Watching our figures

How mathematical psychology visualises cognitive modelling, and how it could improve

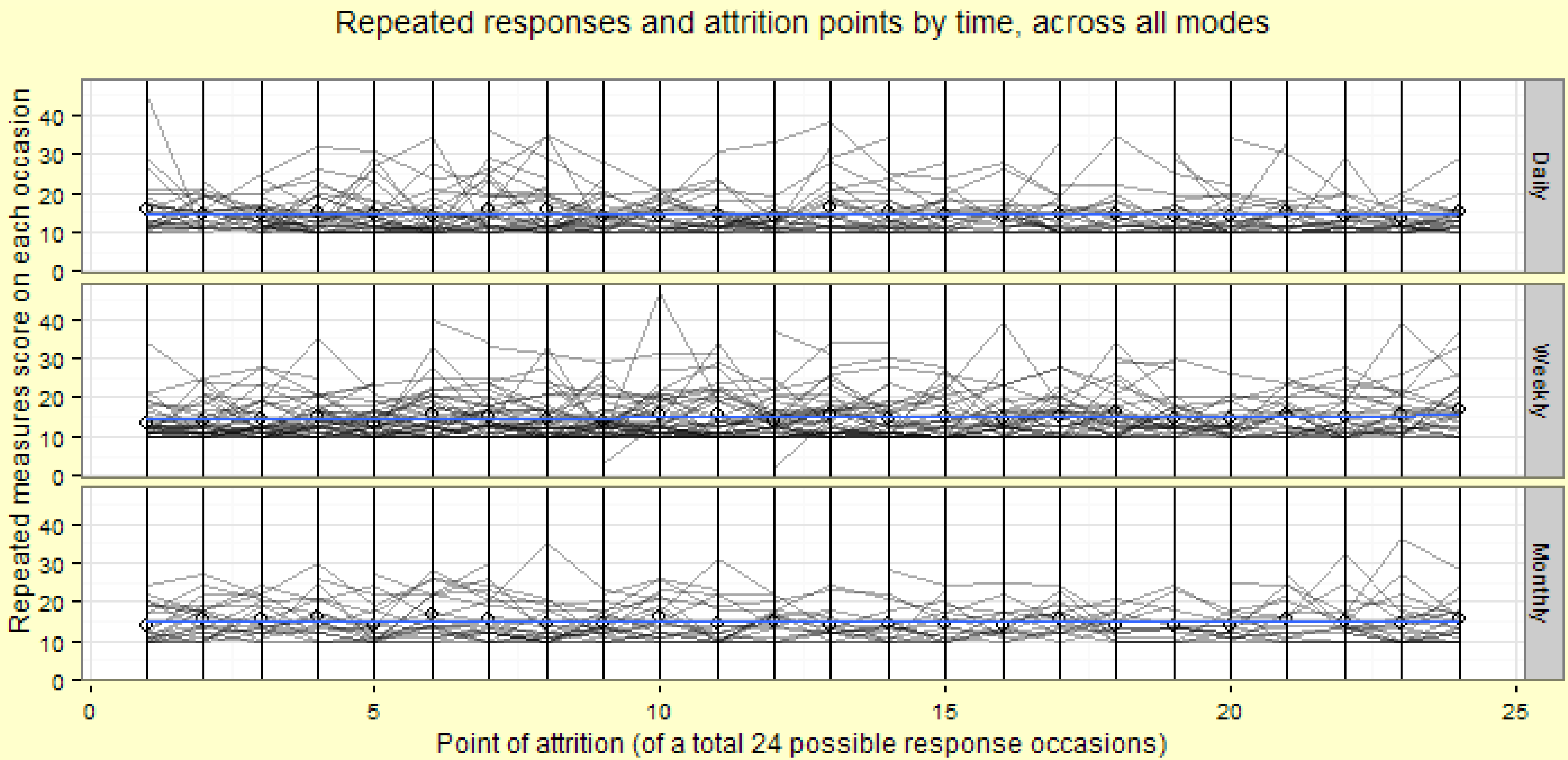
Erin Walsh

Australian National University

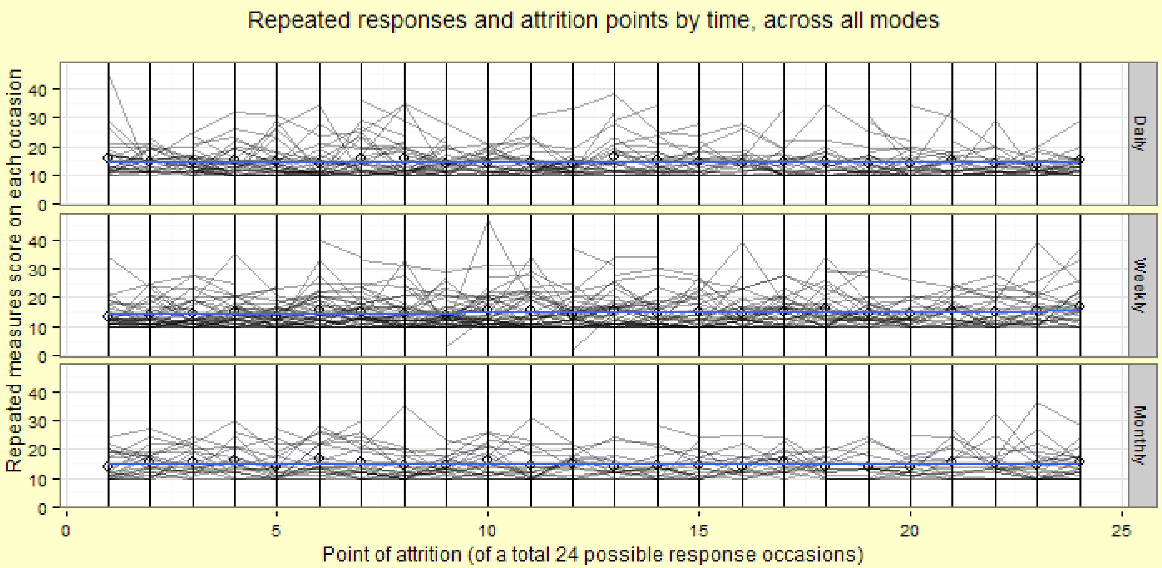
Importance of visualisations

- Visualisations simplify and clarify (Wood, 1994)
- **Scientific** visualisation should communicate information in the most efficient, unambiguous manner possible (Tufte & Graves-Morris, 1983)

Those in glasshouses...



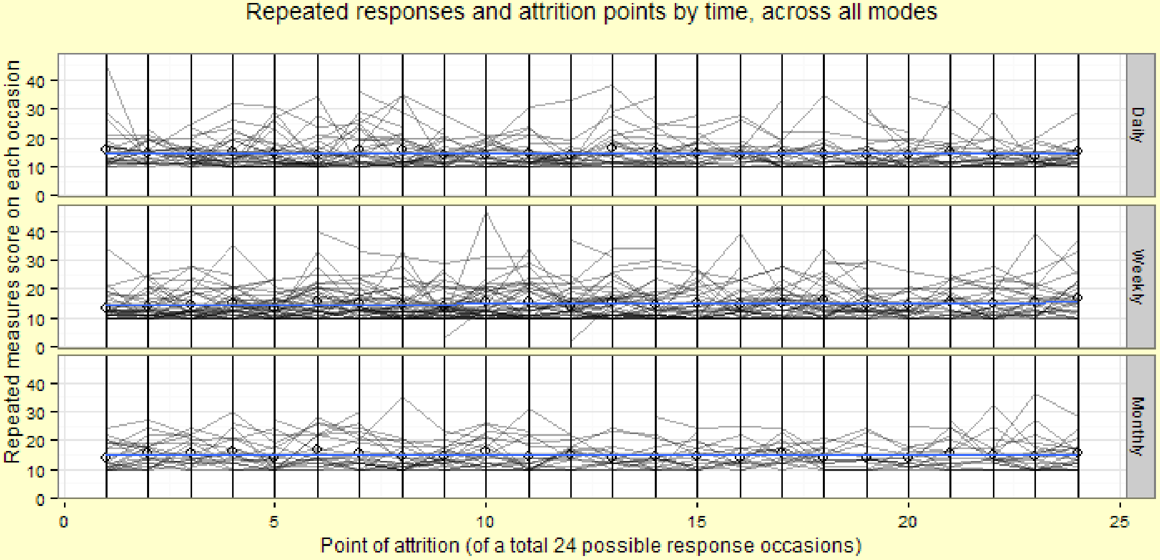
Those in glasshouses...



Even ggplot can't save this one

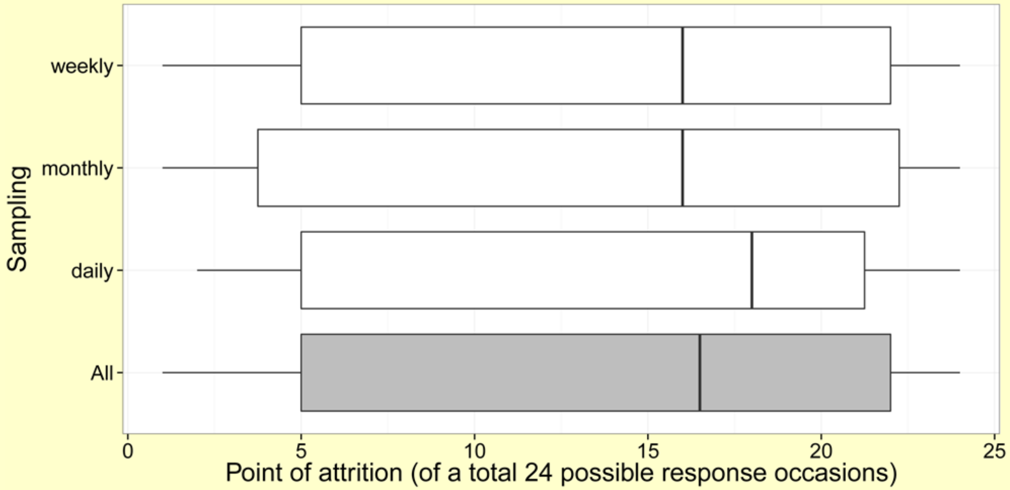
Watching Our Figures

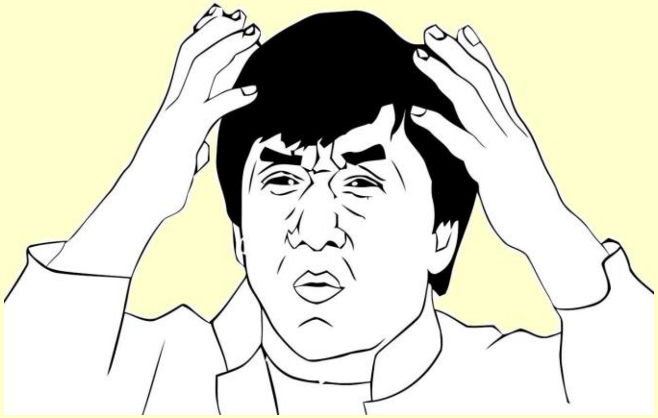
This is a hot mess.



Even ggplot can't save this one

Same information, better visualisation.

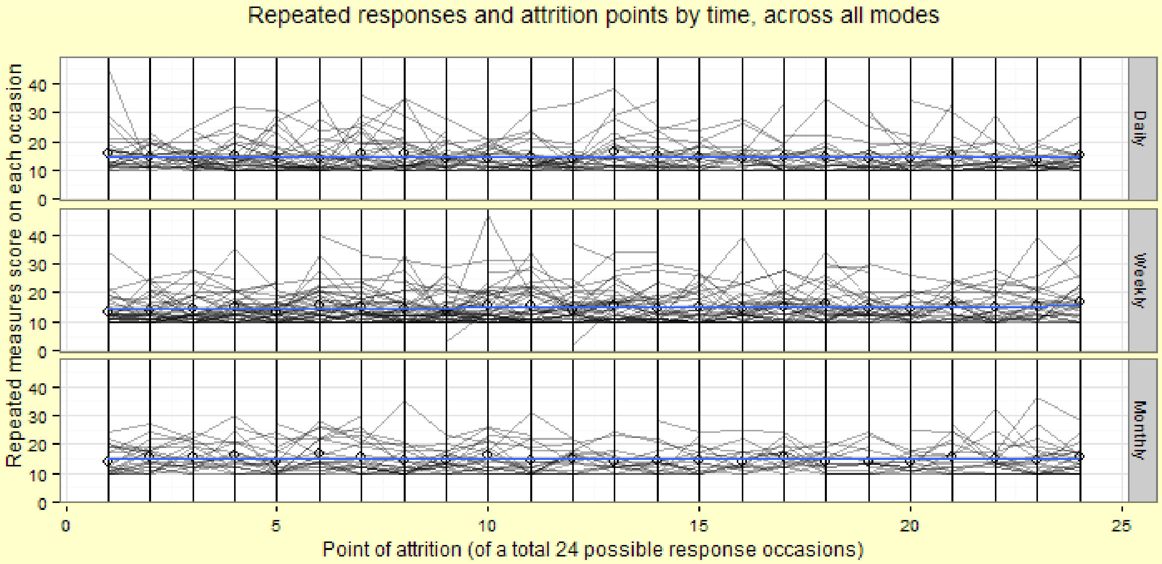




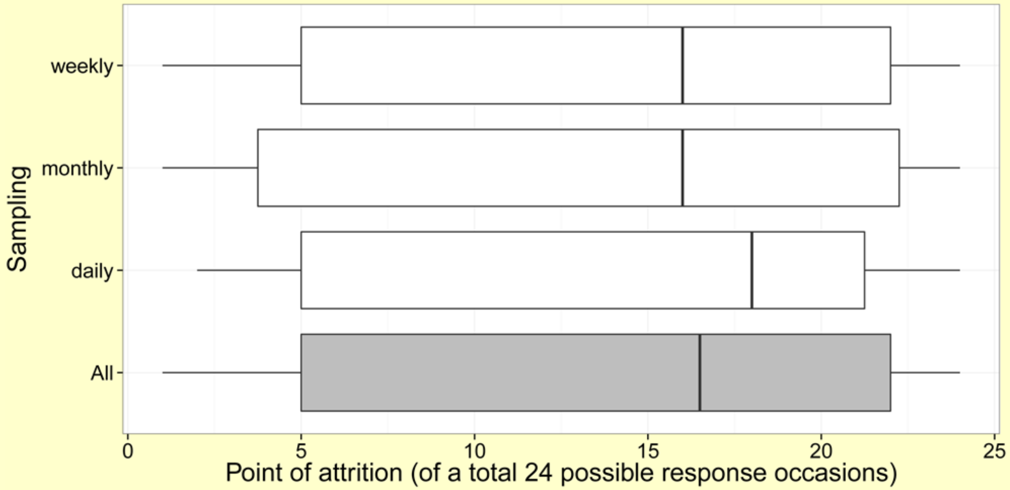
Why didn't you just say so!?

Watching Our Figures

This is a hot mess.



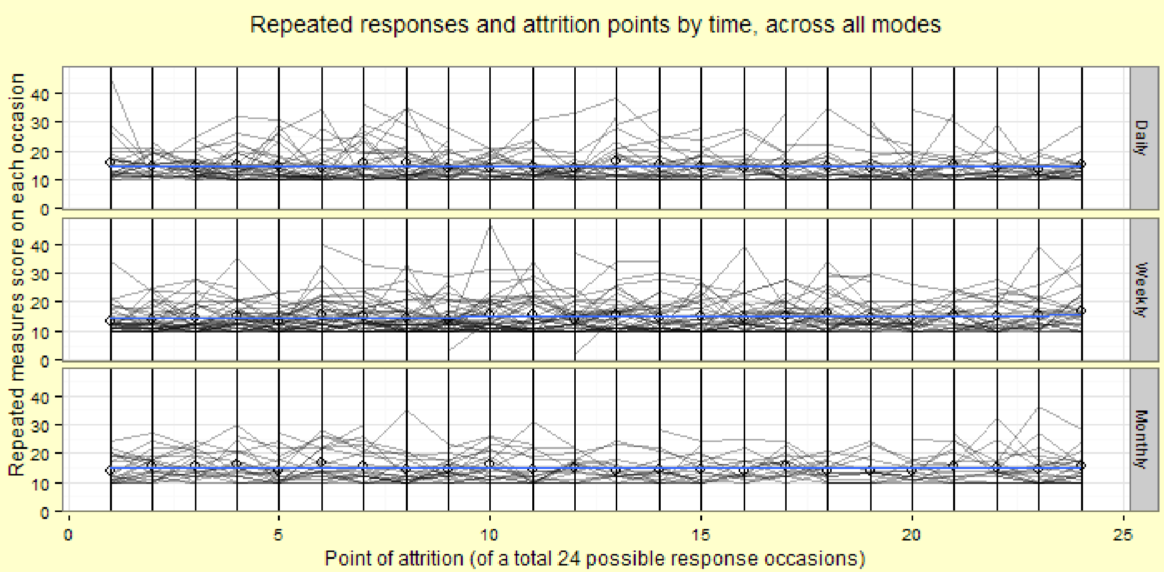
Same information, better visualisation.



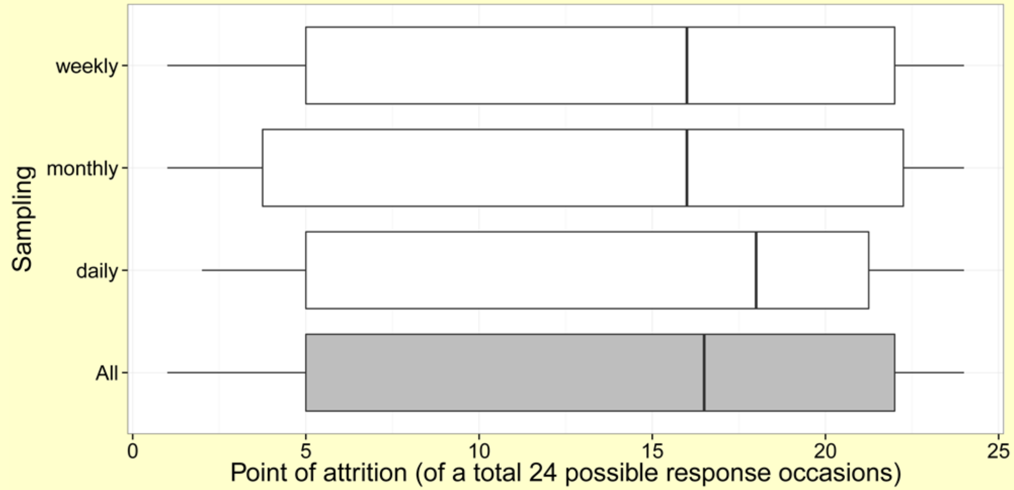
Description is important to clarify what works and what doesn't.

Watching Our Figures

This is a hot mess.



Same information, better visualisation.



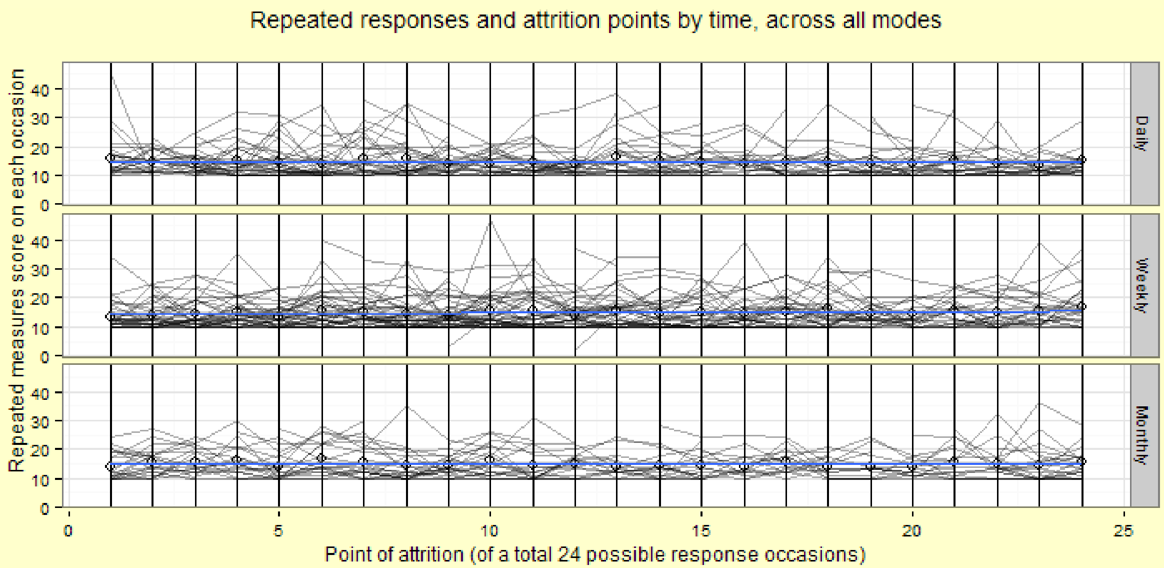
Description is important to clarify what works and what doesn't.

How should we describe these visualisations?

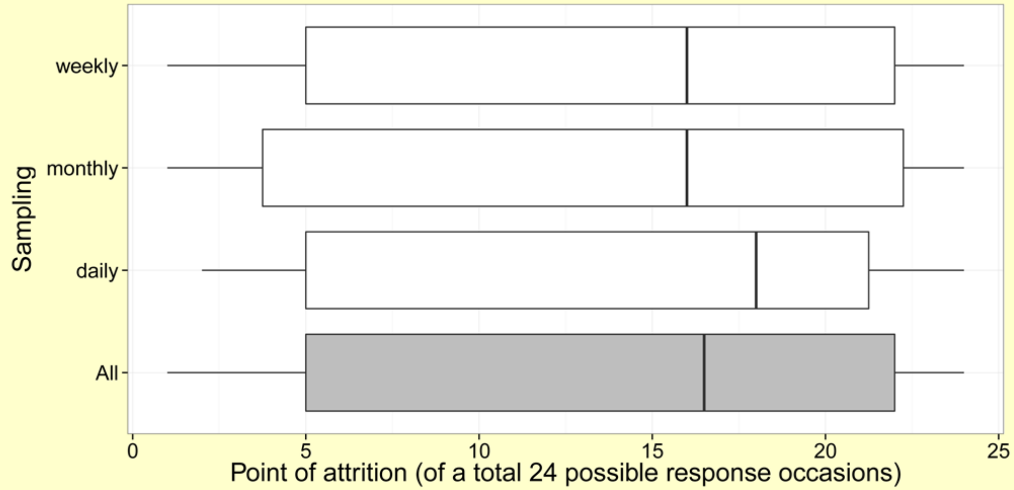
- Value judgement
  - Unsystematic, informal
- Surface features
  - Inconsistent, ungeneralisable

Watching Our Figures

This is a hot mess.



Same information, better visualisation.



Description is important to clarify what works and what doesn't.

- Just like scientific language, scientific visualisations have a form of syntax and grammar (Dimopoulos et al., 2003; Kelleher & Wagener, 2011; Mathai & Ramadas, 2009; Tversky, 2011).
- Visualisations are complex, so description schemes are also complex

- The science visualisation literature is fractured.

- The science visualisation literature is fractured.
  - Theoretical foundations, i.e.
    - Dual-code model (Levie, 1987)
    - Mental model construction (Glenberg & Langston, 1992; Hegarty & Just, 1993; Subramaniam & Padalkar, 2009)
    - Based on image appearance or function (Clark & Lyons, 2010)

Watching Our Figures

- The science visualisation literature is fractured.
  - Theoretical foundations, i.e.
    - Dual-code model (Levie, 1987)
    - Mental model construction (Glenberg & Langston, 1992; Hegarty & Just, 1993; Subramaniam & Padalkar, 2009)
    - Based on image appearance or function (Clark & Lyons, 2010)
  - Many aren't clearly operationalised (Clark & Lyons, 2010; Goldsmith, 1987).
  - Focus more on how to construct a visualisation, but not how to describe it

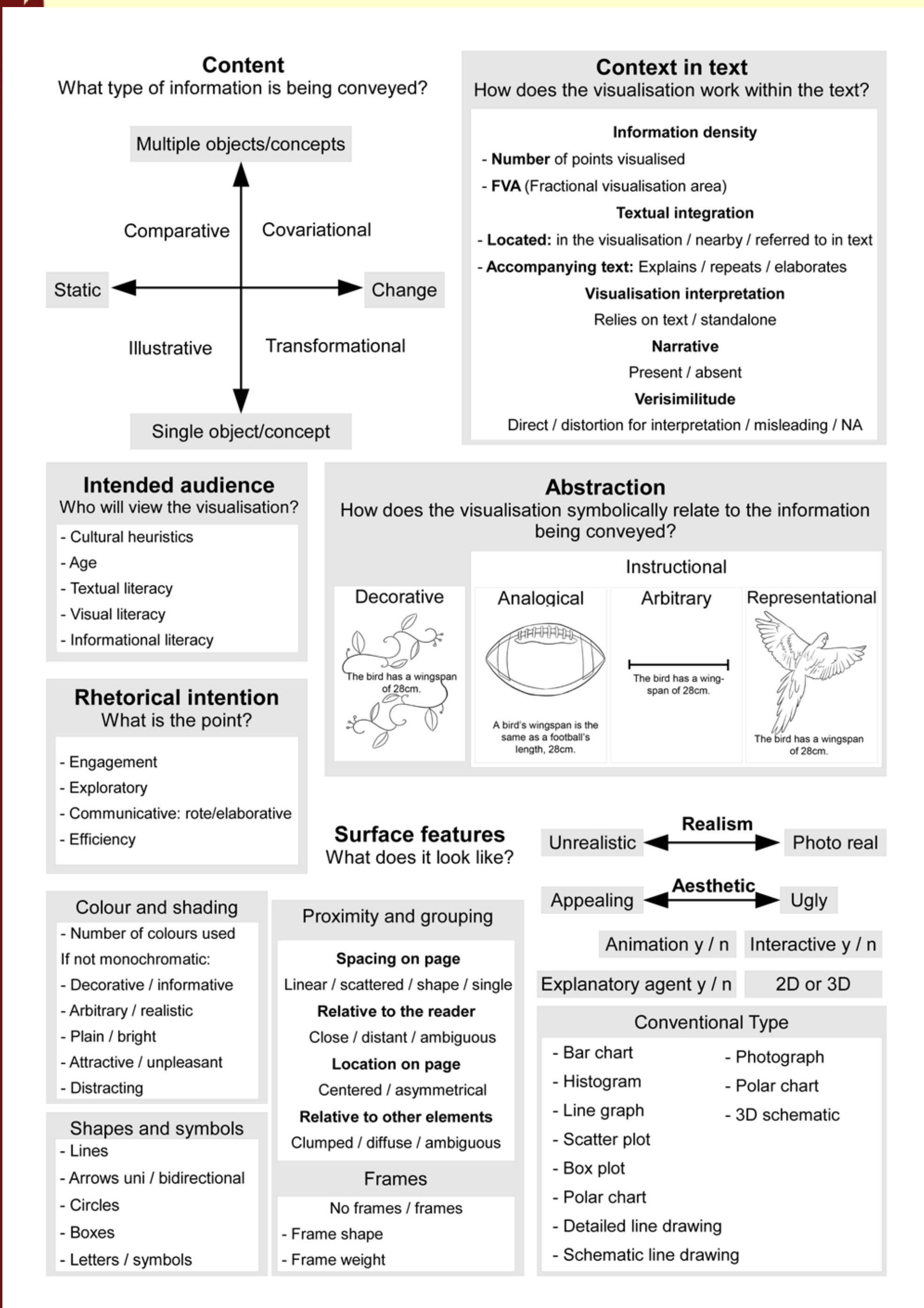
Typologies

- The current state of typologies
  - Typologies based on a grab-bag of descriptors, i.e. Descriptive lists by Clark & Lyons, 2010
  - Lack theoretical basis
  - Tend to be artificially reductionist

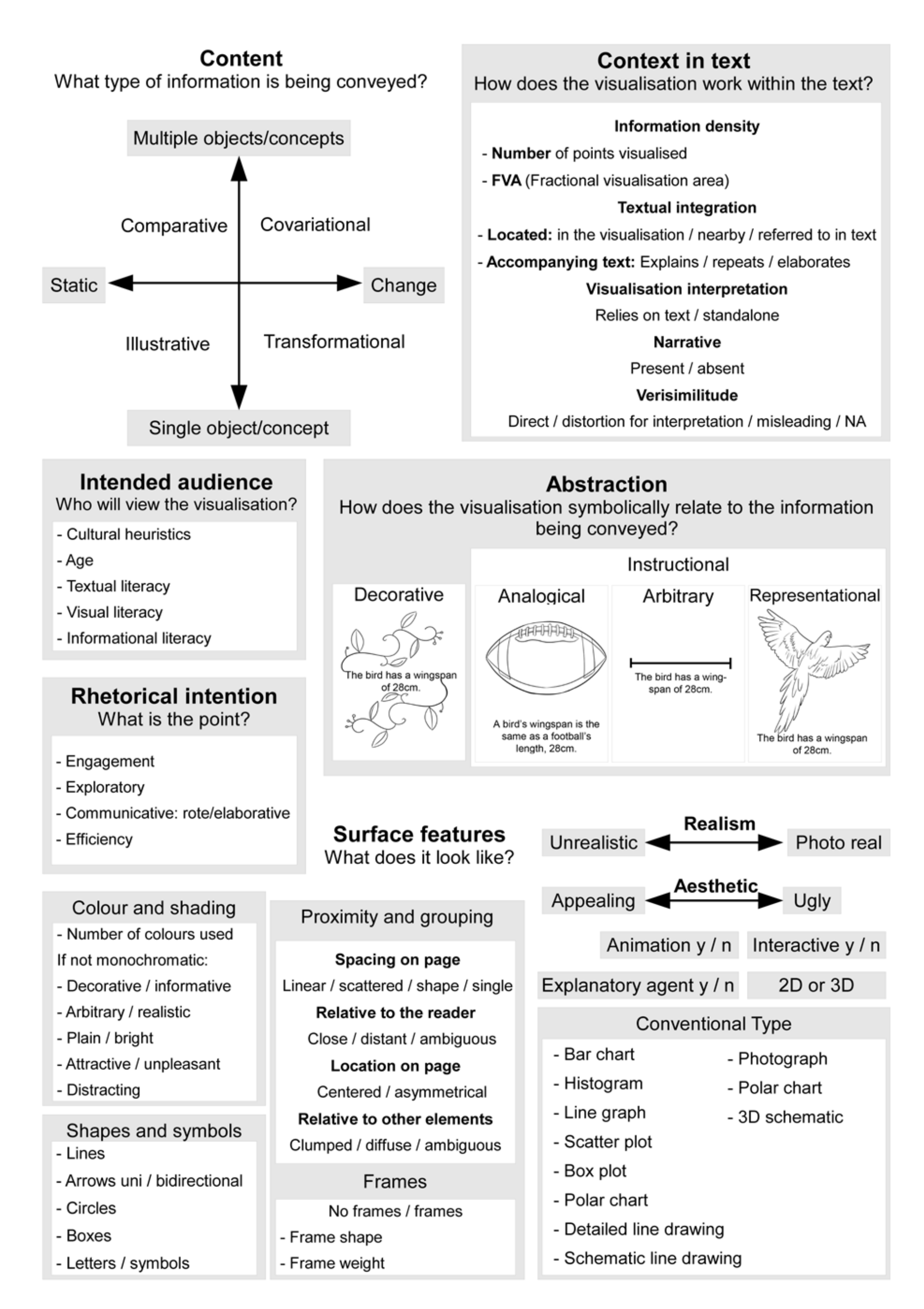
Typologies

- The current state of typologies
  - Typologies based on a grab-bag of descriptors, i.e. Descriptive lists by Clark & Lyons, 2010
  - Lack theoretical basis
  - Tend to be artificially reductionist
- Theories don't talk to each other
- Descriptors are artificially specific
- Synthesis of multiple viewpoints needed (Gahegan, 1999)

The CCAIRS typology

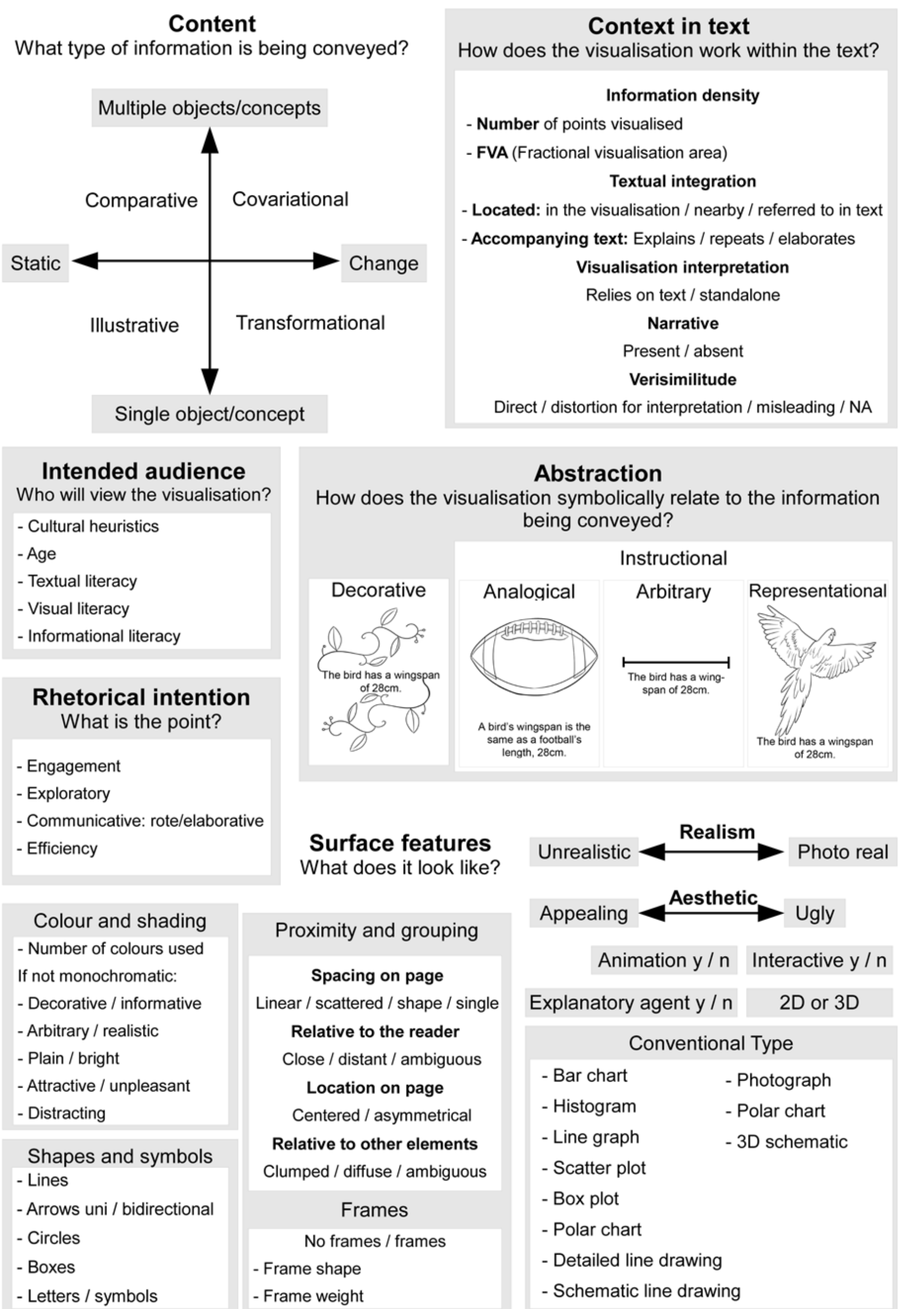


Watching Our Figures



The CCAIRS typology

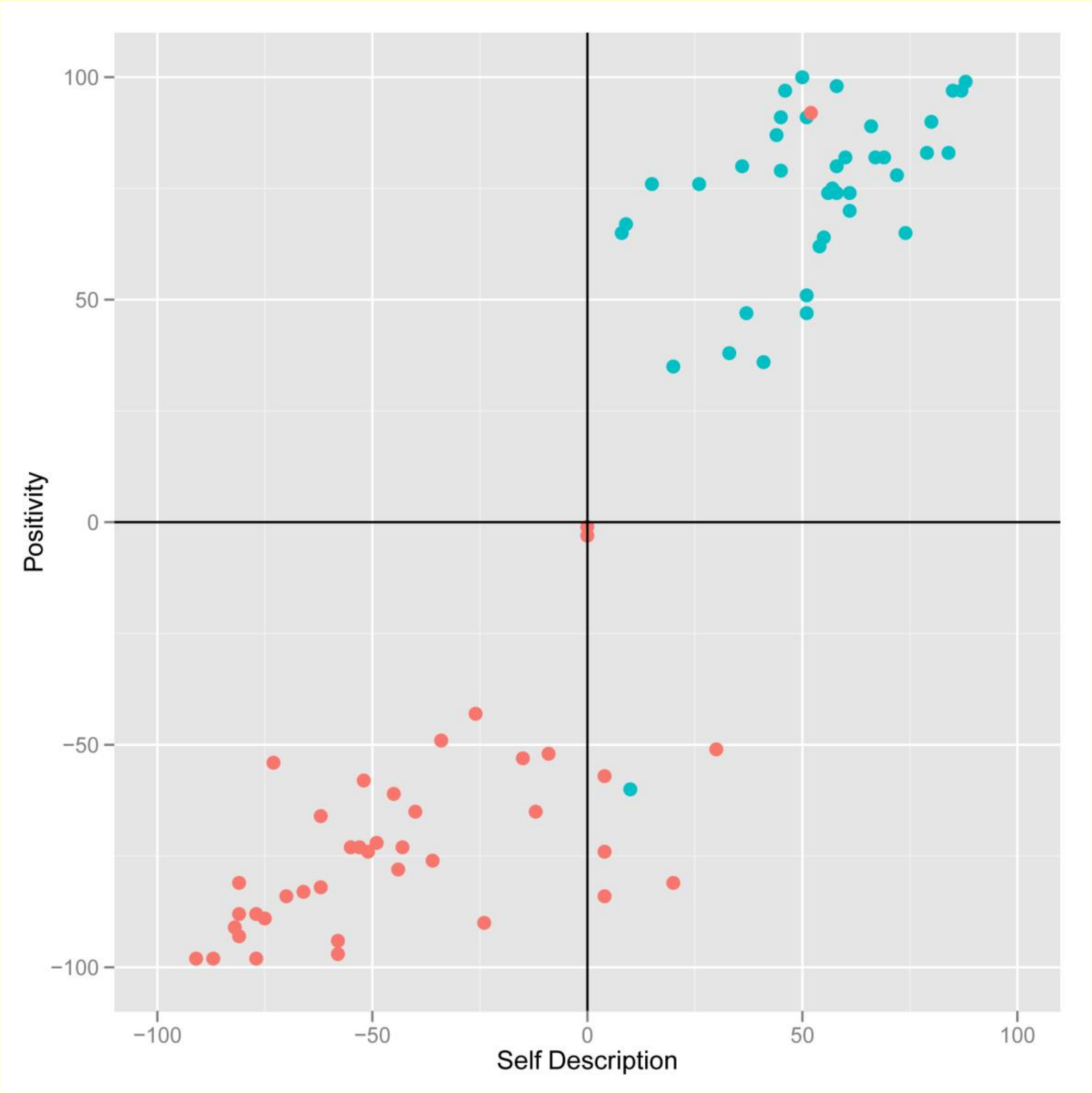
- Qualitative coding scheme for visualisations
  - Content
  - Context in text
  - Abstraction
  - Intended Audience
  - Rhetorical Intention
  - Surface features



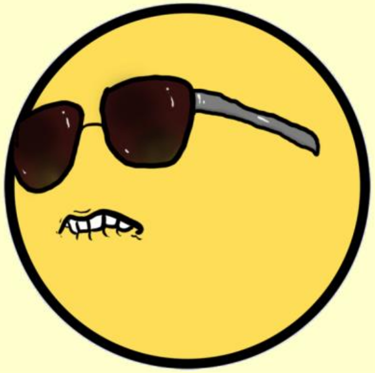
The CCAIRS typology

- Draws from psychology (from perception through persuasion theory), science communication, advertising, and art theory
- Combines a number of descriptor-based typologies
- Explicitly built to be practically applied as a descriptive tool
- BROADLY applicable

The CCAIRS typology



Still some subjectivity



'Dat graph!

Method

- Select some likely journals
  - Psychometrika
  - The Journal of Mathematical Psychology

Method

- Select some likely journals
  - Psychometrika
  - The Journal of Mathematical Psychology
- Discard any paper without a visualisation
- Discard any paper on an irrelevant topic
- Say farewell to friends and loved ones (the next bit will take a while)
- Isolate pages with visualisations

Method

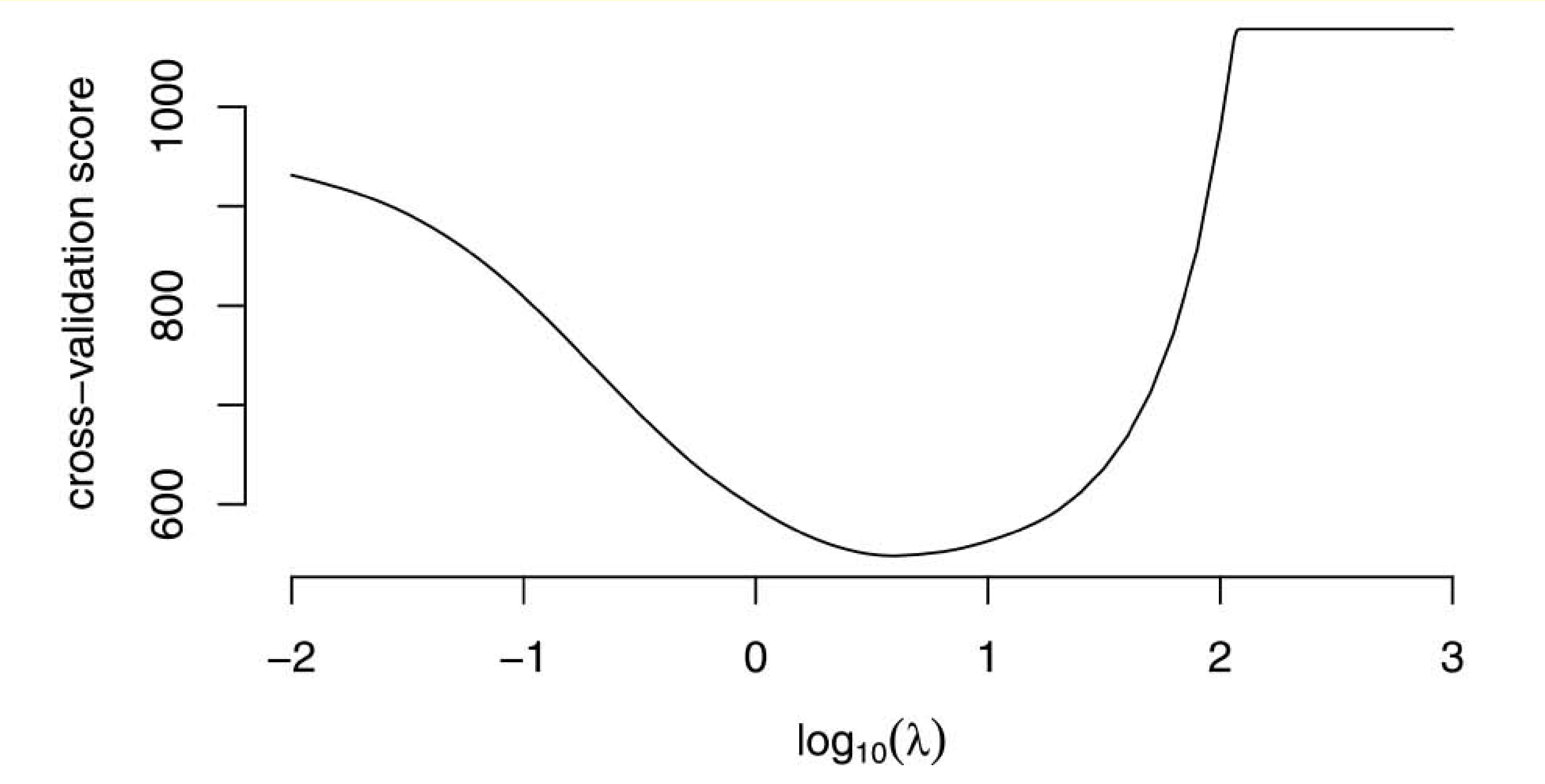
- Select some likely journals
  - Psychometrika
  - The Journal of Mathematical Psychology
- Discard any paper without a visualisation
- Discard any paper on an irrelevant topic
- Say farewell to friends and loved ones (the next bit will take a while)
- Isolate pages with visualisations
- Systematically apply the CCAIRS to the visualisations

Method

- Select some likely journals
  - Psychometrika
  - The Journal of Mathematical Psychology
- Discard any paper without a visualisation
- Discard any paper on an irrelevant topic
- Say farewell to friends and loved ones (the next bit will take a while)
- Isolate pages with visualisations
- Systematically apply the CCAIRS to the visualisations
- Random (ish) sample of 150 images from 2000-2014

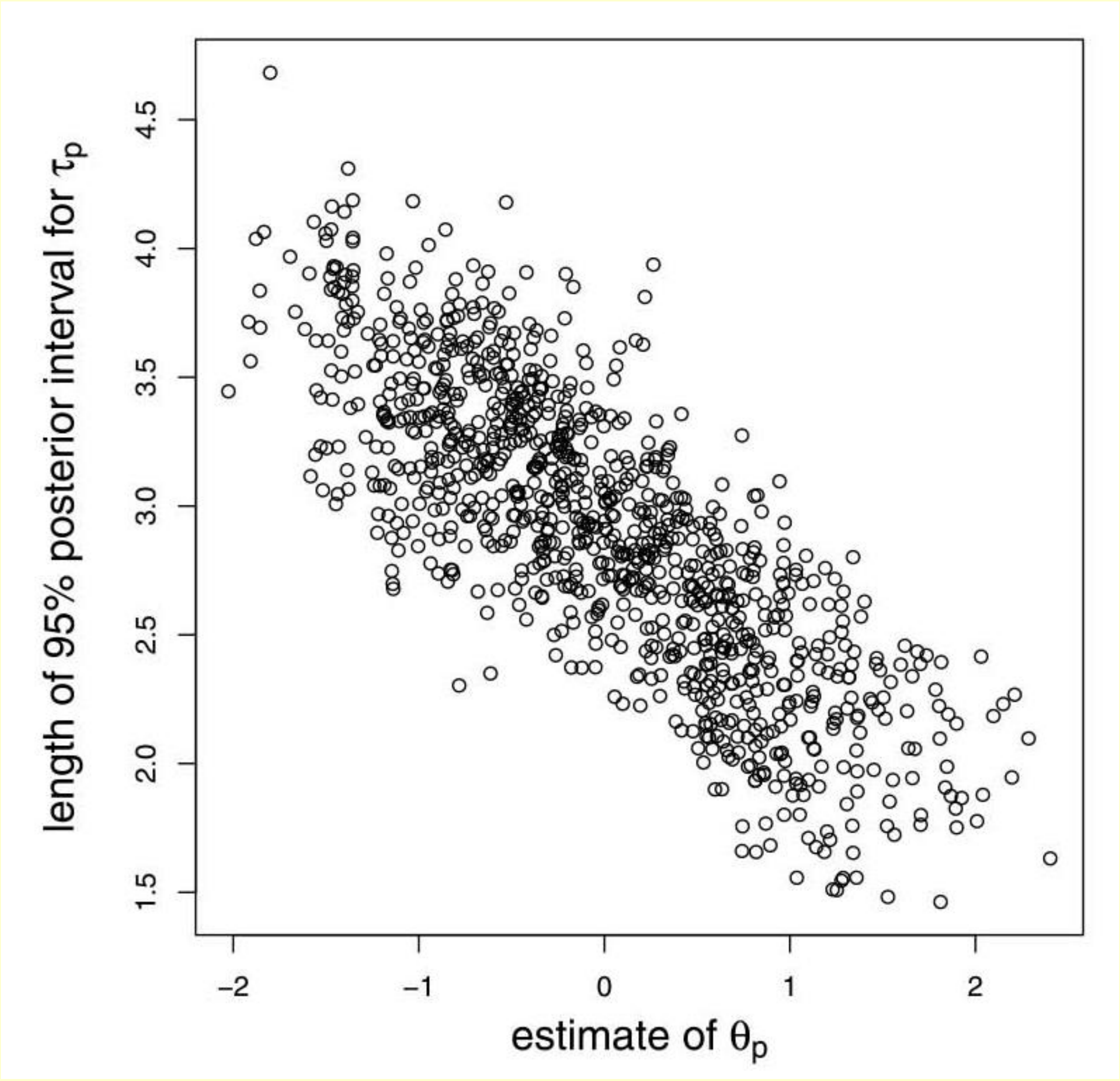
Common visualisation types

Line (73%)



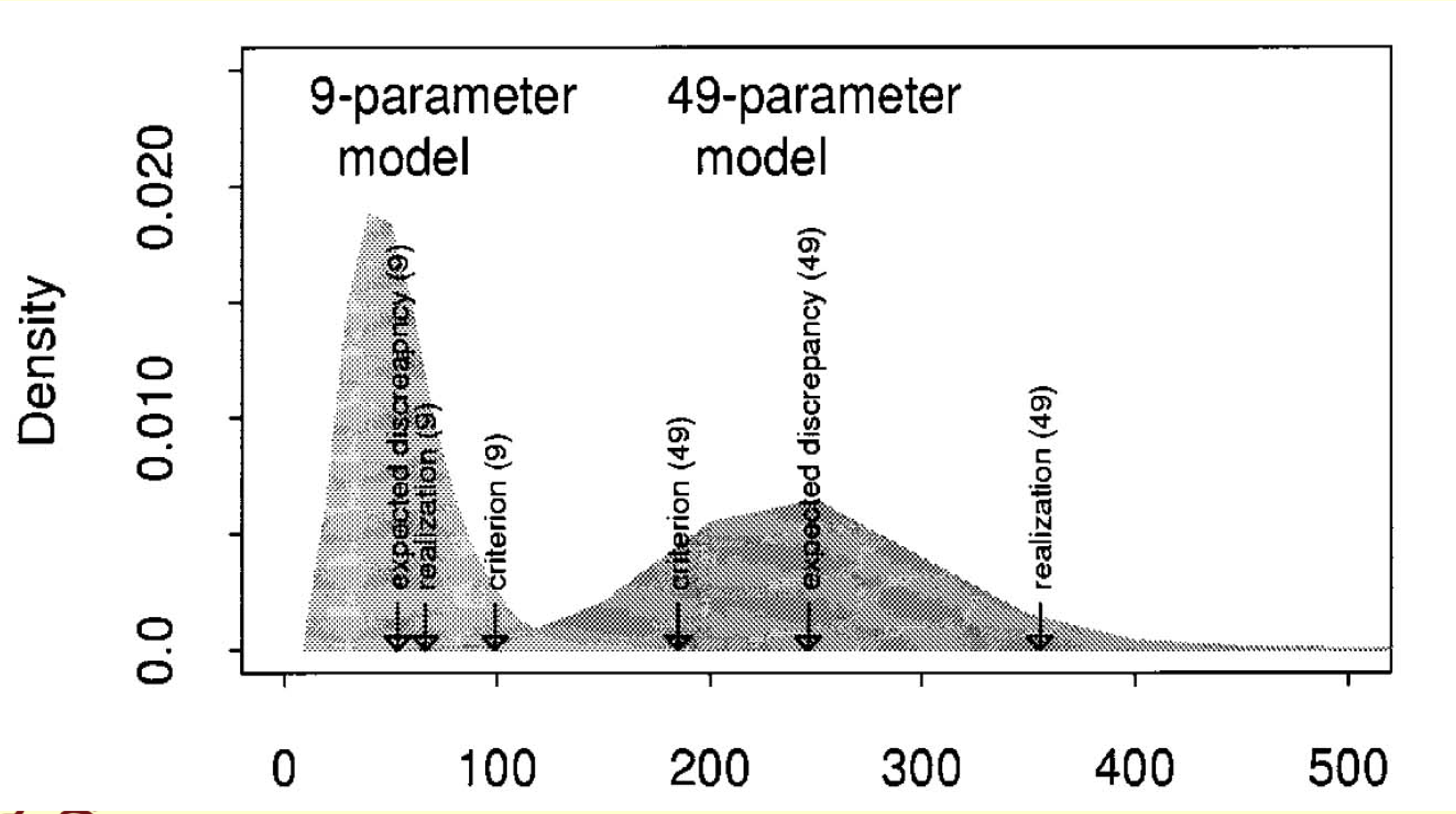
Common visualisation types

Scatter (22%)



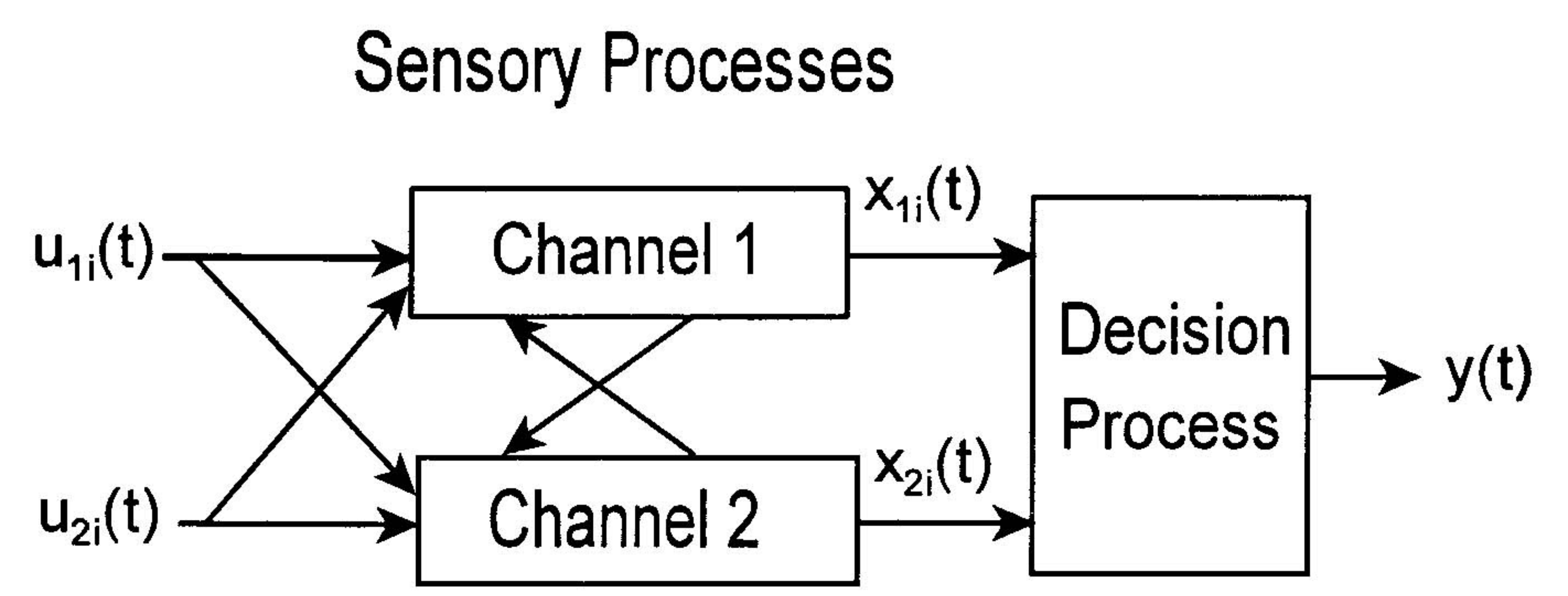
Common visualisation types

Density (17%)



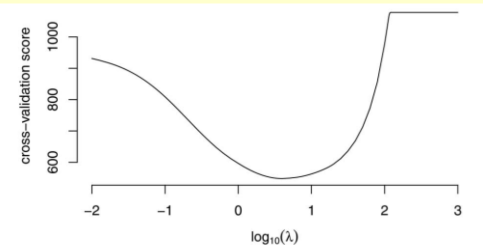
Common visualisation types

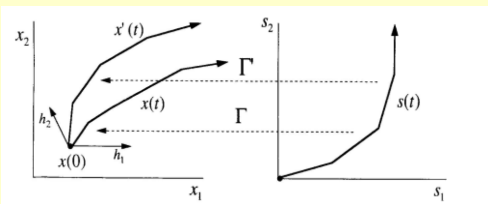
Structural (16%)

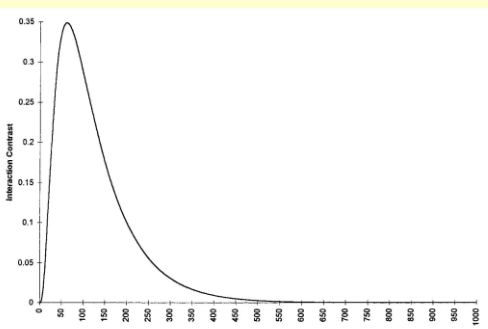


Common visualisation types

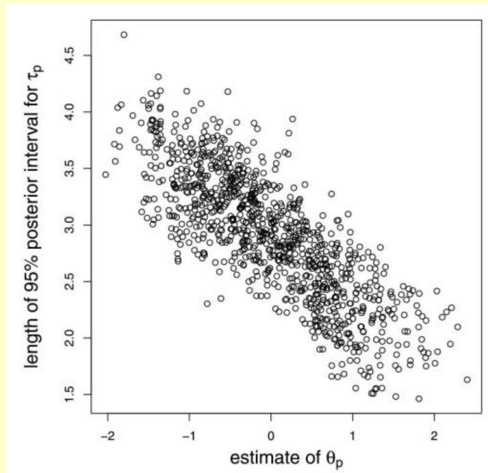
Line (73%)

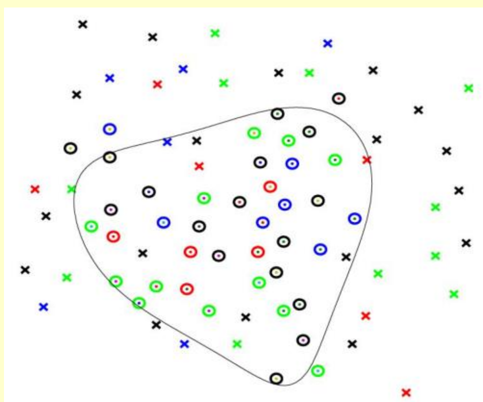






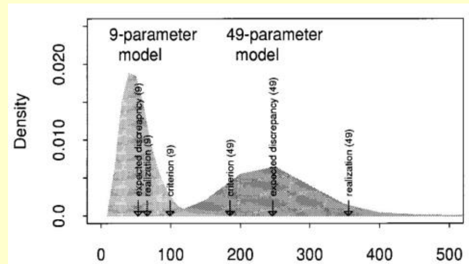
Scatter (22%)

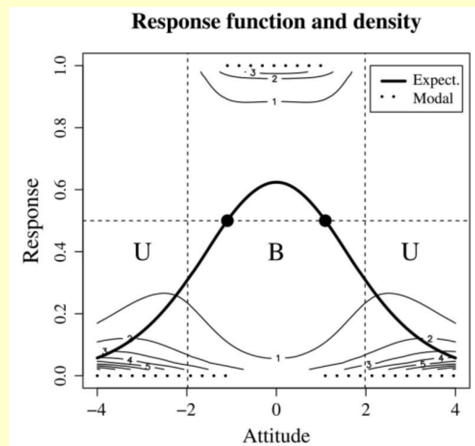




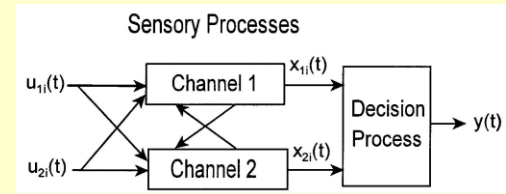
(f) All Sites (Controlled for Site Effect)

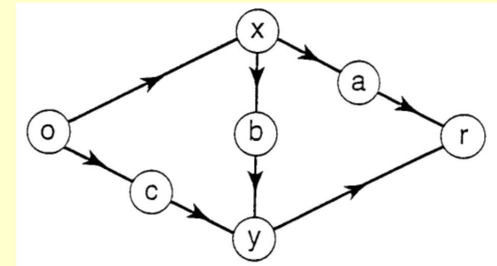
Density (17%)

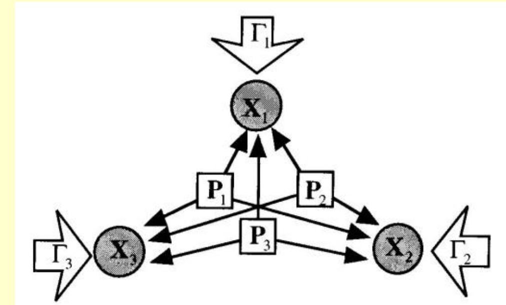




Structural (16%)

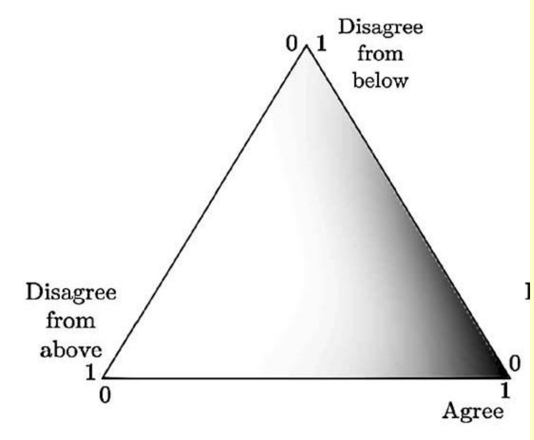




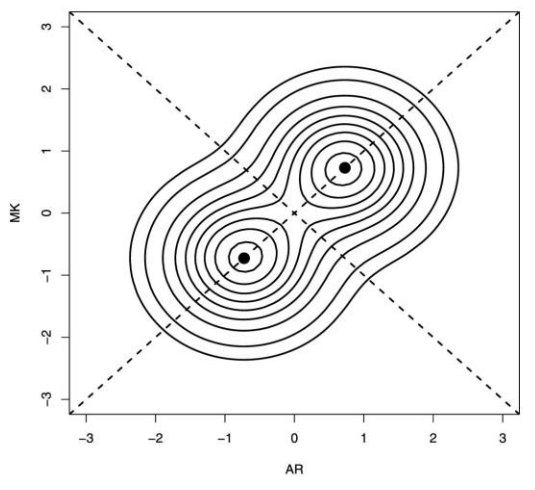


Uncommon visualisation types

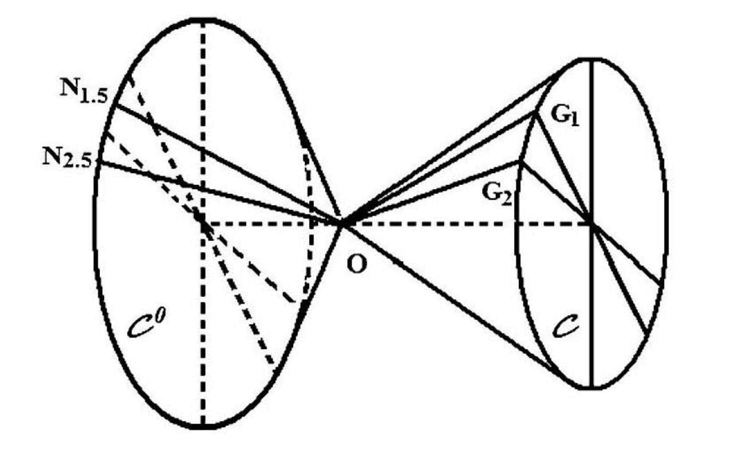
Ternary



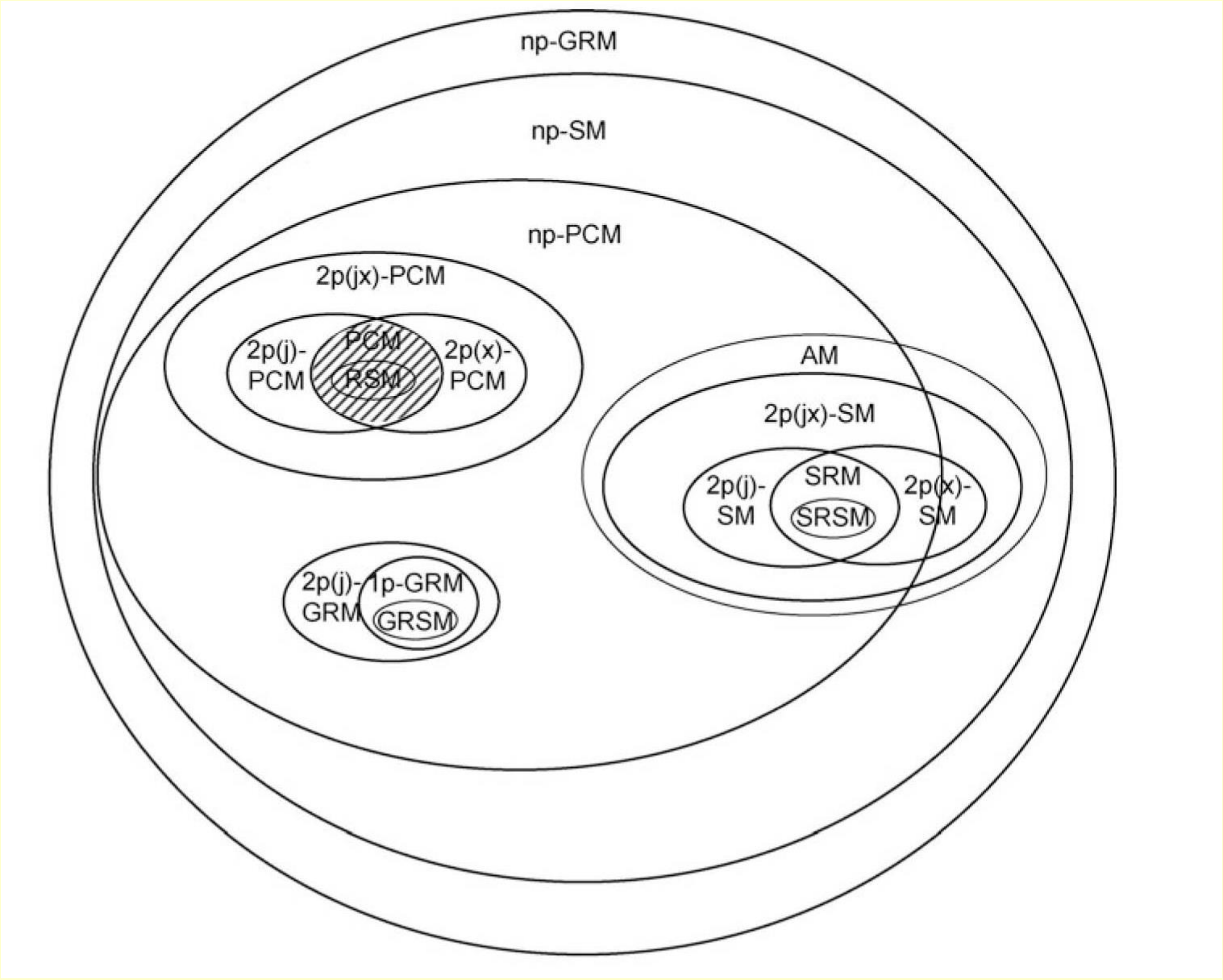
Contour plot



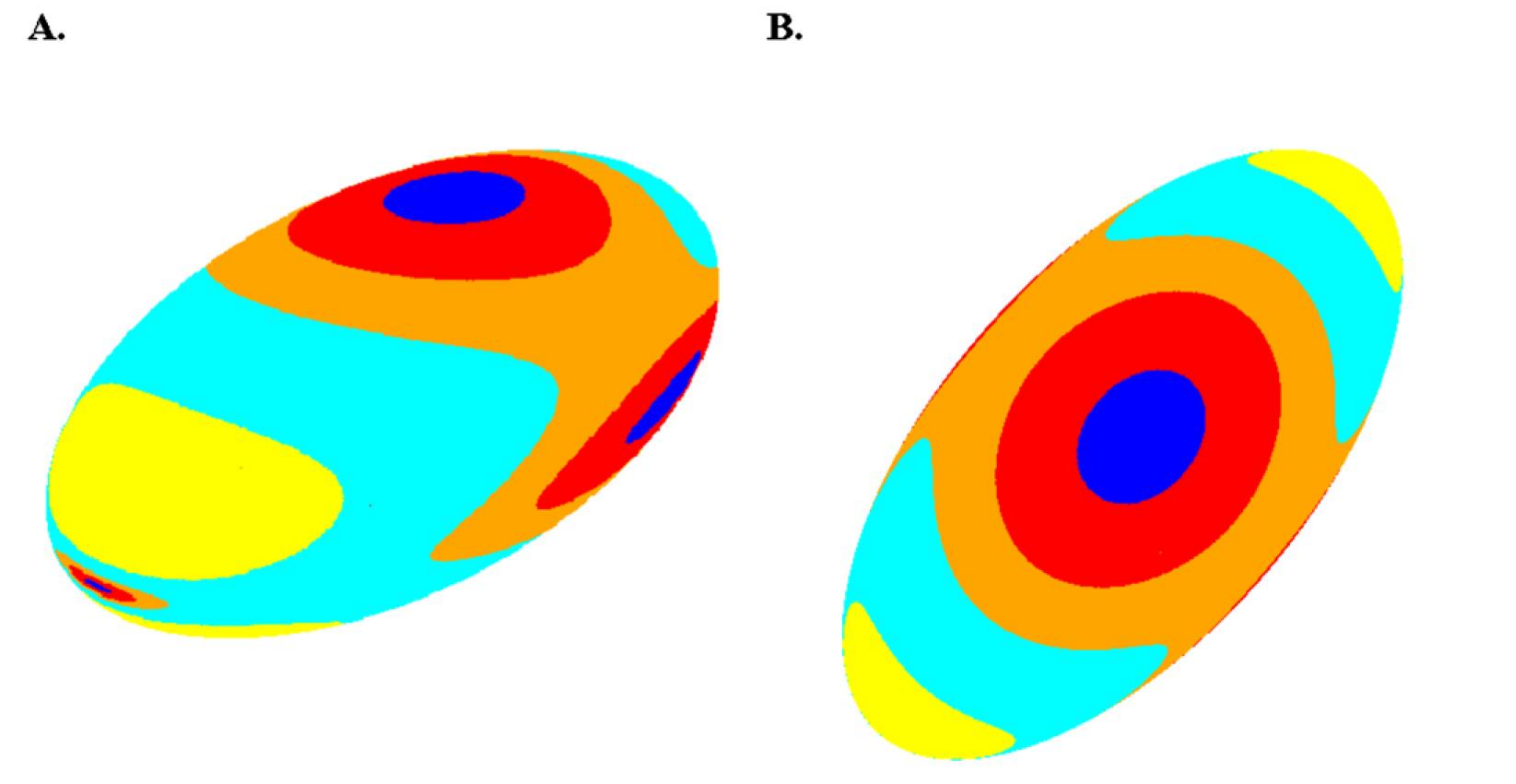
Polyhedral cone



Uncommon visualisation types

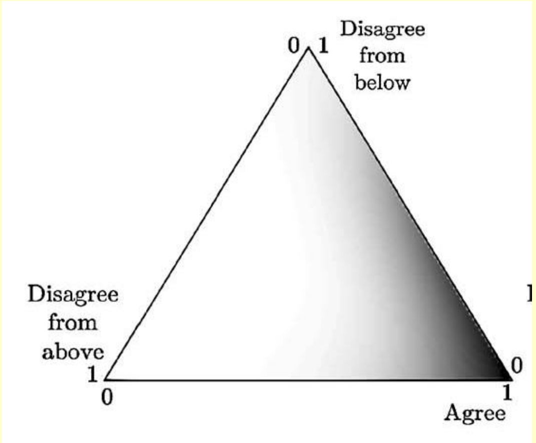


Uncommon visualisation types

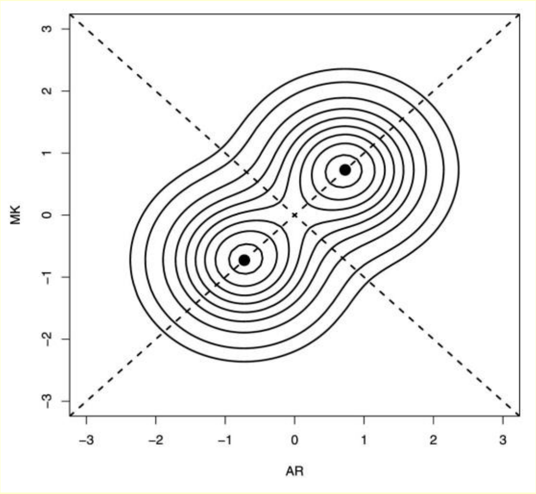


Uncommon visualisation types

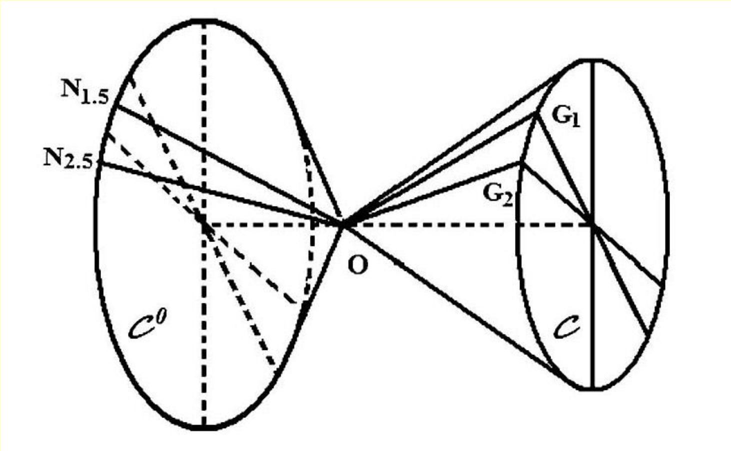
Ternary



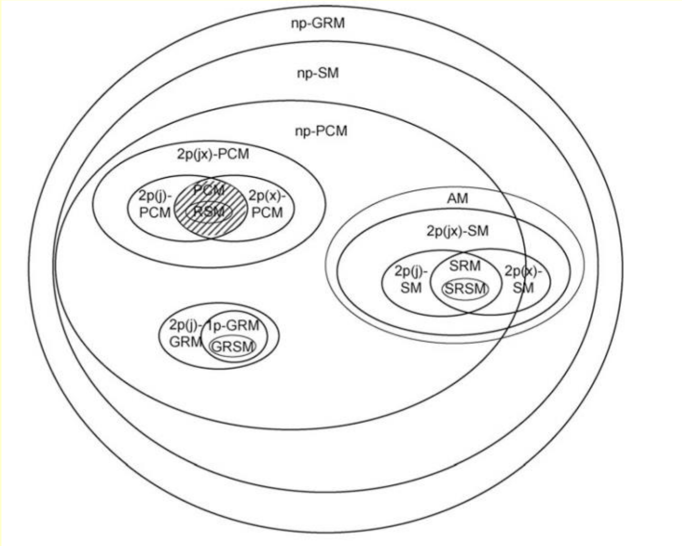
Contour plot



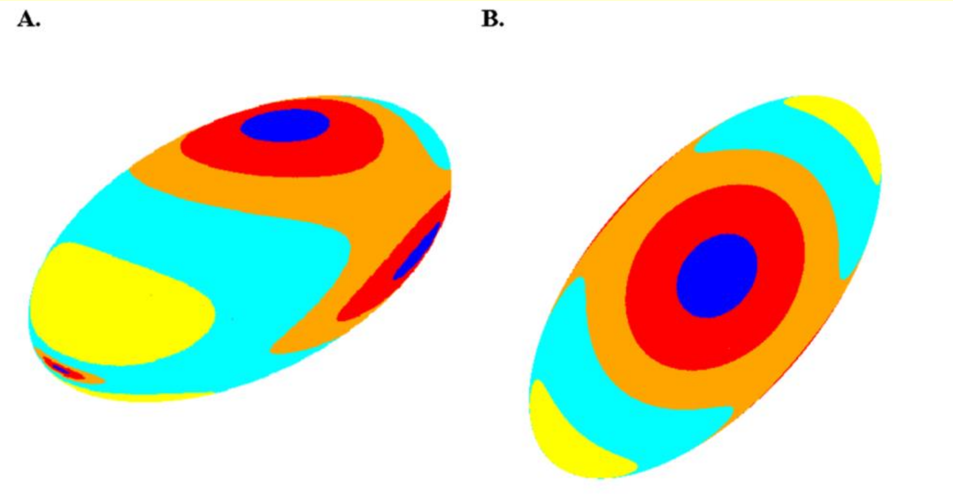
Polyhedral cone



Contour and structural plot had a baby



Correlation weight... heat maps



The good

- Supportive of visual literacy

The good

- Supportive of visual literacy
  - 62% would need high visual literacy (familiarity with the figure) to obtain meaning

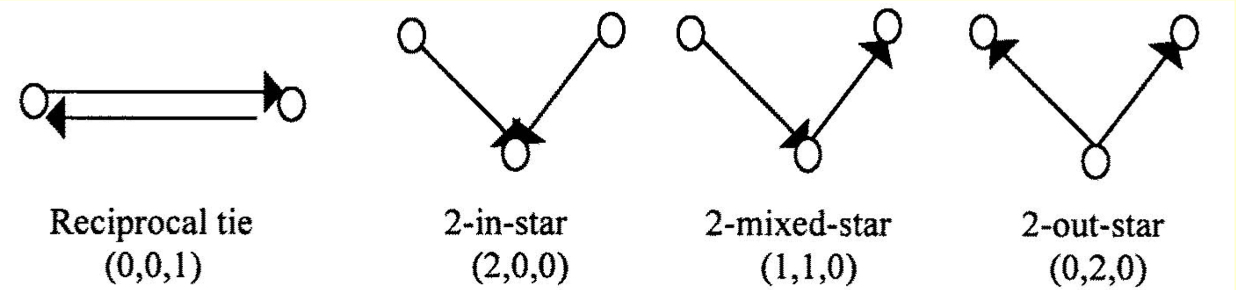


FIGURE 1.
Stars of order 2.

The good

- Supportive of visual literacy
  - 62% would need high visual literacy (familiarity with the figure) to obtain meaning
  - Uncommon figures generally explained well

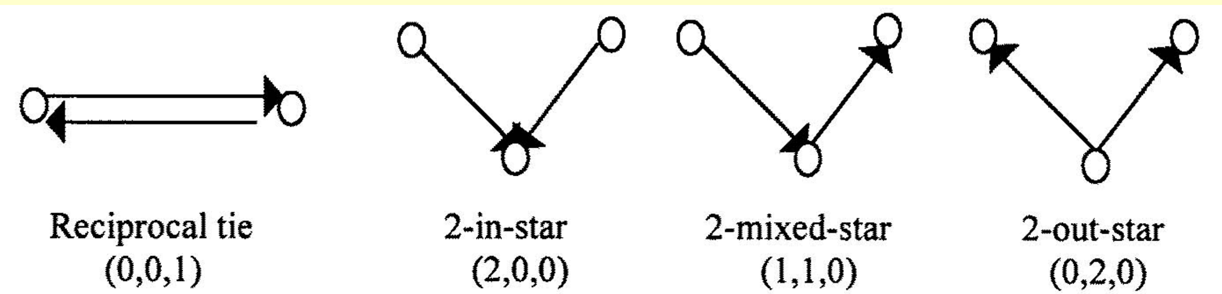
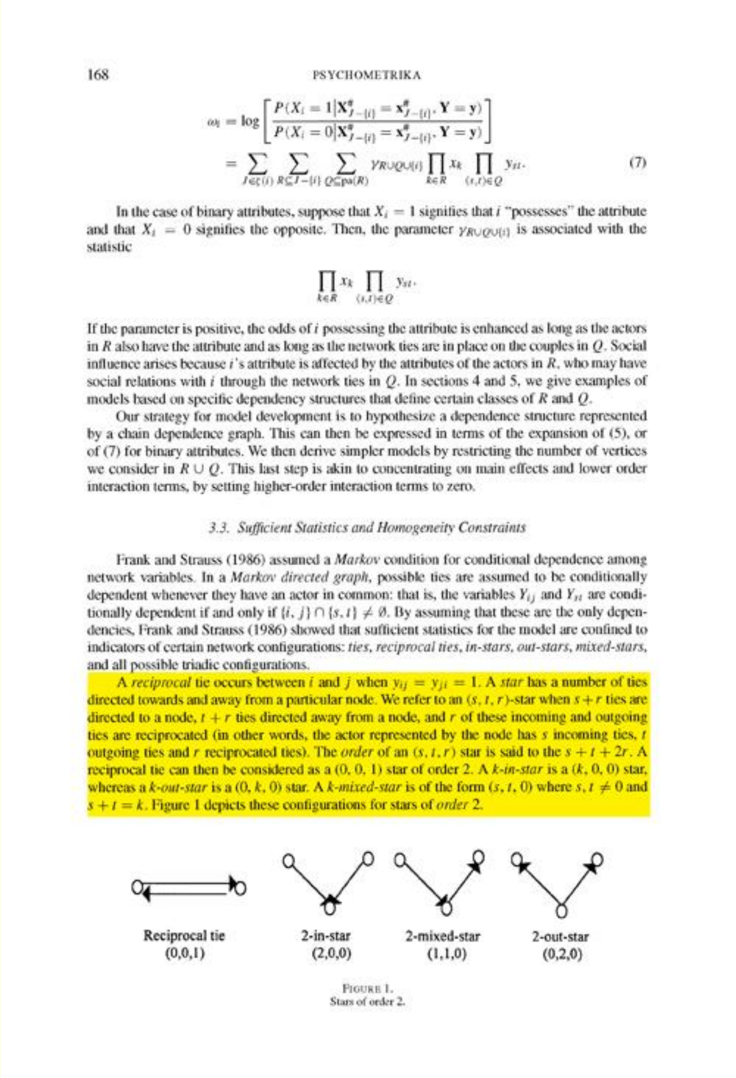


FIGURE 1.
Stars of order 2.



168 PSYCHOMETRIKA

$$\omega_i = \log\left[\frac{P(X_i = 1 \mid \mathbf{X}^{\#}_{J-\{i\}} = \mathbf{x}^{\#}_{J-\{i\}}, \mathbf{Y} = \mathbf{y})}{P(X_i = 0 \mid \mathbf{X}^{\#}_{J-\{i\}} = \mathbf{x}^{\#}_{J-\{i\}}, \mathbf{Y} = \mathbf{y})}\right]$$
$$= \sum_{J \in \xi(i)} \sum_{R \subseteq J-\{i\}} \sum_{Q \subseteq \mathrm{pa}(R)} \gamma_{R \cup Q \cup \{i\}} \prod_{k \in R} x_k \prod_{(s,t) \in Q} y_{st}. \quad (7)$$

In the case of binary attributes, suppose that $X_i = 1$ signifies that i "possesses" the attribute and that $X_i = 0$ signifies the opposite. Then, the parameter $\gamma_{R \cup Q \cup \{i\}}$ is associated with the statistic

$$\prod_{k \in R} x_k \prod_{(s,t) \in Q} y_{st}.$$

If the parameter is positive, the odds of i possessing the attribute is enhanced as long as the actors in R also have the attribute and as long as the network ties are in place on the couples in Q. Social influence arises because i's attribute is affected by the attributes of the actors in R, who may have social relations with i through the network ties in Q. In sections 4 and 5, we give examples of models based on specific dependency structures that define certain classes of R and Q.

Our strategy for model development is to hypothesize a dependence structure represented by a chain dependence graph. This can then be expressed in terms of the expansion of (5), or of (7) for binary attributes. We then derive simpler models by restricting the number of vertices we consider in $R \cup Q$. This last step is akin to concentrating on main effects and lower order interaction terms, by setting higher-order interaction terms to zero.

3.3. Sufficient Statistics and Homogeneity Constraints

Frank and Strauss (1986) assumed a *Markov* condition for conditional dependence among network variables. In a *Markov directed graph*, possible ties are assumed to be conditionally dependent whenever they have an actor in common: that is, the variables Y_{ij} and Y_{st} are conditionally dependent if and only if $\{i, j\} \cap \{s, t\} \neq \emptyset$. By assuming that these are the only dependencies, Frank and Strauss (1986) showed that sufficient statistics for the model are confined to indicators of certain network configurations: *ties*, *reciprocal ties*, *in-stars*, *out-stars*, *mixed-stars*, and all possible triadic configurations.

A *reciprocal* tie occurs between i and j when $y_{ij} = y_{ji} = 1$. A *star* has a number of ties directed towards and away from a particular node. We refer to an (s, t, r)-star when $s + r$ ties are directed to a node, $t + r$ ties directed away from a node, and r of these incoming and outgoing ties are reciprocated (in other words, the actor represented by the node has s incoming ties, t outgoing ties and r reciprocated ties). The *order* of an (s, t, r) star is said to the $s + t + 2r$. A reciprocal tie can then be considered as a $(0, 0, 1)$ star of order 2. A *k-in-star* is a $(k, 0, 0)$ star, whereas a *k-out-star* is a $(0, k, 0)$ star. A *k-mixed-star* is of the form $(s, t, 0)$ where $s, t \neq 0$ and $s + t = k$. Figure 1 depicts these configurations for stars of *order* 2.

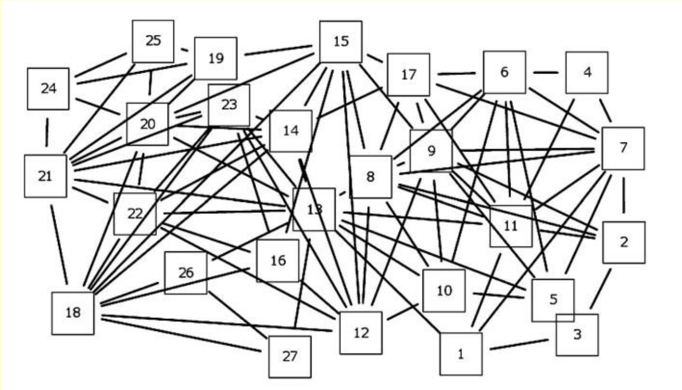
The good

- Generally efficient
  - Efficiency is using the minimum number of visual elements to convey meaning

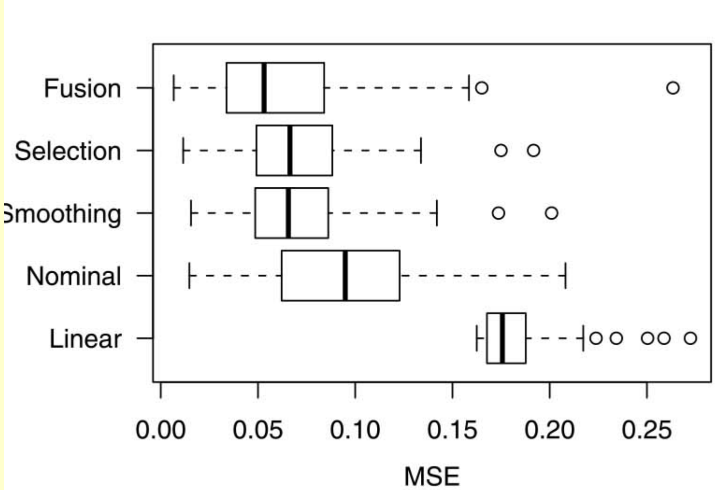
The good

- Generally efficient
  - Efficiency is using the minimum number of visual elements to convey meaning
    - 0 = completely inefficient
    - 100 = completely efficient

Poor efficiency
(only one connection discussed in text)



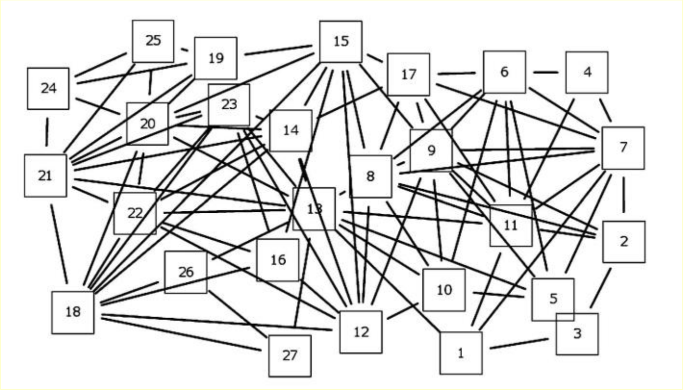
Great efficiency
(meaning clear, all discussed in text)



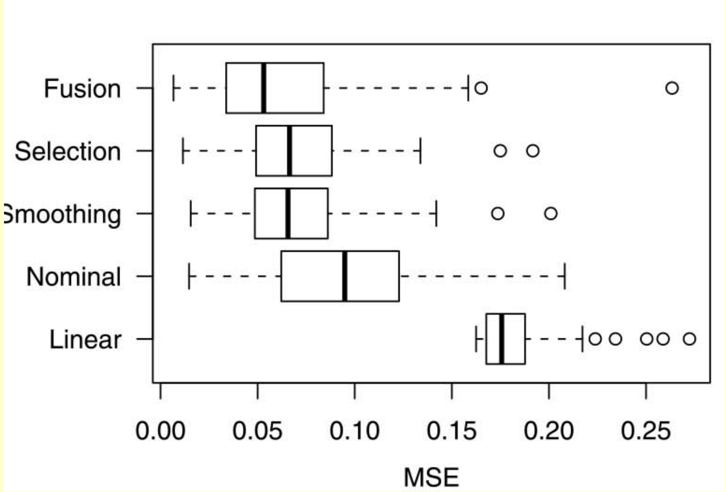
The good

- Generally efficient
  - Efficiency is using the minimum number of visual elements to convey meaning
    - 0 = completely inefficient
    - 100 = completely efficient
  - Average efficiency of 76%

Poor efficiency
(only one connection discussed in text)



Great efficiency
(meaning clear, all discussed in text)



The bad

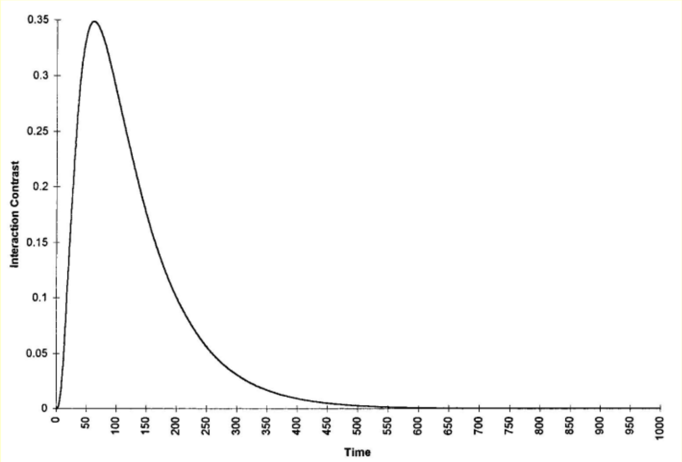
- CLUTTER!
  - Cramming a lot of information into a figure

The bad

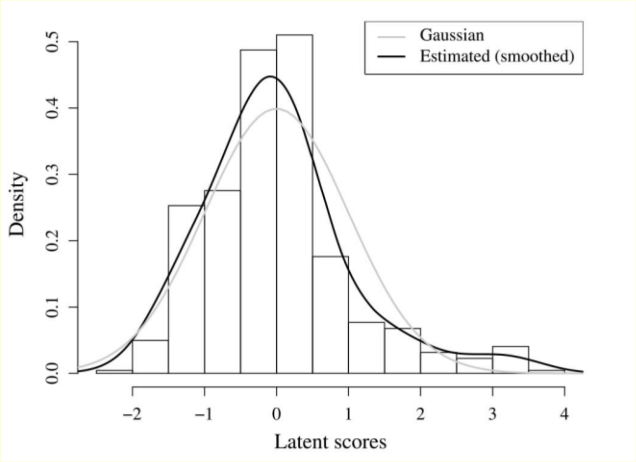
- CLUTTER!
  - Cramming a lot of information into a figure

  (20% had 5 or more points, 10% 10 or more!)

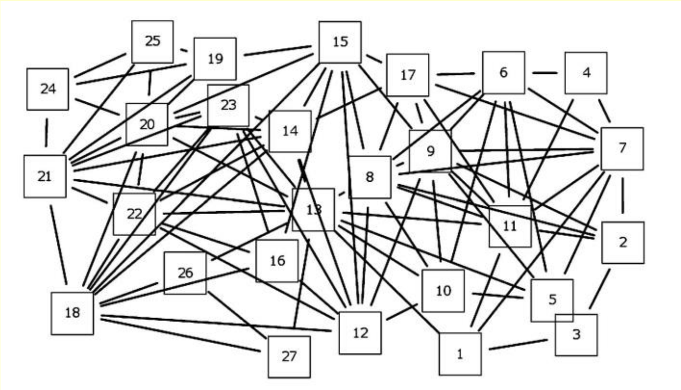
1 point



3 points



Many points

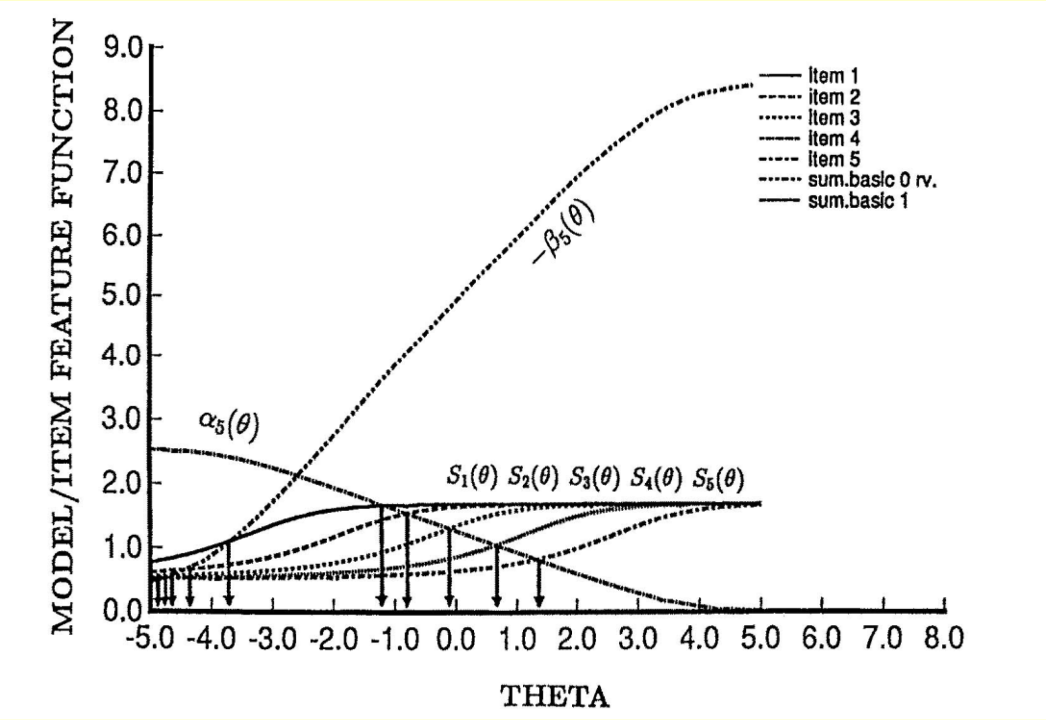


The bad

- CLUTTER!
  - Cramming a lot of information into a figure

(20% had 5 or more points, 10% 10 or more!)

Heavy annotation

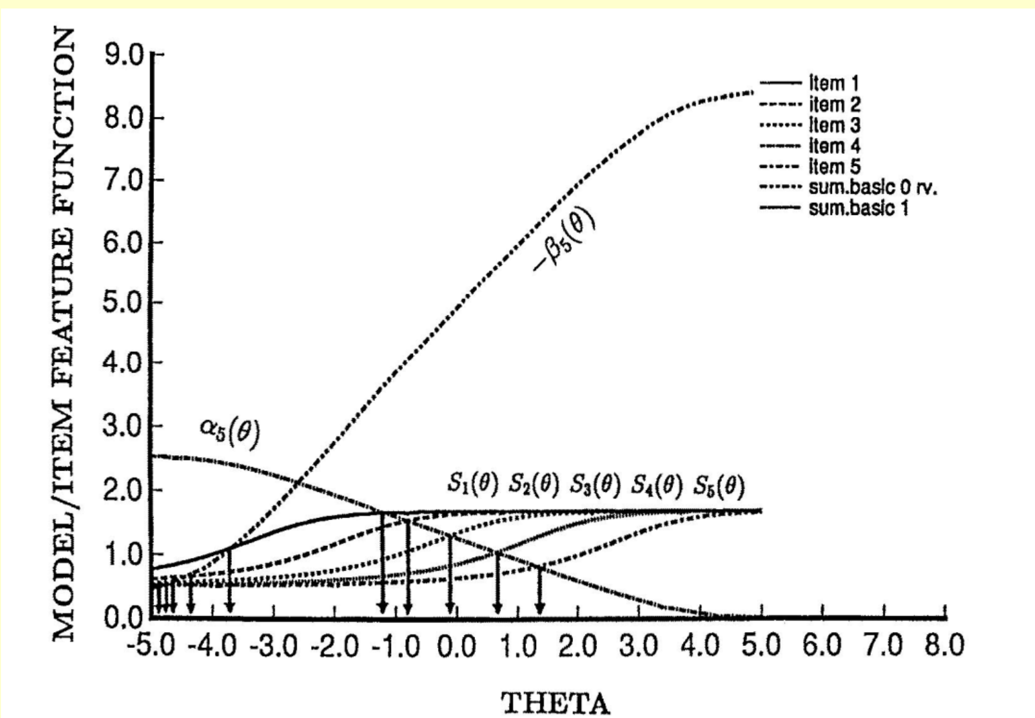


The bad

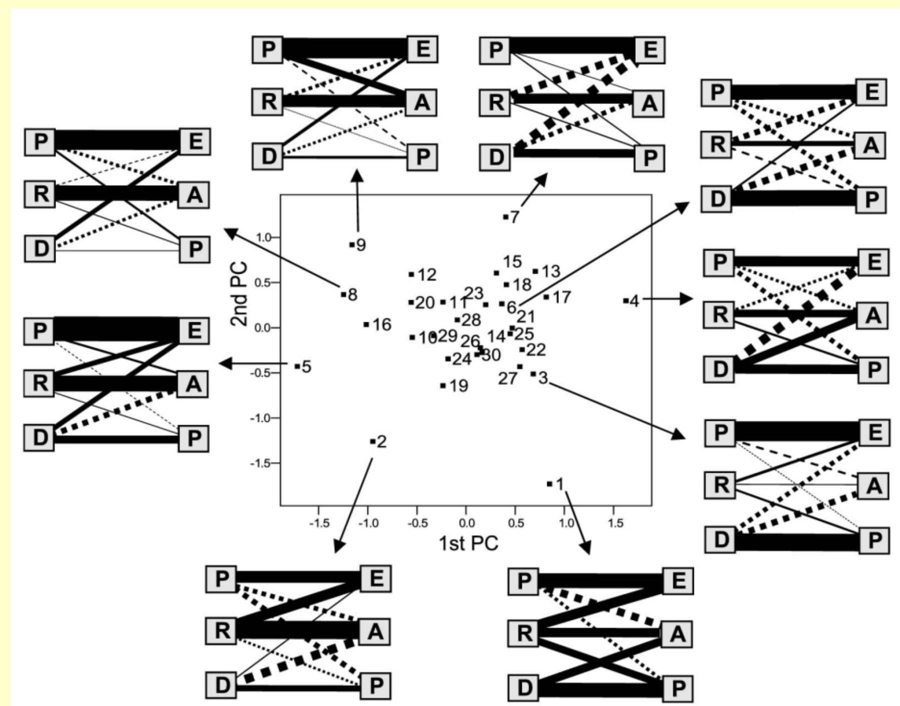
- CLUTTER!
  - Cramming a lot of information into a figure

(20% had 5 or more points, 10% 10 or more!)

Heavy annotation



Hybrid oddity



The bad

- CLUTTER!
  - Cramming too many complex panels into a page

The bad

- CLUTTER!
  - Cramming too many complex panels into a page

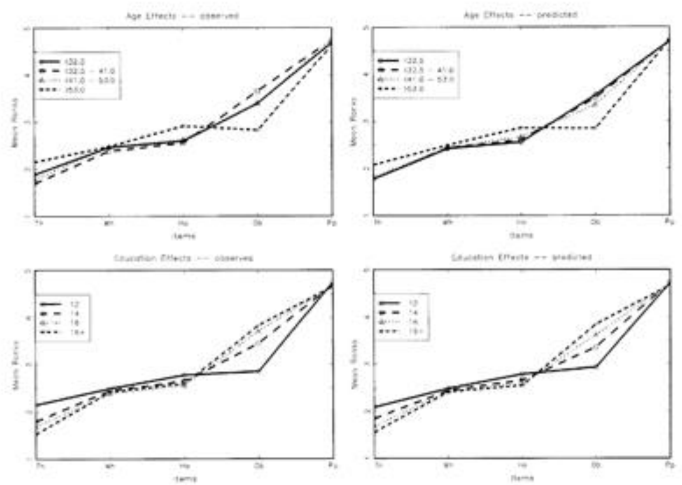


FIGURE 2.
Observed and predicted mean ranks of parental value data for the four quartiles of *Age* and *Education*.

The bad

- CLUTTER!
  - Cramming too many complex panels into a page

Fetch the magnifying glass!

YVONNICK NOEL 667

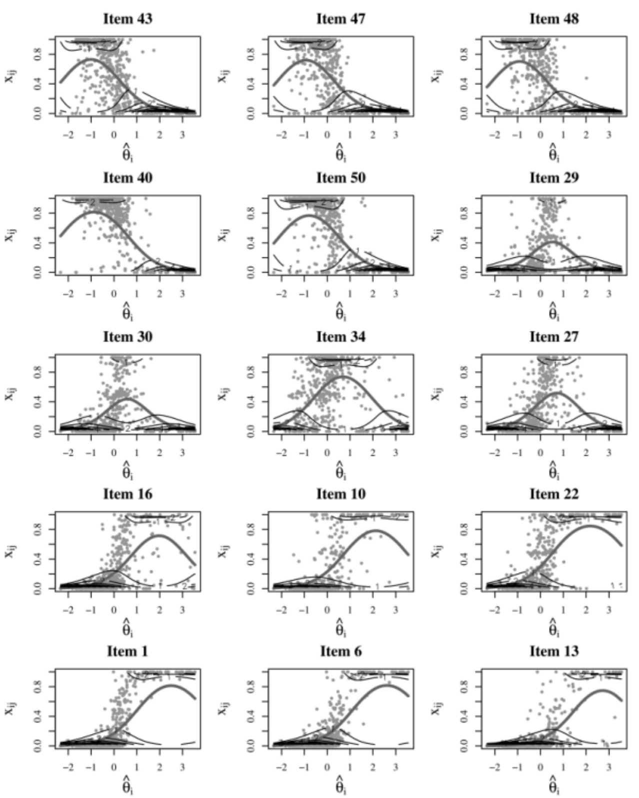


FIGURE 10.
Observed data, response density contours, and expectation function for the abortion data. The *thick plain lines* display the expected rating curve, the *thin lines* the response density contours as functions of attitude. Figures in the contour lines are the density values at that particular level.

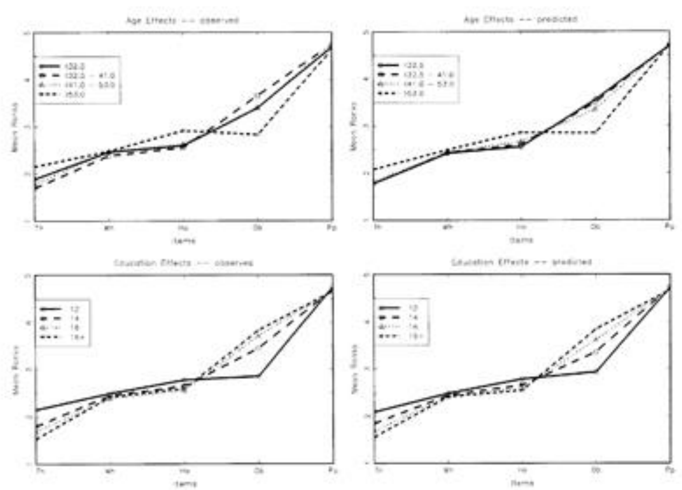
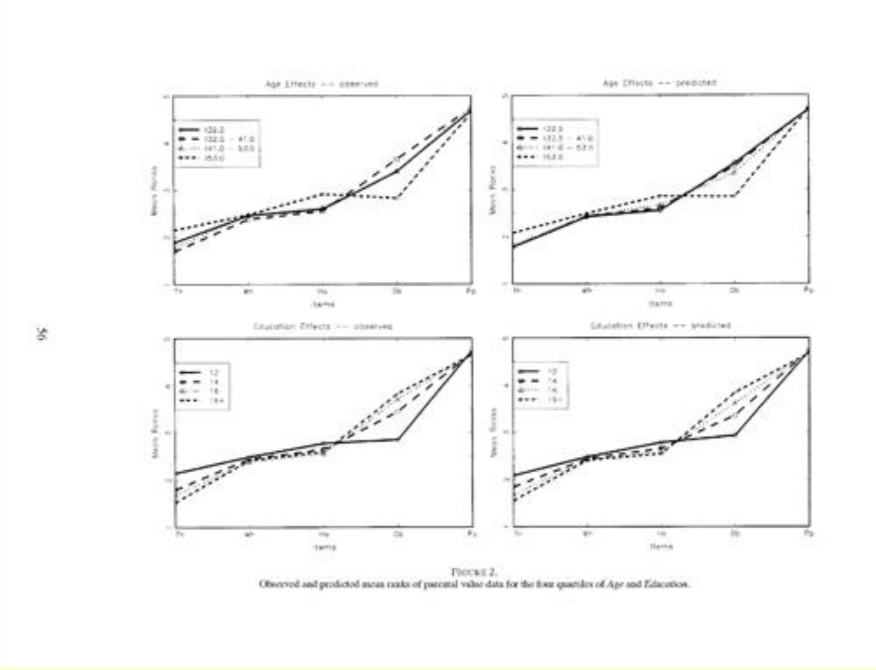


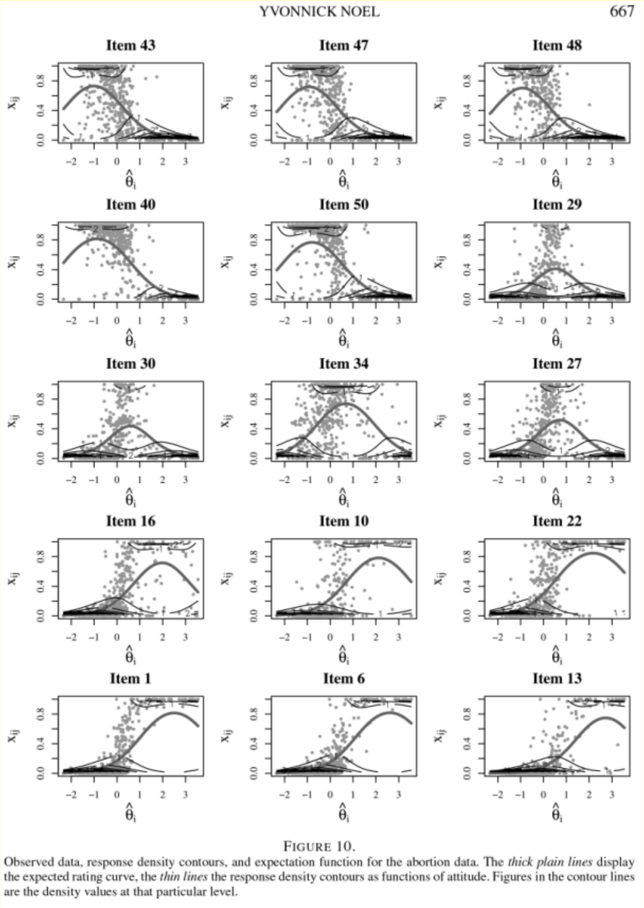
FIGURE 2.
Observed and predicted mean ranks of parental value data for the four quartiles of *Age* and *Education*.

The bad

- CLUTTER!
  - Cramming too many complex panels into a page
    - Mostly 2,4, or 9 panels per figure
    - Ranges from 1 to 15 panels per page

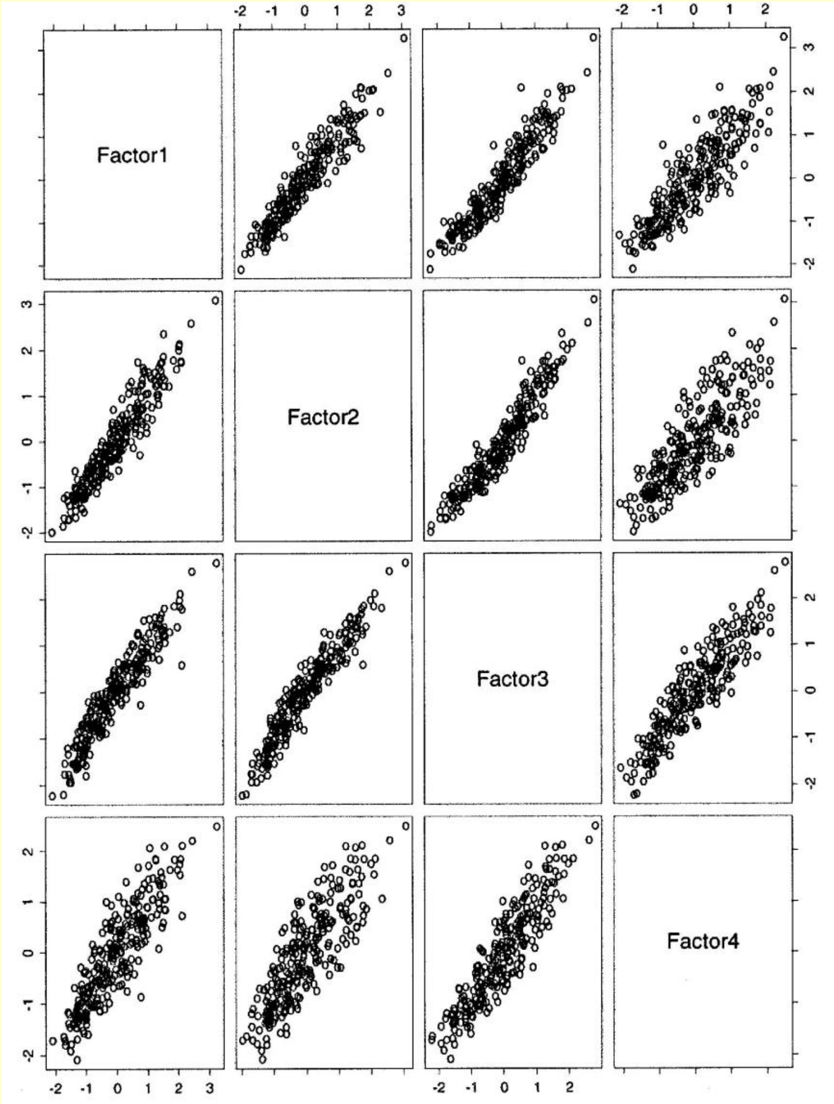


Fetch the magnifying glass!



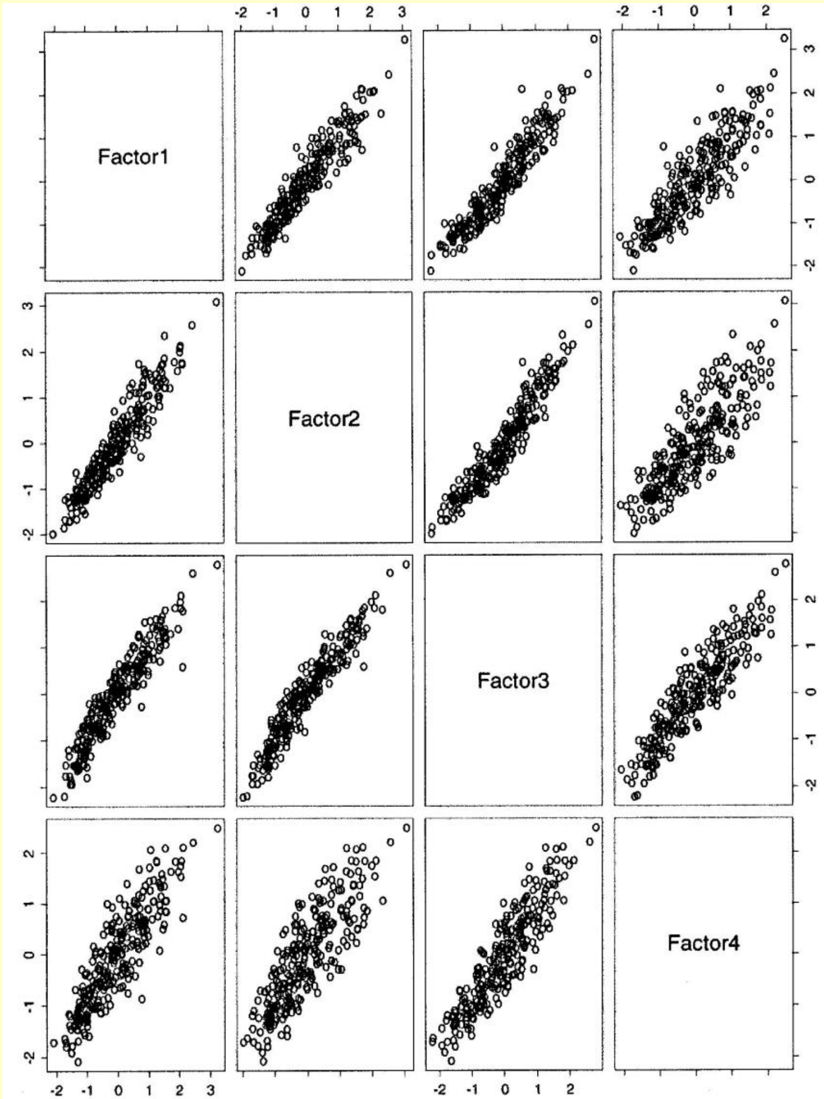
Totally just Figure 1

Theoretically valid

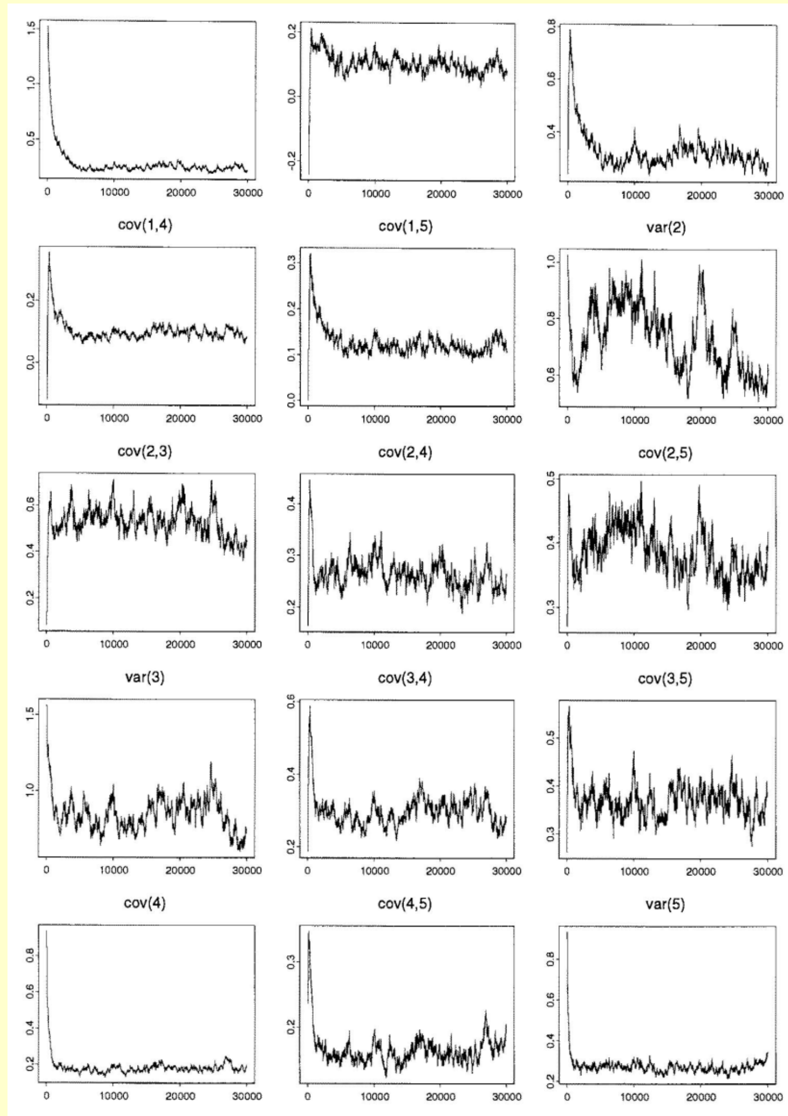


Totally just Figure 1

Theoretically valid



That's just cheating!



The bad

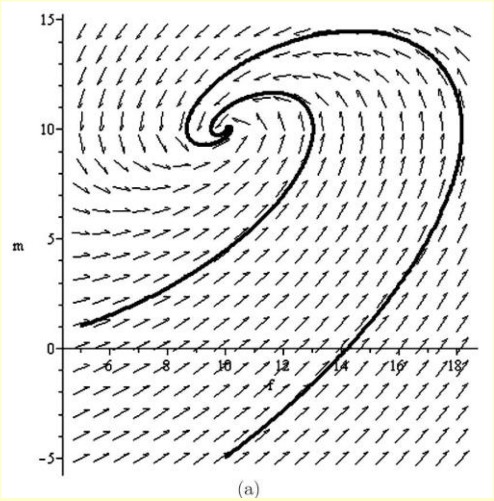
- Cut the clutter

The bad

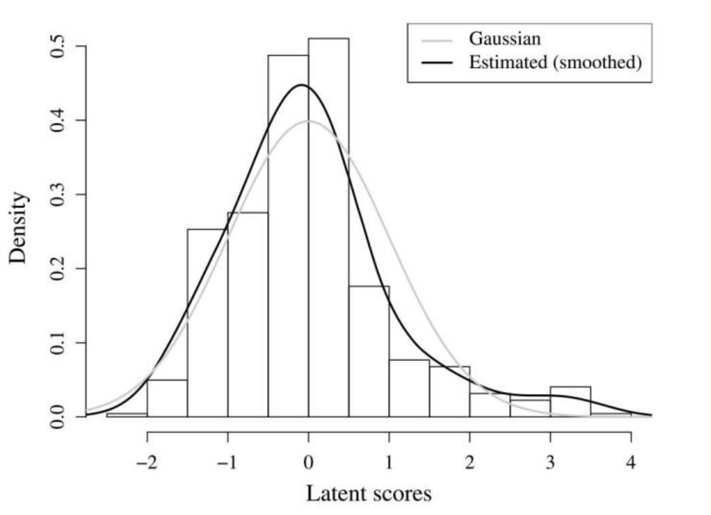
- Cut the clutter
  - Crammed panels
    - If you don't go through it in detail in text, you don't need it

The bad

- Cut the clutter
  - Crammed panels
    - If you don't go through it in detail in text, you don't need it
    - Use shading and colour to differentiate important visual elements

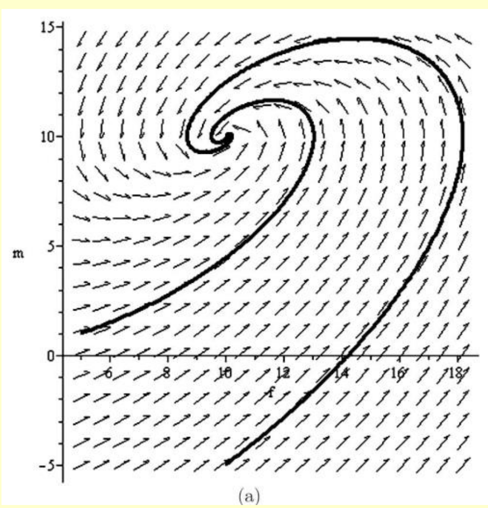


+

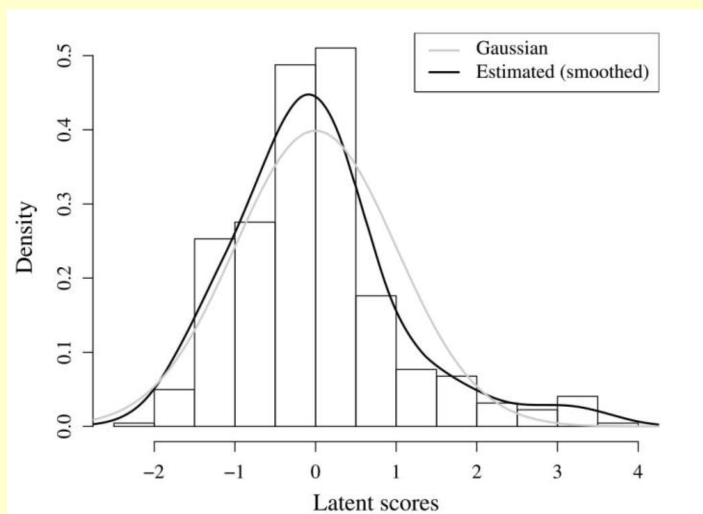


The bad

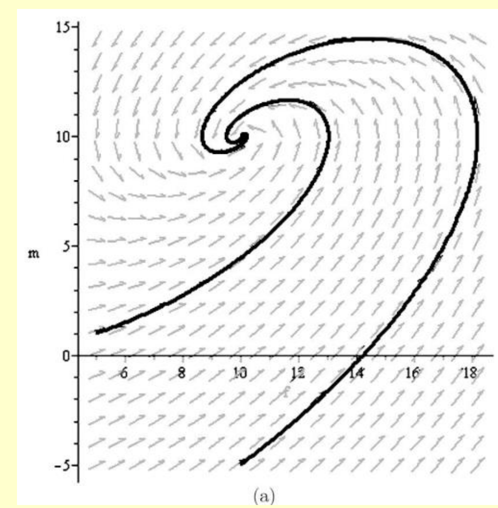
- Cut the clutter
  - Crammed panels
    - If you don't go through it in detail in text, you don't need it
    - Use shading and colour to differentiate important visual elements



+

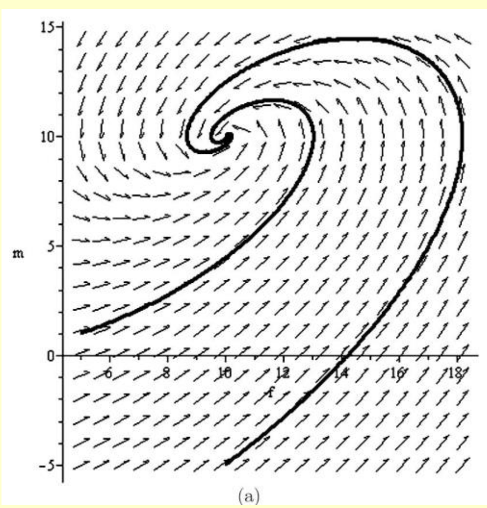


=

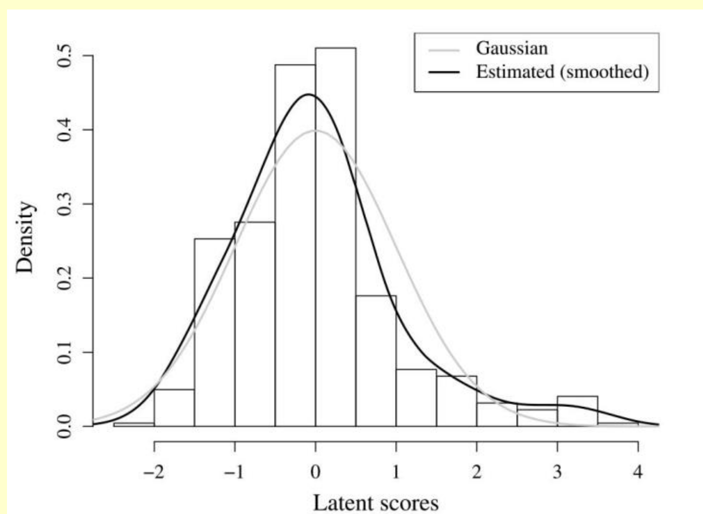


The bad

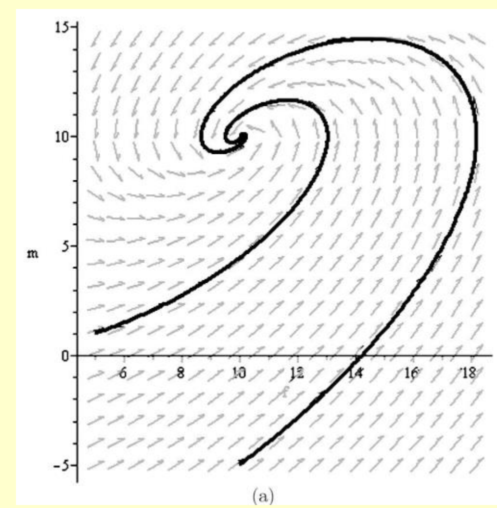
- Cut the clutter
  - Crammed panels
    - If you don't go through it in detail in text, you don't need it
    - Use shading and colour to differentiate important visual elements



+

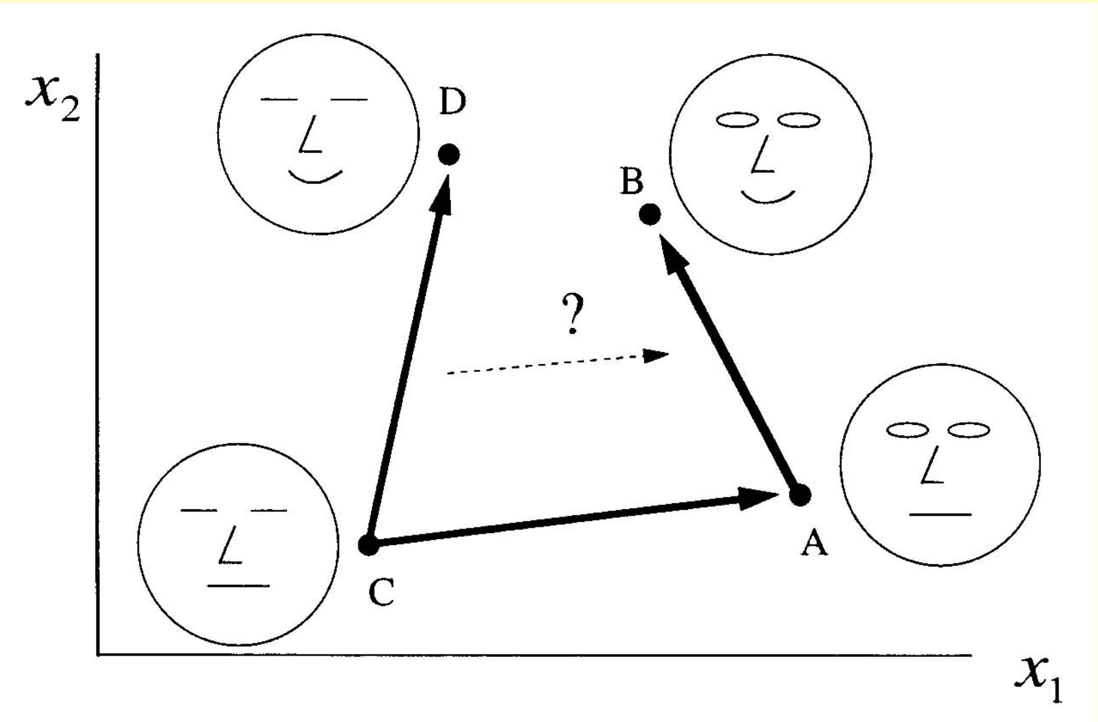


=

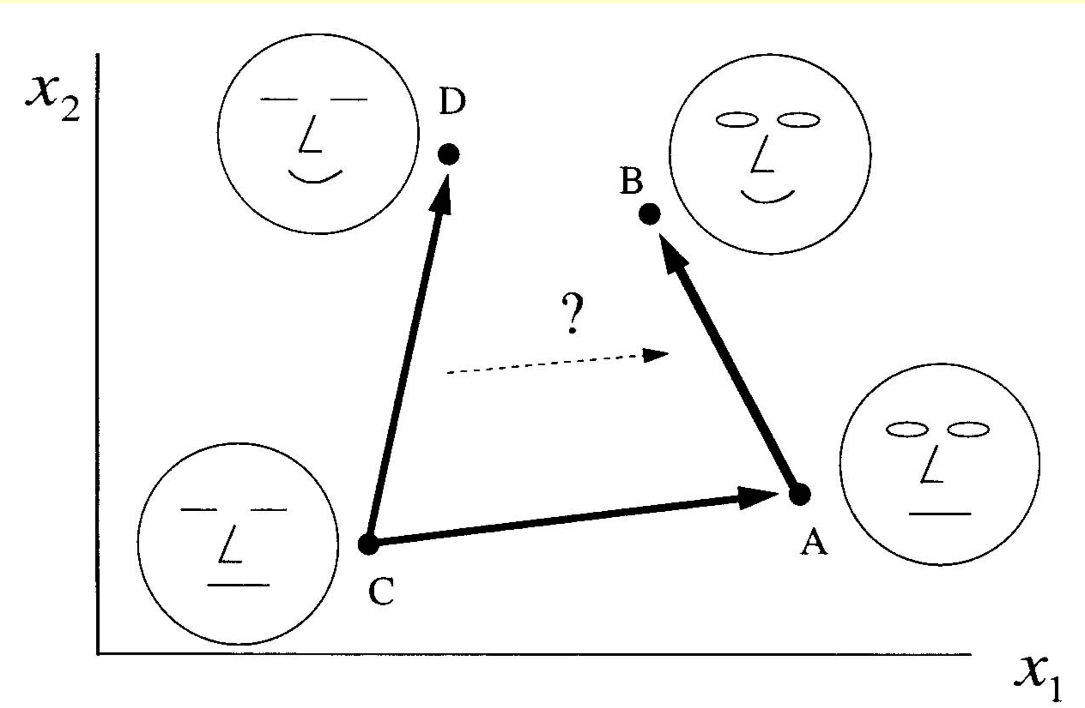


  - Crammed pages
    - Do you *really* need all those panels?

The ugly



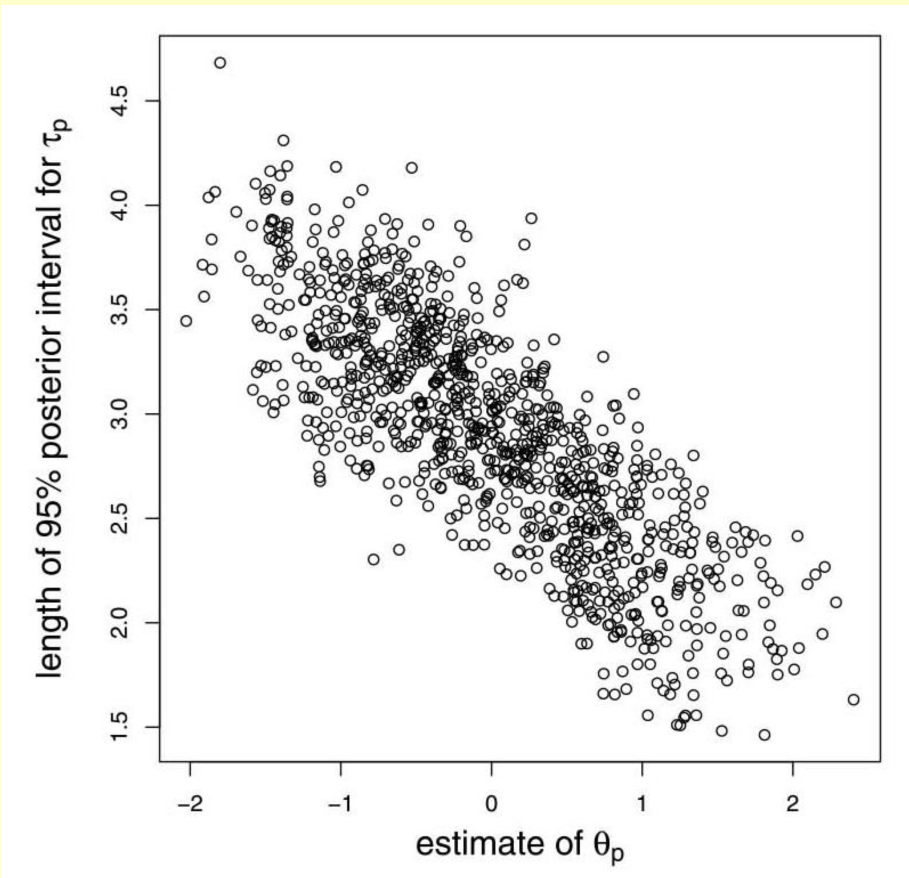
The ugly

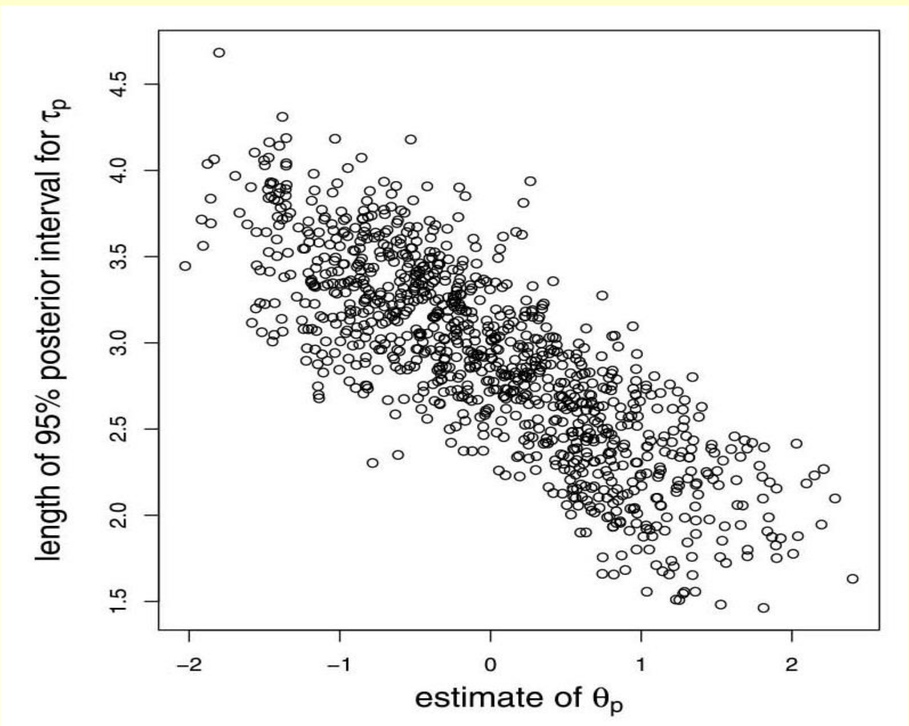


The ugly

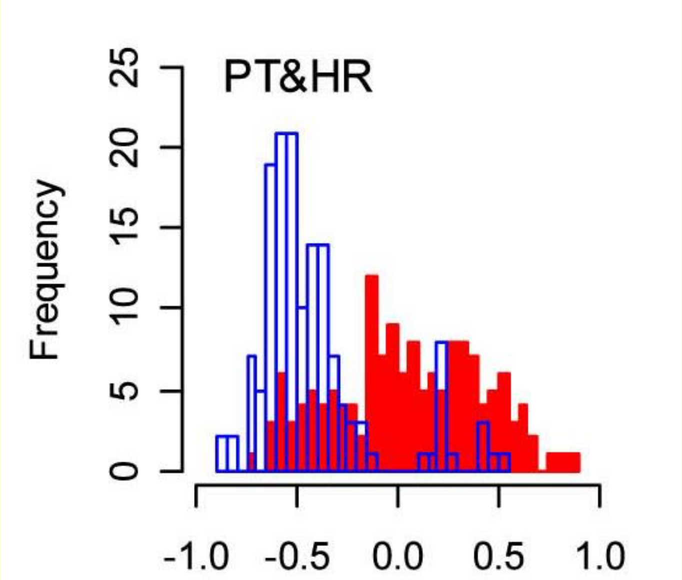
Aspect ratio errors

- Makes words difficult to read
- Can distort interpretation



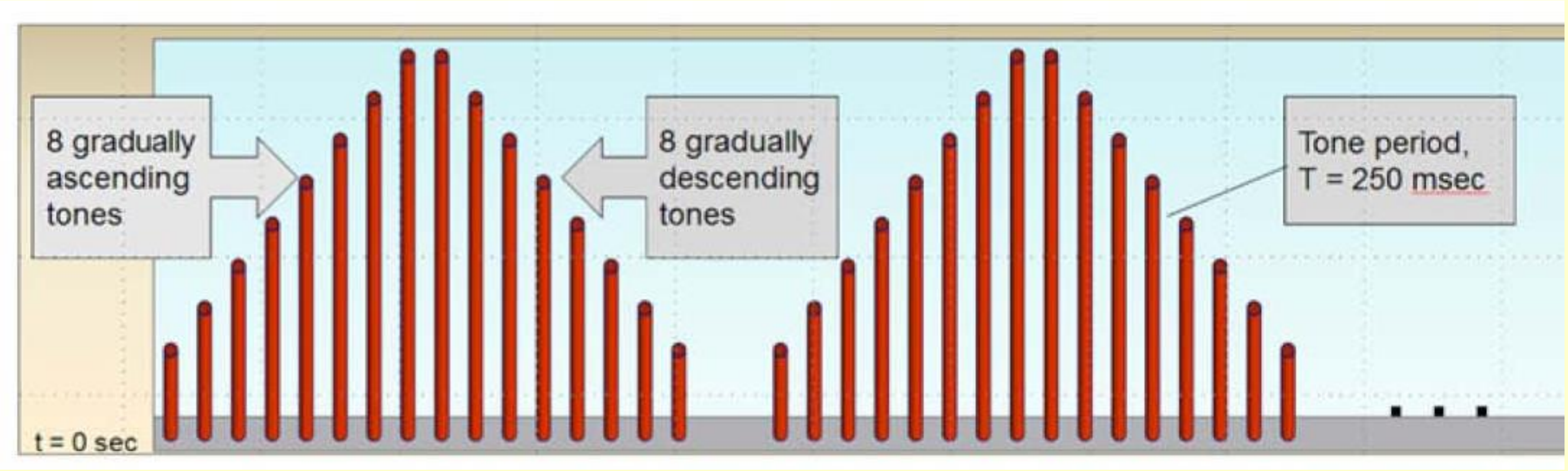


The ugly



Colours without meaning

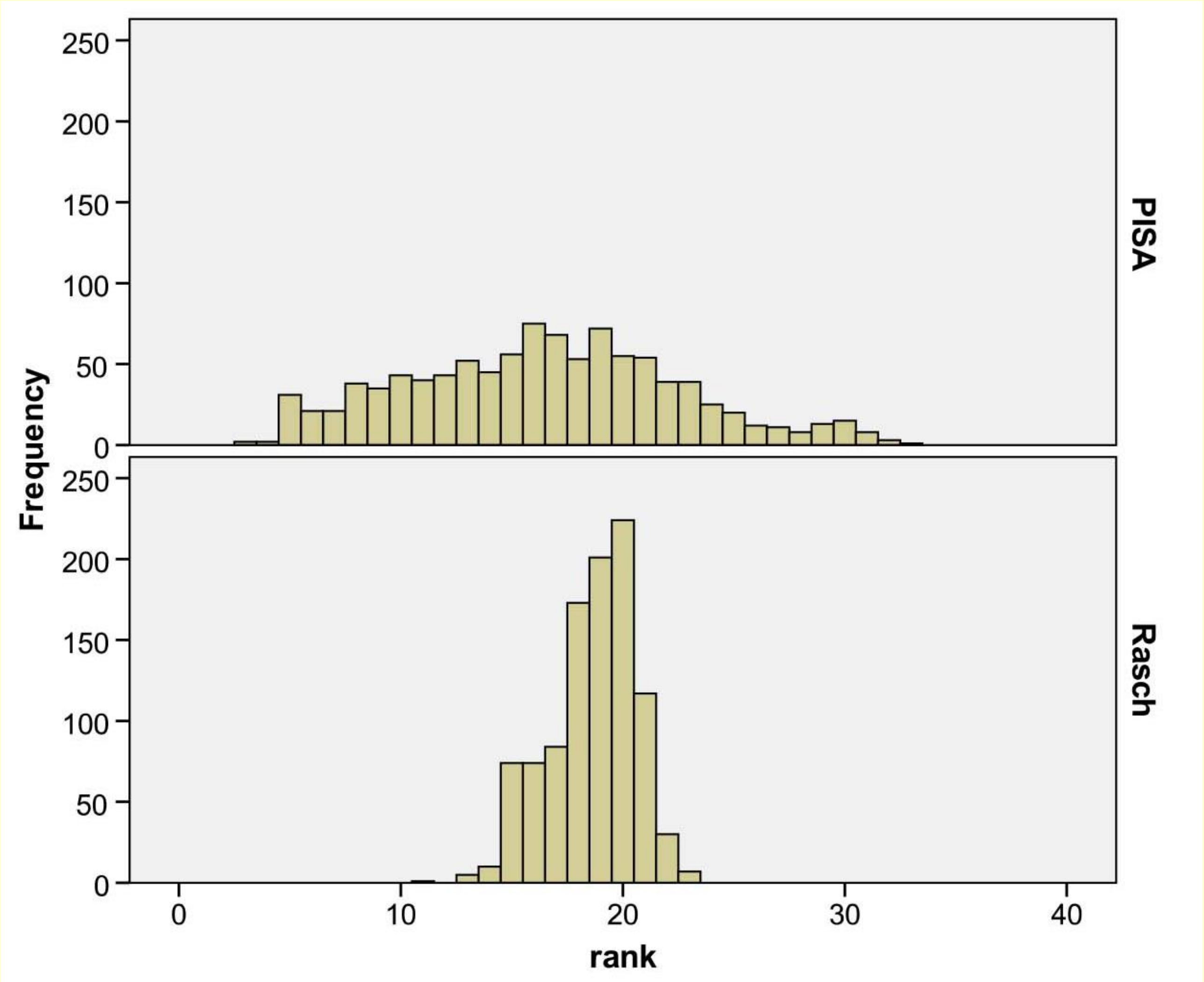
- Distracting
- Make those who can't afford colour budget jealous



The ugly

Straight from SPSS

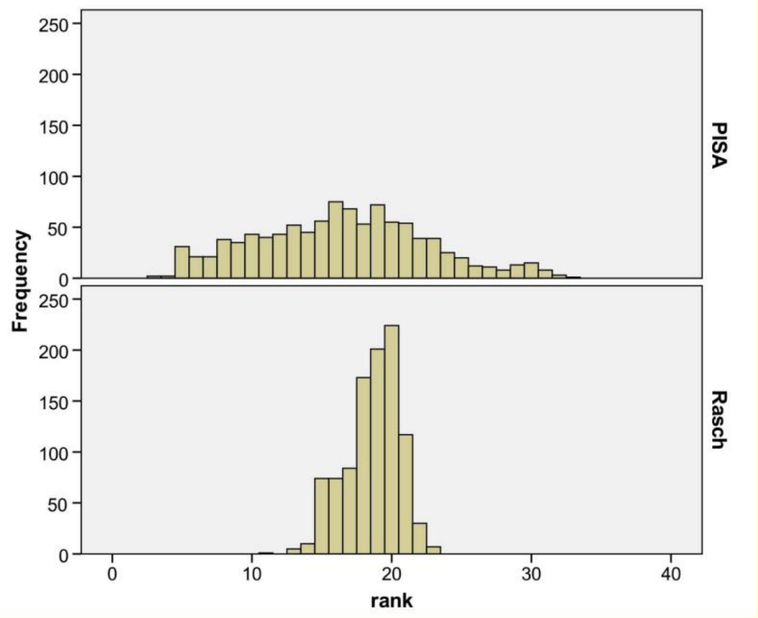
- No.



The ugly

Straight from your analysis

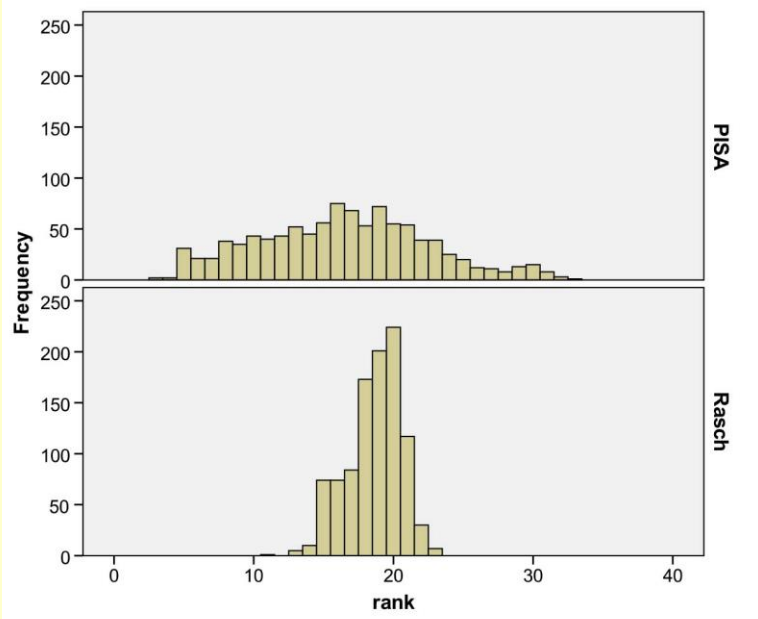
- No.



The ugly

CCAI**R**S = R for rhetorical intention

Straight from your analysis
- No. No.

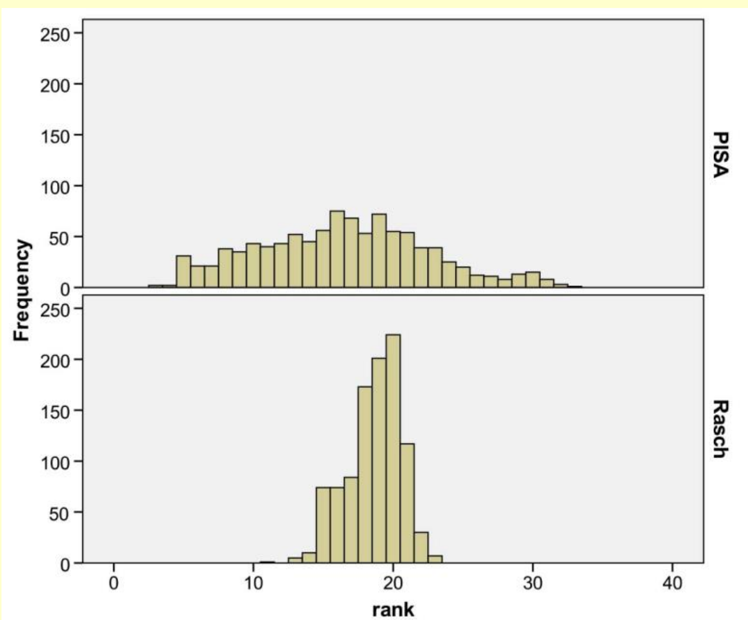


- Engage an audience?
- Facilitate an ongoing process?
- Encourage rote learning?
- Encourage elaborative learning?
- Make a passing point
  (no need for recall)

The ugly

CCAI**R**S = R for rhetorical intention

Straight from your analysis
- No. No.



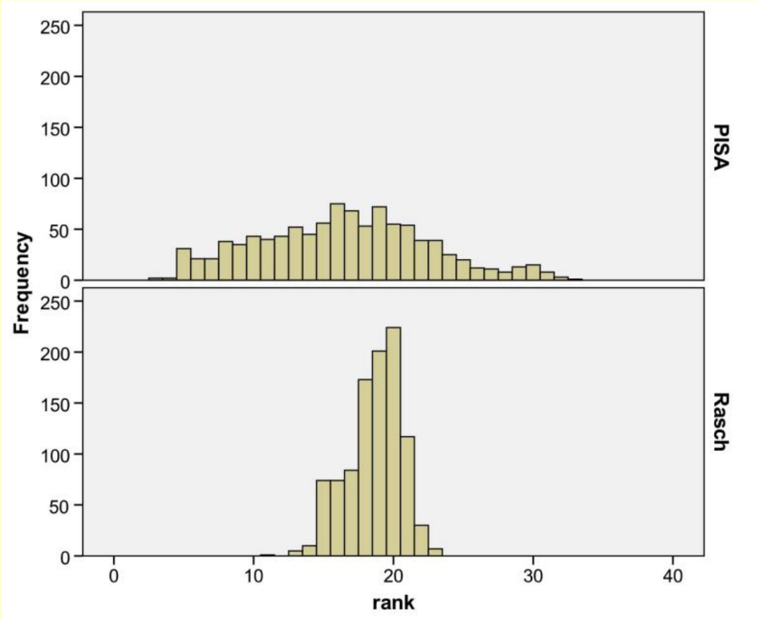
- Engage an audience?
- Facilitate an ongoing process?
- Encourage rote learning?
- Encourage elaborative learning?
- Make a passing point
  (no need for recall)

The ugly

CCAI**R**S = R for rhetorical intention

Straight from your analysis

- No. No.



- Engage an audience?
- Facilitate an ongoing process?
- Encourage rote learning?
- Encourage elaborative learning?
- Make a passing point
  (no need for recall)

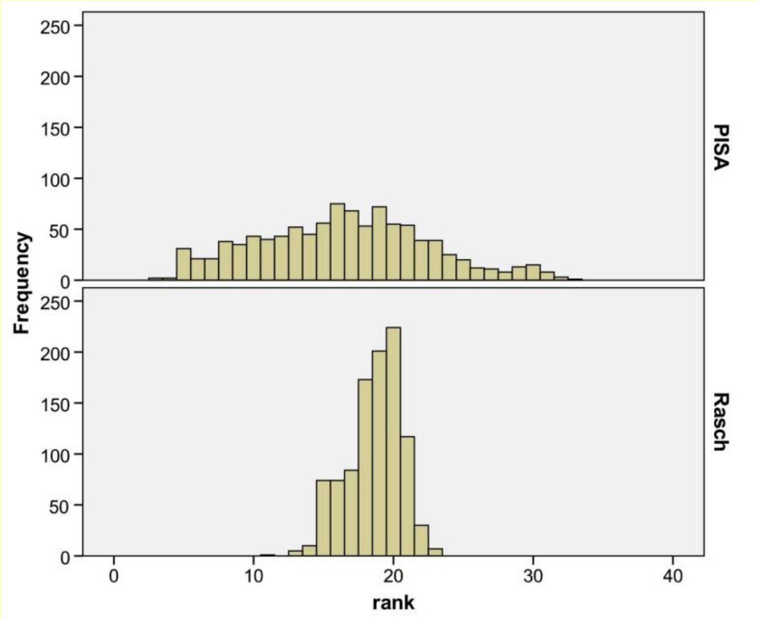
The ugly

Aspect ratio errors

- Makes words difficult to read
- Can distort interpretation

Straight from your analysis

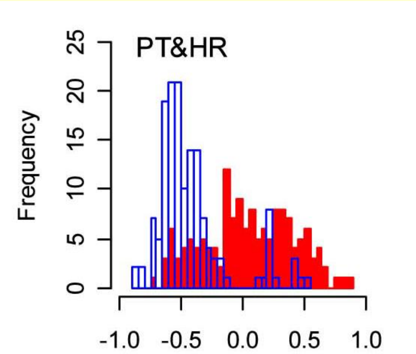
- No. No. No.



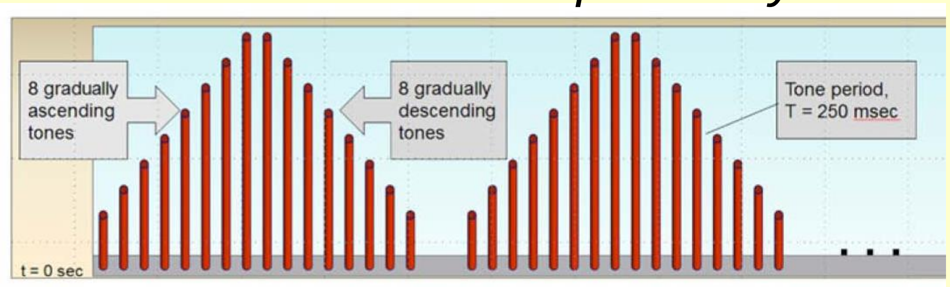
Colours without meaning

- Distracting
- Make those who can't afford colour budget jealous

These colours clarify interpretation



These colours help nobody



Conclusions

Keep doing these things:

- Explain how to interpret unusual figures
- Convey meaning efficiently

Conclusions

Keep doing these things:
- Explain how to interpret unusual figures
- Convey meaning efficiently

Don't do these things:
- Cram your figures with panels
- Cram your panels with visual clutter
- Use colour where it doesn't convey meaning
- Use visualisations straight from your analysis

Conclusions

Keep doing these things:
- Explain how to interpret unusual figures
- Convey meaning efficiently

Don't do these things:
- Cram your figures with panels
- Cram your panels with visual clutter
- Use colour where it doesn't convey meaning
- Use visualisations straight from your analysis

Consider doing these things:
- Use fewer elements and panels, and discuss them in more depth
- Use greys and colours to avoid clutter
- Treat your visualisations like you do your text: craft and draft.

Conclusions

Keep doing these things:

- Explain how to interpret unusual figures
- Convey meaning efficiently

Don't do these things:

- Cram your figures with panels
- Cram your panels with visual clutter
- Use colour where it doesn't convey meaning
- Use visualisations straight from your analysis

Consider doing these things:

- Use fewer elements and panels, and discuss them in more depth
- Use greys and colours to avoid clutter
- Treat your visualisations like you do your text: craft and draft.

Clark, R., & Lyons, C. (2010). Graphics for learning: Proven guidelines for planning, designing, and evaluating visuals in training materials (pp. 23–66).

Dimopoulos, K., Koulaidis, V., & Sklaveniti, S. (2003). Towards an analysis of visual images in school science textbooks and press articles about science and technology. Research in Science Education, 189–216.

Gahegan, M. (1999). Four barriers to the development of effective exploratory visualisation tools for the geosciences. International Journal of Geographical Information Science, 13(4), 289–309.

Glenberg, A., & Langston, W. (1992). Comprehension of illustrated text: Pictures help to build mental models. Journal of memory and language, (2), 129–151.

Goldsmith, E. (1987). The analysis of illustration in theory and practice. The psychology of illustration (pp. 53 – 85). Springer US.

Hegarty, M, & Just, M. (1993). Constructing mental models of machines from text and diagrams. Journal of memory and language, 32, 717 – 742.

Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. Environmental Modelling & Software, 26(6), 822–827.

Levie, W. H. (1987). Research on Pictures : A Guide to the Literature. The psychology of illustration (pp. 1 – 50). Springer US.

Mathai, S., & Ramadas, J. (2009). Visuals and visualisation of human body systems. International Journal of Science Education, (February 2009), 1–20.

Subramaniam, K., & Padalkar, S. (2009). Visualisation and Reasoning in Explaining the Phases of the Moon. International Journal of Science, 1–20.

Tufte, E. R., & Graves-Morris, P. R. (1983). The visual display of quantitative information (Vol. 2). Cheshire, CT: Graphics press.

Tversky, B. (2011). Visualizing thought. Topics in cognitive science, 3(3), 499–535.

Wood, P. (1994) Scientific Illustration (2nd Edition). New York: Van Nostrand Reinhold.

Skeleton Wikimedia commons

Chimp http://blogs.houstonzoo.org/2013/06/protecting-the-zoos-chimpanzee-counterparts-in-the-wild/

Ecosystem http://eschooltoday.com/ecosystems/images/ecosystem-illustration.jpg

Alaska gulf http://celebrating200years.noaa.gov/breakthroughs/ecosystems/gulf_of_alaska_lme_650.jpg