared (countries) at two time points
- Difficult to read: Bad plot for indicators!
- Main question: difference between categories
- Secondary question: change significant
- funny candle ????

### Change of RD expenditure of industry (%GDP)



## Funnelplot for two time points



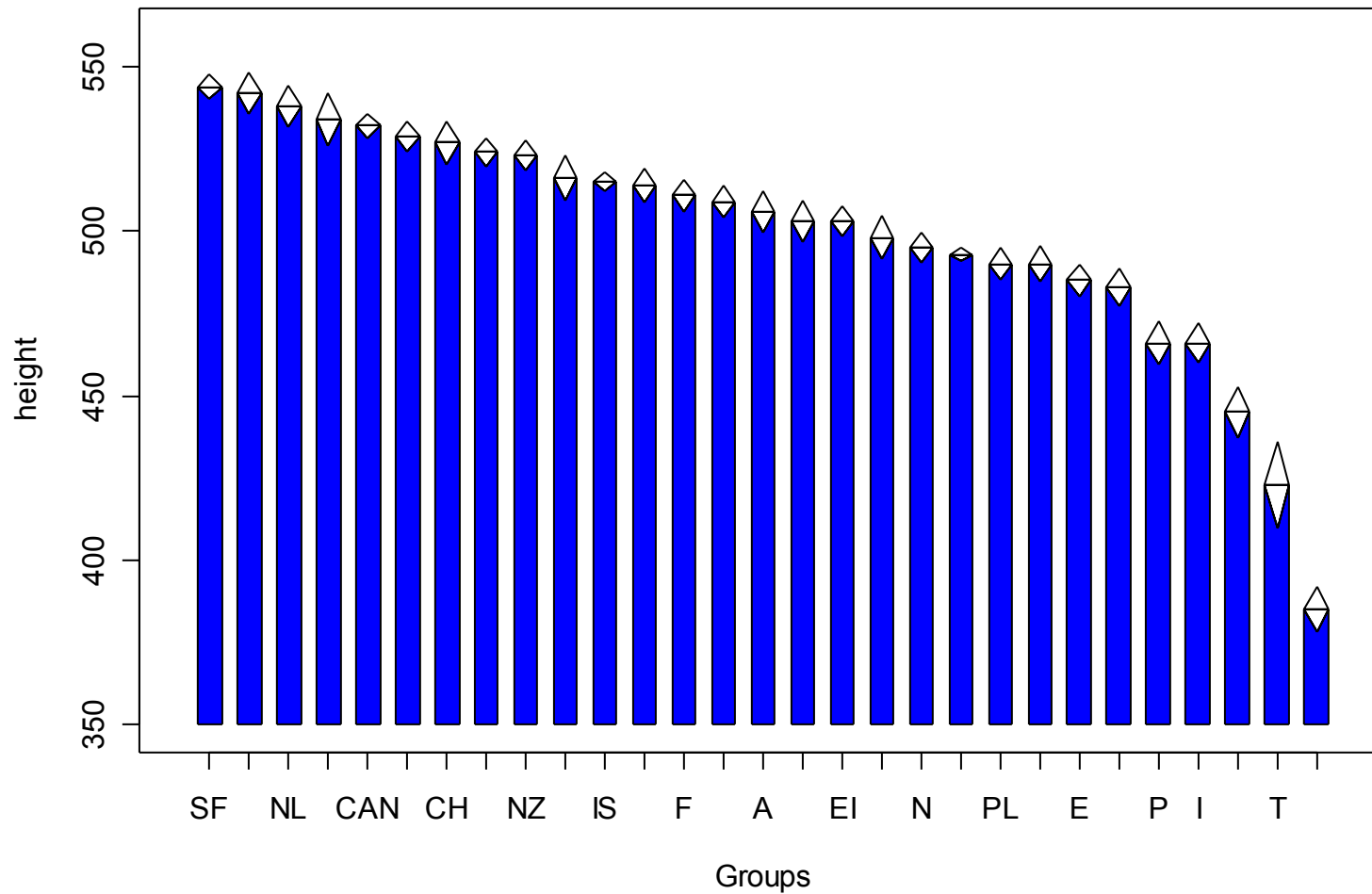
(Without funnels: displays in HDR )

- Lines show evolution well
- Interesting test: evolution of the differences between categories!
 
$$\Delta = (\bar{X}_2 - \bar{Y}_2) - (\bar{X}_1 - \bar{Y}_1)$$
- Test for interactions between time and category!
- Funnels: Narrowing of the gap cannot be excluded, neither widening!

## League Tables

- League tables (rankings) use relative benchmarks
- League tables are poor statistical summaries
- Replace relative with absolute benchmarks
- Plot confidence intervals

### PISA 2003 mean performance per country



## Some conclusions

- It is possible to include information on accuracy in displays and the graphs remain fancy without too large ink to information ratio.
- Graphs with accuracy become more complex
- The message of an indicator changes when its accuracy is displayed.
- Complex tests cannot be displayed simply
- Need more published variances!