

Best Practices – Principles of Data Visualization



© 2018 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of NUS ISS, other than for the purpose for which it has been supplied.

Contents

- Pre-Attentive Attributes
- Gestalt Principles
- Graphical Excellence
 - Well designed presentation of interesting data
 - Complex ideas communicated with clarity, precision and efficiency
 - The truth about data
- Graphical Integrity
 - Lie factor
 - Show data variation, not design variation



Human Perception

**Understanding and
communicating patterns
in raw data can be
difficult...**

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

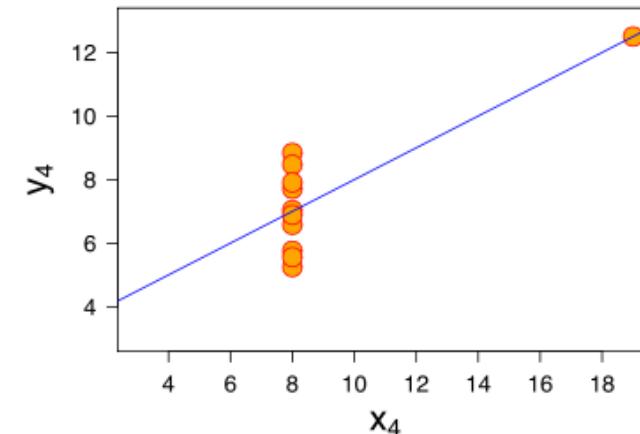
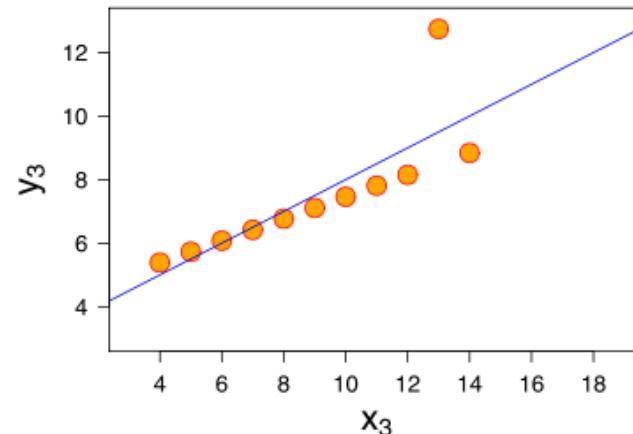
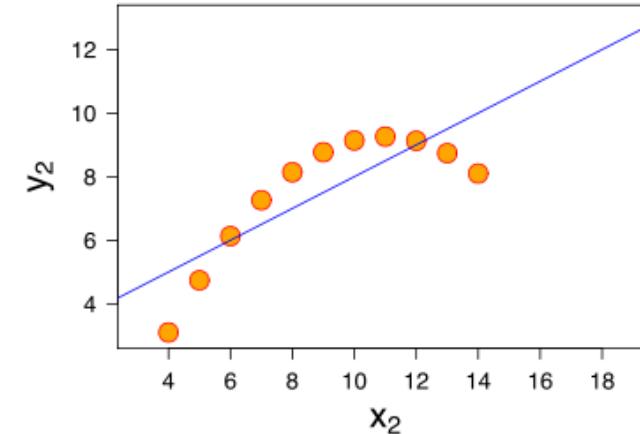
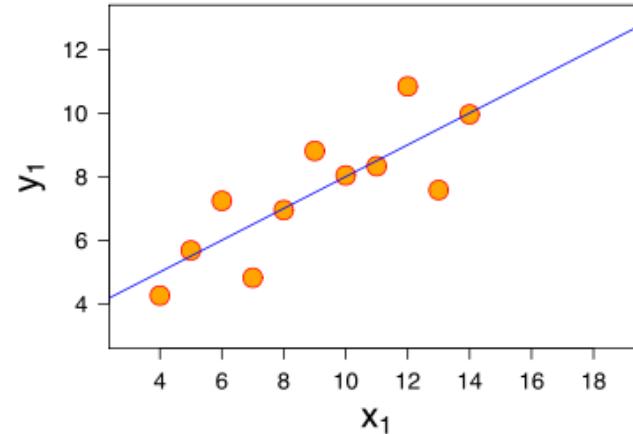




Human Perception



**The same information
expressed 'visually' is far
easier to understand,
interpret, and communicate..**



Visual + Narrative = Engagement



- Aesthetics look and feel
- Visual art
- Psychology
- Cognitive

Pre-Attentive Attributes



Data visualisation works because they use
'Pre-Attentive Attributes' to code the data...



Pre-Attentive Attributes are detected and processed immediately by the brain, without the need for focused attention. They are...

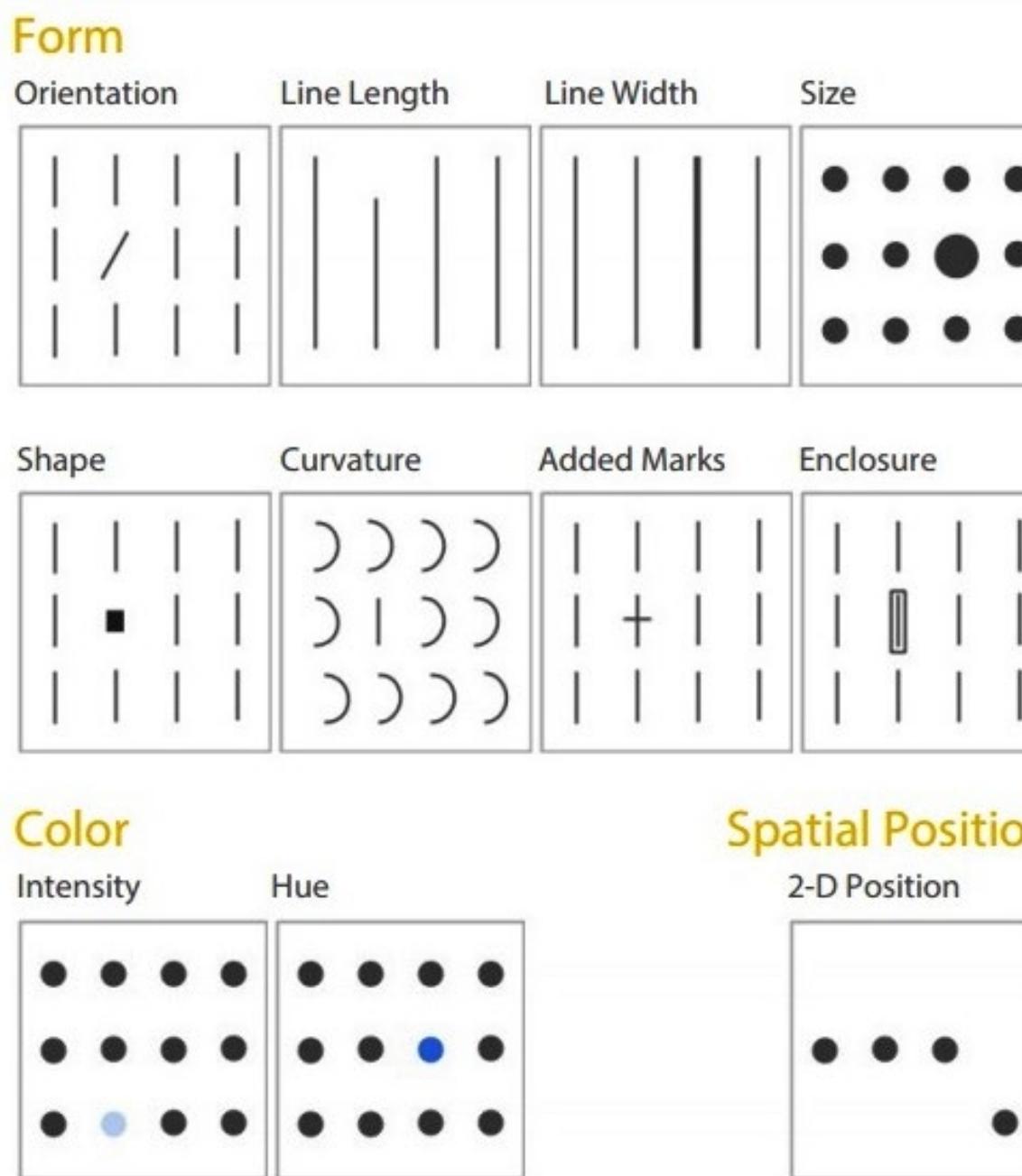


Perceived in less than 10 milliseconds



Unconsciously processed

Pre-Attentive Attributes



Visual Variables

	Quantitative	Ordinal	Nominal
More Accurate	Position Length Angle Slope Area Density Saturation Hue Shape	Position Density Saturation Hue Length Angle Slope Area Shape	Position Hue Density Saturation Shape Length Angle Slope Area
Less Accurate	• • • = = = < < / / / • • • • • • • • • • • • ▲ □	• • • • • • • • • • • • = = = < < / / / • • • ▲ □	• • • • • • • • • • • • • ▲ □ = = = < < / / / • • •

Count the 3s example with pre-attentive attributes

756395068473

658663037576

860372658602

846589107830

Count the 3s example with pre-attentive attributes

756**3**9506847**3**

65866**3**037576

860**3**72658602

8465891078**3**0

No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Color

What are we doing well? Great Products. **These products are clearly the best in their class.**

Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Bold

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Italics

What are we doing well? Great Products. These products are clearly the best in their class.

Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets

without me having to

ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Outline (enclosure)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly.

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Underline (added marks)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

**Aoccdrnig to a rscheearch at
an Elingsh uinervtisy, it deosn't
mttaer in waht oredr the ltteers
in a wrod are, the olny iprmoetnt
tihng is taht frist and lsat ltteer
is at the rghit pclae. The rset can
be a toatl mses and you can stil
raed it wouthit porbelm. Tihs is
bcuseae we do not raed ervey
lteter by itslef but the wrod as a
wlohe.**



Special Effect vs Noise

- Bold
- *Italics*
- Underline
- CAPITAL CASE
- Typeface
- Color
- Inversing Elements
- Size
- Blinking & Flashing

Typefaces for Dyslexia

- Microsoft Office Typefaces.

Arial. Comic Sans. Century Gothic.

Verdana. Trebuchet. Calibri.

- 2. Free fonts designed for dyslexia:

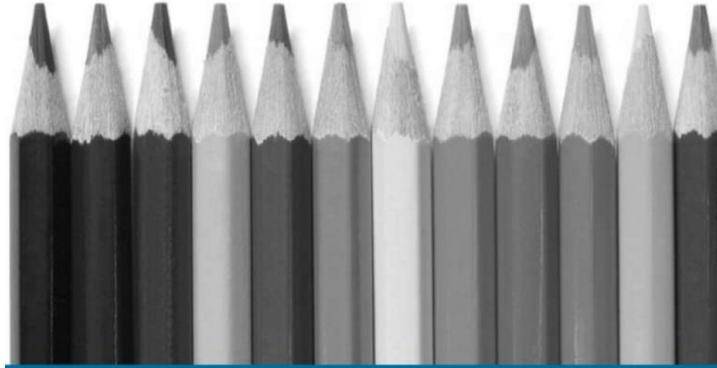
Lexia Readable.

Dyslexie

Open Dyslexic, Open-Dyslexic, OpenDyslexic.

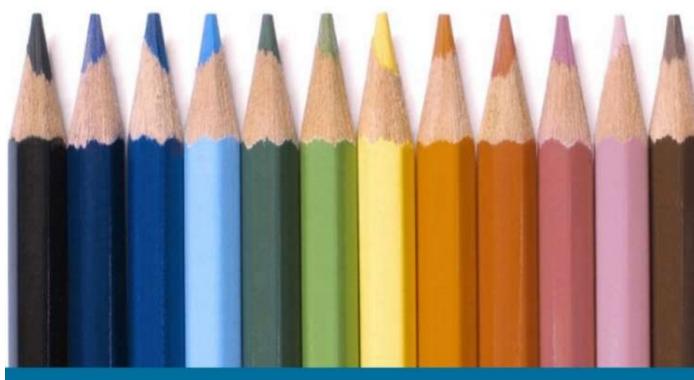
<https://bdanewtechnologies.files.wordpress.com/2011/03/typefaces6.pdf>

<http://opendyslexic.org/>



TOTAL COLOR BLINDNESS

Only able to differentiate
black/dark vs white/ bright



DEUTERANOMALIA

Not able to differentiate
green, dark red, purple



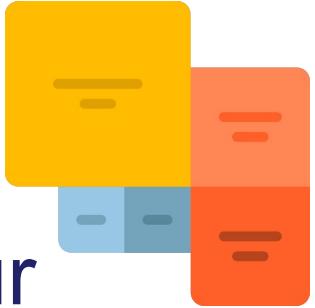
PROTANOPIA

Not able to differentiate between
**red, dark green, blue,
purplish-red** colours

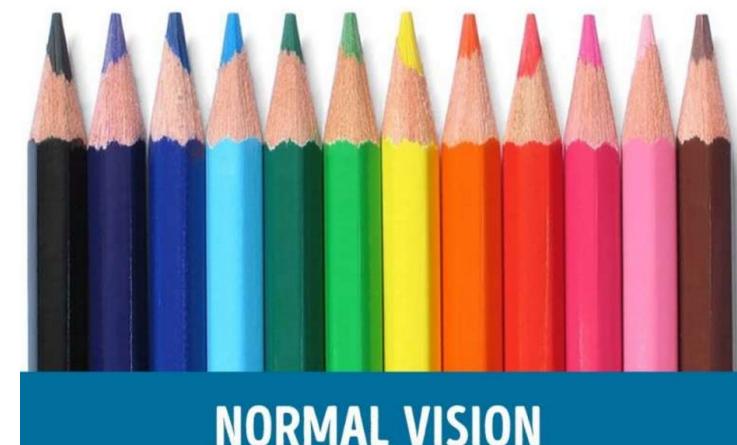


TRITANOPIA

Not able to differentiate **blue**
and **yellow**



Colour Blindness



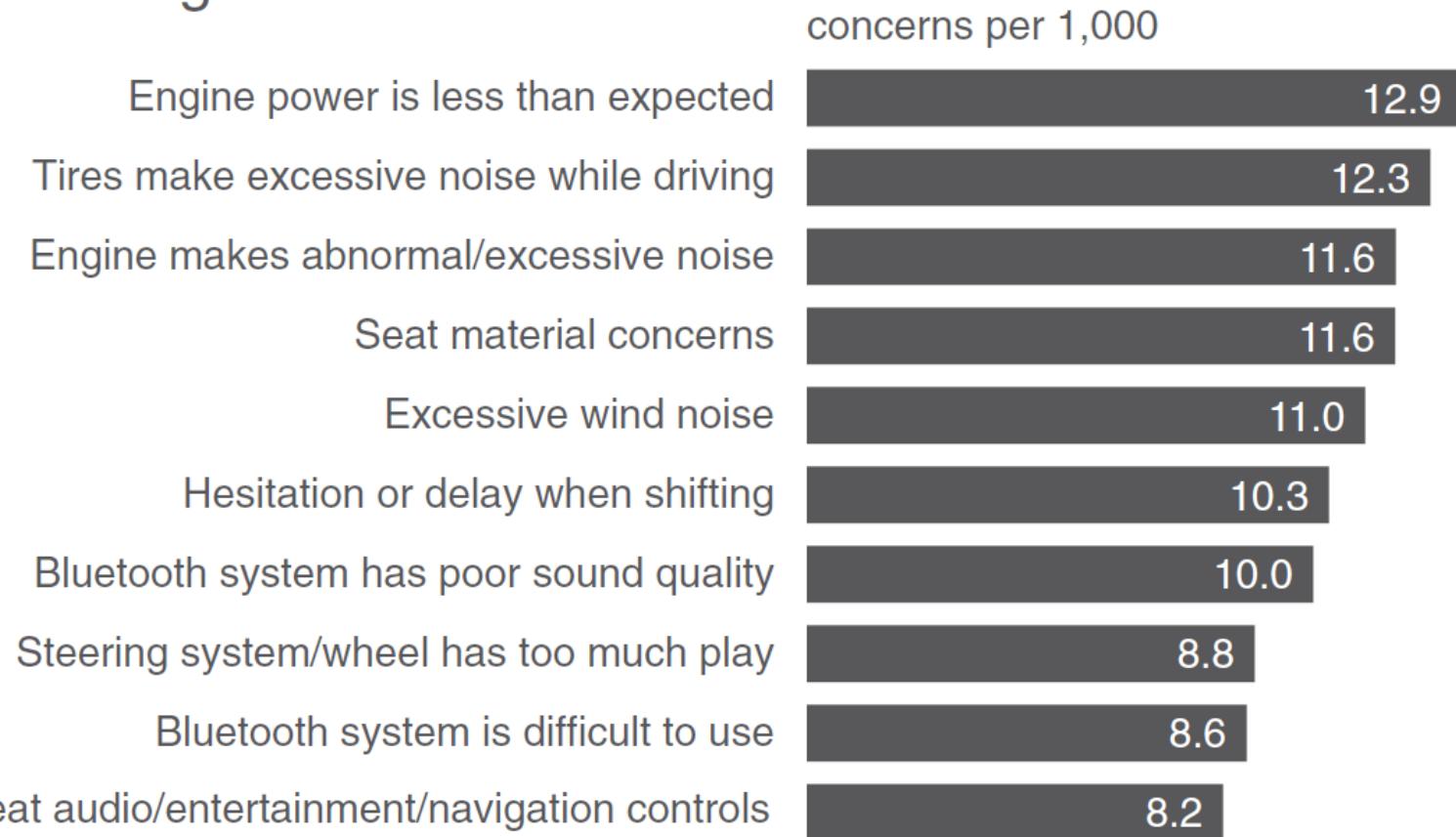
NORMAL VISION

Example of Pre-attentive Attributes in Graphs

Original graph, no pre-attentive attributes



Top 10 design concerns



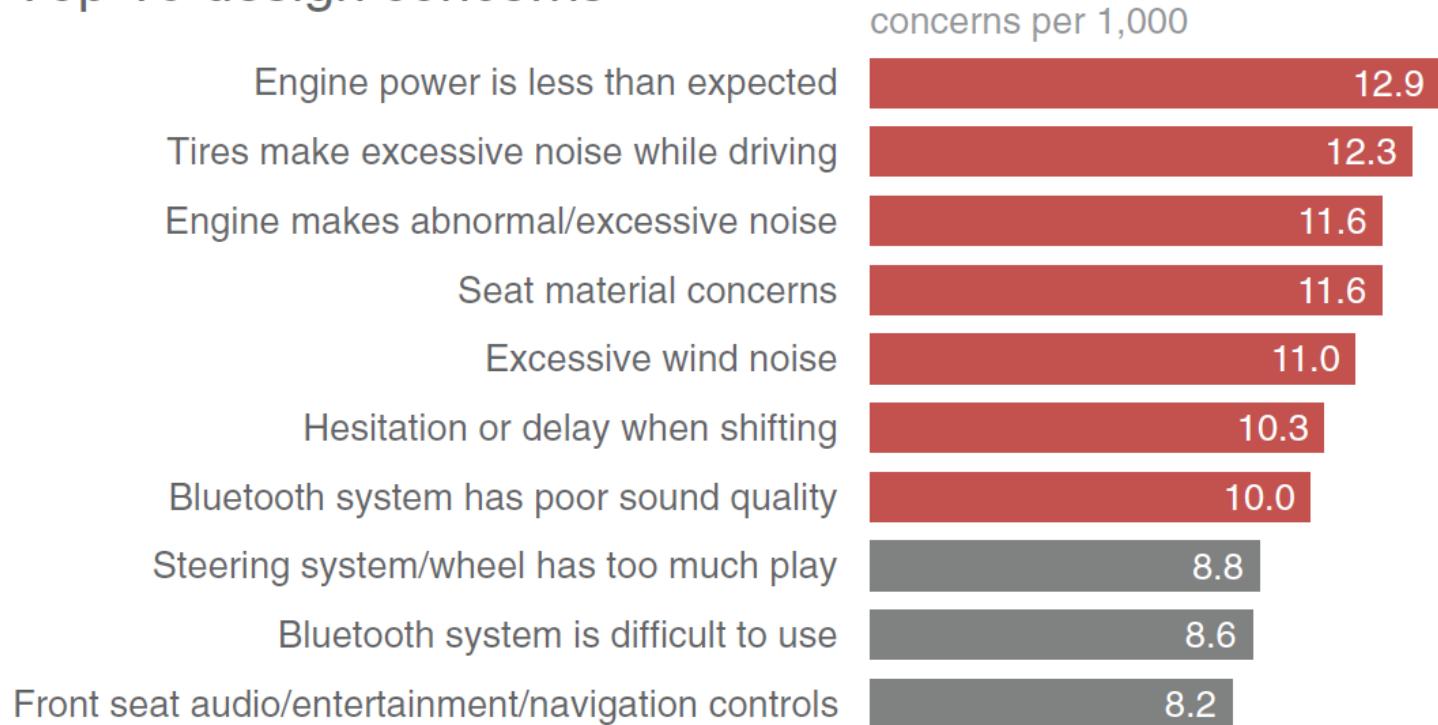
Leverage colour to draw attention



7 of the top 10 design concerns have 10 or more concerns per 1,000.

Discussion: is this an acceptable default rate?

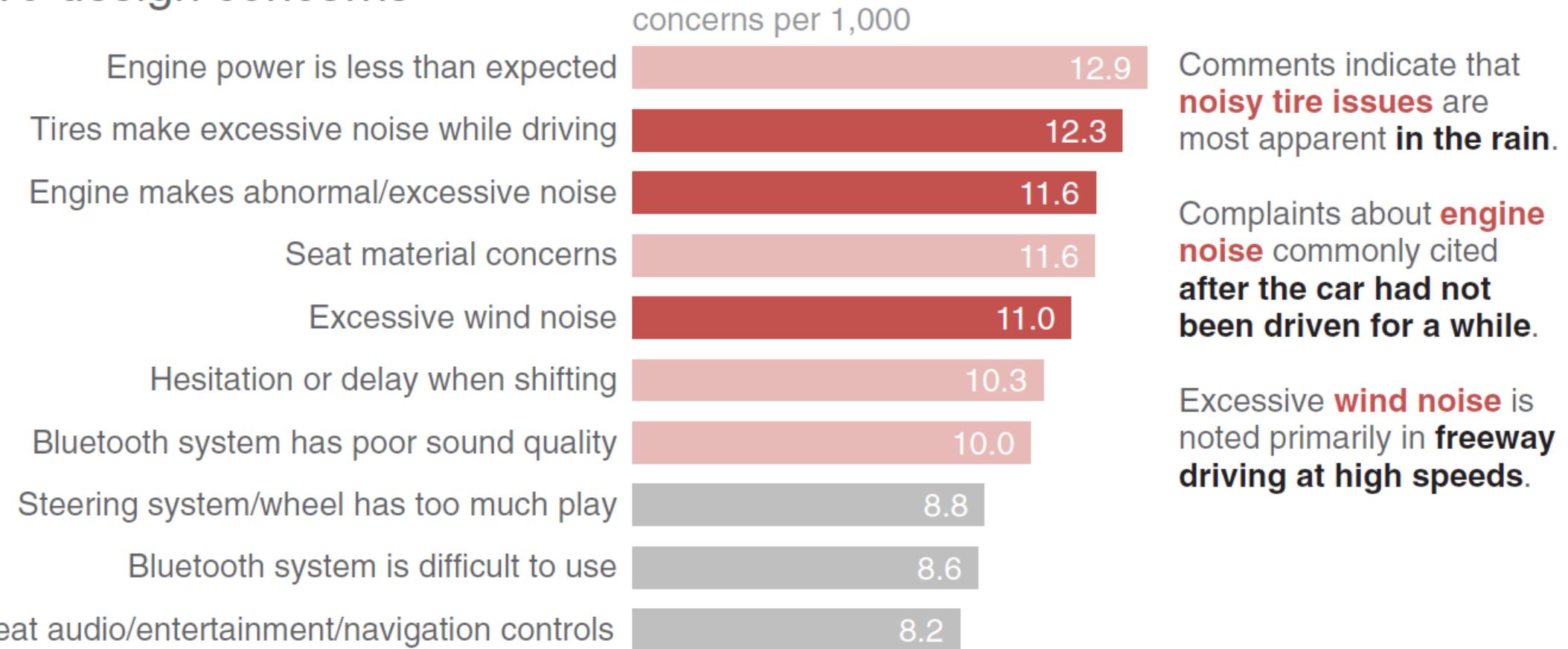
Top 10 design concerns



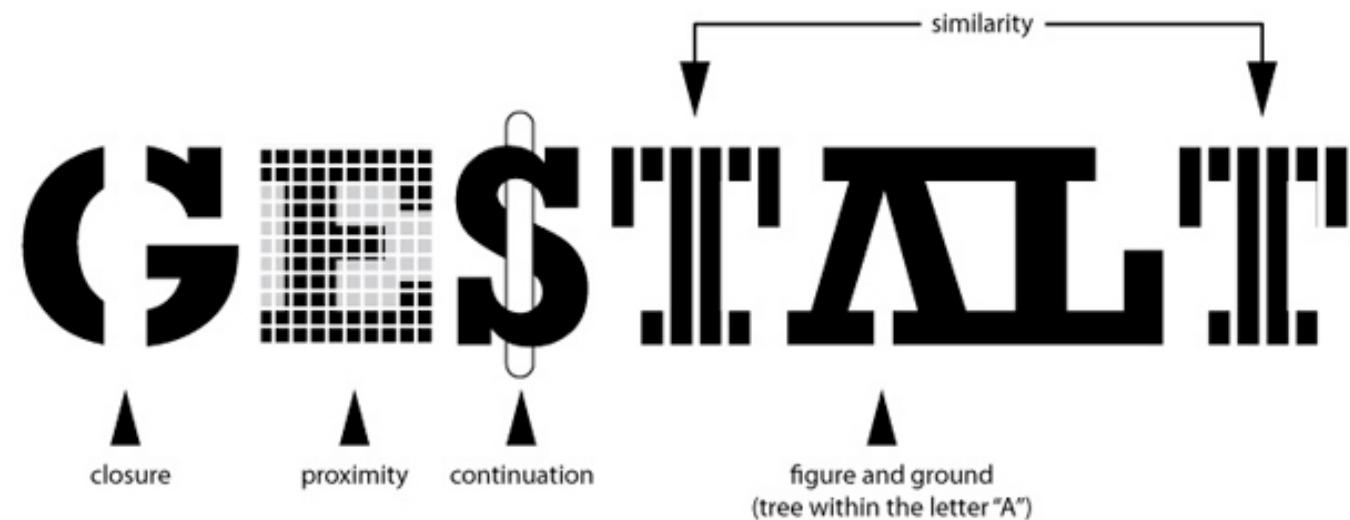
Create a visual hierarchy of information

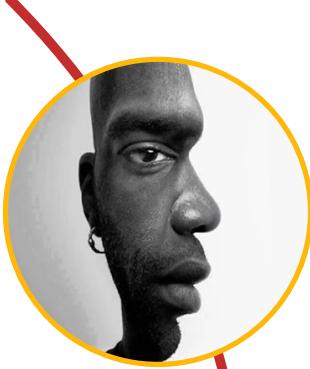


Top 10 design concerns



Gestalt Principles





Gestalt Principles

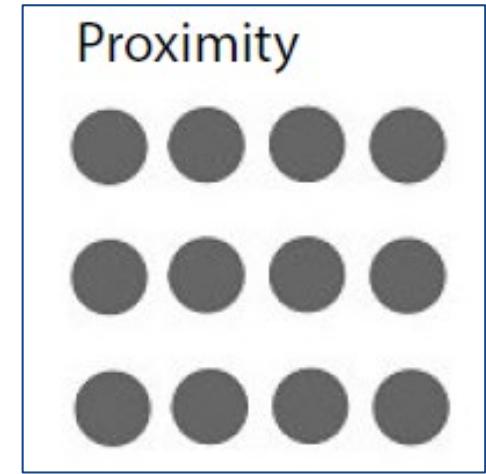
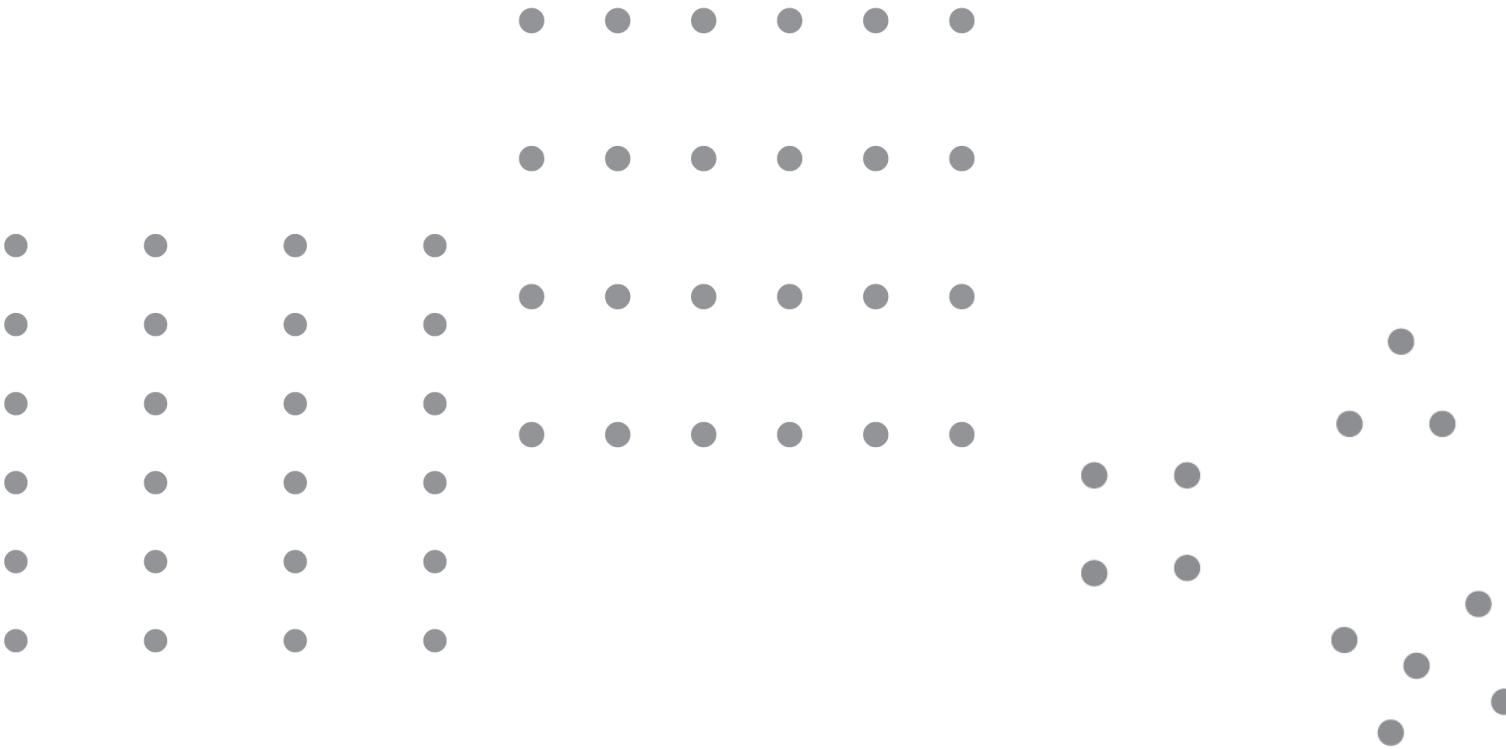


Our brains organize and group visual elements into groups or unified wholes when certain principles are applied...

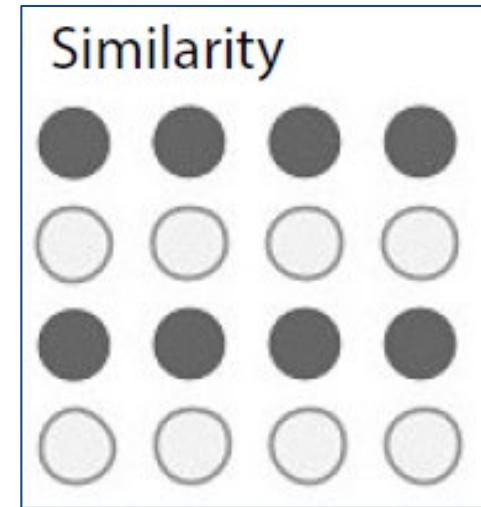
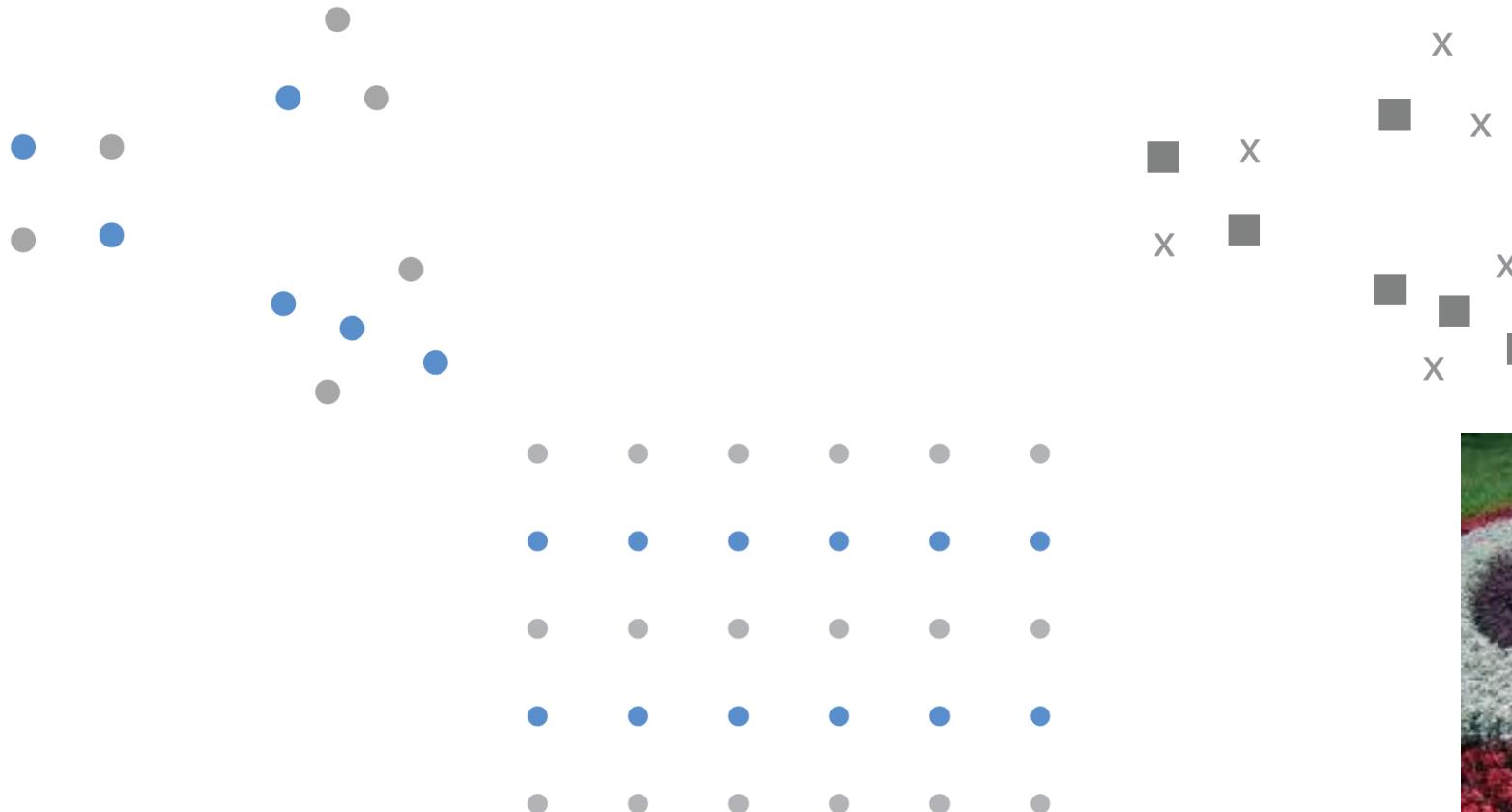


We can use these principles to **highlight patterns** that are important, and **downplay** other **patterns**.

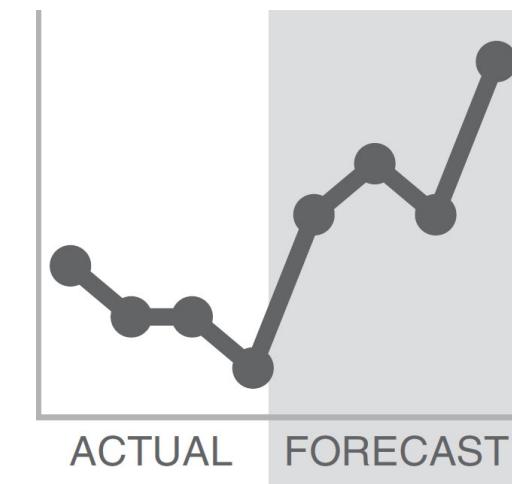
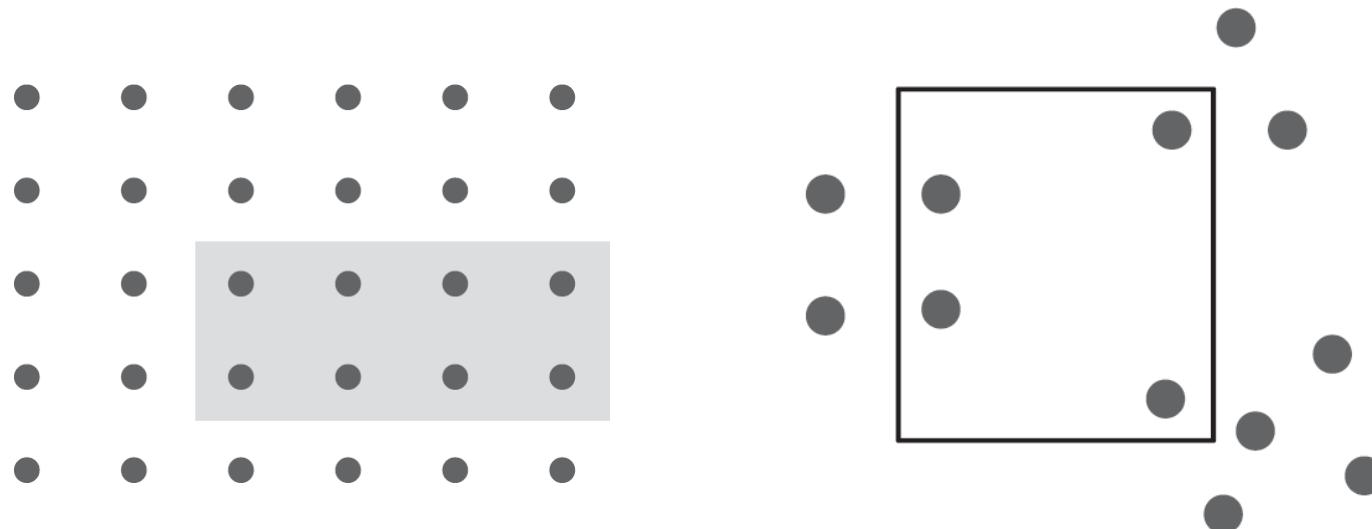
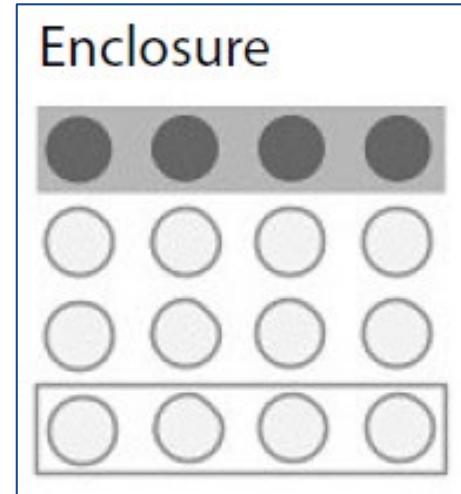
- We see three rows of dots instead of four columns of dots because they are closer horizontally than vertically.



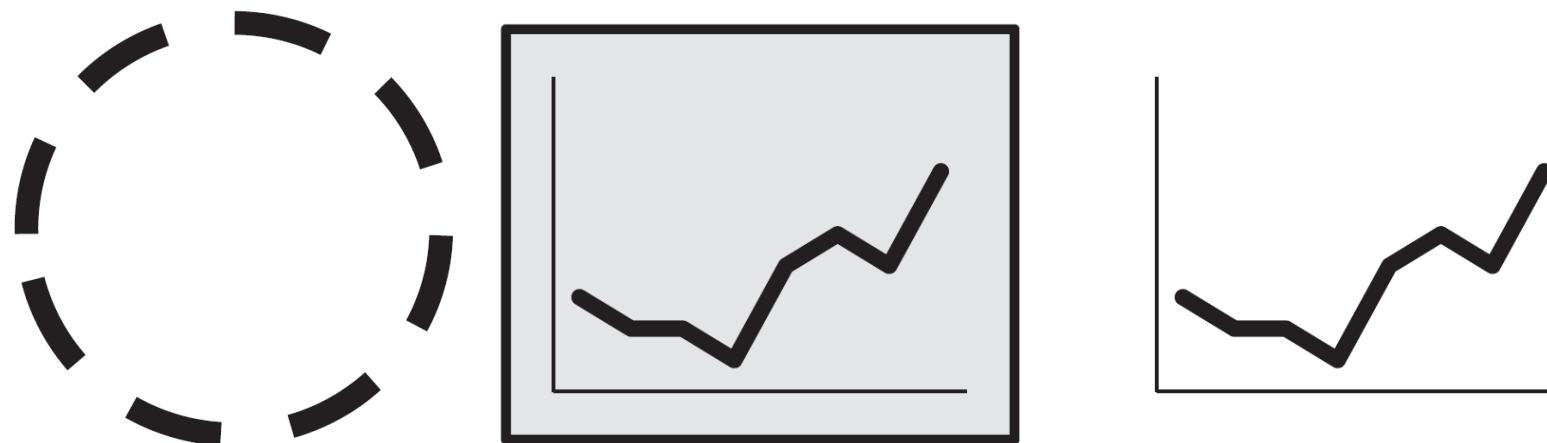
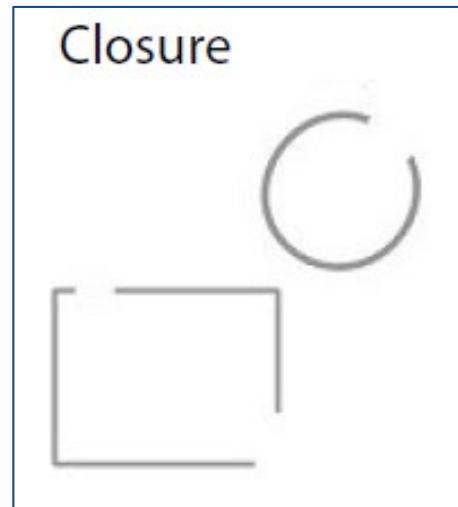
- We see similar looking objects as part of the same group.



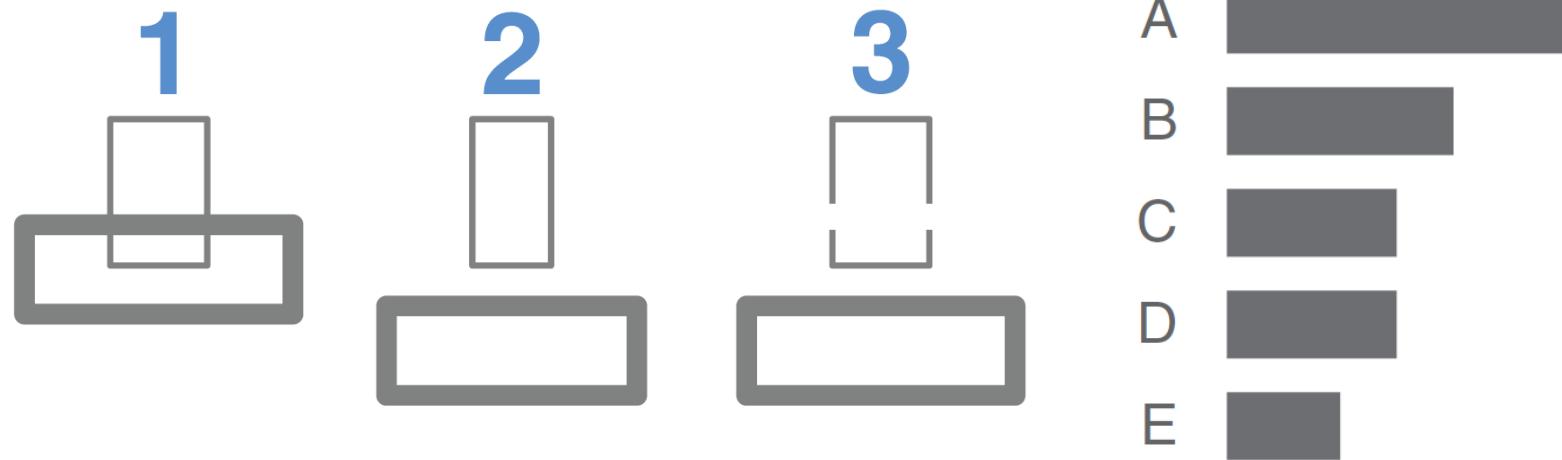
- We group the first four and last four dots as two rows instead of eight dots.



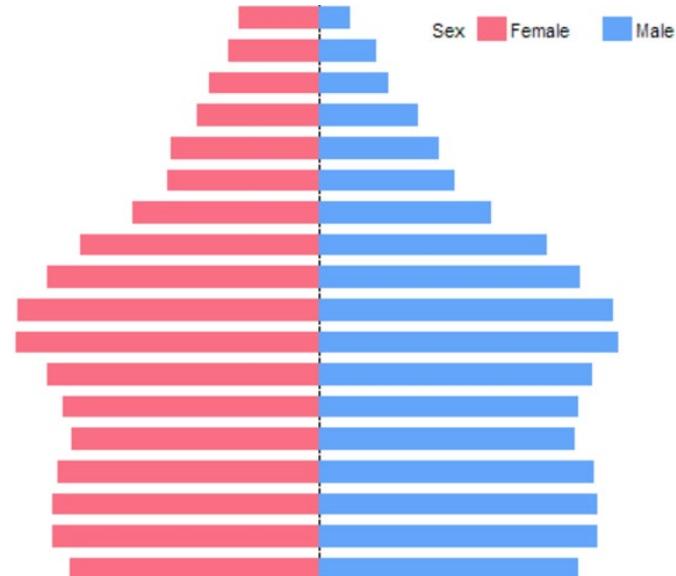
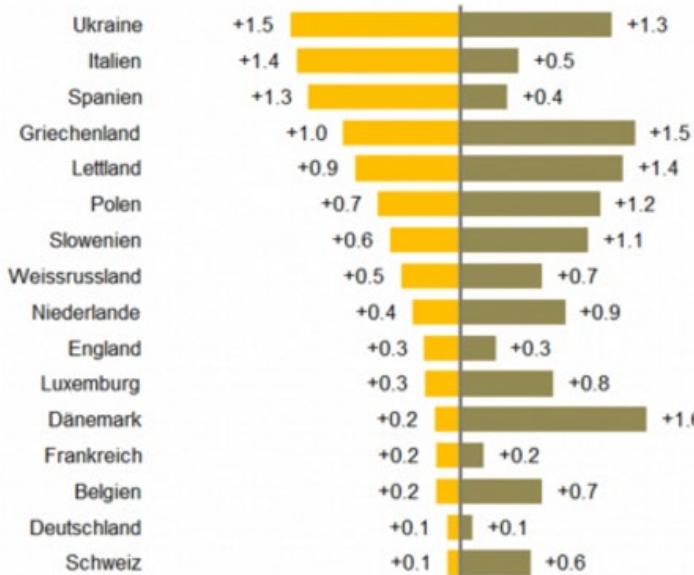
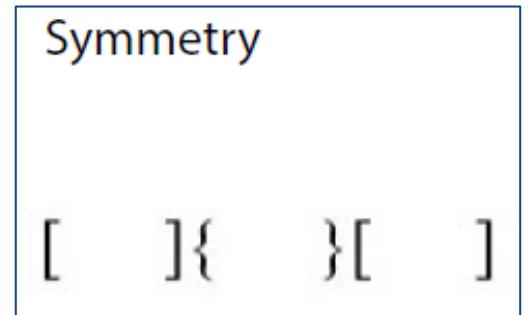
- Closure is the concept that our minds fill in gaps to complete objects and shapes.
- We automatically close the square and circle instead of seeing three disconnected paths.
- We complete the shape made by the hand and combine it with the bottom of a light bulb to make the shape of a full light bulb even though nothing is there.



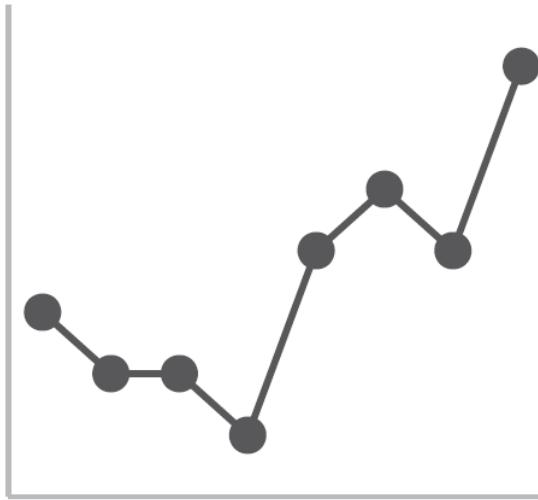
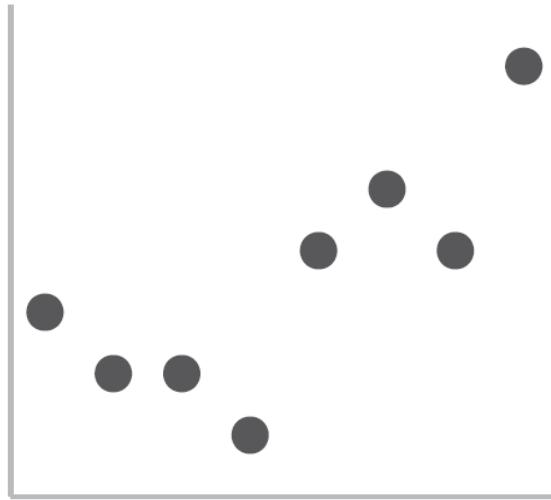
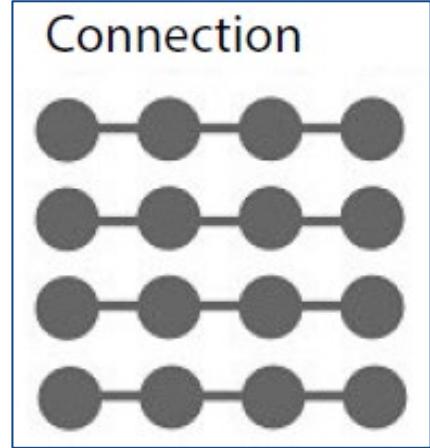
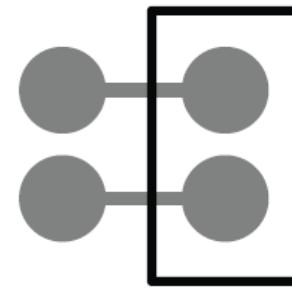
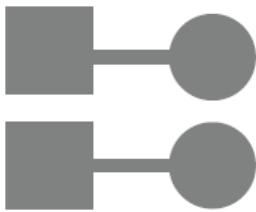
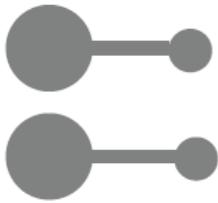
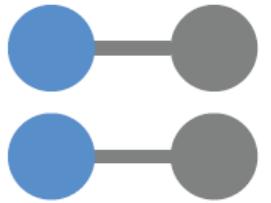
- We see one continuous path instead of three arbitrary ones.
- Continuity is the concept that rather than seeing many discontinuous patterns, we tend to see one continuous pattern. Here rather than seeing the tips of the shoes as individual parts of an ellipse, we see a large circle made with all of the tips of the shoes combined.



- We see three pairs of symmetrical brackets rather than six individual brackets.

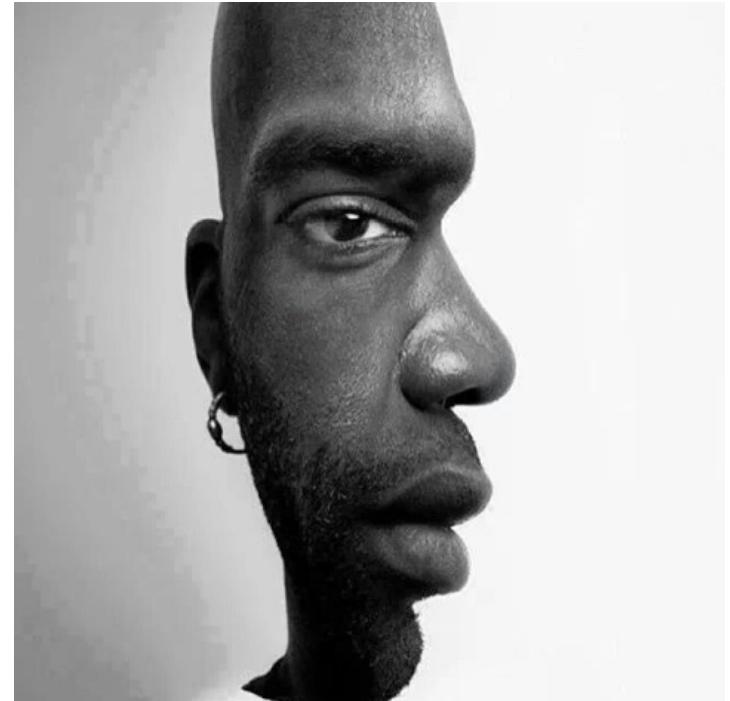


- We group the connected dots as belonging to the same group.



- We either notice the two faces, or the vase. Whichever we notice becomes the figure, and the other the ground
- Reversible Figures are optical illusions that uses graphical similarity between objects and shapes to cause us to be able to see two or more images. This image is an example that has half of a man's face but it can look as if you are looking at the side of the man's face or half of the front of his face.

Figure & ground



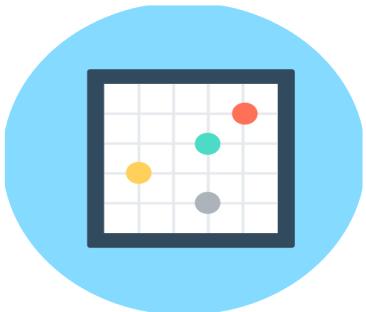
Graphical Excellence



the
greatest
number
of ideas



in the
shortest
time



with
the
least
ink



in the
smallest
space



Edward Tufte, The Visual Display of Quantitative Information

Data Visualisation

that have achieved 'Graphical Excellence'...

-  Show the data and help the audience think about the important messages
-  Avoid distorting the data or its message
-  Present many numbers in a small space
-  Emphasize the important numbers
-  Make large data sets coherent
-  Encourage the audience to compare different pieces of data
-  Reveal the data at several levels of detail (overview to fine detail)

Graphical Excellence **Data:Ink & Chartjunk**



Graphical Excellence is high when the **Data:Ink** is high and **Chartjunk** is low



Data:Ink = ink used to represent data / total ink used to print the graphic
To **maximise** the Data:Ink

- we must erase ‘non-data-ink’ and ‘redundant data-ink’

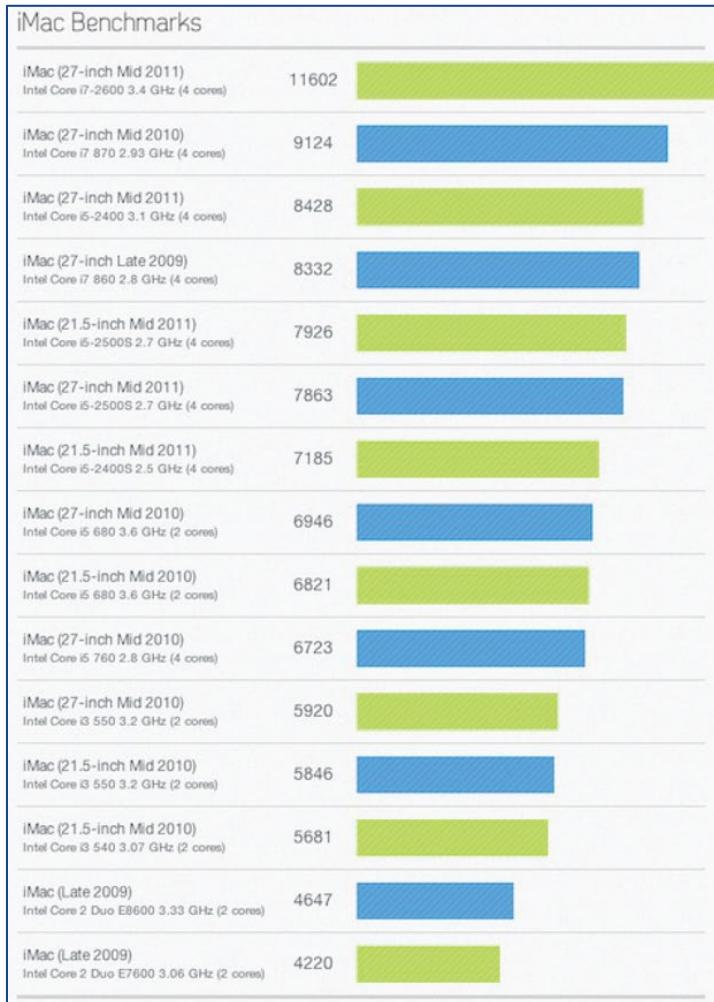


Chartjunk = superfluous, decorative, or diverting ink

To **minimize** Chartjunk

- we must remove ‘**moiré vibration**’, **grids**, and ‘**the duck**’

Data:Ink & Chartjunk Moiré Vibration



Source: <http://www.primatelabs.ca/blog/2011/05/imac-benchmarks-mid-2011/>



Data:Ink & Chartjunk

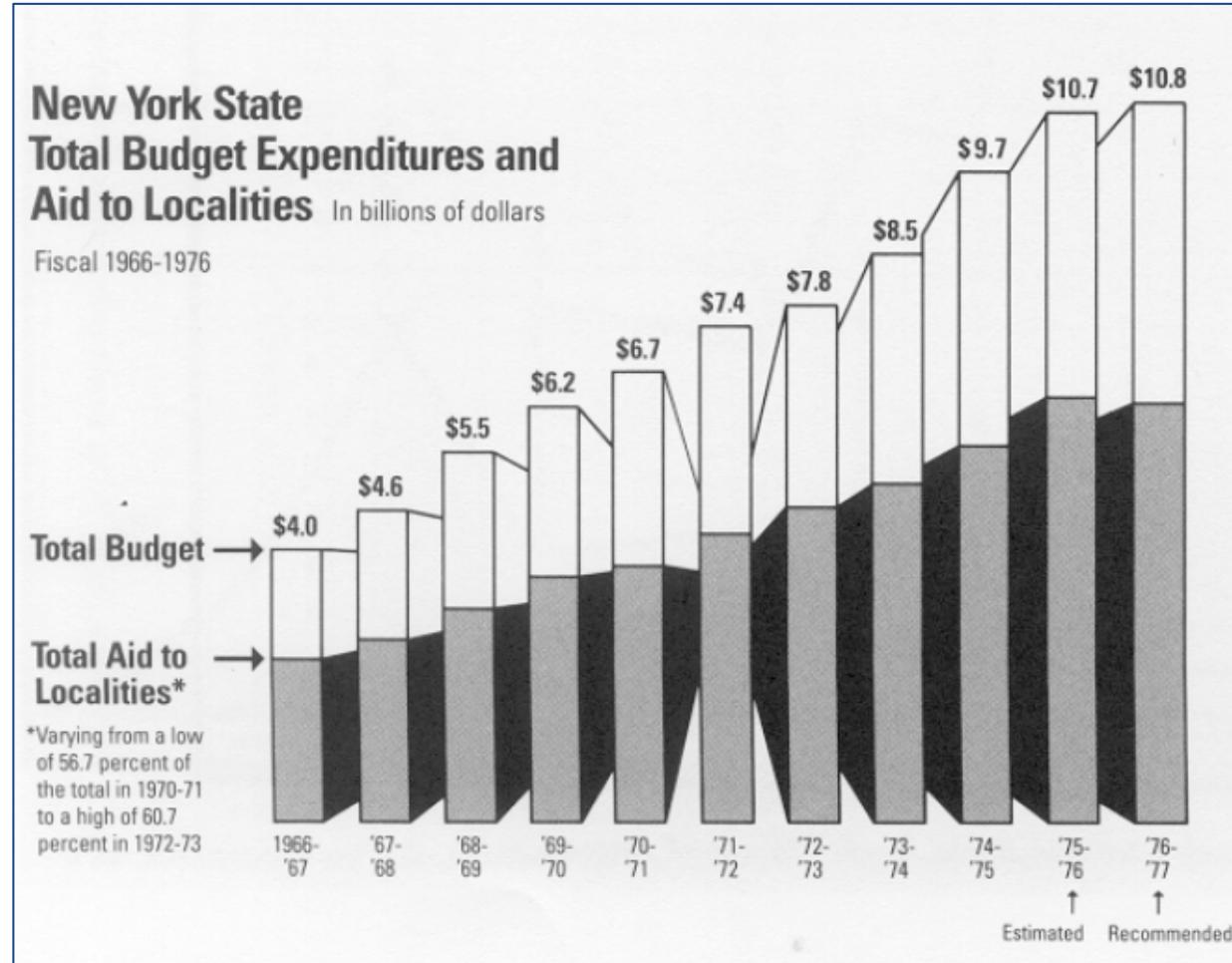
Grid

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Data:Ink & Chartjunk the Duck



Graphical Integrity

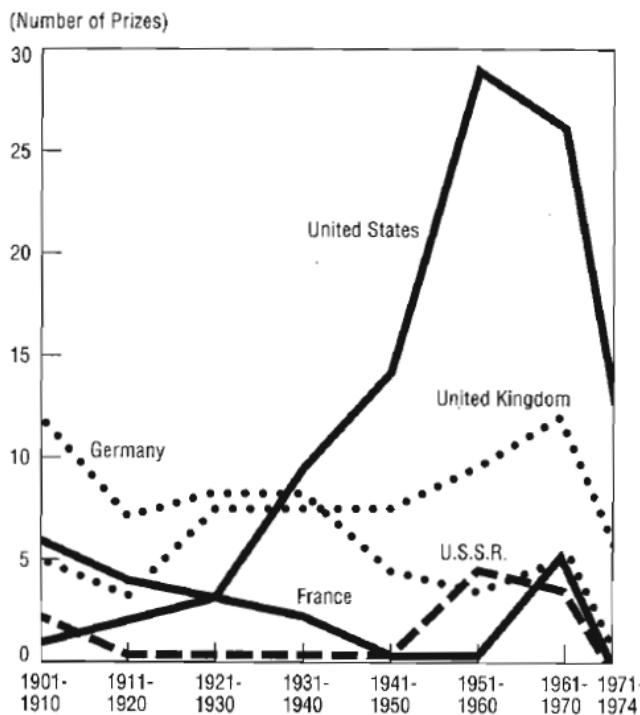


Graphical Excellence is high when the **Graphical Integrity** is high...

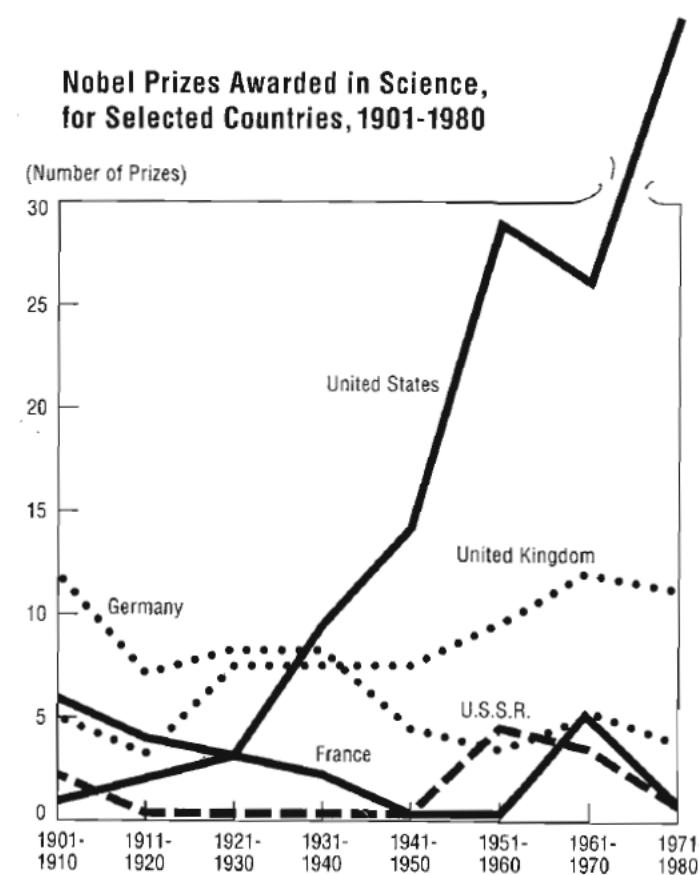
-  Graphical Integrity requires an absence of misrepresentation, persuasion, masking, or 'lies'...
-  Consistency in baselines and labels
-  Consistency in compared scales
-  Partial data to be highlighted
-  Compared data to be normalised
-  Context to be represented

Graphical Integrity

Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1974

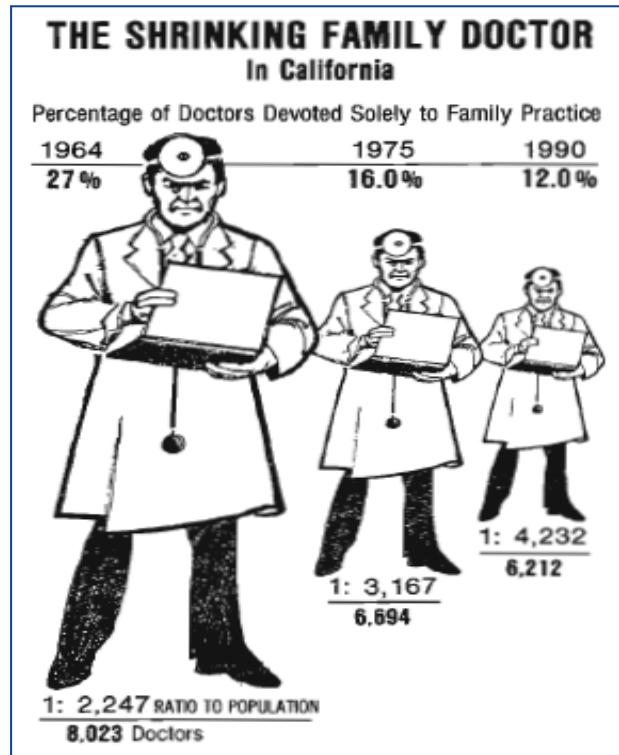


Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1980



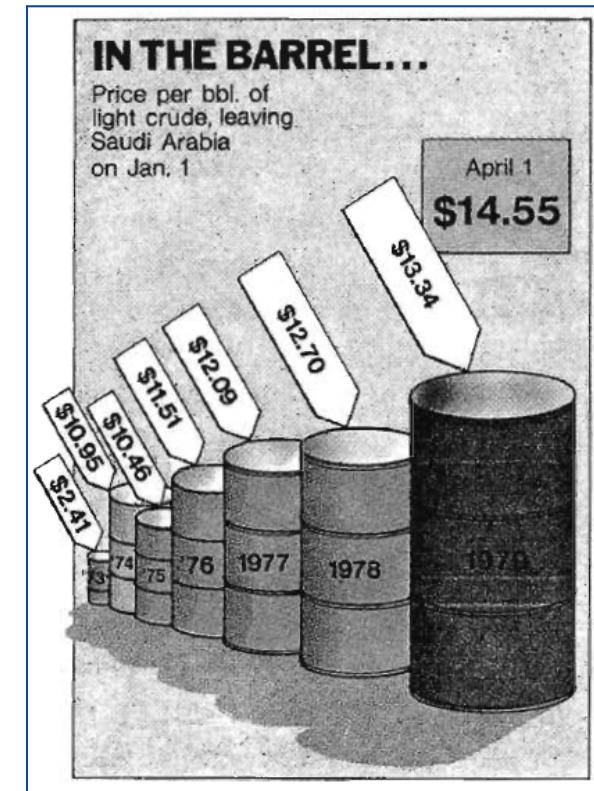
$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Lie Factor = c.3 (area)



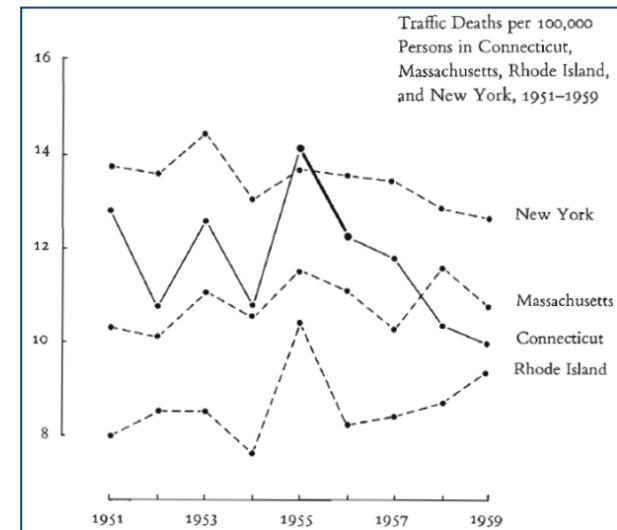
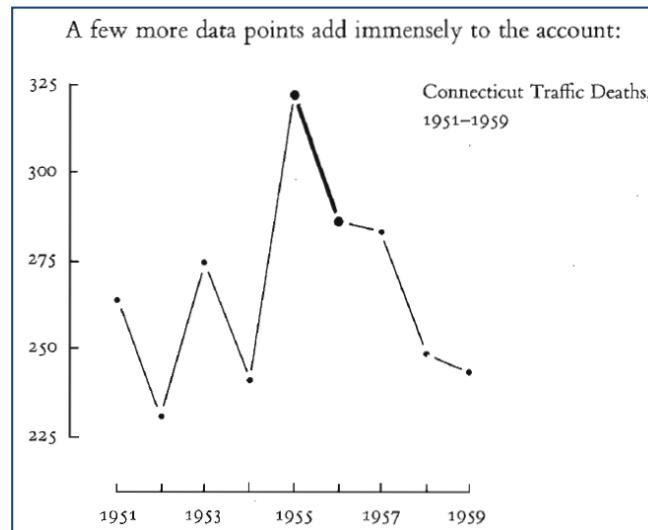
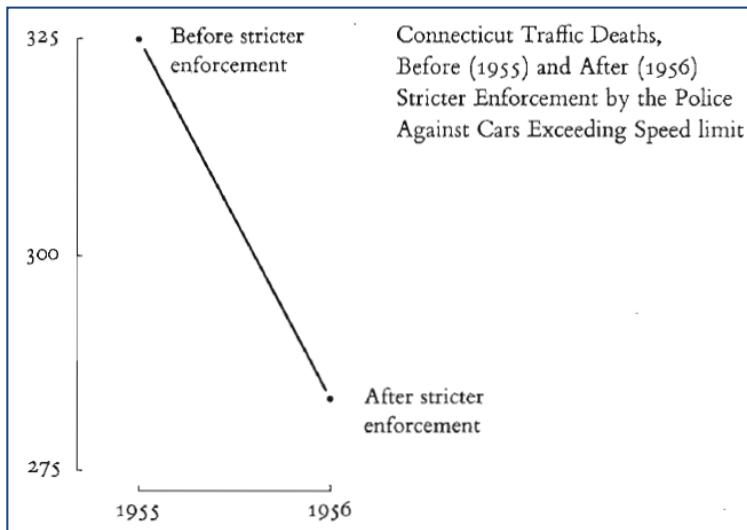
Los Angeles Times, August 5, 1979, p. 3.

Lie Factor: 59.4
(volume)



Time, April 9, 1979, p. 57.

Graphical Integrity



The message evolves as the context
increases...

Graphical Integrity

Six principles to ensure ‘Graphical Integrity’...



Make the representation of numbers proportional to quantities



Use clear, detailed, and thorough labeling



Show data variation, not design variation



Use standardized units, not nominal values

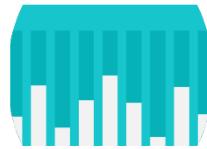


Depict ‘n’ data dimensions with less than or equal to ‘n’ variable dimensions



Quote data in full context

Graphical Integrity



Graphical Excellence is high when the Data Density is high...



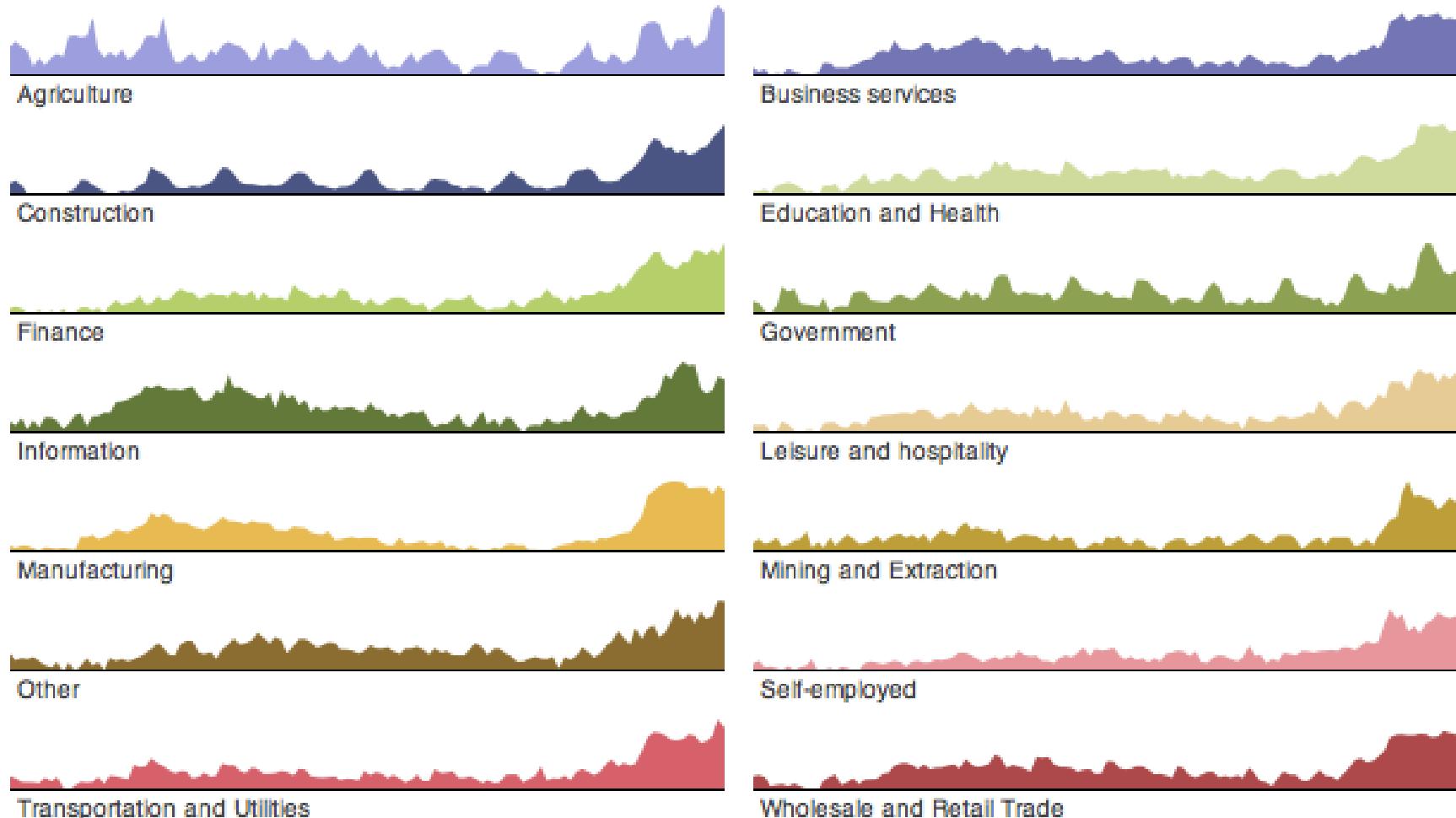
Data Density = number of data entries in graphic / area of graphic



To maximise Data Density we should...

- Consider the use of 'small multiples'
- Maximize space by removing unnecessary 'ink'

Small Multiples

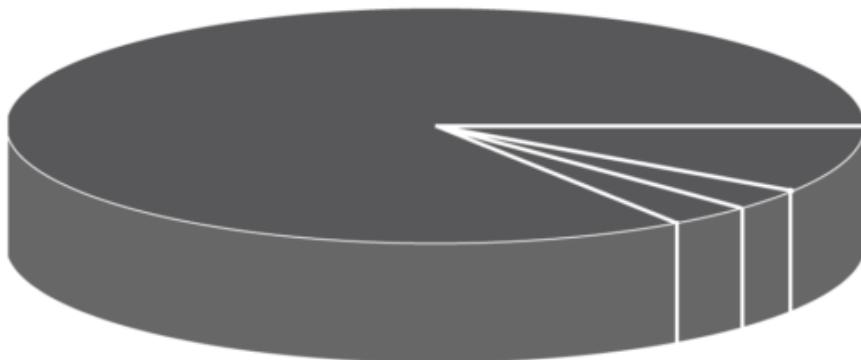


How to Spot Visualization Lies

Keep your eyes open

EXTRA DIMENSION JUST BECAUSE

Just say no.



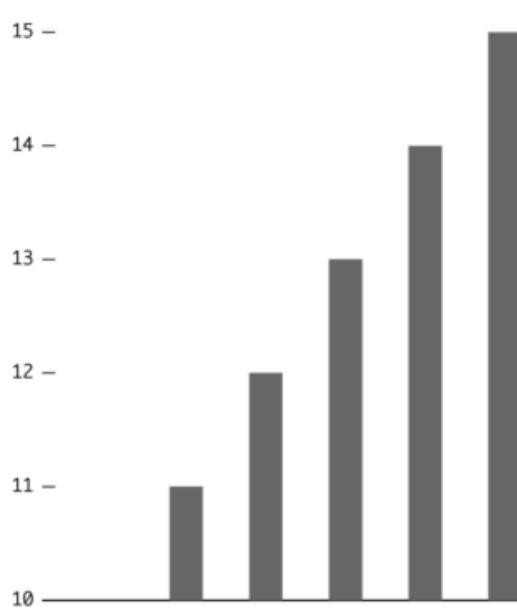
PUTZING AROUND WITH AREA DIMENSIONS

These fill the same amount of area, but they look very different.

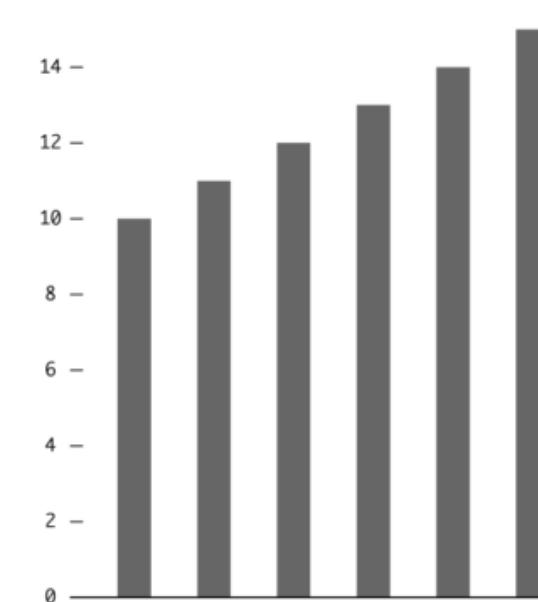


TRUNCATED AXIS

The value axis starts at ten. Liar, liar, pants on fire.

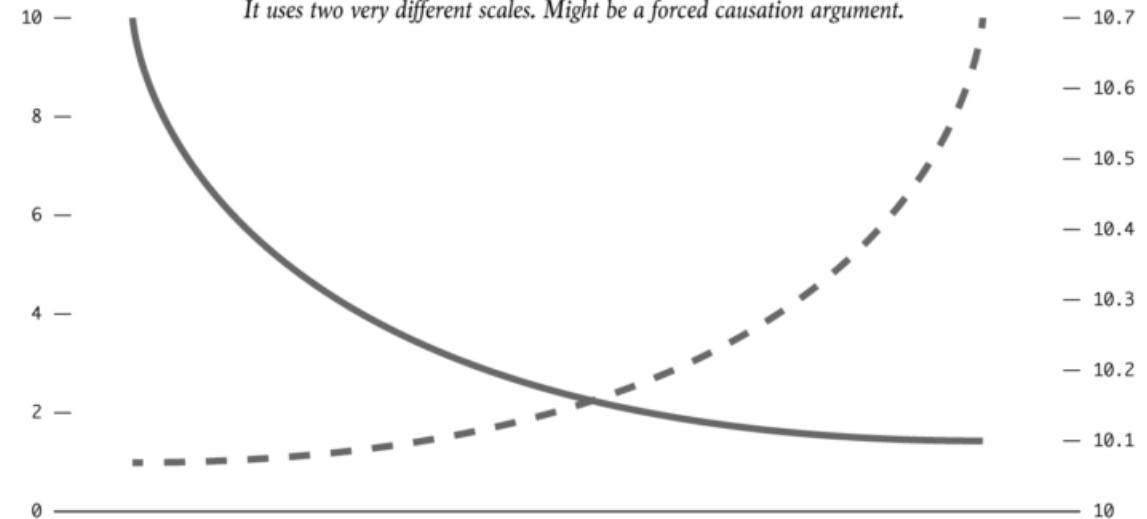


The value axis starts at zero. Good.



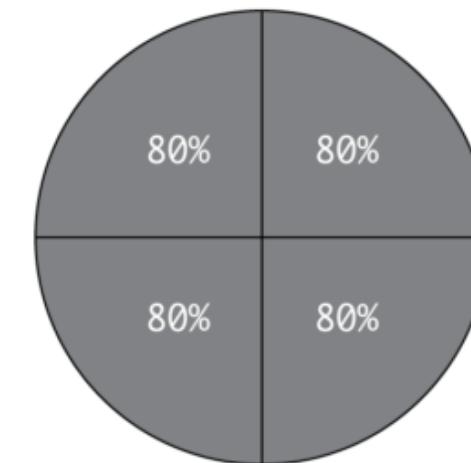
DUAL AXES

It uses two very different scales. Might be a forced causation argument.



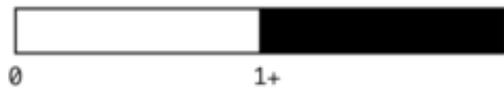
IT DOES NOT ADD UP

*The parts add up to more than the whole, which is 100%.
For my next trick, I will turn this rabbit into a big bag of money.*



ODD CHOICE OF BINNING

Two bins. What's really in the 1+ category?
Might be hiding something.

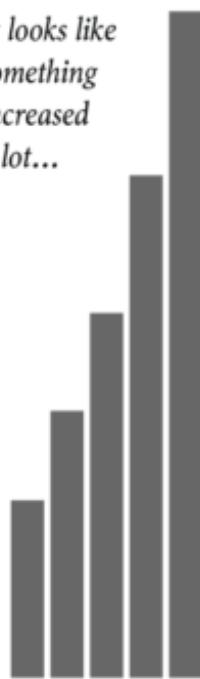


That's better. It can show more variation.

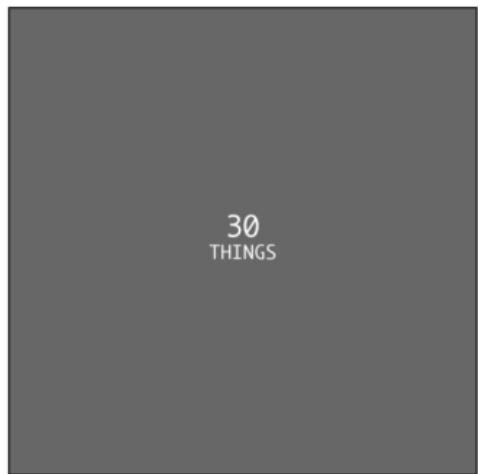
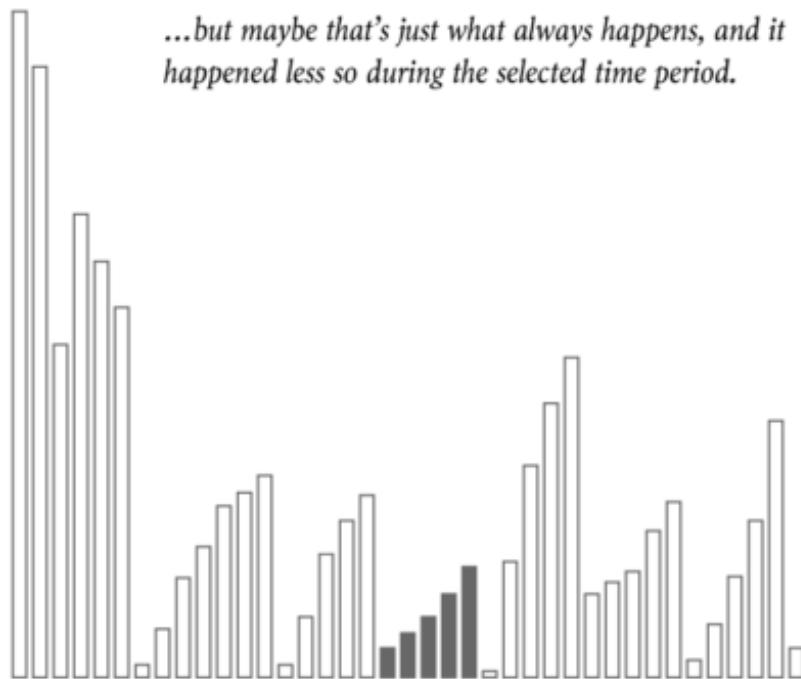


LIMITED SCOPE

It looks like something increased a lot...



...but maybe that's just what always happens, and it happened less so during the selected time period.



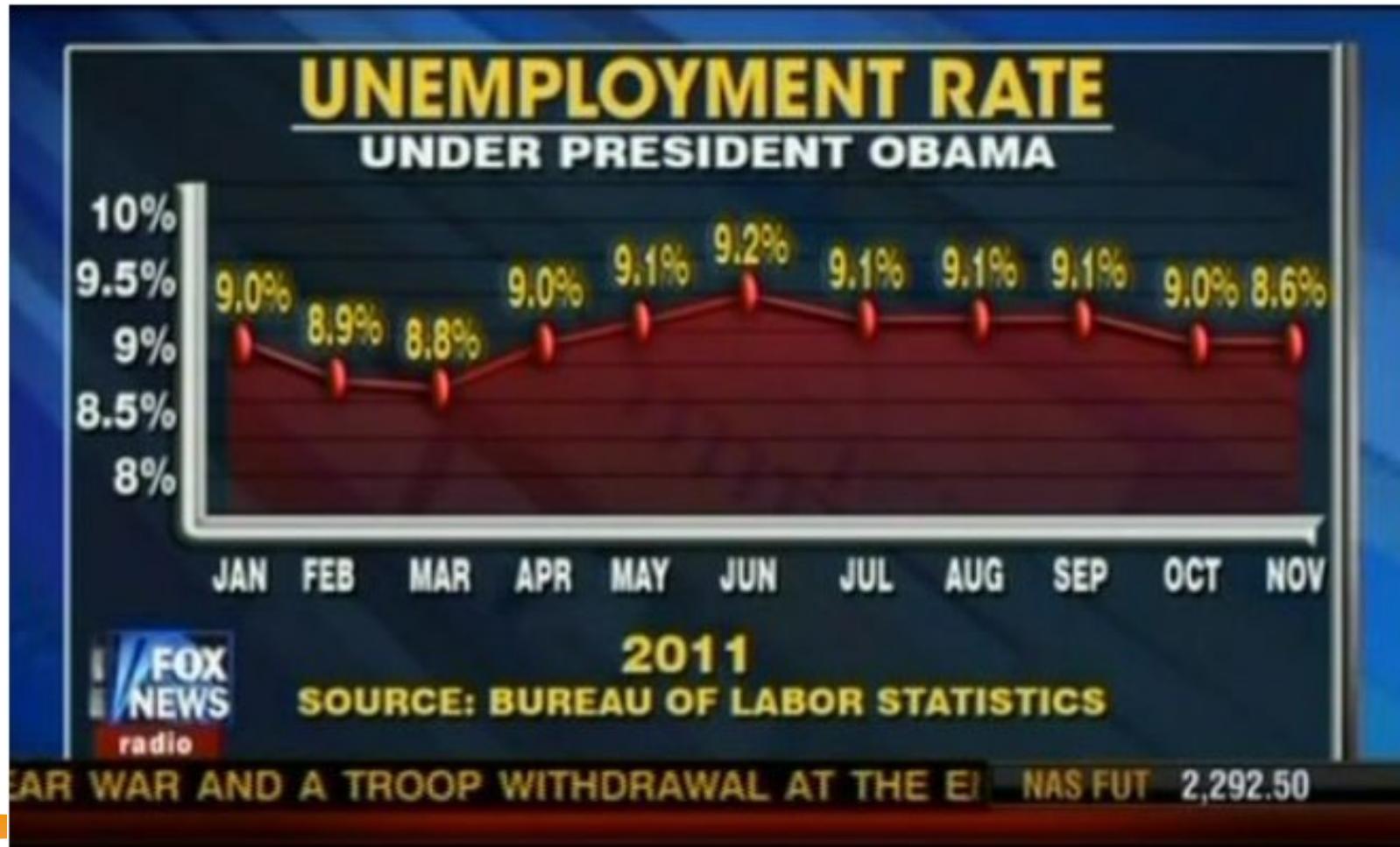
AREA SIZED BY SINGLE DIMENSION

Thirty is three times ten, but that third rectangle looks a lot bigger than the first.
Might be trying to inflate significance.

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

Misleading Graphics

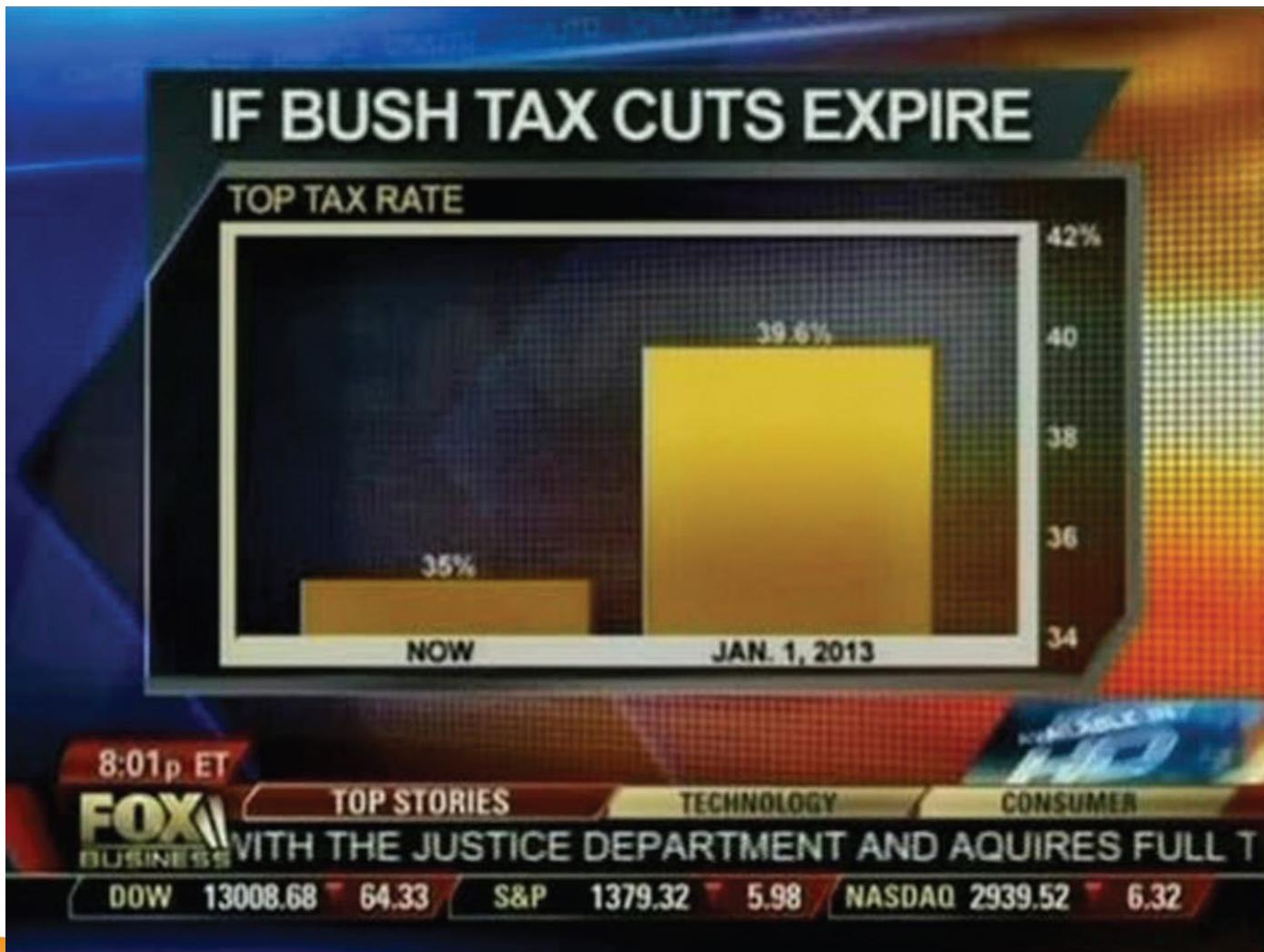
Poor Graphical Integrity & Graphical Excellence



High Chartjunk

Low Data:Ink
ratio

Poor Graphical Integrity & Graphical Excellence



High Chartjunk

Low Data:Ink ratio

Summary

- Pre-Attentive Attributes
- Gestalt Principles
- Visualisation Excellence
 - Graphical Excellence
 - Graphical Integrity
- Three key aspects of Graphical Excellence
- Six principles to ensure Graphical Integrity
- Graphical Excellence is high when the Graphical Integrity is high...

