Group beats Trend!?
Testing feature hierarchy in statistical graphics

Susan VanderPlas, Heike Hofmann*

February 16, 2015

Abstract

abstract goes here

1 Introduction and background

Discussion of pre-attentive visual features (Healey and Enns, 2012) - with a focus on hierarchy of pre-attentive features: color trumps shape - do we also see this in our results, and if so, by how much?

Intro to lineups (Buja et al., 2009; Majumder et al., 2013; Wickham et al., 2010; Hofmann et al., 2012)

The change to lineups we make is to introduce a second target to each lineup. We then keep track of how many observers choose any one of the two targets (to assess the difficulty of a lineup), and additionally we record how often observers choose one target over the other one. This is information that we can use to evaluate how strong the signal of one target is compared to the other one.

A further extension of this testing framework are the use of color (in a qualitative color scheme), the use of shapes, and additional density lines - we anticipate that all of these features are going to emphasize the clustering component. On the other hand, regression lines should emphasize any linear trends in the data.

2 Design Choices

We choose colors and shapes for the lineups in our study to be the most different from a set of ten choices as evaluated by participants in the study by Çağatay Demiralp et al. (2014) on the so called perceptual kernels.

Unfortunately, this limits the choice to the set used in the Tableau software. In order to produce experimental stimuli accessible to the approximately 4% of the population with red-green color deficiency (Gegenfurtner and Sharpe, 2001), we removed the grey hue from the palette. This modification produced maximally different color combinations which did not include red-green combinations, while also removing a color (grey) which is difficult to distinguish for those with color deficiency.

*Department of Statistics and Statistical Laboratory, Iowa State University

Shapes - there were some problems with reliable unicode representations in the lineups, which led to using slightly modified shapes. The left and right triangle shapes (available only in unicode using R) were excluded due to size differences between unicode and non-unicode shapes.

All shapes are non-filled shapes, which means that they are consistent with one of the simplest solutions to overplotting of points in the tradition of Tukey (1977) and Few (2009). For this reason we abstained from the additional use of alpha-blending of points to diminish the effect of overplotting in the plots.

How do the shapes we picked compare to the shapes discussed in Robinson (2003) see http://stat-computing.org/newsletter/issues/scgn-14-1.pdf?

3 Generating Model

We are working with two models M_C and M_T to generate data for the target plots. The null plots are showing data generate from a mixture model M_0. Both models generate data in the same range of values. We made also sure that data from the clustering model M_C shares the same correlation with the null data, while data from model M_T exhibits a similar amount of clustering as the null data.

3.1 Regression Model M_T

This model has the parameter σ_T to reflect the amount of scatter around the trend line.

Isn't there some centering or scaling missing? How do we make sure, that the regression model and the cluster model are on the same x scale? Only for $K = 3$ does the scaling in the clustering model boil down to values between [-1,1]. For $K = 5$ the values are further out.

Algorithm 3.1
Input Parameters: N points, σ_T standard deviation around the line, slope a (1 by default)
Output:

specify output

1. *Generate x_i, $i = 1, ..., N$, as a sequence of evenly spaced points from $[-1, 1]$ (σ_T added and subtracted to match the range of cluster points in x)*

   I don't understand the text in the parenthesis

2. *Jitter x_i by adding small uniformly distributed perturbations to each of the values: $x_i = x_i + \eta_i$, $\eta_i \sim$ Unif$(-z, z)$, $z = 1/5 * (2/(N - 1)$*

   There's something going on with the parentheses: two opening, only one closing; we also cannot write $x_i = x_i + \eta_i$ – that's programming, not math. so rename any of the pre-jittered xs to $\tilde{x}$.

3. *Generate y_i: $y_i = ax_i + e_i$, $e_i \sim N(0, \sigma_T^2)$*

We compute the correlation coefficient for all of the plots to assess the amount of linearity in each panel, computed as

$$r = 1 - RSS/TSS,$$

where TSS is the total sum of square, $TSS = \sum_{i=1}^{N}(y_i - \bar{y})^2$ and $RSS = \sum_{i=1}^{N} e_i^2$. The expected correlation coefficient ρ in this scenario is

$$\rho = \frac{\frac{1}{3}a^2}{\frac{1}{3}a^2 + \sigma_T^2},$$

because $E[RSS] = N\sigma_T^2$ and $E[TSS] = \sum_{i=1}^{N} E\left[y_i^2\right]$ (as $E[Y] = 0$), where

$$E\left[y_i^2\right] = E\left[a^2x_i^2 + e_i^2 + 2ax_ie_i\right] = \frac{1}{3}a^2 + \sigma_T^2.$$

Include some examples here.

3.2 Cluster Model M_C

We begin by generating K cluster centers on a $K \times K$ grid, then we generate points around the cluster center.

Algorithm 3.2
Input Parameters: N points, K clusters, σ_C cluster standard deviation
Output:
specify output

1. *Generate cluster centers (c_i^x, c_i^y) for each of the K clusters, $i = 1, ..., K$:*
   (a) *in form of two vectors c^x and c^y of permutations of $\{1, ..., K\}$, such that*
   (b) *the correlation between cluster centers $Cor(c^x, c^y)$ falls into a range of $[.25, .75]$.*
2. *Center and standard-normalize cluster centers (c^x, c^y):*

$$\tilde{c}_i^x = \frac{c_i^x - \bar{c}}{s_c} \quad \text{and} \quad \tilde{c}_i^y = \frac{c_i^y - \bar{c}}{s_c},$$

   where $\bar{c} = K(K+1)/2$ and $s_c^2 = \frac{K(K+1)(2K+1)}{6} - \frac{K^2(K+1)^2}{4}$ for all $i = 1, ..., K$.
3. *For the K clusters, we want to have nearly equal sized groups, but uphold some variability. Group sizes are therefore determined as a draw from a multinomial distribution: determine group sizes $g = (g_1, ..., g_K)$, with $N = \sum_{i=1}^{K} g_i$, for clusters $1, ..., K$ as a random draw*

$$g \sim Multinomial(K, p) \text{ where } p = \tilde{p} / \sum_{i=1}^{K} \tilde{p}_i, \text{ for } \tilde{p} \sim N(\frac{1}{K}, \frac{1}{2K^2}).$$

4. *Generate points around cluster centers:*

(a) $x_i = c^x_{g_i} + e^x_i$, *where* $e^x_i \sim N(0, \sigma^2_C)$
(b) $y_i = c^y_{g_i} + e^y_i$, *where* $e^y_i \sim N(0, \sigma^2_C)$

As a measure of clustering we use a coefficient to assess the amount of variability taken care of by including the grouping variable, compared to total variability. Note that for the purpose of clustering, variability is measured as the variability in both x and y from a common mean, i.e. we implicitly assume that the values in x and y are on the same scale (which we can easily achieve by an appropriate re-scaling).

some examples? with colour?, K = 3 ... 6? maybe two different standard deviations?

3.3 Null Model M_0

The generative model for null data is created as a mixture model M_0 that draws $n_c \sim B_{N,\lambda}$ observations from the cluster model, and $n_T = N - n_c$ from the regression model M_T.

and a set of examples, here, with varying lambda

4 Experimental Setup

4.1 Design

Factors:

Parameter	Description	Choices
K	# Clusters	3, 5
N	# Points	$15 \cdot K$
σ_T	Scatter around trend line	.25, .35, .45
σ_C	Scatter around cluster centers	.20, .25, 0.30, .35

Table 1: Parameter settings for Data Generation.

Emphasis	Aesthetics
Control	–
Group	Color, Shape
	Color + Shape, Color + Ellipse, Color + Shape + Ellipse
Trend	Line
	Line + Error band
Conflict	Color + Trend Line,
	Color + Ellipse + Trend Line + Error band

Table 2: Aesthetics and add-on design choices.

Based on simulations from the null we get a distribution for each of the two quality measures for the targets under each parameter setting. This gives us an objective measure to assess the difficulty of detecting each of the targets. An overview of the results can be seen in the appendix.

Should we keep these two histograms to explain in more detail how the simulation from the null model works?

Figures 1 and 2 show histograms of the marginal densities ... The red lines show ten samples each from the trend model and the cluster model. The lines for the cluster are, relatively, further to the right of the overall distribution than the red lines for the trend model, indicating that $\sigma_C = 0.25$ is producing target plots that are a bit easier to spot than trend targets with a parameter value of $\sigma_T = 0.30$.

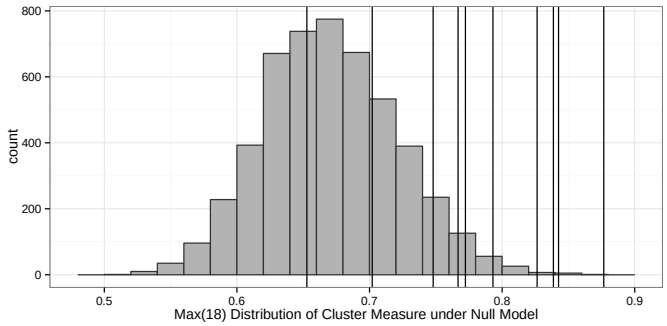


Figure 1: Histogram of R^2 values from 5000 data sets of the null model ($K = 3, N = 45, \sigma_C$ =0.25, σ_T =0.3). The lines in black are R^2 values of ten sample data sets from the Trend model M_T.

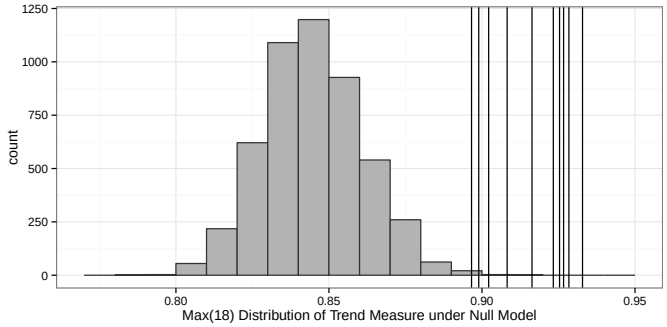


Figure 2: Histogram of the distribution of the cluster measure based on 5000 data sets from the null model ($K = 3, N = 45, \sigma_C$ =0.25, σ_T =0.3). The lines in black correspond to the cluster measure from ten sample data sets from the Clustering model M_C.

Design choices

1. Plain: two targets with data from one of each of the two generative models are included in a set of eighteen panels of null data.
2. Color/Shape: points in each of the panels are colored/marked based on the results of a hierarchical clustering .
3. Trend line: a line of the least square fit is drawn through the points.
4. Color & Shape
5. Color & trend line: this emphasises both the clustering and the regression - it is not clear, which signal will be stronger.
6. Color & Ellipsoids: around the groups of the same color, ellipsoids are drawn to reflect the 95% density estimate.

4.2 Hypothesis

The plot most identified as the "target" will change based on plot aesthetics which emphasize linear features or cluster features. This effect will be mediated by the signal strength of the line and cluster features.

Could you re-phrase these items so they always give an increase in the visual outcome? (or always a decrease, but the emphasis is on the outcome not on increasing the parameters).

- Increasing N will increase signal strength for both line and clusters
- Increasing K will decrease signal strength for clusters (at the same variability, there's less space to spread clusters out resulting in less visual separability)
- Increasing σ_T will decrease signal strength for lines
- Increasing σ_C will decrease signal strength for clusters

Plot features will emphasize either lines or clusters as follows:

- None (control)
- Color (cluster emphasis)
- Shape (cluster emphasis)
- Color + shape (double cluster emphasis)
- Ellipse + color (double cluster emphasis)
- Line
- Line + Prediction Interval (double line emphasis)
- Color + line (conflict)

- Color + line + Prediction Interval (conflict)

The primary purpose of the study is to detect how using visual aesthetics emphasizes one of two features and lead to detection of one feature over another feature.

A secondary purpose of the study is to relate signal strength to detection in a visualization by a human observer.

In a more organized representation:

		Line Emphasis		
		0	1	2
Cluster Emphasis	0	None	Line	Line + Prediction
	1	Color, Shape	Color + Line	
	2	Color + Shape Color + Ellipse		Color + Ellipse + Line + Prediction
	3	Color + Shape + Ellipse		

4.3 Experimental Design

Initially, assume a fully factorial, balanced design, with r unique datasets per parameter set (replicates) and P evaluations per (aesthetic|dataset). The experiment is conducted at three levels: parameter sets (with replication, so EUs are data sets), plot types (i.e. a certain set of aesthetics), and participant evaluations. At the first level, there are three parameters: $K \in \{3, 5\}$, $\sigma_T \in \{.3, .4, .5, .6\}$, and $\sigma_C \in \{.2, .25, .3, .35, .4\}$. At the second level, there are blocks (by data set), and then 10 aesthetic combinations.

We'll have to use contrasts to measure the effect of color individually, etc., for now let's just consider the ANOVA evaluation

Finally, at the lowest level, there are participant effects.

At the participant level, we need to decide if we're going to fully randomize, try to block, etc. - are participants going to get 10 different data sets? 5? Not sure how to conceptualize that, and I would imagine it will affect how we organize model evaluation. Grr, I hate mixed models.
HH: yes, I would assume that participants get ten plots each, one from each of the designs in a random order. (We have the data base set up that way).

Modified from Table 10.6 (pg 181) of Design of Experiments by Dr. Morris. The table in the book has a four-factor split plot design with three levels (randomized, block, block).

We have a couple of options:

- keep the full factorial experiment, use one (at most two) replicates, and use higher level factorial effects to beef up any error variance terms.
- Do a full factorial experiment for $K = 3$ and use a subset of the factorial experiment for $K = 5$ (either using a subset of cases for σ_T and σ_C, or a subset of combinations of the two cases/fractional factorial.)

Level	Factor	Source	DF	Sum of Squares
Dataset	K	α	1	$\sum_i (4)(5)(r)(10P)(\overline{y}_{i\cdots} - \overline{y}_{\cdots})^2$
	σ_T^2	β	3	$\sum_j (2)(5)(r)(10P)(\overline{y}_{\cdot j\cdots} - \overline{y}_{\cdots})^2$
	σ_C^2	γ	4	$\sum_i (2)(4)(r)(10P)(\overline{y}_{\cdot\cdot k\cdots} - \overline{y}_{\cdots})^2$
		$(\alpha\beta)$	3	$\sum_{ij} (5)(r)(10P)(\overline{y}_{ij\cdots} - \overline{y}_{i\cdots} - \overline{y}_{\cdot j\cdots} + \overline{y}_{\cdots})^2$
		$(\alpha\gamma)$	4	$\sum_{ik} (4)(r)(10P)(\overline{y}_{i\cdot k\cdots} - \overline{y}_{i\cdots} - \overline{y}_{\cdot\cdot k\cdots} + \overline{y}_{\cdots})^2$
		$(\beta\gamma)$	12	$\sum_{jk} (2)(r)(10P)(\overline{y}_{\cdot jk\cdots} - \overline{y}_{\cdot j\cdots} - \overline{y}_{\cdot\cdot k\cdots} + \overline{y}_{\cdots})^2$
		$(\alpha\beta\gamma)$	12	$\sum_{ijk} (r)(10P)(\overline{y}_{ijk\cdots} - \overline{y}_{i\cdots} - \overline{y}_{\cdot j\cdots} - \overline{y}_{\cdot\cdot k\cdots} + \overline{y}_{ij\cdots} + \overline{y}_{\cdot jk\cdots} + \overline{y}_{i\cdot k\cdots} - \overline{y}_{\cdots})^2$
		Resid.	$(2)(4)(5)(r-1)$	$\sum_{ijkl} (10P)(\overline{y}_{ijkl\cdot\cdot} - \overline{y}_{i\cdots} - \overline{y}_{\cdot j\cdots} - \overline{y}_{\cdot\cdot k\cdots} + \overline{y}_{ij\cdots} + \overline{y}_{\cdot jk\cdots} + \overline{y}_{i\cdot k\cdots} - \overline{y}_{ijk\cdots} - \overline{y}_{ij\cdot l\cdot\cdot} - \overline{y}_{i\cdot kl\cdots} - \overline{y}_{\cdot jkl\cdots} + \overline{y}_{\cdots})^2$
	Total		$(2)(4)(5)(r) - 1$	$\sum_{ijkl} (10P)(\overline{y}_{ijkl\cdot\cdot} - \overline{y}_{\cdots})^2$
Plot	Dataset	blocks	$(2)(4)(5)(r) - 1$	$\sum_{ijkl} (10P)(\overline{y}_{ijkl\cdot\cdot} - \overline{y}_{\cdots})^2$
	Aes.	δ	9	$\sum_m (2)(4)(5)(P)(\overline{y}_{\cdots m\cdot} - \overline{y}_{\cdots})^2$
	Aes x K	$(\alpha\delta)$	9	$\sum_{im} (4)(5)(P)(\overline{y}_{i\cdots m\cdot} - \overline{y}_{i\cdots} - \overline{y}_{\cdots m\cdot} + \overline{y}_{\cdots})^2$
	Aes x σ_T	$(\beta\delta)$	27	$\sum_{jm} (2)(5)(P)(\overline{y}_{\cdot j\cdot m\cdot} - \overline{y}_{\cdot j\cdots} - \overline{y}_{\cdots m\cdot} + \overline{y}_{\cdots})^2$
	Aes x σ_C	$(\gamma\delta)$	36	$\sum_{km} (2)(4)(P)(\overline{y}_{\cdot\cdot k\cdot m\cdot} - \overline{y}_{\cdot\cdot k\cdots} - \overline{y}_{\cdots m\cdot} + \overline{y}_{\cdots})^2$
	Others		9(31)	
	Resid		40(rP-1)-(40r-1)	difference
	Total		$400r - 1$	$\sum_{ijklm} (P)(\overline{y}_{ijklm\cdot} - \overline{y}_{ijkl\cdot\cdot})^2$
Trial	Picture	Sub-blocks	$400r - 1$	$\sum_{ijklm} (P)(\overline{y}_{ijklm\cdot} - \overline{y}_{ijkl\cdot\cdot})^2$
	Participants	τ	$P - 1$	$\sum_n (2)(4)(5)(r)(10)(\overline{y}_{\cdots\cdots n} - \overline{y}_{\cdots})^2$
	Resid		$(400r - 1)(P)$	difference
	Total		$400(r)(P) - 1$	$\sum_{ijklmn} (y_{ijklmn} - \overline{y}_{\cdots})^2$

Table 3: Evaluation of sources of error in a full factorial version of the experiment, with r replicates of each parameter combination and P participant evaluations of each plot(data/aesthetic combination).

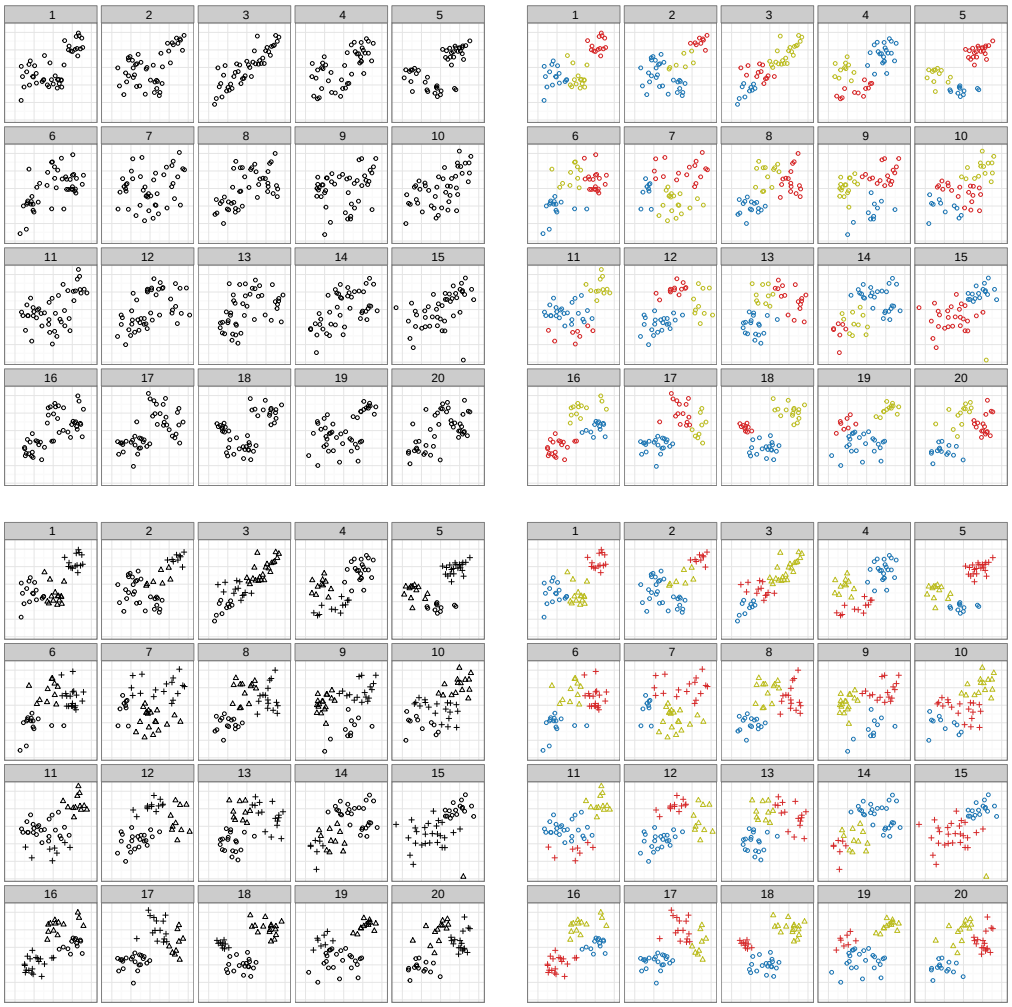
The fractional factorial option will be a pain to explain when we write things up; it will be simpler to explain using a subset of cases. Given that we don't particularly care about the third-order effects (and possibly not even the second-order effects) for the parameters, I'm inclined to say that the single-replicate option is the easiest way to go (and lets us keep the simple SSQ in the table, which is a huge bonus in my opinion). Even if we just use the third-order interaction effect as error, we still have 12 degrees of freedom; that should be plenty - we'd only need F=2.69 to get a significant result for even the $(\sigma_T\sigma_C)$ test.
HH: We might not care about interpreting the two-way interactions, but unfortunately they will be there (see comment at the back). So I would suggest to go with a full factorial design in σ_C, σ_T, and K, with three replications each (we need the replicates, also explained in the back). This gives us 18 parameter settings, and $18 \cdot 3 = 54$ data sets. In case you still want to consider the effect of the number of datapoints N, we could switch from fully factorial to fractional factorial and replace the three-way interaction of σ_C, σ_T, and K by the settings of N. That way we will keep the 18 settings.

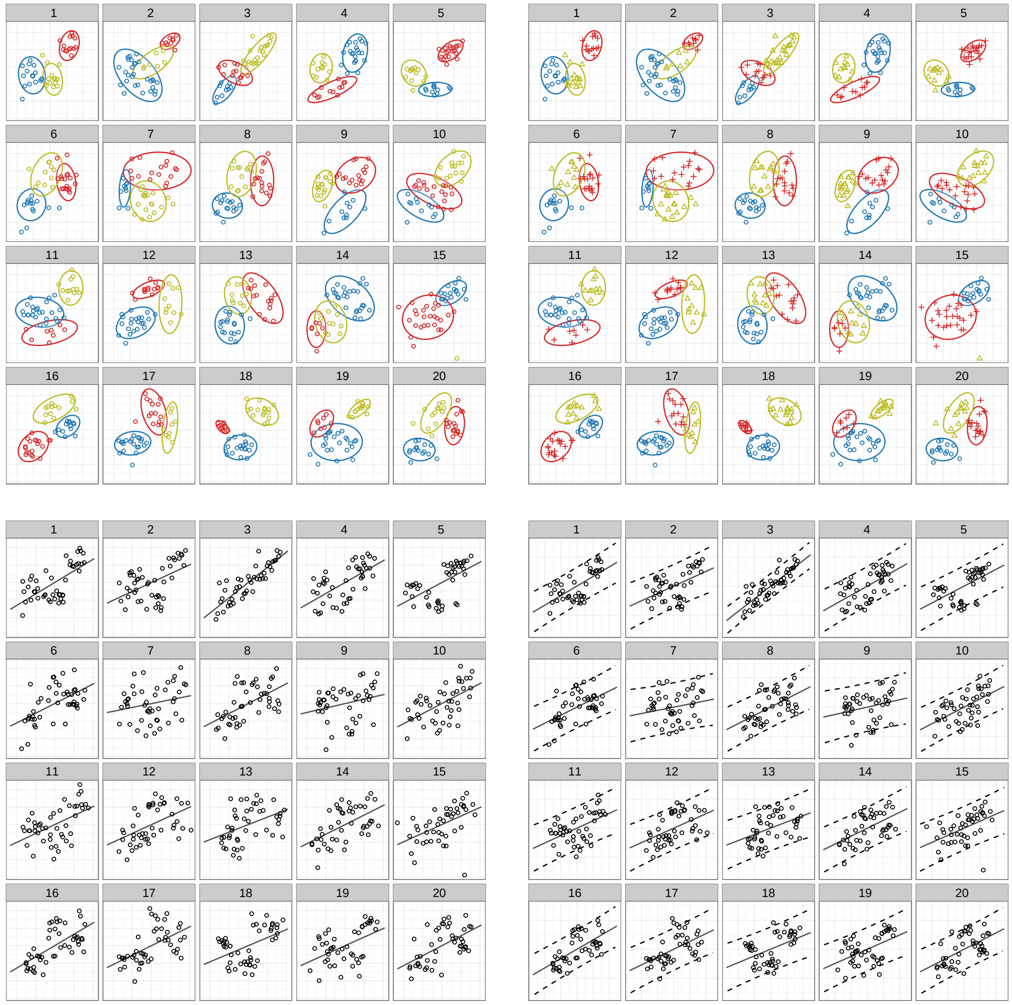
Table 4: ANOVA table - only one replicate. Evaluation of sources of error in a full factorial version of the experiment, with one replicate of each parameter combination and P participant evaluations of each plot(data/aesthetic combination).

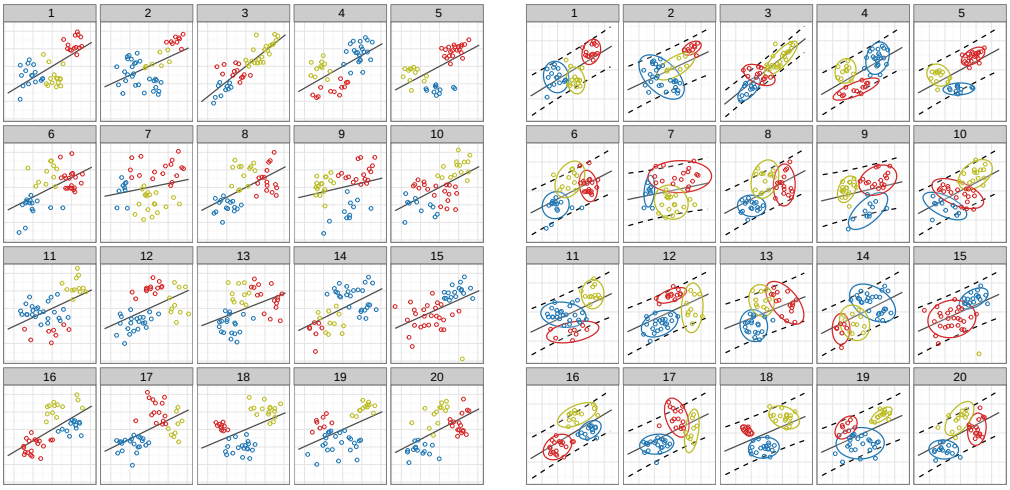
Level	Factor	Source	DF	Sum of Squares
Dataset	K	α	1	$\sum_i (4)(5)(10P)(\overline{y}_{i\ldots} - \overline{y}_{\ldots\ldots})^2$
	σ_T^2	β	3	$\sum_j (2)(5)(10P)(\overline{y}_{\cdot j\ldots} - \overline{y}_{\ldots\ldots})^2$
	σ_C^2	γ	4	$\sum_i (2)(4)(10P)(\overline{y}_{\cdot\cdot k\cdot\cdot} - \overline{y}_{\ldots\ldots})^2$
		Resid.	22	difference
	Total		39	$\sum_{ijk} (10P)(\overline{y}_{ijk\cdot\cdot} - \overline{y}_{\ldots\ldots})^2$
Plot	Dataset	blocks	39	$\sum_{ijk} (10P)(\overline{y}_{ijk\cdot\cdot} - \overline{y}_{\ldots\ldots})^2$
	Aes.	δ	9	$\sum_m (2)(4)(5)(P)(\overline{y}_{\ldots m\cdot} - \overline{y}_{\ldots\ldots})^2$
	Aes x K	$(\alpha\delta)$	9	$\sum_{im} (4)(5)(P)(\overline{y}_{i\cdot\cdot m\cdot} - \overline{y}_{i\ldots\cdot} - \overline{y}_{\ldots m\cdot} + \overline{y}_{\ldots\ldots})^2$
	Aes x σ_T	$(\beta\delta)$	27	$\sum_{jm} (2)(5)(P)(\overline{y}_{\cdot j\cdot m\cdot} - \overline{y}_{\cdot j\ldots} - \overline{y}_{\ldots m\cdot} + \overline{y}_{\ldots\ldots})^2$
	Aes x σ_C	$(\gamma\delta)$	36	$\sum_{km} (2)(4)(P)(\overline{y}_{\cdot\cdot km\cdot} - \overline{y}_{\cdot\cdot k\cdot\cdot} - \overline{y}_{\ldots m\cdot} + \overline{y}_{\ldots\ldots})^2$
	Resid		9(31)	difference
	Total		399	$\sum_{ijkm} (P)(\overline{y}_{ijkm\cdot} - \overline{y}_{ijk\cdot\cdot})^2$
Trial	Picture	Sub-blocks	399	$\sum_{ijkm} (P)(\overline{y}_{ijkm\cdot} - \overline{y}_{ijk\cdot\cdot})^2$
	Participants	τ	$P-1$	$\sum_n (2)(4)(5)(10)(\overline{y}_{\ldots\cdot n} - \overline{y}_{\ldots\ldots})^2$
	Resid		$399(P-1)$	difference
	Total		$400P-1$	$\sum_{ijkmn} (y_{ijkmn} - \overline{y}_{\ldots\cdot})^2$

4.4 Sample Pictures

The following plots use $\sigma_T = .4$ and $\sigma_C = .3$.







References

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E. K., Swayne, D. F., and Wickham, H. (2009), "Statistical inference for exploratory data analysis and model diagnostics," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367, 4361–4383.

Çağatay Demiralp, Bernstein, M., and Heer, J. (2014), "Learning Perceptual Kernels for Visualization Design," *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

Few, S. (2009), *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Burlingame, CA: Analytics Press, 1st ed.

Gegenfurtner, K. R. and Sharpe, L. T. (2001), *Color vision: From genes to perception*, Cambridge University Press.

Healey, C. G. and Enns, J. (2012), "Attention and Visual Memory in Visualization and Computer Graphics," *IEEE Transactions on Visualization and Computer Graphics*, 18, 1170–1188.

Hofmann, H., Follett, L., Majumder, M., and Cook, D. (2012), "Graphical Tests for Power Comparison of Competing Designs," *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis)*, 18, 2441–2448, 25% acceptance rate.

Majumder, M., Hofmann, H., and Cook, D. (2013), "Validation of Visual Statistical Inference, Applied to Linear Models," *Journal of the American Statistical Association*, 108, 942–956.

Robinson, H. (2003), "Usability of Scatter Plot Symbols," *ASA Statistical Computing & Graphics Newsletter*, 14, 9–14.

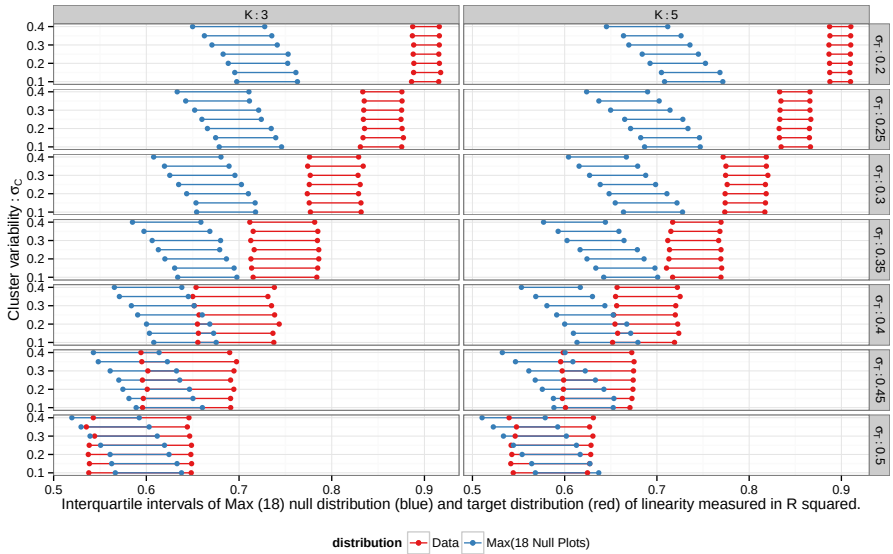
Tukey, J. W. (1977), *Exploratory Data Analysis*, Lebanon, IN: Addison Wesley.

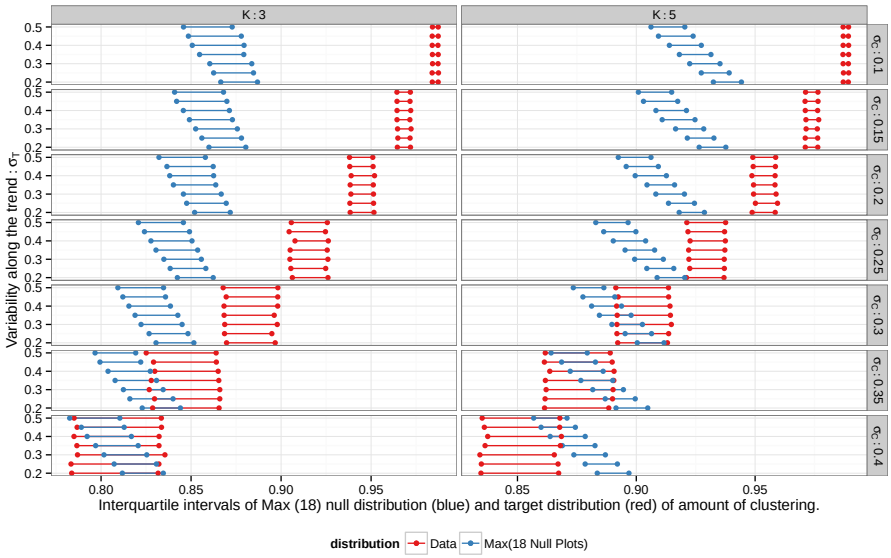
Wickham, H., Cook, D., Hofmann, H., and Buja, A. (2010), "Graphical inference for infovis," *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis)*, 16, 973–979, 26% acceptance rate. Best paper award.

Simulation Studies of Parameter Space

Using 1000 simulations for each of the 98 combinations of parameters ($K = \{3, 5\}$, $\sigma_C = \{.1, .15, .2, .25, .3, .35, .4\}$, $\sigma_T = \{.2, .25, .3, .35, .4, .45, .5\}$), we explored the effect of parameter value on the distribution of summary statistics describing the line strength (R^2) and cluster strength (short description here) for null and target plots. The plot below shows the 25th and 75th percentiles of the distribution of these summary statistics for each set of parameter values. These plots guided our evaluation of "easy", "medium" and "hard" parameter values for line and cluster tasks.

What we also see from these plots, is that we do have a $\sigma_C\sigma_T$ interaction: the distinction between target and null on a fixed setting of clustering becomes increasingly difficult as the standard deviation for the linear trend is increased, and vice versa. We might also have a three-way interaction between σ_C, σ_T, and K: the size of the blue intervals (bottom figure) changes in size between different levels of K, it changes for different levels of σ_C and σ_T. I am not sure whether that is an actual three-way interaction or just all three two-way interactions, but it doesn't matter, at this point we are just talking about potentially saving one parameter. What is clear, is that we need to block by parameter setting. We do so, by blocking on each dataset. Each dataset is non-deterministic, though, because we have a random process generating from different parameter settings, not a deterministic run setting as in an engineering setting. We therefore need repetitions of the data generation to be able to separate the variabiltiy coming from within the parameter setting from the additional variability introduced by the subjects' evaluations of the lineups.




K : 3
K : 5
σc : 0.1
σc : 0.15
σc : 0.2
σc : 0.25
σc : 0.3
σc : 0.35
σc : 0.4
Variability along the trend : σT
0.2
0.3
0.4
0.5
0.80
0.85
0.90
0.95
Interquartile intervals of Max (18) null distribution (blue) and target distribution (red) of amount of clustering.
distribution
Data
Max(18 Null Plots)