

Tasks to Tease Apart Scatterplot Design Decisions

Alper Sarikaya

University of Wisconsin-Madison
sarikaya@cs.wisc.edu

 @yelperalp



Michael Gleicher

University of Wisconsin-Madison
gleicher@cs.wisc.edu

Abstract

Scatterplots are among the most common methods for exploring and presenting data, covering a wide range of tasks and designs. The variety of scatterplot designs has created a proliferation of potential design decisions to consider when constructing a scatterplot. However, there remain many unexamined assumptions in respect to the trade-offs between these decisions. Here, we begin the process of synthesizing recent work to build descriptive knowledge of how design decisions affect the analysis tasks viewers perform with scatterplots. Through deriving twelve abstracted scatterplot tasks, we can start to tease apart the different affordances of design decisions, and begin to formulate a basis for prescriptive scatterplot design.

Procedure

Building off of Schulz, *et al.*'s [1] concretization of the relation between data, task, and visualization design, we strive to enumerate how this combination of factors affect how scatterplots are designed through a bottom-up approach:

Data + Task = Visualization

To capture tasks, we survey scatterplot papers in the visualization research literature. With four visualization researchers (med. 6 years experience), we performed a card sort to arrive at a list of 12 tasks.

With these tasks, we can evaluate how instantiations of scatterplots support these tasks, and start to match design decisions to their affordances and evaluate their trade-offs, dependent on the characteristics of the visualized data.

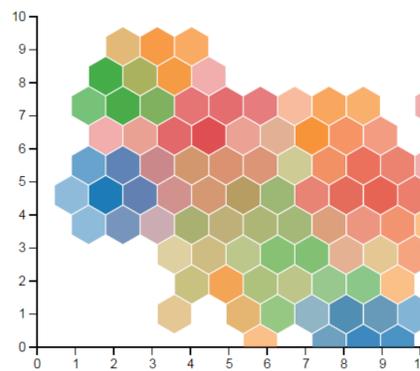
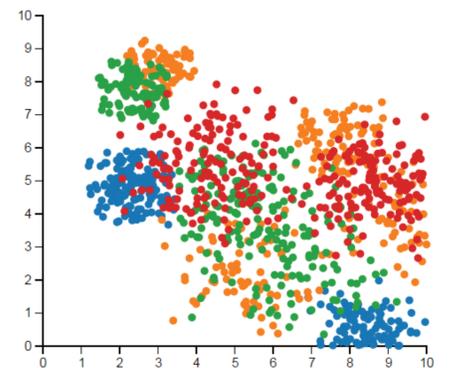
Tasks

#	Task	Description
1	Identify object	Identify the referent from the representation
2	Locate object	Find a particular object in its new spatialization
3	Verify object	Reconcile attribute of an object with its spatialization (or other visual encoding)
4	Search for known motif	Find a particular known pattern (cluster, correlation)
5	Browse data	Look for things that look unusual, global trends
6	Identify outliers	Find an object that doesn't match the 'modal' distribution
7	Characterize distribution	Do objects cluster? Part of a manifold? Range of values?
8	Identify correlation	Determine level of correlation
9	Explore neighborhood	Explore the properties of objects in a neighborhood
10	Numerosity comparison	Compare the numerosity/density in different regions of the graph
11	Object comparison	Do objects have similar attributes? Are these objects similar in some way?
12	Understand distances	Understanding a given spatialization (e.g. relative distances)

Table 1: The collected list of tasks performed with scatterplots.

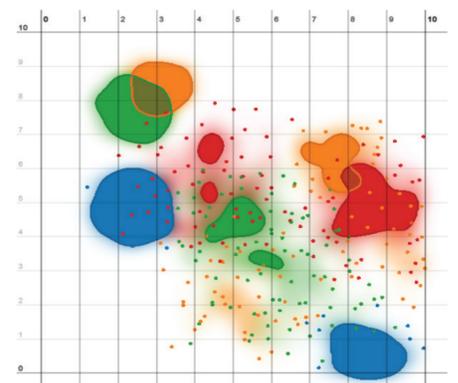
These tasks allow for teasing apart the affordances of design decisions, and provide rationale for the trade-offs between these decisions.

Can **identify** objects
Overdraw masks
distribution



Cannot **identify** objects
Binning identifies
maximum **distribution**

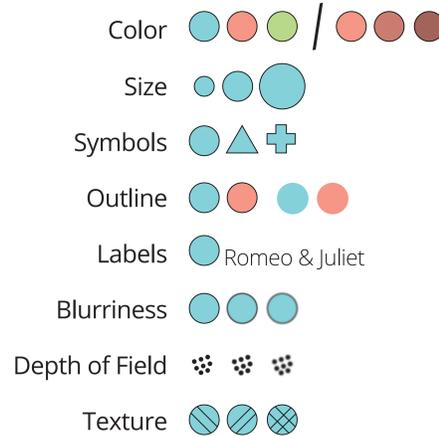
Can **identify** just outliers
Clustered points
highlight **distribution**



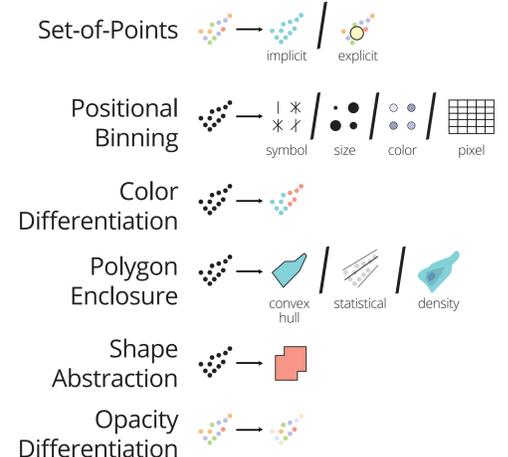
Design Decisions

We evaluated visualization papers within the last six years that contained the word *scatter* in their abstract or title. For each paper, we collected the tasks supported by the scatterplot along with the design strategies. With this collection, we can start to identify recurring design patterns for scatterplots and link their efficacy in supporting the list of scatterplot tasks, as well as identify missing design patterns for under-supported tasks and data.

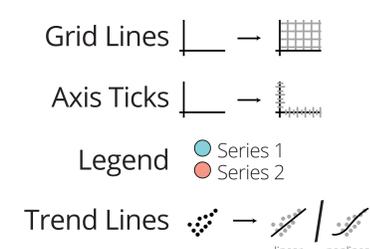
Point Encodings



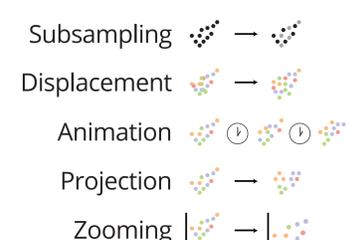
Point Groupings



Graph Amenities



Point Position



[1] H. Schulz, *et al.* "A Design Space of Visualization Tasks." IEEE TVCG (InfoVis 2013), 19(12), 2366-2375, 2013.

Supplemental material for this work is available: <http://graphics.cs.wisc.edu/Papers/2016/SG16/>. This work is supported by NSF award IIS-1162037.