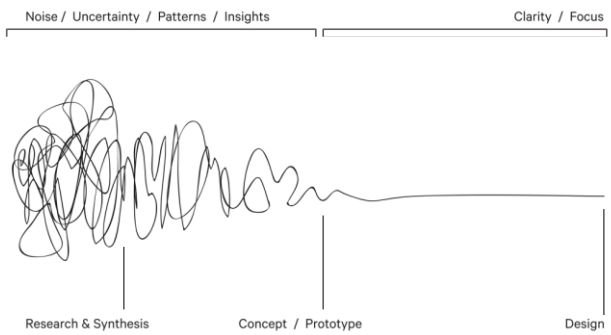
Data Visualization

VISUALIZATION DESIGN

Visualization design



https://thedesignsquiggle.com/

Overview

The 7 steps of visualization design

Basic charts

Multivariate/multidimensional data visualization

Visualizing uncertainty and missing data

Interactivity

Storytelling

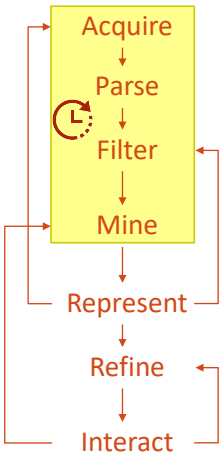
Tools

3

The 7 steps of visualization design

4

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

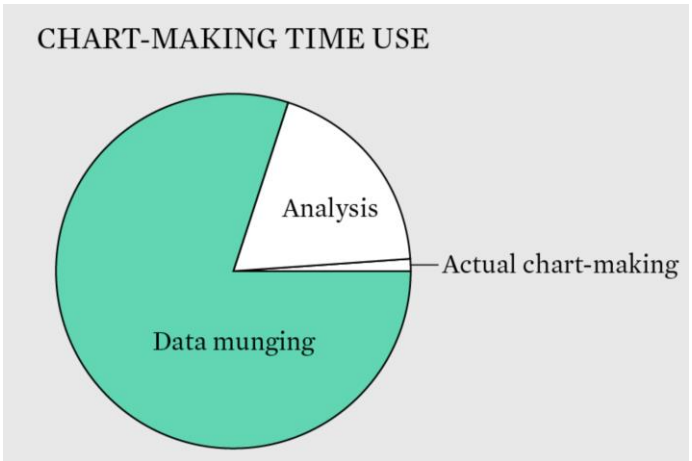
Choose a basic visual model and draw the data

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

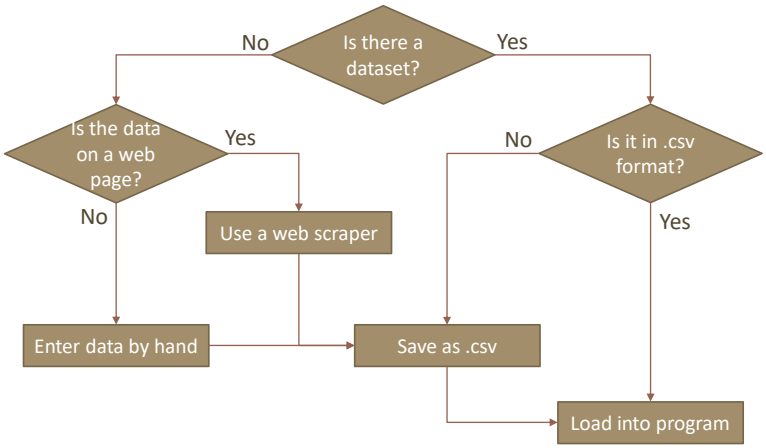
B. Fry. *Visualizing data*. O'Reilly Media, 2008 5

The 7 steps of visualization design



https://flowingdata.com/2022/09/15/process-206-munging/ 6

Acquire the data



https://itunes.apple.com/us/course/data-literacy-and-data-visualization/id693097601 7

Parse the data

Check for errors

Change type
- For example, ordinal to categorical

Choose the level for hierarchical data
- Temporal data: day of the week, day of the month, …
- Spatial data: countries, regions, municipalities, …

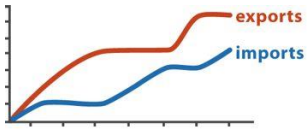
(Dis)aggregate data

8

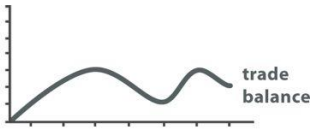
Parse the data

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



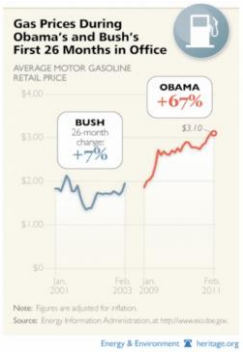
trade balance = exports – imports

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

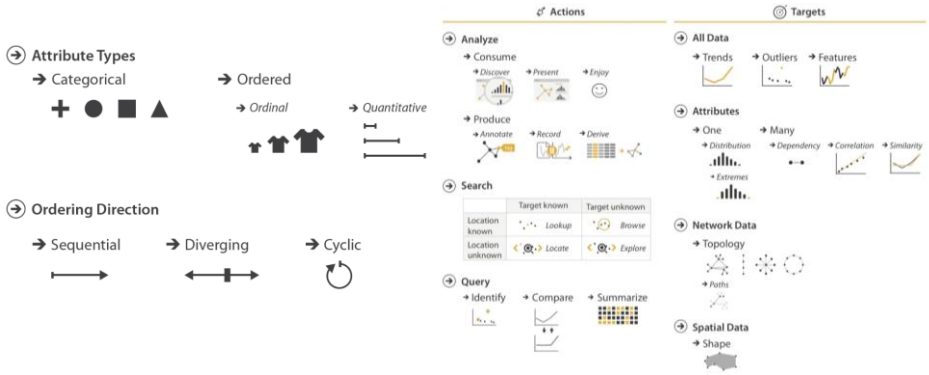
- Look for important features and patterns
- Look for any striking deviations (outliers)
- Interpret your findings

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

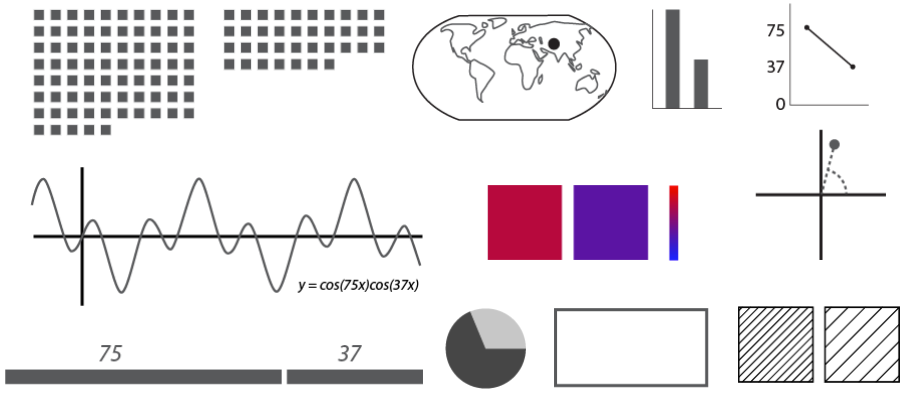
Choice depends on the data and the task



T. Munzner. *Visualization Analysis & Design*. CRC Press, Boca Raton, 2014

Represent the data

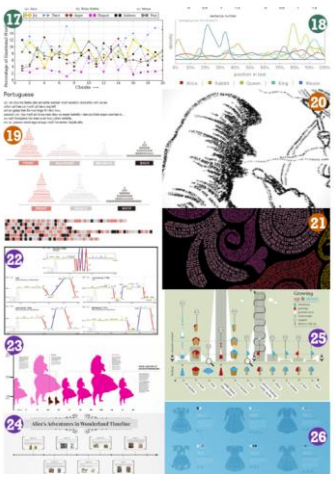
45 ways to communicate two quantities (75 and 37)



https://visual.ly/blog/45-ways-to-communicate-two-quantities/ 13

Represent the data

58 ways to visualize
Alice in Wonderland



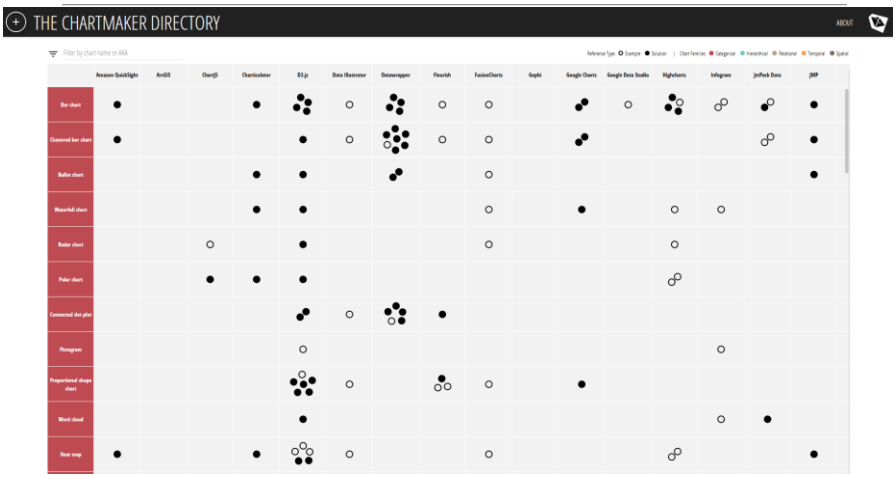
https://richardbrath.wordpress.com/2021/10/31/58-ways-to-visualize-alice-in-wonderland/ 14

Represent the data

https://datavizcatalogue.com/search.html

15

Represent the data



https://chartmaker.visualisingdata.com/

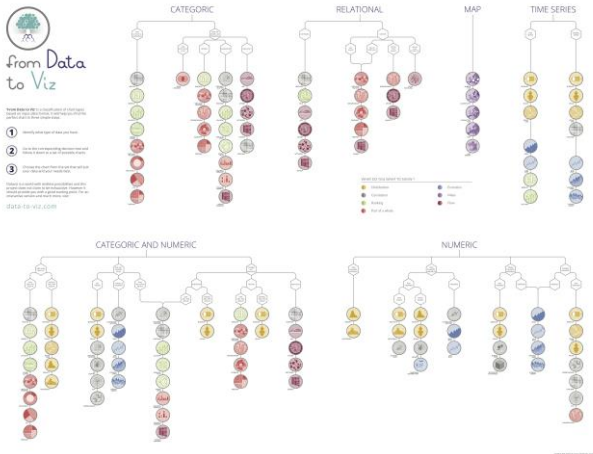
16

Represent the data

https://github.com/ft-interactive/chart-doctor/tree/master/visual-vocabulary
https://ft-interactive.github.io/visual-vocabulary/

17

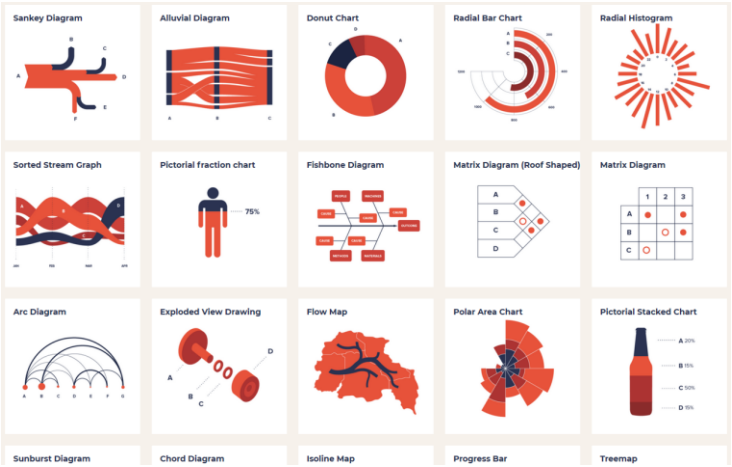
Represent the data



https://www.data-to-viz.com/poster.html

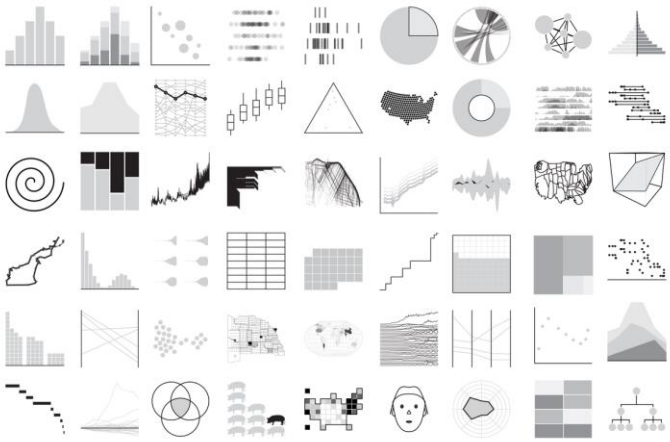
18

Represent the data



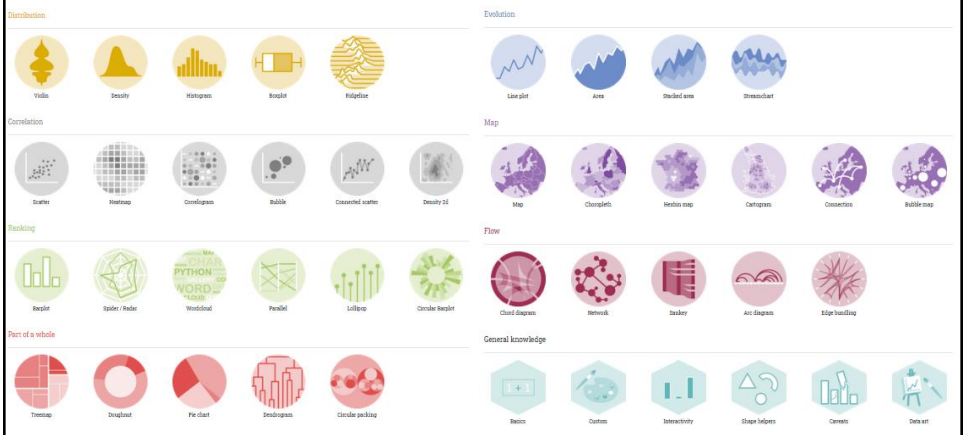
https://datavizproject.com/ 19

Represent the data



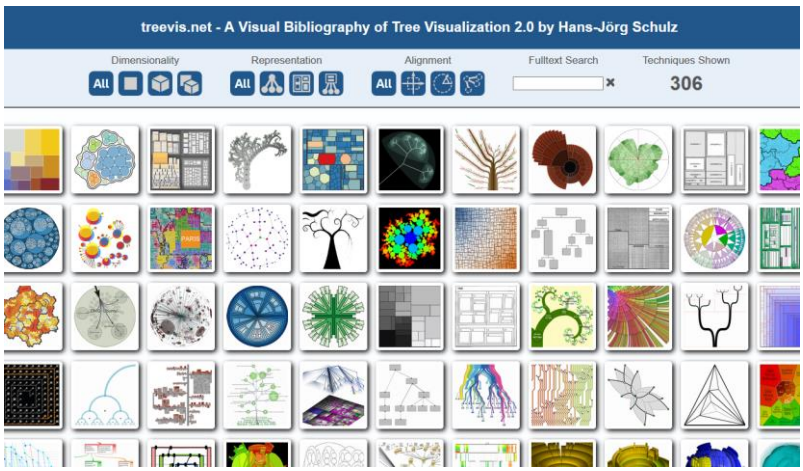
https://flowingdata.com/chart-types 20

Represent the data



https://www.d3-graph-gallery.com/ 21

Represent the data

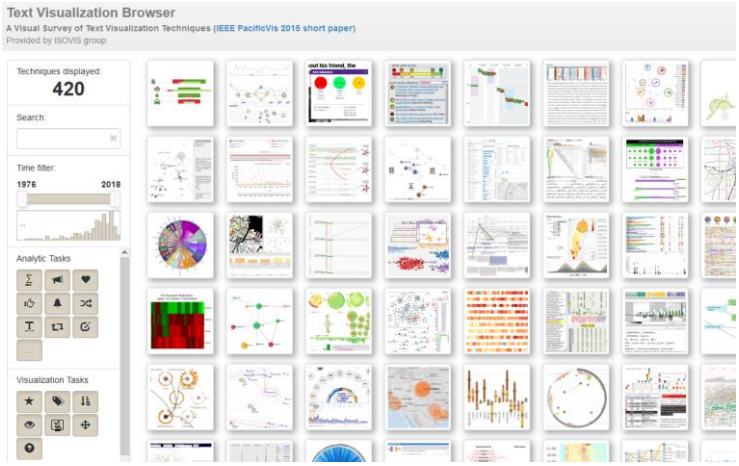


http://treevis.net/ 22

Represent the data

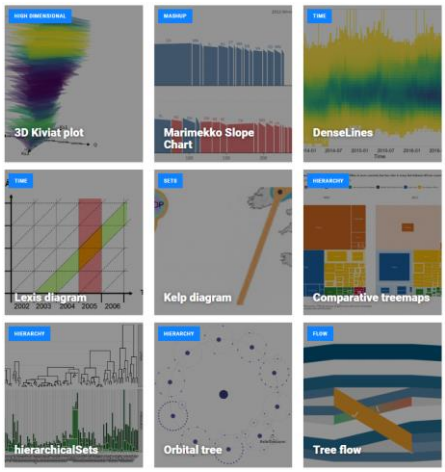
https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html

Represent the data



http://textvis.lnu.se/

Represent the data

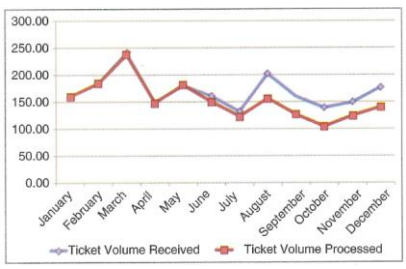


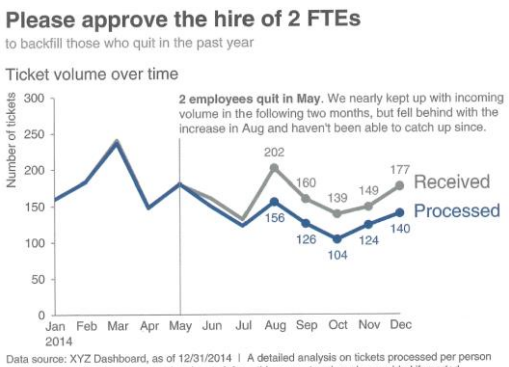
https://www.xeno.graphics/

25

Refine the visualization

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





C. Nussbaumer Knaflic. *Storytelling with data*. Wiley, 2015

26

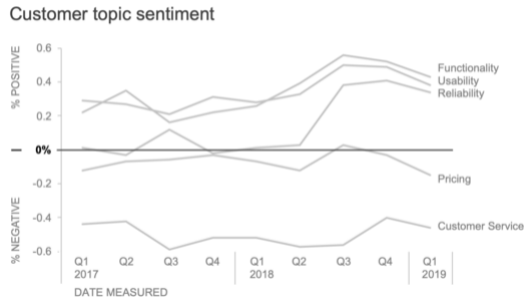
Refine the visualization: Use takeaway titles

Customer Service ranks lowest

Action needed to address **recent decline**

Q1 2019: all topics declined

Success: efforts to increase **reliability** worked!



Use action-oriented titles

Aim for one line

Refine the visualization: Tools

Proprietary vector graphics editors

- Adobe Illustrator https://www.adobe.com/it/products/illustrator.html
- CorelDRAW https://www.coreldraw.com/en/

Free vector graphics editors

- Inkscape https://inkscape.org/
- Gravit Designer (now Corel Vector?) https://www.techspot.com/downloads/7062-gravit-designer.html

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*

29

Basic charts

30

Basic charts

Bar charts

Line chart

Pie charts

Geographical data
- Dot maps
- Choropleth maps
- (Hexagon) tile maps

Networks and trees
- Node-link diagrams
- Matrices

Bar charts

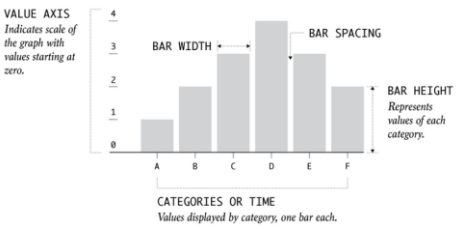
Use them to show values per categories (or discrete time)

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



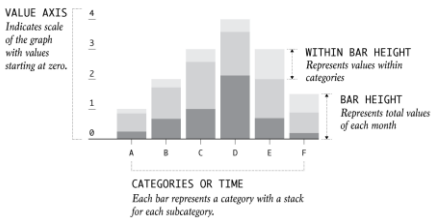
https://flowingdata.com/charttype/bar-chart-type/

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



https://flowingdata.com/charttype/stacked-bar-chart/ 33

Line charts

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

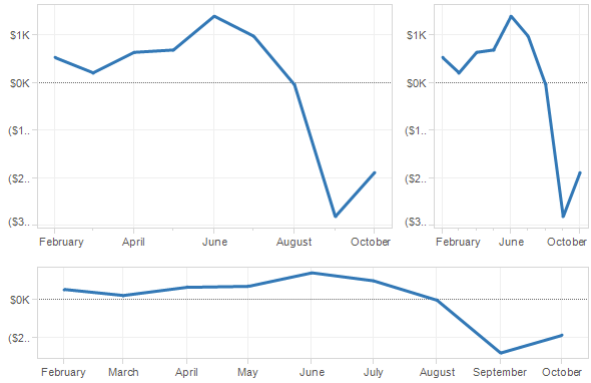
Do not use them for categories

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

https://blog.datawrapper.de/line-charts/ 34

Line charts: Aspect ratio bias

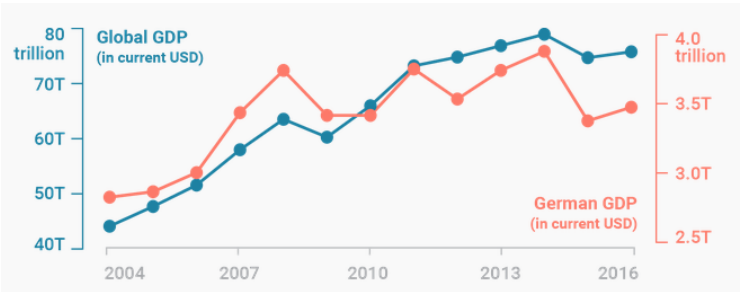


Banking to 45 Degrees

https://eagereyes.org/basics/banking-45-degrees

35

Line charts: Avoid dual axis charts

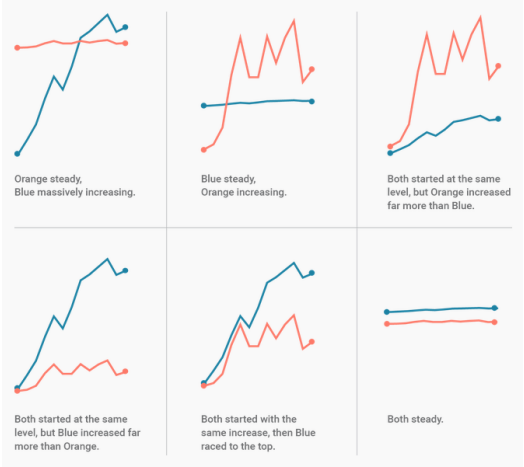


Easily manipulated to mislead
Hard to read

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

36

Line charts: Avoid dual axis charts

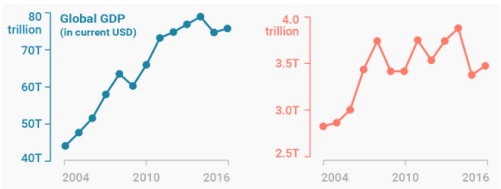


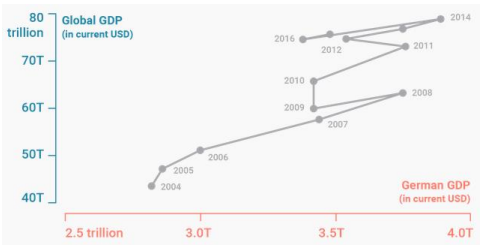
https://blog.datawrapper.de/dualaxis/ 37

Line charts: Avoid dual axis charts

Alternatives

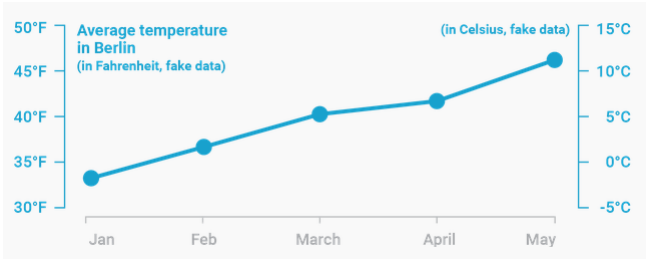
- Side-by-side charts
- Connected scatter plots





https://blog.datawrapper.de/dualaxis/ 38

Line charts: Dual axes sometimes okay



https://blog.datawrapper.de/dualaxis/ 39

Pie charts

Use them to show parts that sum up to 100%

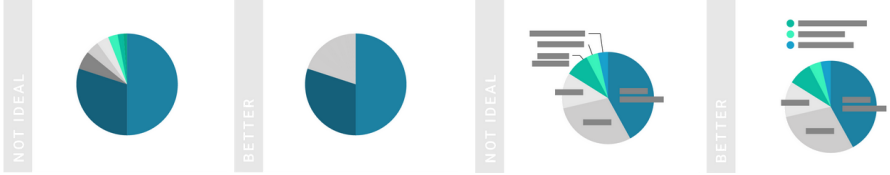
Show the values for each slice

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



https://blog.datawrapper.de/pie-charts/ 40

Geographical data

Use maps only when the spatial relationship is important

Space is the most effective visual channel and you do not want to waste it for spatial information if not relevant

41

Dot maps

Also called *dot distribution maps*

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

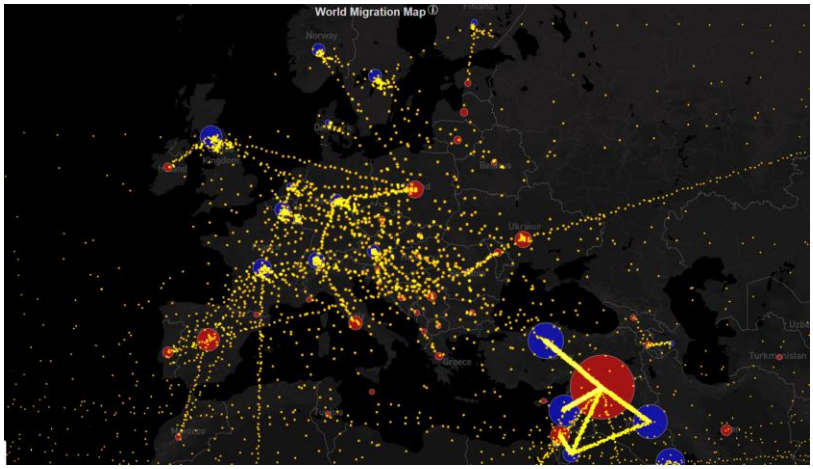
Can reveal patterns when the points cluster on the map

Could just be showing population density (!)

Use size and color to convey additional information

42

Dot maps



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

43

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

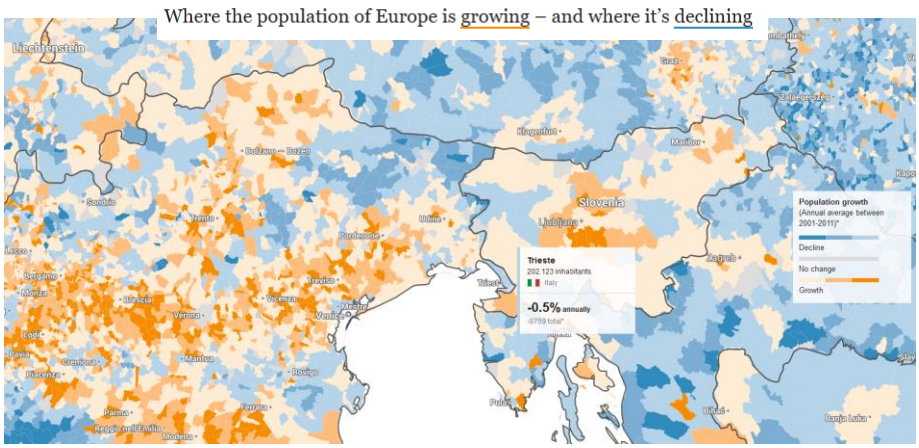
Show relative instead of absolute data

Be careful in choosing bin size

Be careful in choosing colors

44

Choropleth maps



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

45

(Hexagon) tile maps

Use them to show spatial relationship of categorical or numerical data where the area size is not important

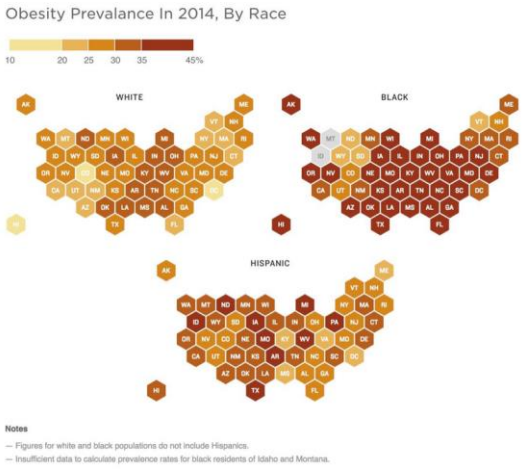
Tile represents a state/province

Often hexagonal or square

Harder to locate the given state/province

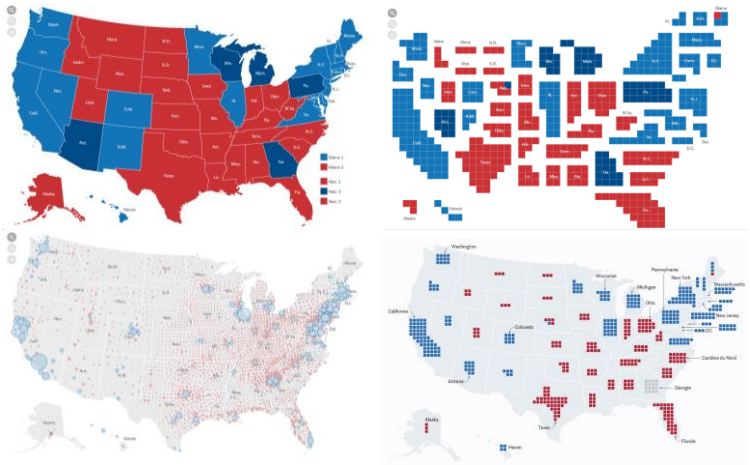
46

(Hexagon) tile maps



https://policyviz.com/2016/05/05/hexagon-tile-map-excel/ 47

Comparisons of maps



https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html
https://www.lemonde.fr/international/article/2020/11/04/elections-americaines-2020-suivez-la-carte-des-resultats-en-direct_6058394_3210.html 48

Networks and trees

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

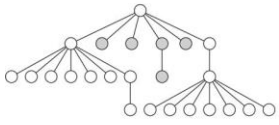
Nodes and links can also have attributes associated to them

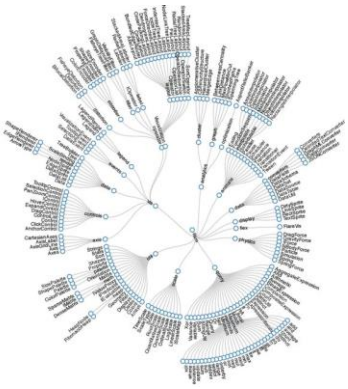
Node-link diagrams

Trees

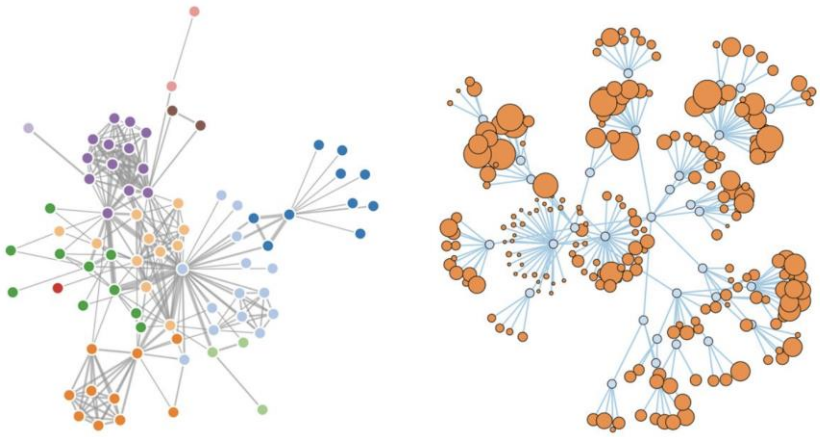
Layout depends on size

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



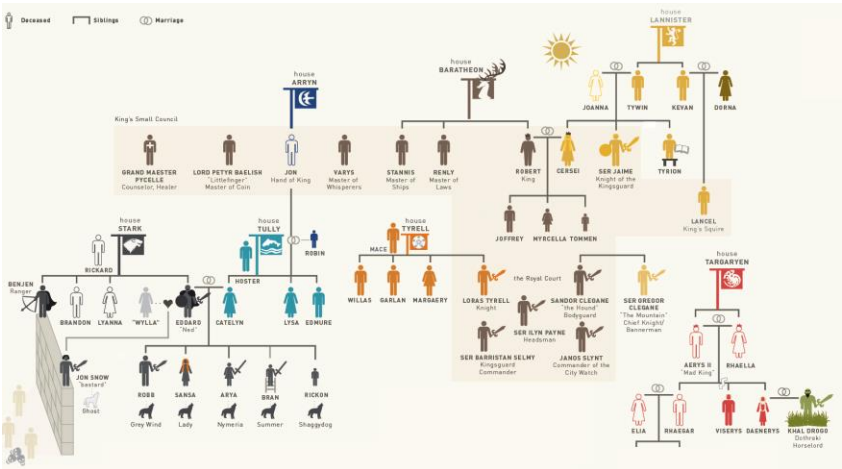


Node-link diagrams



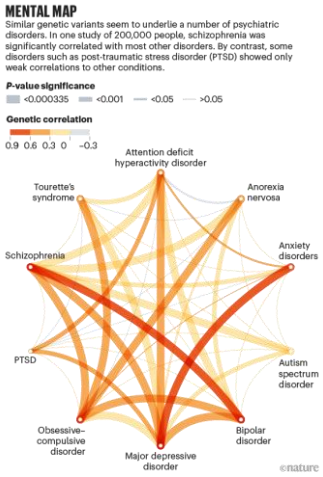
T. Munzner. *Visualization Analysis & Design*. CRC Press, Boca Raton, 2014 51

Node-link diagrams



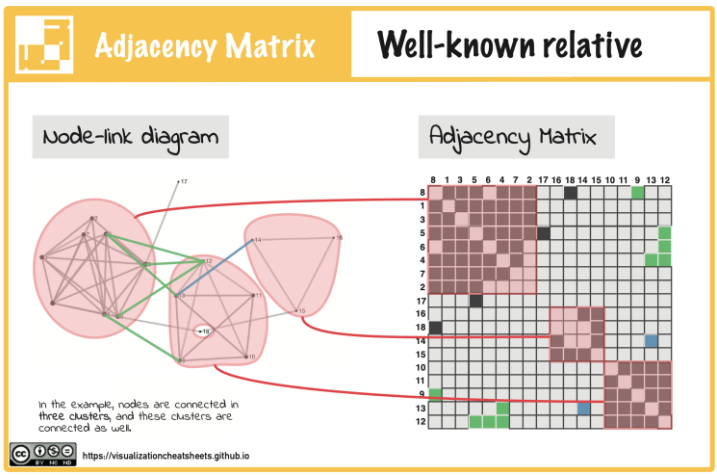
http://www.mahina.se/2011/05/infographic-game-of-thrones-family-tree/ 52

Node-link diagrams



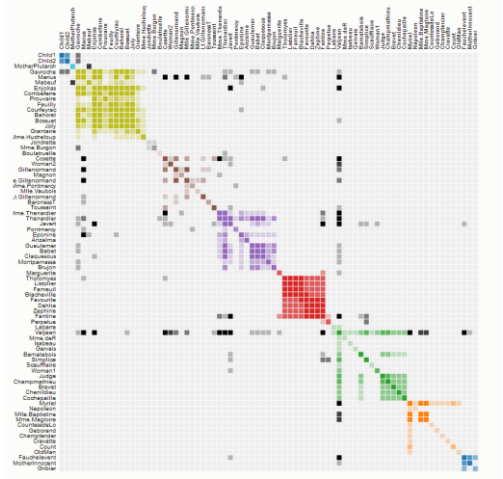
https://www.nature.com/articles/d41586-020-00922-8

Adjacency matrix



https://visualizationcheatsheets.github.io/

Adjacency matrix



Co-occurrence of characters in Les Misérables

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

55

Multivariate/ multidimensional data visualization

56

Multivariate/multidimensional data visualization

Visualize all variables at the same time

- Glyphs
- 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

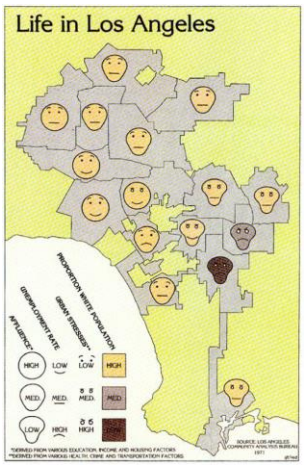
Chernoff faces

A type of glyphs

Can present up to 18 distinct variables

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

Questionable generalization



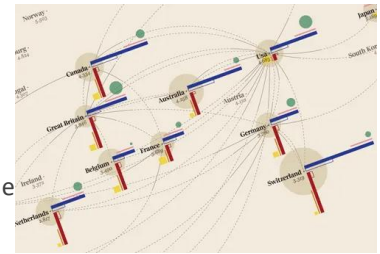
https://mapdesign.icaci.org/2014/12/mapcarte-353365-life-in-los-angeles-by-eugene-turner-1977/

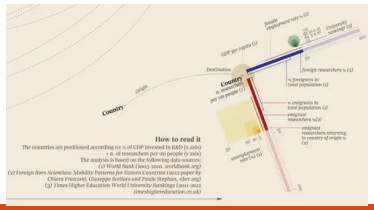
Custom glyphs

Exploring the global “brain drain” in science

Variables shown

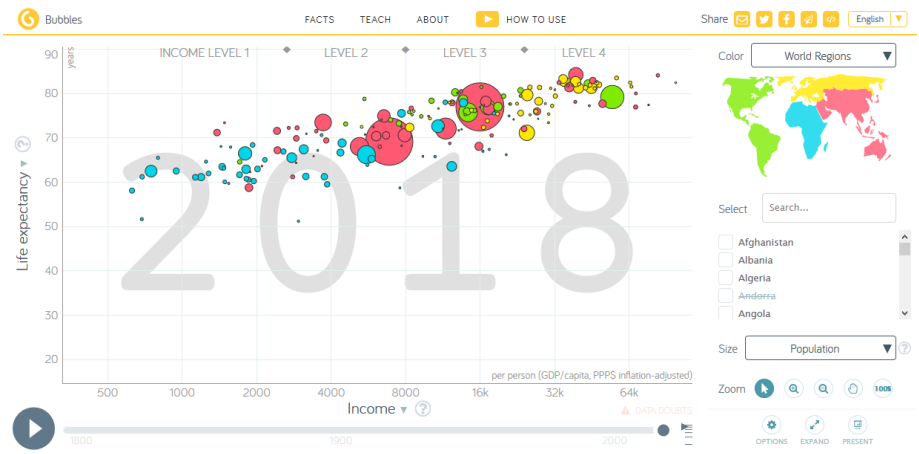
- Percentage of GDP devoted to R&D
- Number of researchers per million people
- Unemployment rate
- Female unemployment rate
- Percentage of foreigners in population
- Percentage of emigrants in population
- Emigrant researchers
- Emigrant researchers returning to their country of origin.





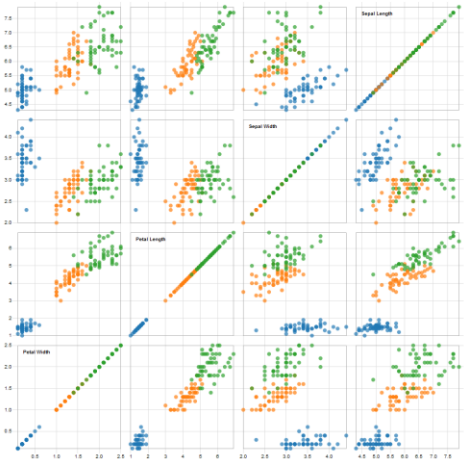
https://www.themarginalian.org/2013/02/13/giorgia-lupi-brain-drain/ 59

Bubble chart



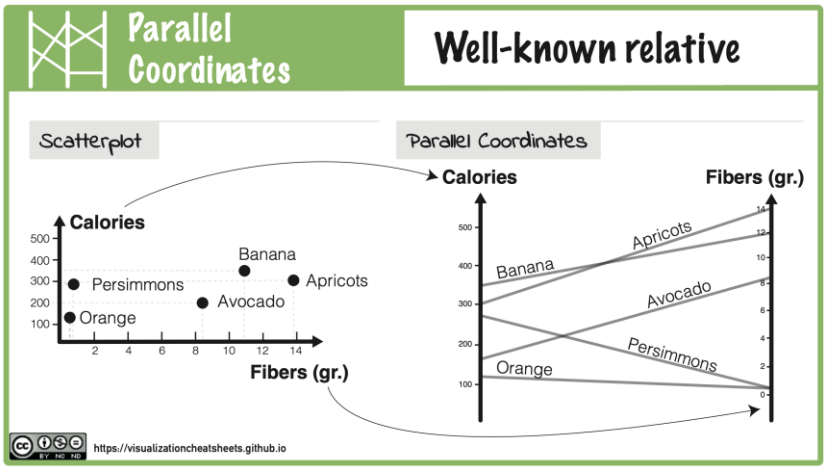
https://www.gapminder.org/tools/ 60

Scatter plot matrix



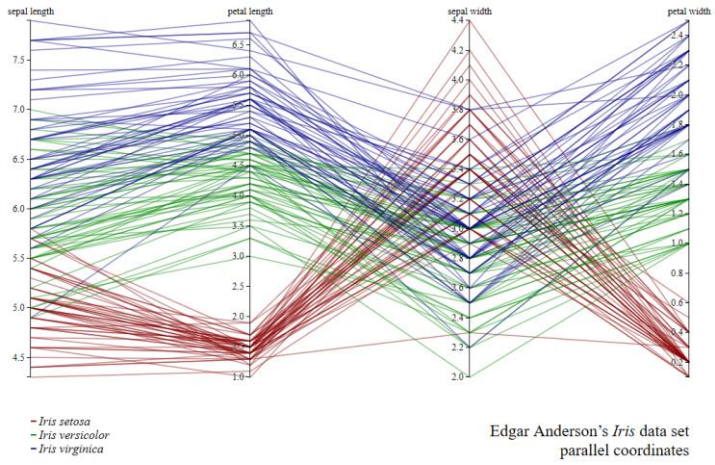
https://bl.ocks.org/mbostock/4063663 61

Parallel coordinate plot



https://visualizationcheatsheets.github.io/ 62

Parallel coordinate plot



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

63

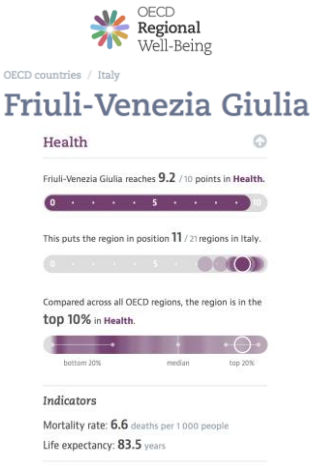
Radar chart

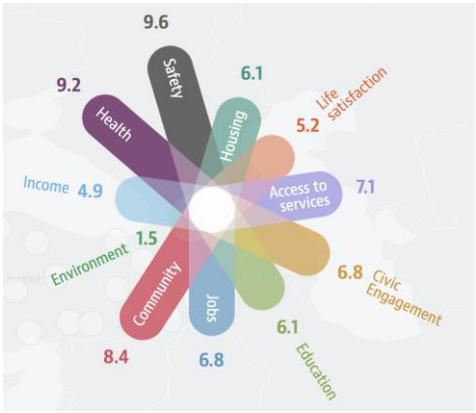
Can be misleading because the size of areas depends on neighboring categories

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

64

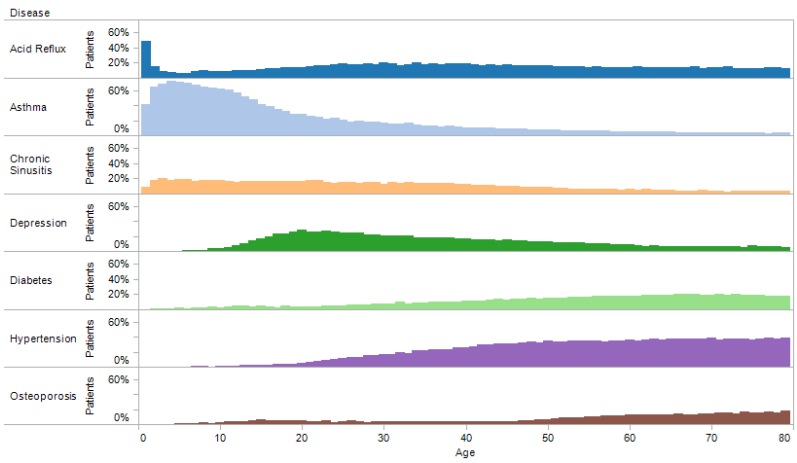
Radial histogram





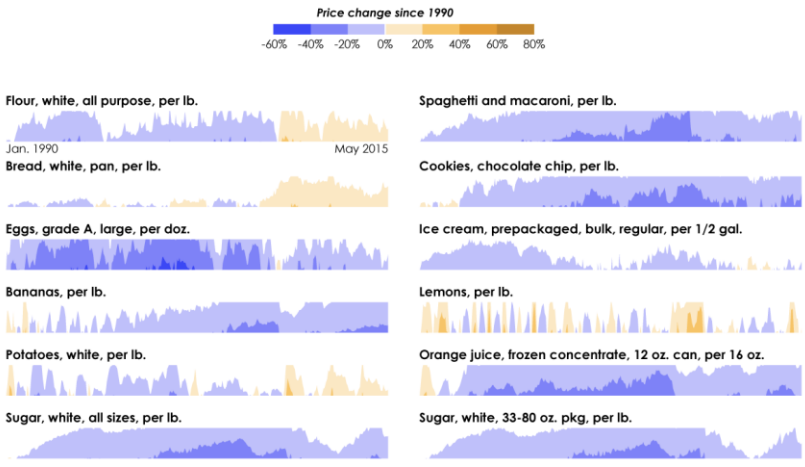
https://www.oecdregionalwellbeing.org/ITH4.html

Small multiples



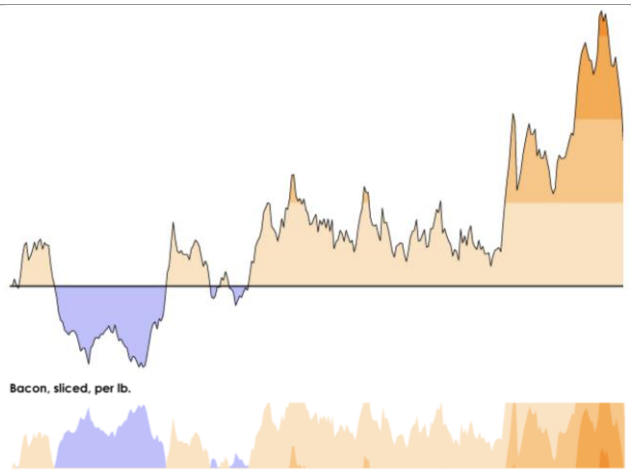
https://eagereyes.org/criticism/curing-a-sick-chart

Horizon charts



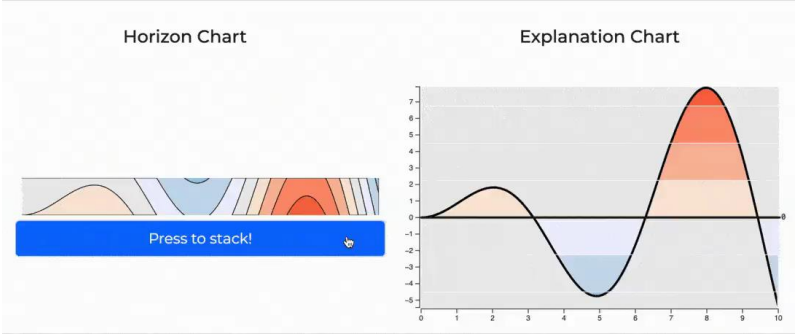
https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/ 67

Horizon charts



https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/ 68

Horizon charts



https://twitter.com/Dev_Lange/status/1521171397736730624

69

Multivariate/multidimensional data visualization

Perform dimensionality reduction and visualize the results

- Principal component analysis
- Multidimensional scaling

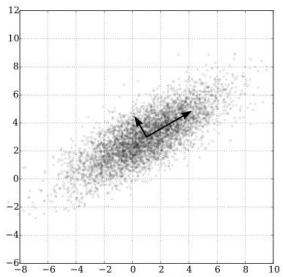
Transformation $R^n \rightarrow R^2$

70

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



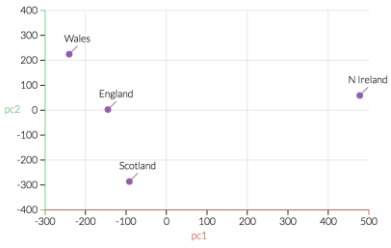
https://en.wikipedia.org/wiki/Principal_component_analysis

71

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

http://setosa.io/ev/principal-component-analysis/

72

Multidimensional scaling

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

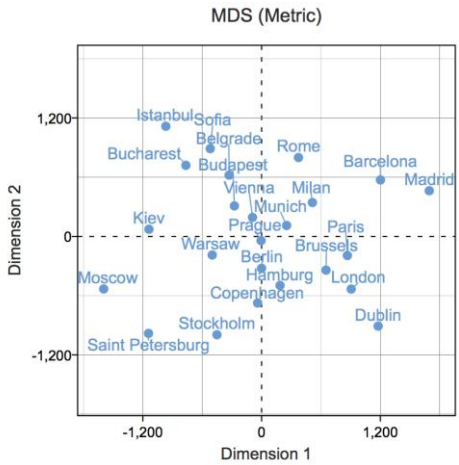
Useful for visualizing similarity matrices or graphs where you wish to preserve distances between nodes

Minimize the stress function

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

Solve with any method for optimizing nonlinear functions

Multidimensional scaling

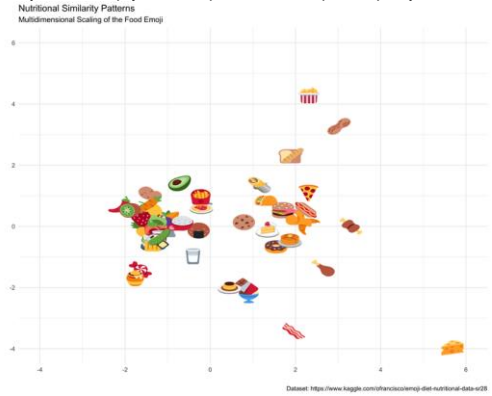


https://www.displayr.com/goodness-of-fit-in-mds-and-t-sne-with-shepard-diagrams/

Multidimensional scaling

Nutritional information for 58 types of food

- 35 variables (calories, protein, vitamin a, zinc, …)



https://grace.rbind.io/post/2019-02-22-emoji-mds/

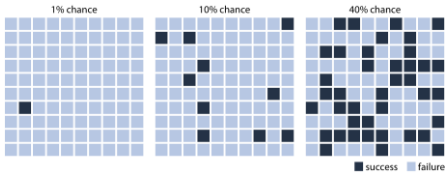
75

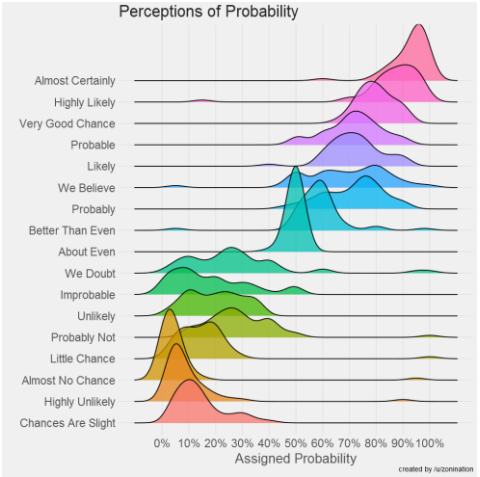
Visualizing uncertainty and missing data

76

Visualizing uncertainty

Uncertainty (confidence intervals, etc.) hard to understand





https://clauswilke.com/dataviz/visualizing-uncertainty.html
https://github.com/zonination/perceptions

77

Visualizing uncertainty

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

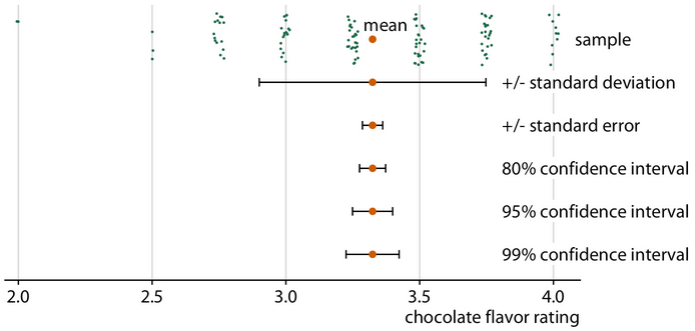
Techniques to show uncertainty
- Ranges
- Distributions
- Multiple outcomes
- Obscurity

https://www.iqt.org/wp-content/uploads/2017/09/Uncertainty-Report_PUBLIC.pdf
https://flowingdata.com/2018/01/08/visualizing-the-uncertainty-in-data/

78

Visualizing uncertainty with ranges

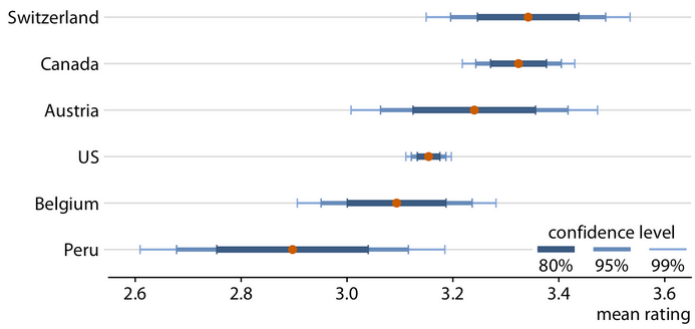
Specify what the range represents



https://serialmentor.com/dataviz/visualizing-uncertainty.html 79

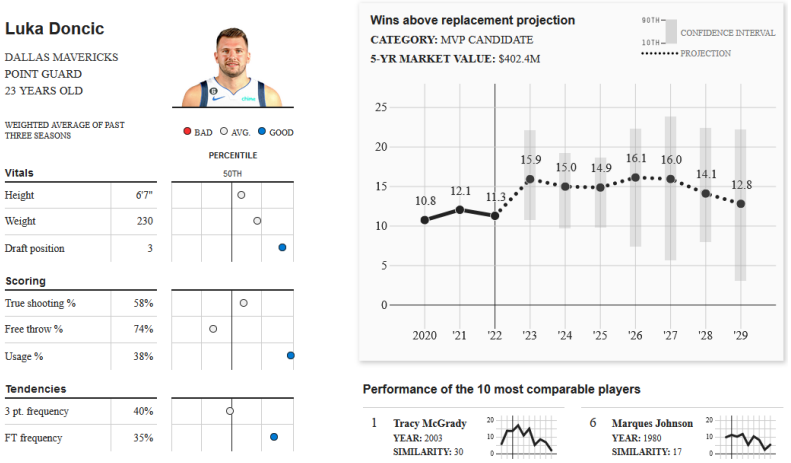
Visualizing uncertainty with ranges

Specify what the range represents



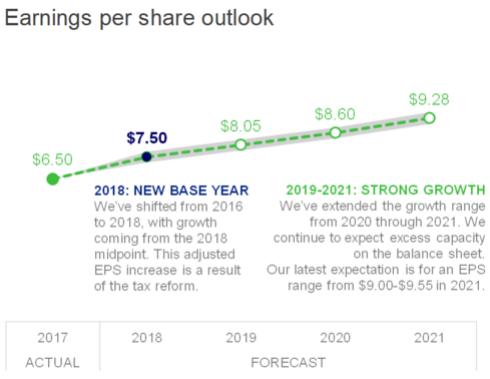
https://serialmentor.com/dataviz/visualizing-uncertainty.html 80

Visualizing uncertainty with ranges



https://projects.fivethirtyeight.com/2023-nba-player-projections/luka-doncic/ 81

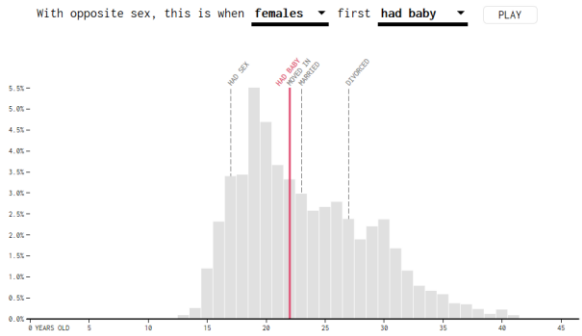
Visualizing uncertainty with ranges



https://www.storytellingwithdata.com/blog/2018/6/27/visualizing-uncertainty 82

Visualizing uncertainty with distributions

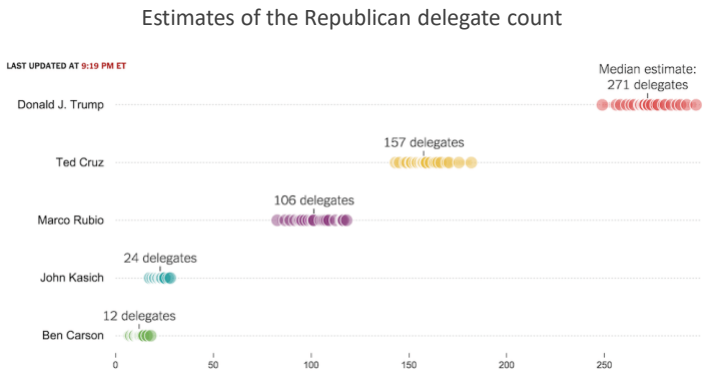
Show the spread of possible values with a histogram (or a variant of it)



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

Visualizing uncertainty with multiple outcomes

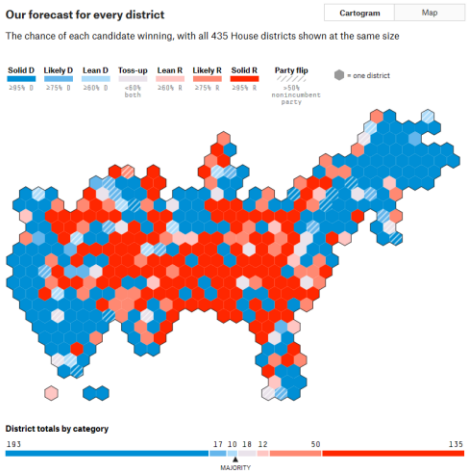
Show the various outcomes



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

Visualizing uncertainty with obscurity

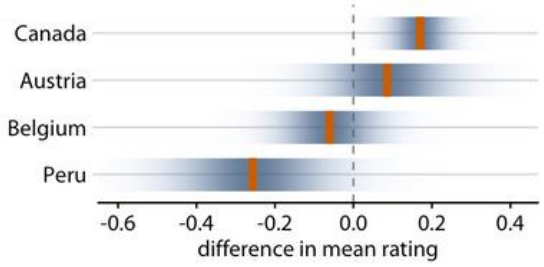
Use transparency, color scale, or blurriness to show uncertainty



https://projects.fivethirtyeight.com/2018-midterm-election-forecast/house/ 85

Visualizing uncertainty with obscurity

Use transparency, color scale, or blurriness to show uncertainty



https://serialmentor.com/dataviz/visualizing-uncertainty.html 86

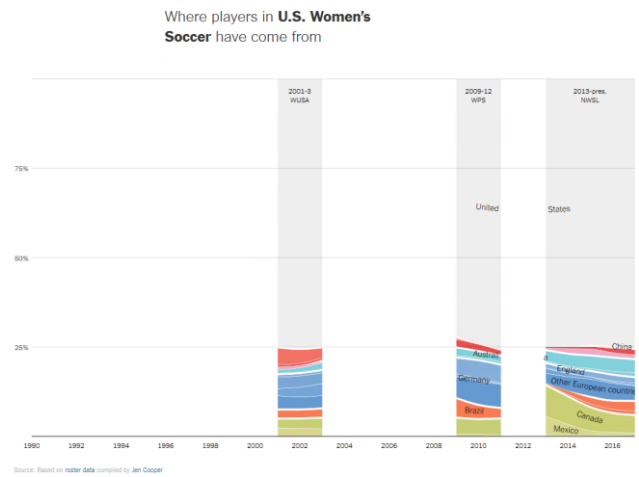
Visualizing missing data

Techniques to handle missing data

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

https://flowingdata.com/2018/01/30/visualizing-incomplete-and-missing-data/ 87

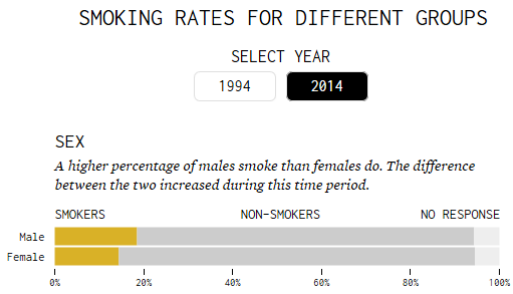
Visualizing missing data by showing the gaps



https://www.nytimes.com/interactive/2017/12/29/upshot/internationalization-of-pro-sports-leagues-premier-league.html 88

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/

89

Interactivity

90

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

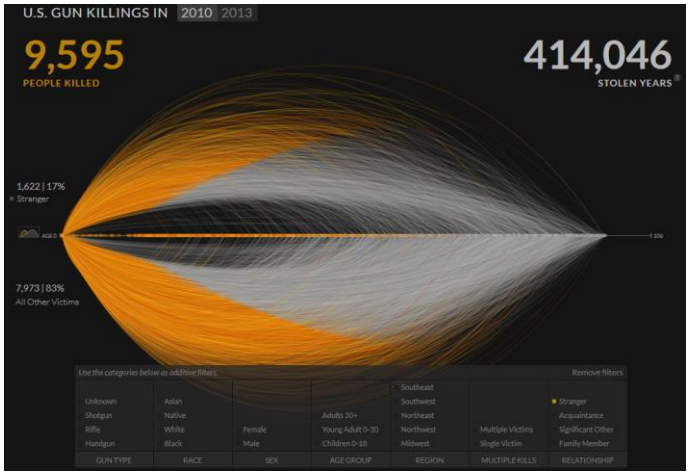
Can affect

- 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

93

Navigating

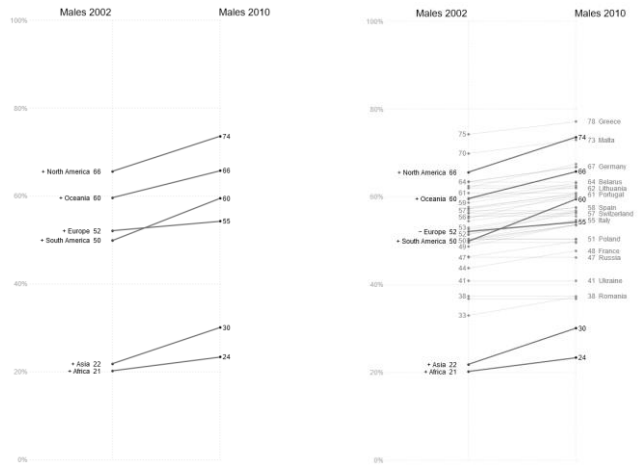
Road orientation map

https://mourner.github.io/road-orientation-map/#10.65/40.6758/-74.0034

94

Navigating

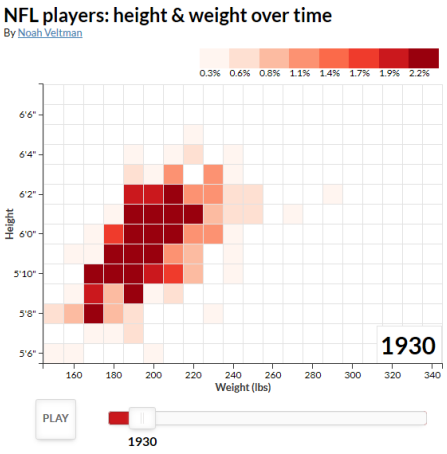
Obesity around the world

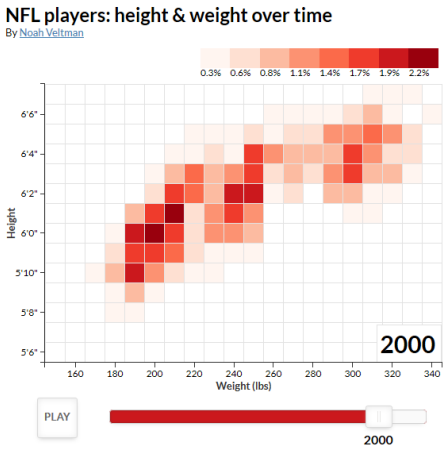


http://neoformix.com/Projects/ObesitySlope/

95

Animating

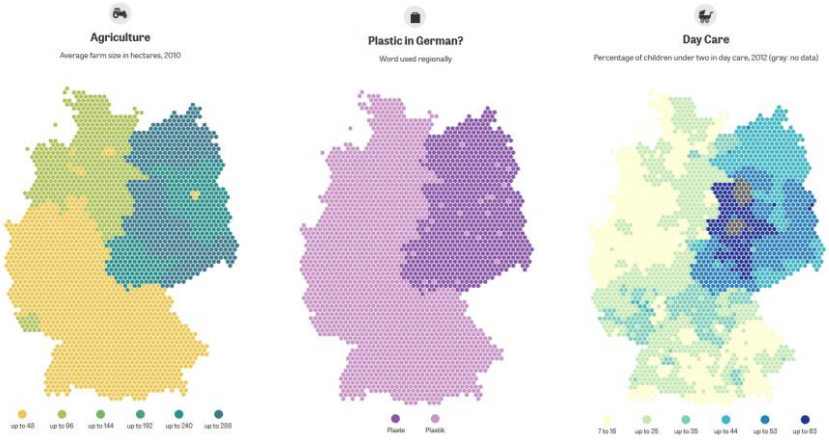




https://noahveltman.com/nflplayers/

96

Sequencing



https://www.zeit.de/feature/german-unification-a-nation-divided 97

Contributing

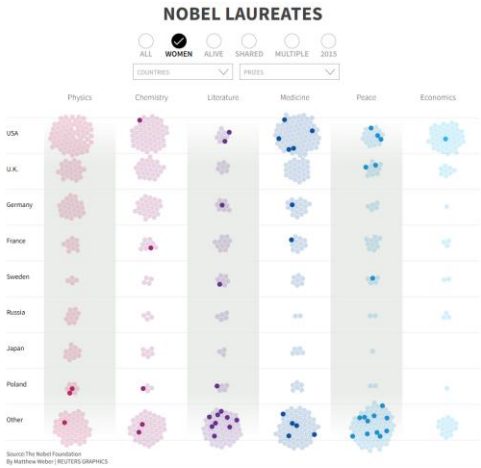
http://www.informationisbeautiful.net/visualizations/who-old-are-you/ 98

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

A. Kirk. *Data Visualization*, SAGE Publications, 2016. 99

Focusing



http://graphics.thomsonreuters.com/15/nobels/index.html 100

Focusing

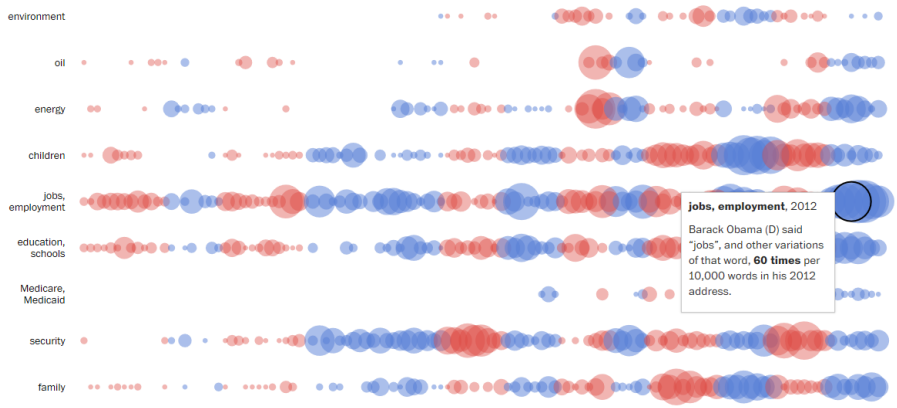
Using brushing and linking

https://observablehq.com/@d3/brushable-scatterplot-matrix

101

Annotating

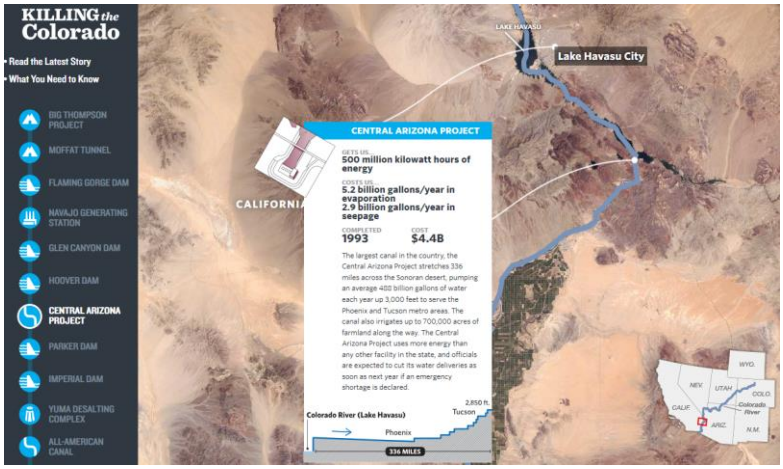
History through the president's words



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

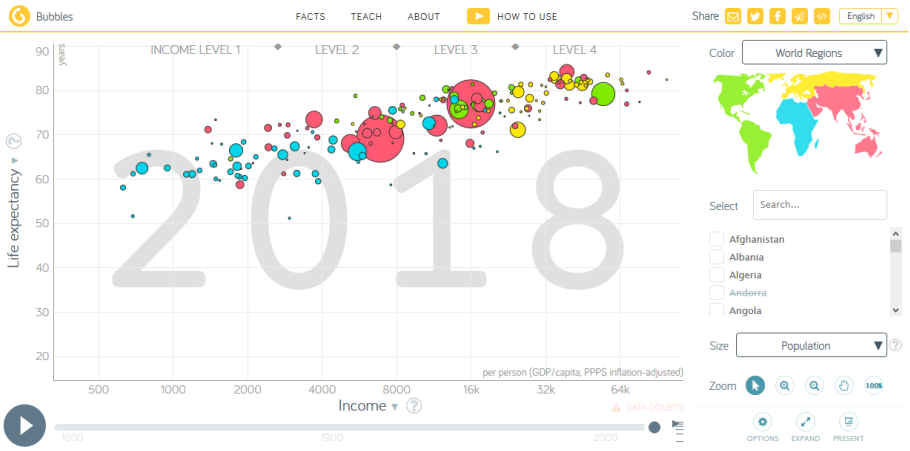
102

Orientating



https://projects.propublica.org/killing-the-colorado/explore-the-river#central-arizona-project 103

Interactivity example: Gapminder



https://www.gapminder.org/tools/ and https://www.youtube.com/watch?v=hVimVzgtD6w 104

Storytelling

Storytelling

Storytelling ≠ making something up

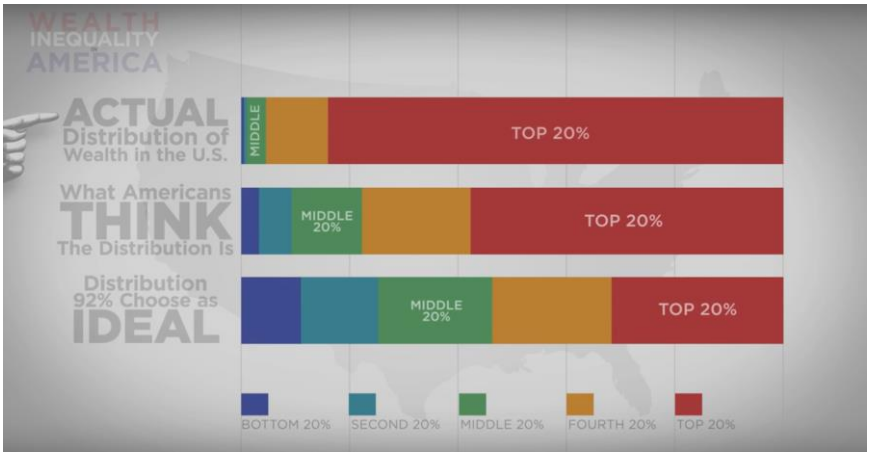
Visualization can be used to tell a story

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

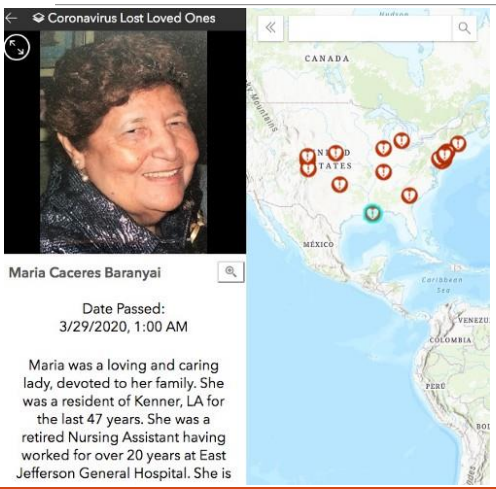
http://www.thefunctionalart.com/2014/04/annotation-narrative-and-storytelling.html

Storytelling example



https://www.youtube.com/watch?v=QPKKQnijnsM 107

Enforce emotions



Emotions can be enforced by providing stories about individual experiences

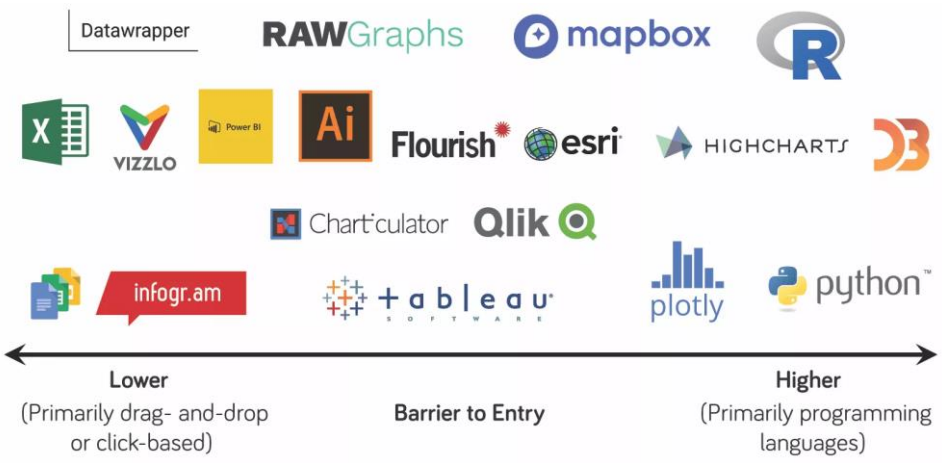
They move us more than statistics about a large number of people

Best: do both

https://medium.com/nightingale/ten-ways-cognitive-biases-impact-data-design-work-be83f86d4274 108

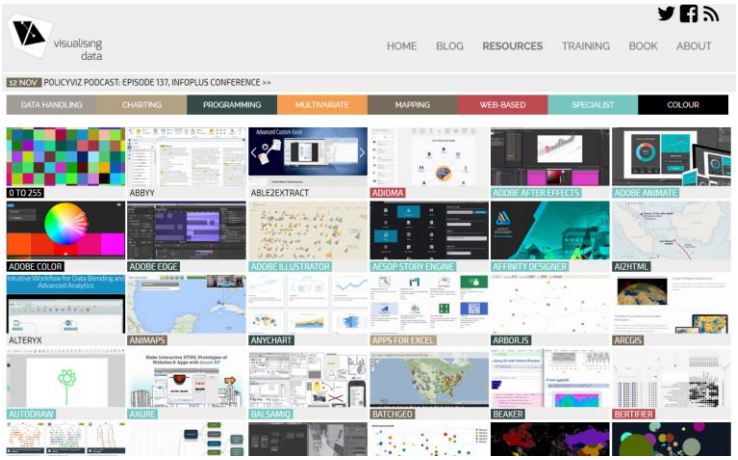
Tools

Tools



https://policyviz.com/2022/02/01/the-data-visualization-tools-wars/

Tools



http://www.visualisingdata.com/resources/ 111

D3

Data-Driven Documents

A JavaScript library

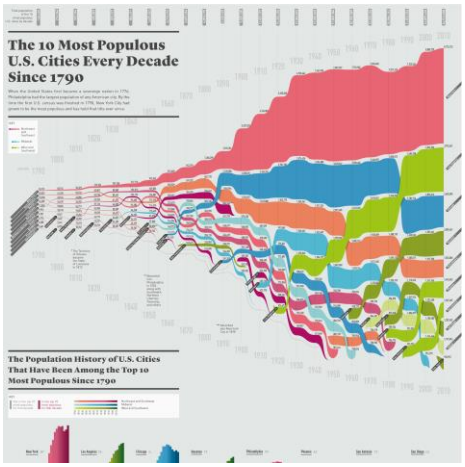
Emphasis on interactivity

https://d3js.org/ 112

RAW Graphs

Open source tool for data visualization

Politecnico di Milano



https://www.rawgraphs.io/

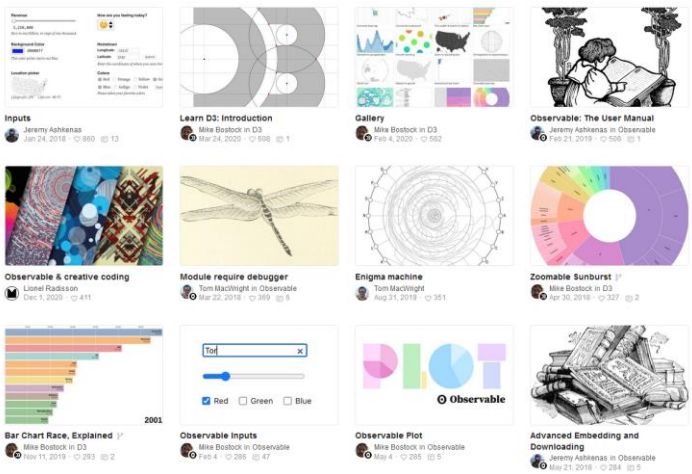
113

Observable

Interactive notebooks supporting collaboration

JavaScript

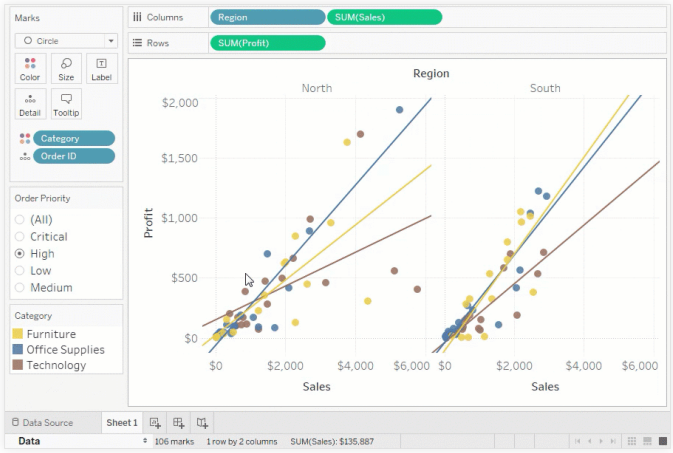
Free plan



https://observablehq.com/

114

Tableau Public



Does not require programming skills

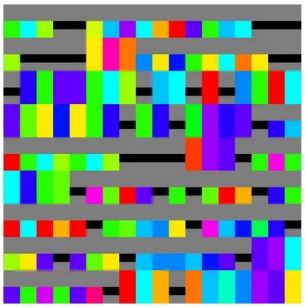
Visualizations created with the free version are public

Students can get a free version with private use

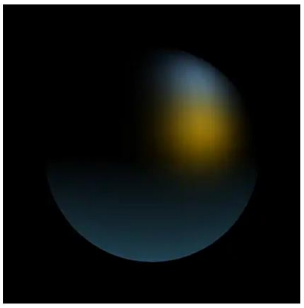
https://public.tableau.com/s/

115

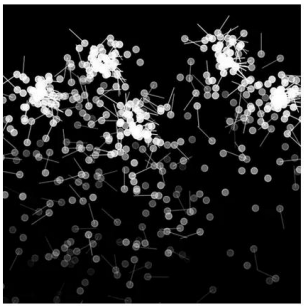
Processing



Keyboard Functions



Spot



Multiple Particle Systems

Software for coding within the context of the visual arts

Free

Interfaces for JavaScript, Python, Android and Raspberry Pi

https://processing.org

116