

Data Visualisation and
Dashboarding

Week 3 – Design Principles

UNIVERSITY OF
WESTMINSTER



**“Clutter and confusion are
failures of design, not
attributes of information.”**

— Edward Tufte

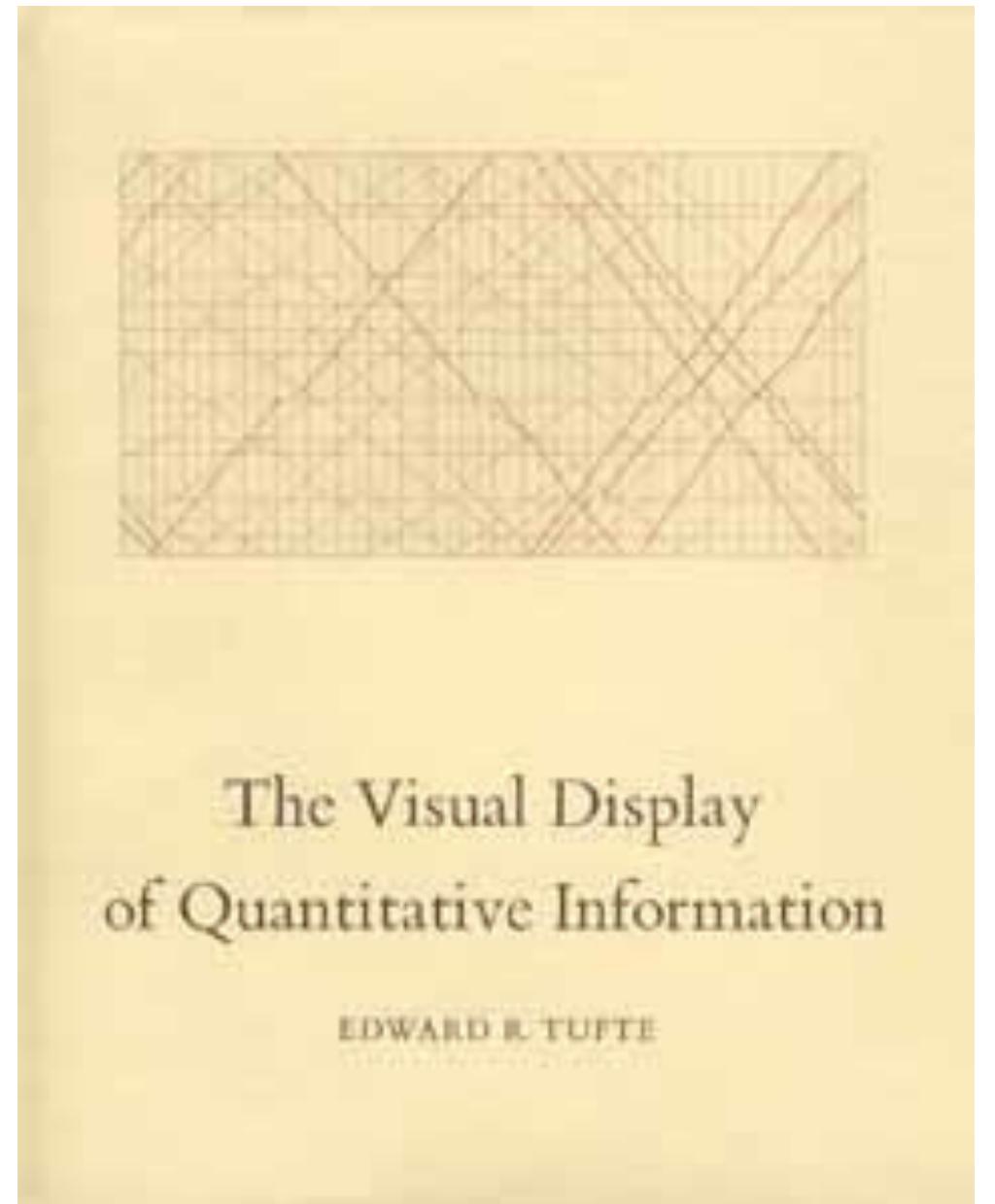
Key Principles

Strive for graphical integrity (lie factors)

Theory of Data Graphics and Data Ink

Avoid Chart Junk

Use Design strategies



Graphical Integrity

No lying

Clear mapping

Clear labelling

Show data variation

Use deflated and standardised figures

Do not quote data out of context

Lie factor

The figure on the left has an end line at 27.5 which is 5.3 inches wide and starts at 18 with a width of 0.6 inches

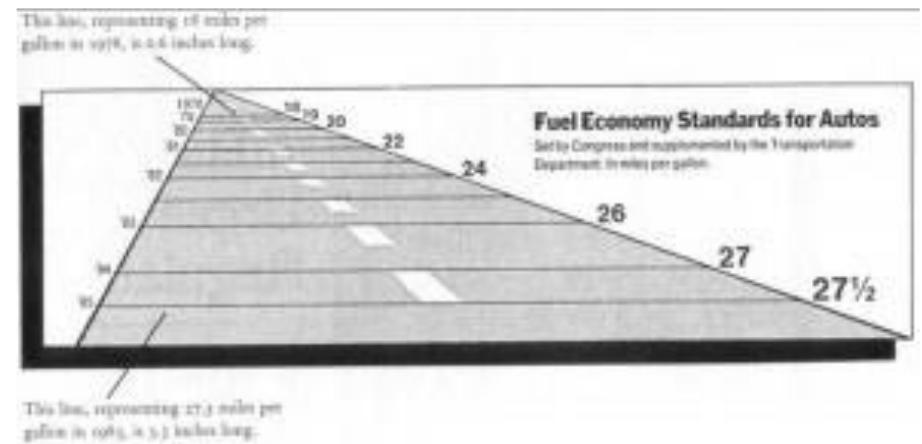
If we work this out the size of effect in graphic is $\frac{5.3 - 0.6}{0.6} = 7.8333 \dots$

The size of the effect in the data is $\frac{27.5 - 18}{18} = 0.52777 \dots$

The Lie Factor is $\frac{7.8333}{0.52777} = 14.8$, which is clearly over 1.0

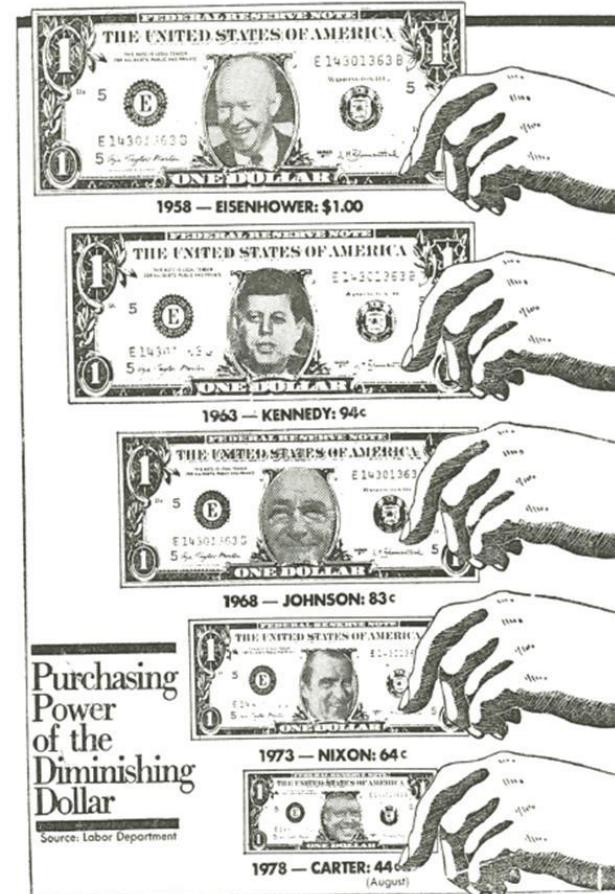
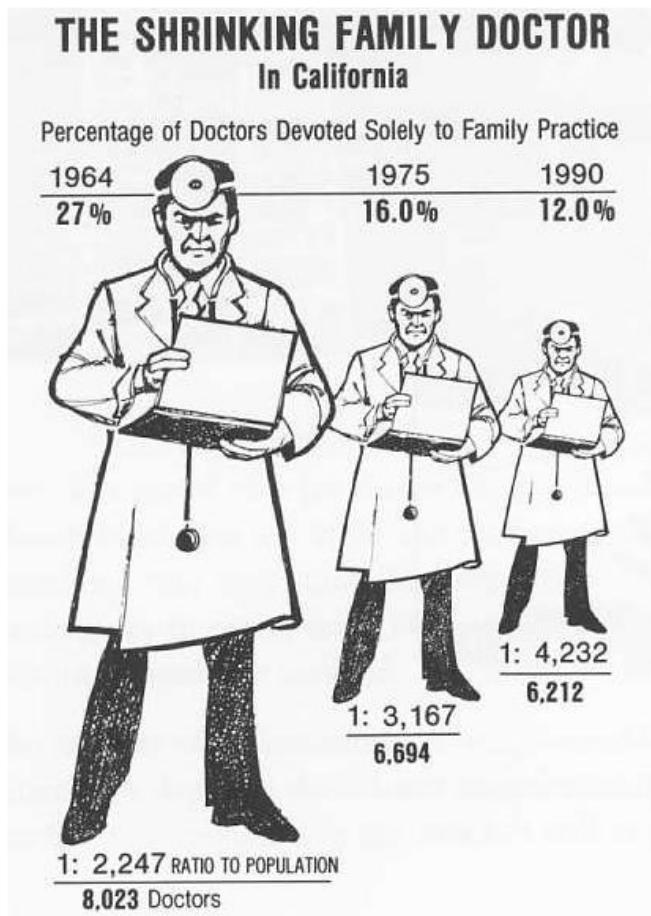
This suggests that this representation is highly misleading.

$$\text{Lie factor} = \frac{\text{size of effect in graphic}}{\text{size of effect in data}}$$



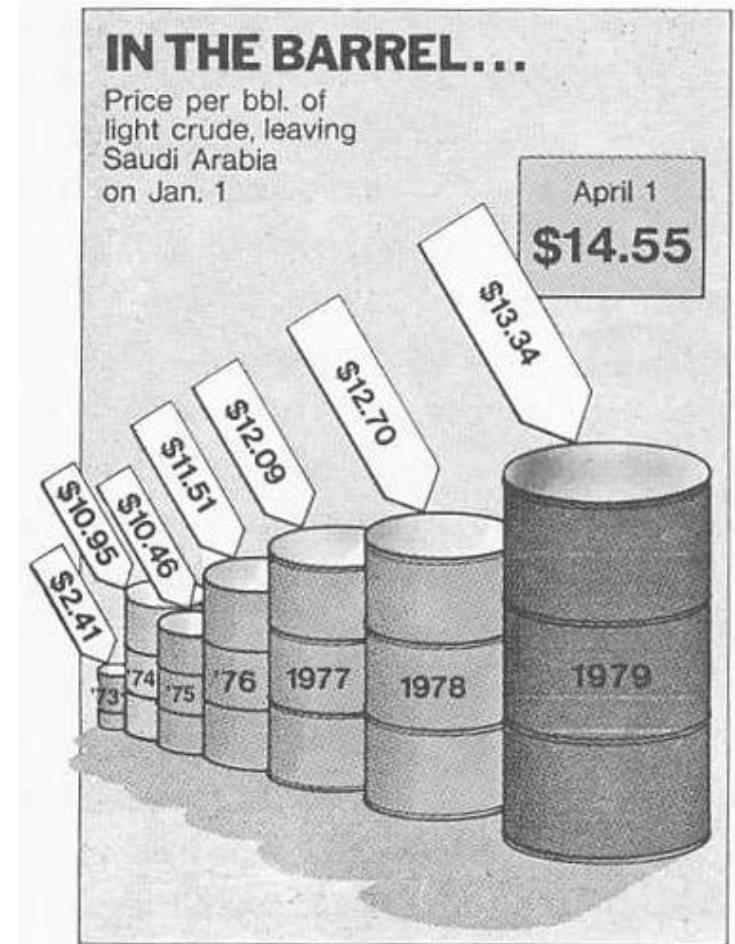
New York Times, August 6, 1998, p. D-6.

Lie factor - areas



Clear mapping

Graphical representations of numbers need to be directly proportional to actual numbers



Clear labelling

Minimise ambiguity with clear, detailed and thorough labelling

Write interpretation directly into graph

Highlight or label interesting data points

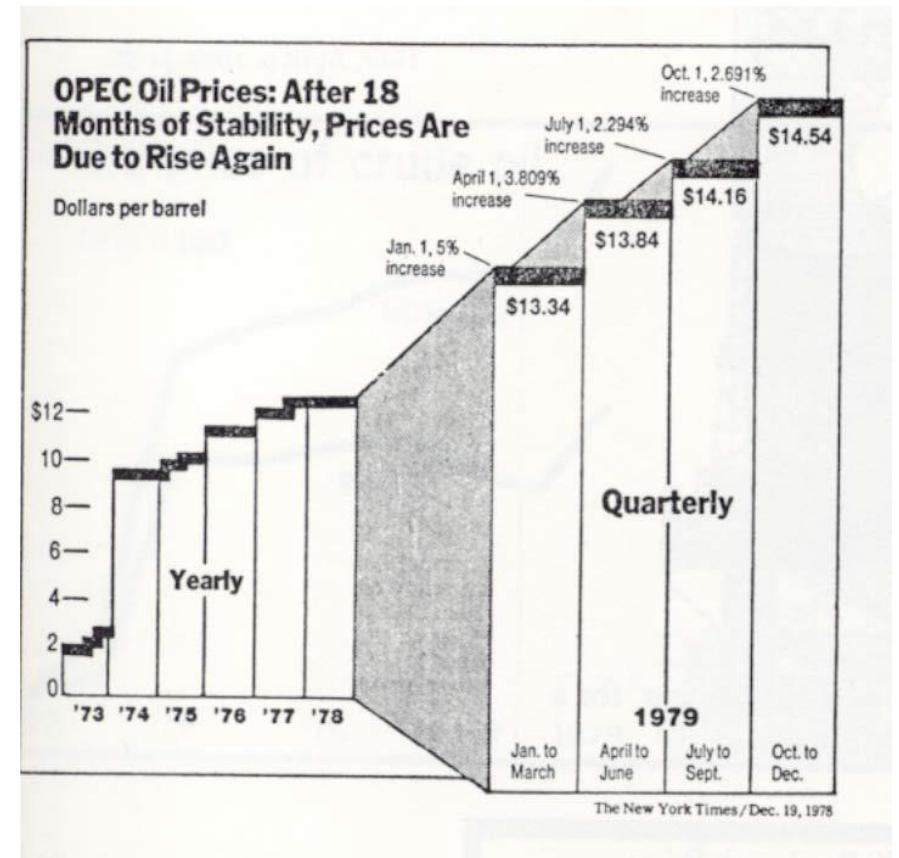


Show data variation (not design variation)

Two vertical scales ('73-'78 and 1979)

Two horizontal scales

Huge variation as both axes shift

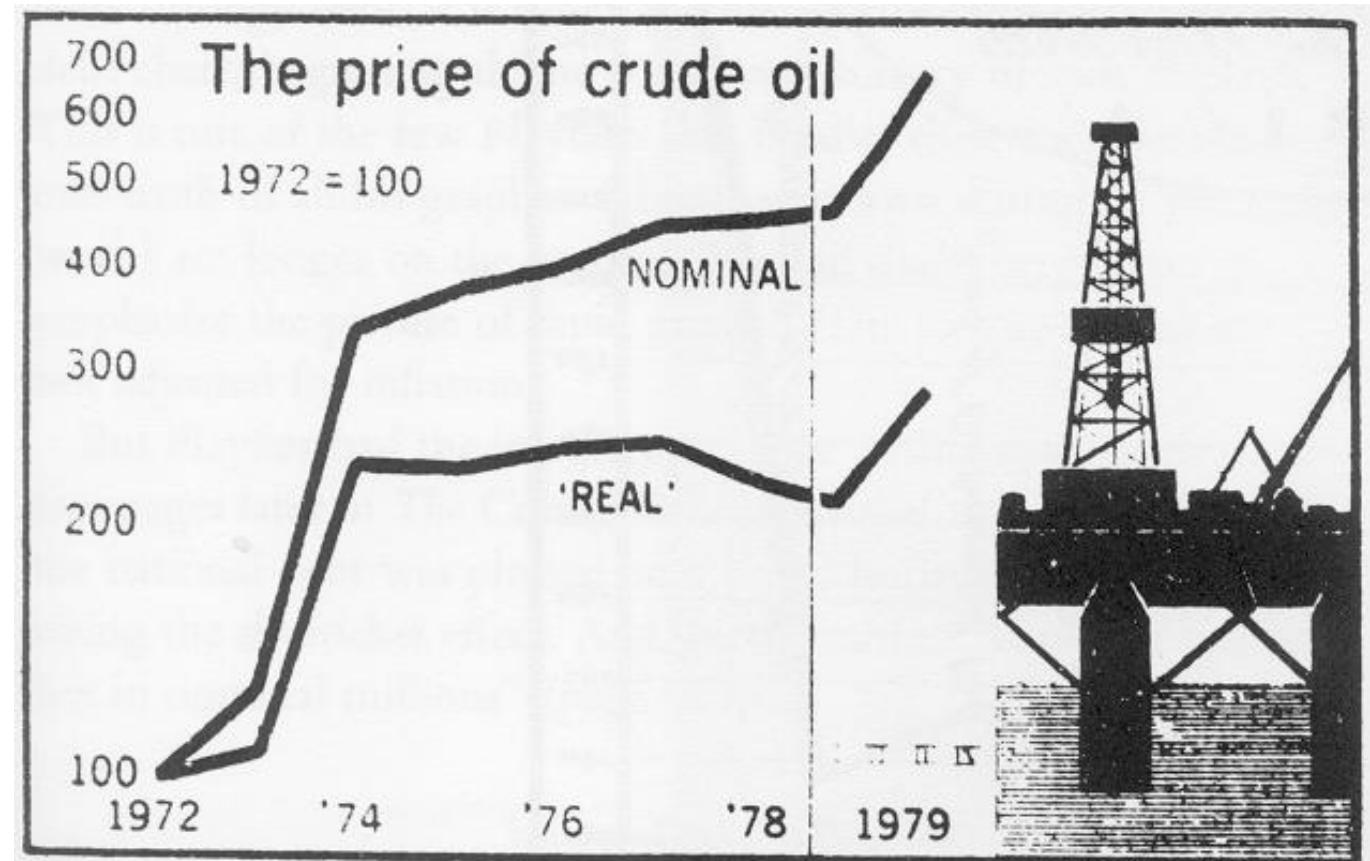


Use deflated and standardised figures

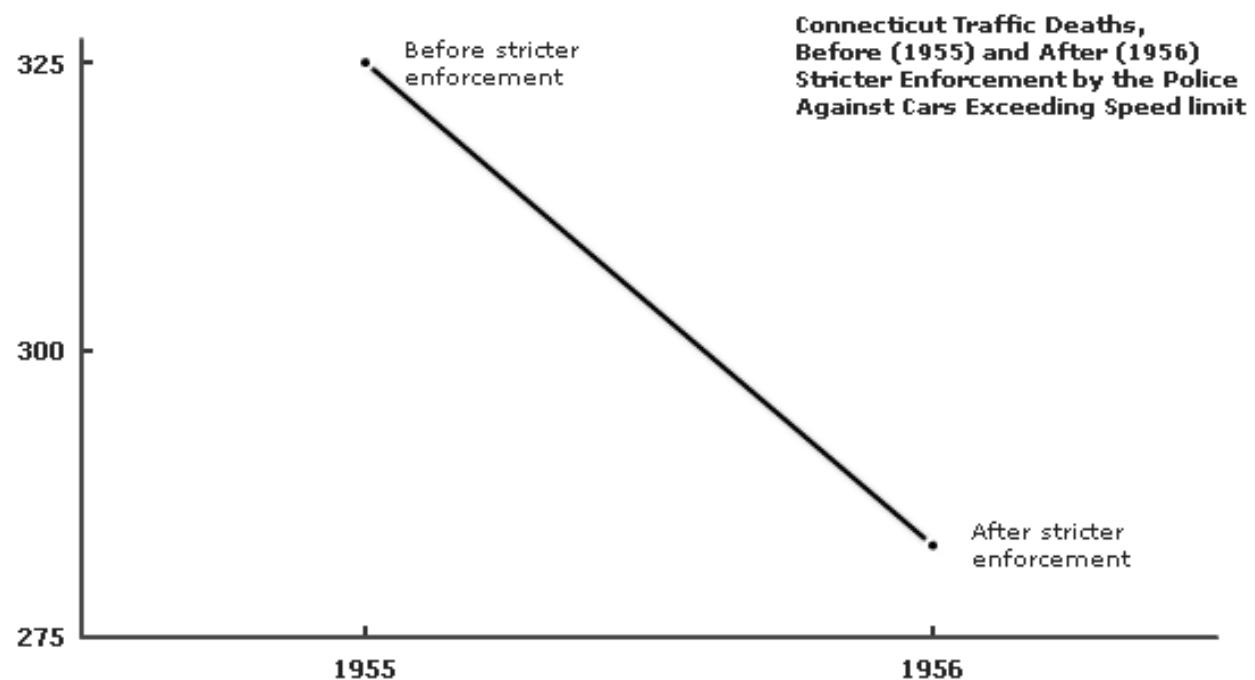
Use 'real' figures when looking at money

Use 'per capita' or indexed figures

Etc...

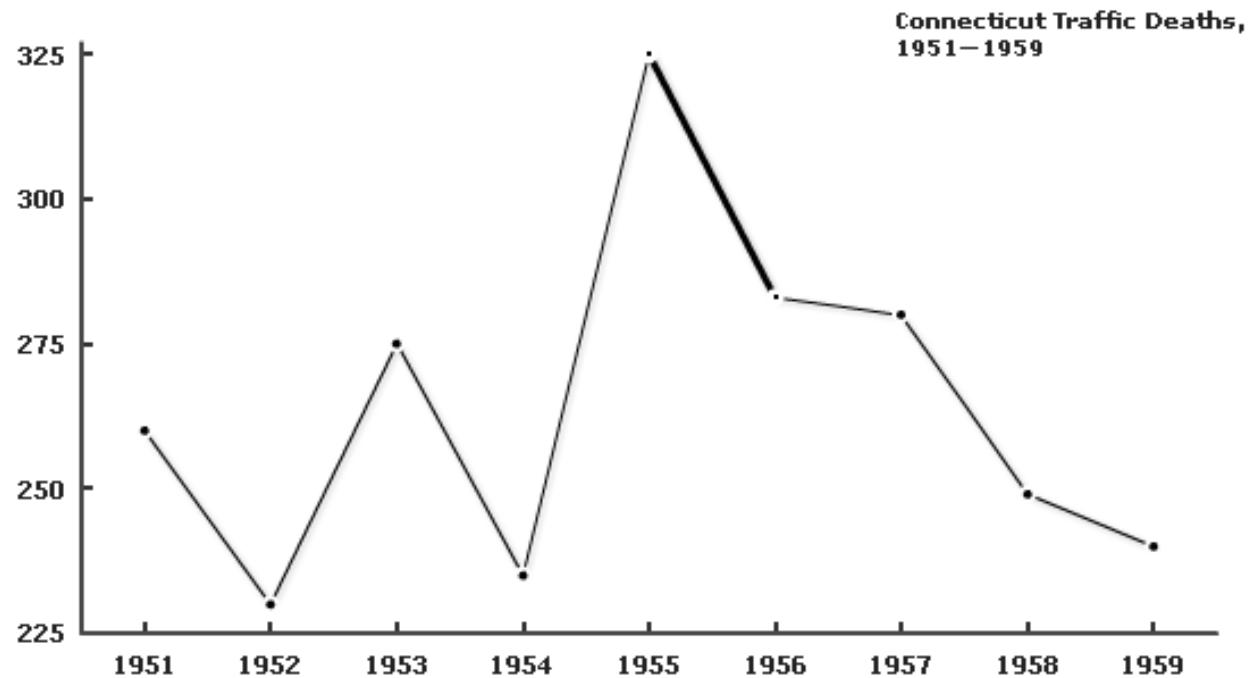


Do not quote data out of context



Rendered by AnyChart

Do not quote data out of context



Rendered by AnyChart

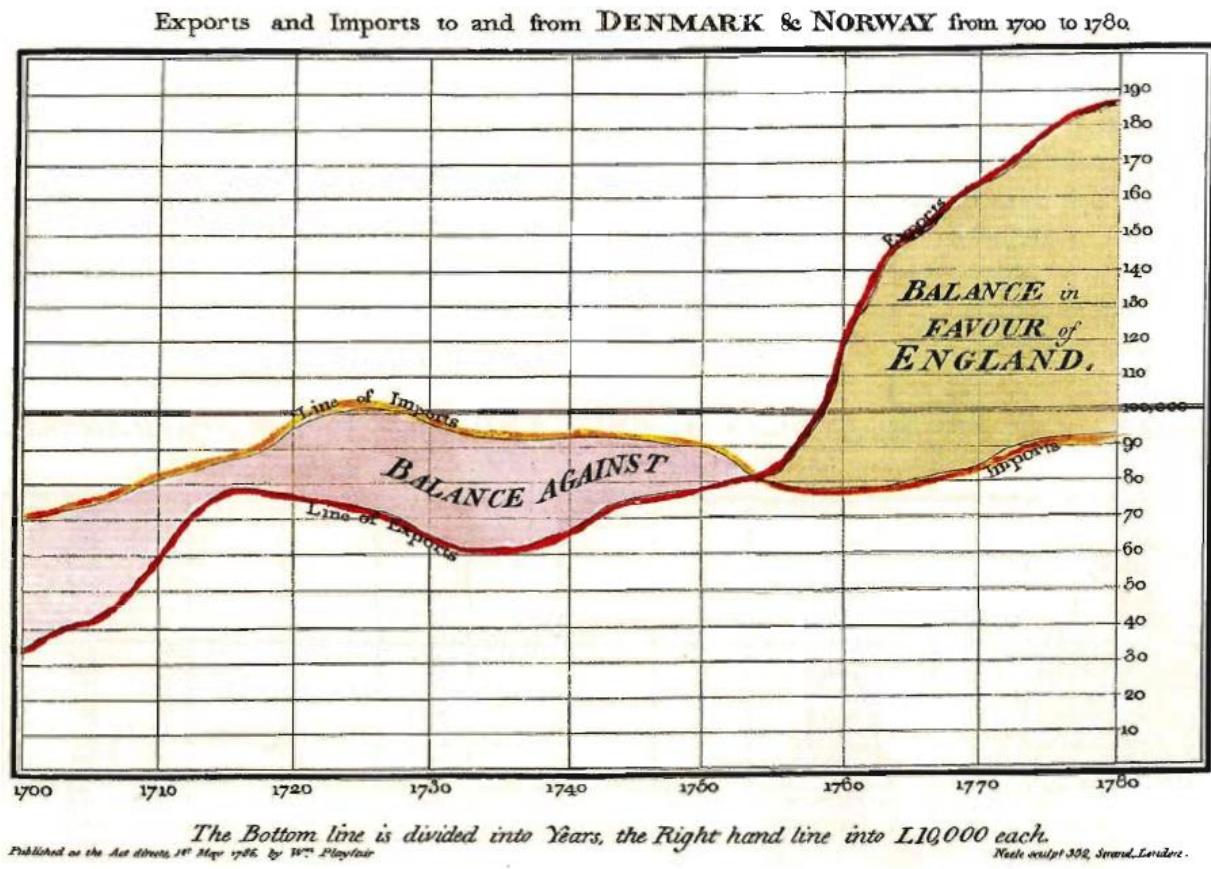
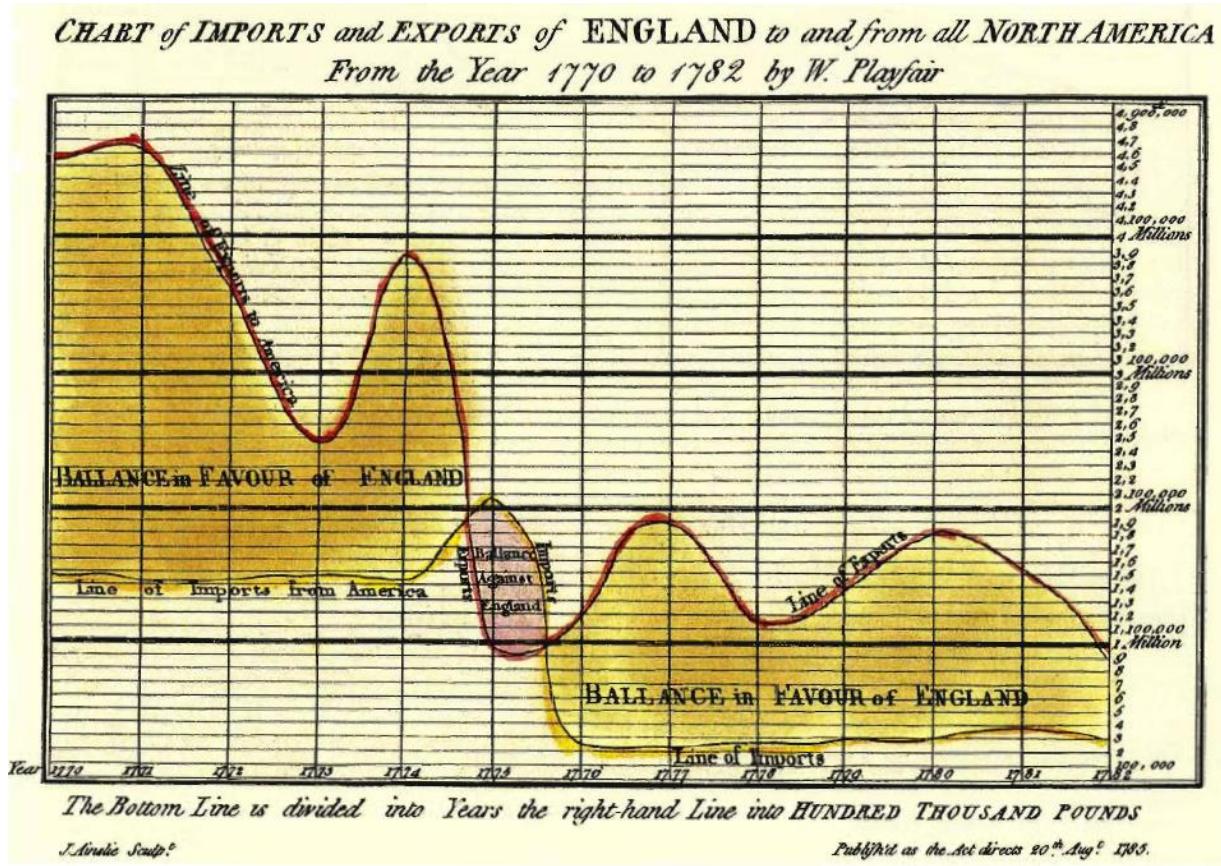
Do not quote data out of context

Surrounding data can give important context

Meaning of graph can be completely changed

Data Ink

Data-Ink and Graphical Redesign



Data Ink

Large share of ink on a graphic should present data/information.

Data ink is non-erasable core of a graphic.

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

= $1.0 - \text{proportion of a graphic that can be erased}$

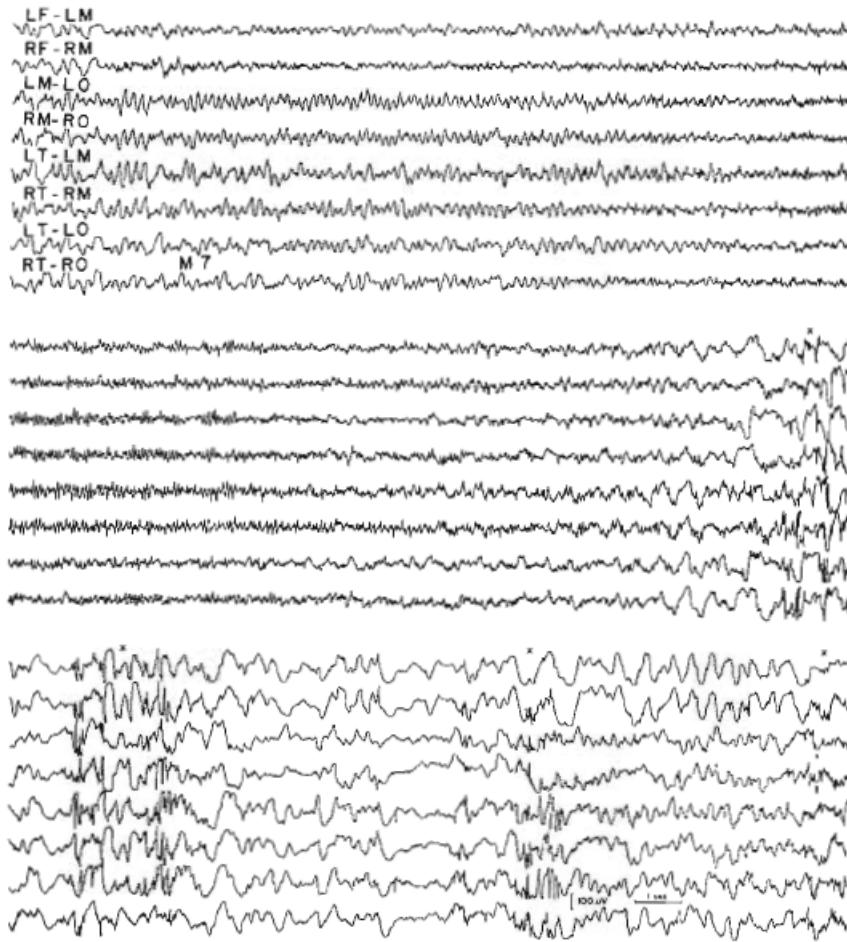
Good Example

High Data-Ink ratio

Note scale in bottom block, lower right.

Source: Kenneth A. Kooi.

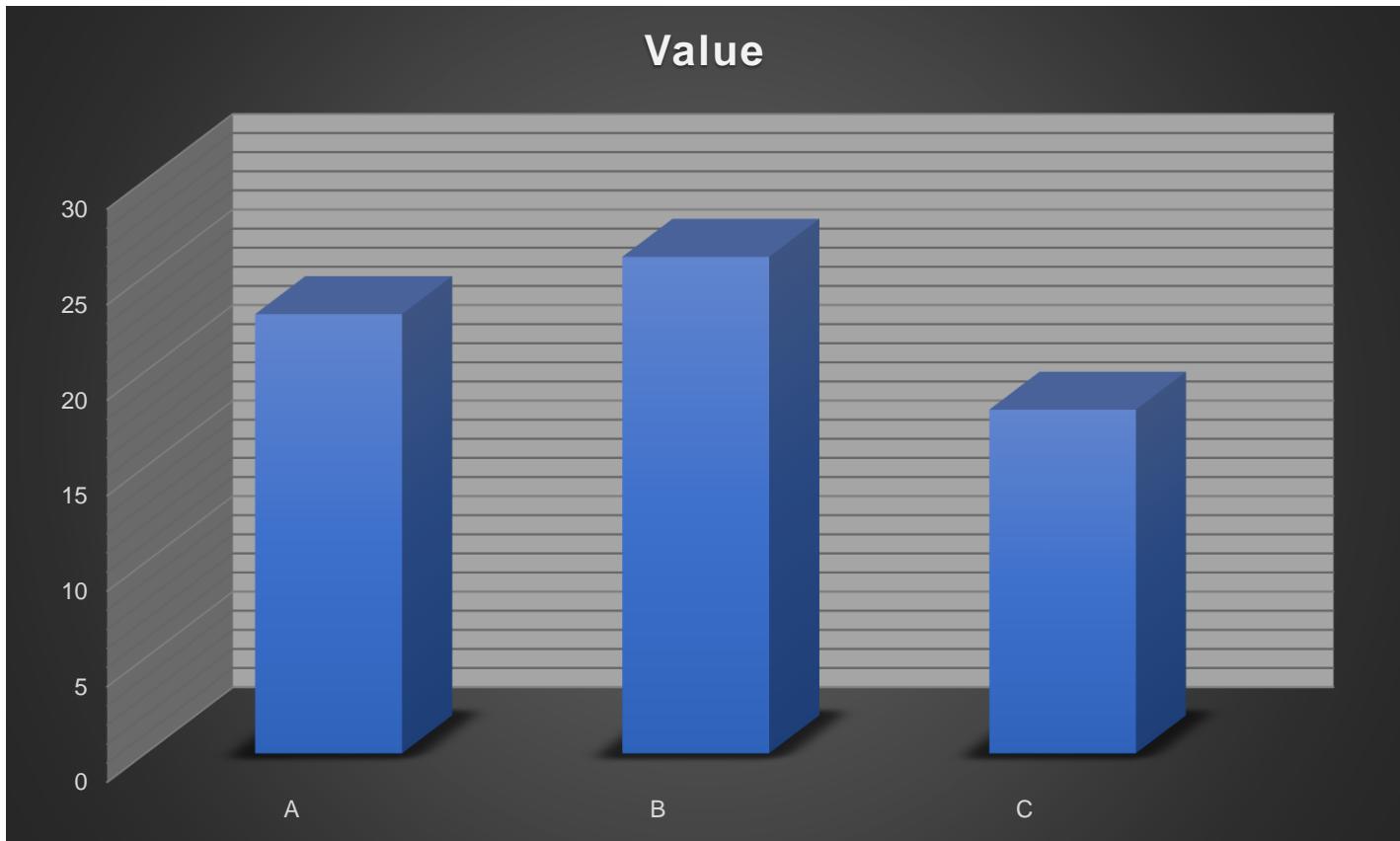
*Fundamentals of
Electroencephalography* (New
York, 1971).



Bad Example

Low data-ink ratio

Superfluous elements include 3D-effect, background, title, axis, grid lines, ...

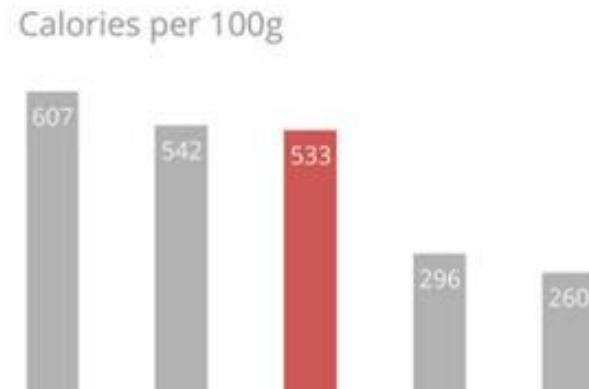


Example: From bad to good

Before



After



<https://www.darkhorseanalytics.com/blog/data-looks-better-naked>

Maximise share of data-ink

Every bit of ink requires a reason.

Almost always, this reason should be to present new information.

Erase non-data ink (within reason)

Ink that fails to depict statistical information clutters data.

Redundant data-ink



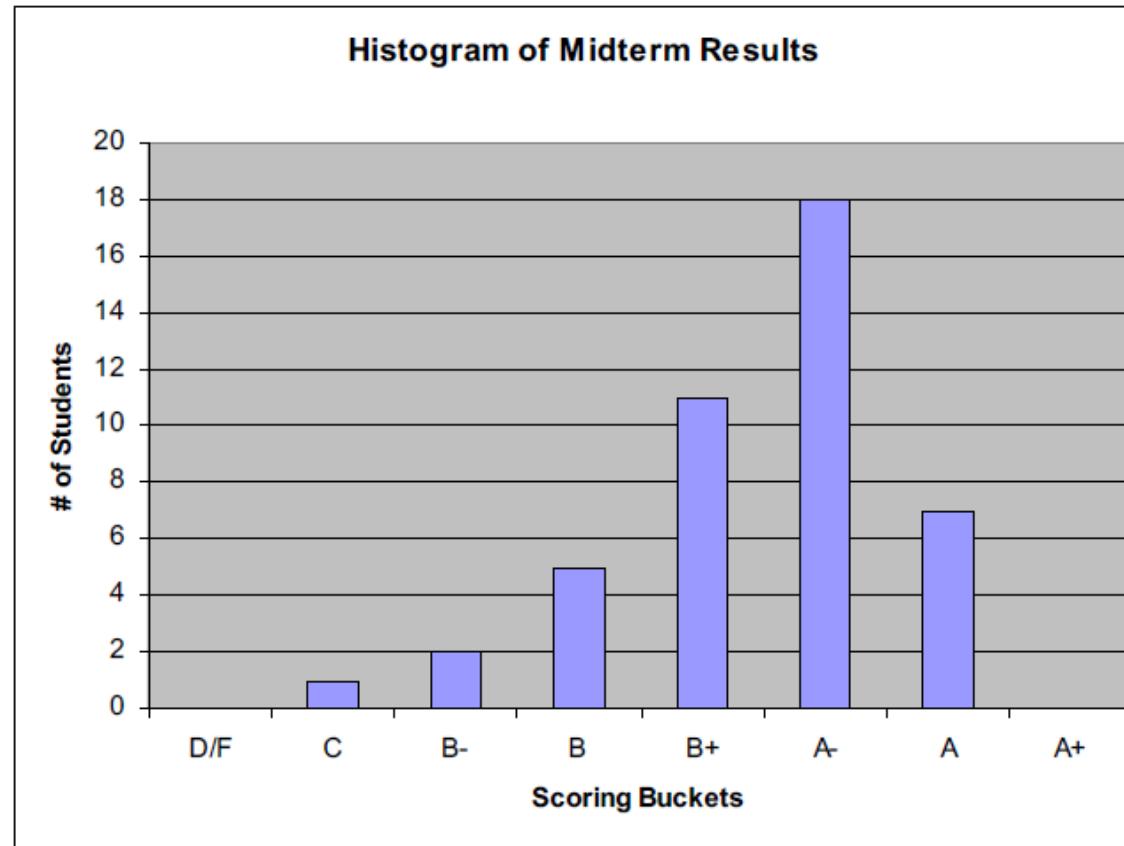
Redundant data-ink

Value is represented in six different ways:

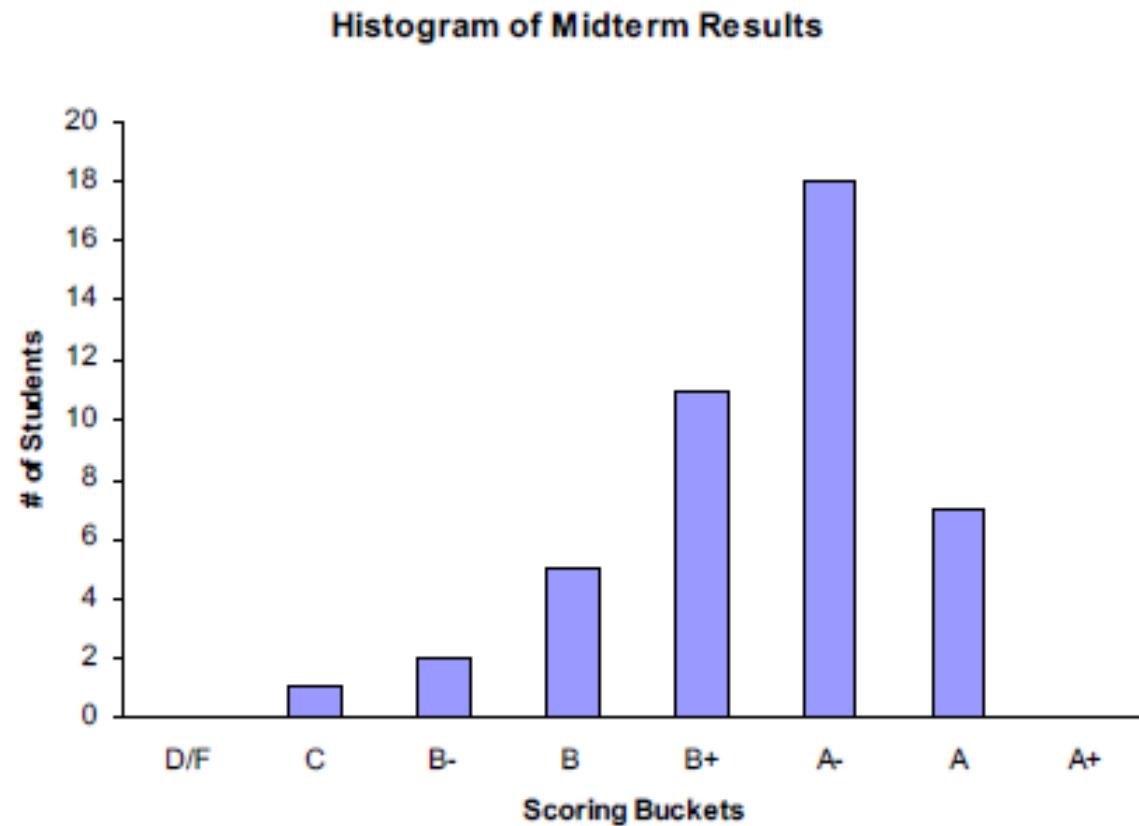
- 1) Height of left line
- 2) Height of colour
- 3) Height of right line
- 4) Position of top horizontal line
- 5) Position of number at column's top
- 6) Number itself



Erase non-data ink

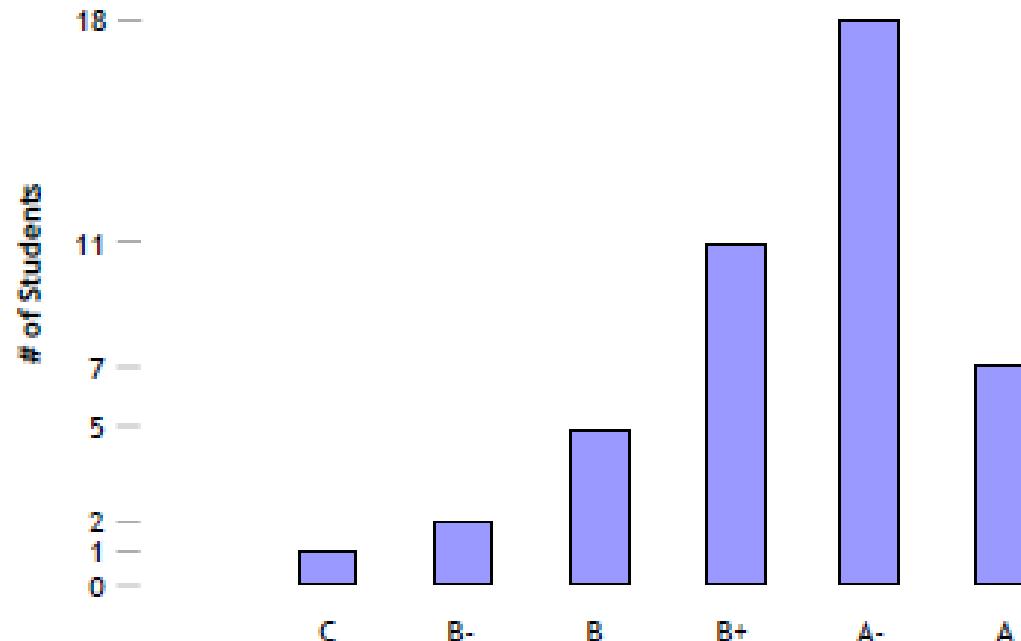


Erase non-data ink



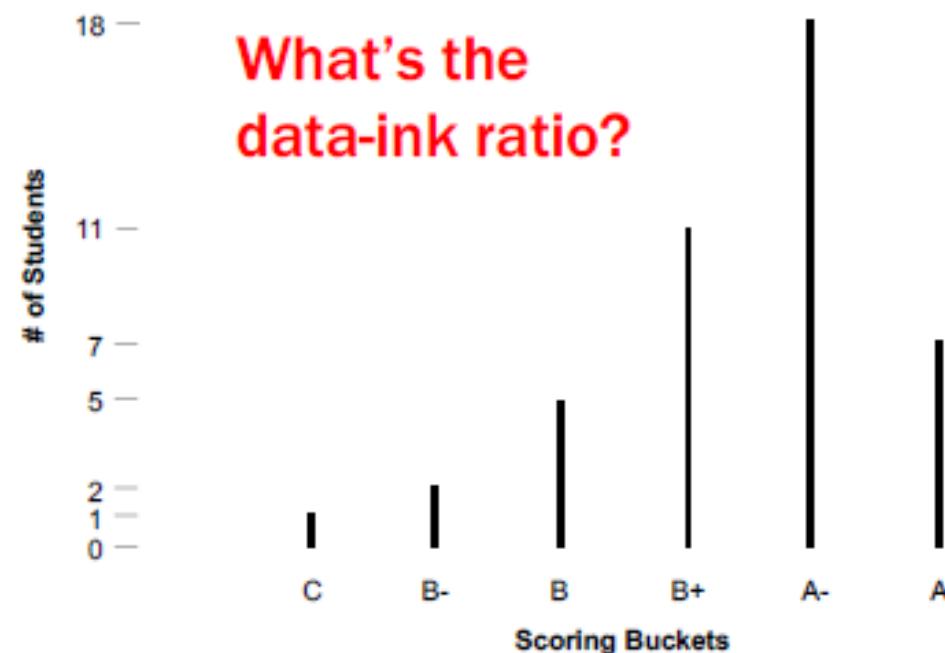
Erase non-data ink

Histogram of Midterm Results



Erase non-data ink

Histogram of Midterm Results



Maximise data-ink

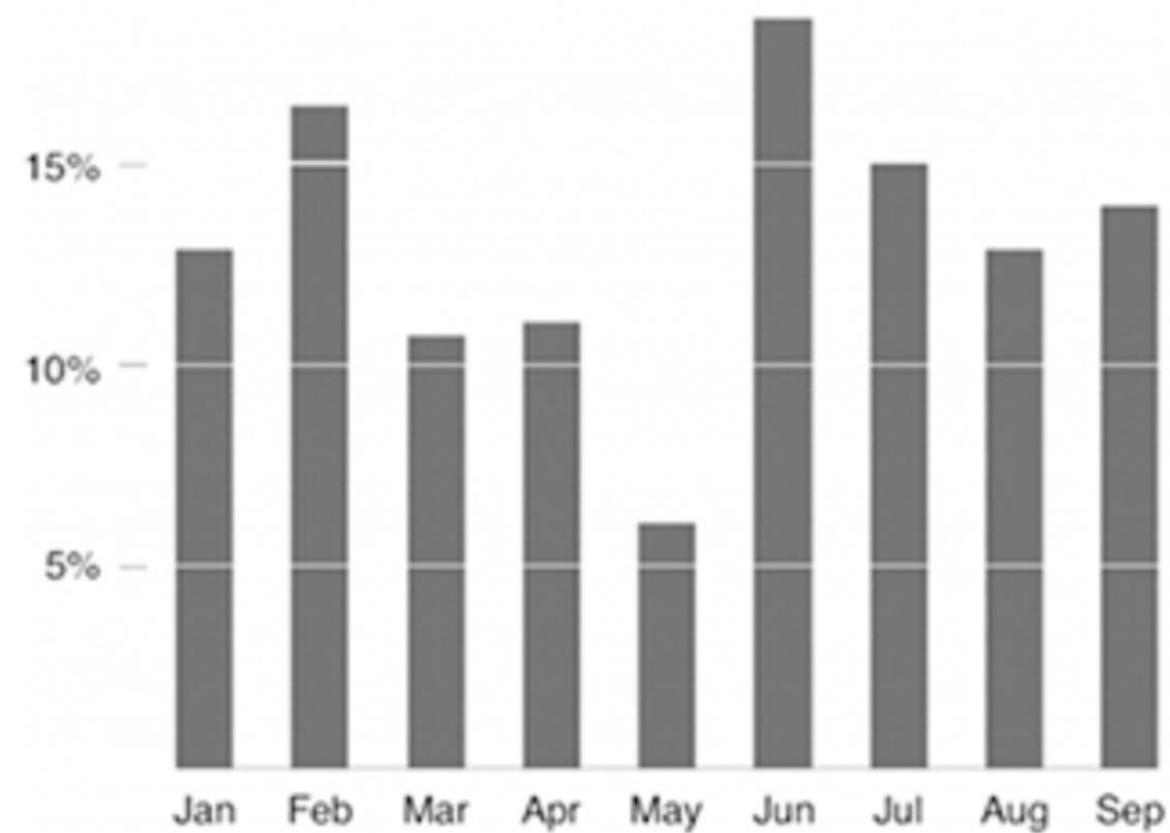
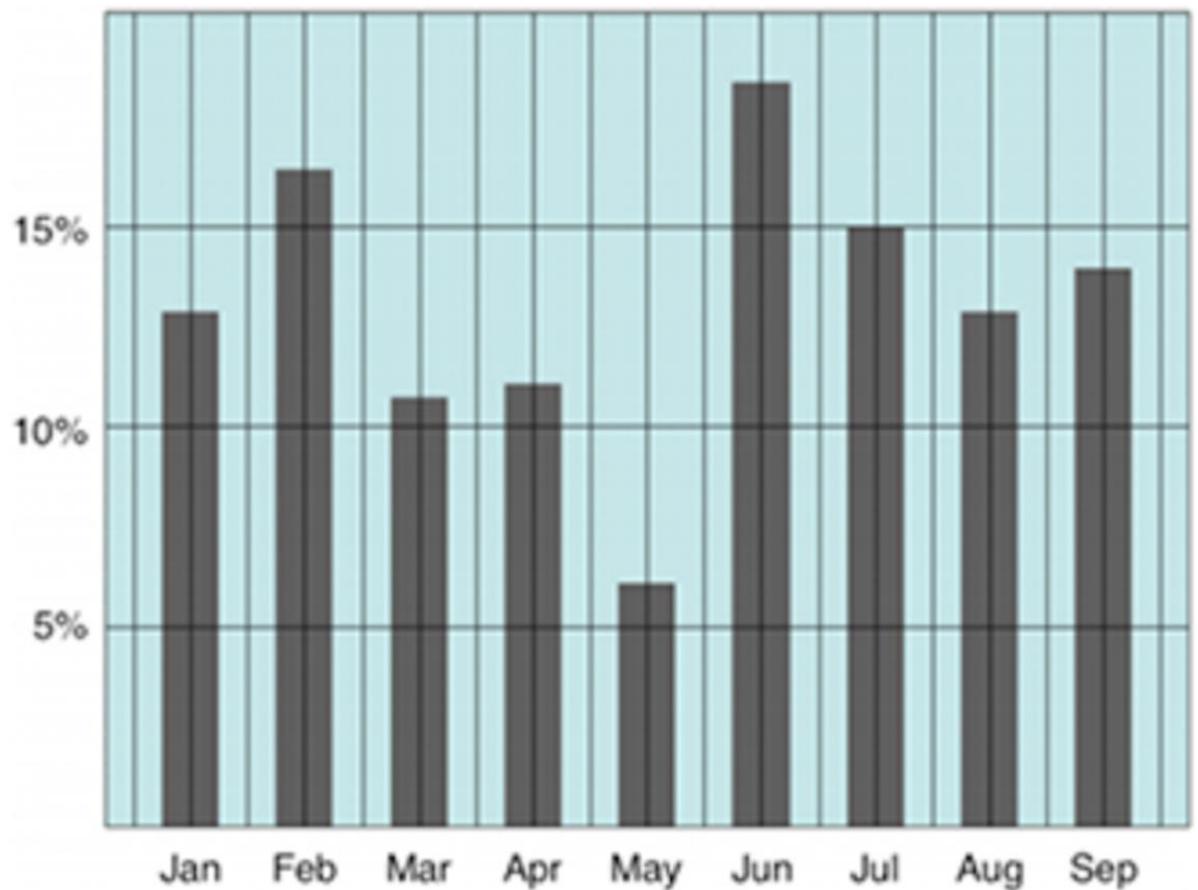


Chart Junk

Chart Junk – Moiré effect

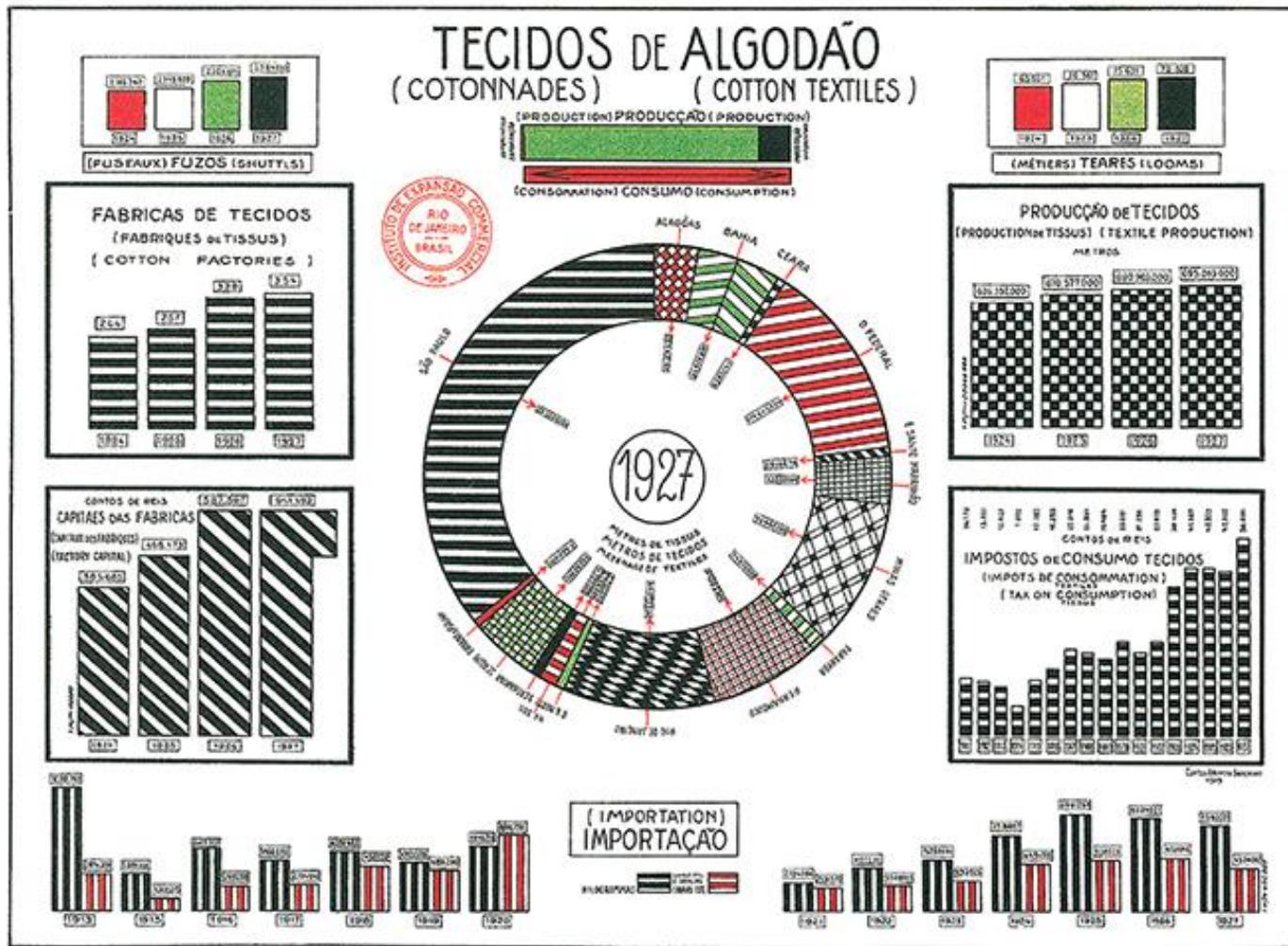


Chart Junk – Dark Gridlines

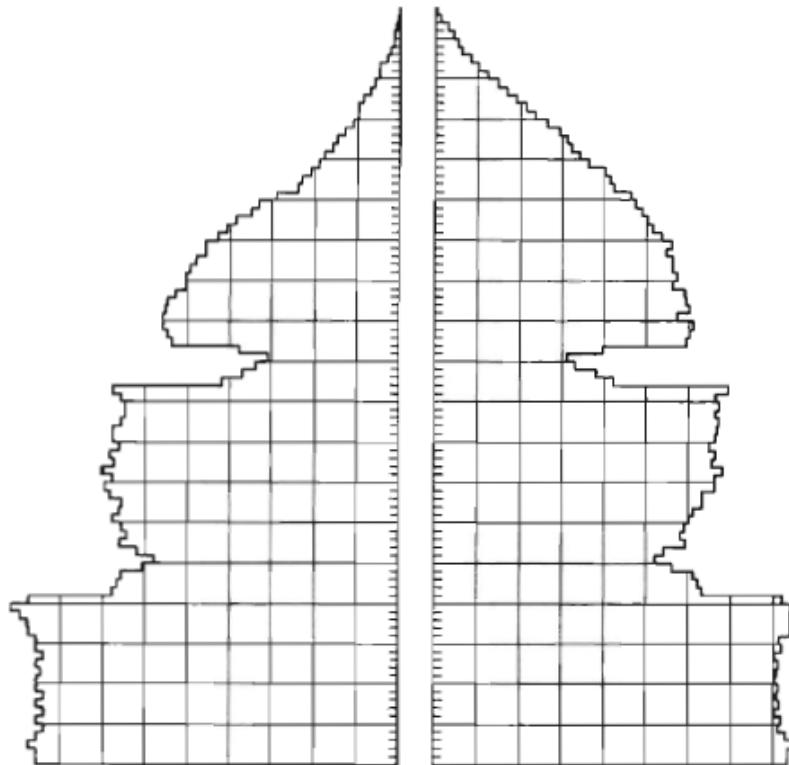
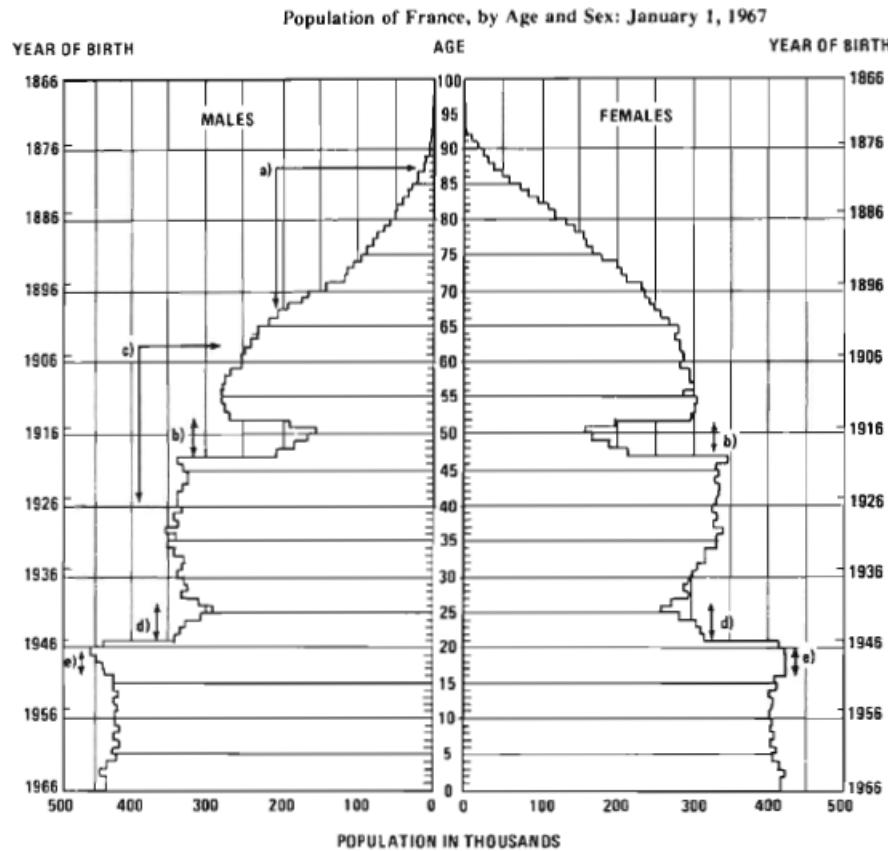


Chart Junk - Gridlines

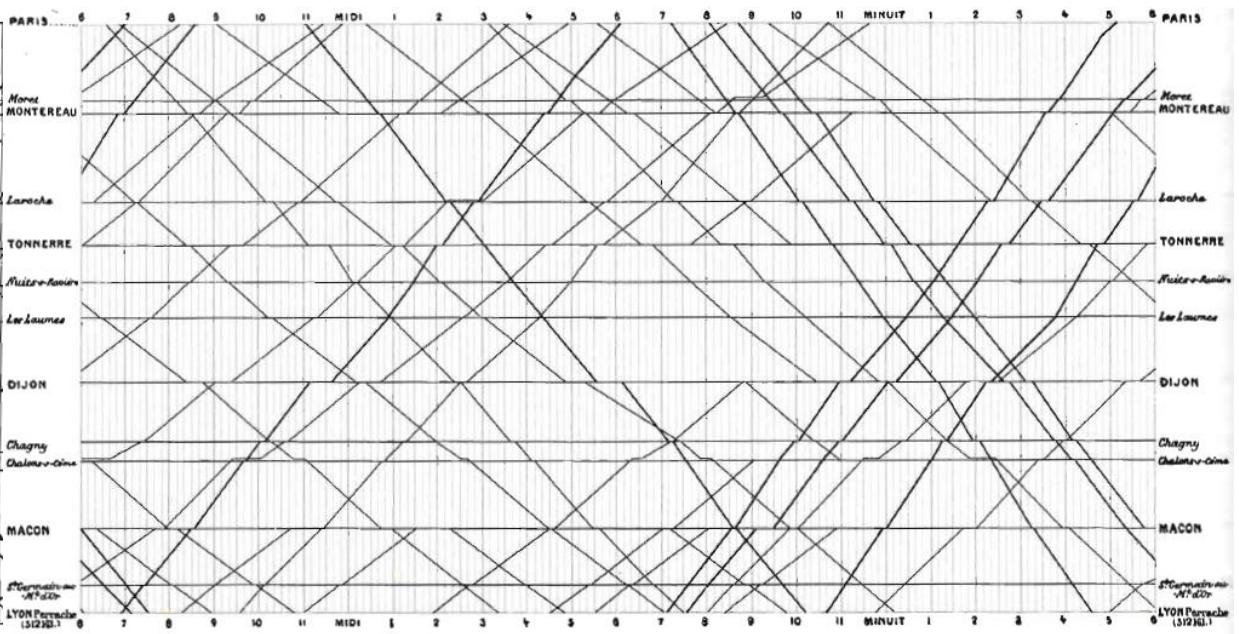
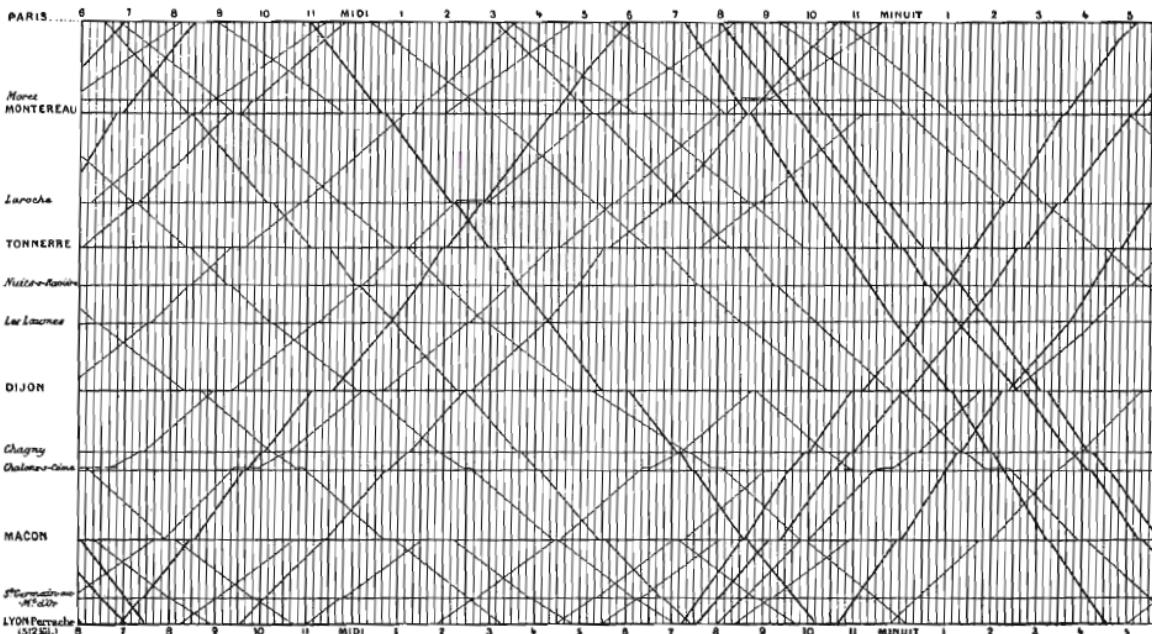


Chart Junk: Decorations

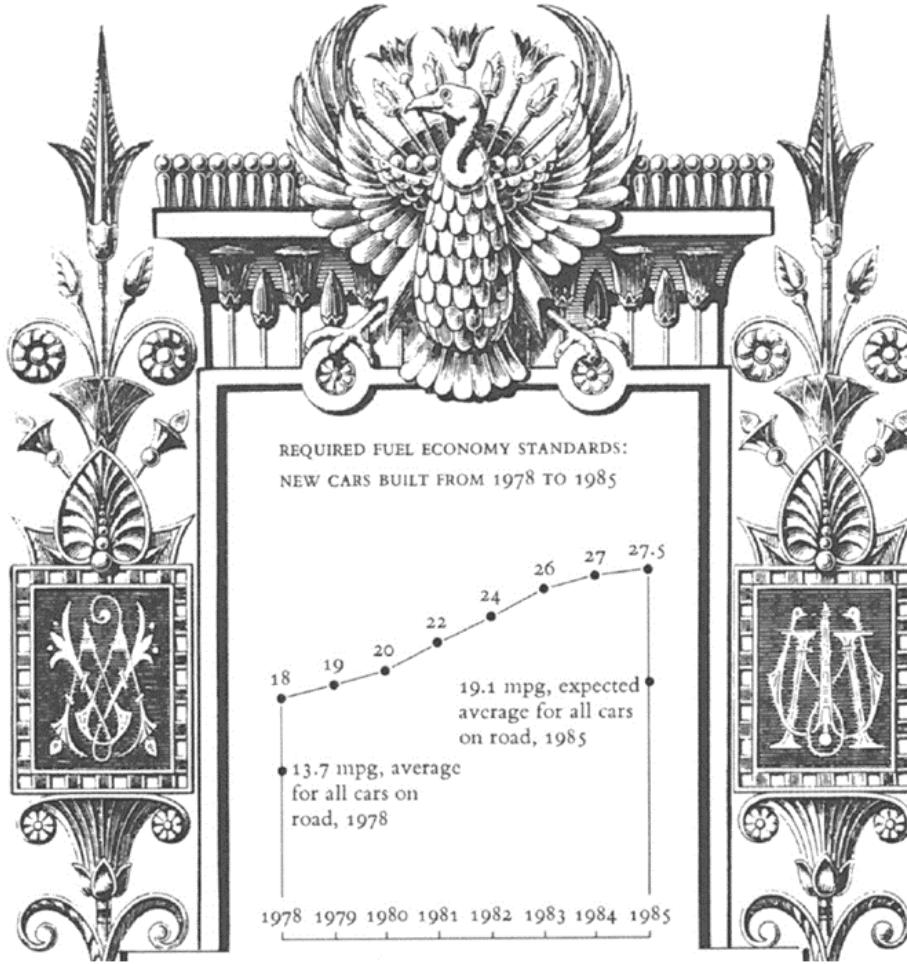


Chart Junk: Decorations

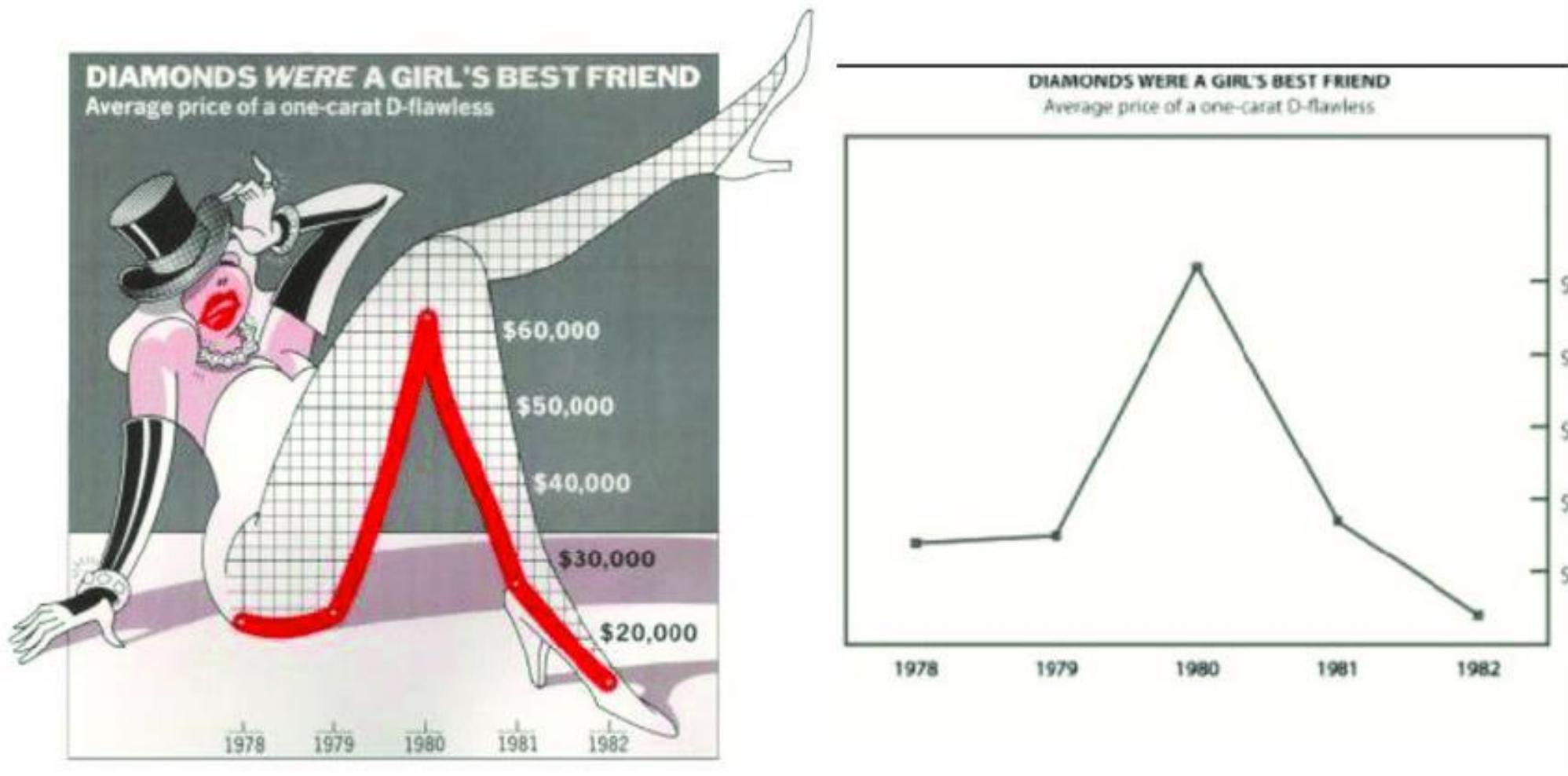
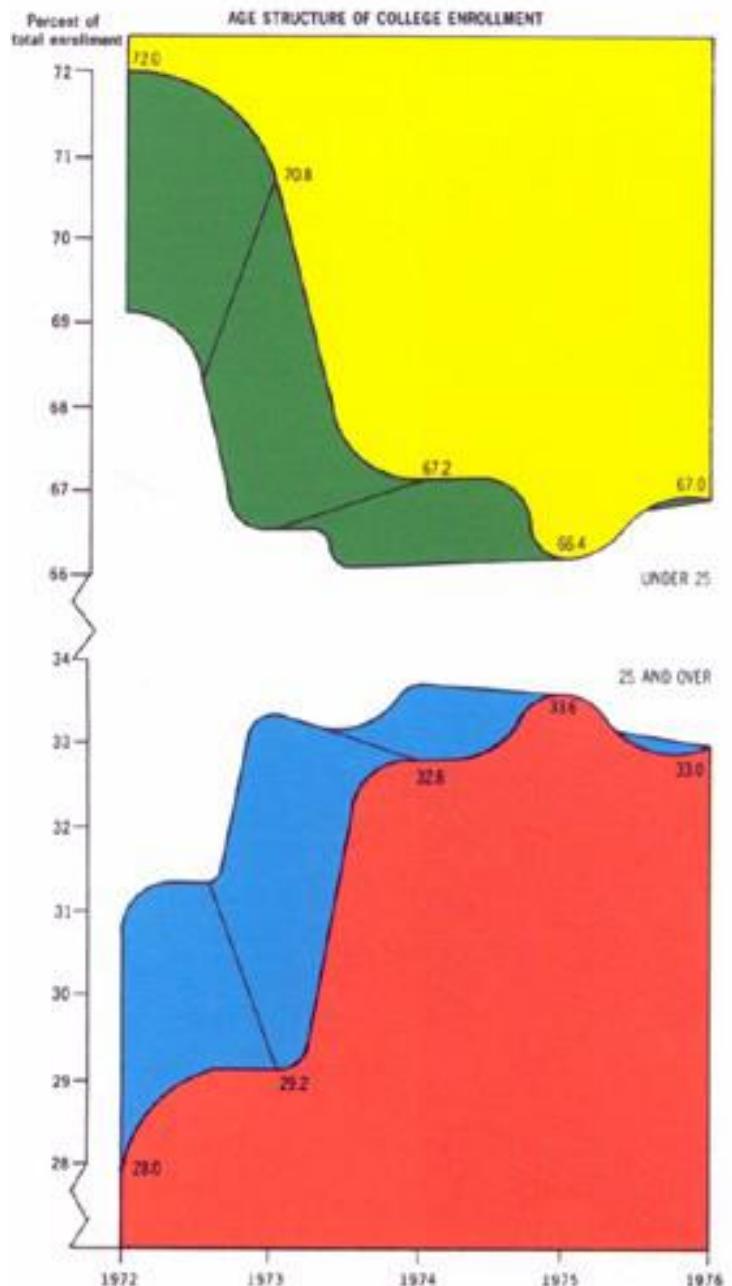


Chart Junk: False Perspective

Tufte: "This may well be the worst graphic ever to find its way into print".

Five colours report five pieces of data.

False perspective distorts data.



Design Strategies

Design strategies

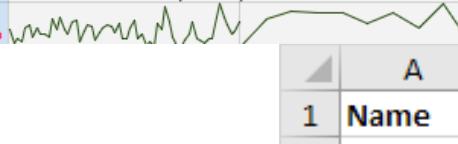
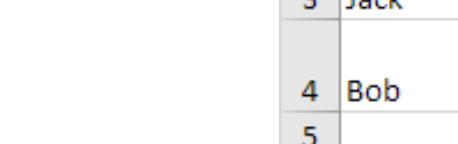
Sparklines

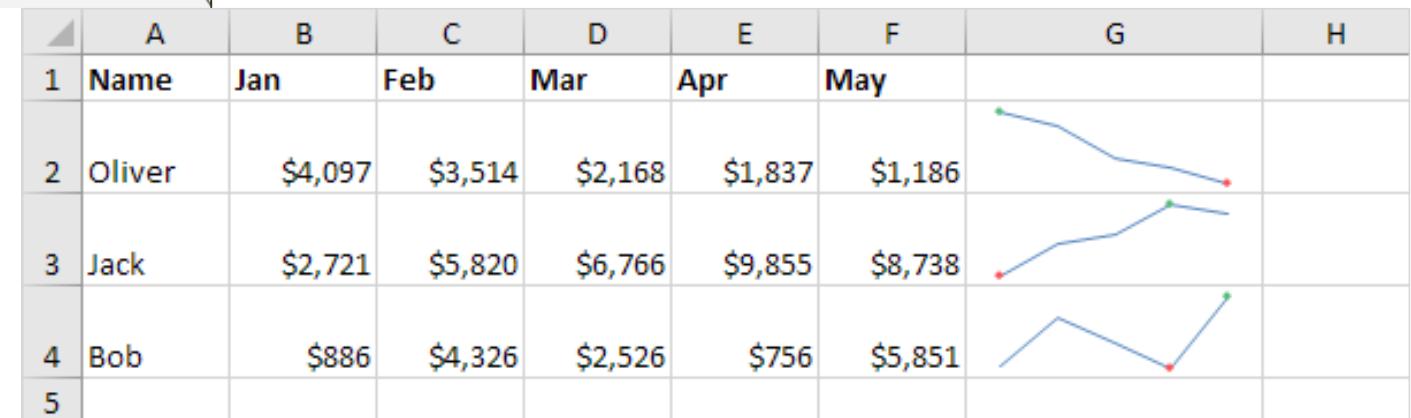
Small multiples

Micro/Macro

Layering

Sparklines

	A	B	C	D	E	F
1	Campaign	Clicks 60daysAgo tot 31daysAgo	Clicks 30daysAgo tot 1daysAgo		Clicks 60daysAgo tot 1daysAgo	Clicks 60daysAgo tot 1daysAgo per week
2		1254	1105	-11.88%		
3		73	45	-38.36%		
4		94	70	-25.53%		
5		145	144	-0.69%		
6		208	203	-2.40%		
7		163	89	-45.40%		
8		428	407	-4.91%		



Sparklines



2,922 Visits



1,694 Absolute Unique Visitors



4,975 Pageviews



1.70 Average Pageviews



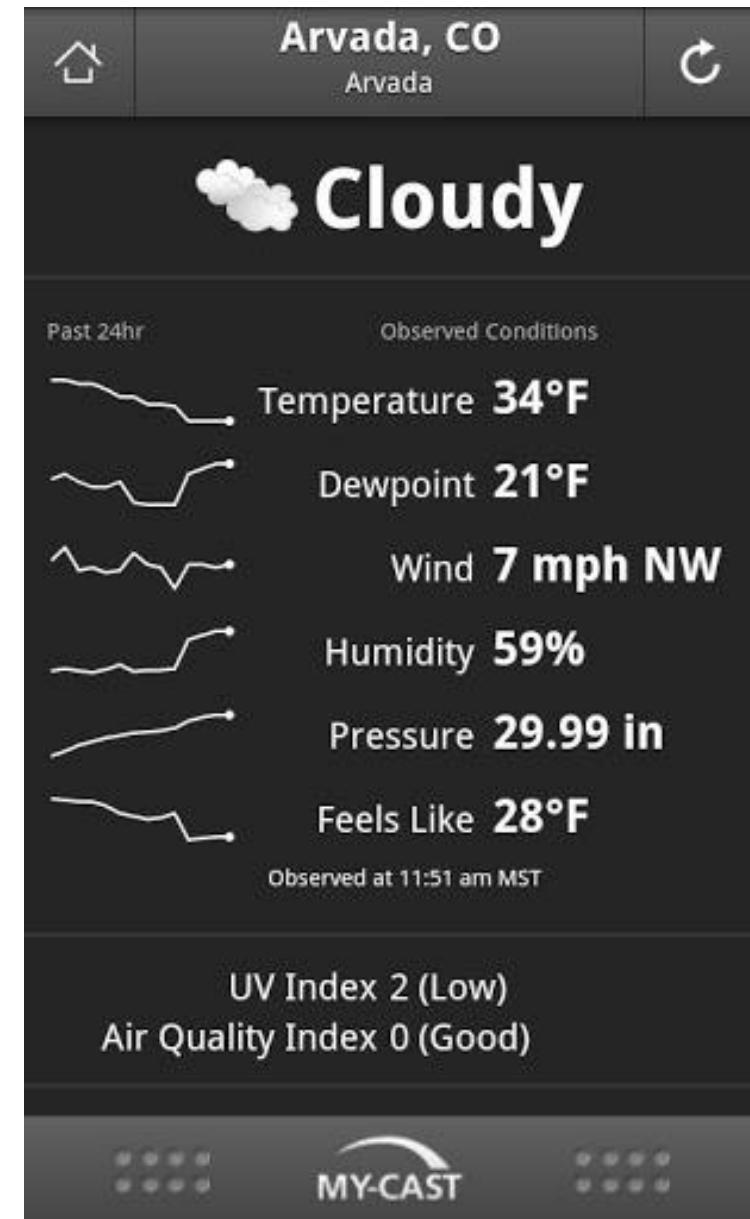
00:02:21 Time on Site



71.59% Bounce Rate

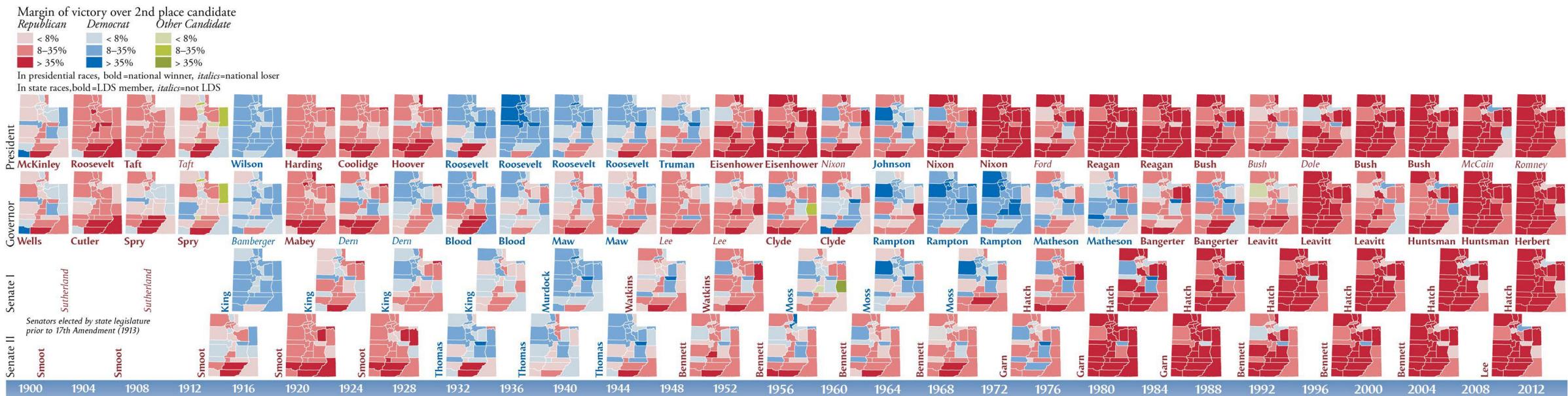


57.97% New Visits

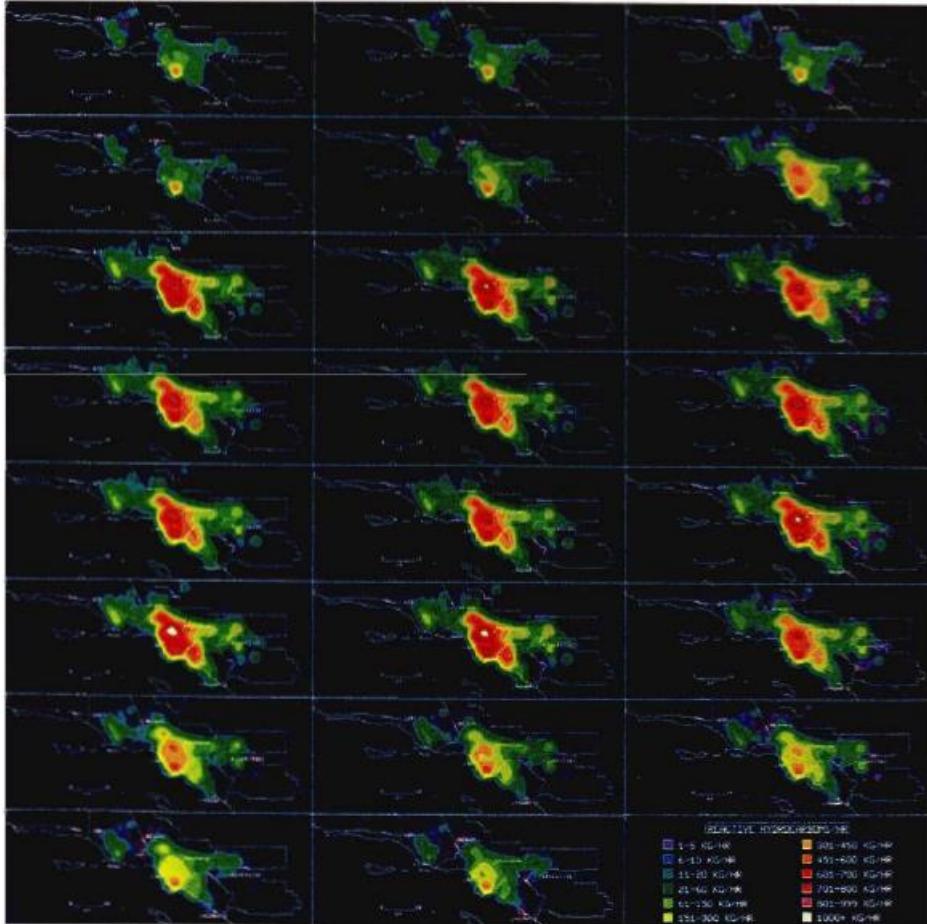


Small multiples

Show multiple, comparable, high-density charts.



Small multiples



Visualisation shows air pollution in Los Angeles for each hour

Micro-Macro

Graphics should be readable in whole images and in parts, and that should help manage levels of detail.



Micro-Macro

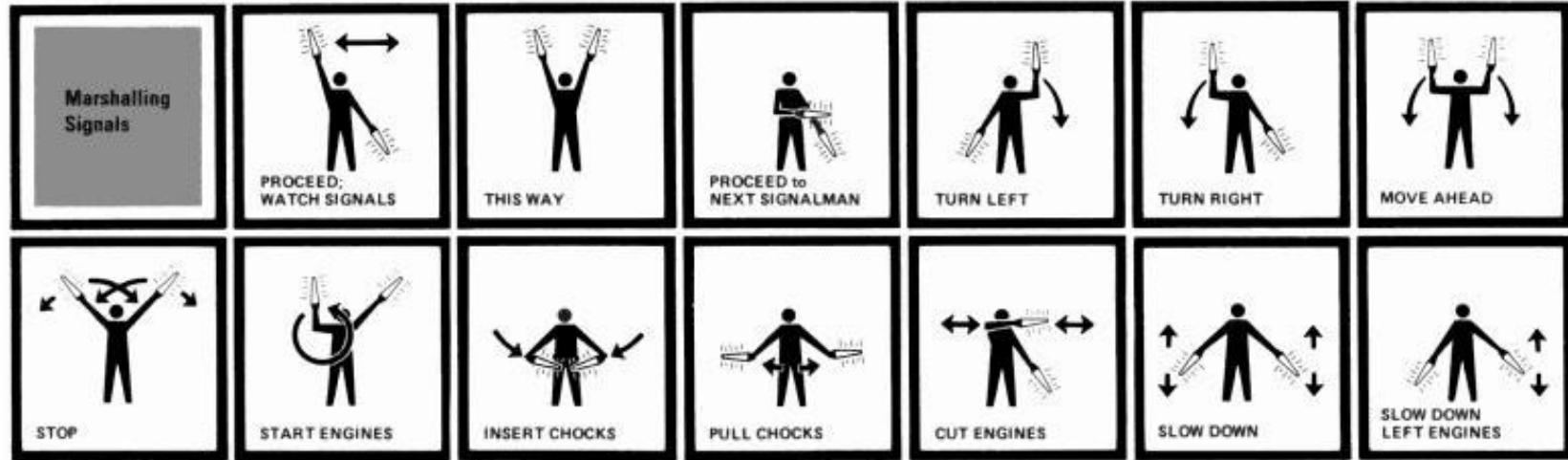
Over 150k hearts

Each representing a life lost to
Covid-19 in the UK

Wall shows the scale of the
tragedy, but also reflects on the
individual victims



Layering

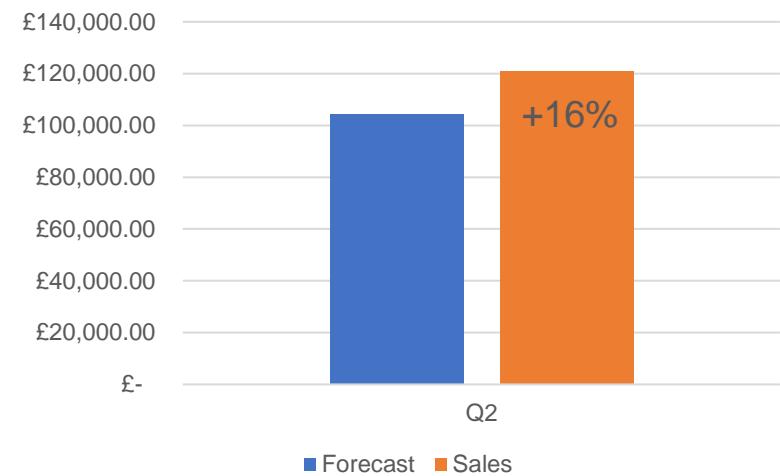


Text, Tables or Graphics

Text, Table or Graphs?

Q2 Sales of £121,065 exceeded the forecast by 16%

Q2 Sales	Compared to forecast
£121,065	+16%



When to use tables?

Need to look up individual values

Used to compare individual values (not series of values)

Precise values needed

Multiple units of measure

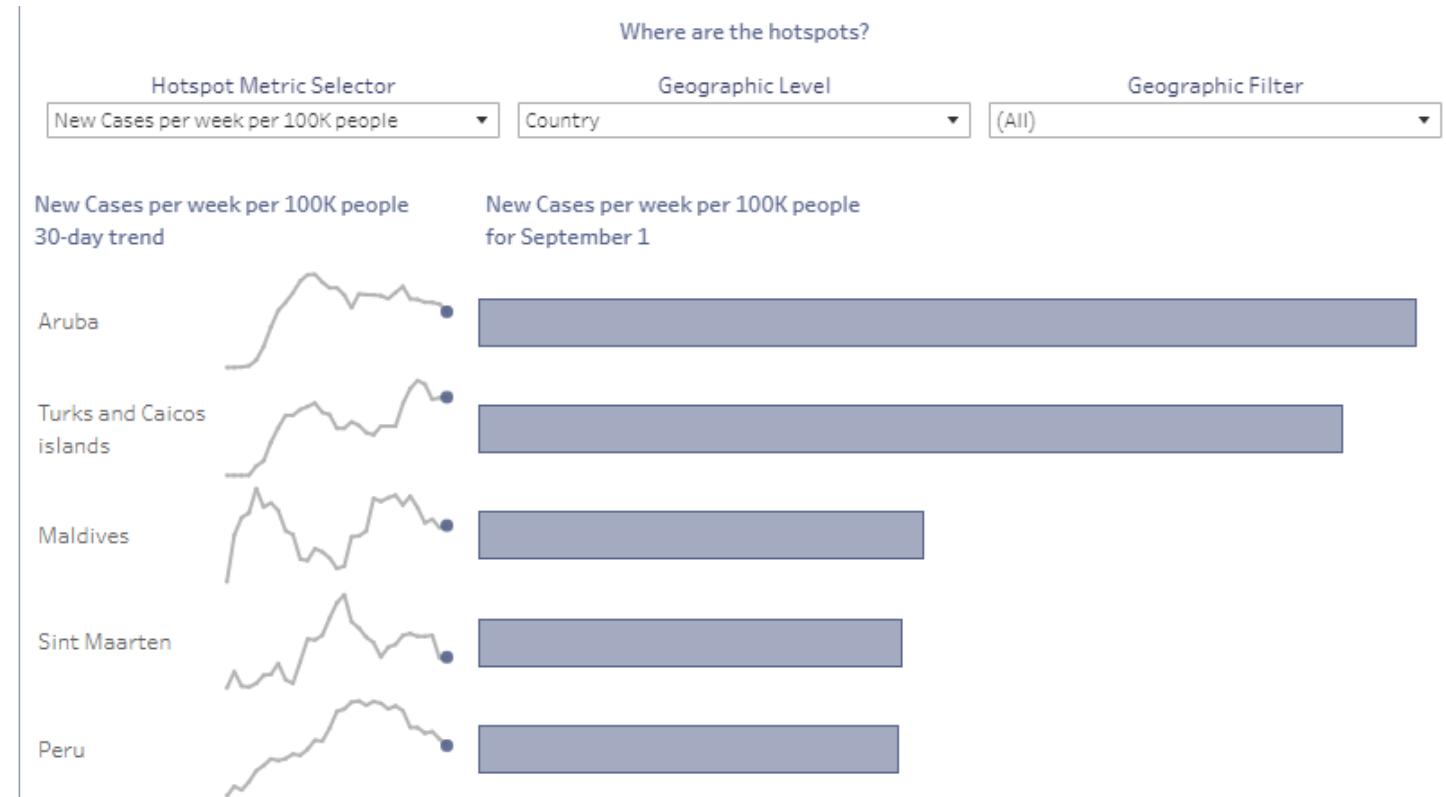
Needs summary and detail

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
2018	164	164	165	166	166	166	167	167	168	168	168	168	167
2019	175	176	176	177	178	178	178	178	178	178	177	177	177
2020	177	178	179	180	180	180	180	181	181	181	181	181	180

When to use Graphs?

Need to look for patterns, trends, or exceptions

Need to reveal relationships among whole sets of values



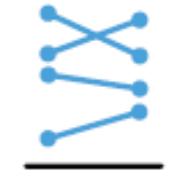
Which graph to use?



Comparison



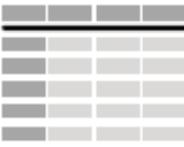
Bar & Column



Slopegraph



Dumbbell

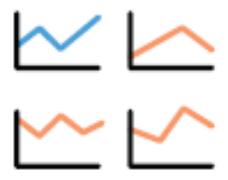


Table

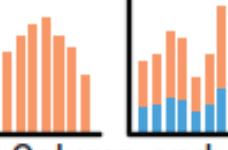
Trend



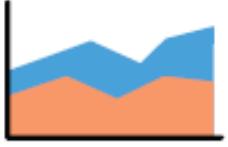
Line



Small Multiple (Line)



Column and Stacked Column



Stacked Area

Composition



Waffle



100% Stacked Column



Pie and Donut



Treemap

Relationship



Scatterplot



Bubble

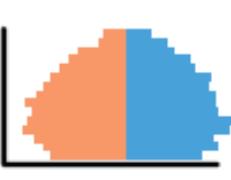


XY Heatmap

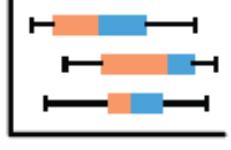
Distribution



Histogram



Pyramid



Box plot



Cloud

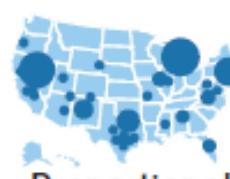
Spatial



Choropleth



Dot Density



Proportional Symbol

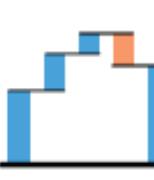


Heatmap

Flow



Funnel



Waterfall

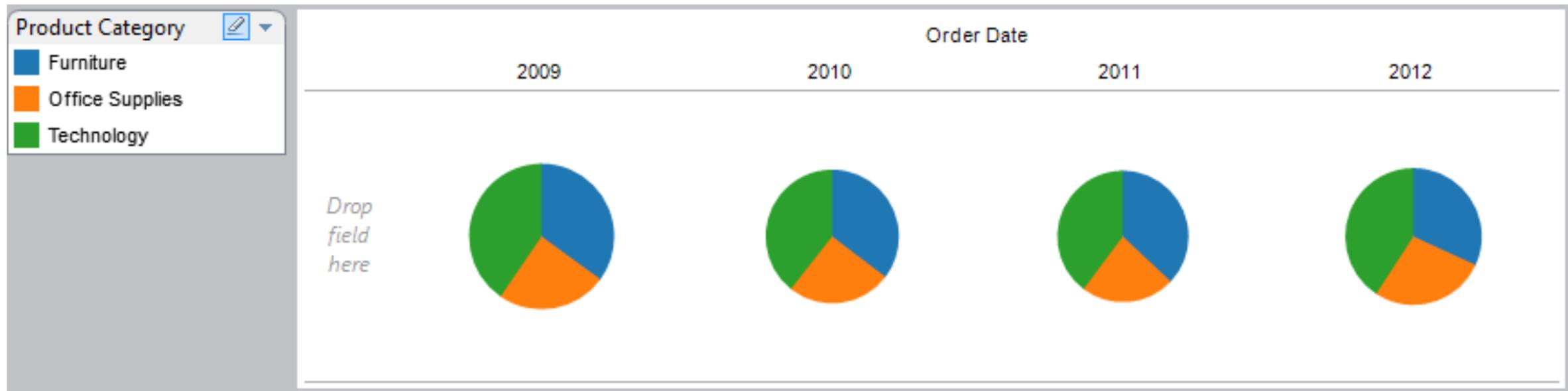


Sankey

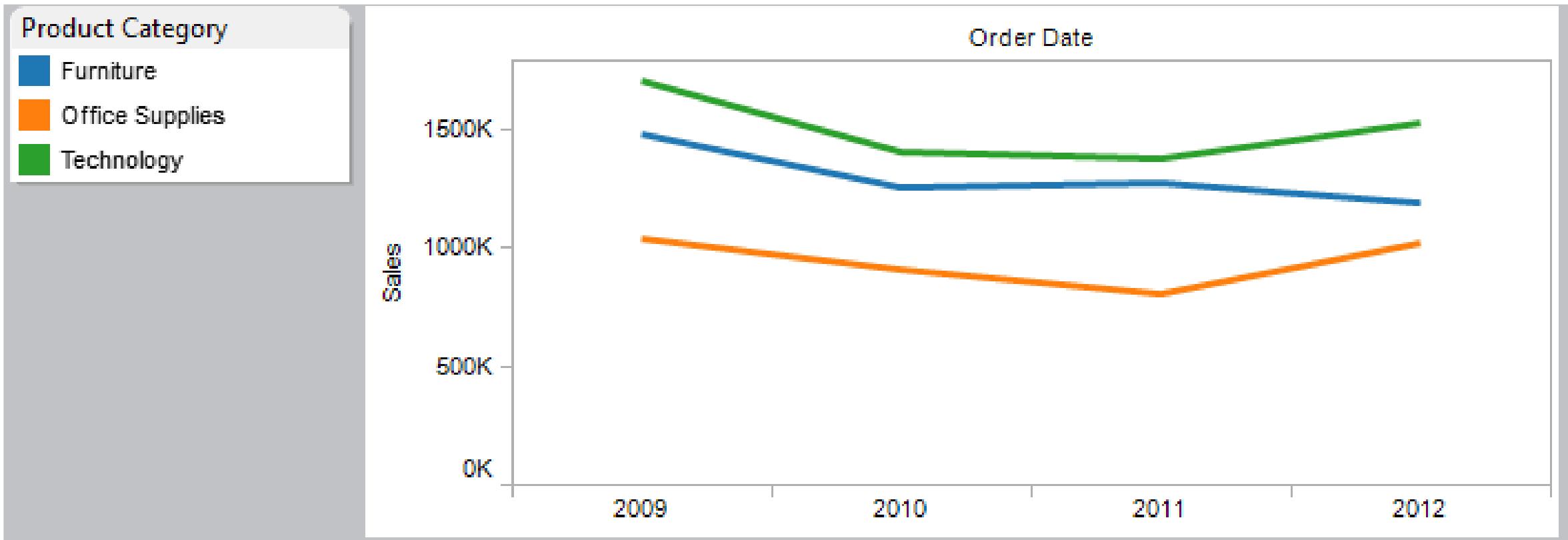


Chord

What category is trending up?



What category is trending up?



Some perceptual tasks are easier than others...

Nominal (qualitative)

1. Position
2. Colour (Hue)
3. Pattern (Texture)
4. Connection/Edge
5. Containment
6. Pattern (Density)
7. Colour (Lightness)
8. Symbol/Shape
9. Size (Length)
10. Angle/Slope
11. Size (Area)
12. Size (Volume)

Ordinal (qualitative)

1. Position
2. Pattern (Density)
3. Colour (Lightness)
4. Colour (Hue)
5. Pattern (Texture)
6. Connection/Edge
7. Containment
8. Size (Length)
9. Angle/Slope
10. Size (Area)
11. Size (Volume)
12. Symbol/Shape

Interval, Ratio (quantitative)

1. Position
2. Size (Length)
3. Angle/Slope
4. Size (Area)
5. Size (Volume)
6. Pattern (Density)
7. Colour (Lightness)
8. Colour (Hue)
9. Pattern (Texture)
10. Connection/Edge
11. Containment
12. Symbol/Shape

Encoding quantitative values

Points

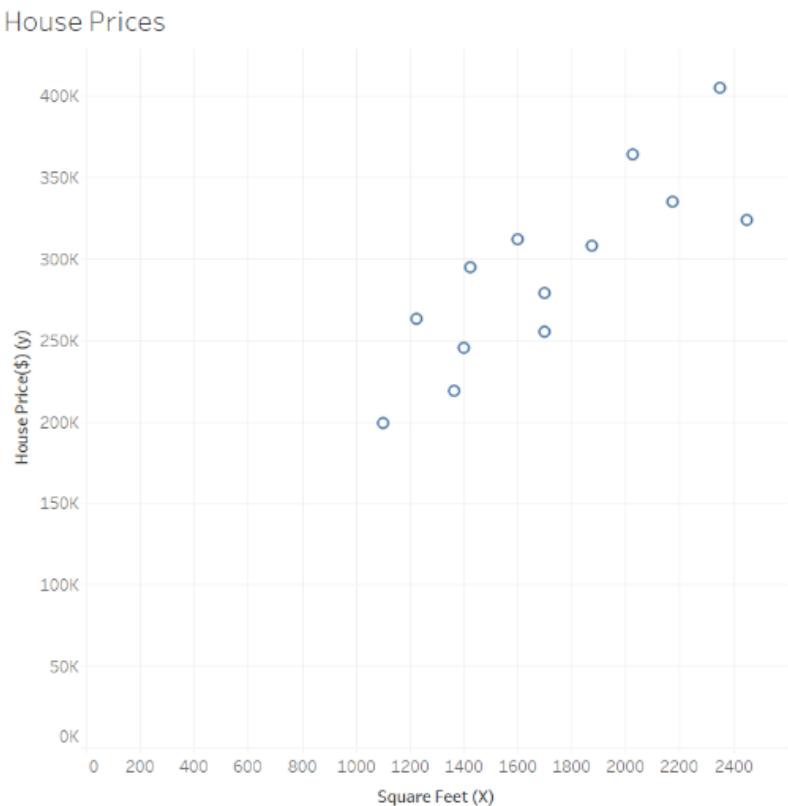
Lines

Bars

Boxes

Shapes with 2-D areas

Shapes with color intensity



Encoding quantitative values

Points

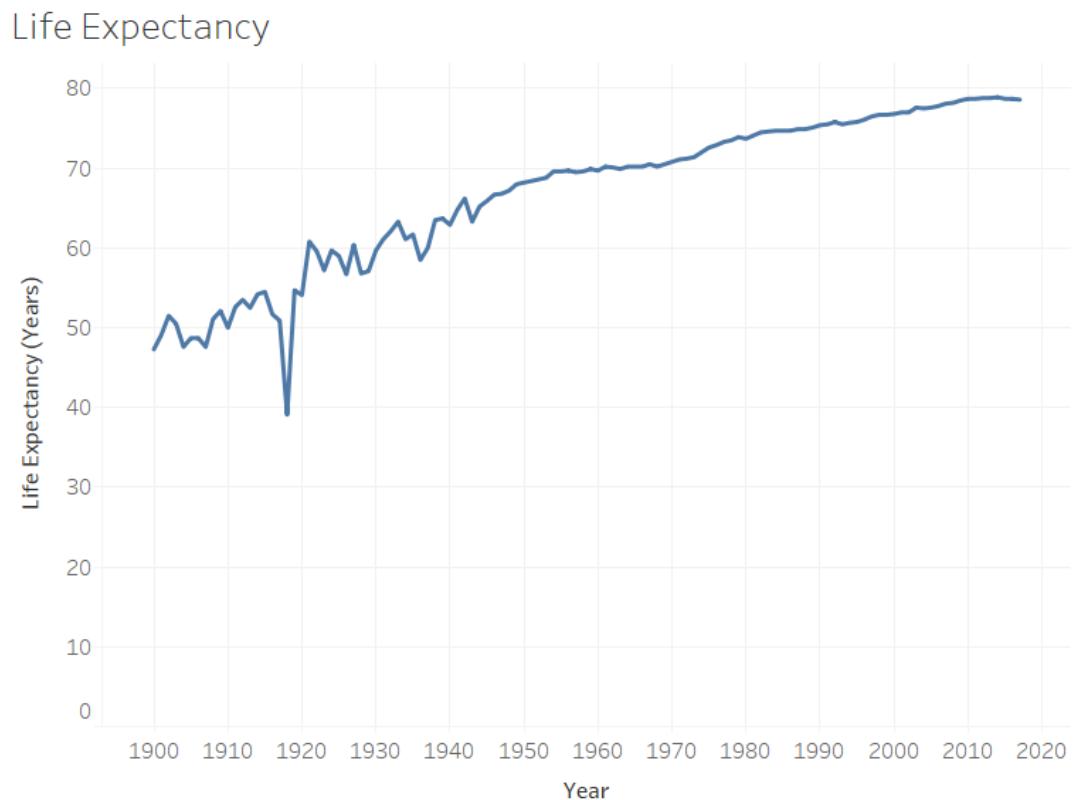
Lines

Bars

Boxes

Shapes with 2-D areas

Shapes with color intensity



Encoding quantitative values

Points

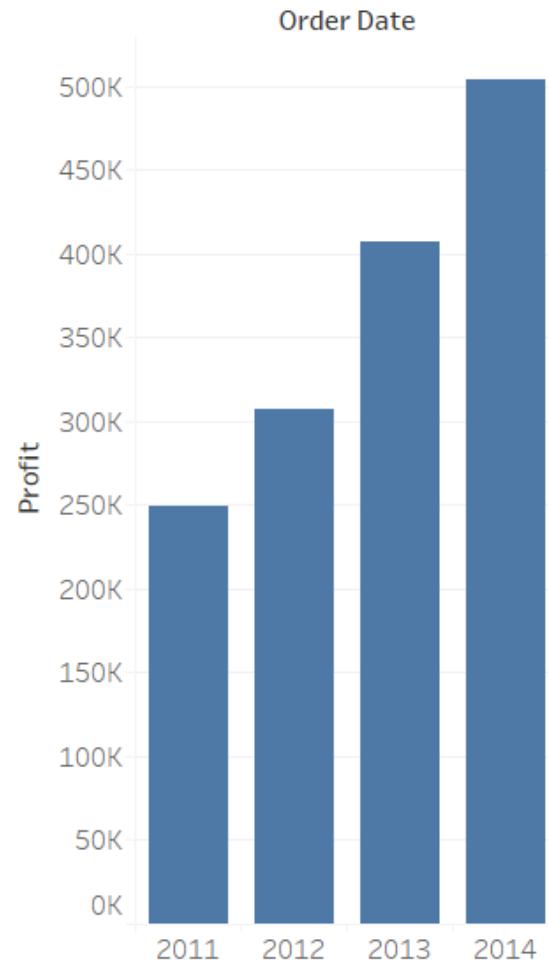
Lines

Bars

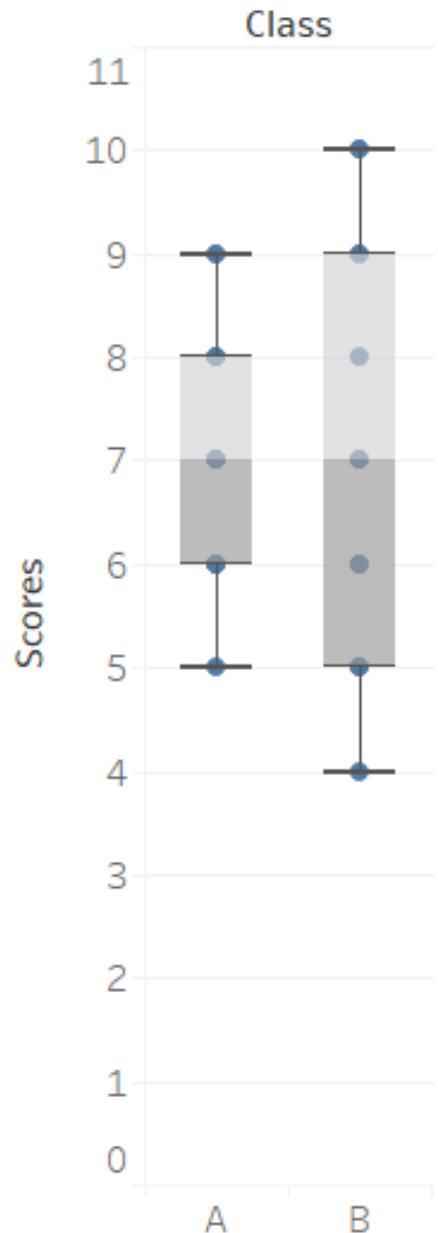
Boxes

Shapes with 2-D areas

Shapes with color intensity



Quiz Scores



Encoding quantitative values

Points

Lines

Bars

Boxes

Shapes with 2-D areas

Shapes with color intensity

Encoding quantitative values

Points

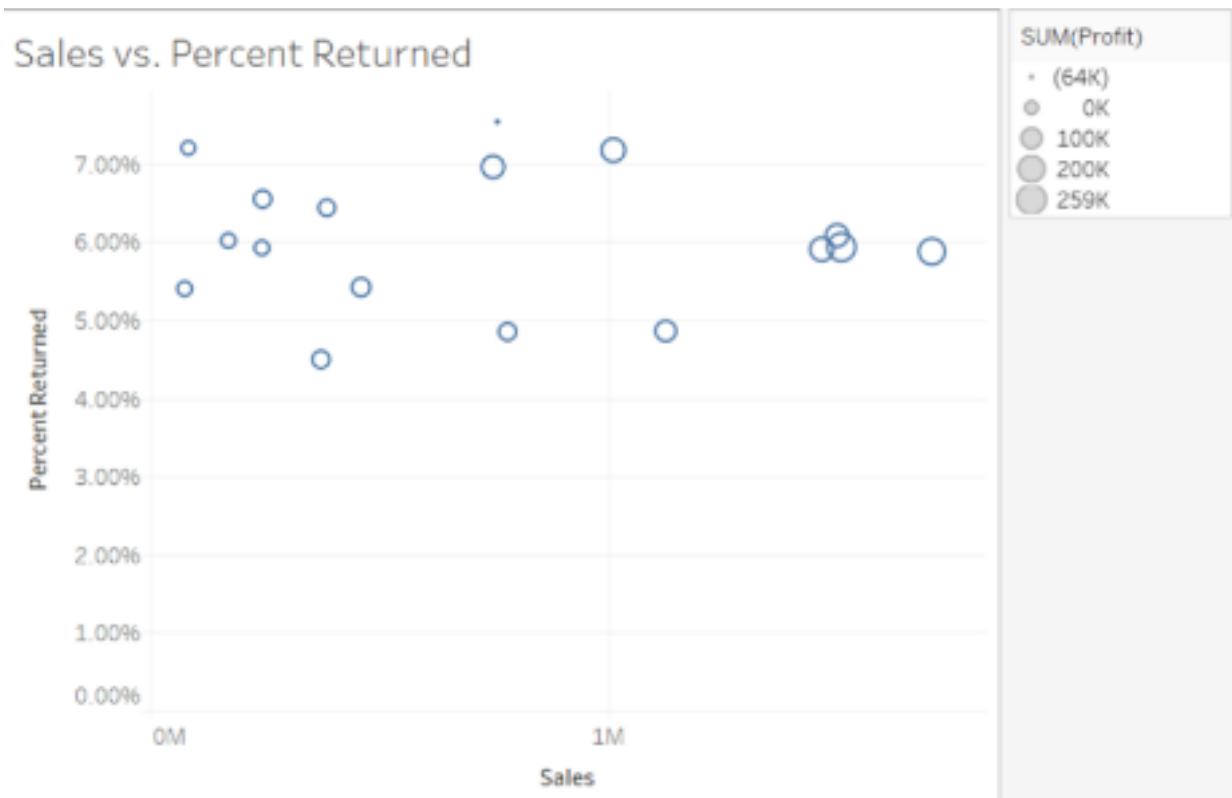
Lines

Bars

Boxes

Shapes with 2-D areas

Shapes with color intensity



Encoding quantitative values

Points

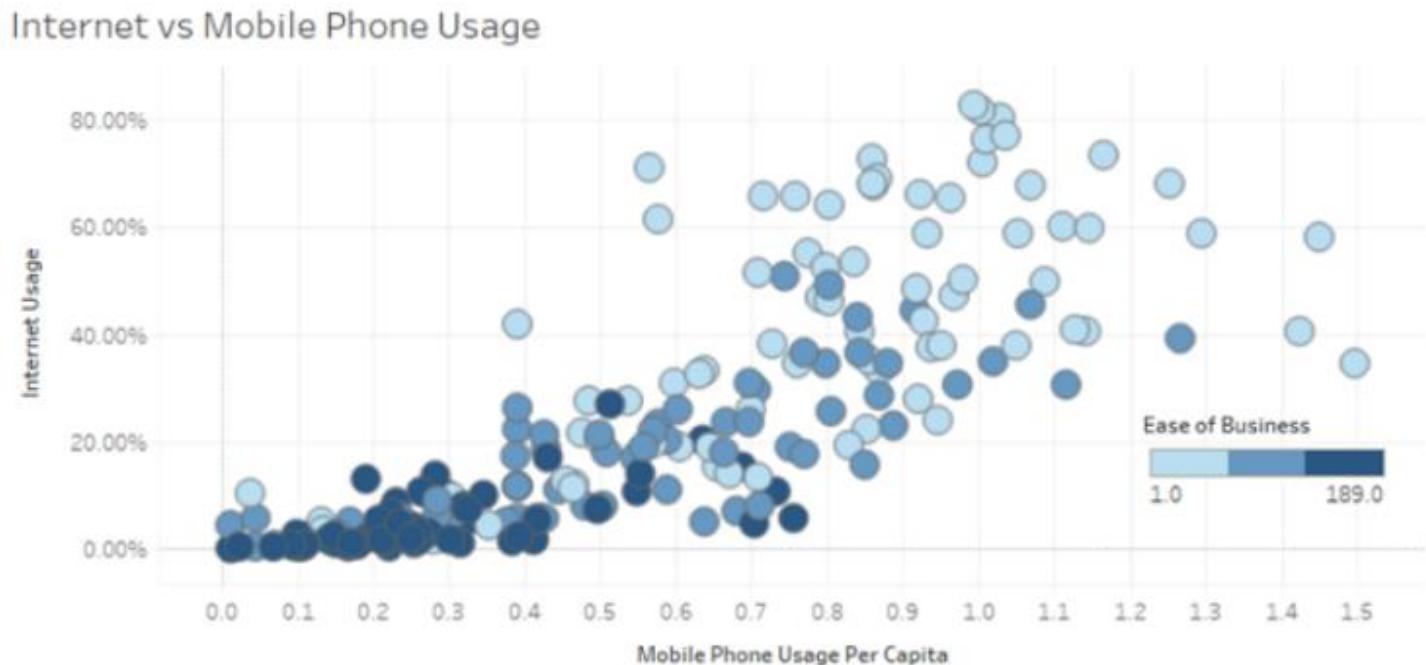
Lines

Bars

Boxes

Shapes with 2-D areas

Shapes with color intensity

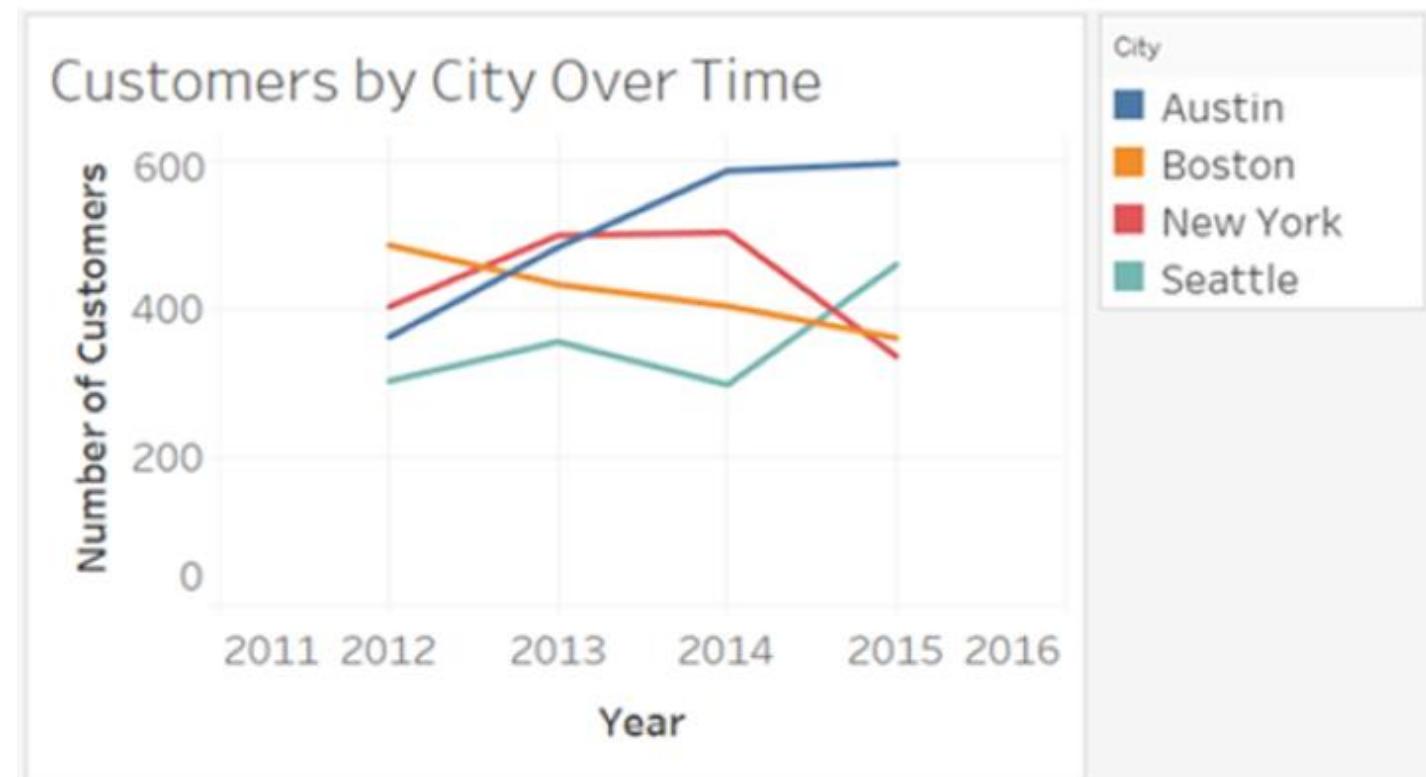


Encoding categorical items

Hue

Shape

Position

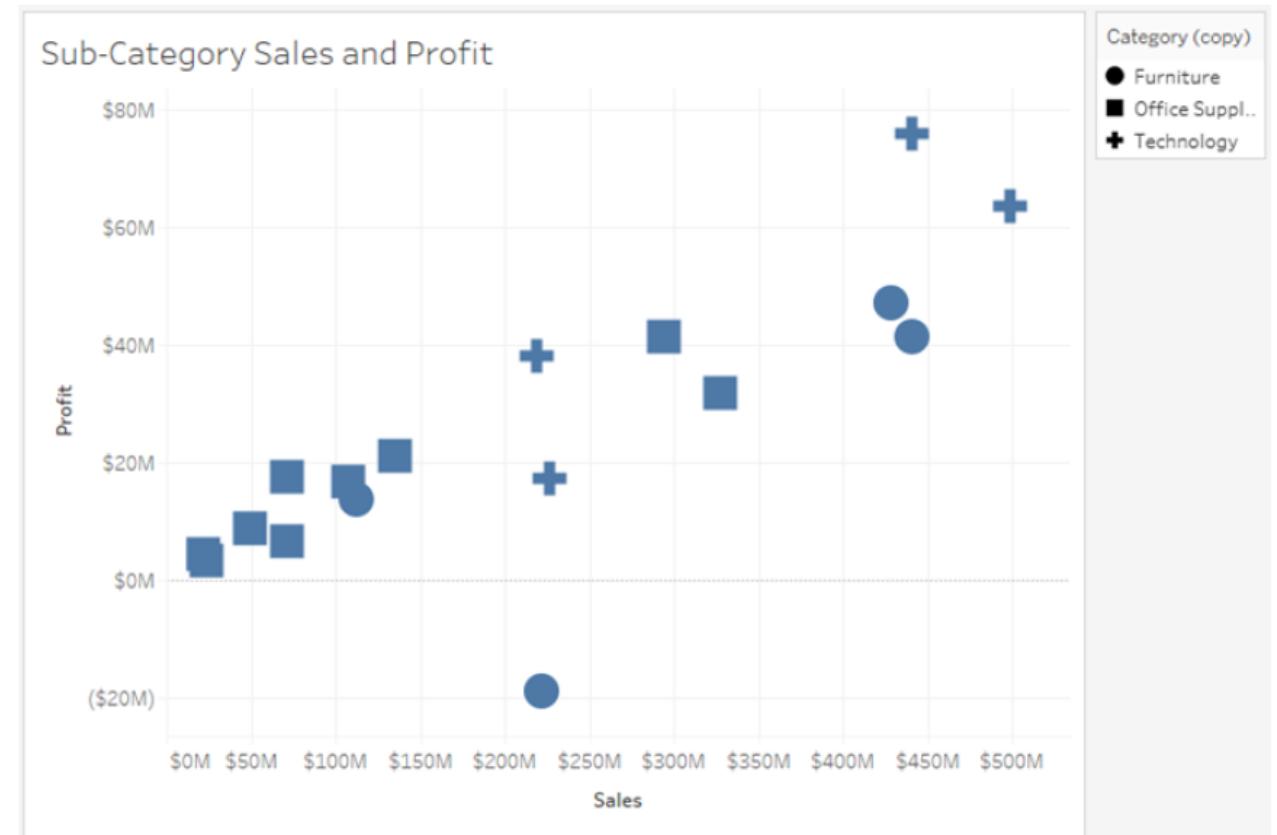


Encoding categorical items

Hue

Shape

Position

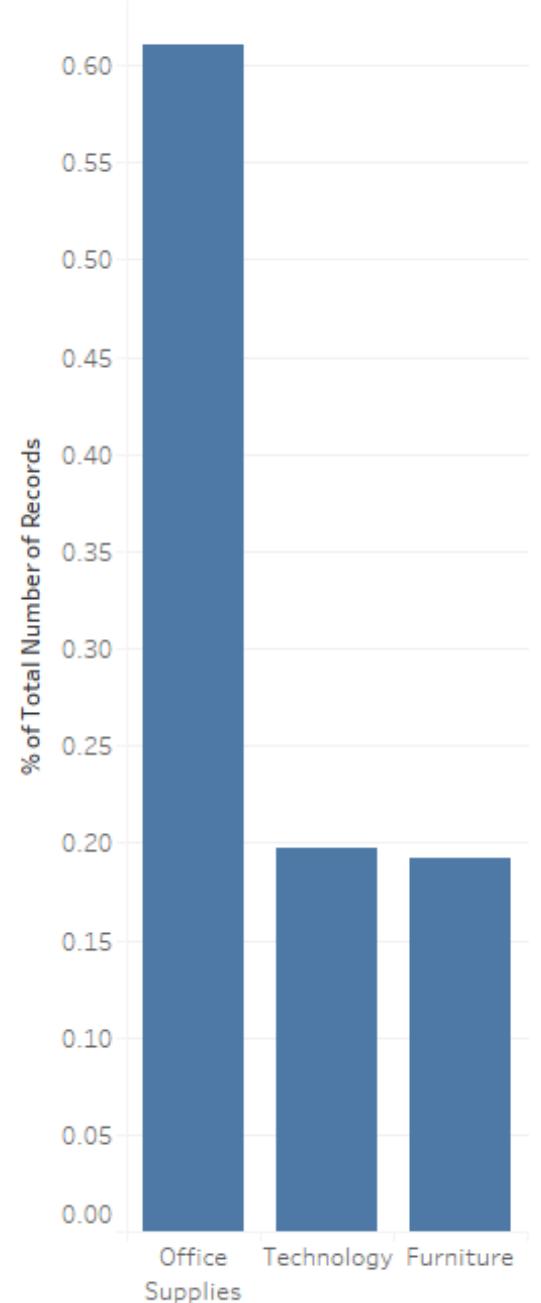


Encoding categorical items

Hue

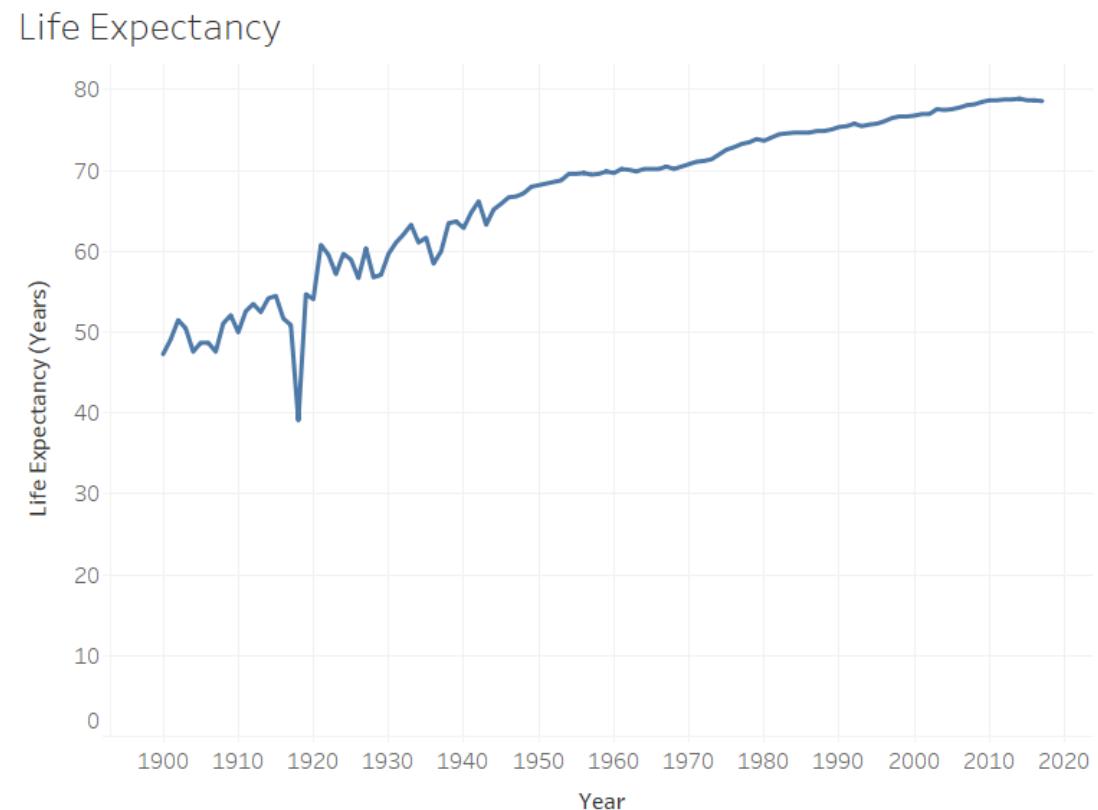
Shape

Position



Time Series

“A series of quantitative values that feature how something changed with time, such as by year, quarter, month, week, day, or hour” (*Show Me the Numbers, page 102*)

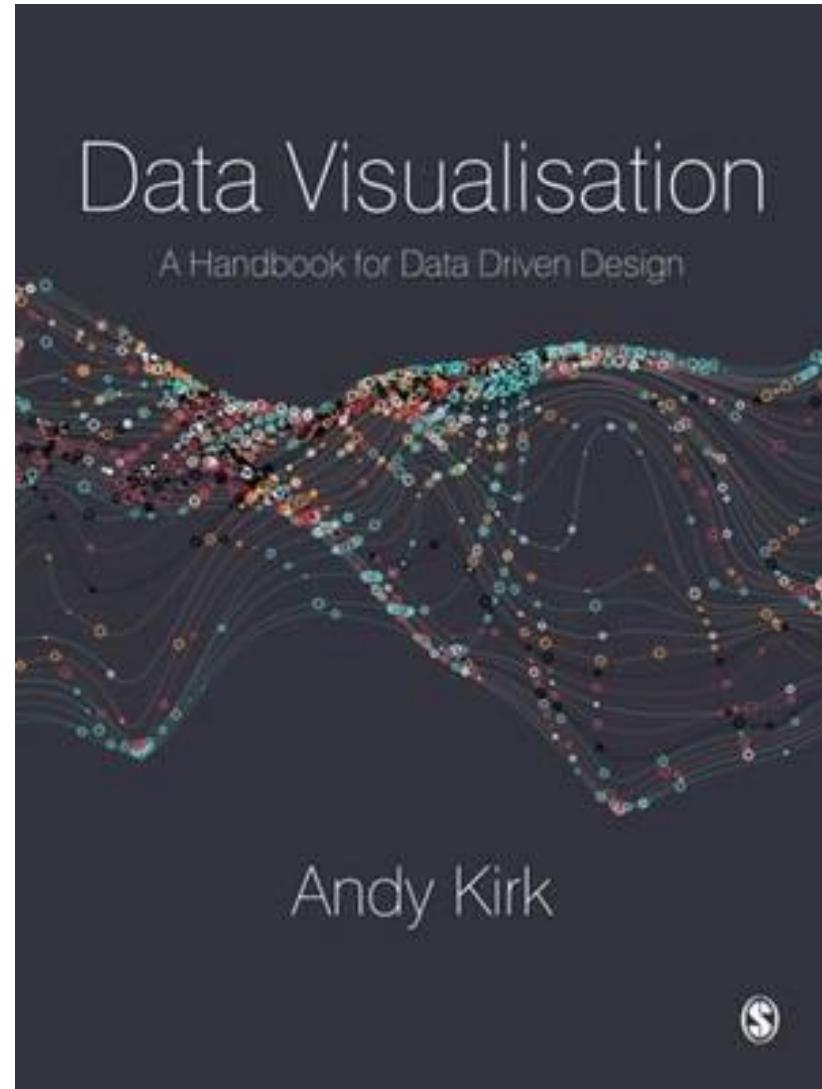


Further reading

Data visualisation :
a handbook for data driven design

Kirk, Andy, author.; 2016

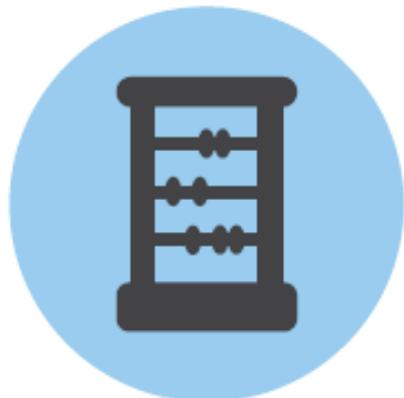
Los Angeles : SAGE



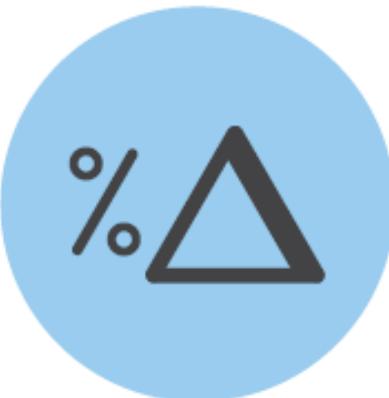
What data to show?

Data variations to consider

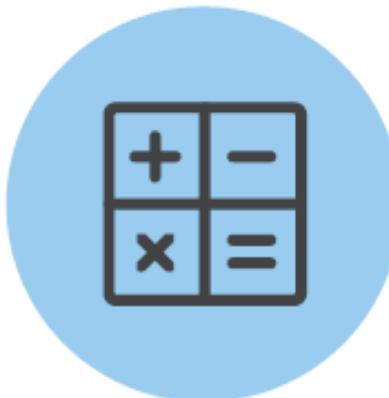
Totals



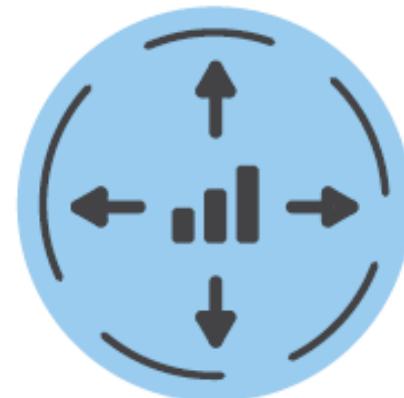
% Change



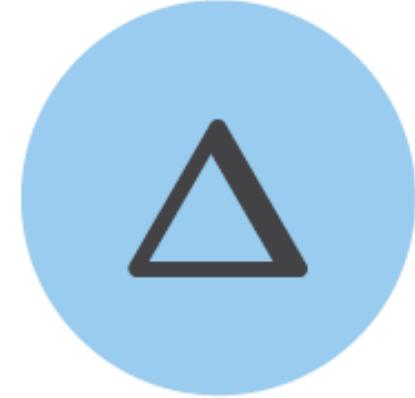
Calculated Metric (Ratio)



Added Context



Variance

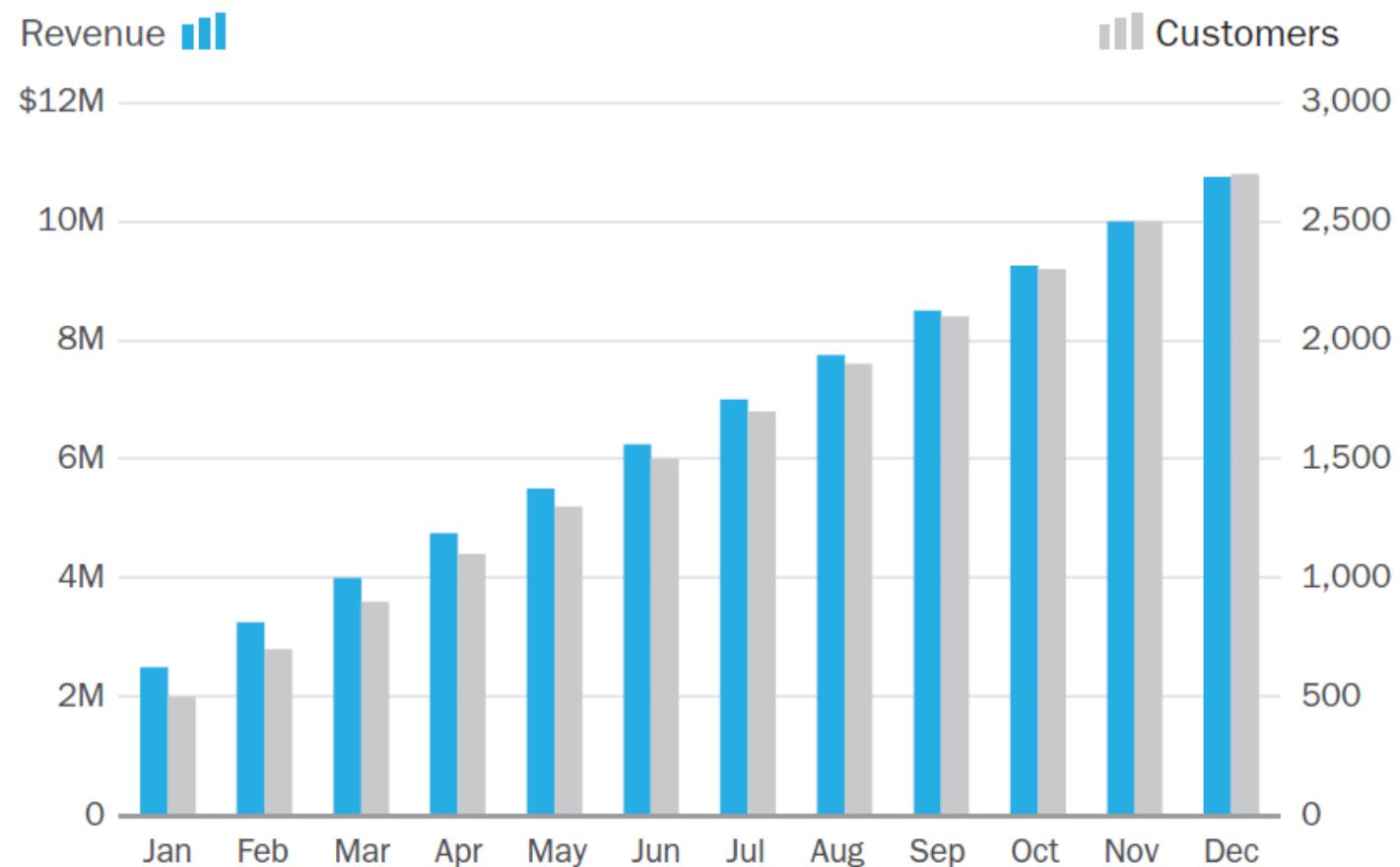


Total values may not communicate
you points as effectively

Dual axis chart

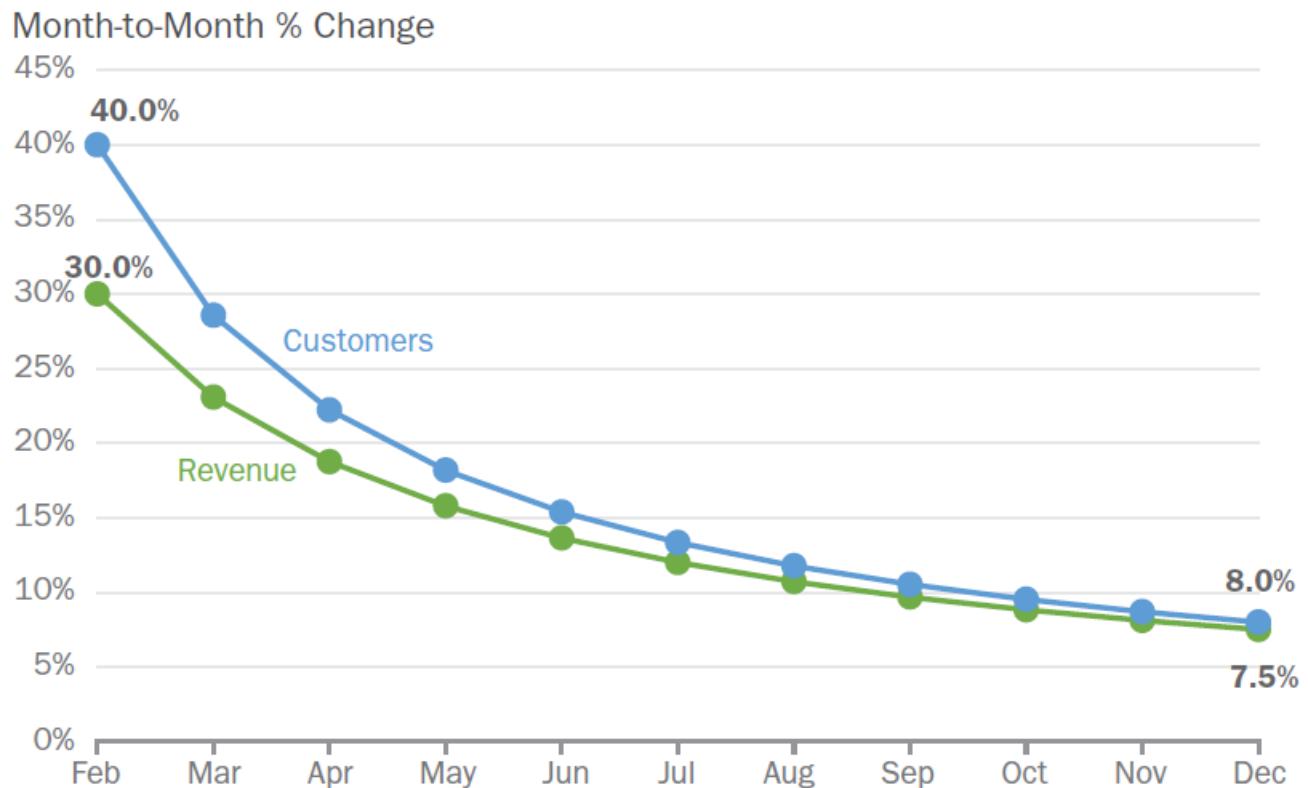
Both values growing

However, revenue isn't growing as fast



Show Percent Change

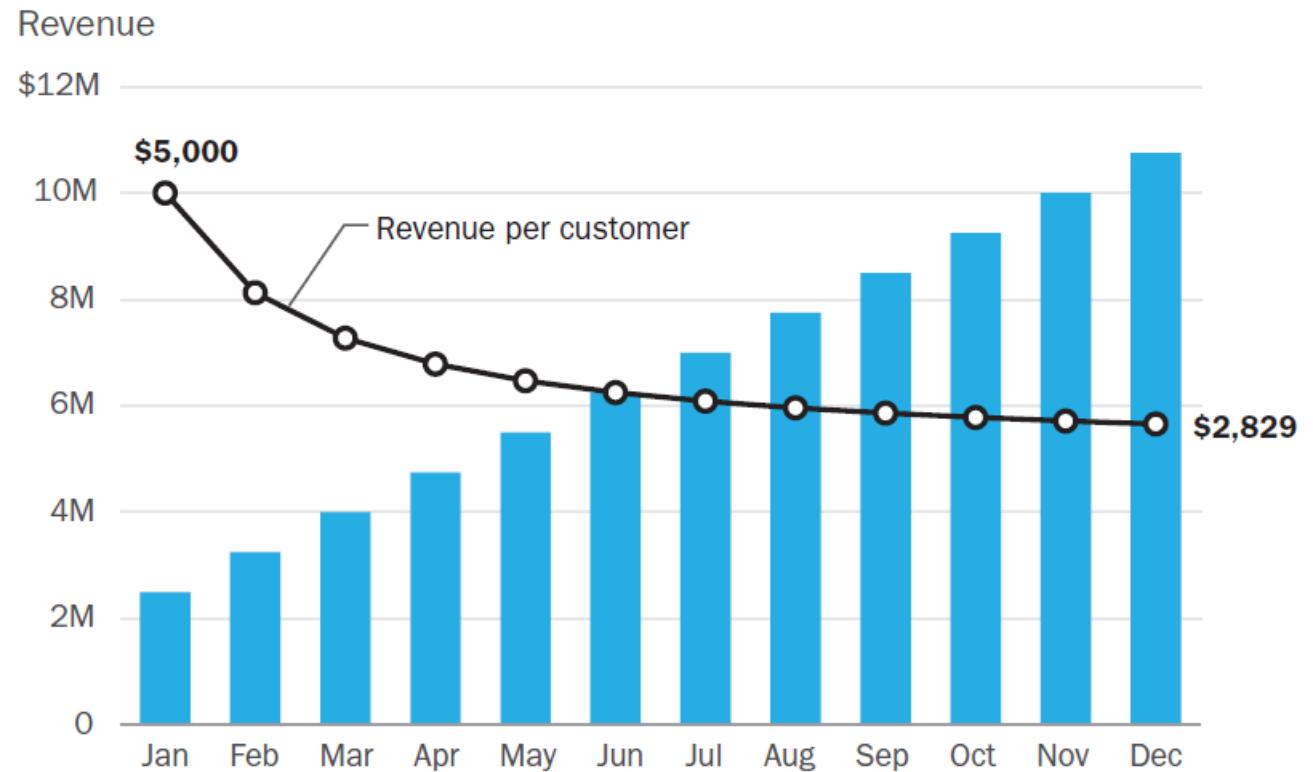
- Puts different metrics on the same %-axis
- Easier to compare
- Needs (some) explanation



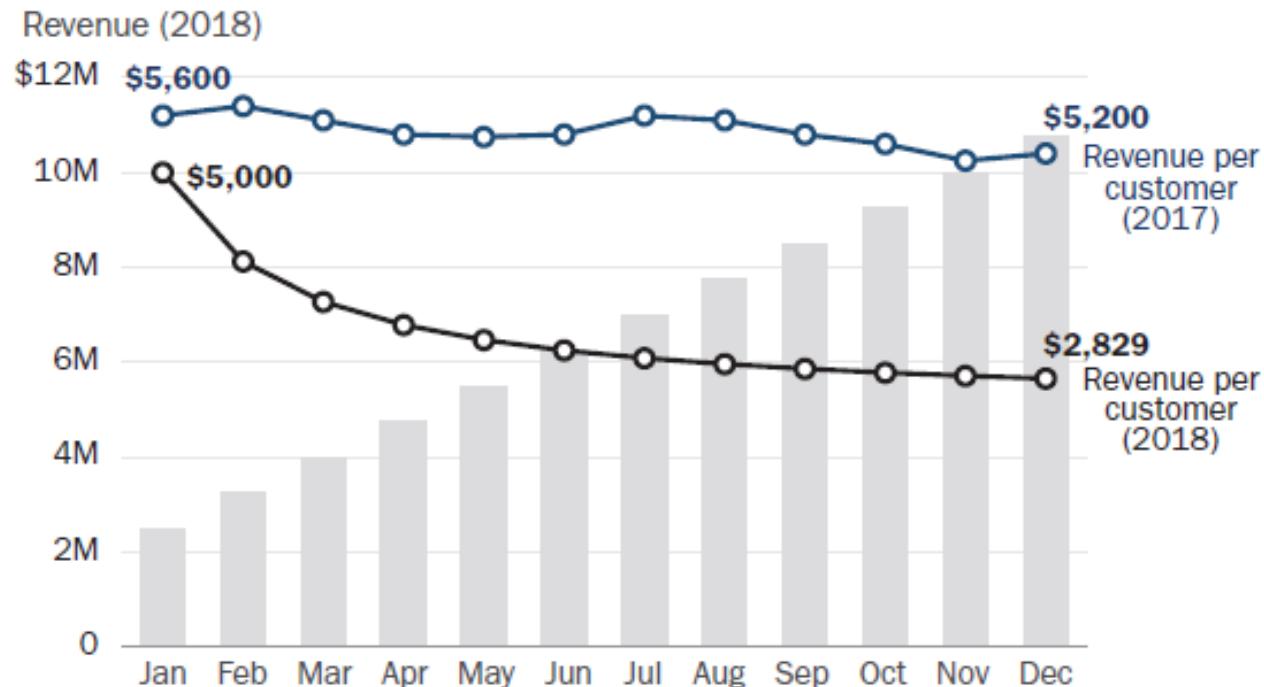
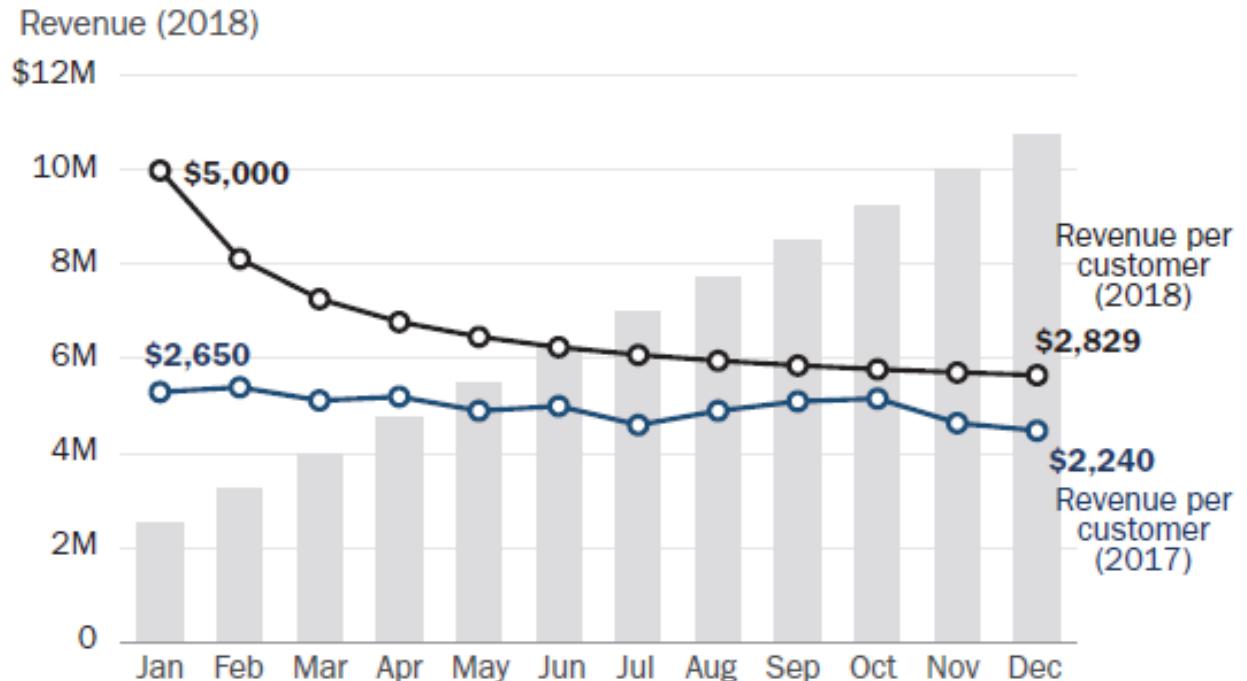
Calculated metrics may help

Combine absolute axis with percent change

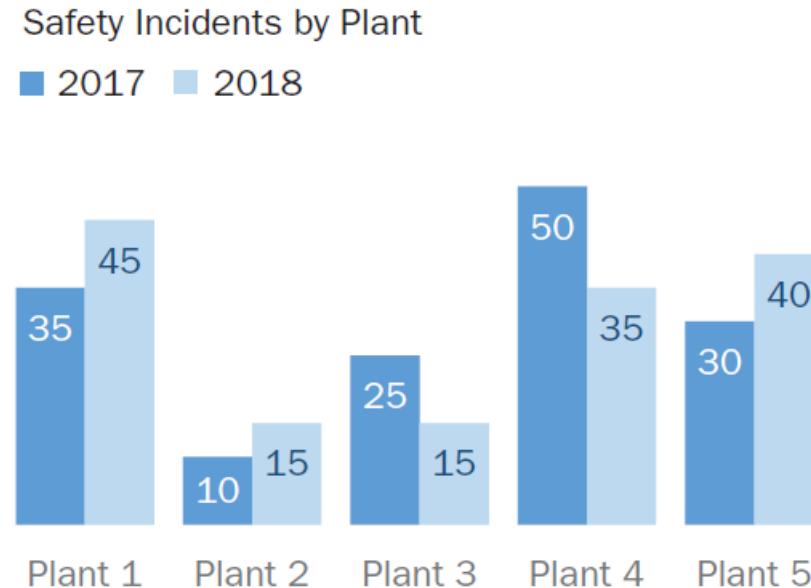
Company is acquiring more customers that spend less



Added context can change message



Variance: Do the maths for your audience



Year-to-Year Variance in Safety Incidents by Plant (2017–2018)



Showing variance can emphasise key point

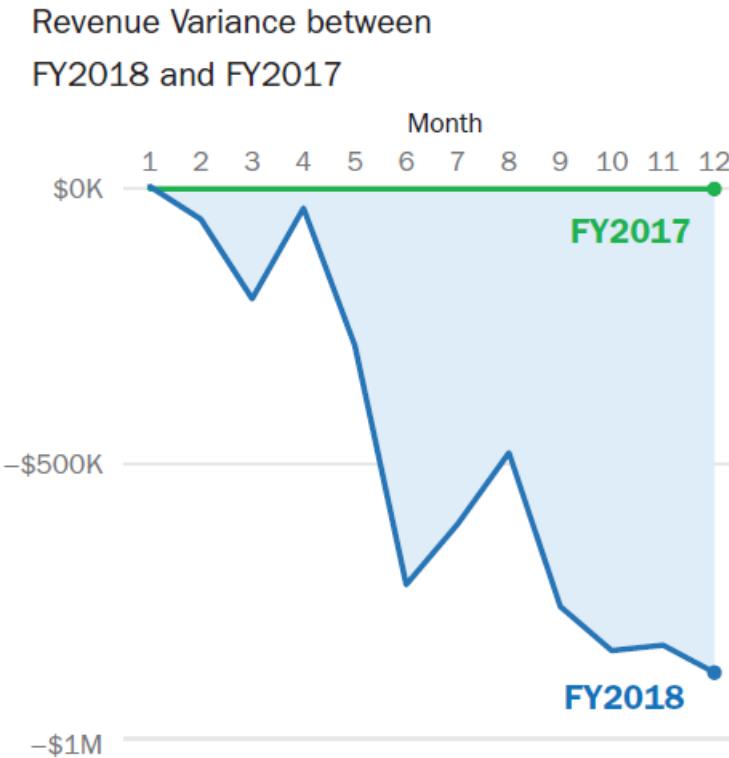
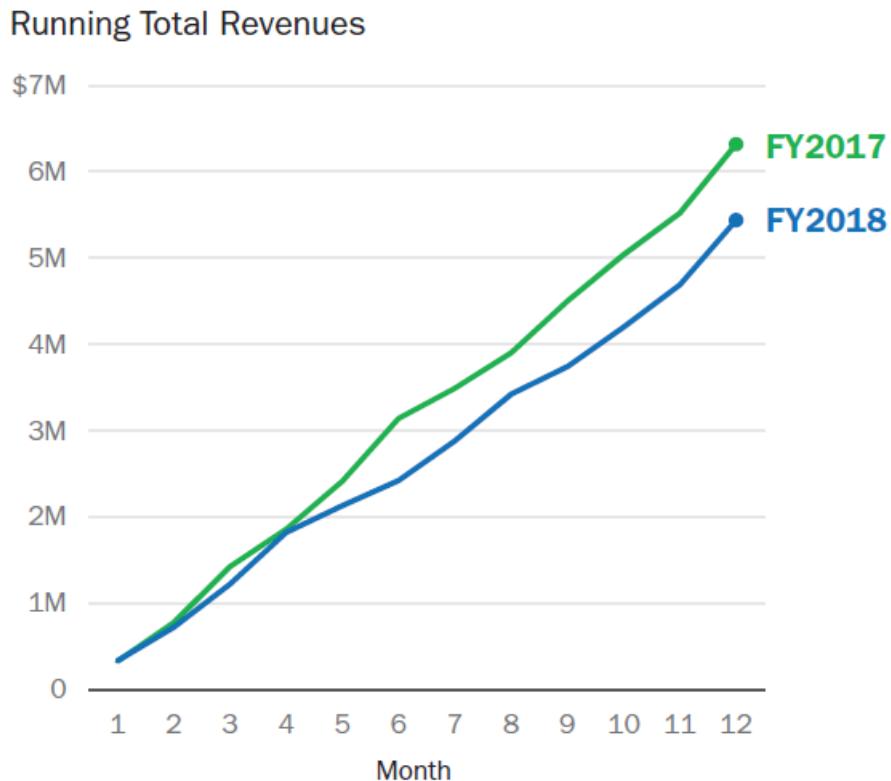
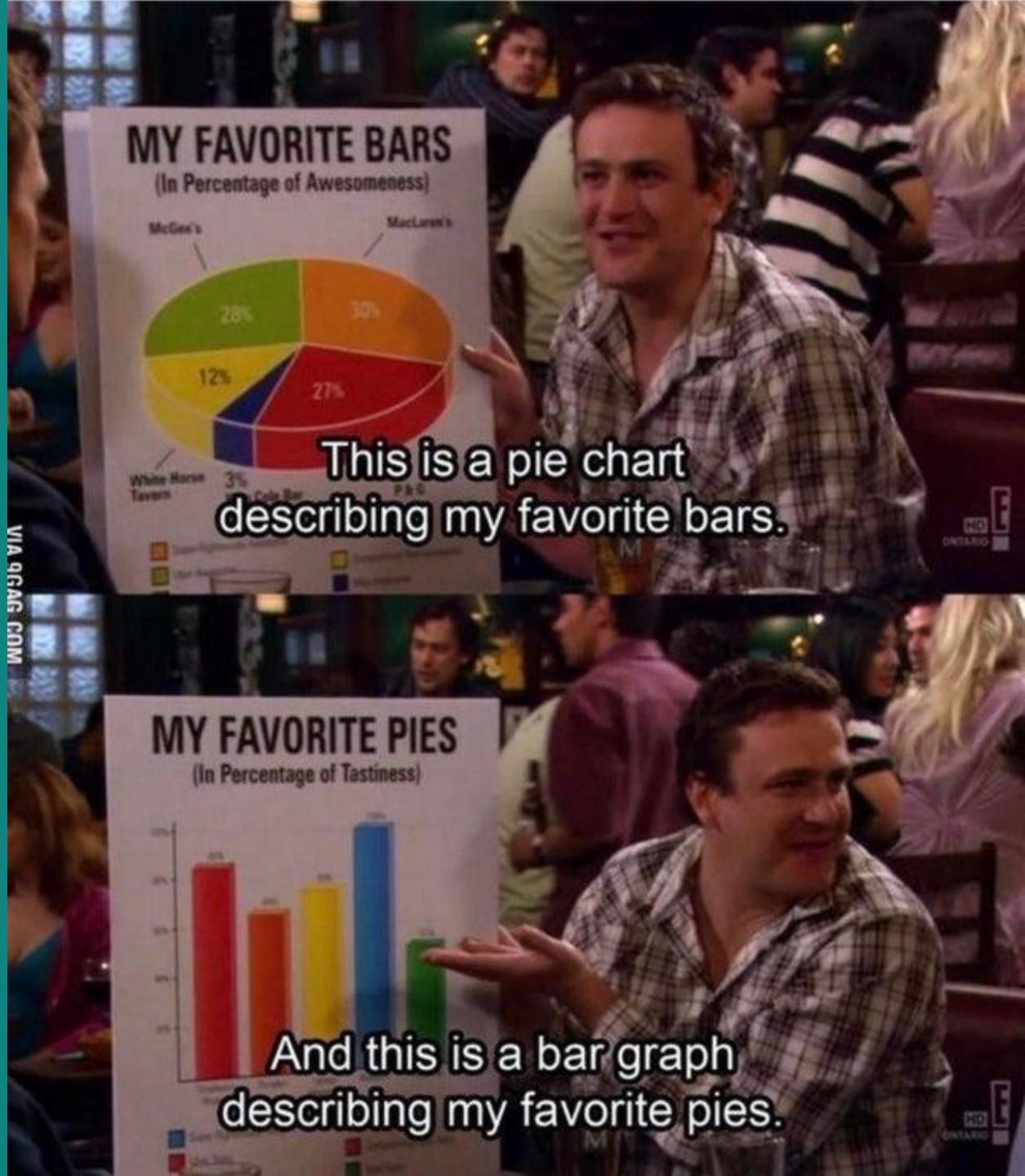
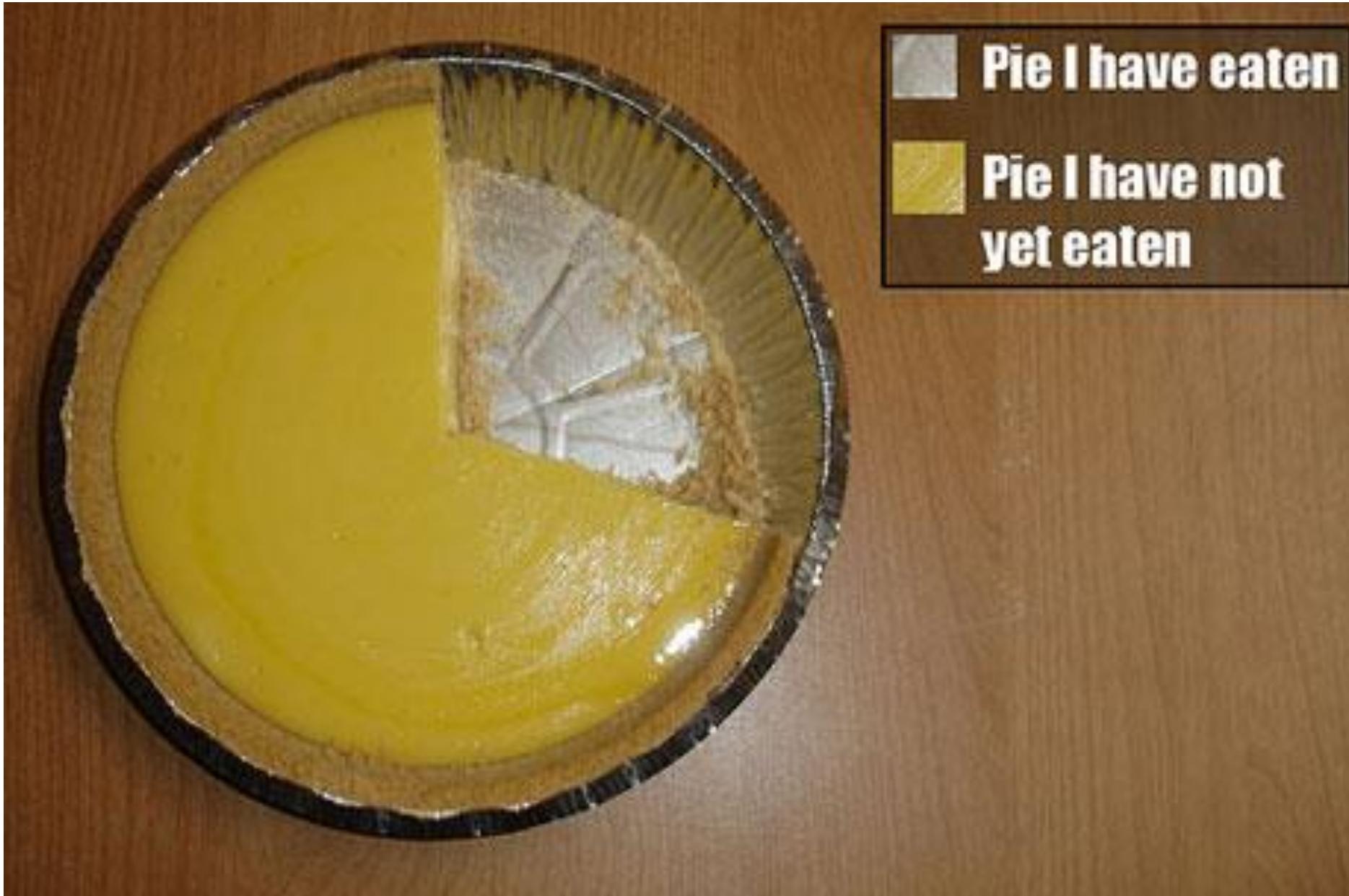


Chart of the week





“Save the Pies for Dessert”- Stephen Few

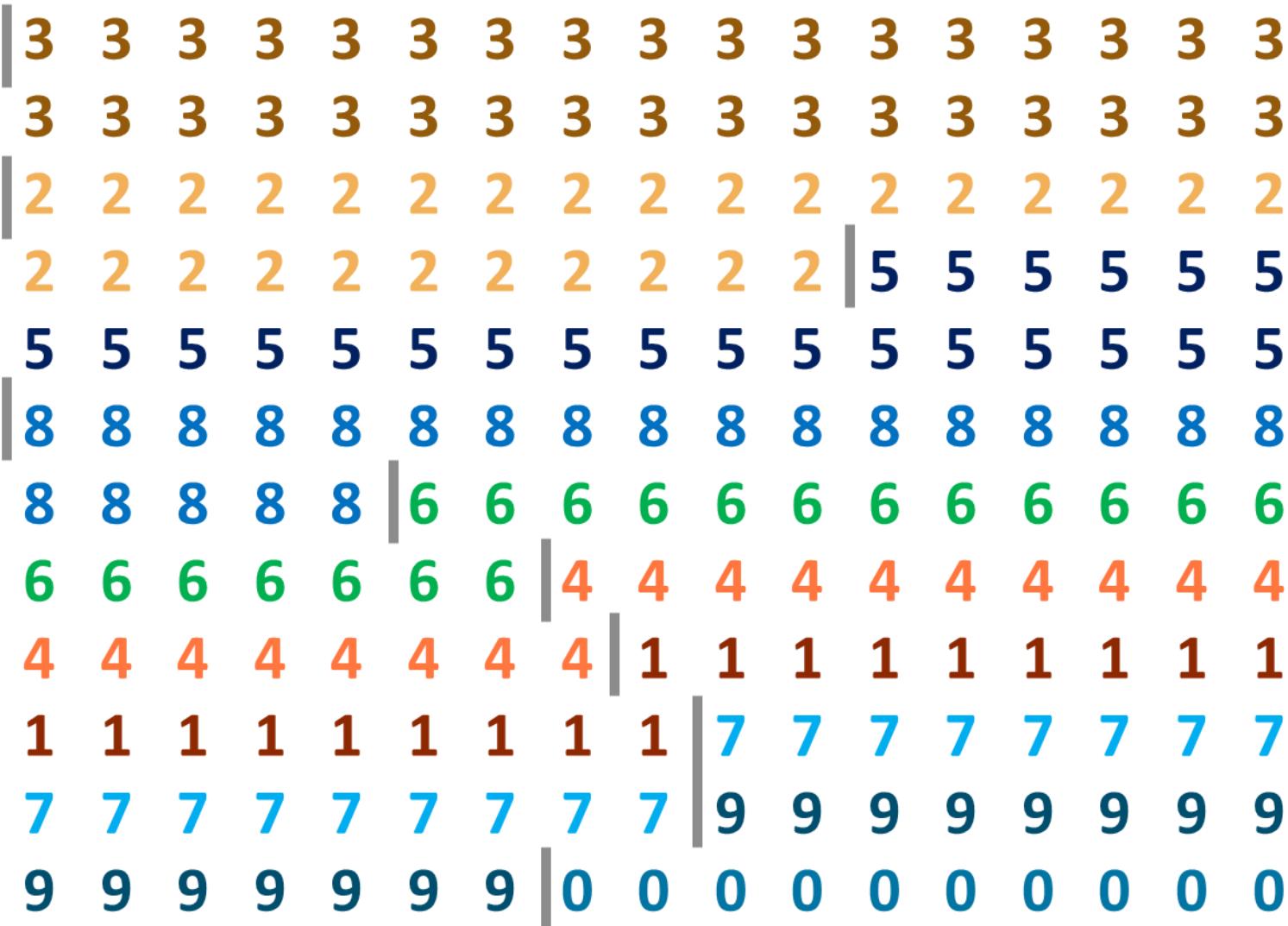
5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6

This is how a Pie Chart Represents the Data

3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6
6	6	6	6	6	6	6	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	7	7	7	7	7	7	7	7	7	7	7
7	7	7	7	7	7	7	7	7	9	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0

Try to quickly compare the totals of any digits.

This is how a Pie Chart Represents the Data



Baseline moves based on previous # of Digits

This is how a Bar Chart Represents that Data

5555555555555555555555555555

888

666666666666666666

44444444444444444444

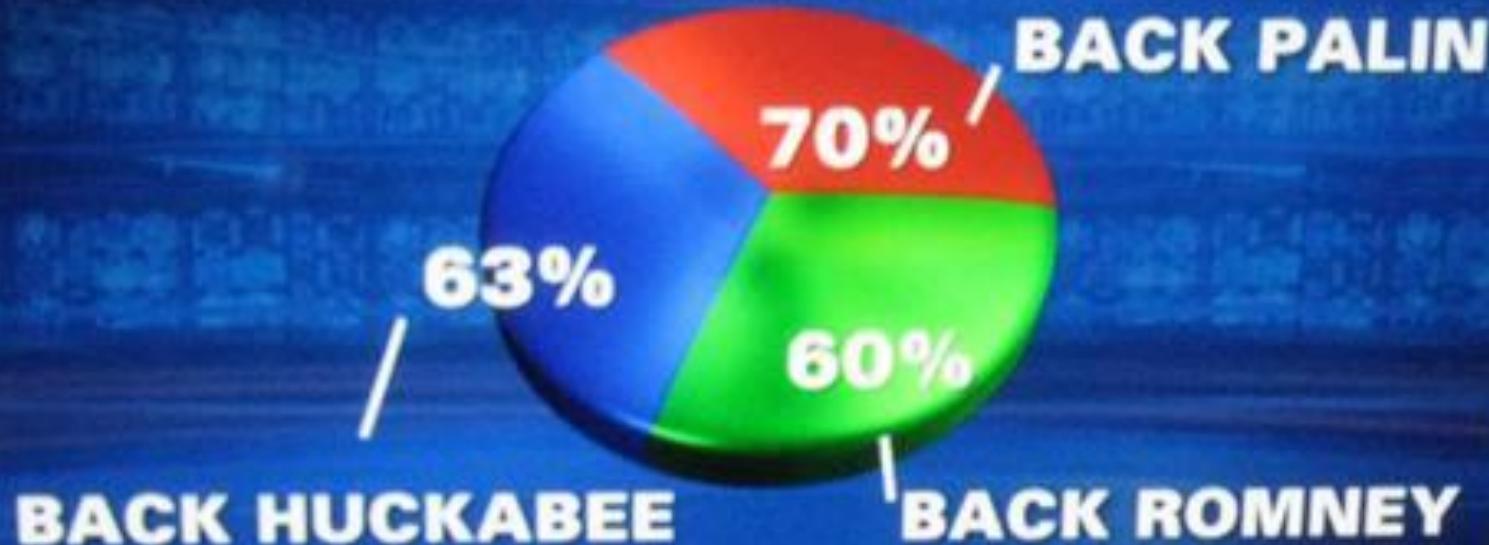
111111111111111111

Easier comparison even without color encoding!

Remove
to improve
the **pie chart** edition

2012 PRESIDENTIAL RUN

GOP CANDIDATES



SOURCE: OPINIONS
DYNAMIC

FOX

47°

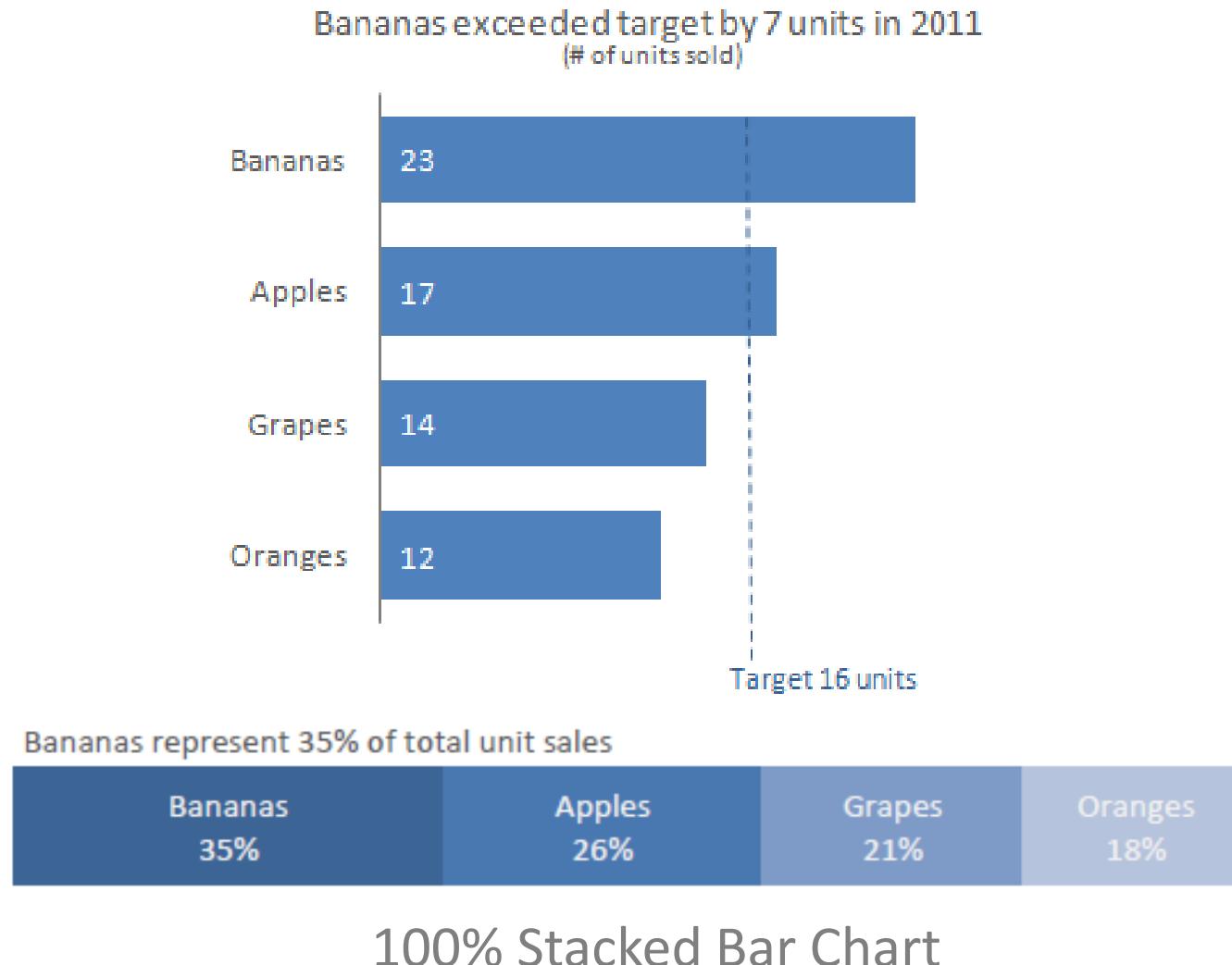
General Rules for Pie Charts

1. Don't Use Pie Charts

If you must break Rule #1 then:

2. Make sure it adds up 100%
3. Only a few categories
4. Start at noon and move clockwise
5. Largest to Smallest Values
6. Add Labels for %
7. Avoid 3D
8. Keep it Simple

Show part-to-whole relationship (in lieu of Pie Charts)



Girl Scout Cookie Sales

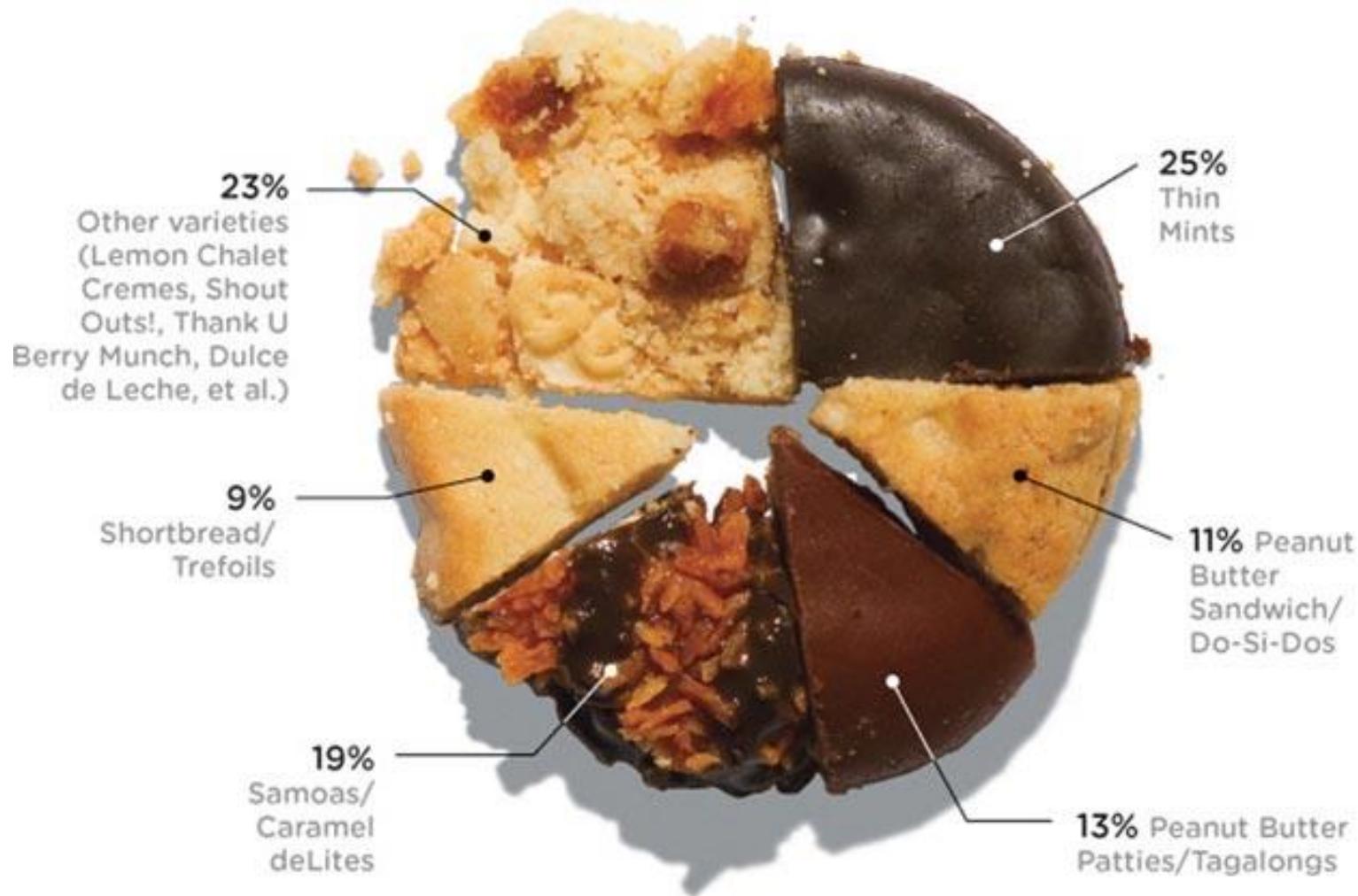
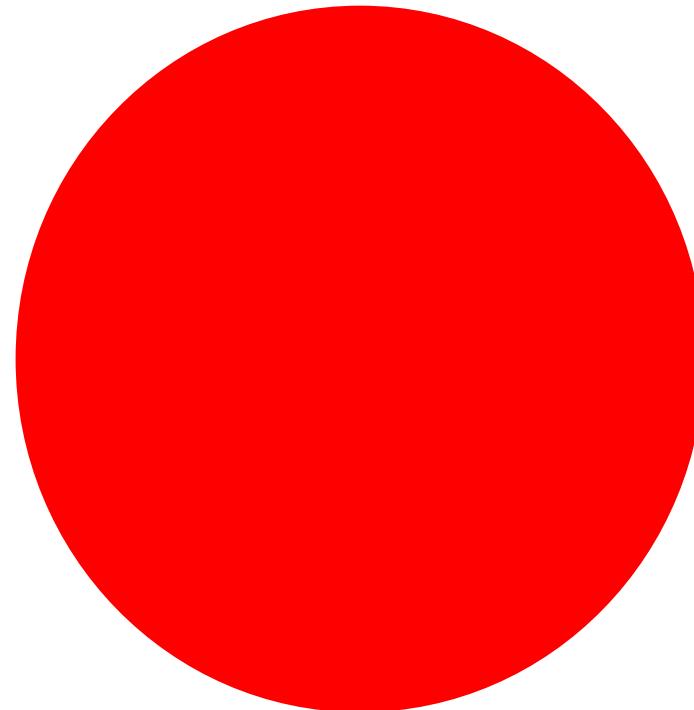


Photo: Celine Grouard

Source: www.wired.com/magazine/2011/08/st_datagirlscoutcookies

Pie Chart of Japan



■ Japan

UNIVERSITY OF
WESTMINSTER™