Clusters beat Trend!? Testing Feature Hierarchy in Statistical Graphics

Susan VanderPlas & Heike Hofmann,

Iowa State University

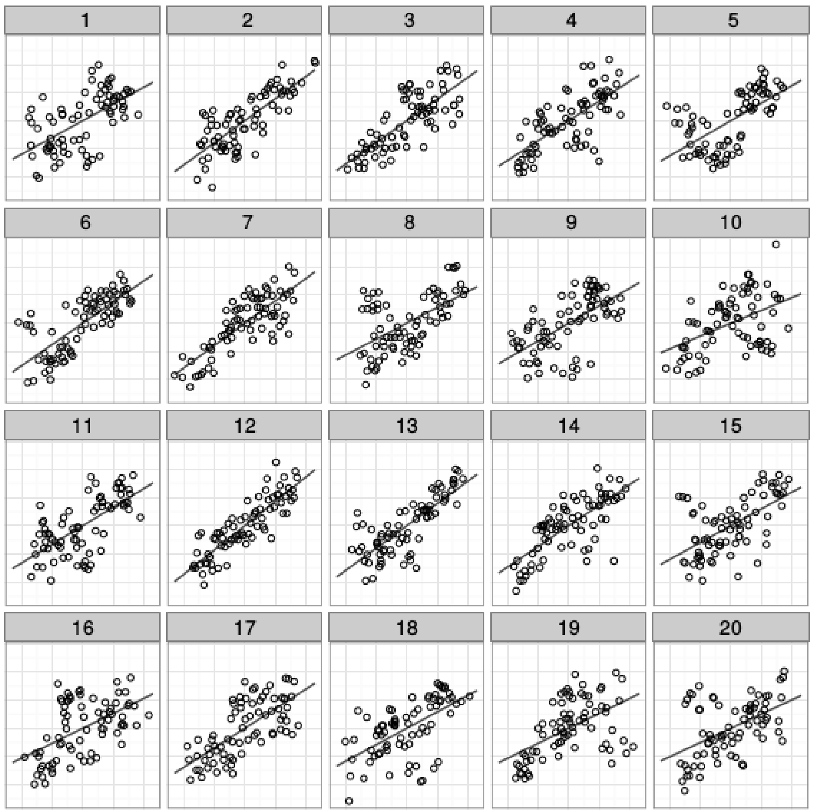
August 1, 2016

Outline

- Introduction
- Experiment Setup
  - Model
  - Plot Aesthetics
- Results
  - Target Identification
  - Clusters vs. Trend
  - Participant Reasoning
- Conclusions & Future Work

Introduction

Which plot is the most different?



Participant Responses

Plot 12: 52.2% (Trend target)

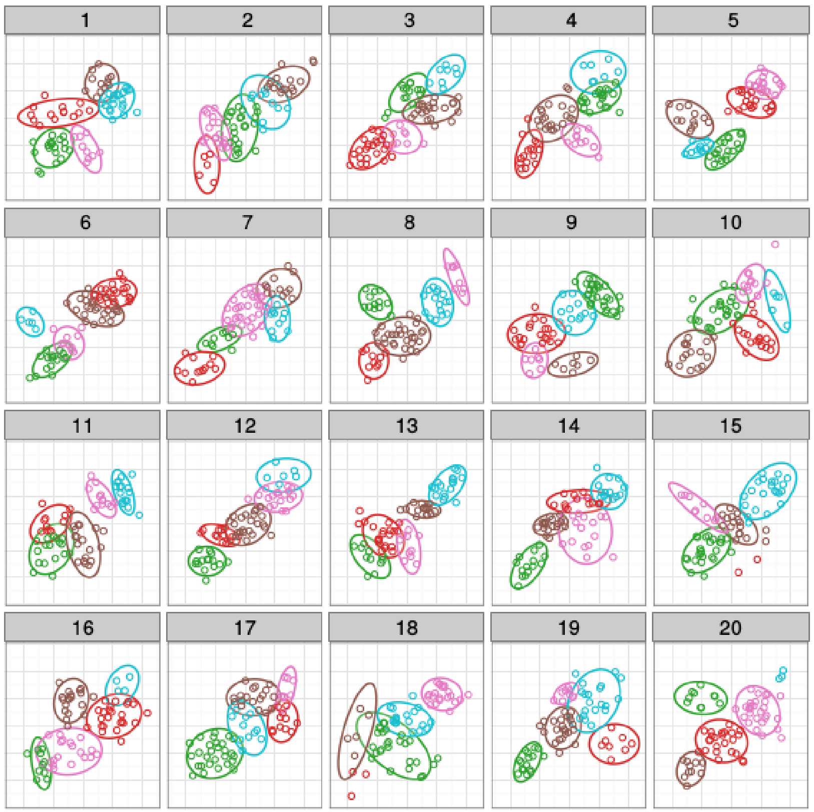
Plot 5: 17.4% (Cluster target)

Other: 30.3%

Sample size: 22

Trend target: 12, Cluster target: 5

Which plot is the most different?



Participant Responses

Plot 12: 9.4% (Trend target)

Plot 5: 28.1% (Cluster target)

Plot 18: 31.2%

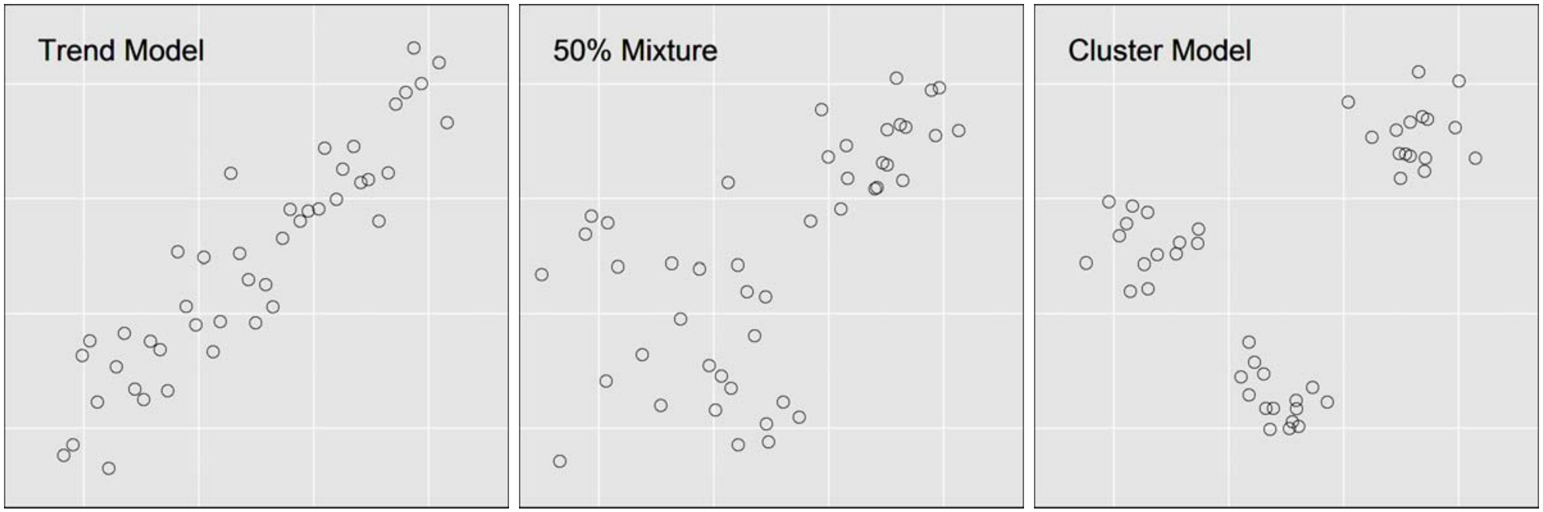
Other: 31.1%

Sample size: 31

Trend target: 12, Cluster target: 5

Experiment Design

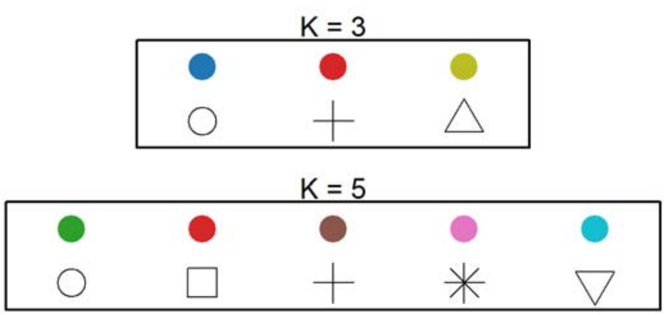
Data-Generating Models



Parameters		
σ_T: Variability in y	λ: Mixing parameter	K: # clusters σ_C: Variability around cluster centers

Plot Aesthetic Combinations

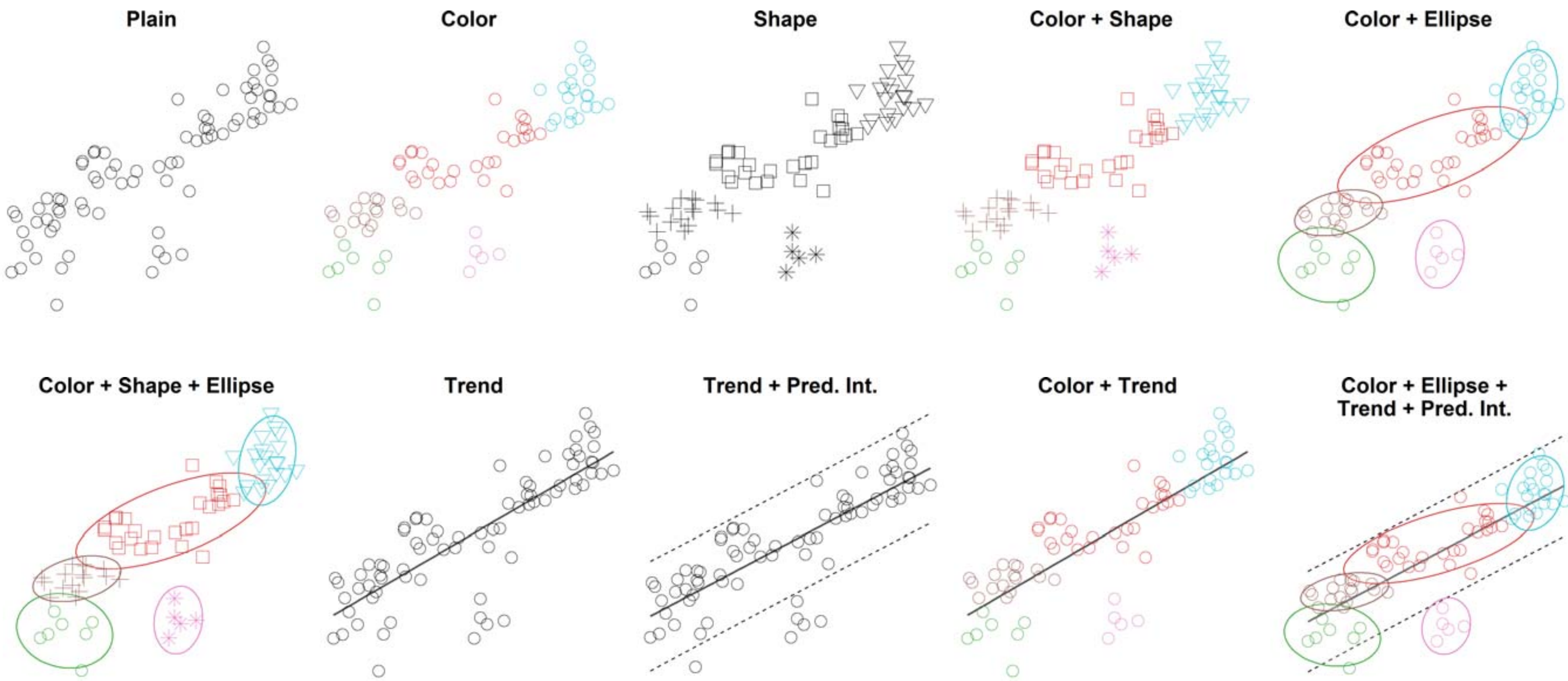
		Trend Emphasis		
	Strength	0	1	2
Cluster Emphasis	0	Plain	Line	Line + Pred. Interval
	1	Color Shape	Color + Line	
	2	Color + Shape Color + Ellipse		Color + Ellipse + Line + Pred. Interval
	3	Color + Shape + Ellipse		



Palettes selected to provide maximum perceptual distance (Ç. Demiralp, et al., 2014).

Shapes conform to guidelines in Robinson (2003) and Lewandowsky & Spence (1989).

Plot Aesthetic Combinations



Experimental Structure

Model Parameters

- Trend Strength $\sigma_T =$ easy, med., hard
- Cluster Strength $\sigma_C =$ easy, med., hard
- Number of Clusters $K = 3, 5$

Plot Aesthetics

- Plain
- Trend
- Trend + Pred. Int.
- Color + Trend
- Color + Ellipse + Trend + Pred. Int.
- Color
- Shape
- Color + Shape
- Color + Ellipse
- Color + Shape + Ellipse

Plot Level

- 18 parameter combinations
- 3 datasets/parameter combination
- 10 plot types for each dataset
  = 540 total plots

Evaluation Level

- Participants evaluate 10 plots:
  - 1 of each aesthetic
  - 1 of each combination of σ_T and σ_C randomized over K

Data Collection

(via Amazon Mechanical Turk)

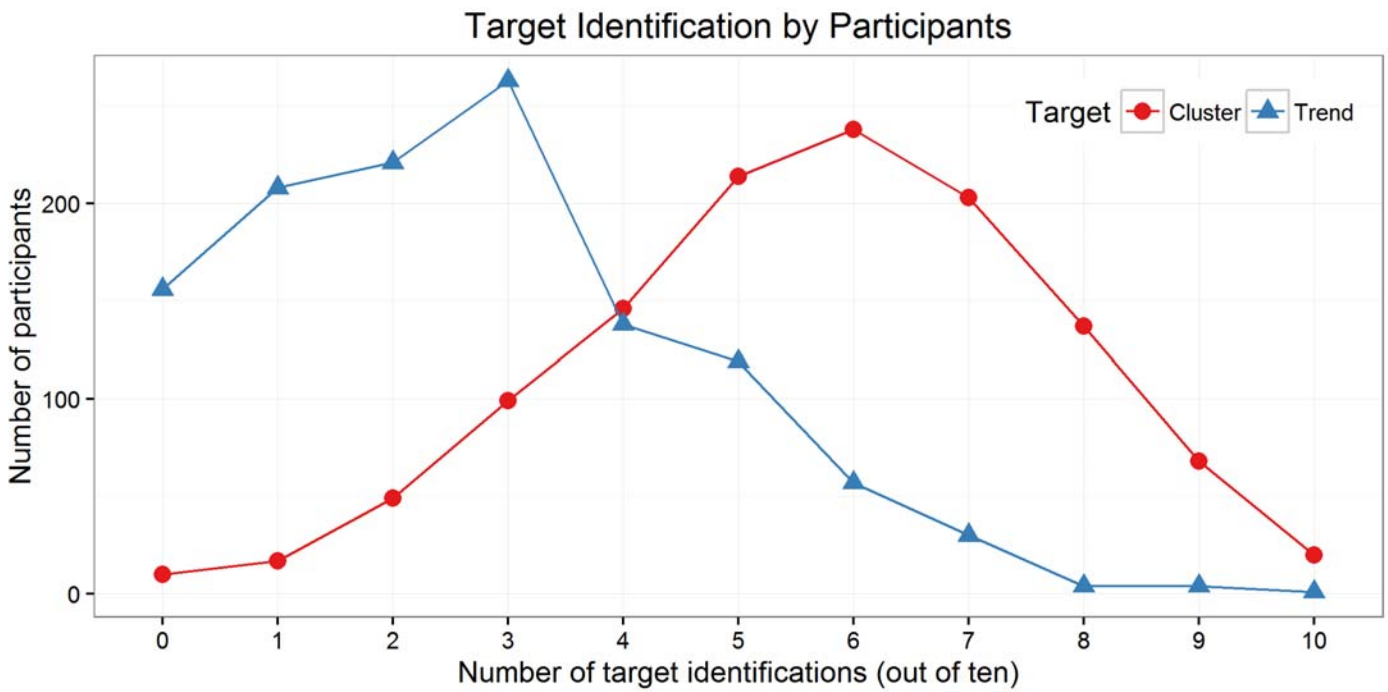
`1201` participants provided:

- Demographic information: age range, gender, education level
- 10 plot evaluations (`12010` total)
  - Target plot identification (one or more sub-plots)
  - Level of confidence in their answer (1 = least, 5=most)
  - Reasoning
    (i.e. "Strongest linear relationship", "Clustered points", "Odd shape")

Results

- Target Identification
- Cluster vs. Trend Target Selection
- Participant Reasoning

Target Identification



Participants selected more cluster targets than line targets.

5 plot types were expected to emphasize clustering; only 2 plot types were expected to emphasize trends.

Faceoff: Cluster vs. Trend?

Cluster vs. Trend

Define C_{ijk} to be the event

{Participant k selects the
cluster target for dataset j
with aesthetic set i},

and T_{ijk} to be the analogous selection of the trend target.

$$\text{logit } P(C_{ijk}|C_{ijk} \cup T_{ijk}) = \mathbf{W}\alpha + \mathbf{X}\beta + \mathbf{J}\gamma + \mathbf{K}\eta$$

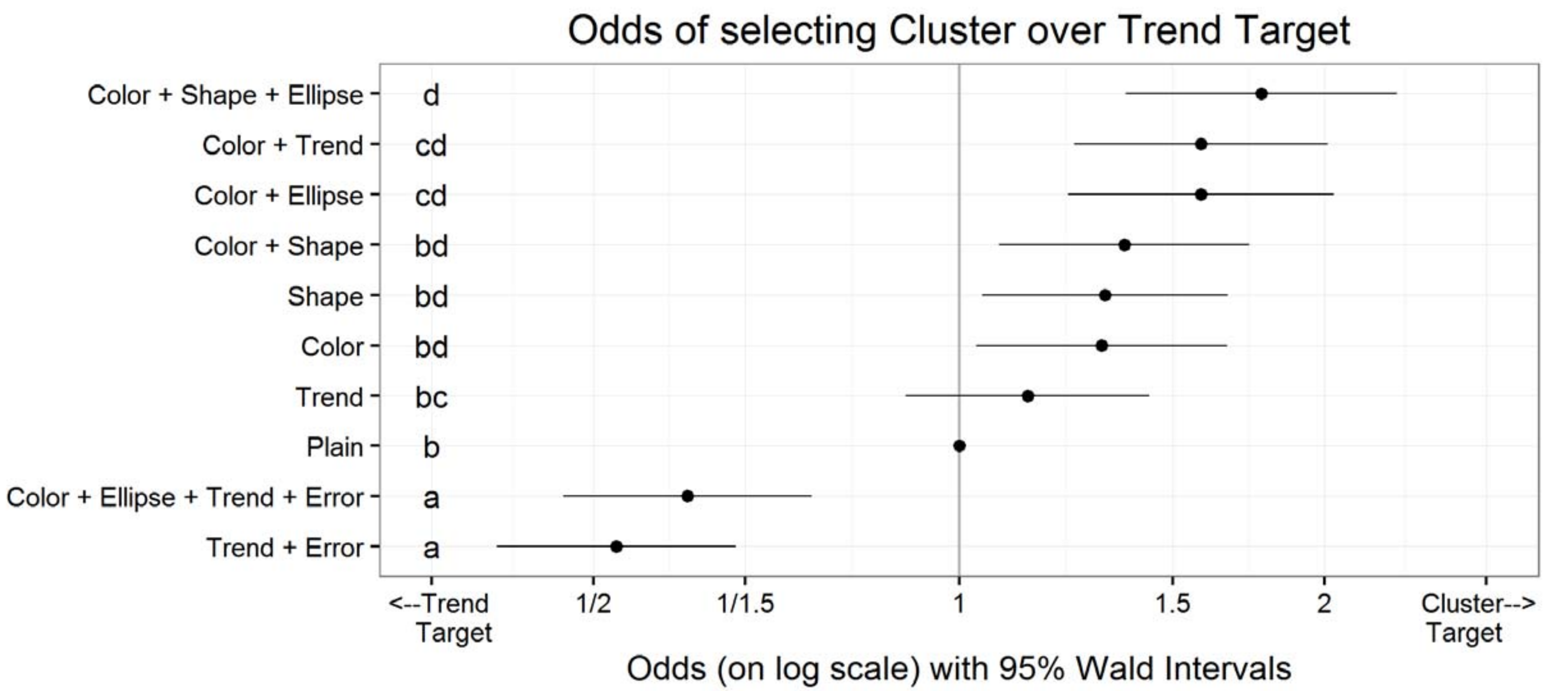
Cluster vs. Trend

Given that participants identified one of the two target plots...

α	data model fixed effects
β	effect of specific plot types
$\gamma_j \overset{iid}{\sim} N\left(0, \sigma^2_{\text{dataset}}\right)$	Dataset random effects
$\eta_k \overset{iid}{\sim} N\left(0, \sigma^2_{\text{participant}}\right)$	Participant random effects
$\epsilon_{ijk} \overset{iid}{\sim} N\left(0, \sigma^2_e\right)$	Individual evaluation errors

Dataset and participant effects are orthogonal by design

Cluster vs. Trend



Plot types are significantly different if they do not share a letter

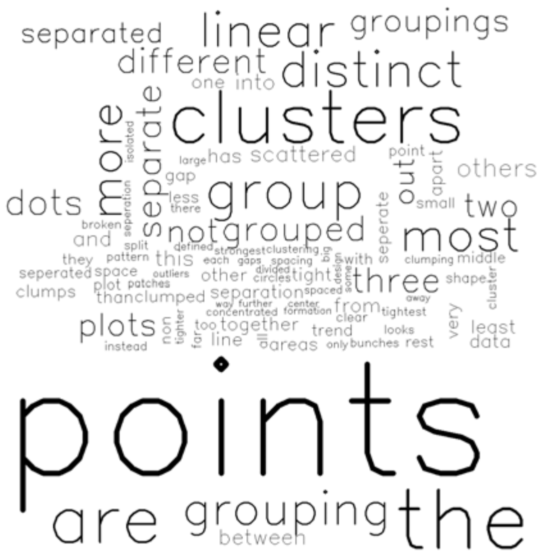
Participants are `0.52` times as likely to select cluster targets when plots have trend line and prediction interval aesthetics.

Participants are `1.77` times as likely to select cluster targets when plots have color, shape, and ellipse aesthetics.

Participant Reasoning

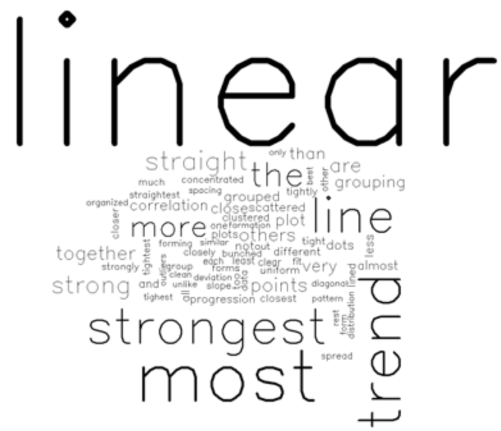
Participant Reasoning

Plain Plots



Neither Target
(N=127)

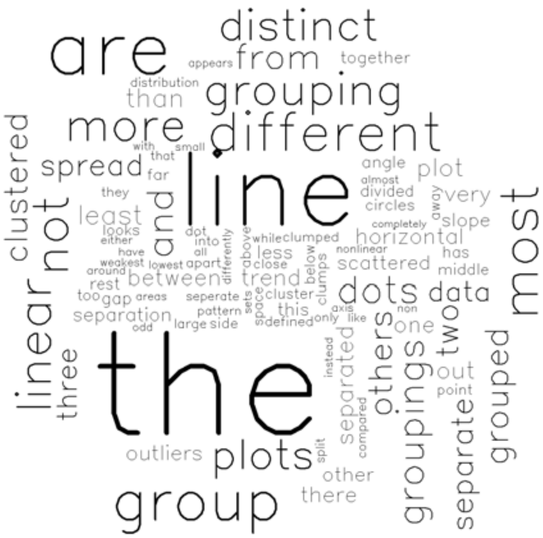
Cluster Target
(N=712)



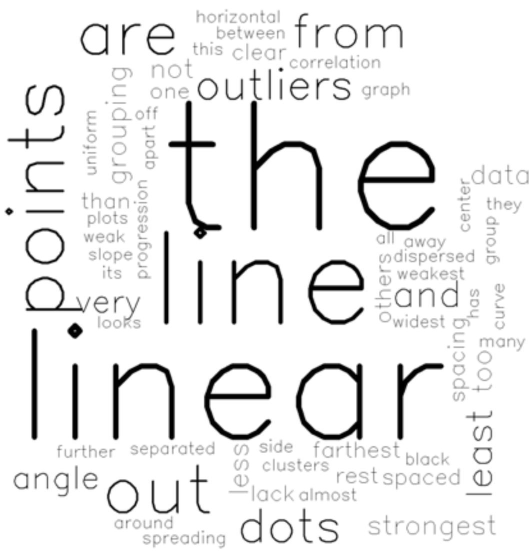
Trend Target
(N=355)

Participant Reasoning

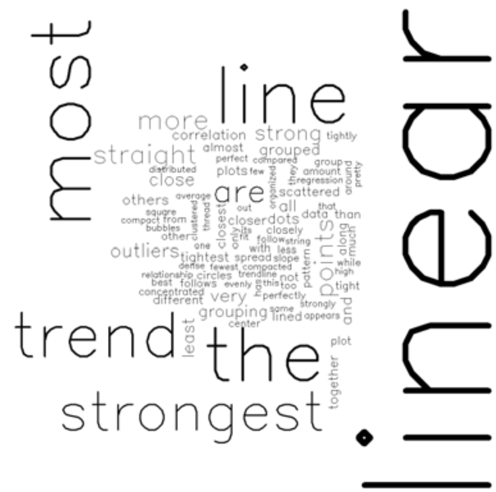
Trend line



Neither Target
(N=159)



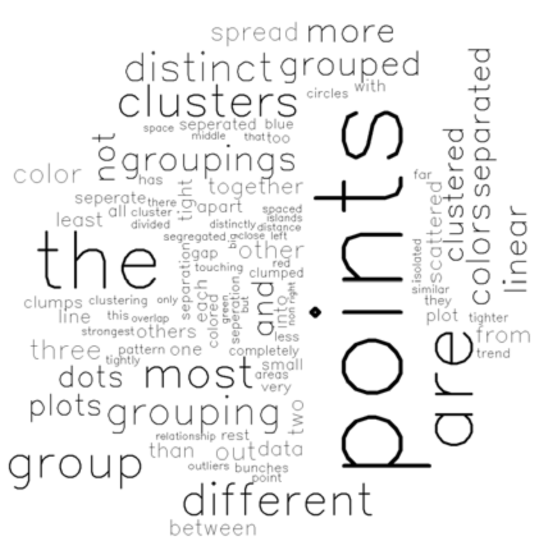
Cluster Target
(N=694)



Trend Target
(N=333)

Participant Reasoning

Color Plots



Neither Target
(N=188)

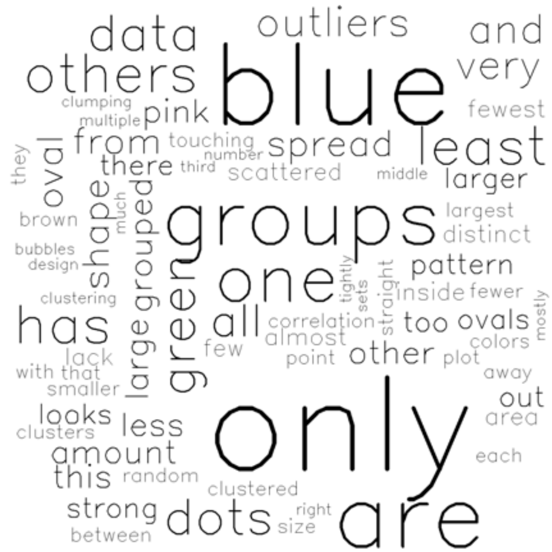
Cluster Target
(N=715)

Trend Target
(N=292)

Participant Reasoning

Color + Ellipse Plots

Neither Target
(N=347)



Cluster Target
(N=621)

Trend Target
(N=222)

Discussion

Conclusions

- Plot aesthetics influence perception of ambiguous data displays
- Aesthetic effects are not additive:
  Conflict conditions don't show similar/neutral results
- Aesthetics which recruit new gestalt heuristics have more influence, and we can quantify the size of that influence

More Information

- Github Repository (Data, paper, code)

  http://github.com/srvanderplas/FeatureHierarchy/

- JCGS Paper: Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics

  http://www.tandfonline.com/doi/abs/10.1080/10618600.2016.1209116

  - Lots of models examining trial completion time, parameter strength, and participant confidence
  - Much more in-depth treatment of gestalt perception
  - Simulation-based modeling of null plot characteristics

Future Work

- Restrict group sizes so null plots have the same objects as target plots
- Explore the effect of different types of common region for error bands and ellipses - shading, bounding boxes, etc.
- Test ellipse and error band aesthetics alone and with trend lines and color to examine interaction effects
- Test plotted statistics (trend line, ellipses, error bands) with and without data points to examine interactions between heuristics from the data and heuristics from summary statistics