ELECTRONIC PUBLISHING
AND DIGITAL STORYTELLING
Lesson 2

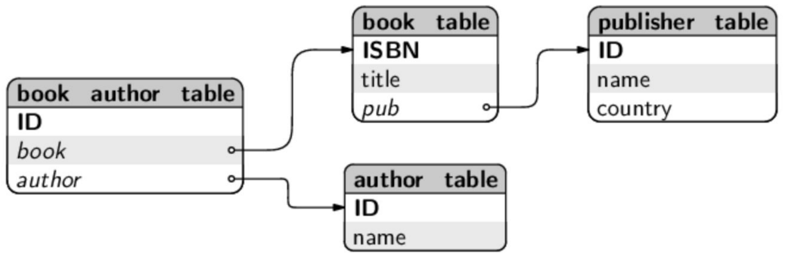
PRELIMINARY CONCEPTS

TABLE OF CONTENTS

DATA

DATA

```
<book id="b002" year="2000">
   <title>Data on the Web</title>
   <author><last>Abiteboul</last><first>Serge</first></author>
   <author><last>Buneman</last><first>Peter</first></author>
   <author><last>Suciu</last><first>Dan</first></author>
   <publisher>Morgan Kaufmann</publisher>
</book>
```



A datum is an abstraction of a real-world **entity** (e.g. a person, a book, an artwork).

Data are characterised by **attributes** (also called features or variables). *For instance: author, title, ISBN are all attributes of a book.*

DATASET

	ISBN	bookTitle	bookAuthor	yearOfPublication	publisher	imageUrlS
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

A dataset is a collection of entities and their attributes.

A dataset can be seen as a **n*m matrix**, where n are the entities (rows) and m are the attributes (columns).

DATA ATTRIBUTES

TYPES OF DATA ATTRIBUTES

Measurable quantities represented by integers (e.g. temperature, price).

NUMERIC

Or (Nominal) are names from non-overlapping sets of categories, classes, or states (e.g. title, sex)

CATEGORICAL

Nominal or numeric attributes that can be ranked (e.g. days, years, Likert: “strongly like” to “strongly dislike”).

ORDINAL

* NUMERIC ATTRIBUTES

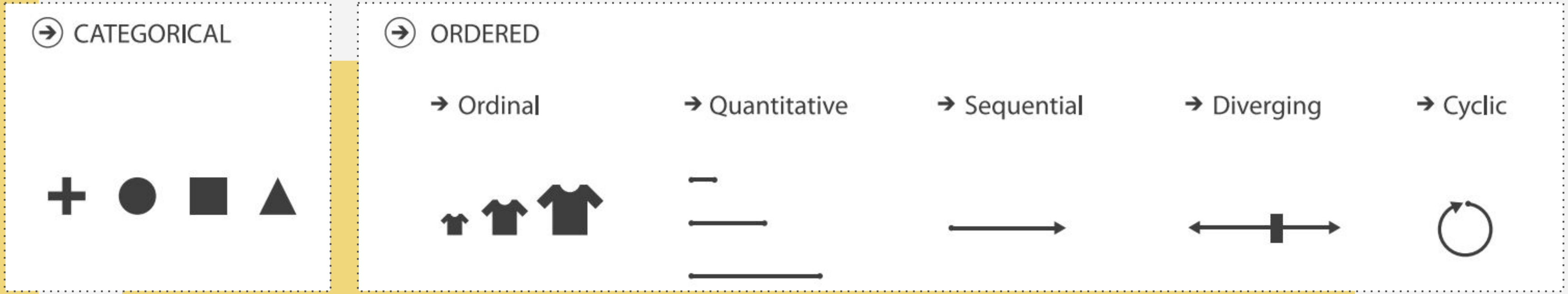
Interval attributes are measured on the basis of a scale with an interval and an origin (e.g. time, temperature). They have **no true-zero**.

INTERVAL SCALE

Ratio attributes have a true-zero - meaning no quantity is measured in that point (e.g. price)

RATIO SCALE

ANOTHER WAY OF SEEING IT



Cannot be ordered

Can be ordered according to their magnitude, quantity, order in a sequence, etc.

WHAT OPERATIONS CAN BE DONE ON DATA ATTRIBUTES

interval attributes cannot be multiplied or divided

Potentially, can be **ordered** (interval) and arithmetic operations can be performed (ratio).

NUMERIC

Can be **sorted** (e.g. alphabetically). Cannot be ordered and arithmetic operations cannot be performed.

CATEGORICAL

Can be naturally **ordered**, but arithmetic operations are not possible

ORDINAL

WHICH ANALYTICAL METHODS CORRESPOND TO DATA ATTRIBUTES

Statistical analysis

NUMERIC

Counting and proportion of occurrences (distribution) in the dataset

CATEGORICAL

Counting and proportion of occurrences (distribution) in the dataset

ORDINAL

OTHER CHARACTERISTICS OF DATA

A parenthesis on higher level aspects

STRUCTURED AND UNSTRUCTURED DATA

Can be stored in a **table**. Data have a similar (attributes) structures and are easy to query, store, merge with other structured data.

STRUCTURED

Every instance of the dataset may have a different internal structure (e.g. web pages are semi-structured).

UNSTRUCTURED

RAW AND DERIVATIVE DATA

Raw abstractions of events and objects (e.g. weight, height).

RAW

Derived from the **interaction** of other pieces of data (e.g. percentage, average).

DERIVATIVE

* RAW DATA

Collected by means of direct measurement or observation (e.g. surveys).

CAPTURED

By-product of the original collected data (e.g. social media interactions) or **metadata**.

EXHAUST

DATABASES

Based on relations between tables.

SQL

Based on object-oriented flexible structures

No SQL

INTRODUCTION TO DATA VISUALISATION

DATA VISUALISATION

At the end of the 18th century, **William Playfair** invented statistical graphics.

He designed line charts and area charts (time-series data), bar charts (quantitative or categorical comparison), and pie charts (proportions within a set)

DATA VISUALISATION

Since visualisations allow to have a view of large and complex datasets, these are fundamental aid in data science in all the phases, such as **exploring**, **understanding** and **presenting** results.

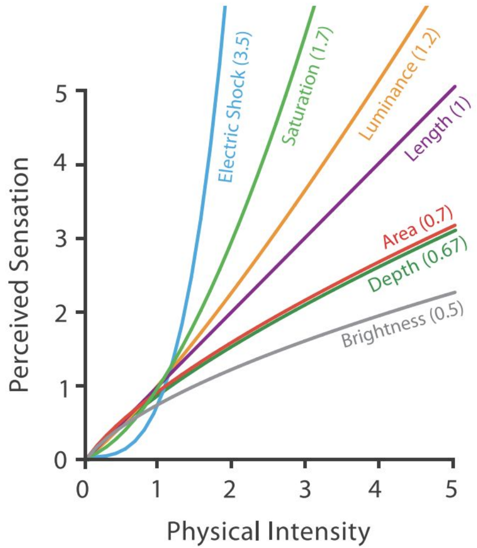
A Disclaimer

Visualisations rarely give precise answers to a question (e.g. low-level precision of areas in bubble charts). Rather, they frame phenomena and contribute to give a precise answer.

DATA VISUALISATION DIMENSIONS

What are the recurring graphical aspects and how these interact with the viewer's perception and data attributes

PERCEPTION OF VISUAL ASPECTS



Steven's Psychophysical Power Law: $S = I^N$

Stevens, 1975

Visual aspects have a **functional** role in the interpretation of data rather than aesthetic.

The idea of visualising data in graphical forms is to **replace cognition with perception**.

The human eye **perceives** visual dimensions in different ways, with more or less effort, and users' **knowledge background** can affect the perception.

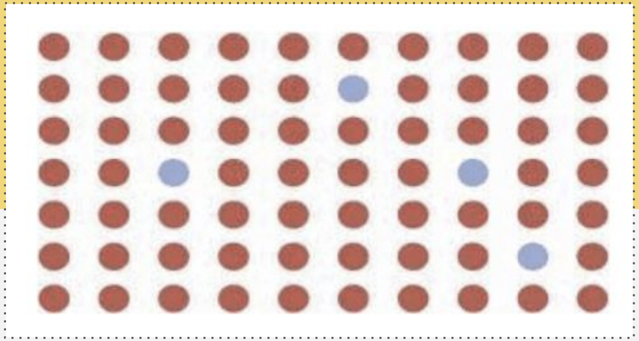
https://en.wikipedia.org/wiki/Stevens%27s_power_law

TIDWELL’S PREATTENTIVE VARIABLES

Some visual aspects work preattentively, meaning they are able to communicate something **before** the user pays conscious attention to it.

Effective and impactful data visualisations work extensively on these aspects.

0.103	0.176	0.387	0.300	0.379	0.276	0.179	0.321	0.192	0.250
0.333	0.384	0.564	0.587	0.857	1.064	0.698	0.621	0.232	0.316
0.421	0.309	0.654	0.729	0.228	0.529	0.832	0.935	0.452	0.426
0.266	0.750	1.056	0.936	0.911	0.820	0.723	1.201	0.935	0.819
0.225	0.326	0.643	0.337	0.721	0.837	0.682	0.987	0.984	0.849
0.187	0.586	0.529	0.340	0.829	0.835	0.873	0.945	1.103	0.710
0.153	0.485	0.560	0.428	0.628	0.335	0.956	0.879	0.699	0.424



Tidwell et al, Designing interfaces
https://www.oreilly.com/library/view/designing-interfaces-3rd/9781492051954/

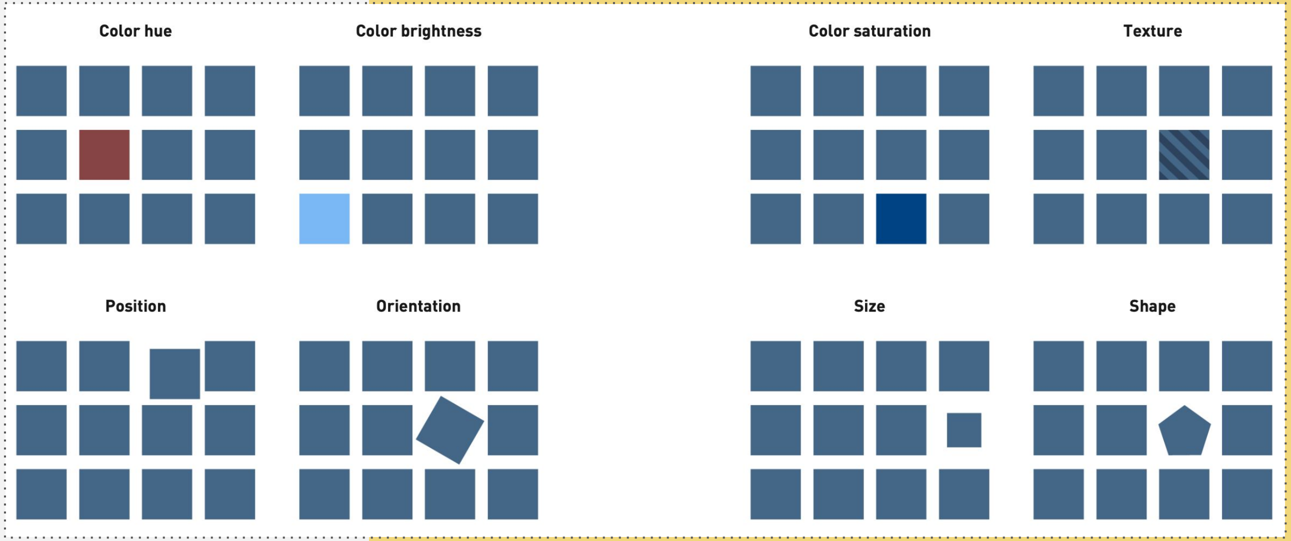
TIDWELL'S PREATTENTIVE VARIABLES

Color HSL
(hue, saturation, lightness)

Form
(size, shape)

Movement

Positioning



GRAPHICAL PROPERTIES

Graphical properties are properties which can be applied to the graphical elements which make them more (or indeed less) noticeable/valuable to the user of the representation.

Axes (Orientation)

Color

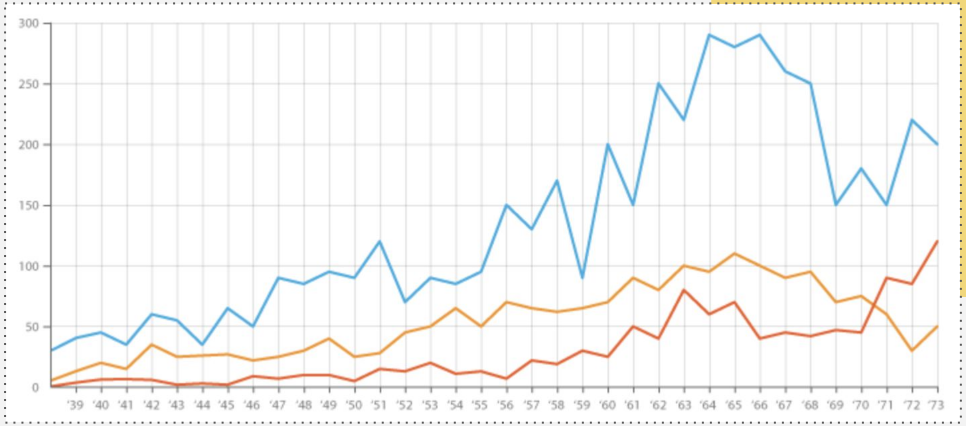
Layout

Size

Shape

Typography

AXES



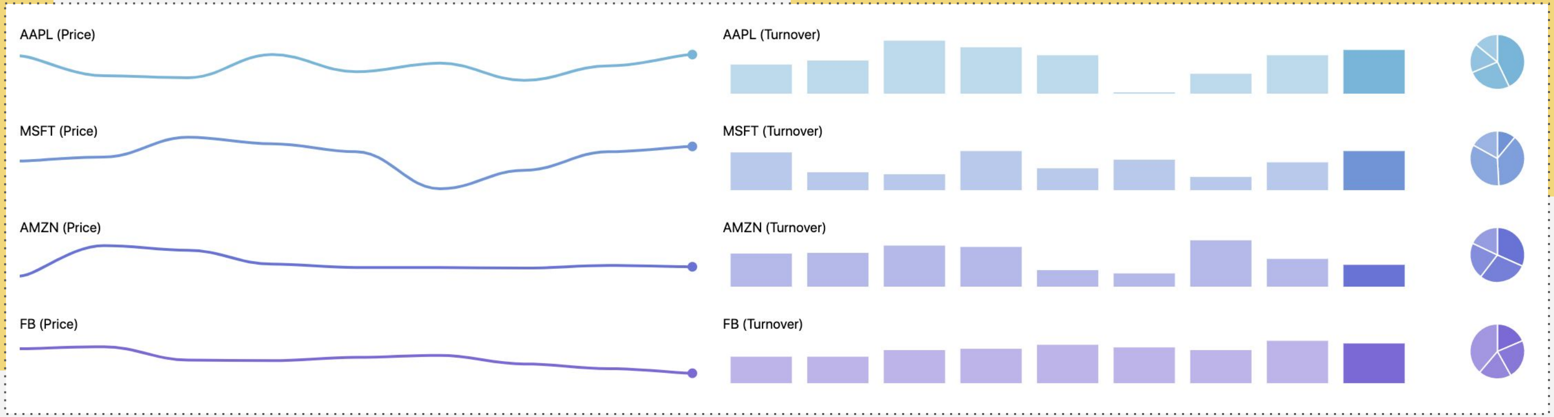
A visual guide for the **placement** of elements composing the visualisation.

LAYOUT

The **format** and **symmetry** of the visualisation according to the volume of data.

(e.g. include one or more visualisations in the canvas, like in a comic)

some relations might not be evident in multiple smaller visualisations!



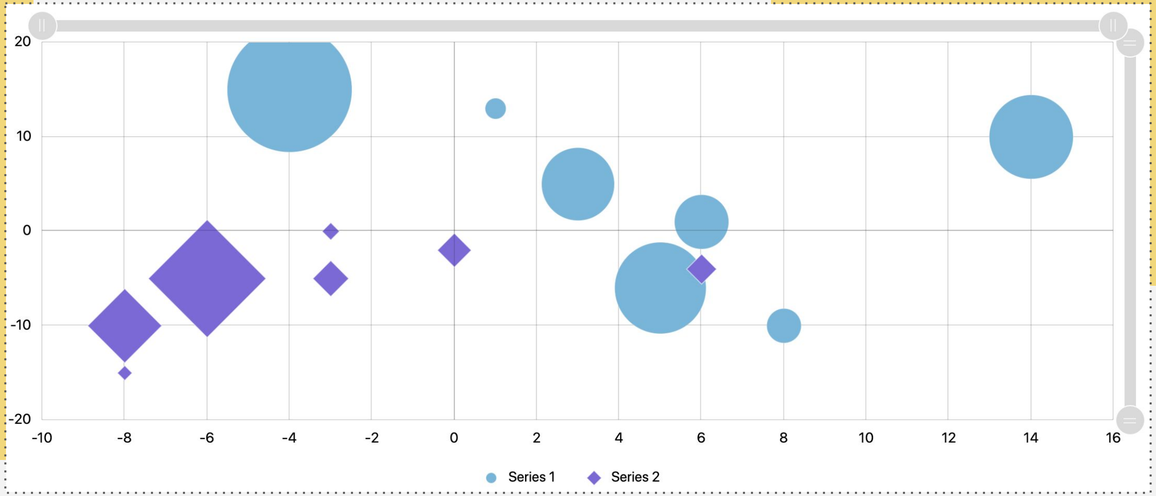
SHAPE

require supporting text to explain the meaning and may be less immediate

Can be **realistic** (e.g. icons) or abstracts (circles, rectangles)

Shapes and glyphs demand more attention (icons can mitigate the effort)

Lines are particularly effective shapes that can demonstrate a fact (e.g. a trend)



COLOR

may require high resolution and attention for colorblindness

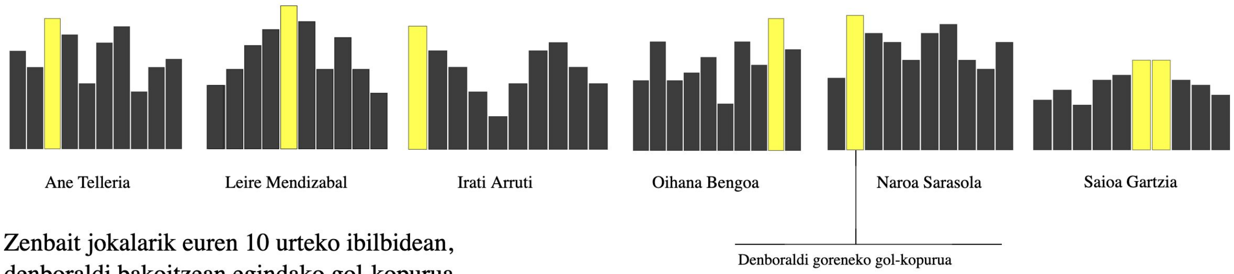
works better with quantitative data dimensions

Used in big datasets to **distinguish** elements or patterns, it's less useful in small datasets.
In smaller datasets **one** color is sufficient. When data dimensions are only two, **2-dimension color palette** is sufficient.

Color differences are detected in <200 milliseconds (**preattentive** perception)

Colors must be **natural** (e.g. those found in nature)

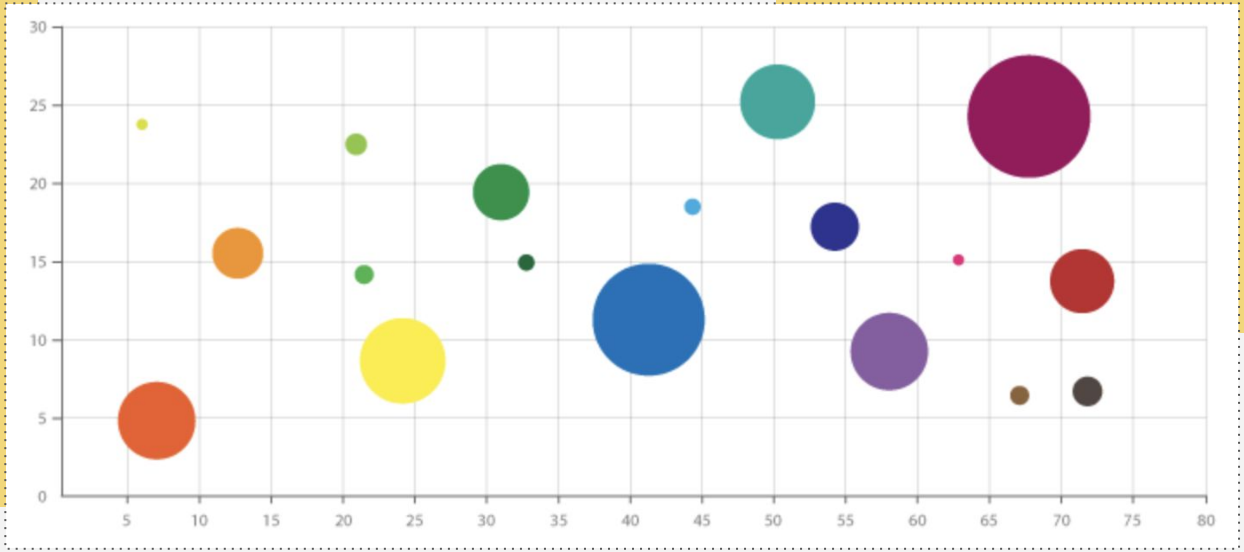
Luminosity immediately highlights relevant patterns (bright colors pop out, while dark colors recede)



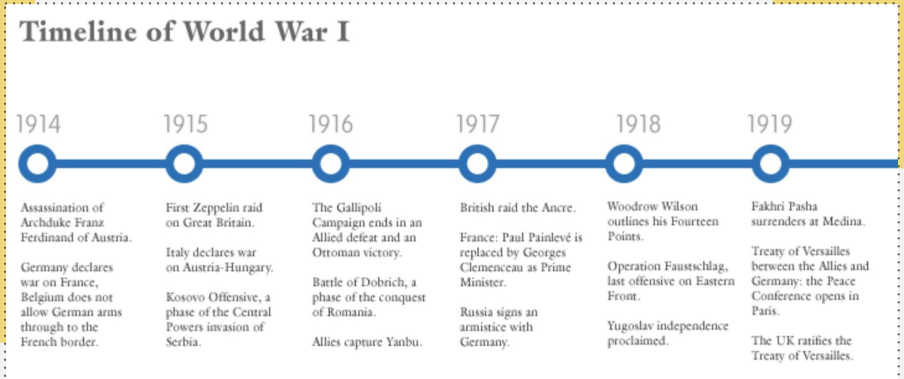
Zenbait jokalarik euren 10 urteko ibilbidean, denboraldi bakoitzean egindako gol-kopurua.

SIZE

Identify **variations** between similar elements by size (regardless other dimensions) is the quickest way.



TYPOGRAPHY



Can be a primary aspect (e.g. wordle) or secondary along with other dimensions (e.g. size).

INTERACTION OF VISUAL DIMENSIONS

Visual dimensions can be summed up to convey the same data dimension (**redundancy** helps to understand faster) and can **interact** with each other in different ways.

CONTEXTUAL ASPECTS

intuitive but not always carefully constructed

can be replaced by animations, vertical/horizontal layout help to contextualise sequences

The **placement** of data in maps (geographical or not, e.g. airplane seats) help to familiarise with contents.

LOCATION

Construct **connections** between data rather focusing on data points.

NETWORK

An axis-based visualisation of other dimensions (e.g. timeline).

SEQUENCE

VISUAL ASPECTS + CONTEXT = VISUAL PATTERN

THE ULTIMATE GOAL:
IDENTIFY VISUAL PATTERNS

By means of graphical and contextual aspects, what should emerge from visualisations is mainly a **visual pattern**. Several co-occurring dimensions can contribute to highlight patterns.

VISUAL PATTERNS

size and colors work better with 2-dimensional data

Contrast boolean dimensions (e.g. sold VS unsold paintings)

DIFFERENTIATION

size does not work well when differences are not significative

A gradation of quantitative information (e.g. temperatures).

GRADATION

color and location highlight anomalies

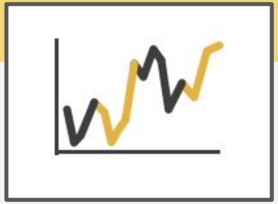
Data points that do not follow a pattern (this are the interesting data)

ANOMALIES

WHAT TO LOOK FOR IN PATTERNS

Overall picture of data over time

TRENDS



Some sample of overall data

FEATURES

Some data point in the dataset

OUTLIERS

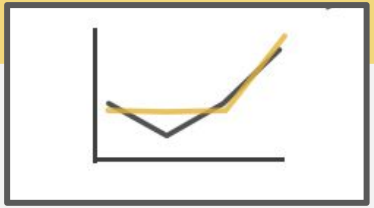
WHAT TO LOOK FOR IN PATTERNS

Presence of an attribute in the dataset

DISTRIBUTION

Relations between attributes

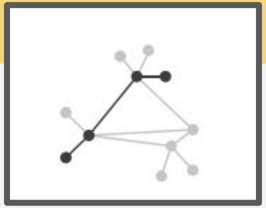
CORRELATION



Common features of data points

SIMILARITY

WHAT TO LOOK FOR IN PATTERNS



Relations between data points or position of clusters

PATHS

VISUAL THINKING

Visual patterns stimulate three types of thinking that cannot be achieved by looking at numbers only: **recursive, relational, functional**

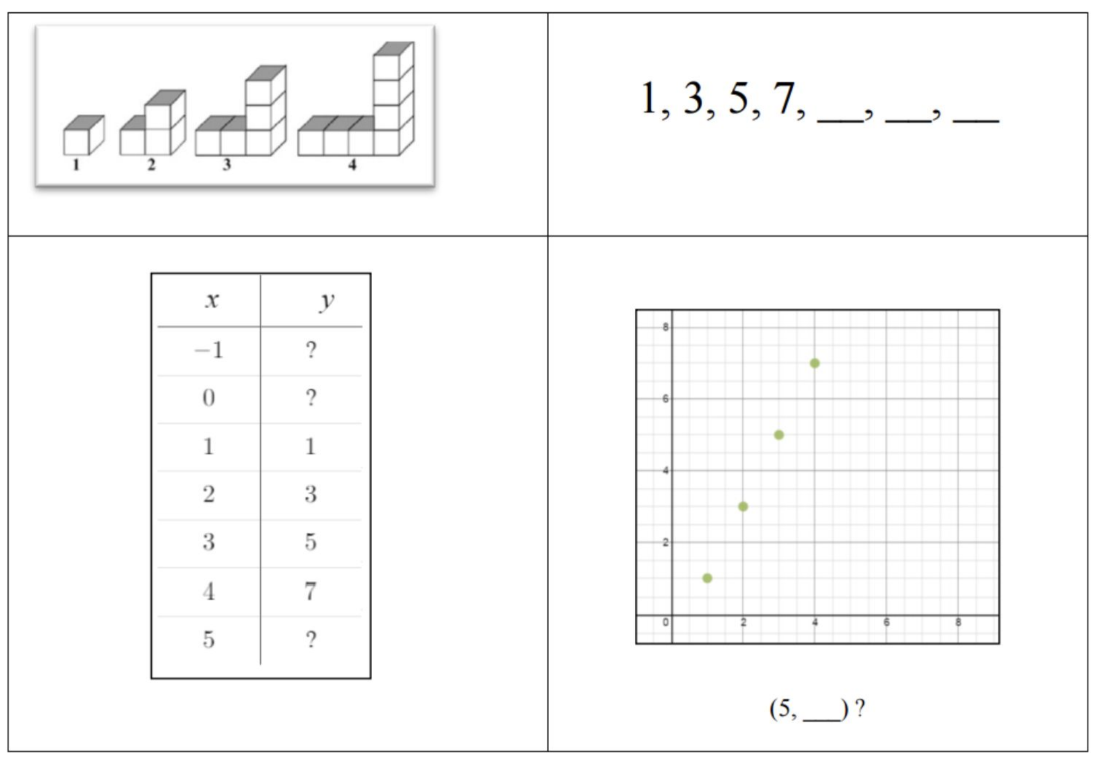
AN EXAMPLE OF VISUAL THINKING

Look at the sequence of numbers (top-right).

Relational thinking is possible when visual aids can be compared (top-left).

Recursive thinking is when you understand how many cubes to add to the next (fifth) L-shape.

Functional thinking is when you know how to compute the number of cubes of the, say, 10th L-shape without drawing it.



x	y
−1	?
0	?
1	1
2	3
3	5
4	7
5	?

Thinking about thinking: Visual patterns
https://problemproblems.files.wordpress.com/2016/02/on-visual-patterns.pdf

DATA VISUALISATION BEST PRACTICES

OBJECTIVES OF DATA VISUALIZATION

Effective visualisations are able to **identify**, **summarise and prioritise** information.

These can be aimed at **exploring** a dataset or to **discover** knowledge

KNOWLEDGE DISCOVERY

some visualisations provide an intuitive validation but do not allow a precise assessment

The designer has a **question** to answer, and a **hypothesis** and wants to demonstrate.

HYPOTHESIS

The validation consists of the **visual evidence** of a trend, a behaviour, or a relationship.

VALIDATION

YOU CAN SEE EXPLORATION AS A FORM OF DISCOVERY

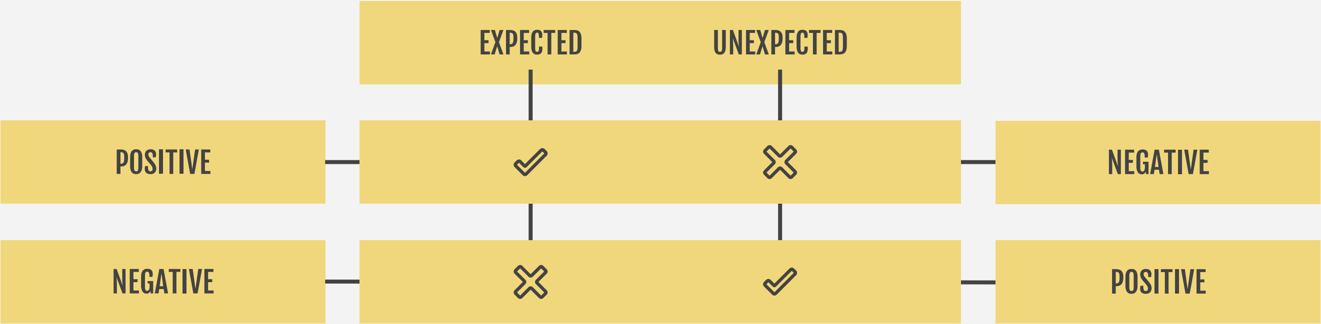
Select data that **may characterise** the dataset (the added value).

A PRIORI INTEREST

Find what data are **available, significant, interesting and at scale** for further analysis.

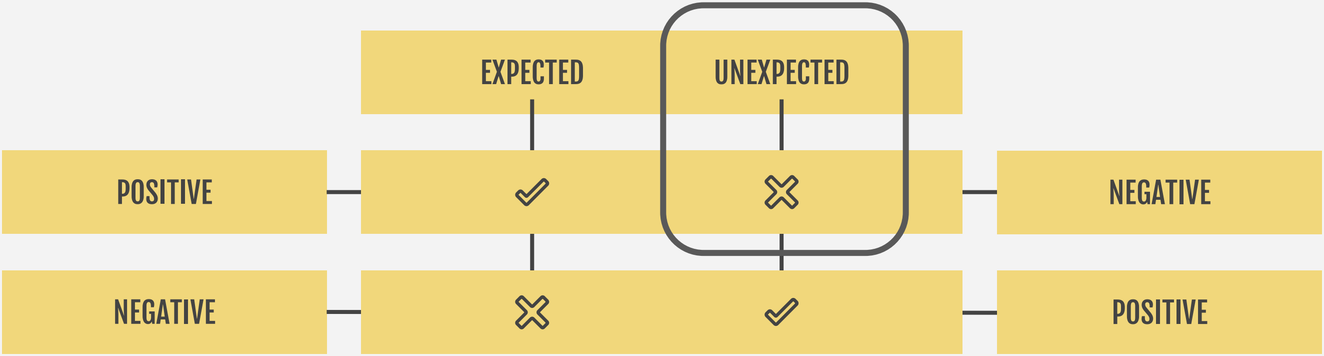
DATA DIVERSITY

DISCOVERY MATRIX



Outcomes of data analysis/discovery techniques fall into two categories: expected/unexpected, positive/negative

DISCOVERY MATRIX



Unexpected negative results are the most insightful patterns that can effectively support decision-making

HOW GOOD VISUALISATIONS SHOULD BE

NOVEL

A byproduct of informative and efficient visualisations

INFORMATIVE

Make explicit an **intended message** and the **context of use** of data.

EFFICIENT

Free from non informative data (noise). The learning time to understand the graphics is low.

* INFORMATIVE

The knowledge conveyed, telling a **story** or answering a **question**. It corresponds to the **title** of the graph.

INTENDED MESSAGE

How the knowledge is going to be used by users, such as a **presentation** of known information, the **exploration** or the **discovery** of unknown information.

CONTEXT OF USE

WHAT VISUALISATIONS SHOULD ALWAYS ACHIEVE

SHOW PATTERNS

Arrange data in a way that reveals relations between data.

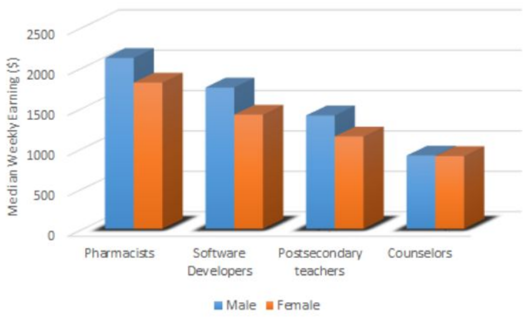
ACCESS PERTINENT DATA

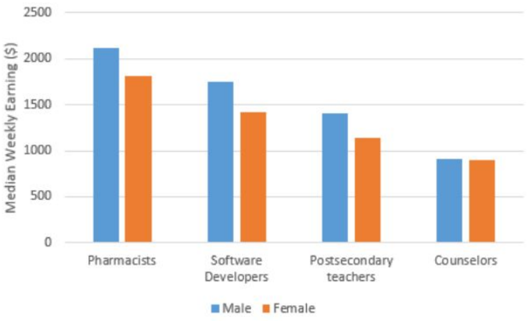
Strip away irrelevant information so that pertinent data is highlighted.

GET EXCEPTIONAL DATA

Relevant, exceptional data should be easy to read.

EDWARD TUFTE'S PRINCIPLES

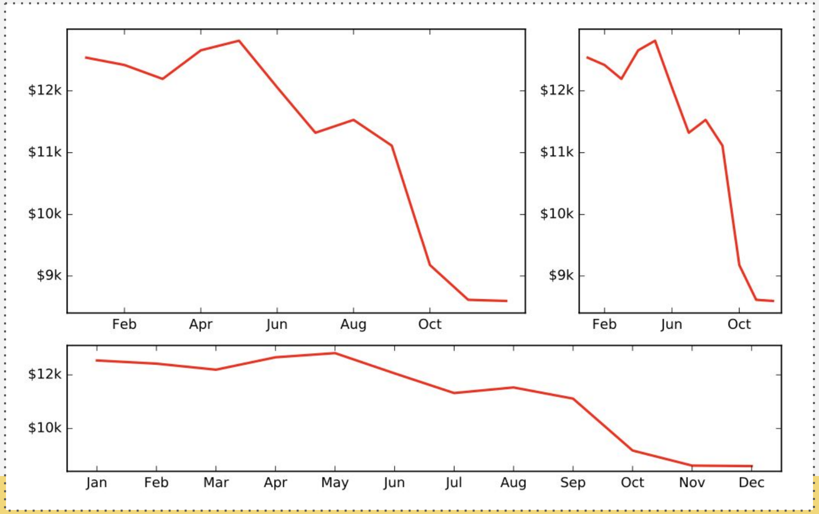




The ink used to represent data should be minimal, so that the intended message is clear.

MINIMIZE DATA-INK RATIO

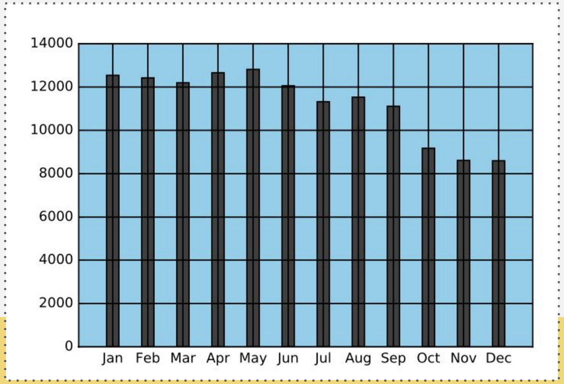
EDWARD TUFTE'S PRINCIPLES

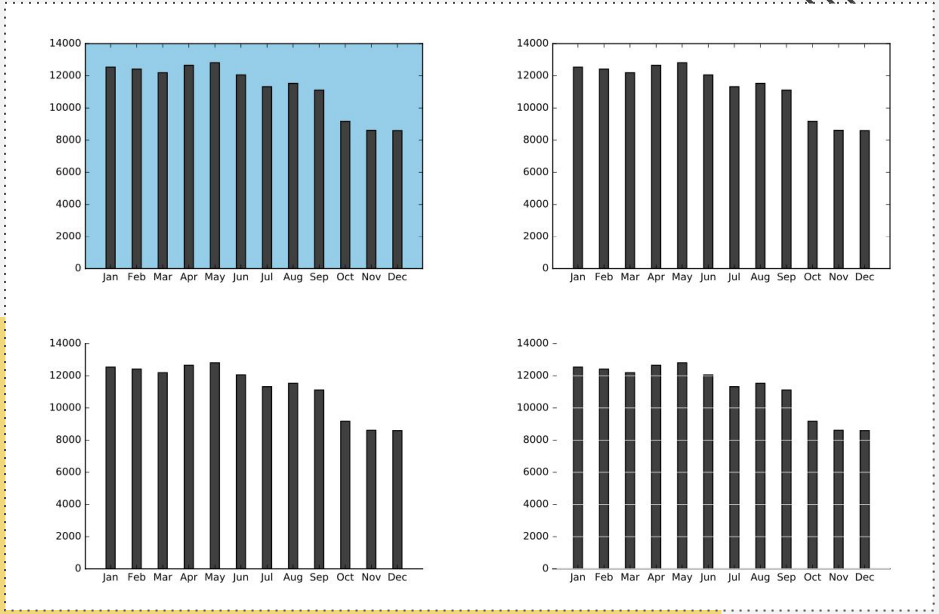


Omitting data or changing proportions in the visualisation can tell different stories.

MINIMIZE THE LIE FACTOR

EDWARD TUFTE'S PRINCIPLES

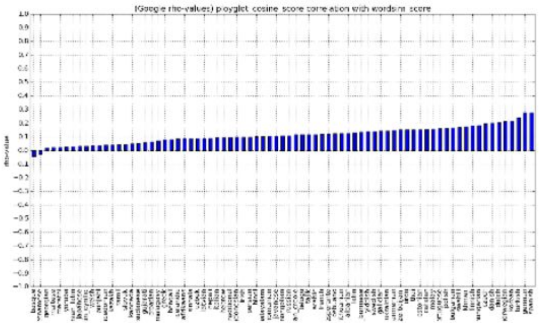


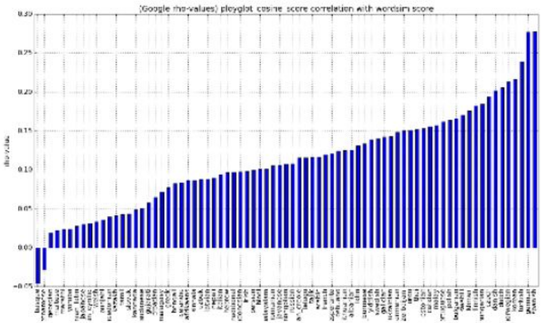


Visual elements should not hide the data.

MINIMIZE CHARTJUNK

EDWARD TUFTE'S PRINCIPLES





Data should be scaled so as to fill all the space allotted to it. Labels should be high resolution.

PROPER SCALES AND LABELS

EDWARD TUFTE'S PRINCIPLES

Use colors to distinguish values and represent **difference** - rather than representing linear, numerical scales.

EFFECTIVE USE OF COLOR

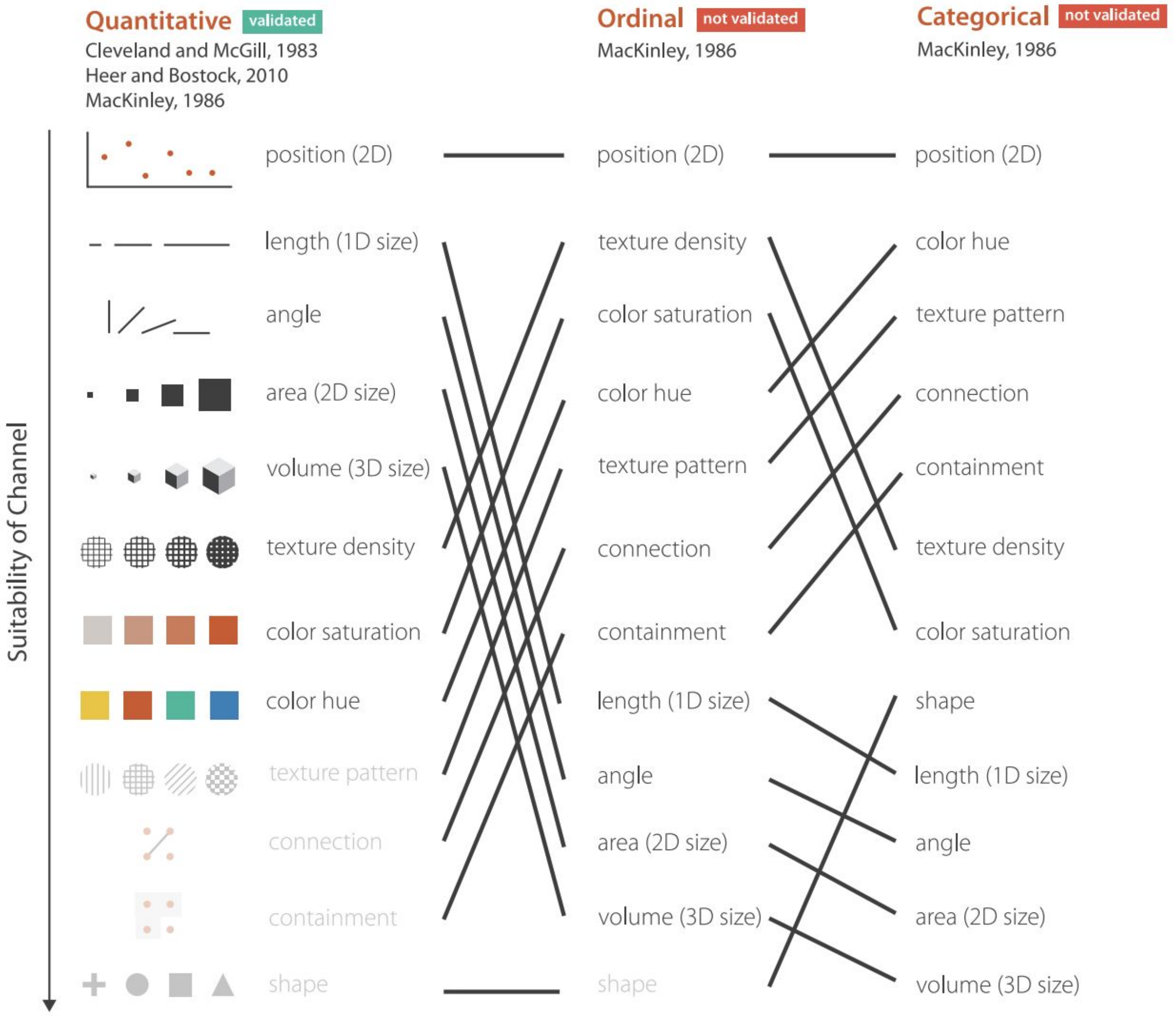
Use multiple smaller charts to facilitate comparison (while losing the big picture)

REPETITION

VISUAL MAPPING

The process of associating visual aspects (such as preattentive properties) to data attributes is called **visual mapping**.

VISUAL MAPPING



THE TOP BEST PRACTICES

Design several **different visualisations** to understand which one works better.

Redundancy of encodings (e.g. using color and shape for one dimension) helps users' perception.

Use **familiar** colors, icons, layouts as visual hooks and reduce users' effort in understanding the intended message.

ELECTRONIC PUBLISHING
AND DIGITAL STORYTELLING
Lesson 2

AN OVERVIEW OF DATA VISUALISATION SOLUTIONS

From simple to complex plots, an
inspirational journey rather
than a complete survey

SUMMARY OF TYPES OF DATA VISUALISATIONS

TYPES OF DATA VISUALISATION
static
animated
interactive
direct manipulation

VISUALISATIONS BASED ON DATA ATTRIBUTES
quantitative data
categorical data
mixed data

TASKS
exploration
presentation
discovery

AFFORDANCE
identification
comparison
summarisation

Visualization atlas
https://visualisation.trespass-project.eu/wp-content/uploads/2016/10/Visualisation-atlas.pdf

Dataviz catalogue
https://datavizcatalogue.com/

MOST COMMON VISUALISATION TYPES

CHARTS

TABLES

GRAPHS

MAPS

NETWORK GRAPHS

TABLES

Based on carbon-12

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.00794 H 1																	4.0026 He 2
2	6.941 Li 3	9.0122 Be 4											10.811 B 5	12.011 C 6	14.0067 N 7	15.9994 O 8	18.9984 F 9	20.1797 Ne 10
3	22.9898 Na 11	24.3050 Mg 12											26.9815 Al 13	28.0855 Si 14	30.9738 P 15	32.066 S 16	35.4527 Cl 17	39.948 Ar 18
4	39.0983 K 19	40.078 Ca 20	44.9559 Sc 21	47.88 Ti 22	50.9415 V 23	51.9961 Cr 24	54.9381 Mn 25	55.847 Fe 26	58.9332 Co 27	58.69 Ni 28	63.546 Cu 29	65.39 Zn 30	69.723 Ga 31	72.61 Ge 32	74.9216 As 33	78.96 Se 34	79.904 Br 35	83.80 Kr 36
5	85.4678 Rb 37	87.62 Sr 38	88.9059 Y 39	91.224 Zr 40	92.9064 Nb 41	95.94 Mo 42	98.9063 Tc 43	101.07 Ru 44	102.9055 Rh 45	106.42 Pd 46	107.8682 Ag 47	112.411 Cd 48	114.82 In 49	118.710 Sn 50	121.75 Sb 51	127.60 Te 52	126.9045 I 53	131.29 Xe 54
6	132.9054 Cs 55	137.327 Ba 56	138.9055 La a 57	178.49 Hf 72	180.9479 Ta 73	183.85 W 74	186.207 Re 75	190.2 Os 76	192.22 Ir 77	195.08 Pt 78	196.9665 Au 79	200.59 Hg 80	204.3833 Tl 81	207.2 Pb 82	208.9804 Bi 83	208.9824 Po 84	209.9871 At 85	222.0176 Rn 86
7	223.0197 Fr 87	226.0254 Ra 88	227.0278 Ac b 89	261.1087 Unq 104	262.1138 Unp 105	263.1182 Unh 106	262.1229 Uns 107	Uno 108	Une 109									

a n = 6	140.115 Ce 58	140.9077 Pr 59	144.24 Nd 60	146.9151 Pm 61	150.36 Sm 62	151.965 Eu 63	157.25 Gd 64	158.9253 Tb 65	162.50 Dy 66	164.9303 Ho 67	167.26 Er 68	168.9342 Tm 69	173.04 Yb 70	174.967 Lu 71
b n = 7	232.0381 Th 90	231.0359 Pa 91	238.0289 U 92	237.0482 Np 93	244.0642 Pu 94	243.0614 Am 95	247.0703 Cm 96	247.0703 Bk 97	251.0786 Cf 98	252.0829 Es 99	257.0951 Fm 100	258.0986 Md 101	259.1009 No 102	260.1053 Lr 103

IPTE9007

0.103	0.176	0.387	0.300	0.379	0.276	0.179	0.321	0.192	0.250
0.333	0.384	0.564	0.587	0.857	1.064	0.698	0.621	0.232	0.316
0.421	0.309	0.654	0.729	0.228	0.529	0.832	0.935	0.452	0.426
0.266	0.750	1.056	0.936	0.911	0.820	0.723	1.201	0.935	0.819
0.225	0.326	0.643	0.337	0.721	0.837	0.682	0.987	0.984	0.849
0.187	0.586	0.529	0.340	0.829	0.835	0.873	0.945	1.103	0.710
0.153	0.485	0.560	0.428	0.628	0.335	0.956	0.879	0.699	0.424

Simple and effective. They allow a **precise comparison** of heterogeneous data.

If you want to demonstrate and convince anybody of your scientific results, you are very likely to need a table.

You can always start from a table and then experiment more sophisticated and suitable visualisations

TABLES

Good for **multivariate values**, heterogeneous data (categorical, numeric, ordinal, etc.)

The **order of rows** should invite comparison (e.g. apply descending order to a column).

The **order of columns** highlight importance.

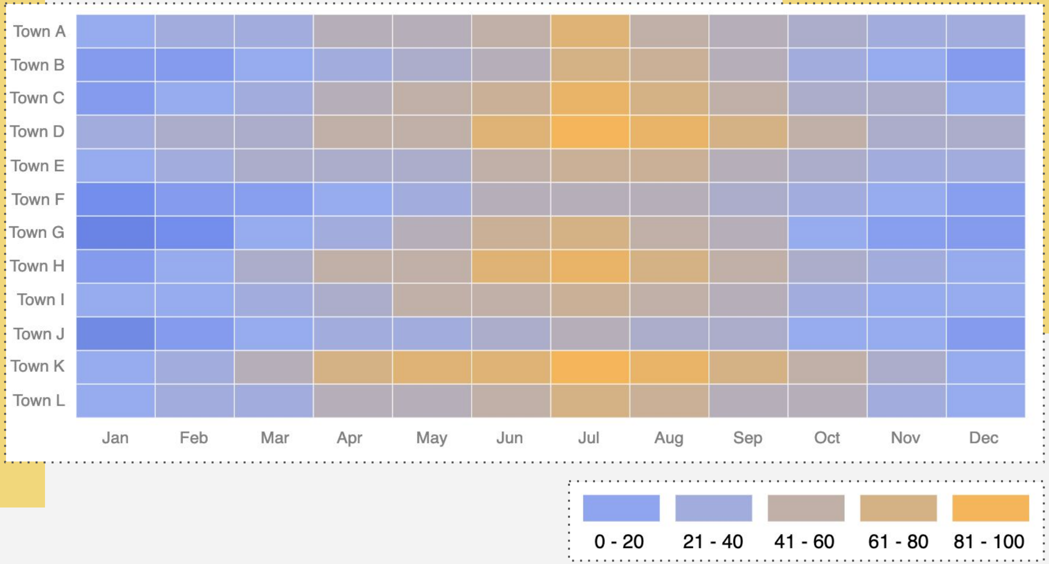
Can include icons, colors, changes in typography to add **emphasis**.

Frequency (Hertz)	Scientific Pitch	Helmholtz/ German	Octave Name	Pipe Length
8.176	C-1	CCCC		64'
16.352	C0	CCC	sub-contra	32'
32.703	C1	CC	contra	16'
65.406	C2	C	great	8'
130.81	C3	c	small	4'
261.63	C4	c'	1-line	2'
523.25	C5	c''	2-line	1'
1046.5	C6	c'''	3-line	1/2'
2093.0	C7	c''''	4-line	
4186.0	C8	c'''''	5-line	
8372.0	C9	c''''''		

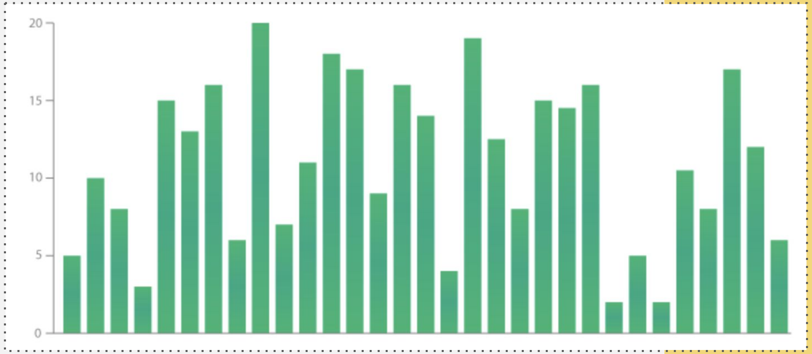
HEATMAPS

A special type of tables where **multivariate data** are mapped to colours. The map highlights the **variance across variables**, patterns like correlation.

While it gives a good overview, it's hard to get a precise comparison of underlying numerical values.



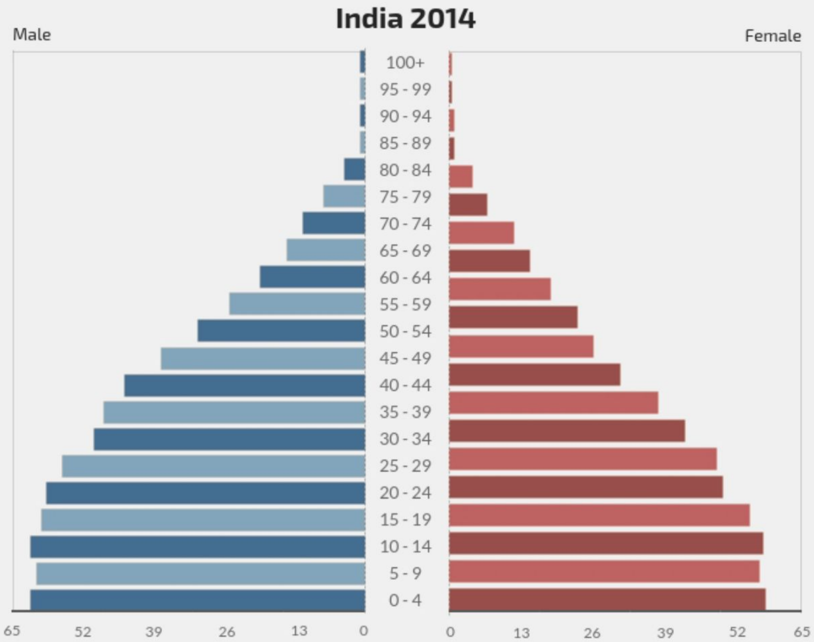
BAR CHARTS



Often used with **bivariate** values, especially to combine a categorical value and a numeric value, and **highlight differences between categories** (their occurrence in the dataset).

Categorical data need to be **independent** (if there are changes in one of those, these do not affect the other categories).

BAR CHARTS
>> PYRAMIDS

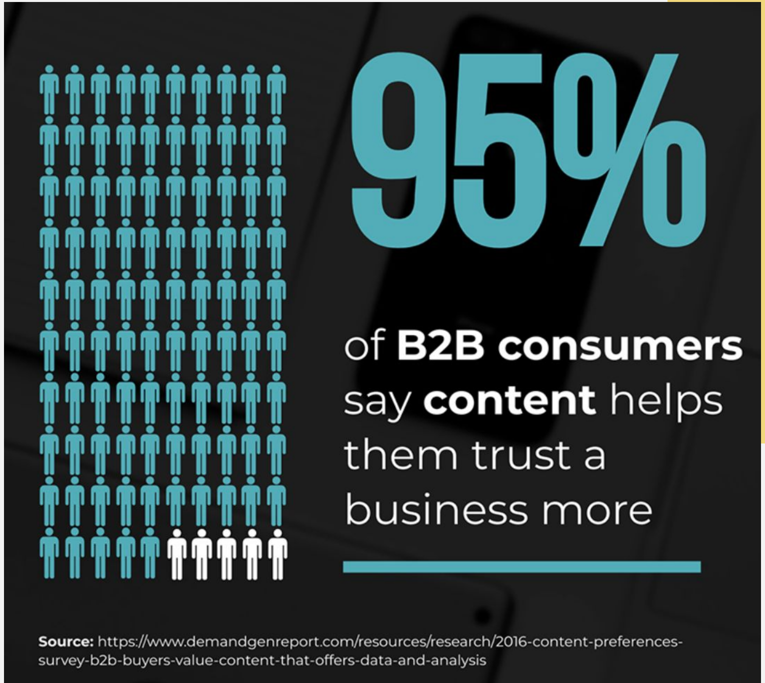


Source: https://simconblog.wordpress.com/2015/08/19/insurance-industry-analysis/

Vertical or horizontal layout usually does not matter much.

In few cases, sorting variables by ascending/descending order (e.g. by age) helps to **grasp a better overview in the comparison**.

BAR CHARTS
>> PICTOGRAMS

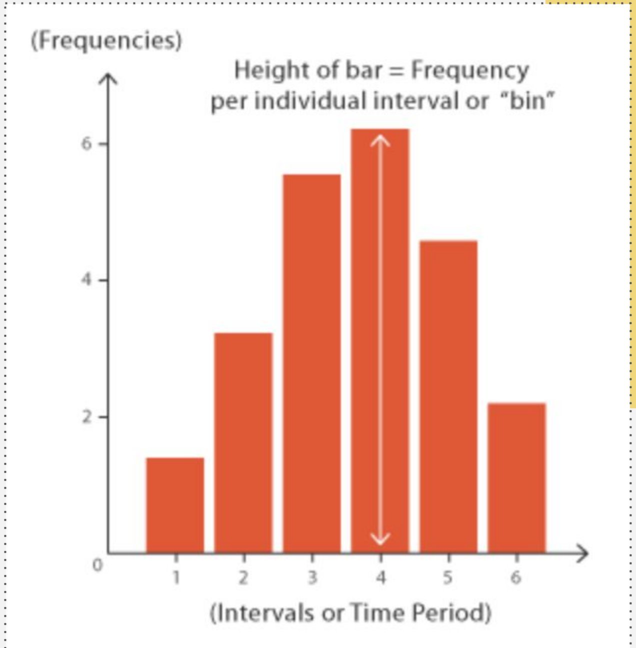


A special type of bar charts where one of the axis is replaced by a certain number of icons.

The number can be representative of the actual value on that axis, or being a proportional approximation.

Particularly engaging when **comparing categories**.

BAR CHARTS
>> HISTOGRAMS

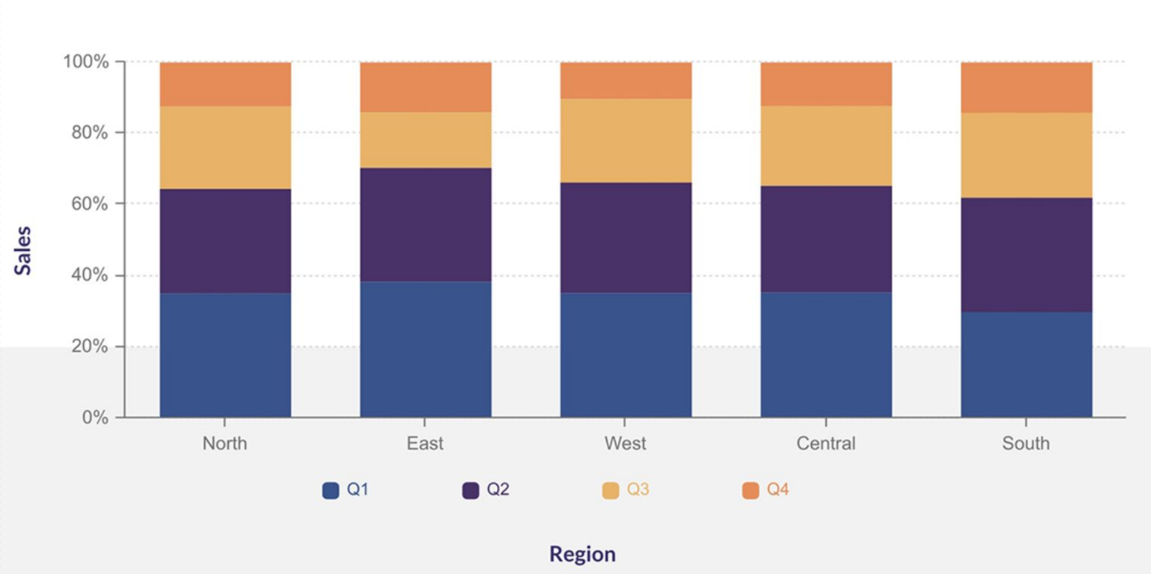


A special type of bar chart where categories are replaced by numeric continuous data (e.g. time).

Usually used to show **distribution** of one variable in a dataset, meaning where that variable concentrates, whether there are gaps, and what are the extreme points of occurrence.

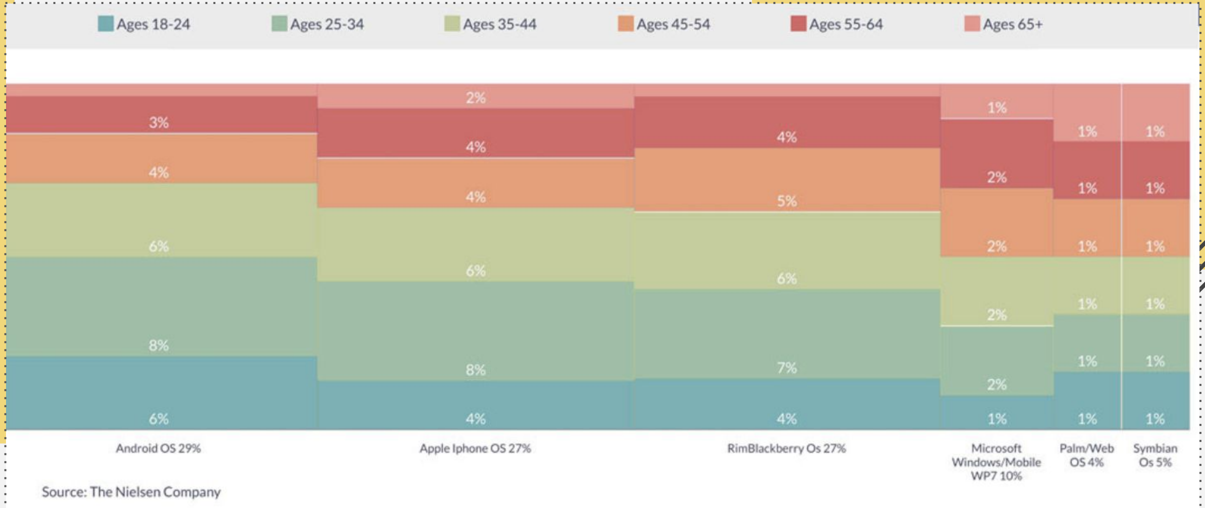
BAR CHARTS
>> STACKED BAR CHARTS

A special type of bar chart where **multiple categories are grouped and compared** against a numeric value.

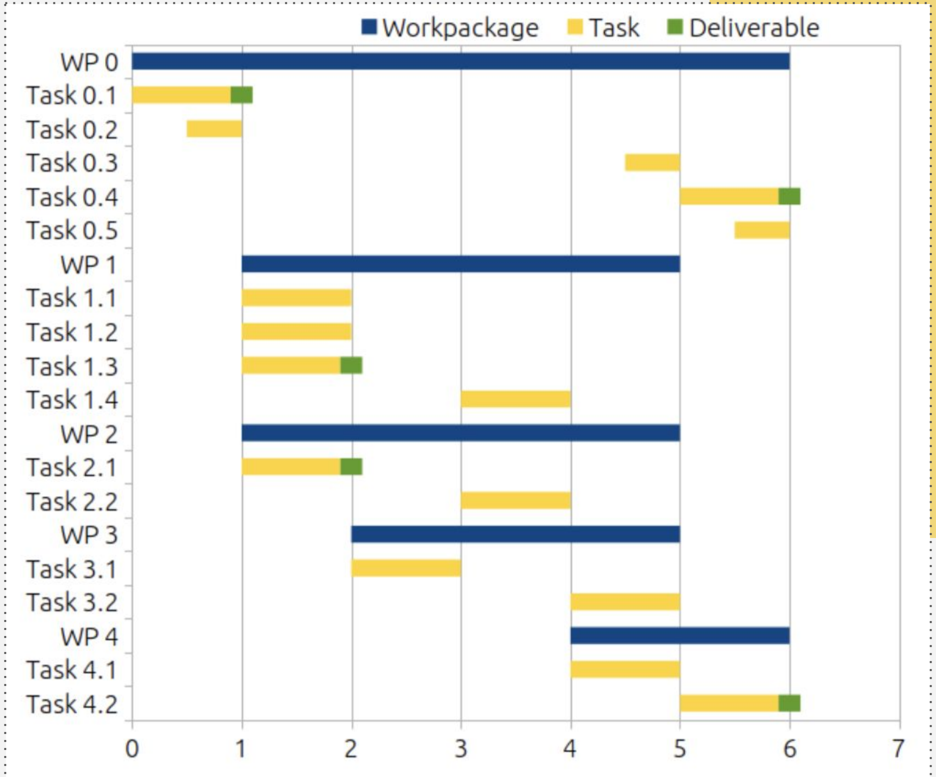


BAR CHARTS
>>MOSAIC OR MARIMEKKO CHARTS

A special type of bar chart where you can **compare multivariate categories against multiple variables**.

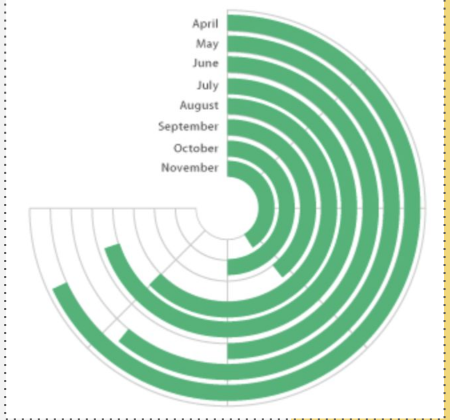


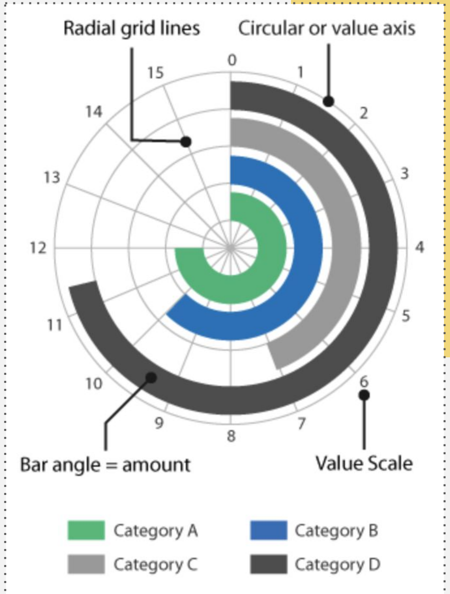
BAR CHARTS
>> GANTT CHART



Halfway between a table and a bar chart, gantt charts allow to see **dependencies between variables** (e.g. activities) with respect to another (e.g. time).

BAR CHARTS
>> RADIAL BAR CHARTS

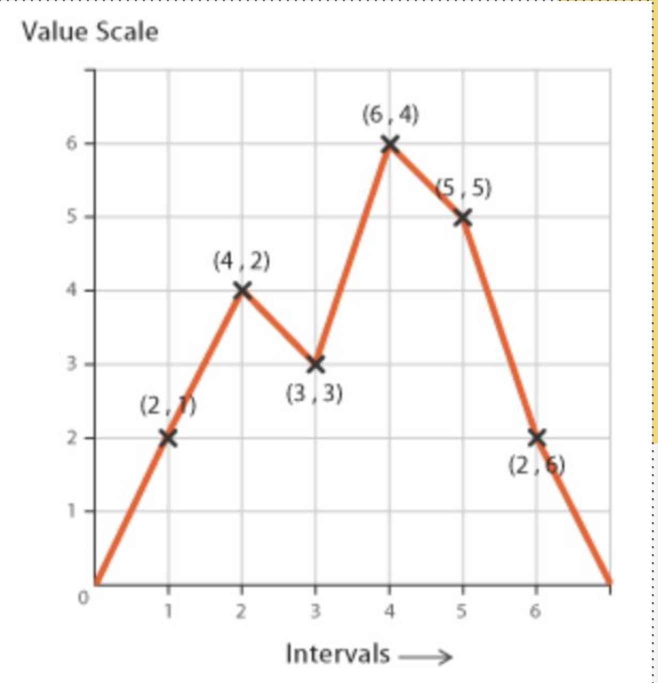




A bar chart where on of the axes is radial, meaning the cartesian system is substituted by a polar system.

Reading these charts can be misleading, since the outer bars tend to be seen as bigger/longer.

DOT AND LINE PLOTS

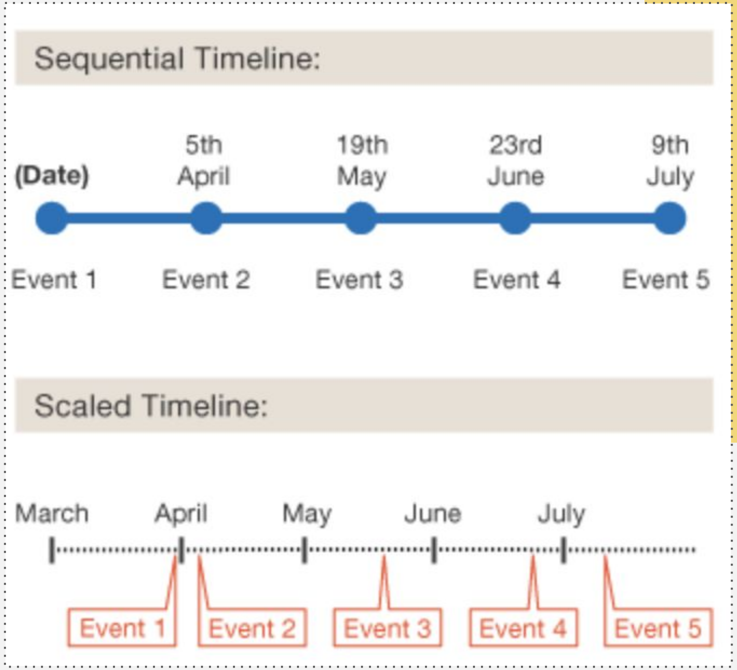


Line plots are used to show **dependent data**, usually between a categorical value and a numeric one (when changes occur in one variable, these affect the other variables).

For each category is shown a point value.

Useful to show **trends**.

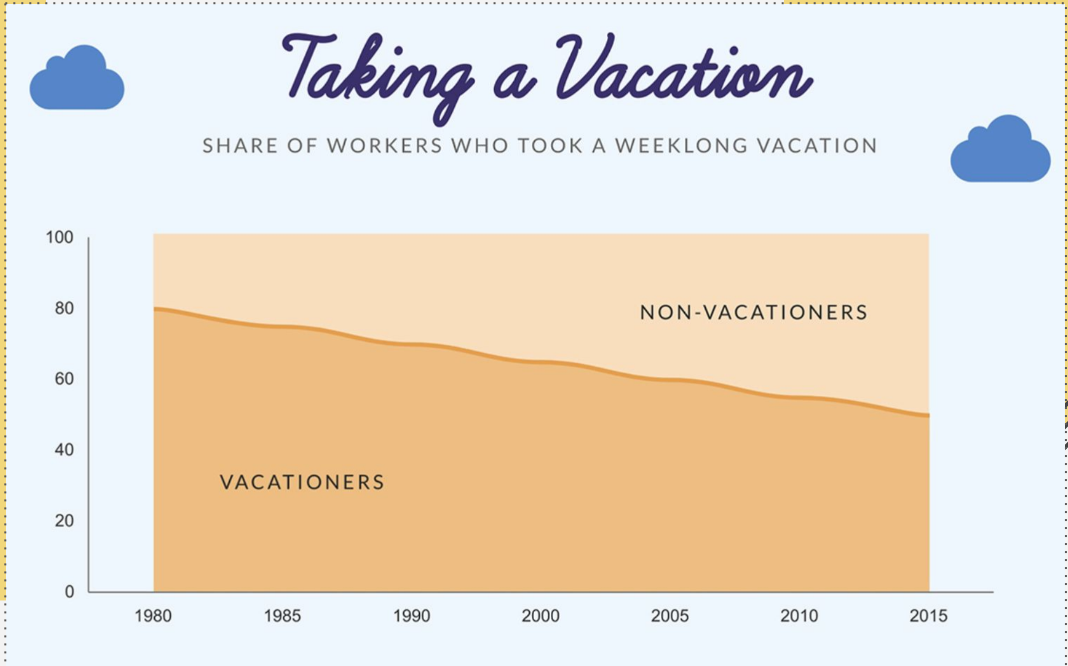
TIMELINES



A list of data points displayed on a line, **according to a scale or in sequence**.

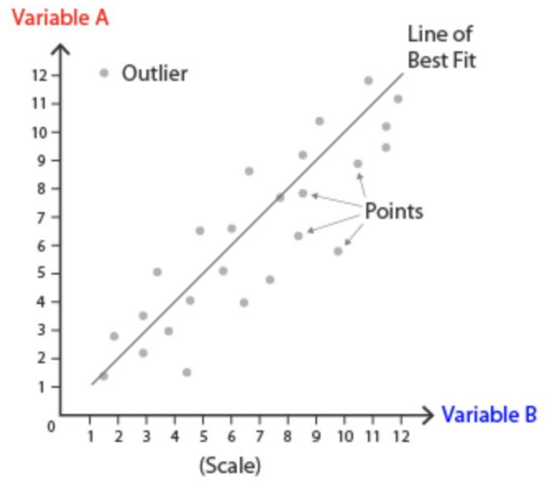
In a scale-based timeline it's possible to see the distribution of data points.

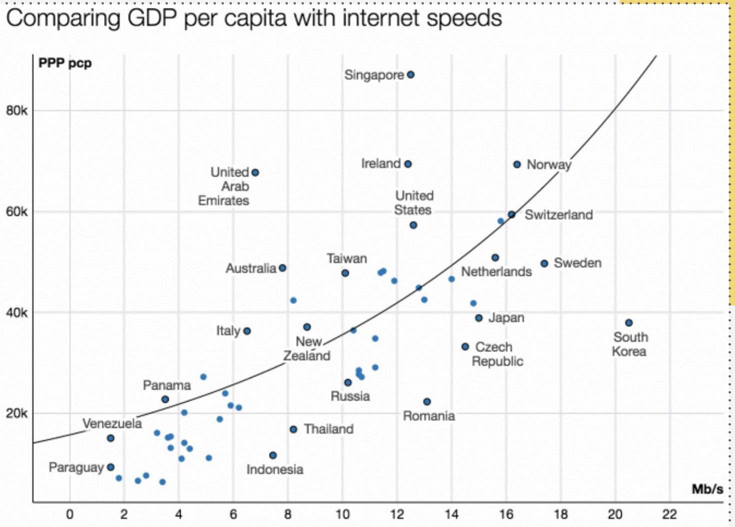
AREA CHARTS



A special type of line chart where areas divided show a **comparison of variables and their dependency** (e.g. inverse proportion) and of the “size” of a phenomenon.

SCATTER PLOTS

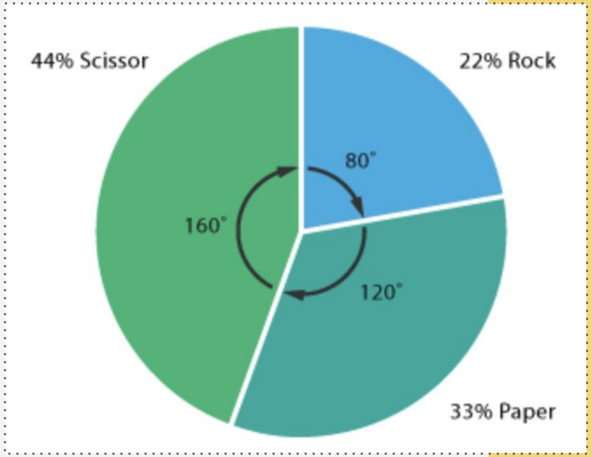


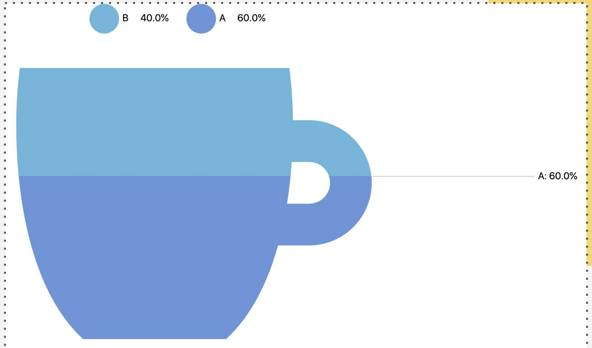


Allows to show the **correlation between two independent numeric variables** and the **outliers**.

Patterns (highlighted by a best fit line) include: **positive** (values increase together), **negative** (one value decreases as the other increases), **null** (no correlation), **linear**, **exponential** and **U-shaped**.

PIE CHARTS



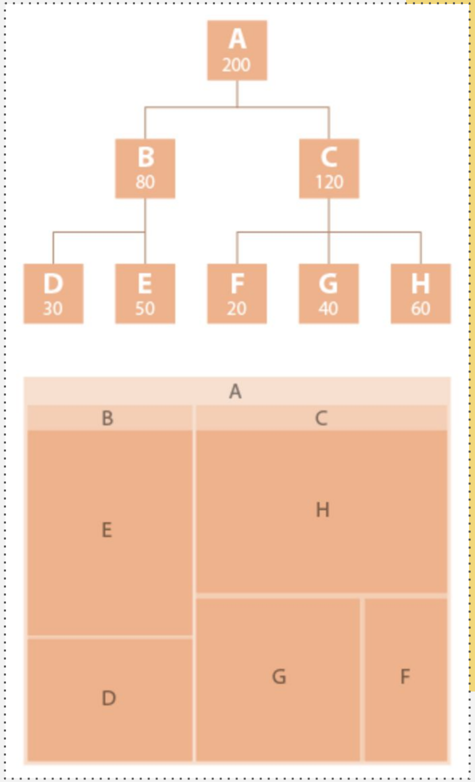


Circles cut in segments to **show parts and their proportion** with respect (1) to the whole, and (2) to other parts.

Used to show **univariate dependant data**.

Circles can be replaced with pictograms.

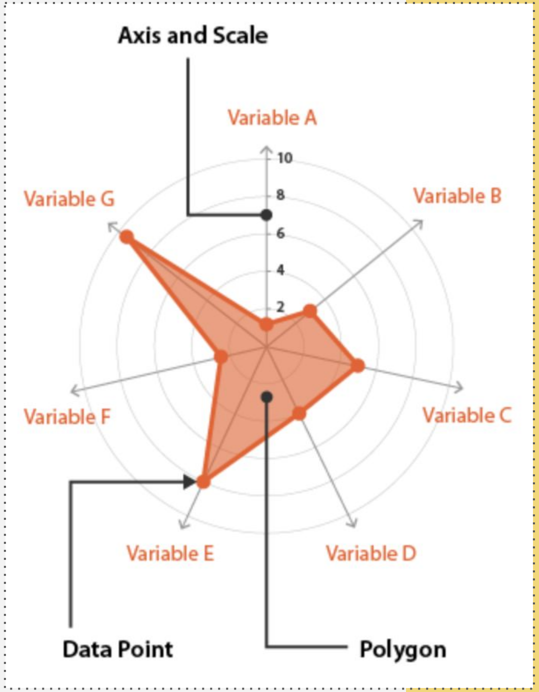
TREEMAPS



Like a pie chart, a treemap allows to **show proportions** of categorical variables in a dataset in a space-efficient way.

It is a different way to visualise tree diagrams (even though the hierarchy is not shown).

RADAR CHARTS



Compare multiple quantitative variables and give them a unique shape.
Useful to **show similarities between variables and outliers**.

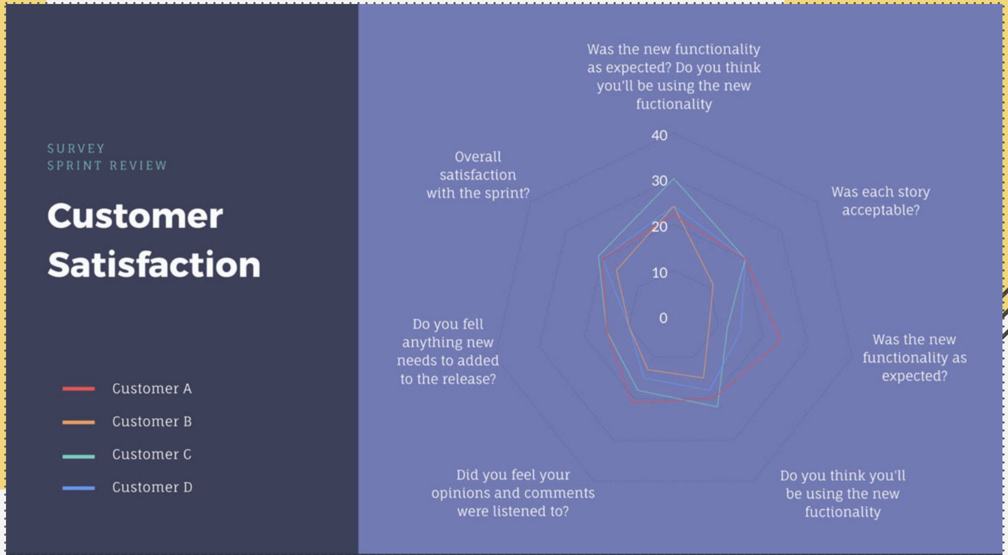
Each **ordinal variable** (e.g. a question in a survey) has one axis that starts from the center. All axes are displayed radially, and a grid (e.g. the circles) connects the axes for the sake of readability.

The polygon represents the collection of variables grouped by another variable (e.g. the person that answered the survey).

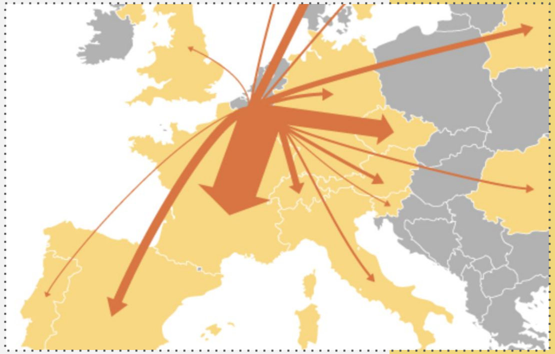
RADAR CHARTS

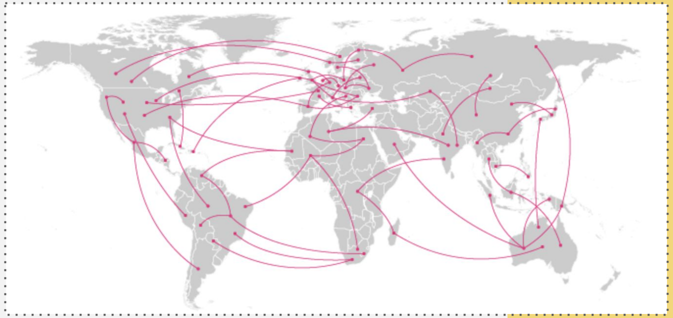
Having multiple polygons helps to compare variables and show outliers.

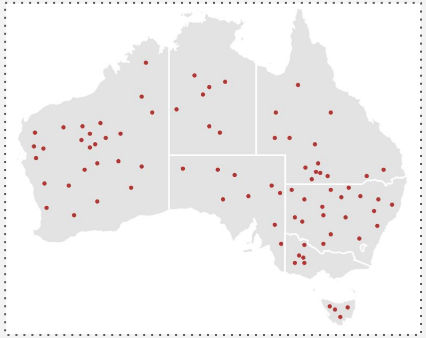
But too many polygons and too many axes make the chart **too cluttered** and unreadable. Plus it is not easy to compare values across each variable.

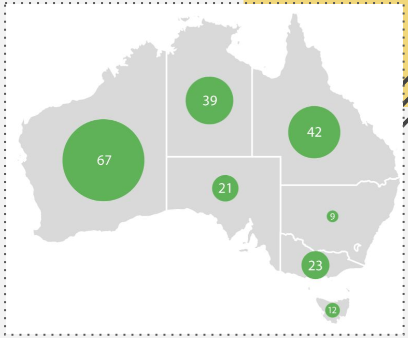


MAPS





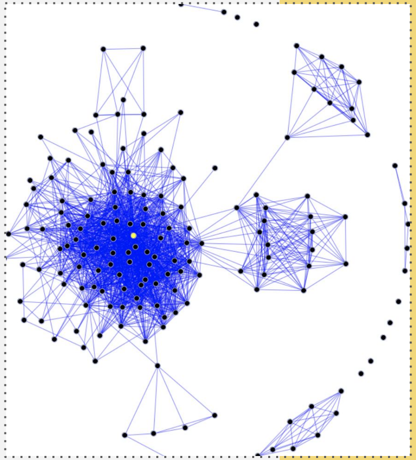


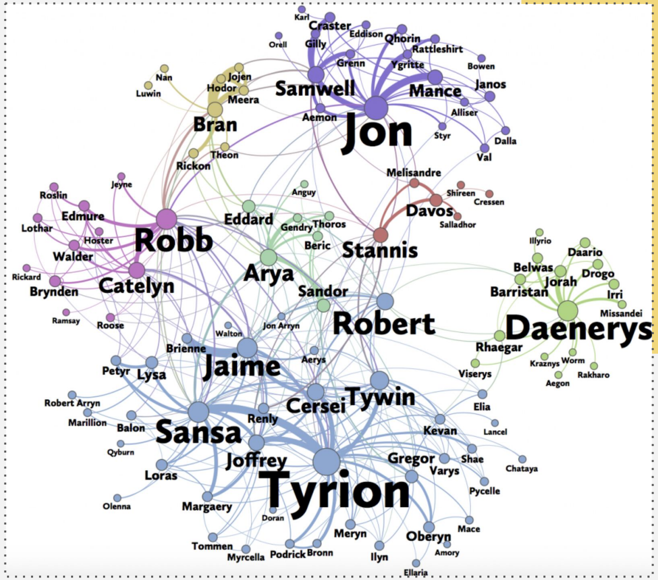


If one of the data variables regards a geolocation, geographical maps can be more intuitive than other axes systems.

Maps can be **combined with other visual strategies** (e.g. flows, trajectories, dots, bubbles).

NETWORKS





Categorical data are mapped to nodes and relations (co-occurrence) become links between nodes.

The number of links related to a node is important.

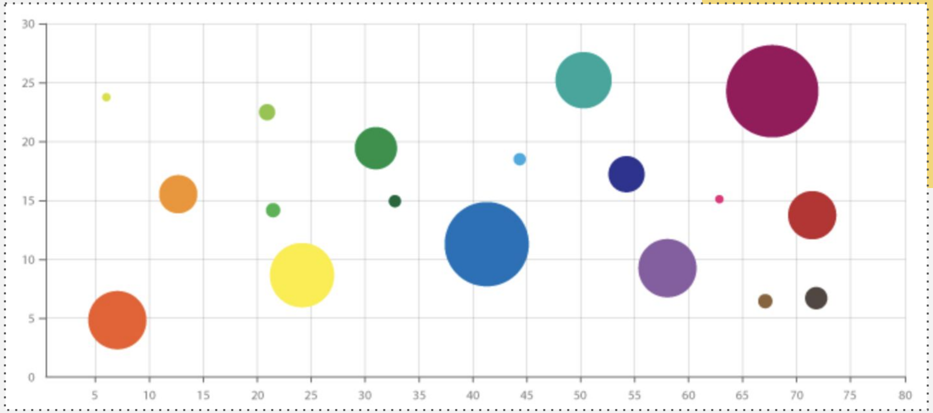
Nonetheless, is often the **location** of clusters of nodes to be more relevant, showing **similarities between groups of data**.

The drawback of networks is that they are **unreadable** (label placement problem) and can only intuitively suggest patterns.

BUBBLE CHARTS

Halfway between a scatterplot and an area chart, it allows to compare **multivariate** values in a cartesian system, show relationships such as **correlation**, by means of position and proportion.

There are two numeric variables (x,y axes), the points are labelled with a categorical data, and the area represents another numeric variable. Colors may represent another categorical value or sum up to the label.



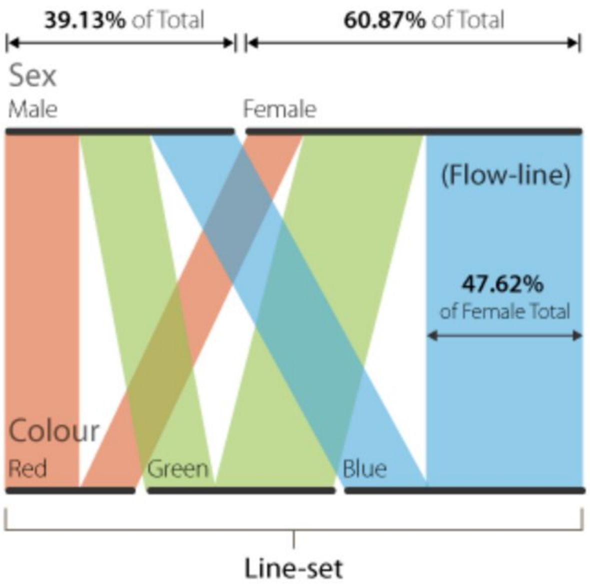
WORDLE

A tag explorer where the **size** of words represents a univariate quantitative aspect of categories.

The **layout** is based on a randomised greedy algorithm that positions words so that they fill empty spaces and words do not overlap.

Tag clouds can be misleading (long words, letters with ascending lines seem bigger) and do not allow precise comparison.

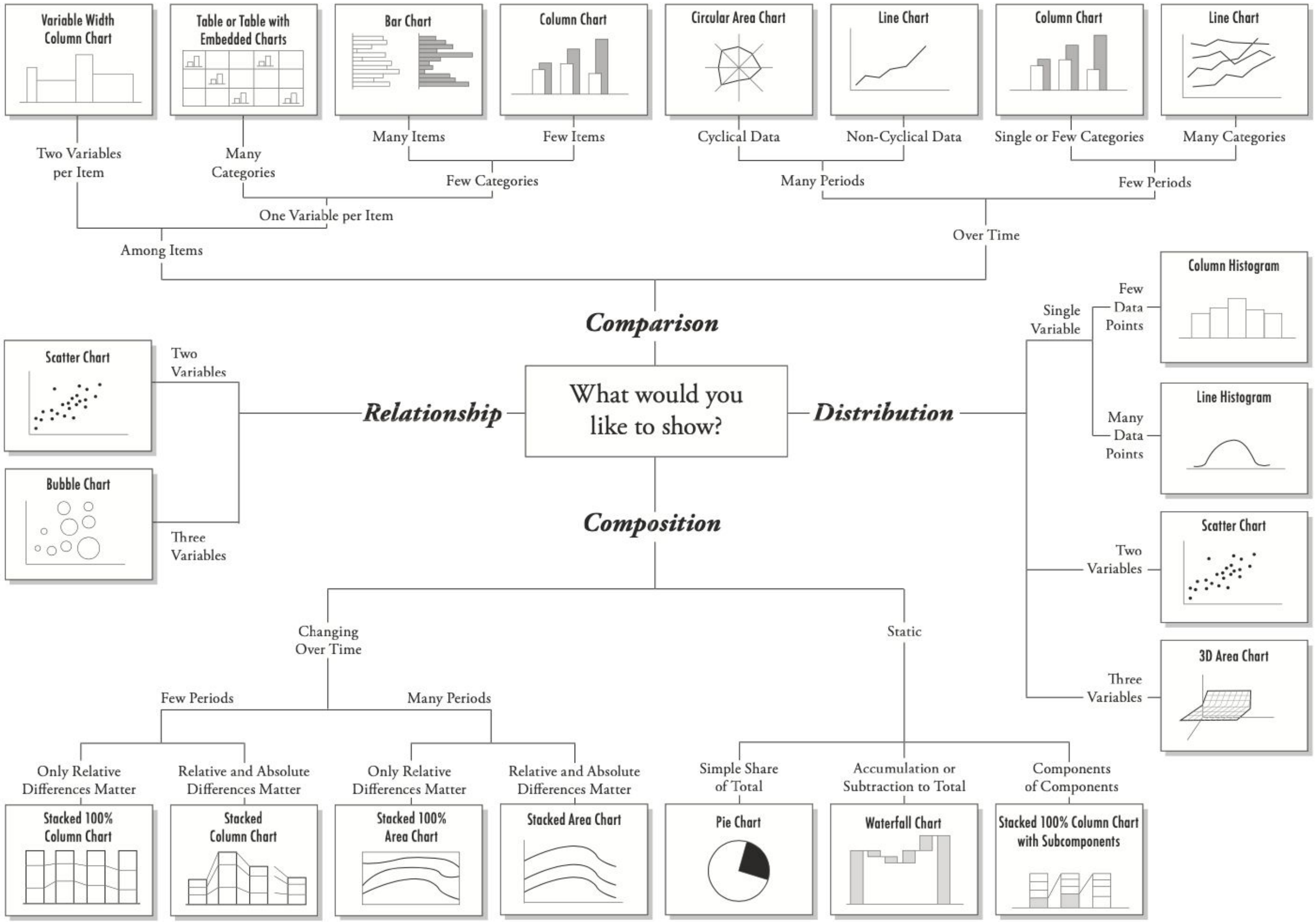
PARALLEL SETS



Sex	Colour	Count	% of Sex TOTAL
Male	Red	35	32.41
	Green	33	30.56
	Blue	40	37.04
Female	Red	28	16.67
	Green	60	35.71
	Blue	80	47.62

Allow to compare proportions and **relations between multivariate categorical values**. Relations are shown as flows between pairs of values.

A DECISION TABLE



ELECTRONIC PUBLISHING
AND DIGITAL STORYTELLING
Lesson 2

DYNAMIC DATA VISUALISATION

DYNAMIC DATA VISUALISATION

https://wattenberger.netlify.app/

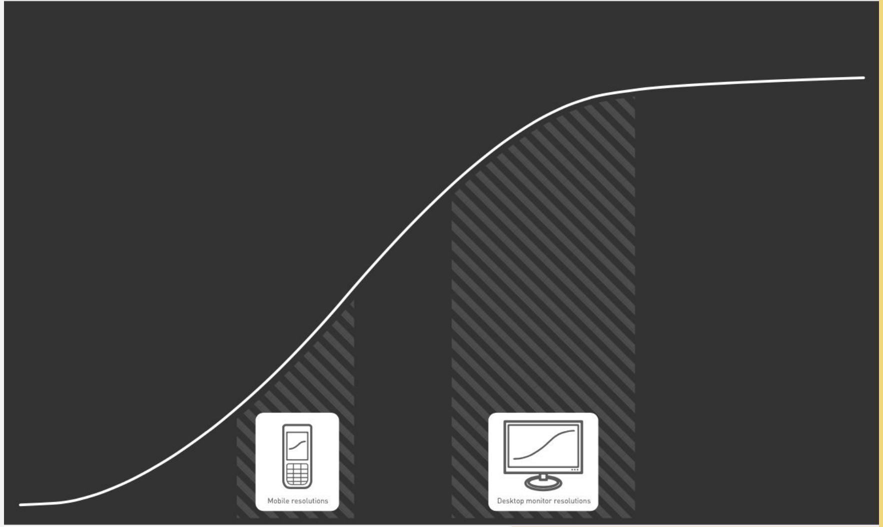
In dynamic data visualisations users can:

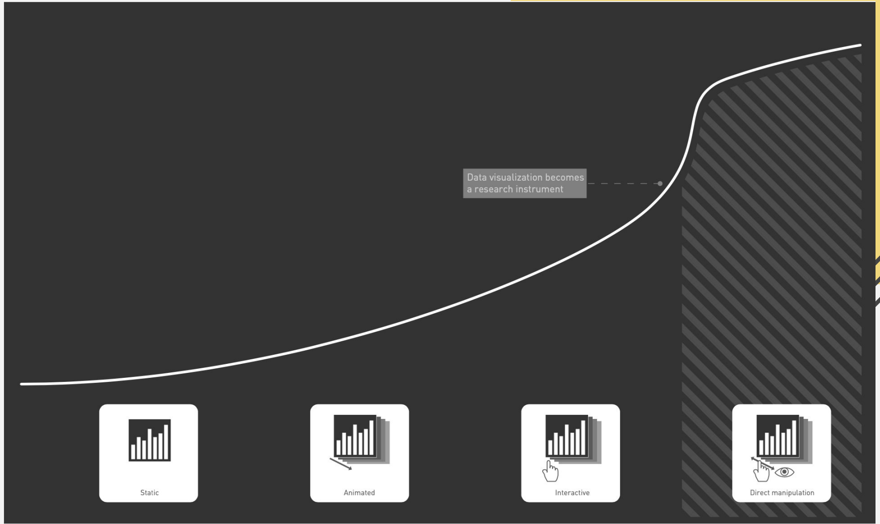
select and filter data to be displayed
interactive

select, filter and change the view
direct manipulation

Read more:
https://medium.com/@GT_Vis/evaluating-interactive-graphical-encodings-for-data-visualization-b9b8b4bf47f8

DYNAMIC DATA VISUALISATION





The **effectiveness** of the data visualisation is directly proportional to the resolution of the screen.

The **meaningfulness** of data is directly proportional to the interactivity allowed by the visualisation.

HOMEWORK

FILL IN THE QUESTIONNAIRE
AND GET READY FOR THE NEXT CLASS

FILL IN THE QUESTIONNAIRE BY THE NEXT CLASS

https://forms.gle/d8UbNJfVoMxF9iz86

GET READY FOR THE FIRST HANDS-ON CLASS:

- Come with your **laptop**
- Be sure you have installed **Python3** and a **IDE** (such as PyCharm, Sublime Text, Atom)
- Be sure you are able to run a python script via **shell** (read here)
- Be sure you know you are able to install a Python library with **pip** (or directly install RDFlib by the next class)

THANKS

Does anyone have any questions?

marilena.daquino2@unibo.it
github