

# FIT3179 Data Visualisation

Week 07: More Idiom Design Choices



# Idiom: ISOTYPE

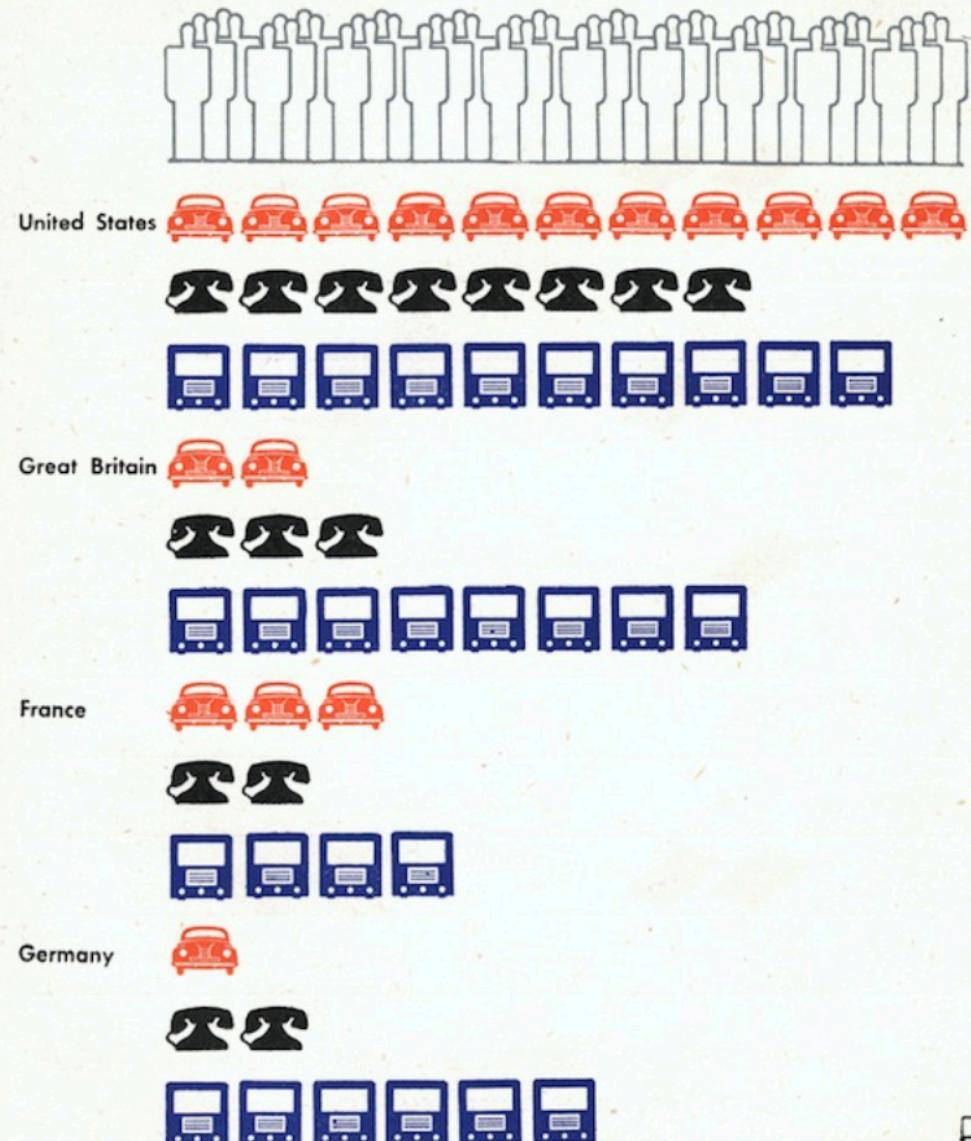
- **What?**
  - 1 quant attribute
- **Why?**
  - Tasks: Compare, lookup values, Find biggest / smallest
- **How?**
  - Marks: pictographs
  - Channels: number of pictographs

ISOTYPE: International System Of Typographic Picture Education

Made popular by Otto Neurath in the 1920s and later.

One symbols (pictograph or pictorial symbol) represents a fixed quantity; stacked to show total.

**Motor Cars, Telephones, Radio Sets 1937**  
per 50 population



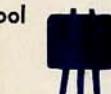
Britain is close to America in radio sets per head, but in motor cars and telephones European countries lag far behind American standards.

# The Young Population aged 13-20

## United States

How many out of ten in each age group

stay at school



are gainfully occupied



Age

13

14

15

16

17

18

19

20

blue: at school

red: gainfully occupied

grey: others

about 1930

Starting at age thirteen and comparing the next age groups 14 to 15, 16 to 17, and 18 to 20, the proportion going to work for money in both countries steps up, while the proportion

## Great Britain

How many out of ten in each age group

stay at school



are gainfully occupied



Age

13

14

15

16

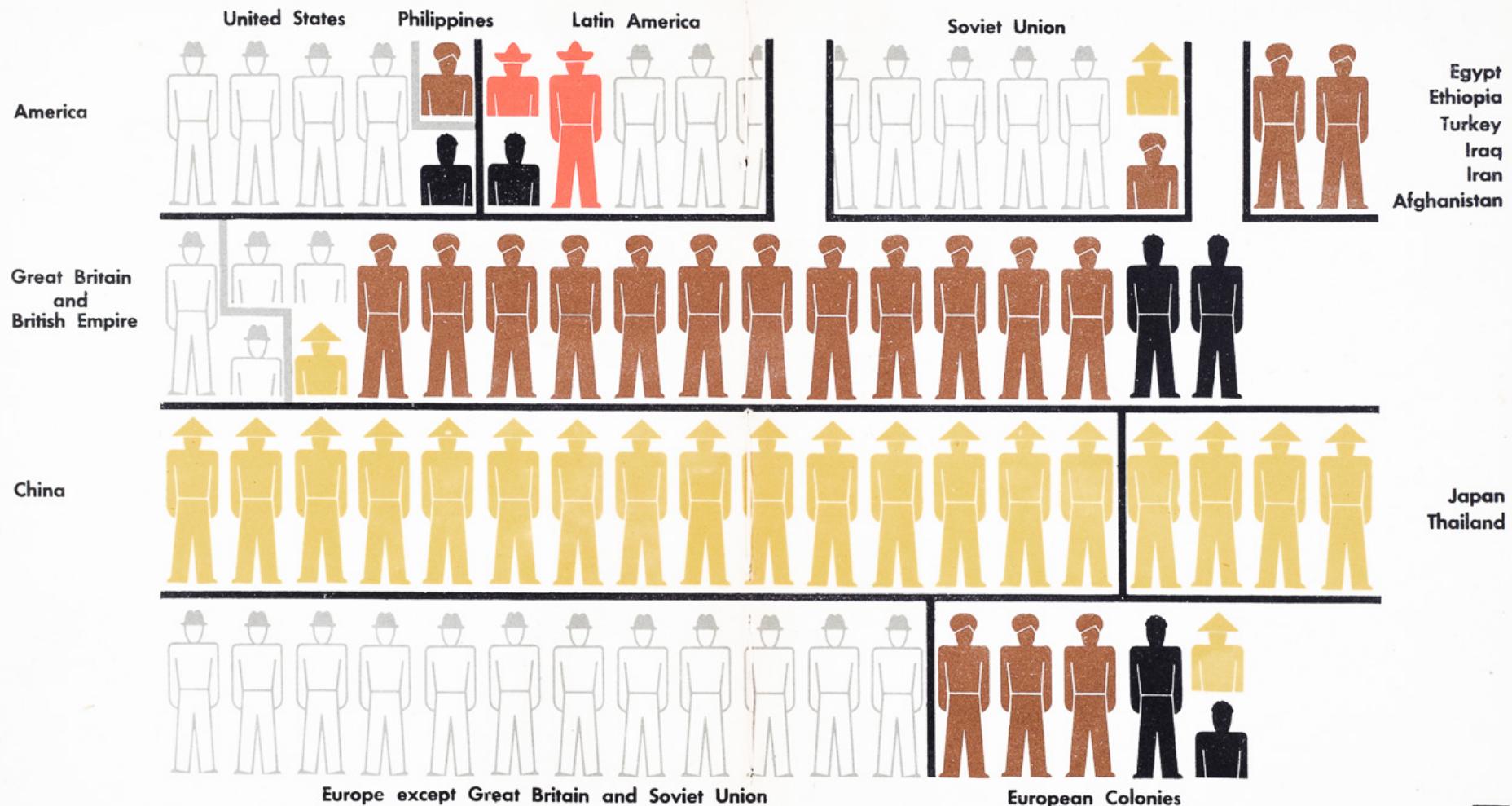
17

18

19

20

# United States and Great Britain in the World



Each complete figure represents 30 million population

Here we are—the peoples of the world—black, white, yellow and brown. The man and a half in Latin America are Red Indians; there are so few left in the United States that it would be difficult to carve a man up small enough to show them. The brown men are mostly Indians. It is surprising to find that there are more Negroes in the British Empire than in the United States.

ISOTYPE

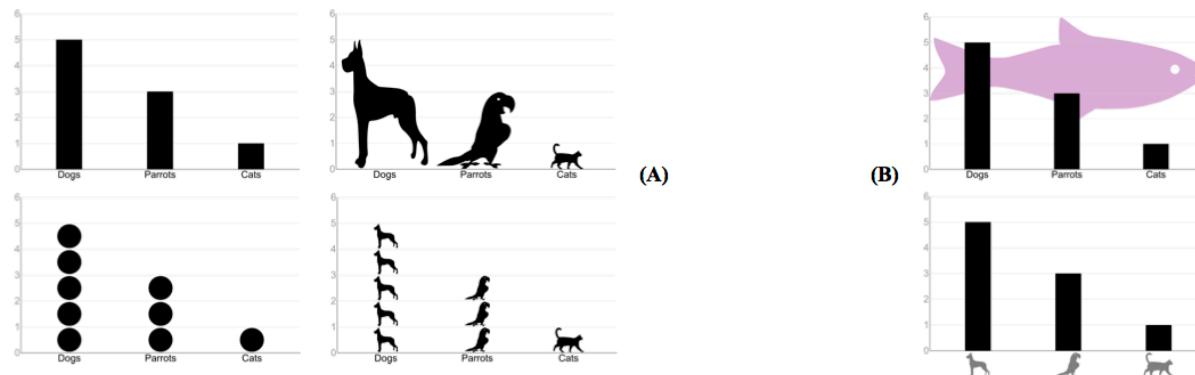


# ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs

**Steve Haroz**  
Northwestern University  
isotype@steveharoz.com

**Robert Kosara**  
Tableau Research  
rkosara@tableau.com

**Steven L. Franconeri**  
Northwestern University  
franconeri@northwestern.edu



**Fig. 1.** Pictographic charts have been used for decades. (A) Which chart above most effectively conveys information? Which data is easiest to remember during a demanding task? Which is most engaging? (B) How integrated must a pictograph be to benefit the user? Do purely decorative background images offer the same benefits as simple axis labels? Or must they be used to convey data?

## ABSTRACT

Although the infographic and design communities have used simple pictographic representations for decades, it is still unclear whether they can make visualizations more effective. Using simple charts, we tested how pictographic representations impact (1) memory for information just viewed, as well as under the load of additional information, (2) speed of finding information, and (3) engagement and preference in seeking out these visualizations. We find that superfluous images can distract. But we find no user costs – and some intriguing benefits – when pictographs are used to represent the data.

## Author Keywords

Visualization; Psychophysics; Working Memory; User Performance; Pictograph; Embellishment; ISOTYPE

## INTRODUCTION

The International System Of TYPographic Picture Education (ISOTYPE) uses simple pictographic elements to convey many types of information, including numerical data. Otto and Marie Neurath defined the term in the 1920s [21], though this type of chart was first described by Willard Brinton in

1914 [6]. Together with Gerd Arntz, the Neuraths created many ISOTYPE designs over several decades [1].

The goal was a universally understandable system for communicating quantities of commercial, social, or economic information (e.g., automobile production or number of children born per year). Symbols, each representing a fixed quantity, were stacked to provide an intuitive representation of a total amount (Fig. 2). Gerd Arntz's pictographs – simplified icons with minimal color – are highly recognizable and are still used in signs, traffic icons, and warning labels.

While the design community has largely embraced the simple style of ISOTYPE for pictographic embellishments [7, 17], the visualization and HCI communities tend to regard pictographs as ‘chart junk’ – a distraction from the data itself [24]. Here we examine how ISOTYPE-style embellishment affects viewer memory, speed, and engagement within simple visualizations.

Recent work suggests that extraneous pictographic information can indeed improve the effectiveness of visualizations. Bateman et al. found that visualizations that integrate

**GUIDELINES**

Based on our findings in the studies, we suggest the following guidelines for using ISOTYPE displays.

**(1) Superfluous pictographs are a distraction**

Pictographs do not impair the viewer as long as they are used to represent data. But including an unnecessary background image in a visualization appears to be distracting, and it may divert attention away from the data. Even replacing text labels with pictographs makes encoding less efficient (at least when the text labels are unambiguous).

**(2) Redundantly code length and (small) number**

Break up large length-defined objects (such as the bars in a bar chart) into a few smaller items. One way of doing so is to use ‘Tufte-style’ gridlines [24], which are white lines superimposed over a (black) bar chart. This approach divides the bars at regular intervals allowing a user to also make a number estimate rather than only a length judgment. For small values below 4-5, number estimation is quick and accurate. However, because this performance diminishes rapidly for larger values, gridlines that break the bars into more than a few sections are unlikely to be beneficial.

**(3) Use pictographs for demanding tasks**

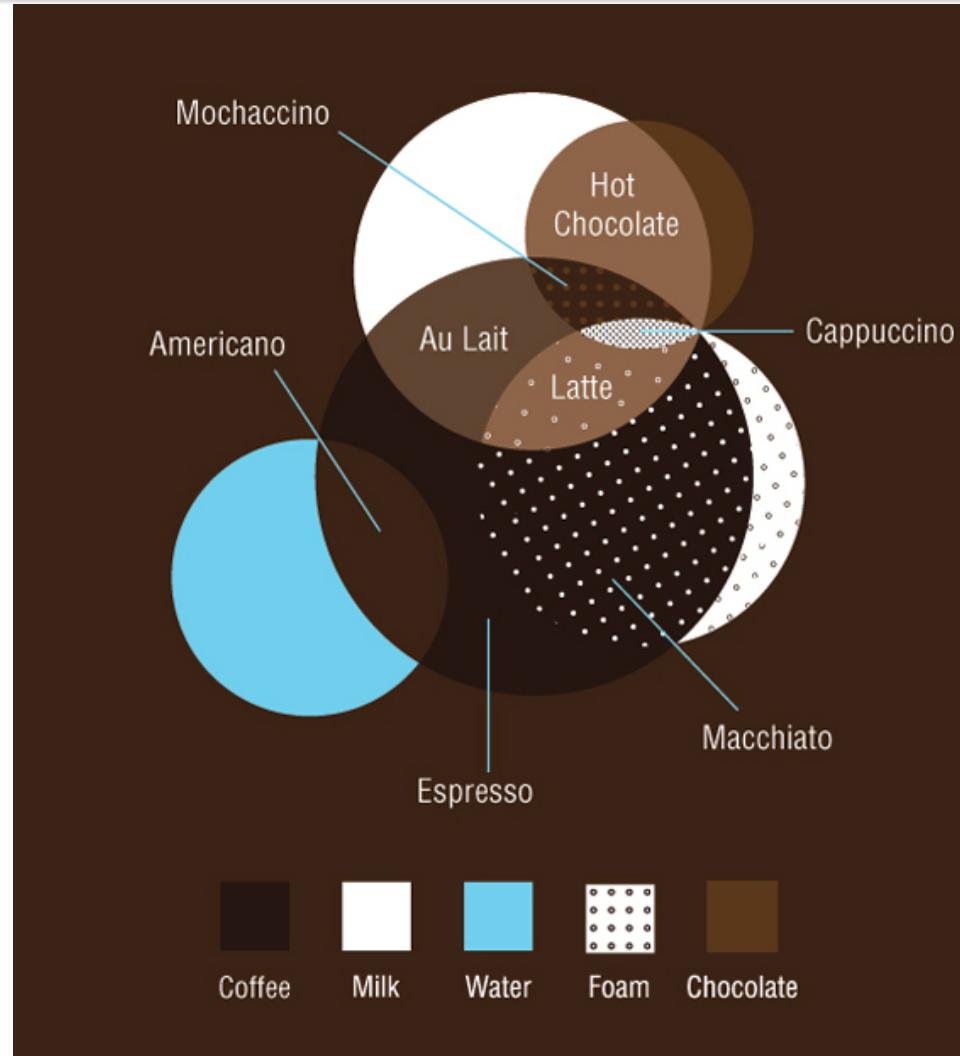
When working memory is under load, the data in ISOTYPE visualizations is recalled more accurately than with simple bar charts. Presenting successive visualizations of different information (such as visualizing sales of different products or showing food preferences in different regions) may benefit from ISOTYPE. In spite of the additional visual complexity, the information is recalled more accurately.

**(4) ISOTYPE engages readers**

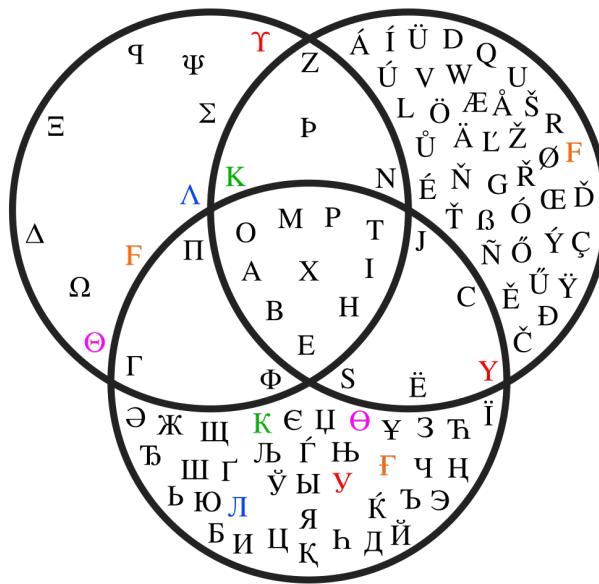
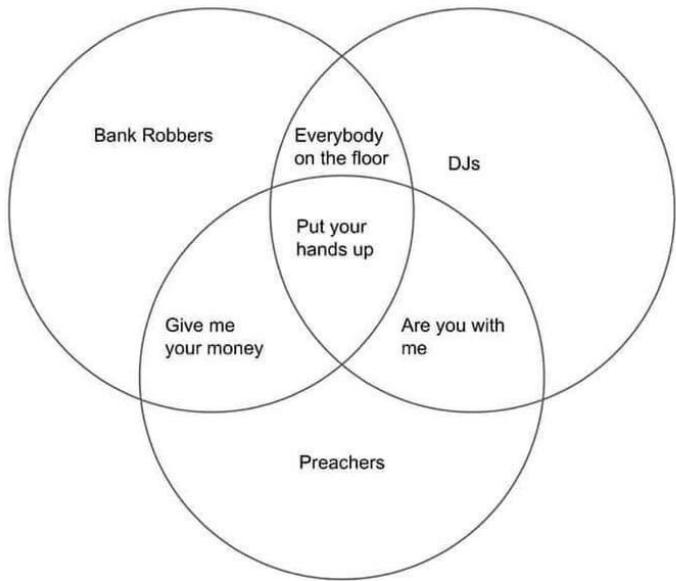
Visualizations rarely exist in isolation. They are often embedded among additional content such as text and other visualizations that compete for a user, reader, or viewer’s attention. ISOTYPE visualizations offer a way through this assortment to engage with a potential viewer. People are inclined, at least initially, to direct their focus towards a visualization with pictographic data compared with a simple bar chart or text.

- **What?**
  - Logical relations among sets
- **How?**
  - Marks: overlapping circles.
  - Channels: position

Show set relationships, not quantities.



# Idiom: Venn diagram (or Set Diagram)



Uppercase letter glyphs shared by the Greek, Latin and Cyrillic alphabets.  
Source: [https://en.wikipedia.org/wiki/Venn\\_diagram](https://en.wikipedia.org/wiki/Venn_diagram)

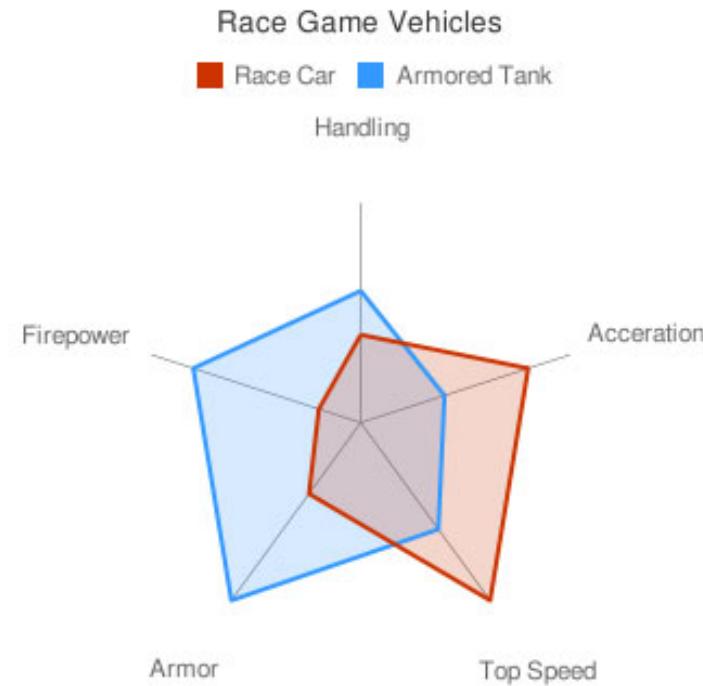
# Idiom: Radar chart

- **What?**
  - Table (multiple attributes)
- **How?**
  - Marks: lines on radial axes.
  - Channels: hue

Relative position and order of axes often unrelated.

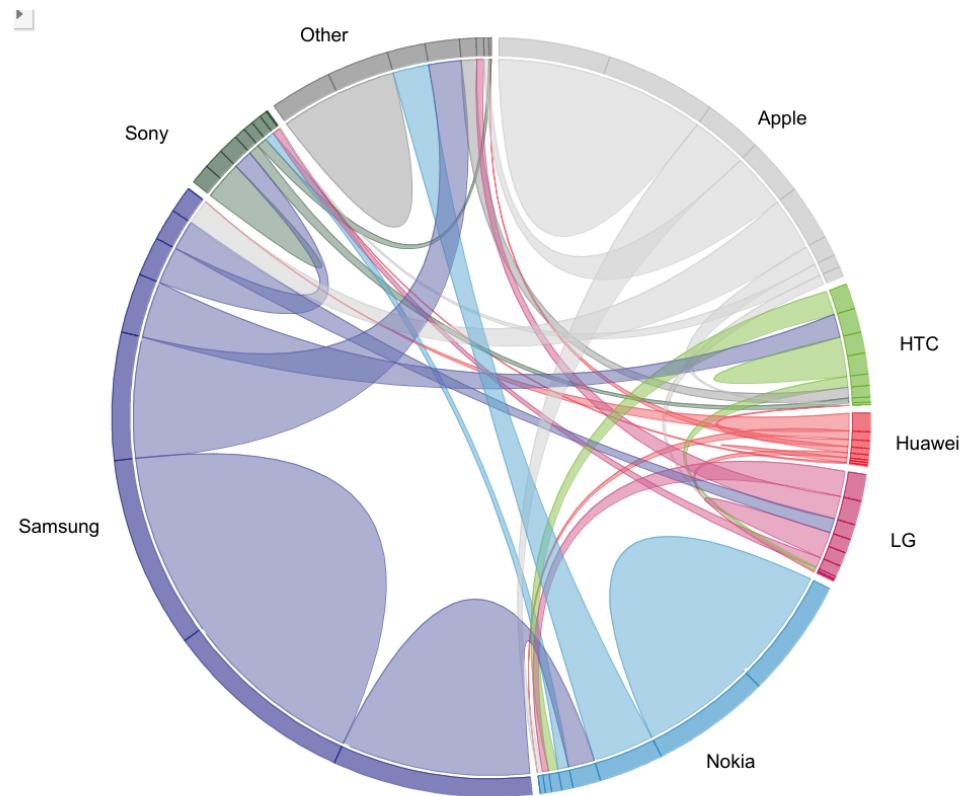
Alternative names: web chart, spider chart, star plot, cobweb, polar chart.

Like parallel coordinates but with polar arrangement.



# Idiom: Chord Diagram

- **What?**
  - Network, edges can be quantitative
- **How?**
  - Marks: proportional curved lines (edges) and areas for nodes along the circle.
  - Channels: size, hue
  - Clutter can become a problem; use line bundling.

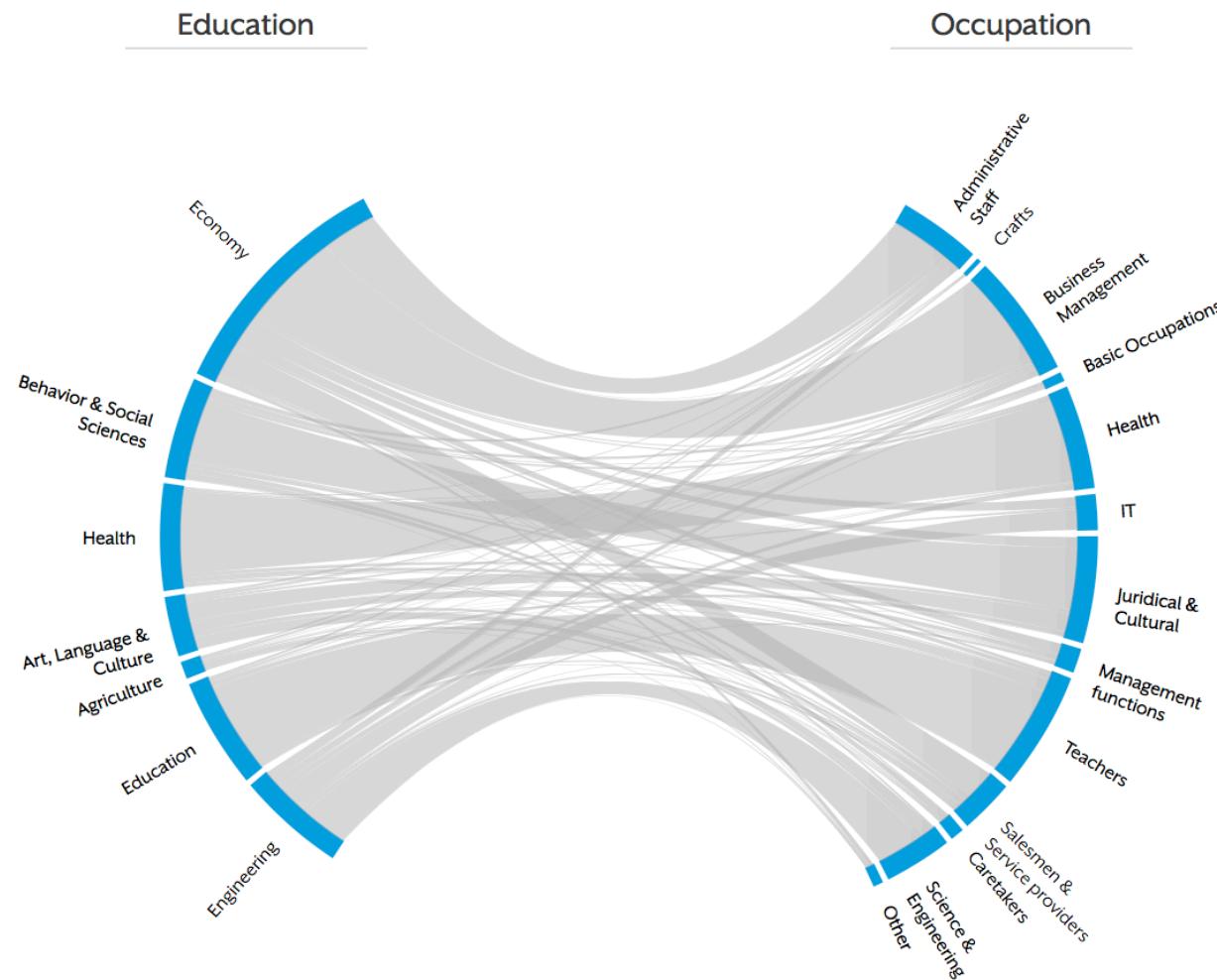


Chord diagram in Tableau:

<http://www.datablick.com/blog/2015/08/27/diy-chord-diagrams-in-tableau-by-noah-salvaterra>

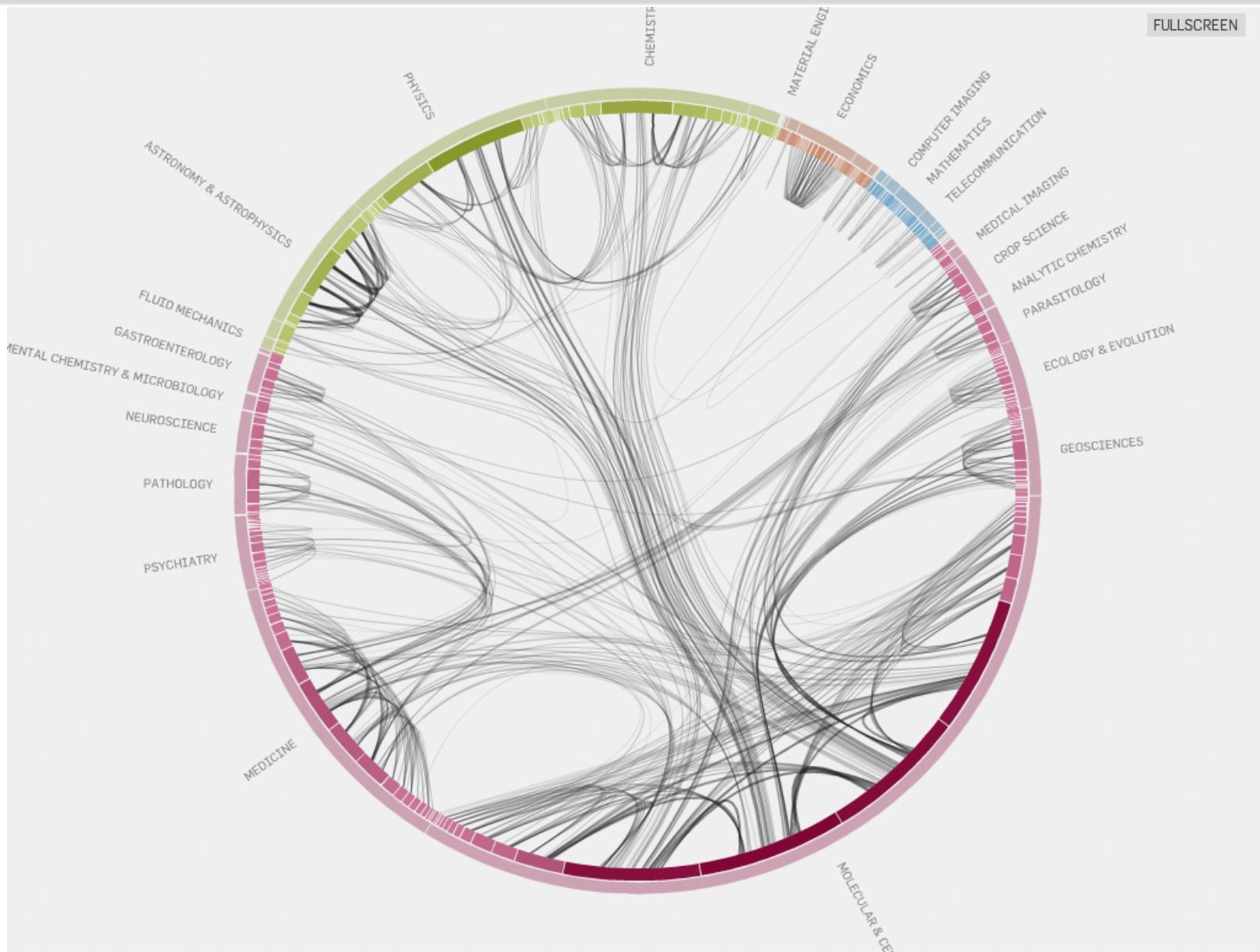
<https://public.tableau.com/profile/nsalvate#!/vizhome/PhoneChord/ChordDiagramBarChart>

# Idiom: Chord Diagram



<https://www.visualcinnamon.com/2015/08/stretched-chord.html>

# Idiom: Chord Diagram with Bundling



# Idiom: Sankey

## ■ What?

- Flows in a network

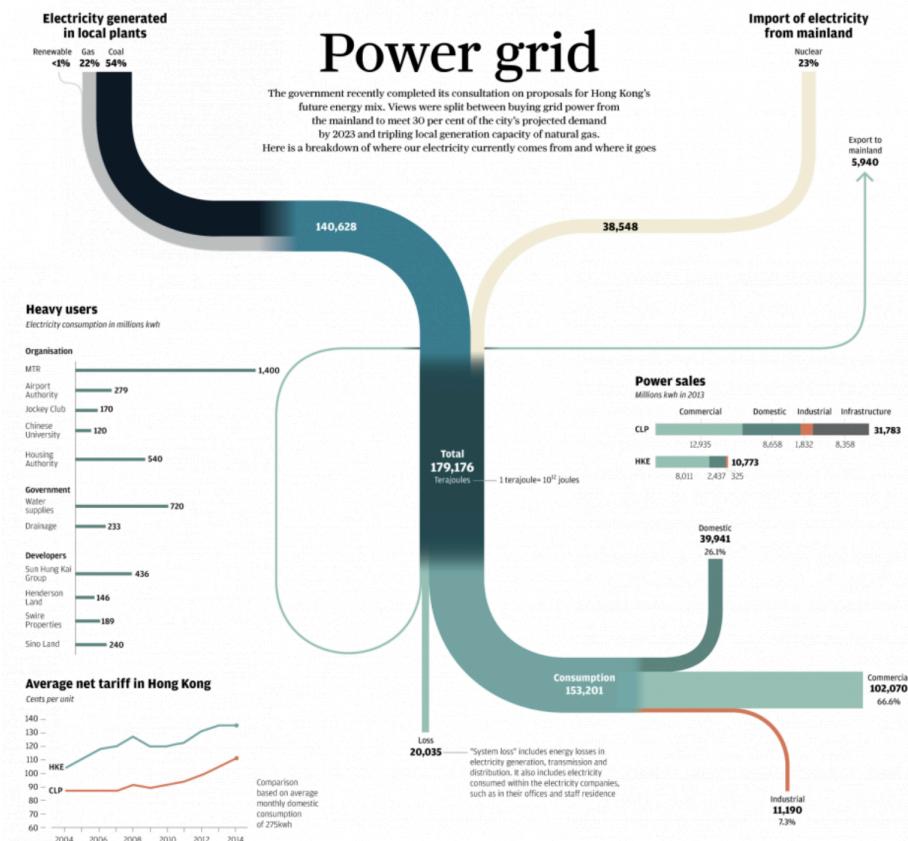
## ■ How?

- Marks: proportional lines for edges, sometimes arrows.
- Channels: size

Like flow maps but without geospatial reference.

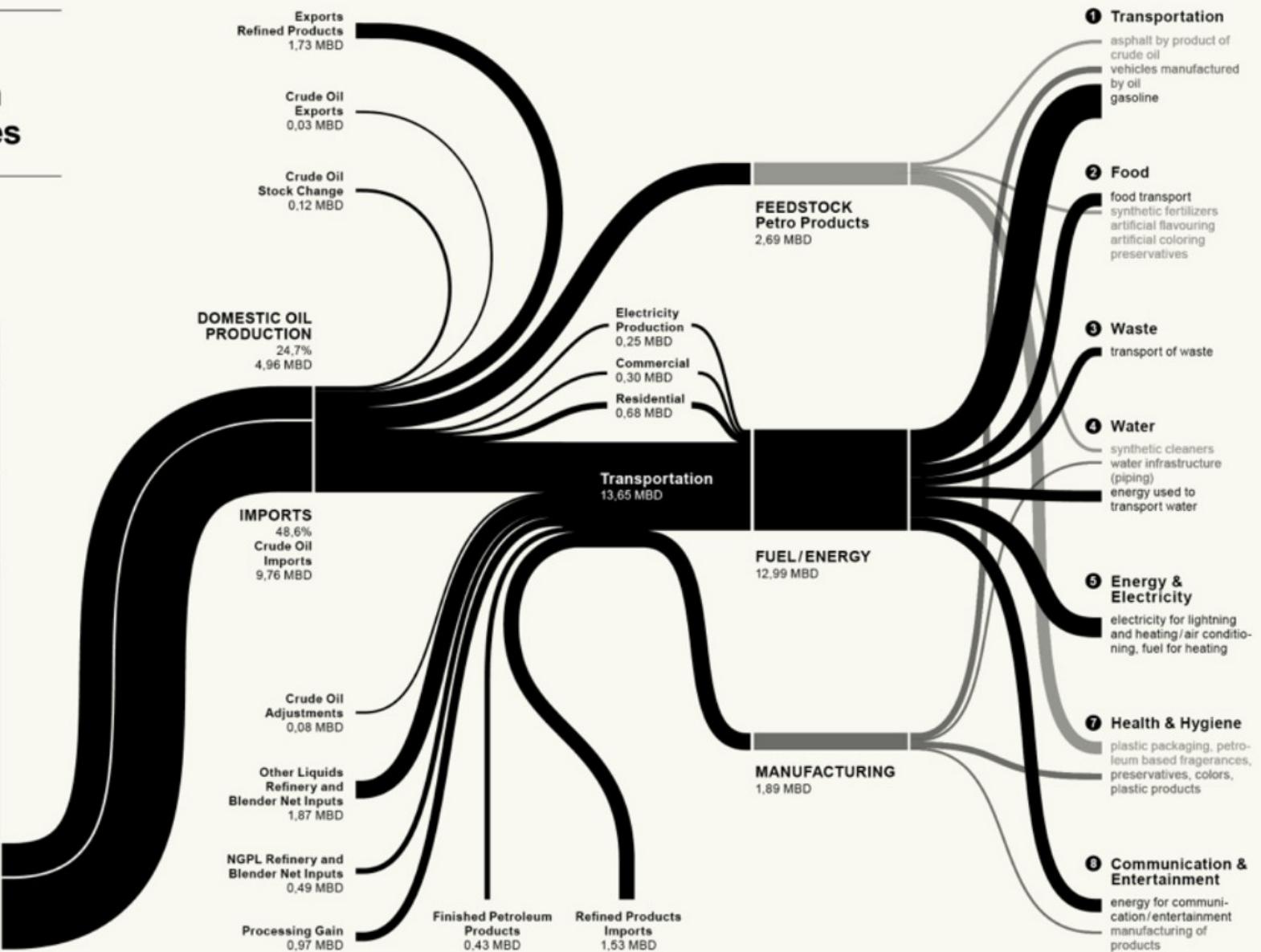
Flows split and merge.

Sankey diagram in Tableau:  
<https://www.youtube.com/watch?v=1HwCzIA9hI4>



## Where is petroleum in our daily lives

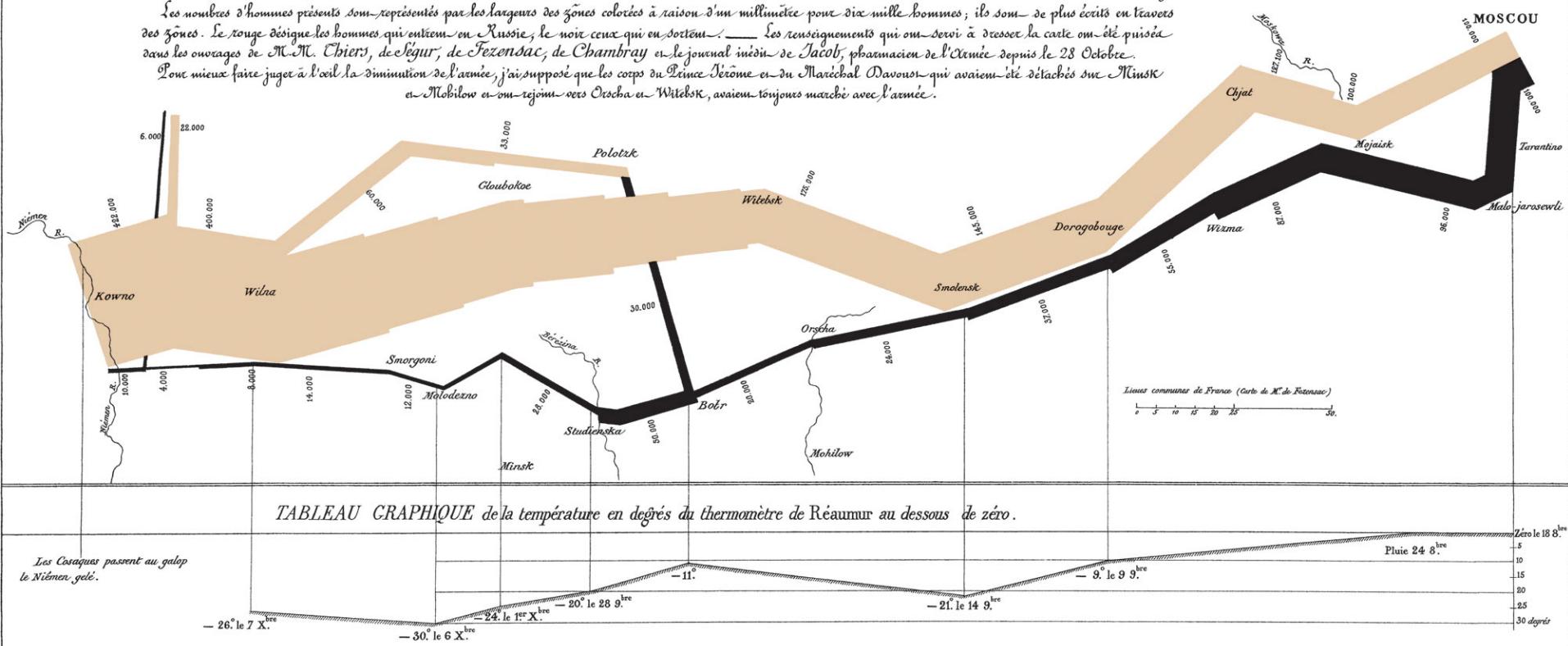
MBD = million barrels a day  
1 Barrel = 42 Gallons



*Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.*  
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.  
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chier, de Clément, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

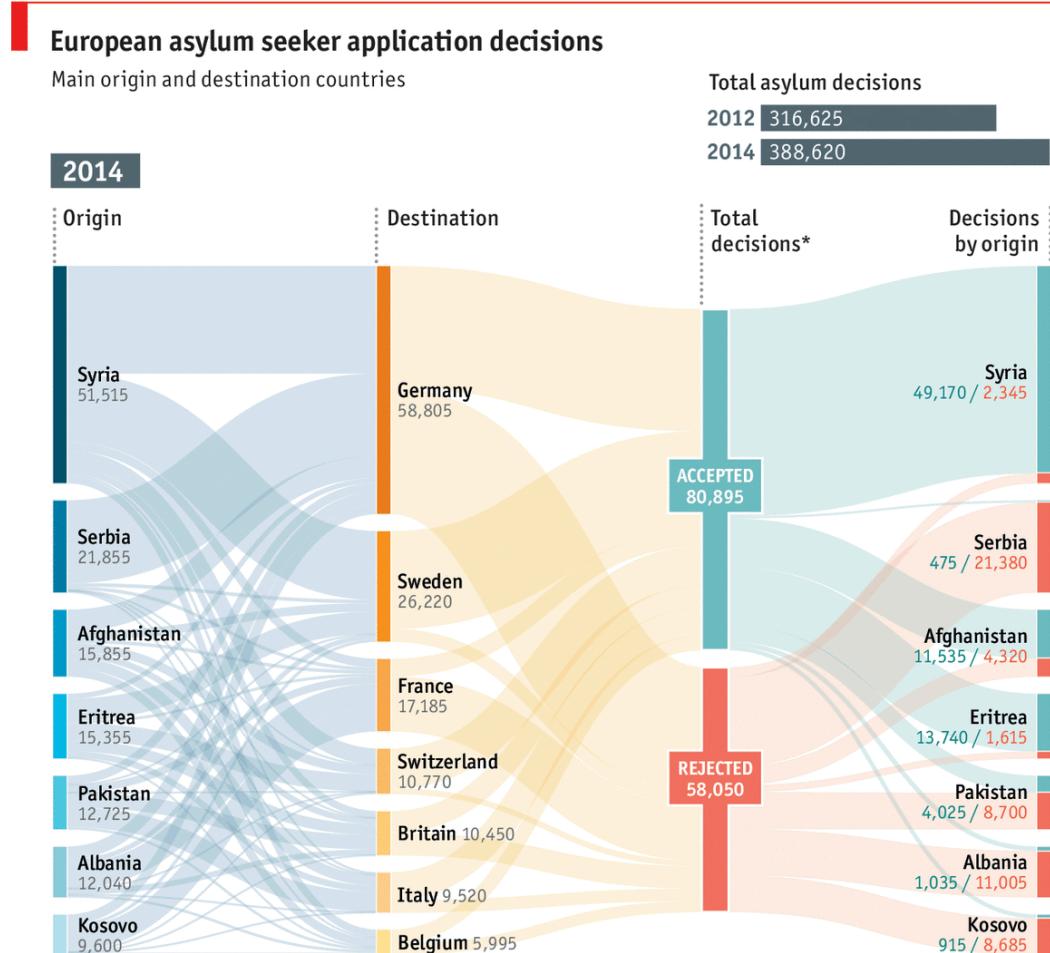
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et de Maréchal Davout qui avaient été détachés sur Minsk et Mohilow se sont rejoints vers Orscha et Witelsk, avaient toujours marché avec l'armée.



# Idiom: Alluvial diagram

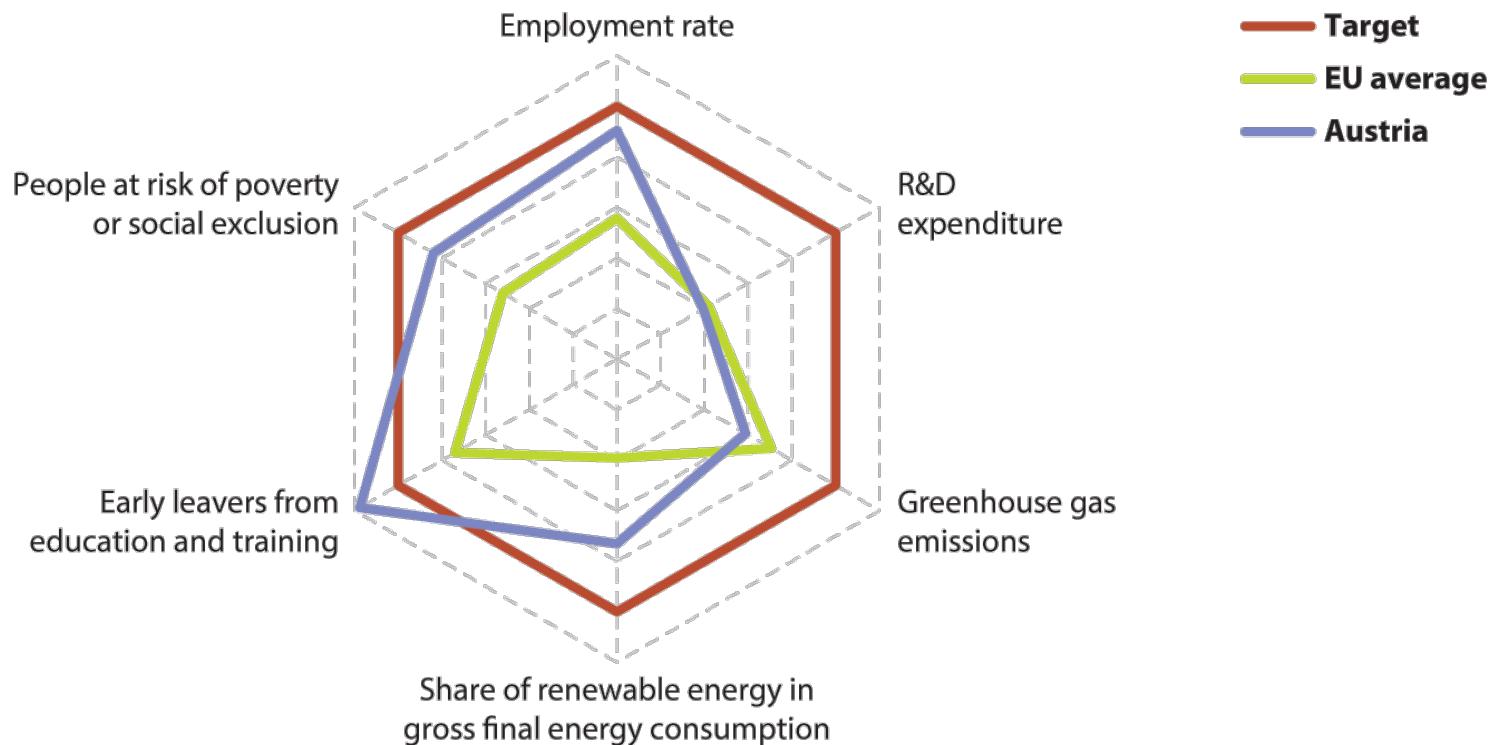
- **What?**
  - Flows in a network
- **How?**
  - Marks: proportional lines for edges, nodes arranged along vertical lines.
  - Channels: size

Like Sankey, but more rigid layout.



Alluvial diagram in Tableau:  
<https://community.tableau.com/thread/152115>

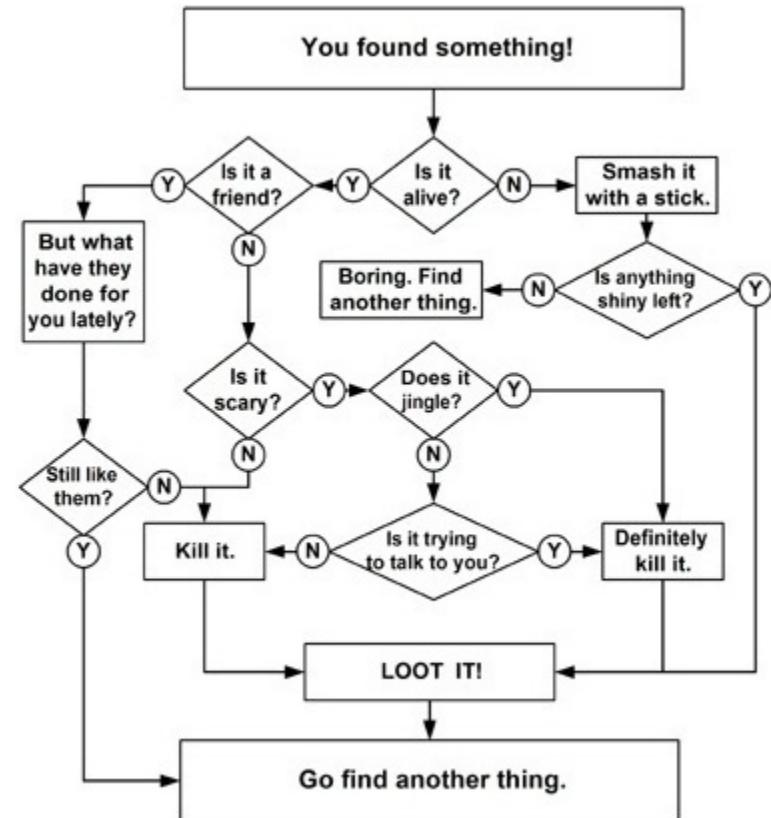
# Idiom: Radar chart



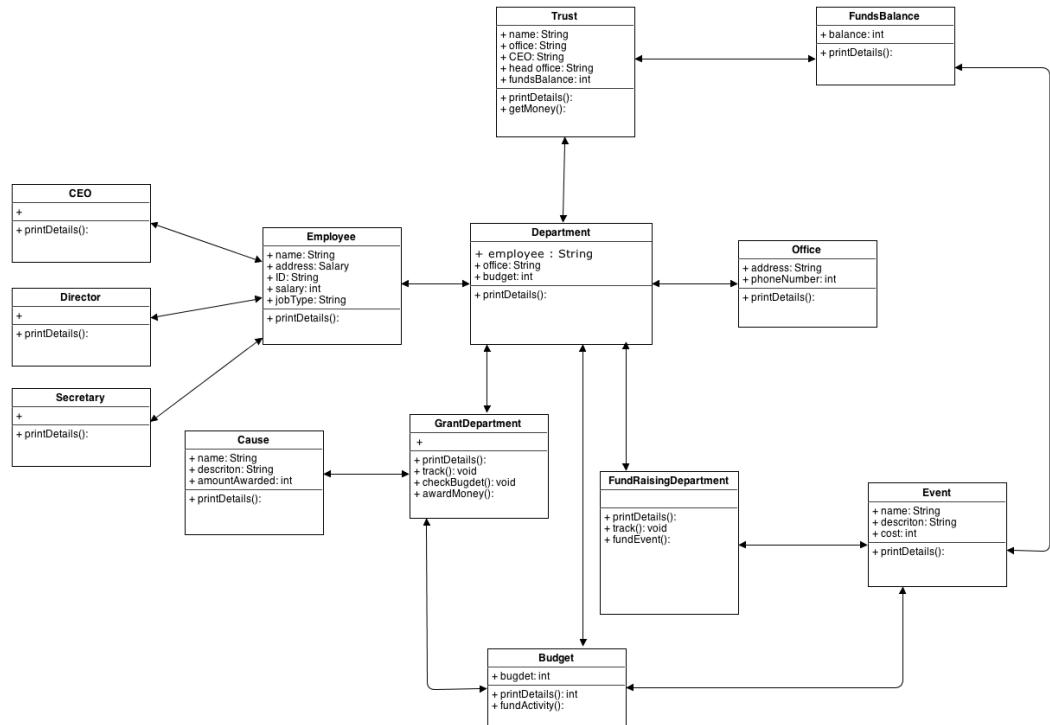
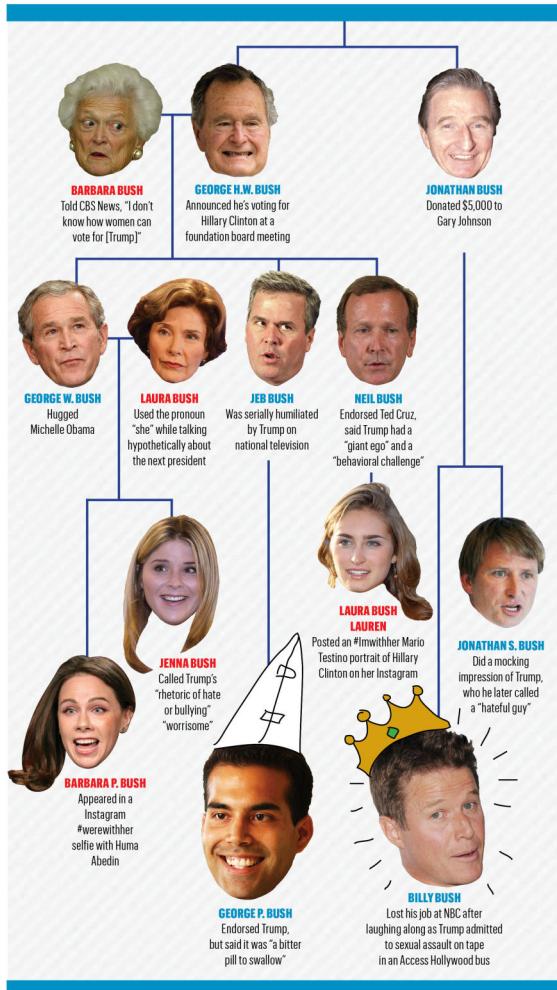
- **What?**
  - Network with attributes
- **How?**
  - Marks: lines for edges, points/areas for nodes
  - Channels: colour, size (line width and point size), shape

Family-tree, flow-charts, node-link diagram, UML diagram, etc.

Show relations among discrete entities.



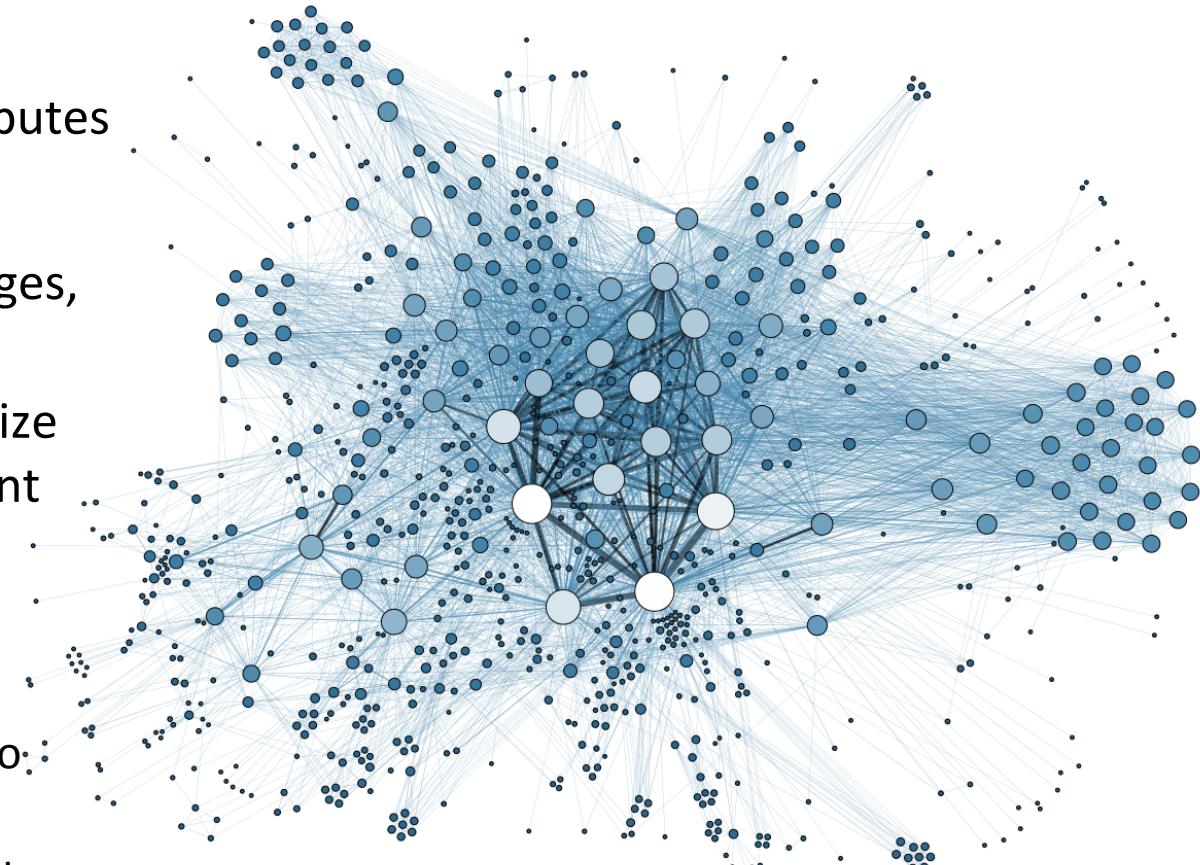
# Network Diagrams



- **What?**
  - Network with attributes
- **How?**
  - Marks: lines for edges, points for nodes
  - Channels: colour, size (line width and point size)

Graph drawing: methods for arranging nodes and edges to depict a graph/network.

Very different layouts possible.  
An important class: force-based layout methods. Repulsing and attracting forces between nodes and edges.



# Selection of an Idiom

## Selection by dataset type



# Selection of an Idiom

Selection by attribute type and table layout

datavizproject.com/#

A project in beta by **ferdlio** FAMILY  INPUT  FUNCTION  SHAPE

D	V
P	

X	Y <sub>1</sub>	Y <sub>2</sub>	
1	30	28	>
2	34	22	
3	38	26	
	▼		

A	2010	2015	
A	1	3	>
B	2	1	
C	3	2	
	▼		

A	60%	

X	Y	
0-2	30	
2-4	34	
4-6	38	
	▼	

Event	Time	
A	1-4-2015	
B	10-4-2015	
C	12-4-2015	
	▼	

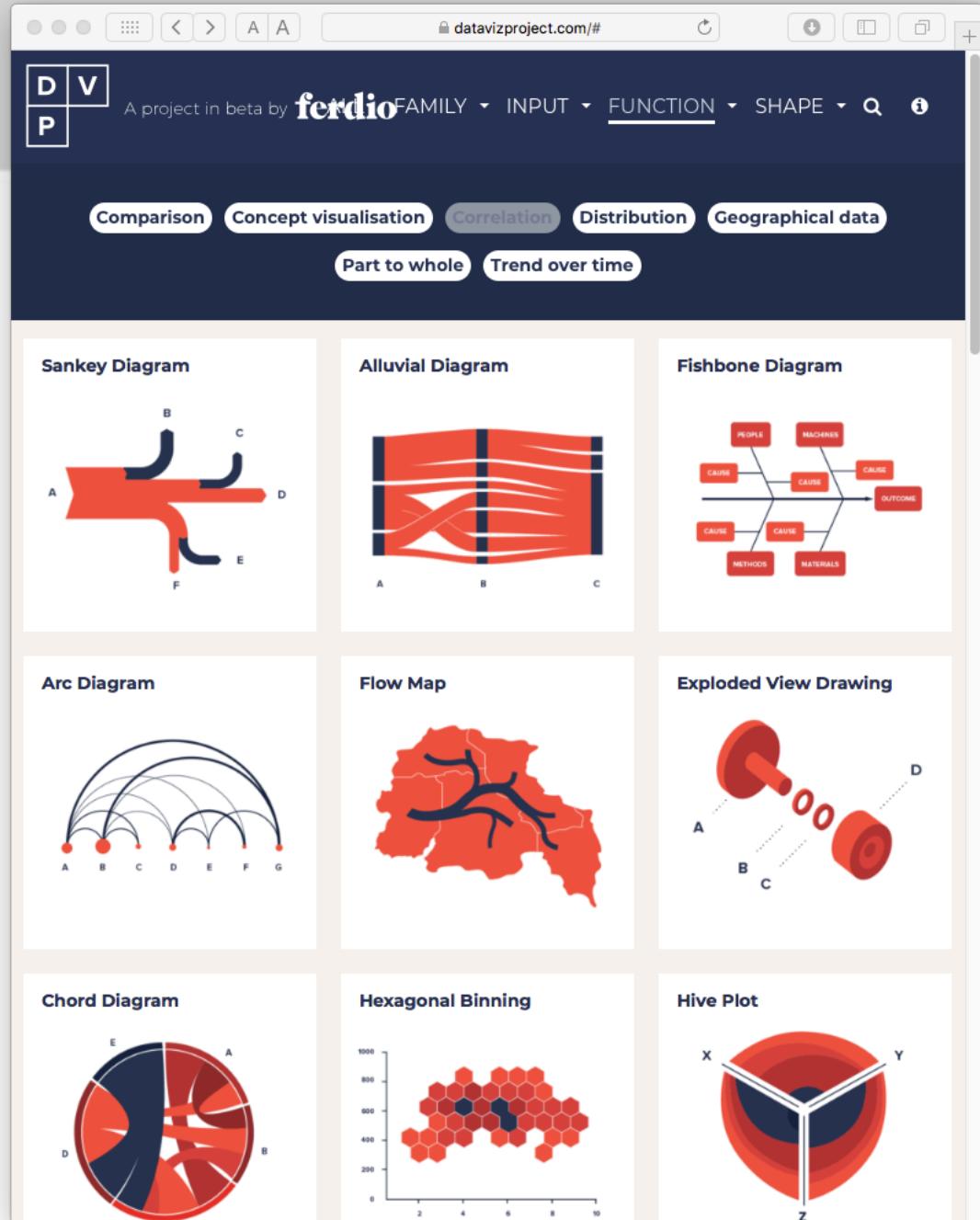
Bump Chart

Stacked Ordered Area Chart

Do you have any feedback or suggestions for Data Viz Project? Let us know

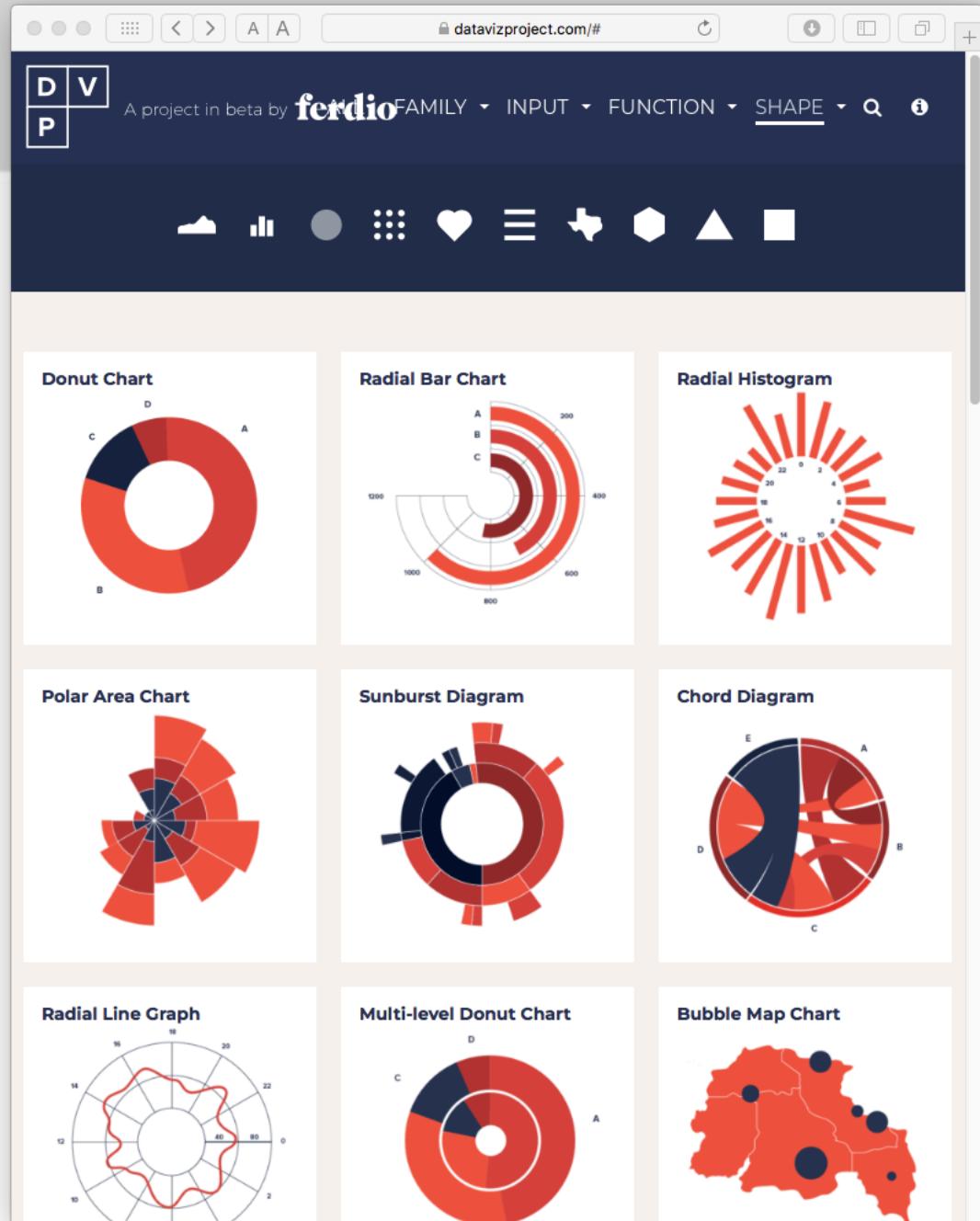
# Selection of an Idiom

Selection by action  
and target



# Selection of an Idiom

## Selection by shape



# Selection by Action and Target

## Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is the mean, median or target or a long-term average. Can also be used to show sentiment (optimistic/negative).

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

### Diverging bar



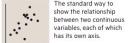
A simple standard bar chart that handles both negative and positive magnitude values.

## Correlation

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

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

### Scatterplot



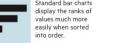
The standard way to show the relationship between continuous variables, each of which has its axis.

## 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

### Ordered bar



Standard bar charts are good for showing the distribution of values much more easily when sorted into order.

## Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution is a meaningful way of highlighting the lack of uniformity or equality in the data.

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

### Histogram



The standard way to show the distribution - keep the gaps between columns equal to highlight the shape of the data.

## Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or long (months, years, traversing decades or centuries). Choosing the correct time period is crucial to provide suitable context for the reader.

**Example FT uses**  
Price movements, economic time series

### Line



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

### Column



Columns work well for showing change over time, especially if the data is best with only one series at a time.

## Magnitude

Show size comparisons. These can be relative (just being able to see larger/smaller) or absolute (able to see five differences). Usually these show a 'counted' number (for example, bars, dots, etc.) and require some calculation or per cent.

**Example FT uses**  
Commodity production, market capitalisation

### Stacked column



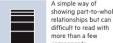
A simple way of showing the size of components relative to the whole, starting at 0 on the axis.

## Part-to-whole

Show how a single entry can be broken down into its component elements. If the reader needs to know the size of the components, consider using a magnitude-type chart instead.

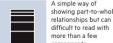
**Example FT uses**  
Fiscal budgets, company structures, national election results

### Bar



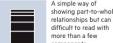
See above: Good when the data are not time series and labels have long category names.

### Marimekko



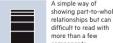
A good way of showing the size and proportions of data at the same time – as long as the data is not too complicated.

### Pie



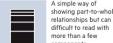
A common way of showing parts-to-whole relationships – but beware that it's difficult to read and compare the size of the segments.

### Donut



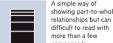
Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg totals).

### Treemap



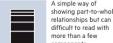
Use for hierarchical data – relationships can be difficult to read when there are many small segments.

### Equalised cartogram



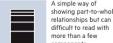
Converting each unit on a map to a regular and equal size – good for representing regions with equal value.

### Contour map



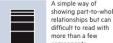
For showing a set of equal values on a map. Can use elevation colour schemes for showing > values.

### Scaled cartogram (value)



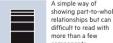
Stretching and shrinking a map so that areas are sized according to a particular value.

### Dot density



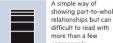
Used to show the location of individual events/locations – can create any patterns the reader should see.

### Gridplot



Good for showing % data points in a grid – works best when used on whole numbers and doesn't need a single layout form.

### Heat map



Grid-based data values with an intensity colour scale. As choropleth map – good for relating to an admin/political unit.

## Spatial

Aside from locator maps only used when precise location or geographical patterns are more important to the reader than results.

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

### Basic choropleth (rate/ratio)



The standard approach for putting data on a map – should be rates rather than totals as they're a sensible base geography.

### Proportional symbol



Used to relate other rates – but rates are better than totals as small differences will be hard to see.

### Waterfall



Designed to show the sequencing of data flow between locations – process, typically budgets. Can include > components.

### Chord



A complete but powerful diagram which can make complex (and not winner) in a matrix.

### Network



Used for showing the strength and inter-connectedness of relationships of varying types.

## Flow

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

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

### Sankey



Shows changes in flows from one node to another – less one other; good for tracing the eventual outcome of a complex process.

# 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 Sterns, Chris Campbell, Ian Britt, Liu Fearn, Giadae Parish, Billy Oberberg, Paul McCullagh, Marlin Stabe

Inspired by: Graphic Catechism by Jon Saksida and Giovanni Salata



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



© Financial Times

# Selection of an Idiom

## Vis catalogues

- <https://datavizproject.com>
- <https://datavizcatalogue.com>
- <https://ft.com/vocabulary>

## What do you want to show?

Here you can find a list of charts categorised by their data visualization functions or by what you want a chart to communicate to an audience. While the allocation of each chart into specific functions isn't a perfect system, it still works as a useful guide for selecting chart based on your analysis or communication needs.



Comparisons



Proportions



Relationships



Hierarchy



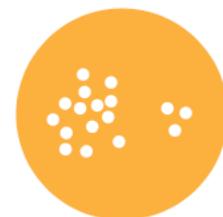
Concepts



Location



Part-to-a-whole



Distribution



How things work



Processes & methods



Movement or flow



Patterns



Range



Data over time



Analysing text



Reference tool