

# Beyond Basics- Advanced Topics in Visualization

Nam Wook Kim

Mini-Courses – January @ GSAS  
2018

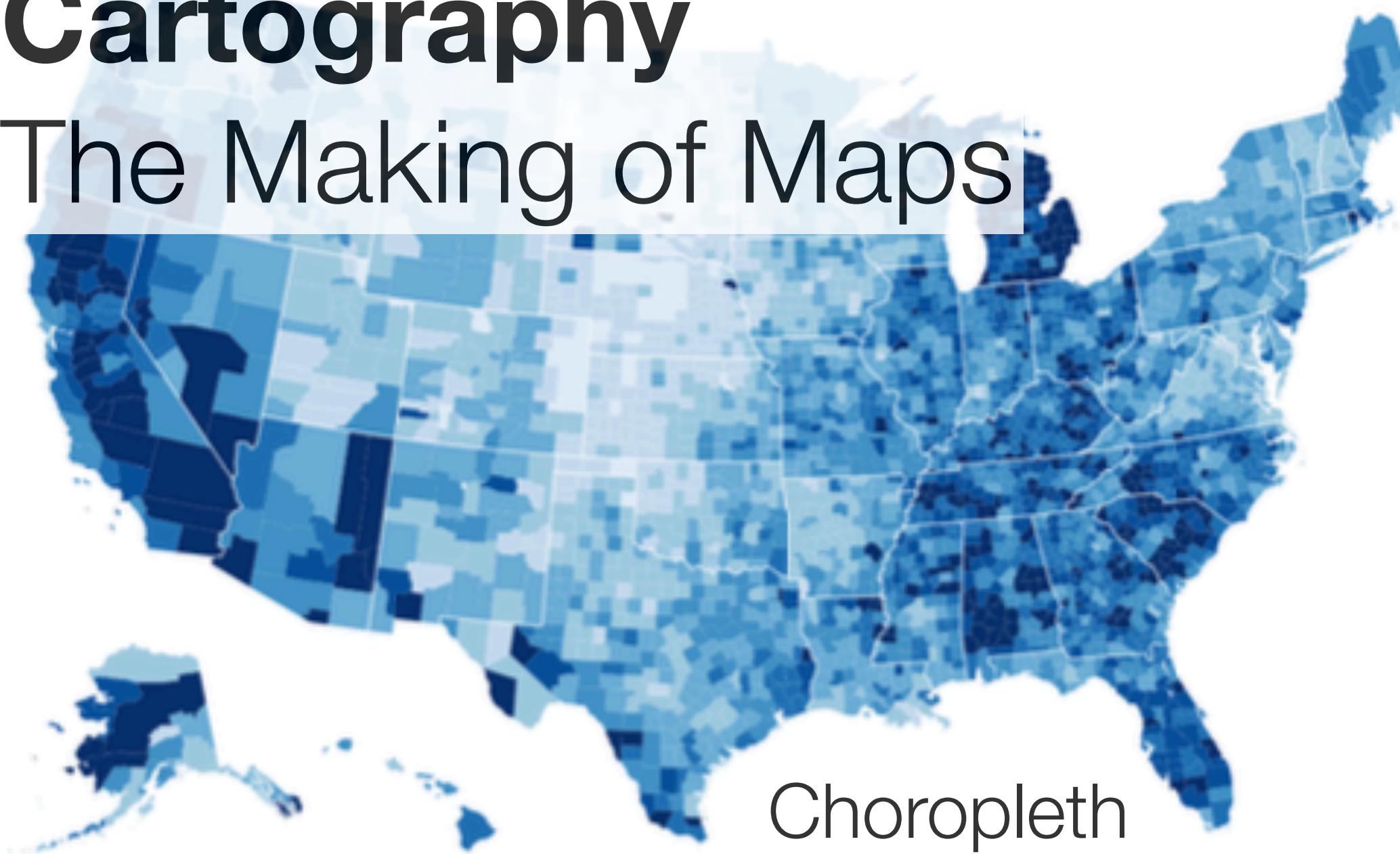
# Topics

- Maps
- Networks & Trees
- High-Dimensional Data
- Text

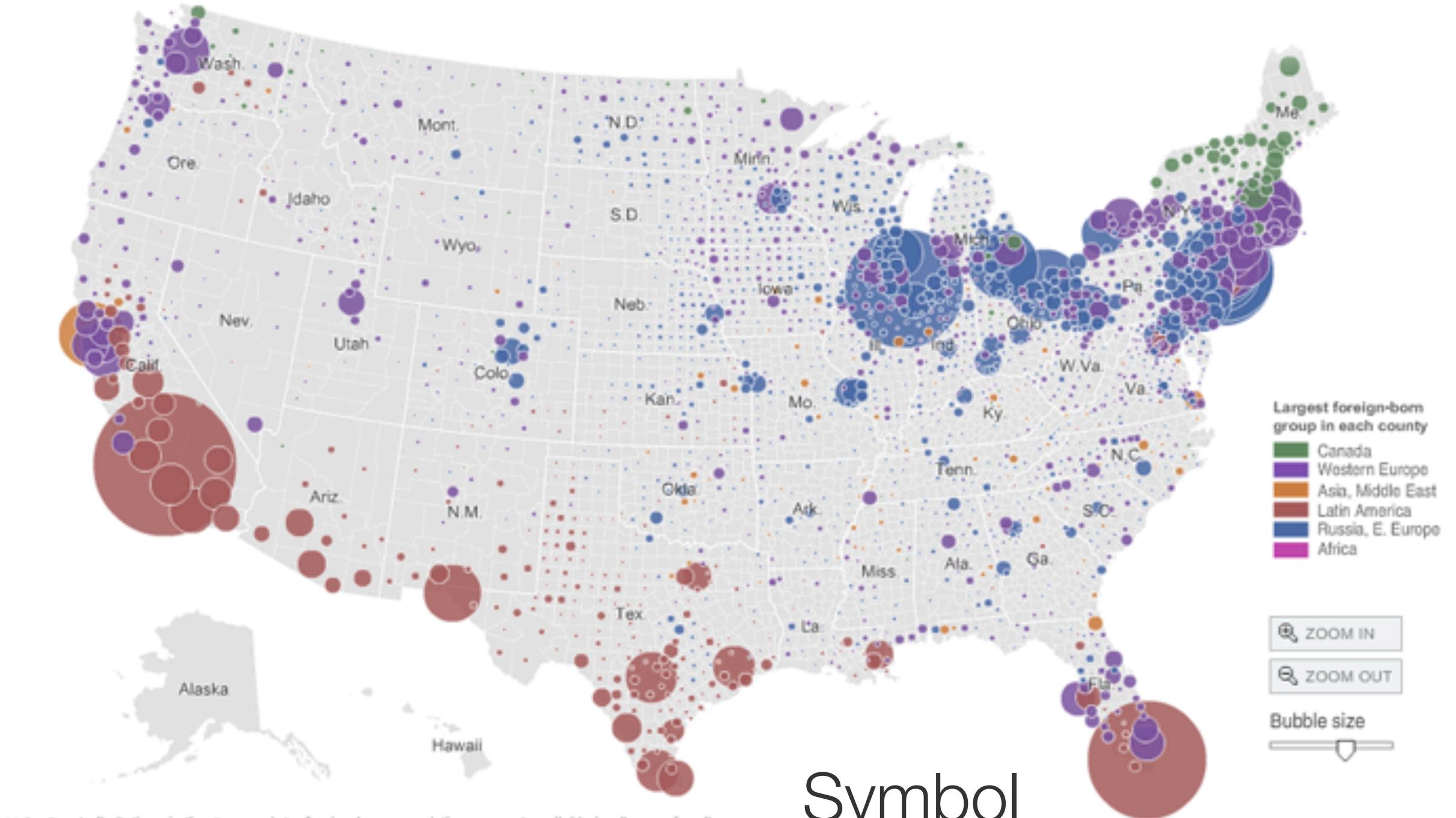
# Maps

# Cartography

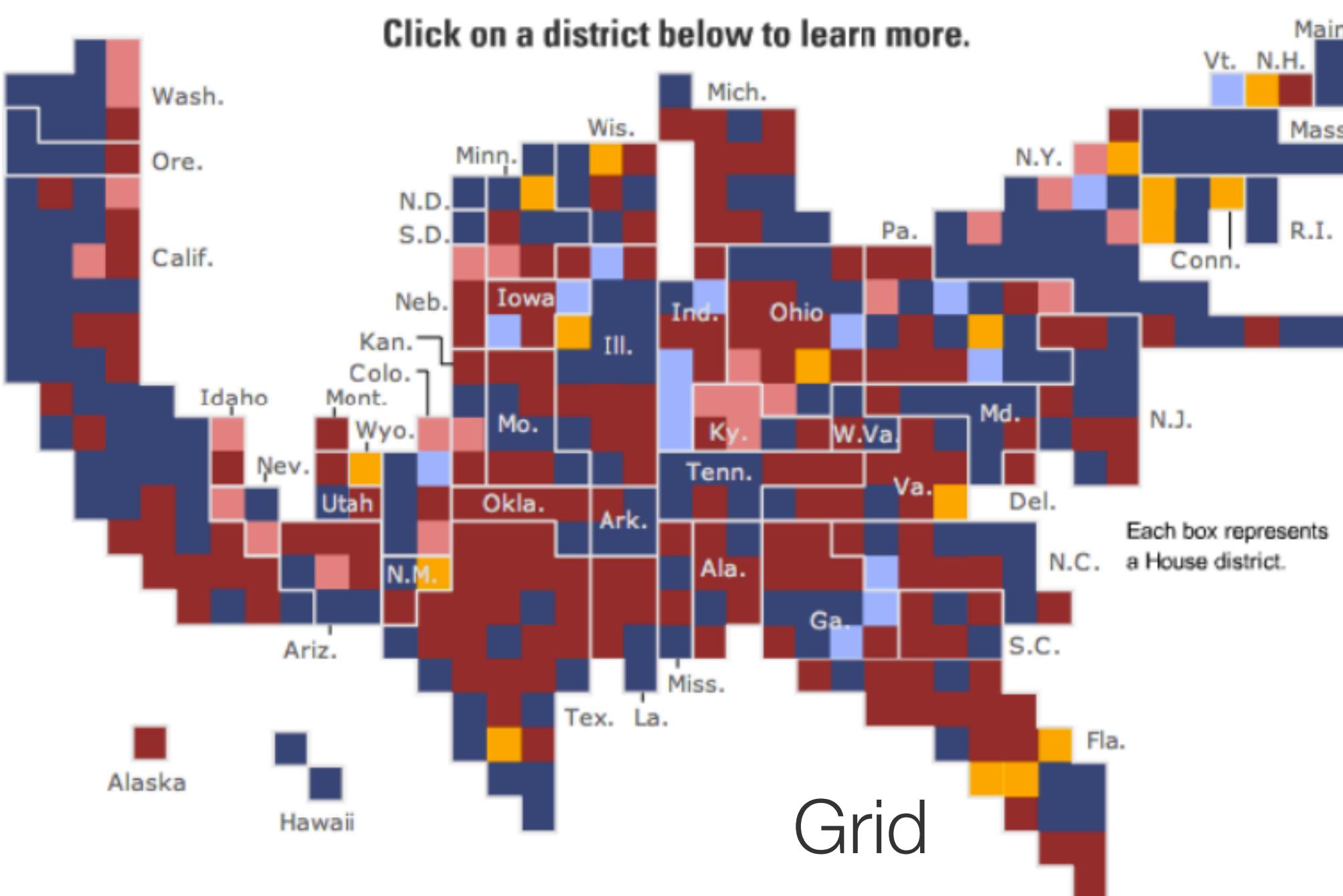
## The Making of Maps



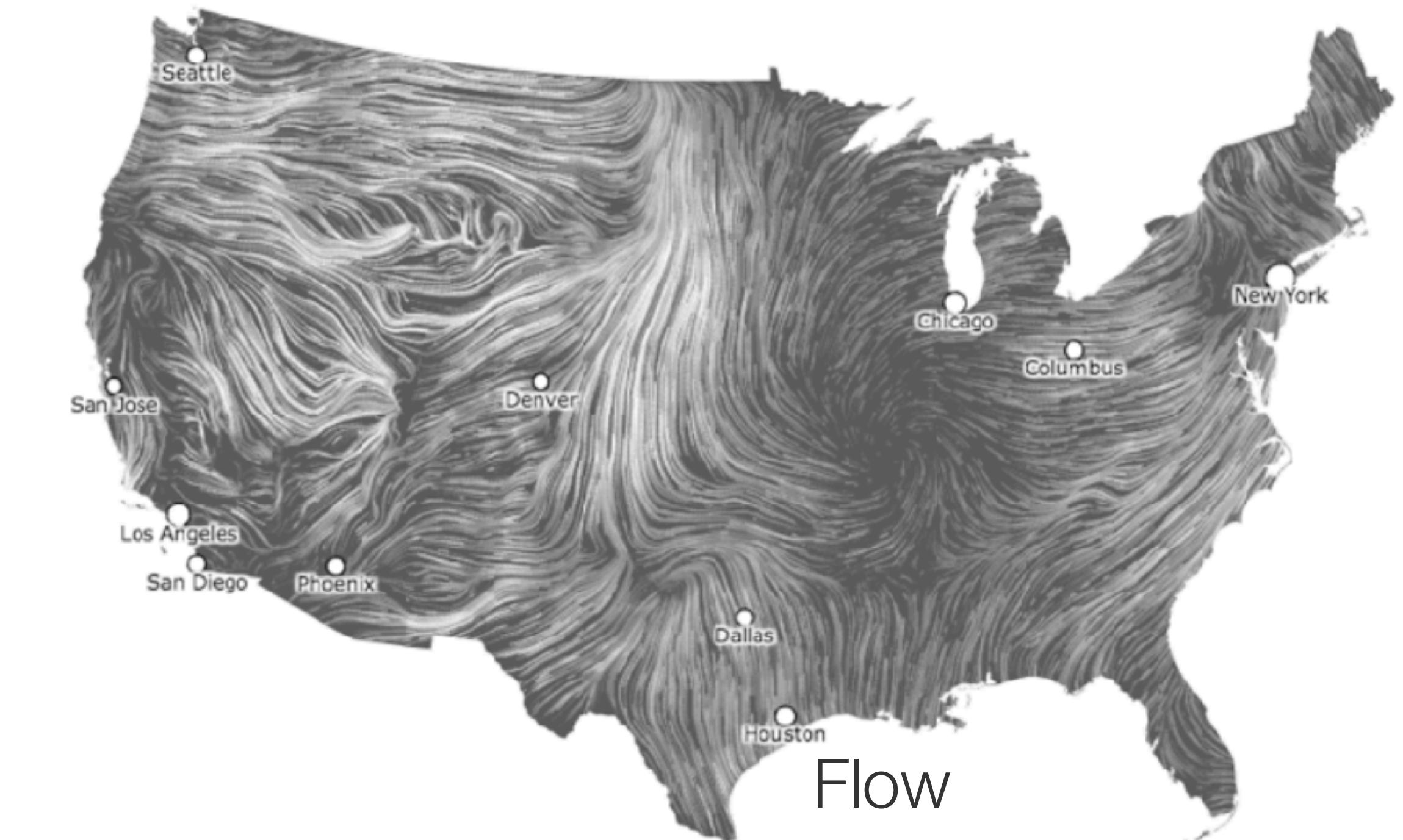
Choropleth



Symbol



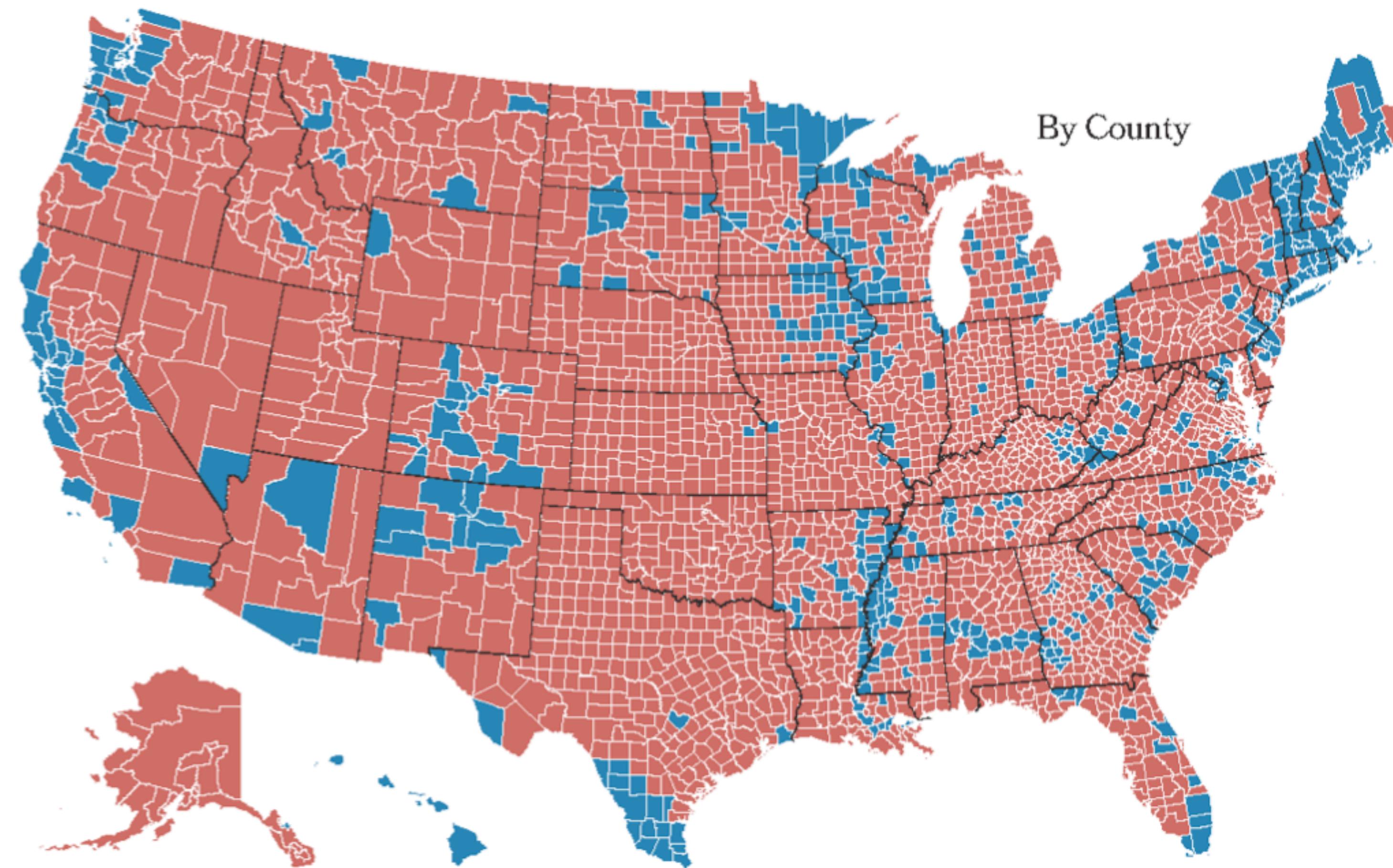
Grid



Flow

# Problem with Choropleth

Majority vote for Kerry or Bush in 2004 (Binary data)



[Matthew Ericson, NY Times]

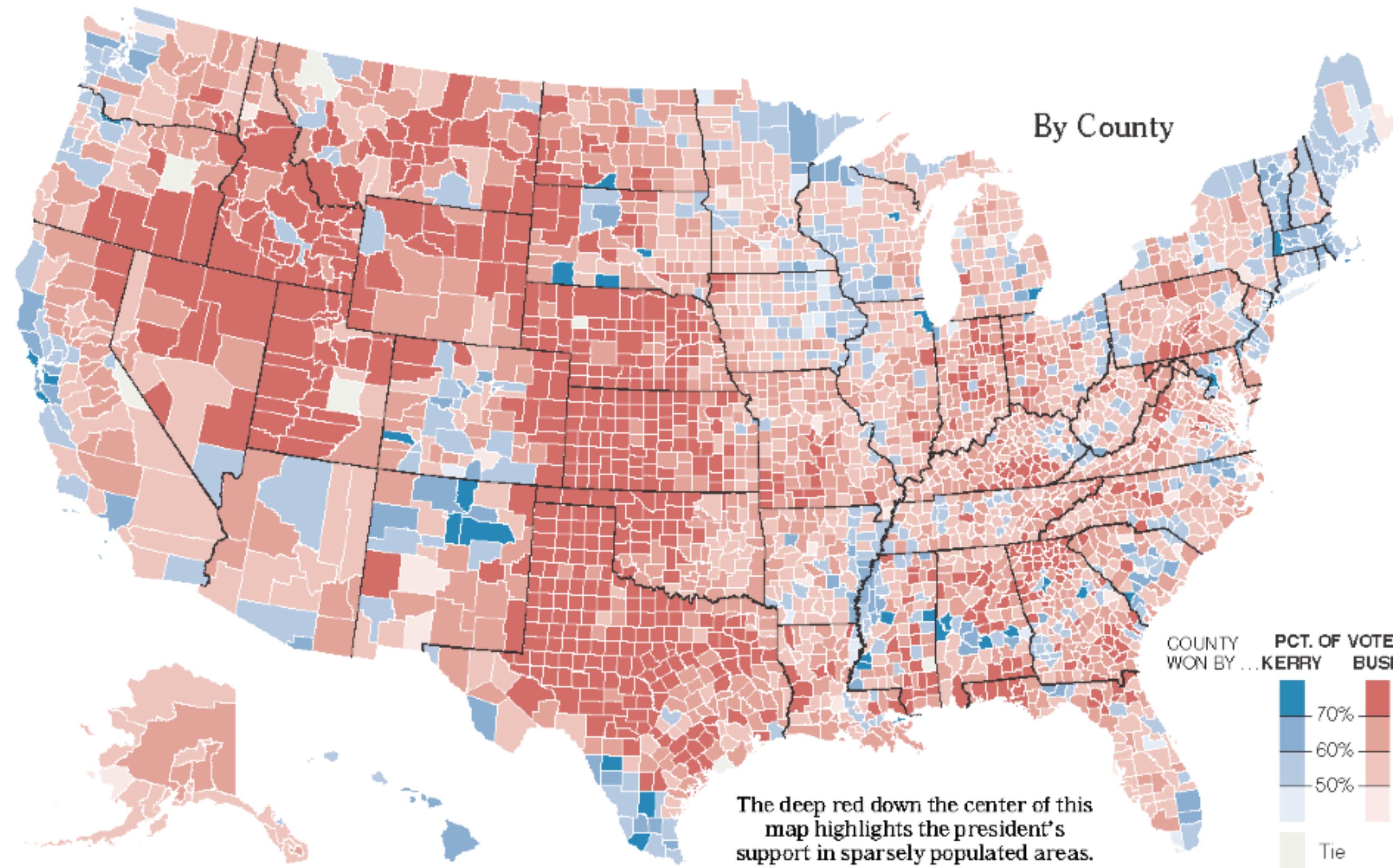
2004 Popular Vote



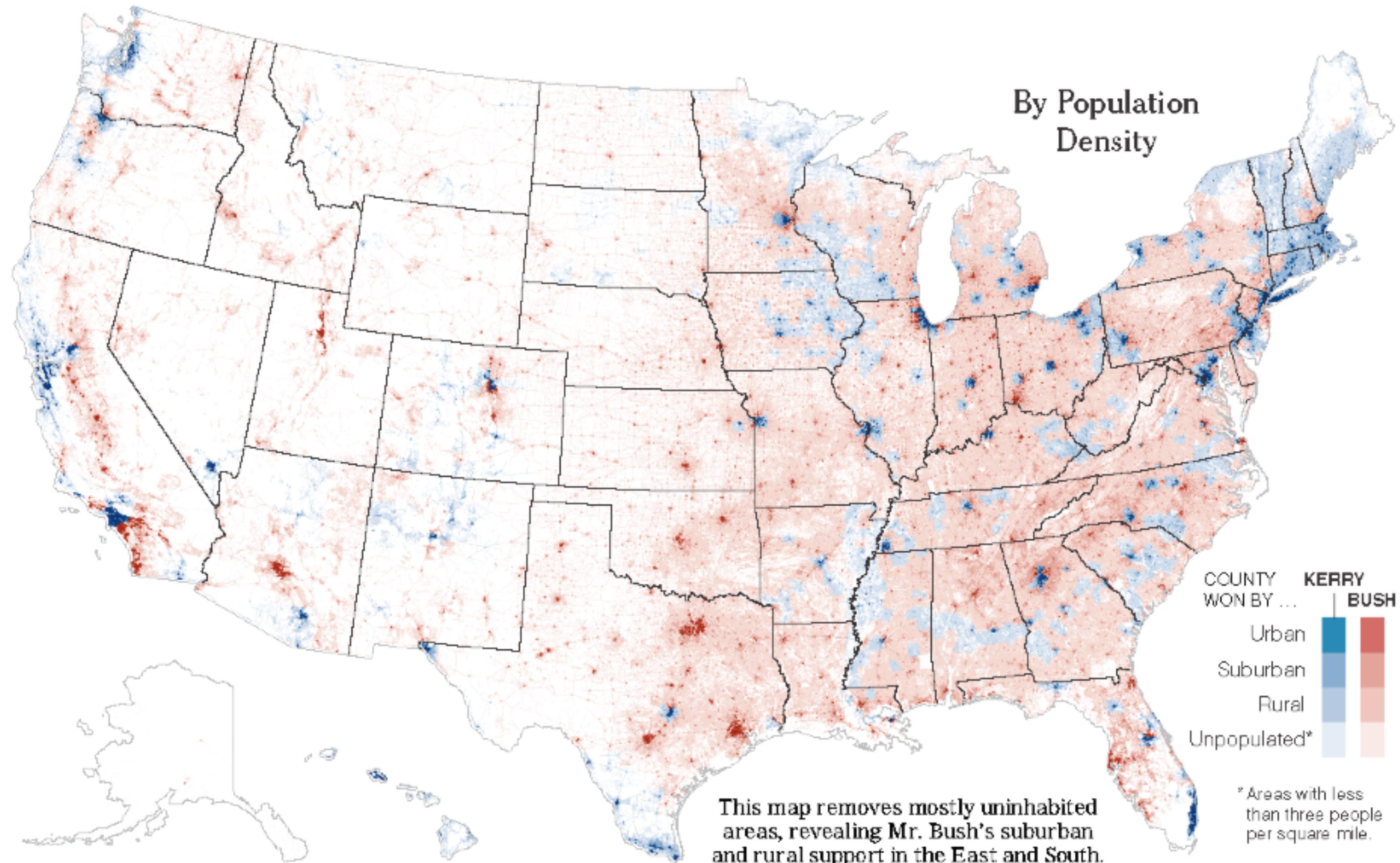
Amount of red and blue shown on map



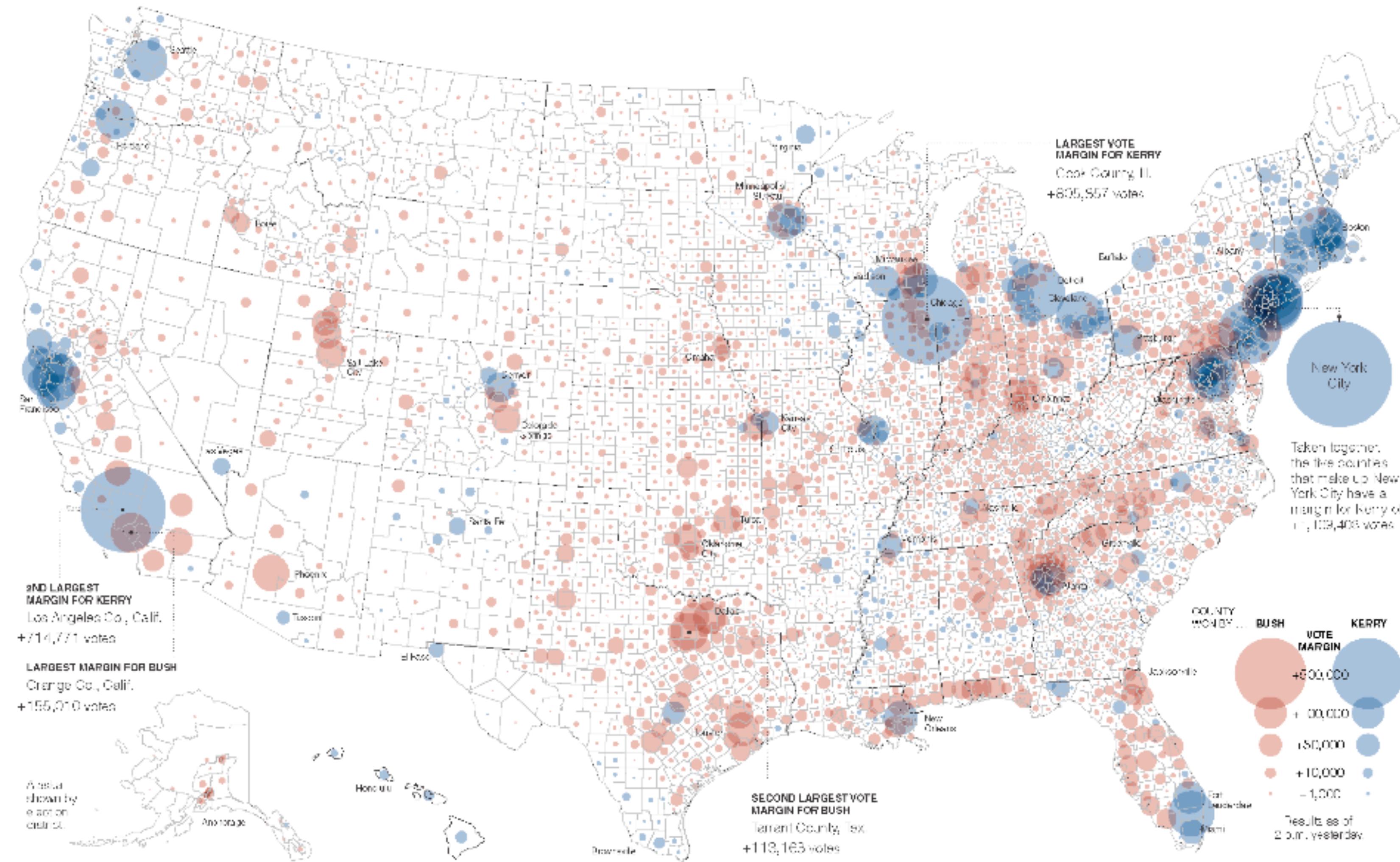
# Take into account for the number of votes.



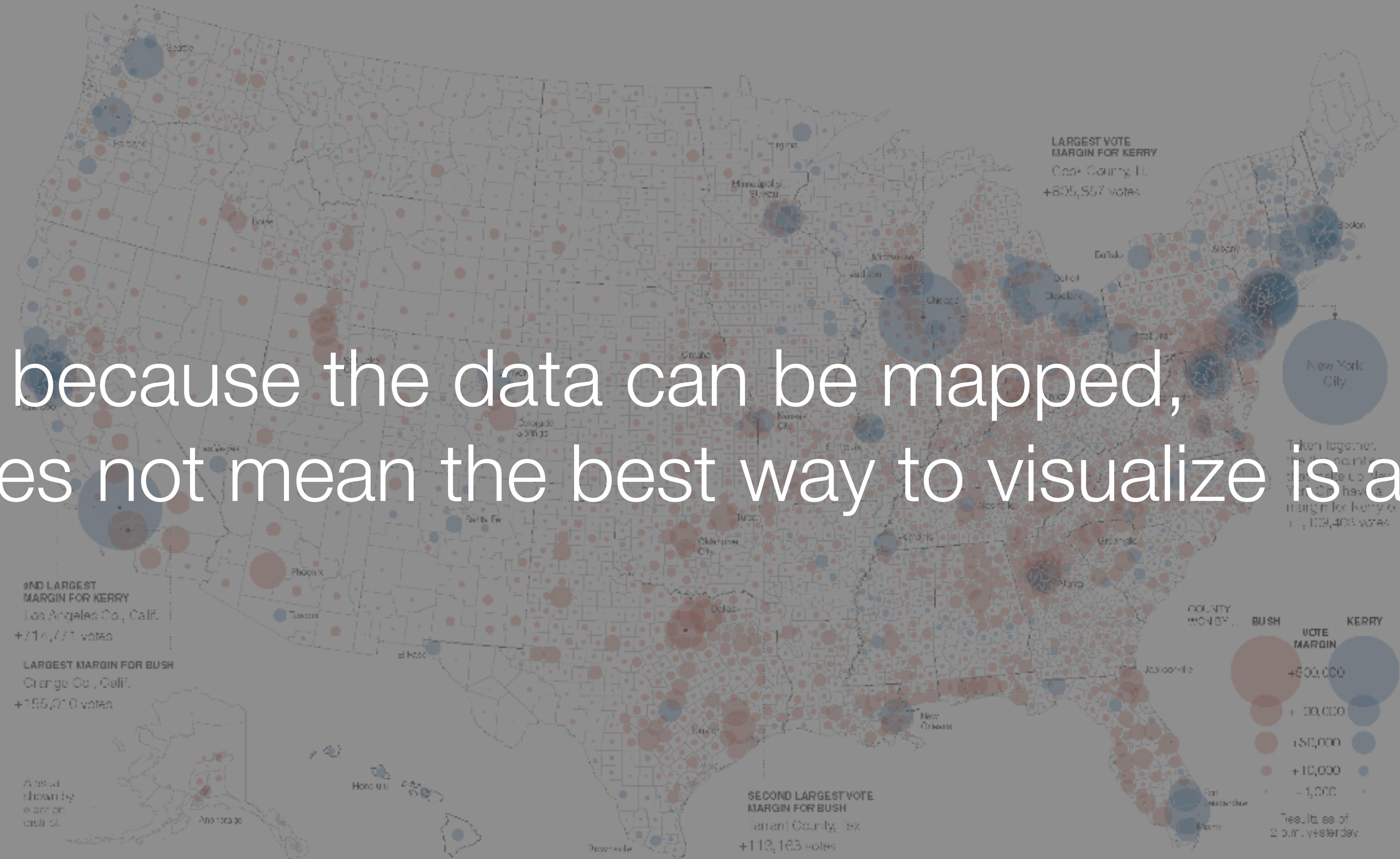
# Take into account for population density



# Symbol Map



The background of the slide is a grayscale map of the United States, specifically highlighting the state of New York. The map shows county-level election results, with many counties colored red and others blue. Overlaid on this map is a large, semi-transparent white rectangular area that contains the main text of the slide. In the top right corner, there is a circular callout with the text "New York City" and some smaller, partially visible text below it.



# Map Projections



## THE TRUE SIZE OF ...

eg...Ghana



About   Clear Map



North  
Atlantic  
Ocean

United States  
(Contiguous 48)

Total size in km<sup>2</sup> (sq mi) 9,526,461  
(3,678,190) – 9,826,675 (3,794,100)

**United States (Contiguous 48)**

Total size in km<sup>2</sup> (sq mi) 9,526,468  
(3,678,190) – 9,826,675 (3,794,100)

Iceland

Sweden

Norway

Finland

Russia

United Kingdom

Ireland

Denmark

Poland

Belarus

Ukraine

Austria

Germany

France

Italy

Spain

Portugal

Morocco

Tunisia

Algeria

Western Sahara

Mauritania

Mali

Niger

Guinea

Burkina Faso

Nigeria

Ghana

Gulf of Guinea

DR Congo

Gabon

Angola

Zambia

Namibia

Botswana

Zimbabwe

Tanzania

South Sudan

Ethiopia

Somalia

Kenya

Mozambique

Madagascar

South Africa

Kazakhstan

Uzbekistan

Kyrgyzstan

Turkmenistan

Iran

Afghanistan

Pakistan

Nepal

India

Oman

Arabian Sea

Laccadive Sea

Myanmar (Burma)

Thailand

Vietnam

Gulf of Thailand

Malaysia

Indonesia

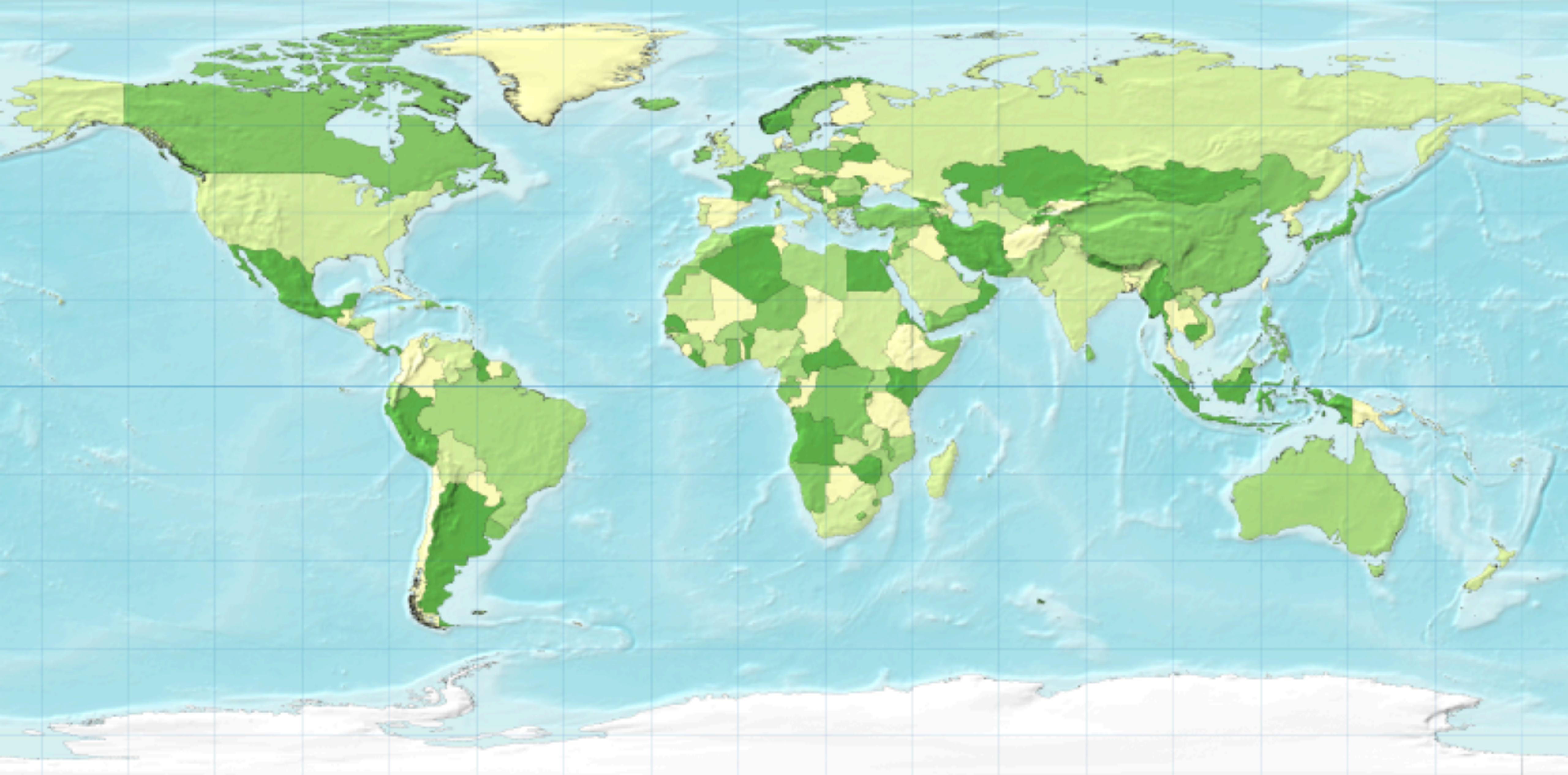
South China Sea

Bay of Bengal

Indian Ocean

South Atlantic Ocean

<http://thetruesize.com/>

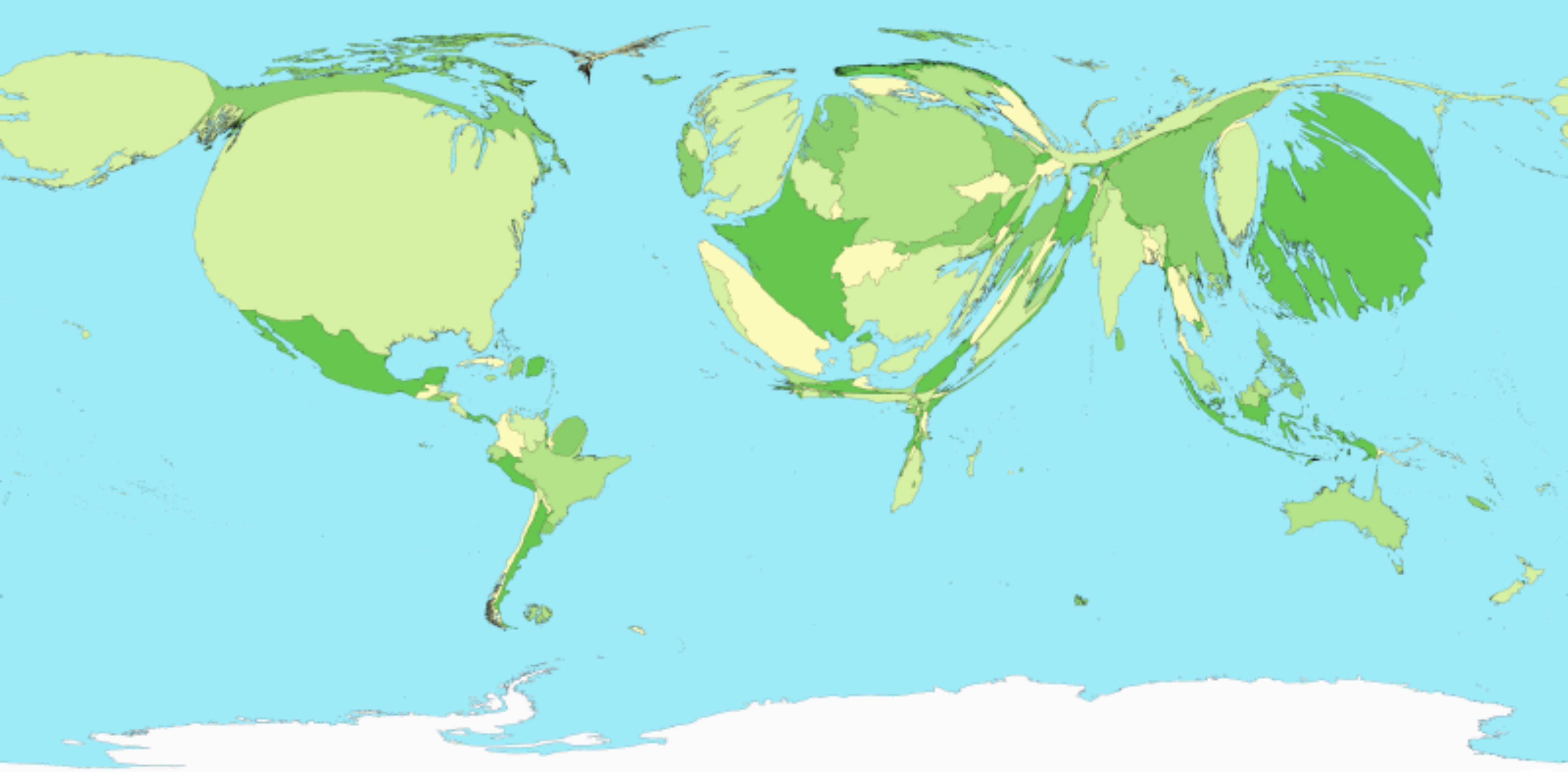


Ordinary map

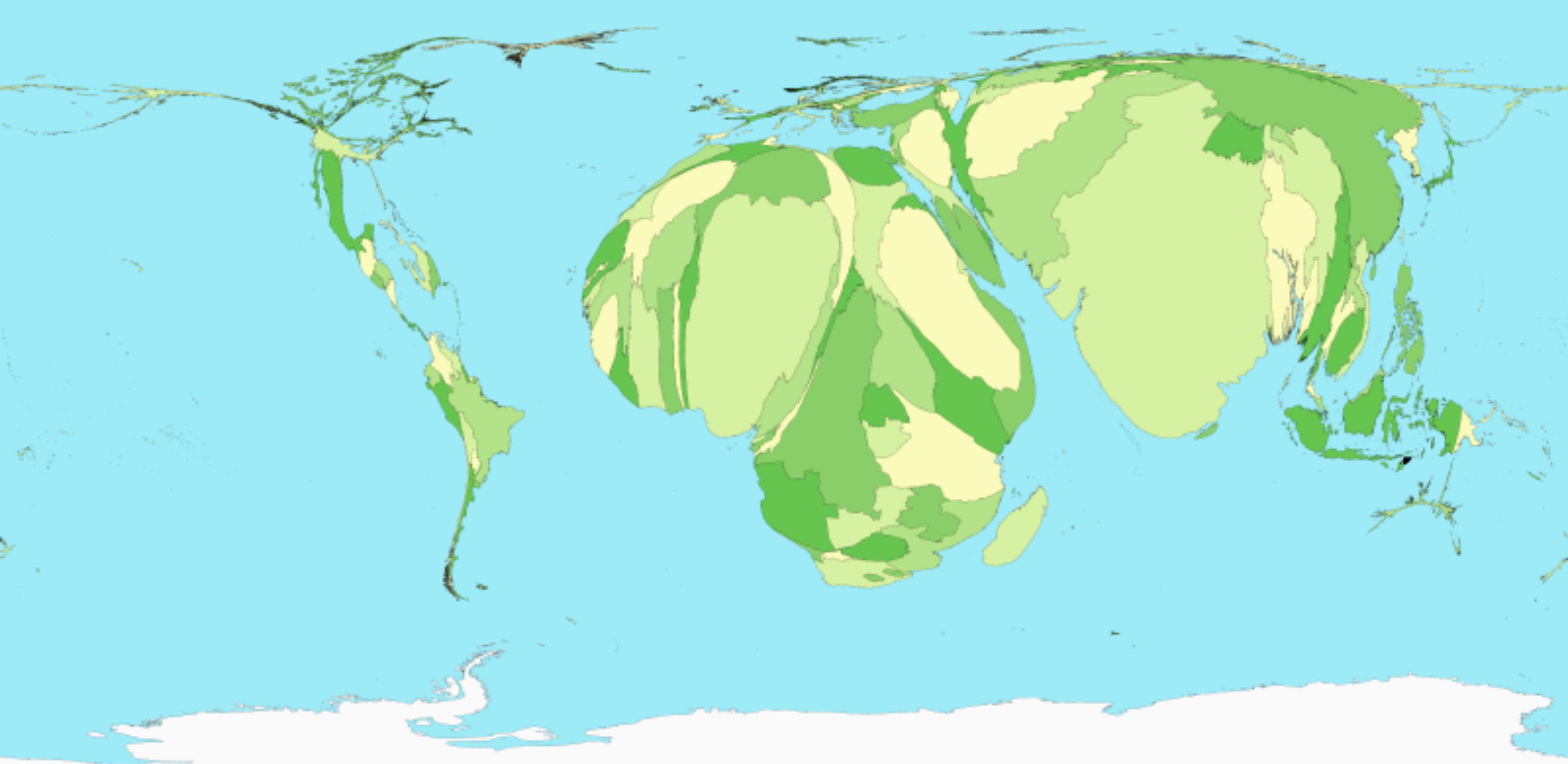
[\[http://www-personal.umich.edu/~mejn/cartograms/\]](http://www-personal.umich.edu/~mejn/cartograms/)



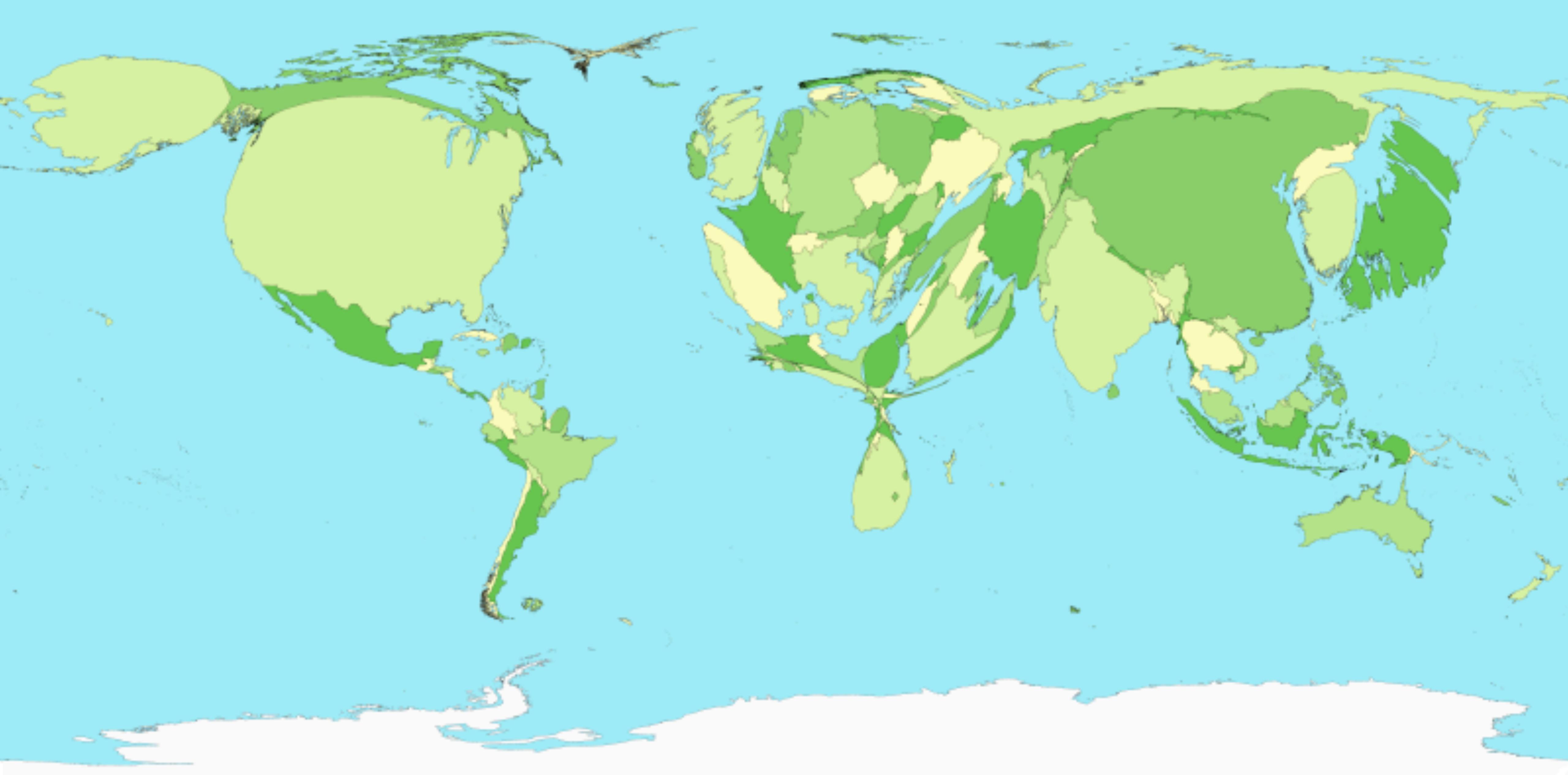
# Population



Gross domestic product



Child mortality



Greenhouse gas emissions

# Problem with Rainbow Colormap

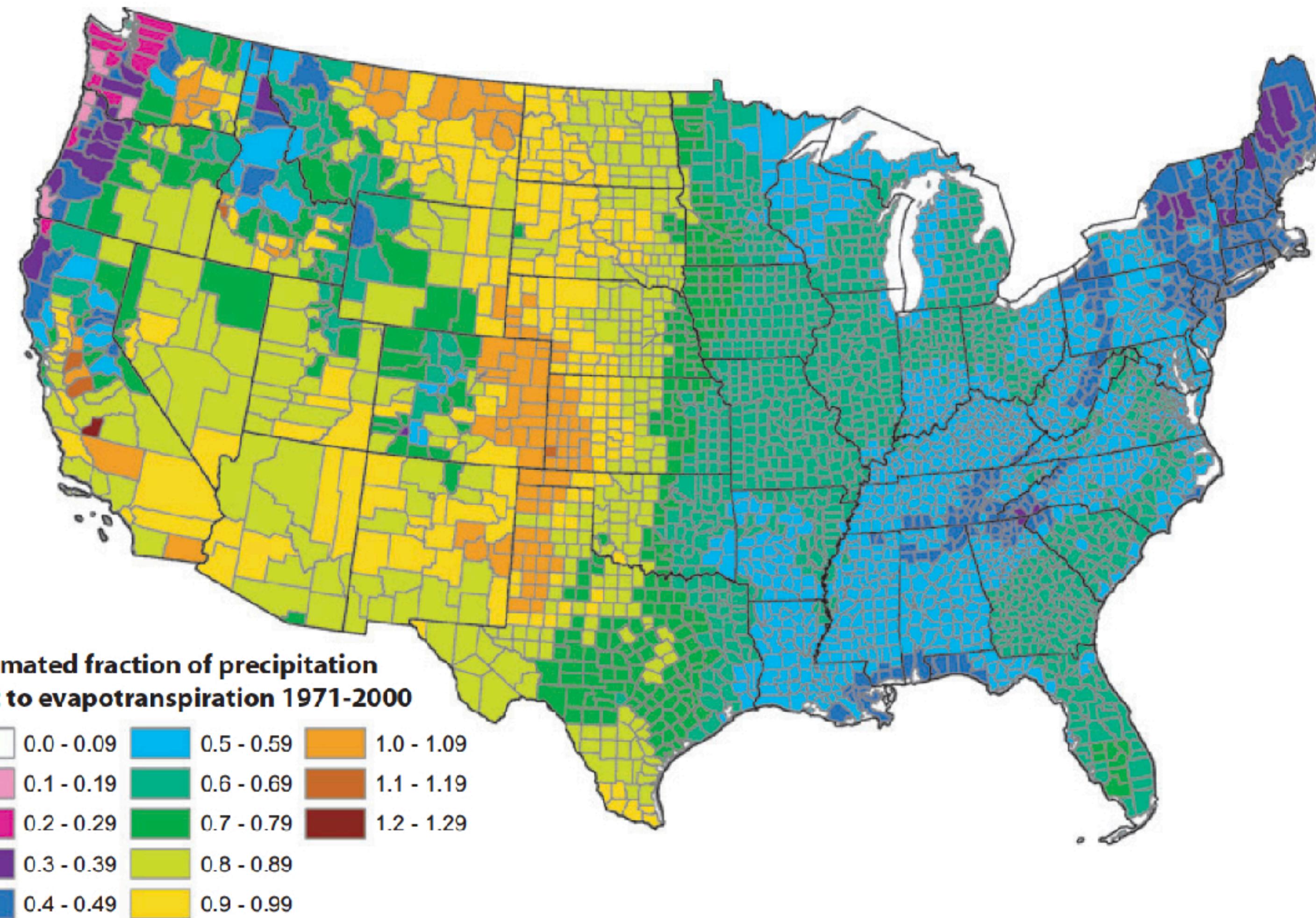


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions  $>1$  are agricultural counties that either import surface water or mine deep groundwater.

Number of data classes: 3



how to use | updates | downloads | credits

Nature of your data:



sequential  diverging  qualitative

Pick a color scheme:

Multi-hue:



Single hue:



Only show:



- colorblind safe
- print friendly
- photocopy safe

Context:



- roads
- cities
- borders



Background:

solid color

terrain



color transparency

i

3-class Greens

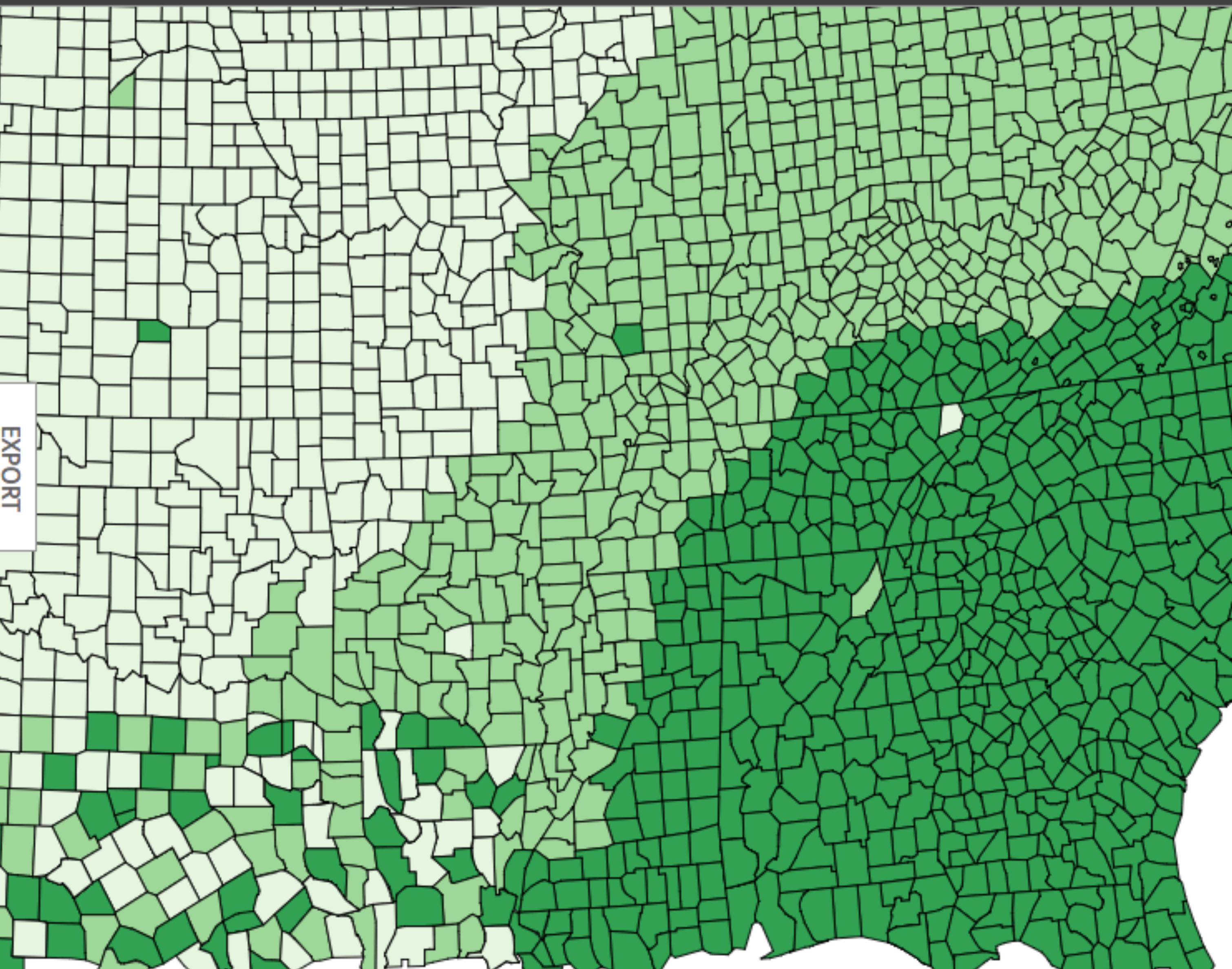


HEX



- #e5f5e0
- #a1d99b
- #31a354

EXPORT

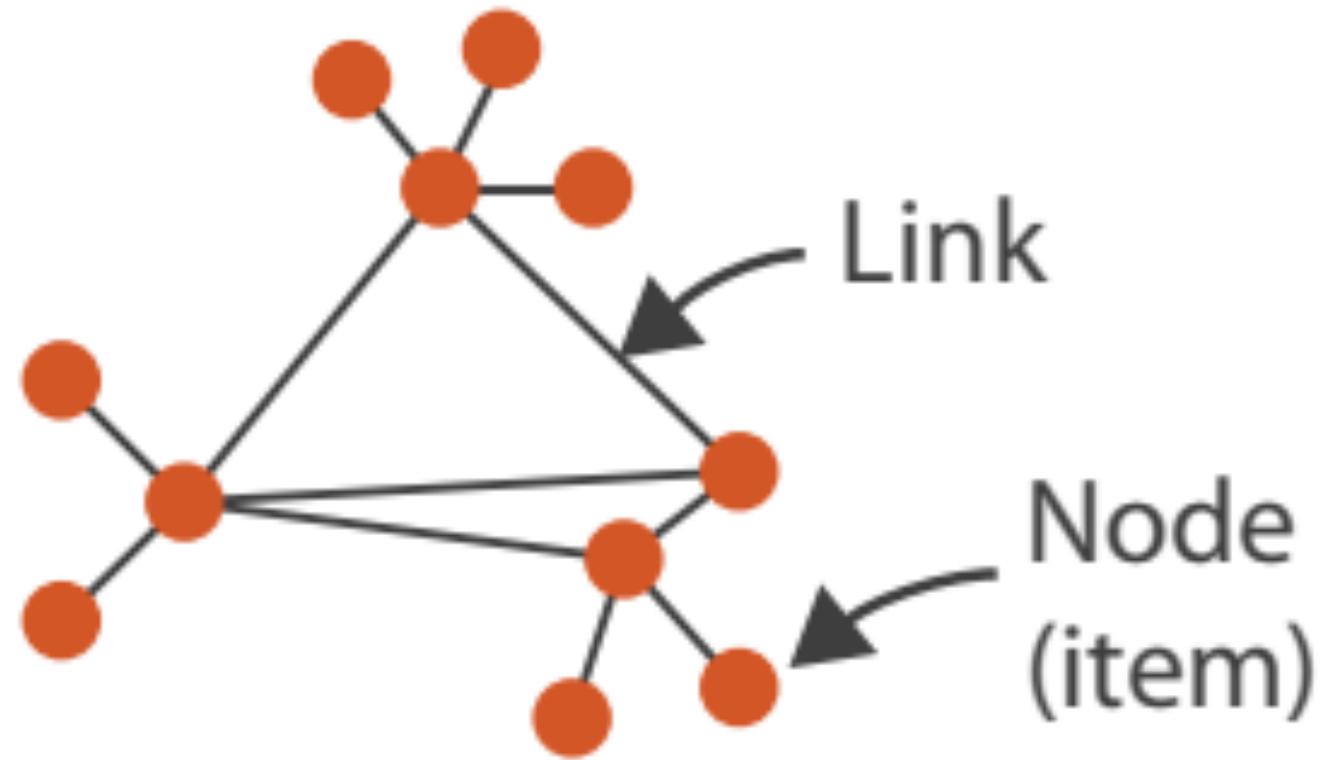


COLORBREWER 2.0  
color advice for cartography

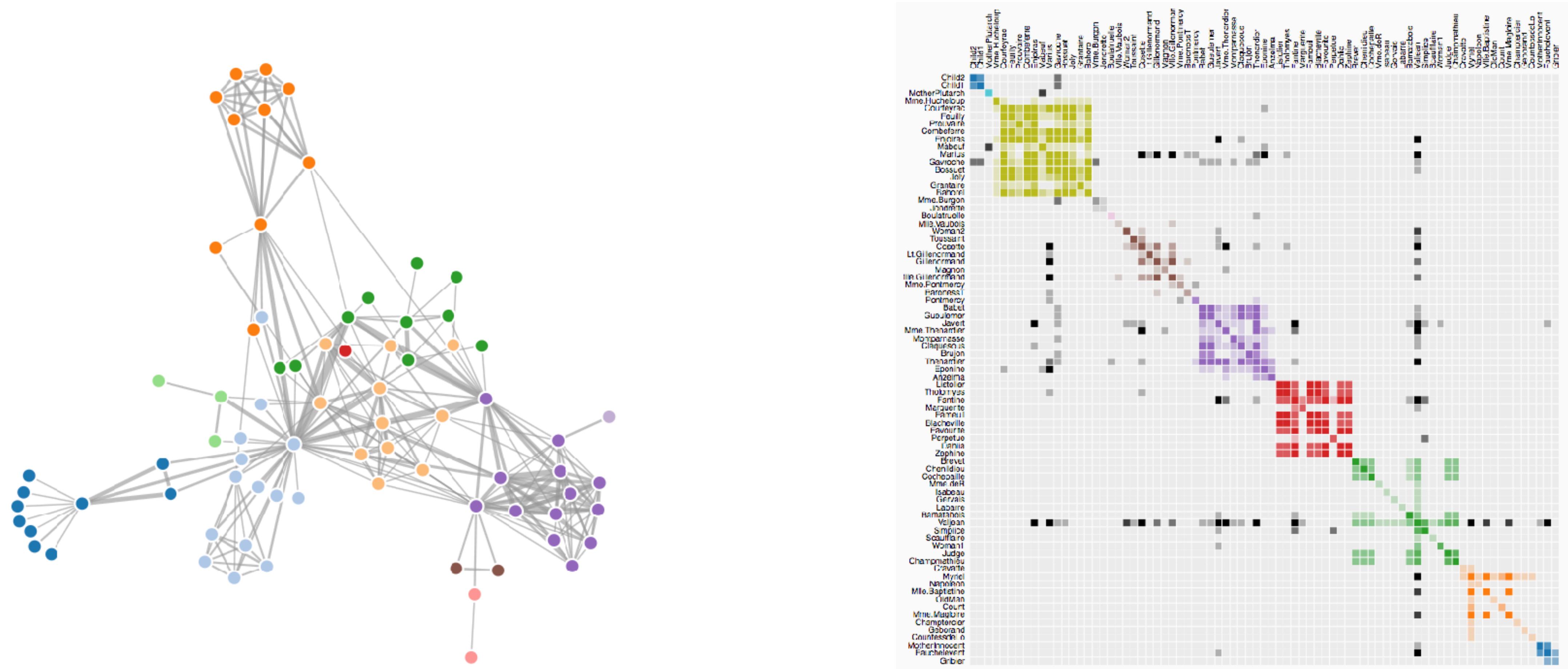
# Networks & Trees

# Networks

Model **relations** among data  
**Nodes** and **links**

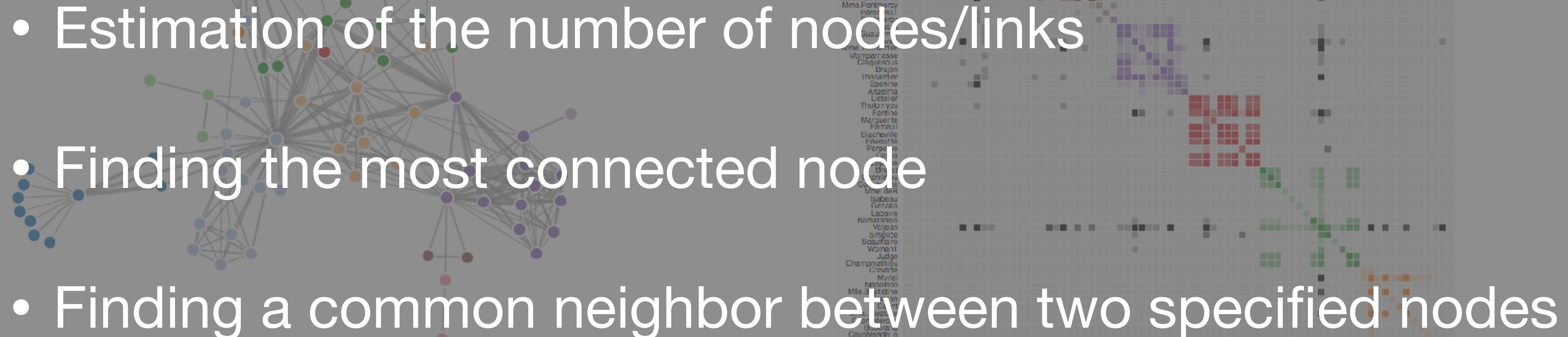


# Node-Link Diagram vs Matrix Diagram



# Les Misérables Character Co-occurrence

Matrix Diagram is found more readable  
in most tasks except path finding. [M. Ghoniem 2004]

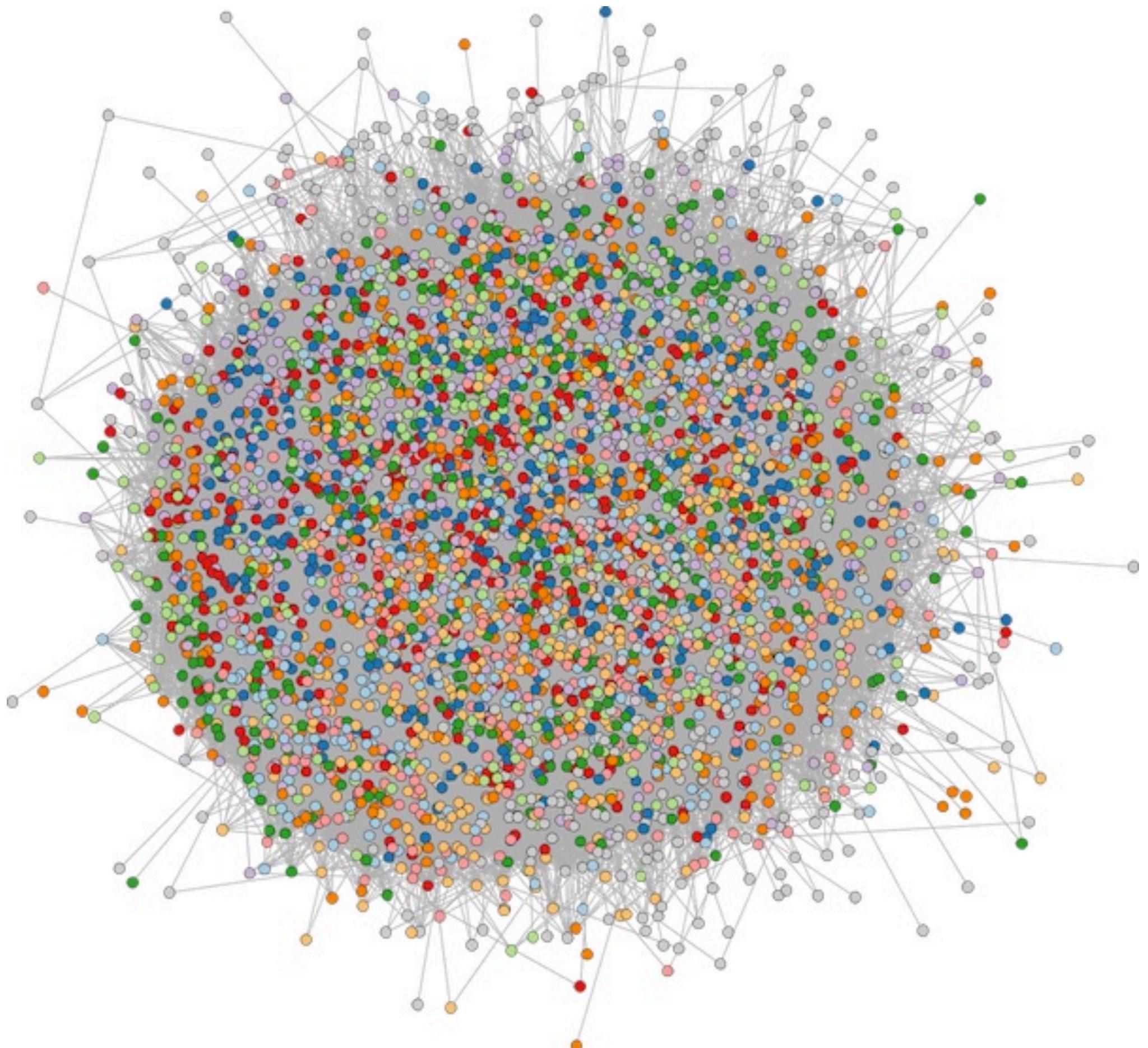
- Estimation of the number of nodes/links
  - Finding the most connected node
  - Finding a common neighbor between two specified nodesThe background features a network graph with numerous nodes of various colors (green, blue, orange, purple) connected by a dense web of gray lines. To the right of the graph is a vertical column of names, some of which are also present as nodes in the network. To the far right is a large, square grid composed of smaller squares in shades of gray, red, green, and orange, forming a pattern that suggests a heatmap or a matrix representation of data.

# Node-Link Diagram

# Matrix Diagram

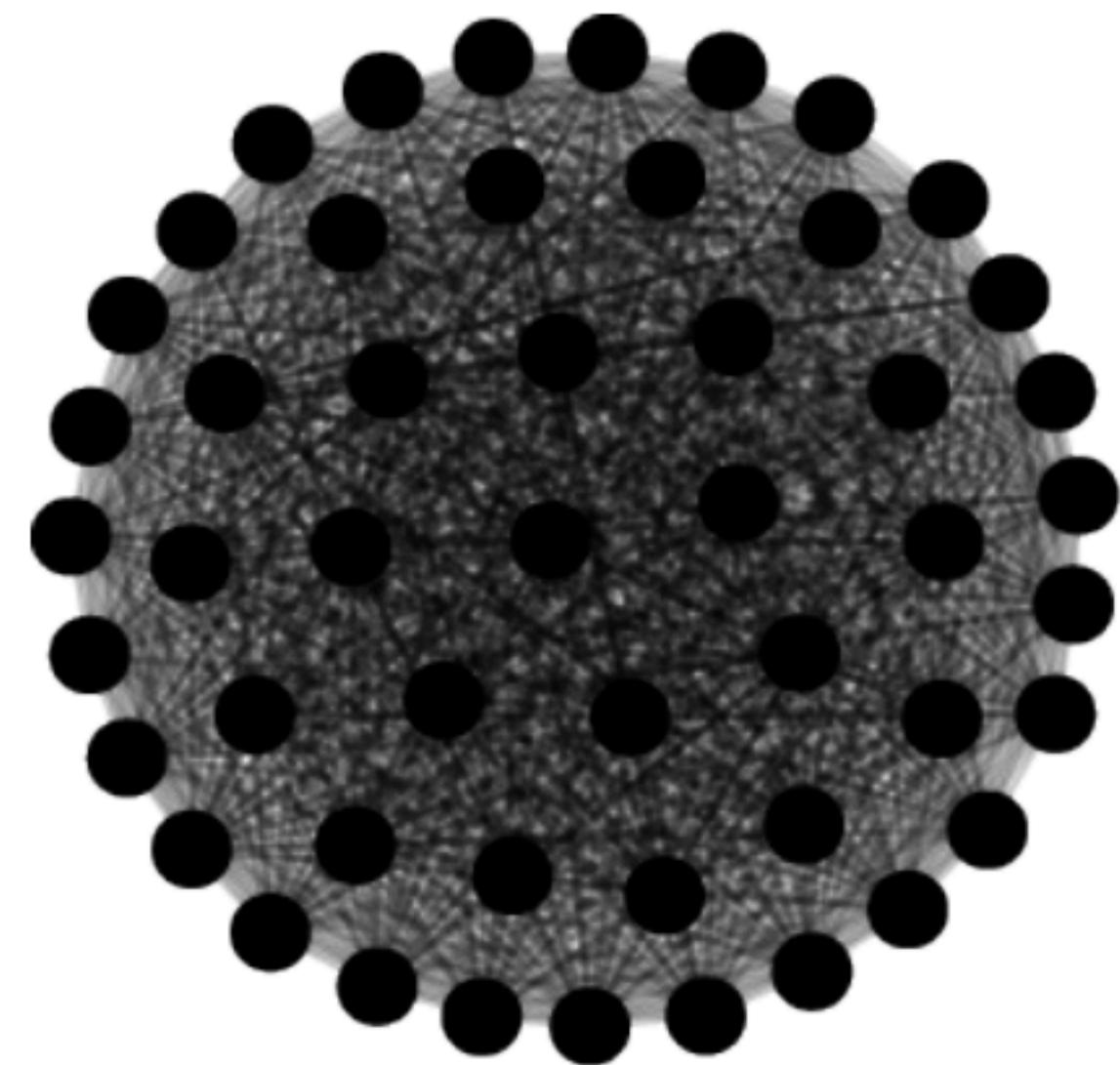
# Problem with Node-link diagram

A giant **hairball**: clutter when the number of nodes and links becomes too high

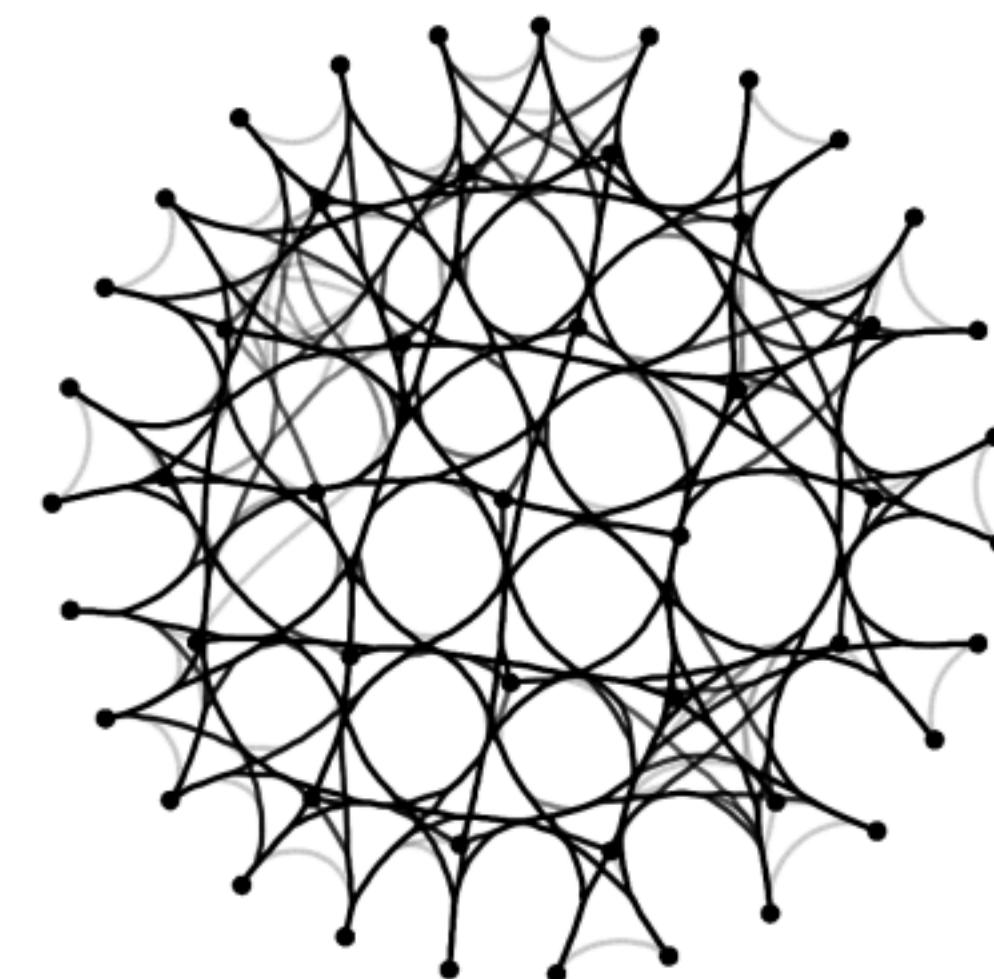


[A. Nocaj 2015]

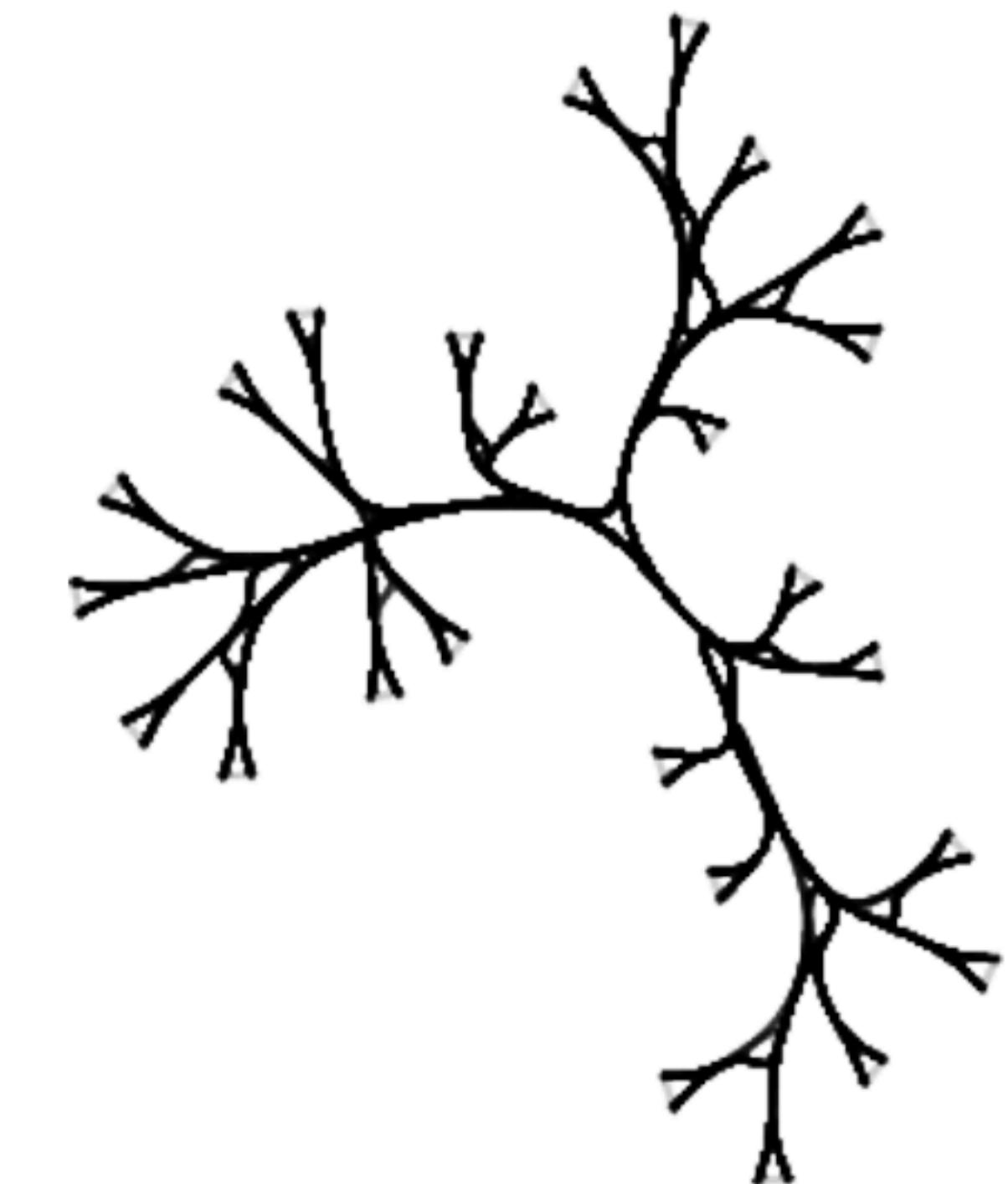
Original Network



Edge Bundling

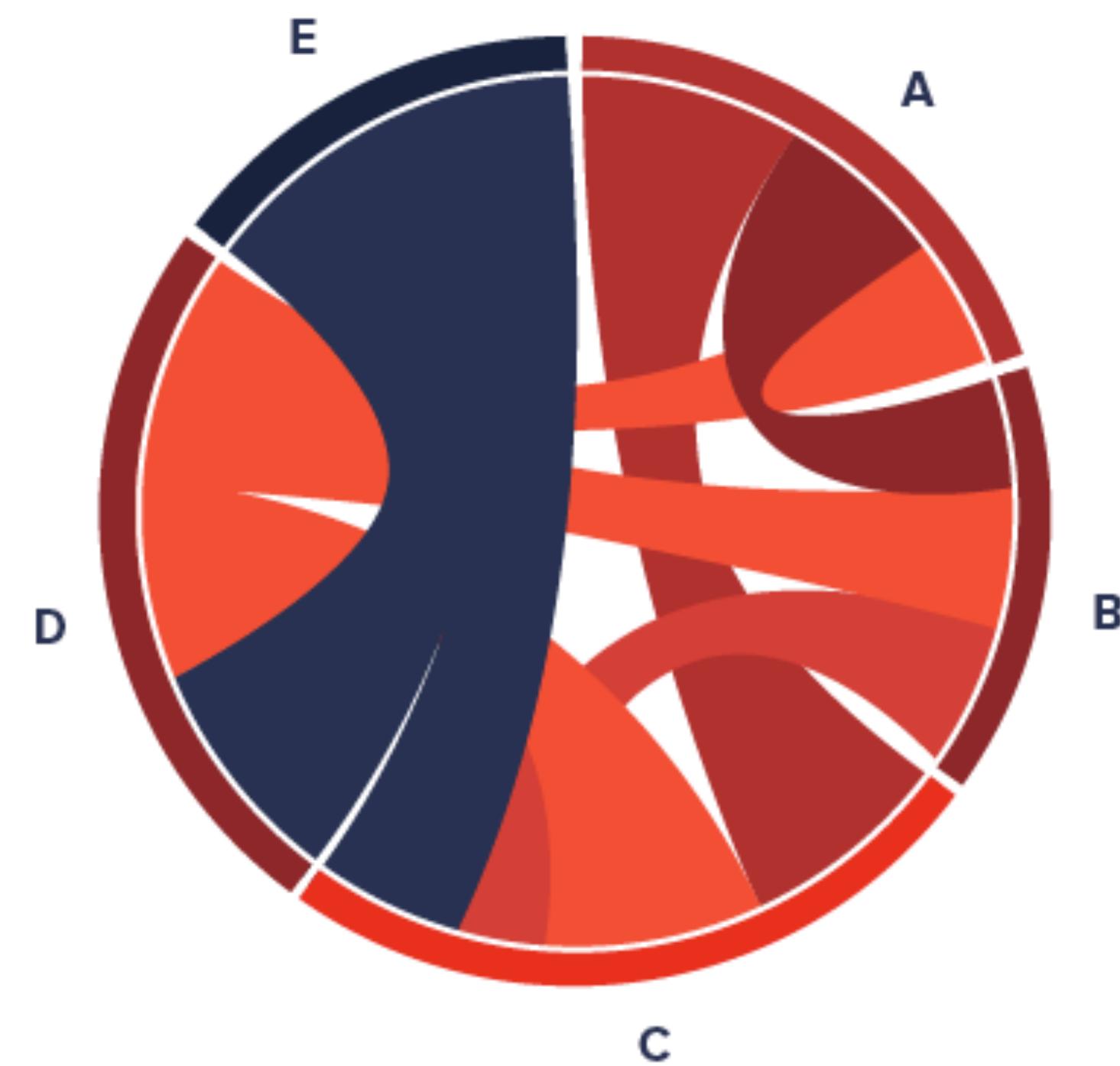


Confluent Drawing

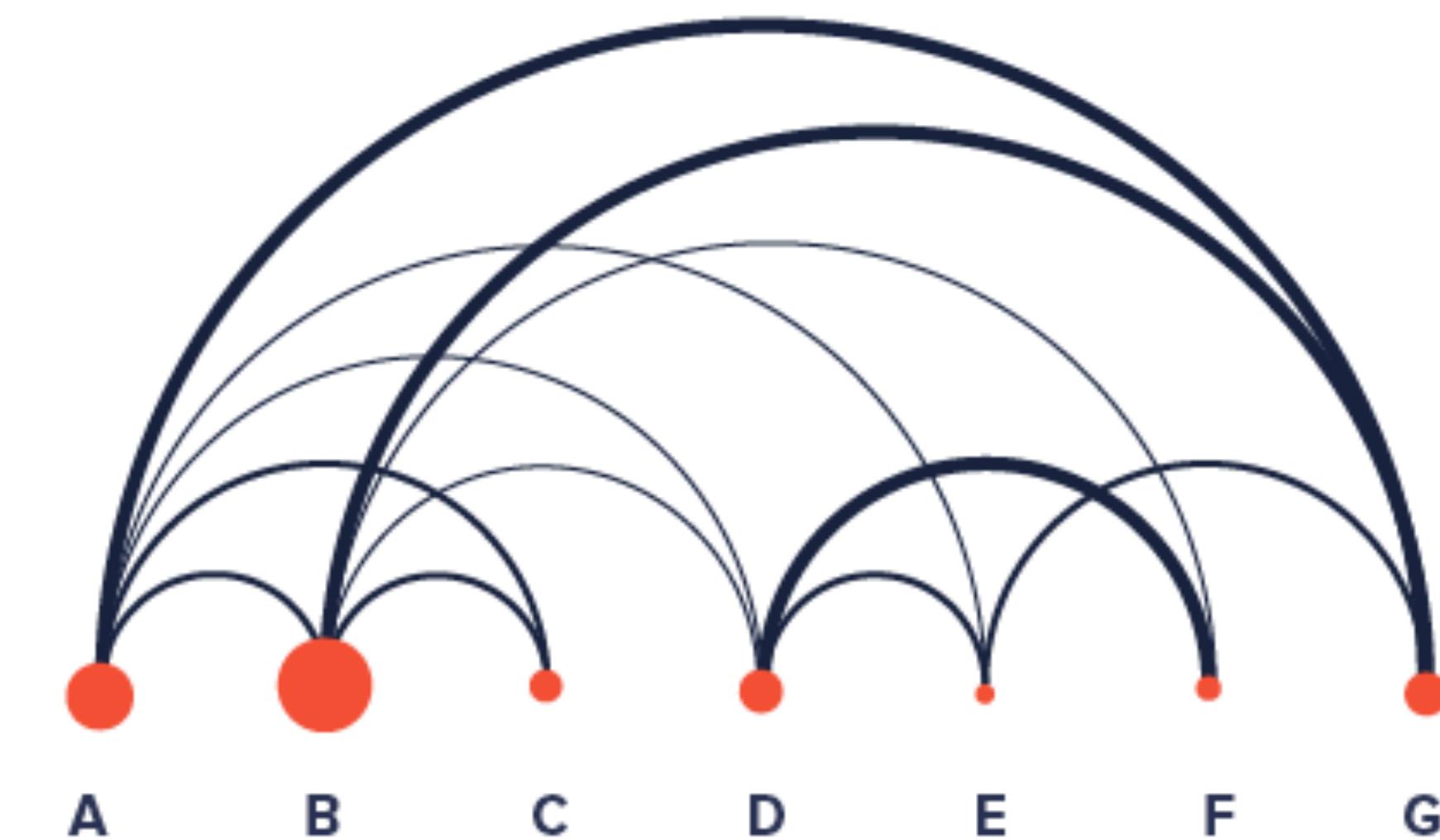


[B. Bach 2017]

# Alternatives. Are these better?



Chord Diagram



Arc Diagram

# Networks & Trees

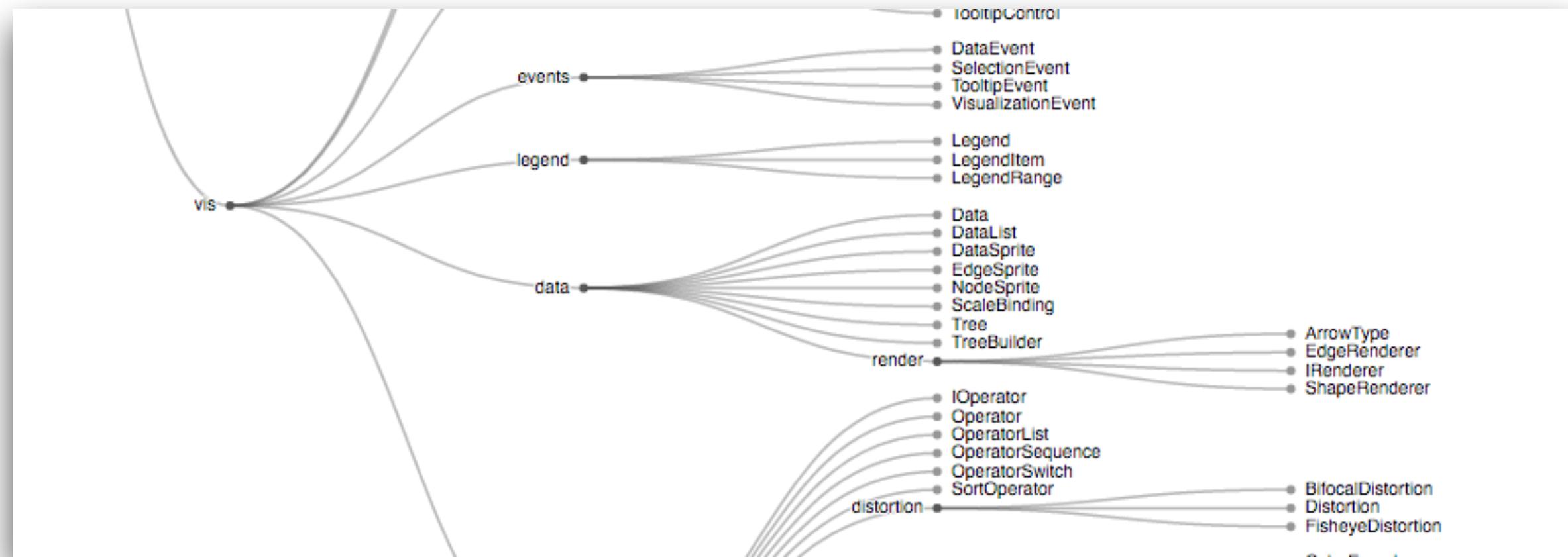
# Trees

Graphs with **hierarchical** structure

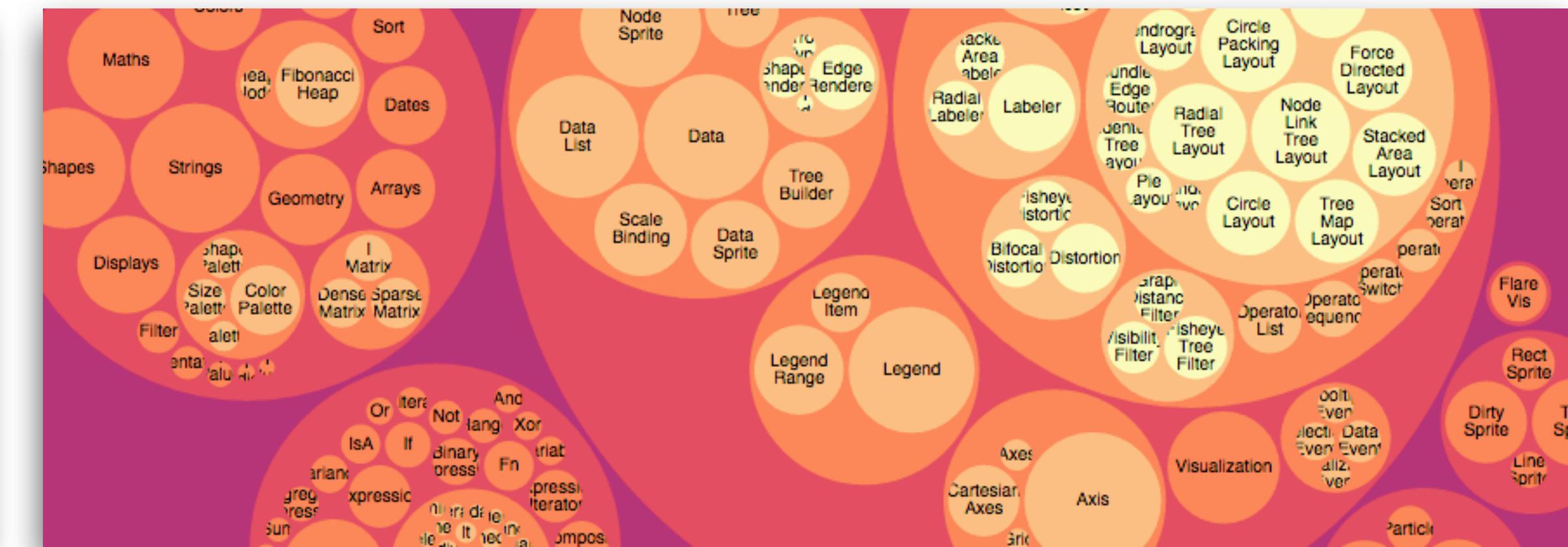
Nodes as **parents** and **children**



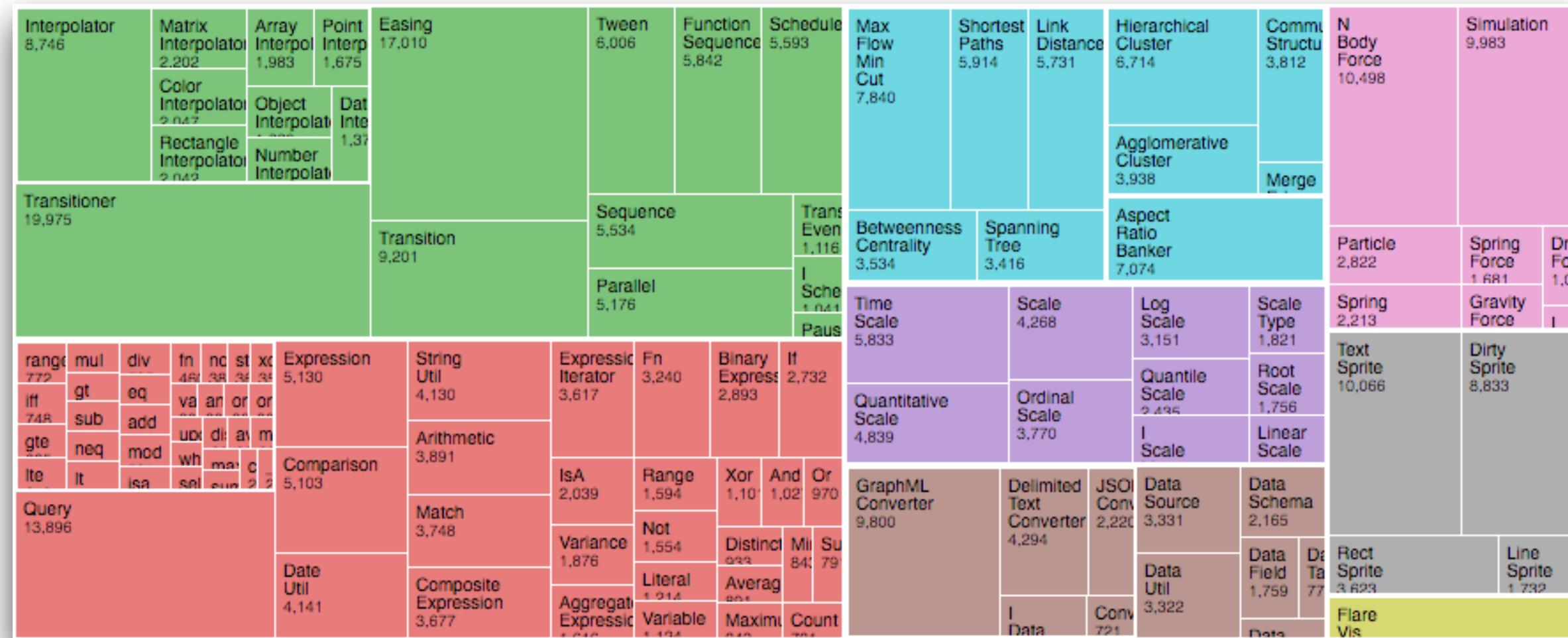
# Dendogram



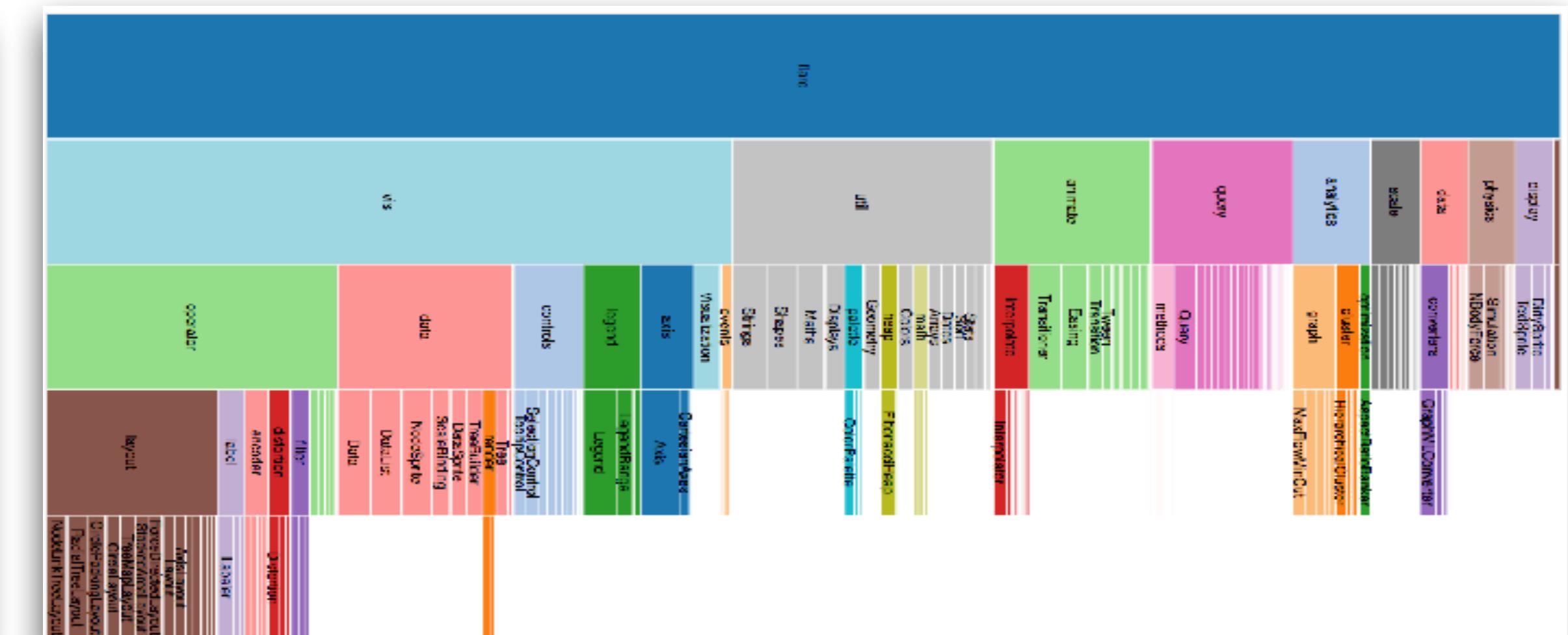
# Circle Packing



# Treemap



# Icicle



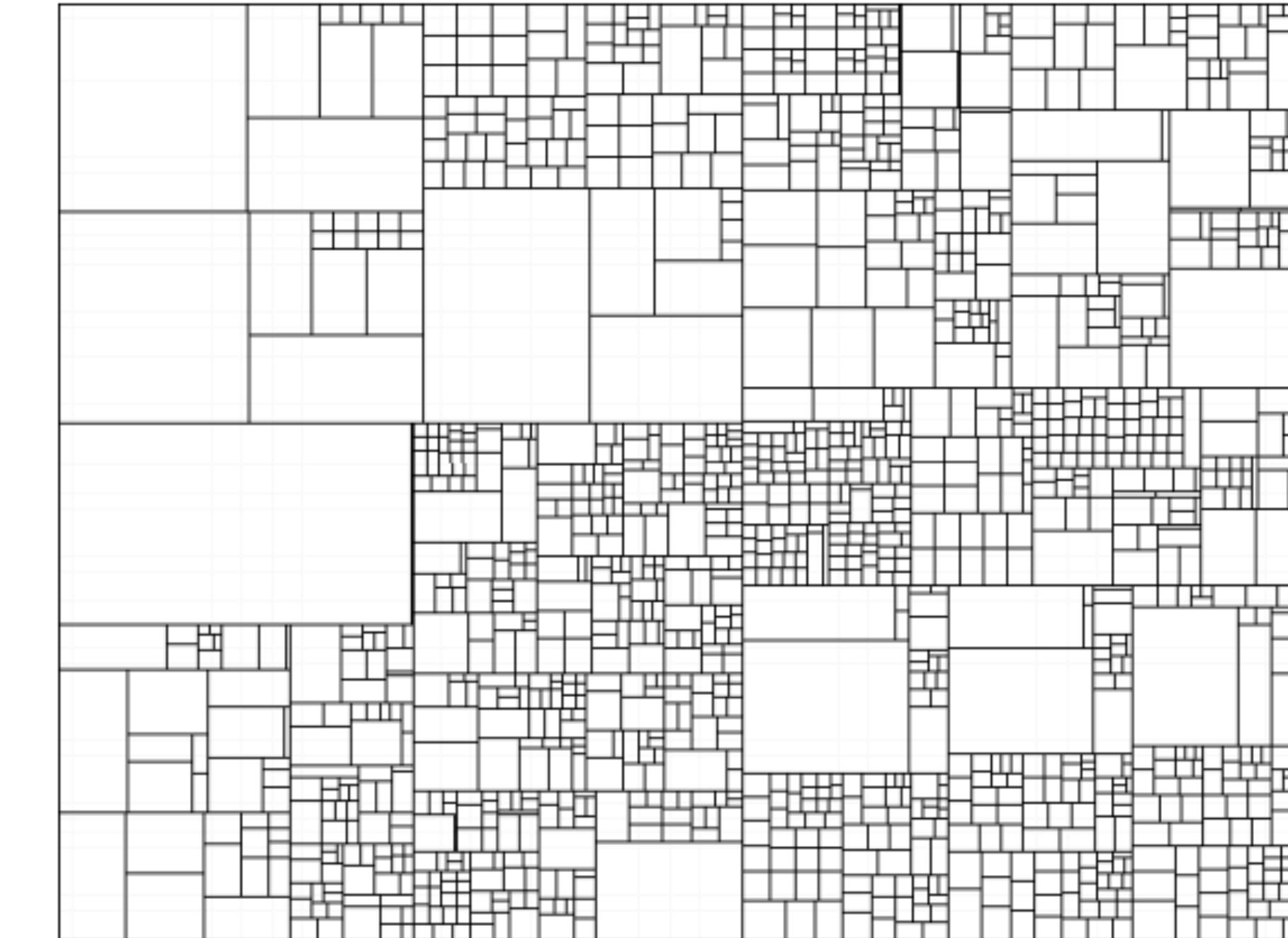
# Squarified Treemaps

“The original treemap method often gives thin, elongated rectangles. As a result, rectangles are difficult to compare and to select.”

[M. Bruls et al 2000]

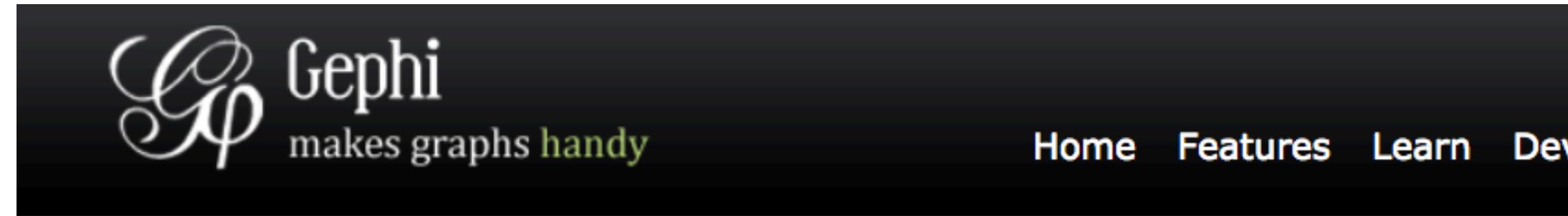


Before



After

# Tools for Graph Analysis



## The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

Runs on Windows, Mac OS X and Linux.

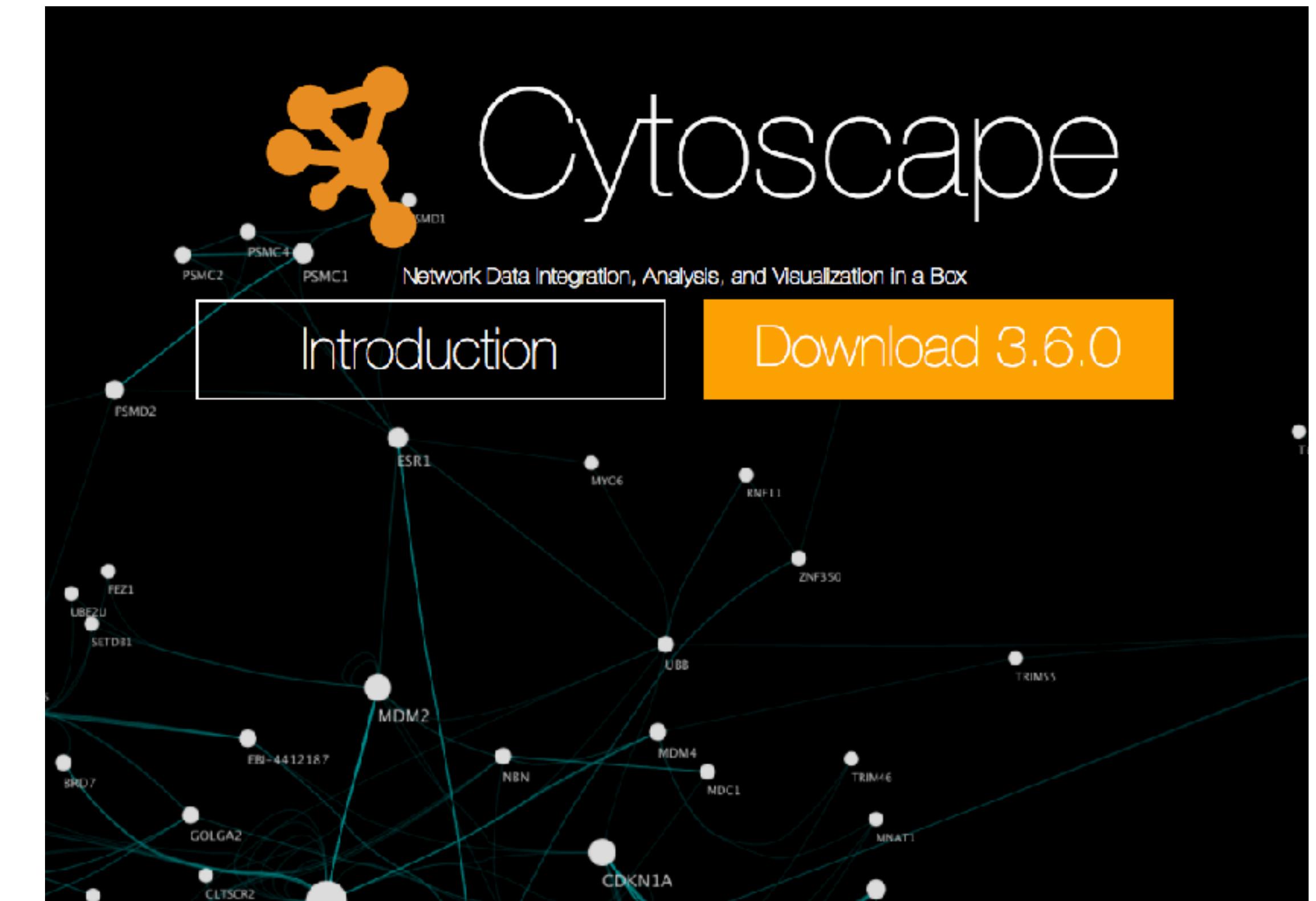
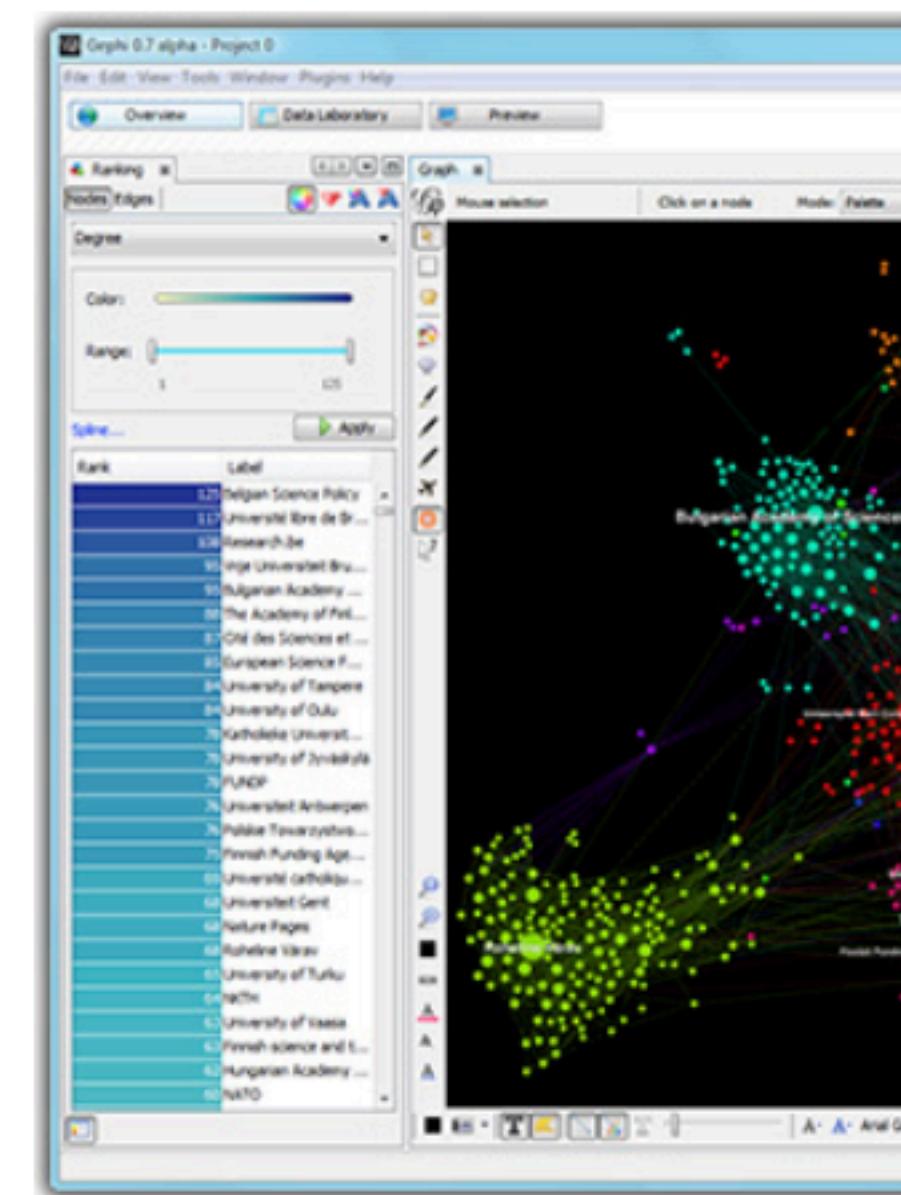
[Learn More on Gephi Platform »](#)

 Download FREE  
Gephi 0.9.2

[Release Notes](#) | [System Requirements](#)

► [Features](#)  
► [Quick start](#)

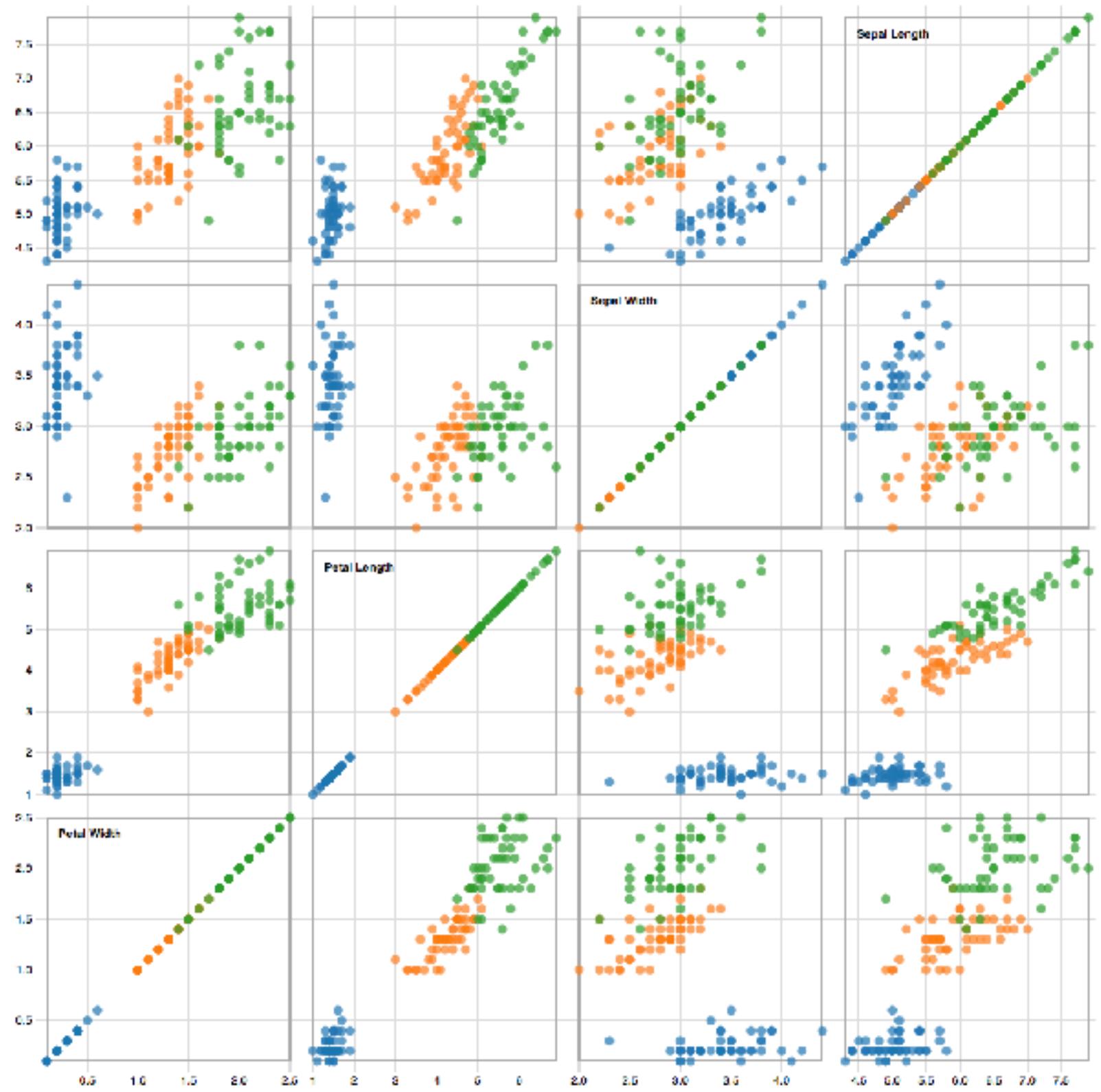
► [Screenshots](#)  
► [Videos](#)



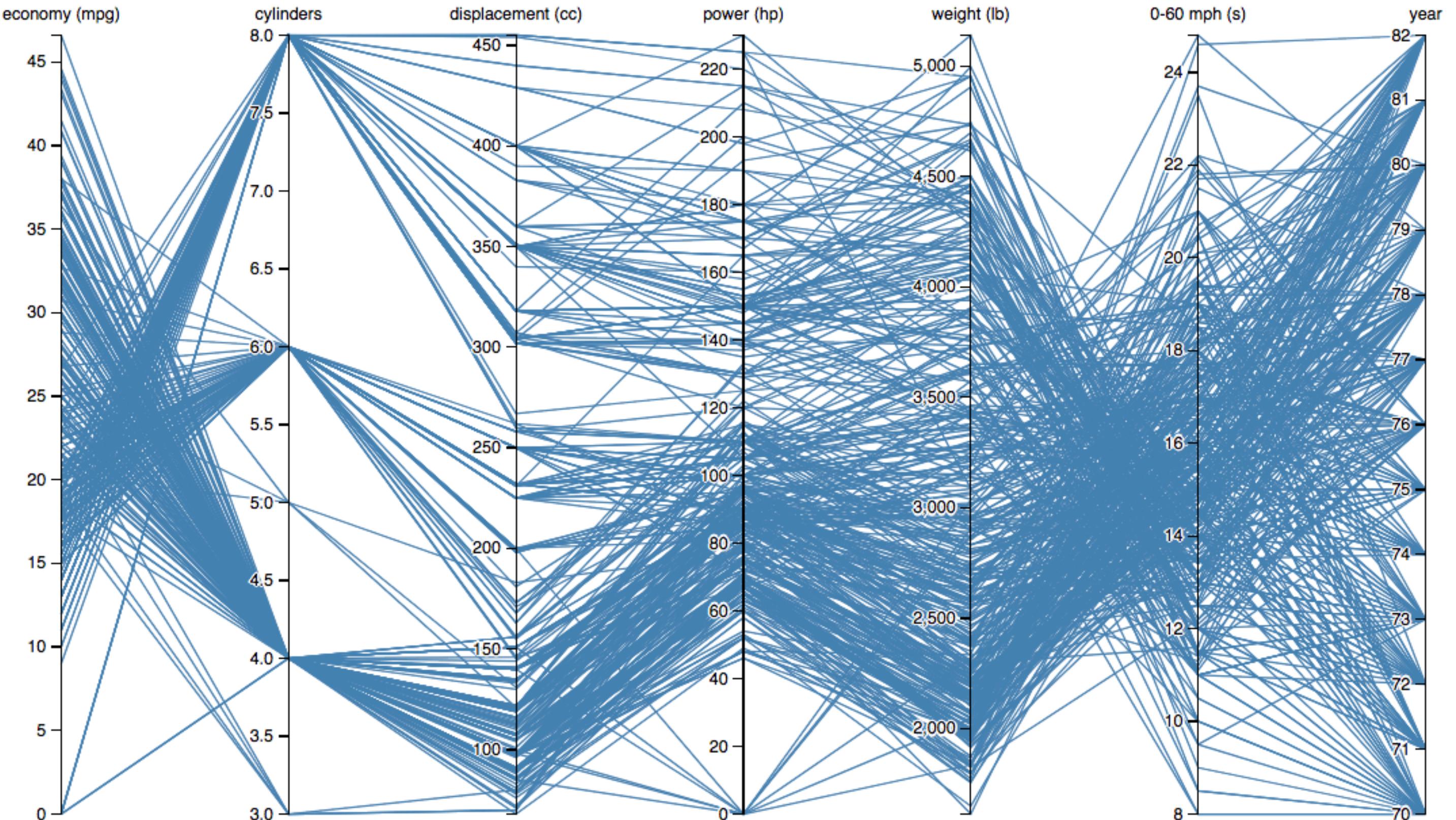
## Gephi

## Cytoscape

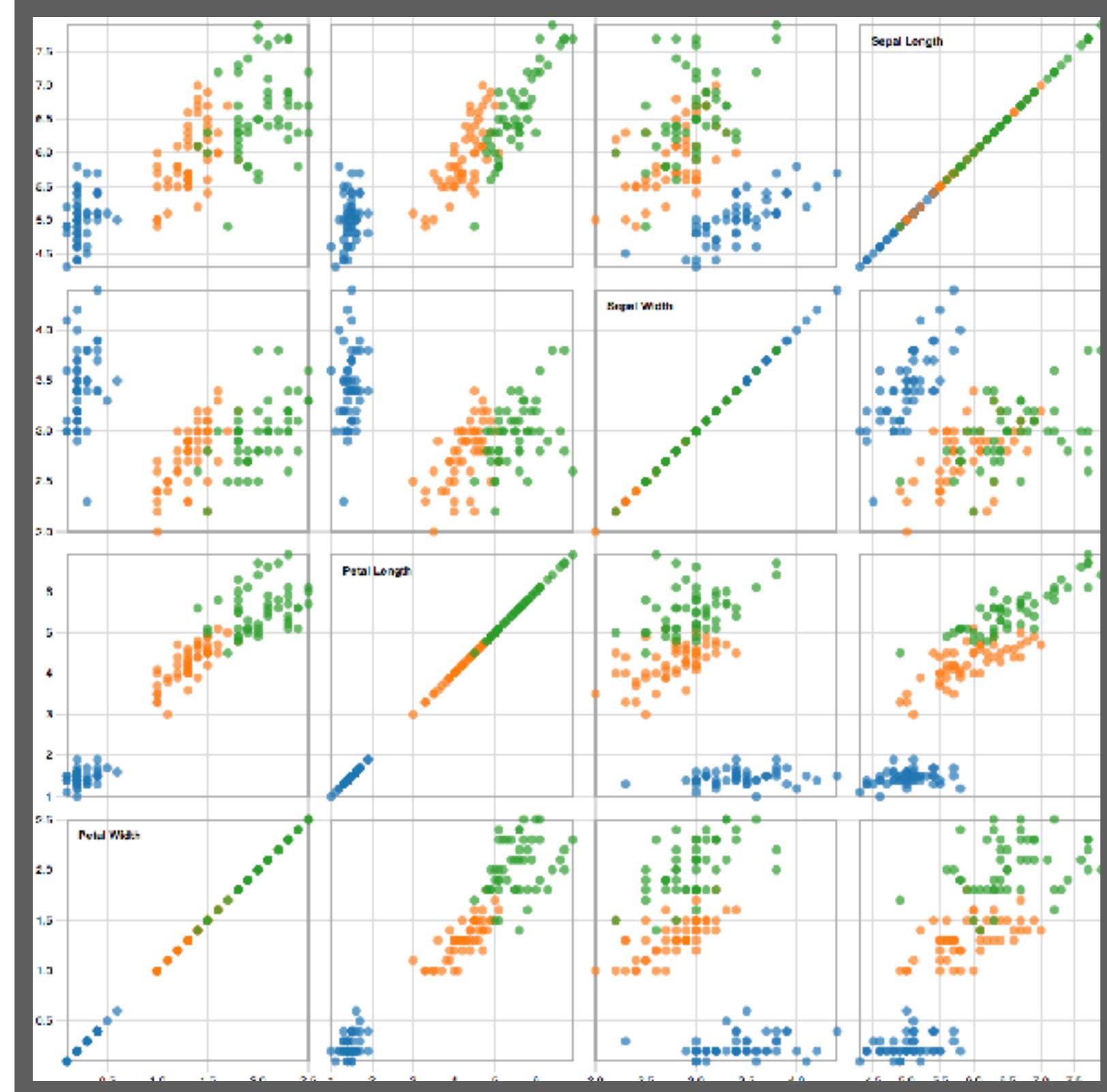
# High-Dimensional Data



## Scatterplot Matrix

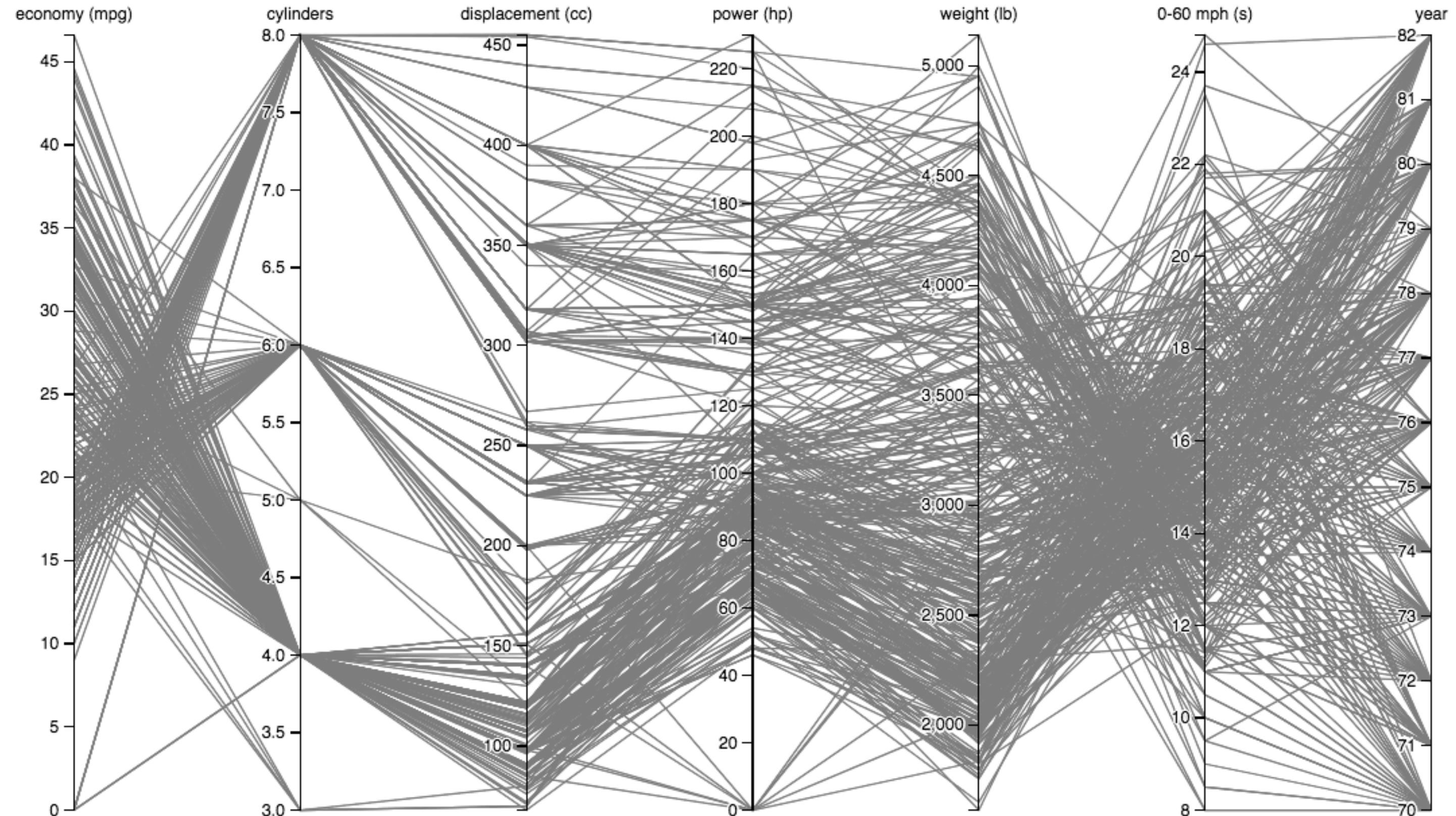


## Parallel Coordinates

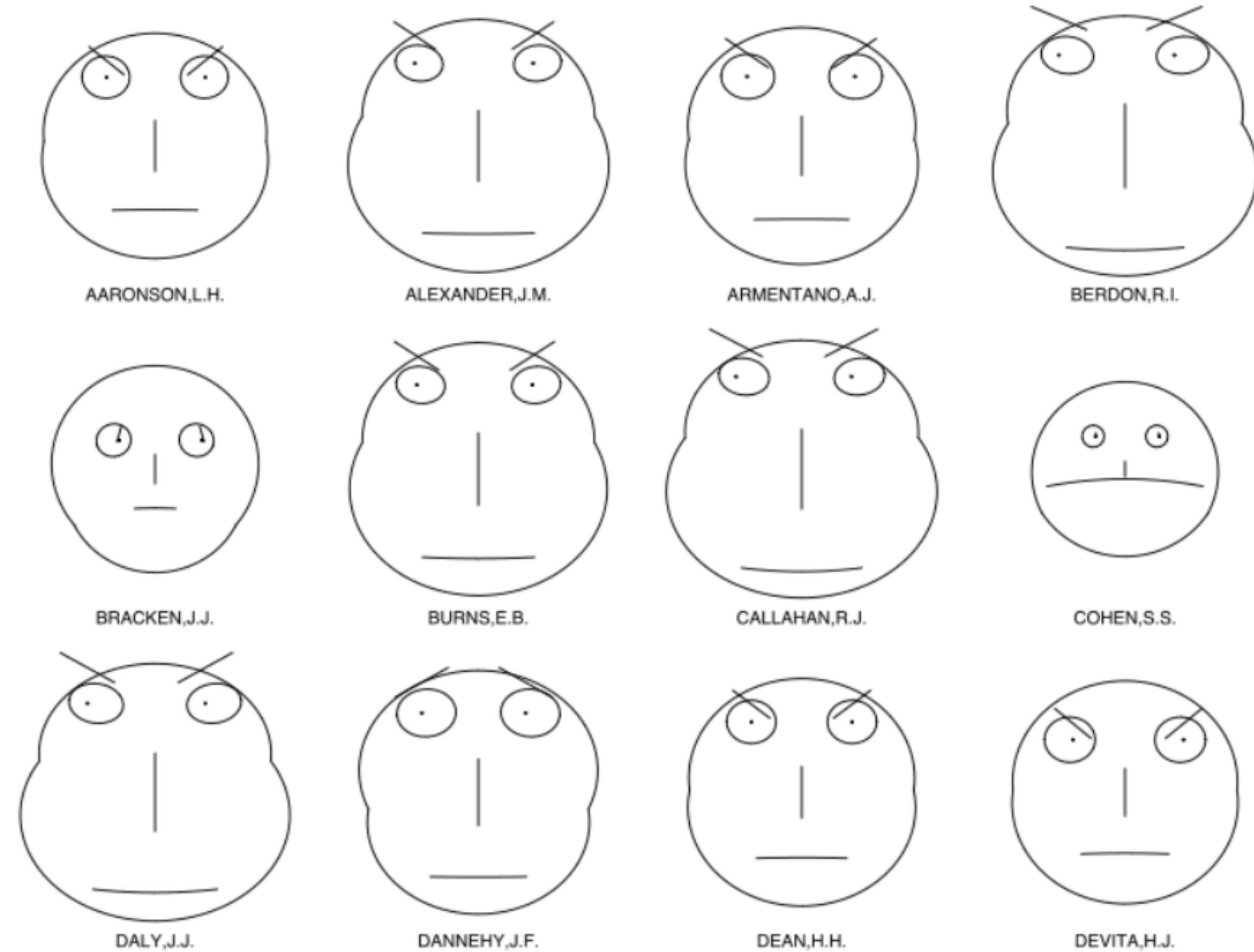
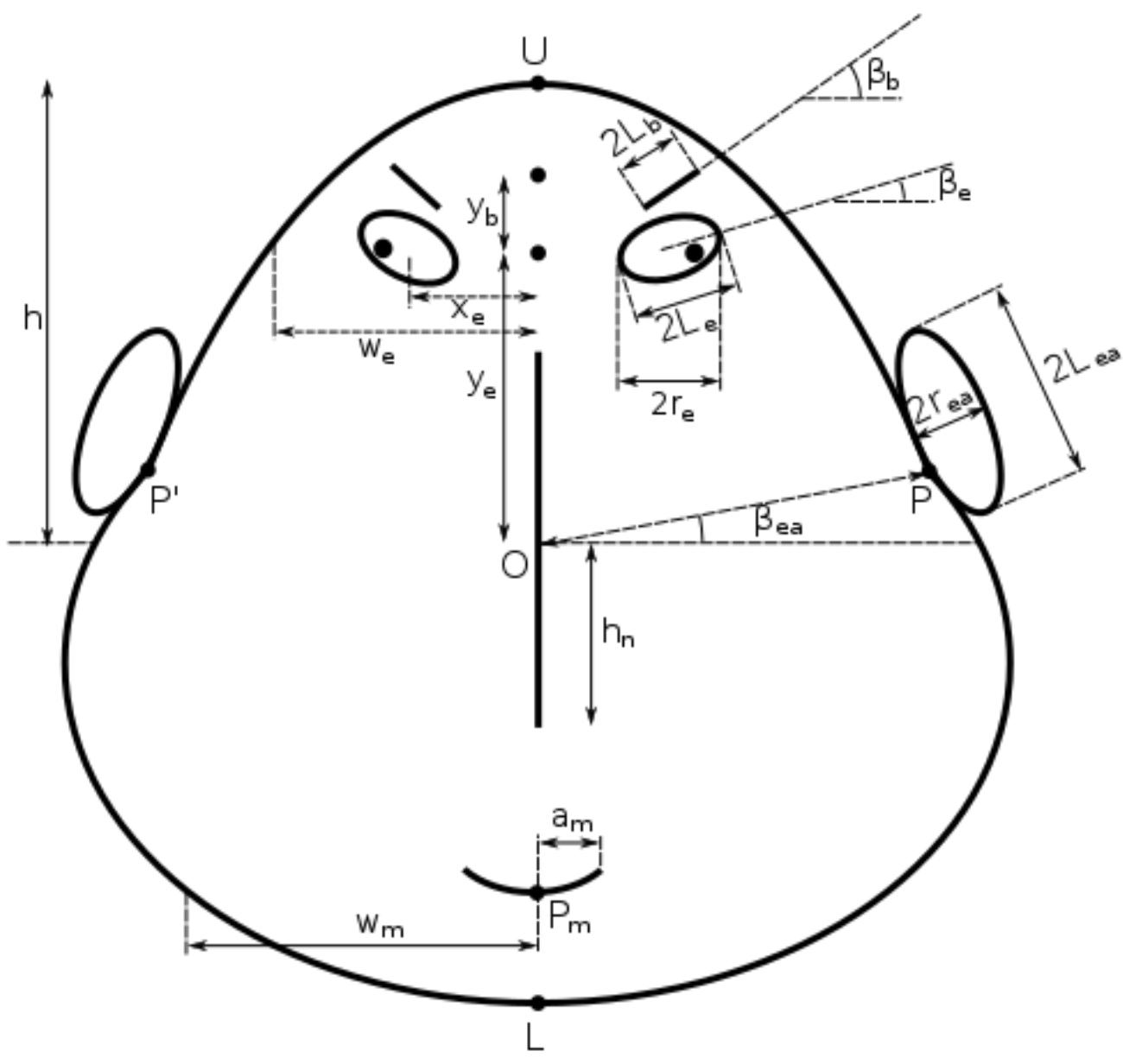


## Scatterplot Matrix

Winner for analyzing correlations between multiple variables [L. Harrison 2014]

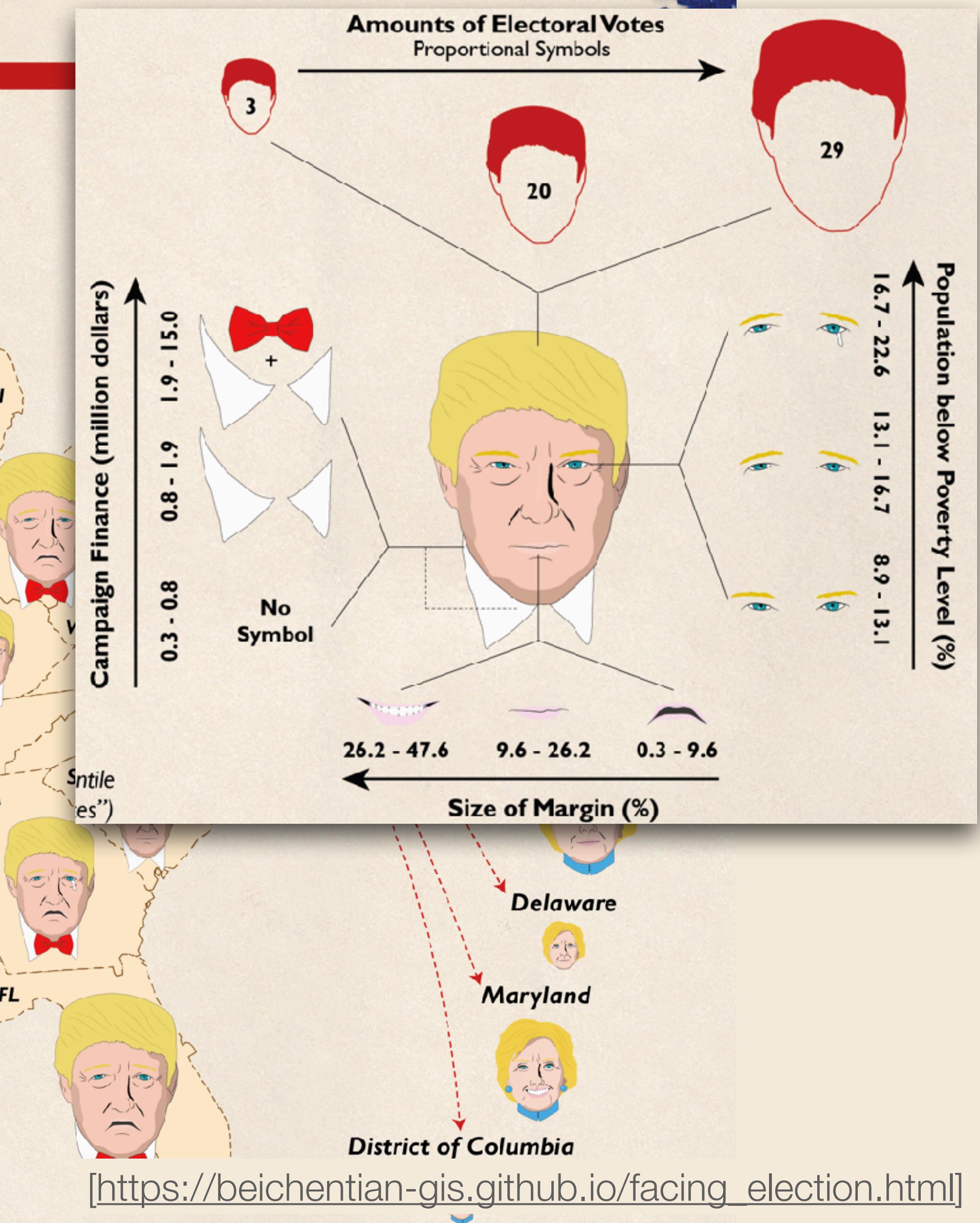
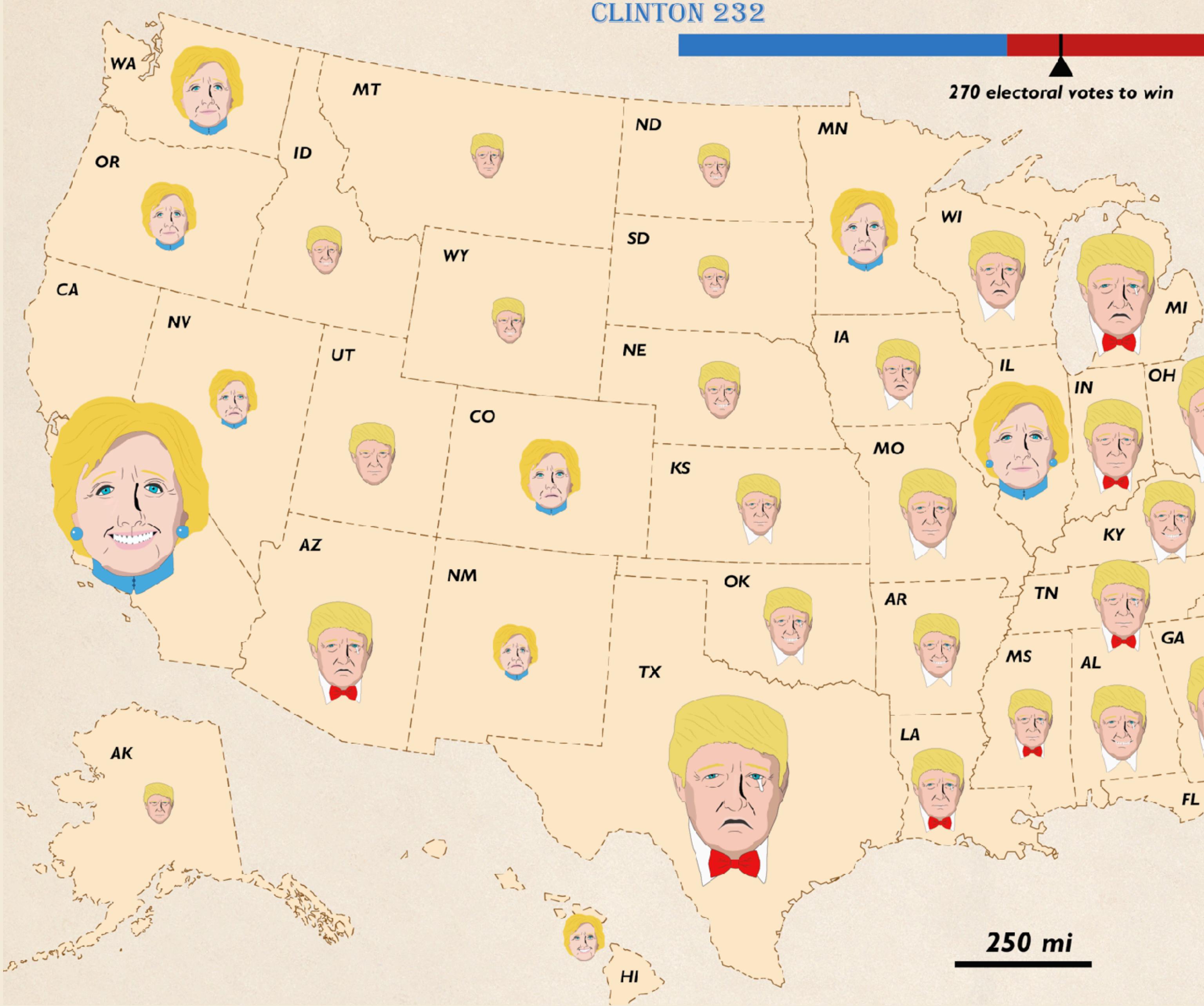


## Parallel Coordinates

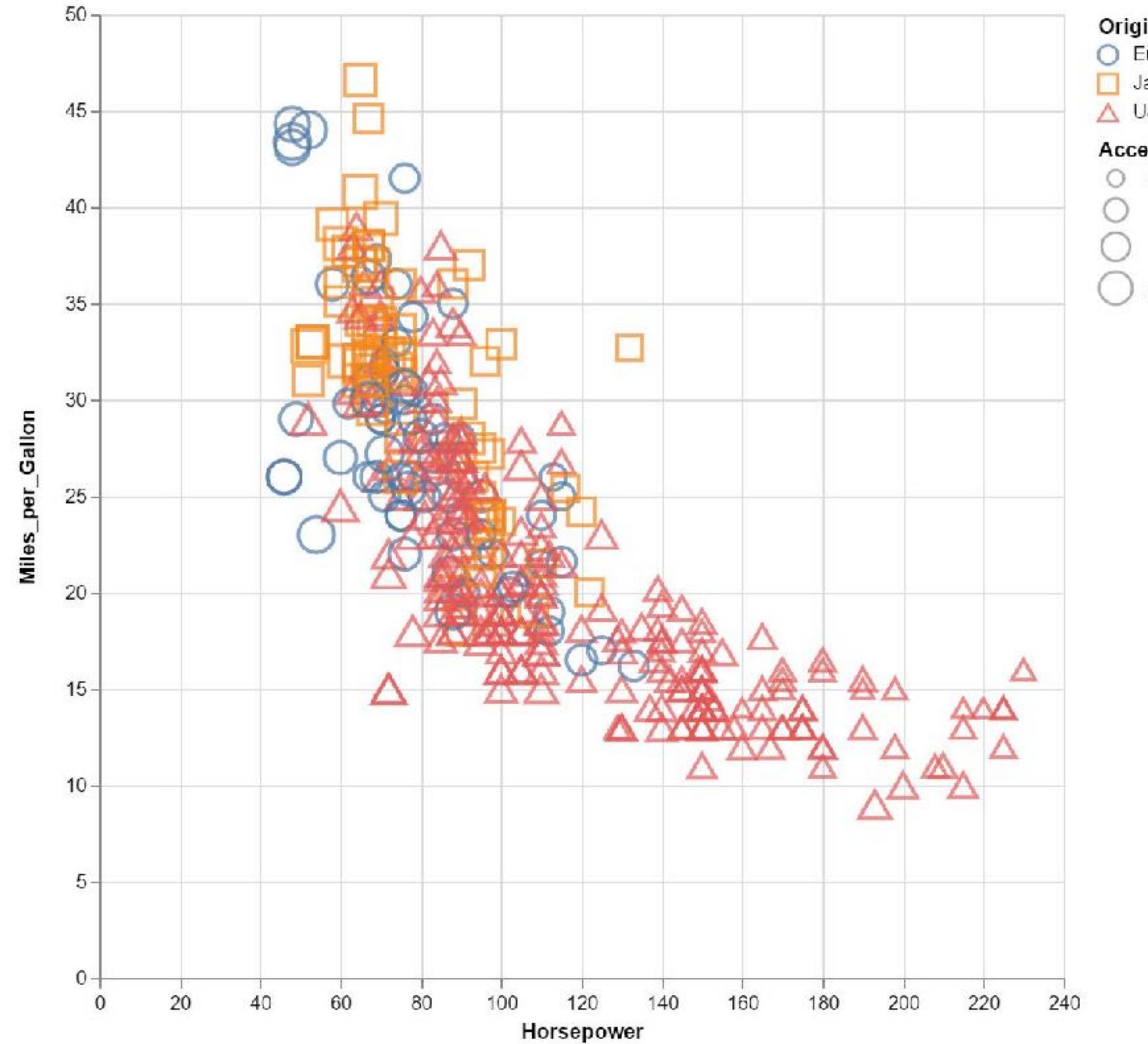


**Lawyers' ratings of twelve judges**

# FACING THE YEAR 2016 PRESIDENTIAL ELECTION



# Don't overload visual encodings



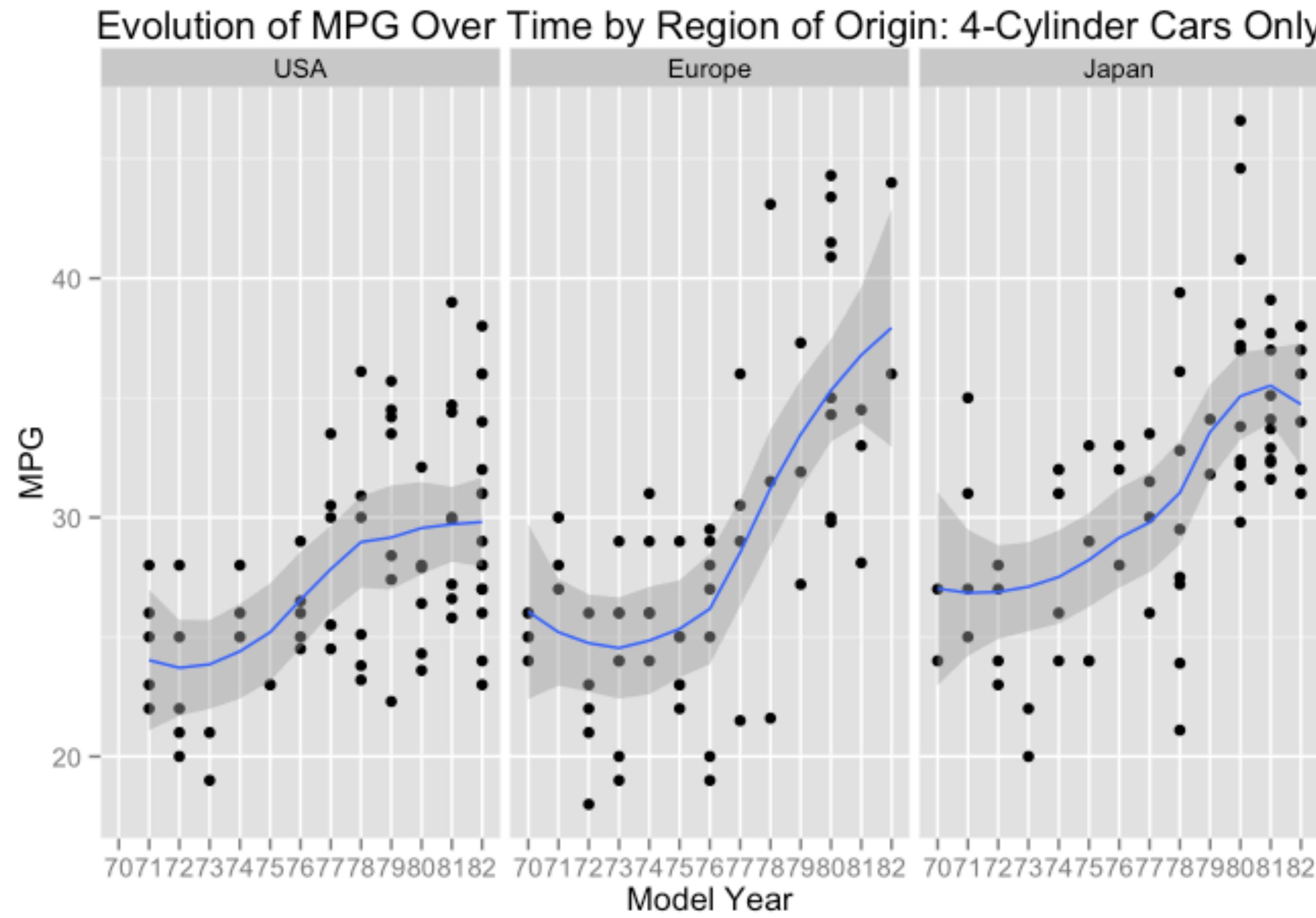
**Origin**  
Europe  
Japan  
USA

**Acceleration**  
5  
10  
15  
20

Origin: Shape  
Acceleration: Size  
X-axis: Horsepower  
Y-axis: Miles per Gallon  
Size: Acceleration

Five visual variables in one chart!

# Use facets or small multiples



Orange and green colors correspond to states where support for vouchers was greater or less than the national average.  
The seven ethnicreligious categories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.  
Where a category represents less than 1% of the voters of a state, the state is left blank.

[Andrew Gelman]

# Dimension Reduction

Mapping from high-dimensional space to two or three dimensions

## How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.

[<https://distill.pub/2016/misread-tsne/>]



Text

# Tag Clouds: Word Count

September 10, 2009

TEXT

## Obama's Health Care Speech to Congress

Following is the prepared text of President Obama's speech to Congress on the need to overhaul health care in the United States, as released by the White House.

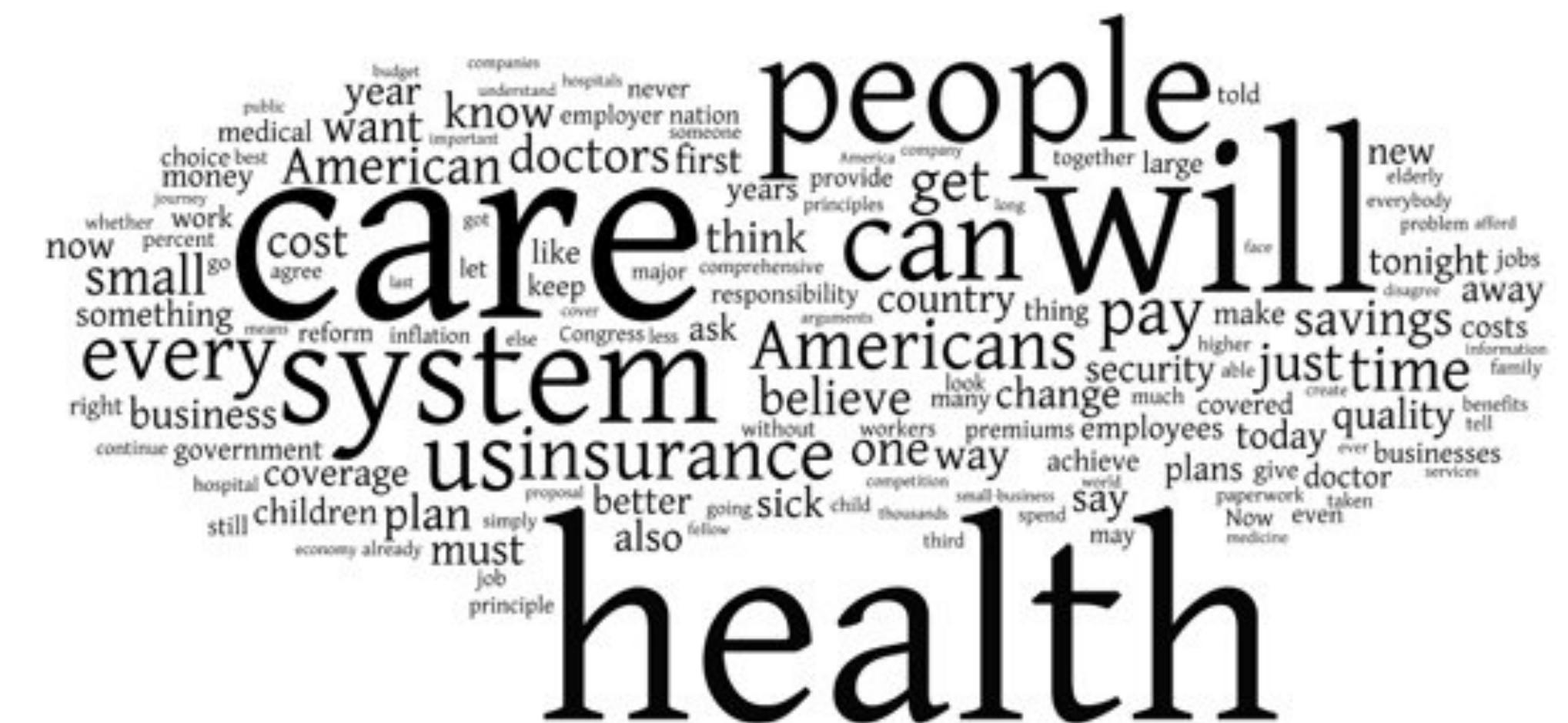
Madame Speaker, Vice President Biden, Members of Congress, and the American people:

When I spoke here last winter, this nation was facing the worst economic crisis since the Great Depression. We were losing an average of 700,000 jobs per month. Credit was frozen. And our financial system was on the verge of collapse.

As any American who is still looking for work or a way to pay their bills will tell you, we are by no means out of the woods. A full and vibrant recovery is many months away. And I will not let up until those Americans who seek jobs can find them; until those businesses that seek capital and credit can thrive; until all responsible homeowners can stay in their homes. That is our ultimate goal. But thanks to the bold and decisive action we have taken since January, I can stand here with confidence and say that we have pulled this economy back from the brink.

I want to thank the members of this body for your efforts and your support in these last several months, and especially those who have taken the difficult votes that have put us on a path to recovery. I also want to thank the American people for their patience and resolve during this trying time for our nation.

But we did not come here just to clean up crises. We came to build a future. So tonight, I return to speak to all of yo



# Tag Clouds

## Strengths

Getting **overview** and initial query formation.

## Weaknesses

Sub-optimal visual encoding (**size vs. position**)

Inaccurate size encoding (**long words are bigger**)

May not facilitate comparison (unstable layout)

Term frequency may not be meaningful

Does not show the structure of the text

# EdWordle [Y. Wang 2017]

EdWordle Home Create FAQ Reference

## EdWordle

EdWordle is a tool for editing "word clouds" based on the [Wordle](#). The initial word cloud can be generated from the input text or read from an existing one. You can re-font, re-color, resize, move, rotate, add and delete words to create custom visualizations.

EdWordle's main benefit is that it allows a neighborhood-preserving editing process, which keeps words at predictable and close locations during and after the editing process. Like Wordle, the images you create with Wordle are yours to use however you like. You can save them to your own desktop to use as you wish.

[Create Now »](#)

Some examples created by others and you can further edit them:



[Edit it!](#)



[Edit it!](#)



[Edit it!](#)

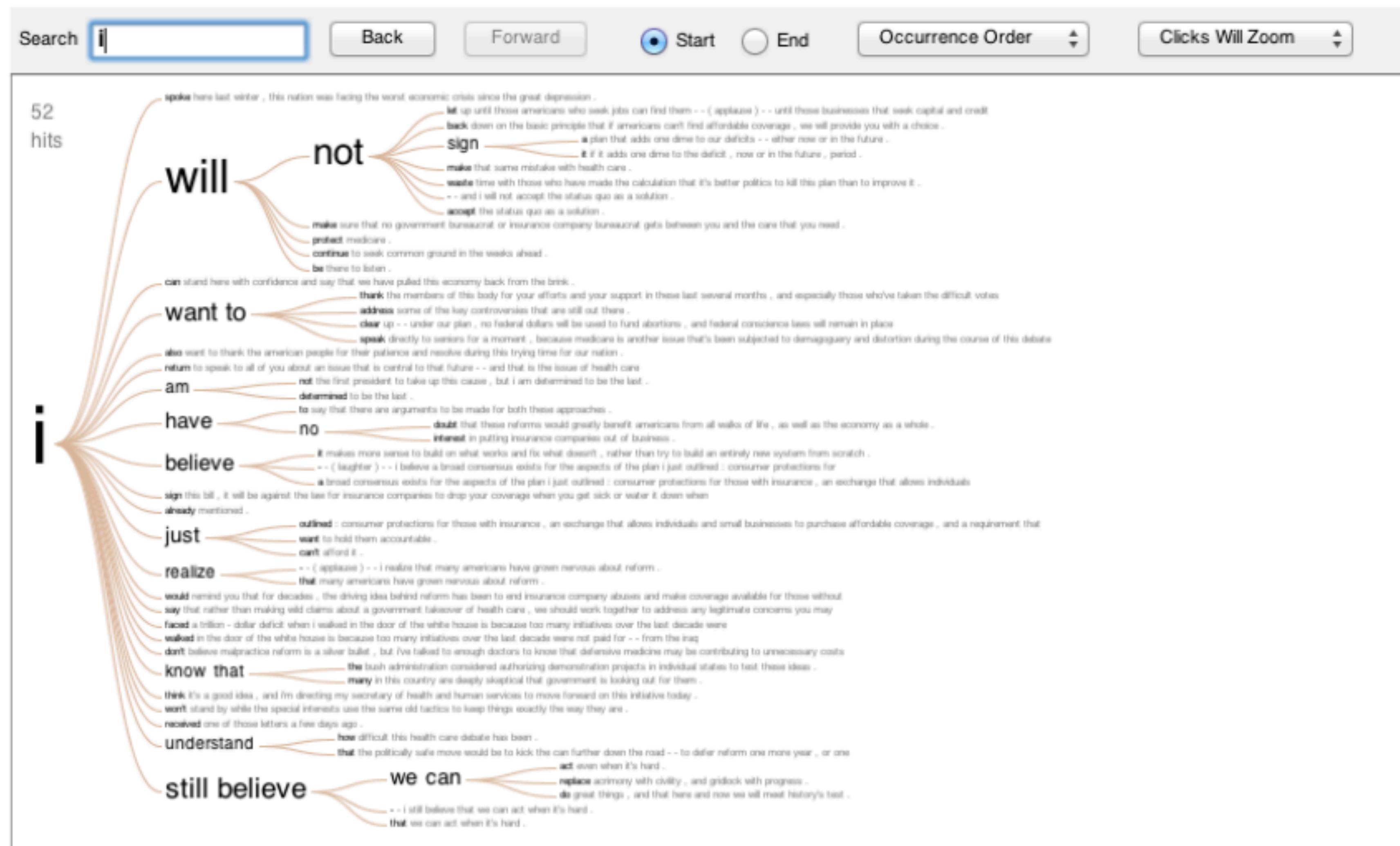


[Edit it!](#)

[<http://www.edwordle.net/>]

# Word Tree: Word Sequences [Wattenberg et al.]

## Visualizations : Word Tree President Obama's Address to Congress on Health Care



if love be rough with you , be rough with love .

if love be blind , love cannot hit the mark .

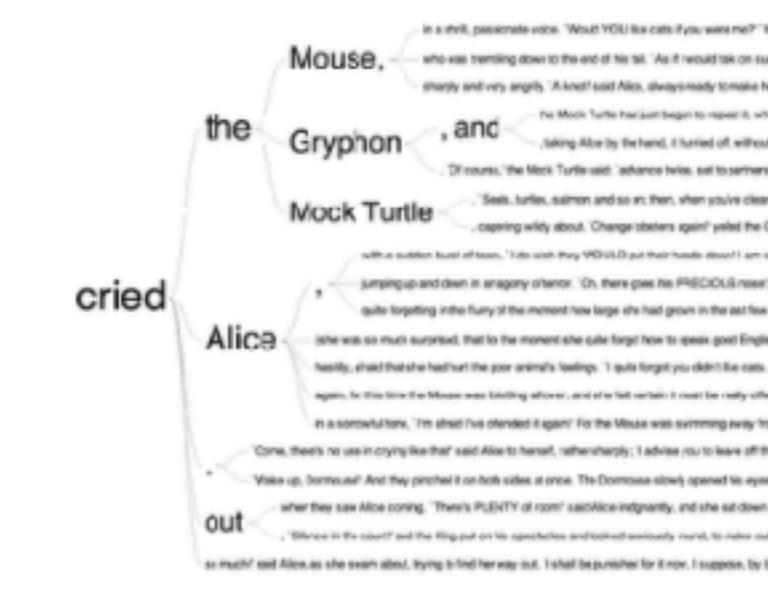
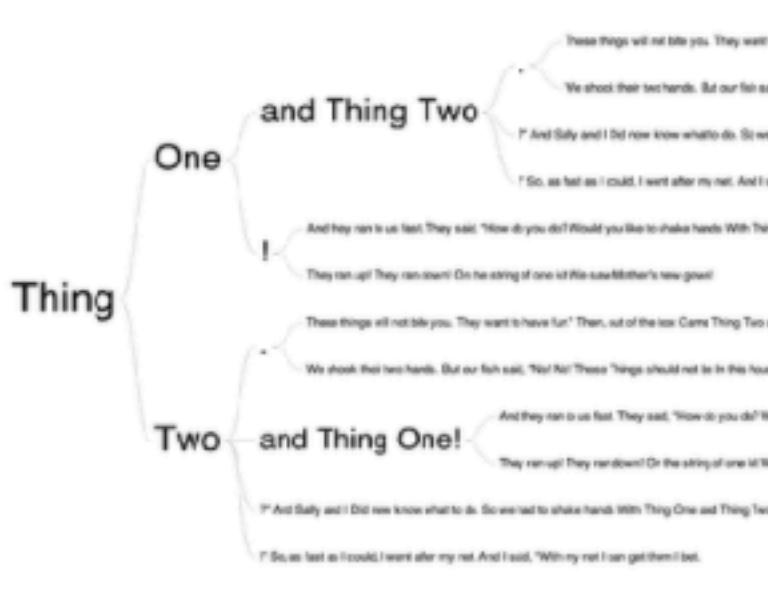
if love be blind , it best agrees with night .

**if love be**

rough with you , be rough with love .

love cannot hit the mark .  
it best agrees with night .

## word tree



# **Obama's Inauguration Speech**

the Administration finally went against its own rhetoric, and pursued diplomacy as its first policy, and from acknowledging our enemies, left us right back in the same place. America's enemies than any strategic choice that we have made in decades preceding arguments we hear from those who have failed to demonstrate the Administration have failed to presents an overarching strategy – the words people. In addition to freeing up resources to take the fight to al Qaeda, ending Taliban, which has returned to strength since we took our eye off of Afghanistan. Above all, country that had no al Qaeda here. Maybe that is why he ~~argued~~. He fails to understand that there are others when I was so serious, when the enemies of the American Perry – because since terrorist principles of the Afghan people. It is not too late to prevail in Afghanistan. But we can't prevail until we rededicate ourselves to the principles of the Afghan people. We will never succeed in Afghanistan if we don't stand by the principles we have in Iraq today, we can remove all of them 150 months. After this deployment, we will stand by Iraq. We can't succeed in Afghanistan. I've committed myself to success. It was a commitment to the principles of the Afghan people.

In the year since President Bush announced the surge – the 10th anniversary of the war for America – and I will offer another break from the failed policies and policies of the past. Nowhere is this better seen than in the threats of a new century have raised the waters of peace and stability. And yet, America's regime that threatens the peace with weapons of mass murder." Recalling the pain of 9/11, he said, "There is no evidence that Iraq had those capabilities of weapons of mass destruction. There was not any evidence in the areas of our cities." At the time the President uttered those words, there was no hard evidence of mass destruction. There was not any evidence that Iraq was responsible for the attacks of September 11, or a focus in a source of fear and comfort it with hope. Here are just five ways in which a shift in strategy away of al Qaeda, whose recruitment has surged and whose leadership enjoys safe-haven in Pakistan – a thousand miles and a world away from the United States – will immediately begin to reverse the course of the war. First, a targeted strike on the terrorists who carried out the 9/11 attacks, and the name three individuals who supported an invasion of Iraq. We will engage with every country in the region – and the UN – to support the stability and territorial integrity in Iraq, who could do little to reduce the insurrection, and too much time reading public opinion. The lesson of this will be easy. But, do know that we can only begin to realize these changes when we end the mindless flat focuses and prudential that will end a war nearly seven years after it started. My plan to end this war will finally put pressure on resources to deal with nearly every other national security challenge we face. This is why the judgment that matters most at the end of the day – that we must leave Iraq because violence was up, and then argue that this year that we can't be safer because of this war? Another is why Senator McCain can argue – as he did last year – that we could. Now, we know what we will hear from these men like John McCain who support open-ended war. They will argue that leaving Iraq will undermine regional stability, leaving it to re-emphasize that ground, to ensure that our interests in countries with maintaining the threat military in the history of the world. As Commander in Chief, I will begin to give a military overextended by a region – and the UN – to support its stability and territorial integrity of Iraq. And as we will launch a major humanitarian initiative to support a full Iraqi leadership that will stay as long as it takes, they have an iron as they want. We need to send a different message. We all help to end this war will put pressure on Iraq's leaders to take responsibility for their future. Because we've learned that when we end a war, and whose leadership enjoys safe-haven in Pakistan – a housebroken Iraq. The central front in the war against terror is not Al Qaeda. Petraeus has himself acknowledged, the Iraqis are not achieving the political progress needed to end their civil war. Beyond will collate the reasons why we waged a war the didn't need to be fought, but less stand by. In 2002, when the failed decisions almost of approach – an approach that puts the onus on the military, and that relies on more than just military power, it is needed to stabilize

# The Cat in the Hat

/www.jason

## Alice in Wonderland

# **Bob Dylan's Blowin' in the Wind**

```
graph TD; dtree["dtree/] -- for --> sharing[sharing]; dtree -- for --> the[the]; dtree -- for --> your[your]; dtree -- for --> posting[posting]; dtree -- so --> much[much]; dtree -- so --> I[I]; dtree -- so --> all[all]; dtree --> verymuch[very much];
```

## **Obama War Speech**

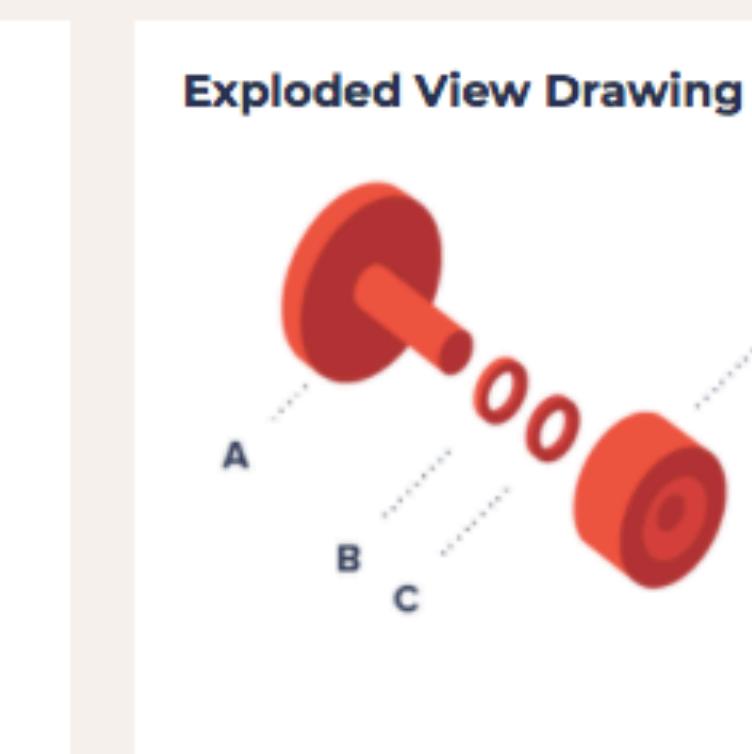
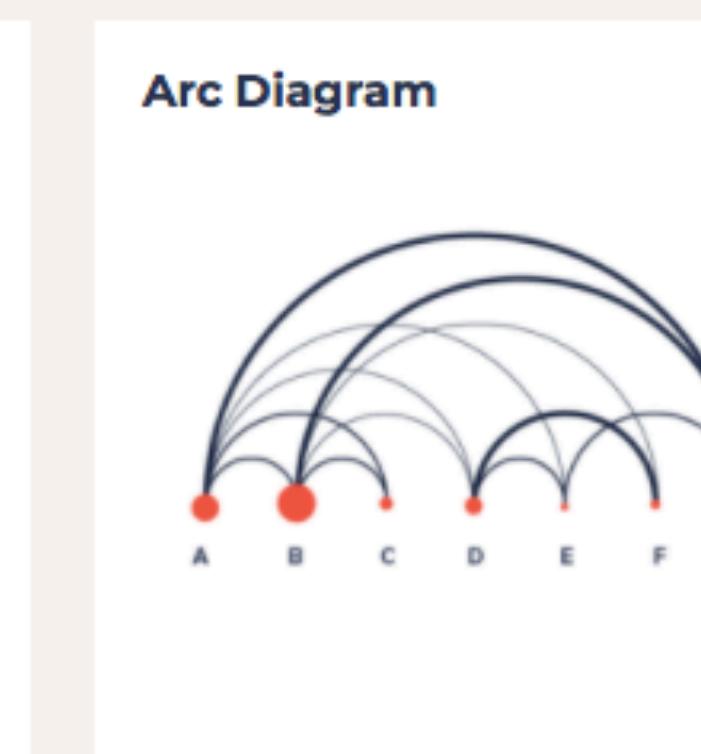
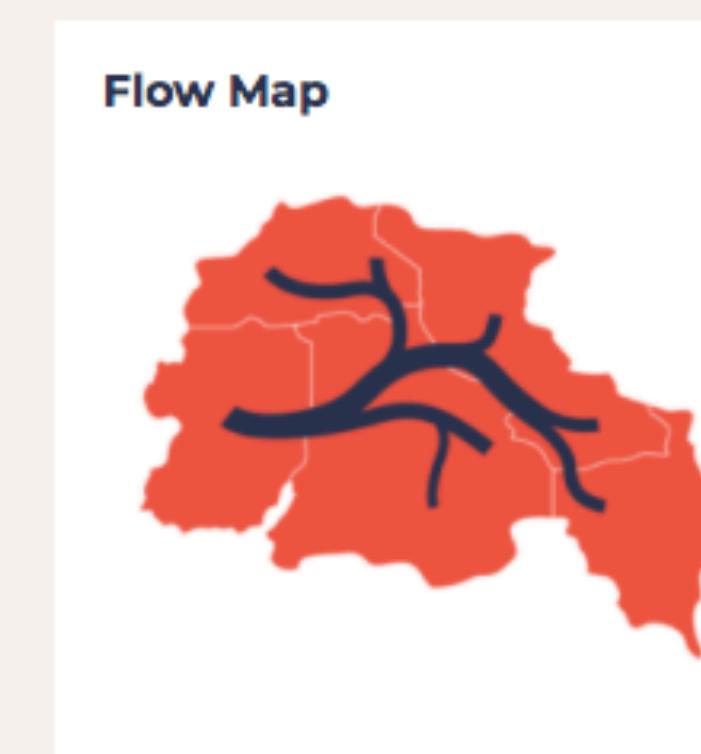
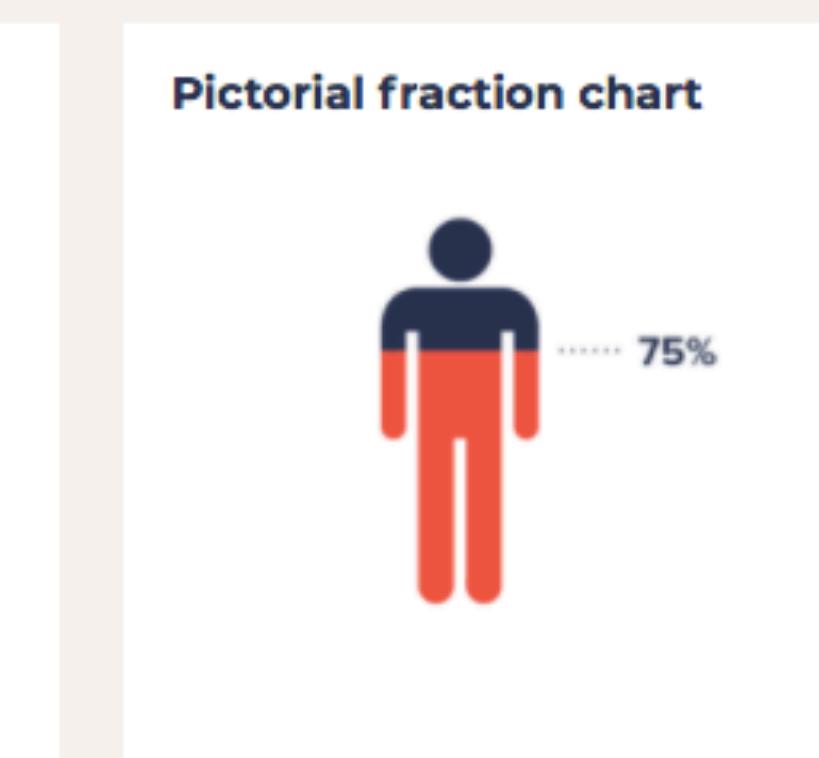
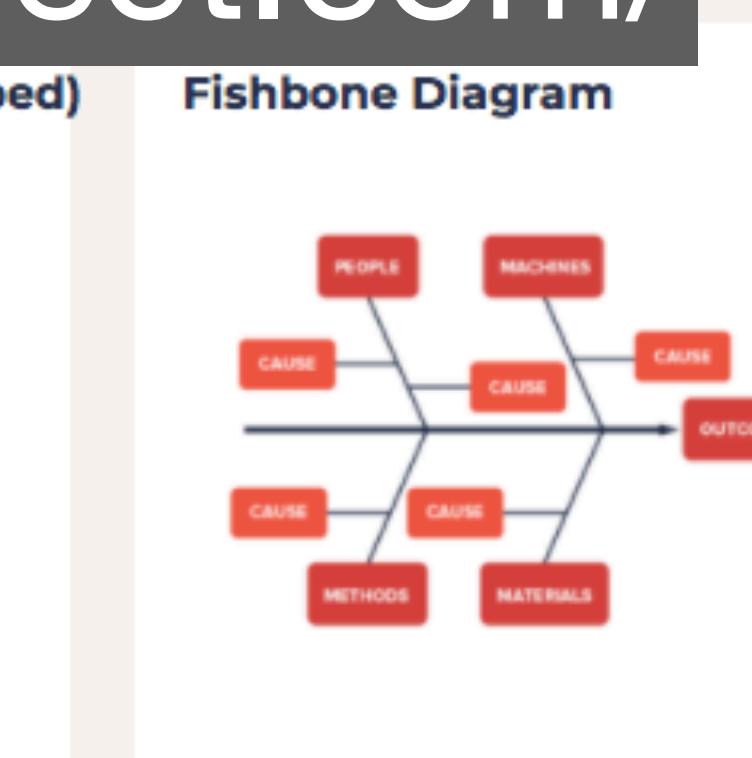
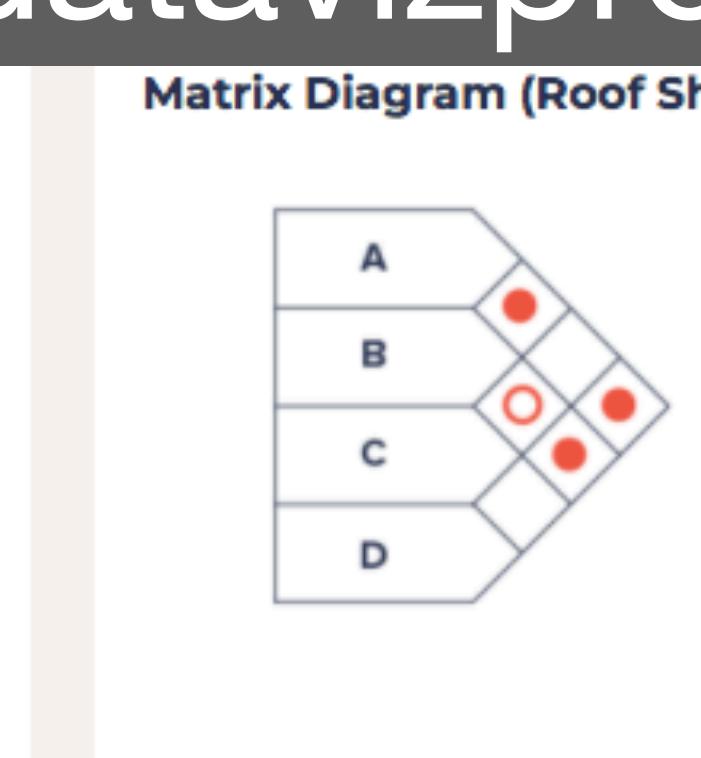
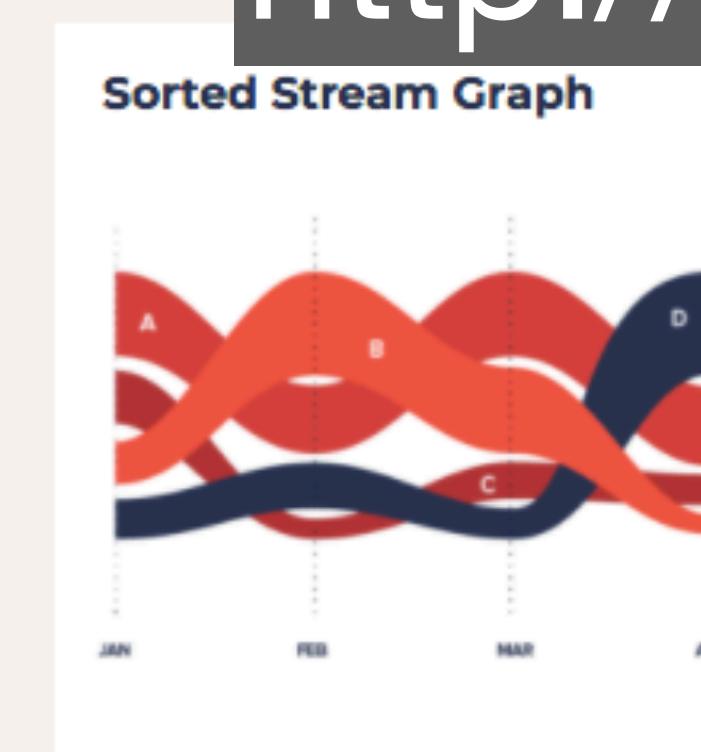
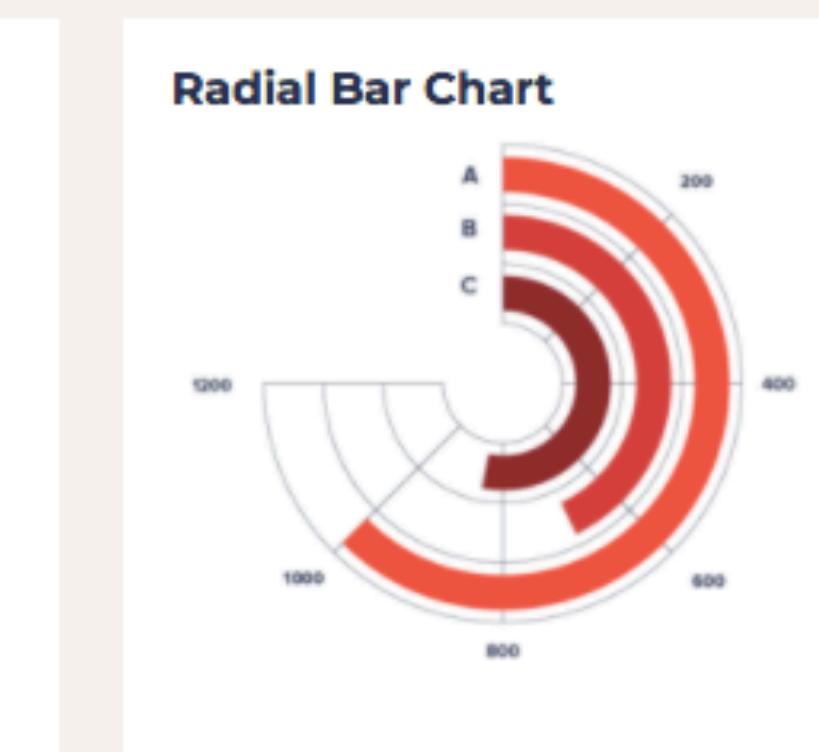
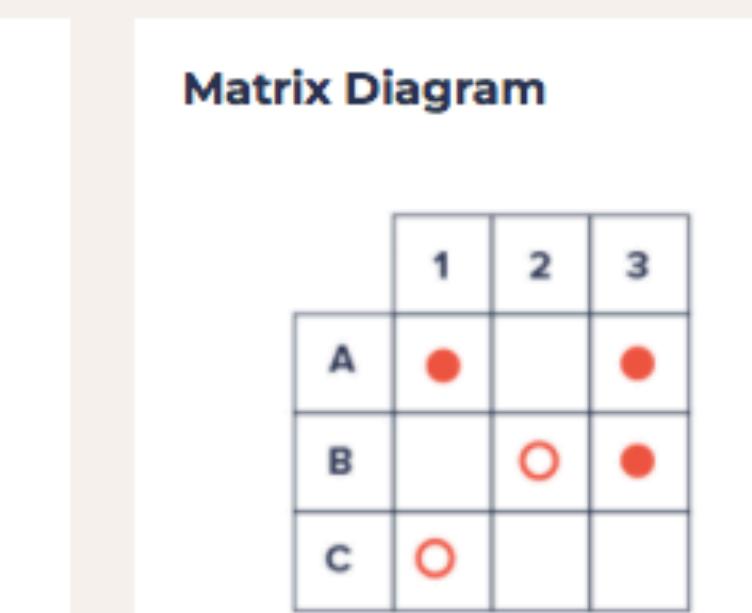
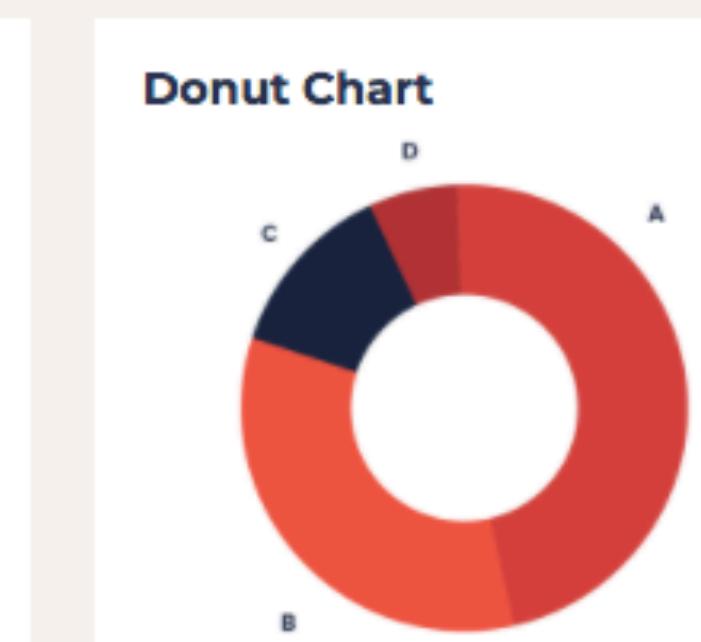
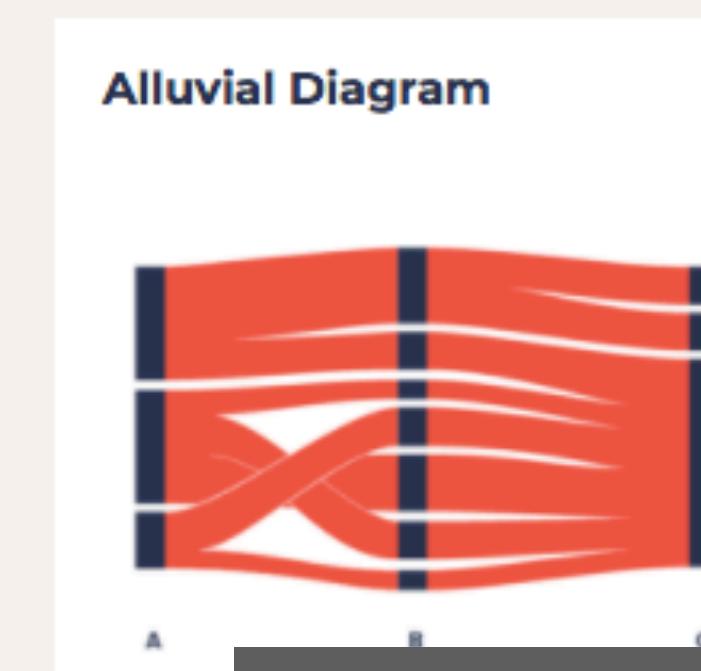
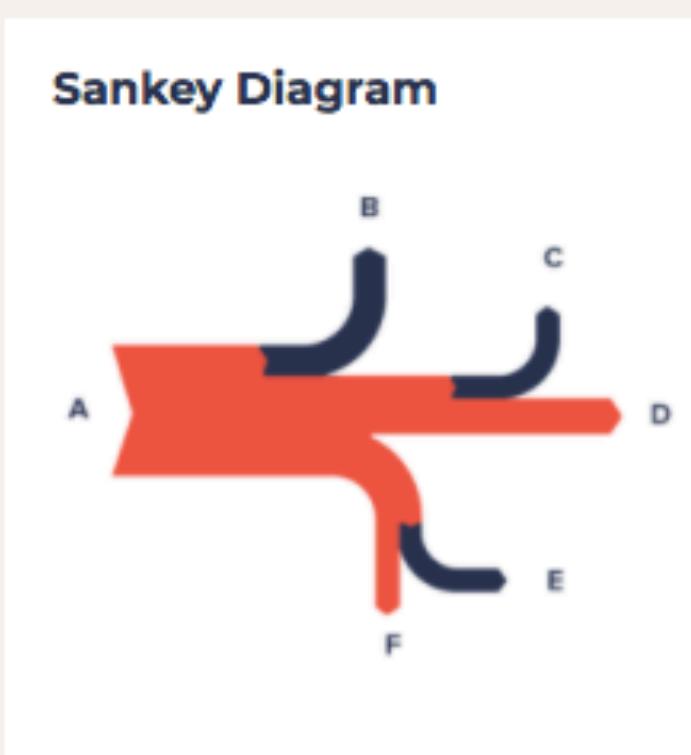
# **Steve Jobs Stanford Commencement Speech**

@jasondavie

## Flickr Comments

## Paste Text

# Others

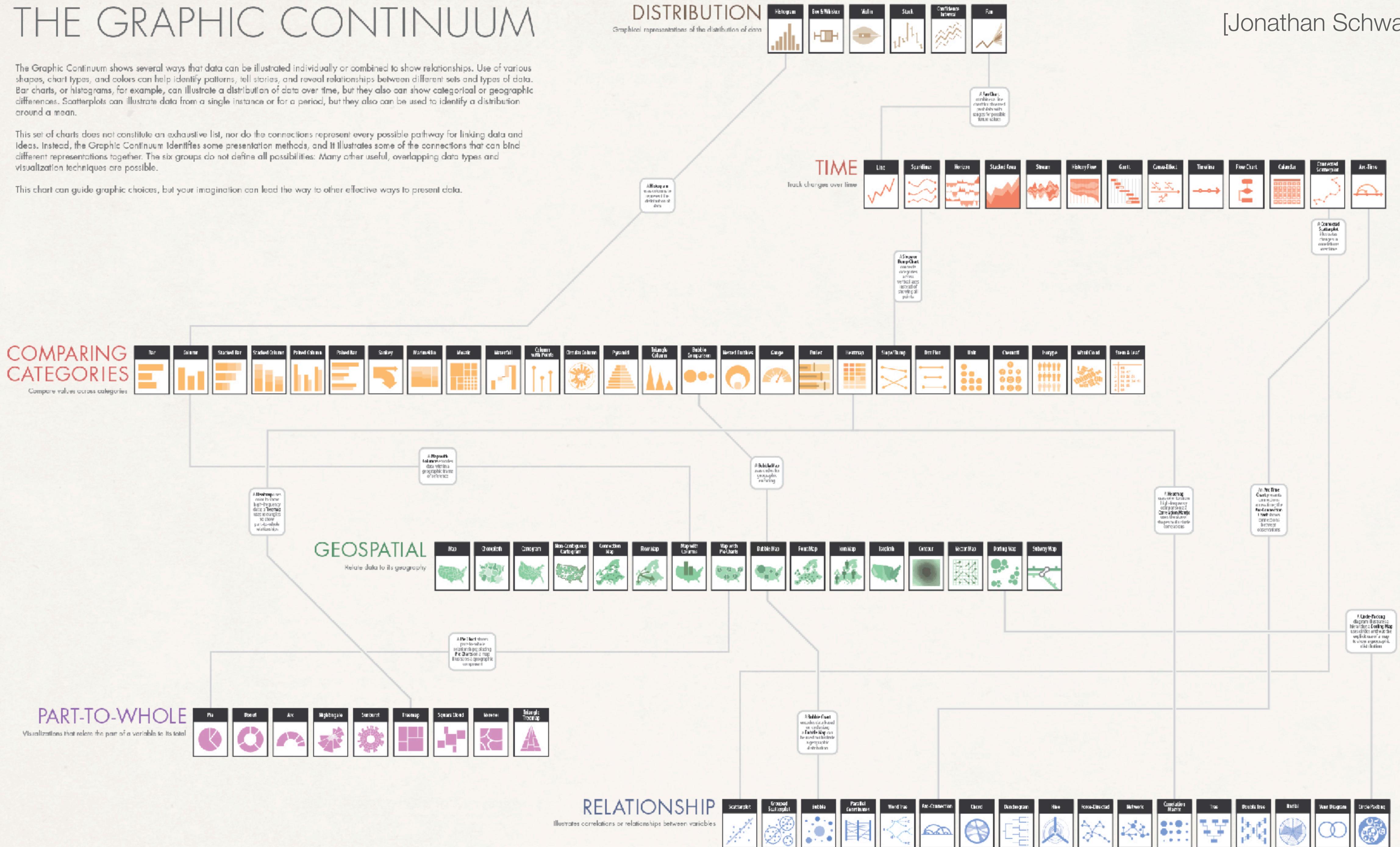


# THE GRAPHIC CONTINUUM

The Graphic Continuum shows several ways that data can be illustrated individually or combined to show relationships. Use of various shapes, chart types, and colors can help identify patterns, tell stories, and reveal relationships between different sets and types of data. Bar charts, or histograms, for example, can illustrate a distribution of data over time, but they also can show categorical or geographic differences. Scatterplots can illustrate data from a single instance or for a period, but they also can be used to identify a distribution around a mean.

This set of charts does not constitute an exhaustive list, nor do the connections represent every possible pathway for linking data and ideas. Instead, the Graphic Continuum identifies some presentation methods, and it illustrates some of the connections that can bind different representations together. The six groups do not define all possibilities: Many other useful, overlapping data types and visualization techniques are possible.

This chart can guide graphic choices, but your imagination can lead the way to other effective ways to present data.

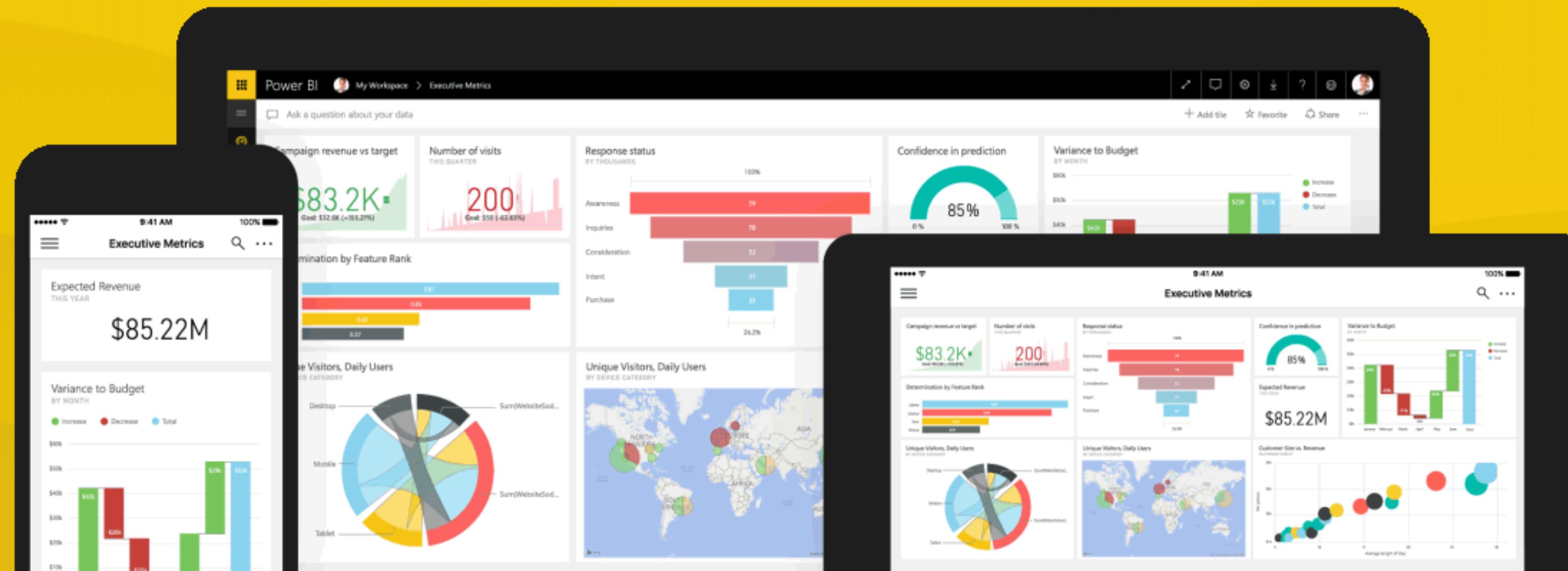


# Tools & More

# Business intelligence like never before

Go from data to insights in minutes.  
Any data, any way, anywhere. And all in one view.

START FREE >



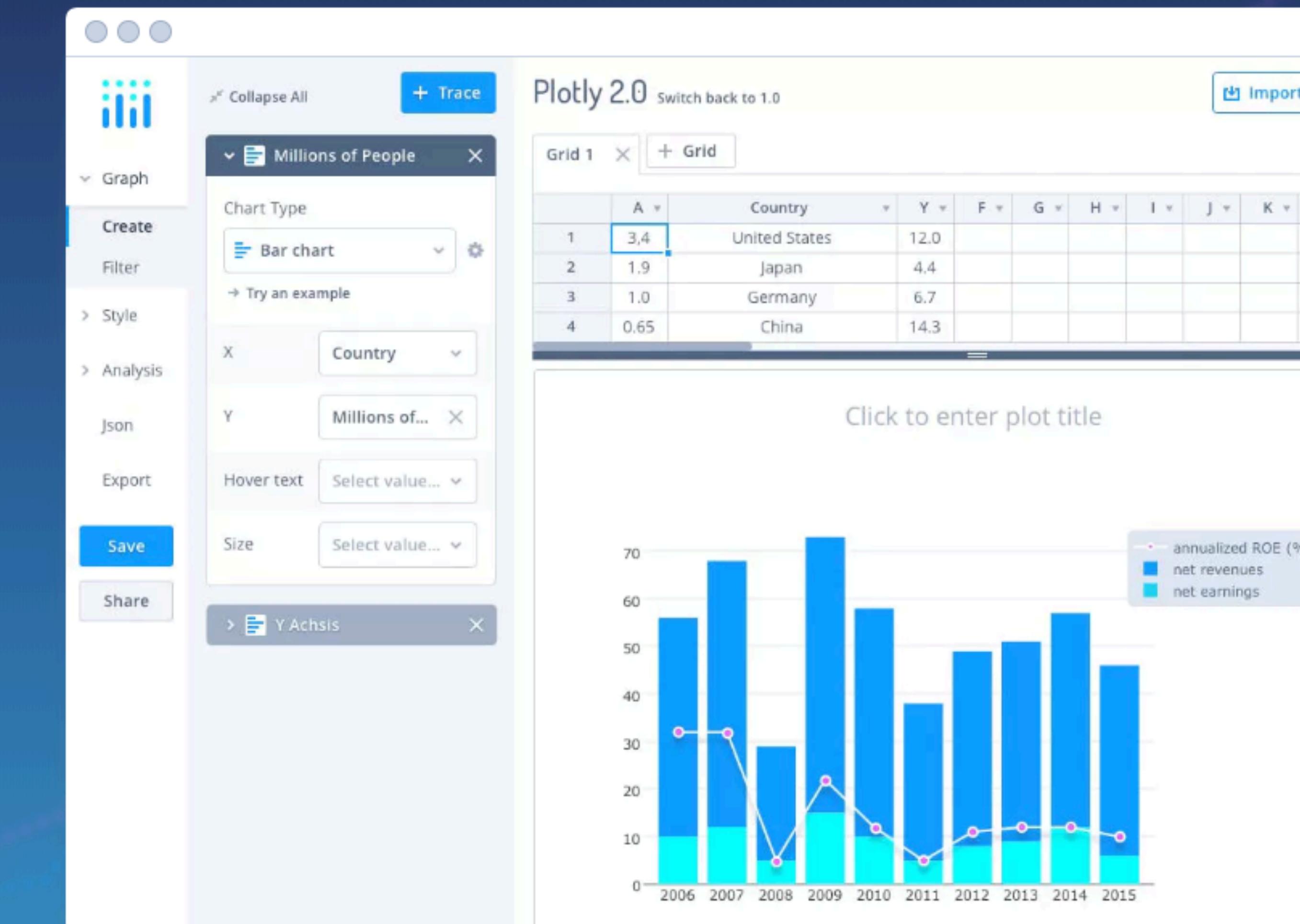
New version of Power BI Report Server is now available. Publish Power BI reports on-premises connecting to any data. [Learn more](#)

# Chart Studio

The world's most sophisticated editor for creating D3.js and WebGL charts.

No coding required.

**GET STARTED**



# RAWGraphs

The missing link between spreadsheets and data visualization.

[USE IT NOW!](#)[FORK IT ON GITHUB](#)



# Turn your data into knowledge

**Quadigram** allows you to engage people by sharing stories that matter.

Take a look at the new account for professionals.

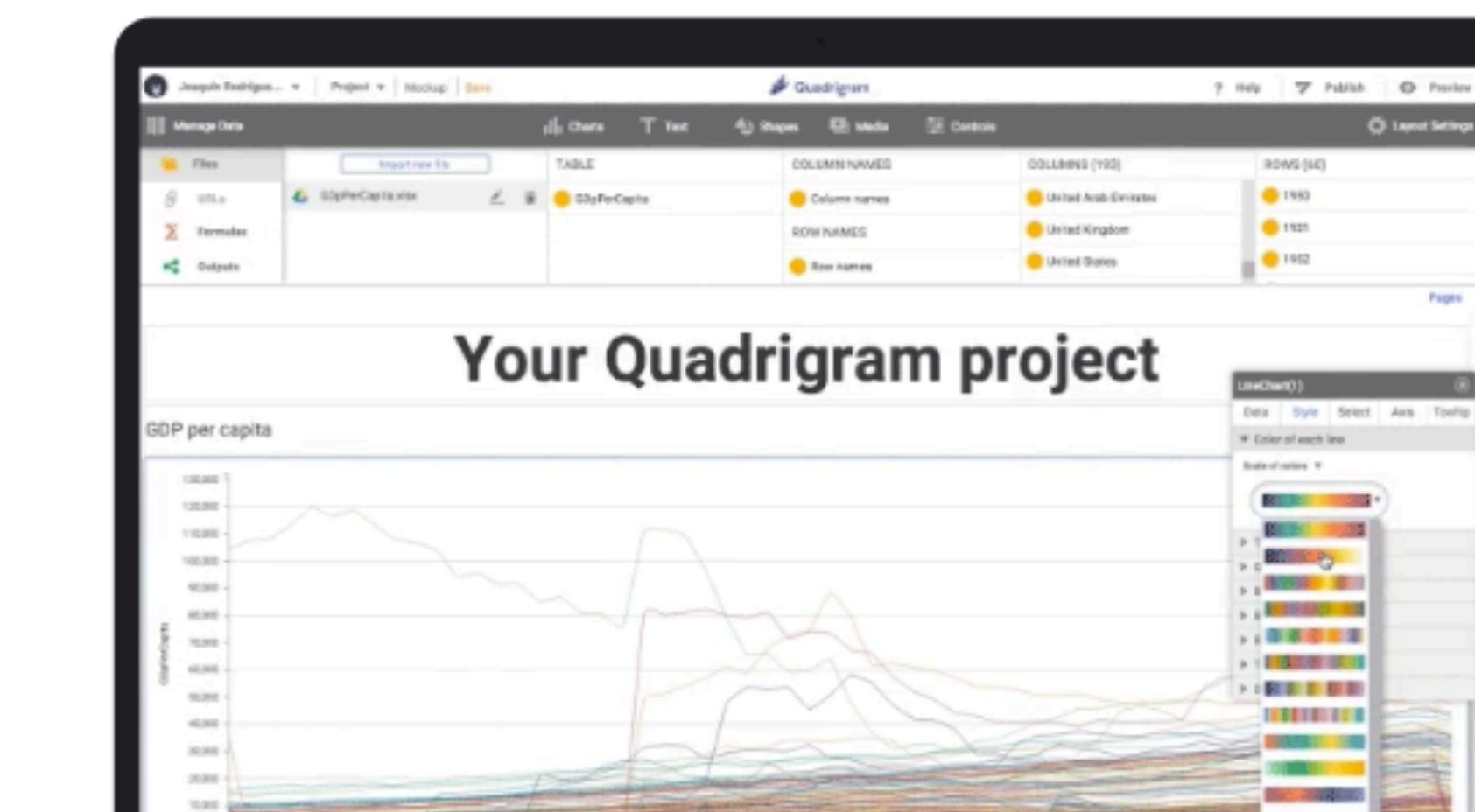
PURCHASE NOW

Or try it for **FREE**

## WHAT

### Easy to use

Quadigram is a visual drag & drop data editor that will transform the stories you bring to the web.

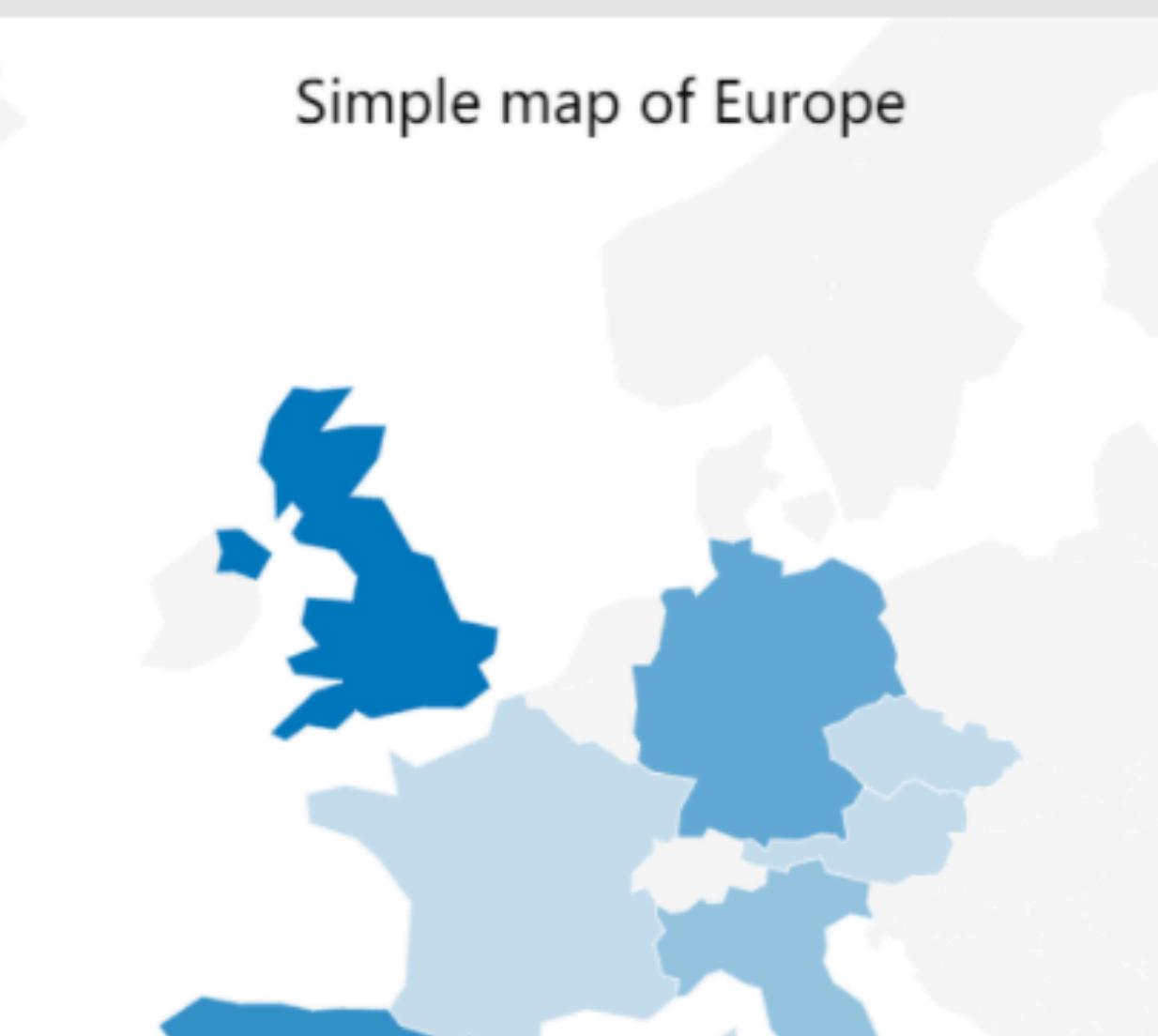


# Content Management System for data visualizations

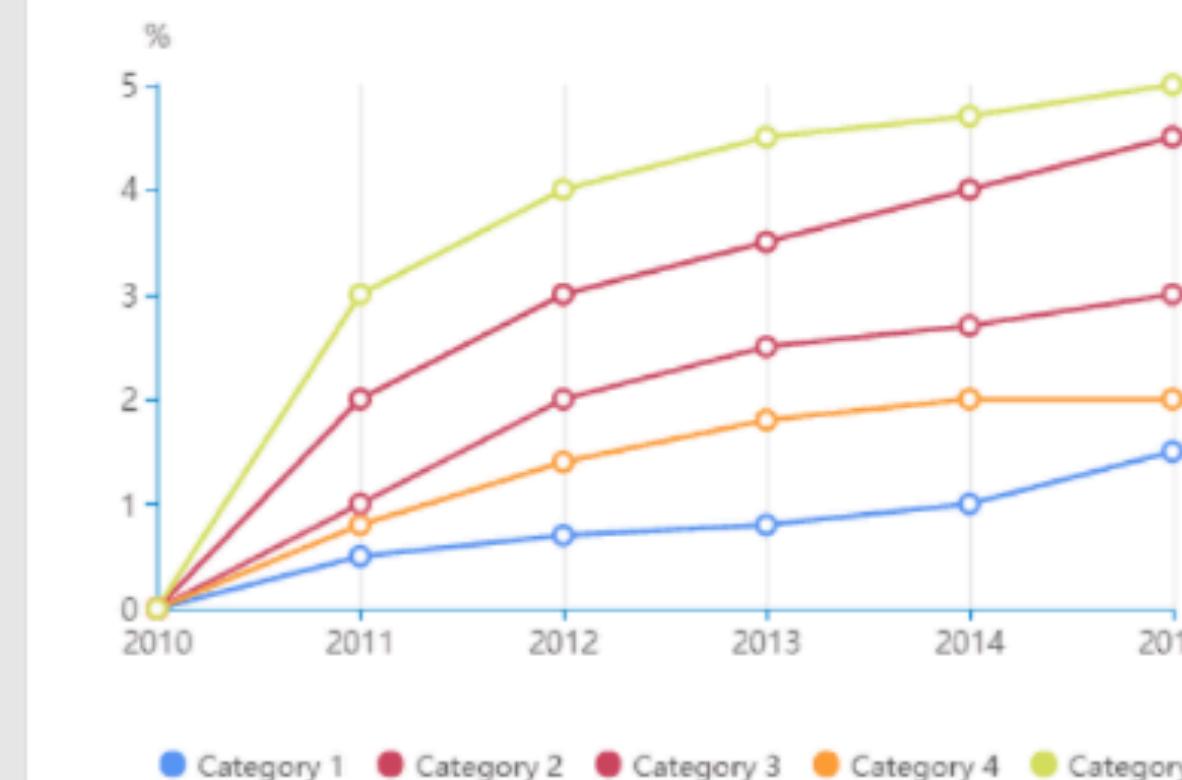
Easily pick a data visualization template, customize it, and publish on your website.

[SIGN UP](#)[or SIGN IN with your Google Account](#)[All](#)[Map](#)[Pie](#)[Area](#)[Line](#)[Bar](#)[Column](#)[Scatter](#)[Misc](#)[All \(74\)](#)[Show PREMIUM](#)

Simple map of Europe



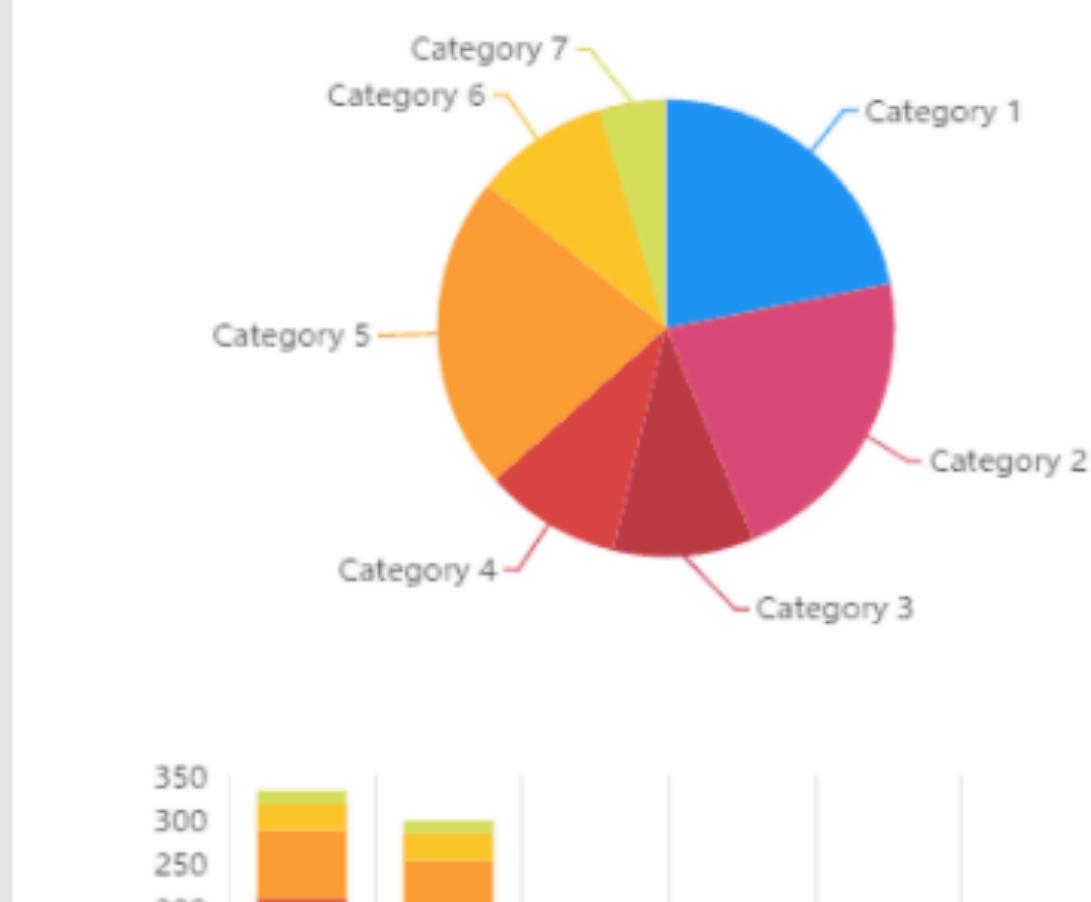
Multiple lines chart

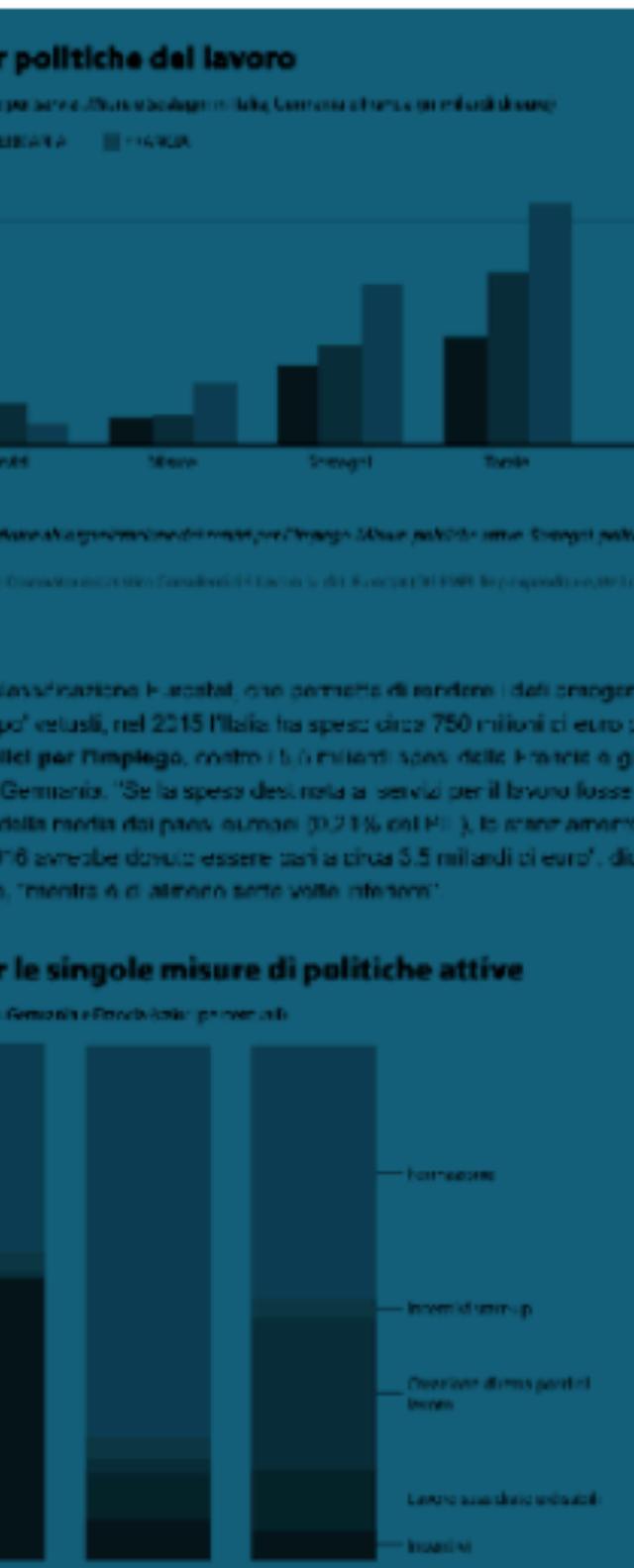


Multiple lines chart

FREE

Stacked column chart with pie chart





# Enrich your stories with charts, in seconds.

[CREATE A CHART](#)
[CREATE A MAP](#)

GOES ITS SURVEY, WHICH KEPT SOME WORKERS AT HOME. ADDITIONALLY, THE "APCALYPSE" OF ANNOUNCED STORE CLOSINGS MEANT THAT MORE JOBS THAN NORMAL LEFT THE ECONOMY DURING THE MONTH. THIS MONTH, THE DISAPPOINTING MARCH NUMBER WAS REVISED DOWN FROM 98,000 TO 79,000. BUT THE APRIL JOBS REPORT PROVIDES A BOUNCE BACK IN PART BECAUSE OF WARMER WEATHER AND FEWER LAYOFFS. THE LABOR DEPARTMENT REPORTED GAINS IN HOTELITY, MINING, HEALTHCARE, AND FINANCE. INCLUDING THE REVISIONS FOR THE FEBRUARY AND MARCH REPORTS, AN AVERAGE OF 174,000 JOBS WERE ADDED PER MONTH OVER THE LAST THREE MONTHS.

## Monthly Changes in U.S. Employment (Non-Farm), 2006-2017



Source: Bureau of Labor Statistics | Get The Data | Embed

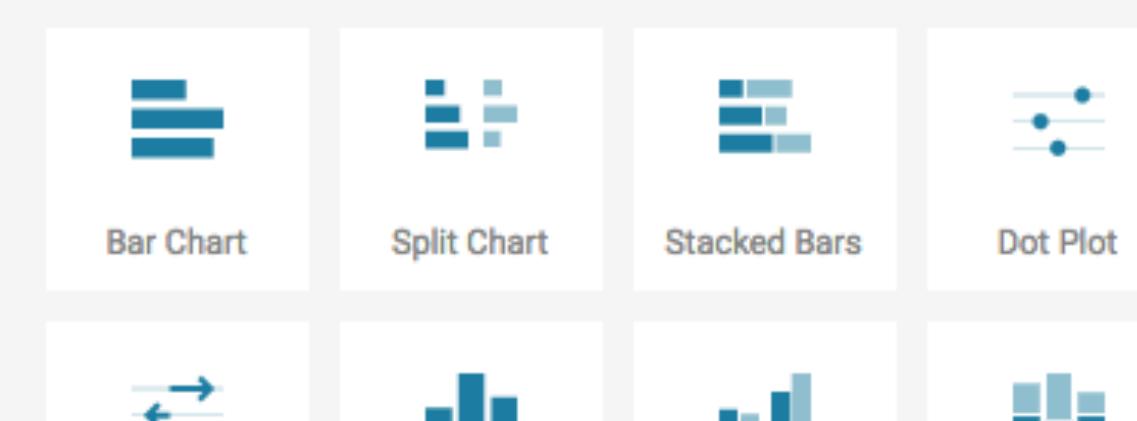
The Atlantic

## 3. An interest-rate hike in June now looks more likely.



## Datawrapper makes it easy to create beautiful charts.

See for yourself how different newsrooms use our charts:



### Evolución de la población mundial

En millones de habitantes

■ Asia ■ África ■ América Latina y Caribe ■ Europa ■ América del Norte (excepto México) ■ Oceanía

11000

10000

Previsión 2050  
9.773 millones

*Recent* [SEE ALL →](#)**Spotting rip currents**

Rip currents are like hidden rivers near the shore that ...



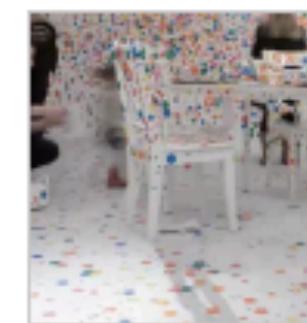
FEED SPONSOR

**Be the data whisperer →**

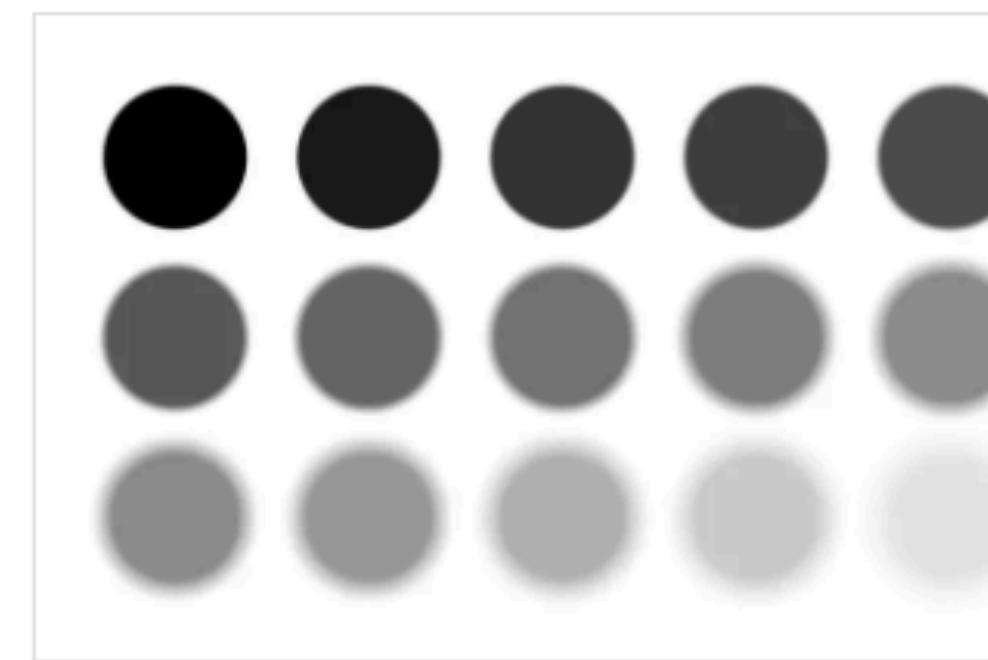
Notre Dame's Mendoza College of Business in Chicago offers a one-year, part-time MS in Business Analytics for ambitious working professionals

**Obliteration Room invited people to put dotted stickers everywhere**

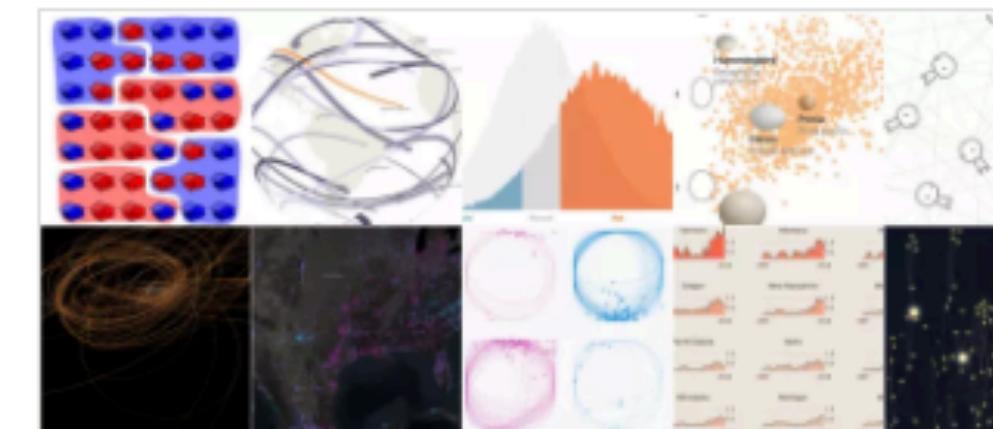
The Obliteration Room (2012) by artist Yayoi Kusama started as ...

**Where athletes in professional sports come from**

Sports are growing more

**Visualizing the Uncertainty in Data**

Data is an abstraction, and it's impossible to encapsulate everything it represents in real life. So there is uncertainty. Here are ways to visualize the uncertainty.

**10 Best Data Visualization Projects of 2017**

It was a rough year, which brought about a lot of good work. Here are my favorite data visualization projects of the year.

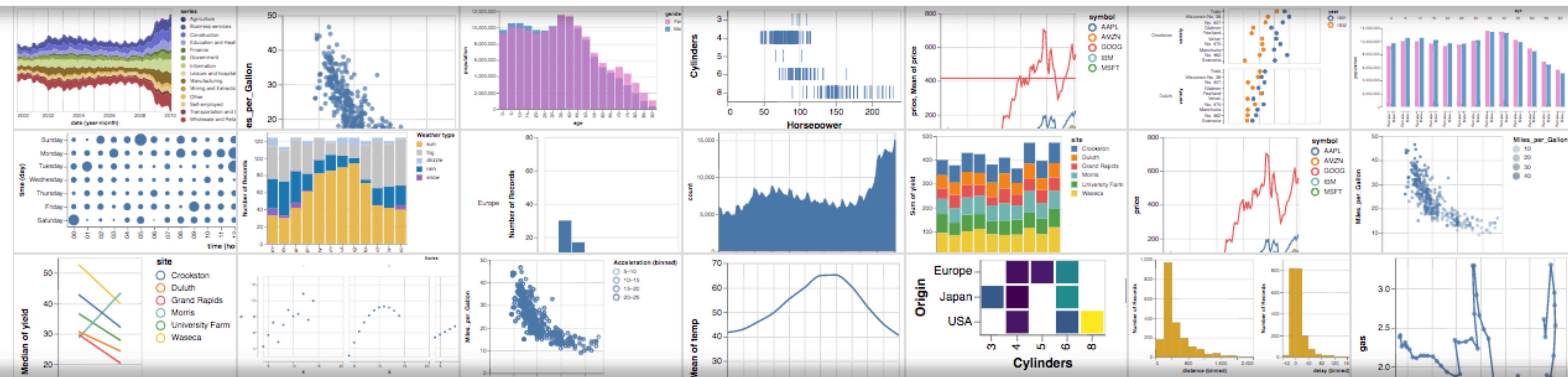
**Become a member.**

*Learn to visualize your data.  
From beginner to advanced.*

[WHAT YOU GET](#)**Categories****Visualization***Seeing data***Statistics***Analyzing data***Maps***Seeing geographic data***Software***Working with data***Infographics***Explaining data***Sources***Getting data***Networks***Connecting data***Design***Making data readable***Favorites**

Most popular porn searches, by state

# Vega-Lite – A Grammar of Interactive Graphics



**Vega-Lite** is a high-level grammar of interactive graphics. It provides a concise JSON syntax for rapidly generating visualizations to support analysis. Vega-Lite specifications can be compiled to [Vega](#) specifications.

Vega-Lite specifications describe visualizations as mappings from data to **properties of graphical marks** (e.g., points or bars). The Vega-Lite compiler **automatically produces visualization components** including axes, legends, and scales. It then determines properties of these components based on a set of **carefully designed rules**. This approach allows specifications to be succinct and expressive, but also provide user control. As Vega-Lite is designed for analysis, it supports **data transformations** such as aggregation, binning, filtering, sorting, and **visual transformations** including stacking and faceting. Moreover, Vega-Lite specifications can be **composed** into layered and multi-view displays, and made **interactive with selections**.

[Get started](#)

Latest Version: 2.0.3

[Try online](#)

Read our [introduction article to Vega-Lite v2](#) on Medium, watch our [OpenVis Conf talk](#) about the new features in Vega-Lite v2, check out the [documentation](#) and take a look at our [example gallery](#).

# Yesterday

## *Fundamental*

---

1. Value of visualization
2. Design principles
3. Graphical perception

# Today

## *Practical*

---

1. Data model and visual encoding
2. Exploratory data analysis
3. Storytelling with data
4. Advanced visualizations

Good luck on your visualizations  
and share your knowledge with others!