

# What is information visualization?

*“To show information in a spatial or graphical representation, in order to facilitate comparison, pattern recognition, change detection, and other cognitive aptitudes of our visual system”*

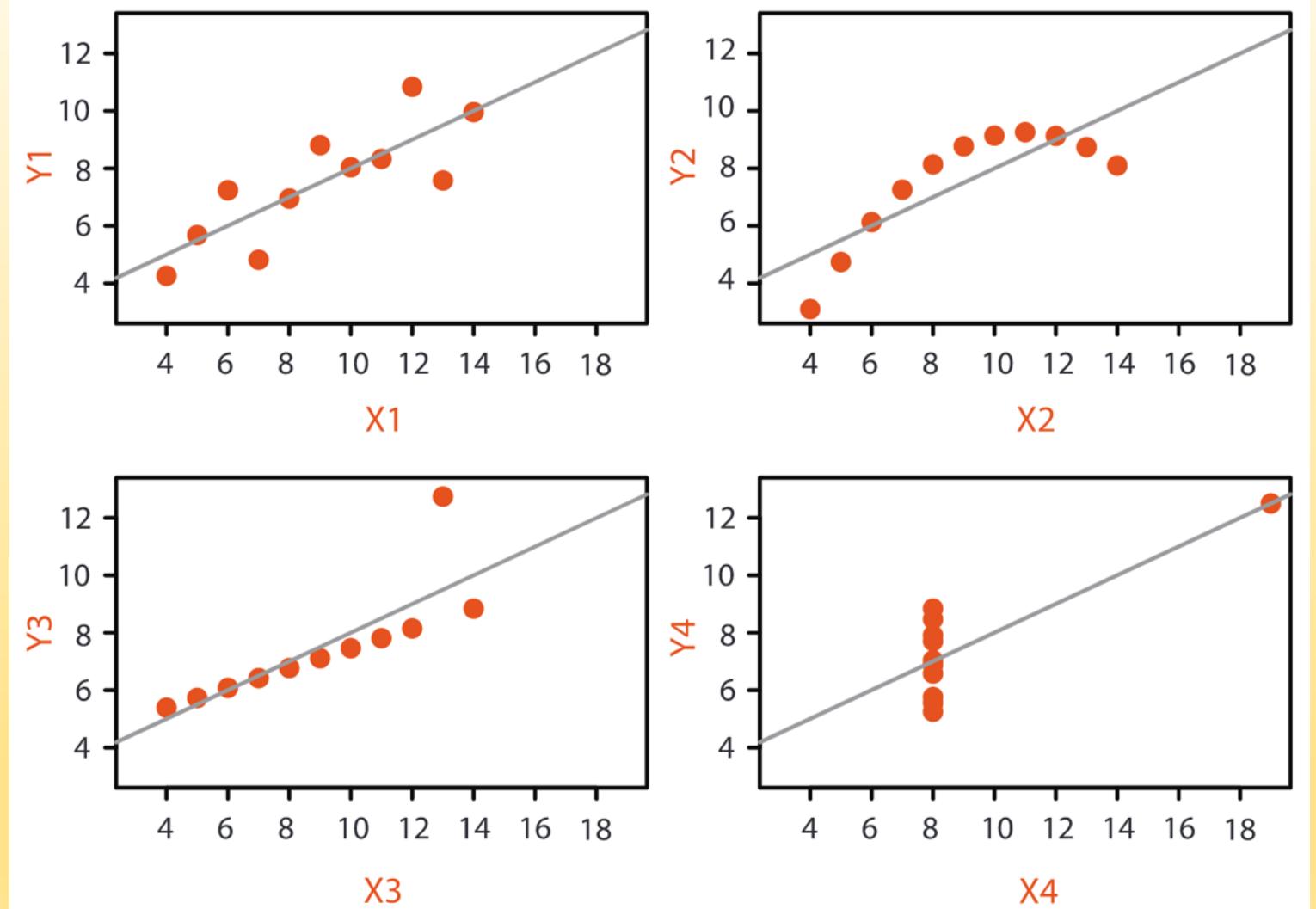
Marti Hearst

# Why

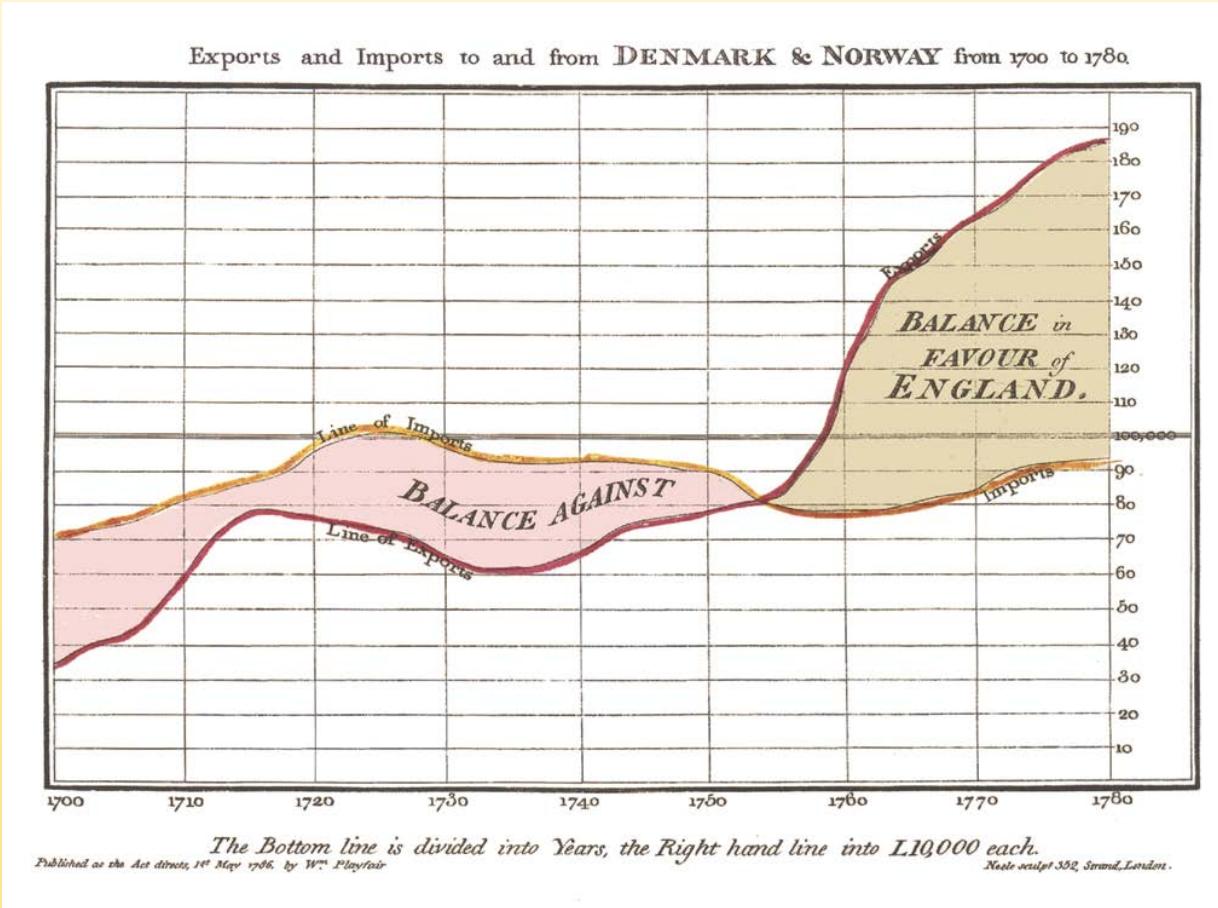
Anscombe's Quartet: Raw Data

	1		2		3		4	
	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
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

# Why

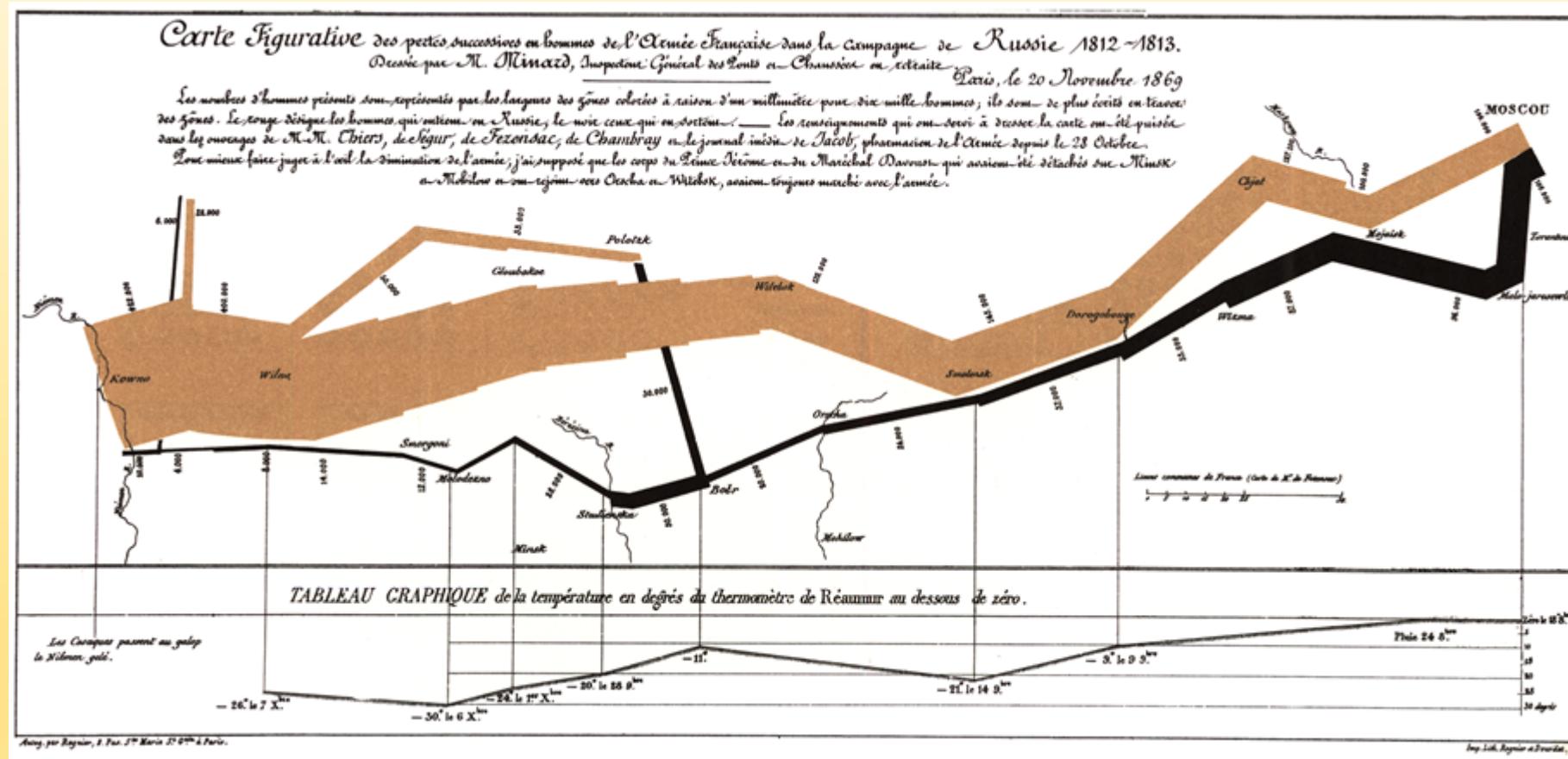


# Example 1: William Playfair



Line, bar and pie charts were all invented by William Playfair.

# Example 2: Joseph Minard

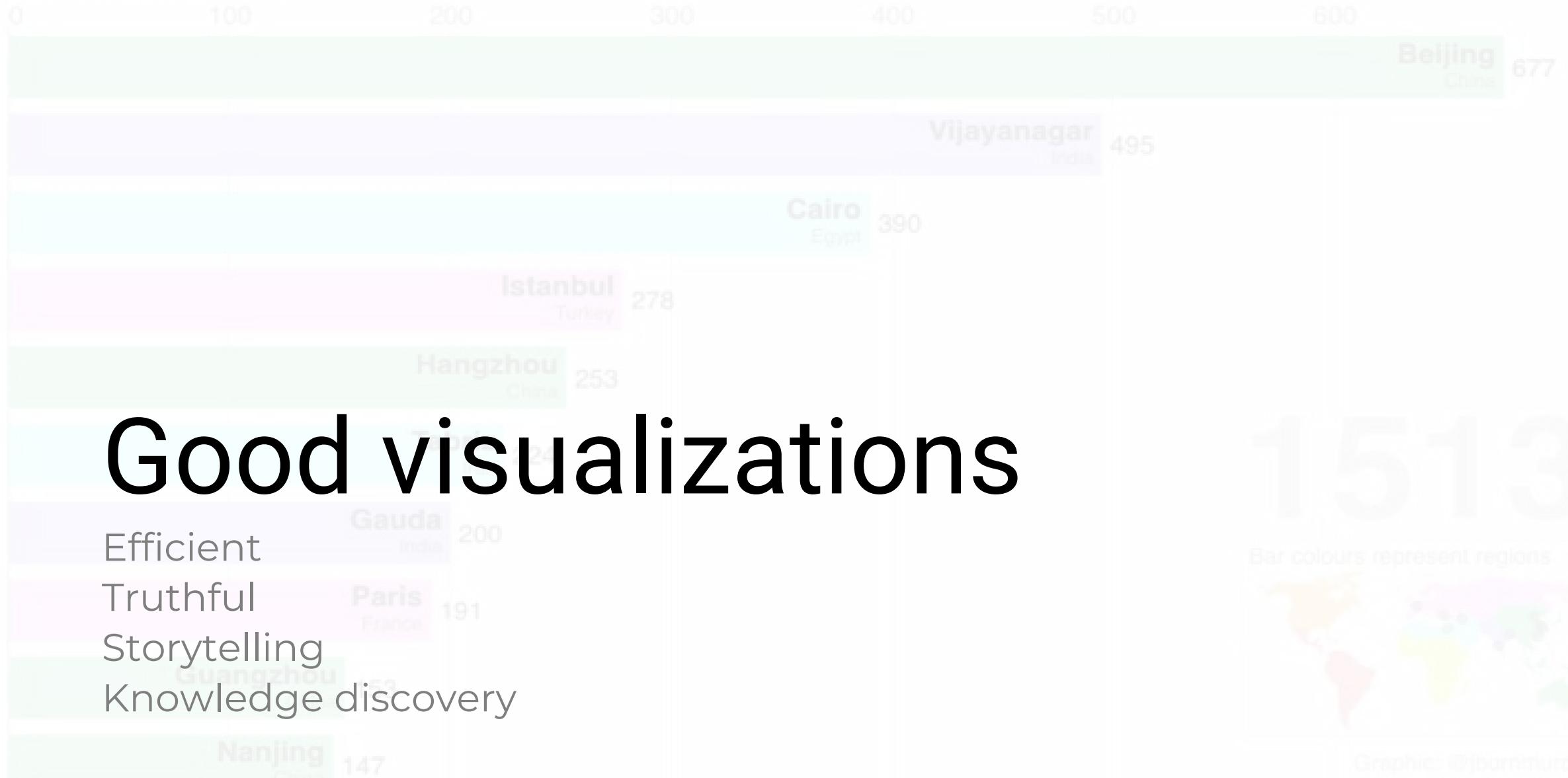


# We use information visualization to...

- Communicate
  - Clearly
  - Efficiently
- Facilitate discovery
  - To identify relations
  - To discover patterns
- Help digest the data

# The most populous cities in the world from 1500 to 2018

Population (thousands)



# Good visualizations

Efficient

Truthful

Storytelling

Knowledge discovery



Bar colours represent regions

Graphic: ©jbmurdoch

Sources: Rebs, M., L. F. Reitsma, and K. C. Seto. 2018. Demographia

# Efficient

To **fit** your users **needs**

To require **low effort**

# Text: plain language

	Consejo 1: <b>Piense antes de escribir</b>	.....Página 3
	Consejo 2: <b>Piense en el lector</b>	.....Página 4
	Consejo 3: <b>Dé forma a su documento</b>	.....Página 5
	Consejo 4: <b>Sea breve y conciso</b>	.....Página 6
	Consejo 5: <b>Dé sentido a sus frases ordenándolas correctamente</b>	.....Página 7
	Consejo 6: <b>Reduzca el número de sustantivos: las formas verbales dan fluidez al texto</b>	.....Página 8
	Consejo 7: <b>Utilice más activas que pasivas y no indique el sujeto si no es preciso</b>	.....Página 9
	Consejo 8: <b>Sea concreto, no abstracto</b>	.....Página 10
	Consejo 9: <b>Cuidado con los falsos amigos, las jergas y las siglas y abreviaturas</b>	.....Página 11
	Consejo 10: <b>Revise y compruebe lo que haya escrito</b>	.....Página 14

# Numbers – Precision (1)

Use only the **number of digits which are necessary and make sense** for the purpose of a clear communication

Rounding of numbers should take place at the latest phase of data processing and analysis.

- Tables with percentages: general rule is to **round to one decimal**
- For tables with absolute numbers, identify **the smallest number**, decide **how many digits to keep** for this number and then round all other entries to those digits

# Numbers - Precision(2)

- For target indicators always use the full precision of the indicator to assess whether the target has been met. The rounding should not change the situation
- Big numbers are difficult to grasp. It may be reasonable to round them and use the words millions, billions, etc.

# Example of rounding

GEO/TIME	1970
Belgium	9660154
Germany (until 1990 former territory of the FRG)	61194591
France	:
Italy	53685300
Luxembourg	338500
Netherlands	12957621

Before

Member State	Population in 1970
BE	9 660 000
DE	61 190 000
FR	:
IT	53 690 000
LU	340 000
NL	12 960 000

After

# Tables



Less  
is **more**  
(effective)

See the bar chart, pie chart, and choropleth maps versions as well

# Truthful

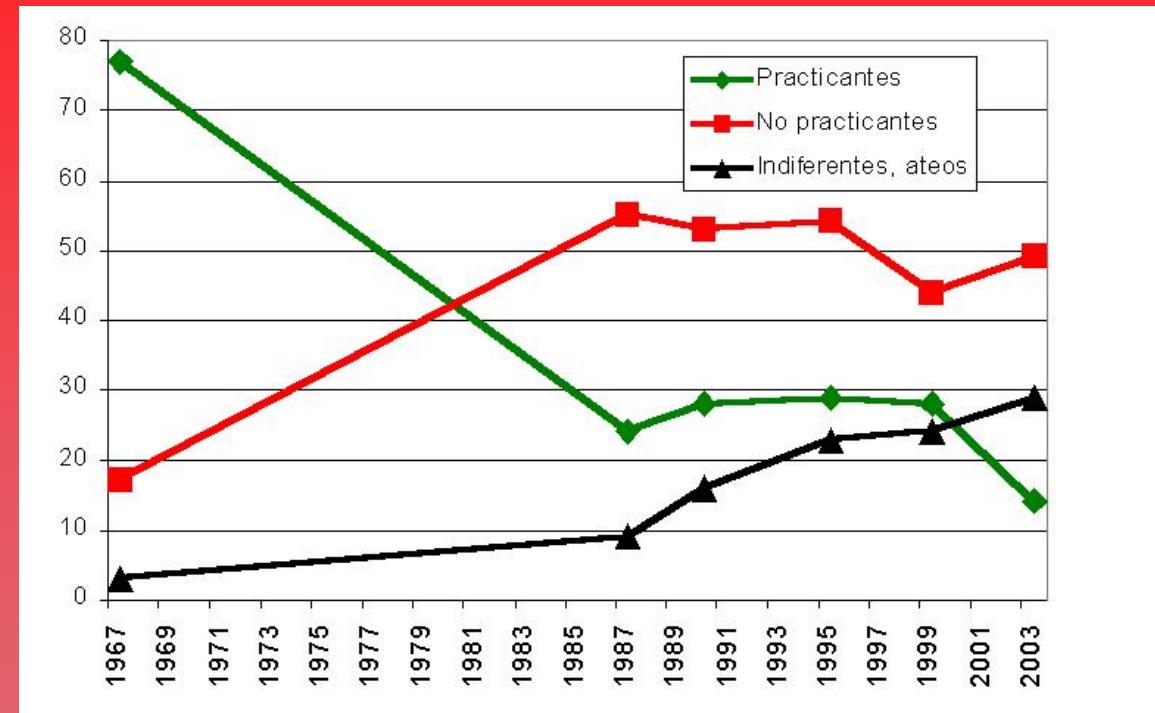
To communicate **clearly** you must avoid some misleading practices

# Misleading practices

# Scales are not regular

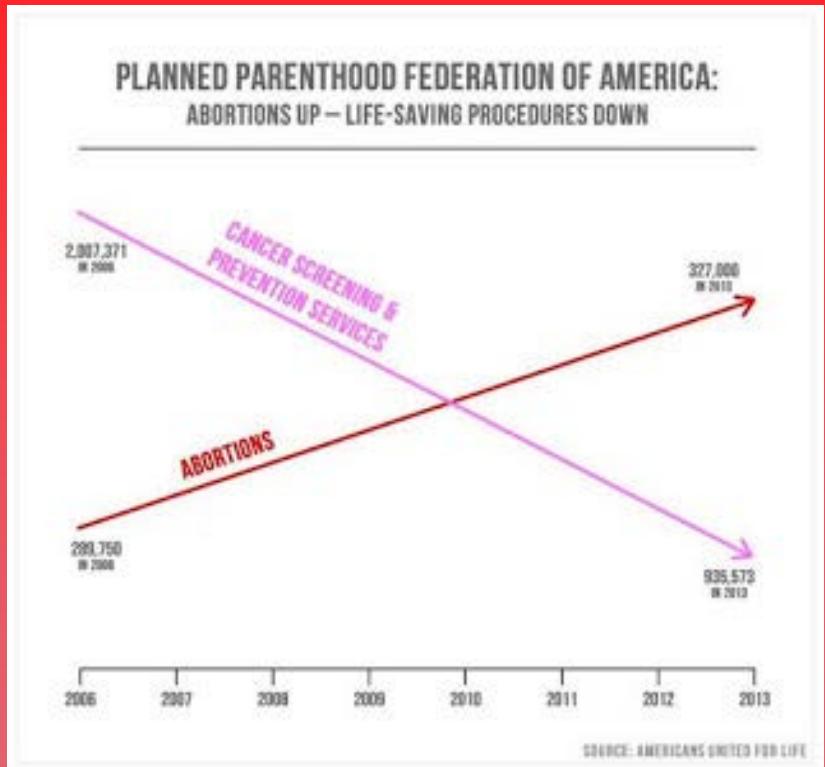


Before

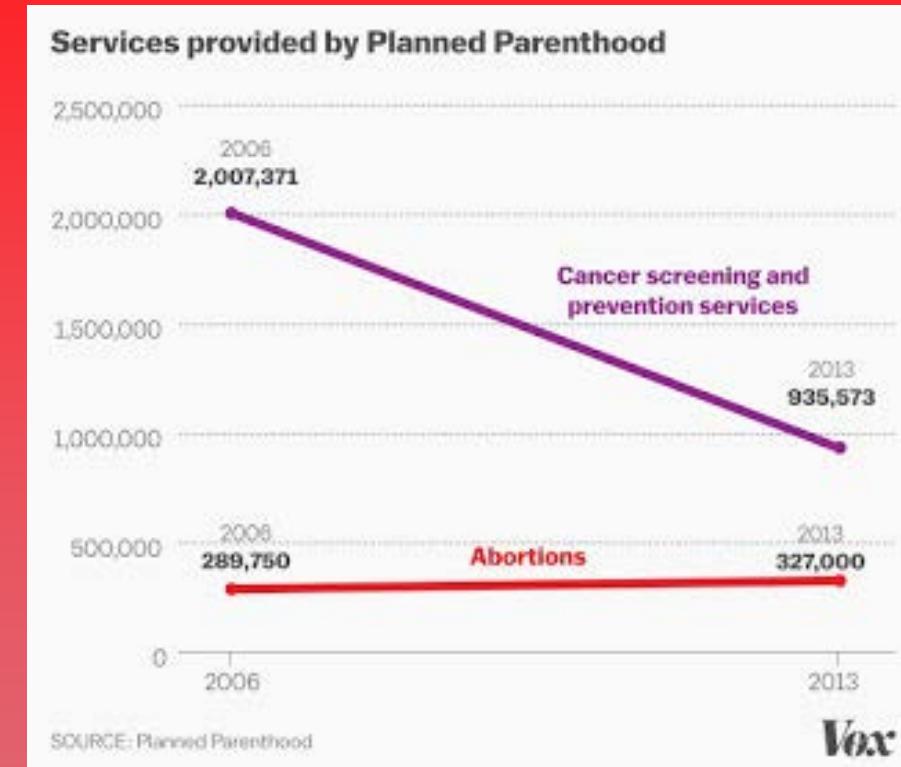


After

# Dual axes

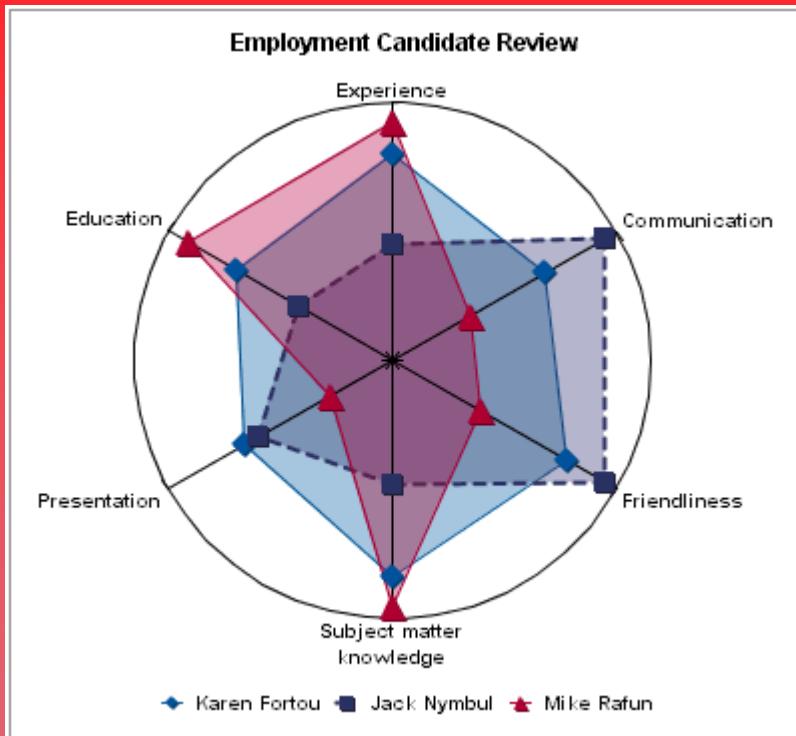


Before

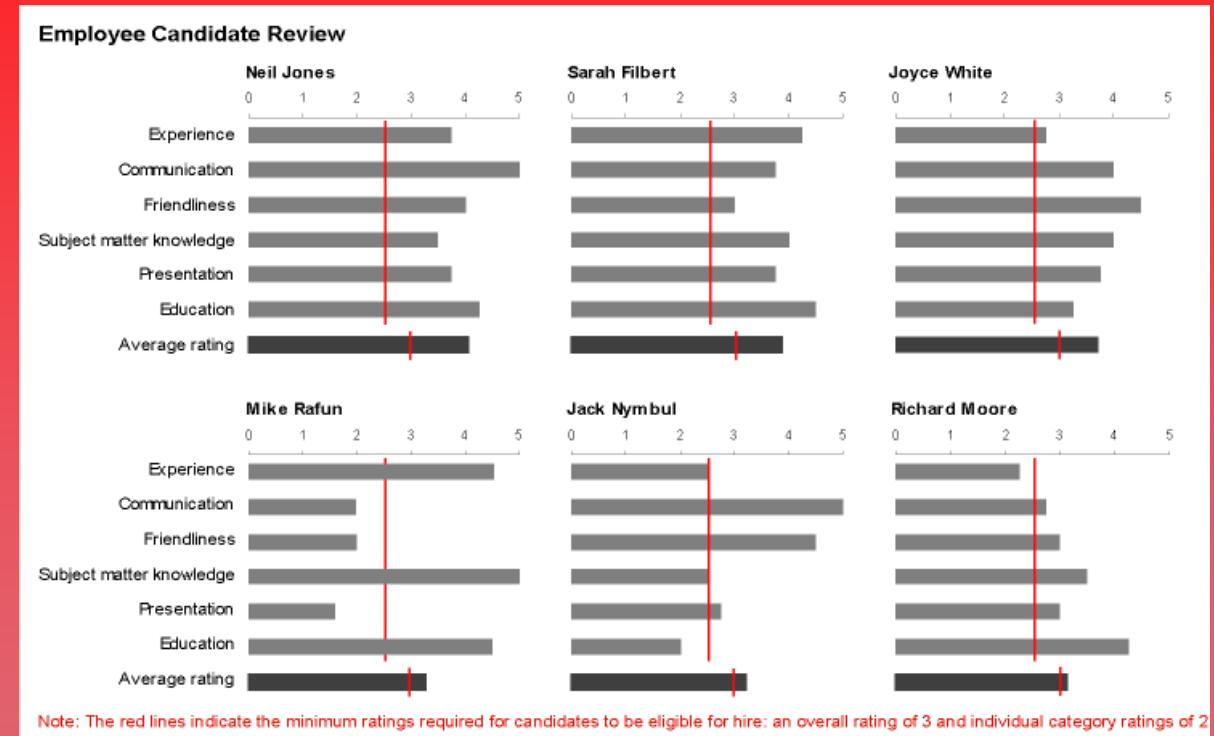


After

# Use the appropriate chart

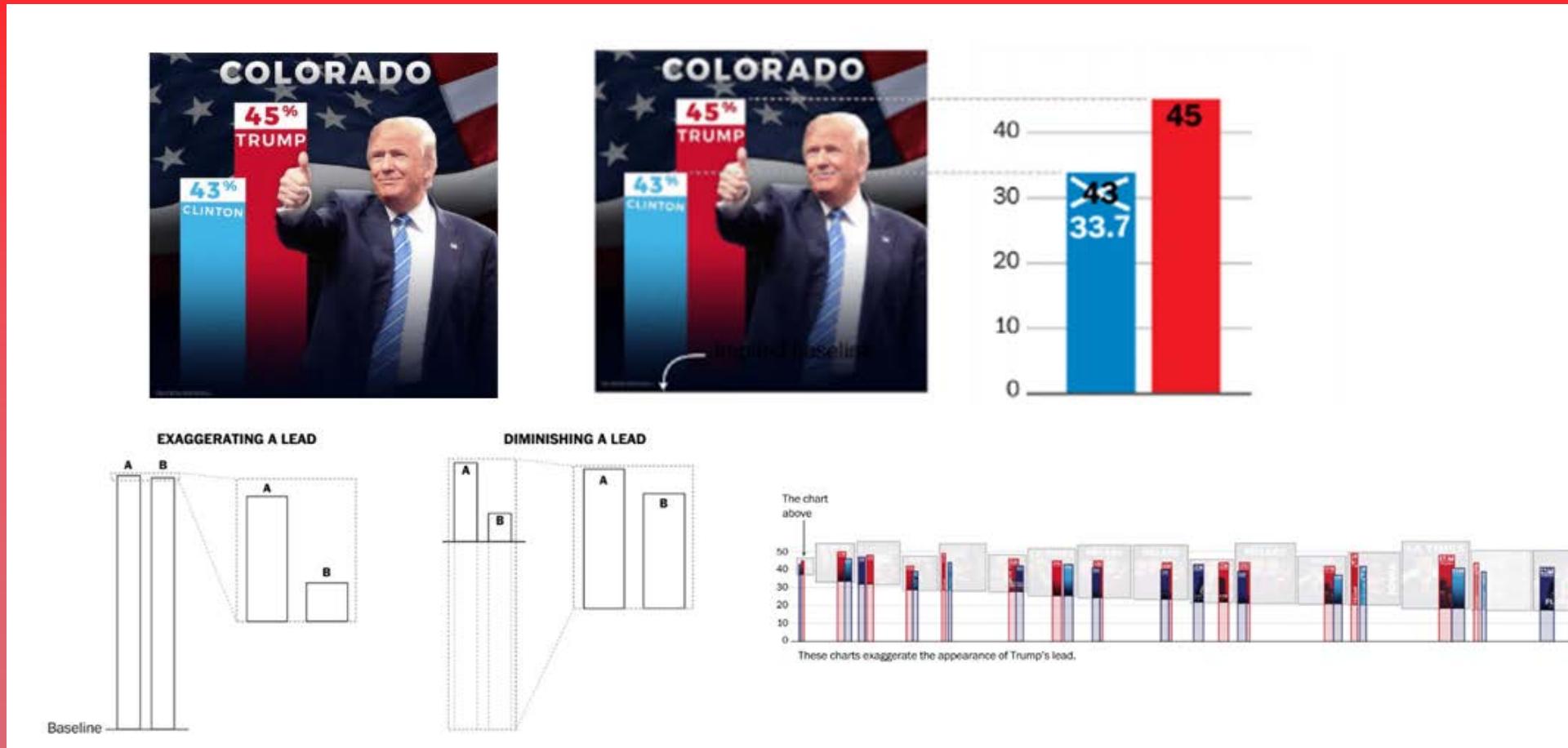


Before

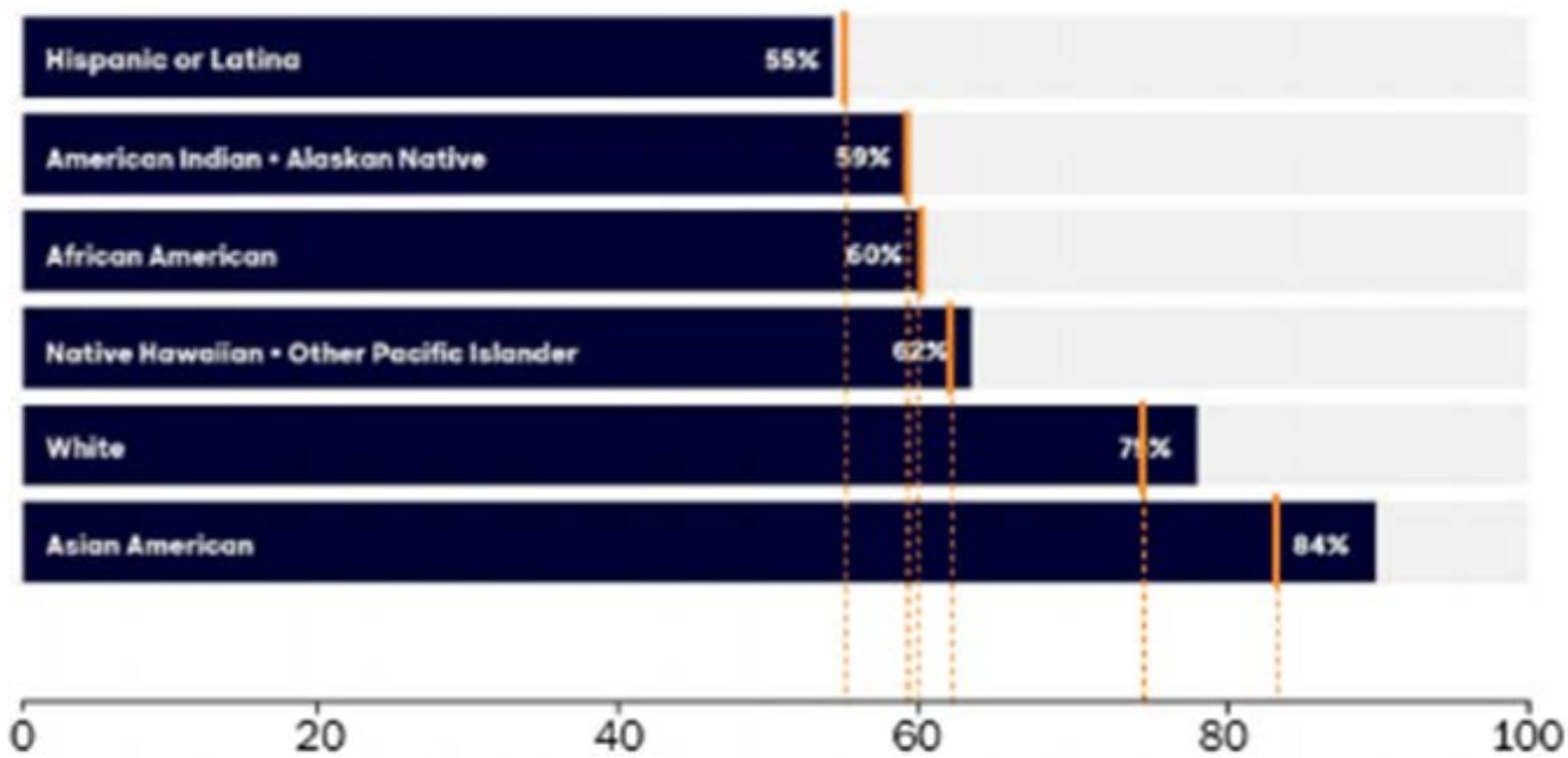


After

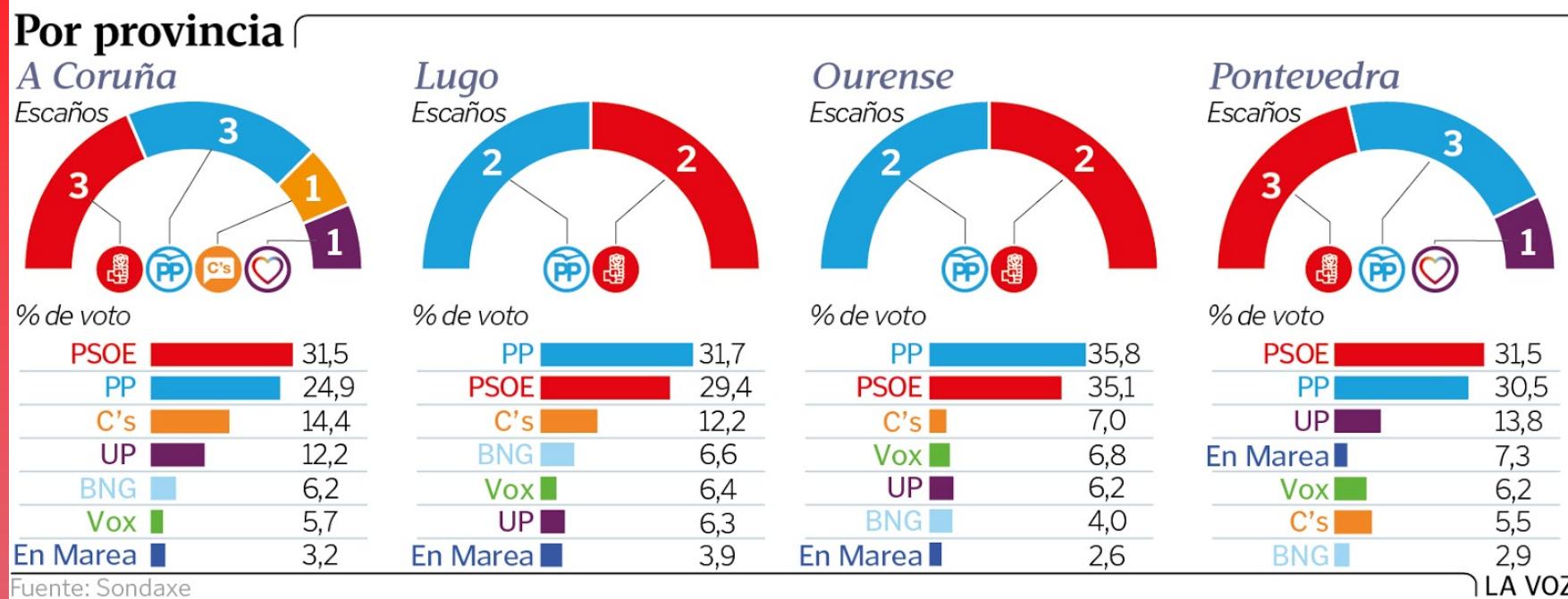
# Bar charts length. Always start at 0



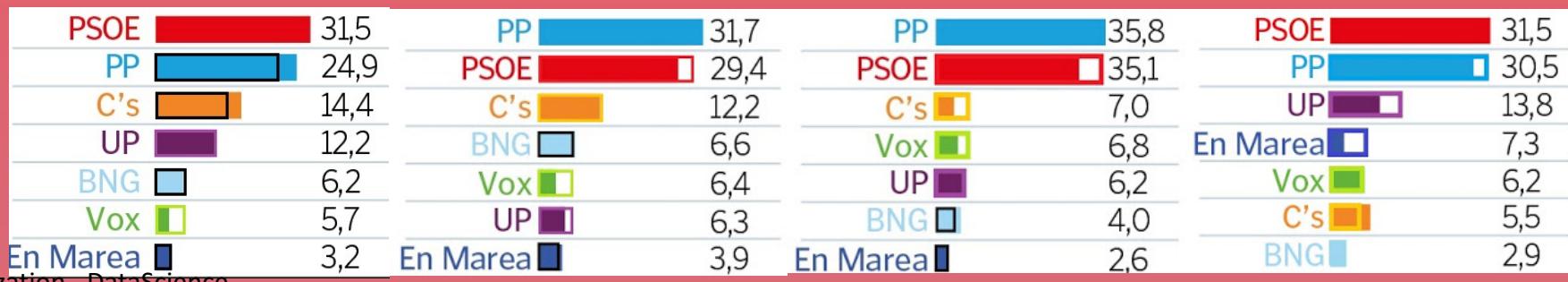
# Use scales wisely (1)



# Use scales wisely (2)

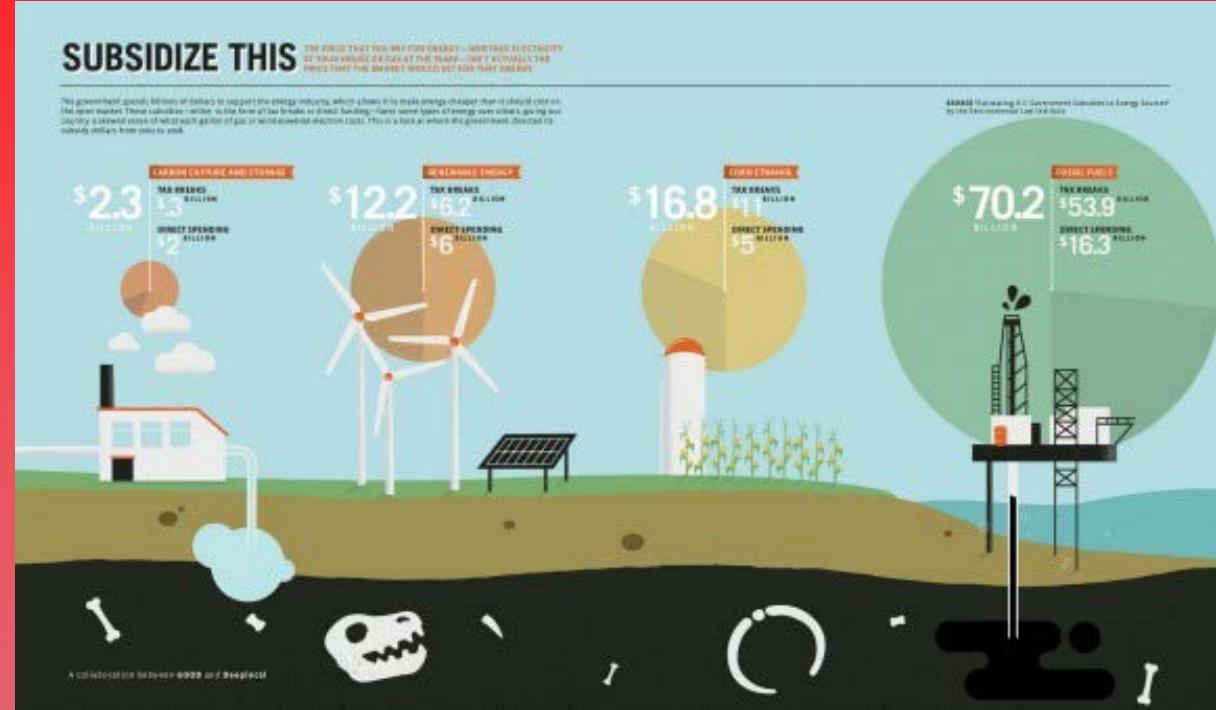
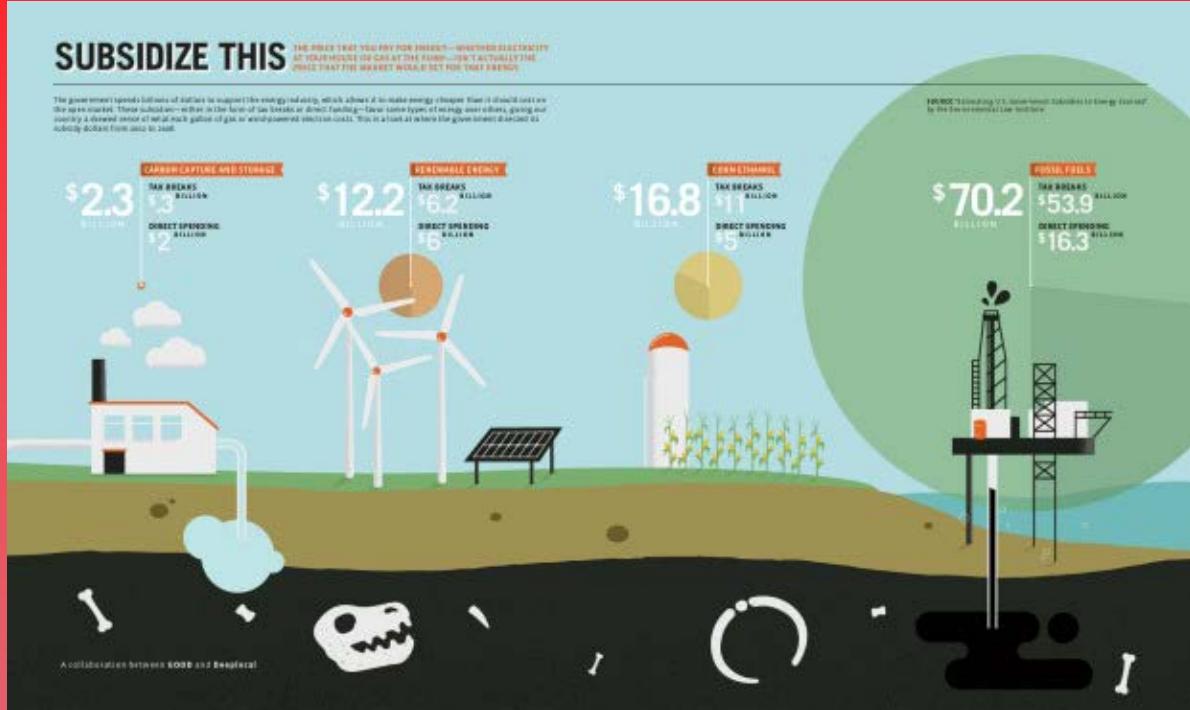


Before

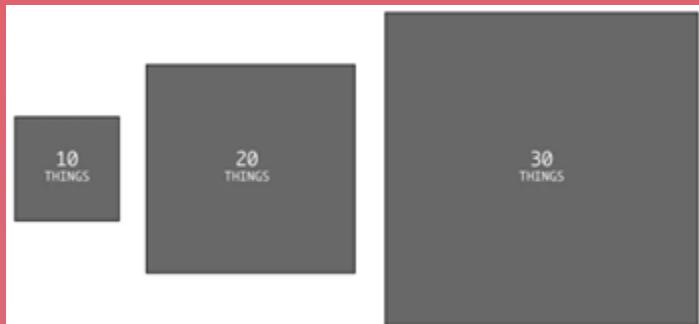


After

# Area vs diameter, be careful!

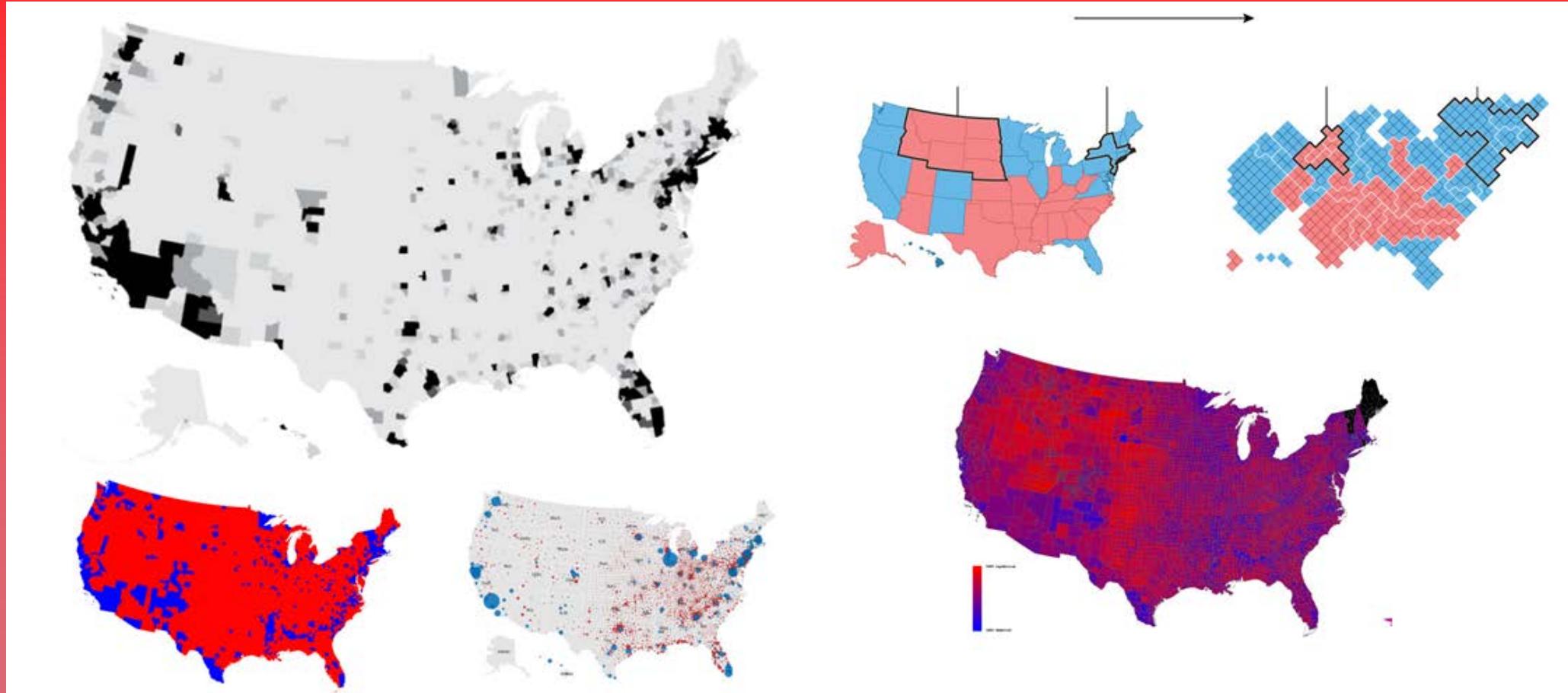


Before

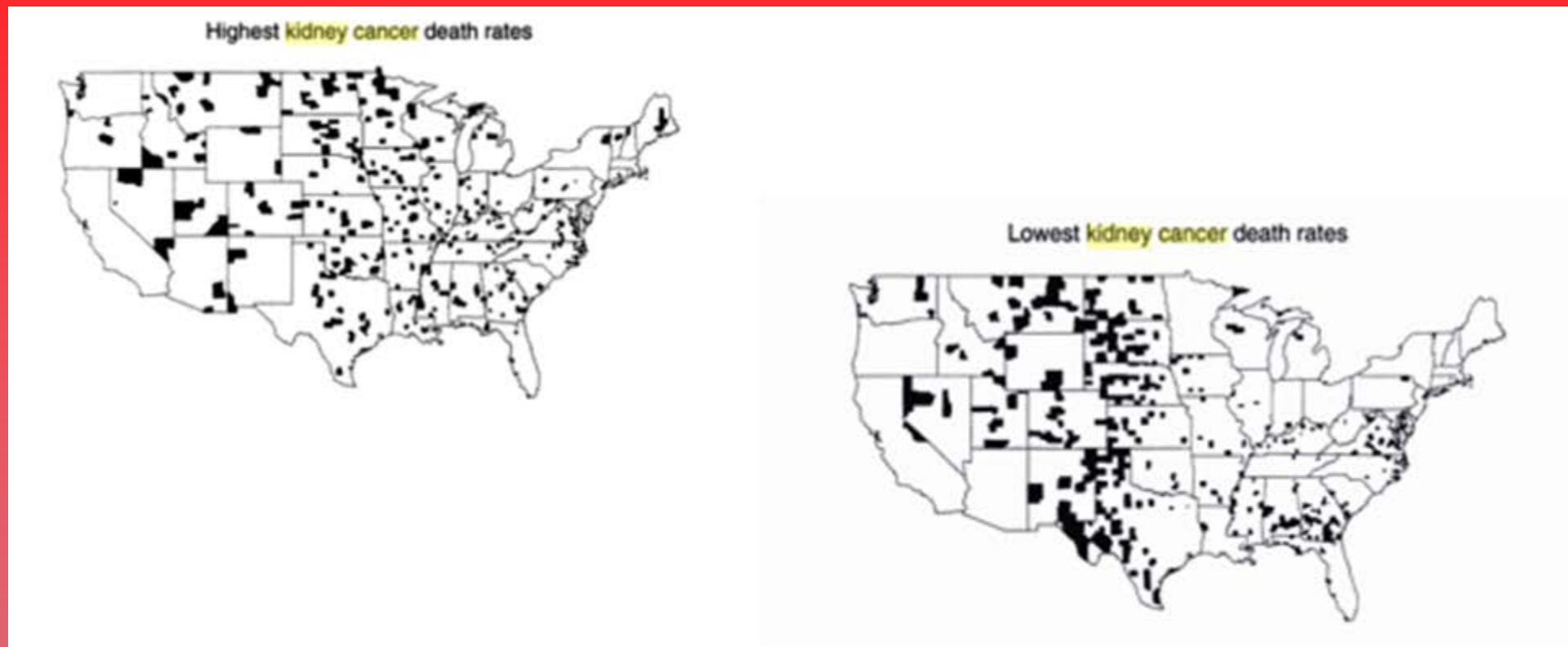


After

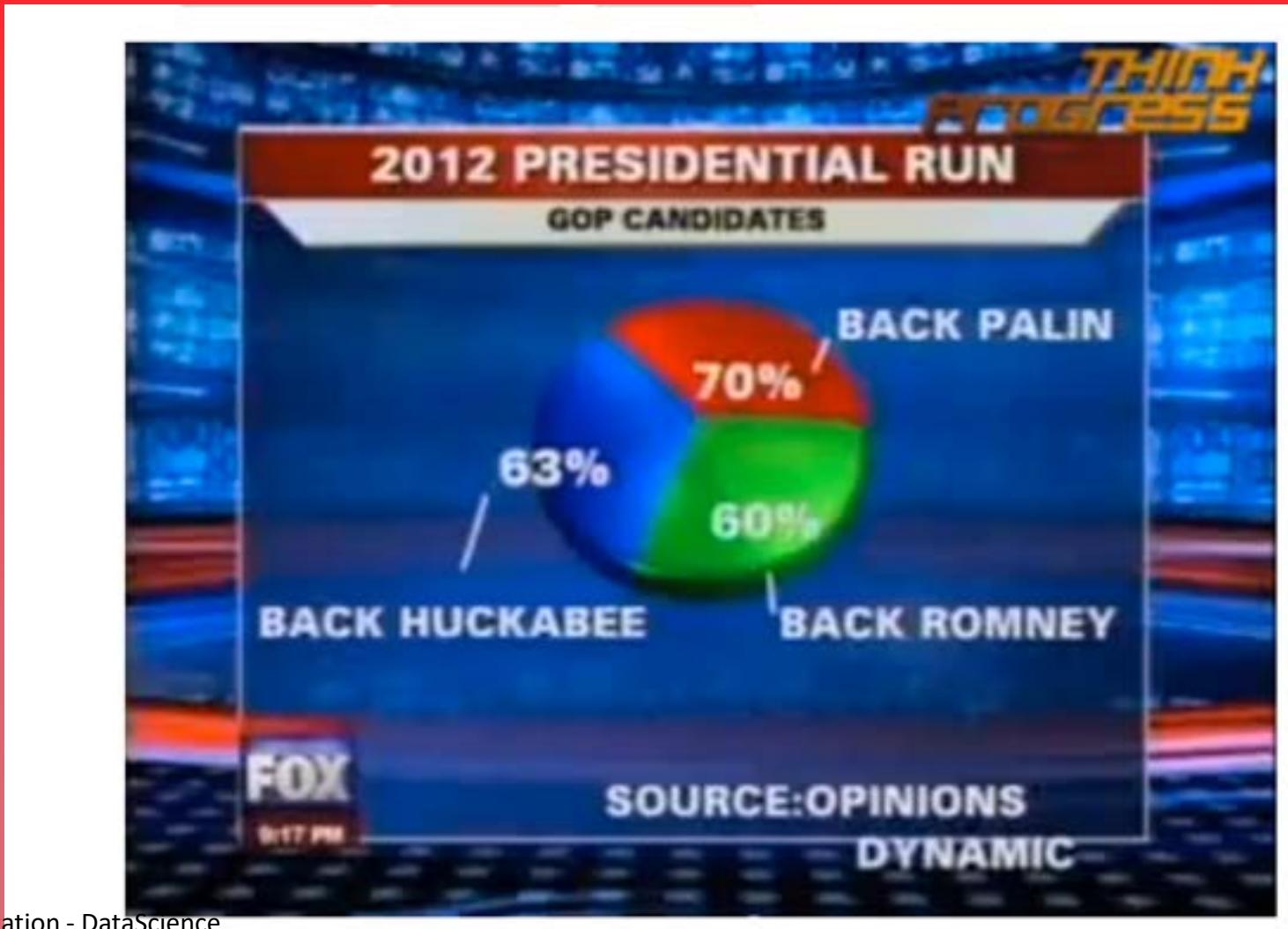
# Don't oversimplify with aggregation



# Small size effect



# Basic arithmetics. Follow conventions

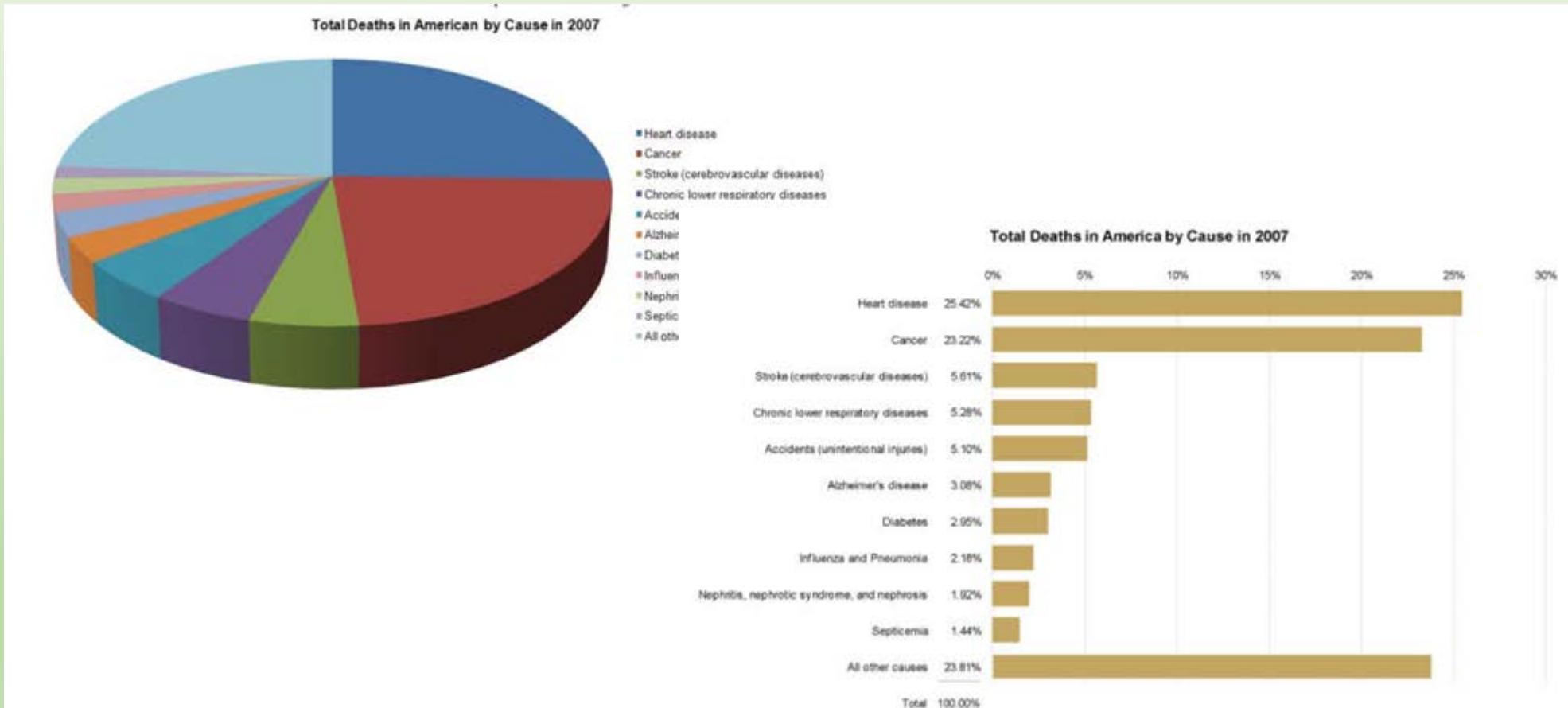


# Truthful

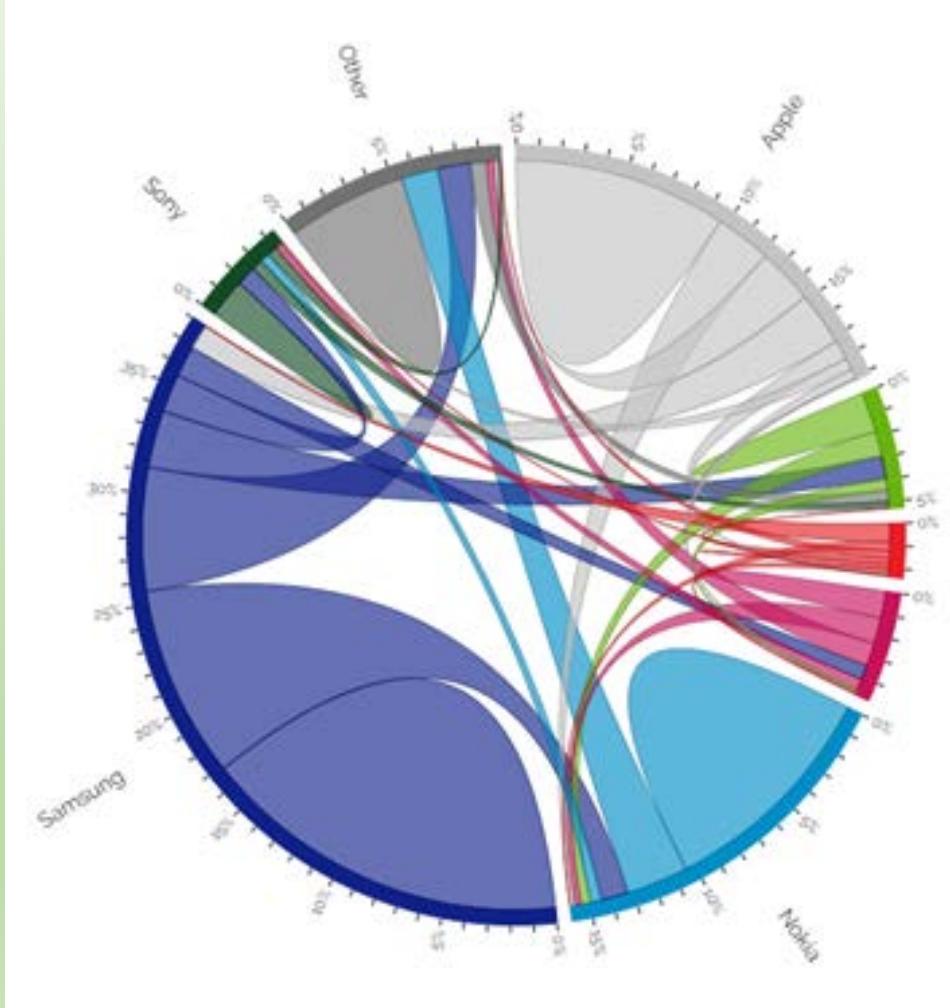
To communicate **clearly** you must follow some best practices

# Best practices

# Use the appropriate chart

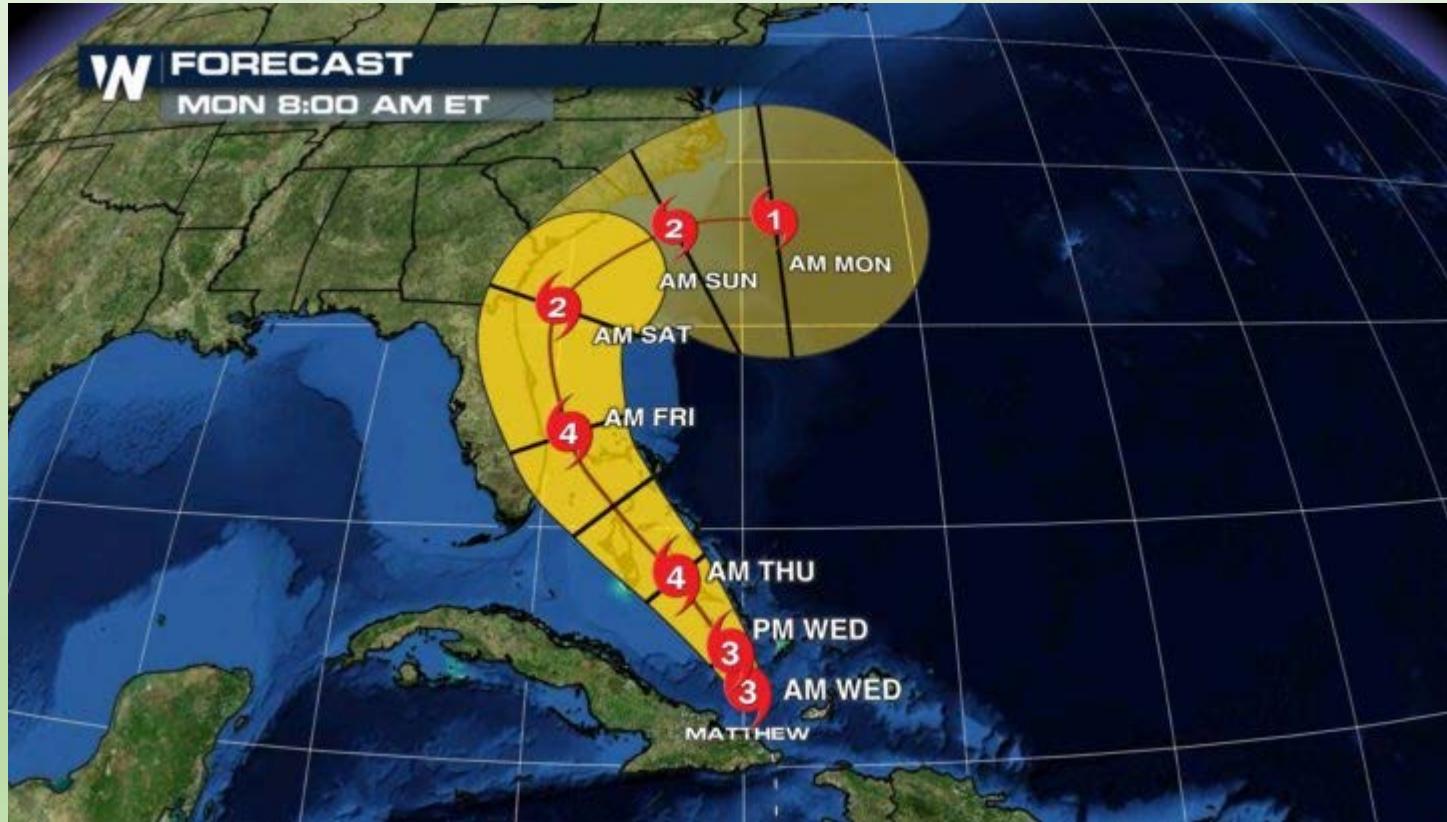


# Offer clear legends: title, key to encodings, axes



You may even [offer a tutorial](#), if the type of chart is not usual.

# Tell uncertainty if it is relevant



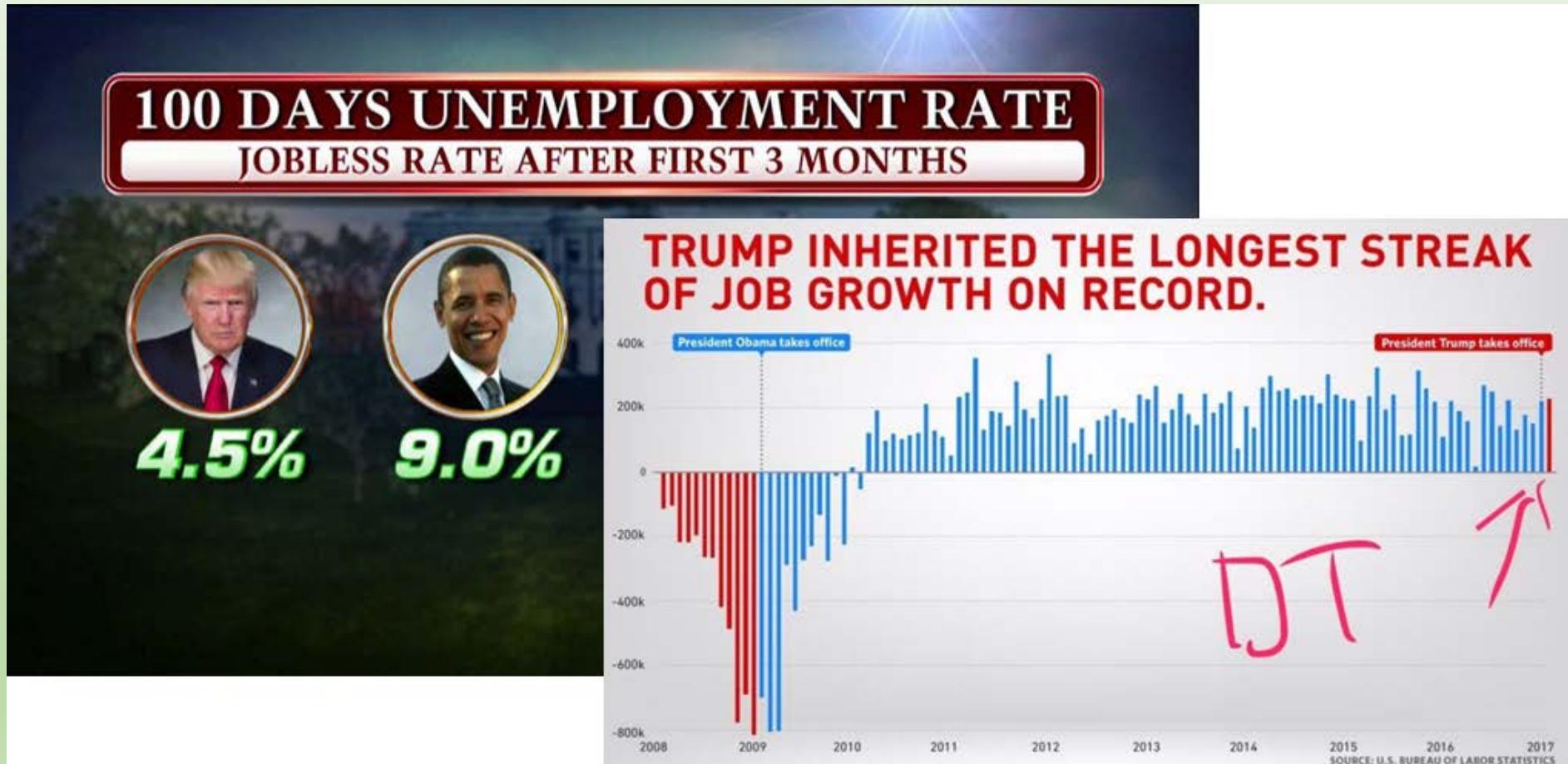
How do you interpret this graph?

Which areas will be affected by the hurricane?

# Disclose the **origin** of your data

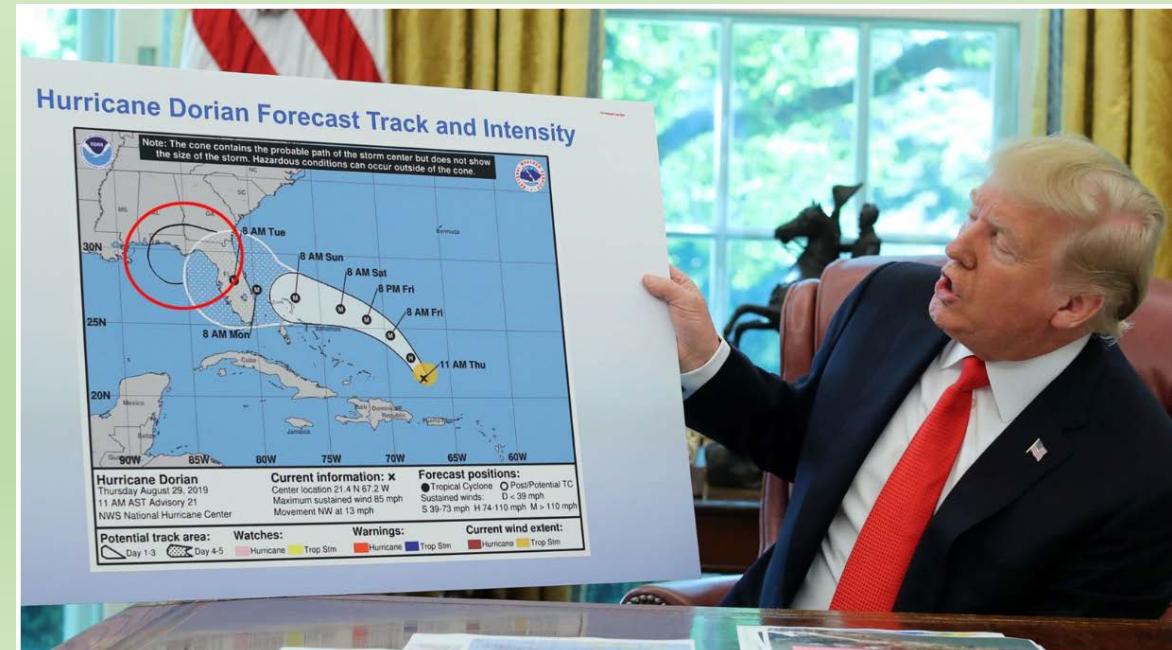
- Readers should be able to go to the facts behind your chart and check them
- Tell your sources
- Link to them if possible

# Give context to your data



# Ethics

- Become aware from your own biases
- Don't twist facts to fit your agenda
- Don't share before checking



# Storytelling

Engage your users in your message

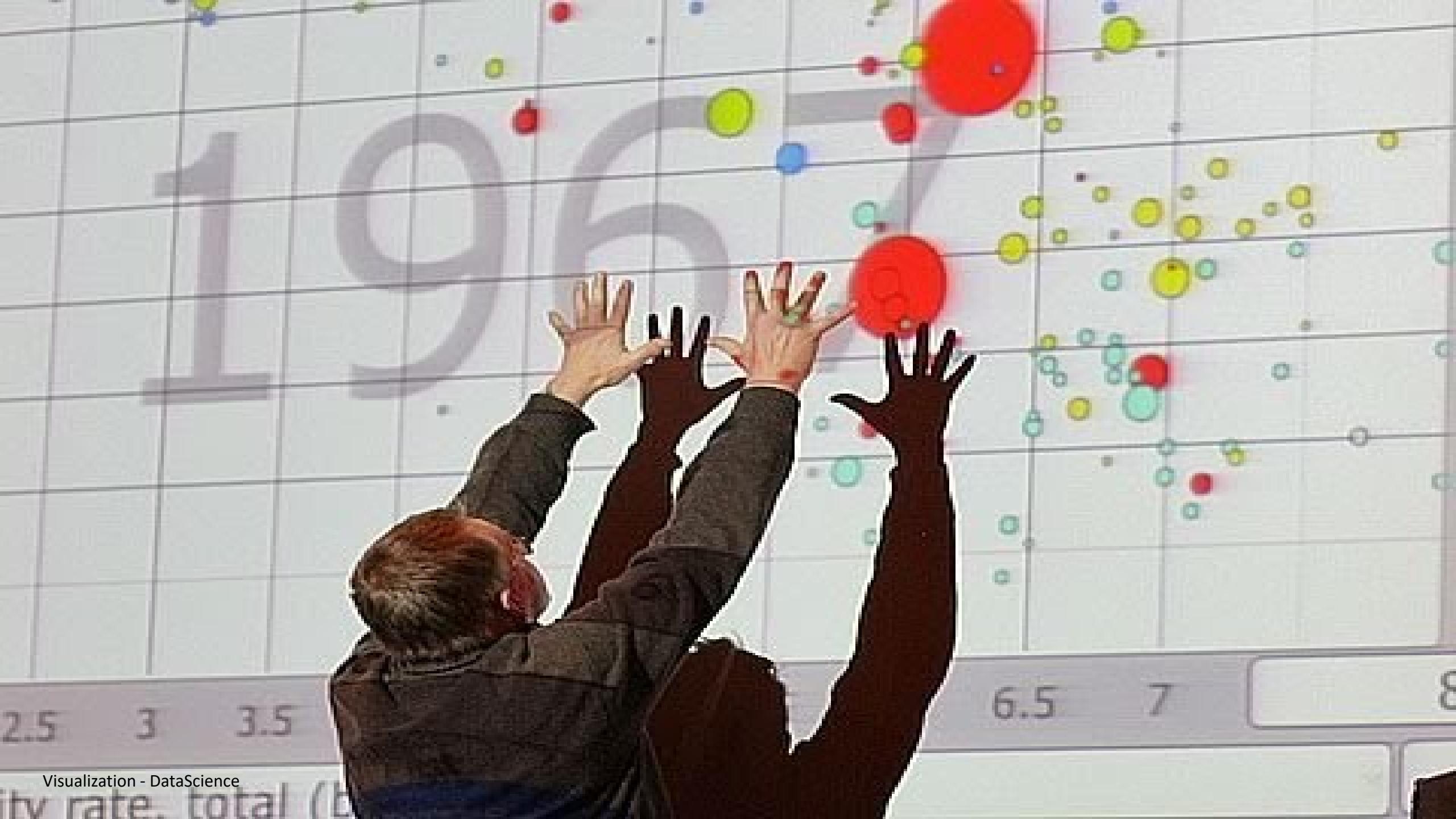


Engagement

Memorability

Emotional  
connection

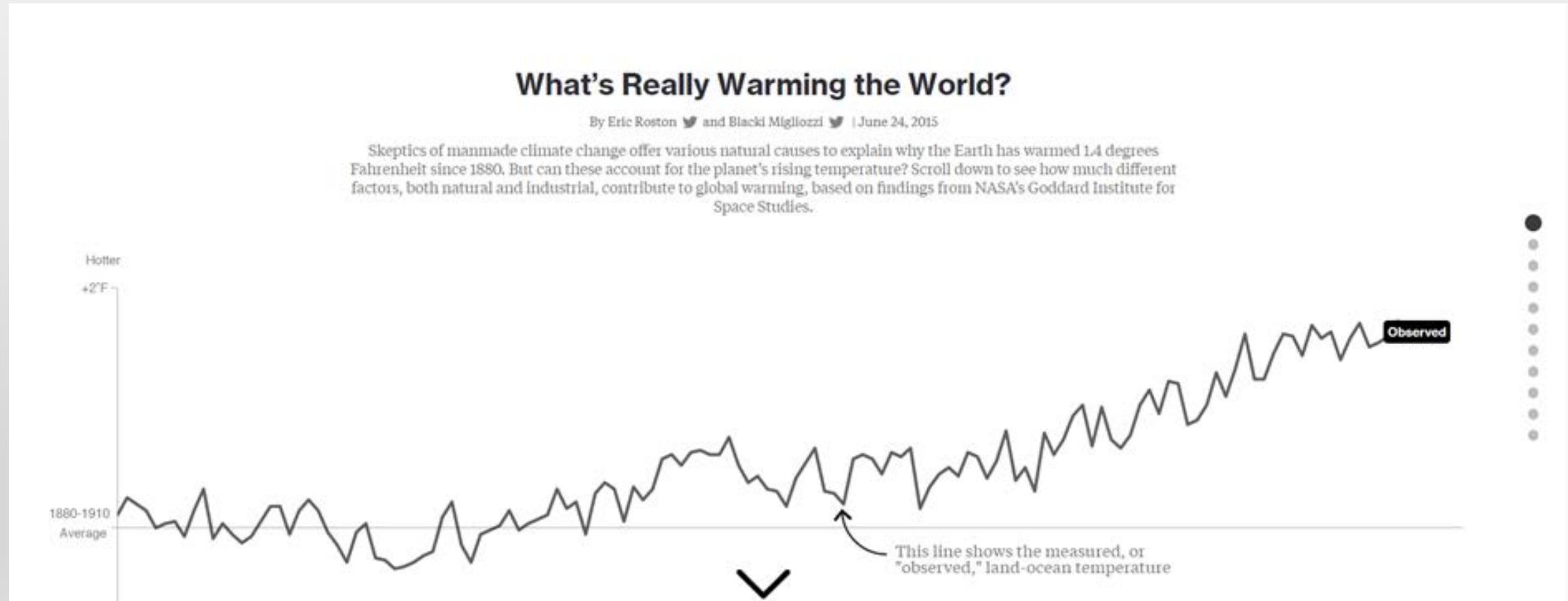
Storytelling goes beyond data understanding



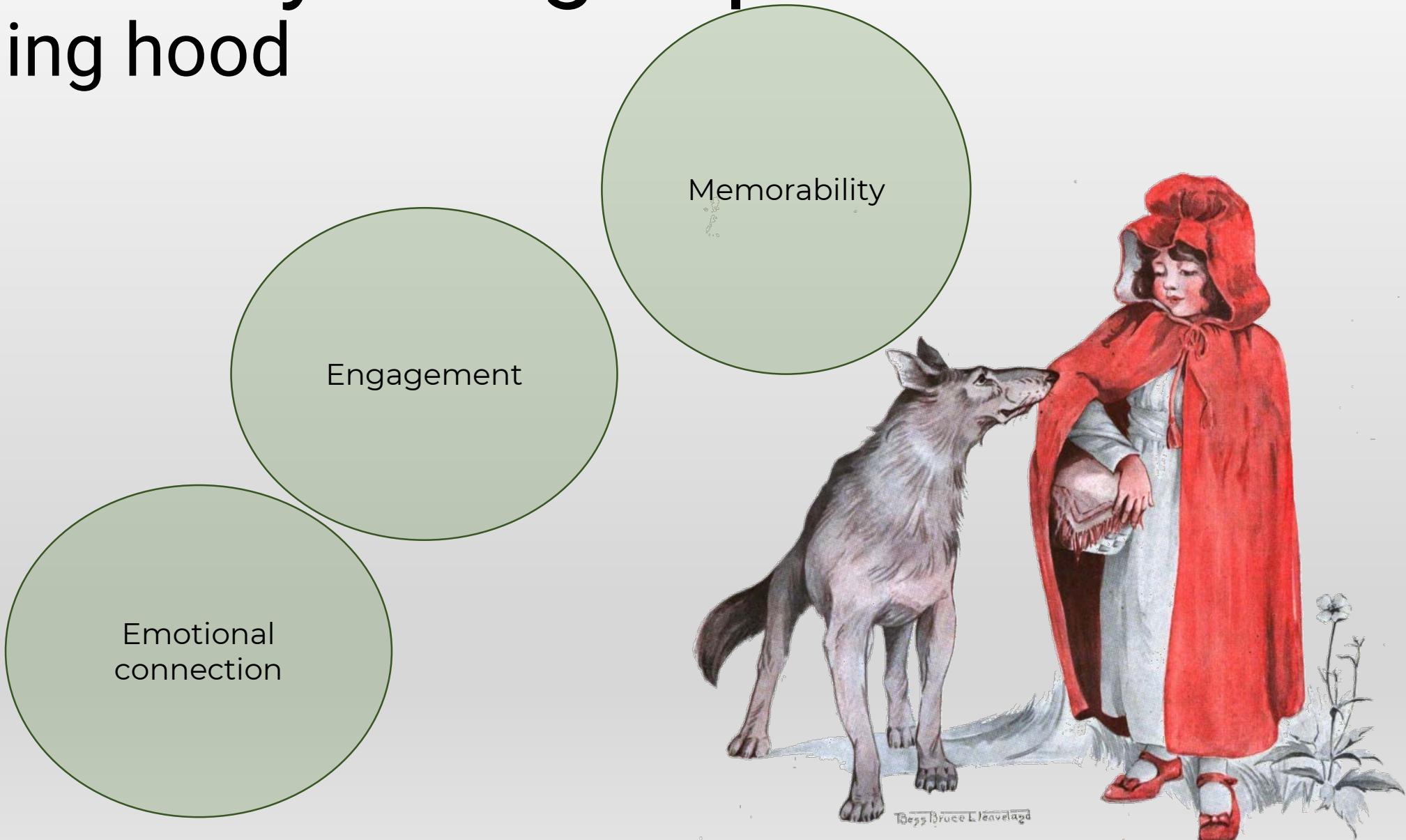
# Wonders of storytelling

- Remembering is easy with a story
- Magic from Plot-Twist-End: data are stories
- Captivation, emotionally engage
- Takes you on a journey, you don't want to turn away

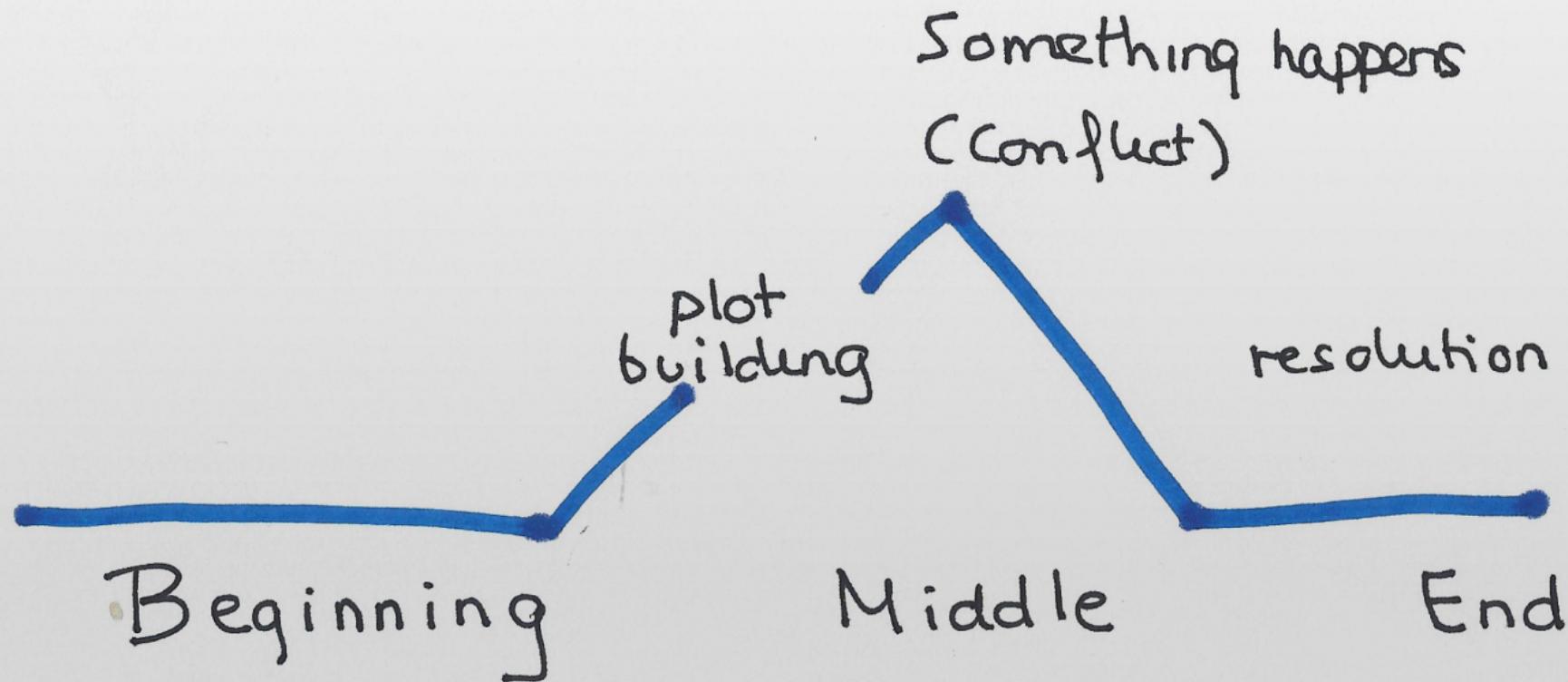
# Storytelling example 2



# A child storytelling experience: Red riding hood



# The story



Identify  
beginning-  
middle-end

What tricks do  
they use for  
engagement?

- The interactive UK energy consumption guide
- How many slaves work for you

What is most  
memorable?

# Wrap-up story-telling in infovis

- Engagement
  - Use striking language / strong imaginary
  - Drive user focus
  - Identify emotional drivers
- Understanding
  - Users lack background information: provide context
  - Unclear visual encodings: explain piecemeal
  - Overwhelming: soft start

# Wrap-up story-telling in infovis

- Memorability
  - People forgets the message: connect symbols to reality by text, iconography or actual photos
  - Offer a plot
- Emotional connection
  - Offer fun, Humanize stories
  - Prioritise authenticity: real stories, use details such as names, real pictures...



# 3-minute story rehearsal

Try to convey your story in 3-min

- It must articulate your unique point of view
- It must convey what's at stake
- It must be a complete sentence.

# Knowledge discovery

Use visualization to help discover **trends, outliers, patterns**

# Knowledge discovery 1

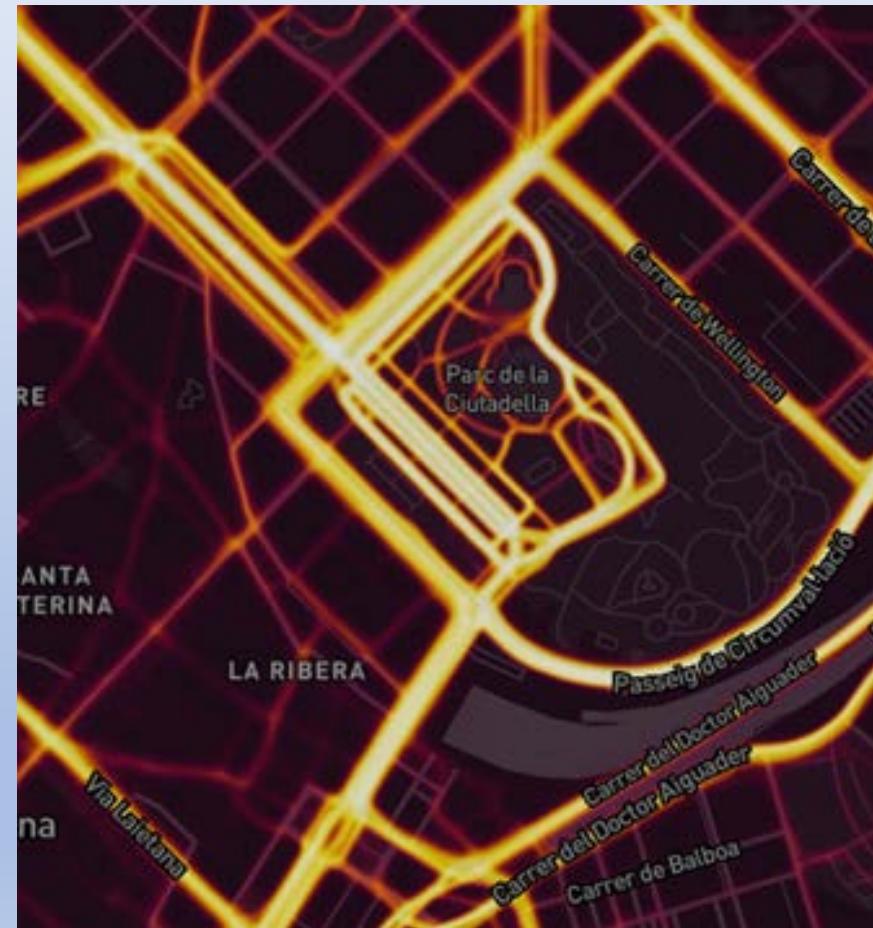
What was the cause of the cholera epidemics in London in 1854?

[John Snow](#) discovered it through visualization



# Knowledge discovery 2 (incidental)

Strava heat map made visible [military bases](#)



# Knowledge discovery 3

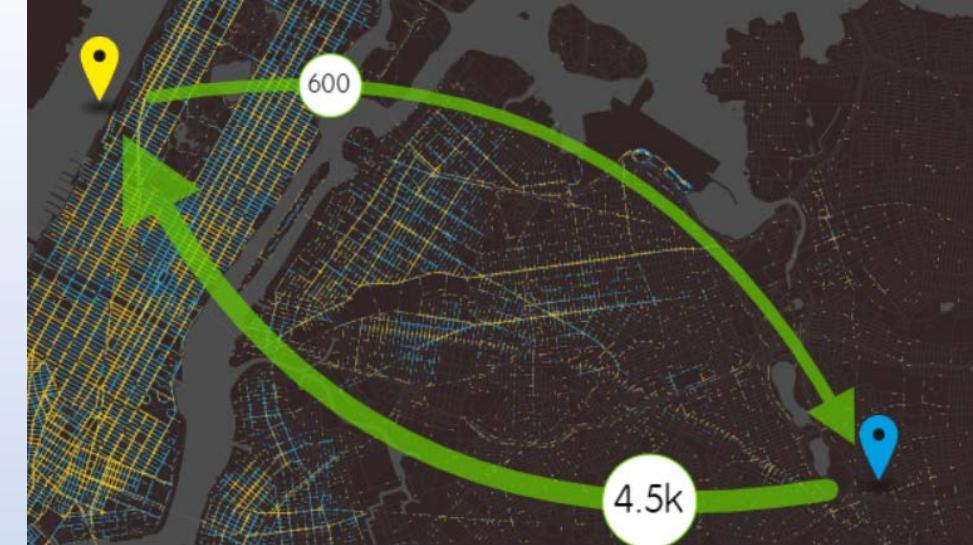
The New York Times

## You Draw It: What Got Better or Worse During Obama's Presidency

By LARRY BUCHANAN, HAEYOUN PARK and ADAM PEARCE JAN. 15, 2017

# NY Cabs open data

- Lots of interesting stuff to learn
  - Taxi GPS data helps researchers study Hurricane Sandy's effect on NYC
  - traffic <https://engineering.illinois.edu/news/article/9717>
  - Exploring New York City taxi trails and sharing our way to a more sustainable urban future  
<http://hubcab.org/#13.00/40.7219/-73.9484>
  - Public NYC Taxicab Database Lets You See How Celebrities Tip  
<http://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546>





# Perception and colours

Gestalt laws

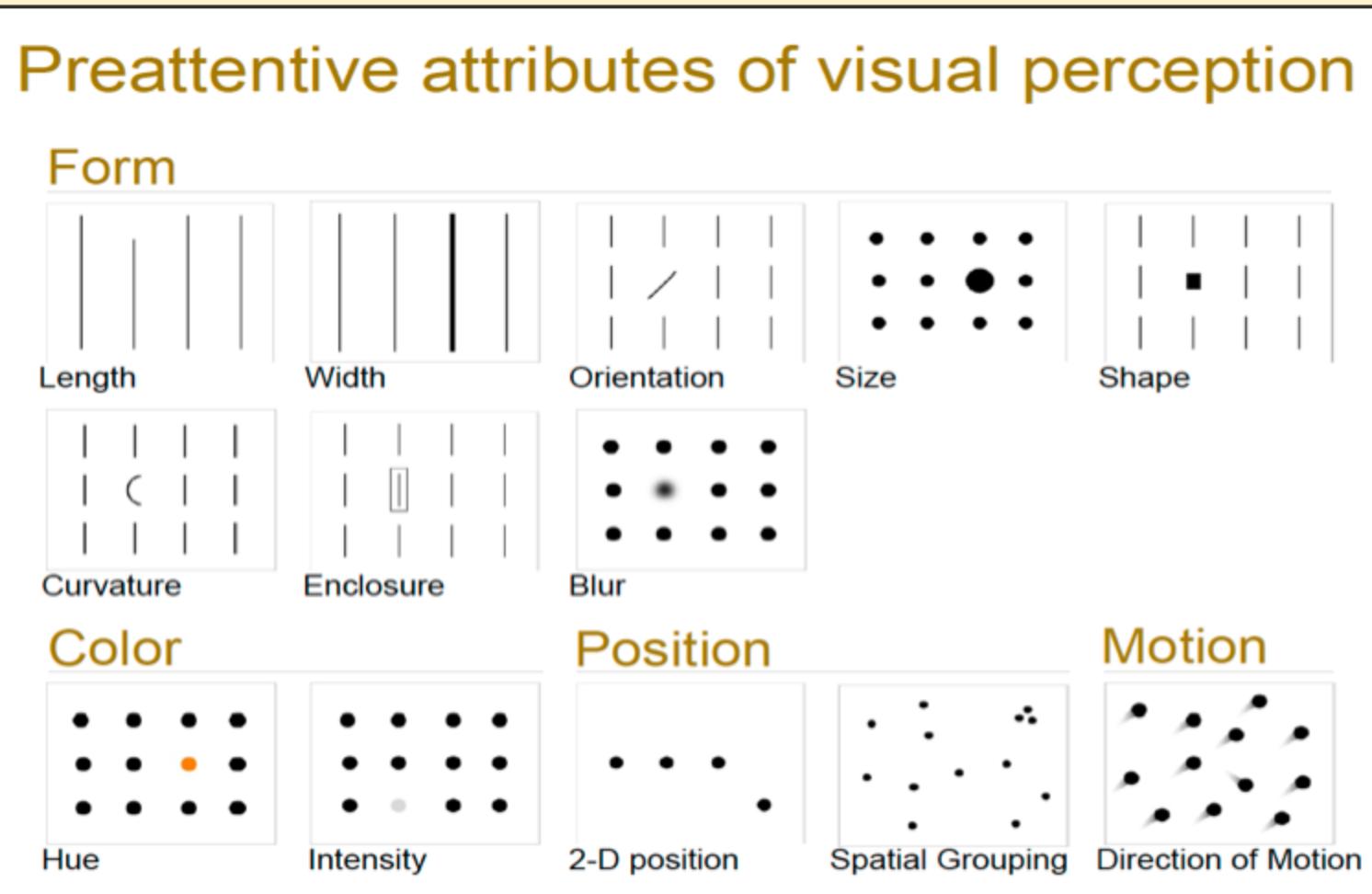
Preattentive properties

Colours

# Preattentive properties

- Certain visual properties are detected immediately by low-level visual system
  - Immediately is <200-250 ms
- They “pop-out” without requiring serial search
- Not affected by distractors

# Preattentive processing and processing channels



# Test your abilities (3)

- Perception in visualization / Christopher G. Healey  
<https://www.csc2.ncsu.edu/faculty/healey/PP/index.html>

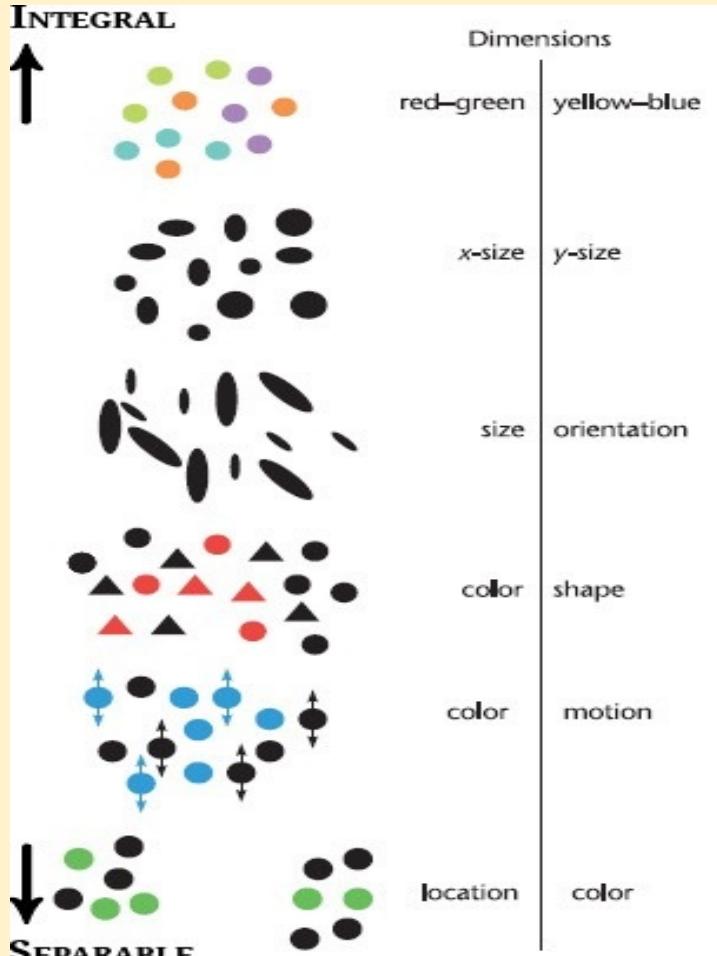
# Design principles

- G5.2 “Use **different visual channels** to display aspects of **data** so that they are visually distinct”
- G5-7 “For **maximum popout** a symbol should be the **only object** in a display that is **distinctive** on a particular feature channel; for example, it might be the only item that is colored in a display where everything else is black and white.”

# Design principles

- G5-8 “Use **positively asymmetric** preattentive cues for **highlighting**”
- G5-9 “For highlighting, use whatever feature dimension **is used least** in other parts of the design”
- G5-10 “When color and shape channels are already fully utilized, consider using **motion or blink highlighting**. Make the motion or blinking as subtle as possible, consistent with rapid visual search”

# Combination of dimensions: integral and separable



- Integral dimensions are seen together
- Separable dimensions are seen individually

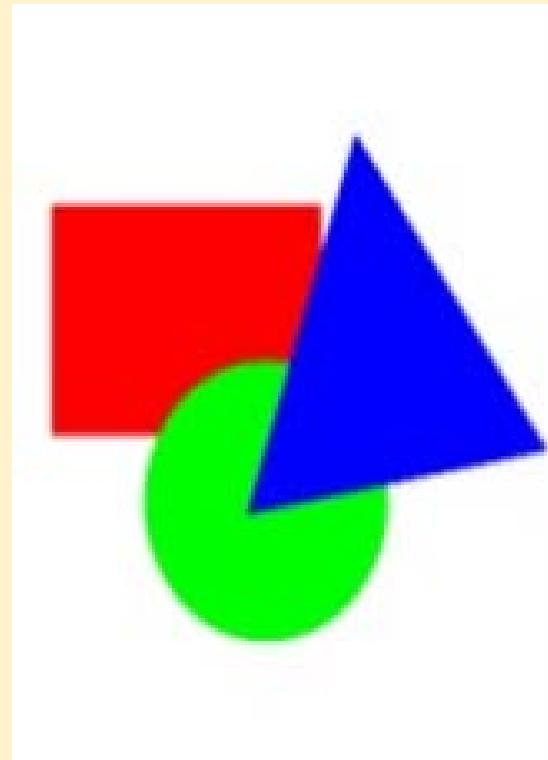
# Design principles

- G5.14 “If it is important for people to respond holistically to a combination of two variables in a set of glyphs, map the variables to integral glyphs properties”
- G5.15 “If it is important for people to respond analytically to a combination of variables, making separate judgments on the basis of one variable or the other, map the variables to separable glyph properties”

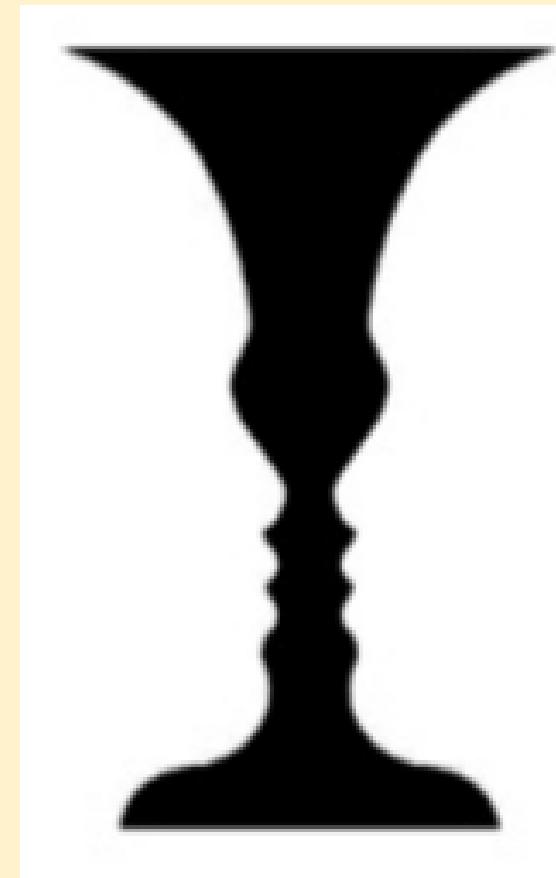
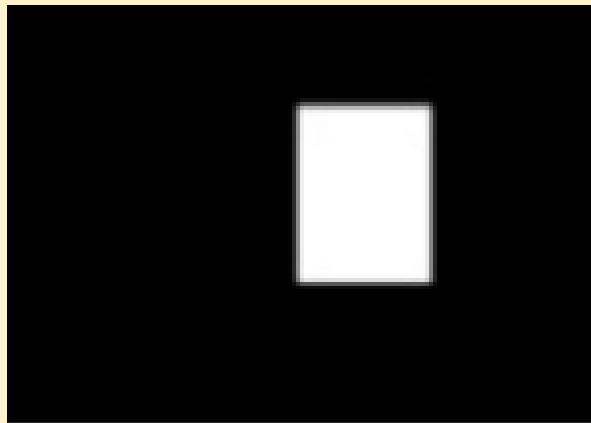
# Design principles

- G5.11 “To make symbols in a set **maximally distinctive**, use **redundant coding** wherever possible; for example, make symbols differ in both shape and color”
- When the visual query implies a **conjunction query** (searching for two attributes at the same time) G5.13 “consider coding one using **motion or special grouping** and the other using a property such as **color or shape**”

# Gestalt laws: Simplest forms



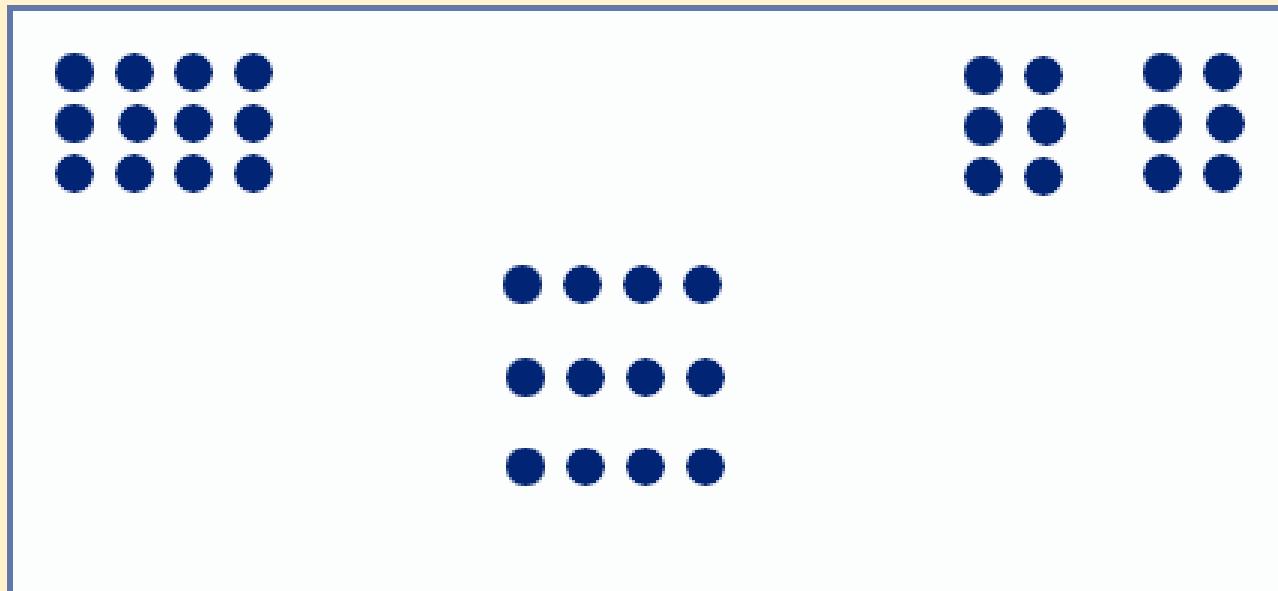
# Figure and ground



# Design principle

- G5.3 “To make *symbols* easy to find, make them *distinct* from their background and from other symbols”.

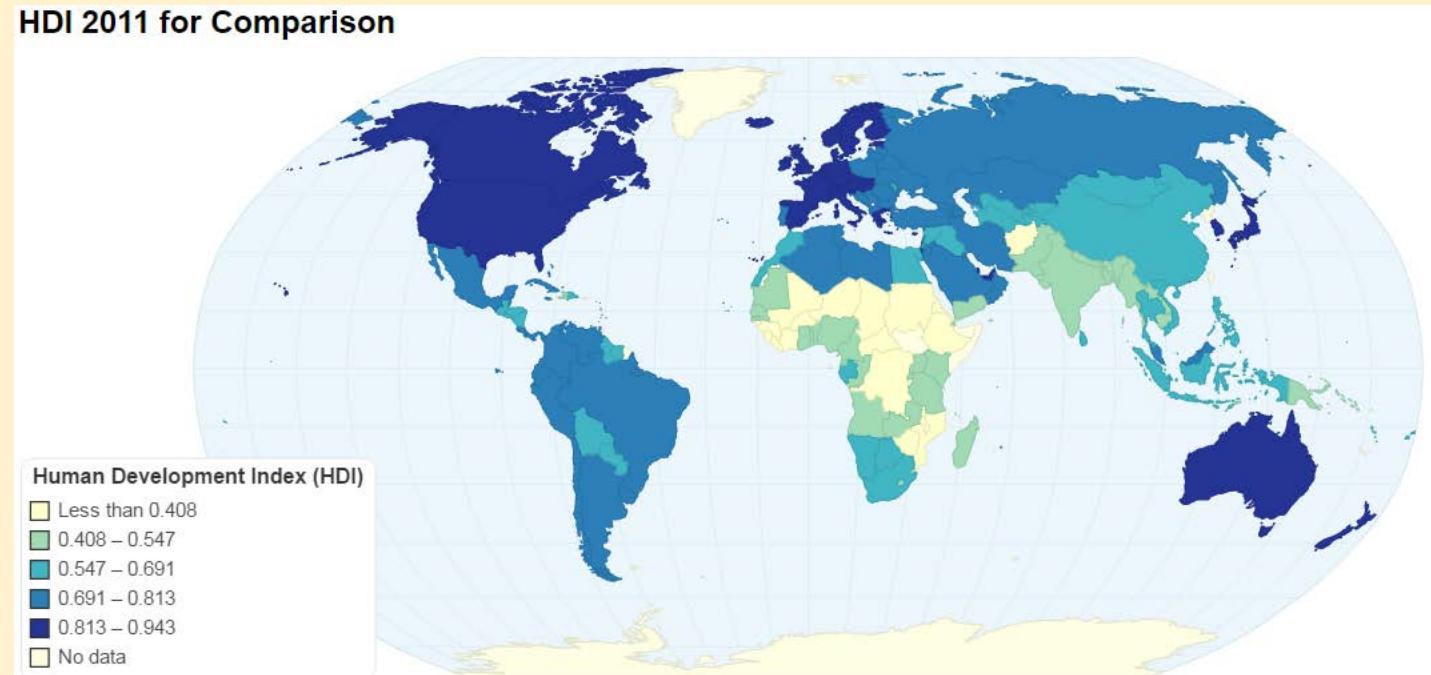
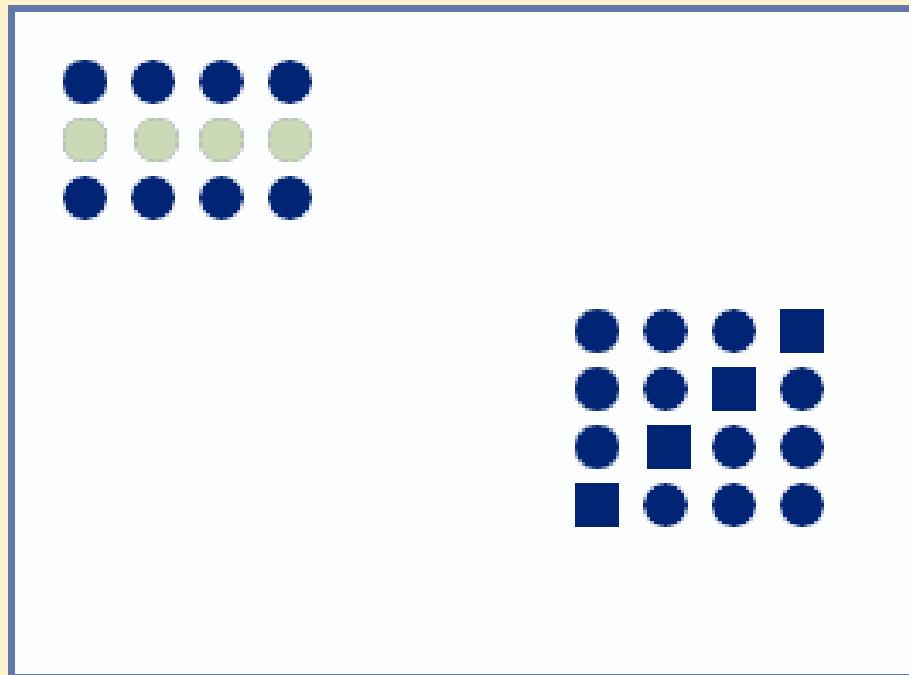
# Proximity



# Design principle

- G6.1 “Place symbols and glyphs representing related information close together”

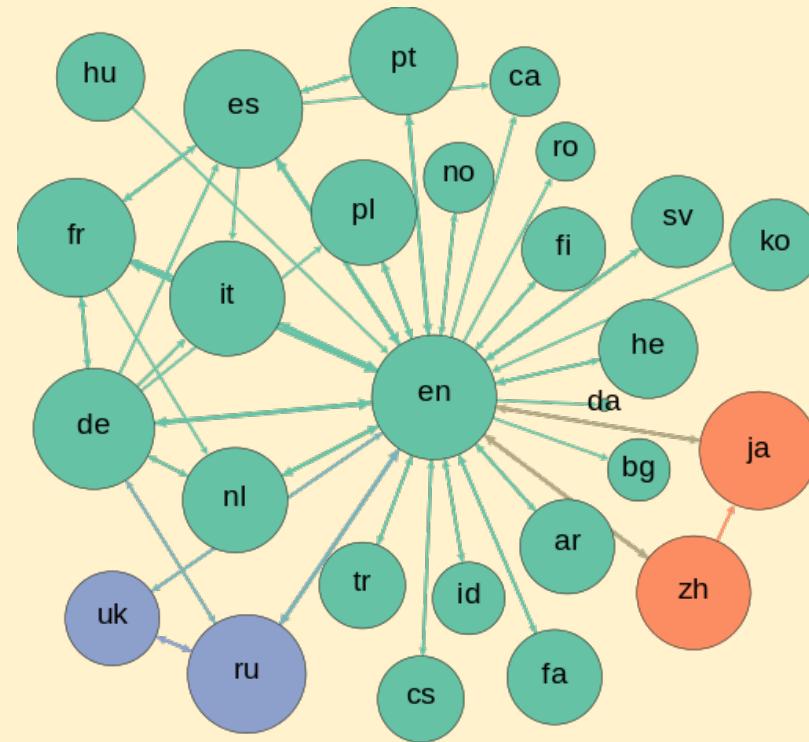
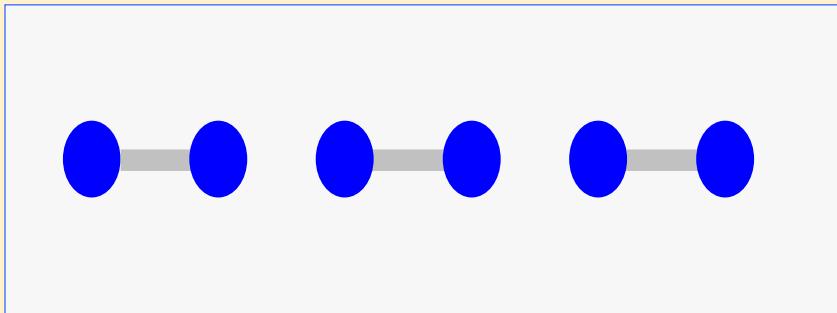
# Similarity



# Design principle

- G6.2 “When designing a **grid layout** of a data set, consider coding rows and/or columns using low-level visual channel properties, such as **color and texture**”

# Connectedness



Source: Computermacgyver (Own work)

[CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons

# Design principle

- G6.3 “To show **relationships** between entities, consider **linking** graphical representations of data objects **using lines or ribbons** of colour”

# Continuity



# Common fate (synchrony)



# Symmetry

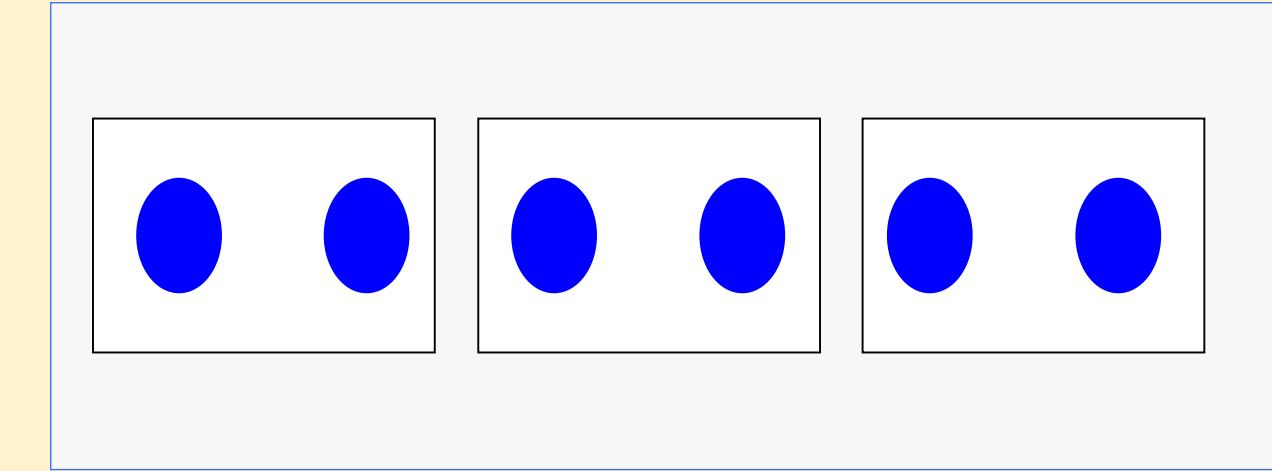
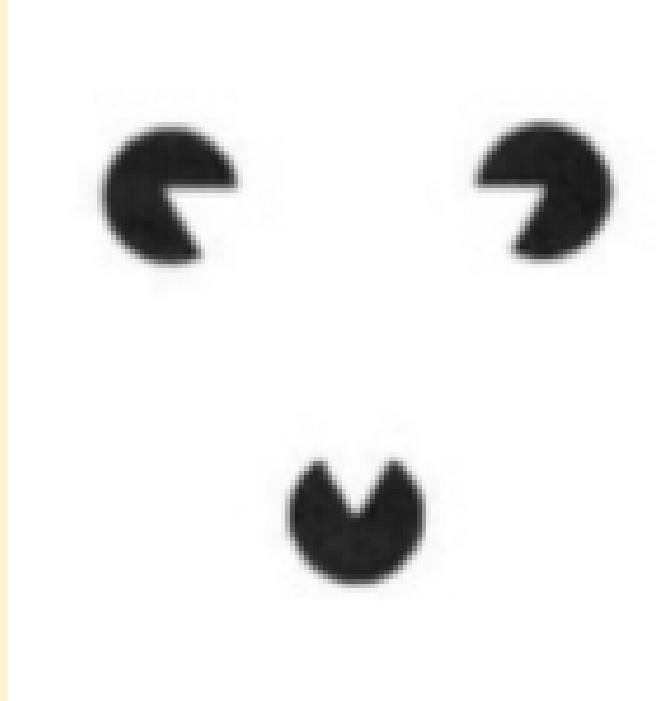
{    } [    ] (    )

# Design principle

- G6.4 Consider using **symmetry** to make **pattern comparisons** easier, but be sure that the patterns to be compared are small in terms of visual angle (<1 degree horizontally and <2 degrees vertically).

Symmetrical relations should be arranged on horizontal or vertical axes unless some framing pattern is used.

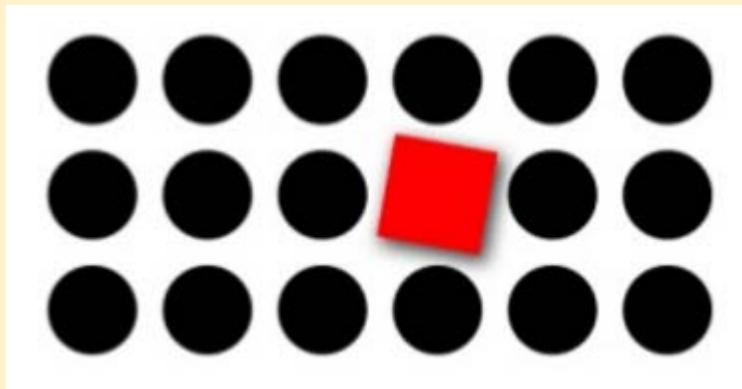
# Closure and common region



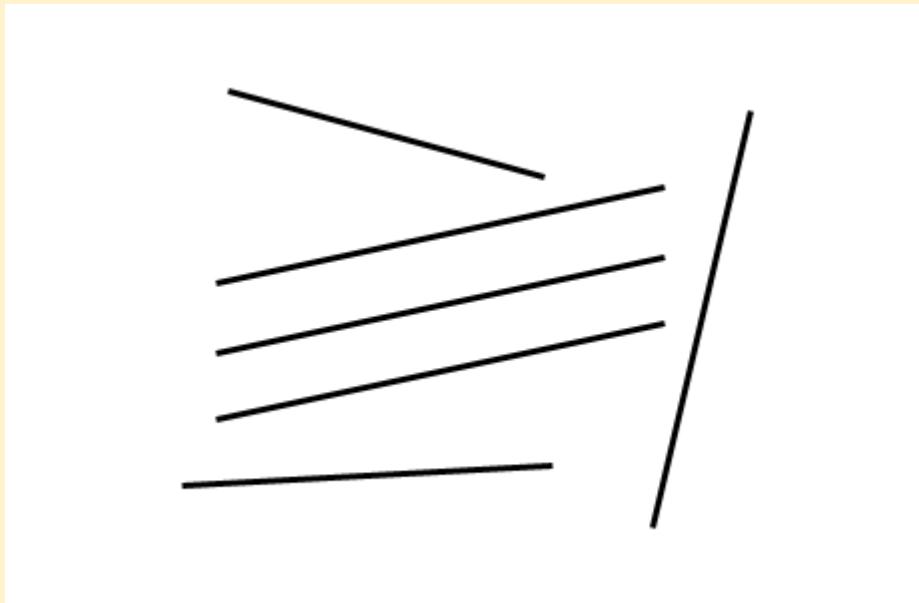
# Design principle

- G6.5 “Consider putting related information inside a closed contour. A line is adequate for regions having a simple shape. Color or texture can be used to define regions that have more complex shapes”.
- G6.6 “To define multiple overlapping regions, consider using a combination of line contour, color, texture, and sweet contours”

# Focal point



# Parallelism



# Design principle

- Combining preattentive processing properties and Gestalt laws we can derive best practices to represent quantity, intensity or to provide visual salience

# Combining to represent quantity

- size:
  - *length or height,*
  - *area* (radio),
  - never *volume*
- lightness, darker = bigger
- hue saturation, saturated = bigger
- vertical position, higher = bigger

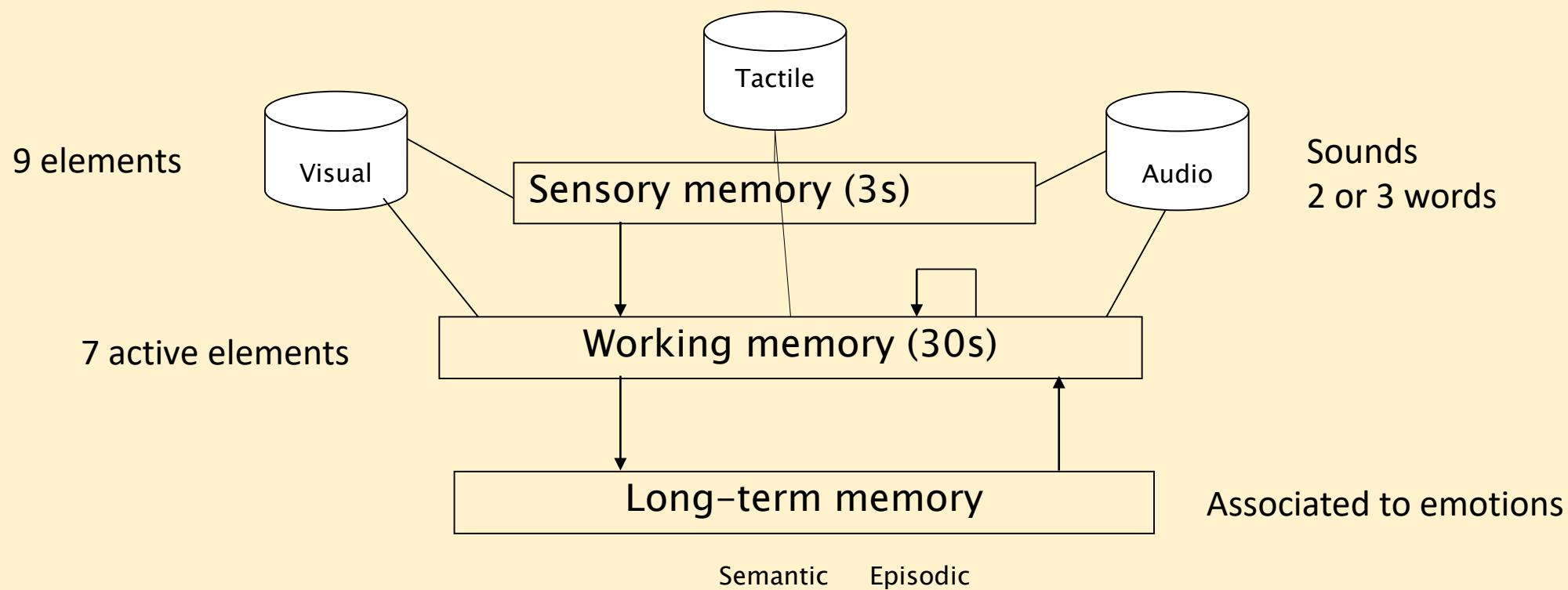
# Combining to represent INTENSITY

- Darker or more saturated
- Bigger
- Thicker

# Combining to represent VISUAL SALIENCE

- Distinct from the norm: in hue, orientation
- Enclosure: by line or background colour
- Added marks

# Memory



# AUTOMATIC teller machine (ATM)

- What do you have to take first: money or card?

WHY?

# Design principles

- Important information should receive focused attention, it shall appear in preeminent locations and have visual salience
- Secondary information may be on secondary locations or hidden, only visible on demand

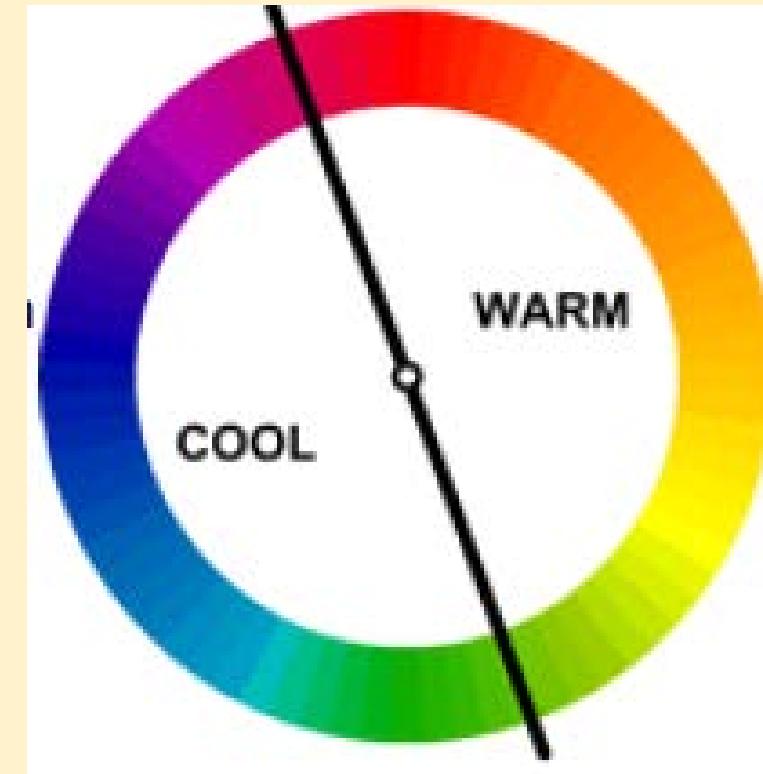
# Specification of a colour: perceptual dimensions HSL

- **Hue** : what we associate to colour names
  - **Saturation (Chroma)**: Purity of the colour (vividness)
  - **Luminance / Lightness / Value** : (it is relative) how much light appears to reflect an object in relation to the White on the scene
- <http://hslpicker.com/>
  - <http://colorizer.org/>

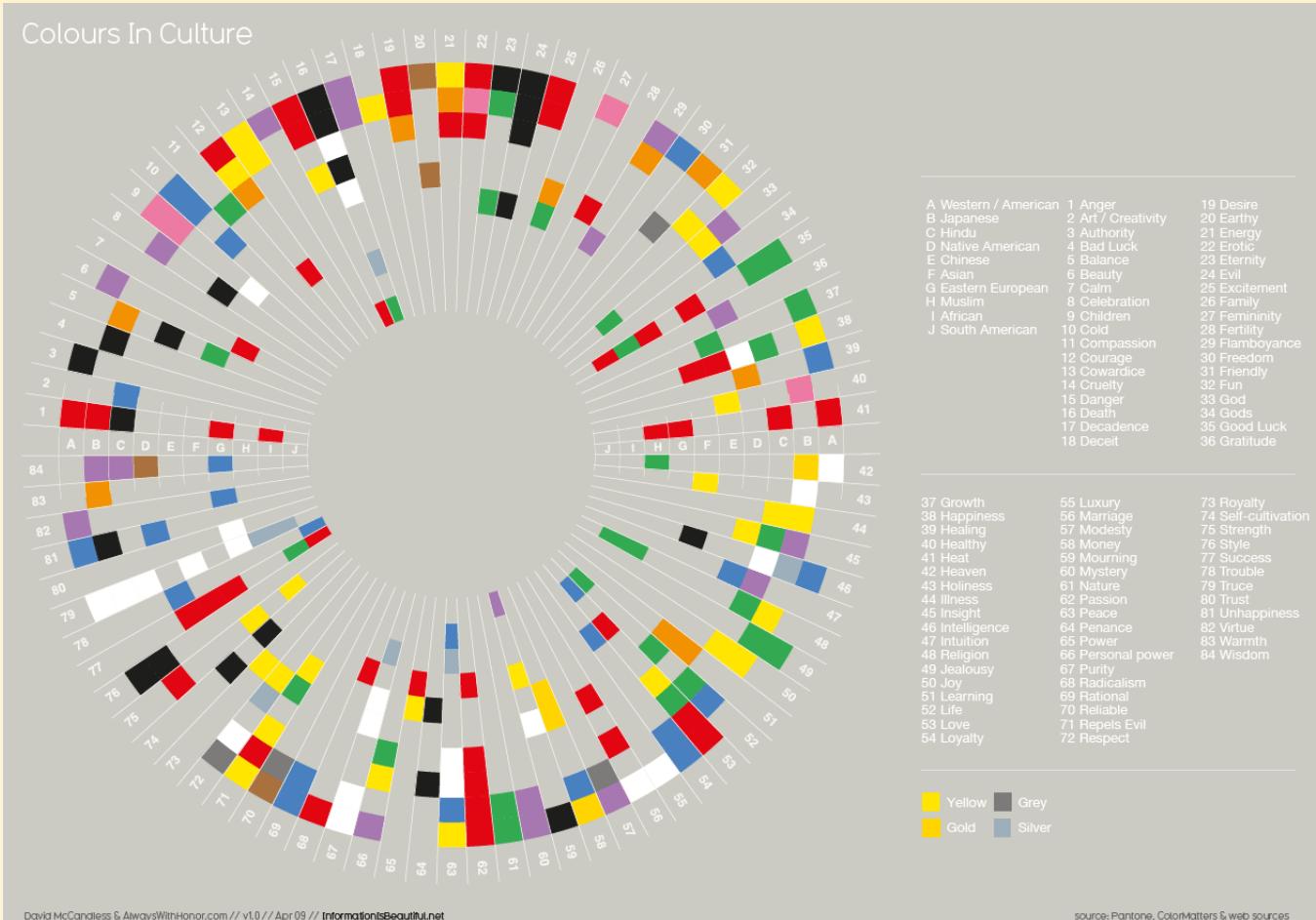


# Colours affect our moods

- Colours **affect us** in numerous ways, both mentally and physically
- Warm colours are **energetic**, and tend to advance in space.
- Cool colours **give calm**, and tend to work better as background.

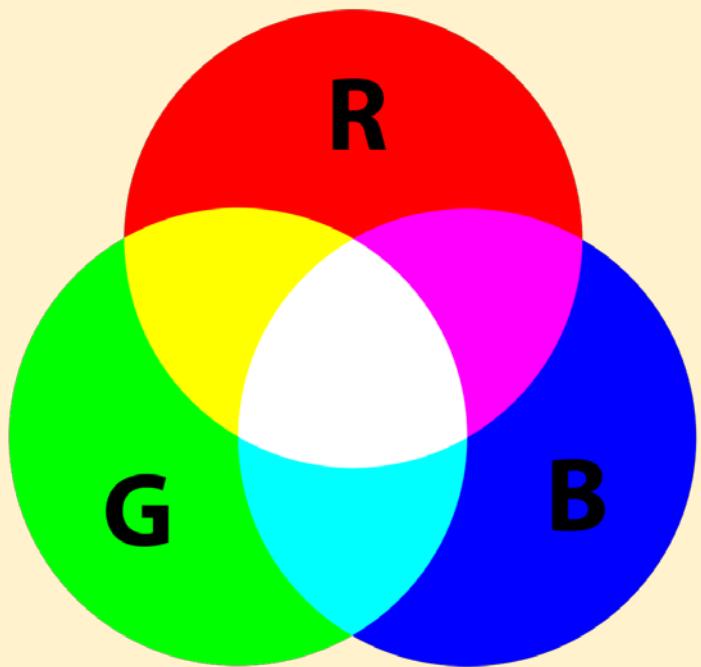


# Colour meaning is a cultural issue

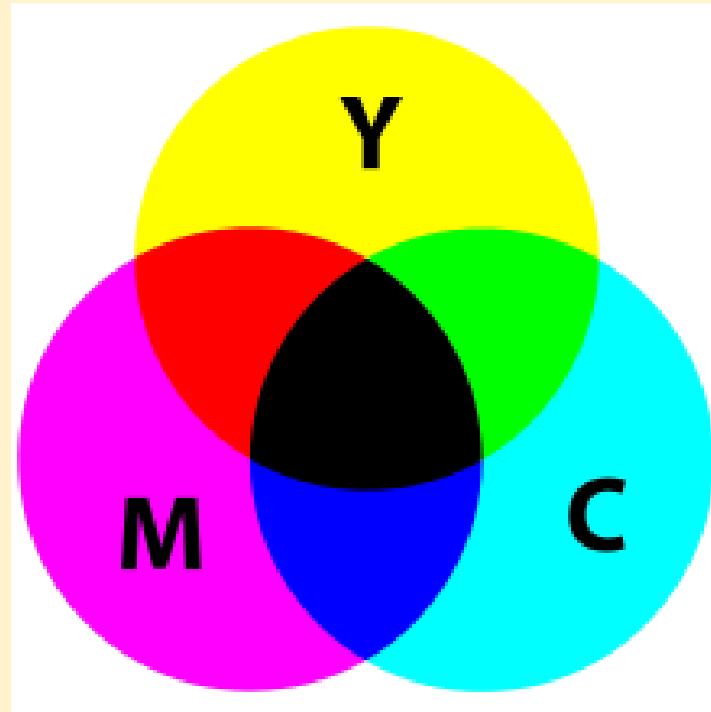


# Specification of a colour

Additive model: screen



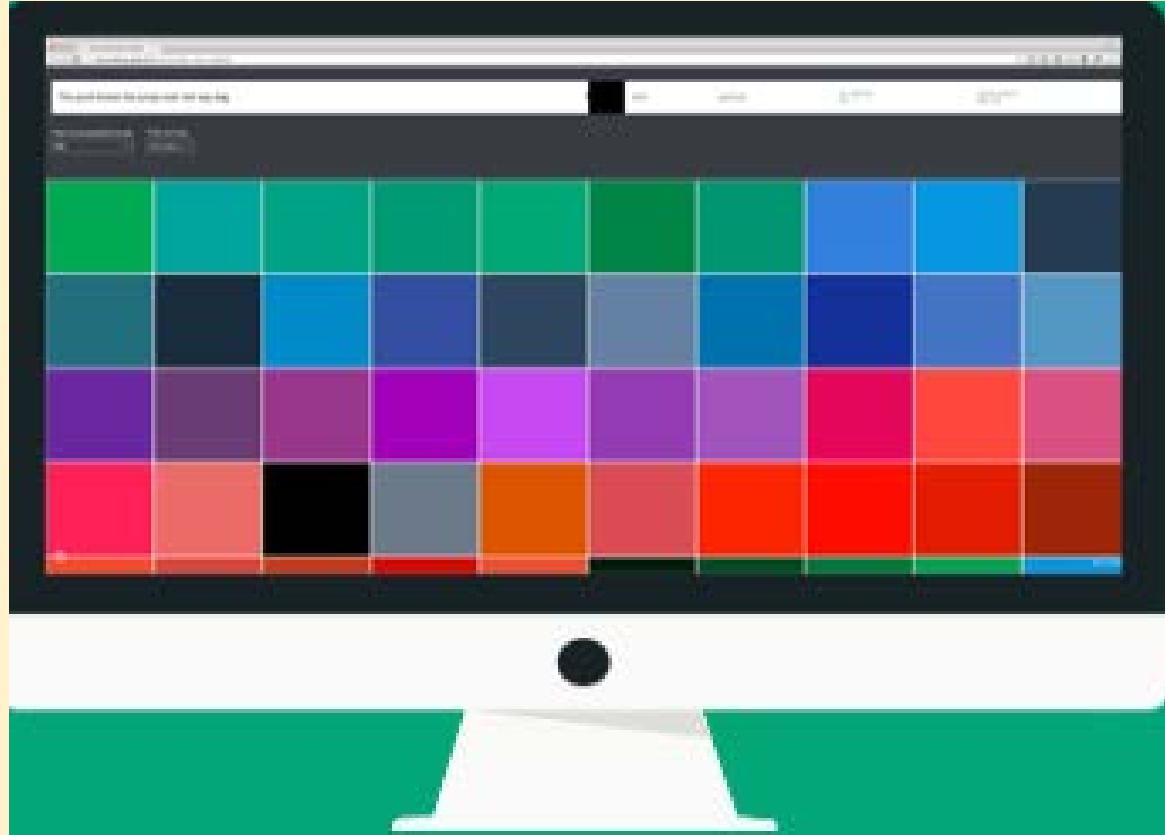
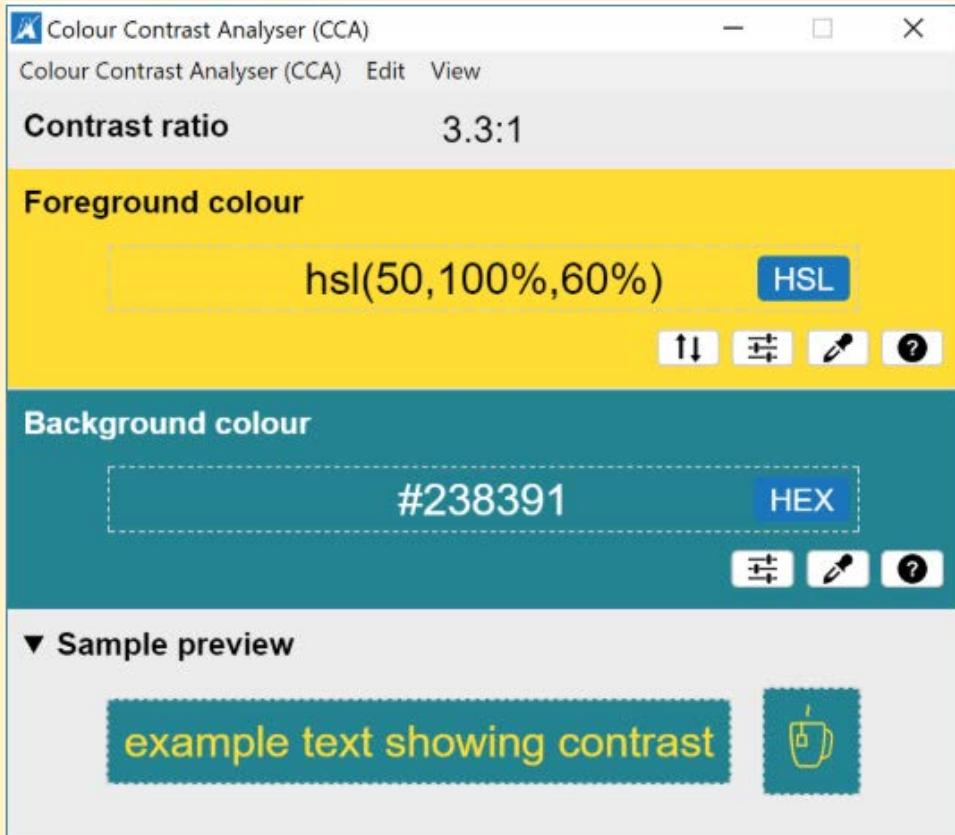
Subtractive model: printing



# Age, colour blindness and contrast

- Colour perception **decreases with age**
- Sight decrease affects both rods and cones
- As we have much more rods, **elders perceive** much better luminance differences than hue differences.
- 
- It is therefore important to **keep contrast differences** in every colour system.

# Contrast: Some tools



LINUX: <https://contrast-ratio.com>

# Colours for charts

- Assign colour according to function:
  - Use contrast to highlight
  - Analogous colours to group
  - Use greys for context and axis when labelling with colour

# Colour principles (1)

- G4.7 If using colour saturation to encode numerical quantity, use **greater saturation** to represent **greater numerical quantities**. Avoid using a saturation sequence to encode more than three values.

# Colour principles (2)

- G4.16 Use low-saturation colours to colour code large areas.  
Generally, light colours will be best because there is more room in colour space in the high-lightness region than in the low-lightness region.

# Colour principles (3)

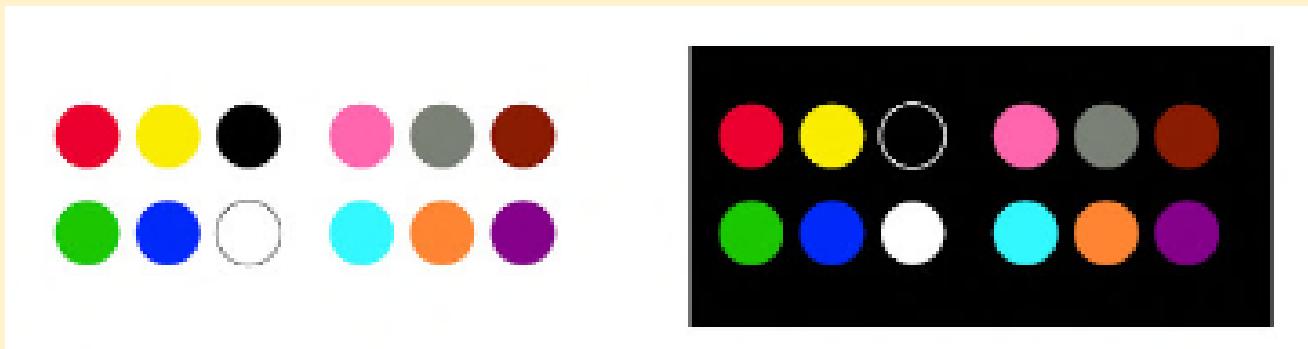
- G4.17 When colour coding large background areas overlaid with small coloured symbols, consider using all low-saturation, high-value (pastel) colours for the background, together with high-saturation symbols on the foreground.

# Colour principles (4)

- G4.18 When **highlighting text** by changing the colour of the font, it is important to **maintain luminance contrast** with the background.

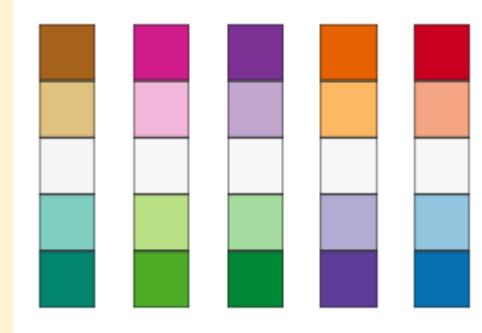
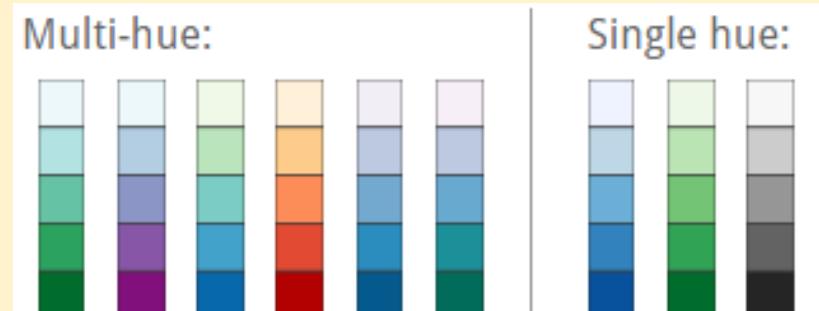
# Colours for labelling. Qualitative

- Small set: red, green, yellow, blue (opponent theory)
- 12 cross-cultural safe colours (see figure)
- Different hues have no order

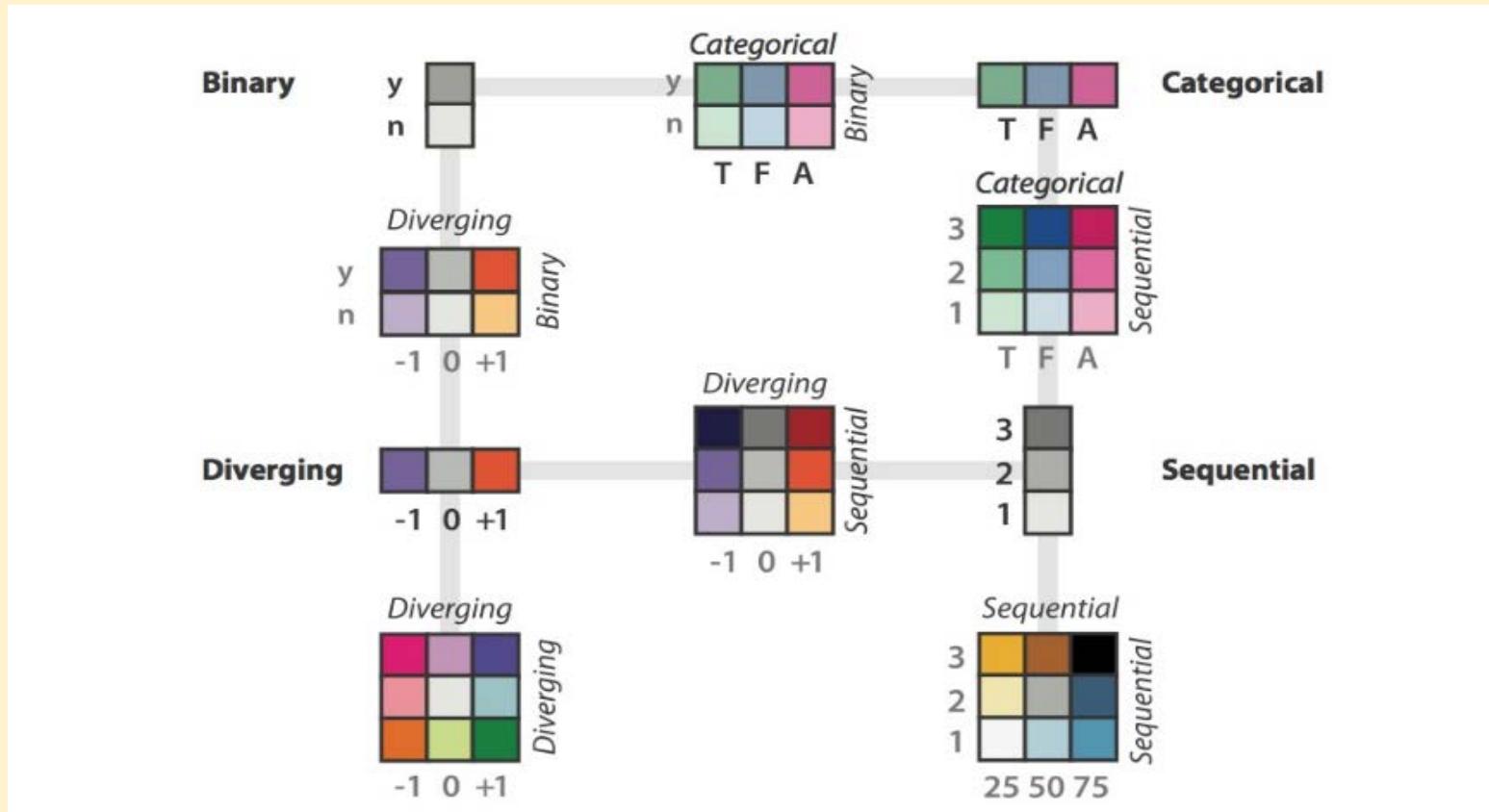


# Colours for labelling. Quantitative Scales

- Sequential: each step differs in saturation or in saturation and lightness
- Diverging: two hues, a neutral hue in the middle. Suited for opposite values



# Colour mapping, combinations



# Colours for text

- For highlighting, contrast is the most relevant issue

# Colours for maps

- Big areas: low saturation;
- Small areas: highly saturated
- Ensure hue and luminance contrast with the background (use a border if needed)
- For colour-blindness assure yellow-blue distinction
- See Cynthia Brewer [ColorBrewer tool](#)

# Choosing the right chart

Visual vocabulary

Charts handout

# Which chart?

## Visual vocabulary

#### Dealing with data

There are so many ways to visualize data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important to your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualizations.

