09. Visualization Design Principles (Heuristics)

Venkatesh Rajamanickam (@venkatrajam)
venkatra@iitb.ac.in
http://info-design-lab.github.io/

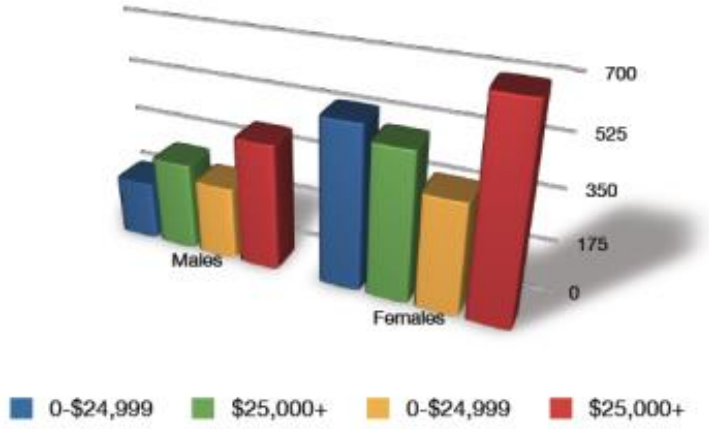
Tufte's design principles for graphical excellence

1. Maximize the data-ink ratio, within reason
2. Mobilize every graphical element, perhaps several times over, to show the data
3. Maximize data density and the size of the data matrix, within reason
4. Establish context
5. Show cause and effect, where possible
6. Compare and contrast, utilize layering & separation
7. Escape flatland, use small multiples, parallel sequencing (reality is multivariate)
8. Show multiple dimensions
9. Utilize narratives of space and time
10. Integrate image, number and text

1. Maximize the data-ink ratio, within reason

- Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented
- It is also the proportion of graphic's ink devoted to the non-redundant display of data-information.

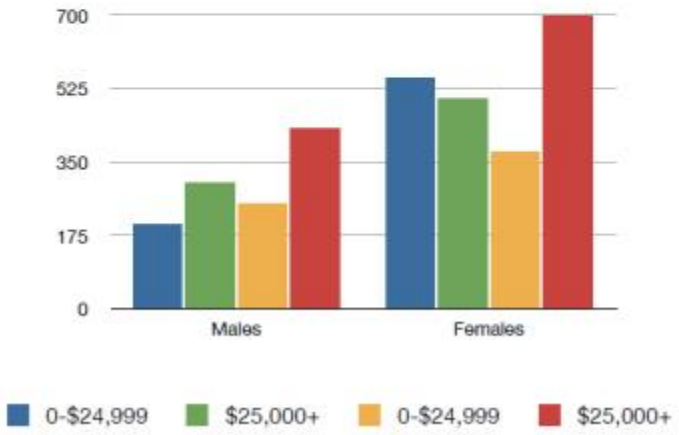
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



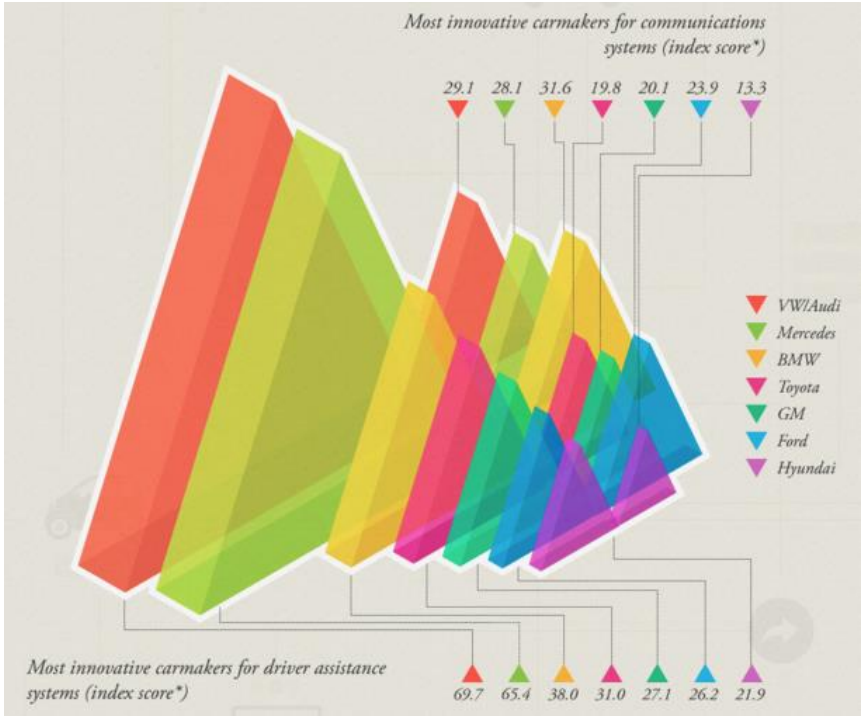
1. Maximize the data-ink ratio, within reason

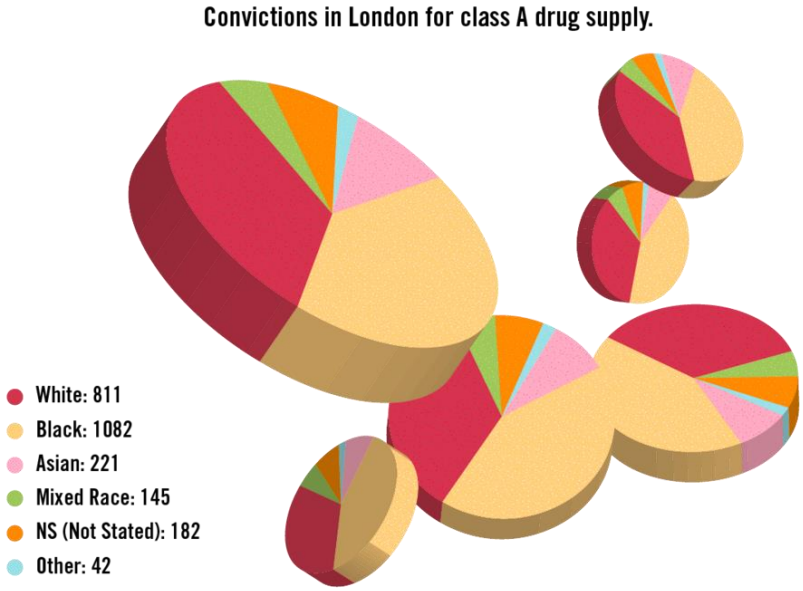
- This rule states that a visualization should contain as much data as possible while also using as little pixels as possible
- Through a comprehensive editing and testing process, any visualization can continually be improved upon

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



Unjustified 3D all too common, in the news and elsewhere



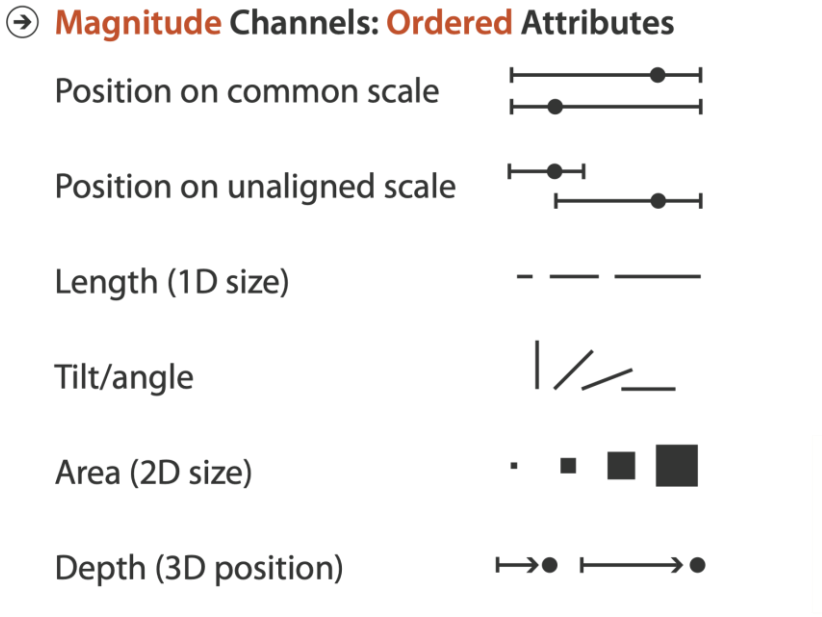


http://viz.wtf/post/137826497077/eye-popping-3d-triangles

http://viz.wtf/post/139002022202/designer-drugs-ht-ducqn

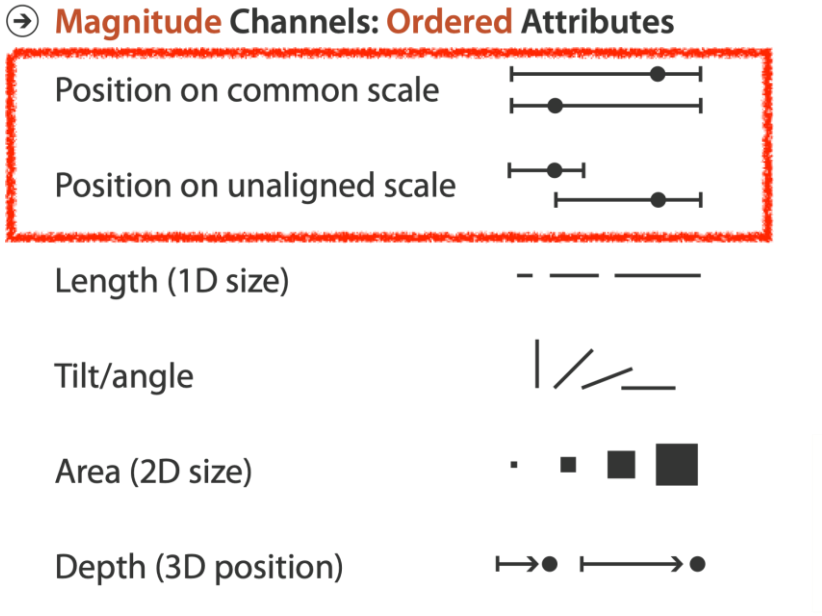
Depth vs power of the plane

- high-ranked spatial position channels: **planar** spatial position
  - not depth!



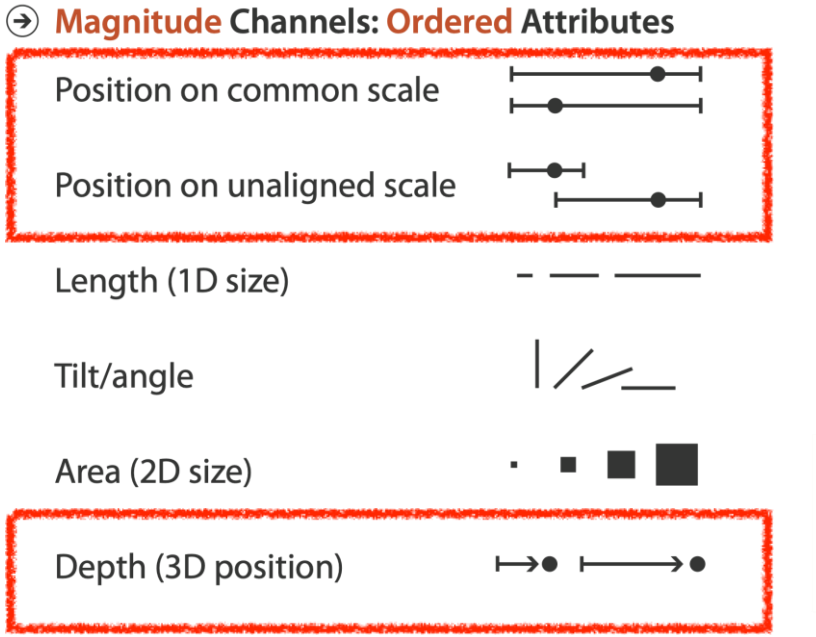
Depth vs power of the plane

- high-ranked spatial position channels: **planar** spatial position
  - not depth!



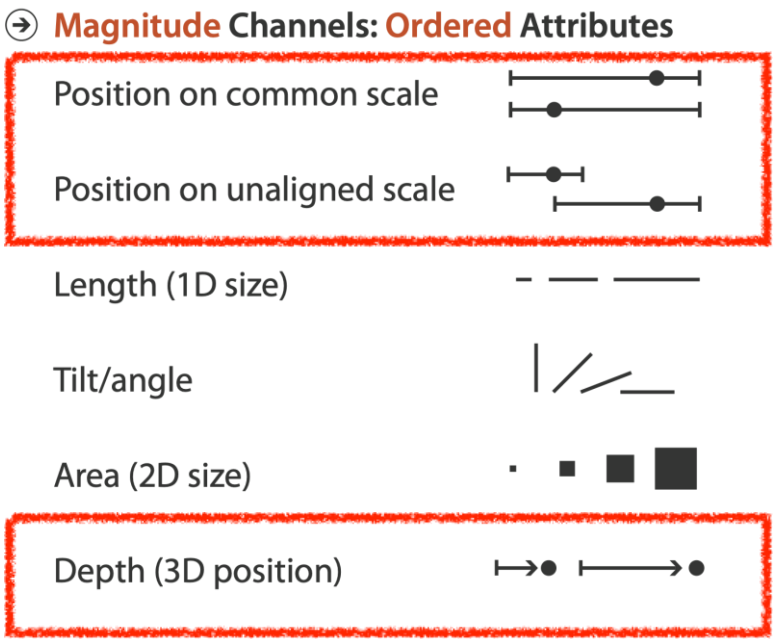
Depth vs power of the plane

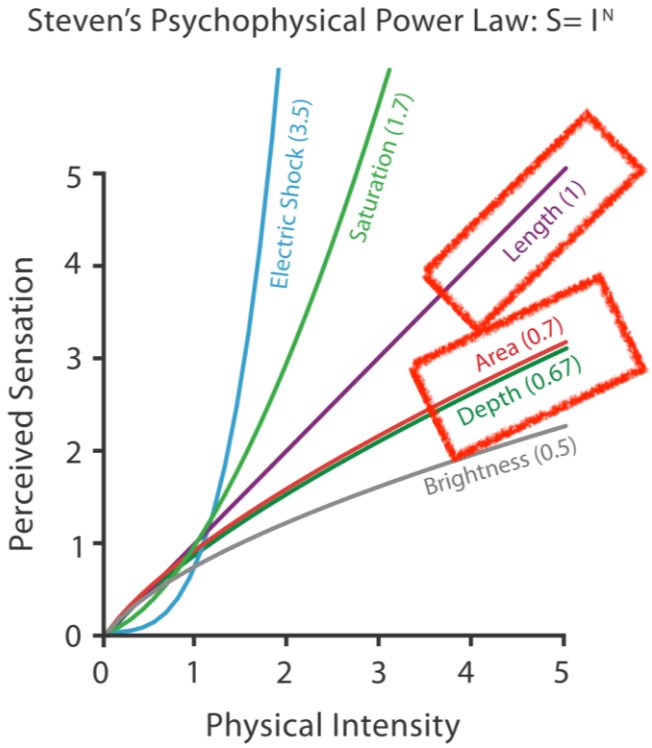
- high-ranked spatial position channels: **planar** spatial position
  - not depth!



Depth vs power of the plane

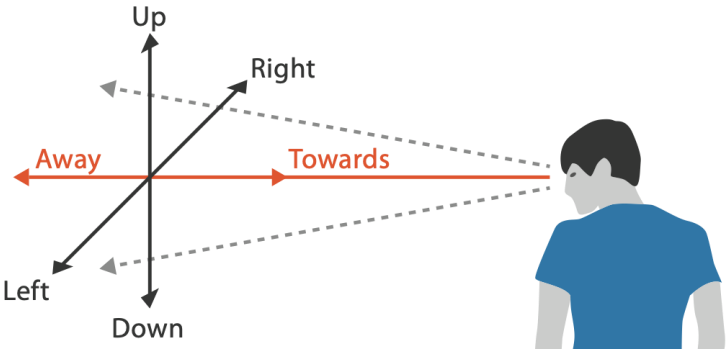
- high-ranked spatial position channels: **planar** spatial position
  - not depth!

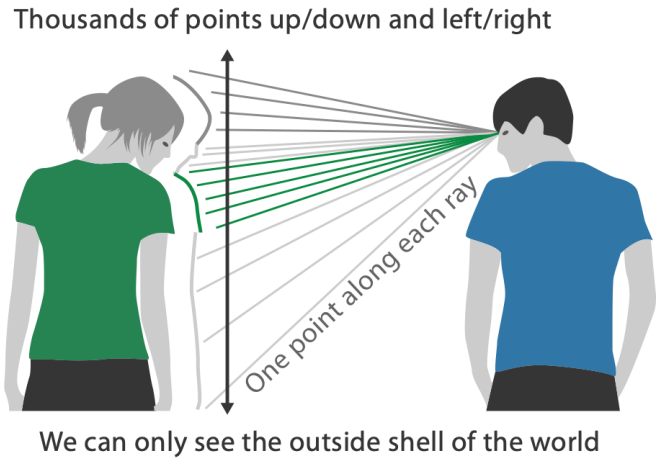




No unjustified 3D: Danger of depth

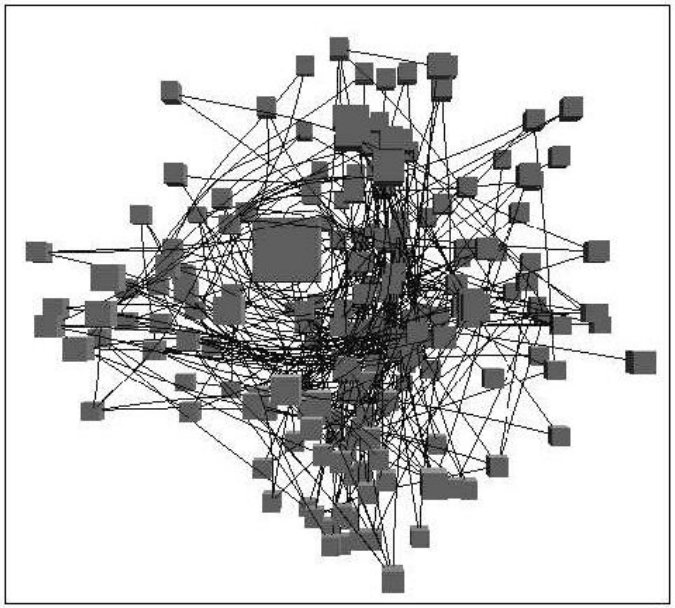
- we don't really live in 3D: we ***see*** in 2.05D
  - acquire more info on image plane quickly from eye movements
  - acquire more info for depth slower, from head/body motion





Occlusion hides information

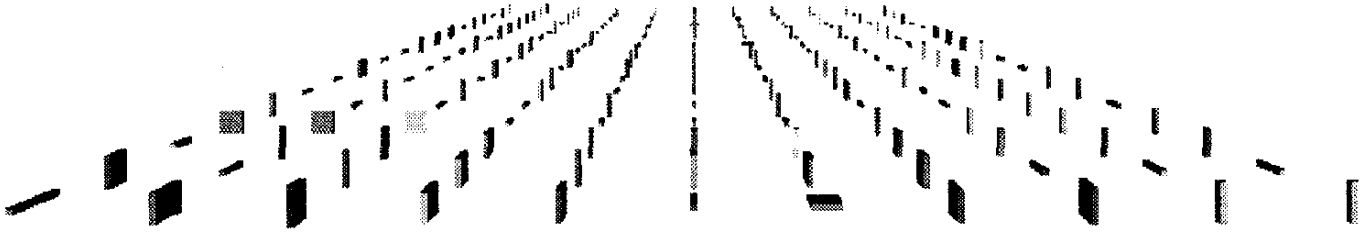
- occlusion
- interaction can resolve, but at cost of time and cognitive load



[Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.]

Perspective distortion loses information

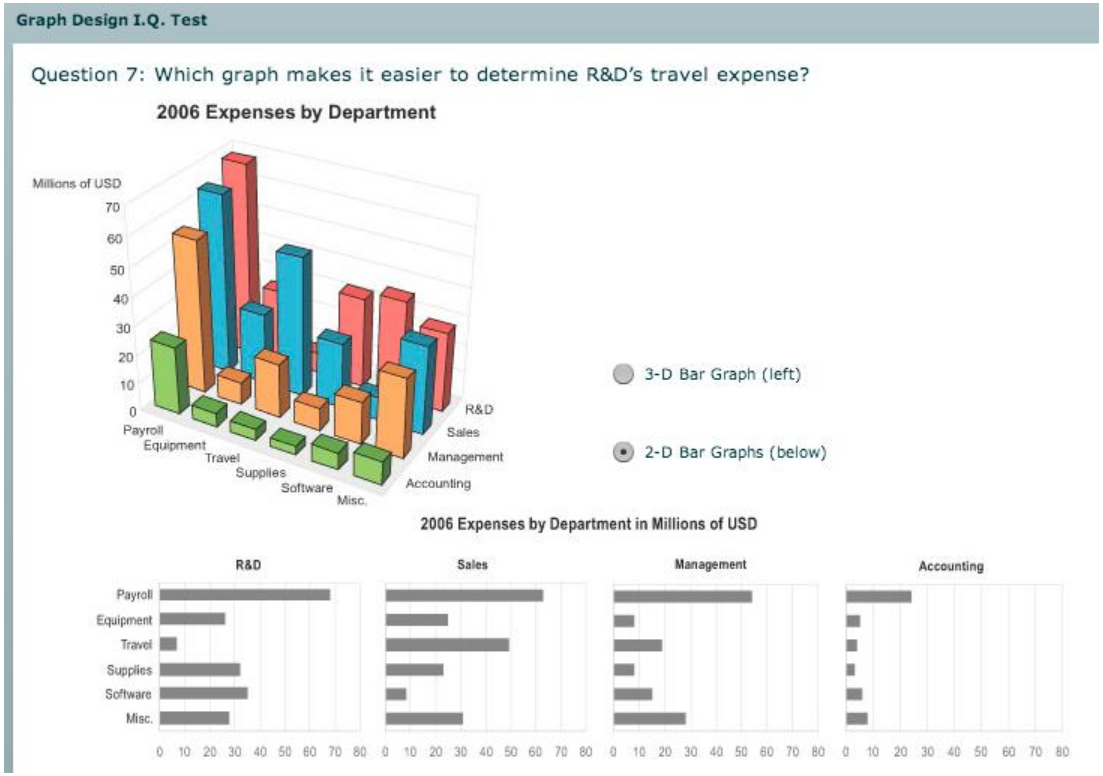
- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!



[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

3D vs 2D bar charts

- 3D bars:
  very difficult to justify!
  - perspective distortion
  - occlusion
- faceting into 2D almost always better choice



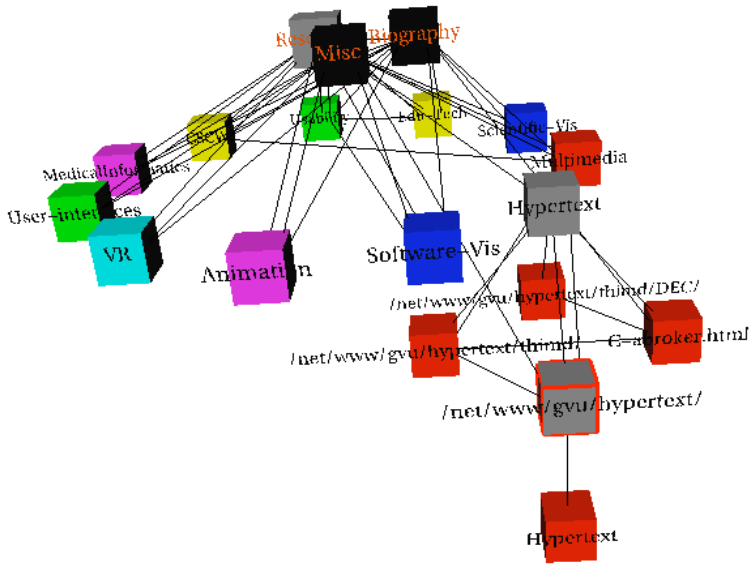
[http://perceptualedge.com/files/GraphDesignIQ.html]

Tilted text isn't legible

- text legibility
  - far worse when tilted from image plane

- further reading

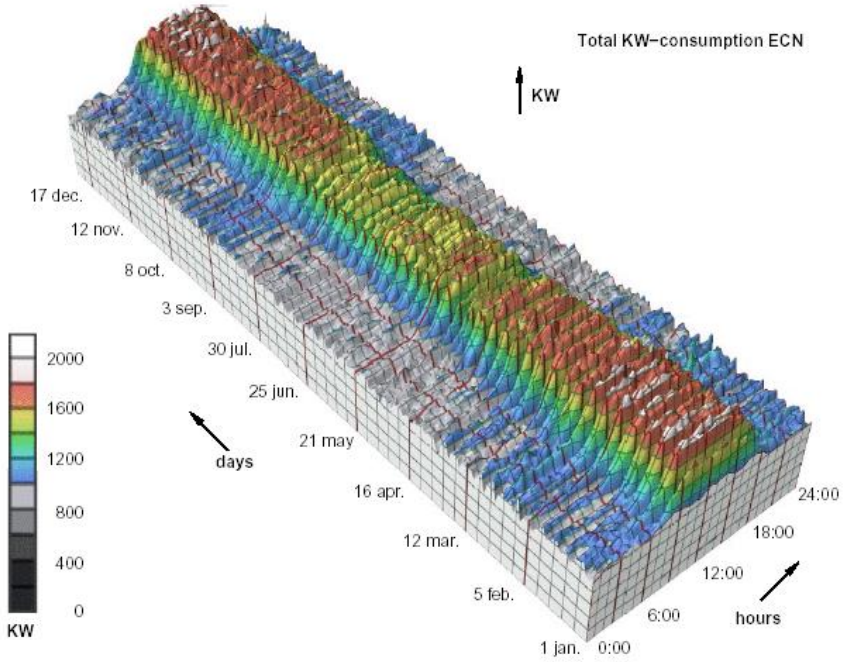
Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007



[*Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.*]

No unjustified 3D example: Time-series data

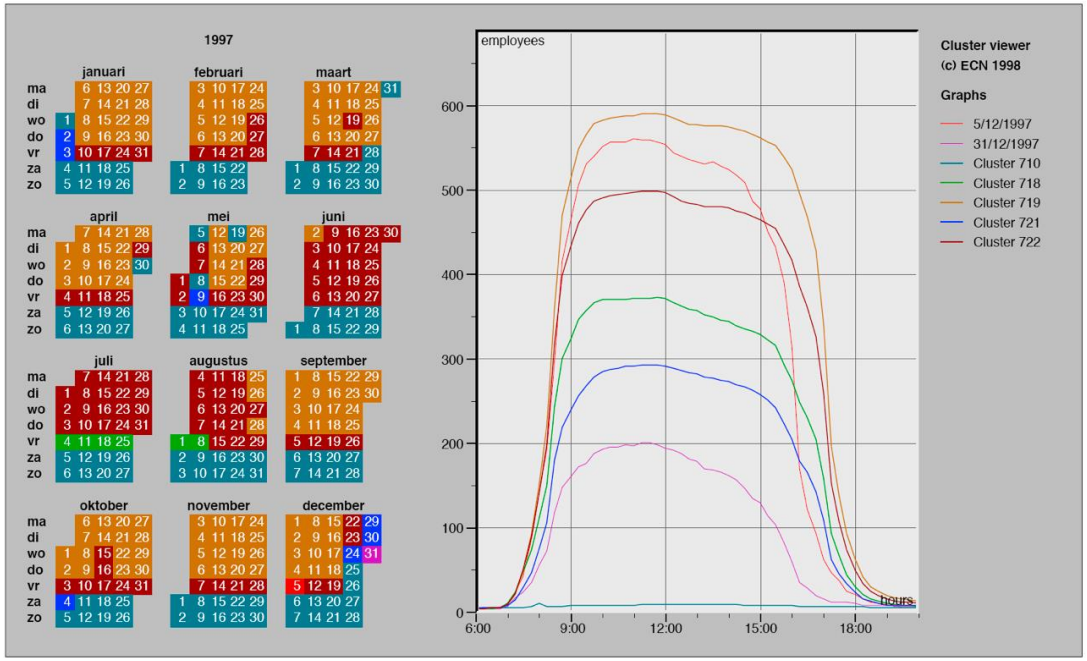
- extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

No unjustified 3D example: Transform for new data abstraction

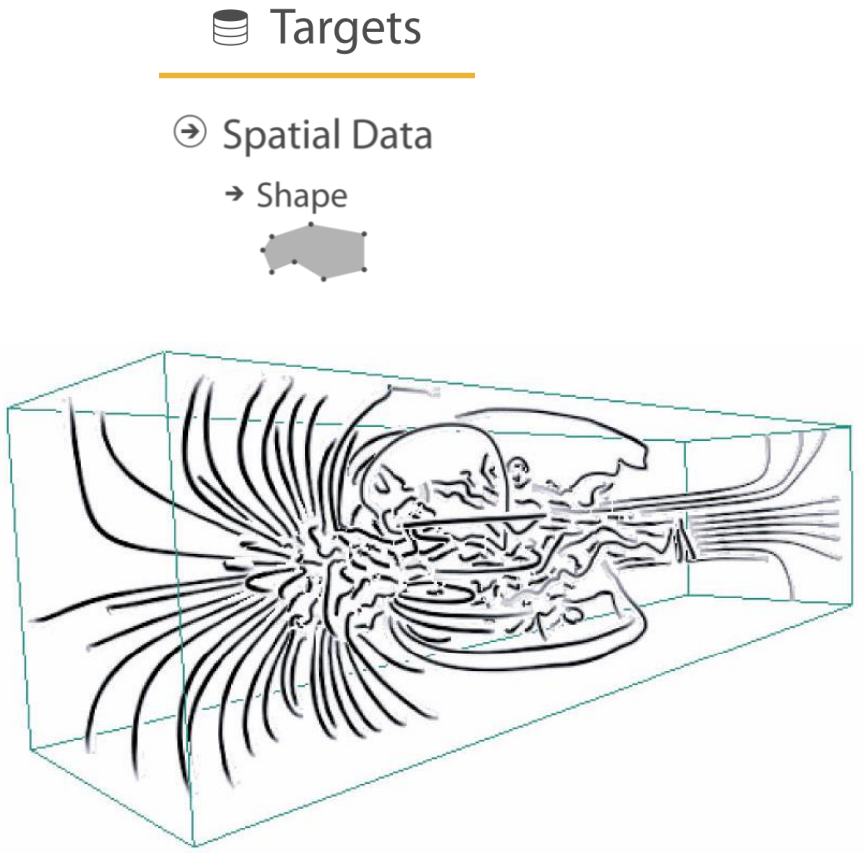
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Justified 3D: shape perception

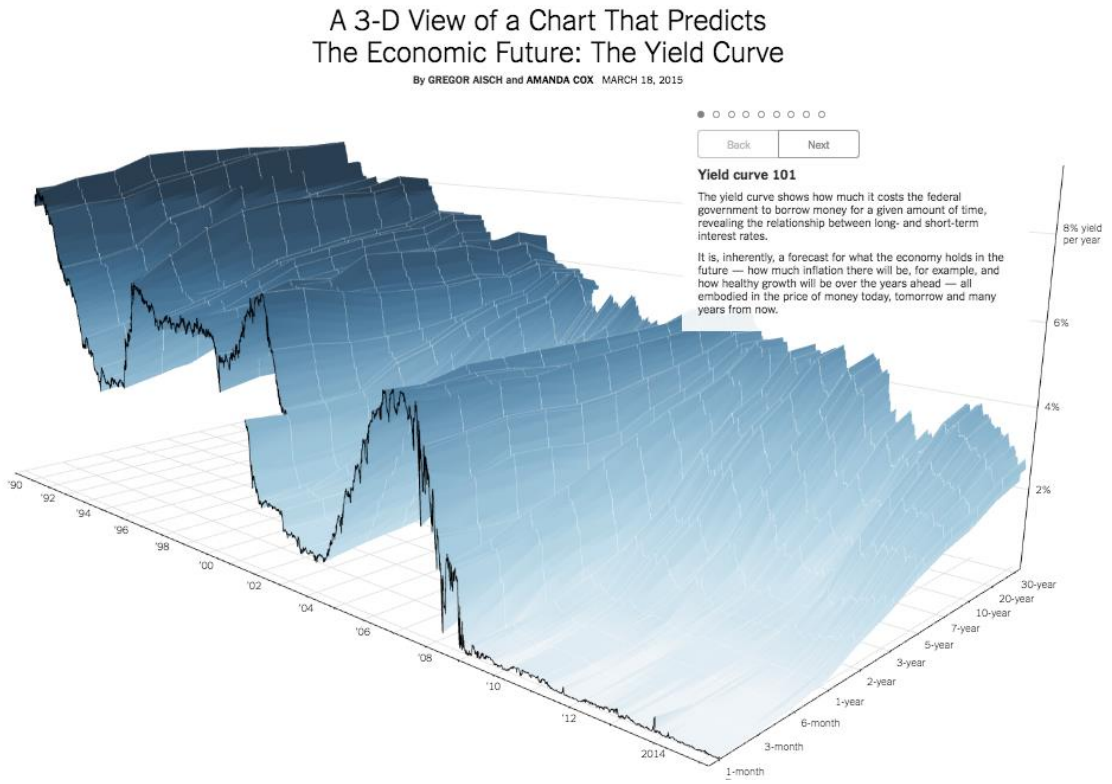
- benefits outweigh costs when task is shape perception for 3D spatial data
  - interactive navigation supports synthesis across many viewpoints



[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630–640.]

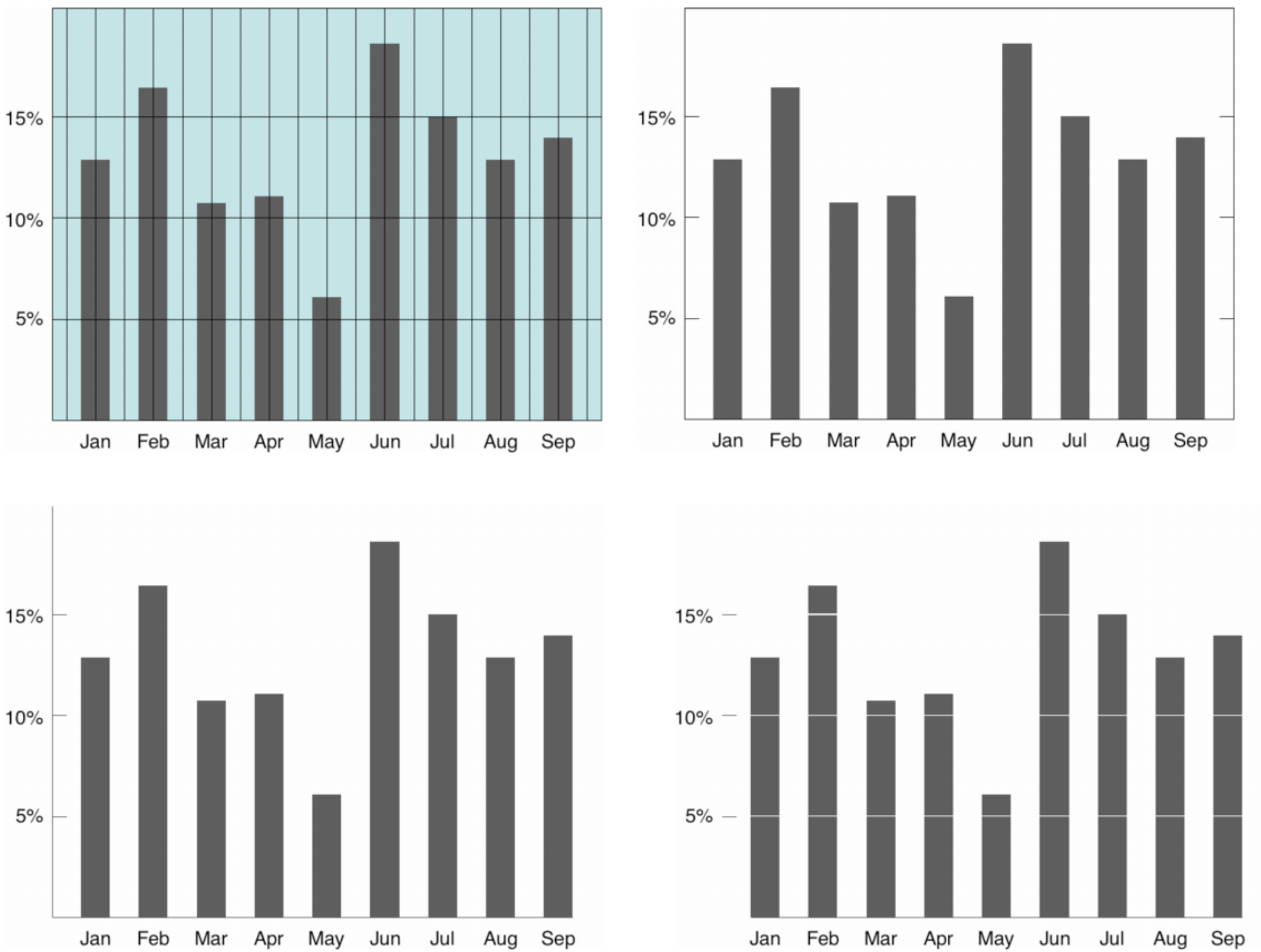
Justified 3D: Economic growth curve

- constrained navigation steps through carefully designed viewpoints

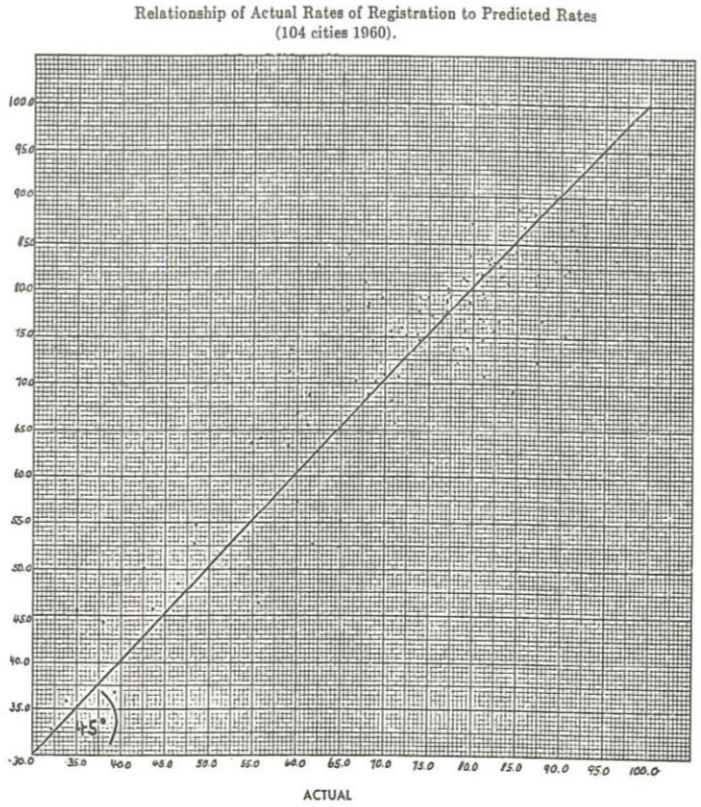


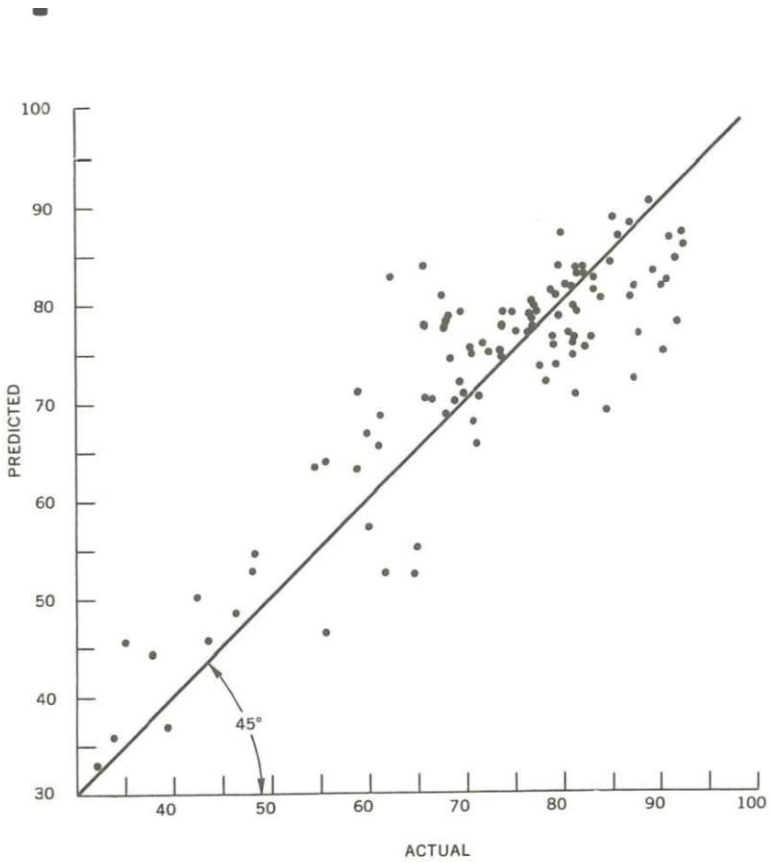
http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

1. Maximize the data-ink ratio, within reason

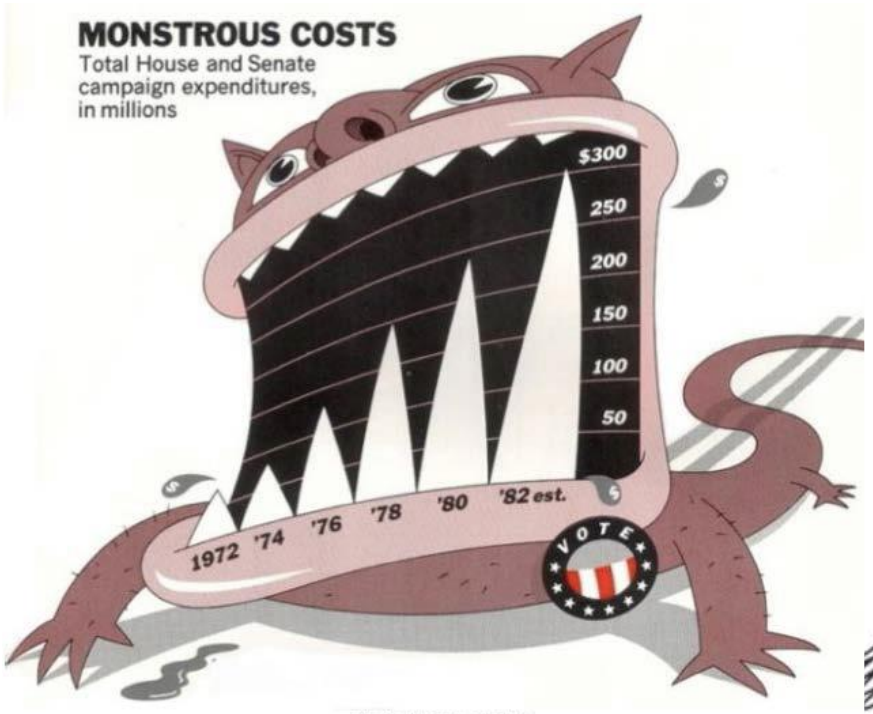


1. Maximize the data-ink ratio, within reason

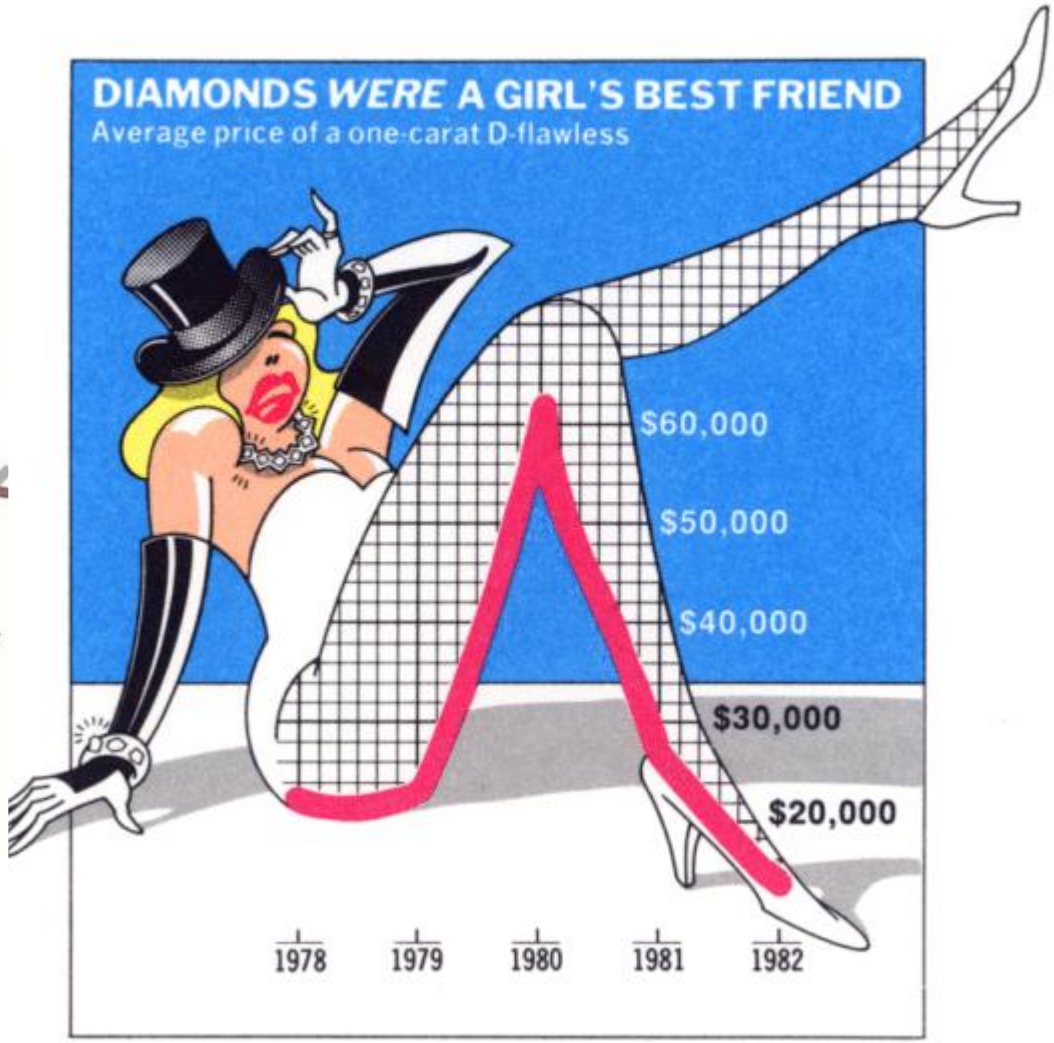




1. Maximize the data-ink ratio, within reason

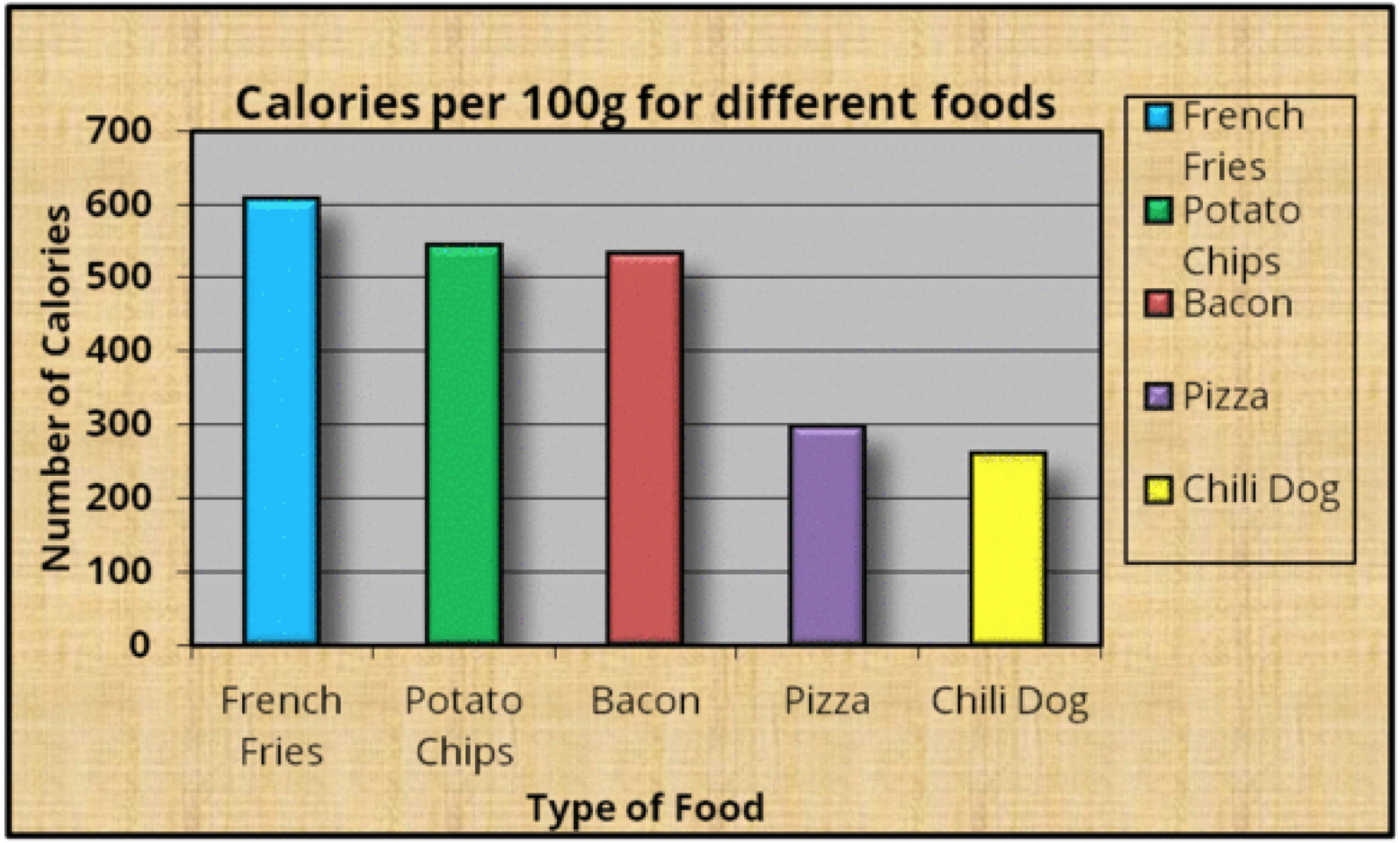


MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

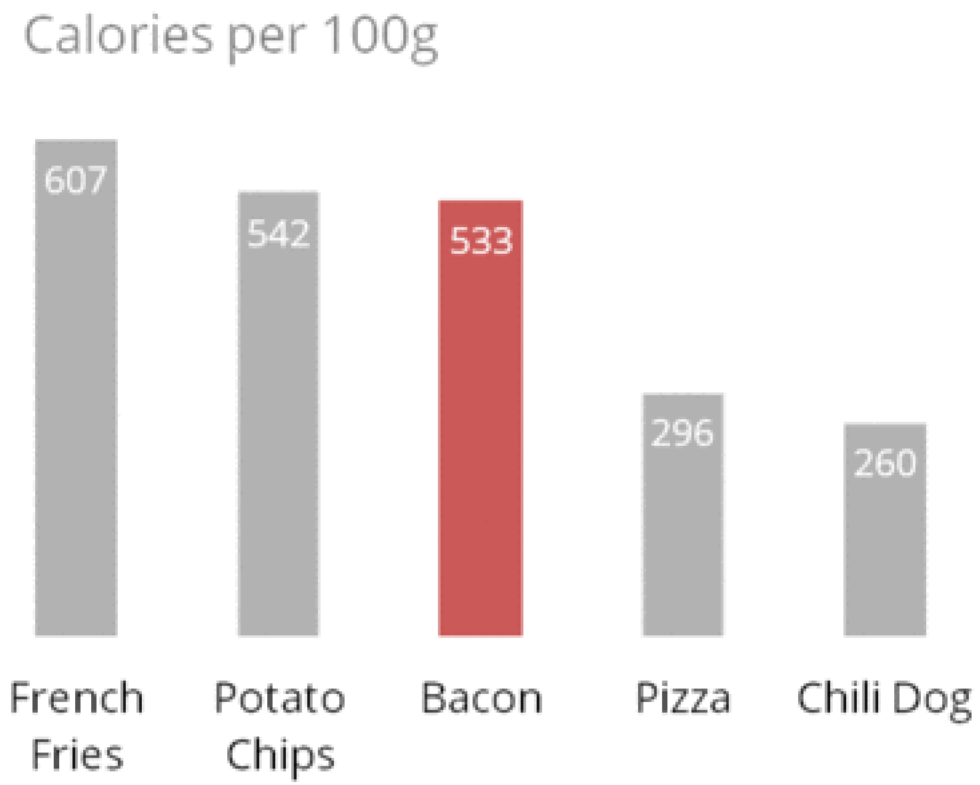


Nigel Holmes

1. Maximize the data-ink ratio, within reason

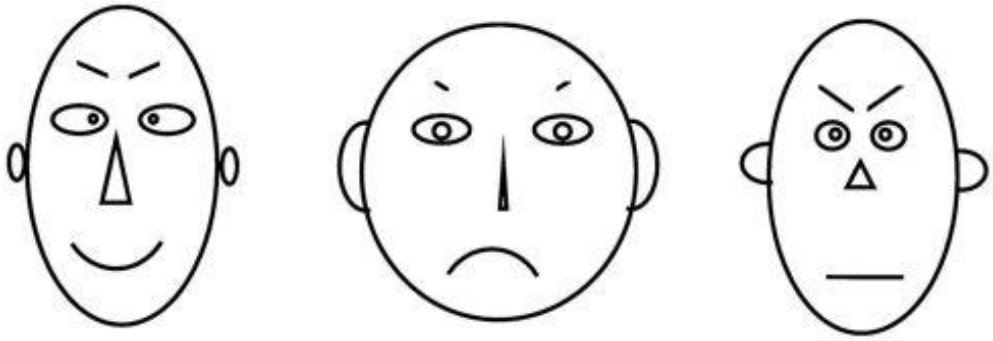


1. Maximize the data-ink ratio, within reason



2. Mobilize every graphical element to show the data

- The danger of multifunctioning elements is that they tend to generate graphical puzzles, with encodings that can only be broken by their designer.
- Thus design techniques for enhancing graphical clarity in the face of complexity must be developed along with multifunctioning elements.
- In other words, we should try to make all present graphical elements data encoding elements. We must make every graphical element effective.



2. Mobilize every graphical element to show the data

- Invented by applied mathematician, statistician and physicist Herman Chernoff in 1973, display multivariate data in the shape of a human face.
- The individual parts, such as eyes, ears, mouth and nose represent values of the variables by their shape, size, placement and orientation.
- The key idea is that humans are well trained to recognize faces and discern small changes without difficulty. They allow for easy outlier detection and pattern recognition despite multiple dimensions of the data.
- Faces are also emotionally expressive, and thus could be used to semantically resonant encode data.

2. Mobilize every graphical element to show the data

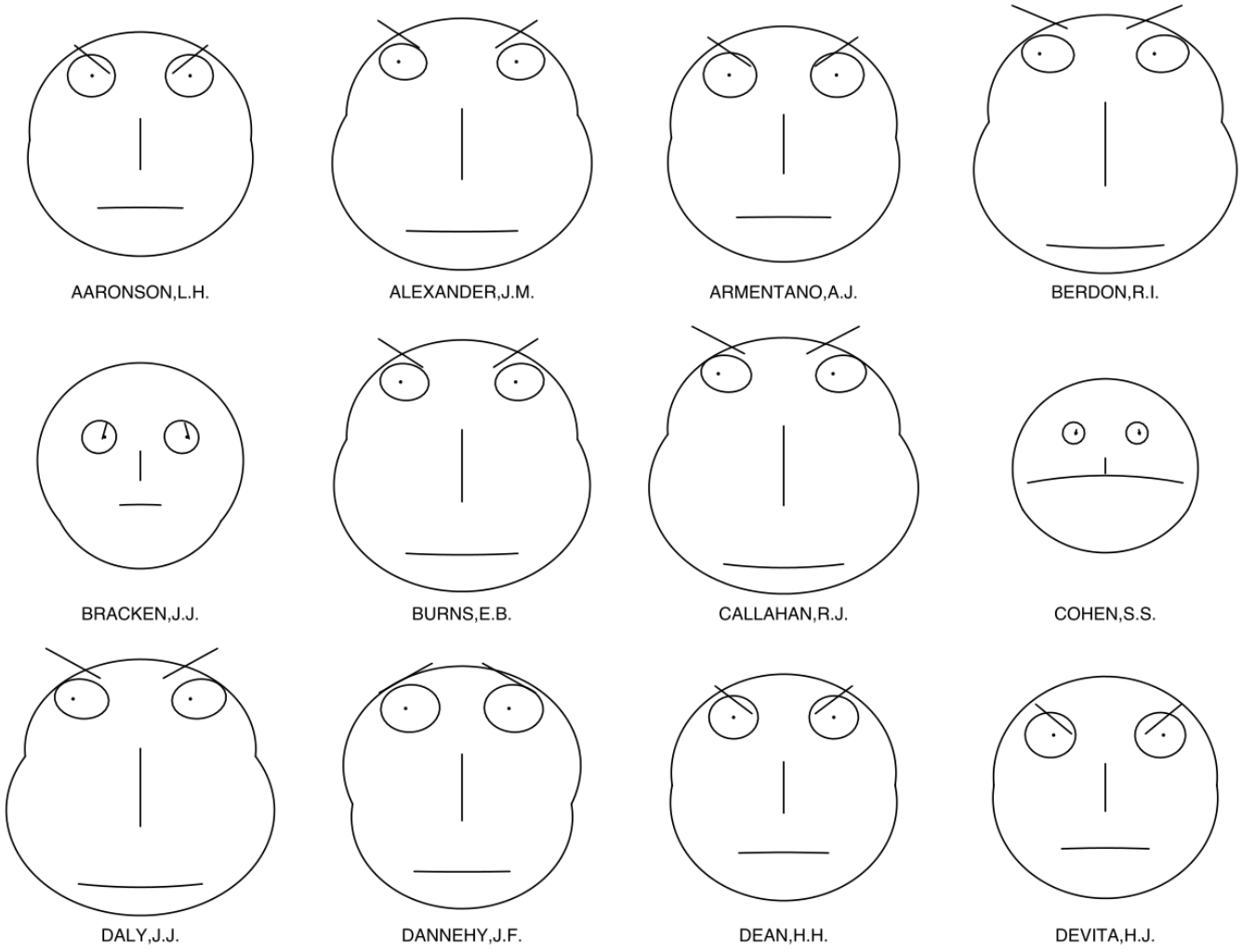
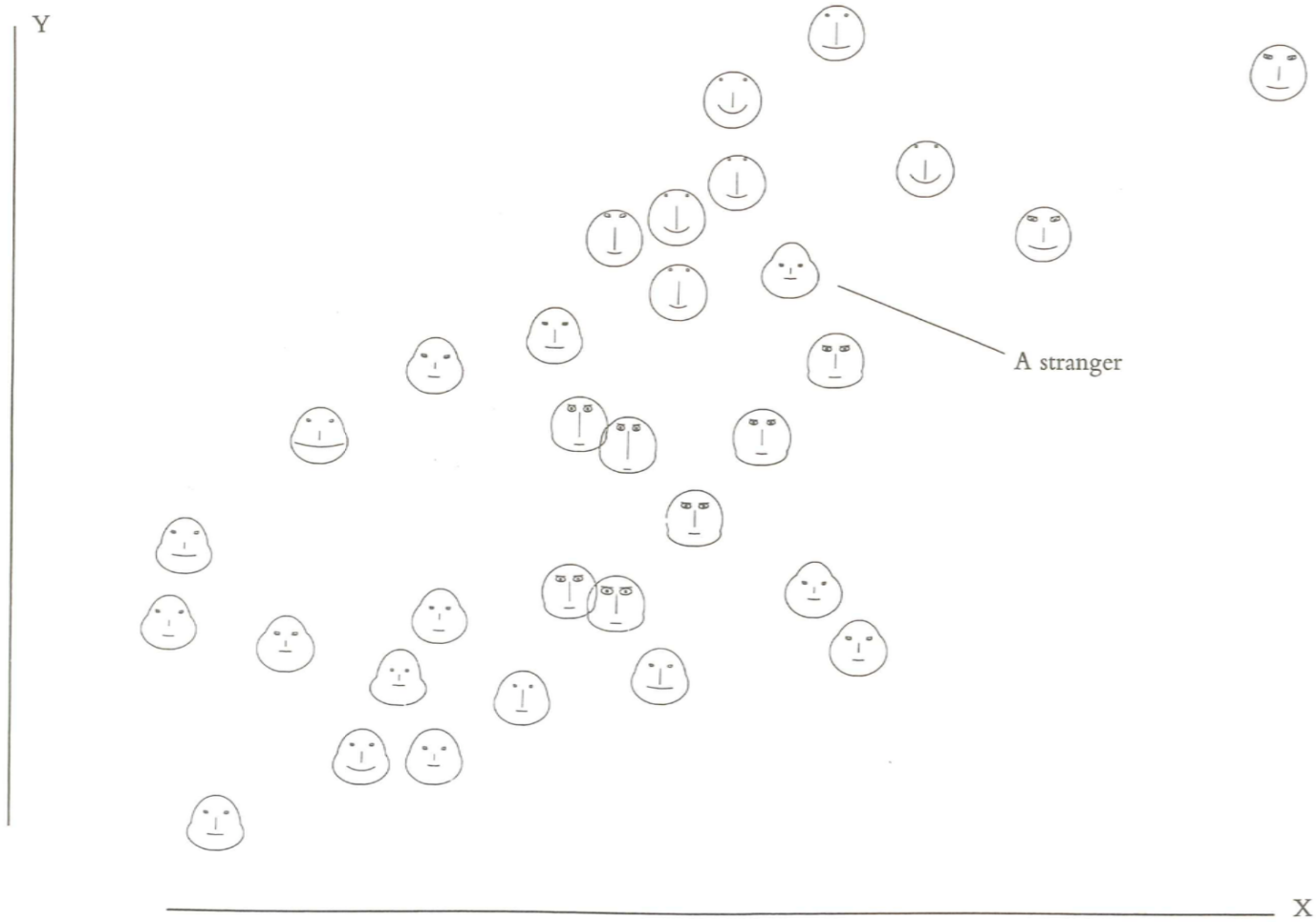


Chart showing Chernoff faces for data selected from a dataset, which contains ratings of state judges in the US Superior Court by lawyers who have had contact with them.

2. Mobilize every graphical element to show the data

Chernoff Faces

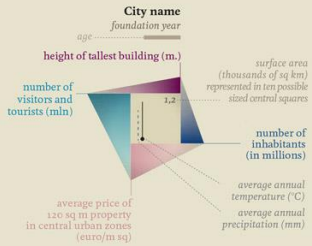


Urban story: Lisbon is on a par with Honolulu

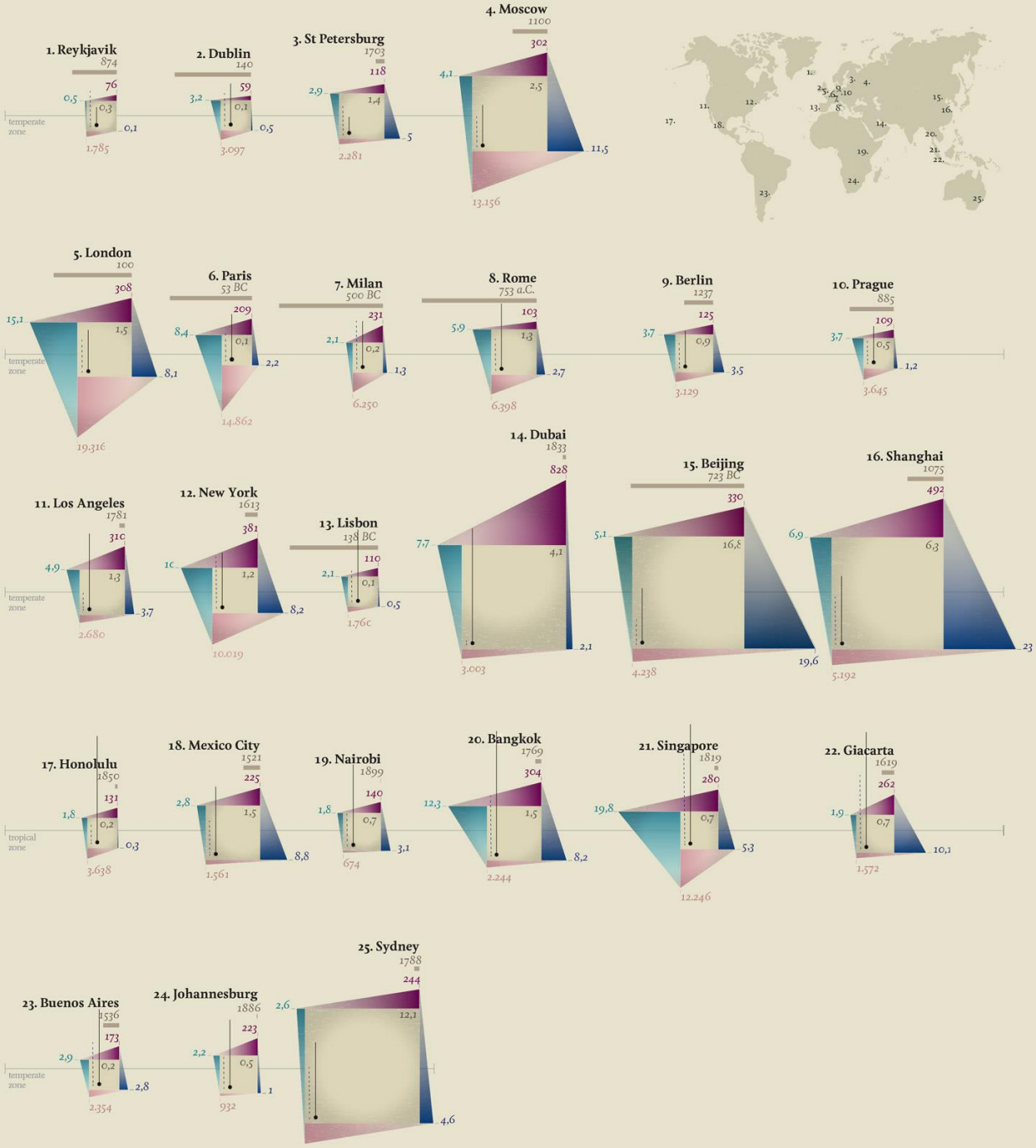
Among 25 of the world's most important cities are visualised here. Each metropolis has been grouped vertically by 5 main temperature zones according to their latitude; and ordered horizontally according to their longitude. Each city identity is visually represented in a multiform polygon correlating to its size in sq m, number of inhabitants, number of tourists per year, height of the tallest building, and the average price of property according to real estate parameters (square metres). Further information is provided for each city "id-card": age of the city, average temperature and rainfall. We discover that latitude and longitude parameters not only determine environmental factors but also several urban characteristics. Surprisingly Lisbon and Honolulu have a lot in common!

Sources:
city-data.com, currentresults.com, euromonitor.com, globalpropertyguide.com, skyscrapercenter.com, weatherbase.com, wikipedia.org.

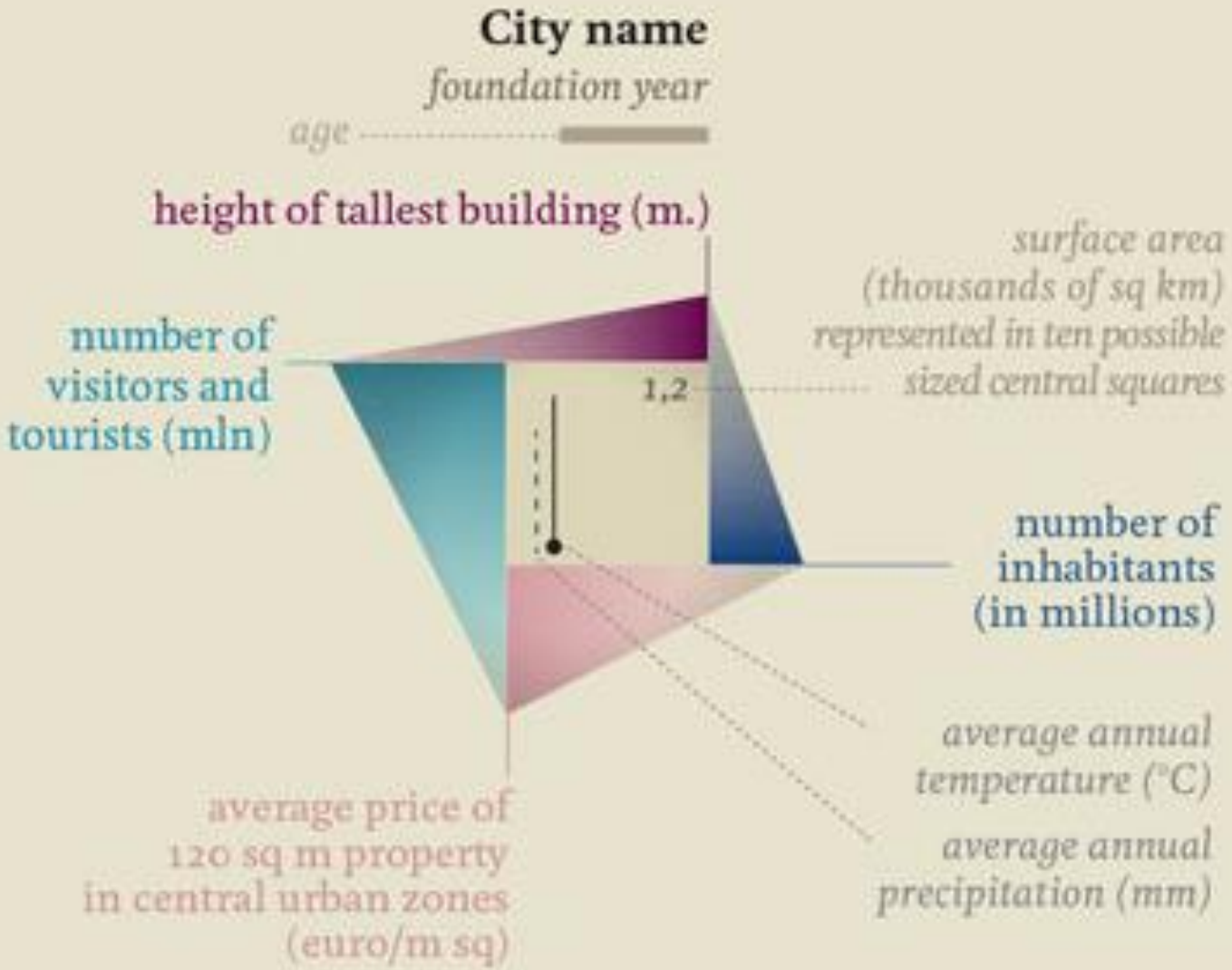
How to read it?



The visualization has been designed and produced by Accurat (www.accurat.it), and was originally published in italian on La Lettura the sunday cultural supplement of Corriere della Sera.



How to read it?



The Pritzker Architecture Prize

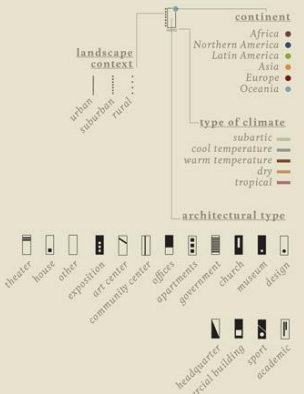
The visualization explores the Pritzker Prize assignment evolution, since 1979 to 2013. For each winner we visualized biographical information such as the continent of origin, the year of birth and death, the prize-awarding year and the age at the moment of the awards. Visualized are also the body of work selected by pritzkerprize.com per each architect, representing the architectural type, the year, the landscape context, the continent and the type of climate.

Sources: archiplanet.org, ees.rochester.edu, greatbuildings.com, pritzkerprize.com

How to read it?

Architects are vertically ordered by age, from the oldest to the youngest

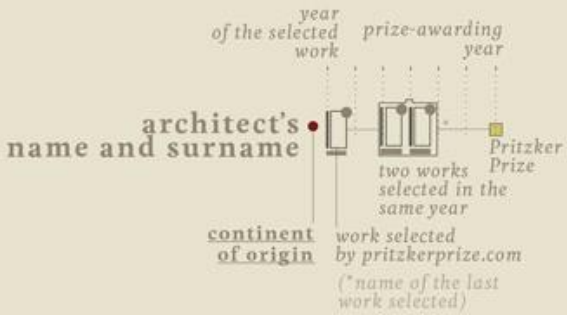
selected work:



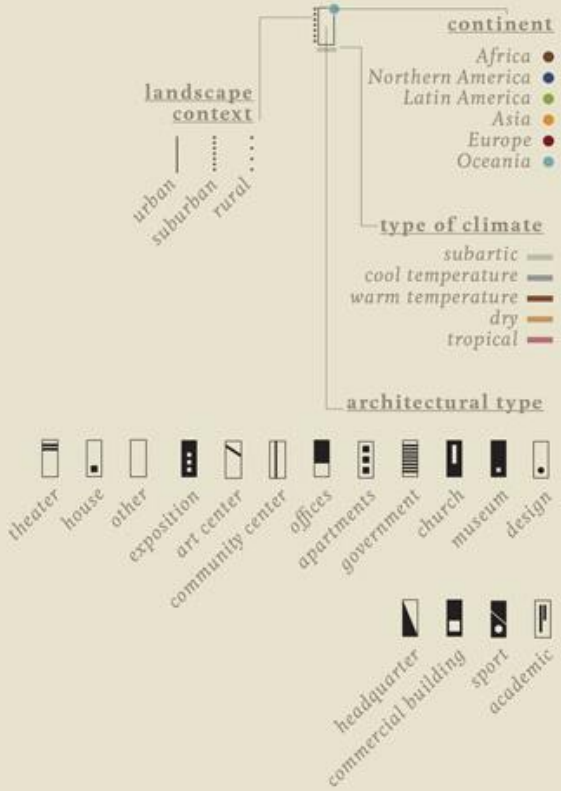
The visualization has been designed and produced by Accurat (www.accurat.it), and was originally published in Italian on La Lettura the sunday cultural supplement of Corriere della Sera.

How to read it?

Architects are vertically ordered by age, from the oldest to the youngest



selected work:



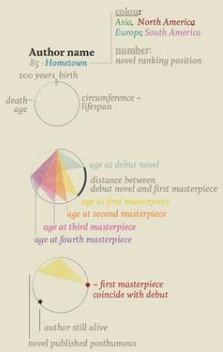
From first published to masterpieces

Visualized are the authors of the 100 best english novels of 1900 according to the Modern Library ranking.
Authors are represented through circles showing their life span where their debut novel, their masterpiece(s) according to the raking are positioned.
Authors' order points out the time-gap between the debut novel and the first published novel selected from the Modern Library ranking (masterpiece).
Debut novel and first masterpiece correspond for the first 22 authors.
Information about authors' hometown and novels' ranking are reported as well.

Sources:
biography.com, britannica.com, modernlibrary.com

How to read it?

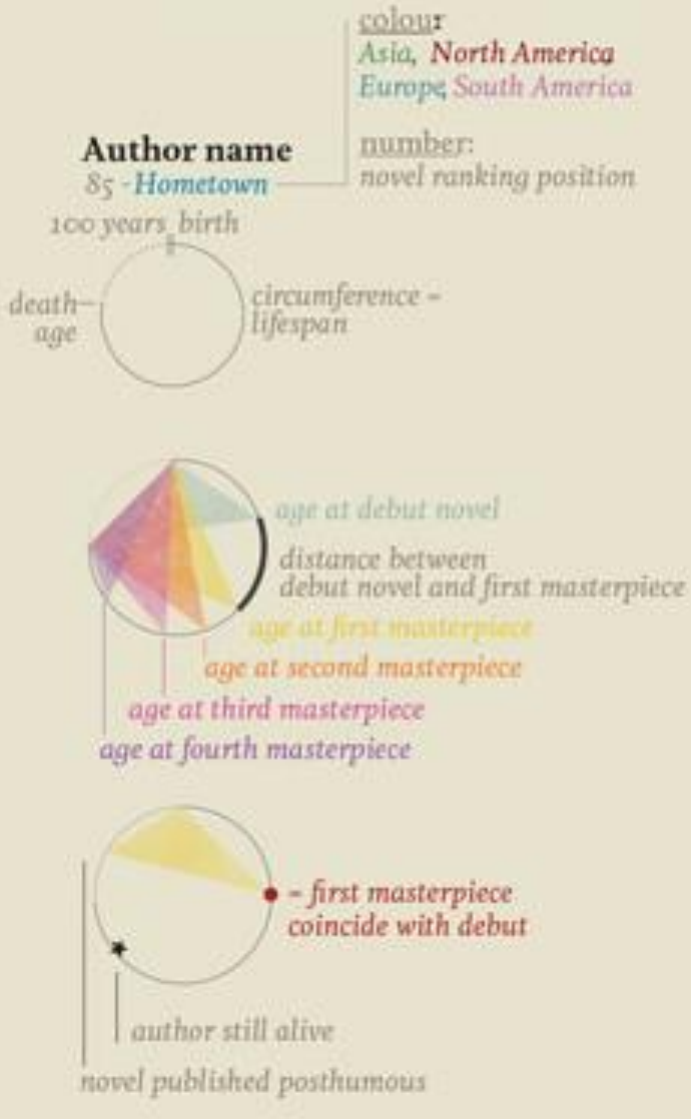
Authors are ordered from the earliest success to the last one.



The visualization has been designed and produced by Accurat (www.accurat.it), and was originally published in italian on La Lettura the sunday cultural supplement of Corriere della Sera.

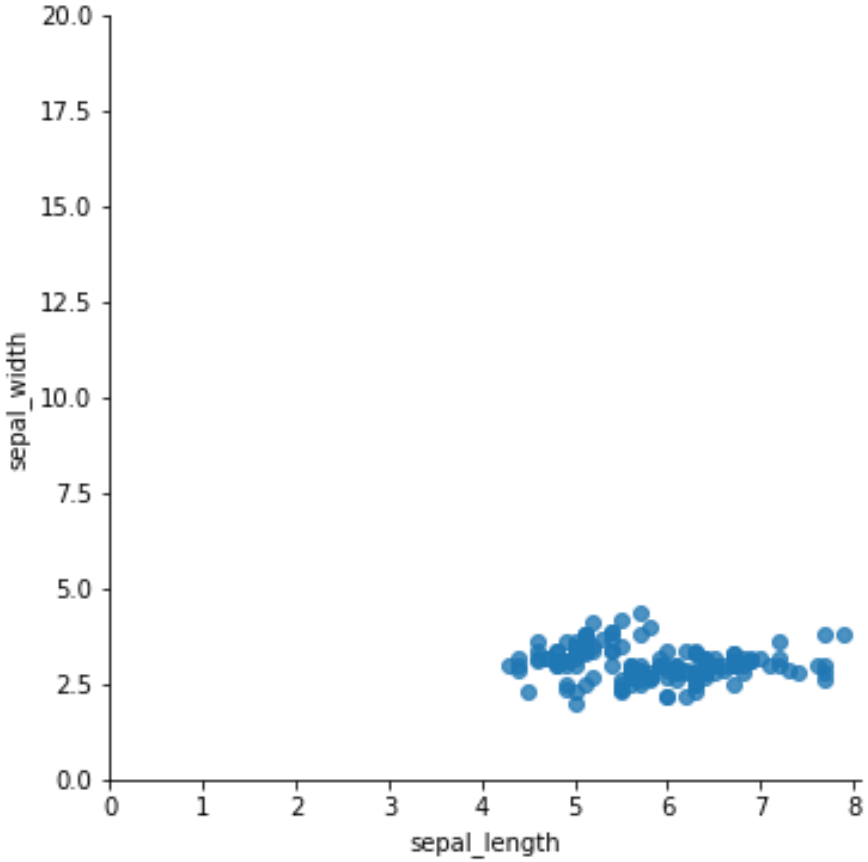
How to read it?

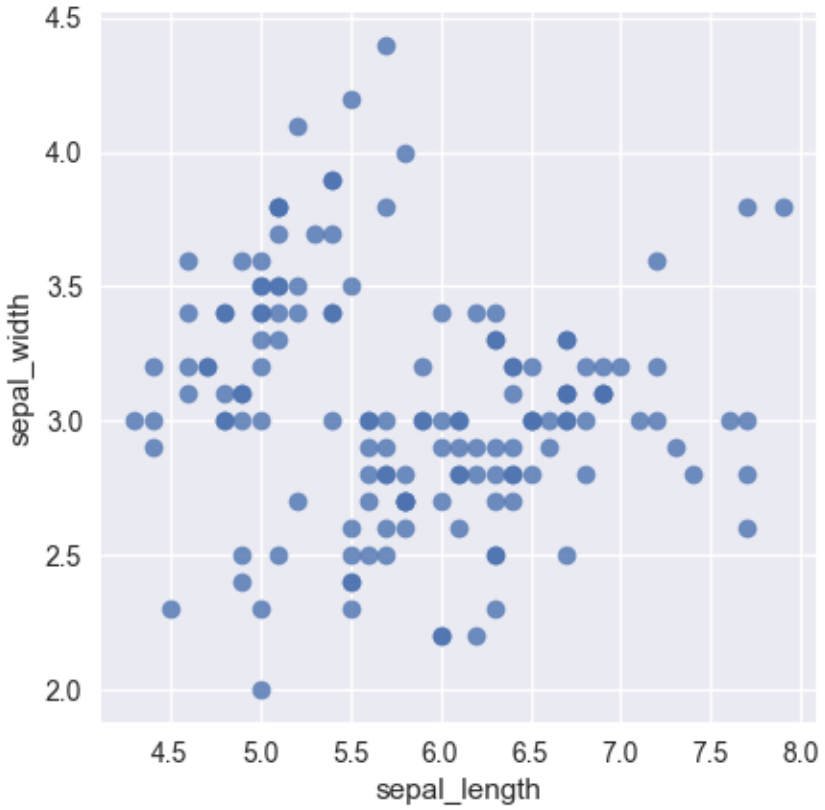
Authors are ordered from
the earliest success to the last one.



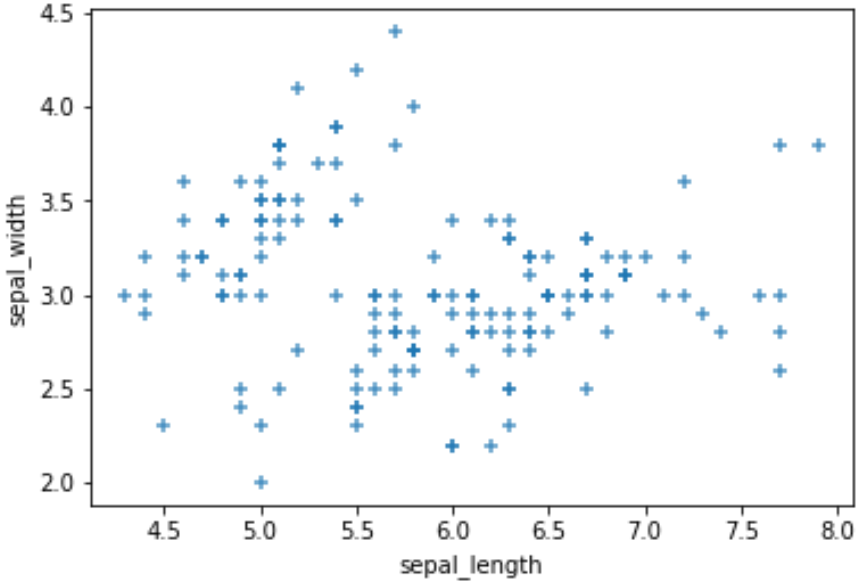
3. Maximize data density and the size of the data matrix, within reason

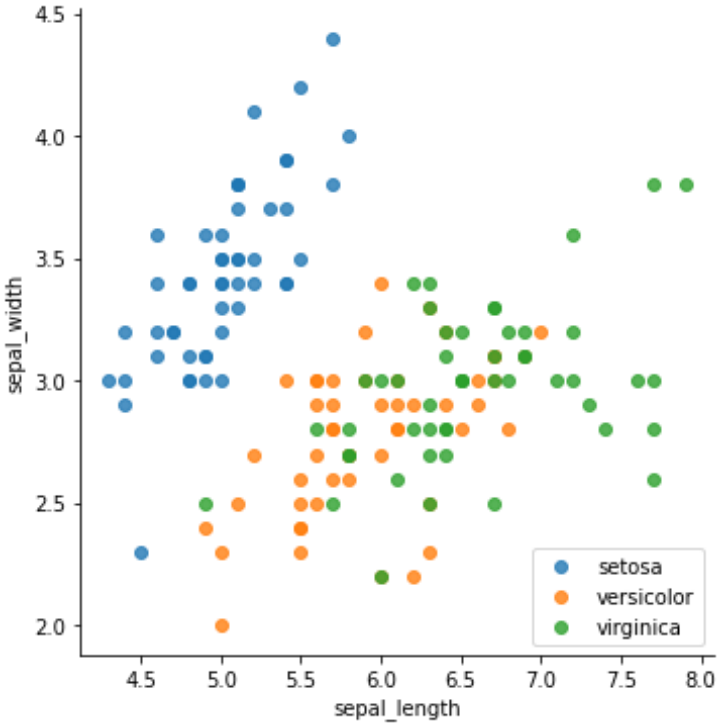
- High preformation graphics should be designed with special care. As the volume of data increases, data measures must shrink (smaller dots for scatters, thinner lines for busy time-series).



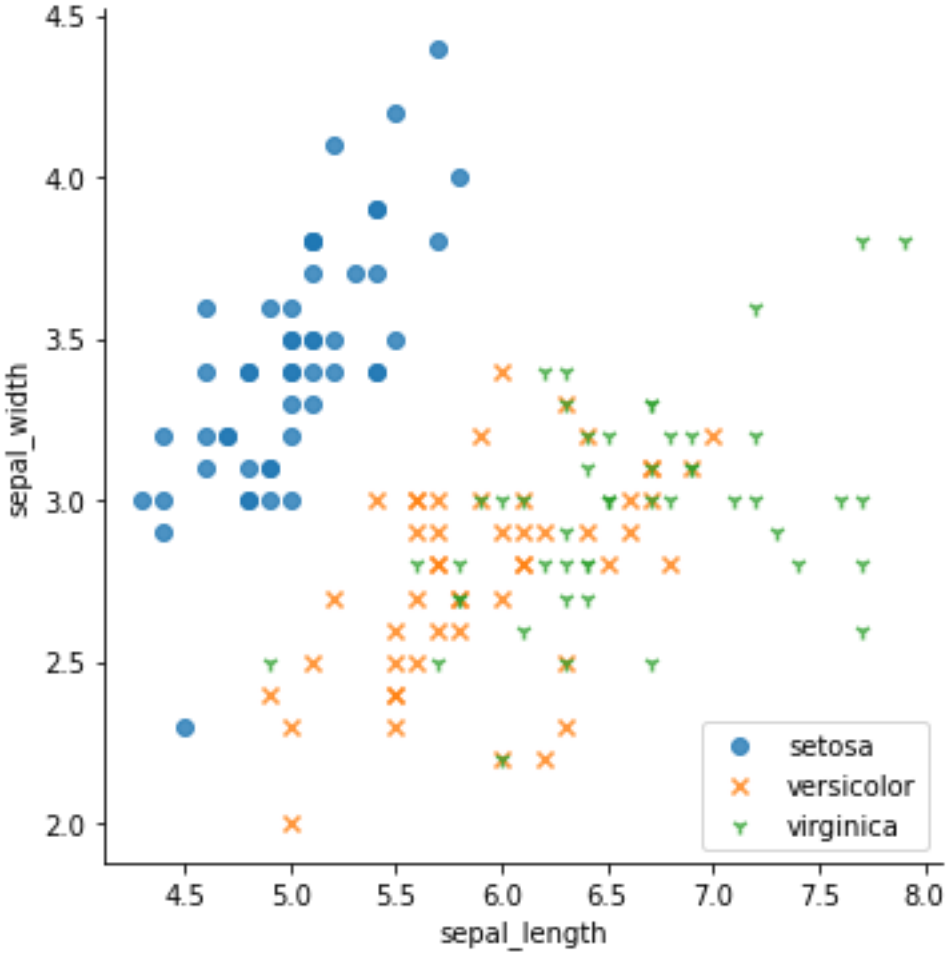


3. Maximize data density and the size of the data matrix, within reason



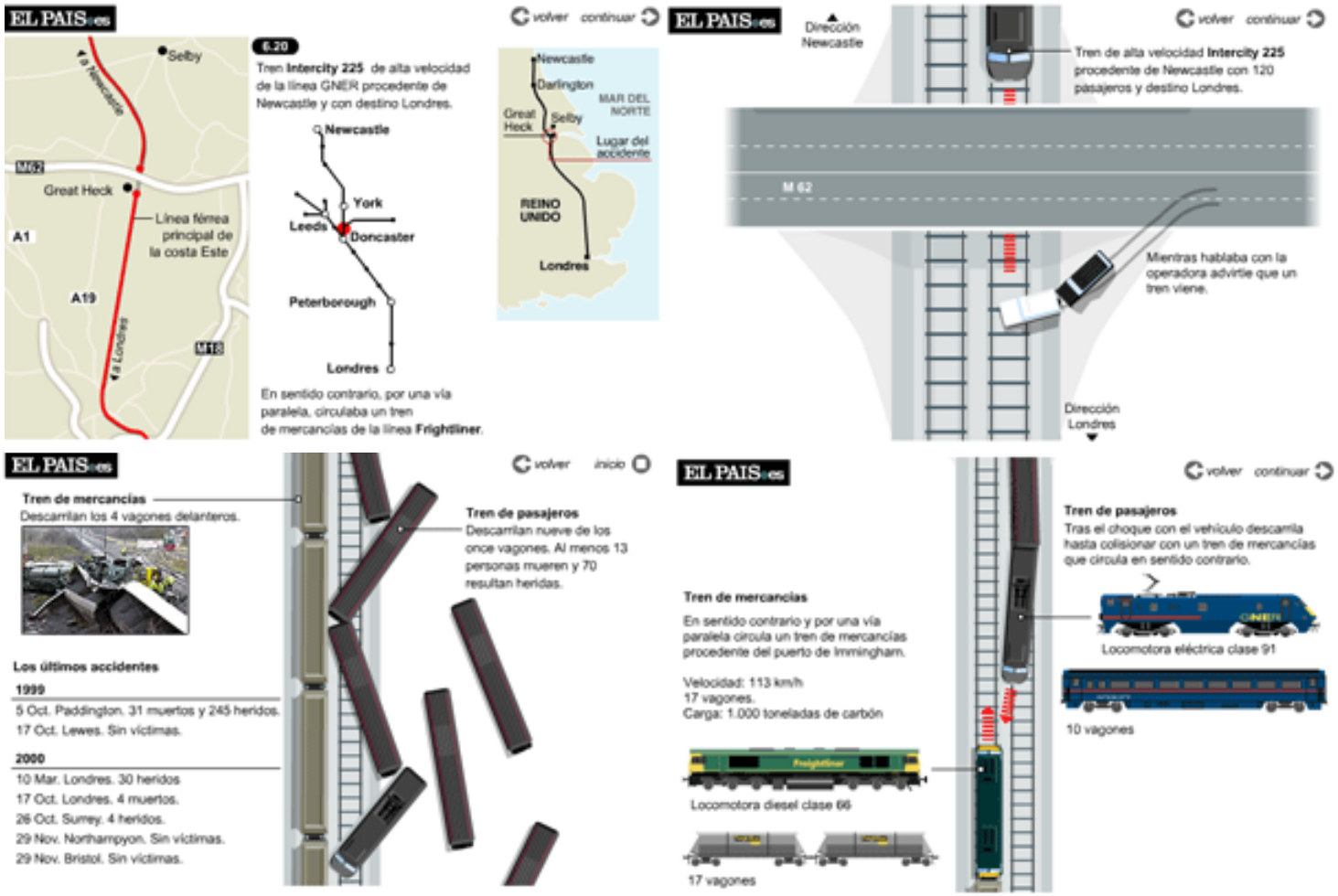


3. Maximize data density and the size of the data matrix, within reason



4. Establish Context

- Provide spatial, temporal and numerical context

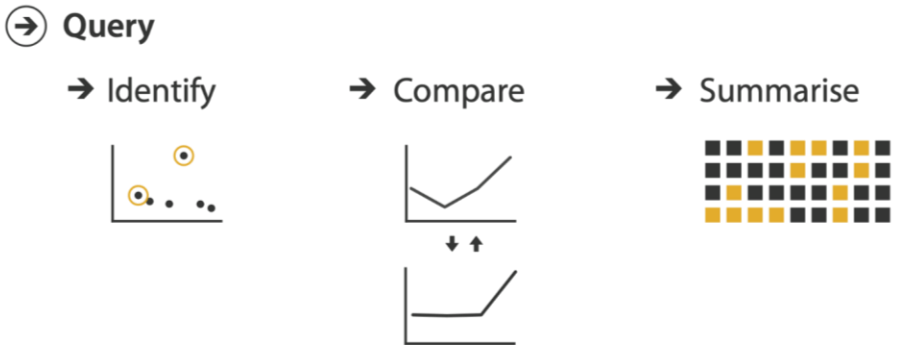


Overview first, zoom and filter, details on demand

- influential mantra from Shneiderman

[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Shneiderman. Proc. IEEE Visual Languages, pp. 336–343, 1996.]

- overview = summary
  - microcosm of full vis design problem

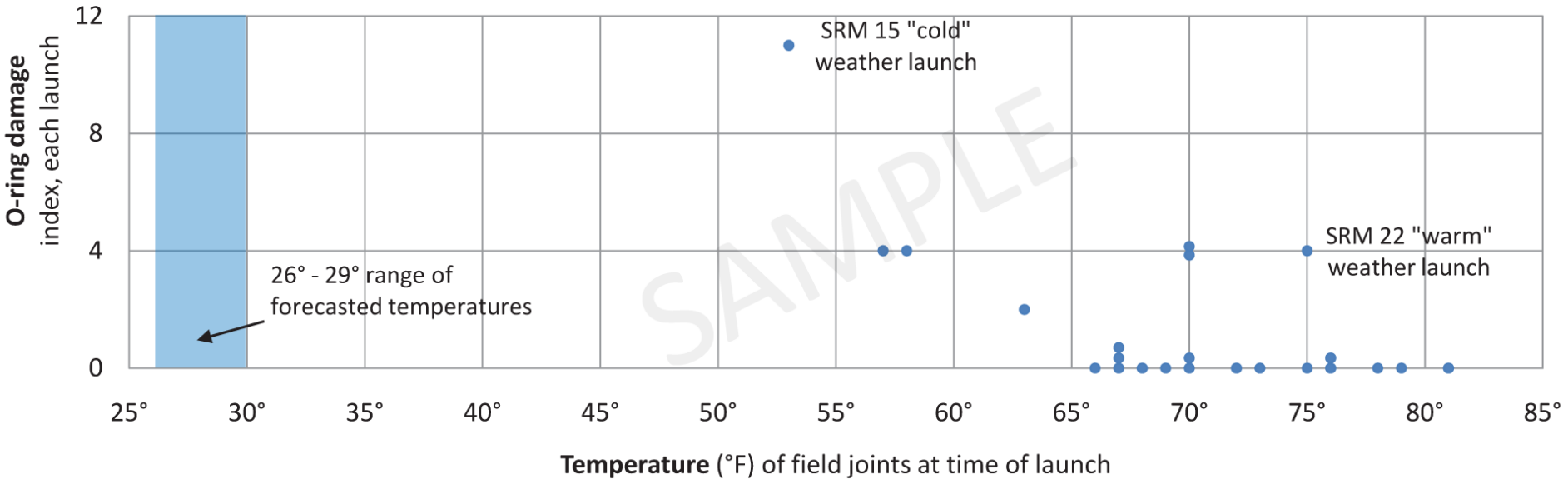


5. Show cause and effect, where possible

Space Shuttle History of Temperature and O-ring Damage

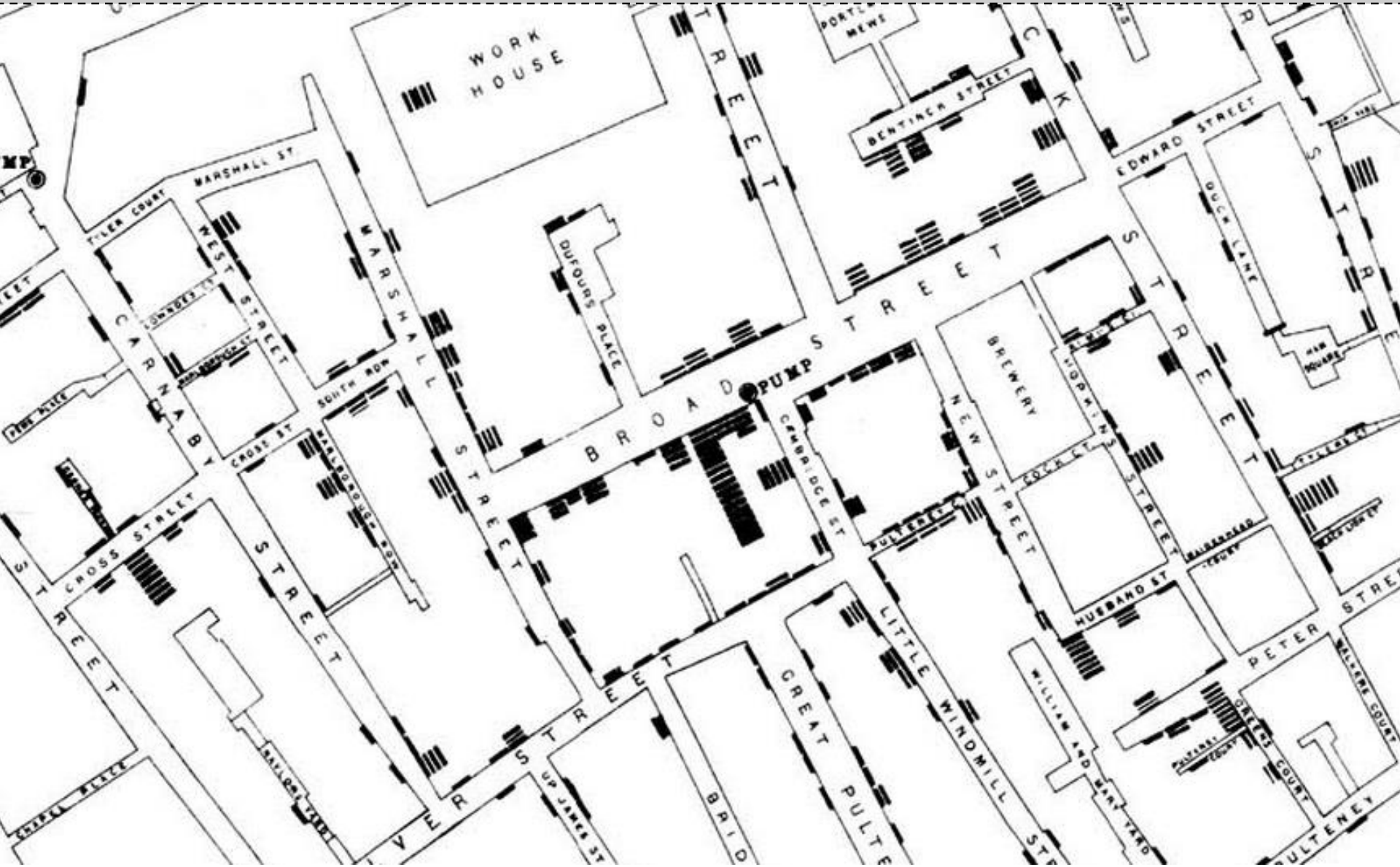
For All 24 Launches Prior to Challenger on January 28, 1986

Solid Rocket Motor (SRM) 15 and SRM 22 were the only prior launches discussed in relation to temperature on the eve of the launch.



Sources: Presidential Commission on the Space Shuttle Challenger Accident (PCSSCA) and Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management as quoted in **Visual and Statistical Thinking** by Edward Tufte.

5. Show cause and effect, where possible

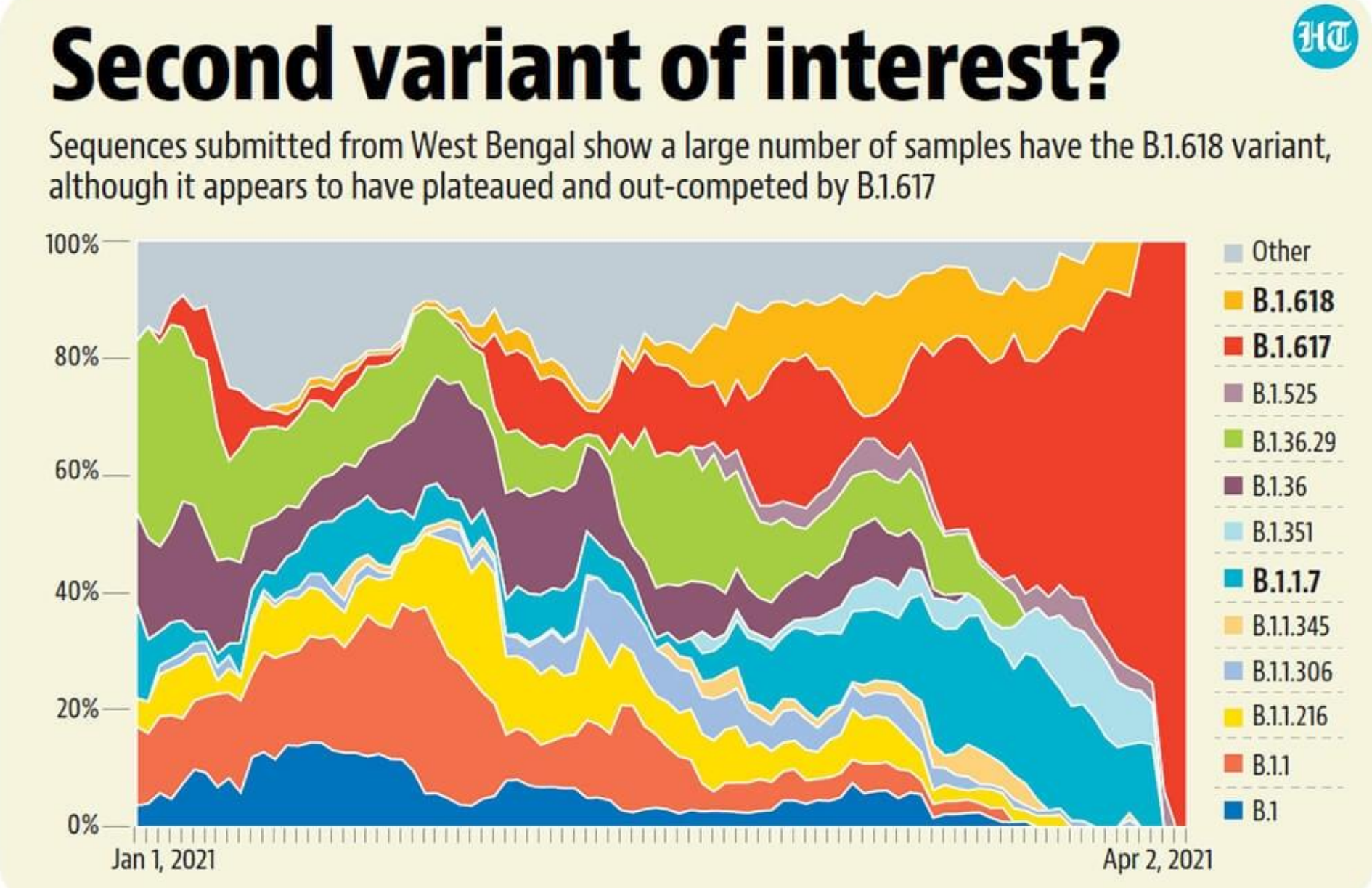


5. Show cause and effect, where possible

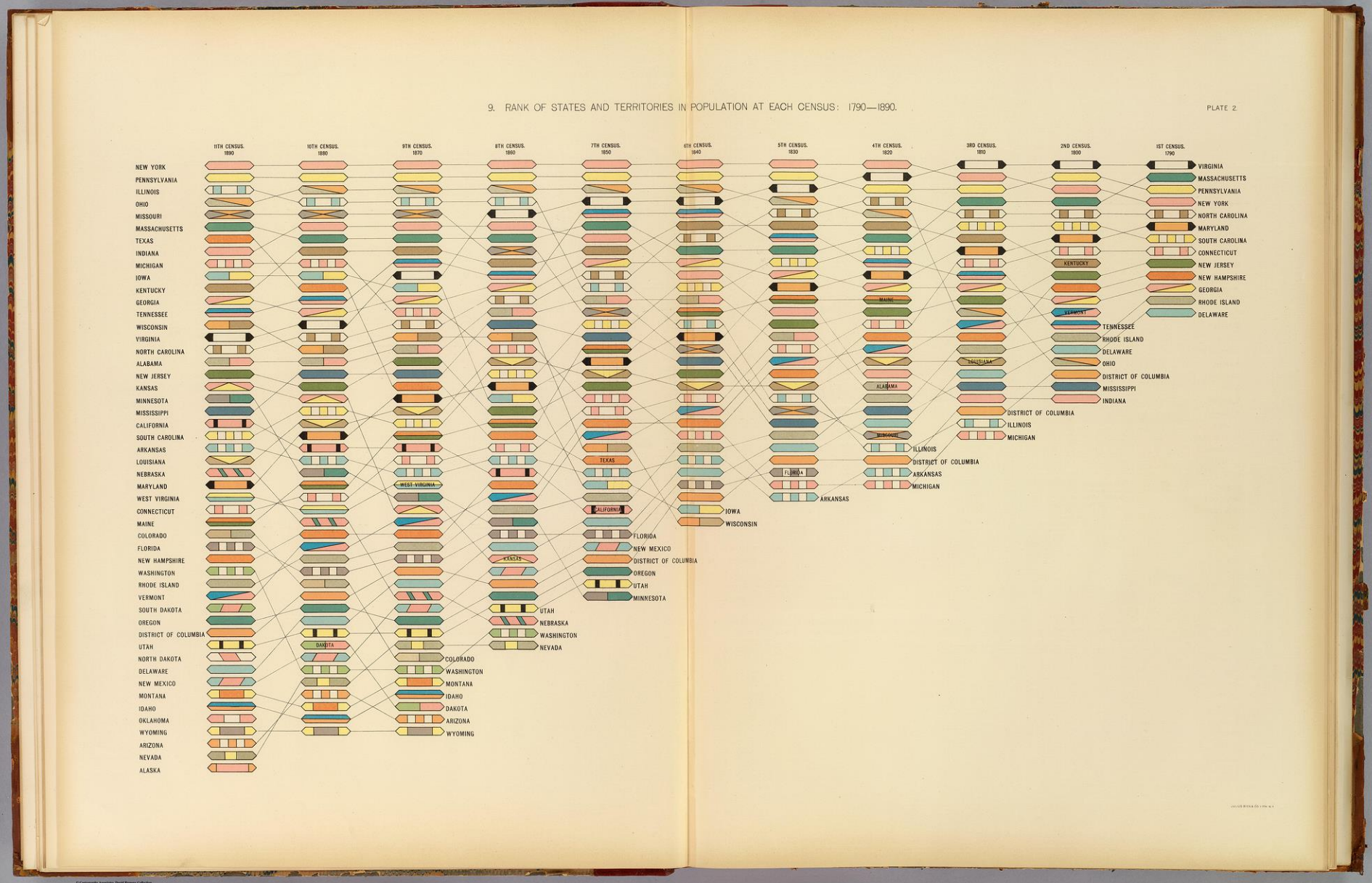
Magician Teller's definition of magic:

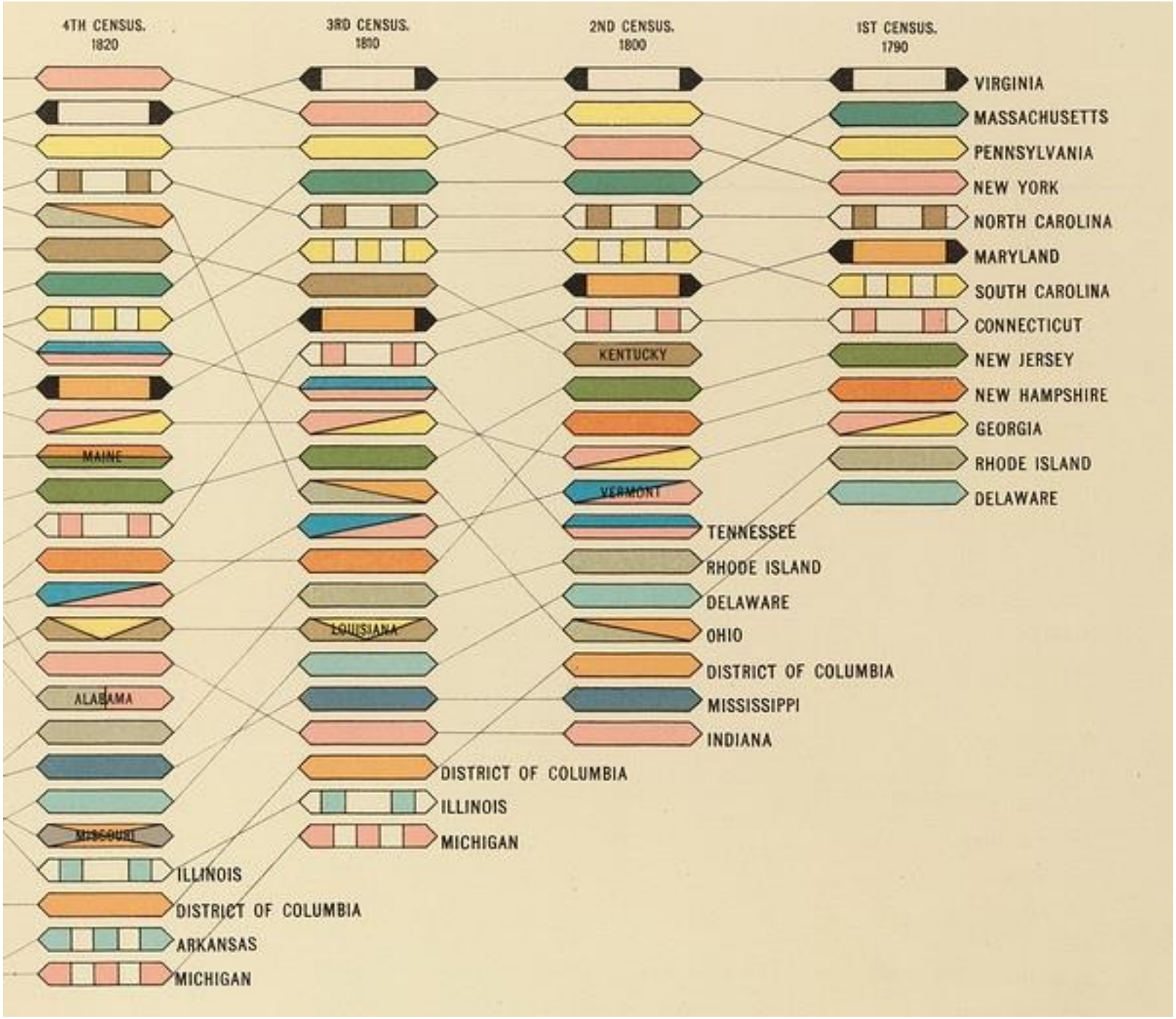
"The theatrical linking of a cause with an effect that has no basis in physical reality, but that — in our hearts — ought to."

6. Compare and contrast, utilize layering & separation

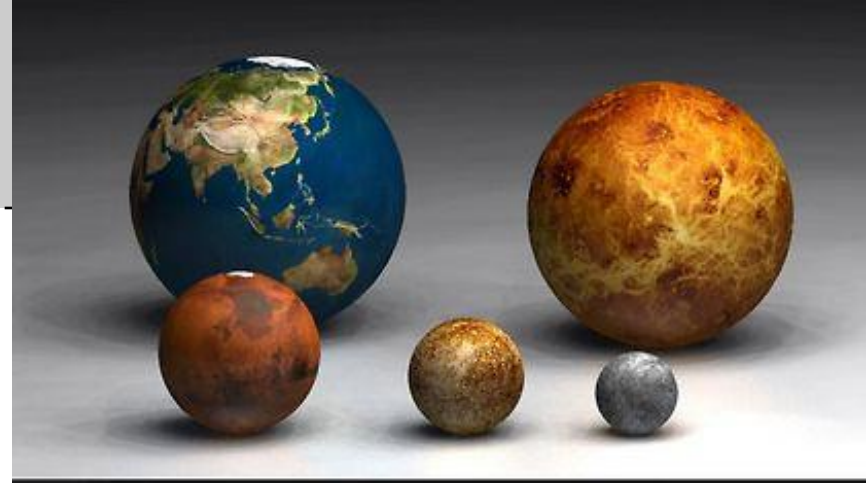


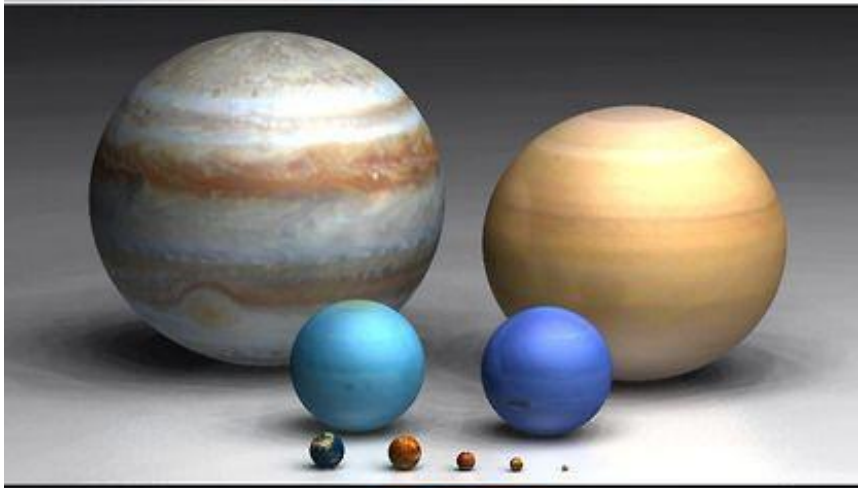
6. Compare and contrast, utilize layering & separation

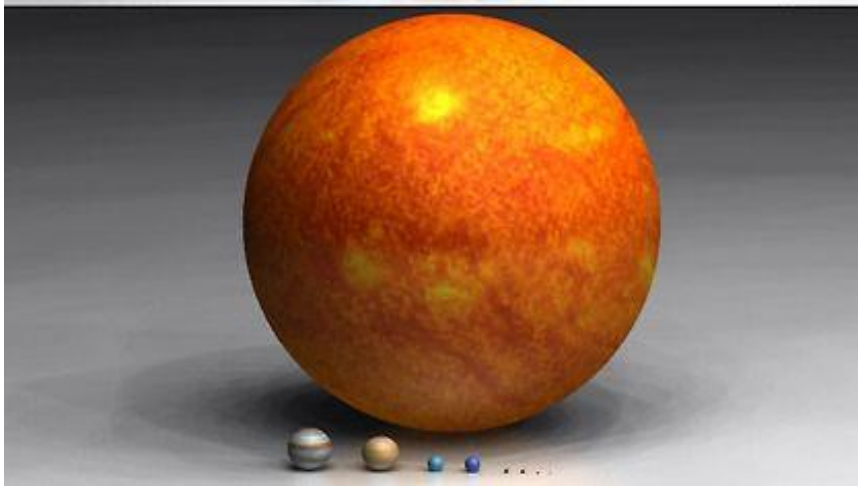



4TH CENSUS.
1820
3RD CENSUS.
1810
2ND CENSUS.
1800
1ST CENSUS.
1790
VIRGINIA
MASSACHUSETTS
PENNSYLVANIA
NEW YORK
NORTH CAROLINA
MARYLAND
SOUTH CAROLINA
CONNECTICUT
NEW JERSEY
NEW HAMPSHIRE
GEORGIA
RHODE ISLAND
DELAWARE
KENTUCKY
VERMONT
TENNESSEE
RHODE ISLAND
DELAWARE
OHIO
DISTRICT OF COLUMBIA
MISSISSIPPI
INDIANA
LOUISIANA
DISTRICT OF COLUMBIA
ILLINOIS
MICHIGAN
MAINE
ALABAMA
MISSOURI
ILLINOIS
DISTRICT OF COLUMBIA
ARKANSAS
MICHIGAN

6. Compare and Contrast

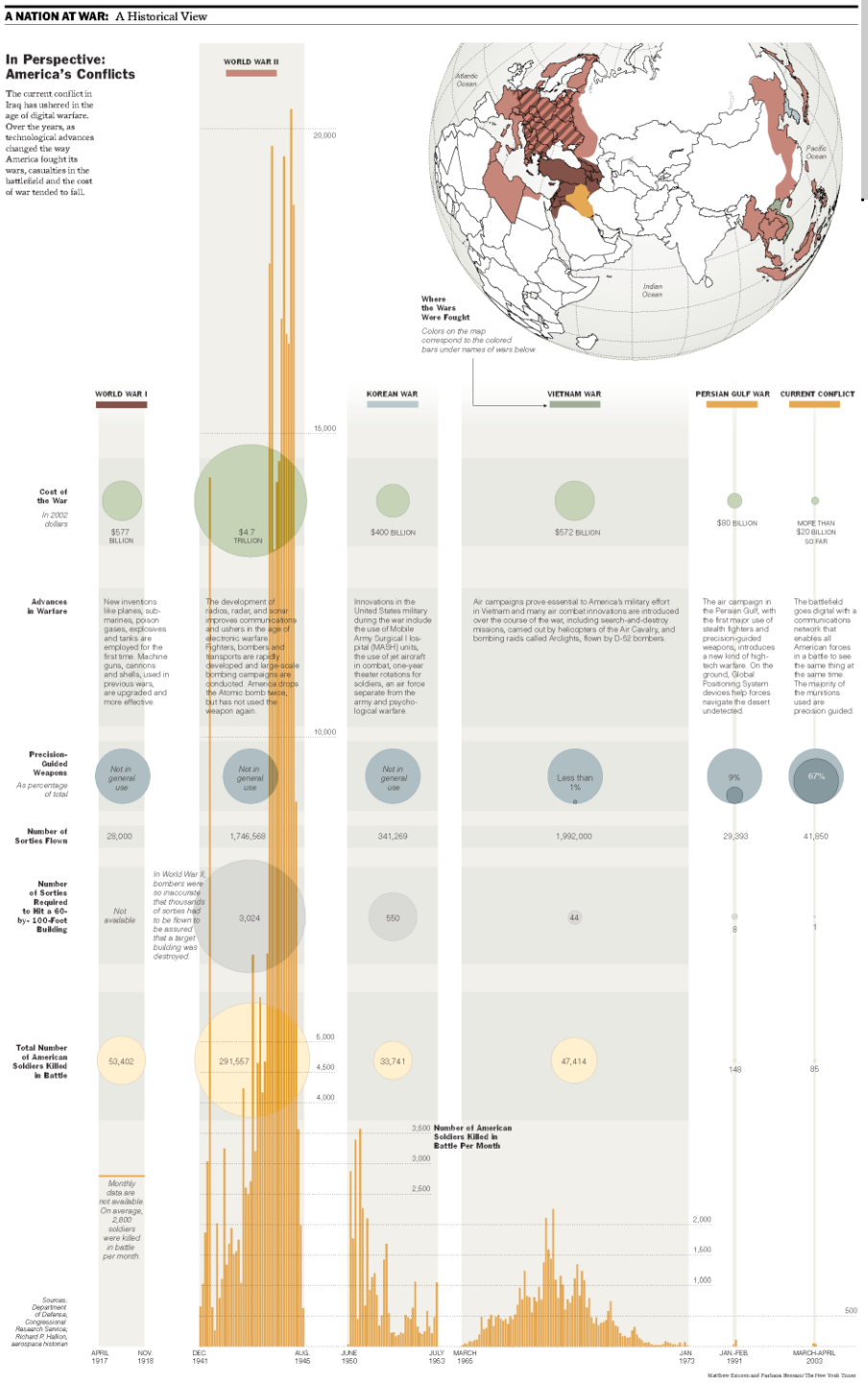






https://www.vox.com/2015/4/17/8432733/space-maps

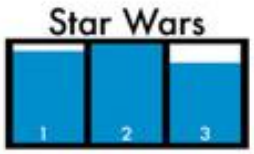
6. Compare and Contrast

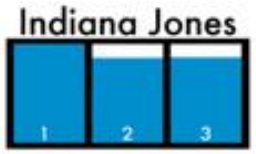


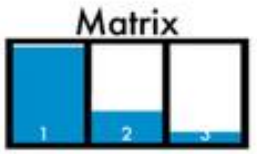
7. Escape flatland, use small multiples, parallel sequencing

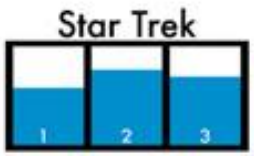
- Data is multivariate doesn’t necessarily mean 3D projection. How can we enhance multivariate data on inherently 2D surfaces? We can use small multiple graphs or parallel sequencing skill.

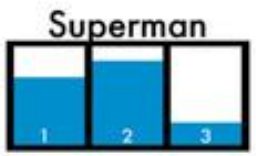
THE TRILOGY METER

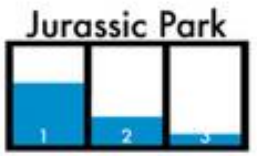


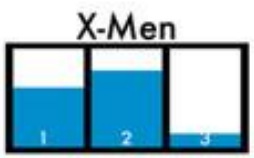


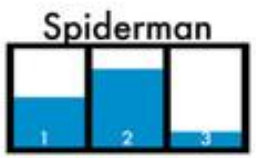


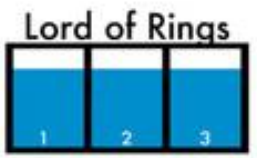


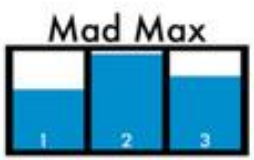


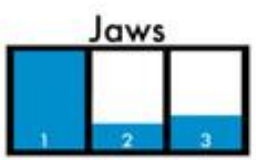


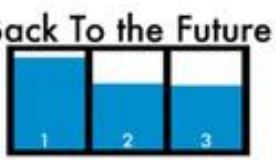


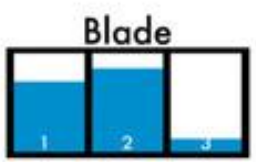


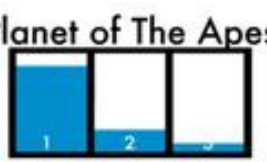


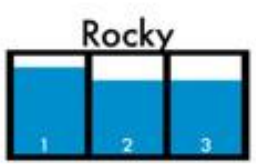


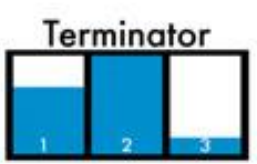


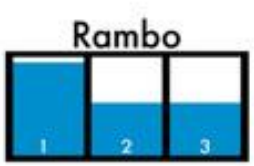


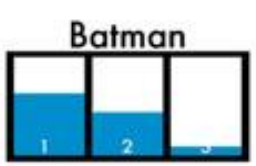


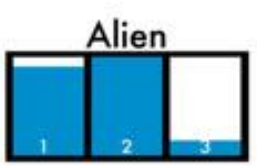










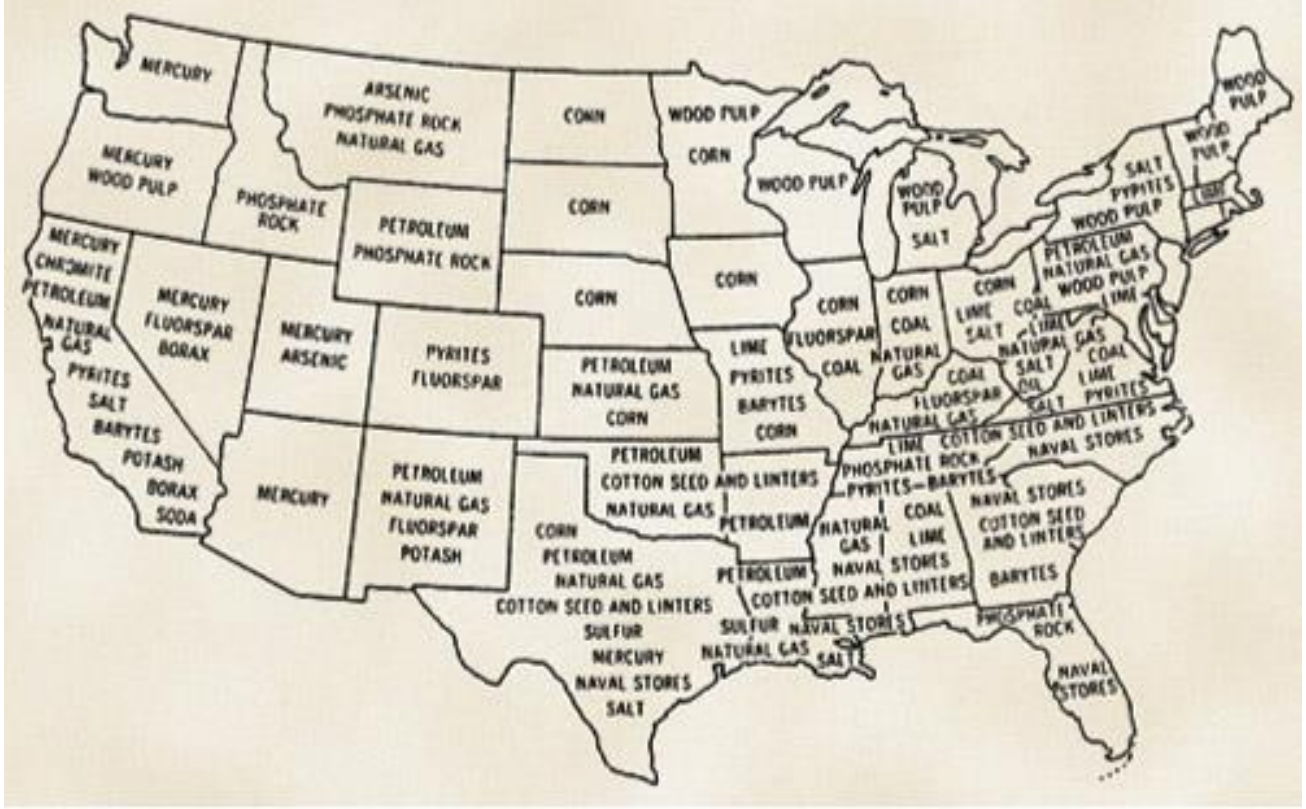


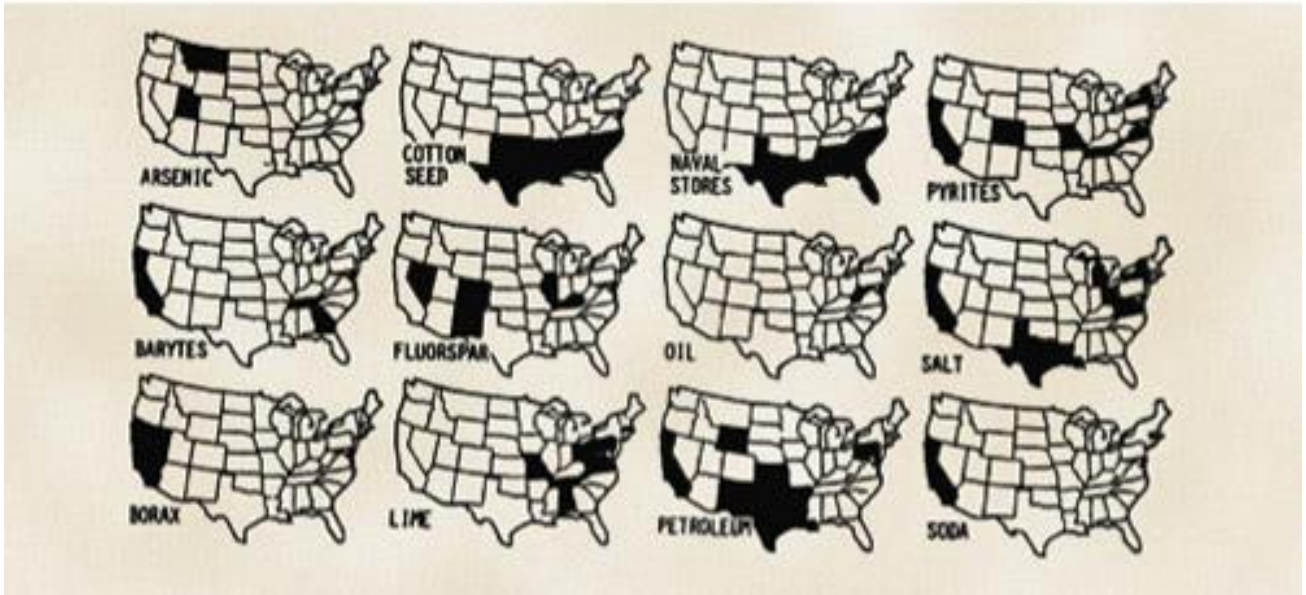
#1 In A Series of Pop Cultural Charts

DANMETH.COM

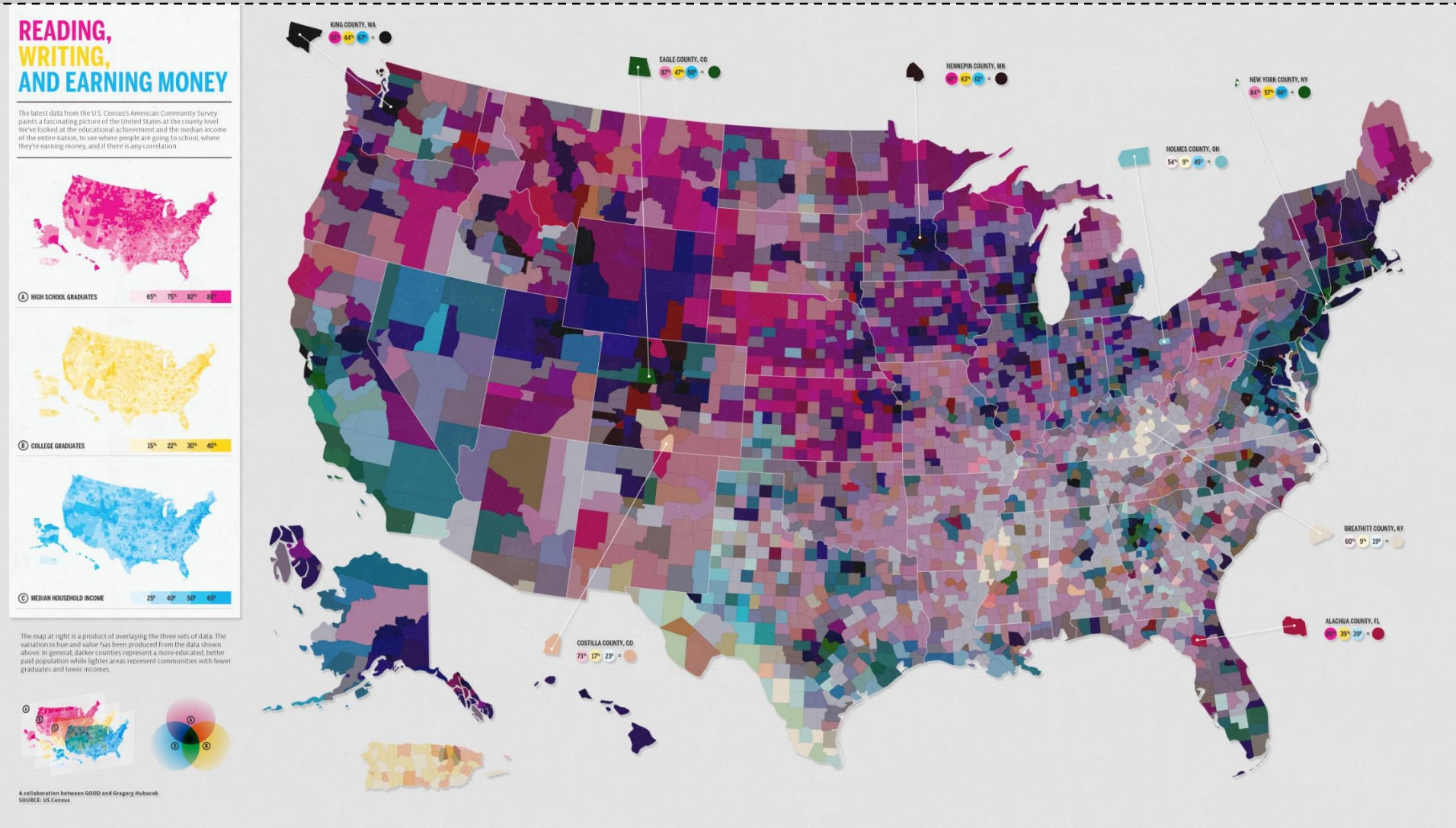
2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support

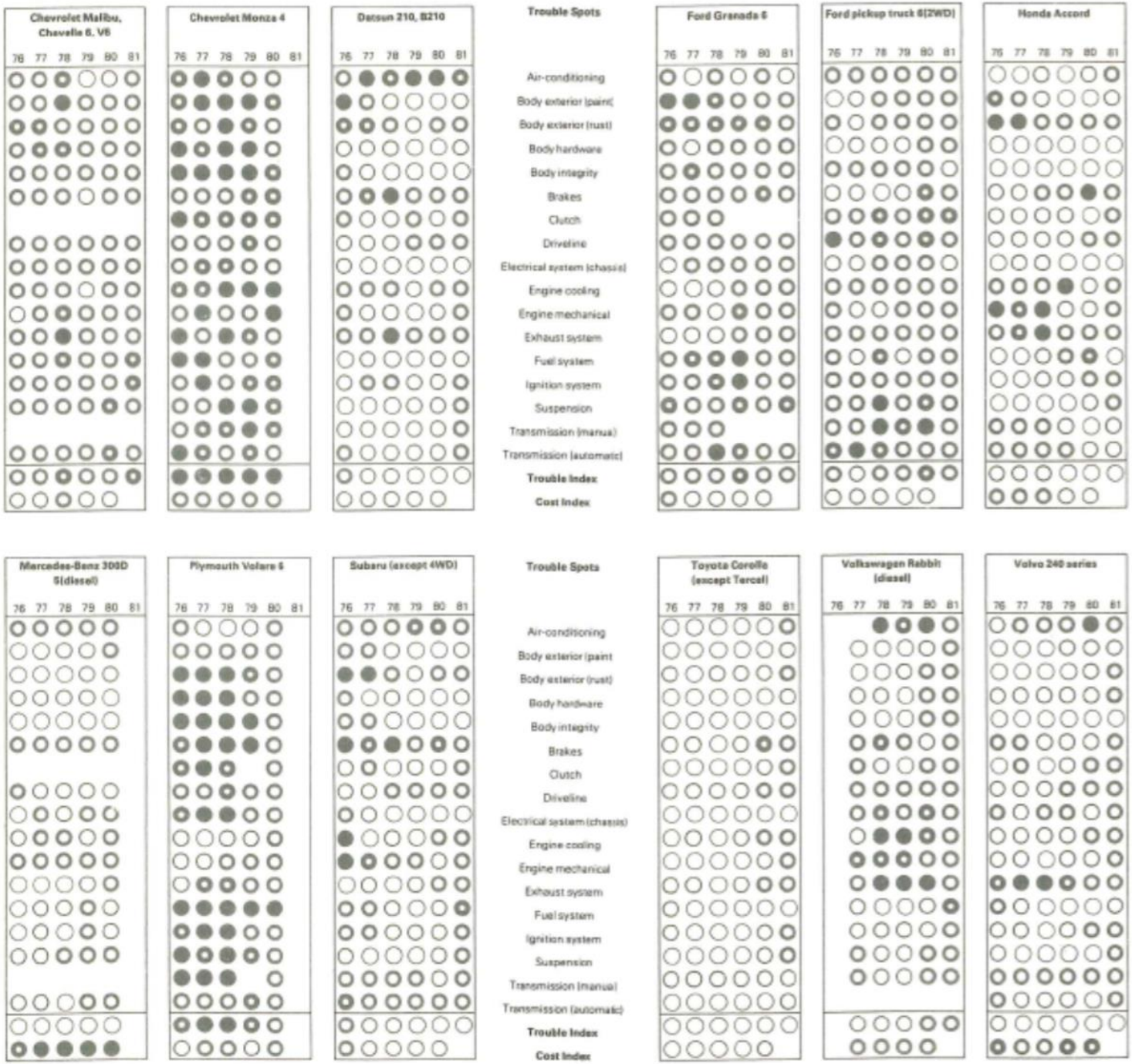
Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnicreligious cagetories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters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


MERCURY
ARSENIC
PHOSPHATE ROCK
NATURAL GAS
CORN
WOOD PULP
CORN
WOOD PULP
WOOD PULP
SALT
WOOD PULP
WOOD PULP
SALT
PYRITES
WOOD PULP
MERCURY
WOOD PULP
PHOSPHATE ROCK
PETROLEUM
PHOSPHATE ROCK
CORN
CORN
CORN
MERCURY
CHROMITE
PETROLEUM
NATURAL GAS
PYRITES
SALT
BARYTES
POTASH
BORAX
SODA
MERCURY
FLUORSPAR
BORAX
MERCURY
ARSENIC
PYRITES
FLUORSPAR
PETROLEUM
NATURAL GAS
CORN
LIME
PYRITES
BARYTES
CORN
CORN
FLUORSPAR
COAL
CORN
COAL
NATURAL GAS
CORN
LIME
SALT
COAL
LIME
NATURAL GAS
SALT
OIL
COAL
FLUORSPAR
COAL
LIME
PYRITES
SALT
NATURAL GAS
COTTON SEED AND LINTERS
NAVAL STORES
LIME
PHOSPHATE ROCK
PYRITES—BARYTES
MERCURY
PETROLEUM
NATURAL GAS
FLUORSPAR
POTASH
PETROLEUM
COTTON SEED AND LINTERS
NATURAL GAS
CORN
PETROLEUM
NATURAL GAS
COTTON SEED AND LINTERS
SULFUR
MERCURY
NAVAL STORES
SALT
PETROLEUM
NATURAL GAS
COAL
LIME
NAVAL STORES
COTTON SEED AND LINTERS
NAVAL STORES
COTTON SEED AND LINTERS
BARYTES
PETROLEUM
COTTON SEED AND LINTERS
SULFUR
NATURAL GAS
NAVAL STORES
SALT
PHOSPHATE ROCK
NAVAL STORES

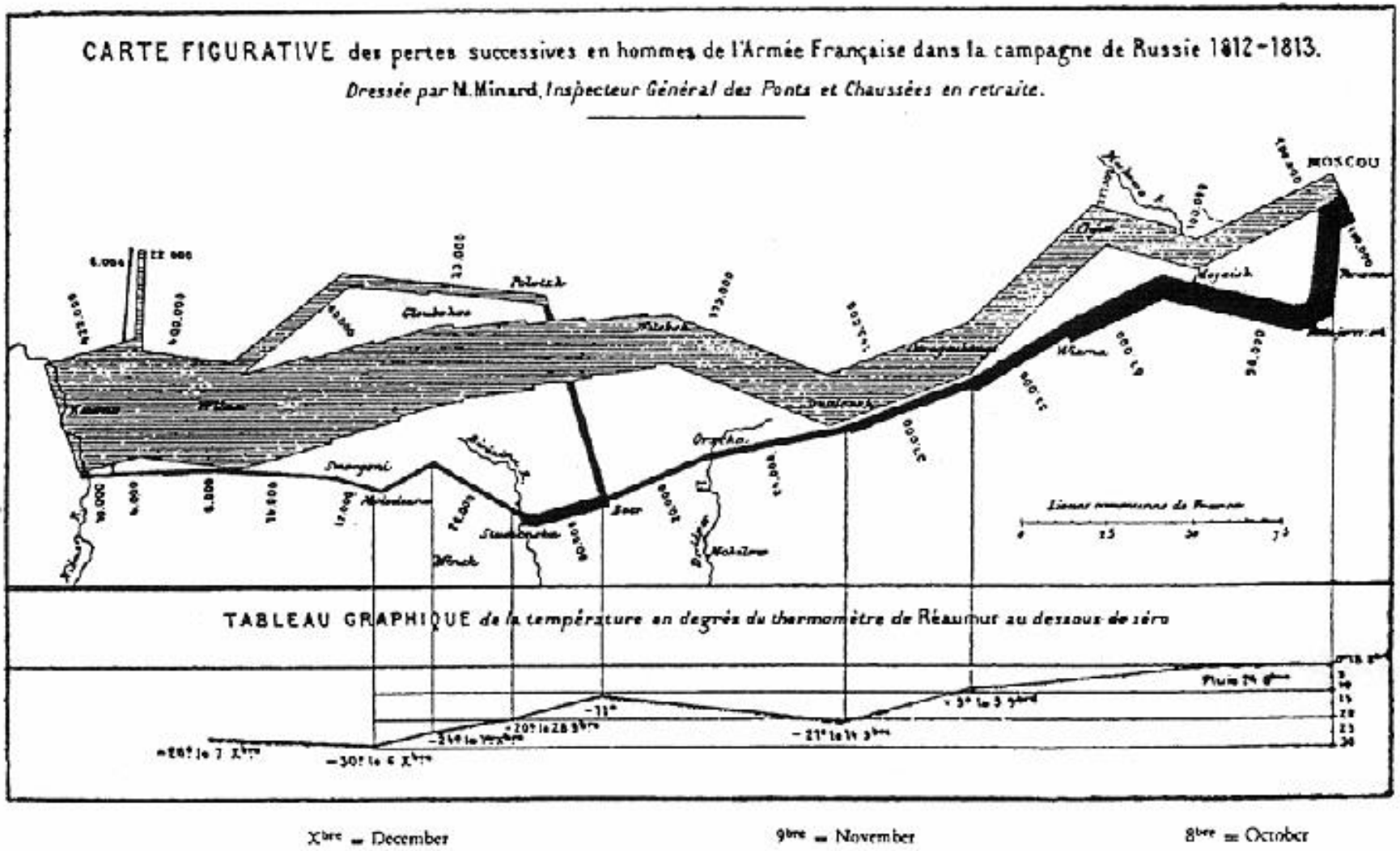

ARSENIC
COTTON SEED
NAVAL STORES
PYRITES
BARYTES
FLUORSPAR
OIL
SALT
BORAX
LIME
PETROLEUM
SODA

Separability vs. Integrality

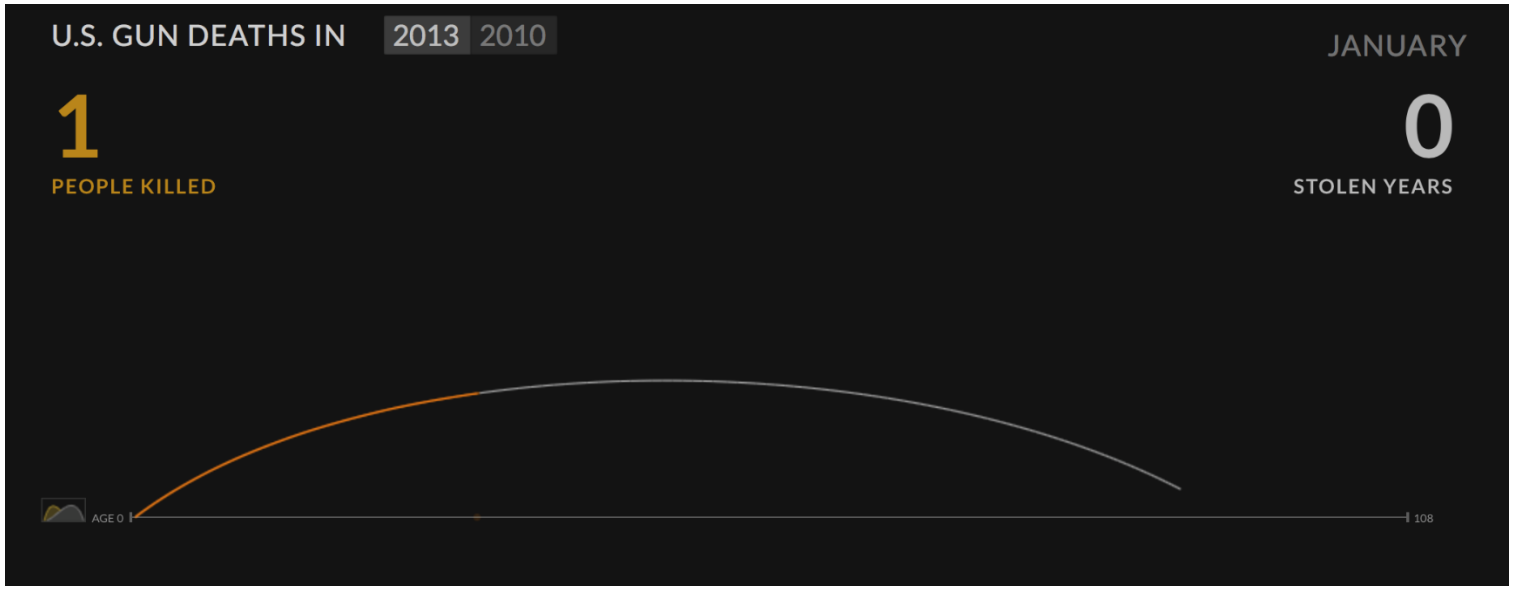


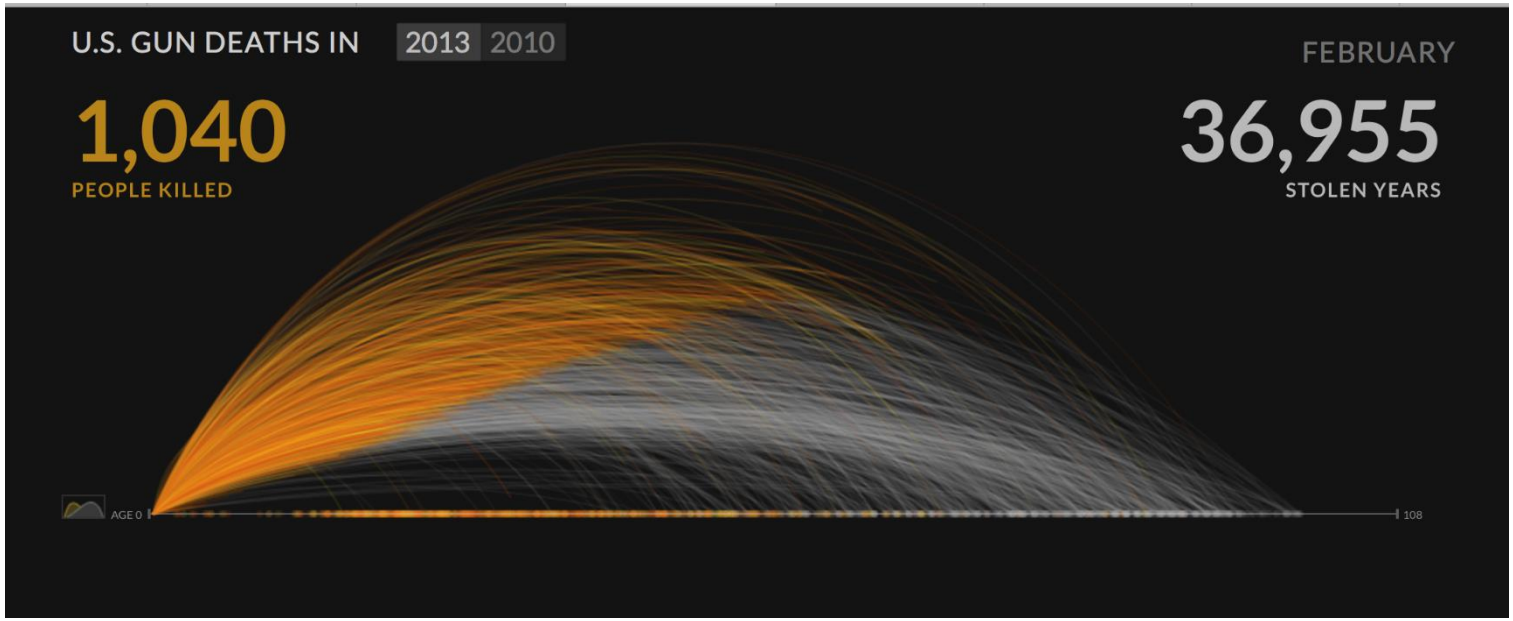

Chevrolet Malibu, Chevelle 6, V8
Chevrolet Monza 4
Datsun 210, B210
Trouble Spots
Ford Granada 6
Ford pickup truck 6(2WD)
Honda Accord
76 77 78 79 80 81
Air-conditioning
Body exterior (paint)
Body exterior (rust)
Body hardware
Body integrity
Brakes
Clutch
Driveline
Electrical system (chassis)
Engine cooling
Engine mechanical
Exhaust system
Fuel system
Ignition system
Suspension
Transmission (manual)
Transmission (automatic)
Trouble Index
Cost Index
Mercedes-Benz 300D 5(diesel)
Plymouth Volare 6
Subaru (except 4WD)
Trouble Spots
Toyota Corolla (except Tercel)
Volkswagen Rabbit (diesel)
Volvo 240 series
76 77 78 79 80 81
Air-conditioning
Body exterior (paint
Body exterior (rust)
Body hardware
Body integrity
Brakes
Clutch
Driveline
Electrical system (chassis)
Engine cooling
Engine mechanical
Exhaust system
Fuel system
Ignition system
Suspension
Transmission (manual
Transmission (automatic)
Trouble Index
Cost Index

8. Show Multiple Dimensions

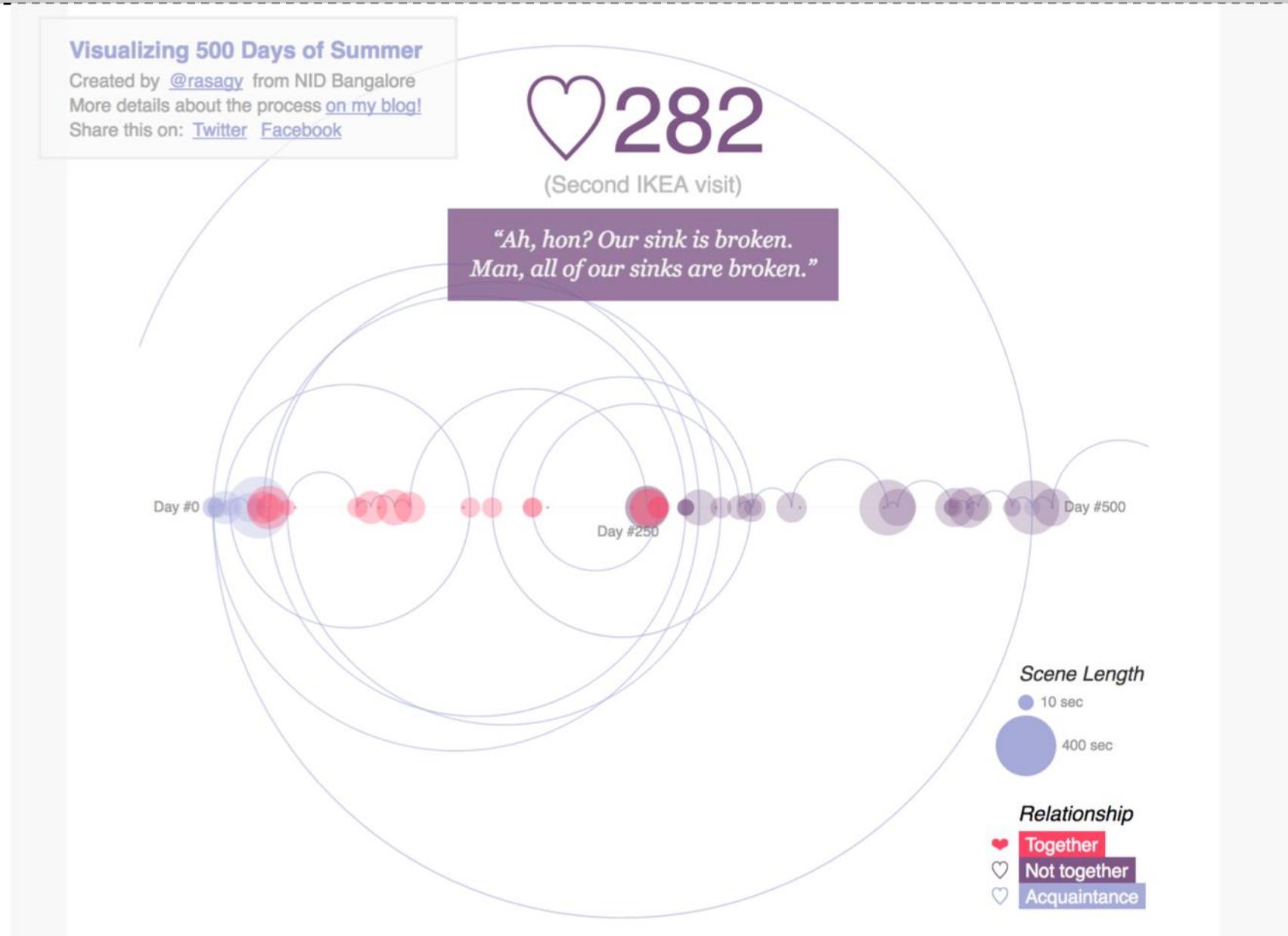


9. Utilize narratives of space and time

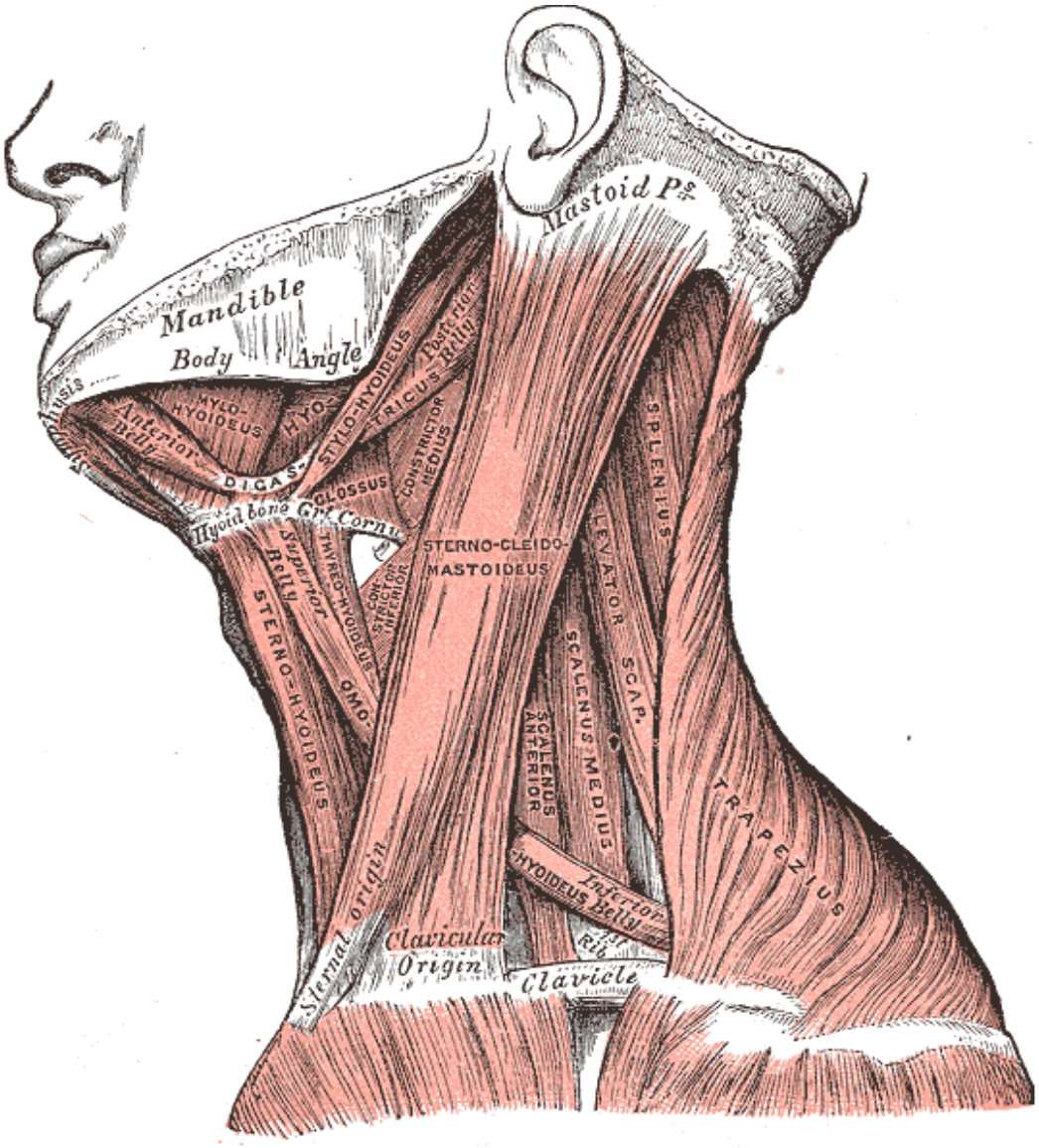




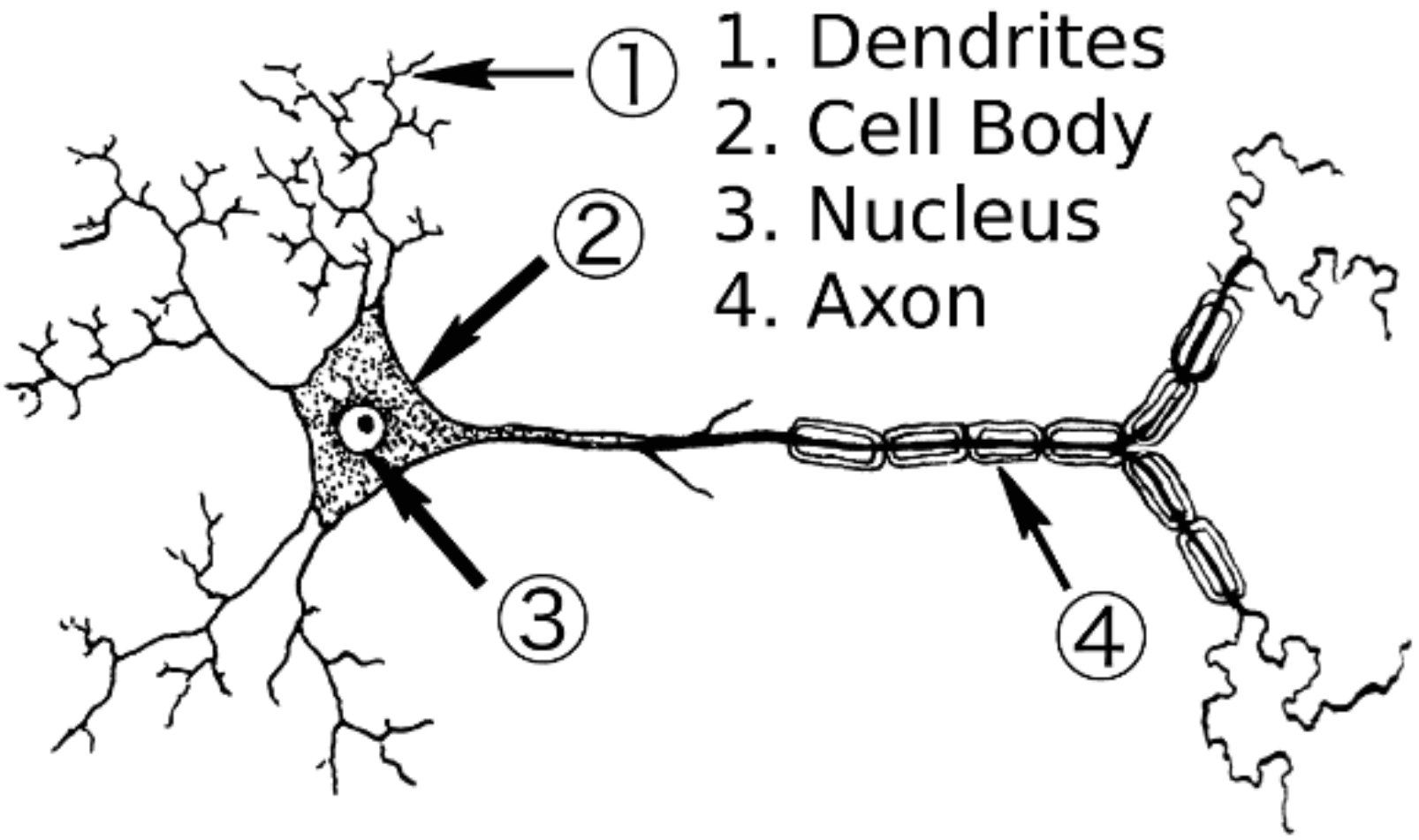
9. Utilize narratives of space and time



10. Integrate image, number and text



10. Integrate image, number and text

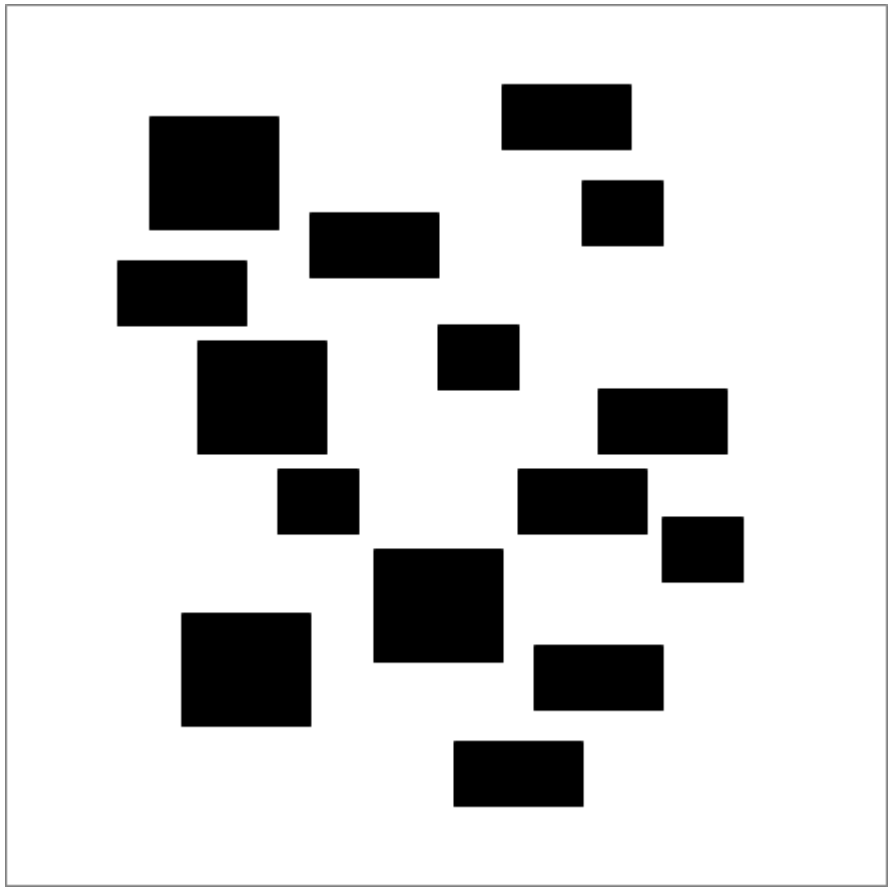


Tufte's design principles for graphical excellence

1. Maximize the data-ink ratio, within reason
2. Mobilize every graphical element, perhaps several times over, to show the data
3. Maximize data density and the size of the data matrix, within reason
4. Establish context
5. Show cause and effect, where possible
6. Compare and contrast, utilize layering & separation
7. Escape flatland, use small multiples, parallel sequencing (reality is multivariate)
8. Show multiple dimensions
9. Utilize narratives of space and time
10. Integrate image, number and text

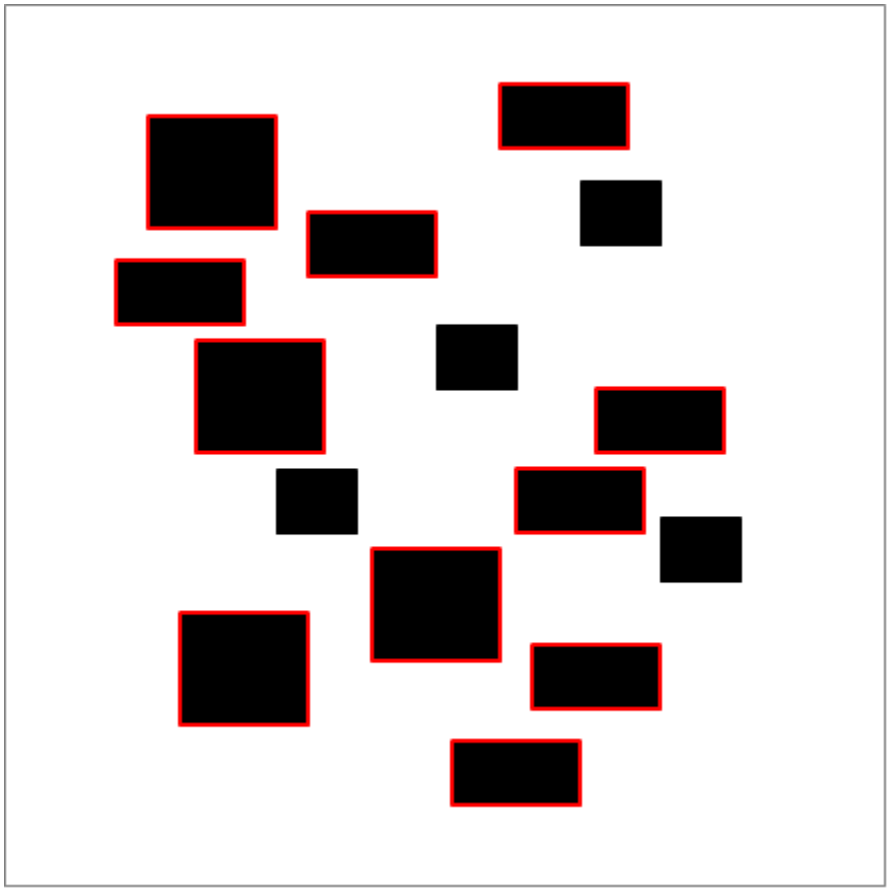
Separability vs. Integrality

Below is a scatter plot where the height is mapped to one data variable and the width to another. Can you spot all the rectangles with the same width?



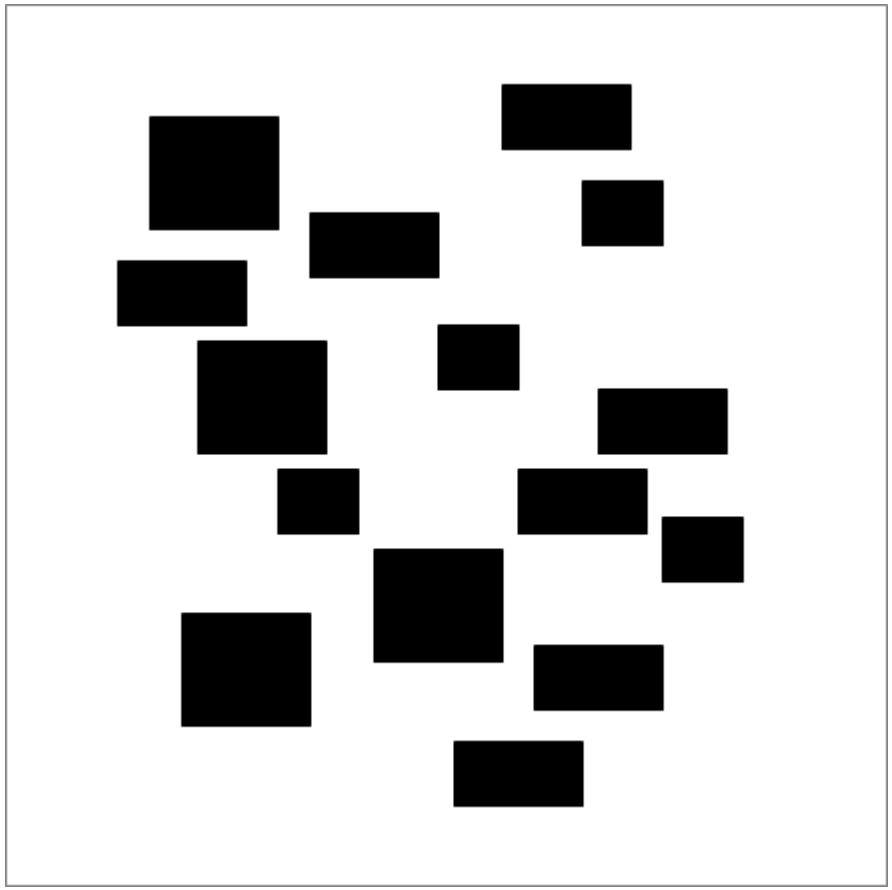
Separability vs. Integrality

Below is a scatter plot where the height is mapped to one data variable and the width to another. Can you spot all the rectangles with the same width?



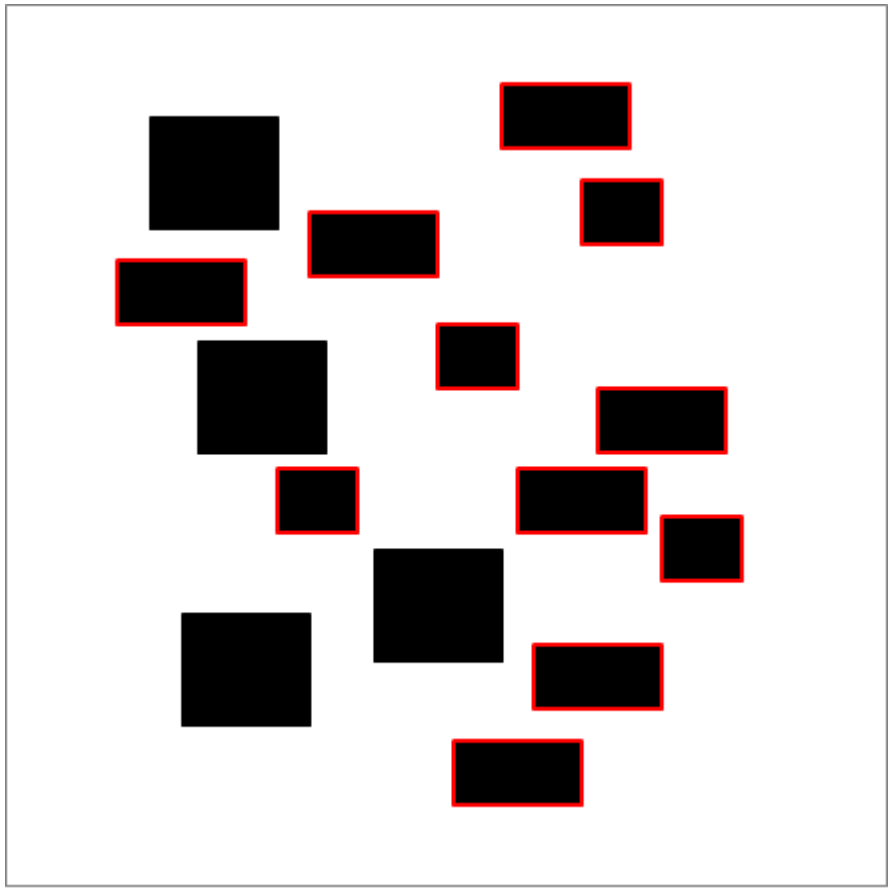
Separability vs. Integrality

Below is a scatter plot where the height is mapped to one data variable and the width to another. Can you spot all the rectangles with the same height?



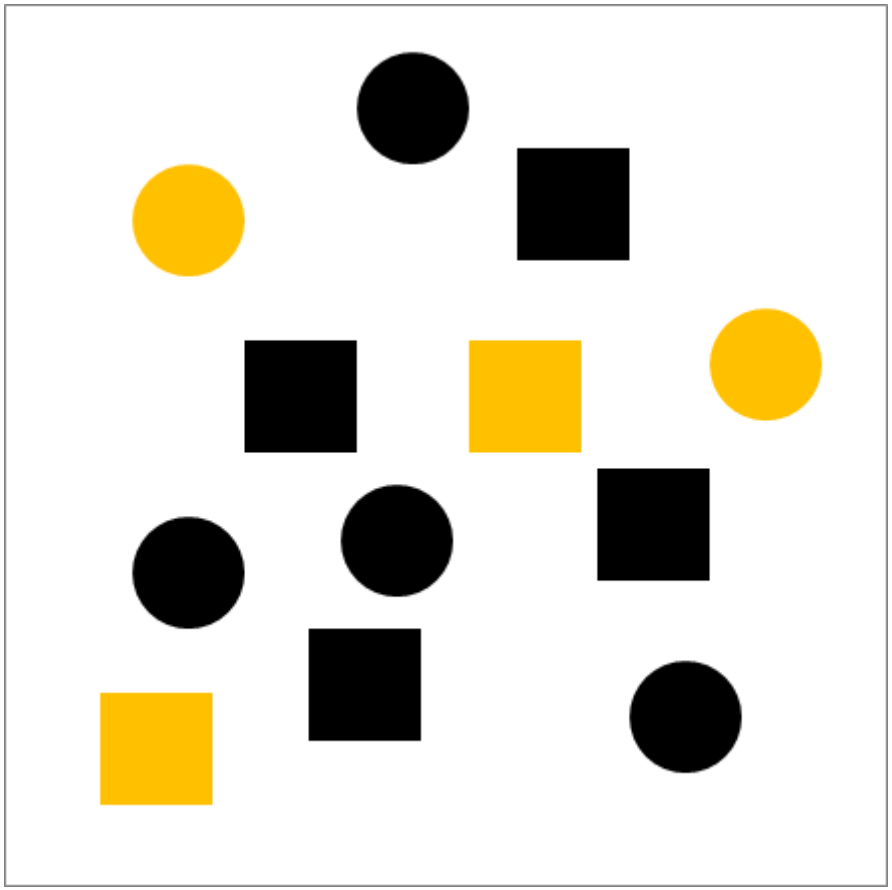
Separability vs. Integrality

Below is a scatter plot where the height is mapped to one data variable and the width to another. Can you spot all the rectangles with the same height?



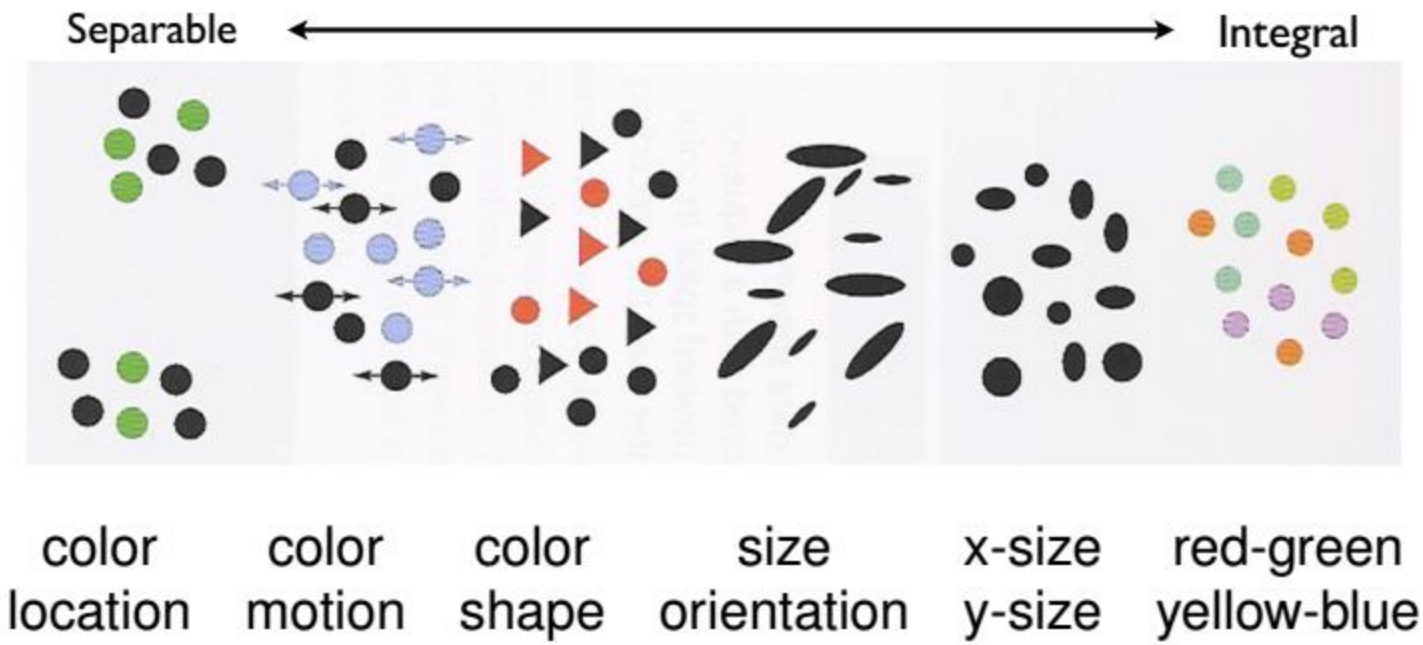
Separability vs. Integrality

On the contrary, you can more easily spot yellow or black dots. And you can also spot circles or squares. Shape and colour are more separable than width and height.



Separability vs. Integrality

Colin Ware ordered the dimensions from the most separable on the left, to the most integral on the right



Ware, "Information Visualization"

Separability vs. Integrality

- The choice of which visual features are used in conjunction to encode the various data features greatly affects the way their are perceived
- All features influence each other to some extent but some more than others. For instance, if you use colour and size to encode two data variables, the way colour is perceived will be affected by the size of the object
- Two data variabes are integral when they are perceived holistically, that is, it's hard to visually decode the value of one independently from the other