

MS in Business Analytics



DV1 Intro to data visualization



Common charts

The 4 chart segments

Core Charts

Common Charts

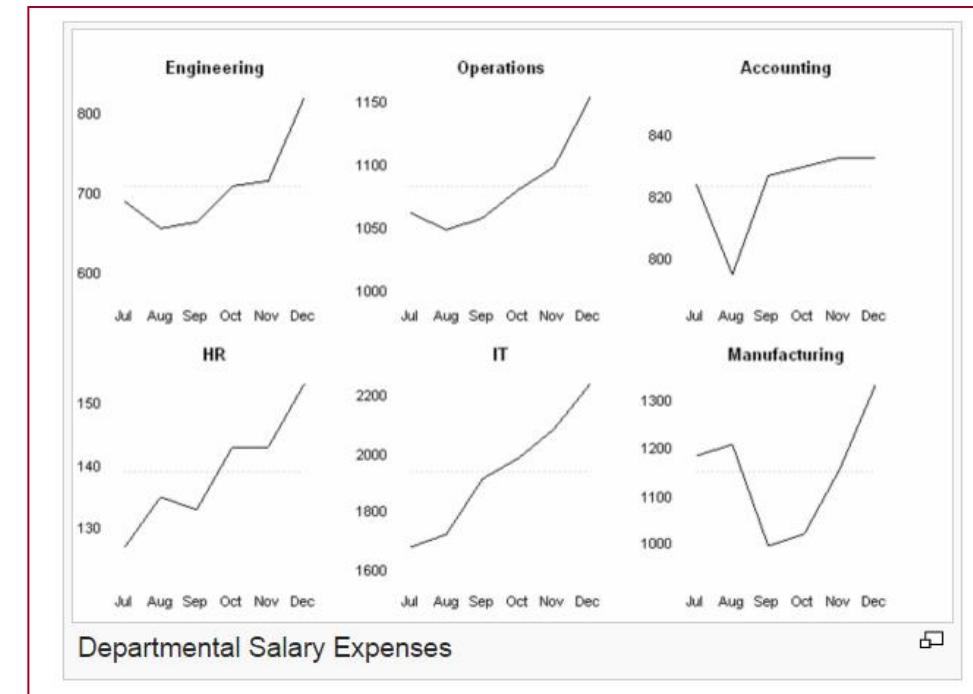
Rare Charts

Risky charts

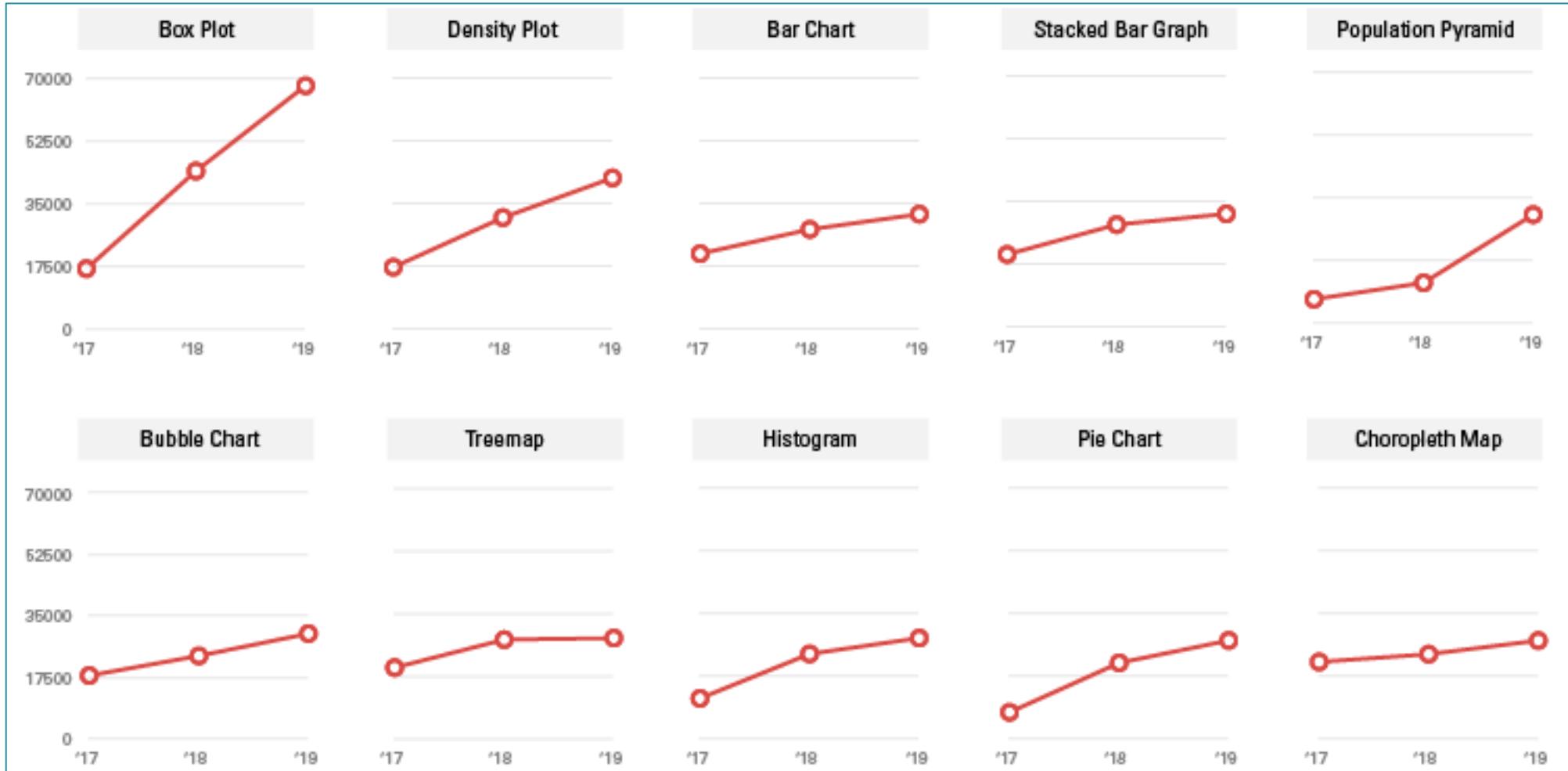
Trellis

Trellis chart

- Trellis
 - Comparison
 - same axis and scaling, different data
 - „small multiples” (Tufte)

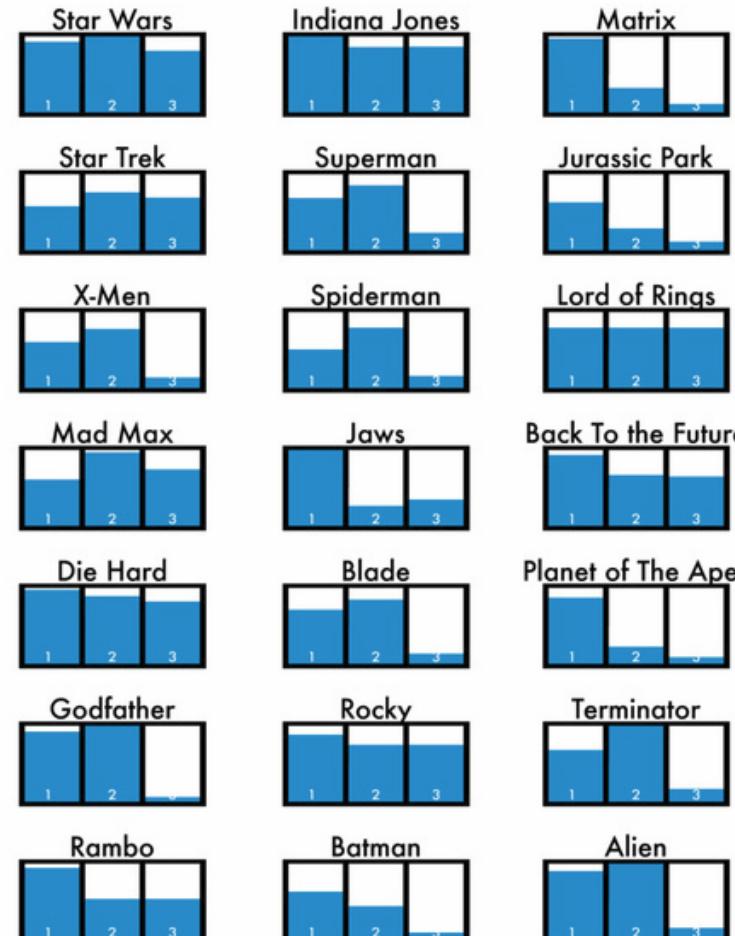


Top 10 Charts in 2019



Trellis chart

THE TRILOGY METER

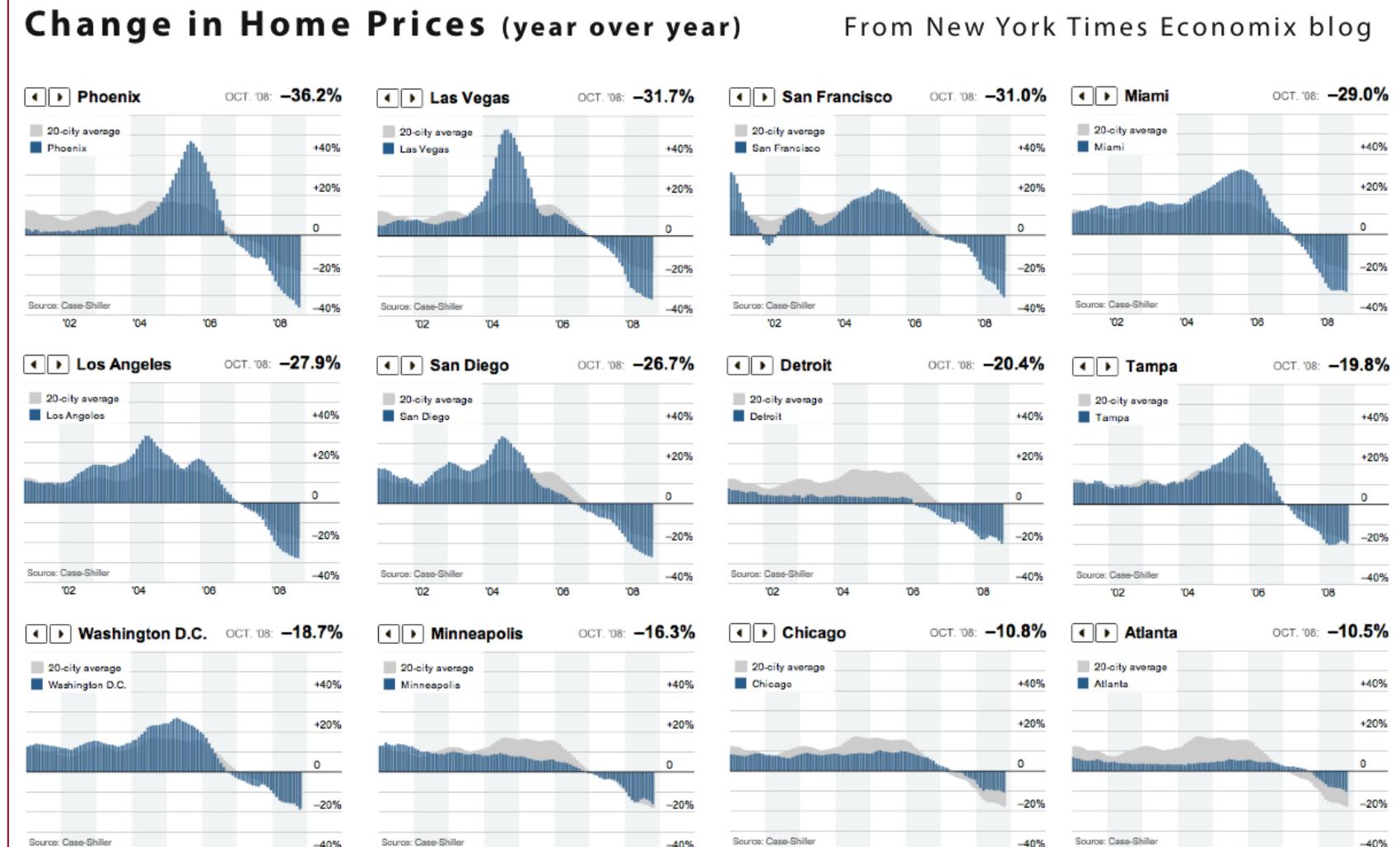


#1 In A Series of Pop Cultural Charts

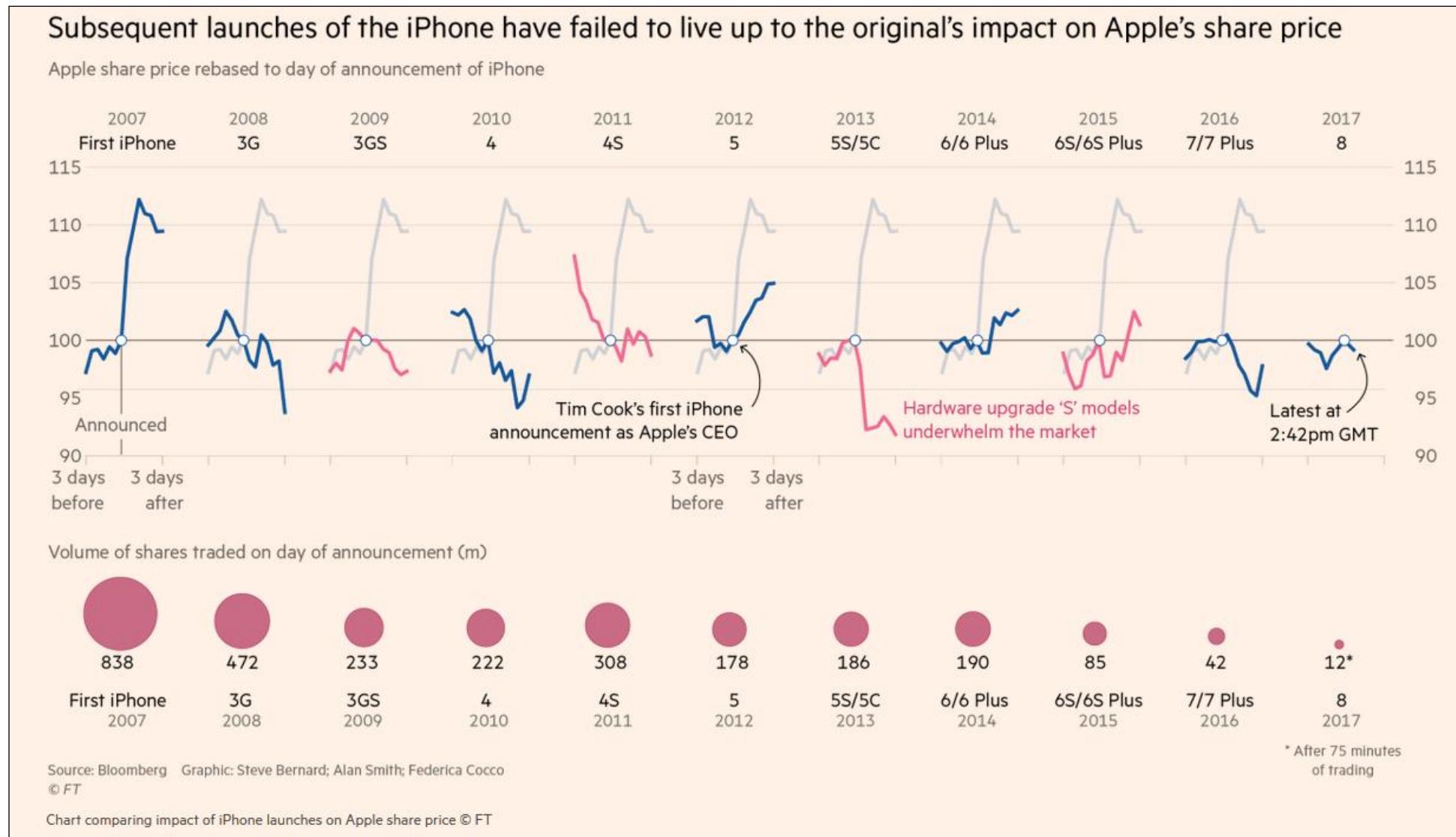
DANMETH.COM

[Trilogy Meter](#) by Dan Meth

Trellis chart



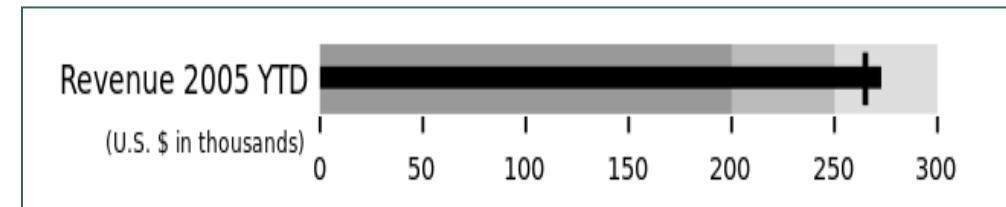
Trellis chart



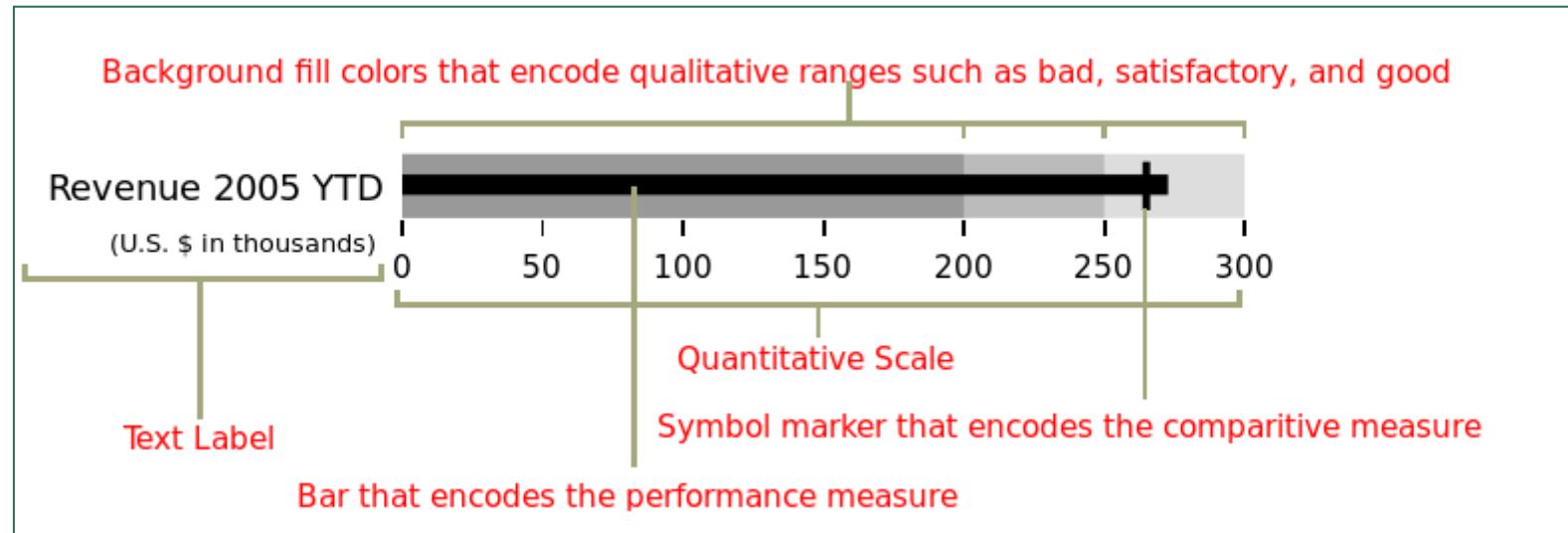
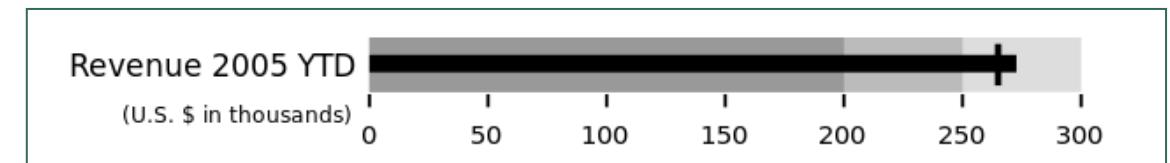
Bullet chart

Bullet chart

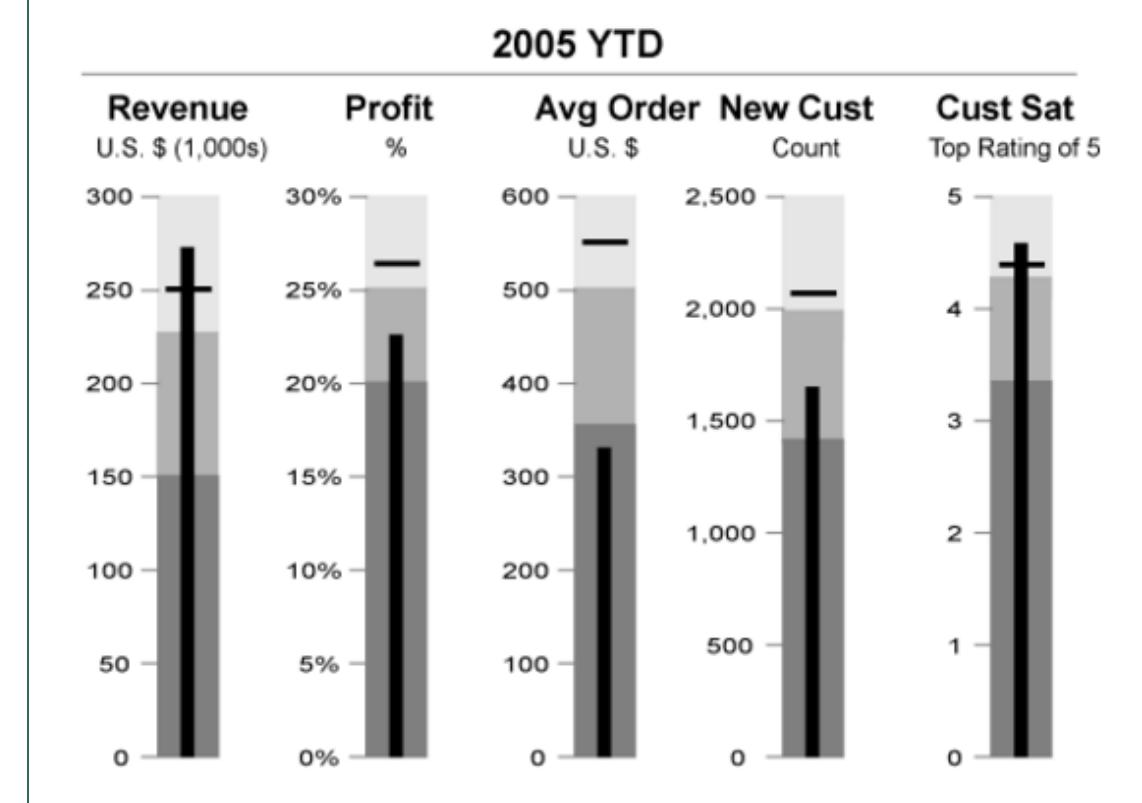
- Bullet chart
 - Actual / Plan / Expectations
 - Fits in a small space, compact, intense
 - Stephen Few, 2005



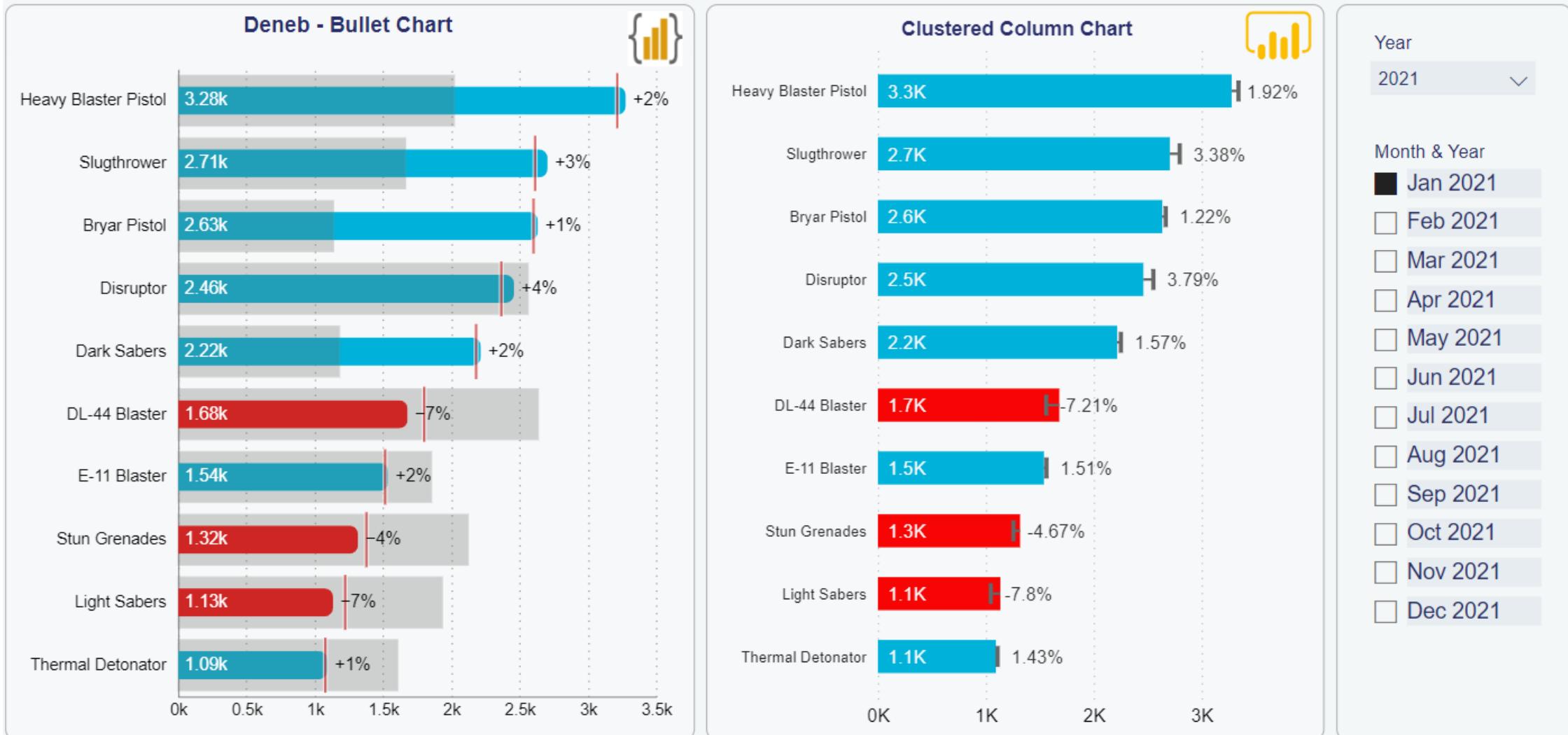
Bullet chart



Bullet chart



The Bullet Chart - Deneb vs PBI Standard Clustered Column Chart



Tree map

Tree map

Globally there are 746 million people in extreme poverty (in 2013)

Extreme poverty is defined as living with less than \$1.90/day.

This is measured in international dollars (i.e. price differences between countries are taken into account).

OurWorld
in Data

Africa (383 million)

Nigeria
86M

Democratic
Republic
of Congo
55.1M

Tanzania
22M

Uganda
12.7M
Malawi
11.4M
Kenya
10.9M
Zambia
9.2M
Niger
8.6M

Ethiopia
20.4M

Mali
8.6M
Burkina
Faso
7.7M
Rwanda
6.7M
Angola
6.4M
Cameroon
5.8M

South Africa
8.5M

Cote d'Ivoire
5.5M
Chad
4.6M
Guinea
4.3M
Central
African
Rep.
3.8M
Togo
3.6M

Madagascar
17.9M

South Sudan
8.1M
Senegal
5.3M
Sudan
3.6M
Zimbabwe
2.6M
Congo
1.6M
Guinea
Bissau
1.2M

Mozambique
15.9M

Burundi
7.9M
Benin
5.2M
Ghana
3.2M
Liberia
2.3M
Lesotho
1.2M
Sierra Leone
2M

Asia (327 million)

India
218M

China
25.1M

Indonesia
24.7M

Brazil
9.9M

Bangladesh
18.4M

Philippines
11.4M

Uzbekistan
7M

Pakistan
12.7M
Vietnam
2.9M
Nepal
2.4M

Colombia
2.9M
Venezuela
2.4M

Haiti
5.5M

Mexico
3.5M

Papua
New
Guinea
2.5 M

Oceania
(2.5 M)

South America (19M)

North America (13M) Europe (0.7M)

Data source: World Bank (PovcalNet)

The interactive data visualization is available at OurWorldInData.org. There you find the raw data and more visualizations on this topic.

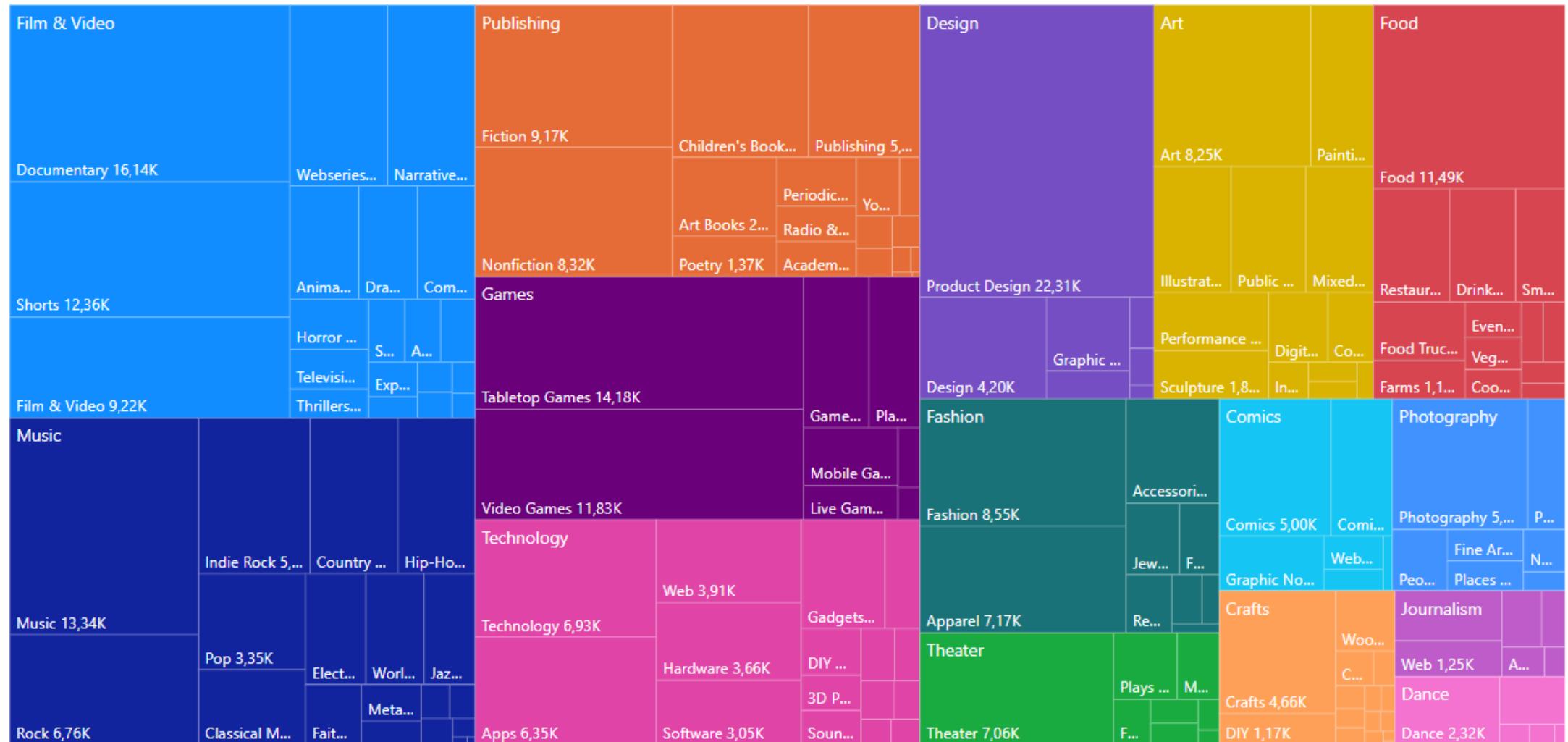
Licensed under CC-BY-SA by the author Max Roser.

Tree map



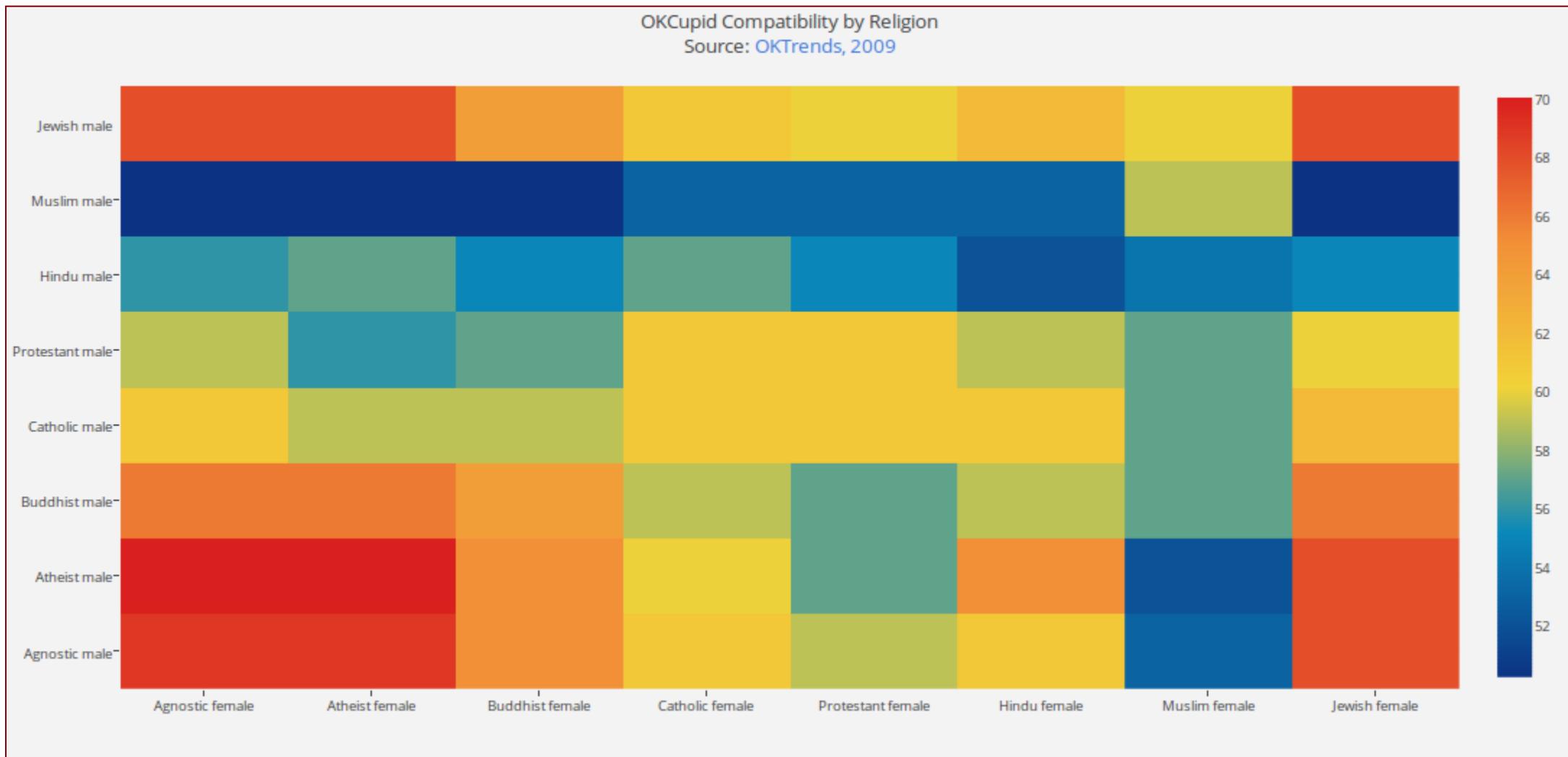
Tree map

Kickstarter projects by category and subcategory



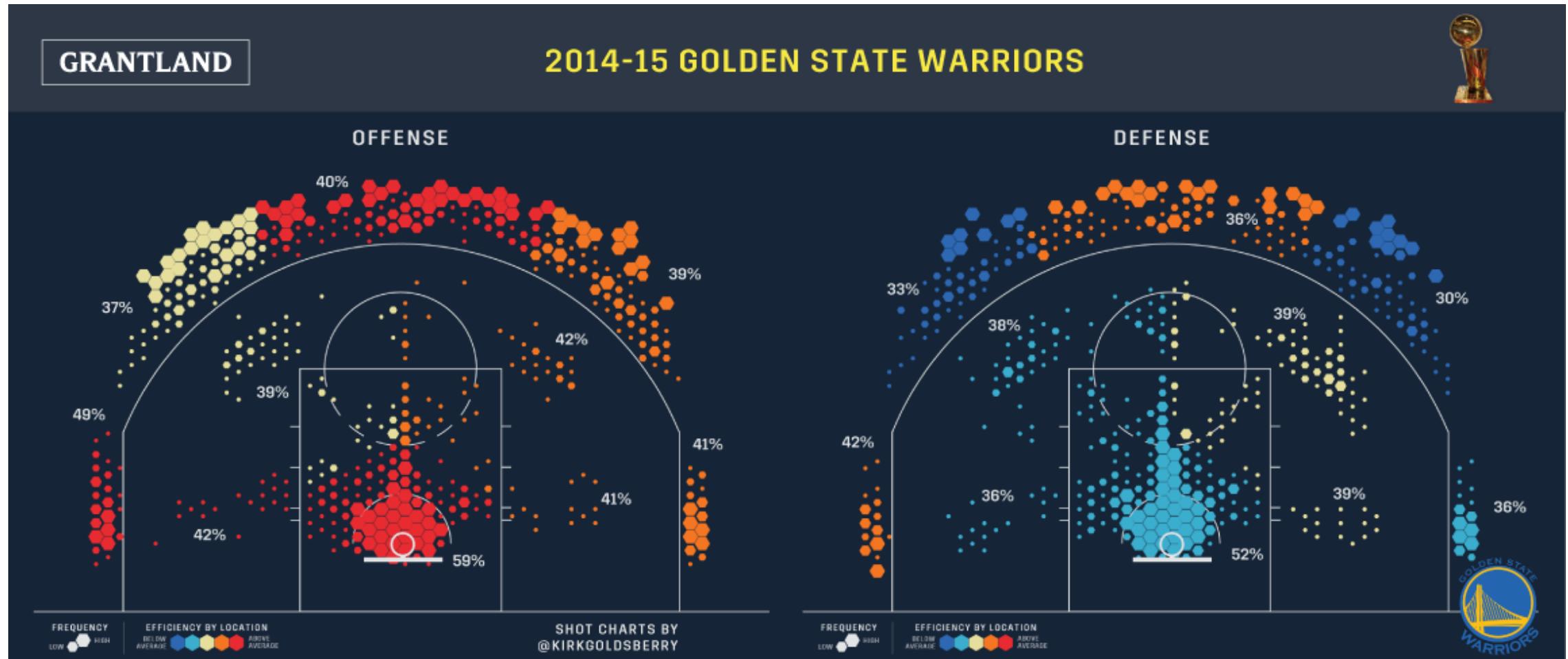
Heat map

Heat map



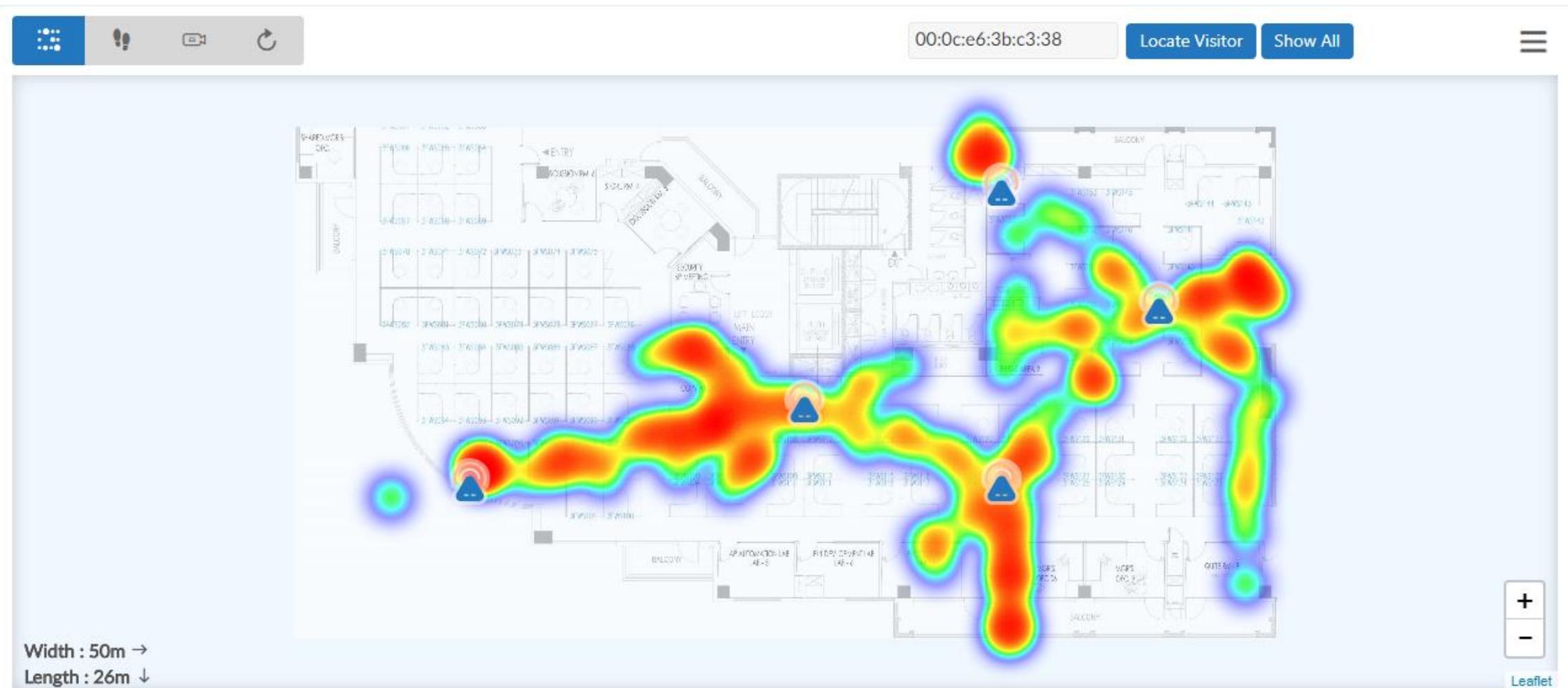
Tableau

Heat map



Shot chart

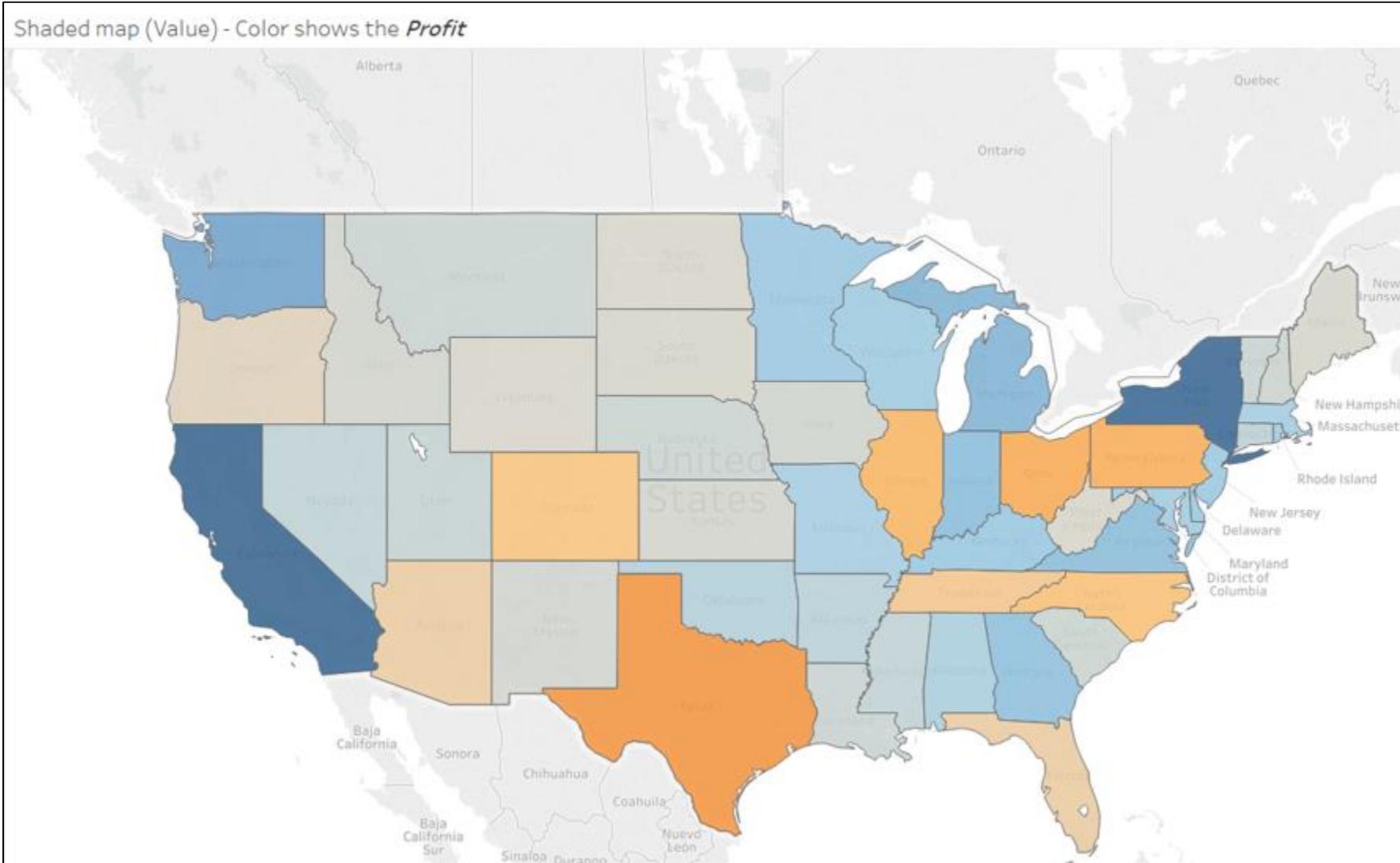
Retail visitor traffic



Visitor traffic in retail space
docs.fortinet.com

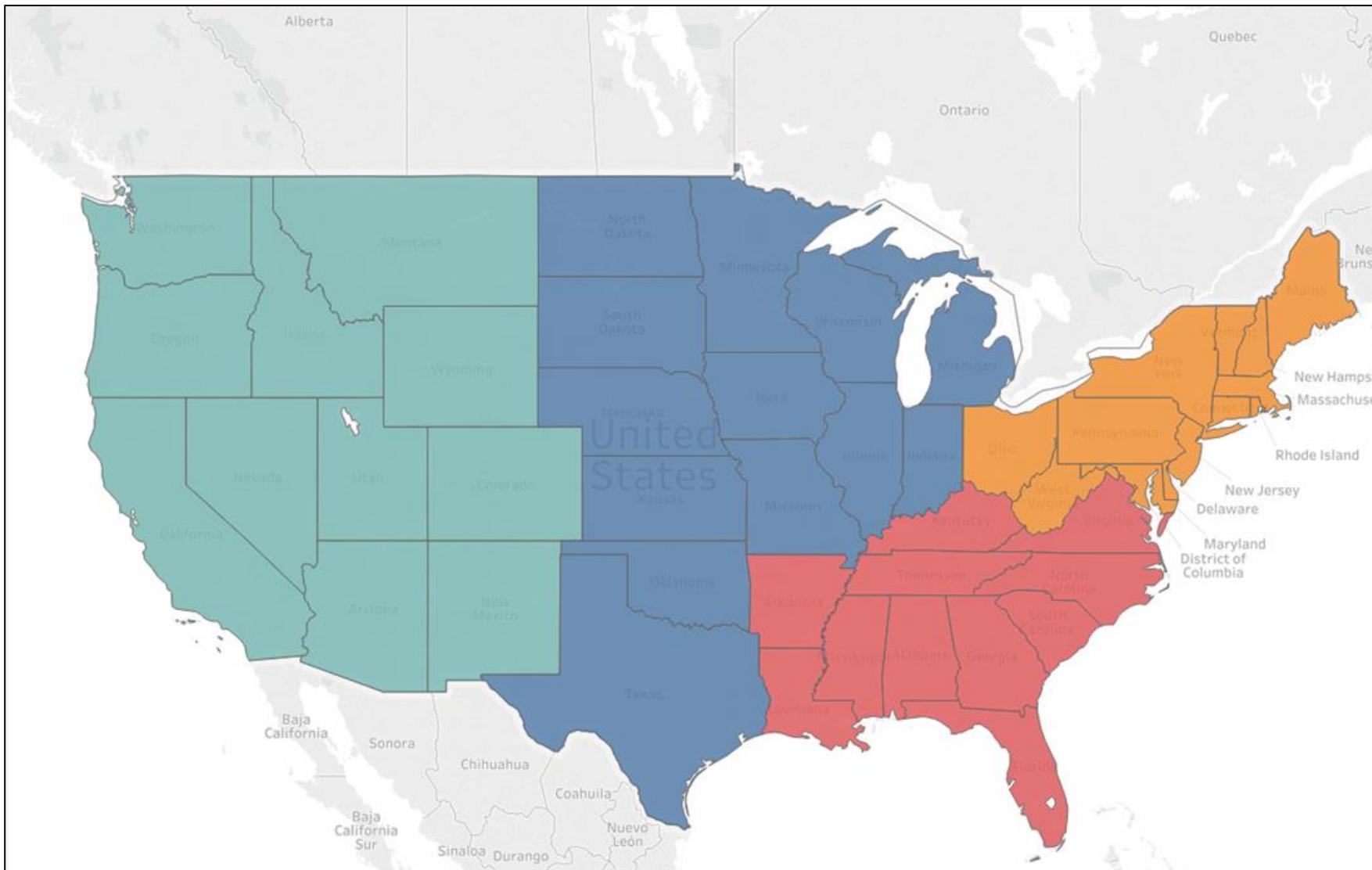
Geomaps

Maps



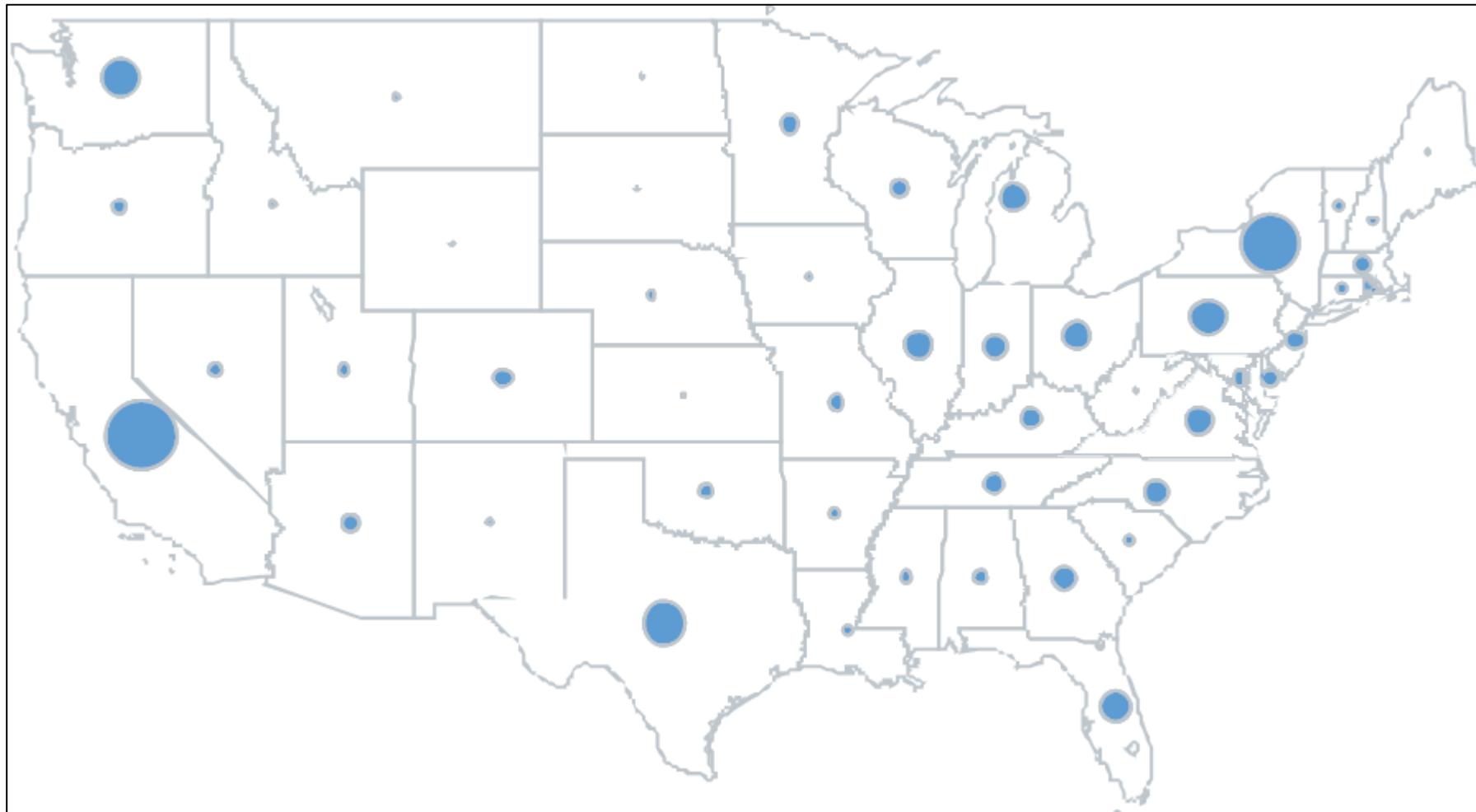
Shaded Map

Maps



Shaded Map

Maps

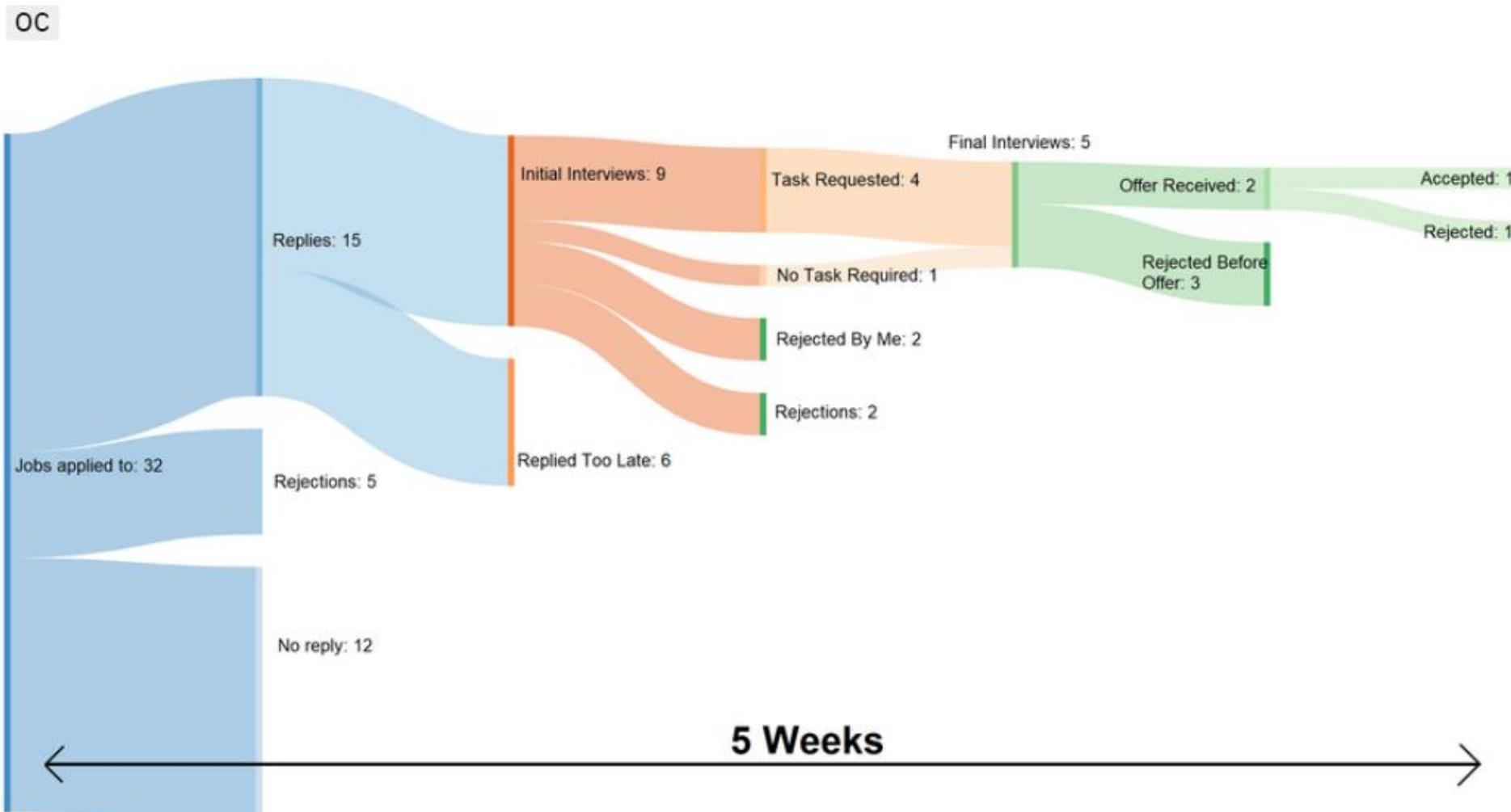


Symbol map

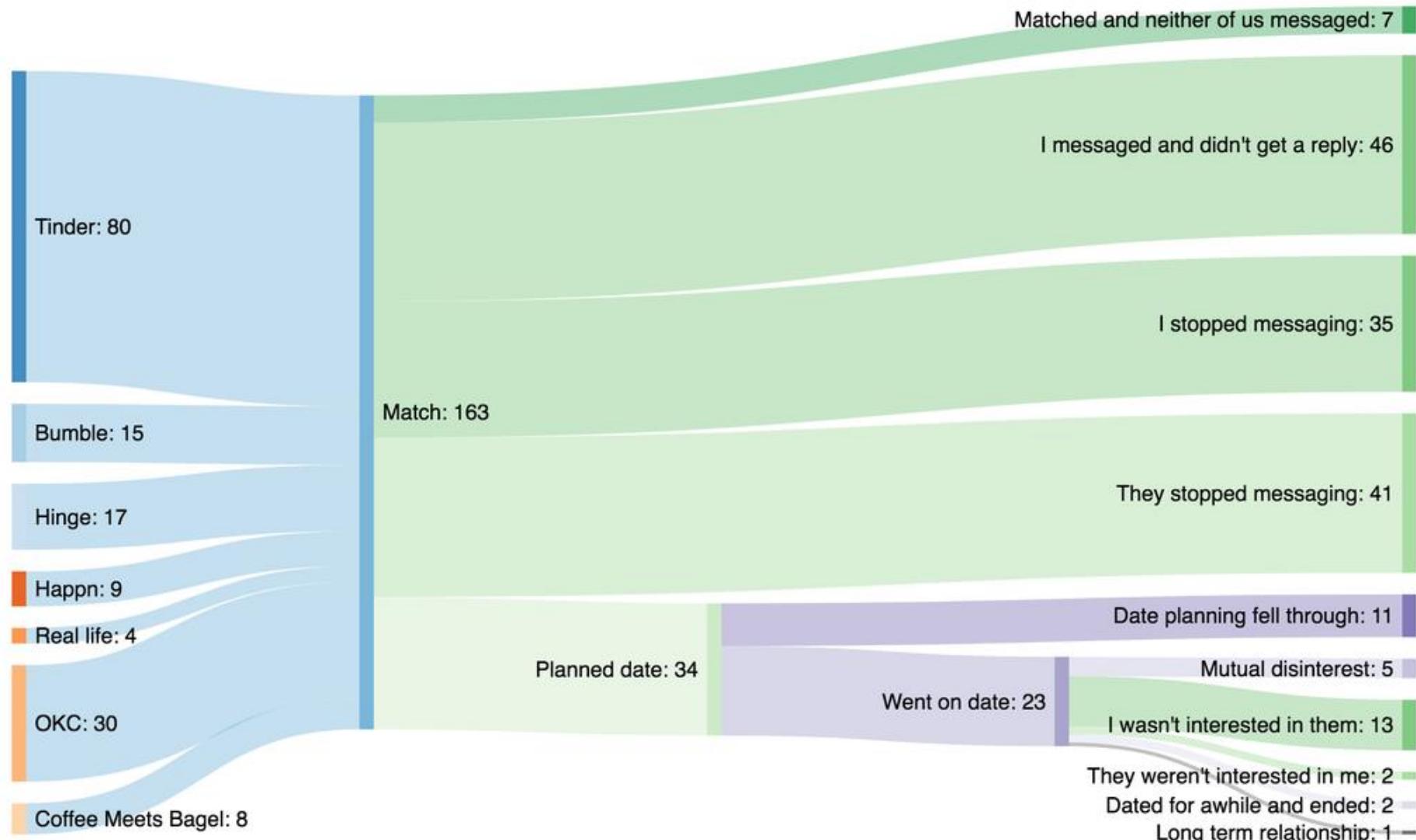
Sankey chart

Sankey chart examples

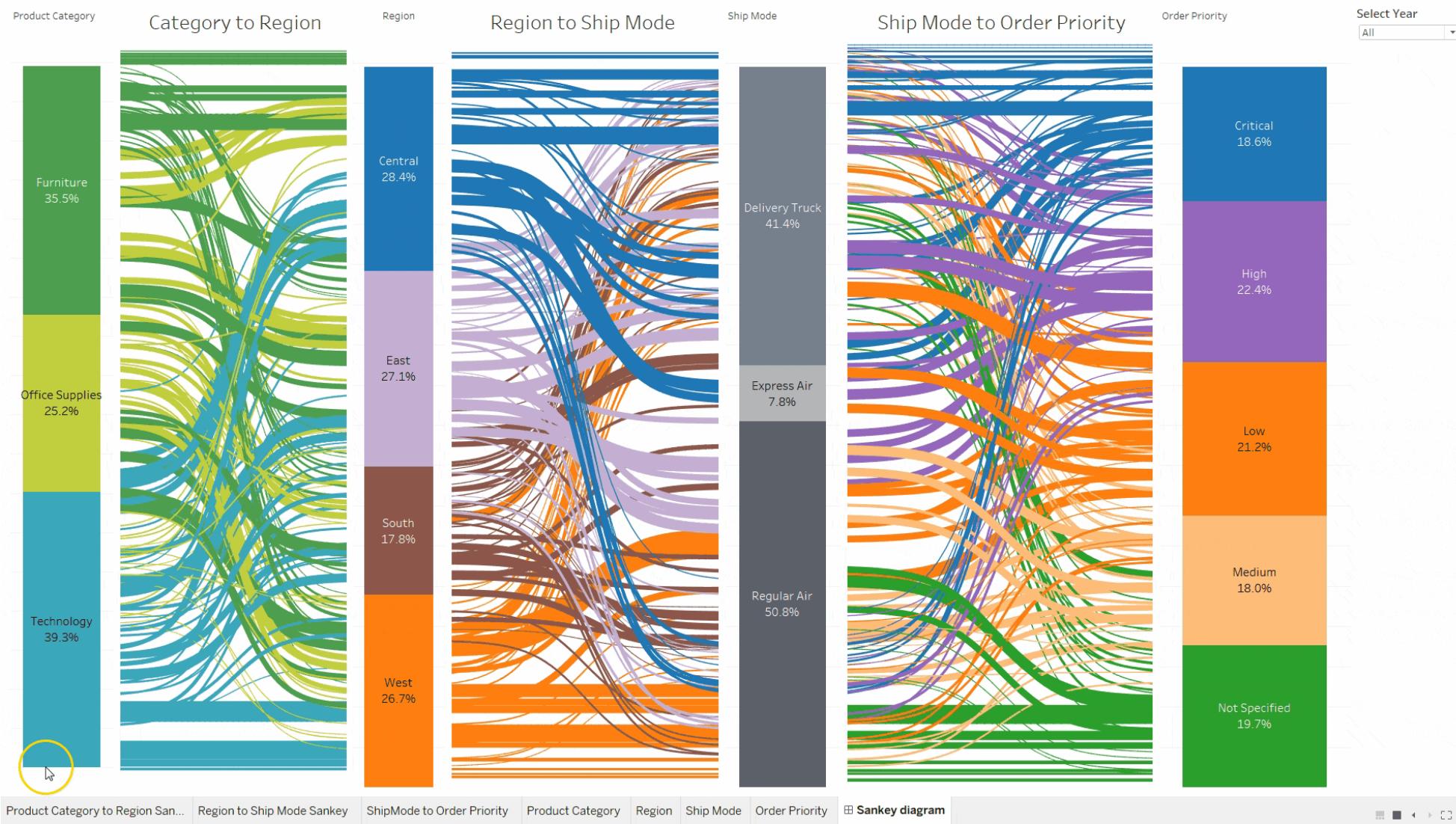
A detailed sankey diagram of my recent job search [OC]



Sankey chart examples



Sankey chart





Risky charts

The 4 chart segments

Core Charts

Common Charts

Rare charts

Risky charts

Pie chart

[HOME](#) > [MARKETS](#)

The Worst Chart In The World

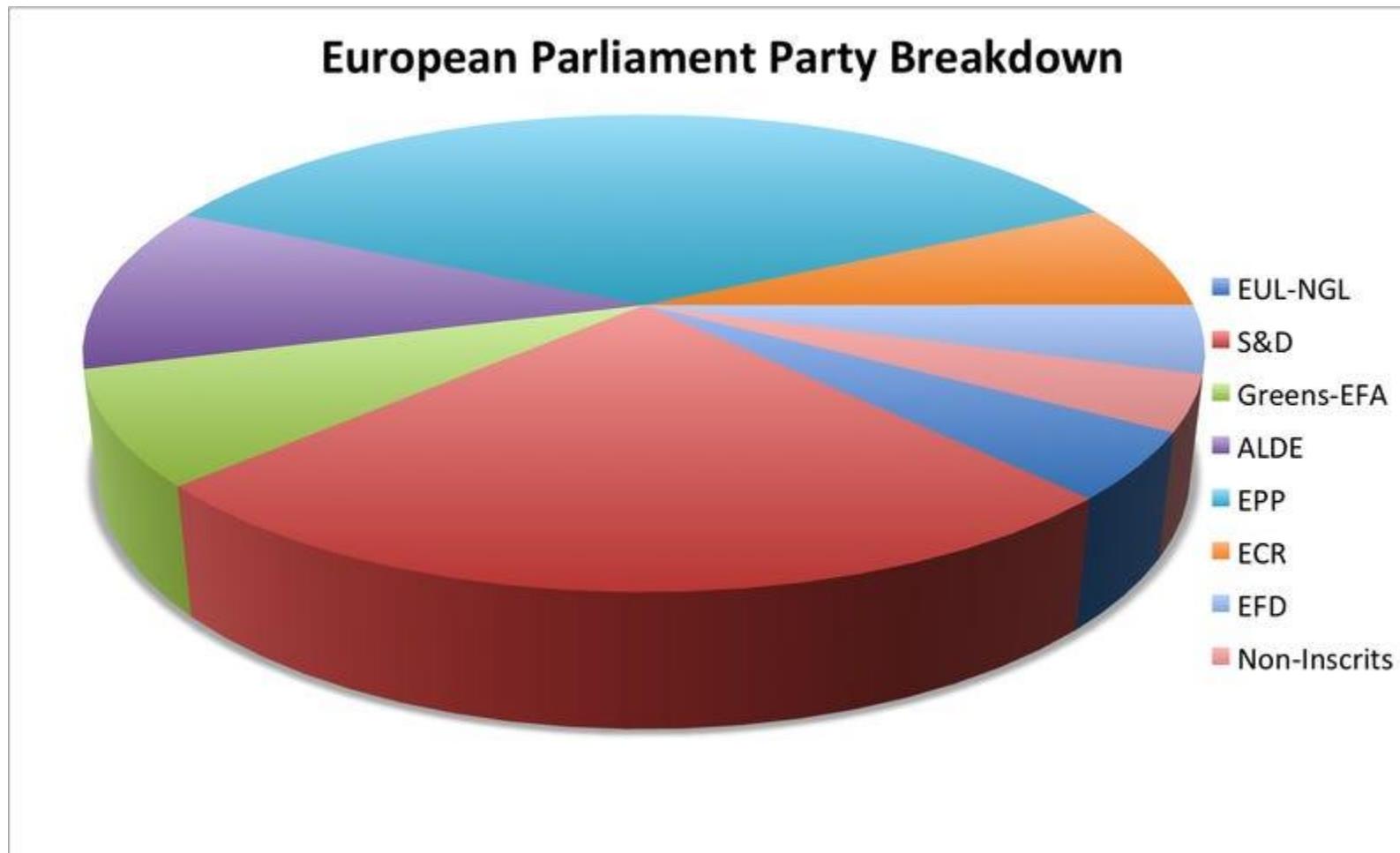
Walt Hickey Jun 17, 2013, 4:39 PM

The pie chart is easily the worst way to convey information ever developed in the history of data visualization.

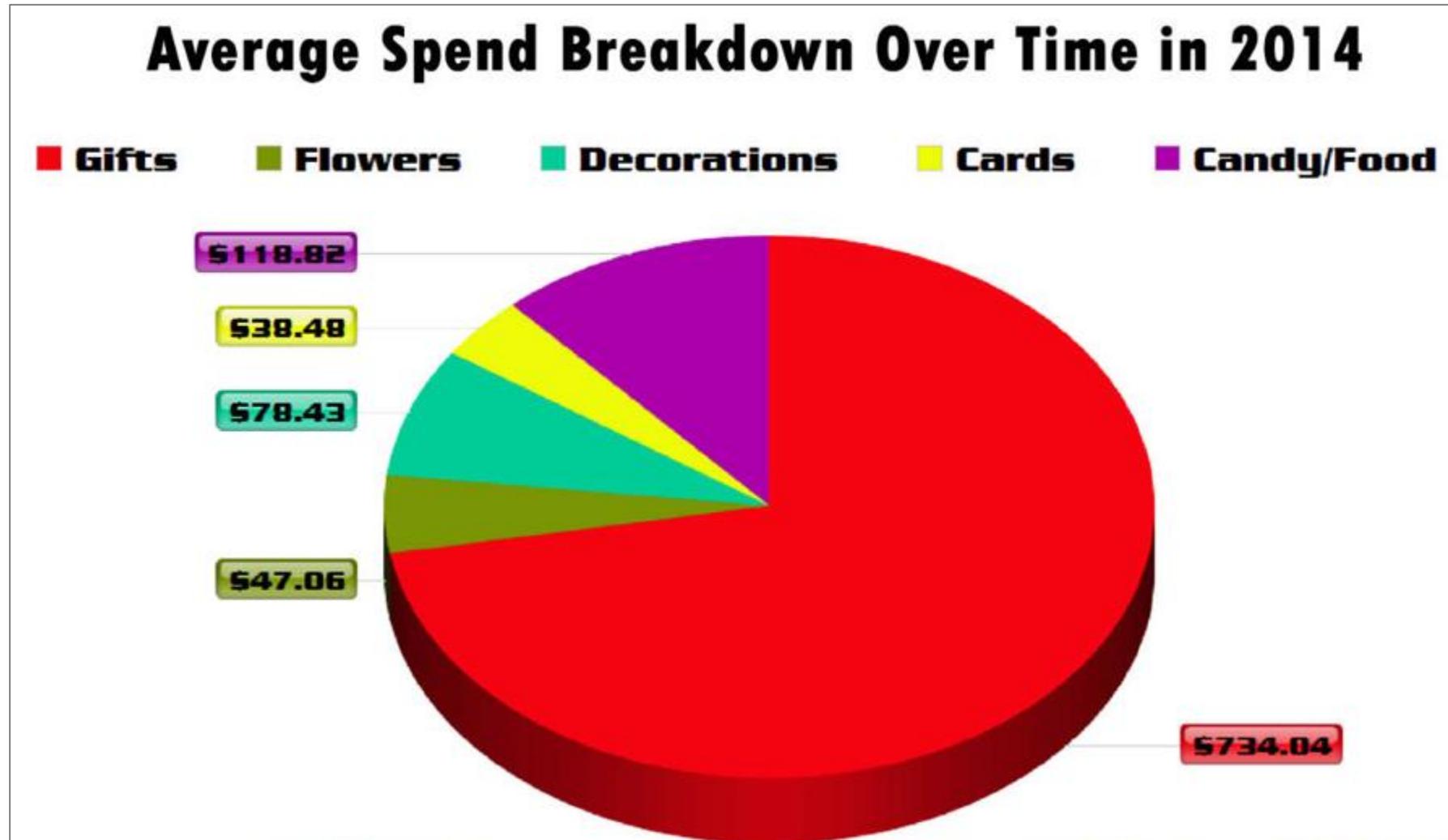
Sure, there are other more cumbersome ways to articulate data. But none have the credibility nor the widespread use that the pie chart has.

Here, I'll explain exactly what's wrong with the pie chart and exactly why you need to stop using them as soon as possible.

Pie chart

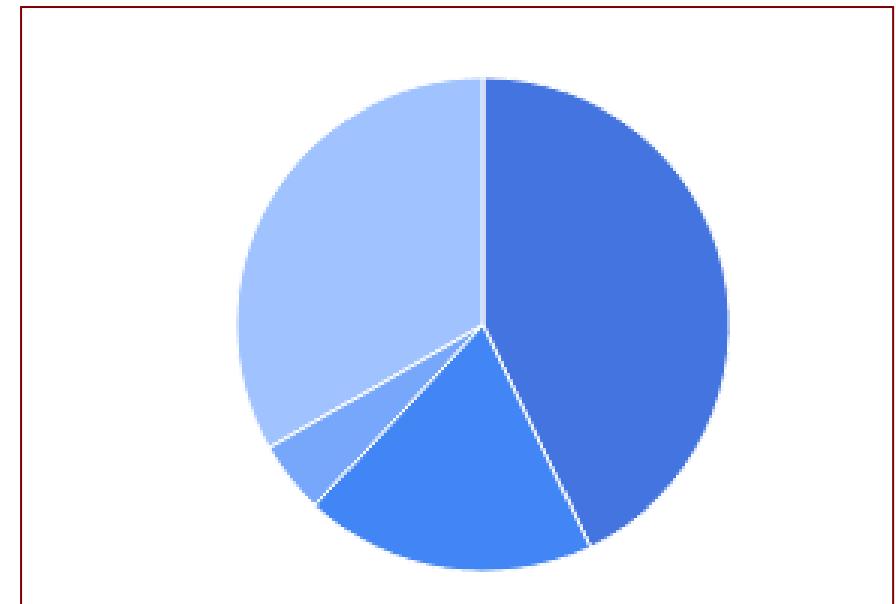


Pie chart



Pie chart

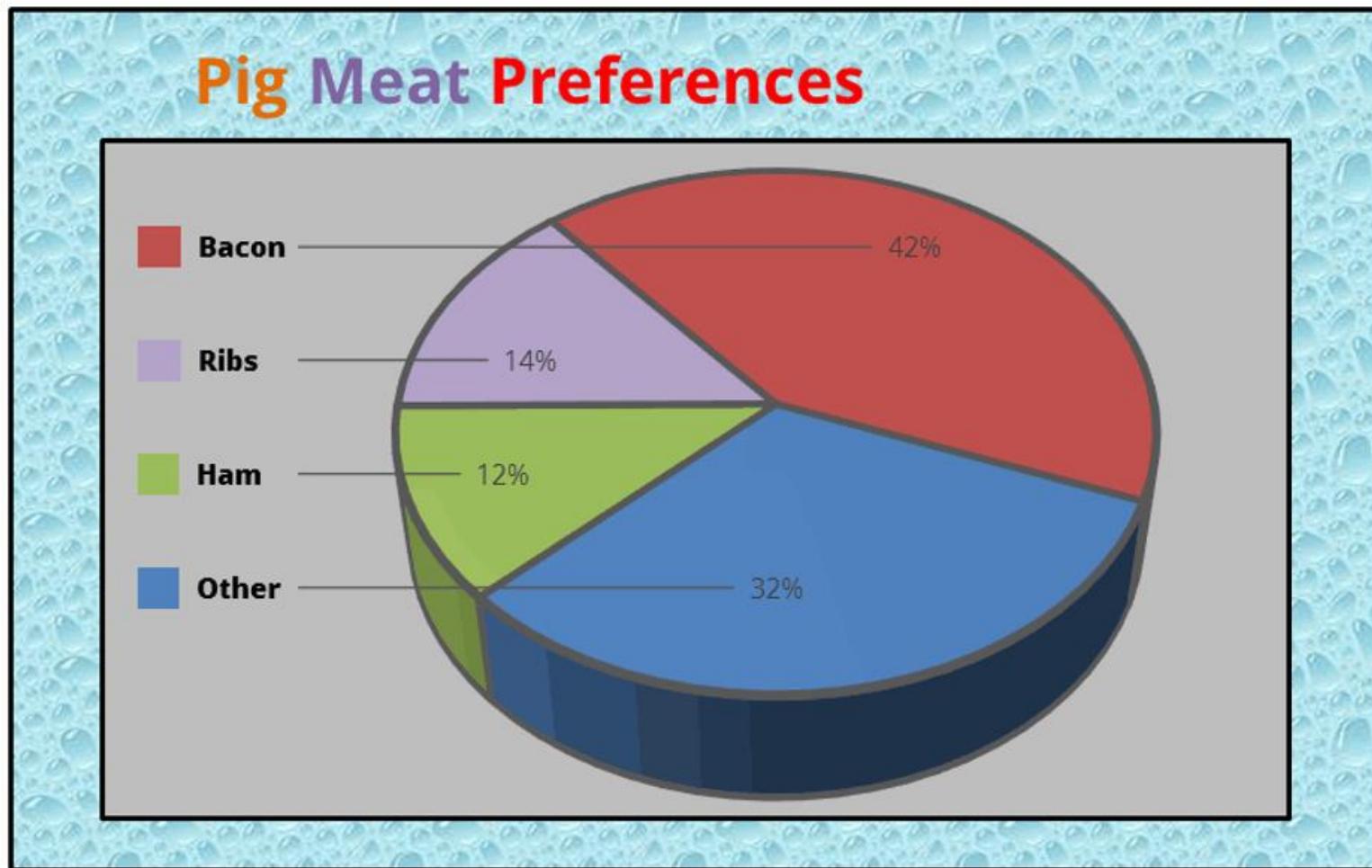
- Pie chart
 - Good for: Comparison, Part-to-whole
- When is it good?
 - Solid, limited formatting
 - Only a few categories
 - No ranking among the categories
 - A single chart (no Trellis)
- And in other cases?



Pie chart

Remove
to improve
the **pie chart** edition

Pie chart

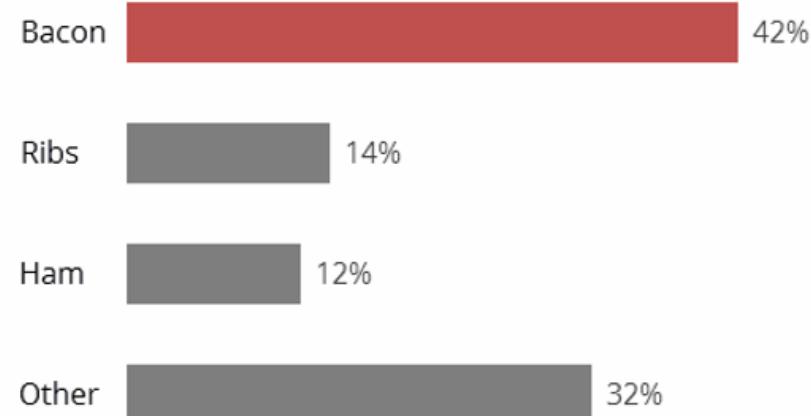


Created by Darkhorse Analytics

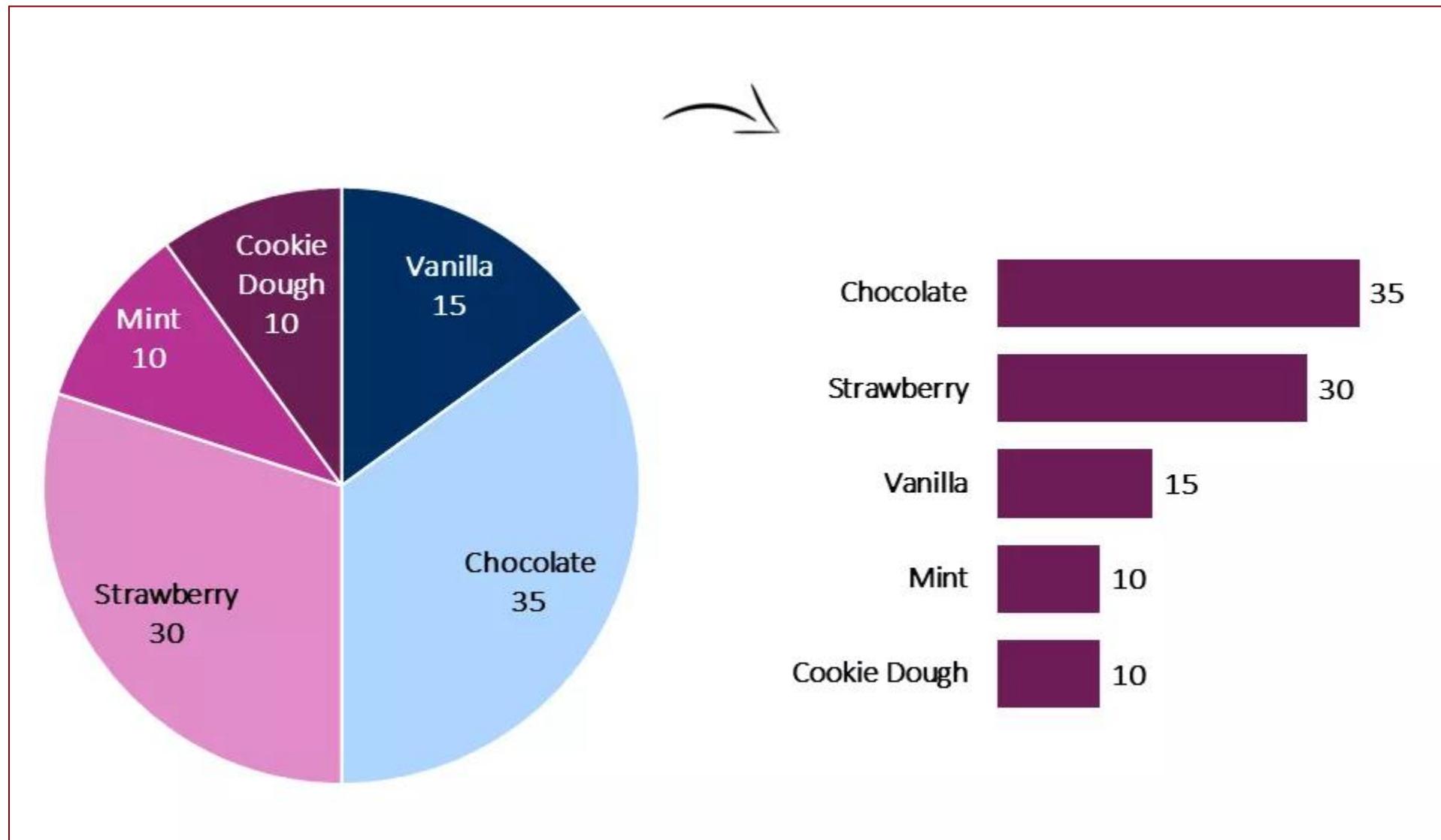
www.darkhorseanalytics.com

Remove
to improve
the **pie chart** edition

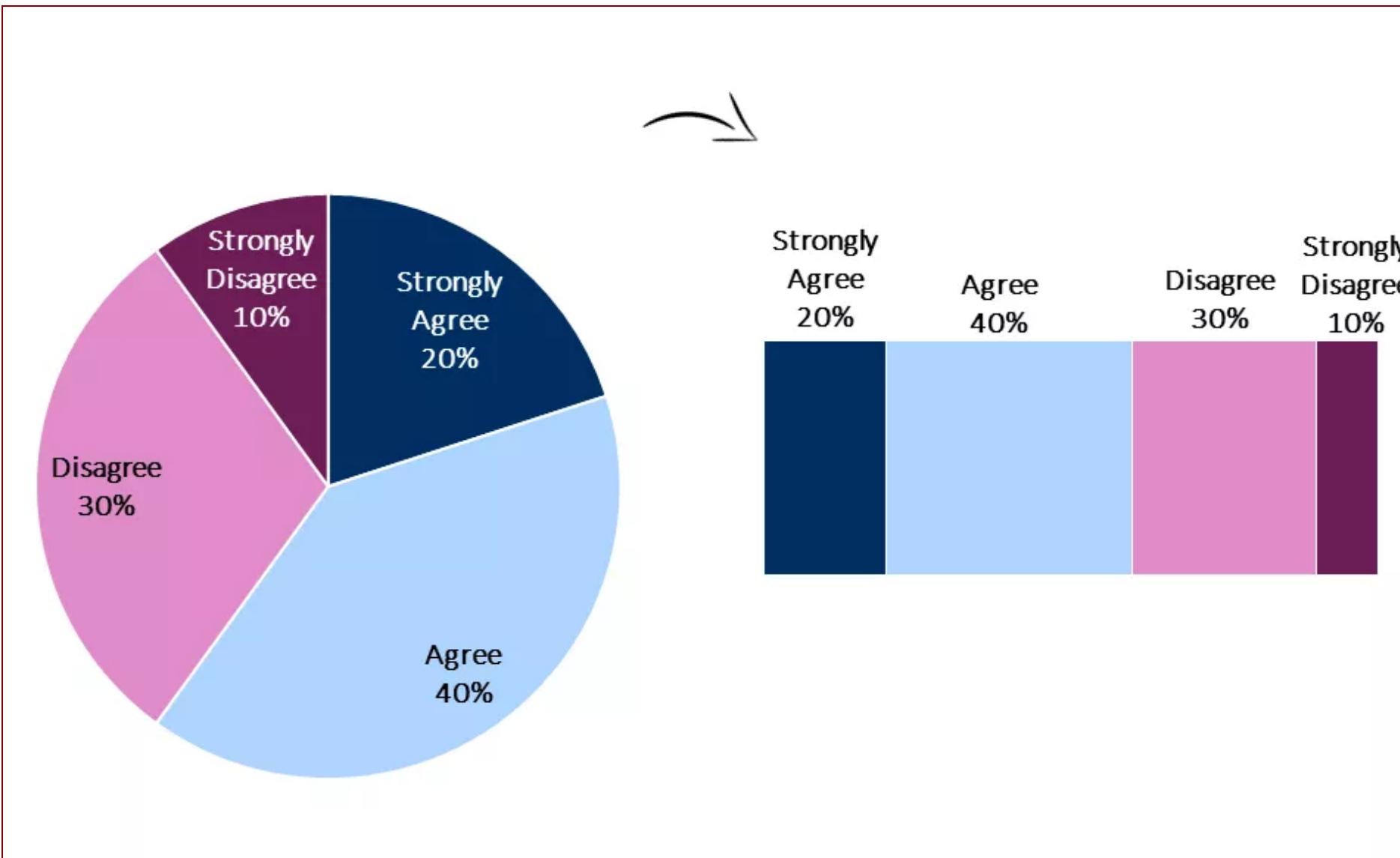
Pig Meat Preferences



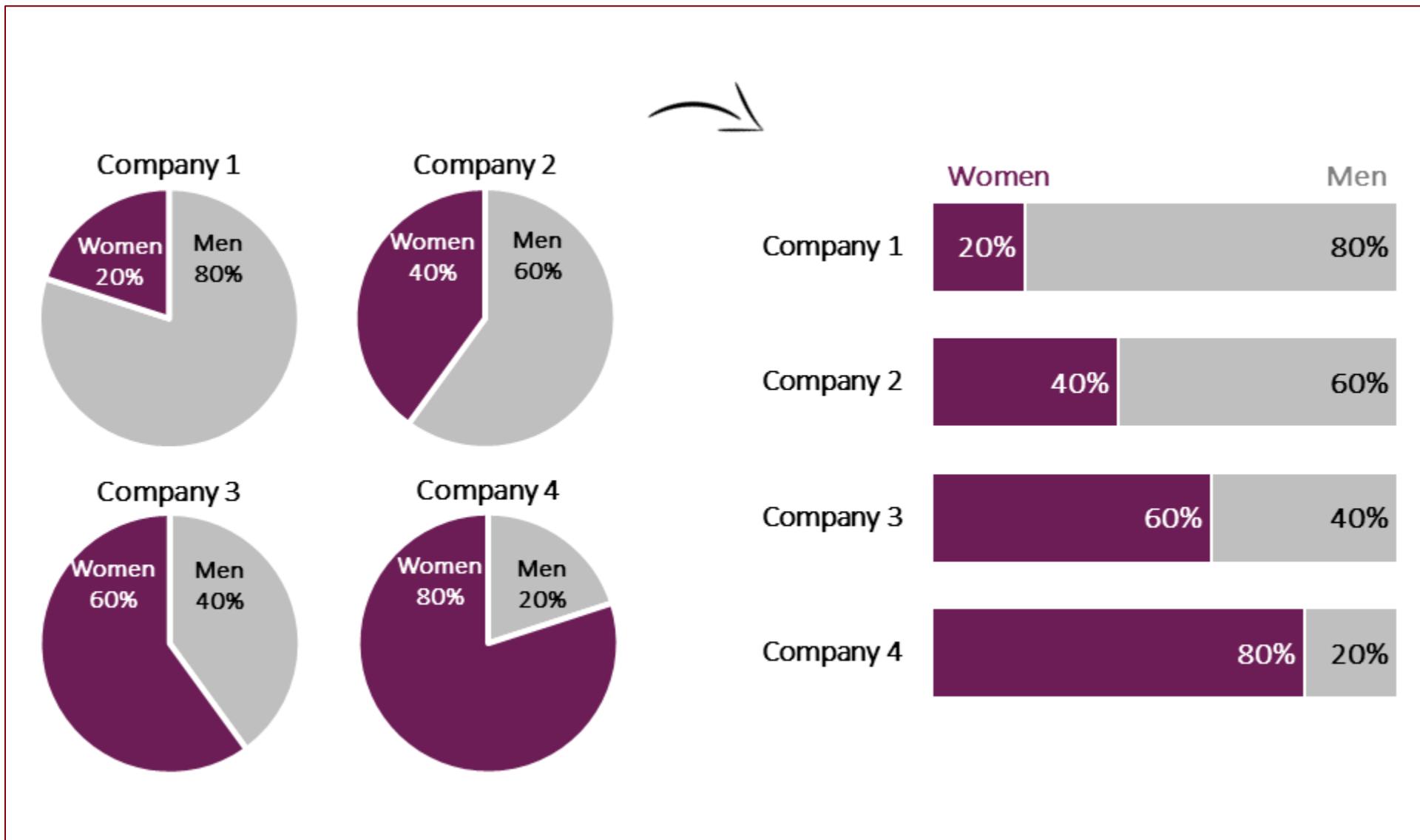
Pie chart



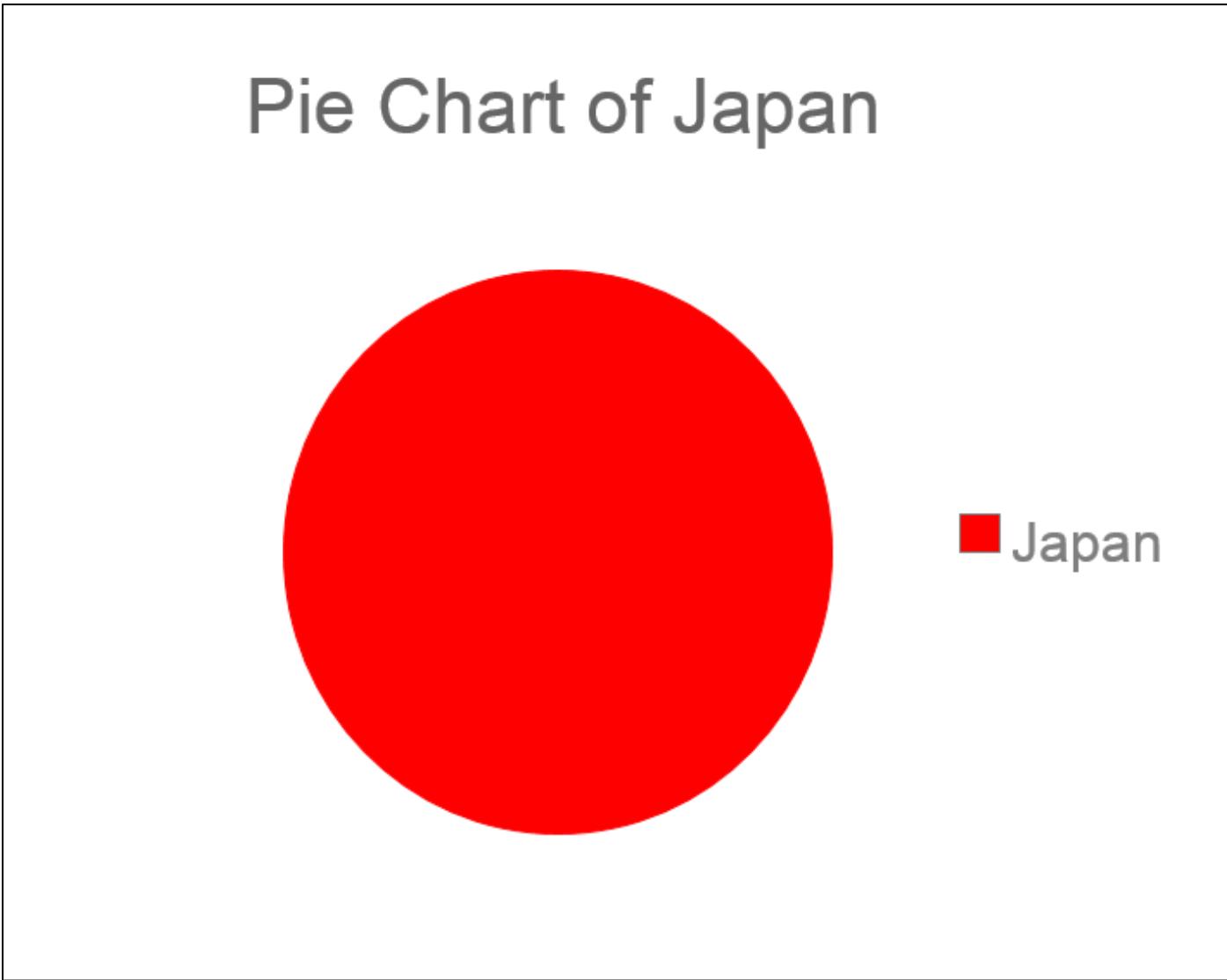
Pie chart



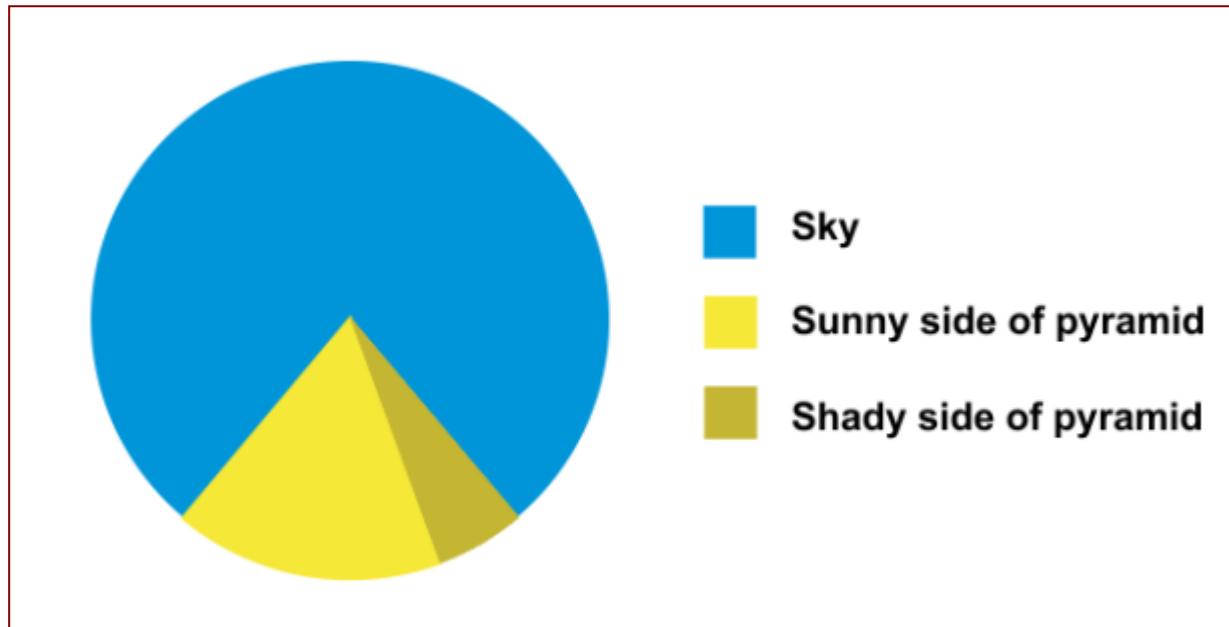
Pie chart



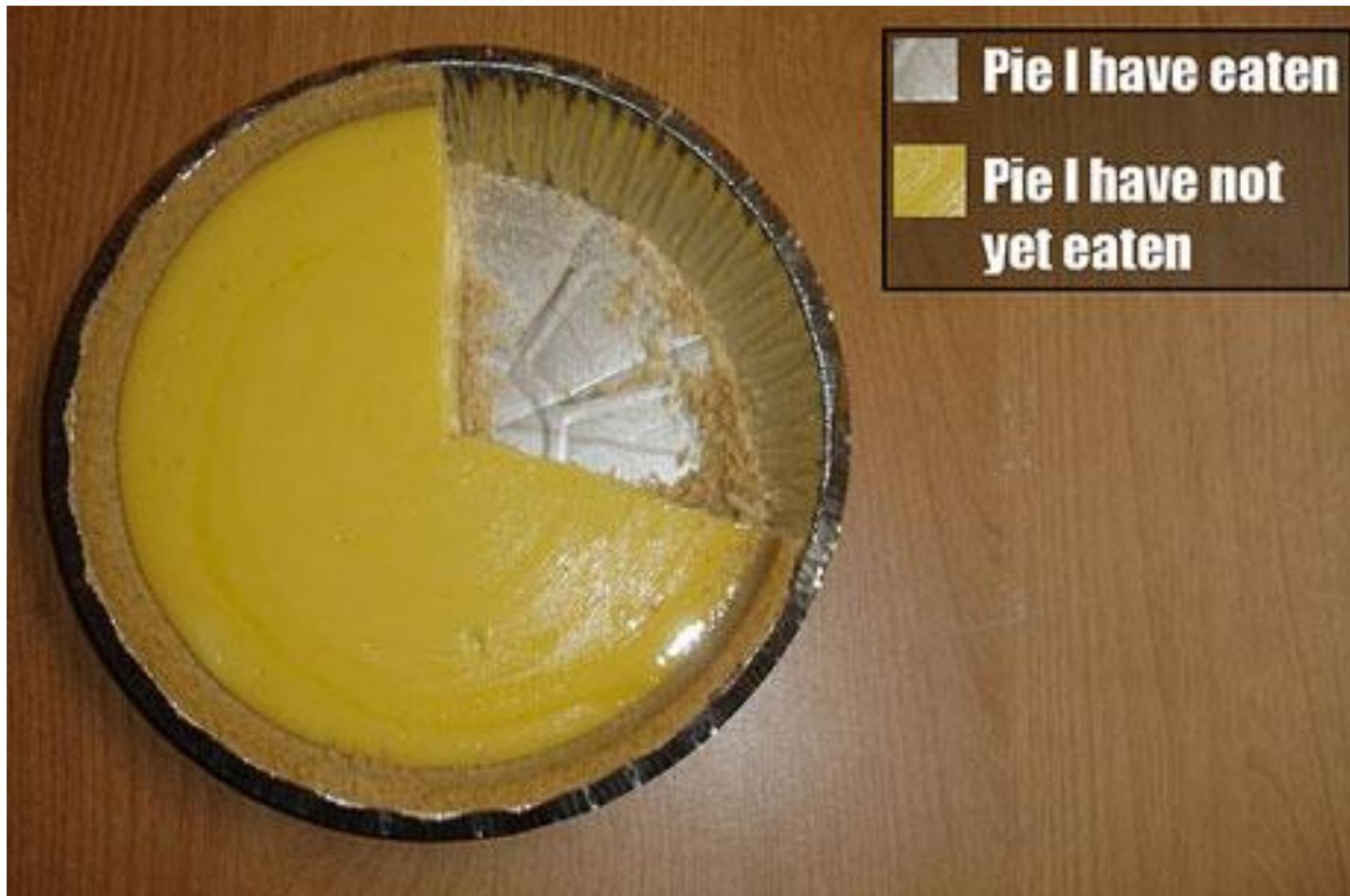
Pie chart



Pie chart

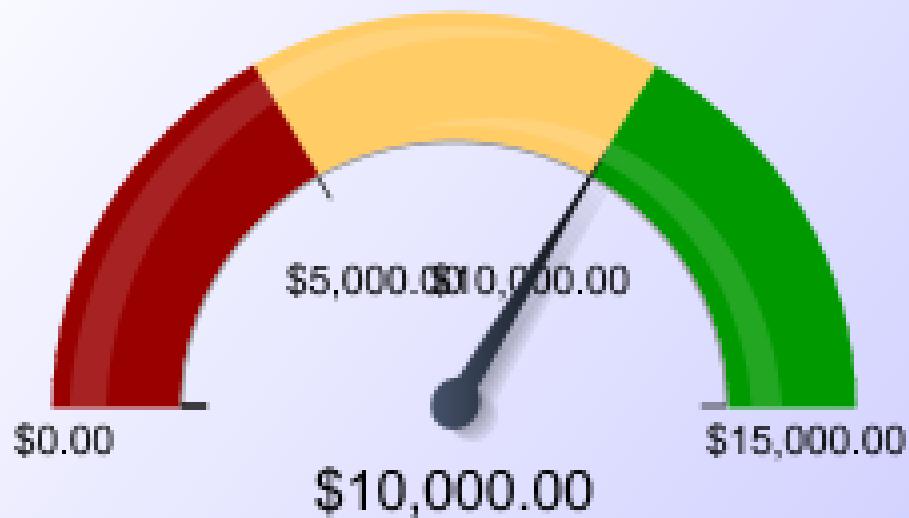


Pie chart



Other risky charts

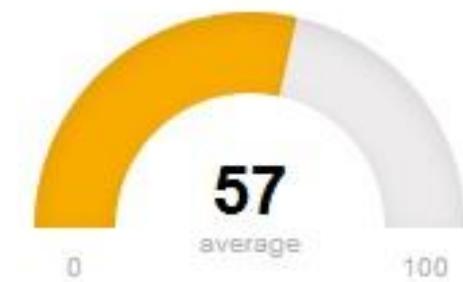
Closed Sales To Date



Visitors



Errors



Timers



Alerts



Events



Bubble Chart

encodes data using size of circle to show comparisons which is difficult for making precise quantitative comparisons.



* Caution this chart type is not recommended.

Concentric Circles

encodes data using arc and area to show comparisons but problematic for many reasons.



* Caution this chart type is not recommended.

Donut Chart

encodes data using arc and area to show a part-to-whole comparison but problematic for many reasons.



* Caution this chart type is not recommended.

Word Cloud

encodes data using size of word to show comparisons which is difficult for making precise quantitative comparisons.

Least Many
Some Most
More Few

* Caution this chart type is not recommended.



How to choose a chart?

How to choose a chart?

- Chart selection tools and visuals
 - Provide a structured way to narrow down the list of charts
 - Many exists
- Visual vocabularies
 - Typically the large media organizations have a standardised way of using charts, including guides on when to use which type
 - Some of them are publicly available

Chart selection

 perceptual edge

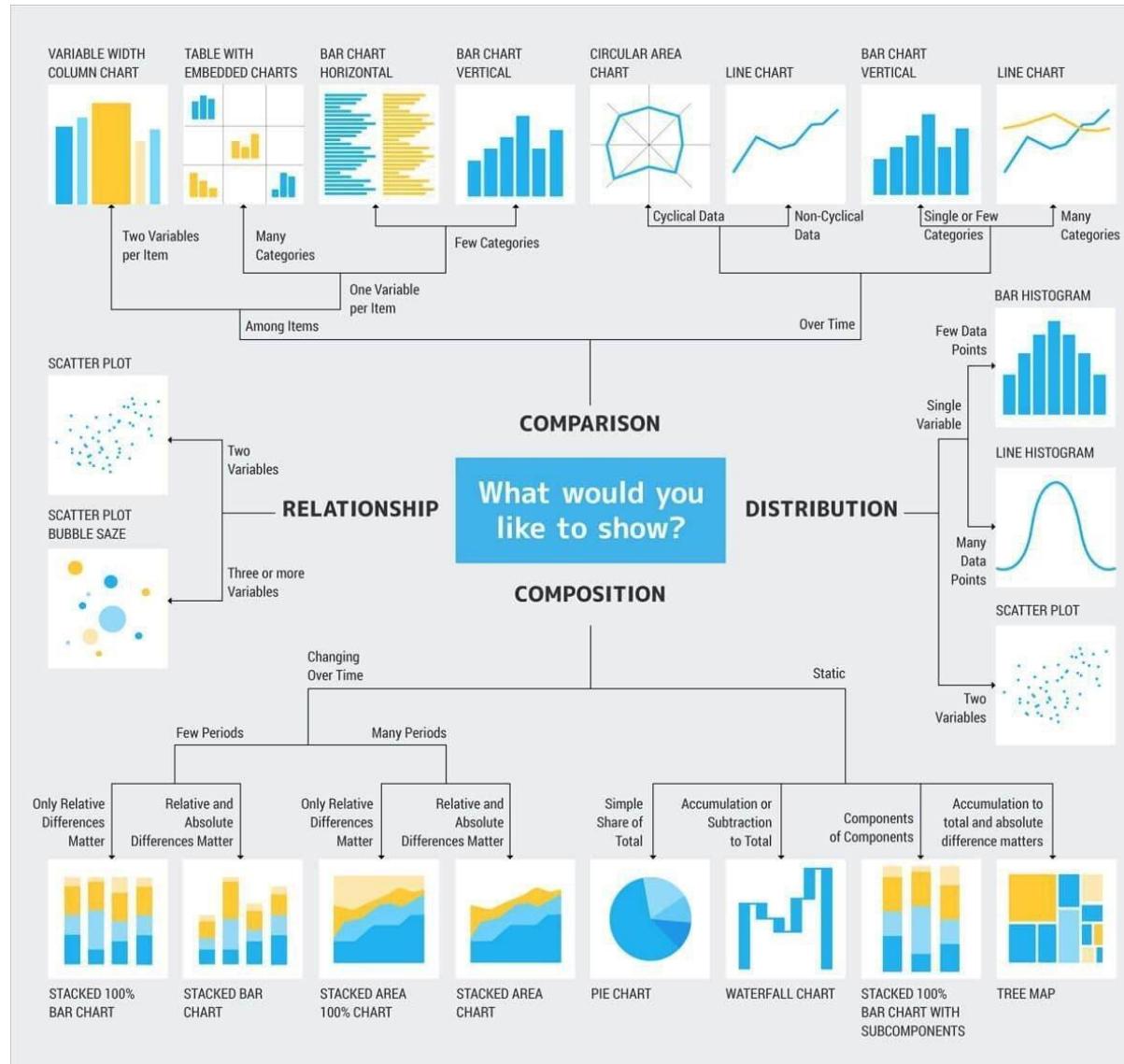
Graph Selection Matrix

Graph	Relationship							
	Time Series	Ranking	Part-to-Whole	Deviation	Distribution	Correlation	Nominal Comparison	
Bar Graph (vertical)	Values display how something changed through time (yearly, monthly, etc.)	Values are ordered by size (descending or ascending)	Values represent parts (ratios) of a whole (for example, regional portions of total sales)	The difference between two sets of values (for example, the variance between actual and budgeted expenses)	Counts of values per interval along a quantitative scale from lowest to highest (for example, counts of people in an organization by age intervals of 10 years each)	Comparison of two paired sets of values (for example, the heights and weights of several people) to determine if there is a relationship between them	A simple comparison of values for a set of unordered items (for example, products or regions)	
Bar Graph (horizontal)								
Line Graph	Values display how something changed through time (yearly, monthly, etc.)	Values are ordered by size (descending or ascending)	Values represent parts (ratios) of a whole (for example, regional portions of total sales)	The difference between two sets of values (for example, the variance between actual and budgeted expenses)	Counts of values per interval along a quantitative scale from lowest to highest (for example, counts of people in an organization by age intervals of 10 years each)	Comparison of two paired sets of values (for example, the heights and weights of several people) to determine if there is a relationship between them	A simple comparison of values for a set of unordered items (for example, products or regions)	
Dot Plot (vertical)	Yes (to feature individual values and support their comparisons; quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)			
Dot Plot (horizontal)	Yes (to feature overall trends and patterns and support their comparisons)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)			
Strip Plot (single)	Yes (when you do not have a value for every interval of time)	Yes		Yes (only when also featuring a time series or single distribution)	Yes (to feature the overall shape of the distribution)			
Strip Plot (multiple)	Yes (when you do not have a value for every interval of time)	Yes						
Scatter Plot	Yes (only when also featuring distributions)							
Box Plot (vertical)	Yes (only when also featuring distributions)				Yes (when comparing multiple distributions)			
Box Plot (horizontal)	Yes (only when also featuring distributions)				Yes (when comparing multiple distributions)			

www.PerceptualEdge.com (Derived from the book *Show Me the Numbers*, Stephen Few, Analytics Press, 2004) Copyright © Stephen Few 2009

Source: Stephen Few

Chart selection



Source: Dr. Abela

Chart selection

WHAT WOULD YOU LIKE TO SHOW?

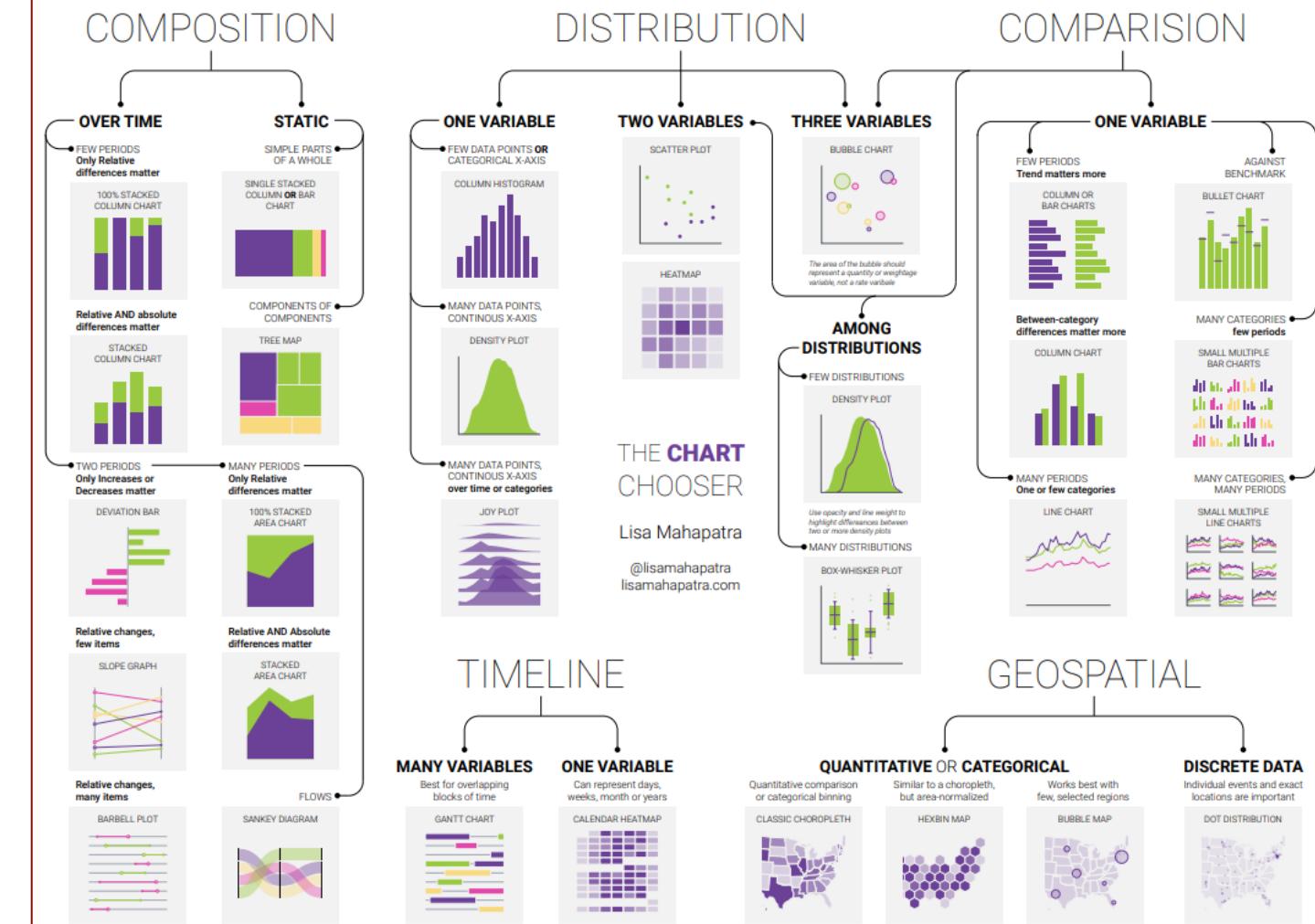
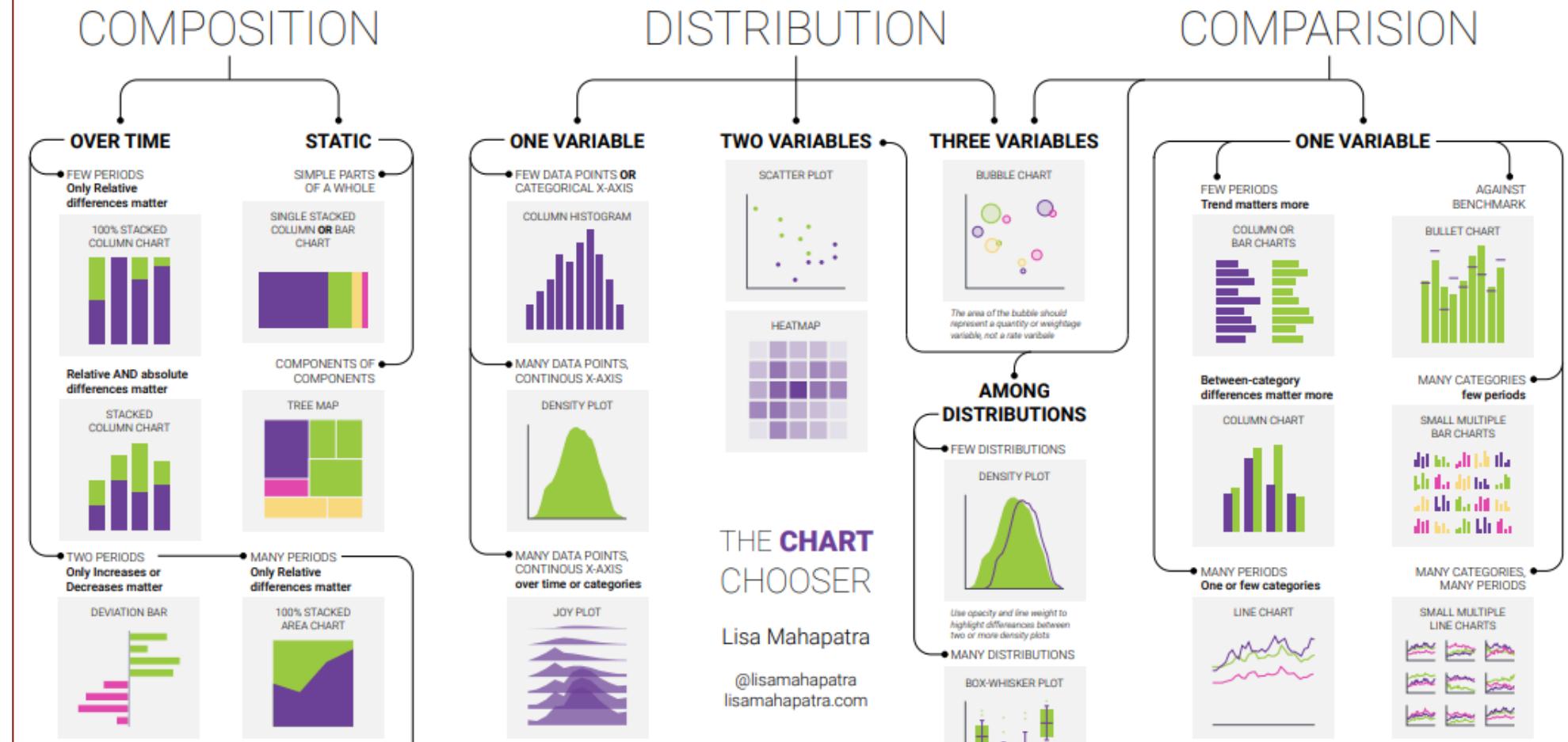


Chart selection

WHAT WOULD YOU LIKE TO SHOW?



Financial Times Visual Vocabulary

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 relationships are 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: Alex Smith; Chris Campbell; Ben Birrell; Li Li; Emanuele Gabriele; Francesco D'Urso; Denebola Gherardi; Paul McCullagh; Henrik Dideriksen; Ingvild Heggelund; Corinne van der Schueren; and Benoit Bozzo

ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Be mindful that the reference point can be a mean, a target or a long-term average. Can also be used to highlight outliers that may be causal (ie. one drives the other).

Example FT uses: Trade surplus/b deficit; climate change

Diverging stacked bar: Perfect for presenting survey results which are split into contrasting divergent/shifted values.

Spine: Splits a single value into two contrasting components (eg. market share).

Surplus/deficit/filled line: The shaded area of these charts allows a comparison between either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that the relationship you show may not be causal (ie. one drives the other).

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

Column + line timeline: A good way of showing the evolution between an amount (column) and a rate (line).

Bubble: Like a scatterplot, but adds additional detail to each point according to a third variable.

XY heatmap: A good way of showing the patterns between 2 categories of data, less effective for showing the differences in amounts.

Ranking

Use when an item's position in an ordered list is more diagnostic than its absolute value. Good for highlighting the lack of uniformity or inequality in the data.

Example FT uses: Wealth, dimension, league tables, constituency election results

Ordered + columns: See above.

Ordered proportional symbol: Used when there are big differences between values and seeing the relationship between data is not so important.

Dot strip plot: Good for showing individual values in a distribution, can be used to show multiple data points have the same value.

Barcode plot: Like a dot strip plot, all data points are shown, they work best when highlighting individual categories.

Slope: Perfect for showing how rates have changed over time or between different categories.

Lollipop: Lollipops draw more attention to the data points than standard bubbles and can be used to show rates and value effects.

Bump: Effective for showing changing numbers across multiple dates, months, years, etc. Consider grouping lines by date.

Distribution

Show values in a dataset and how often they occur. The shape or 'skew' of a distribution can be used to highlight the lack of uniformity or inequality in the data.

Example FT uses: Income distribution, population, geographic location, revealing inequality

Dot plot: A simple way of showing the change or the spread of data across multiple categories.

Dot strip plot: Good for showing individual values in a distribution, can be used to show multiple data points have the same value.

Barcode plot: Like a dot strip plot, all data points are shown, they work best when highlighting individual categories.

Slope: Good for showing changing data as long as it is not too noisy. It can be simplified into a 2 or 3 point slope.

Violin plot: Summarises multiple distributions by showing the median, count and range of the data.

Population pyramid: A standard way for showing the age and sex breakdown of a population. Buttons, effectively, back to back.

Cumulative curve: A good way of showing how something progresses if it exists. It shows cumulative probability and always has a measure.

Frequency polygons: For displaying multiple distributions of data. Like a regular line chart, limited to a maximum of 3 or 4 streams.

Berserker: Used to emphasise individual points. It can be used to emphasise outliers with medium-sized bubbles.

Change over Time

Gives emphasis to changing trends. These can be short (one-day) movements, seasonal trends, training decades or centuries. Choosing the right chart is important to provide suitable context for the data.

Example FT uses: Share price movements, economic time series, artistic changes in art market

Line: The standard way to show a continuous distribution - keep the gaps small, otherwise it will be difficult to highlight the data points.

Column + line: The standard way to compare the things. Must always start at 0 for the axis.

Calories + line: A good way of showing the relationship over time between a calorie count (column) and a rate (line).

Slope: See above.

Area chart: Use with care - these are good at drawing changes in total, but sometimes the components can be very difficult.

Cardioid: Usually focused on day-to-day activity, showing the opening/closing and the point of each day.

Fan chart (projections): A great way to show the uncertainty of future projections - usually drawn with a fan that tapers forward to projection.

Calender heatmap: A great way of showing seasonal patterns (daily, weekly, monthly) and the level of precision in the data.

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

Criss timeline: Good for showing discrete values of varying size across a timeline (eg. earthquakes by intensity).

Vertical timeline: Great for showing the Y axis. Good for showing detailed time series that work well with mobile scrolling on mobile.

Seismograph: Another alternative to the fan chart when there are lots of data or highlighting individual elements is useful.

Streamgraph: A type of area chart: use when seeing the connections over time is more important than the individual values.

Magnitude

Show how entities compare. These can be relative (just being able to see larger/smaller) or absolute (able to see the difference). Usually these show a volume, weight, size, number of items, dollars or people rather than dimensions of space or time.

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

Stacked column/bar: A simple way of showing part-to-whole relationships that can be applied to anything with more than a few components.

Bar: See above. Good when the data is not line and the categories are long category names.

Paired column: As per standard column + line but allows for multiple series. Can sometimes be better than paired with more than 2 series.

Paired bar: See above.

Morimoku: A good way of showing the size and proportion of data at the same time - as long as the data is not too complicated.

Proportional symbol: Use for totals rather than data - be very careful as the data in dots will be hard to read.

Pie: A common way of showing parts-of-a-whole data - just be aware that the slices are proportional to the size of the segments.

Doughnut: Similar to a pie chart - but the centre can be a blank space to include more information. Use radial colour schemes for showing TV+ values.

Treemap: Use for hierarchical pan-to-whole relationships. Can be useful when there are many small regions.

Venn: A way of showing points in areas - any point that is closer to the central area is more often correct.

Arc: A function, often used for visualising parliamentary constituencies by number of seats.

Gridded: Good for showing % information, they work well with whole numbers and percentages and multiple leisure forms.

Radar: A great way of showing values of multiple variables, but need to make sure they are organised in a way that makes sense to readers.

Parallel coordinates: An alternative to radar charts - again, the variables are important and need to highlight the negative.

Waterfall: Can be useful for showing part-to-whole relationships where some of the components are negative.

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

Ground symbol: An alternative to bubble charts when there are lots of data or highlighting individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the entity is large enough, it is good to see the size of the components. Consider a magnitude-based chart instead.

Example FT uses: Company structures, national election results

Spatial

Avoid for location maps only when precise locations or geographical proximity are important to the reader than anything else.

Example FT uses: Location density, name recognition, location diversity mismatch, variation in election results

Flow

Show the volumes or intensity of movement between two or more states in a system, such as time periods or sequences or geographical locations.

Example FT uses: Movement of funds, trade, migrants, loans, information, networking graphs.

Waterfall: Designed to show the sequencing of data from a single process, typically financial.

Chord: A complex bar chart which can illustrate the connection between nodes in a network.

Network: Used for showing the strength and interconnectedness of relationships of varying types.

ft.com/vocabulary

FT

© Financial Times

Financial Times Visual Vocabulary

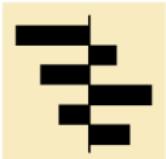
Deviation

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

Example FT uses

Trade surplus/deficit, climate change

Diverging bar



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

Diverging stacked bar



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

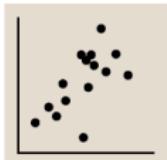
Correlation

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

Example FT uses

Inflation and unemployment, income and life expectancy

Scatterplot



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

Column + line timeline



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

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

Ordered column



See above.

FT Visual Vocabulary – Power BI Edition

Visual Vocabulary

There are so many ways to visualise data - how do we know which one to pick? Click on a category below to decide which data relationship is most important in your story, then look at the different types of charts 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.

Click any section below to view the charts

Deviation

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

Correlation

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

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.

Distribution

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

Change over Time

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

Part-to-Whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Magnitude

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

Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Flow

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

CREATED BY

Jason Thomas | [@501ason](#) | [blaq9](#)

INSPIRED BY

Andy Kriebel | [@drkriebel](#) (including the design / theme template from: [Blog](#))
FT Graphics: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg; Paul McCallum; Martin Stabe
Visual Vocabulary Poster: [ft.com/vocabulary](#)

CREDITS

Power BI Community & Tableau Community - for sharing their data/techniques and learnings

AND IN PARTICULAR

Konstantinos Ioannou | [@koenouloannou](#) - for opening up my mind regarding the potential of R/Python visuals
David Eldersveld | [@datavizd](#) - for being my sounding board
Nujharree | [@nujharree](#) - for creating Violin Plots in R and kickstarting my R visualizations journey
Pragmatic Works | [@pragmaticworks](#) - for blogging about all those custom visualizations in Power BI

CUSTOM VISUALS:

MapBox	Chartulator	Scatter Chart by Akelon	Dot Plot by MAQ
Python	Infographic Designer	Box & Whisker by MAQ	Dumbbell Chart by MAQ
Candlestick by OKViz	Synoptic Panel by OKViz	Mekko Chart	Sunburst
Venn Diagram by MAQ	Waffle Chart	Chord Slicer	Radar Chart
Bullet Chart by OKViz	Flow Map	Heat Map	Heat Map
Sankey	Chord	Network Navigator	Social Network Graph

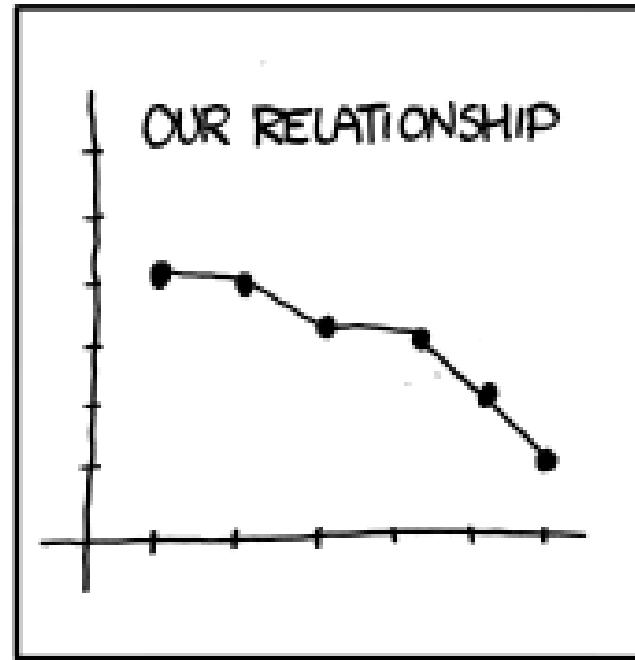
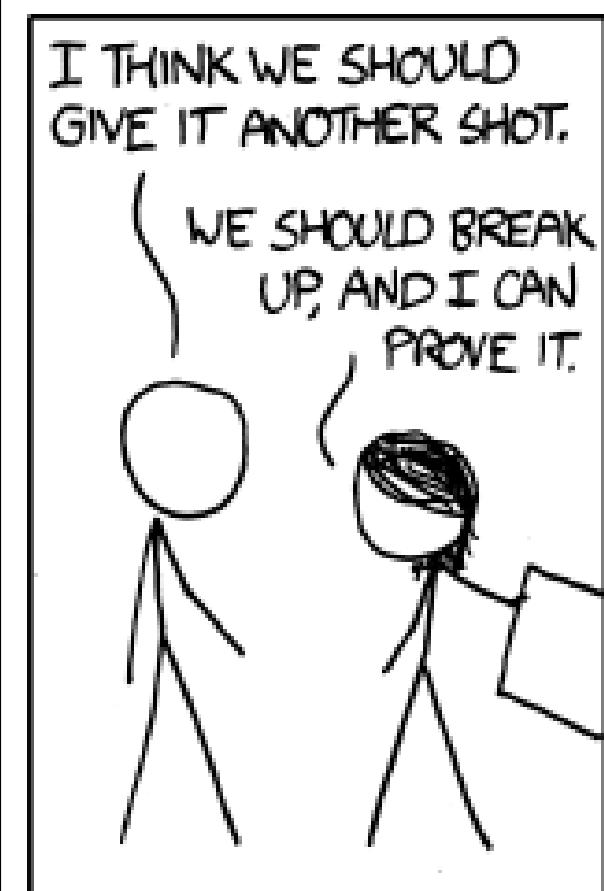
Microsoft Power BI

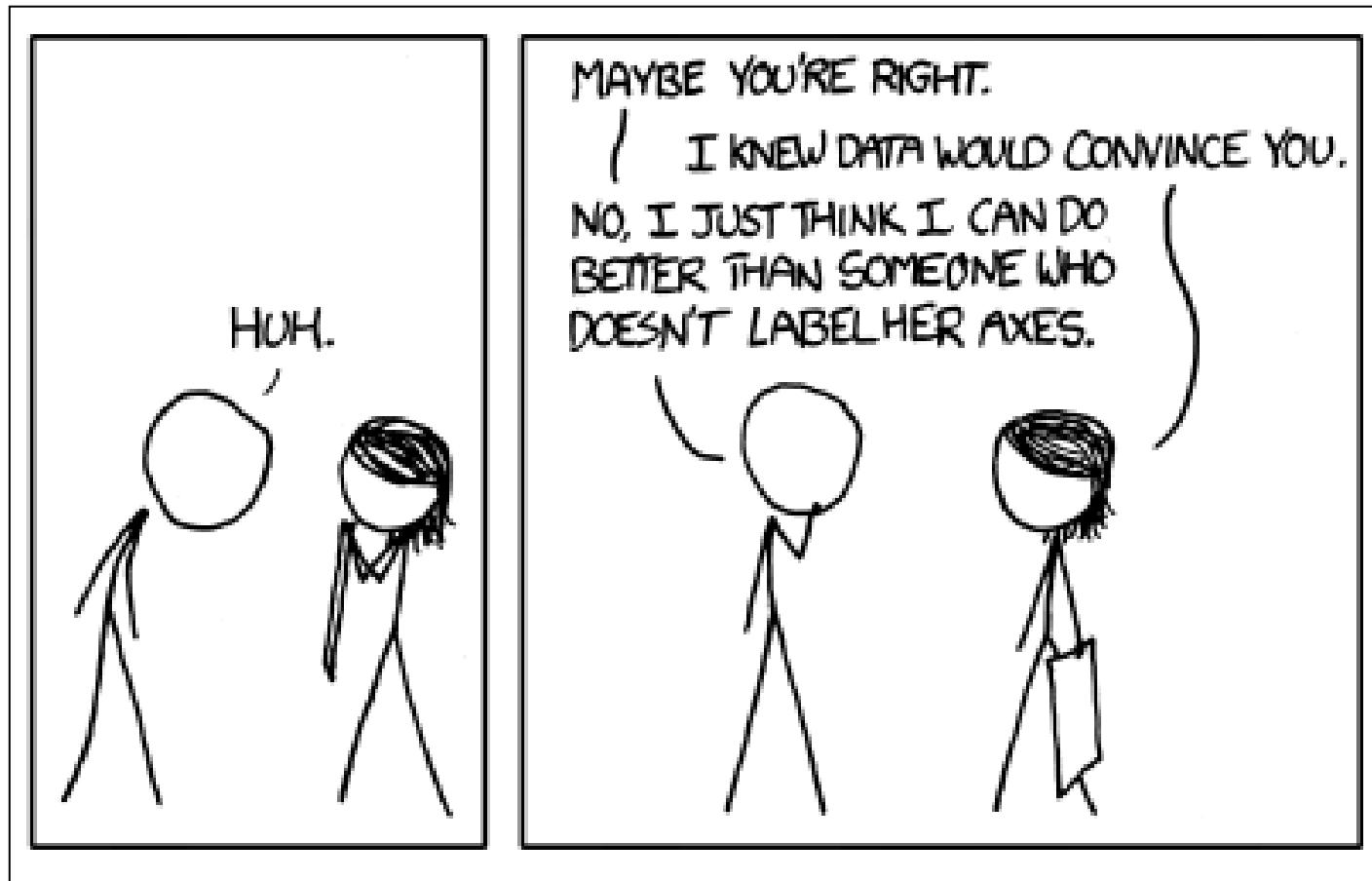
1 of 10

in tw fb li

<https://community.powerbi.com/t5/Data-Stories-Gallery/FT-Visual-Vocabulary-Power-BI-Edition/td-p/584460>

Extras





Thank You

