

DIAGRAM OF THE CAUSES OF MORTALITY  
IN THE ARMY IN THE EAST.

2.  
APRIL 1855 to MARCH 1856.

1.  
APRIL 1854 to MARCH 1855.

# Data Visualization Beyond a simple bar chart

November 2, 2022

Evgheni Polisciuc

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov. 1854 marks the boundary



# **Part 1 - Introduction**

## **Part 2 – Notebook**

<https://bit.ly/3DPPFAP>



# Evgheni Polisciuc

evgheni@dei.uc.pt

Teacher and researcher at DEI/UC (CDV)

[cdv.dei.uc.pt](http://cdv.dei.uc.pt/people/)

The screenshot shows a website for the Computational Design & Visualization Lab (CDV). The header includes the lab's logo (a circular pattern of dots), the title "COMPUTATIONAL DESIGN & VISUALIZATION LAB.", and navigation links for "About", "People", "Publications", "Projects", "News", and "Snapshots". The main content area displays a grid of 21 portrait photographs of lab members. Each portrait includes the name and title of the individual. A red box highlights the portrait of Evgheni Polisciuc, PhD, located in the third row, second column. The individuals in the grid are: Penousal Machado (Scientific Director), João Bicker (Art Director), Filipe Assunção (PhD), Ana Boavida (PhD), João Nuno Correia (PhD), João Miguel Cunha (PhD), António Leitão (PhD), Pedro Martins (PhD), Tiago Martins (PhD), Catarina Maçãs (PhD), Evgheni Polisciuc (PhD) (highlighted), Artur Rebelo (PhD), Francisco Baeta (PhD Student), João Couceiro e Castro (PhD Student), António Cruz (PhD Student), Luis Gonçalo (PhD Student), Daniel Lopes (PhD Student), Solange Margarido (PhD Student), Jéssica Parente (PhD Student), Sérgio Rebelo (PhD Student), and Ana Rodrigues (PhD Student).

People « CDV Lab

cdv.dei.uc.pt/people/

COMPUTATIONAL DESIGN & VISUALIZATION LAB.

About People Publications Projects News Snapshots

Write Something

Penousal Machado  
Scientific Director

João Bicker  
Art Director

Filipe Assunção  
PhD

Ana Boavida  
PhD

João Nuno Correia  
PhD

João Miguel Cunha  
PhD

António Leitão  
PhD

Pedro Martins  
PhD

Tiago Martins  
PhD

Catarina Maçãs  
PhD

Evgheni Polisciuc  
PhD

Artur Rebelo  
PhD

Francisco Baeta  
PhD Student

João Couceiro e Castro  
PhD Student

António Cruz  
PhD Student

Luis Gonçalo  
PhD Student

Daniel Lopes  
PhD Student

Solange Margarido  
PhD Student

Jéssica Parente  
PhD Student

Sérgio Rebelo  
PhD Student

Ana Rodrigues  
PhD Student

# Evgheni Polisciuc

evgheni@dei.uc.pt

Teacher and researcher at DEI/UC (CDV)

[cdv.dei.uc.pt](http://cdv.dei.uc.pt)

The screenshot shows the homepage of the CDV Lab website. At the top, there is a navigation bar with links for About, People, Publications, Projects, News, and Snapshots. On the left, there is a logo consisting of a circular pattern of dots and the text "COMPUTATIONAL DESIGN & VISUALIZATION LAB.". Below the logo, there are several project cards:

- Field of Leaves: An Interactive Installation Sprouting Leaves Out of Public Money** (19/07/2022): A large image of a digital installation where leaves sprout from a network of nodes.
- We are hiring!** (20/06/2022): A photo of a group of people sitting around a table, with one person holding a sign that says "Artificial Intelligence!"
- Qual é o teu poster?** (31/05/2022): A purple-themed card for an informal poster session at the University of Coimbra on June 17th.
- EvoDesigner: Towards aiding creativity in graphic design** (26/04/2022): A grid of small images showing different graphic design prototypes.
- TGPGAN: An Evolutionary approach to Generative Adversarial Networks** (19/04/2022): A card featuring a 10x10 grid of numbers from 0 to 9.
- CDV lab projects exhibited on the National Day of Historic Centres** (25/03/2022): A card showing a collection of abstract, geometric posters.
- ATOVis—a Visualization Tool for the Detection of Financial Fraud** (27/04/2022): A card featuring a circular visualization with red and grey segments.
- Metamorphosis: Experiments in the design of interactive moving posters** (17/03/2022): A card showing a grid of letters and shapes.

On the right side of the page, there is a search bar with the placeholder "Write Something" and a sidebar with various icons.

# Information overload

**The data is produced  
faster than analyzed  
and used!**

We need tools to be able to **visualize** the data,  
**communicate** the data, and **utilize it effectively**.

# Information Visualization

**"COMPUTER-BASED VISUALIZATION SYSTEMS PROVIDE VISUAL REPRESENTATIONS OF DATASETS DESIGNED TO HELP PEOPLE CARRY OUT TASKS MORE EFFECTIVELY."**

Munzner, 2014

# Visualization can make complex problems trivial



# Only if designed properly!



# Why visualization?

## Replace cognition with perception

Which country has the most Hindus?

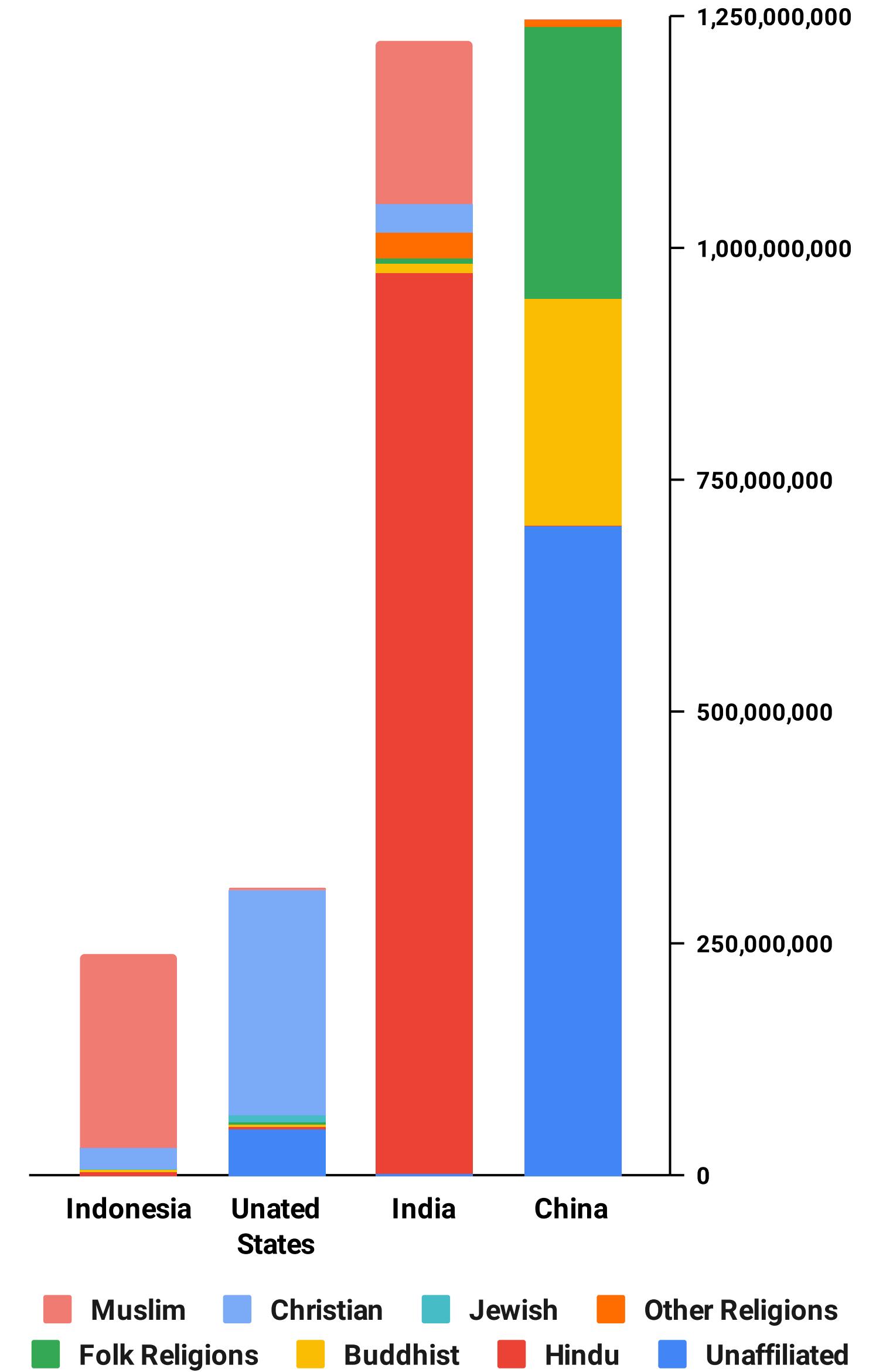
Country	Religion	People
China	Unaffiliated	700,680,000
China	Hindu	20,000
China	Buddhist	244,130,000
China	Folk Religions	294,320,000
China	Other Religions	9,080,000
China	Jewish	0
India	Christian	31,130,000
India	Muslim	176,190,000
India	Unaffiliated	870,000
India	Hindu	973,750,000
India	Buddhist	9,250,000
India	Folk Religions	5,840,000
India	Other Religions	27,650,000
India	Jewish	10,000
United States	Christian	243,060,000
United States	Muslim	2,770,000
United States	Unaffiliated	50,980,000
United States	Hindu	1,790,000
United States	Buddhist	3,570,000
United States	Folk Religions	630,000
United States	Other Religions	1,900,000
United States	Jewish	5,690,000
Indonesia	Christian	23,660,000
Indonesia	Muslim	209,120,000
Indonesia	Unaffiliated	240,000
Indonesia	Hindu	4,050,000
Indonesia	Buddhist	1,720,000
Indonesia	Folk Religions	750,000
Indonesia	Other Religions	340,000
Indonesia	Jewish	0

# Why visualization?

## Replace cognition with perception

Which country has the most Hindus?

Country	Religion	People
China	Unaffiliated	700,680,000
China	Hindu	20,000
China	Buddhist	244,130,000
China	Folk Religions	294,320,000
China	Other Religions	9,080,000
China	Jewish	0
India	Christian	31,130,000
India	Muslim	176,190,000
India	Unaffiliated	870,000
India	Hindu	973,750,000
India	Buddhist	9,250,000
India	Folk Religions	5,840,000
India	Other Religions	27,650,000
India	Jewish	10,000
United States	Christian	243,060,000
United States	Muslim	2,770,000
United States	Unaffiliated	50,980,000
United States	Hindu	1,790,000
United States	Buddhist	3,570,000
United States	Folk Religions	630,000
United States	Other Religions	1,900,000
United States	Jewish	5,690,000
Indonesia	Christian	23,660,000
Indonesia	Muslim	209,120,000
Indonesia	Unaffiliated	240,000
Indonesia	Hindu	4,050,000
Indonesia	Buddhist	1,720,000
Indonesia	Folk Religions	750,000
Indonesia	Other Religions	340,000
Indonesia	Jewish	0



# Why visualization?

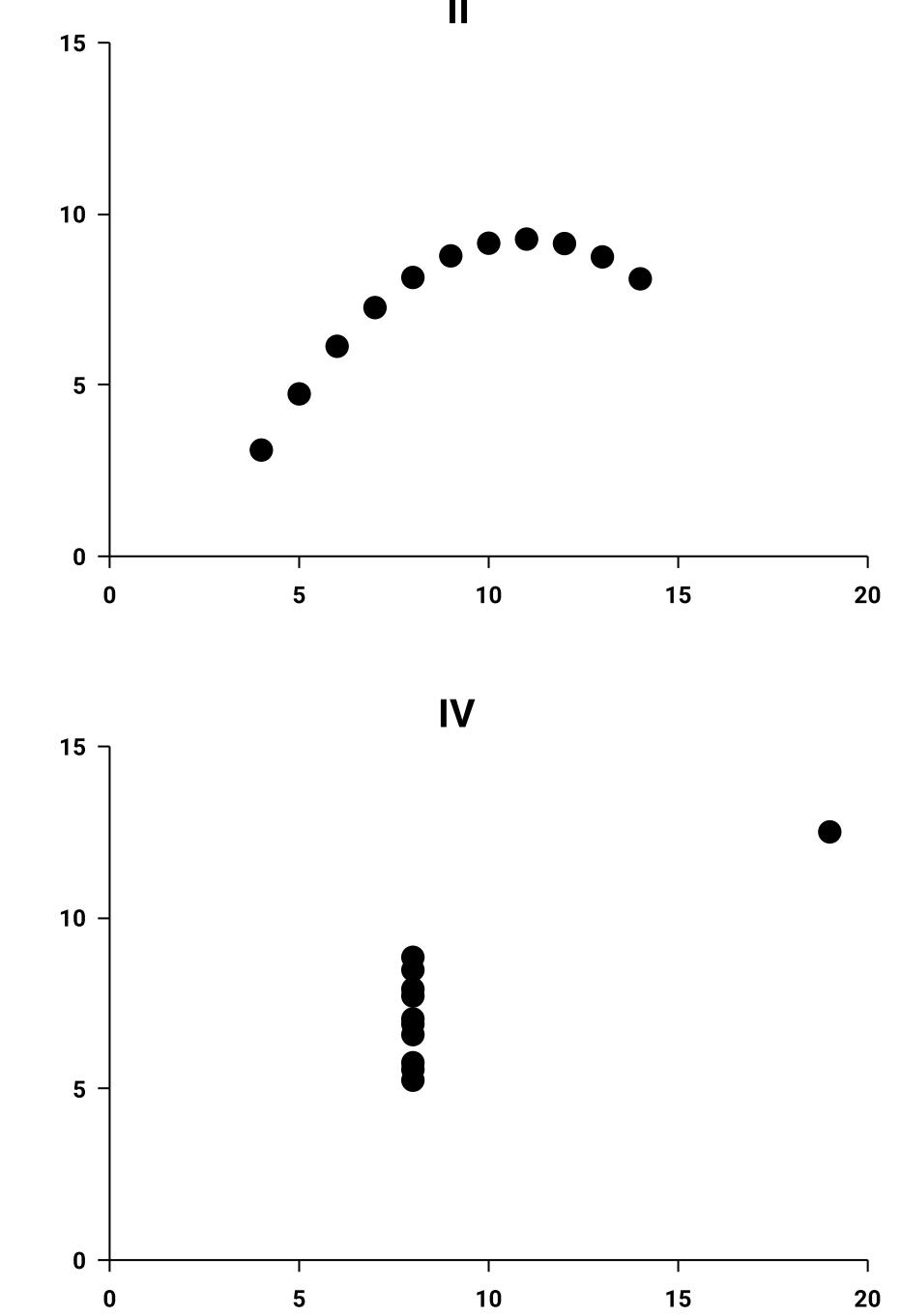
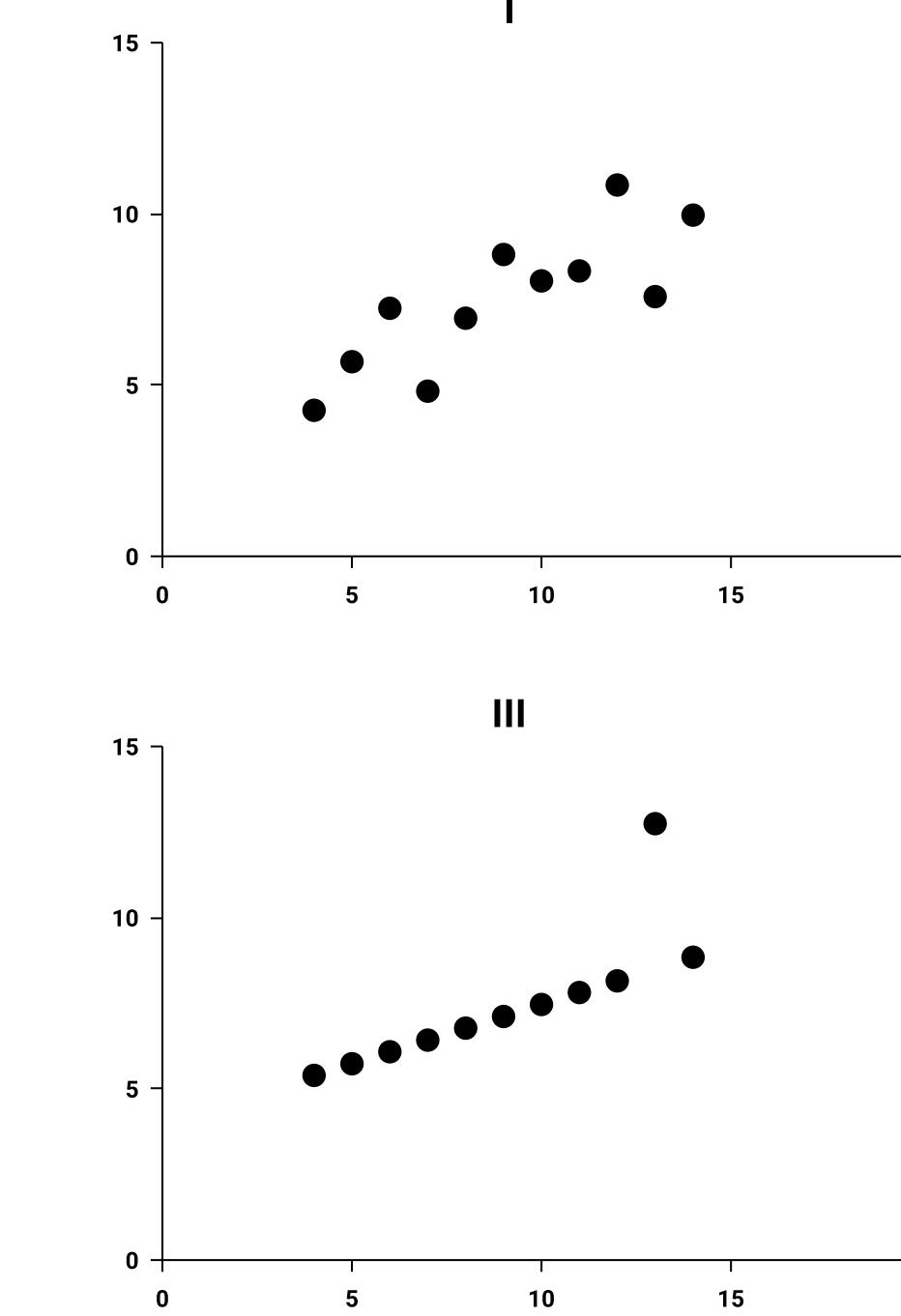
Reveal information

**Visualization can reveal information which conventional statistical approaches may hide.**

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
9		9		9		9	
10.0		10.0		10.0		10.0	
7.50		7.50		7.50		7.50	
3.75		3.75		3.75		3.75	

**Identical statistics**

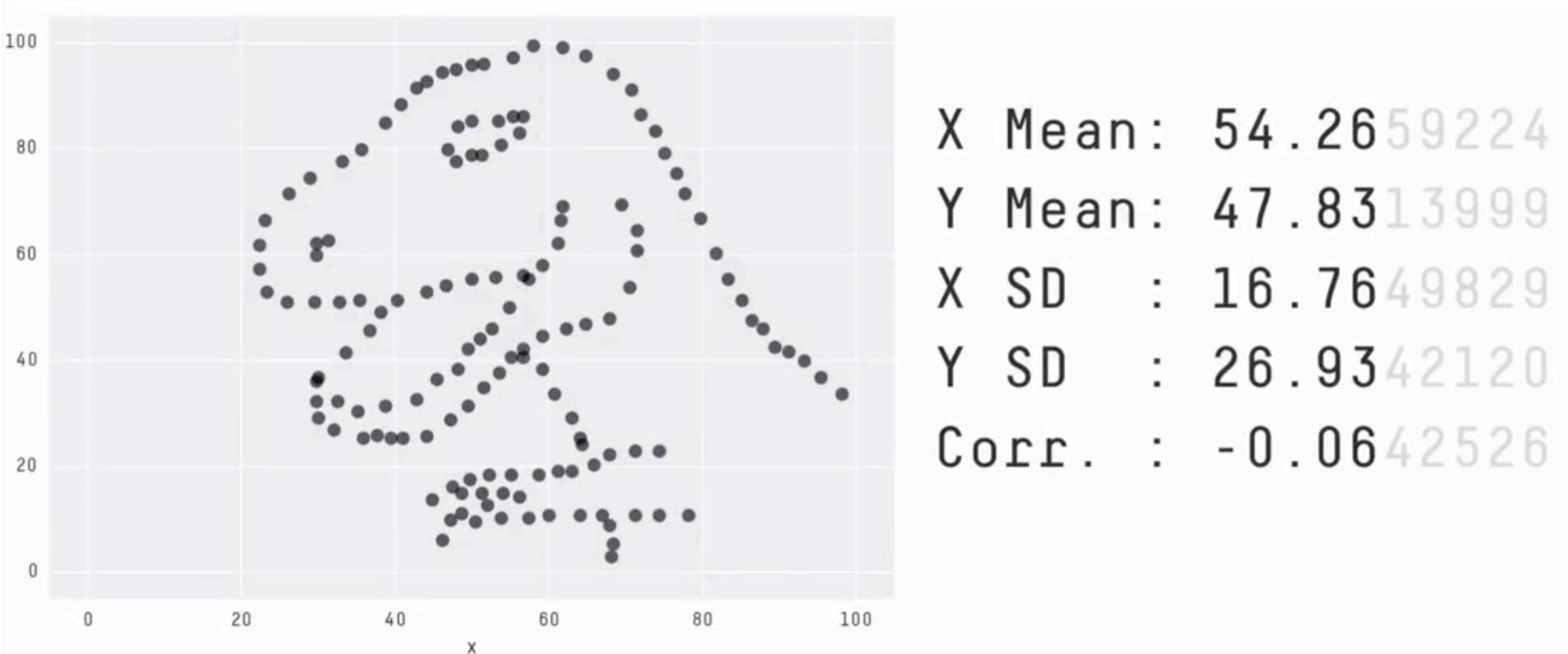
x mean	9.0
x variance	10.0
y mean	7.50
y variance	3.75
x/y corelation	0.816
N	11



**Anscombe's quartet** (Francis Anscombe's 1973 dataset)

# Why visualization?

Reveal information



# Why visualization?

**Human visual system is powerful**

How many 3s?

**1234987205987520198750640391871  
2310984712093875431098712398751  
0934856041487134098710451987123  
9058610938456480193845609871345**

# Why visualization?

Human visual system is powerful

How many 3s?

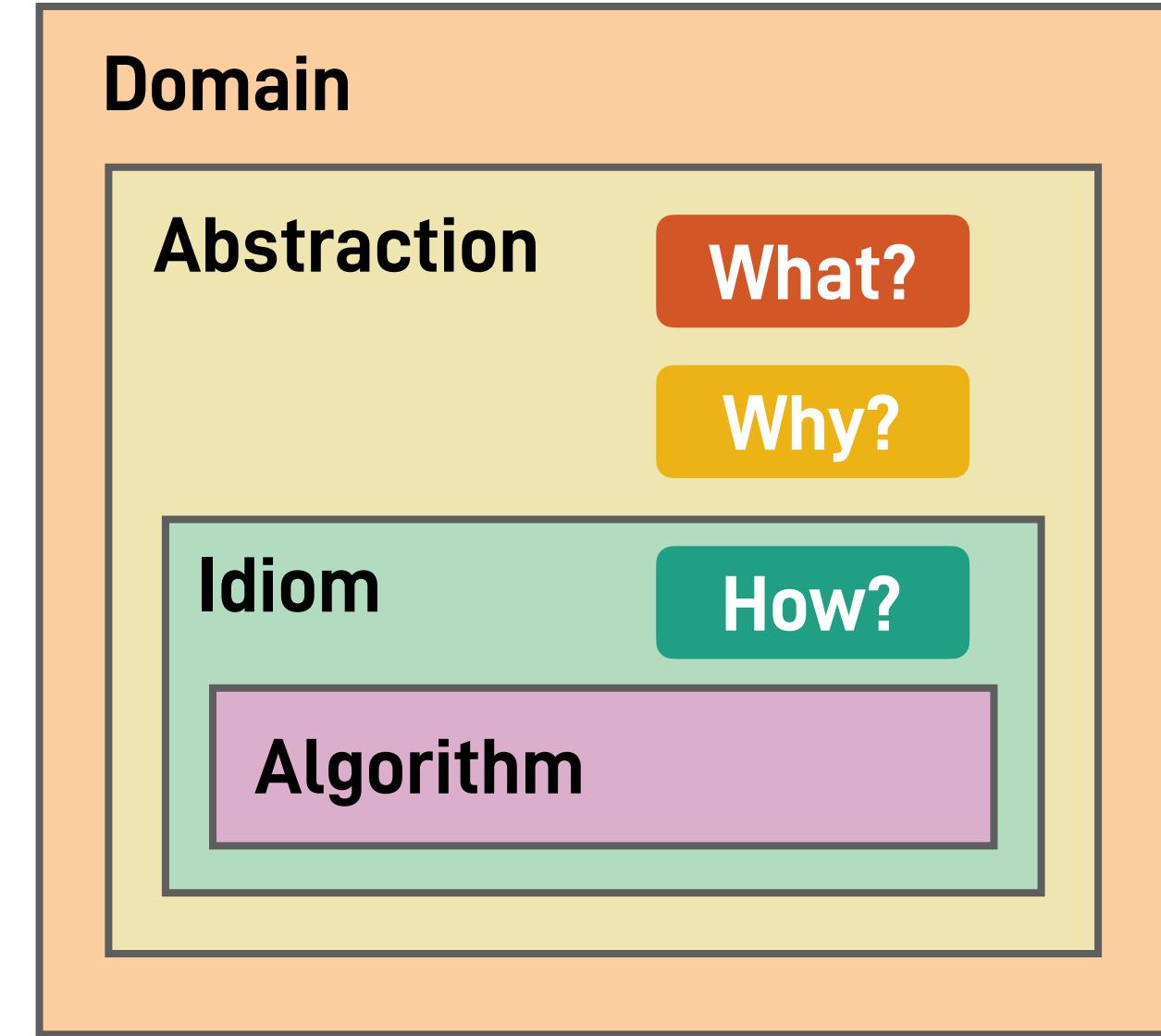
1234987205987520198750640391871  
2310984712093875431098712398751  
0934856041487134098710451987123  
9058610938456480193845609871345

# **What, Why, and How?**

# Nested model

## Four level visualization design

1. **domain** situation
  - Who are the target users?
2. **abstraction**
  - Translate from the language of domain to vocabulary of visualization
  - **What** data is shown? (data abstraction)
  - **Why** is the user using the visualization? (tasks abstraction)
3. **idiom**
  - **How** is the visualization constructed in terms of design choices?
  - Visual encoding idiom (how is it drawn)
  - Interaction idiom (how can it be manipulated)
4. **algorithm**
  - Efficient computation (visualization runs smoothly)



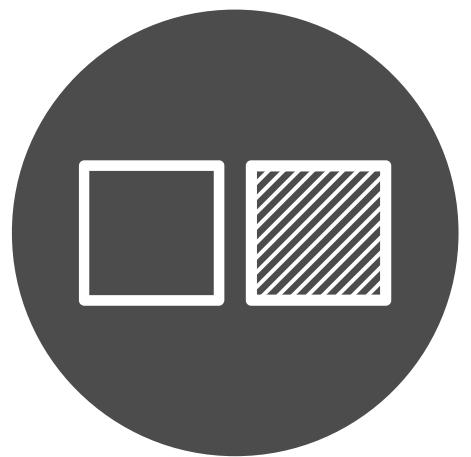
# Let's focus on "How?"

Design visual encoding and interaction

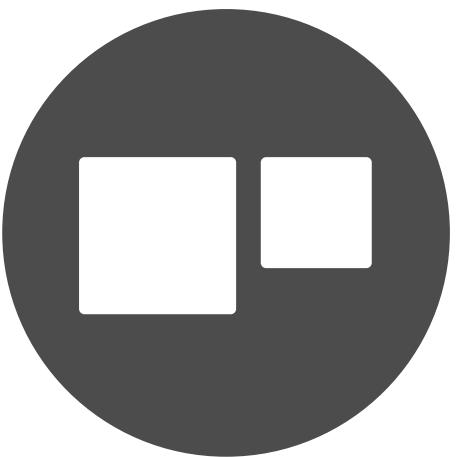
Idiom

How?

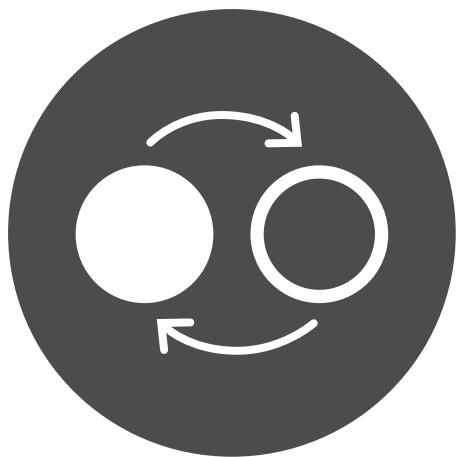
## But instead on How to choose a representation



Comparison



Proportion



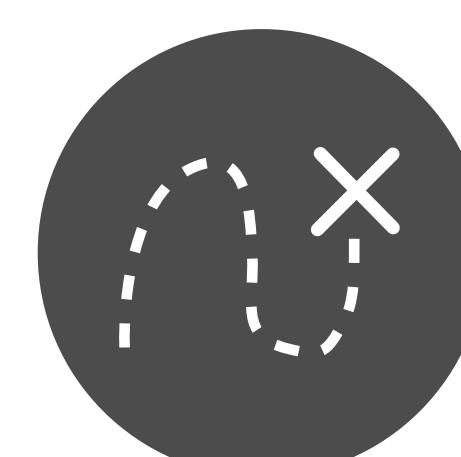
Relationship



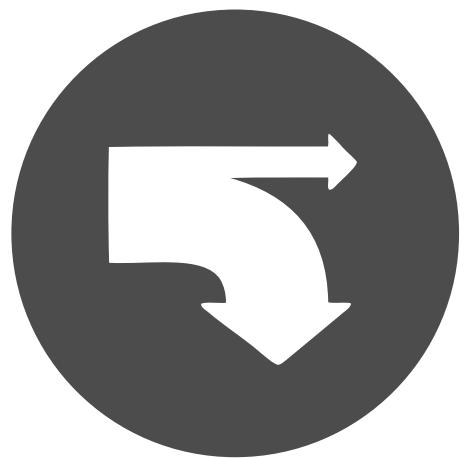
Hierarchy



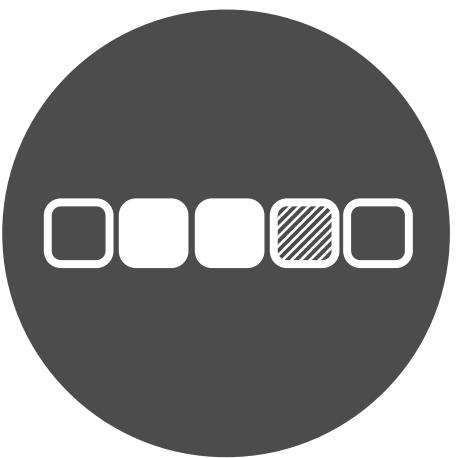
Part-to-whole



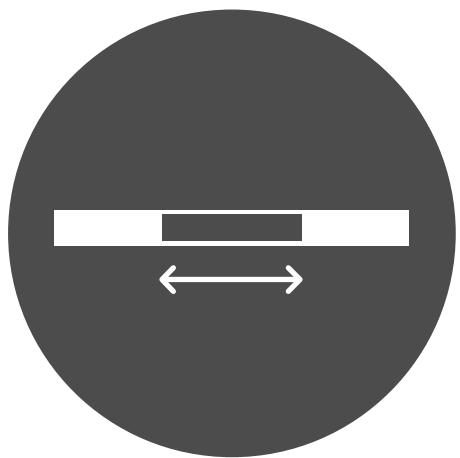
Location



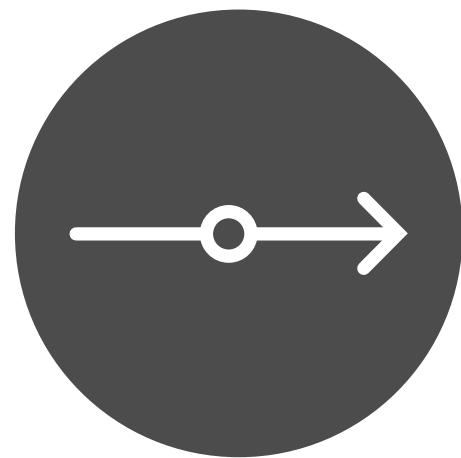
Flow



Pattern



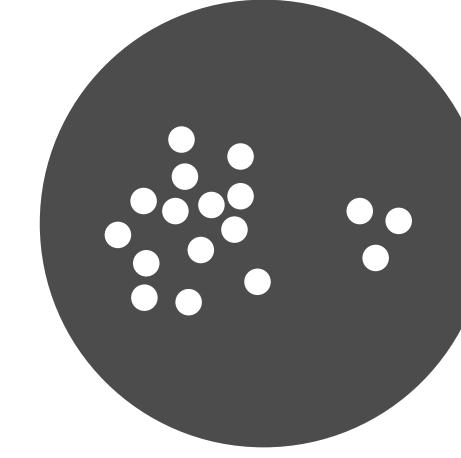
Range



Change over time



Text analysis



Distribution

## Deviation

Emphasise variations (>0) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

**Example FT uses**  
Trade surplus/deficit, climate change

## Correlation

Show the relationship between two or more variables. Be mindful that unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

**Example FT uses**  
Inflation and unemployment, income and life expectancy

## Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

**Example FT uses**  
Wealth, deprivation, league tables, constituency election results

## Distribution

Show values in a dataset and how often they occur. The shape or 'skew' of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

**Example FT uses**  
Income distribution, population (age/sex) distribution, revealing inequality

## Change over Time

Give emphasis to changing trends. These can be short (immediate), movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

**Example FT uses**  
Share price movements, economic time series, sectoral changes in a market

## Magnitude

Show size comparisons. These can be relative (just being able to see larger/biggest) or absolute (need to see fine differences). Usually these show a 'counted' number (for example barrels, dollars or people) rather than a calculated rate or per cent.

**Example FT uses**  
Commodity production, market capitalisation, volumes in general

## Spatial

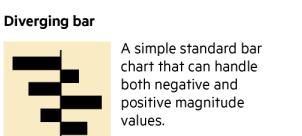
Apart from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

**Example FT uses**  
Flood delivery, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

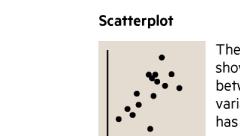
## Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

**Example FT uses**  
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.



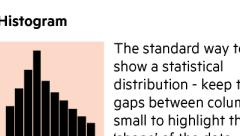
A simple standard bar chart that can handle both negative and positive magnitude values.



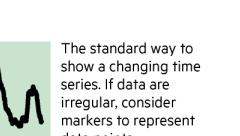
The standard way to show the relationship between two continuous variables, each of which has its own axis.



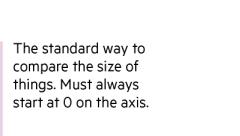
Standard bar charts display the ranks of values much more easily when sorted into order.



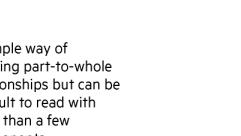
The standard way to show a statistical distribution. Gap the bars between columns small to highlight the shape of the data.



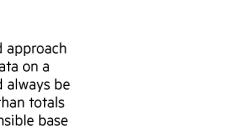
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.



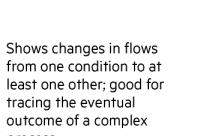
The standard way to compare the size of things. Must always start at 0 on the axis.



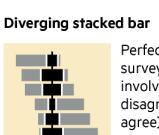
A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.



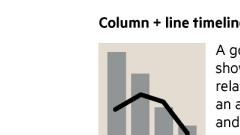
The standard approach for putting data on a map should always be rates rather than totals and use a sensible base geography.



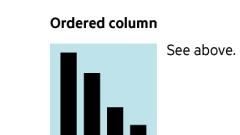
Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.



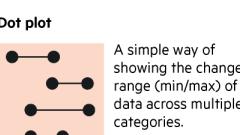
Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).



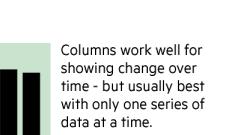
A good way of showing the relationship between an amount (columns) and a rate (line).



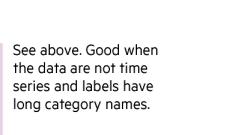
See above.



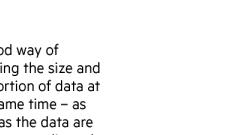
A simple way of showing change or range (min/max) of data across multiple categories.



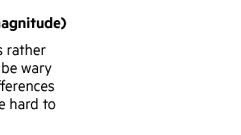
Columns work well for showing change over time, but usually best with only one series of data at a time.



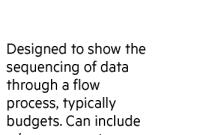
See above. Good when the data are not time series and labels have long category names.



A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.



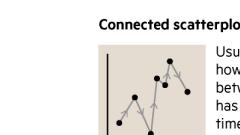
Use for totals rather than rates – be wary that small differences in data will be hard to see.



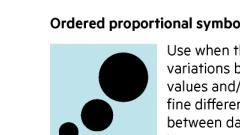
Designed to show the sequence of data through a flow process, typically budgets. Can include +/- components.



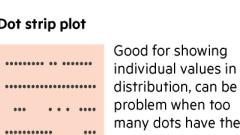
Splits a single value into two contrasting components (eg male/female).



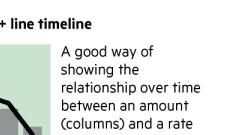
Usually used to show how the relationship between 2 variables has changed over time.



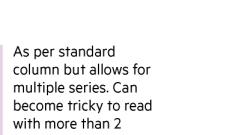
Use when there are big variations between values and/or seeing fine differences between data is not so important.



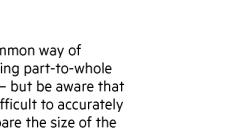
Good for showing individual values in a distribution, can be a problem when too many dots have the same value.



Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.



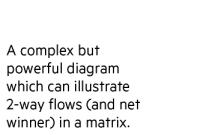
Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of the story.



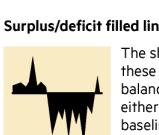
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.



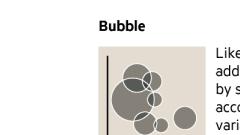
A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.



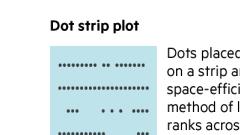
A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.



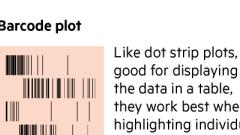
The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.



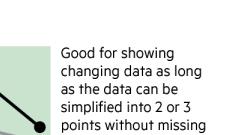
Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.



A good way of showing the pattern between 2 categories of data, less effective at showing fine differences in amounts.



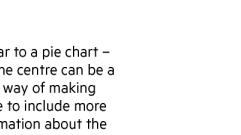
Perfect for showing how ranks have changed over time or vary between categories.



Summarise multiple distributions by showing the median (centre) and range of the data.



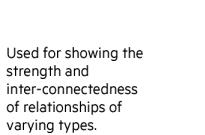
Use with care – these charts are good for showing change over time but seeing change in components can be very difficult.



A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.



Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).



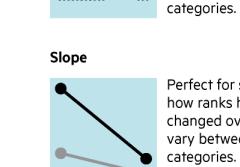
Converting each unit on a map to a regular and equal-sized shape – good for representing voting results with equal value.



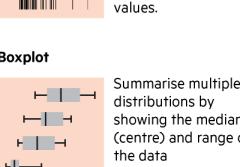
Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.



Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.



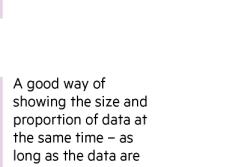
Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.



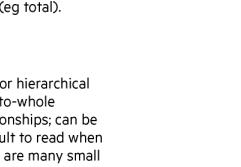
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple averages).



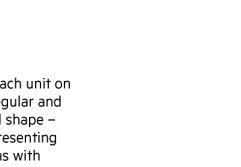
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.



Use to show the uncertainty in future projections – usually this grows the further forward to projection.



Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).



A hemicycle, often used for visualising parliamentary composition by number of seats.



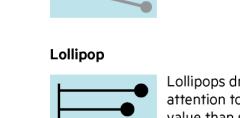
Good for showing % information, they work best when used with whole numbers and work well in small multiple layout form.



For displaying multiple distributions of data.



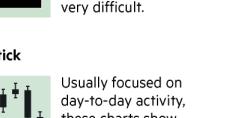
Like a regular line chart, best limited to a maximum of 3 or 4 datasets.



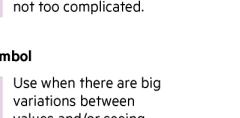
A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.



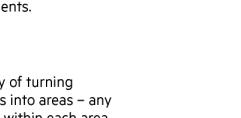
A good way of showing how unequal a distribution is – always cumulative frequency, x axis is always a measure.



A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.



A great way of showing temporal patterns (daily/weekly/monthly) – at the expense of showing precision in quantity.



Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).



A space-efficient way of showing value of multiple variables – but make sure they are organised in a way that makes sense to the reader.



Generally only used for schematic representation.



Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets.



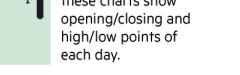
Great when date and duration are key elements of the story in the data.



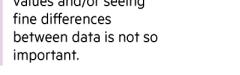
Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).



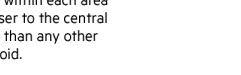
Presents time on the Y axis. Good for displaying parallel timelines that work especially well when scrolling on mobile.



An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.



Good for showing a measurement against the context of a target or performance range.



An alternative to bullet charts when being able to count data or highlight individual elements is useful.



Another alternative to the circle timeline for showing series where there are big variations in the data.



A type of area chart; use when seeing changes in proportions over time is more important than individual values.

# 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 graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe  
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

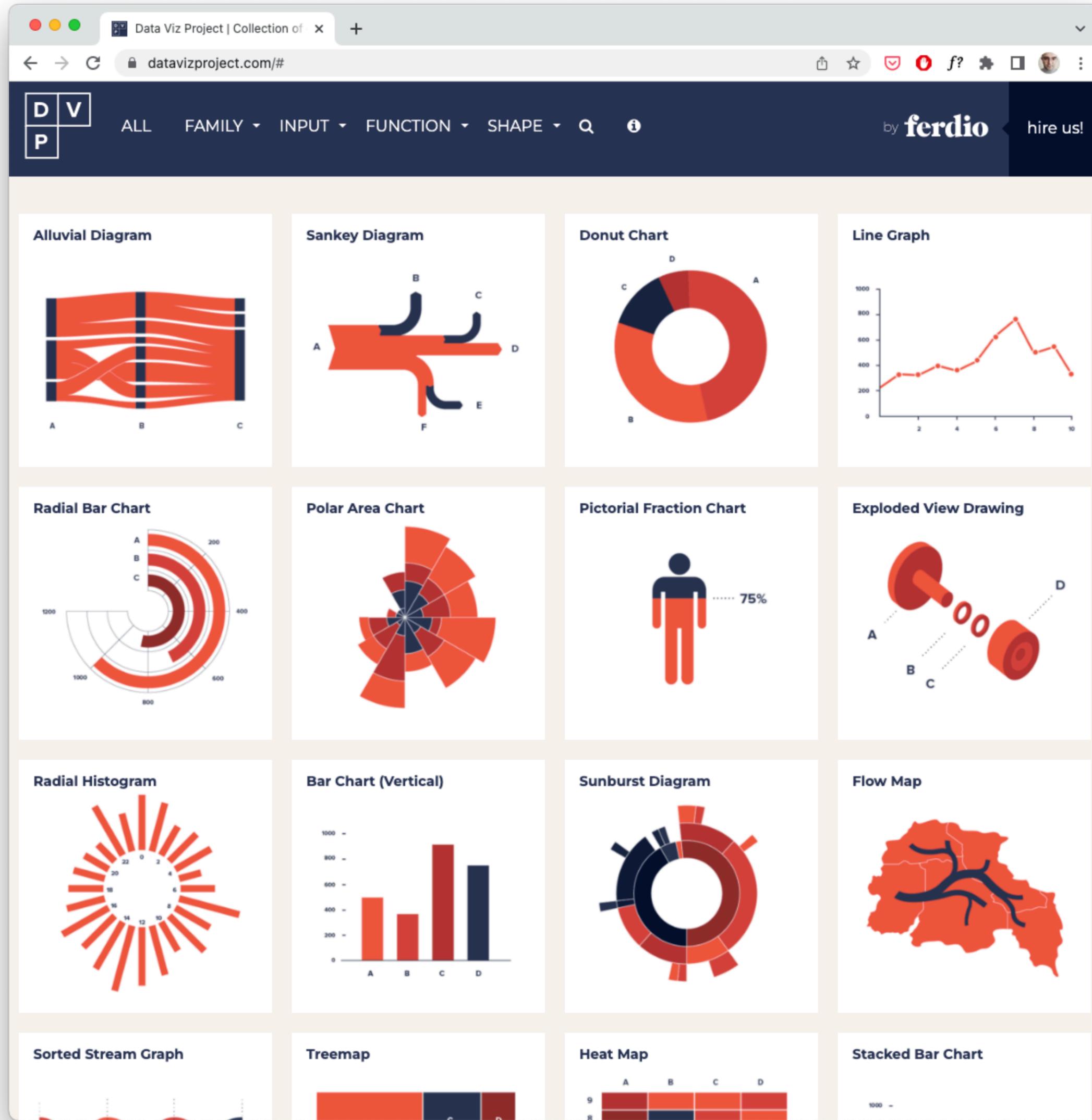


[ft.com/vocabulary](http://ft.com/vocabulary)

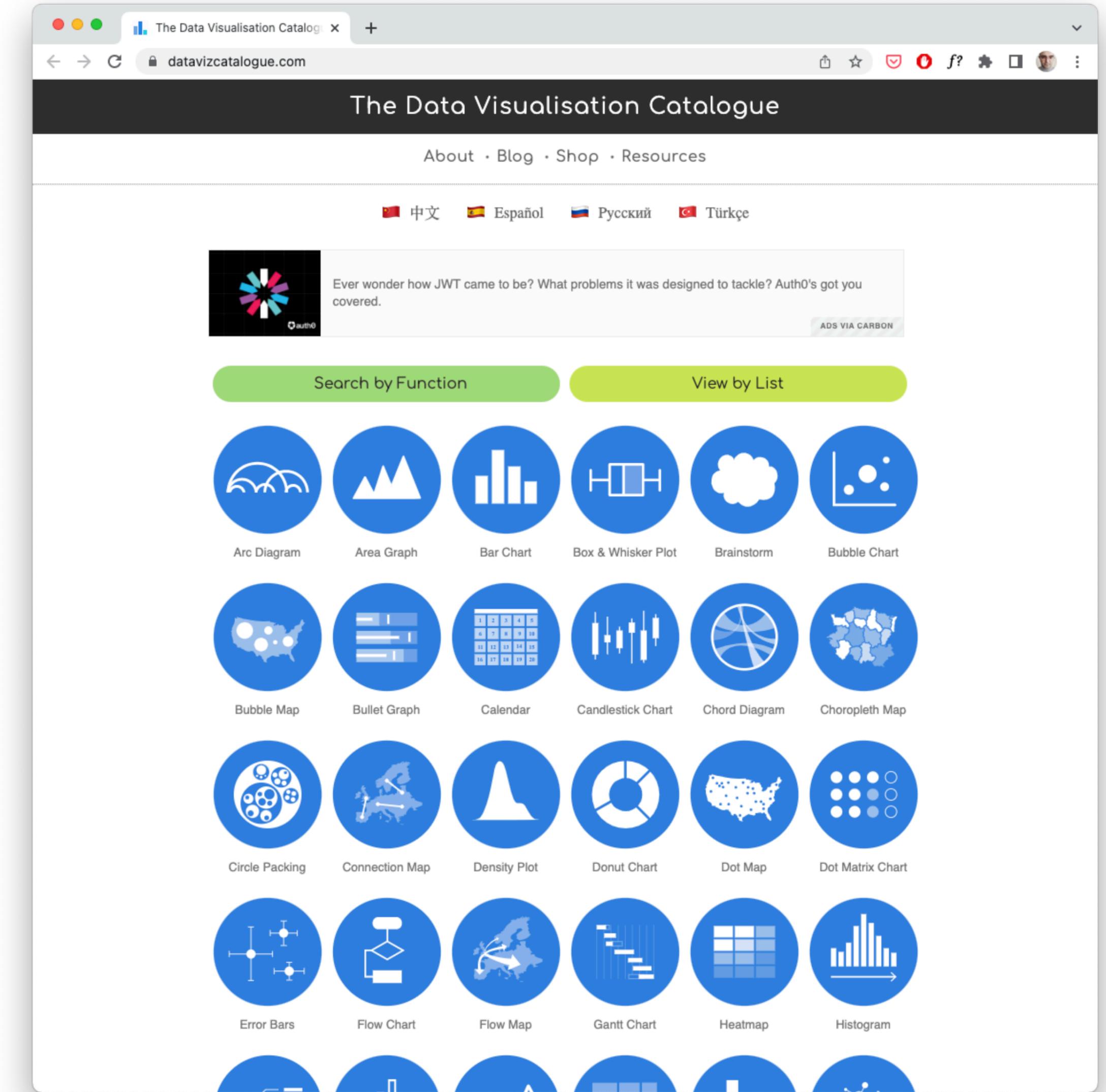


© Financial Times 2016-2020  
This work is licensed under a Creative Commons  
Attribution-ShareAlike 4.0 International License.

# Visualization taxonomies



<https://datavizproject.com>



<https://datavizcatalogue.com/>

# **Design considerations**

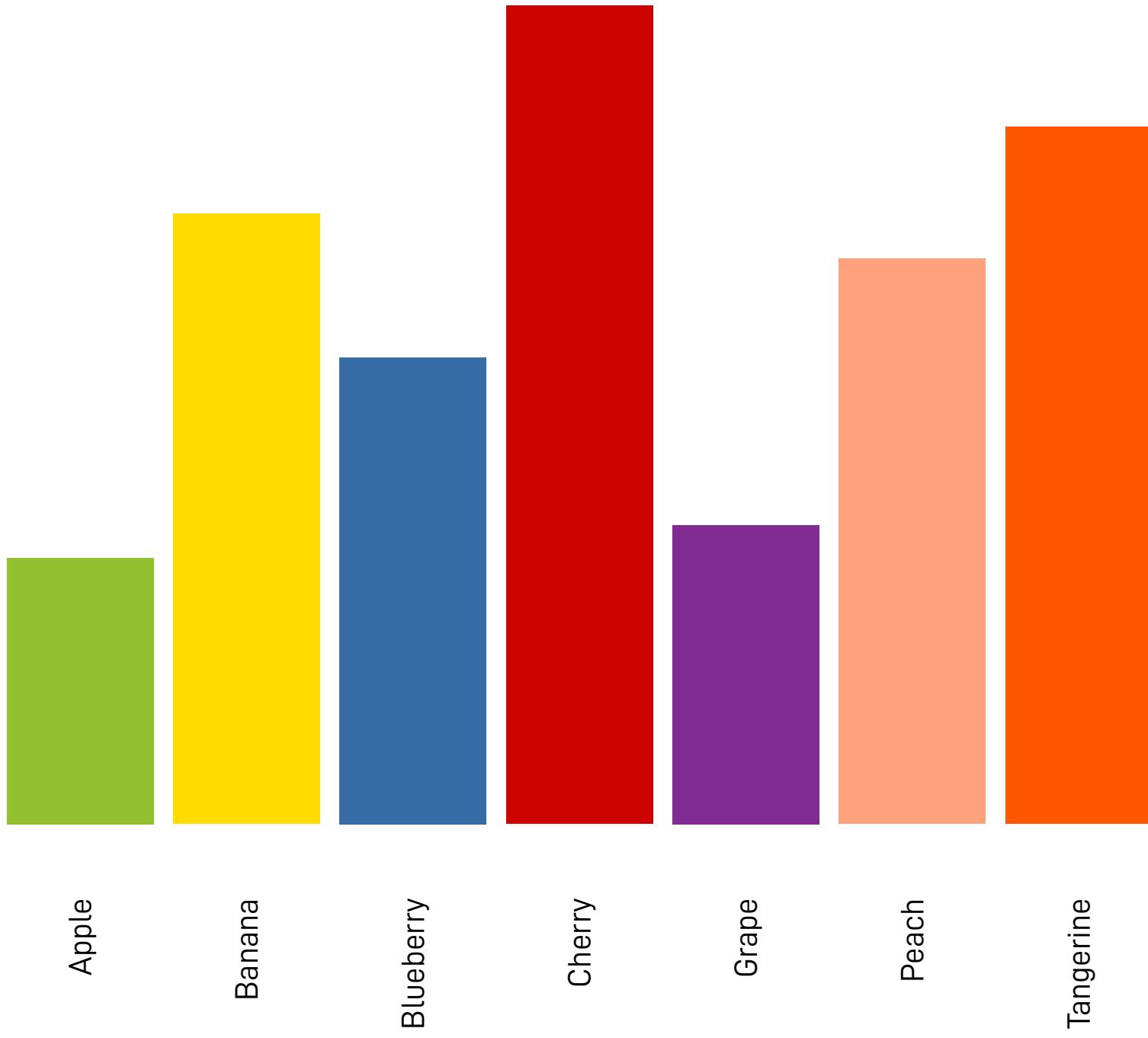
# **Text, labels, and legends**

**Use legends to describe encodings;  
Use text to describe what is in data;  
Do not repeat the numbers in the words.**

# Color

Use with care

**Use color when it corresponds to difference in data;  
Use color schemas that reflects what is in data.**



# Color

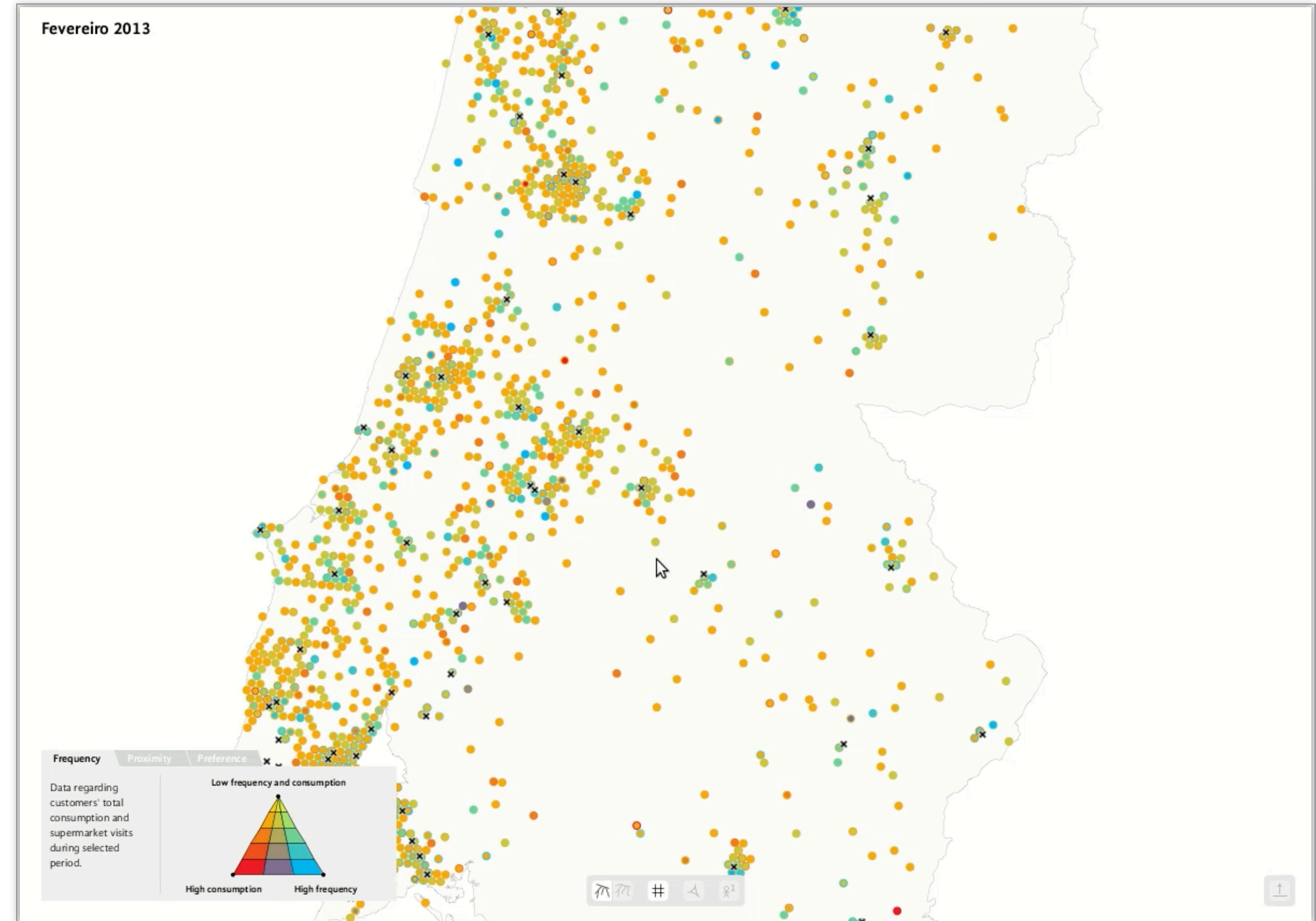
Use with care

Avoid **bad choice of color schemas**

Use **red and green with care**

# Color

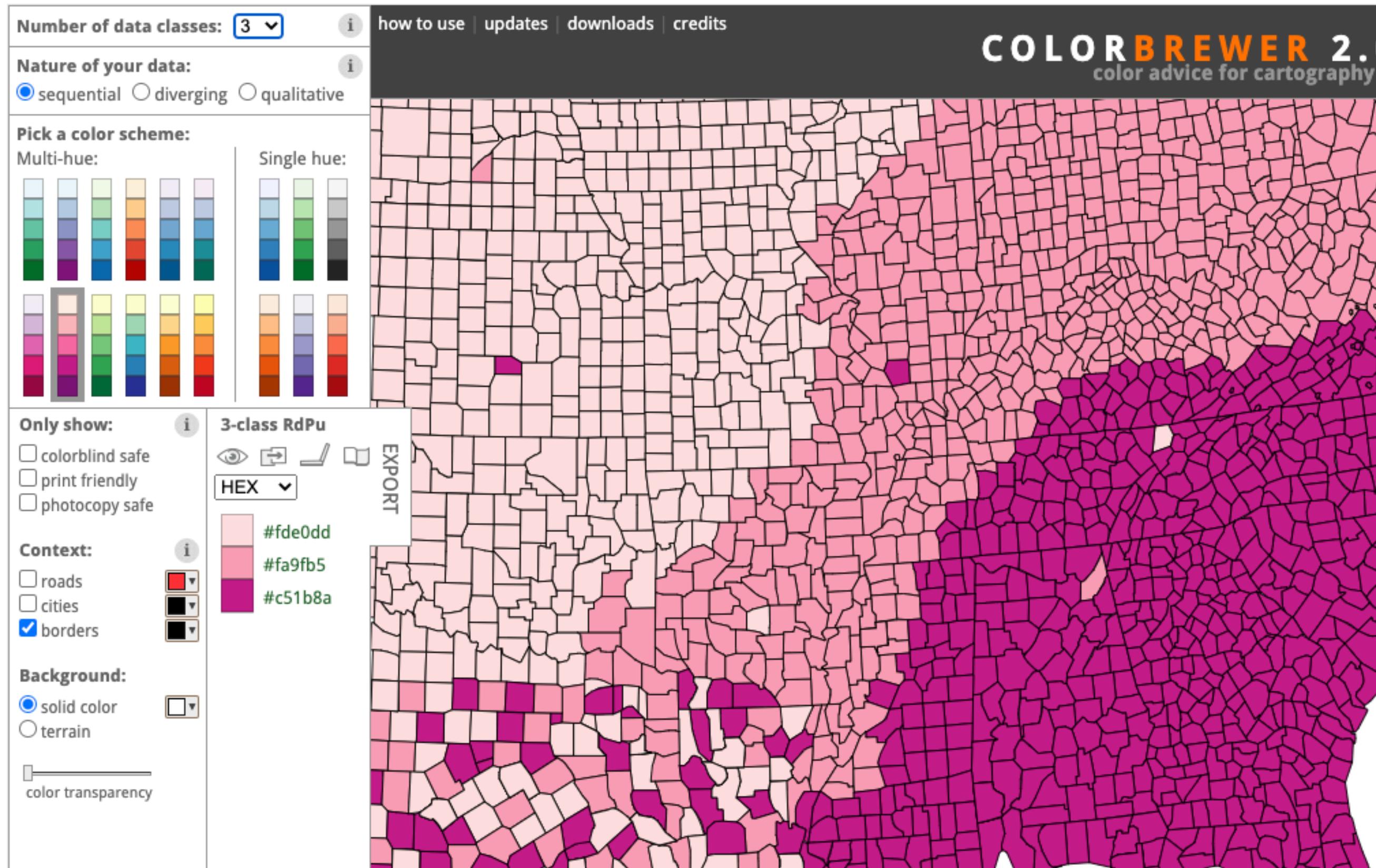
"Where customer shop"



# Color

Use with care

Avoid using colors confused by those with color blindness;  
Avoid using too many colors (max  $\pm 8$ ).



<https://colorbrewer2.org/>

# Typography

**The font family, size, color, direction affect legibility;  
Use sans-serif typefaces.**

Roboto Serif

**Sparrows**

Serifs

Roboto

**Sparrows**

**Georgia**

**Minion**

Serif fonts

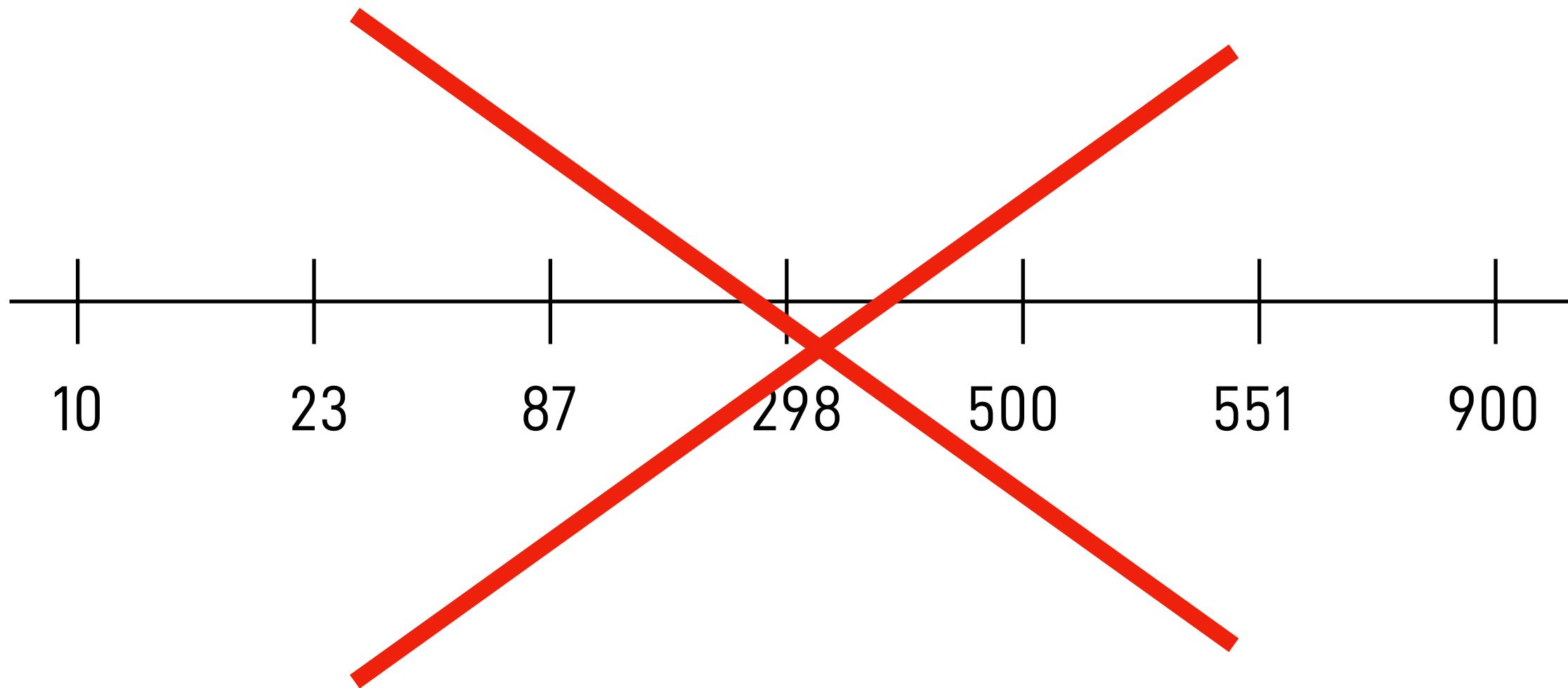
**Helvetica**

**Open Sans**

Sans-Serif fonts

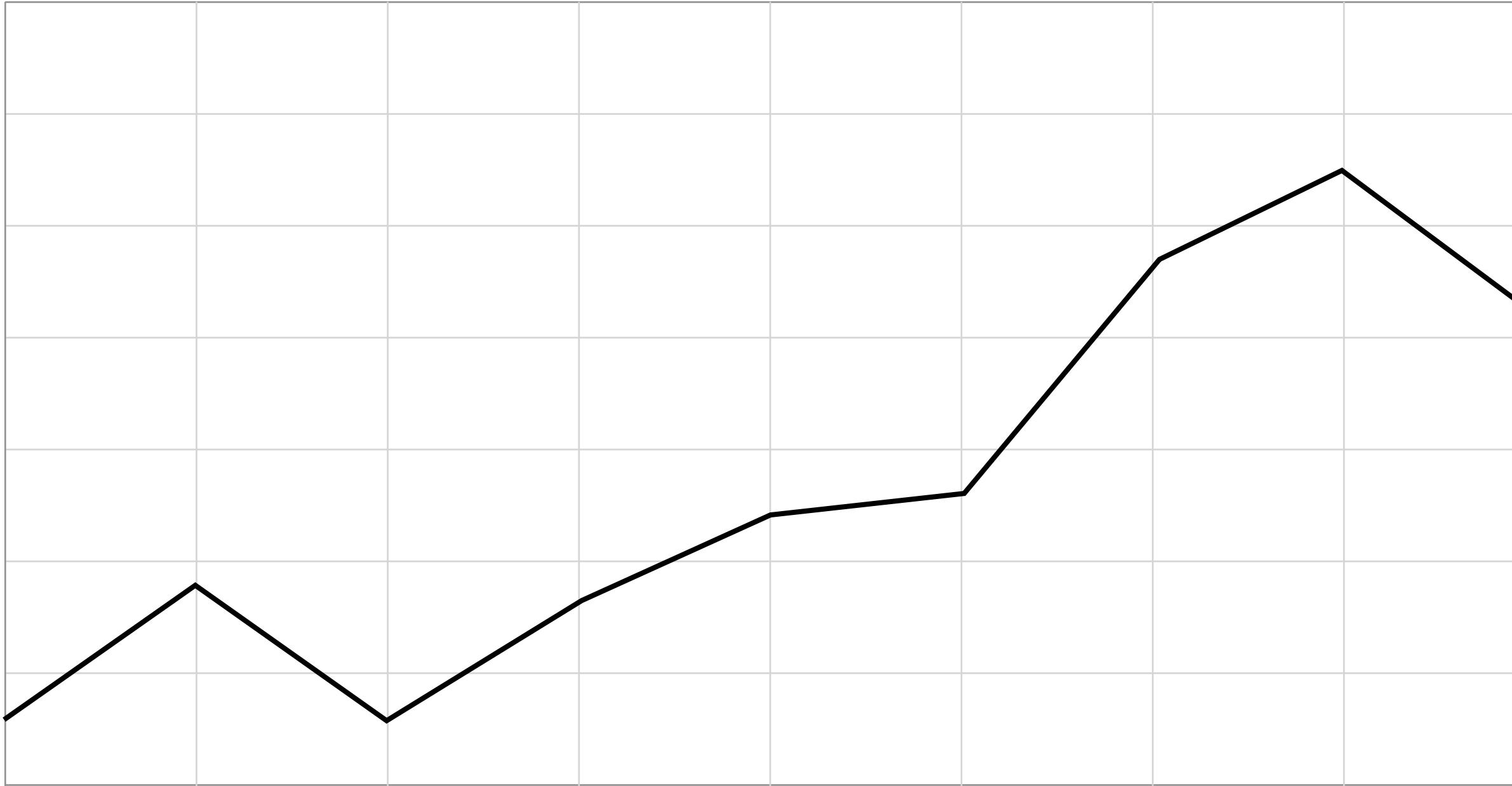
# Scale

**Use natural increments for scales (e.g., 10, 20, 30...);  
Avoid illogical grid scales (260, 273, 285, 298);  
Avoid high dense grids.**



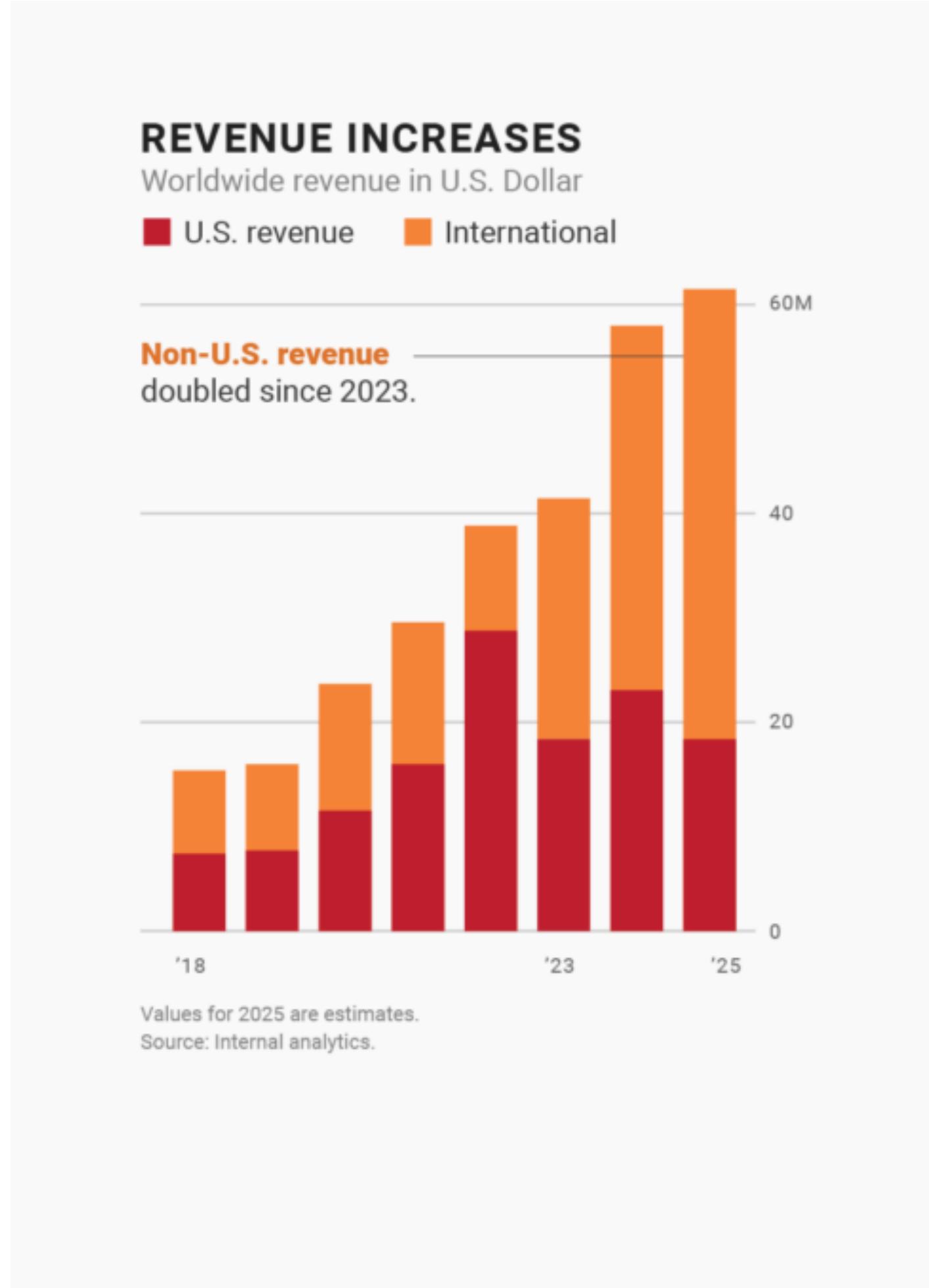
# Proportion

**Graphical elements look better together when their relative proportions are in balance**



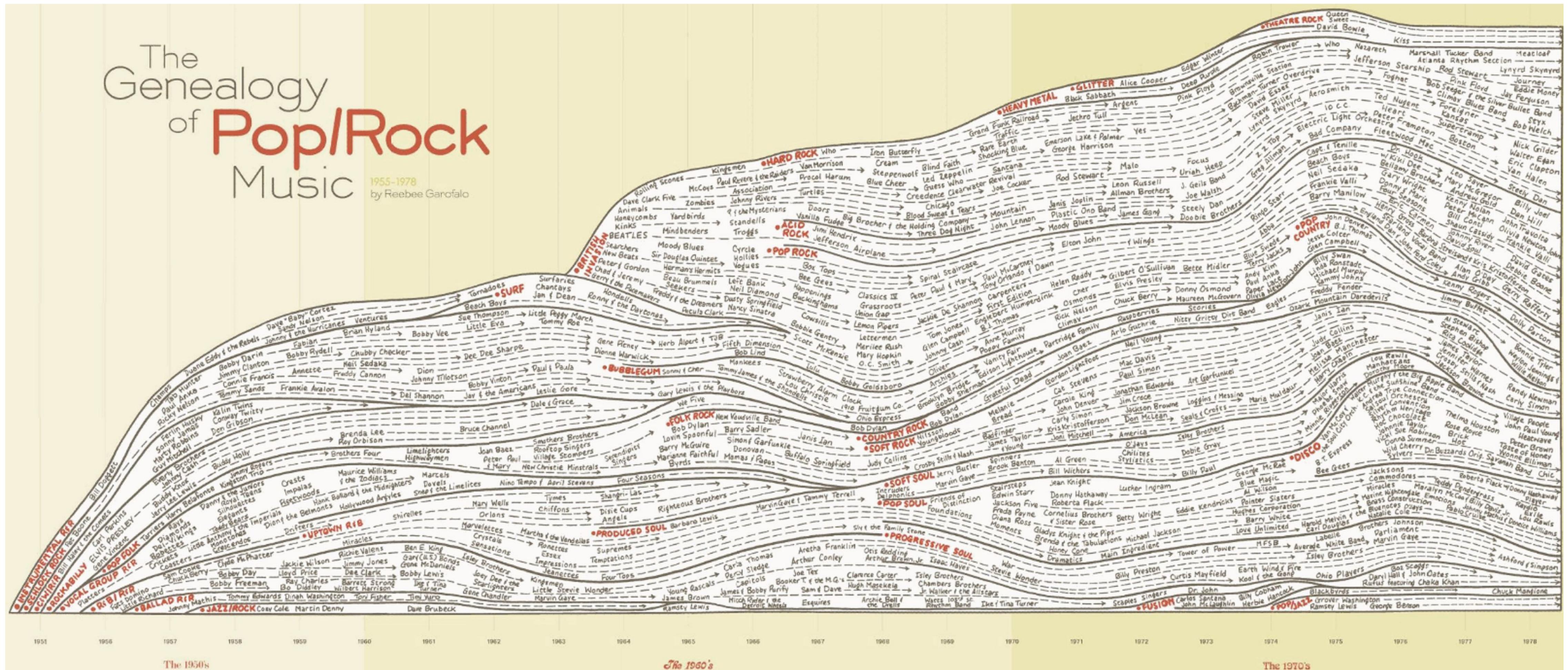
# Proportion

Graphical elements look better together when their relative proportions are in balance



# Data richness

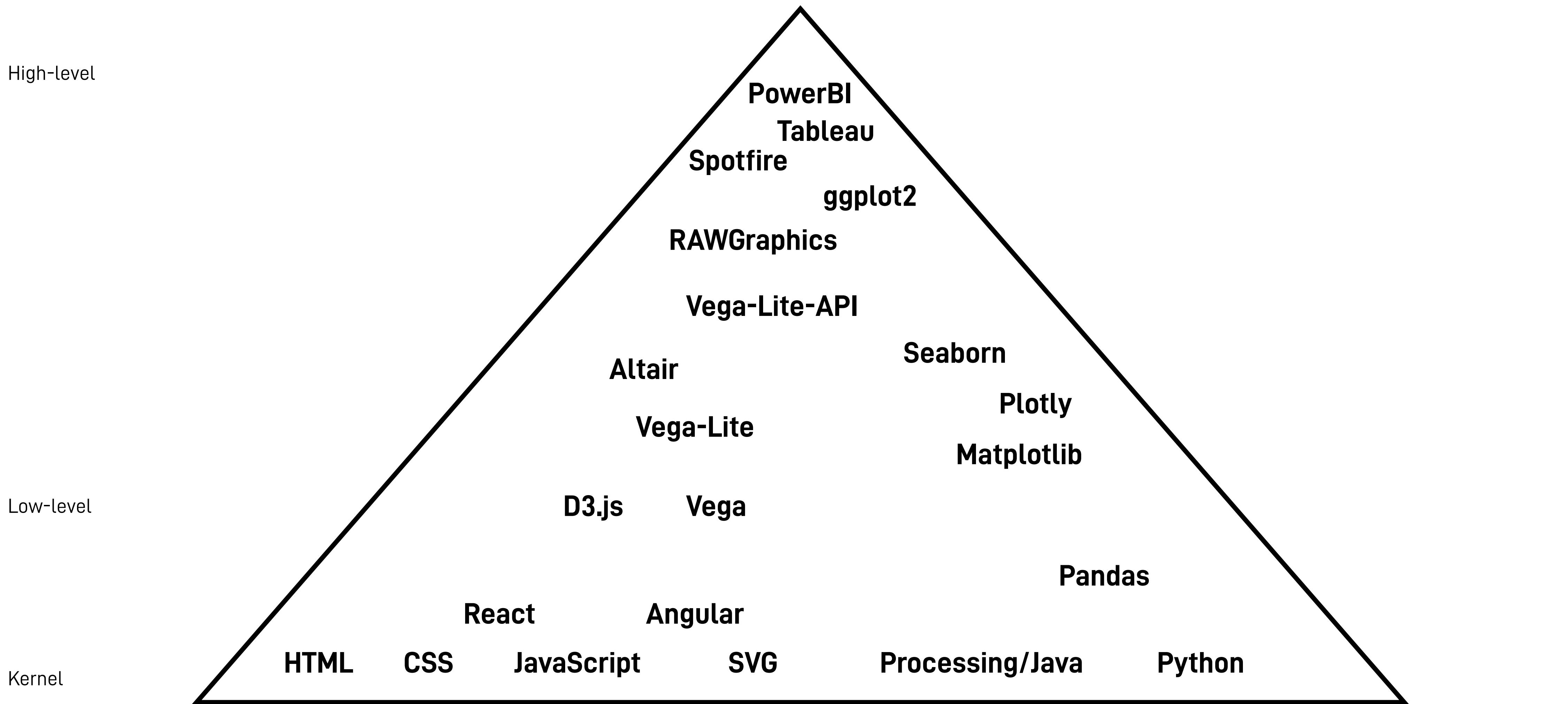
# Maximize the amount of data shown to provide full and accurate picture



[Steve Chappel and Reebe Garofalo in Rock 'N' Roll is Here to Pay: The History and Politics of the Music Industry, 1977]

# Languages, frameworks, and tools

Algorithm



## **Part 2 – Notebook**

<https://bit.ly/3DPPFAP>

