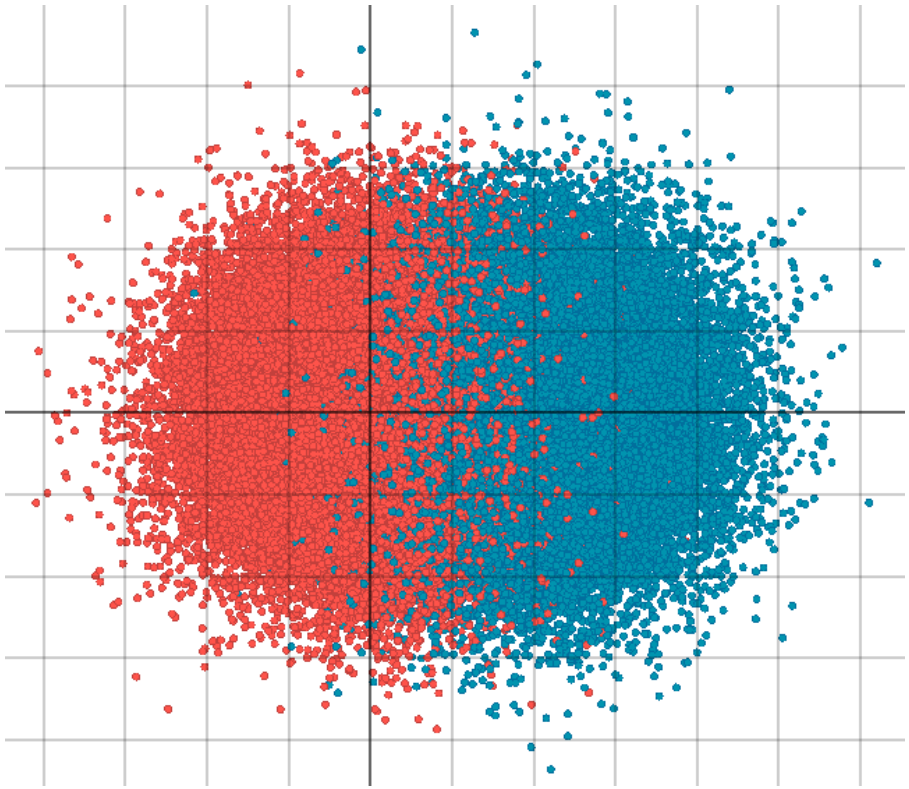
Warning:
This presentation uses the Bariol font, so if you get the PowerPoint, it may look weird. (the PDF embeds the font)
This presentation uses animation and embedded video, so the PDF may seem weird.

Splatterplots:
Overcoming Overdraw in Scatter Plots

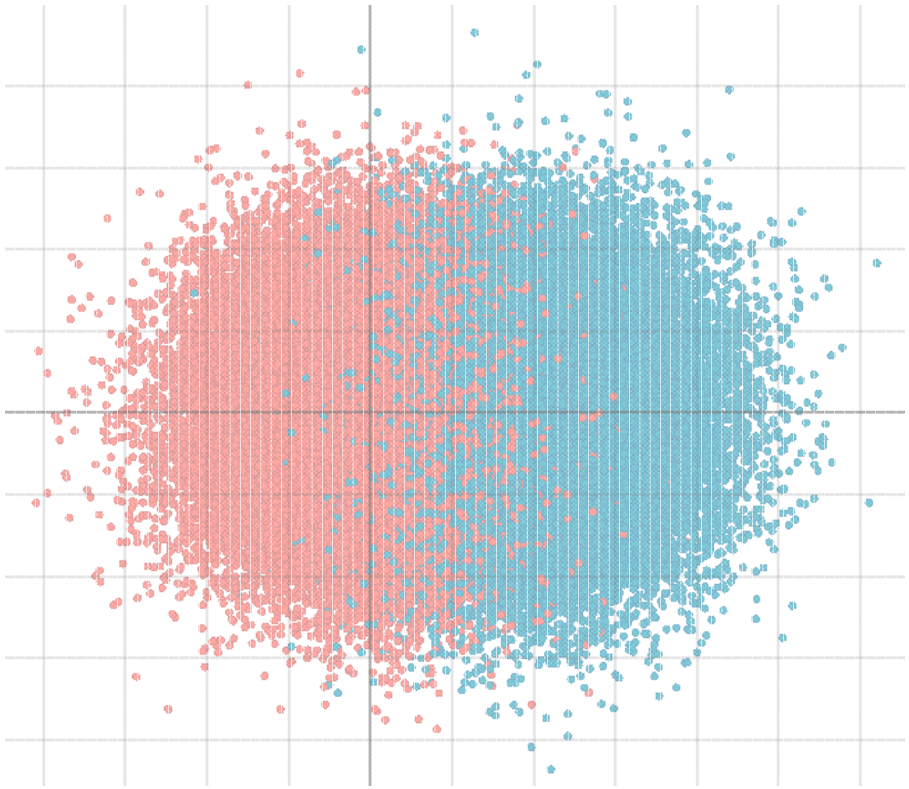
Adrian Mayorga
University of Wisconsin – Madison
(now at Epic Systems)

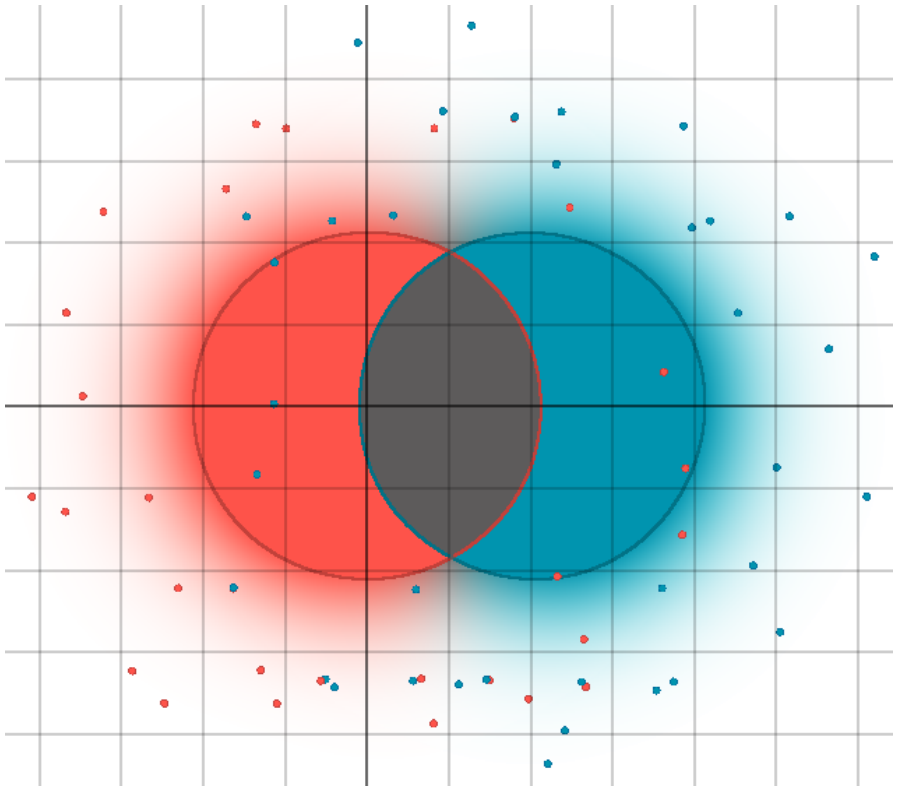
Michael Gleicher
University of Wisconsin – Madison
(on sabbatical at INRIA, Rhone-Alpes)

Problem: Scatterplot with too many points!



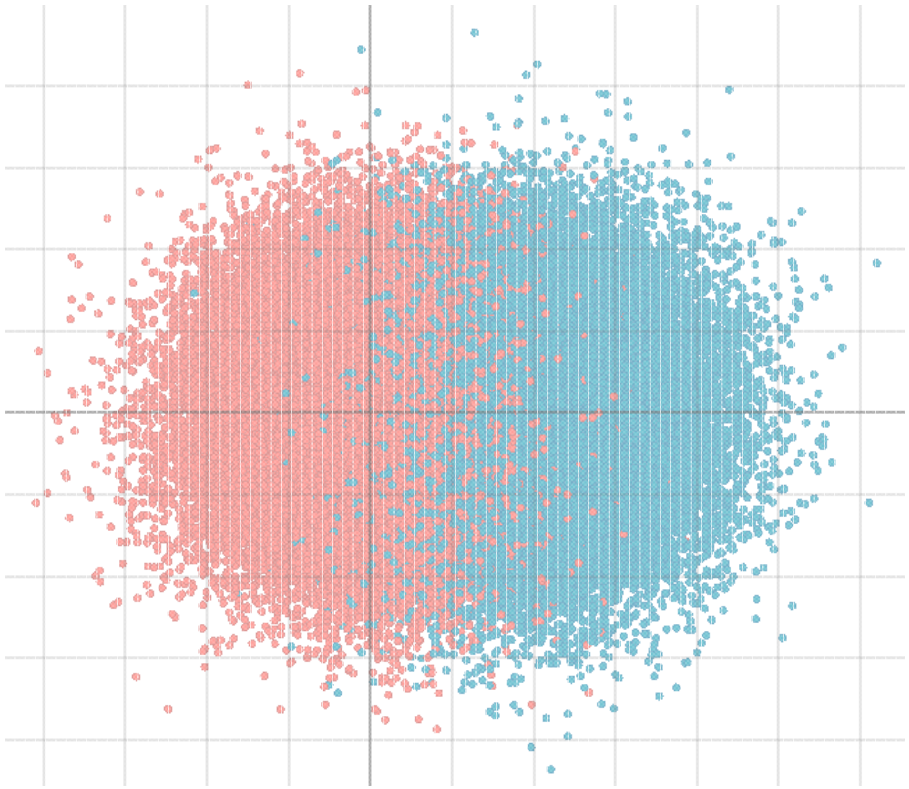
Solution: Splatterplot

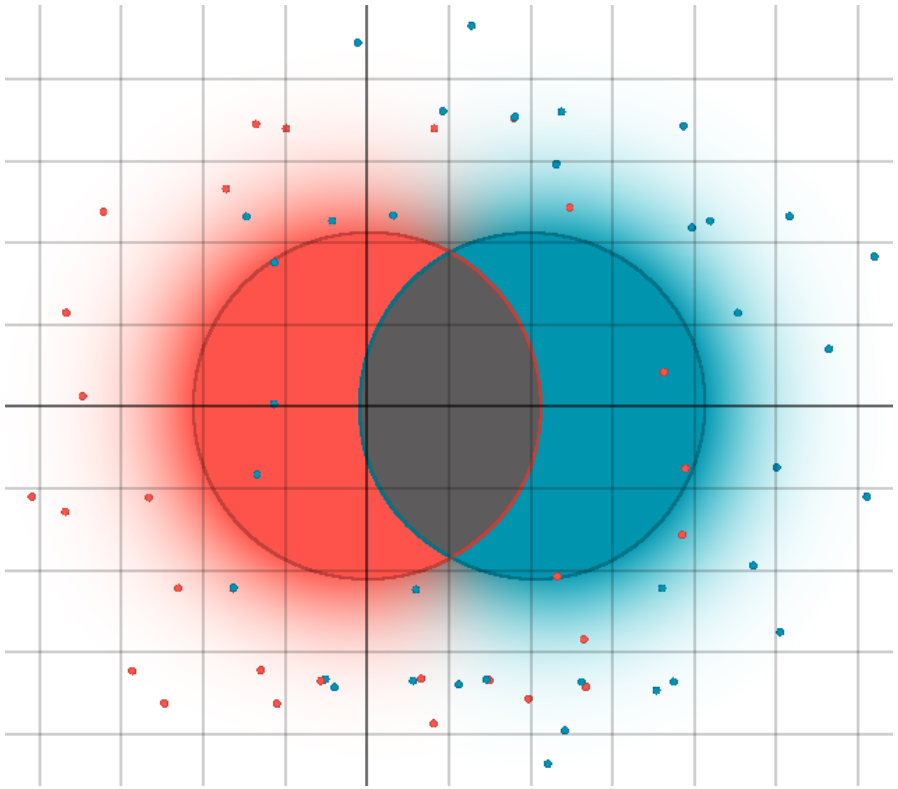


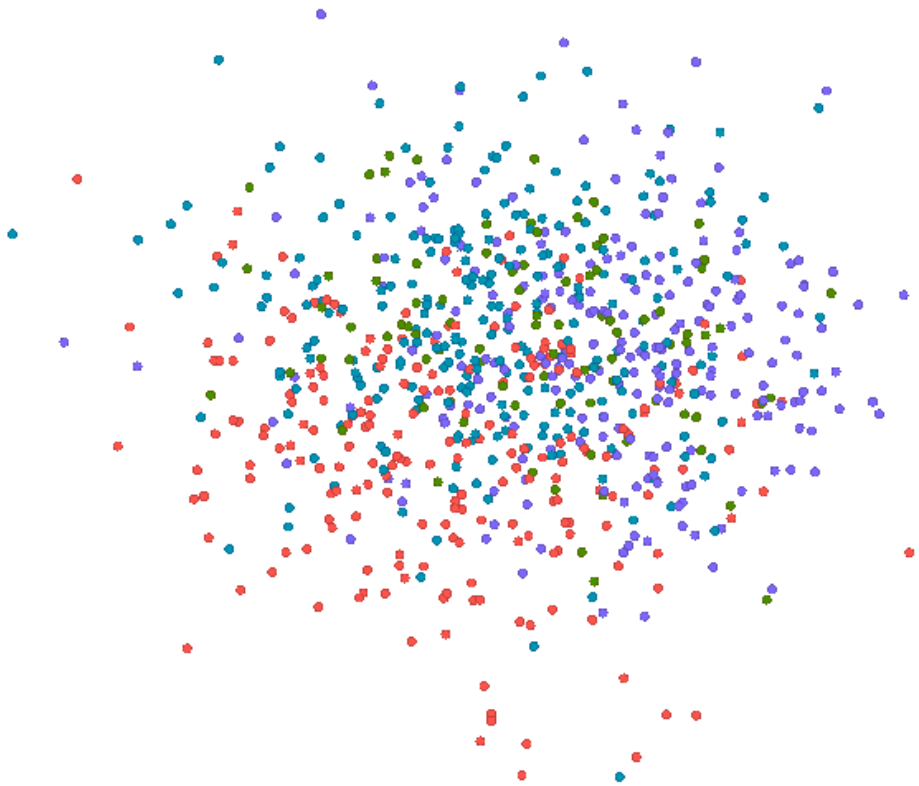


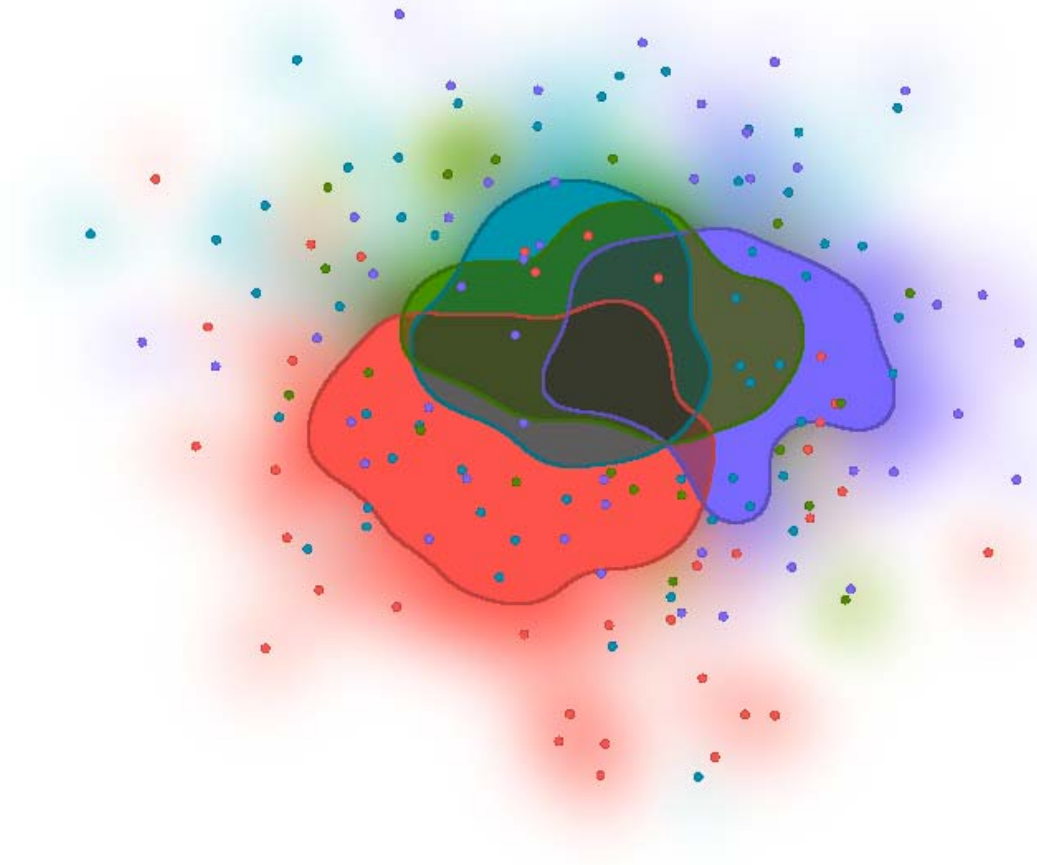
Scatter plots suffer from overdaw.

Solution: Splatterplot

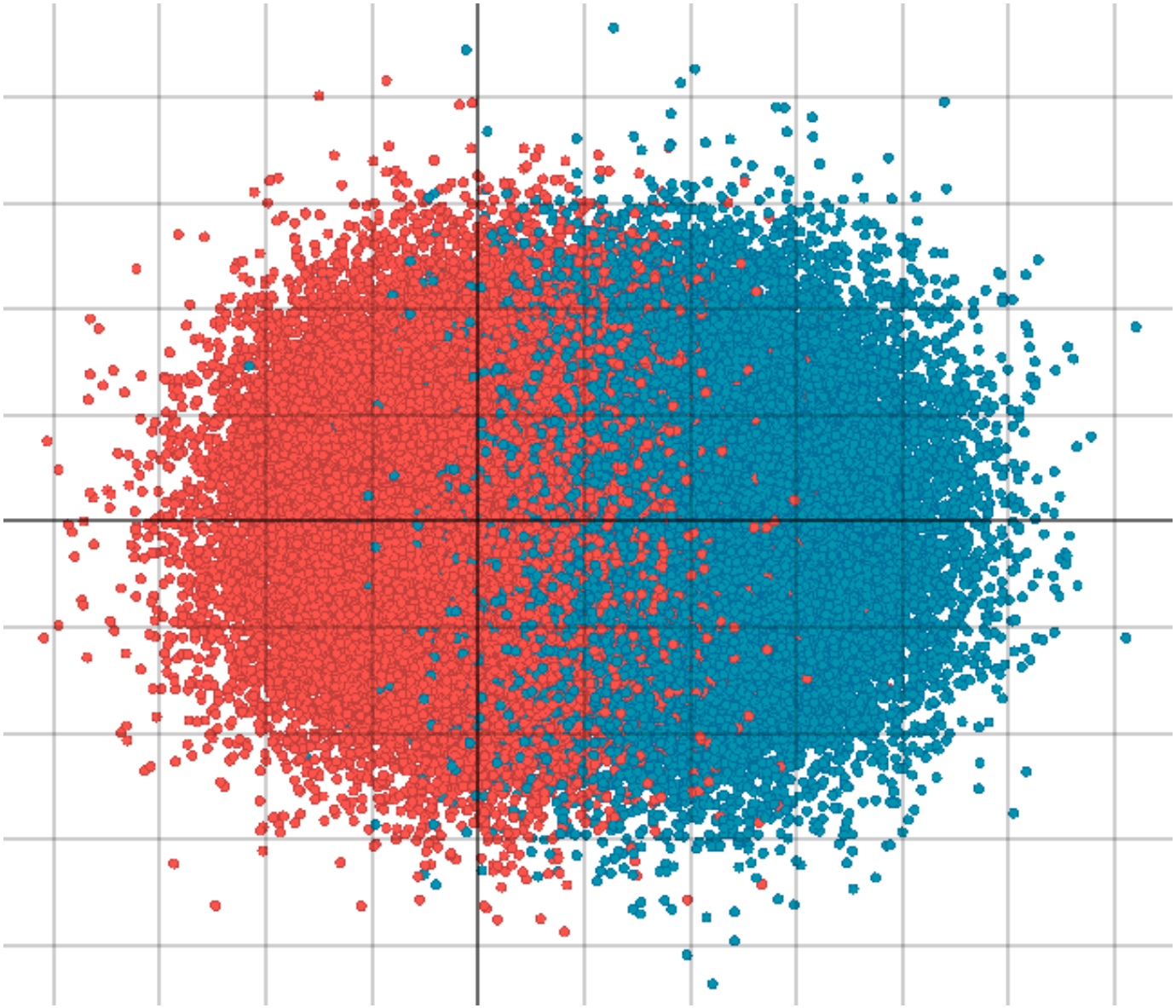




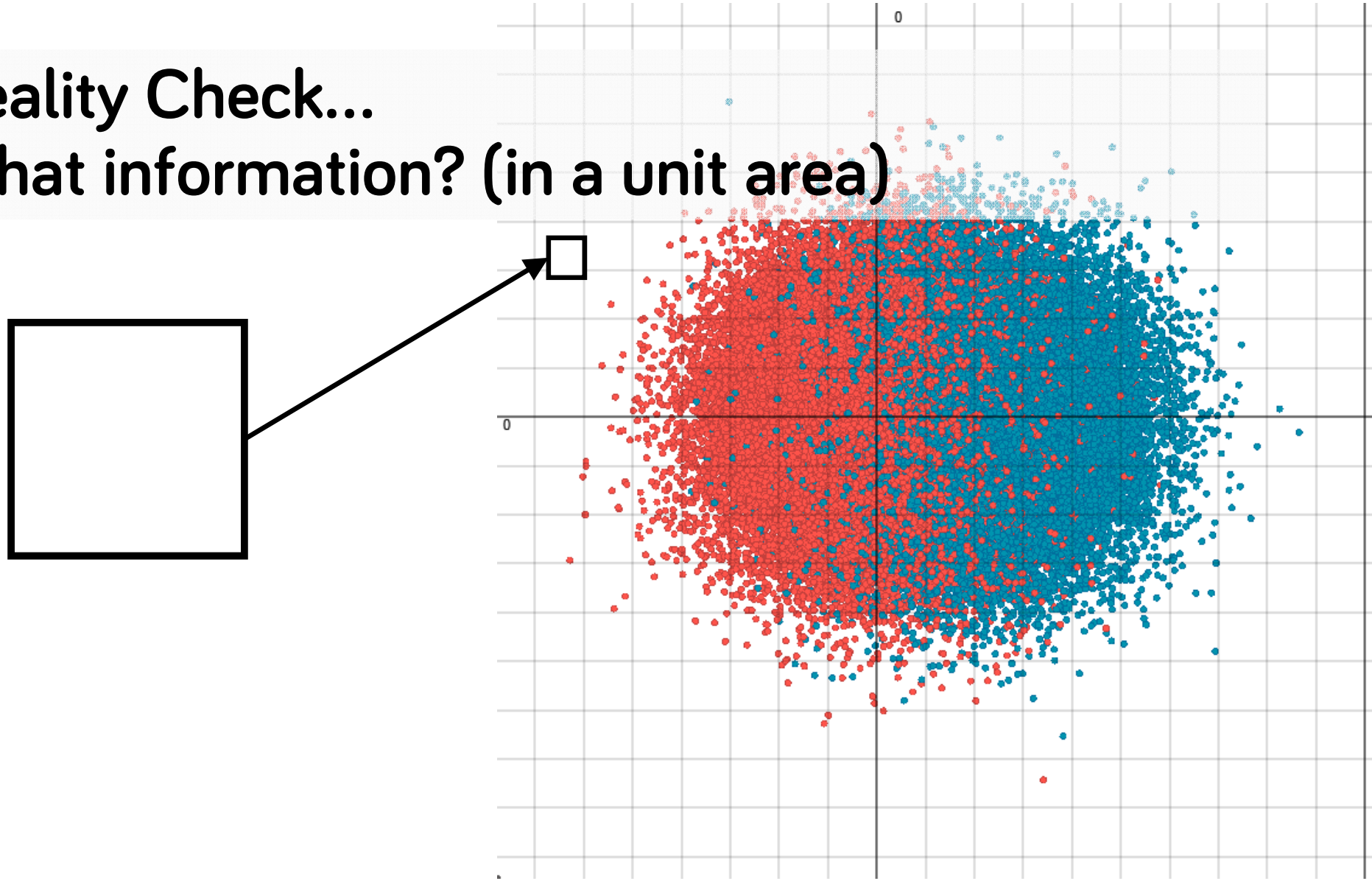




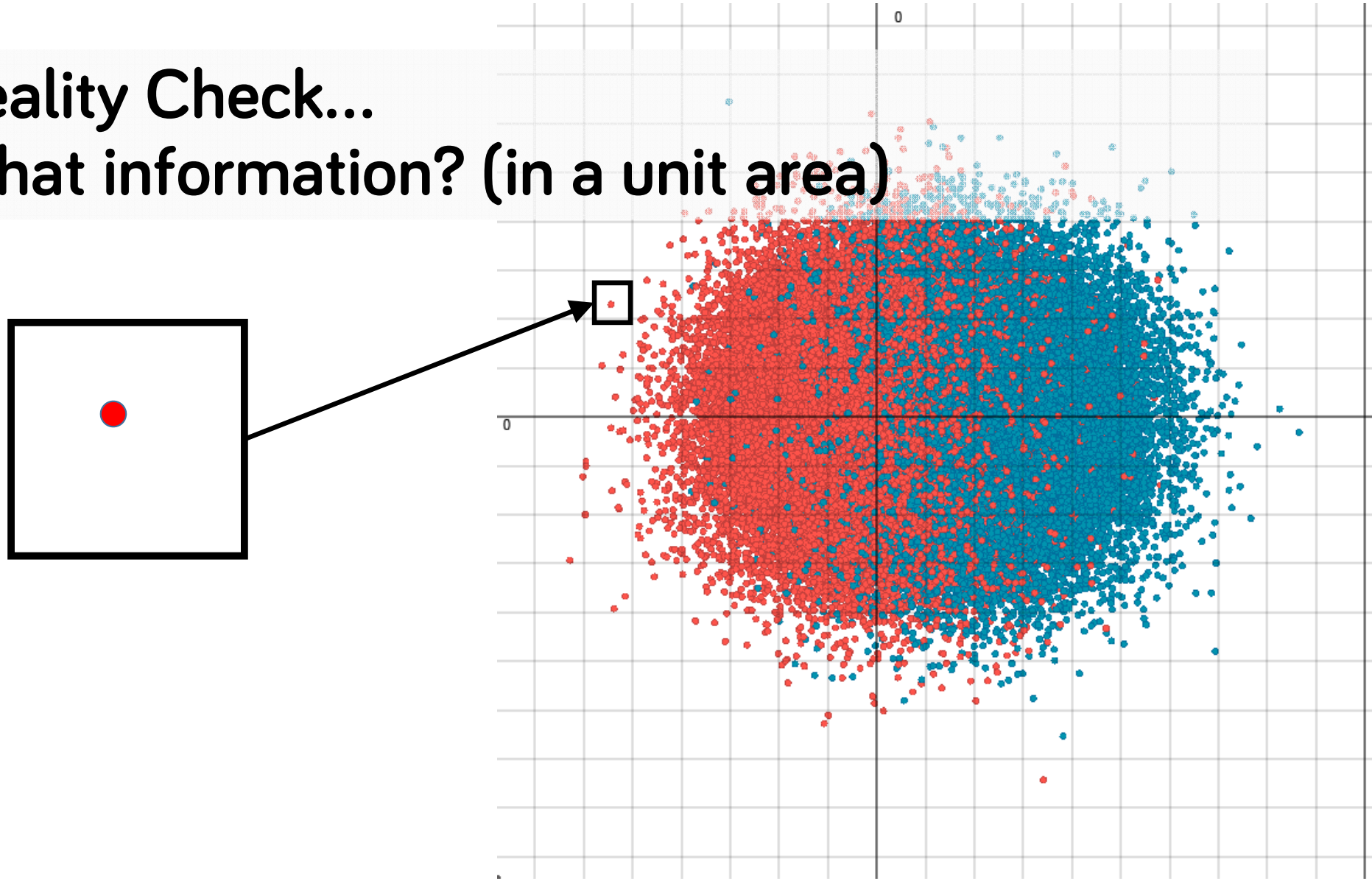
what if you have lots of points?



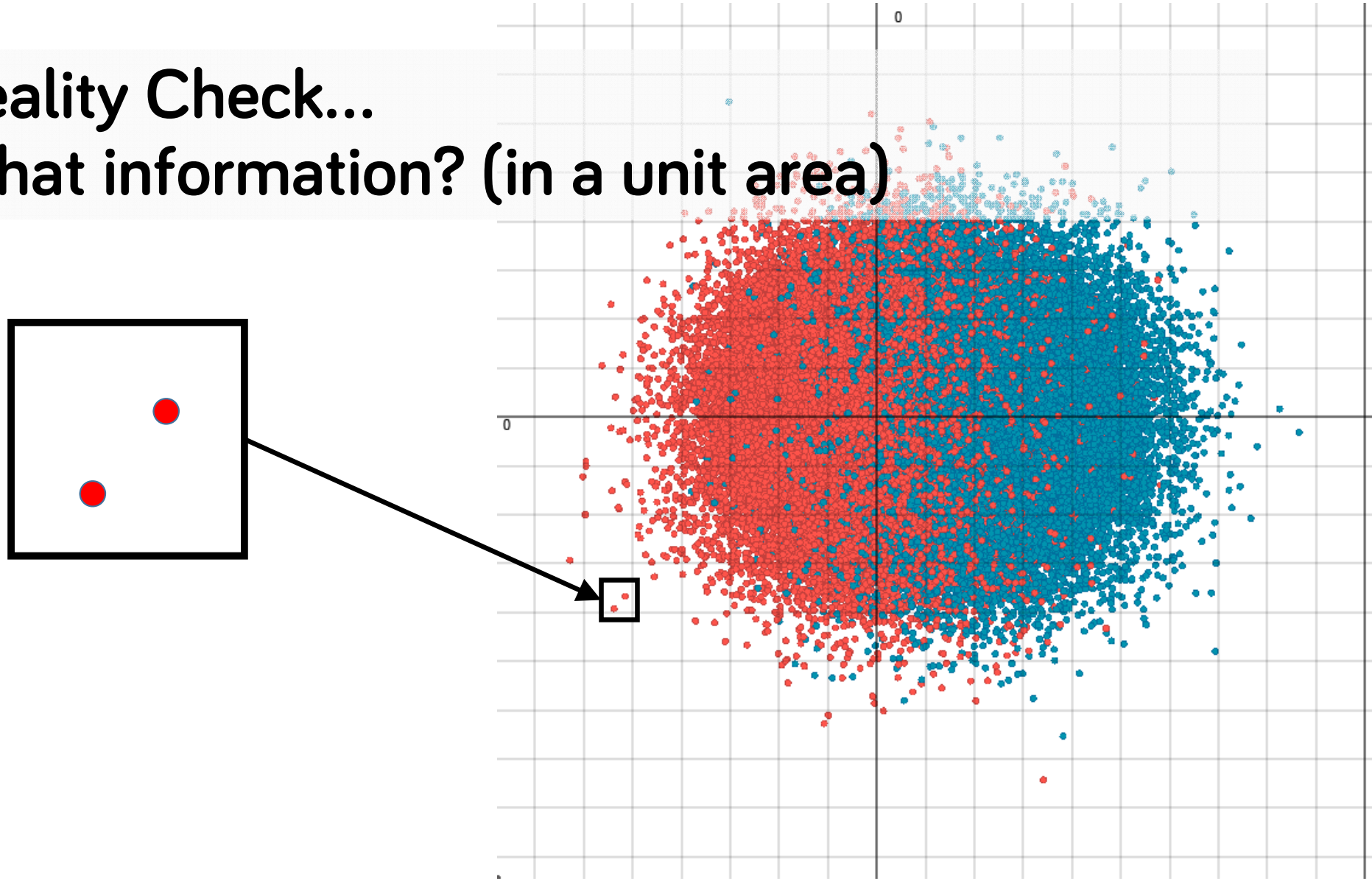
Reality Check...
What information? (in a unit area)



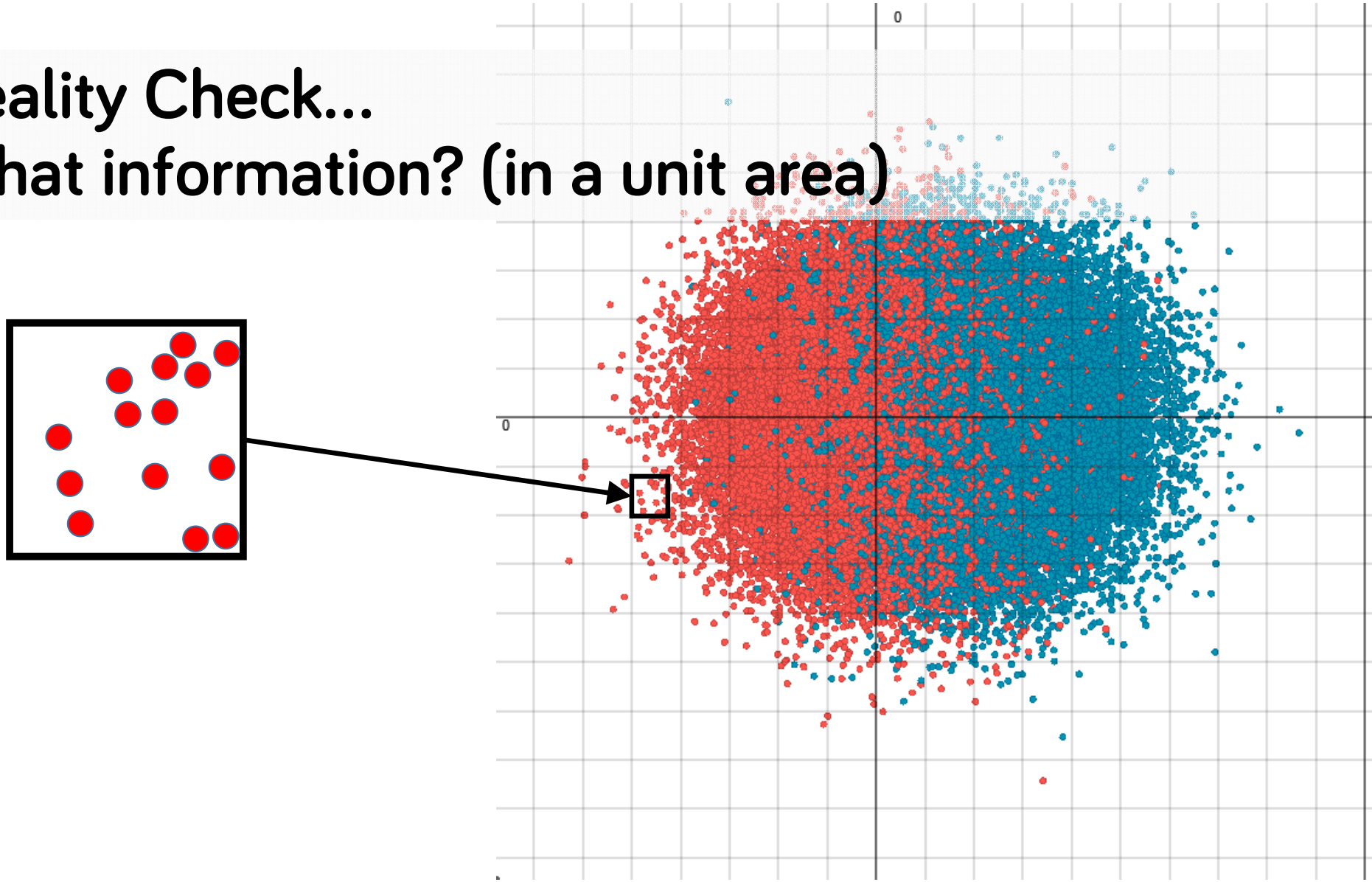
Reality Check...
What information? (in a unit area)



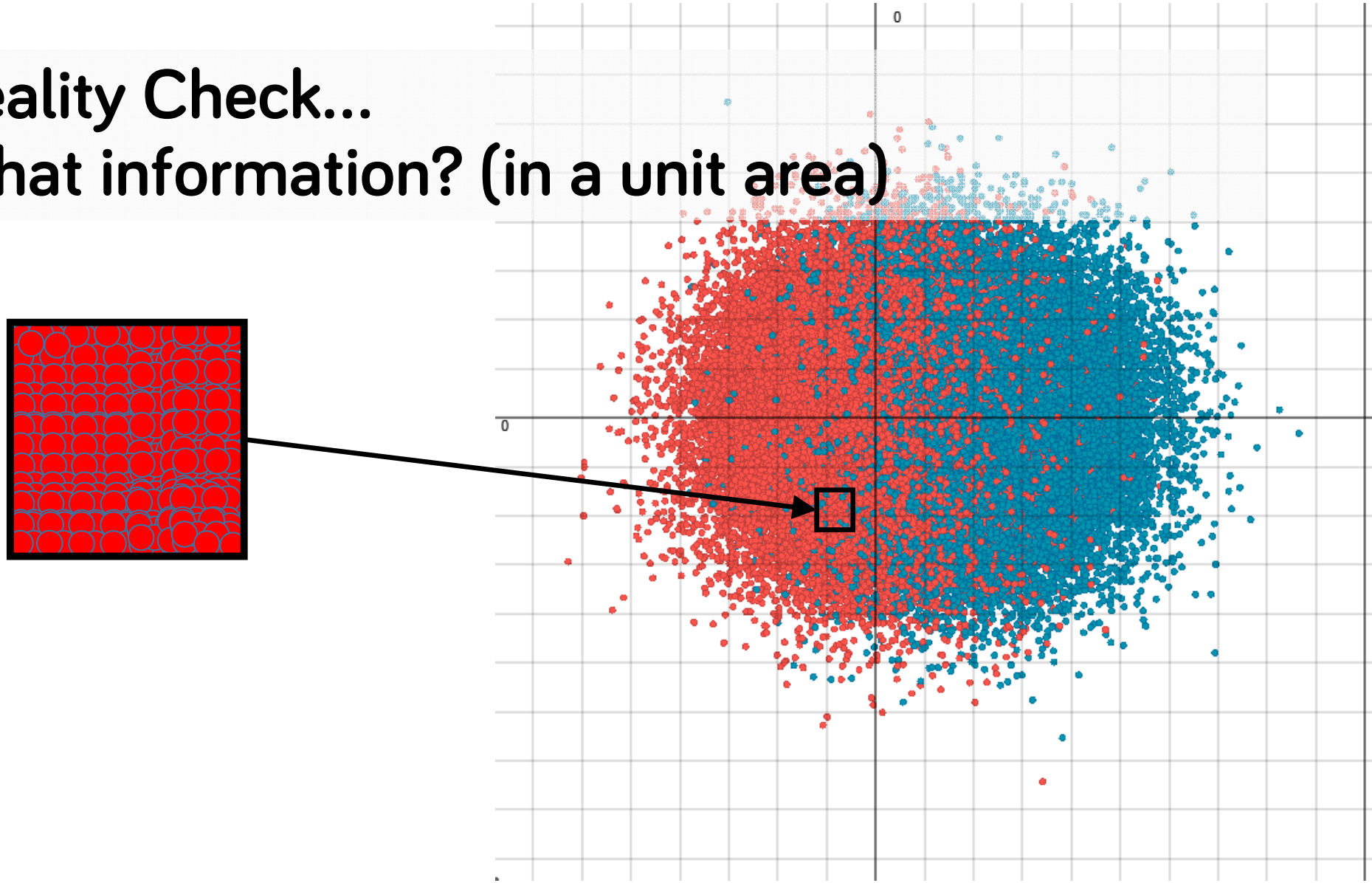
Reality Check...
What information? (in a unit area)



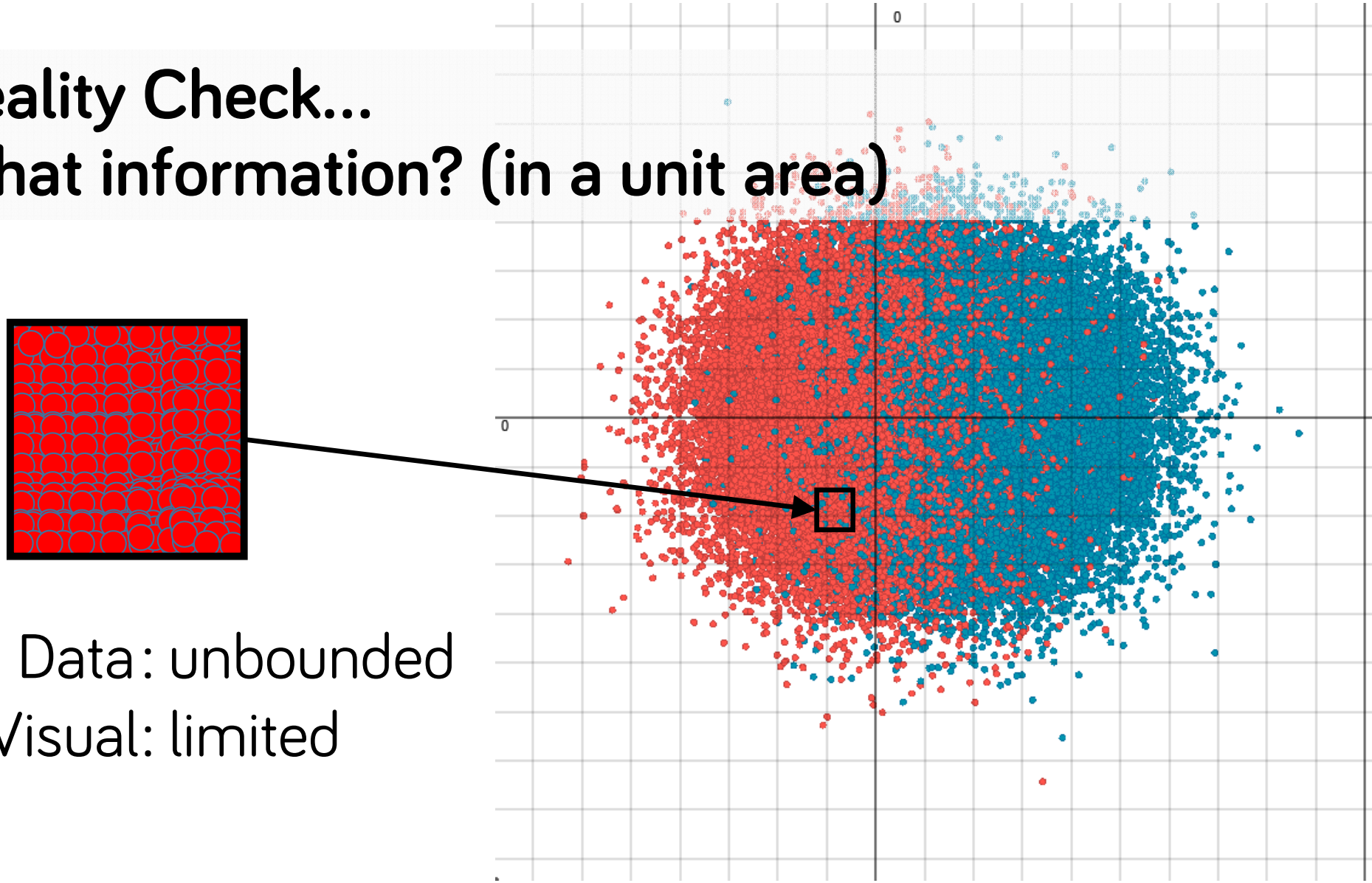
Reality Check...
What information? (in a unit area)



Reality Check...
What information? (in a unit area)



Reality Check...
What information? (in a unit area)



Data: unbounded

Visual: limited

Bounded Information Density

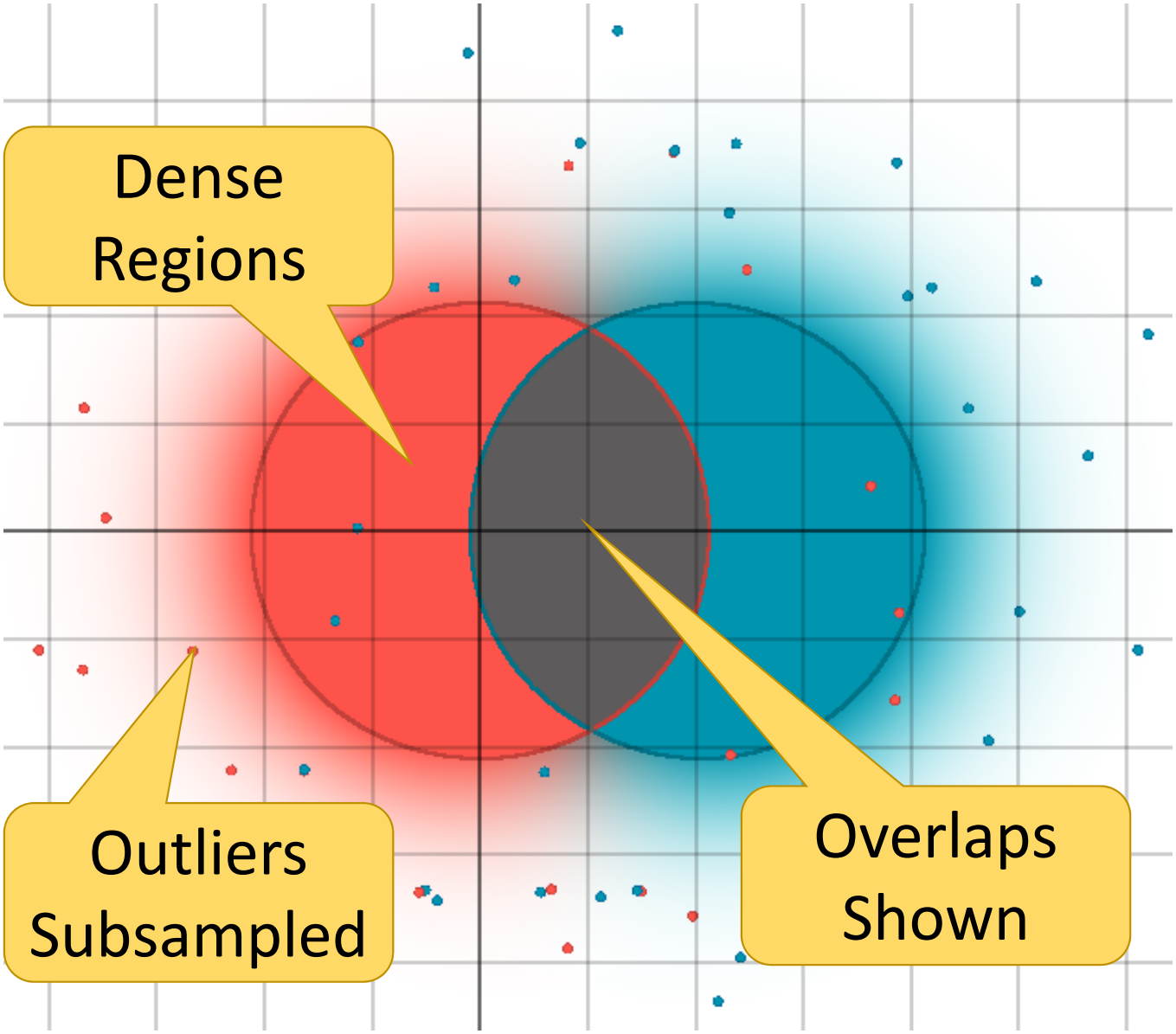
In a Unit of Area:

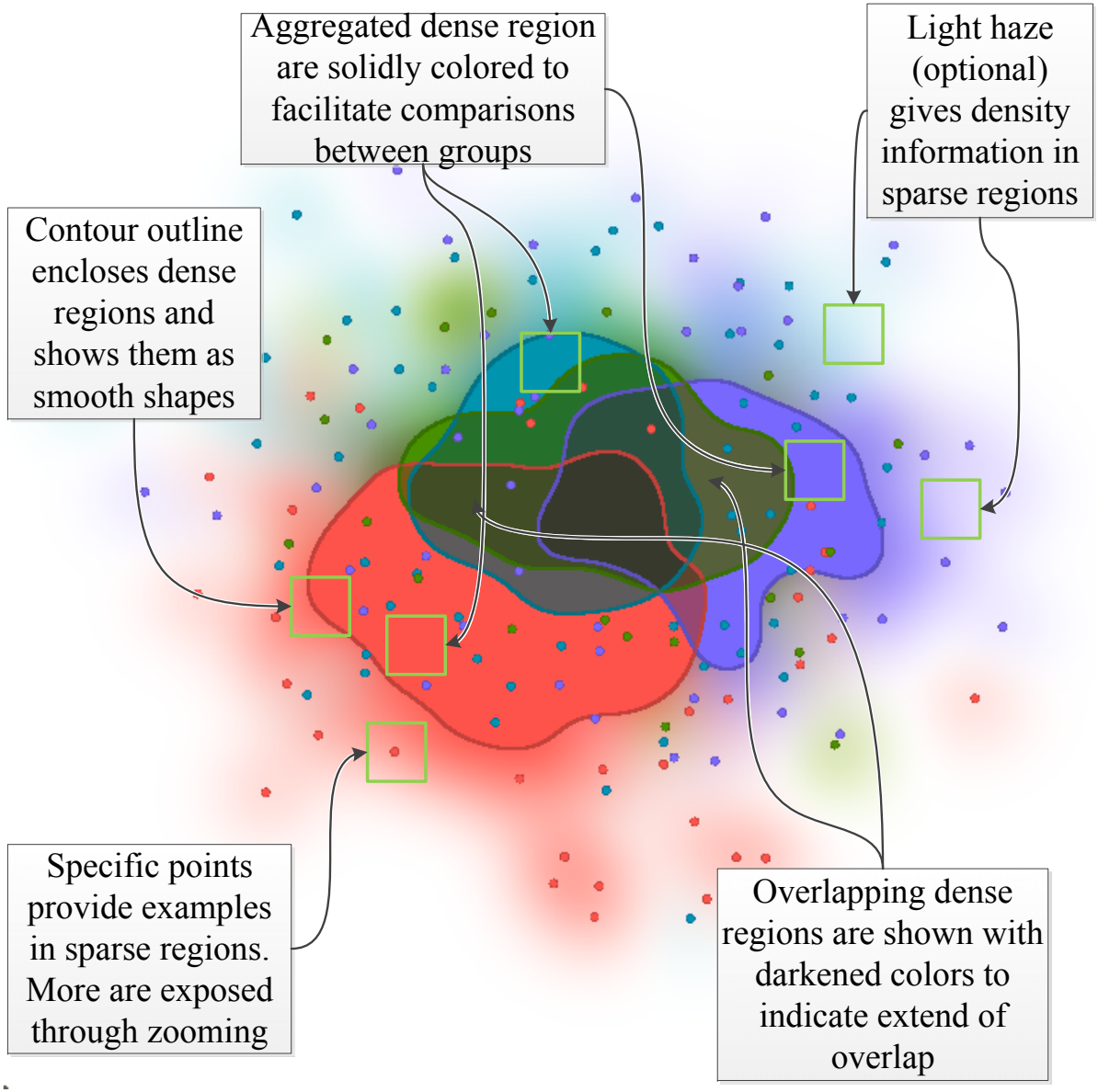
Amount of data is **unbounded**

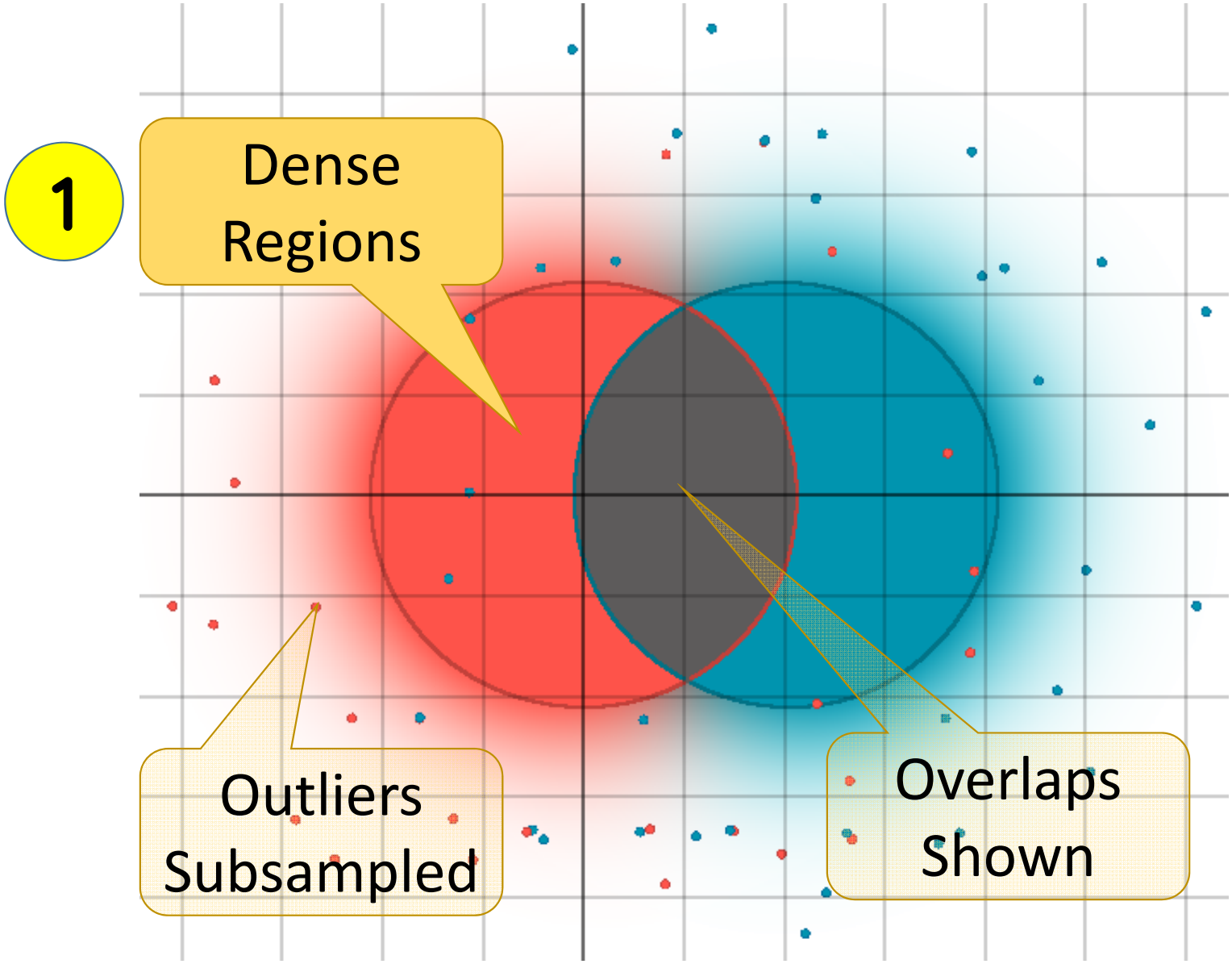
Amount you can see is **limited**

Need to **limit** the amount shown

Choose what to display by **abstracting** the data


Dense Regions
Outliers Subsampled
Overlaps Shown


Aggregated dense region are solidly colored to facilitate comparisons between groups
Light haze (optional) gives density information in sparse regions
Contour outline encloses dense regions and shows them as smooth shapes
Specific points provide examples in sparse regions. More are exposed through zooming
Overlapping dense regions are shown with darkened colors to indicate extend of overlap


1
Dense Regions
Outliers Subsampled
Overlaps Shown

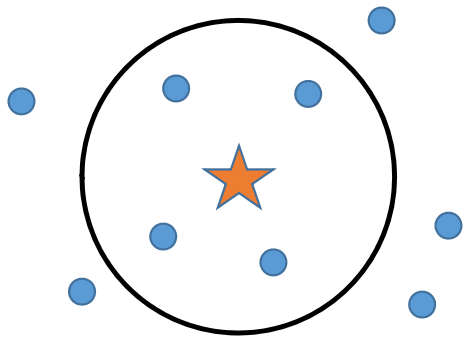
Kernel Density Estimation (KDE)

Count how many points near every position

Weight by distance

Size of kernel (circle) is the bandwidth

Creates smooth fields

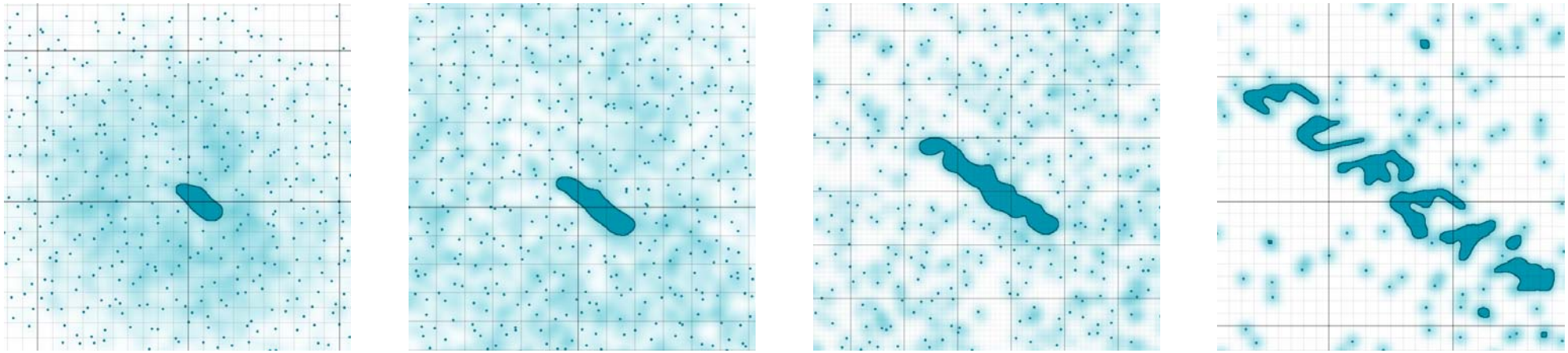


Screen Space KDE

Parameters based on perceptual properties

Independent of data

Does the right thing when you zoom



Discrete dense regions

Threshold

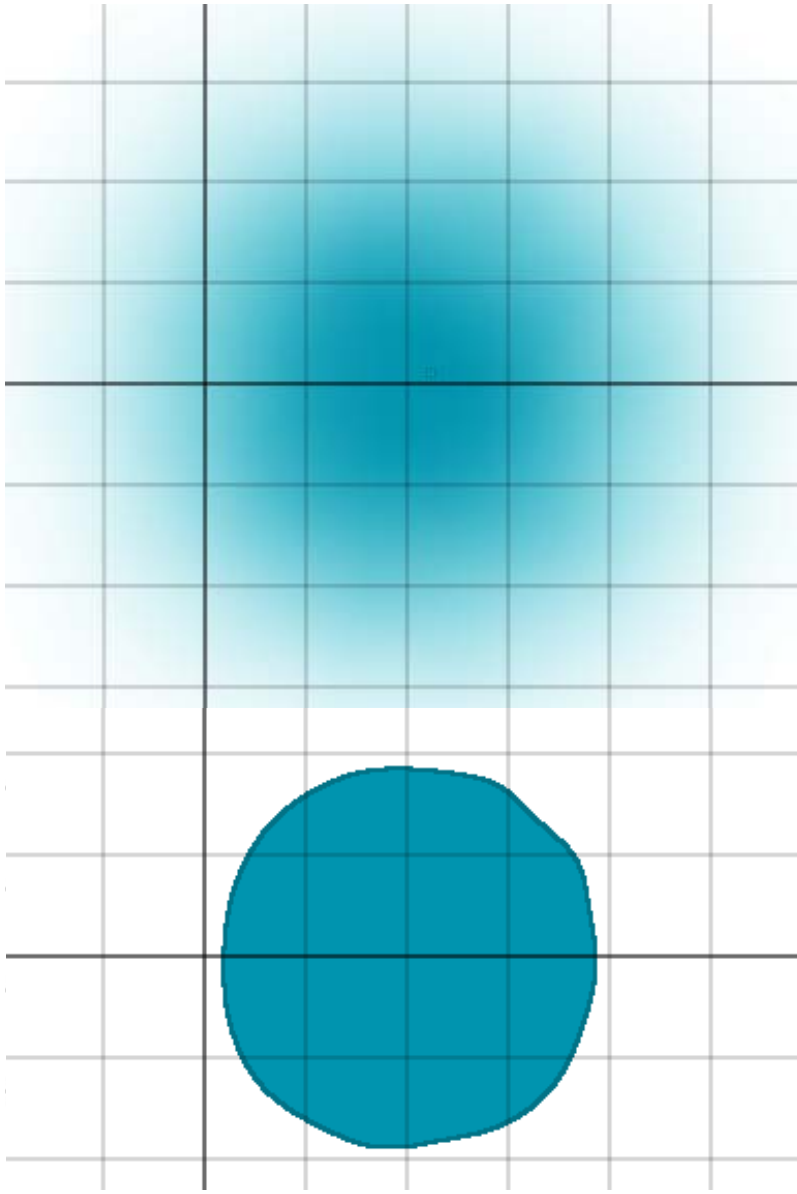
Why? (single set case)

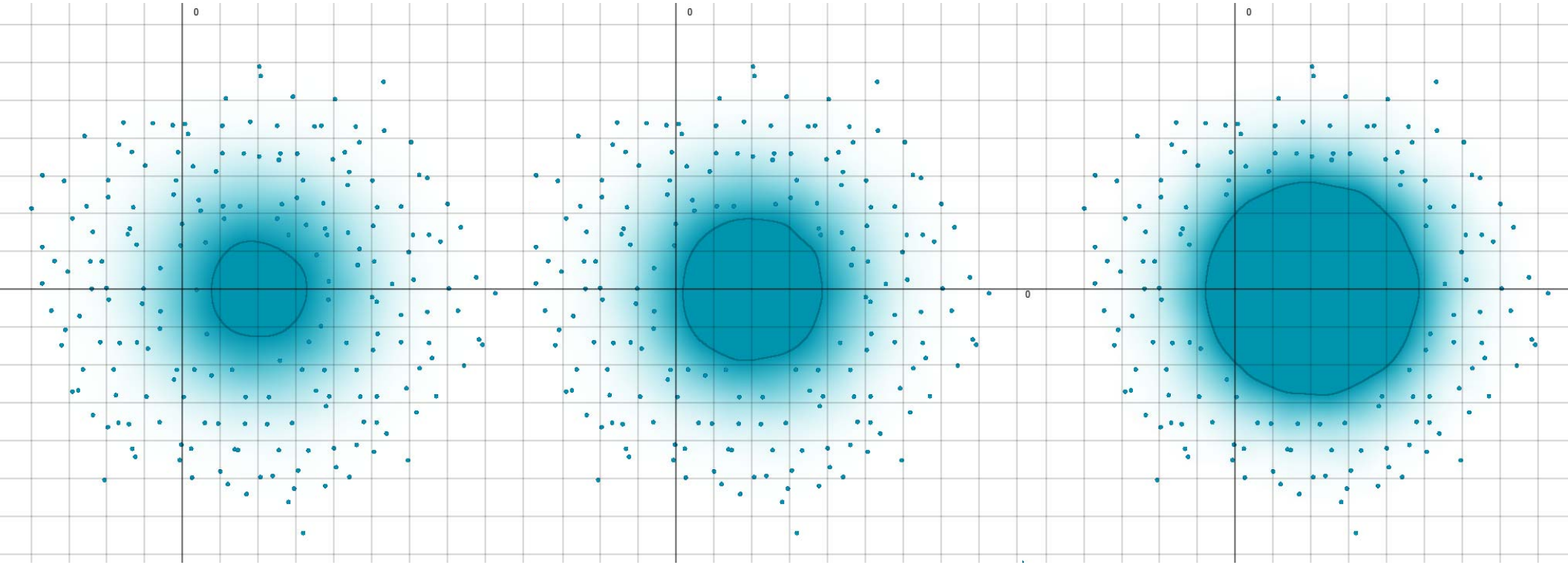
Dynamic range of density may be high
and hard to encode

At some point, it's just "dense"

Crisp boundaries are better visually

Information is thrown away!

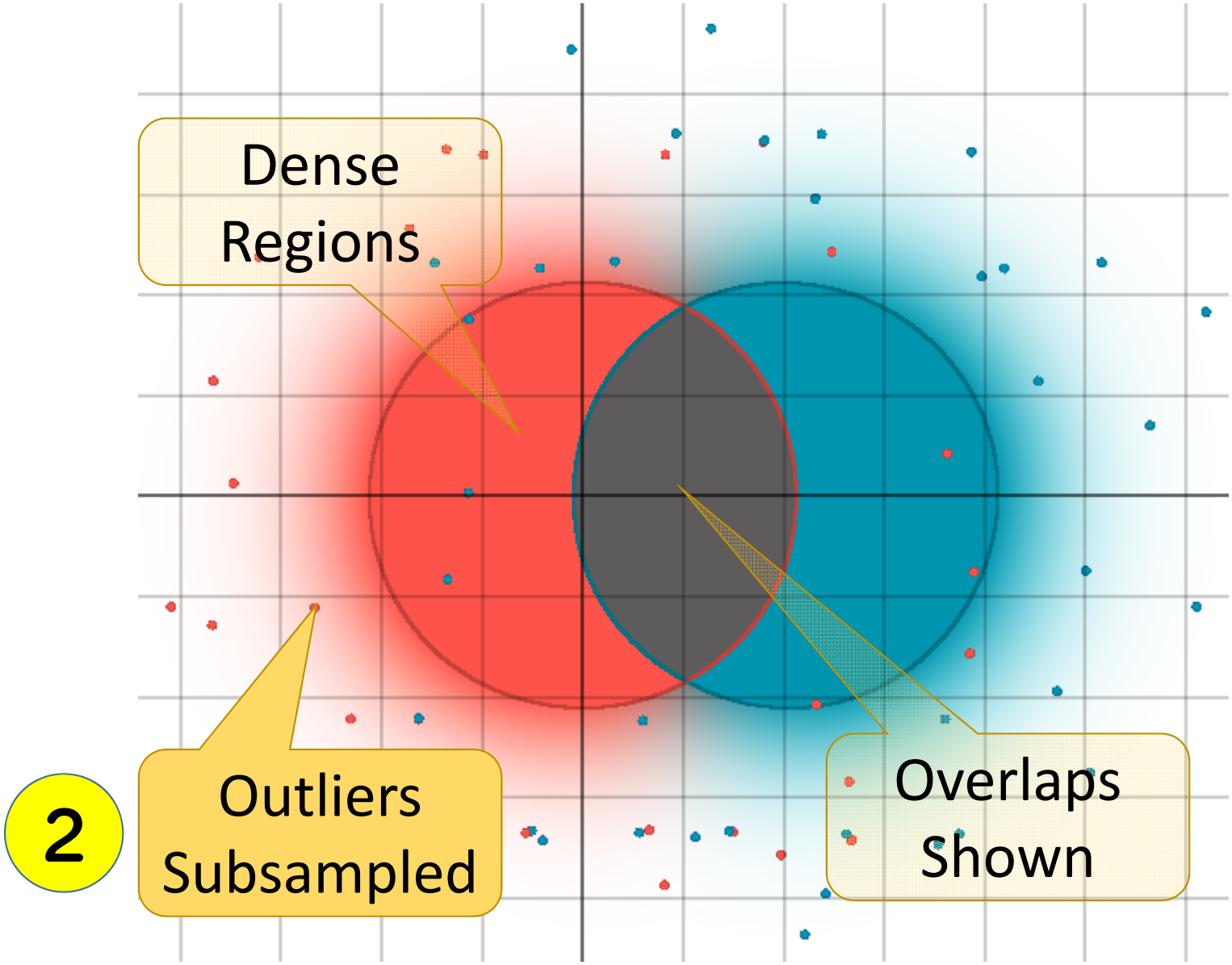




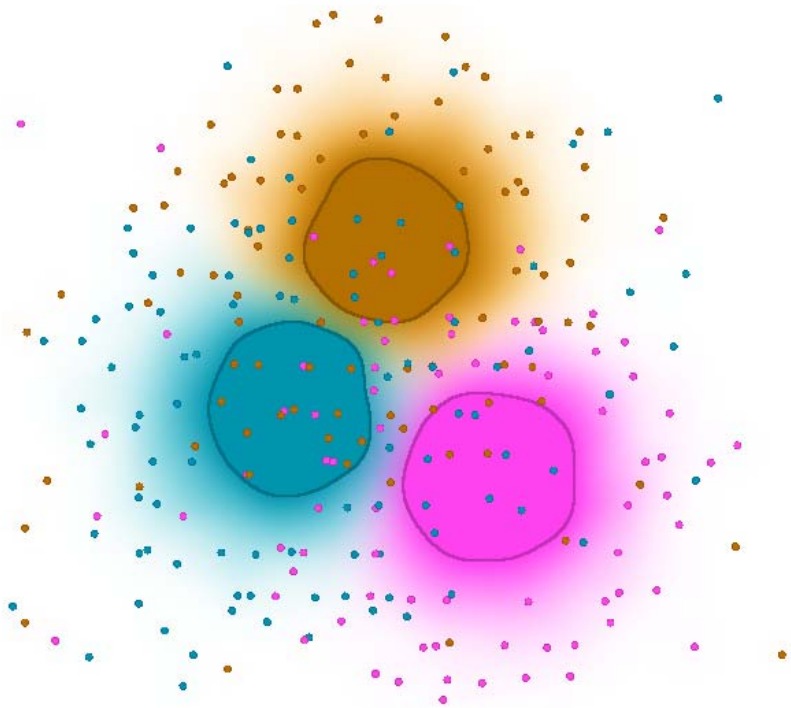
Information is thrown away!

Interactive control of threshold

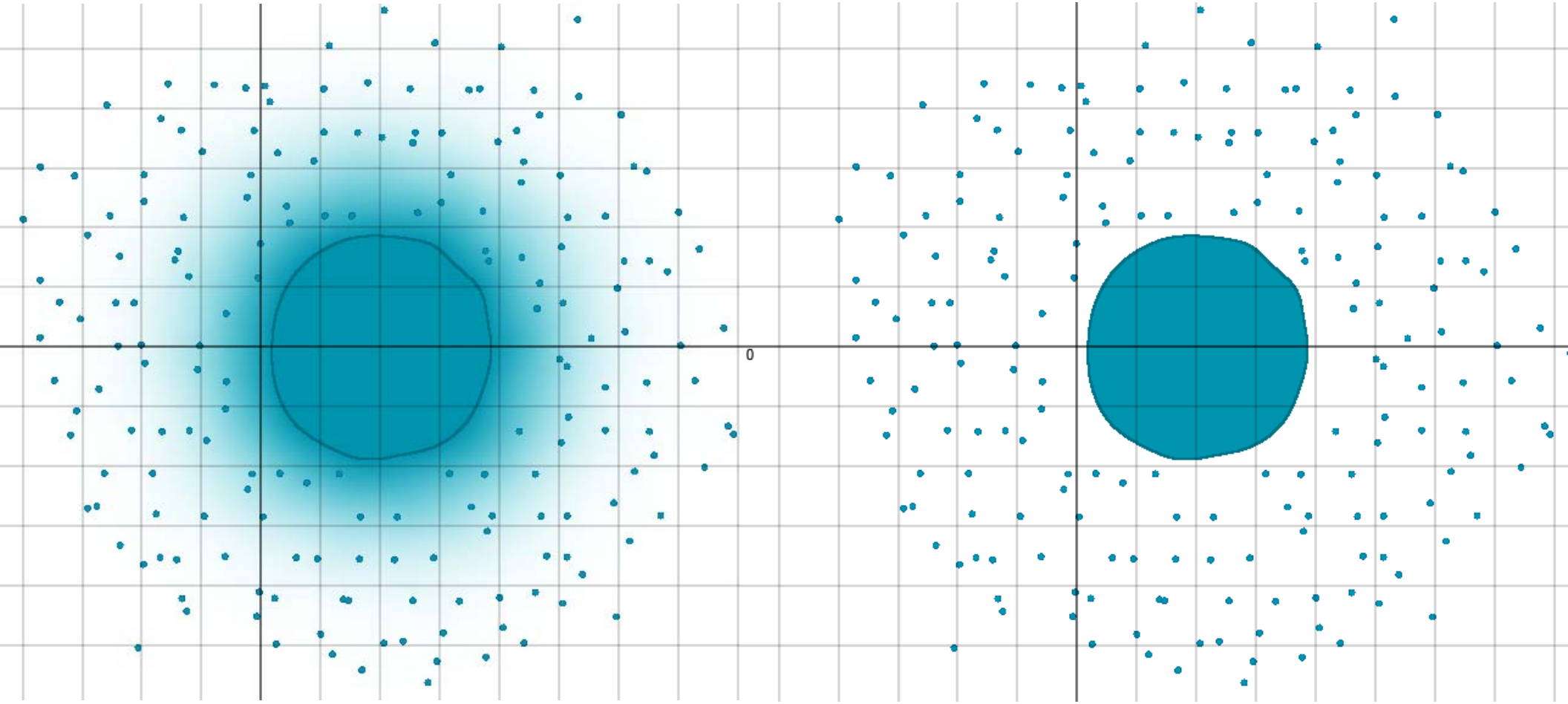
Encode sparse regions differently


Dense Regions
Outliers Subsampled
Overlaps Shown
2

Subsample sparse regions



To Haze or not to Haze?

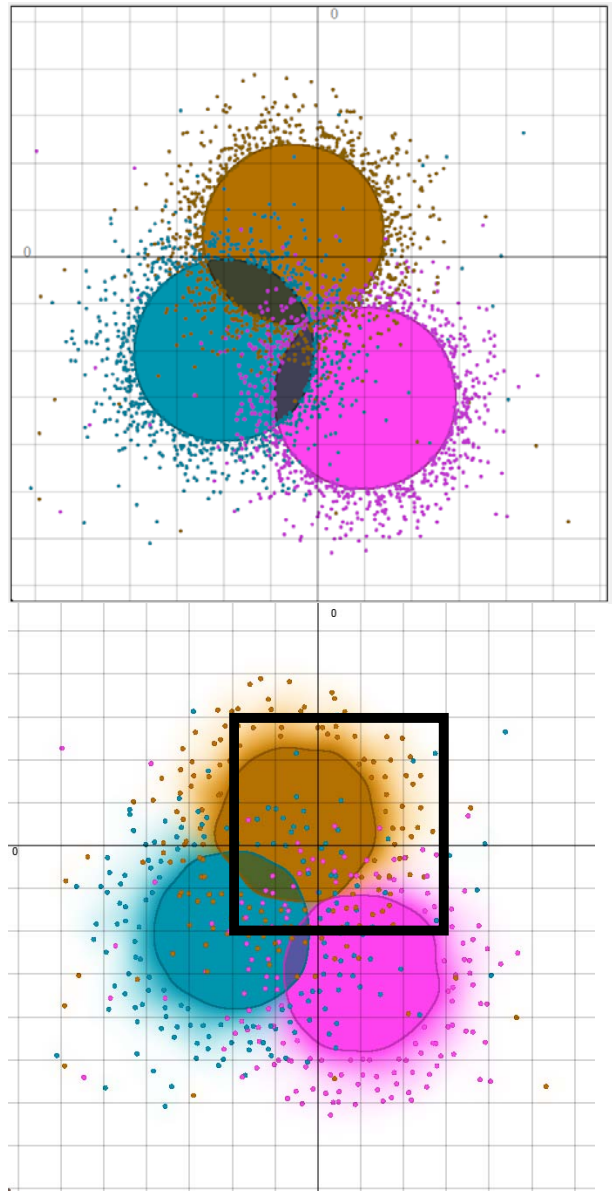


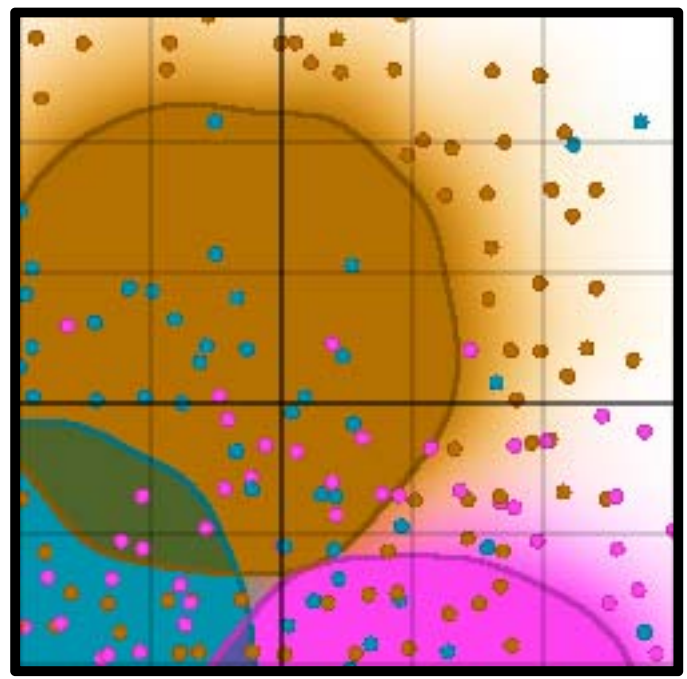
Edges

Strokes

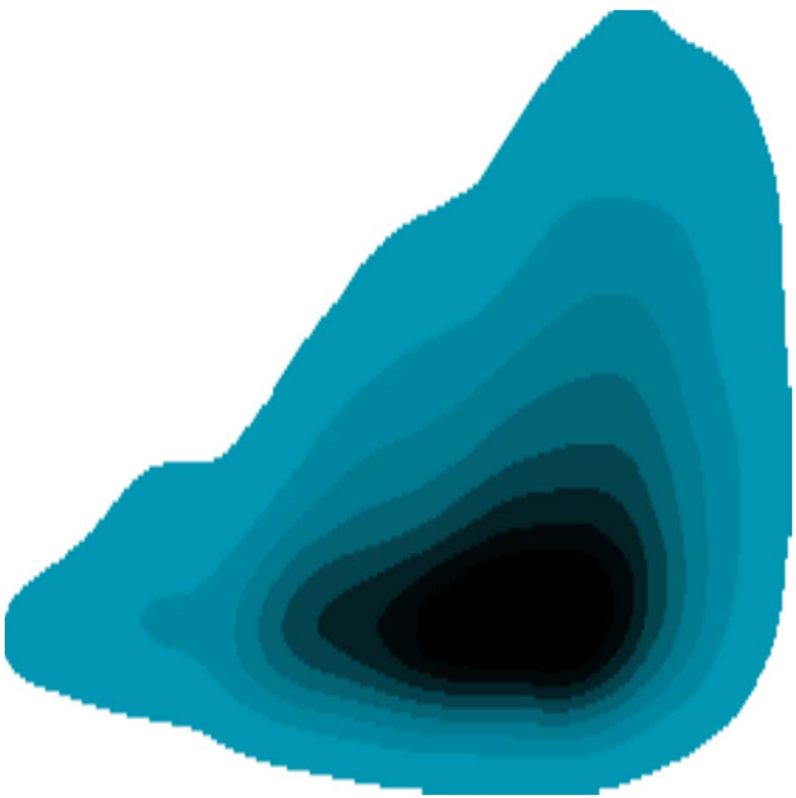
Clear Clutter

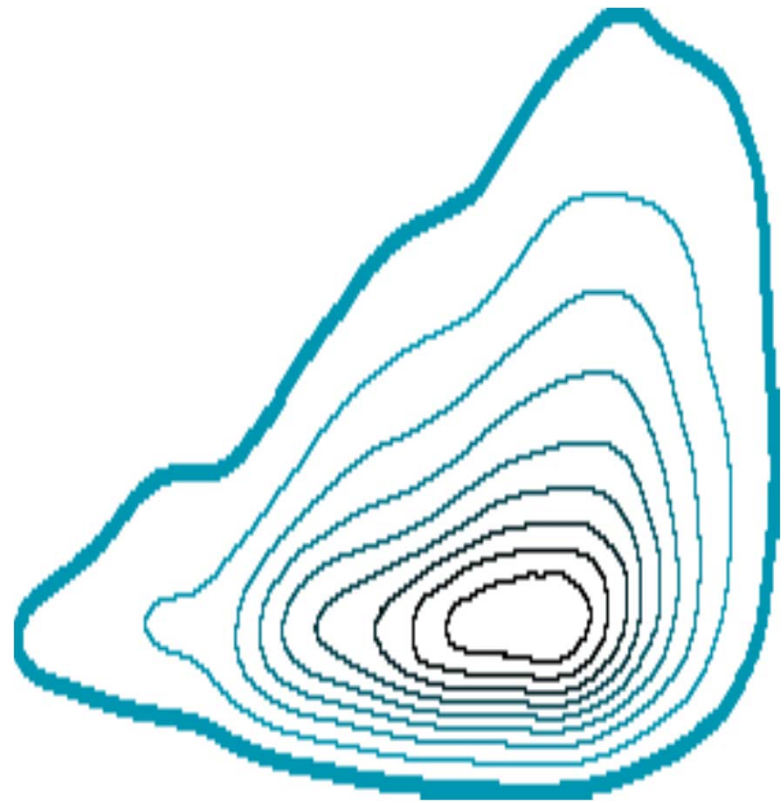
Both require
distance to region



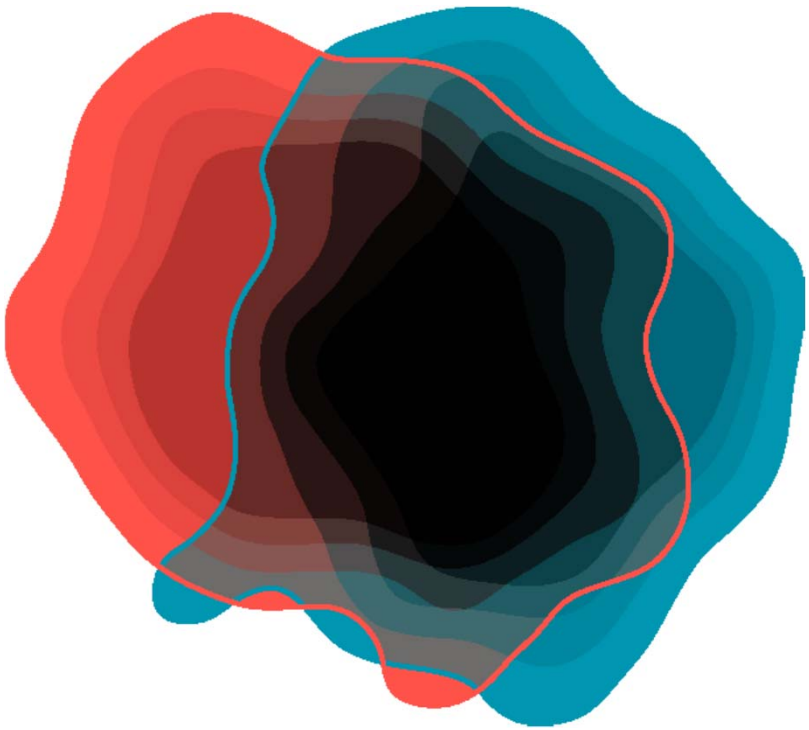


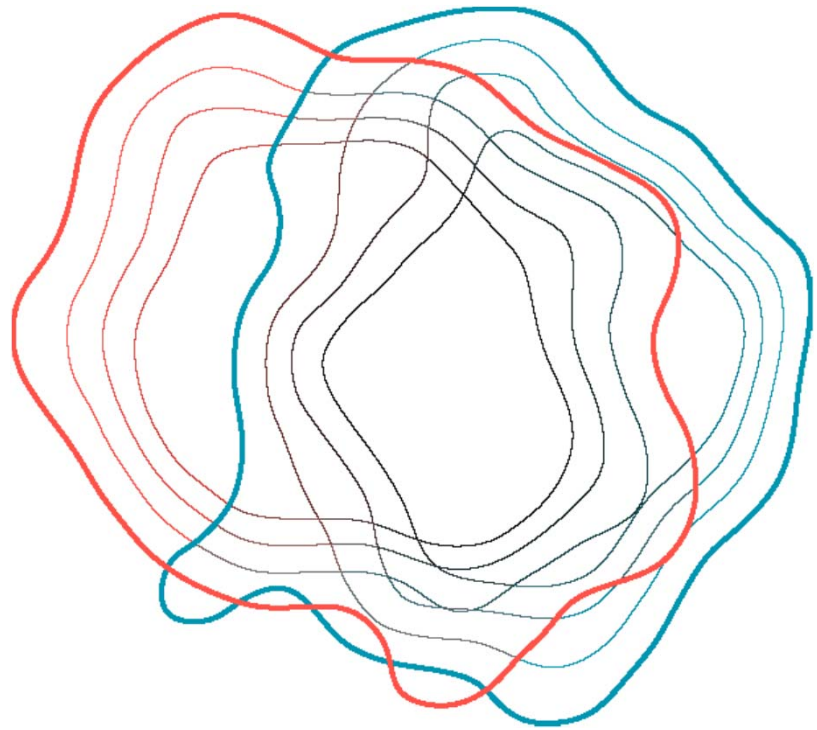
Contours?

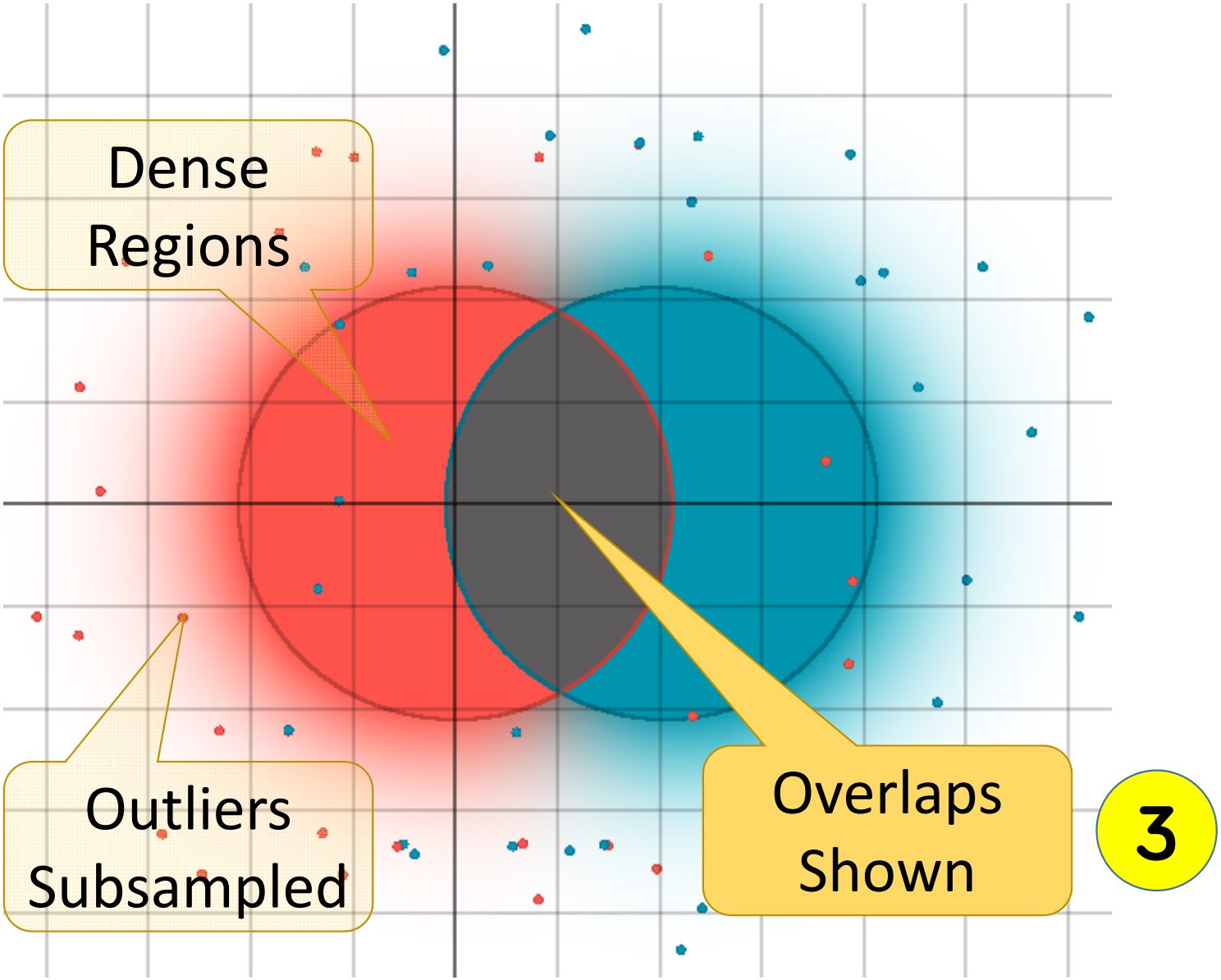




Complicated with multiple groups





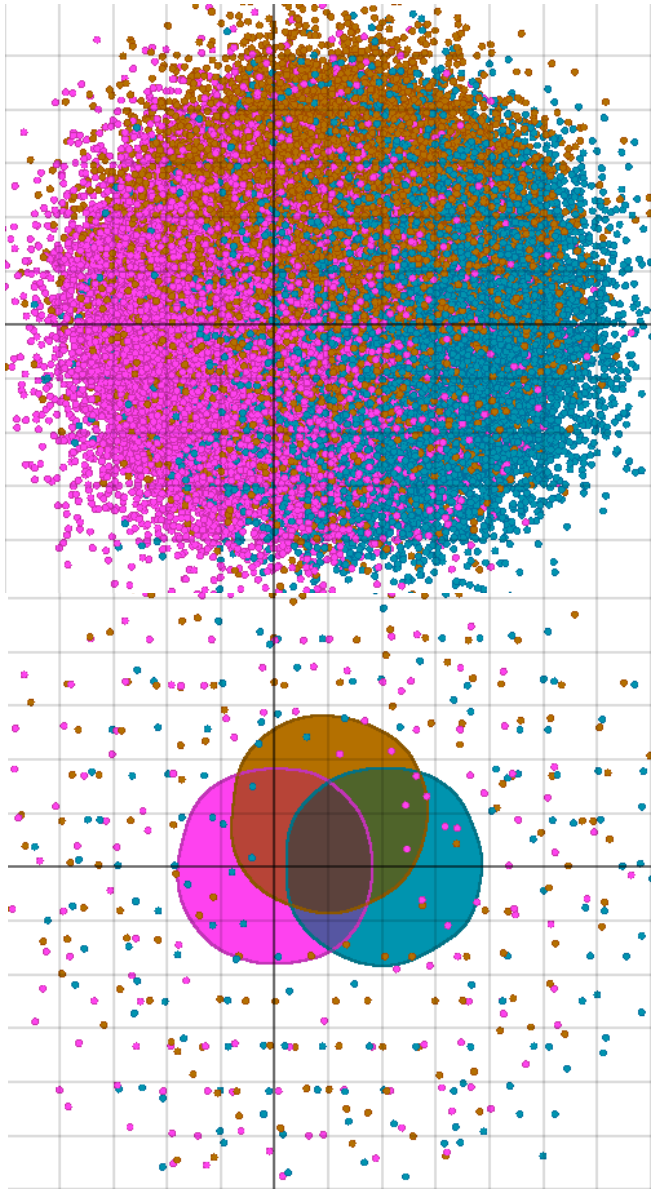

Dense Regions
Outliers Subsampled
Overlaps Shown
3

Multiple Groups

Compute densities independently

Color per group

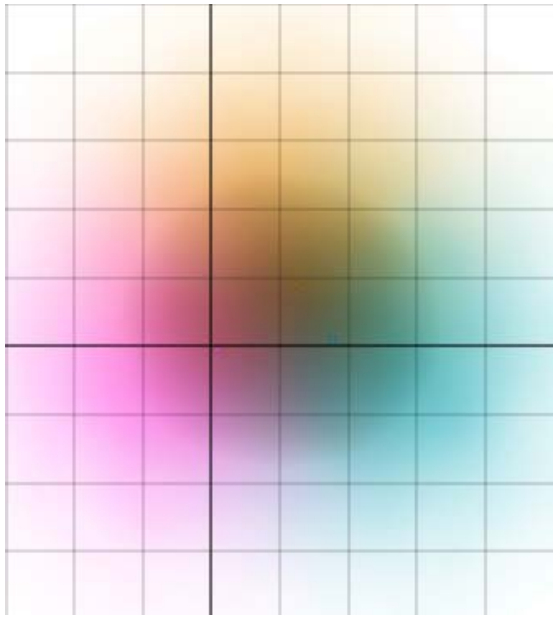
Pick distinctive colors



Colors for combinations

Multi-variate color encoding?

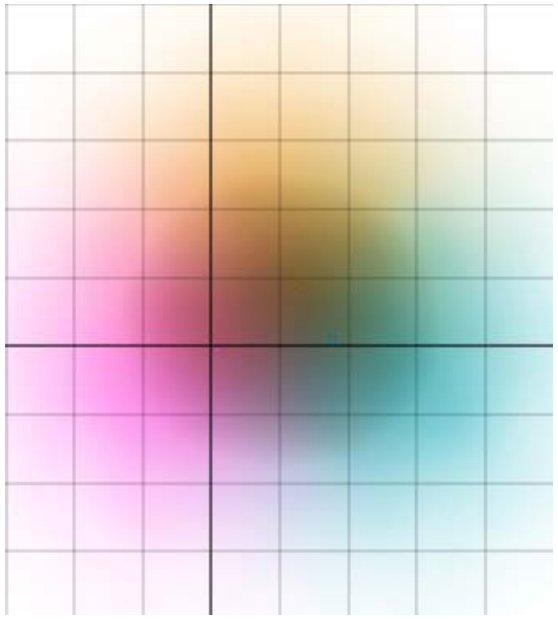
Map R^n to a color



Colors for combinations

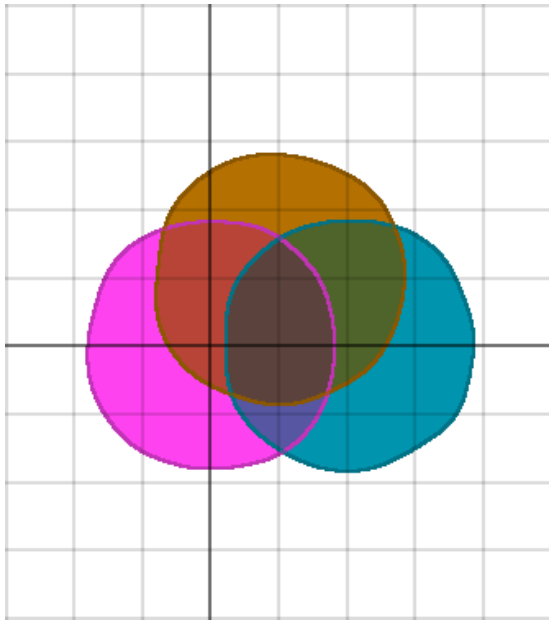
Multi-variate color encoding?

Map R^n to a color



Colors for set combinations

Map 2^n set combinations to colors



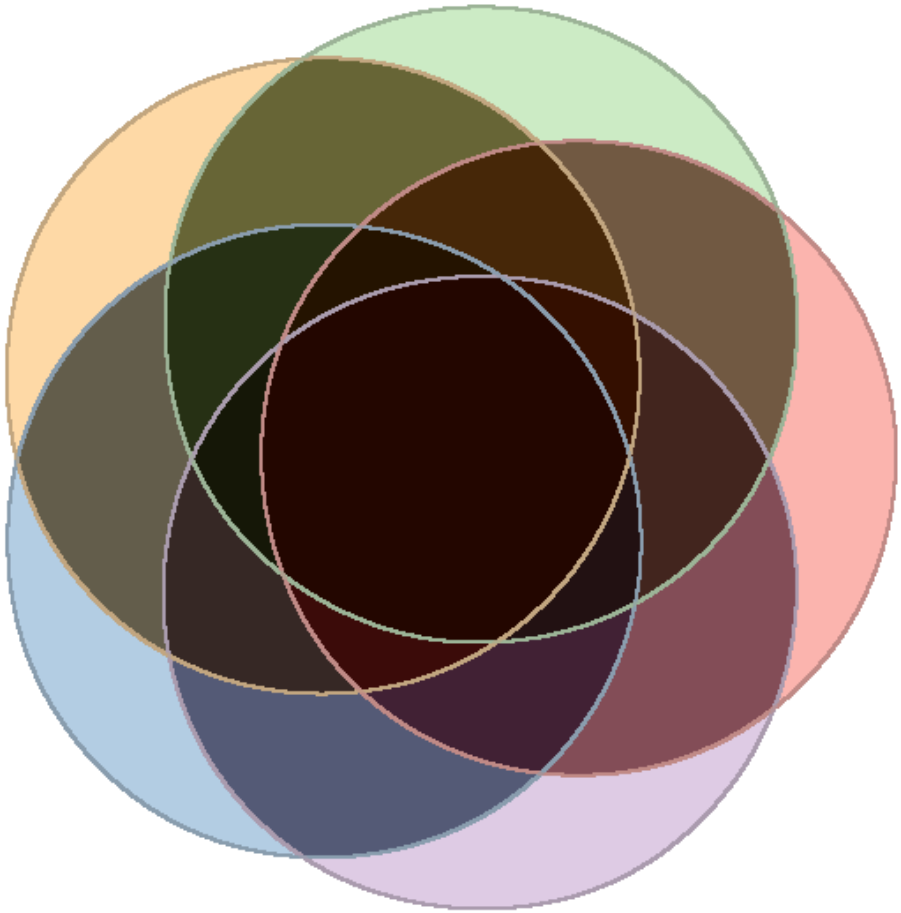
Color Blending

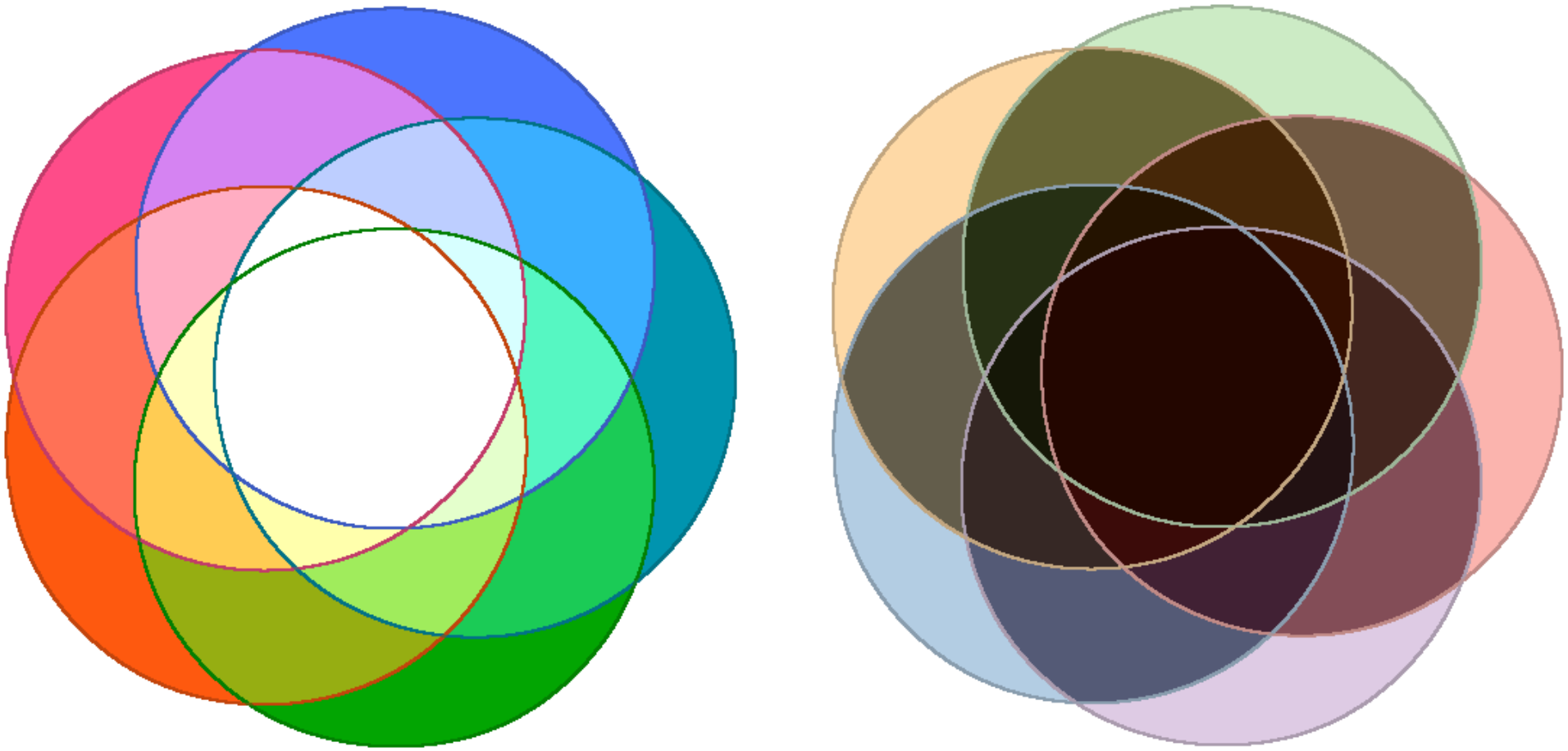
Encode sets with color

Hue = set

Lightness = number of overlaps

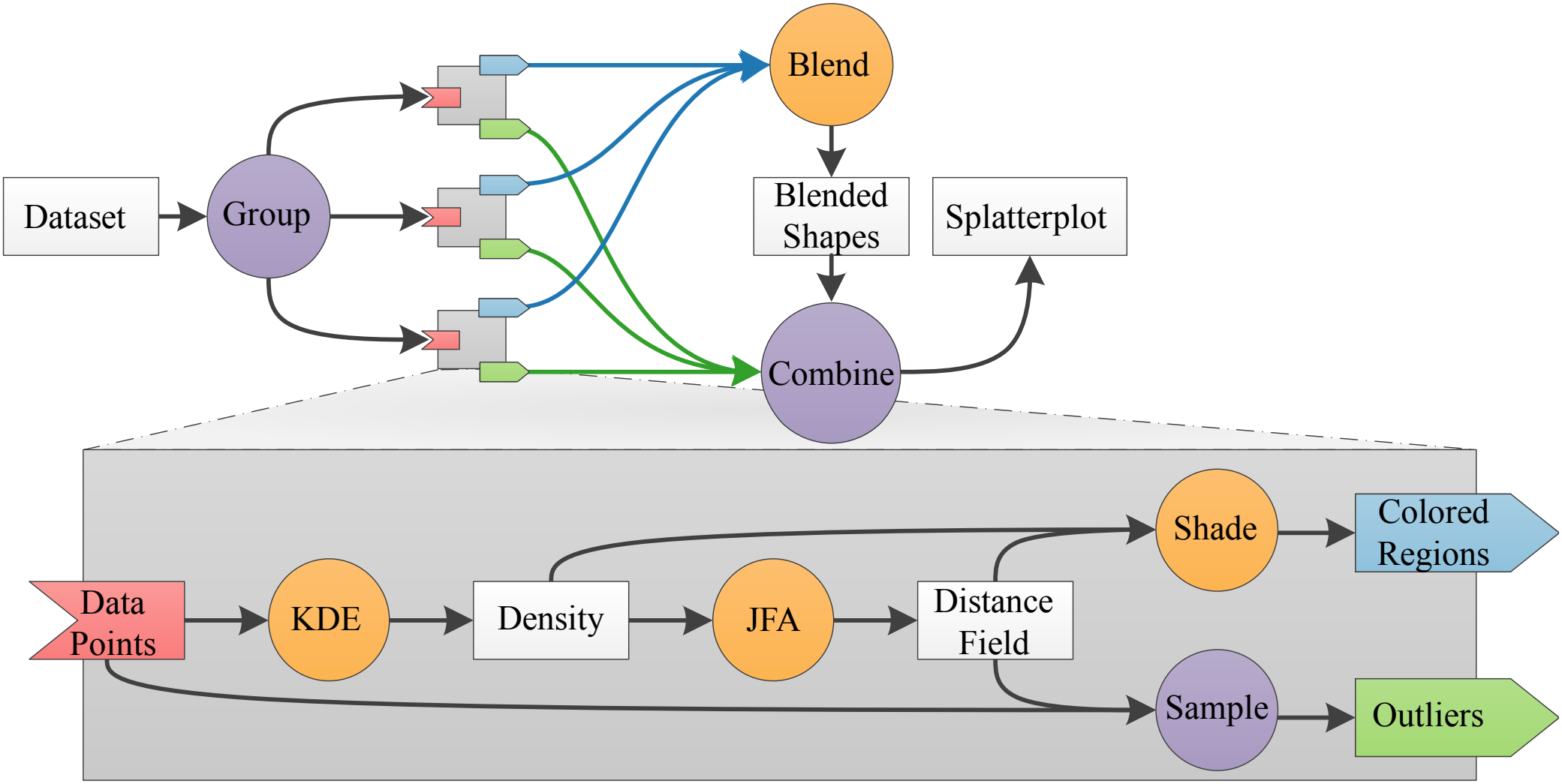
See evaluation in paper





Interactivity is critical!

Implementation


Dataset
Group
Blend
Blended Shapes
Splatterplot
Combine
Data Points
KDE
Density
JFA
Distance Field
Shade
Colored Regions
Sample
Outliers

Performance: Use the GPU

Draw points

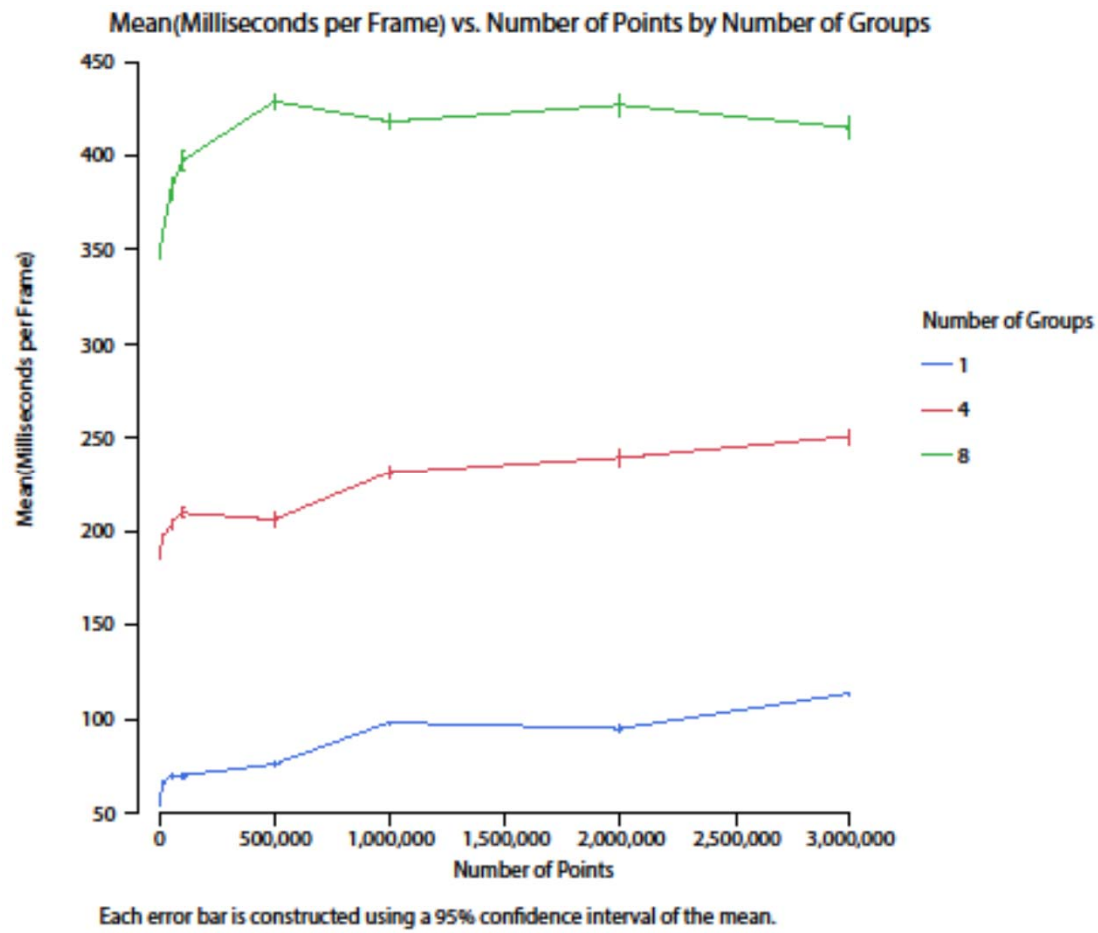
Filter (convolution) for KDE

Jump Flood for distances

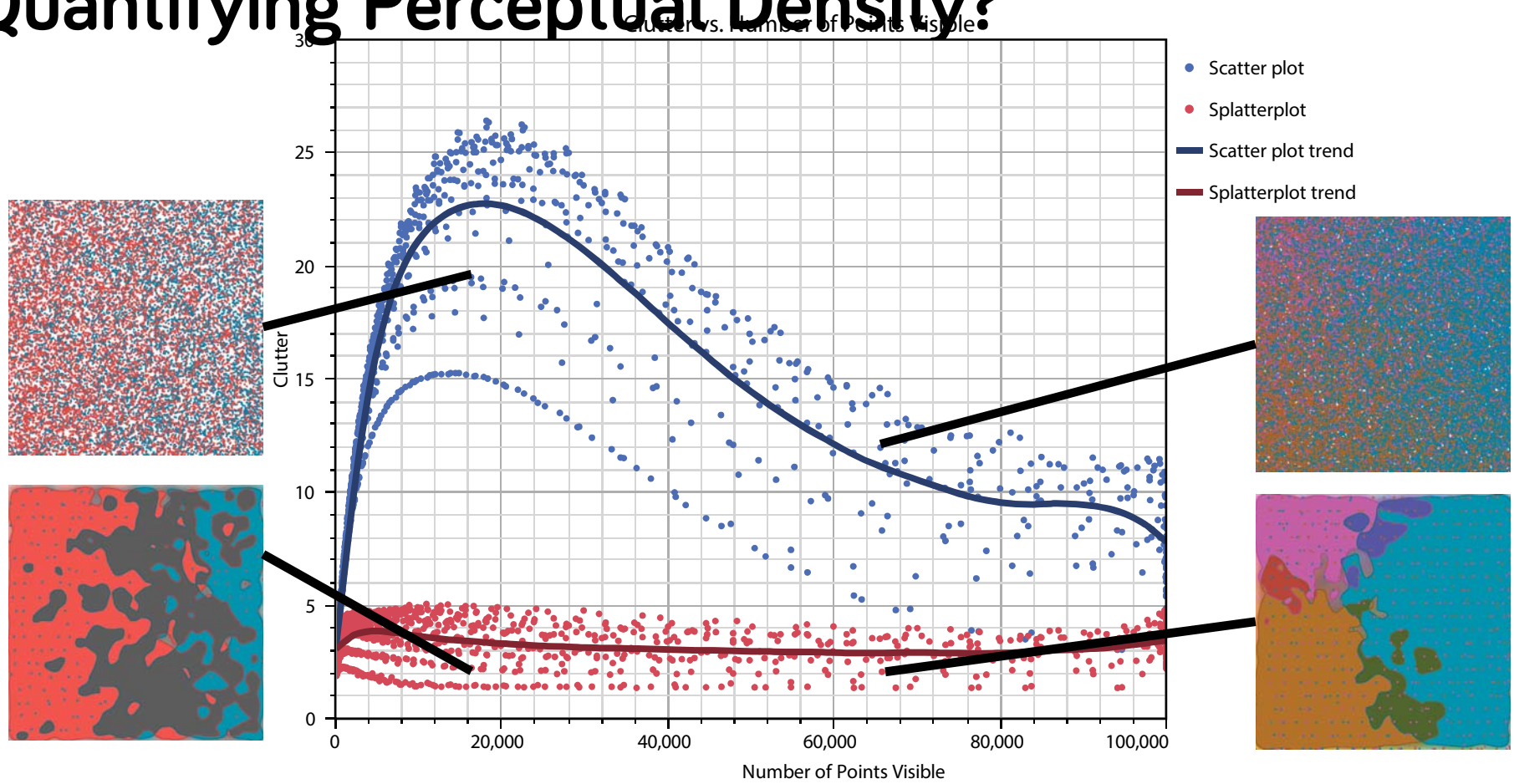
Render each set and combine

Lots of points – fast

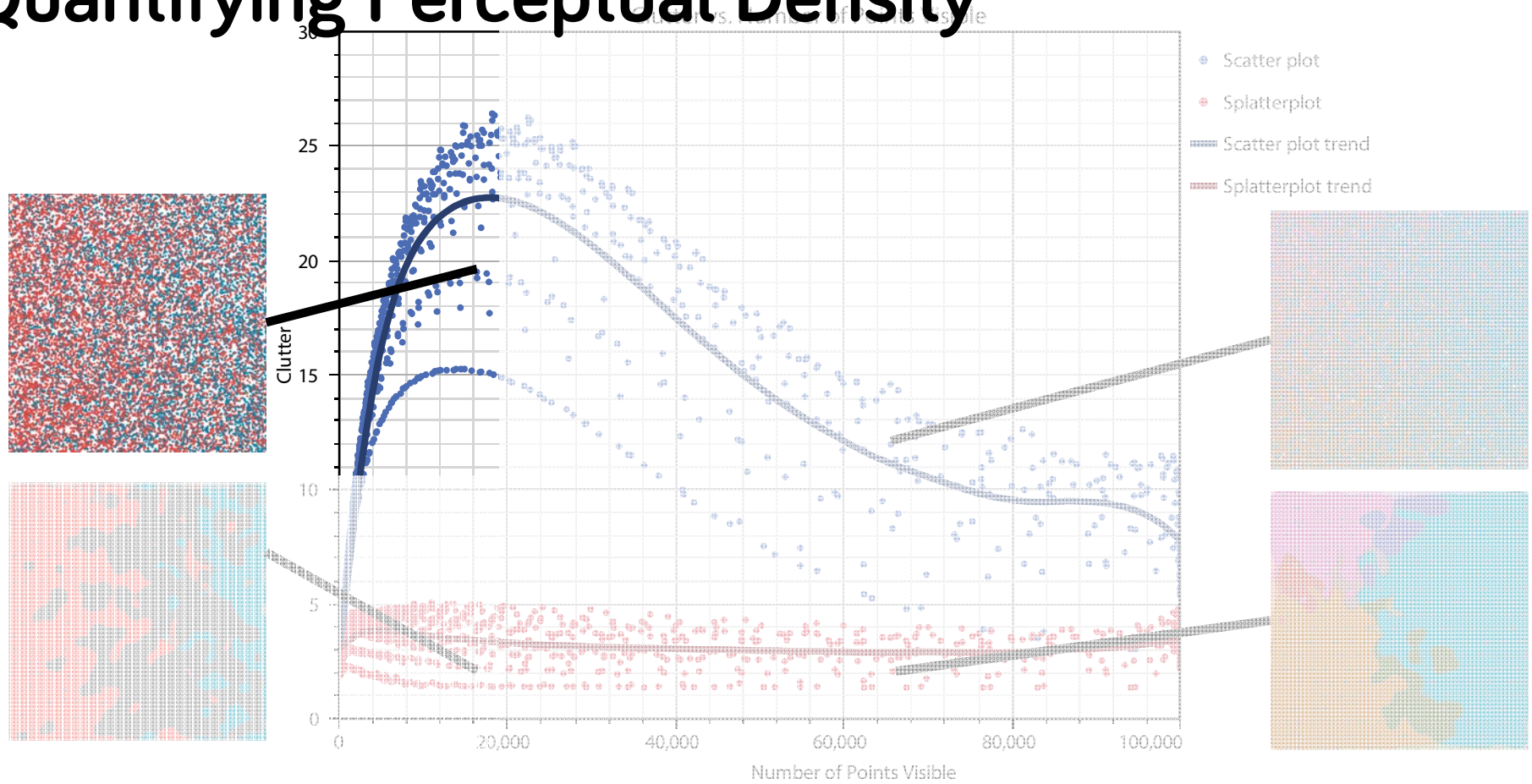
Lots of groups – less fast



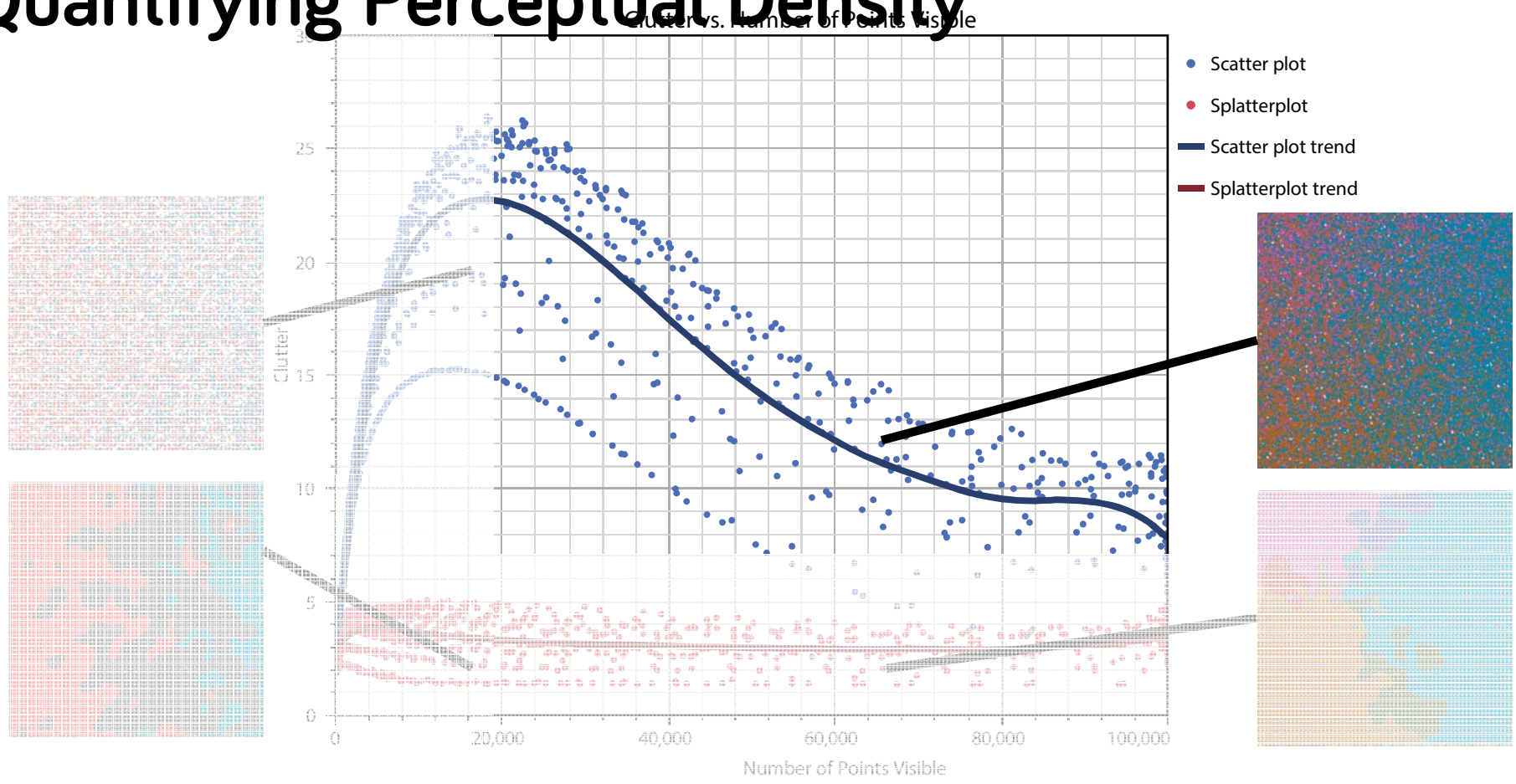
Quantifying Perceptual Density?



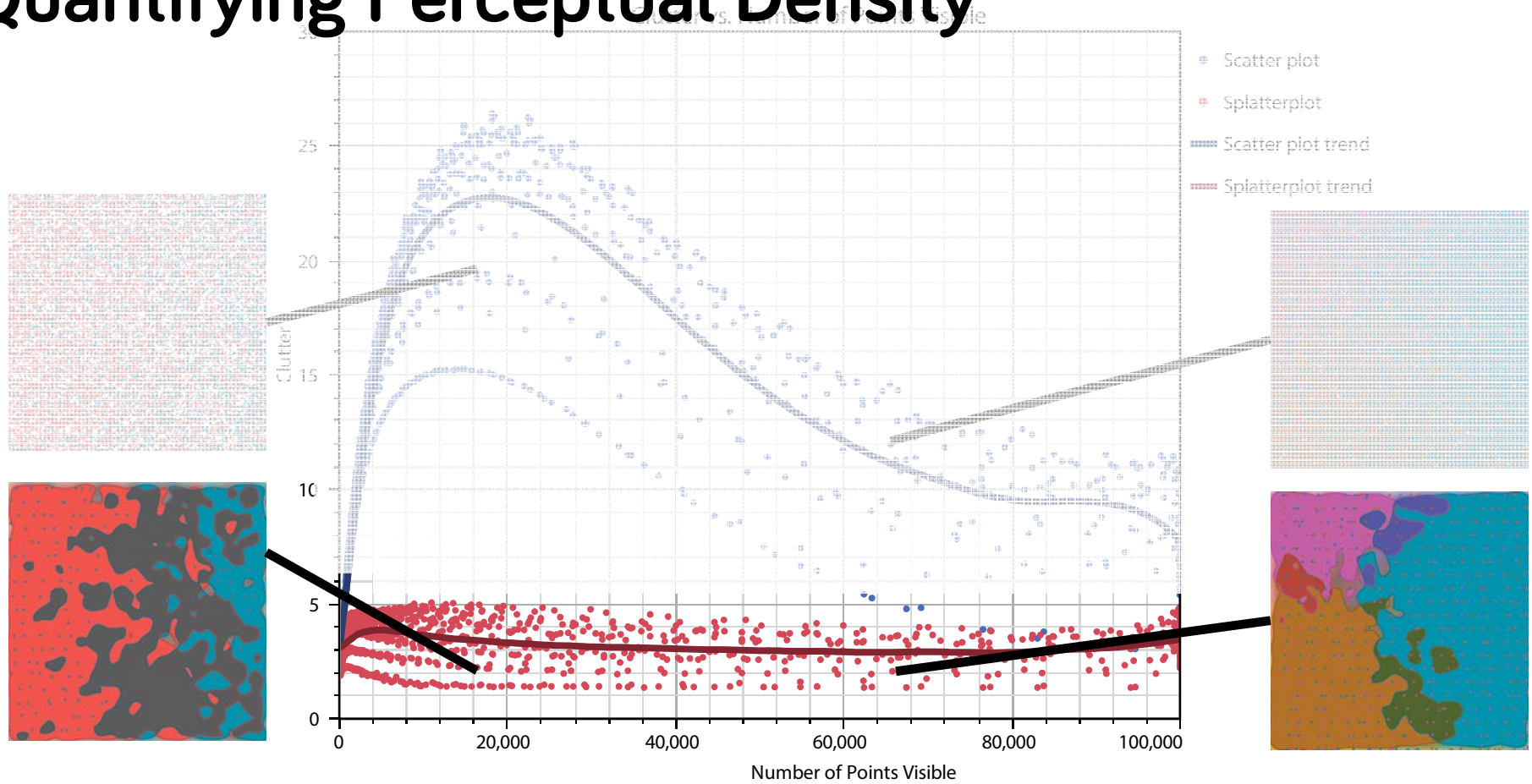
Quantifying Perceptual Density



Quantifying Perceptual Density



Quantifying Perceptual Density



Other ideas?

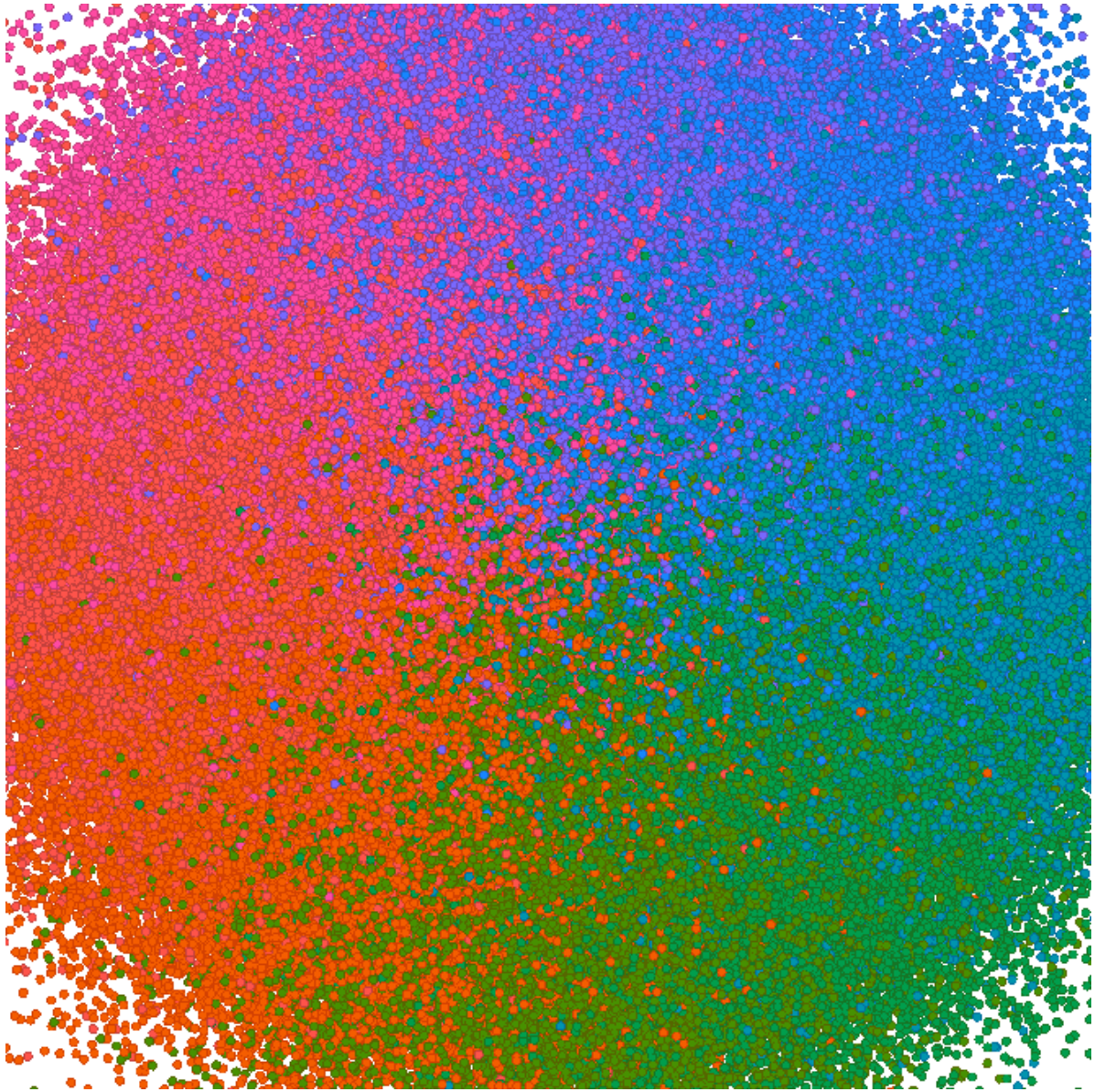
There is plenty of “related work” in research in practice

Key Novelties in Splatterplots

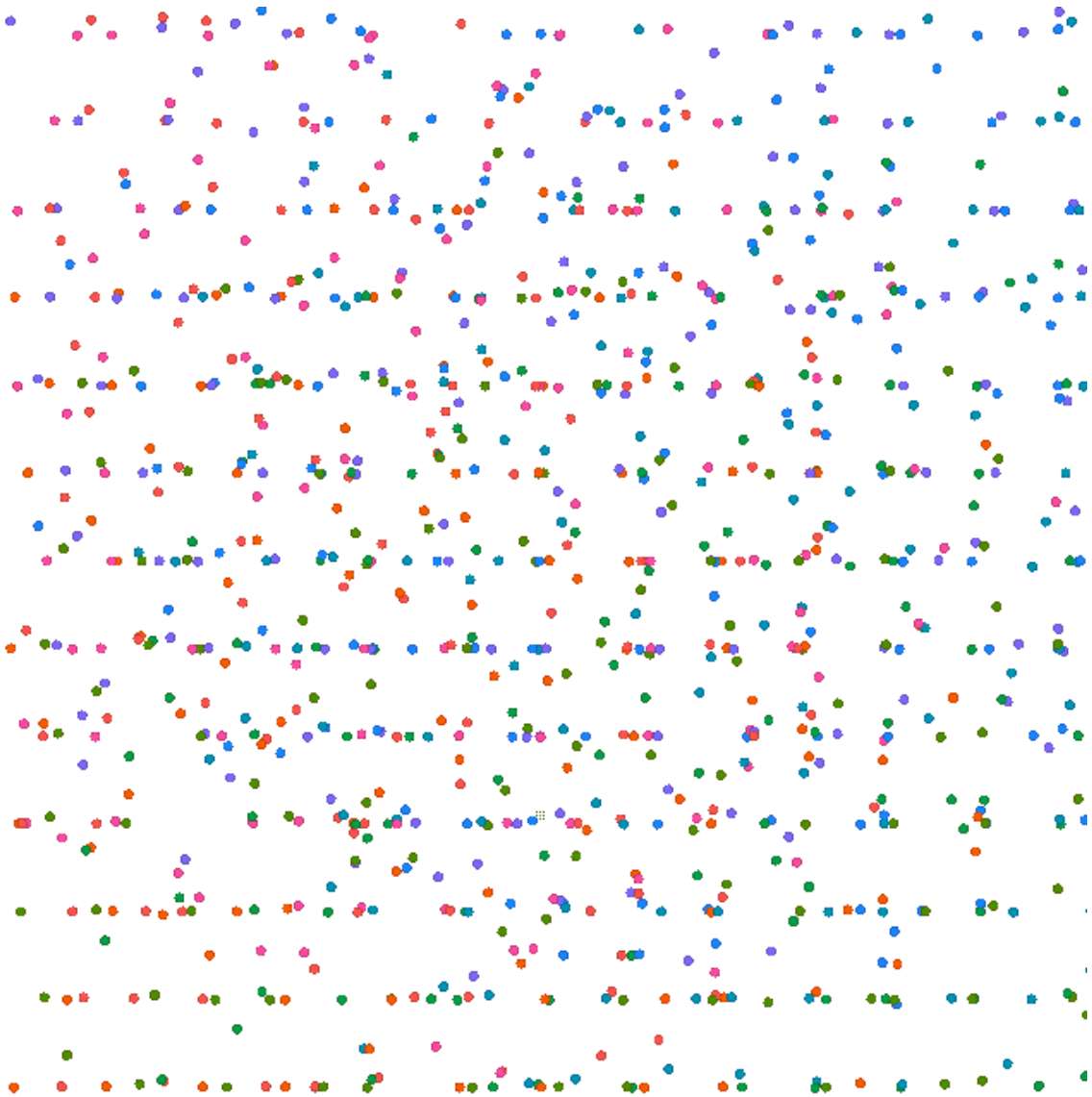
Choose abstractions to understand set relationships

Screen space density estimates

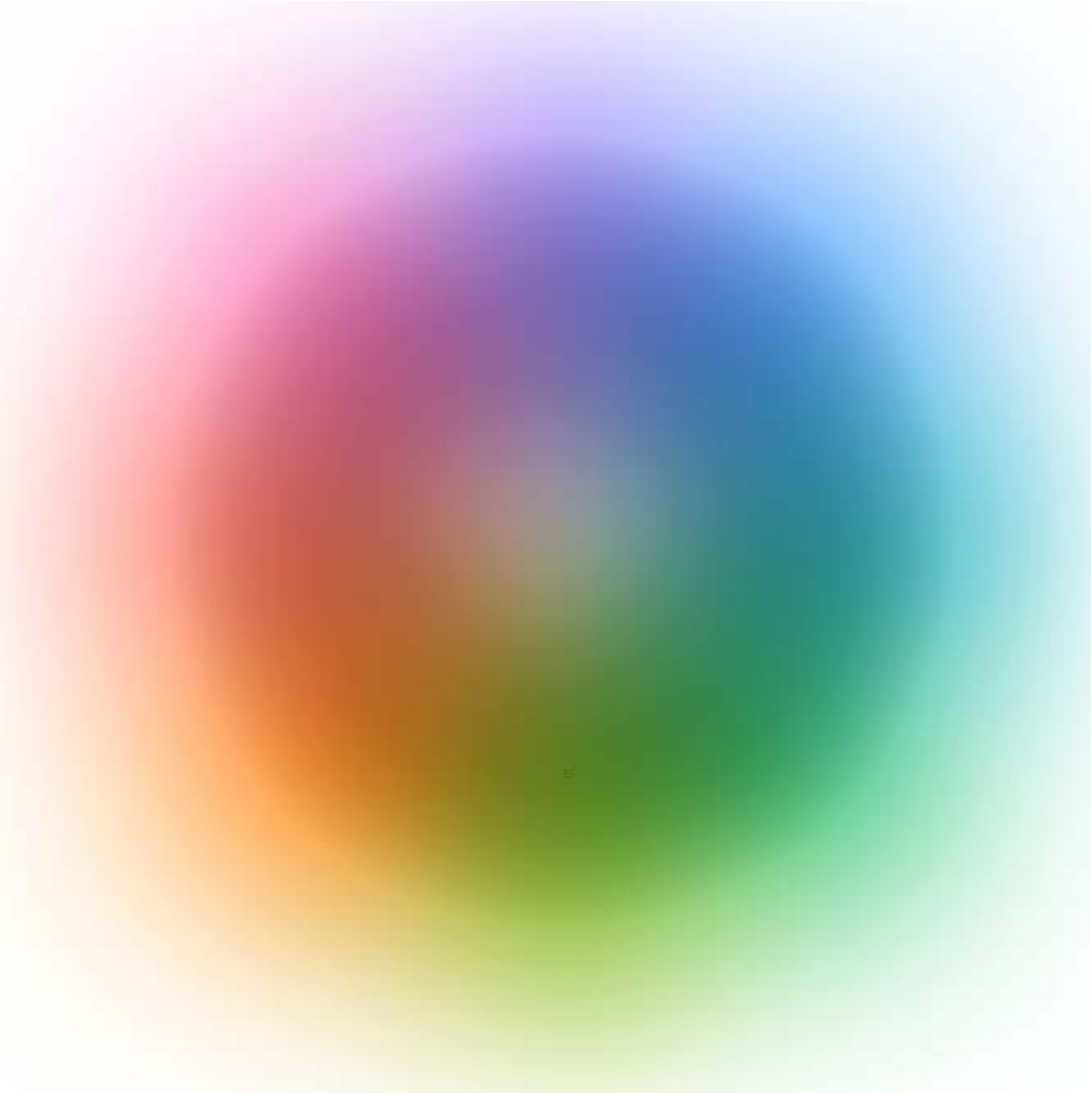
Dual Encodings



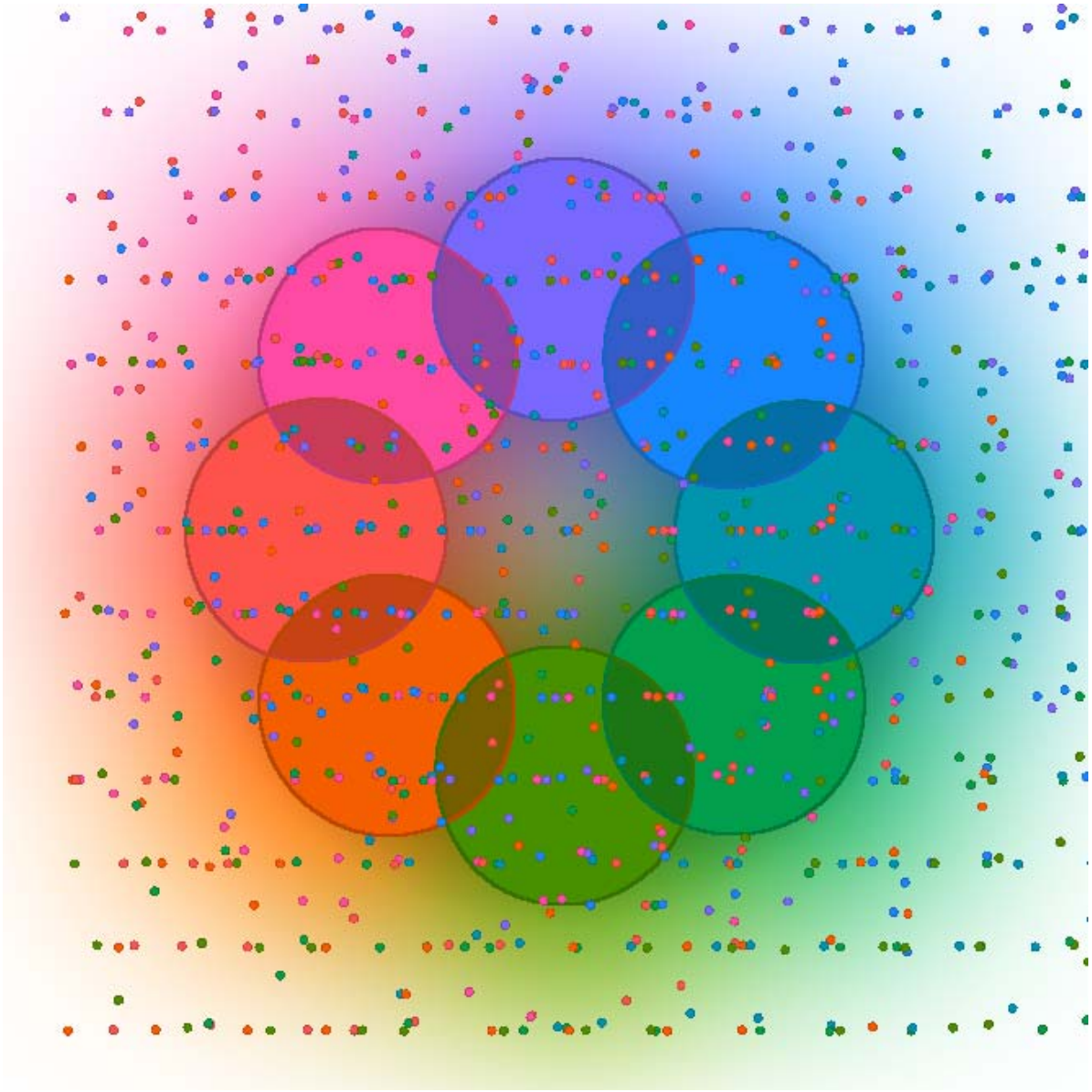
subsample?



histograms and KDEs



Splatterplot!



The synthetic data is pretty but…

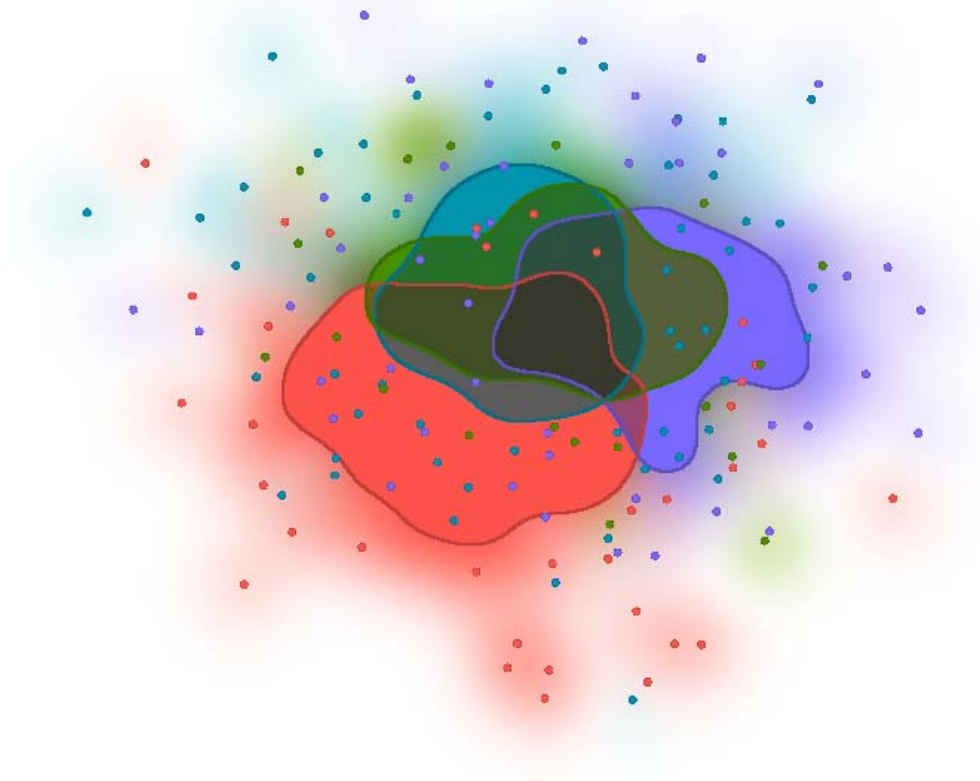
Real (or realistic) Examples

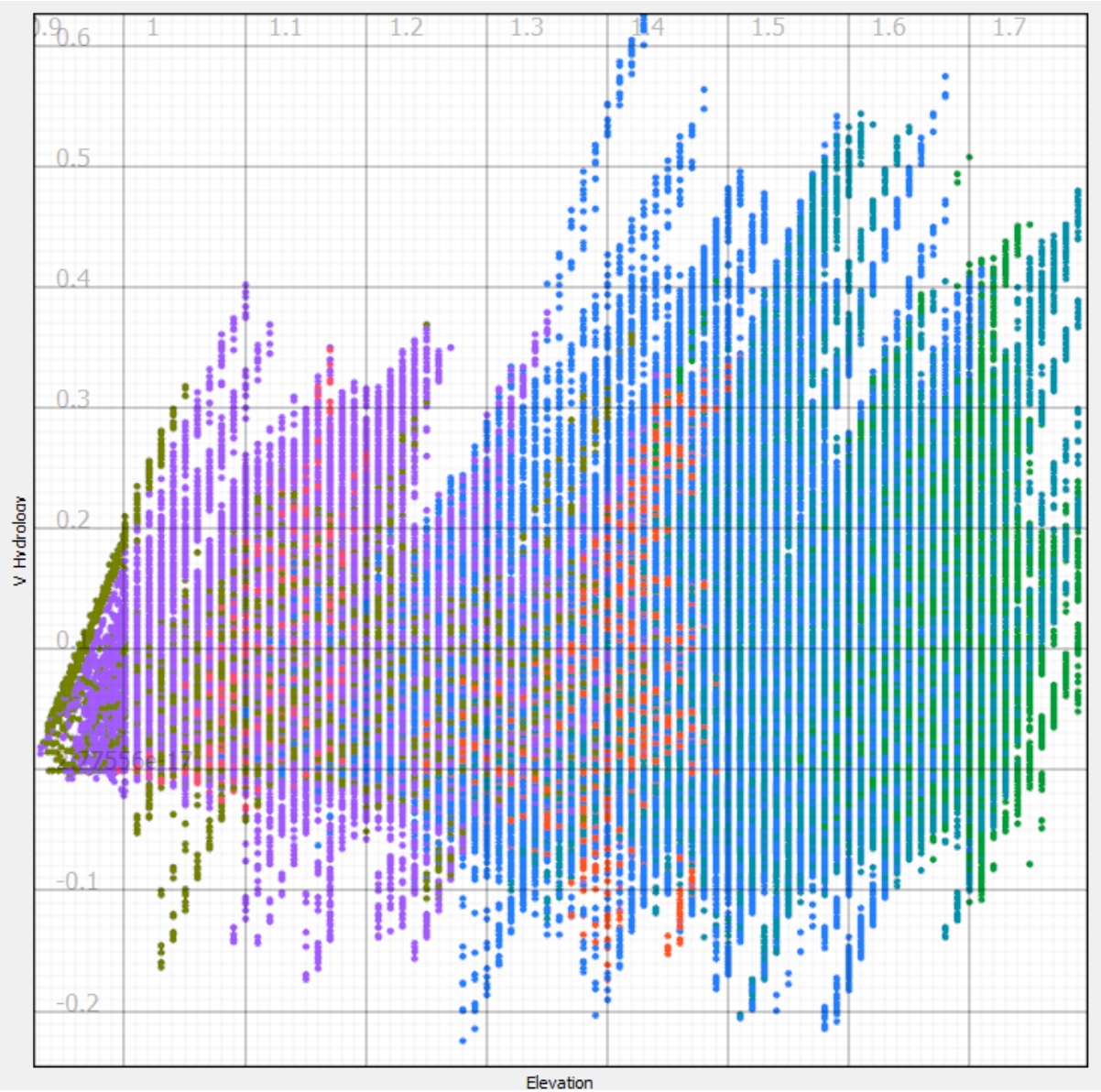
SHAKESPEARE QUARTERLY

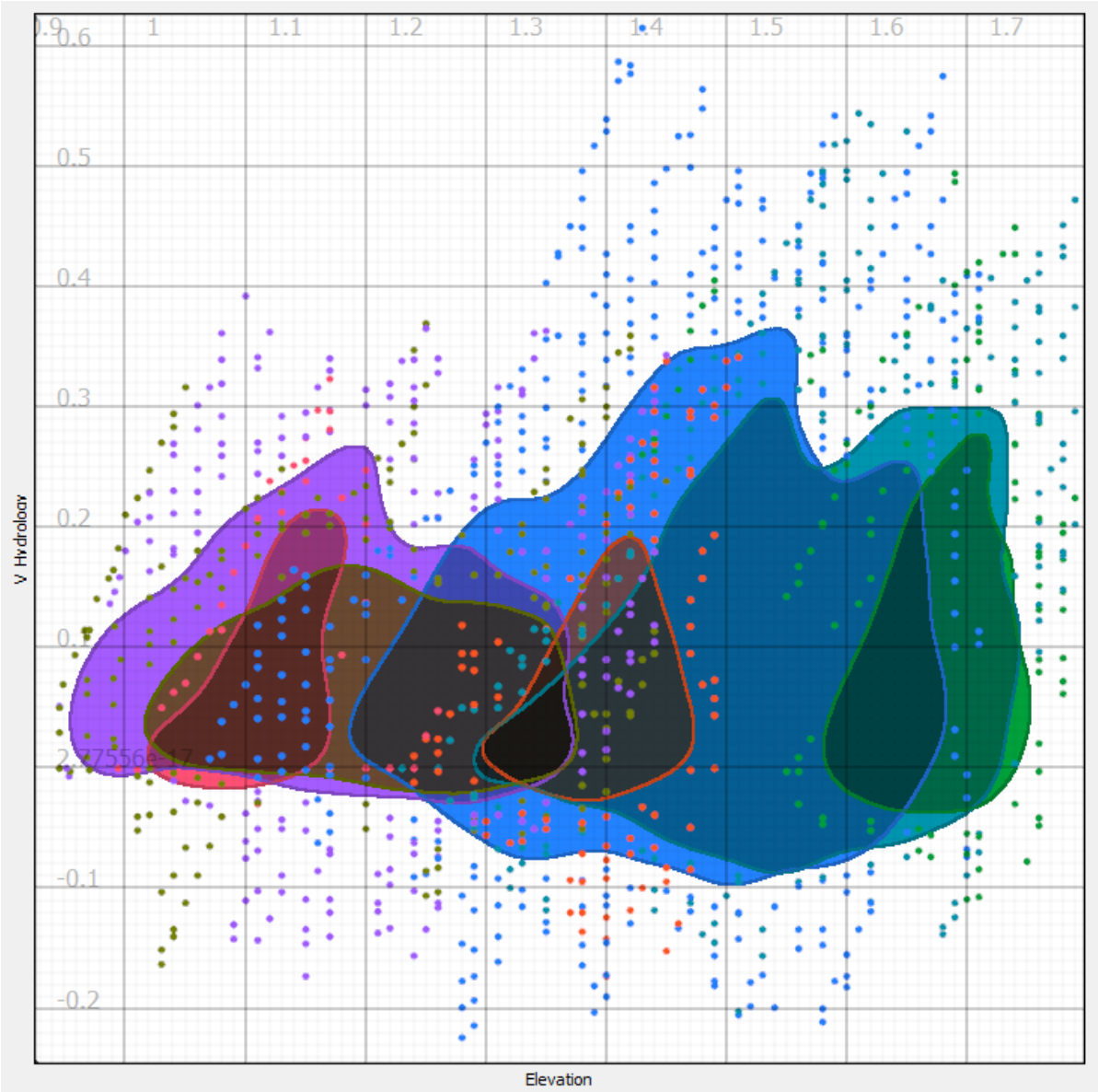
Shakespeare and New Media

Published for the Folger Shakespeare Library
in association with
The George Washington University
by The Johns Hopkins University Press

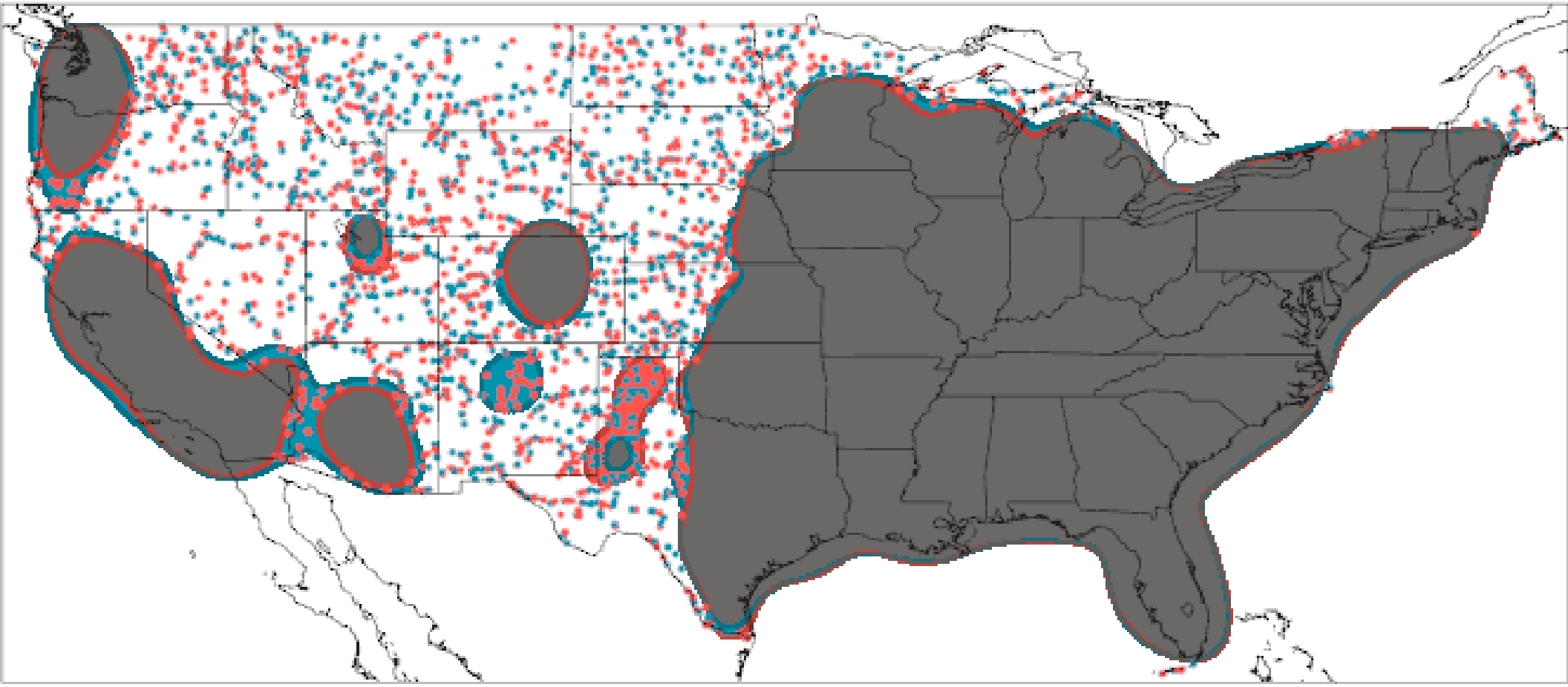
Volume 61 Fall 2010 Number 3

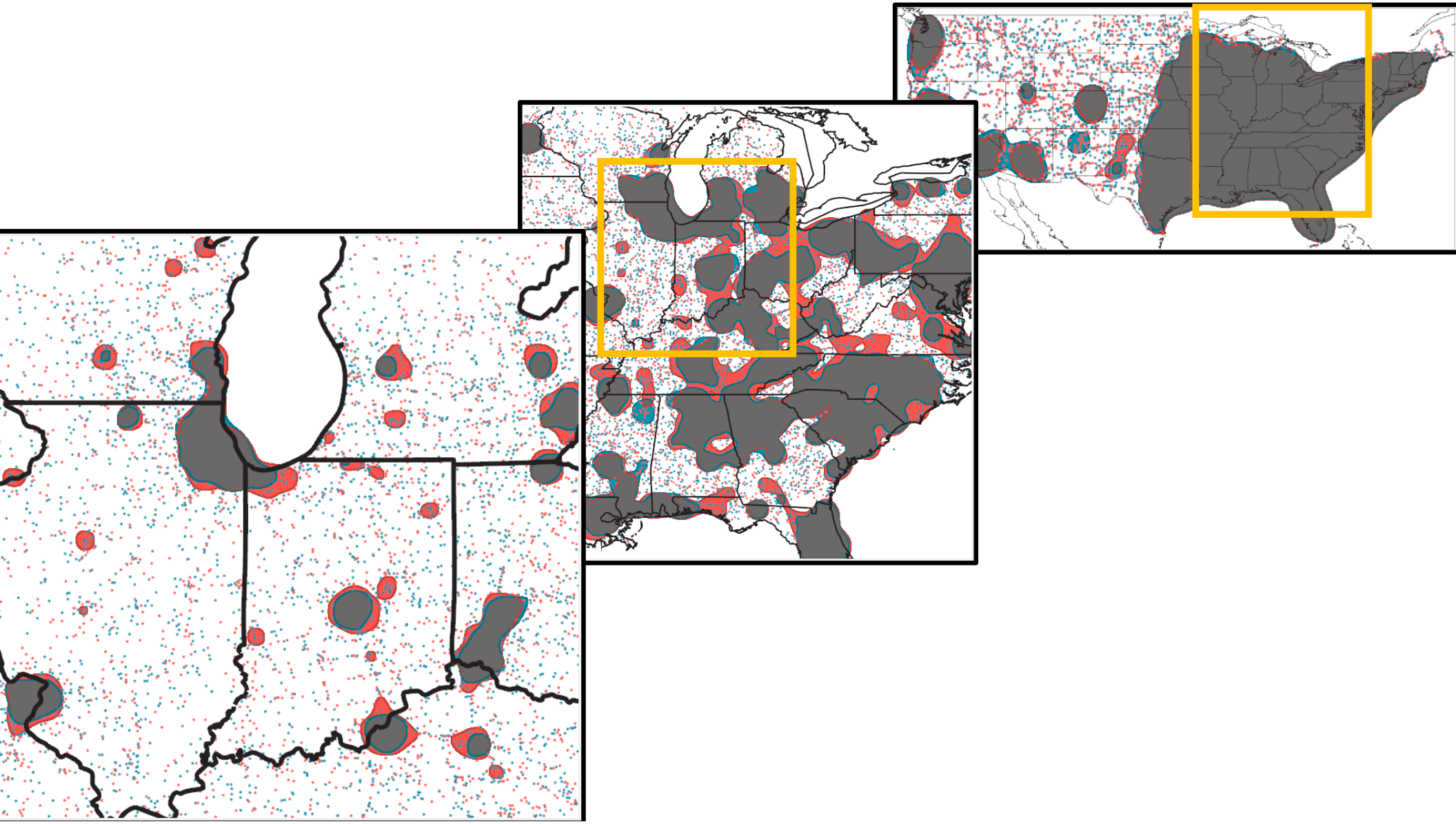



0.9
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
0.6
0.5
0.4
0.3
0.2
0.1
-0.1
-0.2
Y Hydrology
Elevation


0.9
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
0.6
0.5
0.4
0.3
0.2
0.1
-0.1
-0.2
V_Hydrology
Elevation


Aspect
Elevation
Slope
Elevation
Slope
Aspect
H_Hydrology
Elevation
Aspect
H_Hydrology
Slope
V_Hydrology
Elevation
V_Hydrology
Aspect
V_Hydrology
Slope
V_Hydrology
H_Hydrolo
H_Roadways
Elevation
H_Roadways
Aspect
H_Roadways
Slope
H_Roadways
H_Hydrolo
H_Roadways
V_Hydrolo
Shade_9am
Elevation
Shade_9am
Aspect
Shade_9am
Slope
Shade_9am
H_Hydrolo
Shade_9am
V_Hydrolo
Shade_9am
H_Roadw
Shade_Noon
Elevation
Shade_Noon
Aspect
Shade_Noon
Slope
Shade_Noon
H_Hydrolo
Shade_Noon
V_Hydrolo
Shade_Noon
H_Roadw
Shade_Noon
Shade_9a
Shade_3pm
Elevation
Shade_3pm
Aspect
Shade_3pm
Slope
Shade_3pm
H_Hydrolo
Shade_3pm
V_Hydrolo
Shade_3pm
H_Roadw
Shade_3pm
Shade_9a
Shade_3pm
Shade_No
H_Fire_Point
Elevation
H_Fire_Point
Aspect
H_Fire_Point
Slope
H_Fire_Point
H_Hydrolo
H_Fire_Point
V_Hydrolo
H_Fire_Point
H_Roadw
H_Fire_Point
Shade_9a
H_Fire_Point
Shade_No
H_Fire_Point
Shade_3p





More to do!

Theory

Understand Visual Density

Consider other tradeoffs

Other Types of Data

3D (volumes)

Practice

WebGL implementation

Massive Data Handling

Evaluation (see InfoVis paper)

Non-GPU version for my laptop ☹

Splatterplots

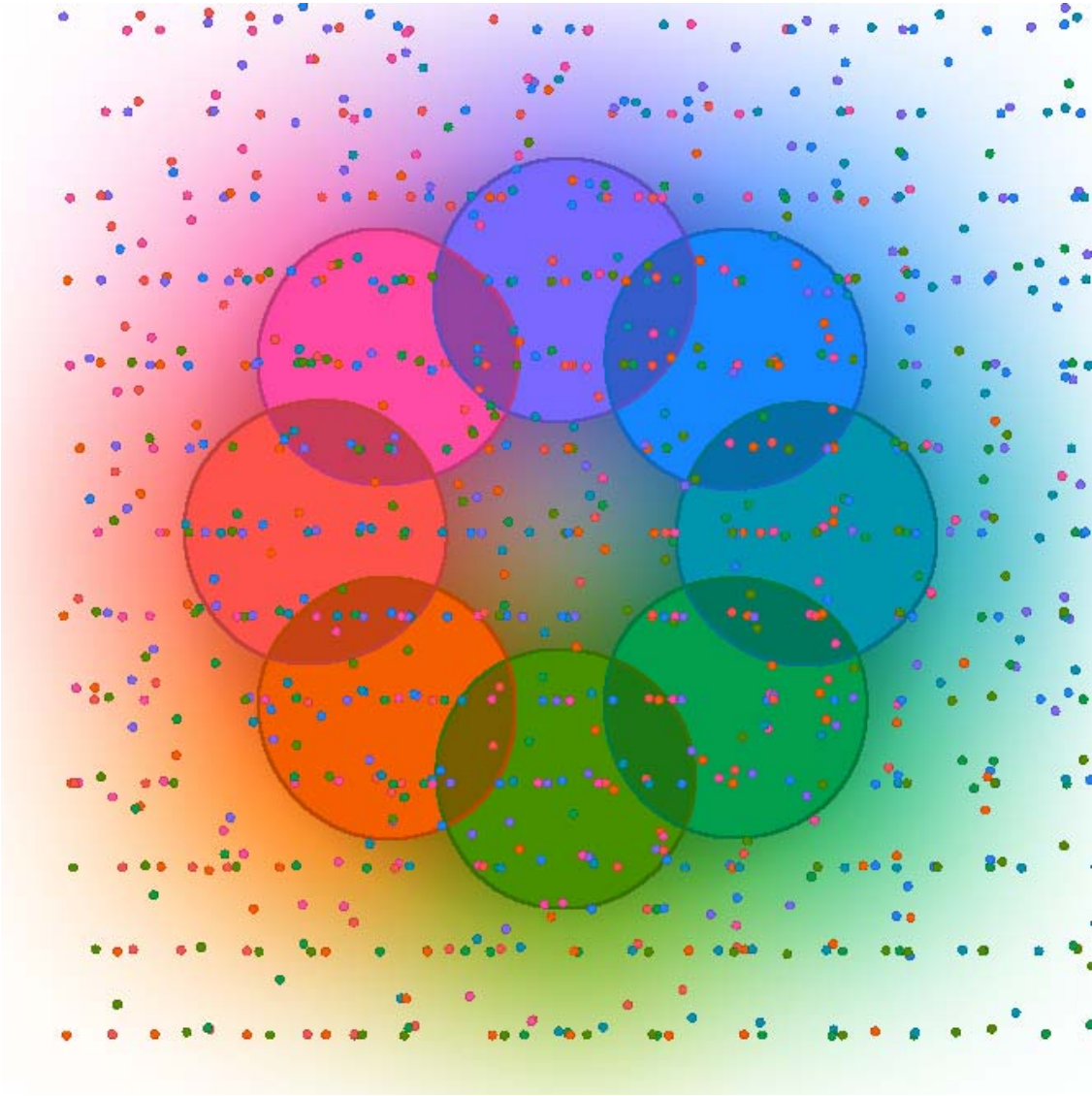
Scalable Display of Scatter Data

Bounded visual complexity

Screen space density estimation

Dual encodings

GPU Implementation



Acknowledgements

This work is supported in part by the Andrew Mellon Foundation through the "Visualizing English Print" project.

This work is supported in part by NSF Awards IIS-1162037, CMMI-094103, and DRL-1247262.