Problems Solved in Information Visualization

or

Finally one Provocative 10min. Talk

Martin Theus

martin@theusRus.de

„Did anybody ask for our great tool?“

- As any other good scientific discipline, InfoVis will produce research on problems which have no application in mind.

- Like in Statistics and/or Data Mining, the validity of methods and algorithms is often shown on data/problems that were sampled according to the method.

- Stimulating input **MUST** come from areas of application.

- Working in applied statistics, we want to analyze data (graphically)

  ➥ visualization techniques are designed to answer a particular question, and that „problem is solved“, once this solution can be applied generally and used by the domain expert.

„Let‘s talk about science …“

- Once we separate DataVis from „arbitrary“ InfoVis, things get a lot easier.

- „Classical“ data can be associated with scales and measures, i.e. we are talking about visualizing distributions in IR^k now.

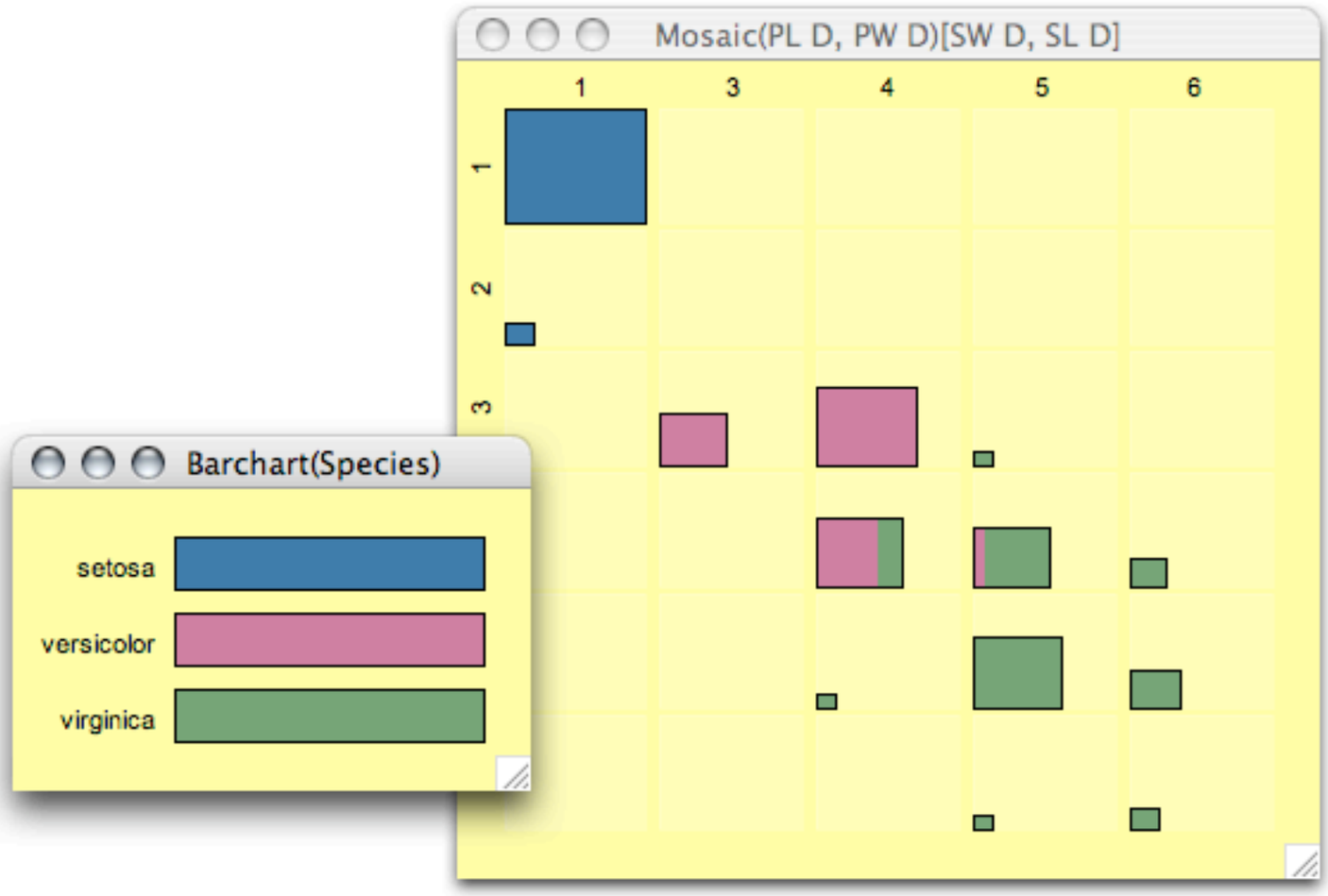
- Visualizing an underlying (univariate) distribution most faithfully is a clear cut task, with fewer degrees of freedom as we might think – there is hardly anything to „invent“ here!

- We can increase dimensionality by either linking further views, or find suitable multivariate plots (PCPs, Mosaic, ...)

Scales imply „Patterns“

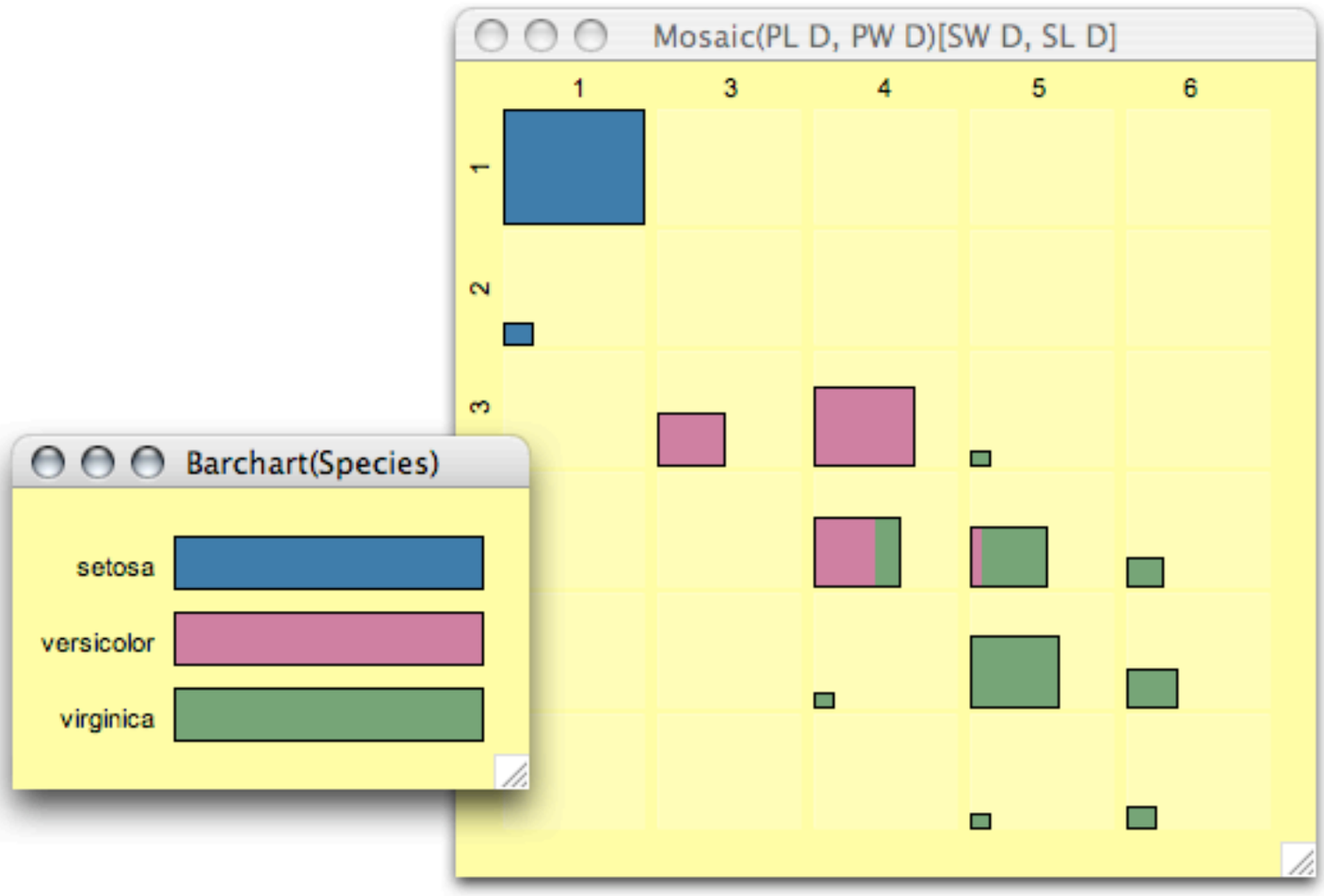
Scales imply „Patterns“

Iris data in a Fluctuation Diagram

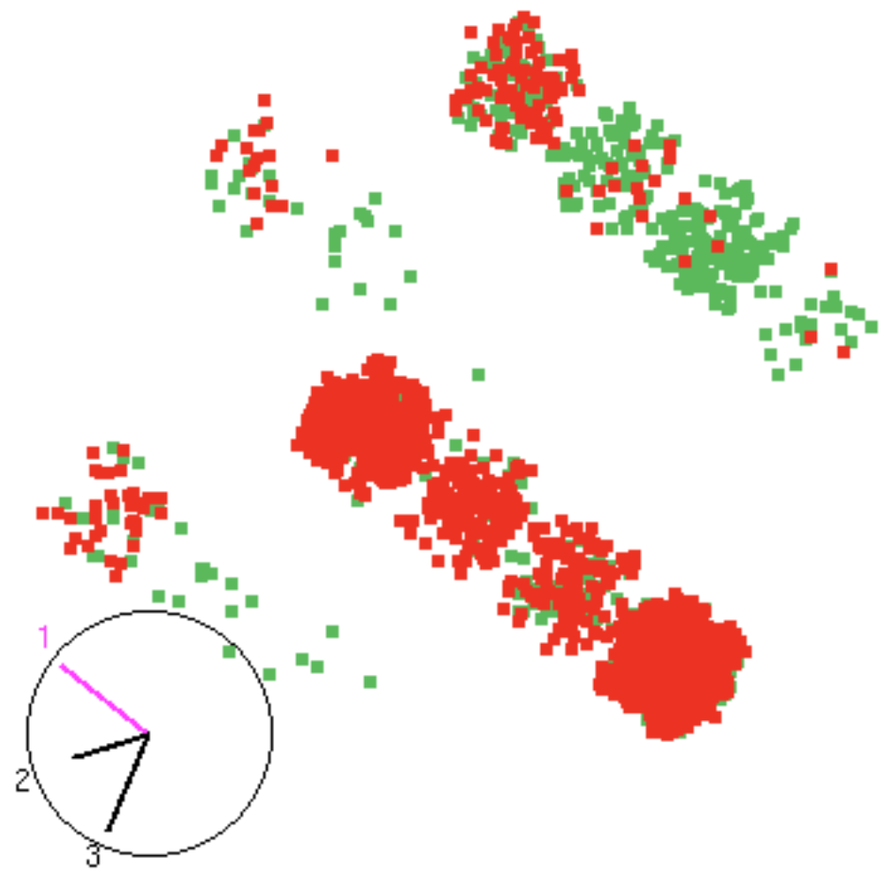


Scales imply „Patterns“

Iris data in a Fluctuation Diagram

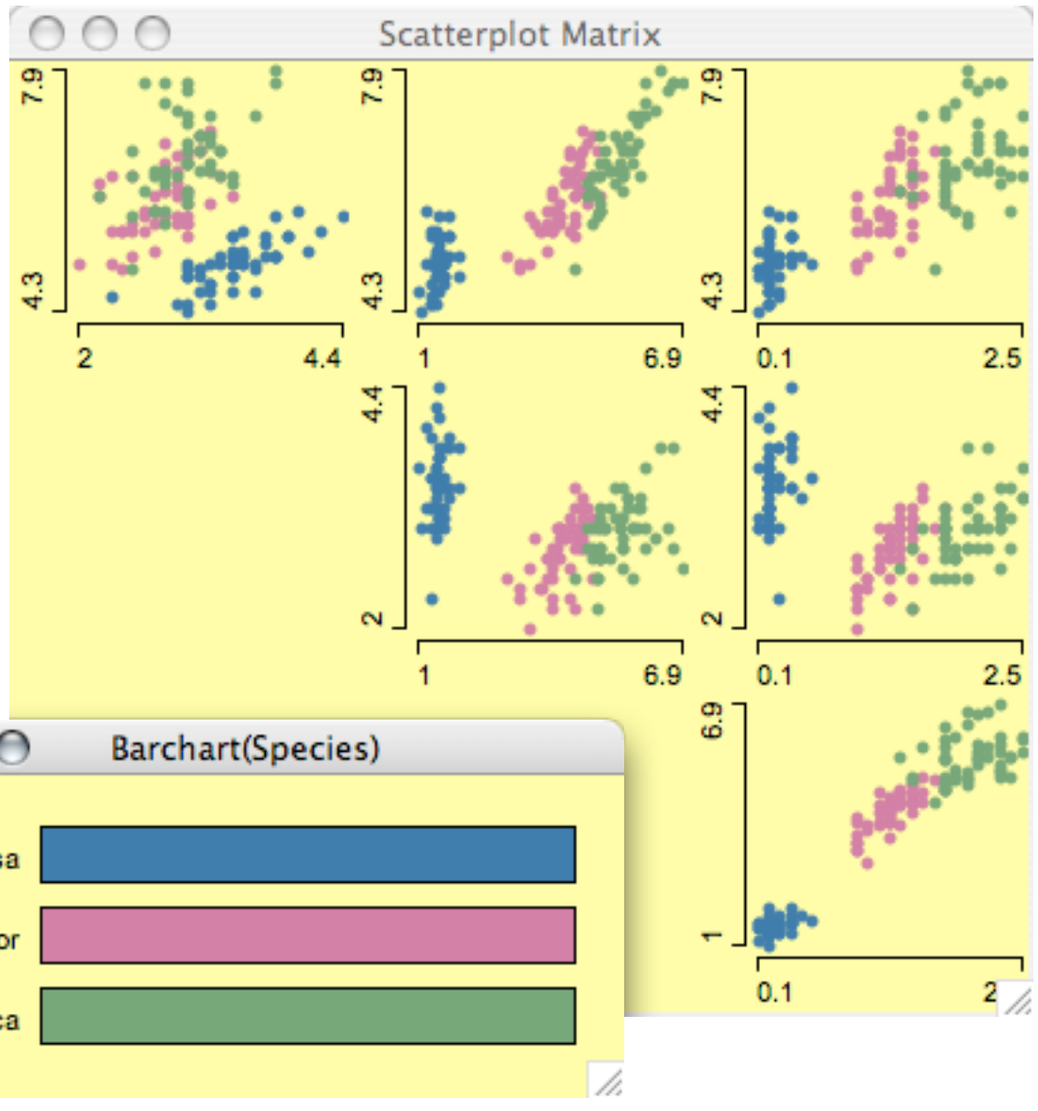


Titanic data in a scatterplot

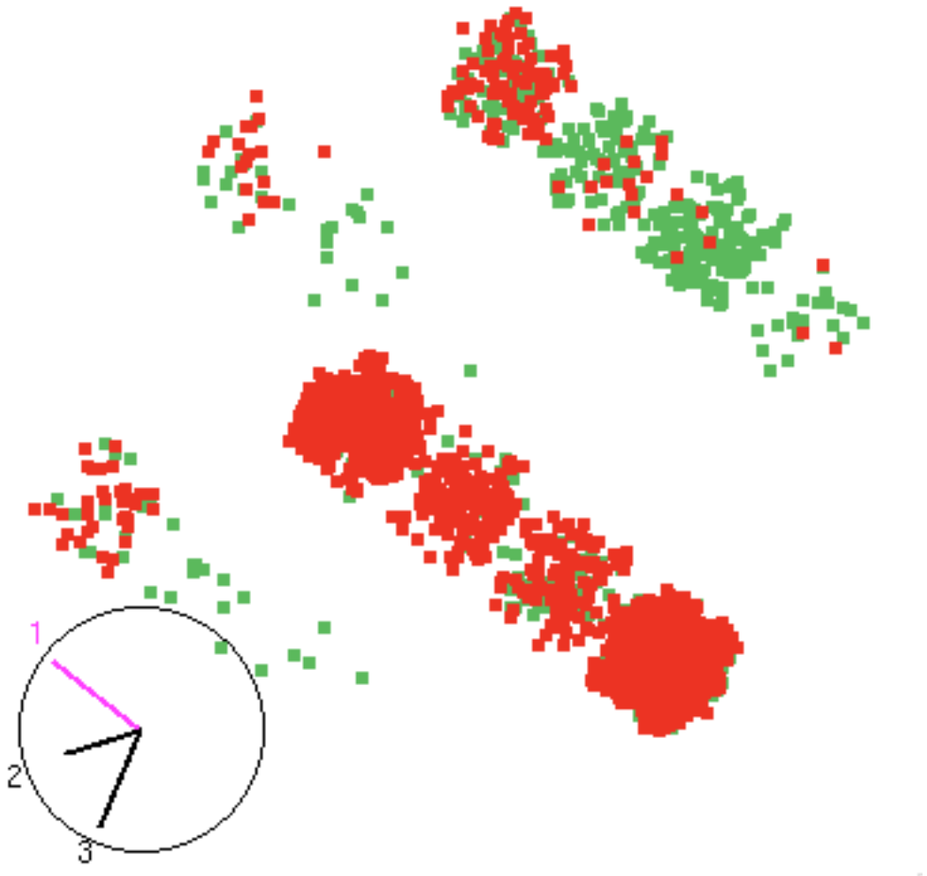


Scales imply „Patterns“

Iris data in a linked SPLOM

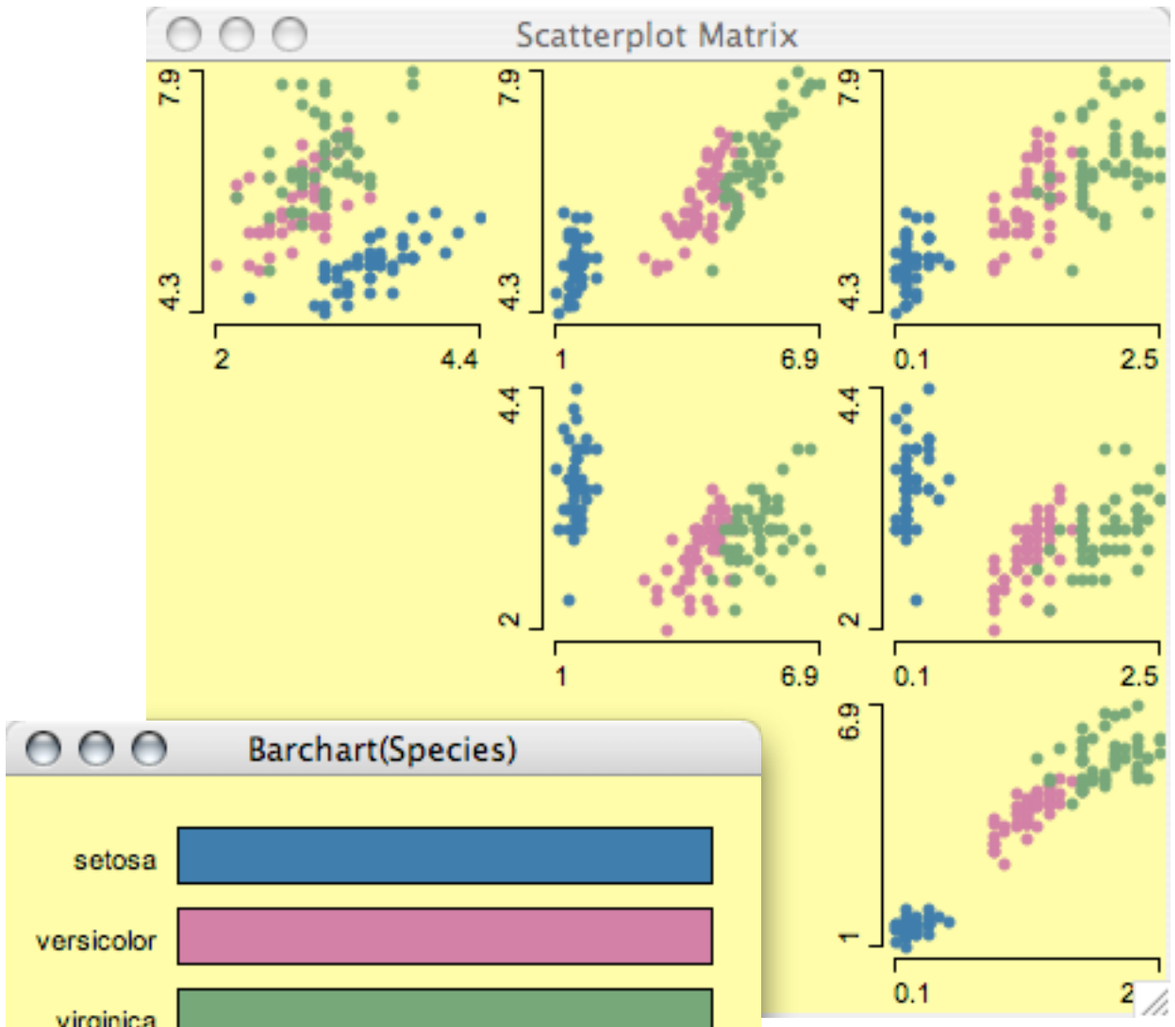


Titanic data in a scatterplot

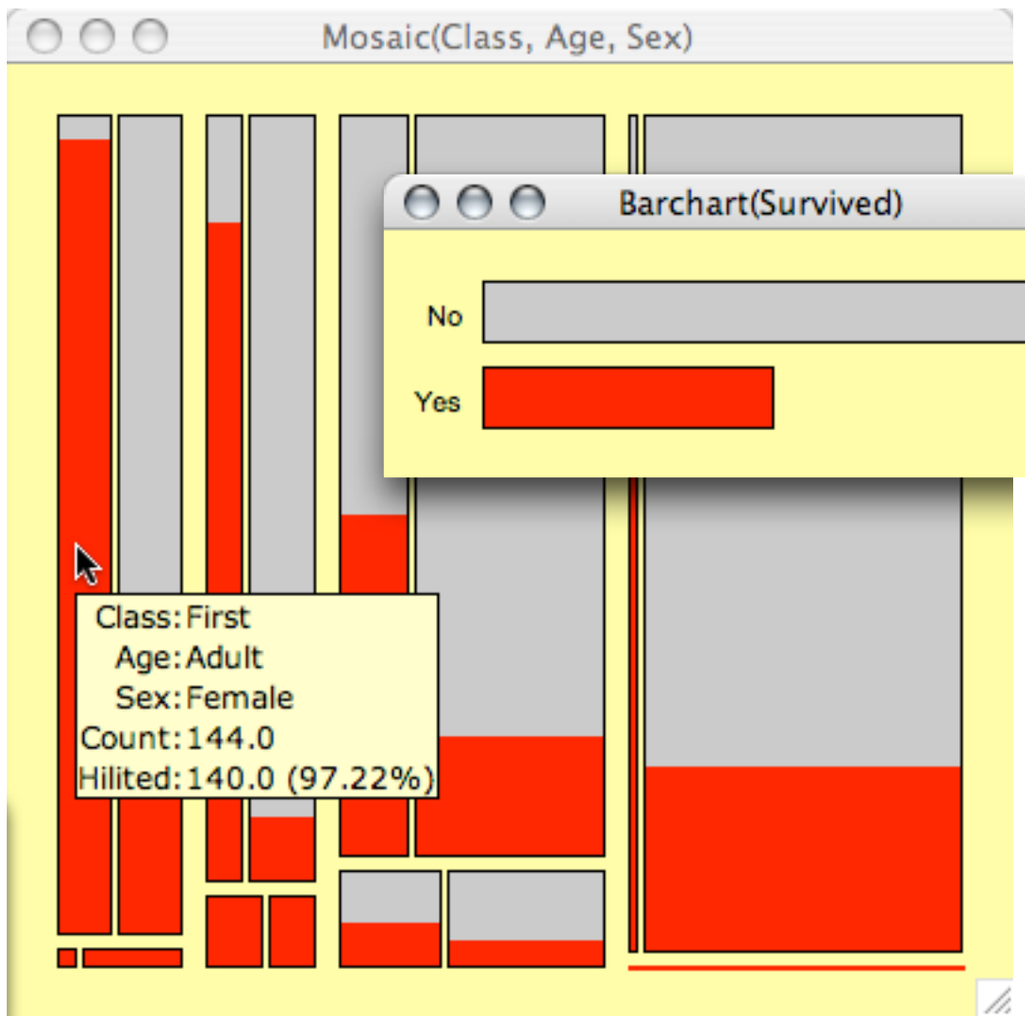


Scales imply „Patterns“

Iris data in a linked SPLOM

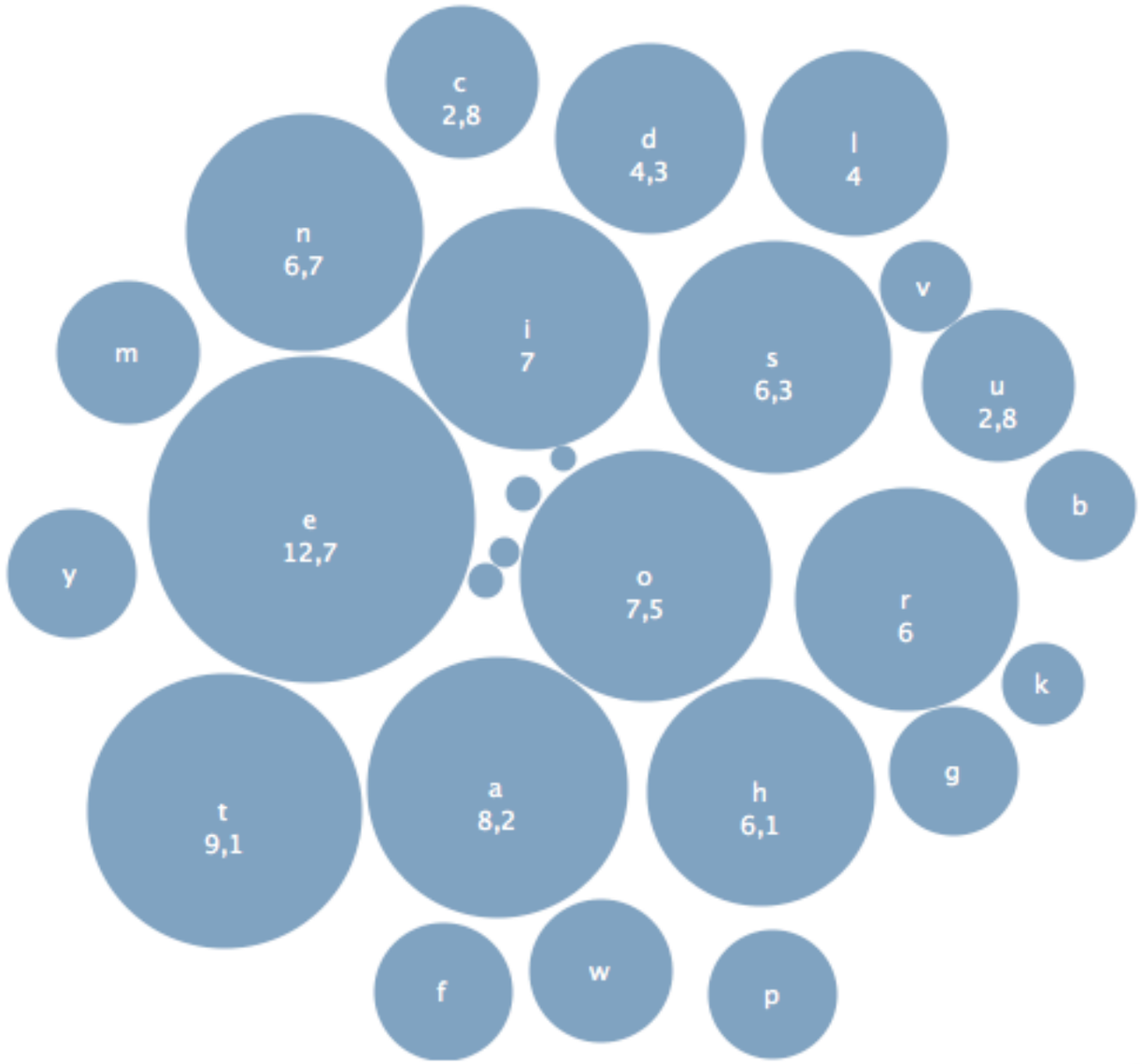


Titanic data in a mosaic plot



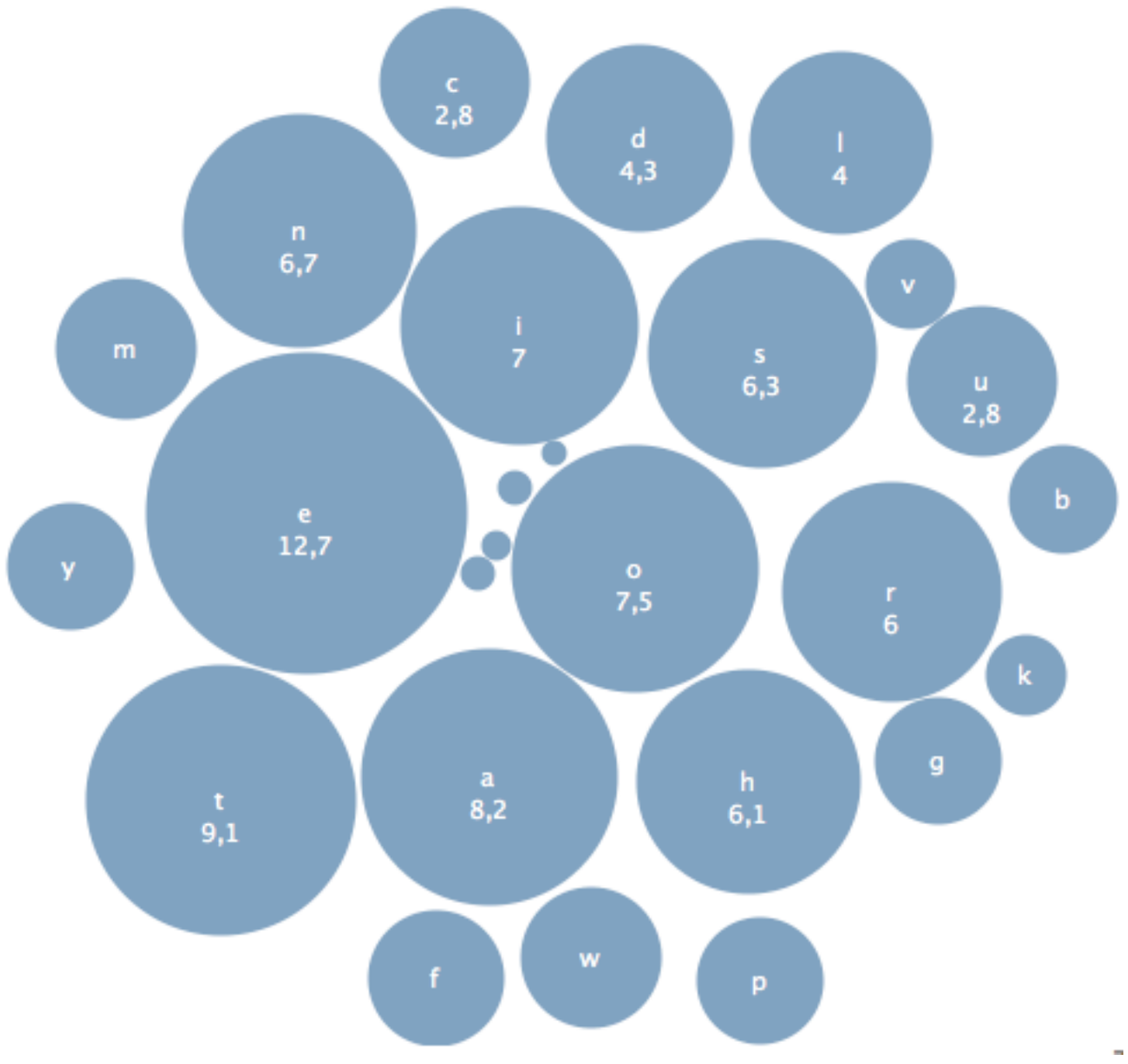
Some Mistakes ...

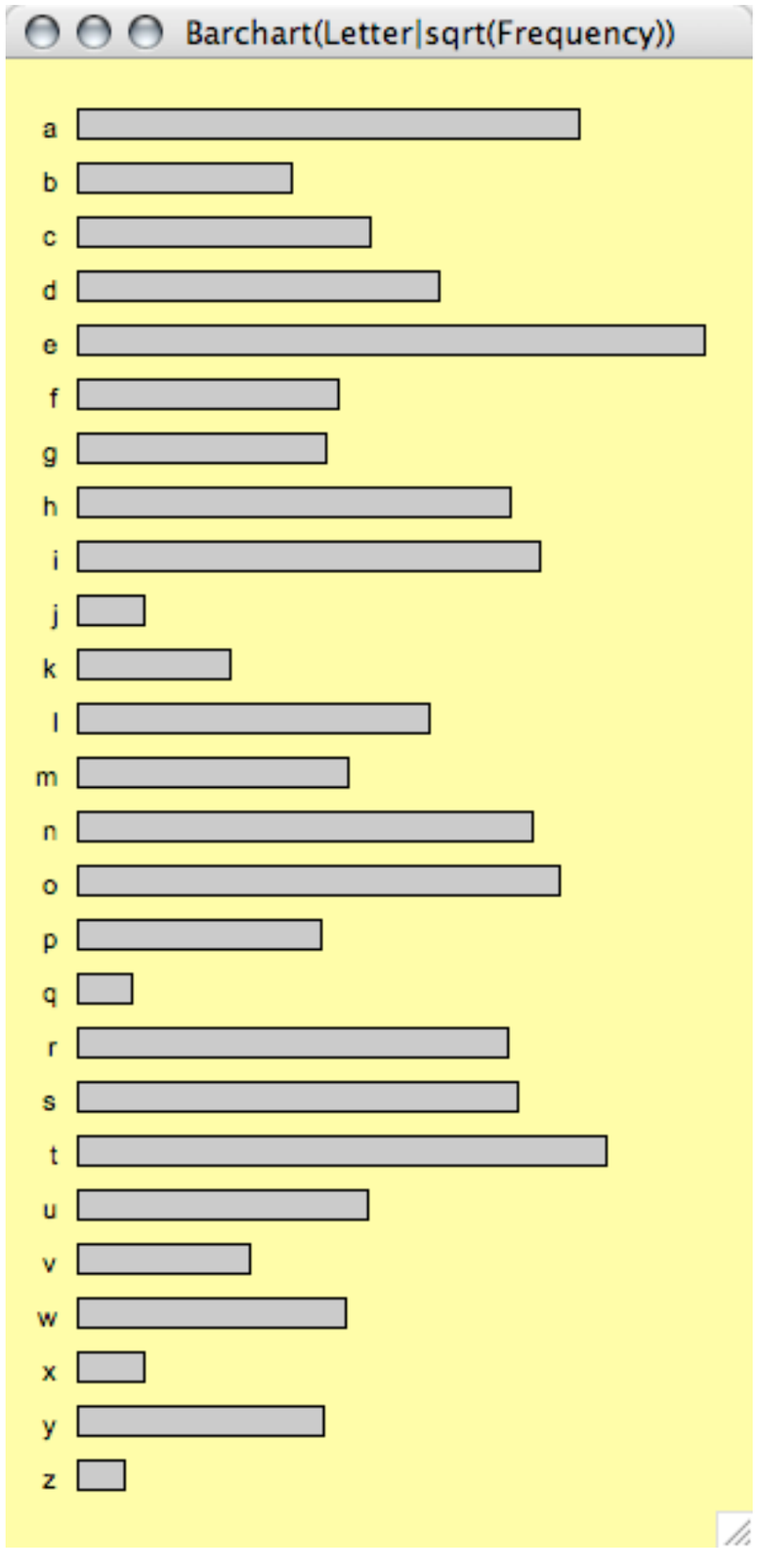
- … what is wrong with the bubble plot?



Some Mistakes ...

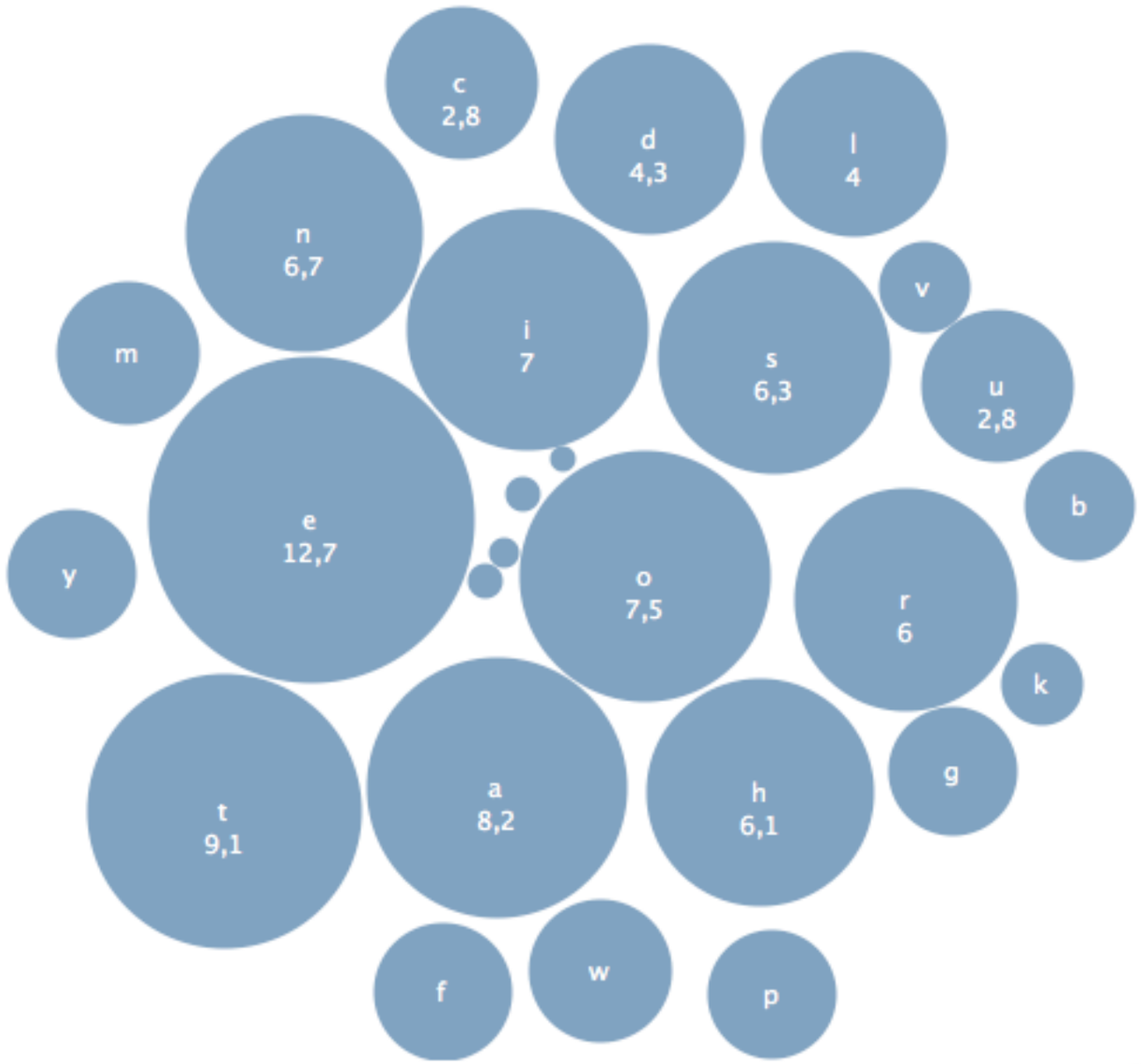
- … what is wrong with the bubble plot?

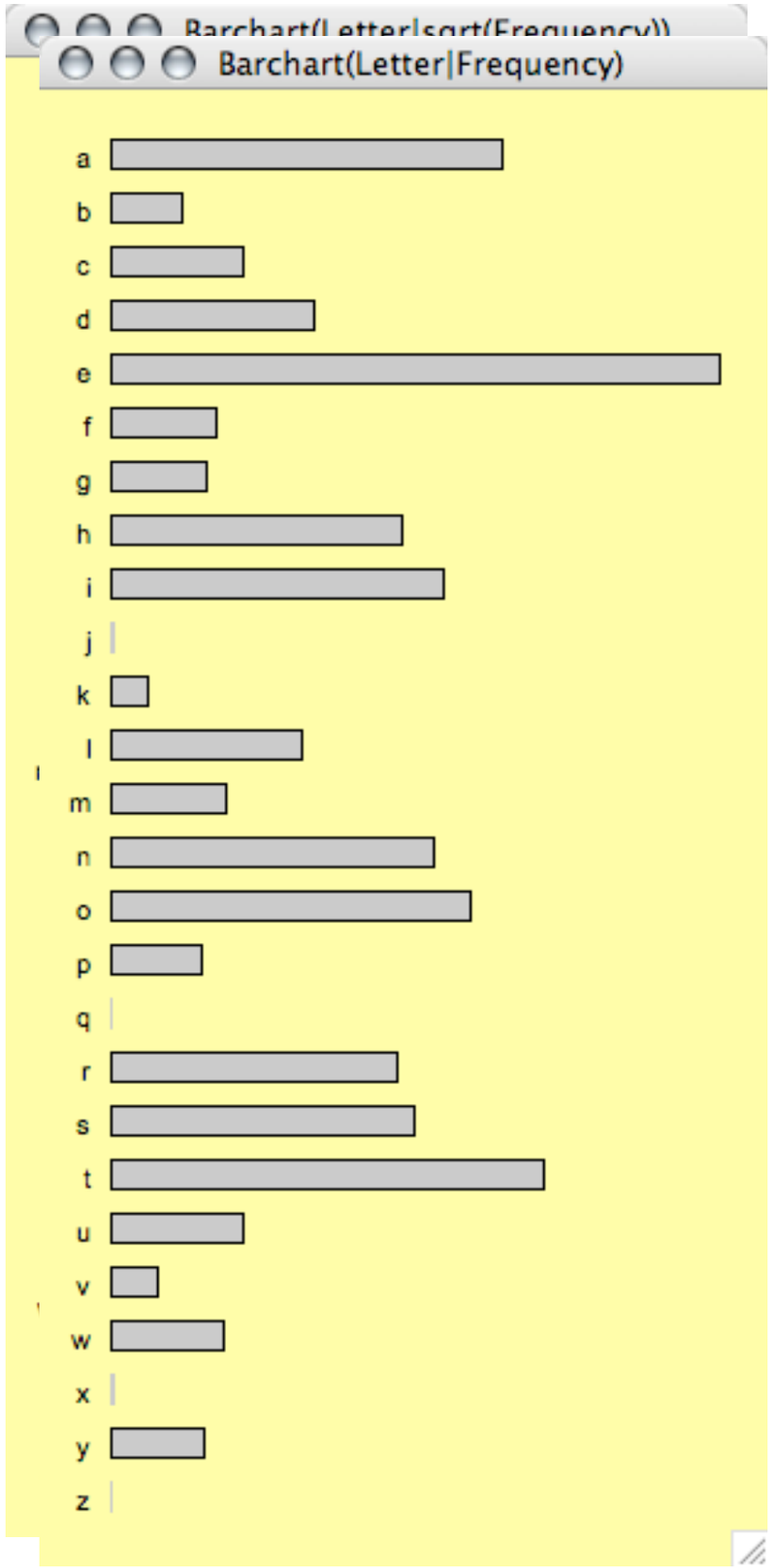




Some Mistakes ...

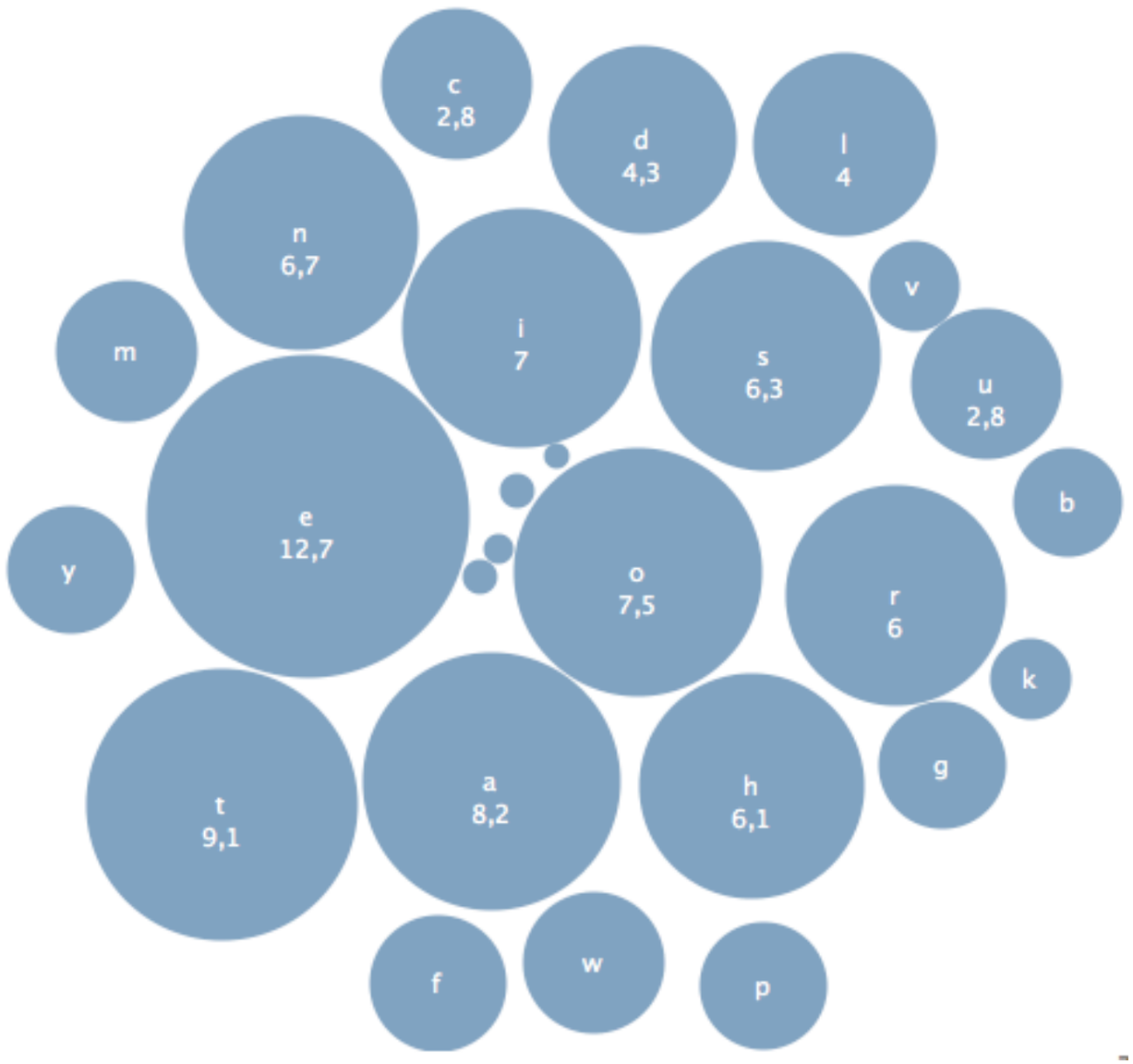
- … what is wrong with the bubble plot?

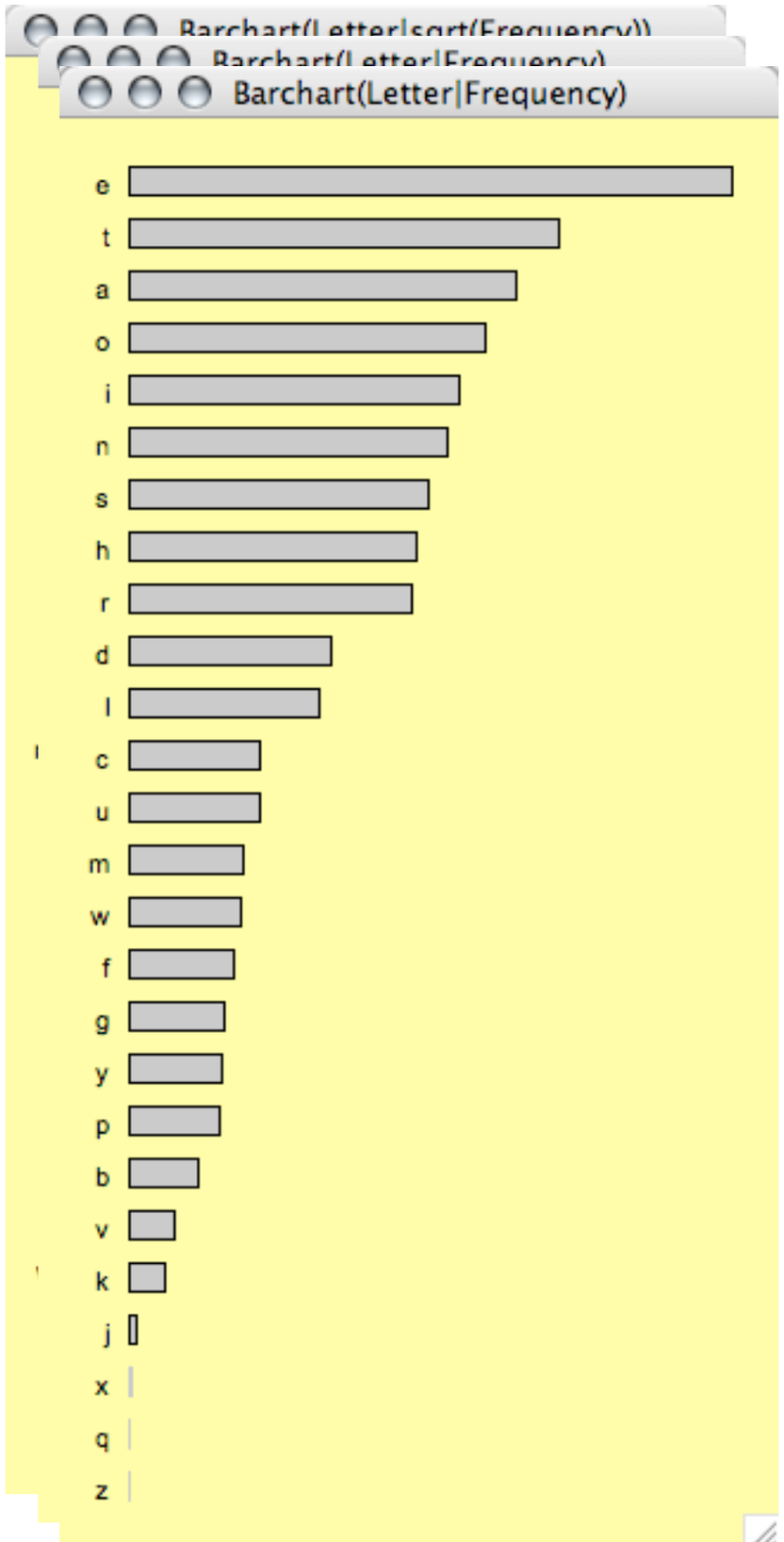




Some Mistakes ...

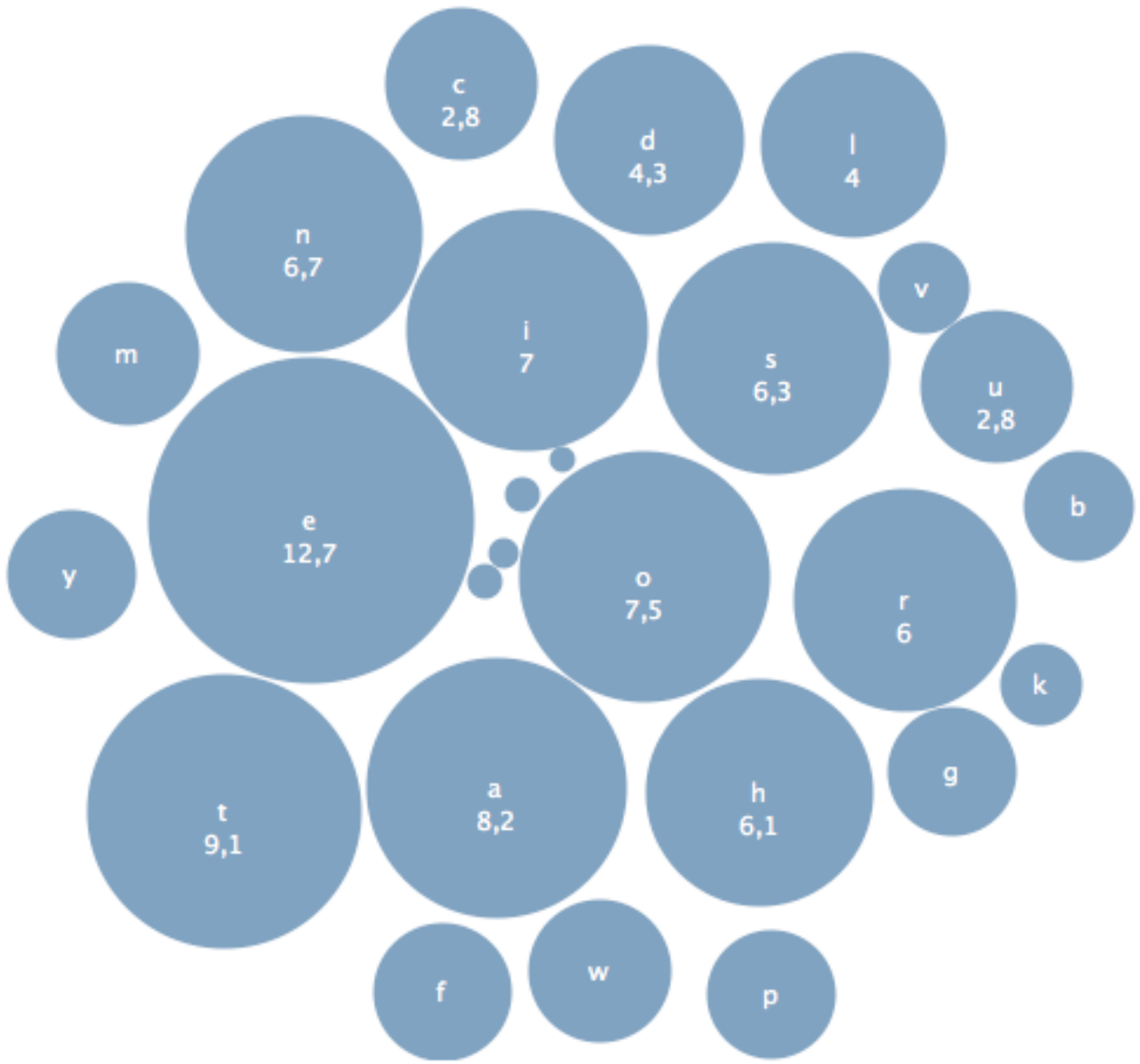
- … what is wrong with the bubble plot?



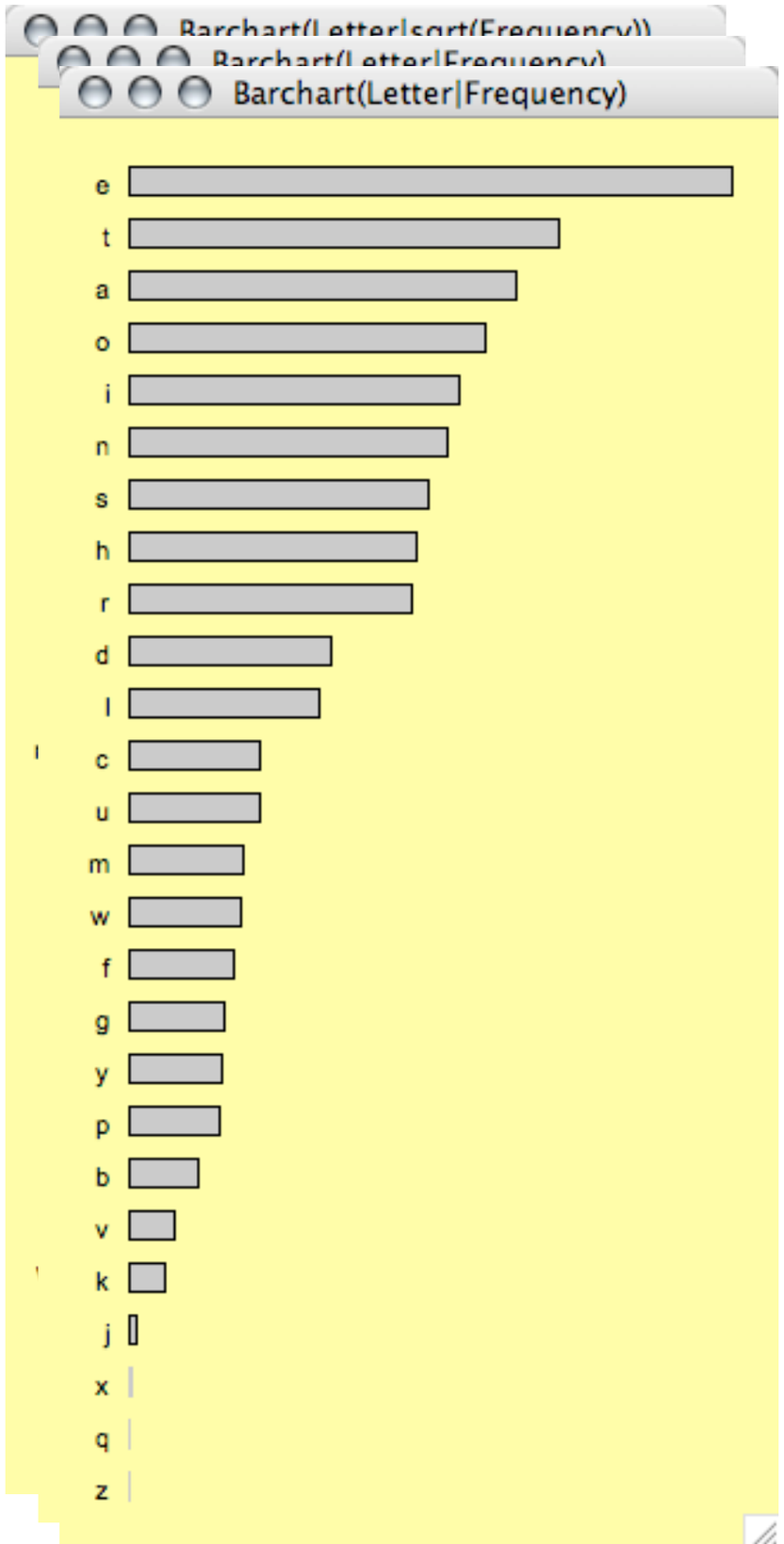


Some Mistakes ...

- … what is wrong with the bubble plot?

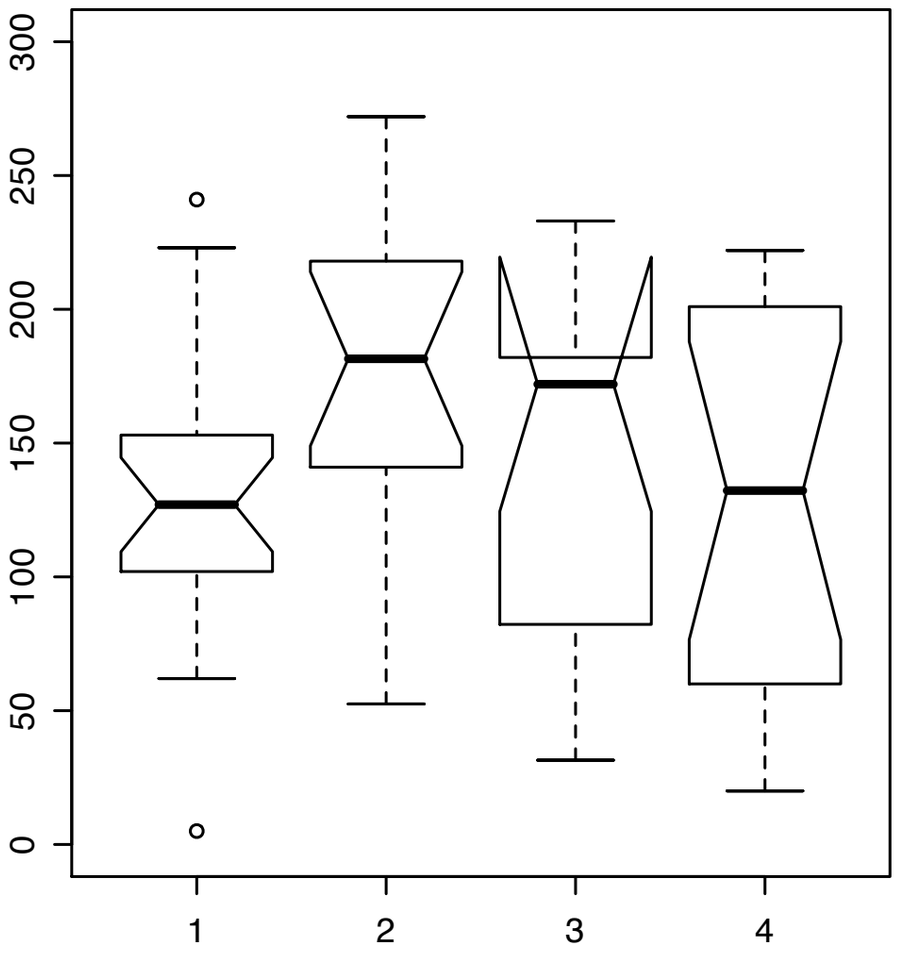


- we actually need to draw conclusions!

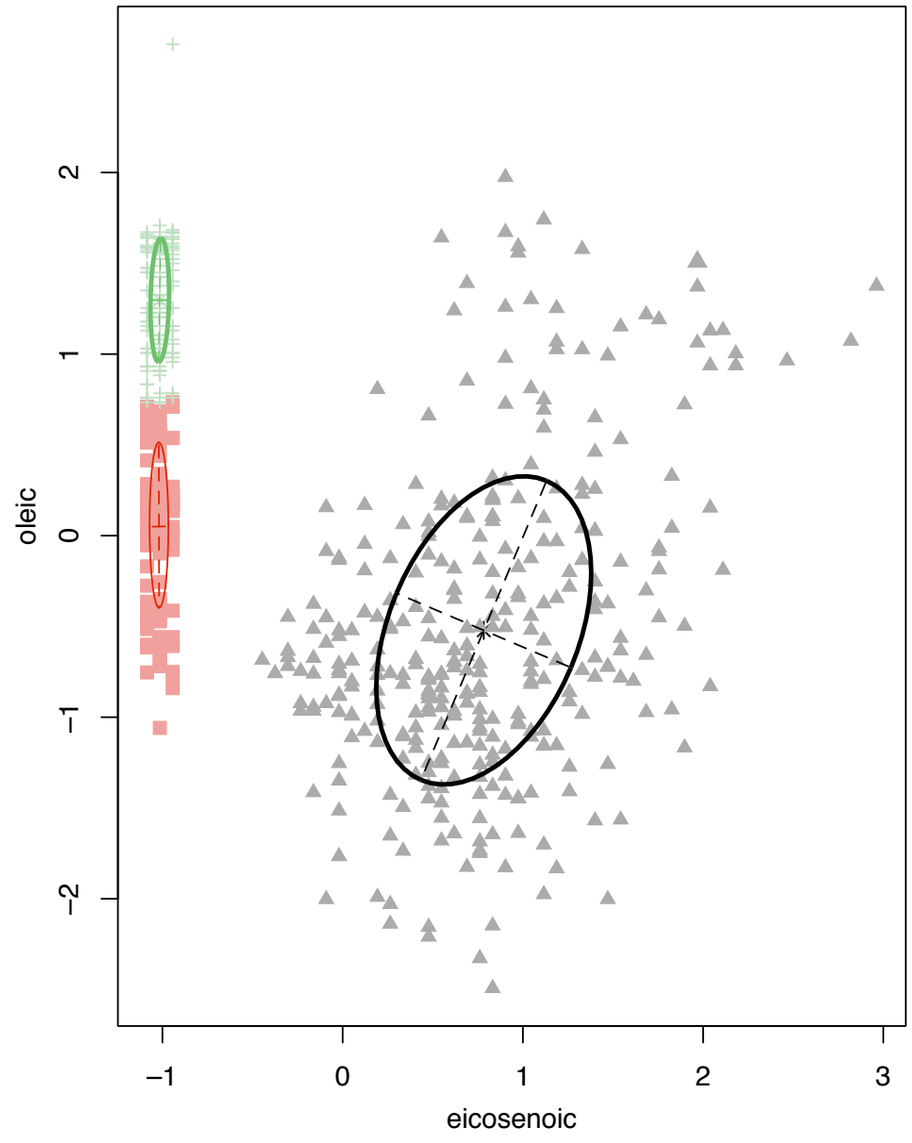


Visualization or Test?

Notched Boxplot

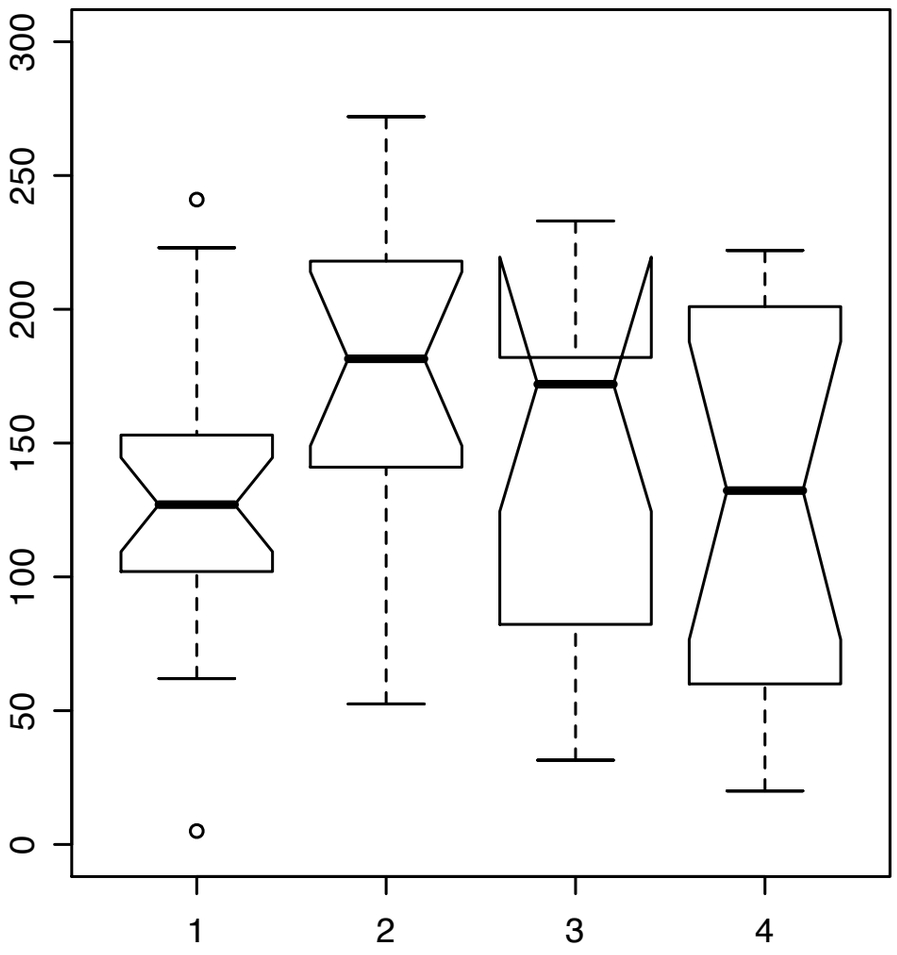


Model-based Clustering

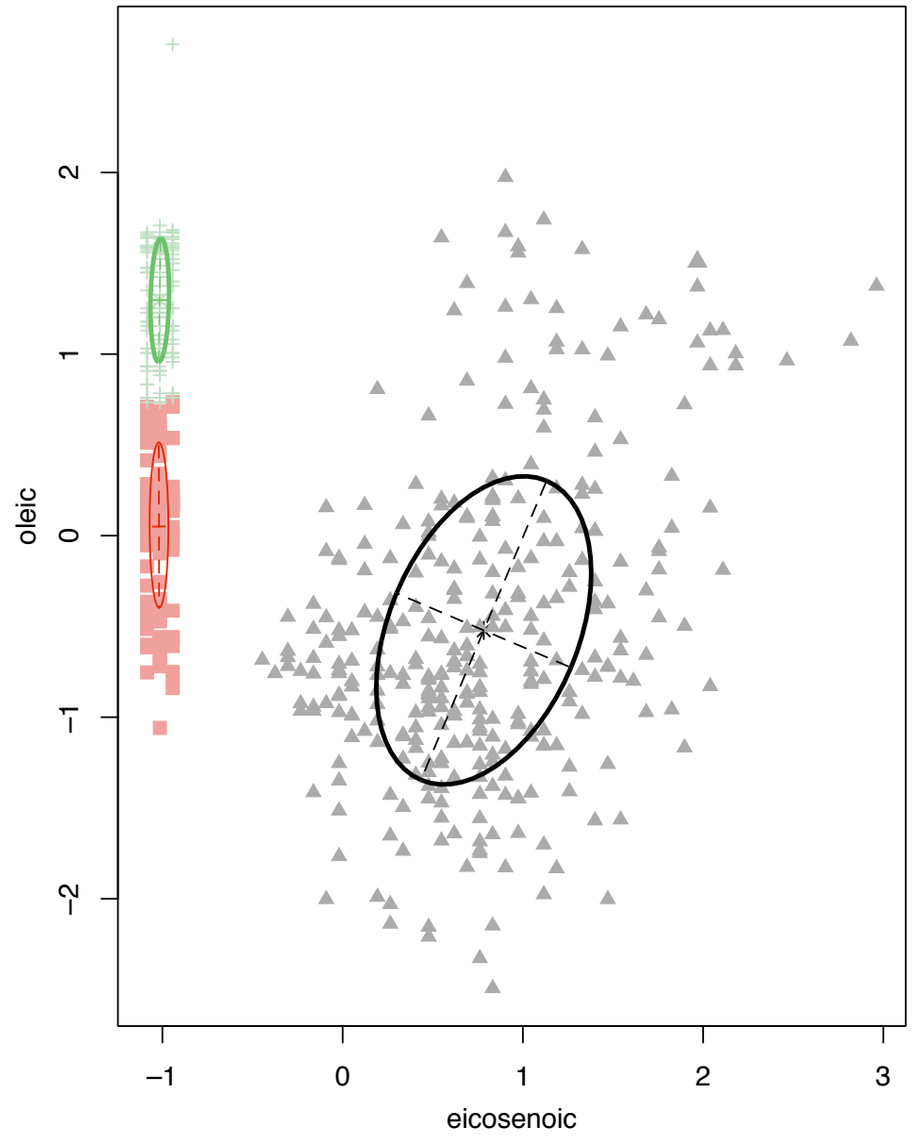


Visualization or Test?

Notched Boxplot



Model-based Clustering



Looking at graphics we are testing, building models and classifying over and over again …

Linked Views ✓

- Linked Views, aka
  - Coordinated Views
  - Synchronized Views
  - …

- The general principle is well known and explored, but
  - the variety of selection and highlighting methods (which usually stand at the ends of linked views) can still be expanded
  - some detangling of terms and concepts has to be done (there is less brushing around than most of us might think)

- Selection -> Linking -> Highlighting
  is looking at conditional distributions in short succession: testing and classifying again.

Interactions and User Interfaces ?

- Should be solved by now ...
  (did you read some Norman or Nielsen, or …?)

- Are key if someone should actually use our tools

- Still, most developer confuse their model with the user‘s model

- Testing the usability of interactive visualizations is still far from being standard – what standard?
  (from a statisticians point of view there is hardly anything that can be measured for a test here.)

- HCI is of some help, but not much ahead of us.

Building Blocks vs. All-in-One Solutions

- or in operational terms:
  Exploration Graphics vs. Presentation Graphics

- Things getting closer today, but the fact that one can use a visualization does not necessarily mean that one understands it.

- A well chosen static presentation may be far more effective than leaving the users out in the dark with compl[ex|icated] tools at hand.

- General principle for all of us:
  „Go from easy to hard, and don‘t start with the most complex view that might capture all informmation!“
  (does not contradict „overview first …“)

Where do we go from here …?

- We can avoid many dead ends when we look more at other people‘s work (inside and outside the community)
- Critics are important
  - if something is crap, call it crap!
  - if something can be improved, tell us how!
- Look for customers (problems) first, and than build your tools (classical business model; most companies who ignored this principle don‘t exist any more – but we are state-run …)
- There will never be a „global“ theory for InfoVis, but many things can be standardized and/or formalized locally.
- Question
  „Is InfoVis *just* a craft that coordinates HCI, GeoVis, DataVis/ StatGraphics, Cognitive Science, … with means of CS?“

Shameless Advertisement

- For those who didn‘t do it yet, stop by at

http://mondrian.theusRus.de

Problems solved in Mondrian 1997 ➟ 2007

- Selection
  - Going from strict to loose by more generality (40% → 90%)
- Highlighting
  - Understanding conditional distributions (50% → 80%)
- Managing large data displays
  - Introducing binning and alpha-transparency (50% → 90%)
- Going high-dim
  - Empowering PCPs and Mosaics (40% → 95%)
- Coping with missingness
  - add it! (0% → 100%)
- Leveraging statistics
  - Connecting to R via Rserve (0% → 20%)