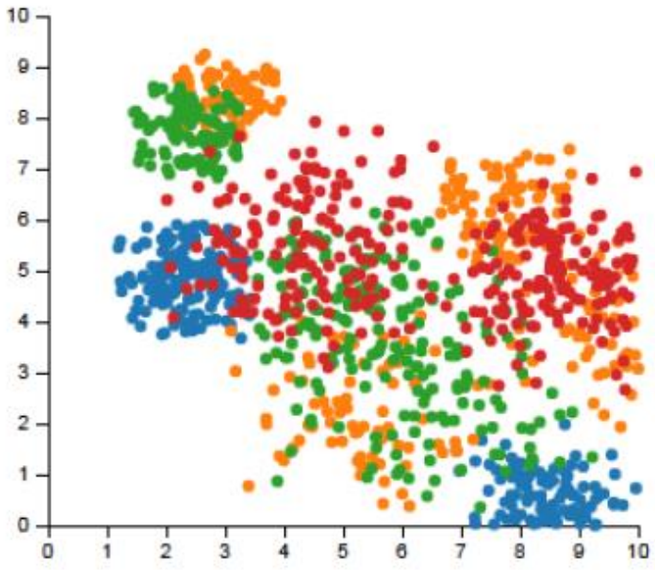
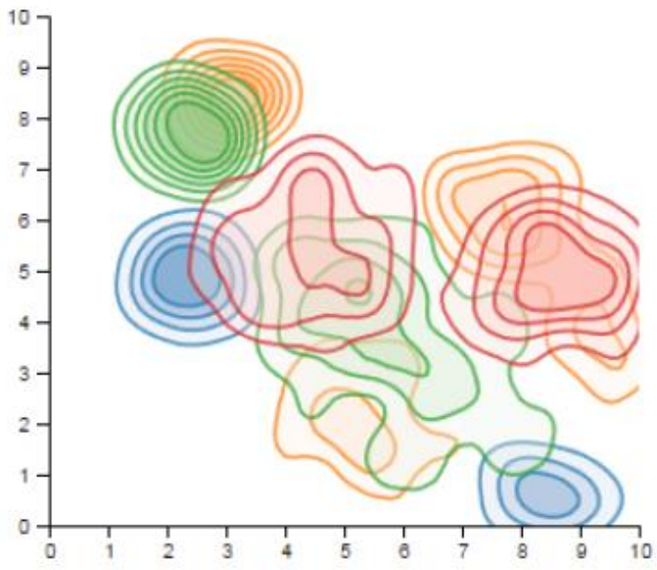
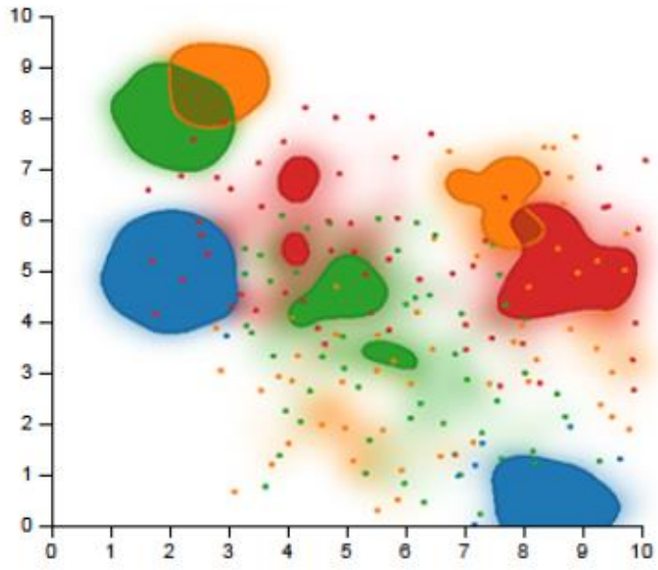
# Scatterplots:
## Tasks, Data, and Designs

Alper Sarikaya - Michael Gleicher

Hailey Báez
CIS6930 - Data Visualization
University of South Florida

# Problem Definition

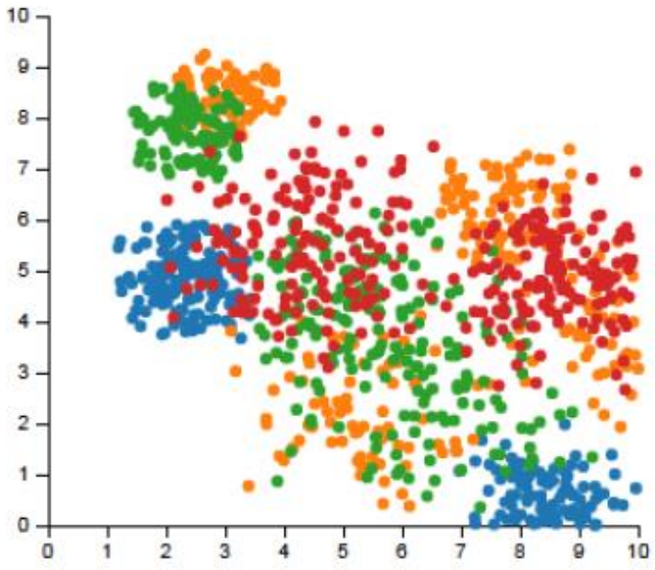
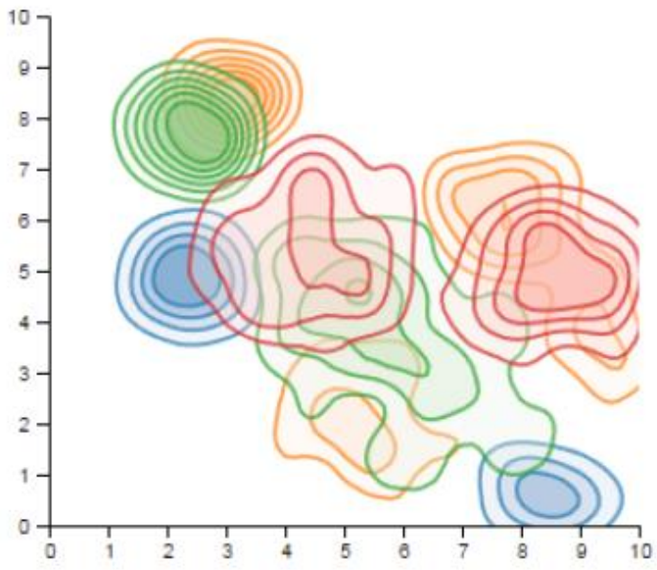
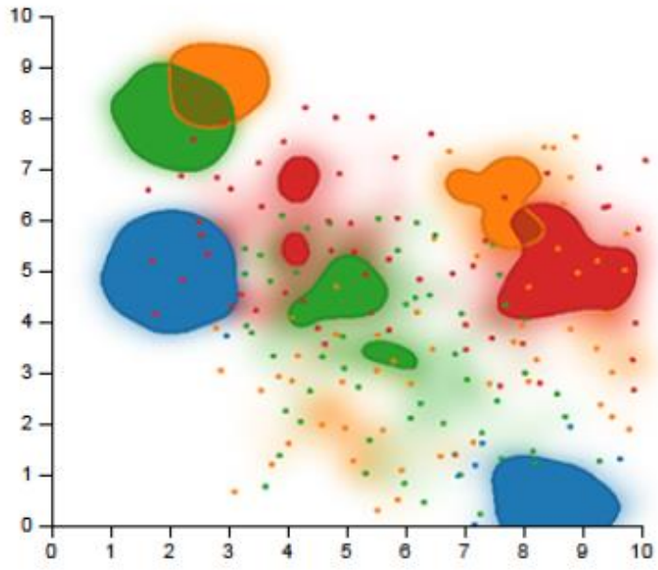
- Scatterplots are a very popular form of visualization designed to emphasize spatial distribution of data in two dimensions.
- Scatterplots can rapidly  fail as data grows in  scale and complexity.
- Designs presented to address this complexity are specific to the data characteristics and task of that plot.

# Problem Definition

- There is little guidance in how to select among different scatterplot design choices.

**“The goal is to design a framework that helps designers select scatterplot designs appropriate for their scenario by identifying the factors that affect the appropriateness of scatterplot designs”**

# Abstractions

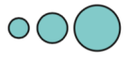
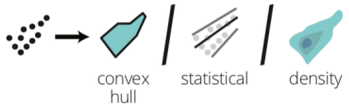
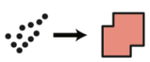
**Scatterplot Tasks:** Specific task analysis made for scatterplots compiled from a variety of sources of data visualization literature.

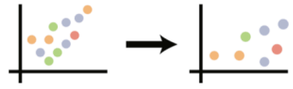
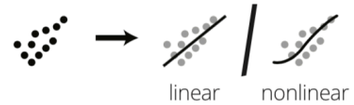
**Data Characteristics:** Which characteristics influence the design of a scatterplot?

| Data Attribute | Possible Values | Relevant Work |
|---|---|---|
| Class label | No class label, 2-4 classes, 5+ classes | Elliott and Rensink [2015], Gramazio et al. [2014], Sips et al. [2009] |
| Num. of points | Small (<10), medium (10–100), large (100–1000), very large (>1000) | Cottam et al. [2013], Gleicher et al. [2013], Keim et al. [2010], Mayorga and Gleicher [2013], Tory et al. [2007] |
| Num. of dimensions | Two continuous, two derived, or >2 dimensions | Best et al. [2006], Chan et al. [2010], Sedlmair et al. [2013] |
| Spatial nature | Dimensions do/do not map to spatial position | MacEachren [1995], Montello et al. [2003] |
| Data distribution | Random, linear correlation, overlap, manifolds, clusters | Bertini et al. [2011], Li et al. [2008], Rensink and Baldridge [2010], Sedlmair et al. [2013], Sips et al. [2009], Tatu et al. [2010], Dang and Wilkinson [2014], Wilkinson et al. [2005] |

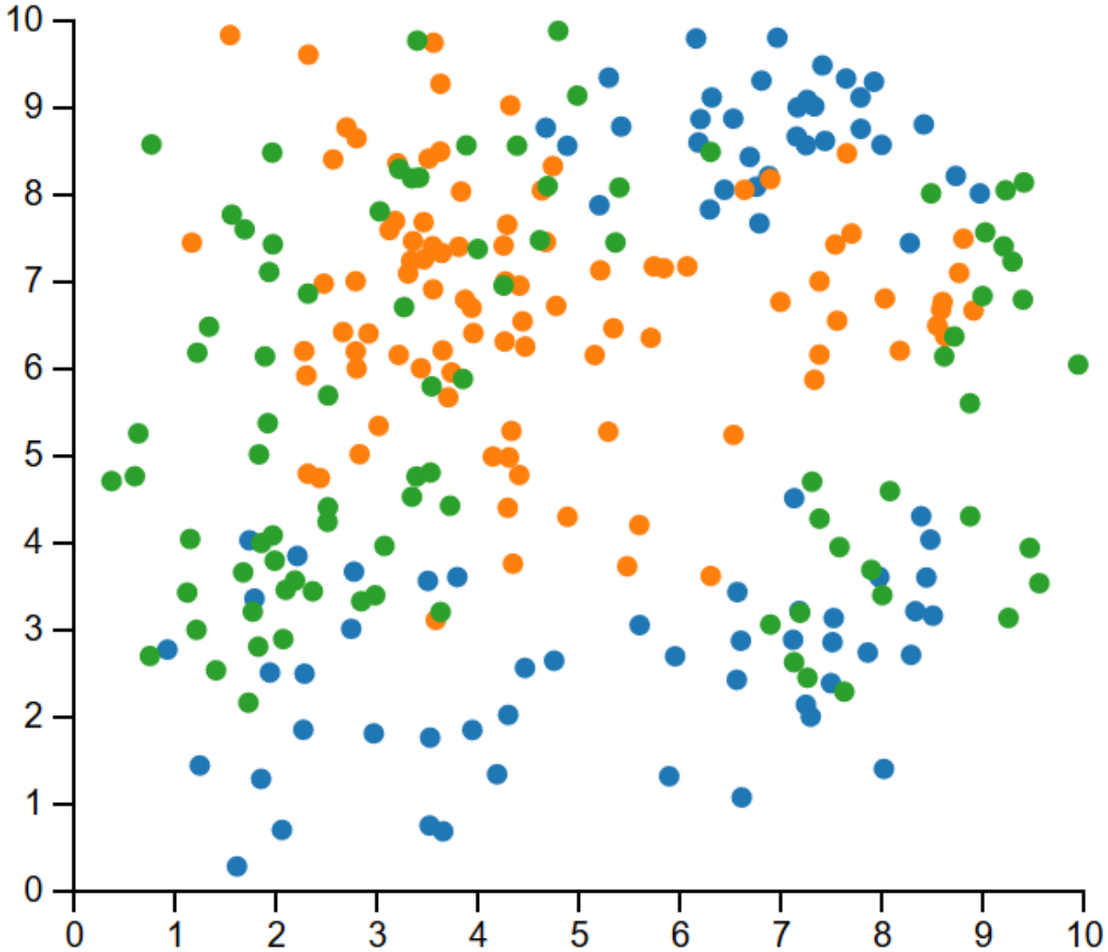
| | # Task | Description |
|---|---|---|
| object-centric | 1 Identify object | Identify the referent from the representation |
| | 2 Locate object | Find a particular object in its new spatialization |
| | 3 Verify object | Reconcile attribute of an object with its spatialization (or other encoding)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

# Design Decisions

| Cluster | Design Choice | Example |
|---|---|---|
| **Point Encoding** | Color |  |
| | Size |  |
| | Symbols |  |
| | Outline |  |
| | Opacity |  |
| | Texture |  |
| | Depth of Field |  |
| | Blurriness |  |
| **Point Grouping** | Representation Type |  |
| | Positional Binning |  |
| | Polygon Enclosure |  |
| | Shape Abstraction |  |

| Cluster | Design Choice | Example |
|---|---|---|
| **Point Position** | Subsampling |  |
| | Displacement |  |
| | Animation |  |
| | Projection |  |
| | Zooming |  |
| **Graph Amenities** | Grid Lines |  |
| | Axis Ticks |  |
| | Legend |  |
| | Trend Lines |  |
| | Annotations |  |

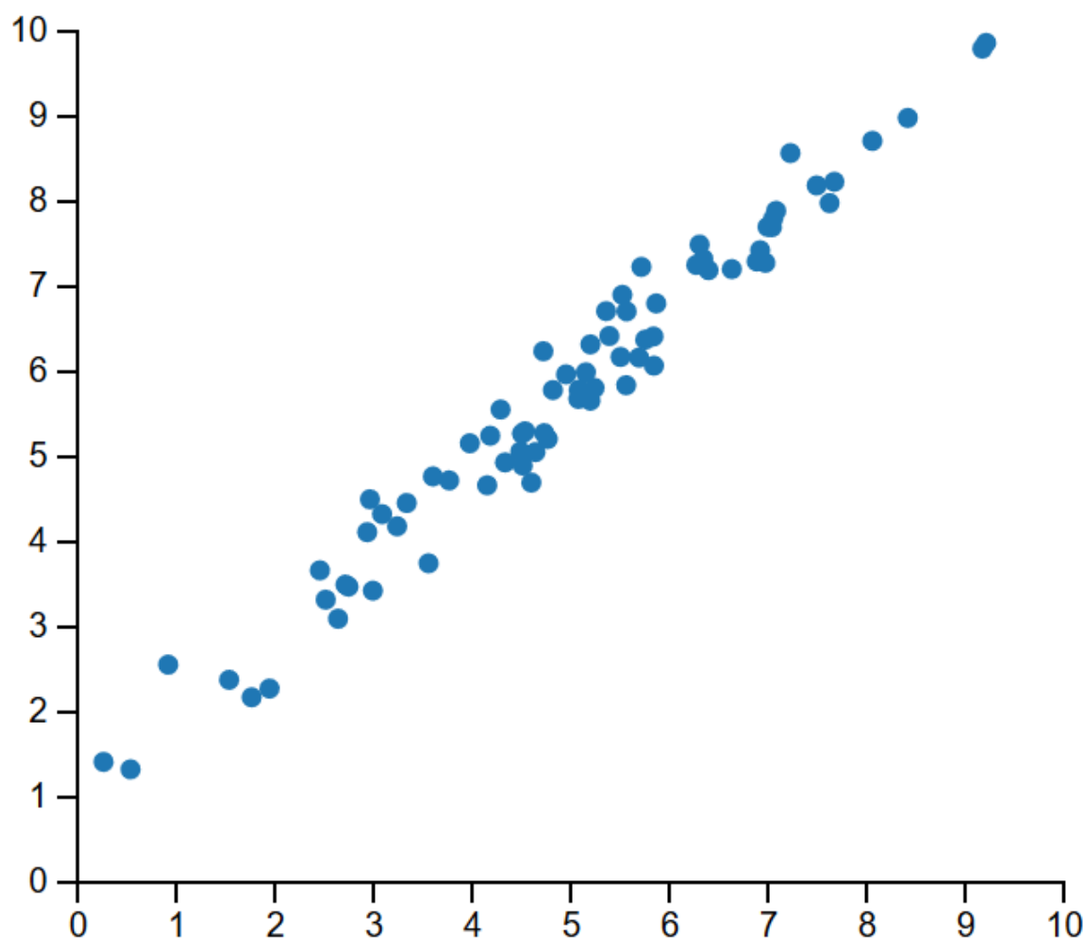
# Results




Identify anomalies ✔*

Identify object ✘

Characterize numerosity ✔




Identify anomalies ✔

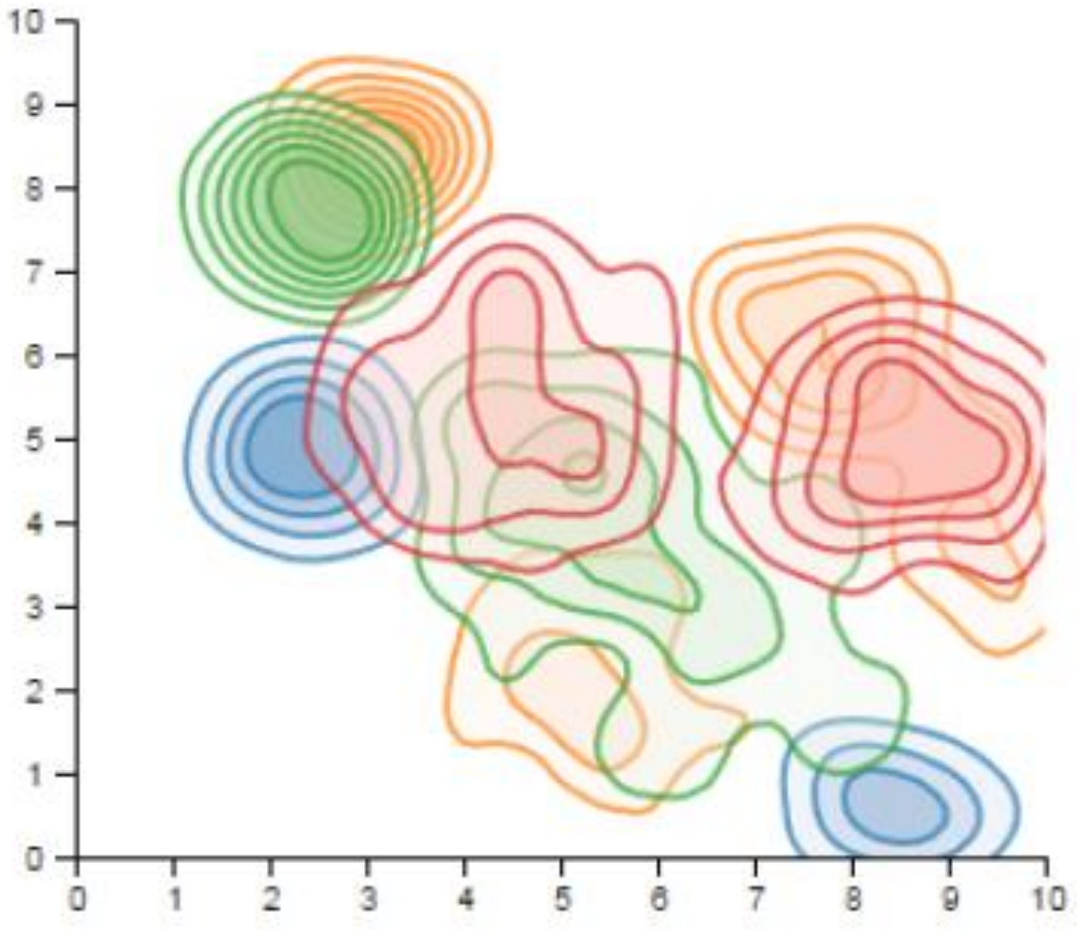
Identify object ✔

Characterize numerosity ✔

# Results (60 cells out of 4300)

| Task | A Point encoding | B Point position | C Point grouping | D Interaction intent | E Graph amenities |
|---|---|---|---|---|---|
| 1 Identify object | ✔ | ✔ | ✧ | ✔ | ✔* |
| 2 Locate object | ✔ | ✧ | ✧ | ✔ | ✔ |
| 3 Verify object | ✔ | ✔* | ✧ | ✔ | ✔ |
| 4 Compare objects | ✔ | ✔ | ✧ | ✔ | ✔ |
| 5 Explore neighborhood | ✔ | ✔ | ✔ | ✔ | ✔ |
| 6 Search for motif | ✔ | ✔ | ✔ | ✔ | ✔* |
| 7 Explore data | ✔ | ✔ | ✔ | ✔ | ✔ |
| 8 Characterize distribution | ✔ | ✔ | ✔ | ✧ | ✔ |
| 9 Find anomalies | ✧ | ✔* | ✧ | ✔* | ✔ |
| 10 Identify correlation | ✘ | ✘ | ✔ | ✘ | ✔ |
| 11 Characterize numerosity | ✘ | ✘ | ✔ | ✘ | ✘ |
| 12 Characterize distances | ✔* | ✔ | ✔* | ✔* | ✔ |

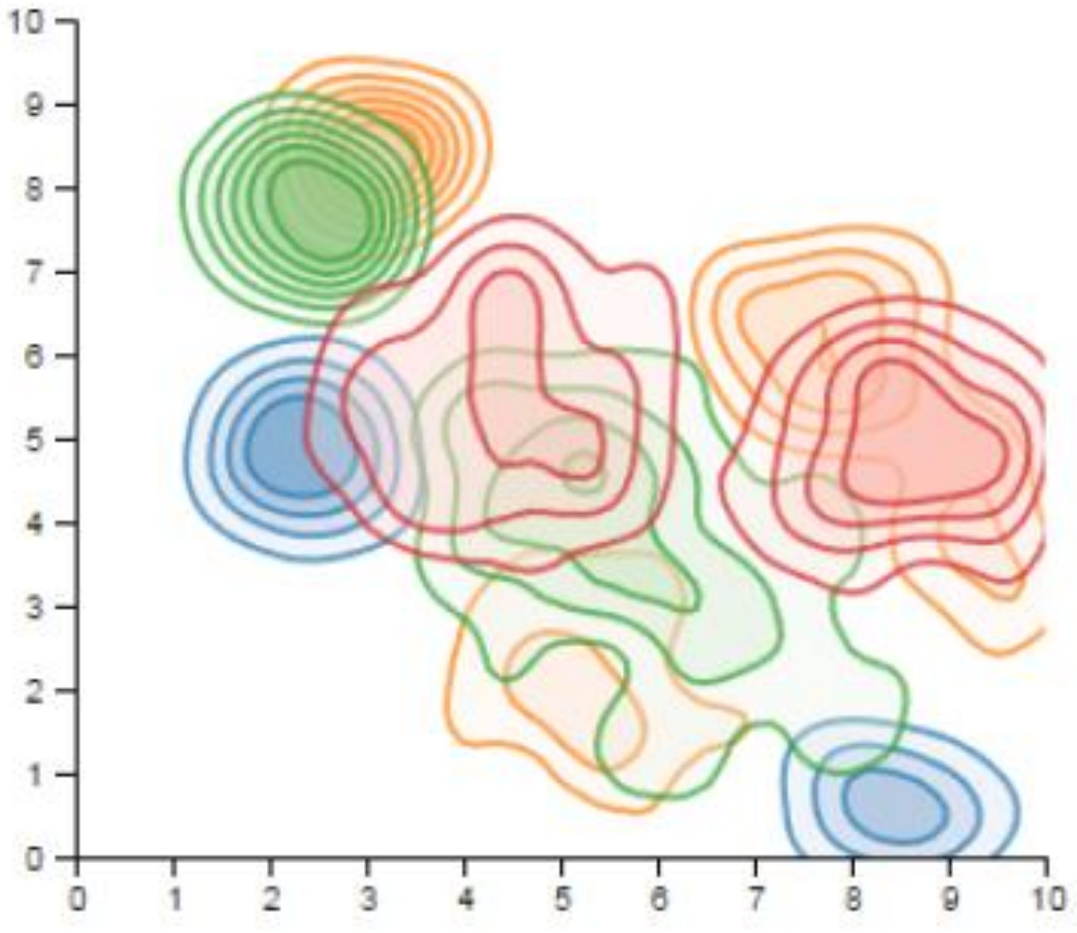
✔ general support
✔* support in particular situations
✧ requires concurrent support from other encodings
✘ no improvement to task support



# Results (60 cells out of 4300)

| Task | A Point encoding | B Point position | C Point grouping | D Interaction intent | E Graph amenities |
|---|---|---|---|---|---|
| 1 Identify object | ✔ | ✔ | ✧ | ✔ | ✔* |
| 2 Locate object | ✔ | ✧ | ✧ | ✔ | ✔ |
| 3 Verify object | ✔ | ✔* | ✧ | ✔ | ✔ |
| 4 Compare objects | ✔ | ✔ | ✧ | ✔ | ✔ |
| 5 Explore neighborhood | ✔ | ✔ | ✔ | ✔ | ✔ |
| 6 Search for motif | ✔ | ✔ | ✔ | ✔ | ✔* |
| 7 Explore data | ✔ | ✔ | ✔ | ✔ | ✔ |
| 8 Characterize distribution | ✔ | ✔ | ✔ | ✧ | ✔ |
| 9 Find anomalies | ✧ | ✔* | ✧ | ✔* | ✔ |
| 10 Identify correlation | ✘ | ✘ | ✔ | ✘ | ✔ |
| 11 Characterize numerosity | ✘ | ✘ | ✔ | ✘ | ✘ |
| 12 Characterize distances | ✔* | ✔ | ✔* | ✔* | ✔ |

✔ general support
✔* support in particular situations
✧ requires concurrent support from other encodings
✘ no improvement to task support



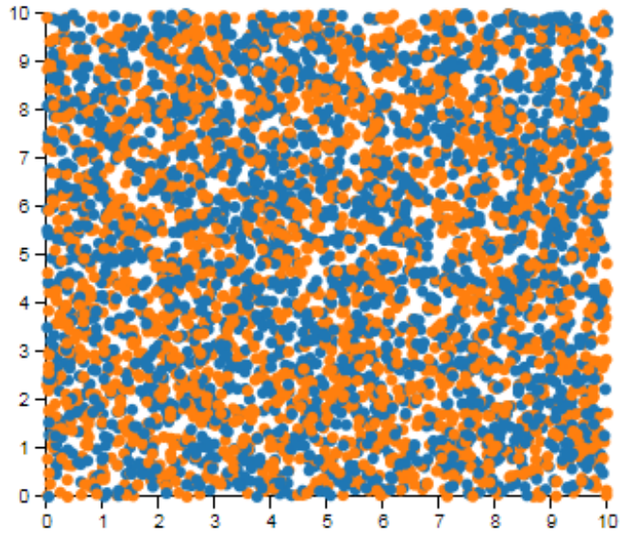
# Results

First data attribute: Class label

First value: 2-4 classes

Second data attribute: Number of objects

Second value: Very large (>1000)




Currently viewing task appropriateness for data characteristics
Class label (2-4 classes) and
Number of objects (Very large (>1000))

Edit Rankings | View Rankings

| # | Task Name | Ranking | Rationale | Strategies |
|---|---|---|---|---|
| 1 | Identify object | Need support | Difficult to identify an object with so many points (though class separation can help identification); aggregation can hurt this task | Provide explicit support to "see what's here", consider jittering/subsampling over aggregation |
| 2 | Locate object | Need support | Difficult to locate an object with so many points (though class separation can help) | Provide explicit support for filtering, selecting an object's representation |
| 3 | Verify object | Doable | Reconciling spatialization of a point is relatively easy, once point is identified (though potentially noisy) | Provide support for emphasizing the object of interest (and any other important landmarks) |
| 4 | Object comparison | Need support | Very difficult to do detailed comparison between objects with so many objects, though categorizing the points help to separate out point spatialization | Filter data to concentrate on those points relevant to analysis (see Shneiderman2015), provide support for selecting class of interest |

# Scatterplots:
## Tasks, Data, and Designs

Alper Sarikaya - Michael Gleicher

Hailey Báez
CIS6930 - Data Visualization
University of South Florida