

FIT3179 Data Visualisation

Week 03: Marks and Channels, Idiom Design Choices



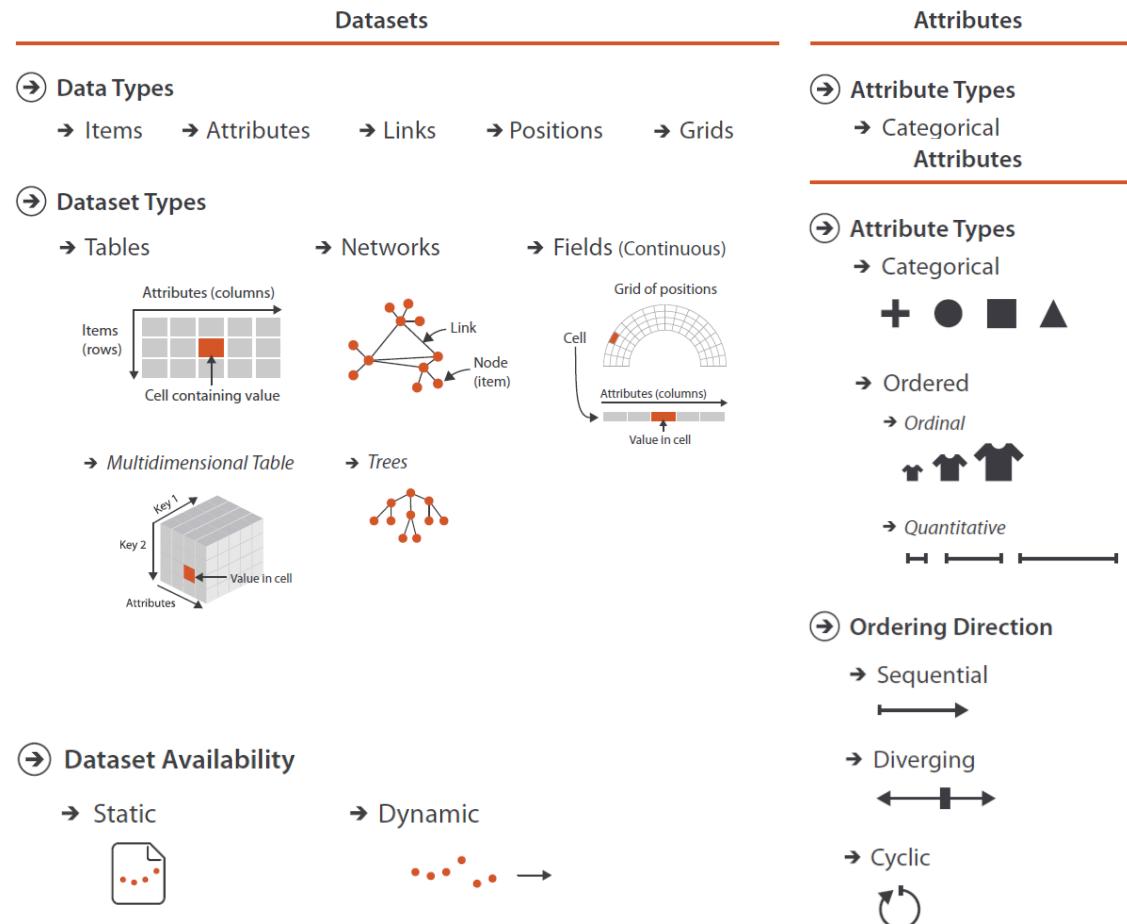
Recapitulation of Weeks 1 and 2

- What?
- Why?

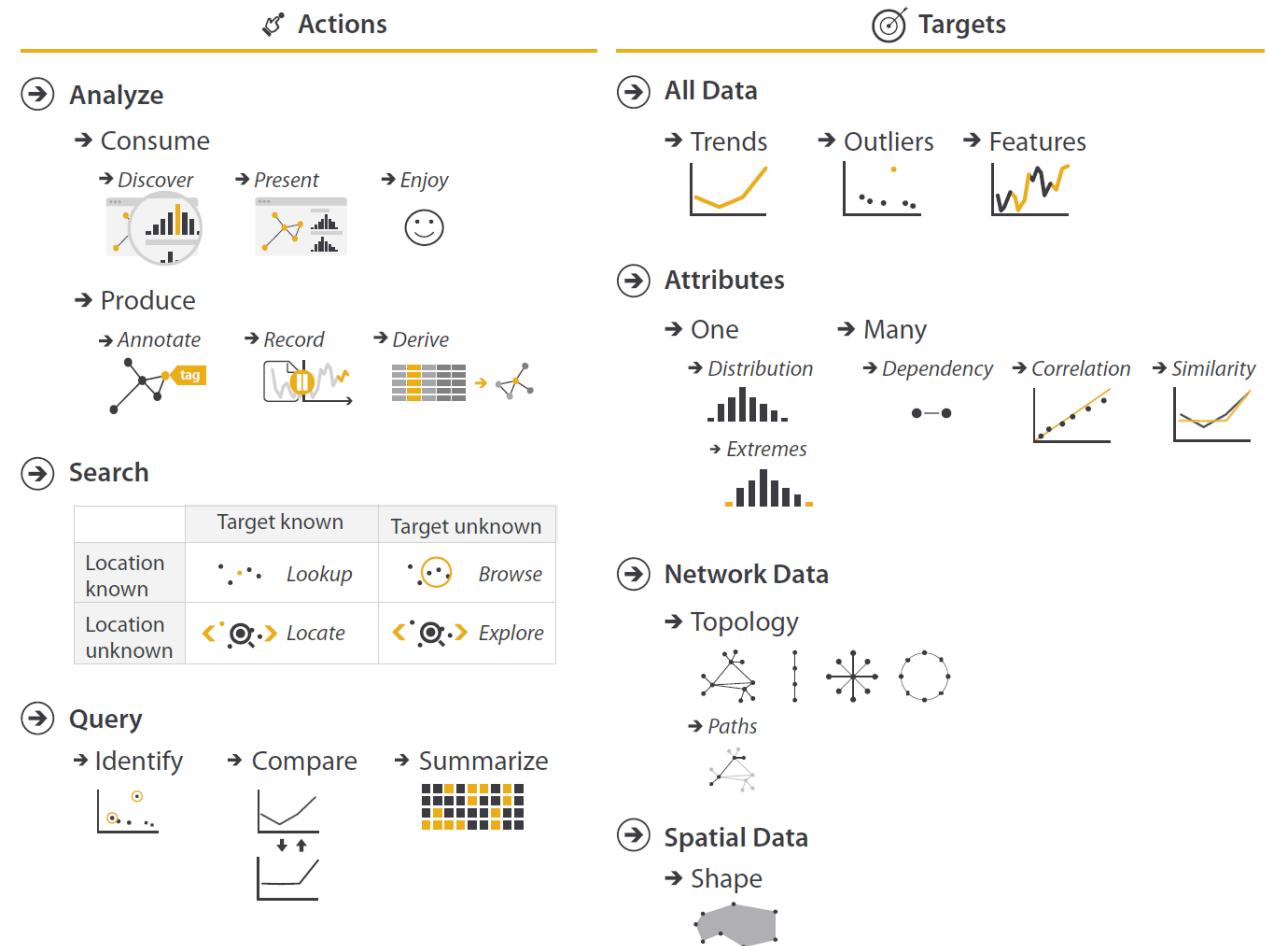
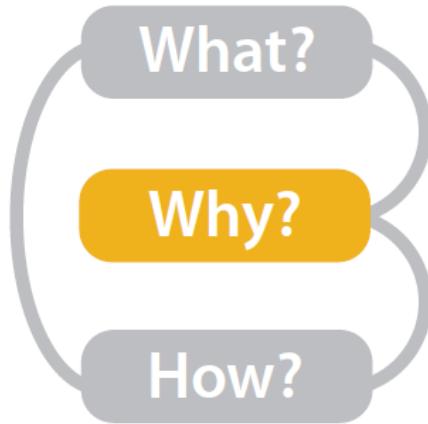
What?

Why?

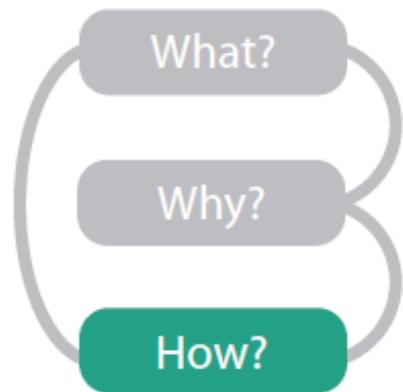
How?



Why?



How?



Presentations next Week!

Based on the Allocate+ tutorial classlist, the students in the CLAYTON Monday 1PM–3PM class will be presenting at the start of the class next week.
No presentations in lab of Monday 13 August, but August 20.

Please read the instructions on the Weekly forums and post your viz. research and analysis there!

- Find a visualisation using an idiom discussed in this lecture in a domain of your choice
- Discuss in a few words each the What / Why / How of your chosen Vis following Munzner's framework

Lecture Overview

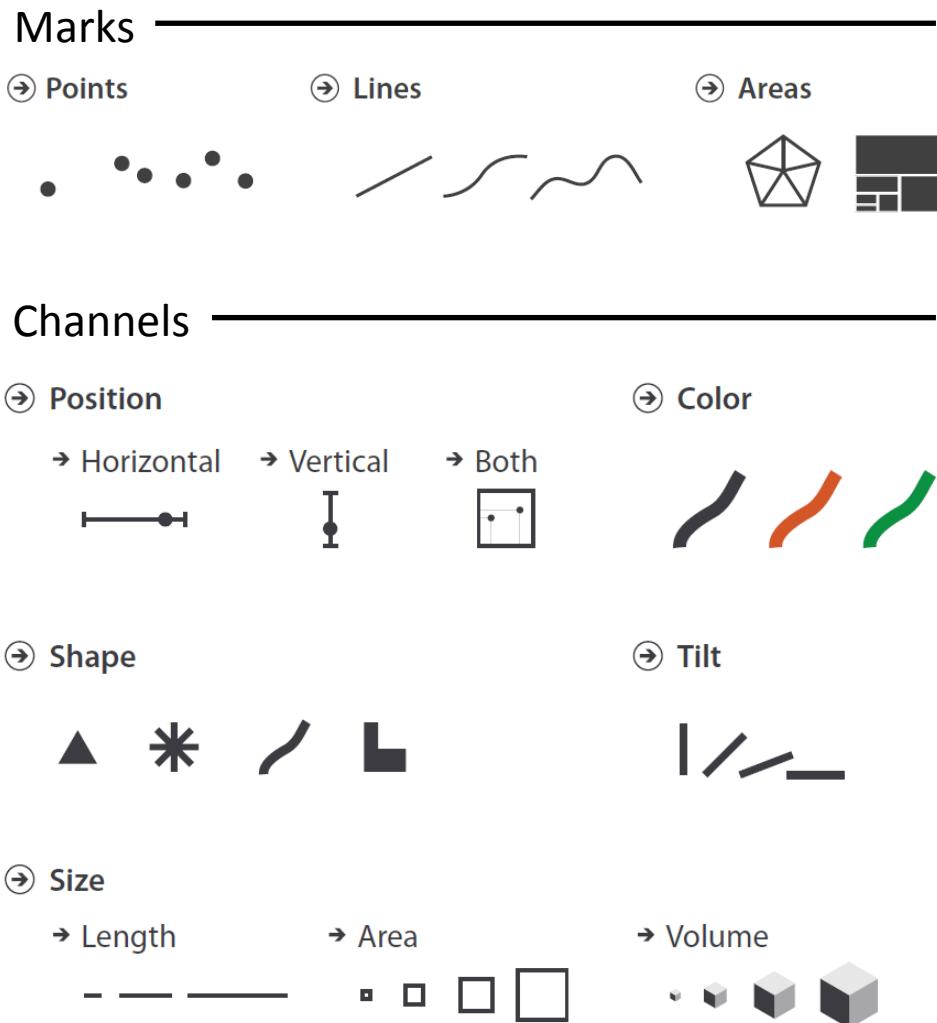
- Marks and Channels (Textbook chapter 5)
- Arrange Tables (Textbook chapter 7)
- Arrange Networks and Trees

VAD chapter 5

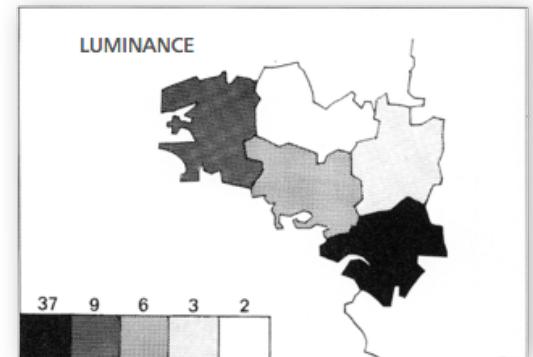
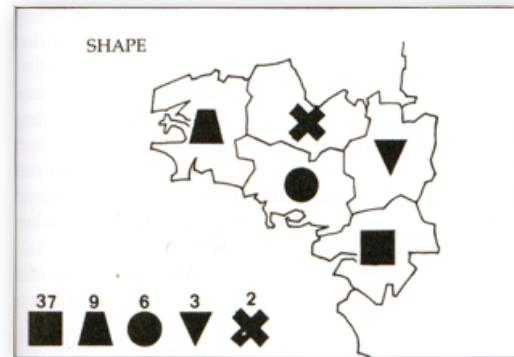
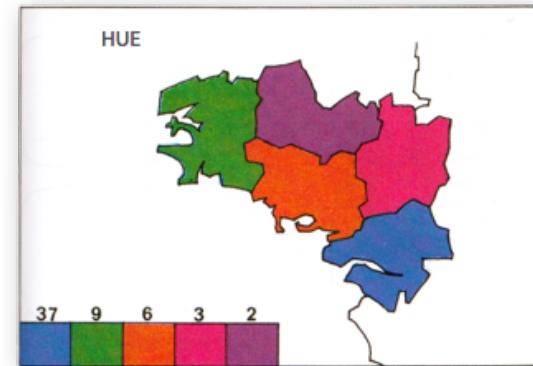
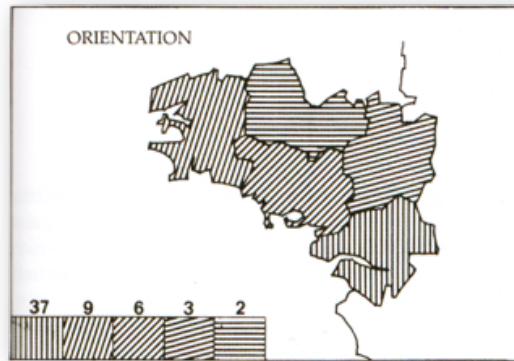
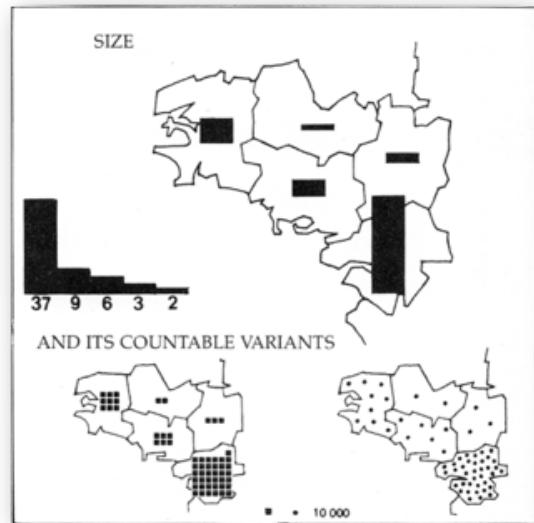
Marks and Channels

Definitions: Marks and Channels

- **Marks**
 - Geometric primitives
- **Channels**
 - Control appearance of marks
 - Can redundantly code with multiple channels
- **Combinations**
 - 2D position of points can represent 2 data attributes
 - Use size/shape/area for additional attributes
 - Example: Line marks encode quantity with width (size) and quality with hue.



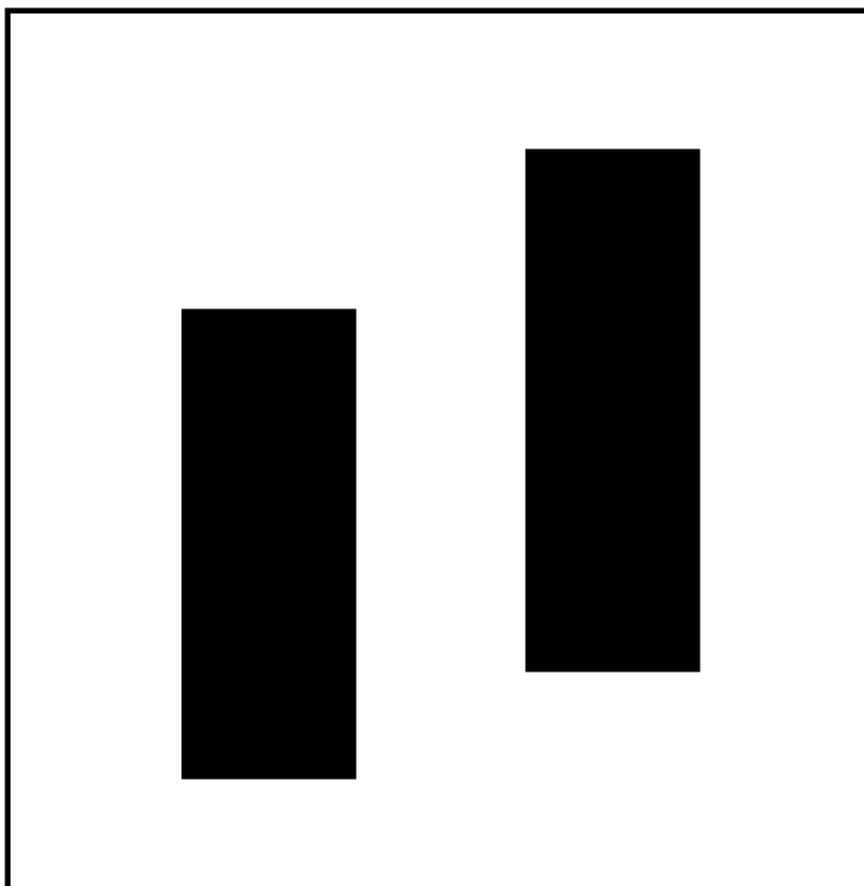
Channels: What works for quantitative data?



Bertin, J. 1983. Semiology of Graphics, p. 60–61

Note: For Bertin and generally in cartography *channels* are *visual variables*.

Accuracy: Relative vs. Absolute Judgements



Bar Chart

Which bar chart is longer? And how much?

Accuracy: Relative vs. Absolute Judgements



Bar Chart



Framed Rectangle

Relative vs. Absolute Judgements

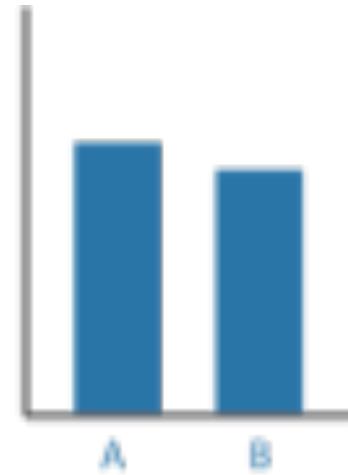
- Perceptual system mostly operates with relative judgements, not absolute - Weber's Law (roughly)



Unframed
Unaligned

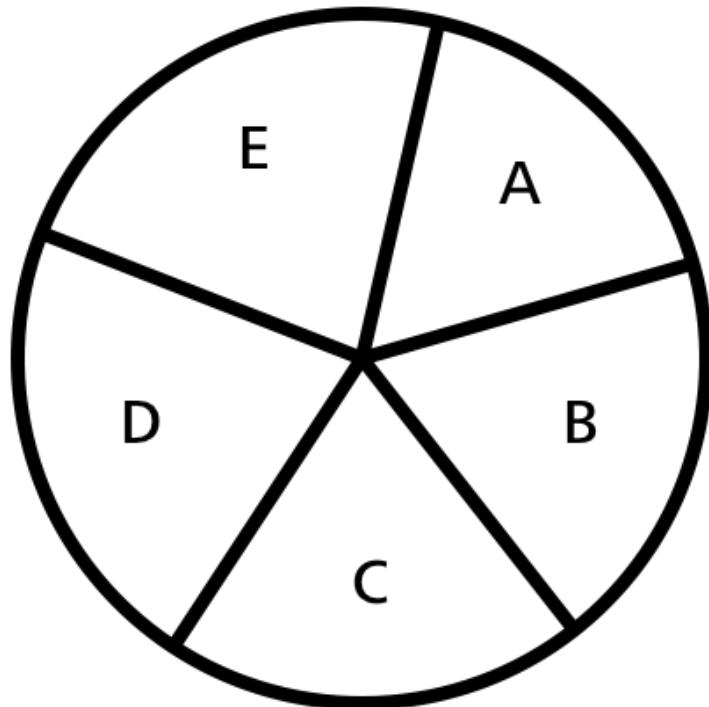


Framed
Unaligned



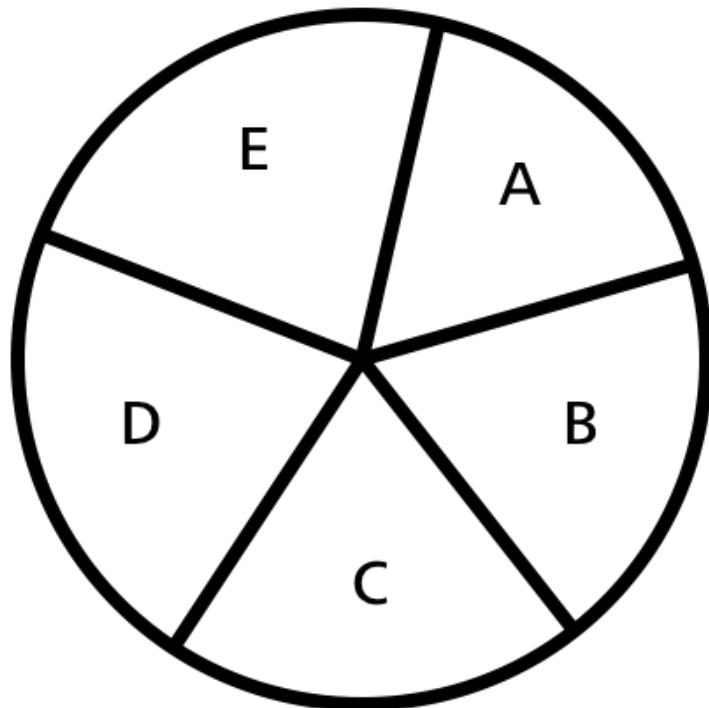
Unframed
Aligned

Accuracy: Angles vs. Length



Is C larger than B?

Accuracy: Angles vs. Length



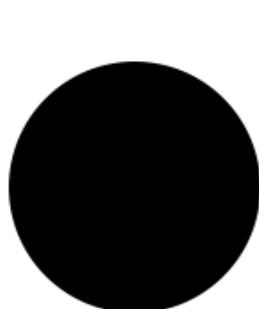
Pie Chart (angle)



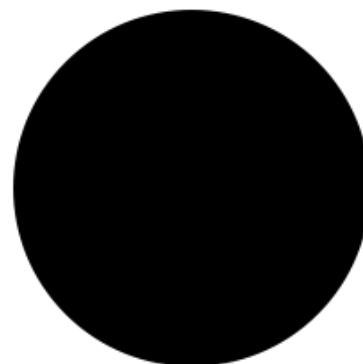
Bar Chart (length)

Accuracy: Area vs. Length

How much larger is B than A?



A



B

Area-proportional Symbols



A

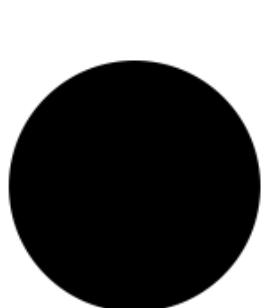


B

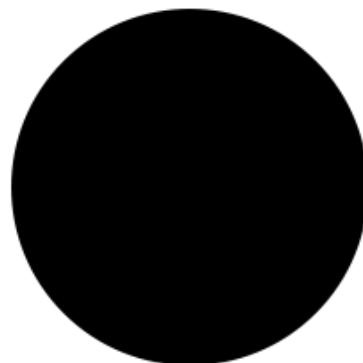
Bar Chart (length)

Accuracy: Area vs. Volume

How much larger is B than A?

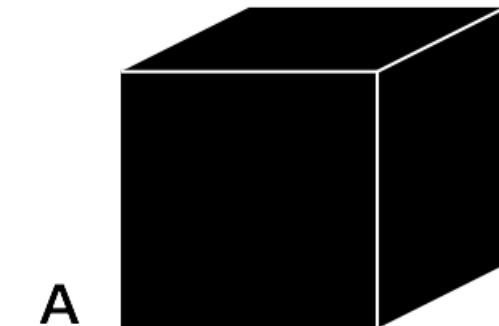


A

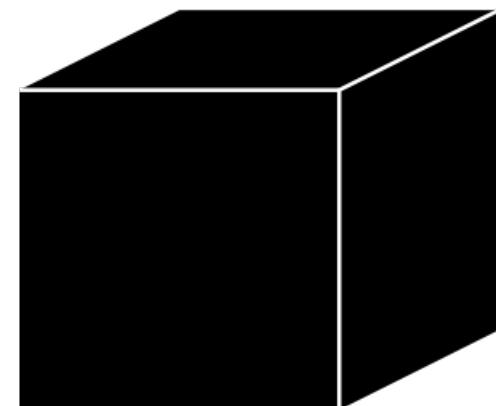


B

Area-proportional Symbols



A



B

Volume-proportional Symbols

Accuracy: Luminance

How much larger is B than A?



A



B

Accuracy: Ranking of Channels

- 1.** Position along a common scale



Values
50 and 75 in
all diagrams

- 2.** Position along nonaligned scales
(multiple diagrams)



- 3.** Length, direction, angle



- 4.** Area



- 5.** Volume, curvature



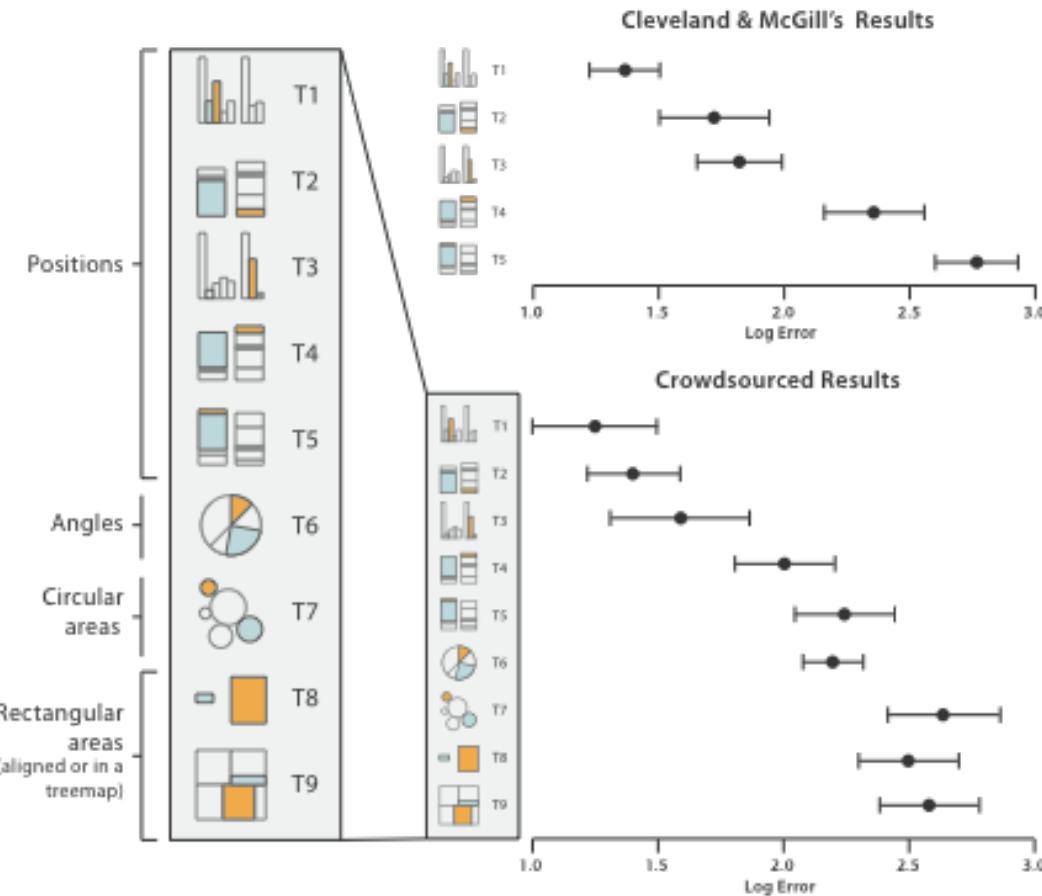
- 6.** Luminance, saturation



Cleveland and McGill 1984.

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

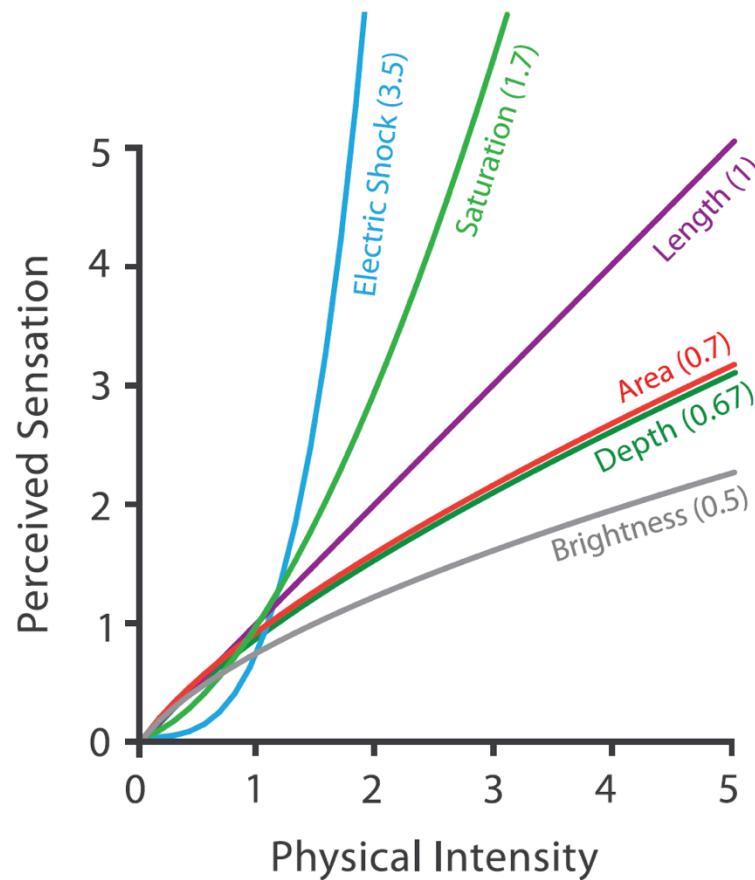
Accuracy: Vis experiments



Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design.
Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.

Accuracy: Physical intensity vs. perceived sensation

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



Channels: Expressiveness types and effectiveness rankings

④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



④ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



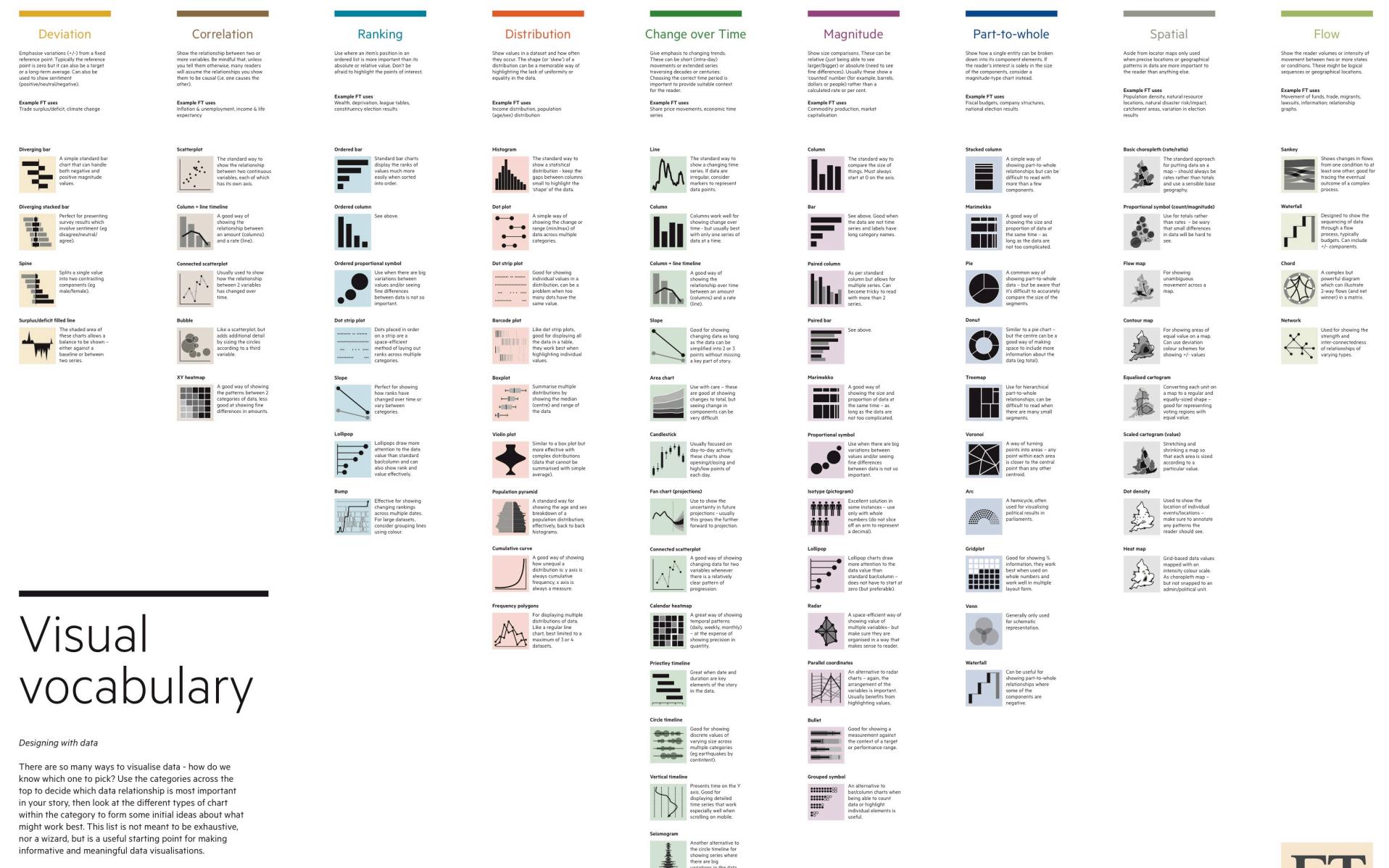
Shape



↑ Most
Effectiveness
↓ Least
Same

VAD chapter 7

Arrange Tables



Visual vocabulary

Designing with data

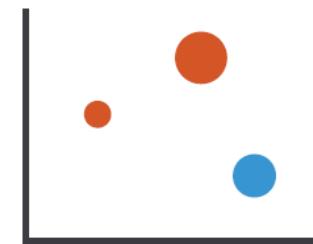
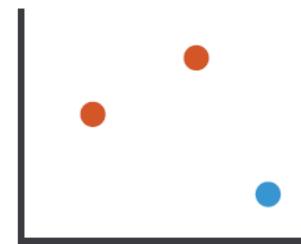
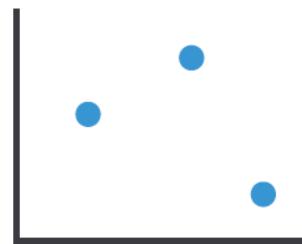
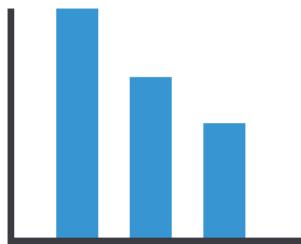
There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list 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; Jon Bert; Liz Fauno; Graham Parish; Billy Ehrenberg; Paul McCallum; Harris Siebel
Inspired by the Graphics Companion by Jon Schwabish and Steven Wolfson

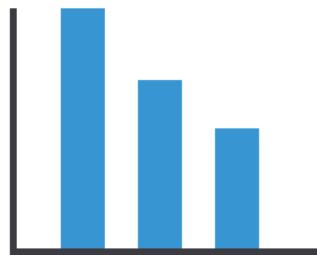


ft.com/vocabulary

- Common data visualisations we see are *idioms*
- What can we say about the structure of these common idioms?

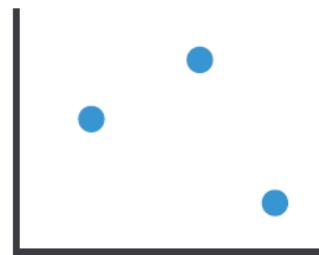


- Structure of common idioms
 - combinations of marks and channels



2:
vertical position
length

mark: line



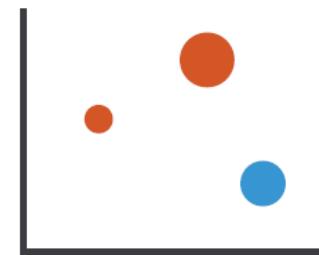
2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

mark: point



4:
vertical position
horizontal position
color hue
size (area)

mark: point

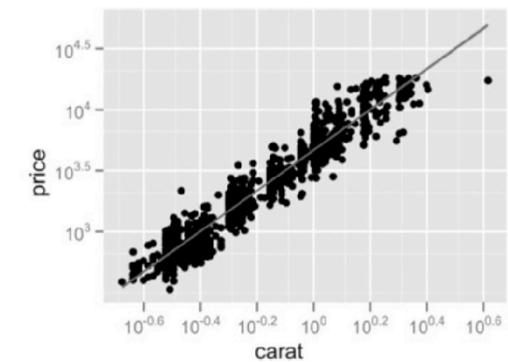
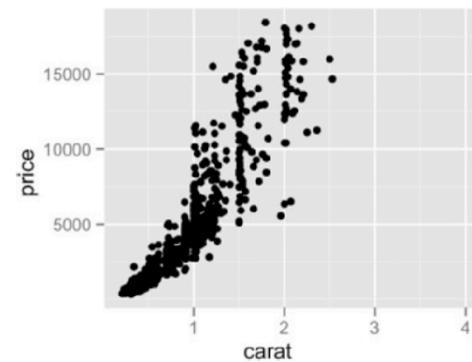
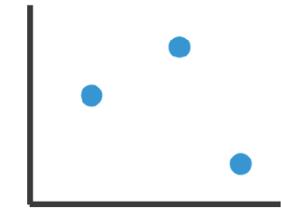
- **What?**
 - Data: 2 quantitative attributes

- **Why?**

- Tasks:
 - Find trends
 - Outliers
 - Distribution
 - Correlation
 - Clusters

- **How?**
 - Marks: points
 - Channels:
 - Horizontal position
 - Vertical position
 - Scalability to hundreds of items

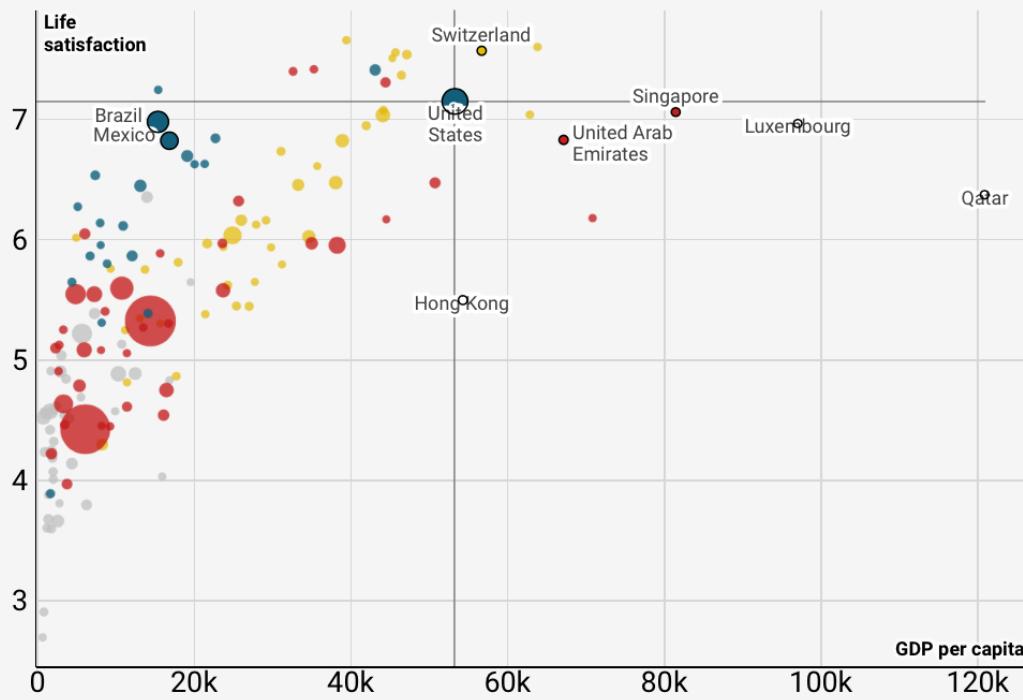
⇒ Express Values



Idiom: scatterplot (with multiple channels)

GDP per capita vs Self-reported Life Satisfaction

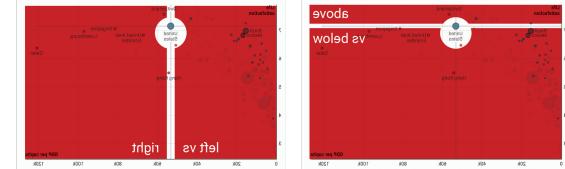
The vertical axis shows national average self-reported life satisfaction on a scale ranging from 0-10, where 10 is the highest possible life satisfaction. The horizontal axis shows GDP per capita based on purchasing power parity (i.e. GDP per head after adjusting for inflation and cross-country price differences). Numbers are from 2014-2016.



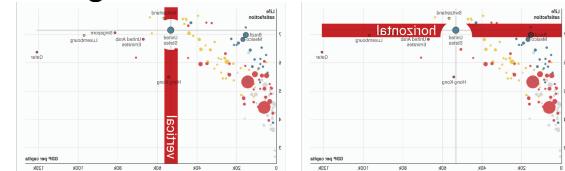
Source: [Our World in Data](#), with data from Gallup World Poll surveys and World Bank • [Get the data](#)

Source: <https://blog.datawrapper.de/weekly-chart-lifesatisfaction/>

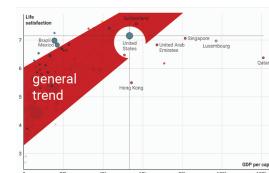
Compare with dots
• above and below
• left and right



- along vertical line
- along horizontal line



- on general trend



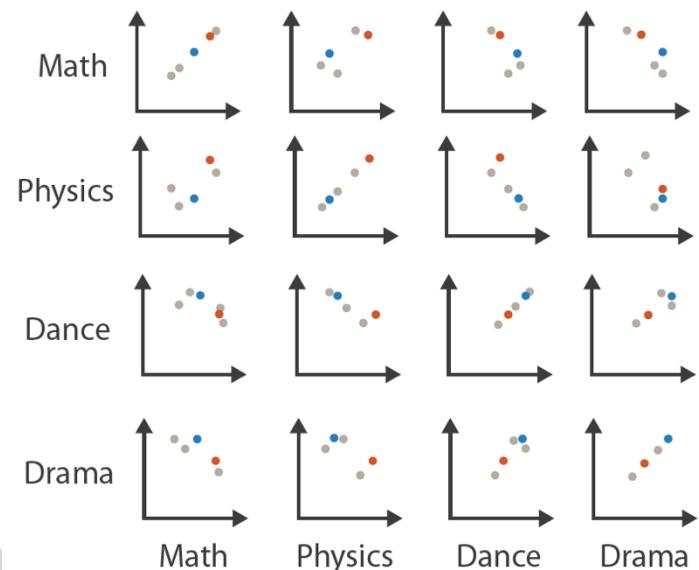
Idiom: scatterplot matrix (SPLOM)

- **What?**
 - Table of quantitative attributes
- **Why?**
 - Task: find correlation, trends, outliers.
- **How?**
 - Scatterplots in 2D grid array (Matrix)
 - Scalability:
up to about a dozen attributes

Table

	Math	Physics	Dance	Drama
	85	95	70	65
	90	80	60	50
	65	50	90	90
	50	40	95	80
	40	60	80	90

Scatterplot Matrix



▪ What?

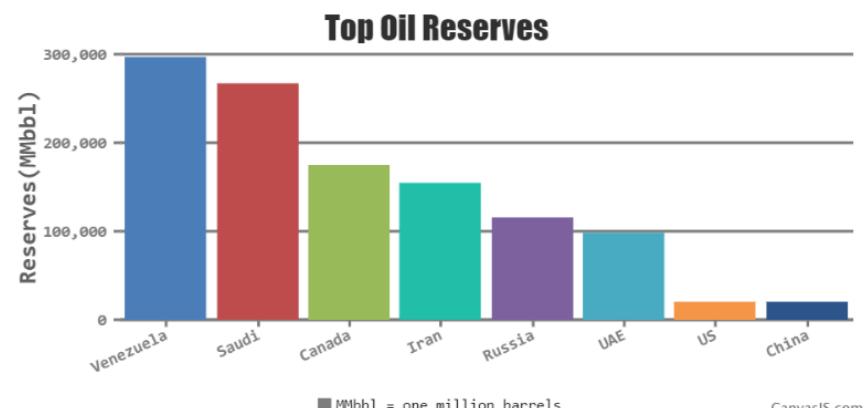
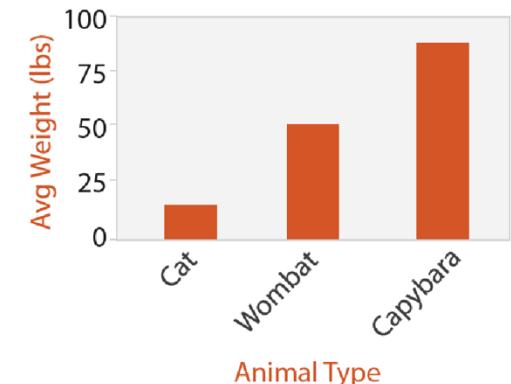
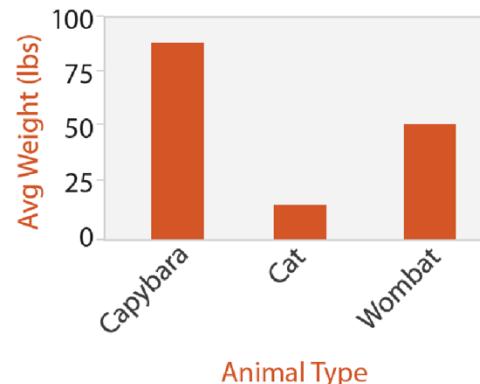
- Data: 1 key, one value
- 1 categ attrib, 1 quant attrib

▪ Why?

- Tasks:
 - Compare
 - Lookup values
 - Find biggest / smallest

▪ How?

- Marks: lines (bars)
- Channels:
 - Length to express quant value
 - Spatial regions: one per mark
 - Order by label or by value
- Scalability: dozens to hundreds of items



- **What?**

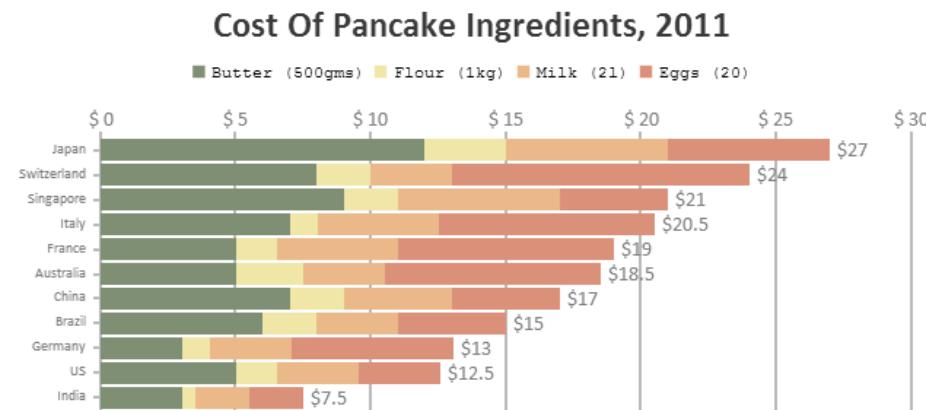
- Data: an extra key
- 2 categ attrib, 1 quant attrib

- **Why?**

- Tasks: Part to whole relationship

- **How?**

- Marks: stack of line marks
 - **Glyph:** composite object
- Channels:
 - Length to express quant value
 - Spatial regions: one per mark
 - Order by label or by value
- Scalability: several to dozens of levels of stacked attributes



<http://canvasjs.com/html5-javascript-bar-chart/>

Feb 13, 2018

by Lisa Charlotte
Rost

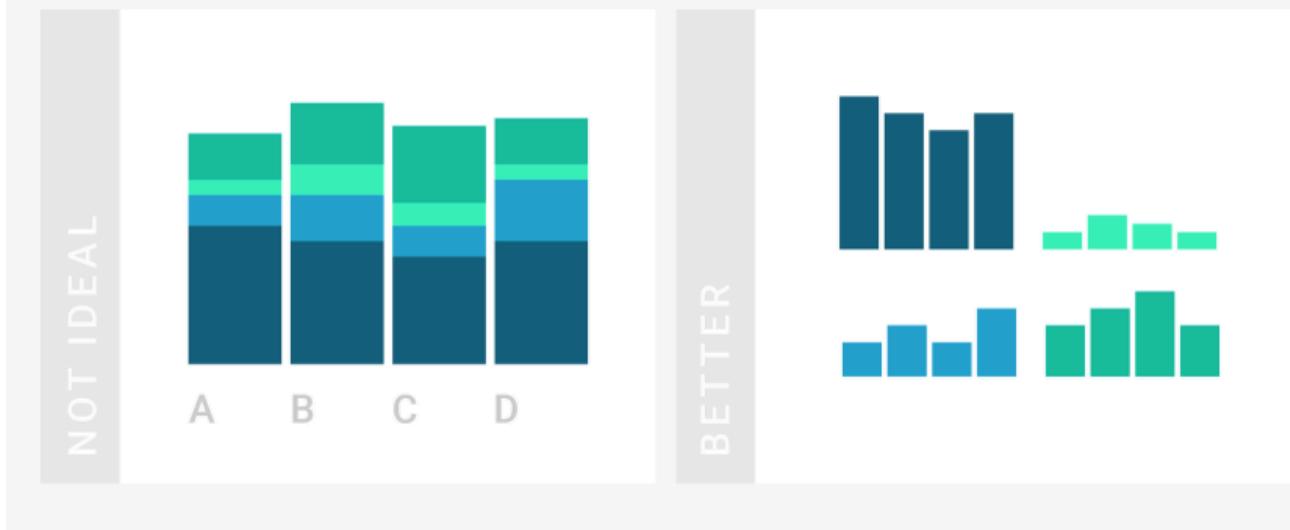
Thoughts & How To's

What to consider when creating stacked column charts

<https://blog.datawrapper.de/stacked-column-charts/>

When to use stacked bar charts?

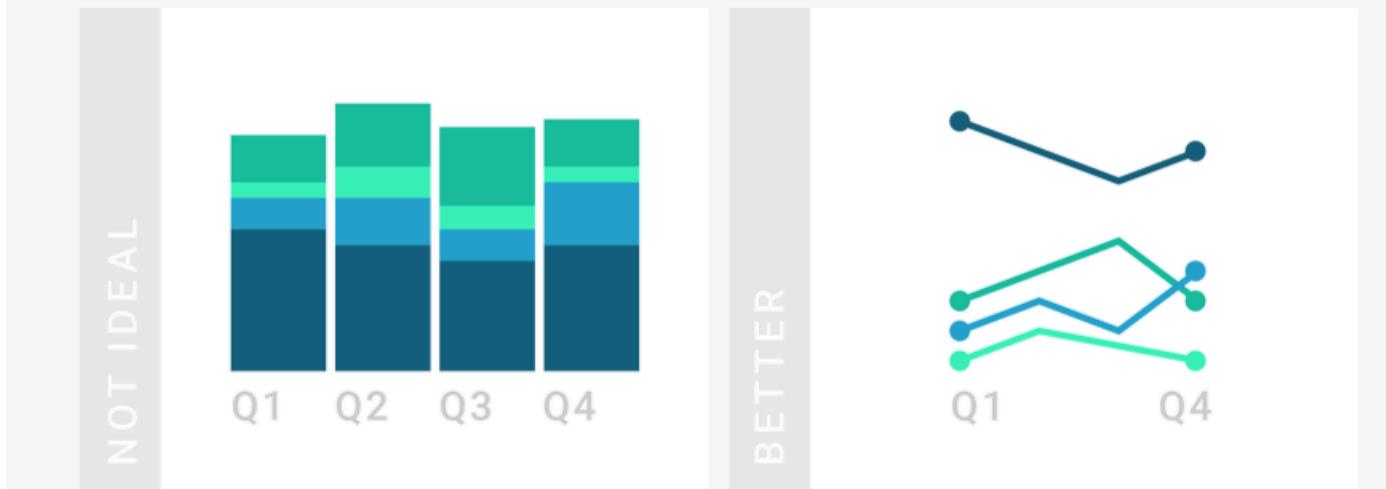
Stacked column charts work well when the focus of the chart is to compare the totals and one part of the totals. It's hard for readers to compare columns that don't start at the same baseline. If the focus of your chart is to compare multiple parts across all your totals with each other, consider split bars or small multiples instead.



Source: List Charlotte Rost, <https://blog.datawrapper.de/stacked-column-charts/>

When to use stacked bar charts?

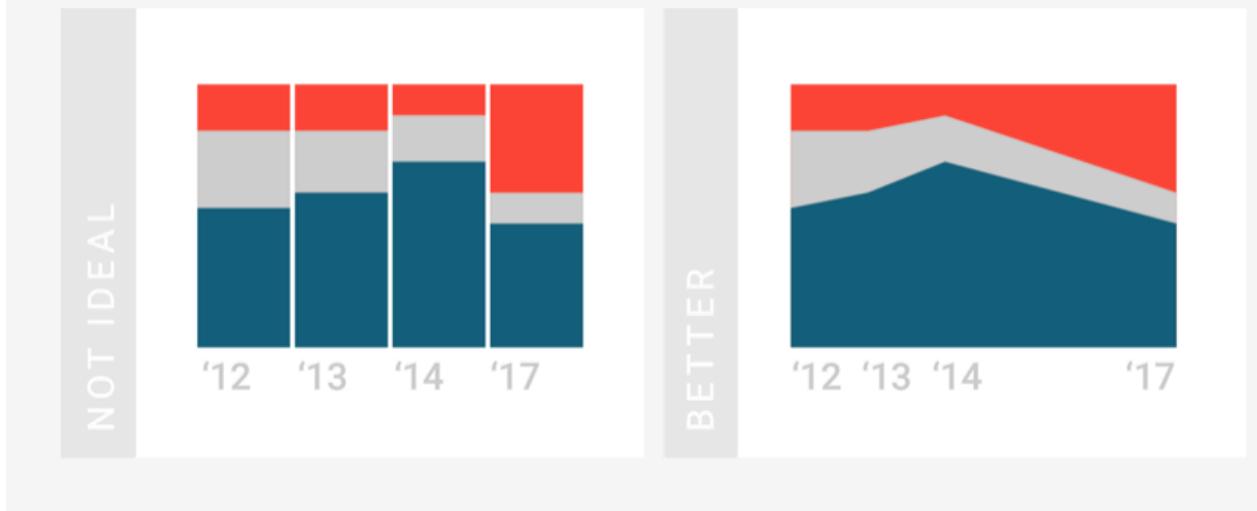
Stacked column chart can work well for dates, but should only be considered if the total of your parts is crucial. If the total of your parts is not important, a line chart is a better choice. It can be quicker decipherable by your readers.



Source: List Charlotte Rost, <https://blog.datawrapper.de/stacked-column-charts/>

When to use stacked bar charts?

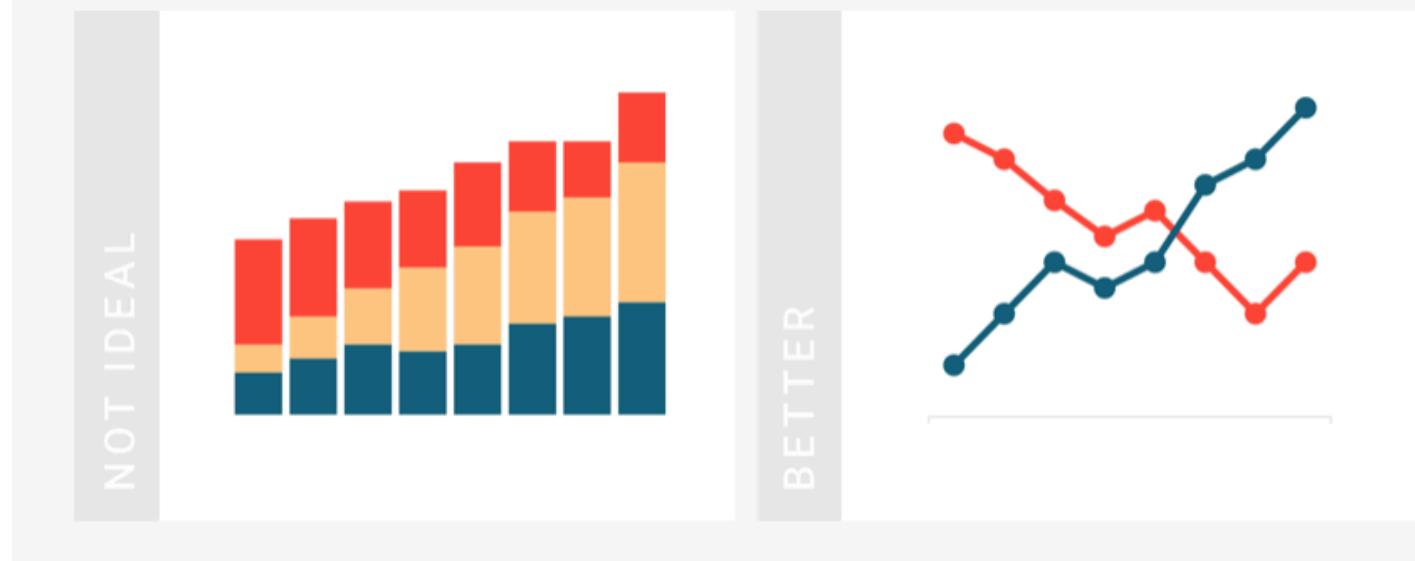
If you want to use stacked column charts for time data, make sure your dates have the same intervals. If you want to choose between area charts and stacked column charts, then the latter one is a better choice if you only have a few dates. Readers will be able to compare the parts of the totals better with each other than on an area chart. However, if the intervals between your dates are different, use an area chart (or line chart). They have x-axes with continuous scales that will show the date intervals correctly.



Source: List Charlotte Rost, <https://blog.datawrapper.de/stacked-column-charts/>

When to use stacked bar charts?

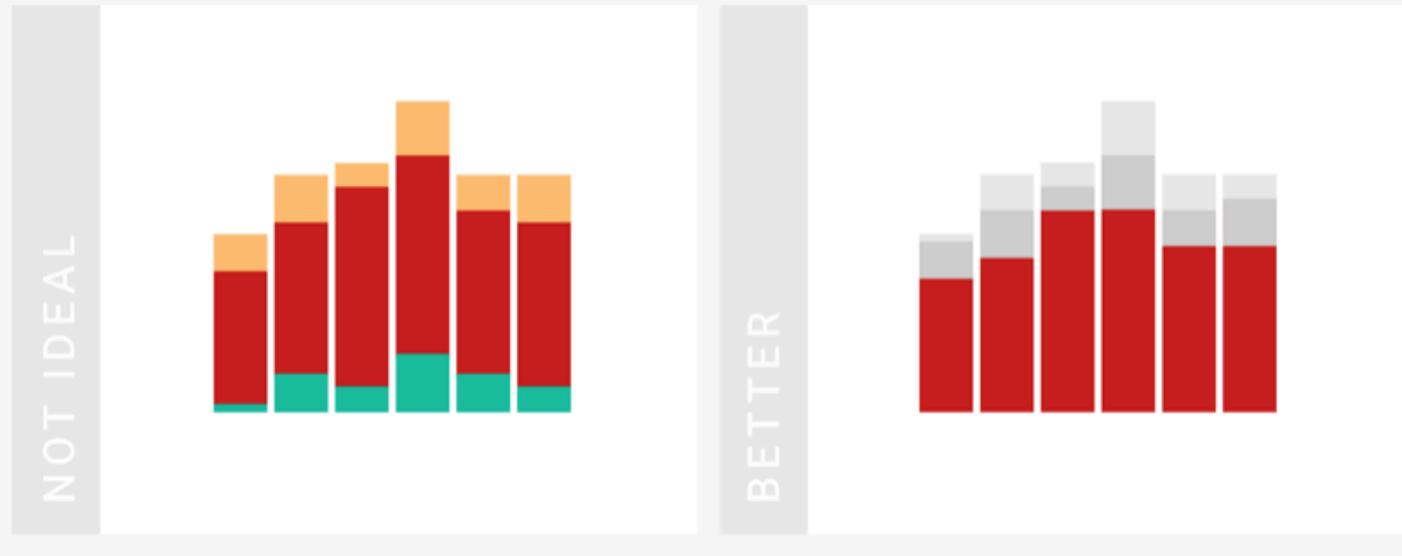
For time data, also consider a line chart if you want to show that one share overtook another one. Readers won't see intuitively that the chart is about parts of a total anymore. However, you will communicate your message more clearly and don't need to show all the shares of the total to do so.



Source: List Charlotte Rost, <https://blog.datawrapper.de/stacked-column-charts/>

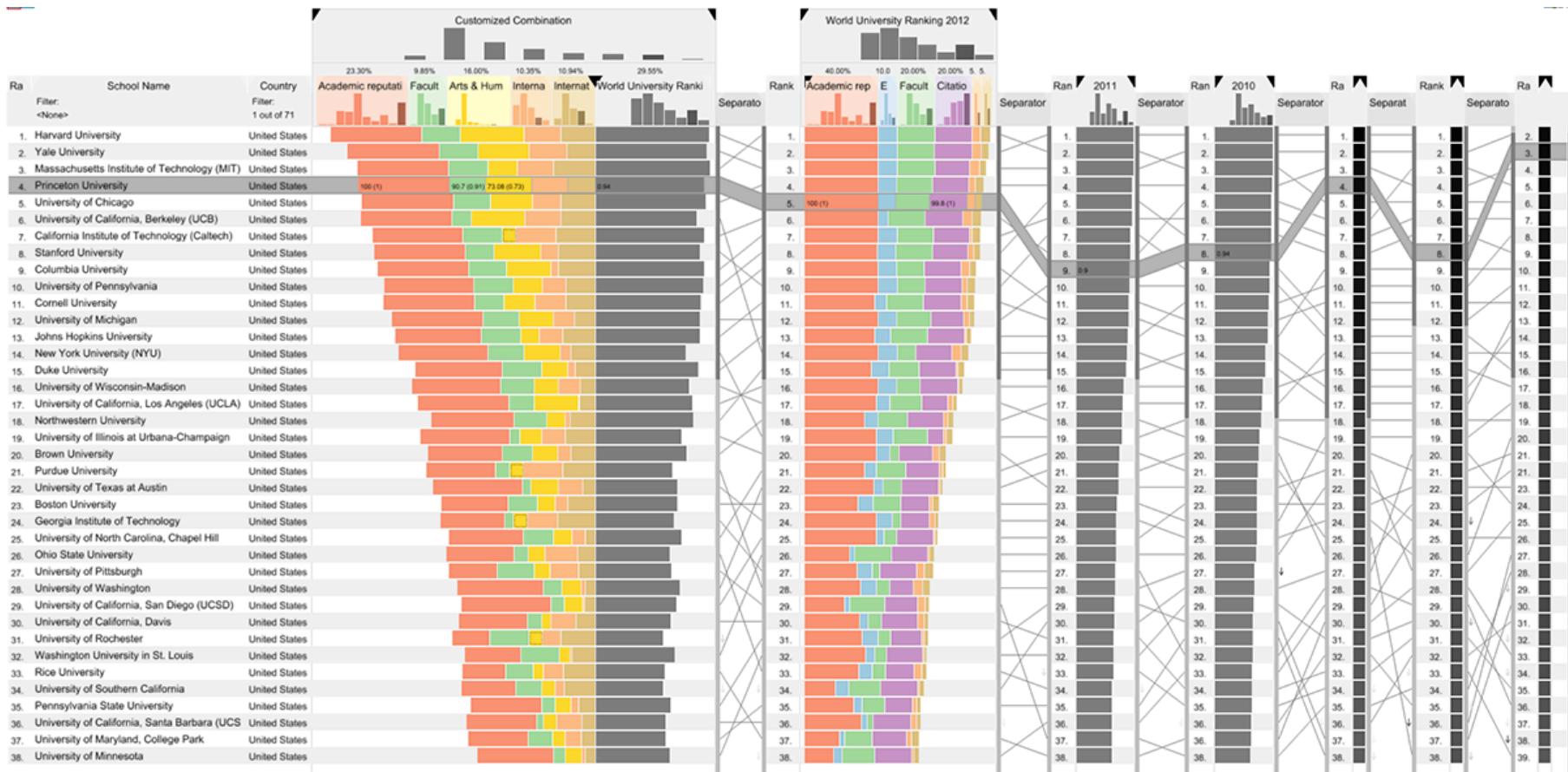
Designing stacked bar charts

Bring the most important value to the bottom of the chart and use color to make it stand out. Your readers can compare values easier with each other if they have the same baseline.



Source: List Charlotte Rost, <https://blog.datawrapper.de/stacked-column-charts/>

LineUp: Blurring the lines between tabular and visual representations



Demo: <http://caleydo.org/tools/lineup/>

▪ What?

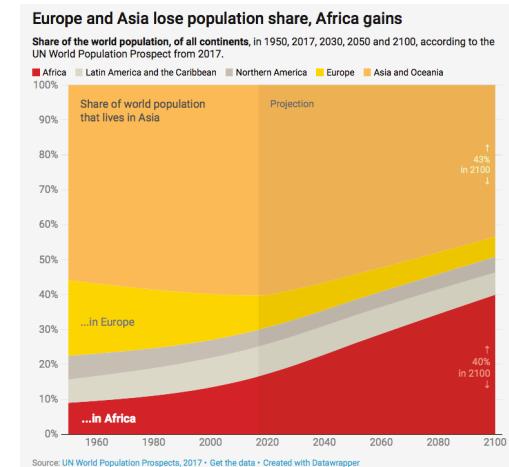
- Data:
 - 1 categ key attrib (continent)
 - 1 ordered key attrib (time)
 - 1 quant value attrib (population)

▪ Why?

- Tasks:
 - Analyse trends over time

▪ How?

- Marks: polygonal layers
- Channels:
 - Height of layer to express quant value
 - Spatial regions: one per mark
 - Order by time
- Scalability: hundreds of items

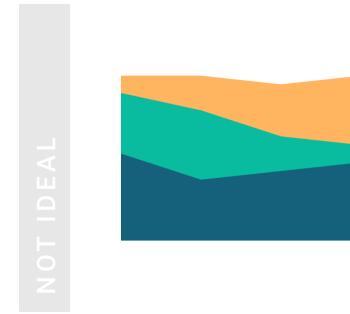
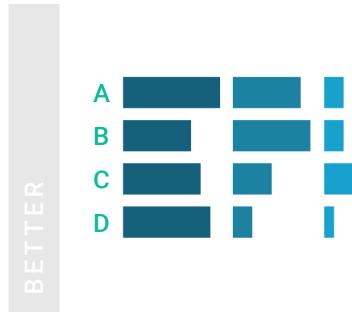


<https://blog.datawrapper.de/area-charts/>

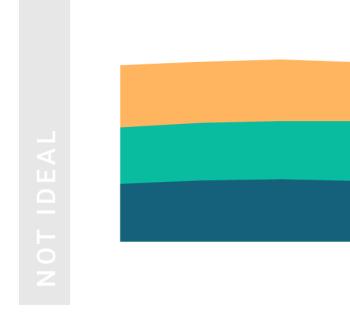
When to use area charts



For time only.

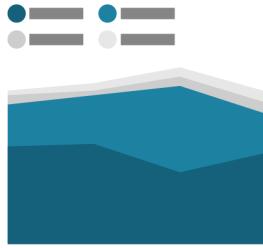


Not good for comparing shares.



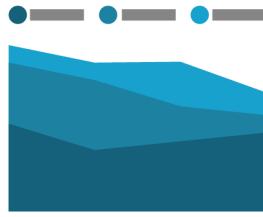
How to design area charts

NOT IDEAL



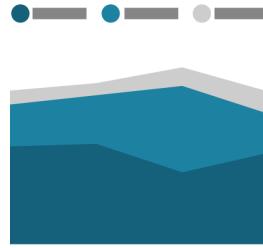
Group tiny values.

NOT IDEAL



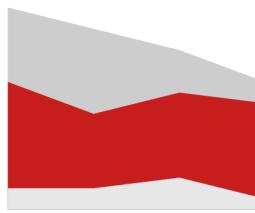
Place labels on chart.

BETTER



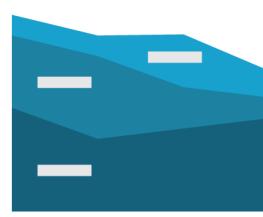
BETTER

NOT IDEAL



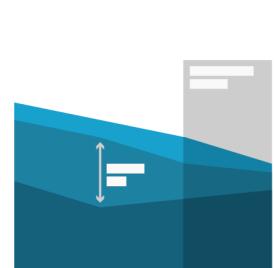
Most important at bottom and highlighted.

NOT IDEAL



BETTER

BETTER



Annotate and highlight ranges, mean values, current time, etc.

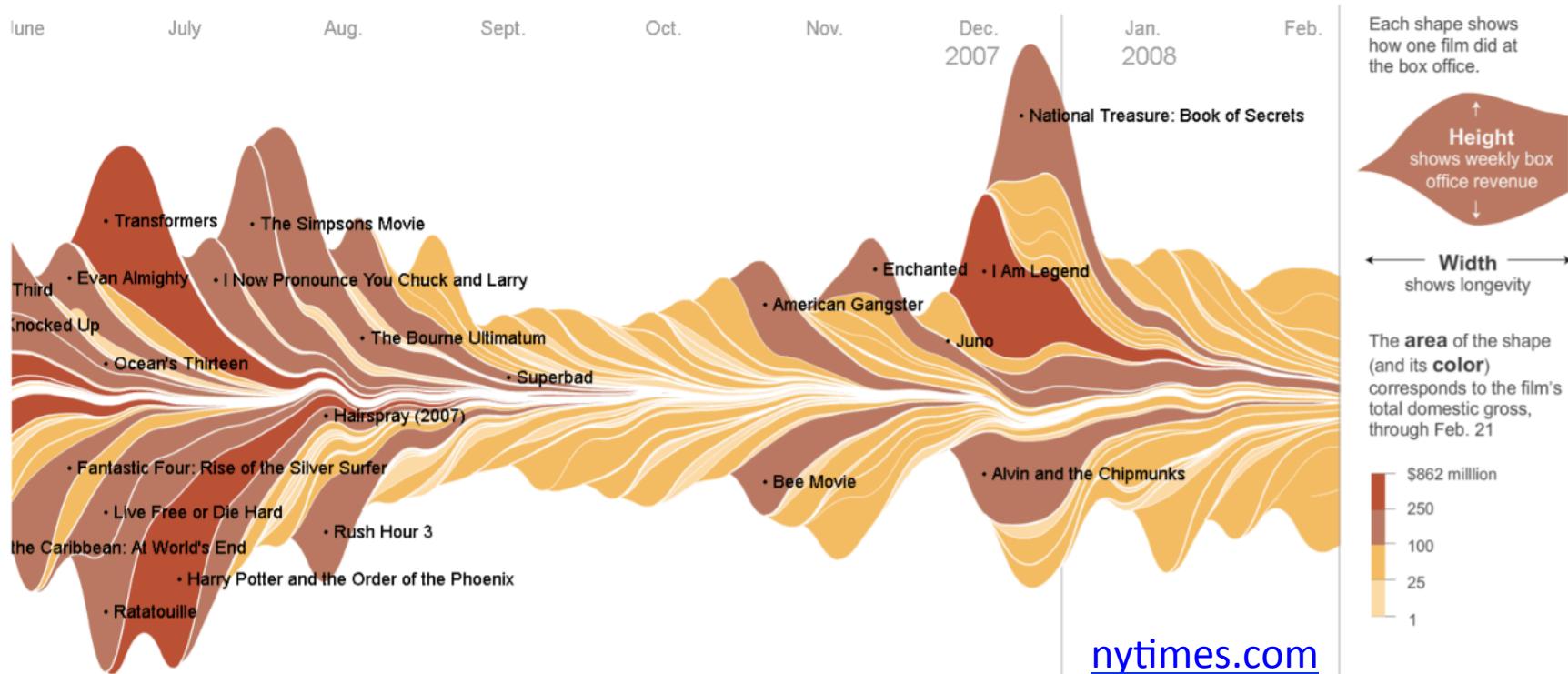
Idiom: streamgraph: a variation of area graphs

▪ What?

— Data:

- 1 categ key attrib (movie)
- 1 ordered key attrib (time)
- 1 quant value attrib (revenue)

Source: *Stacked Graphs Geometry & Aesthetics*. Byron and Wattenberg.
IEEE Trans. Visualization and Computer Graphics, 14(6): 1245–1252, (2008).



- **What?**

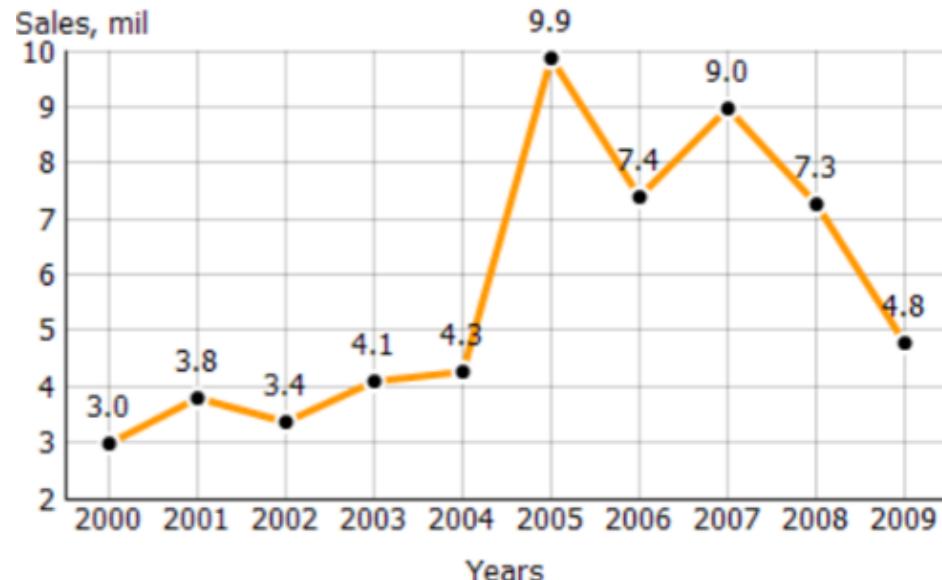
- Data:
 - 1 ordinal attribute (often time)
 - 1 quant attribute

- **Why?**

- Task: find trend, peaks, etc.

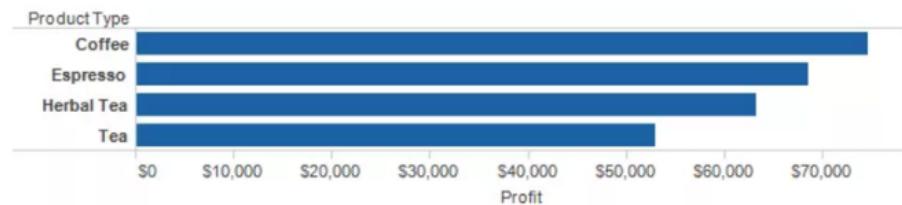
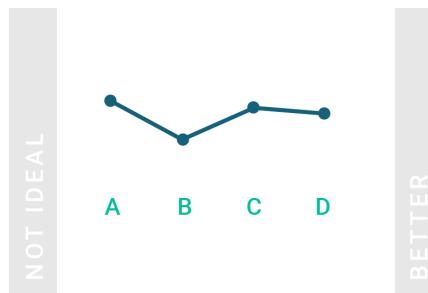
- **How?**

- Marks:
 - Points
 - Line connection
 - Channels:
 - Vertical positions express quantitative values
 - Line emphasises relationship between attributes
 - Scalability: depends on screen real-estate, zoomable interaction etc.



Choosing bar vs line charts

- Is the key attribute:
 - Categorical
 - bar charts
 - Ordered
 - line charts
- Use line charts to show how values develop over time.

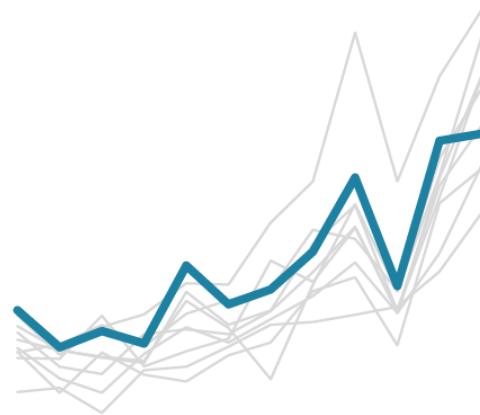


Designing line chart

NOT IDEAL



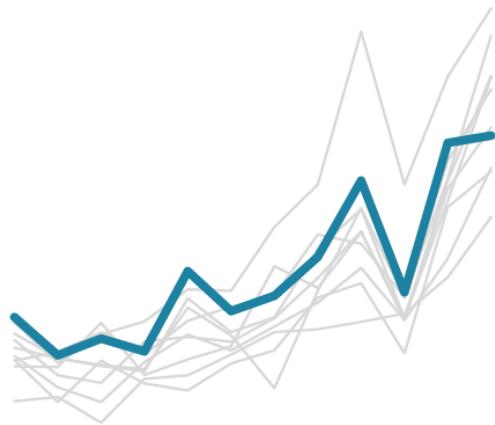
BETTER



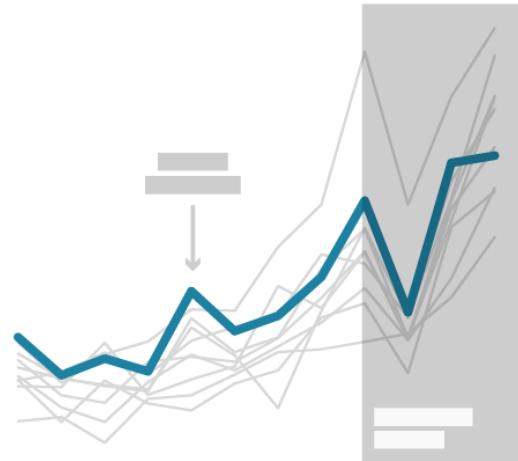
Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/line-charts/>

Designing line chart

NOT IDEAL

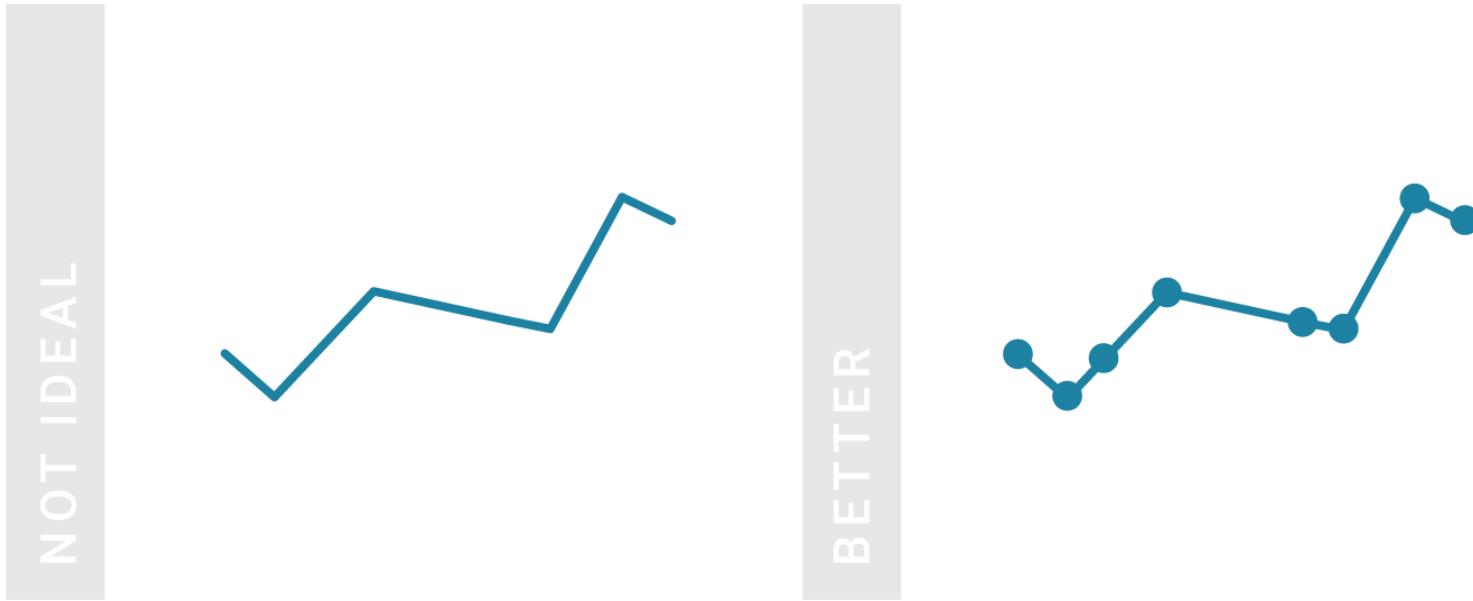


BETTER



Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/line-charts/>

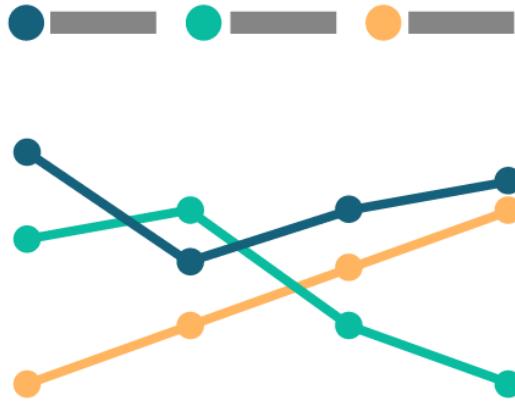
Designing line chart



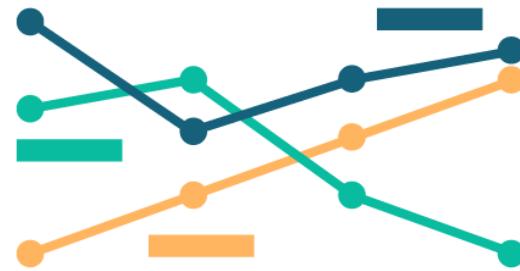
Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/line-charts/>

Designing line chart

NOT IDEAL



BETTER



Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/line-charts/>

- **What?**

- Table of attributes
 - Works with quantitative, categorical, ordered, etc.

- **Why?**

- Task: find correlation, trends, outliers.

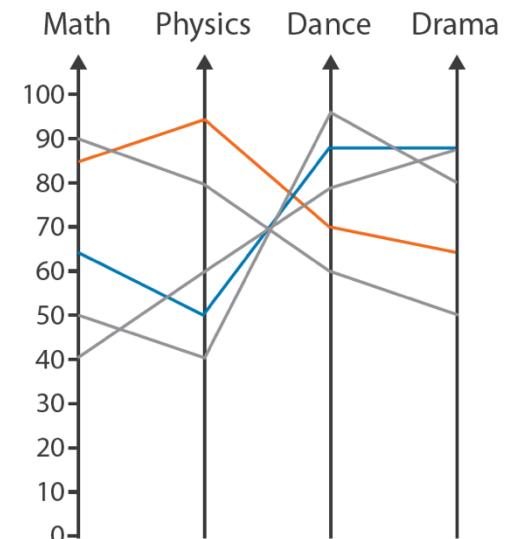
- **How?**

- One axis per attribute
 - marks are one line per data item
 - intersection with axis gives value
- Scalability:
 - hundreds of items (interaction and clever rendering techniques can boost this)

Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Parallel Coordinates



Total rows: 428 Total columns: 15

Data Set: SASHelp.Cars

← ← Rows 1-100 → →

	Make	Model	Type	Origin	DriveT...	MSRP	Invoice	Engin...	Cylind...	Horse...	MPG_...	MPG_...	Weight	Wheel...	Leng...
1	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6	265	17	23	4451	106	189
2	Acura	RSX Type	Sedan	Asia	Front	\$23,820	\$21,761	2	4	200	24	31	2778	101	172
3	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647	2.4	4	200	22	29	3230	105	183
4	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299	3.2	6	270	20	28	3575	108	186
5	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014	3.5	6	225	18	24	3880	115	197
6	Acura	3.5 RL w/I	Sedan	Asia	Front	\$46,100	\$41,100	3.5	6	225	18	24	3893	115	197
7	Acura	NSX coup	Sports	Asia	Rear	\$89,765	\$79,978	3.2	6	290	17	24	3153	100	174
8	Audi	A4 1.8T 4c	Sedan	Europe	Front	\$25,940	\$23,508	1.8	4	170	22	31	3252	104	179
9	Audi	A41.8T co	Sedan	Europe	Front	\$35,940	\$32,506	1.8	4	170	23	30	3638	105	180
10	Audi	A4 3.0 4dr	Sedan	Europe	Front	\$31,840	\$28,846	3	6	220	20	28	3462	104	179
11	Audi	A4 3.0 Qu.	Sedan	Europe	All	\$33,430	\$30,366	3	6	220	17	26	3583	104	179
12	Audi	A4 3.0 Qu.	Sedan	Europe	All	\$34,480	\$31,388	3	6	220	18	25	3627	104	179
13	Audi	A6 3.0 4dr	Sedan	Europe	Front	\$36,640	\$33,129	3	6	220	20	27	3561	109	192
14	Audi	A6 3.0 Qu.	Sedan	Europe	All	\$39,640	\$35,992	3	6	220	18	25	3880	109	192
15	Audi	A4 3.0 cor	Sedan	Europe	Front	\$42,490	\$38,325	3	6	220	20	27	3814	105	180
16	Audi	A4 3.0 Qu.	Sedan	Europe	All	\$44,240	\$40,075	3	6	220	18	25	4013	105	180
17	Audi	A6 2.7 Tur	Sedan	Europe	All	\$42,840	\$38,840	2.7	6	250	18	25	3836	109	192
18	Audi	A6 4.2 Qu.	Sedan	Europe	All	\$49,690	\$44,936	4.2	8	300	17	24	4024	109	193

Total rows: 428 Total columns: 15

Data Set: SASHelp.Cars

← → Rows 1-100

	Make	Model	Type	Origin	DriveT...	MSRP	Invoice	Engin...	Cylind...	Horse...	MPG_...	MPG_...	Weight	Wheel...	Leng...
1	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6	265	17	23	4451	106	189
2	Acura	RSX Type	Sed									31	2778	101	172
3	Acura	TSX 4dr	Sed									29	3230	105	183
4	Acura	TL 4dr	Sed									28	3575	108	186
5	Acura	3.5 RL 4dr	Sed									24	3880	115	197
6	Acura	3.5 RL w/I	Sed									24	3893	115	197
7	Acura	NSX coup	Spo									24	3153	100	174
8	Audi	A4 1.8T 4c	Sed									31	3252	104	179
9	Audi	A41.8T co	Sed									30	3638	105	180
10	Audi	A4 3.0 4dr	Sed									28	3462	104	179
11	Audi	A4 3.0 Qu.	Sed									26	3583	104	179
12	Audi	A4 3.0 Qu.	Sed									25	3627	104	179
13	Audi	A6 3.0 4dr	Sed									27	3561	109	192
14	Audi	A6 3.0 Qu.	Sed									25	3880	109	192
15	Audi	A4 3.0 cor	Sed									27	3814	105	180
16	Audi	A4 3.0 Qu.	Sed	Weight	Cylinders	Horsepower	MPG	Year	Acceleration	Origin		25	4013	105	180
17	Audi	A6 2.7 Tur	Sedan	Europe	All	\$42,840	\$38,840	2.7	6	250	18	25	3836	109	192
18	Audi	A6 4.2 Qu.	Sedan	Europe	All	\$49,690	\$44,936	4.2	8	300	17	24	4024	109	193

- **What?**
 - One quantitative attribute
 - One categorical attribute
- **Why?**
 - Part-whole relationship
- **How?**
 - Area and angle both encode relative quantity
 - Scalability:
a dozen categories



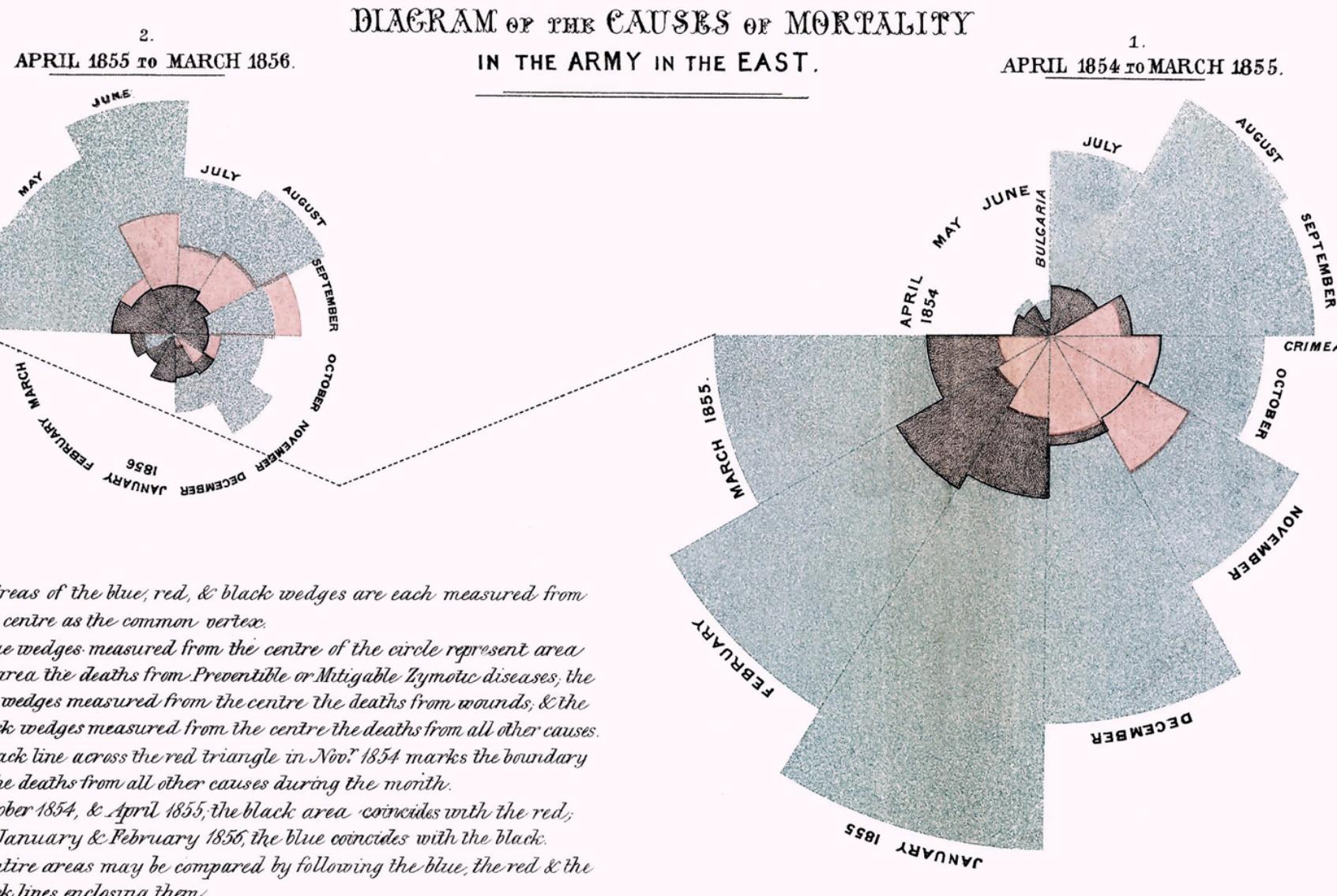
Internet meme: source unknown

Remove
to improve
the **pie chart** edition

When and how to use pie charts?

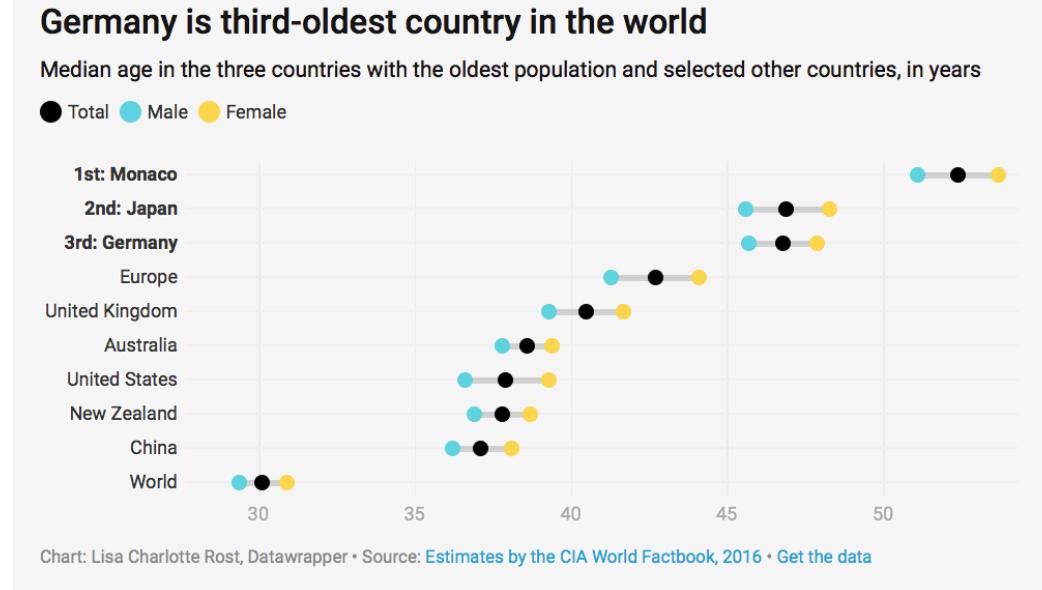


Idiom: polar area chart



Idiom: dot plot

- **What?**
 - Multiple quantitative attributes
 - One categorical attribute
- **Why?**
 - Find trends, outliers, ranges
- **How?**
 - Position
 - Scalability:
a dozen categories



VAD chapter 9

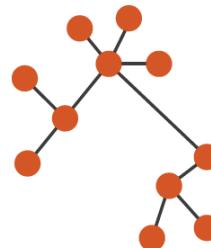
Arrange Networks and Trees

→ Node–Link Diagrams

Connection Marks

NETWORKS

TREES

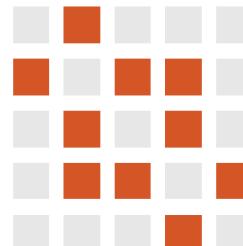


→ Adjacency Matrix

Derived Table

NETWORKS

TREES



→ Enclosure

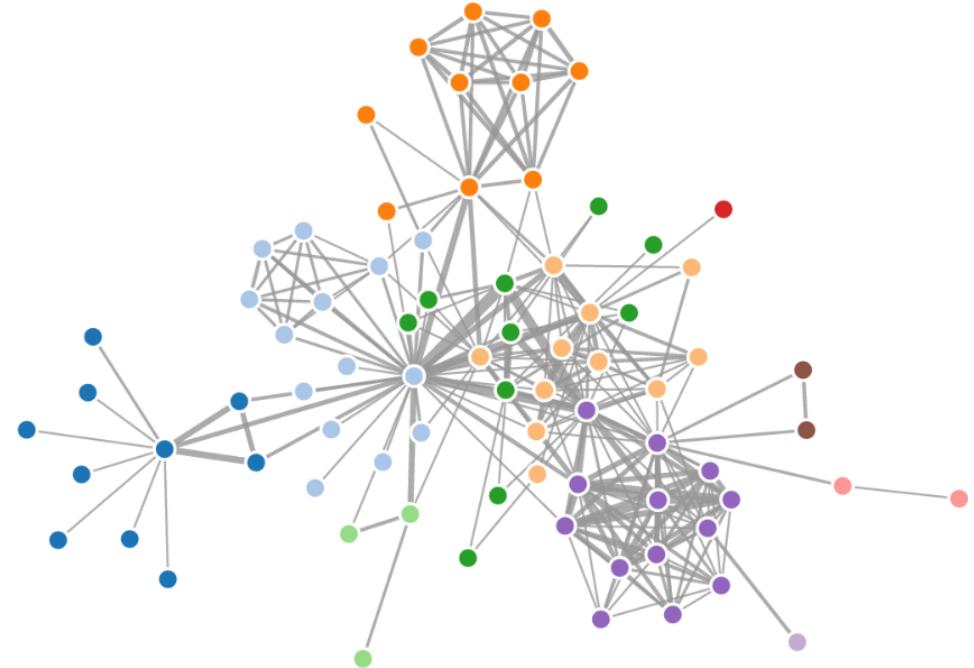
Containment Marks

NETWORKS

TREES



- **What?**
 - Graph / Network
 - Pairs of nodes define links
 - Links can be directed or undirected
- **Why?**
 - Task: Understand connectivity
- **How?**
 - Marks: labelled marks for nodes, lines for links
 - Scalability:
dozens of nodes,
relatively sparse links



<http://marvl.infotech.monash.edu/webcola>

Adjacency Matrix

▪ What?

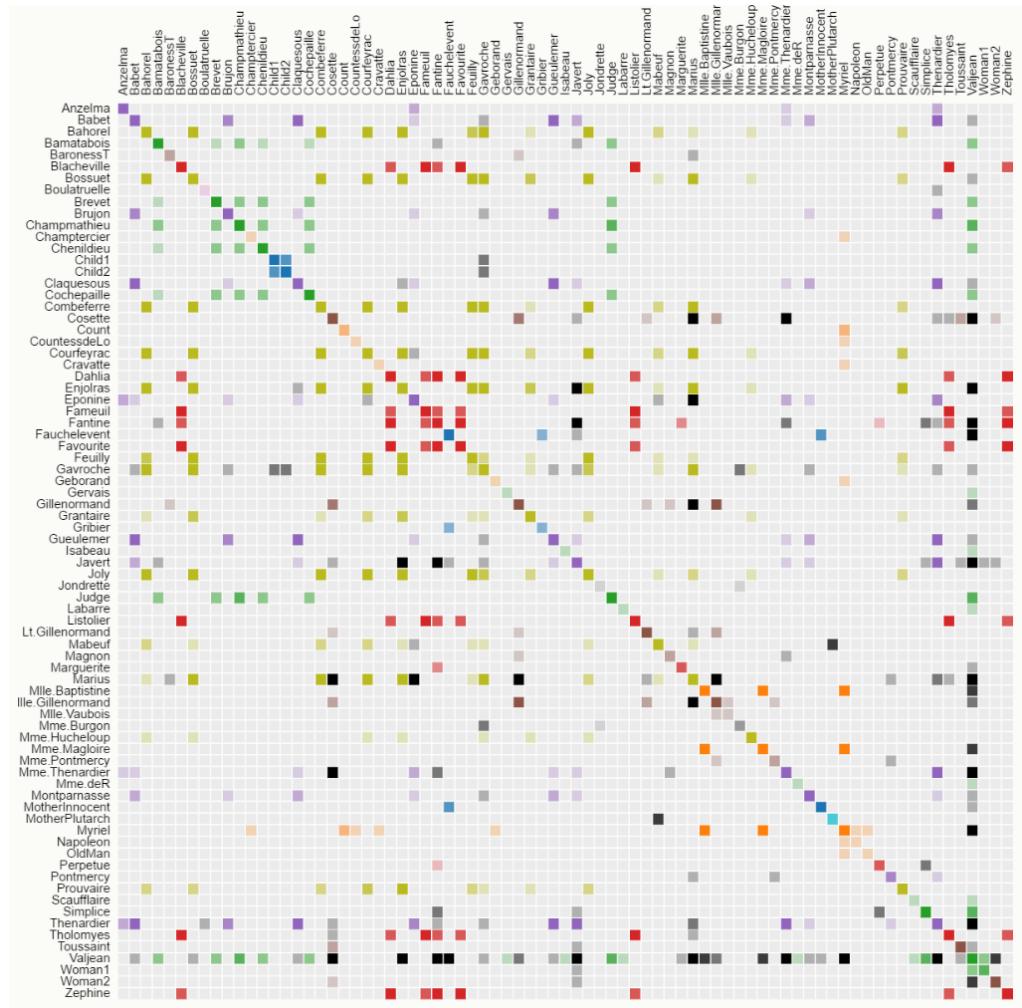
- Graph / Network
- Pairs of nodes define links
- Links can be directed or undirected

▪ Why?

- Task: Understand connectivity

▪ How?

- Marks:
 - Each node is both a column and a row
 - filled cells indicate a link
- Scalability: dozens of nodes, dense connectivity can be shown



<http://bl.ocks.org/jdfekete/raw/a8828c88060a1883b3f4/>
61

▪ What?

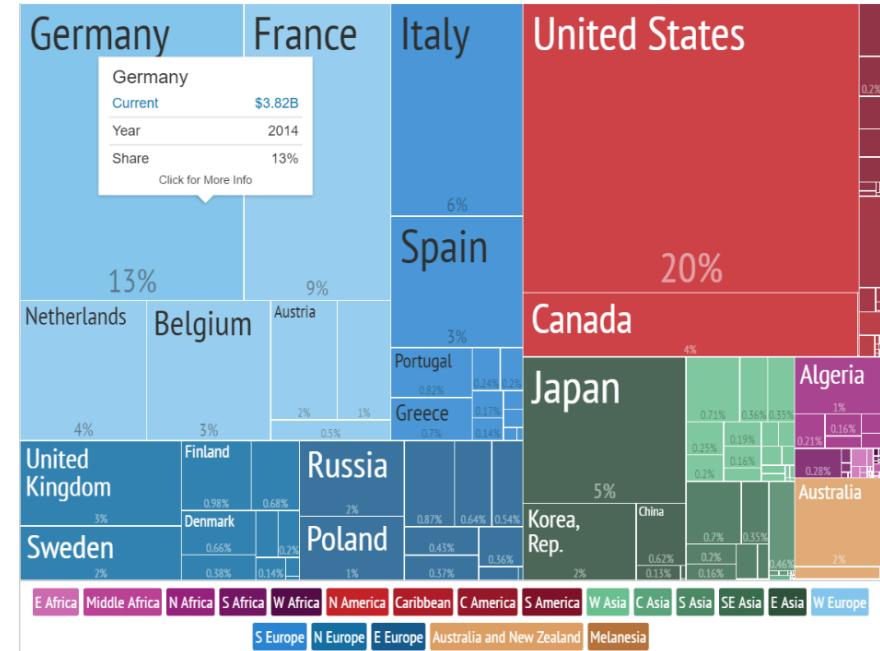
- Hierarchical (tree) structure over categories or items
- Quantitative attributes associated with nodes/items

▪ Why?

- Task: Understand distribution of quantities throughout the hierarchy

▪ How?

- Marks:
 - Areas proportional to quantitative attribute
- Scalability: without [interactivity](#) only larger regions have clearly visible labels



[http://atlas.cid.harvard.edu/
explore/tree_map/import/show/
all/0901/2014/](http://atlas.cid.harvard.edu/explore/tree_map/import/show/all/0901/2014/)