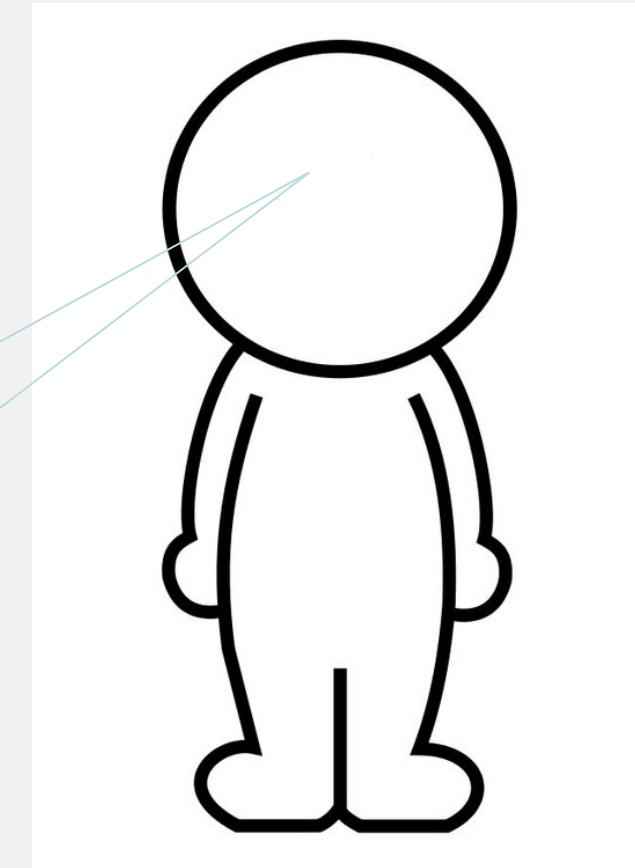
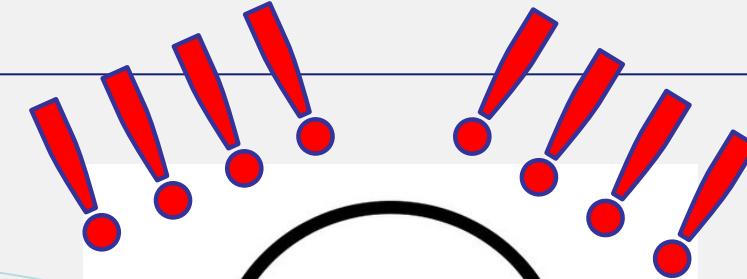




In the past lecture there was a ball coming...





What's that (unconscious processing) that allowing sensation and perception?



Why optical illusions??

There is no just processing the light and a straight coding of the light.

There are biological effects (inhibition) and brain interpretation.

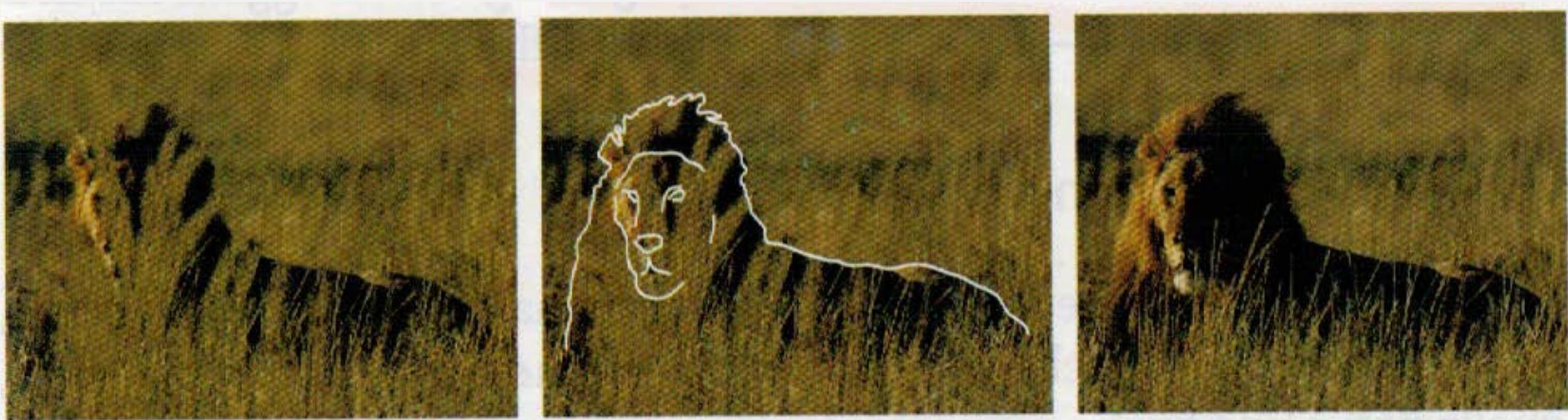
Brains considers also experiences, memories, and cognitive information

(unconscious processing)



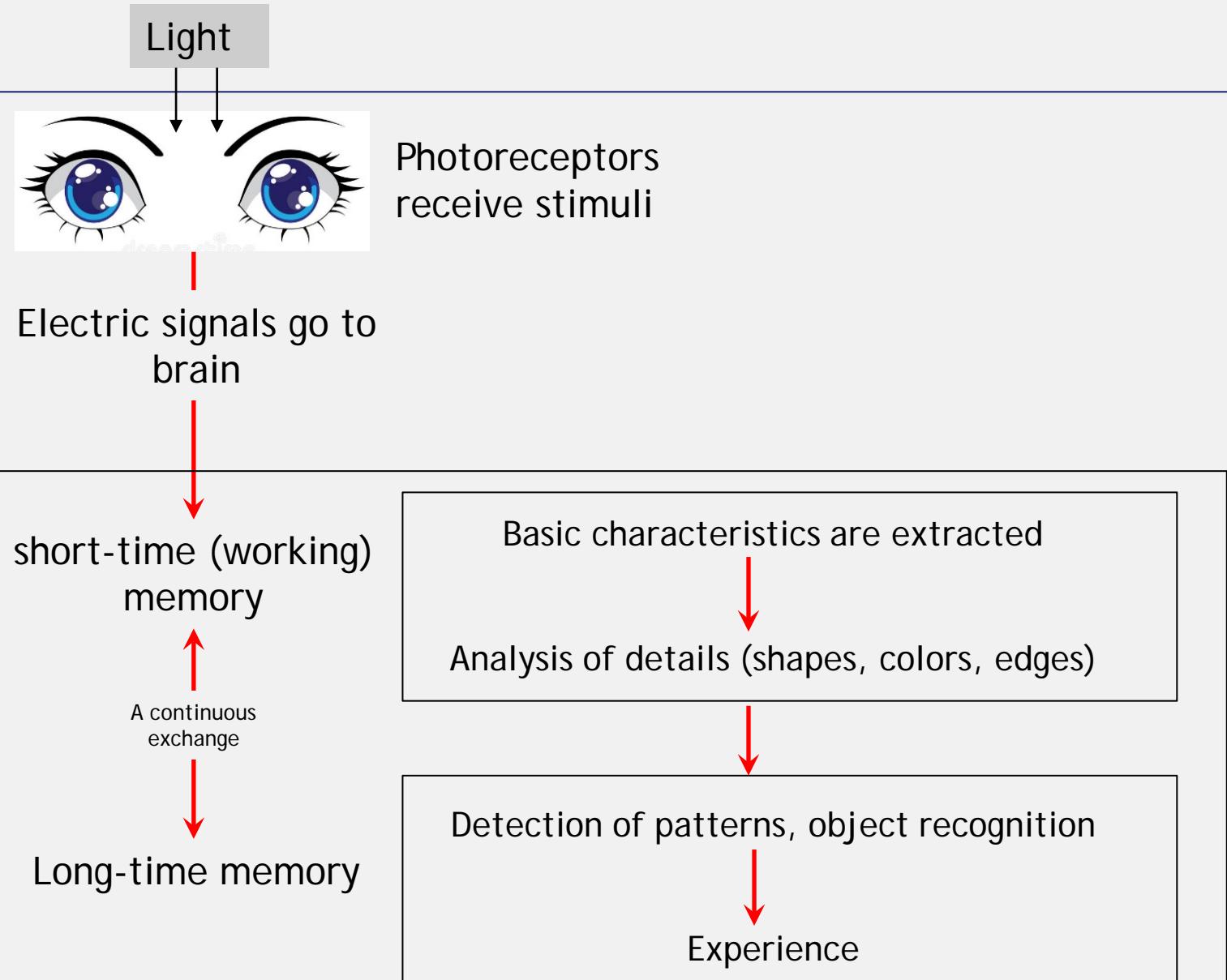
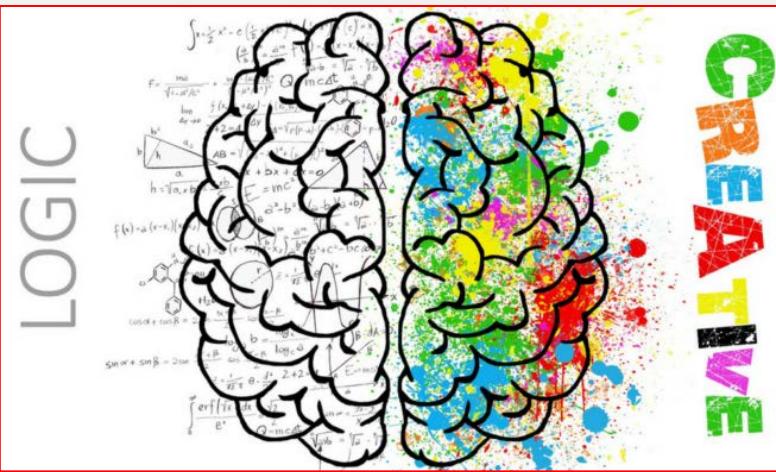


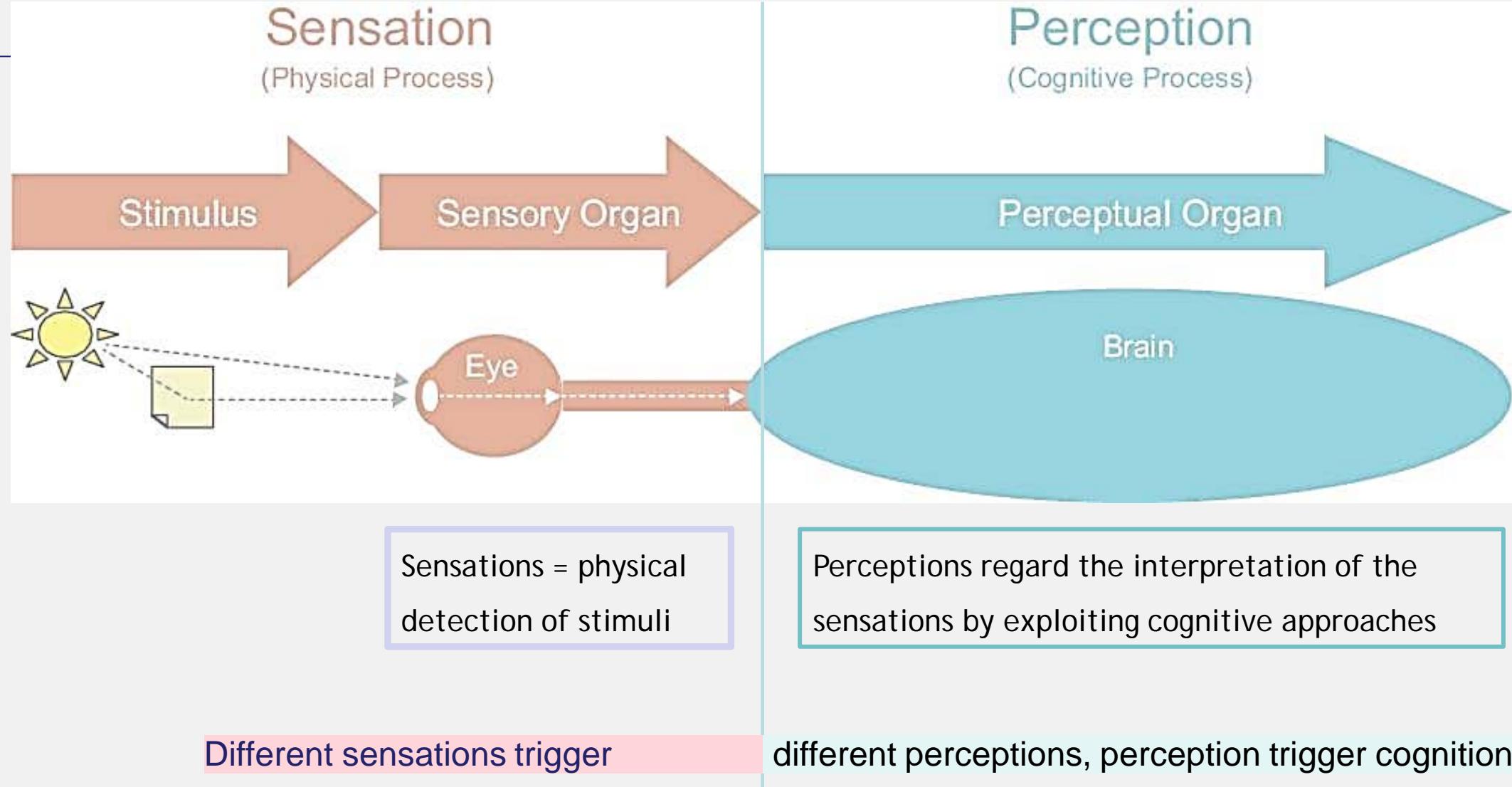
How do we see complex patterns? conscious processing

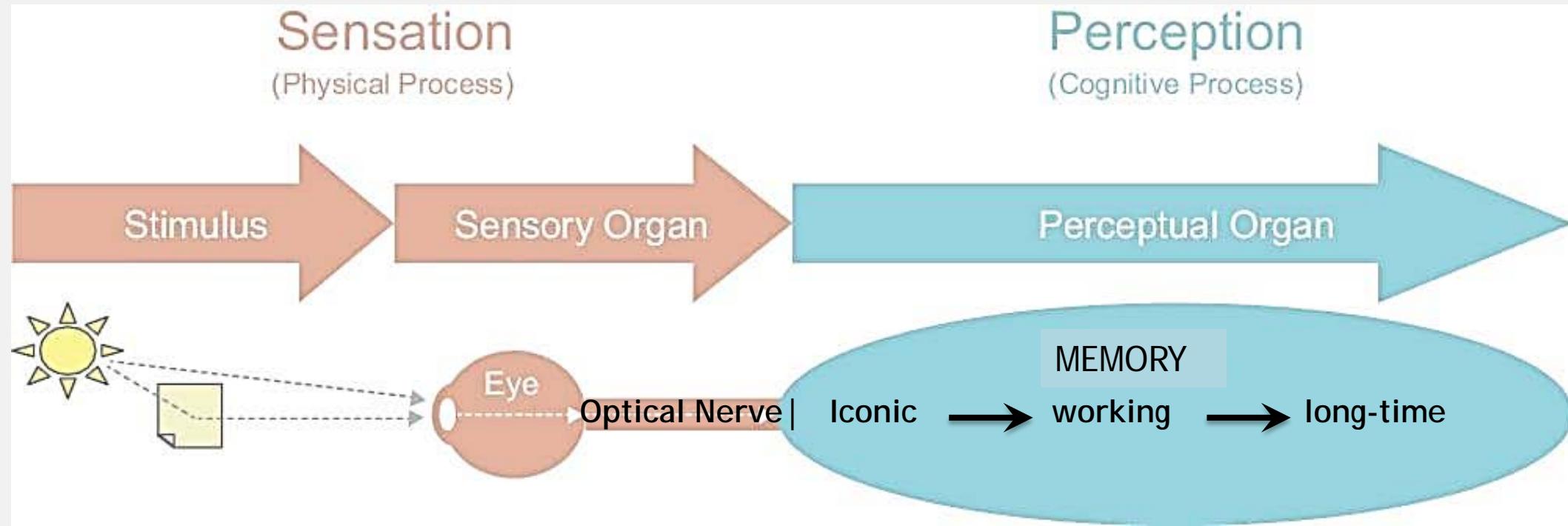


Brain uses memories and says you need to start running, because there is a lion. Brain sees the lion, though it is hidden in the grass









Memory actively builds world representations based on external sensations and on the past experiences.

Tomei, L. (2017). Psicologia della memoria in Aquilar, F., Pugliese, M. (2017). Condividere i ricordi. Psicoterapia cognitiva e funzioni della memoria.
Franco Angeli Editore



Stimulus

Iconic Memory (visual sensory register - Stephen Few):

Stimuli stay there for 1 sec-1 min

Performs the unconscious, *preattentive processing*

finds already known elements or most informative, high frequency, atomic elements: colors, edges, positions of points in space.

Patterns and elements which are further analyzed by the working memory and then by the short-time memory

Visualizations should exploit preattentive stimuli to hit the iconic memory and the following memories!



GAME

I'm reading strings with numbers.

Trye to repeat in your mind the numbers without looking at the slide or writing them down

Example:

If I say "3-8-6", in your mind you repeat: "3-8-6"

I WON'T REPEAT THE NUMBERS.

READY????

Adults can store 7 ± 2 elements

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

The Working memory has a limited space!

We can't store numbers, but we can store pictures (positions of points)!

Could you repeat the third (short)
sequence I said?

.....

As a cache, the working memory is emptied any time new data enters.
To recall the numbers, we should think about them, to bring them in
the long-term memory





Effective visualizations are visualizations that use preattentive attributes hitting the preattentive memory.

Long-term memory is important for the retrieved knowledge:

- I see a red color in a graph with the pre-attentive memory, the long-time memory tells me that color means fear
- The working memory recognizes the shape of a graph, and the long-term memory tells how to interpret the graph





PREATTENTIVE ATTRIBUTES

Quanti 5 sono contenuti nella sequenza qui sotto?

987349790275647902894728624092406037070570279072
808208029007302501270237008374082078720272007083
247802602703793775709707377970667462097094702780
927979709723097230979592750927279798734972608027





E qui?

98734979027**5**647902894728624092406037070**5**70279072
803208029007302**5**01270237008374082078720272007083
24780260270379377**5**709707377970667462097094702780
927979709723097230979**5**927**5**0927279798734972608027

98734979027**5**647902894728624092406037070**5**70279072
803208029007302**5**01270237008374082078720272007083
24780260270379377**5**709707377970667462097094702780
927979709723097230979**5**927**5**0927279798734972608027



Colin Ware organization of pre-attentive attributes

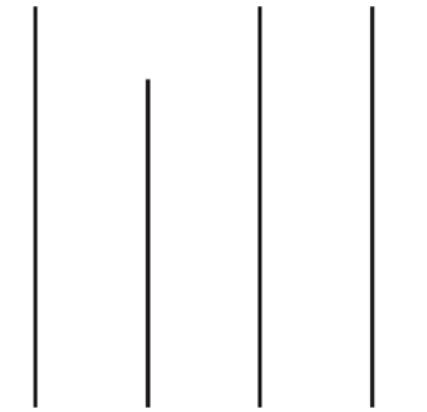
Type	Attribute (variations of)
Form	Length = distance along the dominant dimension Width = distance along the secondary dimension Orientation Size Shape Enclosure
Color	Hue Intensity
Spatial Position	2-D Position



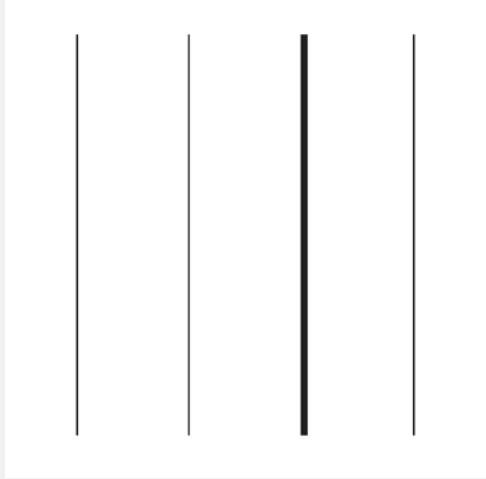


FORM

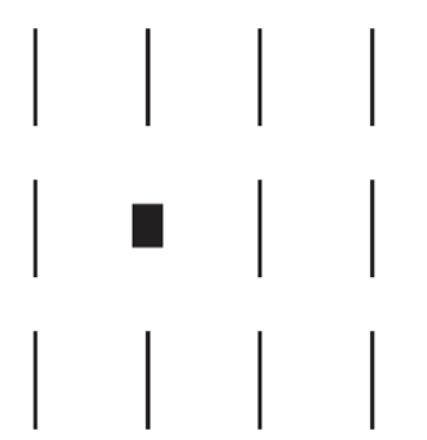
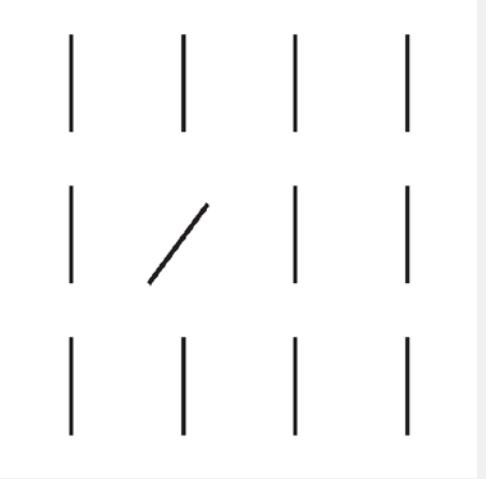
length



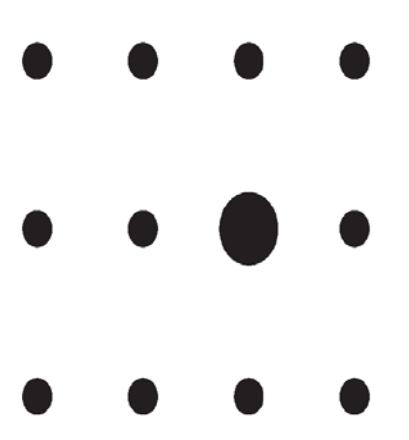
width



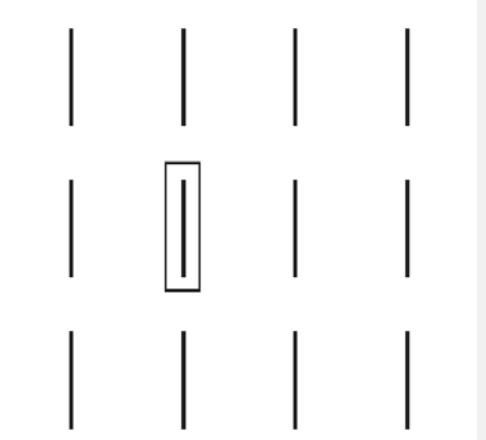
orientation



shape



size



Enclosure

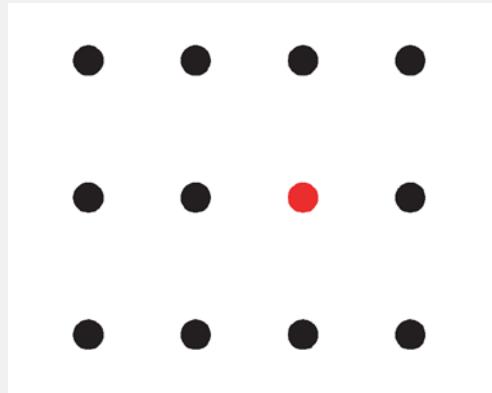




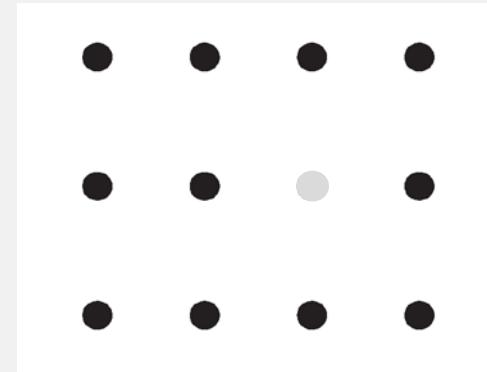
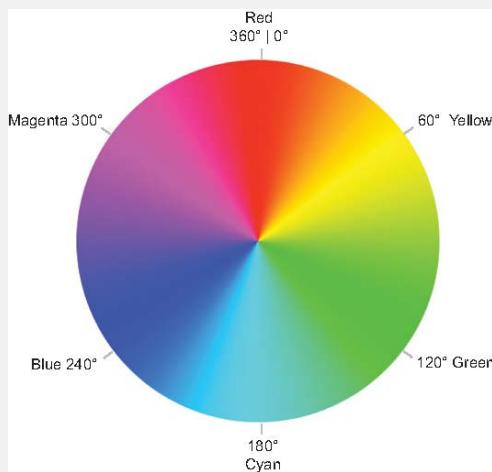
Colors

Hue is the color as we think about it

Intensity influences saturation and lightness



HUE



INTENSITY

saturation



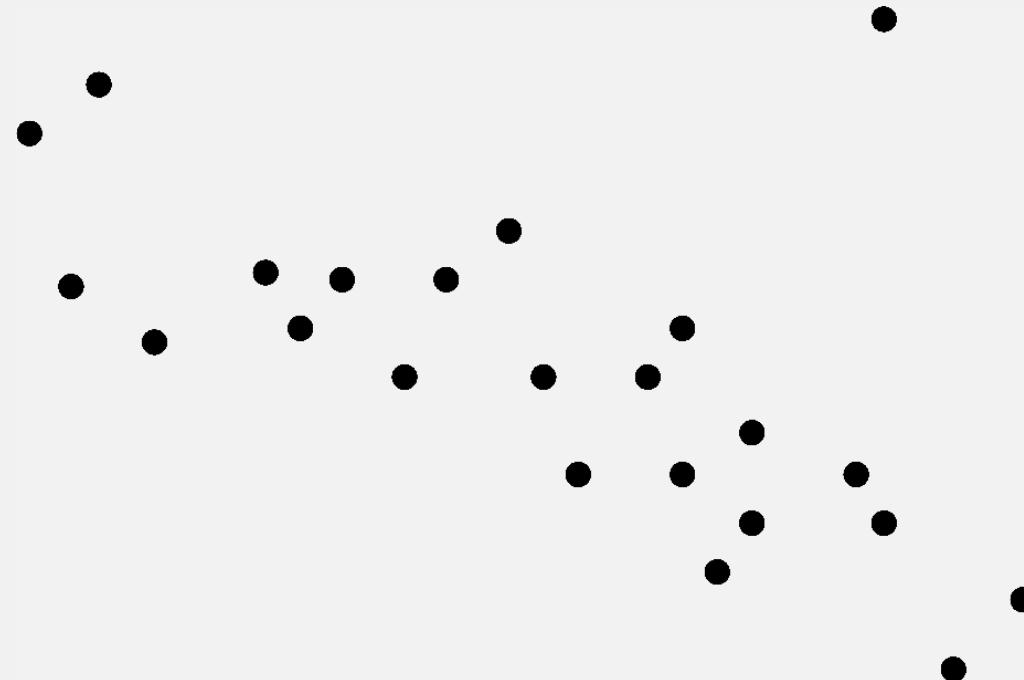
brightness





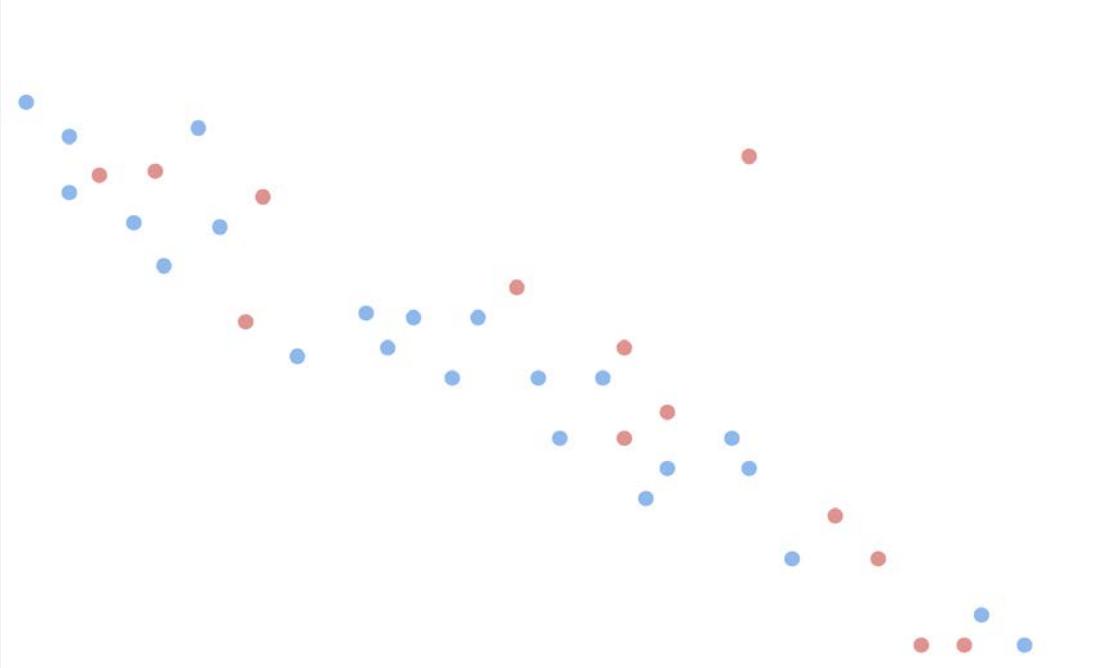
Position

We mainly see differences in 2D positions: horizontal differences in position and vertical differences



Type	Attribute (variations of)	Quantitative/Categorical Perception
Form	Length	Quantitative
	Width	Quantitative BUT LIMITED
	Orientation	Quantitative (ONLY TRENDS)
	Size (Area)	Quantitative BUT LIMITED
	Shape	Categorical
	Enclosure	Categorical
Color	Hue	Categorical
	Intensity	Quantitative BUT LIMITED
Spatial Position	2-D Position	Quantitative (!!! Best one)

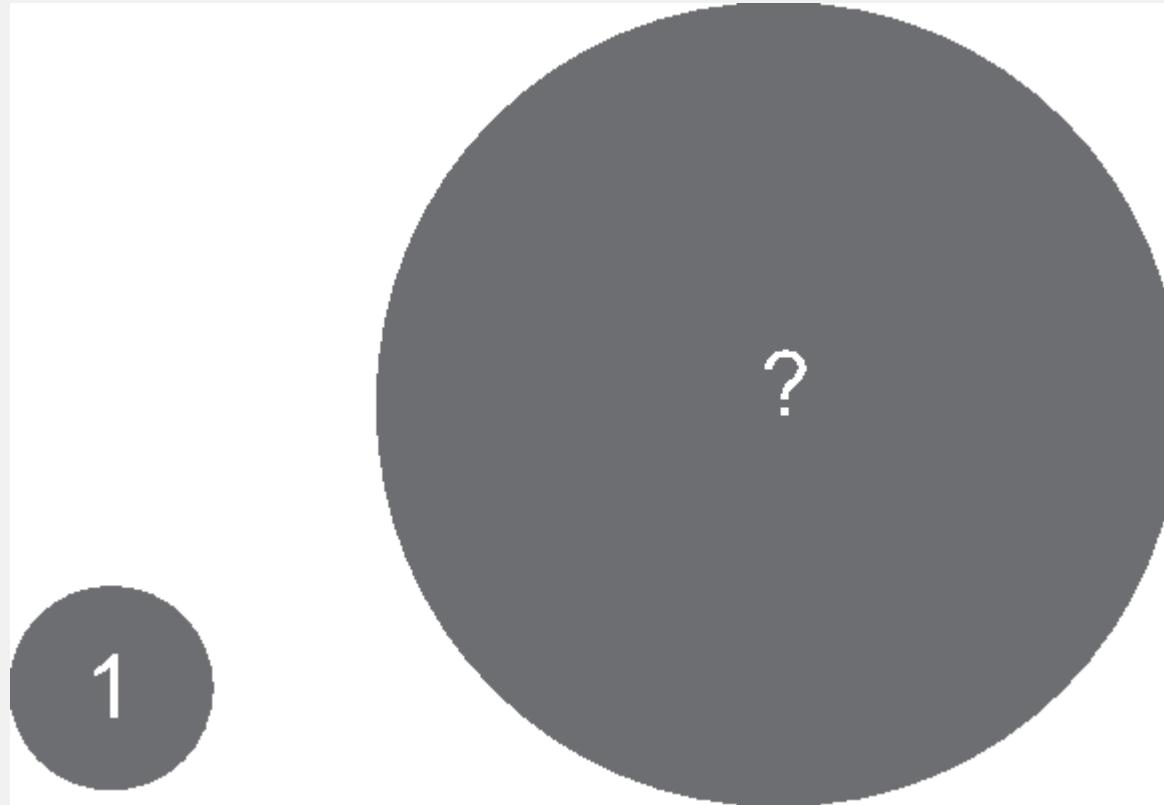


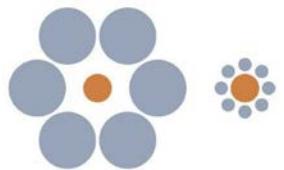


Colors let me perceive two categories

Position let me perceive higher(lower values)







MIND THE OVERALL PICTURE!

"[...] It is easy to spot a hawk in the clear sky
It is easy to spot a hawk in a sky full of pigeons
It is NOT easy to spot a hawk in a sky full of different kind of birds [...]"

Pre-attentive memory is small, for each pre-attentive attributes, it can recognize a limited number of variations.

8 different hues,
4 different sizes,
4 different orientations ...



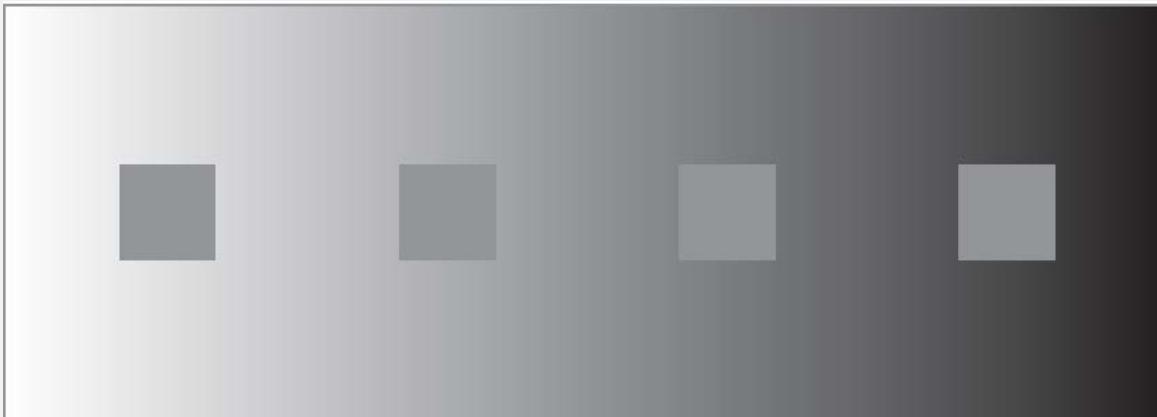


Mind the context!

The context influences the way the output is perceived

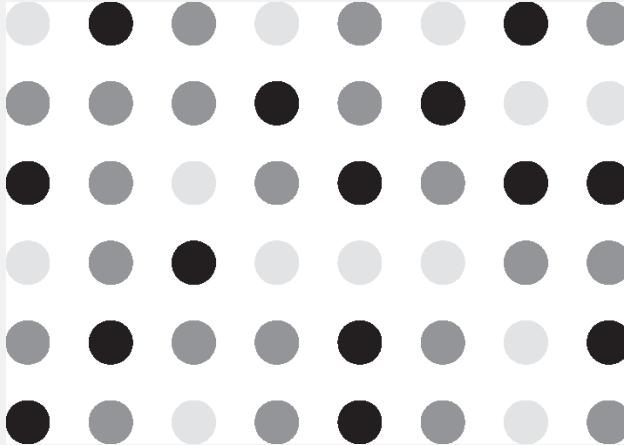
Black text works well here (grey background)

Black text do not work well here (dark blue background)

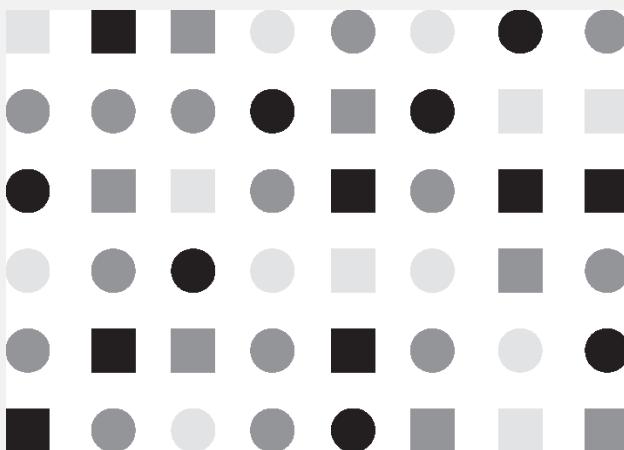




Two attributes alone work well but using them simultaneously may create overload.



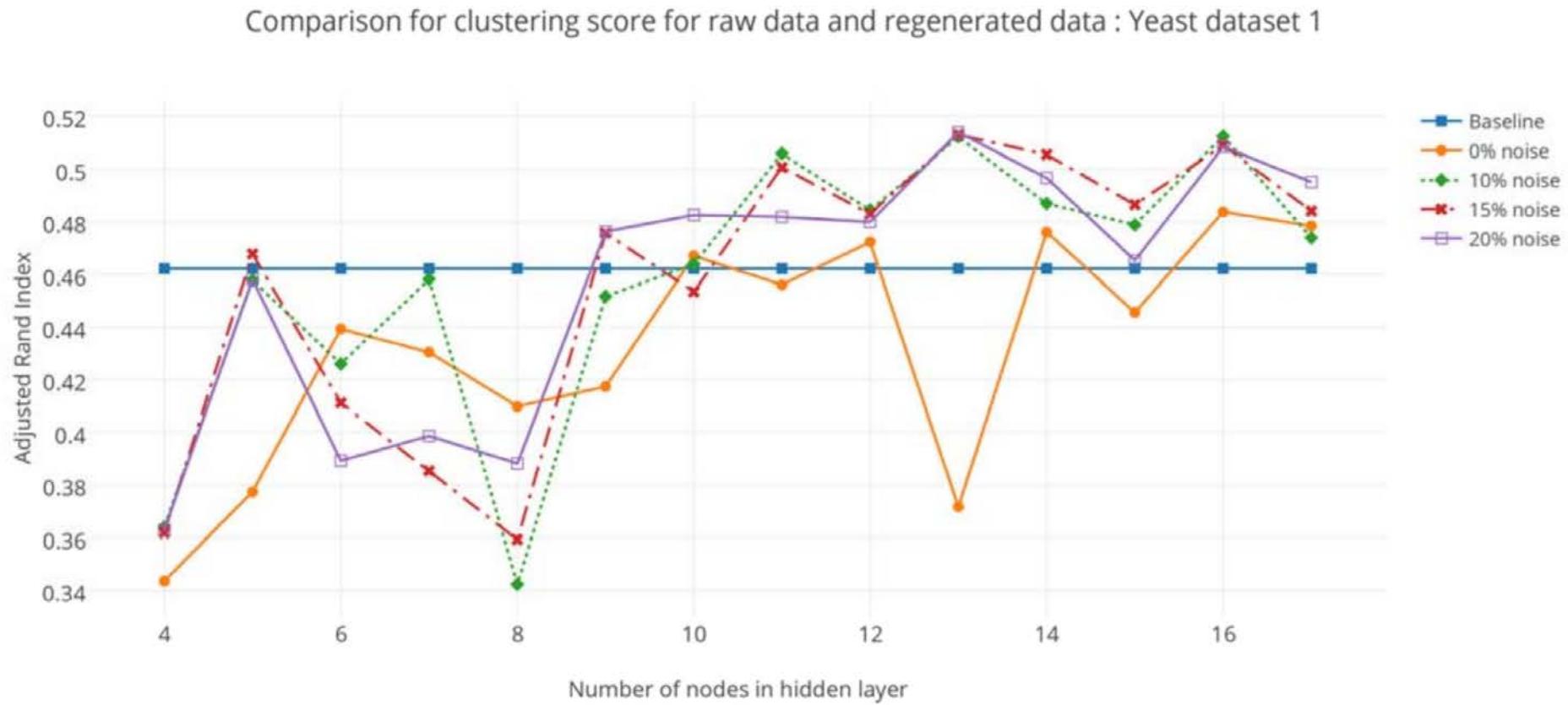
One attribute (intensity), three categories/quantities



Two attributes (shape+intensity), six categories
but picking up the brightest categories is more difficult;
readers get annoyed



Line-plot using different colors and different markers



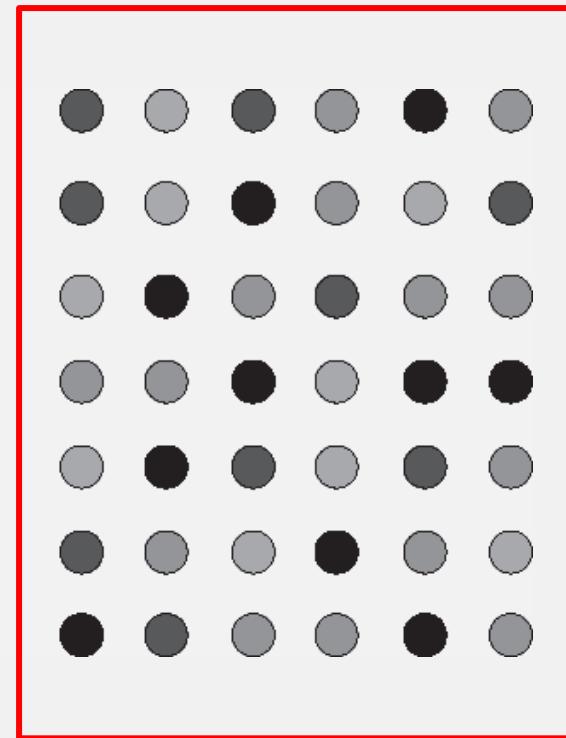
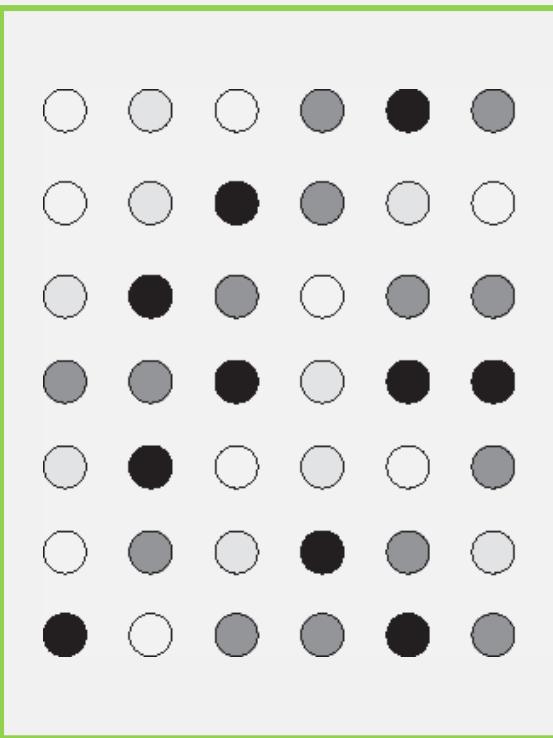
Shapes and colors are used together to discriminate among categories





PICK ATTRIBUTES for which the range is big enough

Differences between different categories/values must reflect in farthest possible differences in the attributes

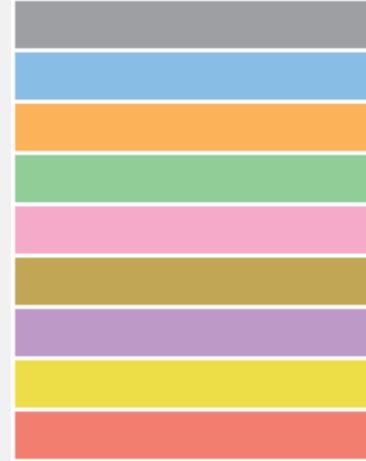


Hues that are seen as different are

- Gray
- Blue
- Orange
- Green
- Pink
- Brown
- Purple
- Yellow
- Red



Colors



Represent different things depending on the cultural background.

Bright, saturated colors are strong and exciting, attracting attention.

Natural colors (generally not fully saturated) are more neutral and soothing.



Use saturated colors to highlight some information.

Use not saturated colors if you want to color tables or for graphs where all the info have similar importancy and nothing need to prevail.

The ability of distinguishing colors diminish with the size of colored objects

Rule of Thumb: whenever you have N categories to be represented, chose a (**colorblind** safe!!) palette and prepare three version of it:

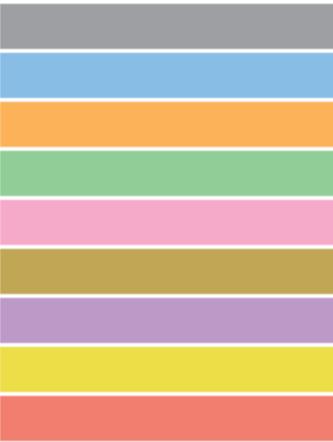
- one with **low saturated** (yet distinguishable!!) colors for **bigger shapes** (e.g. bars, treemaps)
- one with the same colors, but **darker** for **small objects** (e.g. points in scatterplot)
- one with the same **saturated** colors for infos to be highlighted.



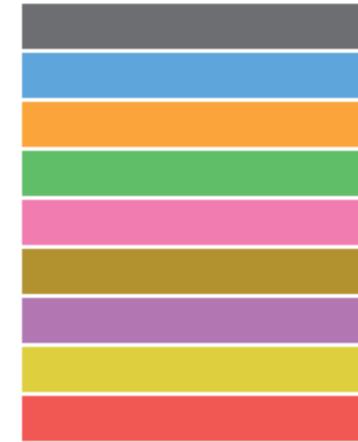


<https://colorbrewer2.org/>

Not saturated



Darker



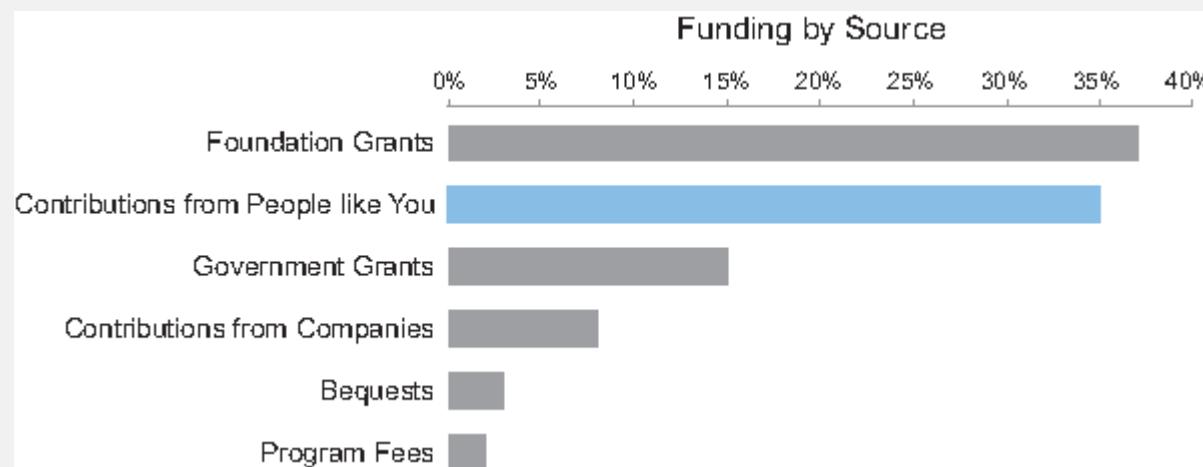
Saturated



In Times Square you are lost... nothing and everything attract you. Everything is important but nothing stand out clear.

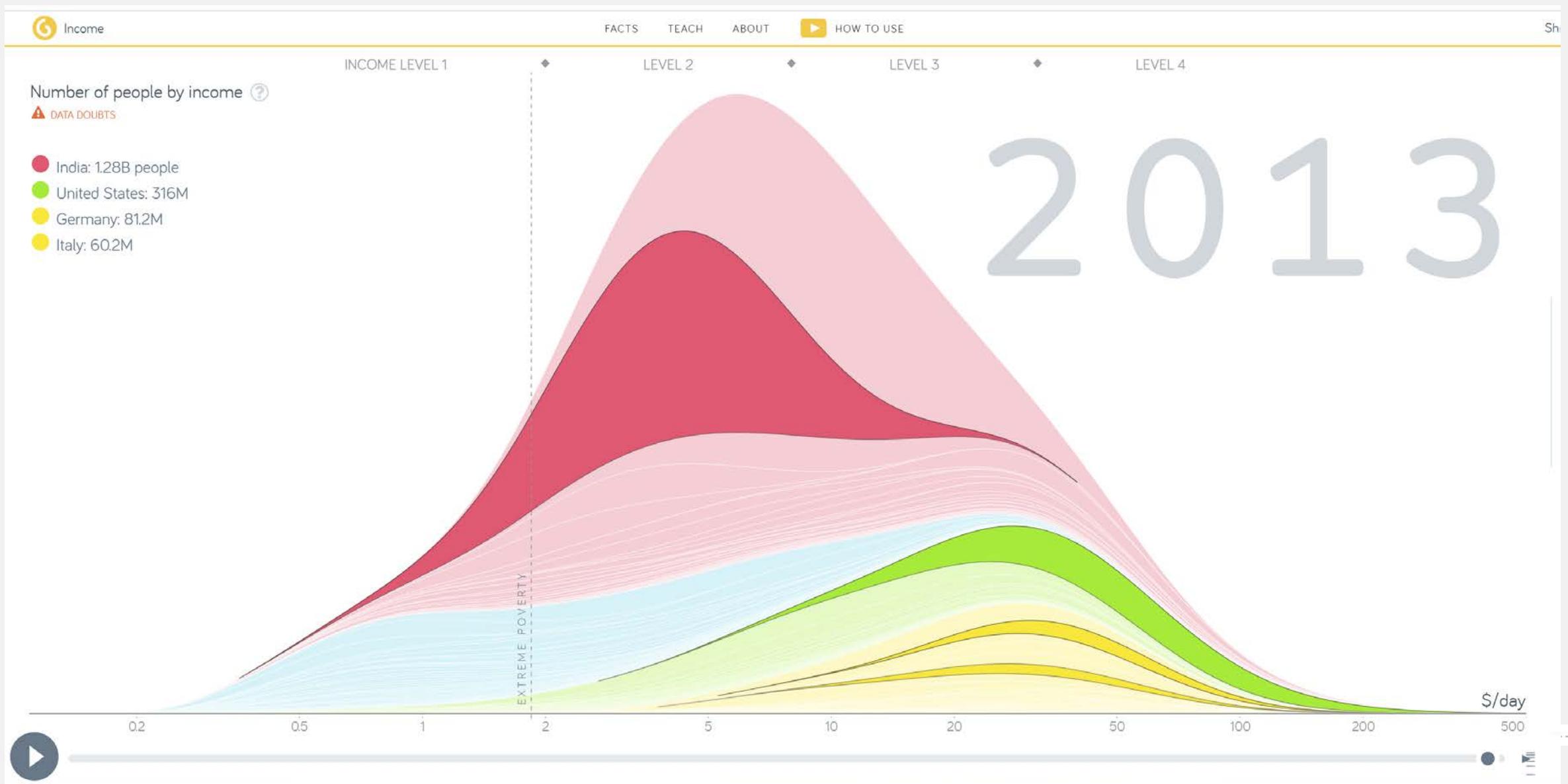
Sensation and Perception like contrast and (color) differences.

BUT too many variations delete the importance of each other.

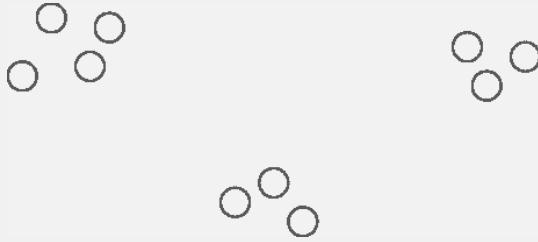




[GapMinder website](#): colors show different geographical areas



GESTALT PRINCIPLES explain pattern perception

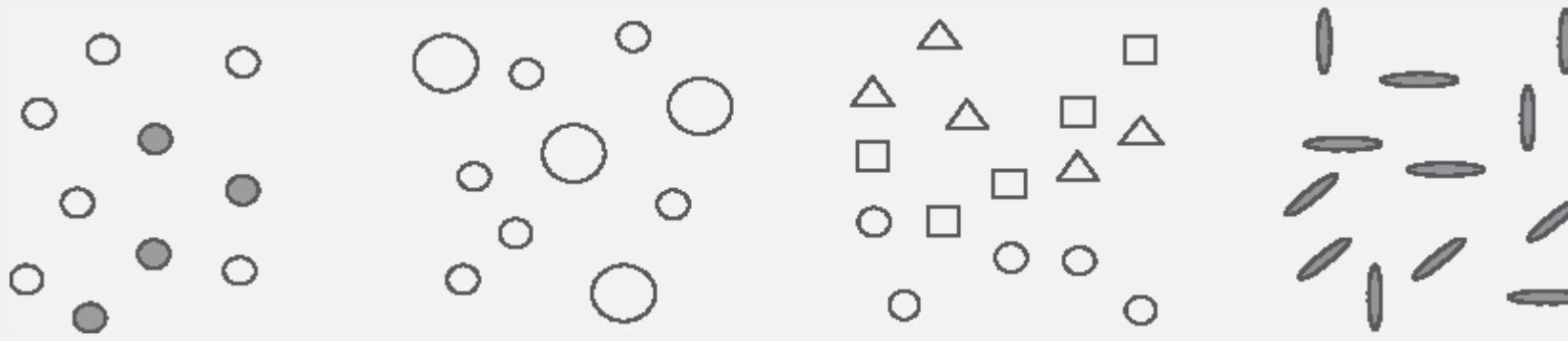


Proximity principle: elements close to each other are perceived as belonging to the same group.

Users that each row is a group;
They scan the table row by row

Users that each column is a group;
They scan the table column by column

Similarity principle: (as we said) similar object (in color, size, shape, ...) are grouped

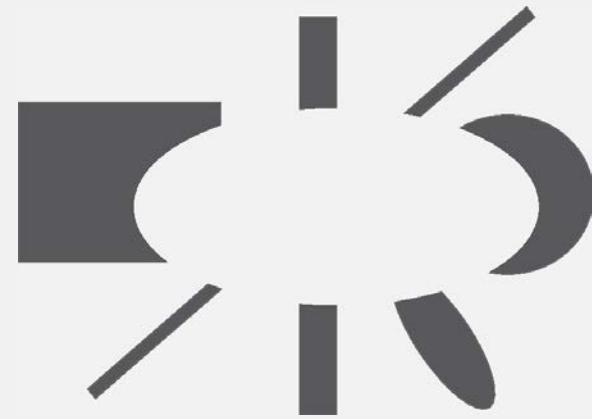
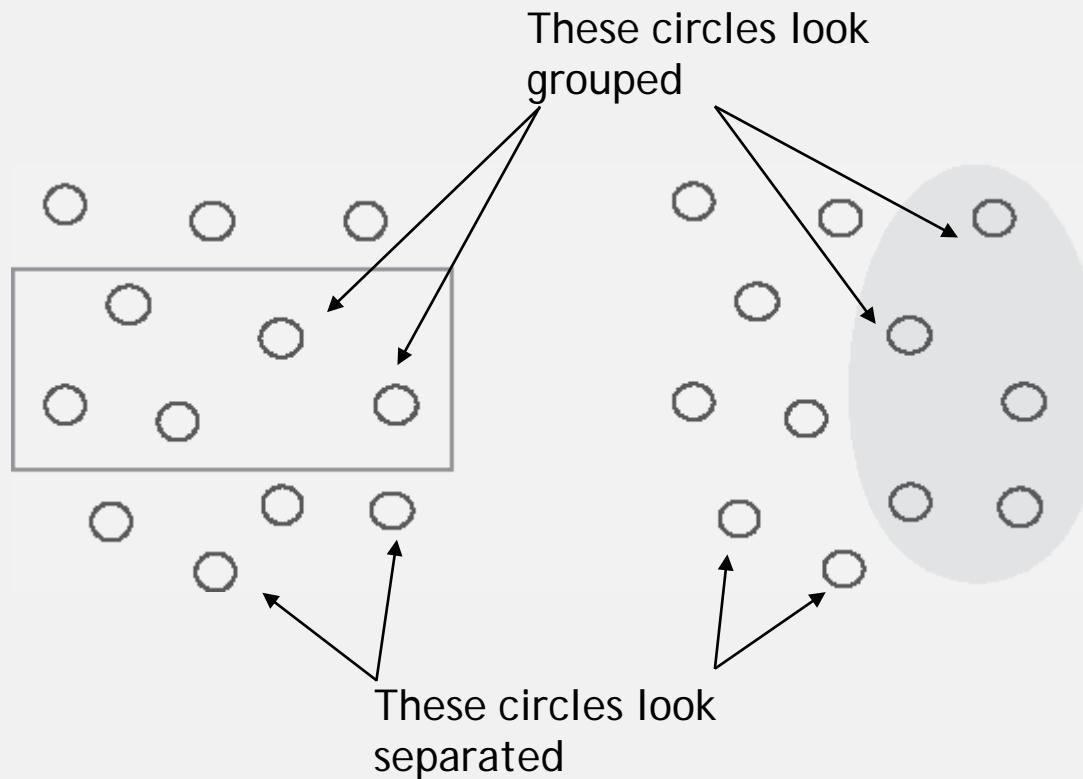


But avoid using too many categories!!!

In tables use differing colors in columns (rows) if you want to orient reader to focus on column (rows)



Principle of enclosure: objects that are enclosed in a way that forms a boundary around them are grouped



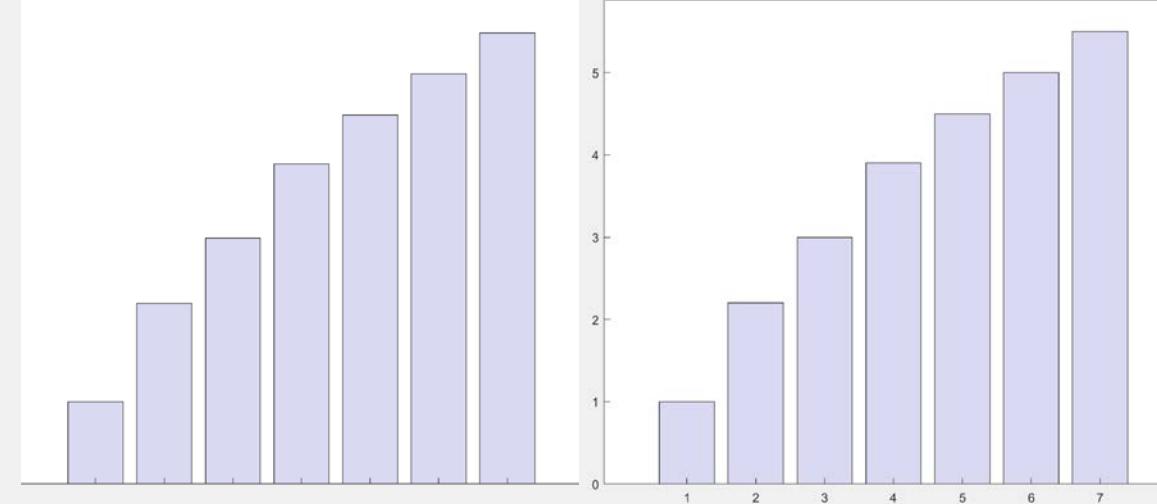
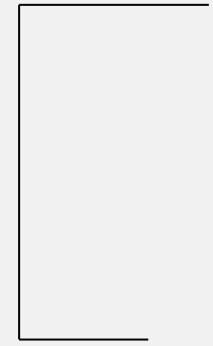
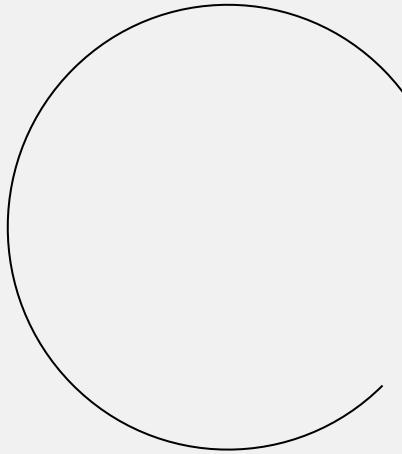
It looks like points in the middle
are together and form an ellipsis

An example from Hans Rosling (TED'S TALKS): "[let my dataset change your mindset](#)"





Principle of closure: human fill the gaps and close open lines





Principle of continuity: aligned objects are perceived as belonging to the same part

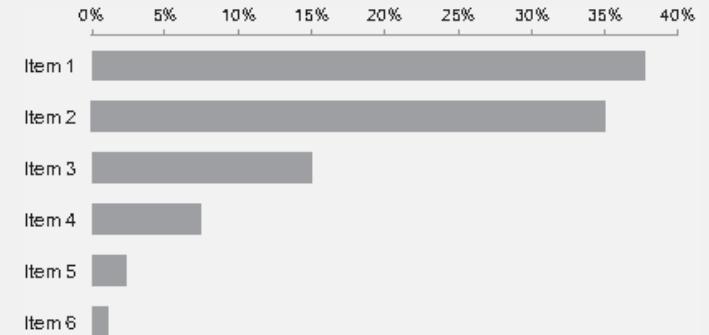
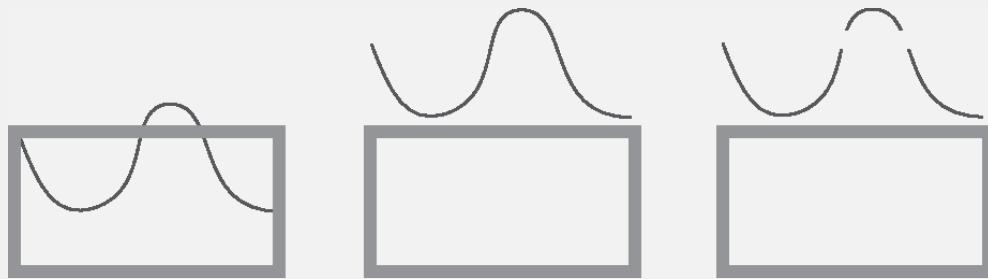
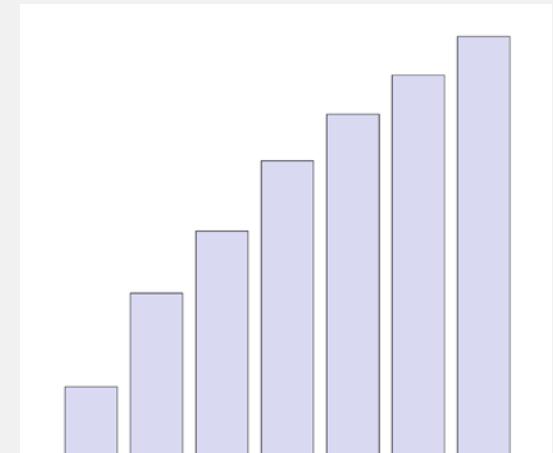
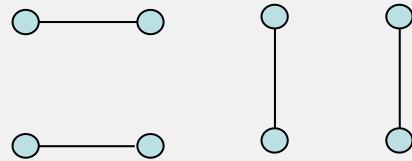


Table A.23: Household financial stress—C10 ‡

	Household comparisons					
	Adult low paid		Other		All households	
	'000s	%	'000s	%	'000s	%
Family finances: optimists						
Poor or very poor	20	1.6	44	1.2	64	1.3
Just getting along	285	23.8	720	19.0	1,005	20.1
Reasonably comfortable	645	53.9	2,039	53.7	2,684	53.8
Prosperous or v comfort	246	20.6	991	26.1	1,237	24.8
Total	1,196	100.0	3,793	100.0	4,990	100.0
Family finances: pessimists						
Poor or very poor	46	3.8	104	2.8	150	3.0
Just getting along	401	33.5	1,054	27.8	1,454	29.1
Reasonably comfortable	645	53.9	2,097	55.3	2,742	55.0
Prosperous or v comfort	105	8.8	539	14.2	644	12.9
Total	1,196	100.0	3,793	100.0	4,990	100.0
Episodes of financial hardship						
Three or more	135	11.3	295	7.8	430	8.7
Two	115	9.7	282	7.5	397	8.0
One	160	13.4	509	13.5	668	13.5
None	781	65.6	2,691	71.3	3,472	69.9
Total	1,191	100.0	3,776	100.0	4,967	100.0
How easily raise \$2000 in one week						
Could not raise it	244	20.4	481	12.7	725	14.6
Have to do something drastic	194	16.2	399	10.5	593	11.9
Raise it, but some sacrifices	321	26.8	949	25.1	1,270	25.5
Easily raise it	436	36.5	1,956	51.7	2,393	48.0
Total	1,196	100.0	3,785	100.0	4,981	100.0



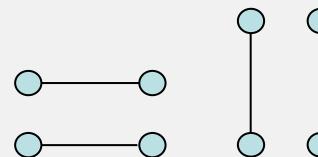
Principle of connection: objects connected by lines are viewed as part of the same group



Points are equally spaced but those connected are perceived as grouped

Connection is more powerful than proximity or similarity but less powerful than enclosure

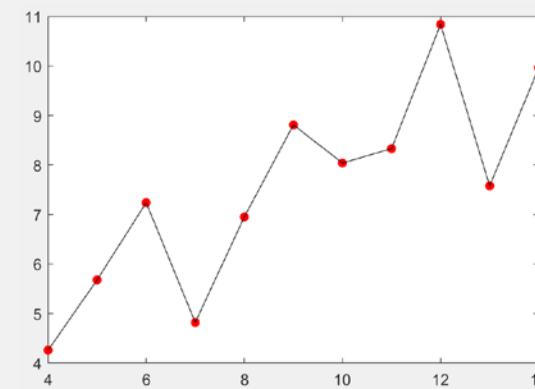
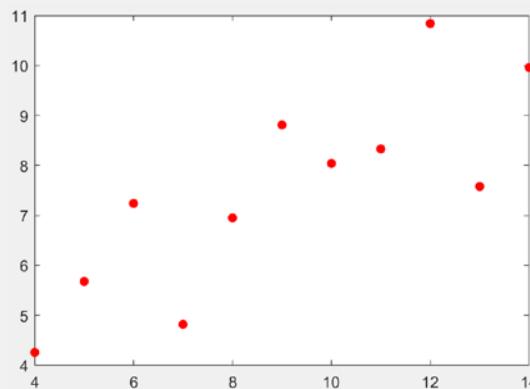
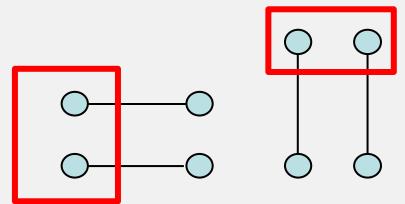
proximity



similarity



enclosure



Point connection with lines in plots let us identify the trend and see patterns

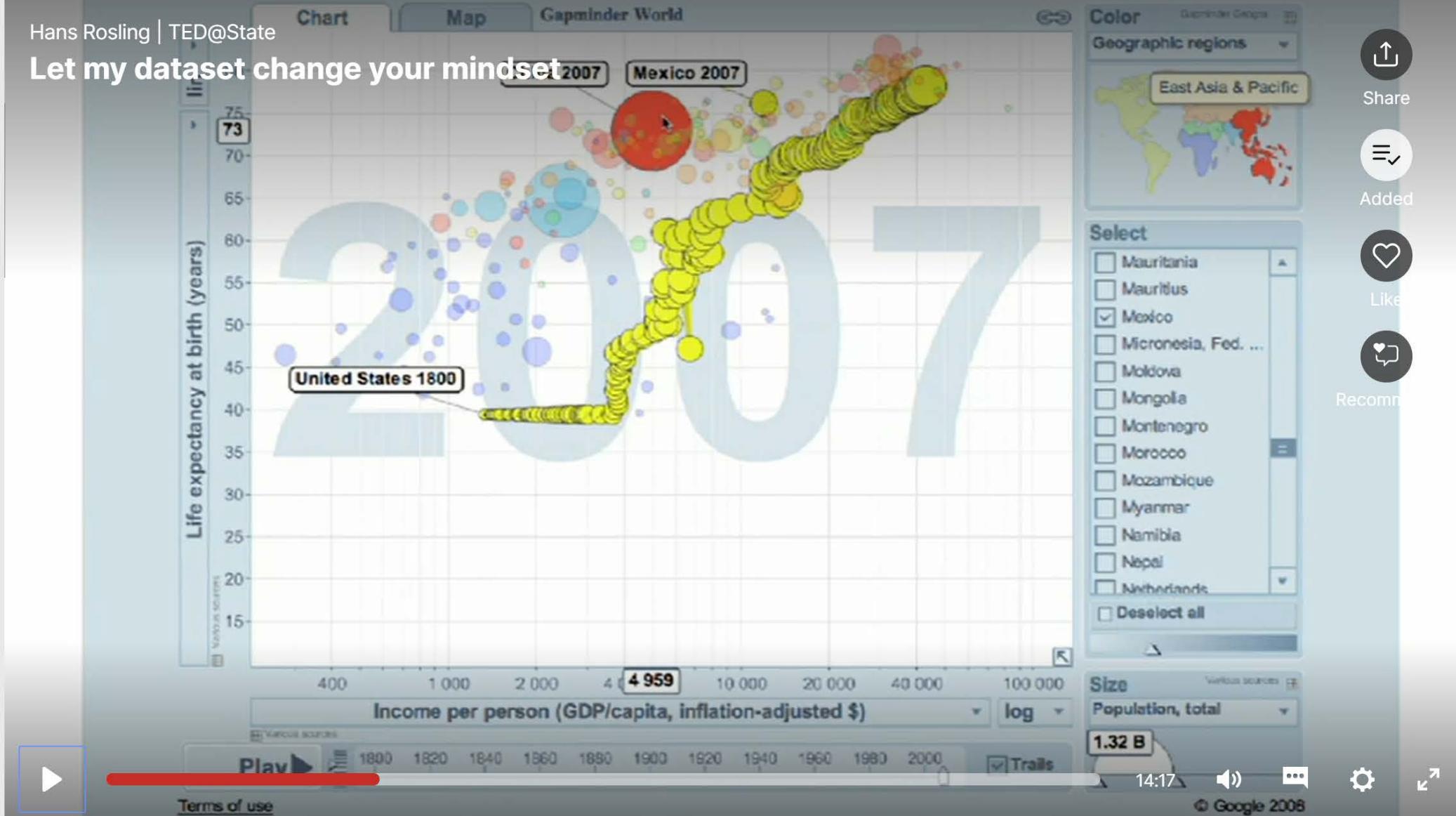




Hans Rosling (TED'S TALKS): "let my dataset change your mindset"



Let my dataset change your mindset



Bubbles are connected to show trends,
Keeping all bubbles allows to show also the



Share



Added



Like

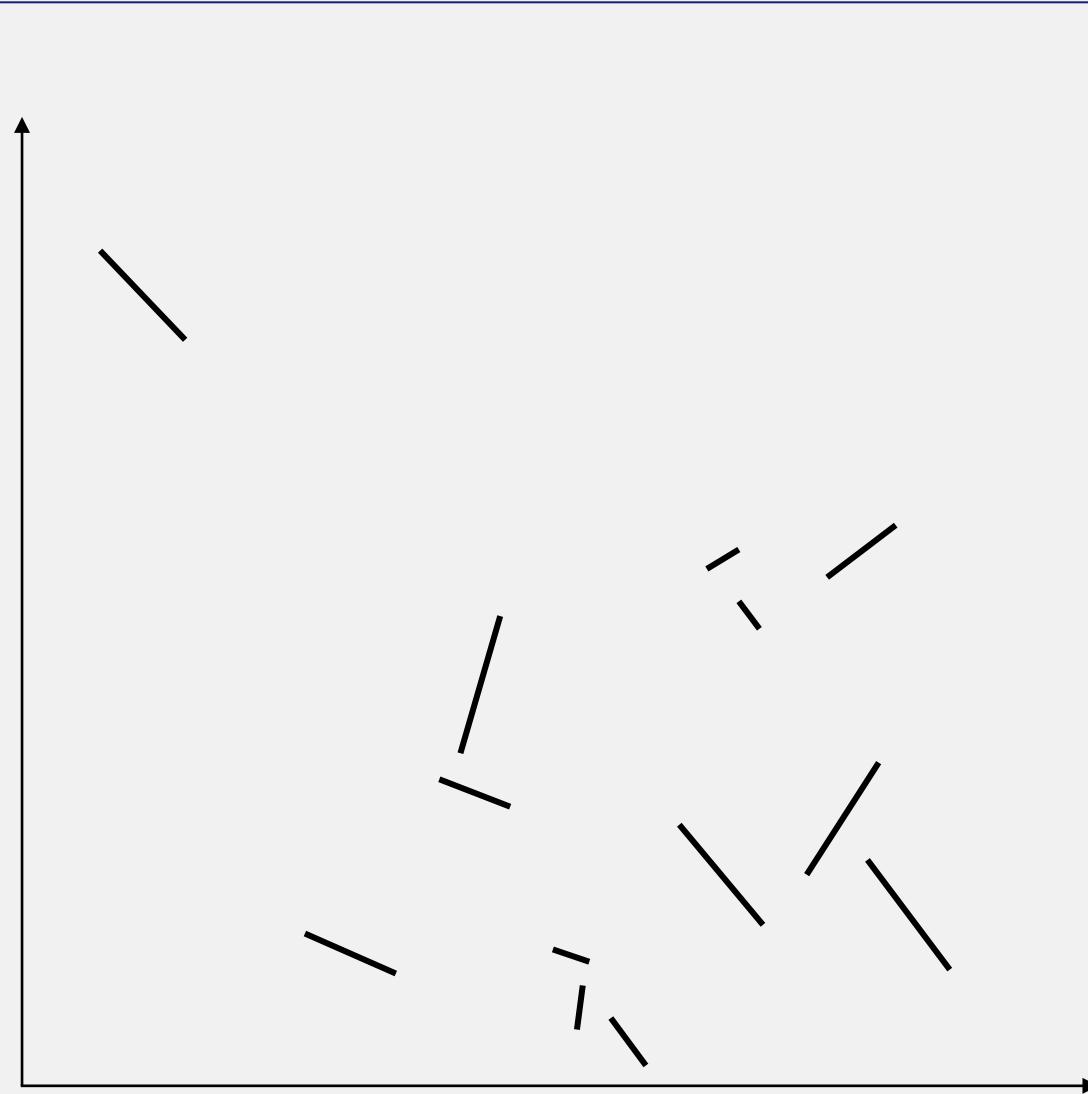


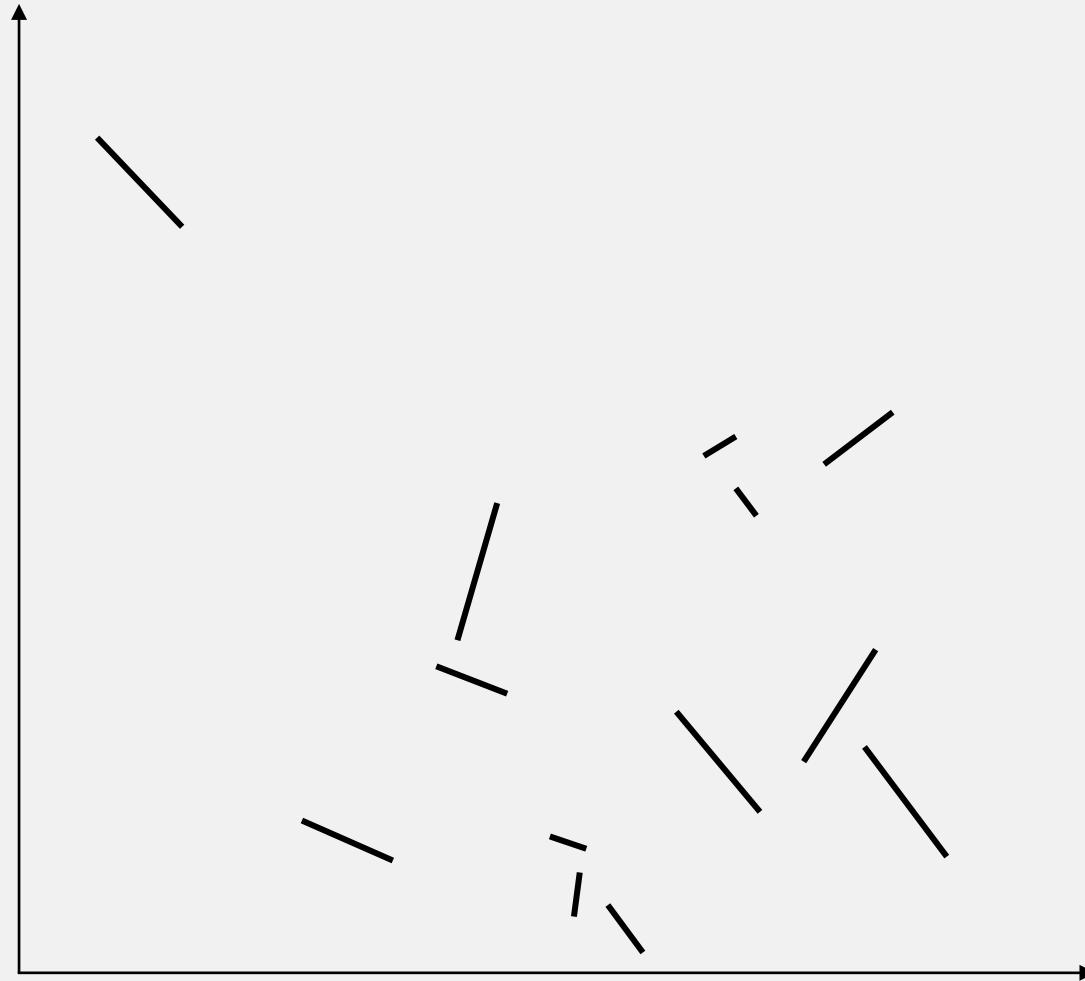
Recomm



Another example:
what do you perceive in the following plot?







Are there categories?

How do you group segments?

- Orientation:

- Orientation does not allow to divide into categories (colors or shapes on the segments would)
- different orientations may allow showing trends

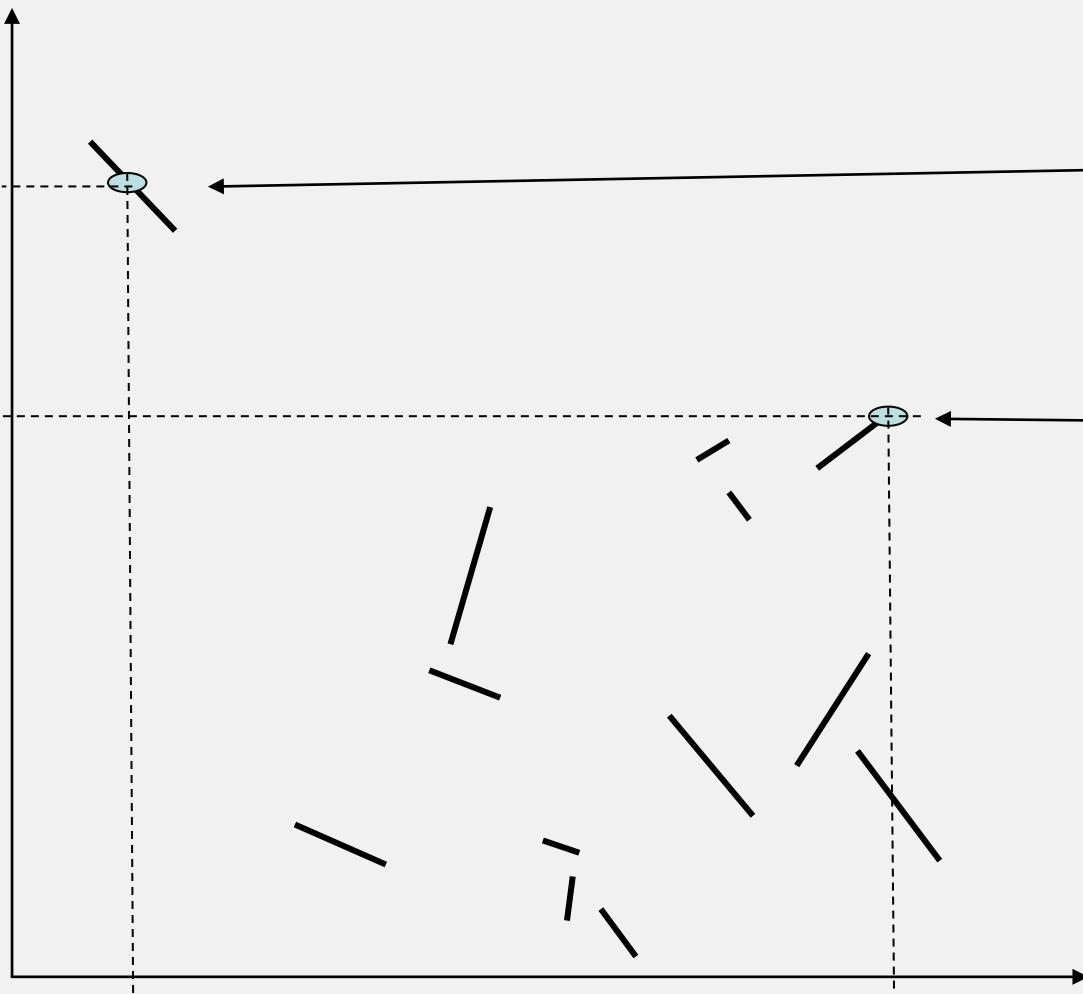
- Length

- Length allows “sizing” each input (we perceive it as related to the strength of the point)





Position: which is the precise position of the points??



Where would you place the center on the x, y axis?

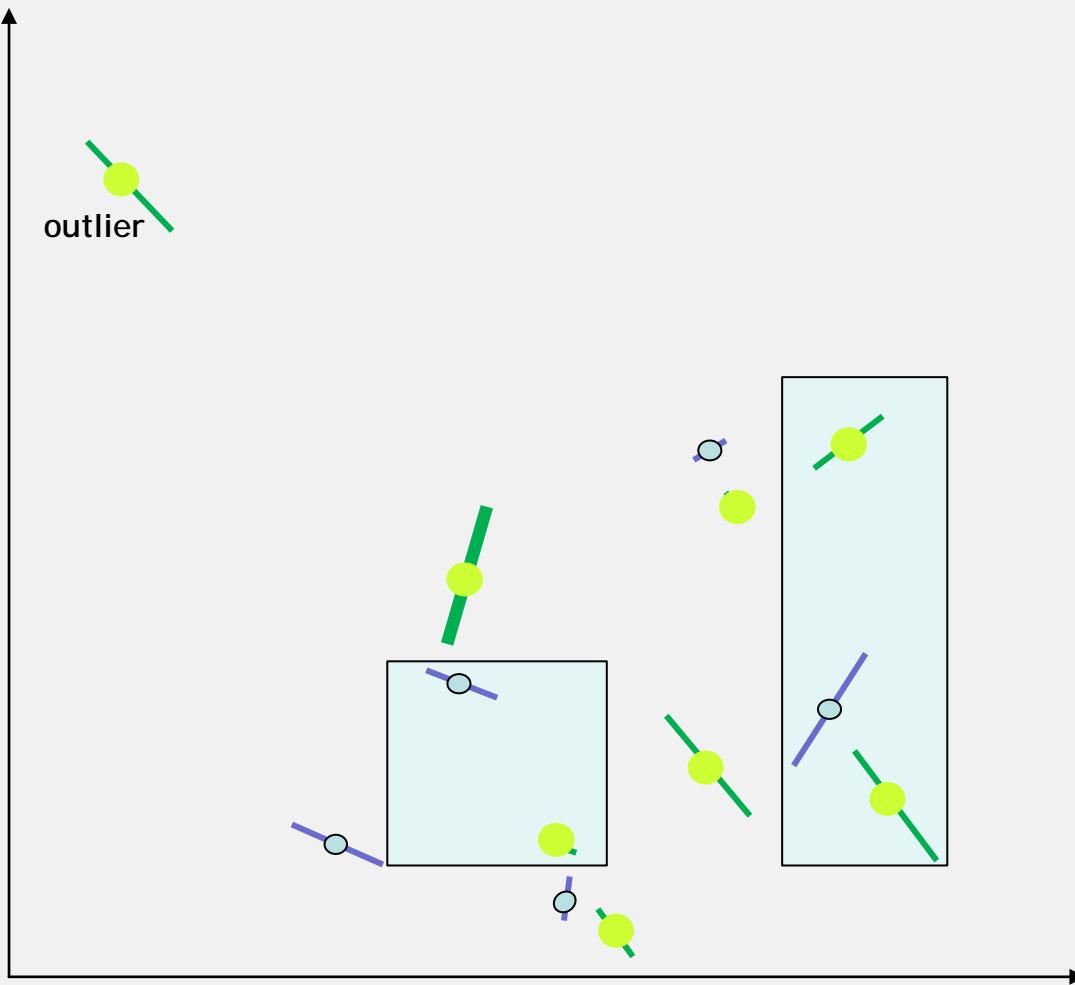
At Center of the segment

At one end of the segment (physics memories)

- the position of the segment is at the extremes of the segment or at the center???
- Extremes of the segments if you recall physics lectures (but you are implicitly searching for an arrow)

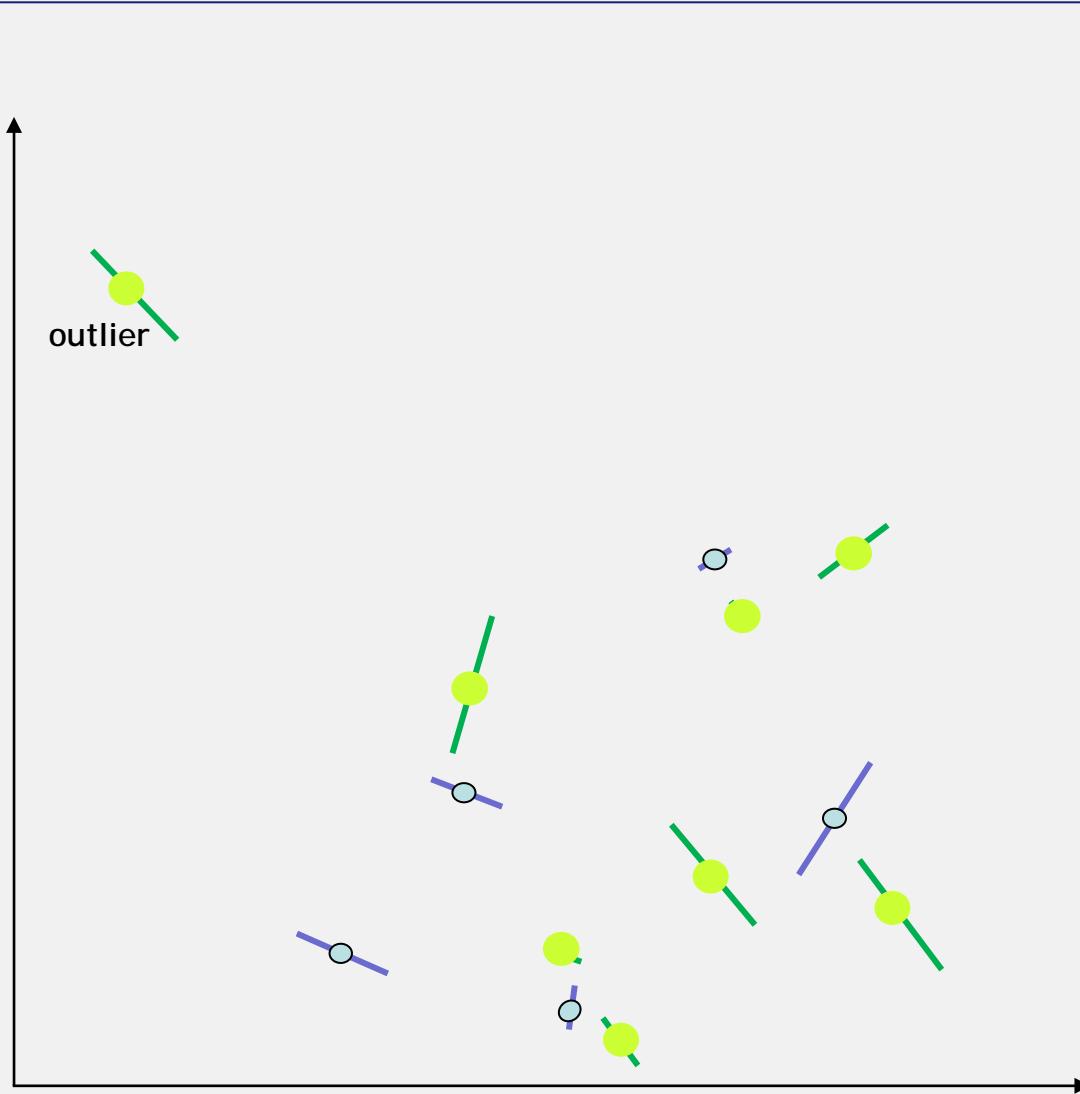


The gestalt principles



- Similarity in Colors and shapes discriminate categories
- Enclosure allows grouping certain elements
- Differences in proximities (Position) helps identifying outliers and further communicates measurements

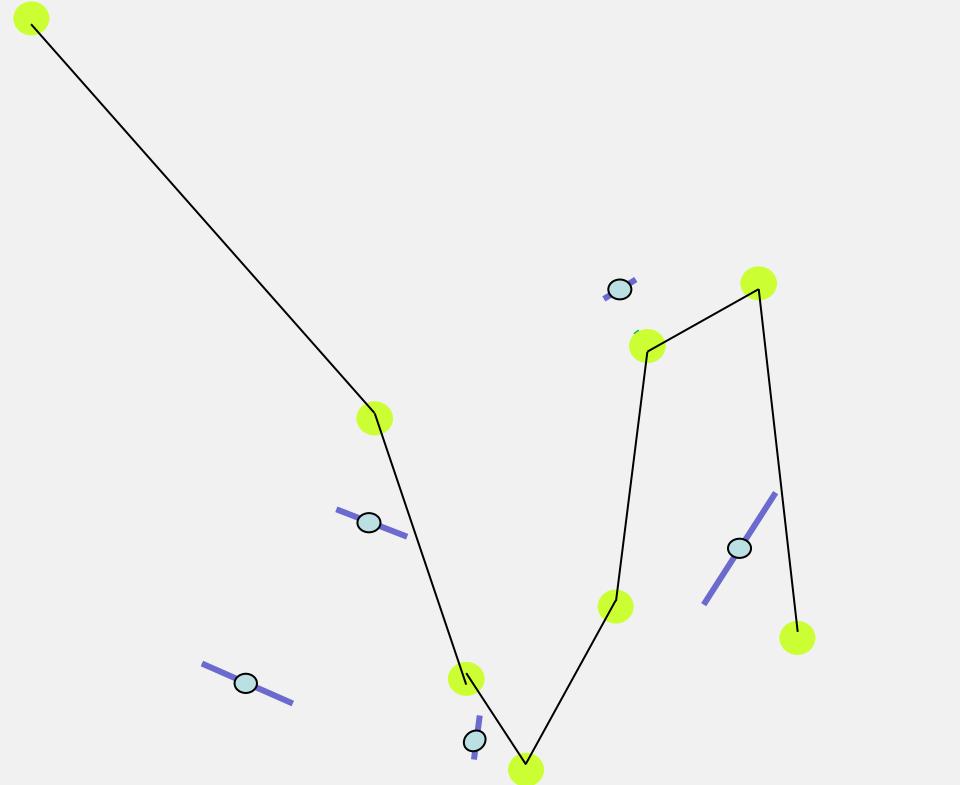




otherwise

- Proximity further drills down connections
(though enclosure is stronger)





- Connection helps viewing the pattern of change





Information Design

exploits knowledge about human perception to create **Infographics** (**information+graphics**): information (data) visualizations combining integrating text, (scientific) data visualizations, and images to tell a whole story and **show results, inform, entertain, persuade** the audience.

(Cool Infographics)



EVERYTHING MUST E CONSIDERED: TEXT

Why should we be interested in Visualization? ...

The visual system has its own rules. We can see patterns presented in certain ways, but if they are presented in other ways, they become invisible...

When data is presented in certain ways, the patterns can be readily perceived.

If we can understand how perception works our knowledge can be translated into rules for displaying information.

Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out.

If we disobey the rules, our data will be incomprehensible or misleading

Colin Ware, «Information Visualization: Perception for Design» (2004)





COLORS....

Why should we be interested in Visualization? ...

The visual system has its own rules. We can see patterns presented in certain ways, but if they are presented in other ways, they become invisible...

When data is presented in certain ways, the patterns can be readily perceived.

If we can understand how perception works our knowledge can be translated into rules for displaying information.

Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out.

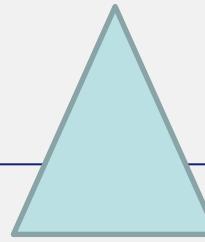
If we disobey the rules, our data will be incomprehensible or misleading

Colin Ware, «Information Visualization: Perception for Design» (2004)





SHAPES



Why should we be interested in Visualization? ...

The visual system has its own rules. We can see patterns presented in certain ways, but if they are presented in other ways, they become invisible...

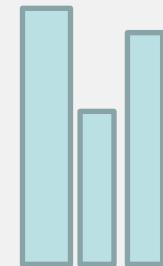
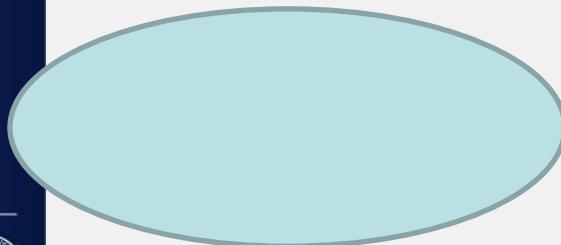
When data is presented in certain ways, the patterns can be readily perceived.

If we can understand how perception works our knowledge can be translated into rules for displaying information.

Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out.

If we disobey the rules, our data will be incomprehensible or misleading

Colin Ware, «Information Visualization: Perception for Design» (2004)



Why should we be interested in Visualization? ...

The visual system has its own rules. We can see patterns presented in certain ways, but if they are presented in other ways, they become invisible...

When data is presented in certain ways, the patterns can be readily perceived.

If we can understand how perception works our knowledge can be translated into rules for displaying information.

Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out.

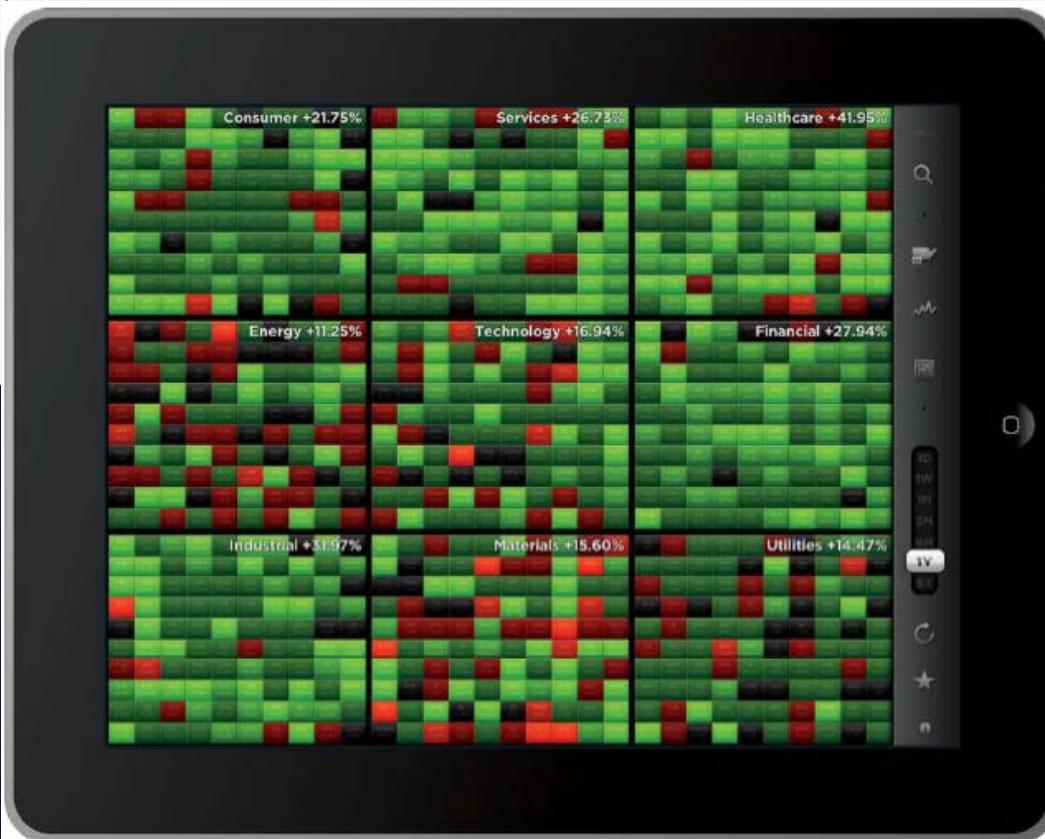
If we disobey the rules, our data will be incomprehensible or misleading.

Colin Ware, «Information Visualization: Perception for Design» (2004)



“Of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful” (Edward Tufte)

For humans, visualization is one of the most natural ways for understanding things.



Good (???) interactive visualization: StockTouch App on iPad.
U.S. Stocks from 9 market sectors.
In each market sector companies are organized in a spiral pattern, from largest company (in the middle of the square) to smallest (on the borders of the square).
Each stock is color-coded based on its stock price performance over the (user-selected) period.
The shades of green (red) show stock prices that have increased (decreased).
Touching a squares shows plots describing the company stock prices.



Though it seems messy it is considered one of the best tools for Business people.

Unfortunately there is no rule.

“the aim of good data graphics is to display data accurately and clearly”



Favorable or Unfavorable View of the U.S.	
Brazil: % with somewhat or very favorable opinion of the U.S.:	52%
Brazil: % with somewhat or very unfavorable opinion of the U.S.:	32%
Mexico: % with somewhat or very favorable opinion of the U.S.:	64%
Mexico: % with somewhat or very unfavorable opinion of the U.S.:	25%
Britain: % with somewhat or very favorable opinion of the U.S.:	75%
Britain: % with somewhat or very unfavorable opinion of the U.S.:	16%
Germany: % with somewhat or very favorable opinion of the U.S.:	61%
Germany: % with somewhat or very unfavorable opinion of the U.S.:	35%
Russia: % with somewhat or very favorable opinion of the U.S.:	61%
Russia: % with somewhat or very unfavorable opinion of the U.S.:	33%
Poland: % with somewhat or very favorable opinion of the U.S.:	79 %
Poland: % with somewhat or very unfavorable opinion of the U.S.:	11%
South Africa: % with somewhat or very favorable opinion of the U.S.:	65%
South Africa: % with somewhat or very unfavorable opinion of the U.S.:	28%
Kenya: % with somewhat or very favorable opinion of the U.S.:	80%
Kenya: % with somewhat or very unfavorable opinion of the U.S.:	15%
India: % with somewhat or very favorable opinion of the U.S.:	54%
India: % with somewhat or very unfavorable opinion of the U.S.:	27%
Japan: % with somewhat or very favorable opinion of the U.S.:	72%
Japan: % with somewhat or very unfavorable opinion of the U.S.:	26%
South Korea: % with somewhat or very favorable opinion of the U.S.:	53%
South Korea: % with somewhat or very unfavorable opinion of the U.S.:	44%
Egypt: % with somewhat or very favorable opinion of the U.S.:	6%
Egypt: % with somewhat or very unfavorable opinion of the U.S.:	69%
Pakistan: % with somewhat or very favorable opinion of the U.S.:	10%
Pakistan: % with somewhat or very unfavorable opinion of the U.S.:	69%
Turkey: % with somewhat or very favorable opinion of the U.S.:	30%
Turkey: % with somewhat or very unfavorable opinion of the U.S.:	55%
Jordan: % with somewhat or very favorable opinion of the U.S.:	25%
Jordan: % with somewhat or very unfavorable opinion of the U.S.:	75%

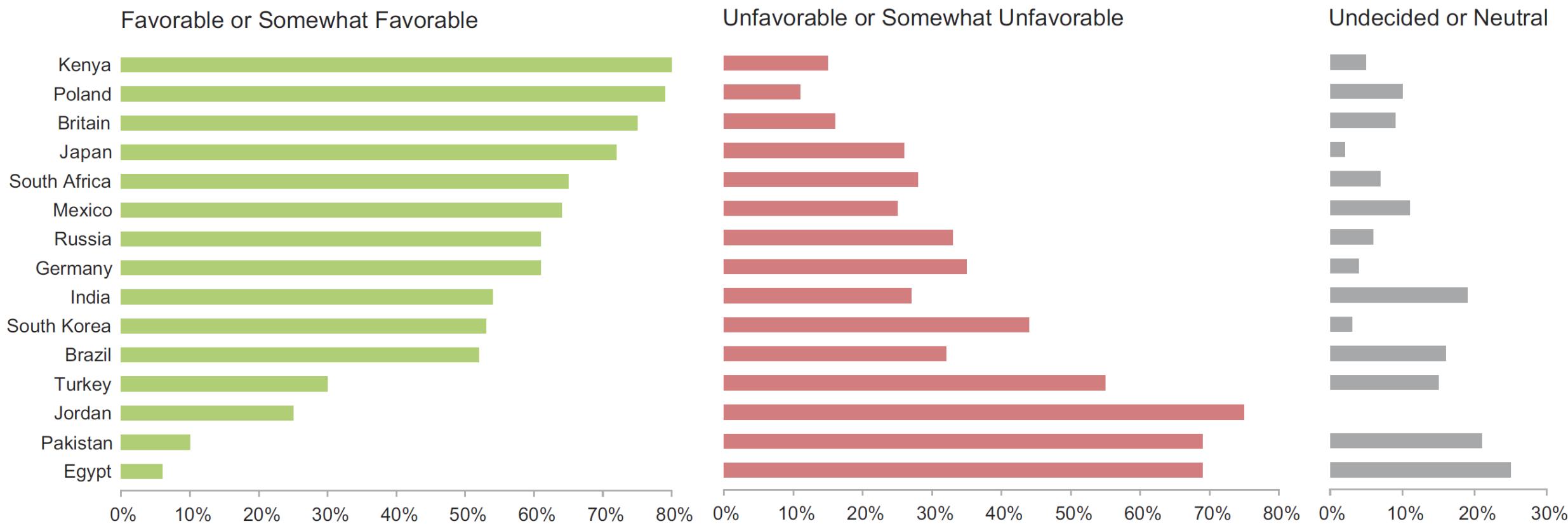
Opinions w.r.t. United States after
11th of September 2001





Much better in this way!

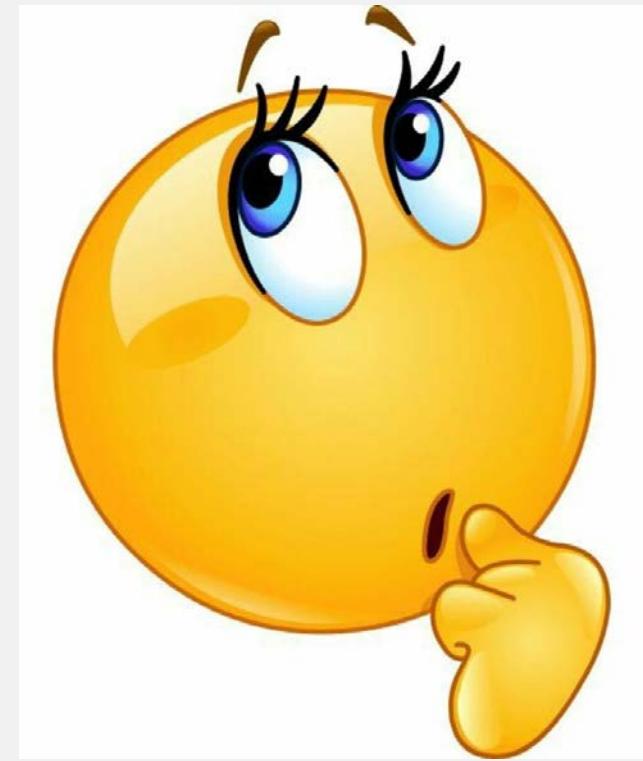
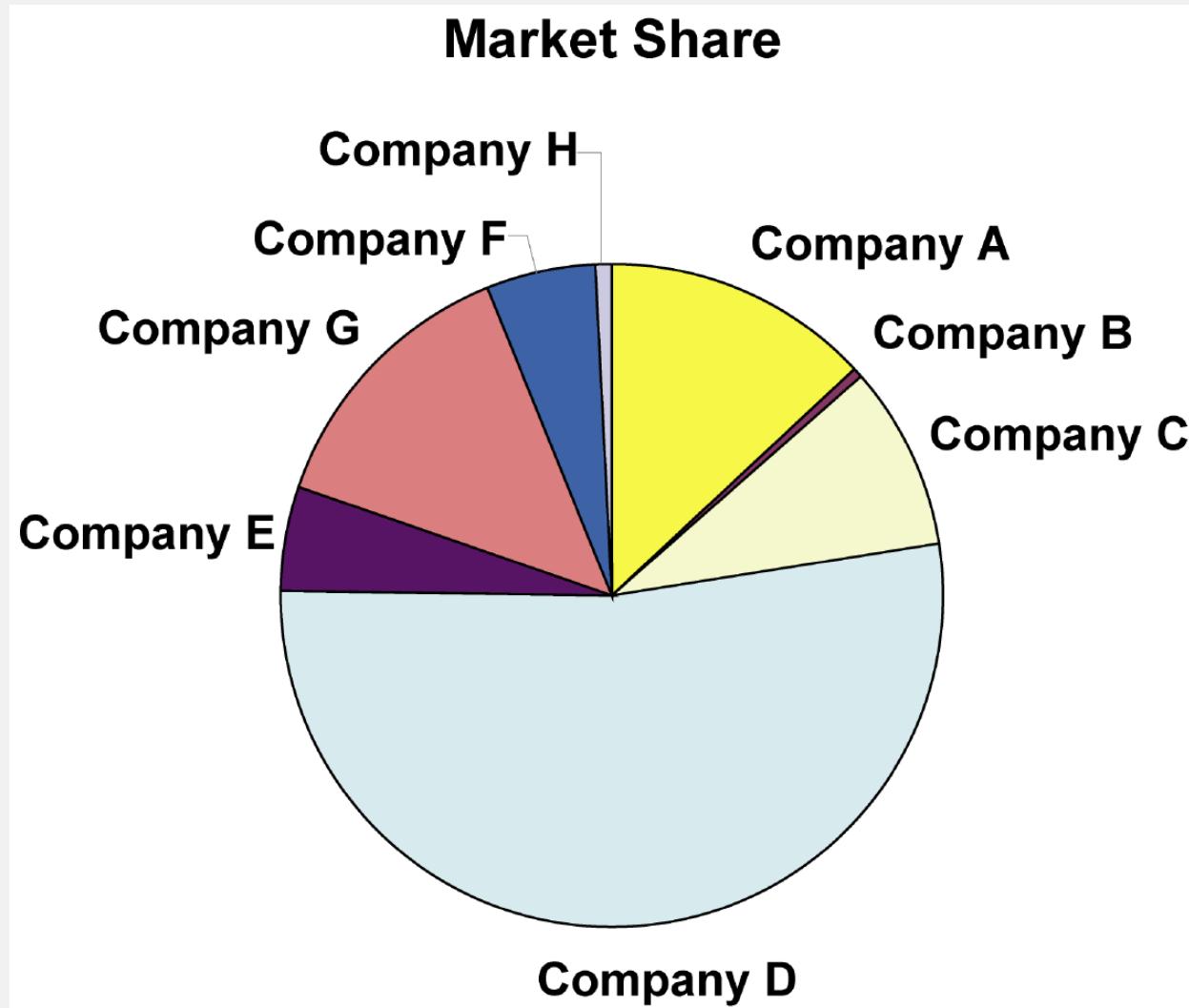
Current World Opinions About the U.S.A



Source: 2004 study conducted by the Pew Research Center, as reported by the PBS television program NOW.



OUR COMPANY IS G: BETTER OR WORSE THAN A?

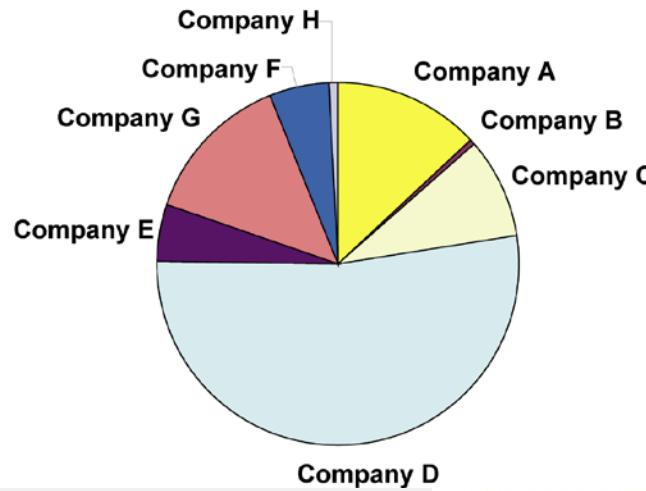


mmmmm.... Is just that there
is not enough dimension...
Let's go in 3D



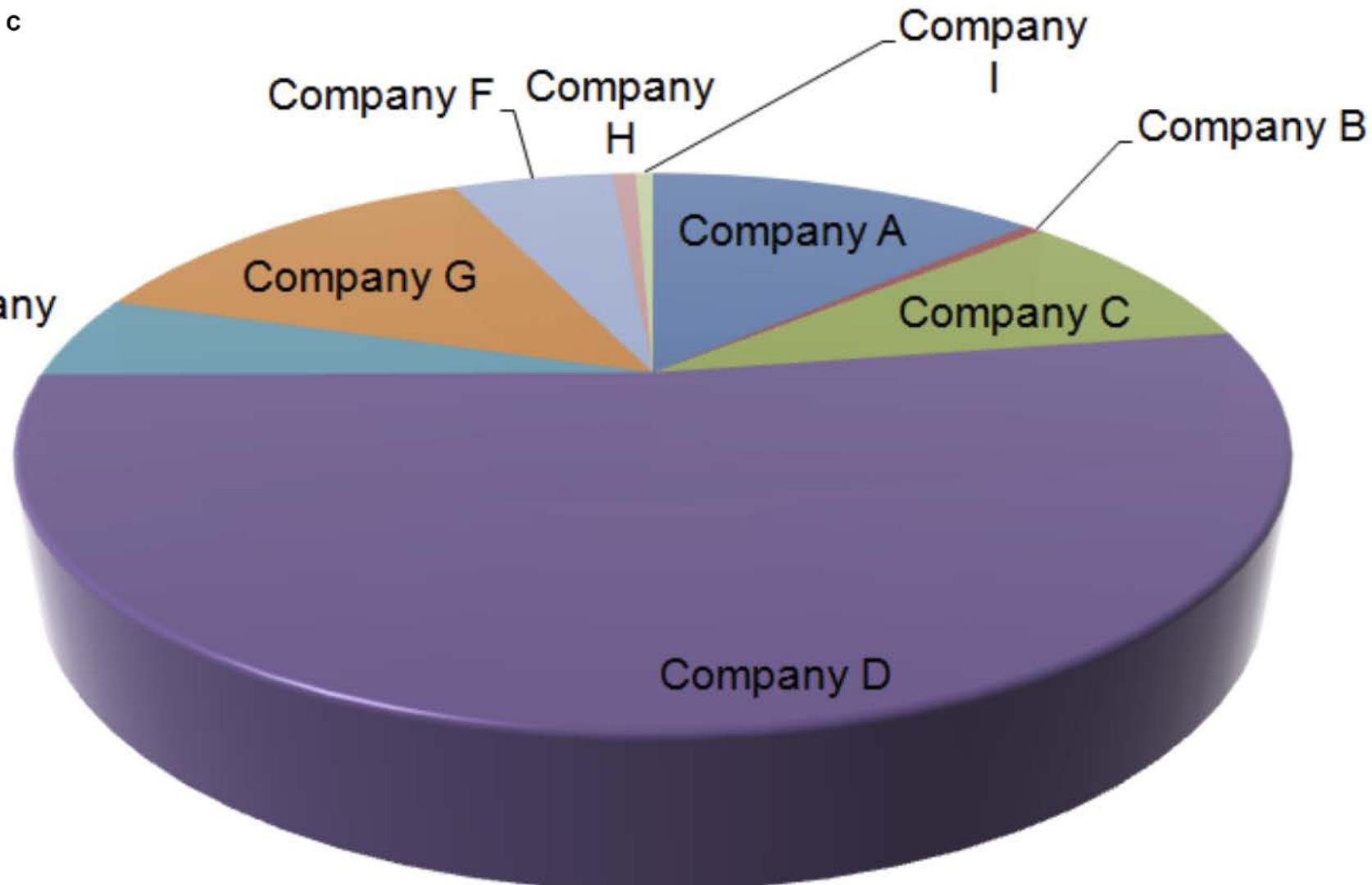


Market Share



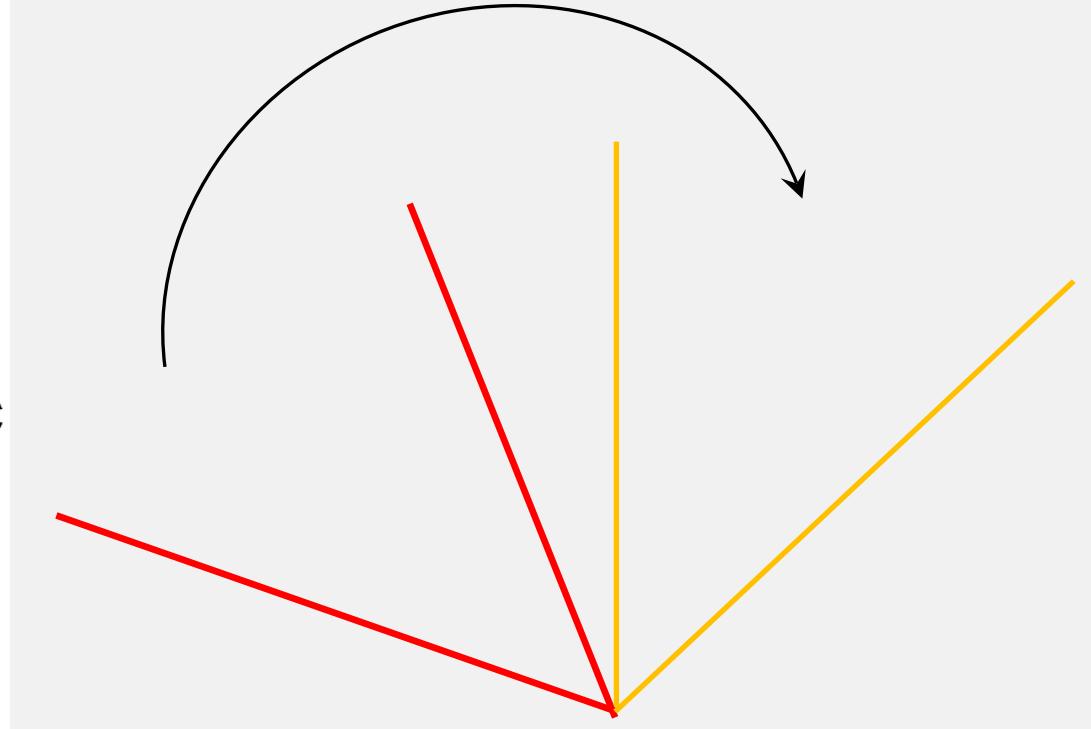
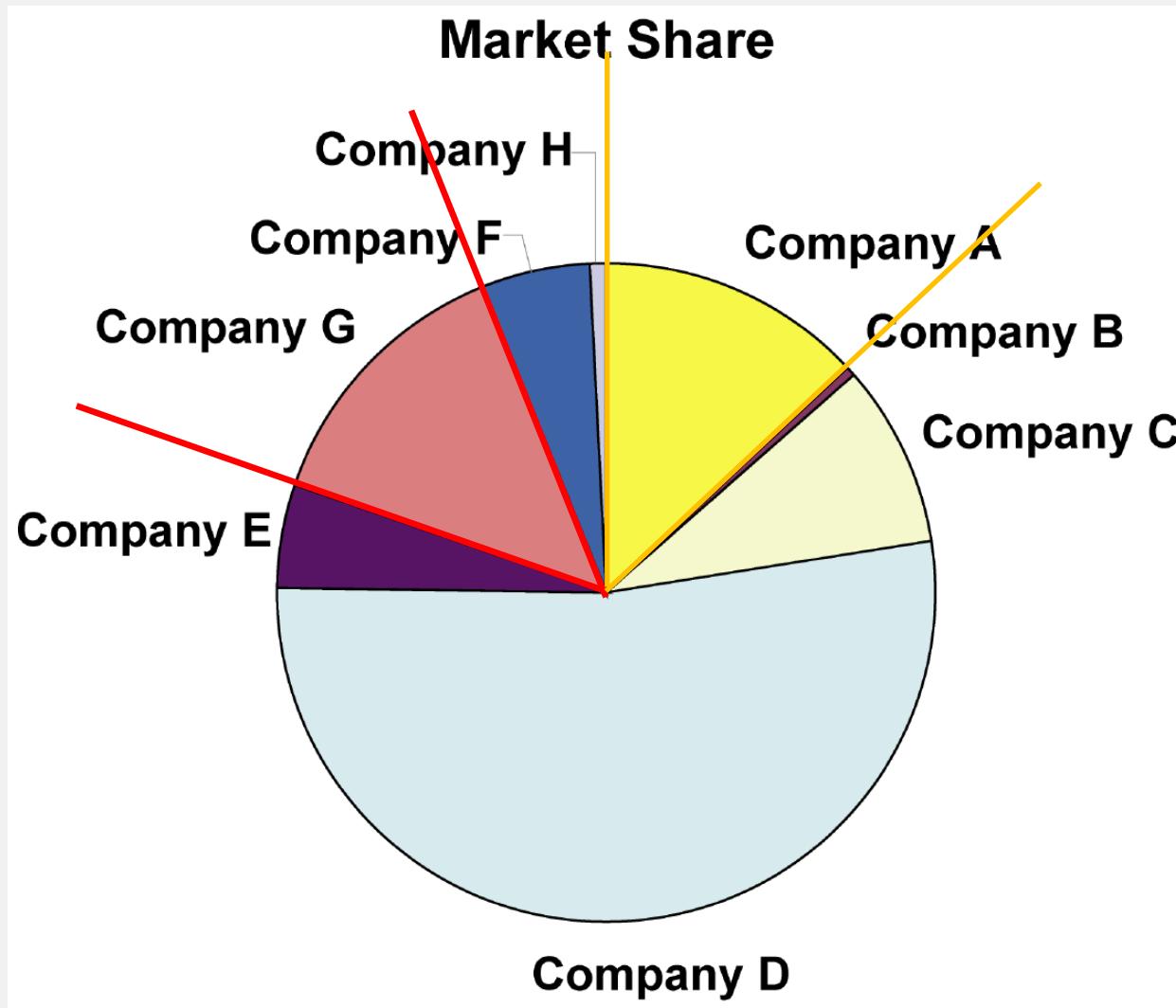
Company
E

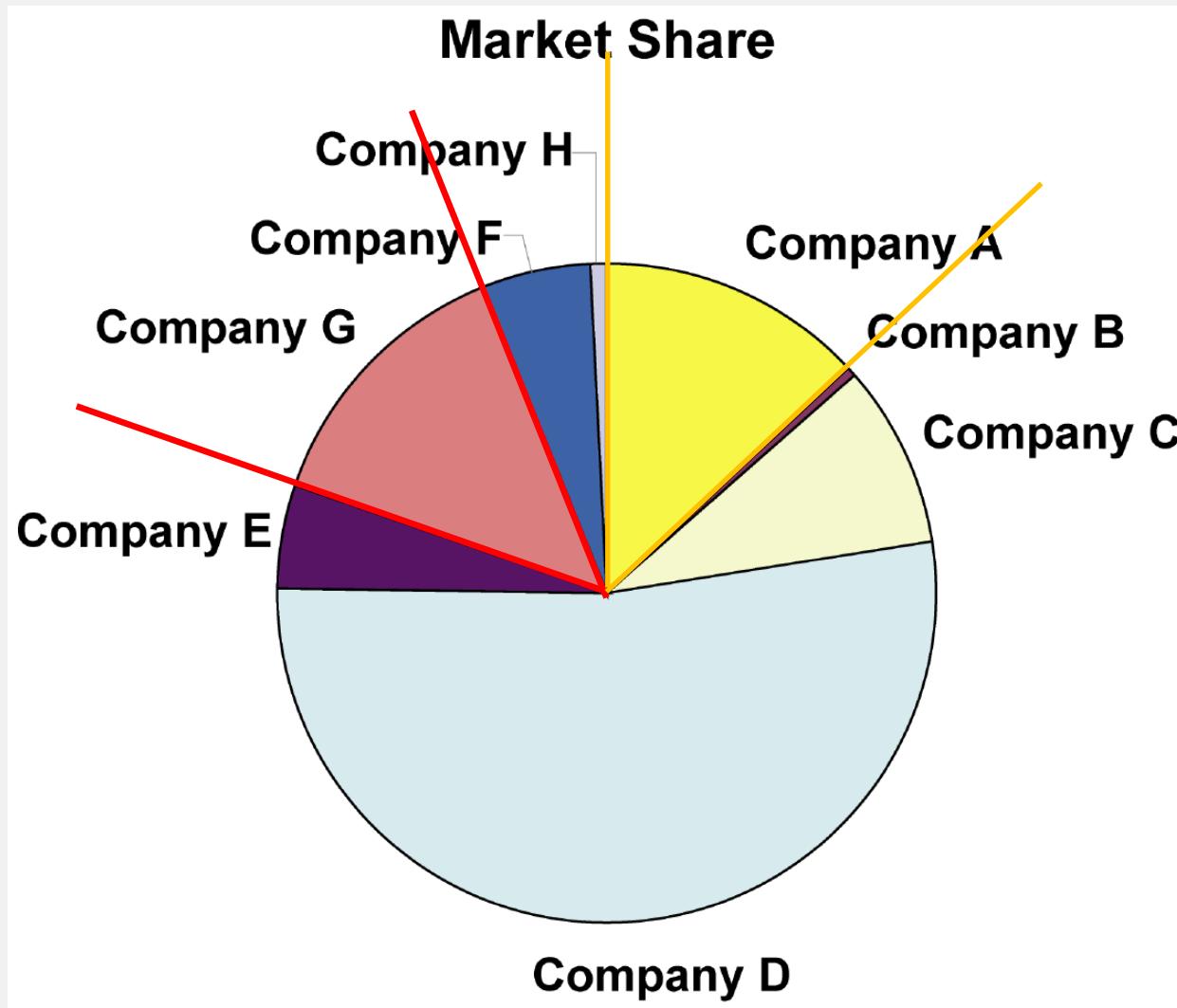
Market Share



Company D

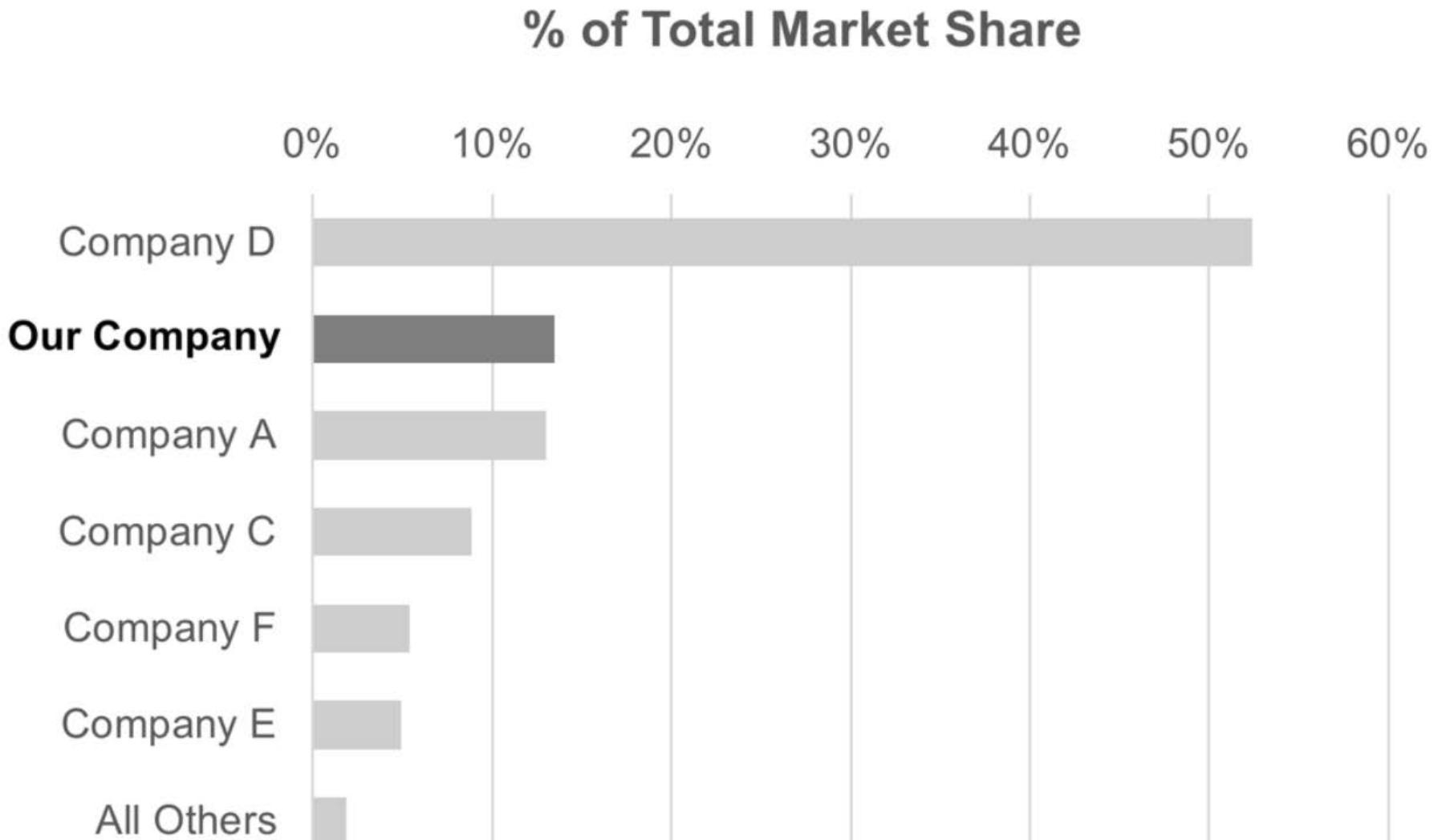






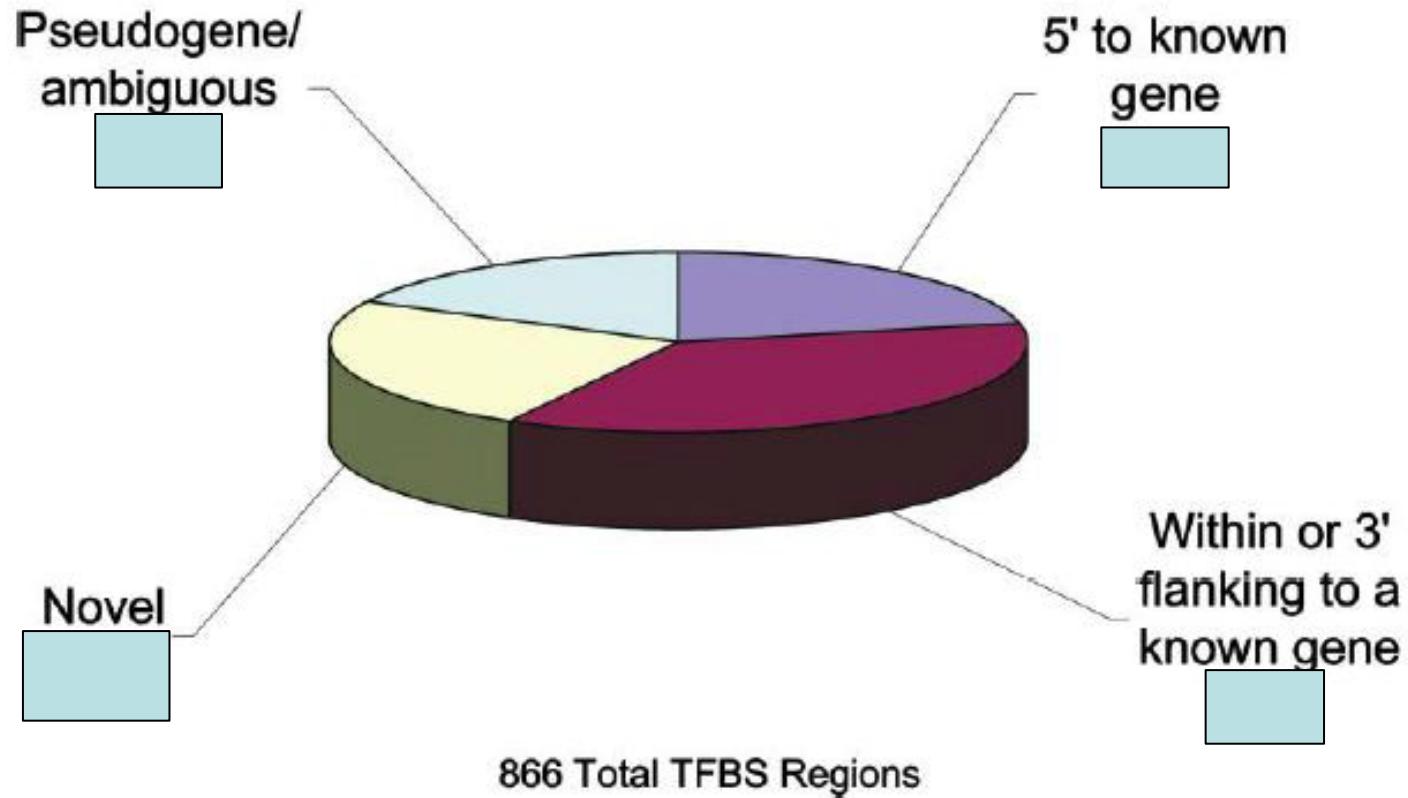
WE ARE BETTER!!!!!!

I'm tired but we are better!!

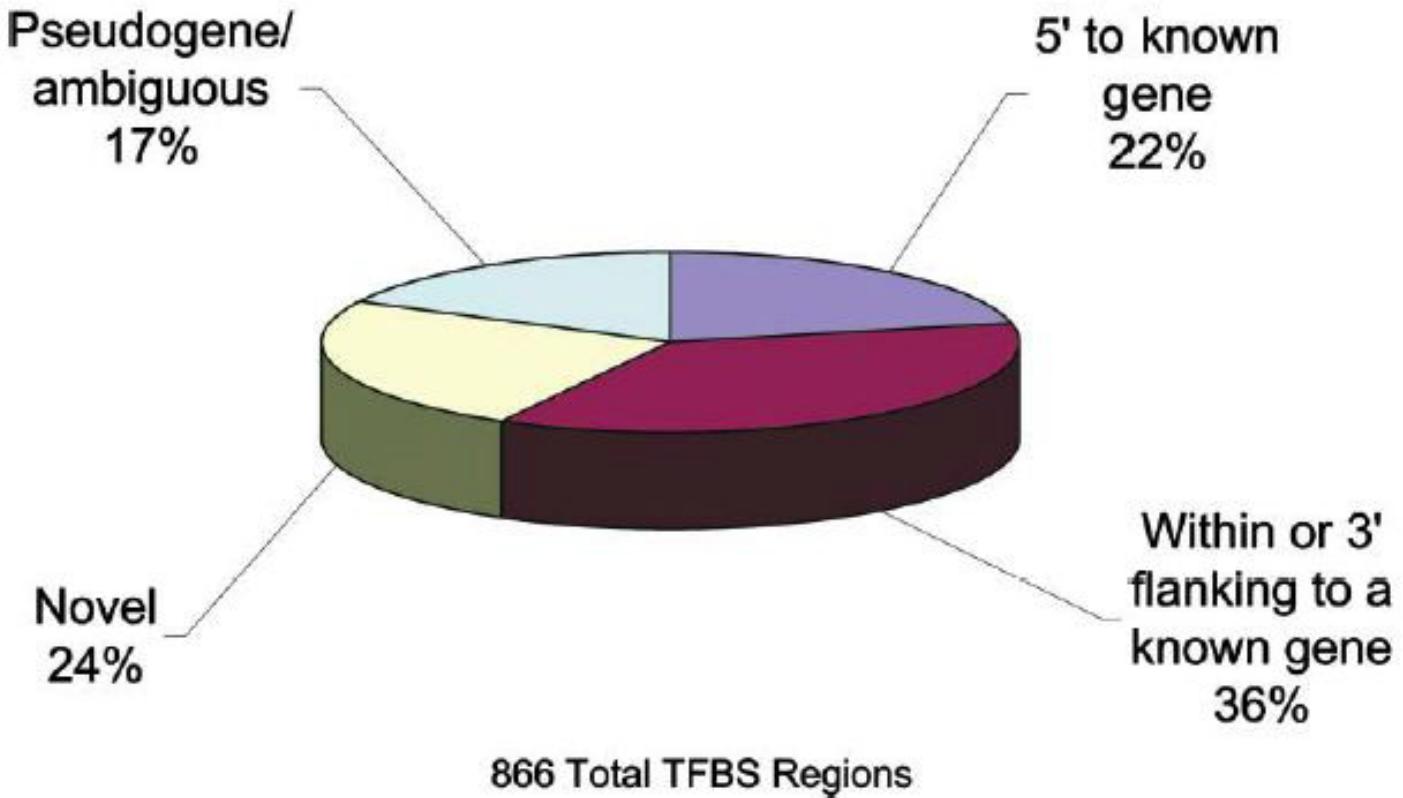


HOW MUCH ARE THE RATES HERE???

Distribution of All TFBS Regions

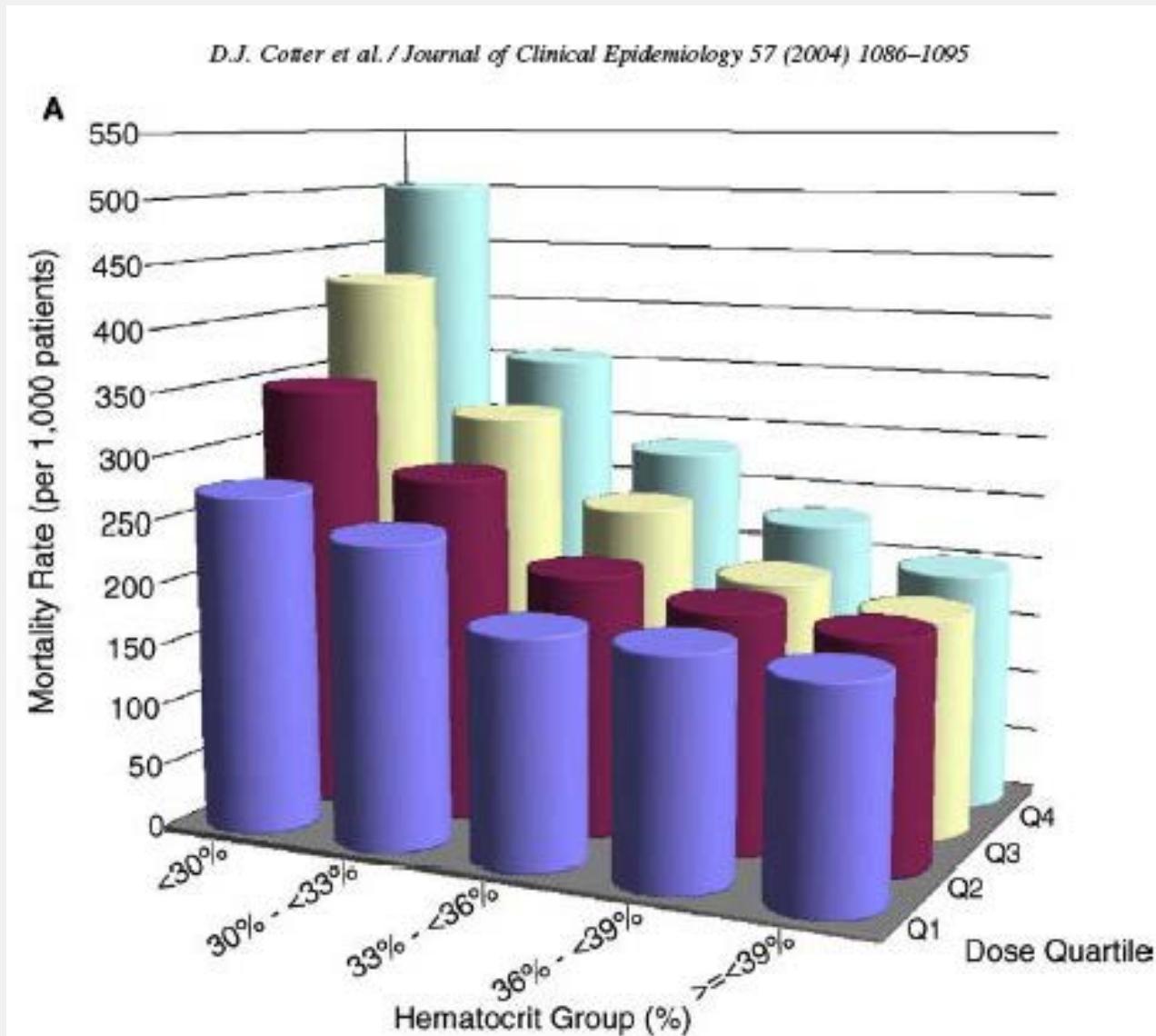
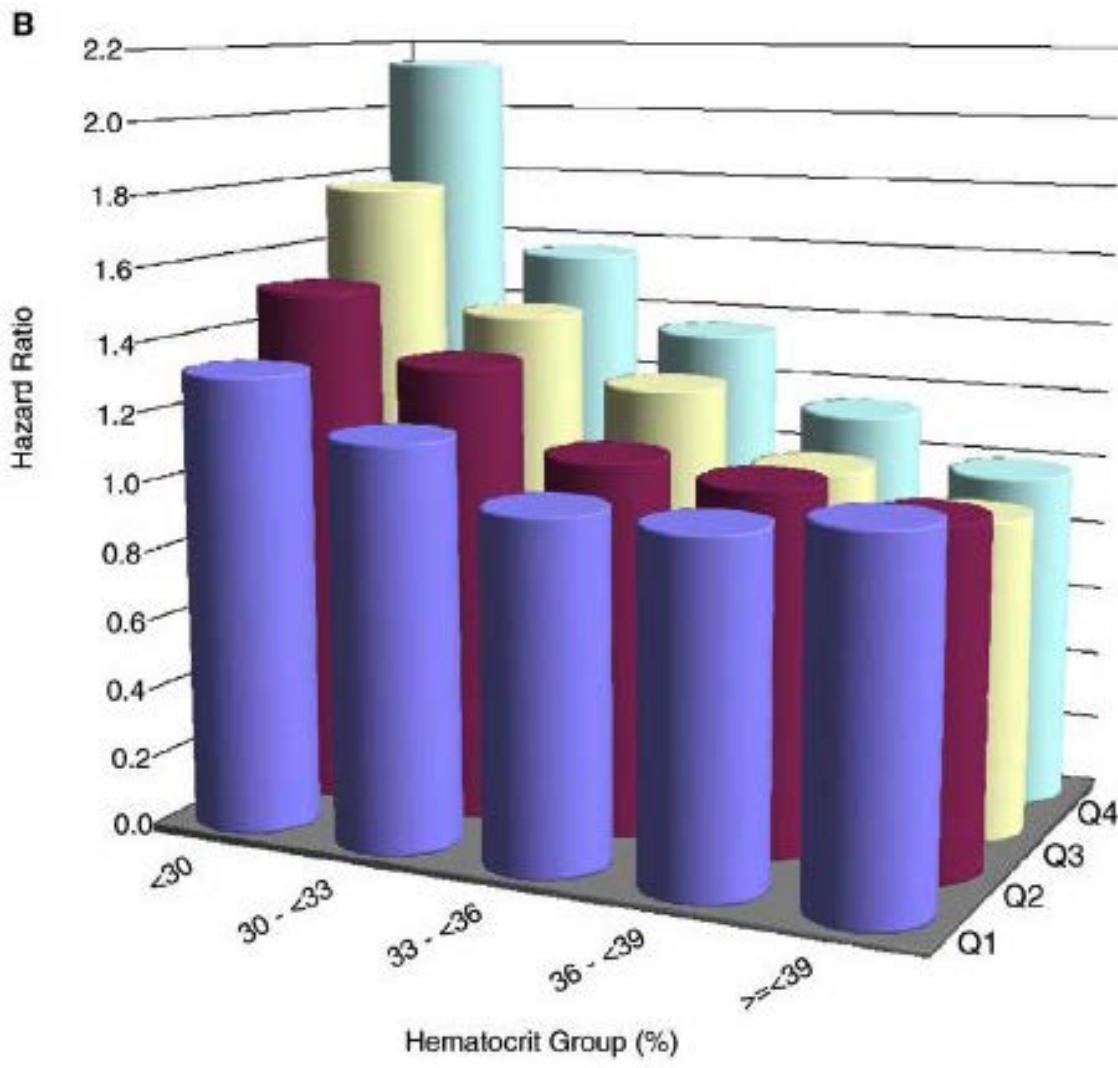


Distribution of All TFBS Regions





3D is almost never good: if there's a difference you hardly see it



"The birth of a word"

[The birth of a word](#) (Deb Roy)

Some 3D data may be mapped to 2D

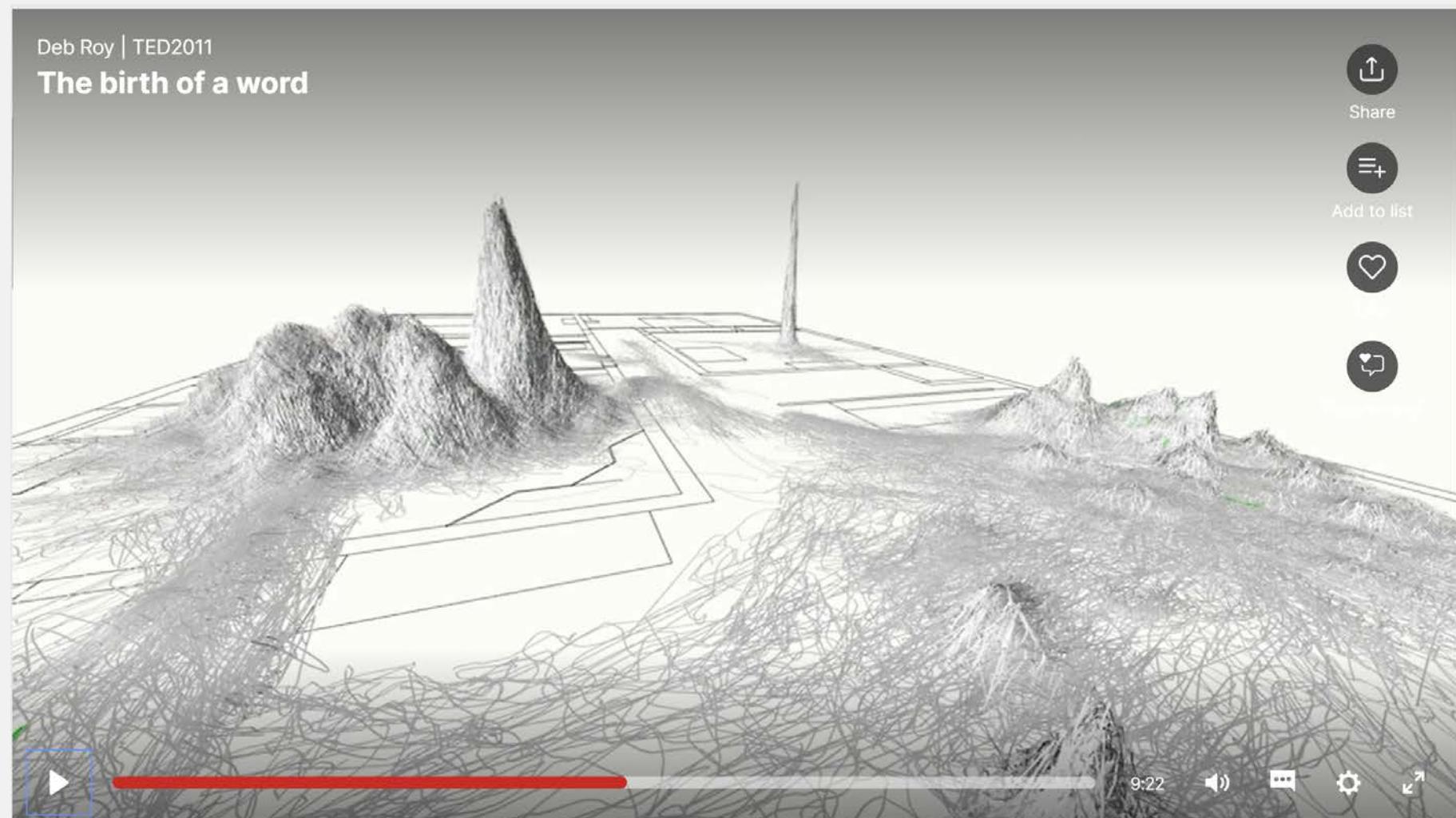
Space-time worms: show the movement of two persons in a 3D space

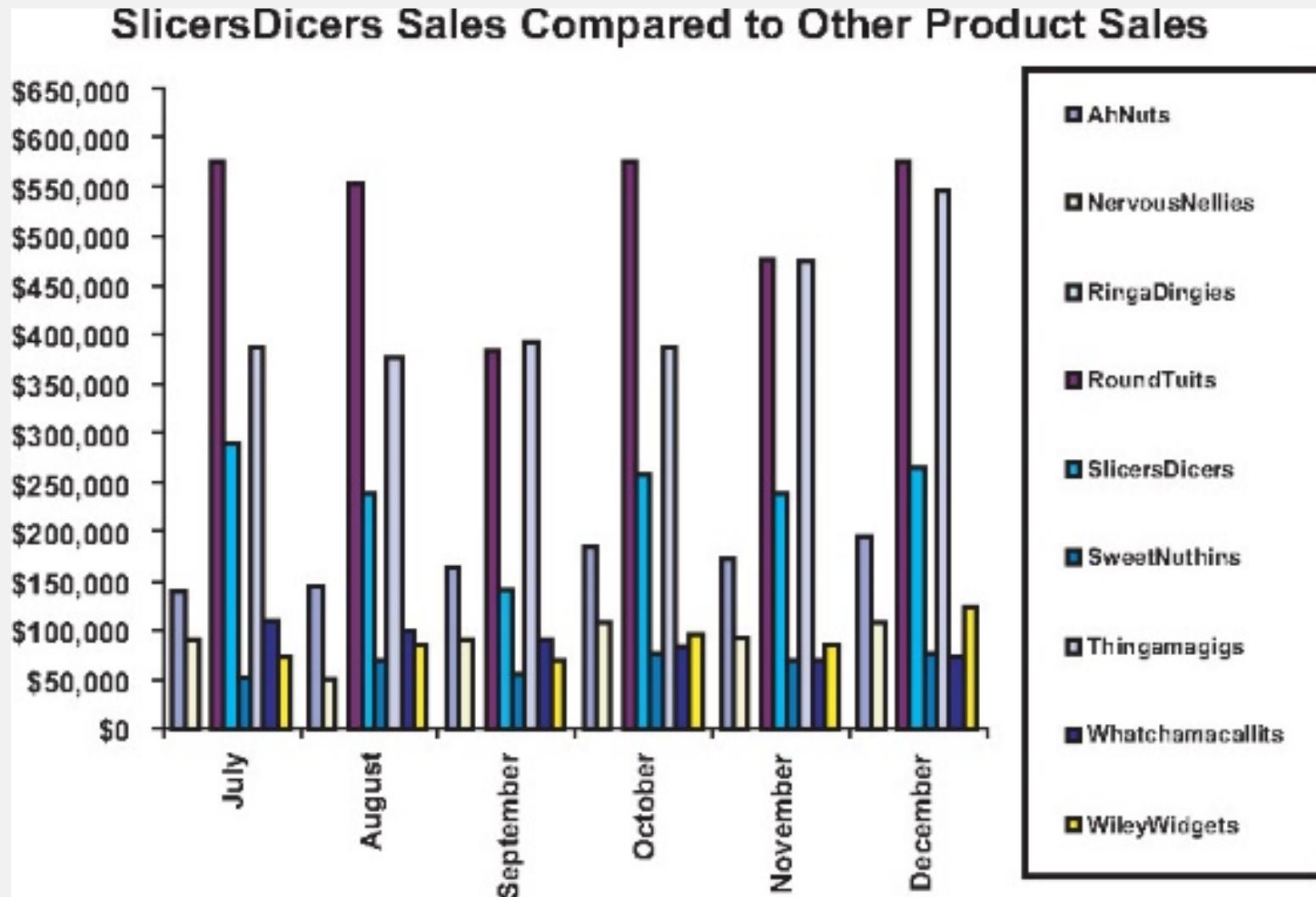




And the 2D data may be summed up to get novel 3D representations which allow analysis

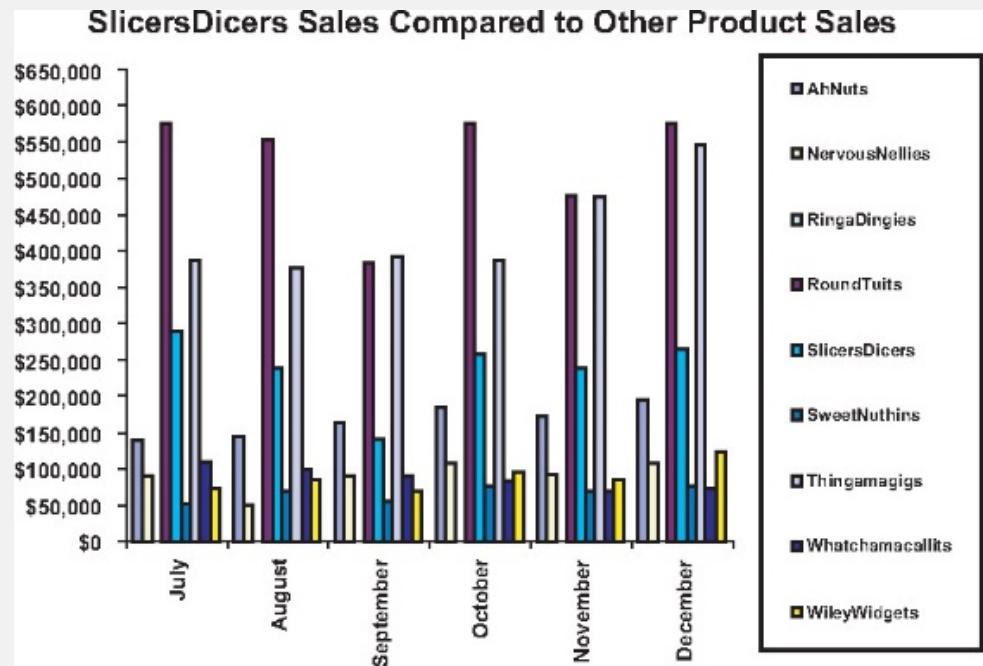
Wordscape: accumulate all the space-time worms related to different persons and sum them. Space-time worms are accumulated when people say a word (in this case the word is "water")





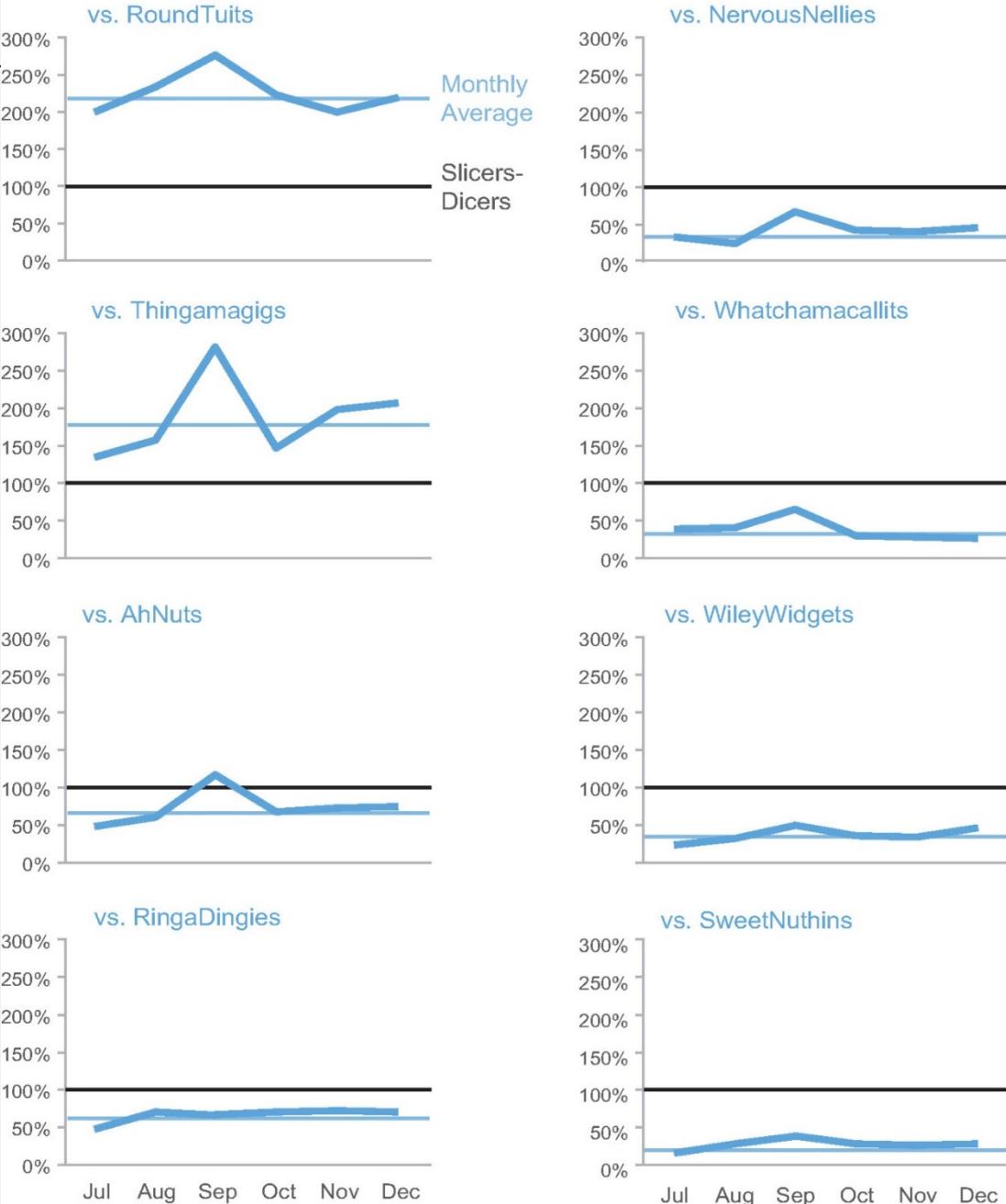
SO NICE TO GO BACK AND FORTH from the legend to the plot!

What about the trend?



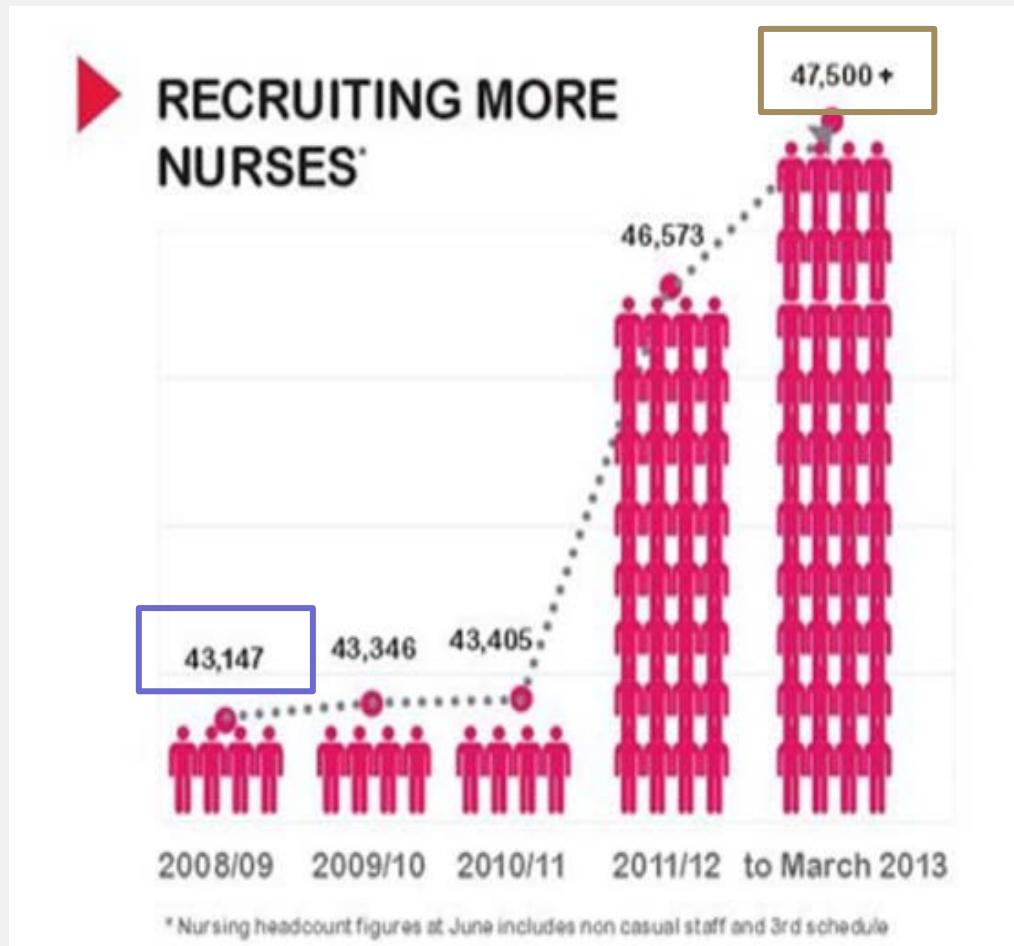
Sales of SlicersDicers Compared to Sales of Other Products
July - December, 2011

Sales of SlicersDicers Compared to Sales of Other Products July - December, 2011





DO NOT LIE



Lowest number/maximum Number =

$$\frac{43147}{47500} = 0.91 \text{ (91\%)}$$

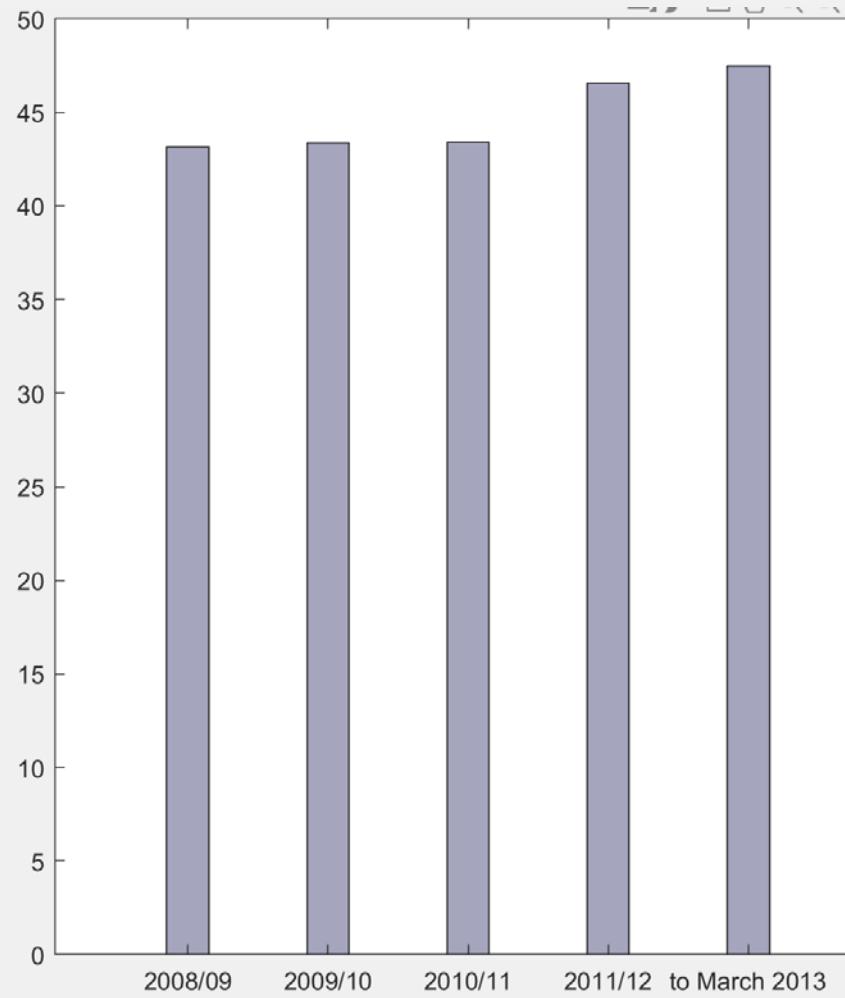


THE RATIO YOU SEE (YOU ARE SOMEHOW PERCEIVING IS):



Ratio you see =

$$0.47'' / 3.5'' = 0.13 \text{ (13\%)}$$



THIS PLOT IS REALISTIC

and

AUGMENTS THE **DATA/INK RATIO**



Too few data in a graph: is the graph really useful?

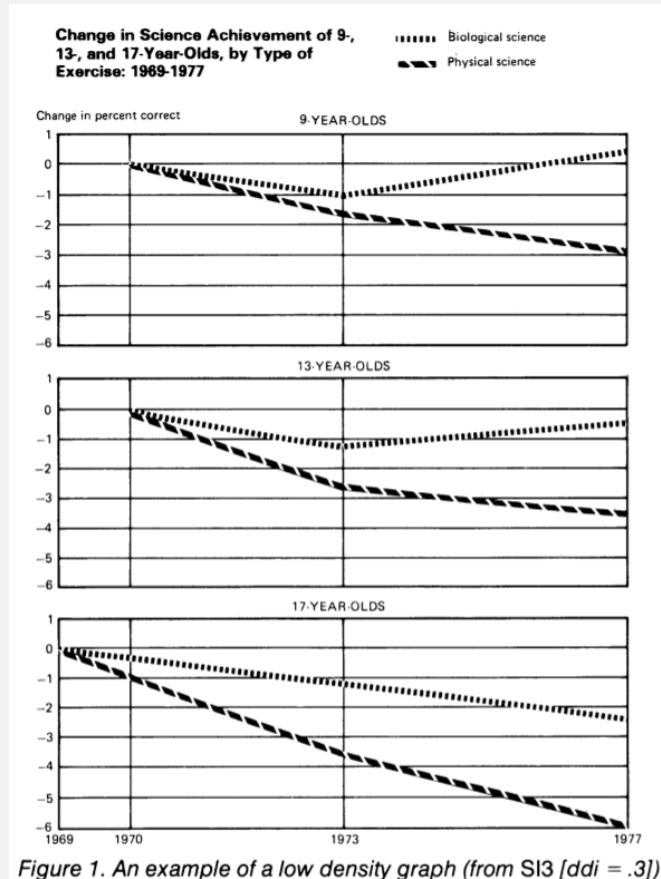


Figure 1. An example of a low density graph (from SI3 [$ddi = .3$]).

headache...

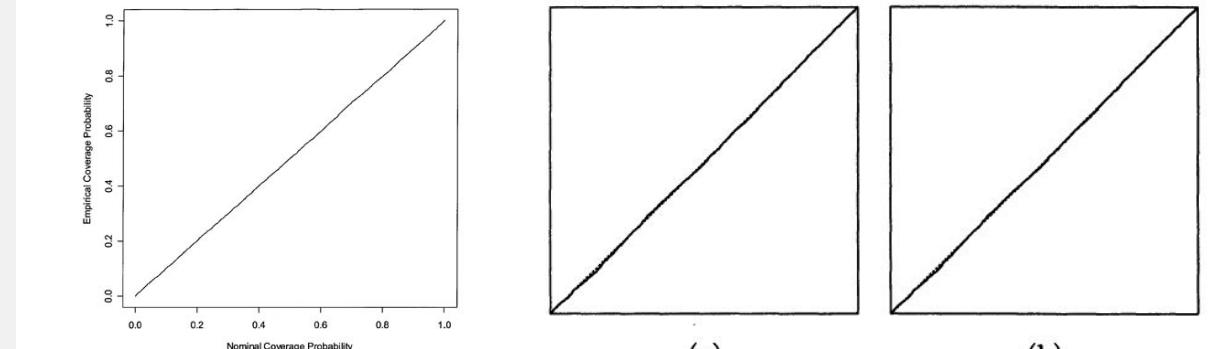


Figure 1. Empirical coverage of CIs for the relative-risk parameter β of haplotype 01100. Results are based on 10,000 simulated data sets with the same haplotype frequencies as the FUSION data. Haplotype 01100 has a multiplicative effect on disease risk, with $\beta = 0.35$.

Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

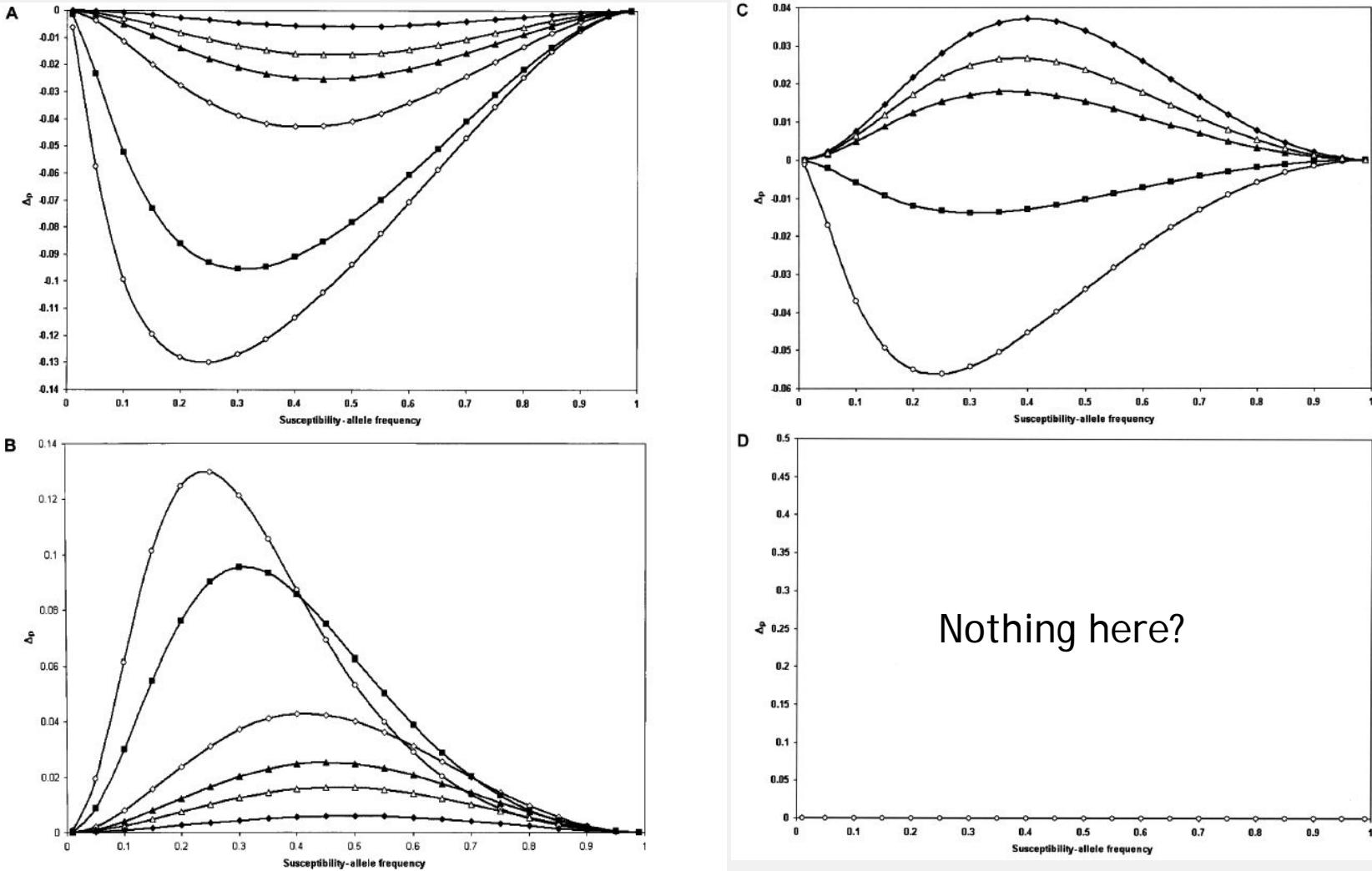


Figure 1 Δ_p plotted versus the susceptibility-allele frequency for patients. A, B, and D, Data points are as follows: $\gamma = 1.1$ (blackened diamonds), $\gamma = 1.3$ (unblackened triangles), $\gamma = 1.5$ (blackened triangles), $\gamma = 2$ (unblackened diamonds), $\gamma = 5$ (blackened squares), and $\gamma = 10$ (unblackened circles). A, Dominant model. B, Recessive model. C, Additive model. Since $\gamma < 2$ would not satisfy our definition of an additive model as $\gamma = 2\beta$ and $\beta > 1$, the data points in C are as follows: $\gamma = 2.2$ ($\beta = 1.1$) (blackened diamonds), $\gamma = 2.6$ ($\beta = 1.3$) (unblackened triangles), $\gamma = 3$ ($\beta = 1.5$) (blackened triangles), $\gamma = 5$ (blackened squares), $\gamma = 2$ (unblackened diamonds). D, Multiplicative model.



GRAPHS MAY HIDE DATA.

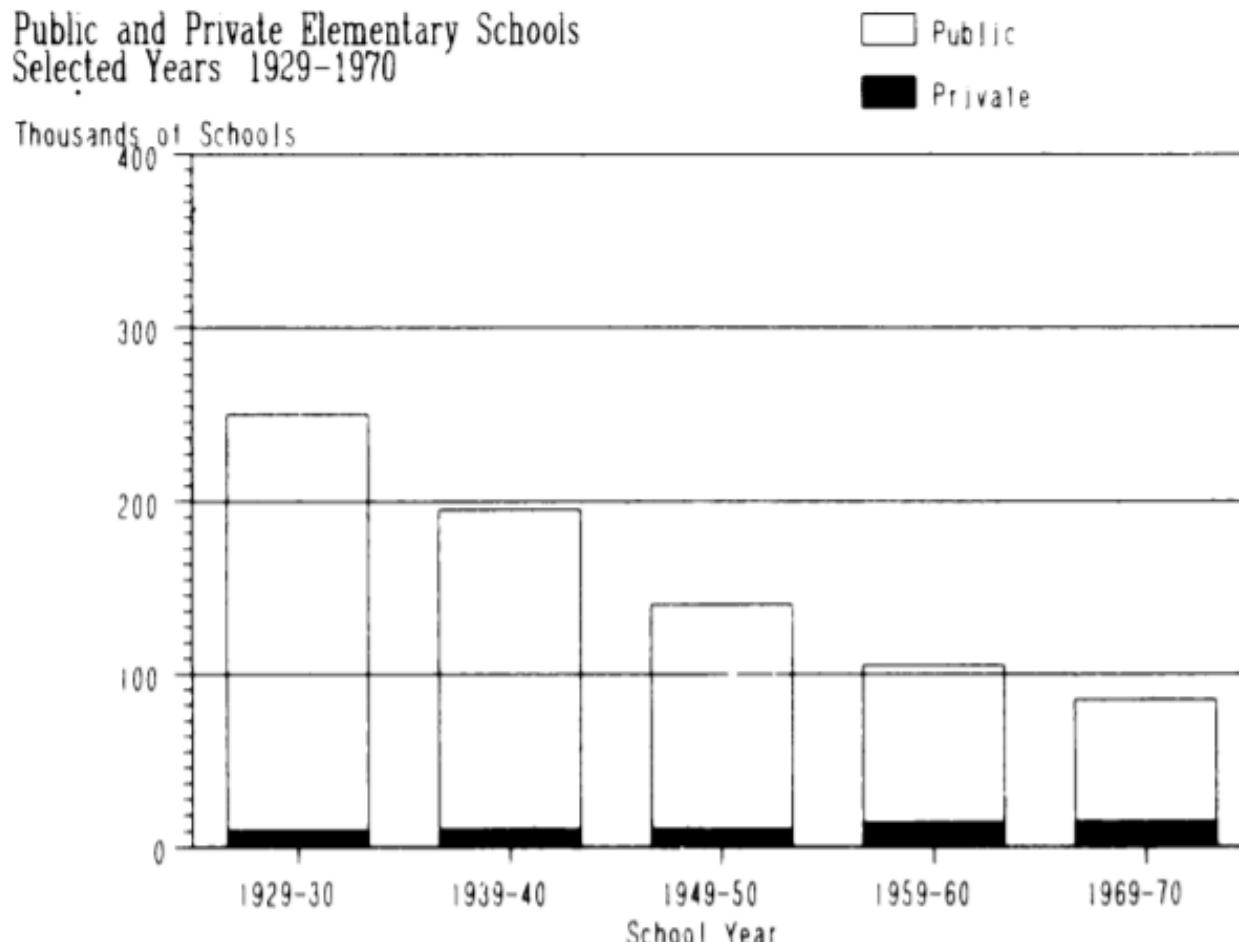


Figure 4. Hiding the data in the scale (from SI3).

What do you see?

Public Schools keep on decreasing.

Private Schools are few though, and their number seems stable

Only notable change is in the decrease of the public education?



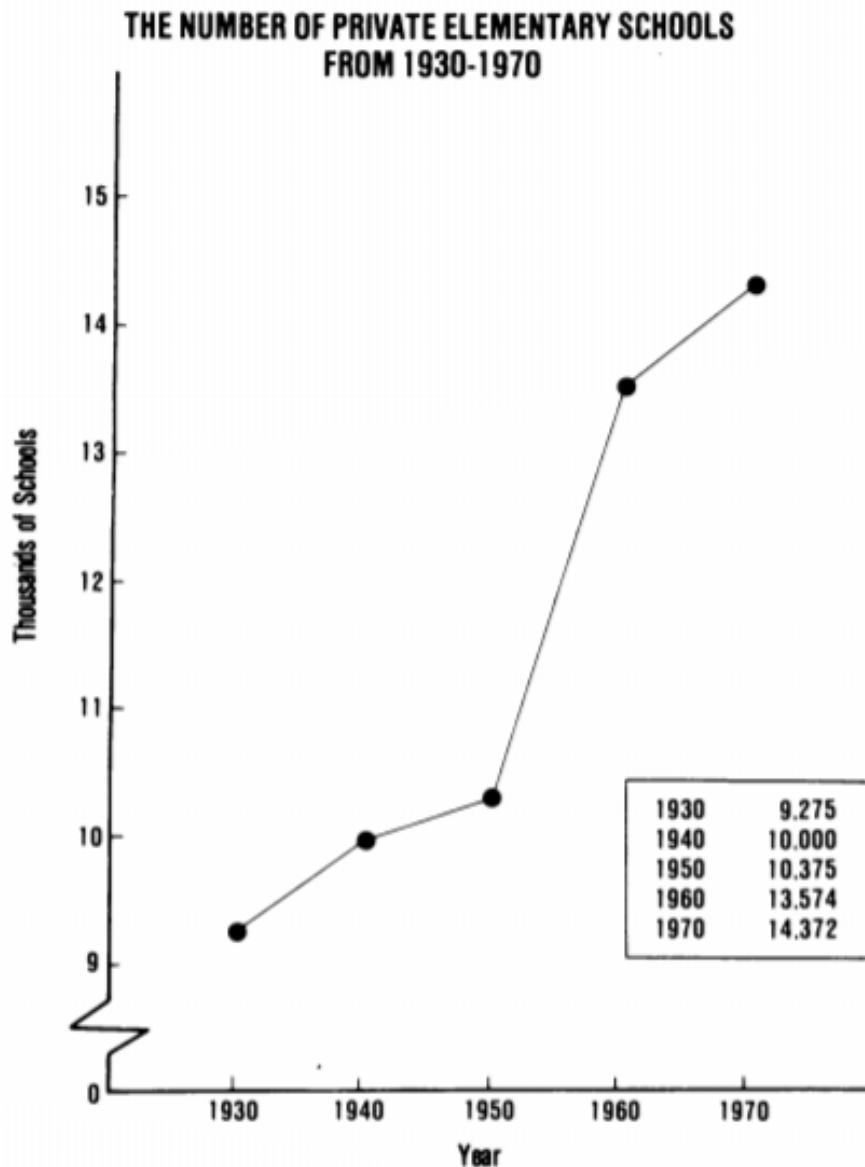


Figure 5. Expanding the scale and showing the data in Figure 4 (from SI3).

Increasing the scale shows a dramatic increase of Private education.

Note that numbers from 0 to 8 (Y axis) are not shown.

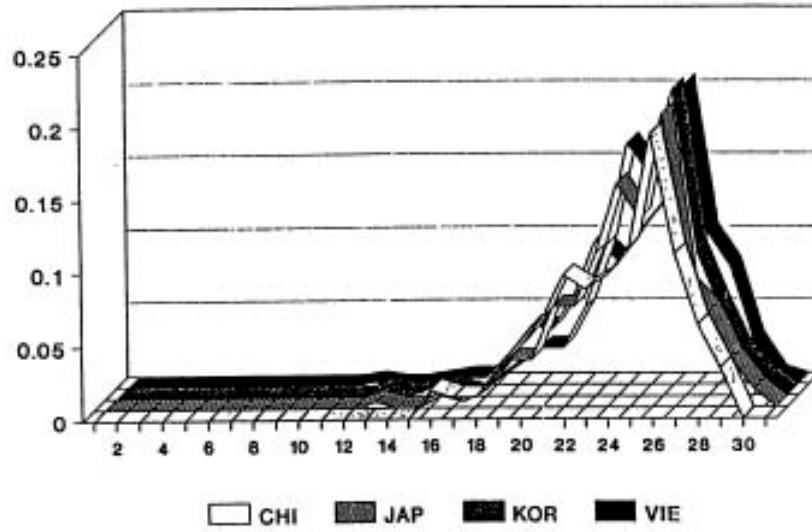
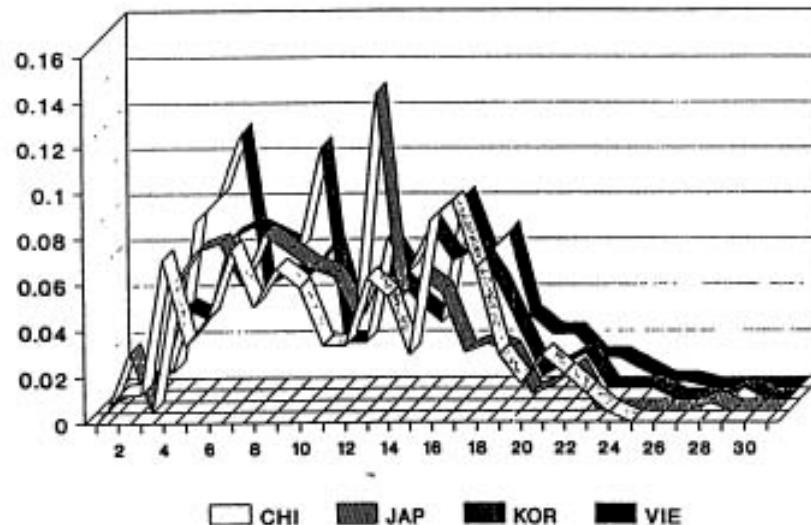
IT'S OK!

Lines: show trends

Points: allow perceiving distributions

Shapes: quantities!



**A**BINNED FREQUENCY DATA - D4S139
CHINESE, JAPANESE, KOREAN & VIETNAMESE**B**BINNED FREQUENCY DATA - D10S28
CHINESE, JAPANESE, KOREAN, VIETNAMESE

Here 3D texture is only visual clutter!

INSTEAD OF USING 3D:

- Line plots with different colors
- Or, if you must print black and white, different line styles (dashed, dash-dot, grey for less important lines, ...)



FOR TRENDS AND TIMELINES USE LINEPLOTS OR POINT PLOTS

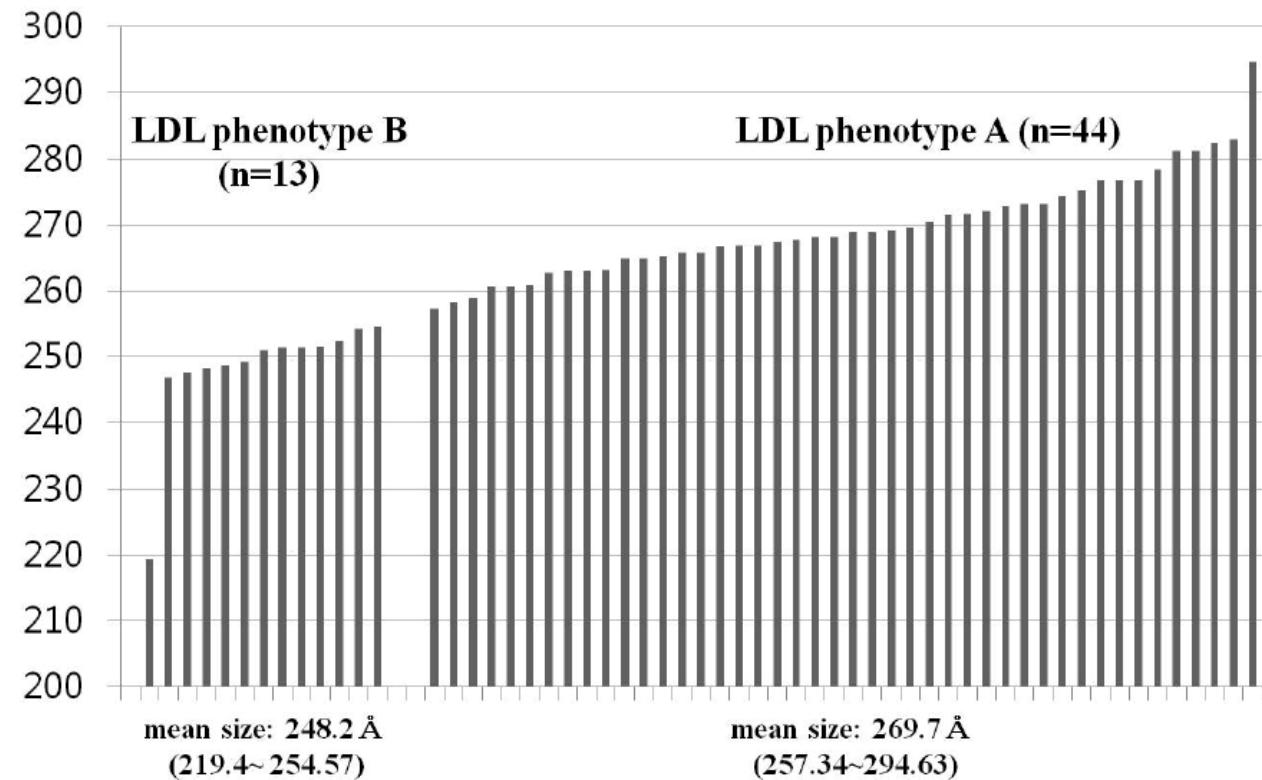


Fig. 1. Distribution of low-density lipoprotein (LDL) particle size in all study subjects (LDL phenotypes A and B). *LDL phenotype A group* (mean size: 269.7 Å, n = 44), subjects with buoyant-mode profiles [peak LDL particle diameter \geq 264 Å] including intermediate LDL subclass pattern [256 Å \leq peak LDL particle diameter \leq 263 Å]; *LDL phenotype B group* (mean size: 248.2 Å, n = 13), subjects with dense-mode profiles [peak LDL particle diameter \leq 255 Å]

Wrong Y axis and
Visual clutter/chart junk





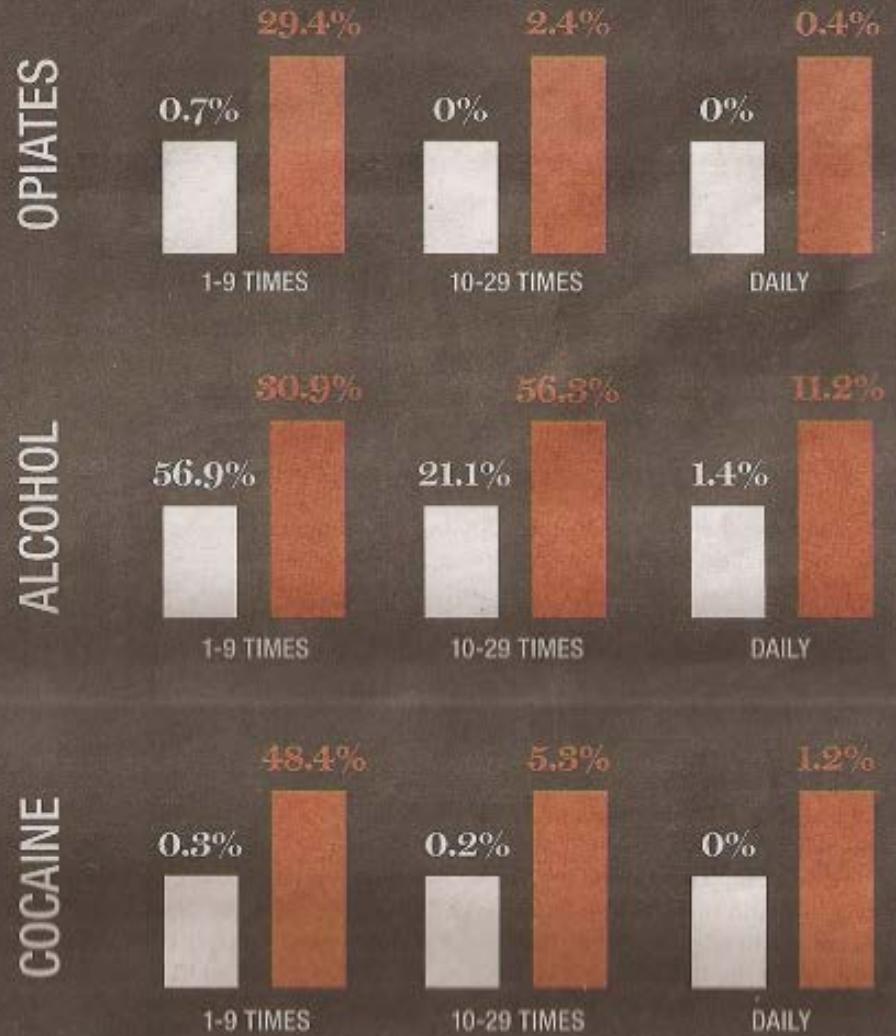
Wainer H (1984) [How to display data badly](#). *The American Statistician* 38:137-147



Bad Infographics

BY THE NUMBERS

The National Collegiate Health Assessment was taken by 1,000 UCSB students in Spring 2009. Participants were asked how frequently they used substances over the past 30 days. Numbers in white reflect actual student use, while red numbers indicate perceived substance use. The average age of participants was 20 years and approximately 99 percent were full-time students.



WHERE WE DONATE VS. DISEASES THAT KILL US

Heart Disease	Jump Rope for Heart (2013)	Suicide	Out of Darkness Overnight Walk (2014)
Diabetes	Step Out: Walk to Stop Diabetes (2013)	HIV / AIDS	Ride to End Aids (2013)
Motor Neuron Disease (including ALS)	ALS Ice Bucket Challenge (2014)	Breast Cancer	Komen Race for the Cure (2012)
Chronic Obstructive Pulmonary Disease	Fight for Air Climb (2013)	Prostate Cancer	Movember (2013)



Source: CDC (2011)

Gun control in America: A state-by-state breakdown

Laws on file

If no colour appears, there is no such law on file

- 2012 election results
- Background check law
- Permit required to purchase
- Licence required to sell
- Records kept on file
- Firearms banned from workplace

Virginia

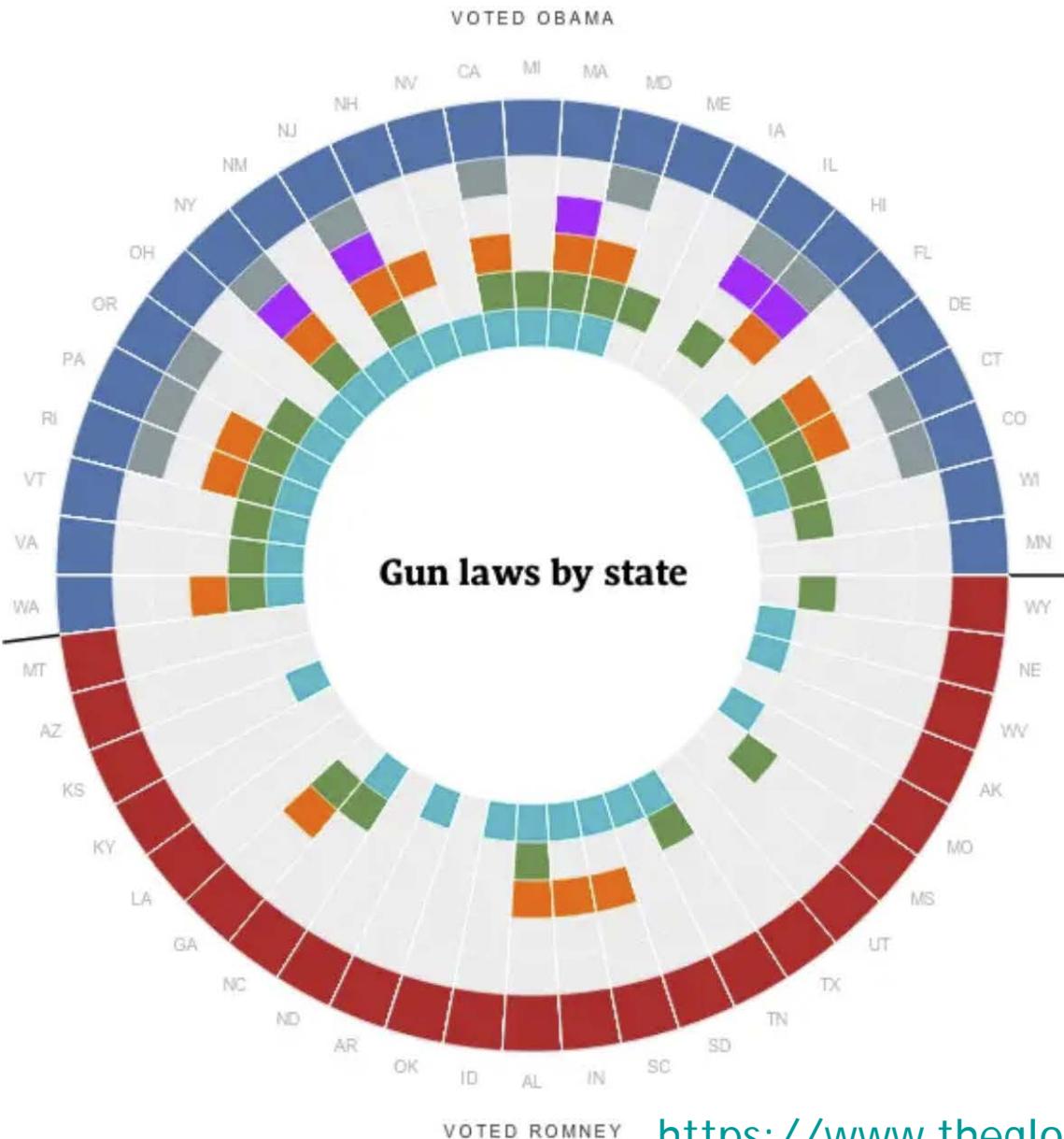
- Voted for Obama in the 2012 election
- **Background check:** not required for handguns
- **Permit:** not required to buy firearms
- **Licence:** not required for dealers
- **Records:** kept on file for handgun owners
- **Workplace:** firearms not allowed in parking lots

Overall gun control score: 12

Virginia has a [Brady Campaign score](#) of 12, which is lower than the national average of 16. The score comes from measuring these and other gun laws according to a weighted points system.

Murder rate: 2.58

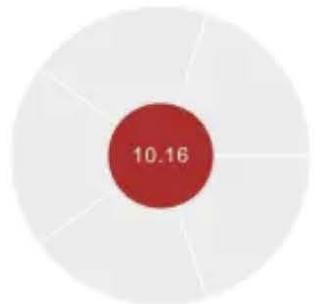
There were 2.58 firearm murders per 100,000 people in Virginia during 2011, which is lower than the national average of 2.77. Overall, it is ranked #27 in murder rates out of 48 states with this data.



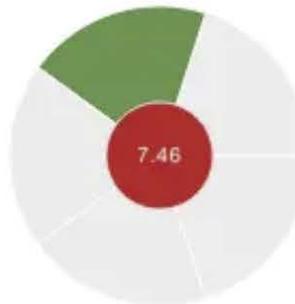


States with the highest firearm murder rate

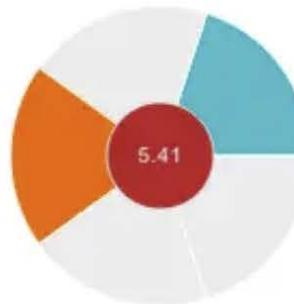
Louisiana scored only two points on the Brady scale for banning guns from college campuses. It also has the highest firearm murder rate per 100,000 people in the country. Overall, Republican states have an average Brady score of 4.6, compared to 26.73 for states that voted for President Obama in the last election.



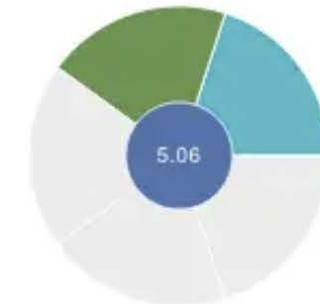
1. Louisiana



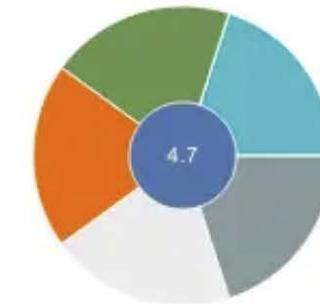
2. Mississippi



3. South Carolina



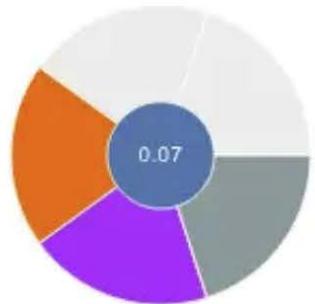
4. Michigan



5. Maryland

States with the lowest firearm murder rate

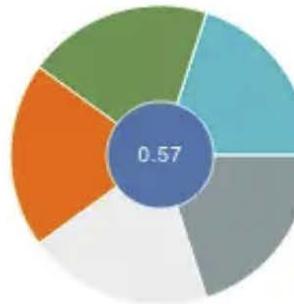
Hawaii has the lowest firearm murder rate in the United States with just 0.07 murders per 100,000 people. South Dakota is the only Republican state to rank on this list. Despite scoring only 7 points on the Brady score and enacting none of the laws highlighted on this chart, Iowa still has one of the lowest firearm murder rates in the country.



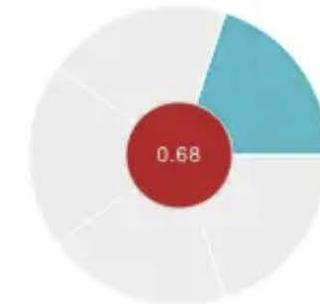
1. Hawaii



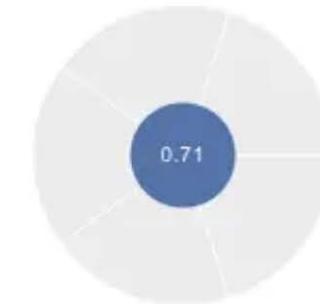
2. New Hampshire



3. Rhode Island



4. South Dakota



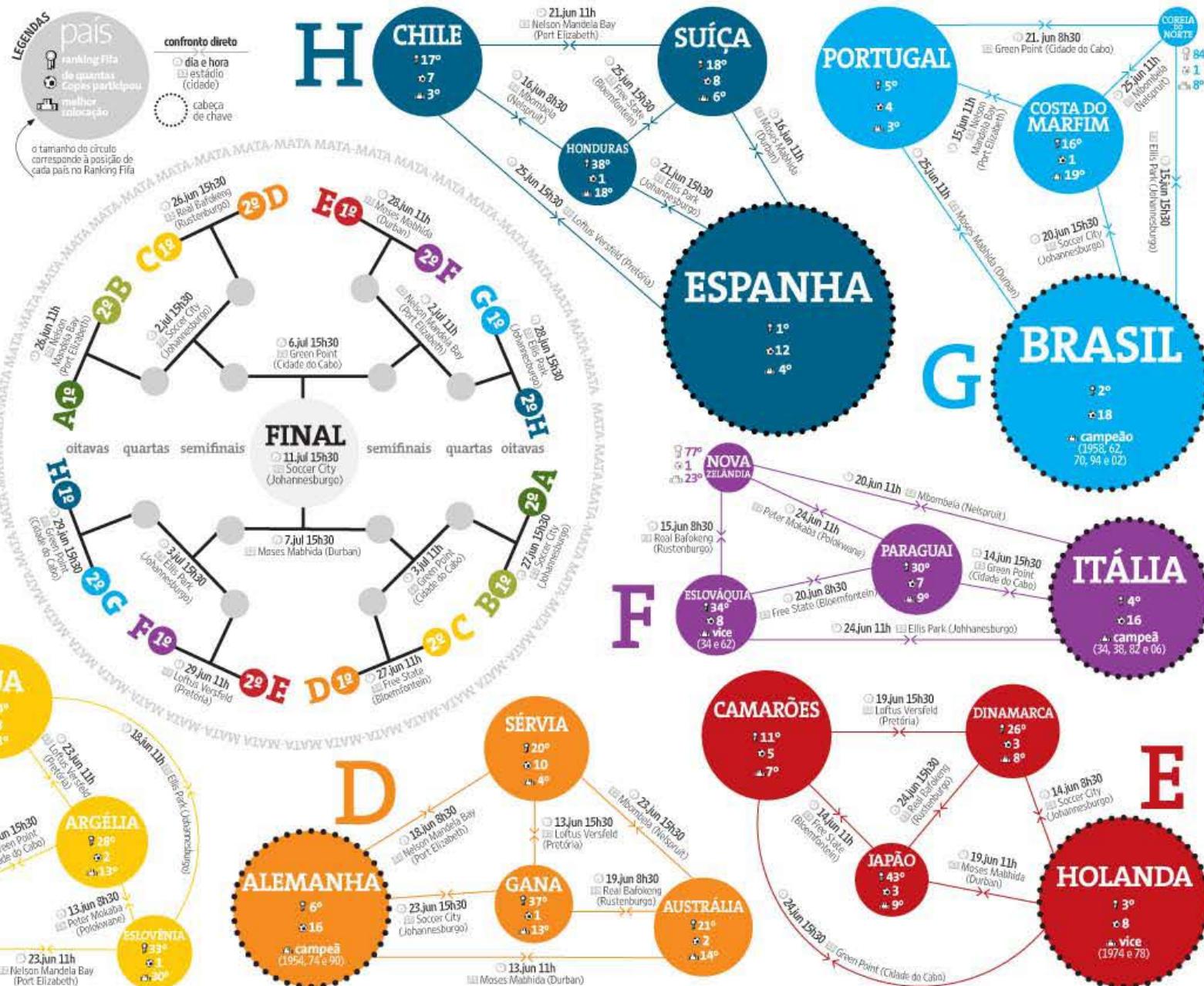
5. Iowa



Folha de São Paulo. This special World Cup Games Table was AWARD OF EXCELLENCE by SND The Best of Newspaper Design 2010.

Todos os grupos e jogos da Copa

Confira a força de cada seleção, quem enfrenta quem na primeira fase e como será o mata-mata na África do Sul



LEGENDAS

- país
- ranking Fifa
- de quantas Copas participou
- melhor colocação
- confronto direto
- dia e hora
- estádio (cidade)
- cabeça de chave

o tamanho do círculo corresponde à posição de cada país no Ranking Fifa



WHAT'S DATA/INK RATIO?

WHAT'S VISUAL CLUTTER?

WHAT'S CHART JUNK?

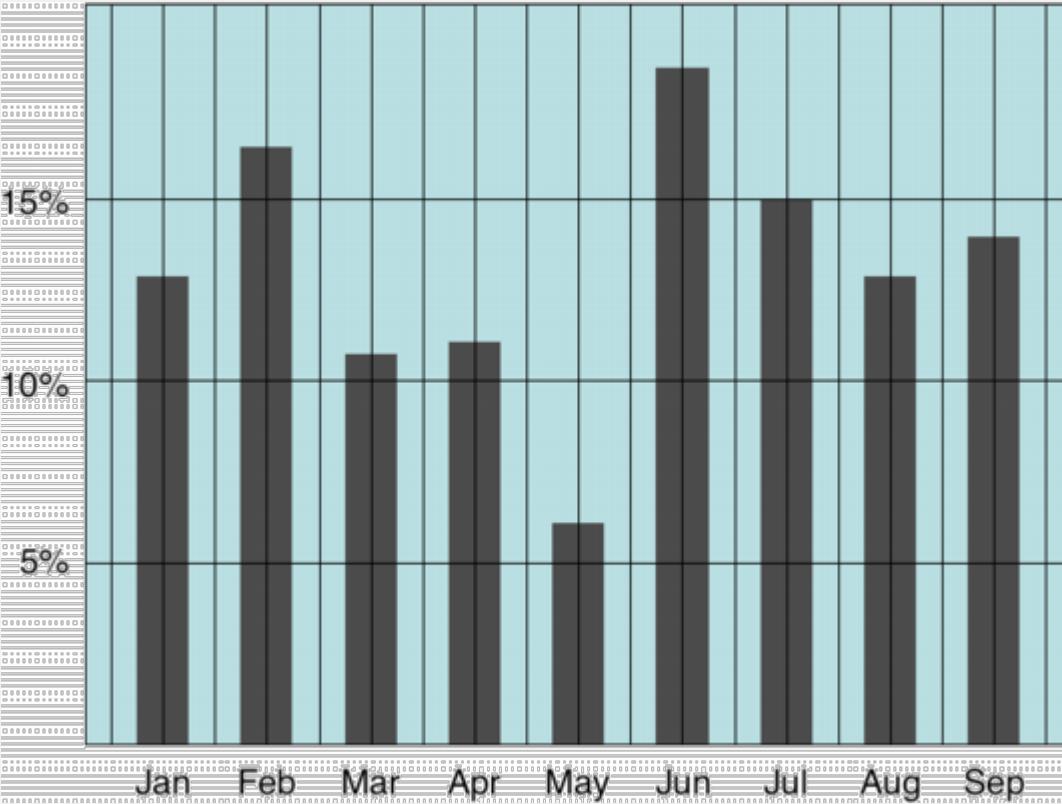


Data-ink ratio =
$$\frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

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

= 1.0 – proportion of a graphic that can be erased





Remove border,

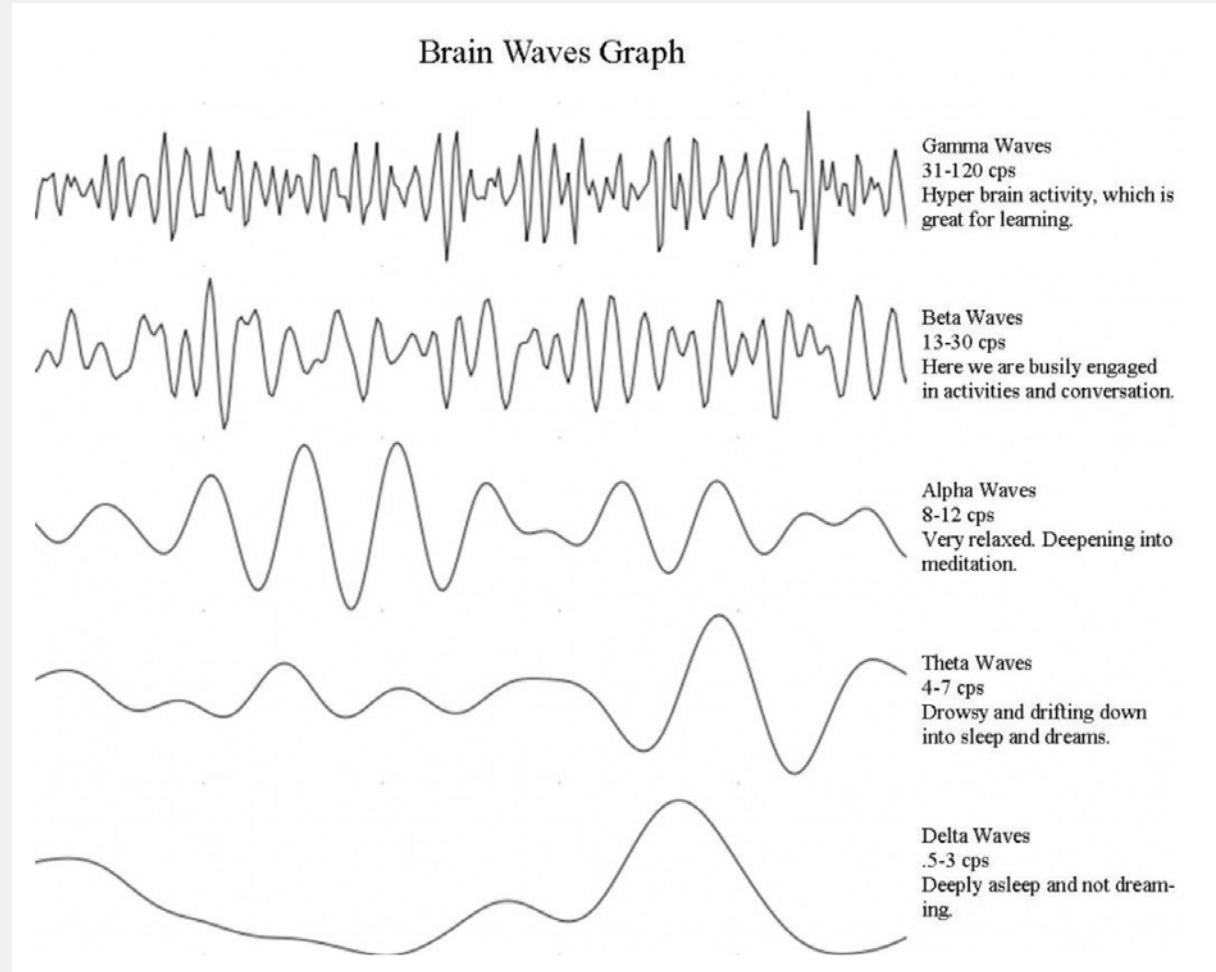
Remove background color

Remove grid lines and eventually draw (grey)
only those helpful for looking up numbers.

DATA-INK PRINCIPLES:

1. Above all else show data
2. Maximize the data-ink ratio
3. Erase non-data-ink (visual clutter)
4. Erase redundant data-ink (visual clutter)
5. Revise and edit

Let's See It in practice



High data-ink ratio



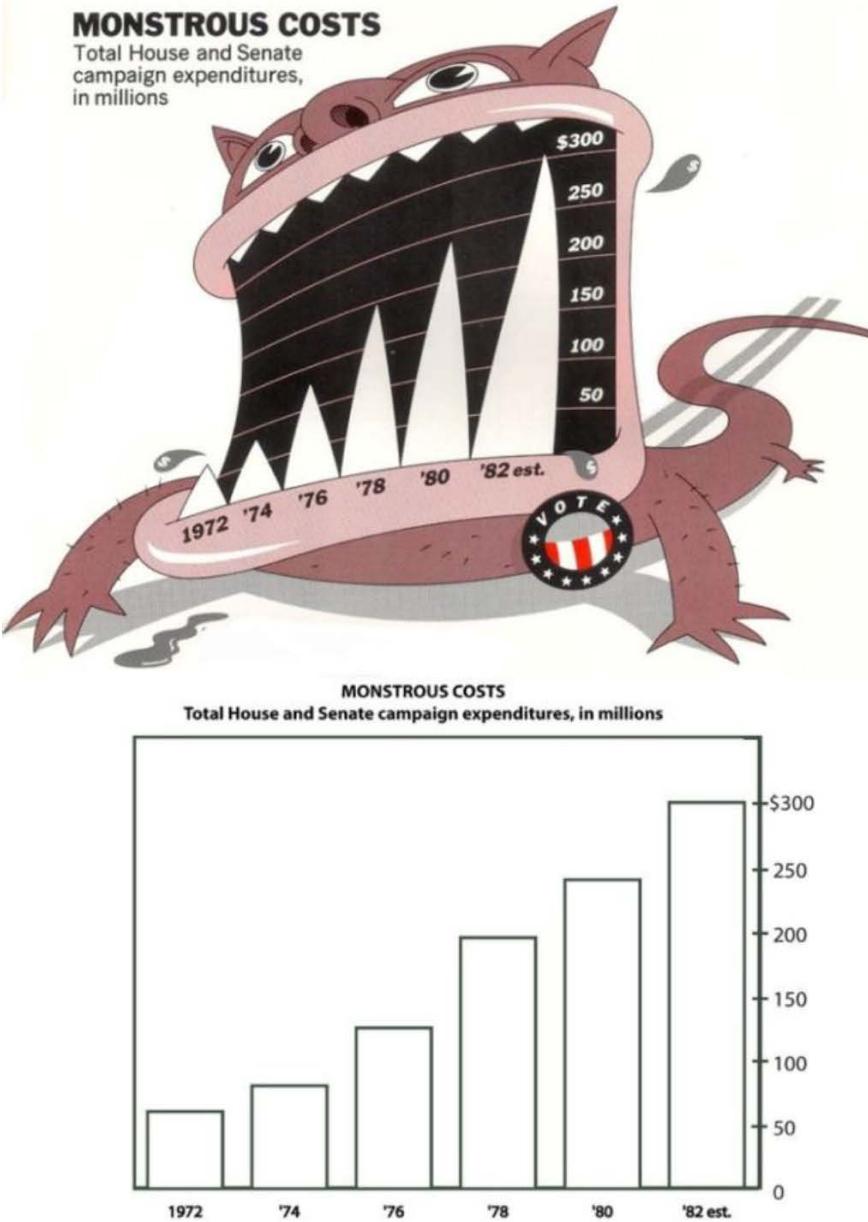


Figure 1. A chart by Holmes [7] (above), and a ‘plain’ version.

[7] Holmes, N. *Designer’s Guide to Creating Charts and Diagrams*, Watson-Guptill Publications, 1984.

CHARTJUNK: the excessive and unnecessary use of graphical effects in graphs.

Though sometimes some artistic view may help interpretation accuracy and long-term recall

[Useful Chart Junk](#)

[NEUVIS](#): a set of guidelines for creative practitioners developing visualizations for Non-Expert Users.





TUFTÉ'S GRAPHICAL INTEGRITY: 6 PRINCIPLES

1. The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented (**DON'T LIE WITH SCALES**)

2. Show data variation, not design variation (**AVOID TOO ARTISTIC EFFECTS**)

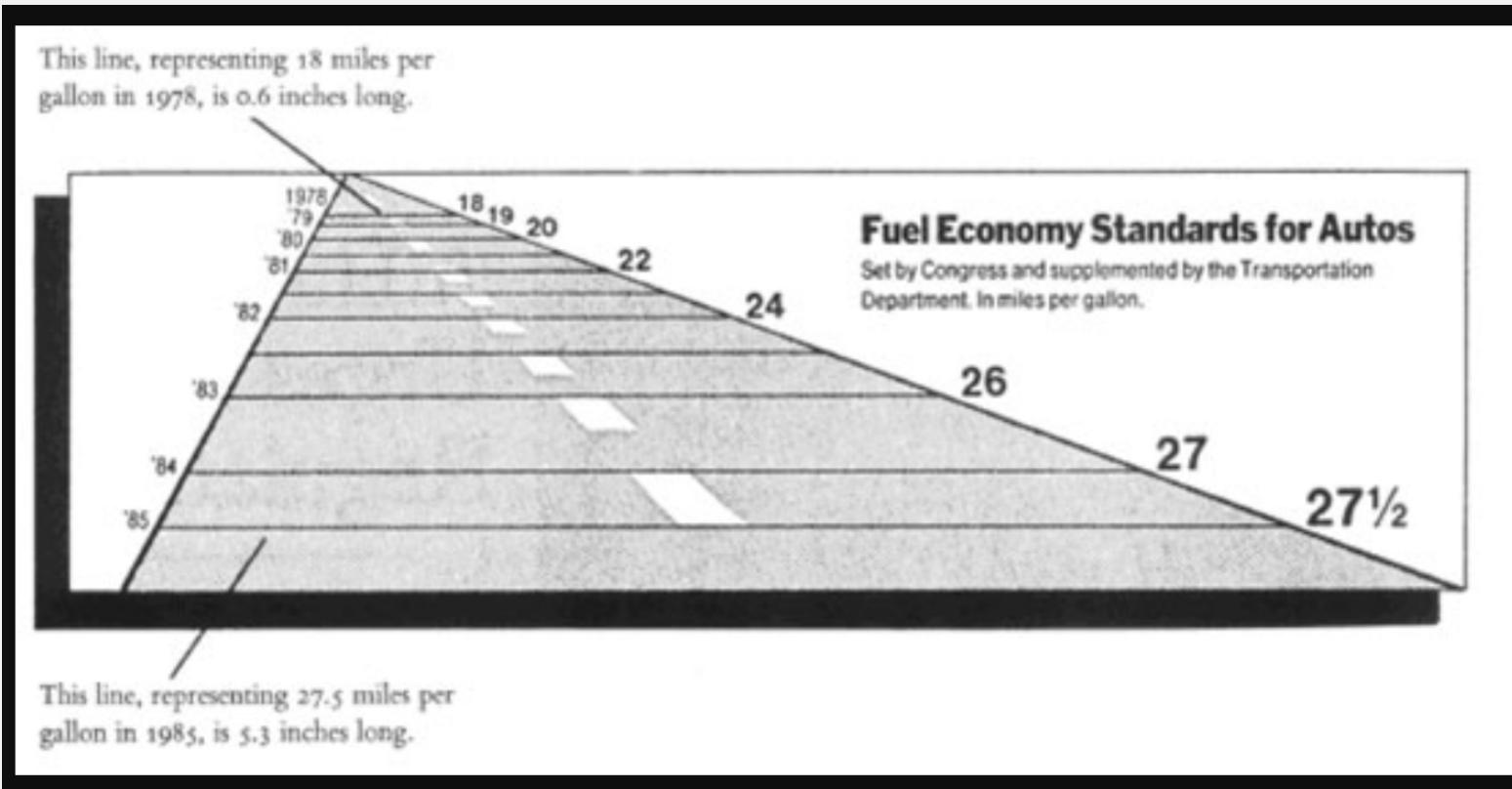
3. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units (**USE "PROPER" – adjusted/normalized - DATA**)





-
4. Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity.
Write out explanations of the data on the graph itself. Label important events in the data. (**THE PLOT SHOULD BE AS EXPLICATIVE AS POSSIBLE**)
 5. Graphics must not quote data out of context. (**THE PLOT SHOULD BE AS AUTO-CONTAINED EXPLICATIVE AS POSSIBLE**)
 6. The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data (**DON'T USE MISLEADING 3D-4D....**)





Example of a graph with low graphical integrity.





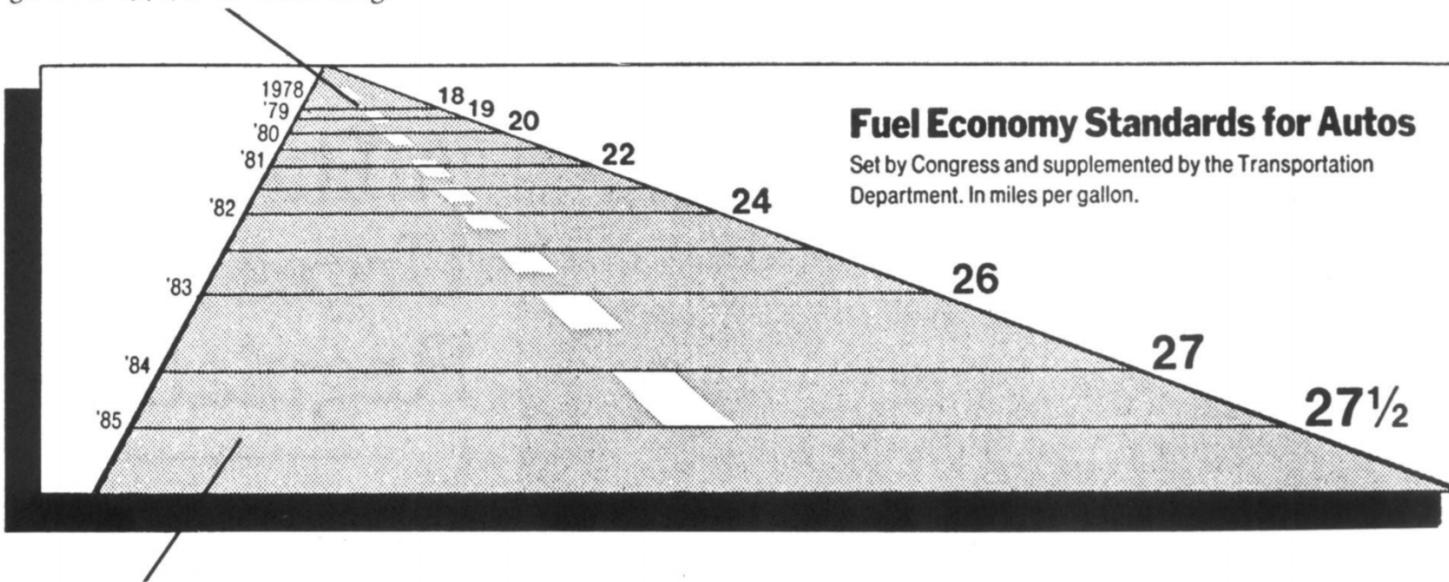
Lie factor (Edward Tufte): amount of distortion on a graph

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

If it is greater than one the graph is lying



This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



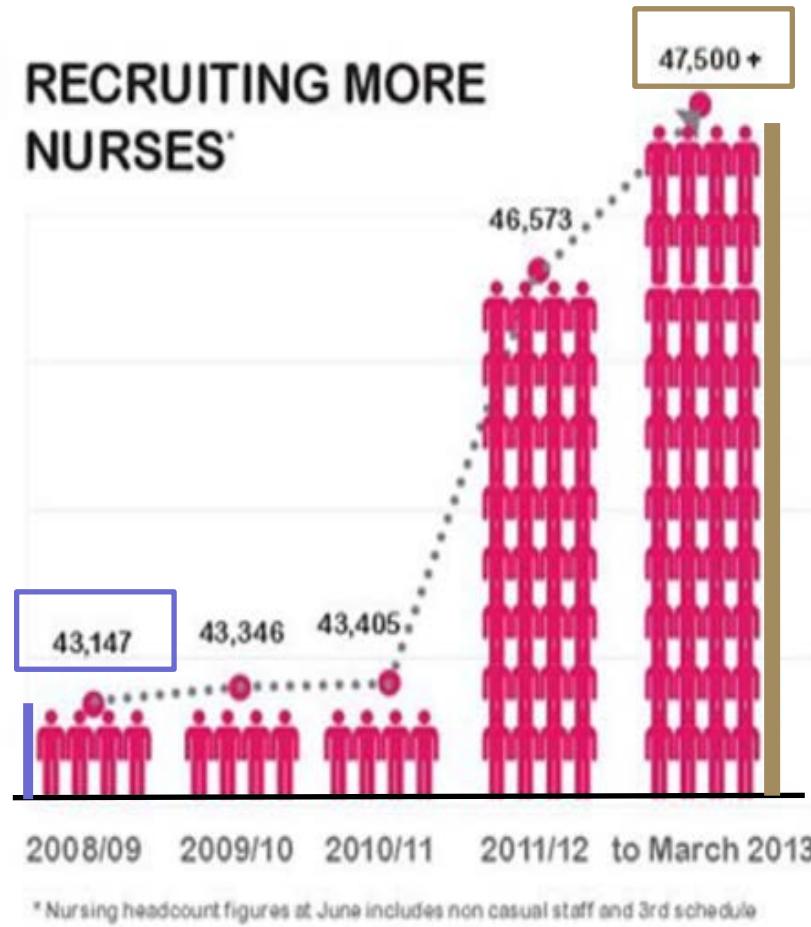
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\text{Data Effect} = \frac{27.5 - 18}{18} = 0.53, \quad \text{Graph Effect} = \frac{5.3 - .6}{.6} = 7.83,$$

$$\text{Lie Factor} = 14.8$$



► RECRUITING MORE NURSES



Ratio you see =

$$\frac{3.5'' - 0.47''}{0.47''} = 6.45 \quad \text{graph}$$

$$\frac{47,500 - 43,147}{43,147} = 0.1 \quad \text{real}$$

Lie Factor = $6.45/0.1 = 64.5$



And someone else has concentrated on the image scale

A plot must be taller or wider??

Cleveland et al. suggest that the mean orientation of all segments should be 45°

[Heer at al.](#) find a scale that makes the mean orientation of the “trend” data 45°

The trend curve is found by applying spectral decomposition to remove high-frequencies in the data.

The trend curve is obtained by smoothing the lower bend (lower frequencies).

This is a bit too complex... Experts suggests using the same scale if the axis are semantically similar.

Otherwise, just make your choice in an objective ways





IT'S THE SAME FOR TABLES!!!





As Tufte states, simple methods help

Video: [Al Gore's CO2 Emissions Chart](#)



An interesting TED'S TALK

https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

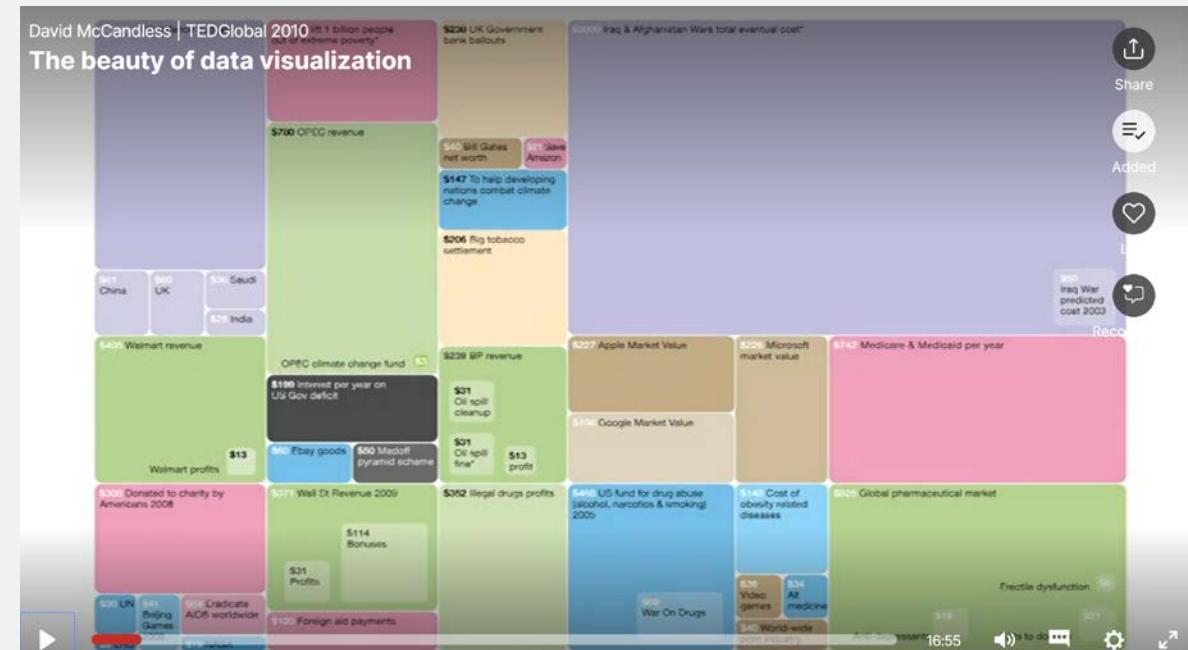
David McCundless uses what he uses the name **information Maps** to define graphs that “map” the information he has scraped from various newspapers into a visible/understandable representation



Treemaps

Treemap representation is used to show money spent/donated/received (as it was mentioned in various newspapers).

- Colors helps define the spending, donating, receiving, ...
 - Text in each rectangle mentions WHO spent/received/donated
 - area shows how much is the money w.r.t. to the others



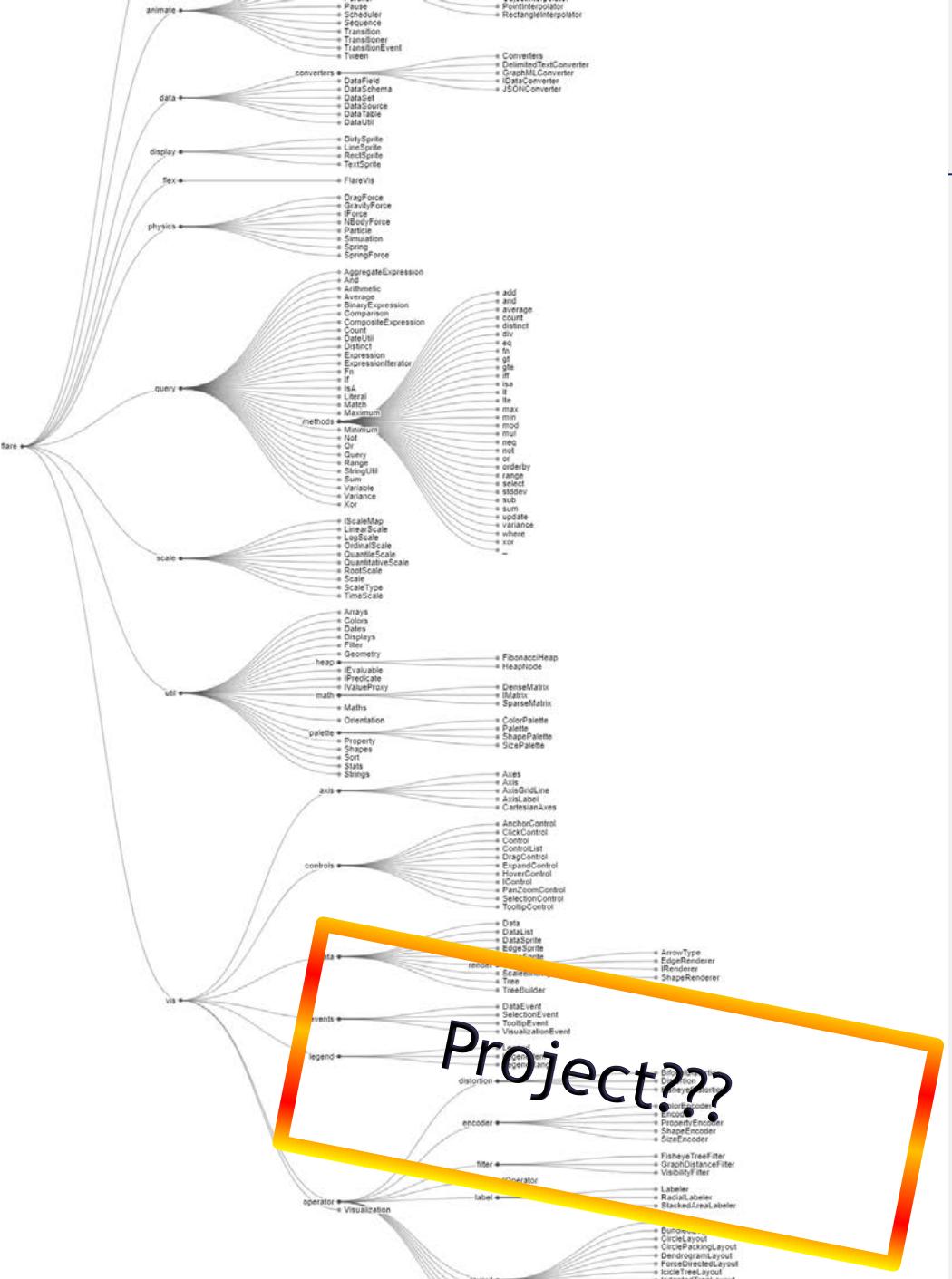


Treemaps: originally developed by Ben Shneiderman for viewing the hard disk content

There are several treemaps version:

- Unordered treemaps
- Nested Treemaps
- Slice-and-dice treemaps
- Hierarchical treemaps





Treemaps are a compact representation of
Trees layout (This image is from the tidy trees
implementation)

"Tidier Drawings of Trees"

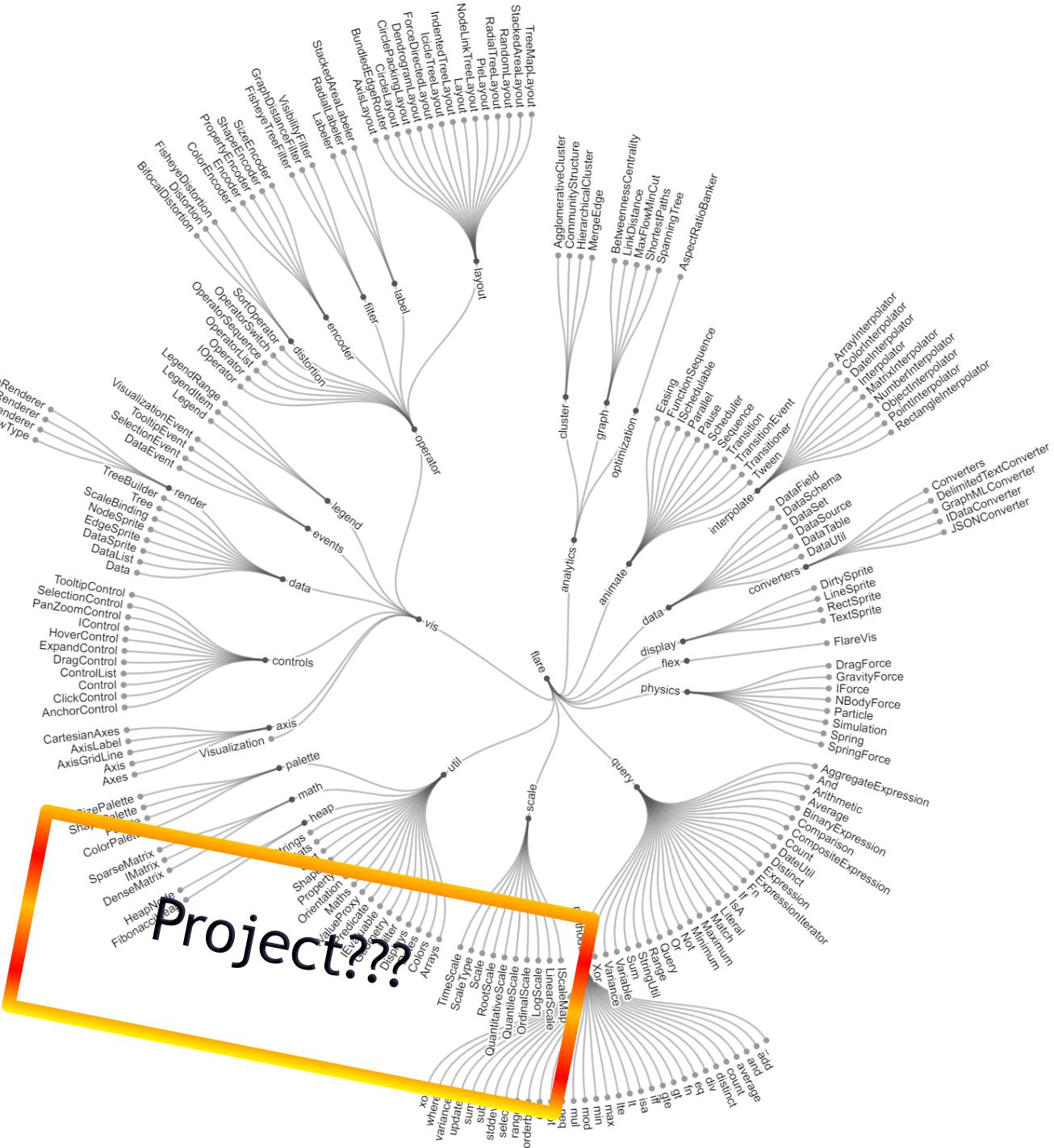
Reingold and Tilford

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING,
VOL. SE-7, NO. 2, MARCH 1981

D3.js implementation

<https://observablehq.com/@d3/tidy-tree>



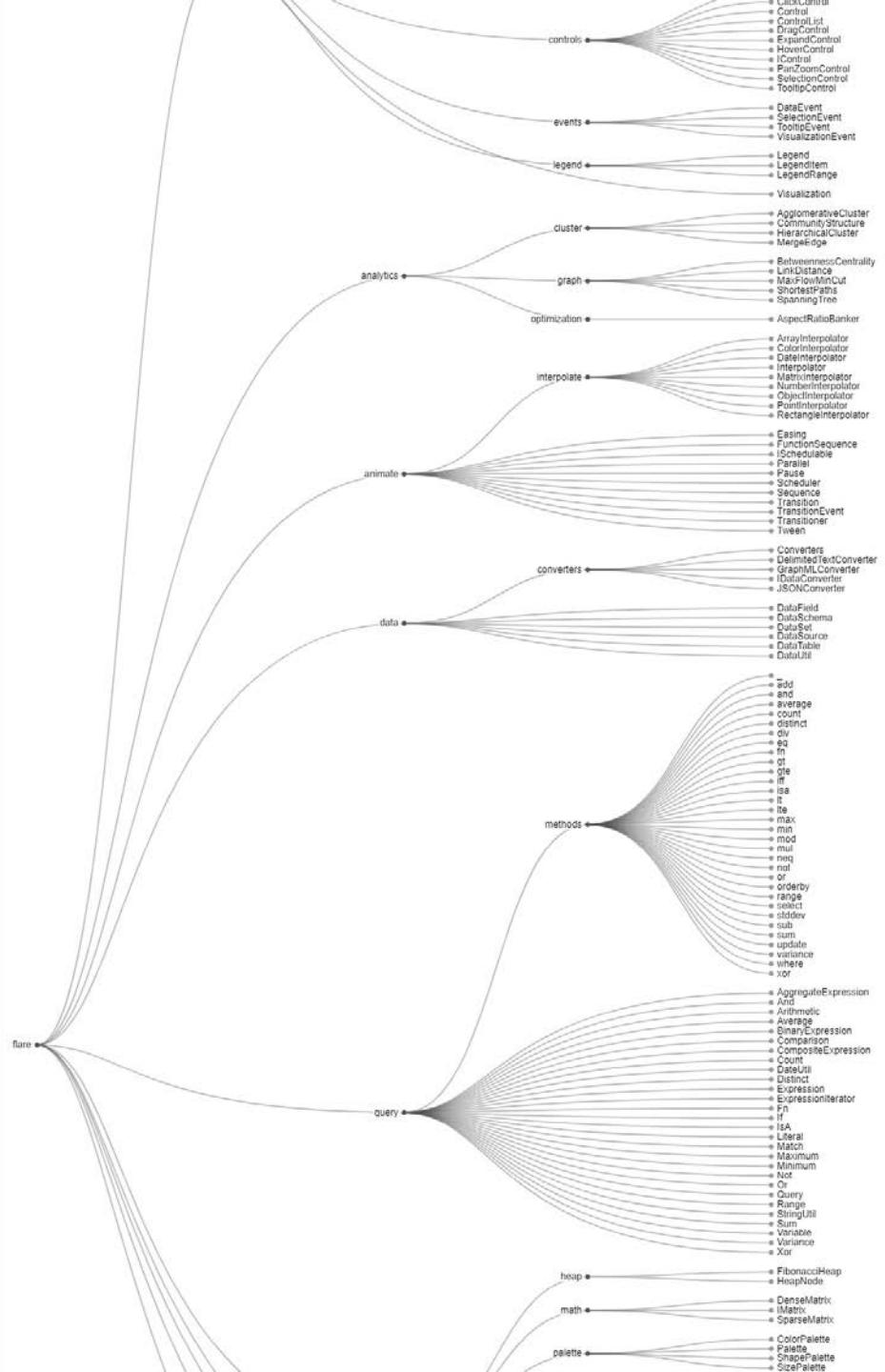


Radial variant of tidy trees:

- it's more compact
- there is not top and bottom: we don't have an unconscious sorting by importance

BUT

- text is rotated and don't horizontal.
- more difficult to see it



Tidy trees are better visualizations than
cluster dendograms which put leafs all at the
same level



D3.js

Here is a d3.js implementation of Treemaps

<https://observablehq.com/@d3/treemap>

Short course at:

<https://observablehq.com/@d3/learn-d3>

Project???





Uses a Frequency trail to view world fears:

David McCandless | TEDGlobal 2010

The beauty of data visualization

MOUNTAINS OUT OF MOTELNITS
A timeline of global media panic



Share



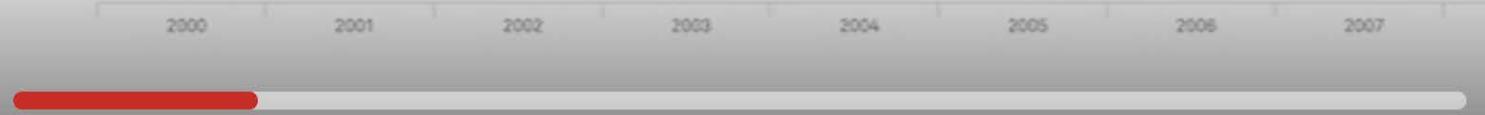
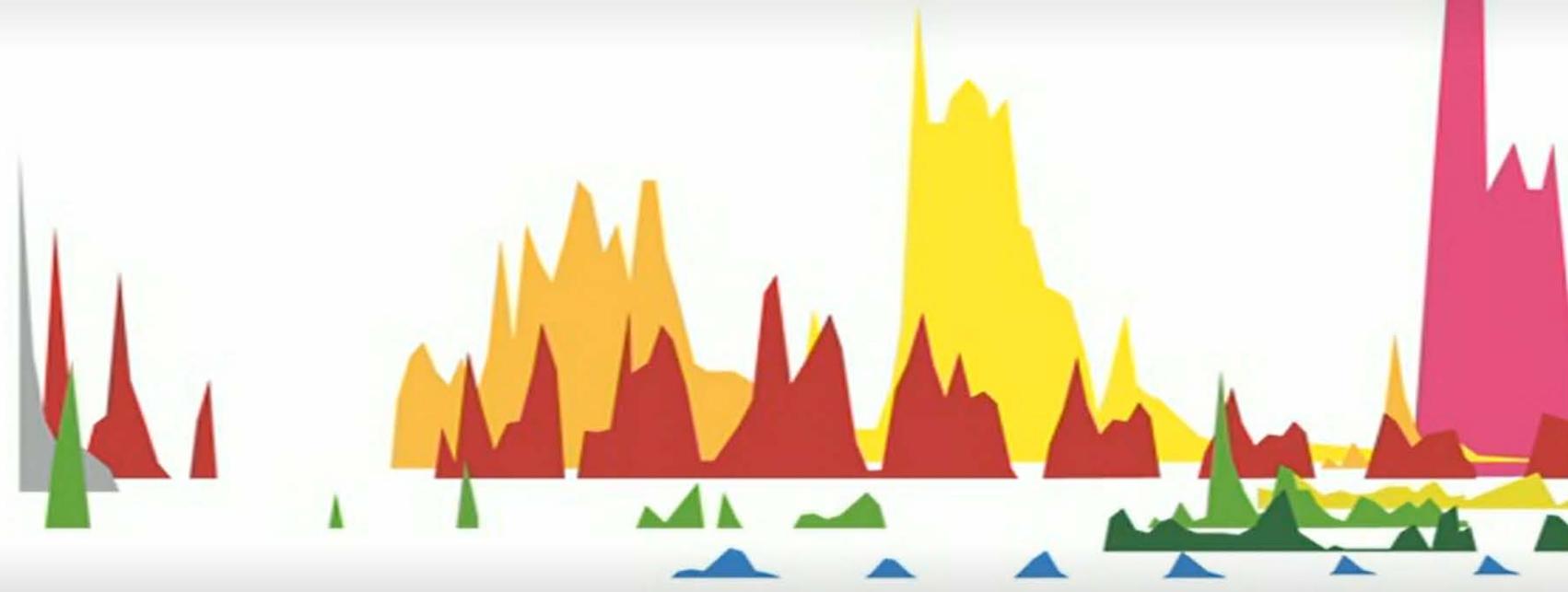
Added



Heart



Comment



15:00



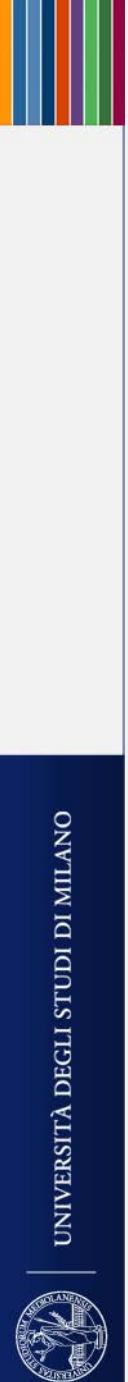


Frequency trails (also known as joy or ridgeline plots) are a method for comparing distributions by vertically offsetting each curve (a 3D-ish way of aligning charts – sometimes 3D helps!).

They are an alternative to:

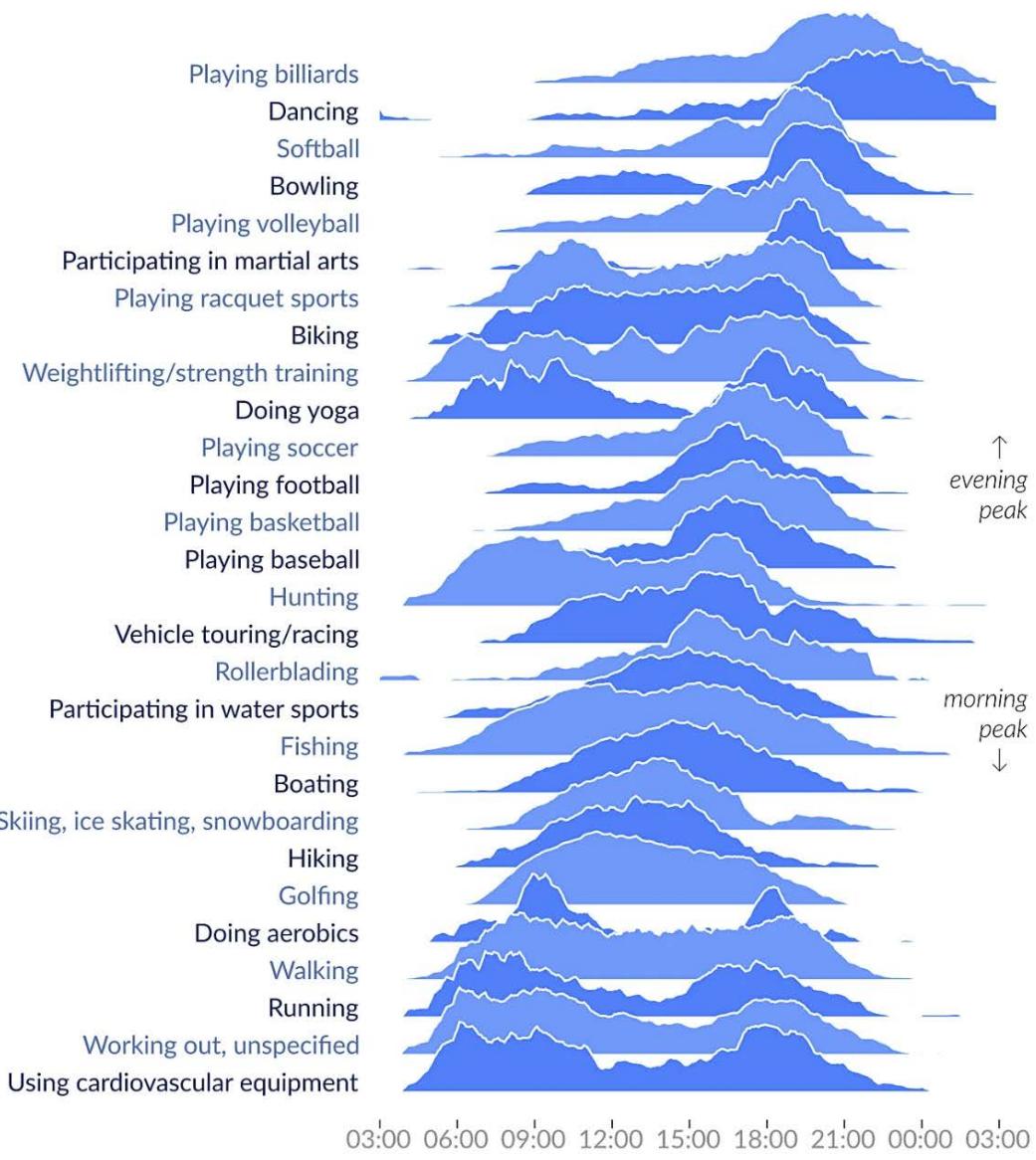
- multi-line charts (where overcrowding increases with the number of series)
- small multiples (which are harder to compare when there are too many lines - preattentive memory limits)





Peak time of day for sports and leisure

Number of participants throughout the day compared to peak popularity.
Note the morning-and-evening everyday workouts, the midday hobbies,
and the evenings/late nights out.



Have been around for some time but they hit peak popularity in 2017 when Henrik Lindberg posted a chart that showed “sports participation by time of the day”

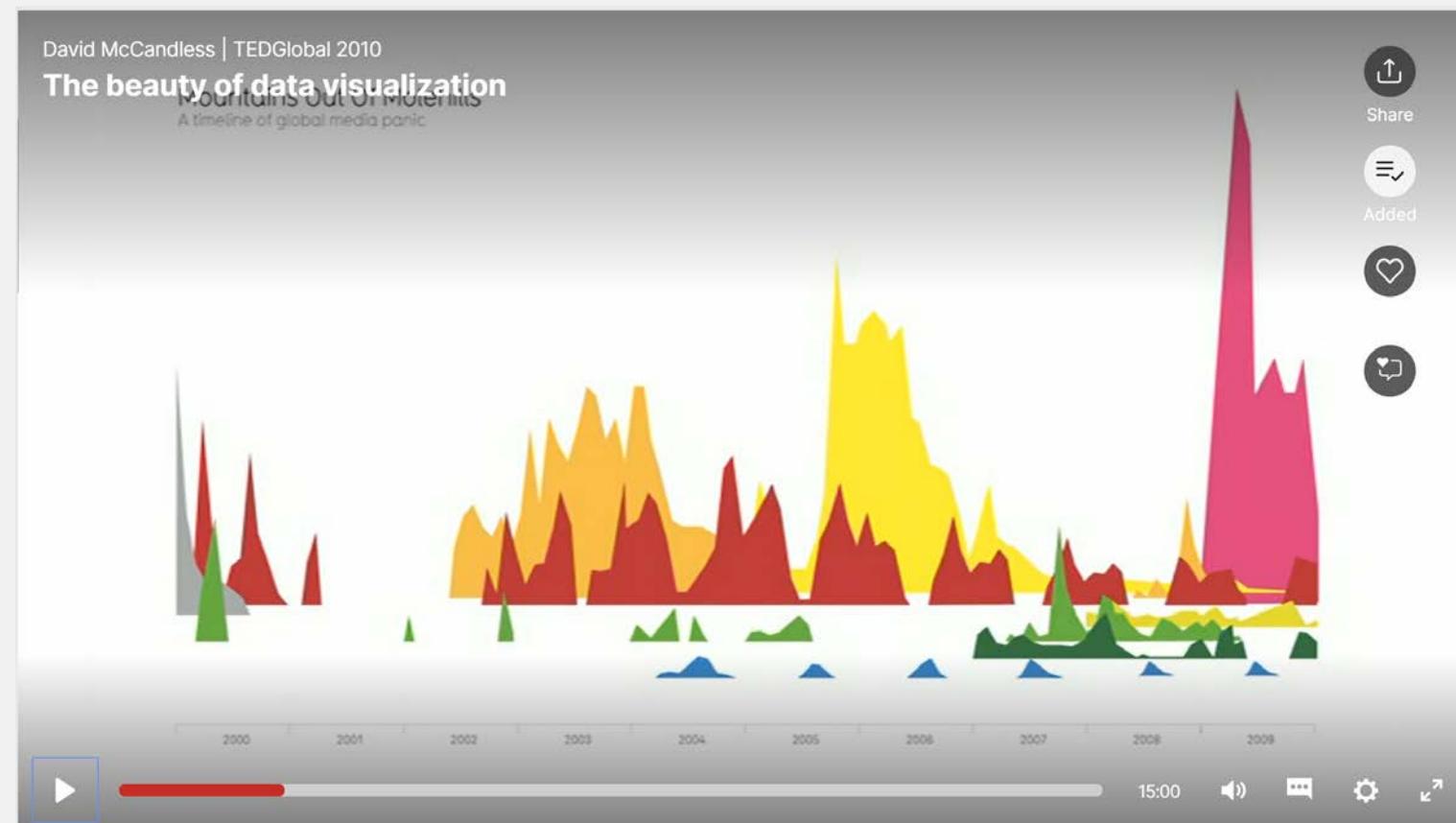
Here colors help distinguishing neighboring lines





Colors help differentiating “important” and (believed) “less important” fears:

- fears due to similar reasons are represented with similar colors
- more important fears are highlighted with hues in red channel
- fears which are less important (not motivated) are marked with colors considered as “positive”, e.g. green (grey)



The beauty of data visualization

MOUNTAINS OUT OF MOTEL HITS
A timeline of global media panic



Share



Added



Millennium bag



SARS

Bird flue

Swine flue

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

15:00



The beauty of data visualization

MOUNTAINS OUT OF MOTEL HITS
A timeline of global media panic



Share



Added



Millennium bug



SARS

Bird flue

Swine flue

Killer wasps

Asteroid collisions

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

15:00



The beauty of data visualization

Mountains Out of Molehills
A timeline of global media panic



Share



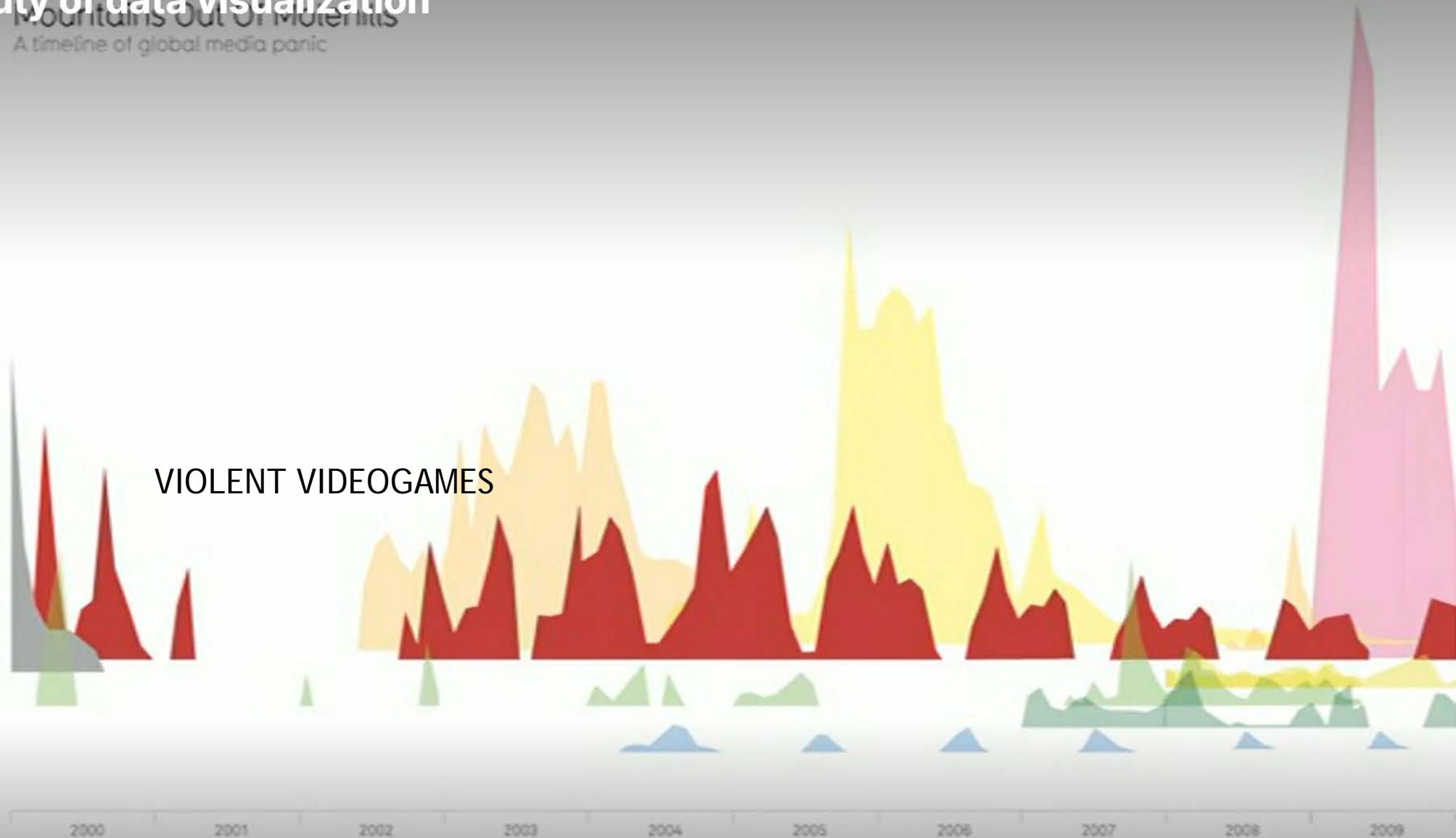
Added



Heart



Comment



14:03



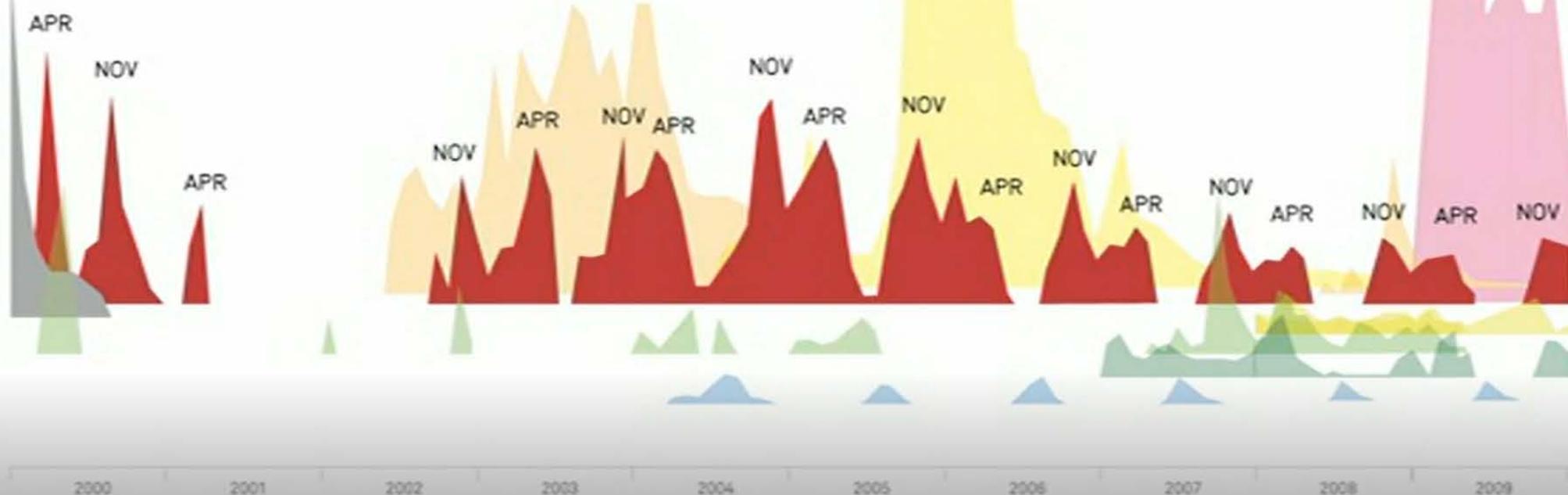
You see a regular pattern: twin peak in November (before Christmas time) and April
(memories of Columbine High School massacre on 20th April 1999)



Share



Added



2000 2001

2002 2003

2004

2005

2006

2007

2008

2009

13:54



Why not using histograms (generally represented by bar charts) instead of lines?

Lines instead of histograms help emphasizing peaks (intensity in media news)

Why FILLED (Area) plots? Filled lines (form mountains) create shapes with areas.

- **Shape/Area an added visual clue** -> enforce and emphasize perception of relative quantities, therefore inherently helping the comparison of their importance.



Direct comparison of trends could also be done with streamgraphs

"A Stream graph is a type of stacked area chart. It displays the evolution of a numeric value (Y axis) following another numeric value (X axis). This evolution is represented for several groups, all with a distinct color.

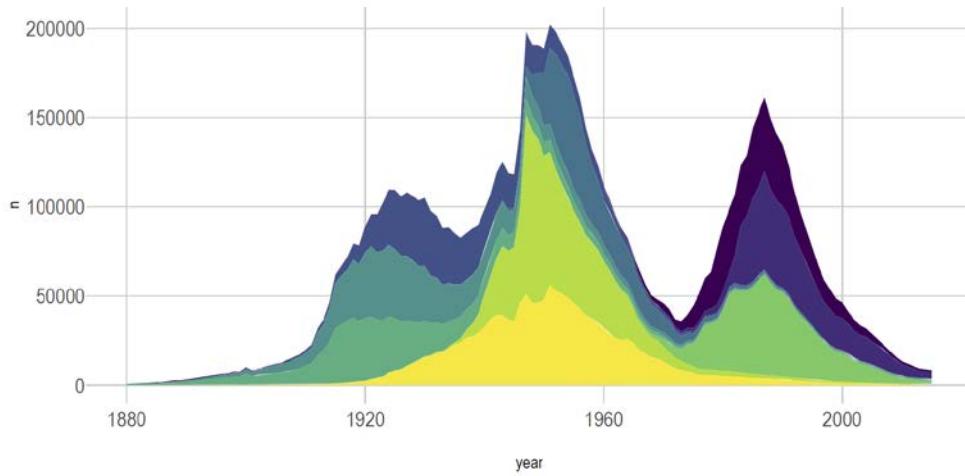
Contrary to a stacked area, there is no corner: edges are rounded what gives this nice impression of flow. Moreover, areas are usually displaced around a central axis, resulting in a flowing and organic shape. "

<https://www.data-to-viz.com/graph/streamgraph.html>





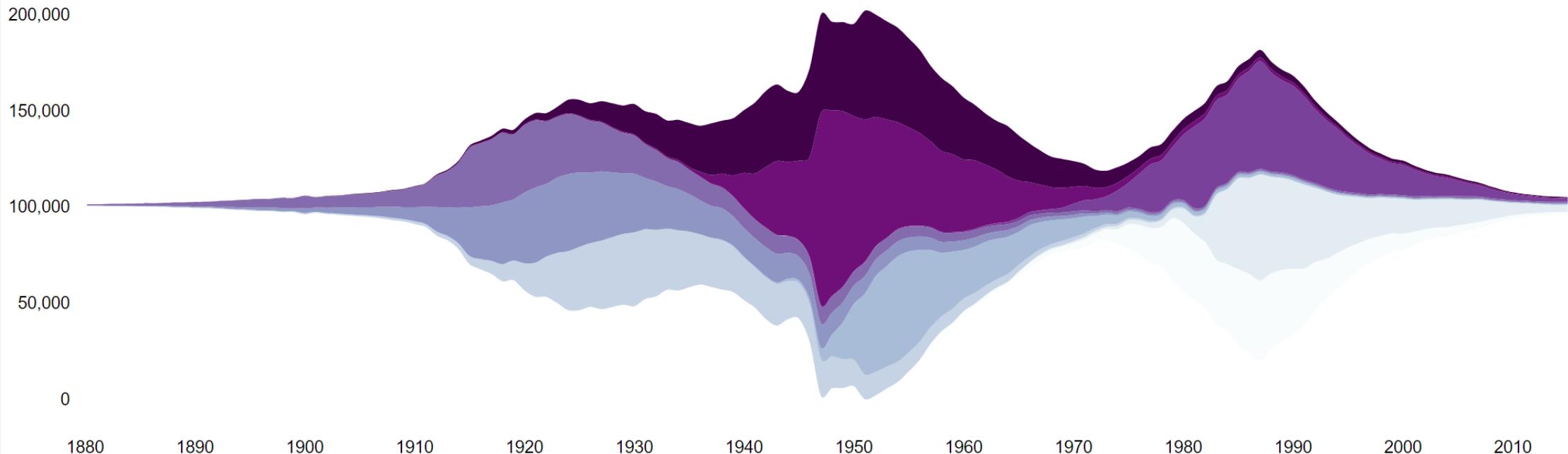
Popularity of American names in the previous 30 years



From stacked area charts

To Stream graphs

CODE



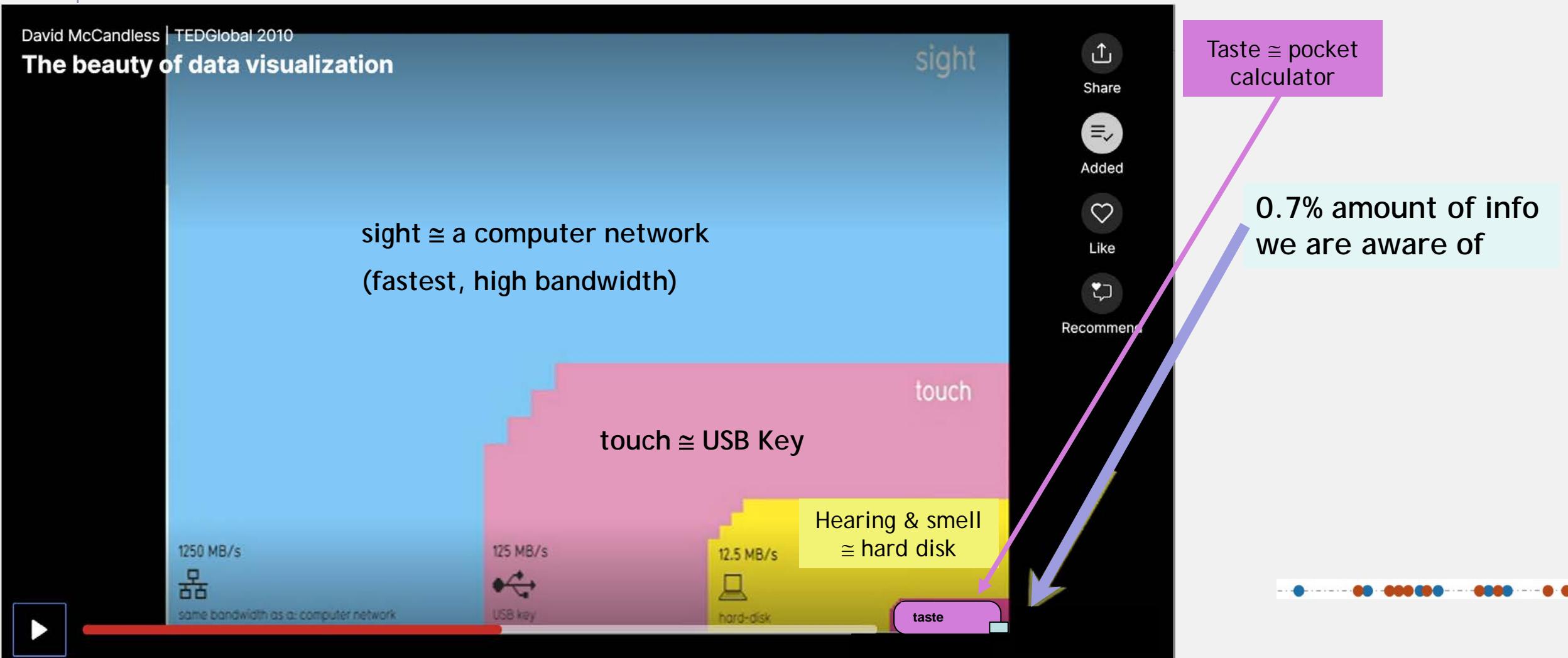
Stream-graphs are good to study the relative proportions of the whole

They are bad to study the evolution of each individual group: it is very hard to subtract the height of other groups at each time point. In this case use small multiples!!!



bandwidth of the senses converted into computer terms: amount of info we get from ...

Tor Norretranders





Snake Oil: scientific evidence for nutritional supplements

<https://www.informationisbeautiful.net/visualizations/snake-oil-scientific-evidence-for-nutritional-supplements-vizsweet/>





FUNCTION OUTLINES THE SHAPE

Here is an example

Which graph may I use to compare population distribution by age?

If I had data such as:

Country (C), age (A), sex (S), number of people in C with age A and sex S...

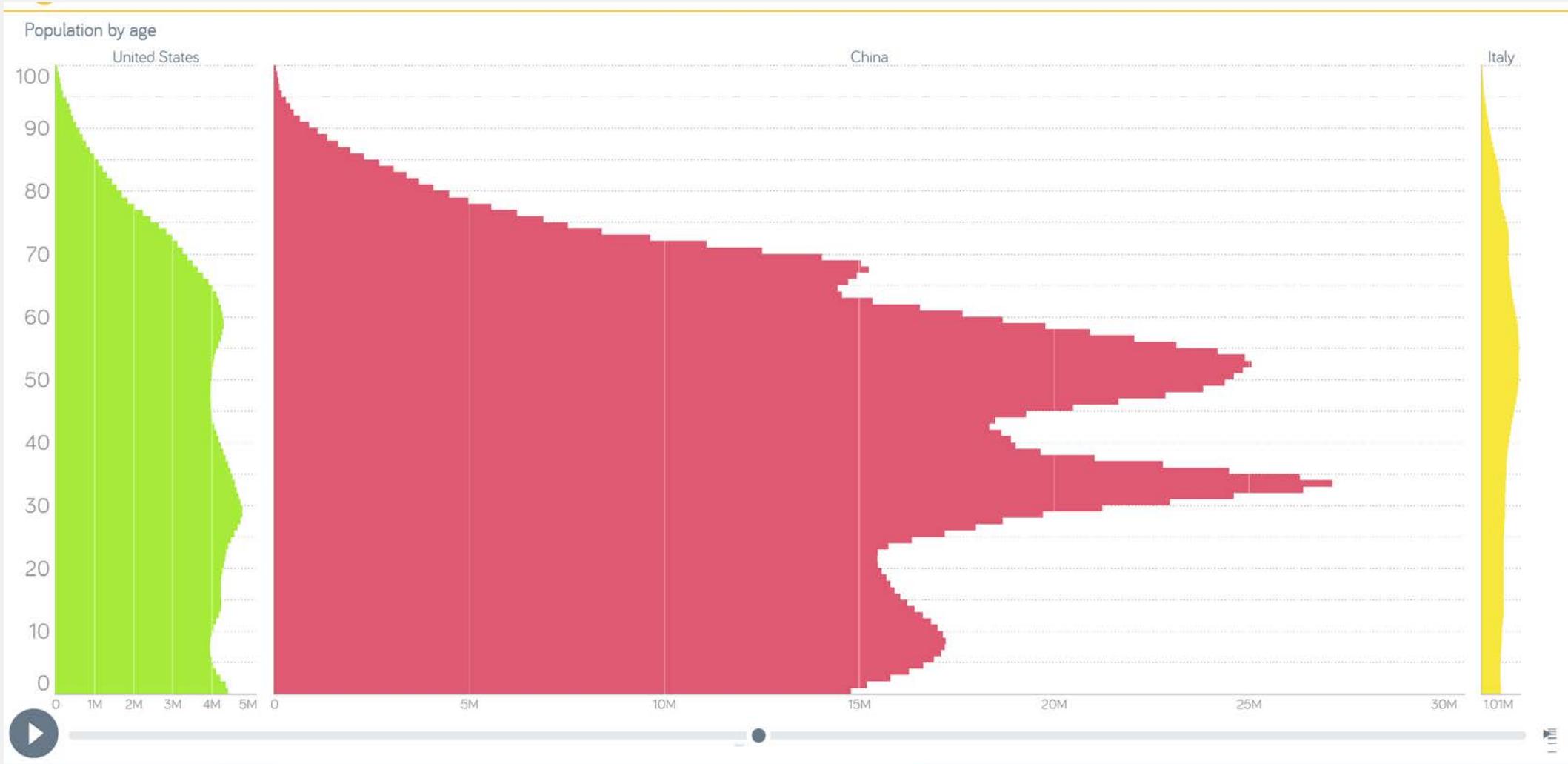
....

...





DataBars example. Distribution of population per age. [Gapminder Tools](#)



Visualizing the Titanic Disaster

