

Data Visualization

VISUALIZATION DESIGN

Overview

The 7 steps of visualization design

Basic charts

Multivariate/multidimensional data visualization

Visualizing uncertainty and missing data

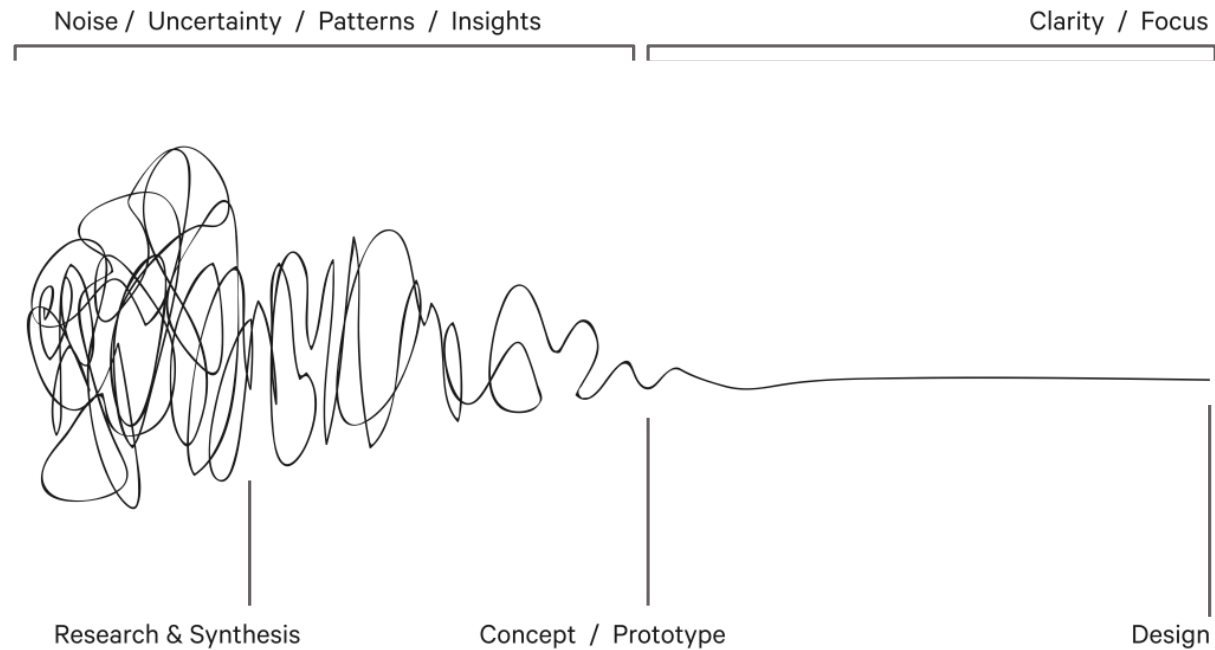
Interactivity

Storytelling

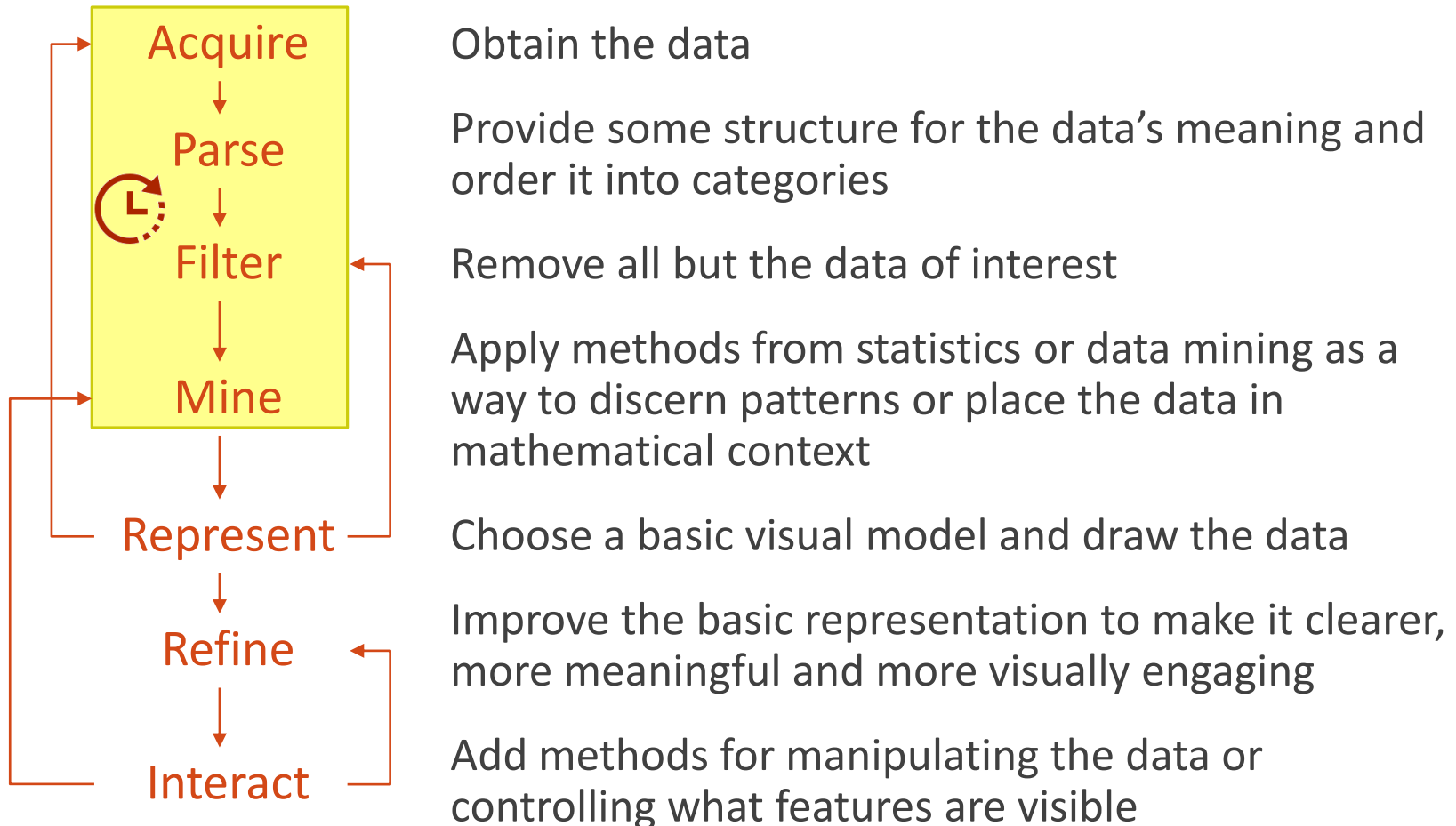
Tools

The 7 steps of visualization design

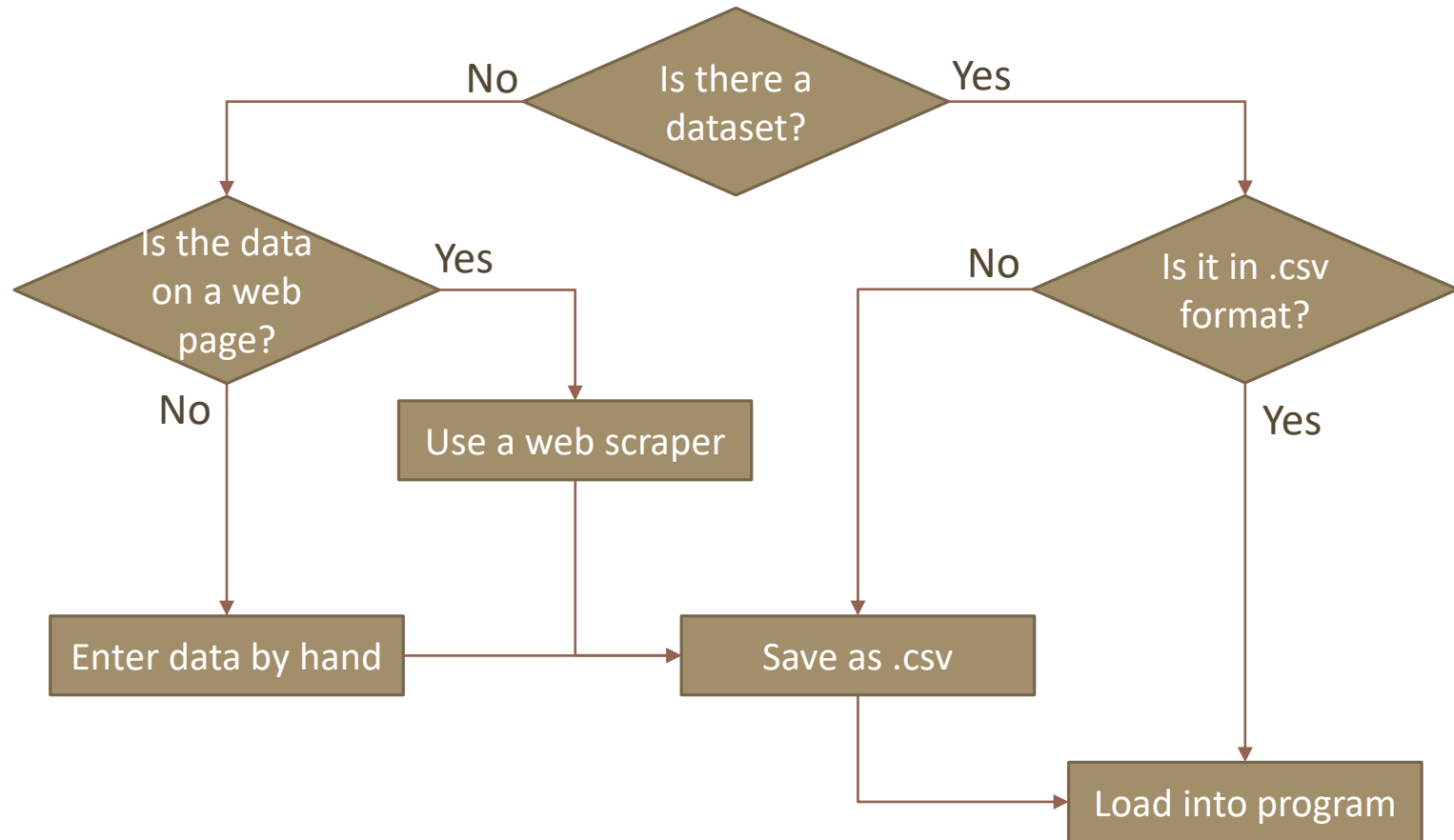
Visualization design



The 7 steps of visualization design



Acquire the data



Parse the data

Check for errors

Change type

- For example, ordinal to categorical

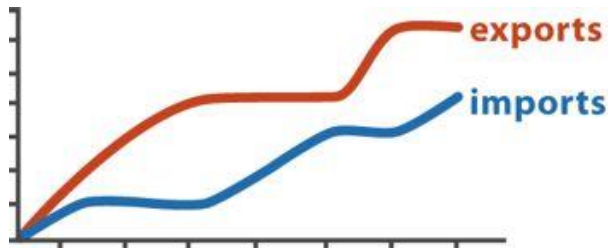
Choose the level for hierarchical data

- Temporal data: day of the week, day of the month, ...

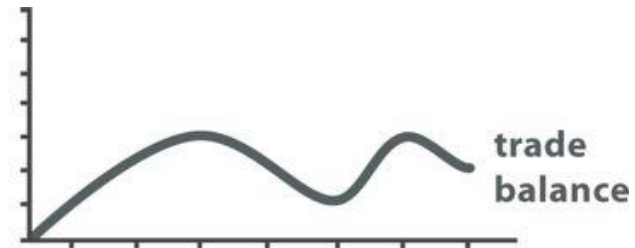
Parse the data

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



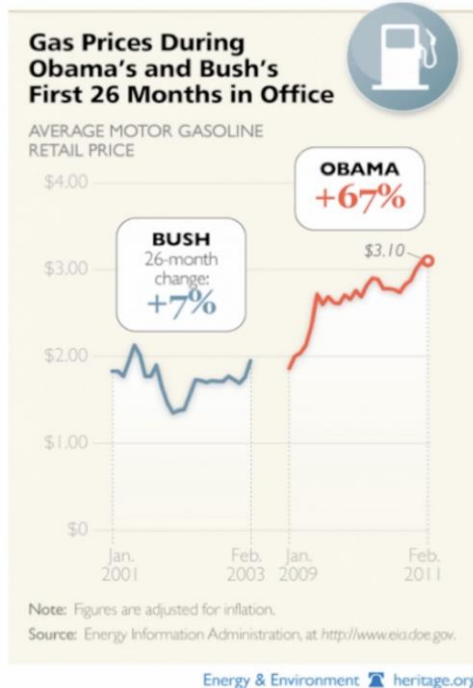
$$\text{trade balance} = \text{exports} - \text{imports}$$

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**

Attribute Types

→ Categorical

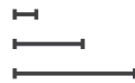


→ Ordered

→ Ordinal



→ Quantitative



Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



Actions

Analyze

→ Consume

→ Discover

→ Present

→ Enjoy



→ Produce

→ Annotate

→ Record

→ Derive



Search

	Target known	Target unknown
Location known	••• Lookup	••• Browse
Location unknown	<•••> Locate	<•••> Explore

Query

→ Identify

→ Compare

→ Summarize



Targets

All Data

→ Trends



→ Outliers



→ Features

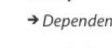


Attributes

→ One

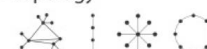


→ Many



Network Data

→ Topology



→ Paths



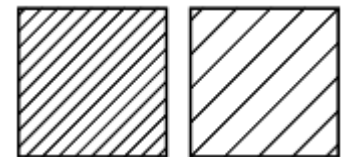
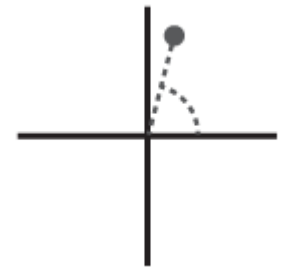
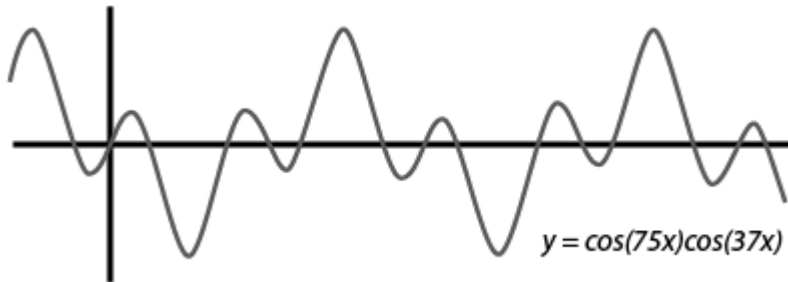
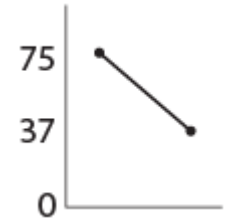
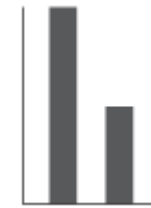
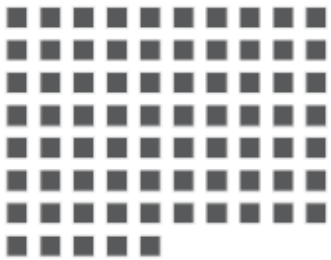
Spatial Data

→ Shape

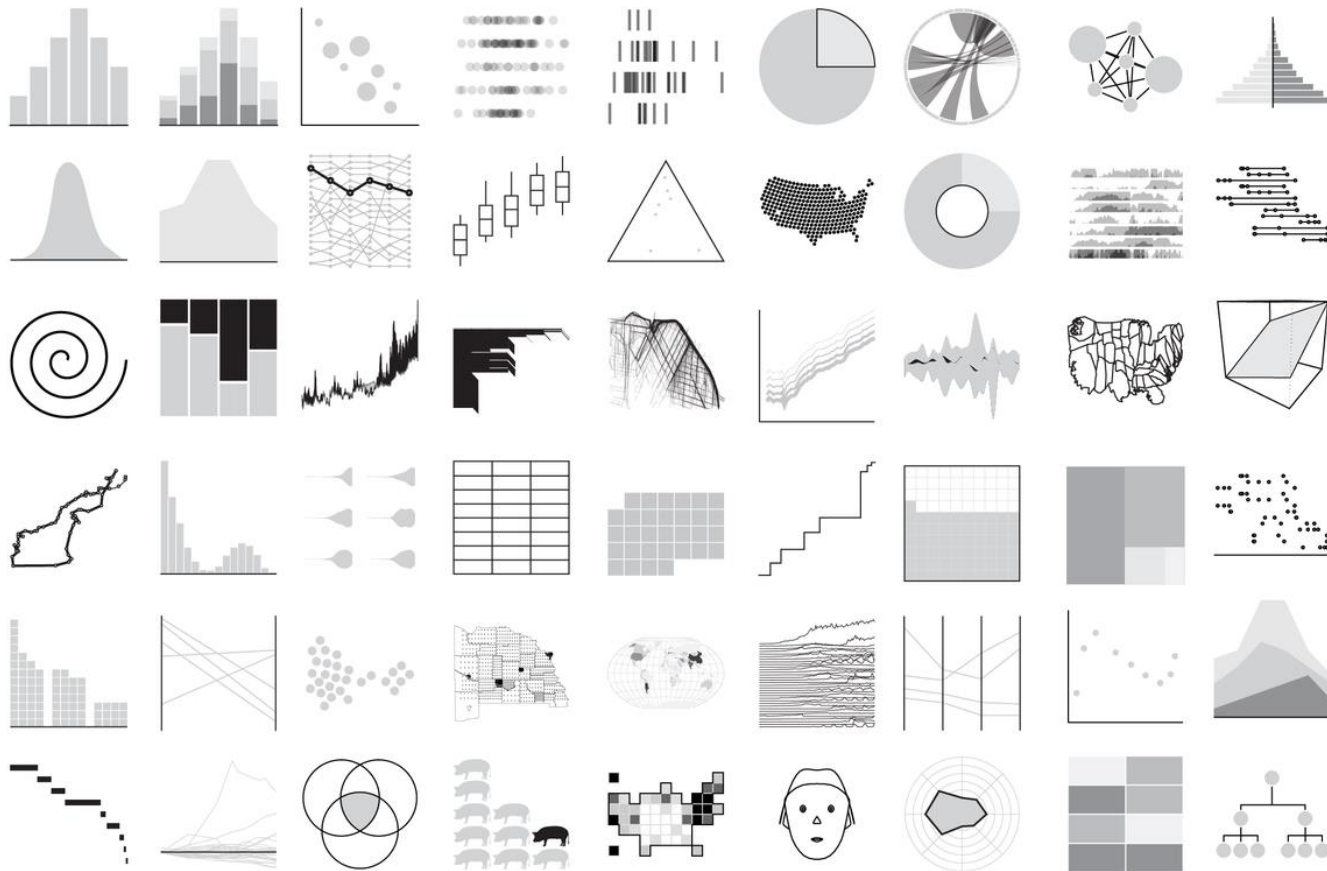


Represent the data

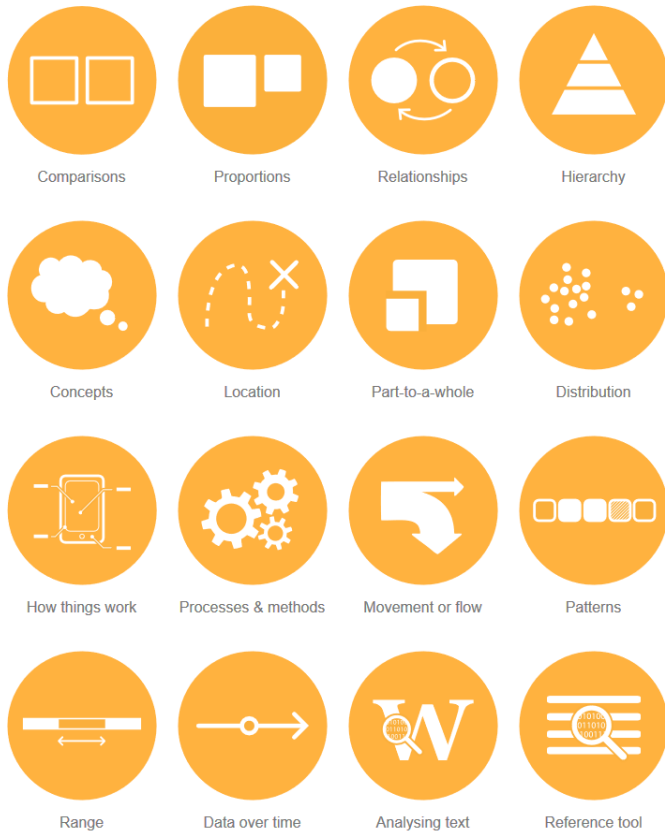
45 ways to communicate two quantities (75 and 37)



Represent the data



Represent the data



Represent the data

150 types of charts and visualizations. A comprehensive glossary of data visualization techniques. The glossary is organized into 15 categories, each with a set of 10 chart types. Each category includes a brief description of the category, a list of 10 chart types with their names and a small icon representing the chart type, and a detailed description of each chart type. The categories are: Deviation, Correlation, Ranking, Distribution, Change over Time, Magnitude, Part-to-whole, Spatial, and Flow. The glossary is designed to be a quick reference guide for data visualization techniques.

Category	Chart Type	Description
Deviation	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Scatter plot	Used to show the relationship between two variables, with each point representing a single observation.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
	Cluster	Used for showing the relationships between different clusters of data points, with lines representing the connections.
Correlation	Scatter plot	The standard way to show the relationship between two variables, with a regression line showing the trend.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
Ranking	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
Distribution	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
Change over Time	Line	Tracks a single value over time with a line connecting the data points.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
	Cluster	Used for showing the relationships between different clusters of data points, with lines representing the connections.
	Scatter plot	Used to show the relationship between two variables, with each point representing a single observation.
Magnitude	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Heatmap	Used for showing the intensity of a value across a grid of categories, with color representing the intensity.
	Network	Used for showing the relationships between different nodes in a network, with lines representing the connections.
Part-to-whole	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
Spatial	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
Flow	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.
	Area	Shows a range of values over time with a shaded area representing the range of values.
	Line	Tracks a single value over time with a line connecting the data points.
	Stacked bar	A simple stacked bar chart that can be used to compare two categories across multiple sub-categories.
	Grouped bar	Two or more vertical bars representing the same category across different sub-categories.
	Bar chart	Shows the relative sizes of different categories, with the height of the bars representing the values.
	Waterfall	Shows changes in a total value over time, with each bar representing a change in a specific component.

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphics team. Thanks to the many contributors who shared their expertise and data.



Copyright © 2015 FT. All rights reserved. FT is a registered trademark of FT. This content is for personal use only. For more information, visit ft.com/vocabulary

Represent the data

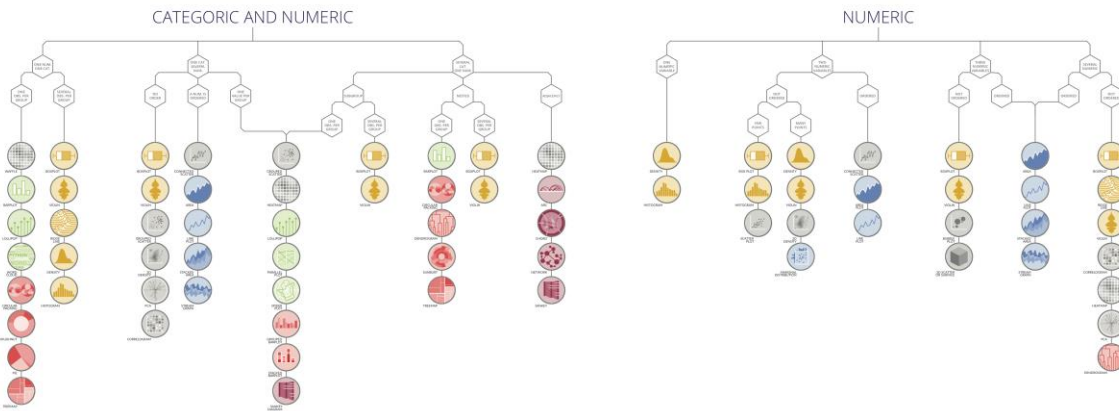
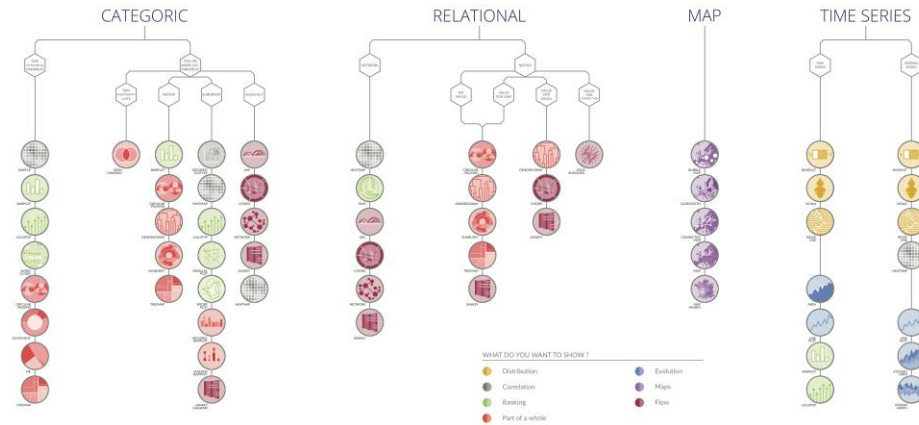


From Data to Viz is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps:

- 1 Identify what type of data you have.
- 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
- 3 Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more visit:

data-to-viz.com

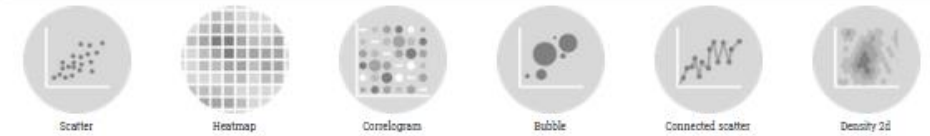


Represent the data

Distribution



Correlation



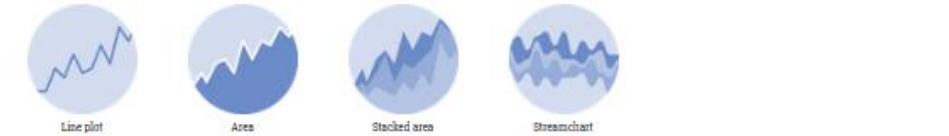
Ranking



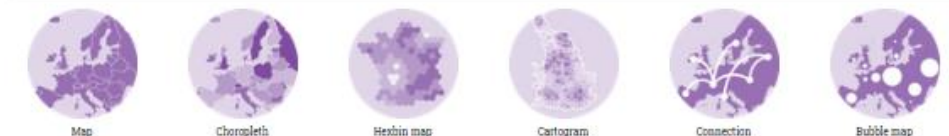
Part of a whole



Evolution



Map



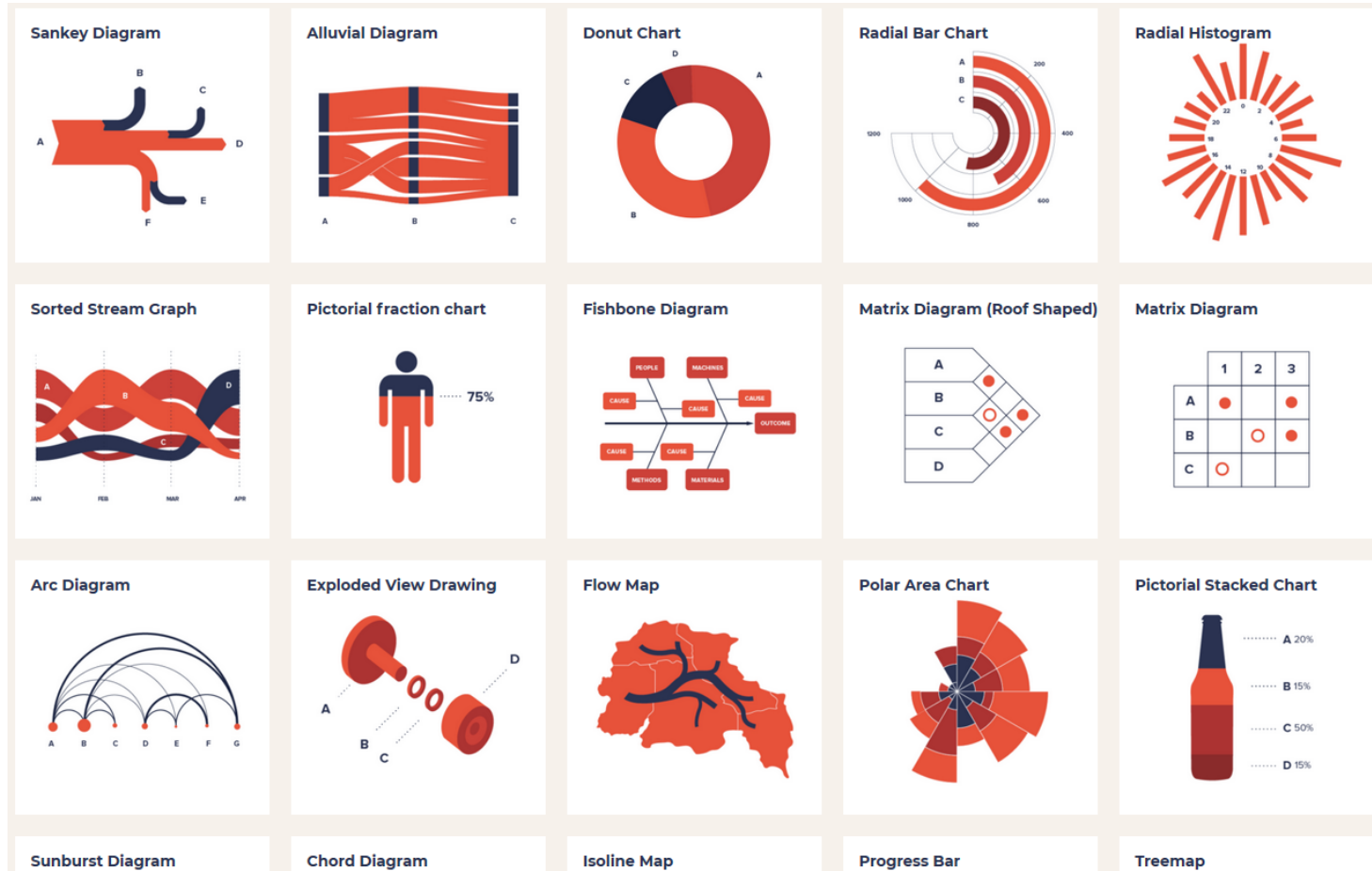
Flow



General knowledge








Represent the data






Represent the data

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

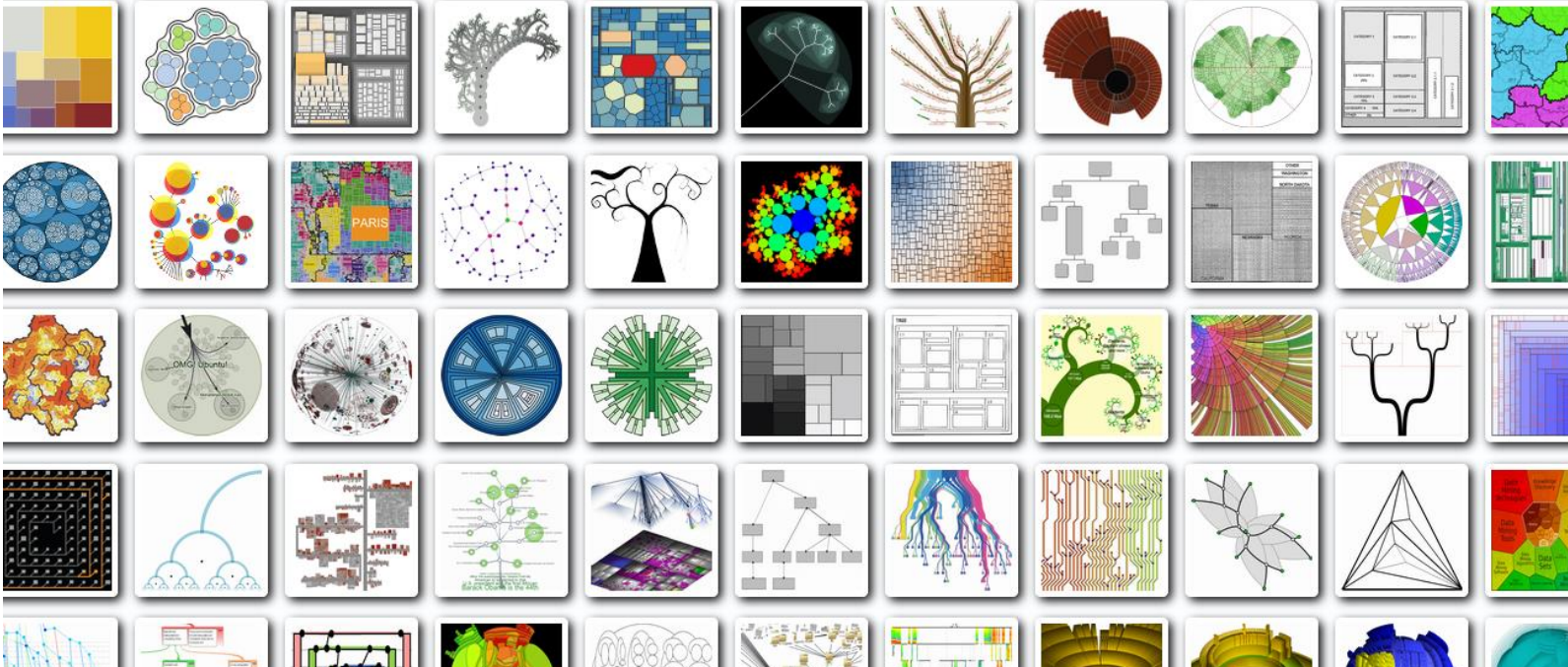
Dimensionality: All   

Representation: All   

Alignment: All   

Fulltext Search: x

Techniques Shown: 306



Represent the data

The TimeViz Browser
A Visual Survey of Visualization Techniques for Time-Oriented Data
by Christian Tominski and Wolfgang Aigner

of Techniques: 115

Search:

How to use filters:

- Want:** Show me!
- Indifferent:** I don't care.
- Hide:** I'm not interested!

Data

Frame of Reference

- Abstract
- Spatial

Number of Variables

- Univariate
- Multivariate

Time

Arrangement

- Linear
- Cyclic

Time Primitives

- Instant
- Interval

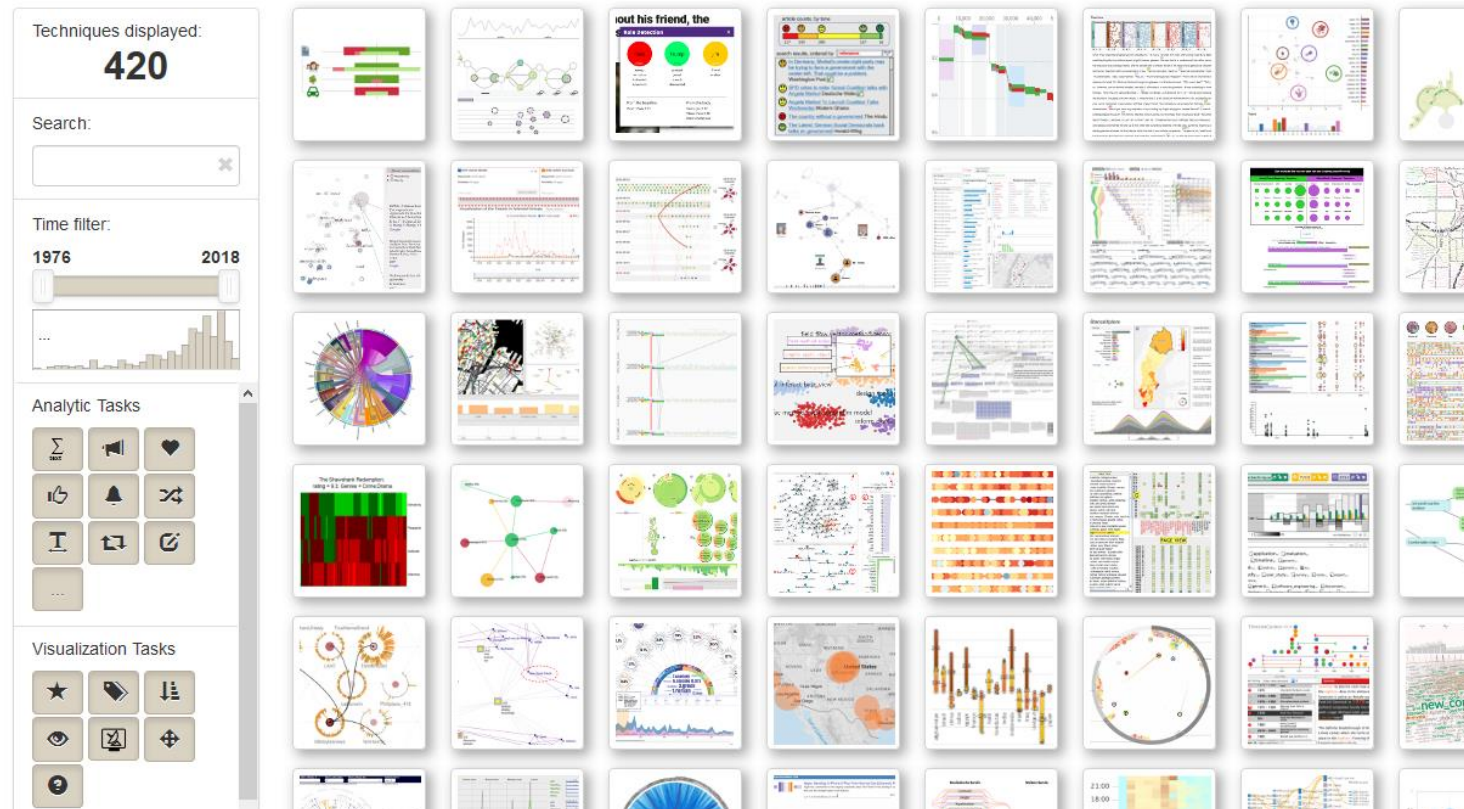


Represent the data

Text Visualization Browser

A Visual Survey of Text Visualization Techniques (IEEE PacificVis 2015 short paper)

Provided by ISOVIS group



Represent the data

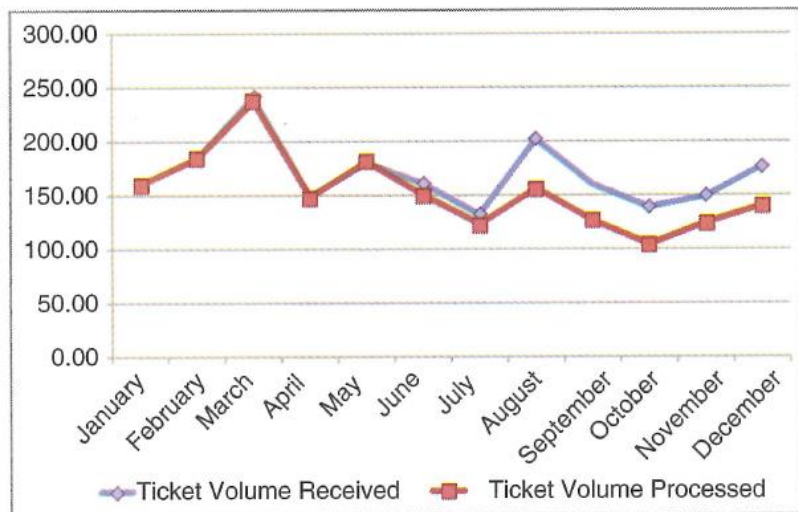
Filter by chart name or AKA

Reference Type:  Example  Solution | Chart Families:  Categorical  Hierarchical  Relational  Temporal  Spatial

	Amazon QuickSight	ArcGIS	ChartJS	Charticulator	D3.js	Data Illustrator	Datawrapper	Flourish	FusionCharts	Gephi	Google Charts	Google Data Studio	Highcharts	Infogram	JetPack Data	JMP
Bar chart	●			●	●●●	○	●●●	○	○		●●	○	●●	○	●○	●
Clustered bar chart	●				●	○	●●●	○	○		●●				○	●
Bullet chart				●	●		●●		○							●
Waterfall chart				●	●				○		●		○	○		
Radar chart			○		●				○				○			
Polar chart			●	●	●								○			
Connected dot plot					●●	○	●●	●								
Pictogram					○									○		
Proportional shape chart					●●●	○		○	○		●					
Word cloud					●									○	●	
Heat map	●			●	○	○			○				○			●

Refine the visualization

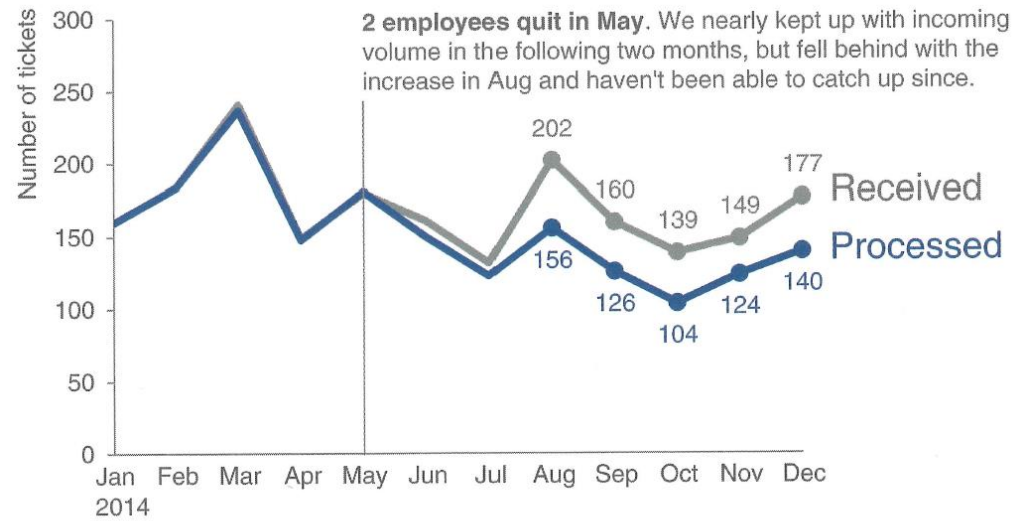
Improve the basic representation to make it clearer, more meaningful and more visually engaging



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.

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
- (Hexagon) tile 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

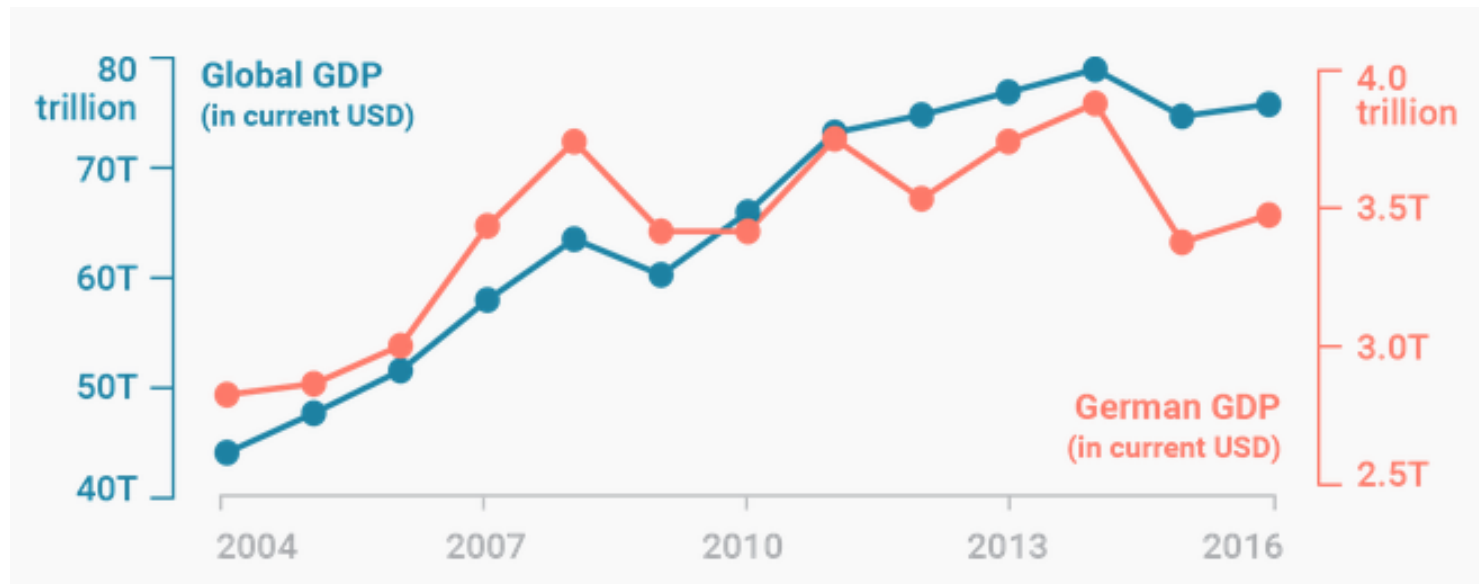
Place the labels close to the data

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

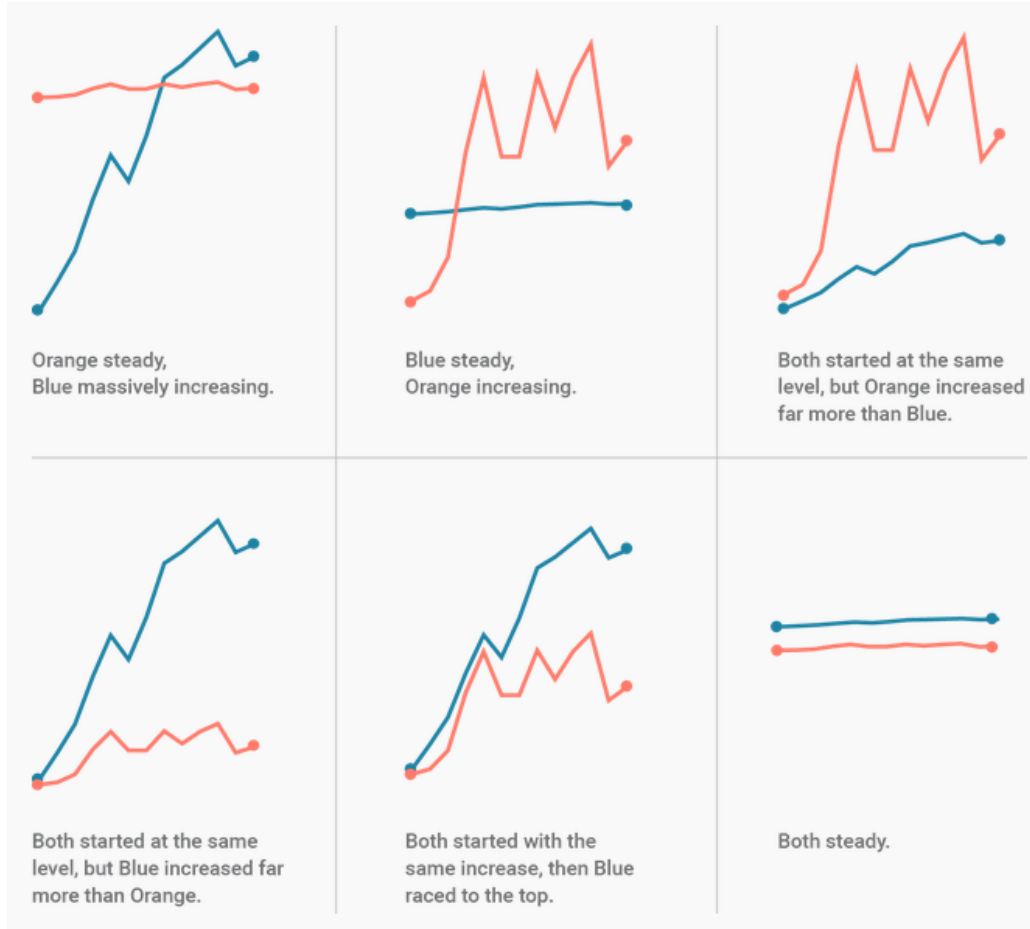
- If the data comes close to 0
- If 0 has a meaning



Line charts: Avoid dual axis charts



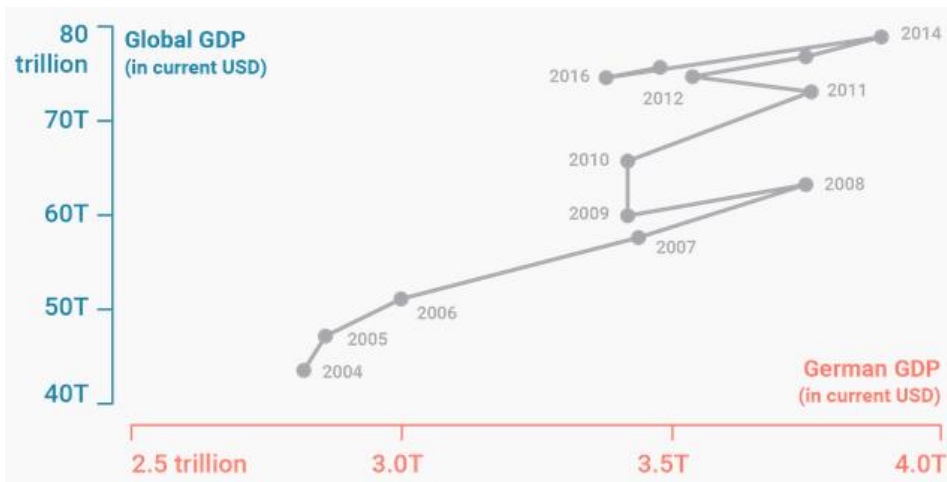
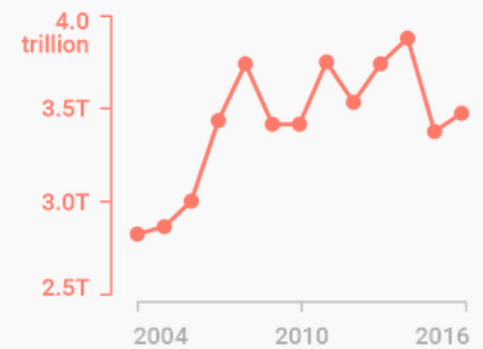
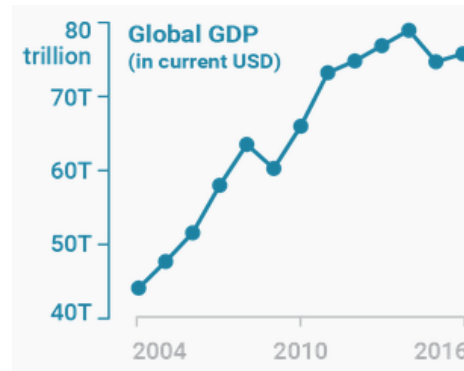
Line charts: Avoid dual axis charts



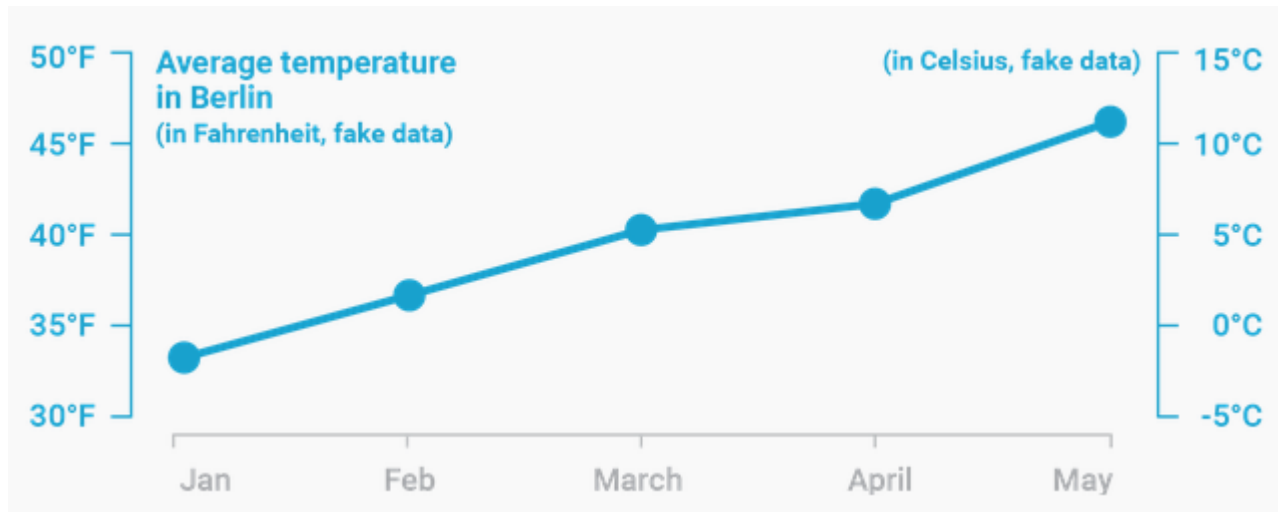
Line charts: Avoid dual axis charts

Alternatives

- Side-by-side charts
- Connected scatter plots



Line charts: Dual axes sometimes okay



Bar charts

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

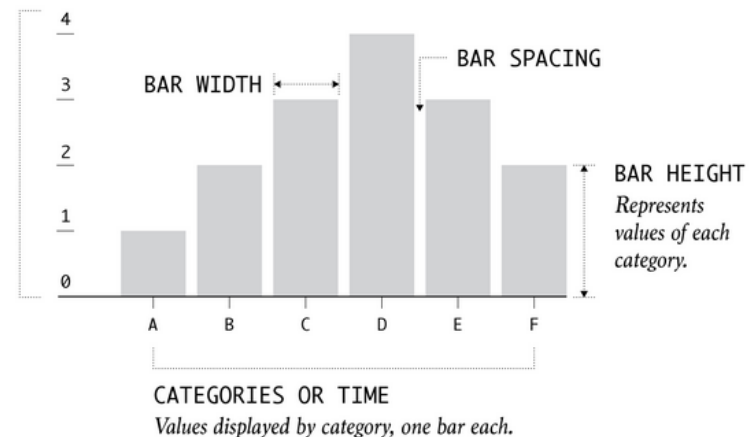
They should always have a 0 baseline

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

VALUE AXIS
Indicates scale of the graph with values starting at zero.

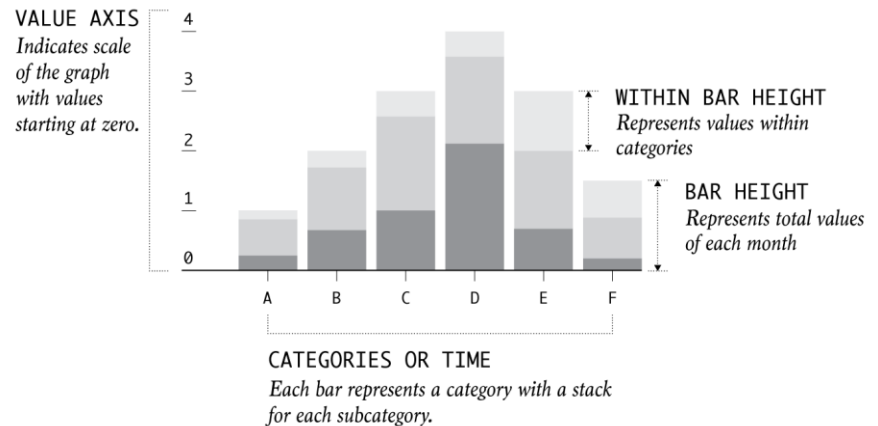


Stacked bar charts

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

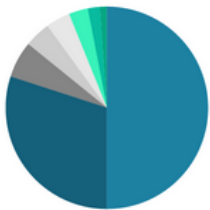
- Group smaller slices together as 'other'
- Label small slices outside of the chart

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

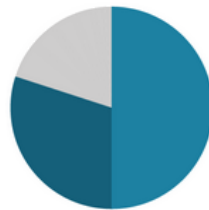
If the slices are of similar size, use a bar chart instead

No 3-D

NOT IDEAL



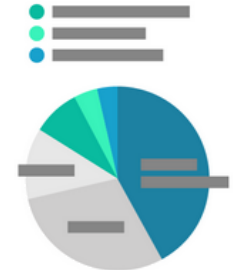
BETTER



NOT IDEAL



BETTER



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*

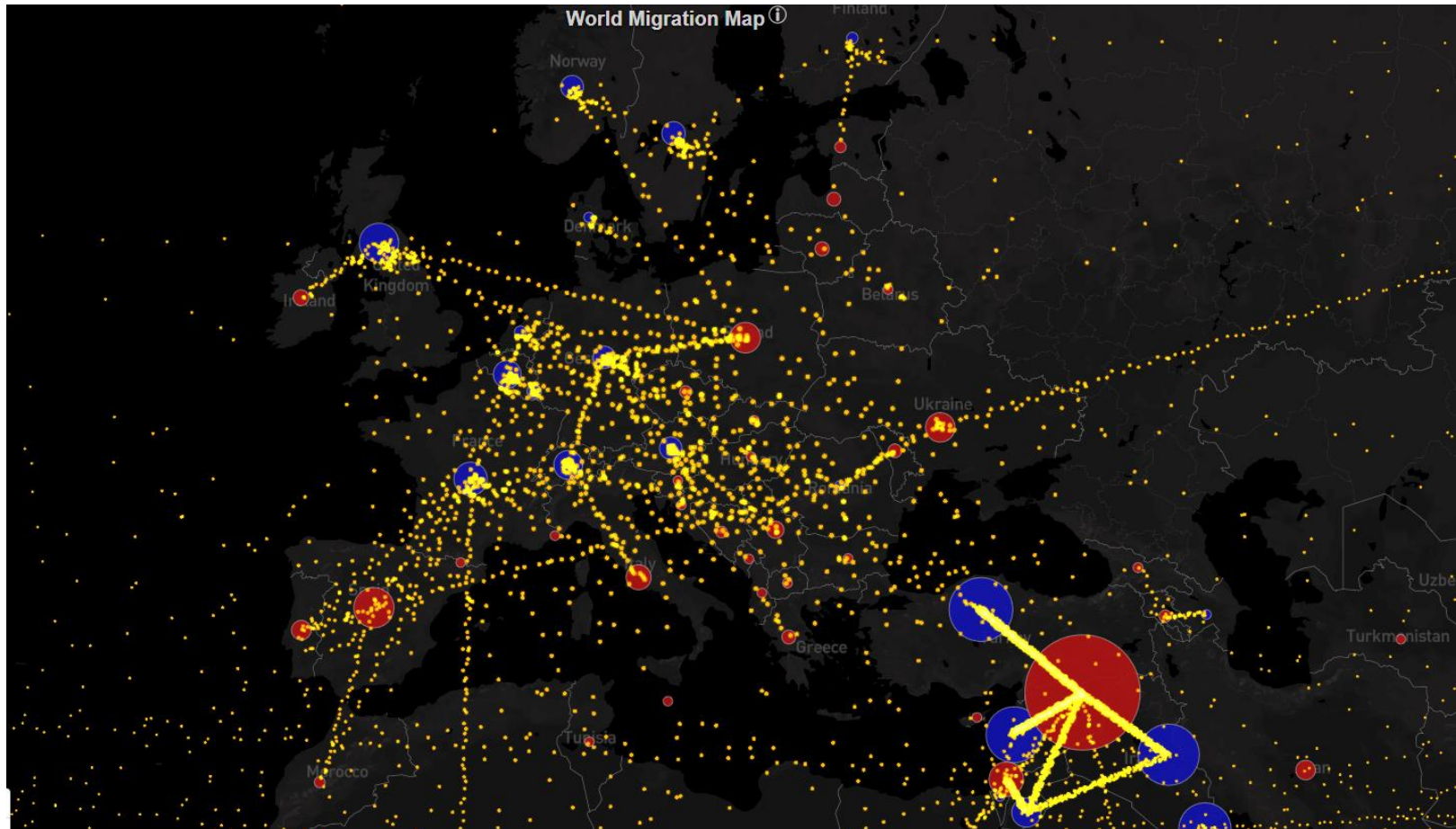
Use them to show how things are distributed over a geographical region

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



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

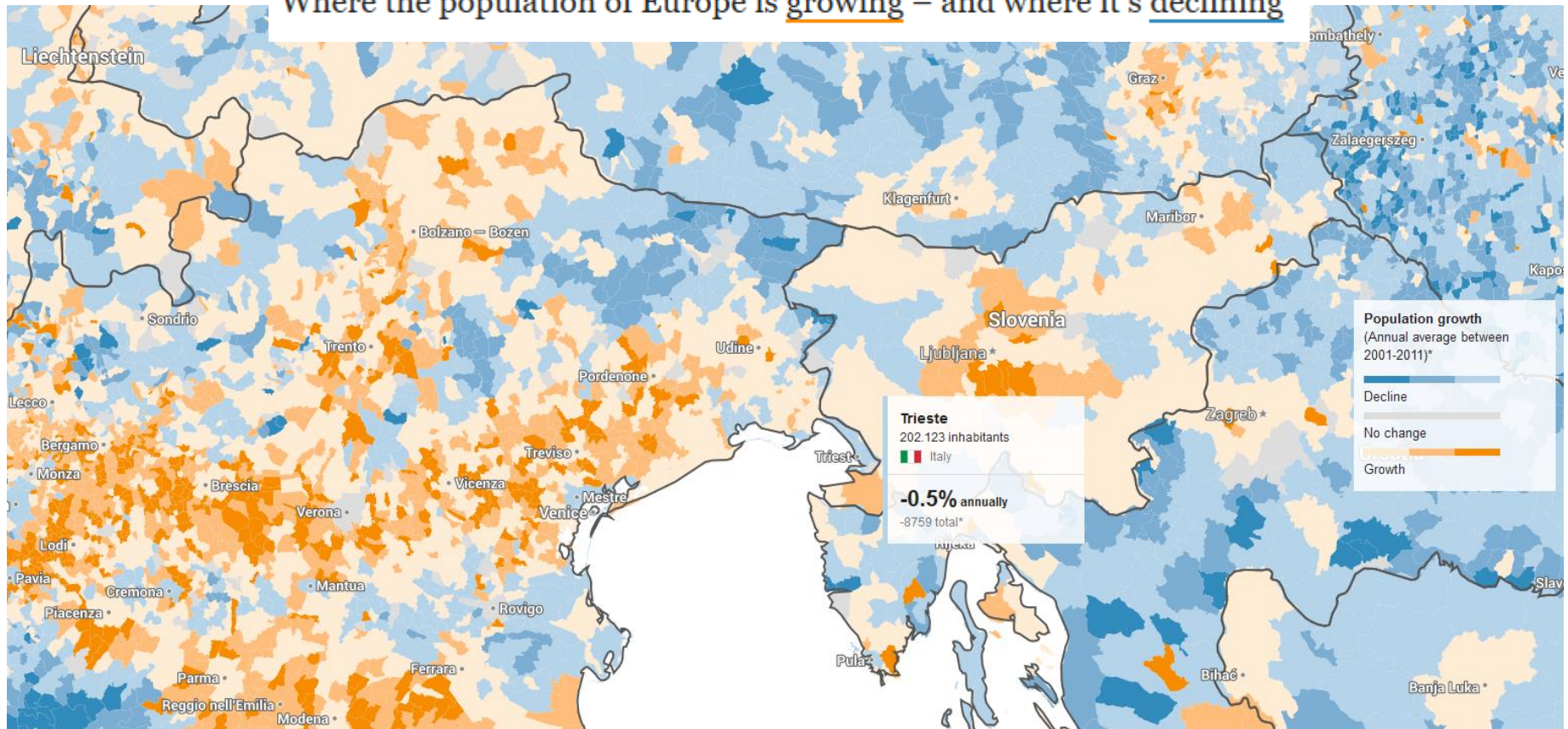
Show relative instead of absolute data

Be careful in choosing bin size

Be careful in choosing colors

Choropleth maps

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



(Hexagon) tile maps

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

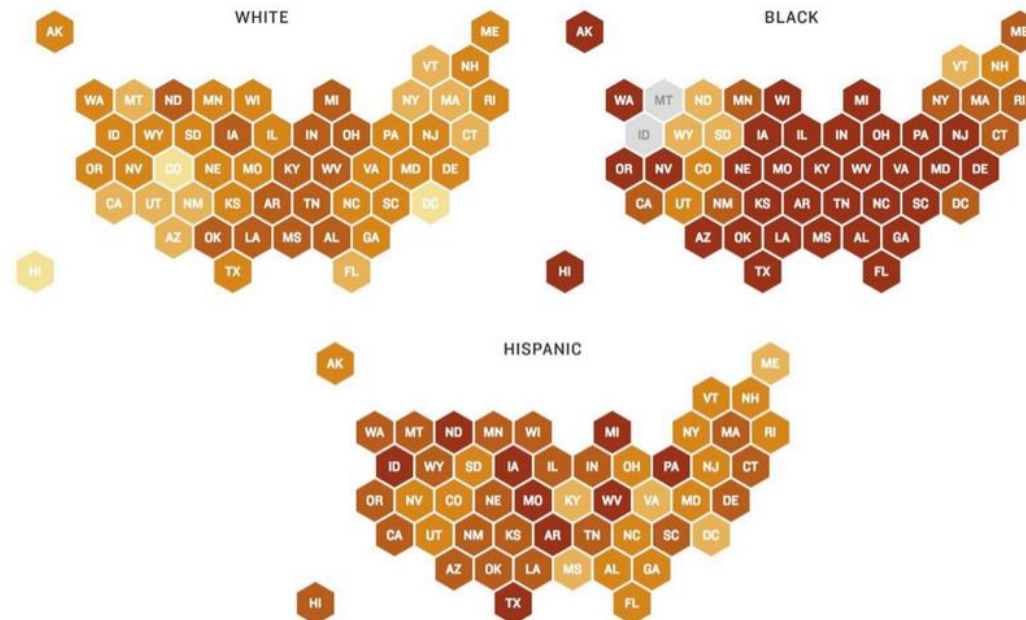
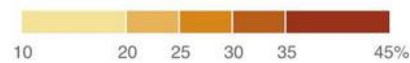
Tile represents a state/province

Often hexagonal or square

Harder to locate the given state/province

(Hexagon) tile maps

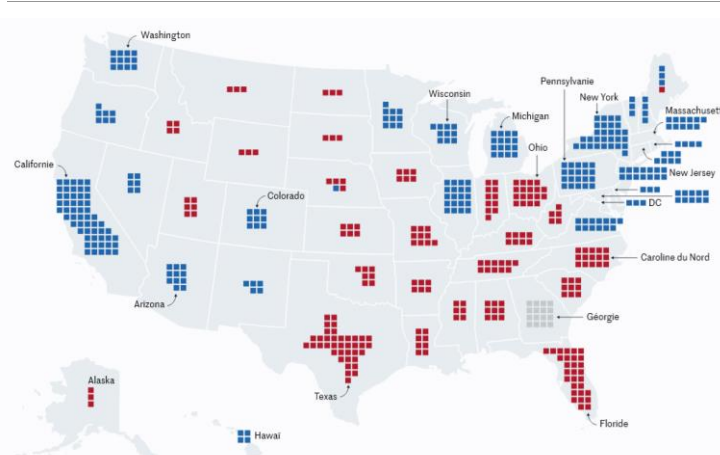
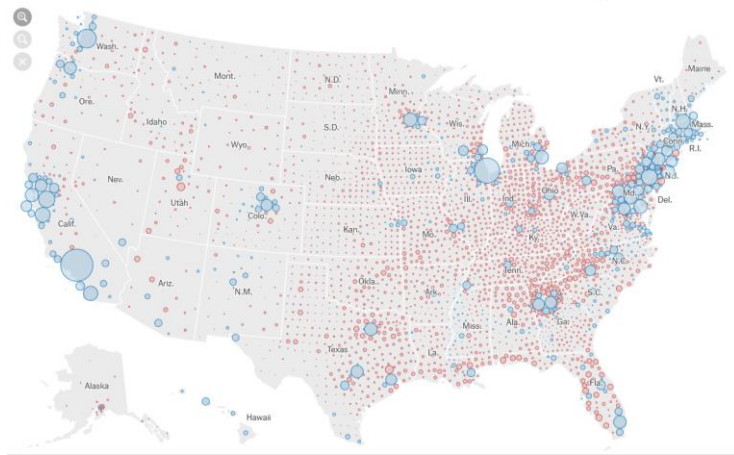
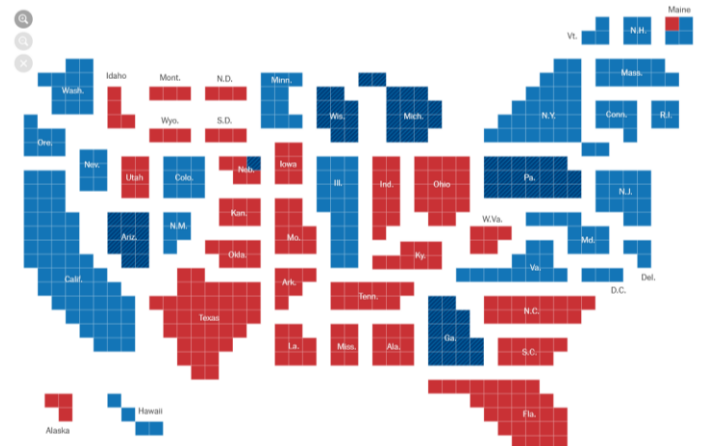
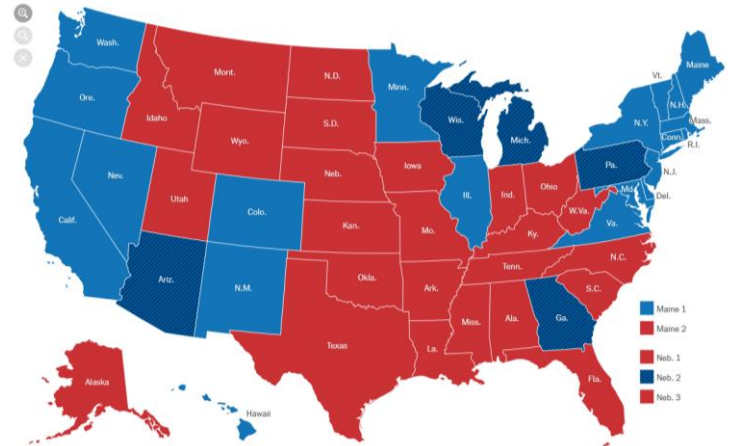
Obesity Prevalance In 2014, By Race



Notes

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

Comparisons of maps

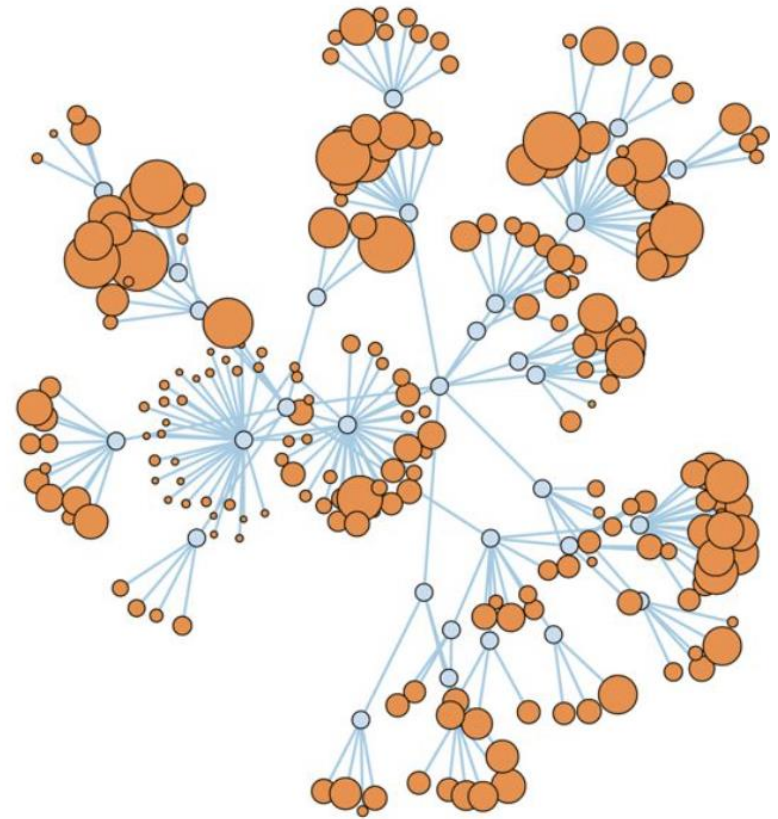
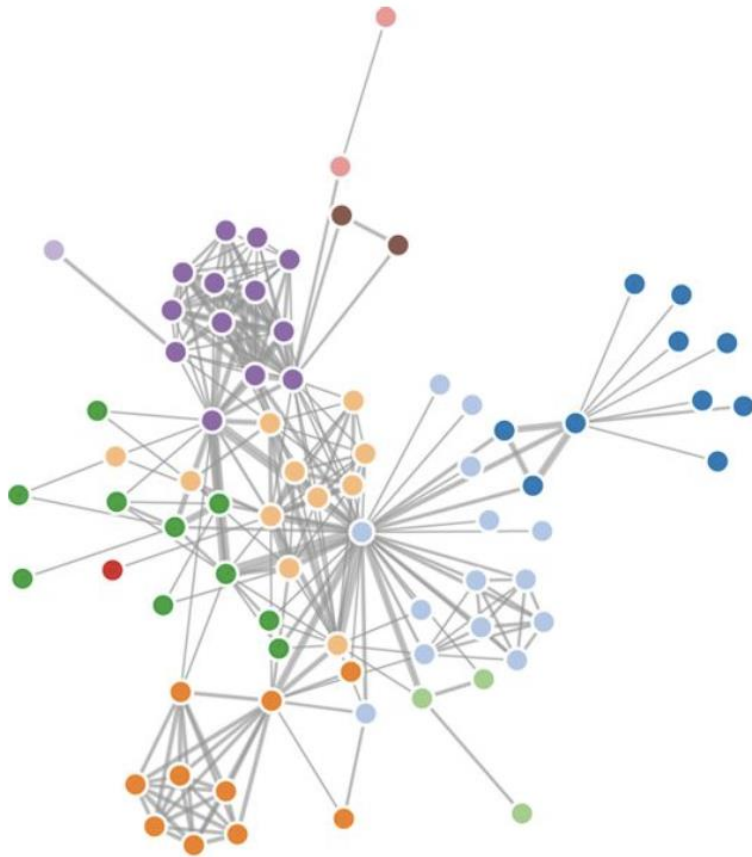


Networks and trees

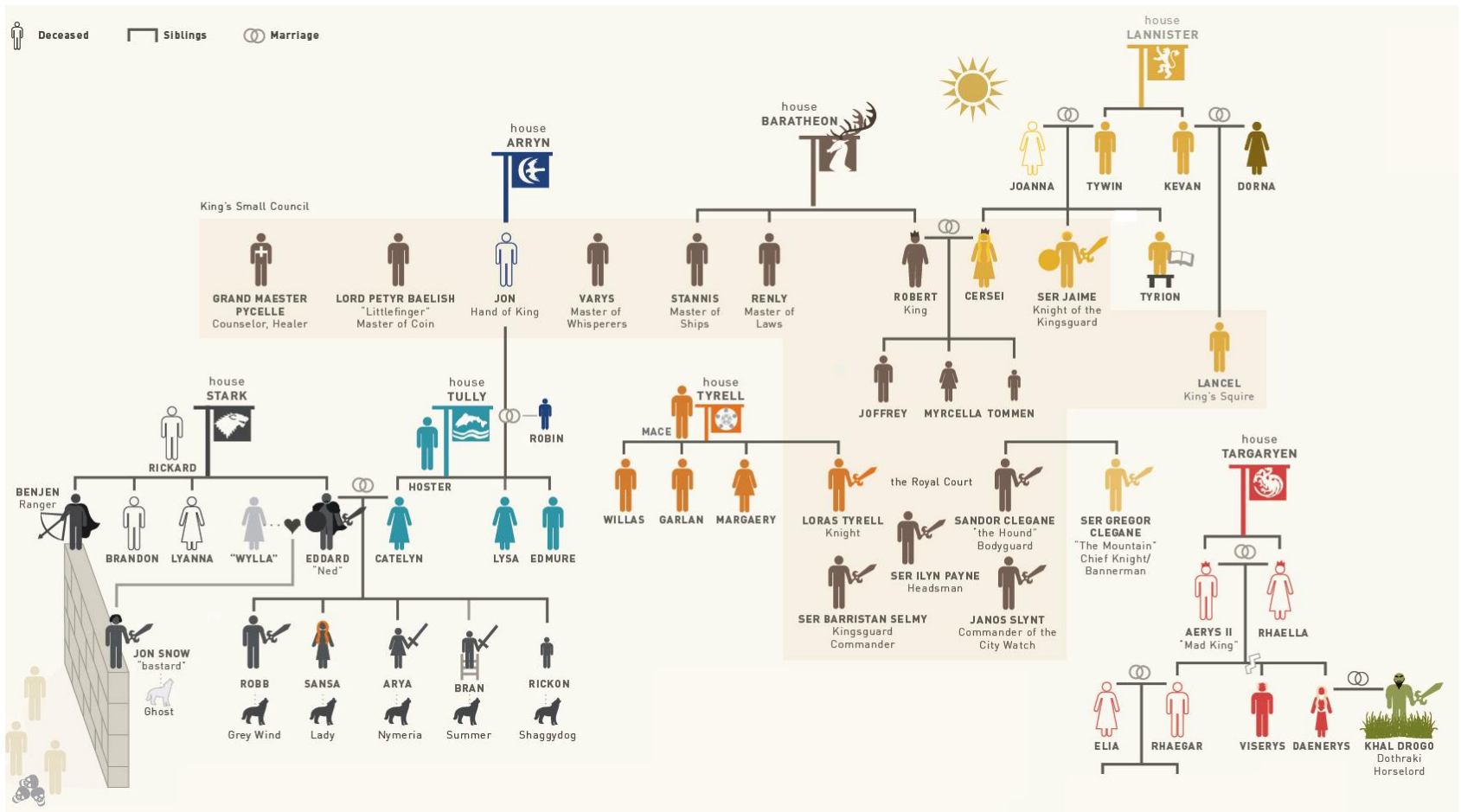
Network and trees are relational structures characterized by a collection of nodes and links that connect the nodes

Nodes and links can also have attributes associated to them

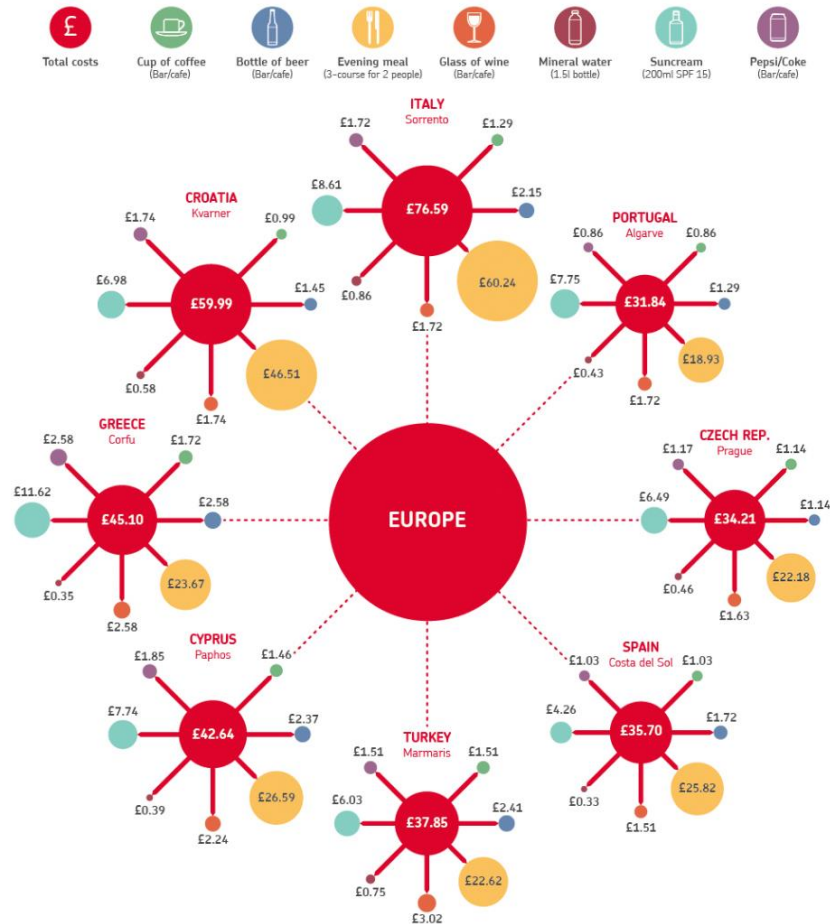
Node-link diagrams



Node-link diagrams



Node-link diagrams



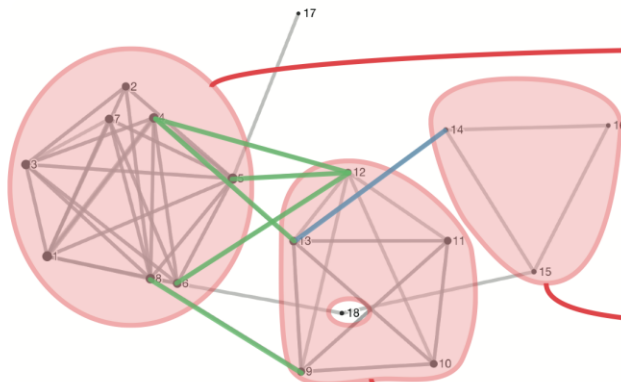
Adjacency matrix



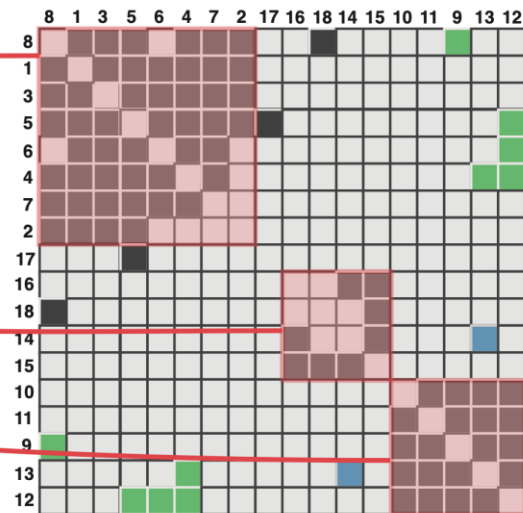
Adjacency Matrix

Well-known relative

Node-link diagram



Adjacency Matrix

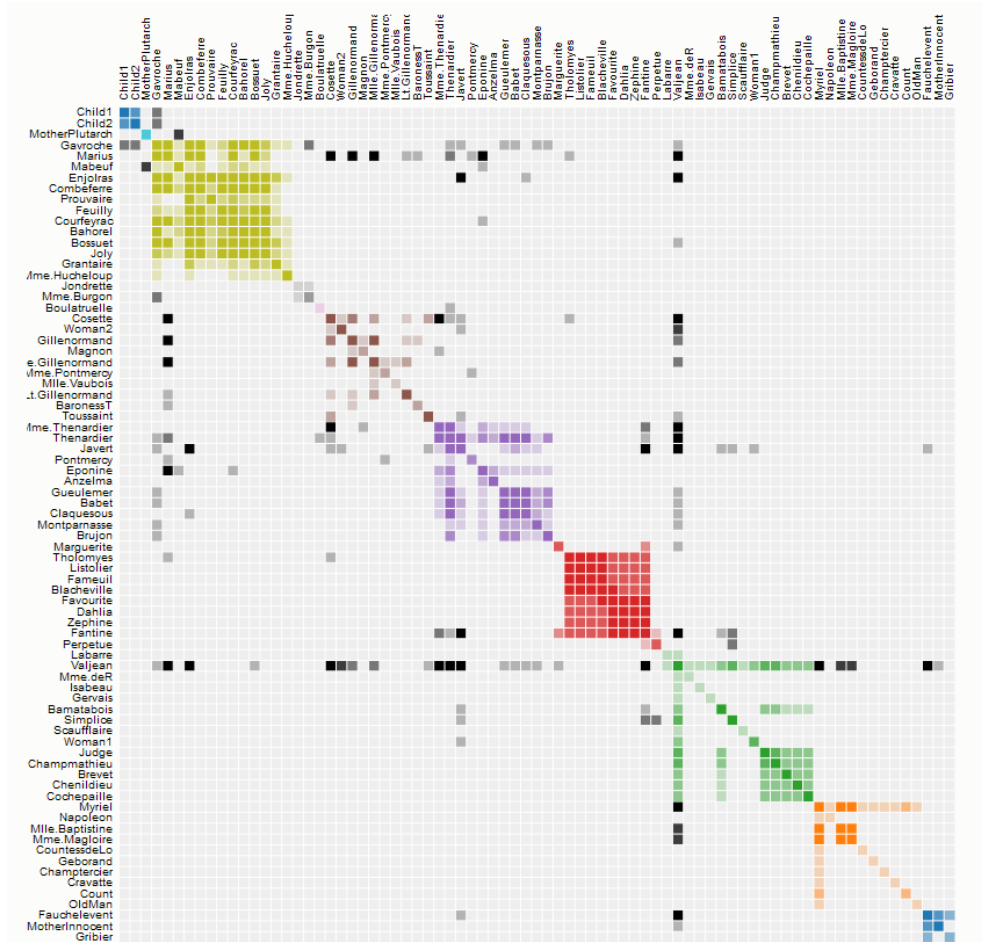


In the example, nodes are connected in three clusters, and these clusters are connected as well.



<https://visualizationcheatsheets.github.io>

Adjacency matrix



Co-occurrence of characters in Les Misérables

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

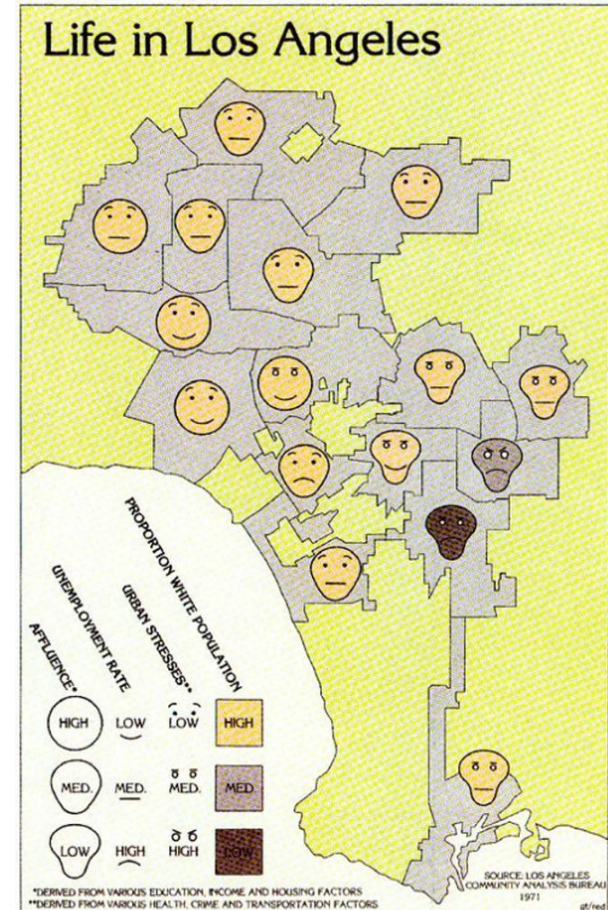
Chernoff faces

Visualization using glyphs

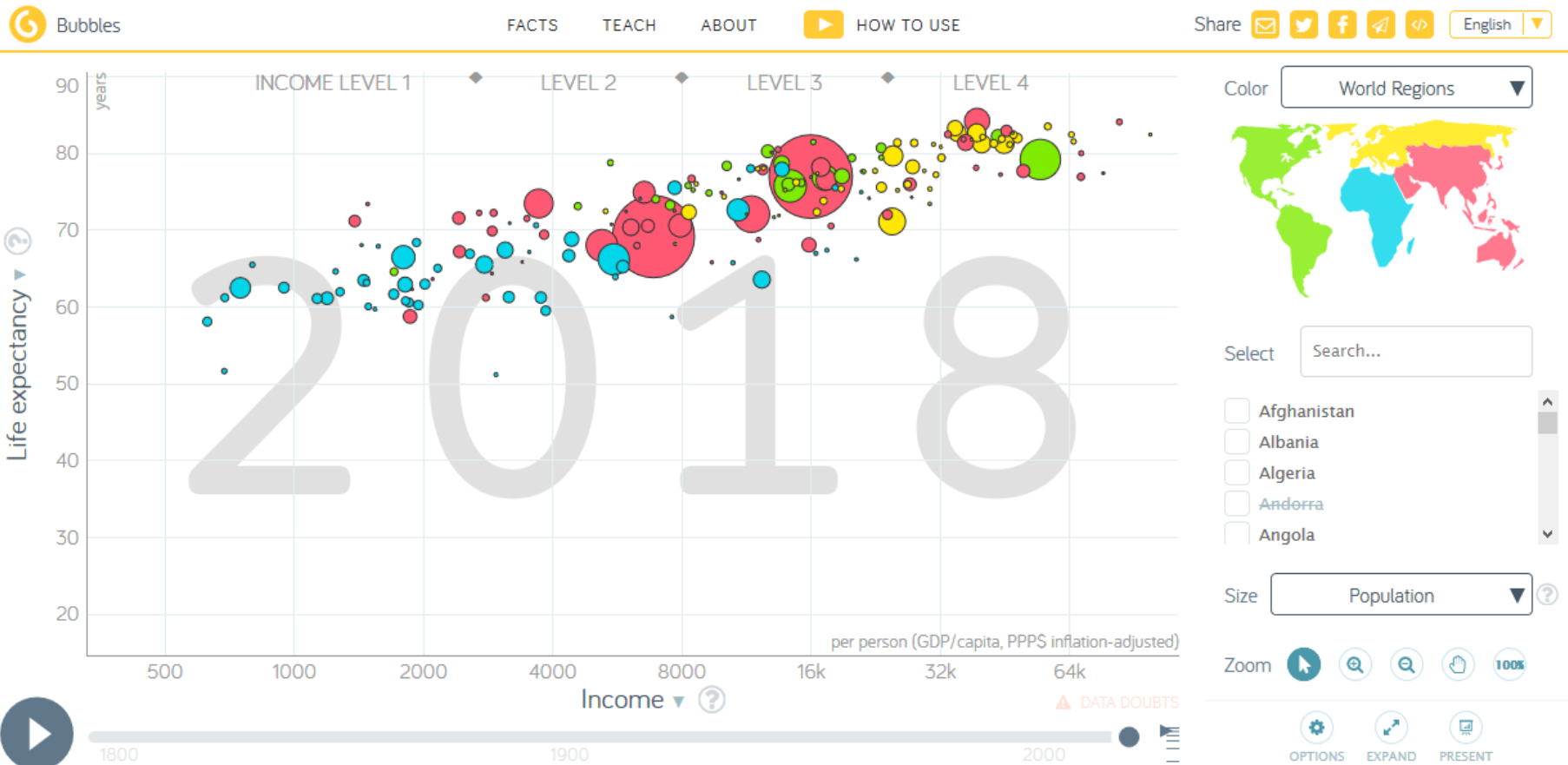
Can present up to 18 distinct variables

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

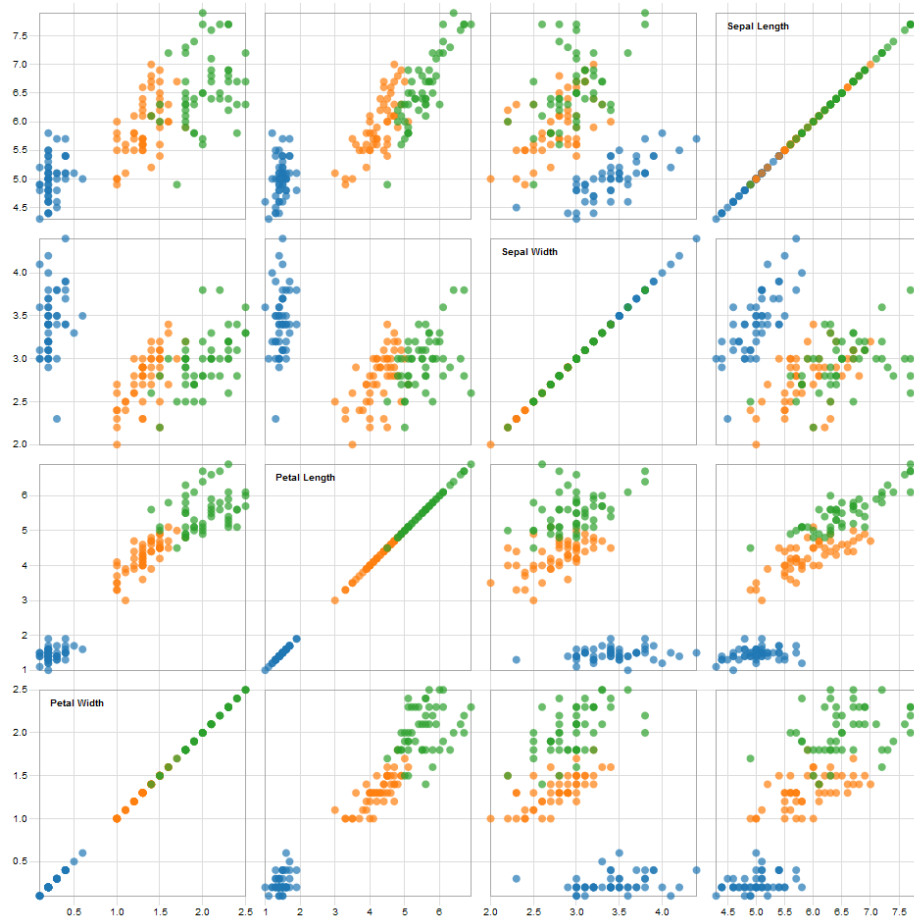
Questionable generalization



Bubble chart



Scatter plot matrix



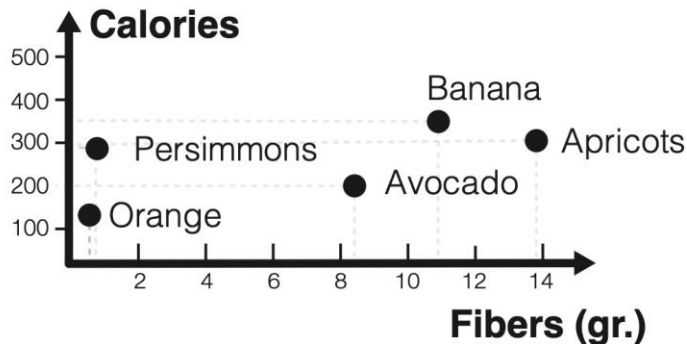
Parallel coordinate plot



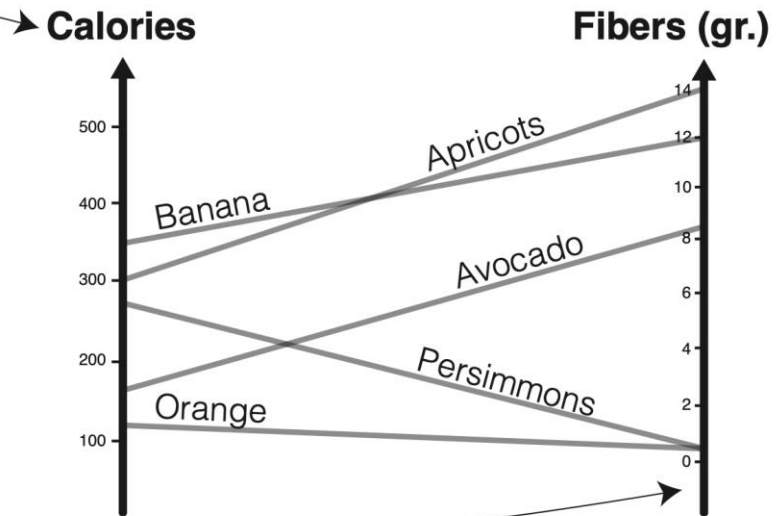
Parallel Coordinates

Well-known relative

Scatterplot

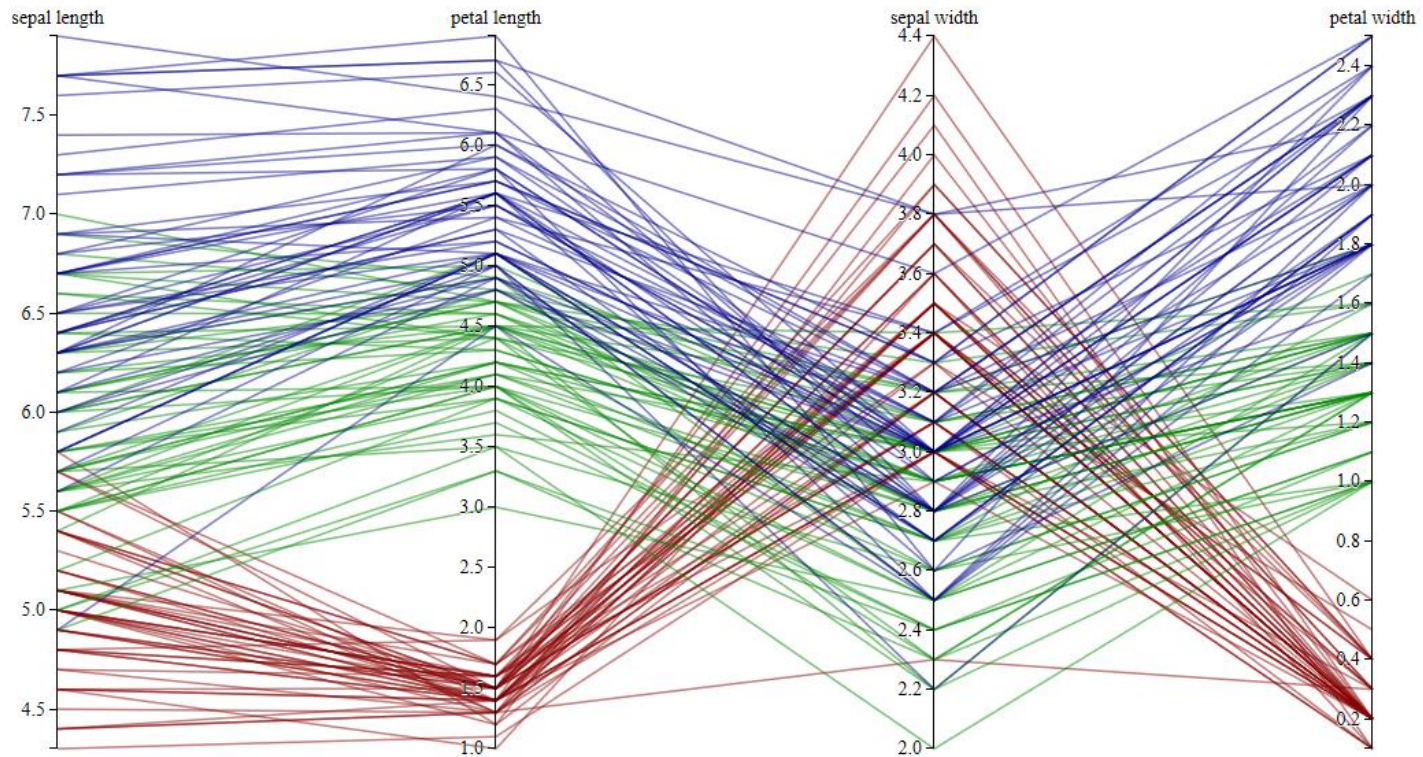


Parallel Coordinates



<https://visualizationcheatsheets.github.io>

Parallel coordinate plot



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

Edgar Anderson's *Iris* data set
parallel coordinates

Radar chart



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

Radial histogram

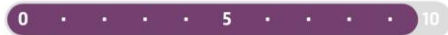


OECD countries / Italy

Friuli-Venezia Giulia

Health

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



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



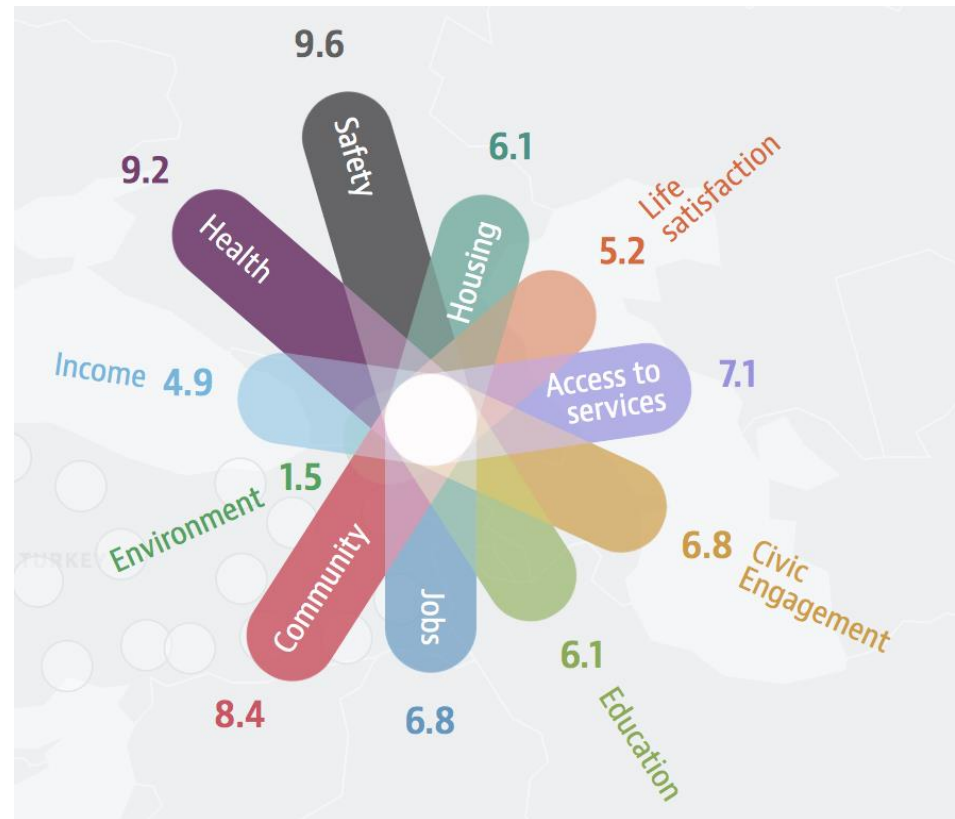
Compared across all OECD regions, the region is in the **top 10%** in **Health**.



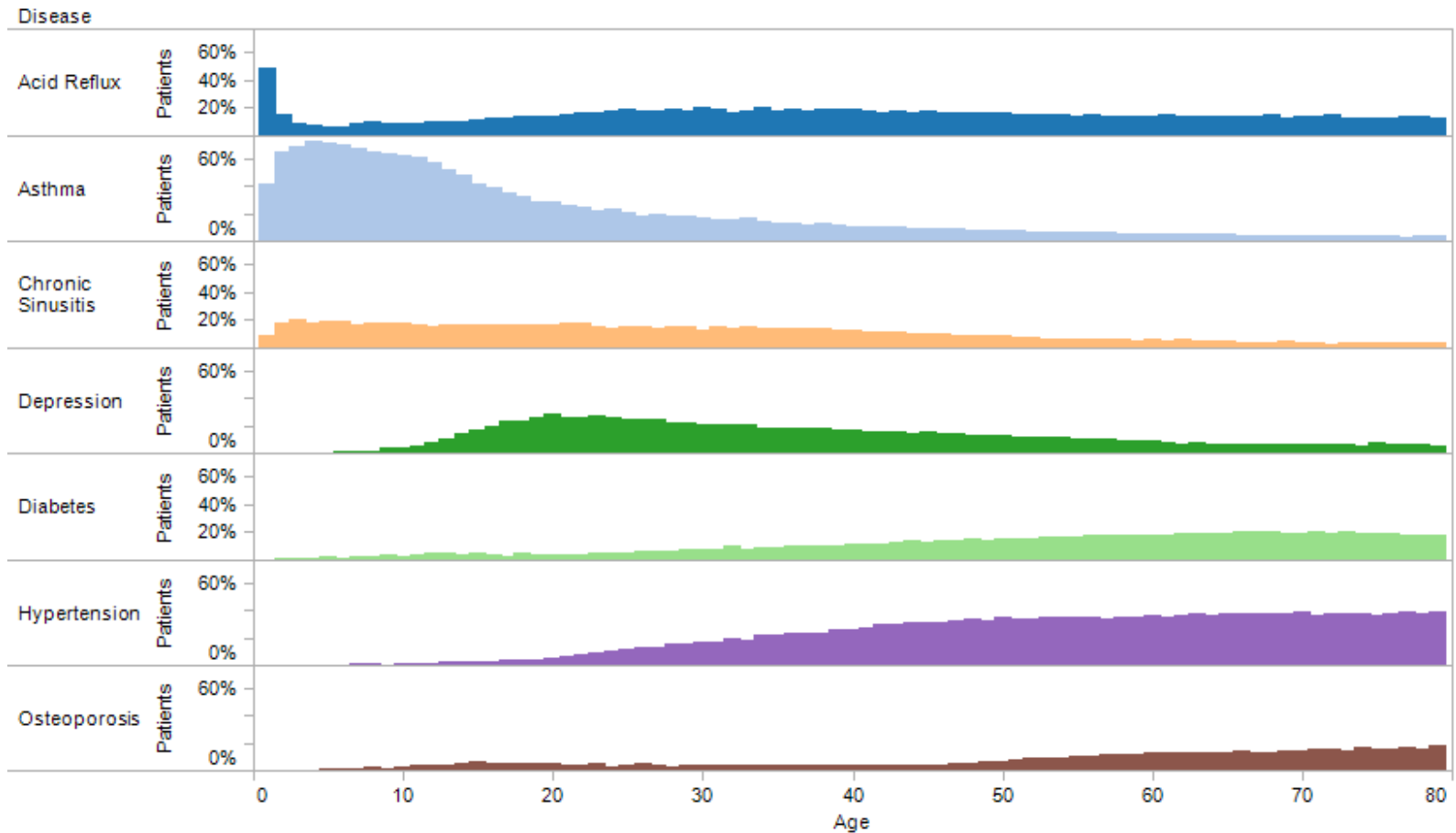
Indicators

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

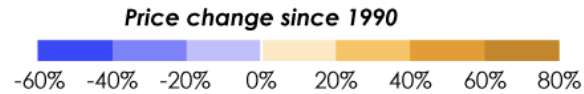
Life expectancy: **83.5** years



Small multiples



Horizon charts



Flour, white, all purpose, per lb.



Bread, white, pan, per lb.



Eggs, grade A, large, per doz.



Bananas, per lb.



Potatoes, white, per lb.



Sugar, white, all sizes, per lb.



Spaghetti and macaroni, per lb.



Cookies, chocolate chip, per lb.



Ice cream, prepackaged, bulk, regular, per 1/2 gal.



Lemons, per lb.



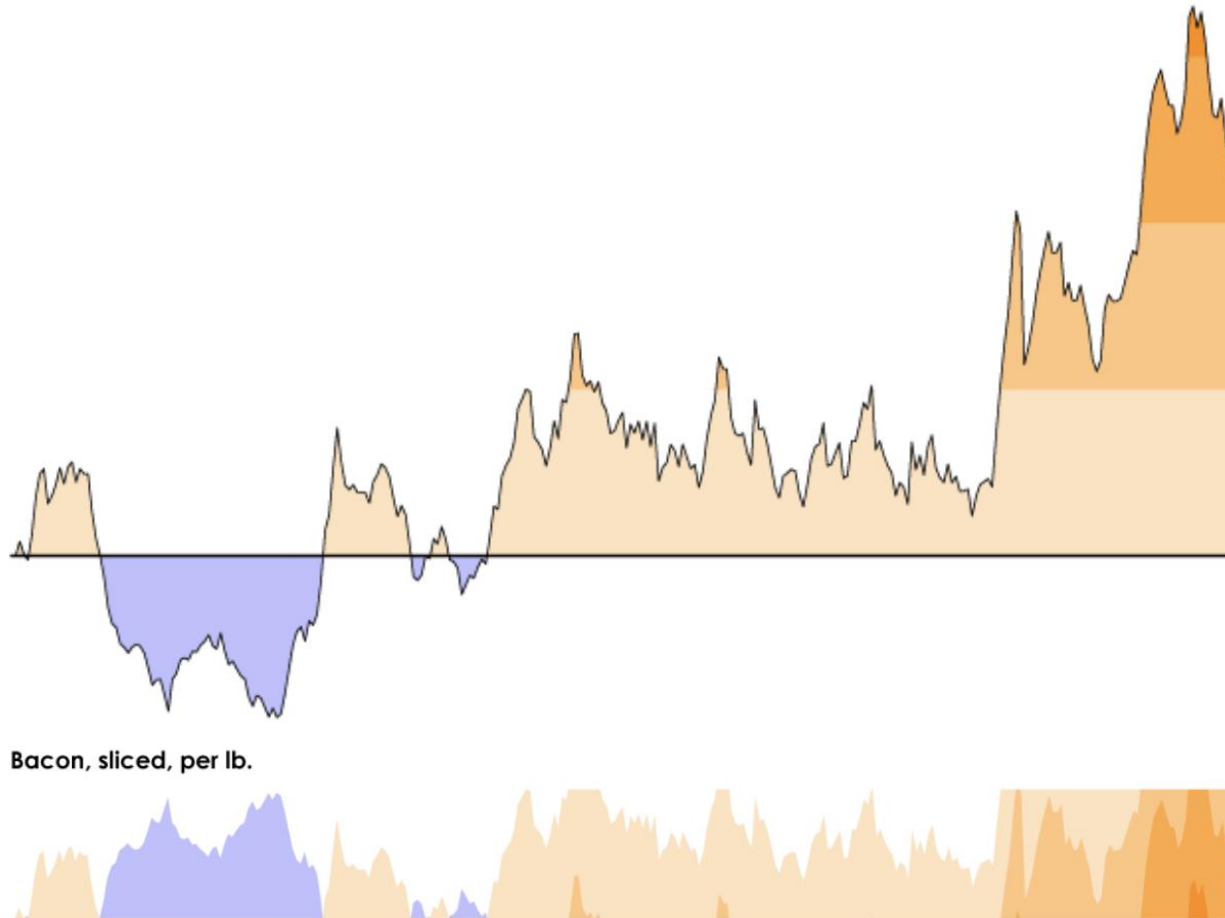
Orange juice, frozen concentrate, 12 oz. can, per 16 oz.



Sugar, white, 33-80 oz. pkg, per lb.



Horizon charts



Multivariate/multidimensional data visualization

Perform dimensionality reduction and visualize the results

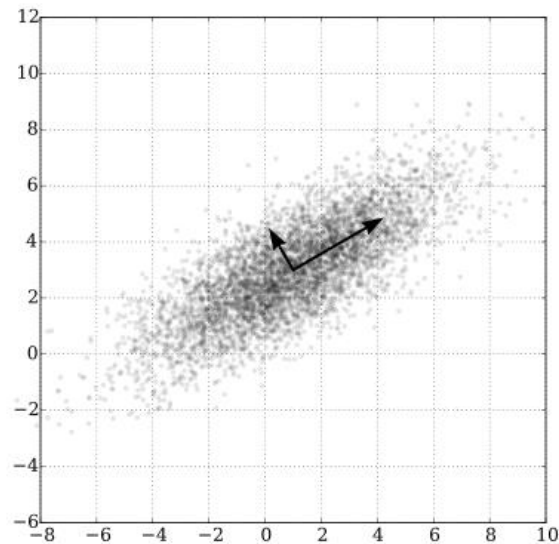
- Principal component analysis
- Multidimensional scaling

Transformation $\mathbb{R}^n \rightarrow \mathbb{R}^2$

Principal component analysis

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

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



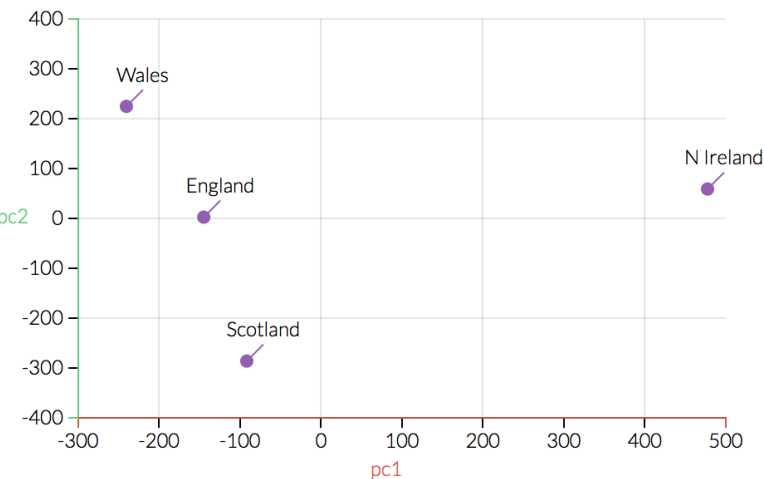
Principal component analysis

Eating in the UK

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

Alcoholic drinks
Beverages
Carcase meat
Cereals
Cheese
Confectionery
Fats and oils
Fish
Fresh fruit
Fresh potatoes
Fresh Veg
Other meat
Other Veg
Processed potatoes
Processed Veg
Soft drinks
Sugars

	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**

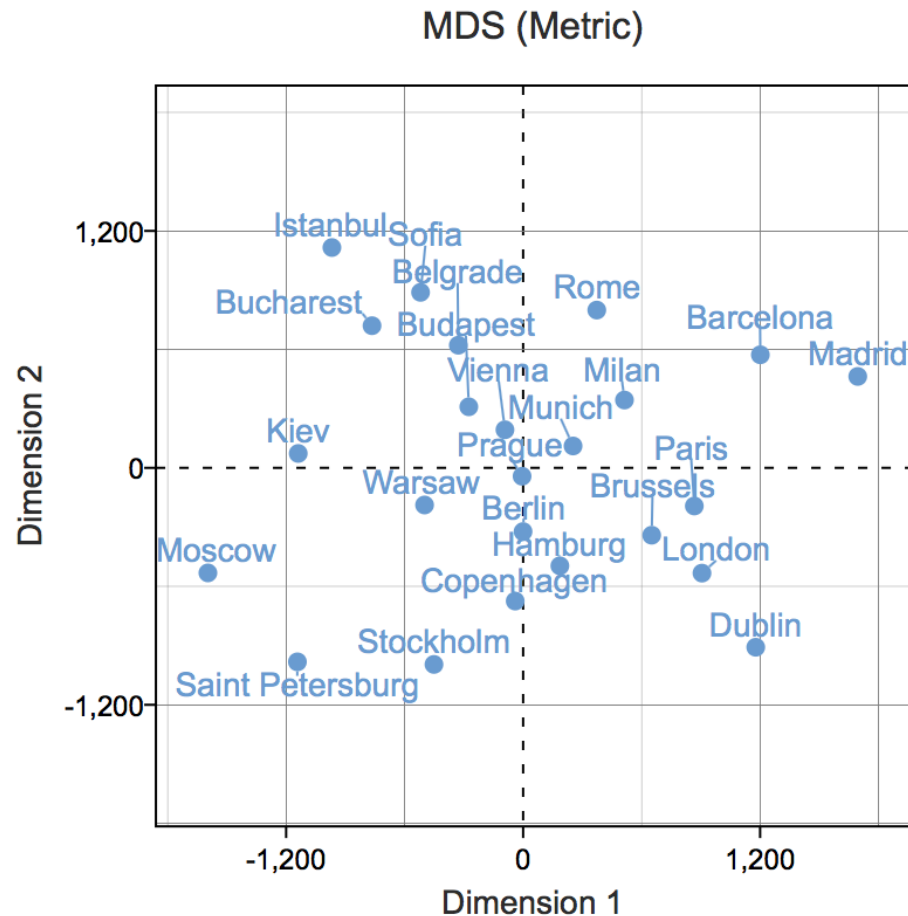
Useful for visualizing similarity matrices or graphs where you wish to preserve distances between nodes

Minimize the stress function

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

Solve with any method for optimizing nonlinear functions

Multidimensional scaling



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

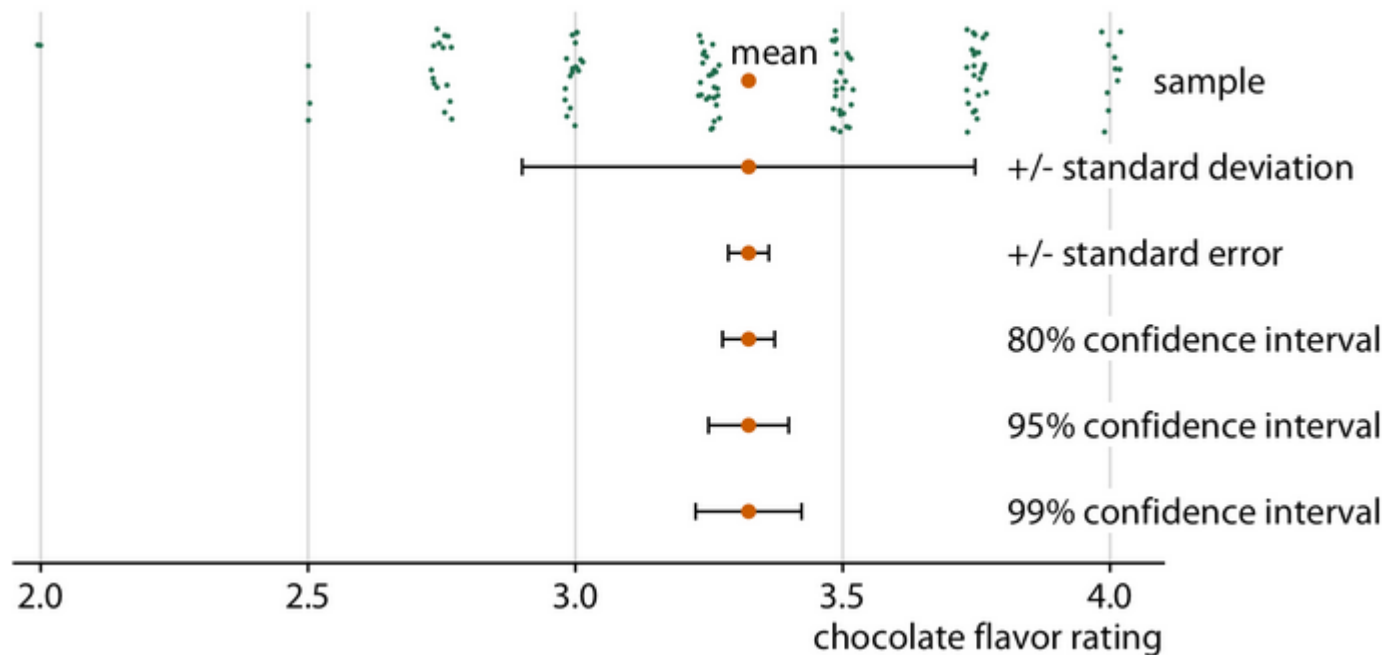
Visualizing uncertainty

Techniques to show uncertainty

- Ranges
- Distributions
- Multiple outcomes
- Obscurity

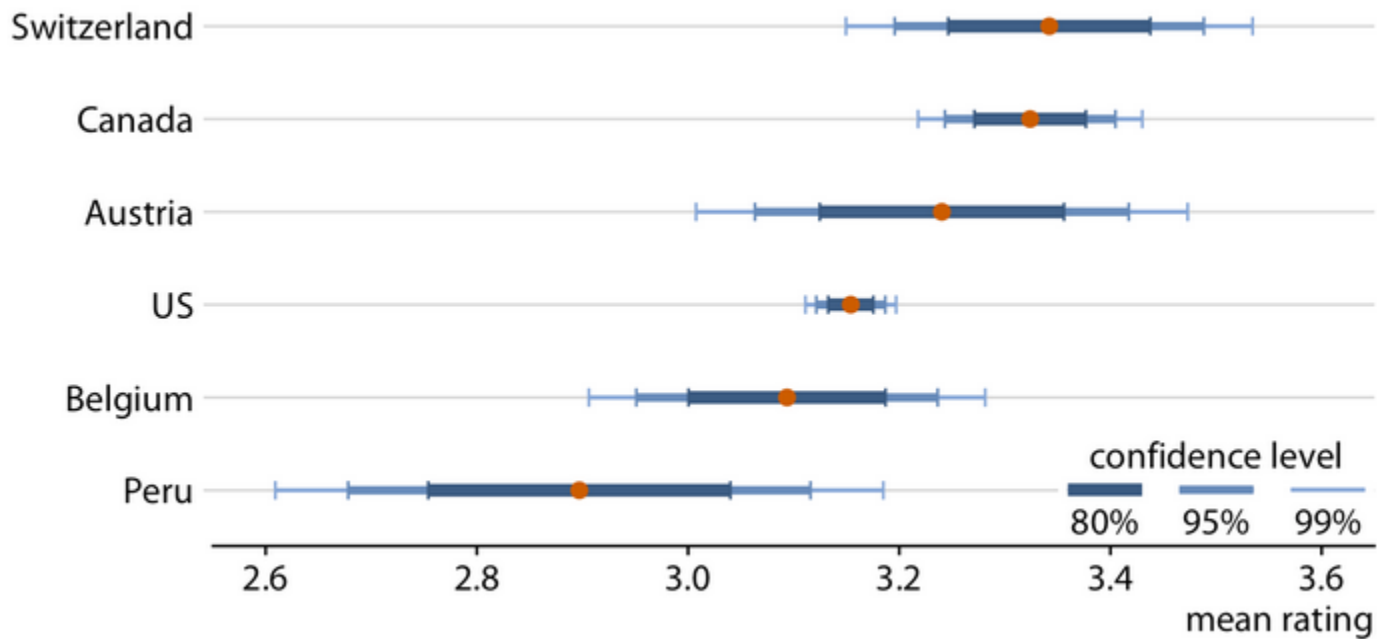
Visualizing uncertainty with ranges

Specify what the range represents



Visualizing uncertainty with ranges

Specify what the range means



Visualizing uncertainty with ranges

Luka Doncic

DALLAS MAVERICKS
SHOOTING GUARD
20 YEARS OLD



WEIGHTED AVERAGE OF
PAST THREE SEASONS

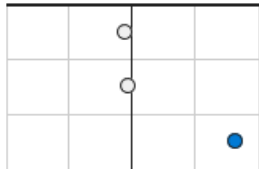
● BAD ○ AVG. ● GOOD

PERCENTILE

50TH

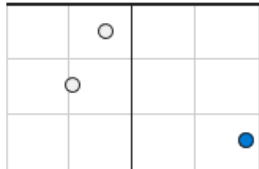
Vitals

Height	6' 7"
Weight	218
Draft position	3



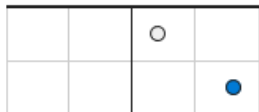
Scoring

True shooting %	55%
Free throw %	71%
Usage %	31%



Tendencies

3 pt. frequency	43%
FT frequency	41%

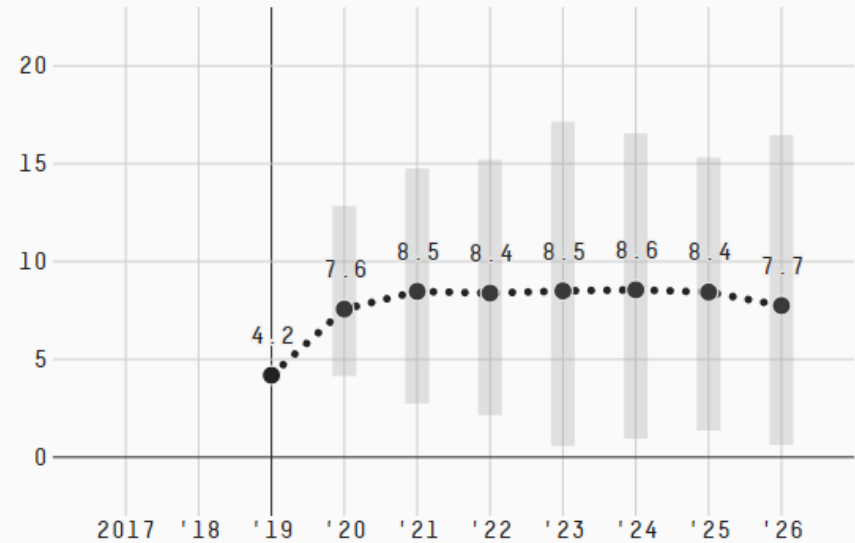


Wins above replacement projection

CATEGORY: ALL-STAR

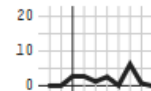
5-YR MARKET VALUE: \$362.3M

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

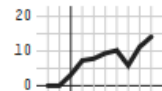


Performance of the 10 most comparable players

1 **Tyreke Evans**
YEAR: 2011
SIMILARITY: 51



6 **R. Westbrook**
YEAR: 2010
SIMILARITY: 36



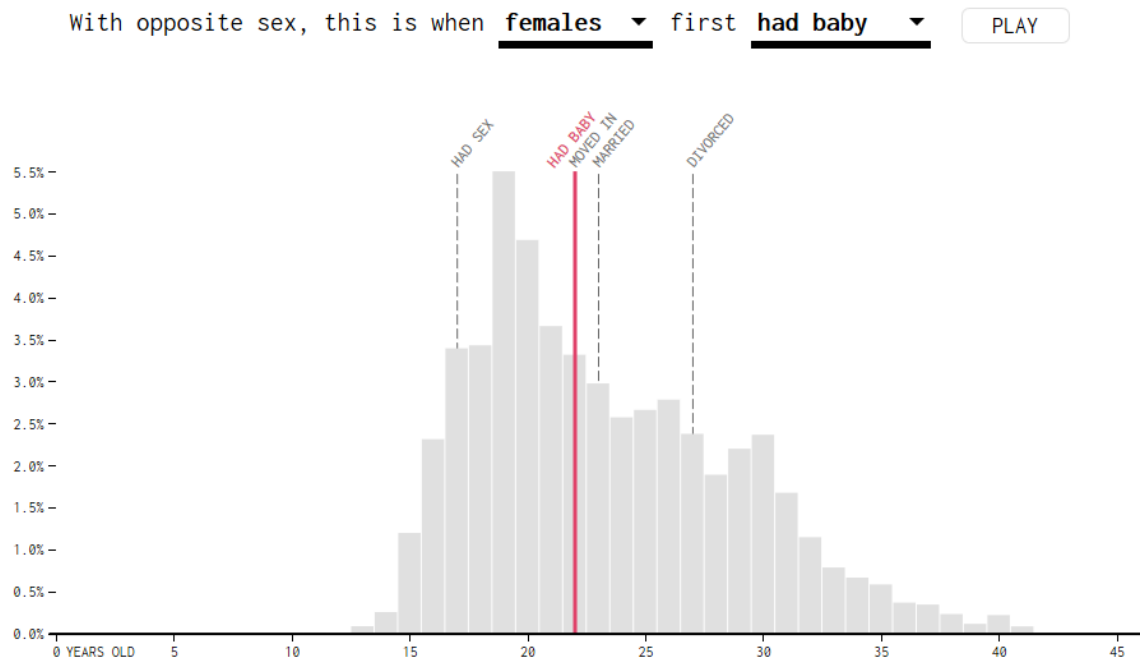
Visualizing uncertainty with ranges

Earnings per share outlook



Visualizing uncertainty with distributions

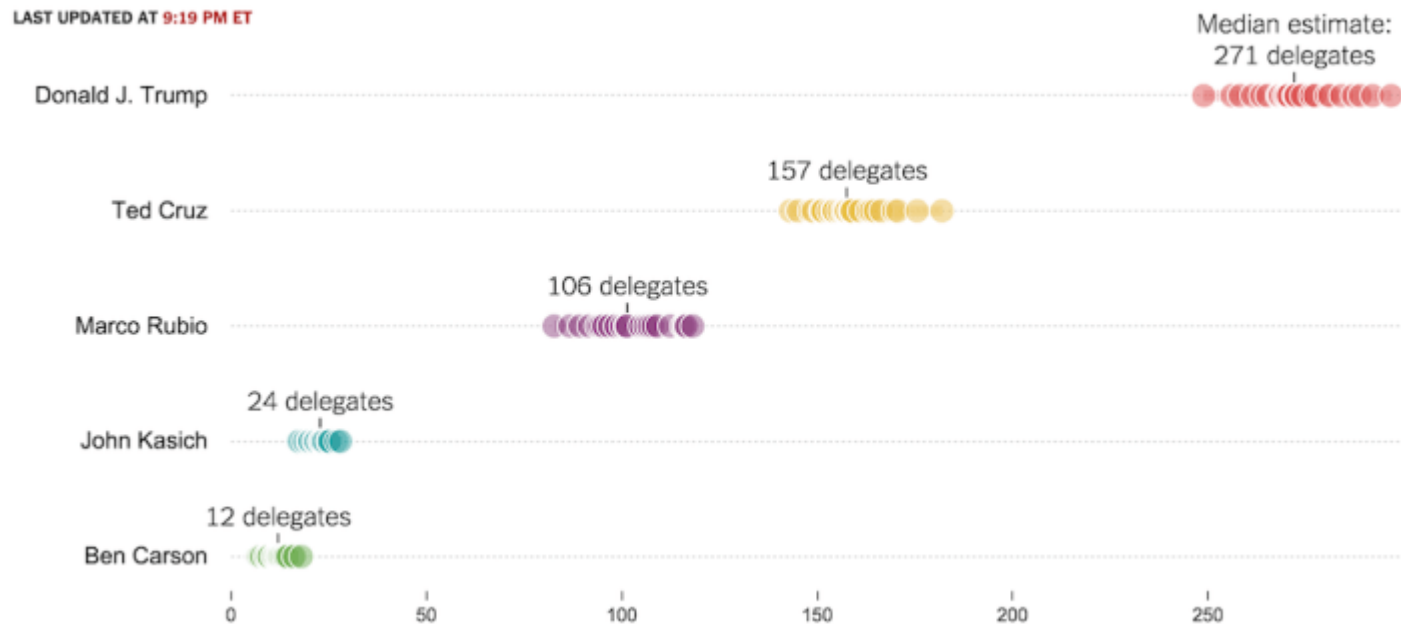
Show the spread of possible values with a histogram (or a variant of it)



Visualizing uncertainty with multiple outcomes

Show the various outcomes

Estimates of the Republican delegate count



Visualizing missing data with obscurity

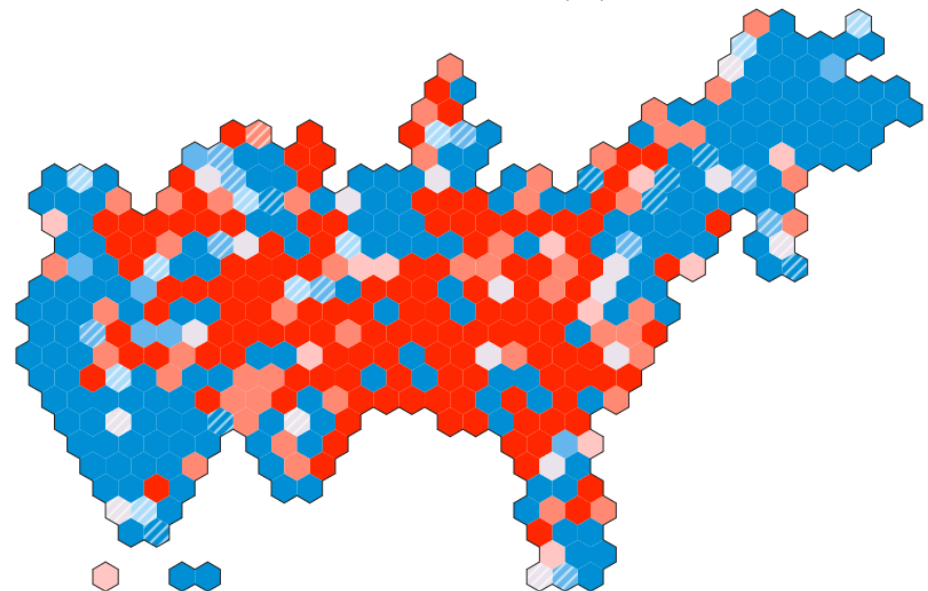
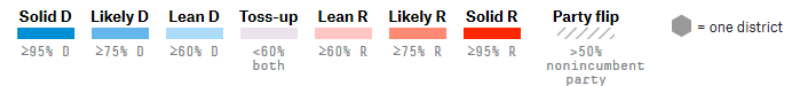
Use transparency, color scale, or blurriness to show uncertainty

Our forecast for every district

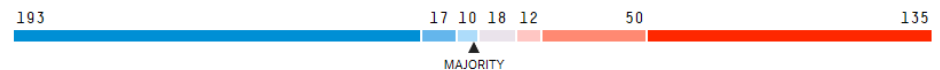
Cartogram

Map

The chance of each candidate winning, with all 435 House districts shown at the same size

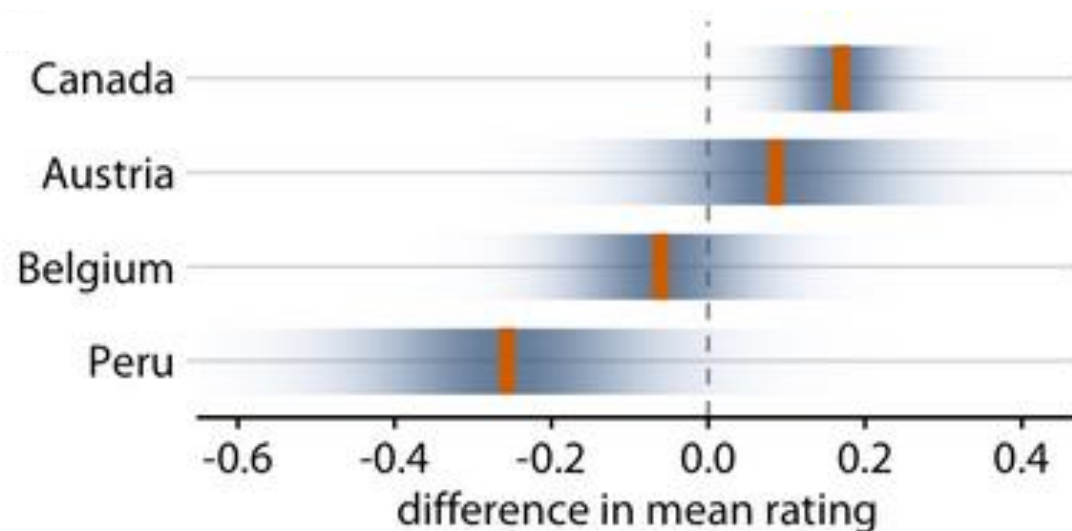


District totals by category



Visualizing missing data with obscurity

Use transparency, color scale, or blurriness to show uncertainty



Visualizing missing data

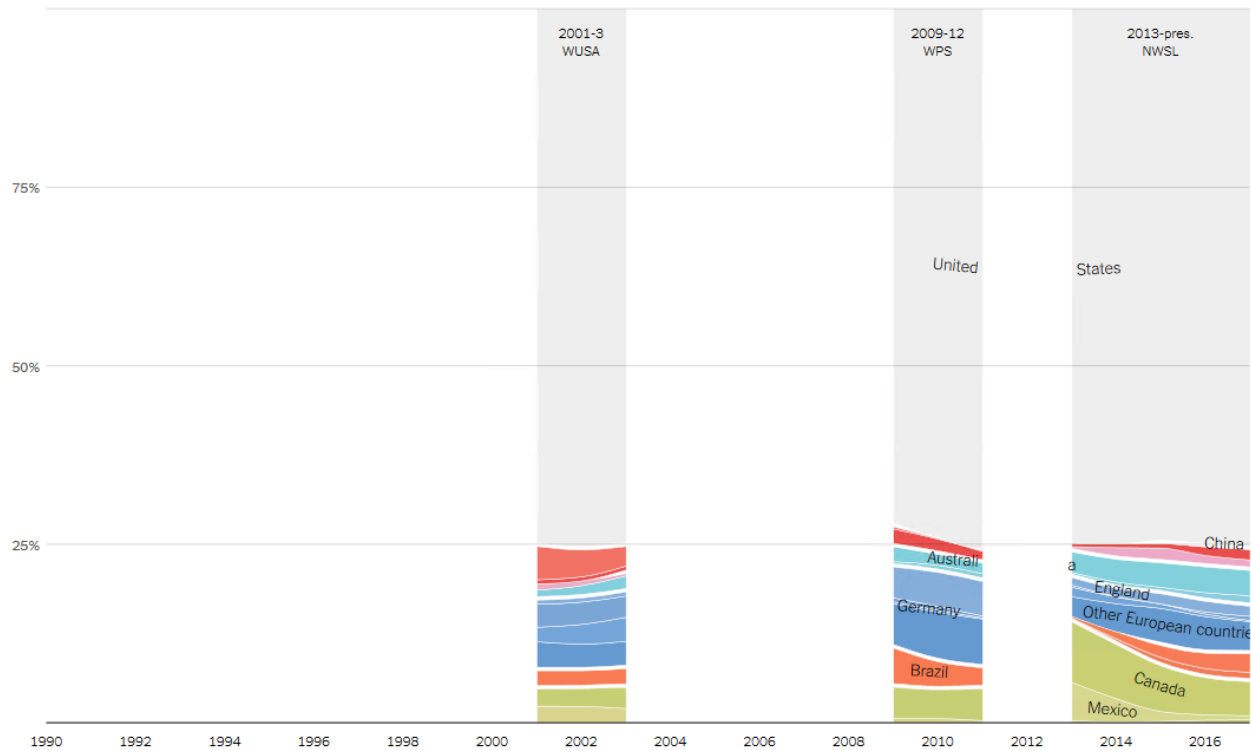
Techniques to handle missing data

- 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

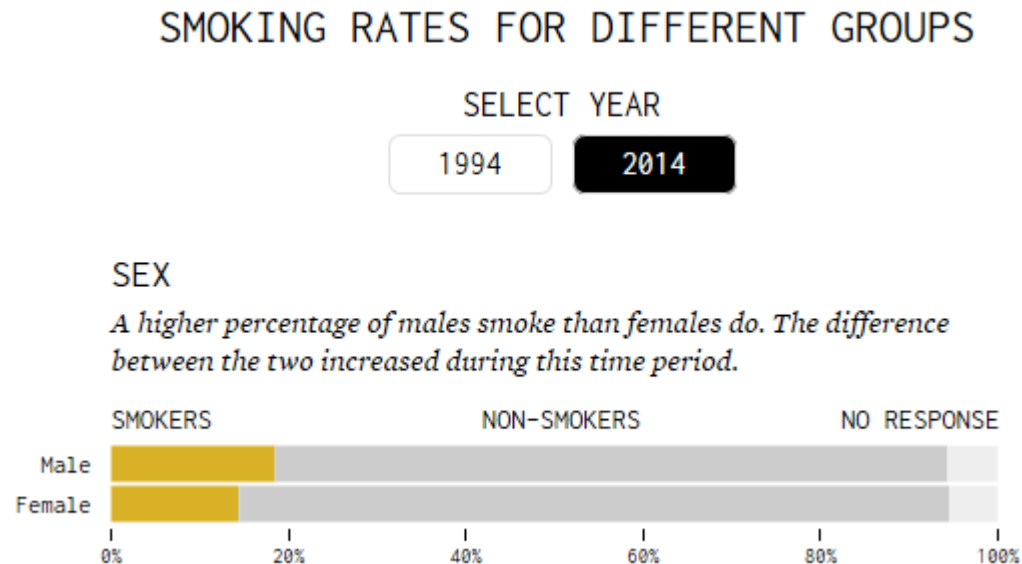
Where players in **U.S. Women's Soccer** have come from



Source: Based on roster data compiled by Jen Cooper

Visualizing missing data as a category

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



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

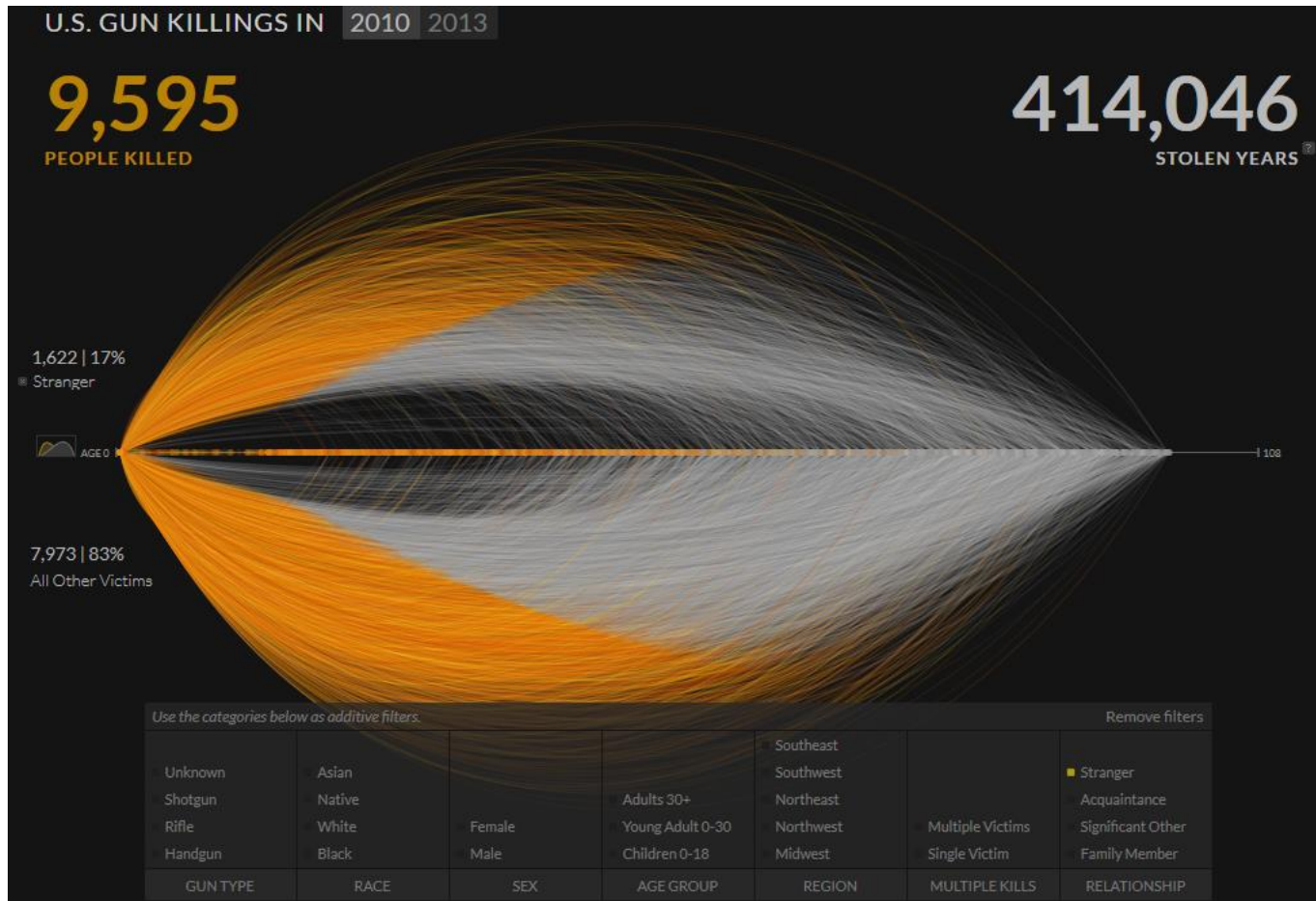
Can affect

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

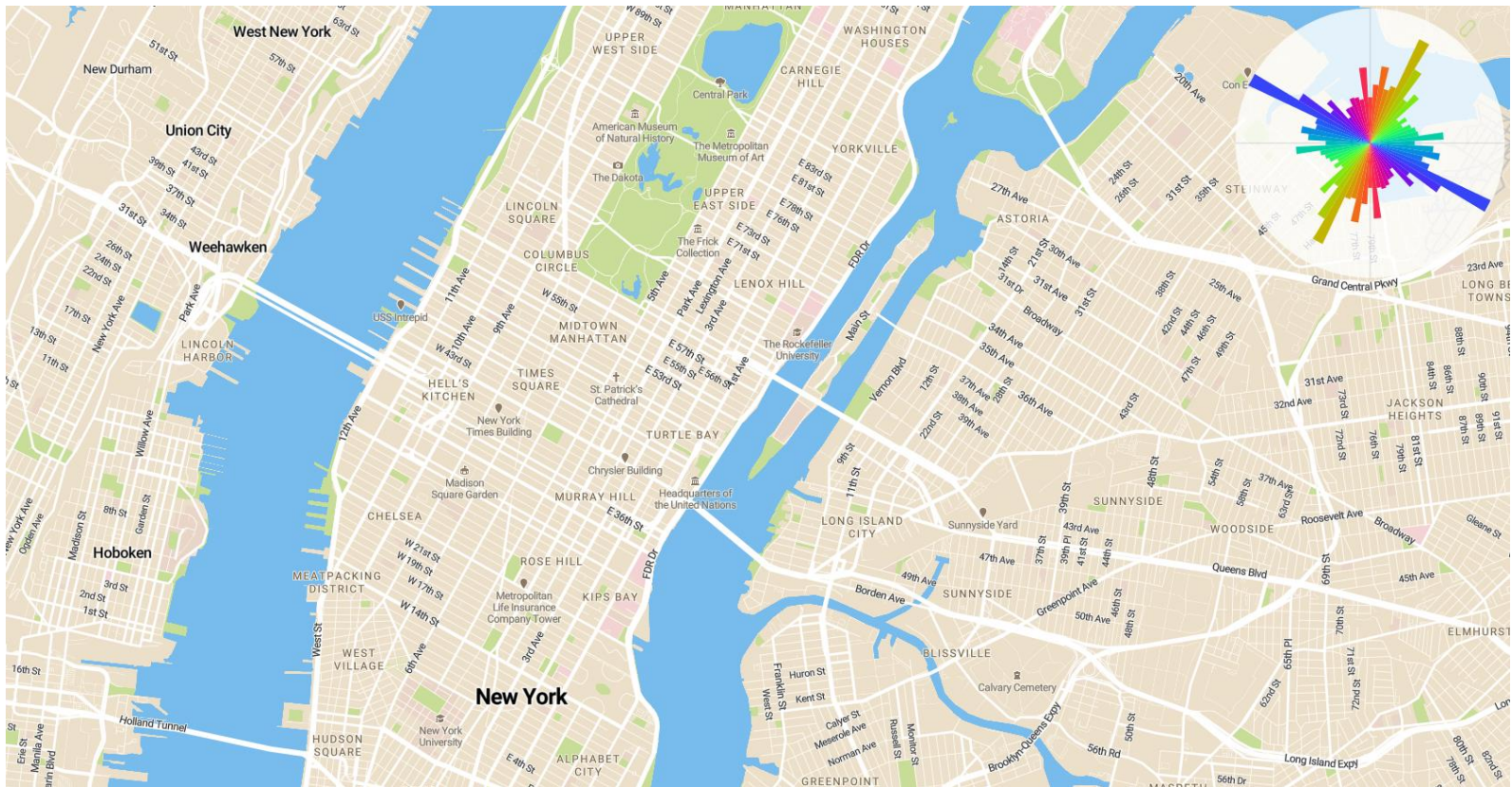
Data adjustments

- **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



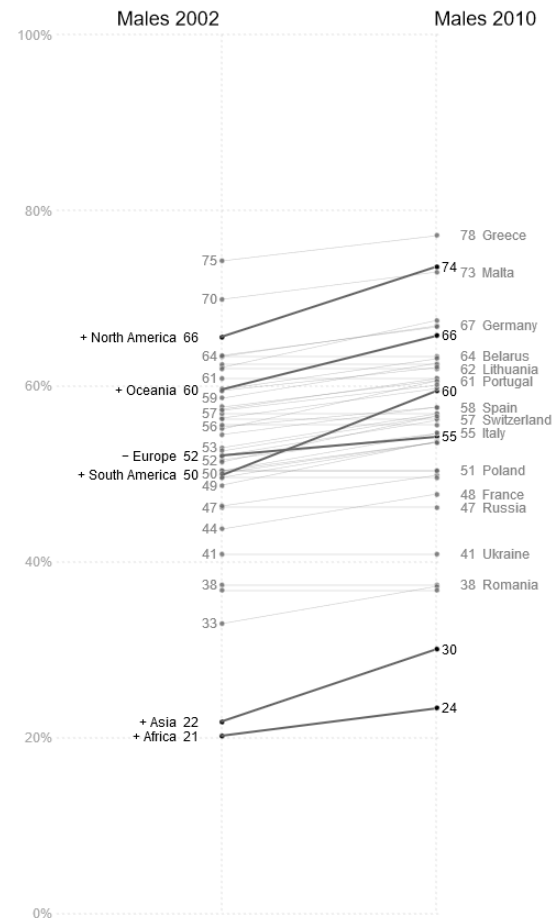
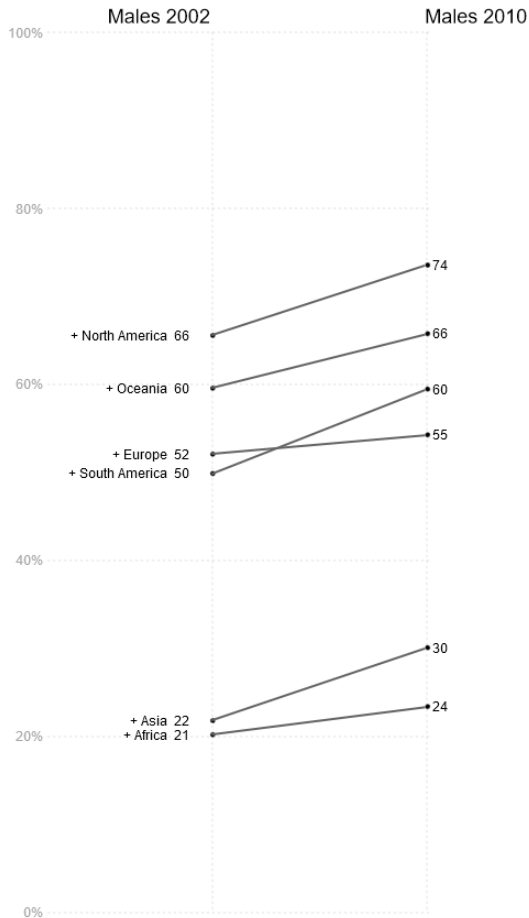
Navigating



Road orientation map

Navigating

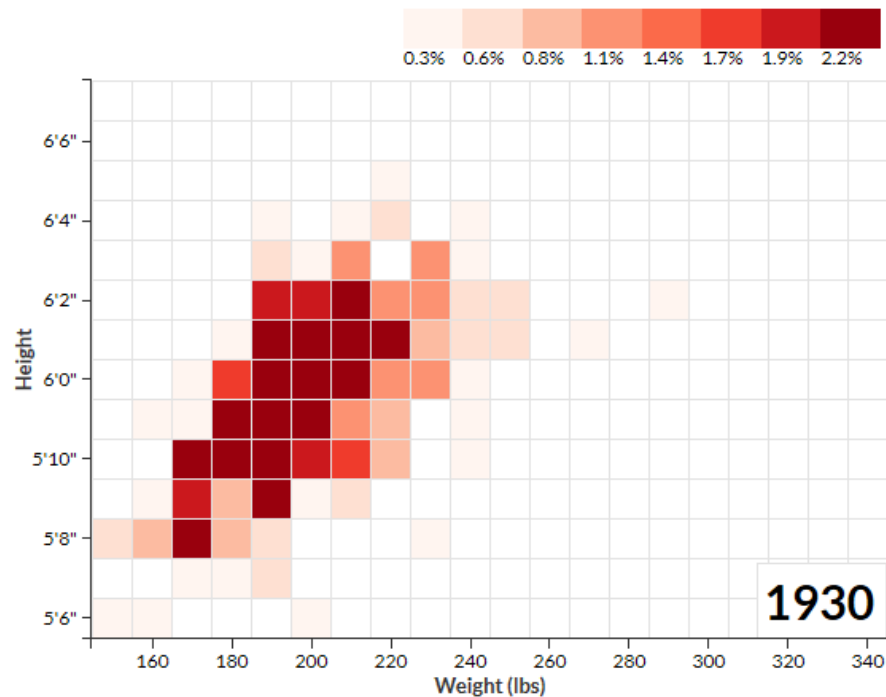
Obesity
around
the world



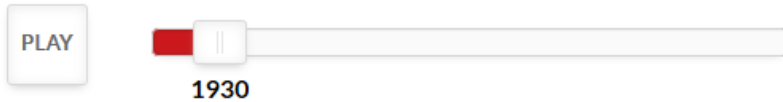
Animating

NFL players: height & weight over time

By [Noah Veltman](#)

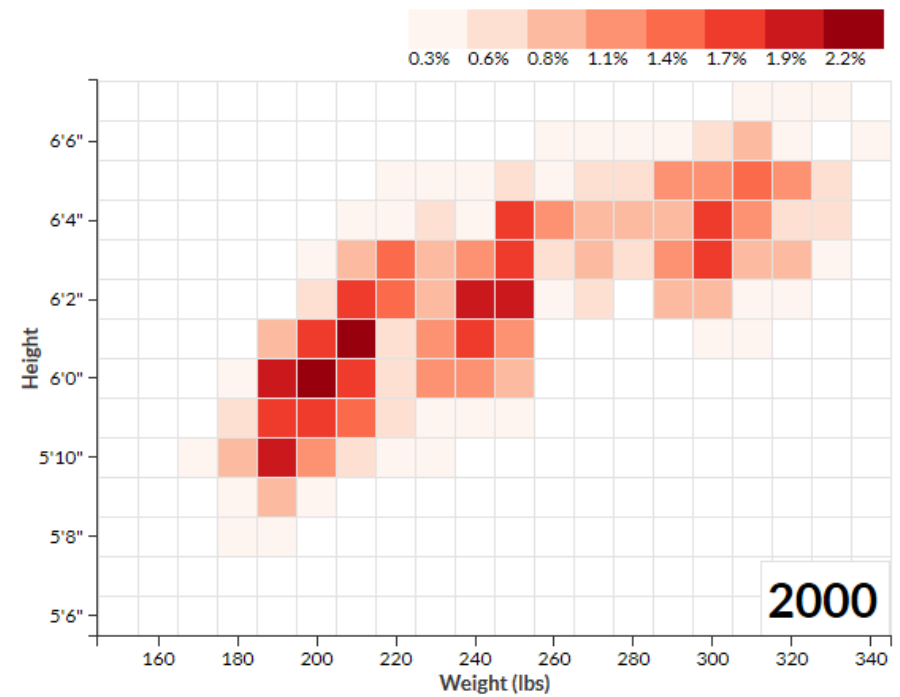


1930



NFL players: height & weight over time

By [Noah Veltman](#)



2000

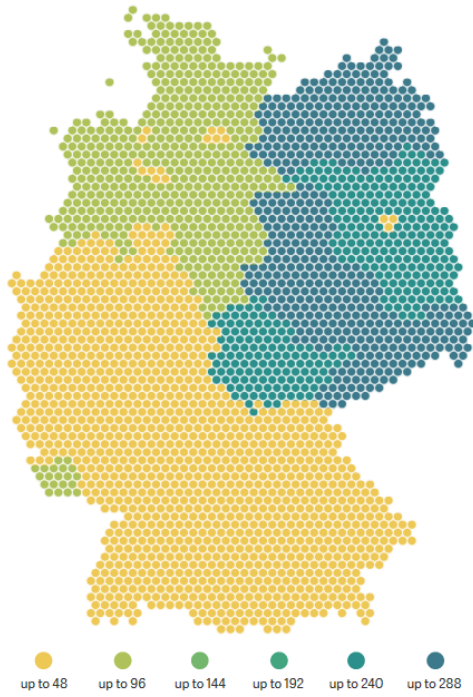


Sequencing



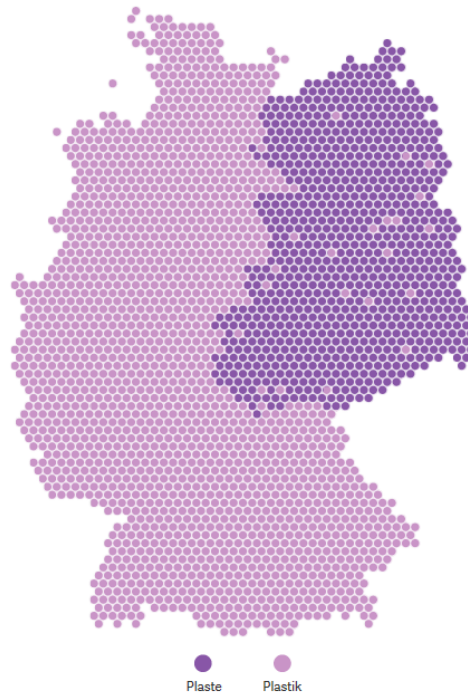
Agriculture

Average farm size in hectares, 2010



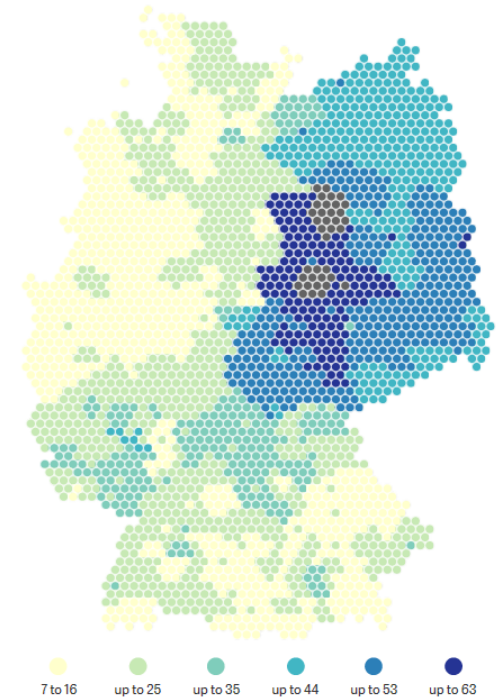
Plastic in German?

Word used regionally

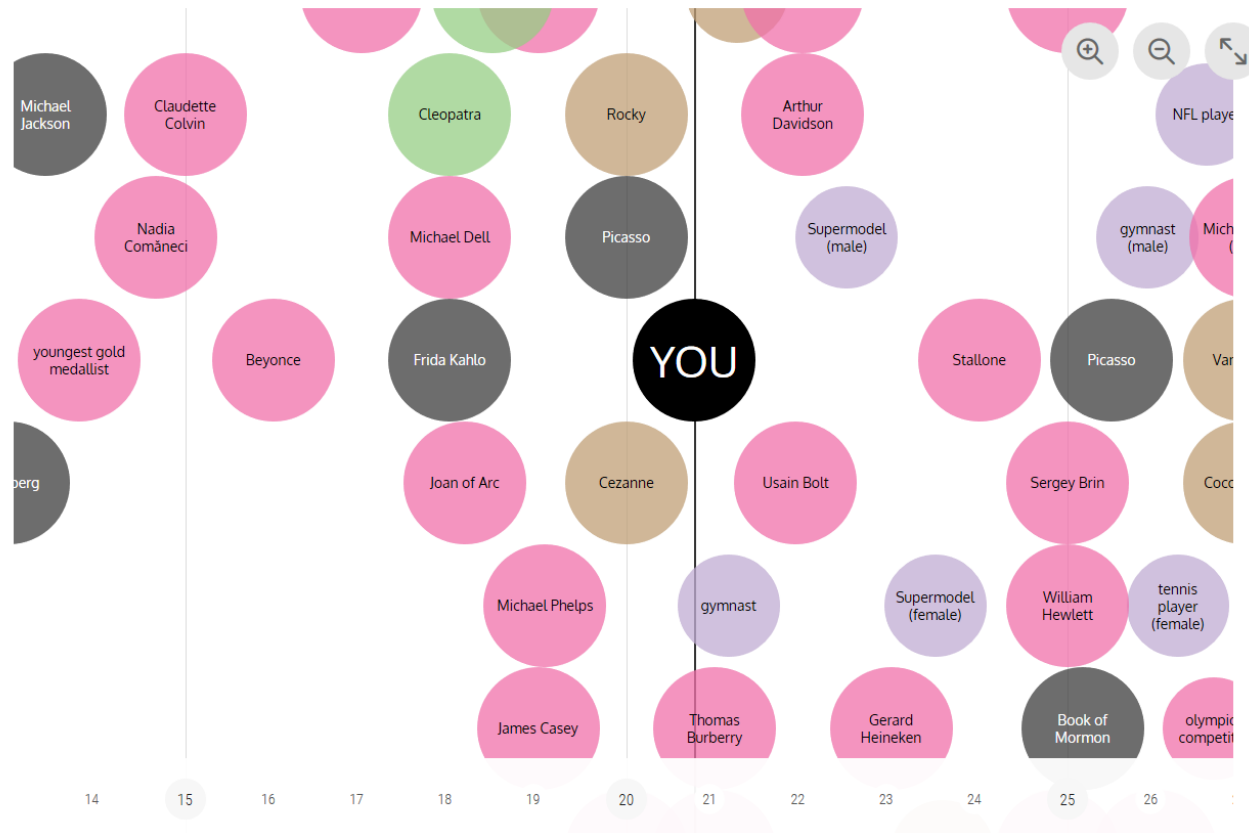


Day Care

Percentage of children under two in day care, 2012 (gray: no data)



Contributing



Who Old Are You?

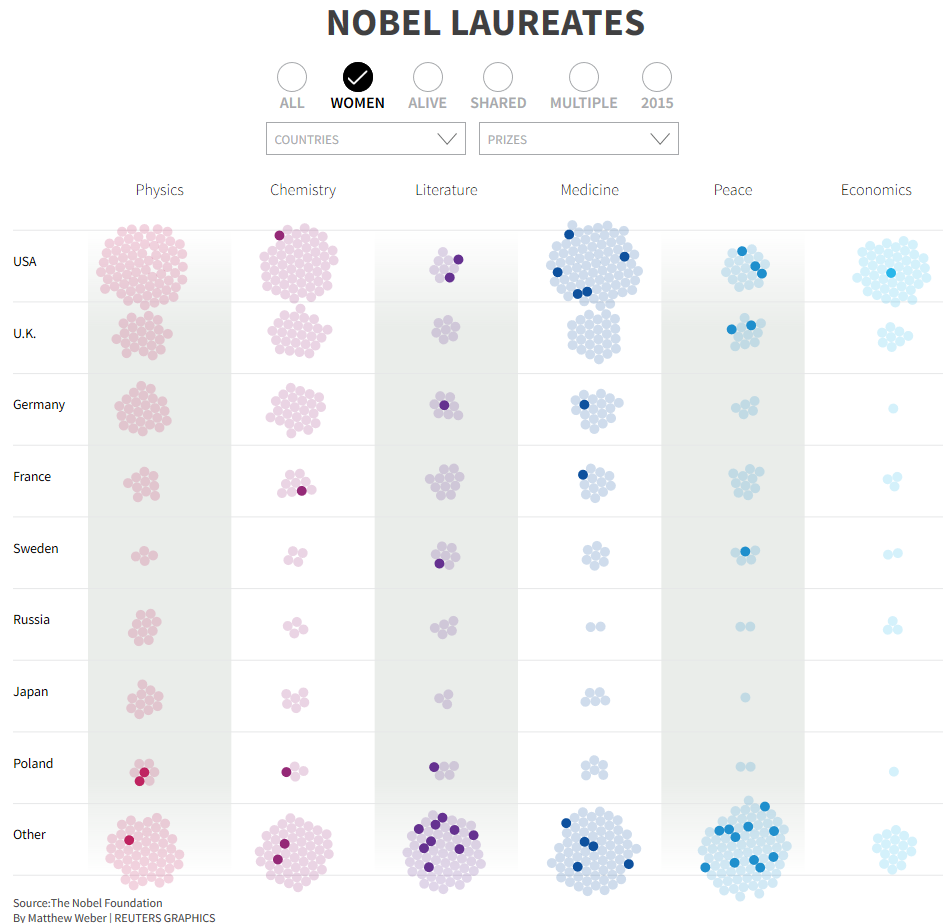
concept & design: David McCandless | code: [Tom Evans](#) | [informationisbeautiful.net](#)

categorise average age career high creative zenith life moment worldly success

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



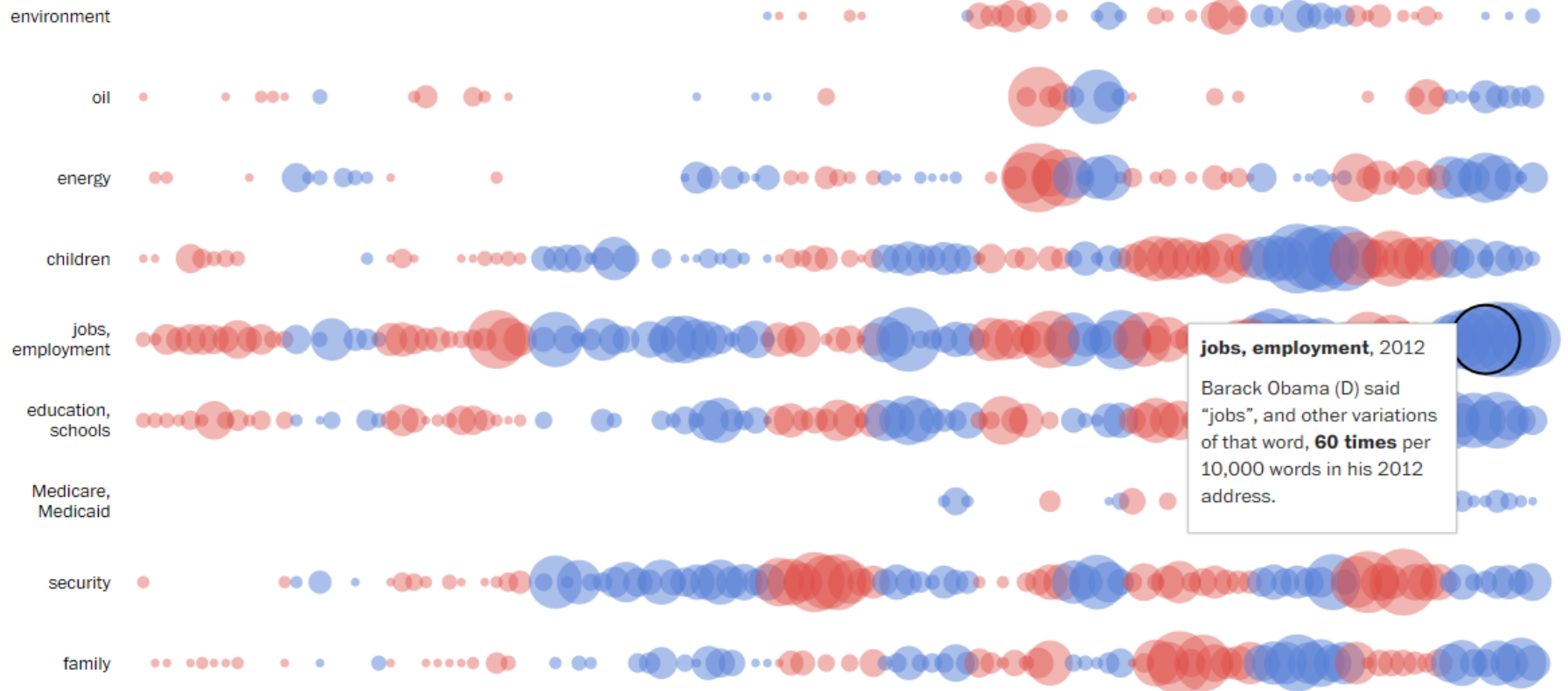
Focusing



Using
brushing
and linking

Annotating

History through the president's words



Orientating

KILLING the Colorado

- Read the Latest Story
- What You Need to Know

- BIG THOMPSON PROJECT
- MOFFAT TUNNEL
- FLAMING GORGE DAM
- NAVAJO GENERATING STATION
- GLEN CANYON DAM
- HOOVER DAM
- CENTRAL ARIZONA PROJECT**
- PARKER DAM
- IMPERIAL DAM
- YUMA DESALTING COMPLEX
- ALL-AMERICAN CANAL

CENTRAL ARIZONA PROJECT

GETS US... **500 million kilowatt hours of energy**

COSTS US... **5.2 billion gallons/year in evaporation**
2.9 billion gallons/year in seepage

COMPLETED **1993** COST **\$4.4B**

The largest canal in the country, the Central Arizona Project stretches 336 miles across the Sonoran desert, pumping an average 408 billion gallons of water each year up 3,000 feet to serve the Phoenix and Tucson metro areas. The canal also irrigates up to 700,000 acres of farmland along the way. The Central Arizona Project uses more energy than any other facility in the state, and officials are expected to cut its water deliveries as soon as next year if an emergency shortage is declared.

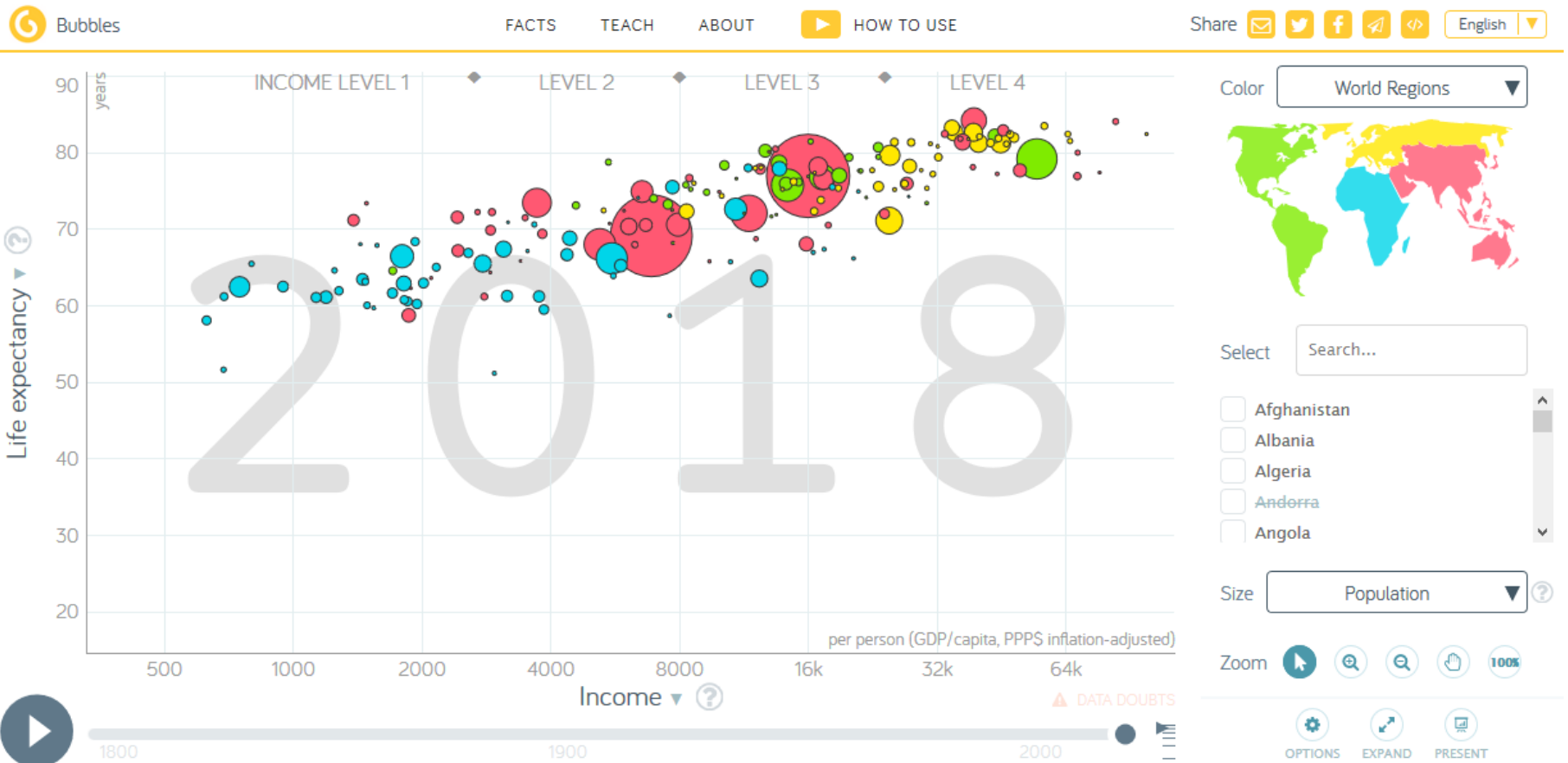
Colorado River (Lake Havasu) → Phoenix → Tucson (2,850 ft.)
336 MILES

CALIFORNIA

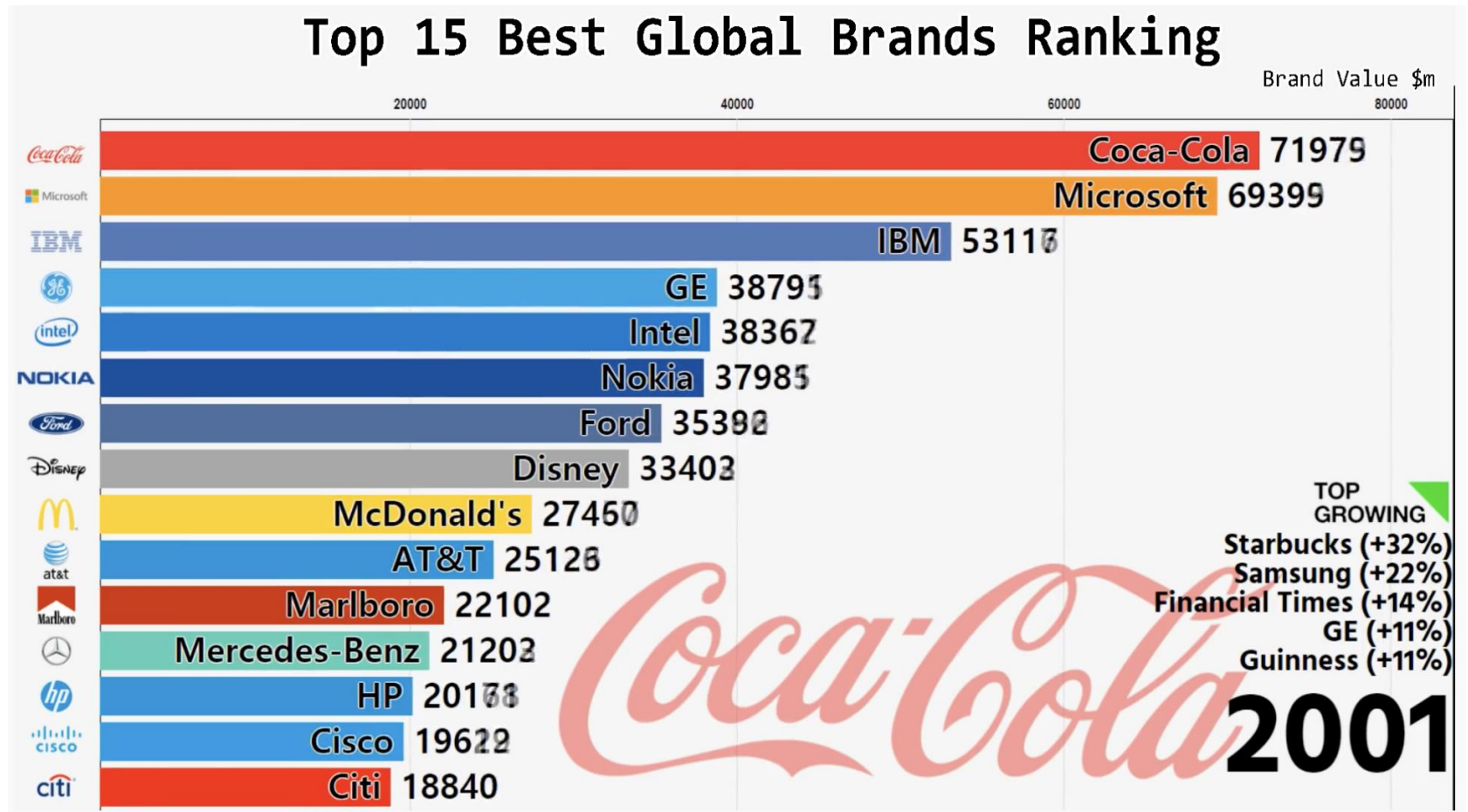
LAKE HAVASU
Lake Havasu City

WYO.
NEV. UTAH COLO.
CALIF. ARIZ. N.M.
Colorado River

Interactivity example



Animation example



Storytelling

Storytelling

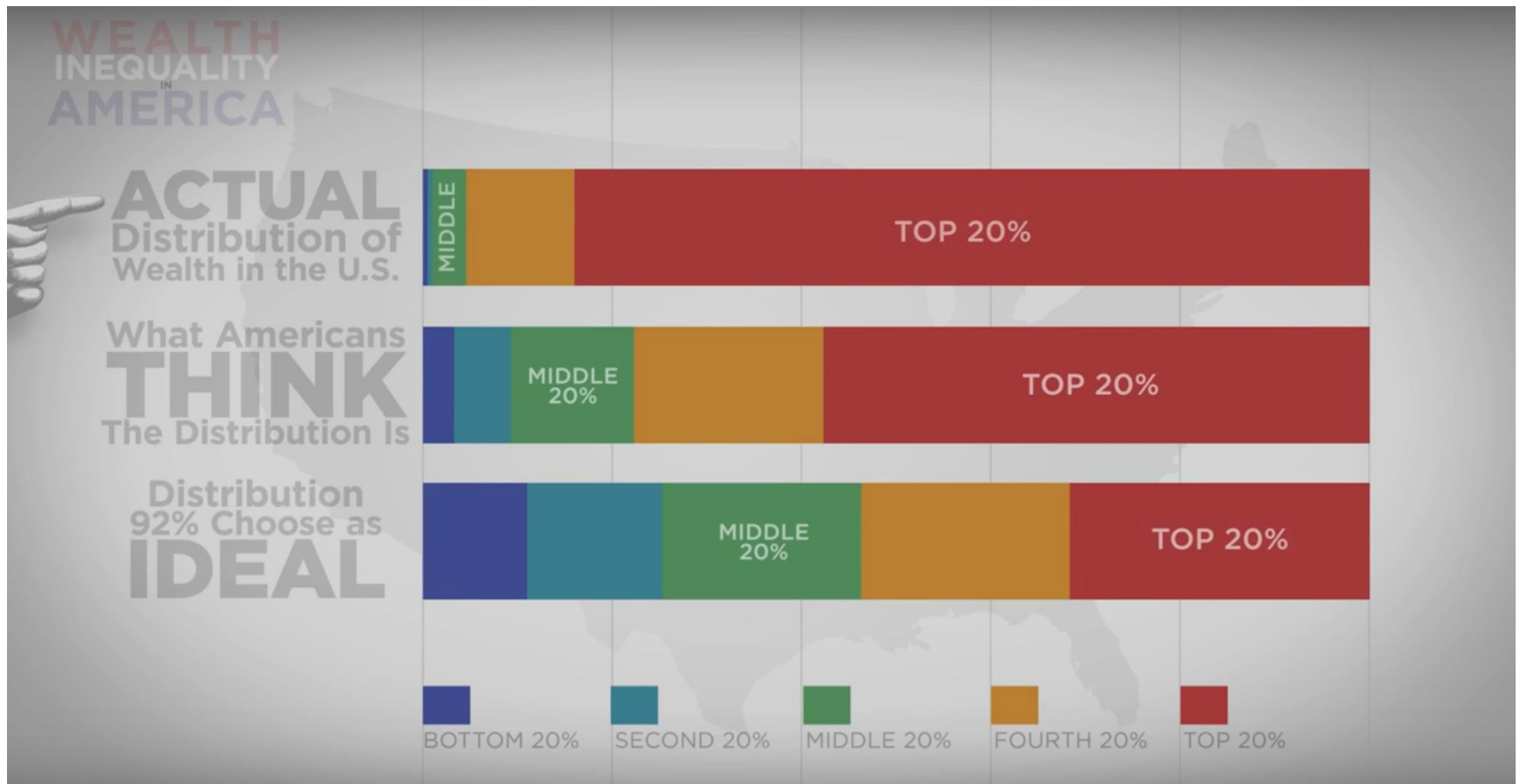
Storytelling ≠ making something up

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



Tools

Tools

visualising data

HOME BLOG RESOURCES TRAINING BOOK ABOUT

12 NOV | POLICYVIZ PODCAST: EPISODE 137, INFOPLUS CONFERENCE >>

DATA HANDLING CHARTING PROGRAMMING MULTIVARIATE MAPPING WEB-BASED SPECIALIST COLOUR

0 TO 255 ABBYY ABLE2EXTRACT ADIOMA ADOBE AFTER EFFECTS ADOBE ANIMATE

ADOBE COLOR ADOBE EDGE ADOBE ILLUSTRATOR AESOP STORY ENGINE AFFINITY DESIGNER AIZHTML

ALTERYX ANIMAPS ANYCHART APPS FOR EXCEL ARBORJS ARCGIS

AUTODRAW AXURE BALSAMIQ BATCHGEO BEAKER BERTIFIER

D3

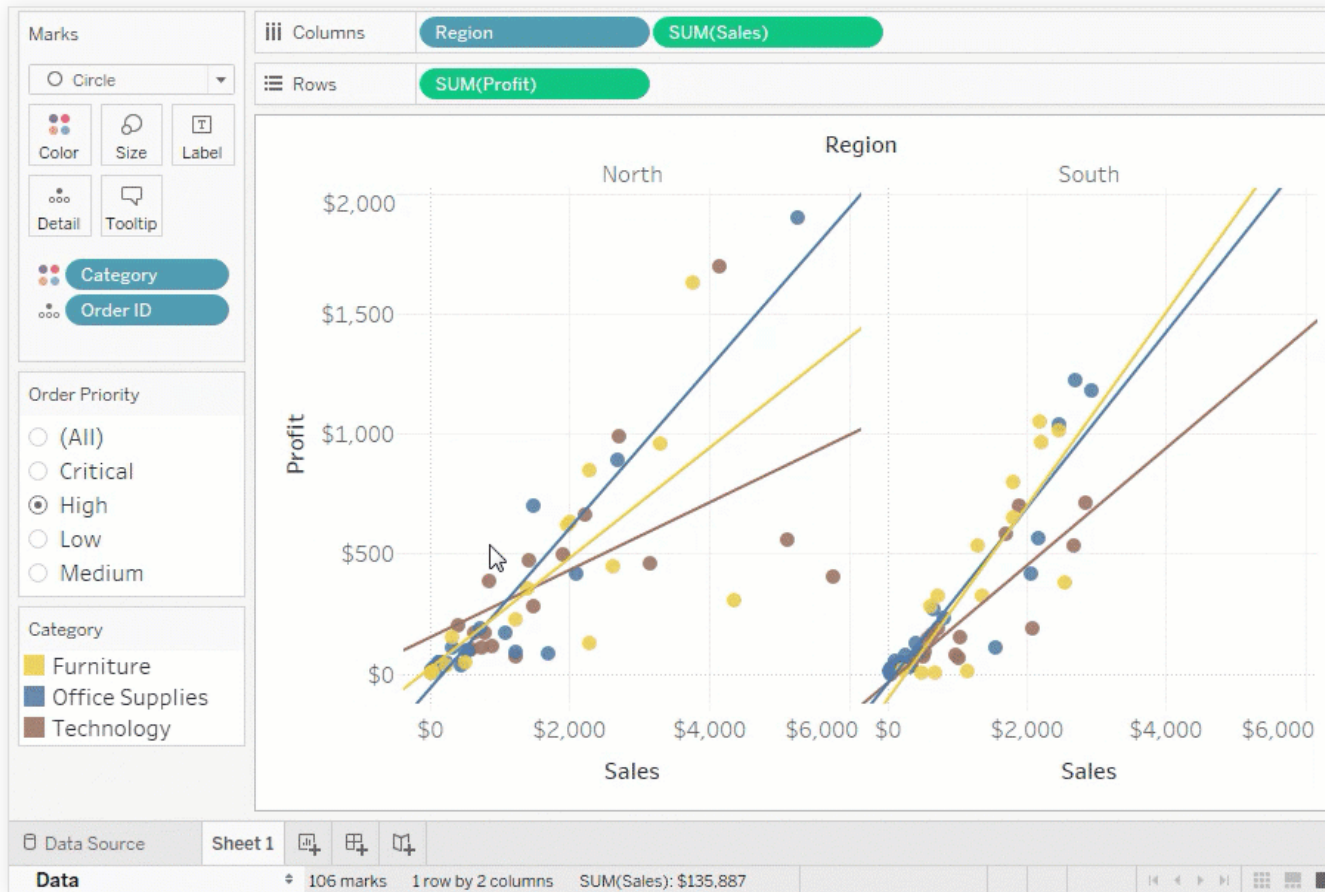
Data-Driven Documents



A JavaScript library

Emphasis on interactivity

Tableau Public



Does not
require
programming
skills

Visualizations
created with the
free version are
public