



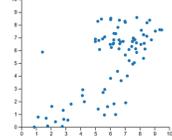
SCATTERPLOTS: TASKS DATA AND DESIGN

A. Sarkisova and M. Chen
IEEE Transaction on Visualization and Computer Graphics

Presented By
Shareen Mahmood

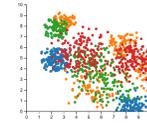
WHAT IS A TRADITIONAL SCATTERPLOT?

- Encodes two quantitative variables using the vertical and horizontal spatial position channels
- Each object in a dataset is represented with a point (mark)
- Effective in providing overviews, finding outliers, and judging correlation

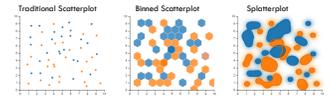


DOES IT FAIL?

- Yell As data grows in scale, traditional scatterplots can become ineffective
- Overdraw is a concern where points overlap one another and masks points drawn under them.



DIFFERENT DESIGNS SOLUTIONS



Designers have little guidance in how to select among choices. Which design to choose?

GOAL OF THE PAPER

- Help designers select scatterplot designs that are appropriate to their scenarios
- Identify factors that affect the appropriateness of scatterplot designs
- Create a framework based on the analysis goal and data characteristics

FACTORS THAT AFFECT THE DESIGN OF SCATTERPLOTS

- Analysis Tasks: What do viewers do with a scatterplot?
- Data Characteristics: How do they prompt changes in design?
- Design Decisions: What design variables need to be constructed?

ANALYSIS TASKS

- Gathered 23 model tasks from various vis literature to capture what viewers do with scatterplots
- Four data visualization experts performed an open card sort where tasks were grouped together based on their similarity
- Refined the categories post hoc to generate a complete picture of the task space

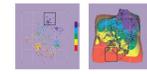


Fig 2. Example task from experiment. Top: Task as it should be presented and used by the participant. Green arrows are highlights on task labels.

Task: Which sector of the graph has the most spots of [blue] color?

M. Teny et al. Scatterplots Design: Comparing points and trajectories. IEEE Transactions on Visualization and Computer Graphics, 1998. 1000-1007

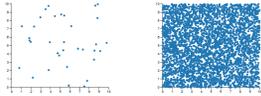
ANALYSIS TASKS

- A final list of 12 tasks split into 3 categories
 - Object Centric
 - Browsing
 - Aggregate Level
- A combination of these tasks can be used as building blocks to achieve an analysis goal

Task	Description
1 Identify object	Identify the selected line for the plot
2 Locate object	Find a particular object in an area
3 Verify object	Verify a particular object in an area with the legend
4 Compare objects	Do objects have similar attributes? Are these objects similar to some other?
5 Explore neighborhood	Explore the properties of objects in a neighborhood
6 Search for known event	Look for a particular known event (known coordinates)
7 Explore data	Look to figure out how overall the data looks
8 Characterize distribution	Describe the distribution of objects
9 Identify correlation	Does the data show any correlation?
10 Identify correlation	Compare the consistency of change in different regions of the graph
11 Unkindsal distribution	Understand a group of data points (e.g., outliers, clusters)

DATA CHARACTERISTICS

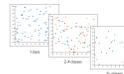
Data characteristics can influence the design of an appropriate scatterplot



https://halper.dawn.us/assets/publications/scatterplots/scatterplot_talk.pdf

DATA CHARACTERISTICS

List of design affecting data characteristics collected from the literature



Data Attribute	Possible Values	Relevant Work
Class label	No class label, 2-4 classes, No classes	Elliott and Rawson [2015], Giammarco et al. [2016], Nye et al. [2016]
Num. of points	Small (< 10), medium (10-100), large (100-1000), very large (> 1000)	Cohen et al. [2011], Chakrabarti et al. [2015], Korte et al. [2015], Mery and Gallette [2013], Teny et al. [2017]
Num. of dimensions	Two continuous, two discrete, one < 12 dimensions	Boat et al. [2006], Chen et al. [2010], Sedore et al. [2012]
Spatial nature	Dimensionality reduction map to spatial position	McDuffin [2005], Muralidhar et al. [2010]
Data distribution	Random, linear correlation, weakly correlated, outliers, multimodal, clusters	Bertoli et al. [2011], Li et al. [2009], Rensink and Baheti [2010], Sedore et al. [2012], Sips et al. [2009], Teny et al. [2017], Teny and Wilhelmsen [2016], Wilhelmsen et al. [2008]

DESIGN DECISION

Identified design decisions by applying a keyword ("scatter") search methodology on 3040 vis papers.

- Clustered the design choices into 4 groups

Point Encoding (Example: Color)

Point Grouping (Example: Binning)

Point Position (Example: Animation)

Graph Aesthetics (Example: Annotations)

Category	Design Choice	Example
Point Encoding	Color	Color, Size, Shape
	Shape	Circle, Square, Triangle
	Size	Small, Medium, Large
	Opacity	Transparent, Opaque
Point Grouping	Aggregation Type	Bin, Cluster, Outlier
	Point Position	Fixed, Animated
	Interaction	Click, Hover, Drag
	Annotation	Label, Legend, Tooltip
Graph Aesthetics	Color	Color, Size, Shape
	Shape	Circle, Square, Triangle
	Size	Small, Medium, Large
	Opacity	Transparent, Opaque

https://halper.dawn.us/assets/publications/scatterplots/scatterplot_talk.pdf

DESIGN SPACE TO EVALUATE APPROPRIATENESS OF DESIGN STRATEGIES

Cross product of these three is huge! Leads to over 4300 discrete scatterplot scenarios

Task	Description	Class	Design Choice	Example
1 Identify object	Identify the selected line for the plot	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
2 Locate object	Find a particular object in an area	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
3 Verify object	Verify a particular object in an area with the legend	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
4 Compare objects	Do objects have similar attributes? Are these objects similar to some other?	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
5 Explore neighborhood	Explore the properties of objects in a neighborhood	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
6 Search for known event	Look for a particular known event (known coordinates)	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
7 Explore data	Look to figure out how overall the data looks	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
8 Characterize distribution	Describe the distribution of objects	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
9 Identify correlation	Does the data show any correlation?	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
10 Identify correlation	Compare the consistency of change in different regions of the graph	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape
11 Unkindsal distribution	Understand a group of data points (e.g., outliers, clusters)	No class label, 2-4 classes, No classes	Color, Size, Shape	Color, Size, Shape

https://halper.dawn.us/assets/publications/scatterplots/scatterplot_talk.pdf

A SLICE OF THE SPACE: TASK & DESIGN STRATEGIES

Framework illustrated with a 2D slice of the entire grid (60 out of 4300 grids)

- Entire set of tasks and design strategies
- Data characteristics fixed to "large" number of points and classes with an unstructured distribution of data



Task	Point Encoding	A	B	C	D	E
1 Identify object	✓	✓	✓	✓	✓	✓
2 Locate object	✓	✓	✓	✓	✓	✓
3 Verify object	✓	✓	✓	✓	✓	✓
4 Compare objects	✓	✓	✓	✓	✓	✓
5 Explore neighborhood	✓	✓	✓	✓	✓	✓
6 Search for event	✓	✓	✓	✓	✓	✓
7 Explore data	✓	✓	✓	✓	✓	✓
8 Characterize distribution	✓	✓	✓	✓	✓	✓
9 Identify correlation	✓	✓	✓	✓	✓	✓
10 Identify correlation	✓	✓	✓	✓	✓	✓
11 Characterize anomaly	✓	✓	✓	✓	✓	✓
12 Characterize distance	✓	✓	✓	✓	✓	✓

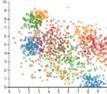
✓ general support
 ✓ support in particular situations
 ✓ requires concurrent support from other knowledge
 # no improvement to task support

13

https://halper.dawn.us/assets/publications/scatterplots/scatterplot_talk.pdf

USING THE FRAMEWORK

- Difficult to support aggregate level tasks such as identifying anomalies, correlations and object density with point encoding and position (9A-11B)



Task	Point Encoding	A	B	C	D	E
1 Identify object	✓	✓	✓	✓	✓	✓
2 Locate object	✓	✓	✓	✓	✓	✓
3 Verify object	✓	✓	✓	✓	✓	✓
4 Compare objects	✓	✓	✓	✓	✓	✓
5 Explore neighborhood	✓	✓	✓	✓	✓	✓
6 Search for event	✓	✓	✓	✓	✓	✓
7 Explore data	✓	✓	✓	✓	✓	✓
8 Characterize distribution	✓	✓	✓	✓	✓	✓
9 Identify correlation	✓	✓	✓	✓	✓	✓
10 Identify correlation	✓	✓	✓	✓	✓	✓
11 Characterize anomaly	✓	✓	✓	✓	✓	✓
12 Characterize distance	✓	✓	✓	✓	✓	✓

✓ general support
 ✓ support in particular situations
 ✓ requires concurrent support from other knowledge
 # no improvement to task support

14

https://halper.dawn.us/assets/publications/scatterplots/scatterplot_talk.pdf

USING THE FRAMEWORK

- Point grouping hurts object-centric tasks (1C-4C, 9C, 11C)
- However, by composing point encoding, point position and interaction intent, object-centric tasks can be supported.

Task	Point Encoding	A	B	C	D	E
1 Identify object	✓	✓	✓	✓	✓	✓
2 Locate object	✓	✓	✓	✓	✓	✓
3 Verify object	✓	✓	✓	✓	✓	✓
4 Compare objects	✓	✓	✓	✓	✓	✓
5 Explore neighborhood	✓	✓	✓	✓	✓	✓
6 Search for event	✓	✓	✓	✓	✓	✓
7 Explore data	✓	✓	✓	✓	✓	✓
8 Characterize distribution	✓	✓	✓	✓	✓	✓
9 Identify correlation	✓	✓	✓	✓	✓	✓
10 Identify correlation	✓	✓	✓	✓	✓	✓
11 Characterize anomaly	✓	✓	✓	✓	✓	✓
12 Characterize distance	✓	✓	✓	✓	✓	✓

✓ general support
 ✓ support in particular situations
 ✓ requires concurrent support from other knowledge
 # no improvement to task support

15

WHAT-WHY-HOW ANALYSIS

Item	Scatterplots (Framework)
What: Data	Vis literature, papers
What: Derived	Table with Tasks, Data characteristics, Design choices
Why: Tasks	Compare design strategies
How: Encode	Multidimensional table, Color highlighting, marks to denote appropriateness of design decisions
How: Reduce	Dimensionality Reduction/Slicing
Scale	4300 scatterplot scenarios

16

STRENGTH AND LIMITATIONS

• Strengths

- First to identify scenarios specific to scatterplot design
- Provides scope to discover potential areas for future innovation in scatterplot design
- Provides a good reference point for designers to get started with scatterplot design

• Limitation

- Infeasible to present the high dimensional grid. Data characteristics were restricted
- Focuses on single scatterplot design. Multi scatterplot tasks were discarded
- Misses the evaluation component in the study. How useful did designers find this framework to be?

17

REFERENCES

Paper: <https://elgar-datavis/assets/publications/scatterplots/scatterplots-preprint.pdf>

Slides: <https://elgar-datavis/assets/publications/scatterplots/scatterplot-talk.pdf>

Project Page: <http://graphics.cs.wisc.edu/Vis/scattertools/>

18