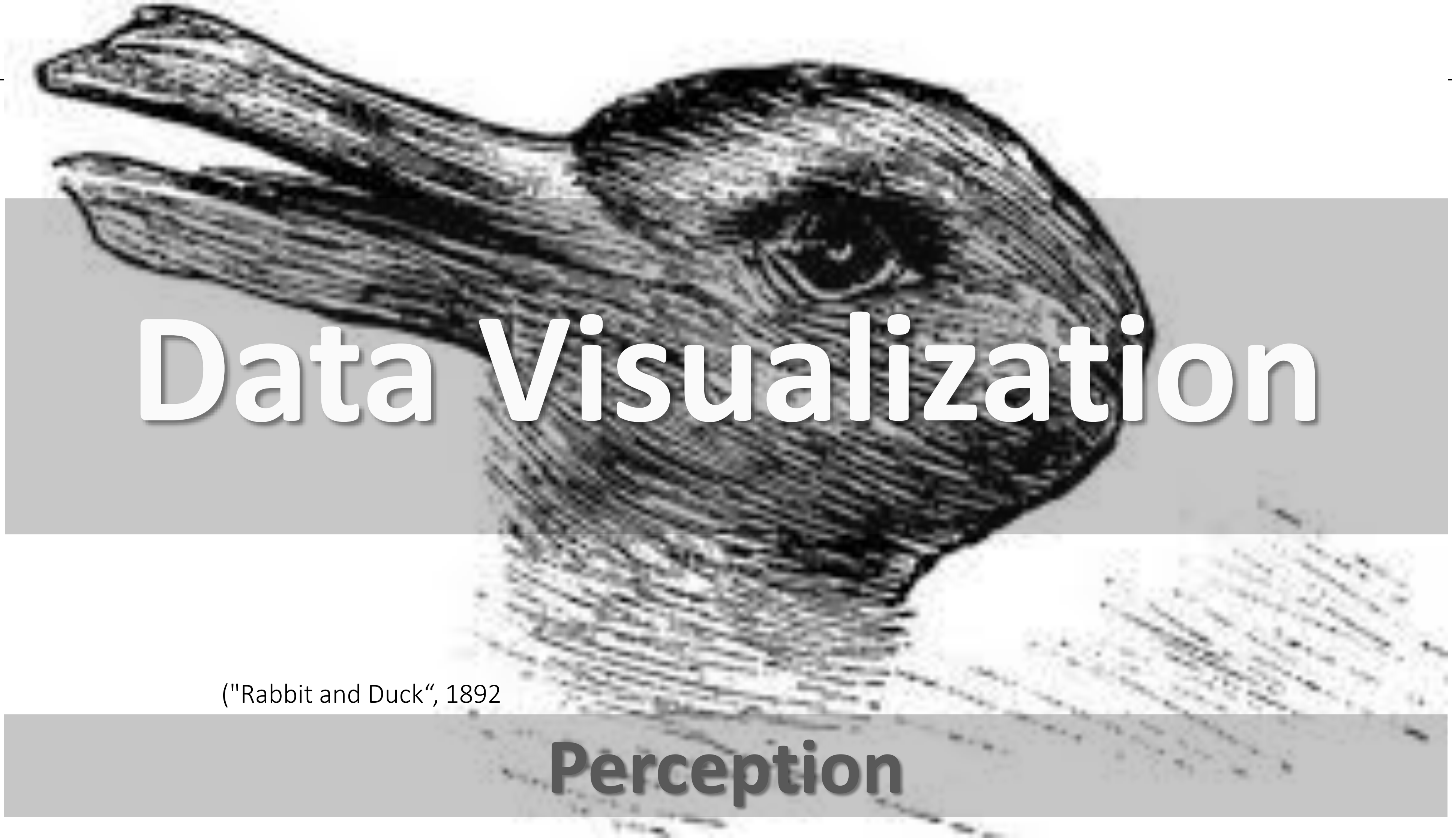
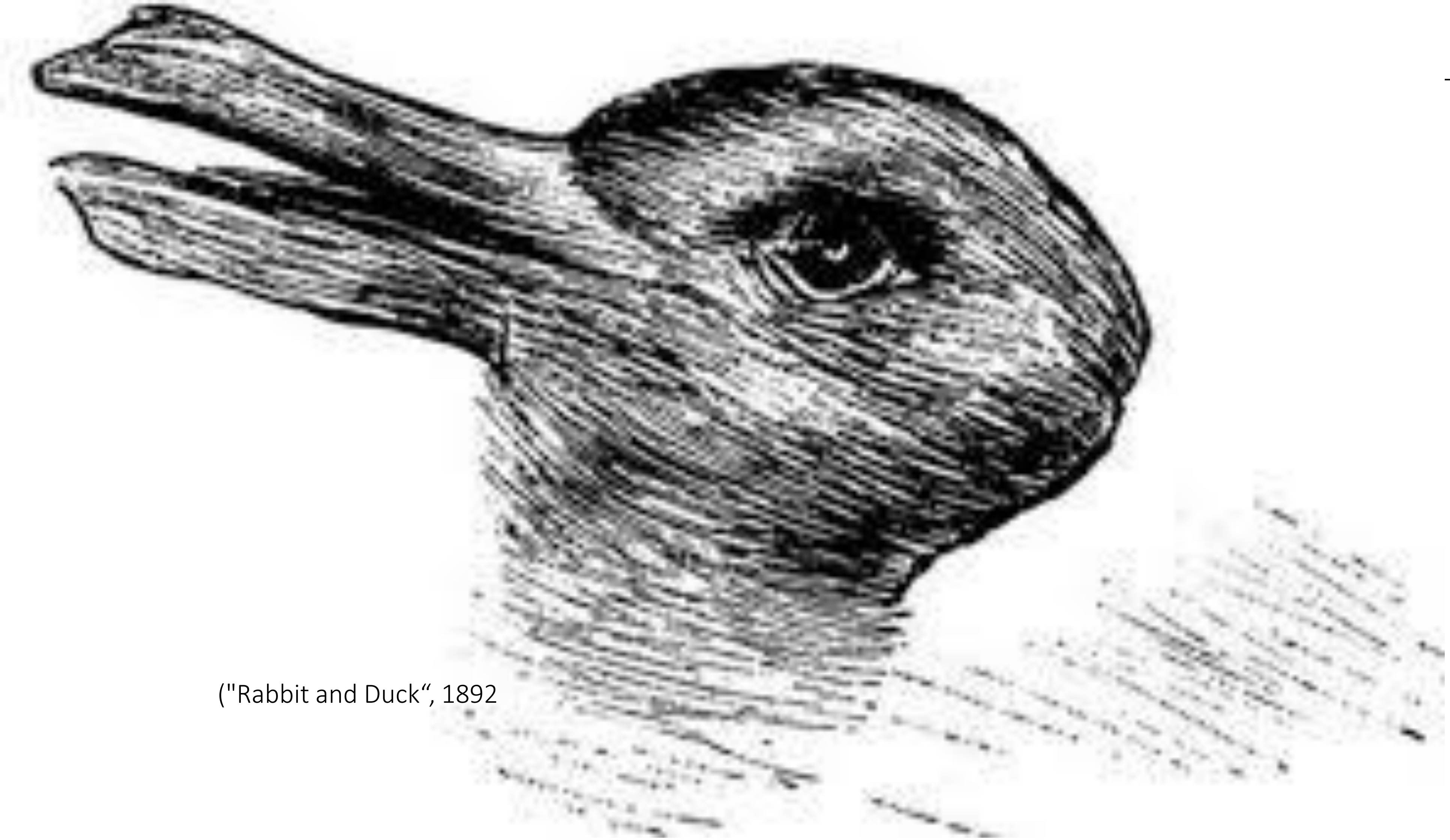


Data Visualization

("Rabbit and Duck“, 1892

Perception

M126 | Maria Roussou



("Rabbit and Duck“, 1892

M126 | Maria Roussou

Hans Rosling - Let my dataset change your mindset - TED@State 2009

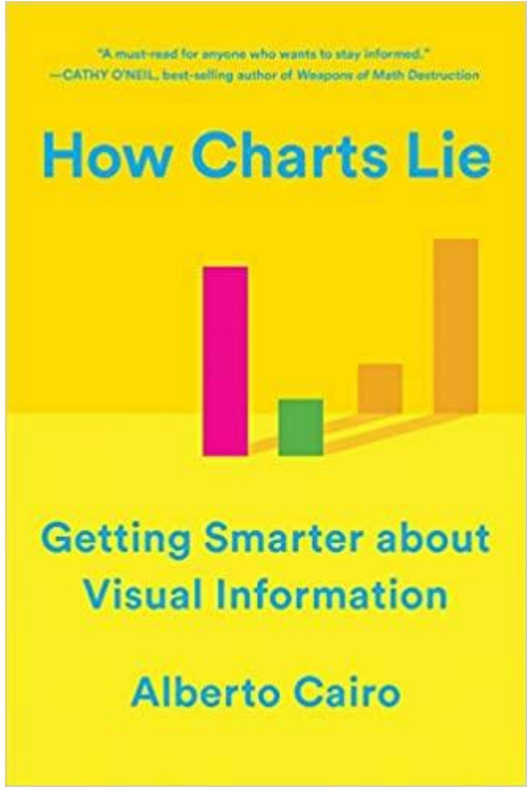
[Let my dataset change your mindset - TED@State 2009](#)

William Allen: The politics of big data, migration, and mobility

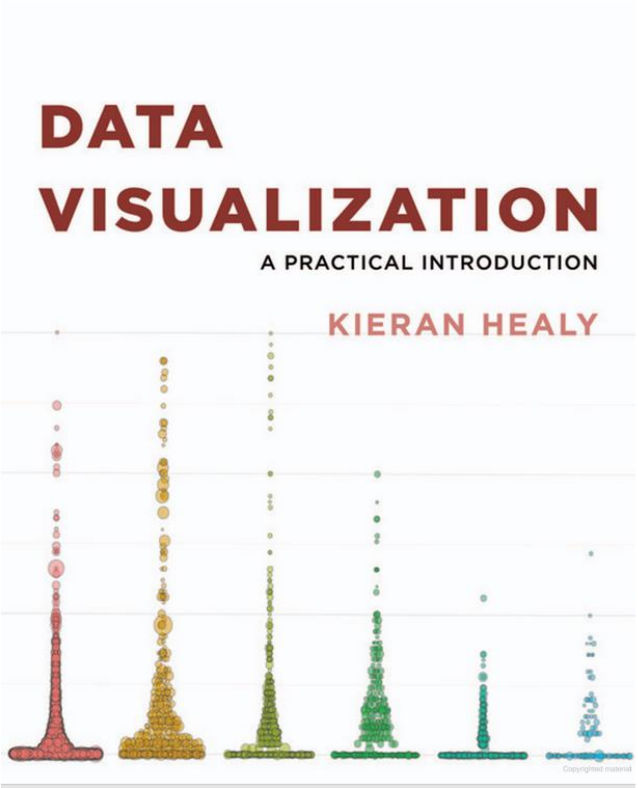
- W. Allen, “The politics of big data, migration, and mobility,” presented at the Ανοικτές Διαλέξεις/Συζητήσεις: Μεγάλα Δεδομένα, Νέα Μέσα, Ζητήματα Τεκμηρίωσης: Μαθαίνοντας από πρωτοπόρα εγχειρήματα, Εθνικό Κέντρο Τεκμηρίωσης (EKT), 08-May-2019.

How charts lie

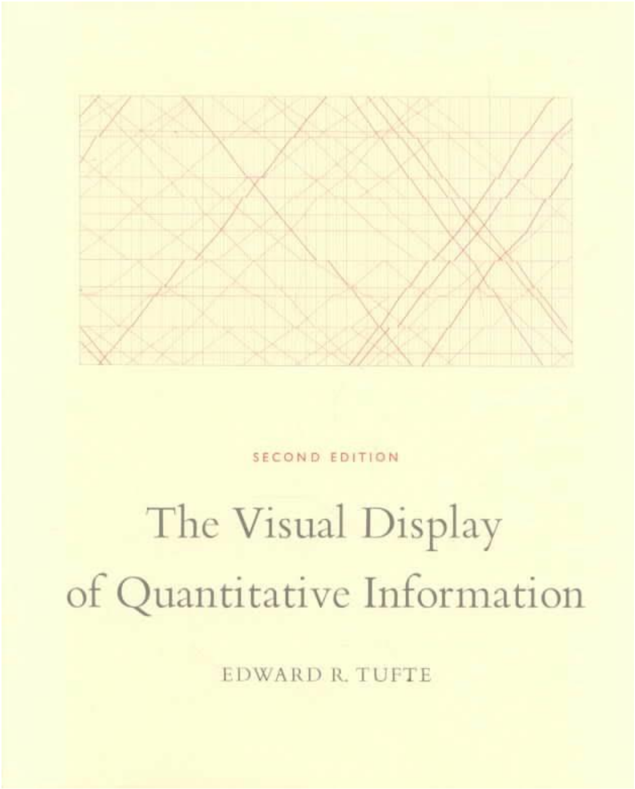
- More about how charts can be misleading



2020



2019



1983

What makes charts “lie”?

- Charts that “lie” by being poorly designed
- Charts that “lie” by displaying dubious data
- Charts that “lie” by displaying insufficient data
- Charts that “lie” by concealing or confusing uncertainty
- Charts that “lie” by suggesting misleading patterns

Cairo, 2020

What makes bad figures “bad”?

- “Bad” aesthetics
  - needless frills (3D, poor/unnecessary colors, chartjunk)
- “Bad” use of data
  - basically cooking the data to make it look how you want
- “Bad” perception
  - encoding between data and visual properties mislead us

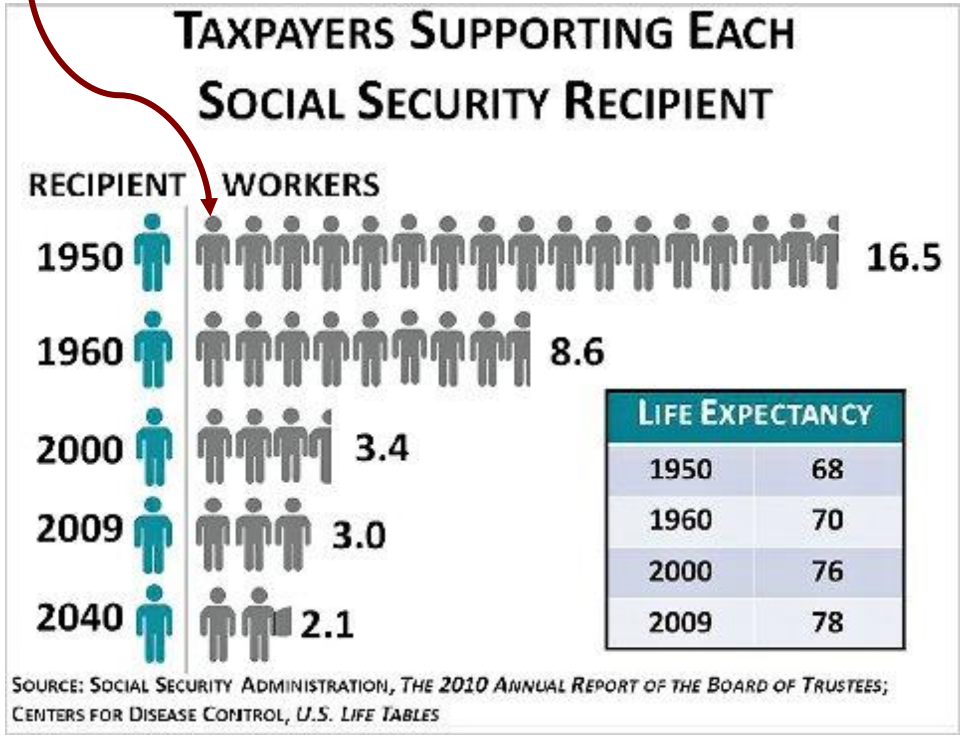
Healy, 2019

Bad aesthetics

- **Chartjunk**: “ornamental and often saccharine design flourishes that impede understanding”

Tufte, 1983

The dot-headed figures to convey quantities clutter the chart without adding information



Bad aesthetics

- Is this a good figure?

- What's good? What's bad?

Bad aesthetics

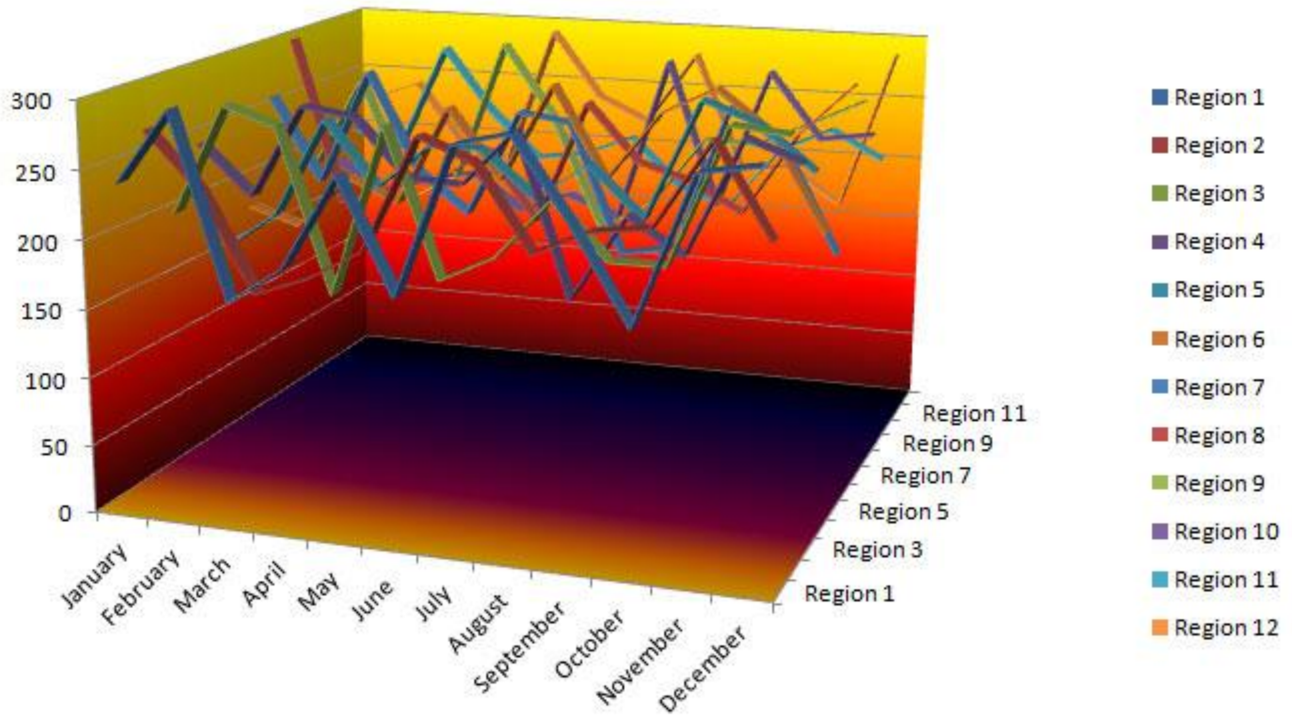
- Is this a good figure?

`Monstrous Costs' by Nigel Holmes

- What's good? What's bad?

Bad aesthetics

- Is this a good figure?



- What's good? What's bad?

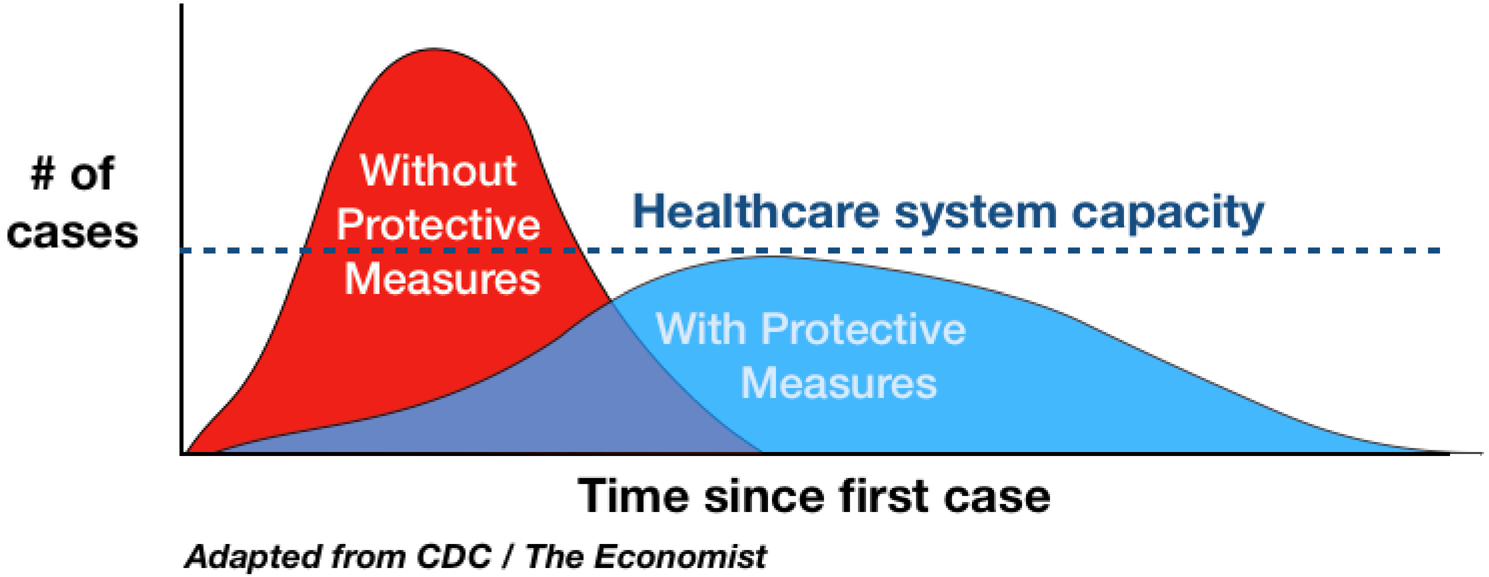
Bad aesthetics

- Is this a good figure?

- What's good? What's bad?

Bad aesthetics

- Is this a good figure?



- What’s good? What’s bad?

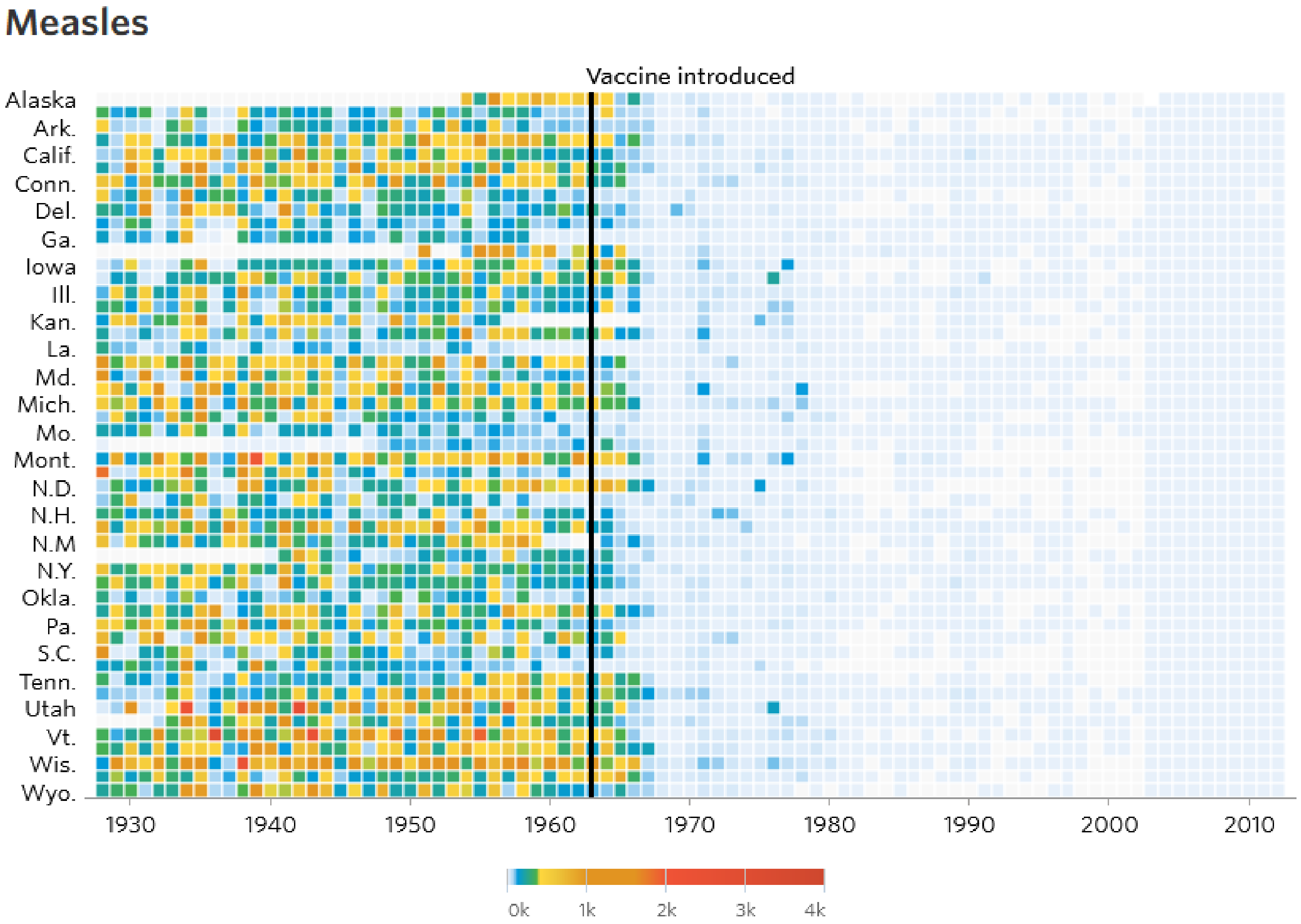
Bad aesthetics

"Above all else, show the data"

Edward Tufte

Data Visualization Aesthetics

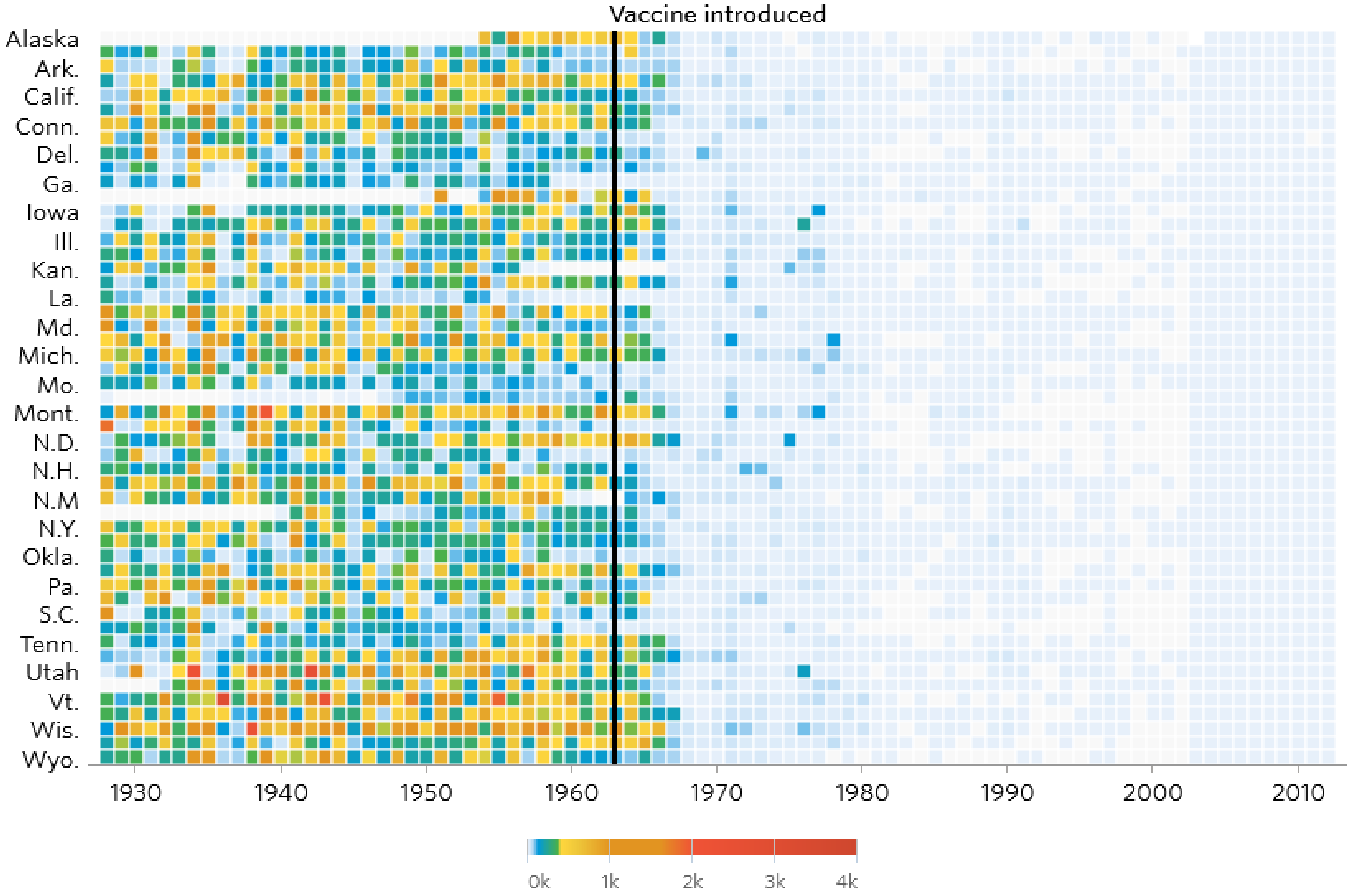
- no chart junk, a clear message



Wall Street Journal, 2015

Data Visu

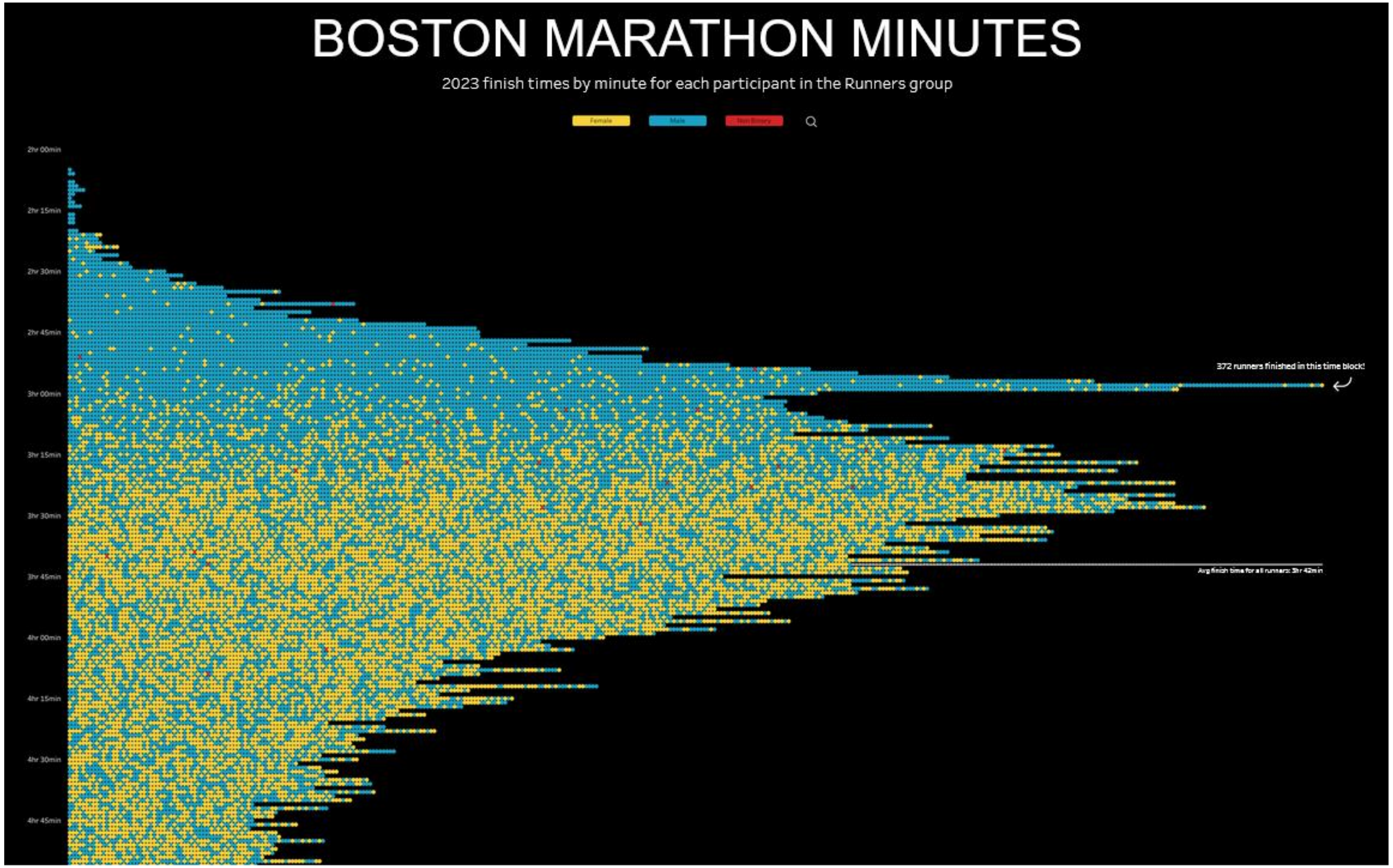
- no cha



Wall Street Journal, 2015

Data Visualization Aesthetics

- Simple, clear message



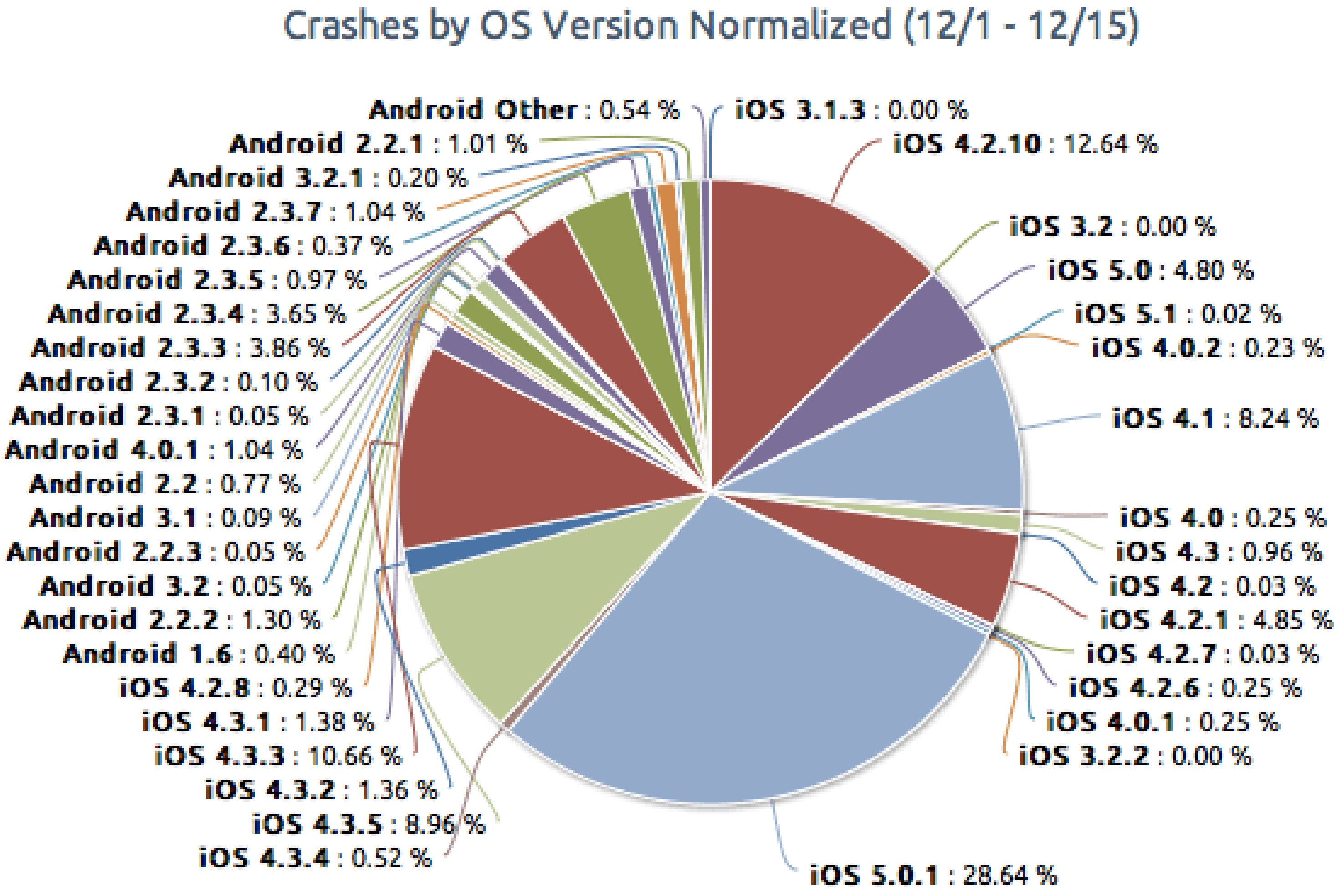
https://public.tableau.com/app/profile/whitney6892/viz/bostonrunners/Dashboard1

“Bad” use of data

- Charts that “lie” by displaying dubious data
- Charts that “lie” by displaying insufficient data
- Charts that “lie” by concealing or confusing uncertainty
- Charts that “lie” by suggesting misleading patterns

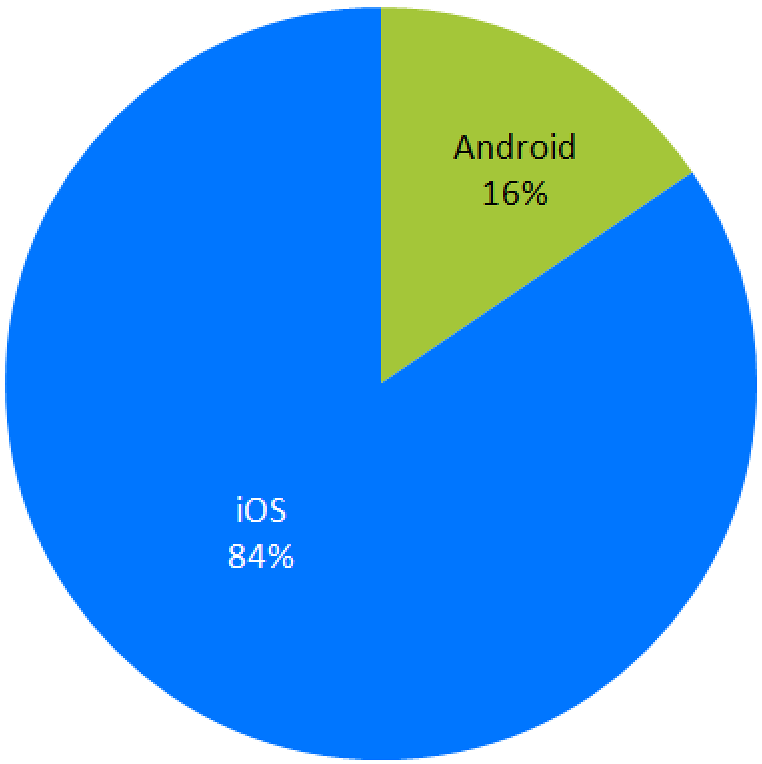
Bad use of data

- Chart junk



Bad use of data

- No chart junk

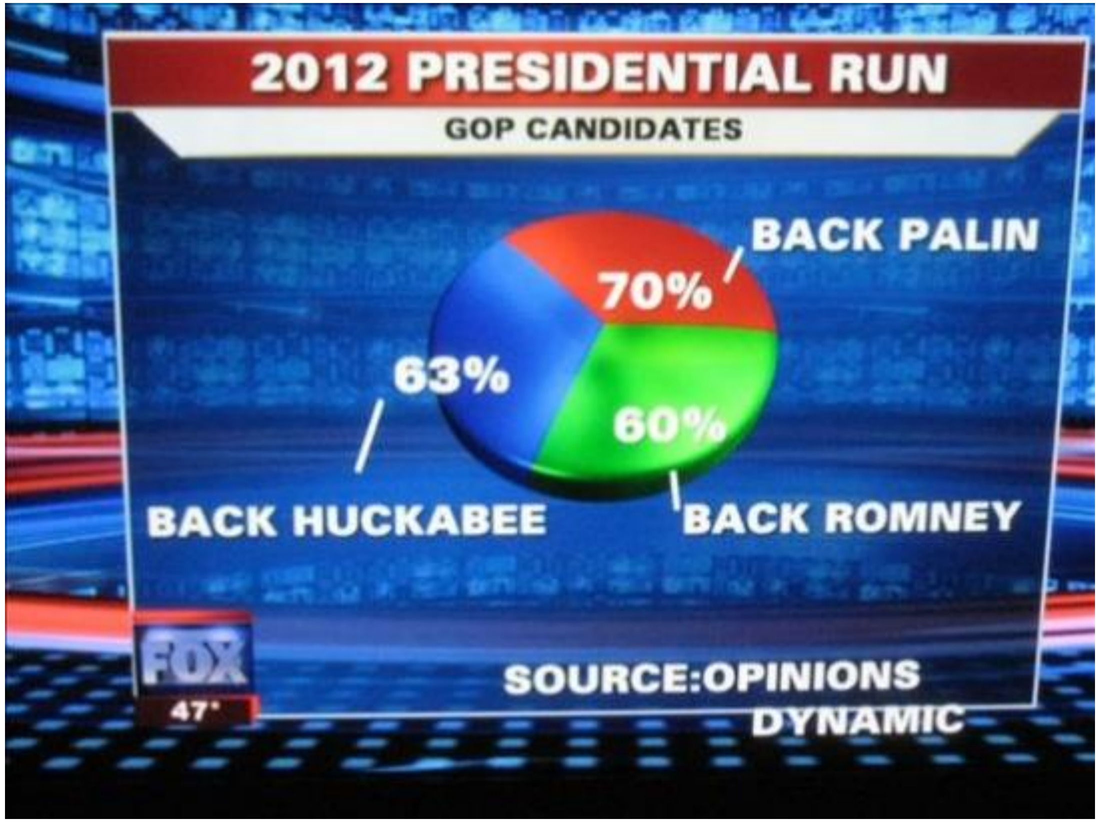


Do iOS Apps Crash More Than Android Apps?

https://www.forbes.com/sites/tomiogeron/2012/02/02/does-ios-crash-more-than-android-a-data-dive/

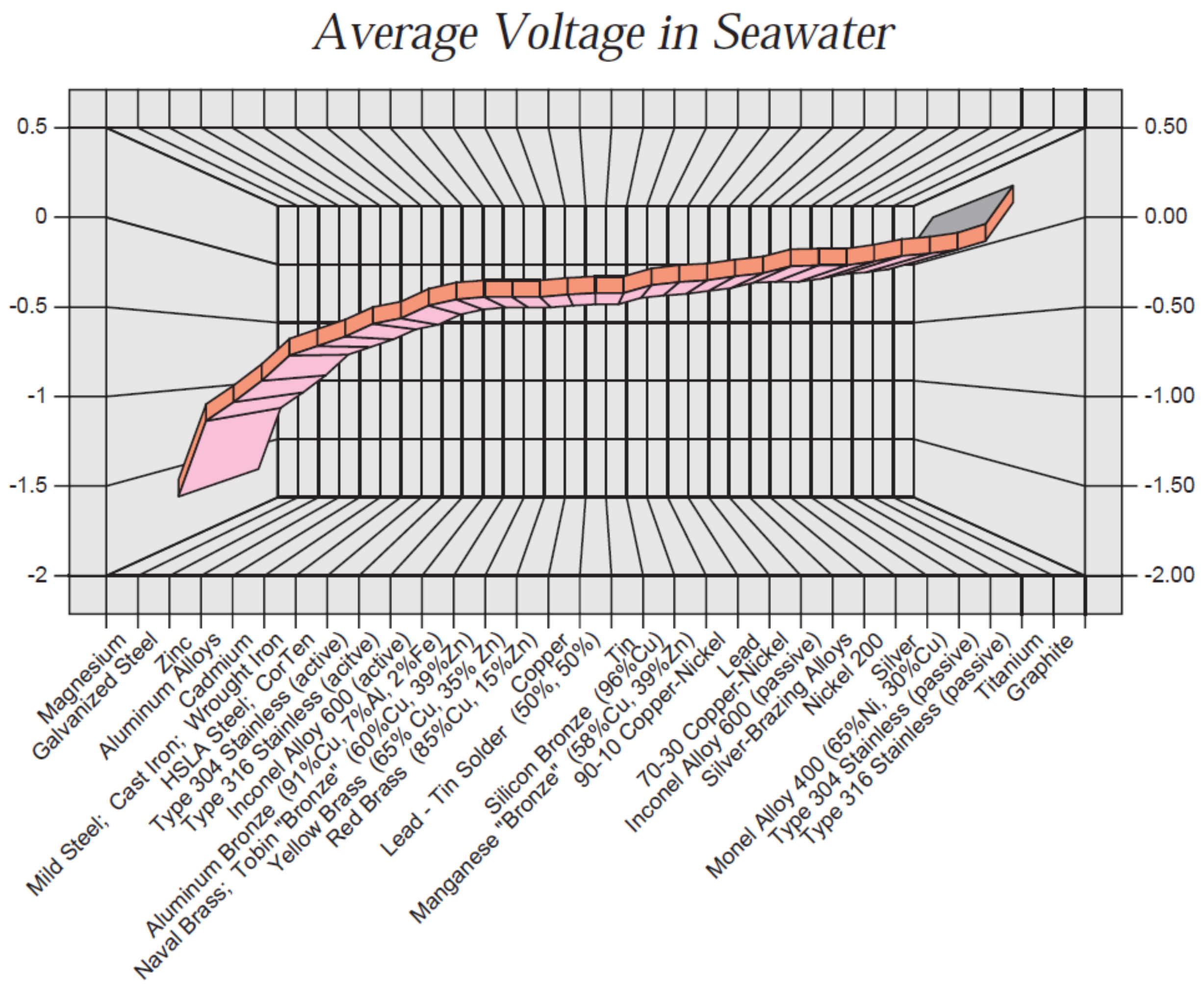
Bad use of data

The “Best Pie Chart Ever”!



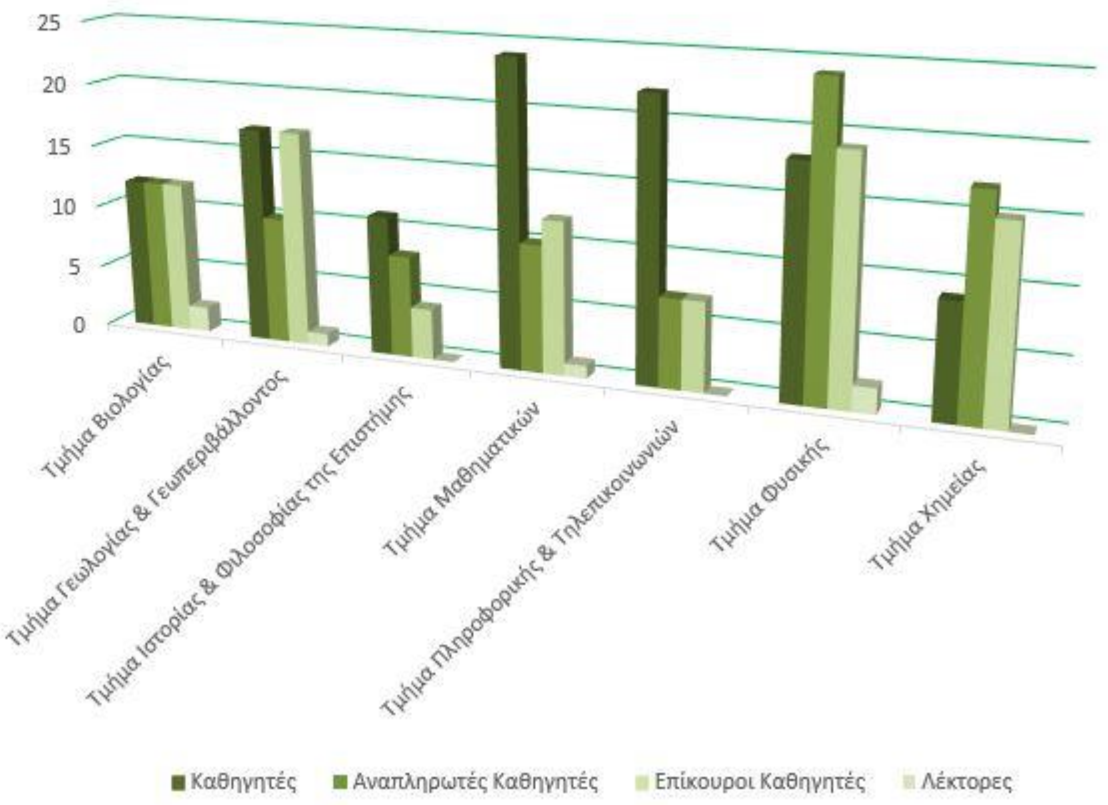
Bad use of data

- horrific use of 3D
- undesirable optical effect of the gridlines
- incorrect display of the data as a series of connected boxes
- item labels misaligned with the tick marks, and taking up far too much of the graphic display area



Bad use of data

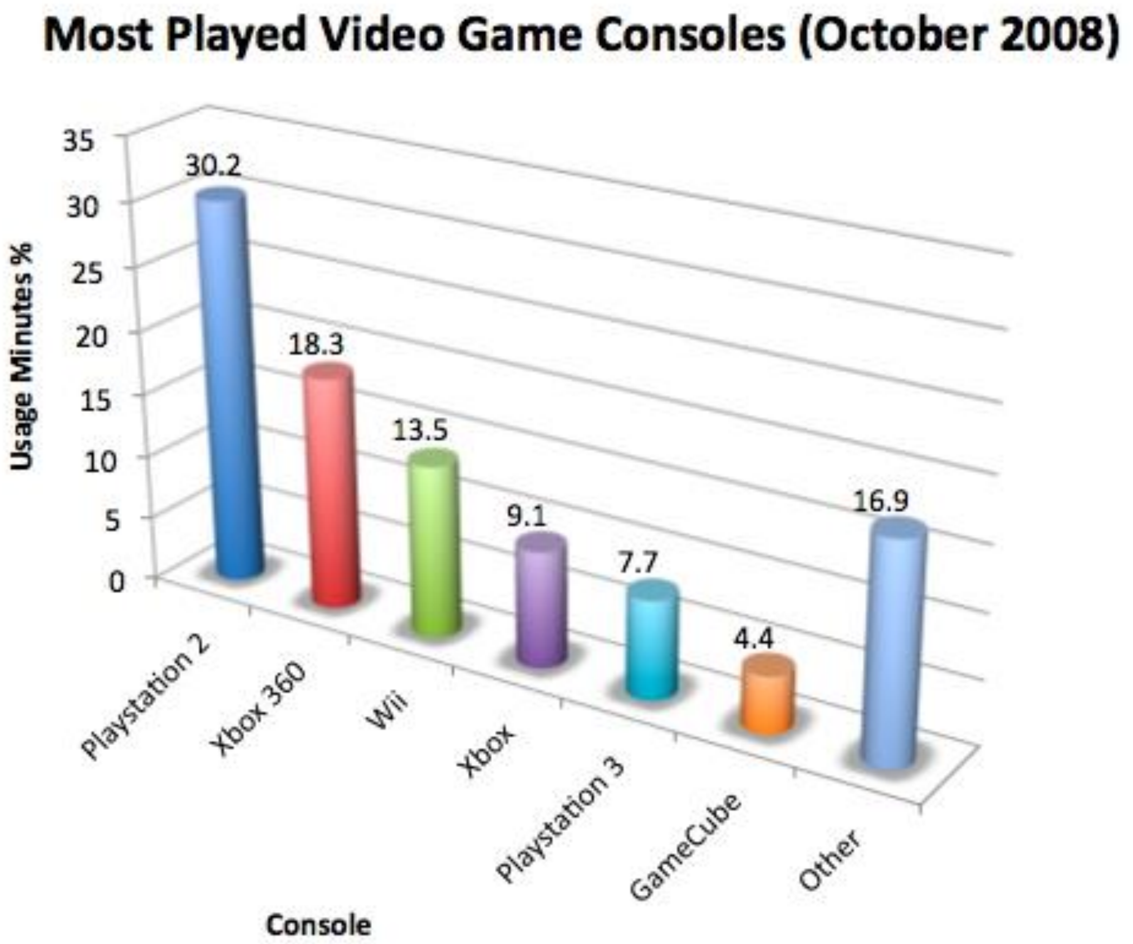
- 3D distorts the data, adds unnecessary details



UoA School of Science, 2020

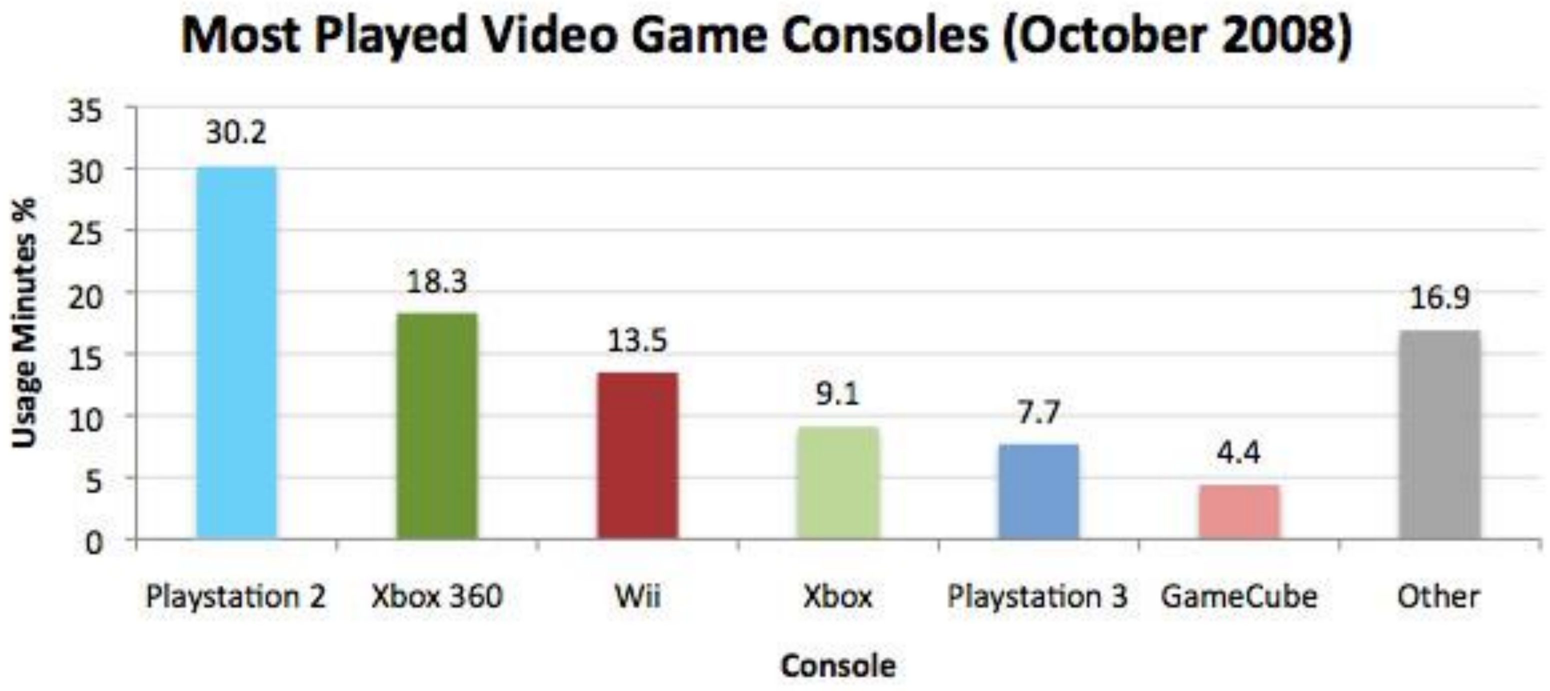
Bad use of data

- 3D distorts the data, adds unnecessary details



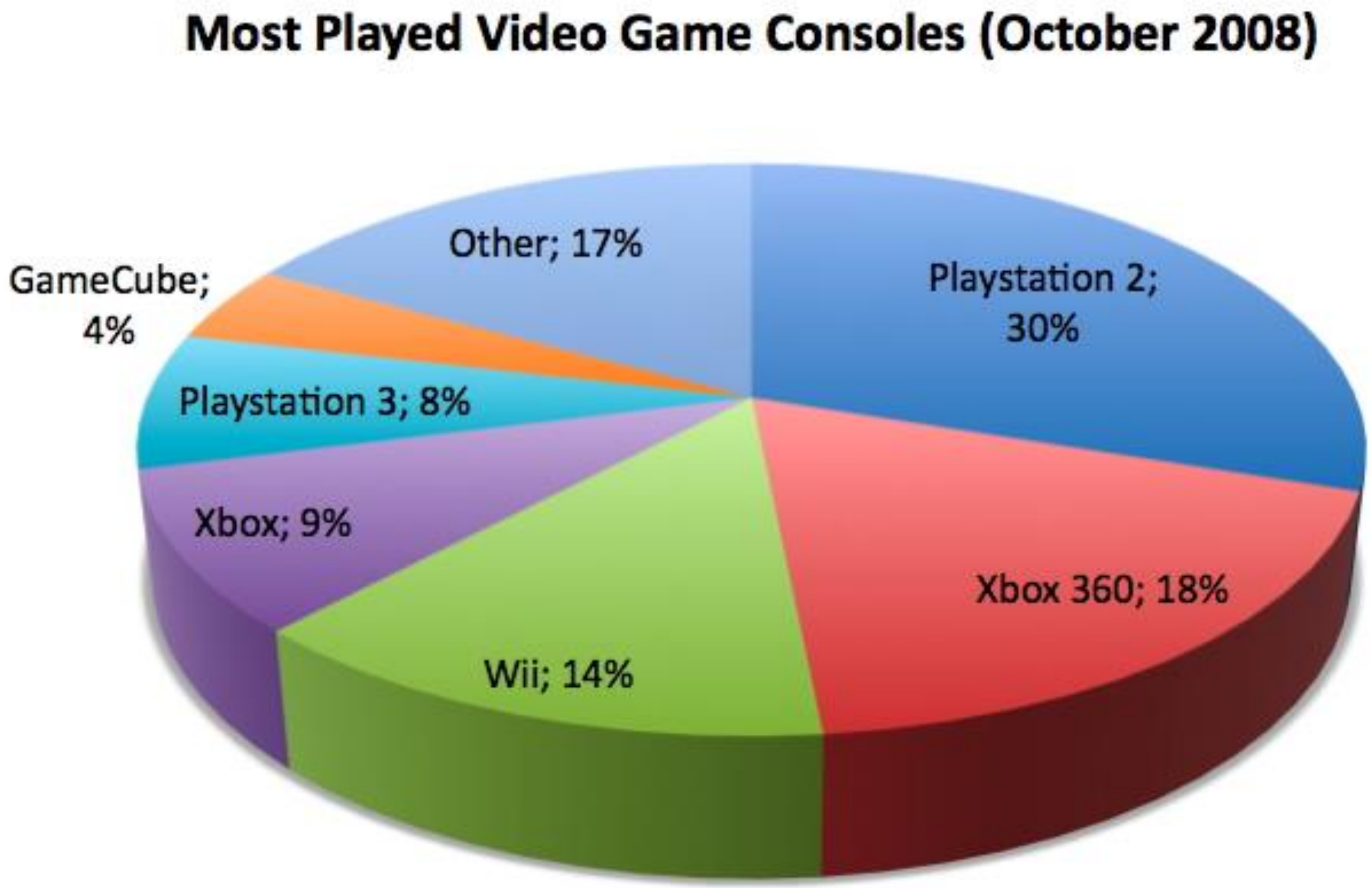
(Bad) use of data

- “boring” but readable



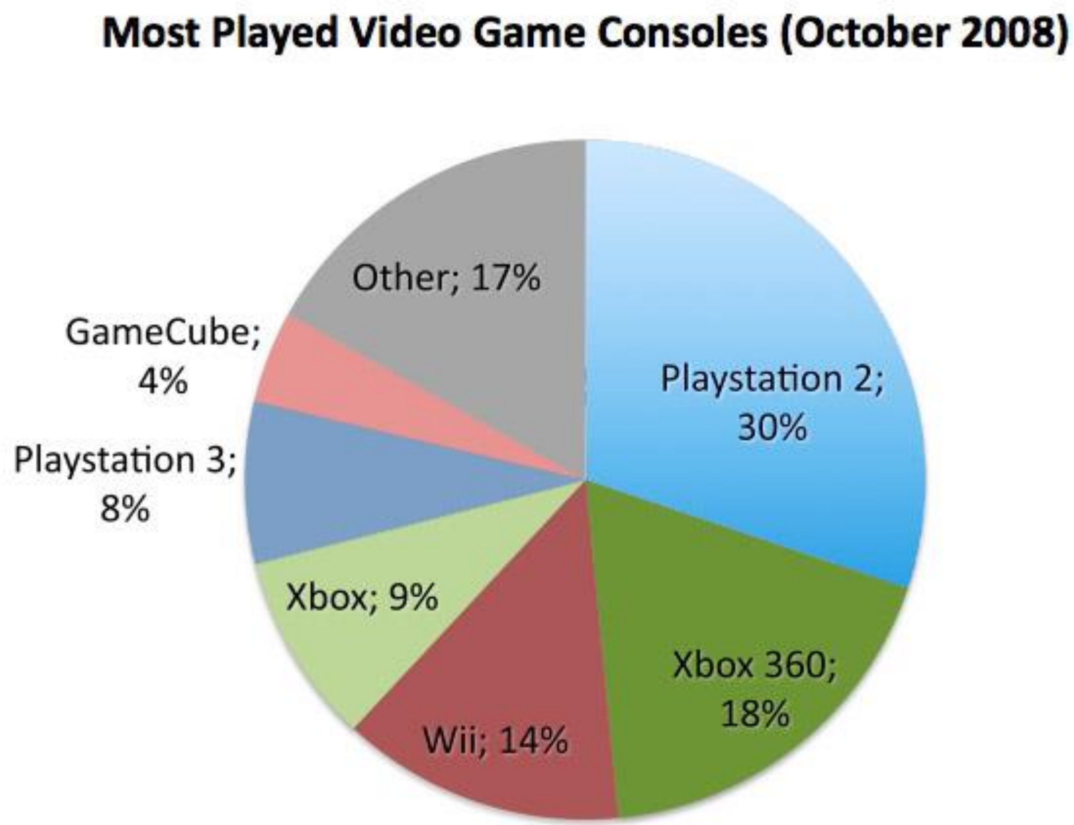
Bad use of data

- 3D distorts the data, adds unnecessary details



(Bad) use of data

- “boring” but readable



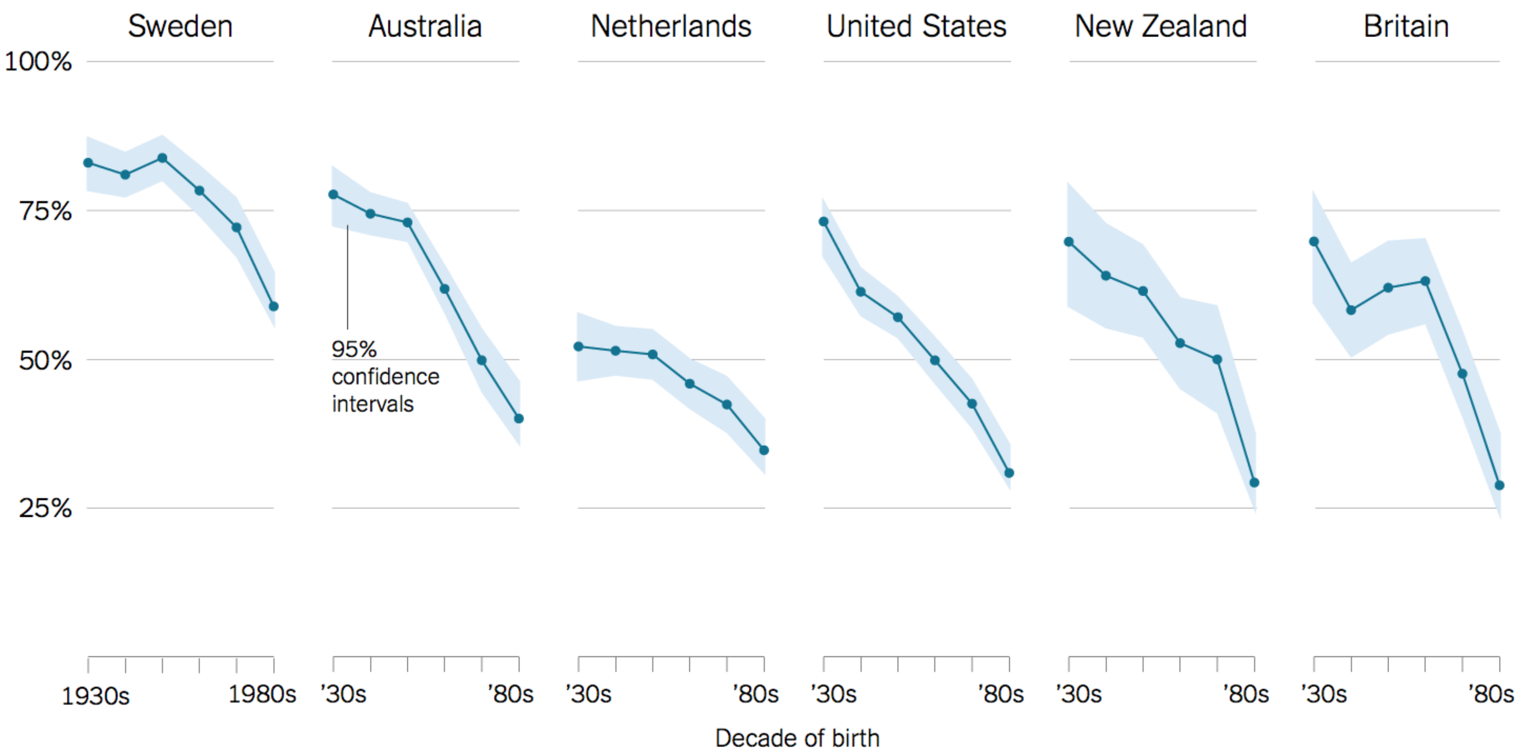
- The bar chart makes it easy to see **how they compare to each other**.
- The pie chart makes it easier to see **how each console compares to the whole**

- A bar chart makes it easier to estimate the **actual amount** compared to a pie chart if you don't have the actual values displayed.
- **Bar charts** are **better the more categories you have** as the slices of the pie get smaller and harder to discern with more and more categories.
- In an analysis tool you may need both views simultaneously, and then additional visualizations to see the values over time.

Bad use of data

- What's wrong with this graph?

Percentage of people who say it is "essential" to live in a democracy

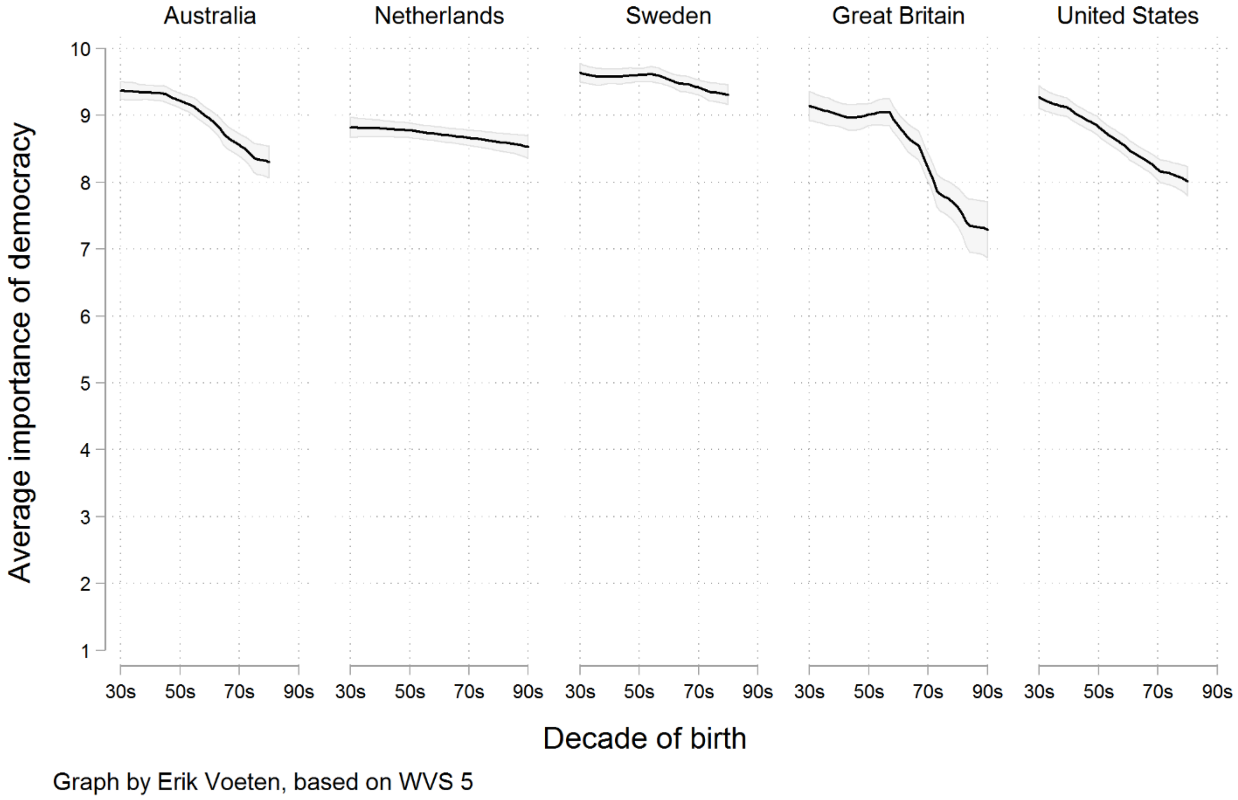


Source: Yascha Mounk and Roberto Stefan Foa, "The Signs of Democratic Deconsolidation," Journal of Democracy | By The New York Times

Foa, Roberto Stefan, and Yascha Mounk. "The Signs of Deconsolidation." *Journal of Democracy* 28, no. 1 (2017): 5–16.

Bad use of data

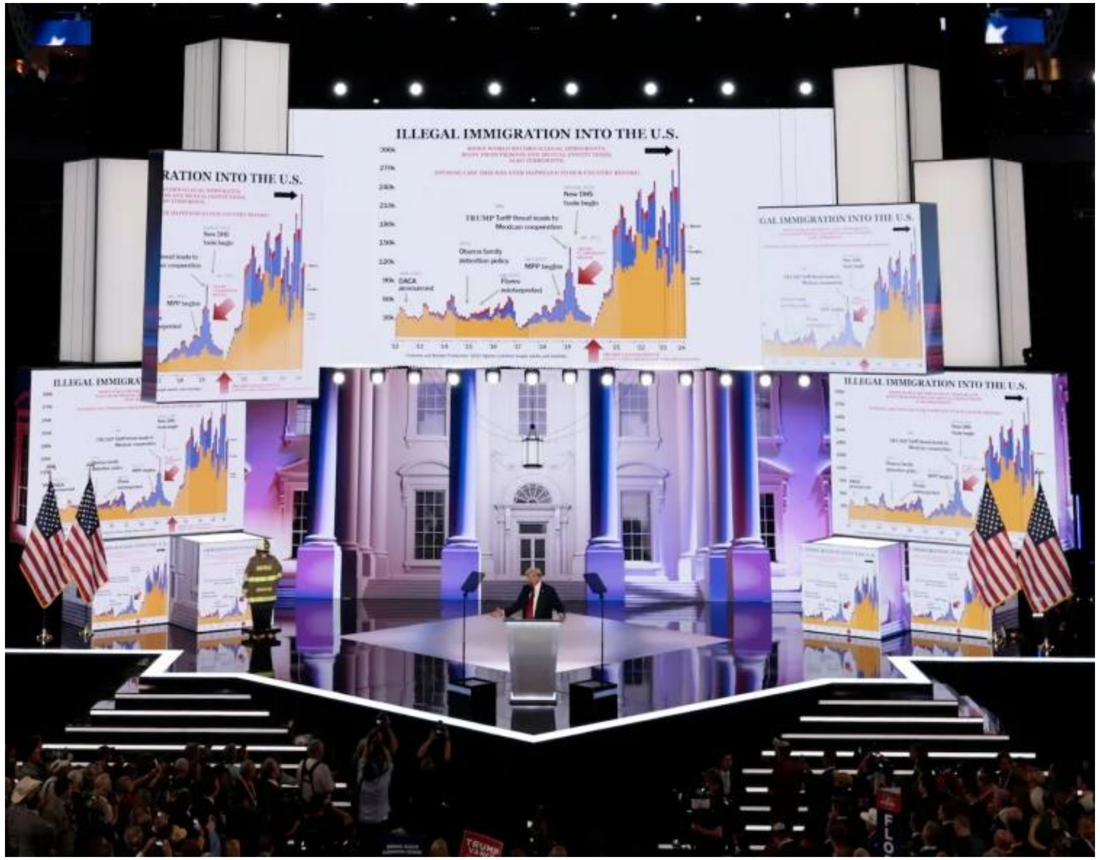
- redrawing based using the average response

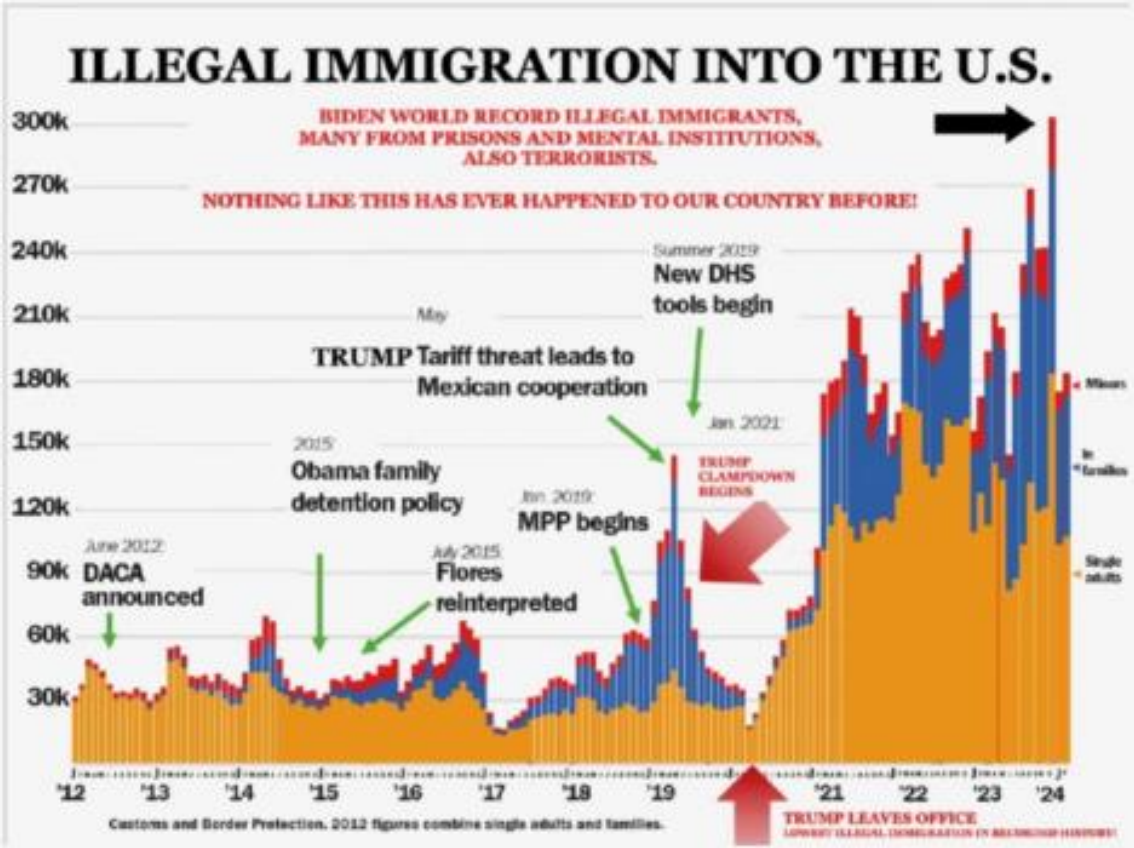


Erik Voeten

Bad use of data

- How Trump uses a deceptive chart to lie about the border



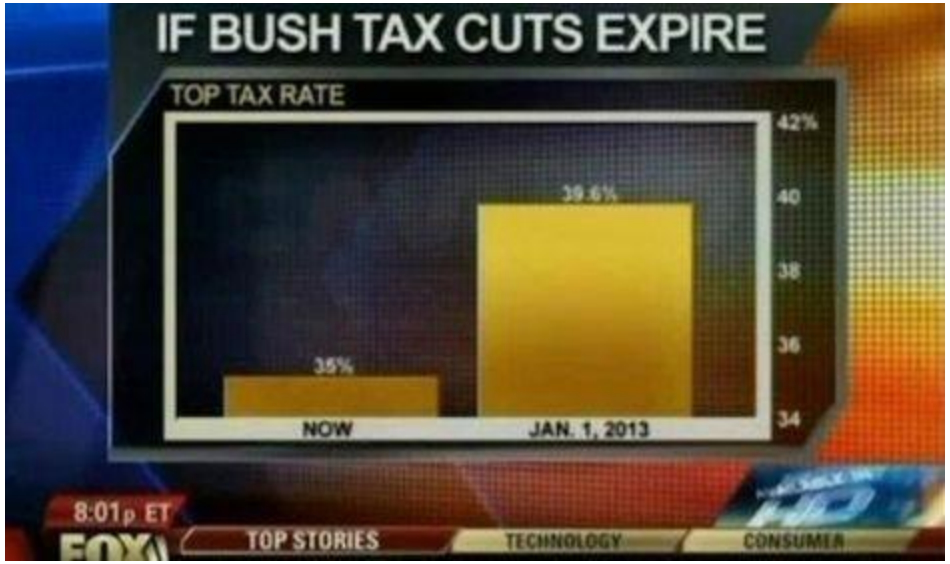


https://www.factcheck.org/2024/04/trumps-misleading-chart-on-illegal-immigration

The data in the chart itself are accurate, but the Trump campaign editorial notes are not. The red arrow at the bottom purports to correspond to the point that “Trump leaves office” and to be the “lowest illegal immigration in recorded history!” But the arrow actually points to April 2020, when there were 16182 apprehensions at the southwest border.

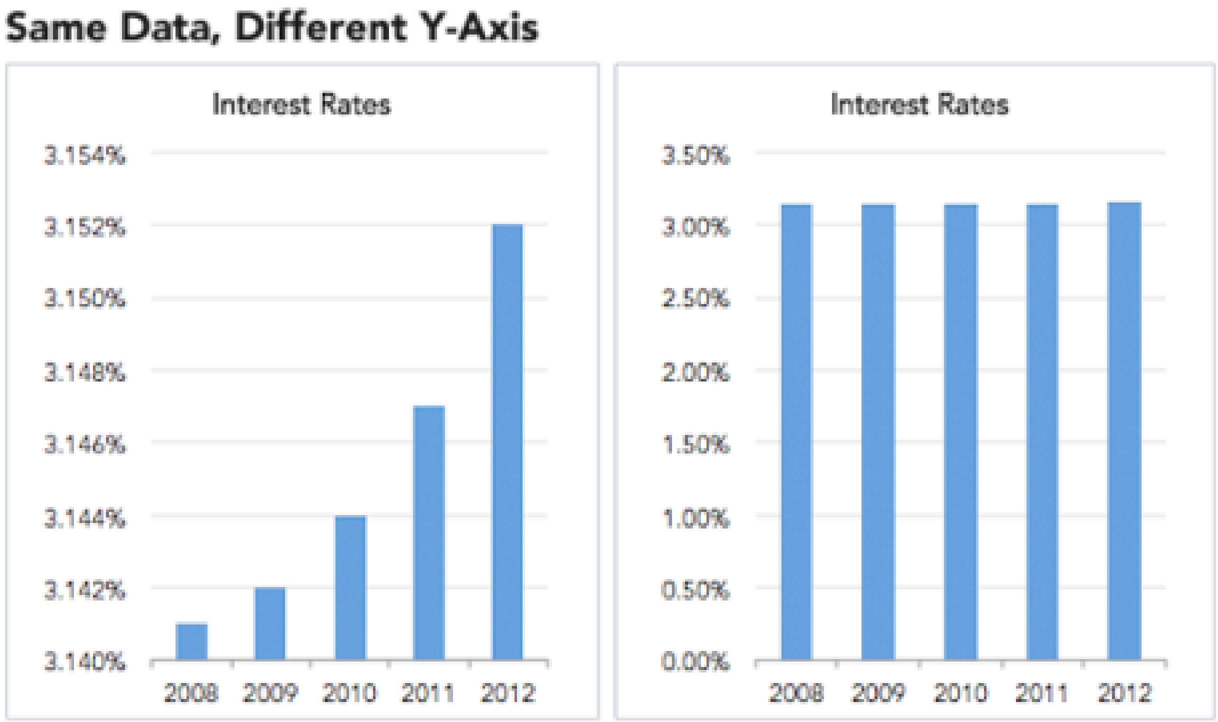
Bad use of data: truncated Y-Axis

- messing with the y-axis of a bar graph, line graph, or scatter plot (e.g., starting the range from a number other than 0)



Bad use of data: truncated Y-Axis

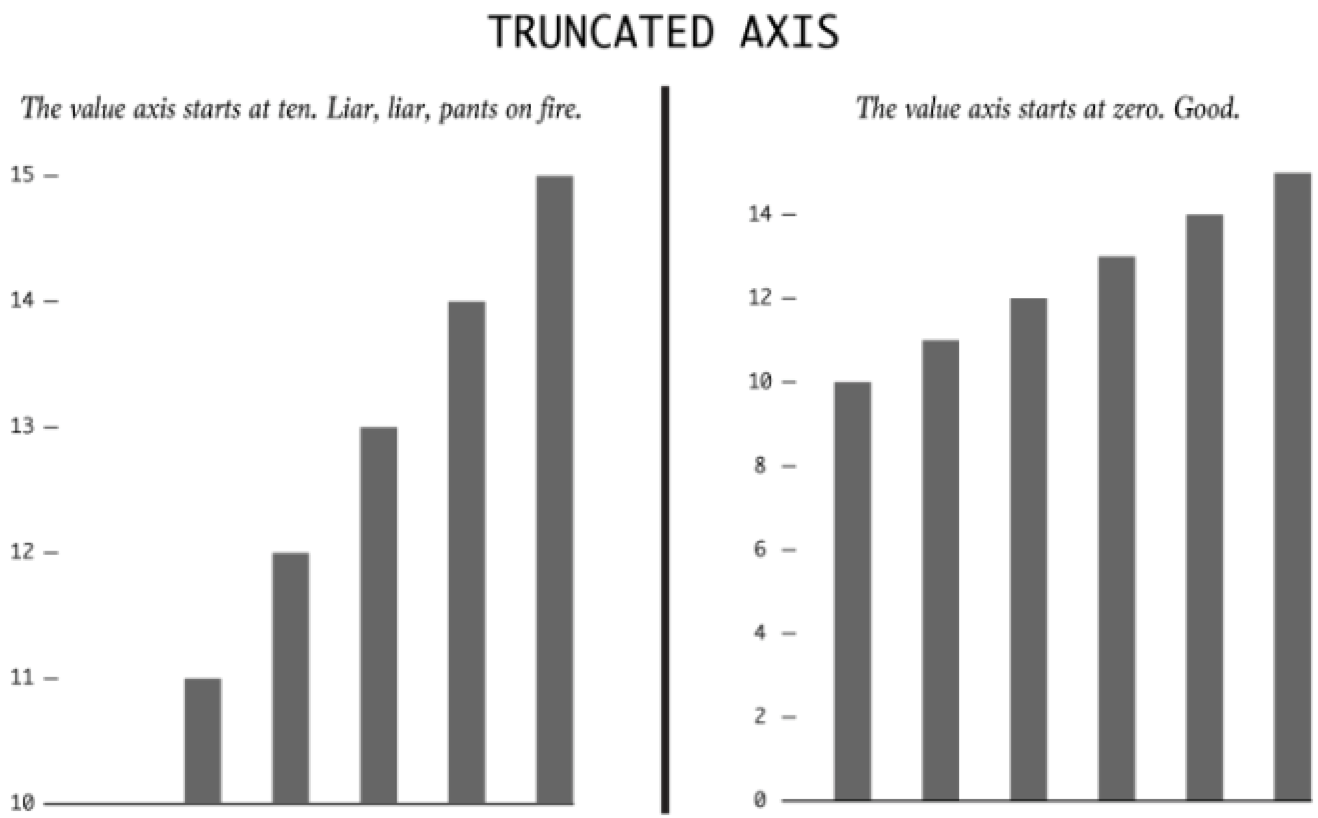
- Exact same data, but different scales for the y-axis



y-axis range 3.140% to 3.154% y-axis with a zero baseline

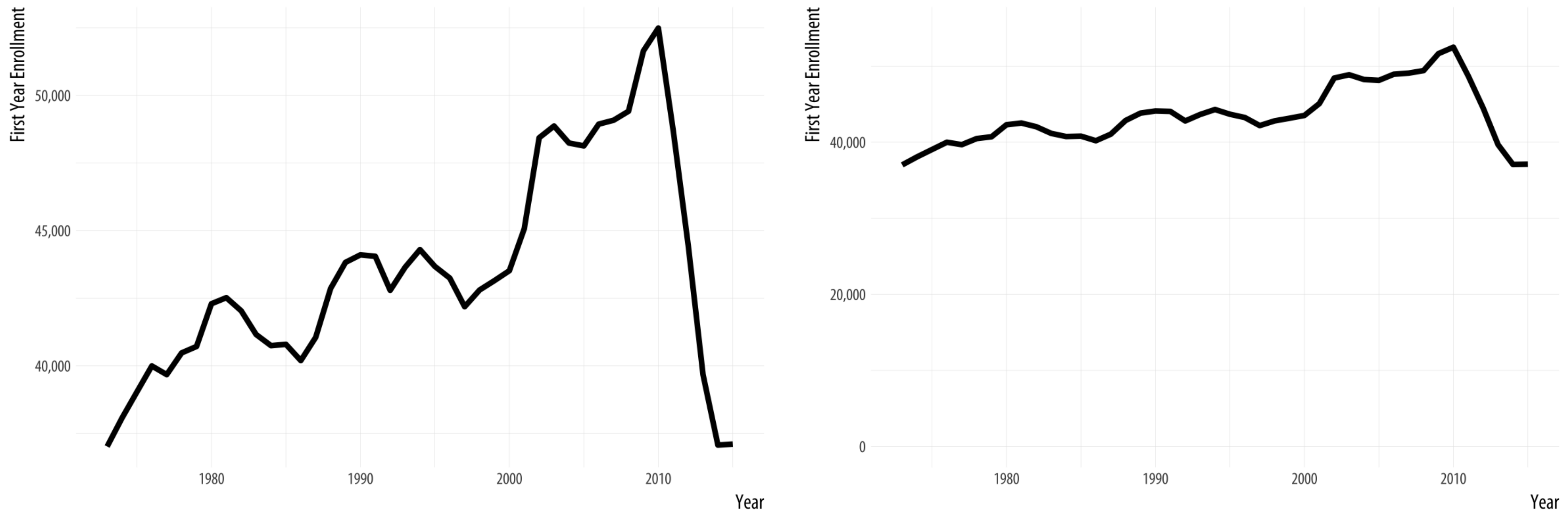
Bad use of data: truncated Y-Axis

- Exact same data, but different scales for the y-axis



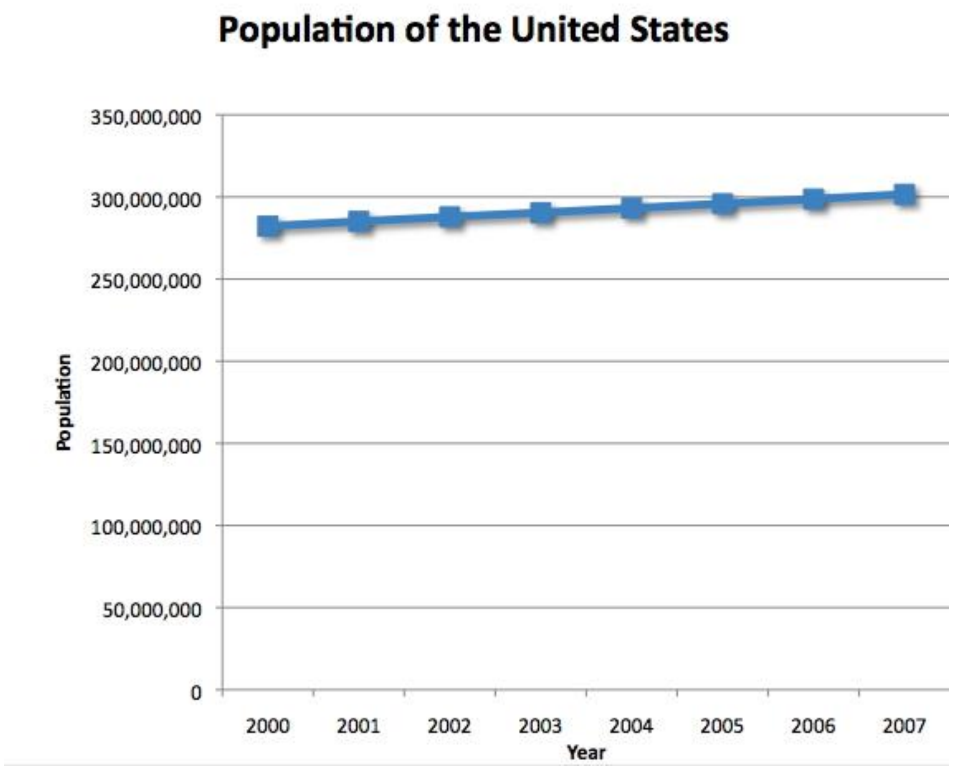
Bad use of data: truncated Y-Axis

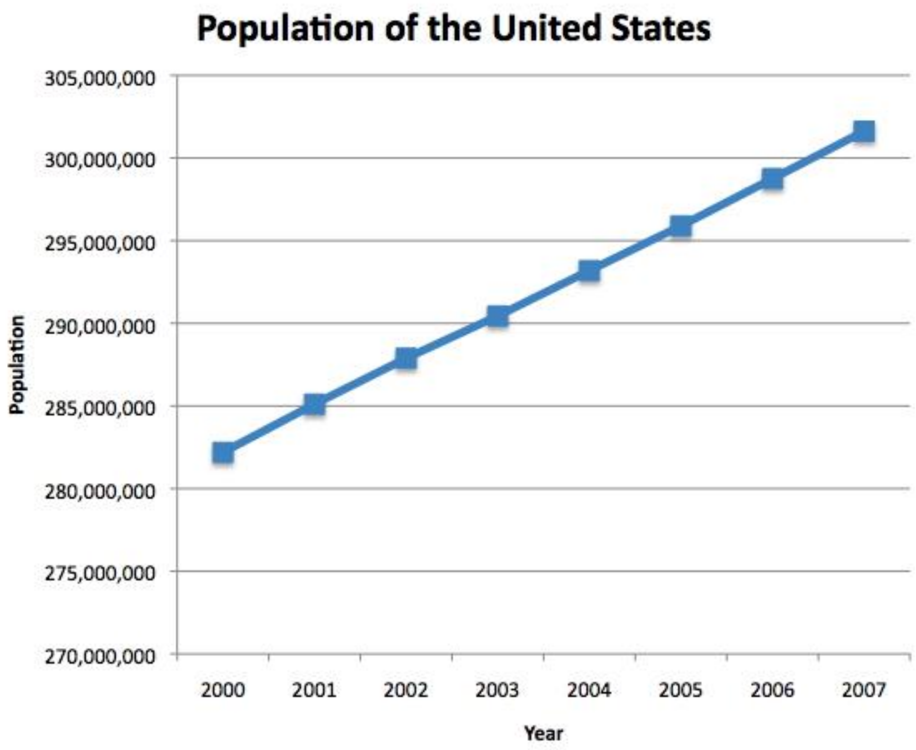
- Exact same data, but different scales for the y-axis



(Bad) use of data: choice of Y-Axis

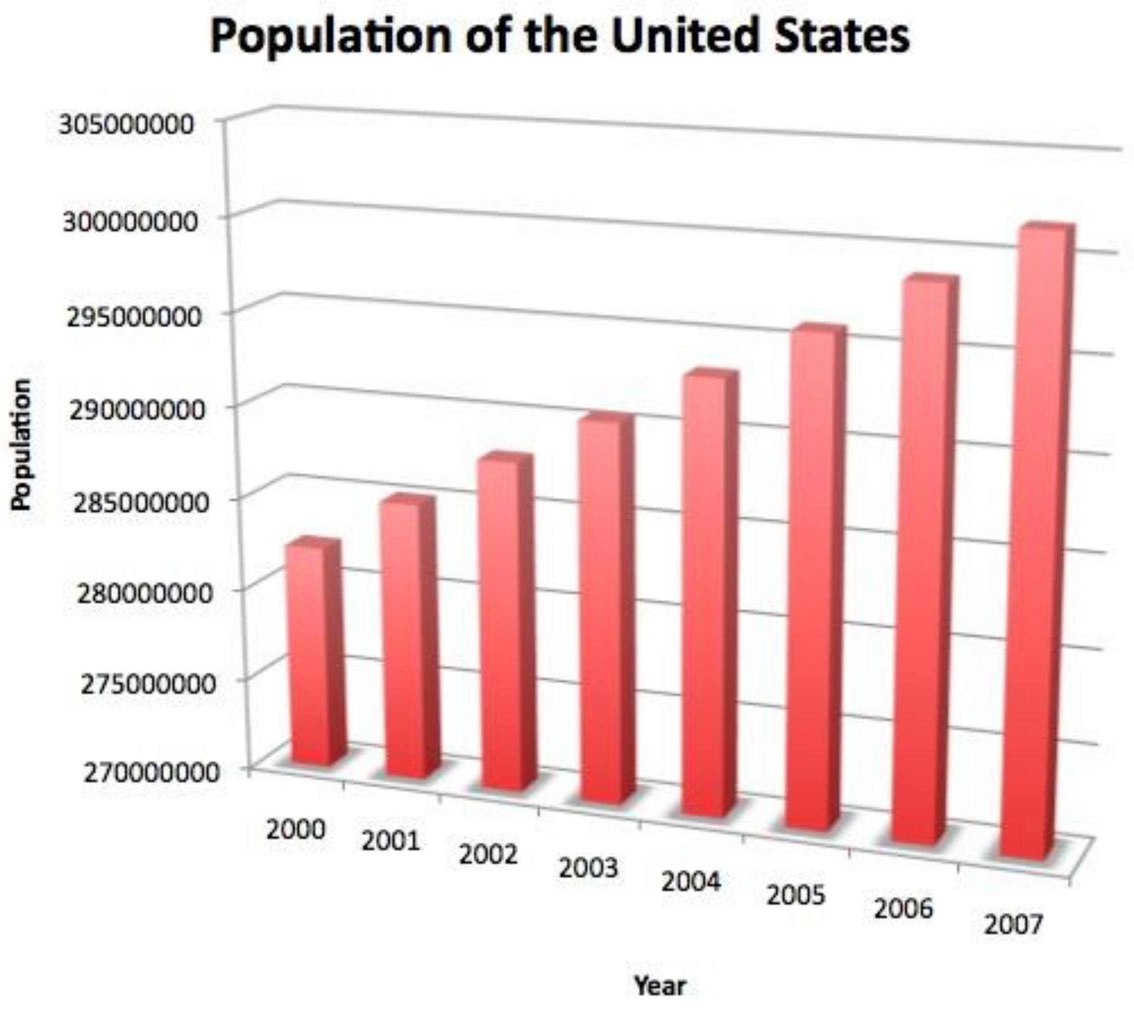
- Exact same data, but different scales for the y-axis





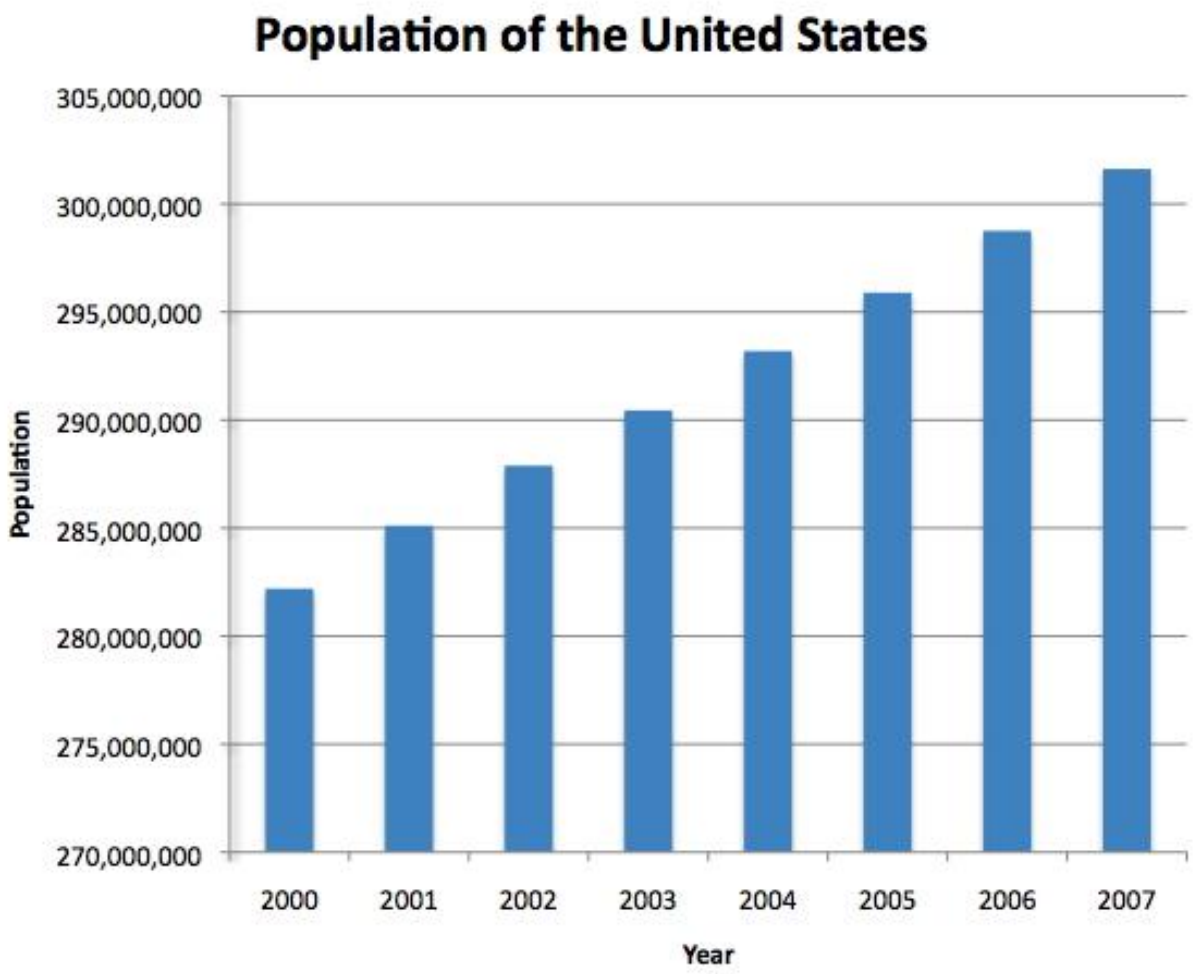
Bad use of data

- Hard to read numbers (plus the unnecessary use of 3D)



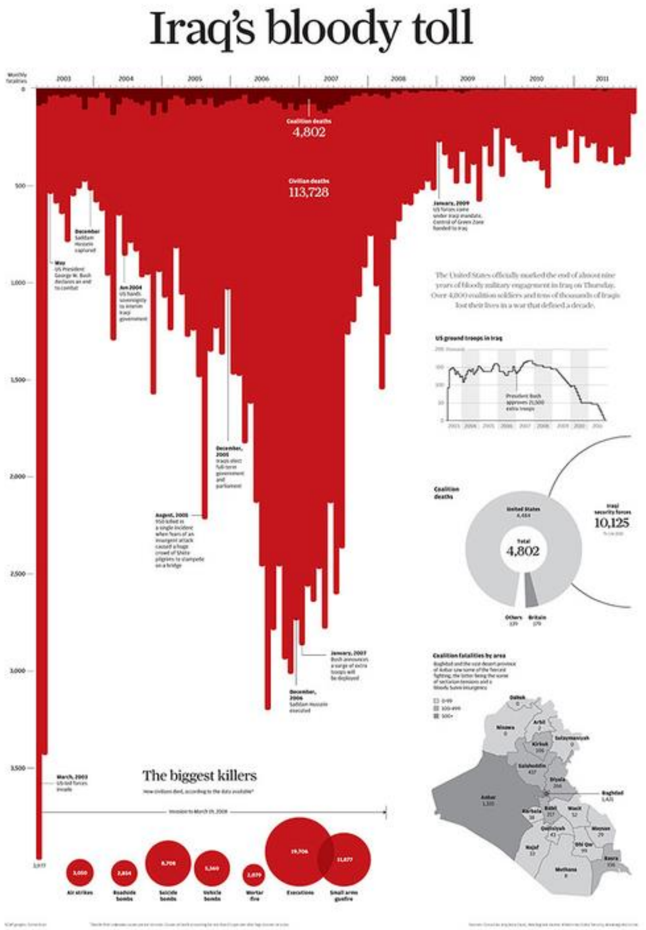
(Bad) use of data

- Same data but population values have commas, in 2D

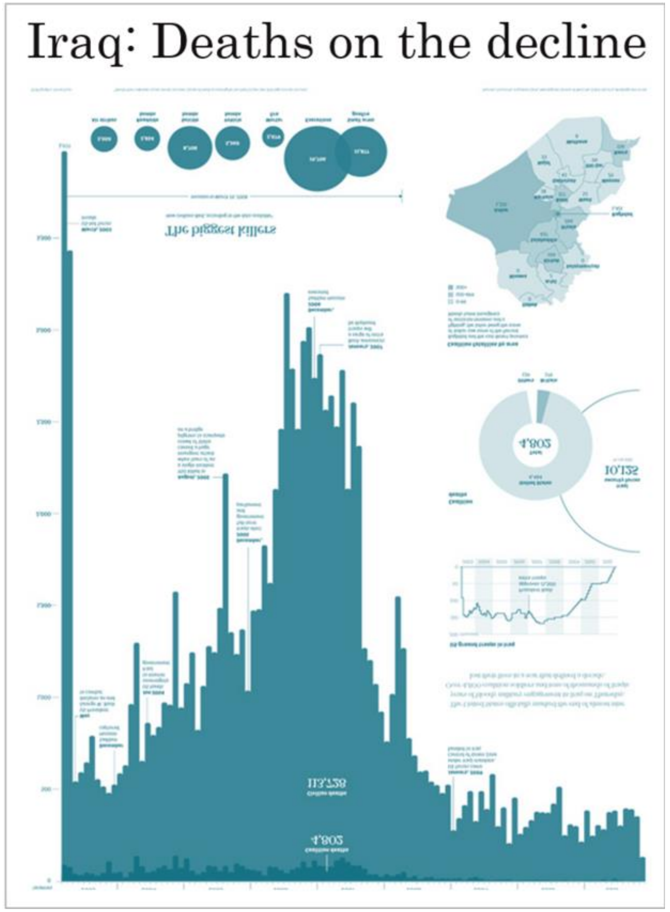


Use of data: choice of design, text

- exact same data, changes in orientation, title, color



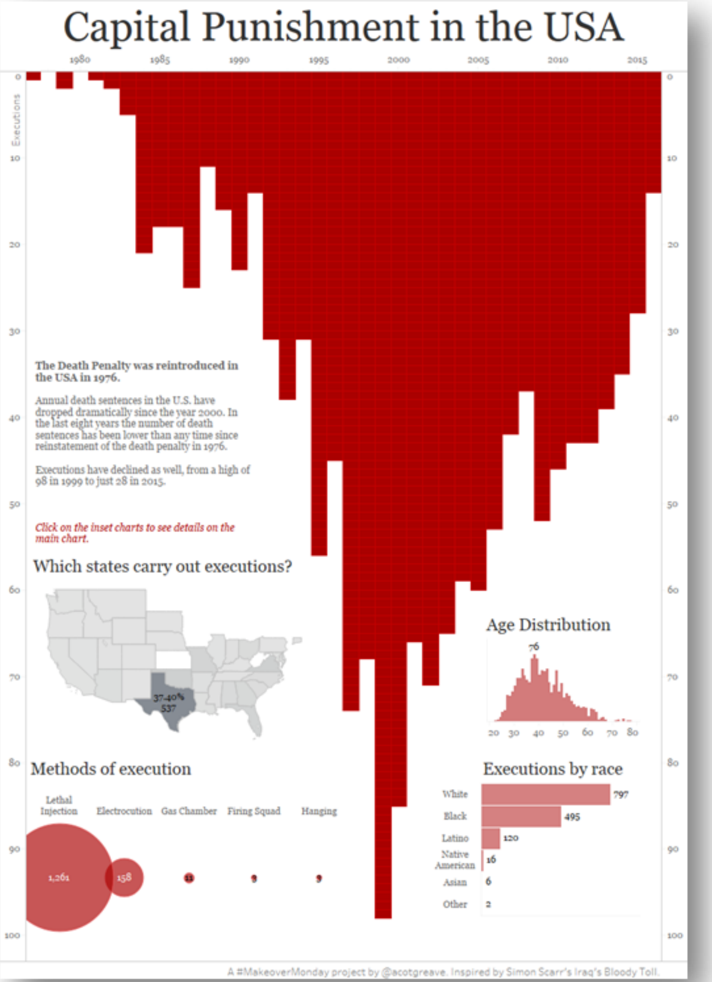
Simon Scarr (2011),
in the South China Morning Post

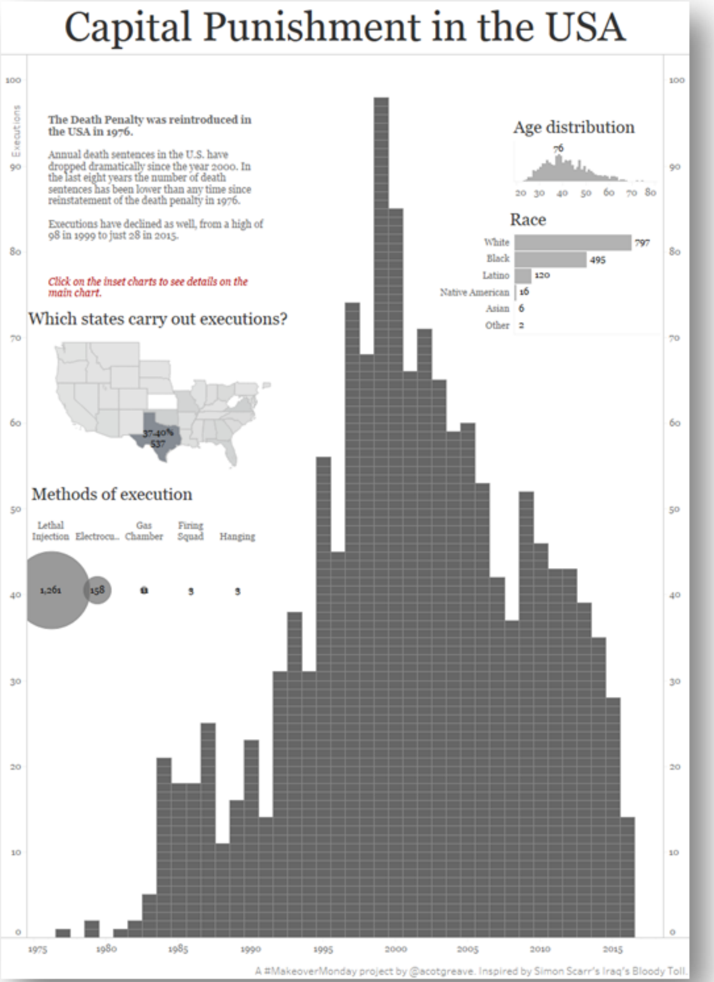


Andy Cotgreave (2016),
on InfoWorld

Use of data: choice of design, text

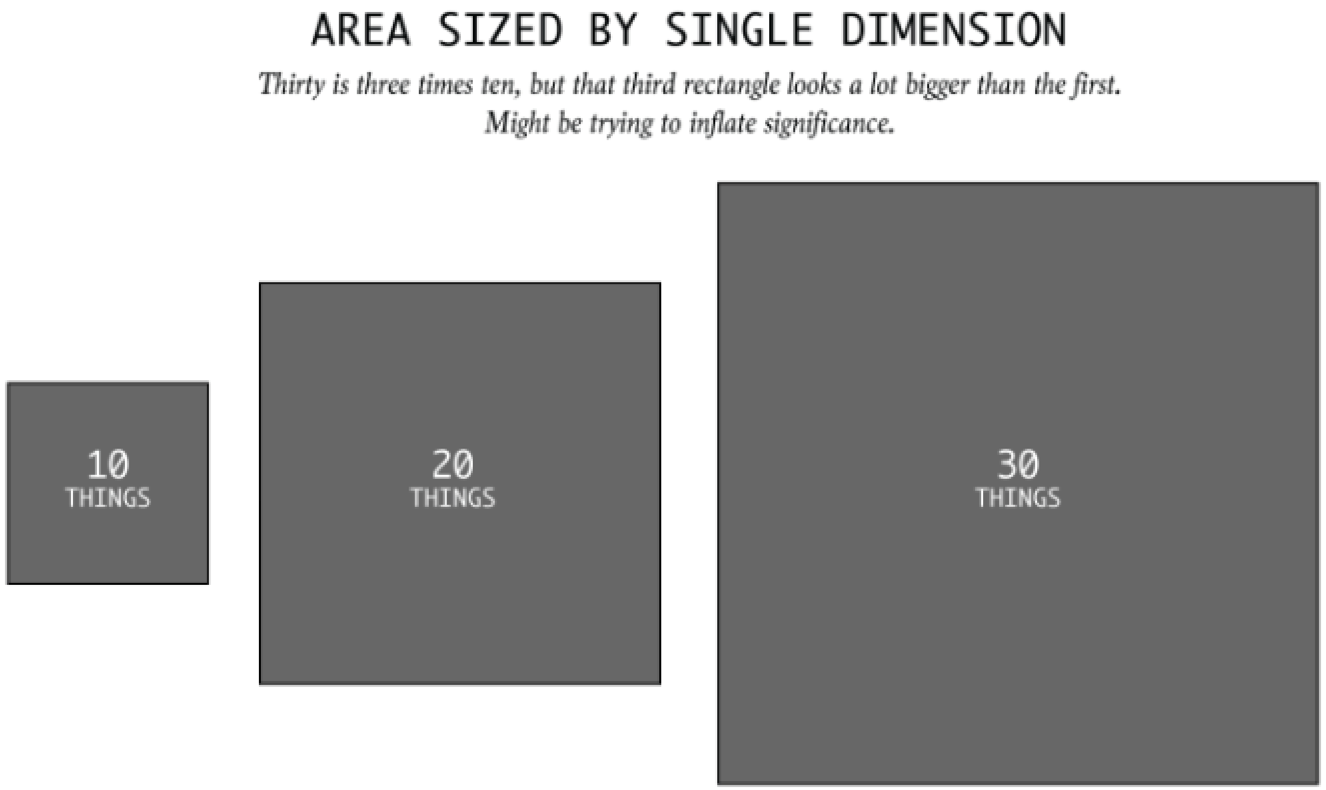
- exact same data, changes in orientation, title, color





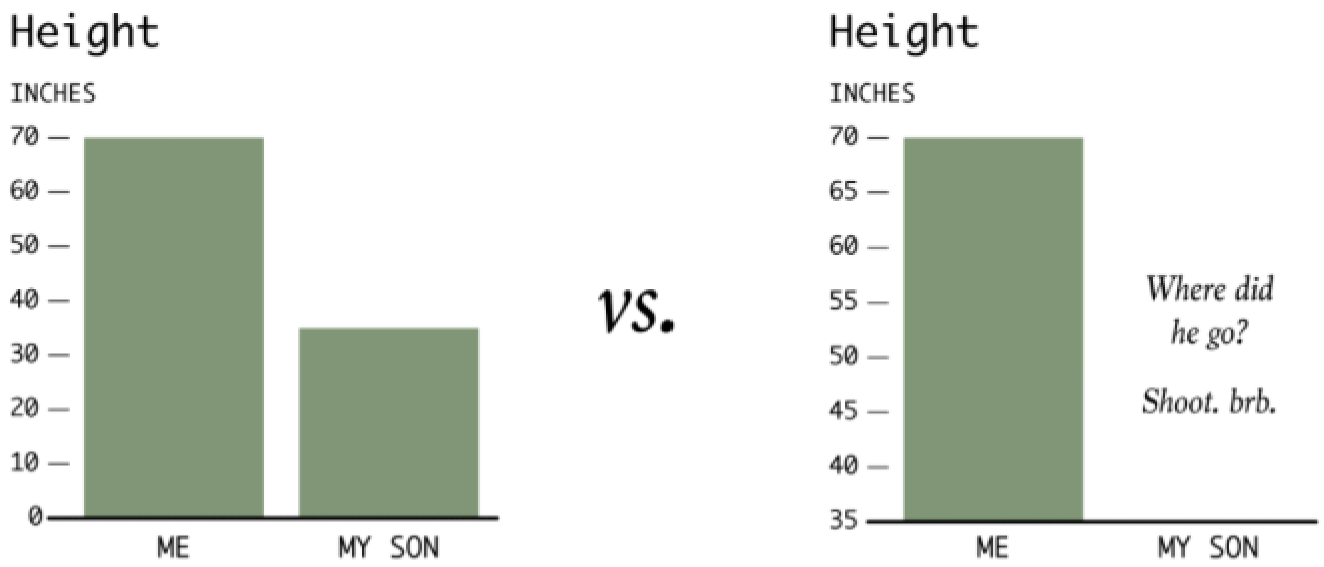
Bad use of data

- Scaling issues…



Bad use of data: Y-Axis height

- Messing with height to make it seem much larger than it is

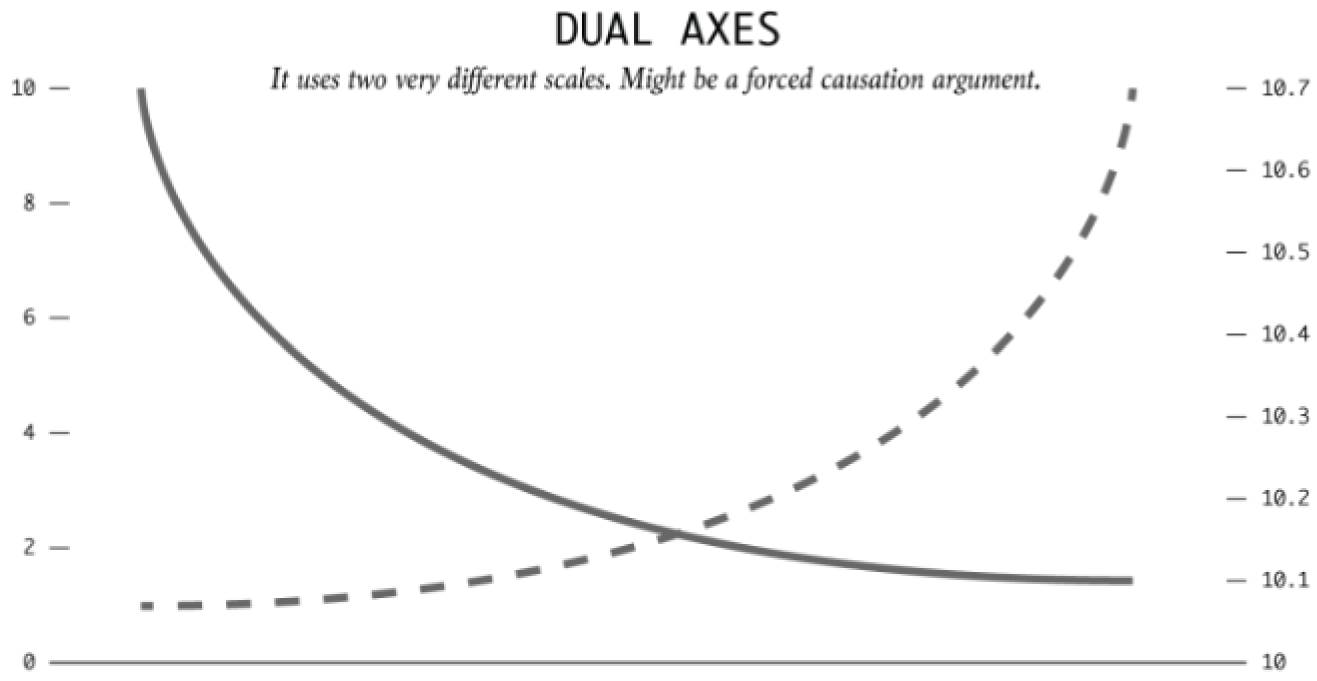


Bad use of data: Y-Axis height

- Messing with height to make it seem much larger than it is

Bad use of data: dual axes

- y-axes on the left and the right



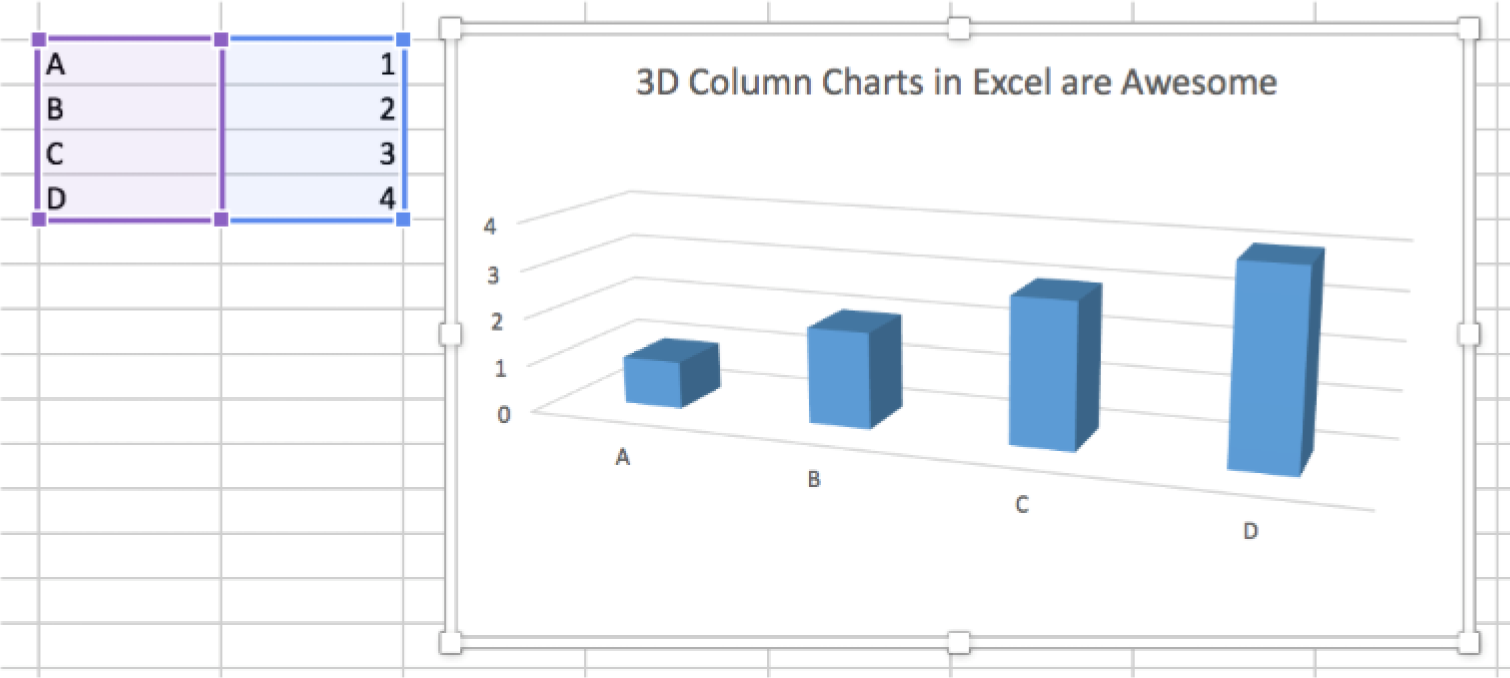
[Spurious correlations](http://www.tylervigen.com/spurious-correlations)
http://www.tylervigen.com/spurious-correlations

Bad perception

- Visualizations encode numbers in lines, shapes, and colors. That means that our interpretation of these encodings is partly conditional on how we perceive geometric shapes and relationships generally.
- Poorly-encoded data can be misleading

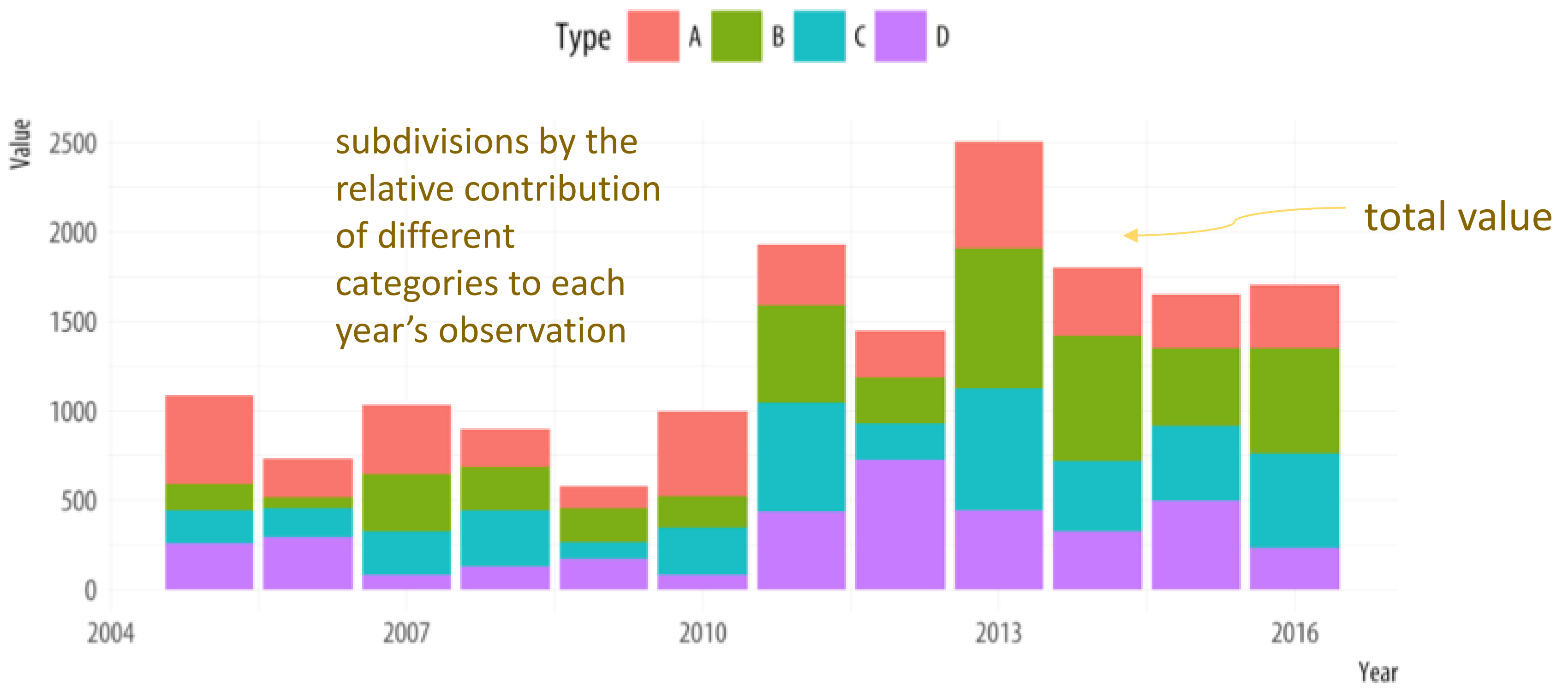
Bad perception

- The 3D columns in combination with the default angle of view for the chart make the values as displayed differ substantially from the ones actually in the cell.
- Each column appears to be somewhat below its actual value.



Bad perception: mapping data to visual encodings

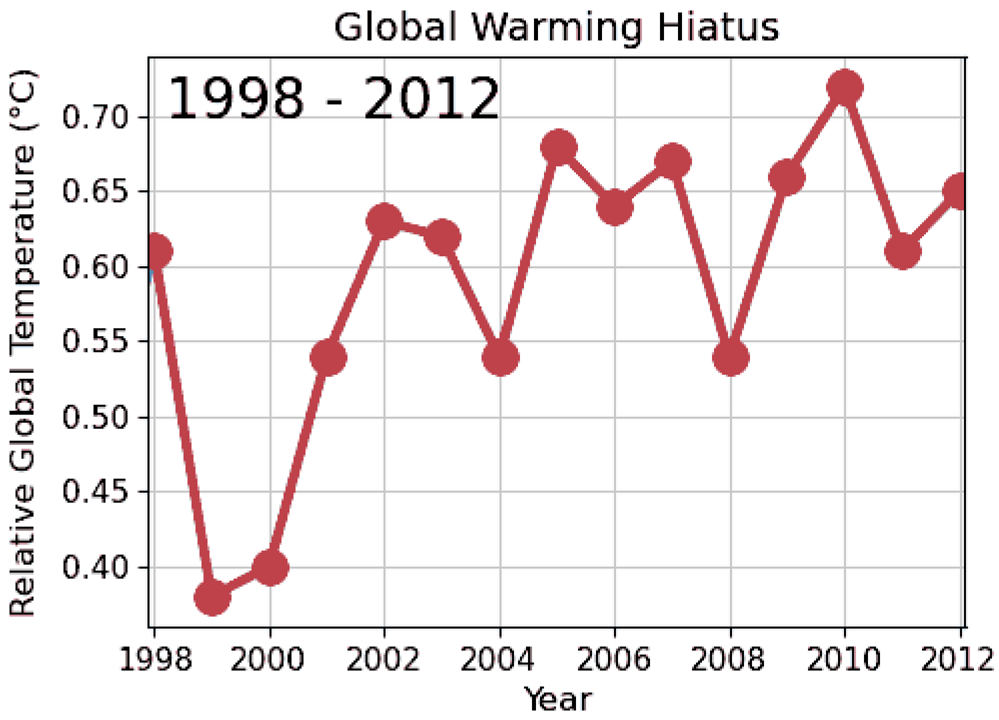
- even without junk, it can be a challenge



The overall trend is readily interpretable, and it is also possible to easily follow the over-time pattern of the category that is closest to the x-axis baseline (Type D, colored in purple). But the fortunes of the other categories are not so easily grasped. Comparisons of both the absolute and the relative share of Type B or C are much more difficult. Relative comparisons need **a stable baseline**.

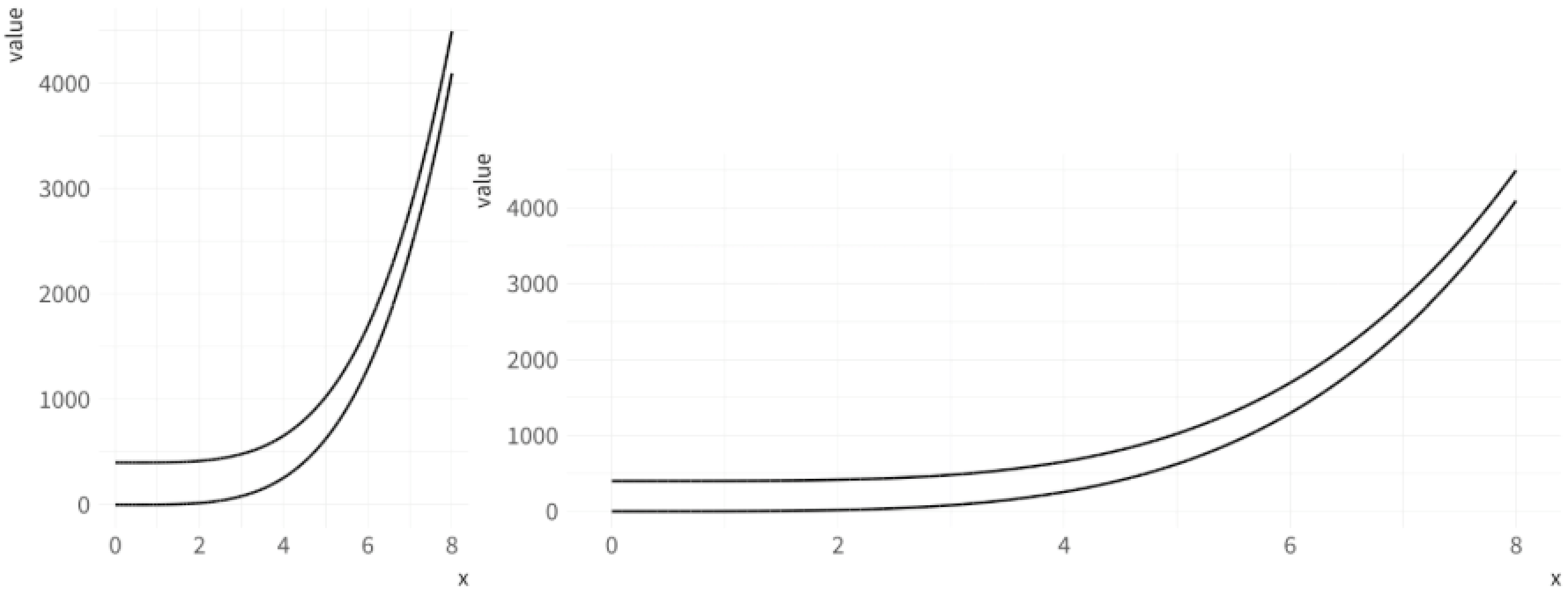
Cherry picking

- Suppressing evidence, or the fallacy of incomplete evidence = pointing to individual cases or data that seem to confirm a particular position while ignoring a significant portion of related and similar cases or data that may contradict that position.



Bad perception: aspect ratios

- Aspect ratios affect our perception of rates of change



The Lie Factor

- Graphs rely on our understanding that a number is represented visually by the magnitude of some graphical element.

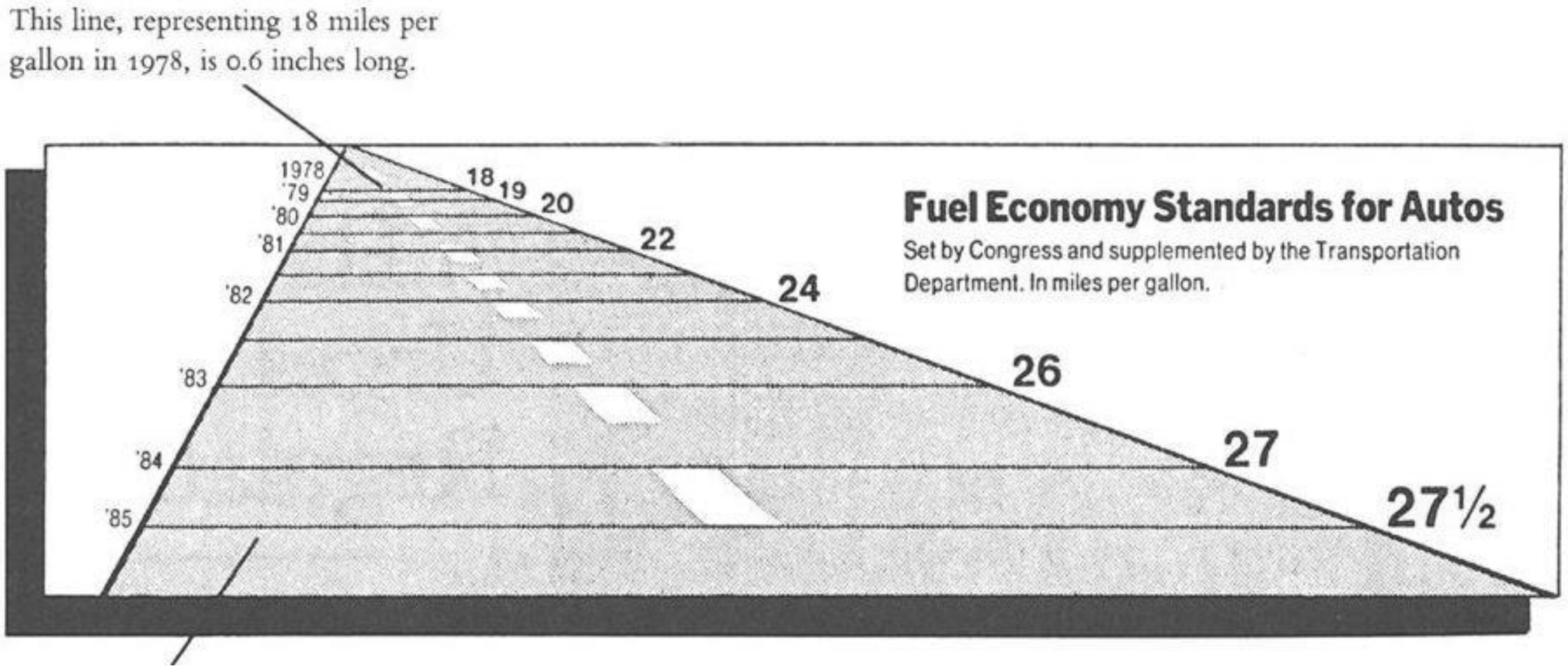
The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented.

Tufte, 1983, p.57

- The violation of this principle is measured by the

$$\textbf{Lie ratio} = \frac{\textbf{size of effect in graphic}}{\textbf{size of effect in data}}$$

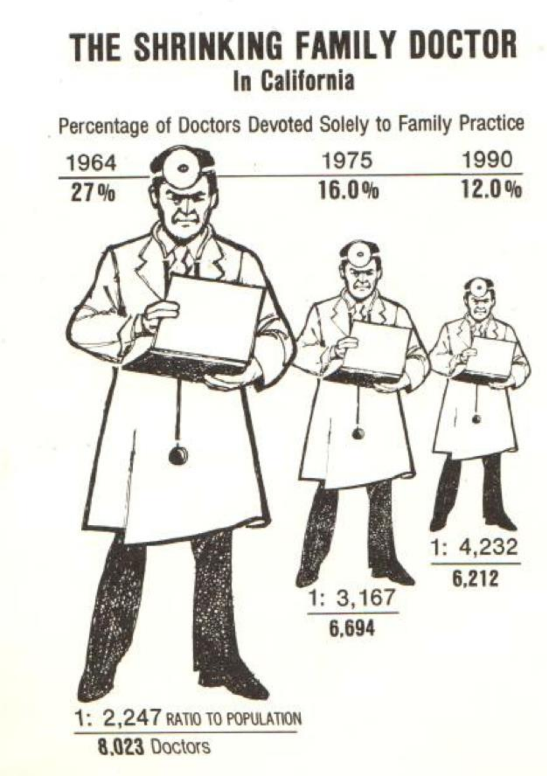
The Lie Factor



This NY Times graph purports to show the mandated fuel economy standards set by the US Dept. of Transportation. The standard required an increase in mileage from 18 to 27.5, an increase of 53%. The magnitude of increase in the graph is 783%

lie factor = (783/53) = 14.8!

The Lie Factor

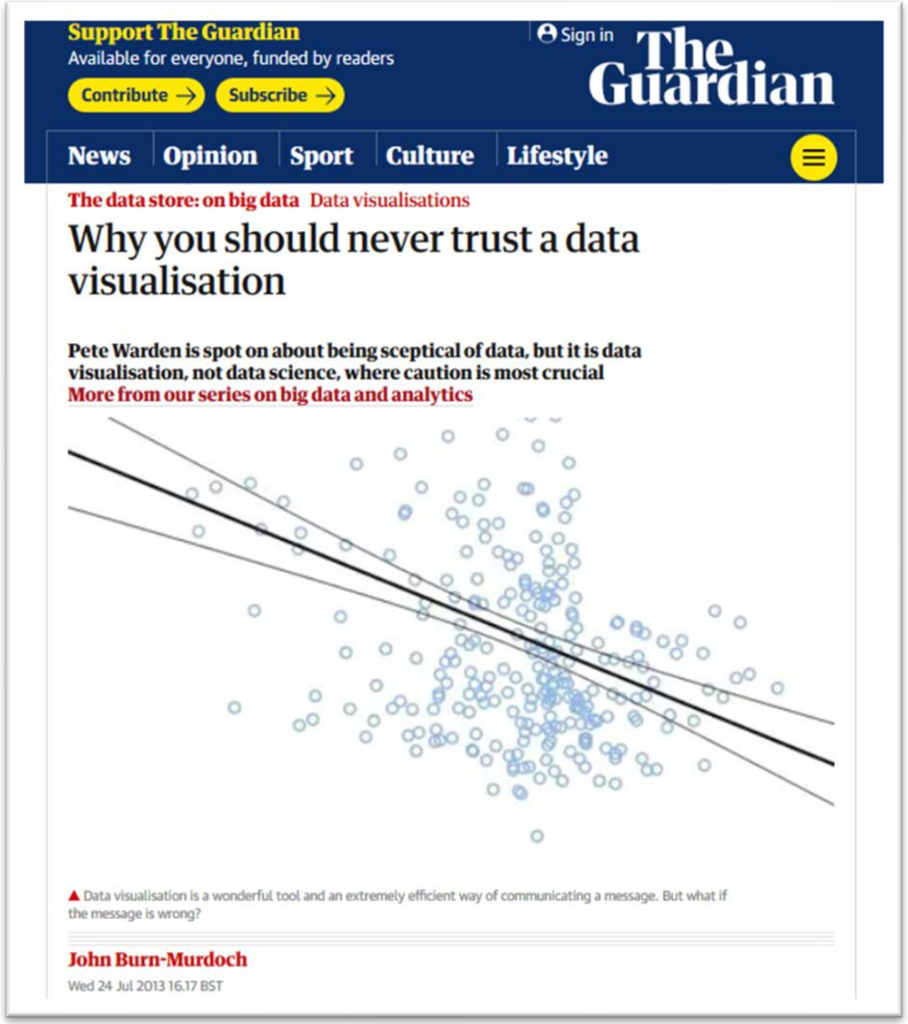


Changes in the scale of the graphic should always correspond to changes in the data being represented. This graph violates that principle by using area to show the data

lie factor = 2.8

So, should we trust a data visualization?

- Debate



Support The Guardian
Available for everyone, funded by readers
Contribute → Subscribe →
Sign in
The Guardian
News | Opinion | Sport | Culture | Lifestyle

The data store: on big data Data visualisations

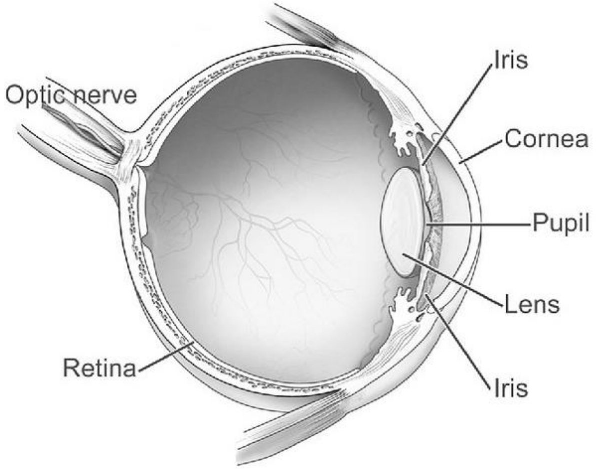
Why you should never trust a data visualisation

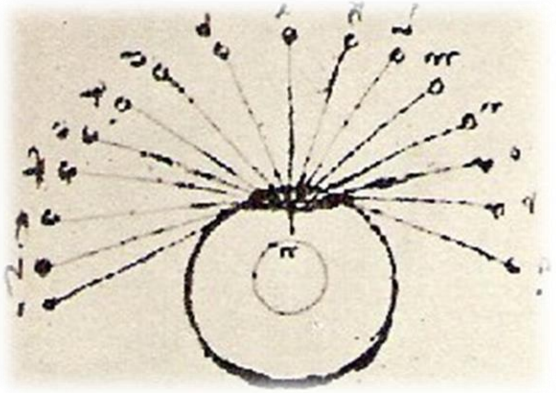
Pete Warden is spot on about being sceptical of data, but it is data visualisation, not data science, where caution is most crucial
More from our series on big data and analytics

▲ Data visualisation is a wonderful tool and an extremely efficient way of communicating a message. But what if the message is wrong?

John Burn-Murdoch
Wed 24 Jul 2013 16.17 BST

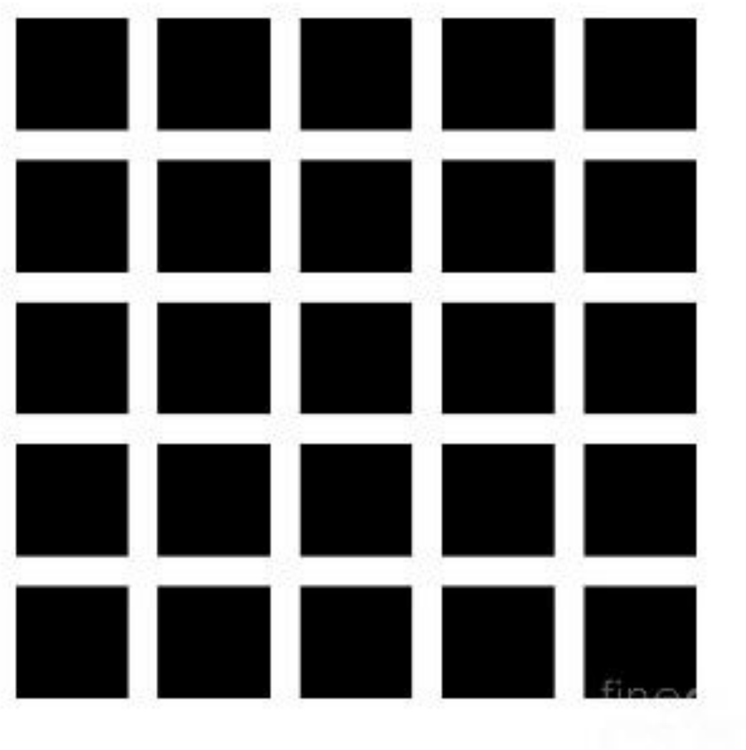
Visual perception





Visual perception

- Looking at pictures of data means looking at lines, shapes, and colors.
- Our visual system works in a way that makes some things easier for us to see than others.



Hermann Grid Effect, 1870

Visual perception

- When the gray bars share a boundary, the apparent contrast between them appears to increase.
- Our visual system is trying to construct a representation of what it is looking at based more on relative differences in the luminance (or brightness) of the bars, rather than their absolute value.

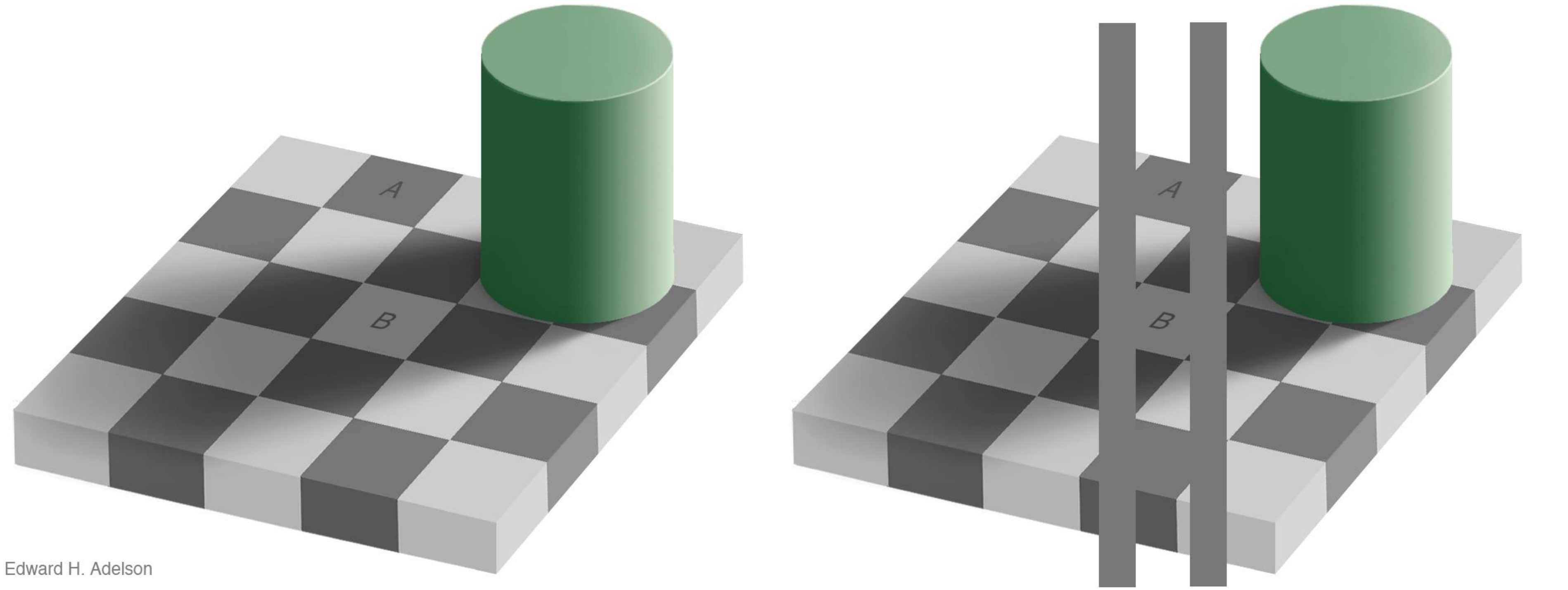
Mach bands

Visual perception

- The role of background contrasts: The same shade of gray is perceived very differently depending on whether it is against a darker background or a lighter one.
- We are better at distinguishing darker shades than we are at distinguishing lighter ones.
- We do better at distinguishing very light shades of gray when they are set against a light background. When set against a dark background, differences in the middle-range of the light-to-dark spectrum are easier to distinguish.

Visual perception

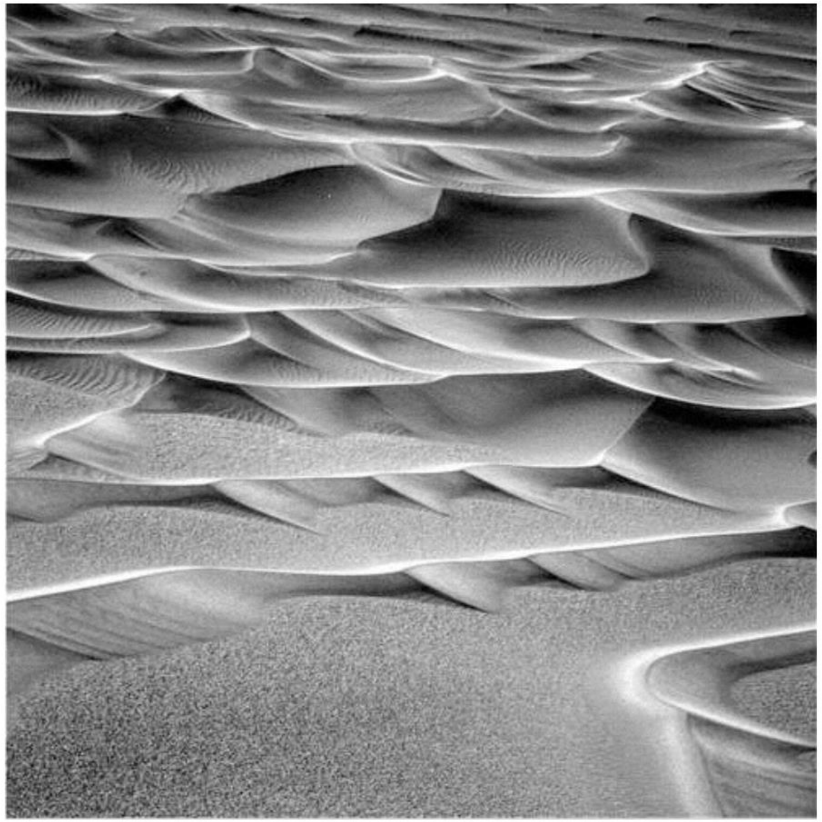
- Our visual system is attracted to edges, and we assess contrast and brightness in terms of relative rather than absolute values.
- To figure out the shade of the squares on the floor, we compare it to the nearby squares, and we also discount the shadows cast by other objects.

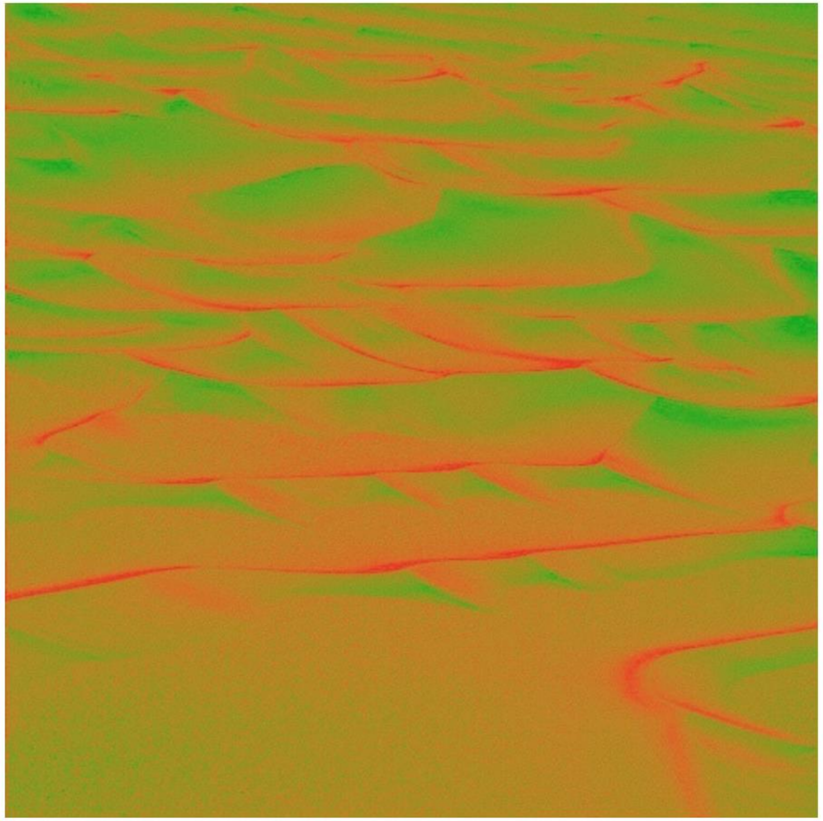


The checkershadow illusion (Edward H. Adelson)

Edges, contrasts, and colors

- Our visual system is attracted to edges
- Our ability to see edge contrasts is stronger for monochrome images than for color.





Ware (2008, p. 71)

Color

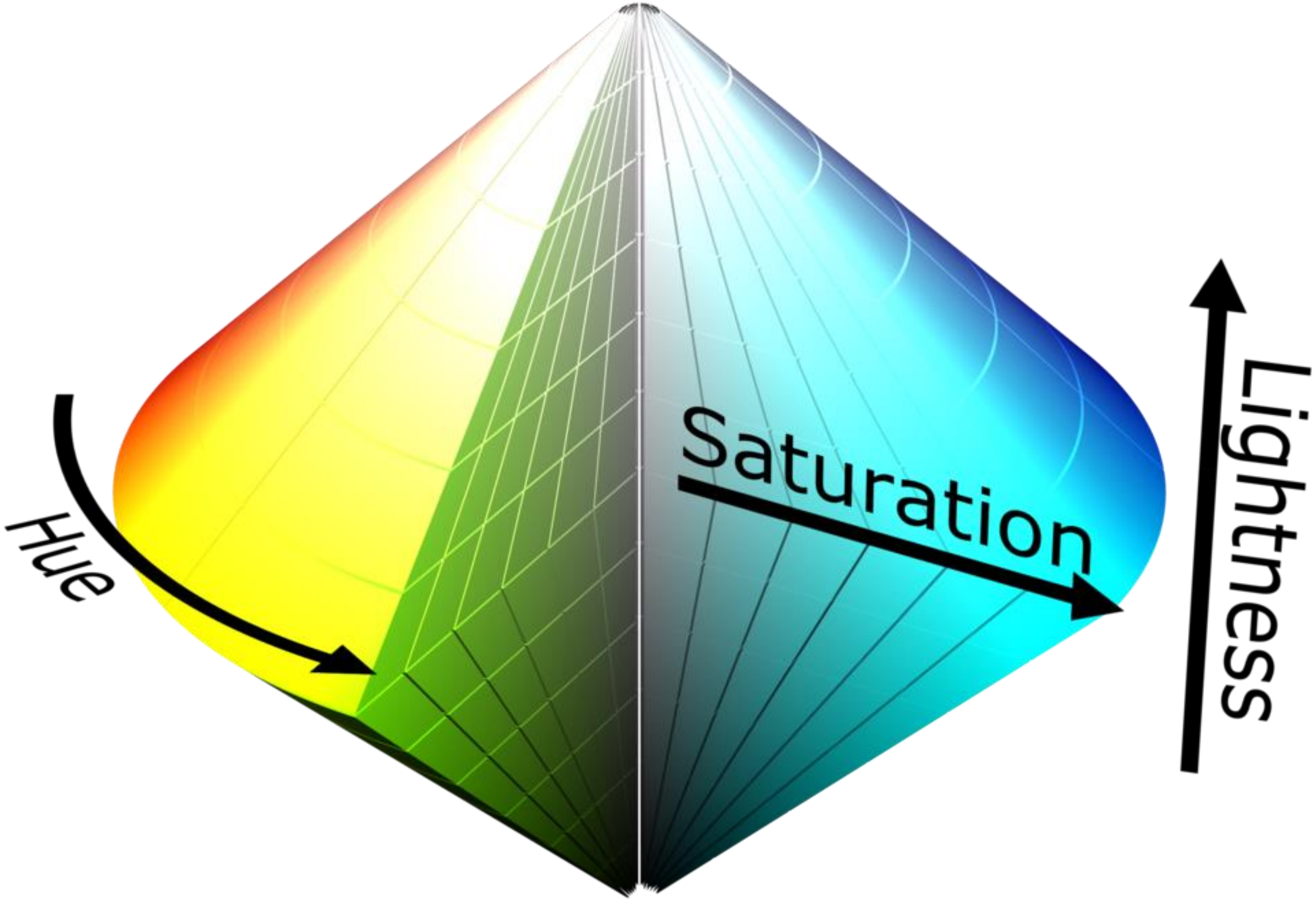
- To label
- To emphasize
- To liven or decorate

Color

- Color in data visualization introduces a number of complications (Zeileis & Hornik, 2006).
- Colors, generally, should be reserved for a 3rd (or 4th, 5th, etc.) variable

Color – HSL scale

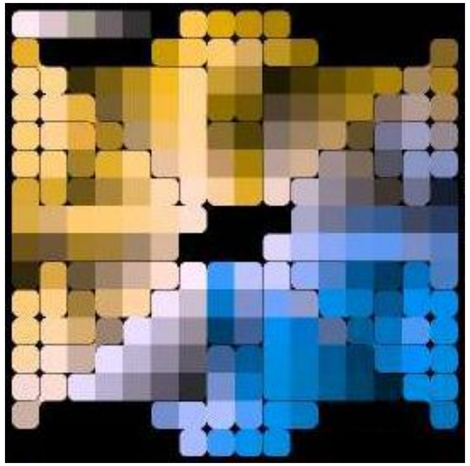
- Hue – Chroma – Luminance color model



Color – HSL scale

Three components: The “name” of the color

- **Hue** = light wave length
  - Blue (B) < Green (G) < Red (R)
  - The average person perceives ~150 hues
- **Luminance** = lightness of color
- **Saturation** = density of color = how “clean” = quantitiy of white light



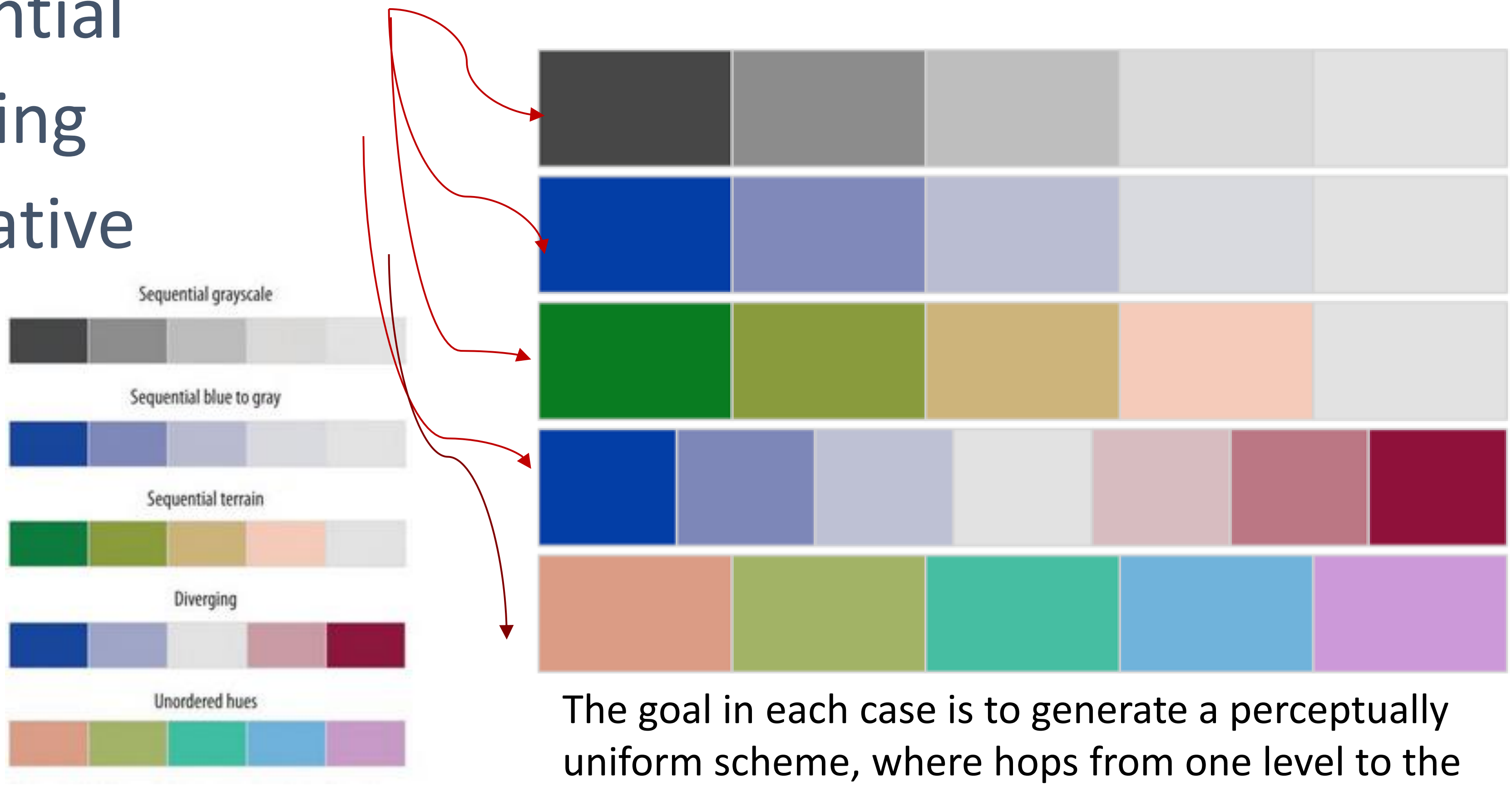
Color

Color is important for:

- Categorization, grouping, distnguishing
- Facilitate search
- perception:
  - Red = danger
  - Blue = cool
  - Orange = warm
- Good for novice users
- Beware of too much color («**color pollution**»)
- Beware of the foreground – background contrast

Color palettes (using the HCL color model)

- Types of color palletes:
  - Sequential
  - Diverging
  - Qualitative

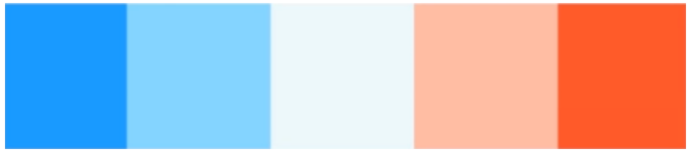


The goal in each case is to generate a perceptually uniform scheme, where hops from one level to the next are seen as having the same magnitude.

Types of data and appropriate color palettes

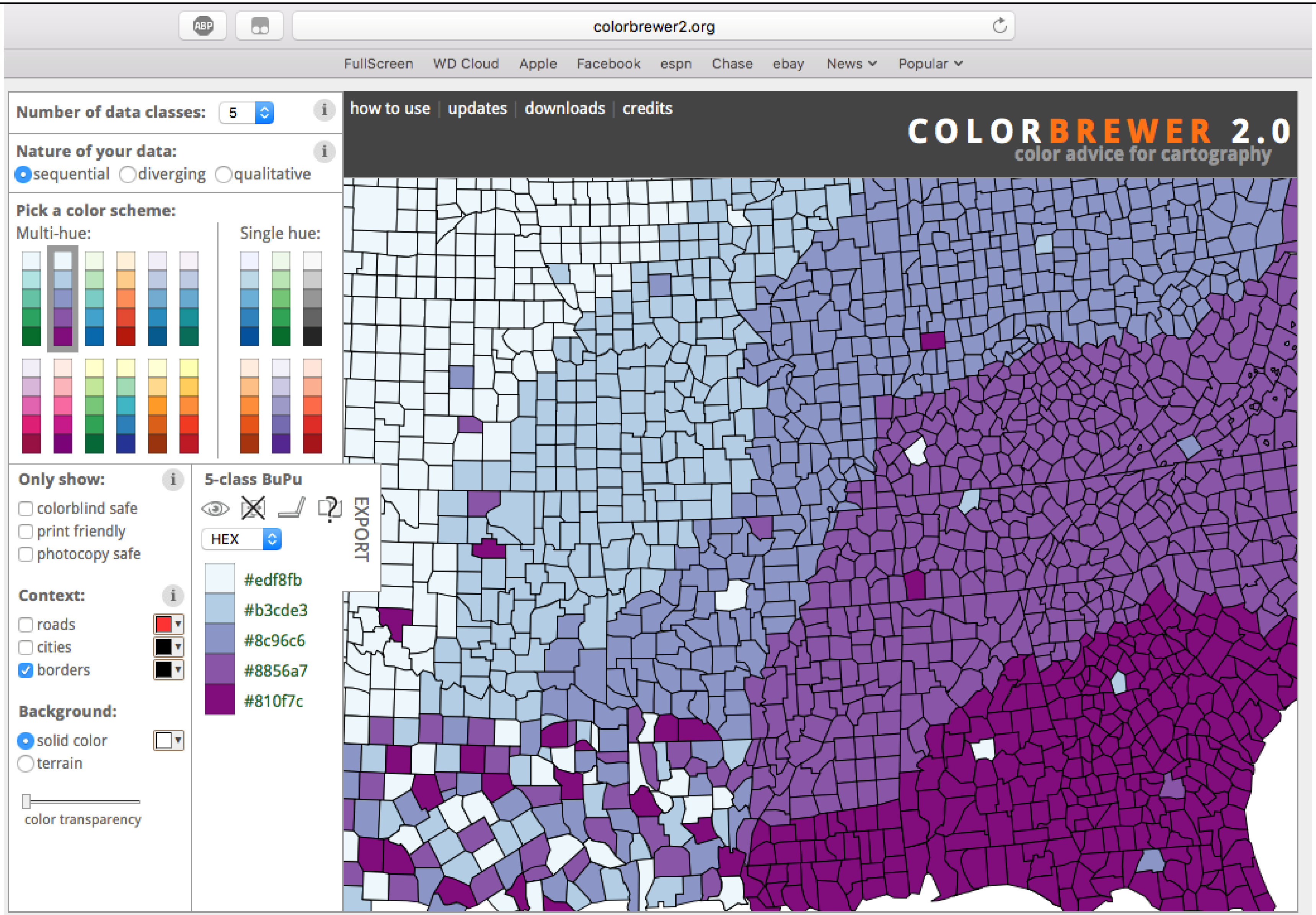
- Sequential
  - Quantitative or ordinal data

- Diverging
  - Quantitative or ordinal data with a meaningful midpoint

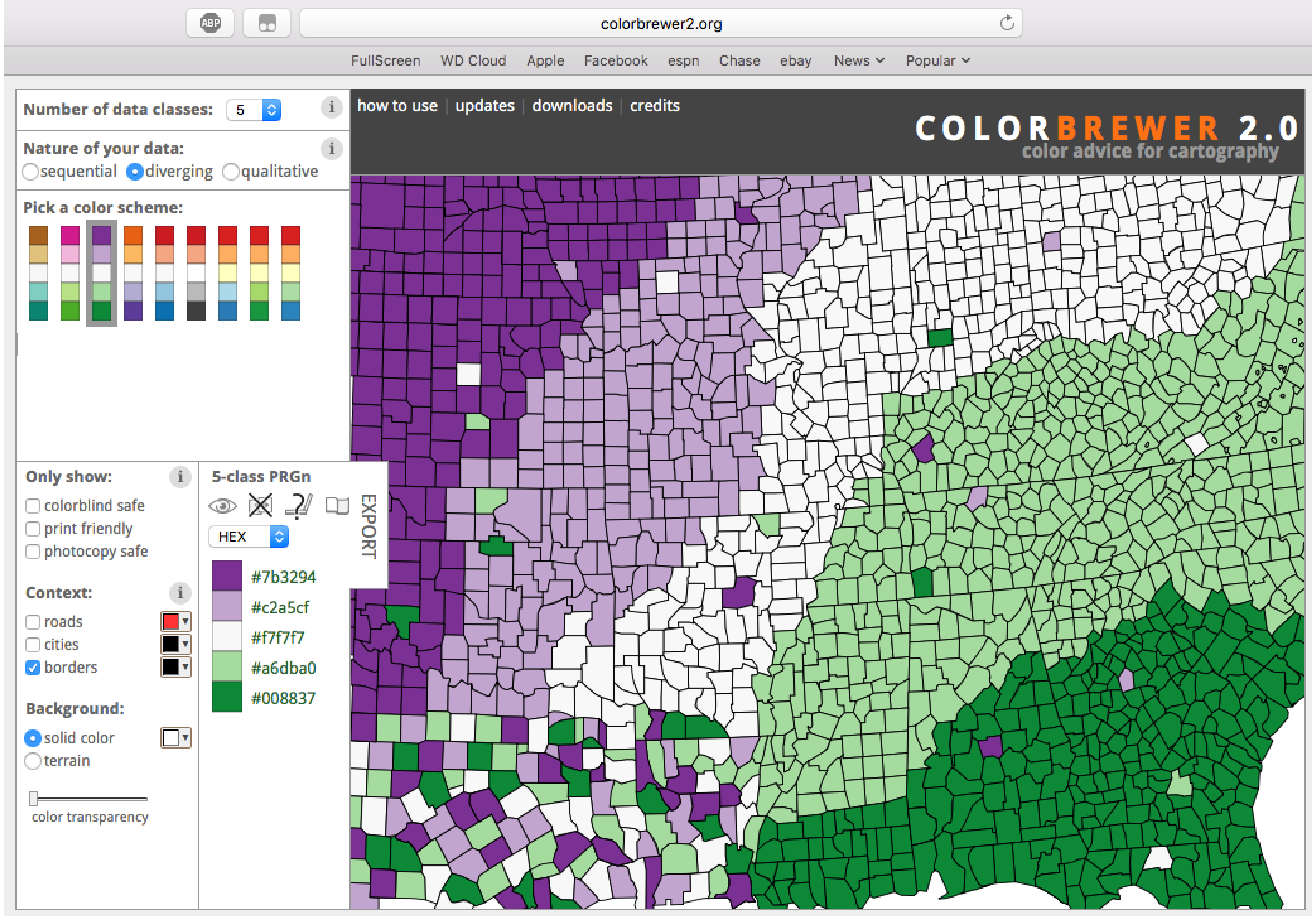


- Qualitative
  - Nominal data

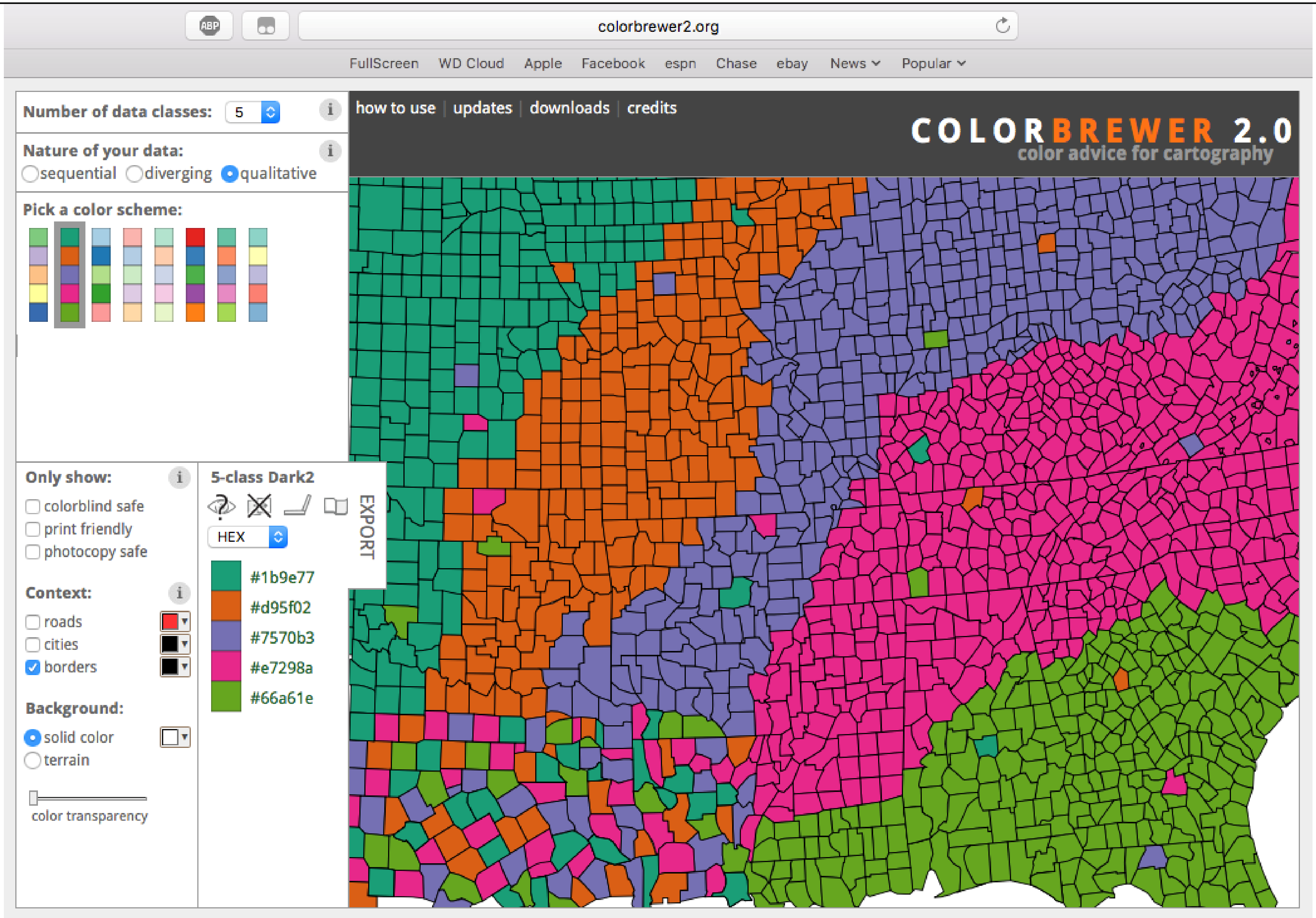
Palletes: Sequential



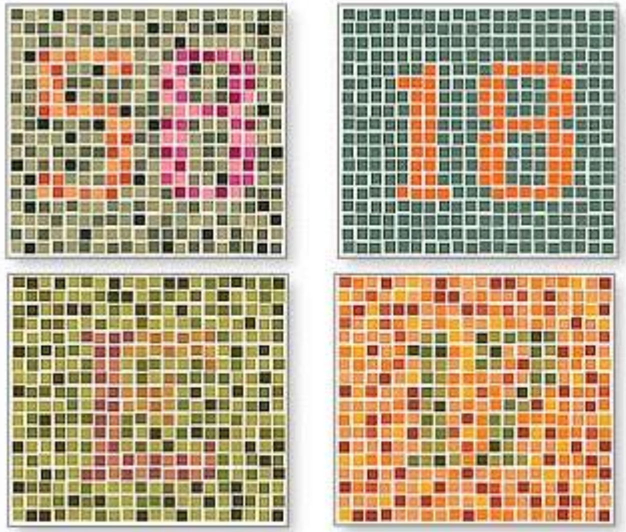
Palletes: Diverging

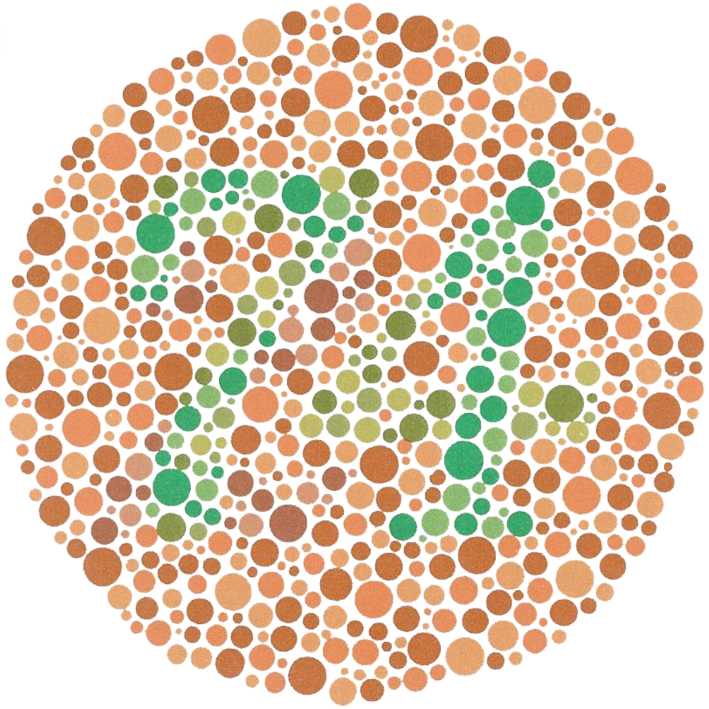


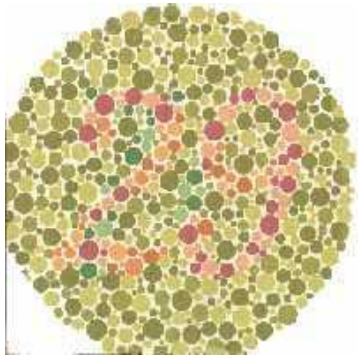
Palletes: Qualitative (unordered)

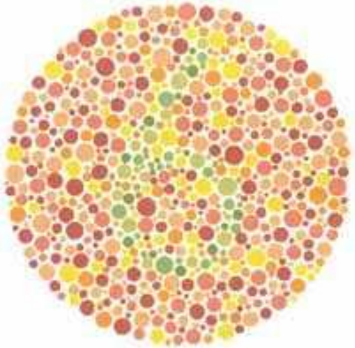


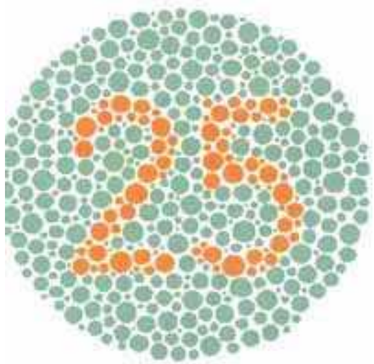
Visual perception: color – color blindness

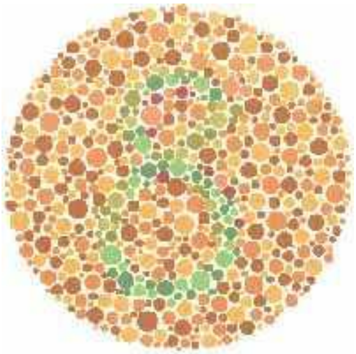








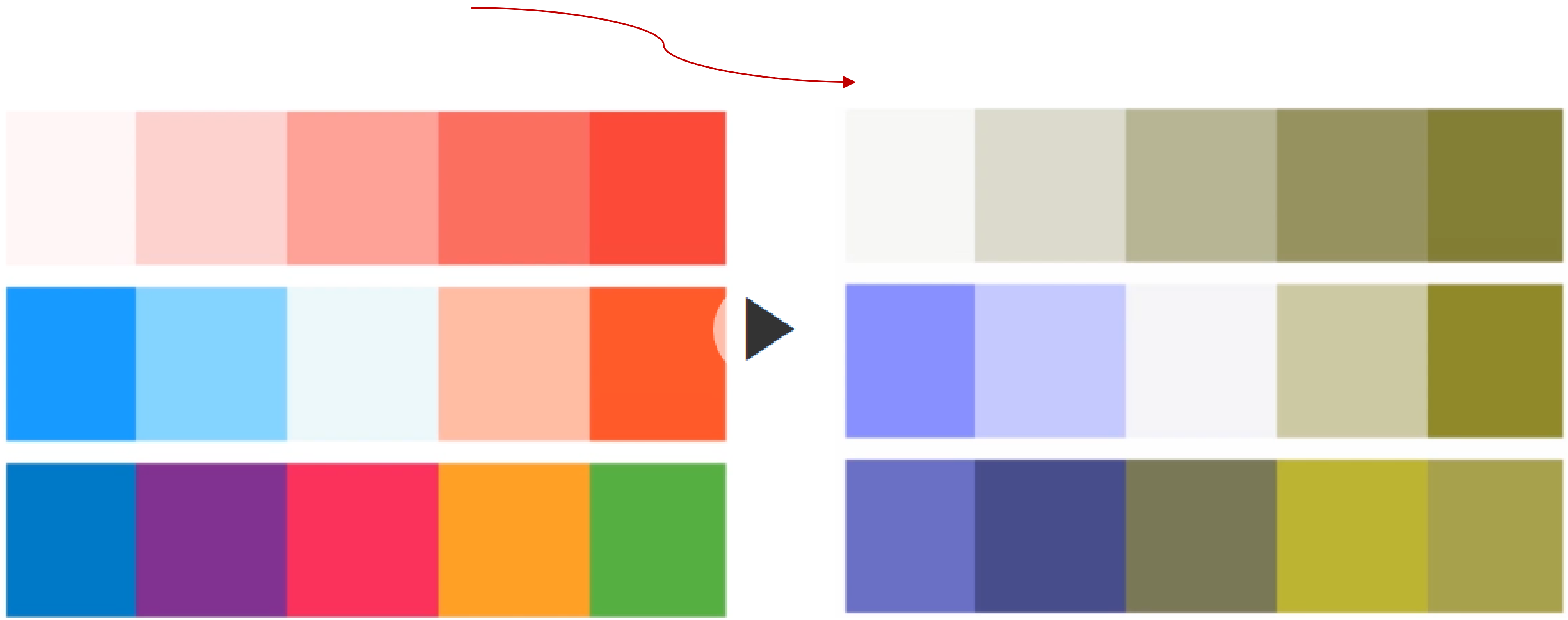




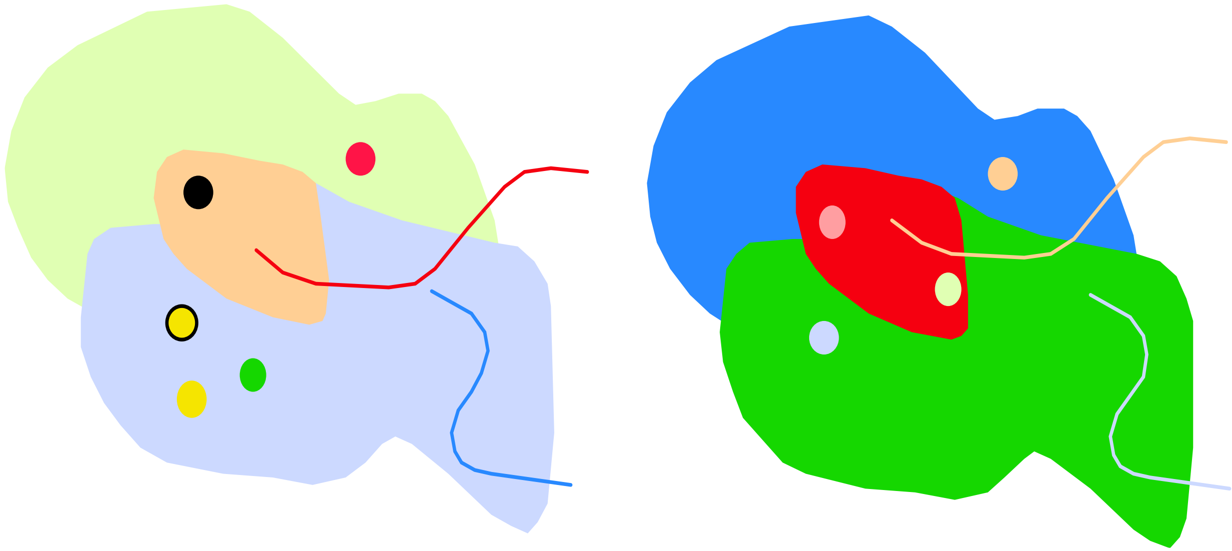
Ishihara color test plates

Visual perception: color – color blindness

- This is what it looks like to someone who is color blind



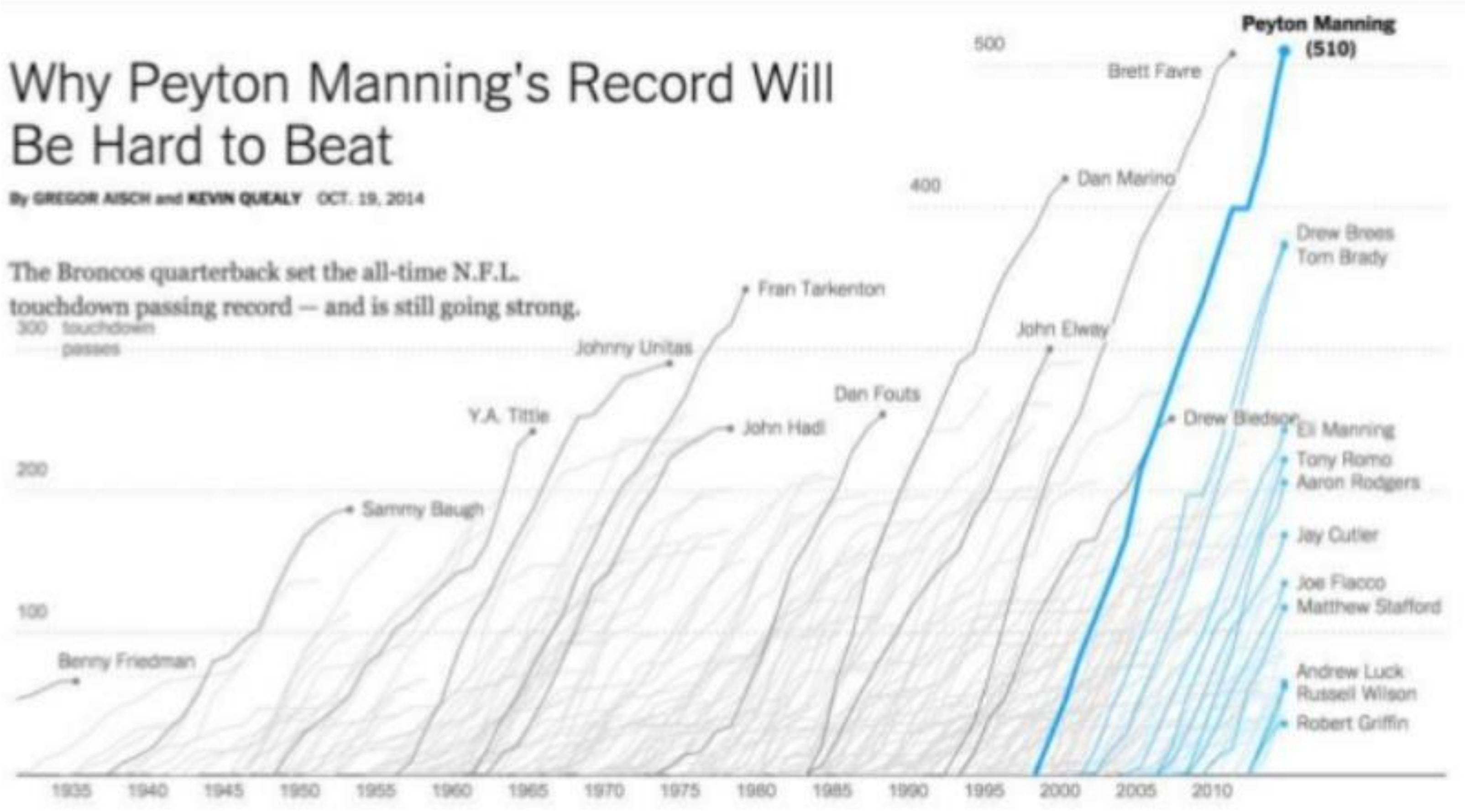
Color Coding



Large areas = low saturation

Small areas = high saturation

Good use of color



Problematic use of color

the values change smoothly, but the colors do not

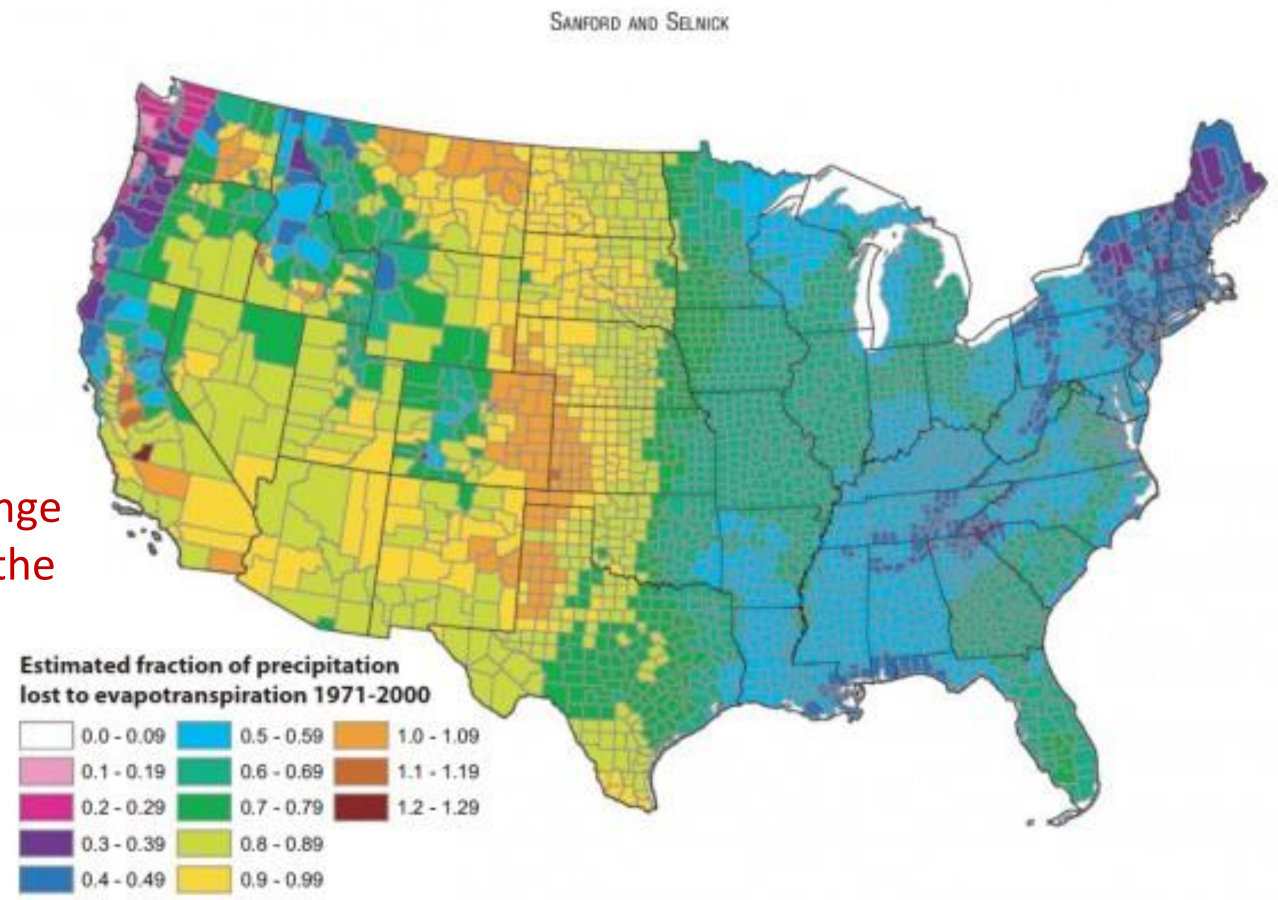


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

guidelines

- Use color only when necessary (avoid unnecessary textures and colors)
- Saturated colors for small areas, labels
- Less saturated colors for large areas, backgrounds
- Do not pick colors in an *ad hoc* way: use tools like ColorBrewer 2.0
- avoid producing plots that confuse people who are color blind

Pre-Attentive Processing

- How many 3s ?

08028085080830802809850–802808
5678472988 72ty4582020947577200
217898438 90r455790456099272188
897594797902855892594573979209

Pre-Attentive Processing (Pops out)

- How many 3s ?

080280850808308028098 50–802808

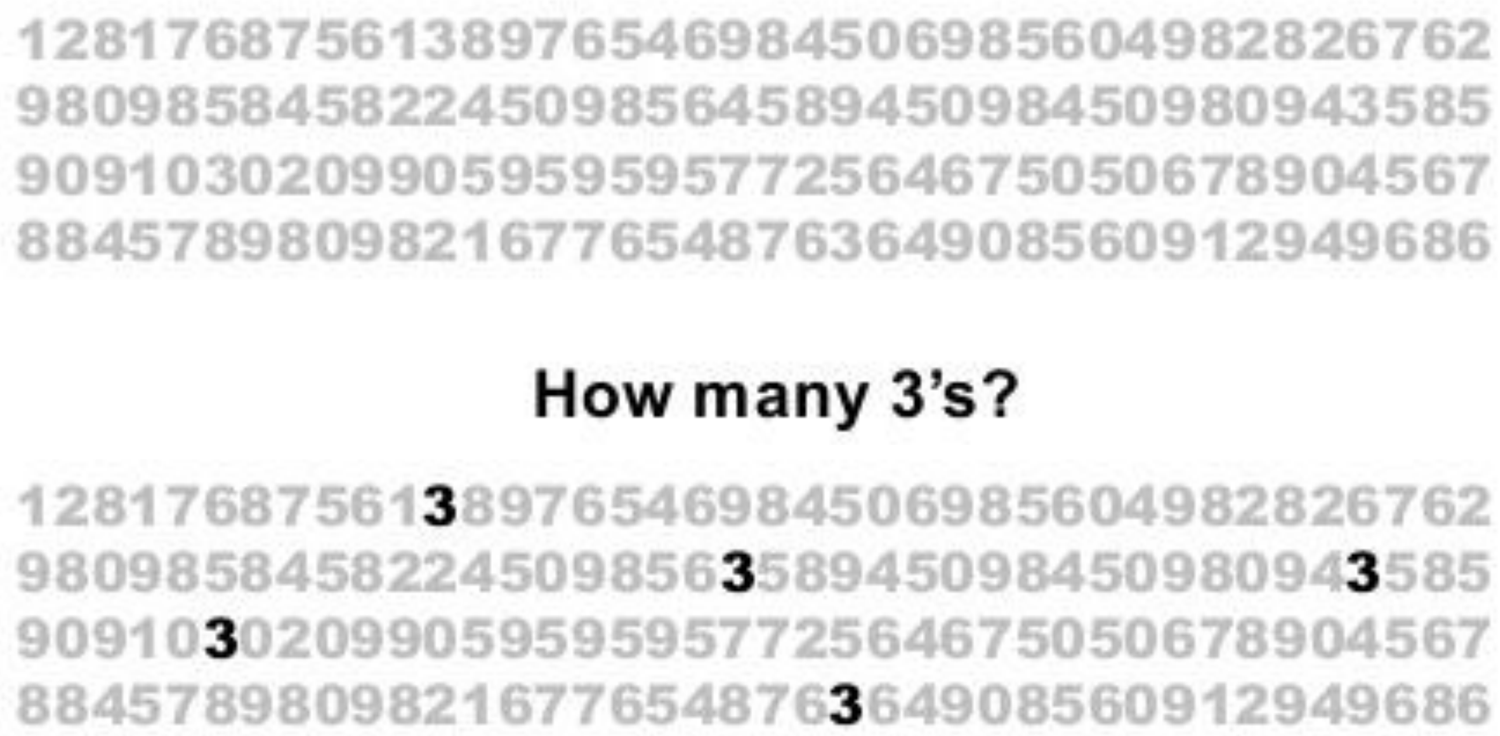
567847298872ty4582020947577200

2178984 3890r4557904560992721 88

89759479790285589259457 3979209

Pre-Attentive Processing (Pops out)

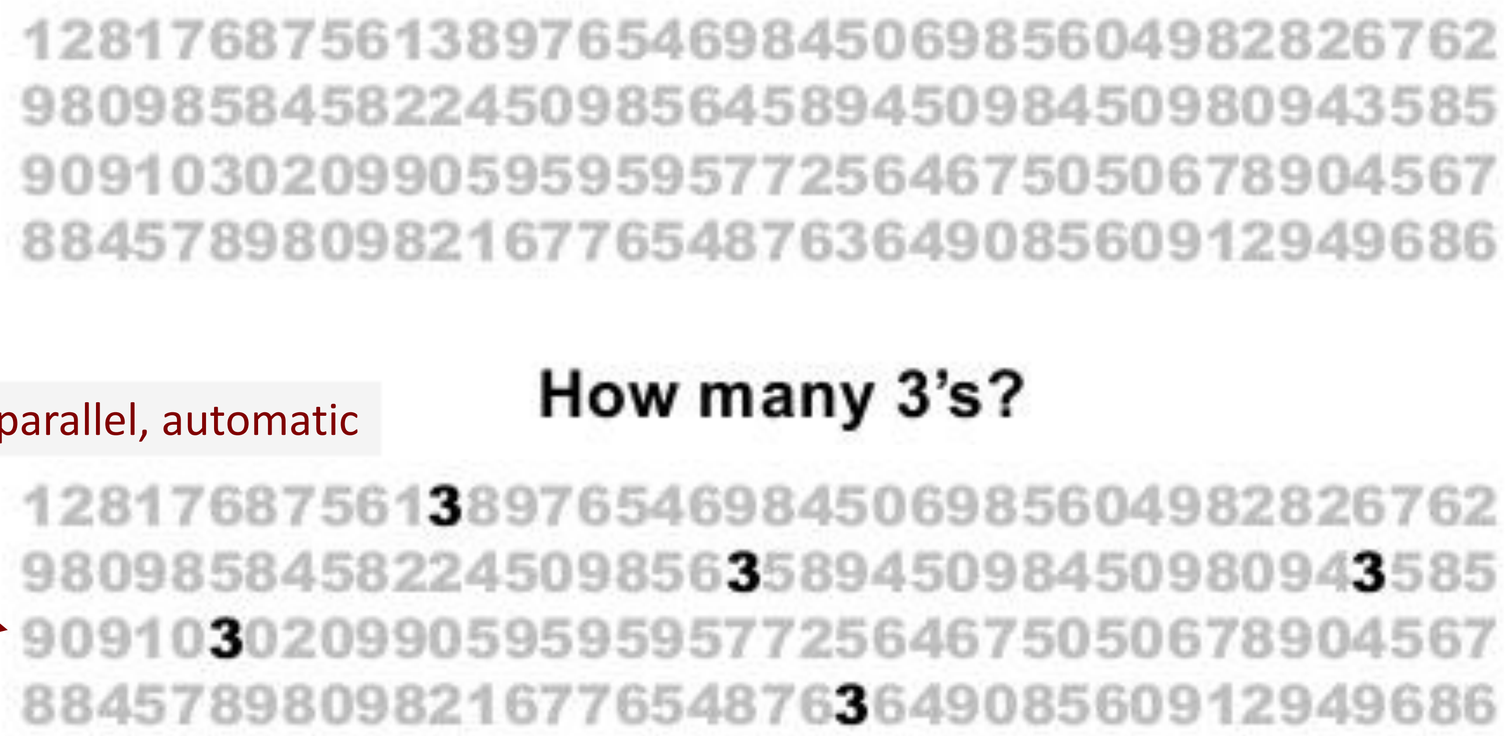
- How many 3s?



- How many 3s?

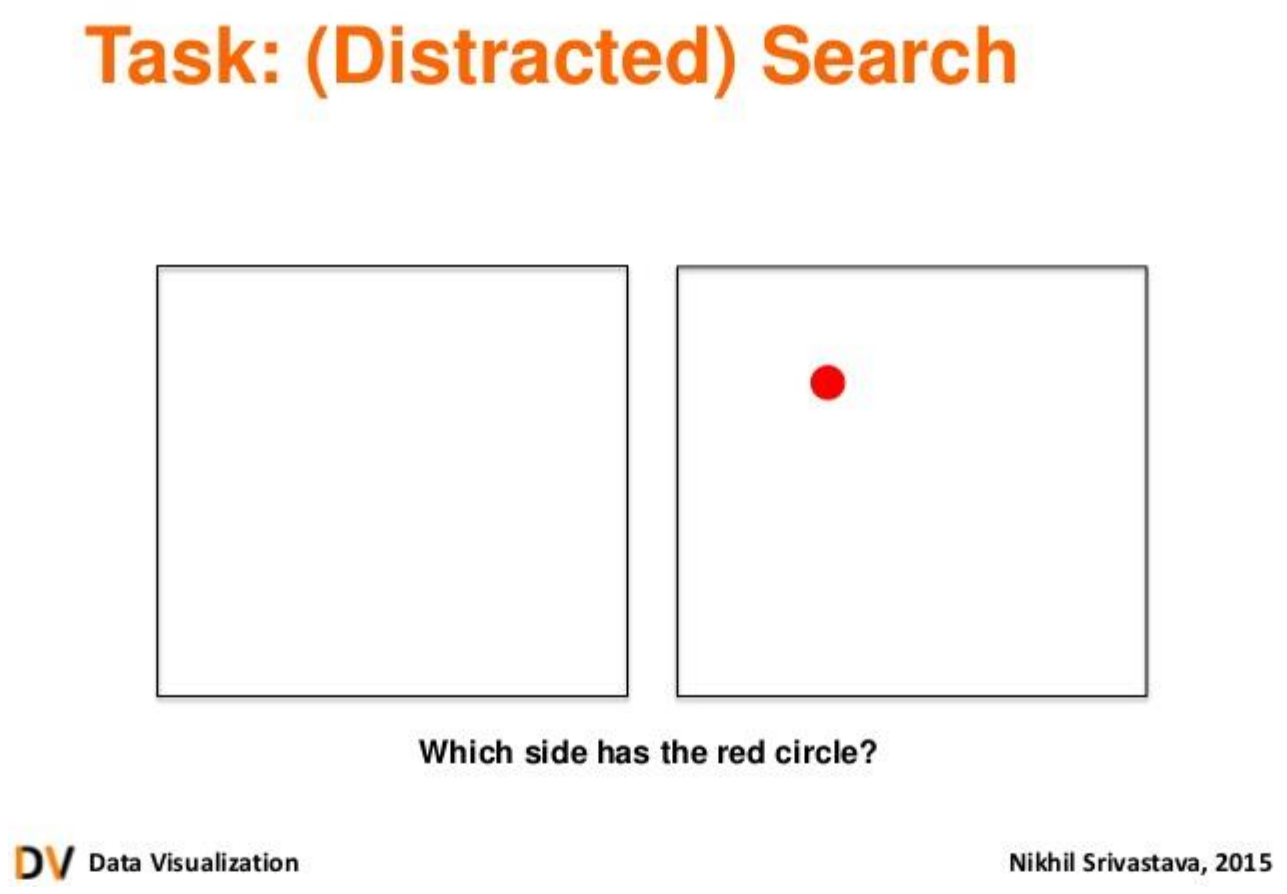
Slow, sequential, conscious

Rapid, parallel, automatic



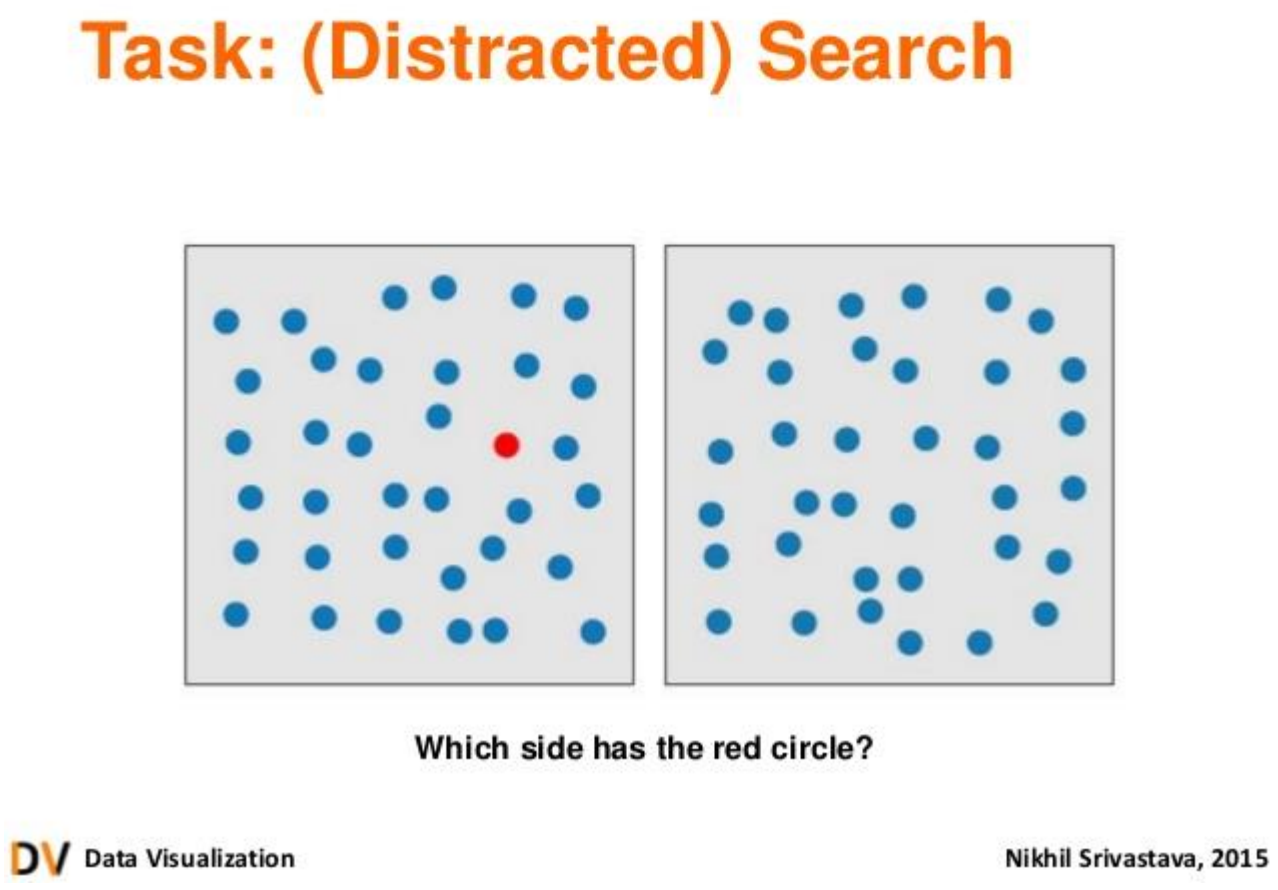
Task: (distracted) search

- Which side has the red circle?



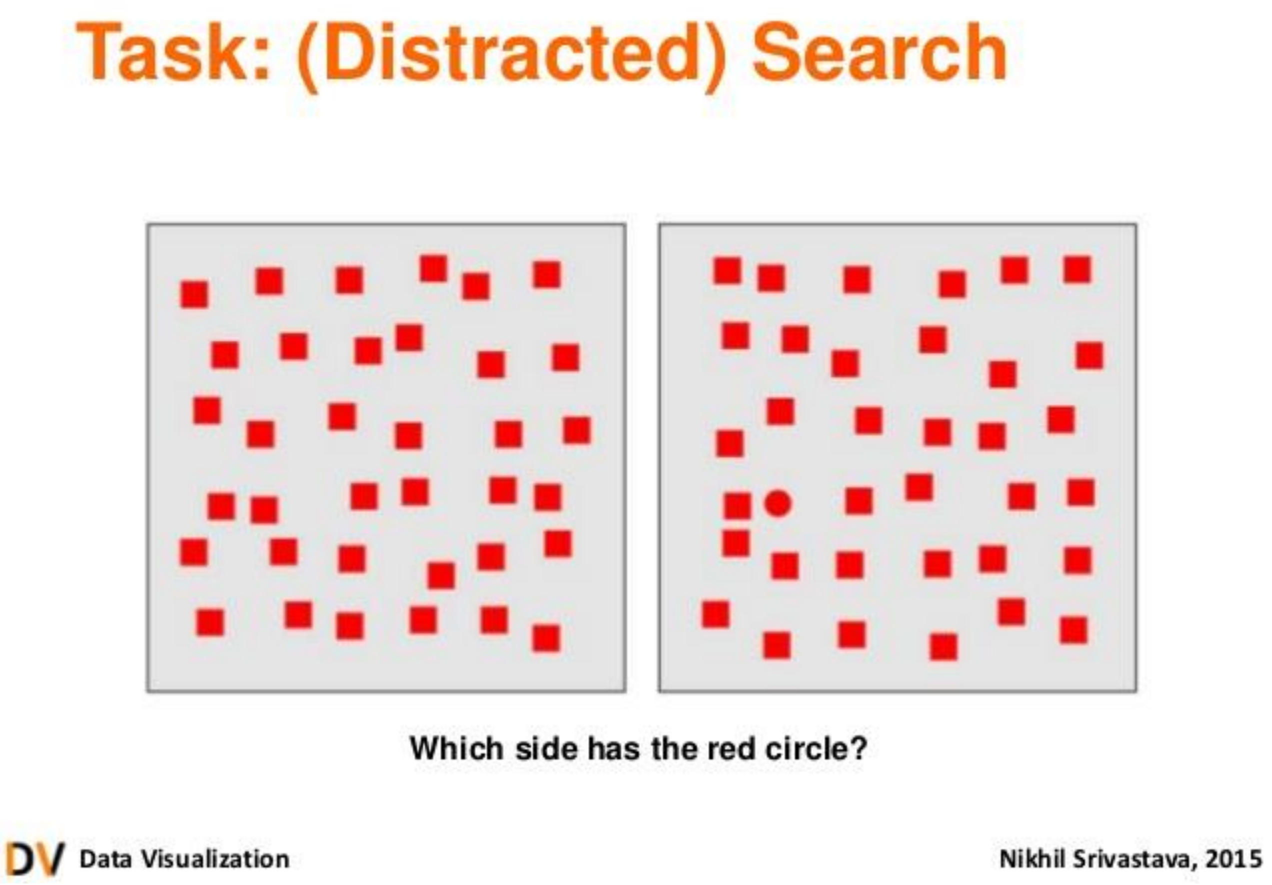
Task: (distracted) search

- Which side has the red circle?



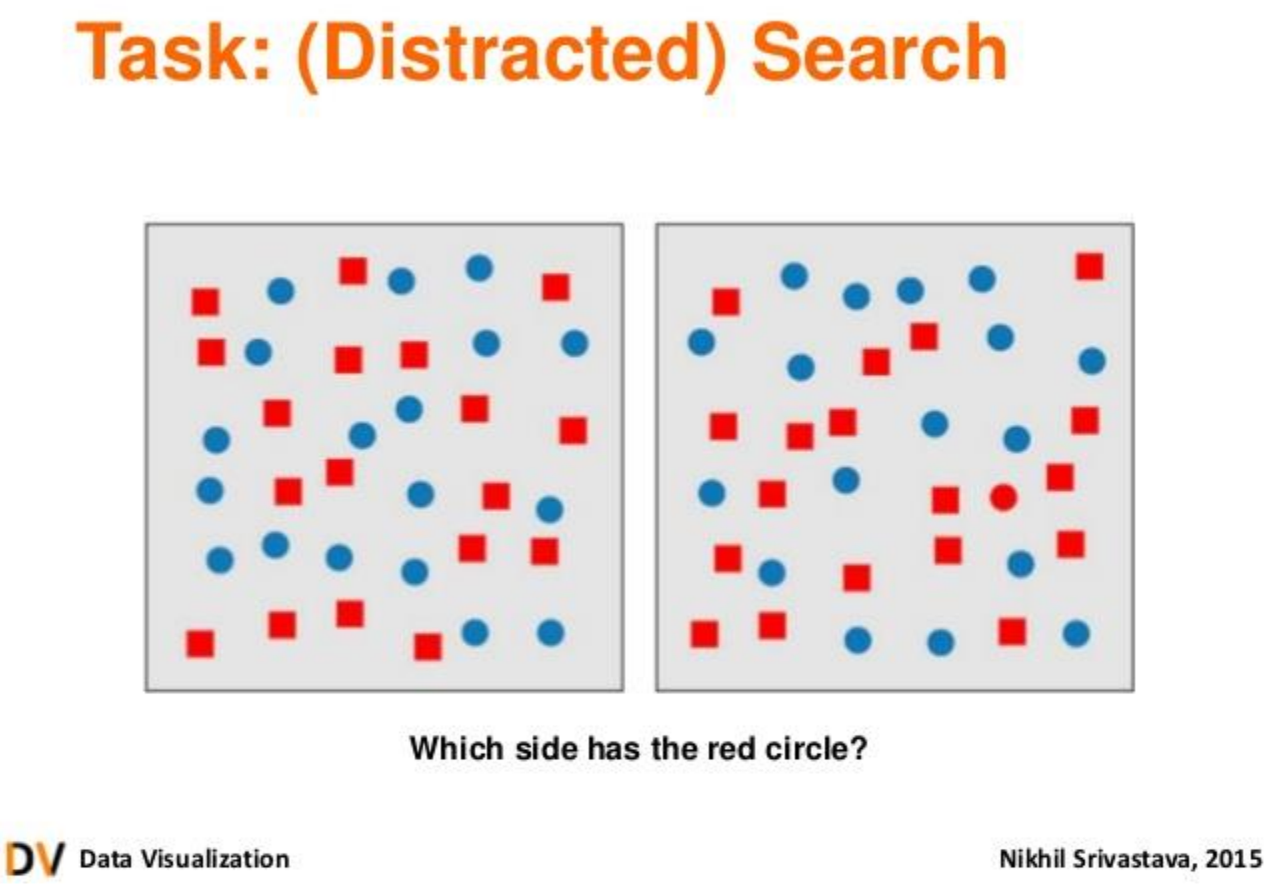
Task: (distracted) search

- Which side has the red circle?



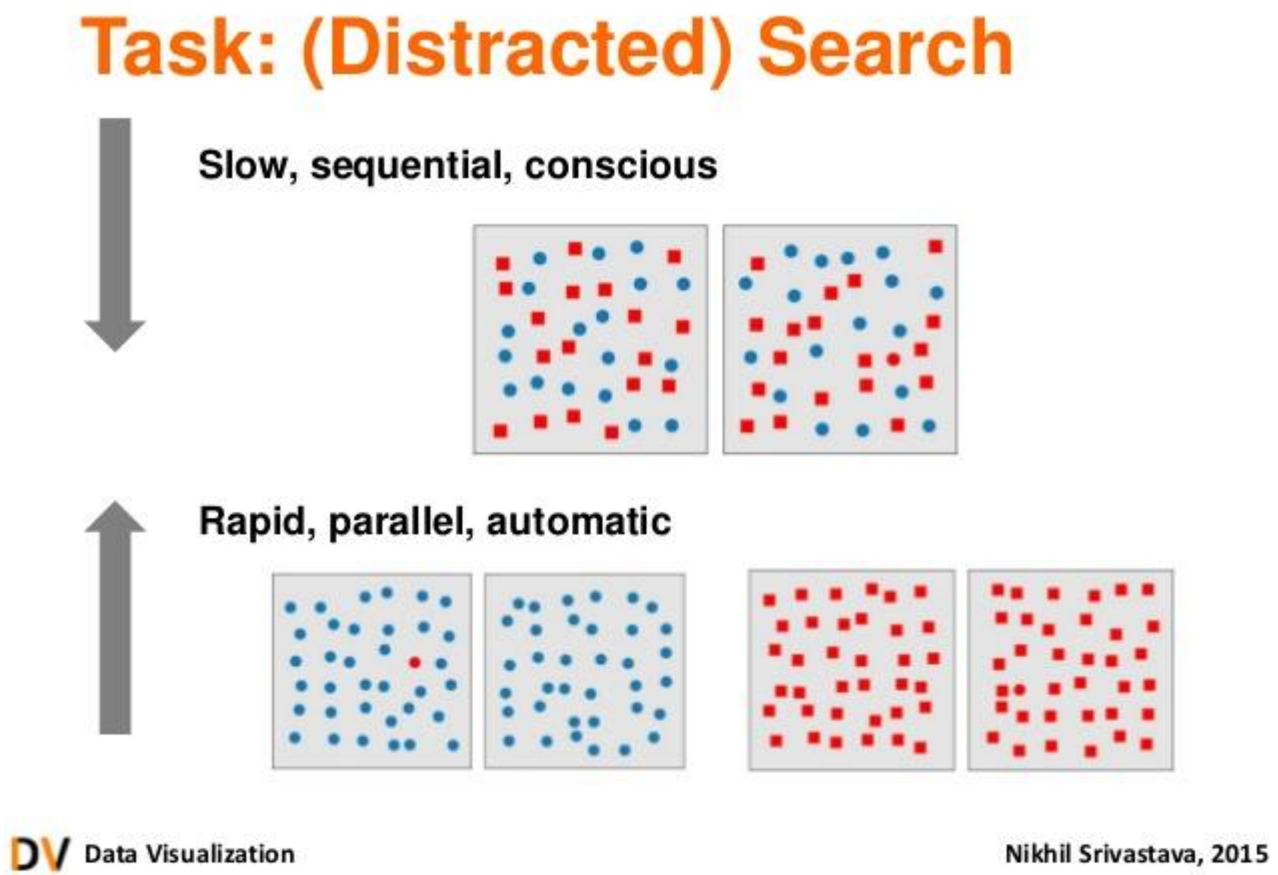
Task: (distracted) search

- Which side has the red circle?



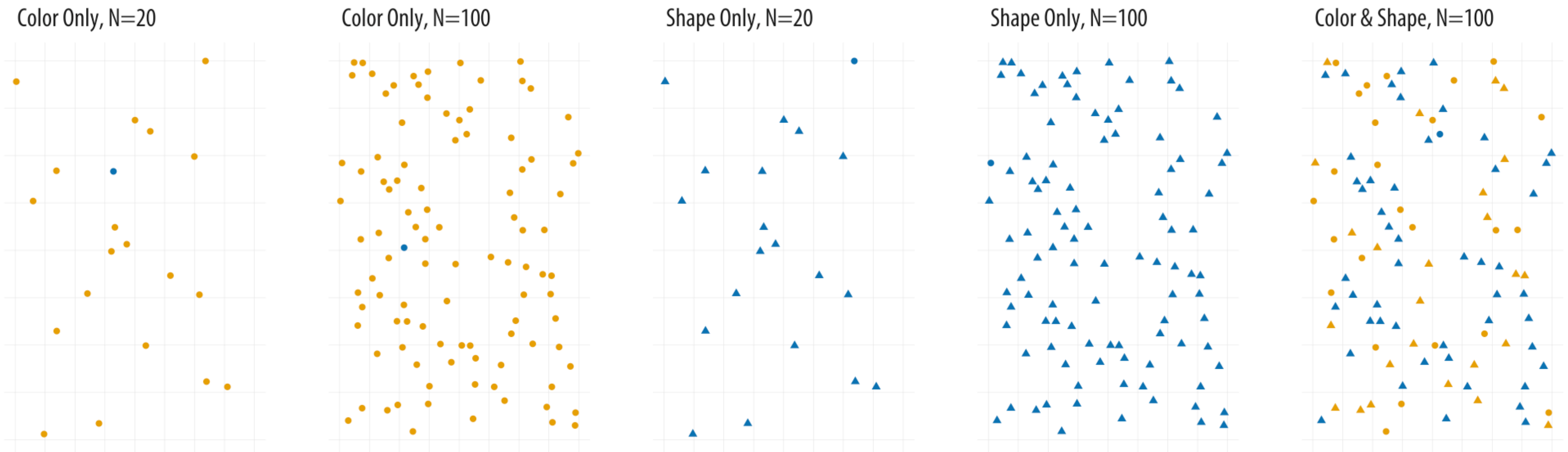
Task: (distracted) search

- Which side has the red circle?



Task: (distracted) search

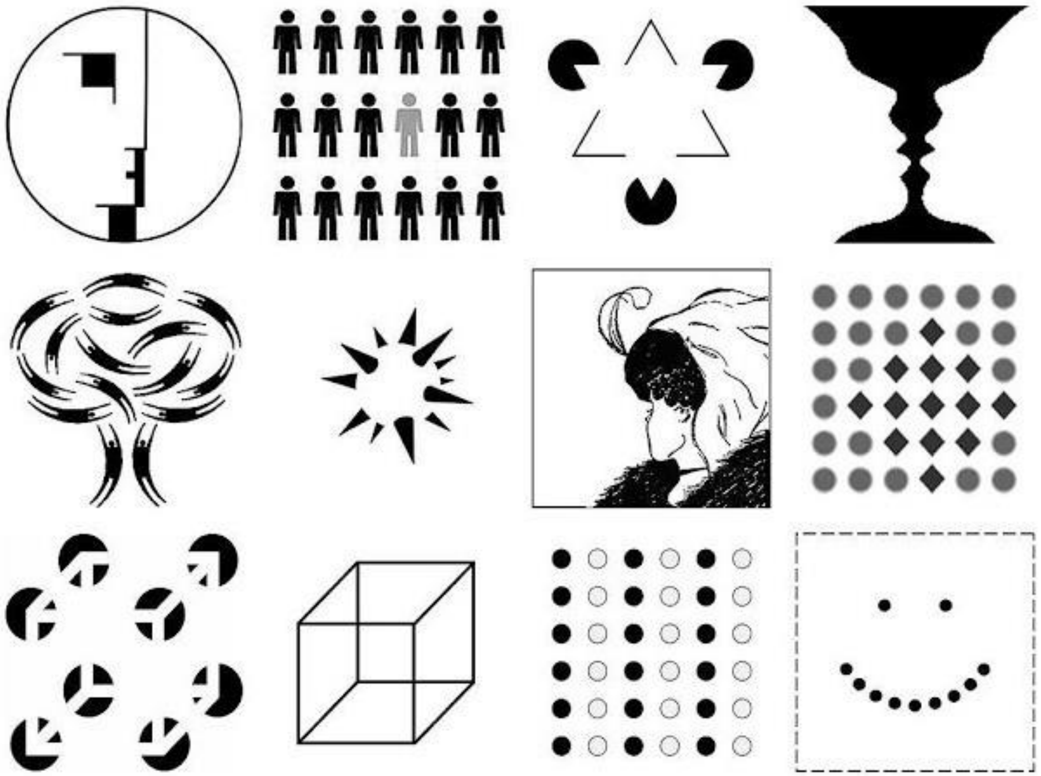
- Searching for the blue circle becomes progressively harder...
- Shape and color are two distinct *channels:* pop-out on **the color channel is stronger than** it is on **the shape channel**



Visual perception: Gestalt principles

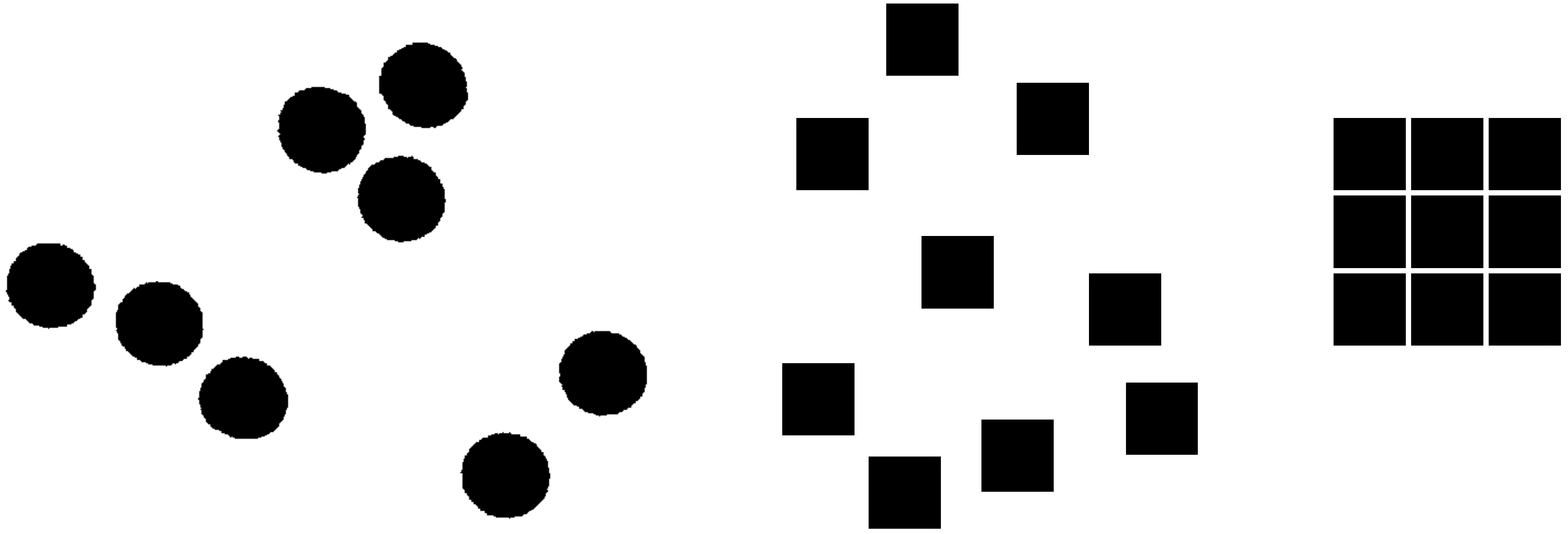
1. Proximity
2. Similarity
3. Continuity
4. Closure
5. Symmetry

6. Figure/ground
7. Common fate



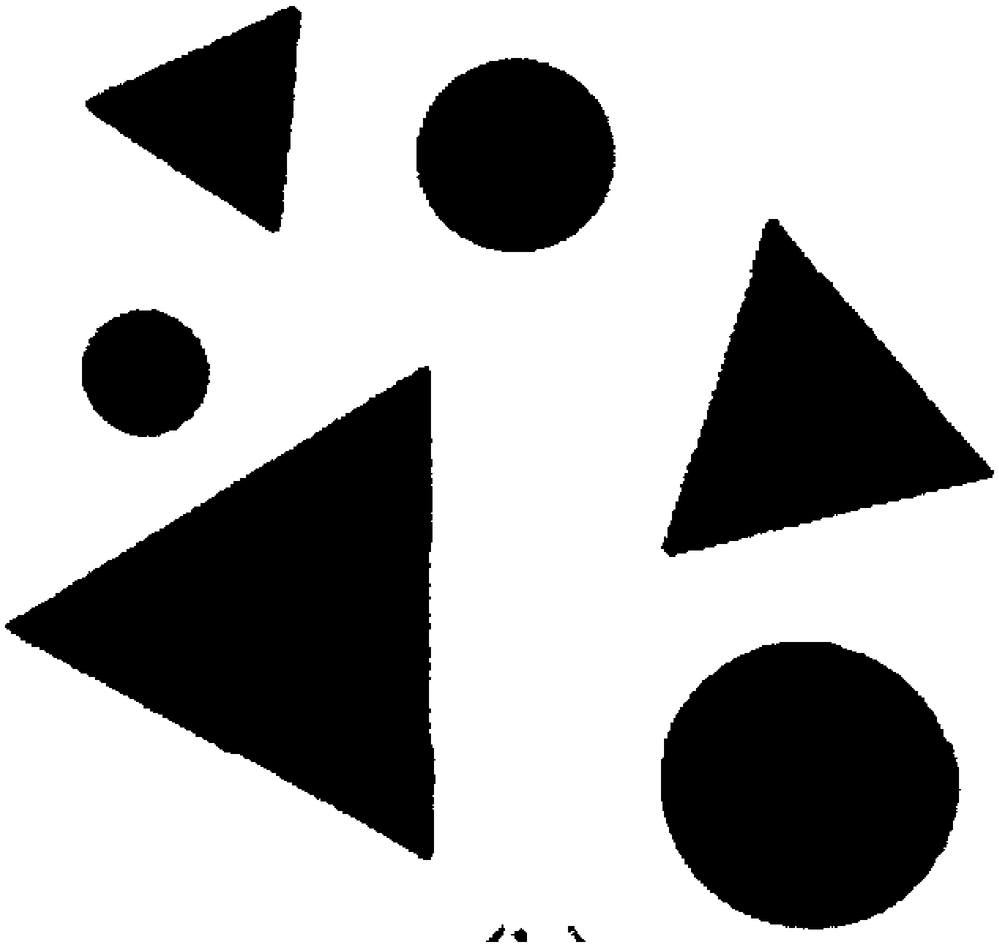
Visual perception: Gestalt principles

- *Proximity*: Things that are spatially near to one another seem to be related.



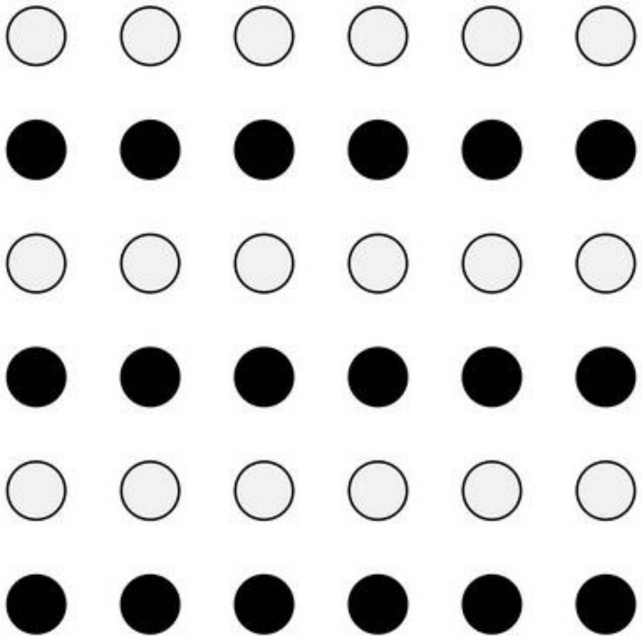
Visual perception: Gestalt principles

- *Similarity*: Things that look alike seem to be related.



Visual perception: Gestalt principles

- *Similarity*: Things that look alike seem to be related.

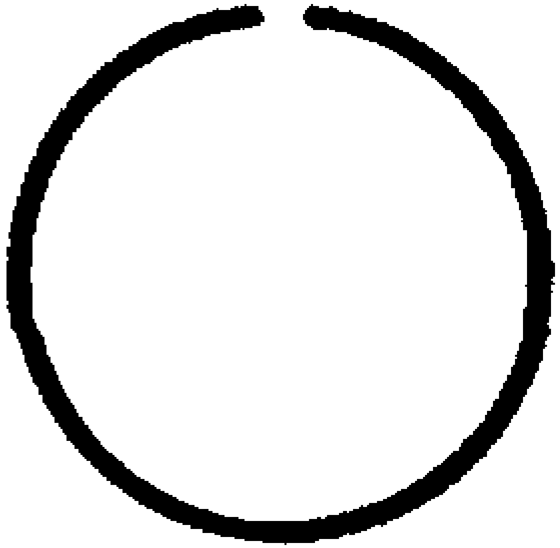


Visual perception: Gestalt principles

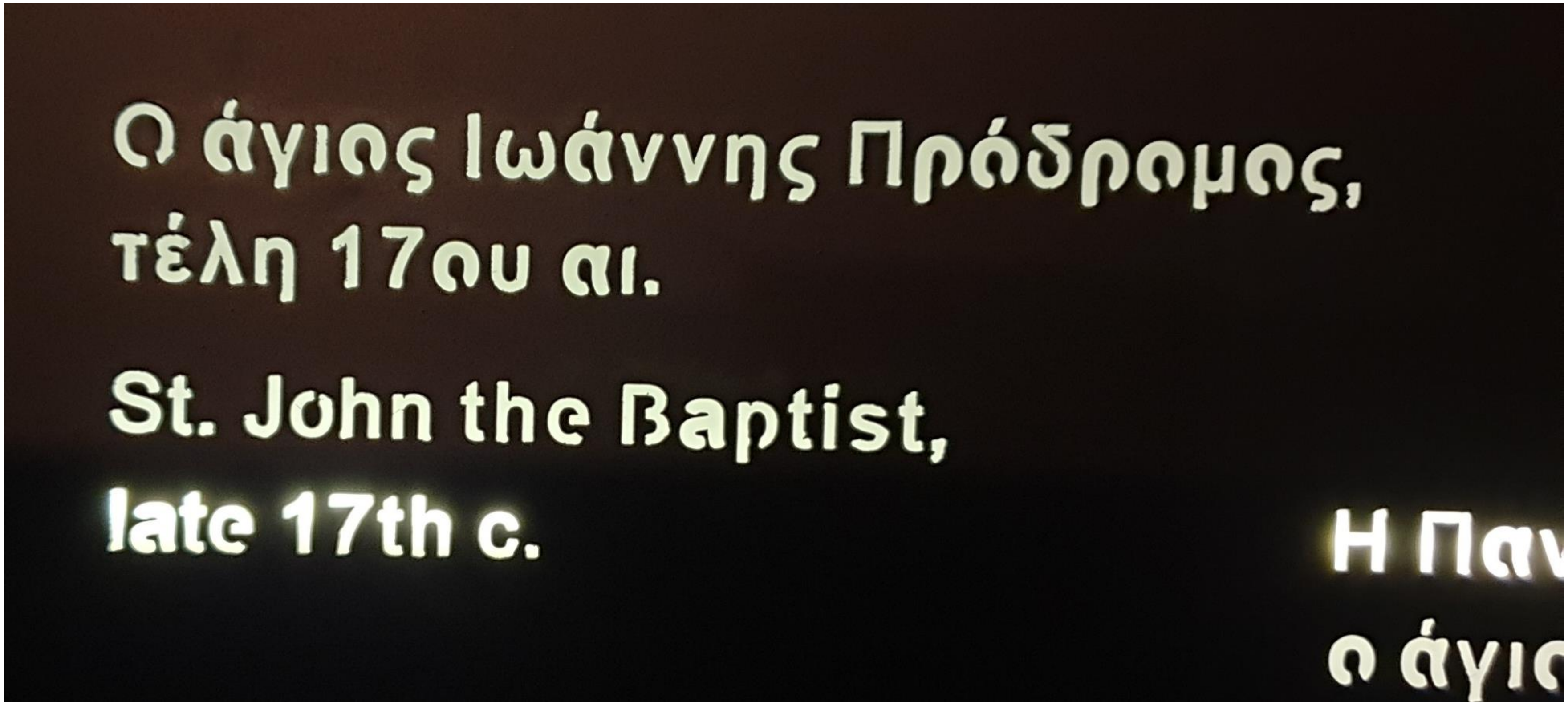
- *Similarity*: Things that look alike seem to be related.

Visual perception: Gestalt principles

- *Closure*: Incomplete shapes are perceived as complete.



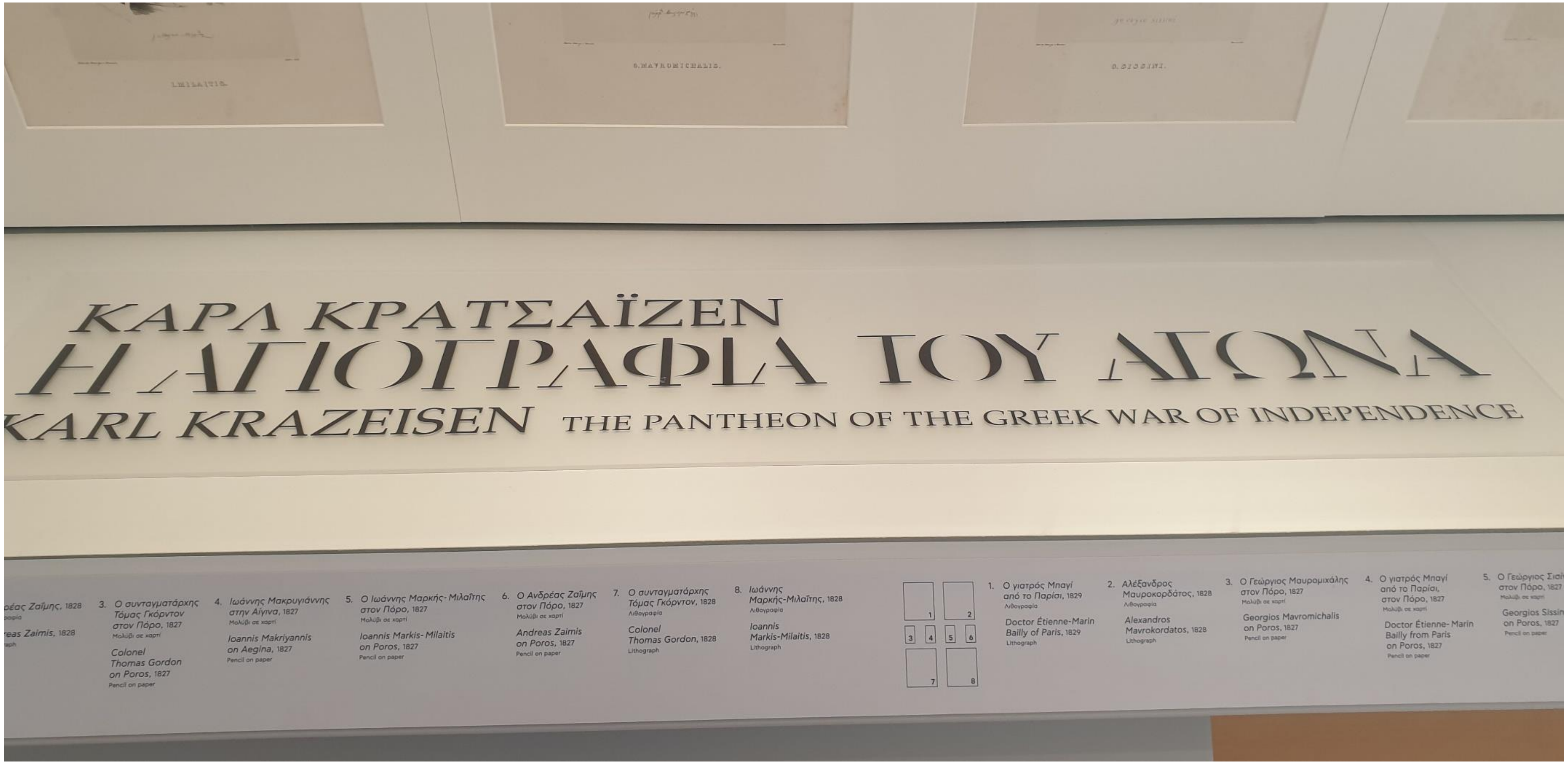
- *Closure*: Incomplete shapes are perceived as complete.



Γραμματοσειρά σε πινακίδα στο Ιστορικό Μουσείο Κρήτης, 2019

Visual perception: Gestalt principles

- *Closure*: Incomplete shapes are perceived as complete.

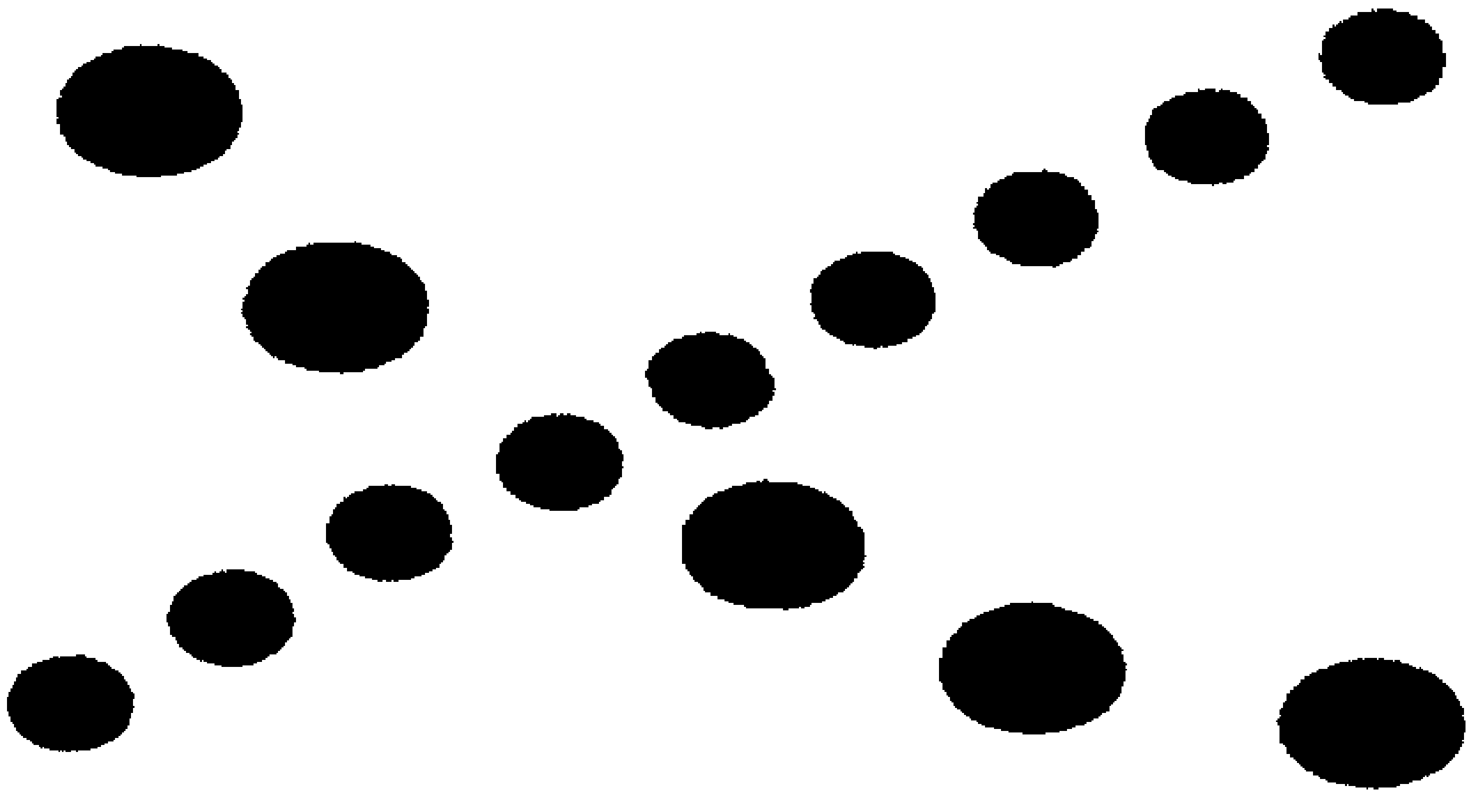


Γραμματοσειρά σε πινακίδα στην Εθνική Πινακοθήκη, 2022

- *Closure*: Incomplete shapes are perceived as complete.

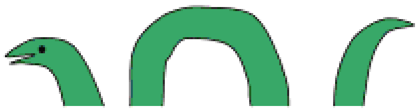
Visual perception: Gestalt principles

- *Continuity*: Partially hidden objects are completed into familiar shapes.

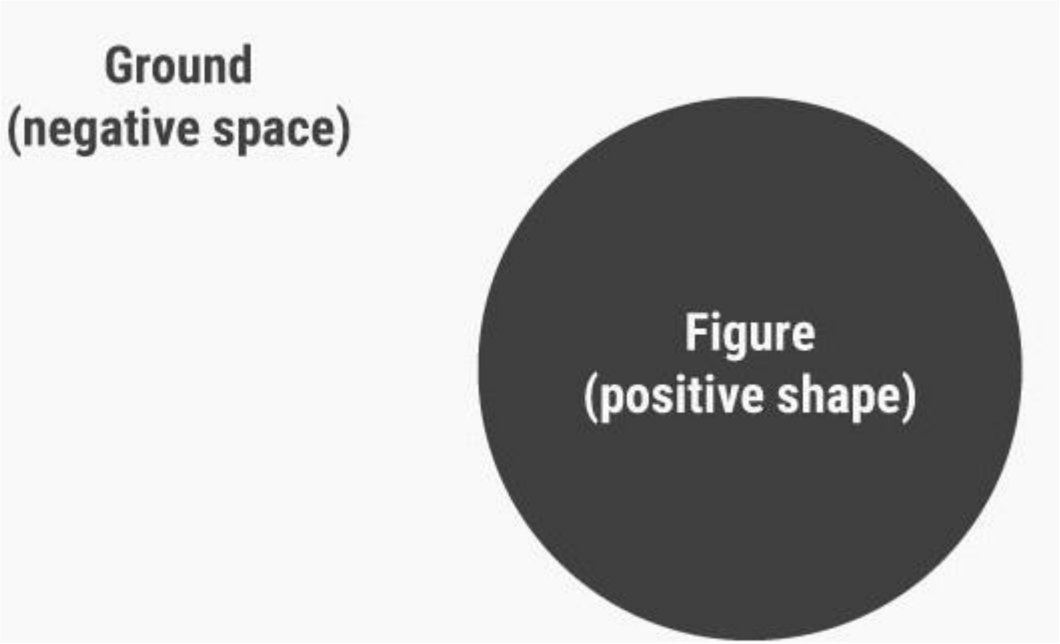


Visual perception: Gestalt principles

- *Continuity*: Partially hidden objects are completed into familiar shapes.

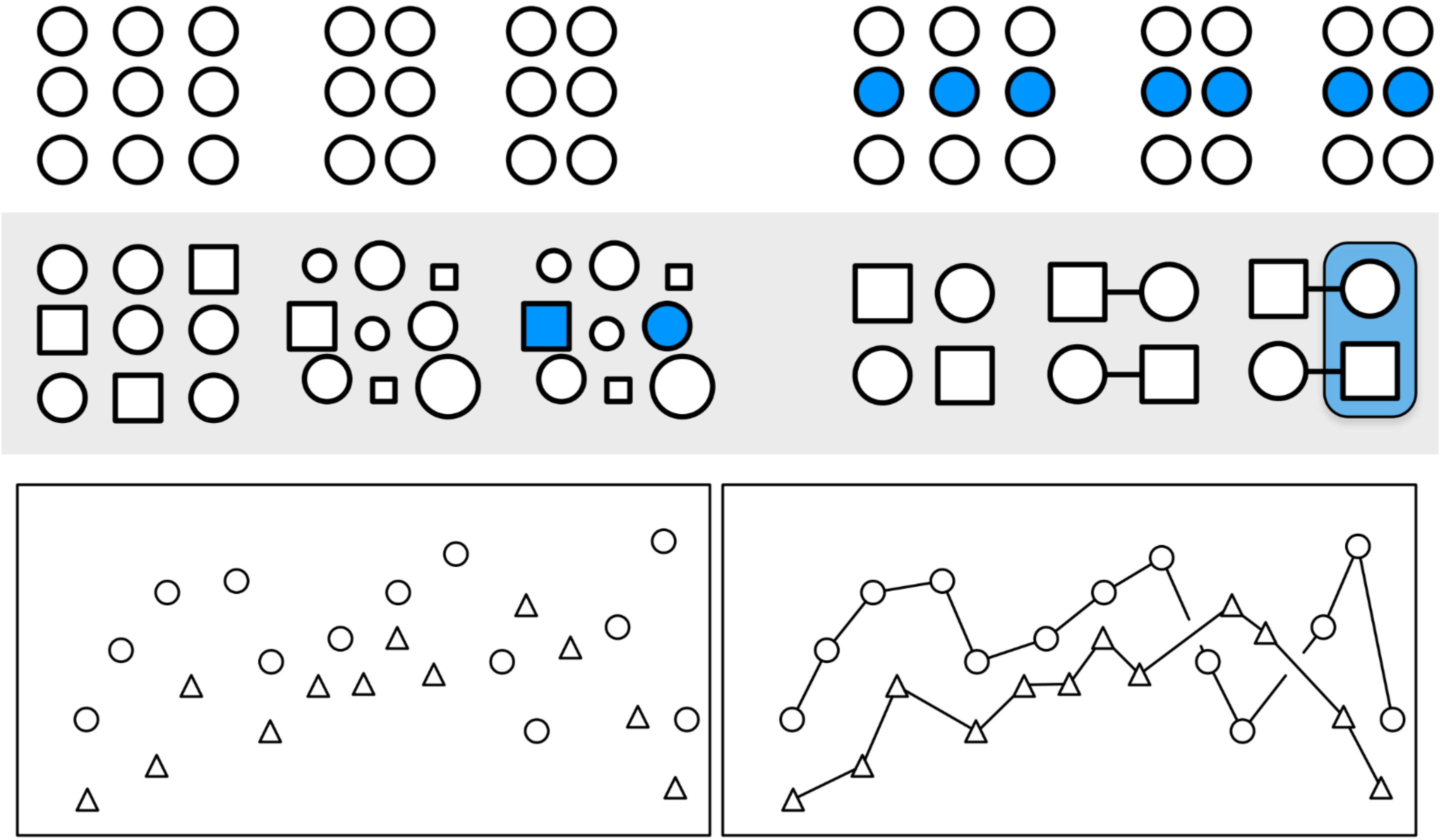


- *Figure and Ground*: Visual elements are taken to be either in the foreground or the background.



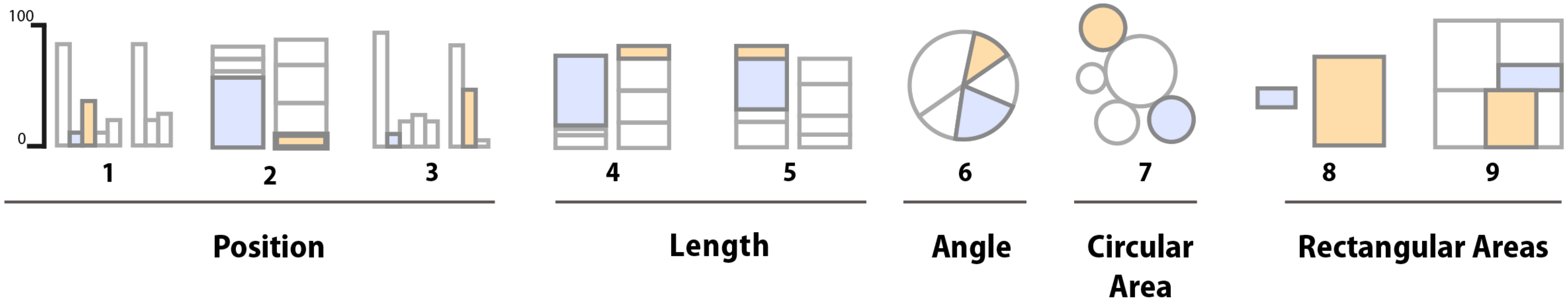
Visual perception: Gestalt principles

- We look for structure all the time



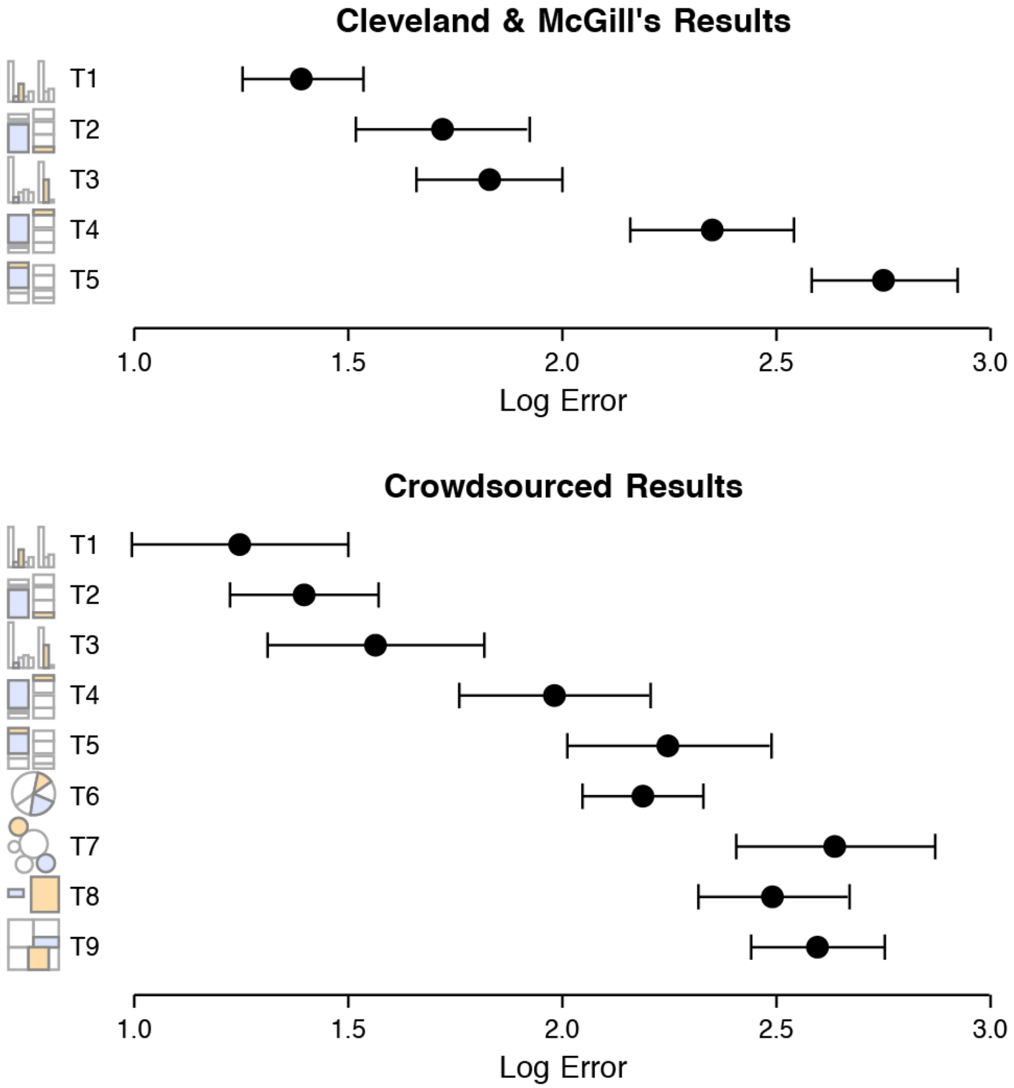
Interpreting and understanding graphs

- Schematic representation of basic perceptual tasks for nine chart types.
- Participants were asked to make comparisons of highlighted portions of each chart type, and say which was smaller.



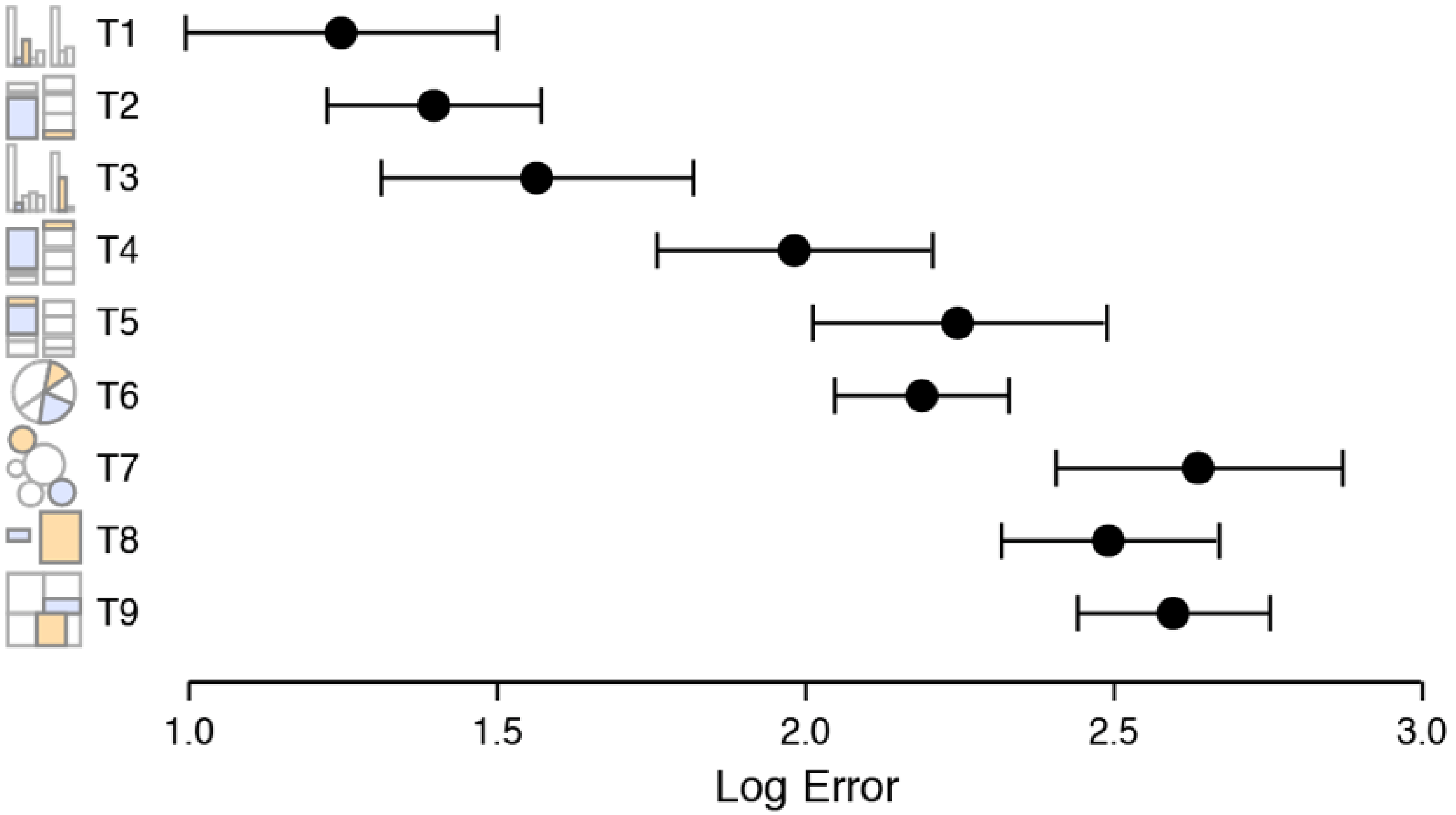
Cleveland & McGill, 1984, 1987
Heer and Bostock

what types of plots do we understand best?



what types of plots do we understand best?

performance worsens substantially as we move away from comparison on a common scale to length-based comparisons to angles and finally areas. Area comparisons perform even worse than the (justifiably) much-maligned pie chart.



Important for Design of Visualizations

- Do not
  - Put too much in two axes (produce separate plots instead)
  - Truncate axes
  - Use 3D unnecessarily
- Do
  - Show the data
  - Be as clear as possible
  - Let the data tell the story

Check out

- Bach, M. 136 Optical Illusions & Visual Phenomena
- Visual Complexity
- Perceptual Edge
- Edward Tufte website

References & further reading

- Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press.
- Ware, C. (2013). *Information Visualization: Perception for Design* (3rd Editio). Morgan Kaufmann.
- Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science*, *21*(2), 179–217. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog2102_3
- Current approaches to change blindness Daniel J. Simons. Visual Cognition 7, 1/2/3 (2000), 1-15.
- Semiology of Graphics, Jacques Bertin, Gauthier-Villars 1967, EHESS 1998

Thank you!

mroussou@di.uoa.gr

http://eclass.uoa.gr/courses/DI411/