Beyond Basics- Advanced Topics in Visualization

Nam Wook Kim

Mini-Courses — January @ GSAS
2018

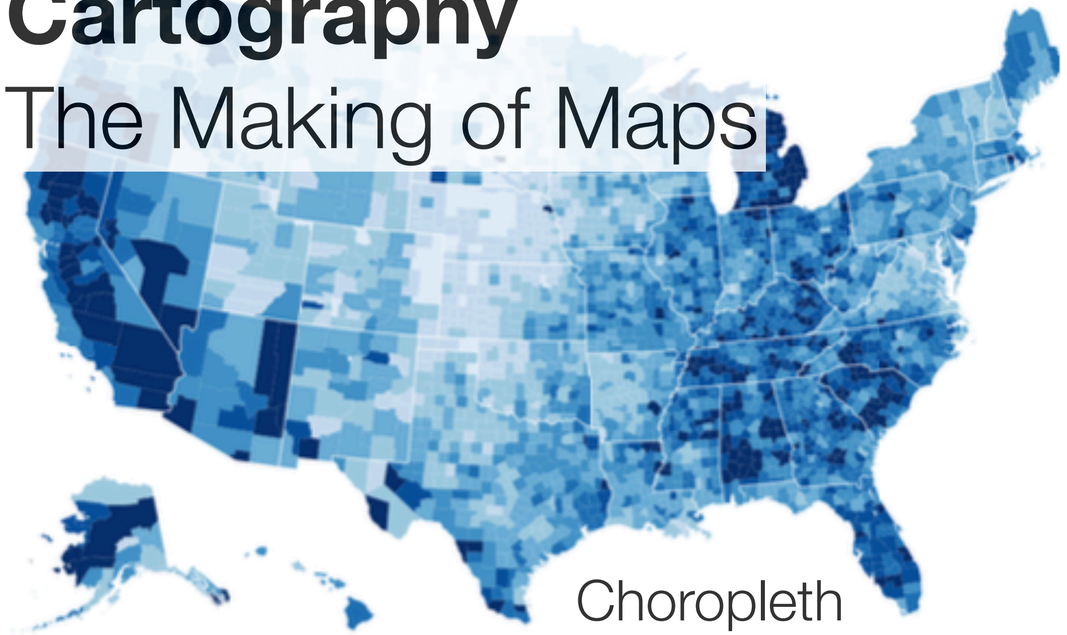
Topics

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

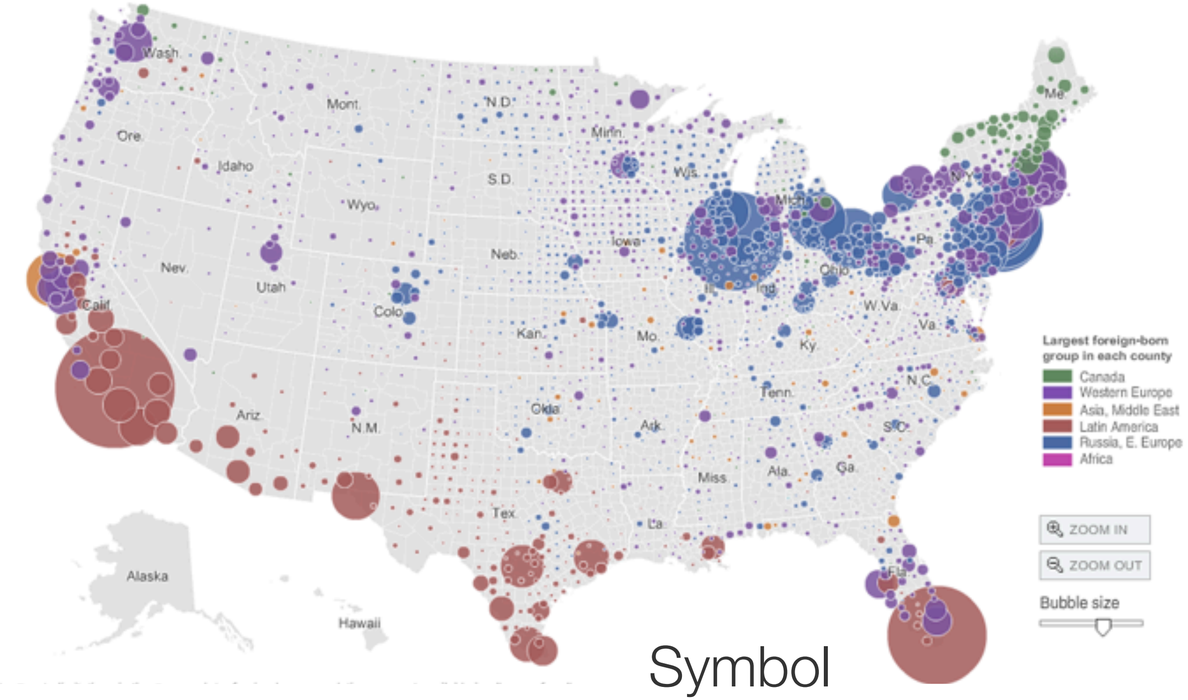
Maps

Cartography

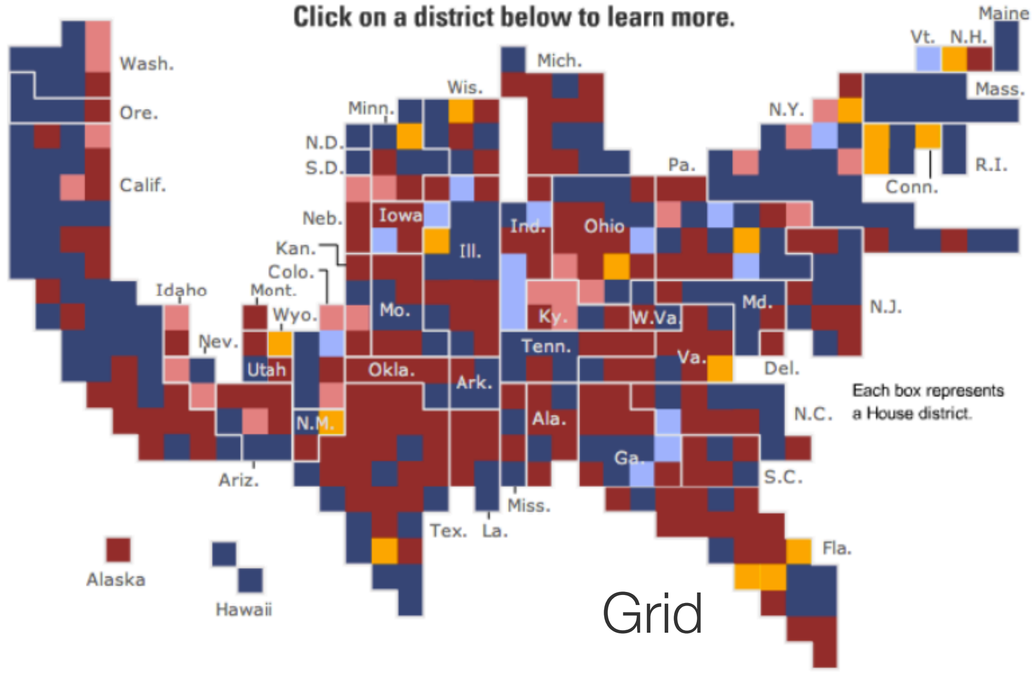
The Making of Maps



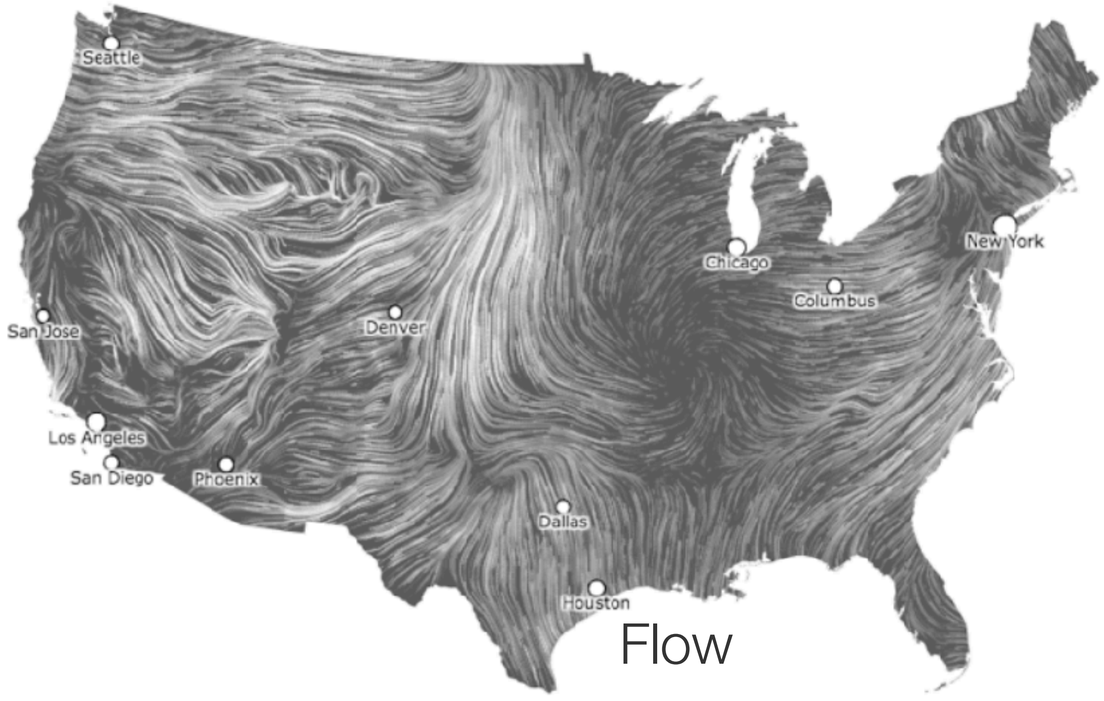
Choropleth



Symbol



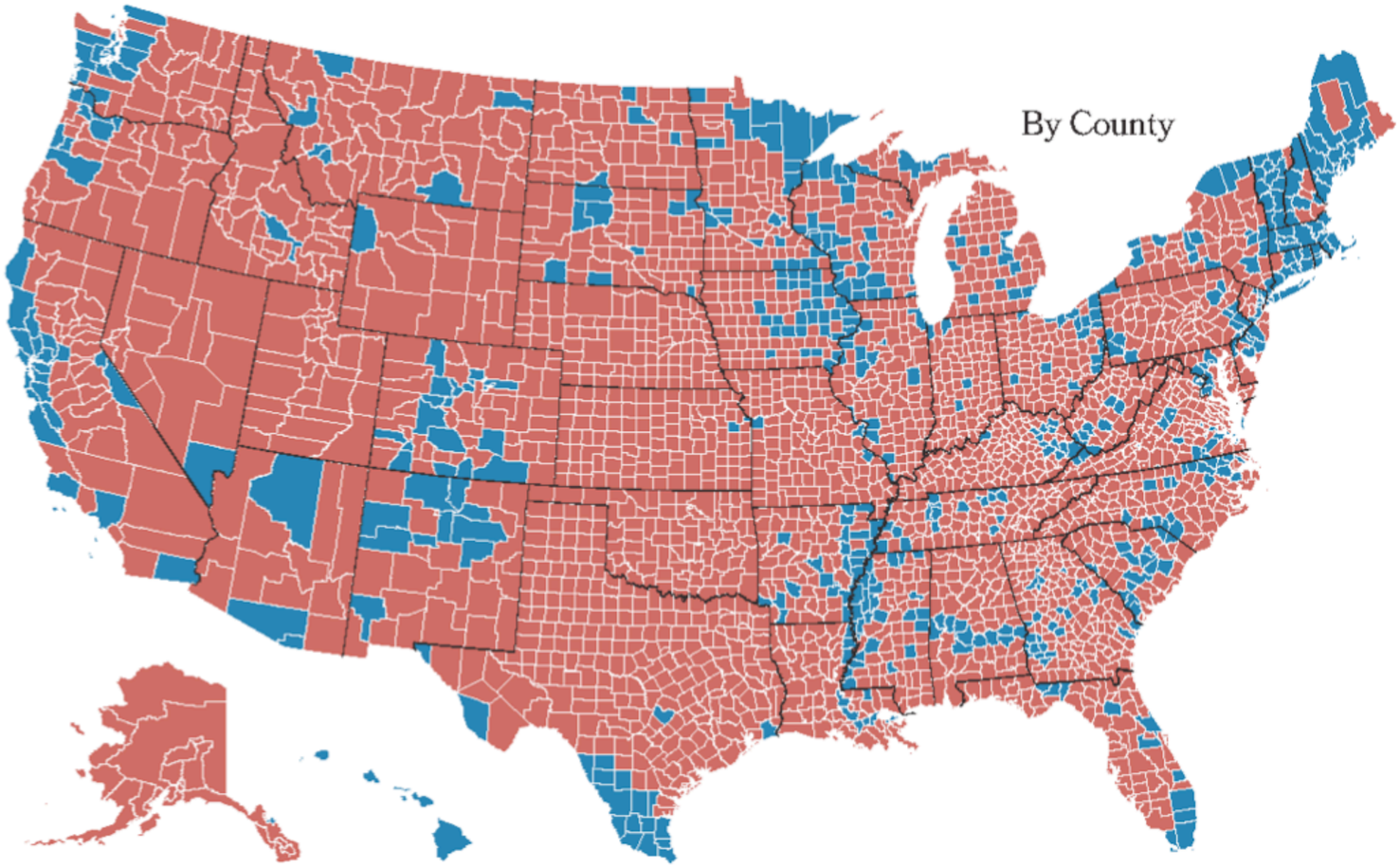
Grid



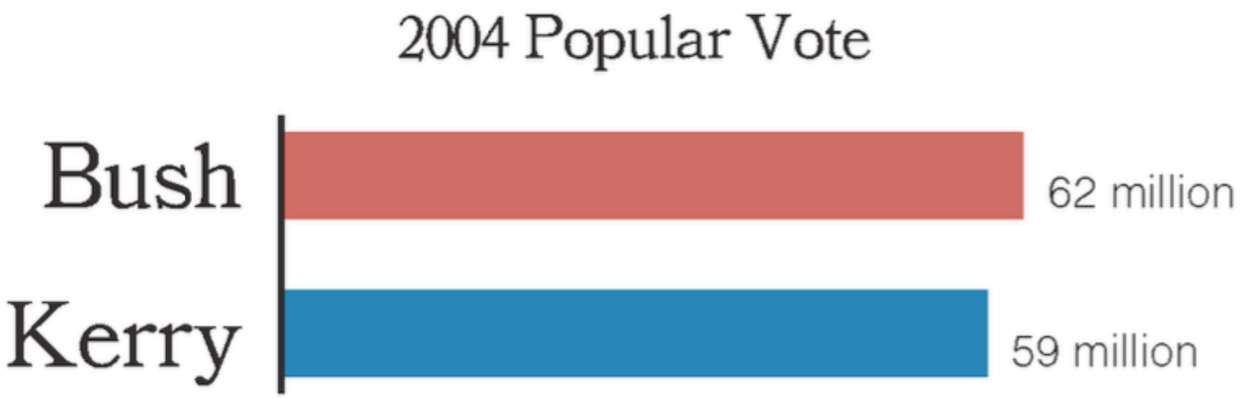
Flow

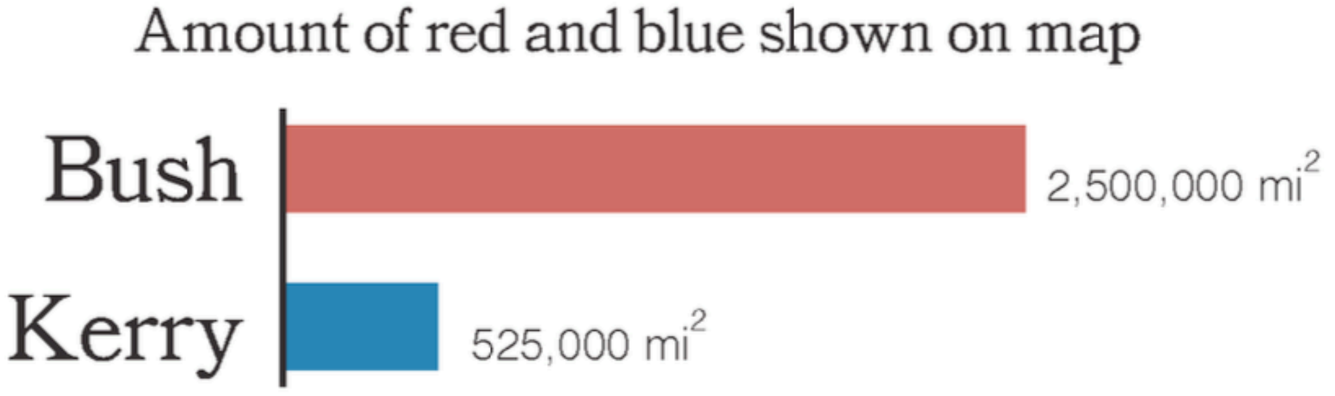
Problem with Choropleth

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

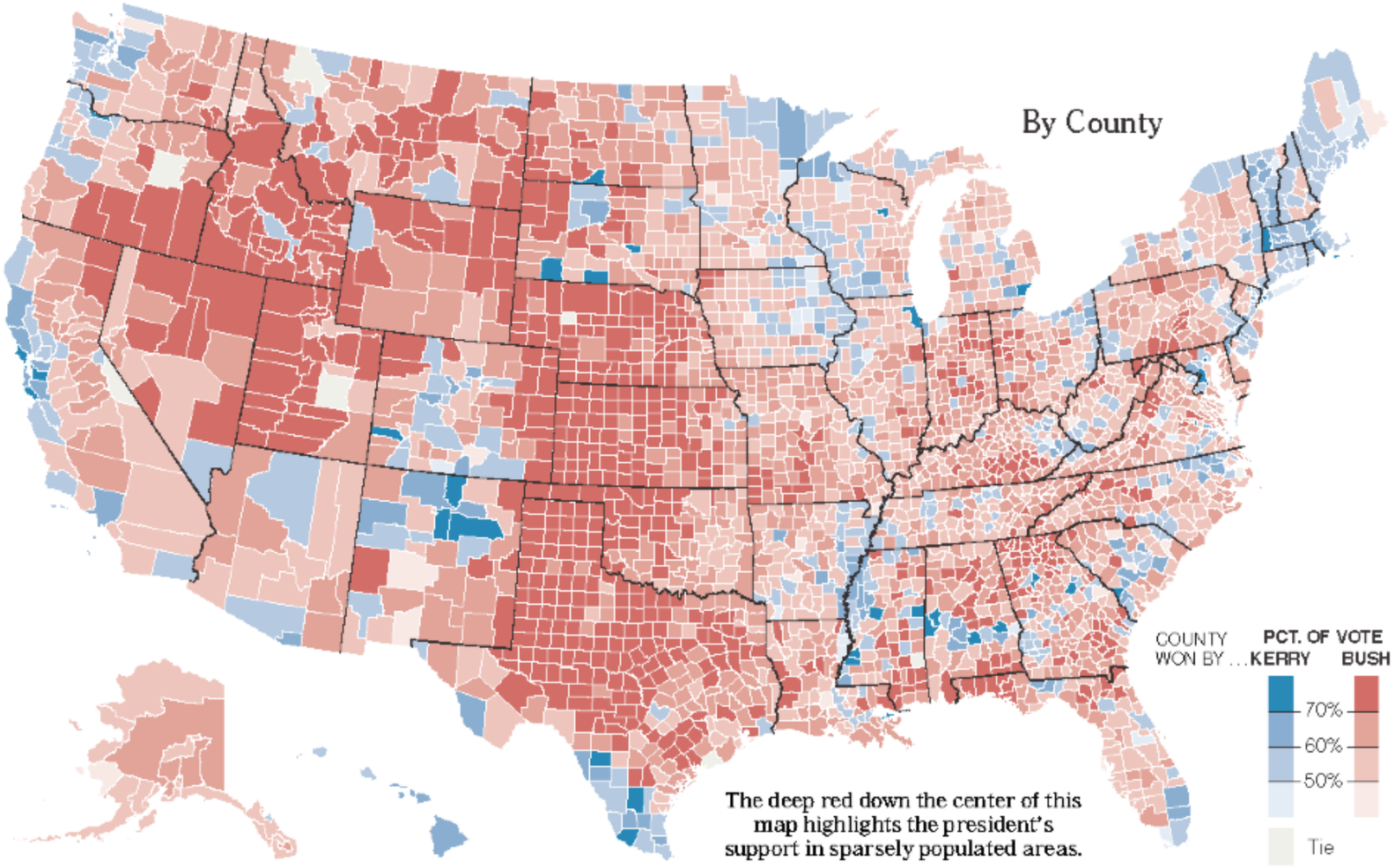


[Matthew Ericson, NY Times]

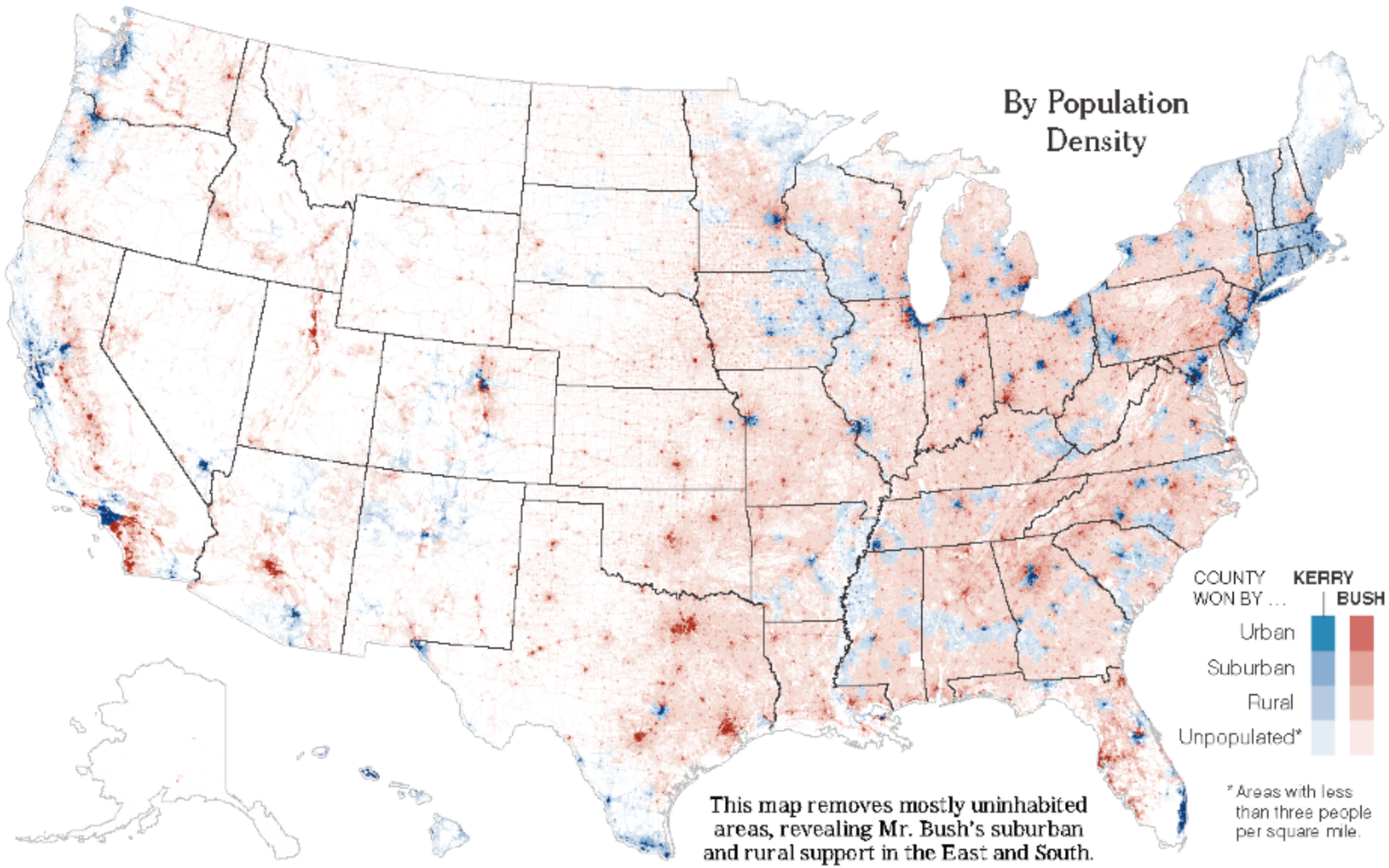

2004 Popular Vote
Bush
62 million
Kerry
59 million


Amount of red and blue shown on map
Bush
2,500,000 mi2
Kerry
525,000 mi2

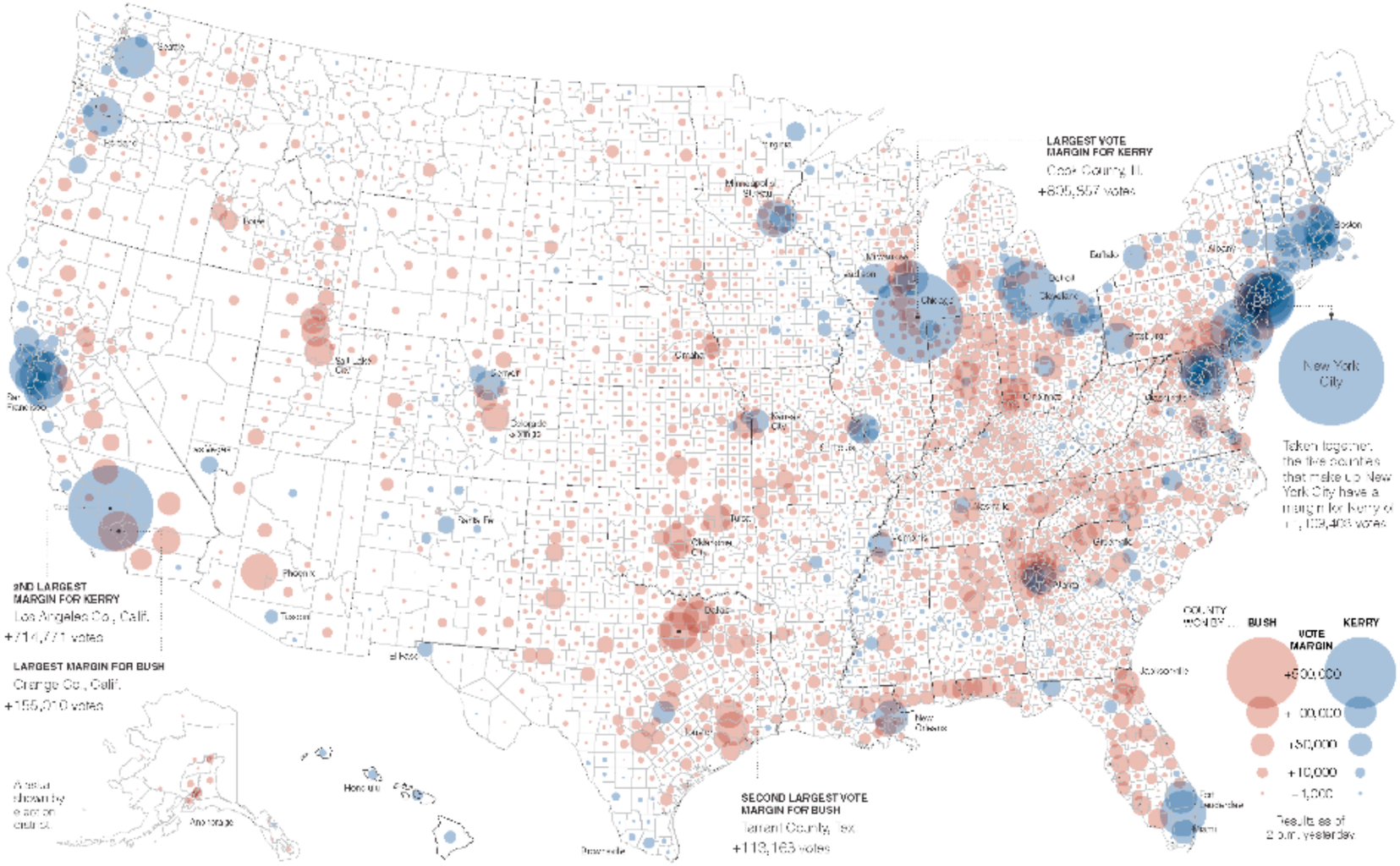
Take into account for the number of votes.

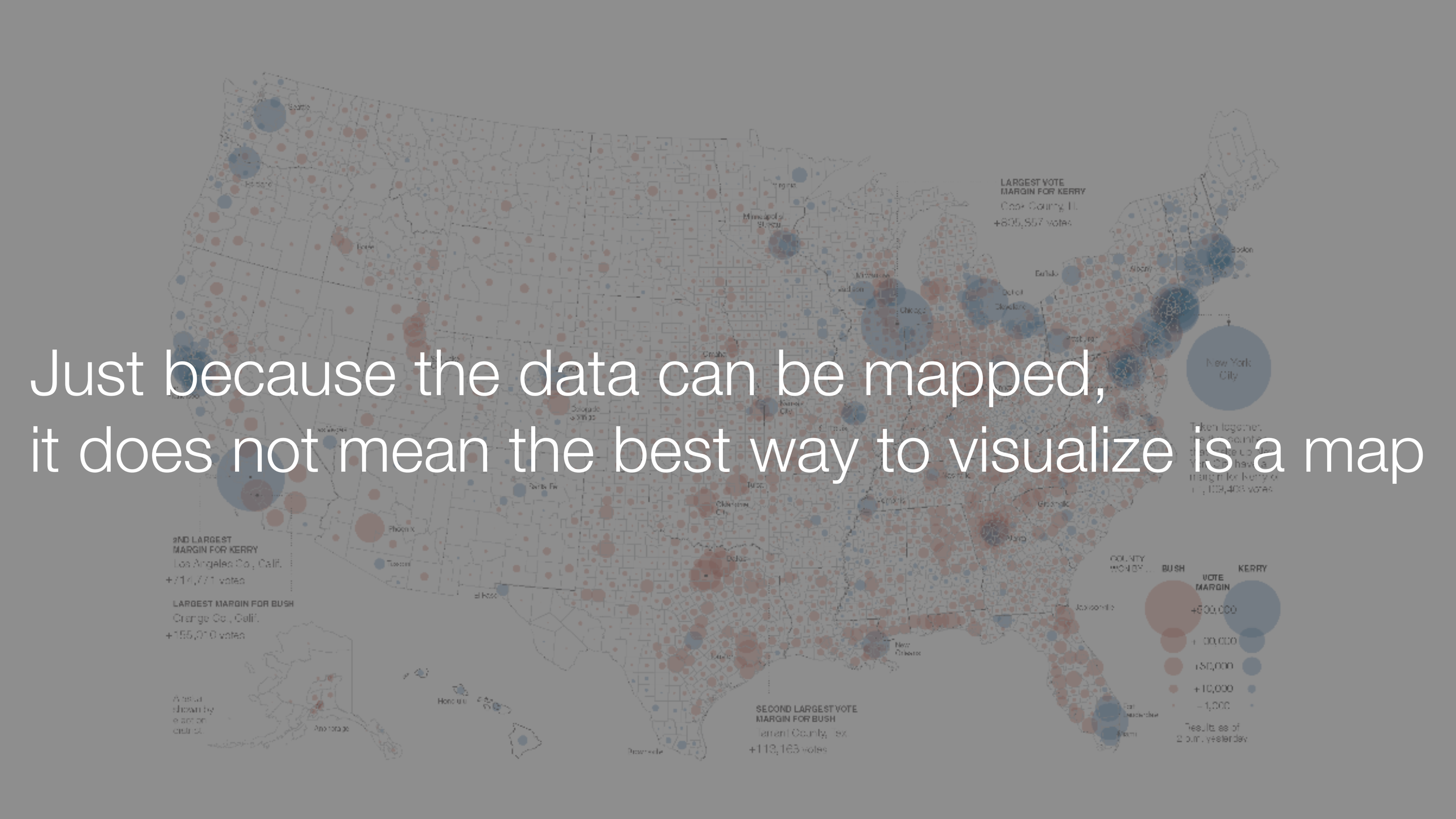


Take into account for population density



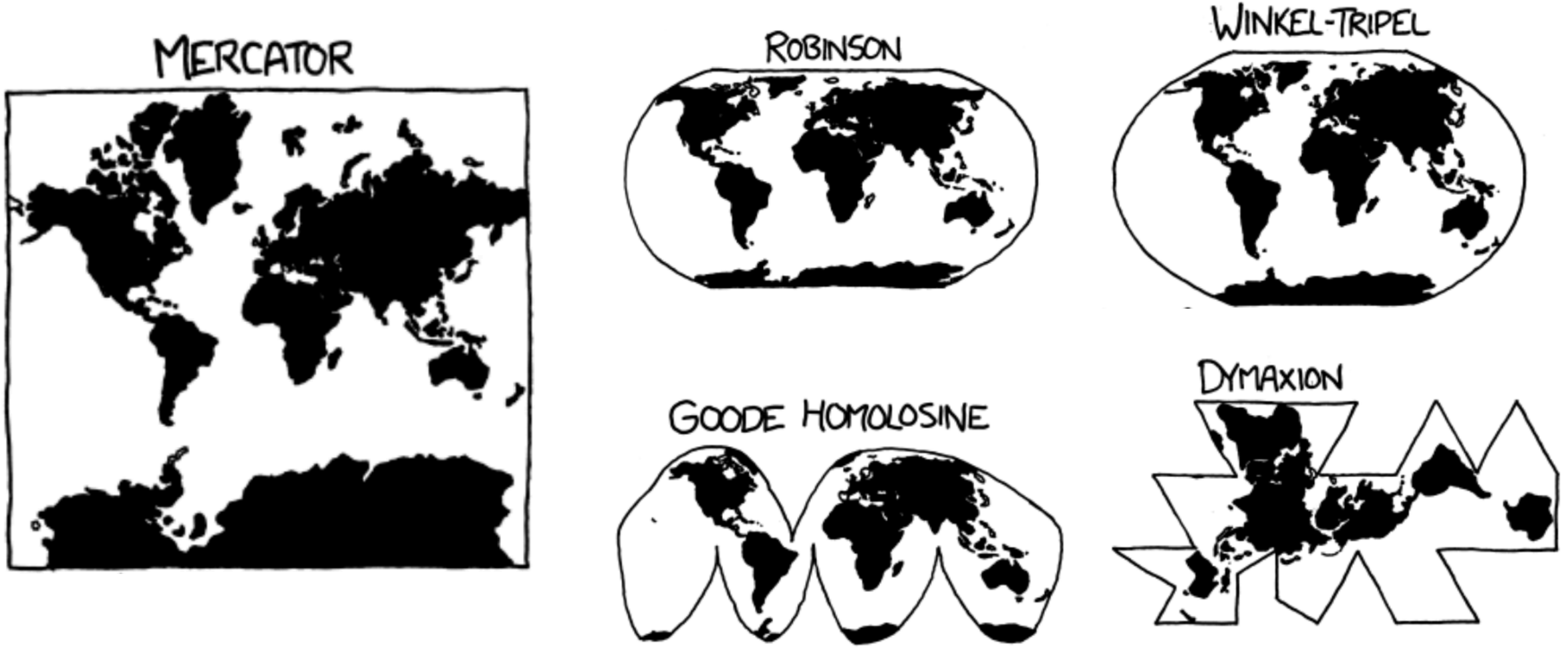
Symbol Map


LARGEST VOTE MARGIN FOR KERRY
Cook County, Ill.
+609,557 votes
New York City
Taken together, the five counties that make up New York City have a margin for Kerry of [illegible],403 votes
2ND LARGEST MARGIN FOR KERRY
Los Angeles Co., Calif.
+714,771 votes
LARGEST MARGIN FOR BUSH
Orange Co., Calif.
+155,210 votes
SECOND LARGEST VOTE MARGIN FOR BUSH
Tarrant County, Tex.
+113,163 votes
COUNTY WON BY...
BUSH
KERRY
VOTE MARGIN
+500,000
+ 100,000
+50,000
+10,000
+ 1,000
Results as of 2 p.m. yesterday

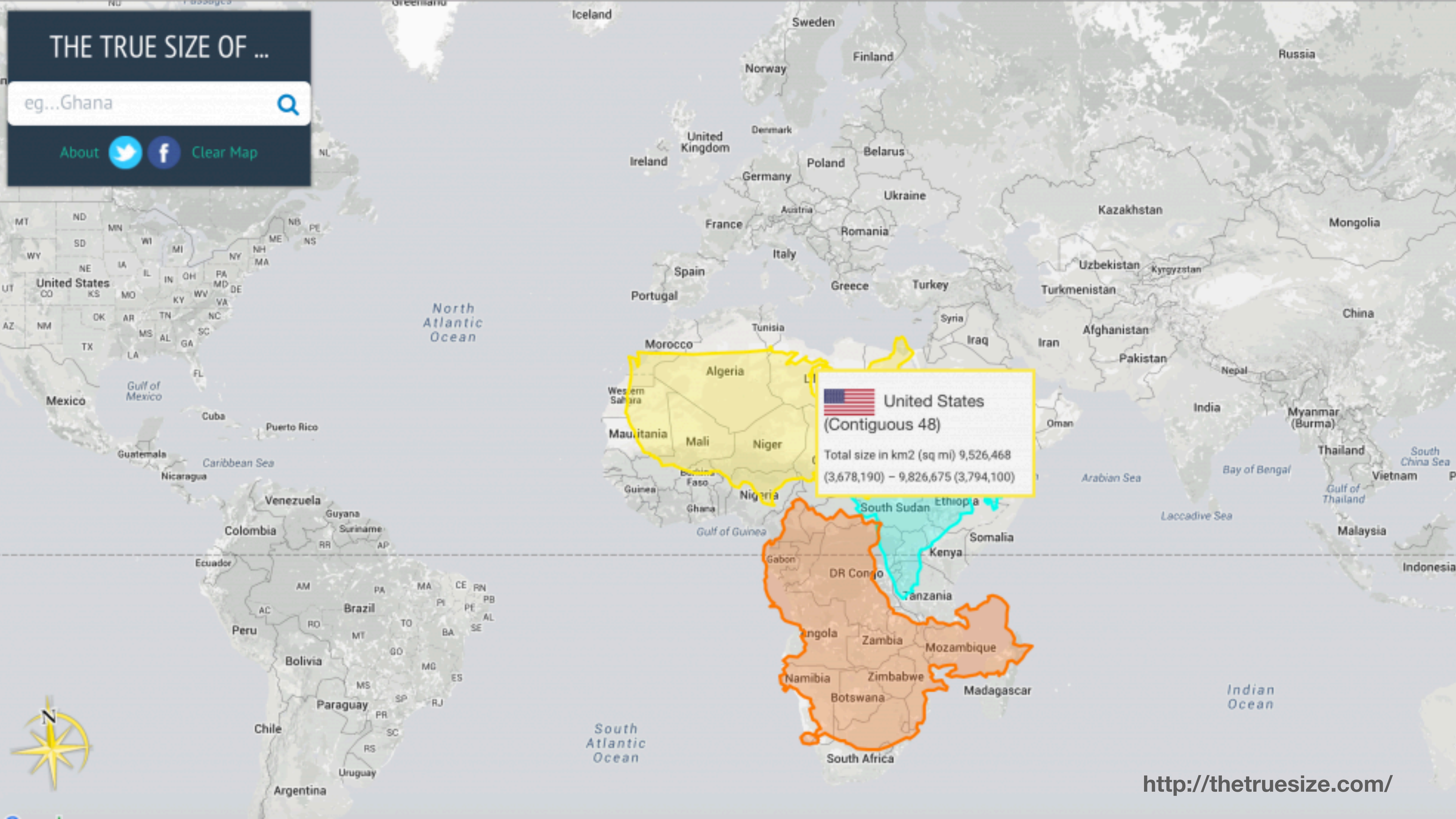


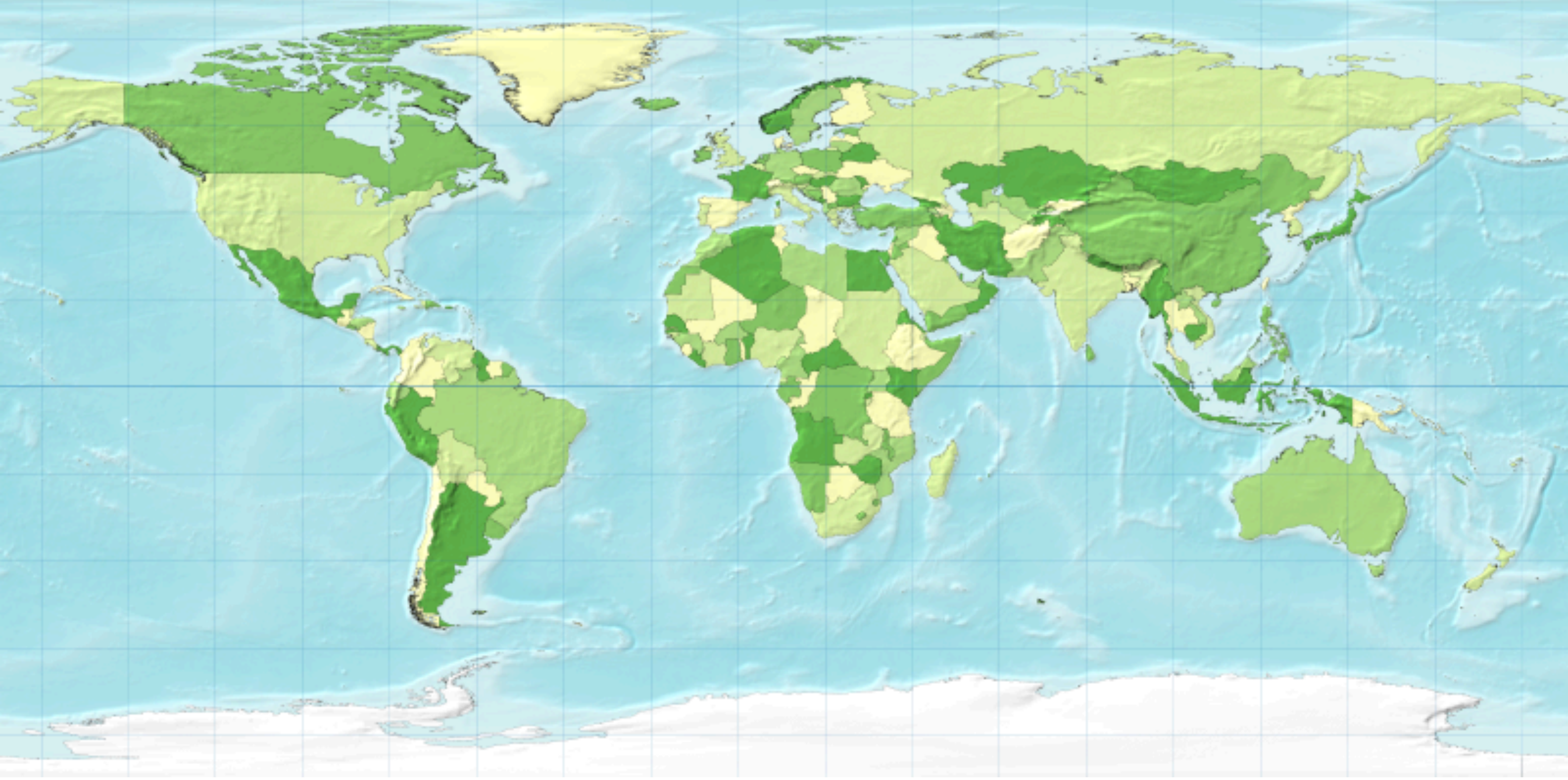
Just because the data can be mapped,
it does not mean the best way to visualize is a map

Map Projections



[http://xkcd.com/977/]


THE TRUE SIZE OF ...
eg...Ghana
About
Clear Map
United States
(Contiguous 48)
Total size in km2 (sq mi) 9,526,468
(3,678,190) – 9,826,675 (3,794,100)
Algeria
Mali
Niger
South Sudan
Ethiopia
Somalia
Kenya
Tanzania
DR Congo
Angola
Zambia
Mozambique
Namibia
Zimbabwe
Botswana
South Africa
Madagascar
Libya
Egypt
Sudan
Chad
Nigeria
Ghana
Guinea
Mauritania
Western Sahara
Morocco
Tunisia
Gabon
Iceland
Sweden
Finland
Norway
Russia
United Kingdom
Ireland
Denmark
Germany
Poland
Belarus
Ukraine
France
Austria
Romania
Italy
Spain
Portugal
Greece
Turkey
Syria
Iraq
Iran
Afghanistan
Pakistan
Kazakhstan
Uzbekistan
Turkmenistan
Kyrgyzstan
Mongolia
China
India
Nepal
Myanmar (Burma)
Thailand
Vietnam
Malaysia
Indonesia
Oman
Arabian Sea
Bay of Bengal
Laccadive Sea
Gulf of Thailand
South China Sea
United States
Mexico
Gulf of Mexico
Cuba
Puerto Rico
Guatemala
Nicaragua
Caribbean Sea
Venezuela
Colombia
Guyana
Suriname
Ecuador
Peru
Brazil
Bolivia
Paraguay
Chile
Uruguay
Argentina
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
Gulf of Guinea
http://thetruesize.com/

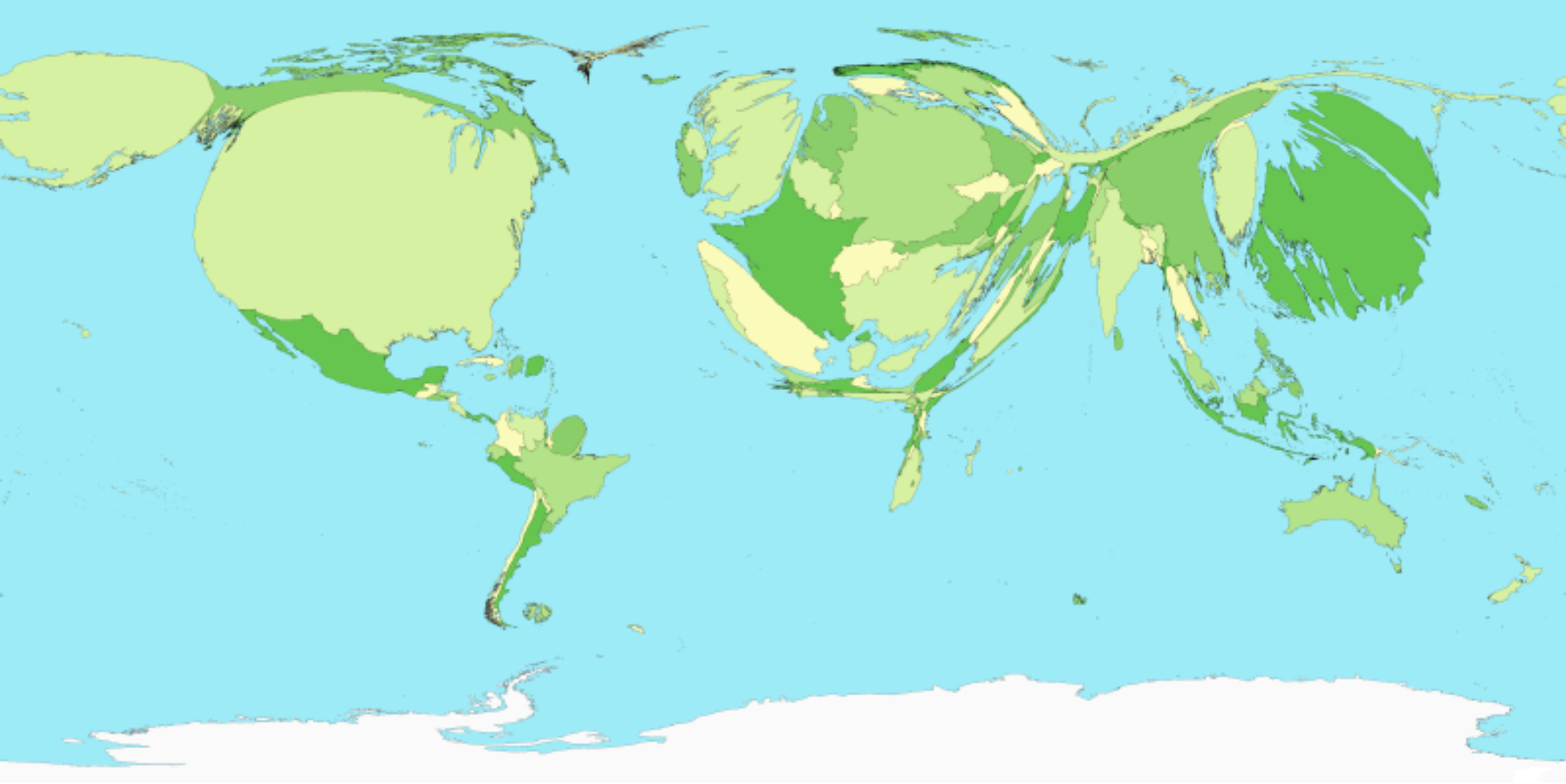


Ordinary map

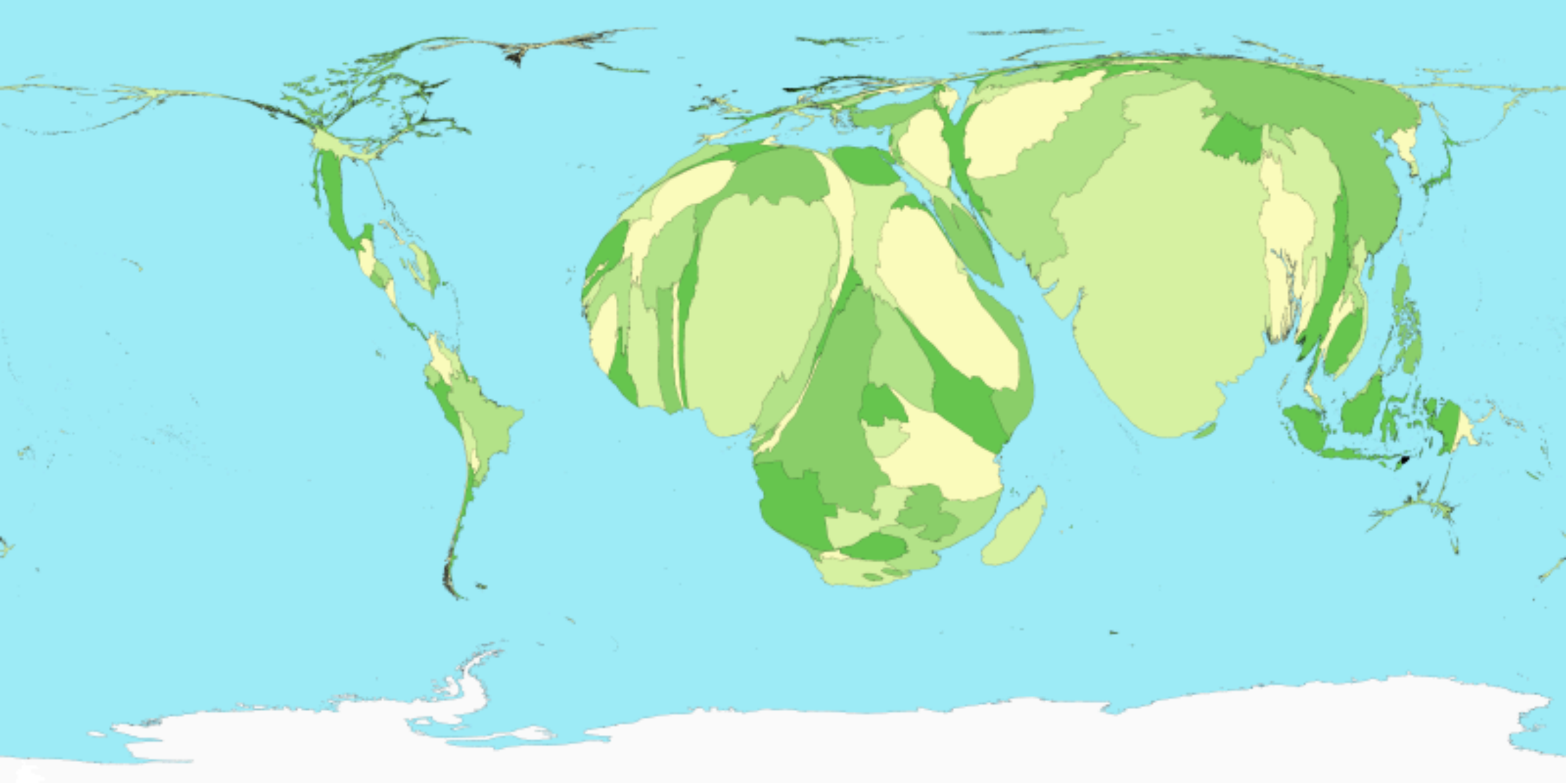
[http://www-personal.umich.edu/~mejn/cartograms/]

RUSSIA
CANADA
UNITED STATES
U.K.
GERMANY
UKRAINE
FRANCE
ITALY
TURKEY
IRAN
PAKISTAN
CHINA
INDIA
S. KOREA
JAPAN
MEXICO
EGYPT
NIGERIA
BANGLADESH
MYANMAR
THAILAND
PHILIPPINES
ETHIOPIA
COLOMBIA
PERU
BRAZIL
D.R. CONGO
INDONESIA
PACIFIC OCEAN
ATLANTIC OCEAN
SOUTH AFRICA
INDIAN OCEAN
AUSTRALIA
NEW ZEALAND
ARGENTINA
ANTARCTICA

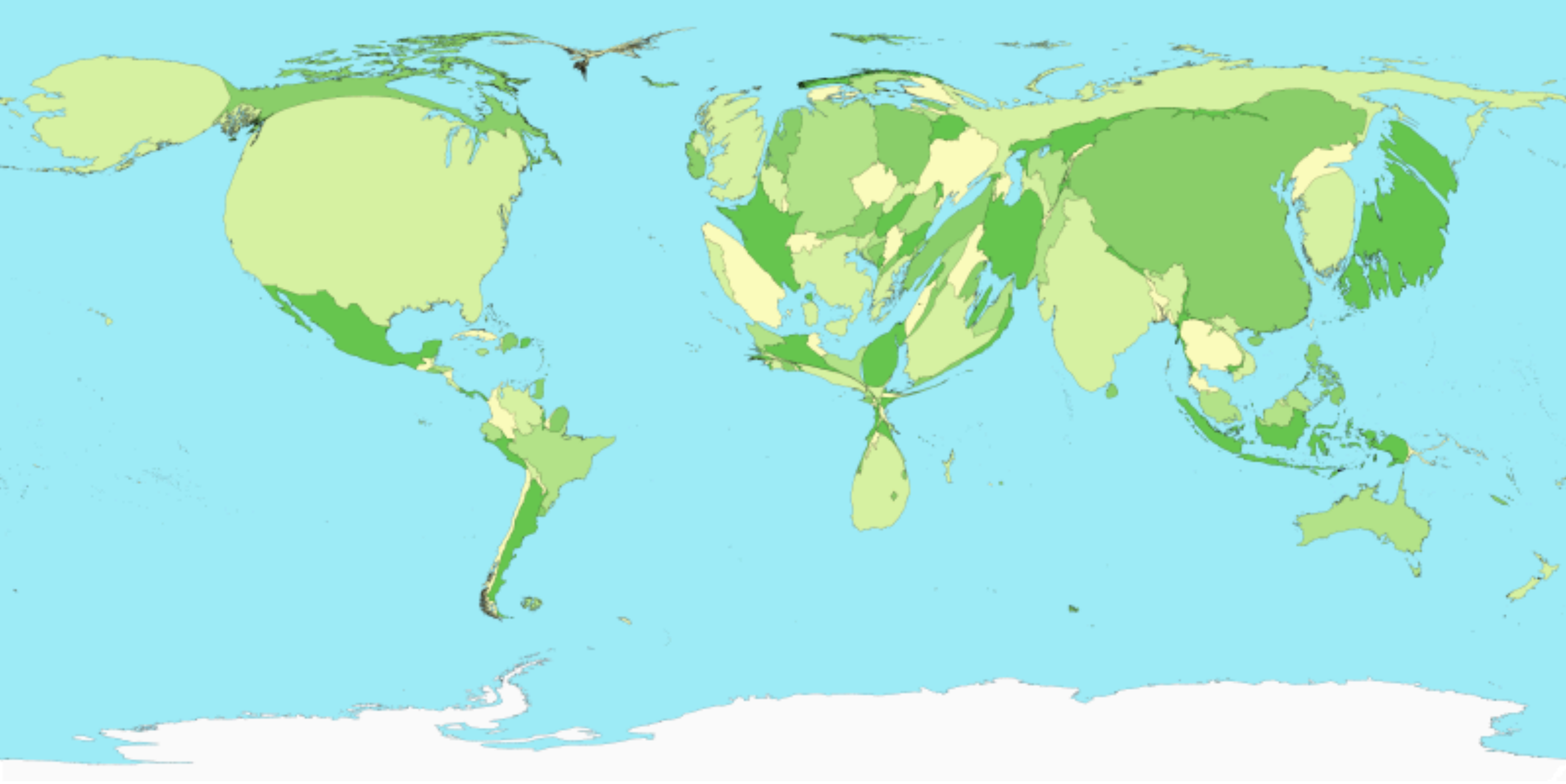
Population



Gross domestic product



Child mortality



Greenhouse gas emissions

Problem with Rainbow Colormap

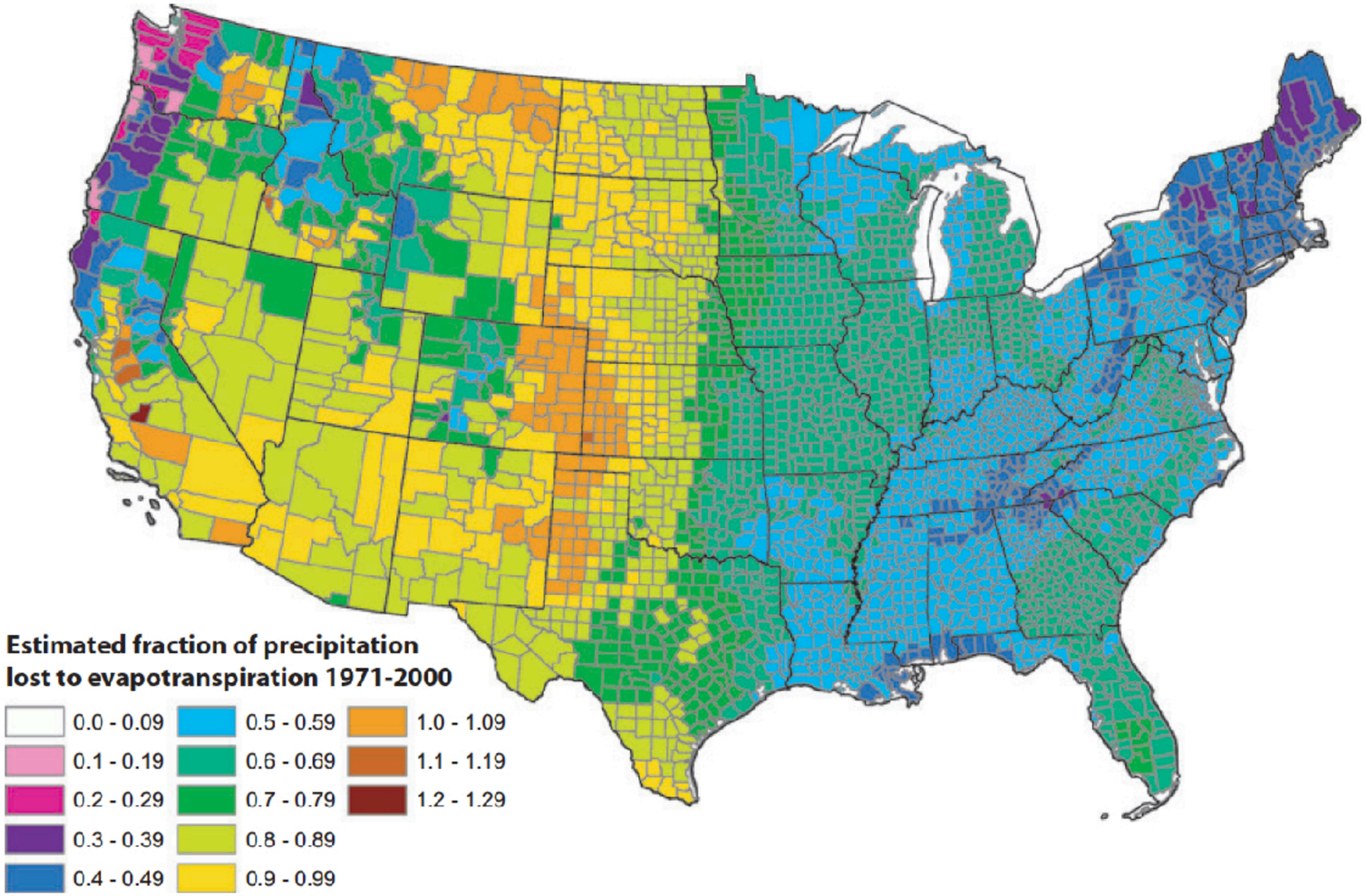
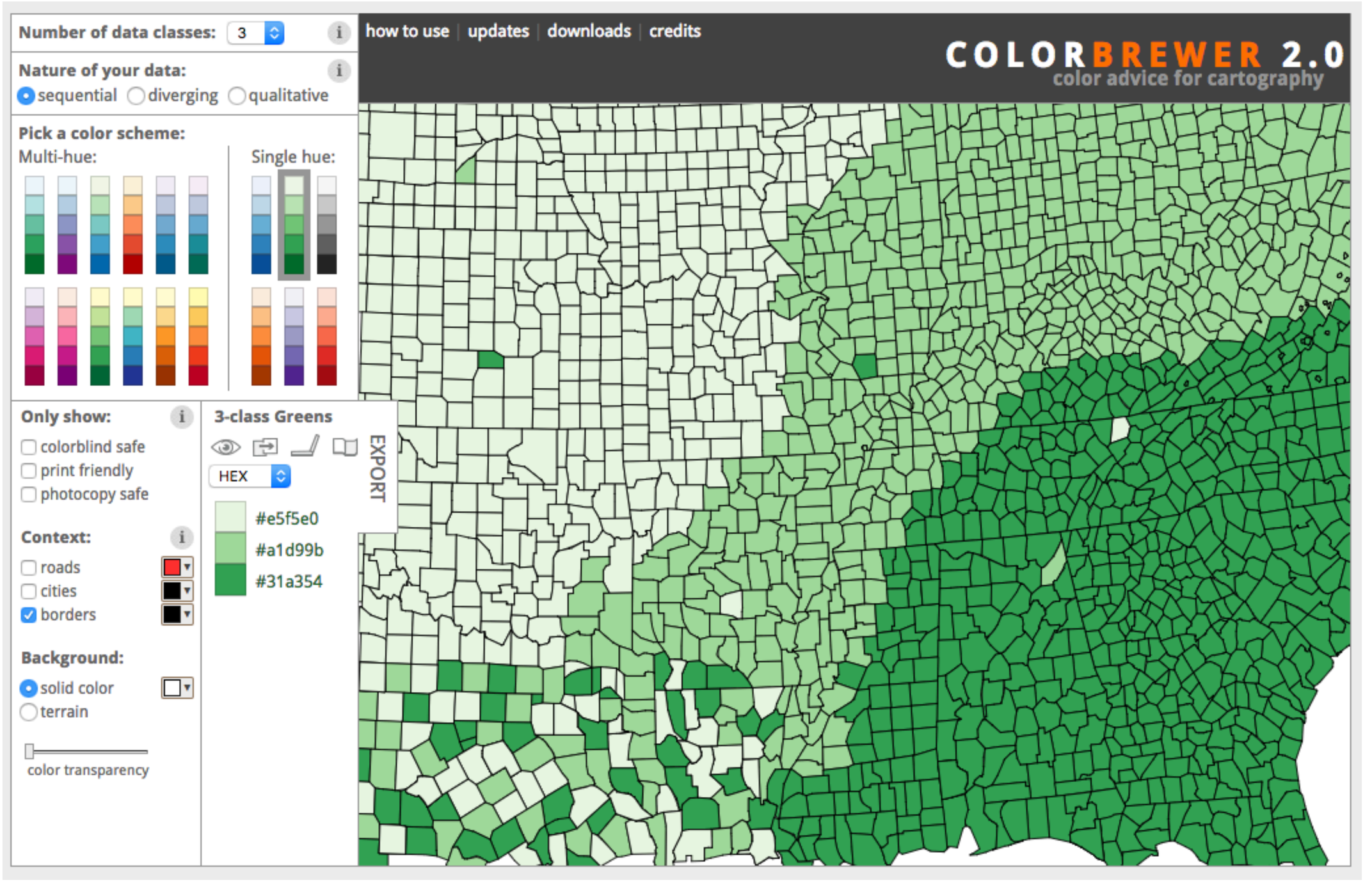


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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.

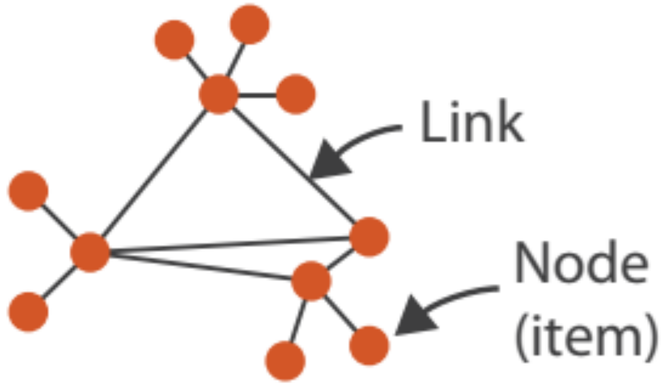

Number of data classes: 3
Nature of your data:
sequential diverging qualitative
Pick a color scheme:
Multi-hue:
Single hue:
Only show:
colorblind safe
print friendly
photocopy safe
Context:
roads
cities
borders
Background:
solid color
terrain
color transparency
3-class Greens
HEX
#e5f5e0
#a1d99b
#31a354
EXPORT
how to use | updates | downloads | credits
COLORBREWER 2.0
color advice for cartography

Networks & Trees

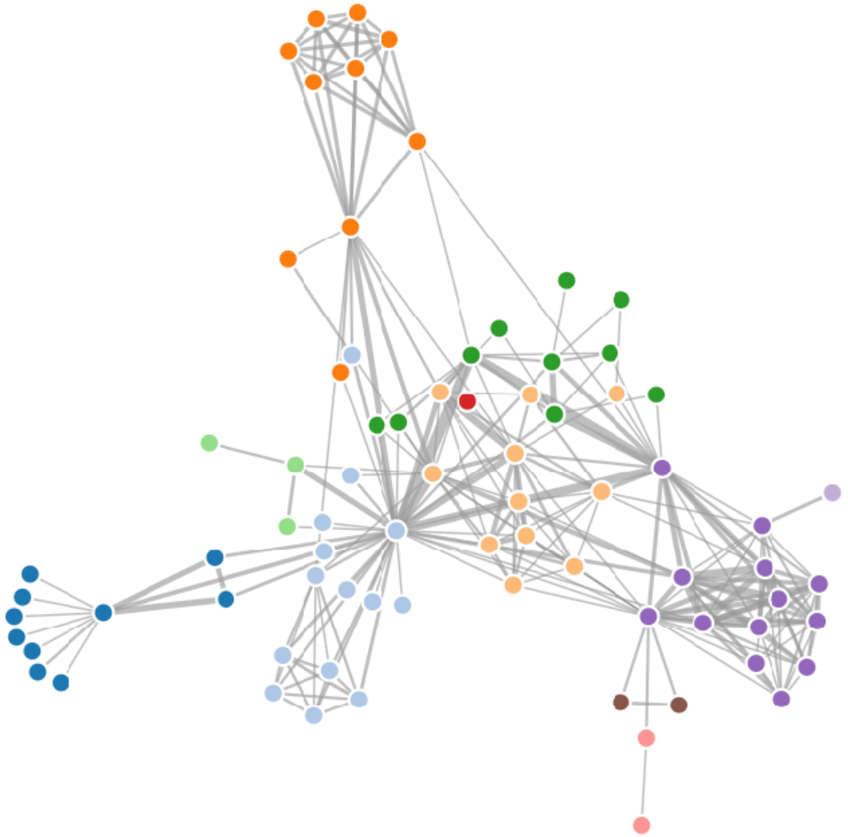
Networks

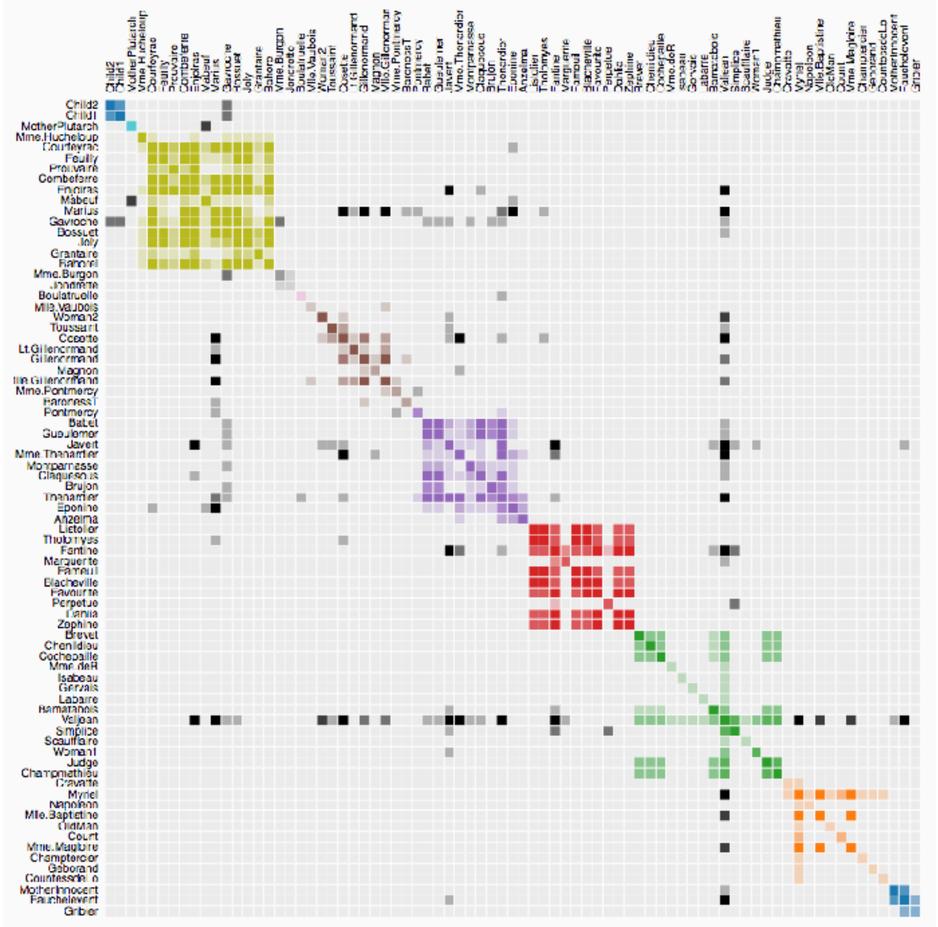
Model relations among data

Nodes and links



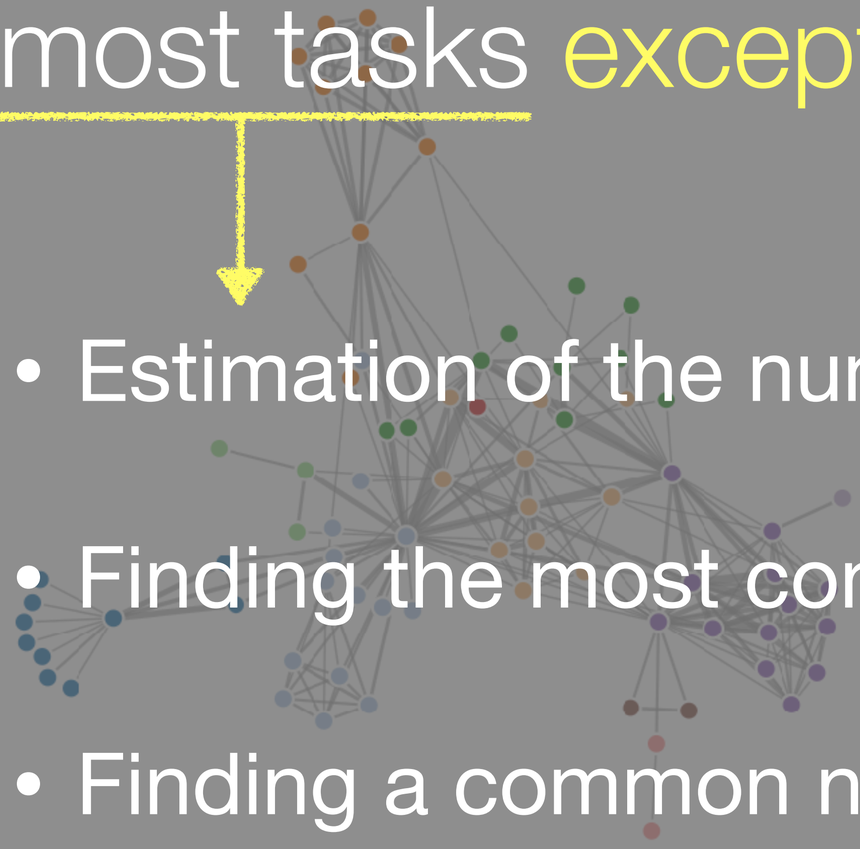
Node-Link Diagram vs Matrix Diagram

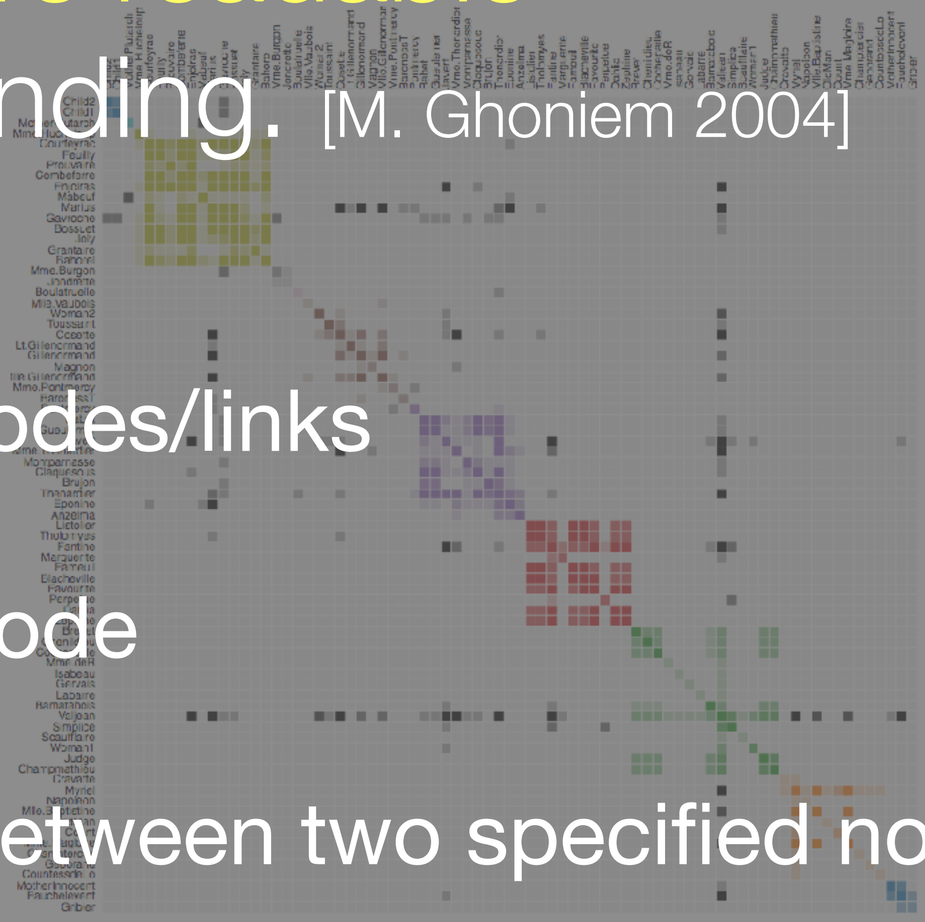




Les Misérables Character Co-occurrence

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





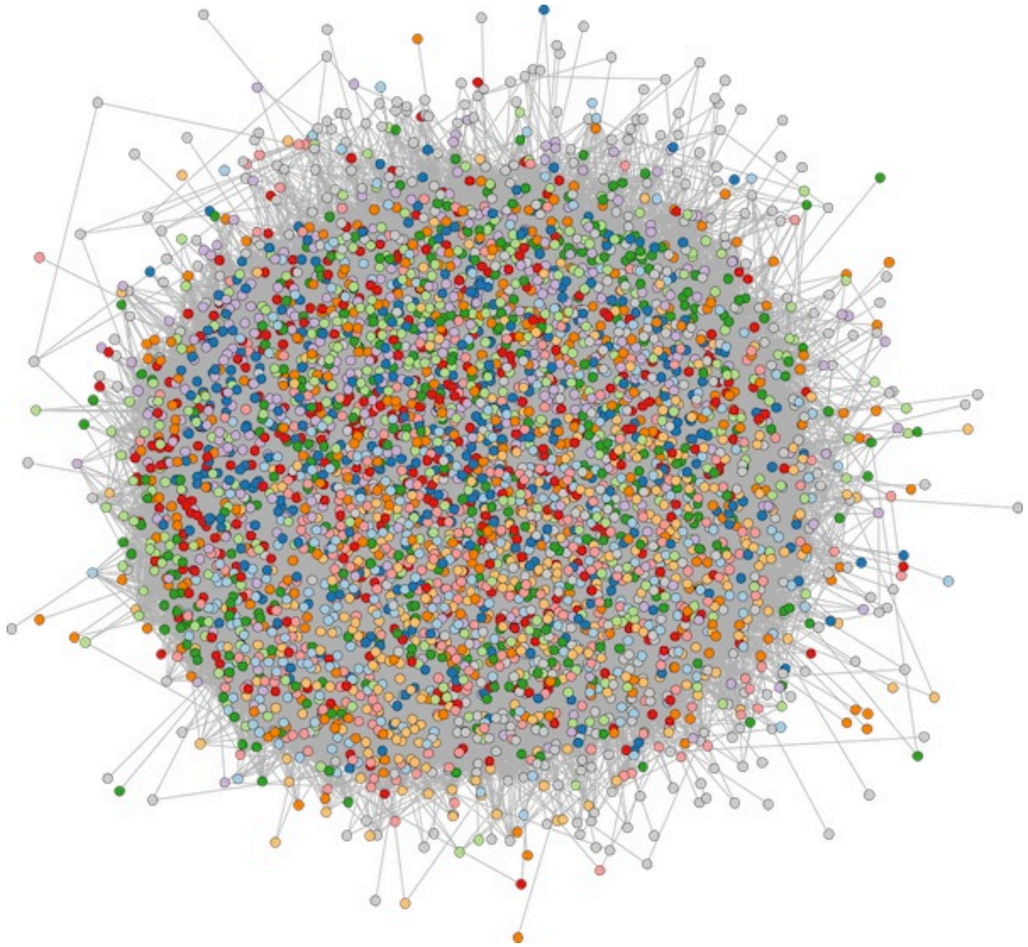
- Estimation of the number of nodes/links
- Finding the most connected node
- Finding a common neighbor between two specified nodes

Node-Link Diagram

Matrix Diagram

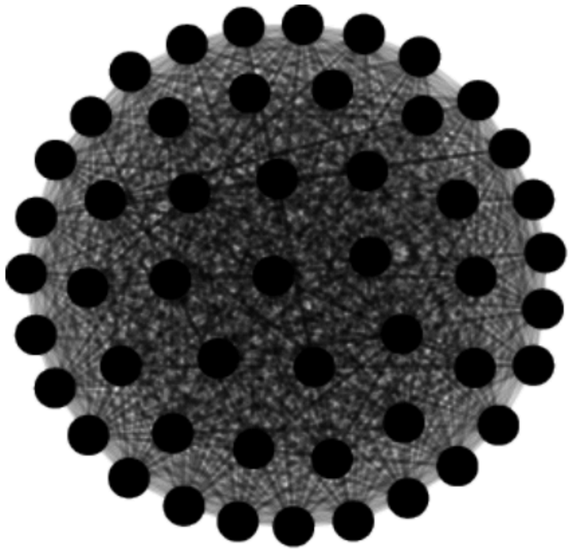
Problem with Node-link diagram

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

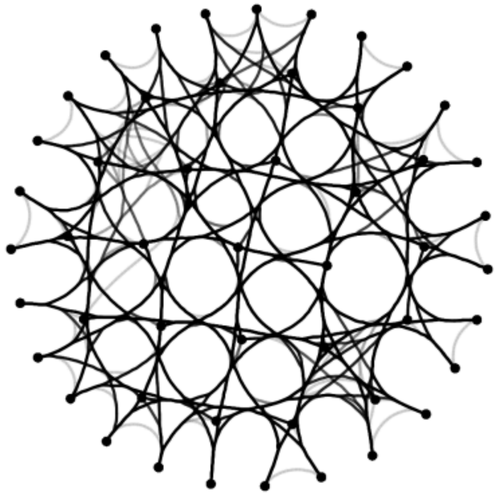


[A. Nocaj 2015]

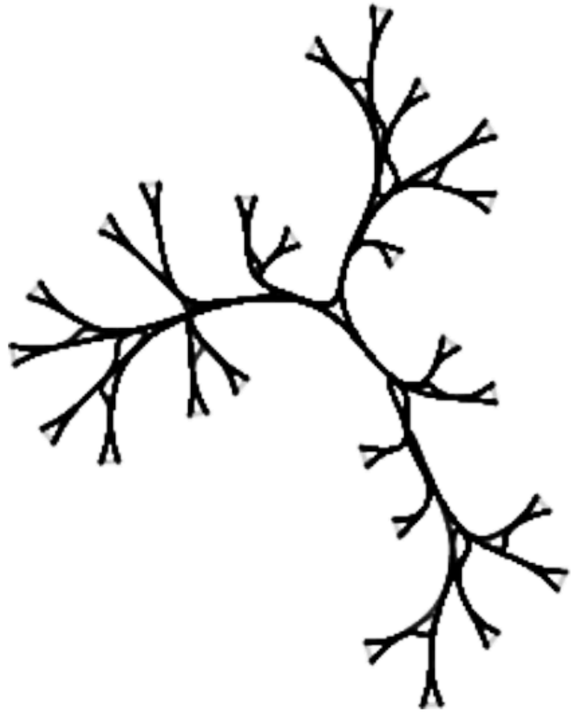
Original Network



Edge Bundling

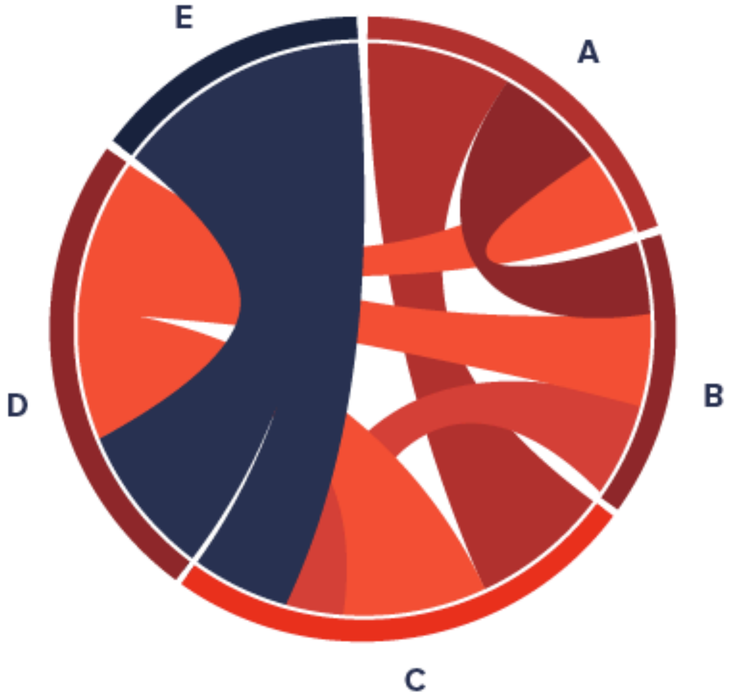


Confluent Drawing

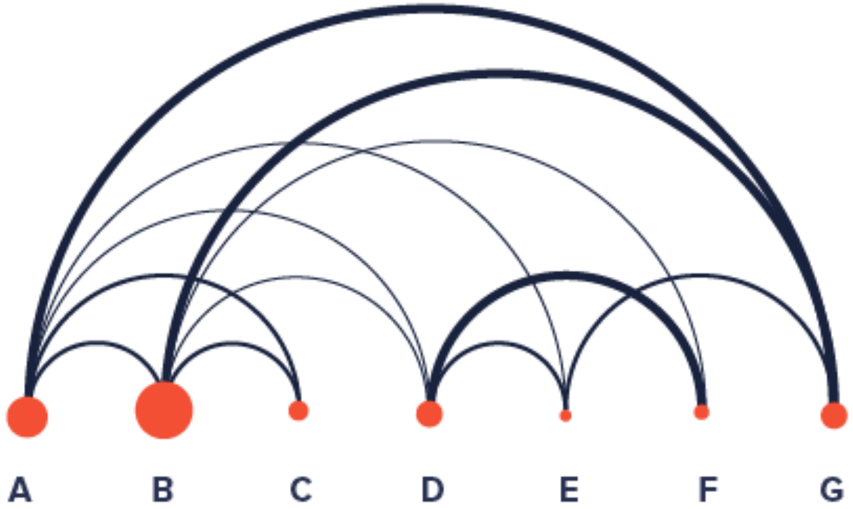


[B. Bach 2017]

Alternatives. Are these better?



Chord Diagram



Arc Diagram

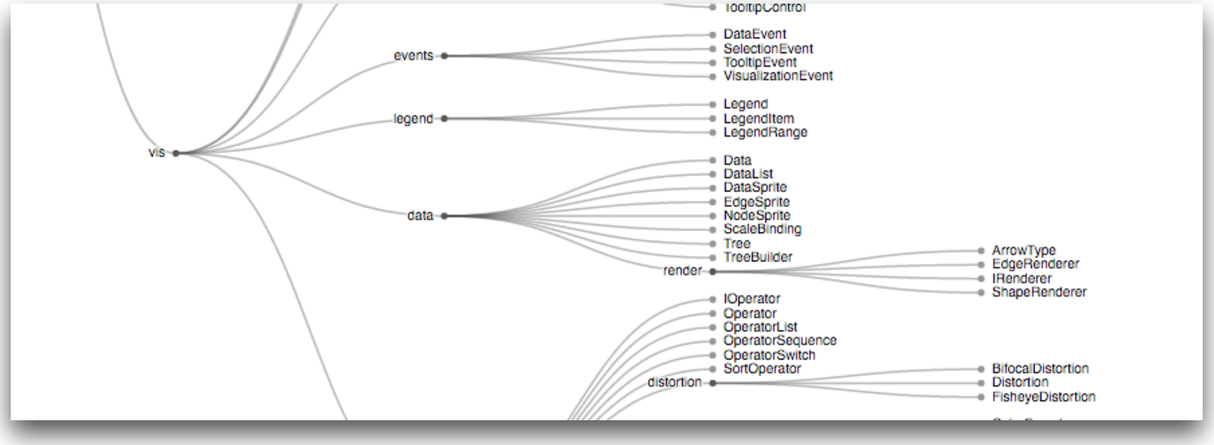
Networks & Trees

Trees

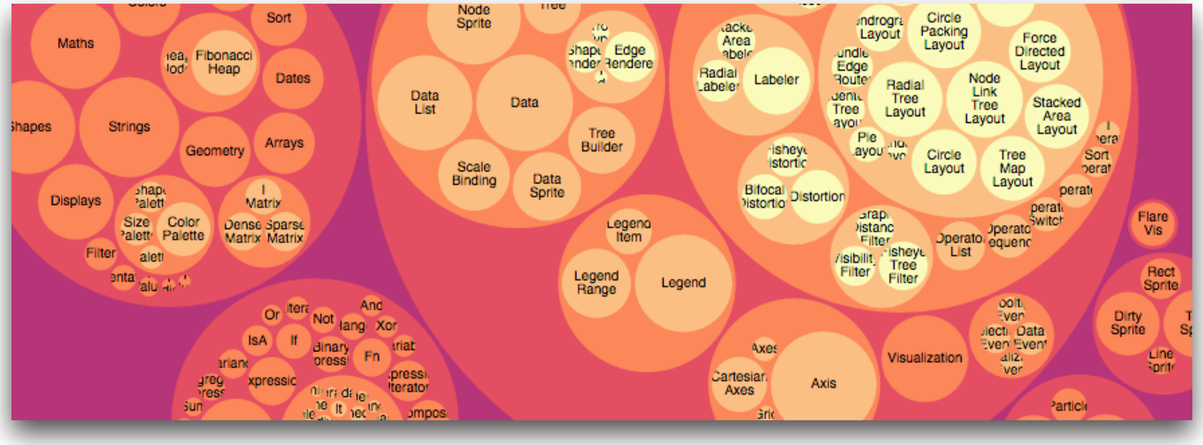
Graphs with hierarchical structure

Nodes as parents and children

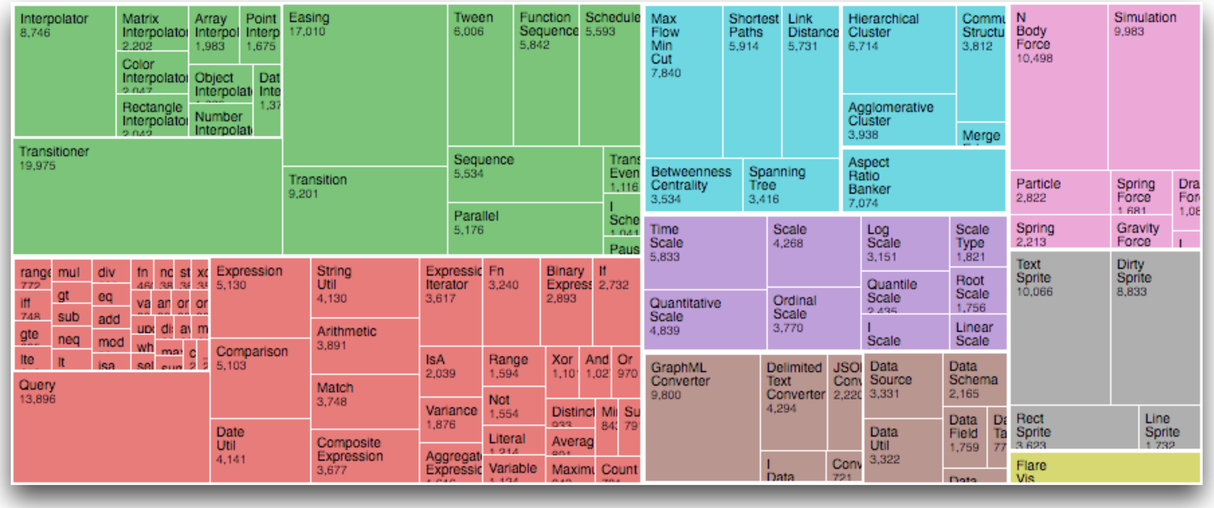
Dendogram



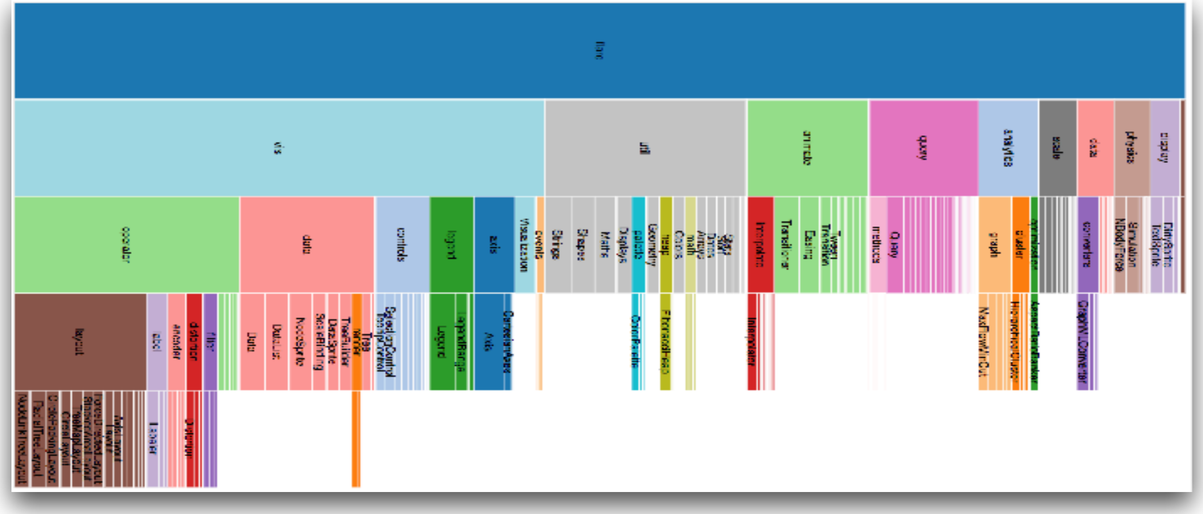
Circle Packing



Treemap

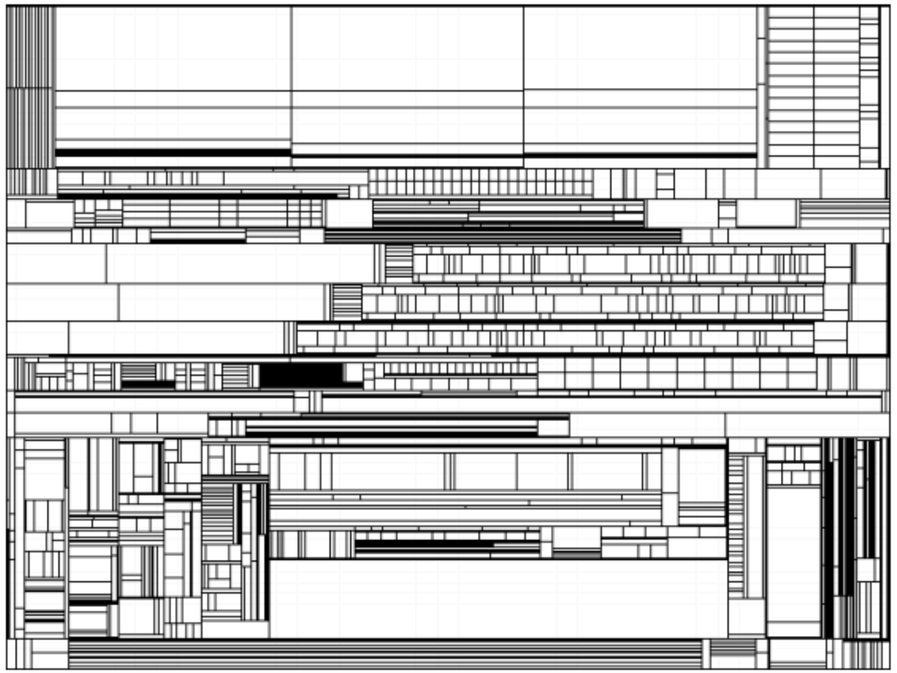


Icicle

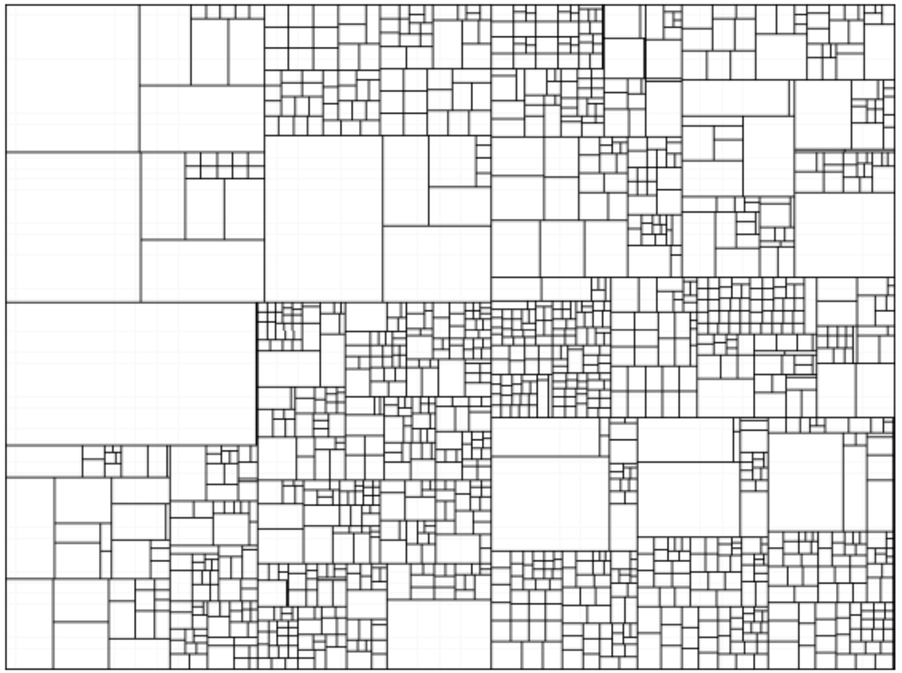


Squarified Treemaps

“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]

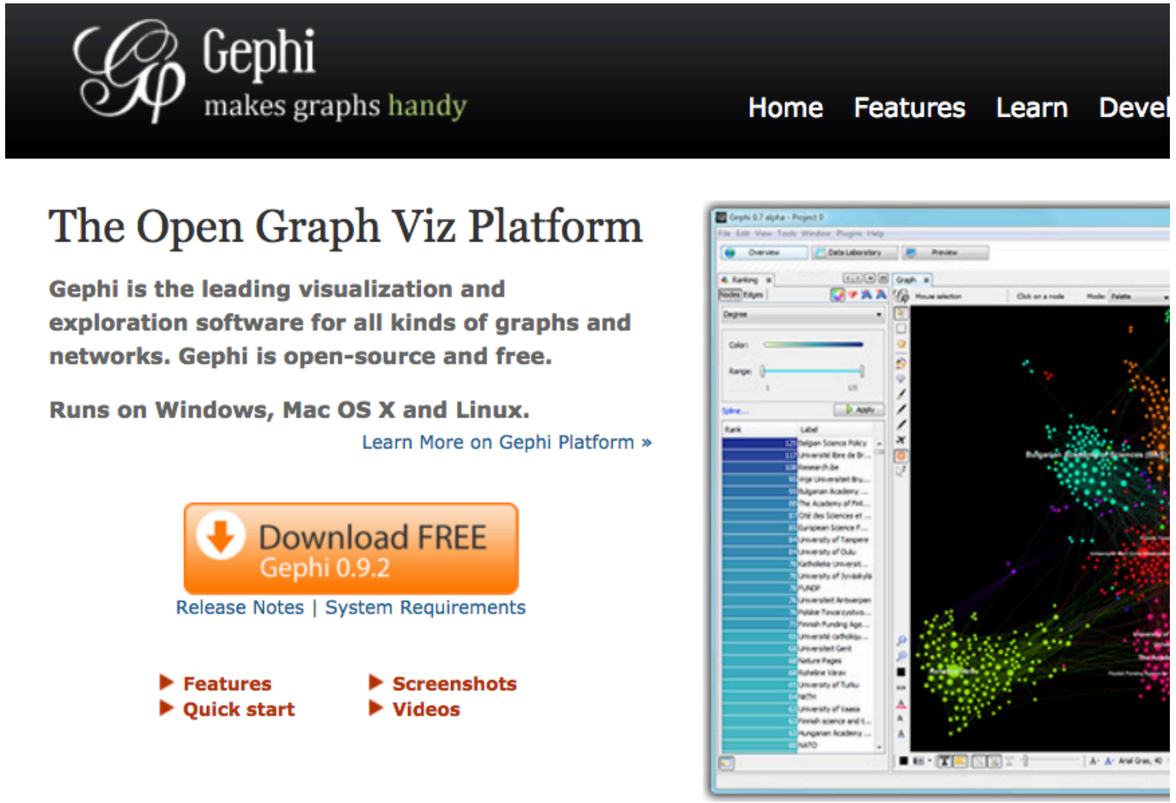


Before

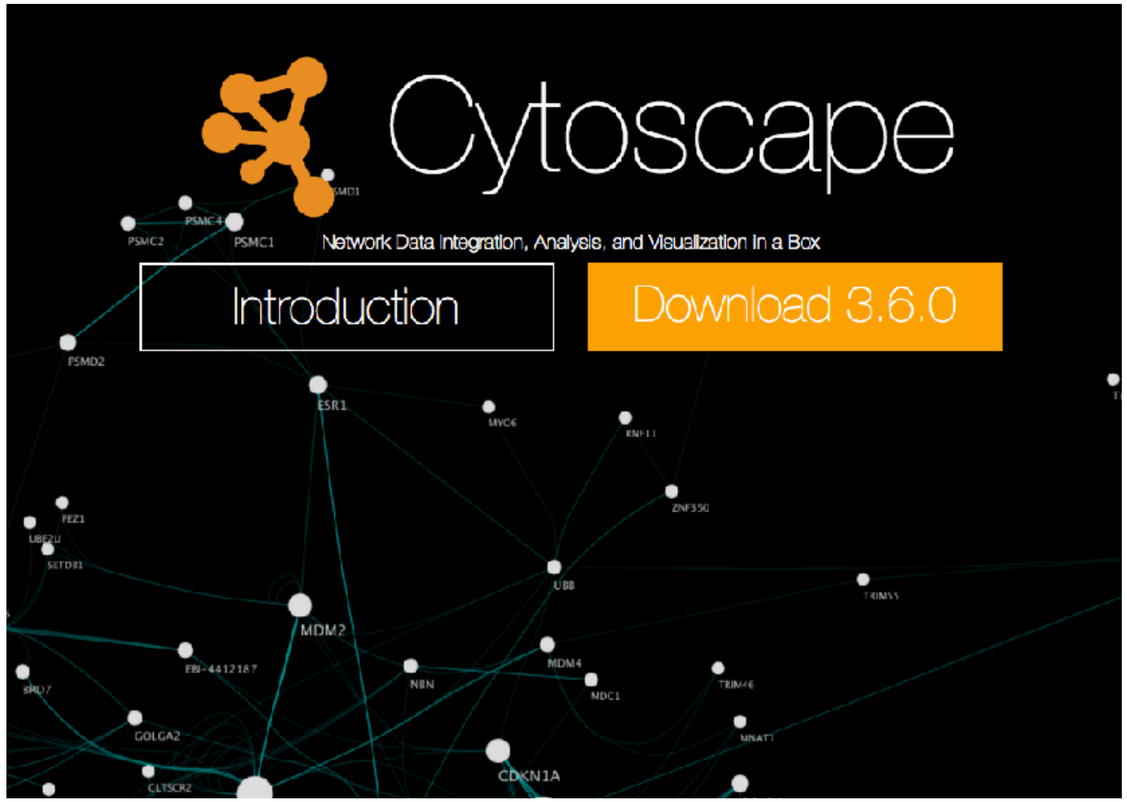


After

Tools for Graph Analysis

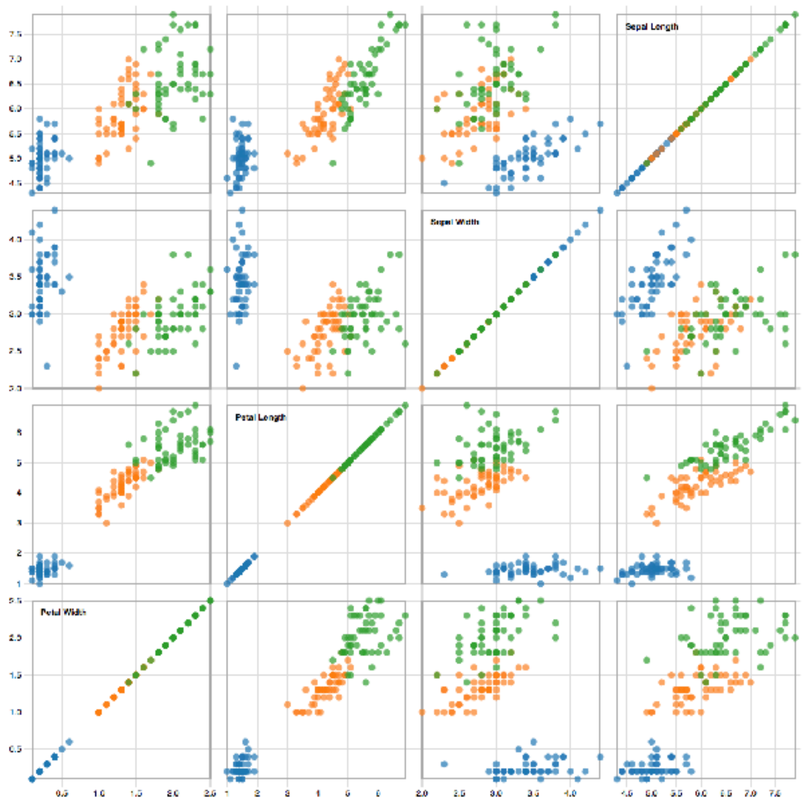


Gephi

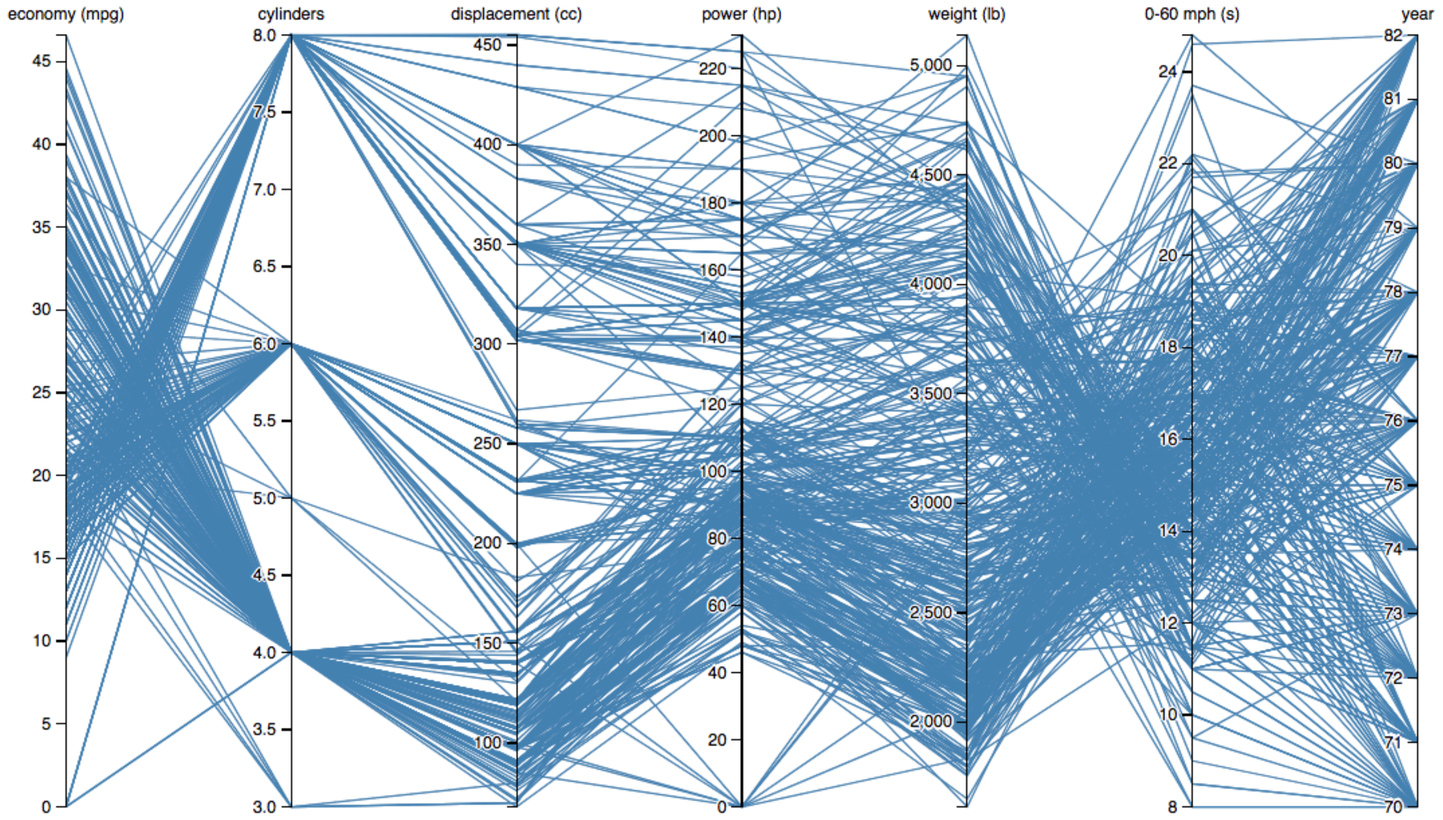


Cytoscape

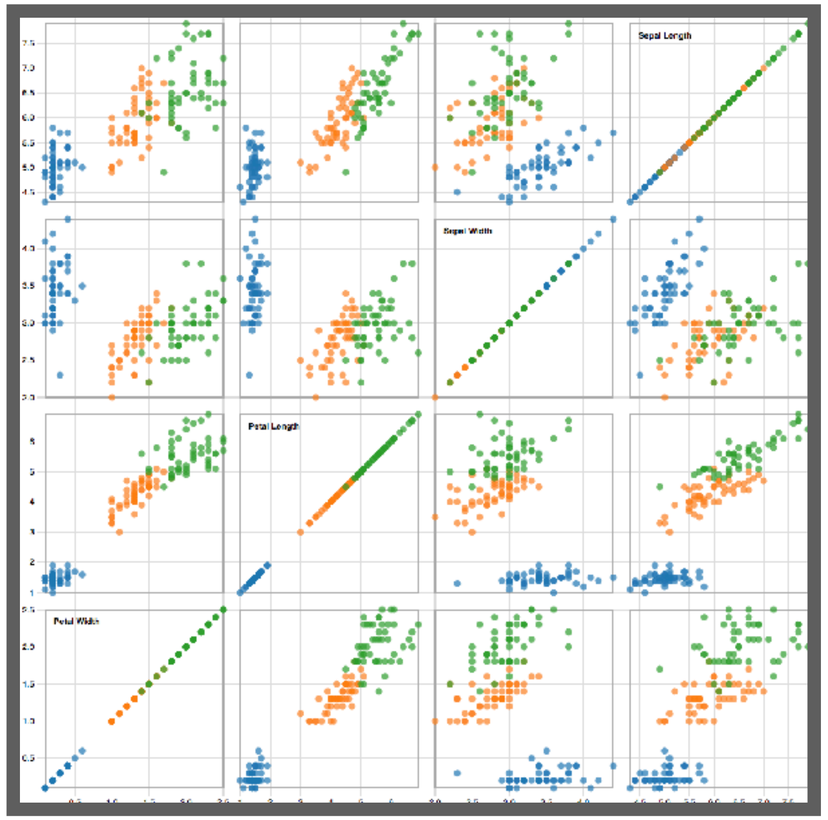
High-Dimensional Data

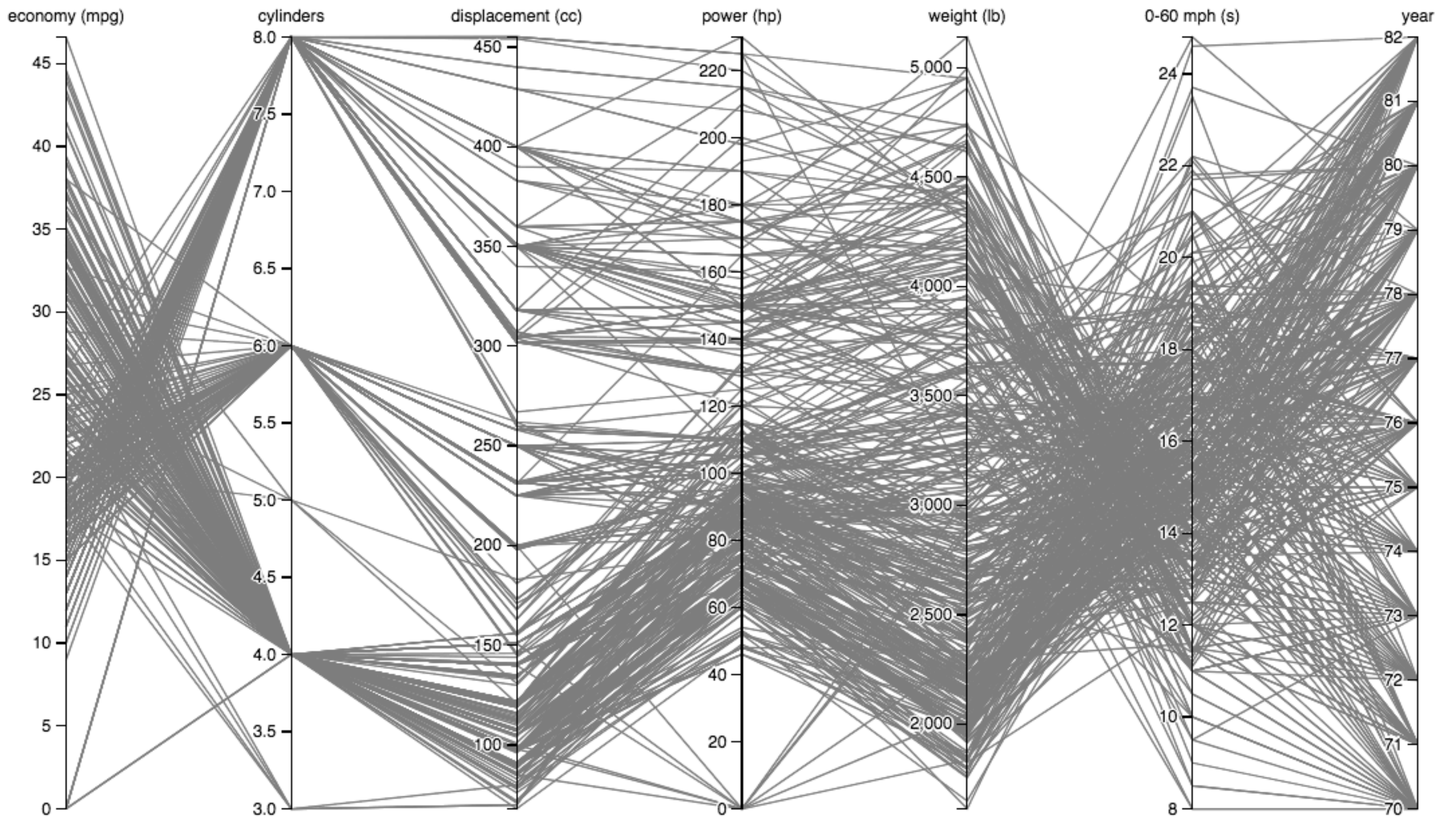


Scatterplot Matrix



Parallel Coordinates

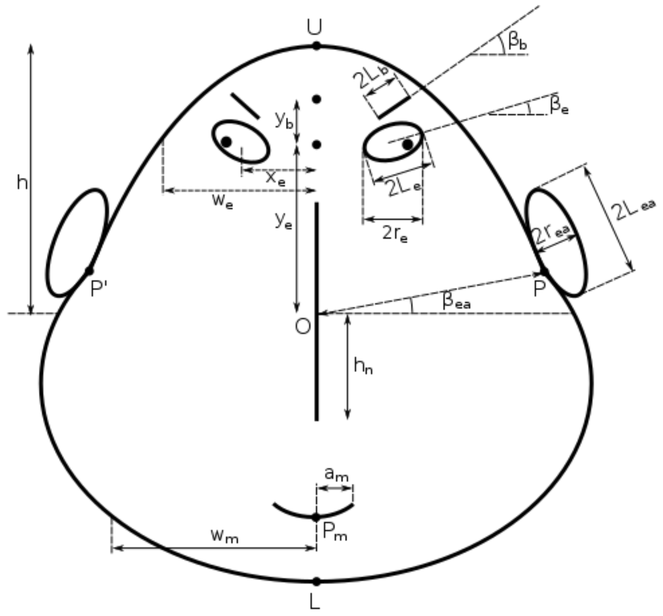


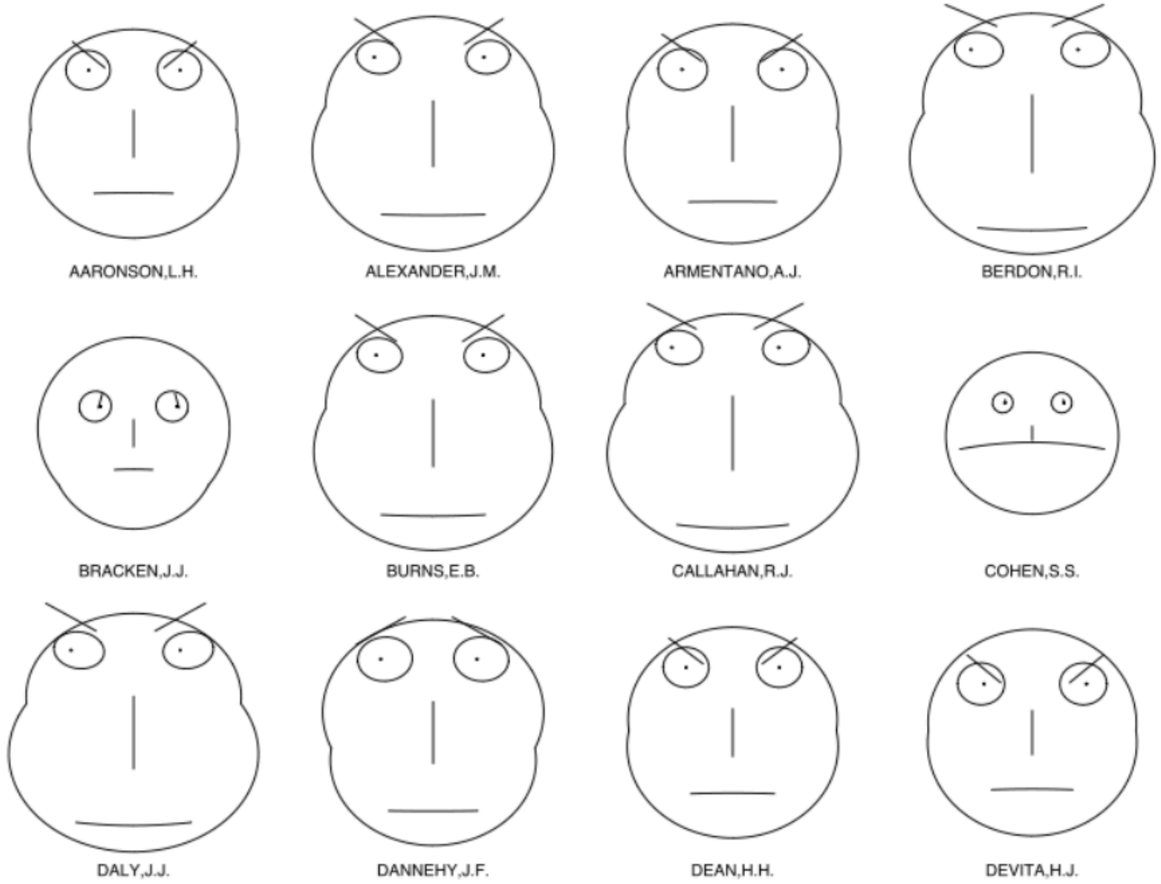


Scatterplot Matrix

Parallel Coordinates

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

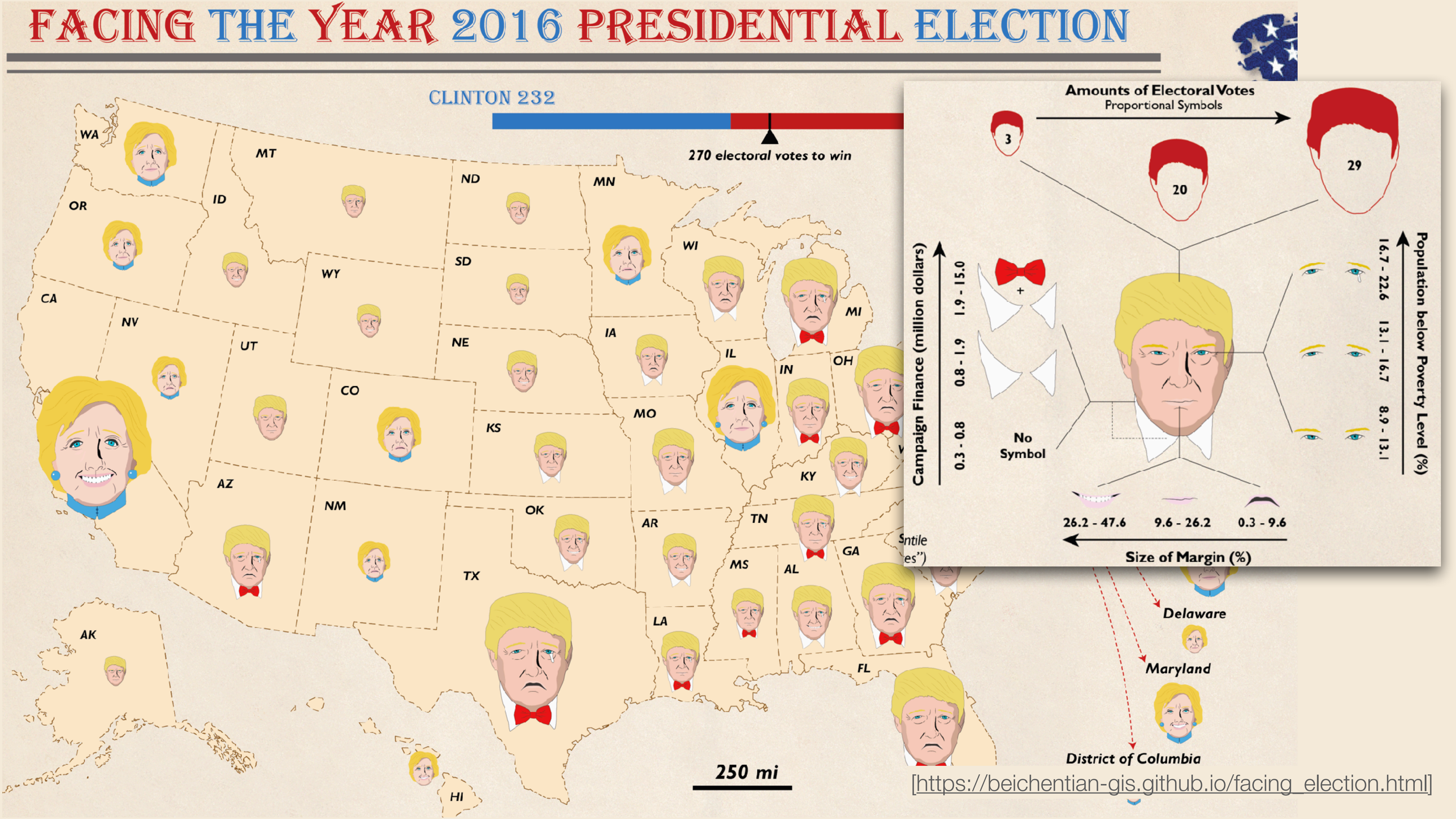




Lawyers' ratings of twelve judges

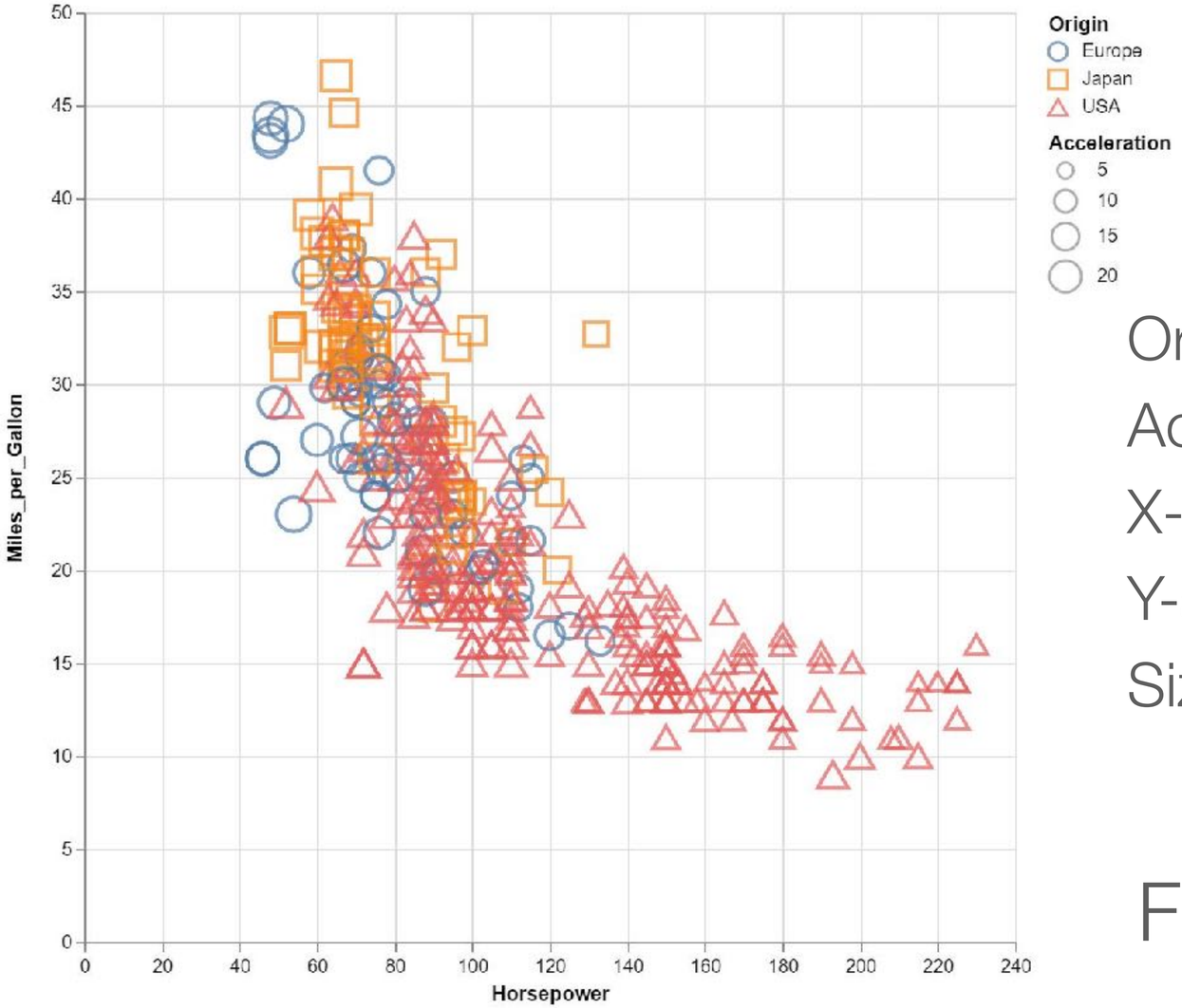
[https://en.wikipedia.org/wiki/Chernoff_face]

FACING THE YEAR 2016 PRESIDENTIAL ELECTION



[https://beichentian-gis.github.io/facing_election.html]

Don’t overload visual encodings



Origin: Shape

Acceleration: Size

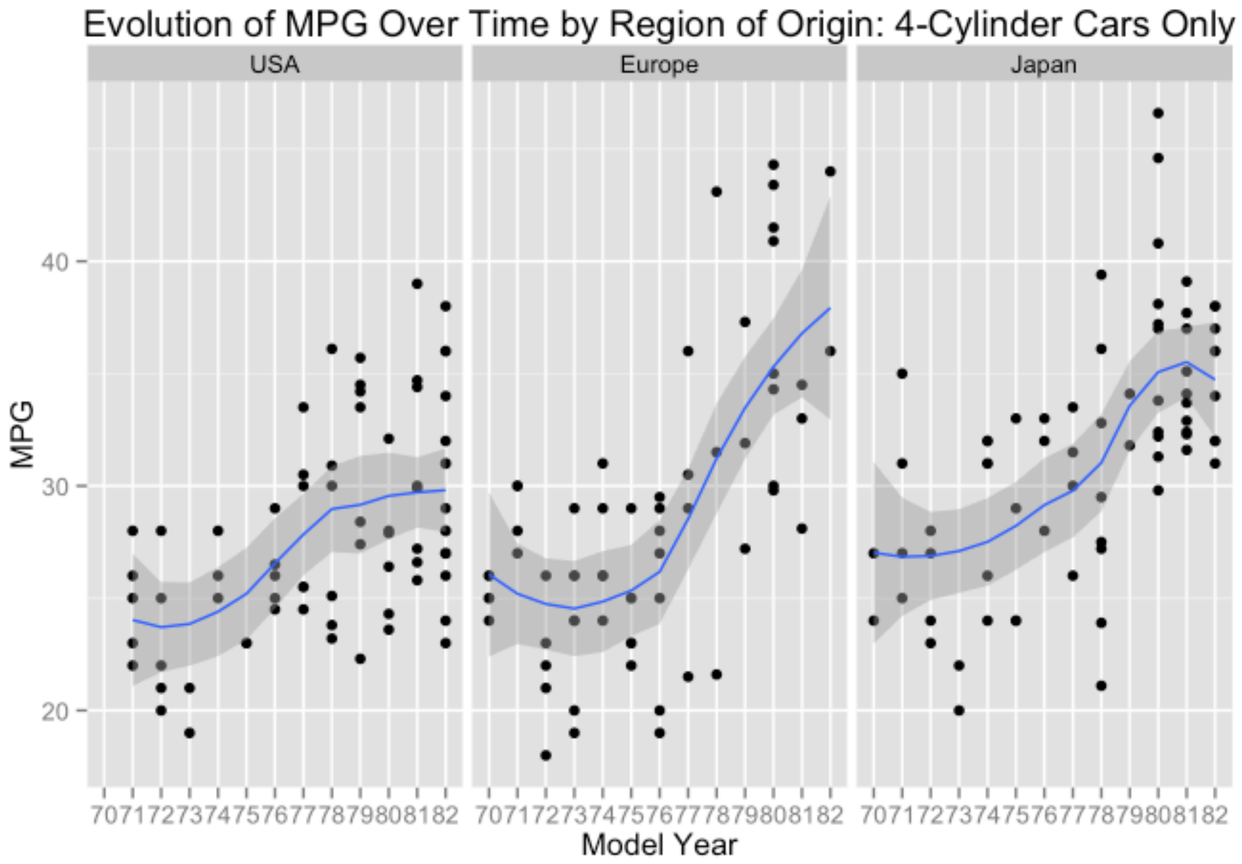
X-axis: Horsepower

Y-axis: Miles per Gallon

Size: Acceleration

Five visual variables in one chart!

Use facets or small multiples



[David K. Smith]

[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.

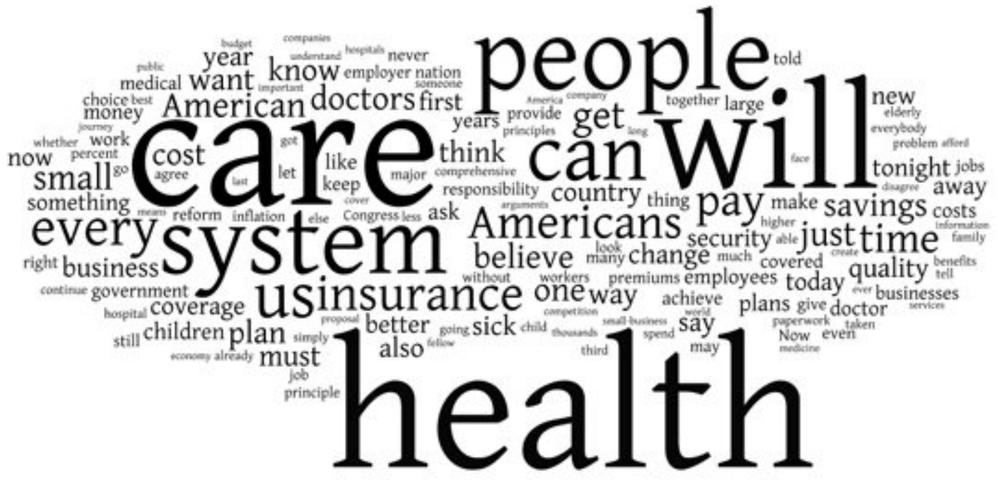
Madame Speaker, Vice President Biden, Members of Congress, and the American people:

When I spoke here last winter, this nation was facing the worst economic crisis since the Great Depression. We wer 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)

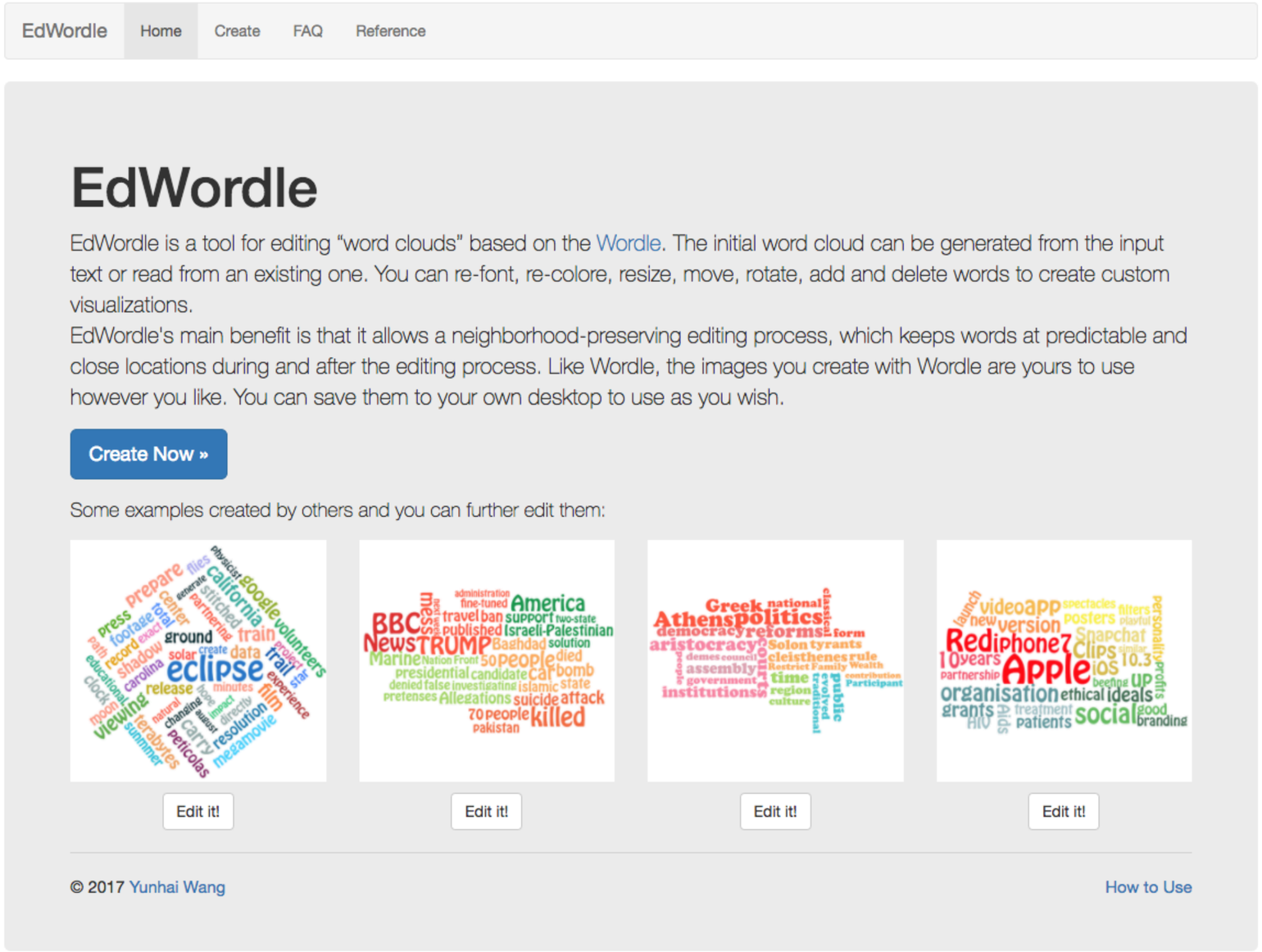
May not facilitate comparison (unstable layout)

Term frequency may not be meaningful

Does not show the structure of the text

EdWordle [Y. Wang 2017]

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

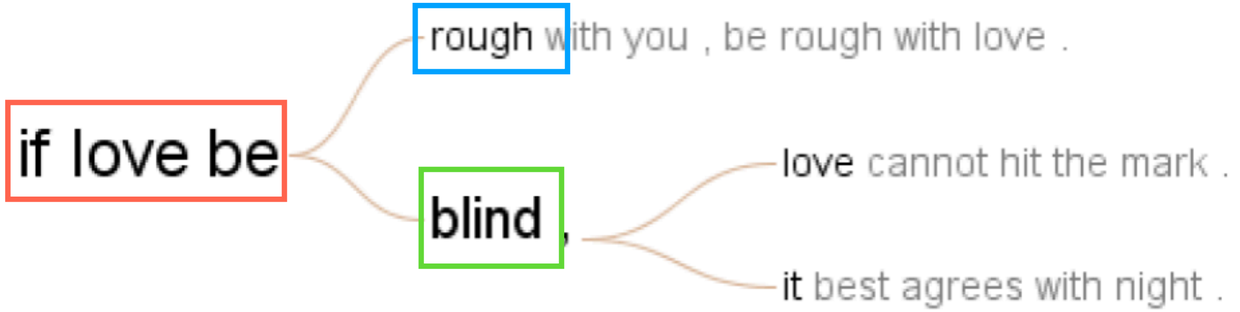


Word Tree: Word Sequences [Wattenberg et al.]

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

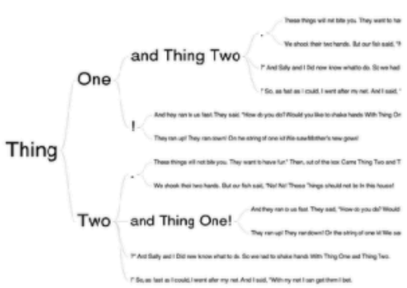
if love be blind , love cannot hit the mark .

if love be blind , it best agrees with night .



word tree

Obama's Inauguration Speech



The Cat in the Hat

Alice in Wonderland

Bob Dylan's Blowin' in the Wind

Obama War Speech

Steve Jobs Stanford Commencement Speech

@jasondavies

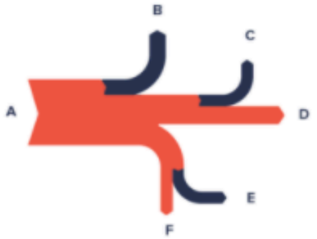
Flickr Comments

[https://www.jasondavies.com/wordtree/]

Paste Text

Others

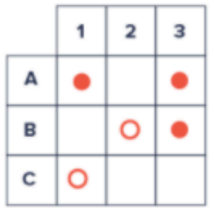
Sankey Diagram



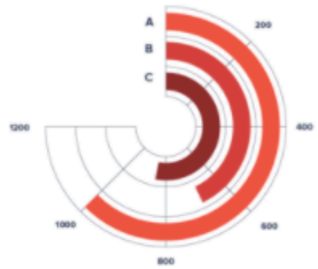
Alluvial Diagram

Donut Chart

Matrix Diagram



Radial Bar Chart

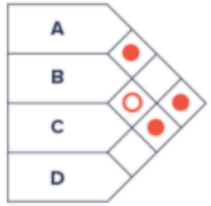


http://datavizproject.com/

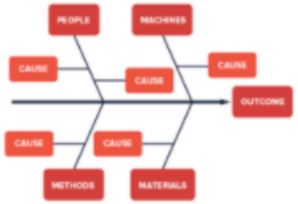
Radial Histogram

Sorted Stream Graph

Matrix Diagram (Roof Shaped)



Fishbone Diagram



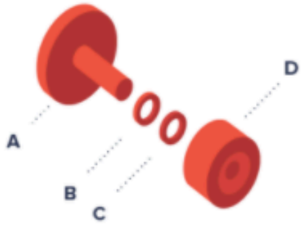
Pictorial fraction chart

Isoline Map

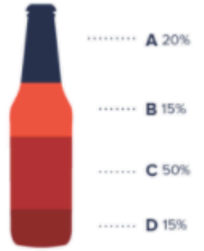
Flow Map

Arc Diagram

Exploded View Drawing



Pictorial Stacked Chart



THE GRAPHIC CONTINUUM

[Jonathan Schwabish 2014]

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 blend 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.

DISTRIBUTION

Graphical representations of the distribution of data

TIME

Track changes over time

COMPARING CATEGORIES

Compare values across categories

GEOSPATIAL

Relate data to its geography

PART-TO-WHOLE

Visualizations that relate the part of a variable to its total

RELATIONSHIP

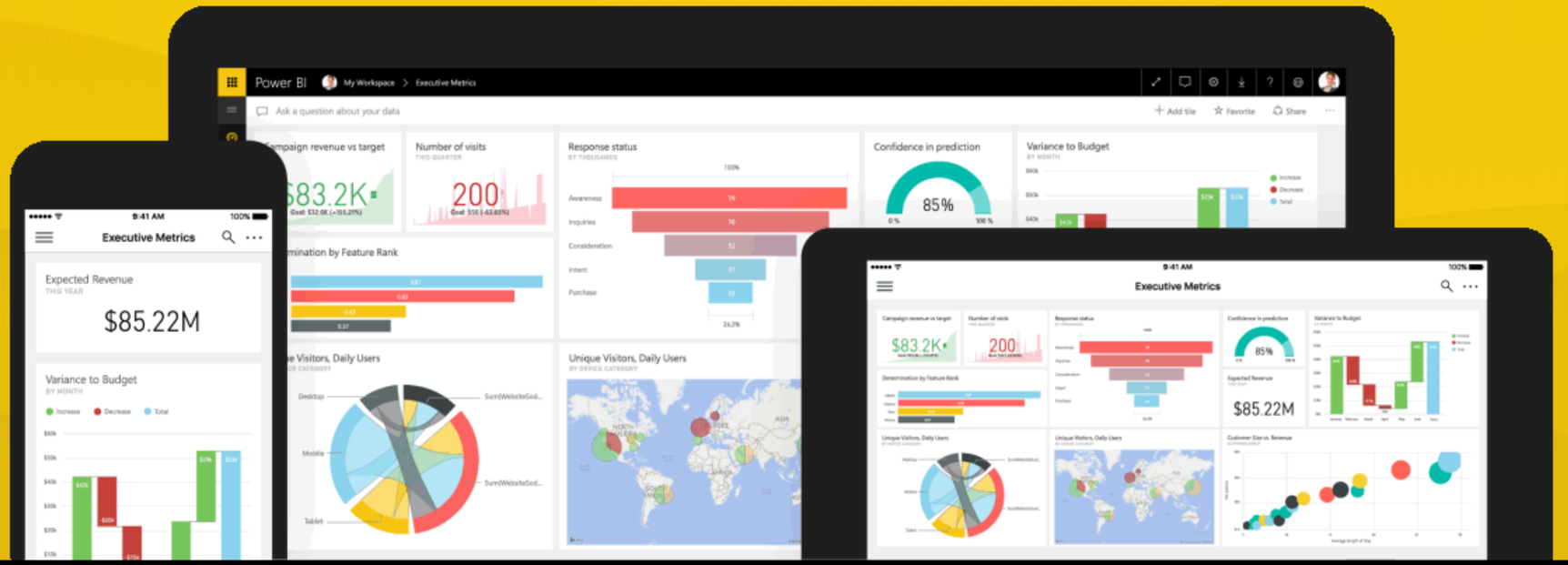
Illustrates correlations or relationships between variables

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 >



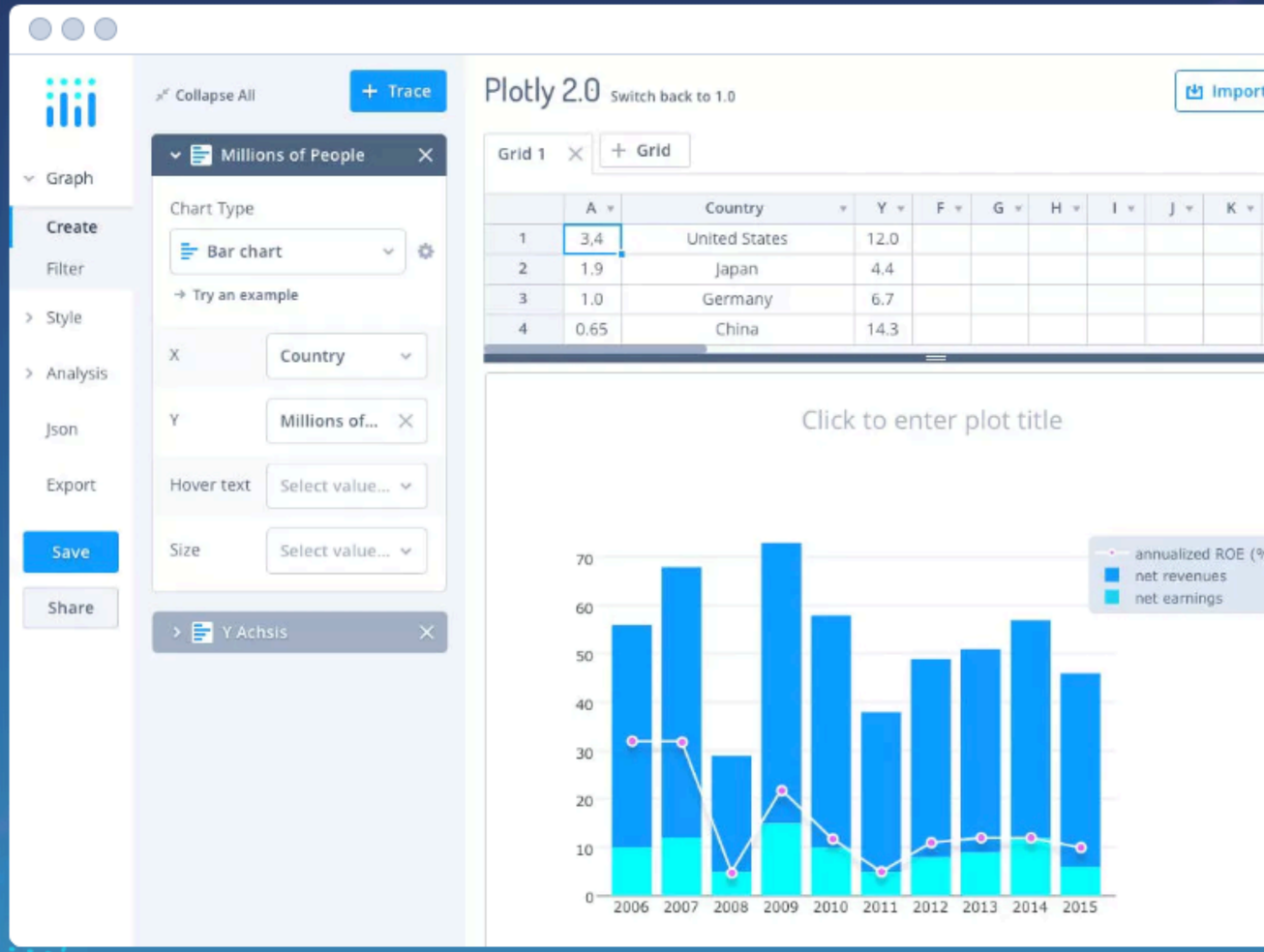
New version of Power BI Report Server is now available. Publish Power BI reports on-premises connecting to any data. Learn more

Chart Studio

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

No coding required.

GET STARTED



redhat. alteryx ComEd.

RAWGraphs

The missing link between spreadsheets and data visualization.

USE IT NOW!

FORK IT ON GITHUB

Turn your data into knowledge

Quadrigram allows you to engage people by sharing stories that matter.

Take a look at the new account for professionals.

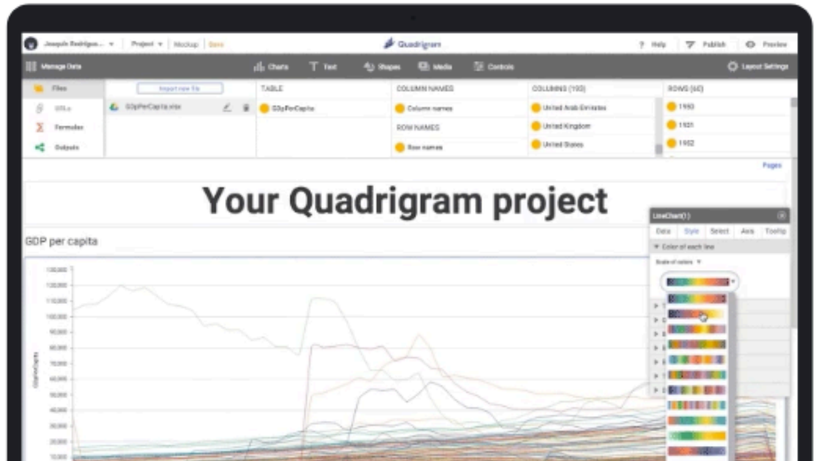
PURCHASE NOW

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.

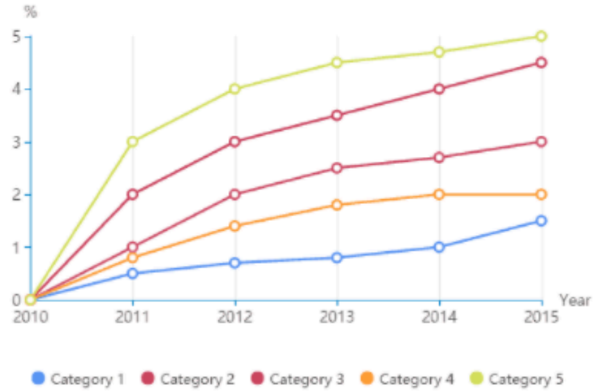
SIGN UP or SIGN IN with your Google Account

All (74)

Show PREMIUM

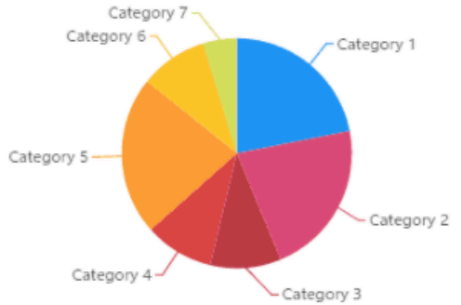
Simple map of Europe

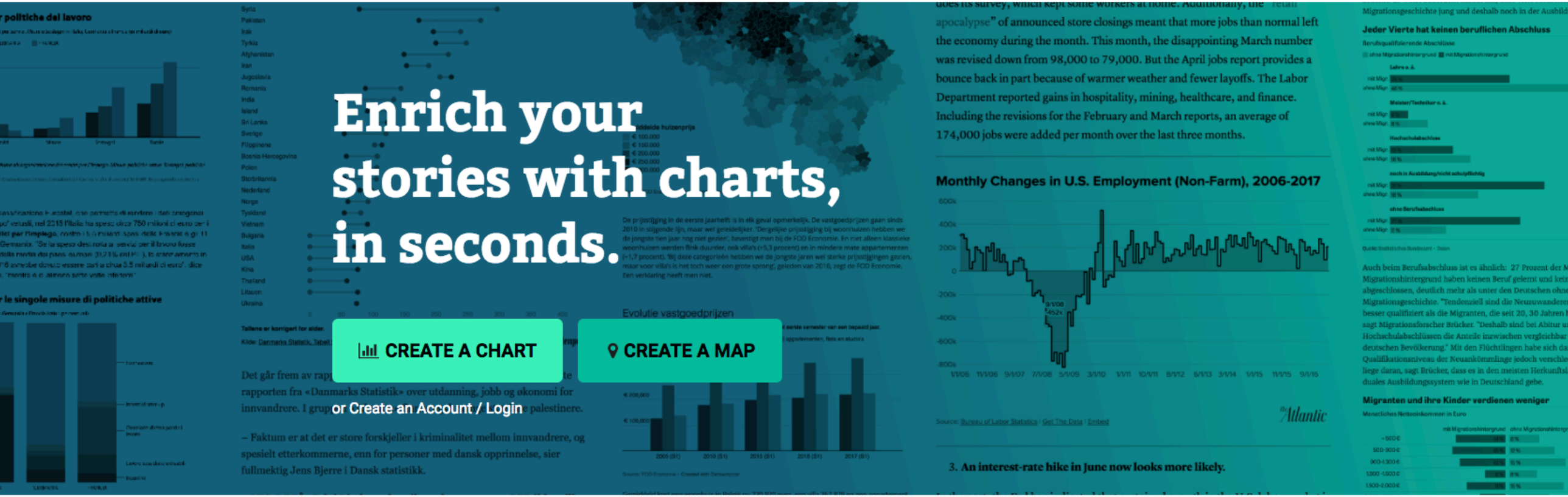
Multiple lines chart



Multiple lines chart FREE

Stacked column chart with pie chart



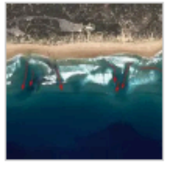


Datawrapper makes it easy to create beautiful charts.

See for yourself how different newsrooms use our charts:

Recent SEE ALL →

Spotting rip currents



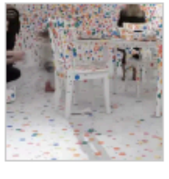
Rip currents are like hidden rivers near the shore that ...

FEED SPONSOR

Be the data whisperer →

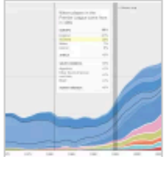
Notre Dame's Mendoza College of Business in Chicago offers a one-year, part-time MS in Business Analytics for ambitious working professionals

Obliteration Room invited people to put dotted stickers everywhere

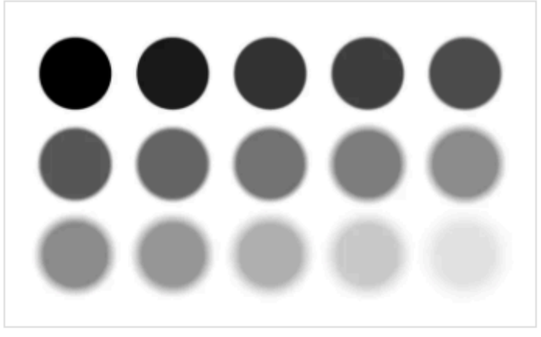


The Obliteration Room (2012) by artist Yayoi Kusama started as ...

Where athletes in professional sports come from

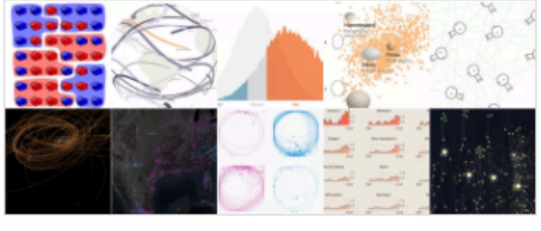


Sports are growing more



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

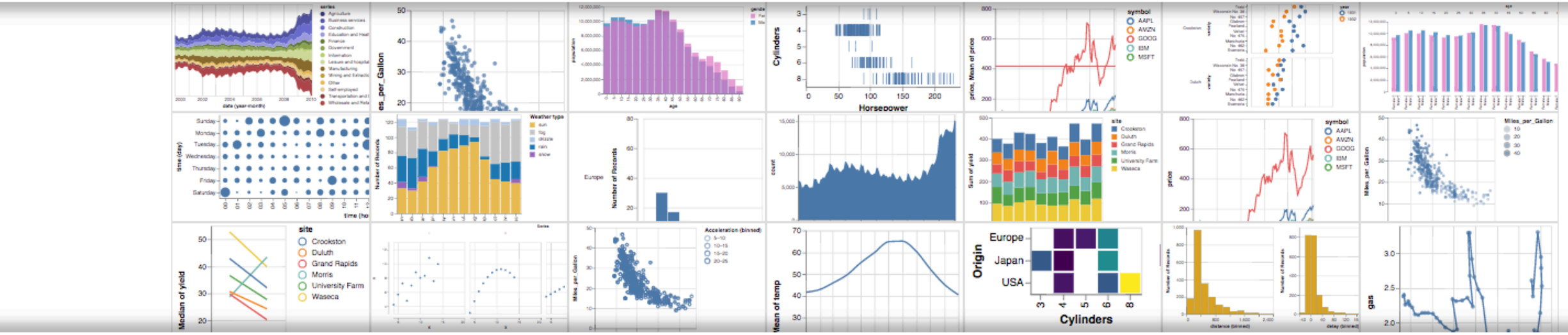
Networks
Connecting data

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!