Data Visualization

VISUALIZATION DESIGN

Overview

The 7 steps of visualization design

Basic charts

Multivariate/multidimensional data visualization

Visualizing uncertainty and missing data

Visual order

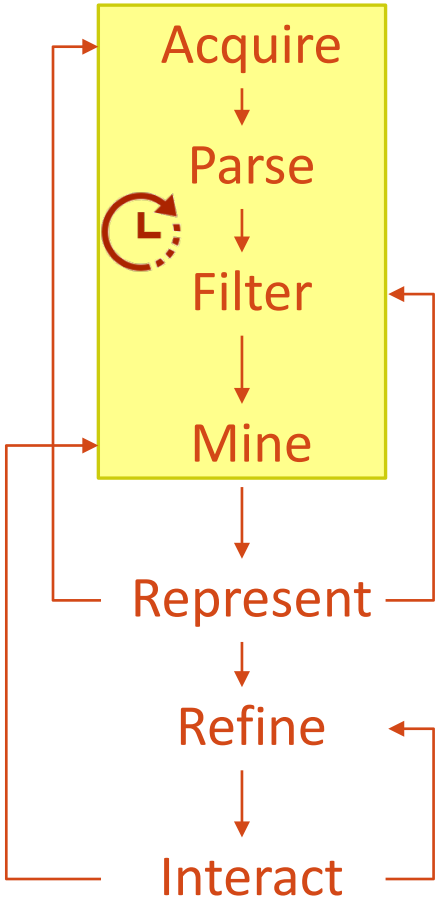
Interactivity

Storytelling

Tools

The 7 steps of visualization design

The 7 steps of visualization design



Obtain the data

Provide some structure for the data's meaning and order it into categories

Remove all but the data of interest

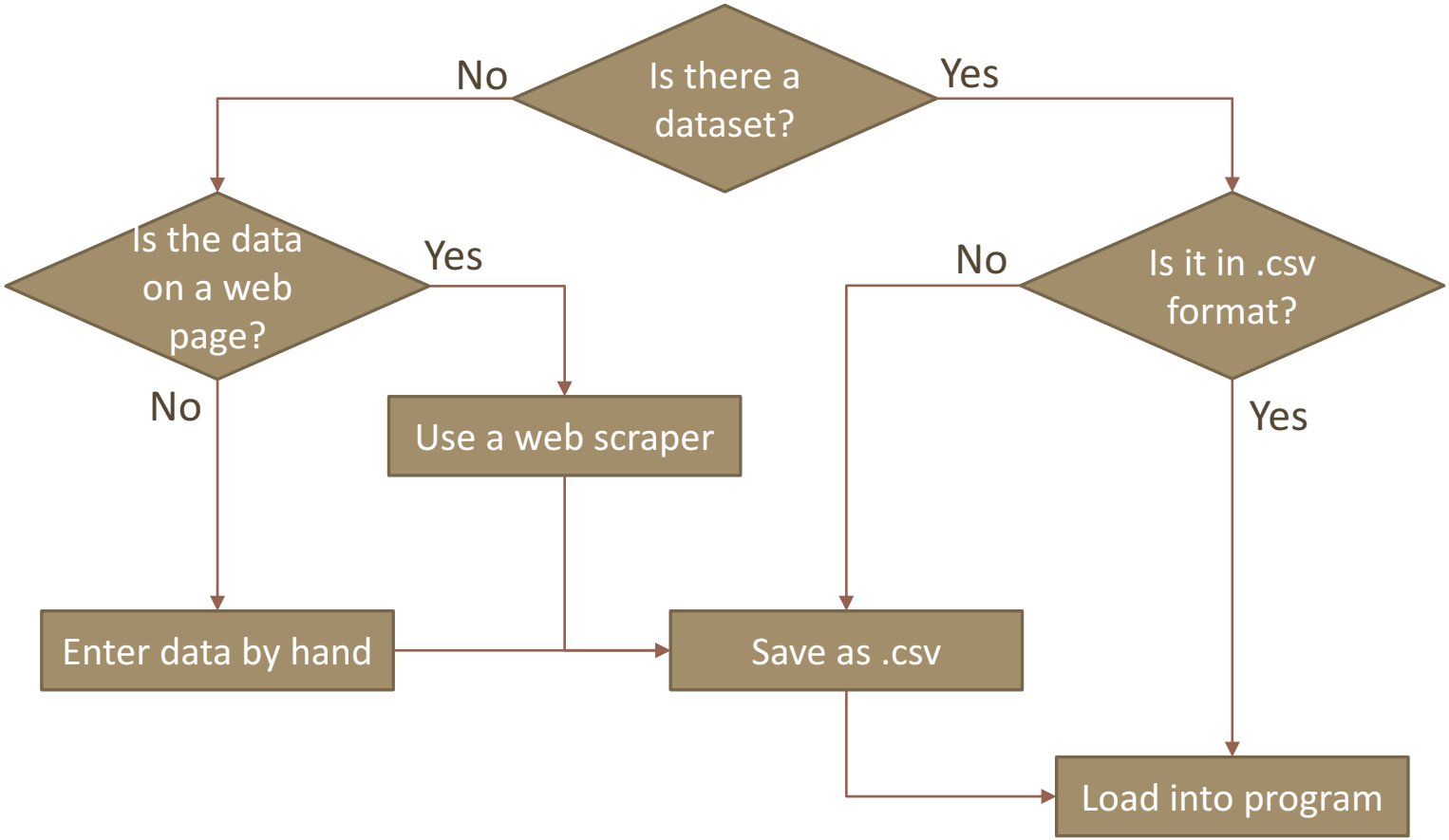
Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context

Choose a basic visual model and draw the data

Improve the basic representation to make it clearer, more meaningful and more visually engaging

Add methods for manipulating the data or controlling what features are visible

Acquire the data



https://itunes.apple.com/us/course/data-literacy-and-data-visualization/id693097601

Acquire the data

Web scrapers

- Scraper (plugin for Chrome) https://chrome.google.com/webstore/detail/scraper/mbigbapnjcgaffohmbkdlecaccepngjd
- Data Scraper (plugin for Firefox) https://addons.mozilla.org/sl/firefox/addon/datascraper/
- Outwit Hub (standalone program, limited functionalities of the free version)

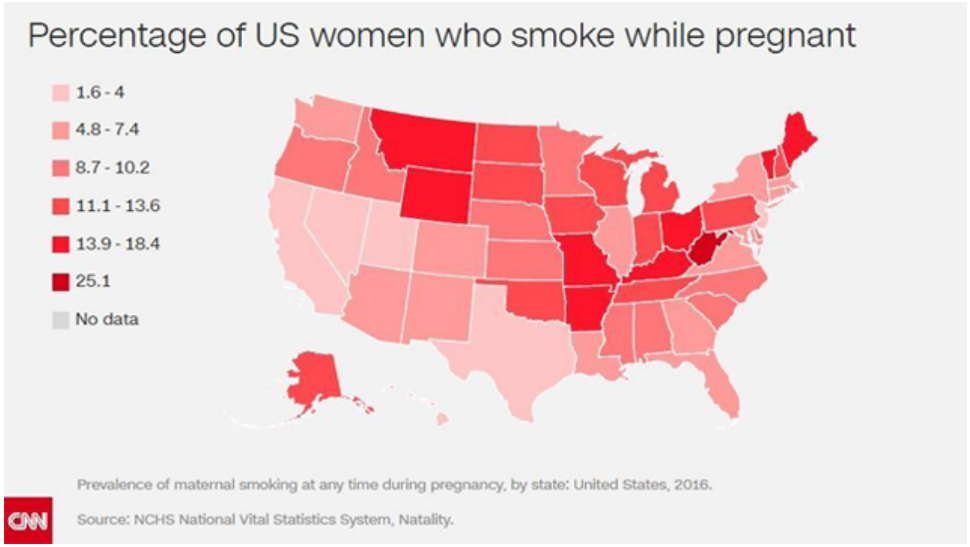
  https://www.outwit.com/

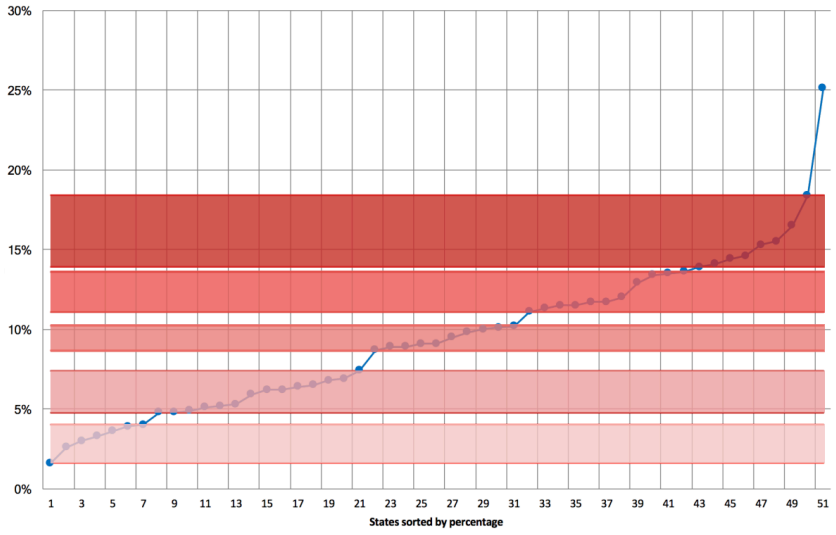
Parse the data

Check for errors

Change type

- For example, ordinal to categorical



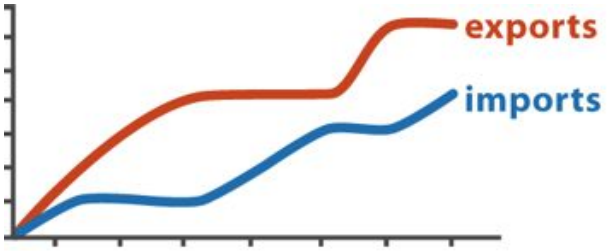


https://meredith.images.worldnow.com/images/16207225_G.png?auto=webp&disable=upscale&width=800

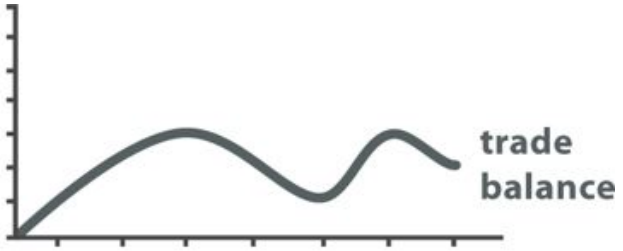
Parse the data

Transform data

- Transform city name to geographical coordinates
- Derive new attributes from existing ones using arithmetic, logical or statistical operations
  - Compute relative data from absolute data
  - Compute cumulative data



Original Data



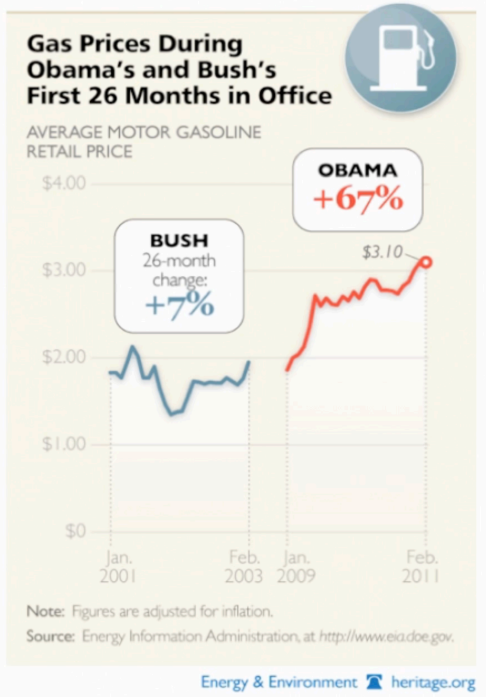
trade balance = *exports* – *imports*

Derived Data

Filter the data

Remove all but the data of interest

Be careful – do not remove relevant data showing patterns!



Mine the data

Exploratory data analysis

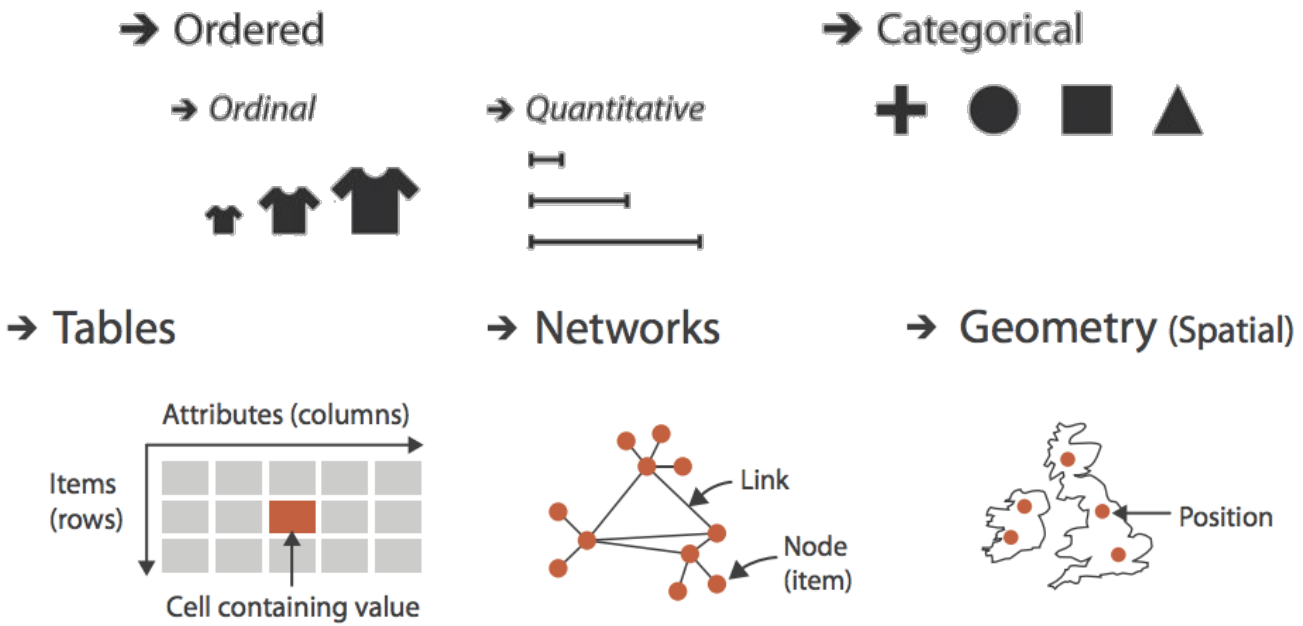
- Look for important features and patterns
- Look for any striking deviations (outliers)
- Interpret your findings

Start with univariate analysis (one variable at a time), continue with multivariate analysis

Represent the data

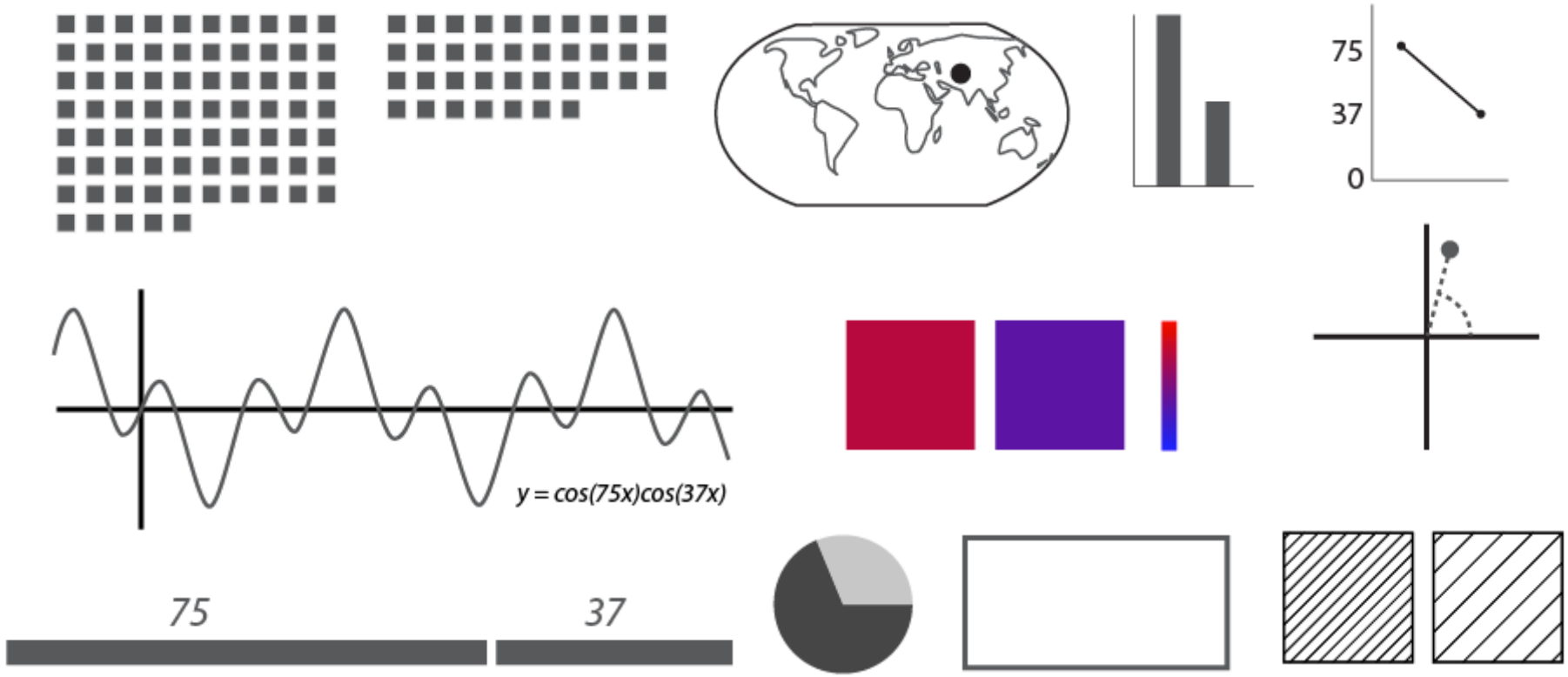
Choose a basic visual model and draw the data

Choice depends on the data and the task



Represent the data

45 ways to communicate two quantities (75 and 37)

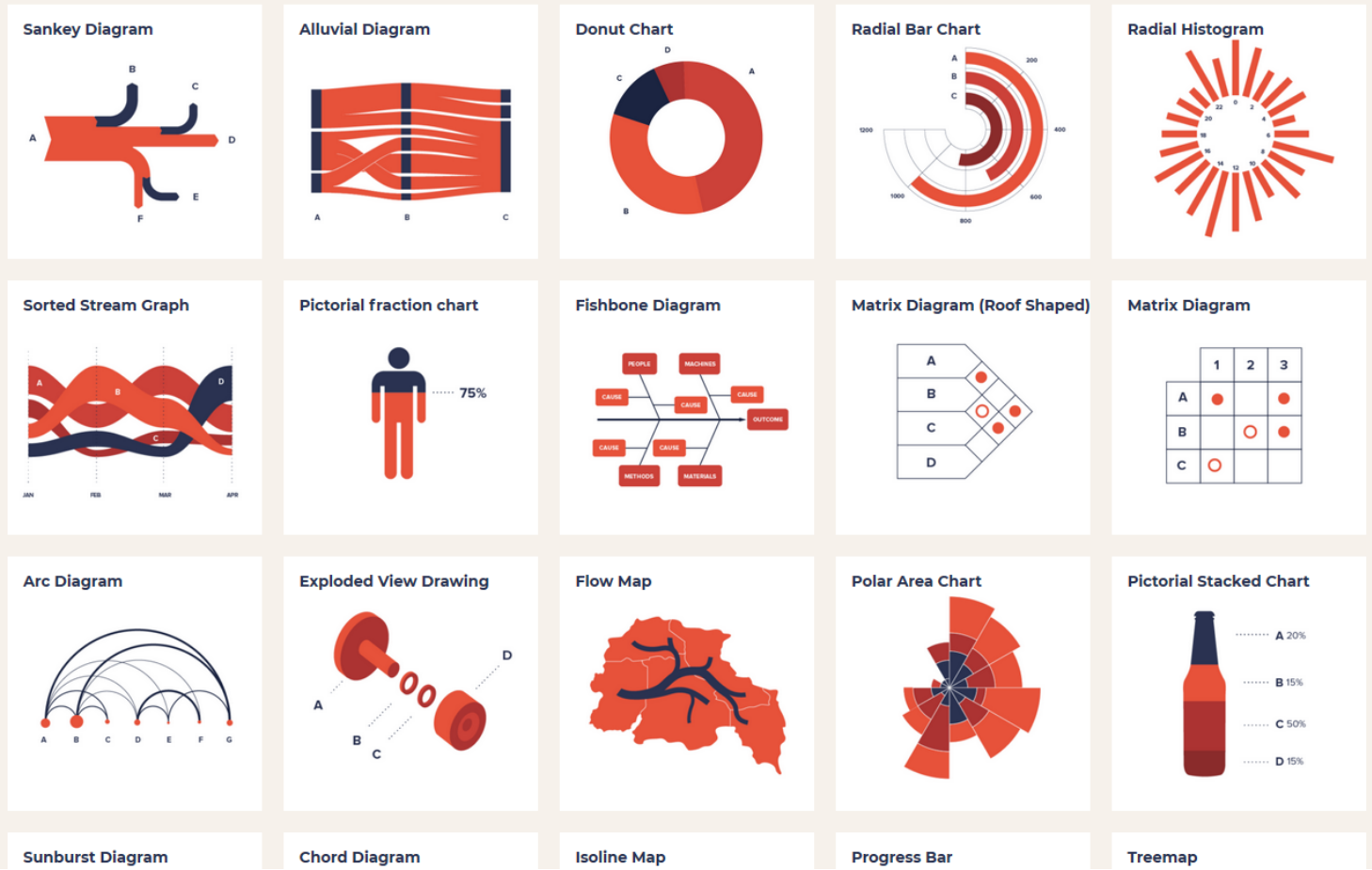


https://visual.ly/blog/45-ways-to-communicate-two-quantities/

Represent the data

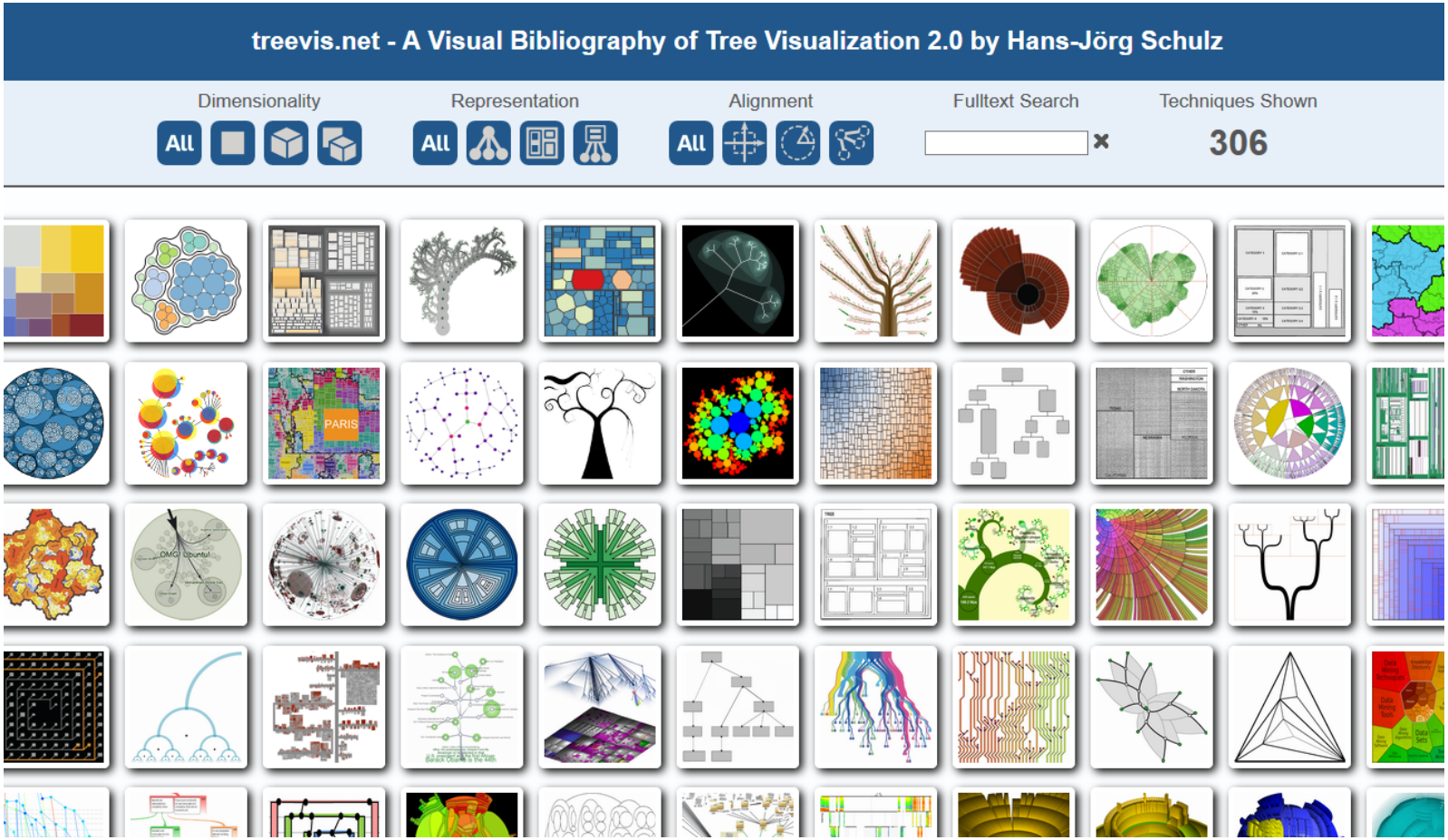
https://datavizcatalogue.com/search.html

Represent the data



https://datavizproject.com/

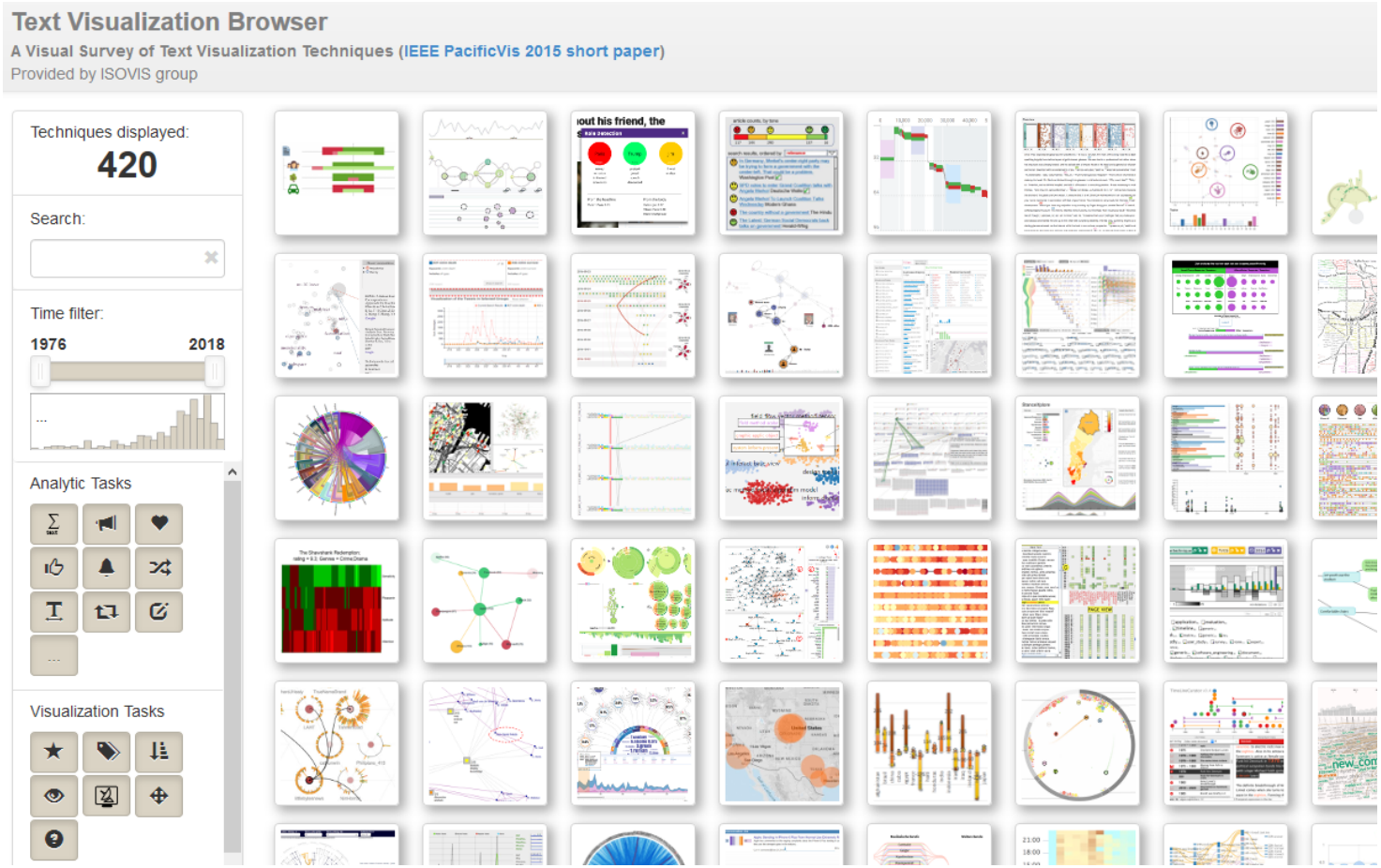
Represent the data



Represent the data

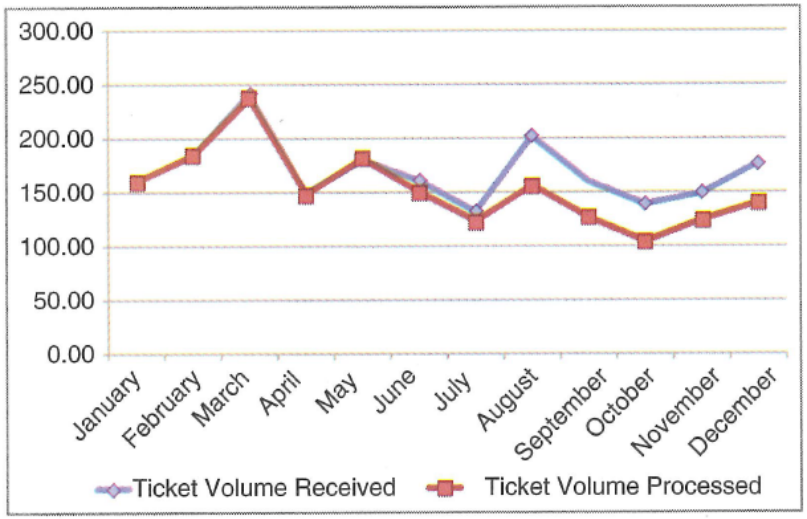
https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html

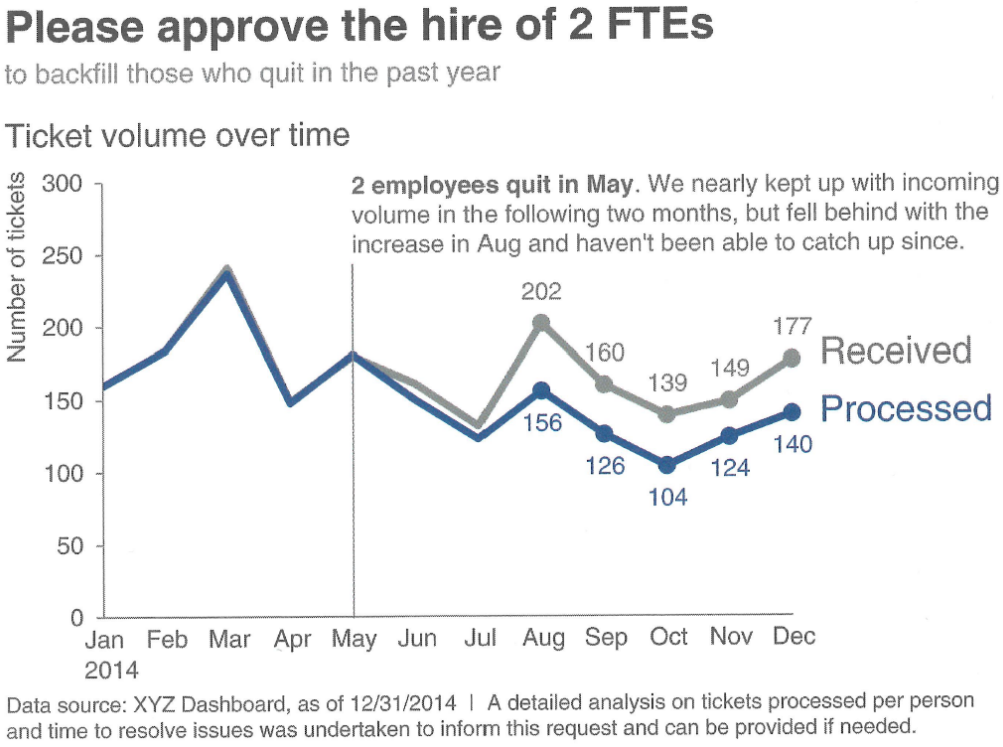
Represent the data



Refine the visualization

Improve the basic representation to make it clearer, more meaningful and more visually engaging





Support interactivity

Optional step (depending also on the format)

Add methods for manipulating the data or controlling what features are visible

Just because you can, doesn't mean you should

Interactivity should support accessibility (help understanding)

Schneiderman's mantra: *overview first, zoom and filter, then details on demand*

Basic charts

Basic charts

- Line chart
- Bar charts
- Pie charts
- Geographical data
  - Dot maps
  - Choropleth maps
- Networks and trees
  - Node-link diagrams
  - Matrices

Line charts

Use them to show how values develop over time (or some other continuous value)

Do not use them for categories

Place the labels close to the data

Extend the y-axis to 0 (or the ‘historic low’ value)

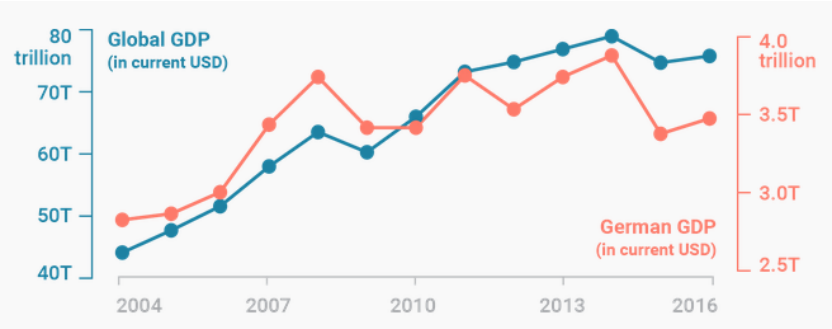
- If the data comes close to 0
- If 0 has a meaning

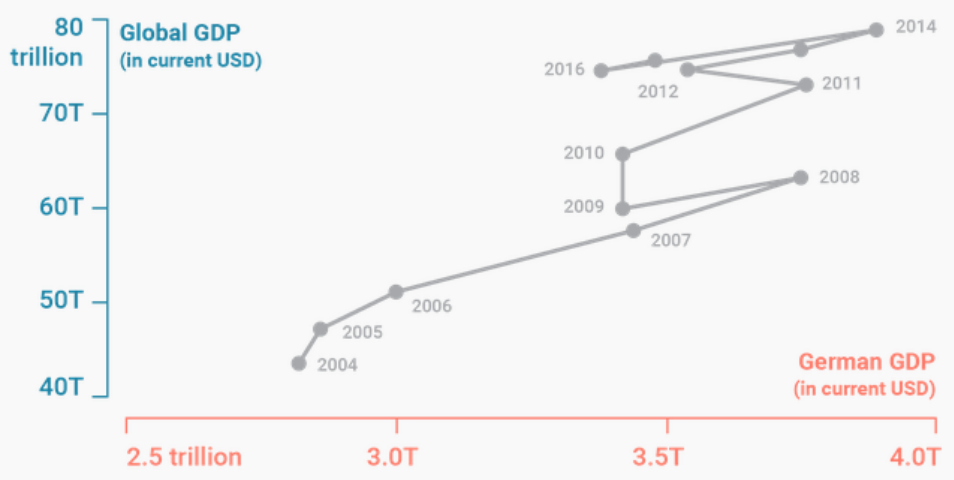
Line charts

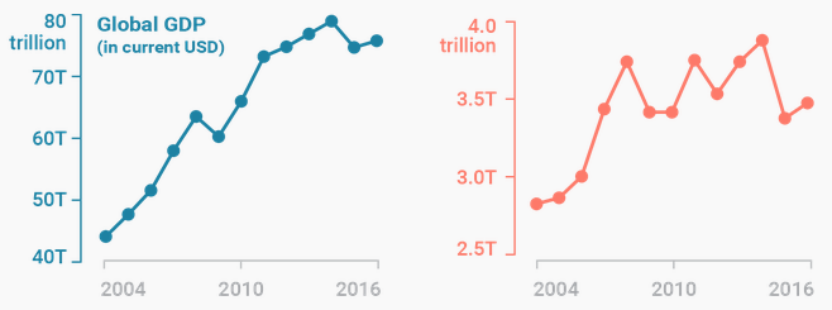
Avoid dual axis charts

Alternatives

- Side-by-side charts
- Connected scatter plots







https://blog.datawrapper.de/dualaxis/

Bar charts

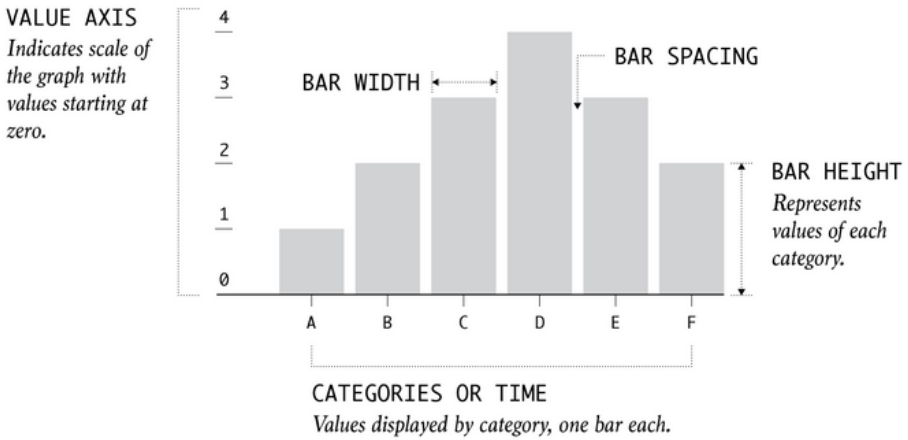
Use them to show values per categories (or discrete time)

They should always have a 0 baseline

If you use (many) categories, sort the bars by value

If the labels are very long, use a horizontal bar chart instead of a vertical one

No 3-D

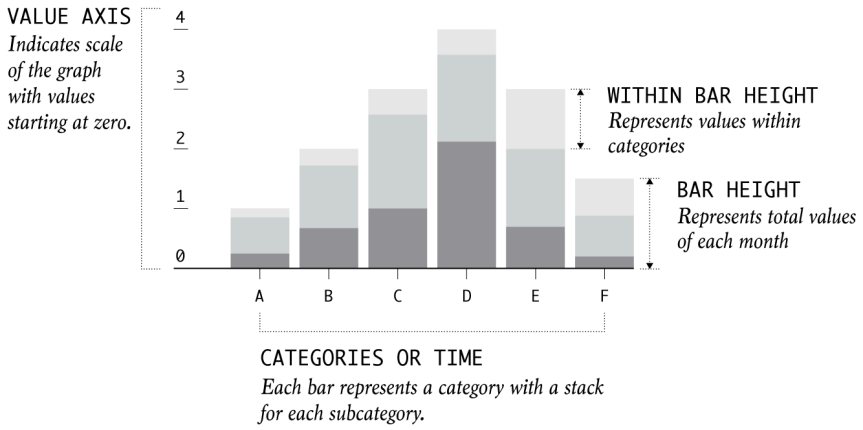


Stacked bar charts

Same rules apply as for regular bar charts

Use them when you are mostly interested in totals (and the bottom category)

If they add up to 100% , you can easily compare only the values in the bottom/top category



Pie charts

Use them to show parts that sum up to 100%

Show the values for each slice

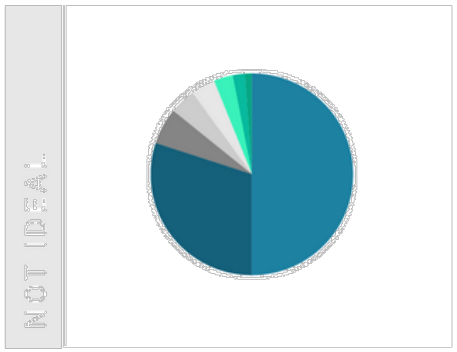
Show only a few (up to 4 or 5) categories

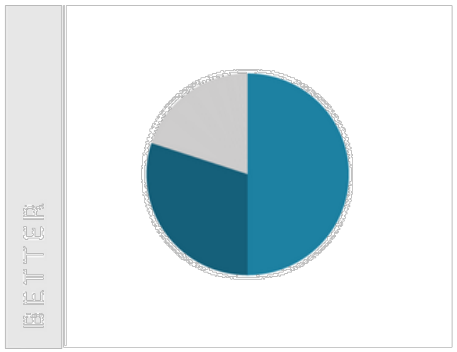
- Group smaller slices together as ‘other’
- Label small slices outside of the chart

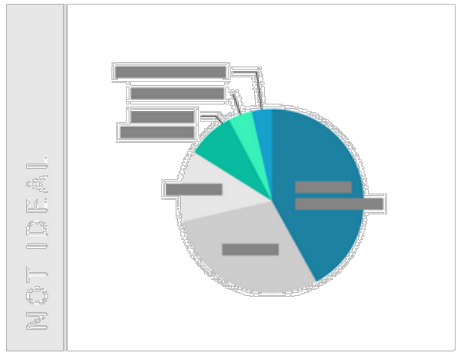
Start on top (at ‘12h’), sort the slices by size

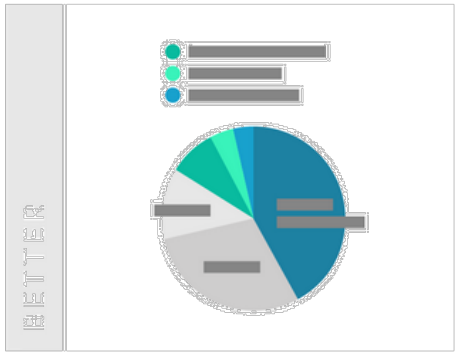
If the slices are of similar size, use a bar chart instead

No 3-D









Geographical data

Use maps only when the spatial relationship is important

Extremely important because space is the most effective visual channel and you do not want to waste it for spatial information if not relevant

Dot maps

Also called *dot distribution maps*

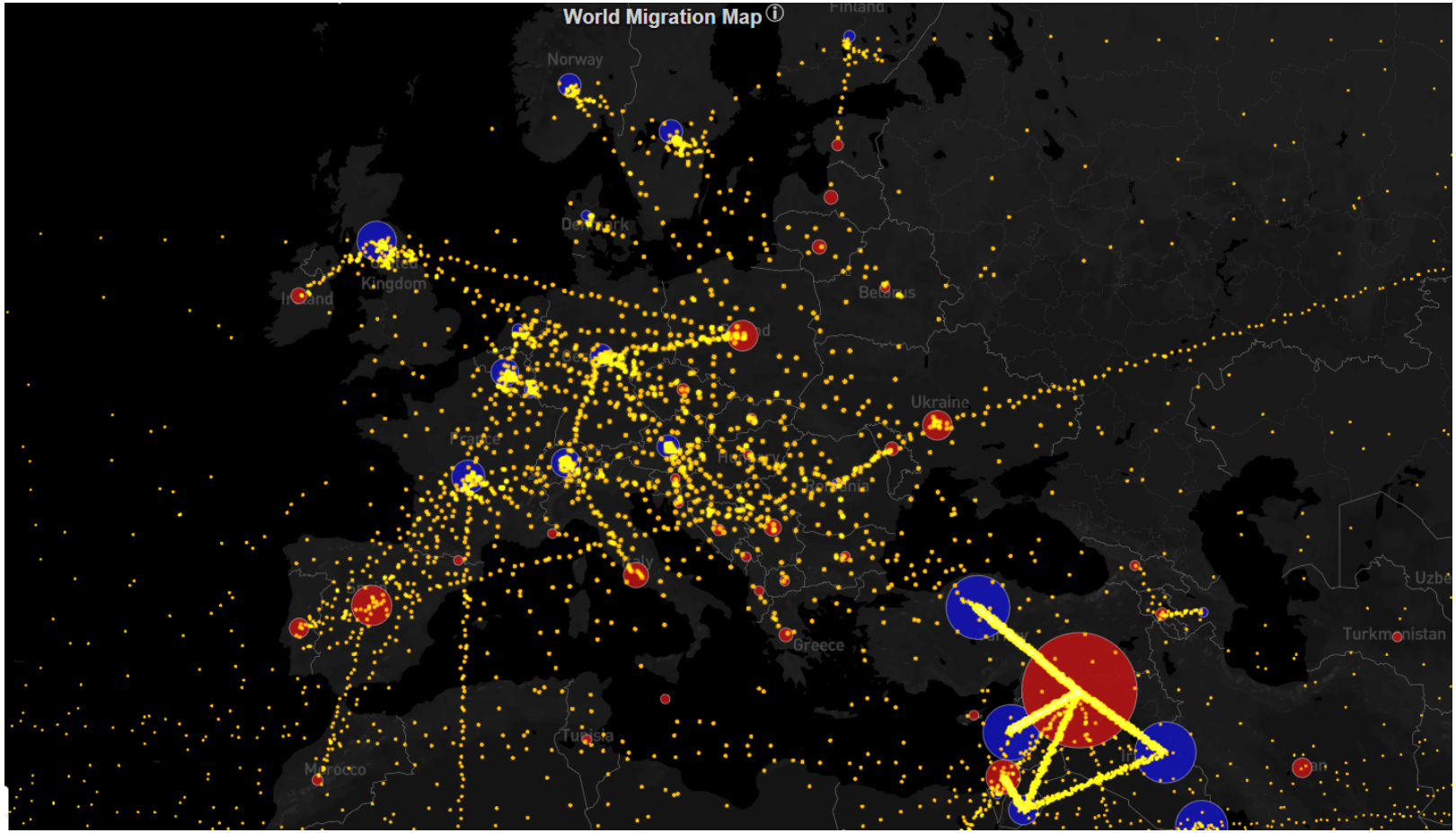
Use them to show how things are distributed over a geographical region

Can reveal patterns when the points cluster on the map

Could just be showing population density (!)

Use size and color to convey additional information

Dot maps



http://metrocosm.com/global-migration-map.html

Choropleth maps

Use them to show the spatial relationship of categorical or numerical data

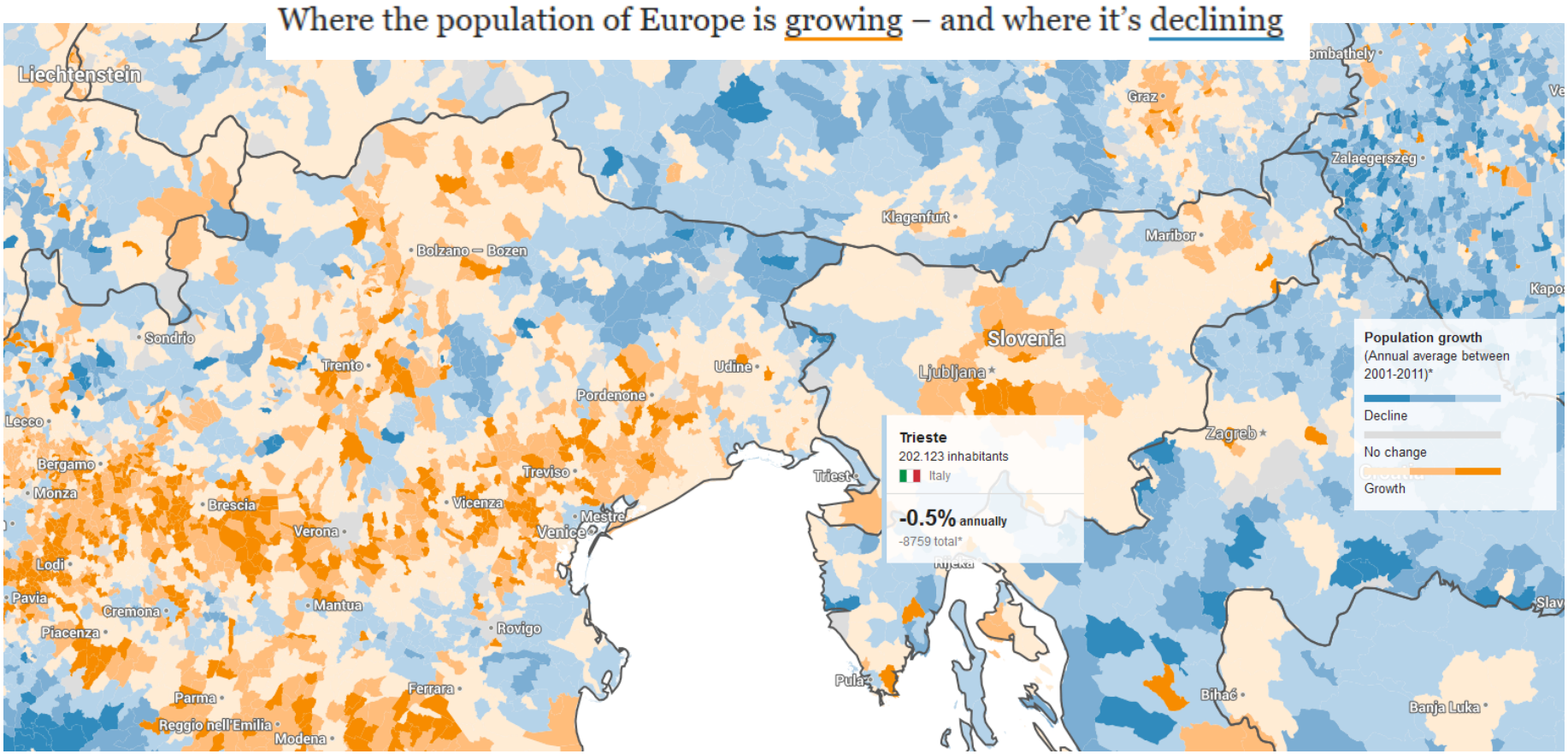
Size of the objects depends on geography not on the variables of interest

Show relative instead of absolute data

Be careful in choosing bin size

Be careful in choosing colors

Choropleth maps



https://interaktiv.morgenpost.de/europakarte/#8/45.783/12.409/en

Networks and trees

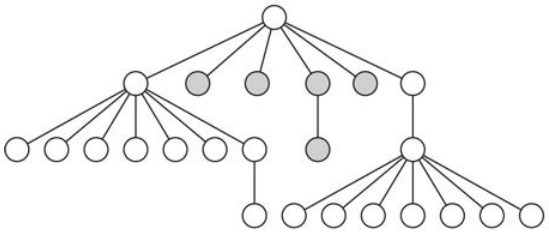
Network and trees are relational structures characterized by a collection of nodes and links that connect the nodes

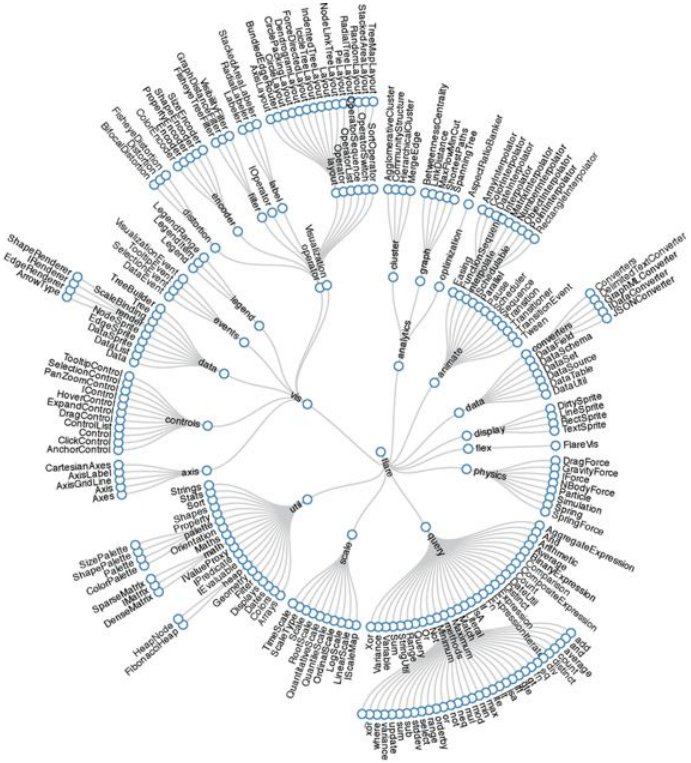
Nodes and links can also have attributes associated to them

Node-link diagrams

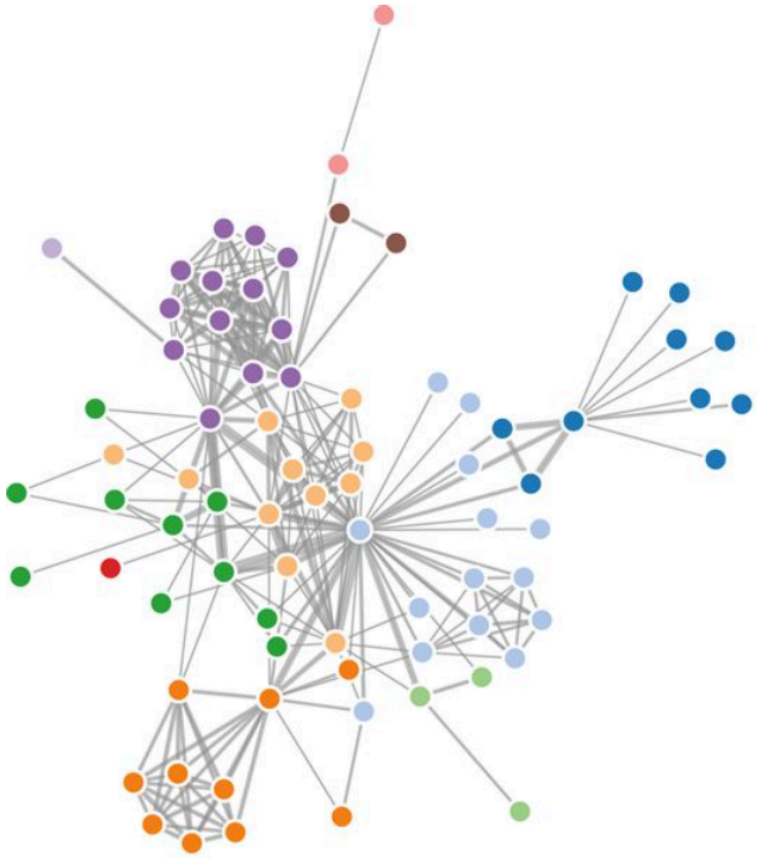
Layout depends on size

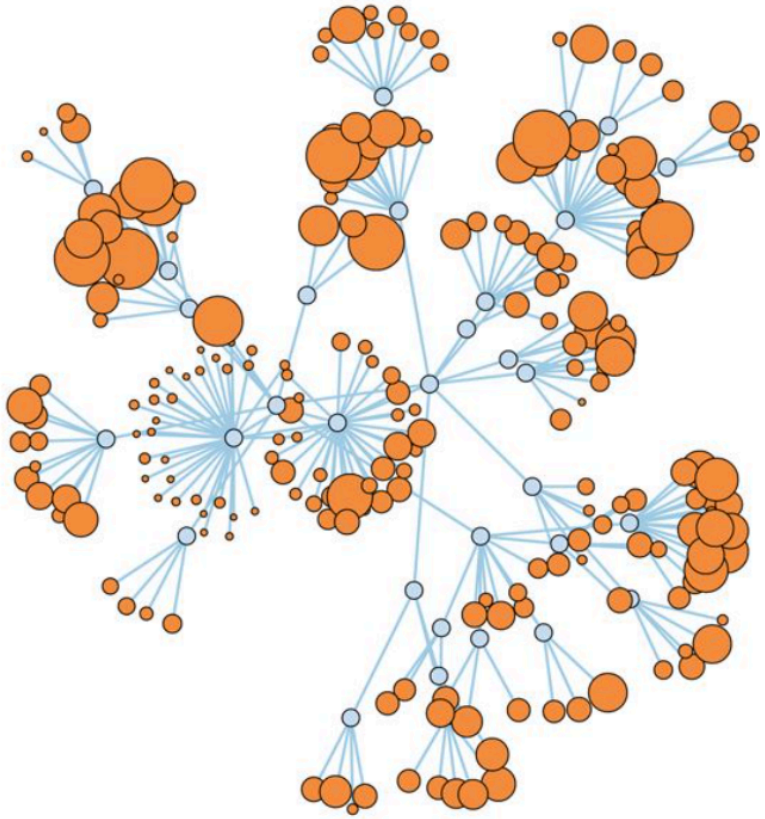
- Triangular vertical (small trees)
- Spline radial (large trees)



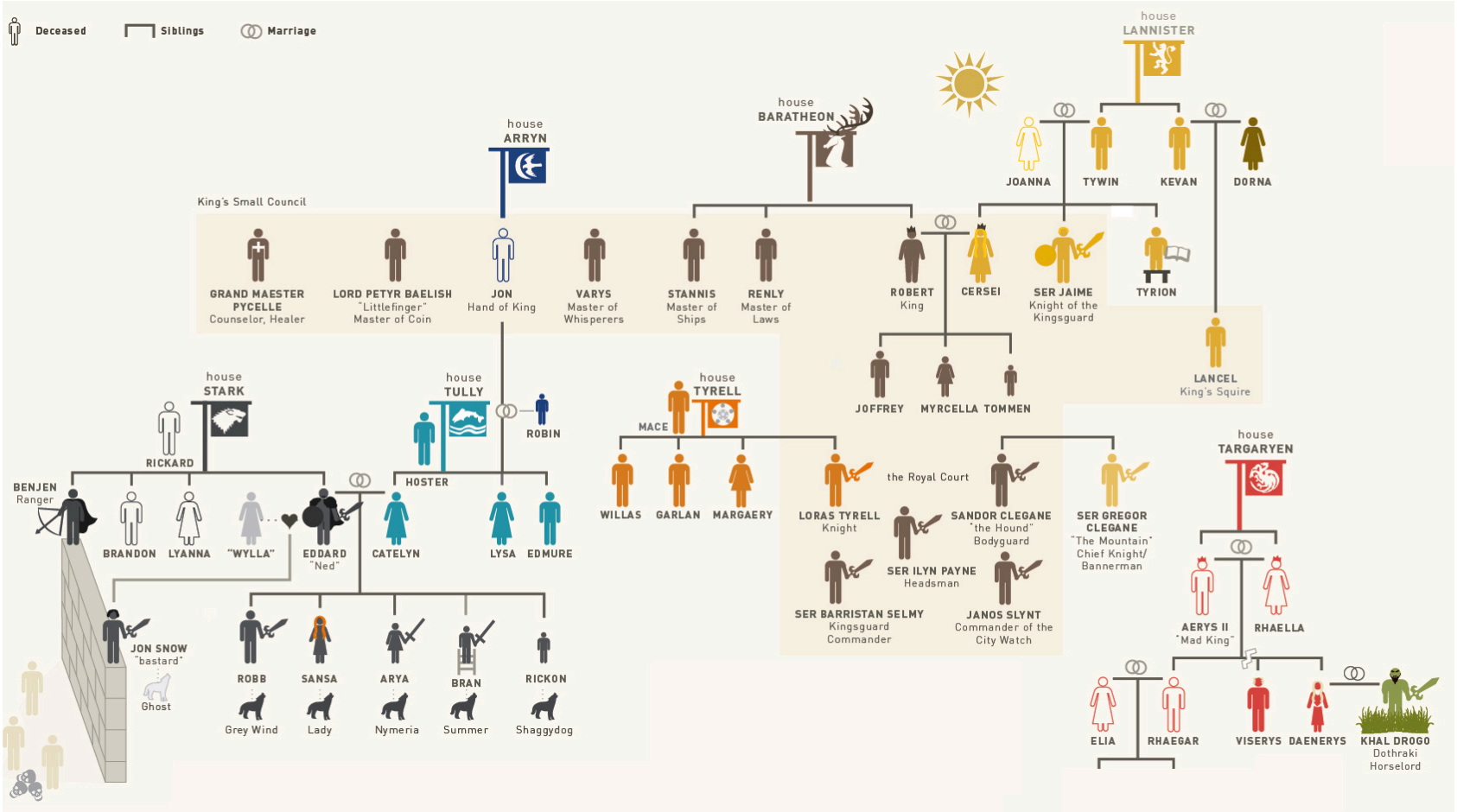


Node-link diagrams



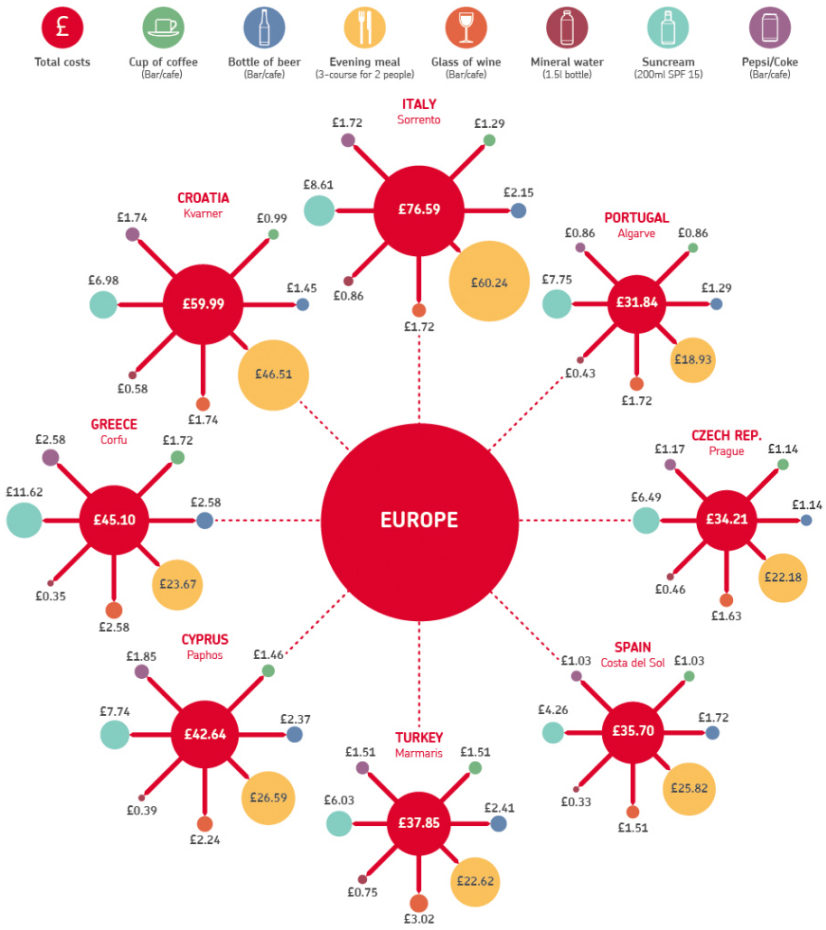


Node-link diagrams



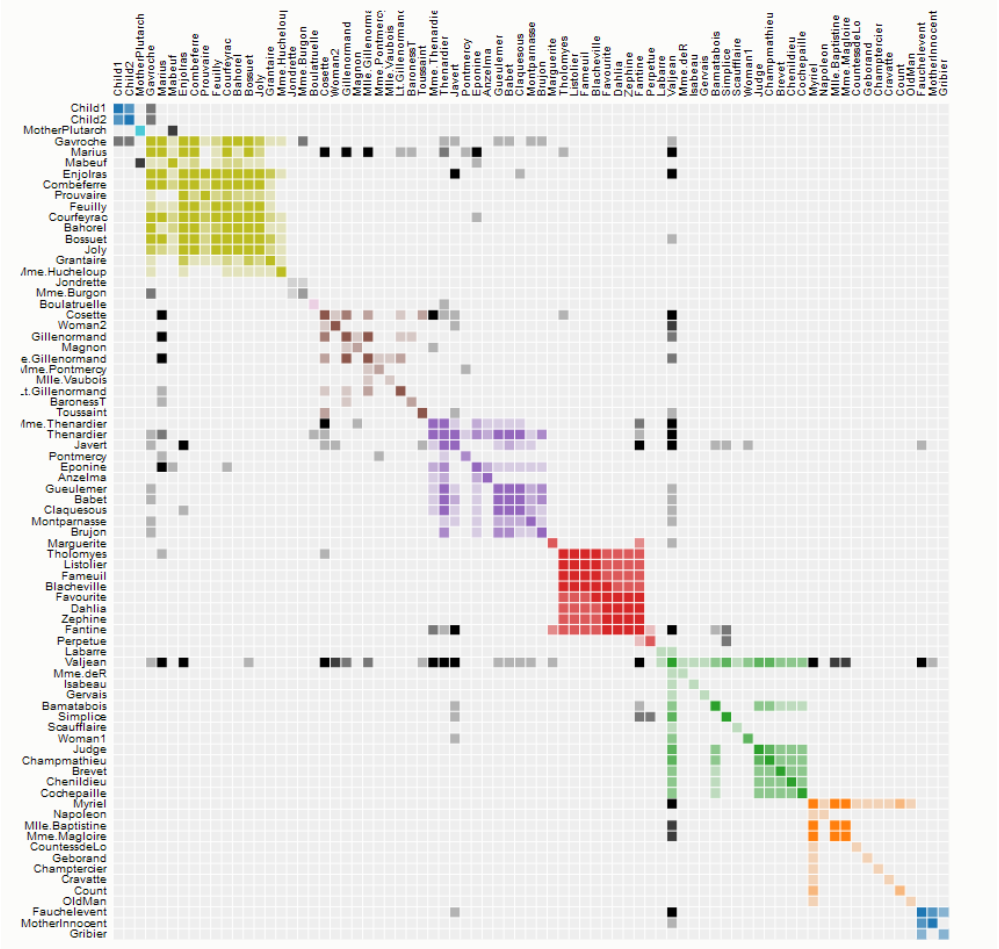
http://www.mahina.se/2011/05/infographic-game-of-thrones-family-tree/

Node-link diagrams



https://www.informationisbeautifulawards.com/showcase/530-worldwide-holiday-costs-barometer-2014

Adjacency matrix



Co-occurrence of characters in Les Misérables

https://bost.ocks.org/mike/miserables/

Multivariate/ multidimensional data visualization

Multivariate/multidimensional data visualization

Visualize all variables at the same time

- Chernoff faces
- Bubble chart (small number of dimensions)
- Scatter plot matrix
- Parallel coordinate plot
- Radar chart
- Radial histogram
- Small multiples
- Horizon charts

Perform dimensionality reduction and visualize the results

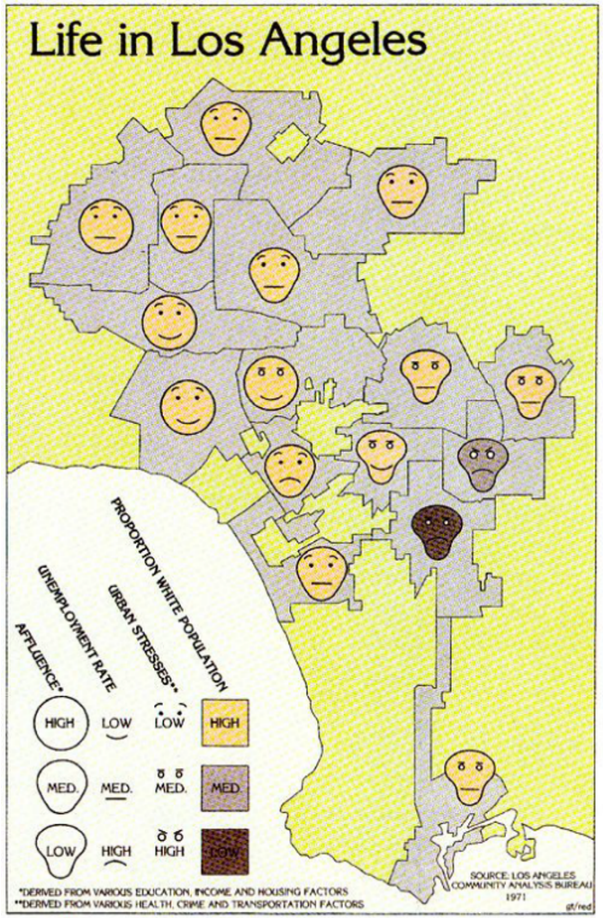
Chernoff faces

Visualization using glyphs

Can present up to 18 distinct variables

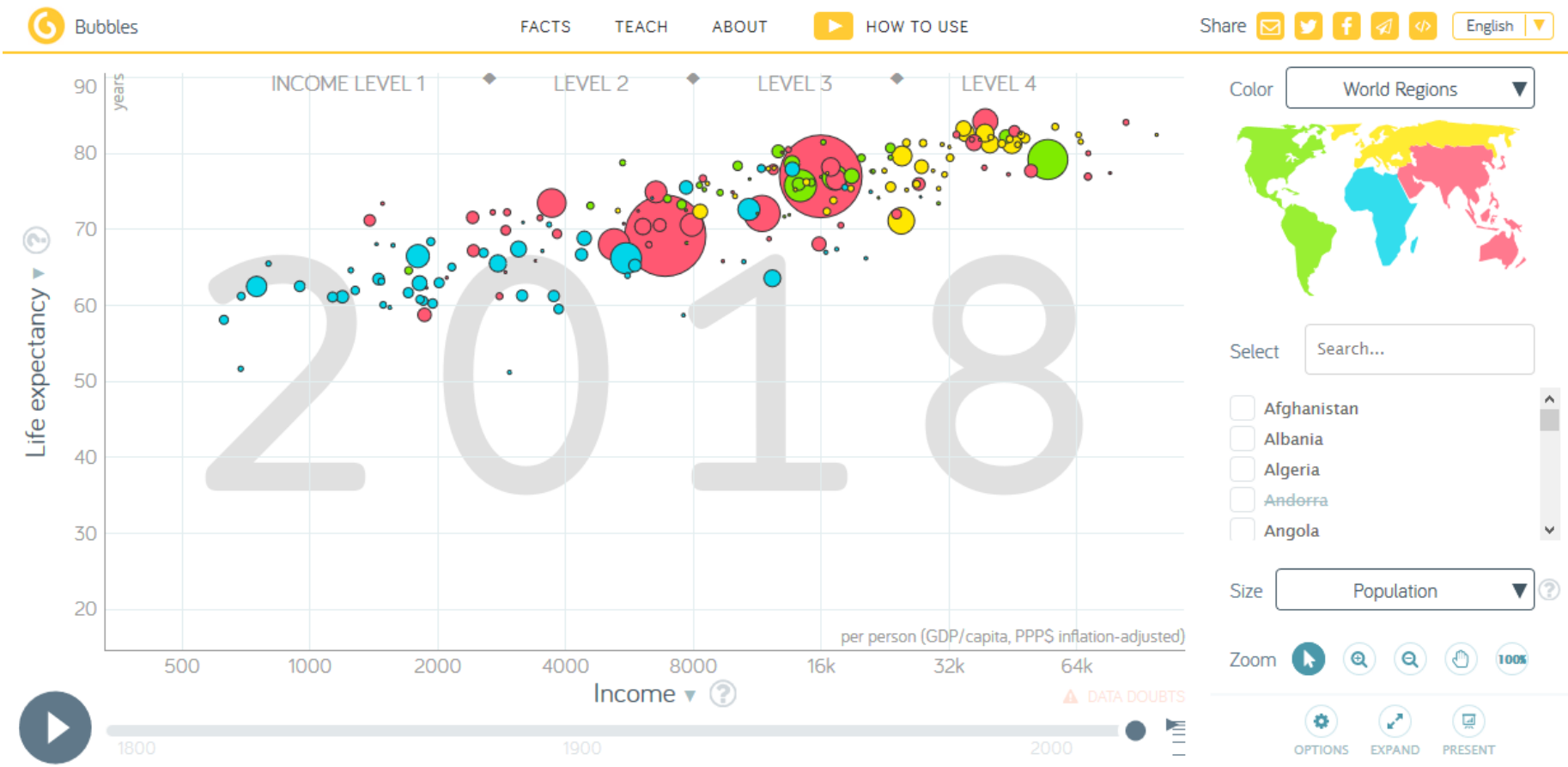
- Size
- Curvature
- Position of the eyes
- Position of the mouth
- …

Questionable generalization



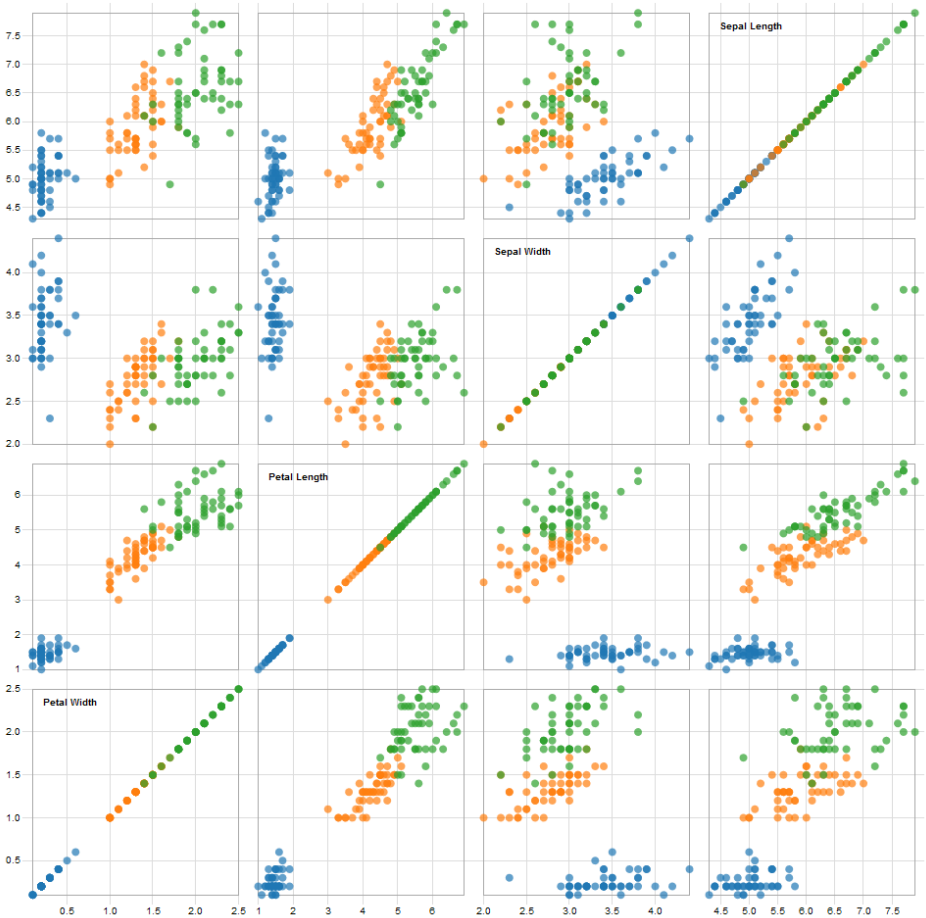
https://mapdesign.icaci.org/2014/12/mapcarte-353365-life-in-los-angeles-by-eugene-turner-1977/

Bubble chart



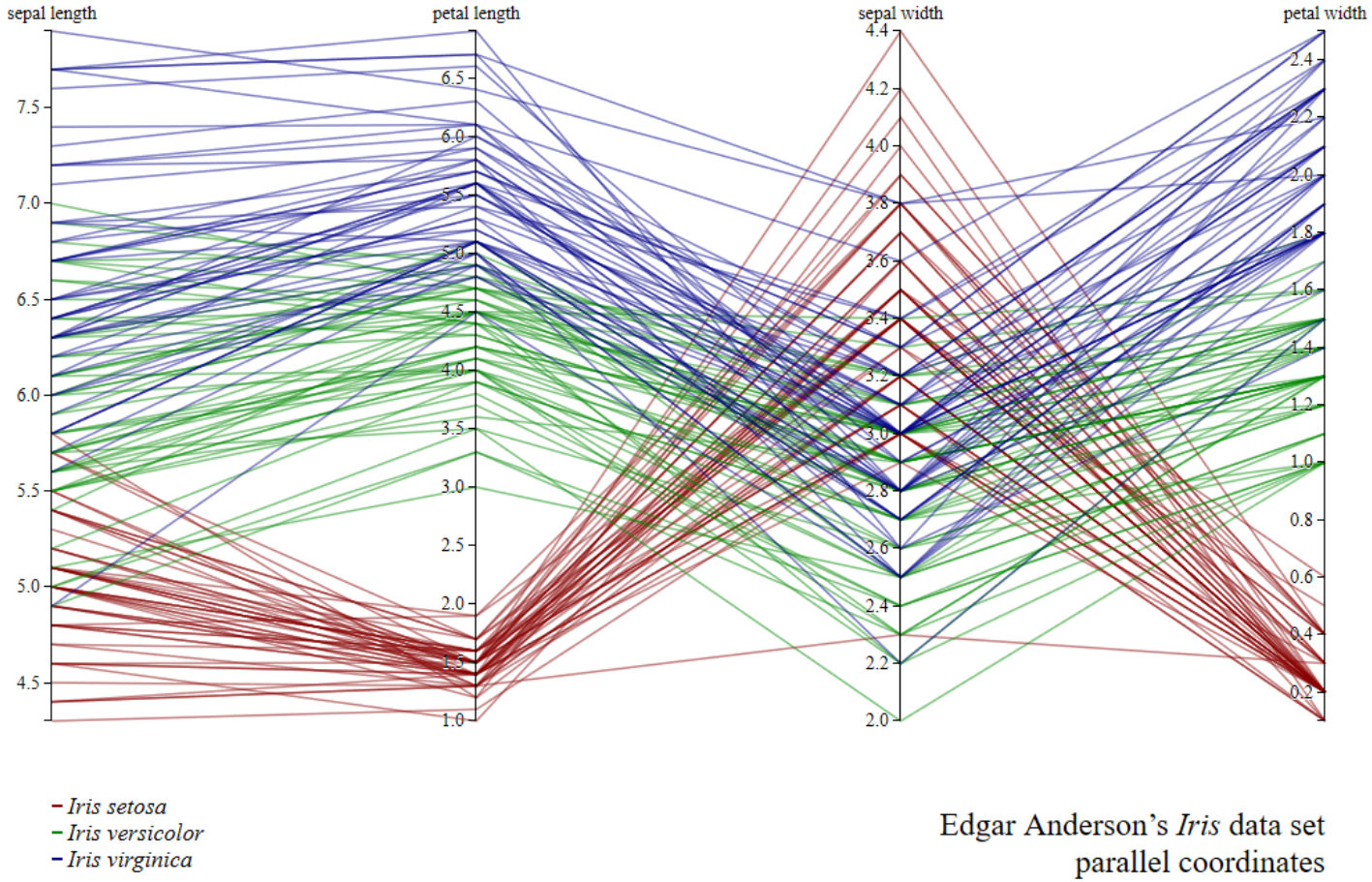
https://www.gapminder.org/tools/

Scatter plot matrix



https://bl.ocks.org/mbostock/4063663

Parallel coordinate plot



http://mbostock.github.com/d3/talk/20111116/iris-parallel.html

Radar chart

https://vizzlo.com/create/radar-chart

Radial histogram

OECD Regional Well-Being

OECD countries / Italy

Friuli-Venezia Giulia

Health

Friuli-Venezia Giulia reaches **9.2** / 10 points in **Health.**

0 · · · · 5 · · · · 10

This puts the region in position **11** / 21 regions in Italy.

0 · · · · 5 · ·

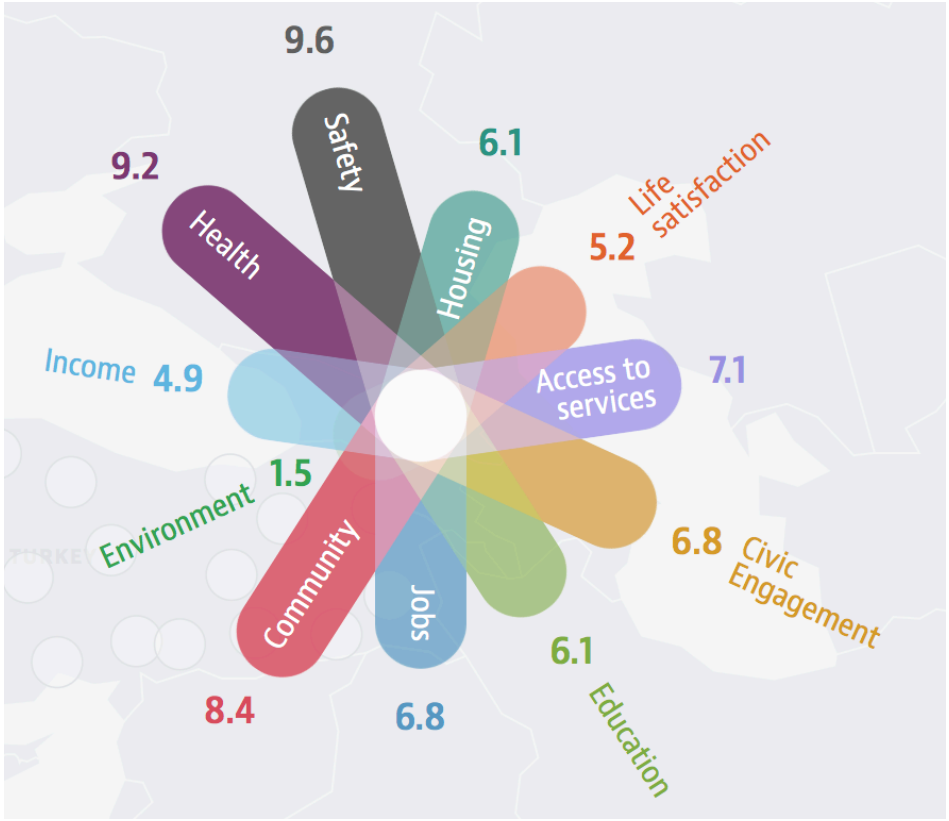
Compared across all OECD regions, the region is in the **top 10%** in **Health.**

bottom 20% median top 20%

Indicators

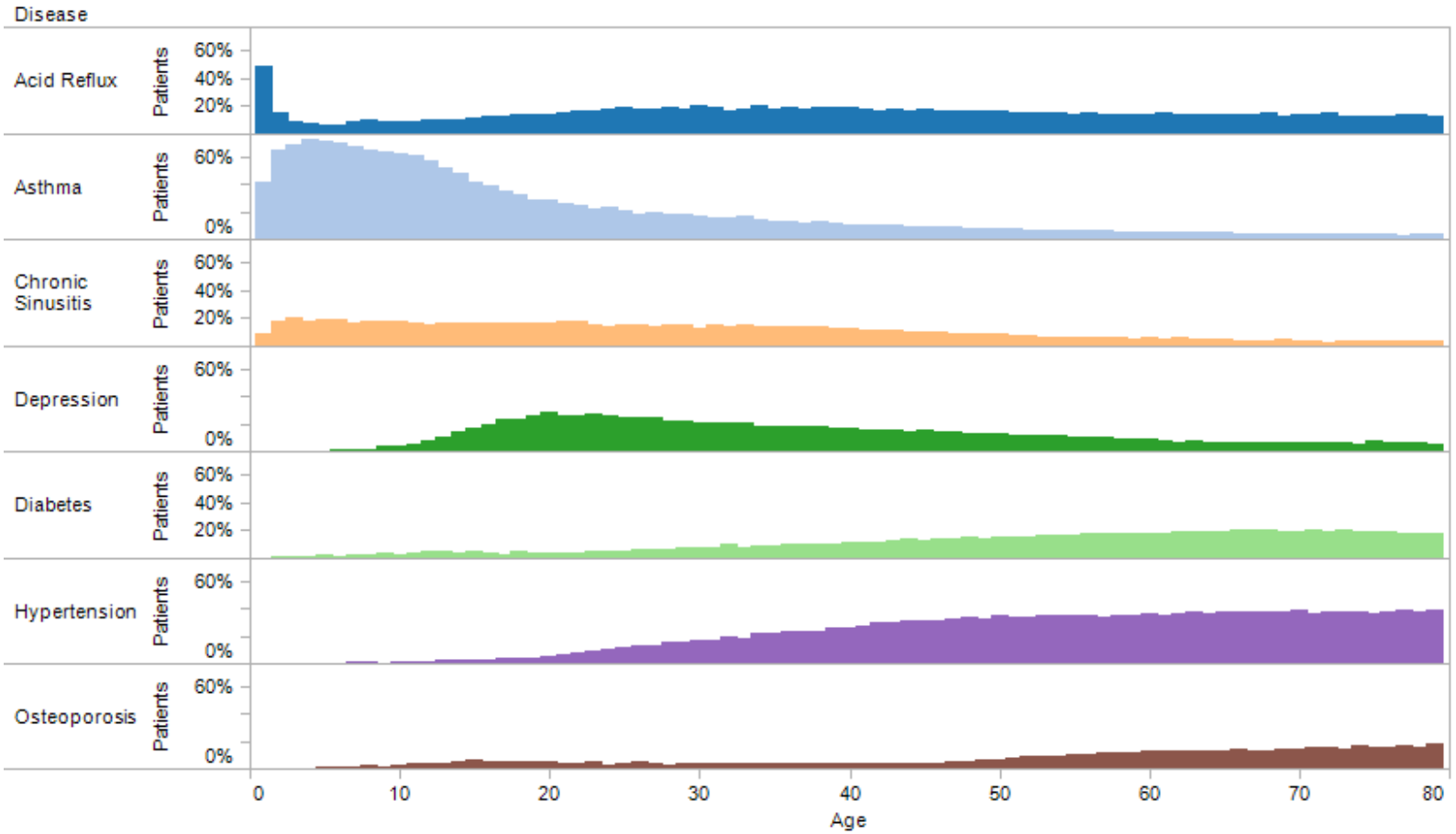
Mortality rate: **6.6** deaths per 1 000 people

Life expectancy: **83.5** years



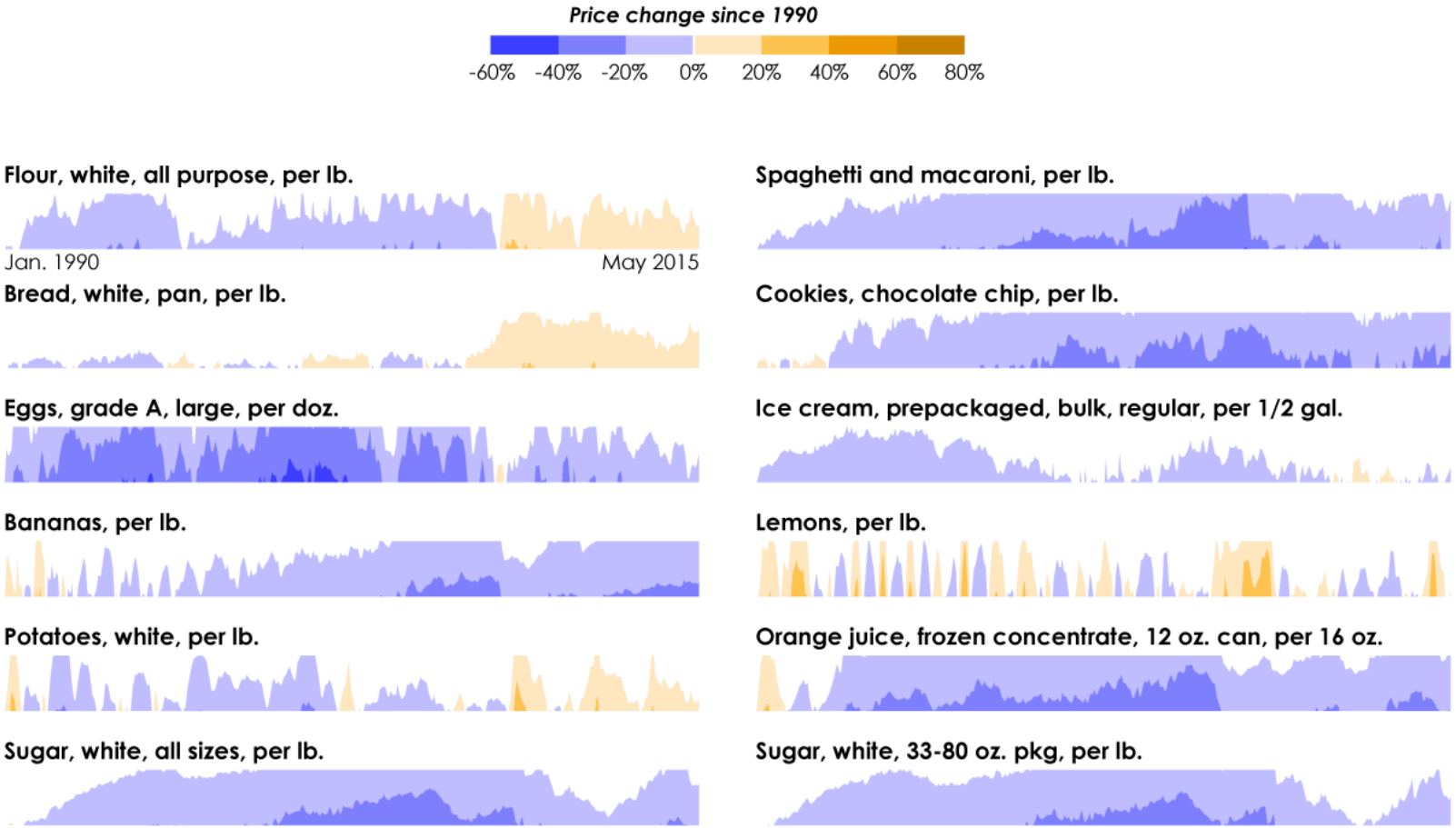
https://www.oecdregionalwellbeing.org/ITH4.html

Small multiples



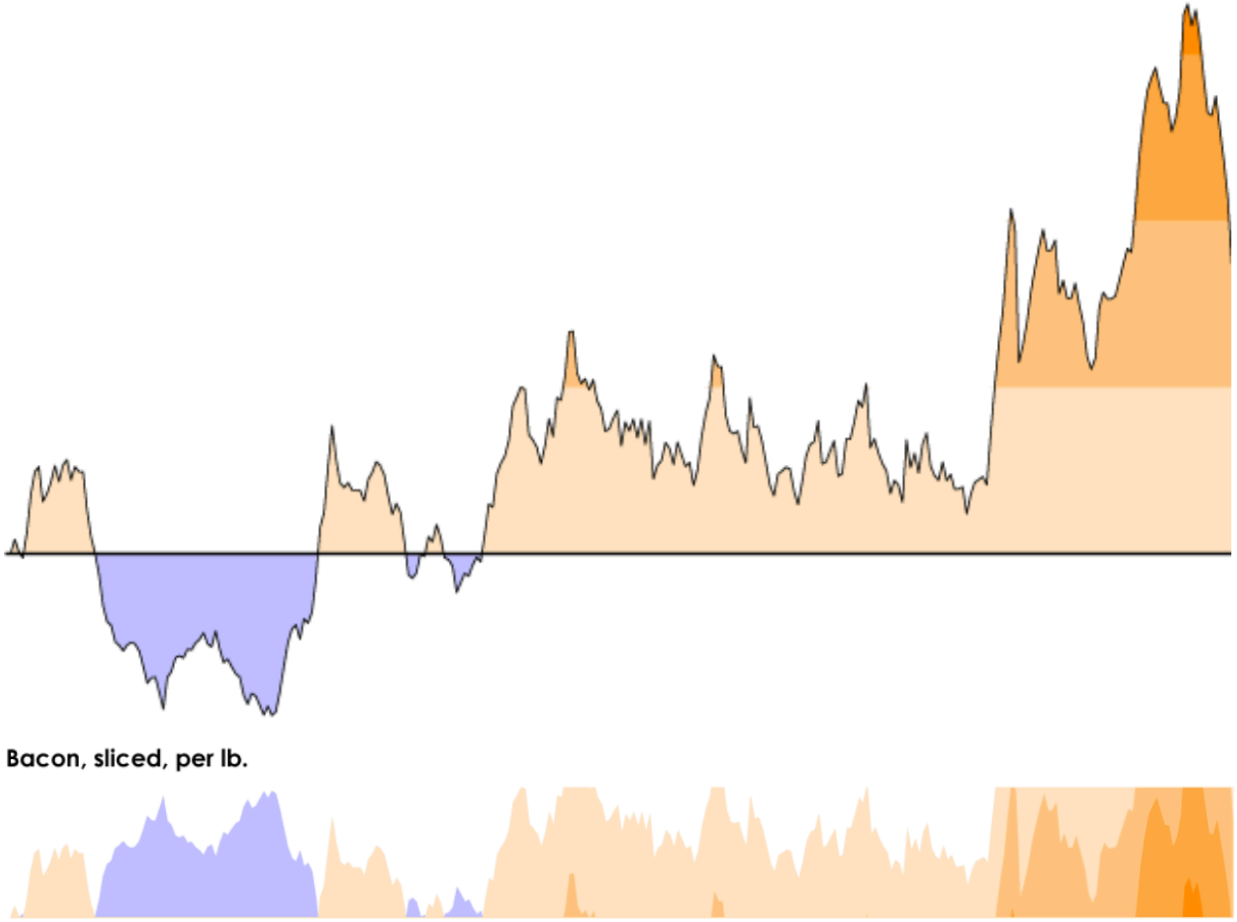
https://eagereyes.org/criticism/curing-a-sick-chart

Horizon charts



https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/

Horizon charts



https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/

Multivariate/multidimensional data visualization

Perform dimensionality reduction and visualize the results

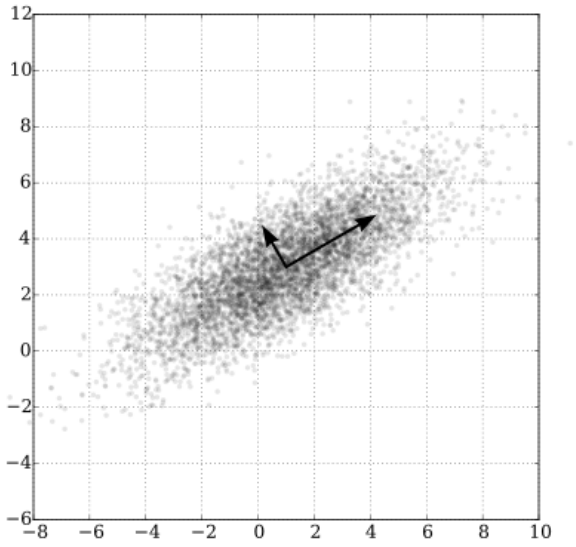
- Principal component analysis
- Multidimensional scaling

Transformation $R^n \rightarrow R^2$

Principal component analysis

PCA uses an orthogonal transformation $R^n \rightarrow R^2$

- First principal component has the largest possible variance
- Second principal component is orthogonal to the first one and has the largest possible variance

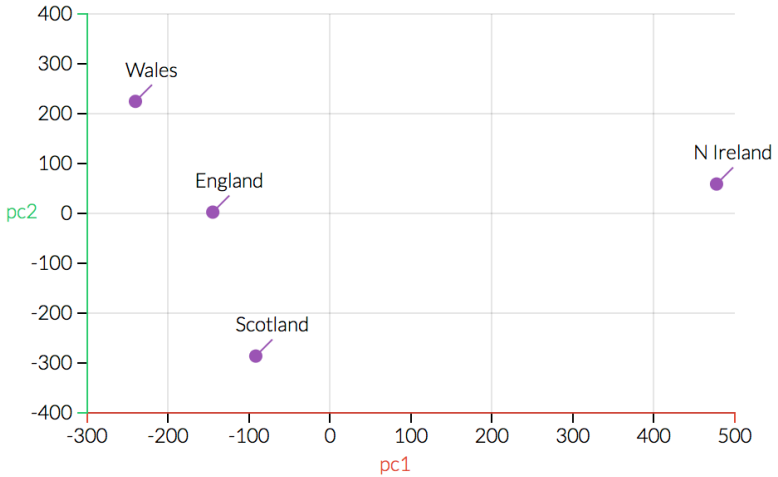


https://en.wikipedia.org/wiki/Principal_component_analysis

Principal component analysis

Eating in the UK

Consumption of 17 types of food in grams per person per week for every country in the UK



	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

Multidimensional scaling

A nonlinear transformation $R^n \rightarrow R^2$ that tries to preserve distances between data points

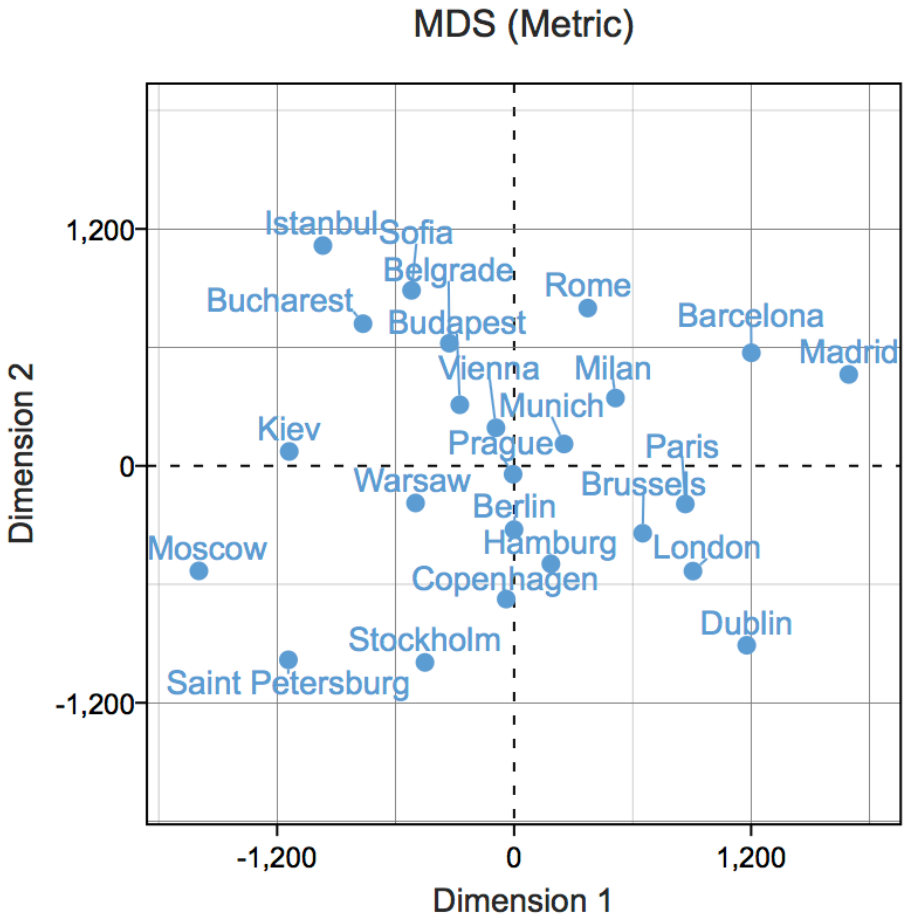
Useful for visualizing similarity matrices or graphs where you wish to preserve distances between nodes

Minimize the stress function

$$S = \sum_{i,j}(d_{ij} - d_{ij}^*)^2$$

Solve with any method for optimizing nonlinear functions

Multidimensional scaling



https://www.displayr.com/goodness-of-fit-in-mds-and-t-sne-with-shepard-diagrams/

Visualizing uncertainty and missing data

Visualizing uncertainty

Uncertainty (confidence intervals, etc.) hard to understand

Uncertainty types

- Spatial uncertainty
- Temporal uncertainty
- Cardinality
- Categorical uncertainty
- Source quality

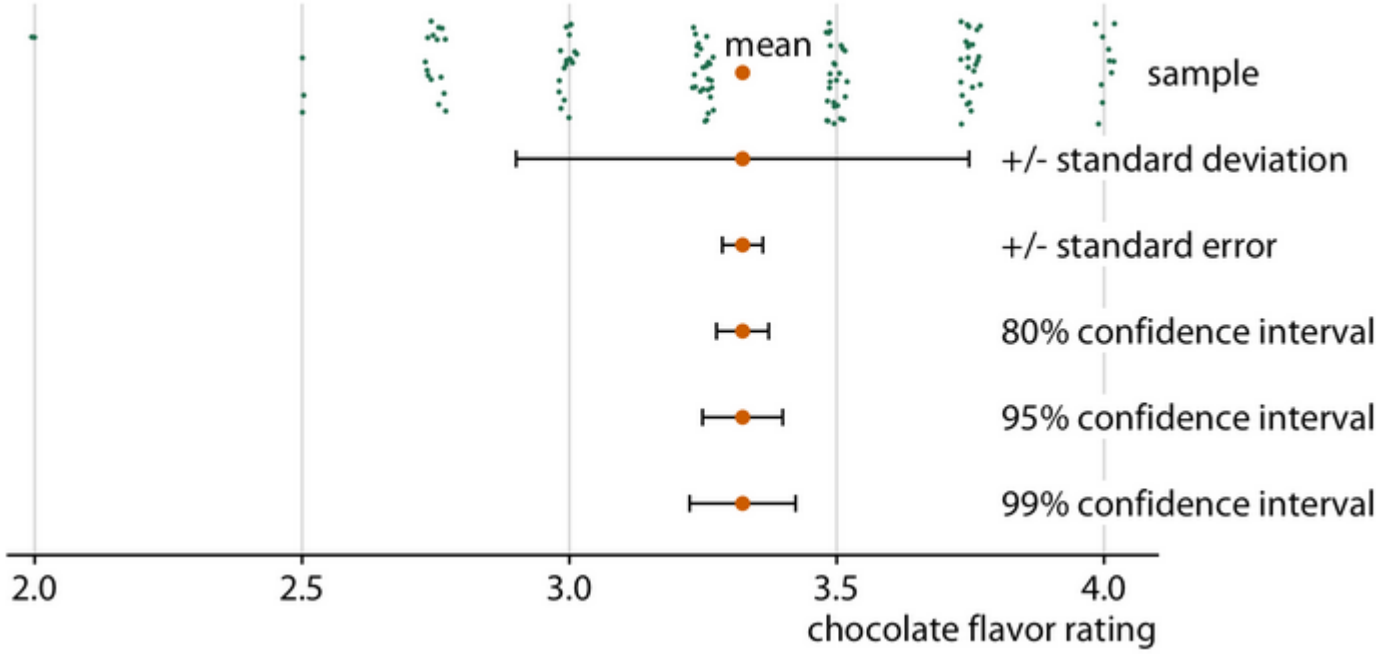
Visualizing uncertainty

Techniques to show uncertainty

- Ranges
- Distributions
- Multiple outcomes
- Obscurity

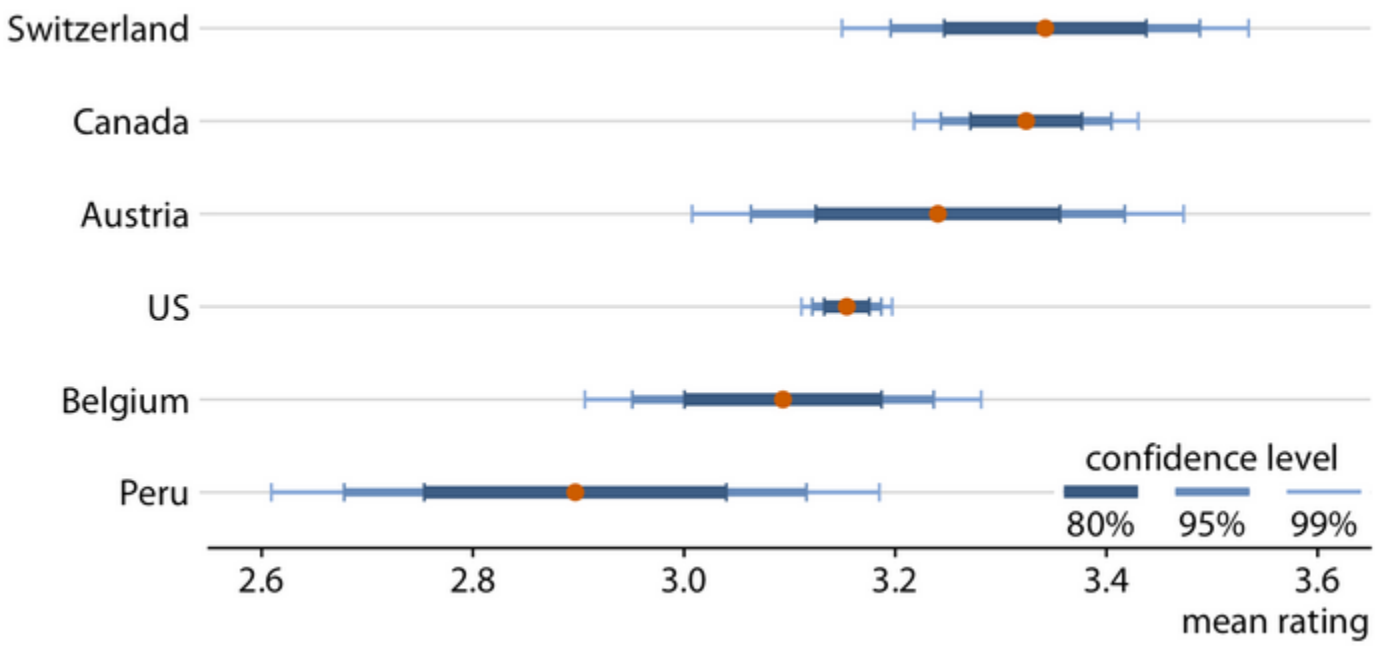
Visualizing uncertainty with ranges

Specify what the range represents

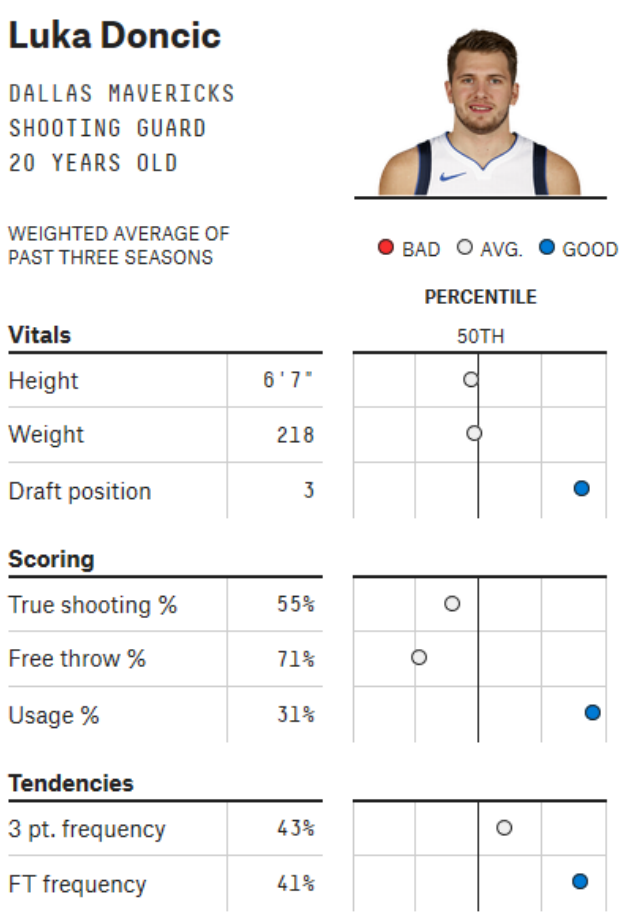


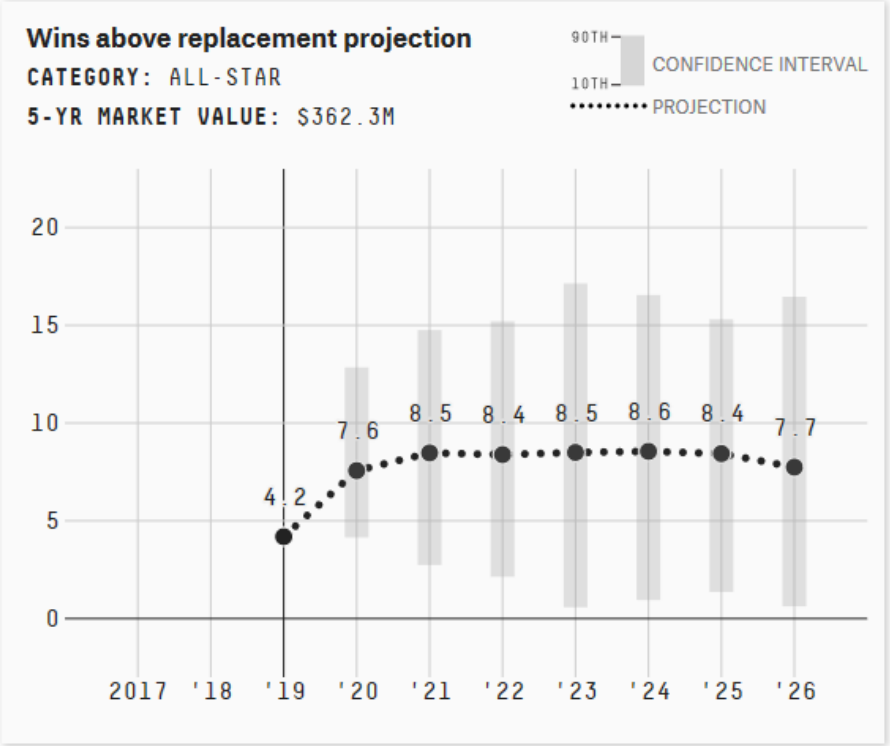
Visualizing uncertainty with ranges

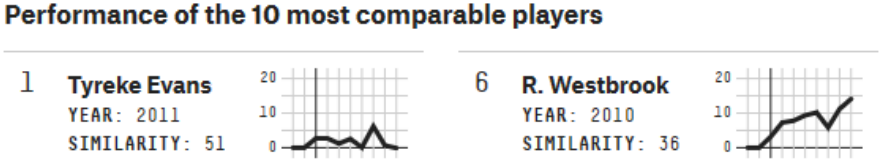
Specify what the range means



Visualizing uncertainty with ranges

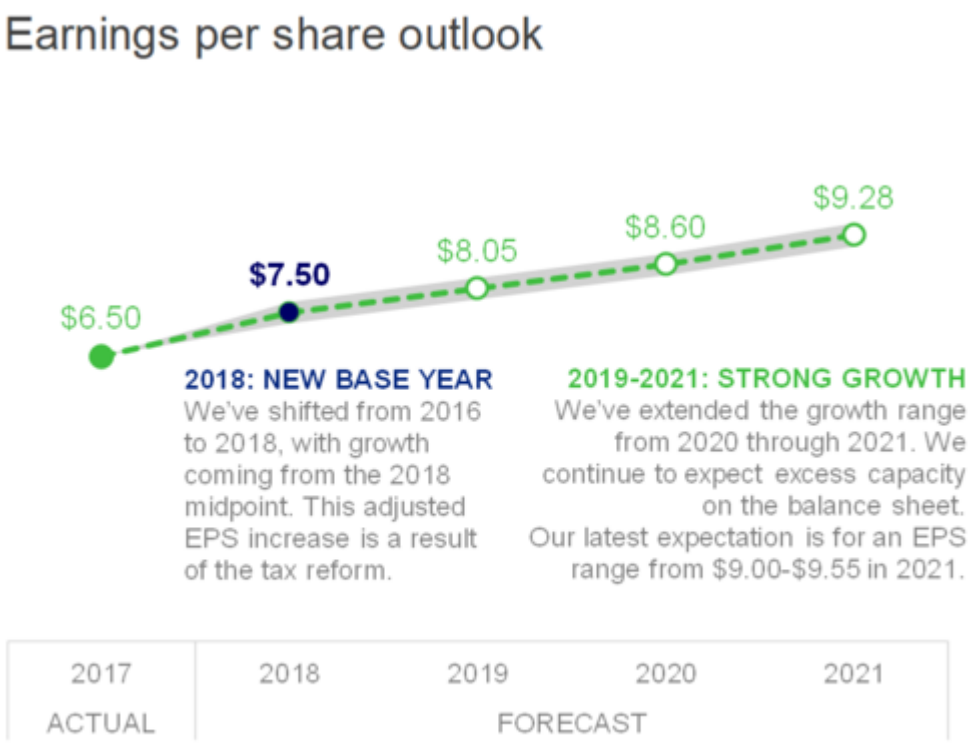






https://projects.fivethirtyeight.com/carmelo/luka-doncic/

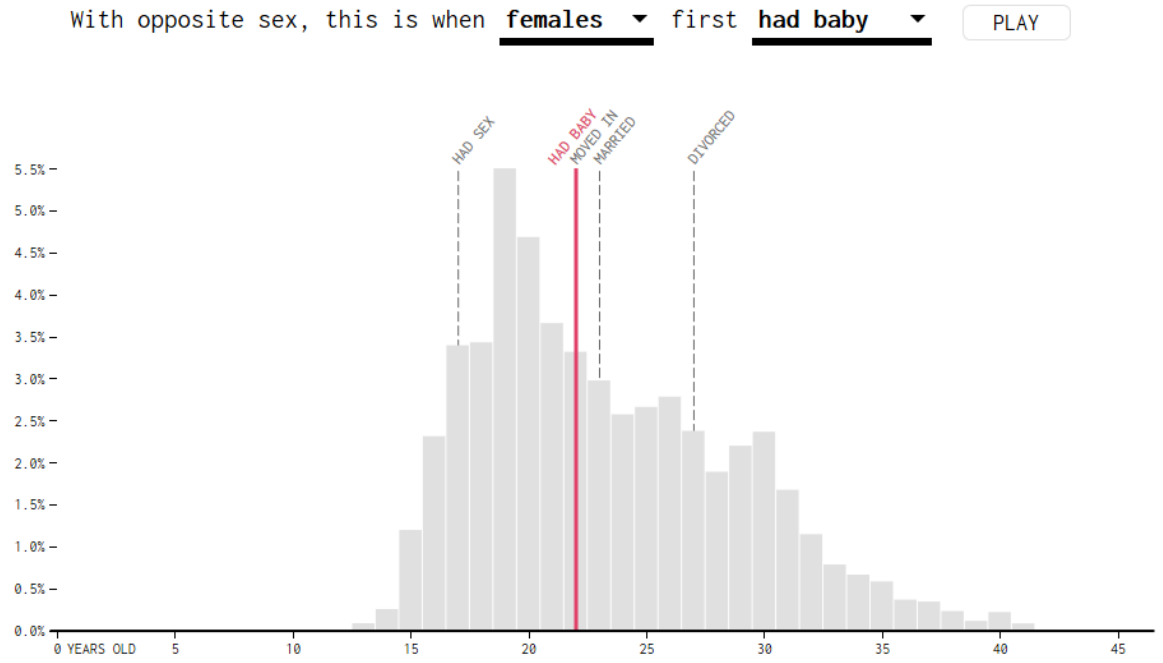
Visualizing uncertainty with ranges



https://www.storytellingwithdata.com/blog/2018/6/27/visualizing-uncertainty

Visualizing uncertainty with distributions

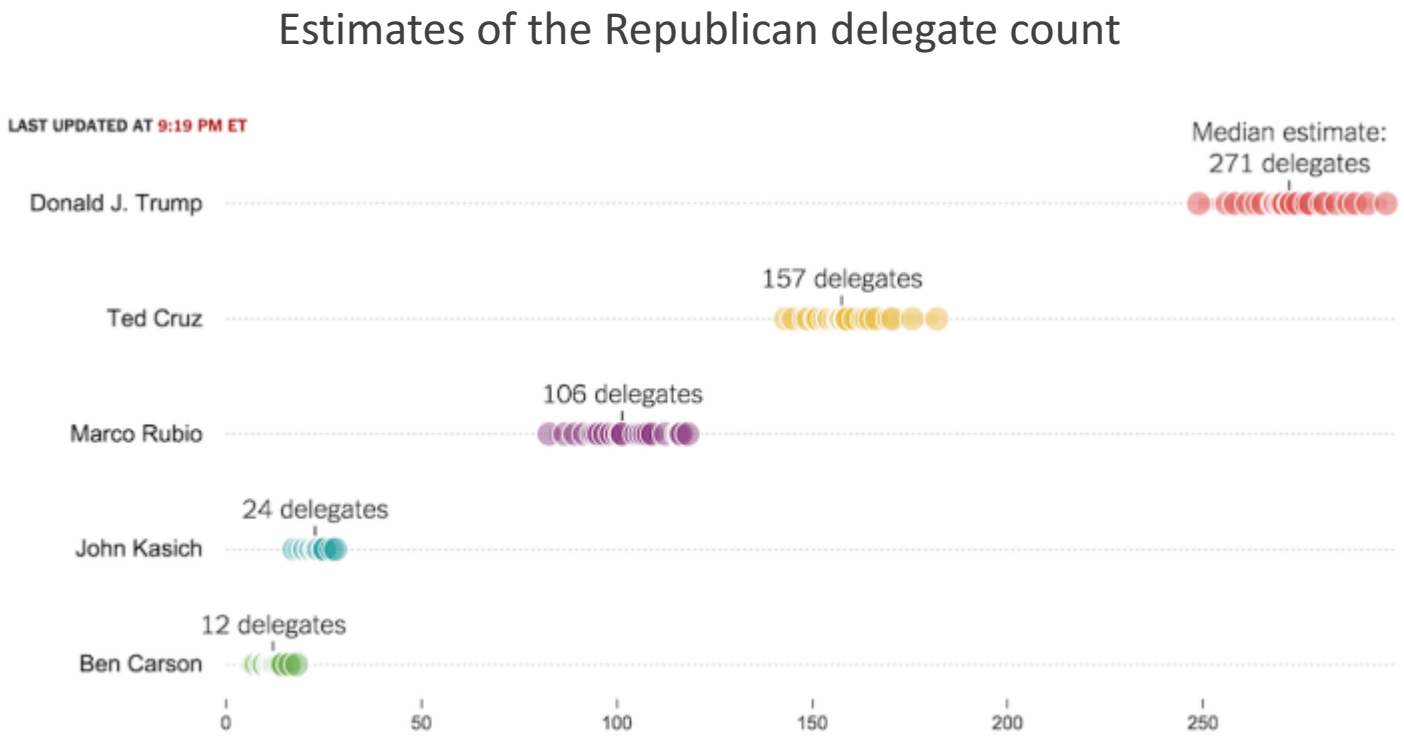
Show the spread of possible values with a histogram (or a variant of it)



https://flowingdata.com/2017/02/23/the-first-time/

Visualizing uncertainty with multiple outcomes

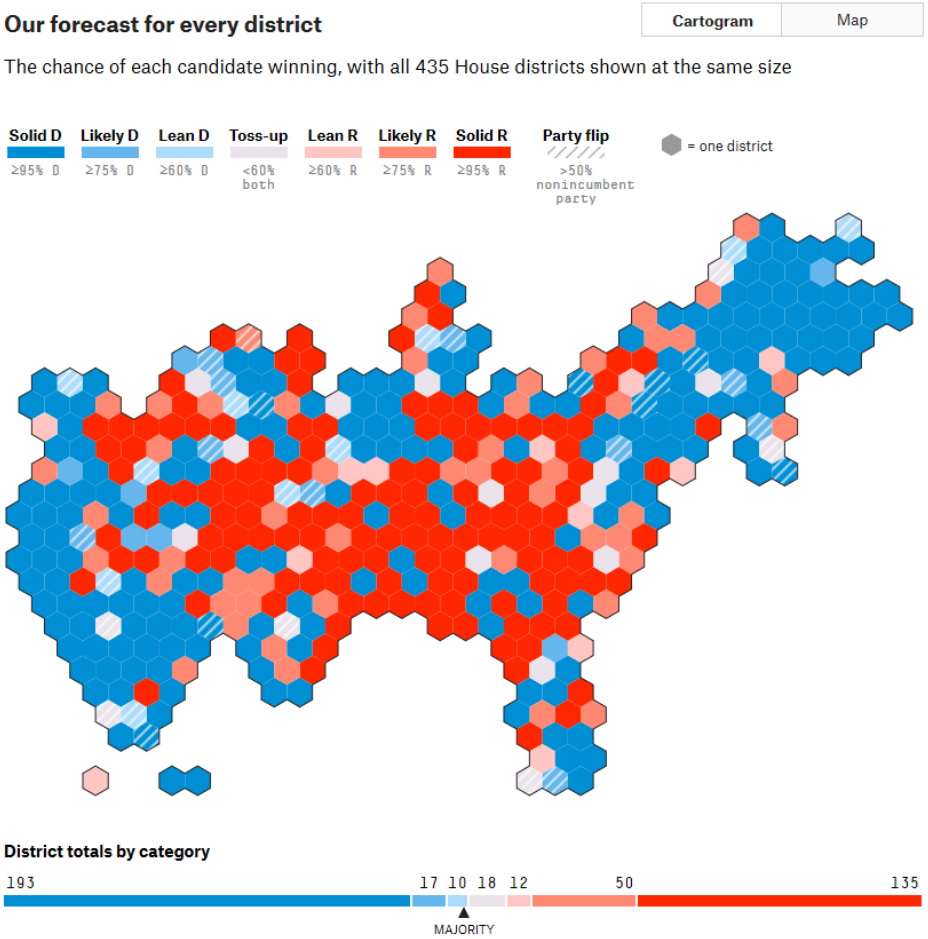
Show the various outcomes



https://www.nytimes.com/interactive/2016/03/01/upshot/super-tuesday-live-republican-delegate-estimates.html?mtrref=undefined

Visualizing missing data with obscurity

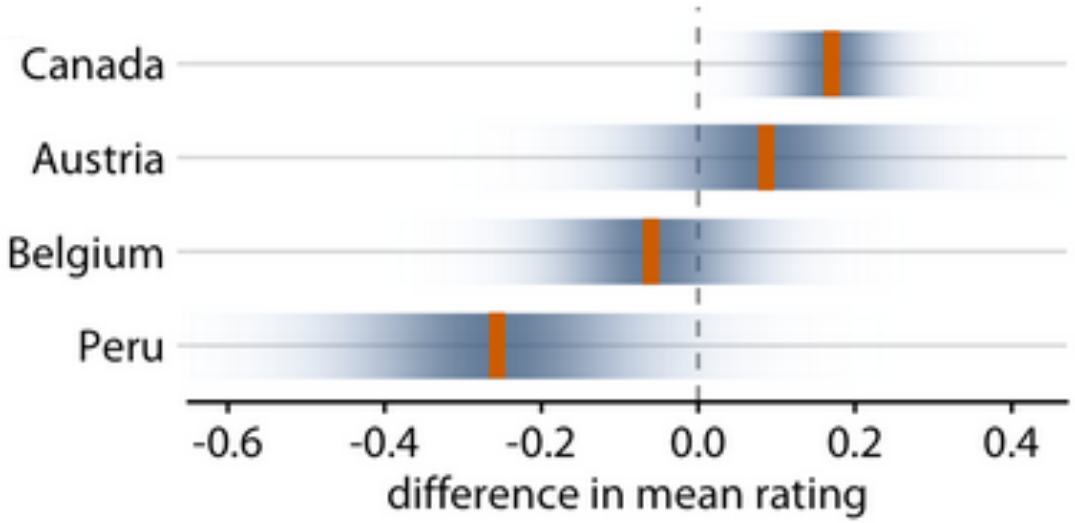
Use transparency, color scale, or blurriness to show uncertainty



https://projects.fivethirtyeight.com/2018-midterm-election-forecast/house/

Visualizing missing data with obscurity

Use transparency, color scale, or blurriness to show uncertainty



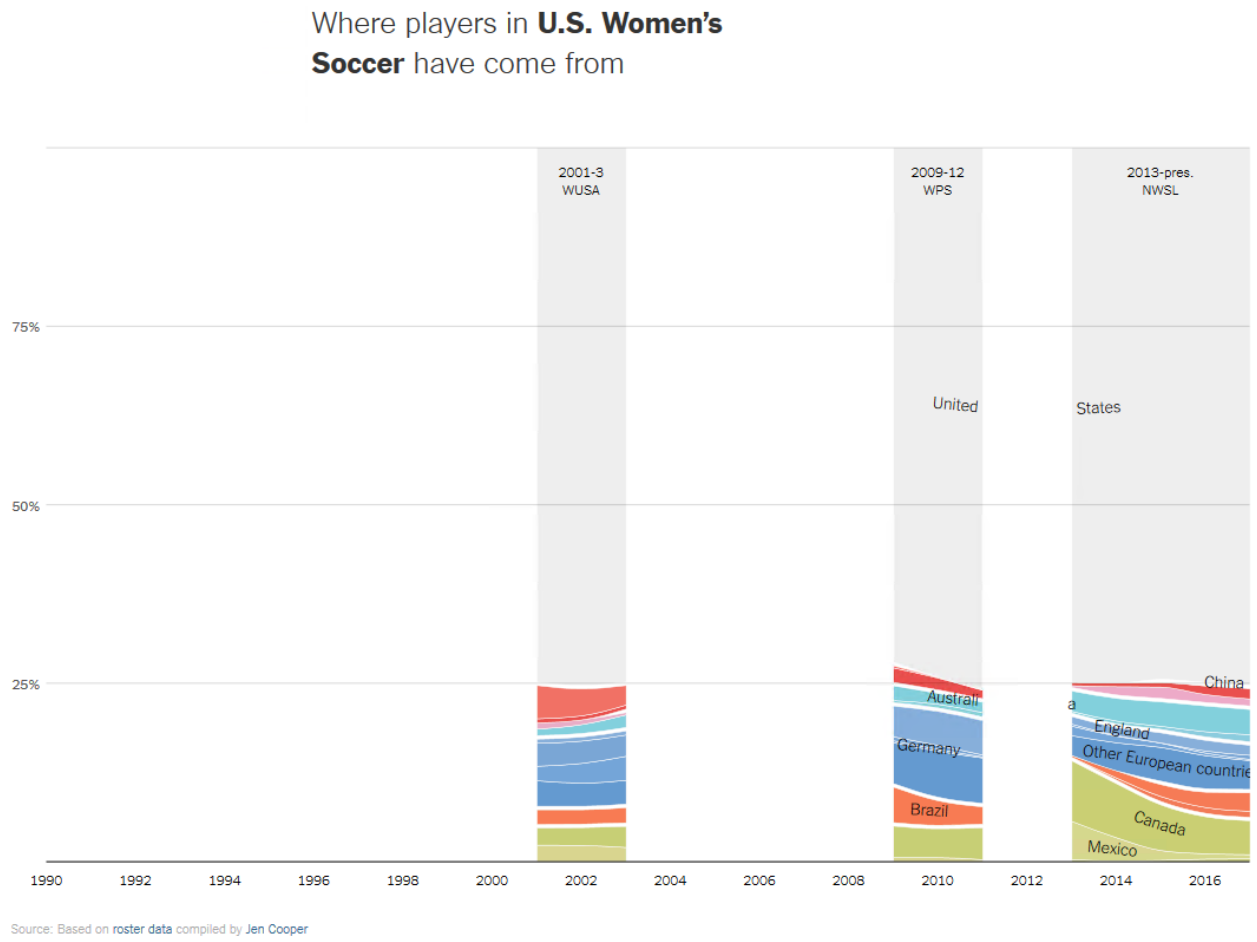
Visualizing missing data

Techniques to handle missing data

- Collect the data
- Show only what you have
- Show the gaps
- Treat it as a category

Visualizing missing data by showing the gaps

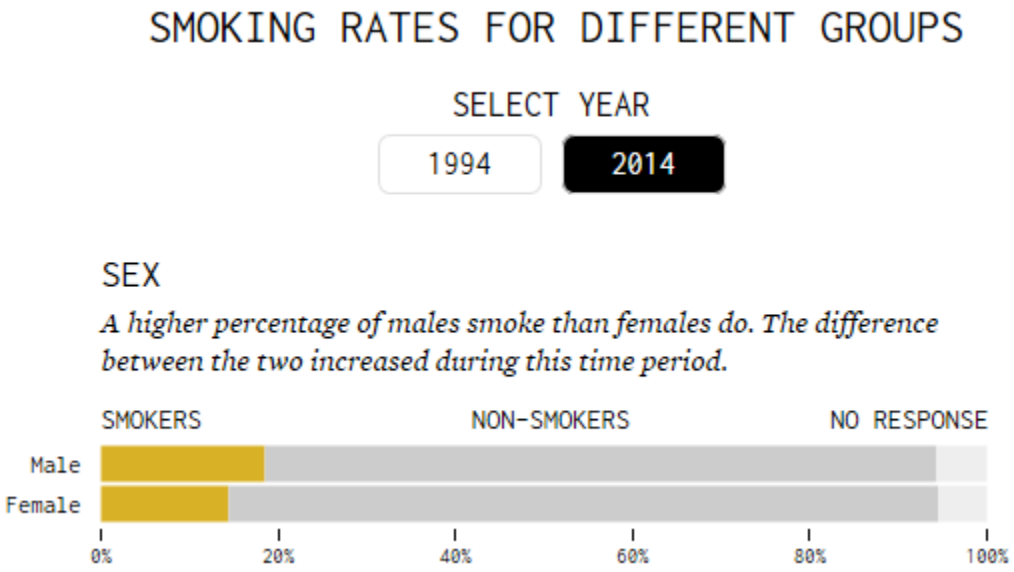
Show only the data you have



https://www.nytimes.com/interactive/2017/12/29/upshot/internationalization-of-pro-sports-leagues-premier-league.html

Visualizing missing data as a category

Use white or neutral color to show the ‘missing data category’

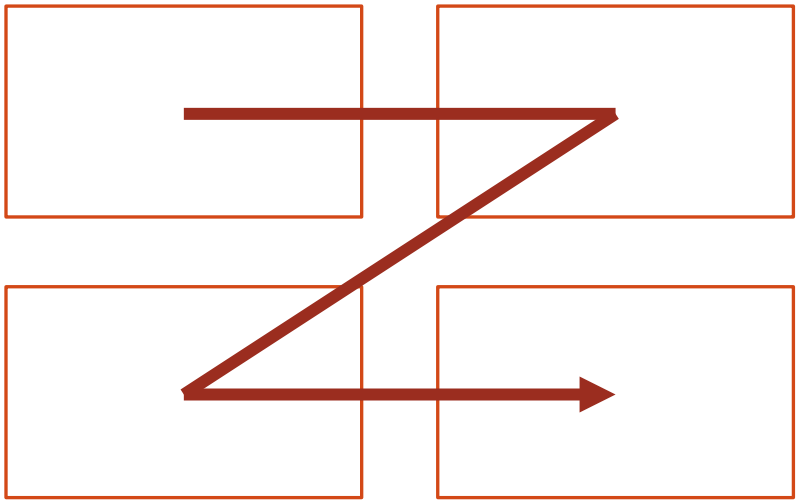


https://flowingdata.com/2016/06/20/who-still-smokes/

Visual order

Visual order

The attention of people follows the Z shape



You should place the important things on the top (left) of the display

Visual order

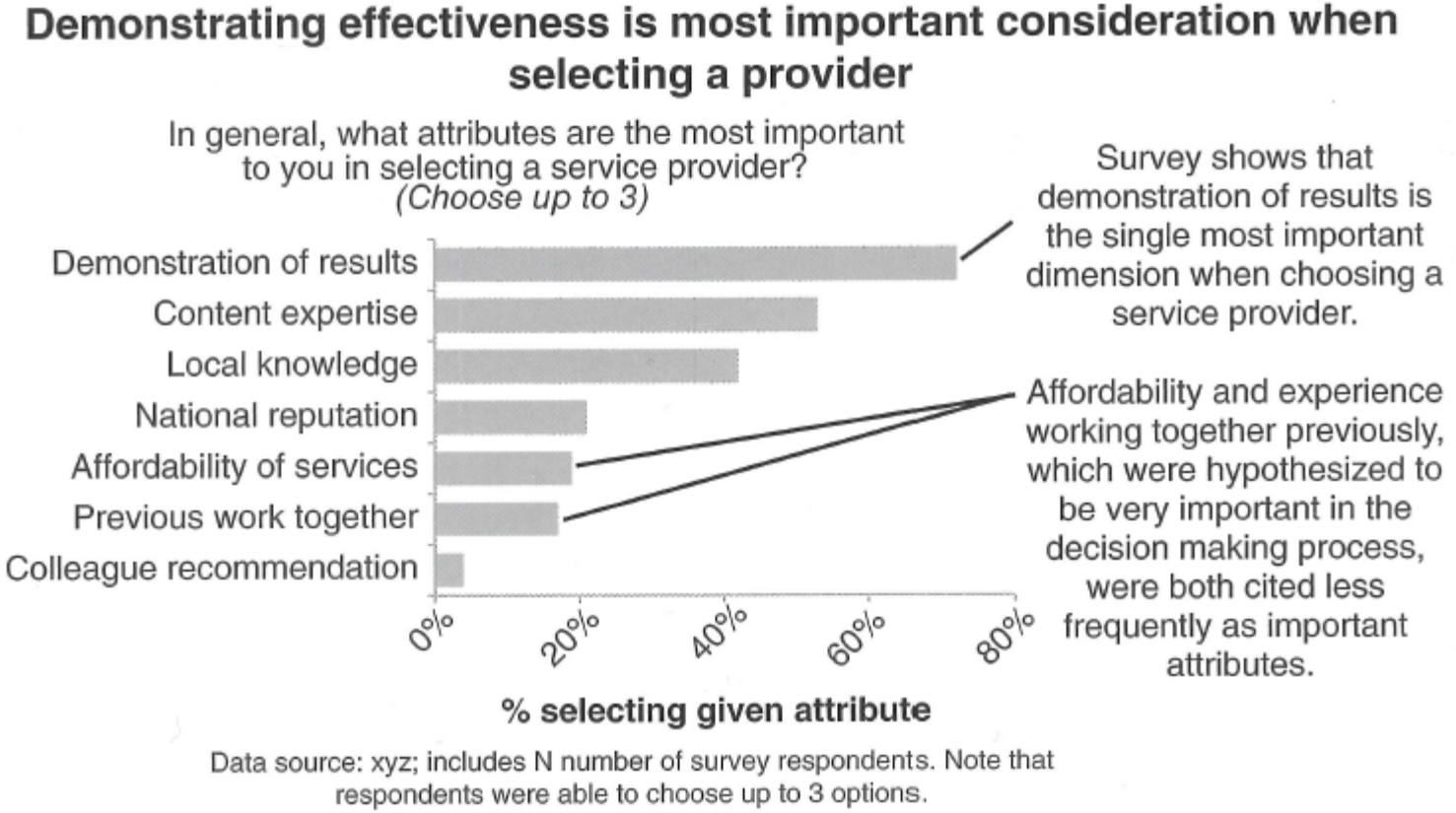
All elements should be aligned – create clean vertical and horizontal ‘lines’ to establish a sense of unity and cohesion

Do not be afraid of white (empty) space – do not add more data (or stretch the graphics) to get rid of it

Stay away of diagonal components (especially text)

- Text rotated by 45 degrees (in either direction) is 52% slower to read than normally oriented text
- Text rotated by 90 degrees (in either direction) is 205% slower to read than normally oriented text

Visual order – an example

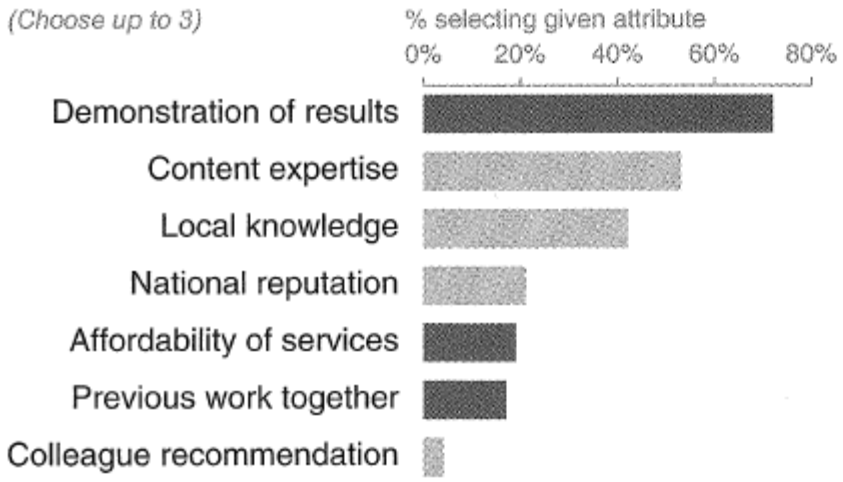


Visual order – an example

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?

(Choose up to 3)



Survey shows that **demonstration of results** is the single most important dimension when choosing a service provider.

Affordability and **experience working together previously**, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

Data source: xyz; includes N number of survey respondents.
Note that respondents were able to choose up to 3 options.

Visual order

Pay attention to details

Avoid

- Too much centered text
- Diagonal components, especially text
- Too many things on a single display

Interactivity

Interactivity

Advantages

- Expands the physical limits of what you can show
- Increases the quantity and broadens the variety of angles of analysis (to serve different purposes)
- Increases control and customization of the experience

Disadvantage

- Requires human time and attention

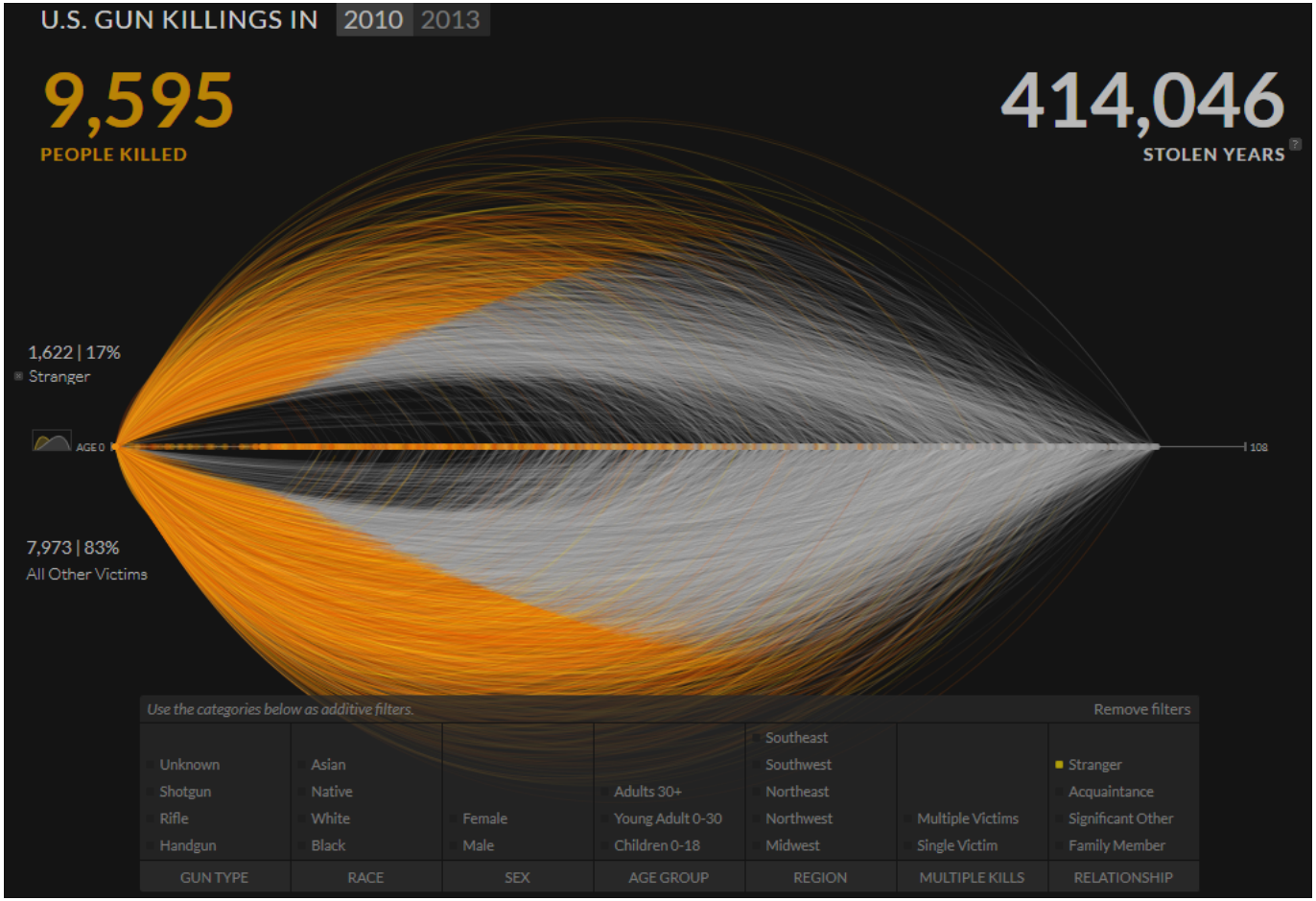
Can affect

- What data is displayed (data adjustments)
- How the data is displayed (presentation adjustments)

Data adjustments

- Framing: Isolate, include or exclude data
- Navigating: Expand or explore greater levels of detail in the displayed data
- Animating: Portray temporal data via animated sequences
- Sequencing: Navigate through discrete sequences of different angles of analysis
- Contributing: Customizing experiences through user-inputted data

Framing



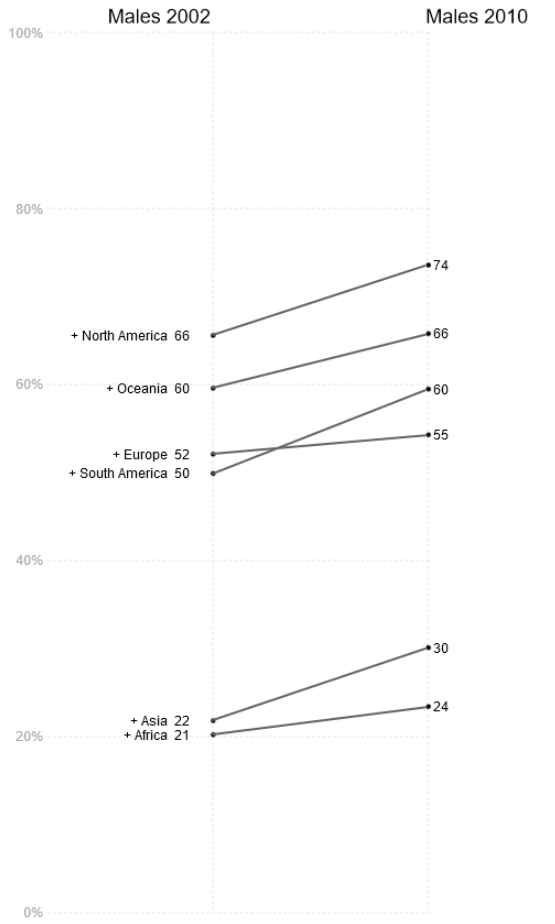
https://guns.periscopic.com/?year=2010

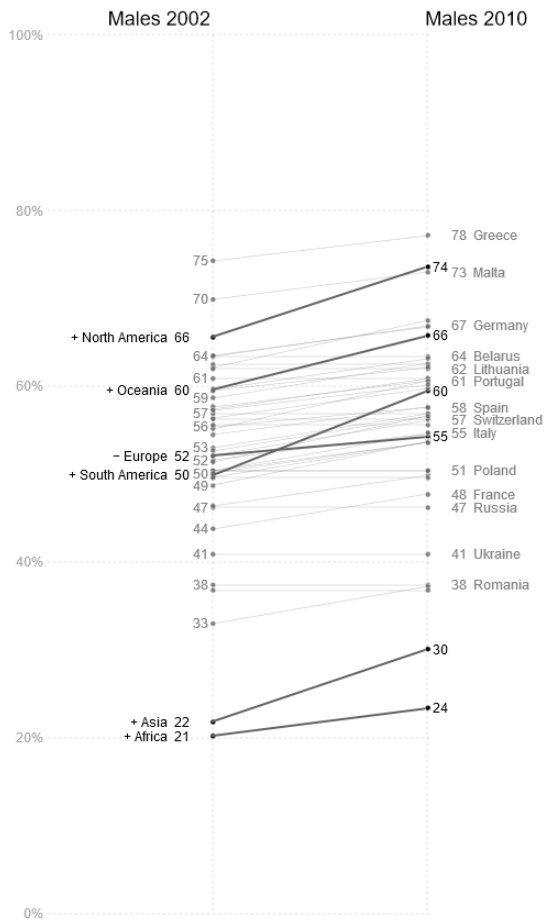
Navigating

Road orientation map

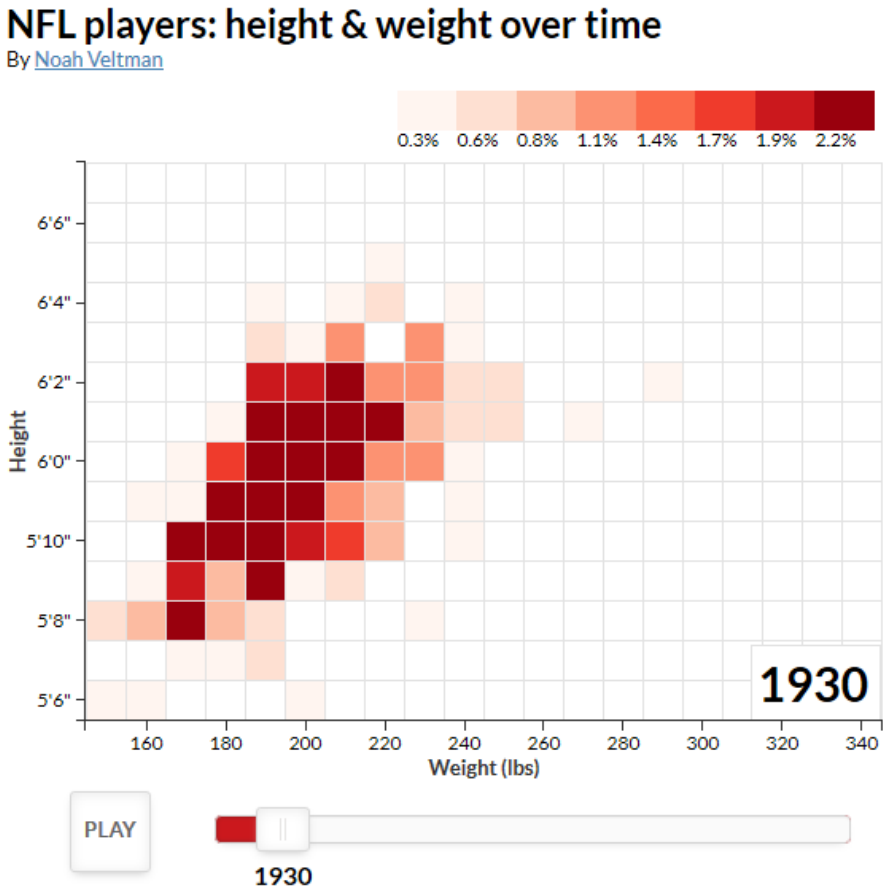
Navigating

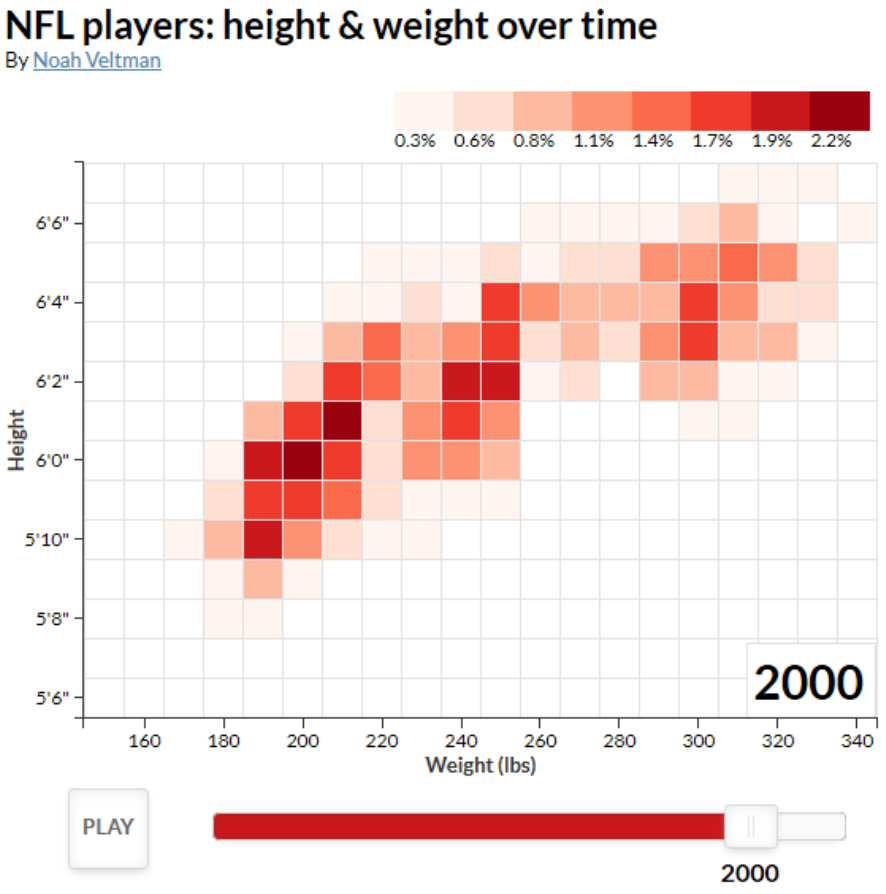
Obesity around the world





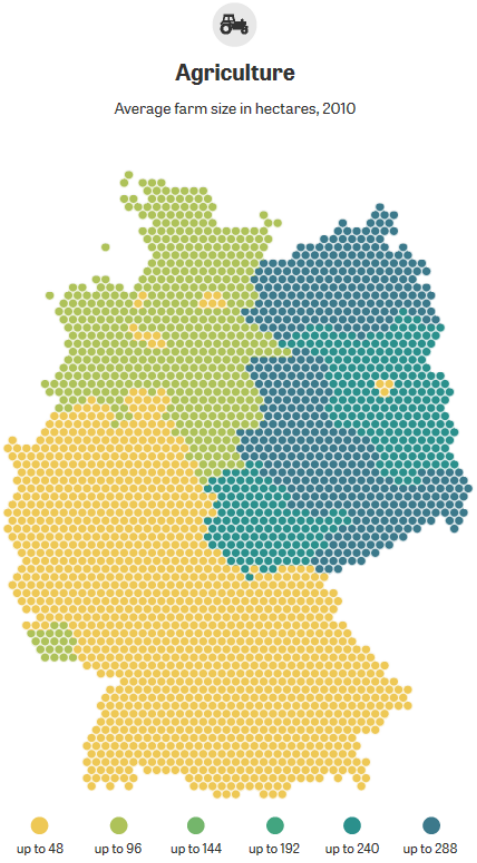
Animating

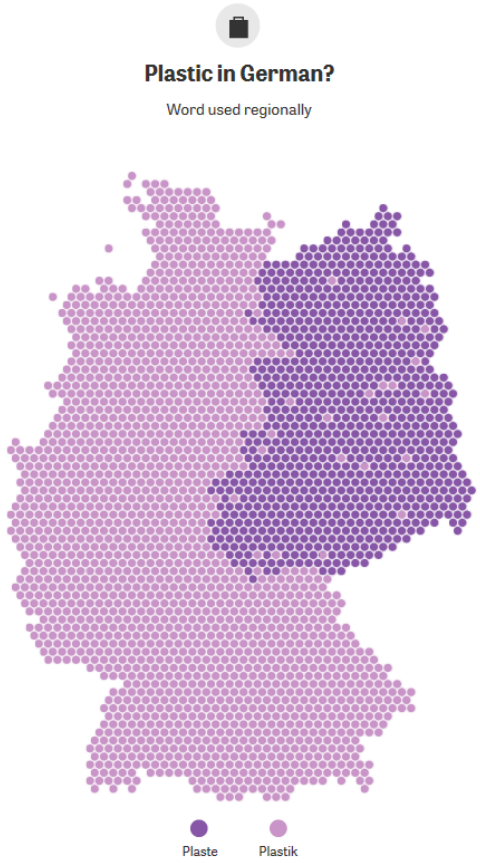


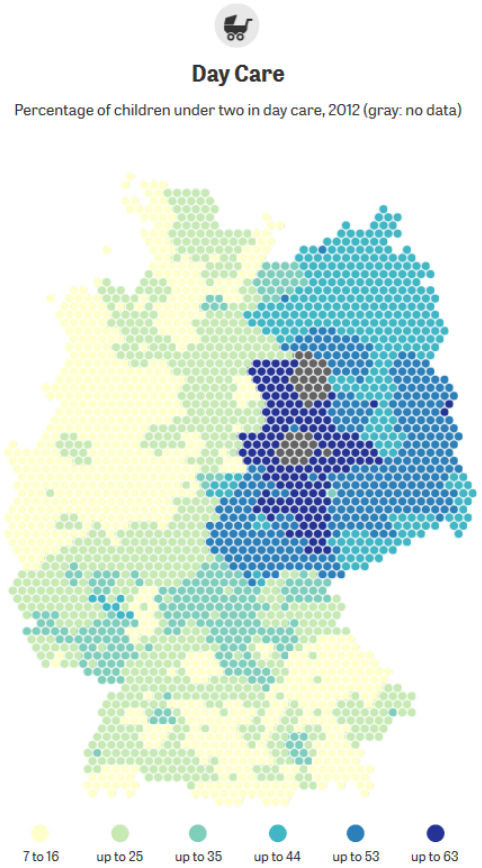


https://noahveltman.com/nflplayers/

Sequencing







https://www.zeit.de/feature/german-unification-a-nation-divided

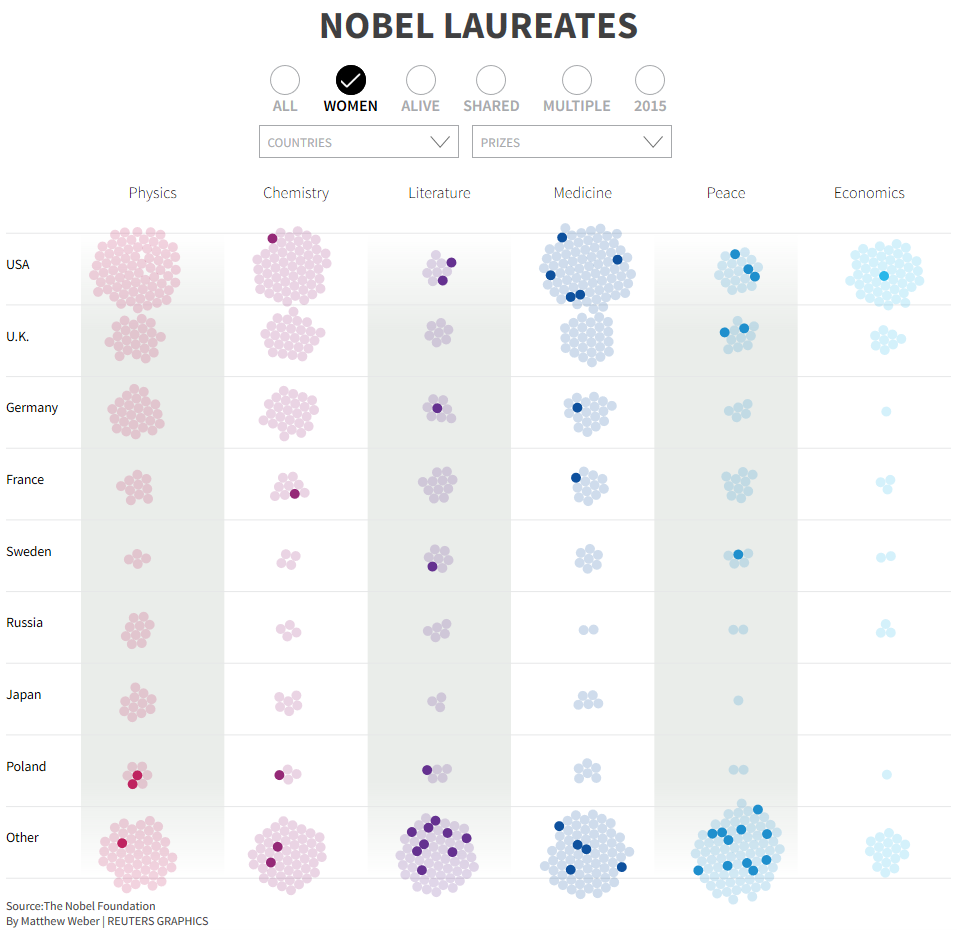
Contributing

http://www.informationisbeautiful.net/visualizations/who-old-are-you/

Presentation adjustments

- Focusing: Control what data is visually emphasized
- Annotating: Interact with marks to bring up more detail
- Orientating: Make better sense of your location within a display

Focusing



http://graphics.thomsonreuters.com/15/nobels/index.html

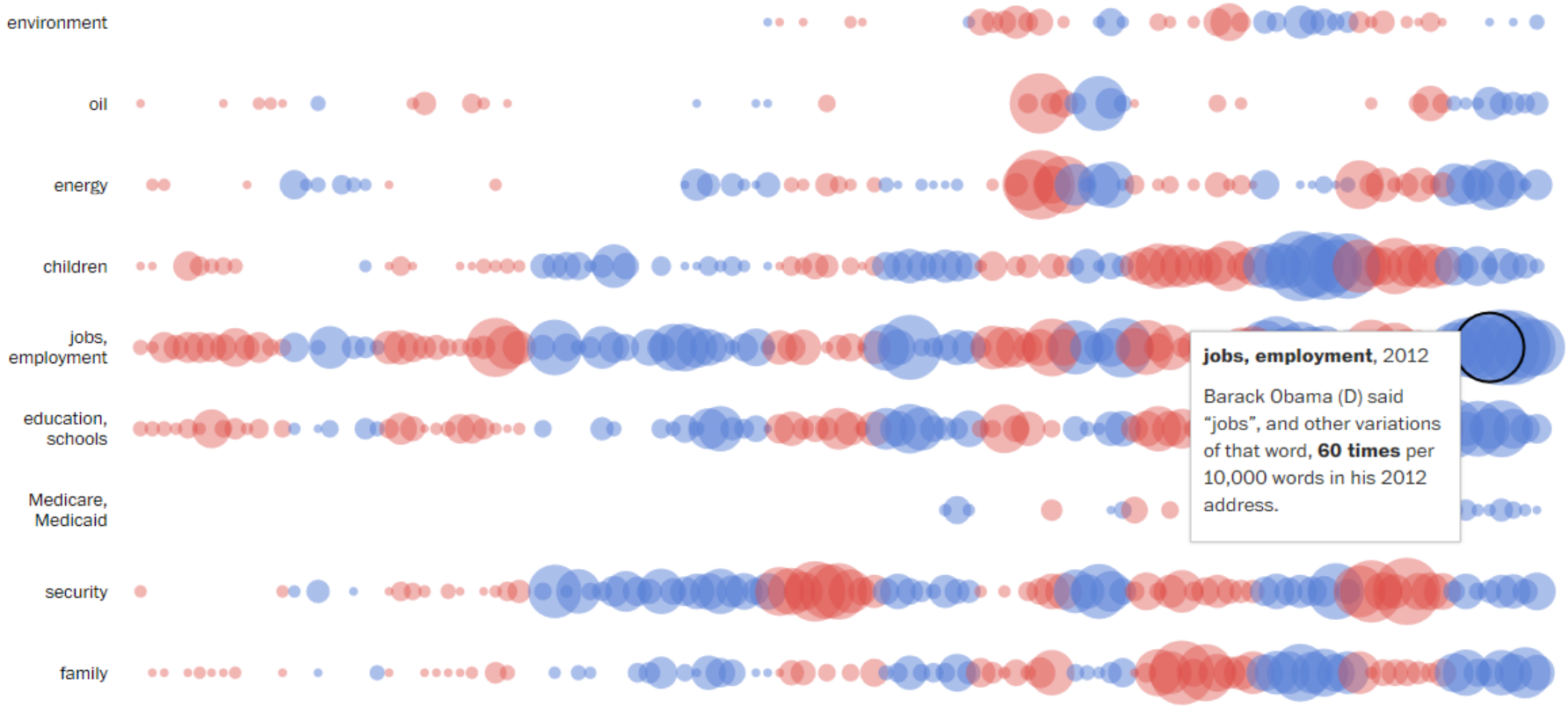
Focusing

Using brushing and linking

https://bl.ocks.org/mbostock/4063663

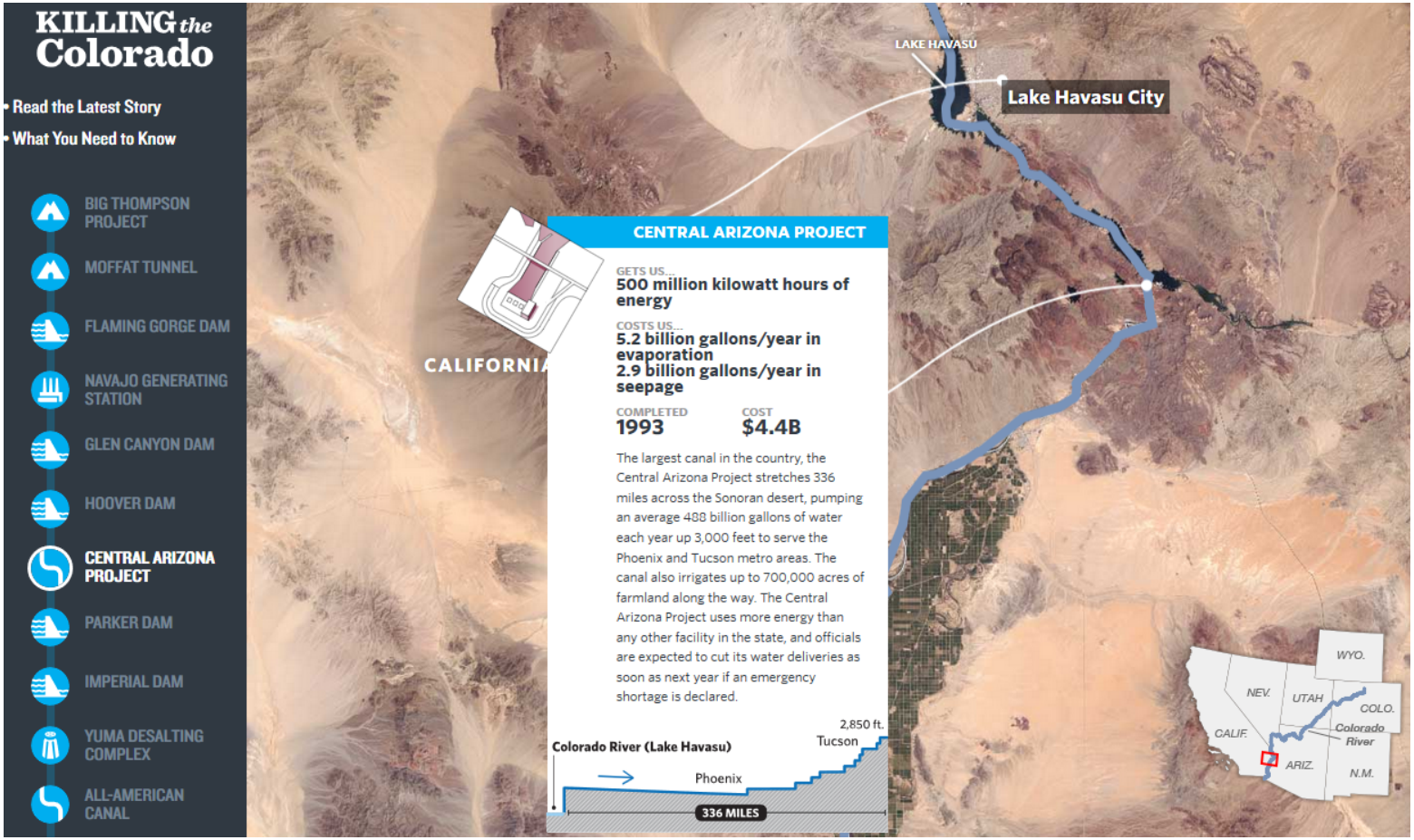
Annotating

History through the president's words



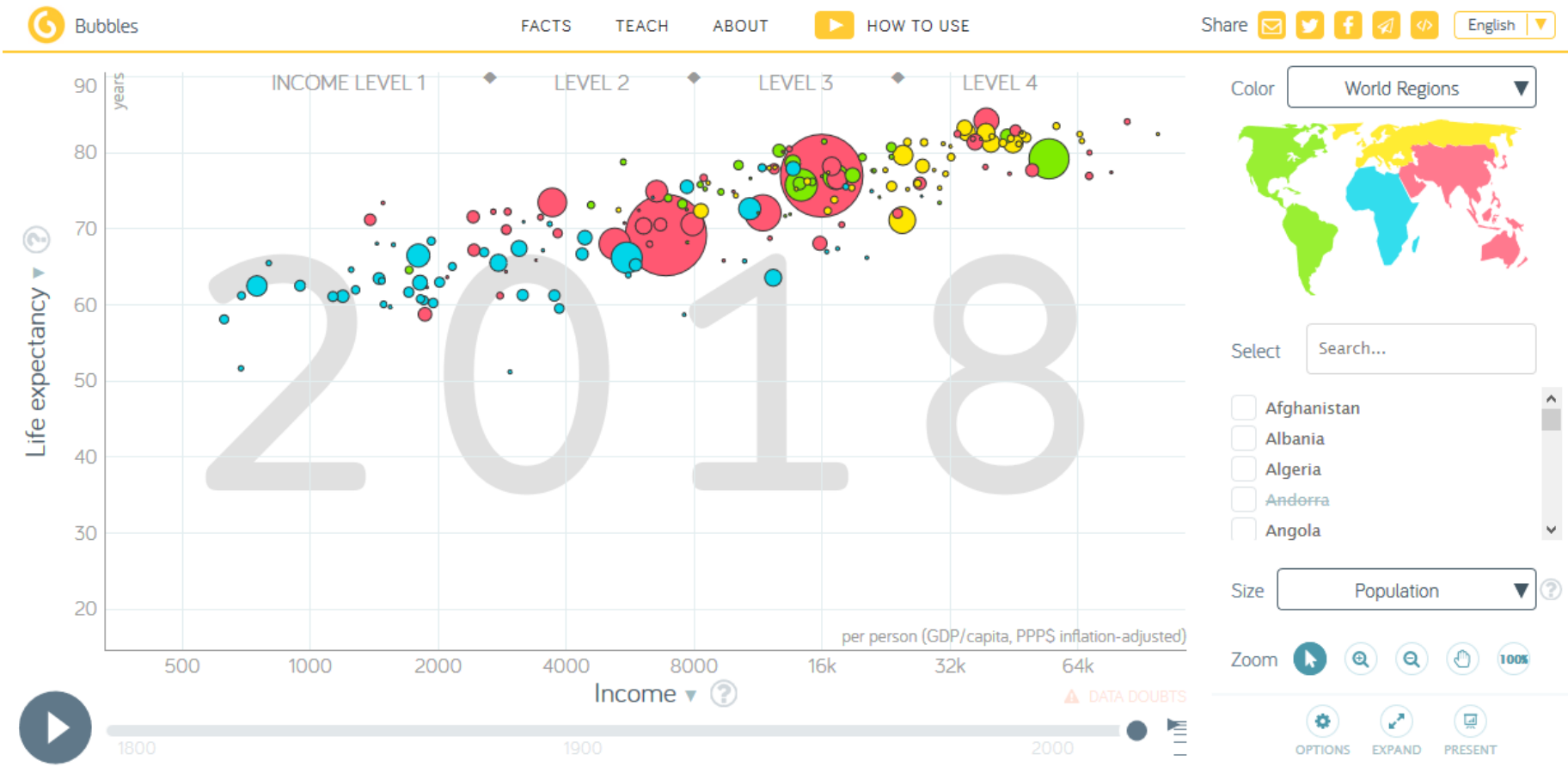
https://www.washingtonpost.com/graphics/politics/2016-sotu/language/

Orientating



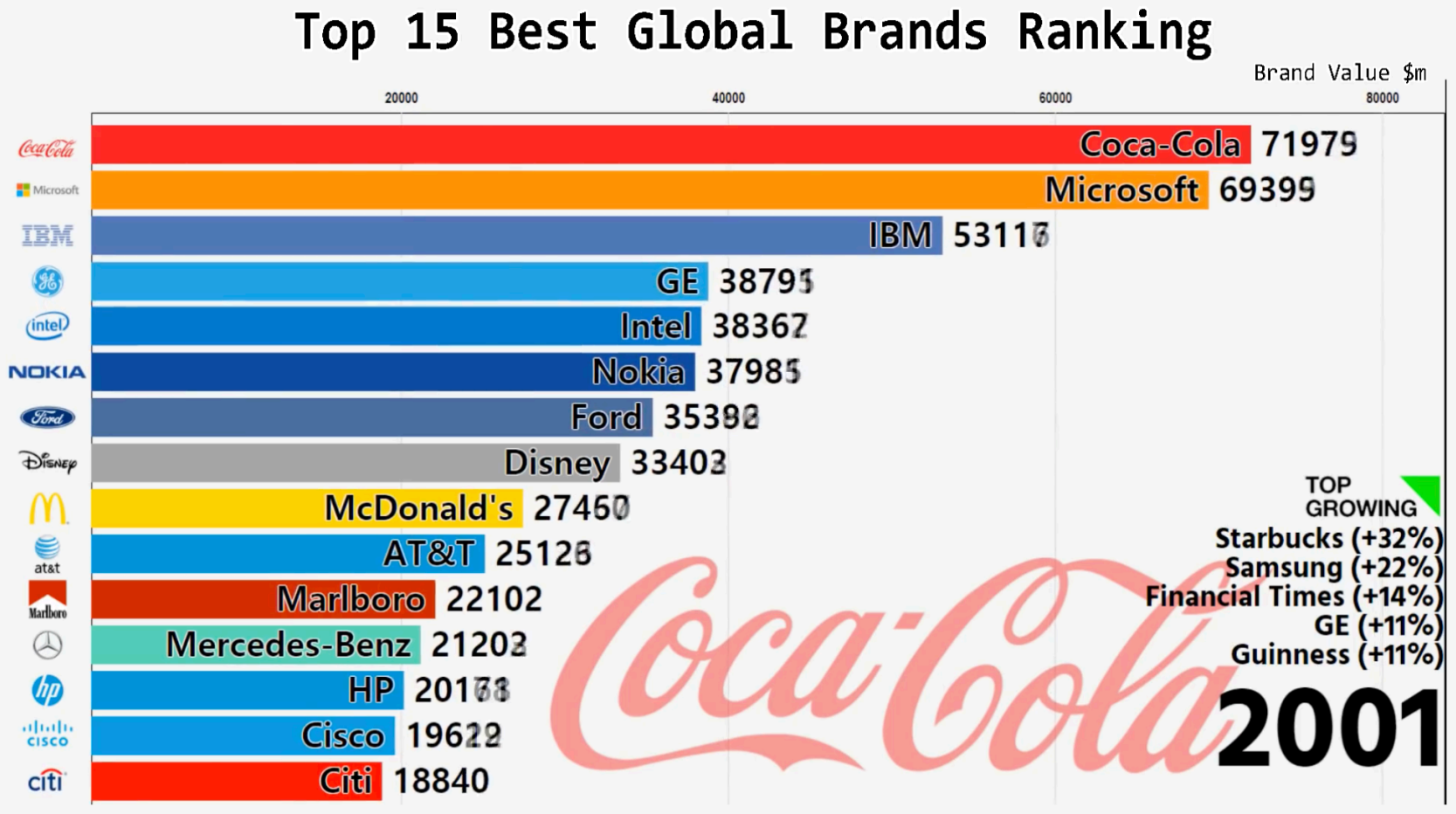
https://projects.propublica.org/killing-the-colorado/explore-the-river#central-arizona-project

Interactivity example



https://www.gapminder.org/tools/ and https://www.youtube.com/watch?v=hVimVzgtD6w

Animation example



https://www.visualcapitalist.com/animation-top-15-global-brands-2000-2018/ and https://www.youtube.com/watch?v=BQovQUga0VE&feature=emb_logo

Storytelling

Storytelling

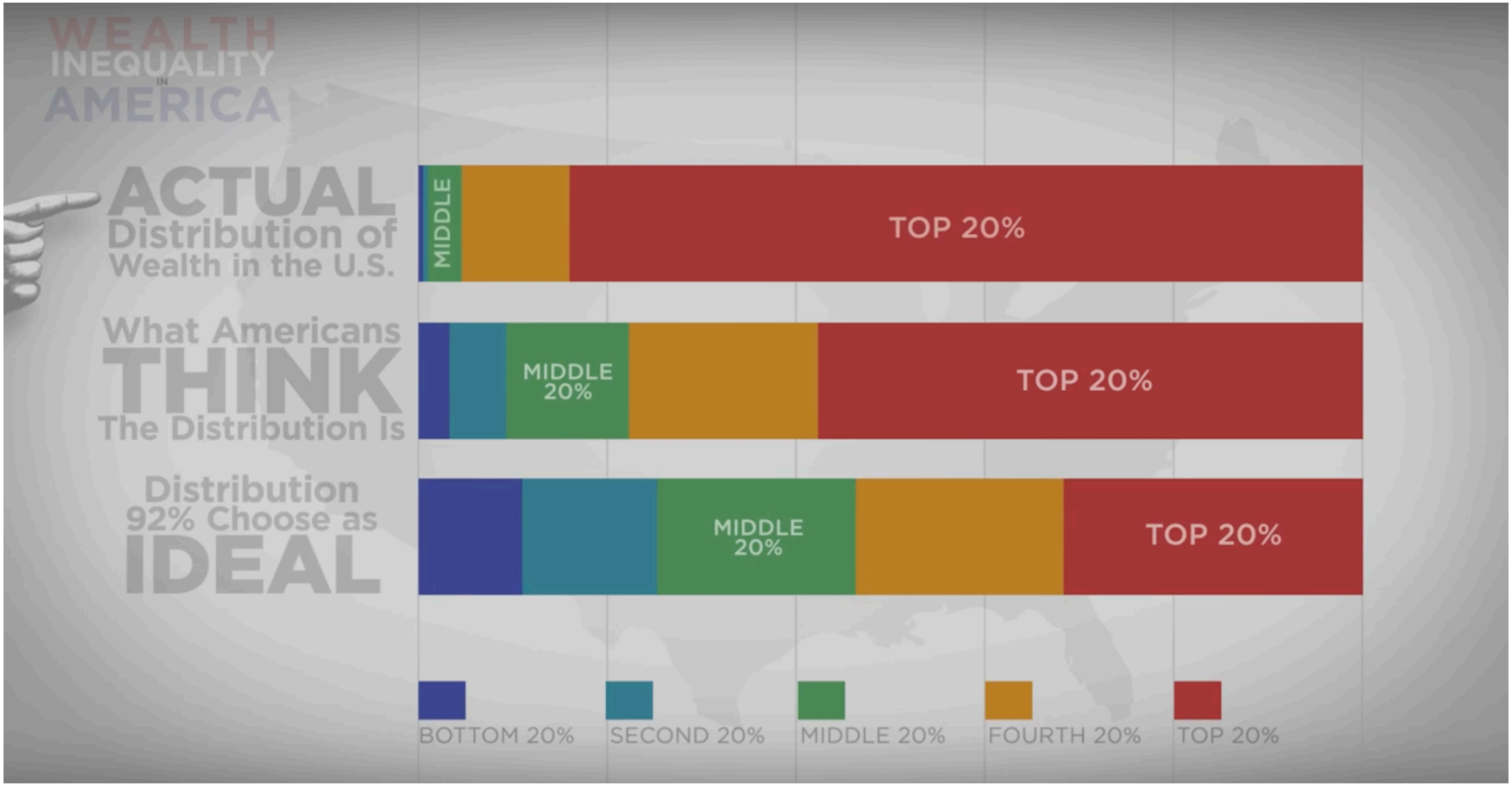
Storytelling ≠ making something up

Visualization can be used to tell a story

Distinctions among terms

- Annotation: Highlighting certain data and putting it in context
- Narration: Arranging your charts in a meaningful sequence intended to display cause and effect relationships
- Storytelling: Narrating with an emotional component

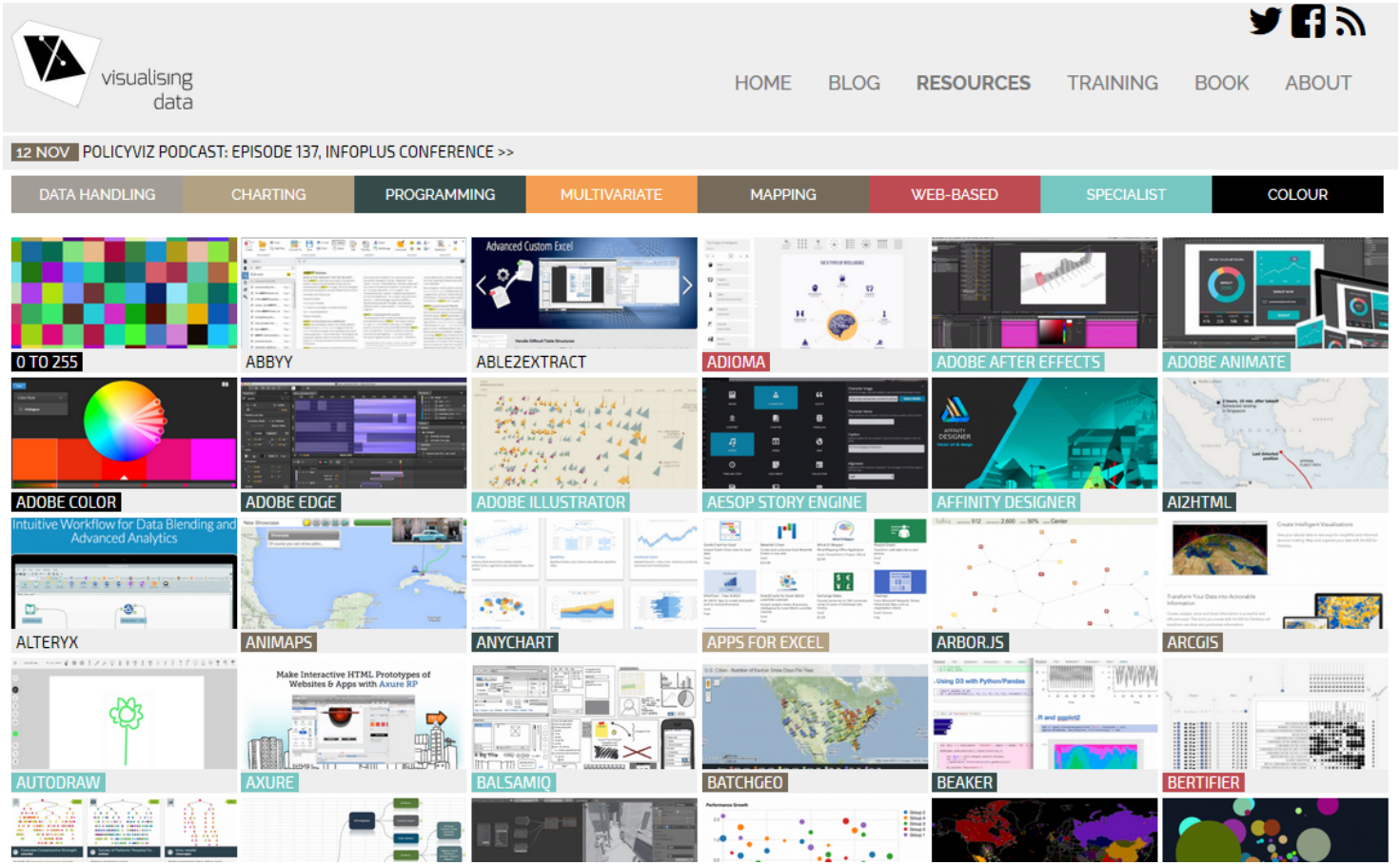
Storytelling example



https://www.youtube.com/watch?v=QPKKQnijnsM

Tools

Tools



http://www.visualisingdata.com/resources/

D3

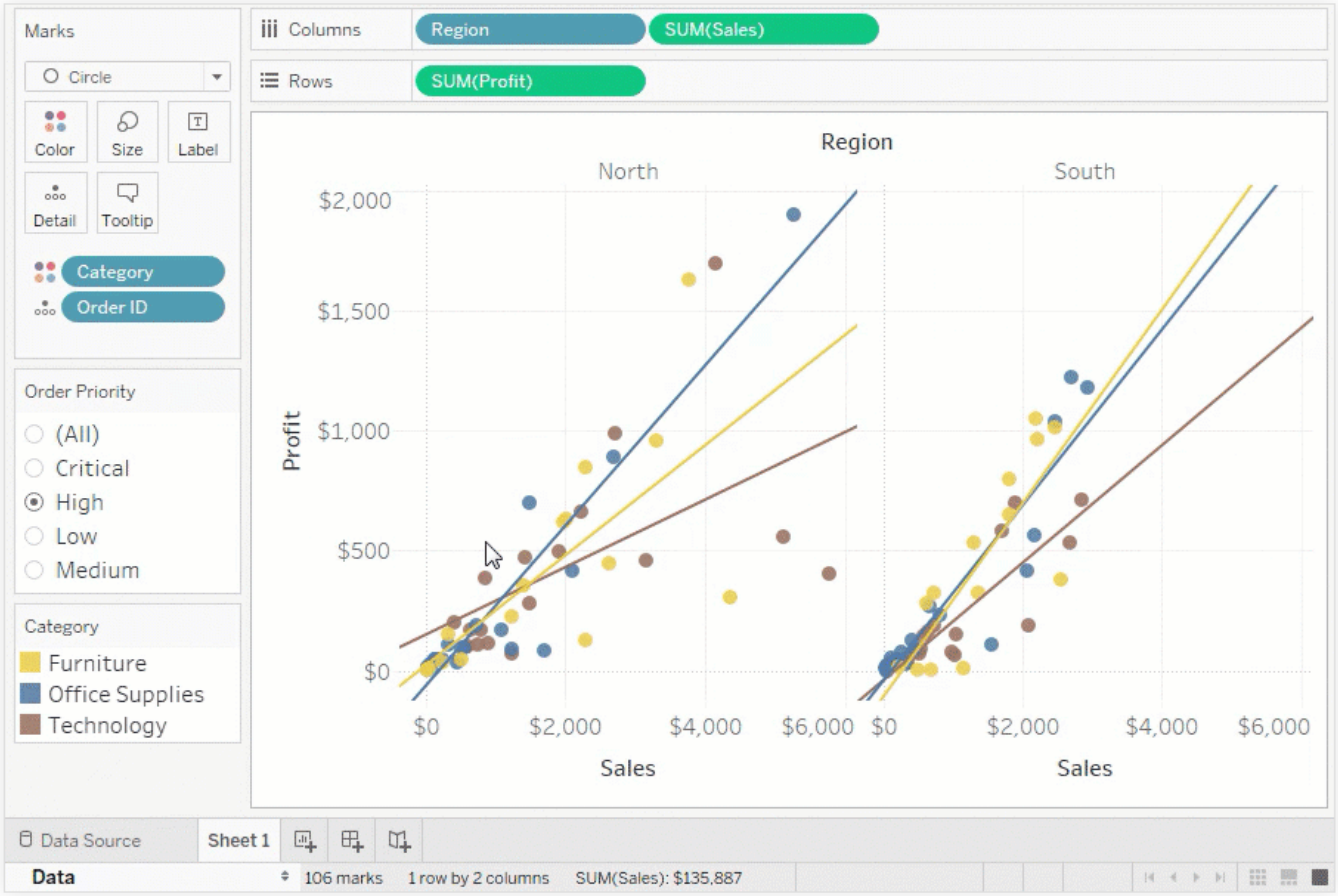
Data-Driven Documents

A JavaScript library

Emphasis on interactivity

https://d3js.org/

Tableau Public



Does not require programming skills

Visualizations created with the free version are public

https://public.tableau.com/s/