ESWC Summer School 2016 Understanding and Communicating with Data

Dr. Chris Phethean

University of Southampton

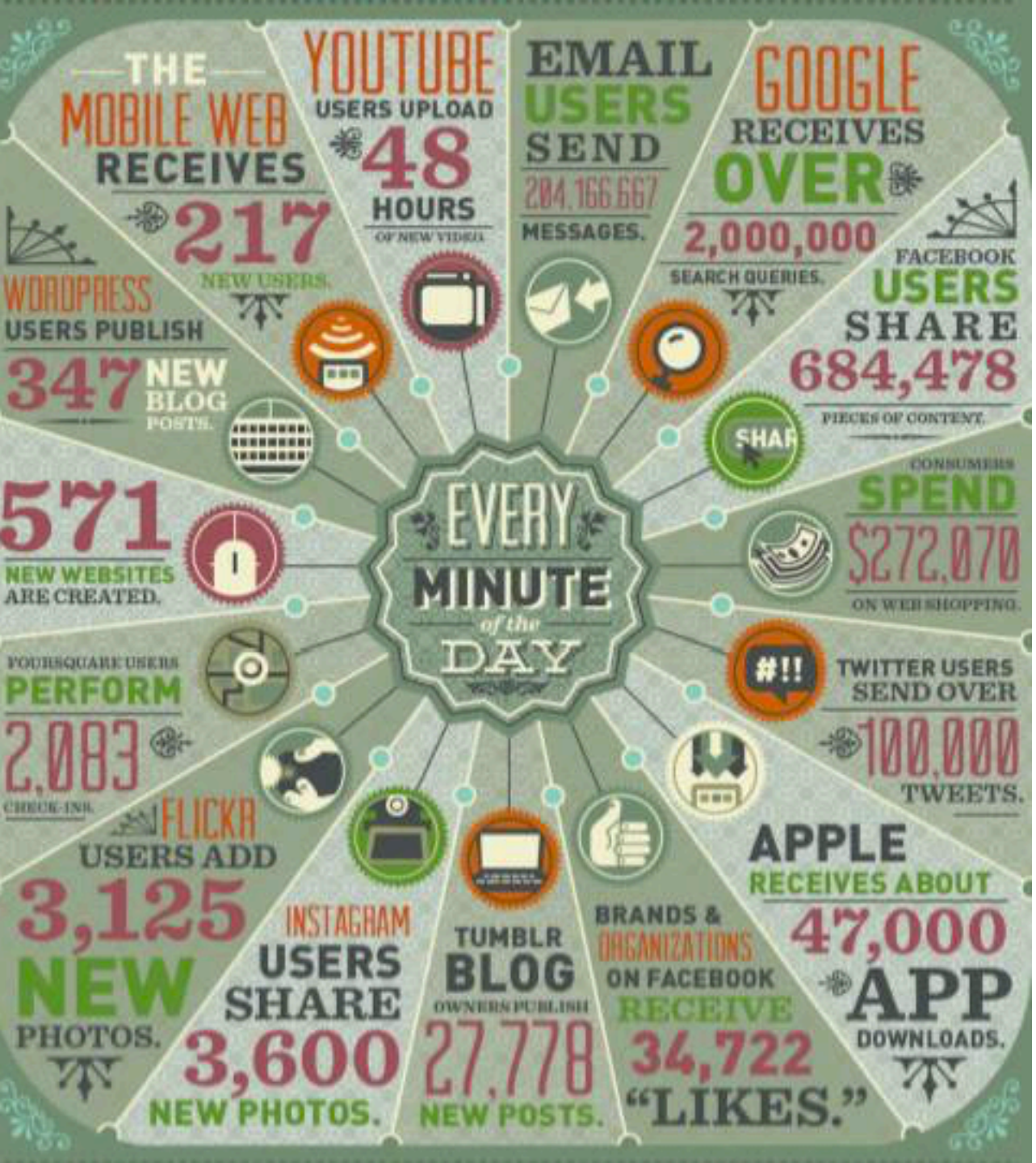
About Me

- Undergraduate, postgraduate and PhD from University of Southampton
- Computer Science -> Web Science -> Data Science
- Now
  - Research Fellow
  - Primarily working on the European Data Science Academy project
  - Teach on MSc Data Science course

Data… there's a lot out there!

- According to IBM:
  - 2.5 Quintillion bytes of data are created each day
  - 90% of the data in the world has been created in the last TWO years.

http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html



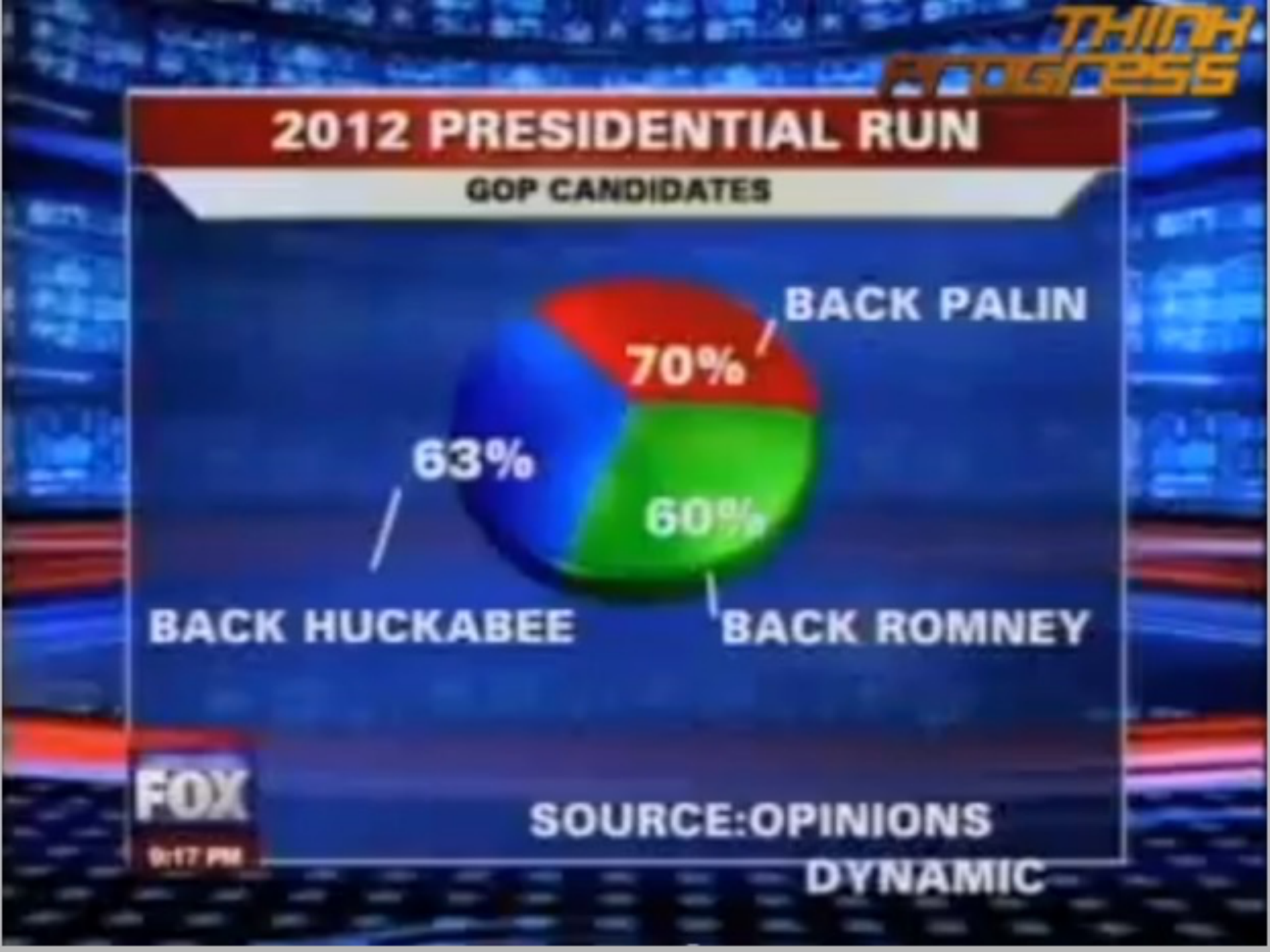
https://www.linkedin.com/pulse/4-big-data-sources-can-used-commercial-lending-risk-modeling-azalone

Communicating with data

- Many people struggle to convey the value of their data
- As the complexity of data increases, so too does the complexity of the analysis done with it
- These results then need to be communicated to a range of audiences

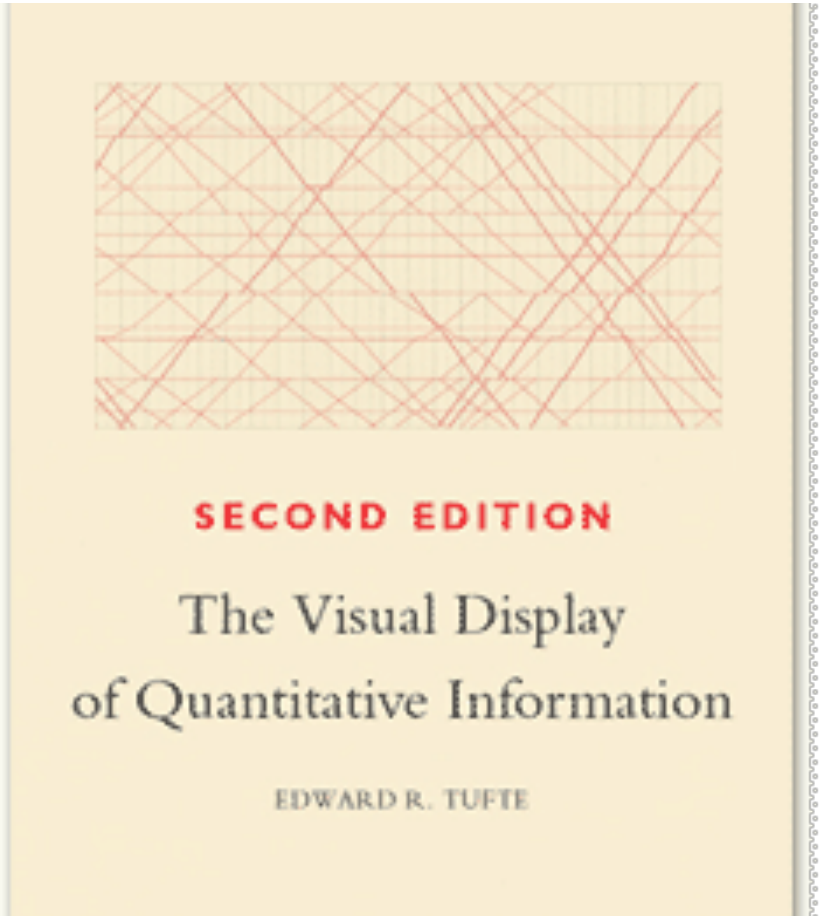
- “Just statistics does not give the full picture” – Claudia on Monday

- Alongside this we can use data visualisation to get a better idea of what the data is saying.


THINK PROGRESS
2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
9:17 PM
SOURCE:OPINIONS DYNAMIC

Data Visualisation is…

- ***"The Visual Display of Quantitative Information"***
- The title of this book provides a good starting definition
- We will see later that Tufte is a big name in the world of visualisation



What about Wikipedia?

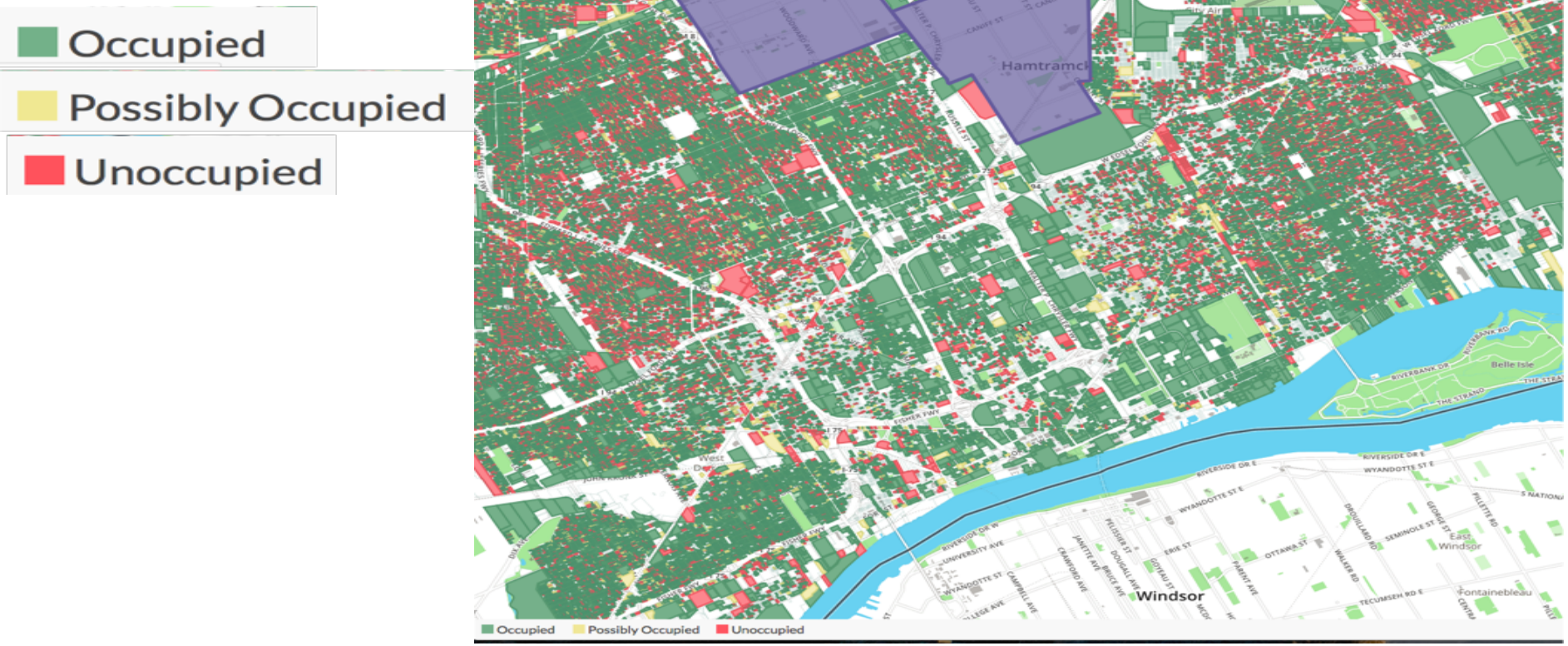
- *"to communicate information clearly and efficiently via statistical graphs, plots and information graphics"*
  - https://en.wikipedia.org/wiki/Data _visualization

- Key points
  - Communicate information
  - Clearly
  - Efficiently
  - Using statistical graphs, plots, information graphics

An example – Detroit (1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	OBJECTID	GEOID10	NAMELSAD10	SqMiles	Acres	CouncilDist	Likely_Occupied	Potentially_Vacant	Likely_Vacant	VeryLikely_Vacant	Total_MCM_Parcel_Count	Pct_Likely_Occupied	Pct_Potentially_Vacant	Pct_Likely_Vacant
2	1	2.6164E+10	Census Tract 5248	0.4317877	276.34413	6	976	123	118	80	1297	75.25057826	9.483423285	9.097918273
3	2	2.6164E+10	Census Tract 5219	0.6534582	418.21325	6	444	42	27	28	541	82.0702403	7.763401109	4.990757856
4	3	2.6164E+10	Census Tract 5218	0.0868371	55.57573	6	25	1			26	96.15384615	3.846153846	
5	4	2.6164E+10	Census Tract 5215	0.5822193	372.62034	6	241	37	29	34	341	70.6744868	10.85043988	8.504398827
6	5	2.6164E+10	Census Tract 5214	0.504105	322.62721	6	310	26	19	15	370	83.78378378	7.027027027	5.135135135
7	6	2.6164E+10	Census Tract 5213	0.3742017	239.4891	6	226	61	49	30	366	61.74863388	16.66666667	13.38797814
8	7	2.6164E+10	Census Tract 5211	0.6177754	395.37627	6	387	38	25	18	468	82.69230769	8.11965812	5.341880342
9	8	2.6164E+10	Census Tract 9853	0.4267861	273.14312	6	45	18	13	5	81	55.55555556	22.22222222	16.04938272
10	9	2.6164E+10	Census Tract 5208	0.5205792	333.1707	6	72	4	3	2	81	88.88888889	4.938271605	3.703703704
11	10	2.6164E+10	Census Tract 5207	0.2963056	189.63559	6	72	15	14	7	108	66.66666667	13.88888889	12.96296296
12	11	2.6164E+10	Census Tract 5204	0.158139	101.20893	6	109	18	13	3	143	76.22377622	12.58741259	9.090909091
13	12	2.6164E+10	Census Tract 5203	0.1741772	111.47344	6	118	18	10	5	151	78.14569536	11.9205298	6.622516556
14	13	2.6164E+10	Census Tract 5169	0.2375026	152.00169	5	31	1			32	96.875	3.125	
15	14	2.6164E+10	Census Tract 5168	0.3796373	242.96785	5	228	80	62	91	461	49.45770065	17.35357918	13.44902386
16	15	2.6164E+10	Census Tract 5167	0.4630091	296.32581	5	101		2		103	98.05825243		1.941747573
17	16	2.6164E+10	Census Tract 5166	0.2668356	170.77478	5	25	6	1		32	78.125	18.75	3.125
18	17	2.6164E+10	Census Tract 5165	0.5169066	330.82024	5	73	17	14	4	108	67.59259259	15.74074074	12.96296296
19	18	2.6164E+10	Census Tract 5164	0.3978771	254.64135	5	209	70	52	54	385	54.28571429	18.18181818	13.50649351
20	19	2.6164E+10	Census Tract 5163	0.239059	152.99777	5	185	65	43	52	345	53.62318841	18.84057971	12.46376812
21	20	2.6164E+10	Census Tract 5162	0.21823	139.66717	5	187	55	30	45	317	58.99053628	17.35015773	9.463722397
22	21	2.6164E+10	Census Tract 5161	0.5209916	333.43461	5	192	65	85	69	411	46.71532847	15.81508516	20.68126521
23	22	2.6164E+10	Census Tract 5157	0.4664307	298.51566	5	103	7	3	3	116	88.79310345	6.034482759	2.586206897
24	23	2.6164E+10	Census Tract 5156	0.333939	213.72095	5	223	49	47	75	394	56.59898477	12.43654822	11.92893401
25	24	2.6164E+10	Census Tract 5154	0.3484437	223.00396	5	409	36	23	27	495	82.62626263	7.272727273	4.646464646
26	25	2.6164E+10	Census Tract 5153	0.3069975	196.47843	5	273	71	35	37	416	65.625	17.06730769	8.413461538
27	26	2.6164E+10	Census Tract 5152	0.3588843	229.68593	5	388	92	76	153	709	54.72496474	12.97602257	10.71932299
28	27	2.6164E+10	Census Tract 5145	0.2177369	139.35162	5	303	55	66	120	544	55.69852941	10.11029412	12.13235294
29	32	2.6164E+10	Census Tract 5202	0.3189737	204.14315	6	87	4	1		92	94.56521739	4.347826087	1.086956522
30	33	2.6164E+10	Census Tract 5188	0.5465149	349.76951	5	206	66	56	72	400	51.5	16.5	14
31	34	2.6164E+10	Census Tract 5186	0.2863341	183.25384	5	177	66	50	56	349	50.71633238	18.91117479	14.32664756
32	35	2.6164E+10	Census Tract 5185	0.3090667	197.80267	5	221	82	74	65	442	50	18.5520362	16.74208145
33	36	2.6164E+10	Census Tract 5184	0.4079347	261.0782	5	122	55	39	56	272	44.85294118	20.22058824	14.33823529
34	37	2.6164E+10	Census Tract 5180	0.3286309	210.32375	5	94	16	3	4	117	80.34188034	13.67521368	2.564102564
35	38	2.6164E+10	Census Tract 5175	0.4252193	272.14037	5	54	9	3		66	81.81818182	13.63636364	4.545454545
36	39	2.6164E+10	Census Tract 5172	0.6302674	403.37112	5	150	21	10	3	184	81.52173913	11.41304348	5.434782609
37	40	2.6164E+10	Census Tract 5171	0.1083773	69.361492	5	25		1		26	96.15384615		3.846153846
38	41	2.6164E+10	Census Tract 5170	0.2217952	141.94894	5	31	5	1		37	83.78378378	13.51351351	2.702702703
39	43	2.6164E+10	Census Tract 5143	0.7796522	498.97742	4	626	151	109	245	1131	55.34924845	13.3510168	9.637488948
40	44	2.6164E+10	Census Tract 5141	0.6214851	397.75046	4	648	184	157	380	1369	47.33382031	13.44046749	11.46822498
41	45	2.6164E+10	Census Tract 5139	0.4966386	317.84869	5	399	95	87	193	774	51.5503876	12.27390181	11.24031008
42	46	2.6164E+10	Census Tract 5136	0.3779577	241.89294	5	231	61	45	106	443	52.14446953	13.76975169	10.15801354
43	47	2.6164E+10	Census Tract 5133	0.6241106	399.43077	4	573	61	44	45	723	79.25311203	8.437067773	6.085753804
44	48	2.6164E+10	Census Tract 5132	0.3582623	229.28785	4	493	95	68	135	791	62.32616941	12.01011378	8.596713021
45	49	2.6164E+10	Census Tract 5129	0.5848656	374.314	4	283	89	48	105	525	53.9047619	16.95238095	9.142857143
46	50	2.6164E+10	Census Tract 5126	0.5342581	341.92515	4	367	110	86	159	722	50.83102493	15.23545706	11.91135734
47	51	2.6164E+10	Census Tract 5124	0.3867265	247.50495	4	283	79	74	139	575	49.2173913	13.73913043	12.86956522
48	52	2.6164E+10	Census Tract 5123	0.3064979	196.15868	4	305	76	49	96	526	57.98479087	14.4486692	9.315589354
49	53	2.6164E+10	Census Tract 5121	0.4607828	294.901	4	686	164	122	253	1225	56	13.3877551	9.959183673
50	54	2.6164E+10	Census Tract 5119	0.4450519	284.83318	5	299	65	90	99	553	54.06871609	11.75406872	16.27486438
51	60	2.6164E+10	Census Tract 9859	0.2414378	154.52018	5	20	5	3	2	30	66.66666667	16.66666667	10
52	61	2.6164E+10	Census Tract 9852	0.940558	601.95711	4	14	1	1	2	18	77.77777778	5.555555556	5.555555556

An example – Detroit (2)



https://www.motorcitymapping.org/#t=parcels&s=detroit&f=all&x=preset2

Why data visualisation?

- Easier to spot key features in the data
  - Act on emerging trends
- Explain the story that the data represents
- Explore what the data shows to find interesting insights
- Easier to communicate findings than with a spreadsheet

Data isn't new: we've always needed to understand and communicate insights

- London cholera outbreak, 1854
- John Snow investigated the outbreak, observing clusters around Broad Street
- Using a map-based visual, he tallied how many case were reported in each location

Cholera in London, 1854

John Snow

- Upon further investigation, he noticed that many cases were clustered around the Broad Street water pump.
- He used the map to convince the authorities and ordered that the handle was removed and the pump investigated, and cases subsequently reduced.
- This helped to change scientific understanding of how diseases are spread.

The First Bar Chart: An apology

This Chart is different from the others in principle, as it does not comprehend any portion of time, and it is much inferior in utility to those that do; for though it gives the extent of the different branches of trade, it does not compare the same branch of commerce with itself at different periods; nor does it imprint upon the mind that distinct idea, in doing which, the chief advantage of Charts consists: for as it wants the dimension that is formed by duration, there is no shape given to the quantities.

Read more: Beniger, James R., and Dorothy L. Robyn. 1978. "Quantitative Graphics in Statistics: A Brief History". *The American Statistician* 32 (1). [American Statistical Association, Taylor & Francis, Ltd.]: 1–11. doi:10.2307/2683467.

Types of visualisation

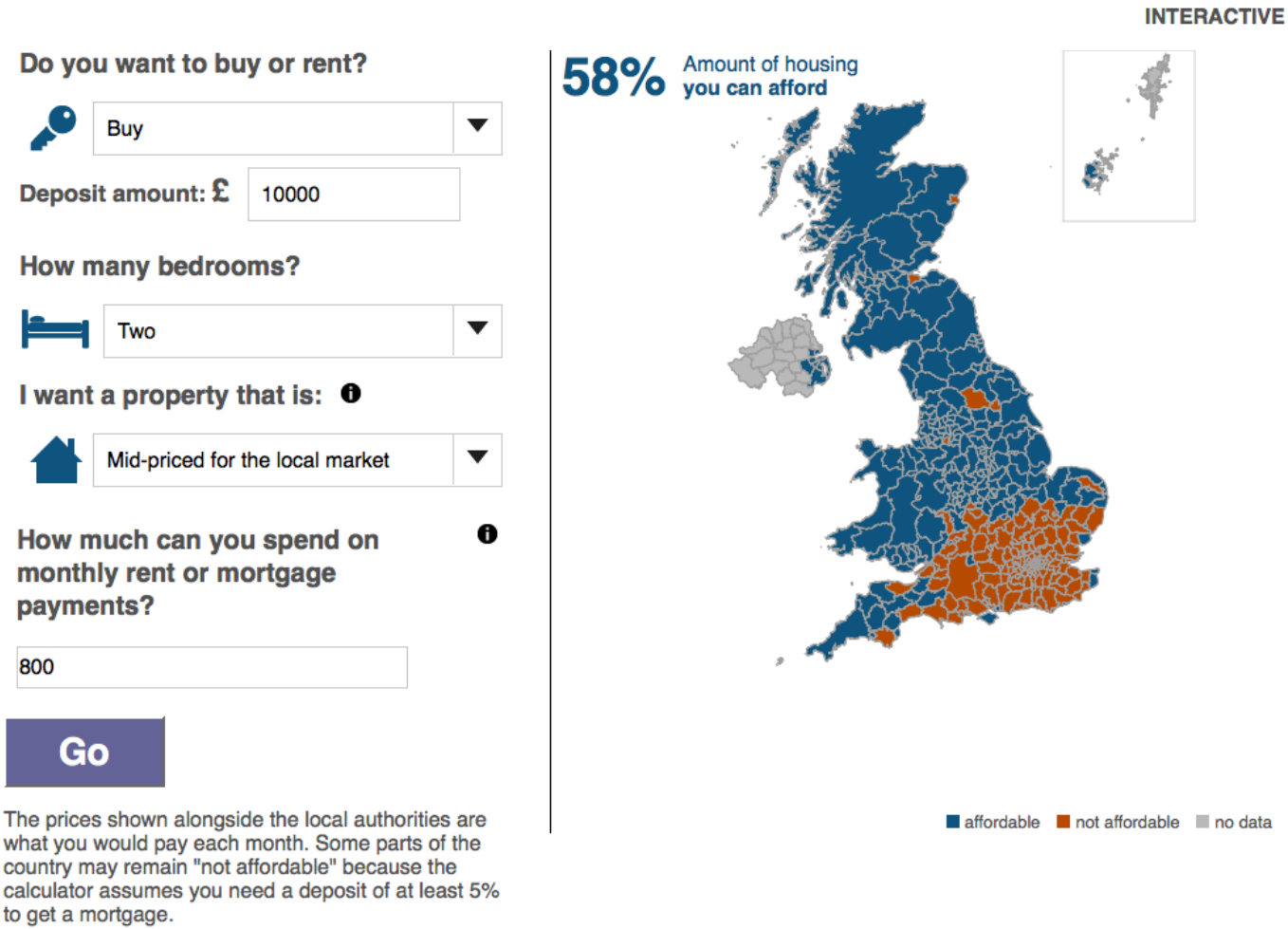
Explore or Explain?

Types of Visualisation

- Data visualisations can generally be split into two main categories:

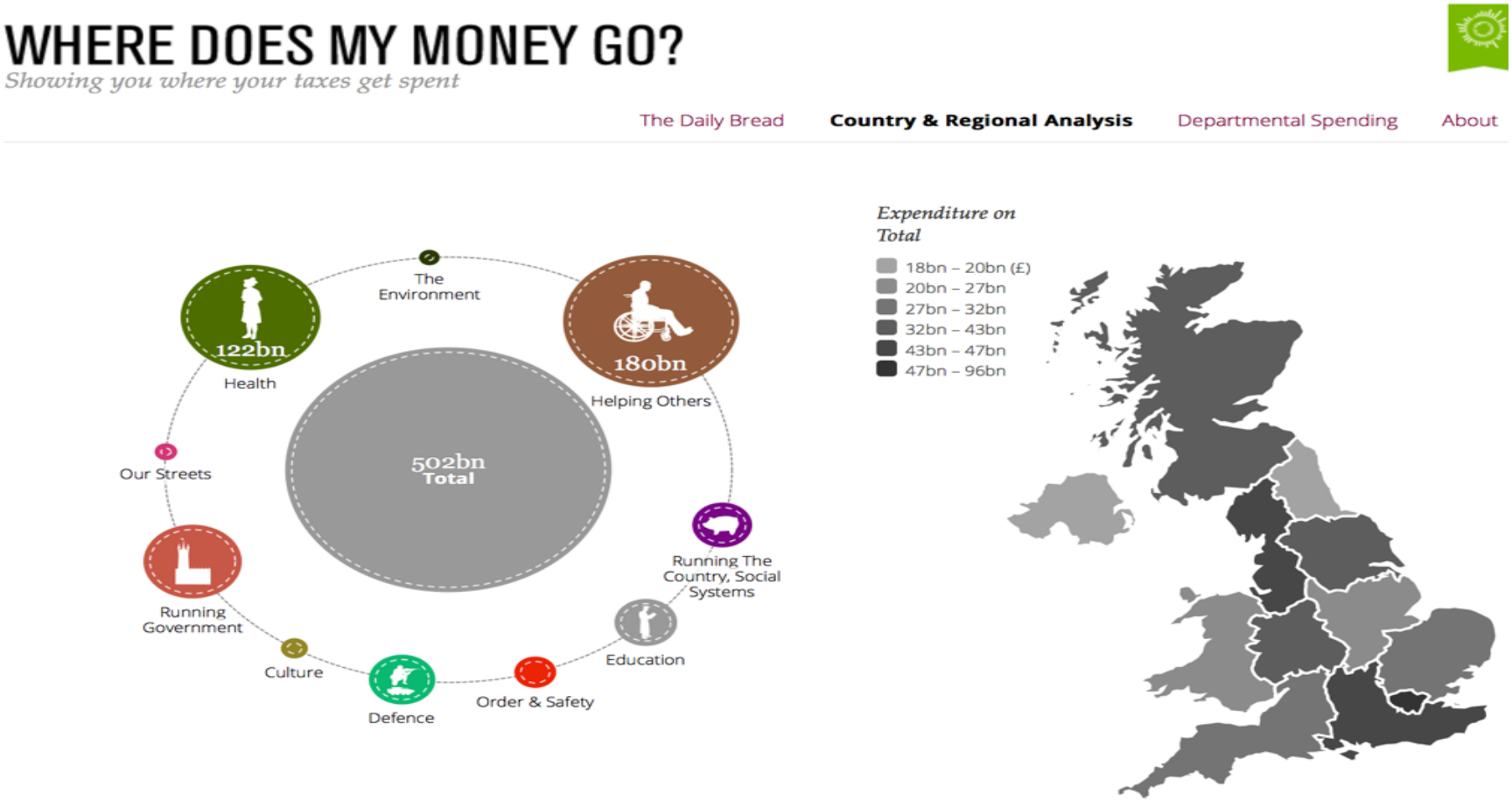
Exploratory	Explanatory
‘Explore’ your data	‘Explain’ your findings
Provide a way to examine your data from different viewpoints or with different filters.	You’ve found the interesting points in the data, now you want to explain or communicate them to others
Generally allows the reader to drive the experience: Let them gain a sense of what the data is about and what it is saying, so that you/they can find the interesting points.	Generally ‘author-driven’: your main focus is on telling the story that you’ve found, so you want to present it as clearly as possible.

Exploratory Visualisation – Where can you afford to live?



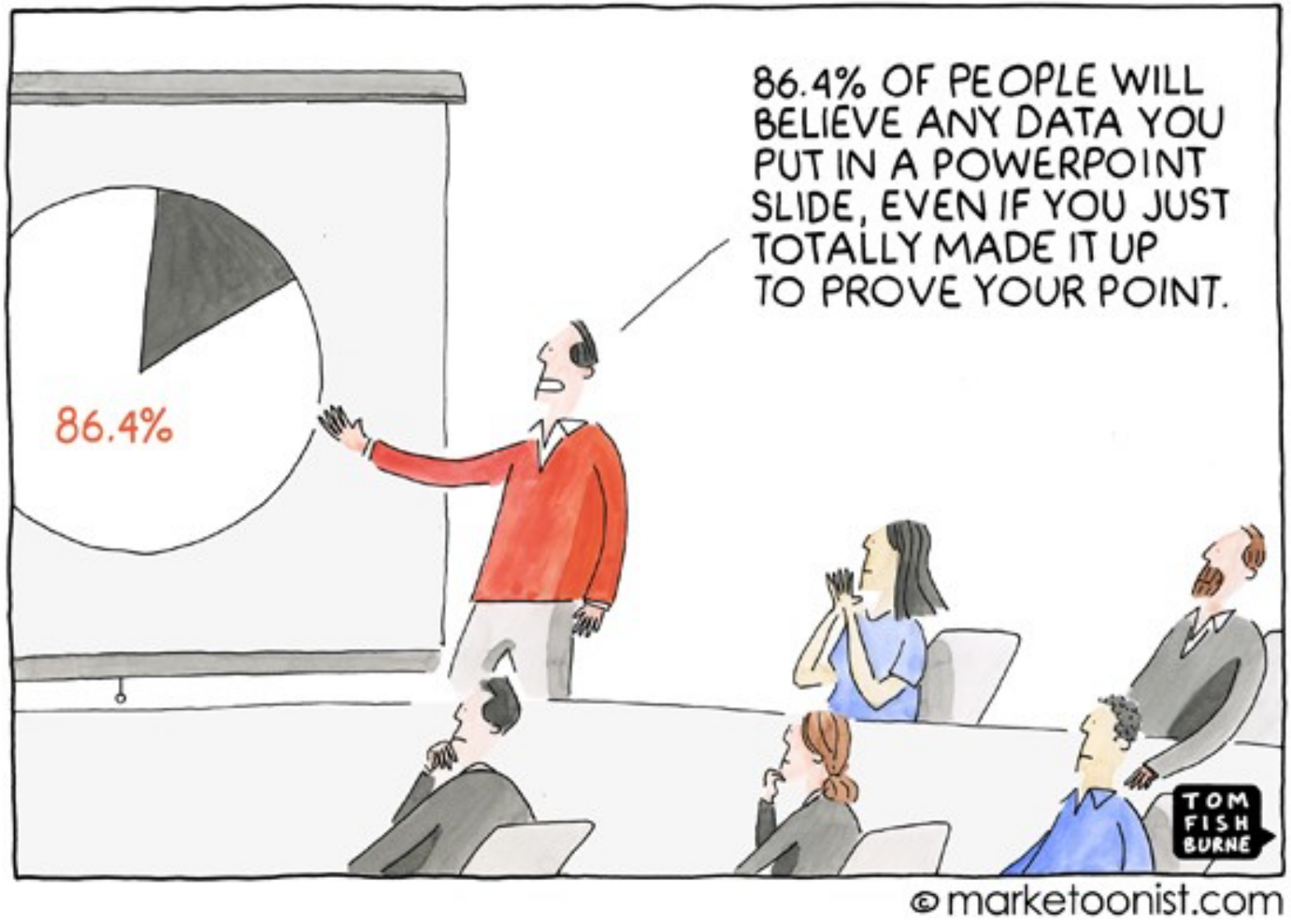
http://www.bbc.co.uk/news/business-23234033

Explanatory Visualisation



http://wheredoesmymoneygo.org/bubbletree-map.html#/~/total

Storytelling


86.4% OF PEOPLE WILL BELIEVE ANY DATA YOU PUT IN A POWERPOINT SLIDE, EVEN IF YOU JUST TOTALLY MADE IT UP TO PROVE YOUR POINT.
86.4%
TOM FISH BURNE
©marketoonist.com

Humans have been telling stories for centuries…

https://annadoherty.wordpress.com/2012/10/06/history-of-animation-cave-paintings/

Storytelling with data visualisation

- Storytelling: "*The conveying of events in words, sound and/or images, often by improvisation or embellishment.*" (Wikipedia)

- Recall that the aim of visualisation is to effectively convey some information from data.
- You want decision makers to be able to see the **results** of your analysis.
  - These should be conveyed in a manner that is suitable for the intended audience to understand.

Considerations for the story

- Claudia covered these a bit on Monday

- Who are your audience?
- What is the story?
- What do you want people to **do** after seeing your visualisation?

1. Your audience

- What **prior knowledge** do they posses?
  - Are they experts in the field? Or general public?
- How much **time** do they have?
  - Will they be able to explore the visualisation in depth, or just glance at it?
- What **format** are they comfortable digesting information?
  - Charts? Infographics? Interaction?

2. The story

- **What is it that the data shows?**
- What is your key result? The message.
  - Not the data itself.
  - Nor the method. **The result**.
- The story is vital – this is what you will be ‘telling’ in your visualisation.
- Why do you want people to look at your visualisation?

2. The story (2)

- You really have to know the story yourself in order to 'sell it'.
- "**Simplify, then exaggerate**".
  - Geoffrey Crowther, The Economist
  - Referring to journalism techniques that have been followed at The Economist since the 1950s
  - Important: this does not mean lying!

3. The action

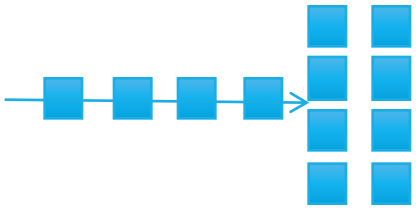
- What do you want people to **do** afterwards?
  - This depends on the audience, and context.
- Change behaviour?
  - E.g.,
    - Implement a new policy
    - Allocate resources
    - Campaign for change
    - Change eating habits
    - …

Structuring your story: author- or reader-driven?

- The type of story will determine the type of visualisation you need to create.
- Explanatory visualisations are typically ‘author-driven’.
  - An author-driven story has a linear ordering, which you walk the reader through, with little or no interactivity.
- Exploratory visualisations are more likely to be ‘reader-driven’.
  - For reader-driven stories, there is no prescribed ordering to the content, allowing the reader to interact and find their own story.

Balancing author- and reader-driven stories

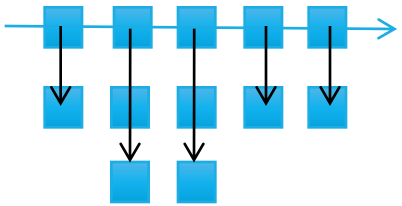
- As a compromise between the two approaches, we can merge them both
  1. Initial author-driven sequence, that opens up into an interactive narrative for exploration.



Often described as the ‘martini glass’.

Balancing author- and reader-driven stories

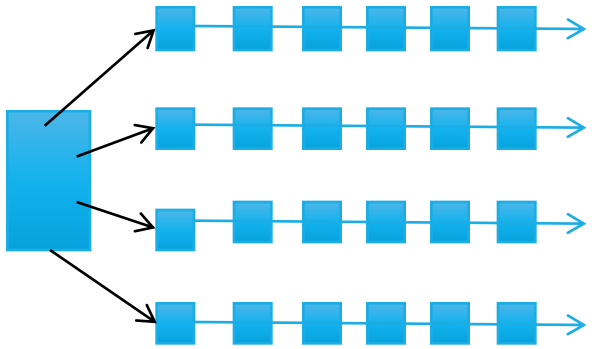
- As a compromise between the two approaches, we can merge them both
  2. Author-driven sequence with interaction allowed ‘mid-narrative’ – user can explore a particular point more before moving ahead.



Often described as the ‘interactive slideshow’.

Balancing author- and reader-driven stories

- As a compromise between the two approaches, we can merge them both
  3. General theme presented allowing reader-driven interaction, but they can then choose particular elements to reveal additional author-driven details.



Often described as the ‘drill-down story’.

Finding a ‘compelling’ narrative

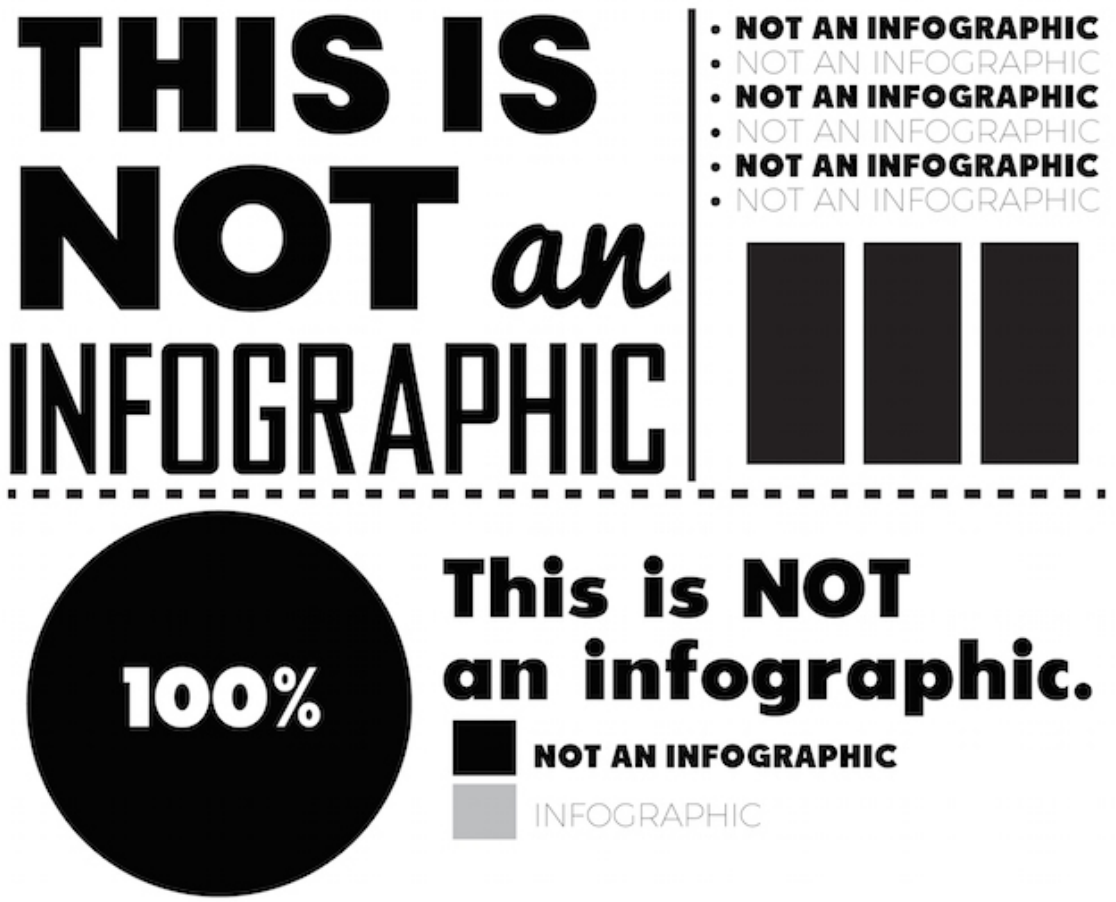
- The narrative should **link the facts** that you’ve found together in an **engaging** way.
- If you can’t find a narrative structure, then your findings may not be as good as you think
  - Or your data may be more suited to an **exploratory data analysis (reader-driven)** style of visualisation, letting the viewer find their own story.

https://hbr.org/2013/04/how-to-tell-a-story-with-data/

Communicating your message

- Two main techniques for ensuring you effectively convey your story:
  - Highlight and emphasise
    - Show the important information that is key to the story, without distractions.
  - Organise
    - Structure information the clearly guide readers to produce the most understanding.

Structure without a story



https://contently.com/strategist/2015/11/25/why-good-infographics-are-more-than-just-pictures-and-numbers/

Highlighting and Emphasising

Tufte’s theories on chart junk

Emphasising key information for the mind (1)

- Making an effective graphic is easier if we know how the mind works
- As an example, look at this image and see how many ‘3’s there are:

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

Emphasising key information for the mind (2)

- We can make it a lot easier to perceive by changing things slightly

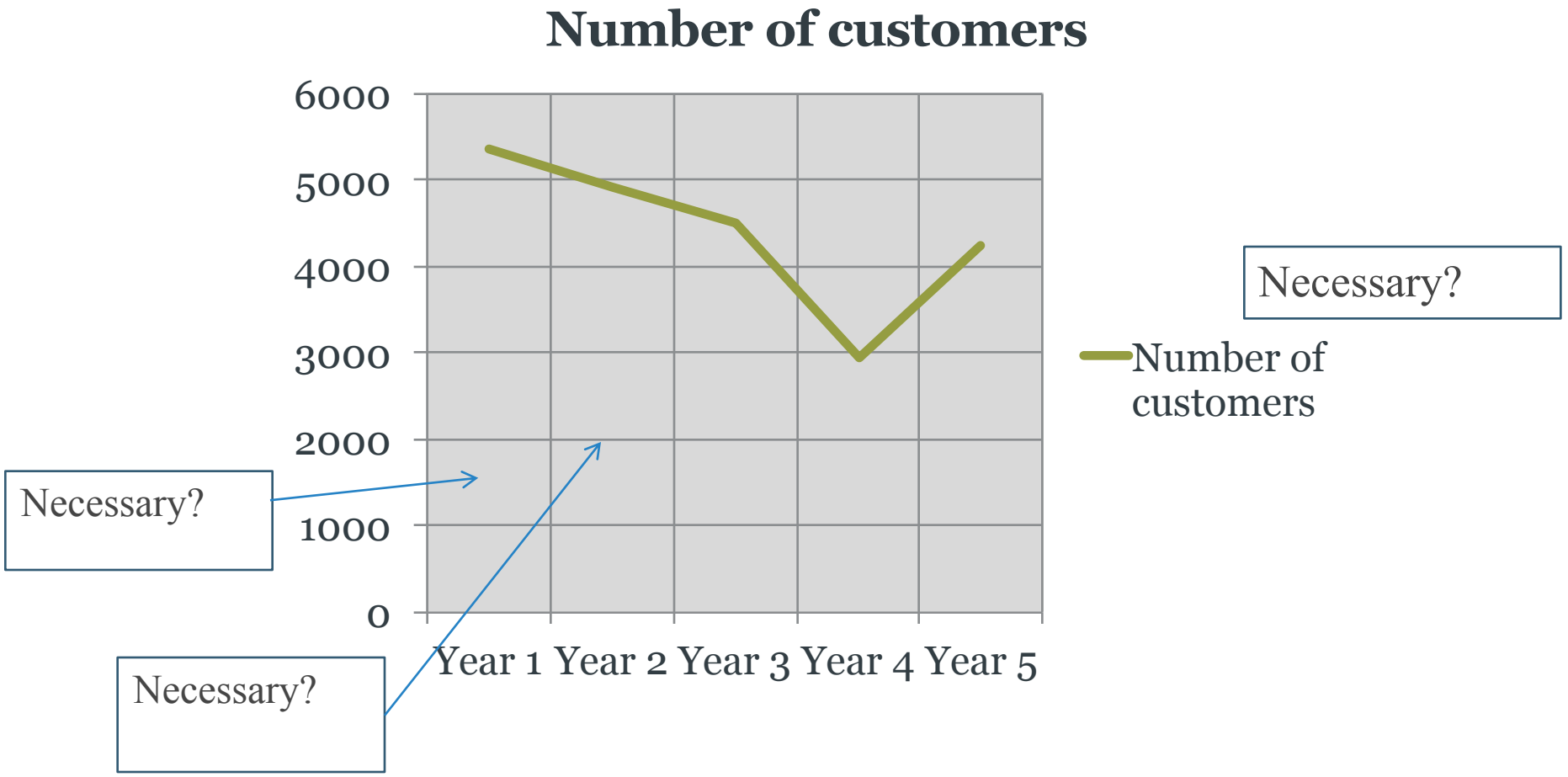
1 9 8 3 5 6 4 2 5 3 7 8 4
5 2 3 8 4 6 8 4 7 9 6 1 2
1 2 6 3 4 8 9 6 5 4 2 3 1
9 8 6 3 5 4 8 7 1 3 1 5 5
4 6 8 2 1 3 5 7 9 1 2 3 4
9 8 7 6 5 4 3 2 1 2 3 4 5

- Shade variations => Easy to perceive
- Shape variations => Difficult to perceive

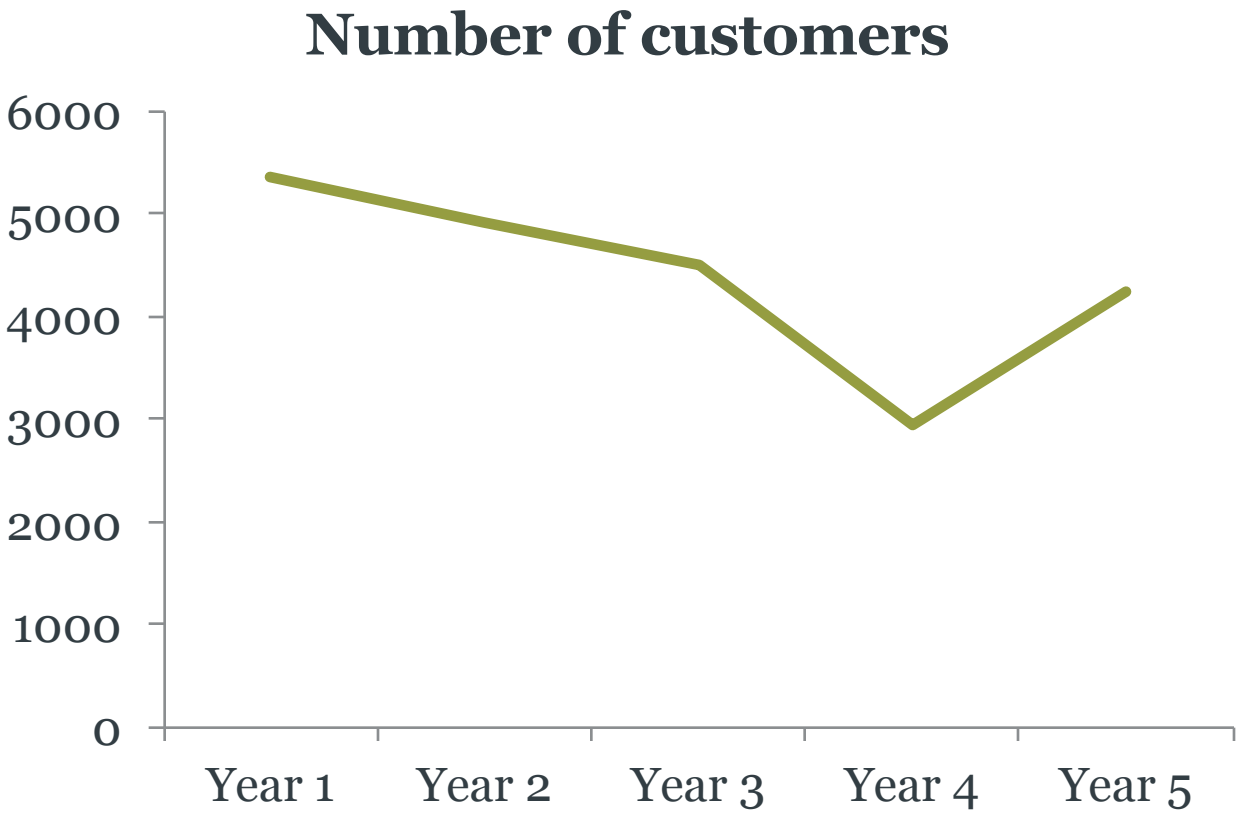
Chartjunk

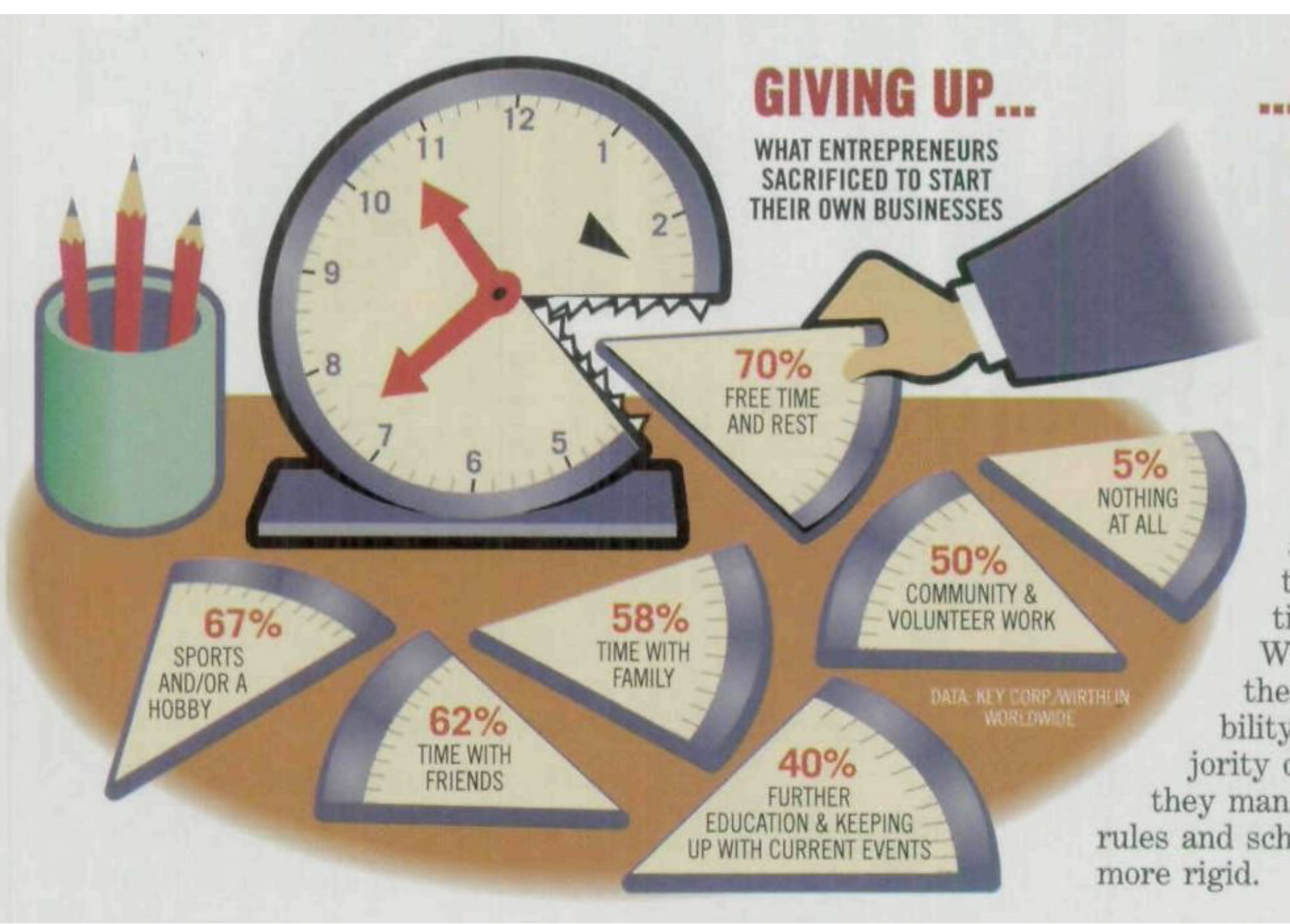
- Chartjunk, according to Tufte, is anything that can be removed from a chart without changing its meaning
  - It obscures the true meaning and story
  - Imagery is not information
  - Imagery draws attention away from data
    - Short-term memory resources are used to identify the images rather than understand the chart's meaning

Chartjunk Example



Chartjunk Example - Fixed



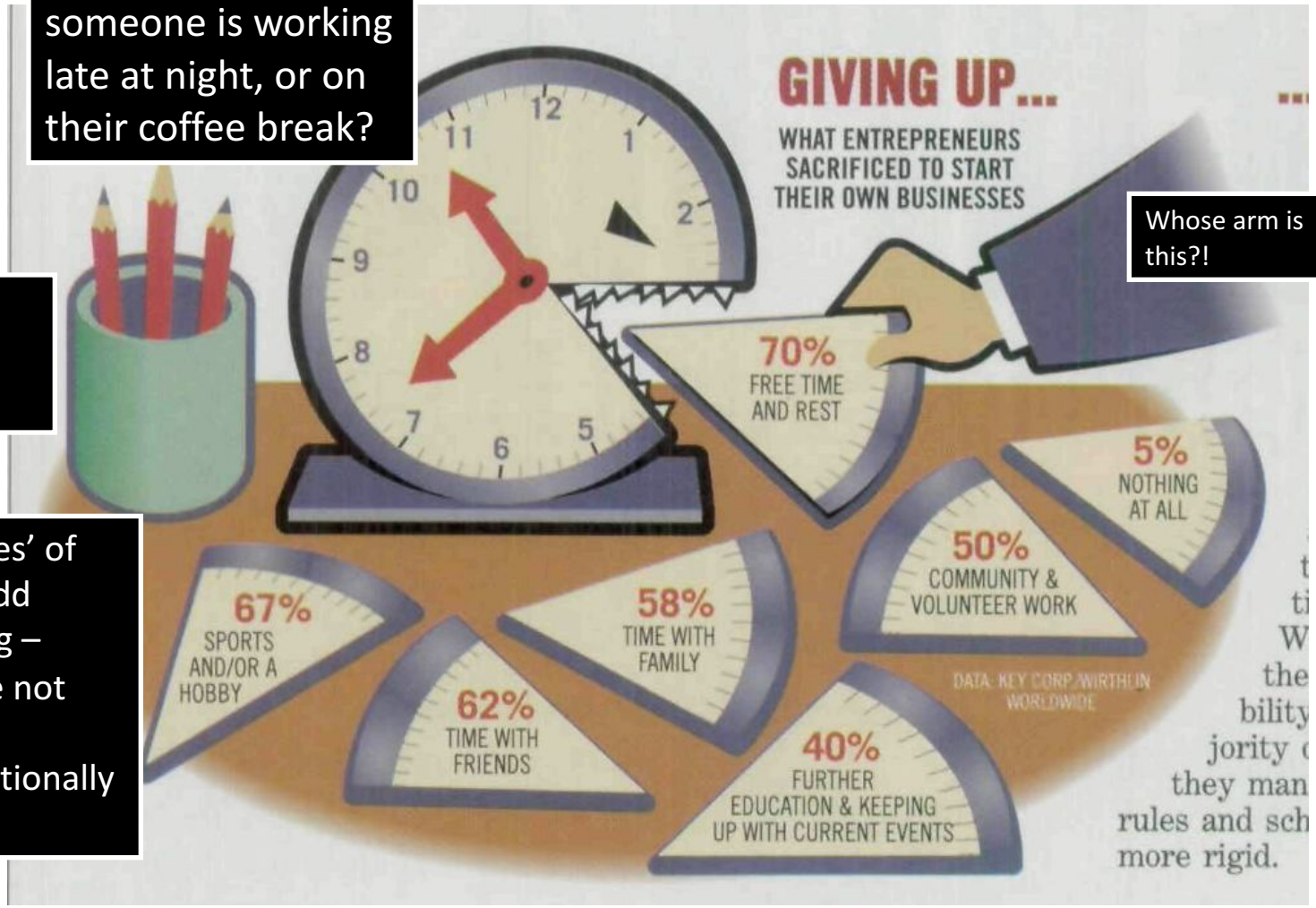


...

t
ti
W
the
bility
jority
they man
rules and sch
more rigid.

Let's fix this horrible example (2)

Is this saying someone is working late at night, or on their coffee break?



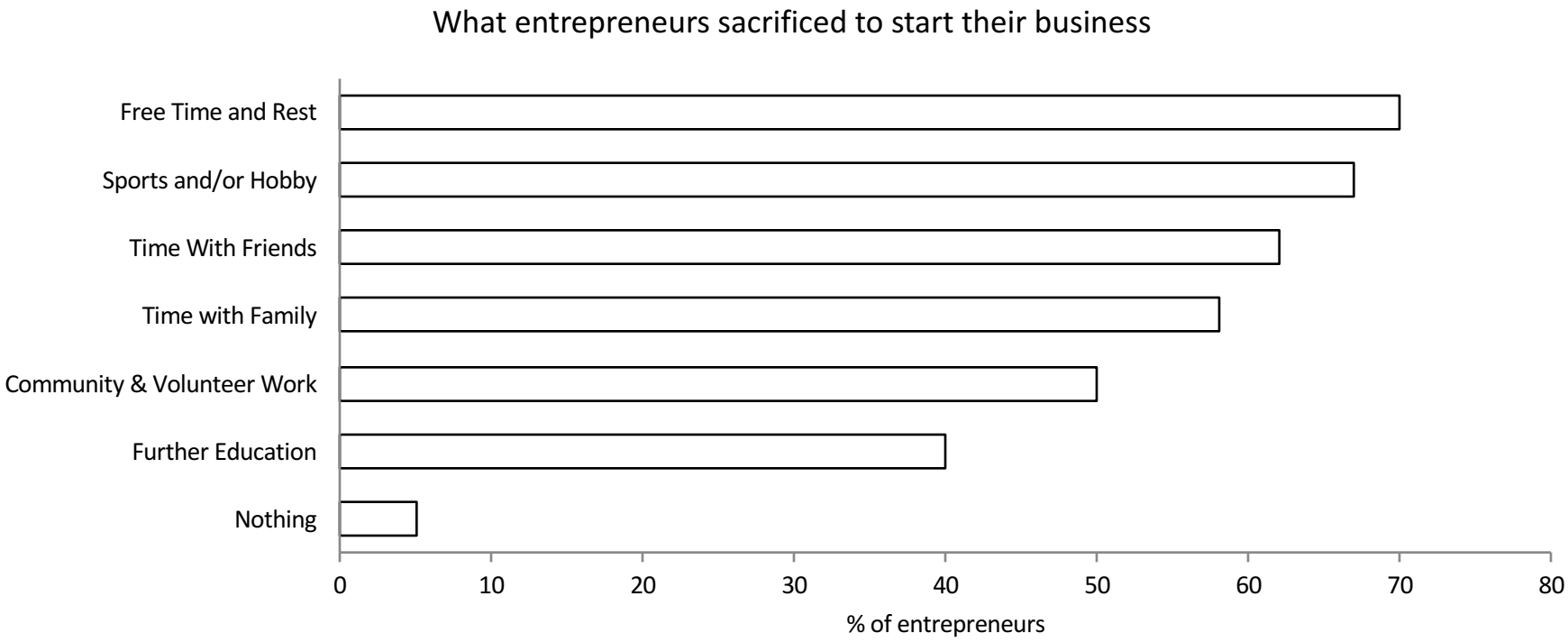
Whose arm is this?!

Pencils… they're not data

'Wedges' of time add nothing – they're not even proportionally sized!

https://datavizblog.com/category/chartjunk/

After removing all the ‘junk’ and keeping only the ‘data’, we get:



The graphic we made is simple, but will it really interest people? Will they remember it?

The Beauty Paradox

- How complex should a graphic be?
  - How much should you show?
- Tufte emphasises minimalism
  - Communicate as much information from as few pixels as possible
- Depending on the audience, this may or may not be the best approach
- When creating a visualisation, you must balance the need to focus on the story of the data, with art and design decisions to help your audience engage with and remember the visual.

Organising and Structuring Information

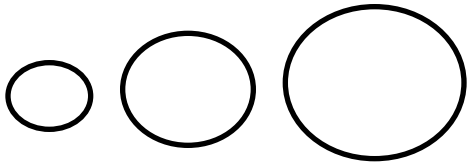
Visual perception and gestalt theory

How is data displayed?

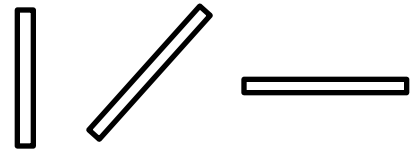
- We turn data into a graphical representation using '**visual encodings**'
- Depending on the **type** of data you have, different encodings will be suitable
- E.g. ordered data (discrete or continuous):
  - **Size** – the larger the circle, the more of something
  - **Orientation/rotation**
  - **Colour** (saturation)
- E.g. nominal data (categories):
  - **Colour** (hue)
    - Red circles to represent Manchester United supporters
    - Blue circles to represent Chelsea supporters

Ordered data

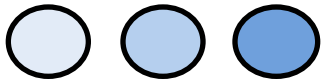
SIZE



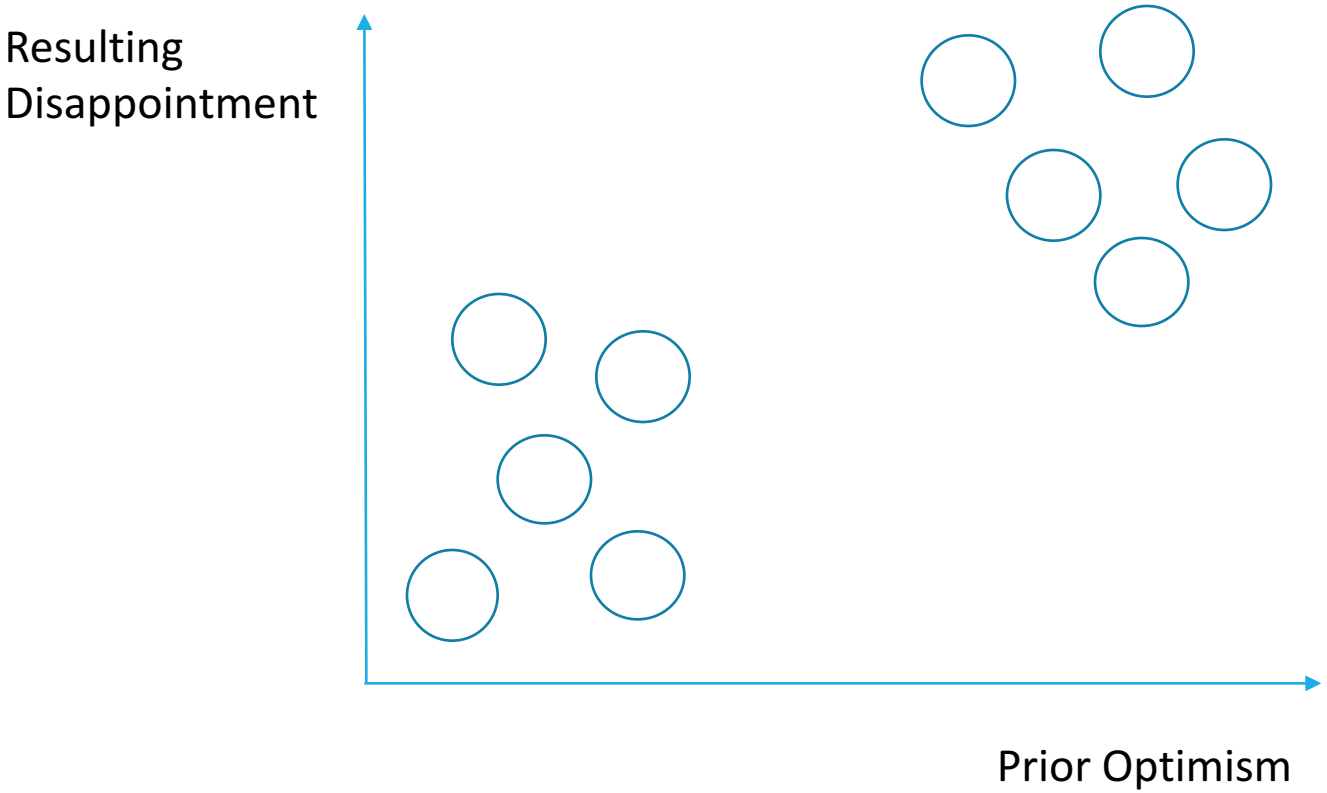
ORIENTATION / ROTATION



COLOUR

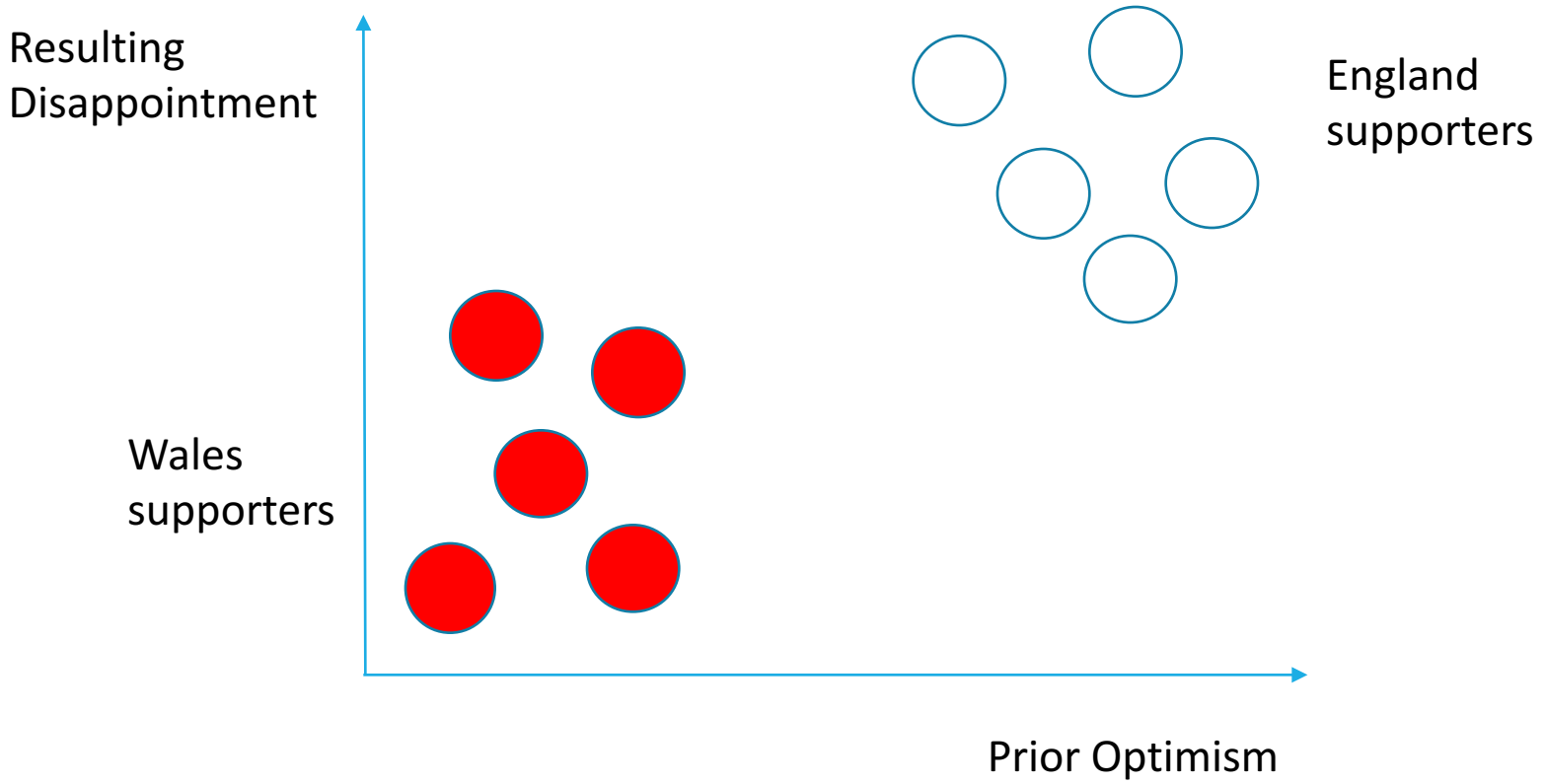


Unordered data



Differences in England and Wales supporters' feelings before and after Euro 2016

Unordered data



Differences in England and Wales supporters' feelings before and after Euro 2016

Ranking Visual Encodings

- In 1985, Cleveland and McGill published in the Journal of the American Statistical Association
- ‘Graphical Perception and Graphical Methods for Analyzing Scientific Data’
- They proposed basic guidelines for choosing an appropriate graphic form (or an appropriate encoding)
- The paper lists and ranks 10 ‘elementary perceptual tasks’

Cleveland and McGill

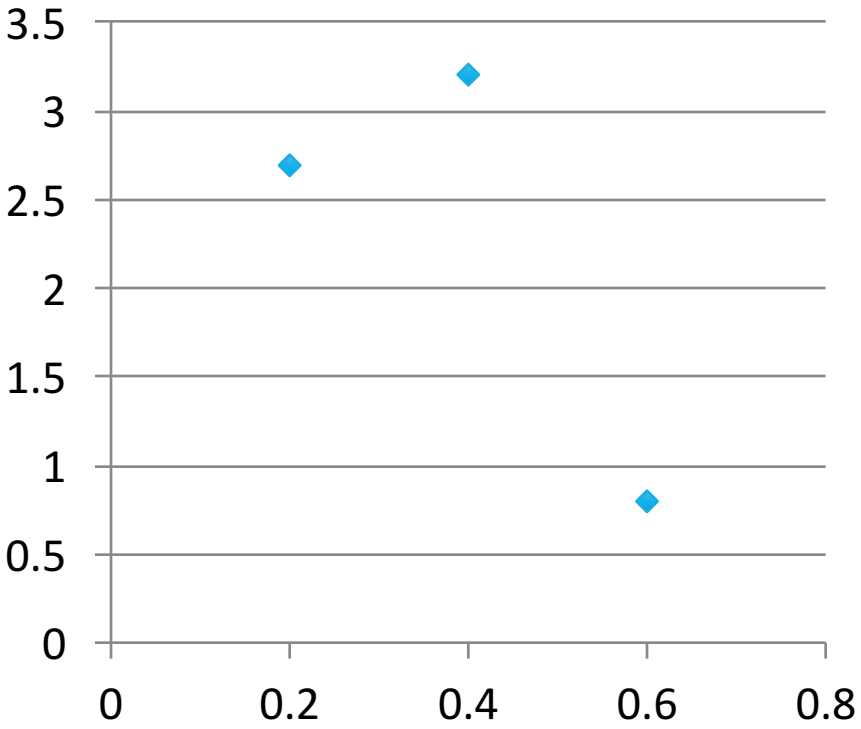
- Rank of graphic properties based on **human ability** to understand information
- These are particularly relevant for **detecting differences and making comparisons**
  - Position along a common scale
  - Position on identical but nonaligned scales
  - Length
  - Angle, slope
  - Area
  - Volume, density, colour saturation
  - Colour hue
- For accurate comparisons, graphical forms from the top of this list should be used.

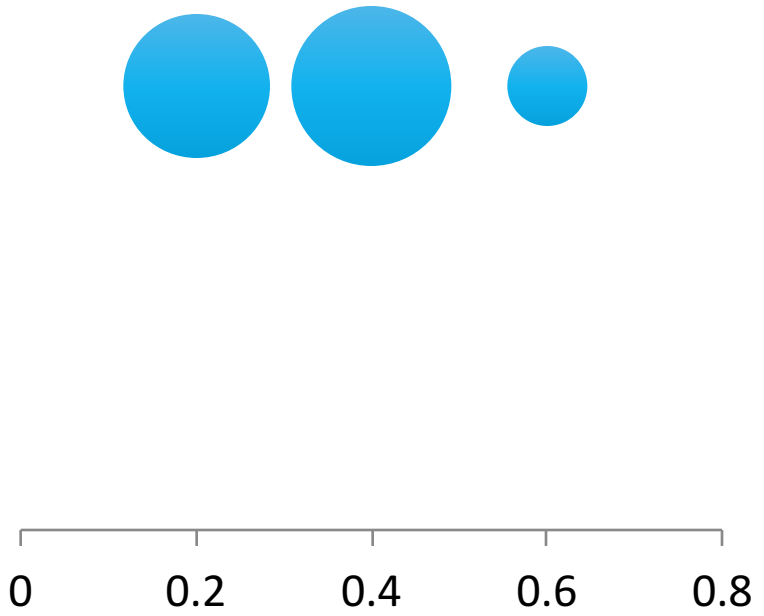
Choosing a Graphic for Visual Perception

- Question:
  - For a precise comparison, which type of chart would be most accurate?
    - Bar chart?
    - Bubble chart?
    - Heat map?

Cleveland and McGill - Example

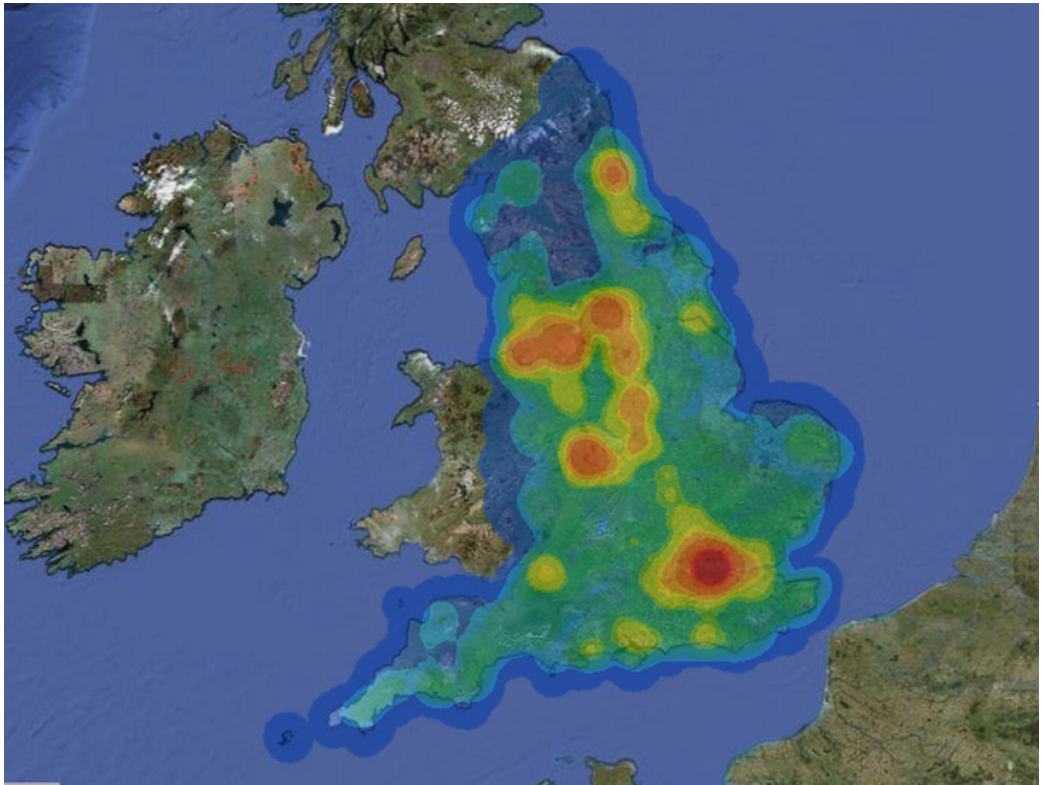
It is easier for humans to perceive the differences the values for X=0.2 and X=0.4 using position along a common scale (on the scatter plot), than by using area (in the bubble chart).



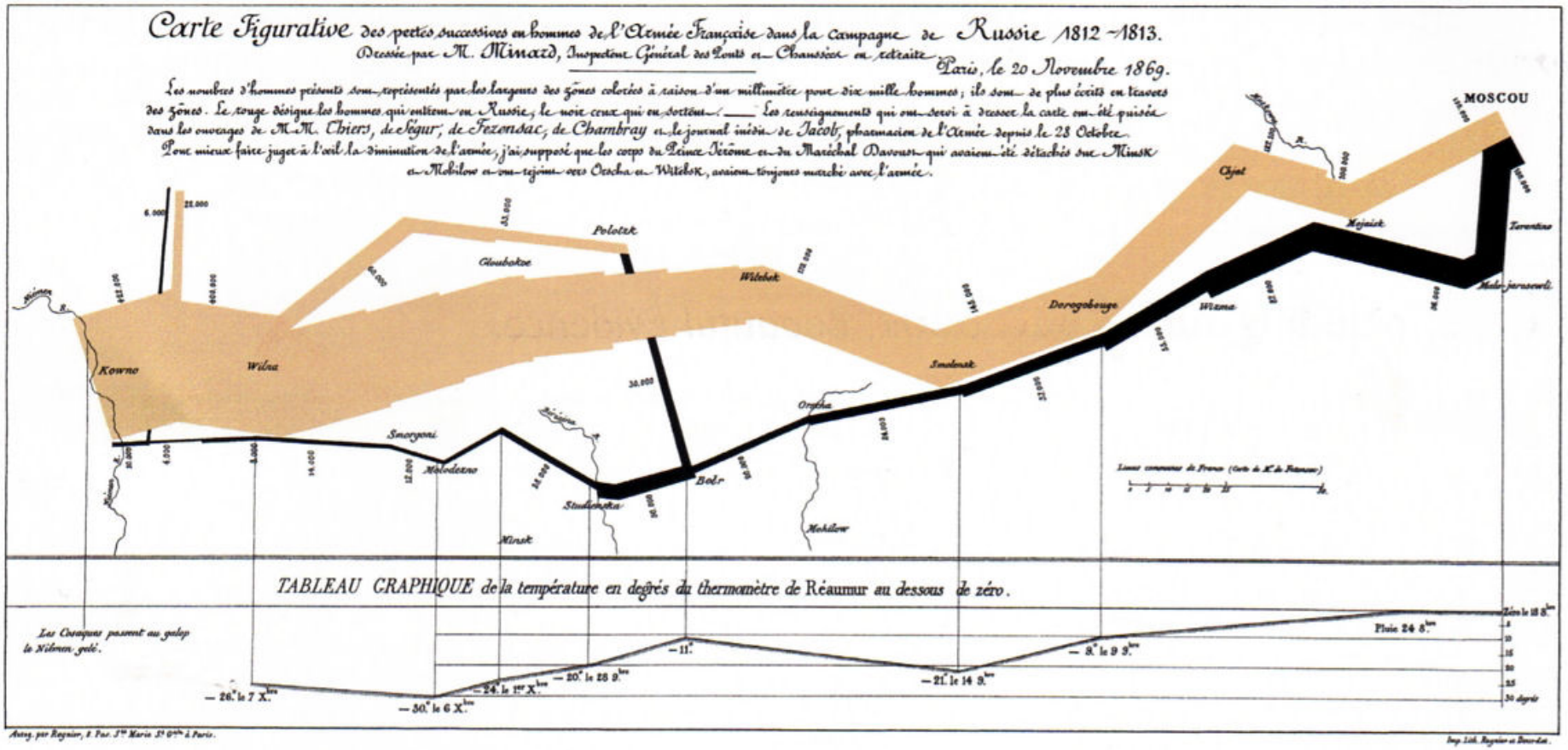


Choosing a Graphic for Visual Perception

- When precision is less important, lower ranked visual forms can be used well
  - E.g., spotting larger patterns



But really difficult to compare two similar areas!



How many dimensions can you find being represented on this map?

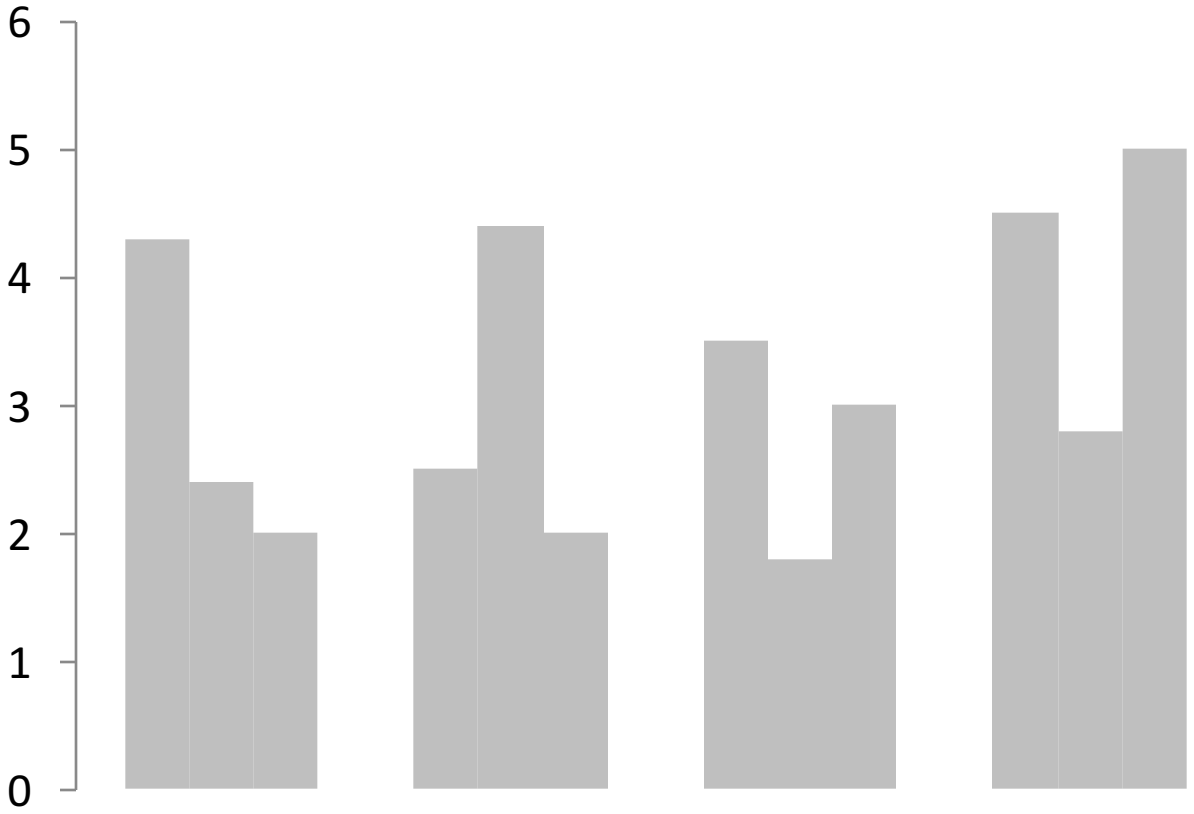
Modern version:
https://en.wikipedia.org/wiki/Charles_Joseph_Minard#/media/File:Minard_map_of_napoleon.png

Pattern Recognition

- As well as the encoding used, human's also perceive information based on how it is organised
- According to Gestalt Theory, the brain and visual system follow a number of principles for perceptual organisation
  - How the brain groups elements into 'patterns'
- This emerged as a school of thought from German psychologists in 1930s/40s

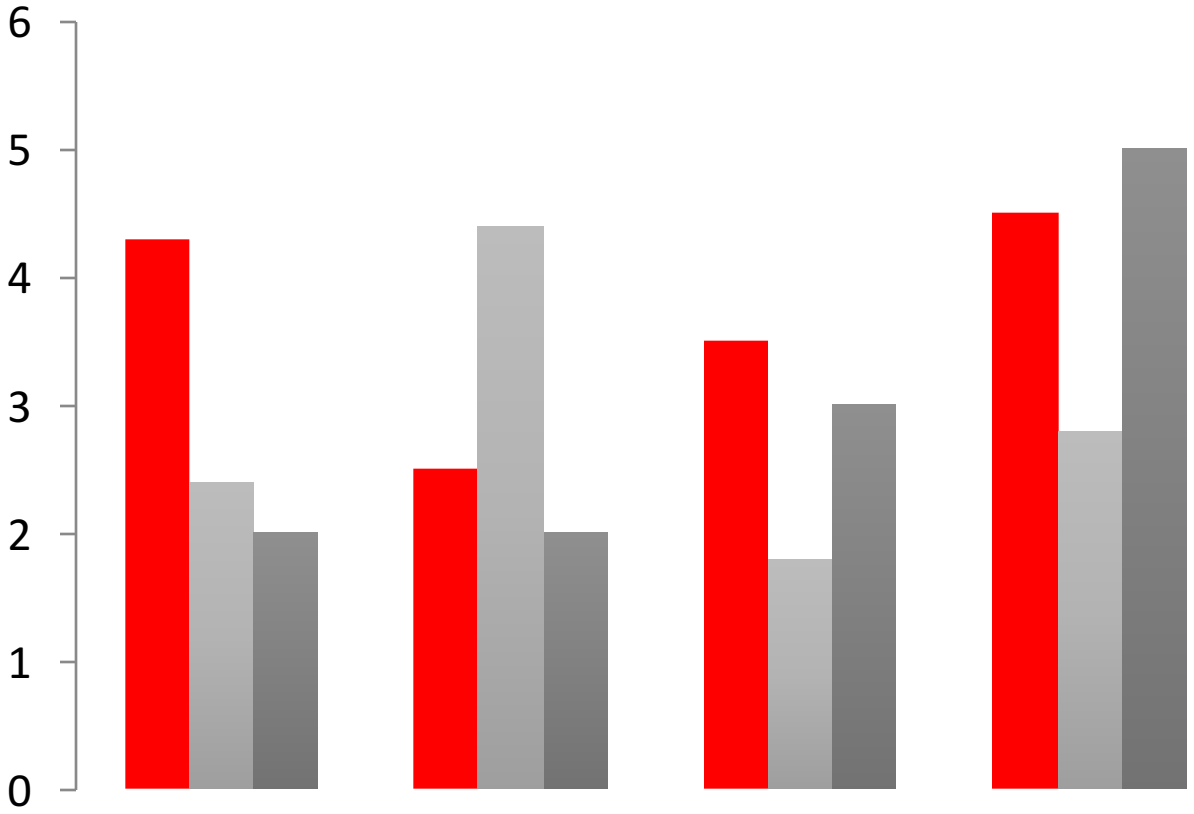
Gestalt Theory: Principles of Organisation

- Proximity – objects that are close are perceived to be natural groups



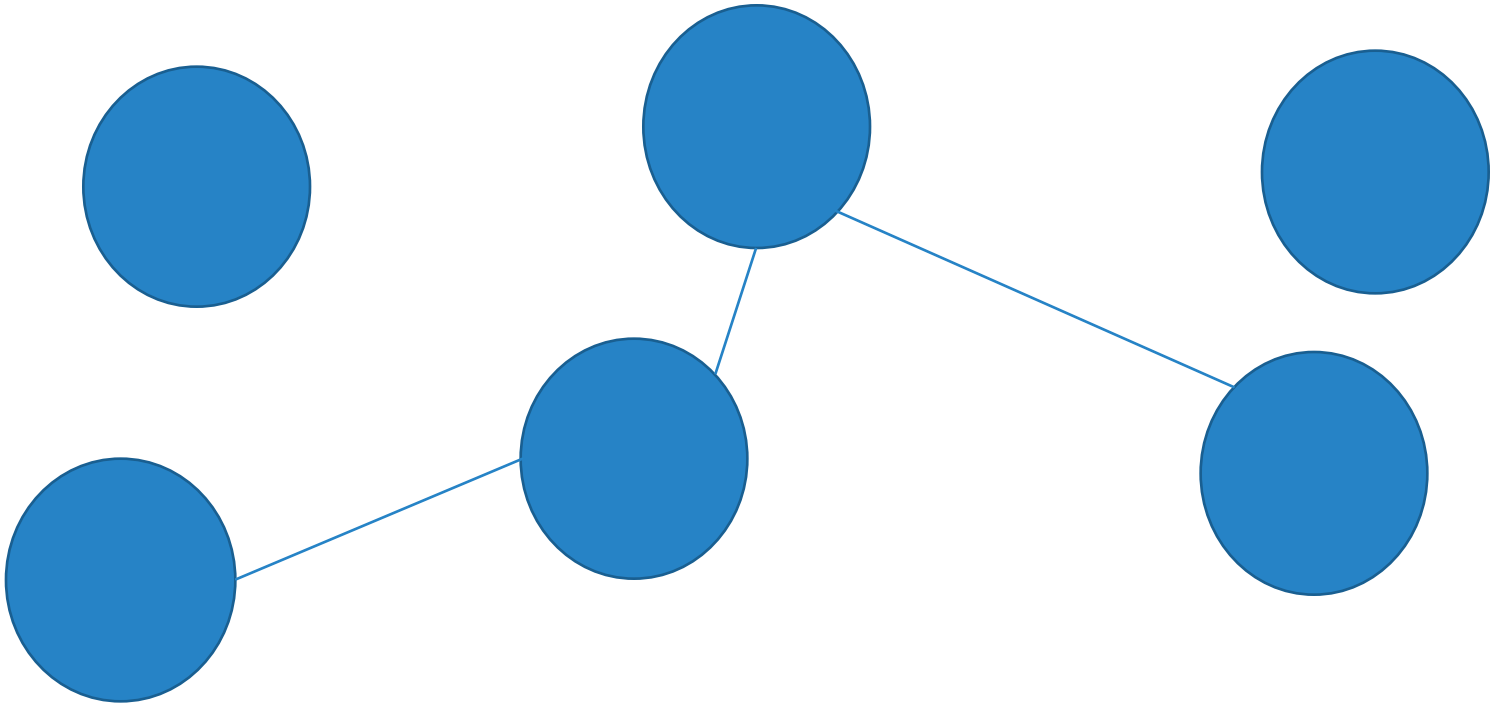
Gestalt Theory: Principles of Organisation (2)

- Similarity – identical or similar objects belong to a group



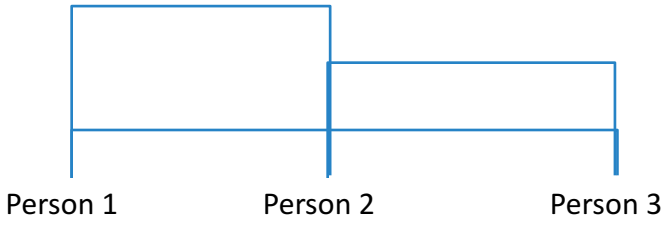
Gestalt Theory: Principles of Organisation (4)

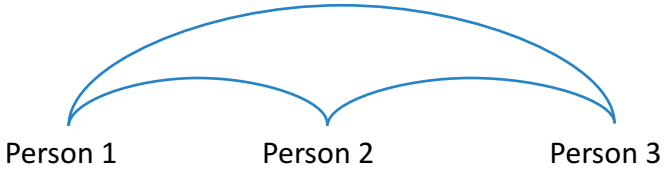
- Connectedness – linking using a line or similar



Gestalt Theory: Principles of Organisation (5)

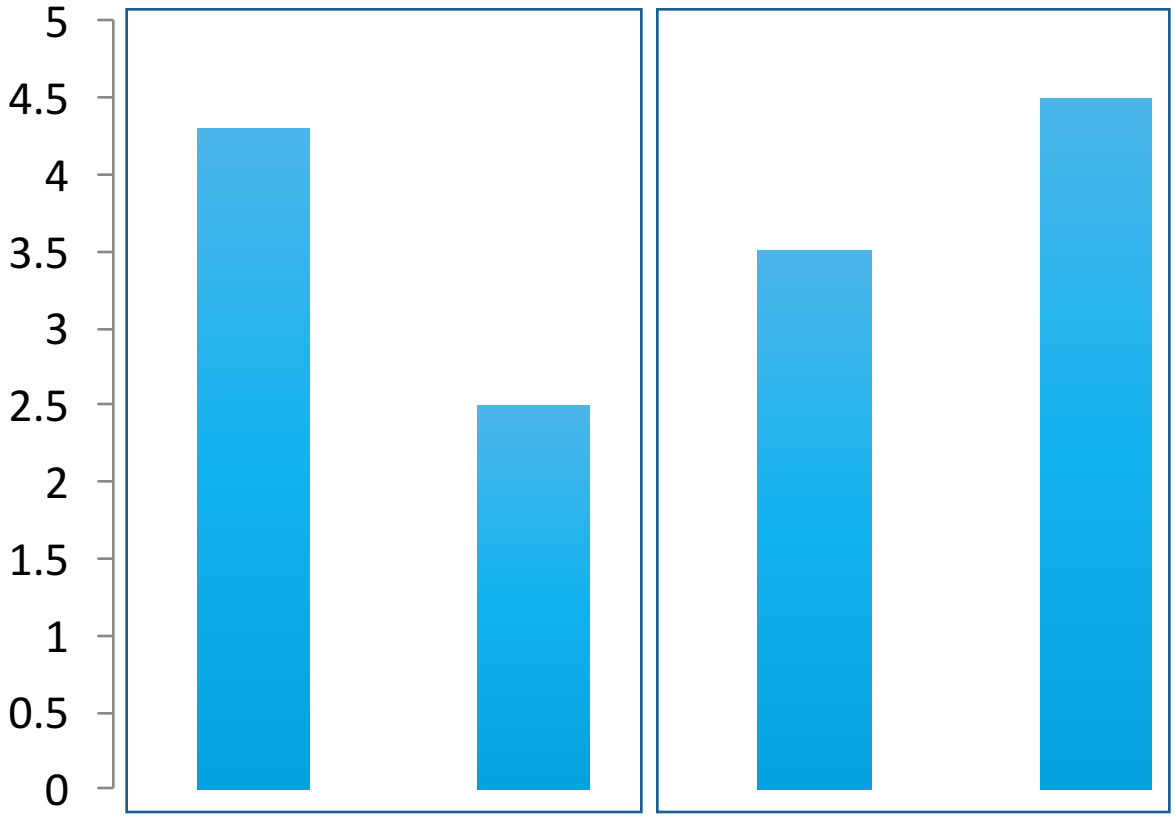
- Continuity: Smooth contours are easier to perceive than sharp angles





Gestalt Theory: Principles of Organisation (6)

- Closure: bounded areas indicate a grouping

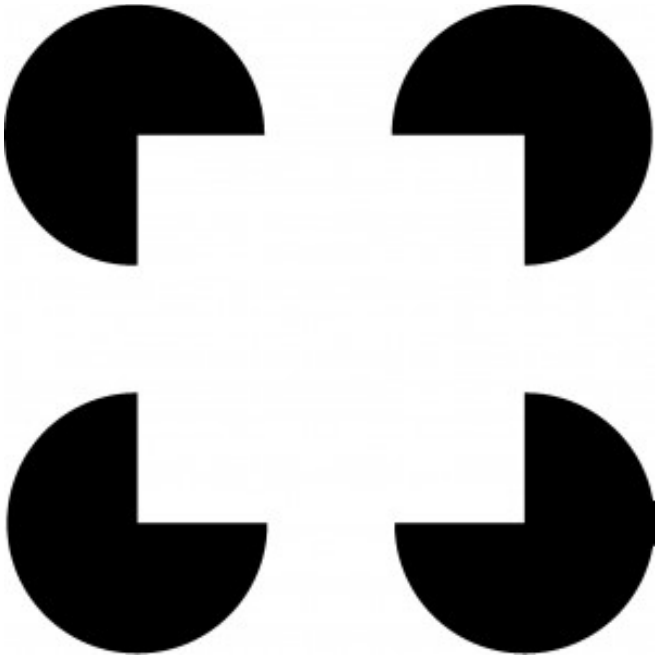


Gestalt Theory: Takeaway Message

- Graphics can be made more functional using simple techniques around organisation and layout of components
- Conversely, not using Gestalt principles correctly might mislead users as they will make wrong assumptions about the components you display

Deceiving your brain

- As we find out more about how the brain works, and how it makes assumptions about what it already knows, we can learn how to lie to it
  - Or more importantly, how to **avoid** lying to it
- E.g. What do you see?



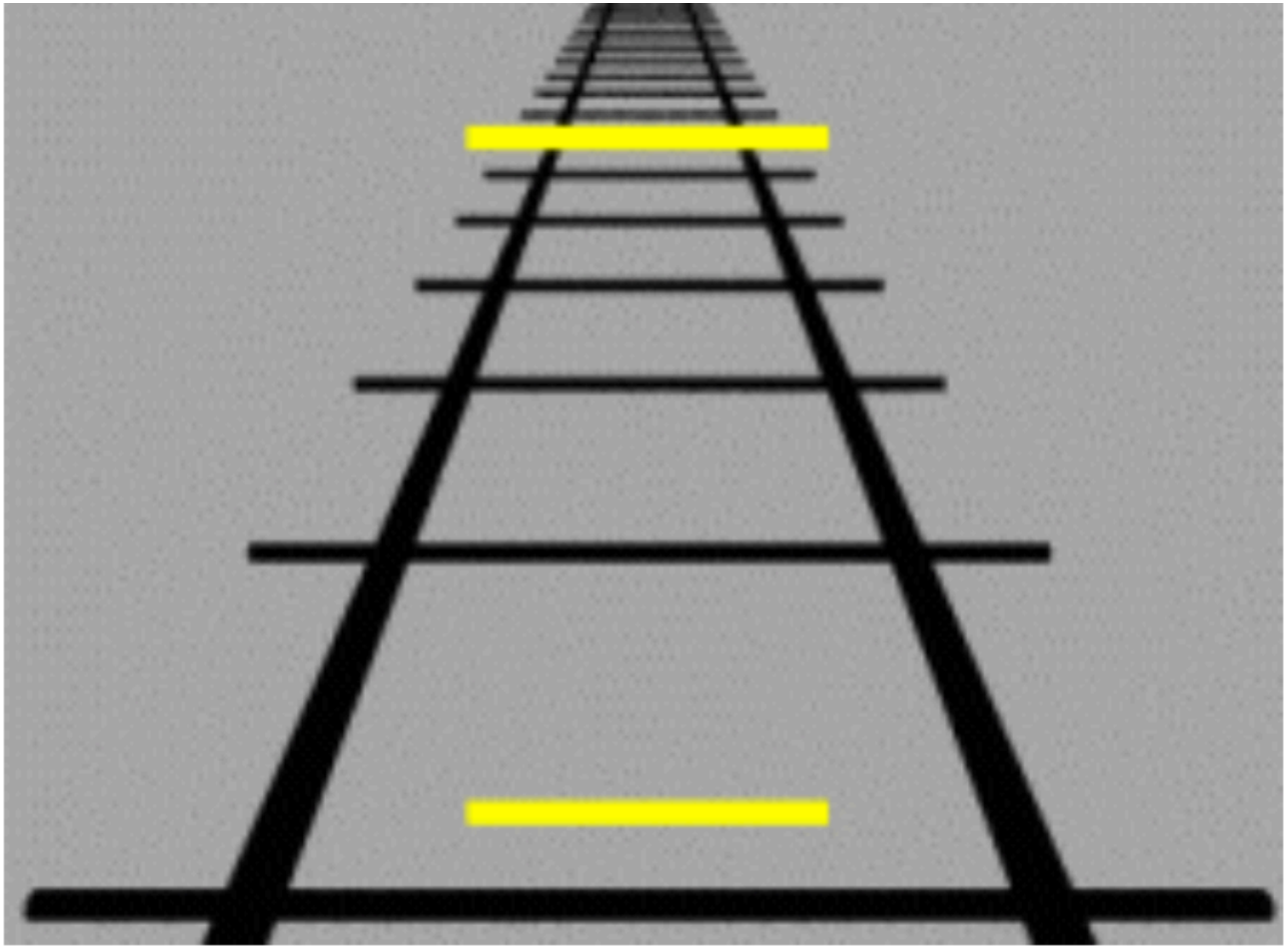
Who saw a white square over four black circles?

Your Brain is Deceiving You!

- There is no white square.
- Remove the 4 ‘pacman’ shapes, and there is nothing there!

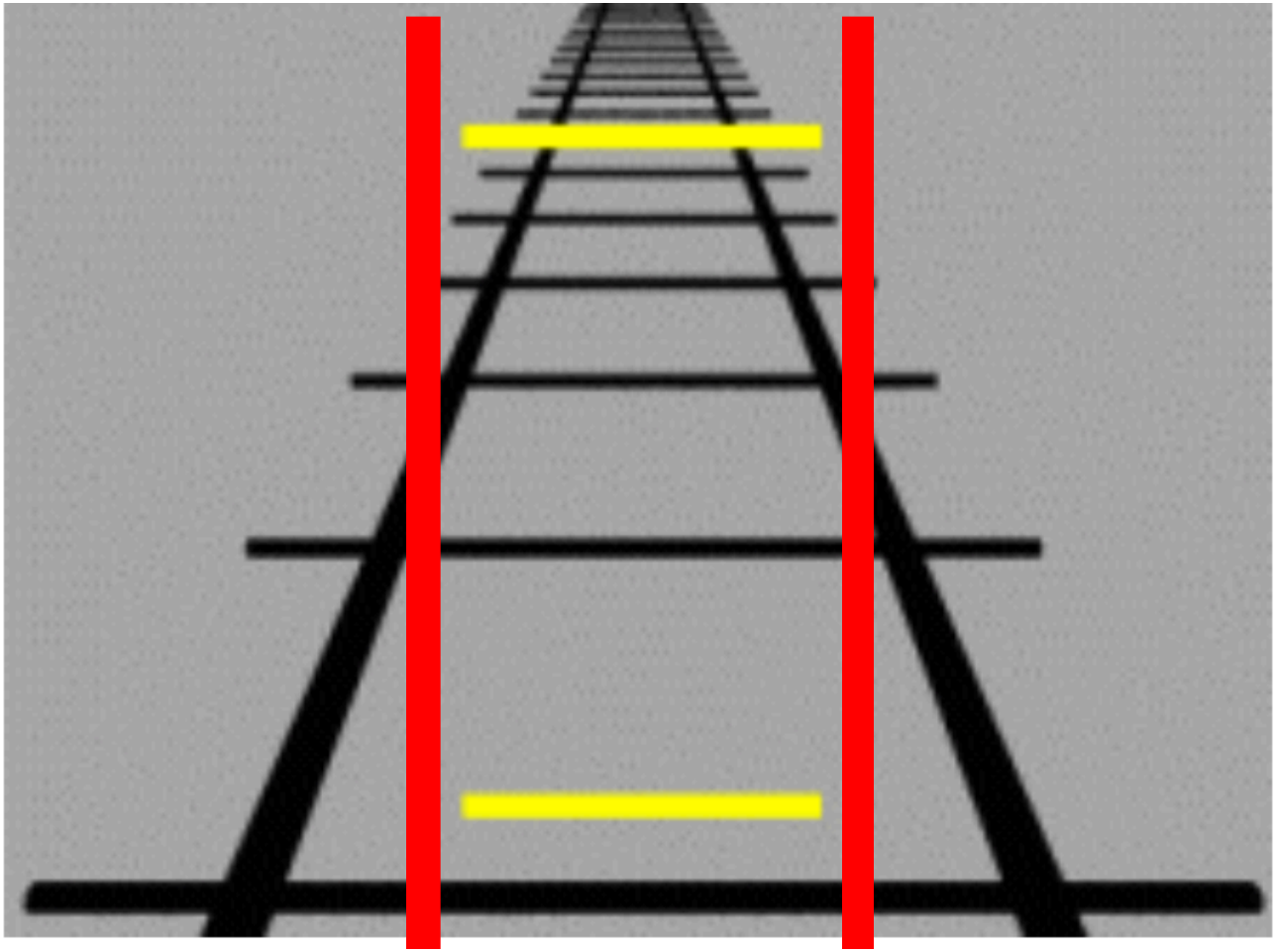
- Your brain is making assumptions based on what it knows about the world already
  - “Empty areas” don’t exist
- The brain ‘perceives’ more than what the retinas in your eyes actually see

Which yellow line is longer?



http://thebrainbank.org.uk/wp-content/uploads/2012/08/Ponzo_illusion.gif

Which yellow line is longer? (2)



http://thebrainbank.org.uk/wp-content/uploads/2012/08/Ponzo_illusion.gif

Don’t stretch the truth!

Tufte’s Lie Factor

Warning!

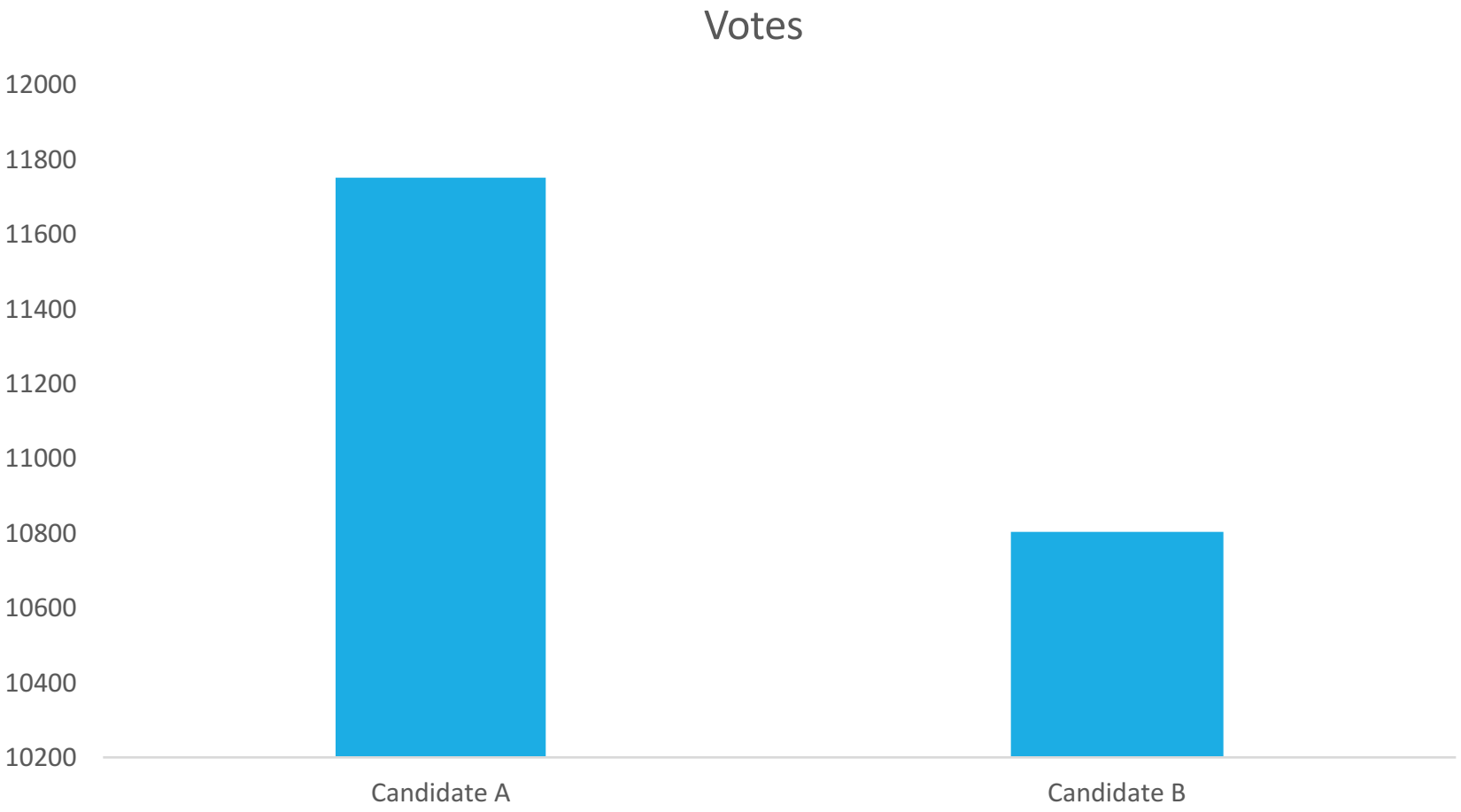
The following slides contain details on how to do “bad” things. Use this knowledge for good, and avoid these mistakes!

“With great power comes great responsibility”

Be Objective

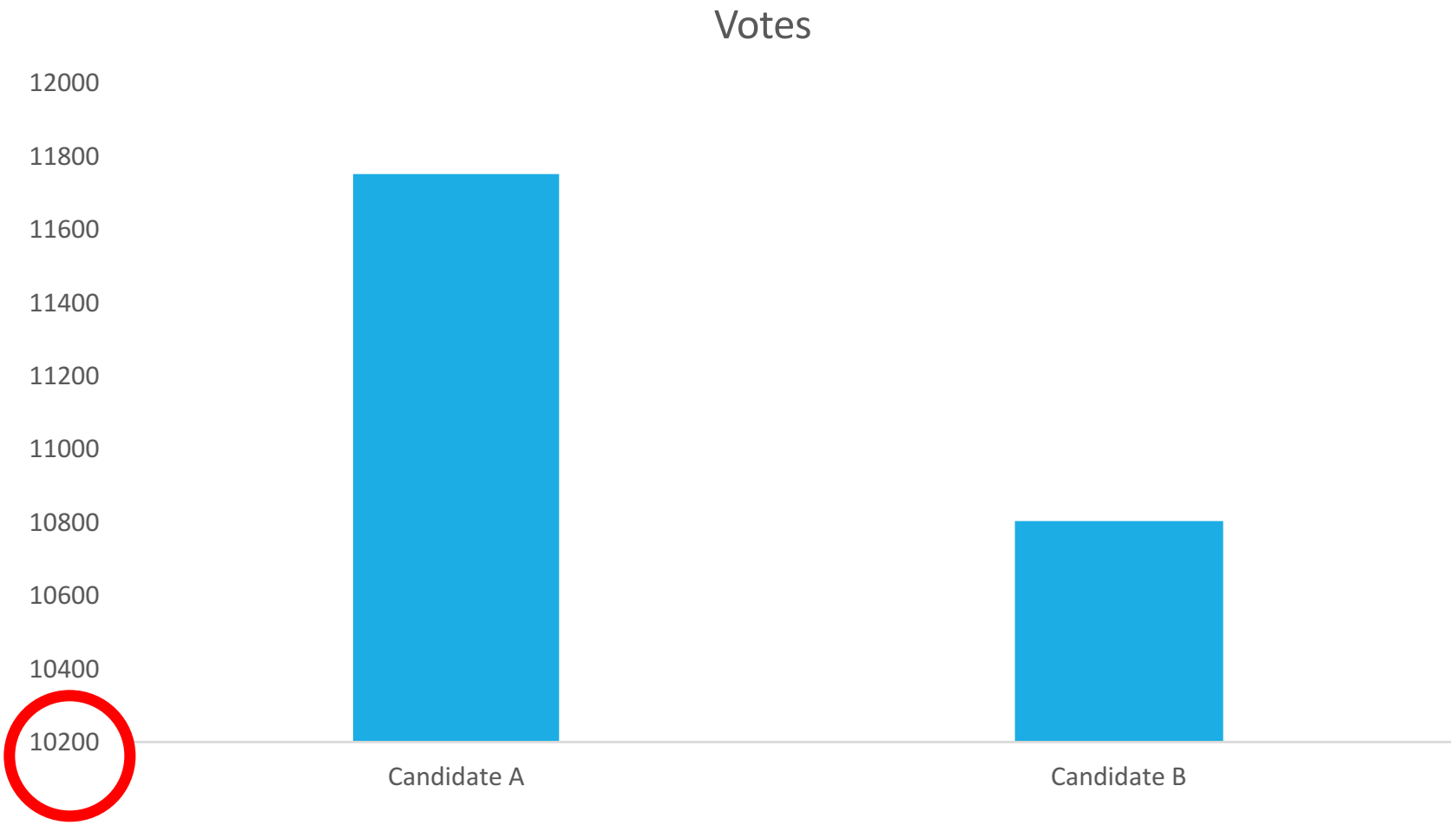
- Tufte created a formula to calculate the 'lie factor' or how 'misleading' a graphic is
- Lie factor = $EffectSizeInGraphic / EffectSizeInData$
  - where EffectSize = $(value2 - value1) / value1$
- A reliable graphic should be between 0.95 and 1.05.

Lie Factor: An Example



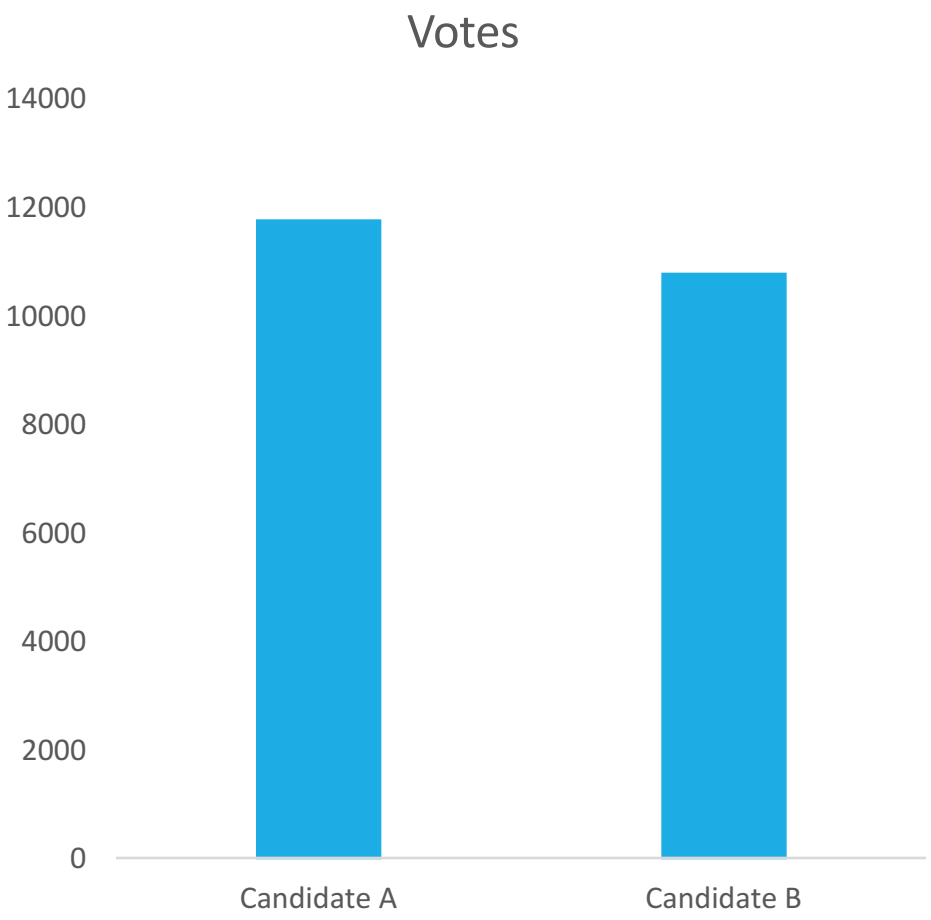
- This suggests that A received around three times as many votes as B.

Lie Factor: An Example



- This suggests that A received around three times as many votes as B.

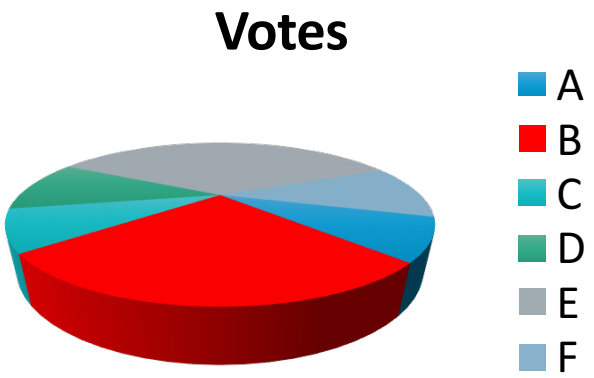
Lie Factor: An Example



- In reality, things are much closer.
- The actual values suggest around 11,750 for A, and 10,800 for B
  - About 13.2% difference
- The original graph however, makes the difference seem 61.3% between the two values
- **0.613 / 0.132 = 4.64**
- The original chart therefore had a high lie factor and is overstating the effect.
  - Well over 1.05!

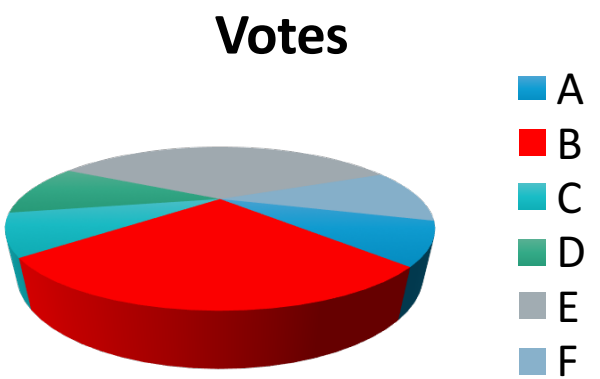
Pie charts + 3D are particularly bad

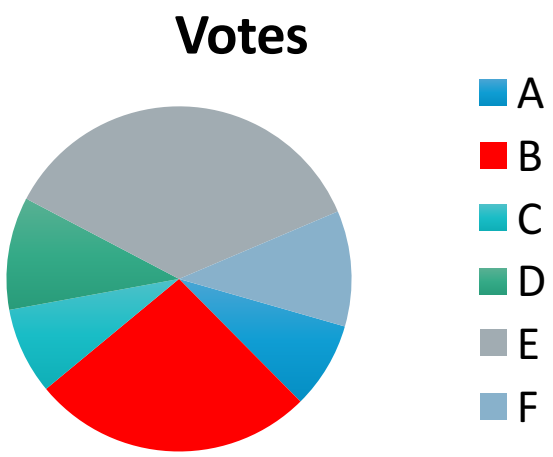
- To communicate a message clearly and effectively.
- A graph or chart should clarify what numbers in a table are saying.
- 3D does NOT help with this.



Pie charts + 3D are particularly bad

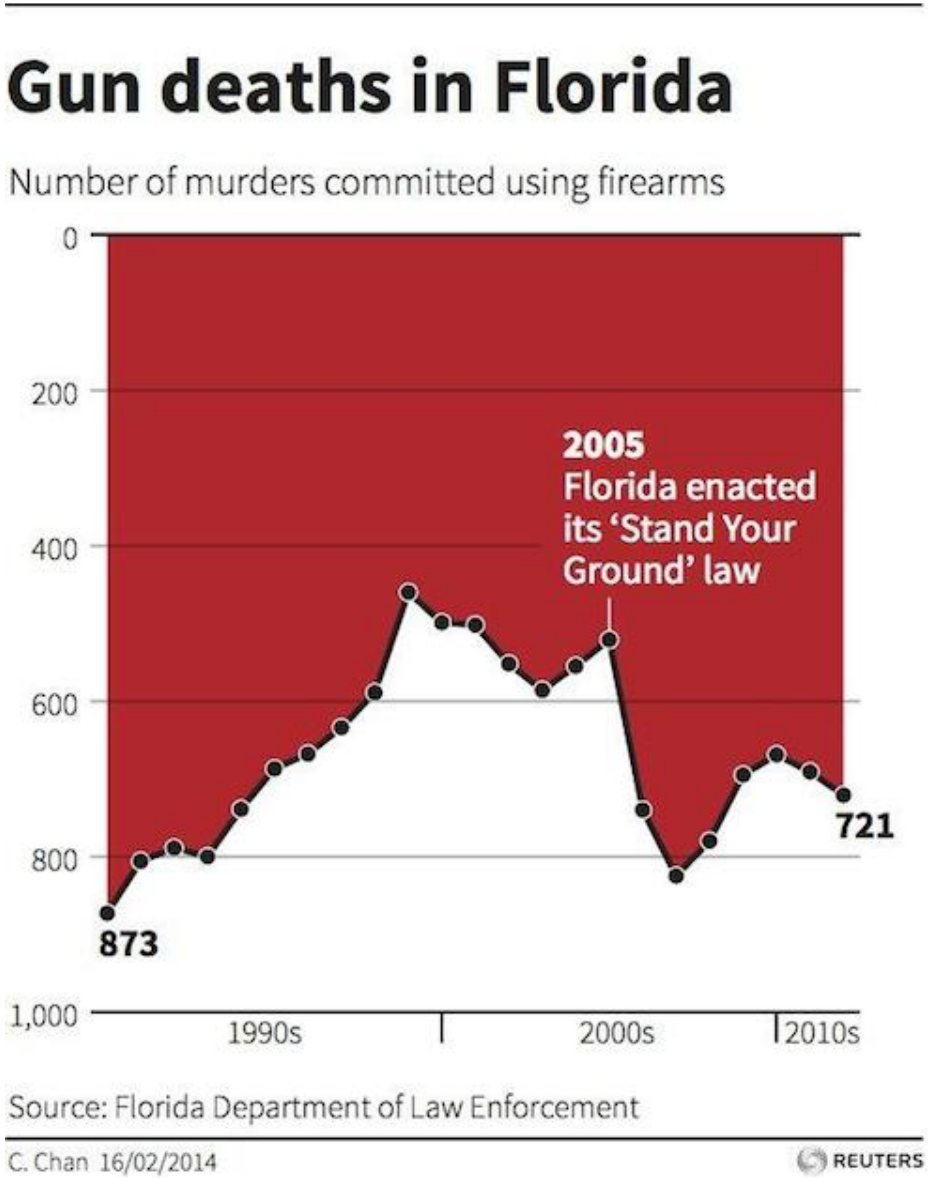
- To communicate a message clearly and effectively.
- A graph or chart should clarify what numbers in a table are saying.
- 3D does NOT help with this.





Party	Votes
A	60
B	195
C	60
D	78
E	265
F	80

What makes a visualisation a ‘bad’ visualisation?

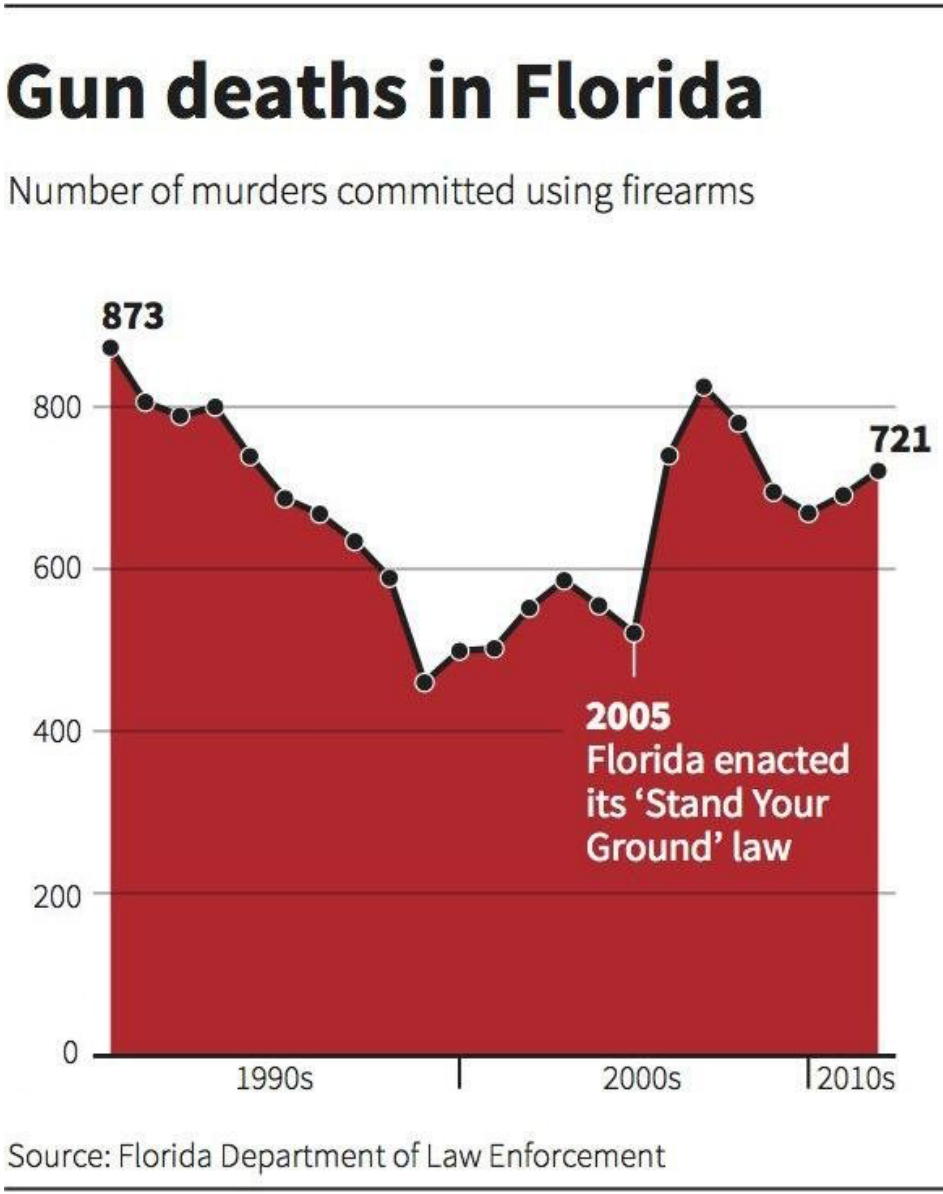


EXERCISE
In pairs

Discuss whether you think the visualisation is good or bad?

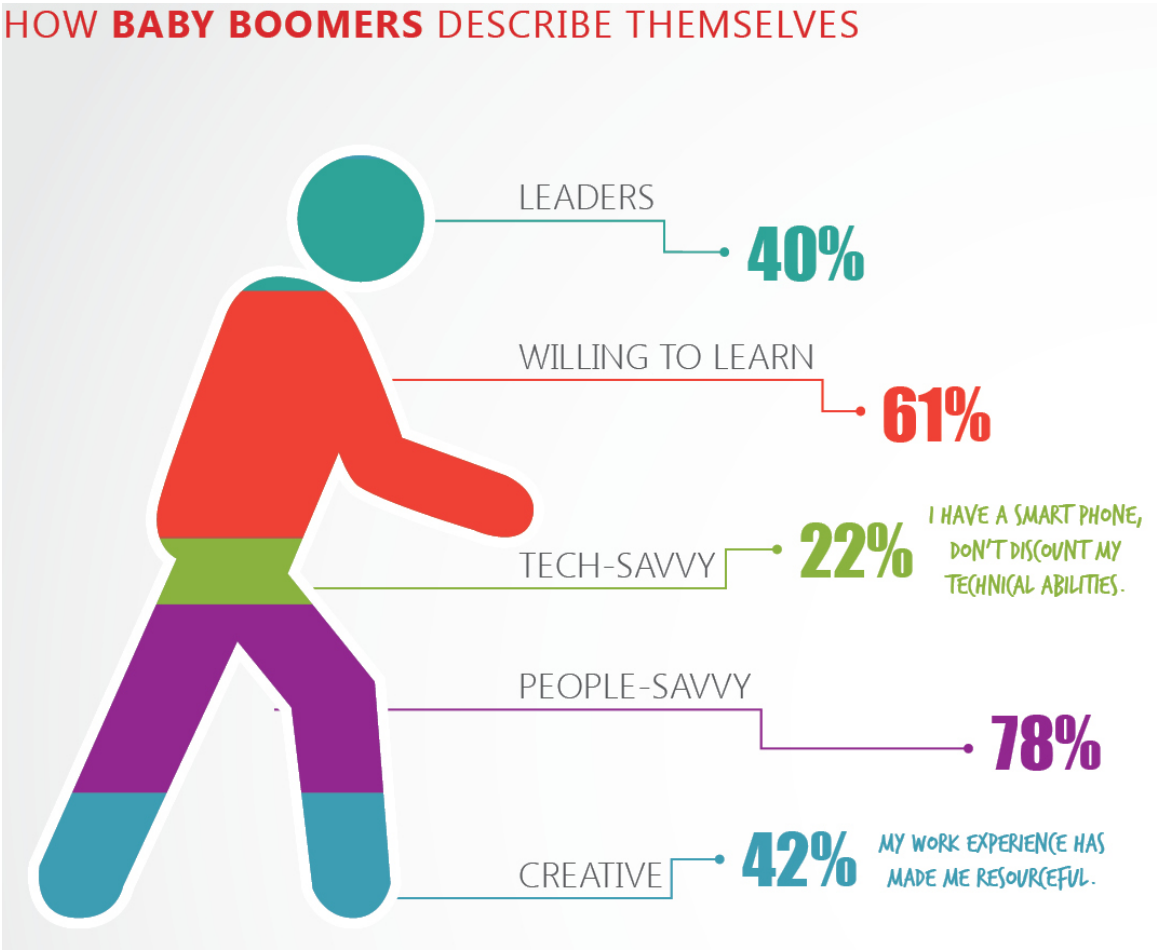
Why?
What is the story that it is trying to tell?

What would you change in order to enhance the story?



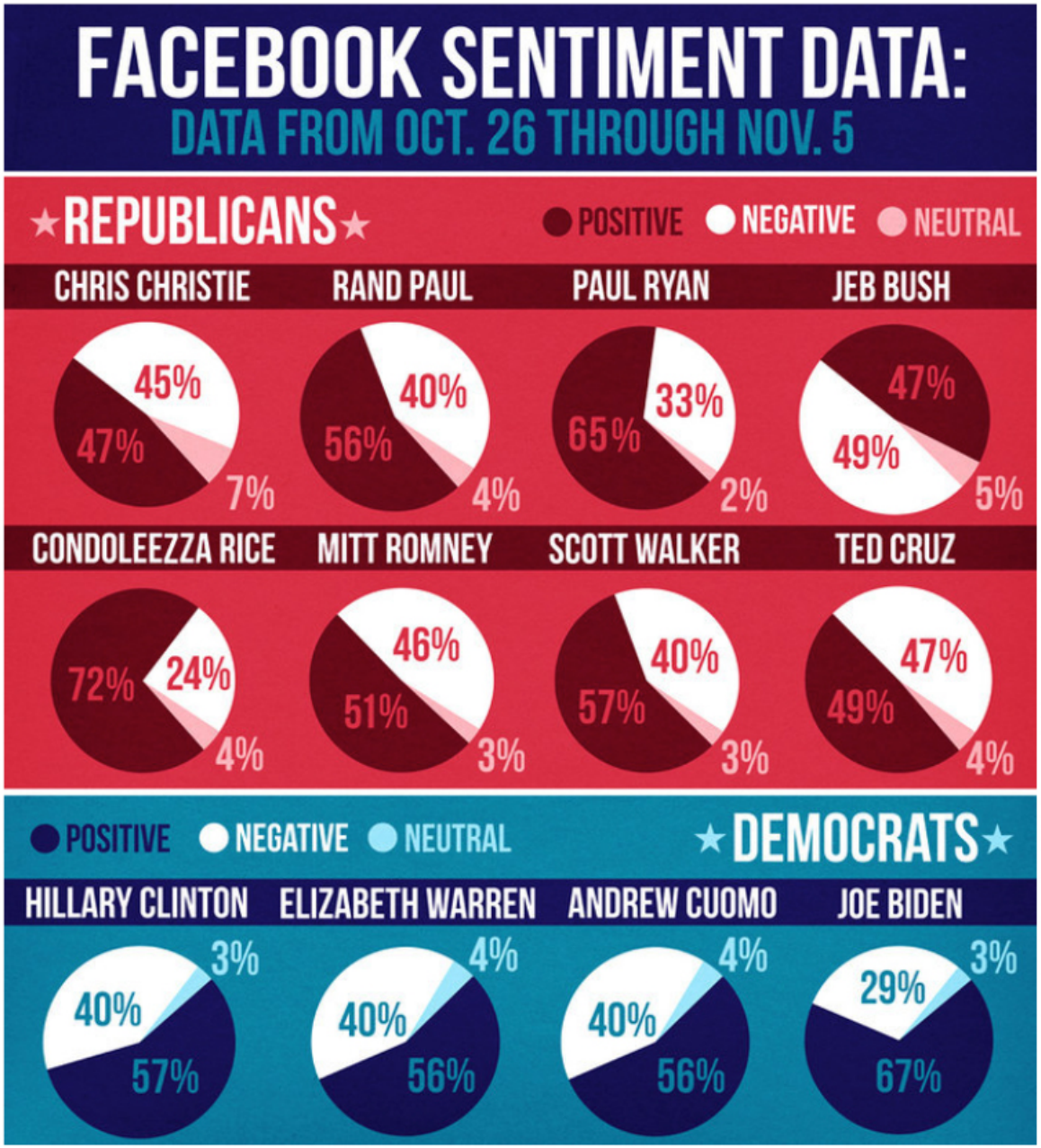
The "fixed" version…

http://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2?IR=T

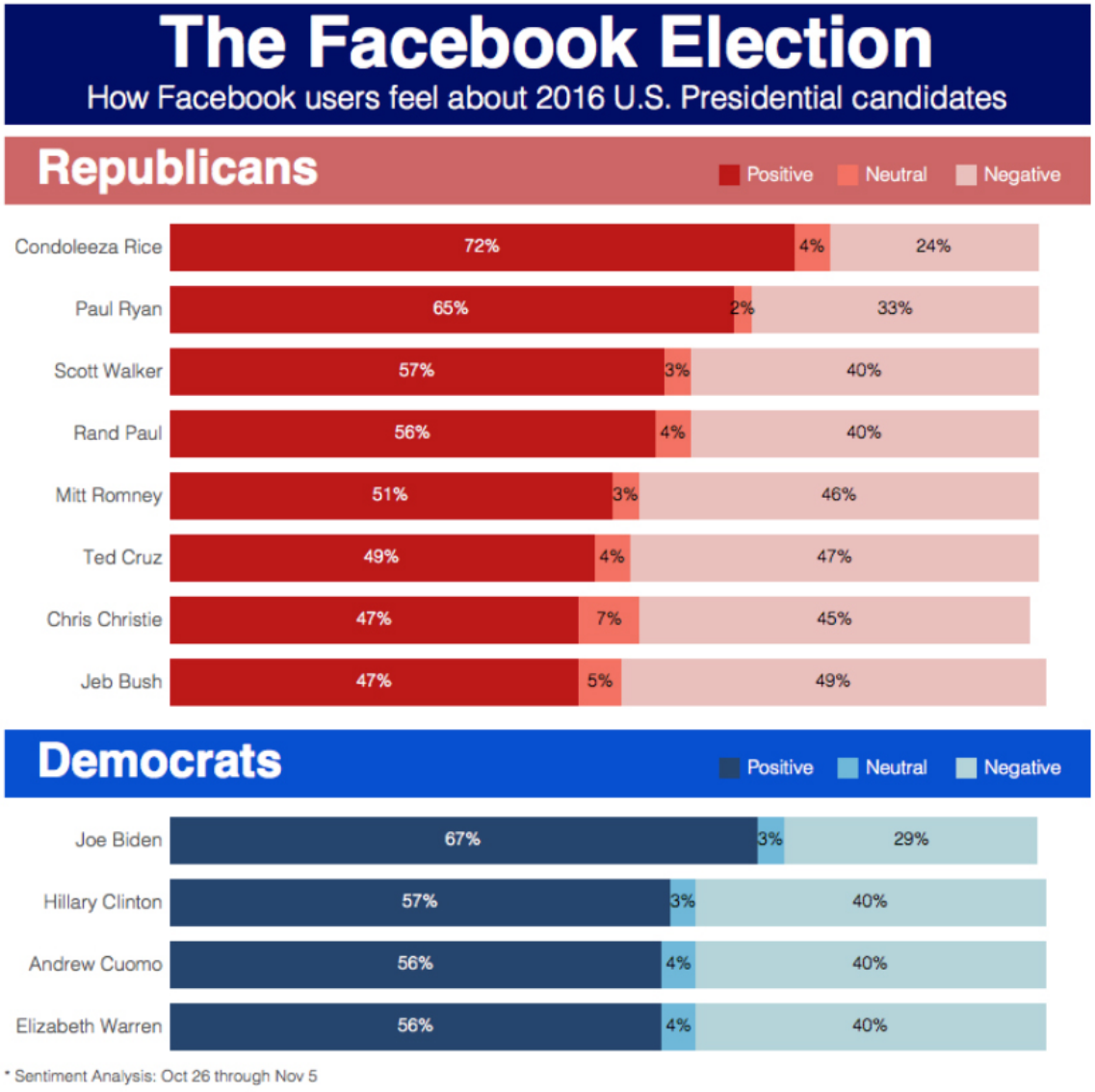


Let's find all the things wrong about this visualisation

https://rstudio-pubs-static.s3.amazonaws.com/70948_b5daeeeecf824047b580c1a101407a98.html



- What problem does this solve?
- Is this the best way to display the information?
- Does *everything* add value?
- Are there clear labels?



Title is clear and addresses the question

Bar charts allow for a clearer comparison of values

Labels are simpler and more consistent

Sorting of bars adds another visualisation element

Thanks to Thierry Driver (Tableau) for this example

Data Vis Technologies

Web technologies and their role in visualisations

- Even if you're using a high-level, online tool that handles the code for you, it is pertinent to understand the underlying technologies of the Web.
- These will be 'manipulated' to create many types of graphics.
- Covering these technologies in depth is not the focus of the course. The following is a (very) brief introduction.

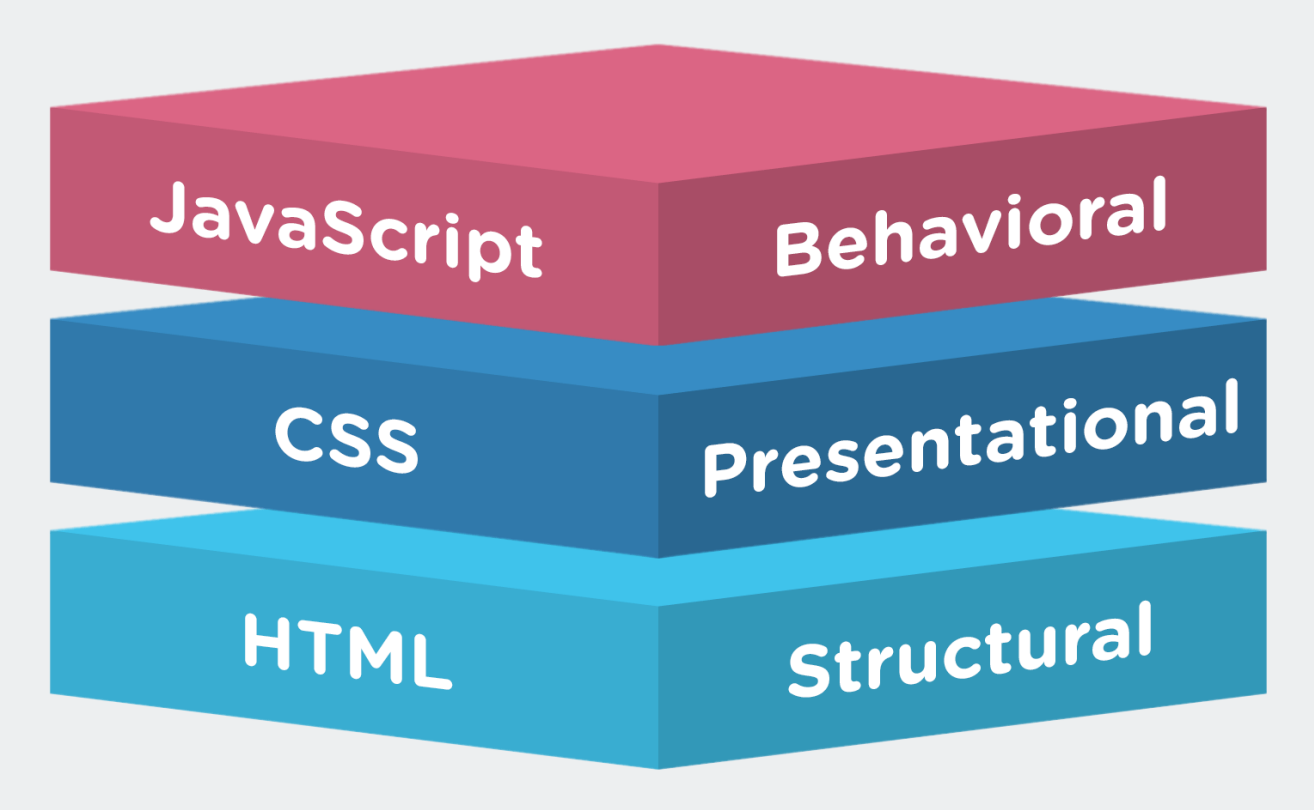


Image from http://blog.teamtreehouse.com/wp-content/uploads/2014/11/progressive-enhancement.png

Hypertext

- Hypertext: text which contains links to other texts.
- HTML (Hypertext Markup Language)
  - A language that defines the basic structure and content of a Web page using **markup** 'tags'.
  - **Each tag describes a different type of content**
    - E.g.,
      - <html></html> For an html page
      - <body></body> For the body of an html page
      -   A paragraph
      - The BBC A link to the BBC
  - Read by a Web Bbrowser, which decides how to display the content.

CSS (Cascading Style Sheets)

- Define the **design and presentation of an HTML document**.
- Tells the Web browser more information about the look and feel of each HTML element – how should it be displayed on the screen?
- E.g.,

```
p {
  font-family: “Times New Roman”;
  color: blue;
  font-size: 12px;
}
```

“Make all paragraphs use Times New Roman font, size 12, and colour the text blue.”

Javascript

- The third essential technology behind the WWW.
- The programming language of the Web.
- Programs the ‘behaviour’ of the Web page.
- Client-side language (runs on the user’s machine, not on the server).
- Define how a page should behave when there is a user event (e.g., clicking on something),

Javascript with HTML

- When the browser displays an HTML page it creates a hierarchical structure of the HTML elements.
  - This is called the Document Object Model (DOM).
- Javascript can access this so that it can change specific styles of particular elements
- The interaction with the DOM is key to how **D3** is used to make visualisations.

D3.js (Data-Driven Documents)

- Javascript library (“.js”).
- Based on standard Web technologies.
- Uses HTML, SVG, and CSS to create visualisations from data.
  - Drives the connection between **data** (provided by user) and **documents** (rendered by the Web browser).

General steps

- Loading the data into browser's memory.
- Binding the data to elements within the document.
- Transforming elements by setting visual properties based on the bound data.
- Transitioning elements between states
  - Response to user input.

D3

- Dynamically access the DOM behind a webpage to apply styles
  - DOM:
    - Document Object Model
    - Hierarchical Structure of HTML elements
    - Allows things like
      - d3.select(“body”).append(“p”).text(“New Paragraph”);

```
<!DOCTYPE html>
▼<html lang="en">
  ▼<head>
      <link href="index.css" rel="stylesheet">
      <title>Playing with D3</title>
    </head>
  ▼<body>
    ▼<div class="tweet-display">
        <div class="author"></div>
        <div class="text"></div>
      </div>
      <script src="bower_components/jquery/dist/jquery.min.js" type="text/javascript"></script>
      <script src="bower_components/underscore/underscore.js" type="text/javascript"></script>
      <script src="bower_components/d3/d3.min.js" type="text/javascript"></script>
      <script src="index.old.js" type="text/javascript"></script>
    ▶<svg>…</svg>
    </body>
</html>
```

D3: Chain Syntax

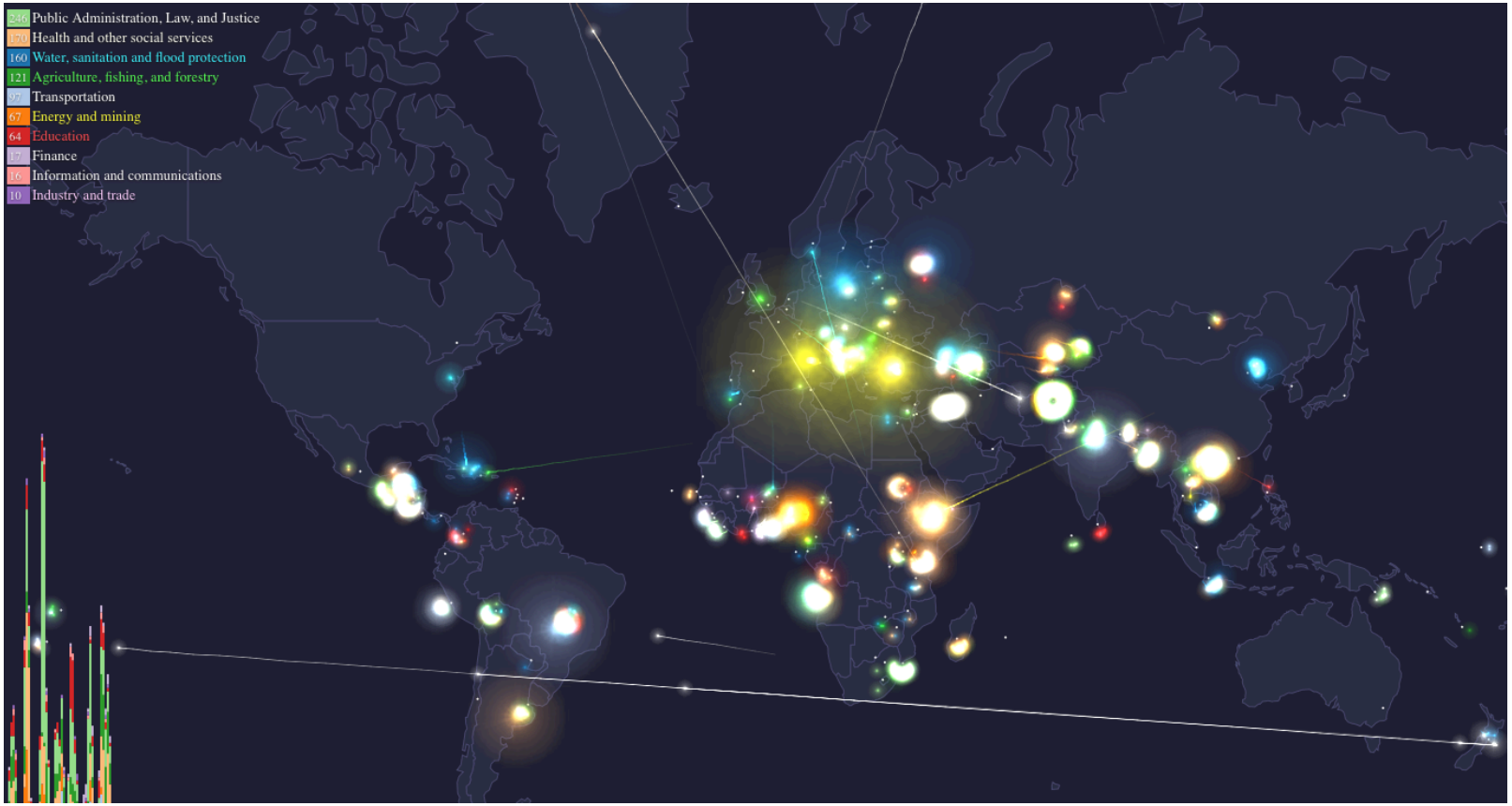
- Using “.”s to chain methods together
- Perform several actions in a single line of code
- Fast and easy, but can cause debugging problems later
- Javascript doesn’t care about whitespace or linebreaks
  - d3.select(“body”)
    .append(“p”)
    .text(“New Paragraph”);

D3: Chain Syntax

- d3.select(“body”).append(“p”).text(“New Paragraph”);
  - First passes ‘select’ a CSS selector – “body”, returning the first element in DOM that matches.
  - Then creates a new DOM element “p” and appends it to the end of the previously selected element. So a paragraph is appended to the body element of the document in this case. The new element is passed on in the chain.
  - To the the text() function which takes a string and inserts it into the currently selected element (the “p” element as this was passed from append).

D3 Example

- http://d3.artzub.com/wbca/
- Interactive visualisation of World Bank contracts



Tableau

- For the hands-on session after lunch we will use Tableau
  - Drag and drop visual analytics package
  - Connect to data from multiple sources without any code
  - Quickly build visuals to explore the data and spot trends
  - The create dashboards and stories to present your data and tell compelling narratives
- There is a license key for you all to use, and you can use Tableau for the rest of the week in your projects
  - Or, if you want to use D3 there are a wide range of tutorials available online

Tableau Public

- Tableau Desktop lets you publish your workbook to ‘Tableau Public’
- Here you can be added to their gallery of projects, with your visualisations hosted live online
  - Remain connected to the data so they are fully interactive
  - Work exactly as they do in TD
  - Can create compelling data stories and share them with your audience simply and easily

Summary

- The point of data visualisation is to communicate what is important in your data, clearly and effectively
- You may have a particular story to tell, or it may be that you want each reader to find their own story in the data
- Regardless, you want to highlight the key elements of the data, and not distract the reader with chartjunk or unnecessary ‘art’
- We can use knowledge about how the brain works to organise content on the page/screen effectively to enhance understanding
- A good visualisation should not **lie**.
- There are a huge range of visualisation technologies. D3 is popular for coding them. Tableau is a popular high-level desktop tool.

Further Reading

- Tufte, Edward. “The Visual Display of Quantitative Information” 2nd Edition
- Cleveland, William S., and Robert McGill. "Graphical perception and graphical methods for analyzing scientific data." *Science* 229.4716 (1985): 828-833.
- Murray, Scott. *Interactive data visualization for the Web*. " O'Reilly Media, Inc.", 2013.
- D3 Tutorials such as https://www.dashingd3js.com/table-of-contents
- The D3 gallery for adaptable examples:
  - https://github.com/mbostock/d3/wiki/Gallery
- Cairo, Alberto. The Functional Art: An introduction to information graphics and visualization. New Riders, 2012