


















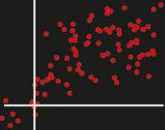


MATCHING VISUALIZATIONS TO DATA

MATCHING VISUALIZATIONS TO DATA

With data displays, we try to highlight:

1. a **relationship** – show a connection or correlation between two or more variables, such as the impact of an aging population on health care;
2. a **comparison** – set some variables apart from others, and display how those two variables interact, such as the number of fans attending hockey games for different teams in a season;
3. a **composition** – collect different types of information that make up a whole and display them together, such as the various search terms that visitors used to land on your site, or how many visitors came from various sources (links, search engines, or direct traffic), and
4. a **distribution** – lay out a collection of related or unrelated information to see how it correlates (if at all), and to understand if there's any interaction between the variables, such as the number of bugs reported during each month after a new software release.

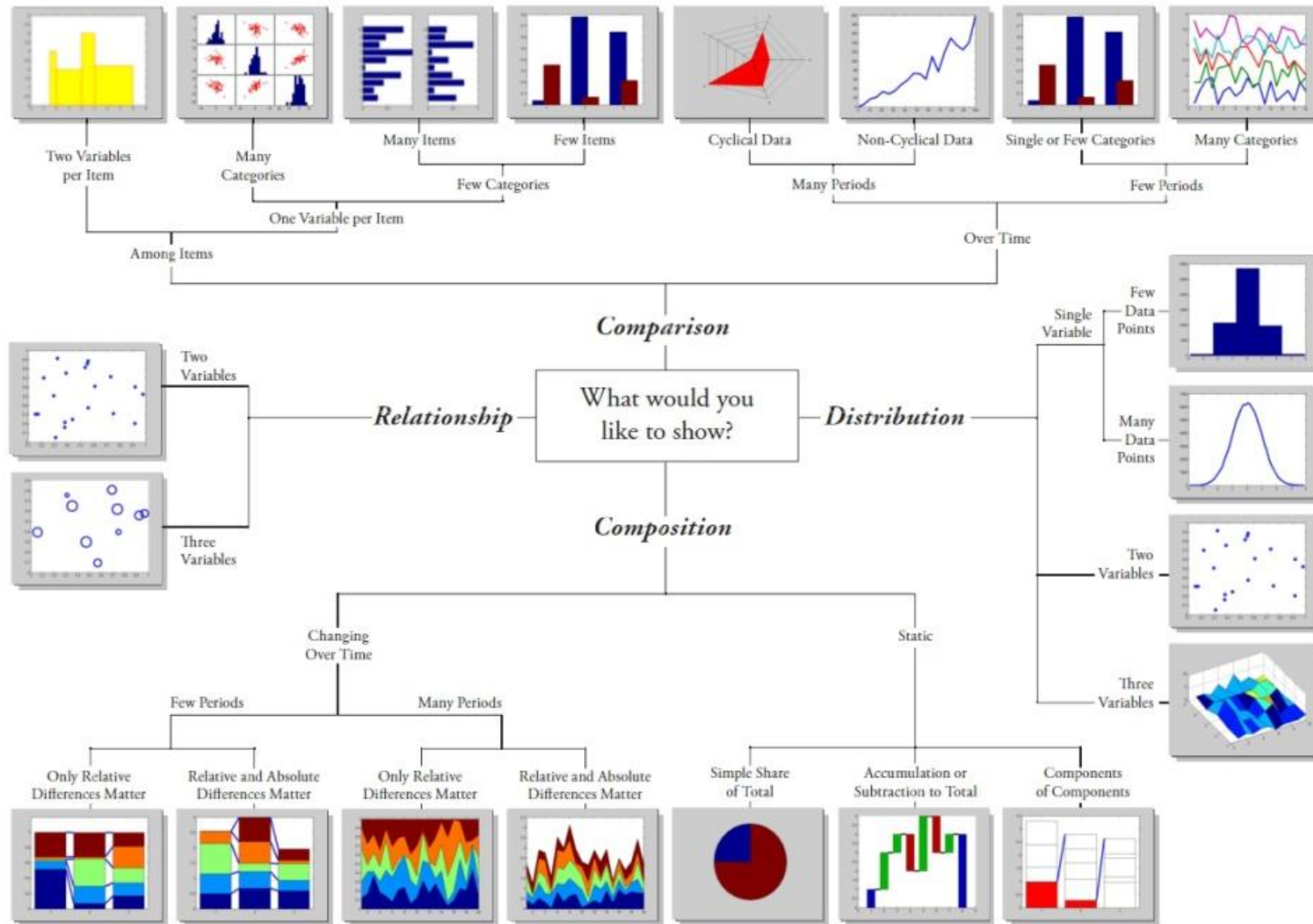
WHICH METHOD SHOULD YOU USE?

	who/which is involved?	where is it?	when did it happened?	what is it about?	how/why does it work?	how much is it?	
	 PROFILE	 LOCATION	 FAMILY TREE	 ORGANOGRAM	 NETWORK DIAGRAM	 VALUE	who
		 POSITION	 TRACK	 PLACES	 CONNECTION	 CHOROPLETH	where
			 TIMELINE	 PERIOD	 EVOLUTION	 CHARTS	when
				 EXPLODED VIEW	 COMIC STRIP	 COMPARISATION	what
					 PROCESS	 RELATIONS	how/why
						 DIAGRAMS	how much

Data visualizations are not just about picking random visualization methods.

The result varies depending on the structure of the data and the (combinations of) questions.

Chart Suggestions—A Thought-Starter



A CLASSIFICATION OF CHART TYPES



Data comparison charts

Data reduction charts

Comparison

Composition

Distribution

Evolution

Relationship

Profiling

Bars



Pie



Histogram



Line



Scatterplot



Grouped bars



Dot plot



Bullet



Pareto



ID Scatterplot



Horizon



Connected Scatterplot



Cycle plot



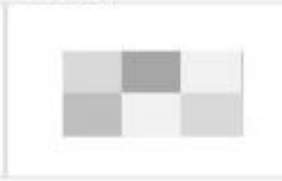
Scatterplot matrix



ID Scatterplot



Heat map



Multidimensional Pie



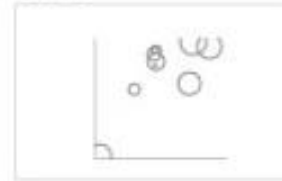
Boxplot



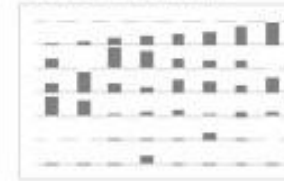
Step



Bubble



Reorderable matrix



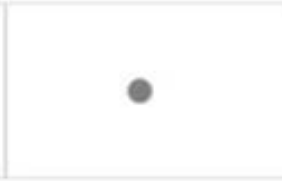
Horizon



Slope



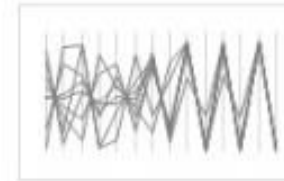
Alert



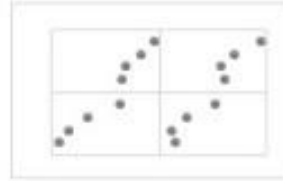
Connected Scatterplot



Parallel Plot



Trellis



VISUALIZATION CATALOGUE

WORKHORSE DATA VISUALIZATIONS

Line Chart/Rug Chart/Number Line (data exploration)

Histogram (data exploration)

Boxplots (data exploration)

Line Graph (data presentation + data exploration)

Bar Chart (data presentation + data exploration)

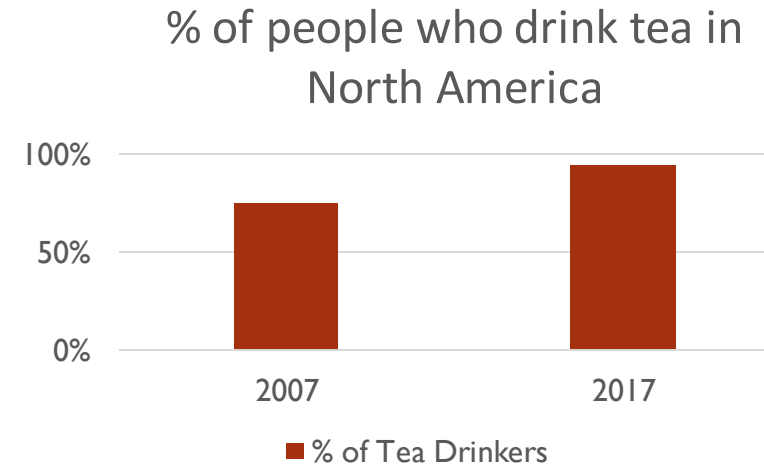
Scatterplot (data presentation + data exploration)

SIMPLE TEXT

One or two numbers to focus on.

Good at “setting the scene”.

Draws focus to an area of the report.



95% of the population
drinks tea today compared to
75% in 2007

TABLE

Tables interact with our **verbal** system, which means we **read** them:

- used to **compare** values
- audiences will look for **their** rows

Table design needs to blend into background

- the data should stand out, not the borders
- dense table/data: use **alternating** row colour

Name	Last Year	This Year
Bob	20	30
Fred	30	40
George	10	15

Name	Last Year	This Year
Bob	20	30
Fred	30	40
George	10	15

TABLE HEATMAP

	Last Year	This Year	Next Year	Optimum
George	20	20	20	20
Peter	40	35	30	25
John	10	10	5	5
Sandra	25	30	35	40

Leverage colour to convey magnitude

- use **single colour saturation** rather than differentiation (different colours)
- with a legend (white = low, blue = high), numbers can be removed without altering the message

	Last Year	This Year	Next Year	Optimum
George	20	20	20	20
Peter	40	35	30	25
John	10	10	5	5
Sandra	25	30	35	40

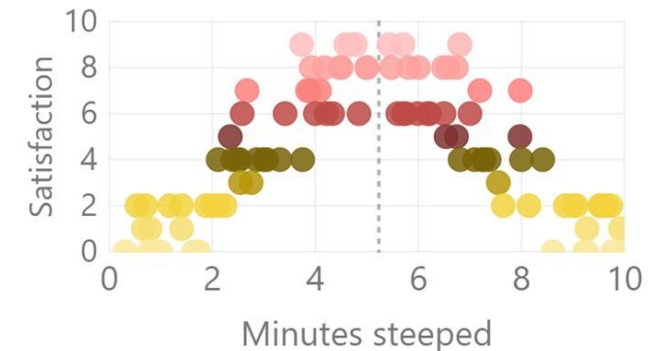
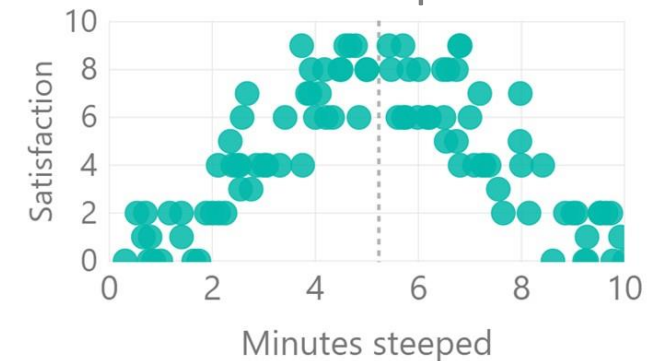
	Last Year	This Year	Next Year	Optimum
George				
Peter				
John				
Sandra				

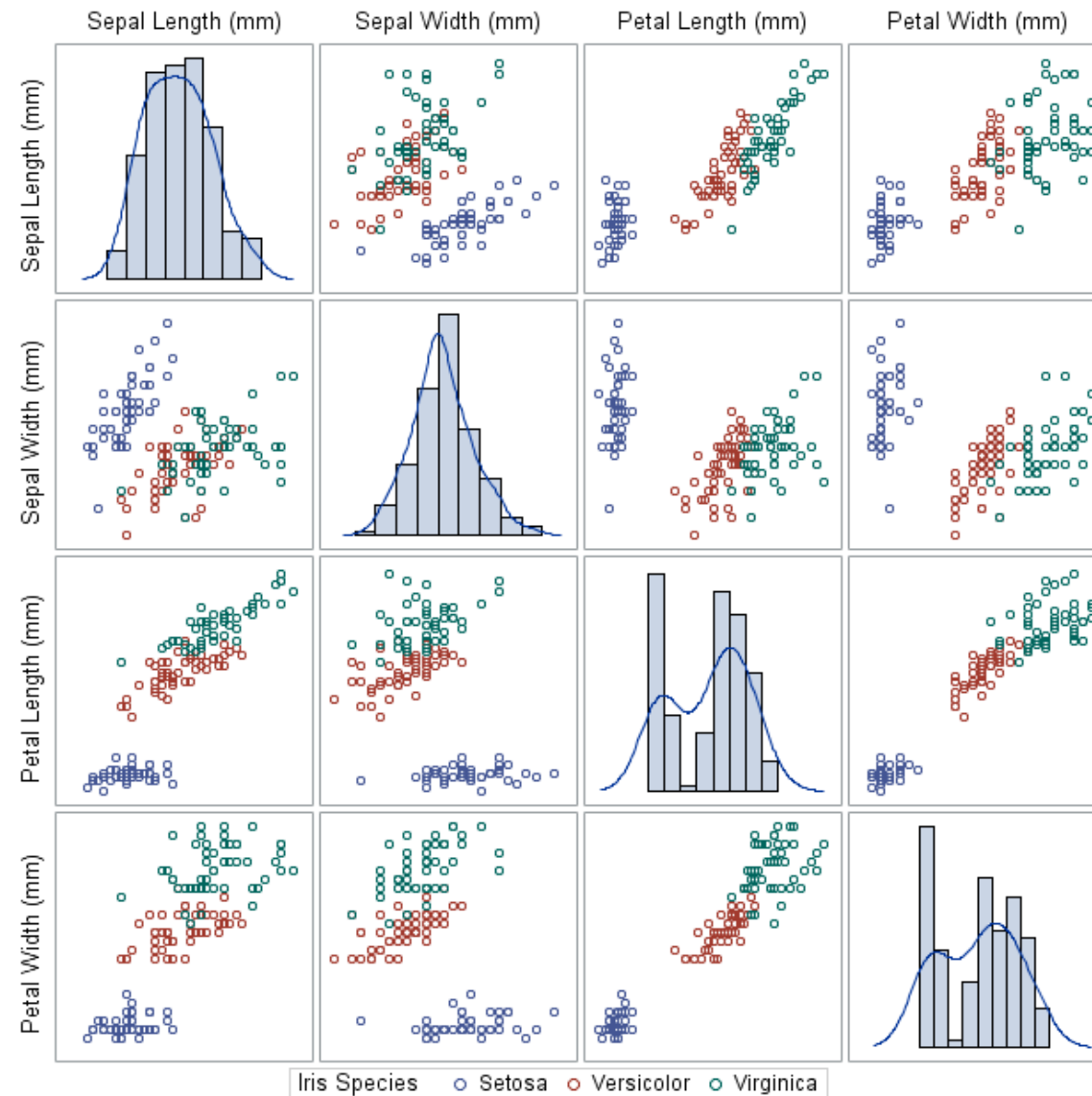
SCATTERPLOT

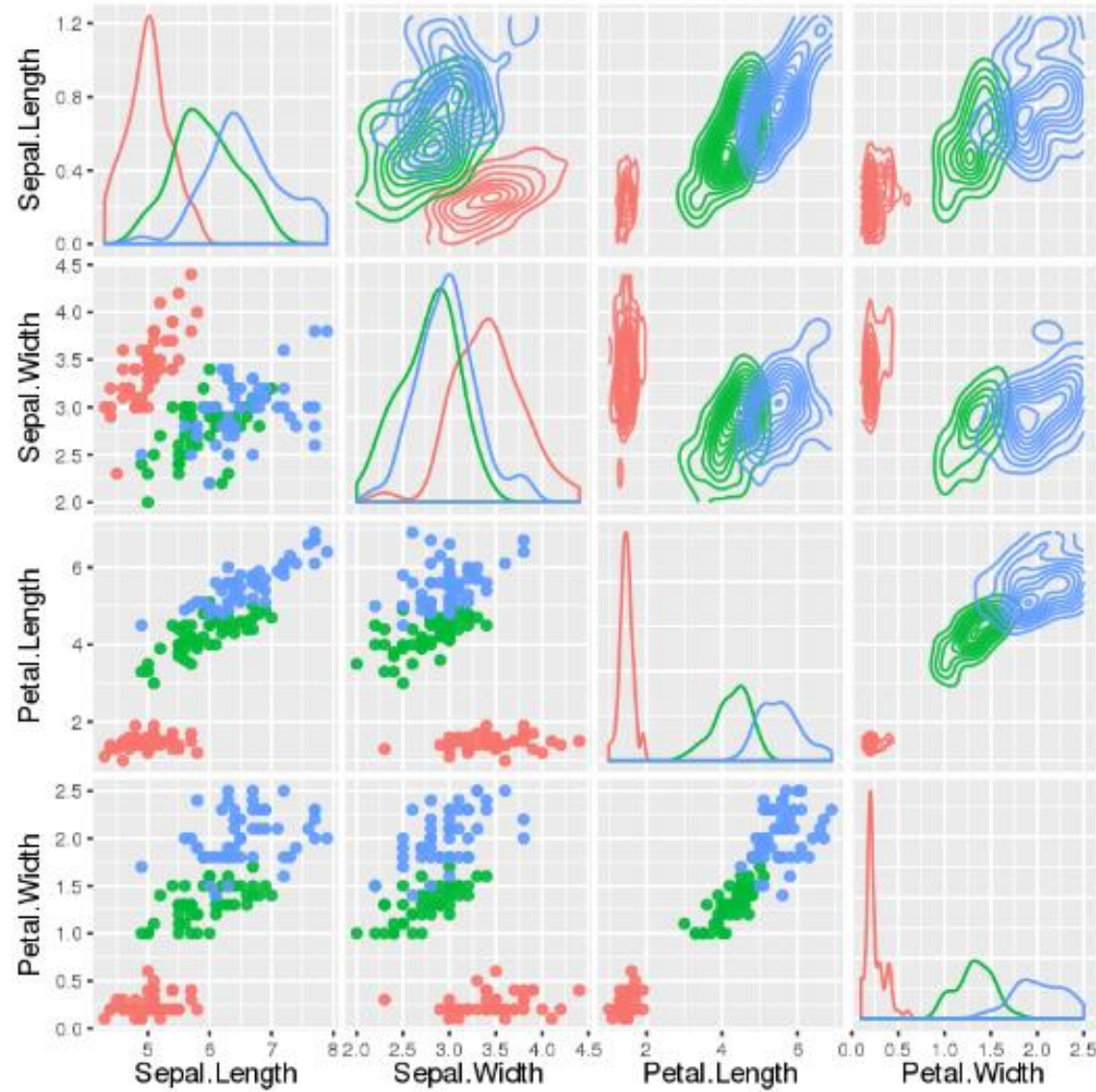
Show relationship between 2 variables (**scatterplot**) or 3 variables (**bubble plot**)

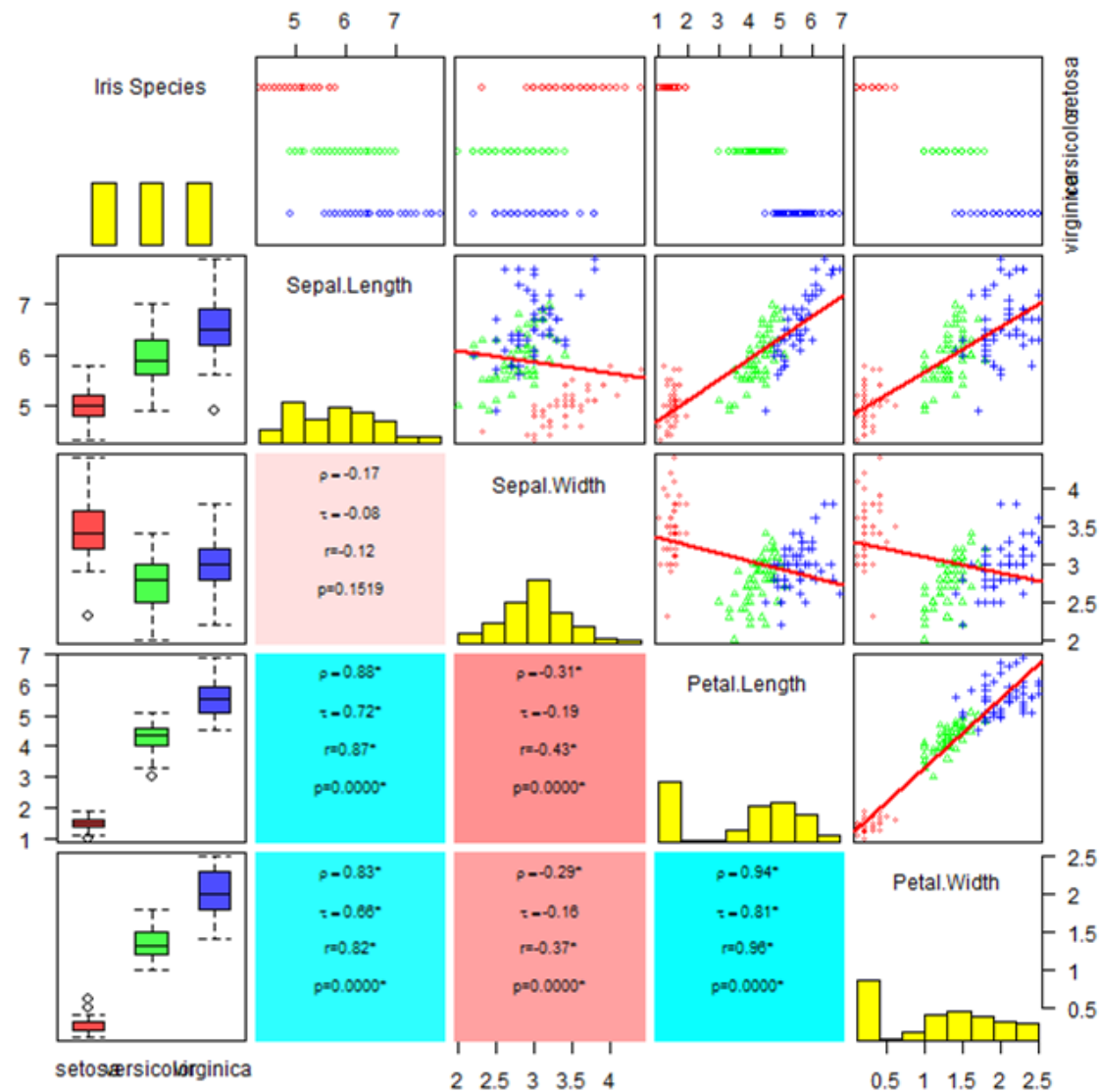
- use average lines (dotted lines) to provide context
- far fewer options in Power BI than Excel
- consider using groupings to add clarity (e.g. **colour gradients**)

How long should the perfect cup of tea be steeped?









Is this starting to get too cluttered?

LINE CHART

Line chart can show a single series or multiple series of data.

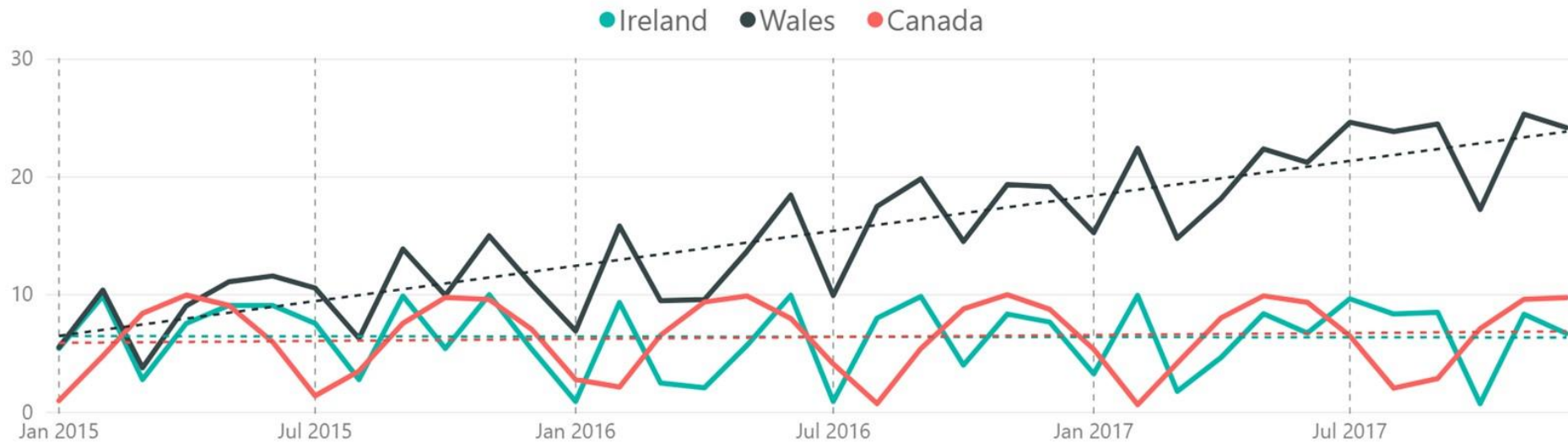
- particularly useful to show time series

Axis scale should be clear and relevant.

May wish to “anchor” y –axis if using dynamic filters

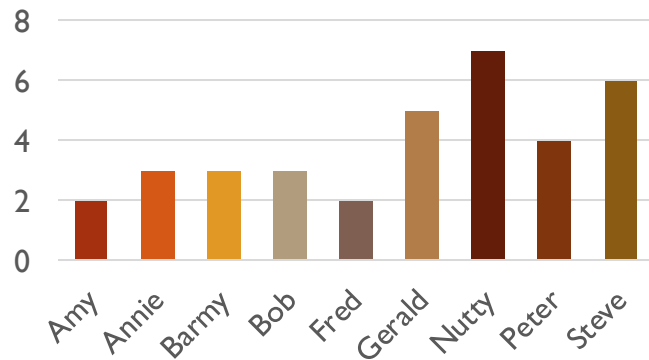
- otherwise the graph can jump around as people interact with it

LINE CHART



Comparison of Countries – cups of tea drunk per week per person

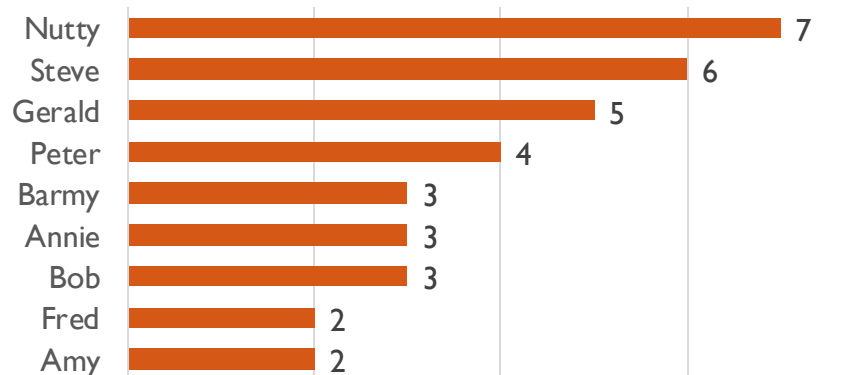
BAR CHART (VERTICAL & HORIZONTAL)



Very versatile and useful.

ALWAYS (?) have a zero baseline.

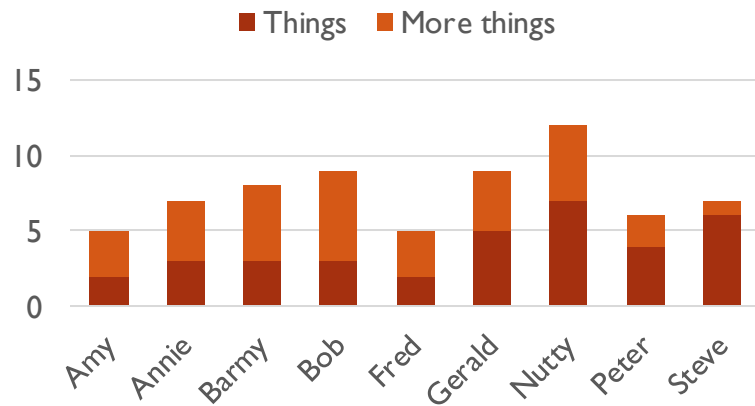
Use graph axis OR data labels. Axis for broad statements, data labels for more detail.



Horizontal charts are apparently easier to read (according to many studies).

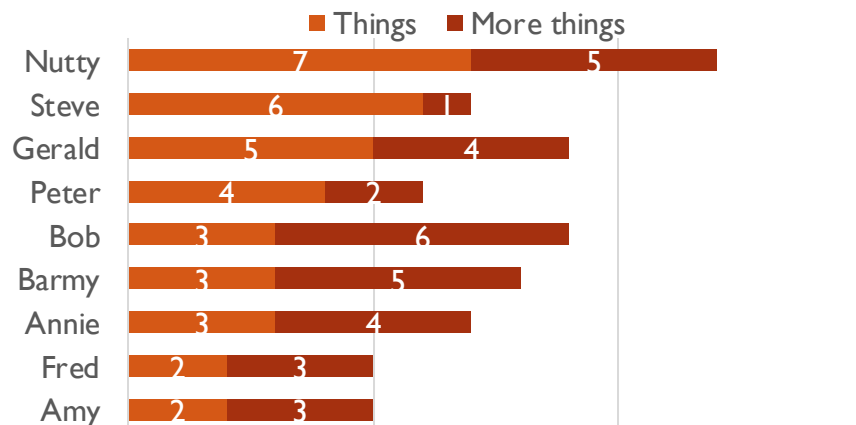
Think about the ordering of categories.

STACKED BAR CHART (VERTICAL & HORIZONTAL)



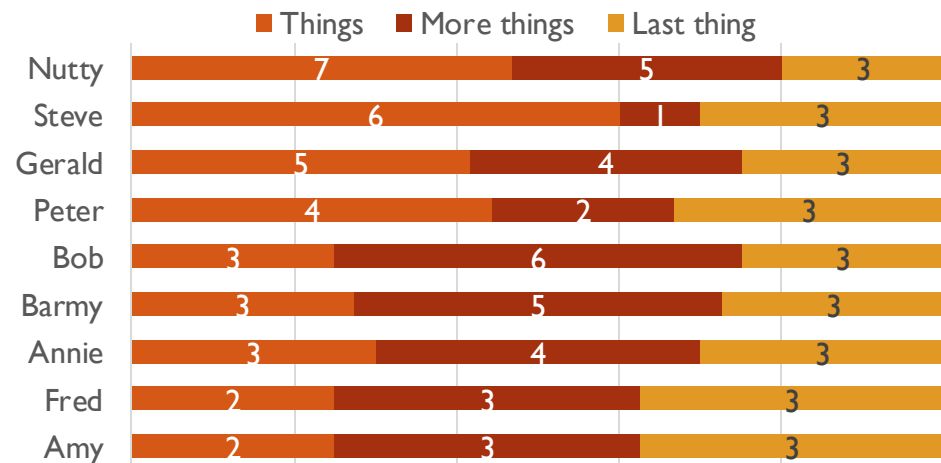
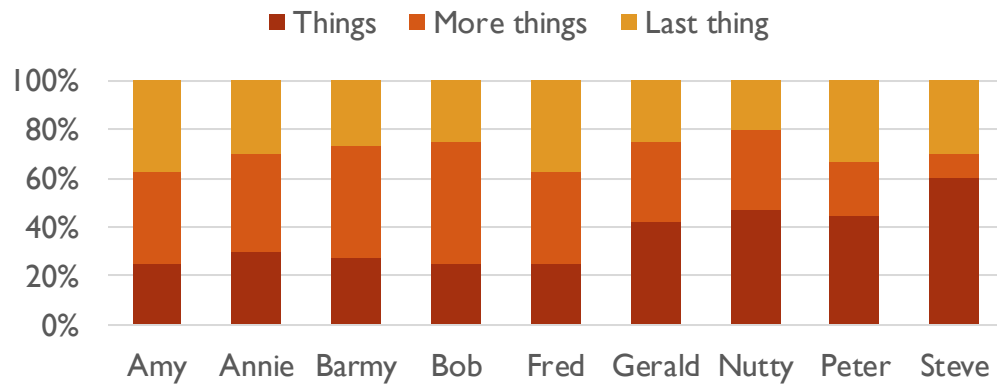
Designed for **comparing totals**, but can quickly become **overwhelming**.

Hard to sort / order.



Filtering is complicated in Power BI (what do you click on & how the chart responds when filter is clicked on?)

100% BAR CHART (VERTICAL & HORIZONTAL)



Work well for visualizing portions of a whole on scale from negative to positive

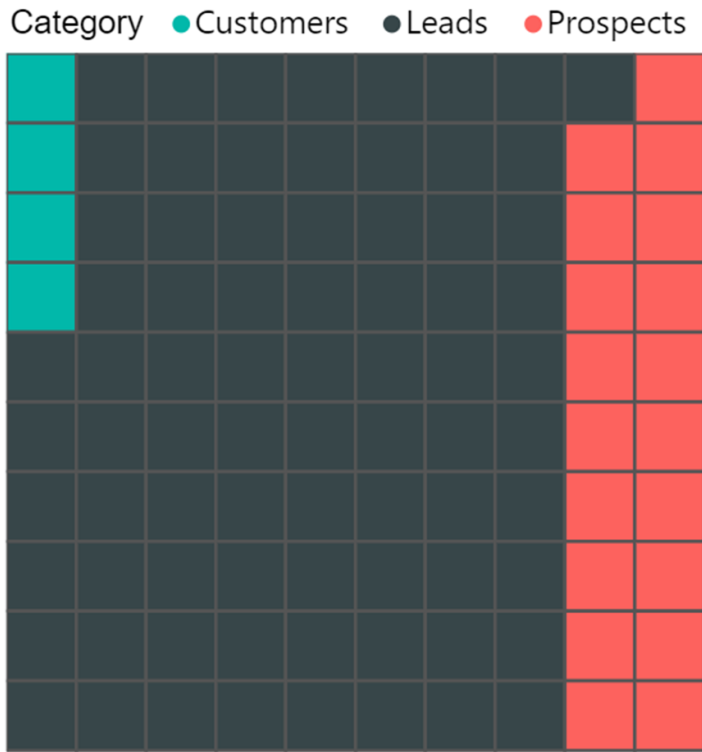
Consistent baseline on far left and right

Easy to compare

Issue is no relative measure to magnitude of data

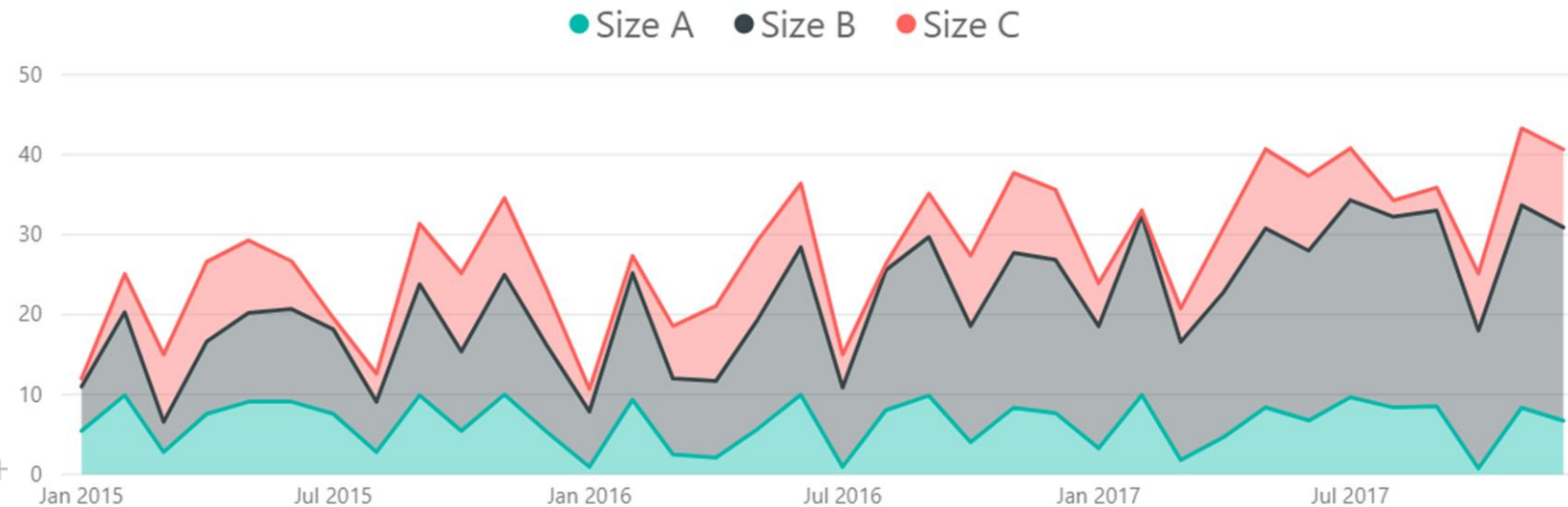
Research shows that horizontal is easier to process than vertical

AREA CHART

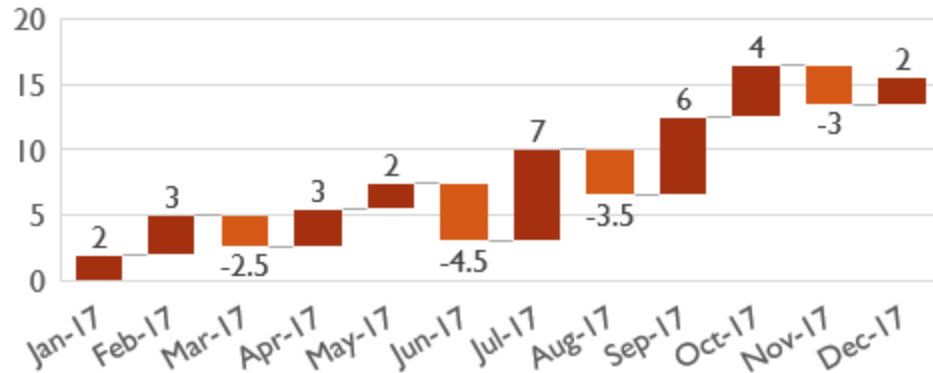


Try to avoid: human brains have a hard time attributing a value to a 2D area...

... except for numbers with **vastly different** magnitudes.



WATERFALL



Number of Units Sold



Shows how initial value increases or decreases using a series of intermediate values.

Different colours can be used for increases and decreases.

Hard to remove elements without removing context (hard to **declutter** the chart).

Large increases / decreases look odd...

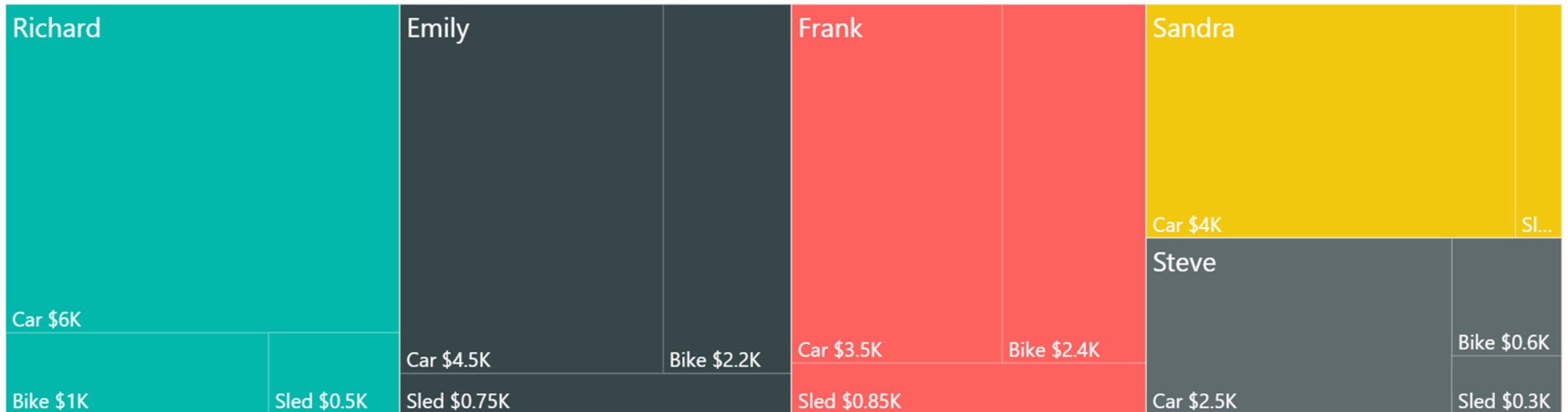
TREEMAP

Simultaneously show big picture and can compare related easily.

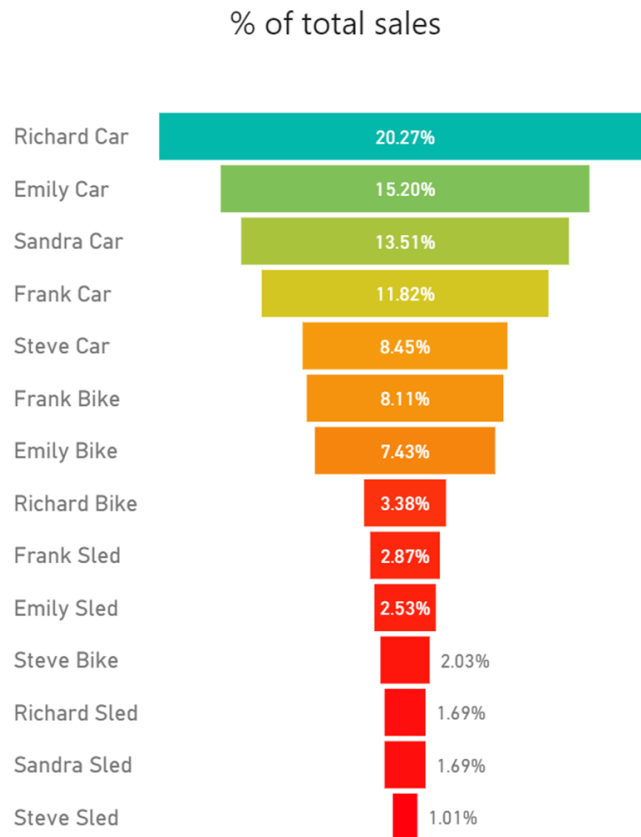
Easy to process data sub-categories.

Useful to prioritize “big ticket items” in dynamic dashboards.

Labeling and colouring are tricky.



FUNNEL CHART



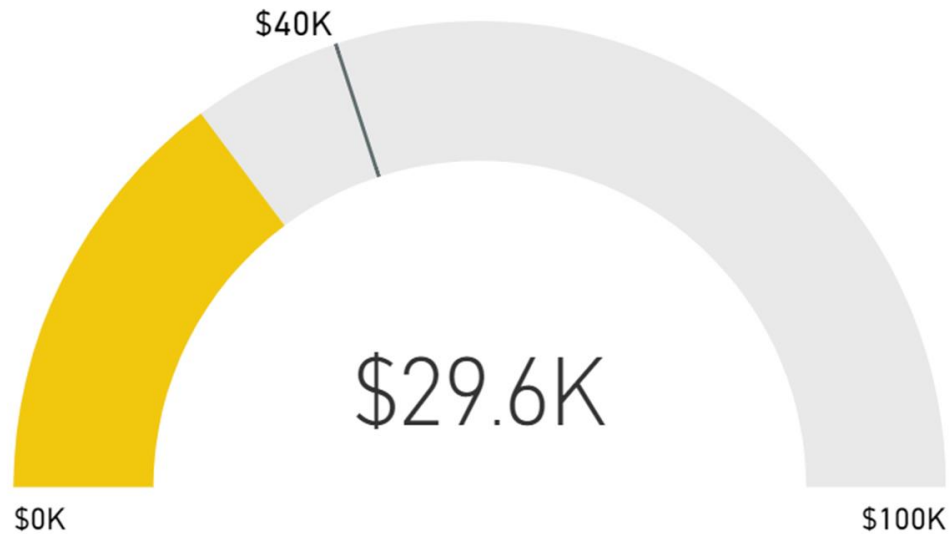
Typically represents **decreasing proportions** amounting to 100% total (not always though).

Power BI does not default sort, so users should **ALWAYS** sort from high to low (otherwise, plot looks messy).

VERY useful to help audience quickly prioritize items without having to actively filter.

GAUGE

Annual sales performance



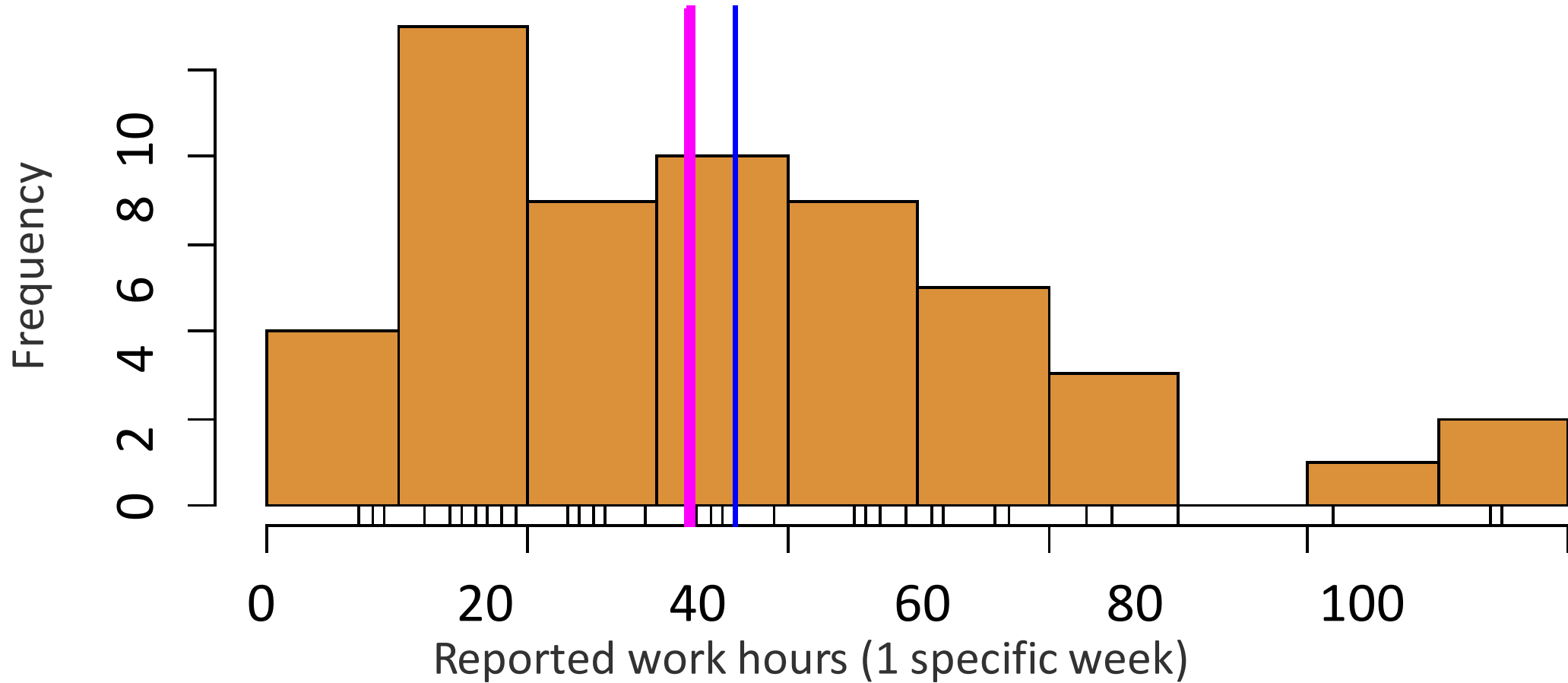
Often used as a dashboard component (with or without needle).

Displays single value measures towards goal / KPI.

Great to show progress (a bit of a management fad, though...)

Displays information that can be quickly scanned and **understood**.

HISTOGRAMS



HISTOGRAMS

Pros:

- known by many non-technical individuals
- easy to read (looks like something right out of high-school)
- can be adorned with added information (median, mean, hairs, etc.)

Cons:

- somewhat depressingly, a vast majority of the population does not know how to read them... yet it's conceivably one of the simplest graphical representations.

MAPS

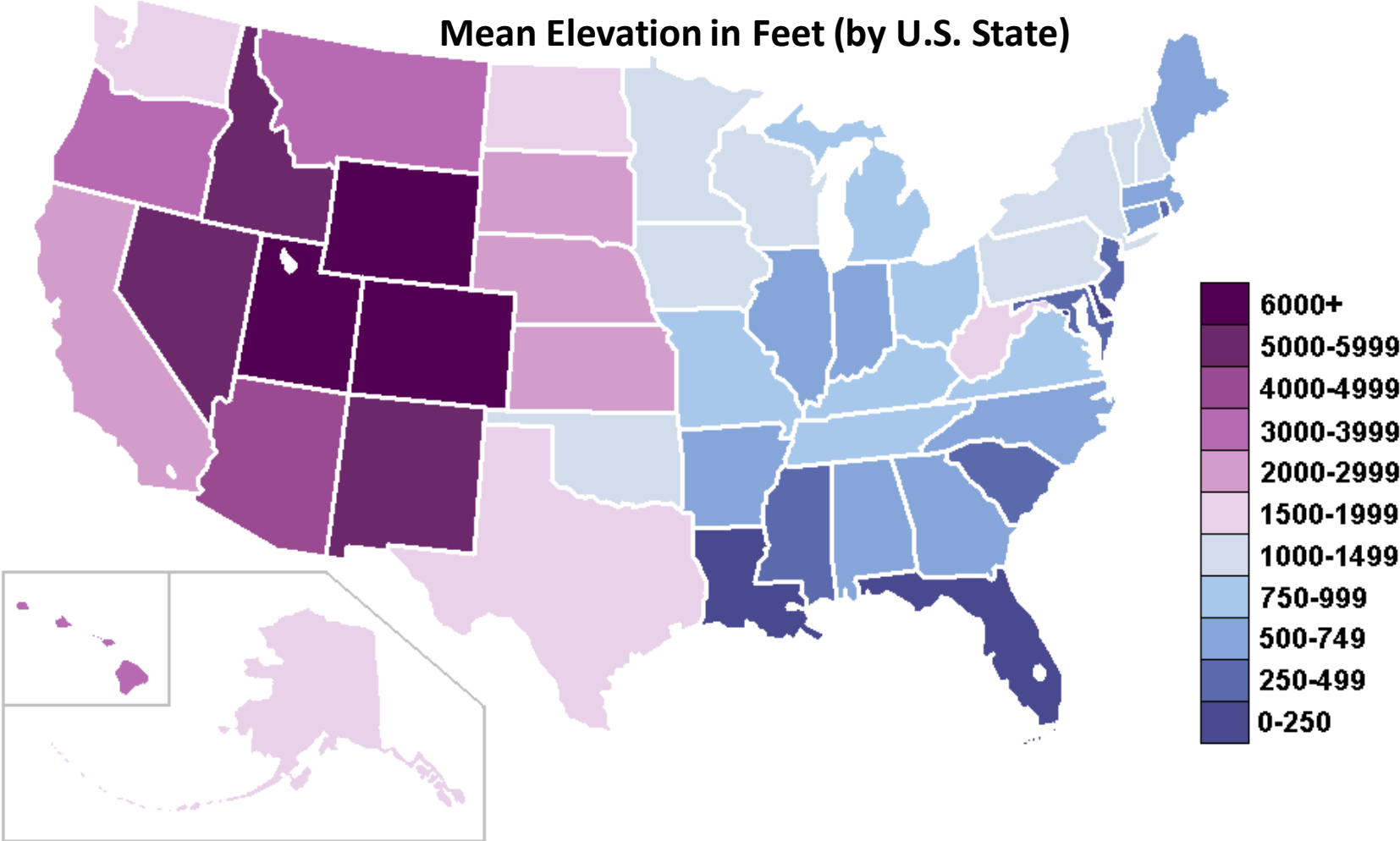
Most of us are quite familiar with geographical maps, so they tend to be easier to interpret.

Can produce a striking effect when the data visualization shows **unexpected results**

- which may mask significant information
- or lack of significant information
- or change the way you view things



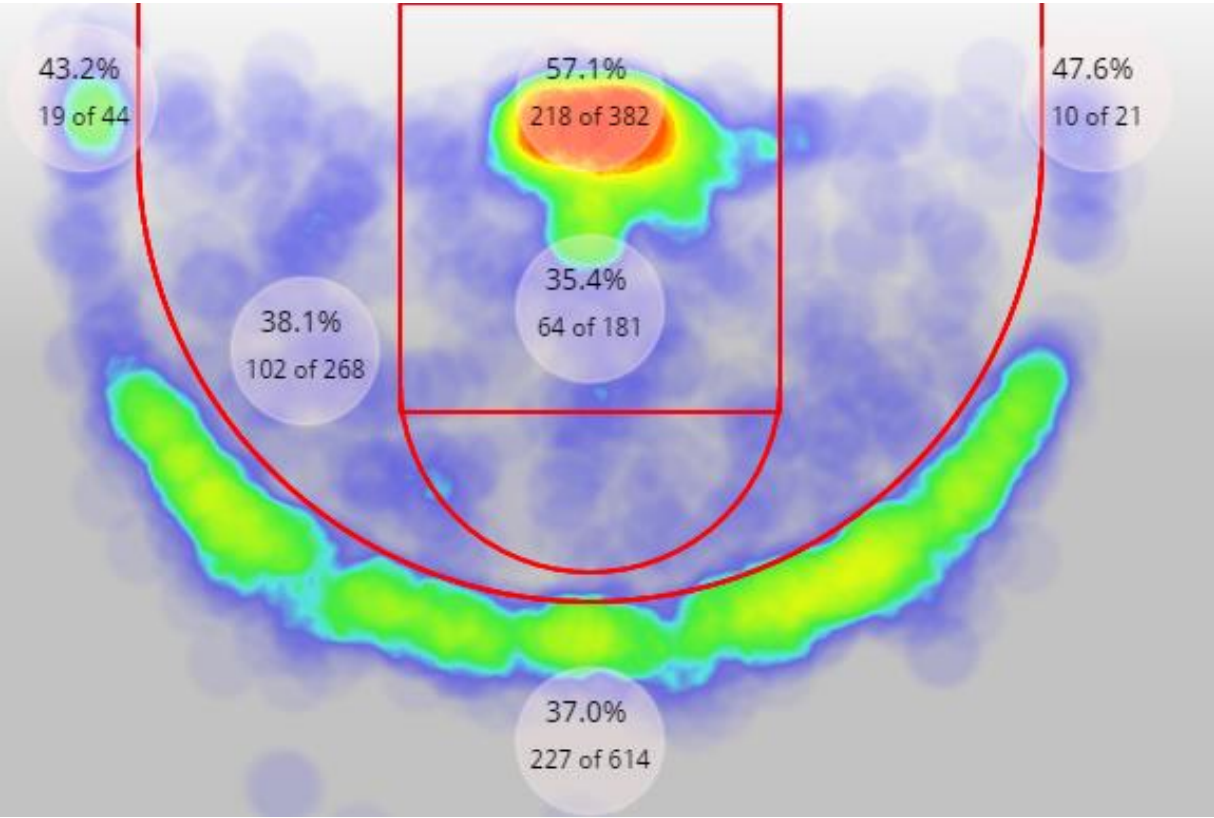
HEAT MAPS (CHOROPLETHS)



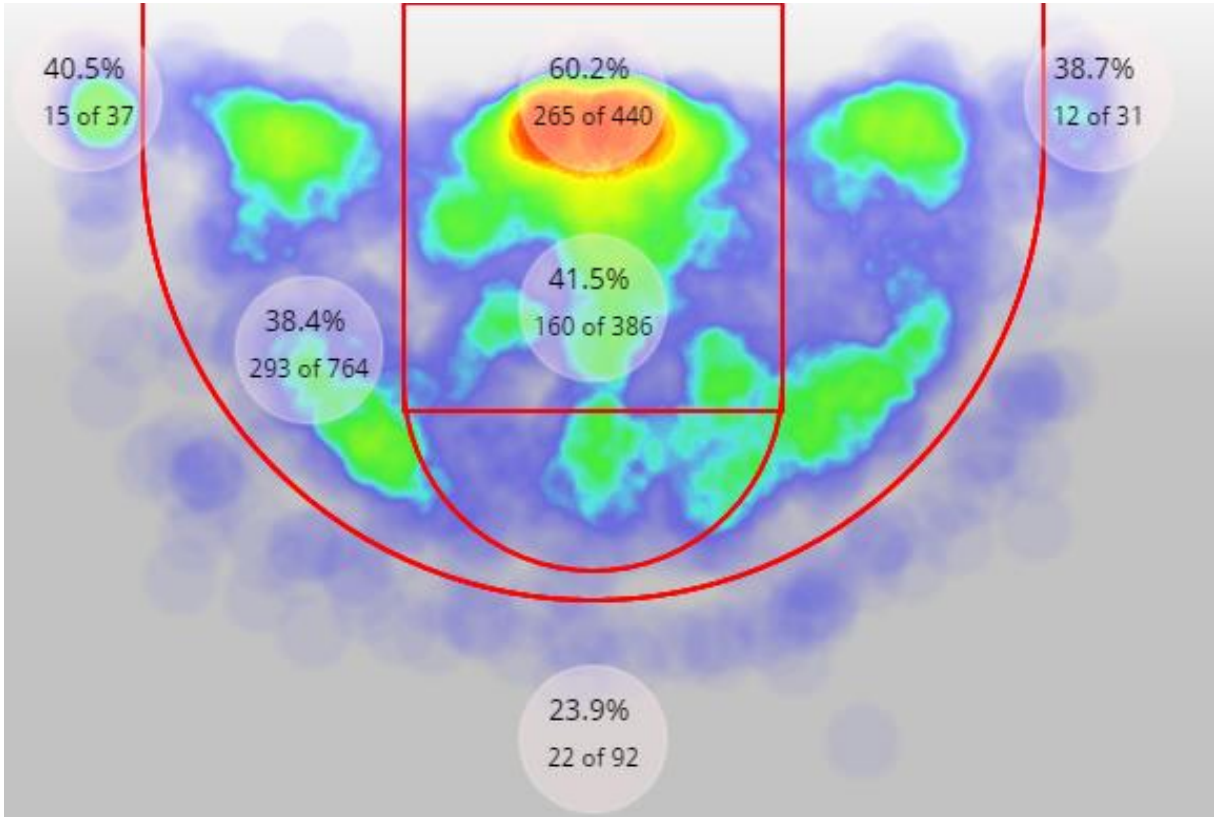
HEAT MAPS

NBA FG% (2015-16)

Kyle Lowry



DeMar DeRozan



HEAT MAPS

Ideal to look at the relationship between 3 or 4 variables

- if one of them represents a percentage or a value within a set range (in order to fix the colour scale, for comparison purposes)
- and the other can act as categorical variables / size variables

Better to **bin the data**, even if the axes variables are continuous (decreases the number of required observations for usefulness)

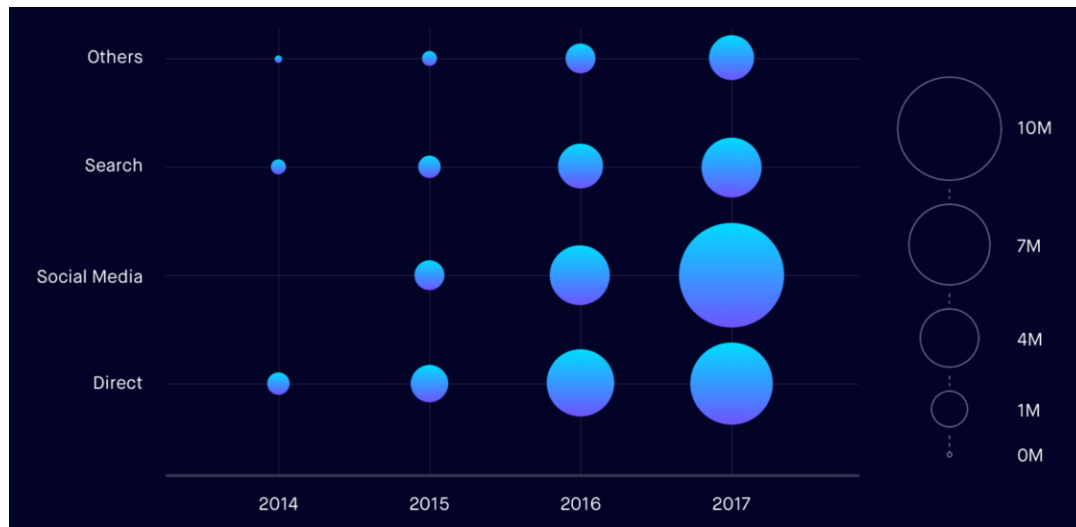
Easier to read if colours are selected along natural colour gradients, such as

Red → **Green** or **Red** → **Yellow** → **Green**

for instance (but that's not ideal if colour blind)

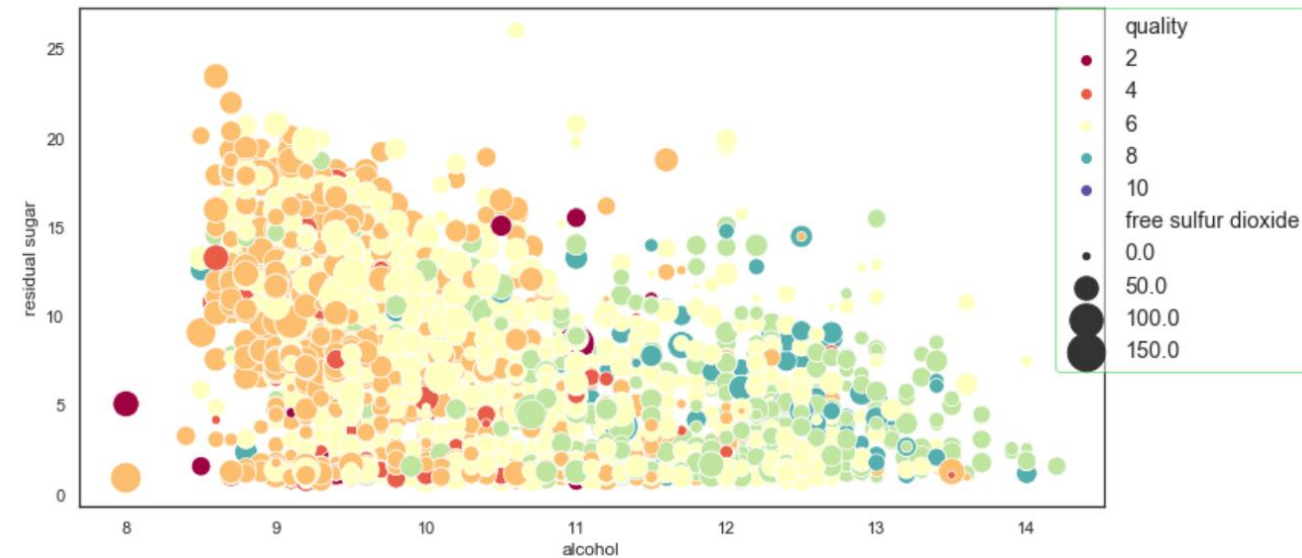
BUBBLE CHARTS

Website Traffic from Different Channels Over a Four-Year Period



from: <https://medium.muz.li/guide-to-data-visualization-comparison-part-1-678382ceef00>

Wine Quality Relative to Three Factors



from: <https://towardsdatascience.com/bubble-charts-why-how-f96d2c86d167>

BUBBLE CHARTS

Colour + geometry allow us to plot (at least) 2 extra variables on a 2D scatter plot

May need to re-scale or bin the available data







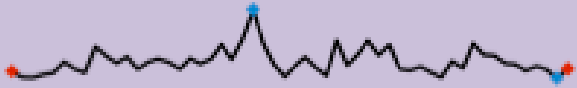

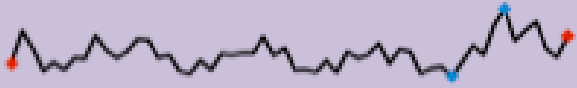
A movie could be used to visualize an additional ordinal variable

Text can also be added to visualize an additional categorical variable

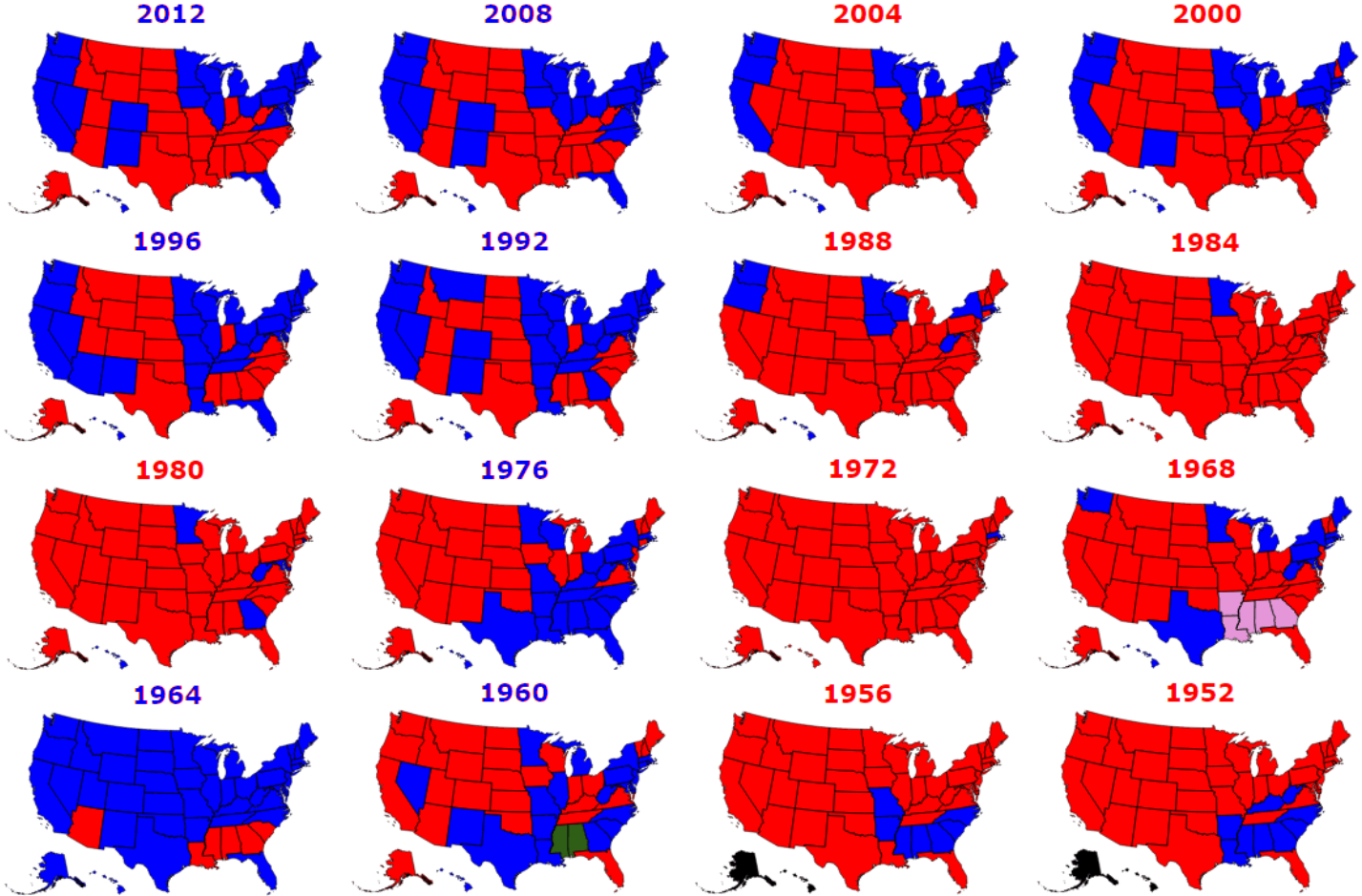
Works best when chart is **not too encumbered**

A **personal favourite** – a good mixture of traditional and modern features

SPARKLINES AND SMALL MULTIPLES

	Start	Monthly Number of Cases	End	Low	High	Mean	Std Dev	Blanks	Zeros	Trend
TOTAL	19502		17265	15150	25072	19903	2612	0.0	0.0	379.2
Hospital #1	46		19	3	46	19	9	0.0	0.0	-1.6
Hospital #2	156		240	101	326	194	60	0.0	0.0	9.7
Hospital #3	16		11	2	76	15	15	0.0	0.0	-2.9
Hospital #4	3		13	0	105	9	15	0.0	0.4	-1.8
Hospital #5	42		50	25	91	61	16	0.0	0.0	1.2
Hospital #6	48		53	34	169	67	25	0.0	0.0	0.6
Hospital #7	0		N.A.	0	0	0	0	2.2	9.8	0.0
Hospital #8	56		104	34	150	73	25	0.0	0.0	4.6

SMALL MULTIPLES



CHARTS TO AVOID

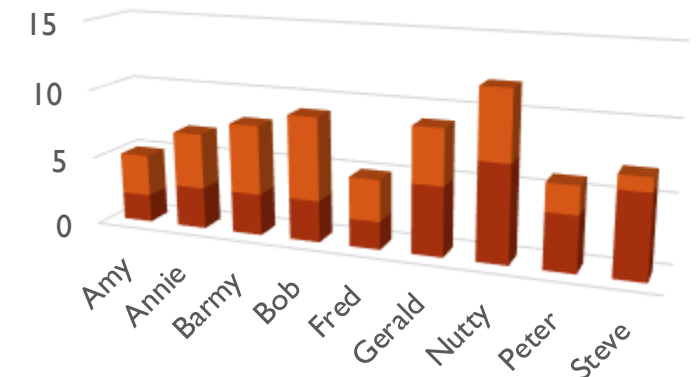
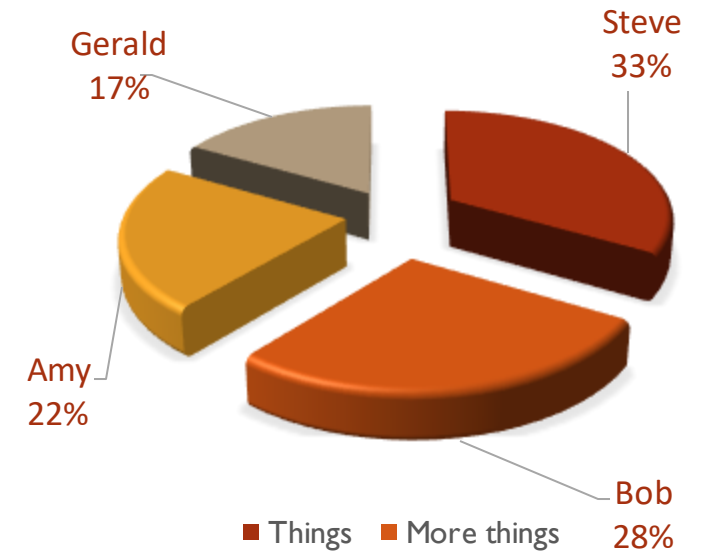
ANYTHING with an arc (except gauge)

- pie
- donut

Brains cannot compare arcs so they can be misleading:
without labels, how easy is it to compare Steve & Bob?

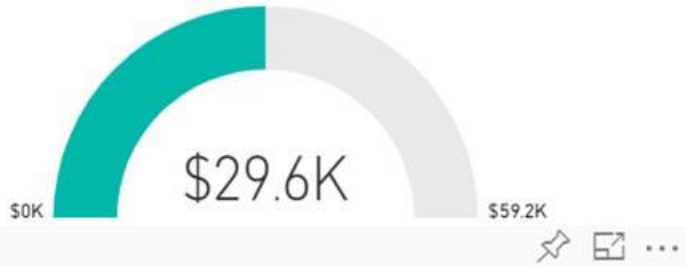
ALL 3D IS EVIL!

- as with arcs, we cannot easily visually compare data series
- adds way too much clutter



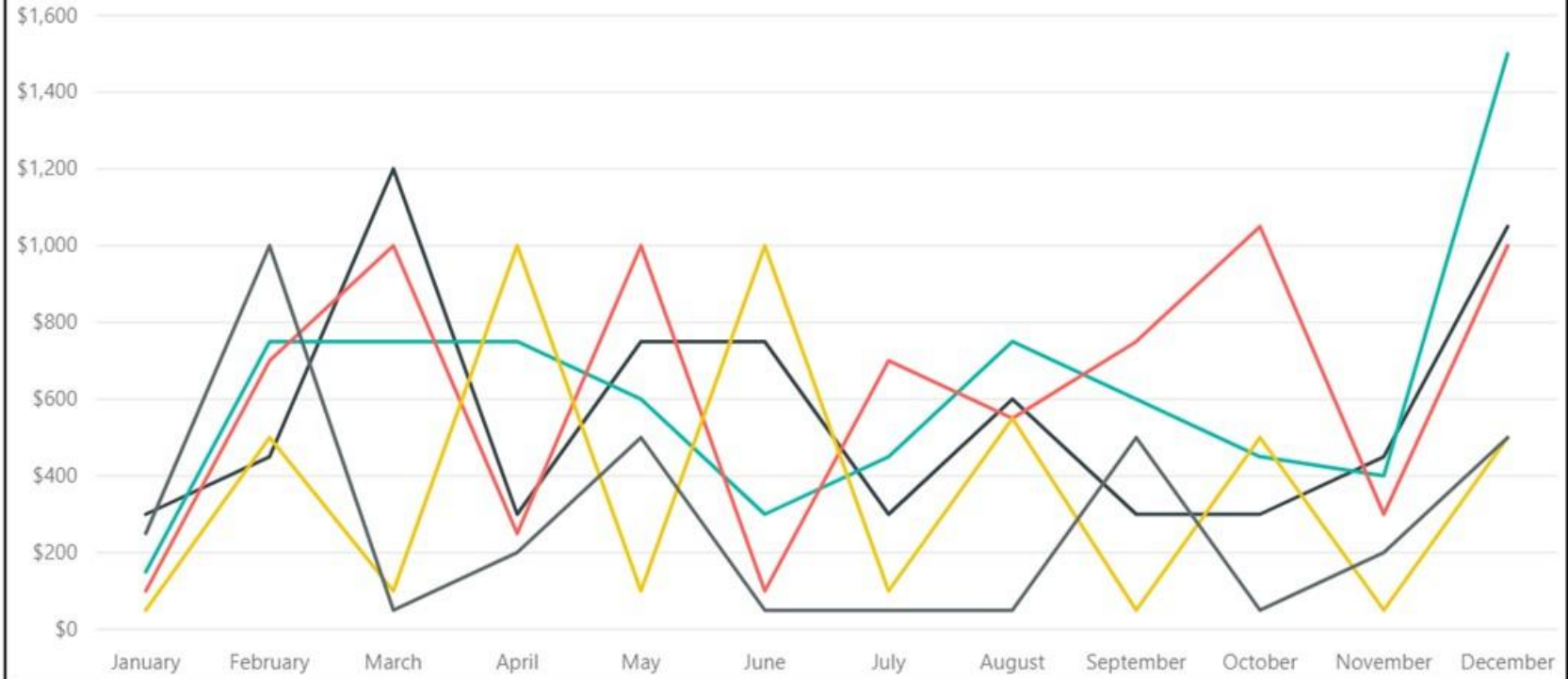
Sales Dashboard

\$ sales



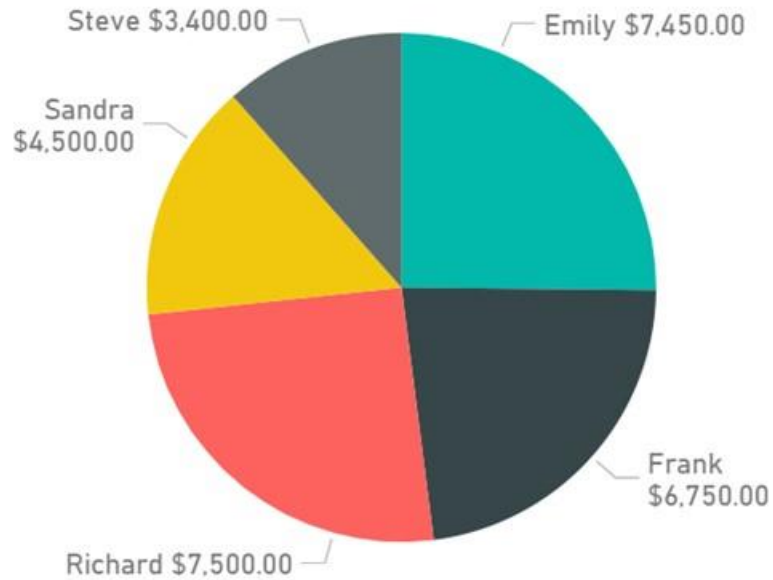
\$ sales by Month and Salesperson

Salesperson ● Emily ● Frank ● Richard ● Sandra ● Steve



\$ sales by Salesperson

Salesperson ● Emily ● Frank ● Richard ● Sandra ● Steve



\$ sales by Product and Salesperson

Product ● Car ● Bike ● Sled

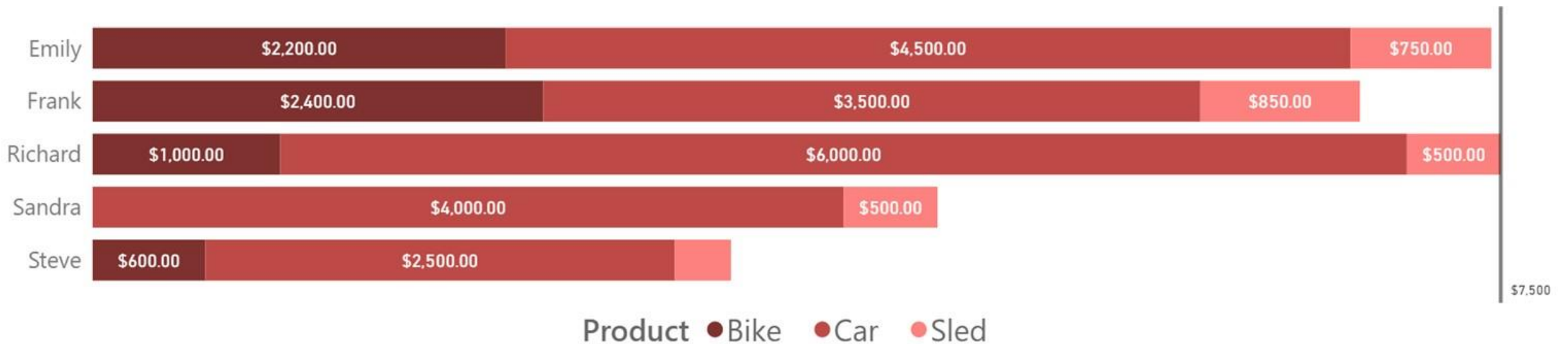


Sales Dashboard

Annual Sales for 2017

Total Sales

\$29.6K



EXERCISE

Find examples of different charts displaying information about the same dataset?

What are the strengths and limitations of the charts, relative to the specific dataset?

INTERACTIVE AND ANIMATED VISUALIZATIONS

INTERACTIVE AND ANIMATED VISUALIZATIONS

Animation **does not always** improve a visualization. What insights can interactivity provide? That depends on the data, and on the visualization method.

Examples:

- [The Clubs That Connect the World Cup](#), NY Times, 2014
- [Who Marries Whom](#), Bloomberg, 2016
- [Hipparcos Star Mapper](#), European Space Agency, 2016
- [The Internet of Things – a Primer](#), Information is Beautiful, 2016
- [The Genealogy and History of Popular Music Genres](#), Musicmap, 2016

INTERACTIVE AND ANIMATED VISUALIZATIONS

Examples (continued):

- [Sequences Sunburst](#), Kerry Rodden, 2015
- [Health and Wealth of Nations](#), Gapminder Foundation
- [Mobius Transformations Revealed](#), Arnold D.N, Rogness, J, 2007
- [Visualizing the Riemann \$\zeta\$ Function and Analytic Continuation](#), 3Blue1Brown, 2016
- [Small Arms and Ammunition – Imports and Exports](#), Google, 2012
- [The Evolution of the Web](#), Google, Hyperakt, Vizzuality, 2012
- [peoplemovin](#), Carlo Zapponi, 2012

DISCUSSION

“There is always a danger that if certain types of visualization techniques take over, the kinds of questions that are particularly well-suited to providing data for these techniques will come to dominate the landscape, which will then affect data collection techniques, data availability, future interest, and so forth.” (P. Boily)

Even when done well, 85% of users don't bother with interactive viz (NY Times).

Take-Away: explore the data and try different methods