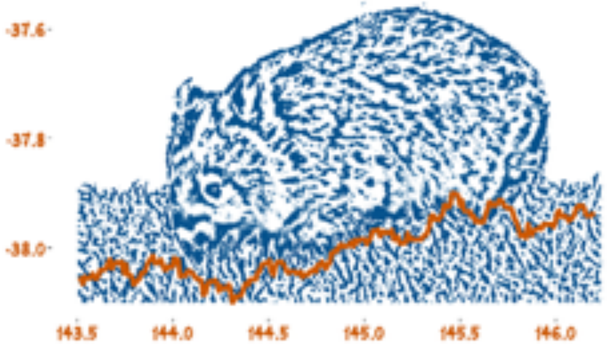
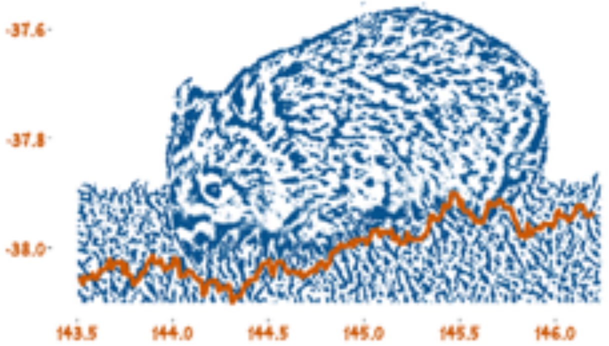
Making data analysis easier



Feature Hierarchy in Graphical Displays

Heike Hofmann*, Susan VanderPlas
Iowa State University

*currently visiting Monash

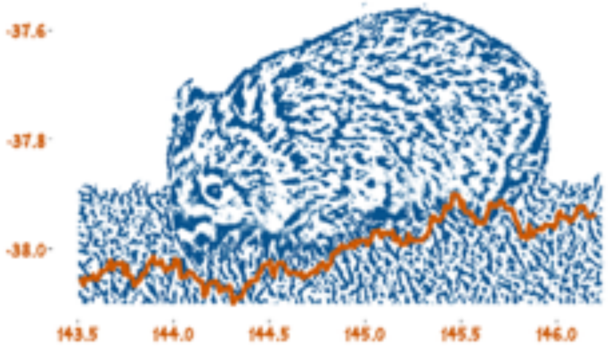


Feature Hierarchy in Graphical Displays

Heike Hofmann*, Susan VanderPlas
Iowa State University

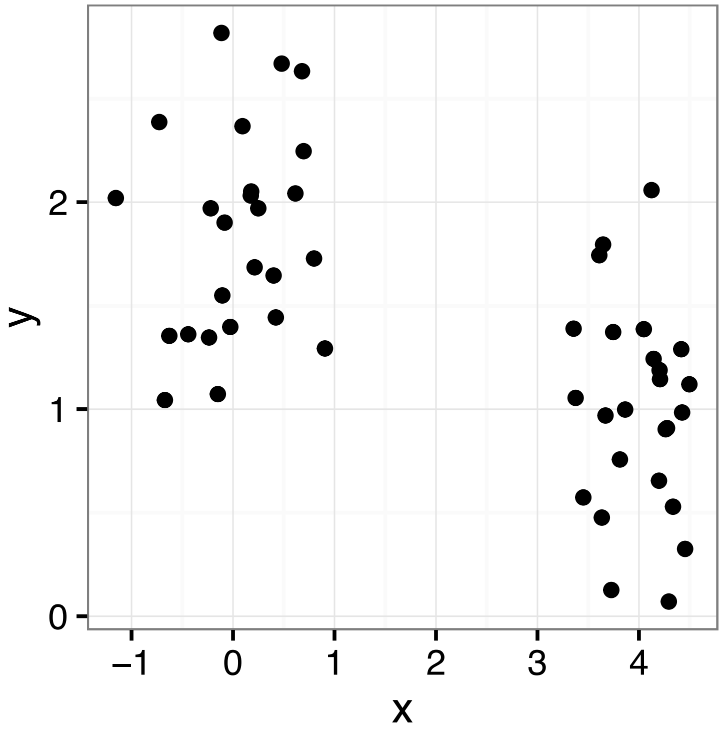
*currently visiting Monash

Outline

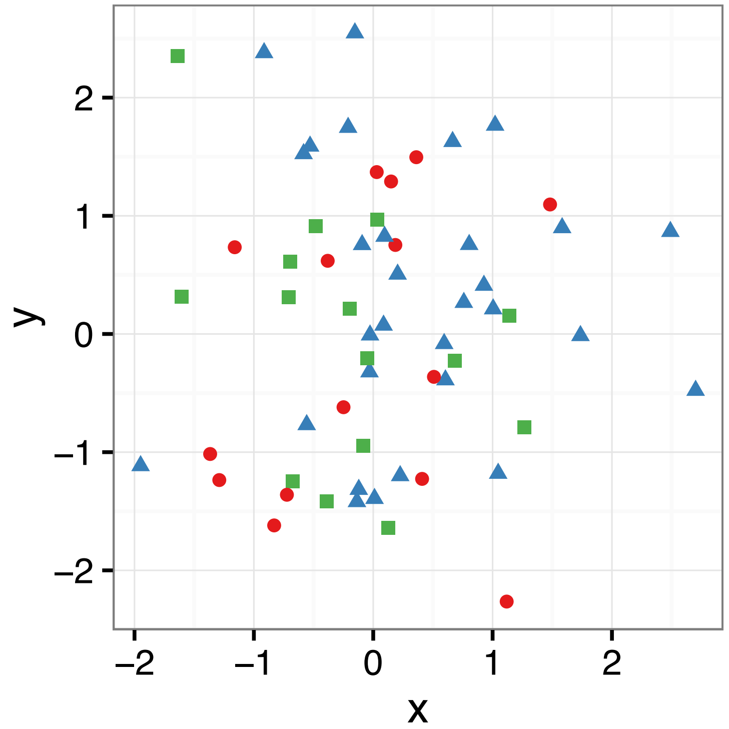


- Cognition and Statistical Graphics
- Lineup Protocol
- Study Design
- Results

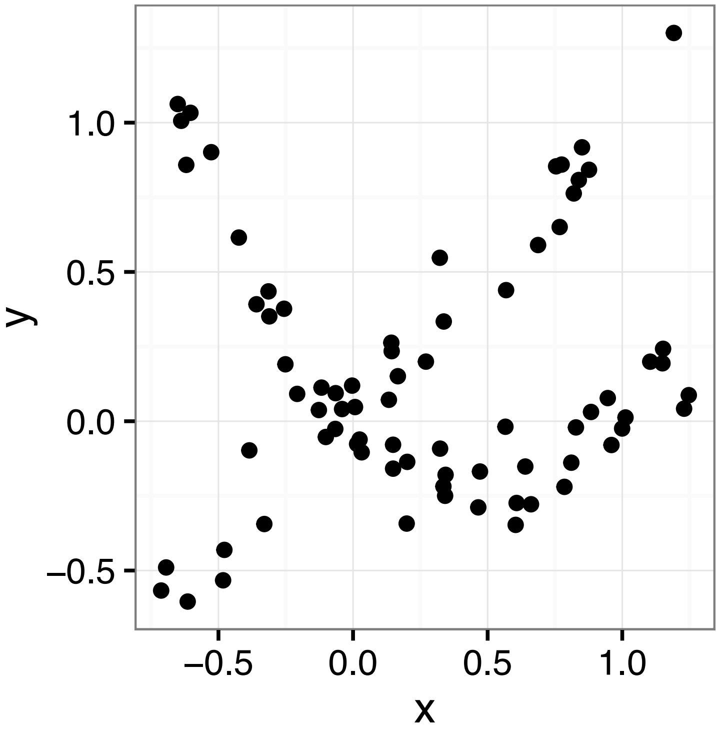
Finding patterns in data



Proximity



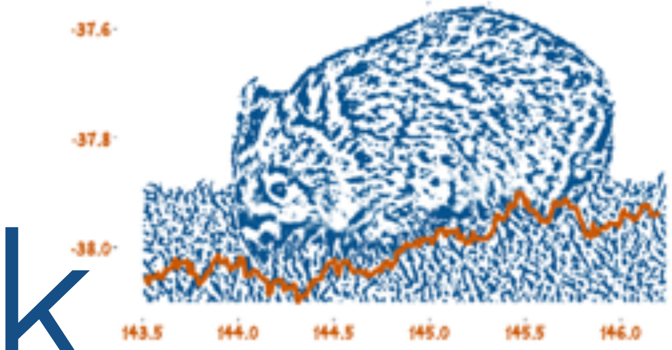
Similarity



Continuity

Cognitive principles for grouping

Missing link

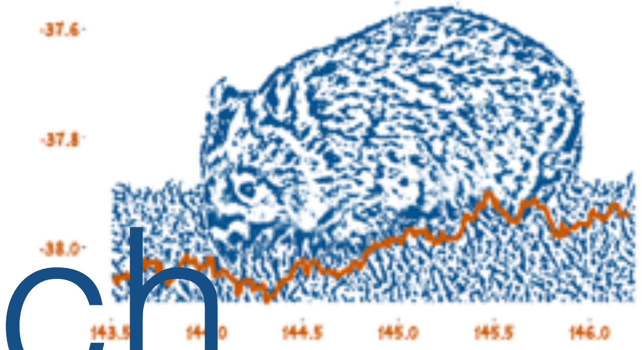


- Cleveland & McGill (1984): hierarchy of basic visual tasks: comparisons along common axis, lengths, area, …

- Hierarchy of pre-attentive features (Healey & Enns, 1999): color, shape, angle, …

- Pre-attentiveness of features does not directly translate to understanding charts … need more direct validation

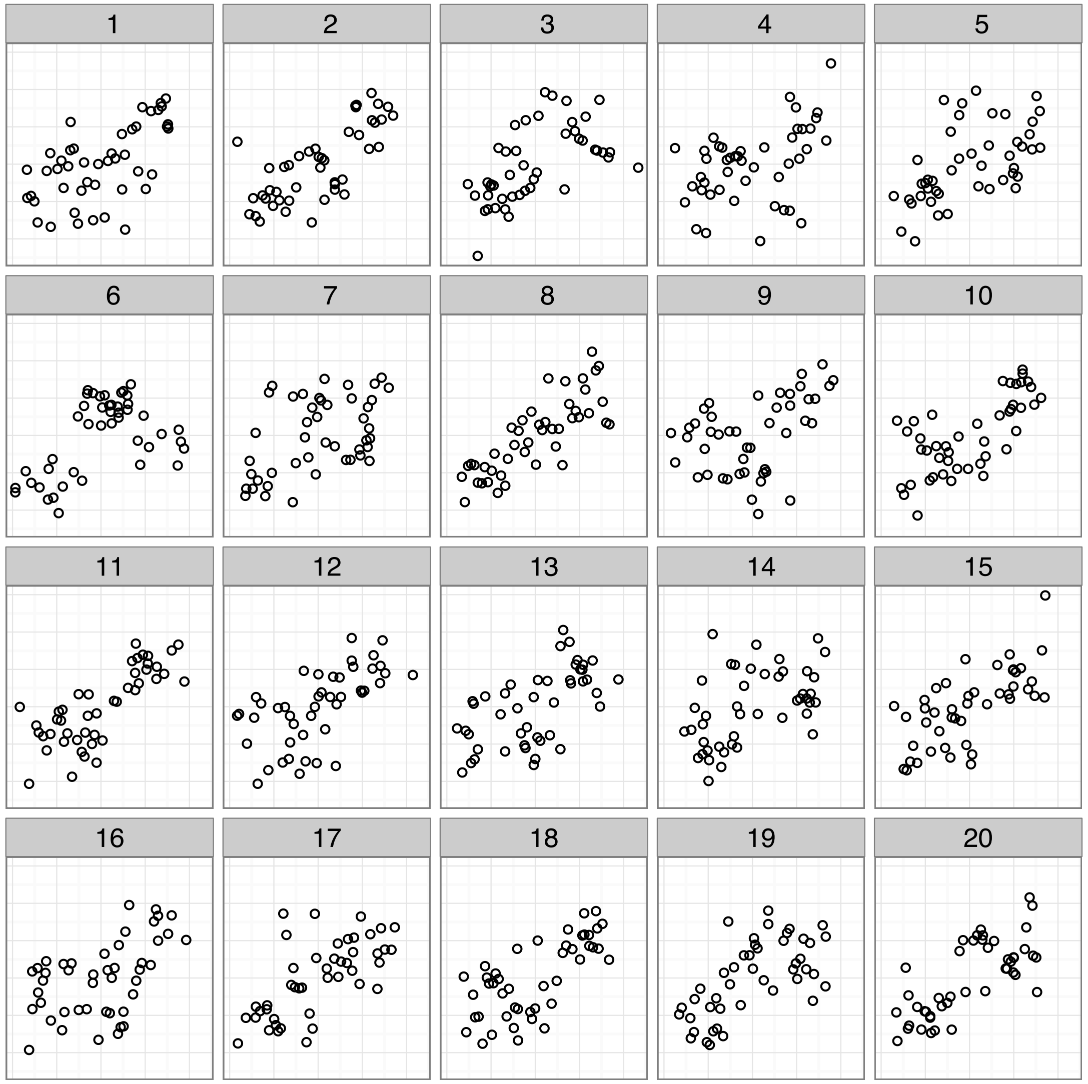
Our approach



- use lineup protocol to investigate charts `in their natural habitat’
- want to quantify how strongly aesthetics such as color and shape and additional features (lines, ellipses) influence pattern detection

The Lineup Protocol

- Buja et al (2009):
  data embedded among a set of 'null' plots
- Visual test of null hypothesis: "data and nulls are generated by the same mechanism"
- Human evaluator: "Which of these plots is the most different?"
- Data plot identification is evidence against the null hypothesis
- p-value based on #data identifications



Which of these plots is the most different?

The Lineup Protocol

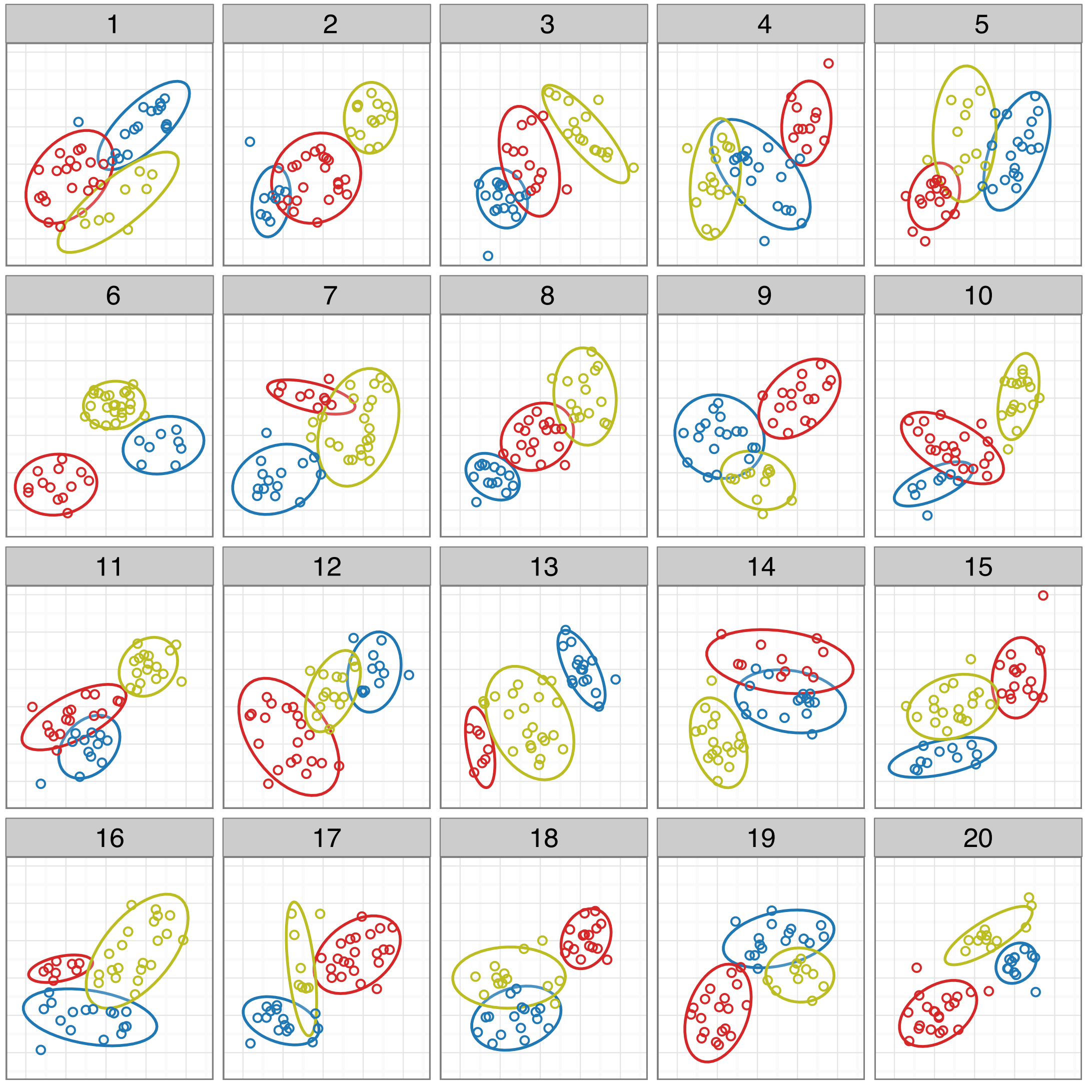
- Buja et al (2009):
  data embedded among a set of ‘null’ plots

- Visual test of null hypothesis: “data and nulls are generated by the same mechanism”

- Human evaluator: “Which of these plots is the most different?”

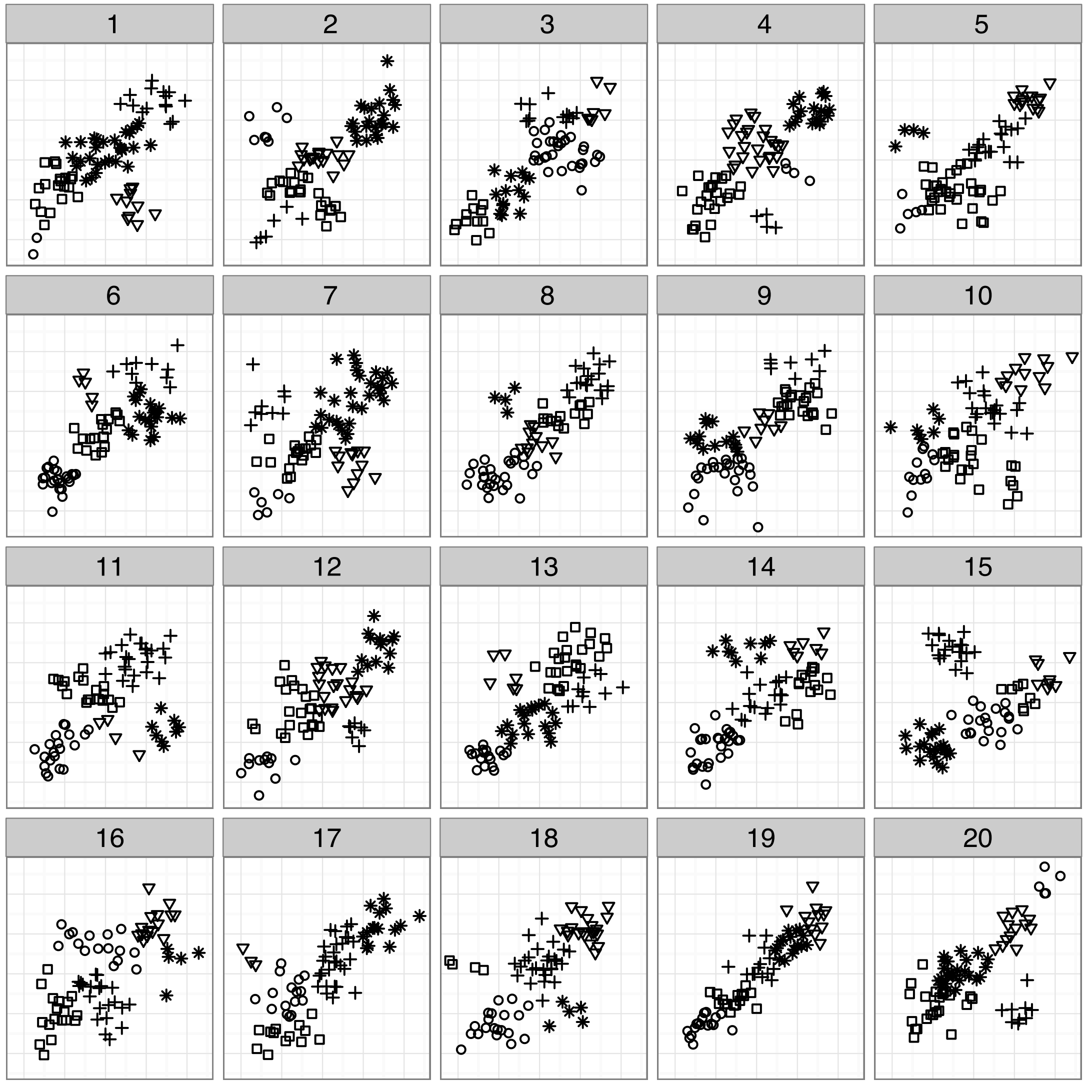
- Data plot identification is evidence against the null hypothesis

- p-value based on #data identifications



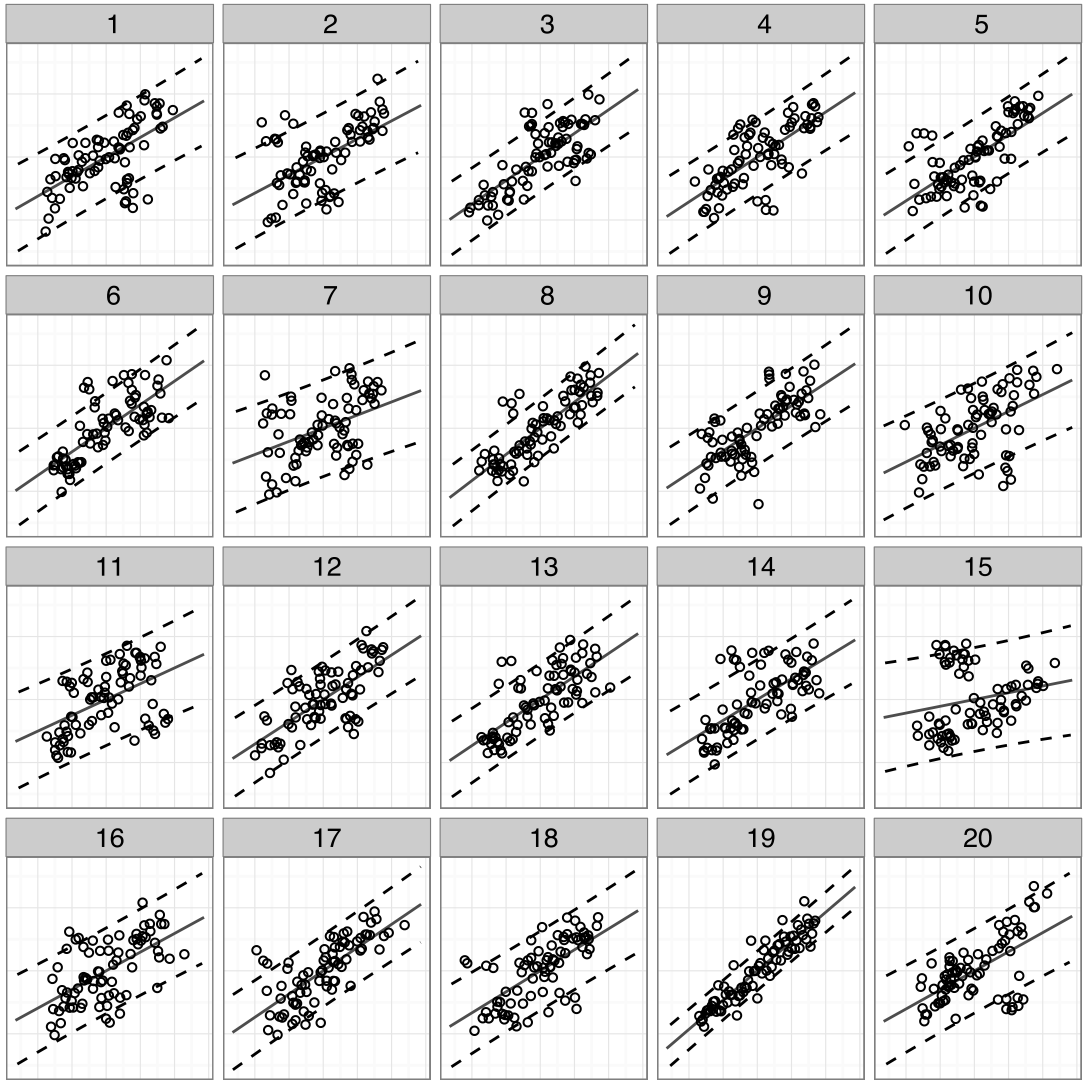
Which of these plots is the most different?

Another Example



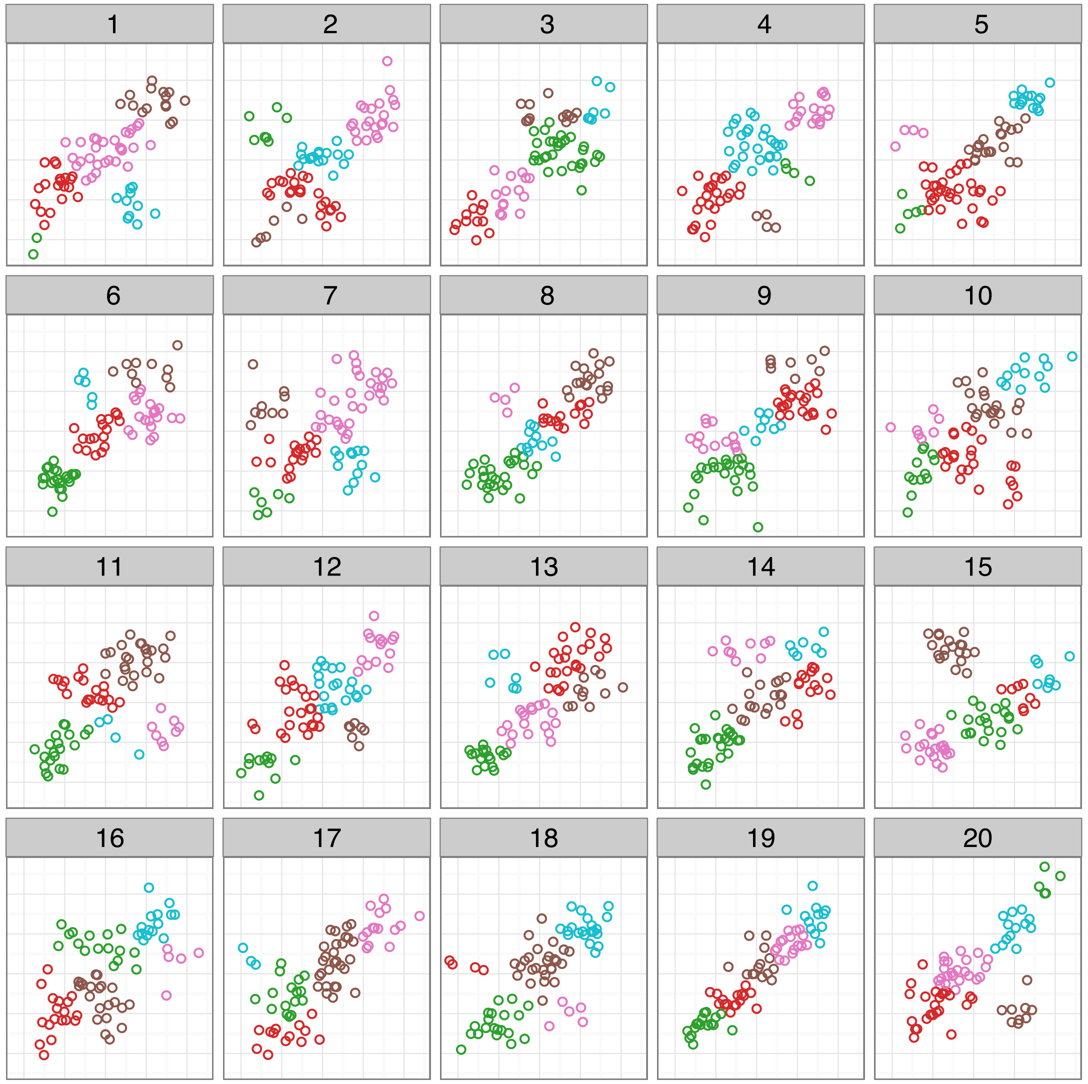
Which of these plots is the most different?

Another Example



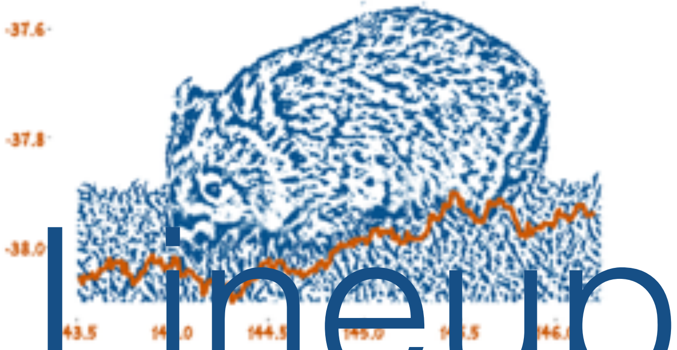
Which of these plots is the most different?

Another Example



Which of these plots is the most different?

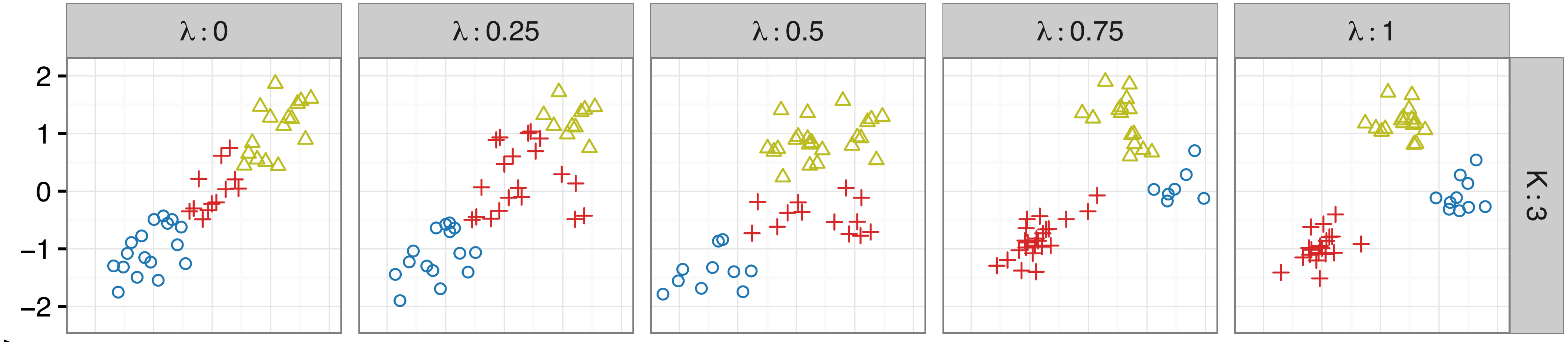
Modified Lineup



trend target

nulls

cluster target



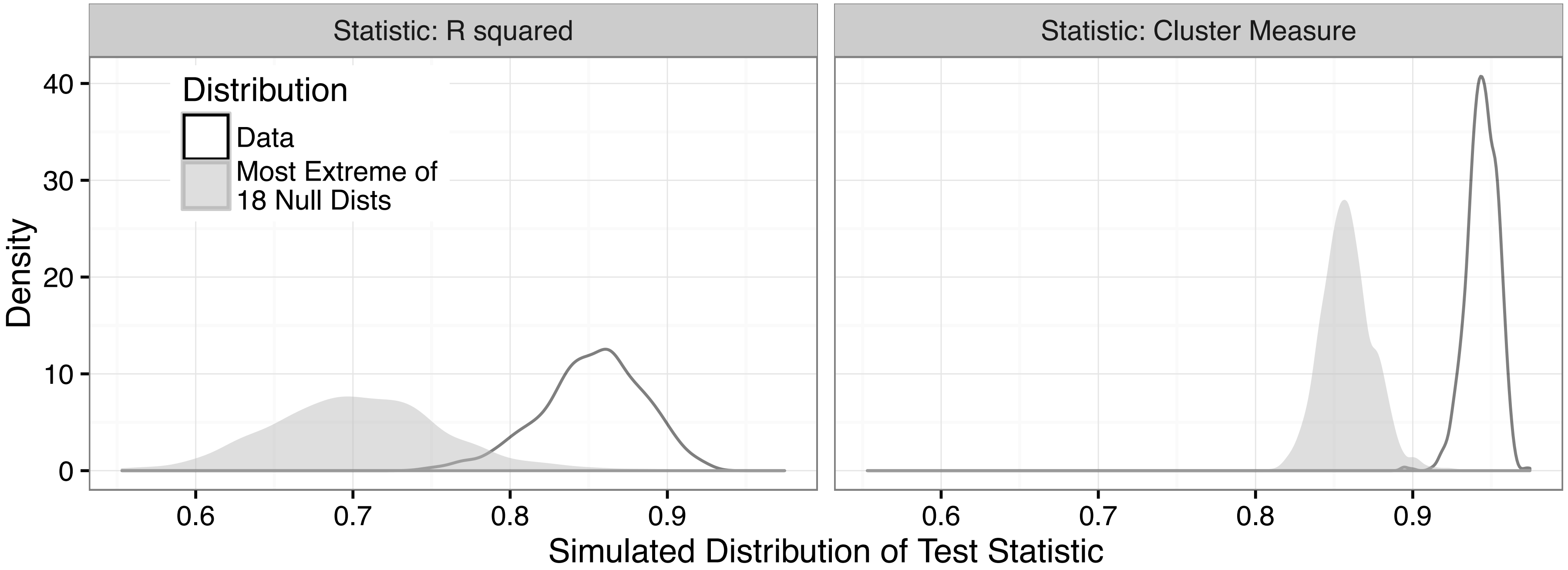
Model M_T
with parameter s_T

mixture

Model M_C
with parameter s_C

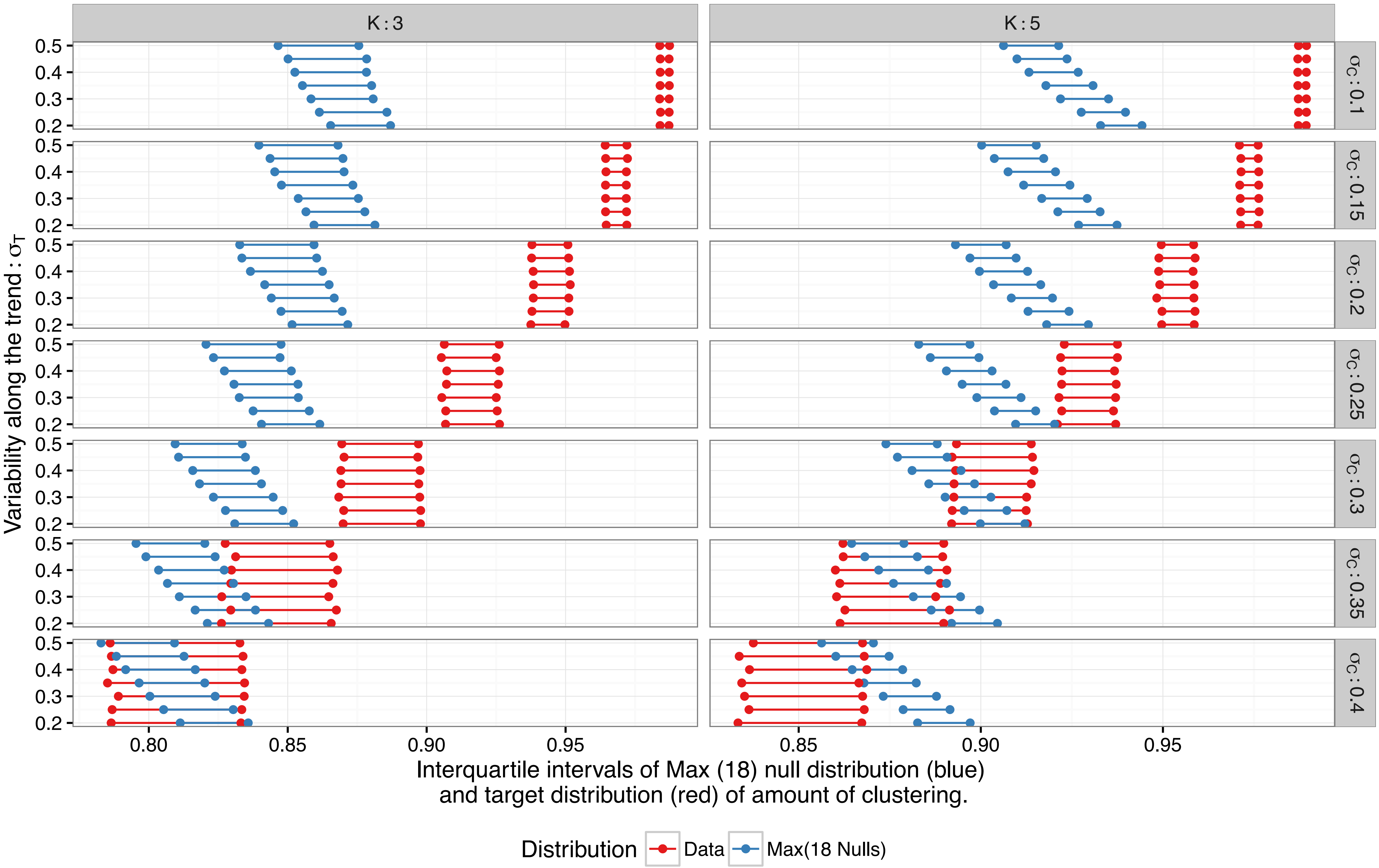
- two targets embedded in the lineup
- allows head-to-head evaluation of signal strength (satisfaction of search, Fleck et al 2010)
- choice of model parameters is tricky

Parameter settings



- Simulation: simulate 1000 data sets for s_T=0.25 and s_C = 0.2
- compute R^2 and cluster measure for data and max null
- we have a good chance of ‘seeing’ the targets in a lineup

Parameter space

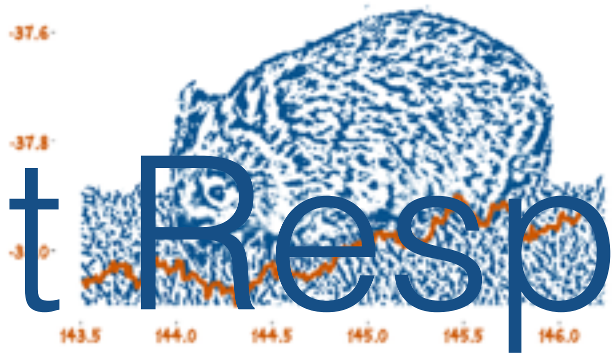


Designs: Cluster vs Trend

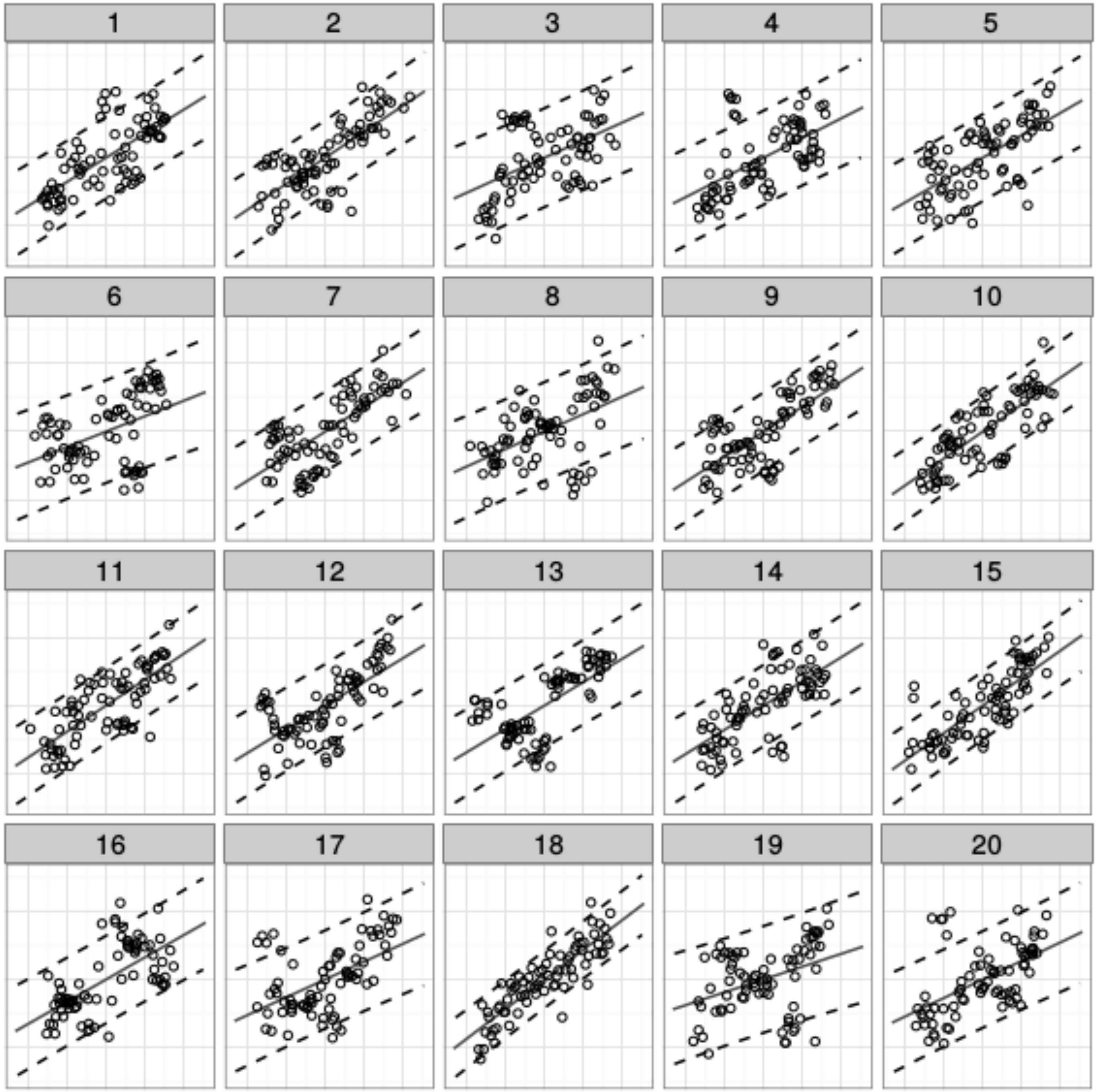
		Trend Emphasis		
	Strength	0	1	2
Cluster Emphasis	0	None	Trend	Trend + Error
	1	Color Shape	Color + Trend	
	2	Color + Shape Color + Ellipse		Color + Ellipse + Trend + Error
	3	Color + Shape + Ellipse		

AMT study

- Using AMT for recruiting participants (https://erichare.shinyapps.io/lineups/)
- requirements: at least 100 HITS, 95% success rate
- two successful pre-trial lineup evaluations
- Ten evaluations:
  one of each design,
  one of each of the nine parameter settings
- Result: 12010 lineup evaluations from 1201 participants

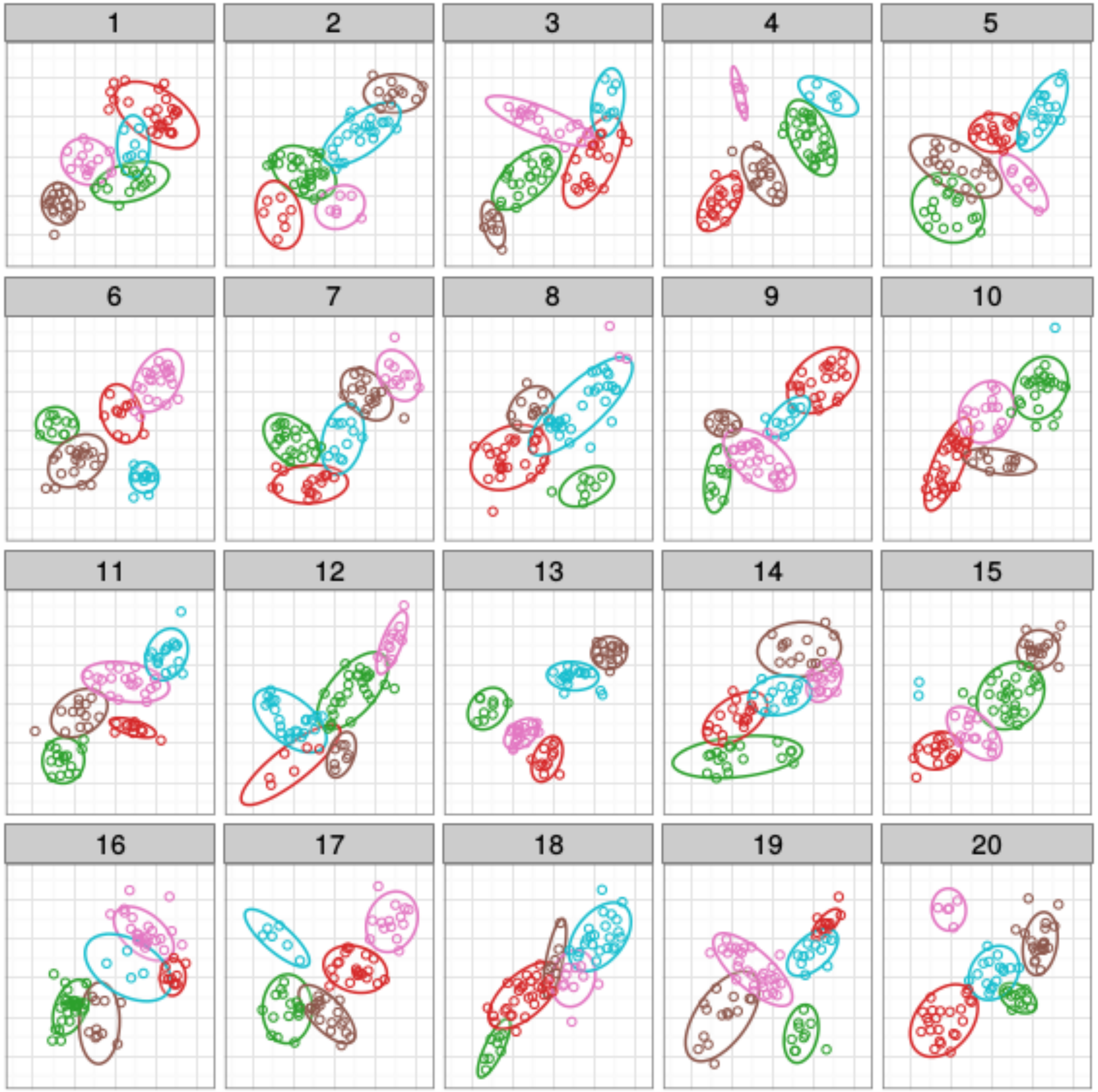


Participant Responses

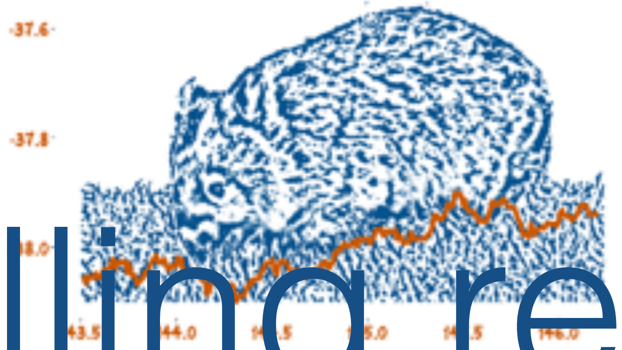


- Sample size: 22
- Trend target: 15
- Cluster target: 2
- Other: 5

Participant Responses



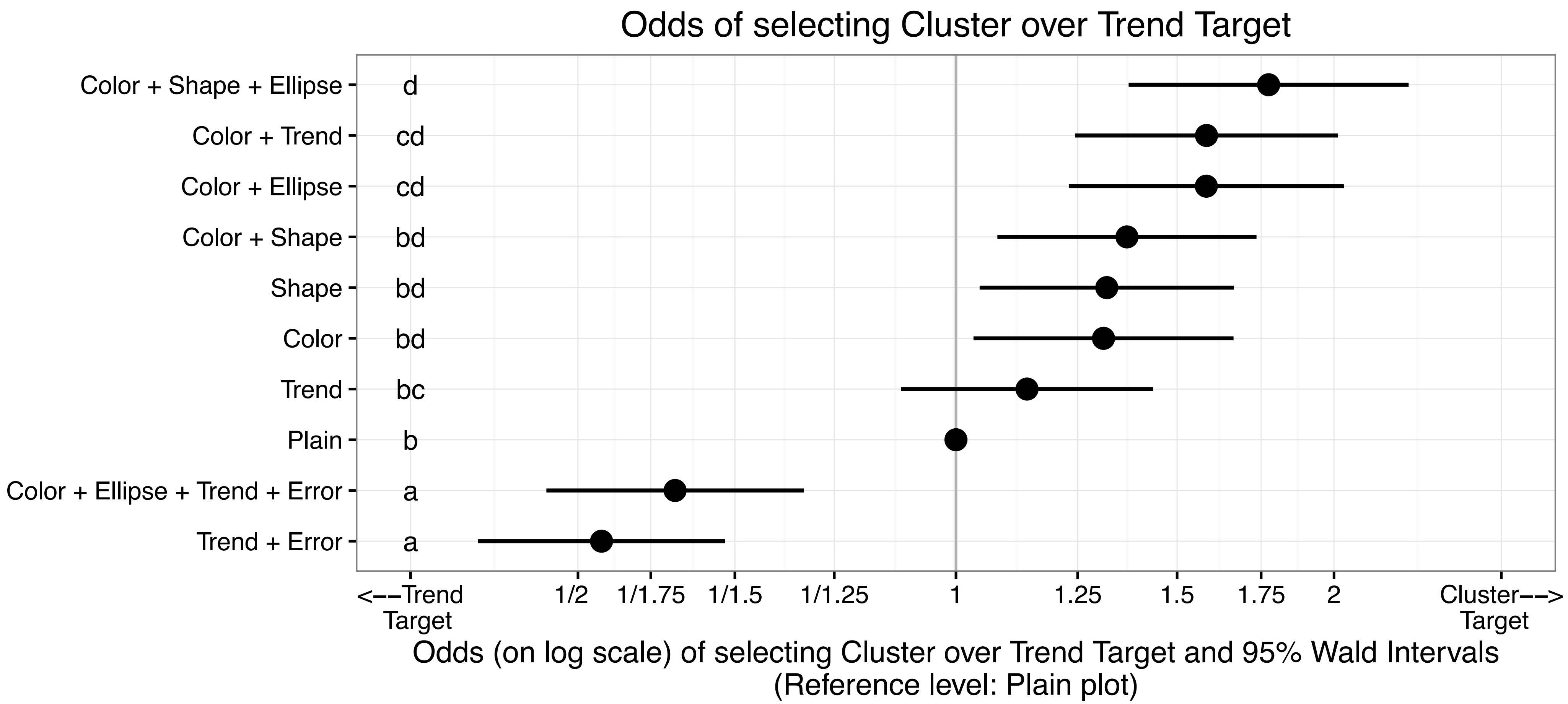
- Sample size: 14
- Trend target: 0
- Cluster target: 11
- Other: 3



Modelling results

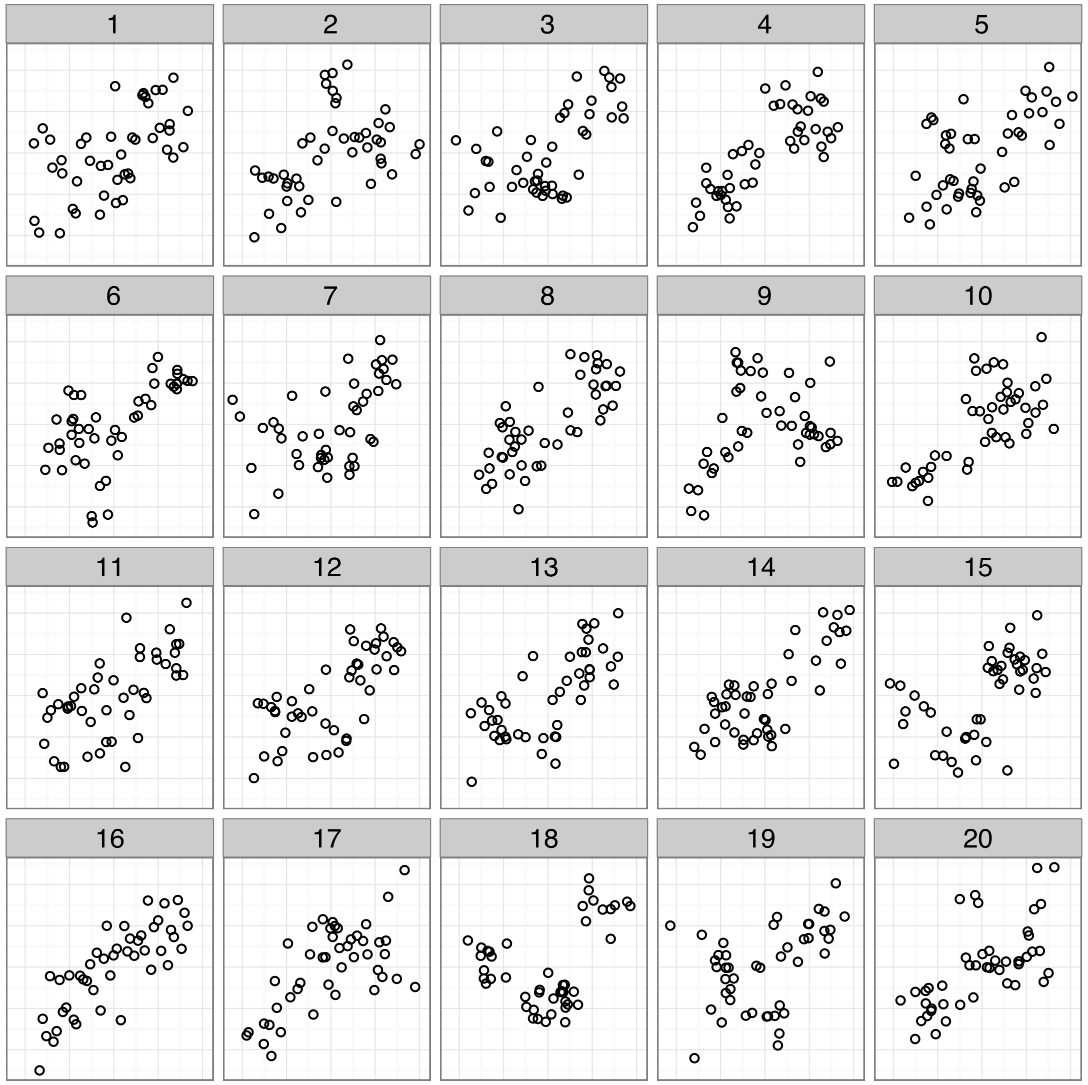
- Modelling balance between targets: subset on lineup evaluations that identified one of the targets (9959 out of 12010 evaluations)
- logistic regression of P(C | C u T)
- with random intercept for individuals' skills
  random intercept for data set difficulty

Cluster vs Trend



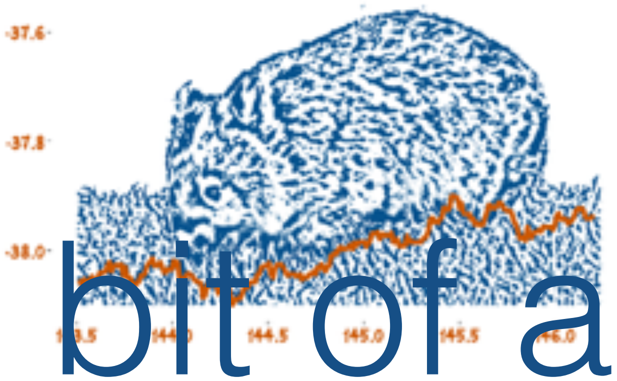
- generally the expected result
- mixed signals have mixed results
- control parameters s_T and s_C work as expected

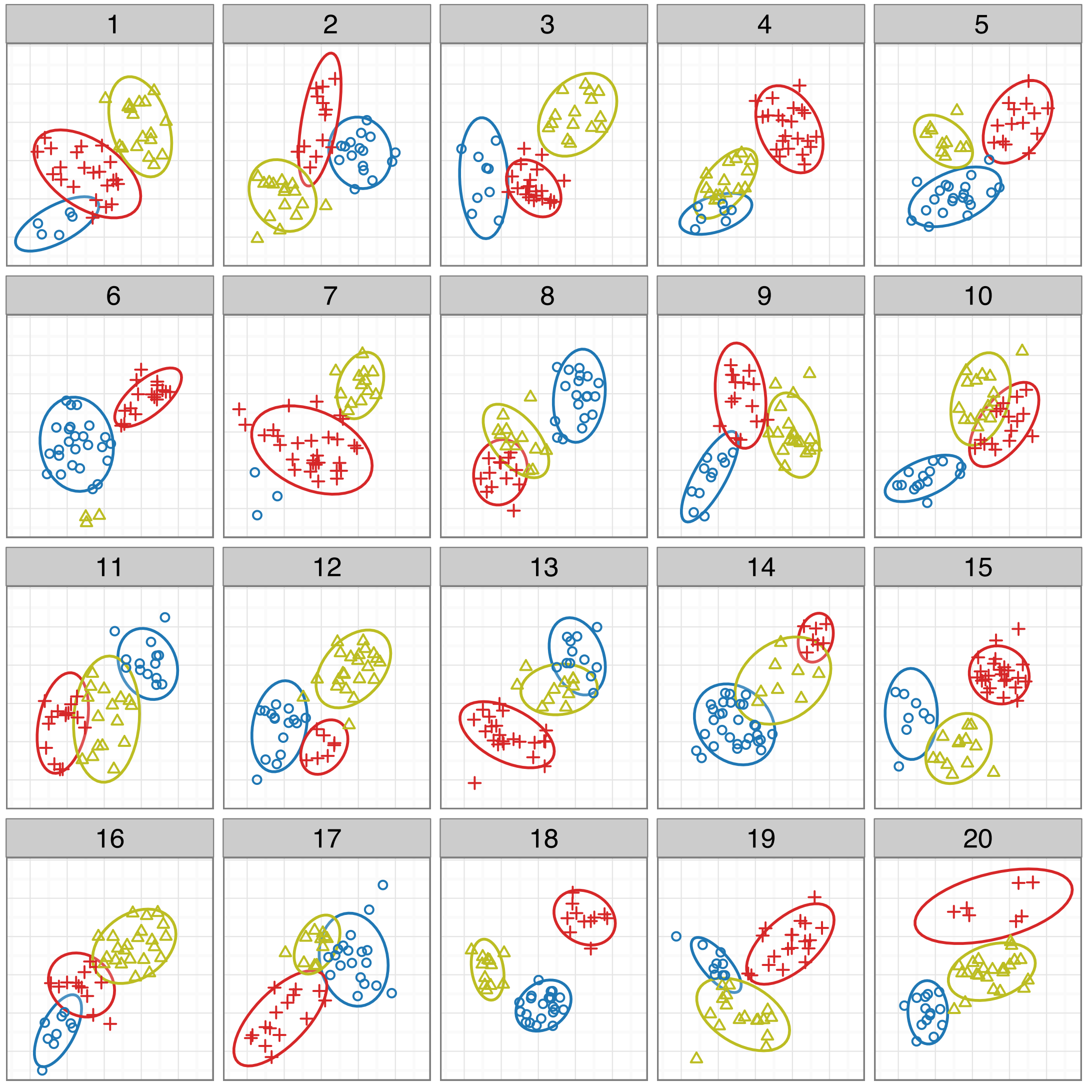
… and a bit of a surprise …



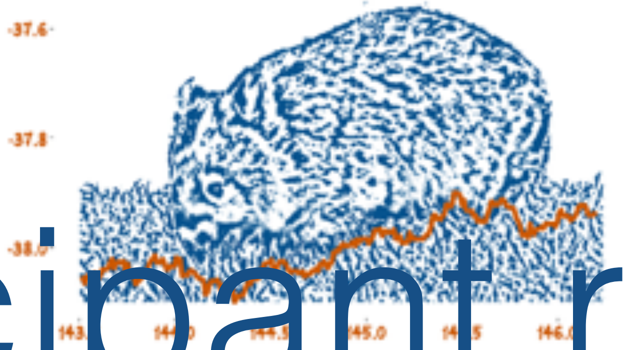
- fairly strong support for cluster target

… and a bit of a surprise …





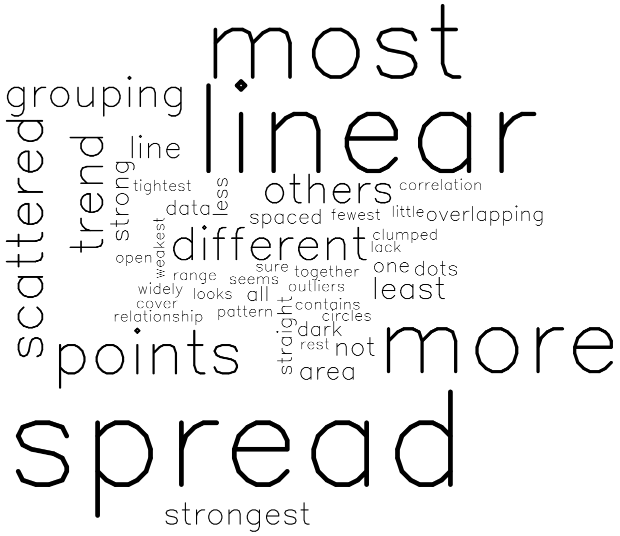
- support for cluster target not as strong???
- instead: #6, #7
- missing ellipses are a strong signal (single missing ellipse cuts probability by 44%)



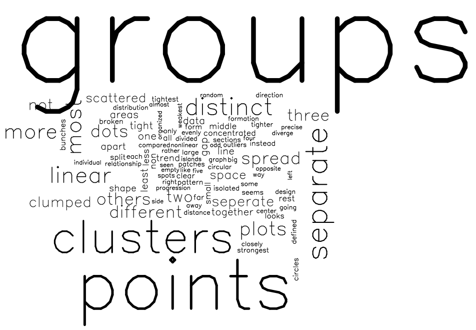
participant reasoning

- word cloud based on reason for choice:

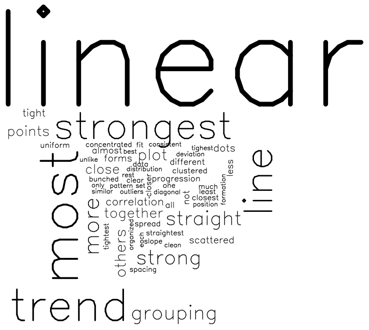
(a) Plain, neither target



(b) Plain, cluster target

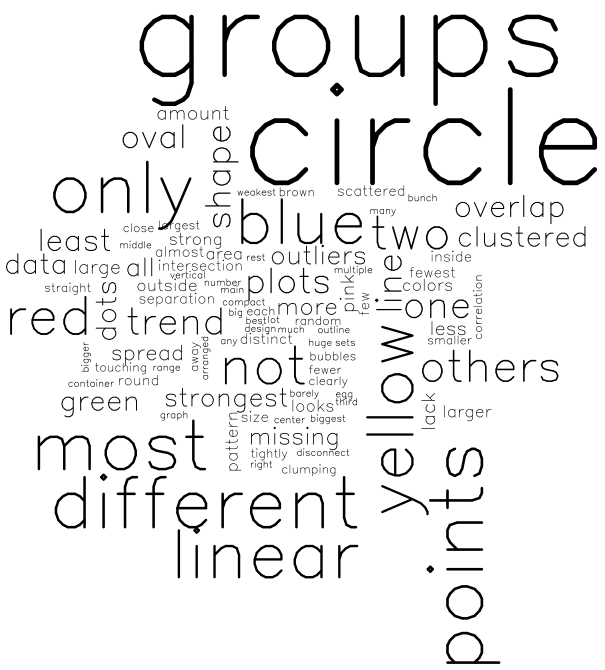


(c) Plain, trend target

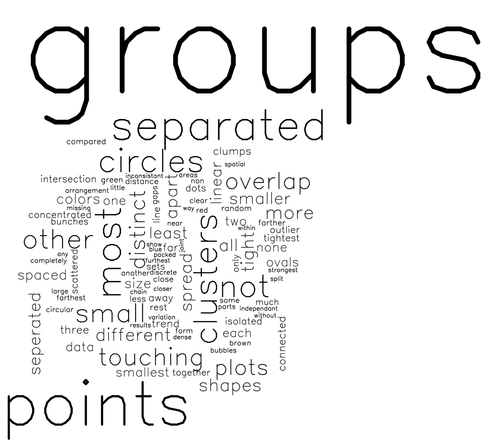


participant reasoning

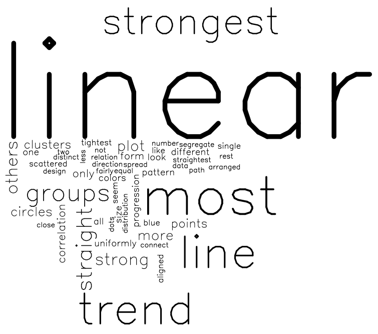
- word cloud based on reason for choice:



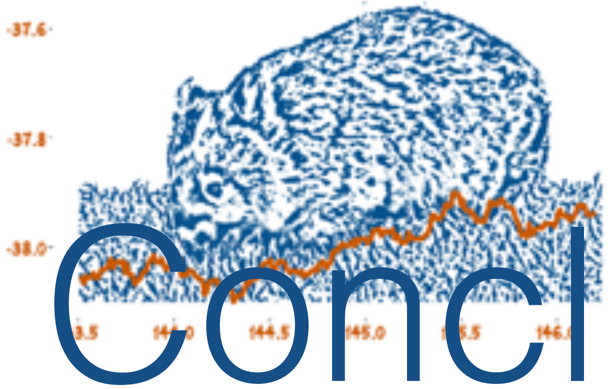
(j) Color + Ellipse, neither



(k) Color + Ellipse, cluster



(l) Color + Ellipse, trend



Conclusions

- Aesthetics matter, while not all significant, the trends follow the expectation:
color, shape and ellipses emphasize clustering
trend-line and predictions emphasize trends

- trend-line by itself might not be a particularly strong signal

- Human observers are extremely good at finding missing groups, if they expected them.