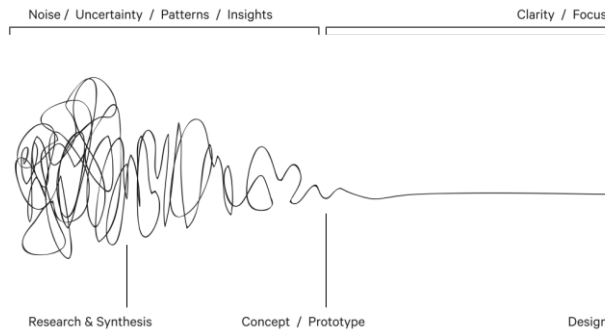


Data Visualization

VISUALIZATION DESIGN

Tea Tušar, Data Science and Scientific Computing, Information retrieval and data visualization

Visualization design



<https://thedesigntquiggle.com/>

2

Overview

The 7 steps of visualization design

Basic charts

Multivariate/multidimensional data visualization

Visualizing uncertainty and missing data

Interactivity

Storytelling

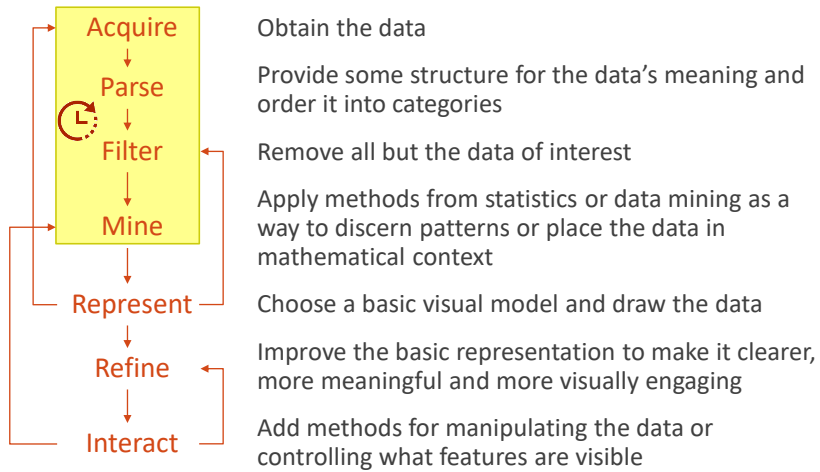
Tools

3

The 7 steps of visualization design

4

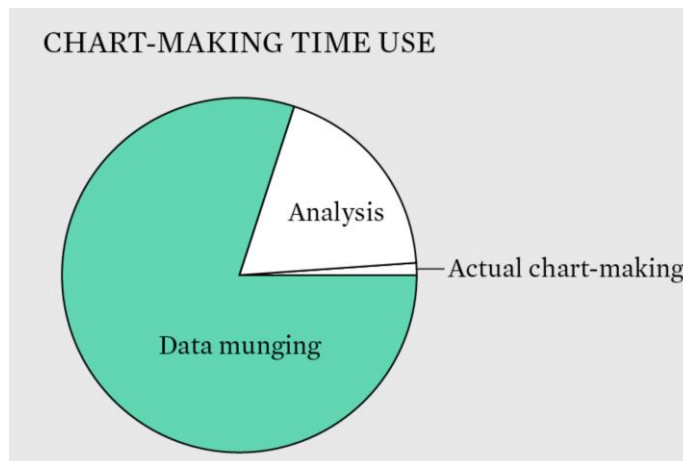
The 7 steps of visualization design



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

5

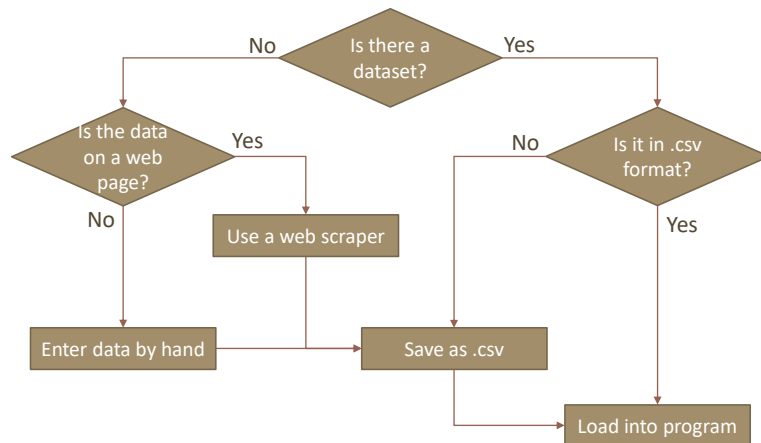
The 7 steps of visualization design



<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, ...

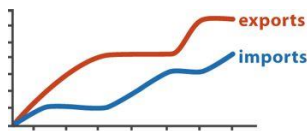
(Dis)aggregate data

8

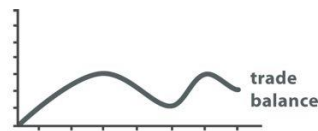
Parse the data

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



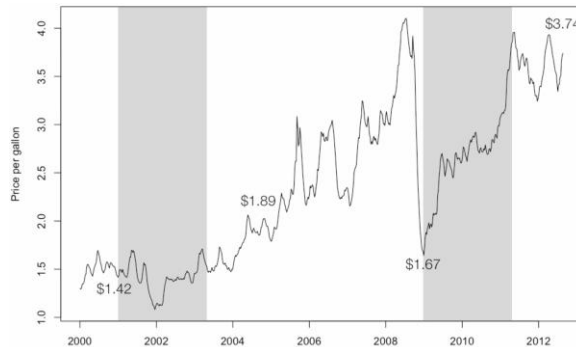
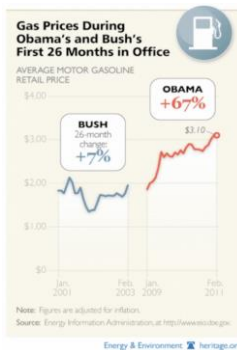
$trade\ balance = exports - imports$

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

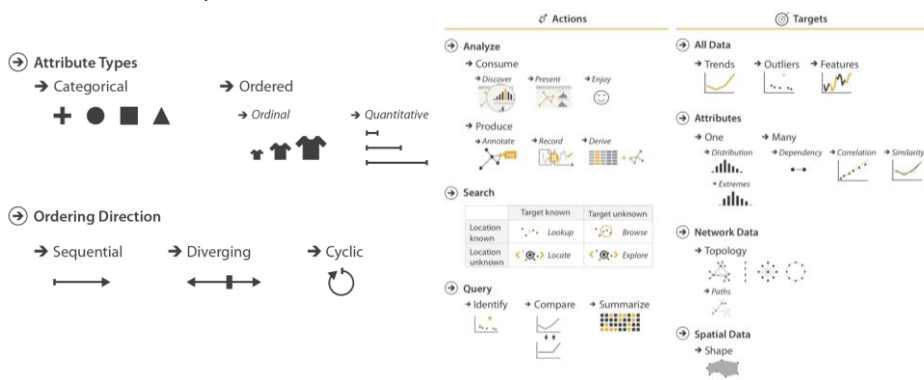
- Look for important features and patterns
- Look for any striking deviations (outliers)
- Interpret your findings

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

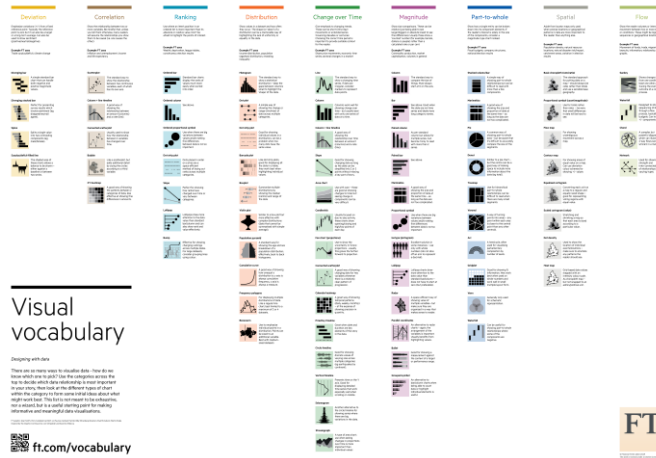
Represent the data

Choose a basic visual model and draw the data

Choice depends on **the data and the task**



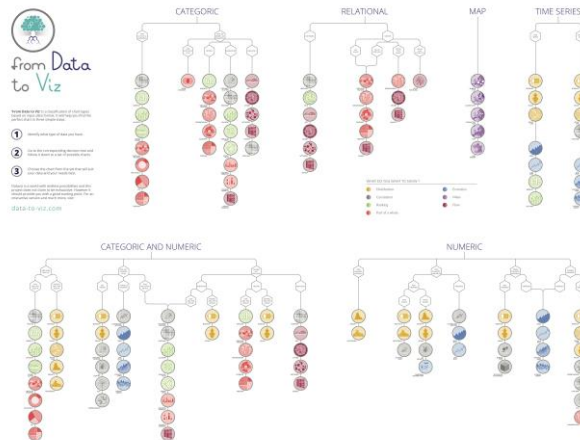
Represent the data



<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

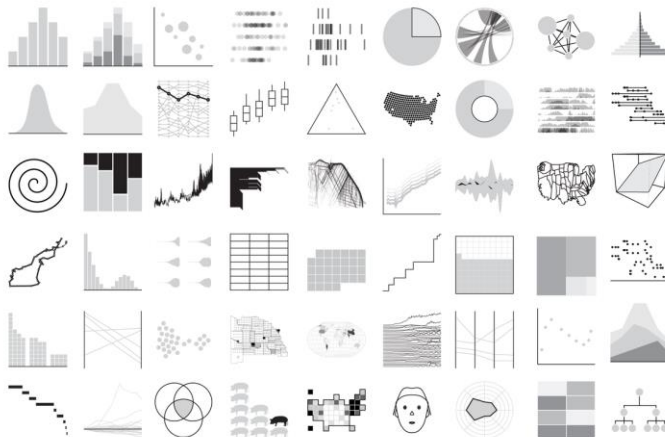
Represent the data



<https://datavizproject.com/>

19

Represent the data



<https://flowingdata.com/chart-types>

20

Represent the data

Distribution

- Violin
- Scatter
- Histogram
- Boxplot
- Wiggleline

Evolution

- Line plot
- Area
- Stacked area
- Streamchart

Correlation

- Scatter
- Heatmap
- Correlation
- Bubble
- Connected scatter
- Density 2D

Map

- Map
- Choropleth
- Heatmap
- Cartogram
- Concentric
- Bubble map

Ranking

- Barplot
- Spider / Radar
- Wordcloud
- Panel
- Lollipop
- Circle layout

Flow

- Cloud diagram
- Network
- Stacked
- Arc diagram
- Edge bundling

Part of a whole

- Timeline
- Donut chart
- Pie chart
- Stacked bar
- Circle packing

General knowledge

- Basic
- Custom
- Interactivity
- Shape helpers
- Context
- Data art

<https://www.d3-graph-gallery.com/>

21

Represent the data

treavis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

Dimensionality: All [Icons]

Representation: All [Icons]

Alignment: All [Icons]

Fulltext Search: [Input] x

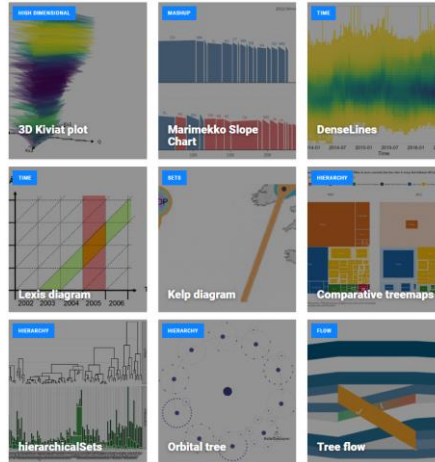
Techniques Shown: 306

[Grid of 40 thumbnail images showing various tree visualization techniques]

<http://treavis.net/>

22

Represent the data

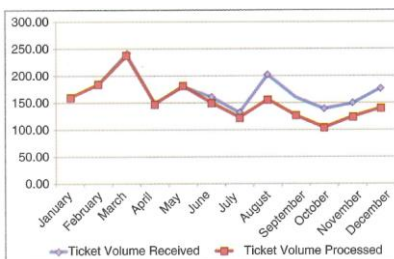


<https://www.xeno.graphics/>

25

Refine the visualization

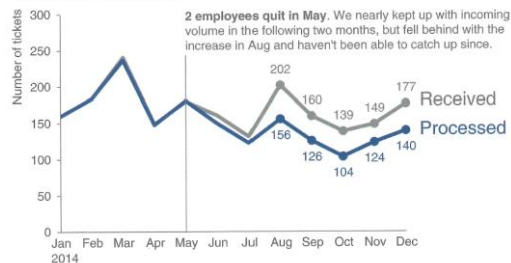
Improve the basic representation to make it clearer, more meaningful and more visually engaging



Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

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

26

Refine the visualization: Use takeaway titles

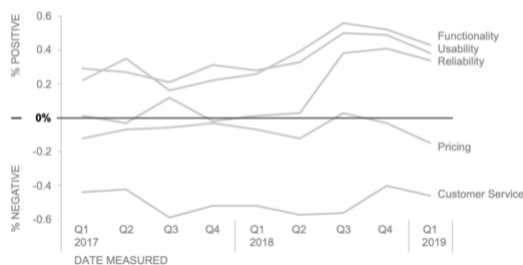
Customer Service ranks lowest

Action needed to address **recent decline**

Q1 2019: all topics declined

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

Customer topic sentiment



Use action-oriented titles

Aim for one line

27

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*

28

Basic charts

29

Basic charts

Bar charts

Line chart

Pie charts

Geographical data

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

Networks and trees

- Node-link diagrams
- Matrices

30

Bar charts

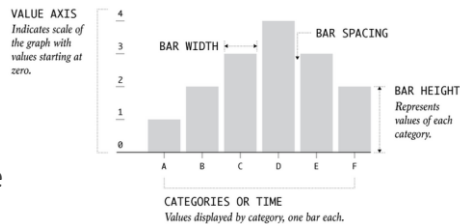
Use them to show values per categories (or discrete time)

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/>

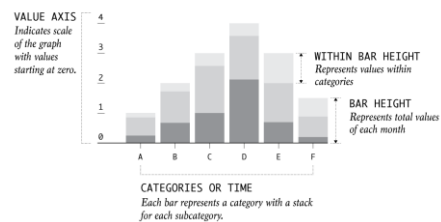
31

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



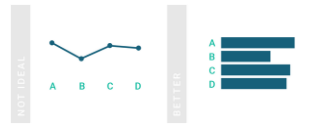
<https://flowingdata.com/charttype/stacked-bar-chart/>

32

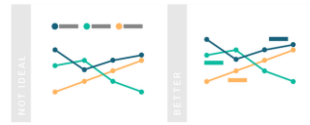
Line charts

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



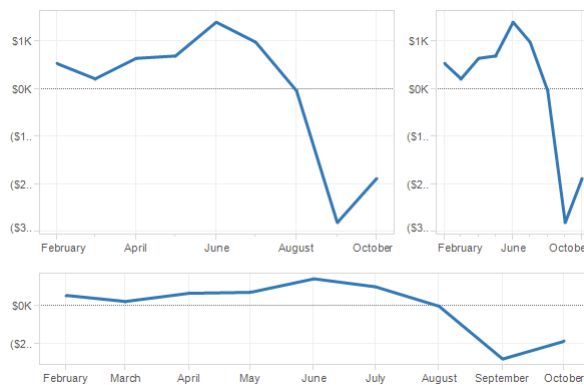
Extend the y-axis to 0 (or the 'historic low' value)

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



<https://blog.datawrapper.de/line-charts/>

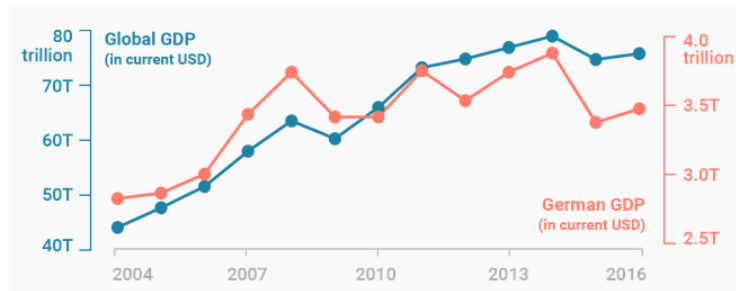
Line charts: Aspect ratio bias



Banking to 45 Degrees

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

Line charts: Avoid dual axis charts

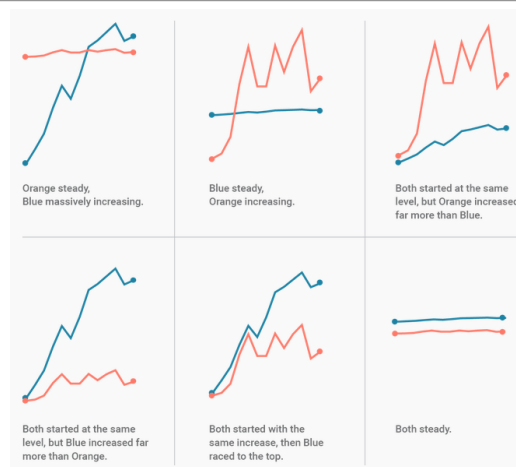


Easily manipulated to mislead
Hard to read

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

35

Line charts: Avoid dual axis charts



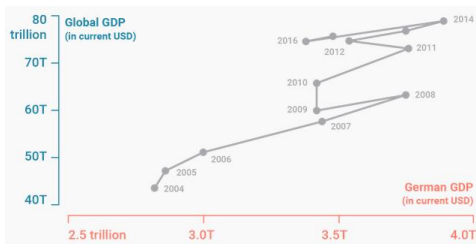
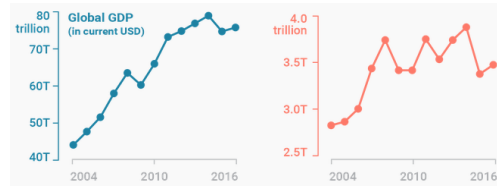
<https://blog.datawrapper.de/dualaxis/>

36

Line charts: Avoid dual axis charts

Alternatives

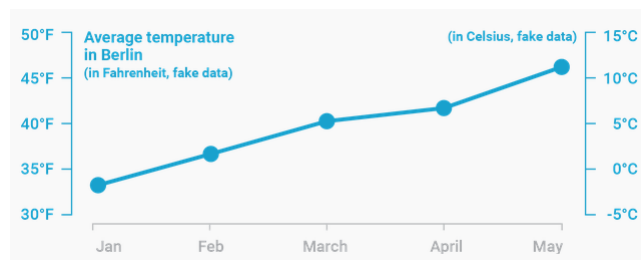
- Side-by-side charts
- Connected scatter plots



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

37

Line charts: Dual axes sometimes okay



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

38

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

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

No 3-D

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



<https://blog.datawrapper.de/pie-charts/>

39

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

40

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

41

Dot maps



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

42

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

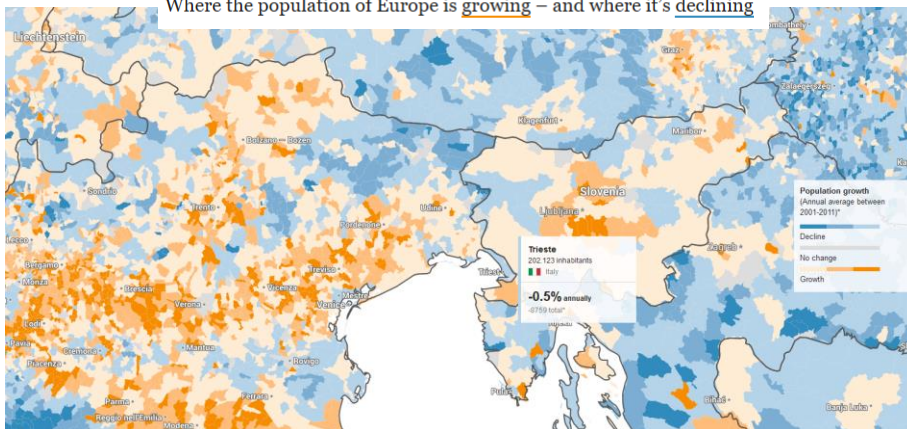
Be careful in choosing bin size

Be careful in choosing colors

43

Choropleth maps

Where the population of Europe is growing – and where it's declining



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

44

(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

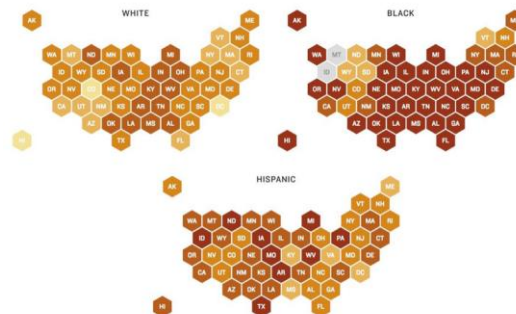
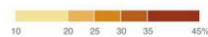
Often hexagonal or square

Harder to locate the given state/province

45

(Hexagon) tile maps

Obesity Prevalance In 2014, By Race



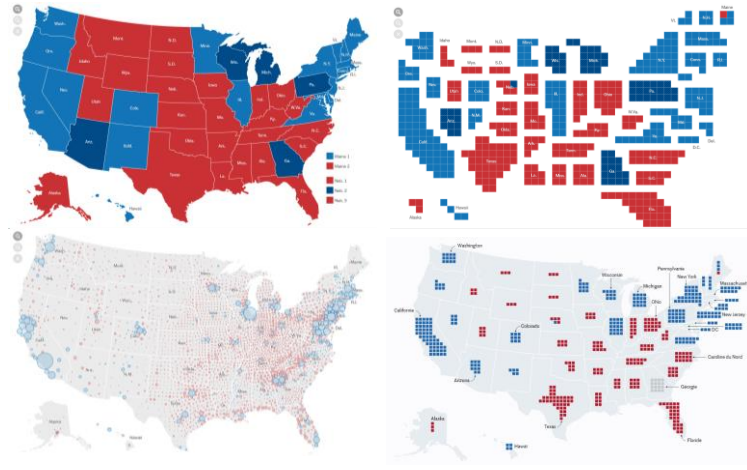
Notes

- Figures for white and black populations do not include Hispanics.
- Insufficient data to calculate prevalence rates for black residents of Idaho and Montana.

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

46

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

47

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

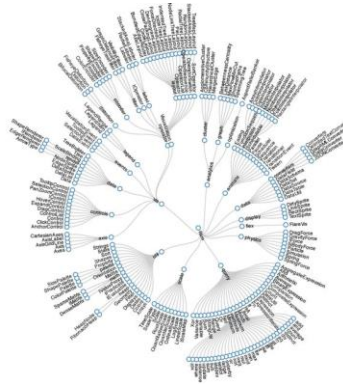
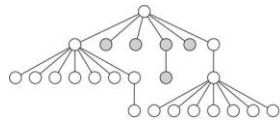
48

Node-link diagrams

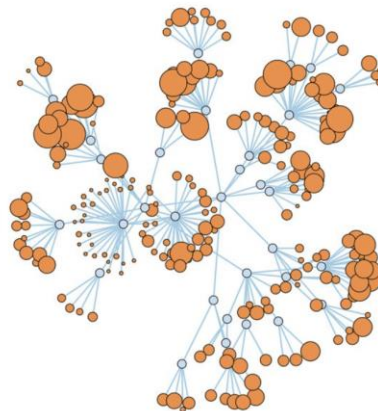
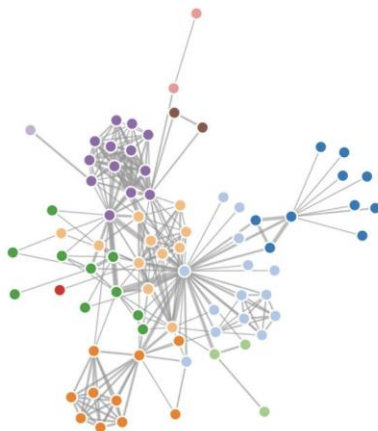
Trees

Layout depends on size

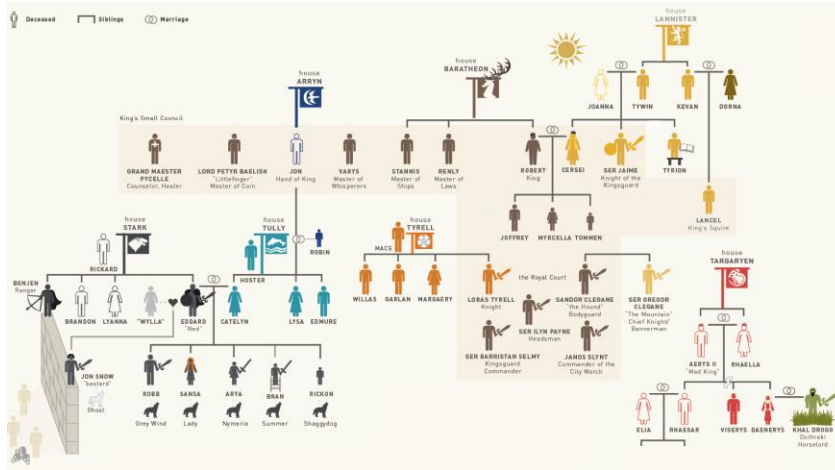
- Triangular vertical (small trees)
- Spline radial (large trees)



Node-link diagrams



Node-link diagrams



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

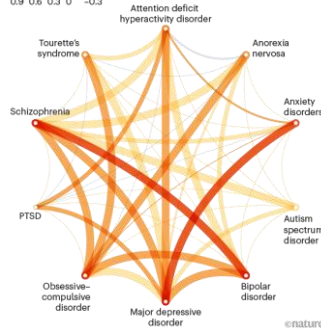
Node-link diagrams

MENTAL MAP

Similar genetic variants seem to underlie a number of psychiatric disorders. In one study of 200,000 people, schizophrenia was significantly correlated with most other disorders. By contrast, some disorders such as post-traumatic stress disorder (PTSD) showed only weak correlations to other conditions.

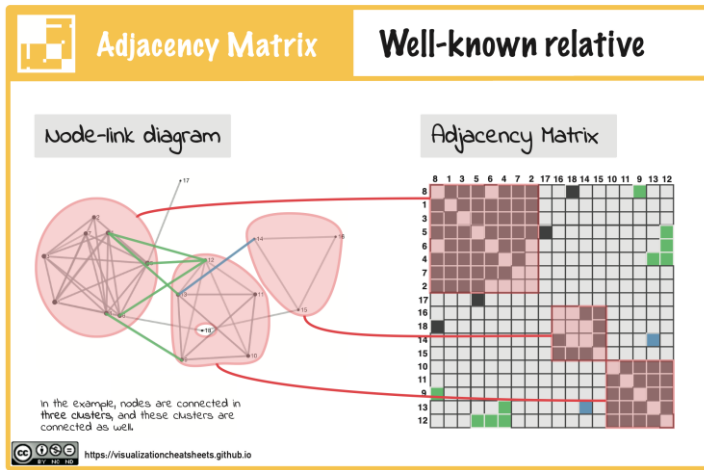
P-value significance
 ■ <0.00035 ■ <0.001 — <0.05 — >0.05

Genetic correlation
 0.9 0.6 0.3 0 -0.3



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

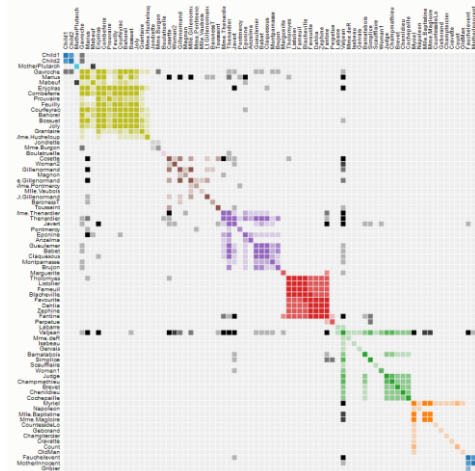
Adjacency matrix



<https://visualizationcheatsheets.github.io/>

53

Adjacency matrix



Co-occurrence of
characters in
Les Misérables

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

54

Multivariate/ multidimensional data visualization

55

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

56

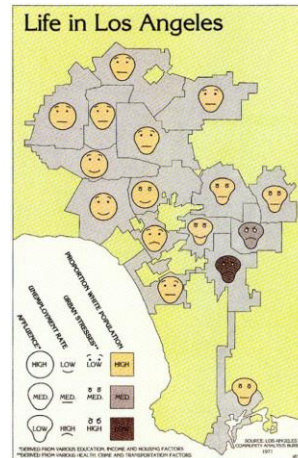
Chernoff faces

A type of glyphs

Can present up to 18 distinct variables

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

Questionable generalization



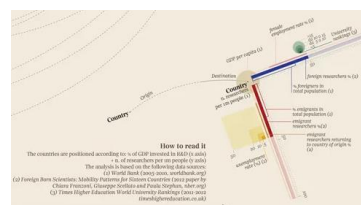
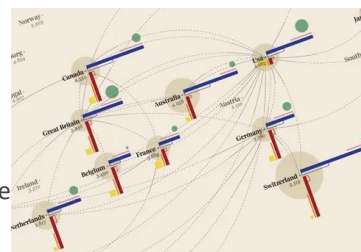
<https://mapdesign.icaci.org/2014/12/mapcarte-353365-life-in-los-angeles-by-eugene-turner-1977/>

Custom glyphs

Exploring the global “brain drain” in science

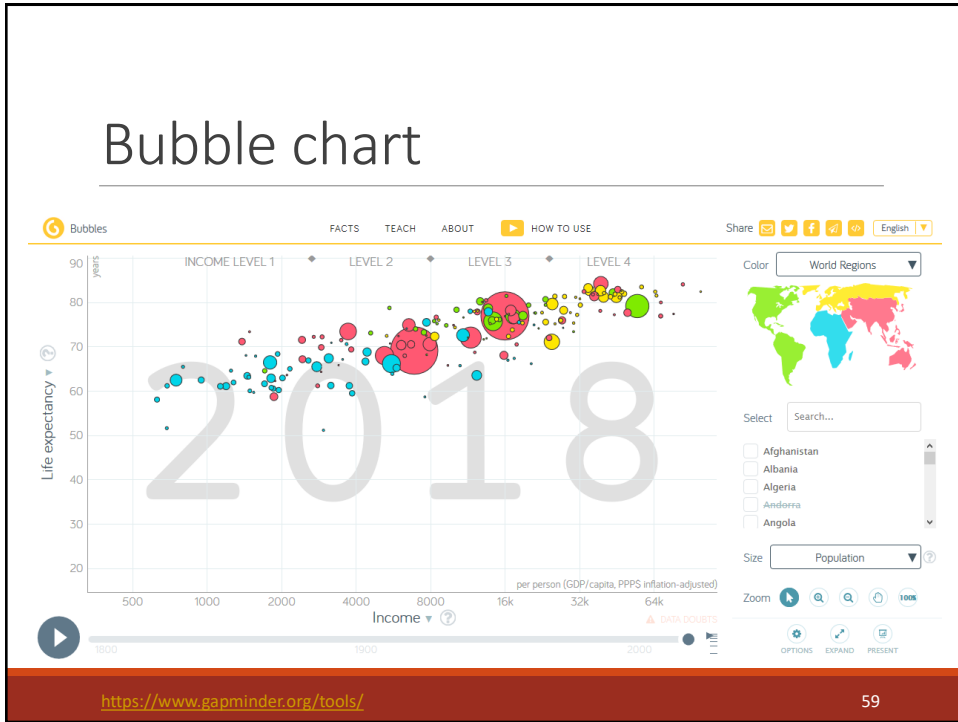
Variables shown

- 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.

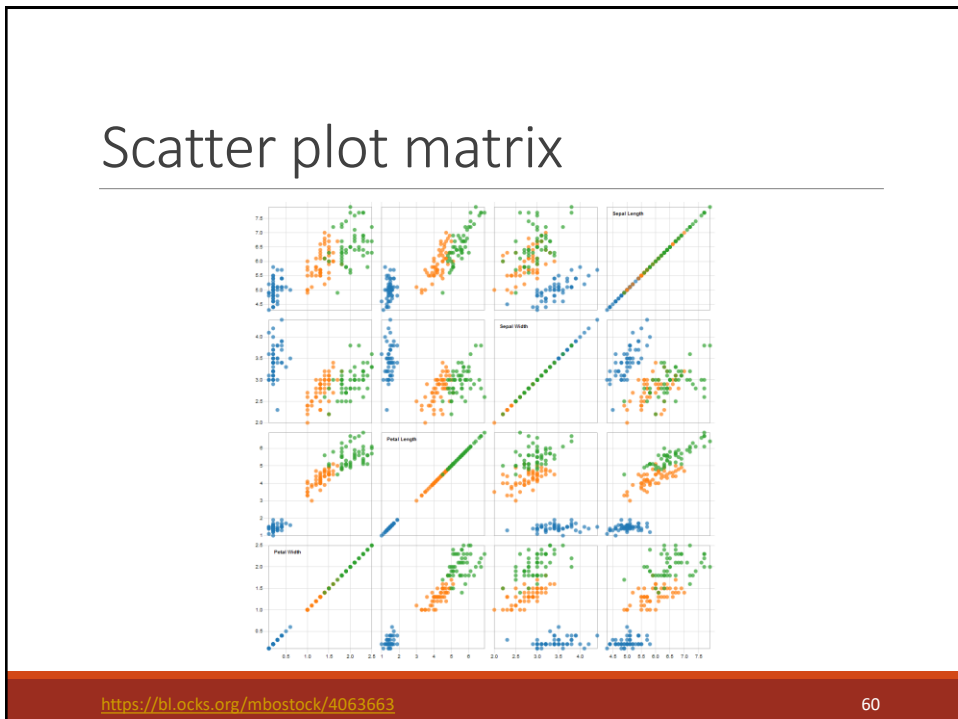


<https://www.themarginalian.org/2013/02/13/giorgia-lupi-brain-drain/>

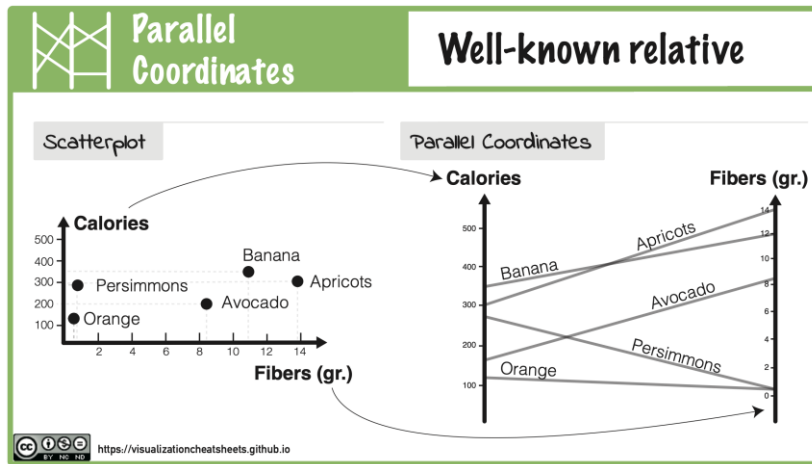
Bubble chart



Scatter plot matrix



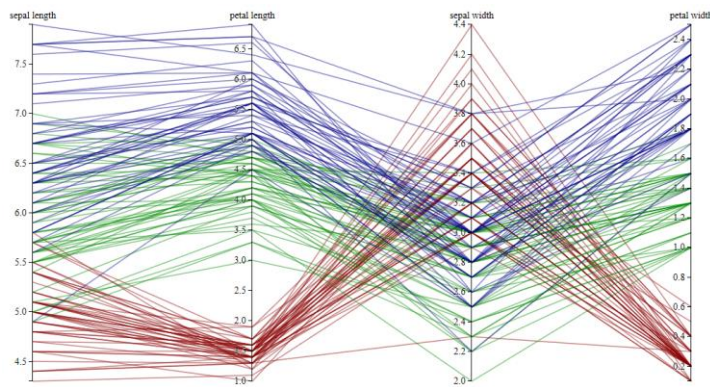
Parallel coordinate plot



<https://visualizationcheatsheets.github.io/>

61

Parallel coordinate plot



- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

Edgar Anderson's *Iris* data set
parallel coordinates

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

62

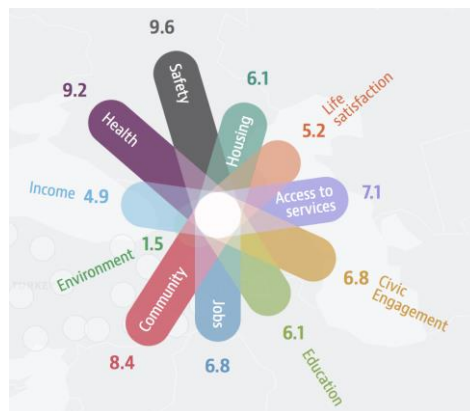
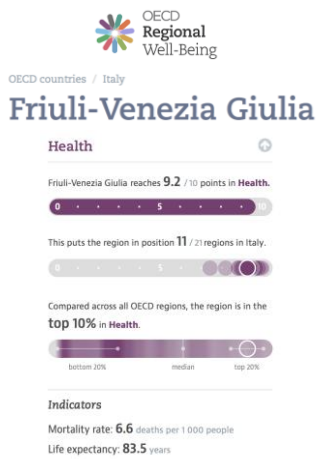
Radar chart



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

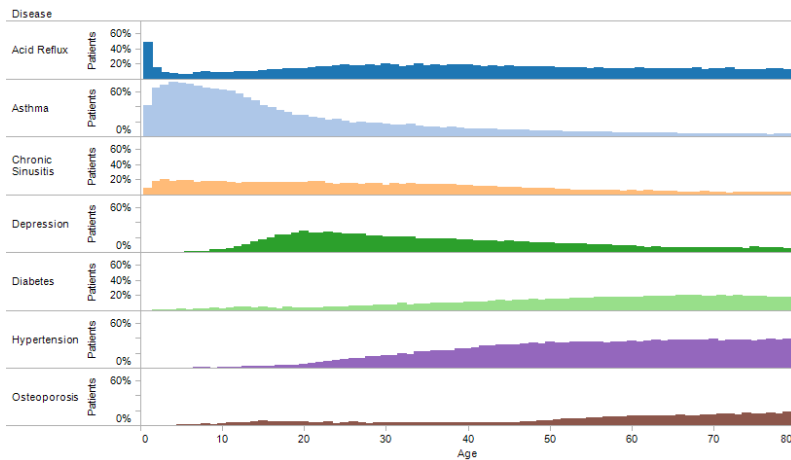
<https://vizzlo.com/create/radar-chart>

Radial histogram



<https://www.oecdregionallwellbeing.org/ITH4.html>

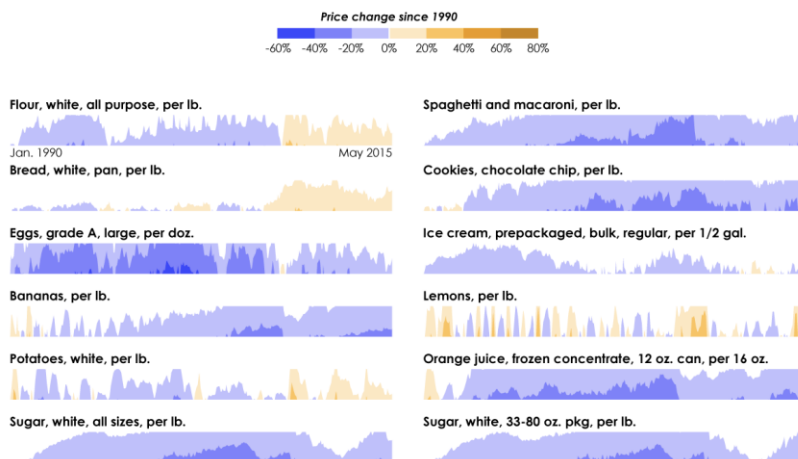
Small multiples



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

65

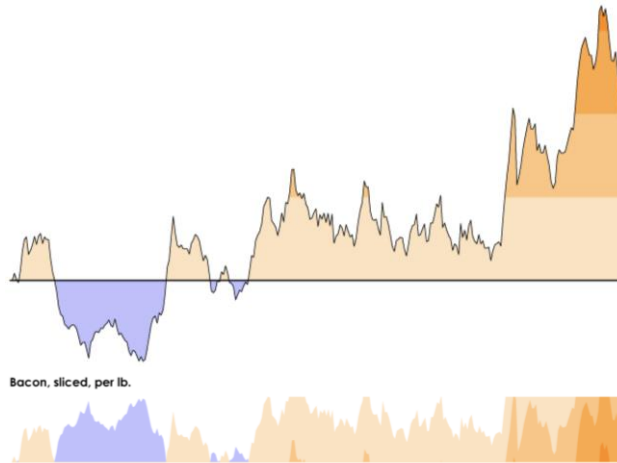
Horizon charts



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

66

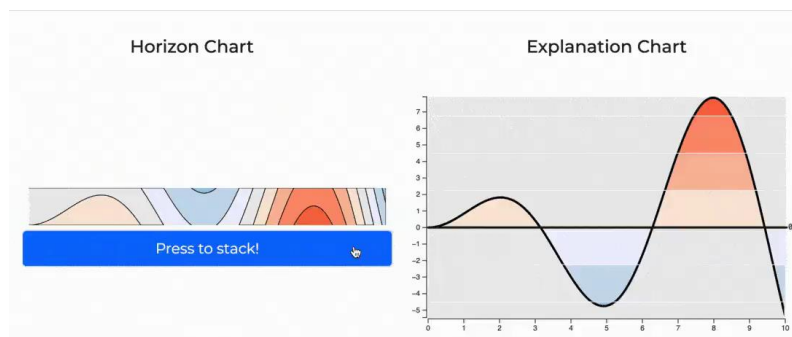
Horizon charts



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

67

Horizon charts



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

68

Multivariate/multidimensional data visualization

Perform dimensionality reduction and visualize the results

- Principal component analysis
- Multidimensional scaling

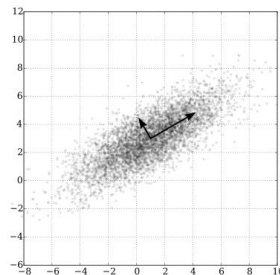
Transformation $R^n \rightarrow R^2$

69

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

70

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/>

71

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

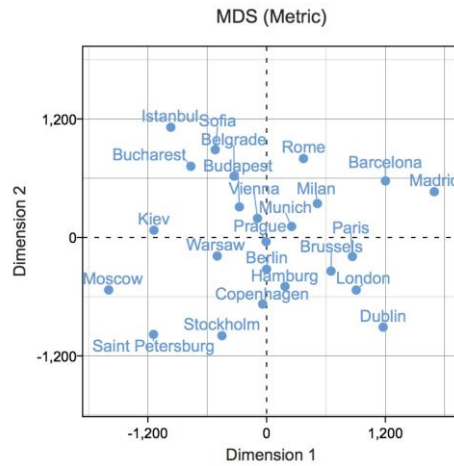
Minimize the stress function

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

Solve with any method for optimizing nonlinear functions

72

Multidimensional scaling



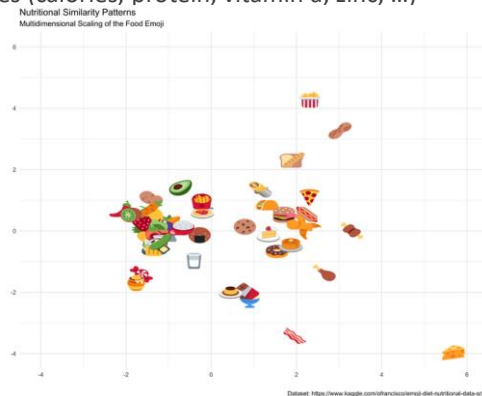
<https://www.display.com/goodness-of-fit-in-mds-and-t-sne-with-shepard-diagrams/>

73

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/>

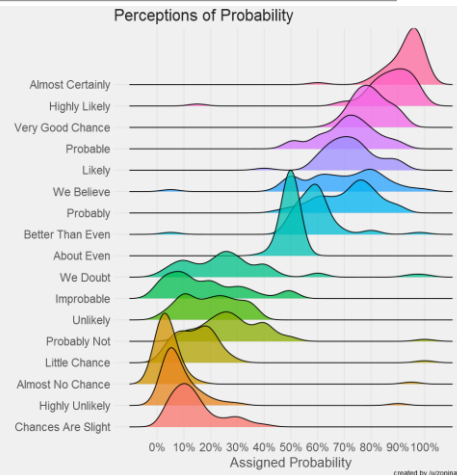
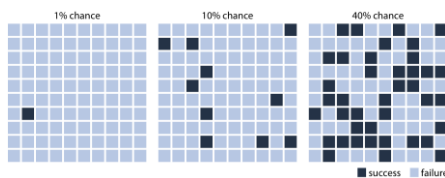
74

Visualizing uncertainty and missing data

75

Visualizing uncertainty

Uncertainty (confidence intervals, etc.) hard to understand



<https://clauswilke.com/dataviz/visualizing-uncertainty.html>
<https://github.com/zoination/perceptions>

76

Visualizing uncertainty

Uncertainty types

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

Techniques to show uncertainty

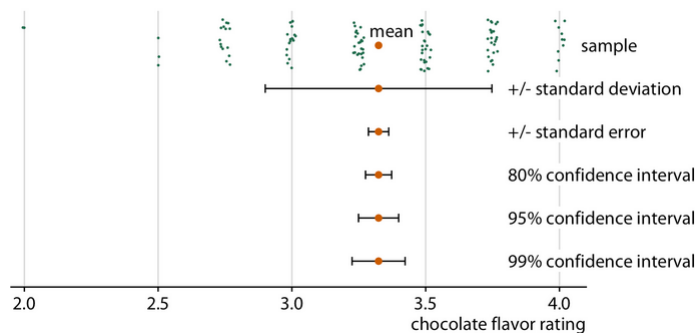
- Ranges
- Distributions
- Multiple outcomes
- Obscurity

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

77

Visualizing uncertainty with ranges

Specify what the range represents

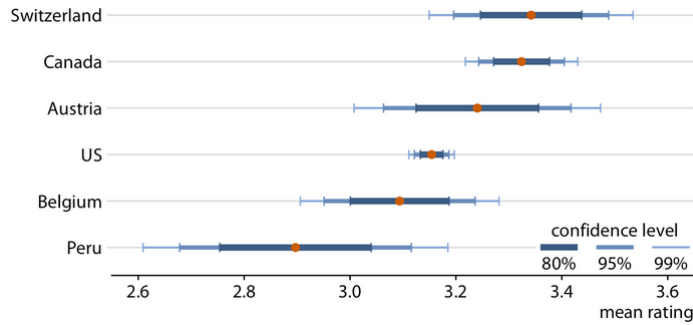


<https://serialmentor.com/dataviz/visualizing-uncertainty.html>

78

Visualizing uncertainty with ranges

Specify what the range represents



<https://serialmentor.com/dataviz/visualizing-uncertainty.html>

79

Visualizing uncertainty with ranges

Luka Doncic

DALLAS MAVERICKS
SHOOTING GUARD
22 YEARS OLD

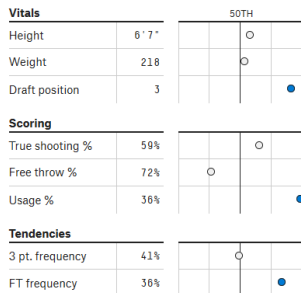


WEIGHTED AVERAGE OF PAST THREE SEASONS

● BAD ○ AVG ● GOOD

PERCENTILE

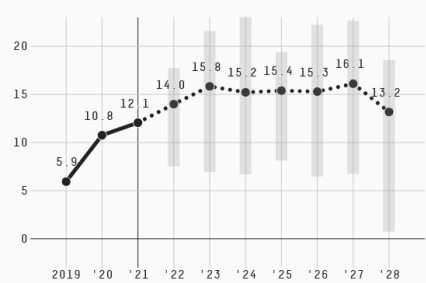
50TH



Wins above replacement projection

CATEGORY: MVP CANDIDATE
5-YR MARKET VALUE: \$379.7M

90TH— CONFIDENCE INTERVAL
10TH—
..... PROJECTION



Performance of the 10 most comparable players



<https://projects.fivethirtyeight.com/2022-nba-player-projections/luka-doncic/>

80

Visualizing uncertainty with ranges

Earnings per share outlook

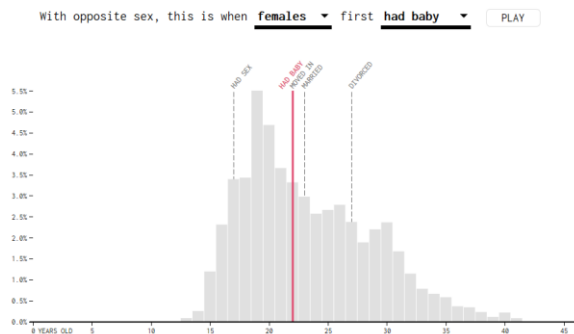


<https://www.storytellingwithdata.com/blog/2018/6/27/visualizing-uncertainty>

81

Visualizing uncertainty with distributions

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

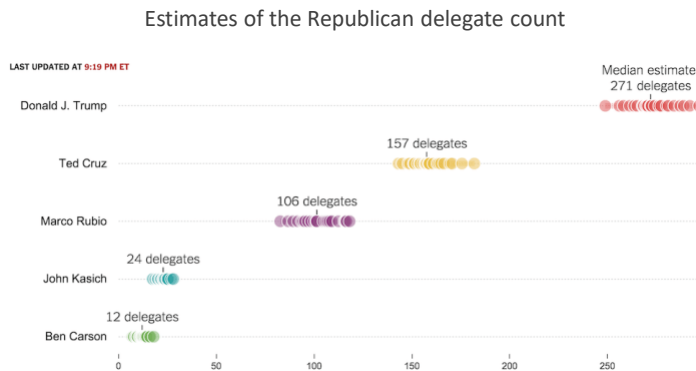


<https://flowingdata.com/2017/02/23/the-first-time/>

82

Visualizing uncertainty with multiple outcomes

Show the various outcomes

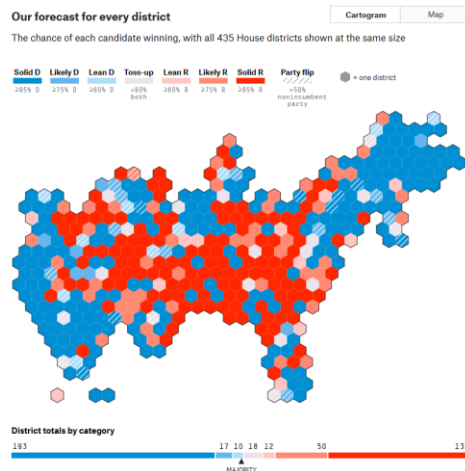


<https://www.nytimes.com/interactive/2016/03/01/upshot/super-tuesday-live-republican-delegate-estimates.html?mtrref=undefined>

83

Visualizing uncertainty with obscurity

Use transparency, color scale, or blurriness to show uncertainty

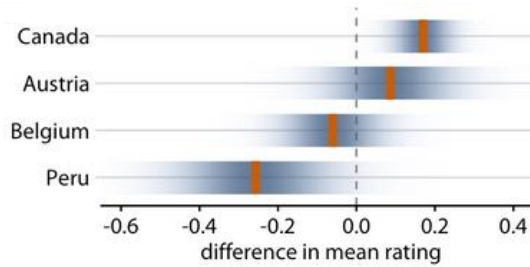


<https://projects.fivethirtyeight.com/2018-midterm-election-forecast/house/>

84

Visualizing uncertainty with obscurity

Use transparency,
color scale, or
blurriness to show
uncertainty



<https://serialmentor.com/dataviz/visualizing-uncertainty.html>

85

Visualizing missing data

Techniques to handle missing data

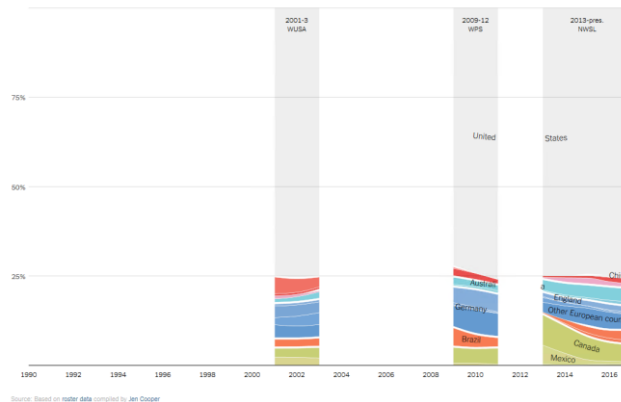
- Collect the data
- Show only what you have
- Show the gaps
- Treat it as a category

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

86

Visualizing missing data by showing the gaps

Where players in **U.S. Women's Soccer** have come from



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

87

Visualizing missing data as a category

Use white or neutral color to show the 'missing data category'

SMOKING RATES FOR DIFFERENT GROUPS

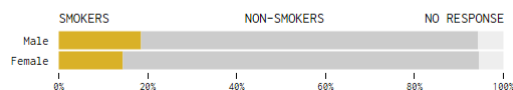
SELECT YEAR

1994

2014

SEX

A higher percentage of males smoke than females do. The difference between the two increased during this time period.



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

88

Interactivity

89

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)

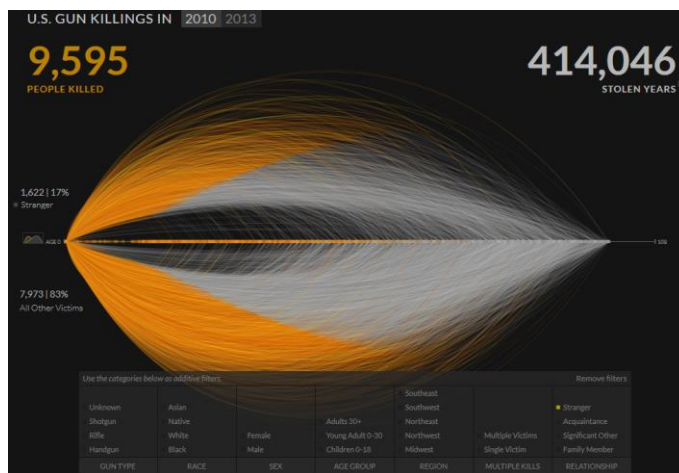
Data adjustments

- **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

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

91

Framing



<https://guns.periscope.com/?year=2010>

92

Navigating



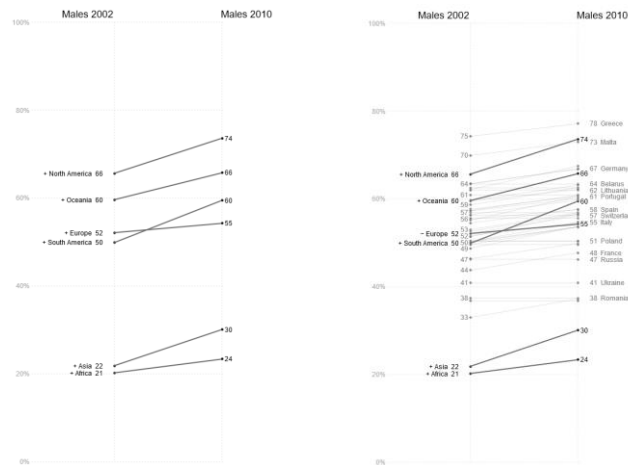
Road orientation map

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

93

Navigating

Obesity
around
the world



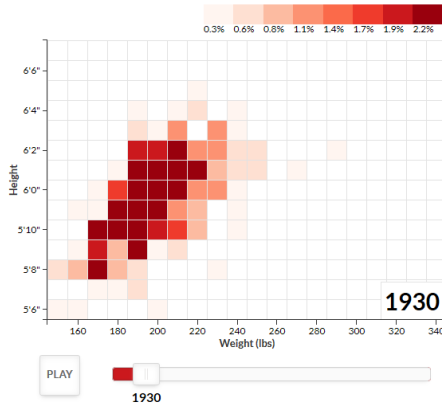
<http://neoformix.com/Projects/ObesitySlope/>

94

Animating

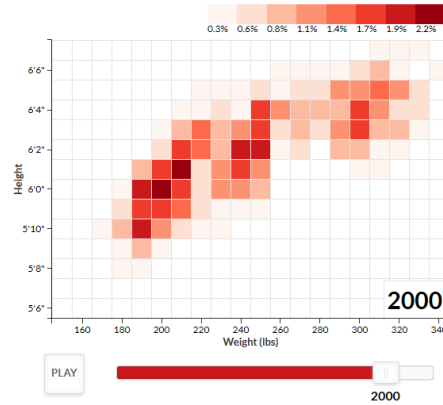
NFL player: height & weight over time

By Noah Veltman



NFL player: height & weight over time

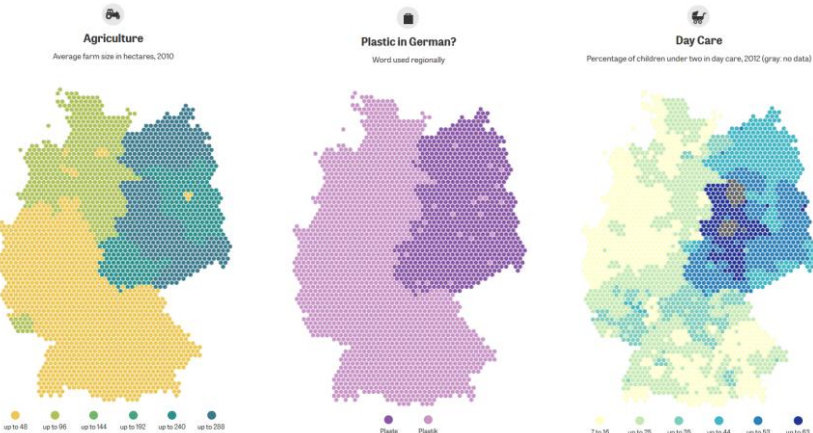
By Noah Veltman



<https://noahveltman.com/nflplayers/>

95

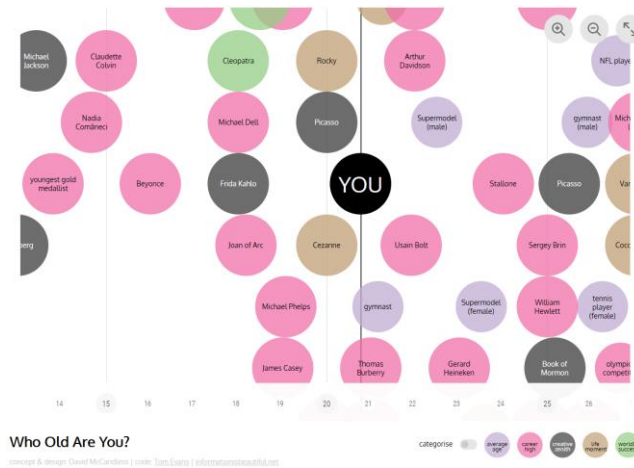
Sequencing



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

96

Contributing



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

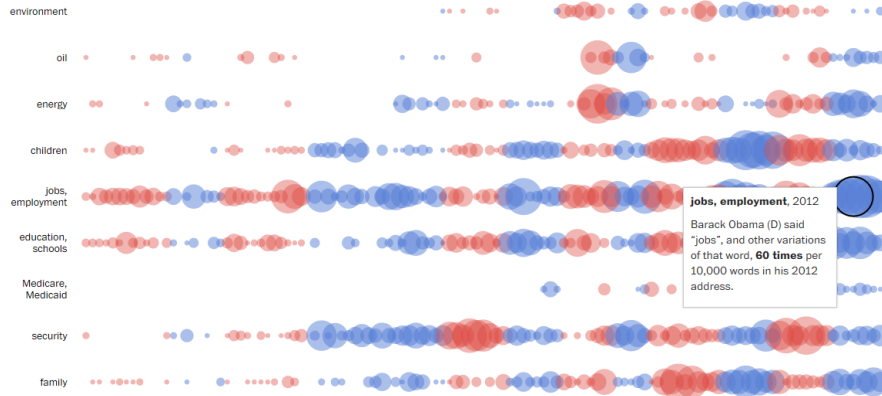
97

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

Annotating

History through the president's words



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

101

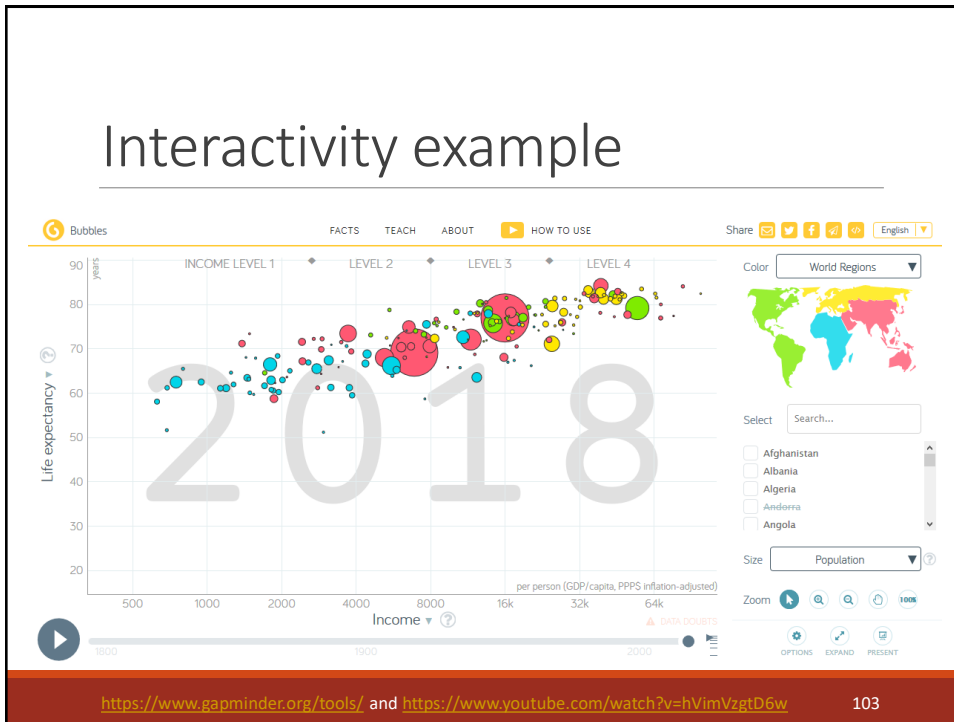
Orientating



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

102

Interactivity example



Animation example



Storytelling

10
5

Storytelling

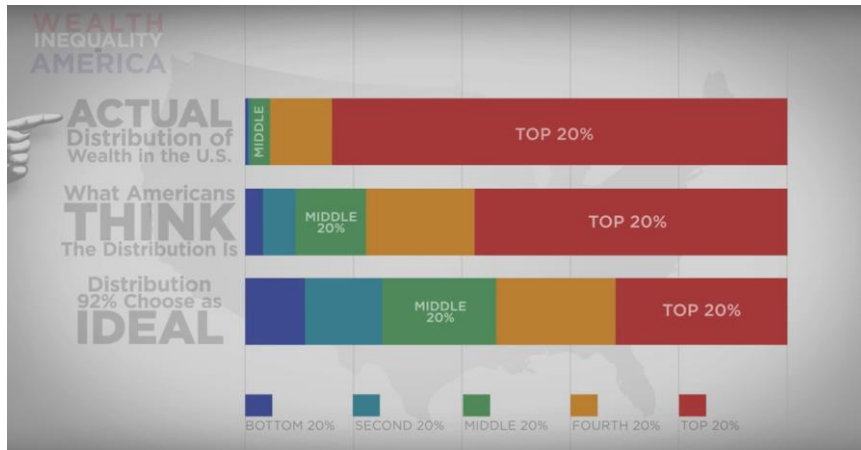
Storytelling ≠ making something up

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=QPKKQnjimsM>

107

Enforce emotions

The screenshot shows a Medium article titled "Coronavirus Lost Loved Ones". On the left is a portrait of Maria Caceres Baranyai. To the right is a map of the United States with several red heart icons indicating locations. Below the portrait, the text reads: "Date Passed: 3/29/2020, 1:00 AM". A paragraph follows: "Maria was a loving and caring lady, devoted to her family. She was a resident of Kenner, LA for the last 47 years. She was a retired Nursing Assistant having worked for over 20 years at East Jefferson General Hospital. She is

Emotions can be enforced by providing stories about individual experiences

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

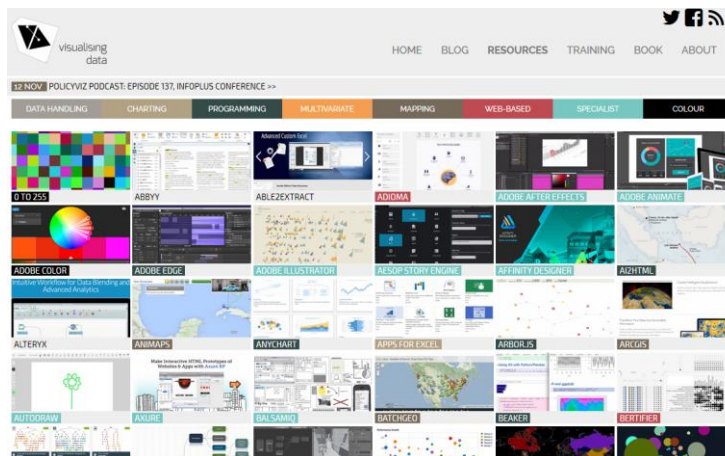
Best: do both

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

108

Tools

Tools



D3

Data-Driven Documents



A JavaScript library

Emphasis on interactivity

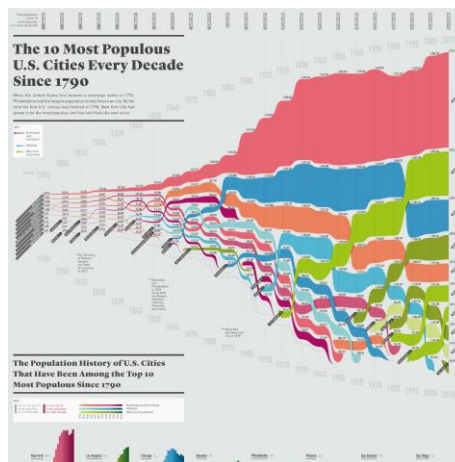
<https://d3js.org/>

111

RAW Graphs

Open source tool for data visualization

Politecnico di Milano



<https://www.rawgraphs.io/>

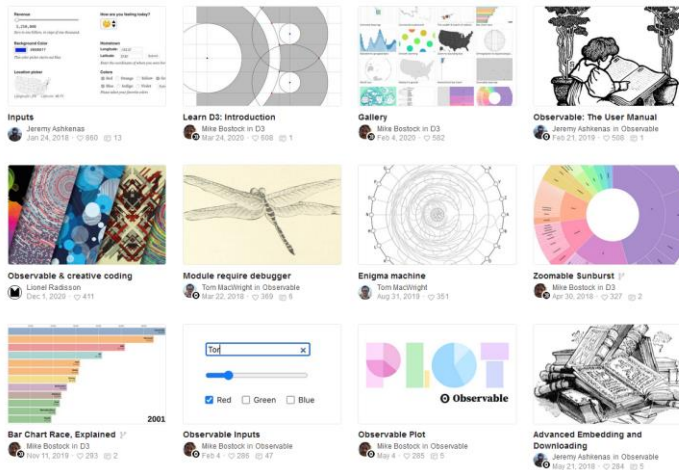
112

Observable

Interactive notebooks supporting collaboration

JavaScript

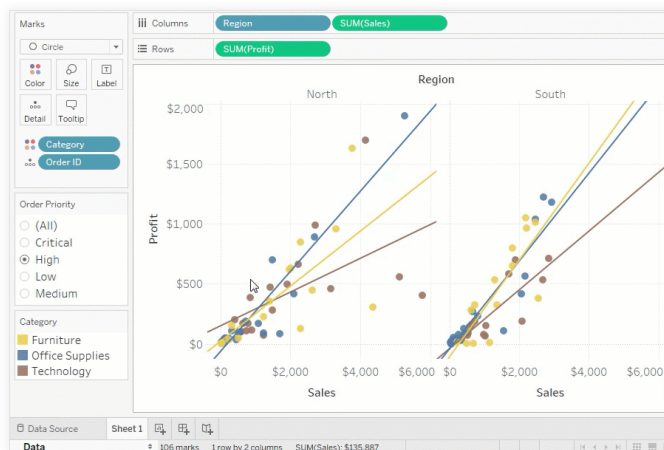
Free plan



<https://observablehq.com/>

113

Tableau Public



Does not require programming skills

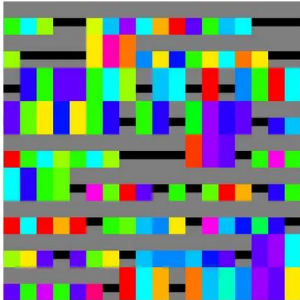
Visualizations created with the free version are public

Students can get a free version with private use

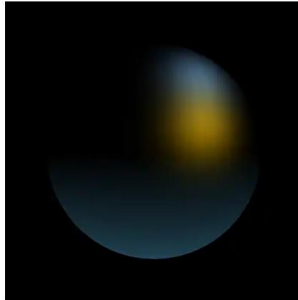
<https://public.tableau.com/s/>

114

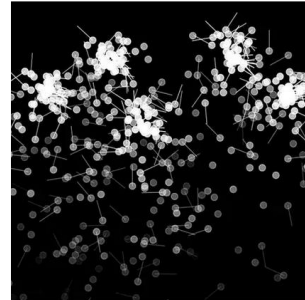
Processing



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>

115