Beyond Basics- Advanced Topics in Visualization

Nam Wook Kim

Mini-Courses — January @ GSAS
2018

Topics

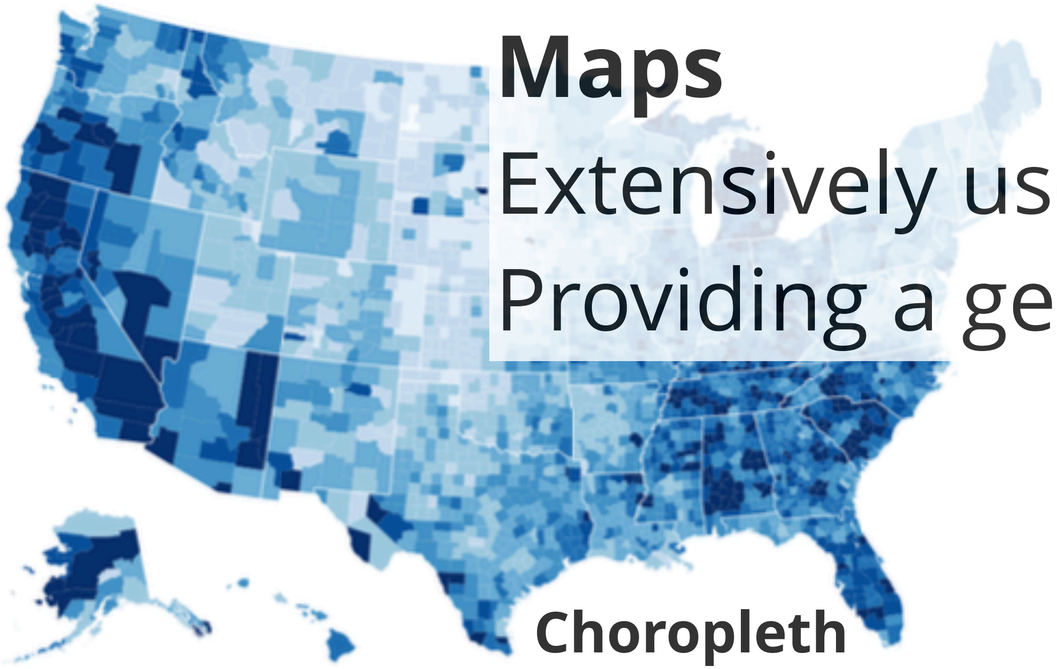
- Maps
- Networks & Trees
- High-Dimensional Data
- Text

Maps

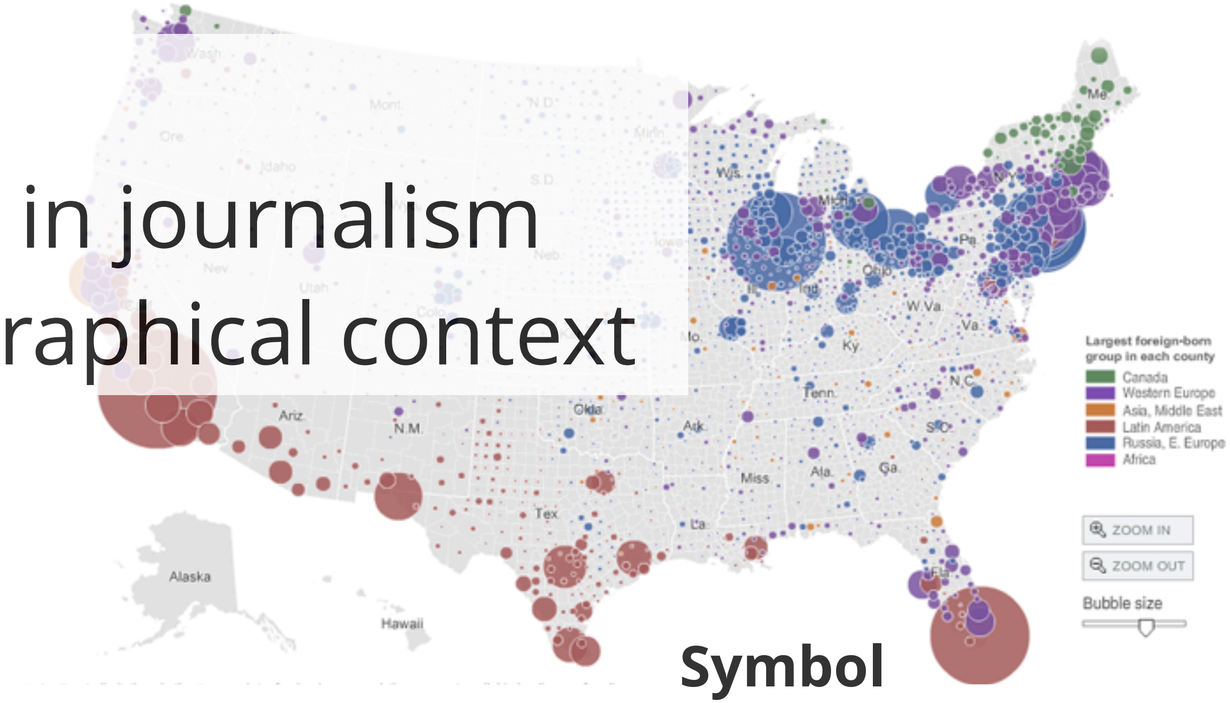
Maps

Extensively used in journalism

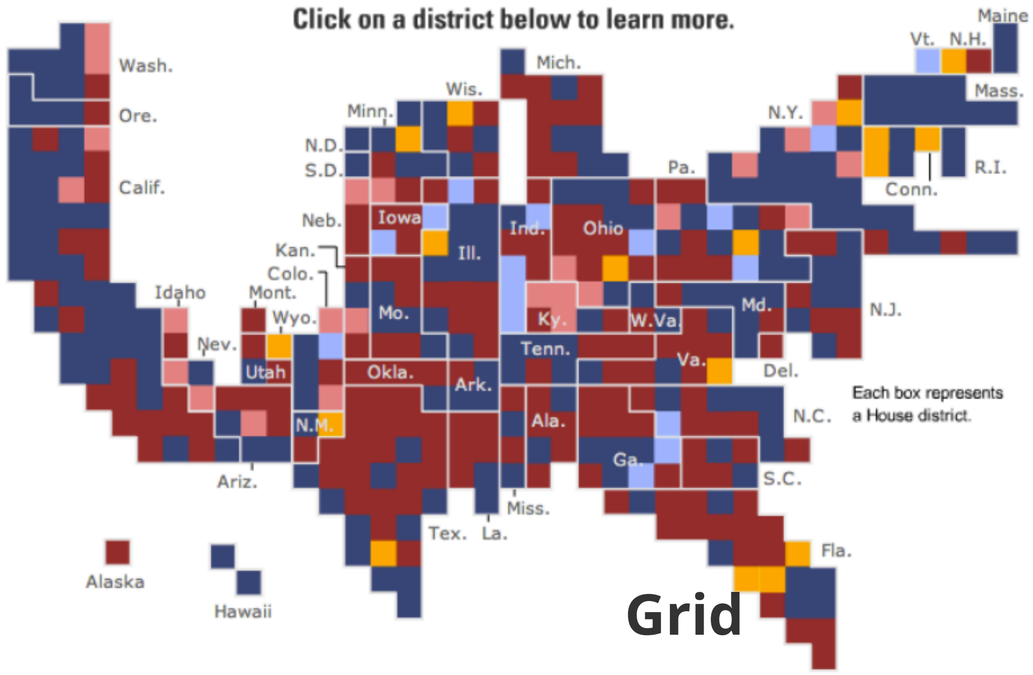
Providing a geographical context



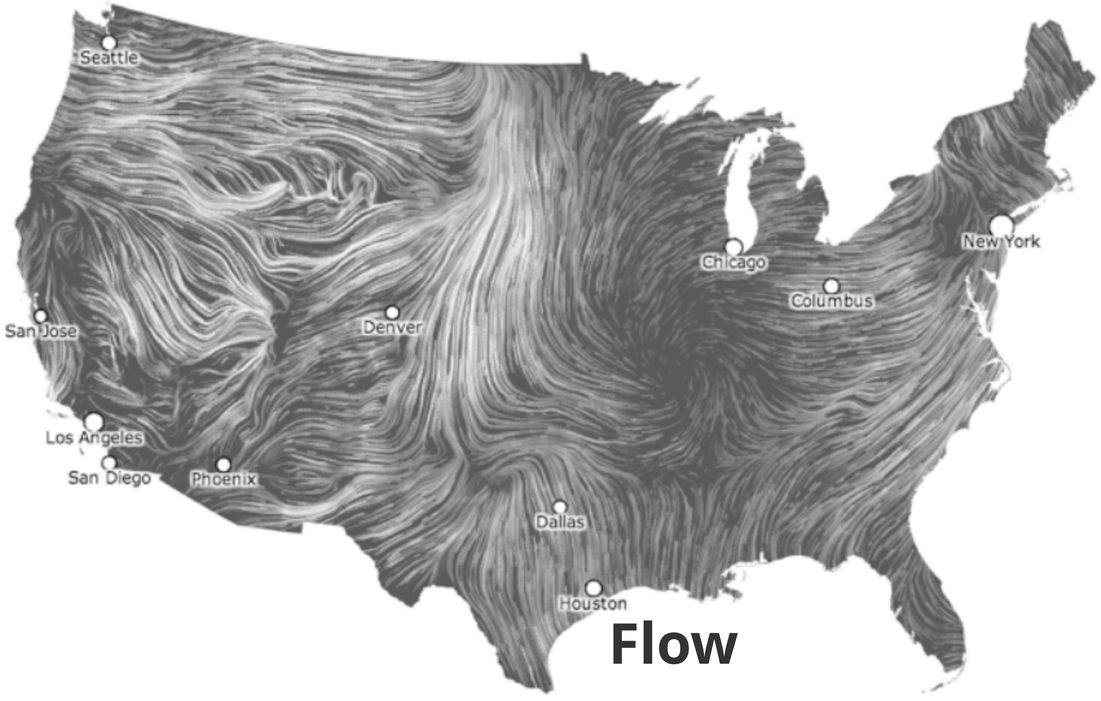
Choropleth



Symbol



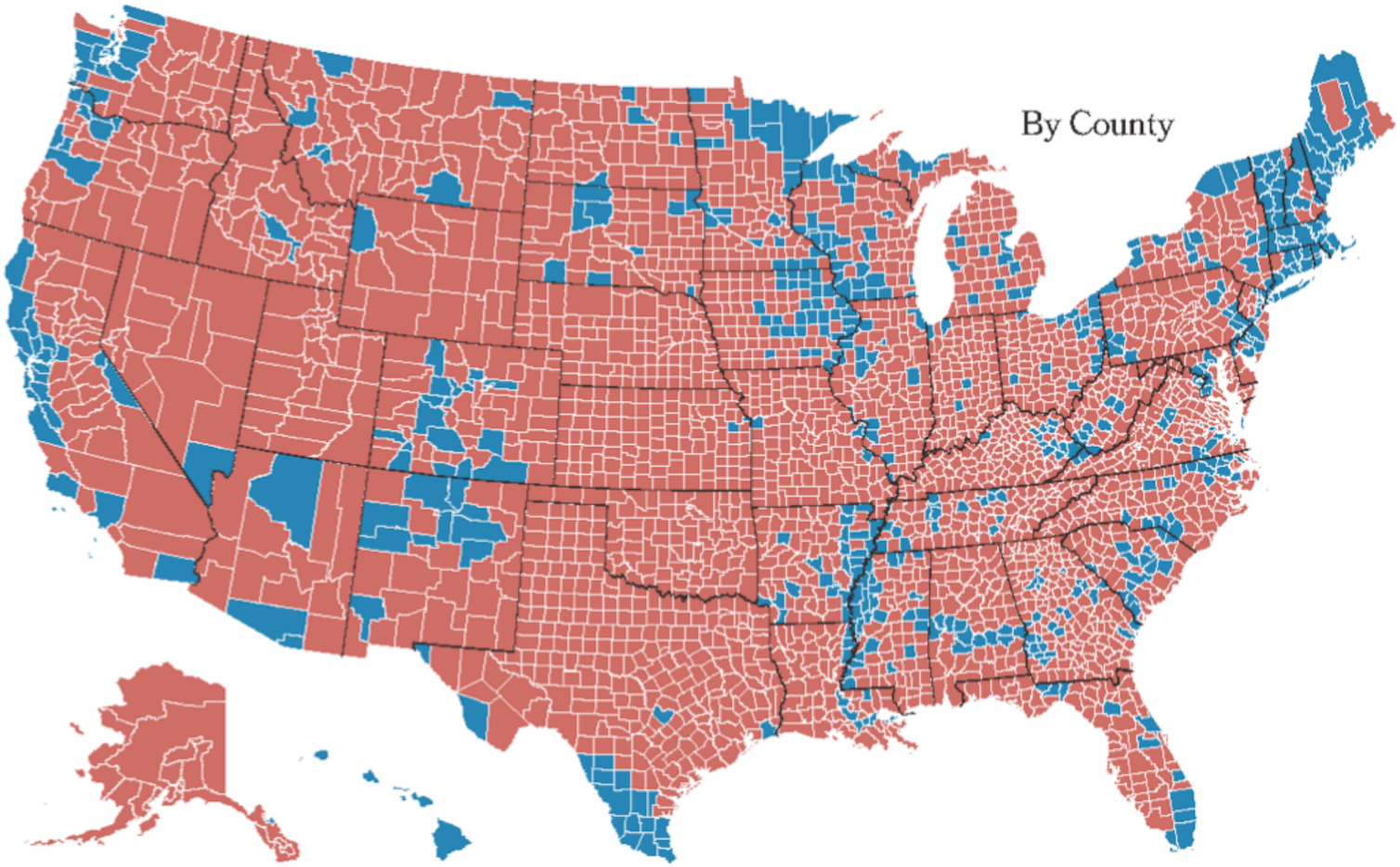
Grid



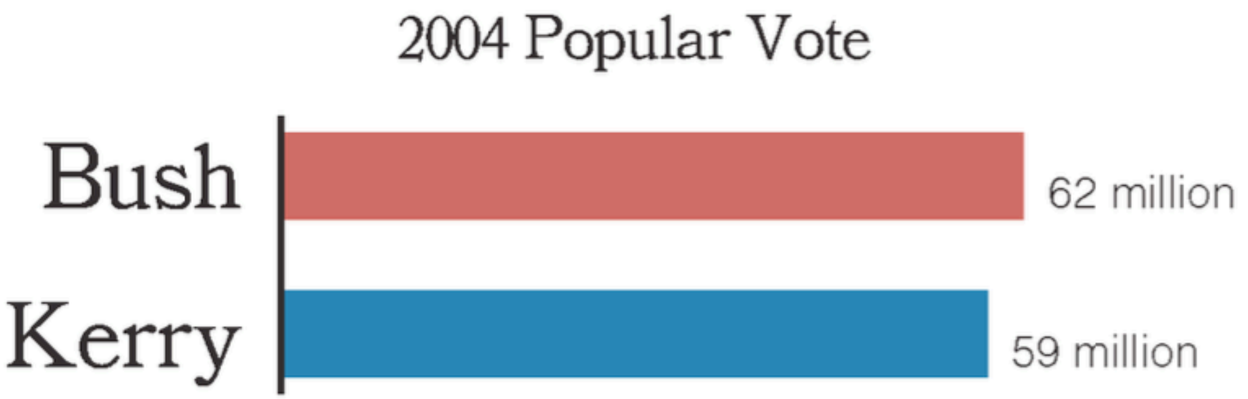
Flow

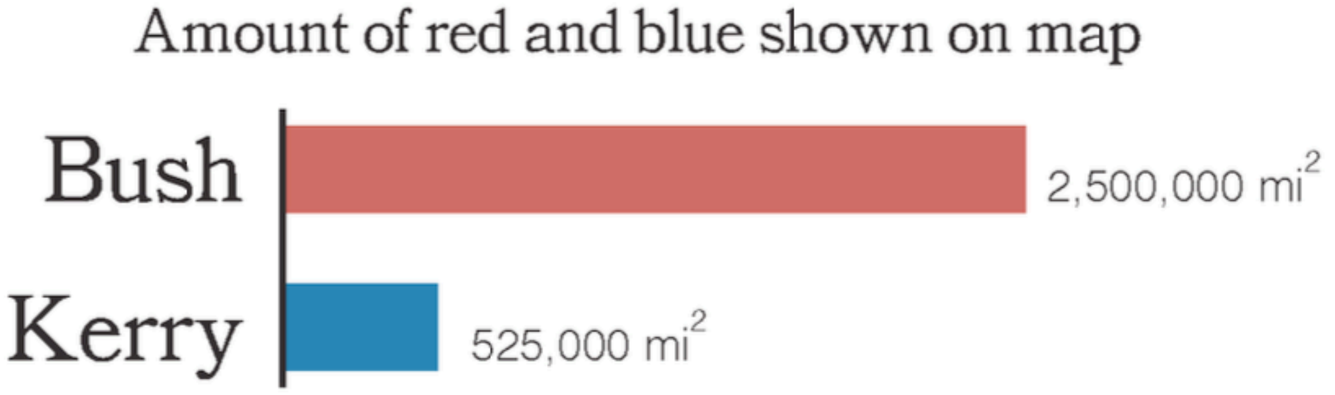
Problem with Choropleth

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

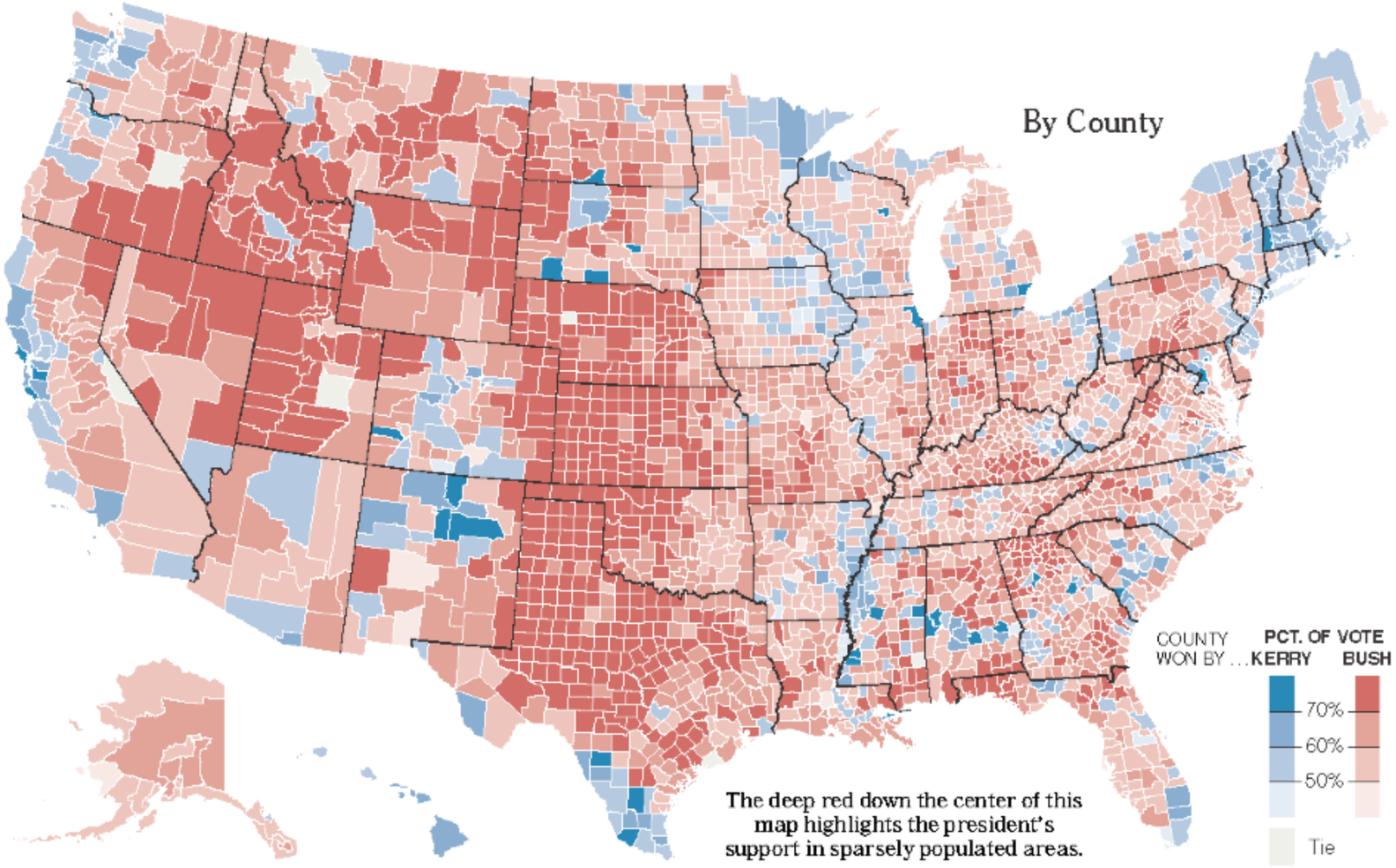


[Matthew Ericson, NY Times]

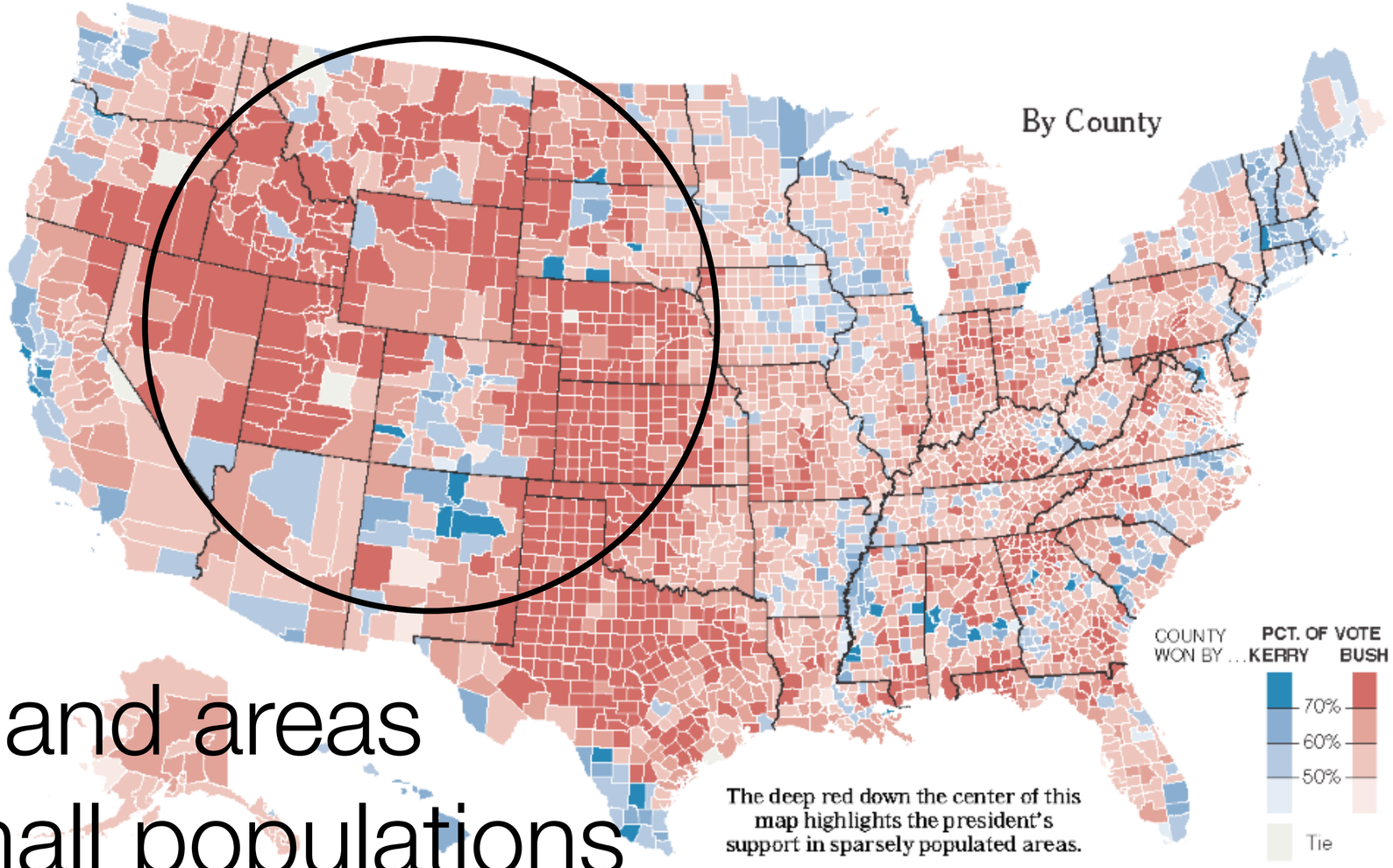

2004 Popular Vote
Bush
62 million
Kerry
59 million


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.

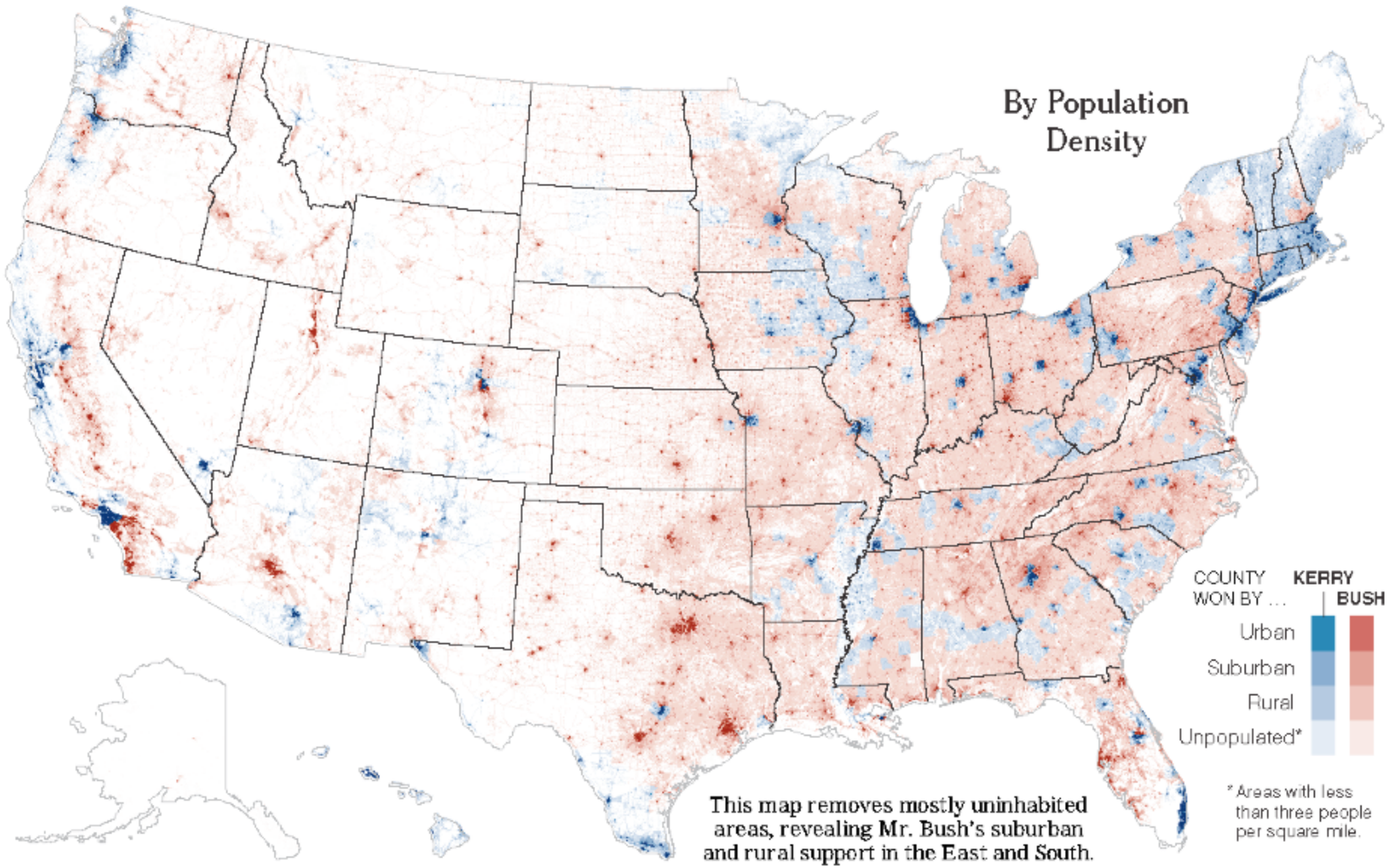


Take into account for the number of votes.

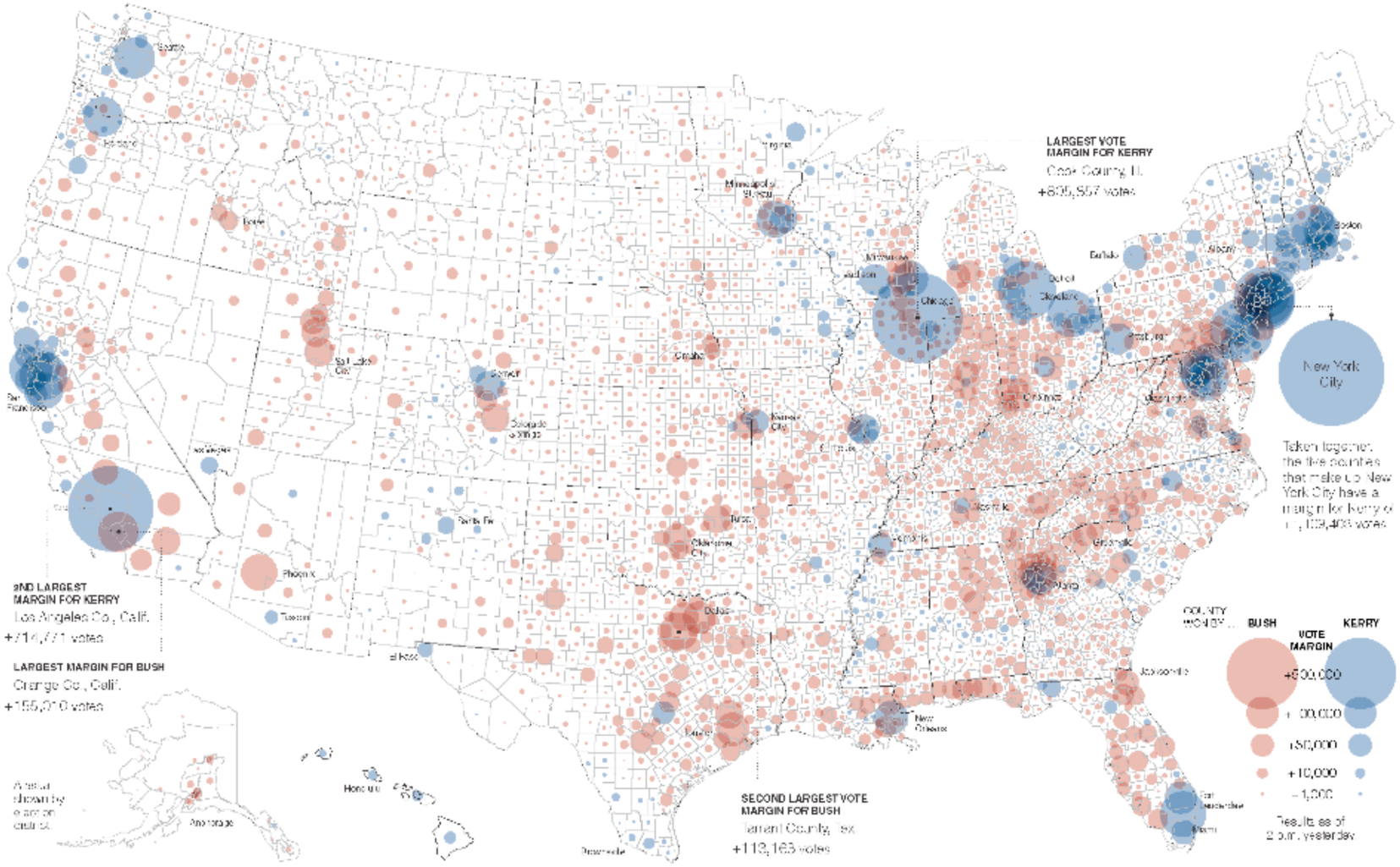


Larger land areas
with small populations

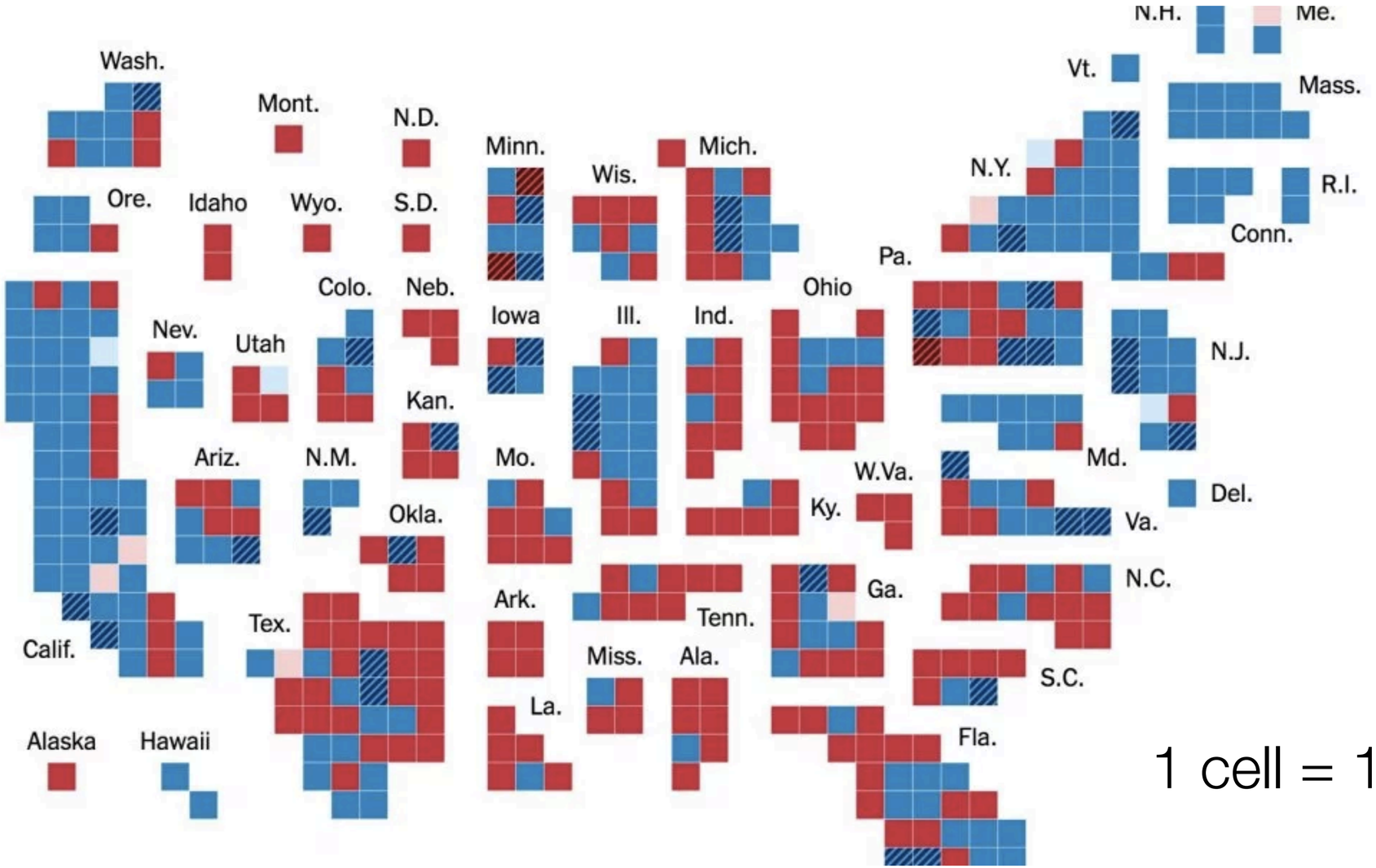
Take into account for population density



Symbol Map


LARGEST VOTE MARGIN FOR KERRY
Cook County, Ill.
+635,557 votes
New York City
Taken together, the five counties that make up New York City have a margin for Kerry of 1,038,493 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

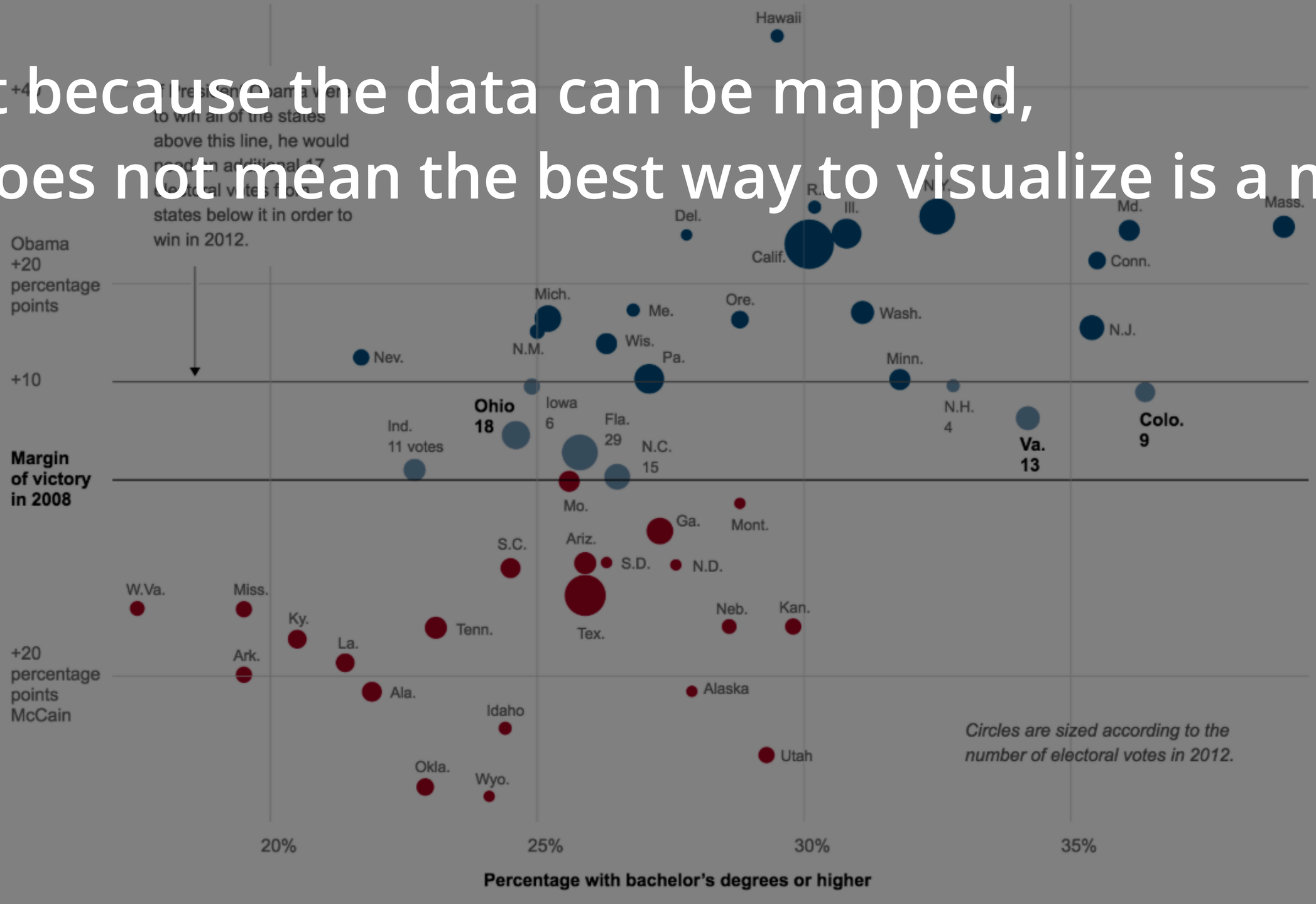
Cartogram: Grid Map



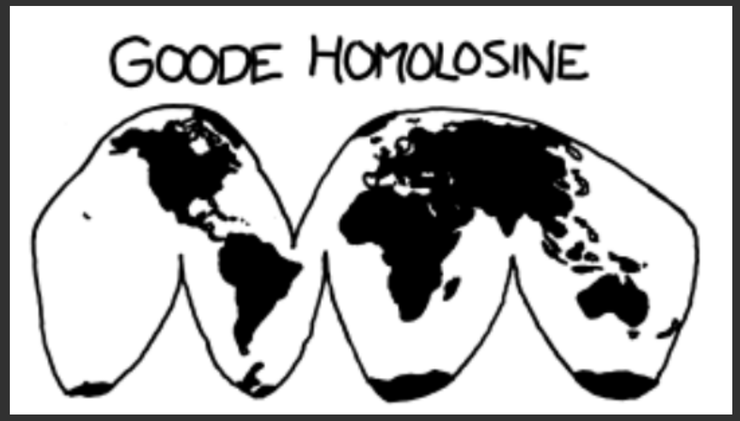
1 cell = 1 seat

https://www.nytimes.com/interactive/2018/11/06/us/elections/results-house-elections.html

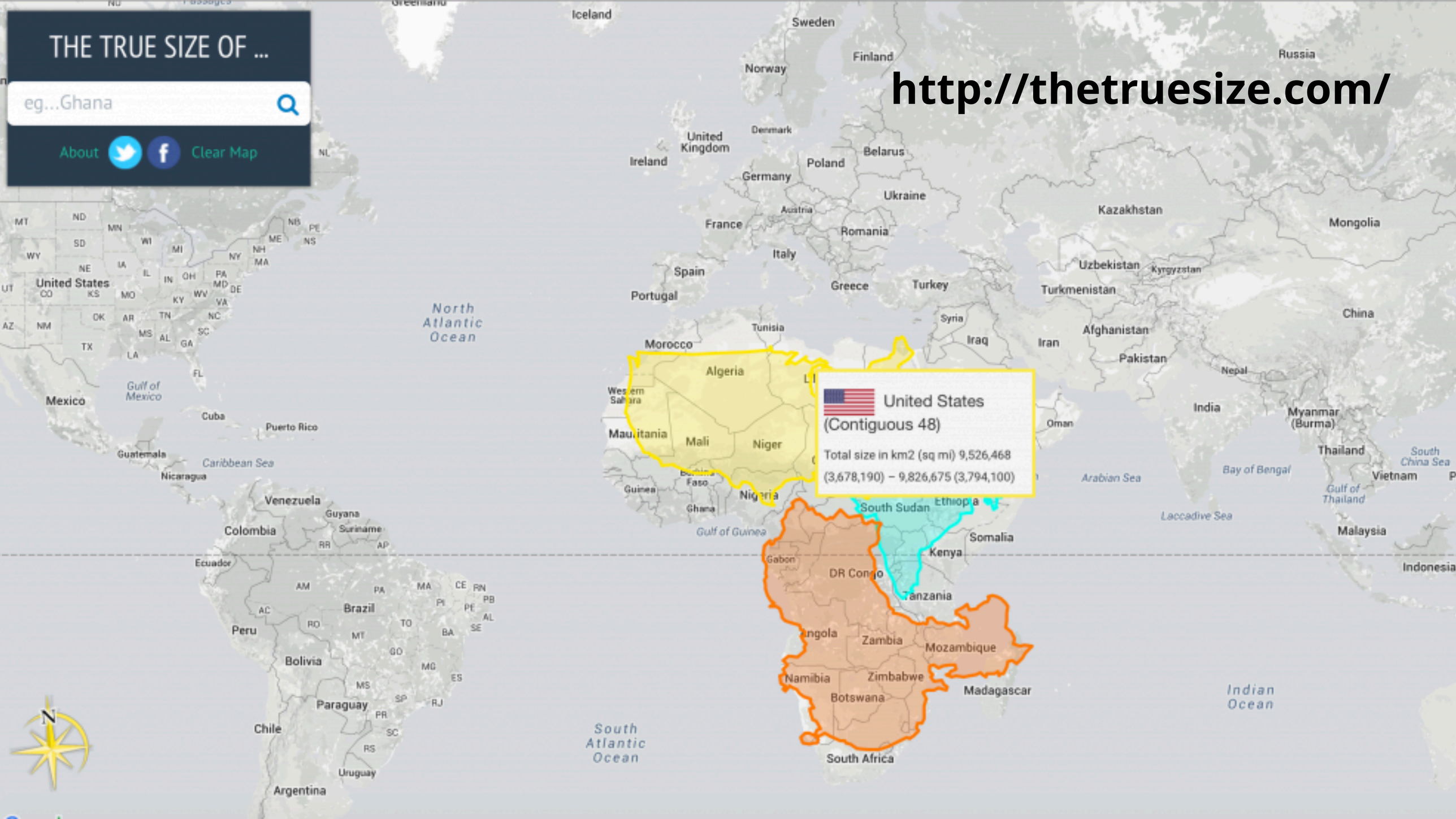
Just because the data can be mapped, it does not mean the best way to visualize is a map

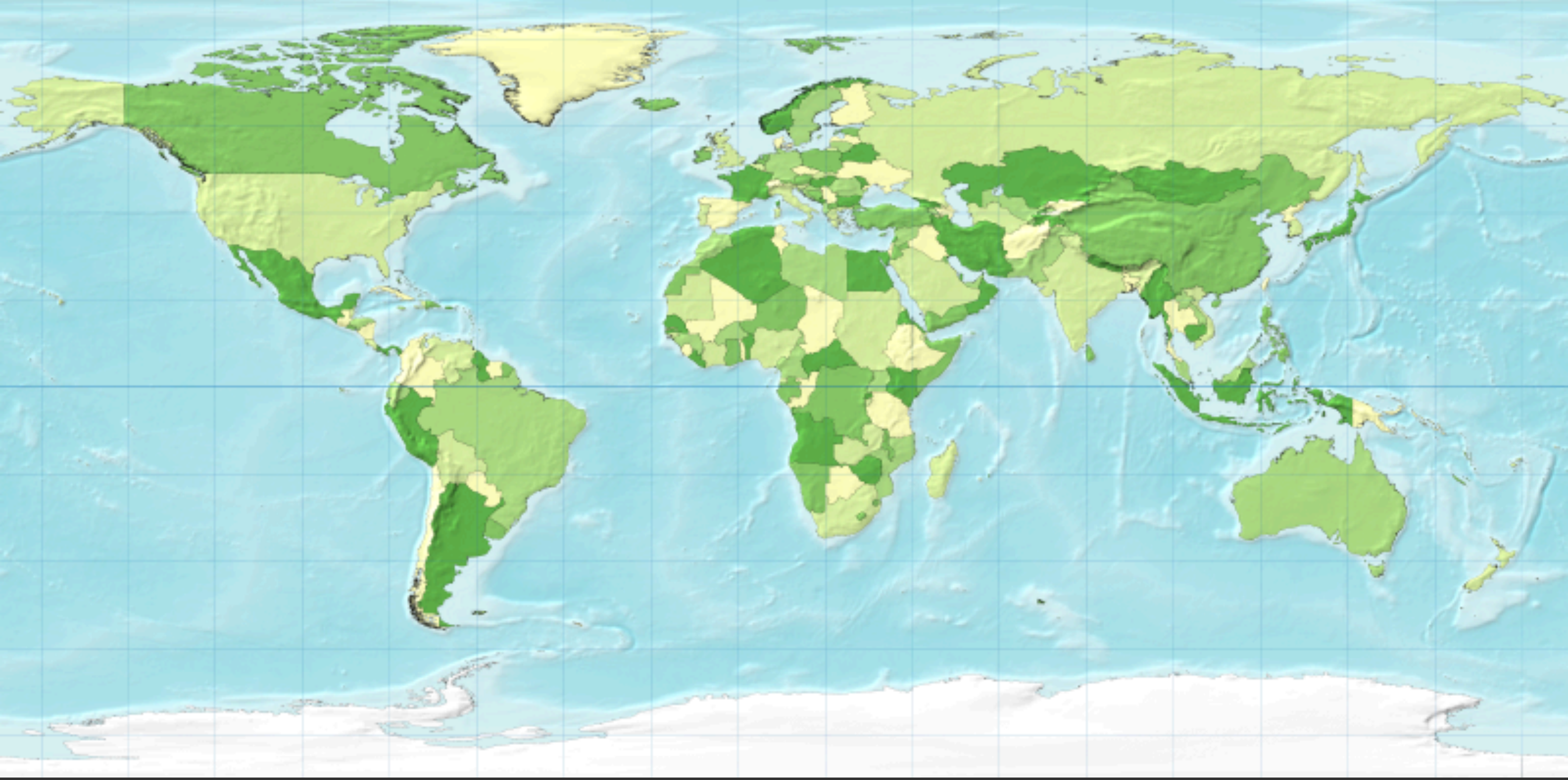


Map Projections



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


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

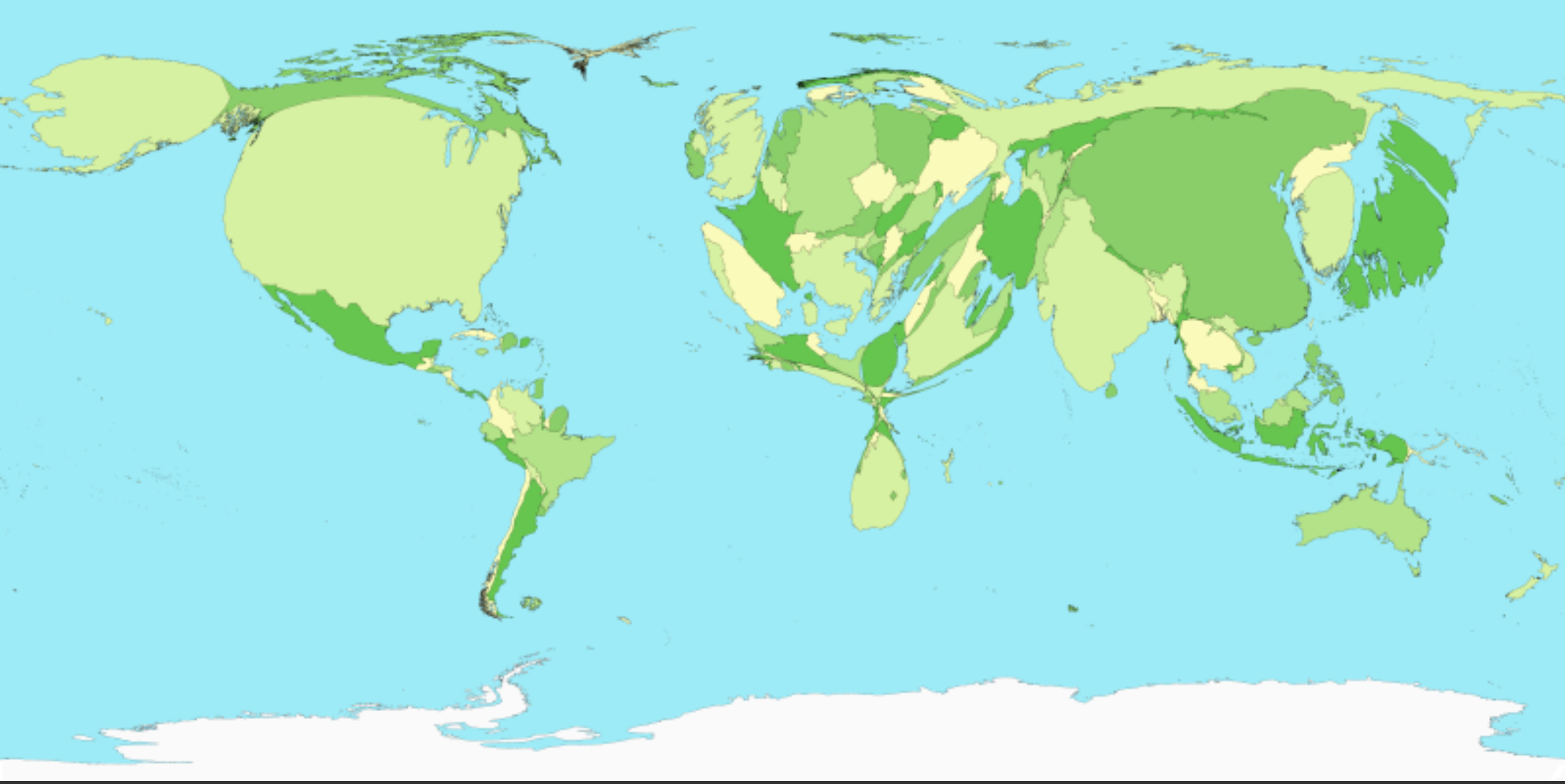


Ordinary map

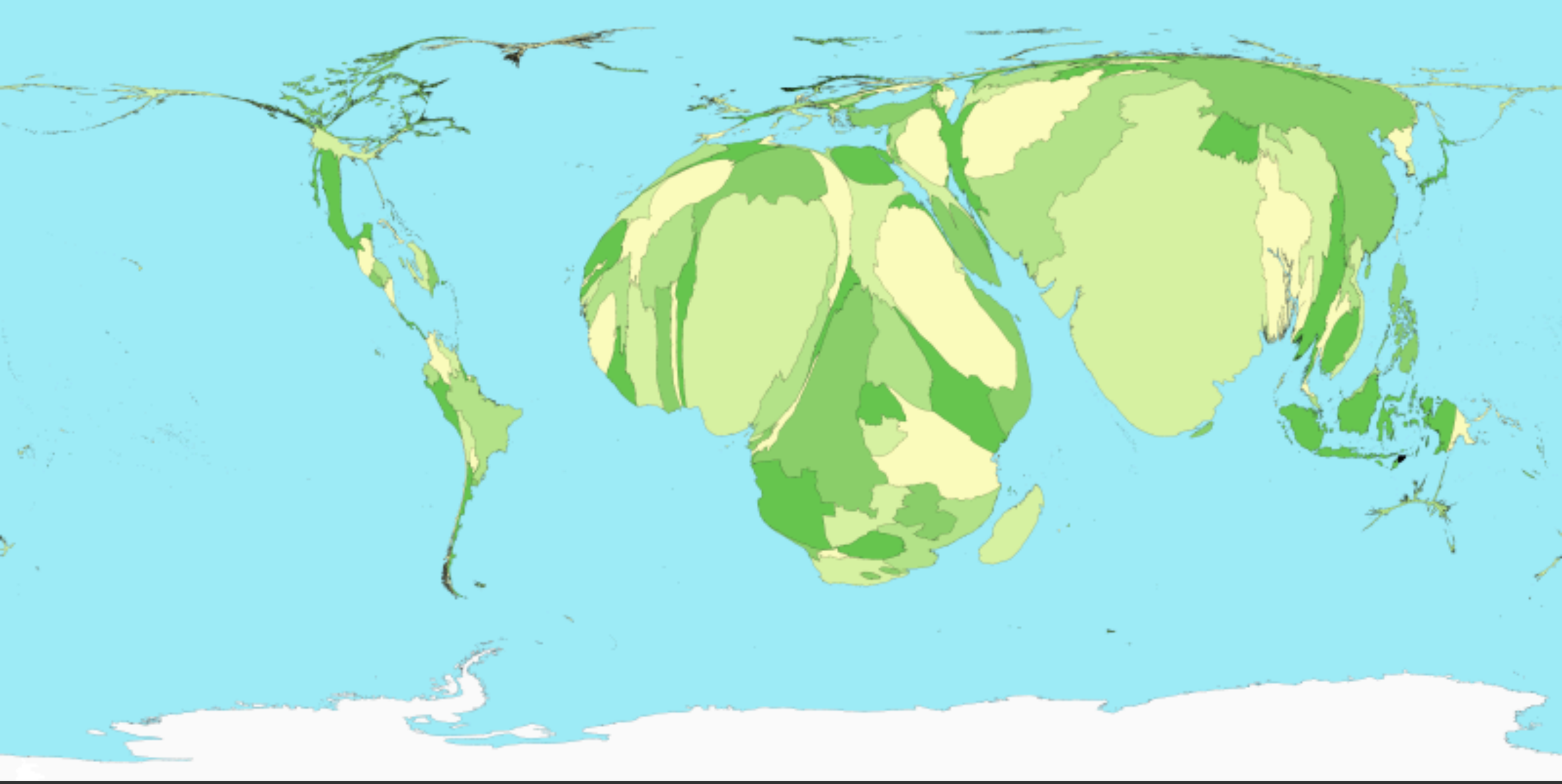
[http://www-personal.umich.edu/~mejn/cartograms/]


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



Greenhouse gas emissions



Child mortality

Problem with Rainbow Colormap

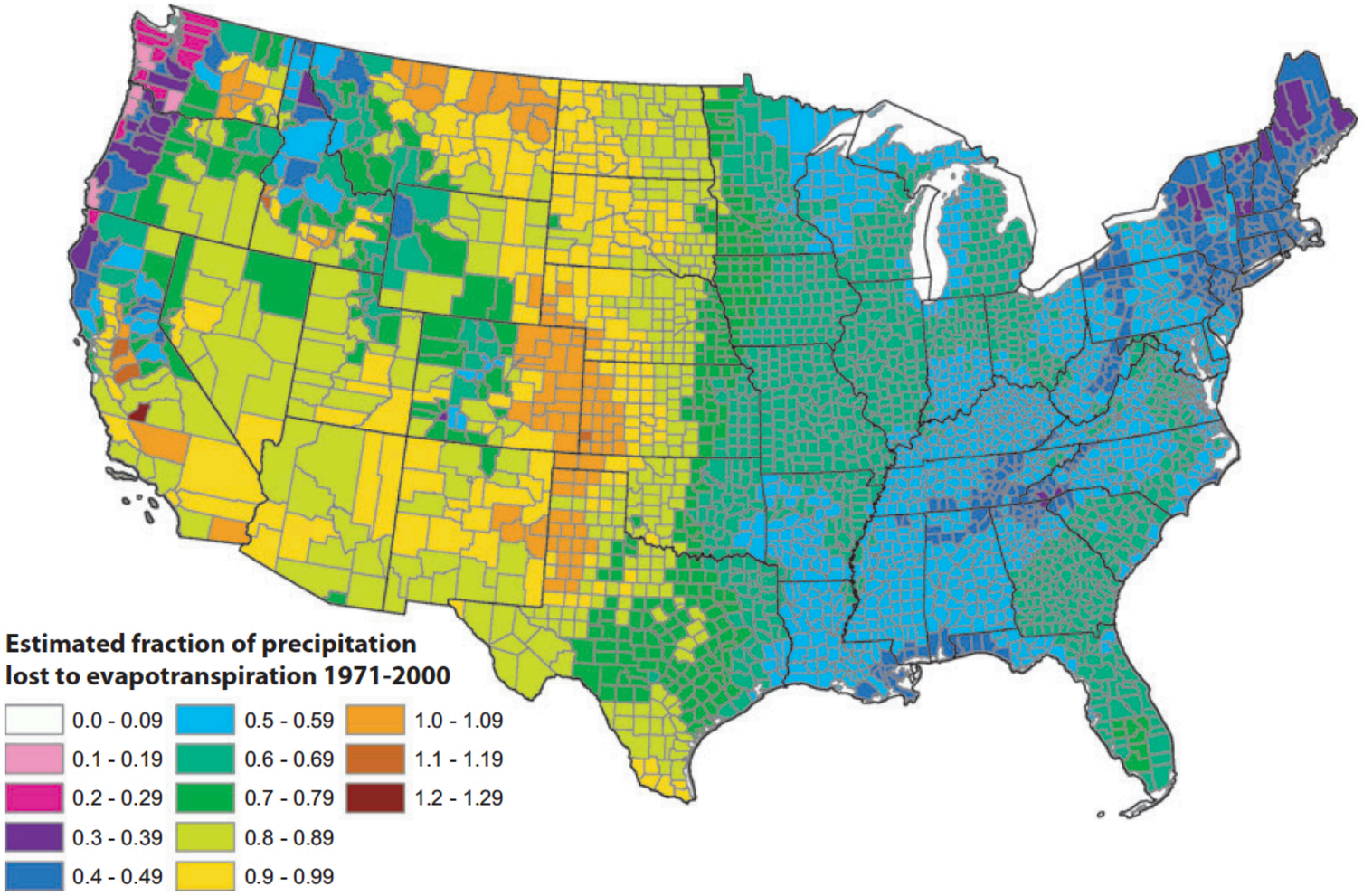
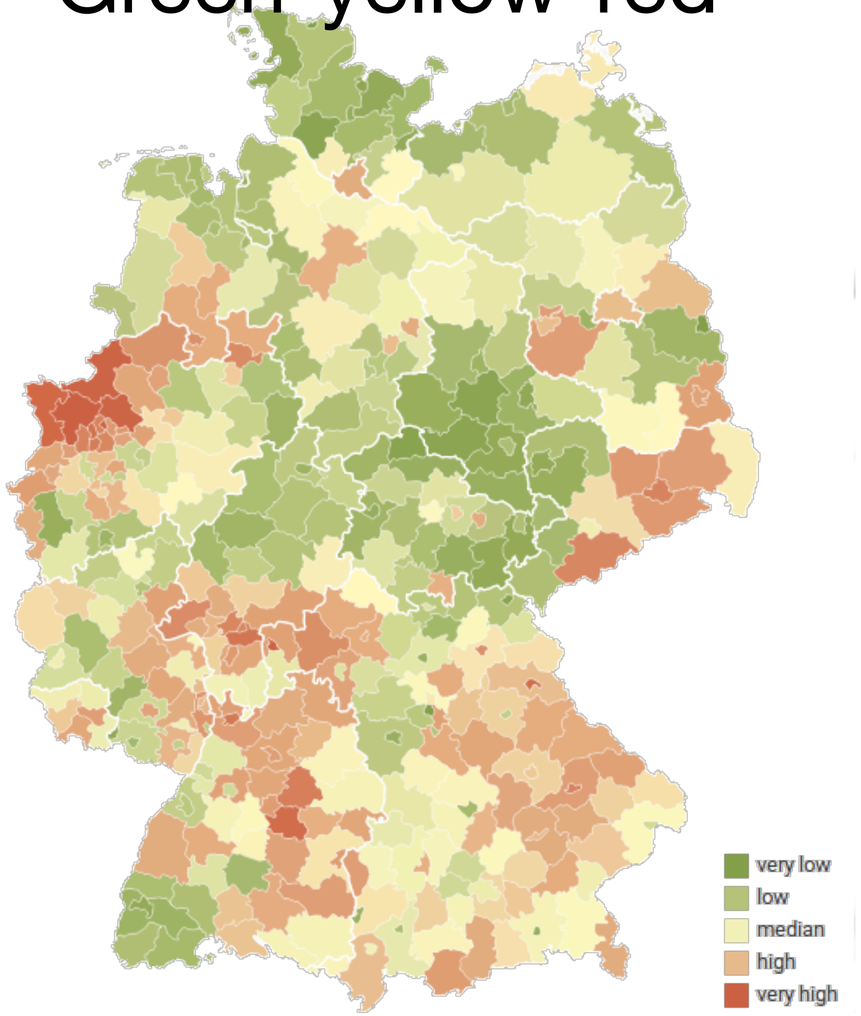


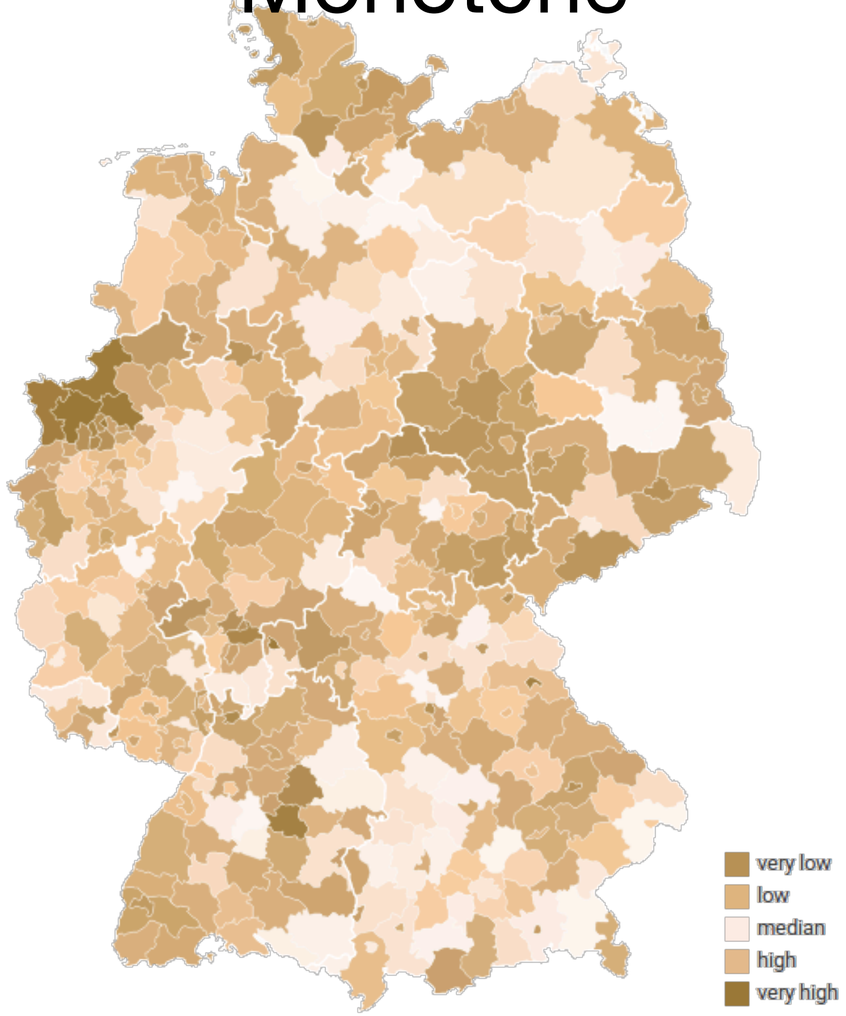
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM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

Color Blindness

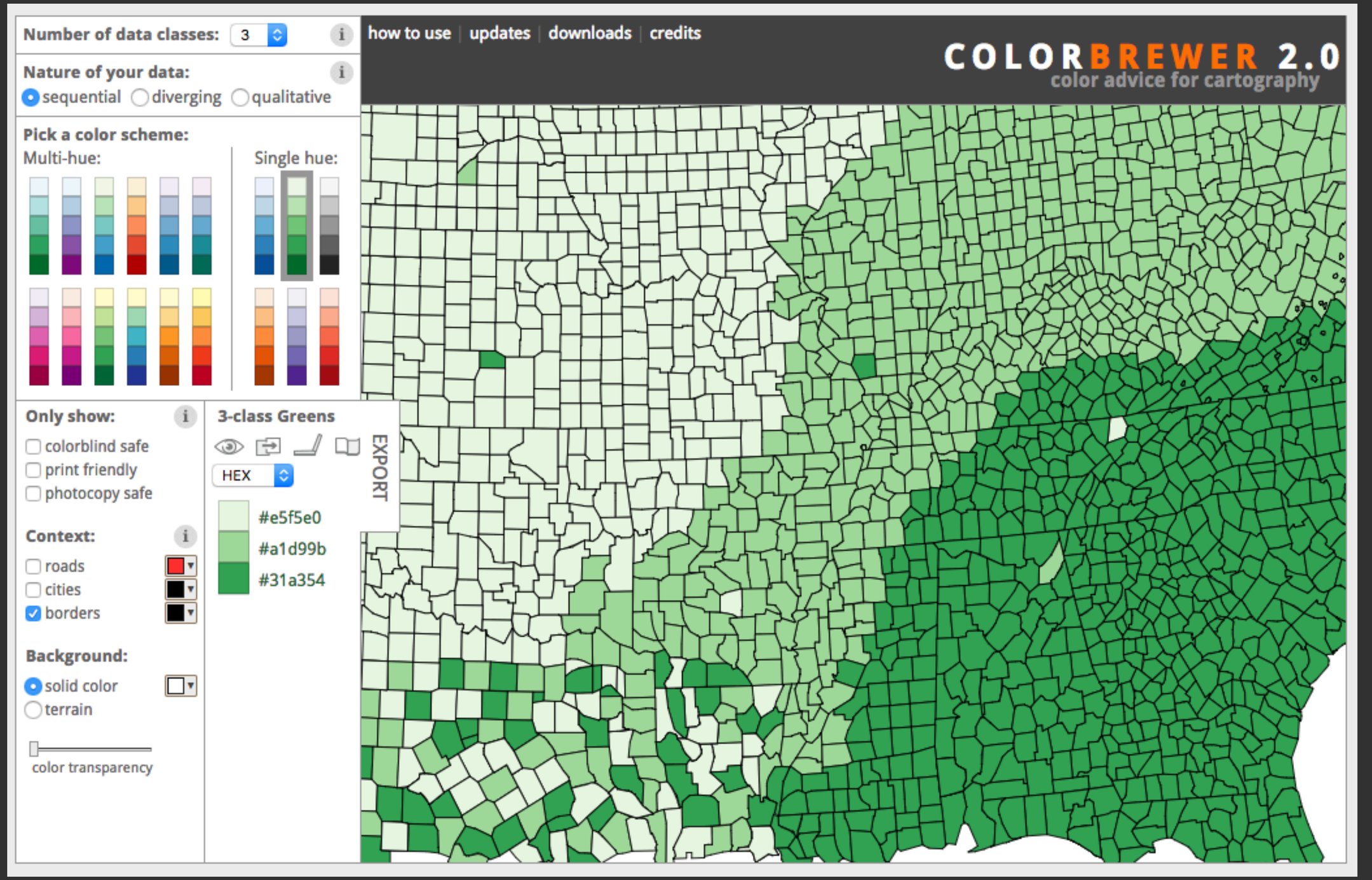
Green-yellow-red



Monotone



https://blog.datawrapper.de/colorblind-check/


Number of data classes: 3
how to use | updates | downloads | credits
COLORBREWER 2.0
color advice for cartography
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

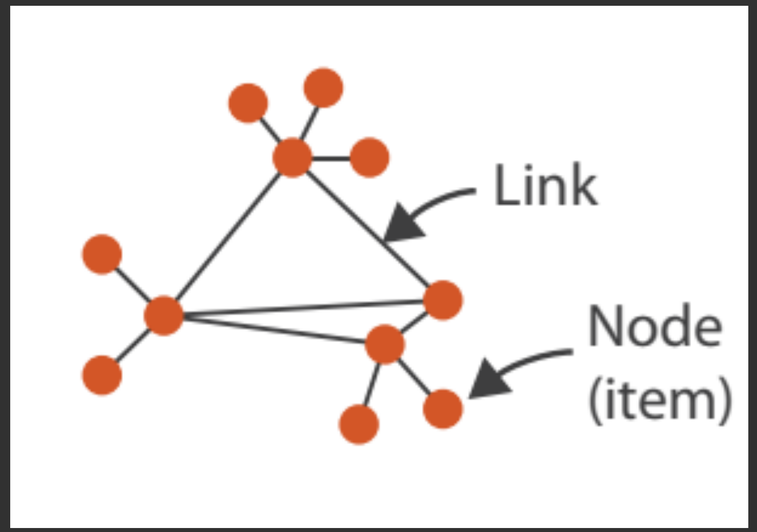
Networks & Trees

Networks

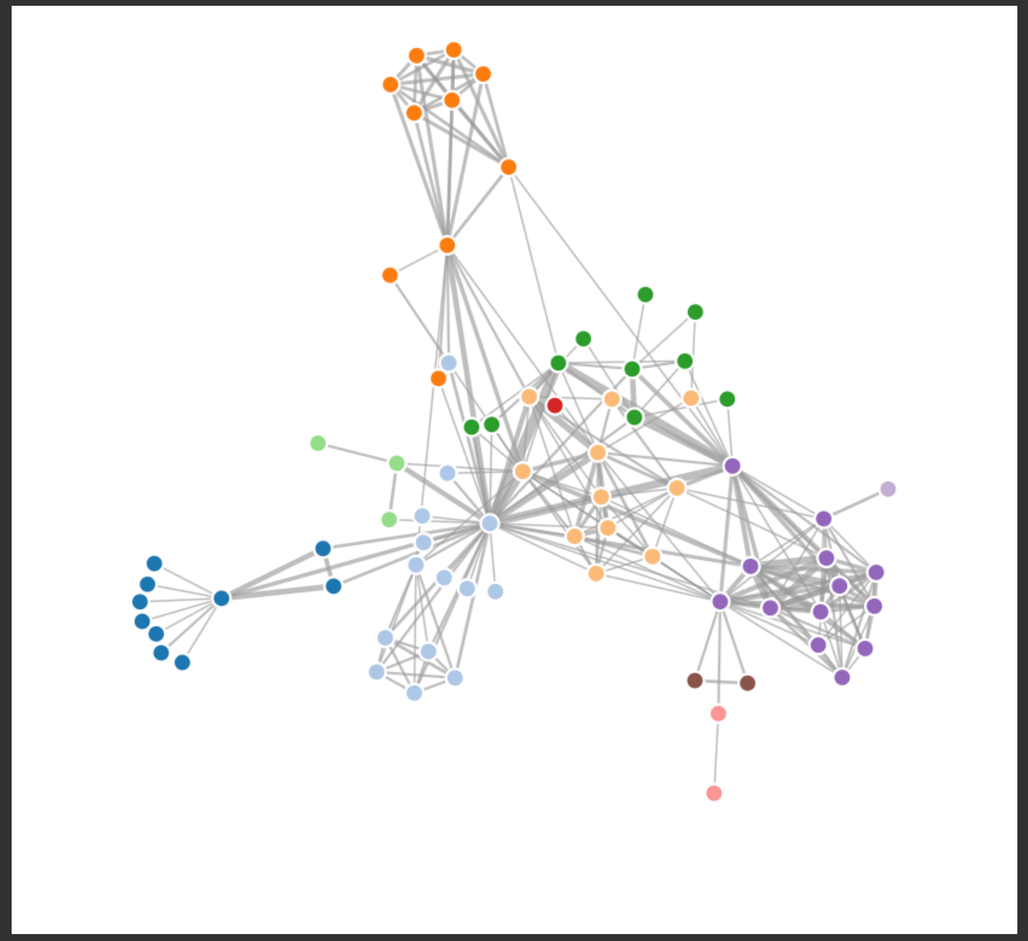
Model relations among data

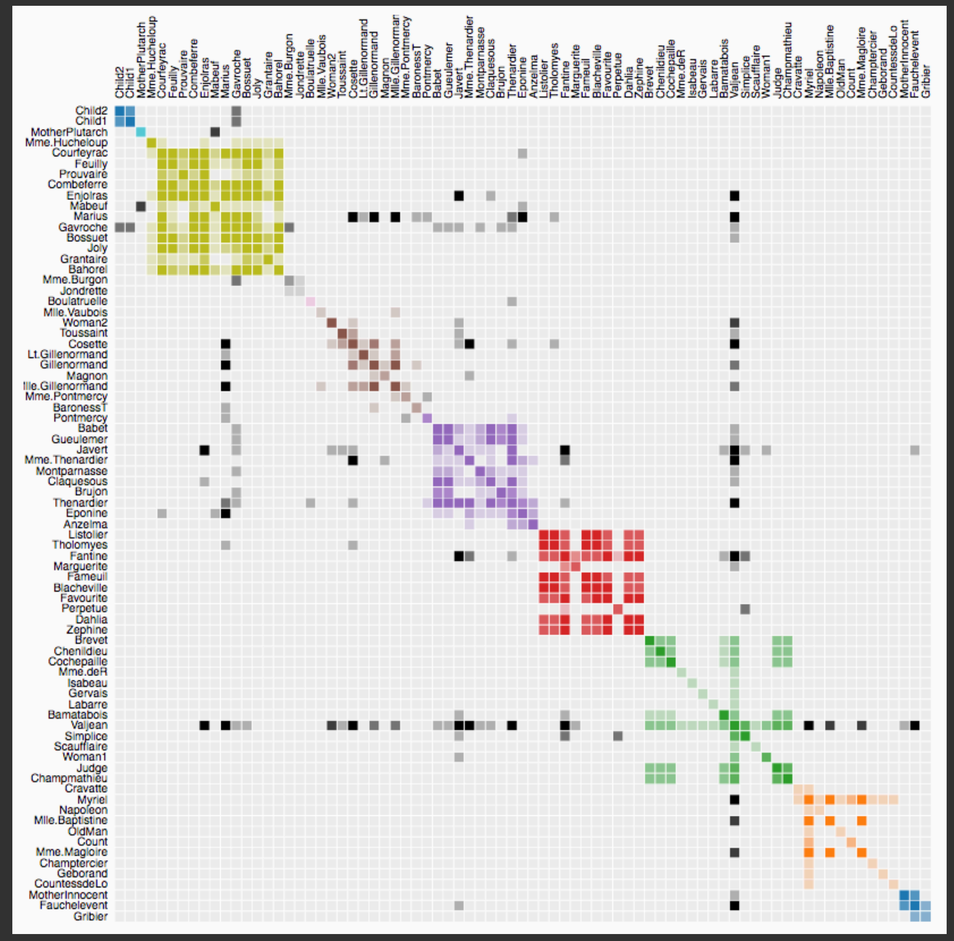
Nodes and links

Node and link attributes



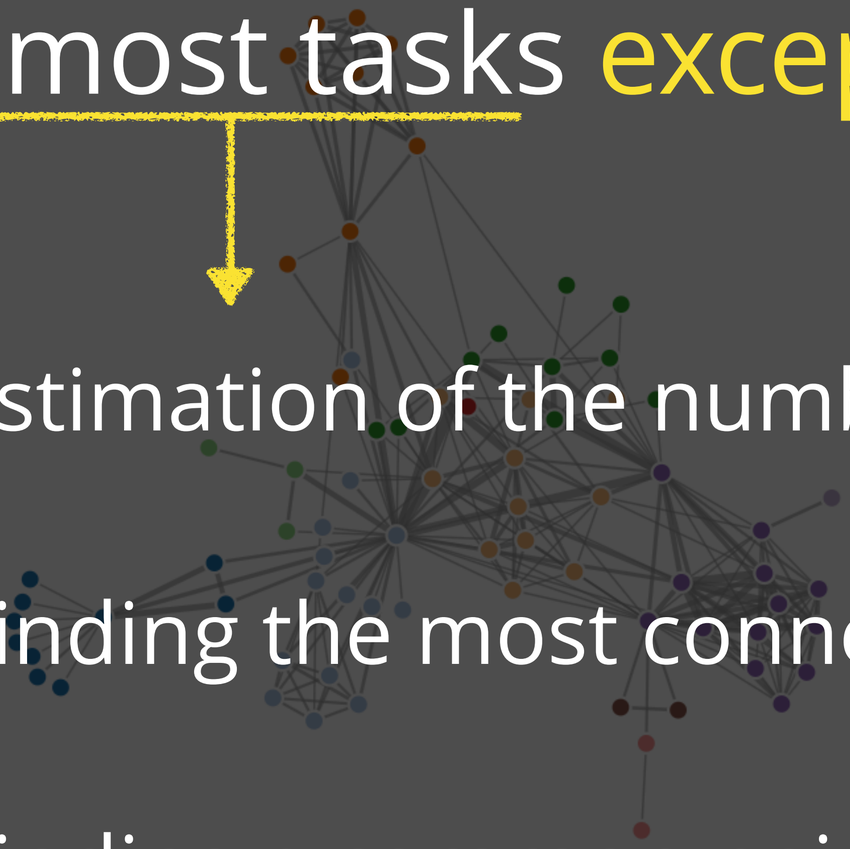
Node-Link Diagram vs Matrix Diagram





Les Misérables Character Co-occurrence

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



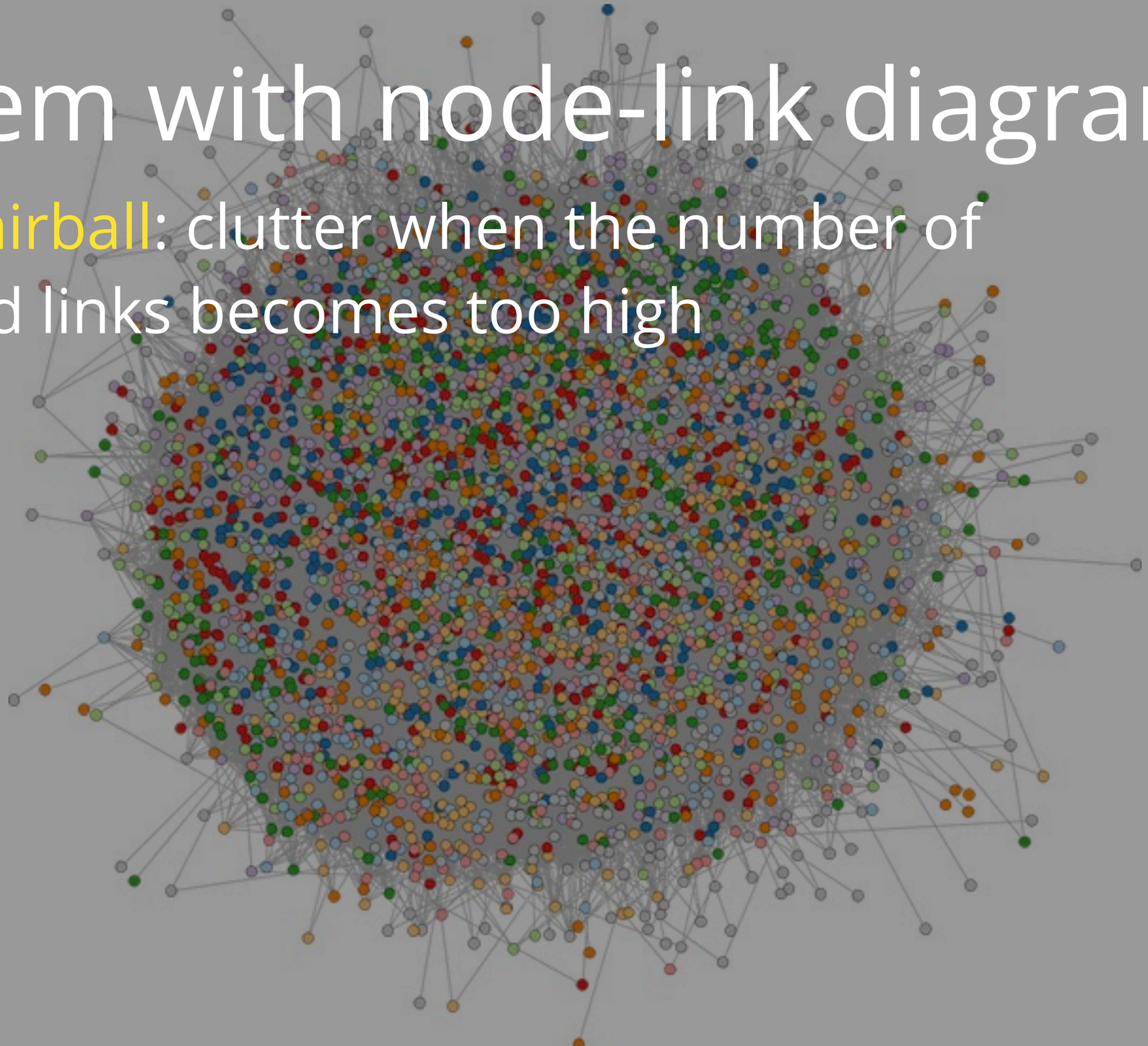
- Estimation of the number of nodes/links
- Finding the most connected node
- Finding a common neighbor between two specified nodes

Node-Link Diagram

Matrix Diagram

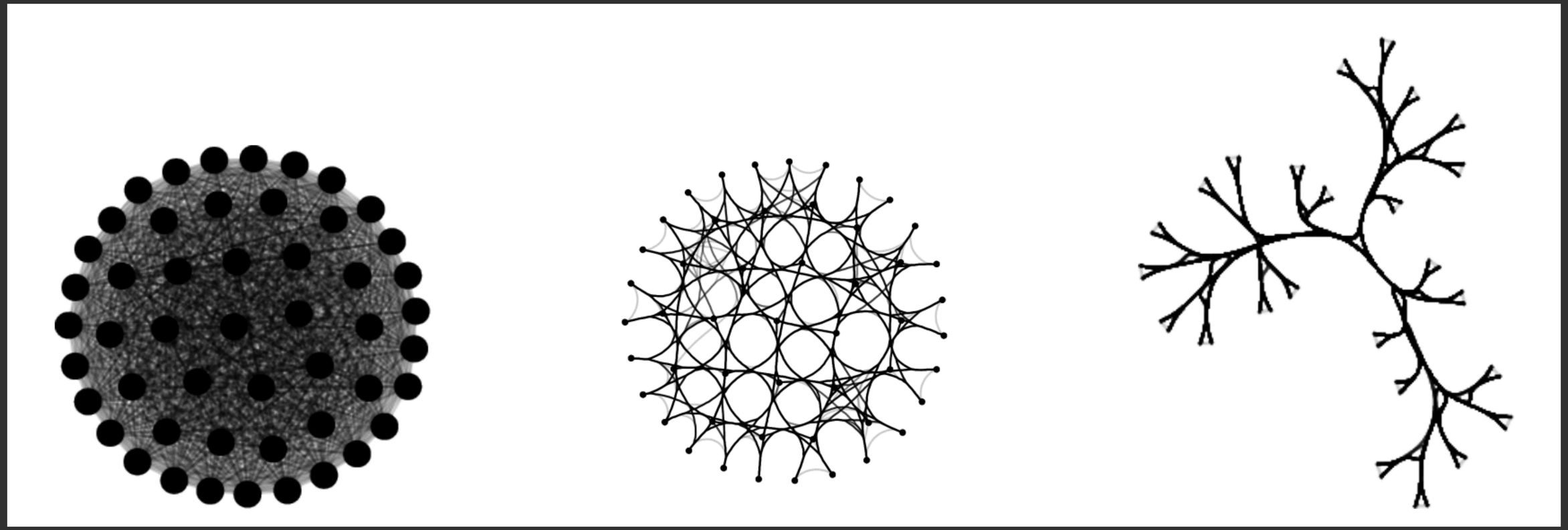
Problem with node-link diagram

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



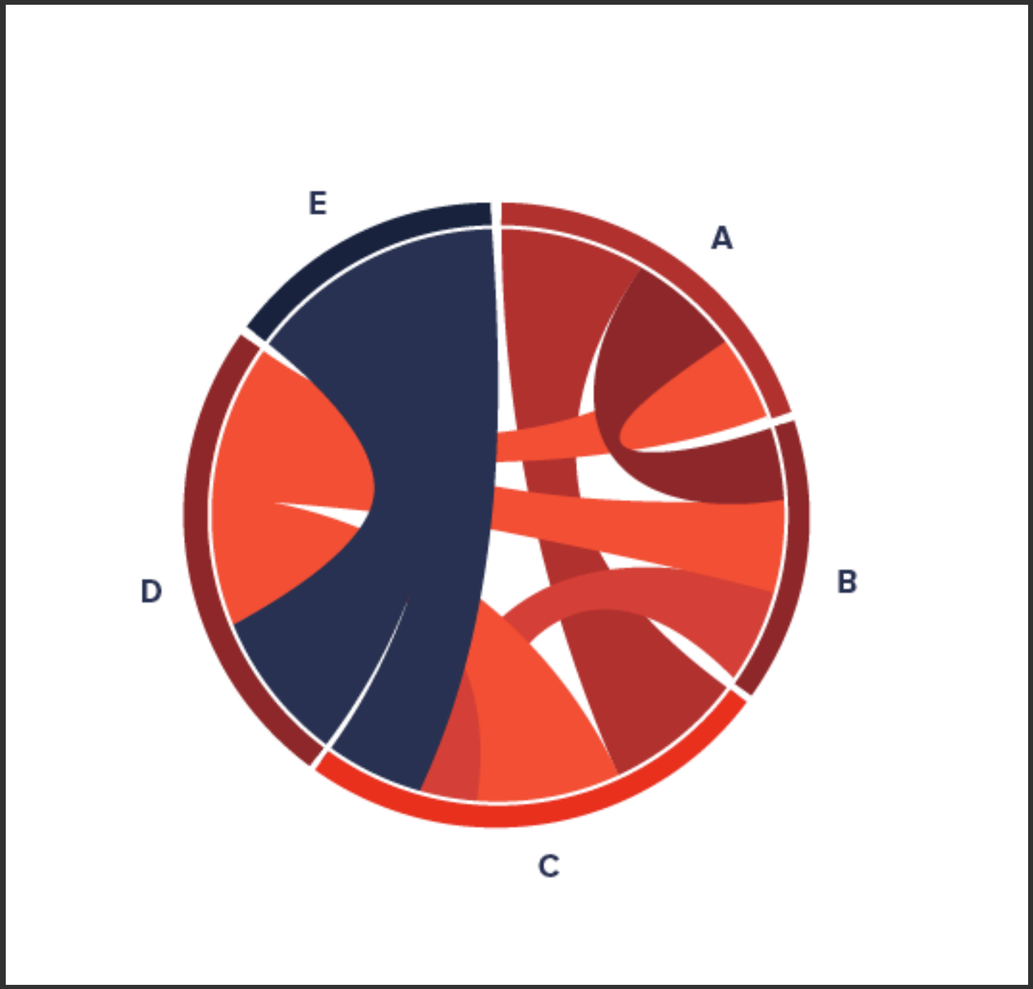
[A. Nocaj 2015]

Original Network Edge Bundling Confluent Drawing

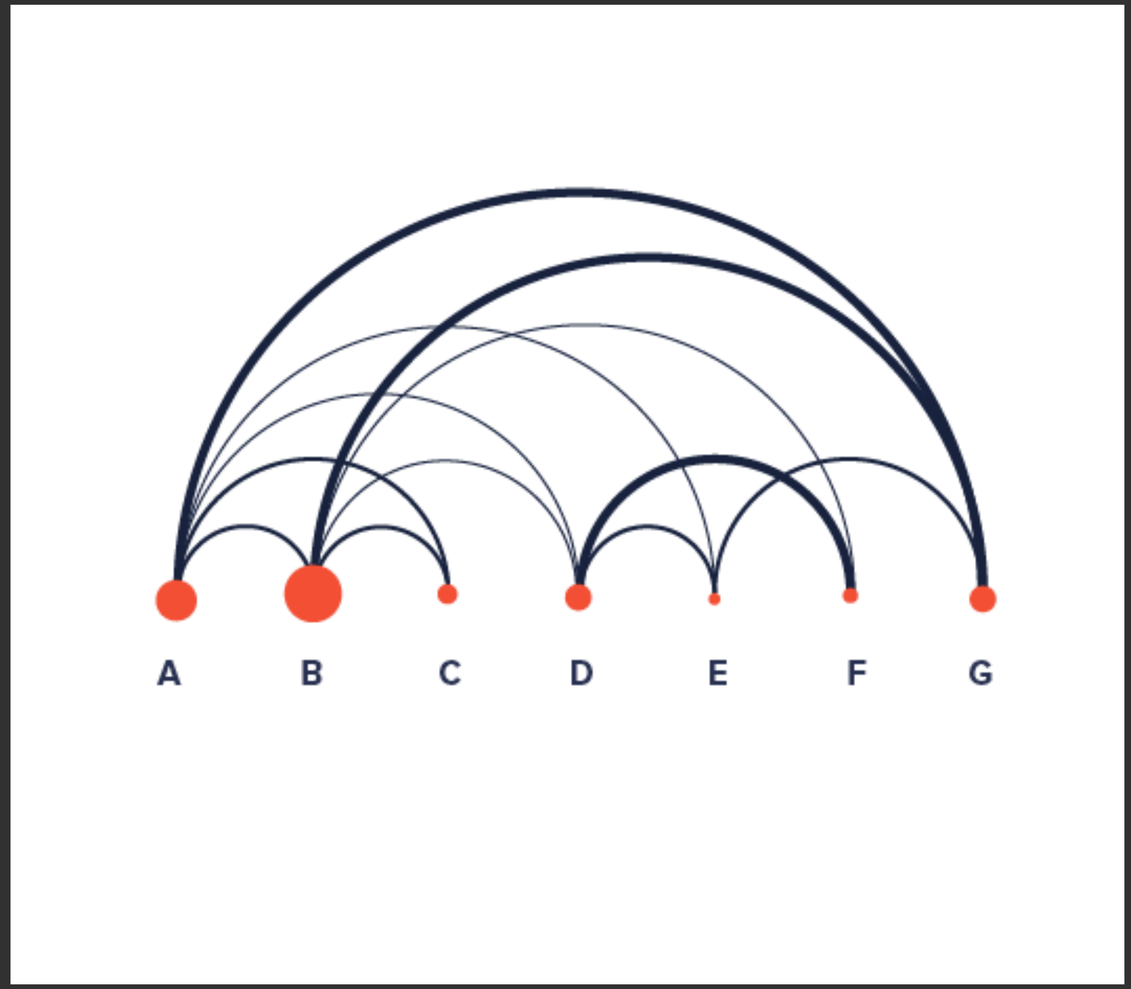


[B. Bach 2017]

Alternatives. Are these better?



Chord Diagram

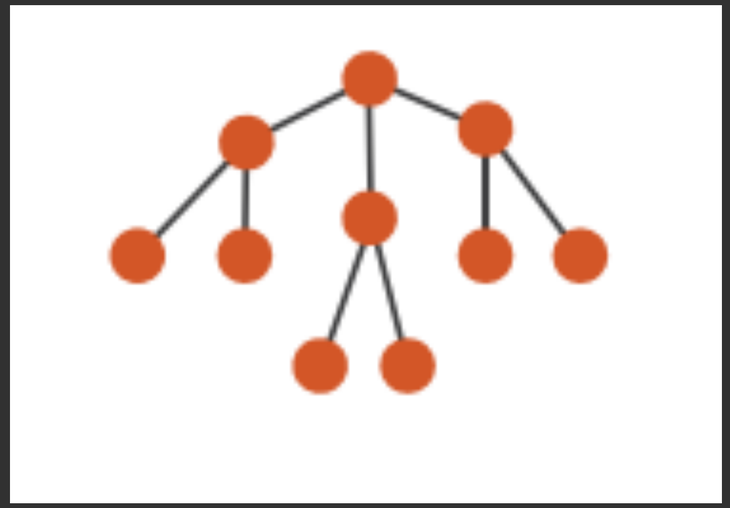


Arc Diagram

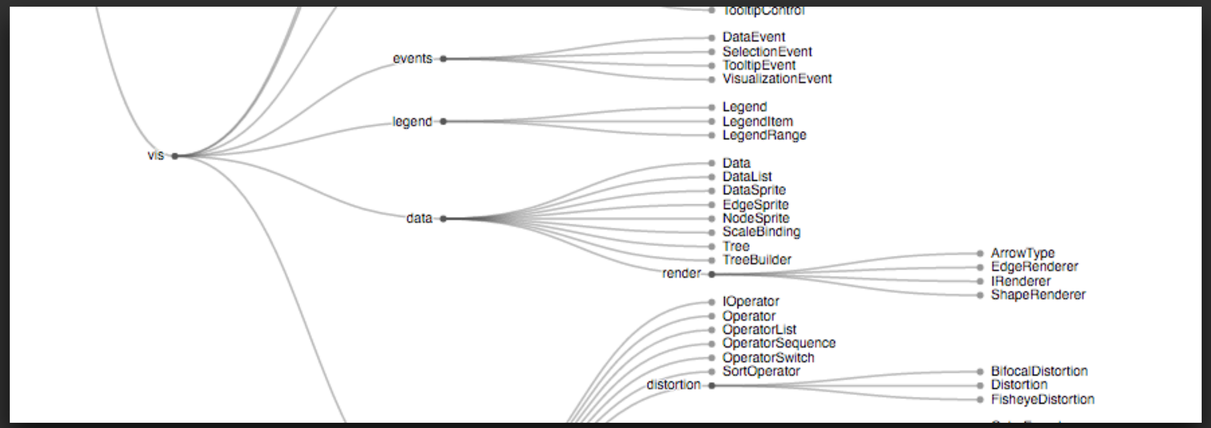
Trees

Graphs with hierarchical structure

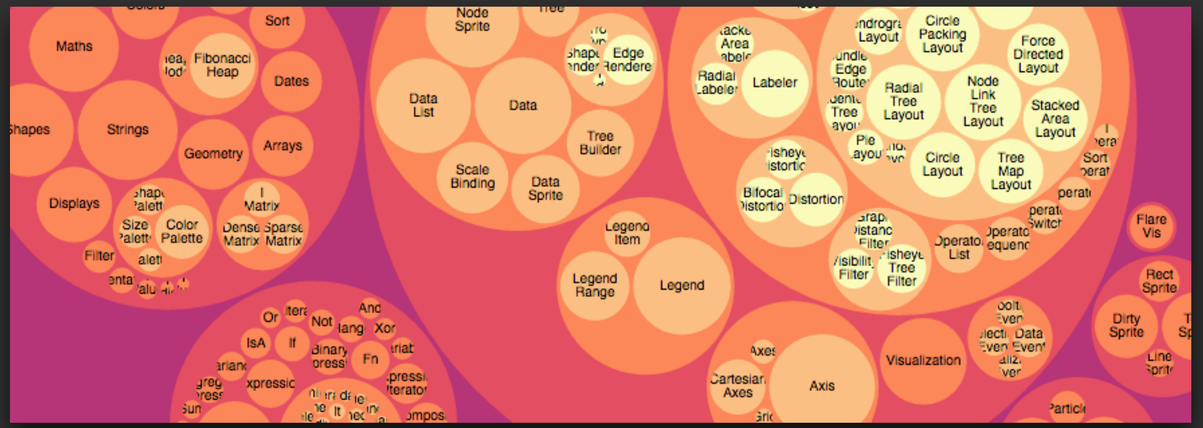
Nodes as parents and children



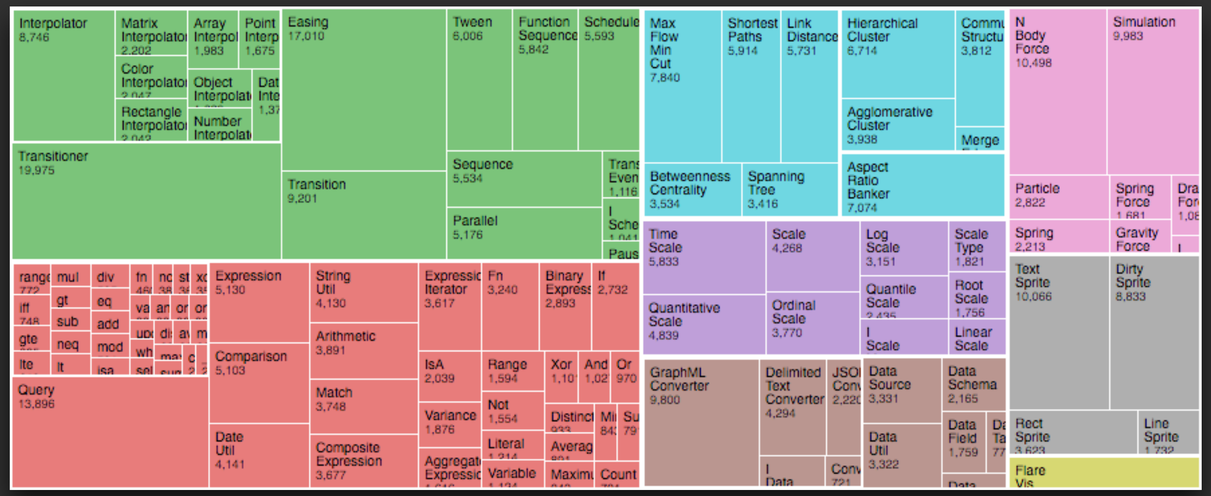
Dendogram



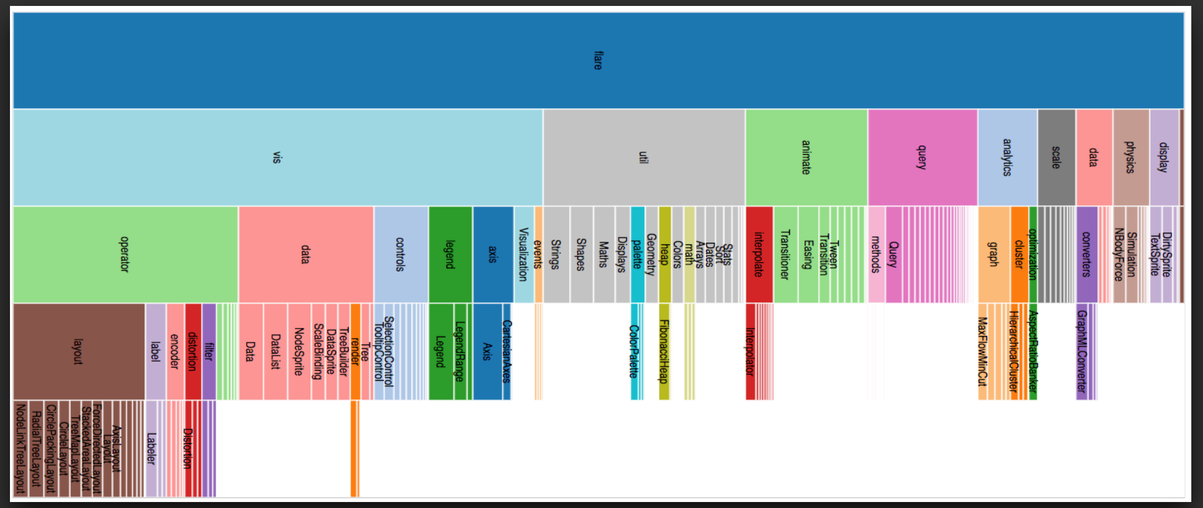
Circle Packing



Treemap

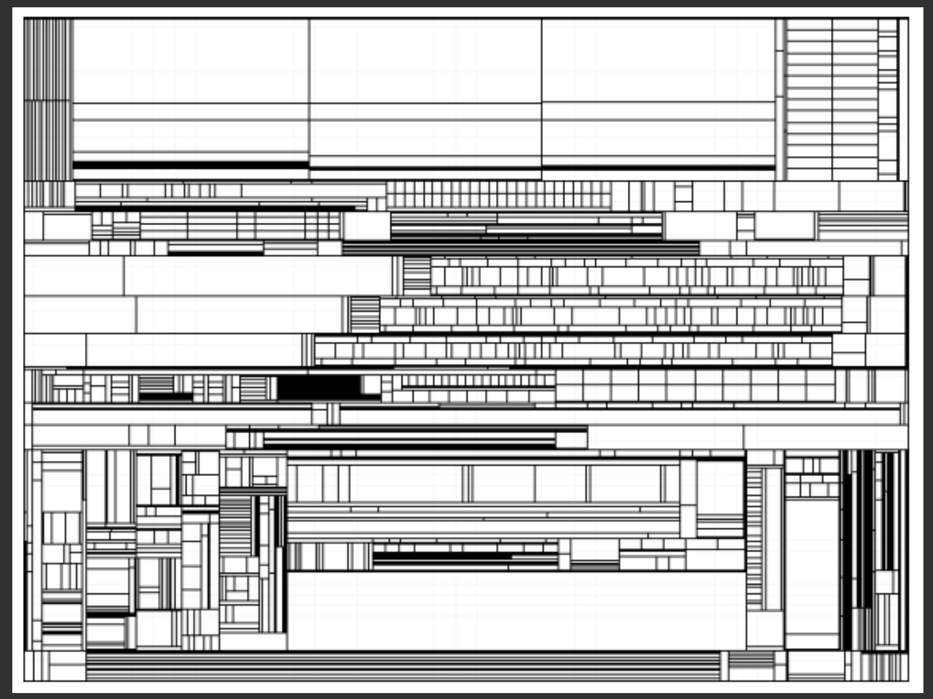


Icicle

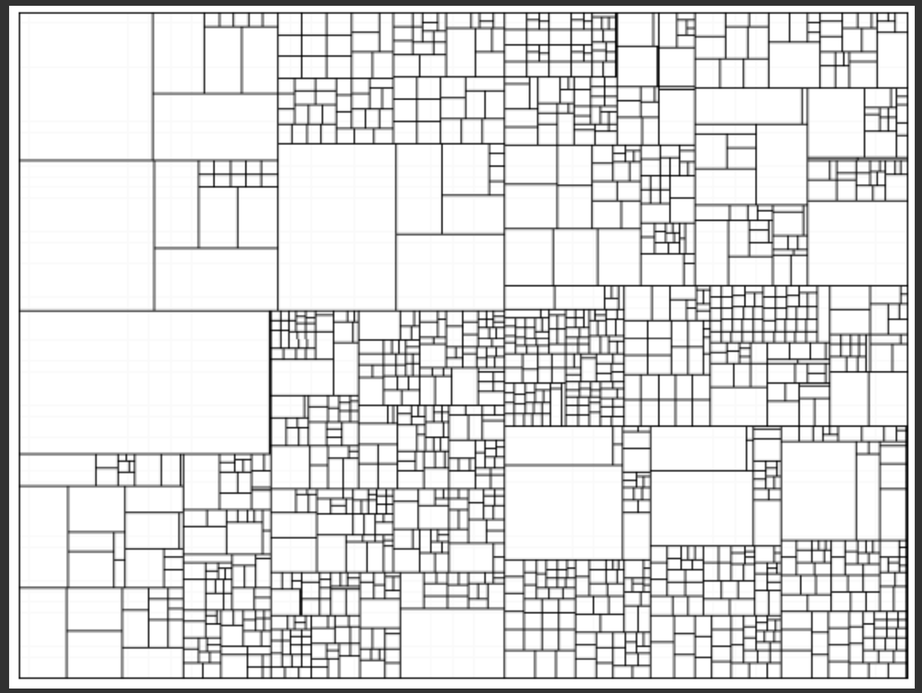


Squarified Treemaps

“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

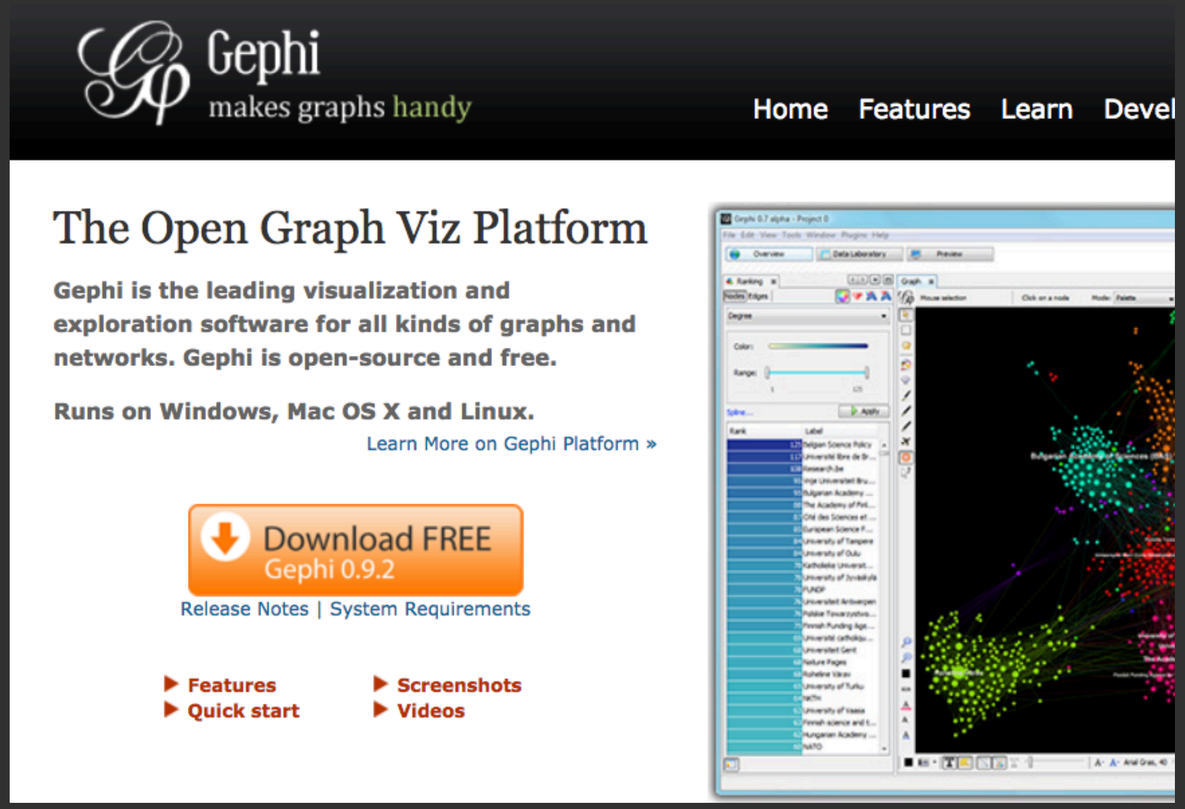


Before

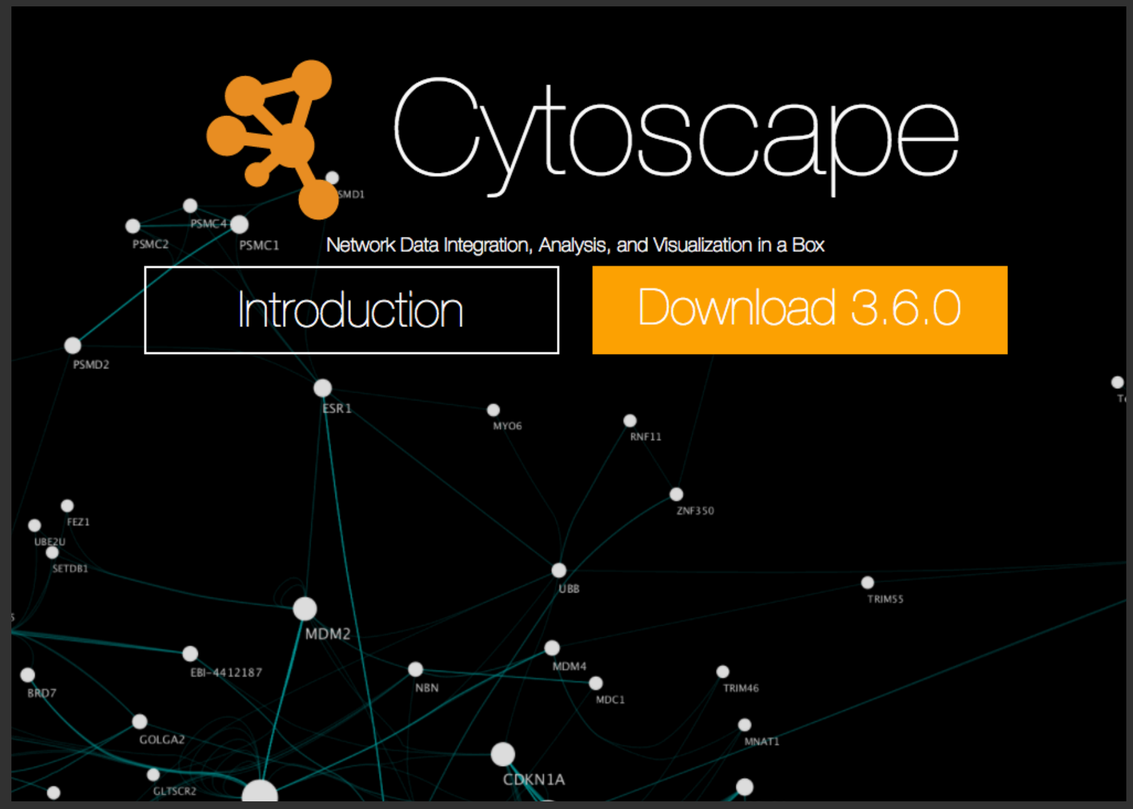


After

Tools for Graph Analysis

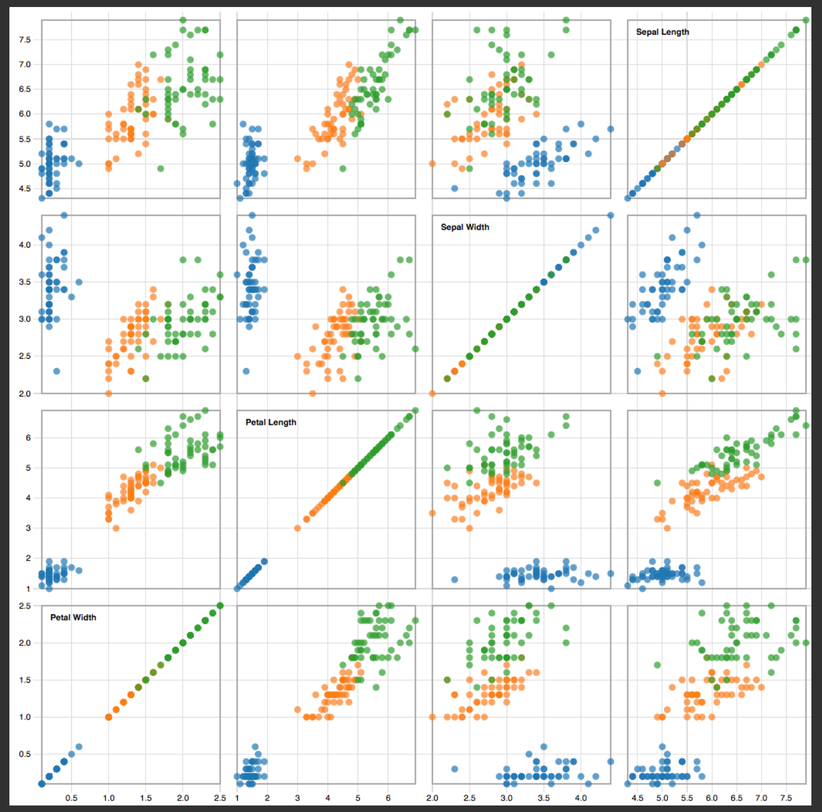


Gephi

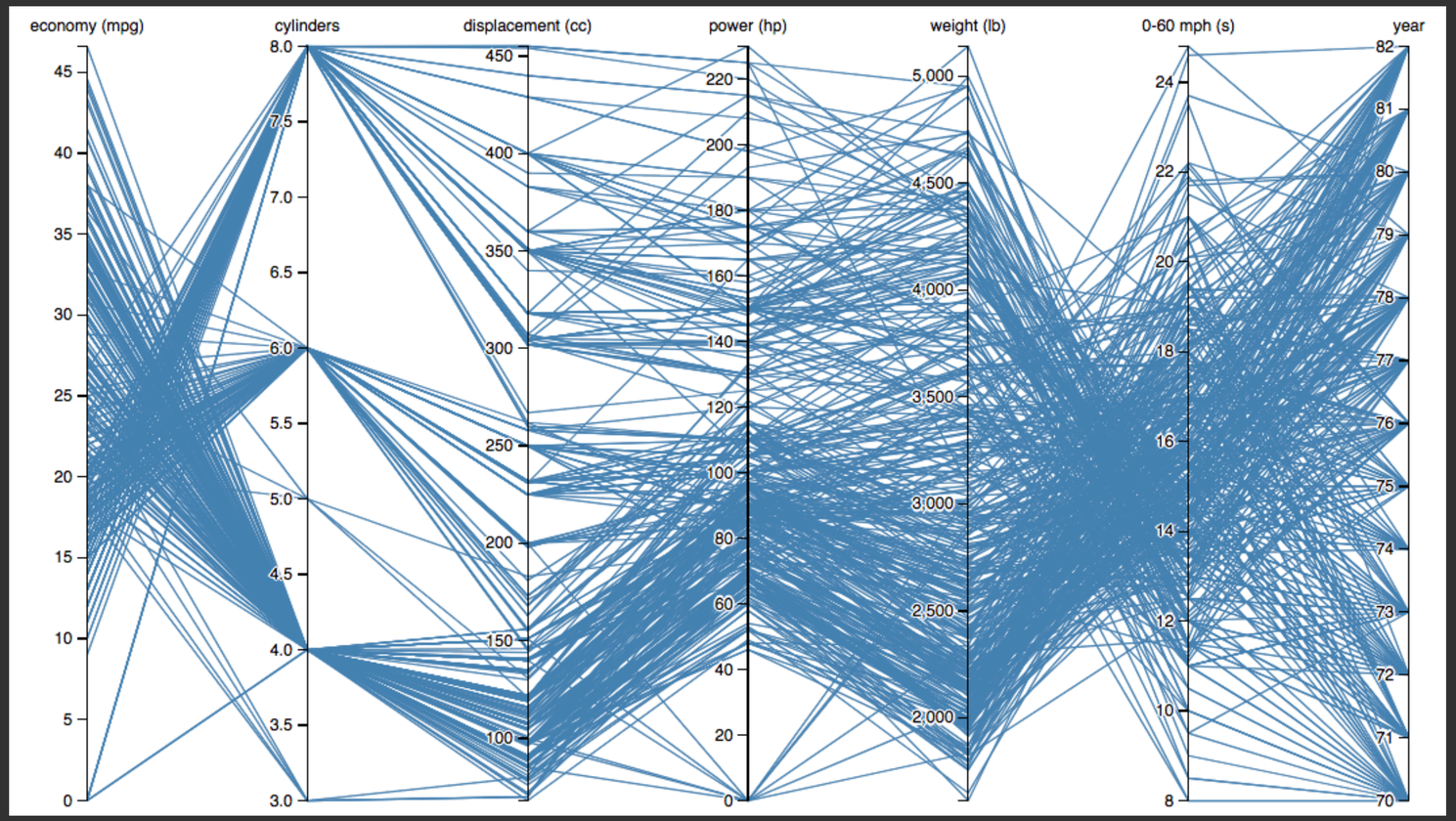


Cytoscape

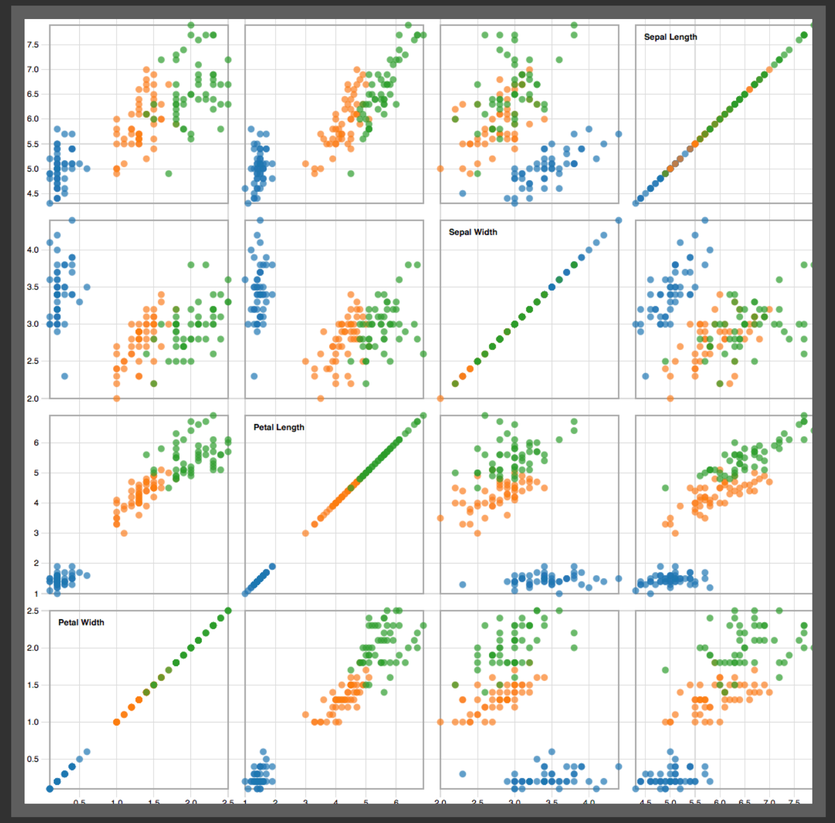
High-Dimensional Data

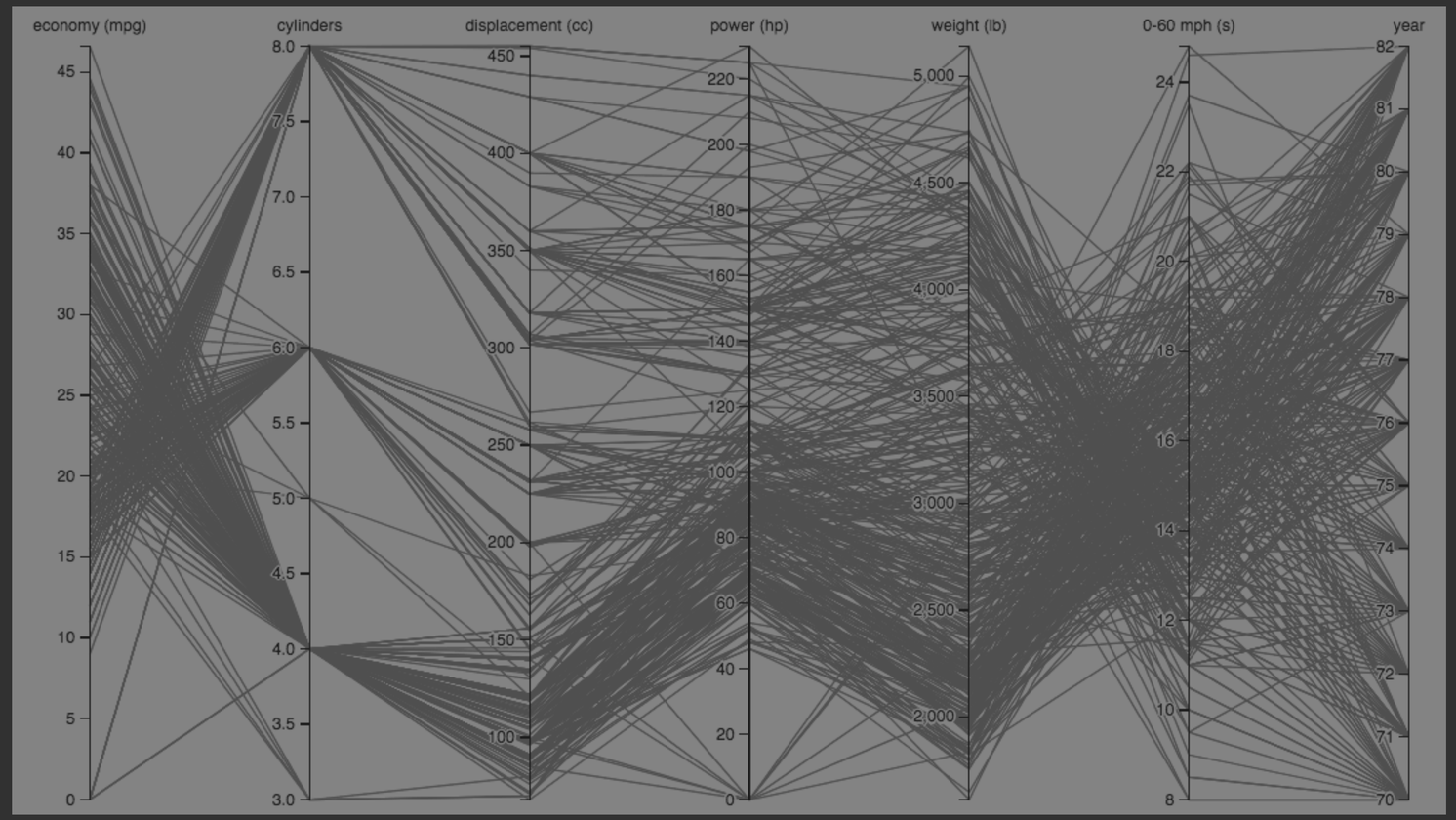


Scatterplot Matrix



Parallel Coordinates





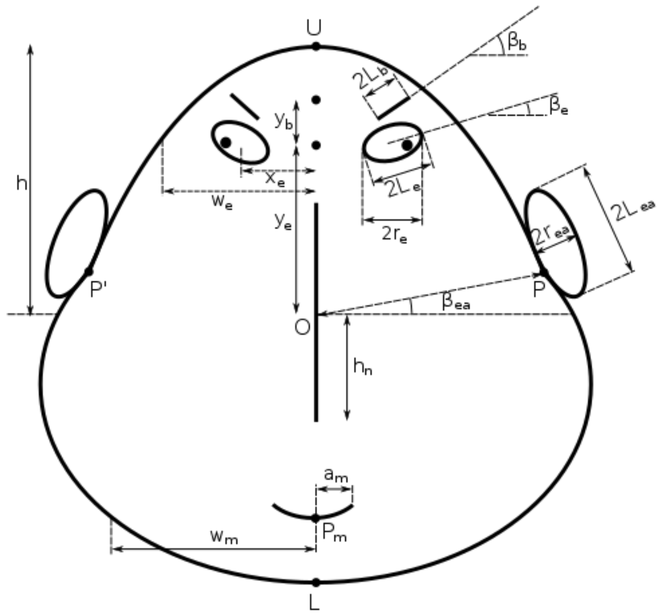
Scatterplot Matrix

Parallel Coordinates

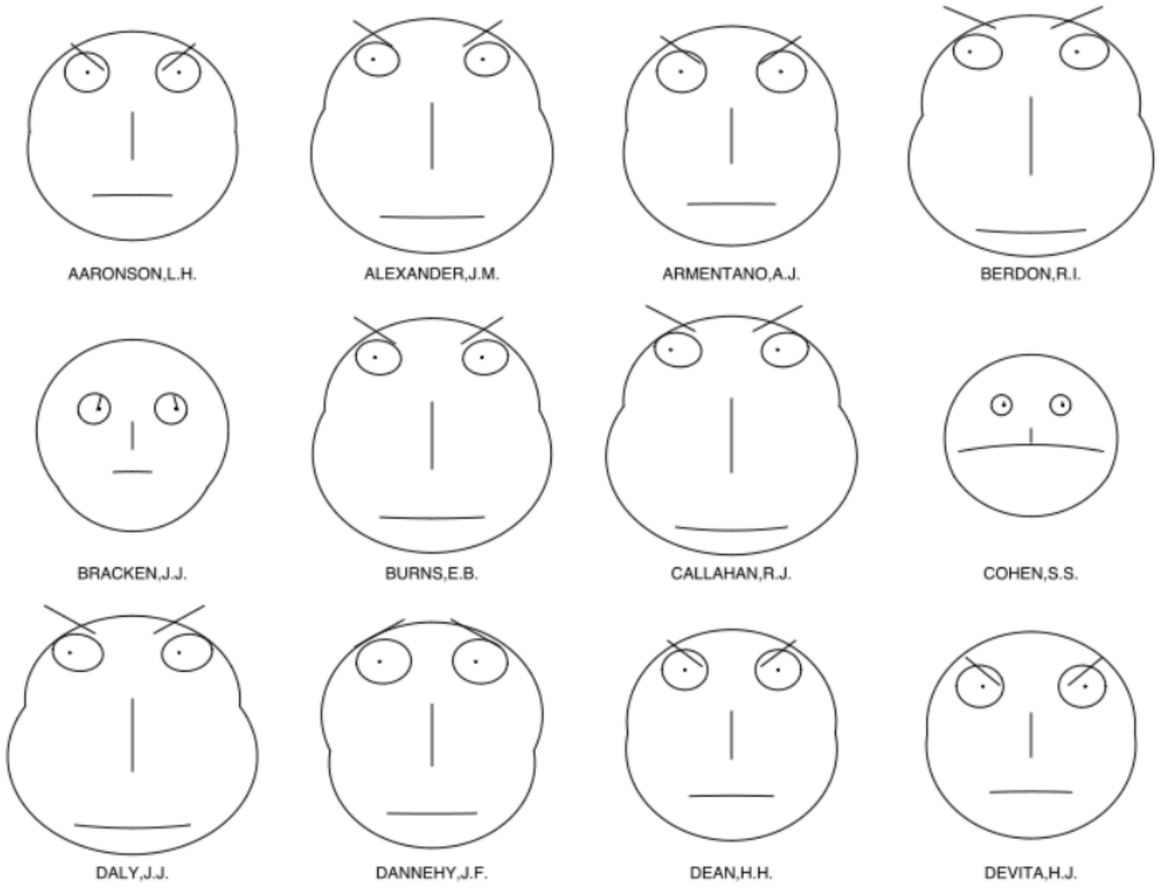
Winner for analyzing correlations between multiple variables

[L. Harrison 2014]

Chernoff Faces

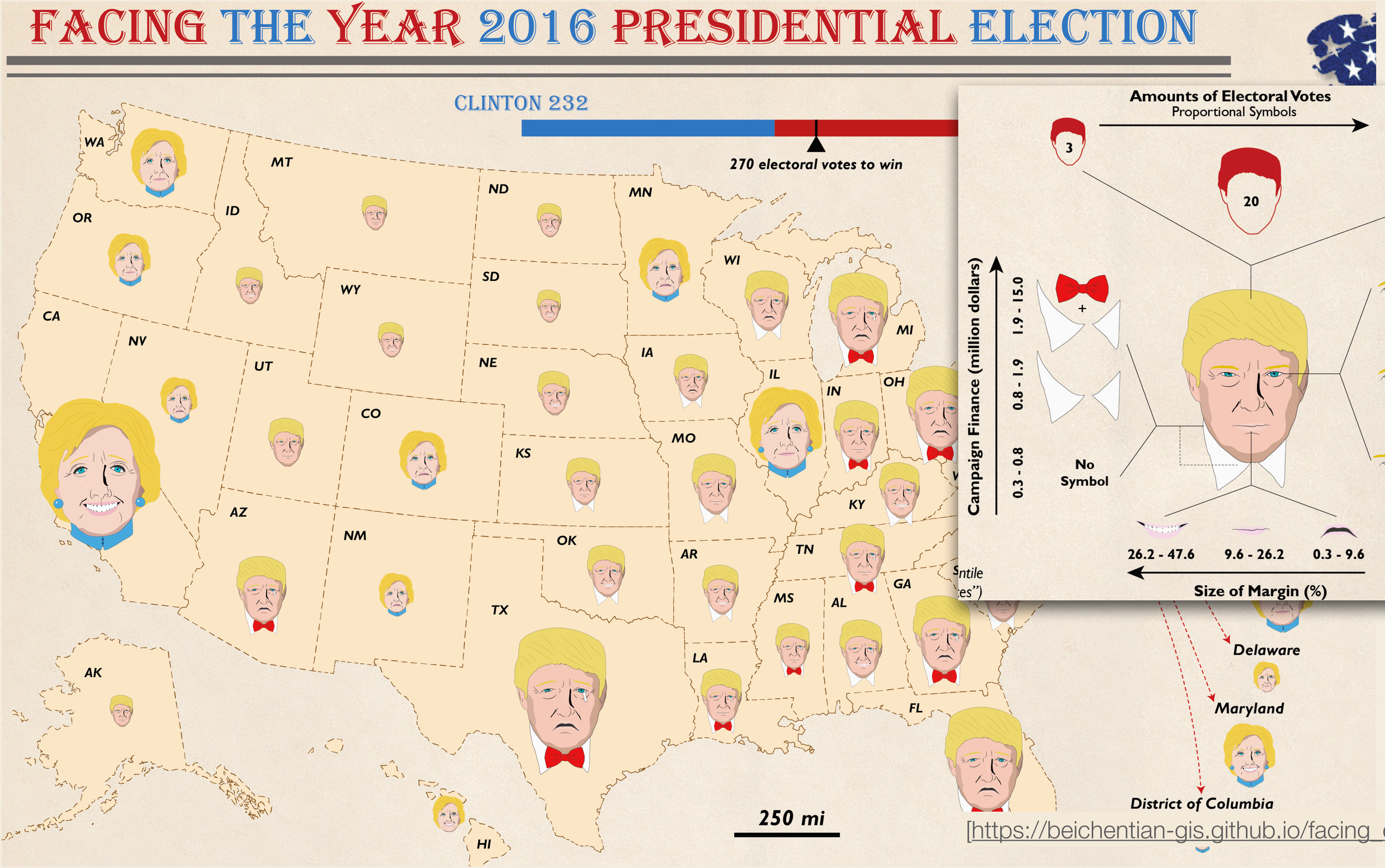


Lawyers' ratings of twelve judges



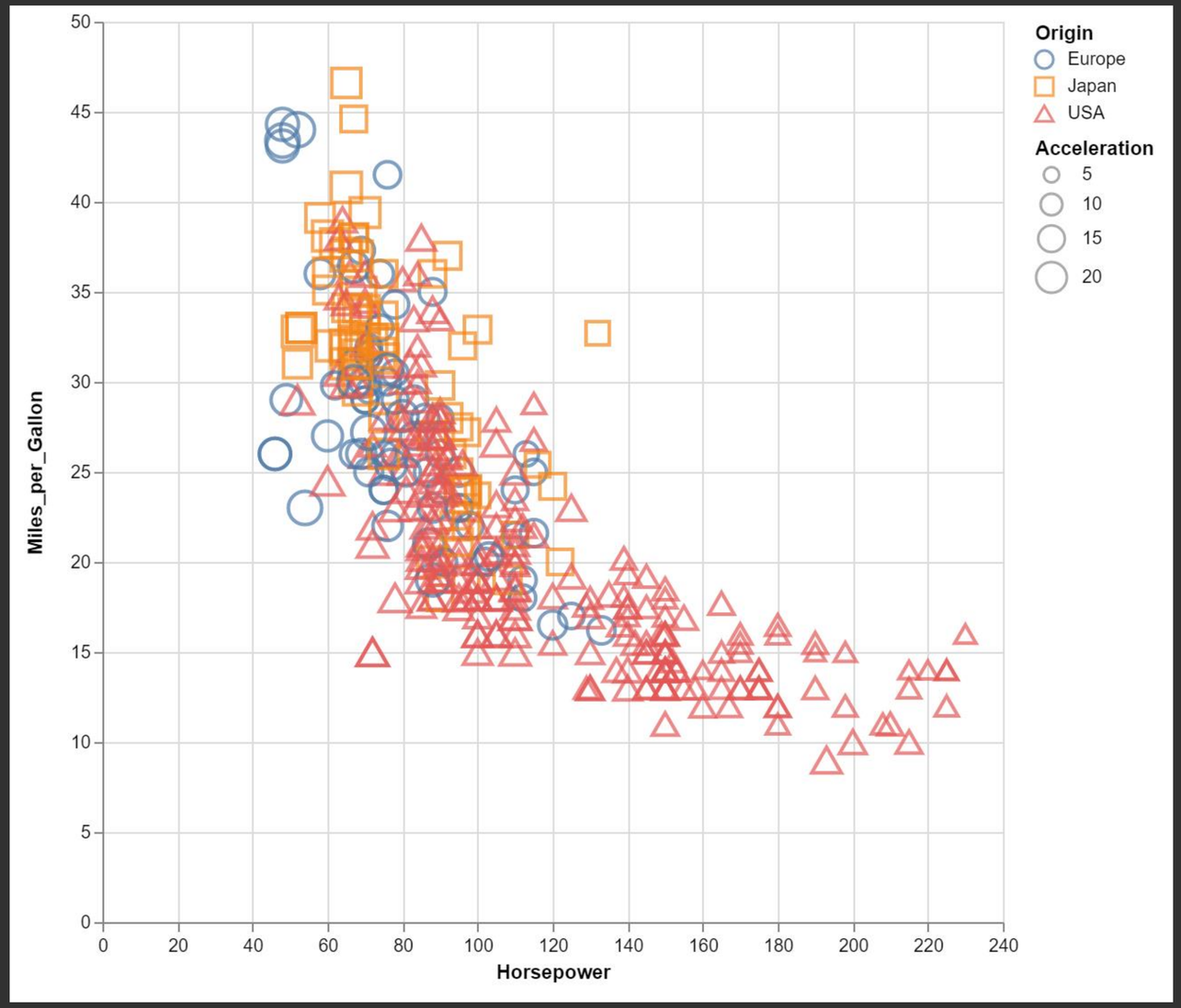
[https://en.wikipedia.org/wiki/Chernoff_face]

FACING THE YEAR 2016 PRESIDENTIAL ELECTION



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

Don’t overload visual encodings



Five visual variables in a single chart!

Origin: Shape

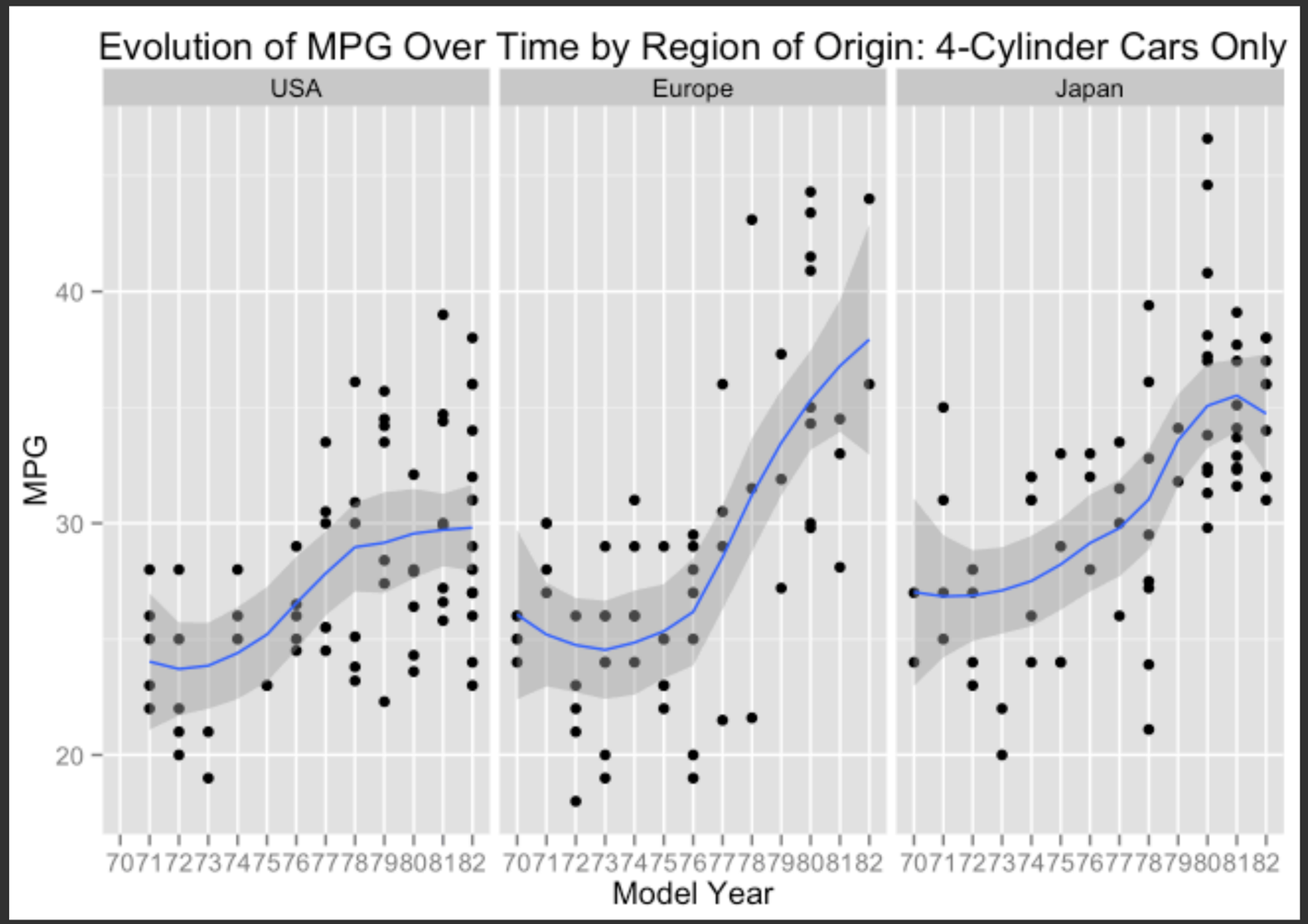
Acceleration: Size

X-axis: Horsepower

Y-axis: Miles per Gallon

Size: Acceleration

Use facets or small multiples



[David K. Smith]

[Andrew Gelman]

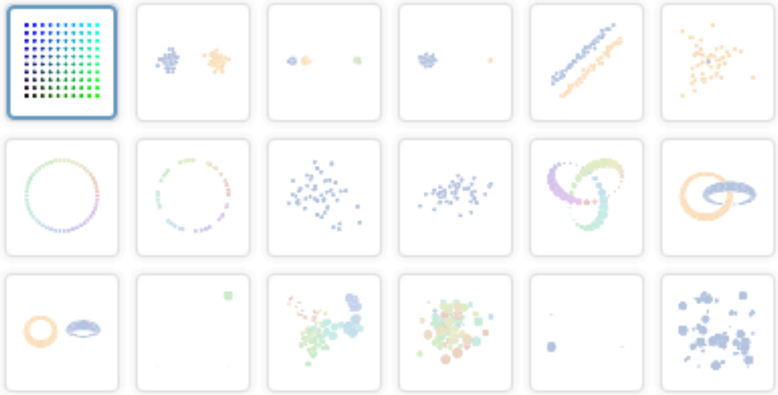
Dimension Reduction

Mapping from high-dimensional space to two or three dimensions

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

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.



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.

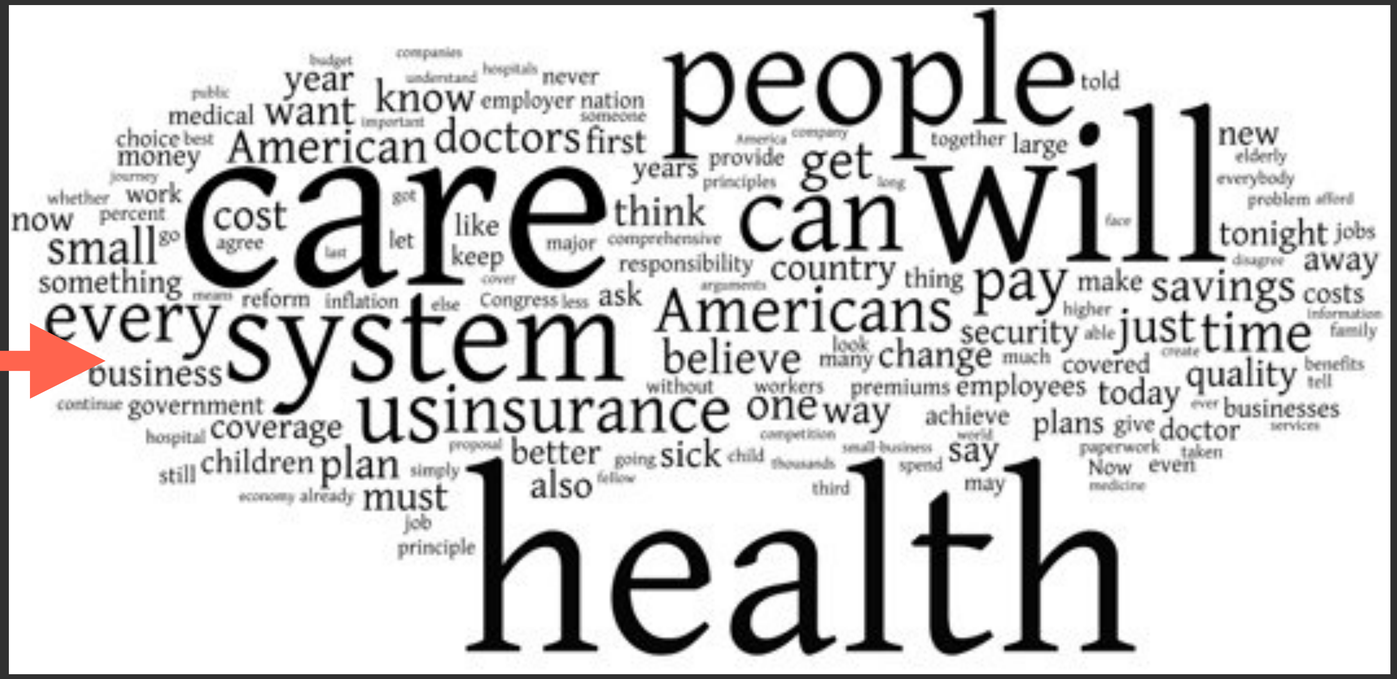
Madame Speaker, Vice President Biden, Members of Congress, and the American people:

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.

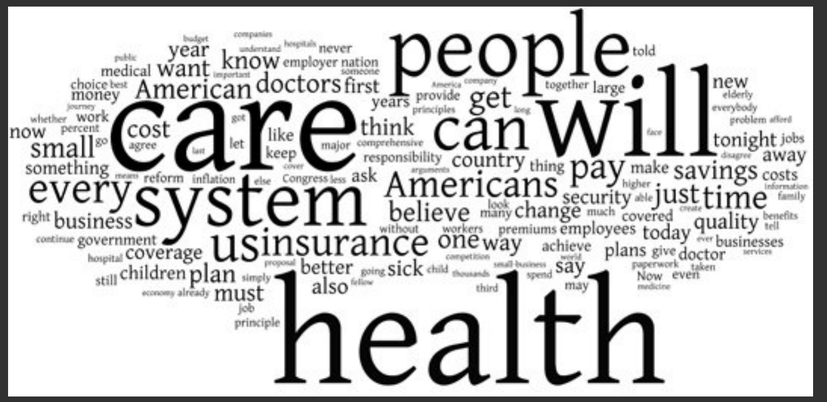
But we did not come here just to clean up crises. We came to build a future. So tonight, I return to speak to all of yo



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 | Home | Create | FAQ | Reference

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

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-colore, 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!

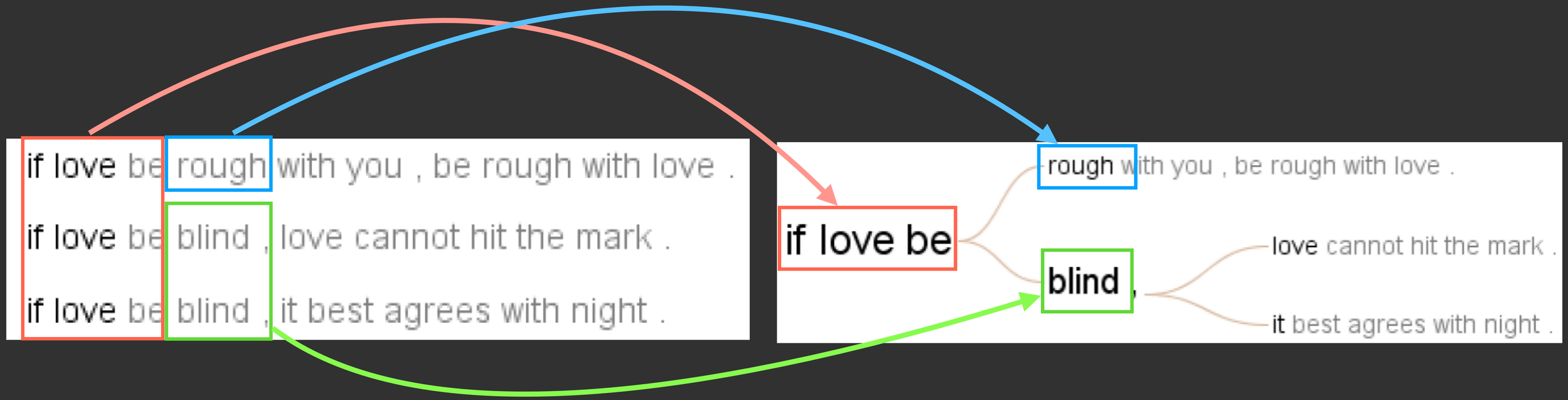
Edit it!

Edit it!

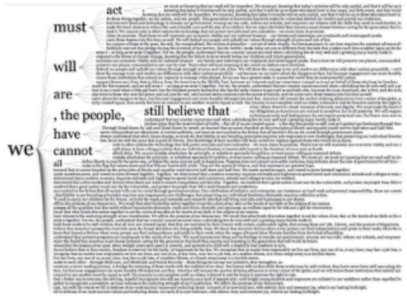
Edit it!

How to Use

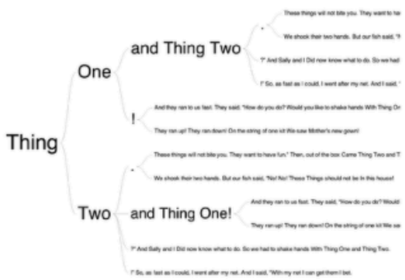
Word Tree: Word Sequences [Wattenberg et al.]


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 .
if love be
rough with you , be rough with love .
blind ,
love cannot hit the mark .
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

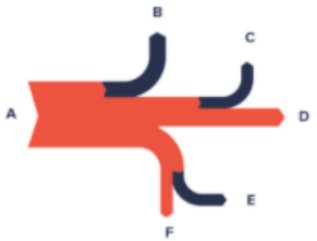
Flickr Comments

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

Paste Text

Other Visualizations

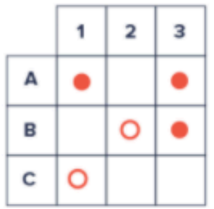
Sankey Diagram



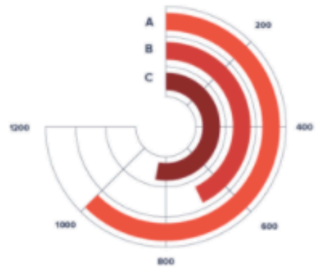
Alluvial Diagram

Donut Chart

Matrix Diagram



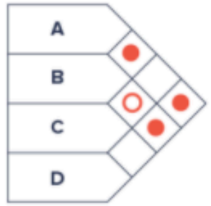
Radial Bar Chart



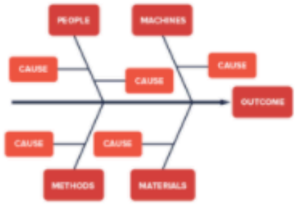
Radial Histogram

Sorted Stream Graph

Matrix Diagram (Roof Shaped)



Fishbone Diagram



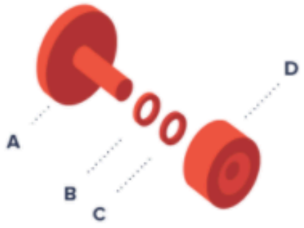
Pictorial fraction chart

Isoline Map

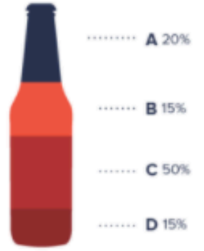
Flow Map

Arc Diagram

Exploded View Drawing



Pictorial Stacked Chart



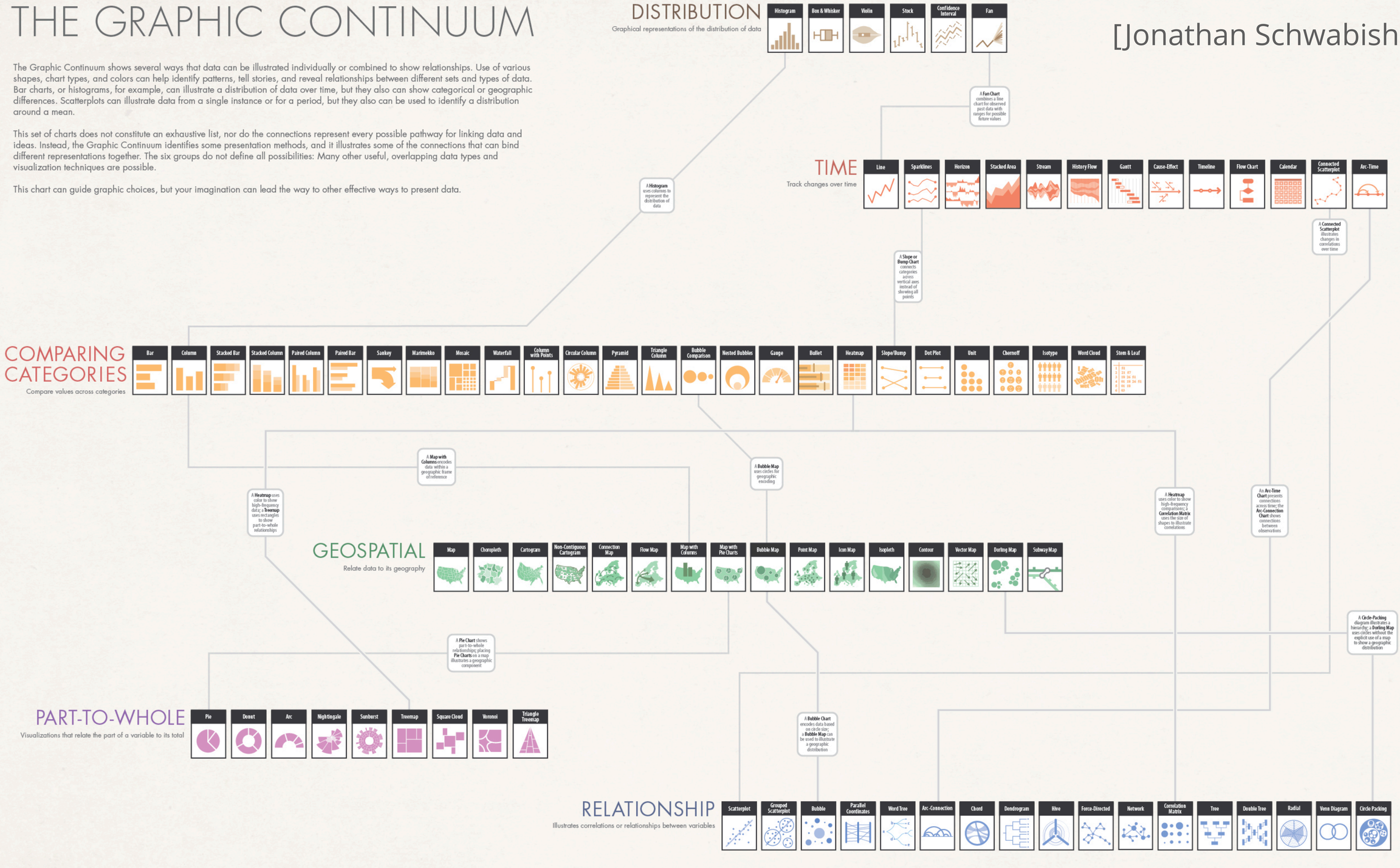
THE GRAPHIC CONTINUUM

[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 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.



Financial Times Visual Vocabulary

Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

Deviation | Correlation | Change v Time | Ranking | Distribution | Part to whole | Magnitude | Spatial | Flow

https://ft-interactive.github.io/visual-vocabulary/

Change v Time

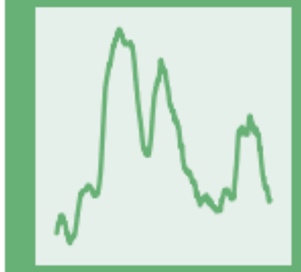
Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader

Examples of use

Share price movements, economic time series

Chart types

line



column-timeline

column-line-timeline

stock-price

slope

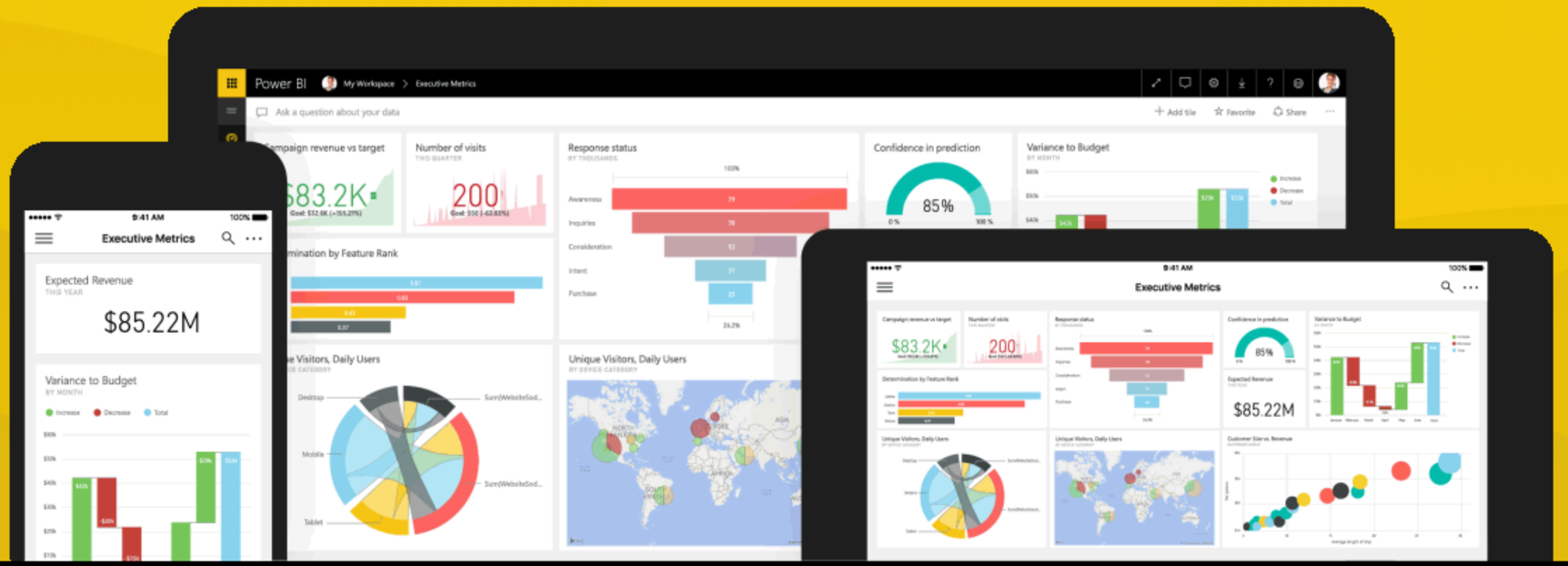
area

Tools

Business intelligence like never before

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

PowerBI



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

plotly

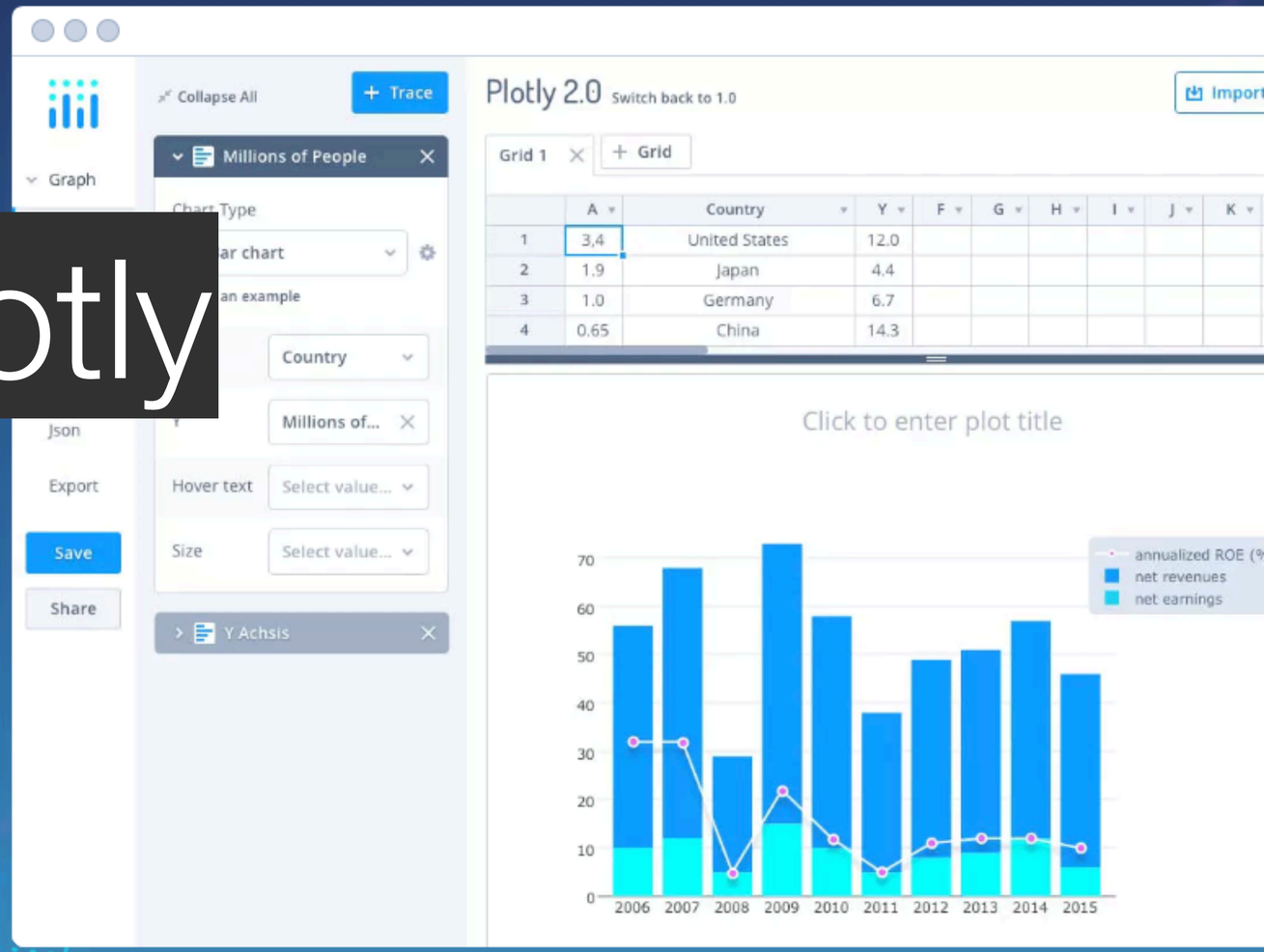
CONSULTING PRICING PRODUCTS MASTER CLASSES

Chart Studio

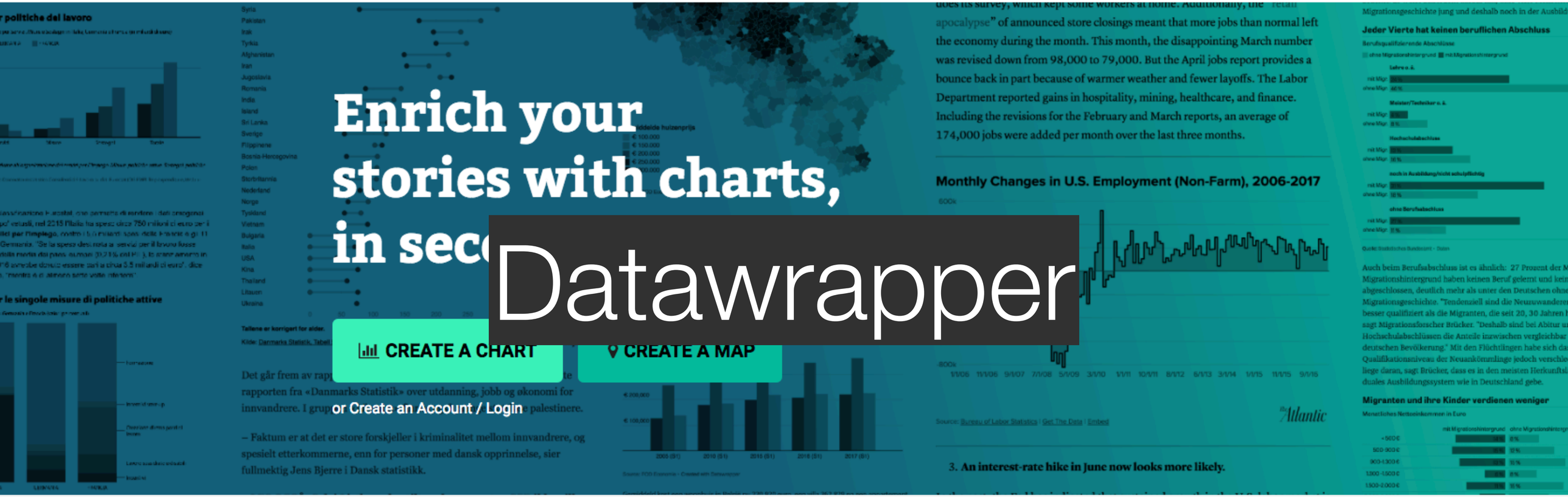
The world's most sophisticated editor for creating D3.js and WebGL charts.

No coding required.

GET STARTED

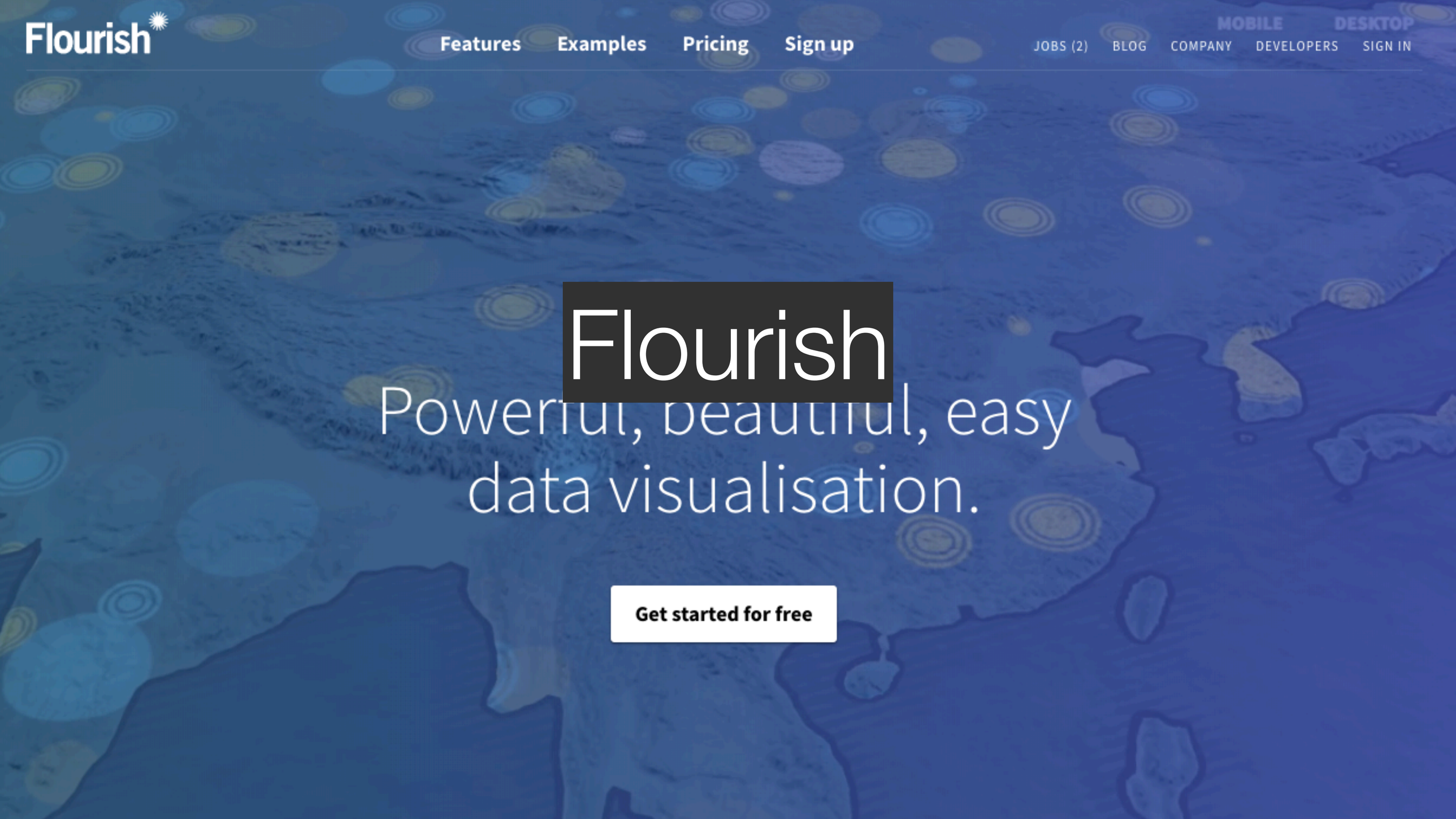


redhat. alteryx ComEd.



Datawrapper makes it easy to create beautiful charts.

See for yourself how different newsrooms use our charts:


Flourish
Features
Examples
Pricing
Sign up
MOBILE
DESKTOP
JOBS (2)
BLOG
COMPANY
DEVELOPERS
SIGN IN
Flourish
Powerful, beautiful, easy
data visualisation.
Get started for free

Turn your data into knowledge

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

Quadrigram

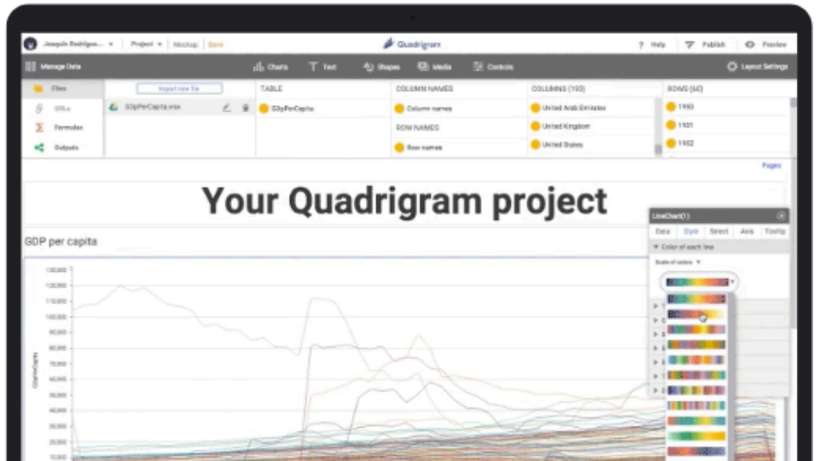
PURCHASE NOW

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.



RAWGraphs

The missing link between spreadsheets and data visualization.

USE IT NOW!

FORK IT ON GITHUB

Blogs

Follow

Visualization and Visual Communication

What's happening in information visualization? Robert Kosara digests new and emerging developments and explores what we know and don't know – so we can be less wrong about visualization.

Recent SEE ALL →

Spotting rip currents

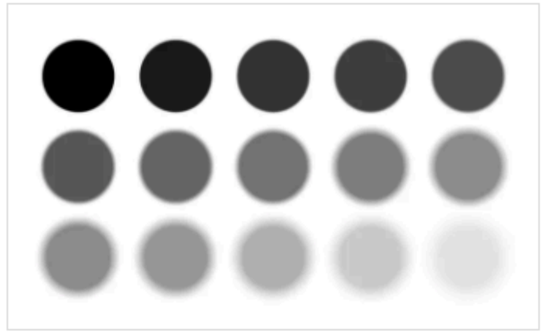
Rip currents are like hidden rivers near the shore that ...

Obliteration Room invited people to put dotted stickers everywhere

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

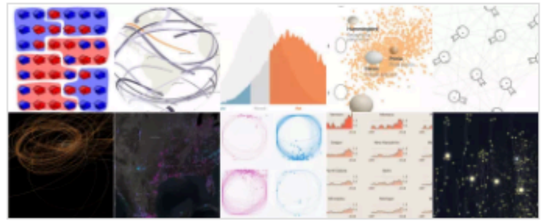
Where athletes in professional sports come from

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.

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

14 JAN THE LITTLE OF VISUALISATION DESIGN: PART 59 >>

18 DEC UPDATED DATA

18 DEC THE LITTLE OF

17 DEC BEST OF THE

14 DEC THE LITTLE OF

13 DEC ANNOUNCING THE

07 DEC BEST OF THE

07 DEC TRANSLATING 'DATA

06 DEC DATA STORIES

06 DEC VISUALISATION

28 NOV BEST OF THE

12 NOV POLICYVIZ PODCAST:

24 OCT LAUNCHING THE

23 OCT THE LITTLE OF

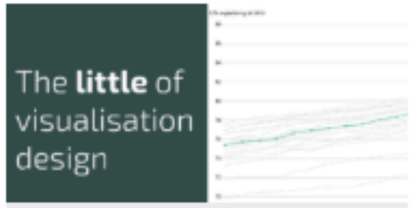
18 OCT BEST OF THE

10 OCT MILESTONE IN DATA

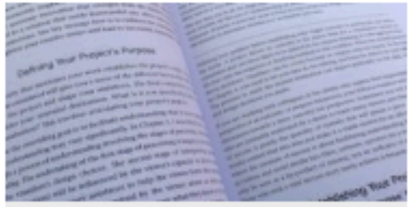
08 SEP BEST OF THE

03 SEP THE LITTLE OF

09 AUG 10 SIGNIFICANT



07 AUG THE LITTLE OF



07 AUG SHORT SURVEY ABOUT

31 JUL BEST OF THE

23 JUL FREE DATA

16 JUL THE LITTLE OF

11 JUL FINAL PUSH TO

Awards Challenges

All 2018 2017 2016 2015 2014 2013 2012

All Shortlist Longlist ★ Winners Gold Silver Bronze Rising Star Outstanding Individual Best-Non-English-Language Outstanding Outfit
Commercial Project Student Community Outstanding Team Most Beautiful

All Arts, Entertainment & Culture Breaking News Humanitarian Leisure, Games & Sport Maps, Places & Spaces People, Language & Identity
Politics & Global Visualization & Information Design Science & Technology Unusual Internal Business Project Paid tool Dataviz Website
Dataviz Project Mini Visualization Dataviz Tool

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!

Thank you!