Information Visualization

Geospatial Visualization

Dr. David Koop

Writing Goals

- Write for the **audience**.
  - "Don't just write what you want to say, write what the audience needs to hear."
- Get your audience to **nod**: if the reviewer doesn't agree or wonders if you're wrong, they are less likely to like your paper
  - Avoid weasel-y words: "Some researchers think…" Who?
- Make your writing **predictable**
  - Readers are lazy
  - "You are not writing a mystery novel"
  - Upside-down pyramid writing: Important things at **start** of paper, section, paragraph

[R. Chang, 2019]

Survey Paper

- "[A]ssist the reader in the hunt for previously published research papers on a given topic"
- Full-length surveys can be 20-30 pages
- Contributions:
  - A novel classification of the literature (how your classification differs from previous surveys, or whether the survey is the first of it's kind in the field).
  - A compilation of future challenges or trends in the domain.
  - The identification of both mature and less explored research directions in the field.

[L. McNabb & R. S. Laramee]

Finding Sources

- A search can yield thousands of papers
- Use known "good" papers to locate more sources
- "Cited by…" in Google Scholar

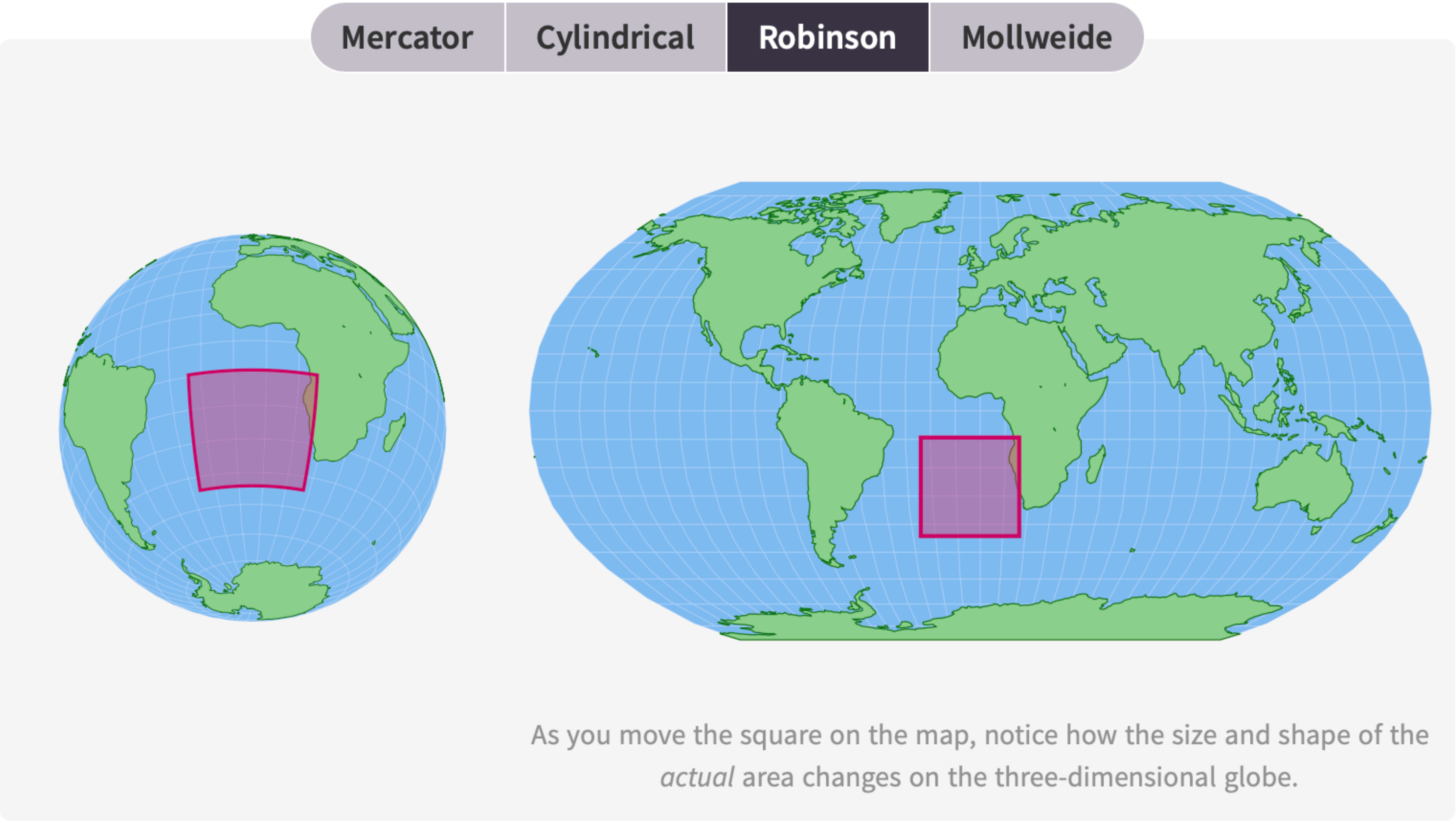
Literature Sources
Google Scholar [Goo16]
IEEE Xplore Digital Library [IEE16]
ACM Digital Library
Vispubdata [IHK*17]
The Annual EuroVis Conference
IEEE TVCG Journal
IEEE Pacific Visualization Symposium
IEEE VAST Conference
The Annual Eurographics Conference
The Eurographics Digital Library
Journal of Visual Languages & Computing
Information Visualization Journal
Computer Graphics Forum
Computer & Graphics
ACM Computing Surveys

[L. McNabb & R. S. Laramee]

Geographic Data

- Spatial data (have positions)
- Cartography: the science of drawing maps
  - Lots of history and well-established procedures
  - May also have non-spatial attributes associated with items
  - Thematic cartography: integrate these non-spatial attributes (e.g. population, life expectancy, etc.)
- Goals:
  - Respect cartographic principles
  - Understand data with geographic references with the visualization principles

Projection Distortion

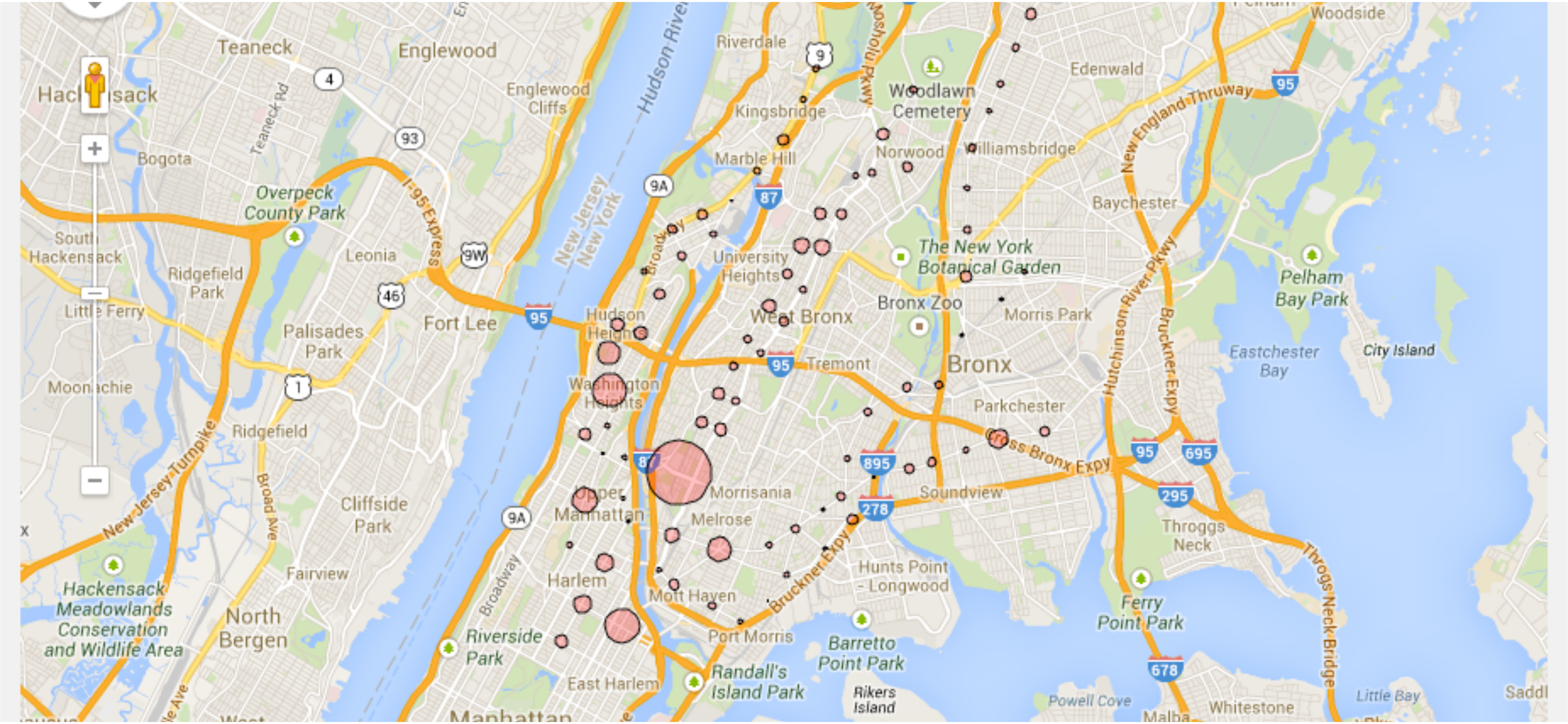


[Mathigon]

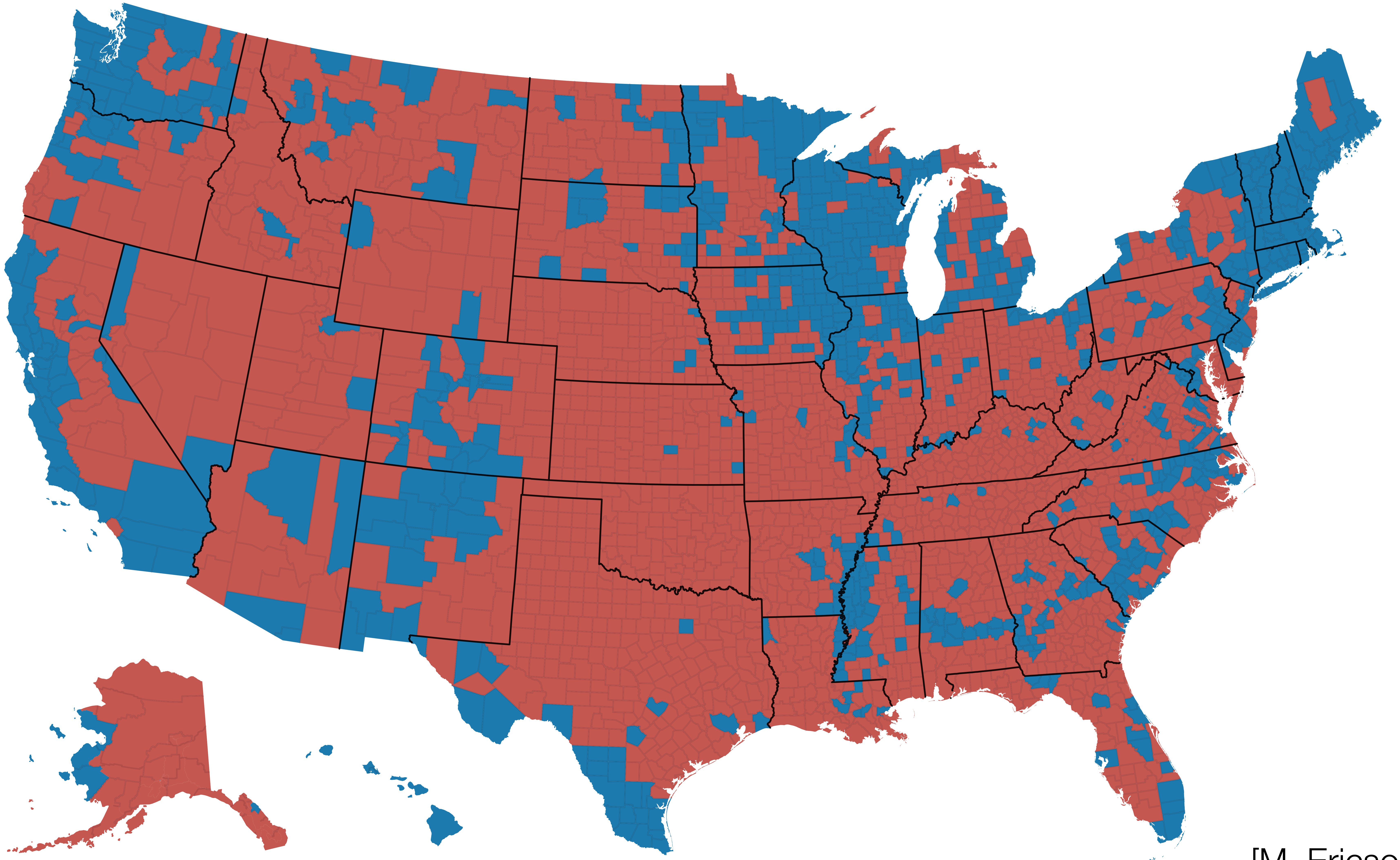
Adding Data to Maps

- Discrete: a value is associated with a specific position
  - Size
  - Color Hue
  - Charts
- Continuous: each spatial position has a value (fields)
  - Heatmap
  - Isolines

Discrete Quantitative Attribute: Size

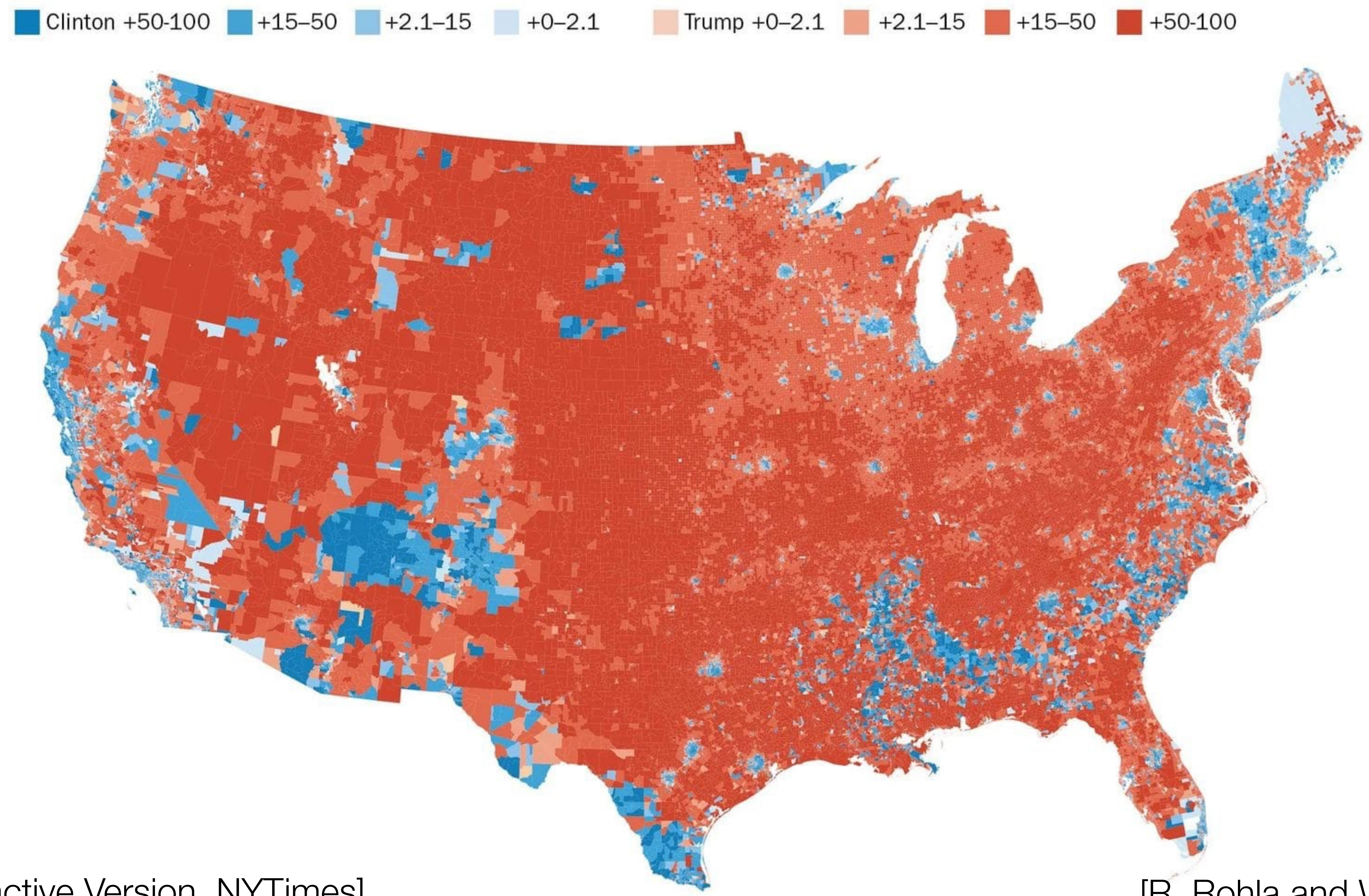


Choropleth (Two Hues)



[M. Ericson, New York Times]

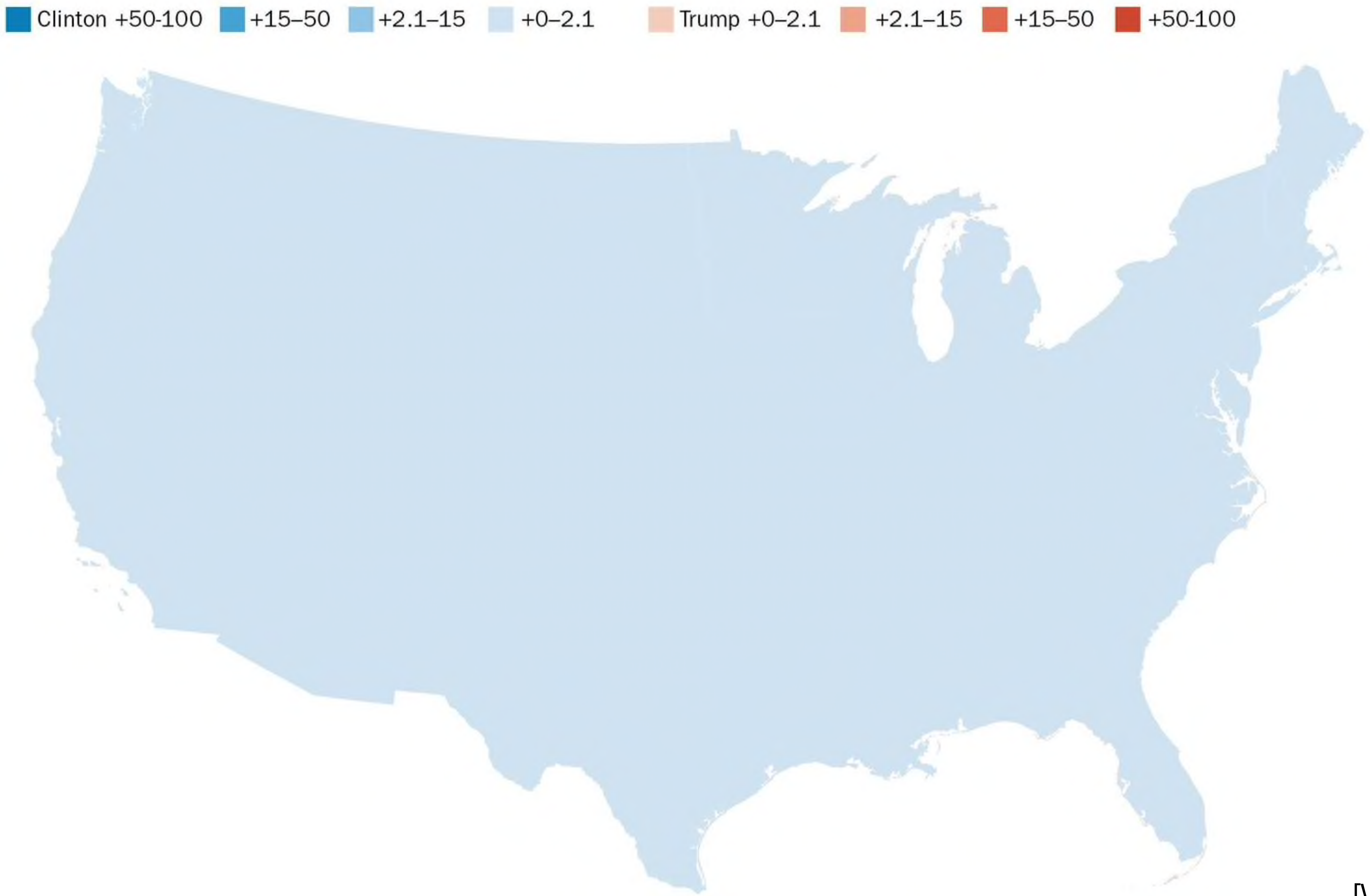
Aggregation: 2016 Election by Precinct



[Interactive Version, NYTimes]

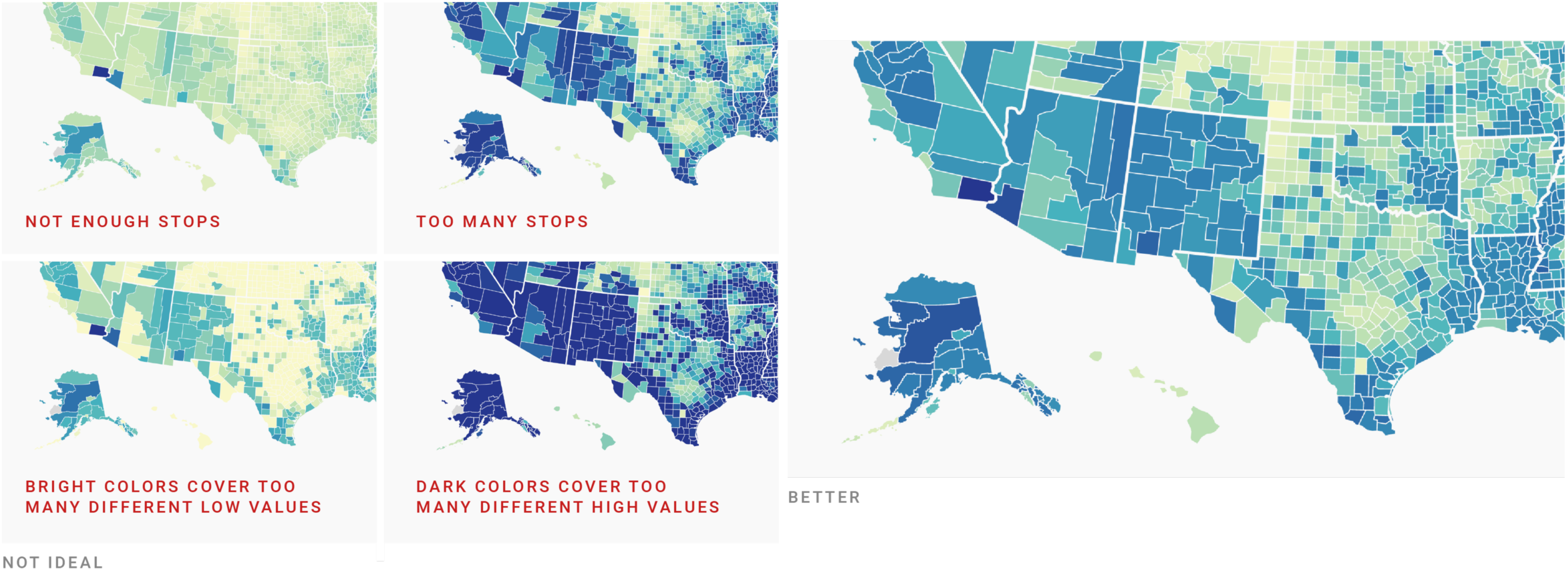
[R. Rohla and Washington Post, 2018]

Aggregation: 2016 Election by Country



[Washington Post, 2018]

When to Use Choropleth Maps



[L. C. Muth]

Paper Presentations

- Primary: Provide necessary background, present core ideas, step through techniques, discuss experiments and results
  - Channel the original authors as much as possible
- Secondary: Provide critique: what is problematic, what could be improved, where could techniques be extended
  - Channel the reviewers as much as possible
- Everyone: Read the paper, come with questions and discussion points

Annotated Bibliography

- Likely related to your project, but can be another subject area
- Wider breadth than just the related work of your project
- Find 30-40 references, and write a few sentences on how they relate to your work/ideas
  - Ok to include papers that show novel variations of a technique, even if the paper is not mostly about the subject area!
  - Your annotations are not the abstract of the paper, include relationship with the subject area you're focusing on
- Due next Thursday

Paper

A Declarative Rendering Model for Multiclass Density Maps

Jaemin Jo, Frédéric Vernier, Pierre Dragicevic, and Jean-Daniel Fekete, *Senior Member, IEEE*

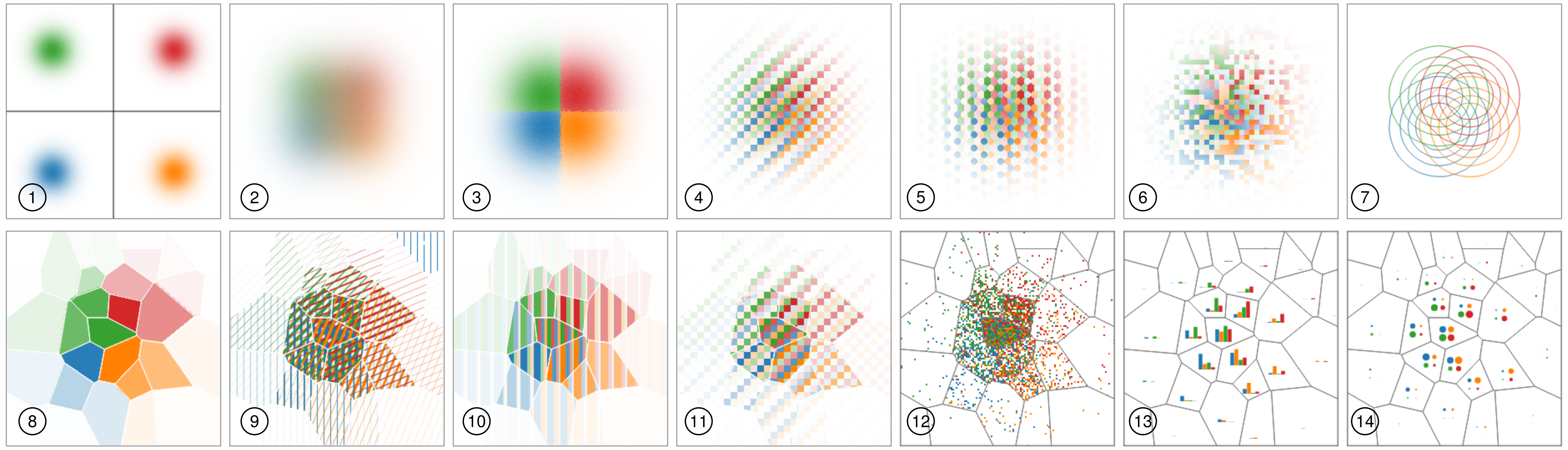


Fig. 1: Design alternatives for a four-class density map. ① shows small multiples where each density map is individually presented with a unique color; ② stacks the density maps and blends the color at each pixel; ③ shows the color of the pixel with the highest density; ④-⑥ use regular and irregular weaving patterns; ⑦ shows a contour plot for each class; and ⑧–⑭ use rebinning (binning and aggregation over the density maps) with tiles produced by a random Voronoï tessellation. The aggregated values are rendered in ⑧ with a flat color showing the highest density, ⑨ with hatching, ⑩ with proportional bars, ⑪ with regular weaving, ⑫ with a dot density plot, ⑬ with bar-chart glyphs, and ⑭ with circle sizes.

Abstract—Multiclass maps are scatterplots, multidimensional projections, or thematic geographic maps where data points have a categorical attribute in addition to two quantitative attributes. This categorical attribute is often rendered using shape or color, which does not scale when overplotting occurs. When the number of data points increases, multiclass maps must resort to data aggregation to remain readable. We present *multiclass density maps*: multiple 2D histograms computed for each of the category values. Multiclass density maps are meant as a building block to improve the expressiveness and scalability of multiclass map visualization. In this article, we first present a short survey of aggregated multiclass maps, mainly from cartography. We then introduce a declarative model—a simple yet expressive JSON grammar associated with visual semantics—that specifies a wide design space of visualizations for multiclass density maps. Our declarative model is expressive and can be efficiently implemented in visualization front-ends such as modern web browsers. Furthermore, it can be reconfigured dynamically to support data exploration tasks without recomputing the raw data. Finally, we demonstrate how our model can be used to reproduce examples from the past and support exploring data at scale.

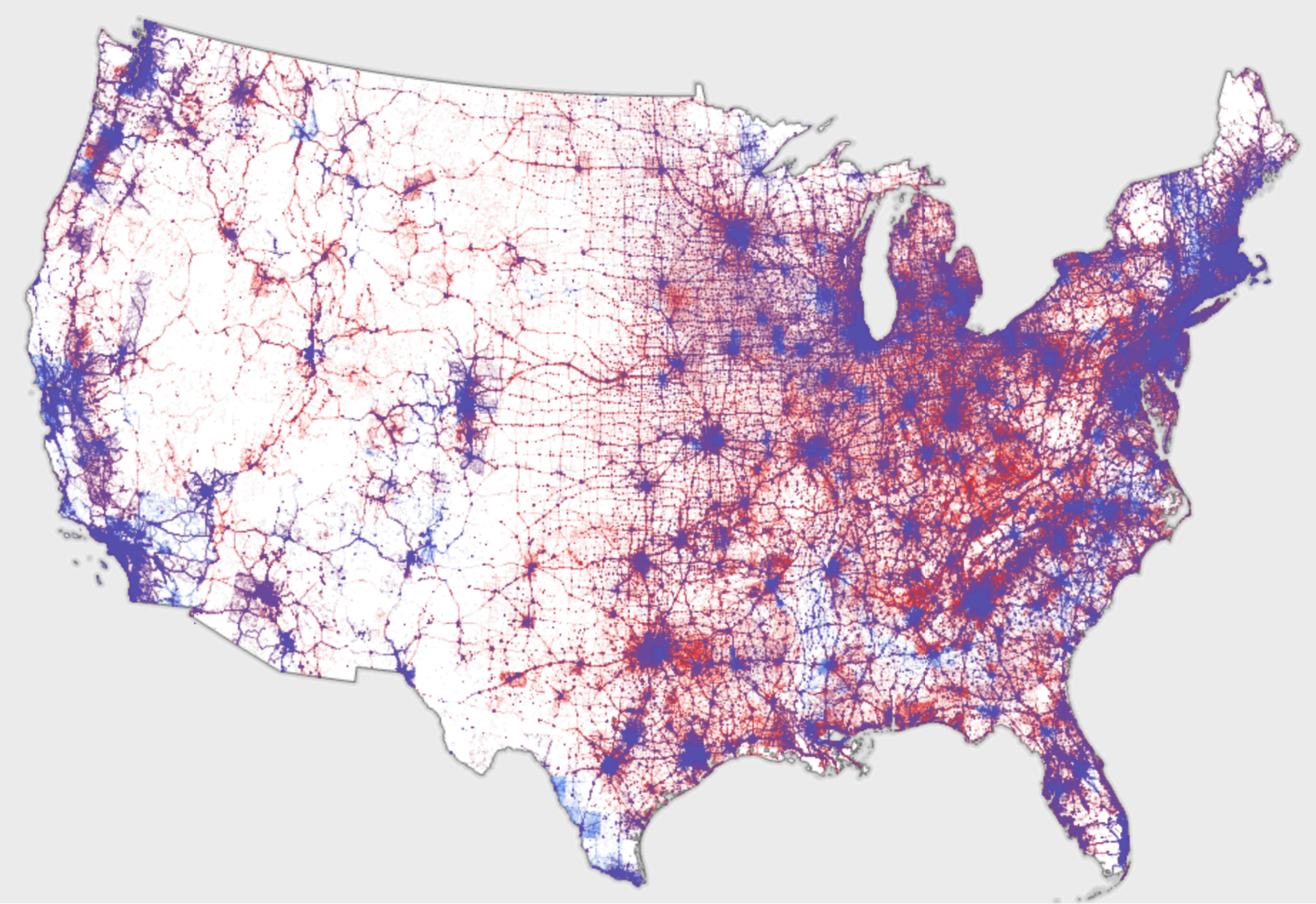
Index Terms—Scalability, multiclass scatterplots, density maps, aggregation, declarative specification, visualization grammar

What is a Multi-class Map?

- Data:
  - 2 quantitative attributes mapped to x-position and y-position
  - 1 categorical attribute
- Can be:
  - (1) scatterplots,
  - (2) multidimensional projections,
  - (3) thematic geography (maps!)

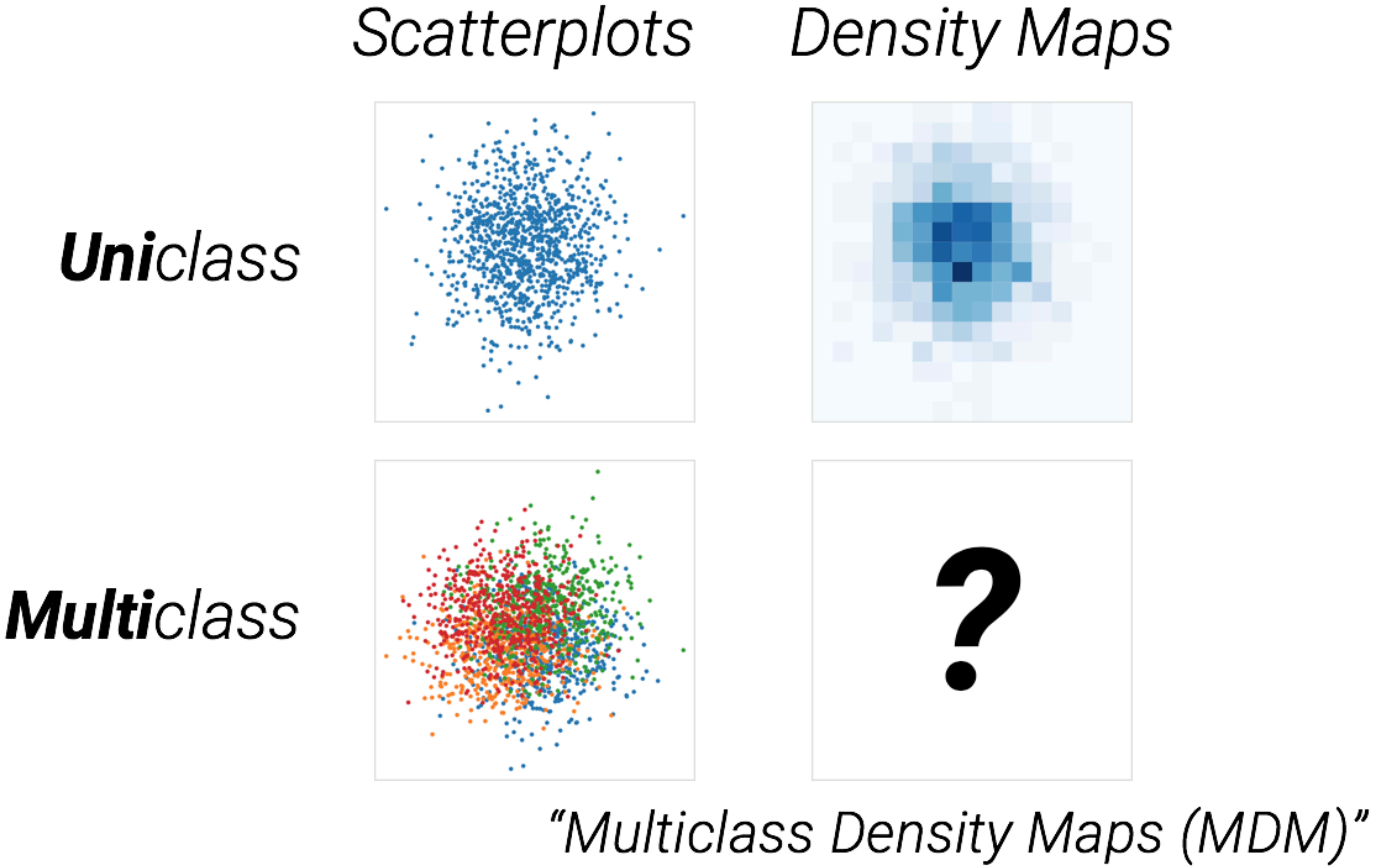
[J. Jo et al.]

Dasymetric Dot Density



[K. Field]

Multiclass Density Maps (MDMs)

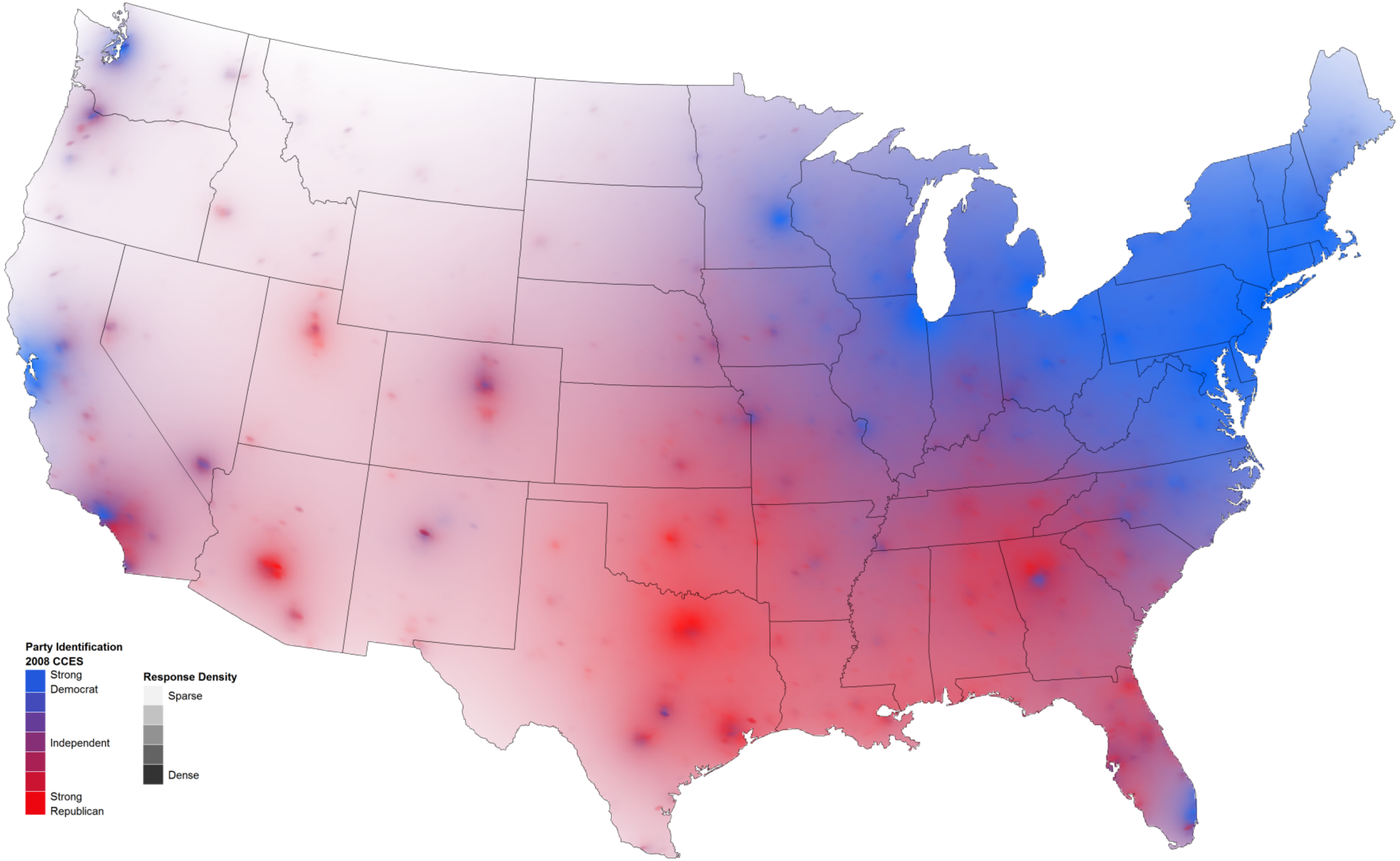


[J. Jo et al.]

What is a Density Map?

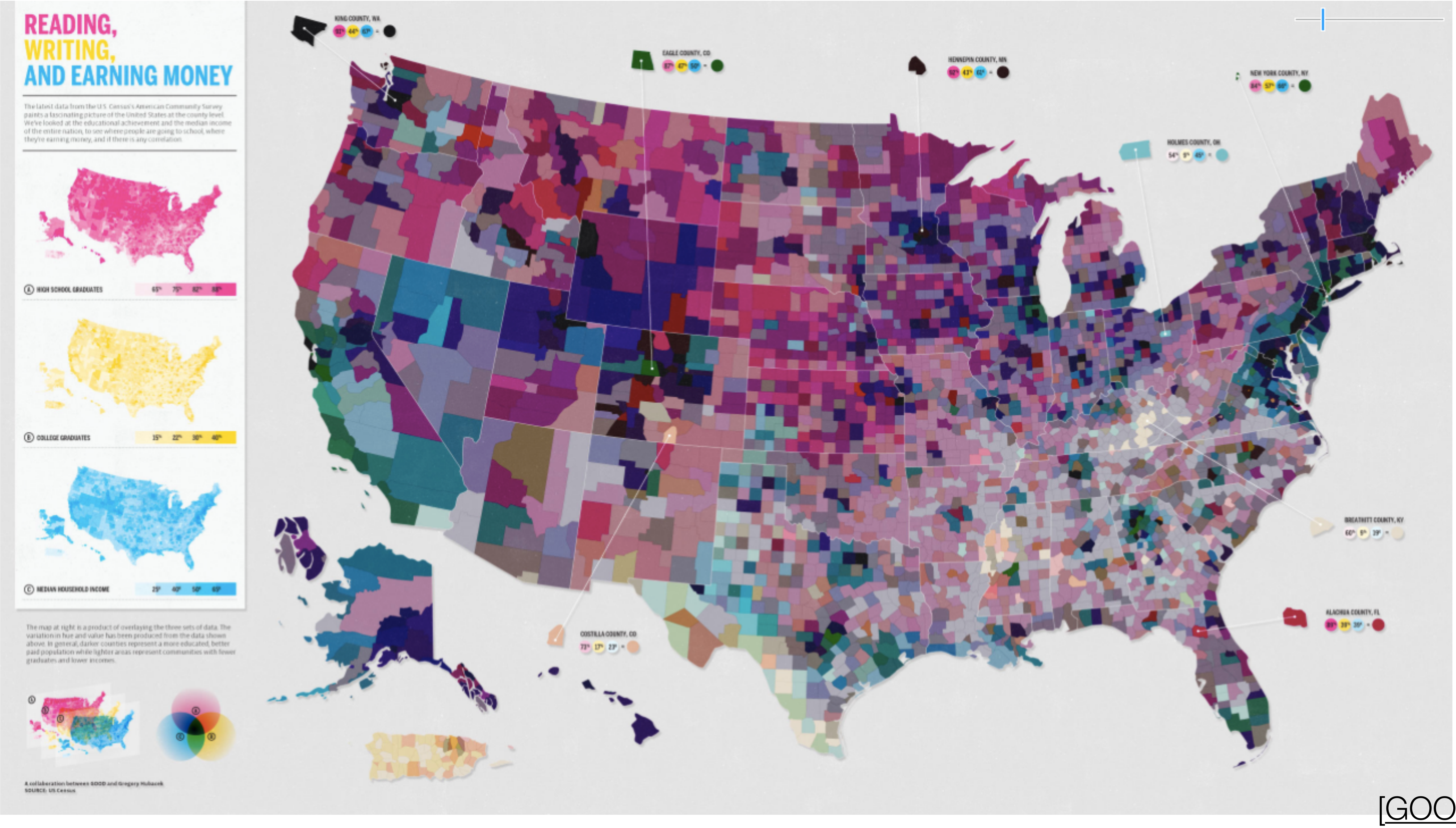
- Problem: too many points
- Solution: Aggregate!
  - In 2D: 2D histogram

Multiclass Density Map "Examples"



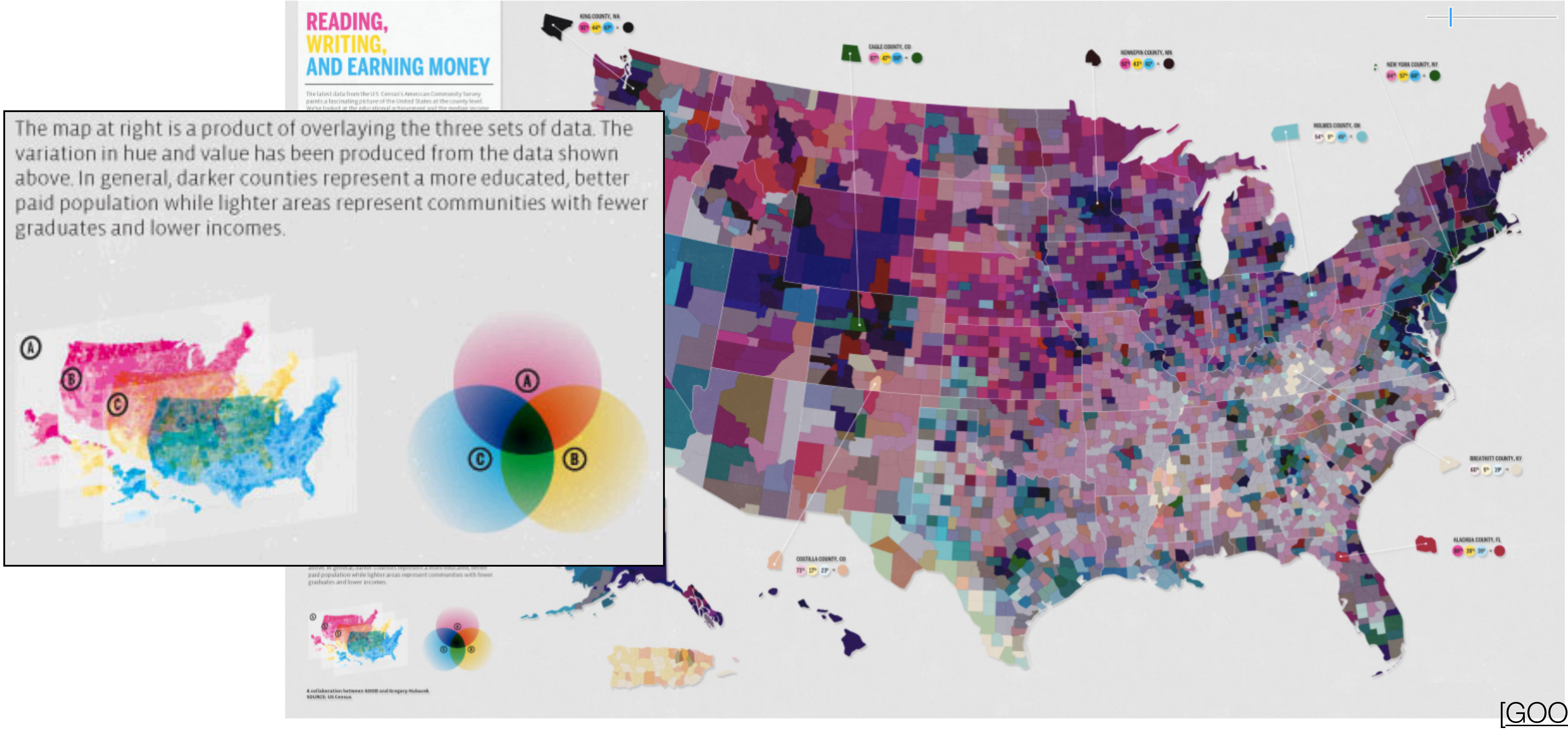
[D. B. Sparks]

Multiclass Density Map "Examples"



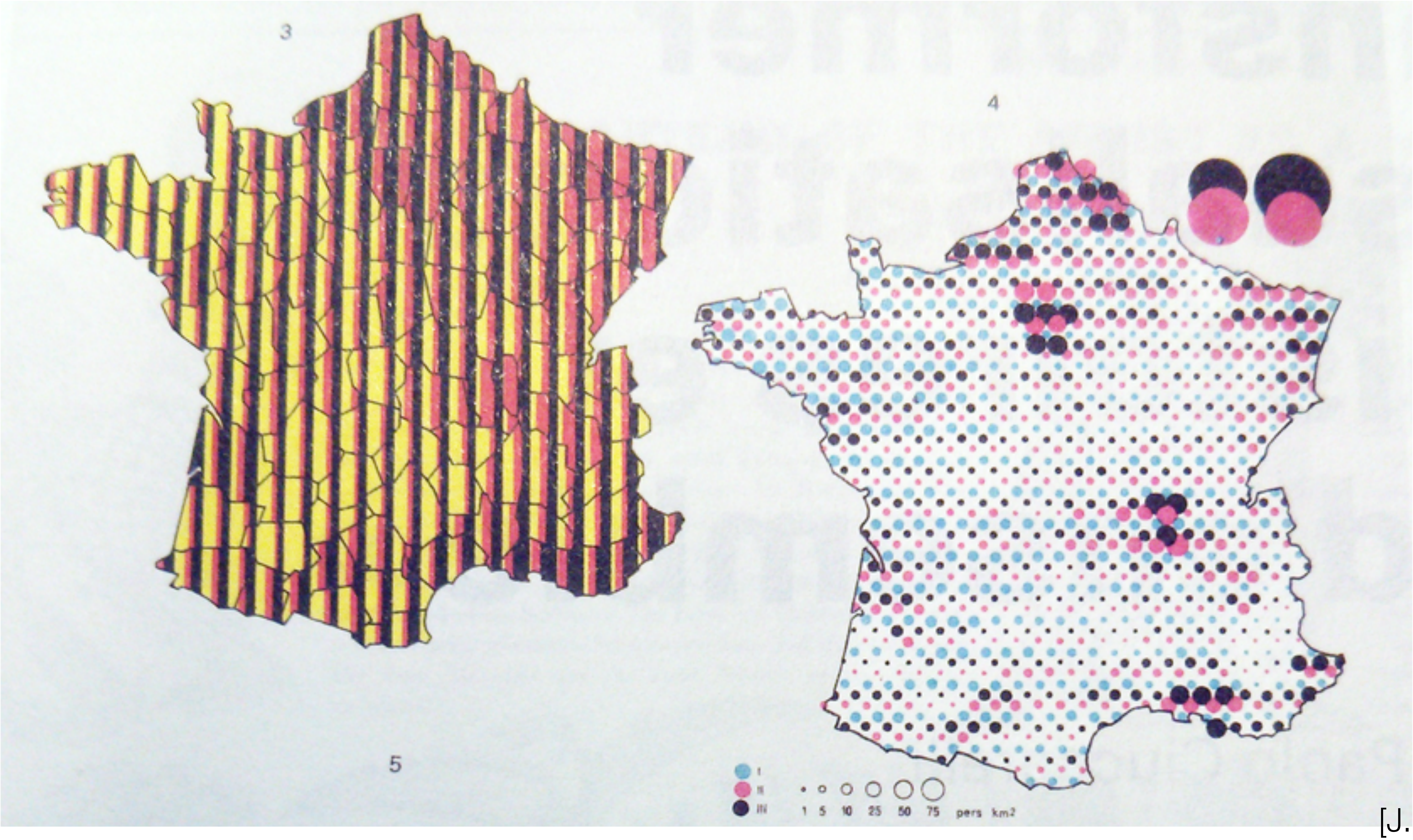
[GOOD]

Multiclass Density Map "Examples"



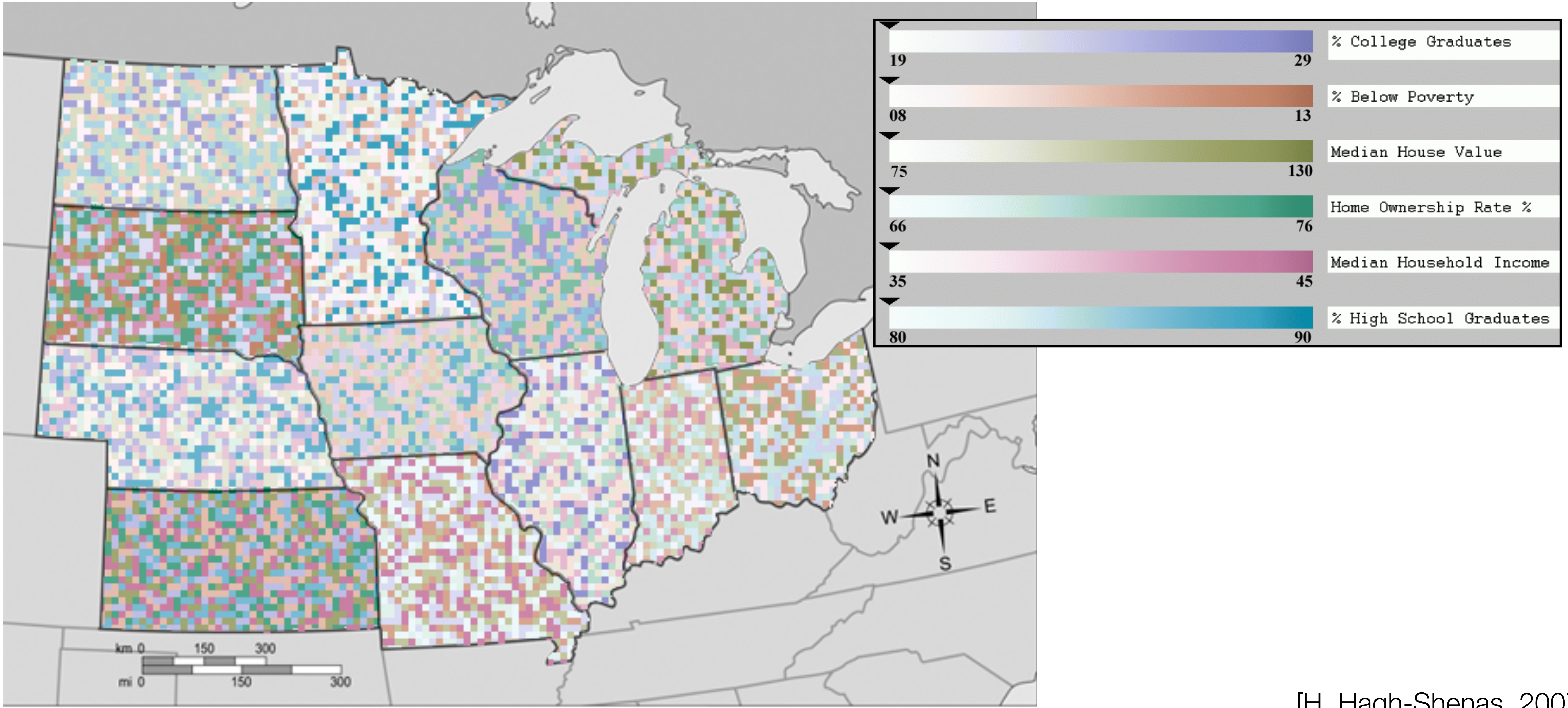
[GOOD]

Multiclass Density Map "Examples"



[J. Bertin via ICACI]

Multiclass Density Map "Examples"



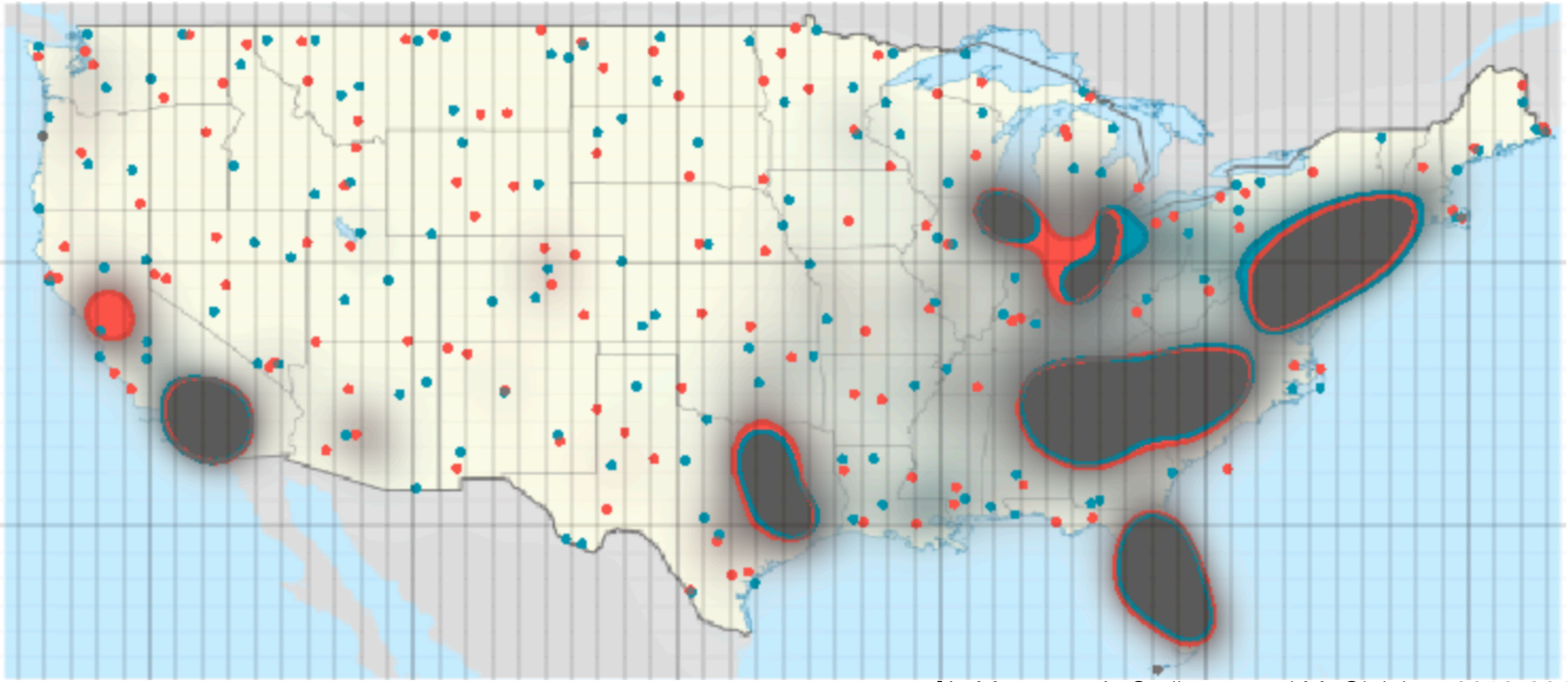
[H. Hagh-Shenas, 2007]

Scalability

- **data size** related to the number of data points and categories,
  - 100 million points or more with tens of categories
- **perceptual processing** related to the ability to perform some tasks efficiently given a data size, and
- **computation speed** related to the time to compute an image from a visualization technique given a data size
  - refresh rate between 25ms to 10s, depending on interaction type

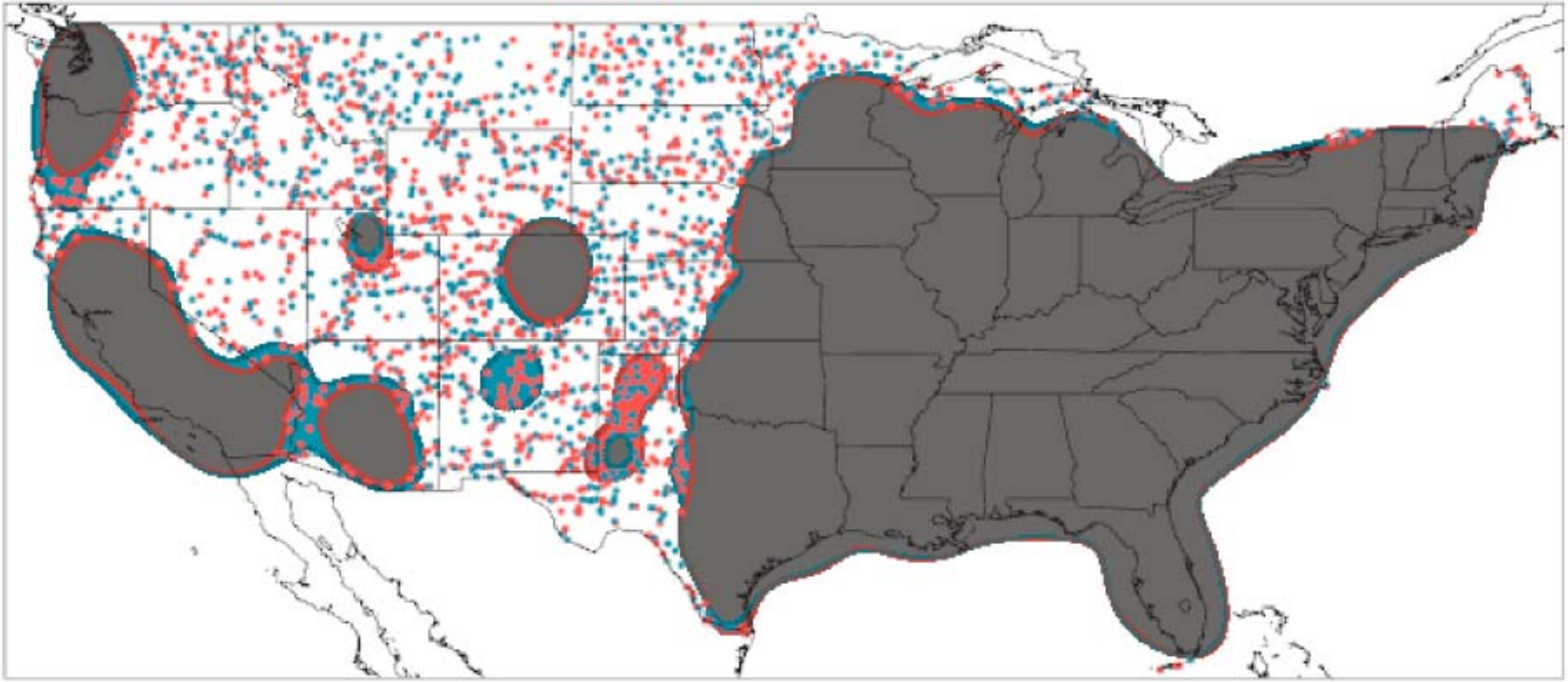
[J. Jo et al.]

Splatterplots: Scatterplot Summarization



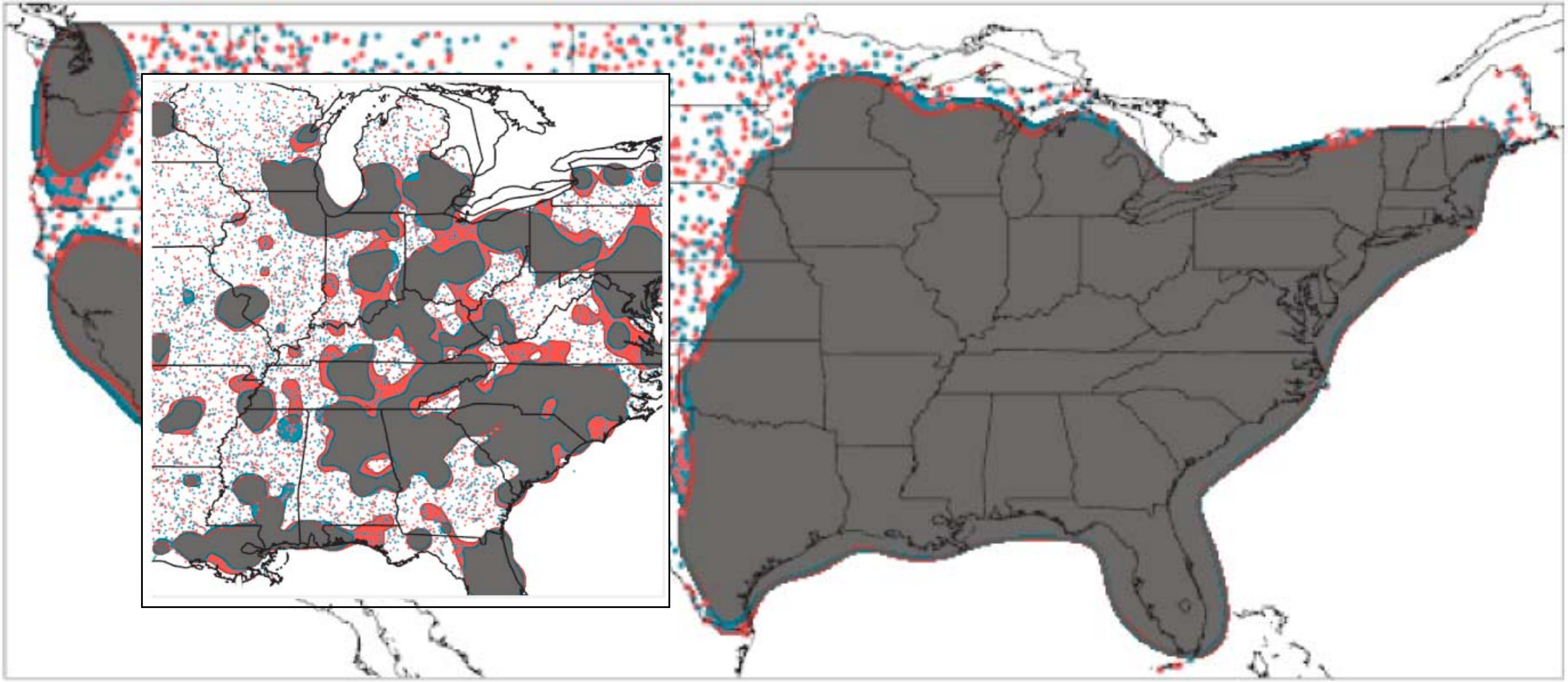
[A. Mayorga, A. Sarikaya, and M. Gleicher, 2013-2015]

Splatterplot: Summarization Changes During Zoom



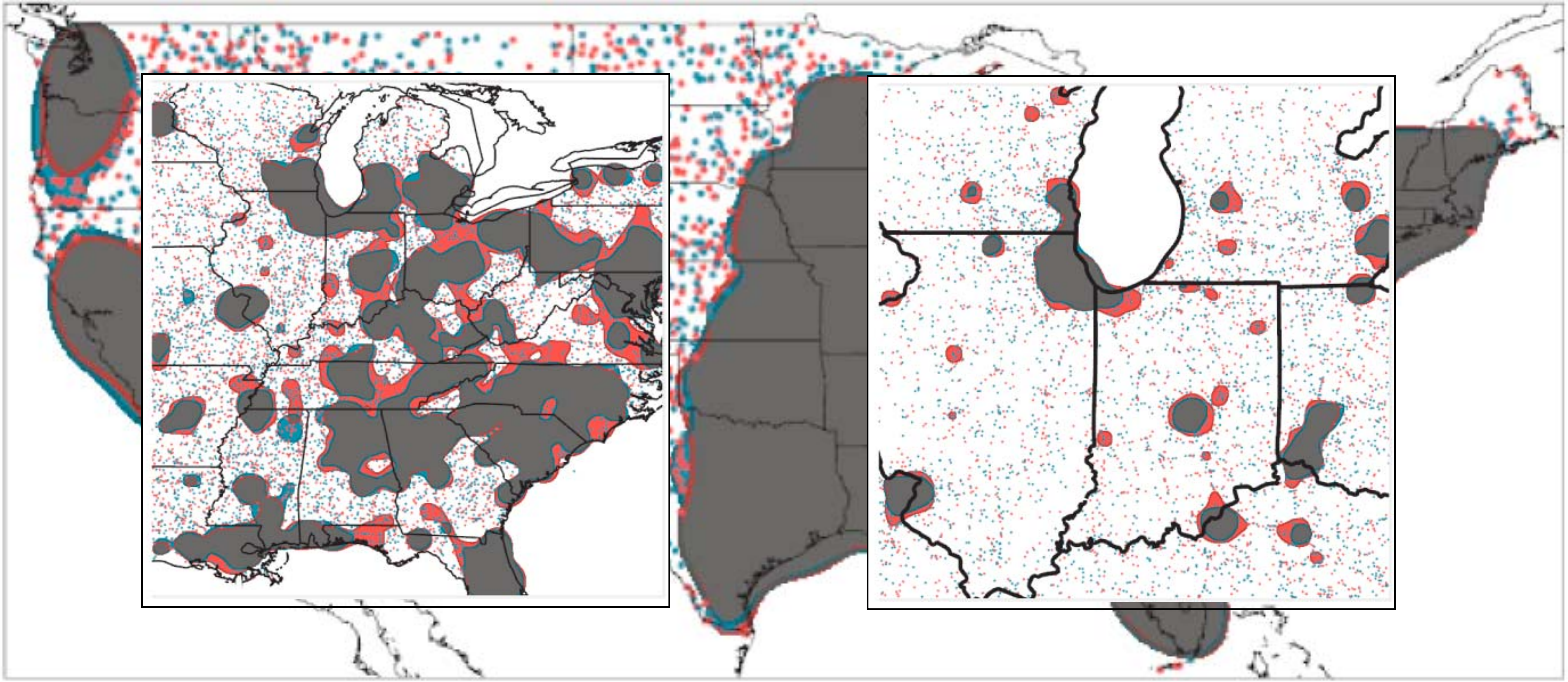
[A. Mayorga and M. Gleicher, 2013]

Splatterplot: Summarization Changes During Zoom



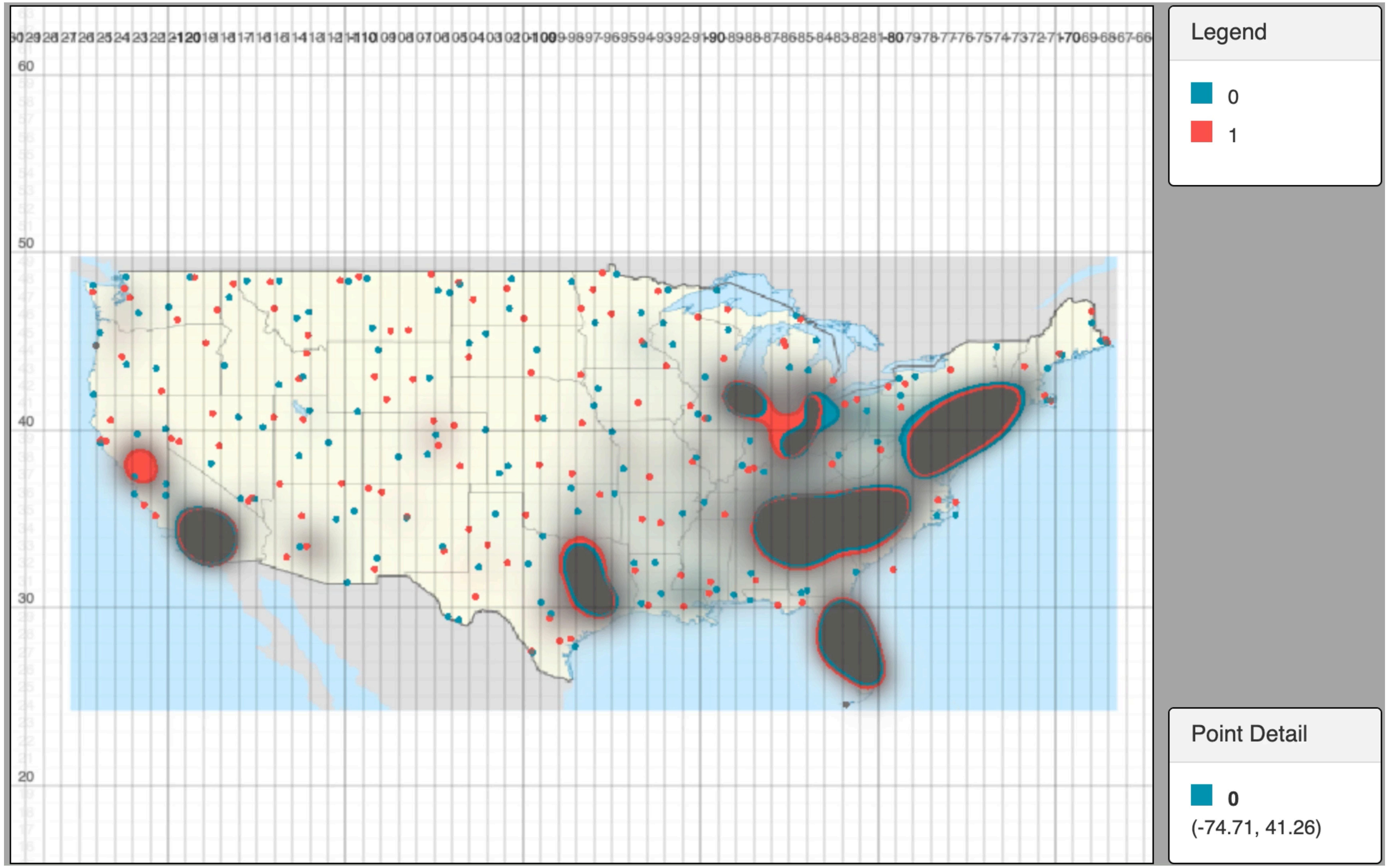
[A. Mayorga and M. Gleicher, 2013]

Splatterplot: Summarization Changes During Zoom

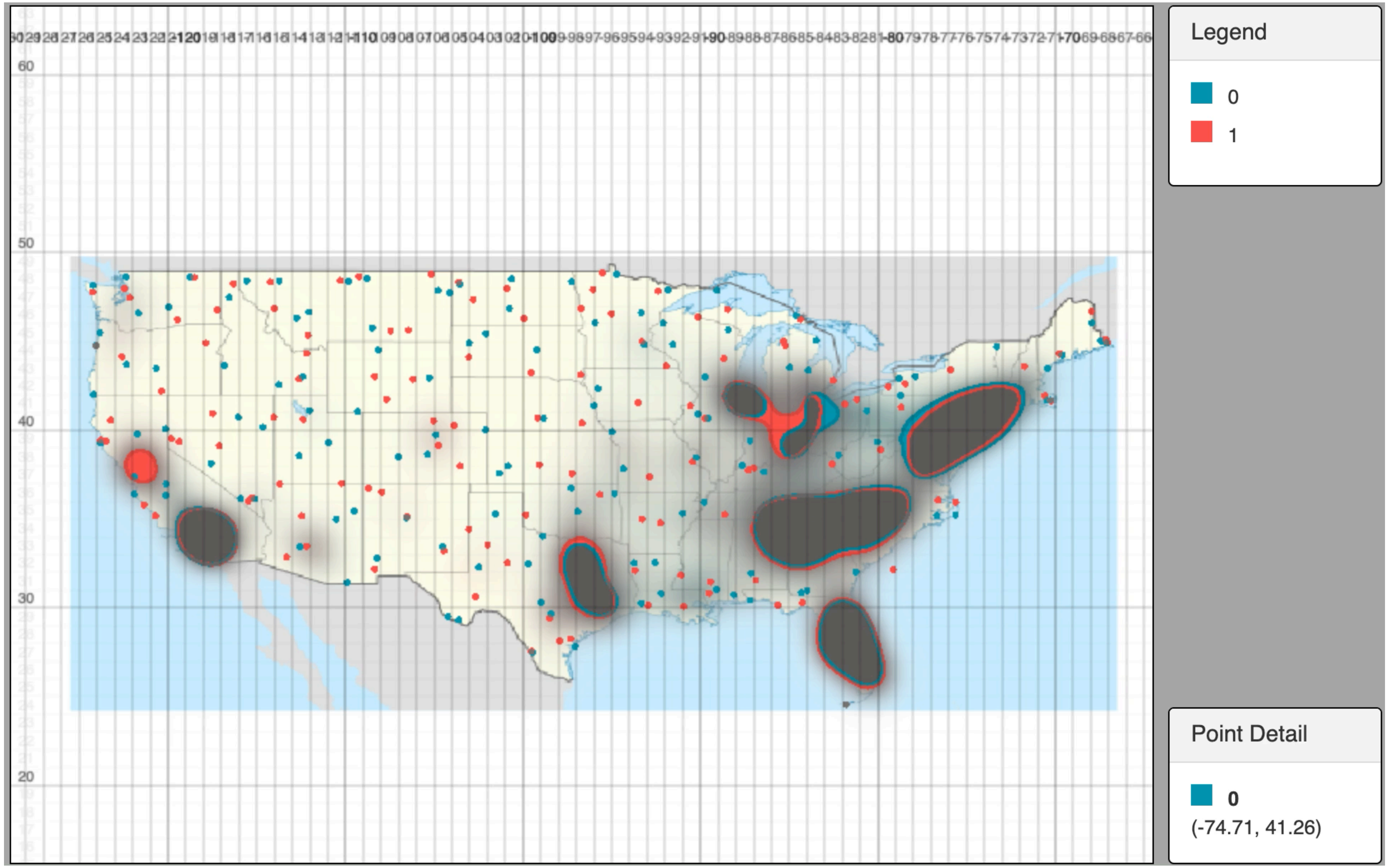


[A. Mayorga and M. Gleicher, 2013]

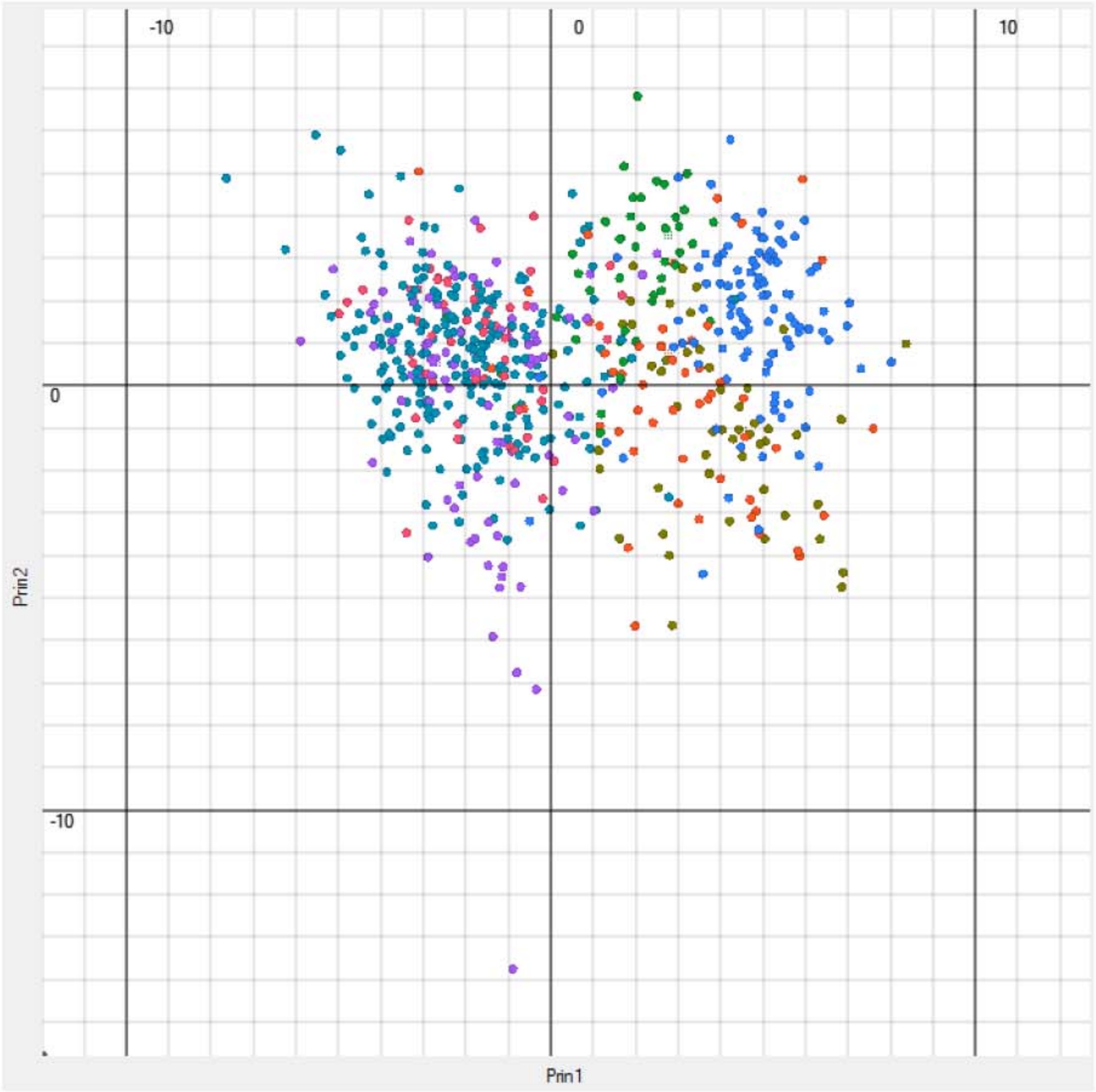
Splatterplots: Representation Changes During Translate!

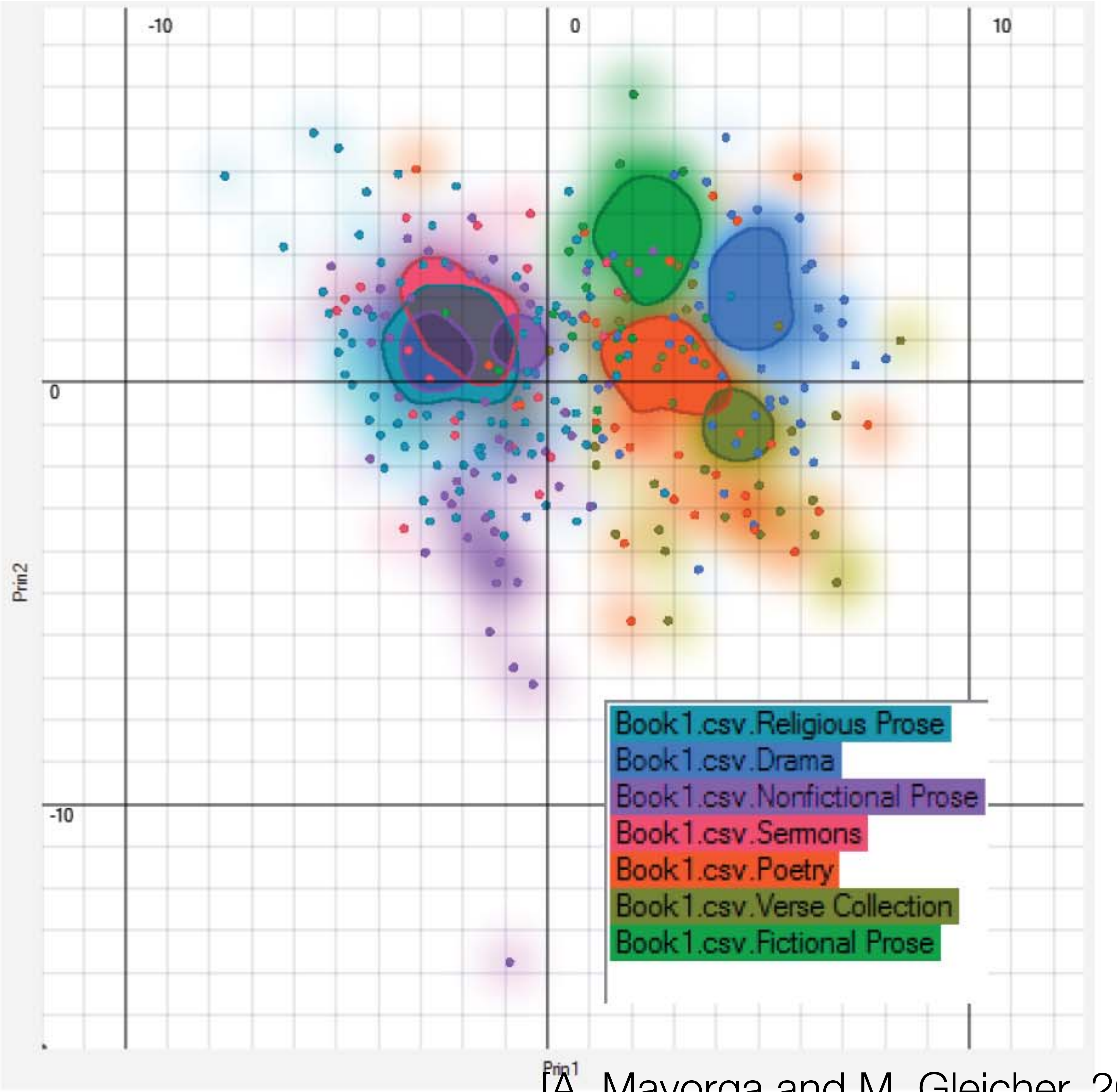


Splatterplots: Representation Changes During Translate!



Multiclass Splatterplots Using Color Blending





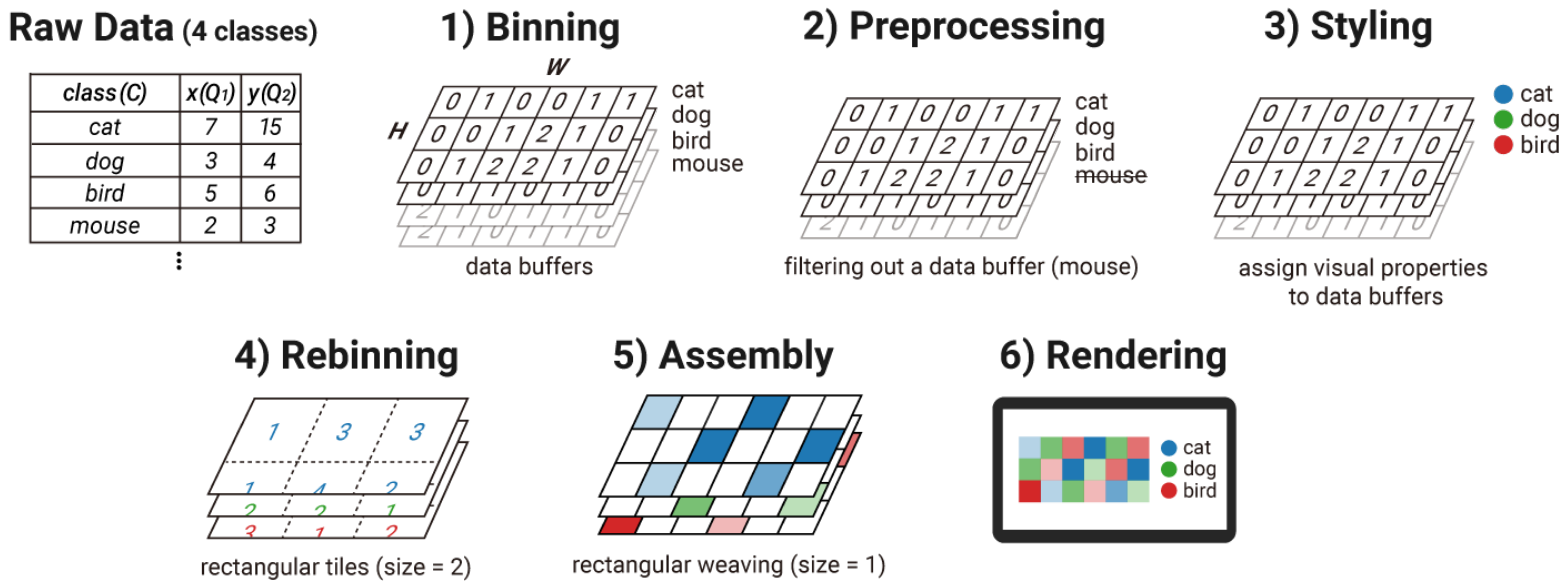
[A. Mayorga and M. Gleicher, 2013]

Class Buffer Model

- Binning (back-end): data buffers aka 2D histograms
- Preprocessing: data buffer operations like smoothing
- Styling: add visual properties to data buffers to get class buffers
- Rebinning: partition into tiles and aggregate
- Assembly: single density map image
  - uses masking, mixing, hatching, glyphs
- Rendering: add backgrounds, legends, etc.

- Most work is done on the front-end (except binning)

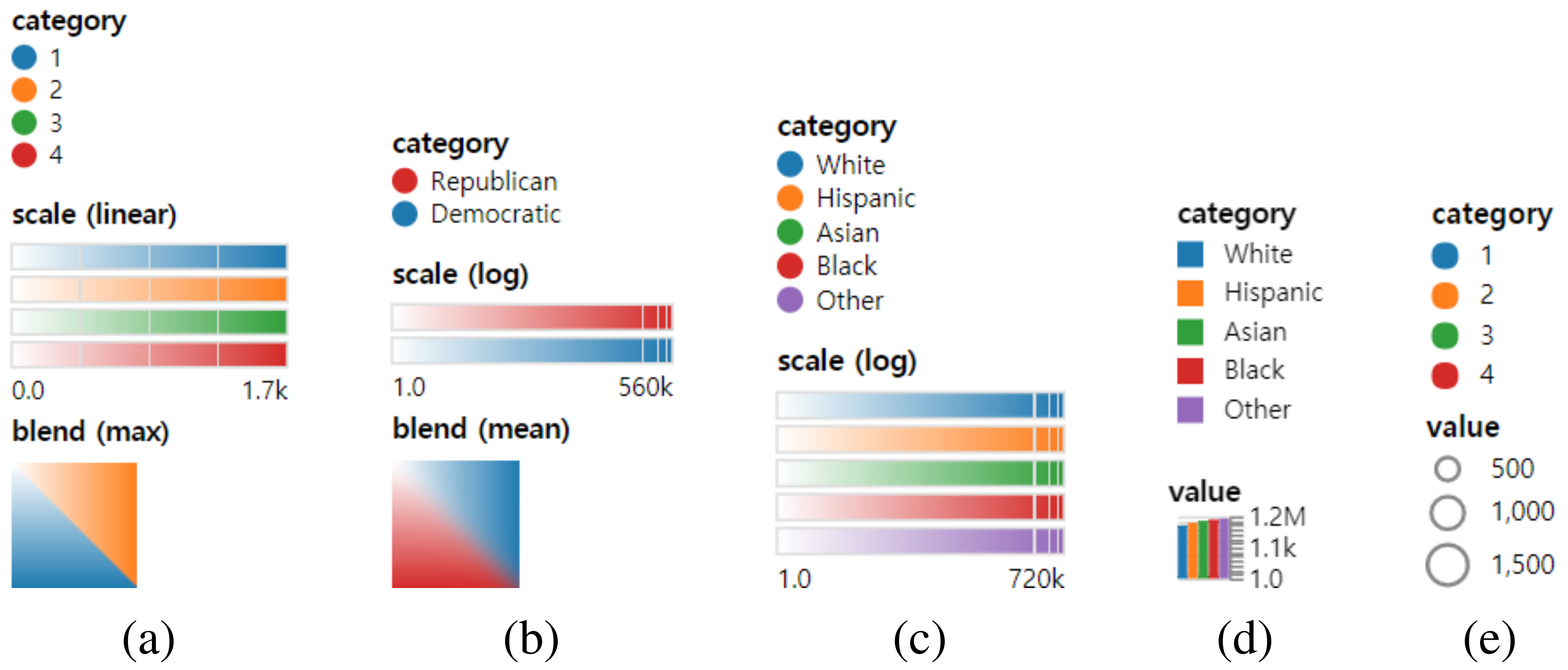
Class Buffer Model



[J. Jo et al.]

Legends

- Automatically generated
- key: name and color for classes
- scale: counts → visual attributes
- explanation: how colors are mixed (not for hatching/masking)



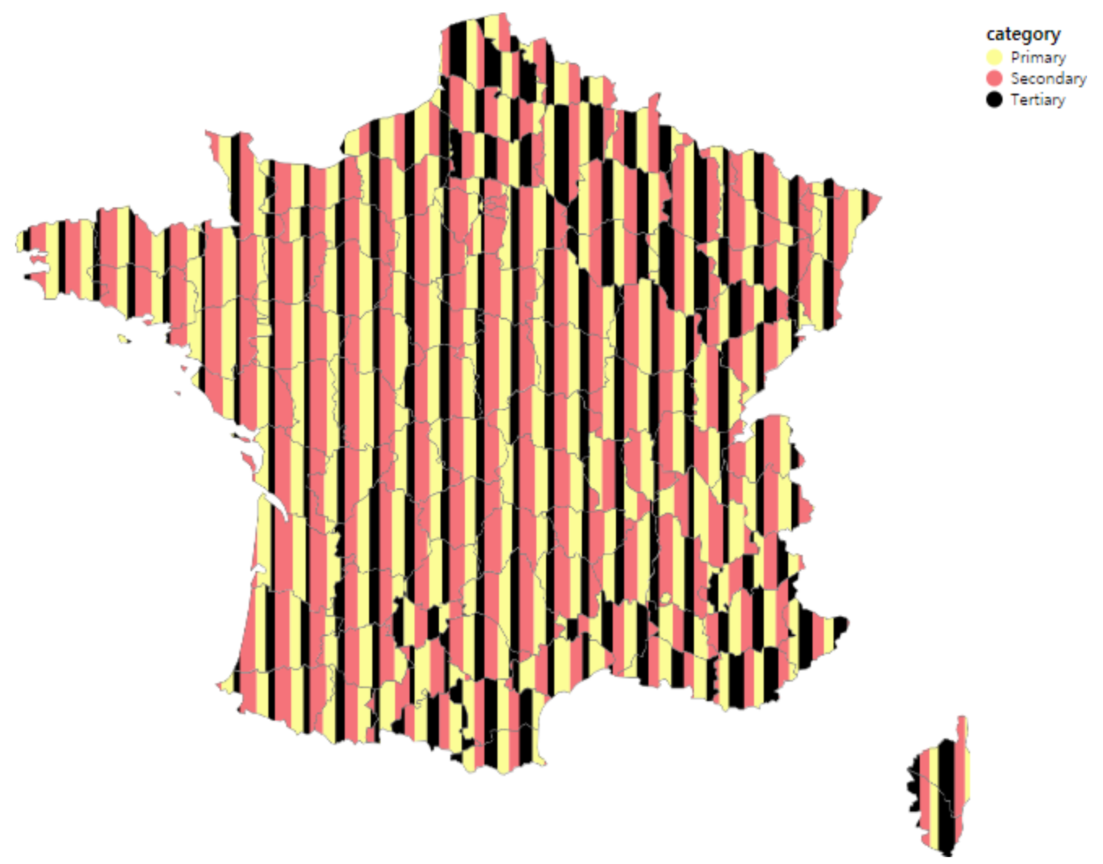
[J. Jo et al.]

Syntax

```
{"description"?: <string>,
 "background"?: <Color>,
 "data": {"url": <url> | "dataSpec": <DataSpec>},
 "smooth"?: {"radius": <number>},
 "reencoding"?: {
   "label"?: <LabelSpec>,
   "color"?: <ColorSpec>,
   "hatching"?: <HatchingSpec>},
 "rescale"?: {
   "type": "linear"|"log"|"pow"|"sqrt"|"cbrt"|"equidepth",
   "rebin"?: {
     "type": "none"|"square"|"rect"|"topojson"|"voronoi",
     "aggregation": "mean"|"max"|"sum"|"min"|"density",
     "width"?: <number>, "height"?: <number>,
     "size"?: <number>,  "topojson"?: <TopoJSONSpec>,
     "url"?: <string>, "feature"?: <string>,
     "points"?: <Point[]>, "stroke"?: <Color>},
   "compose"?: {
     "mix": "none"|"invmin"|"mean"|"max"| "blend"|
            "weavingrandom"|"weavingsquare"|"weavinghex"|
            "weavingtri"|"propline"|"hatching"|"separate"|
            "glyph"|"dotdensity"|"time",
     "mixing"?: "additive"|"subtractive"|"multiplicative",
     "size"?: <number>, "widthprop"?: <string|number>,
     "colprop"?:<boolean>, "order"?: <number[]>,
     "glyphSpec"?: <GlyphSpec>, "interval"?: <number>},
   "levels"?: <number>
 },
 "contour"?: {
   "stroke": <number>, "lineWidth"?: <number>,
   "values"?: <number[]>, "blur"?: <number> },
 "legend"?: <LegendSpec>, "stroke"?: <StrokeSpec>,
 "axis"?: <AxisSpec>}
```

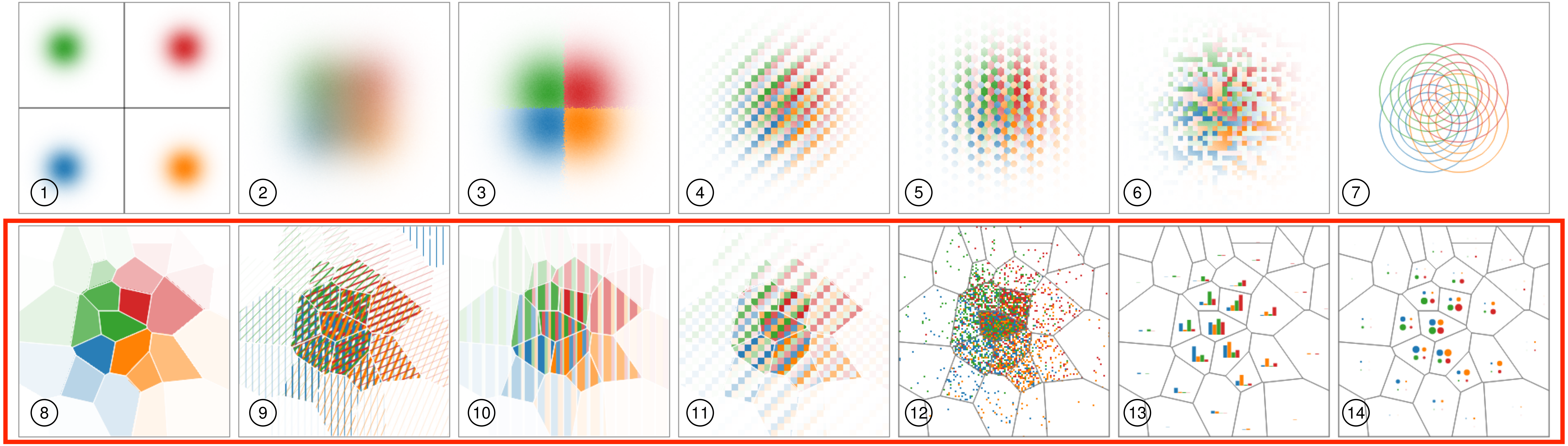
Specification

```
"rebin": {
  "type": "topojson",
  "url": "franceD.json",
  "feature": "poly"},
"compose": {
  "mix": "propline",
  "size": 18,
  "widthprop": "percent"}
```



[J. Jo et al.]

MDM Gallery



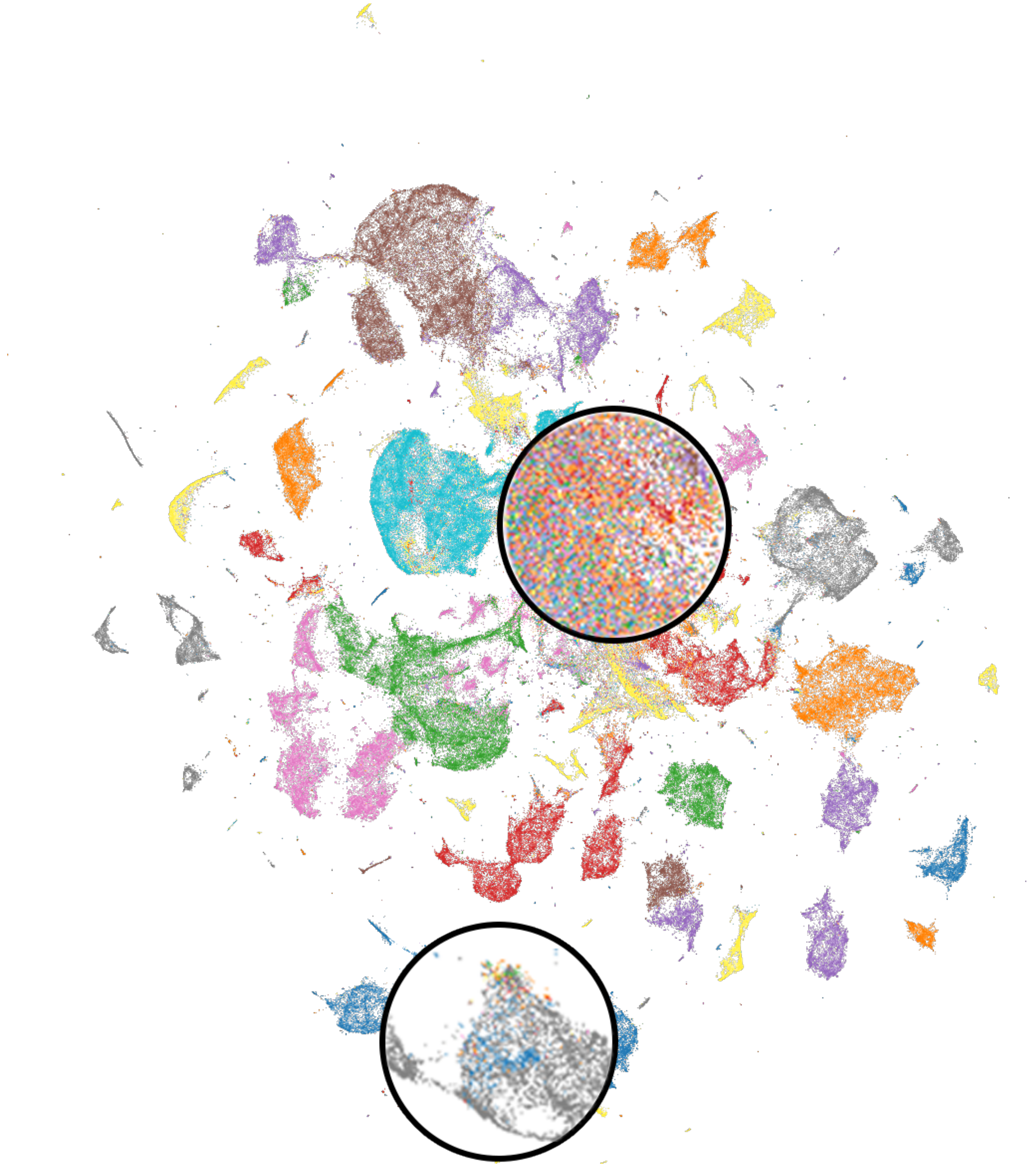
Rebinned and Aggregated

[J. Jo et al., 2019]

High-Dimensional Data

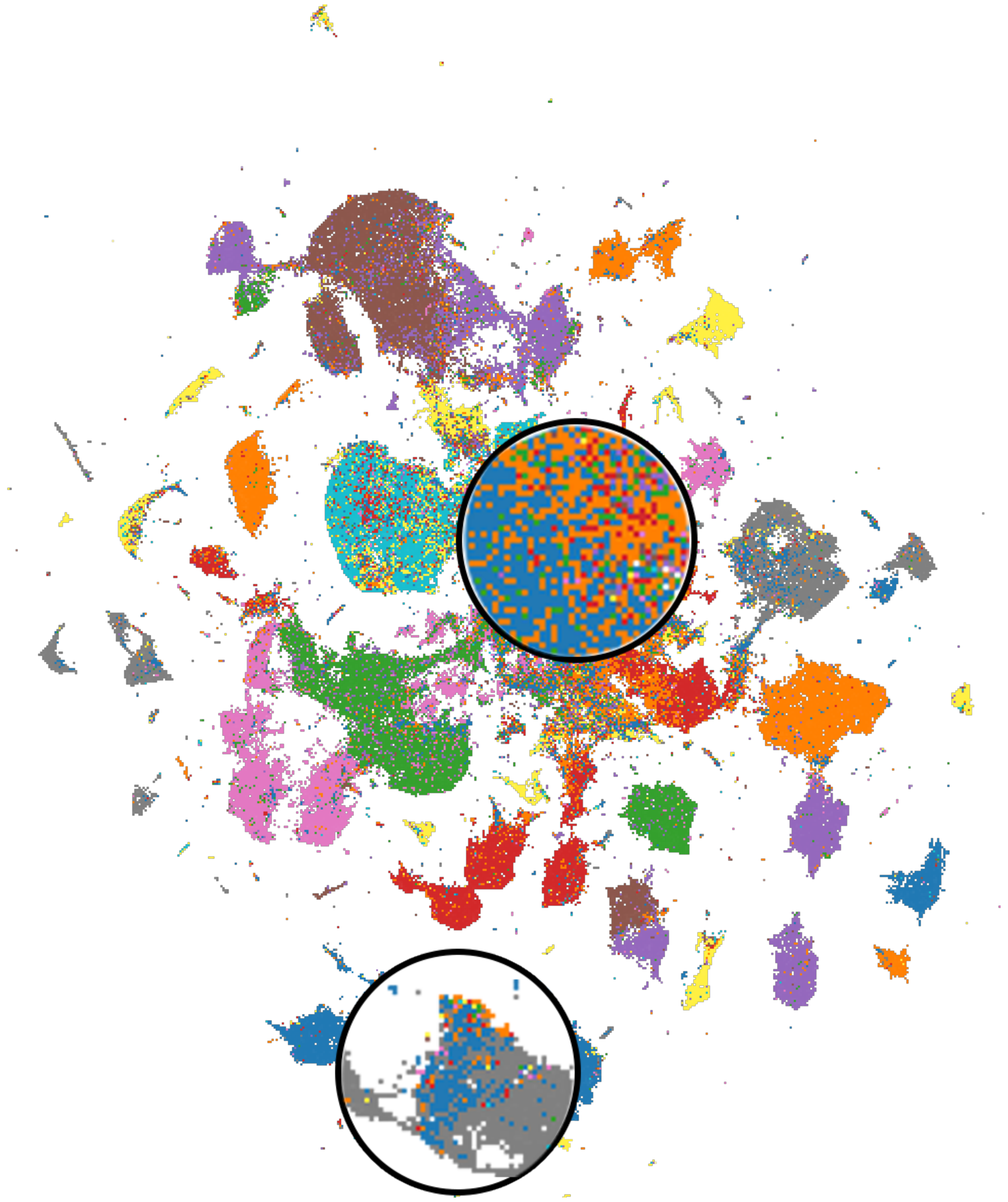
Specification

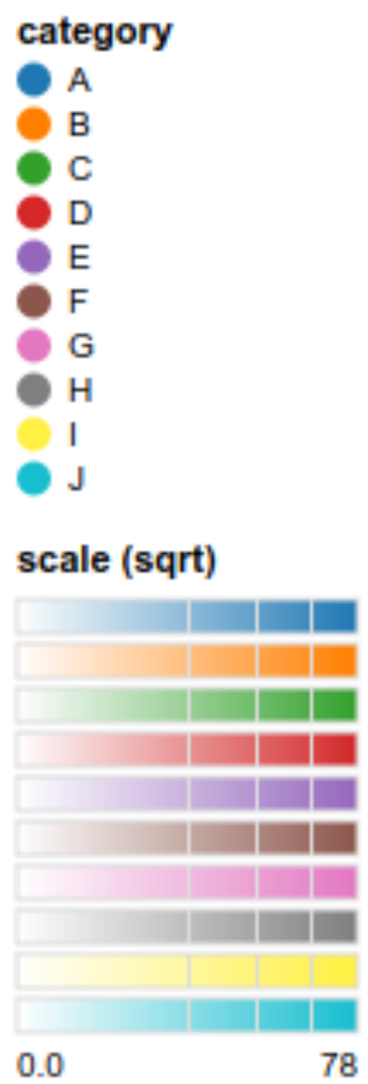
```
"compose": {"mix": "max"},
"rescale": {"type": "equidepth"}
```



Specification

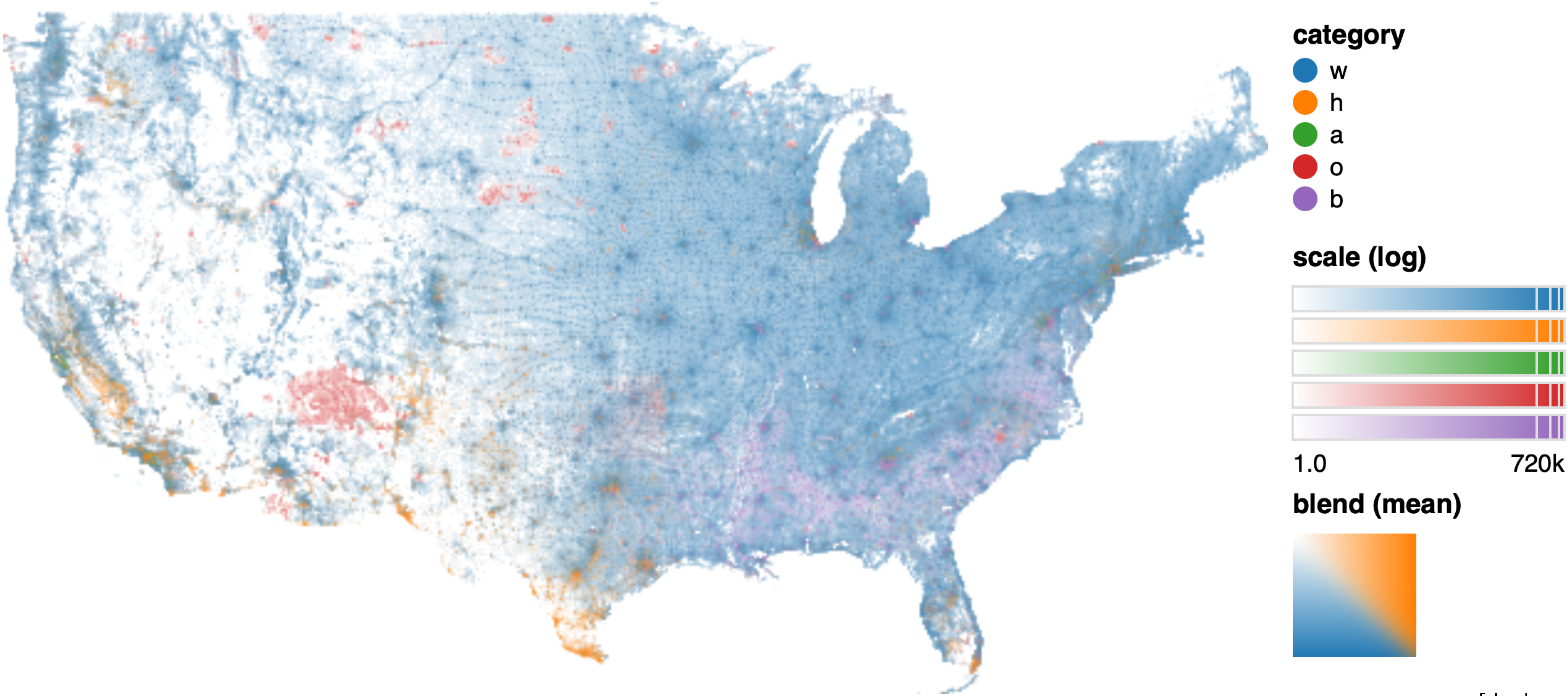
```
"compose": {"mix": "invmin"},
"rescale": {"type": "sqrt"},
"rebin": {"type": "square",
 "size": 2, "aggregation": "min"}
```





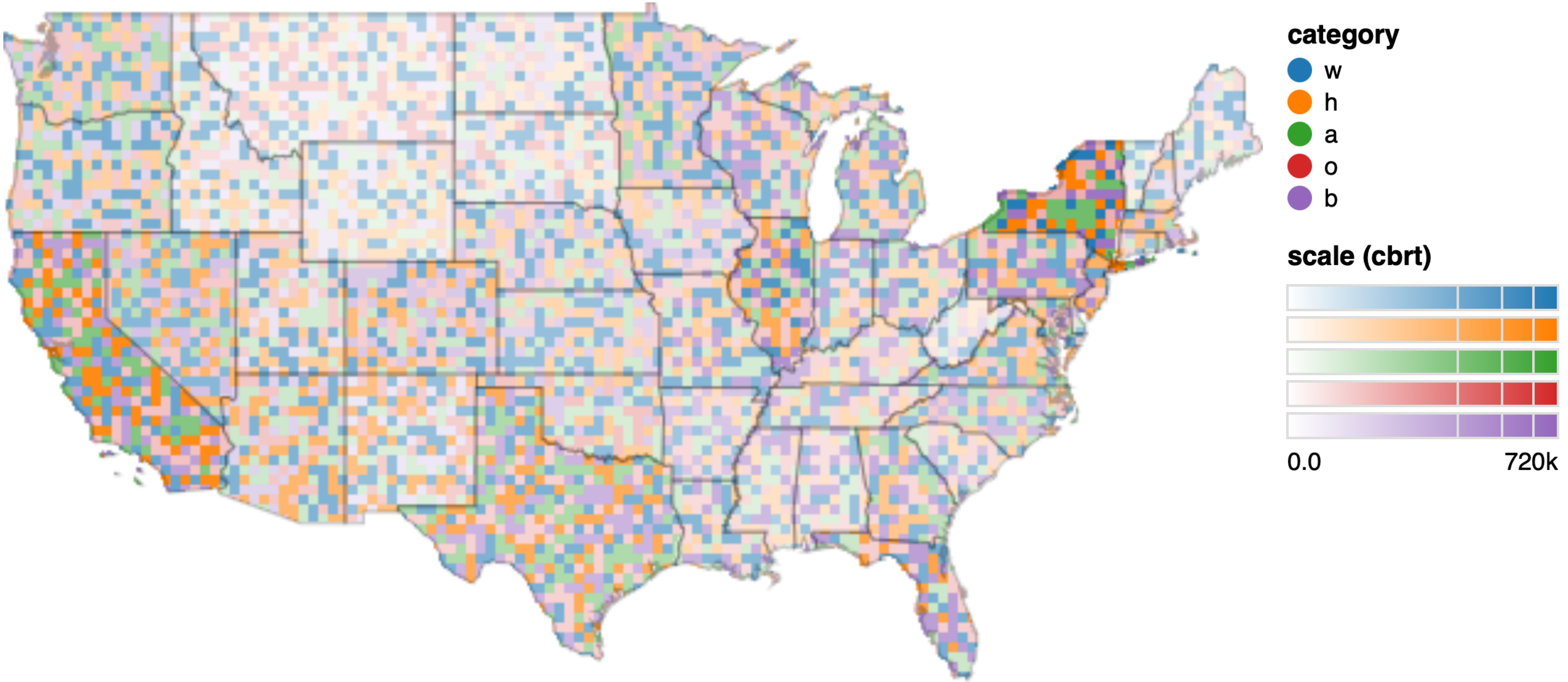
[J. Jo et al.]

Results



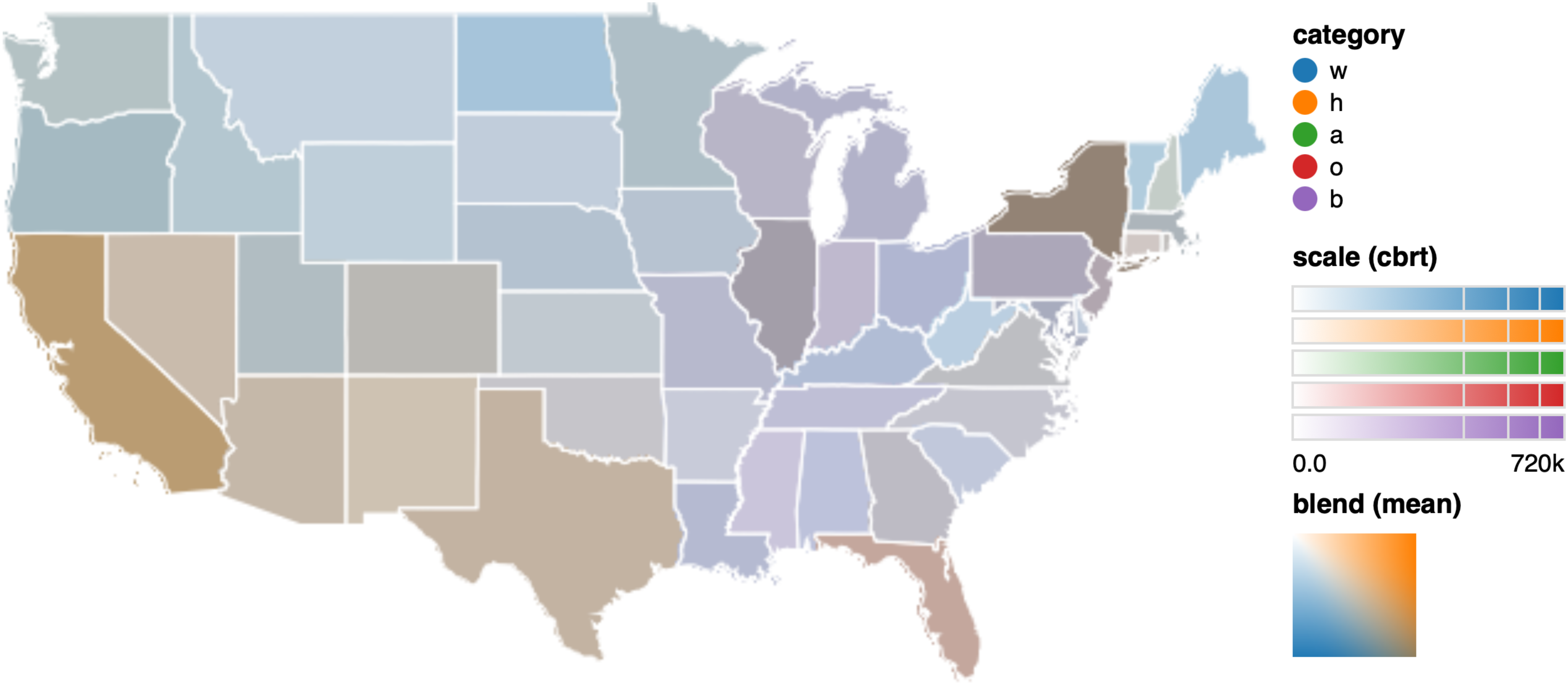
[J. Jo et al.]

Results



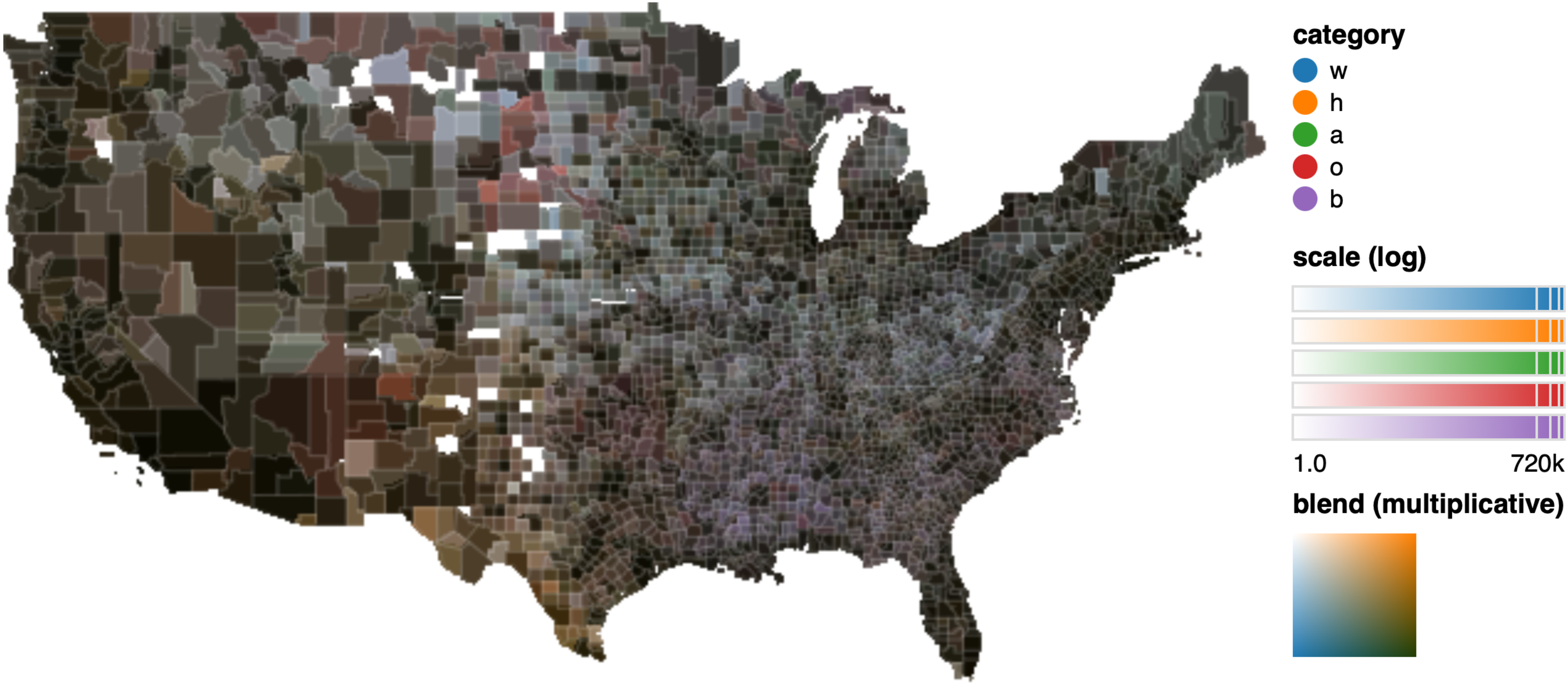
[J. Jo et al.]

Results



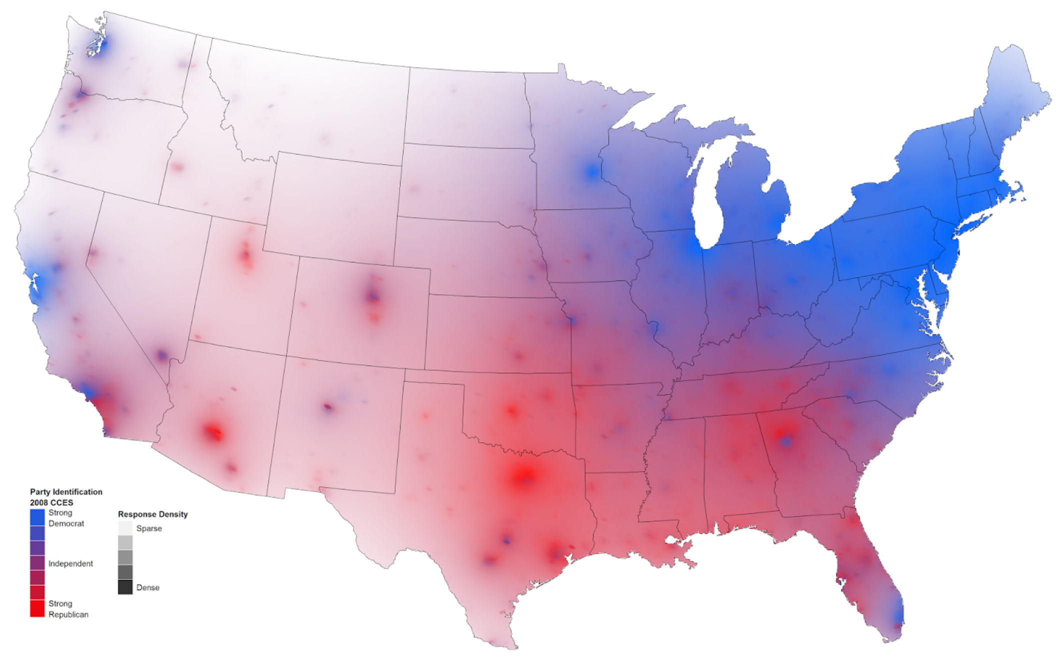
[J. Jo et al.]

Results

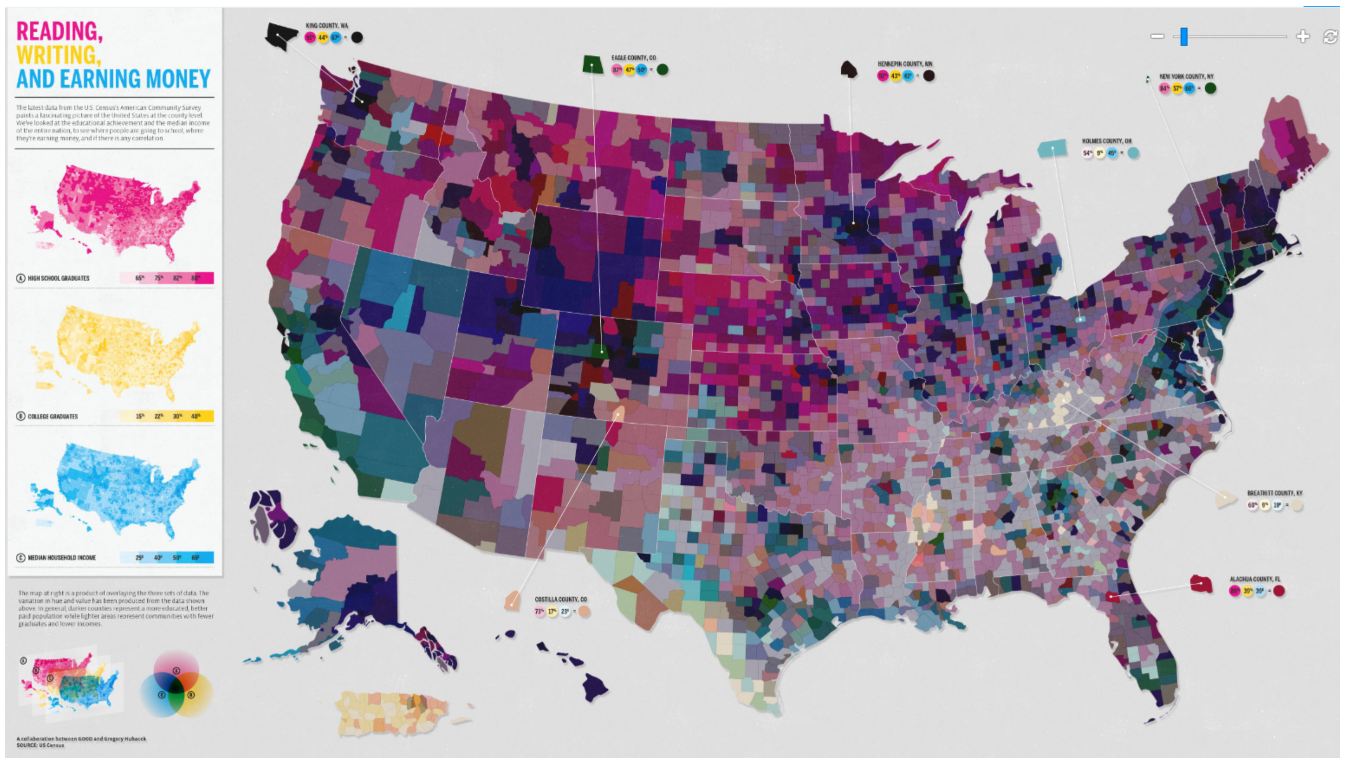


[J. Jo et al.]

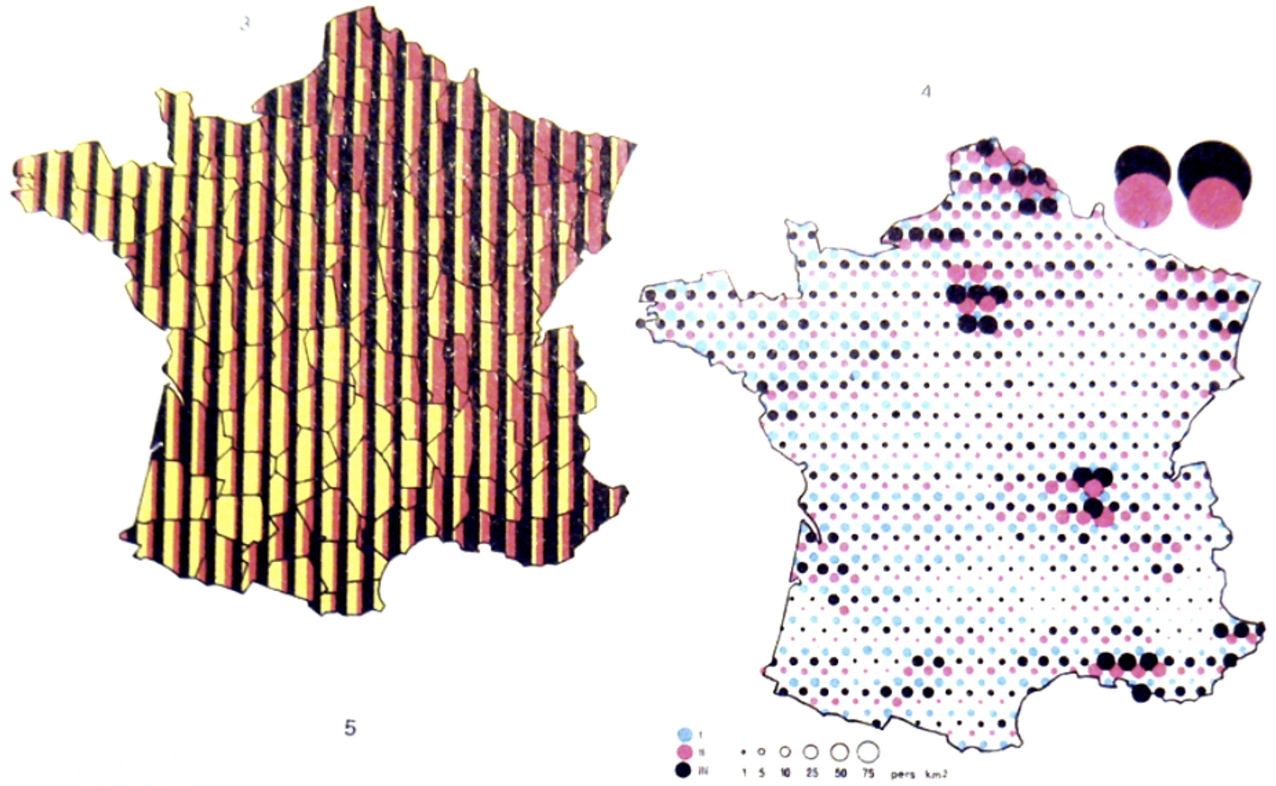
Original Examples



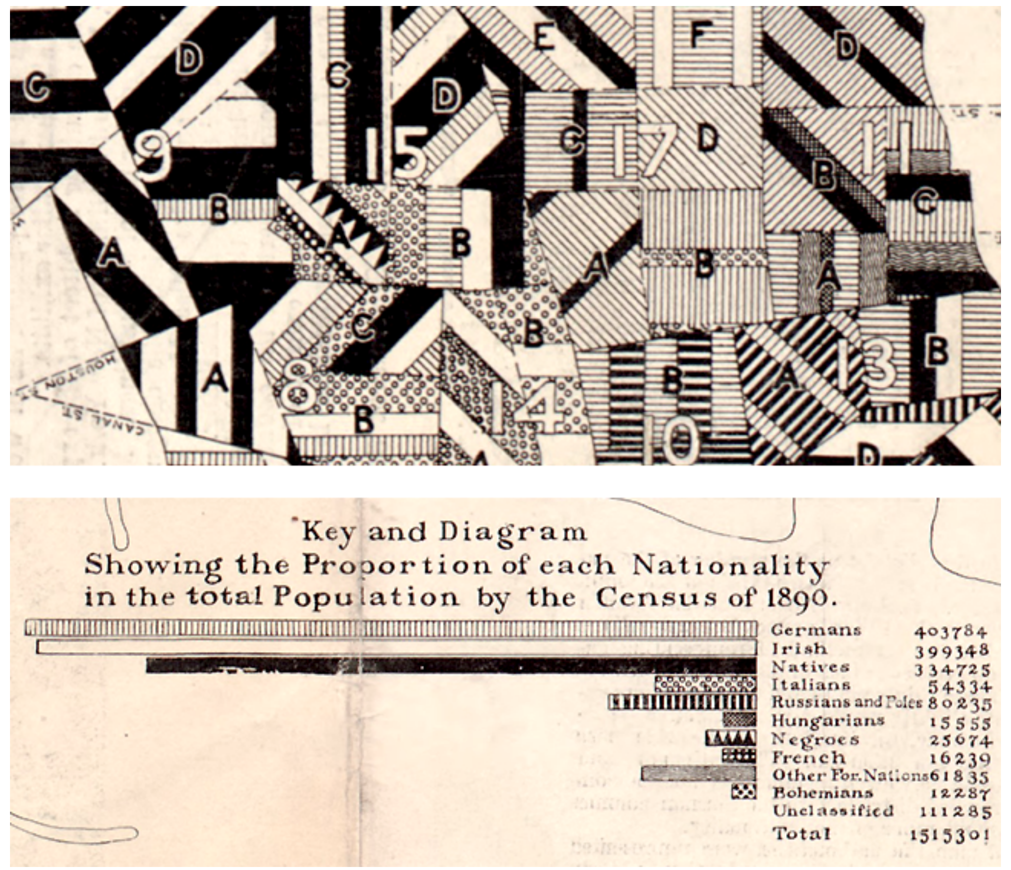
(a) Support for Democratic vs. Republican candidates in 2008 [46].



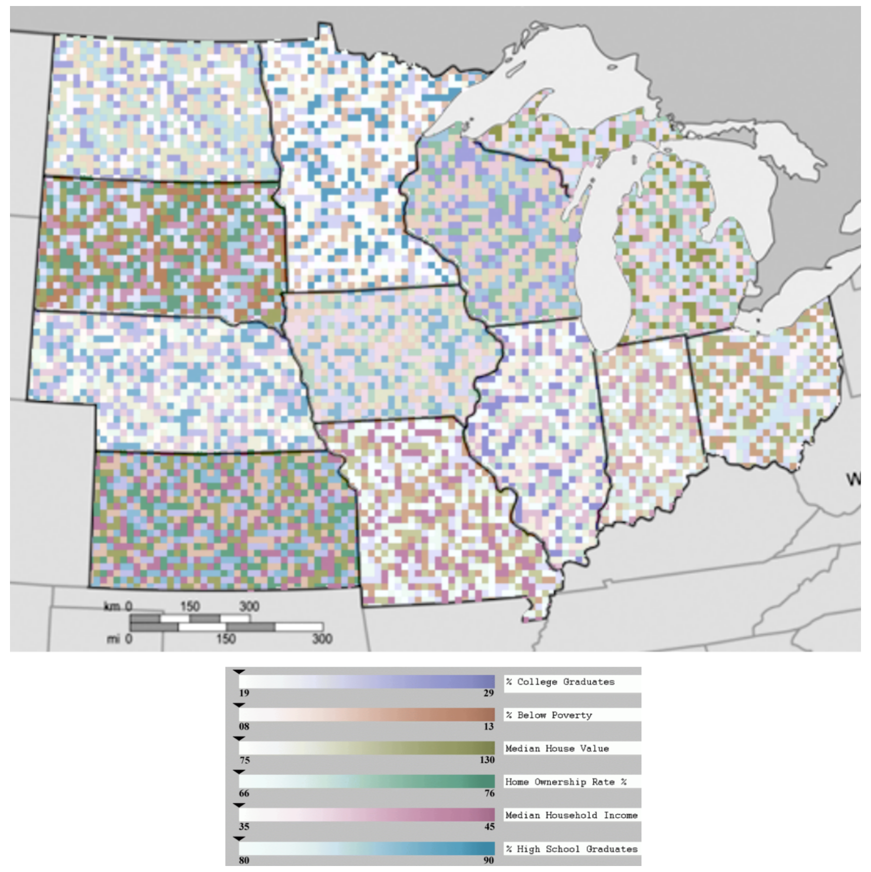
(b) Percentage of high school graduates, of college graduates, and median house income in 2009 [22].



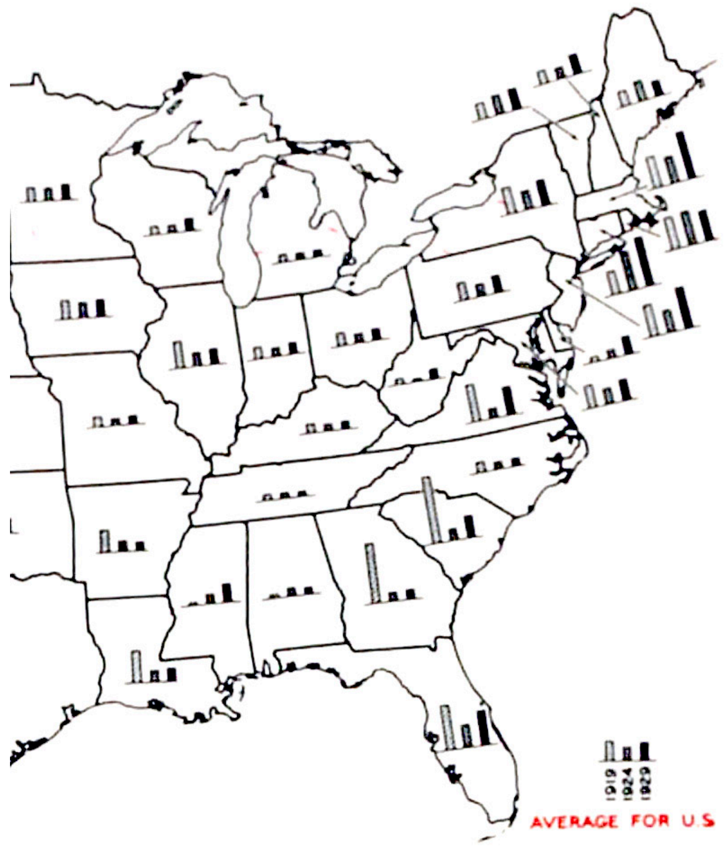
(c) Number of workers per sector of economy (primary, secondary, tertiary) in 1954 by Bertin [4].



(d) Detail of a map of New-York City showing the distribution of nationalities across districts in 1890 [25, 40].



(e) Six socioeconomic indicators in each of the twelve Midwestern US states [19].



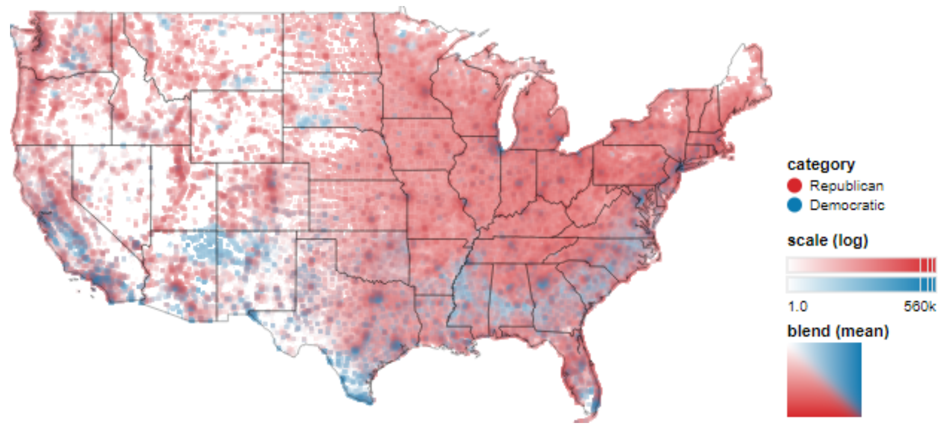
(f) Detail of a map showing average sales per farm for each US state in 1919, 1924, and 1929 [10].

[J. Jo et al.]

Examples using MDM Grammar

Specification

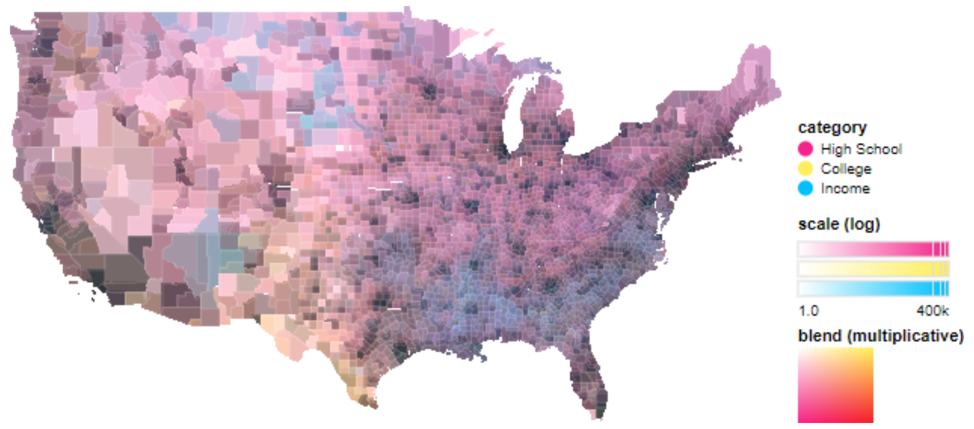
```
"smooth": {"radius": 1},
"rescale": {"type": "log"},
"compose": {"mix": "mean"},
"stroke": {
  "type": "topojson",
  "url": "us.json",
  "feature": "states",
  "color": "rgba(0, 0, 0, 0.3)"}
```



(a)

Specification

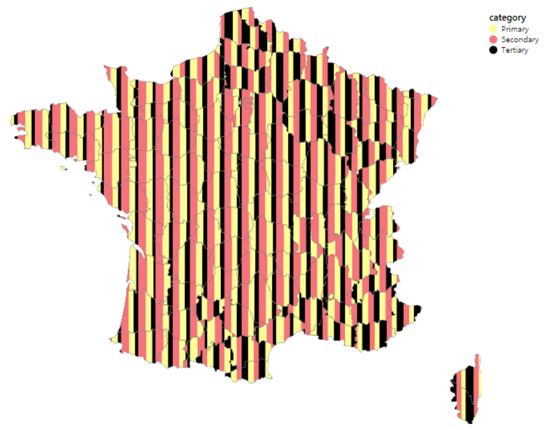
```
"rebin": {
  "type": "topojson",
  "url": "us.json",
  "feature": "counties"},
"rescale": { "type": "log" },
"compose": {
  "mix": "blend",
  "mixing": "multiplicative"}
```



(b)

Specification

```
"rebin": {
  "type": "topojson",
  "url": "franceD.json",
  "feature": "poly"},
"compose": {
  "mix": "propline",
  "size": 18,
  "widthprop": "percent"}
```



(c)

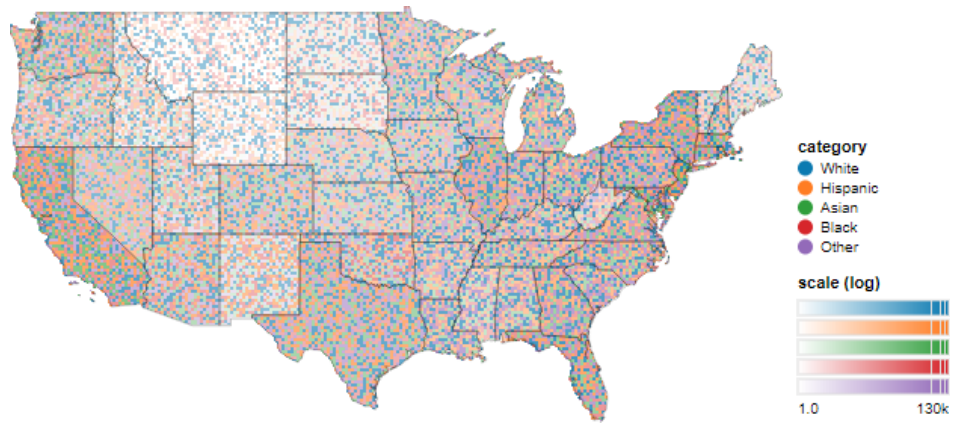
Specification

```
"rebin": // US rebinning
"compose": {
  "mix": "hatching", "size": 4,
  "widthprop": "percent",
  "colprop": true}
```

(d)

Specification

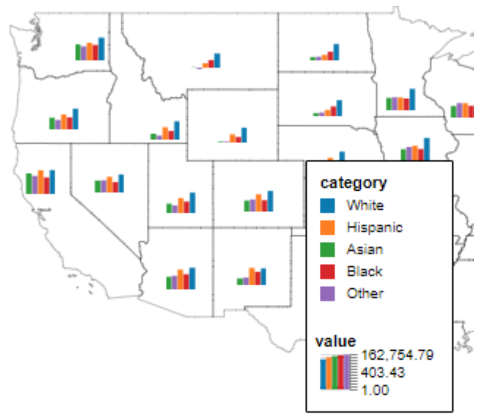
```
"rebin": // US rebinning
"compose": {
  "mix": "weavingrandom",
  "size": 2}
```



(e)

Specification

```
"rebin": // US rebinning
"compose": { "mix": "glyph",
  "glyphSpec": {
    "template": "bars",
    "width": 20, "height": 24}}
```



(f)

[J. Jo et al.]

Benefits

- Expensive Computation (binning) done up front so various visual encodings can be explored
  - Synthetic Census Data: 300 million x 5 classes ~11 seconds on laptop
- Interactive Data Exploration
  - Can filter out rows
  - Deal with more dimensions

[J. Jo et al.]

Limitations

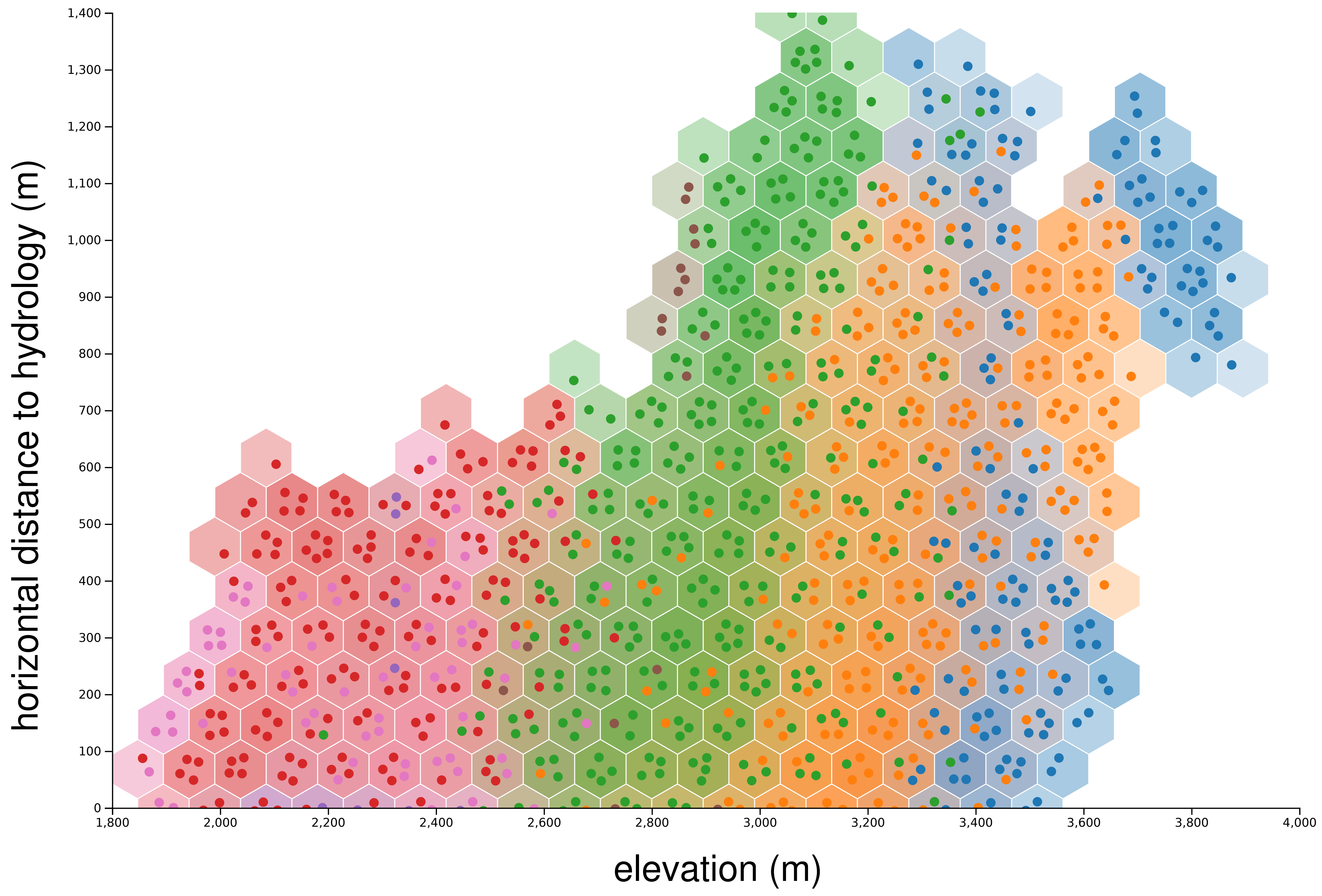
- Other map features missing (e.g. landmarks)
- No guidance on best use
- Uses Canvas instead of WebGL

[J. Jo et al.]

Critique

Grammar Complexity and Coverage?

- Complexity of specification
  - Dependencies between parts?
- Which new techniques work well?
- Do other techniques (e.g. Heimerl's Tree Cover) work with the grammar?

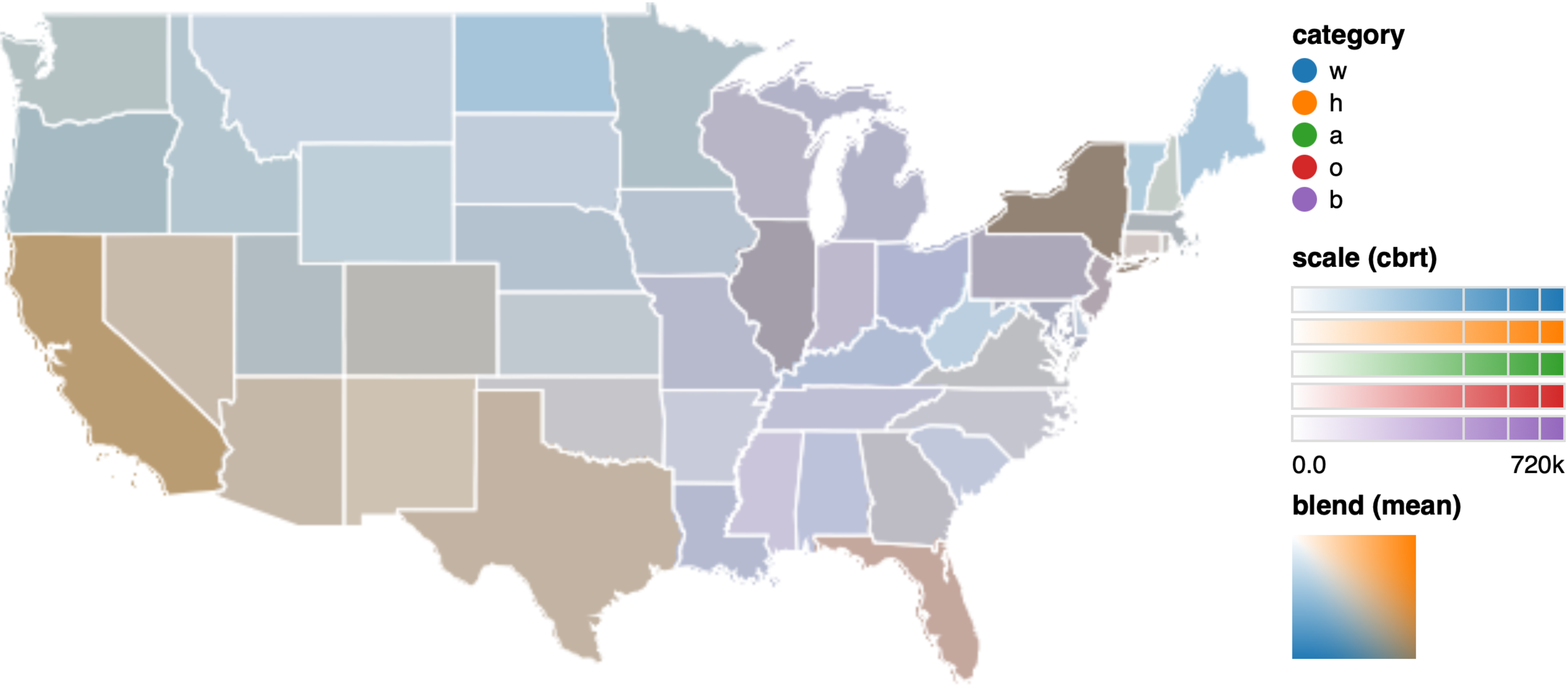


[F. Heimerl et al., Tech Report]

Guidance

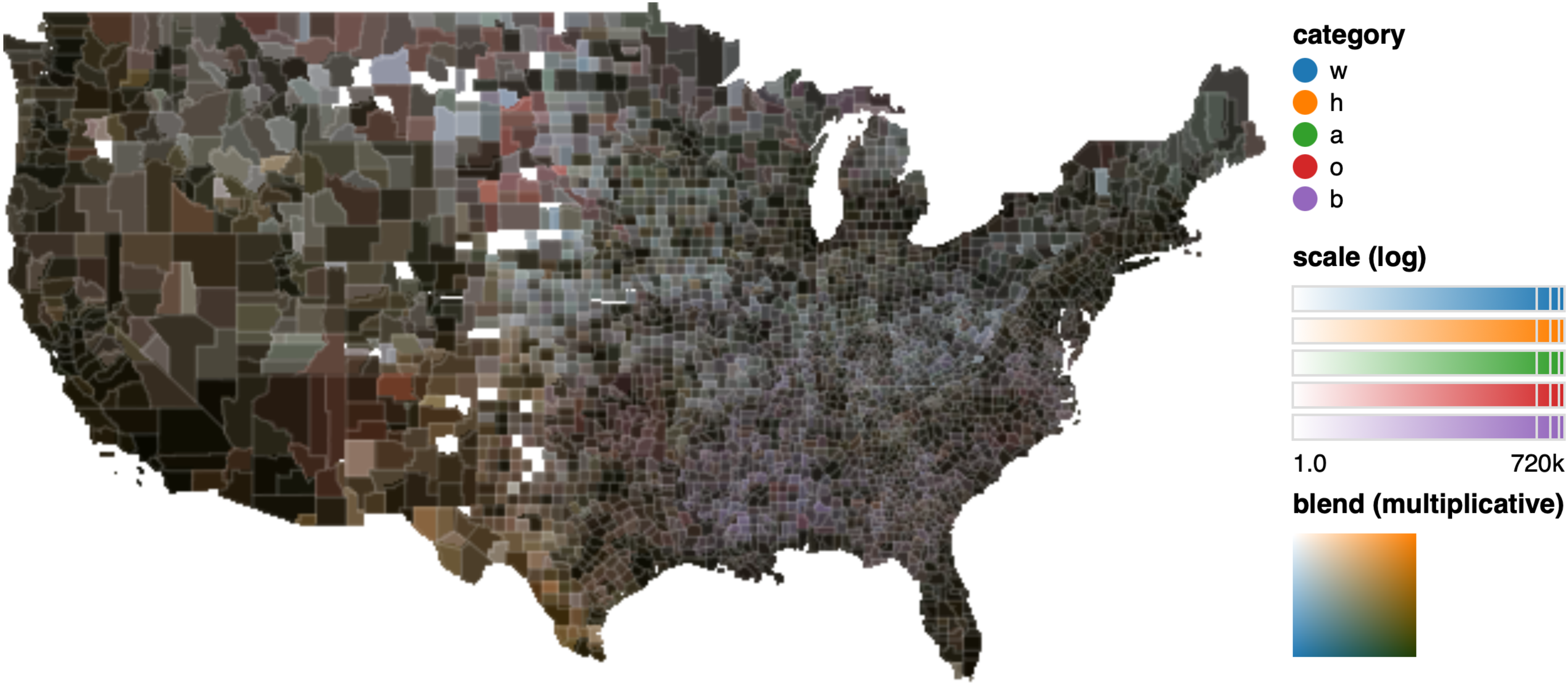
- Vega-Lite has general defaults
- MDM Class Buffer often has paired guidance (hard to know what works)

Meaningful Result?



[J. Jo et al.]

Meaningful Result?



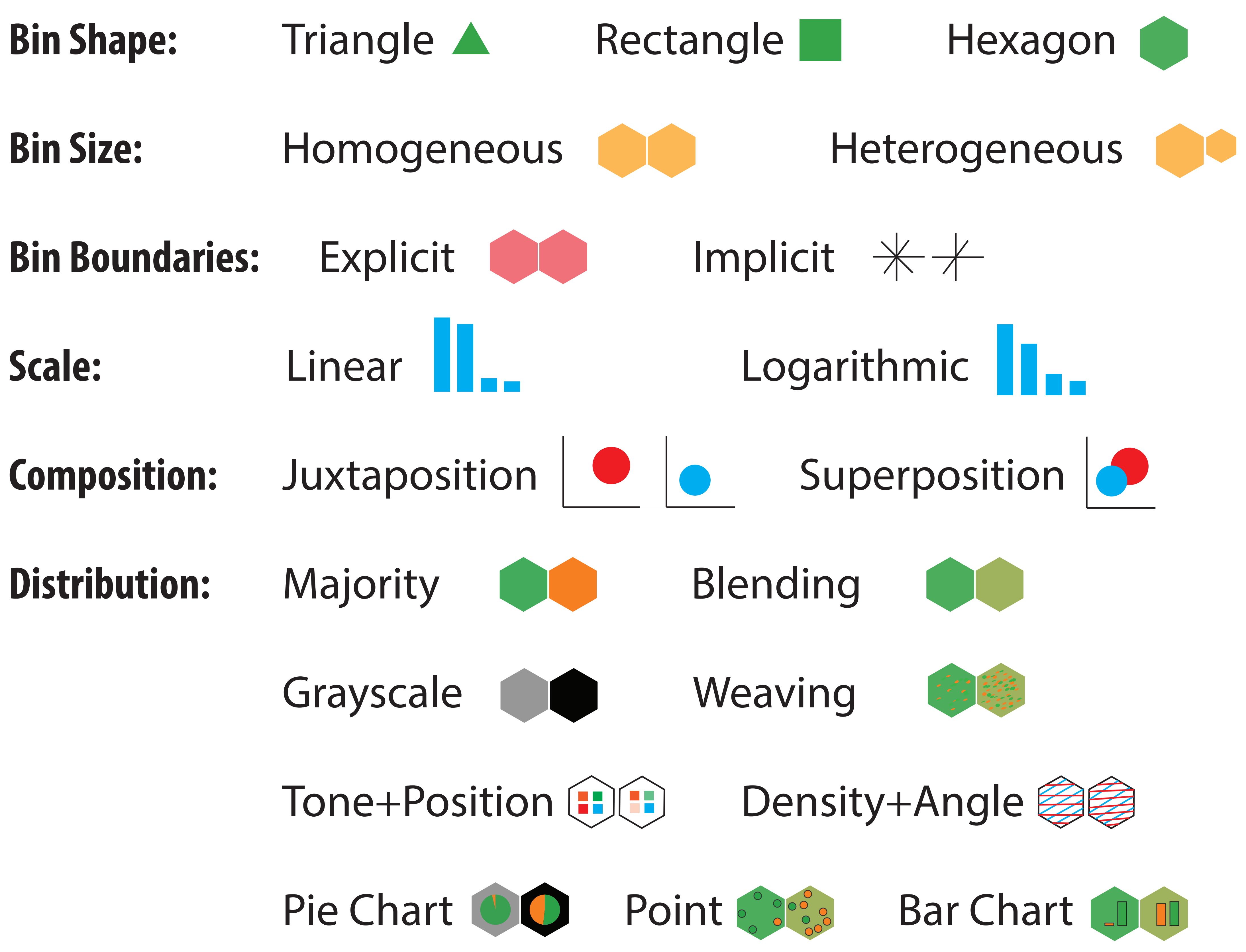
[J. Jo et al.]

What about Tasks? (Bin and Class Tasks)

	Task	Bin-centric		Class-centric	
browsing	Explore neighborhood	1	Explore properties of bins in a neighborhood	2	Explore properties of classes in a neighborhood
	Search for known motif	3	Find known pattern across bins	4	Find known pattern across classes
	Explore data	5	Unusual patterns within or across bins, global trends between bins	6	Unusual patterns within or across classes, global trends within or between classes
aggregate-level	Characterize distribution	7	Do bins close to each other have similar properties? Or within a certain area or range of values?	8	Does a class occupy certain areas of the plot? Does its distribution have a particular shape? Do classes correlate in certain areas?
	Identify anomalies	9	Identify bins that are outliers based on the general distribution	10	Identify classes or subsets of classes that are outliers in a certain region
	Identify correlation	11	Determine level of correlation of bin properties along both dimensions	12	Determine level of correlation for class members along both dimensions
	Numerosity comparison	13	Compare density in different regions of the space	14	Compare class density in different regions of the space
	Understand distances	15	Understand a given spatialization and the coverage of the bins	16	Understand a given spatialization and the coverage of classes

[F. Heimerl et al., Tech Report]

Other Ideas



- What about focusing on tasks?
- Thinking about other visual properties?

[F. Heimerl et al., Tech Report]