

# Data Visualization

Daniela Huppenkothen

*NYU Center for Cosmology and Particle Physics*

*NYU Center for Data Science*

 @Tiana\_Athriel

James Davenport

*Western Washington University*

 @jradavenport

# Why do we make plots?

a.k.a. charts, graphs, visualizations,  
figures, diagrams, maps...

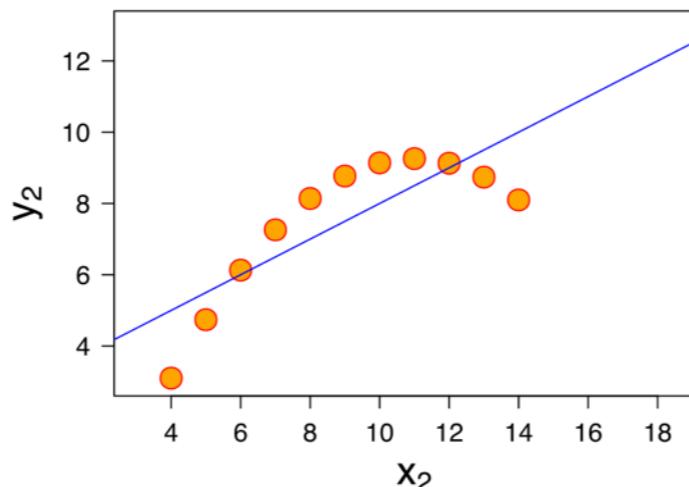
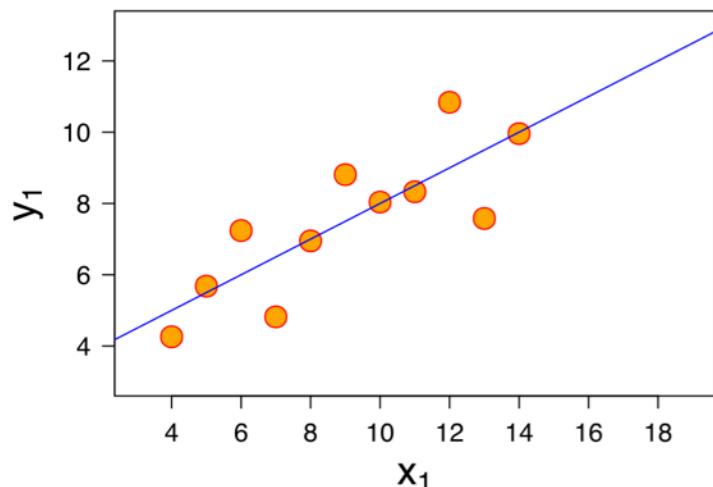
# Why do we make plots?

- **Development**
  - data exploration
  - idea generation
  - debug code
- **Presentation**
  - talks
  - posters
  - papers

Each use/format has unique concerns!

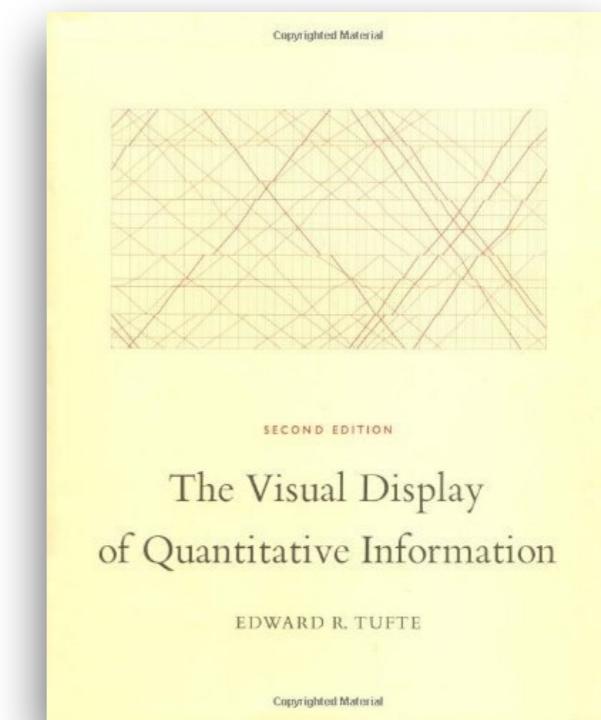
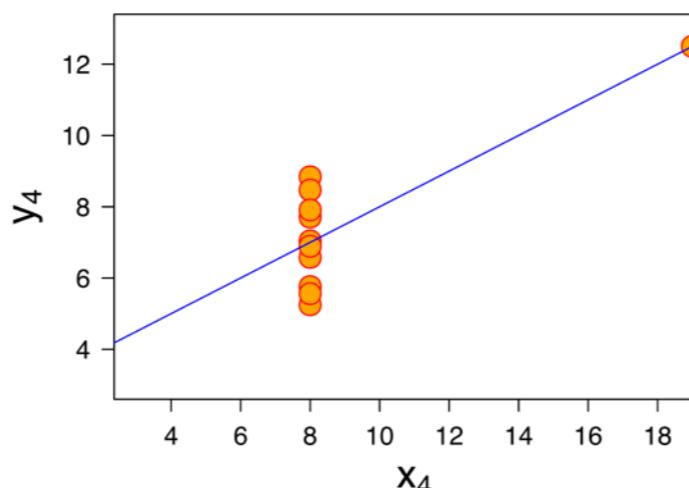
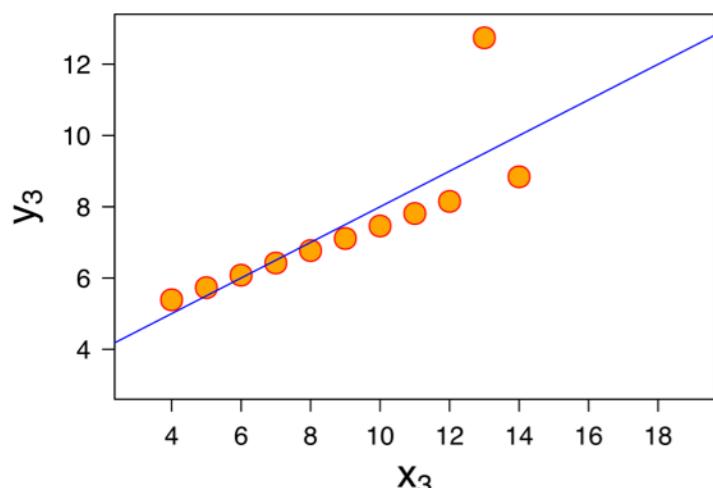
# Development: Exploratory Data Analysis

## Anscombe's Quartet



**All panels have the same**

- mean (in X and Y)
- variance (in X and Y)
- linear regression
- “r” coefficient



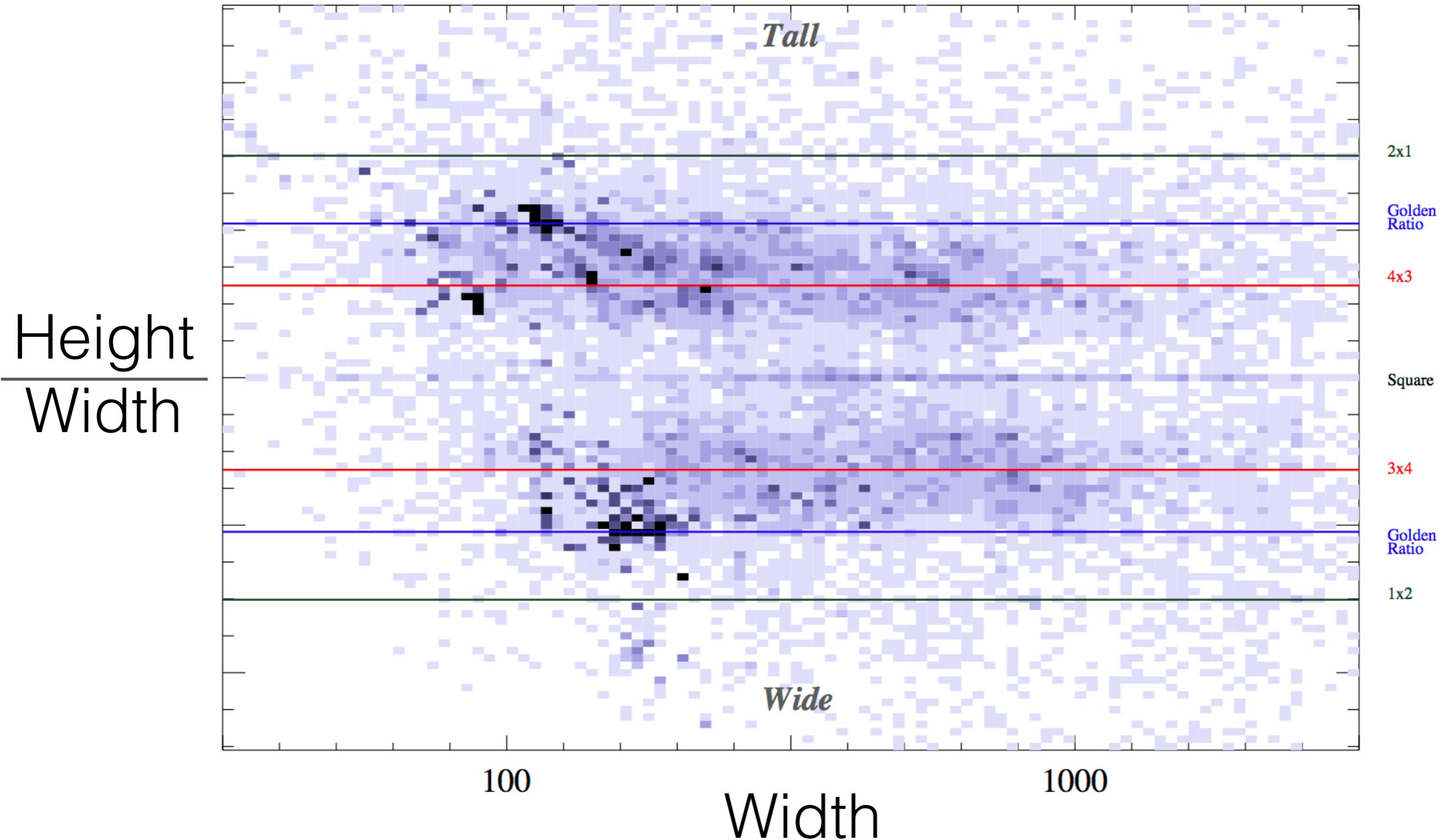
made popular by Tufte

# One plot does not work for all audiences/needs



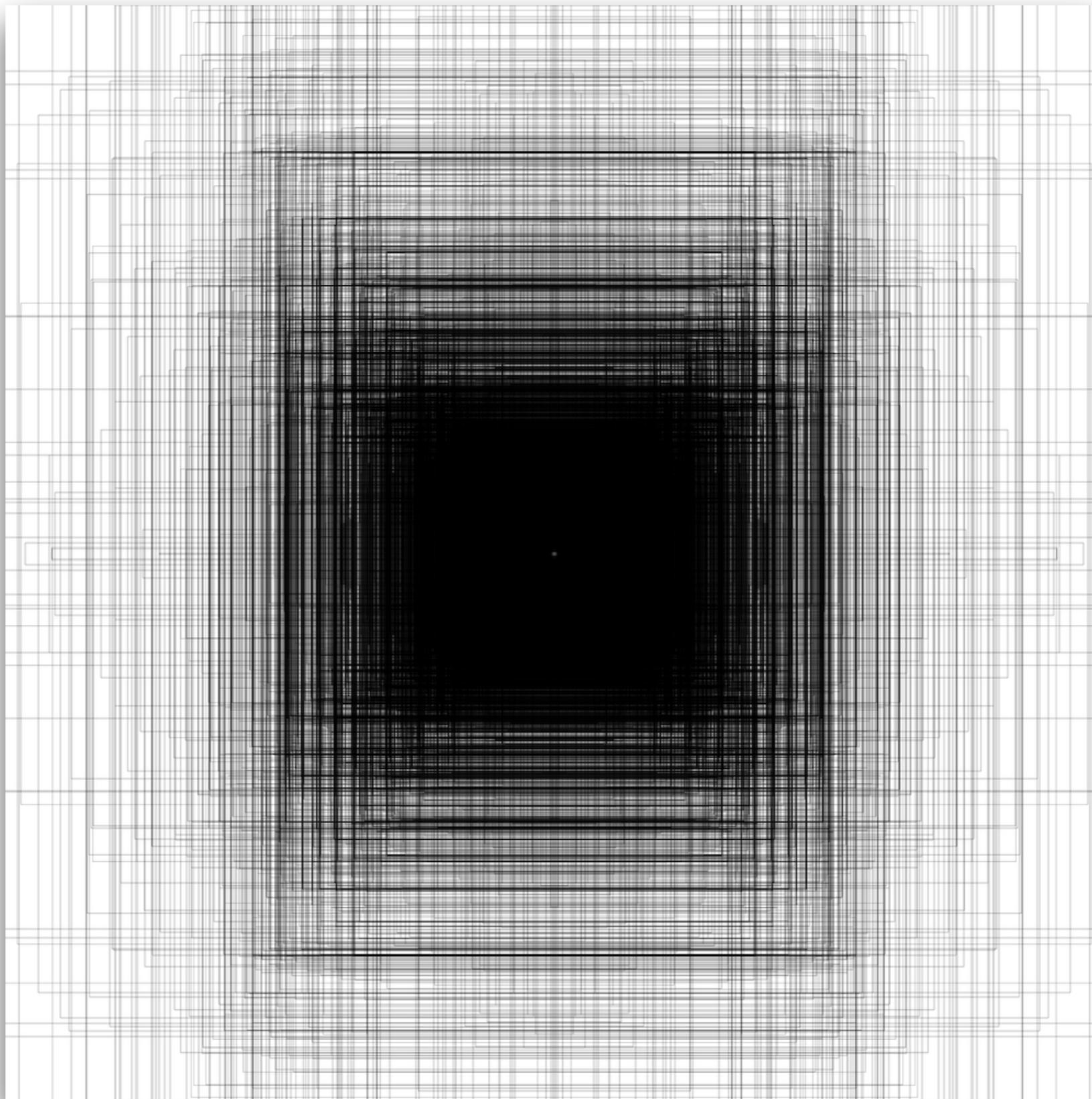
# The Dimensions of Art

65,000 pieces of art from the Tate Modern

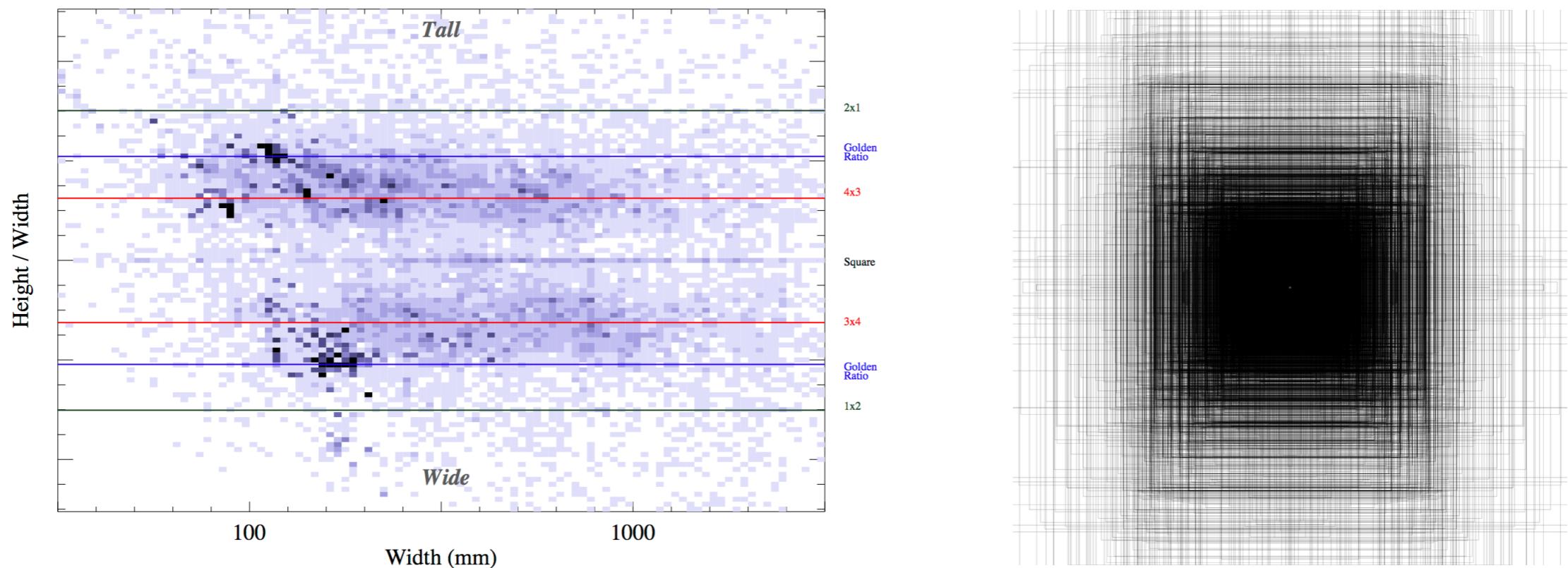


# The Dimensions of Art

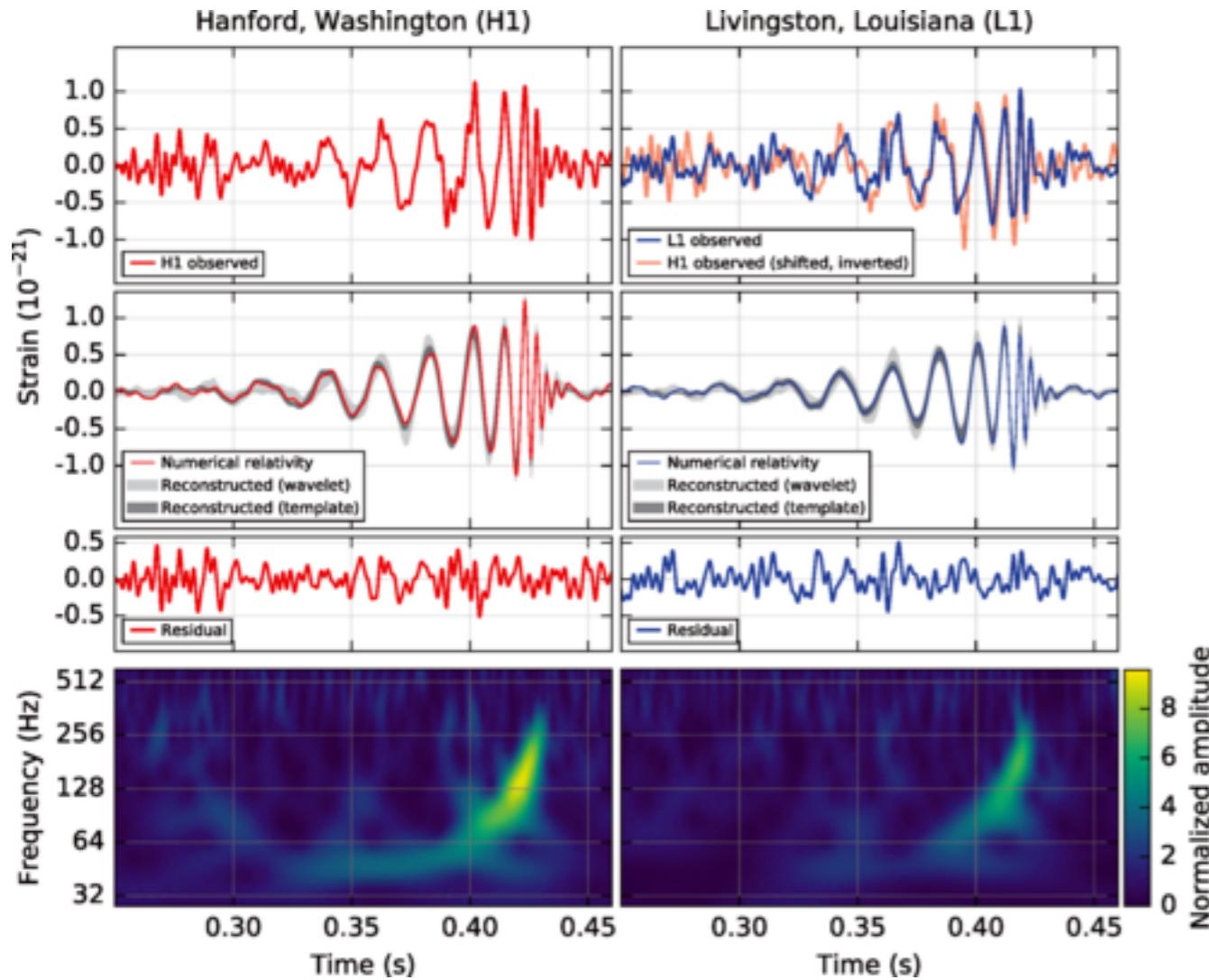
## 65,000 pieces of art from the Tate Modern



# One plot does not work for all audiences/needs



# One plot does not work for all audiences/needs

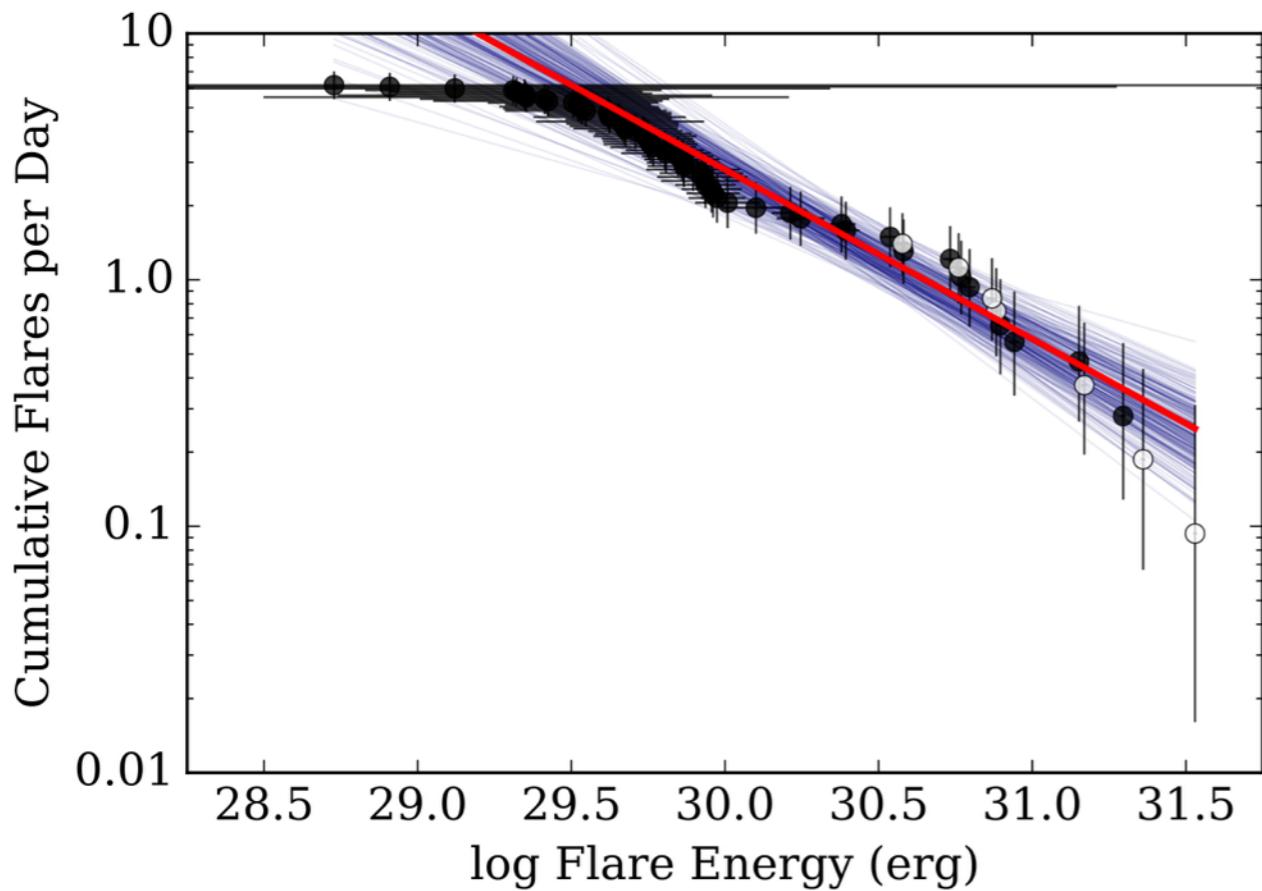


1st LIGO detection, Abbott et al. (2016)

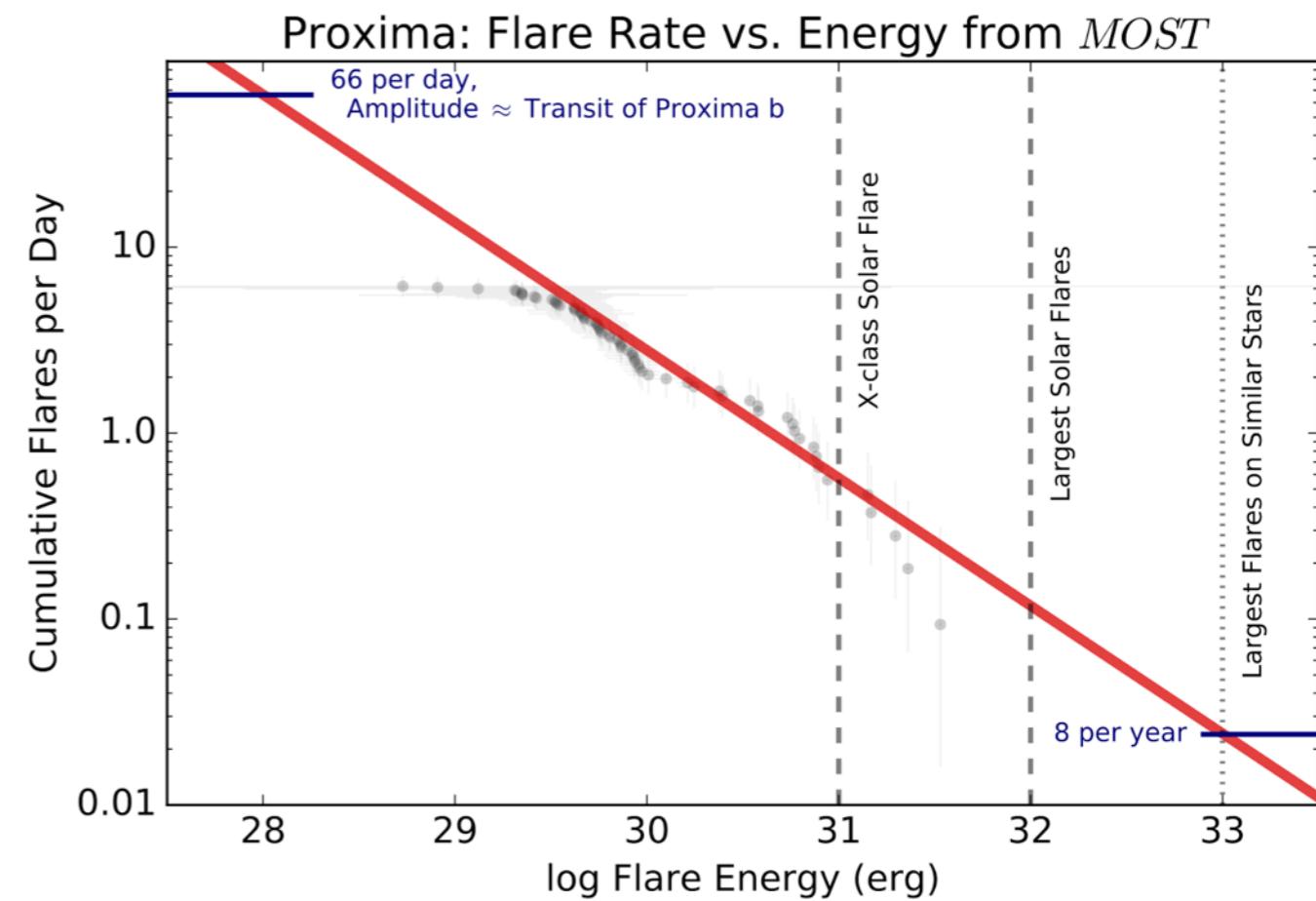


LIGO Magazine, 2016

# One plot does not work for all audiences/needs



Davenport et al. (2016)

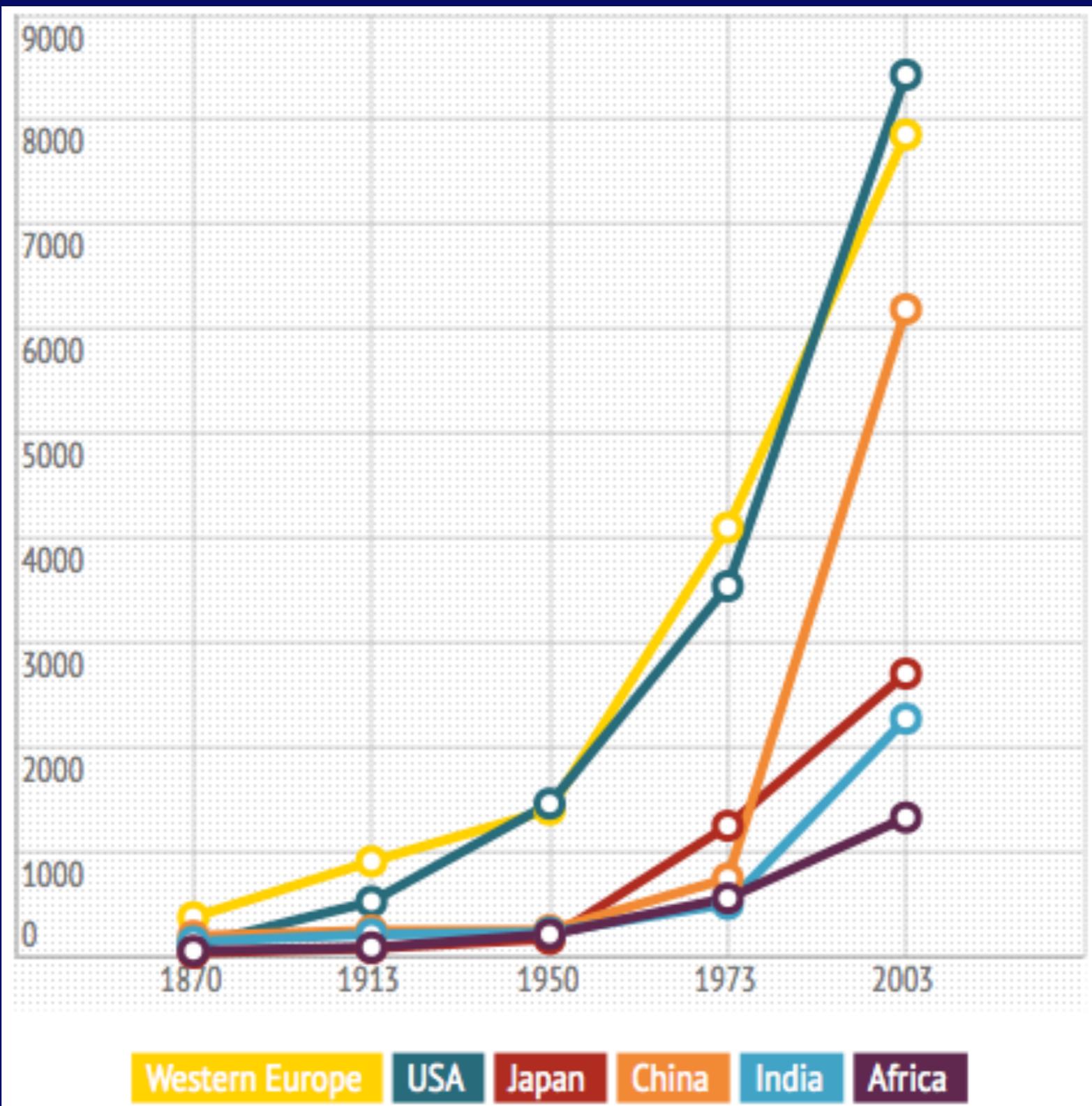


blog post

**Visualization lets you see things  
that would rather go unnoticed**

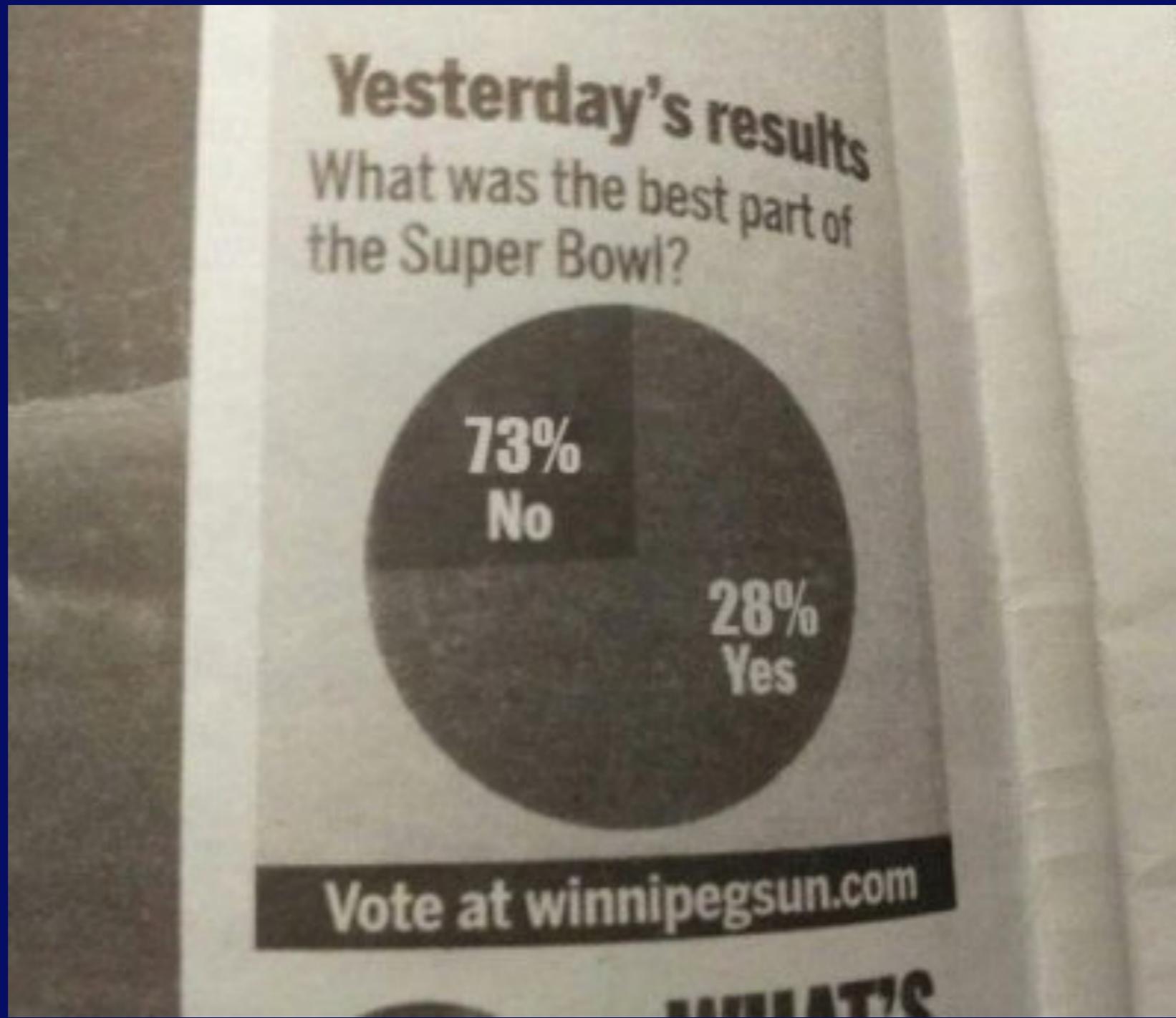
	A	B	C	D	E	F
1	Past GDP	1870	1913	1950	1973	2003
2	Western Europe	367	902	1396	4096	7857
3	USA	98	517	1455	3536	8430
4	Japan	25	71	160	1242	2699
5	China	189	241	244	739	6187
6	India	134	204	222	494	2267
7	Africa	45	79	203	549	1322

<http://www.mulinblog.com/data-visualization-matters/>



**think of visualization as a new set  
of languages you can use to  
communicate**

# How \*not\* to do visualization ...



<http://www.businessinsider.com/the-27-worst-charts-of-all-time-2013-6>

## A brief aside....

Your camera doesn't matter

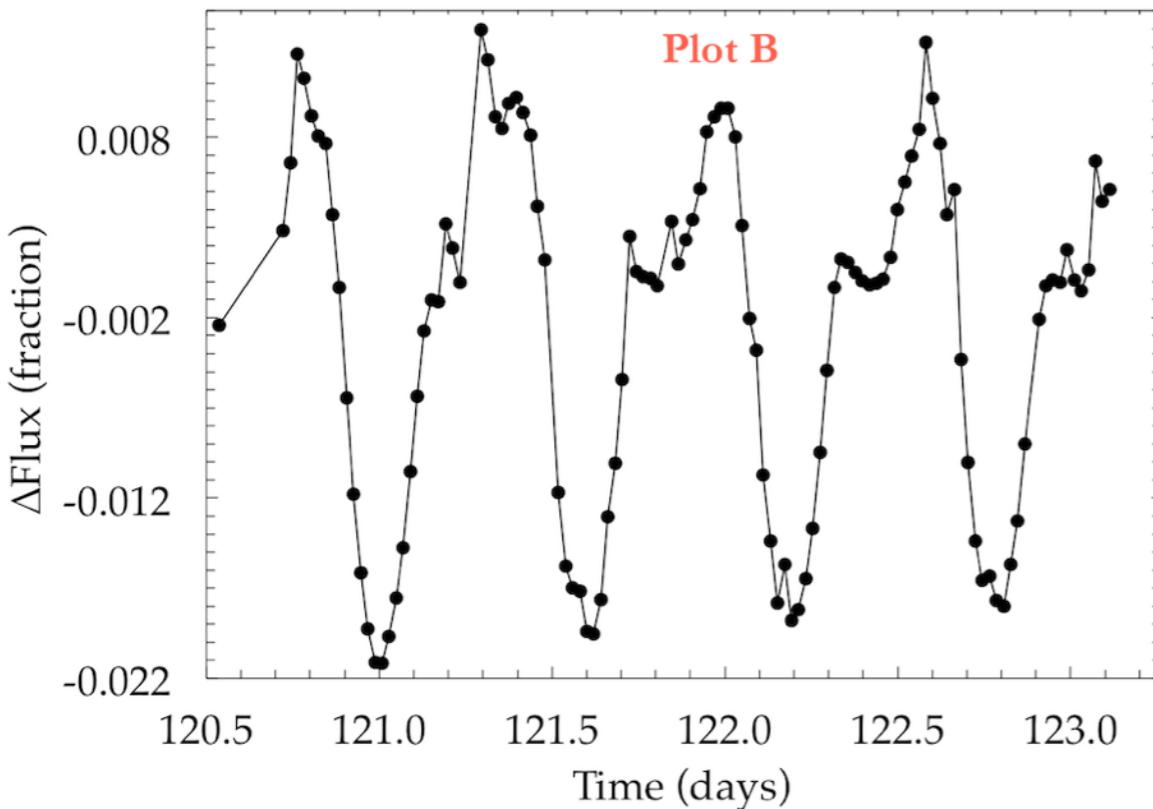
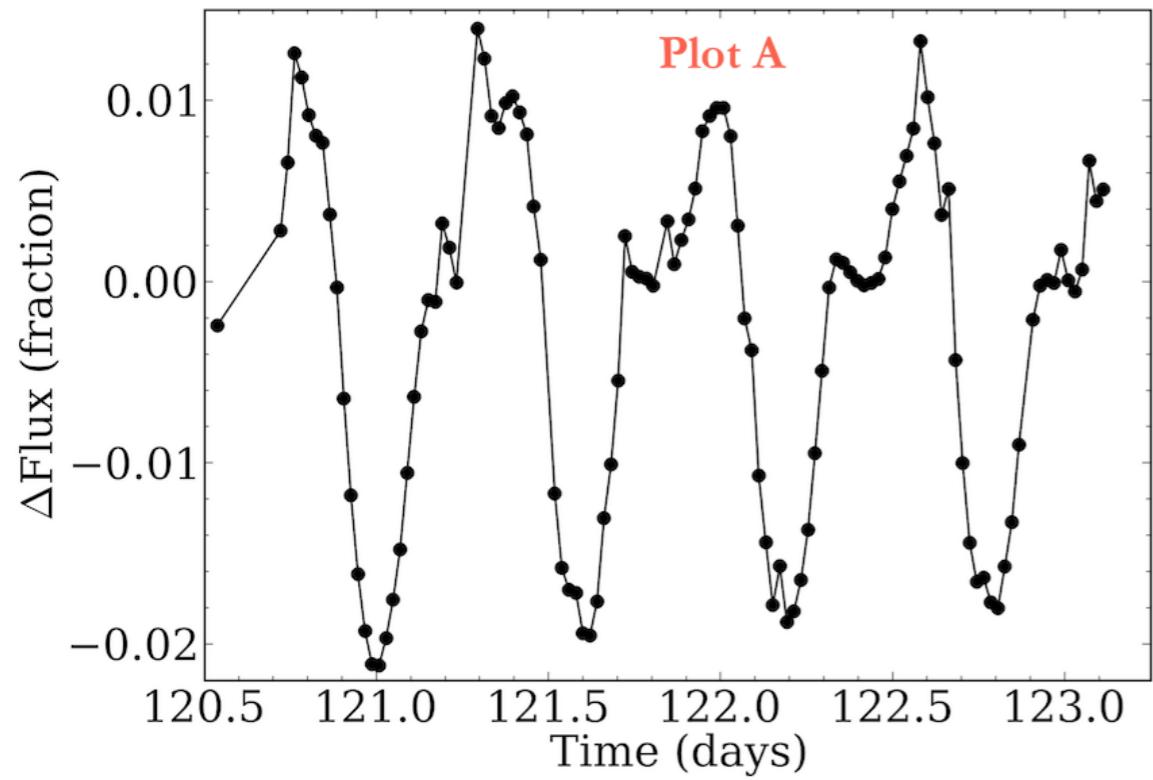
– Ken Rockwell (Photographer)

## **A brief aside....**

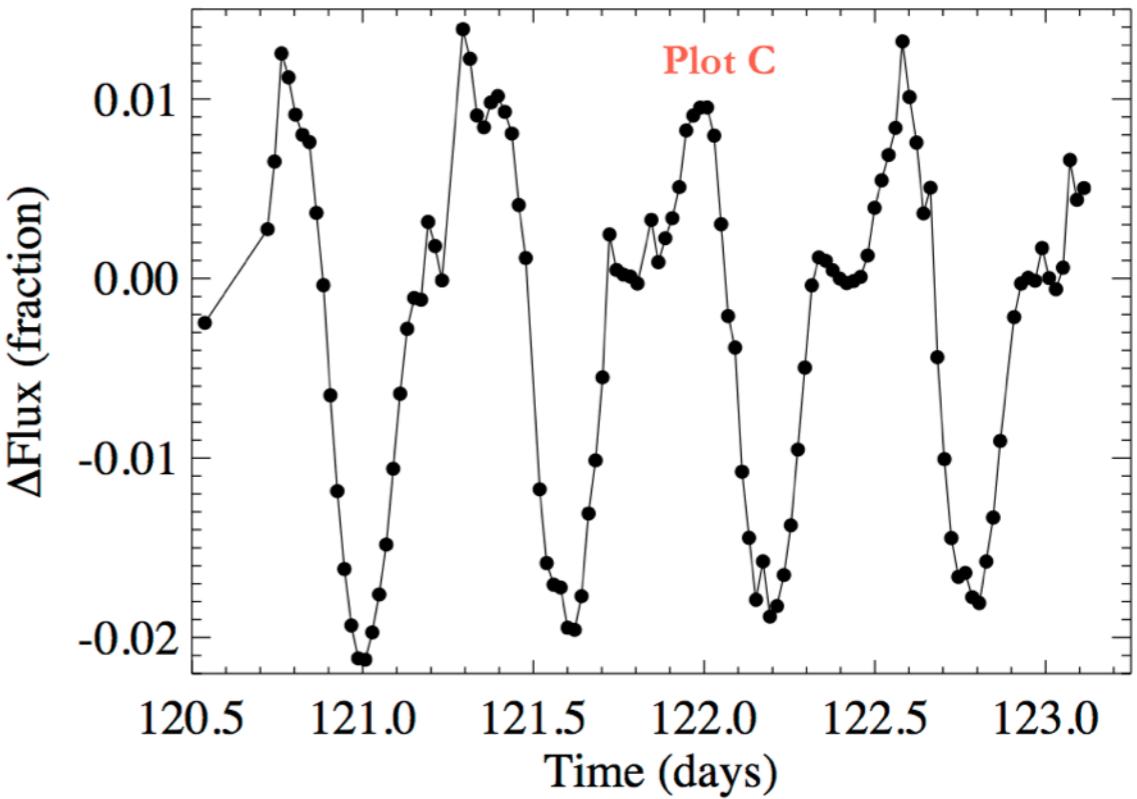
Your graphing language doesn't matter  
– me, just now

A short example:

# A brief aside....



Can you tell which is  
IDL, Python, Excel ?



# Topics:

- perception
- space
- colour + contrast
- visual hierarchy
- types of Data

**A bit of perception ...**

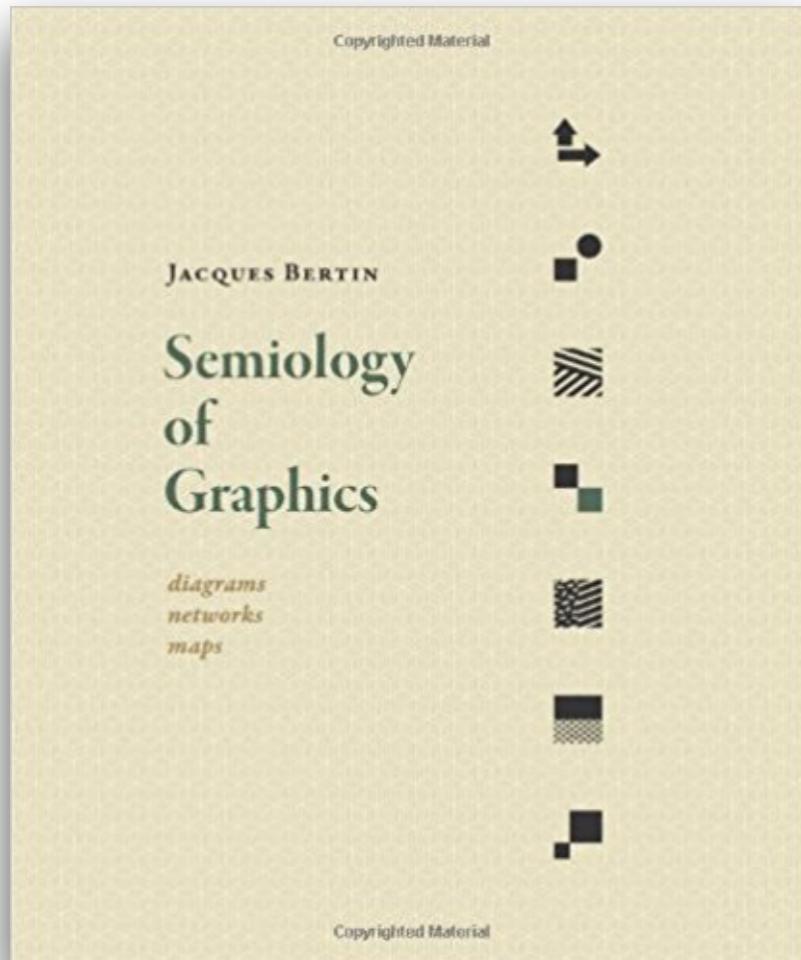
# **Visualization works ...**

# Visualization works ...

“... when data understanding is supported by perceptual rather than cognitive processes”

– Enrico Bertini

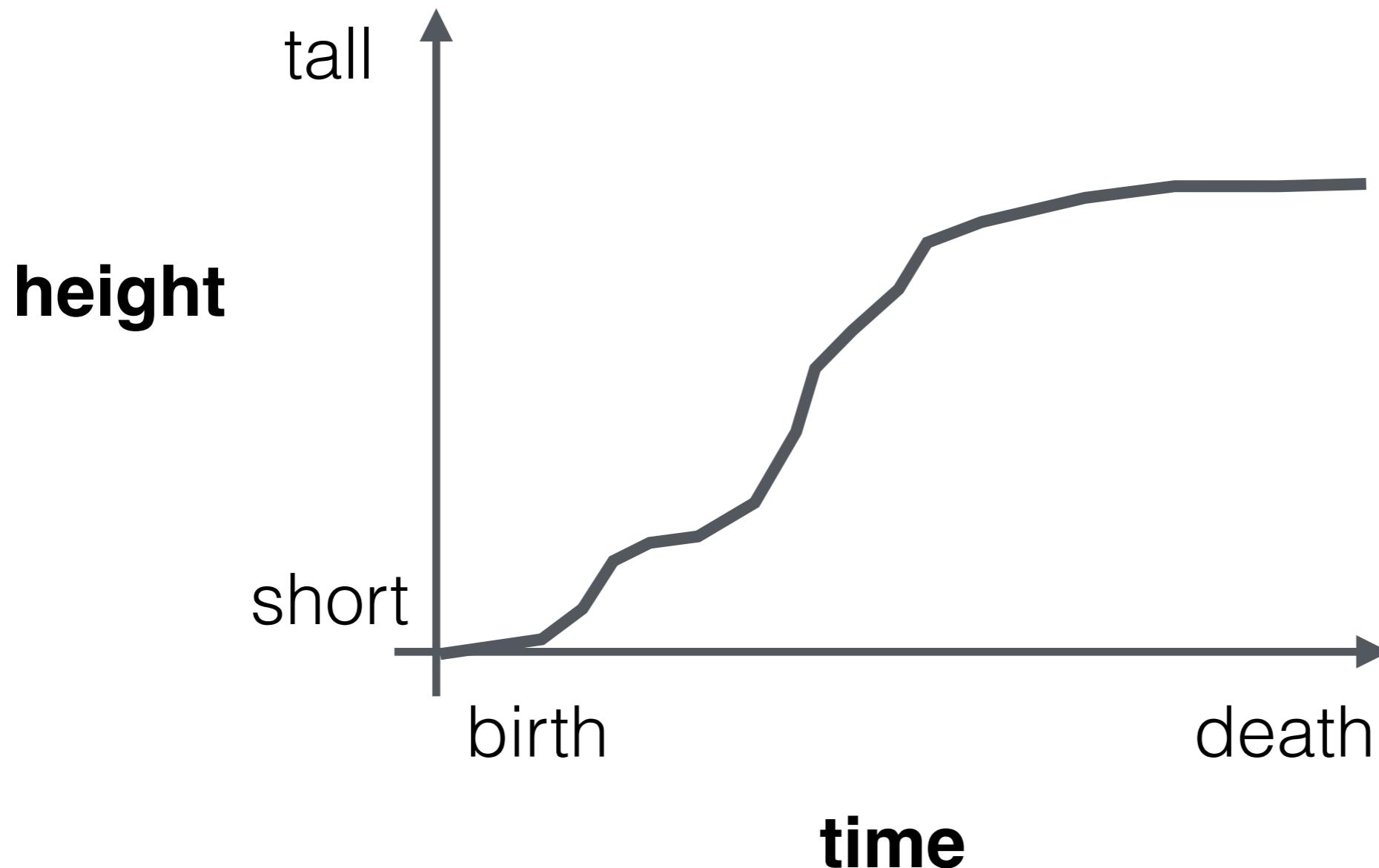
# Everything must have **meaning**



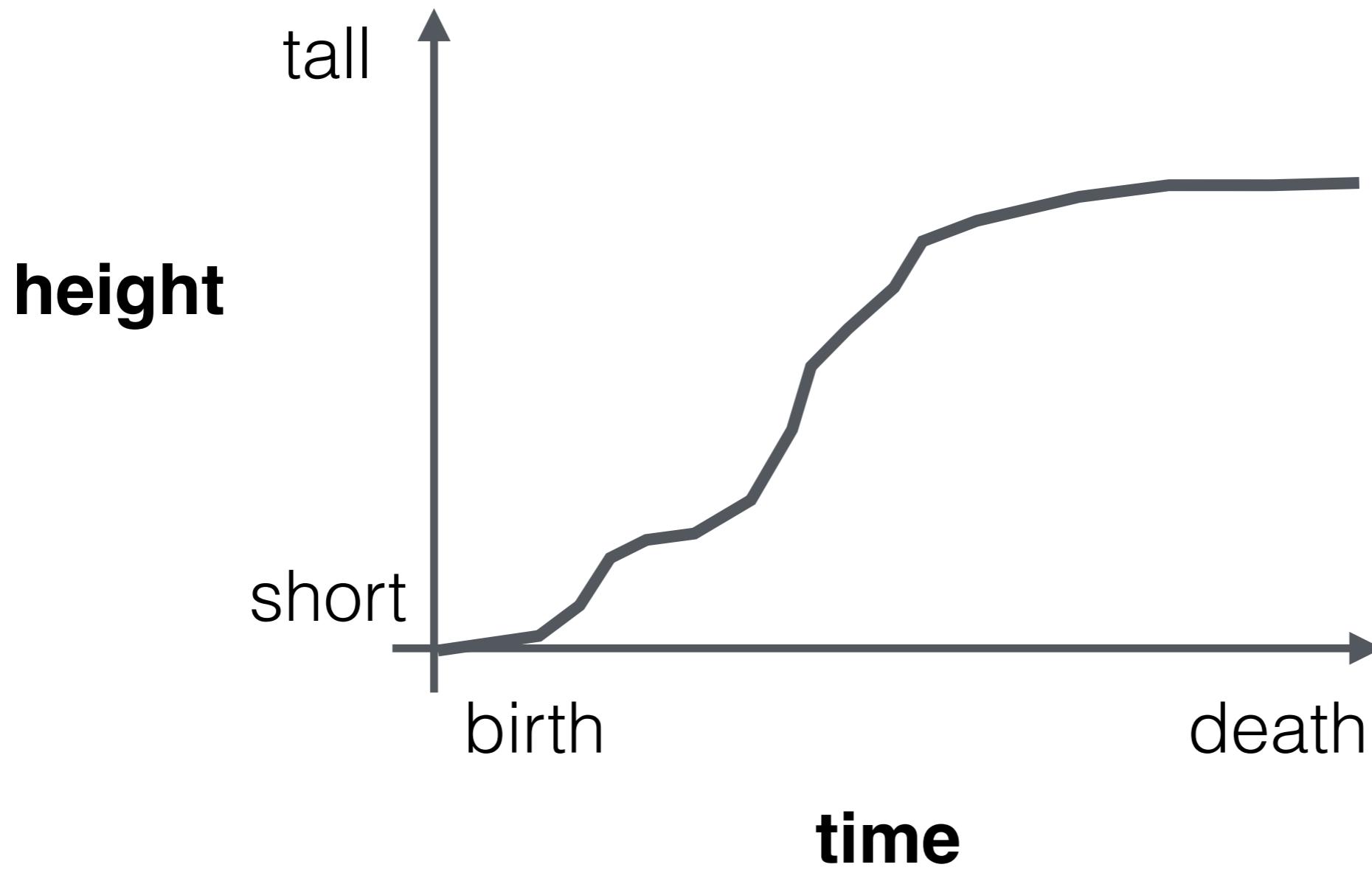
brief vocabulary  
lesson

**Semiotics:** the study  
of signs and symbols  
in communication

# Everything must have **meaning**

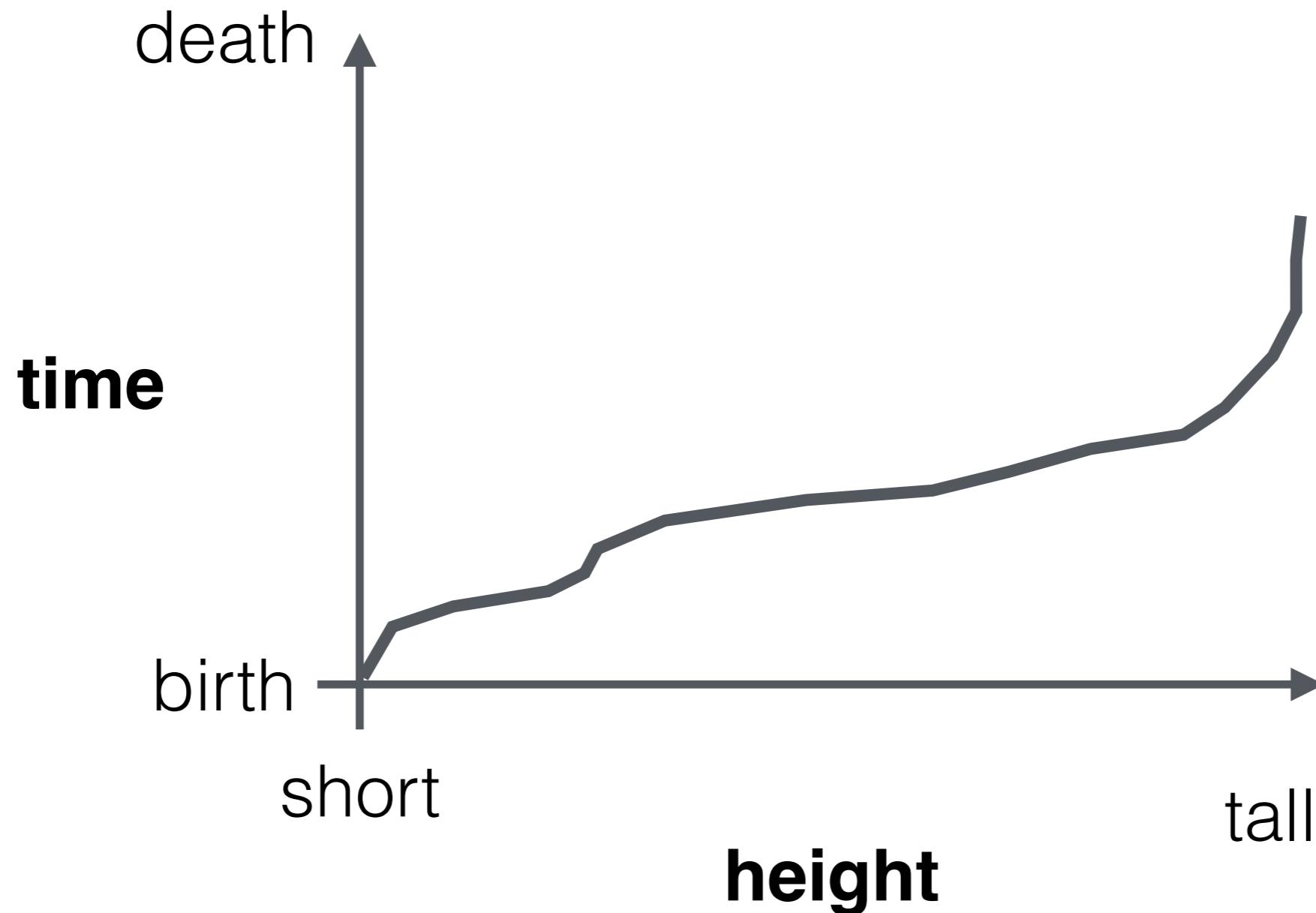


# Everything must have **meaning**

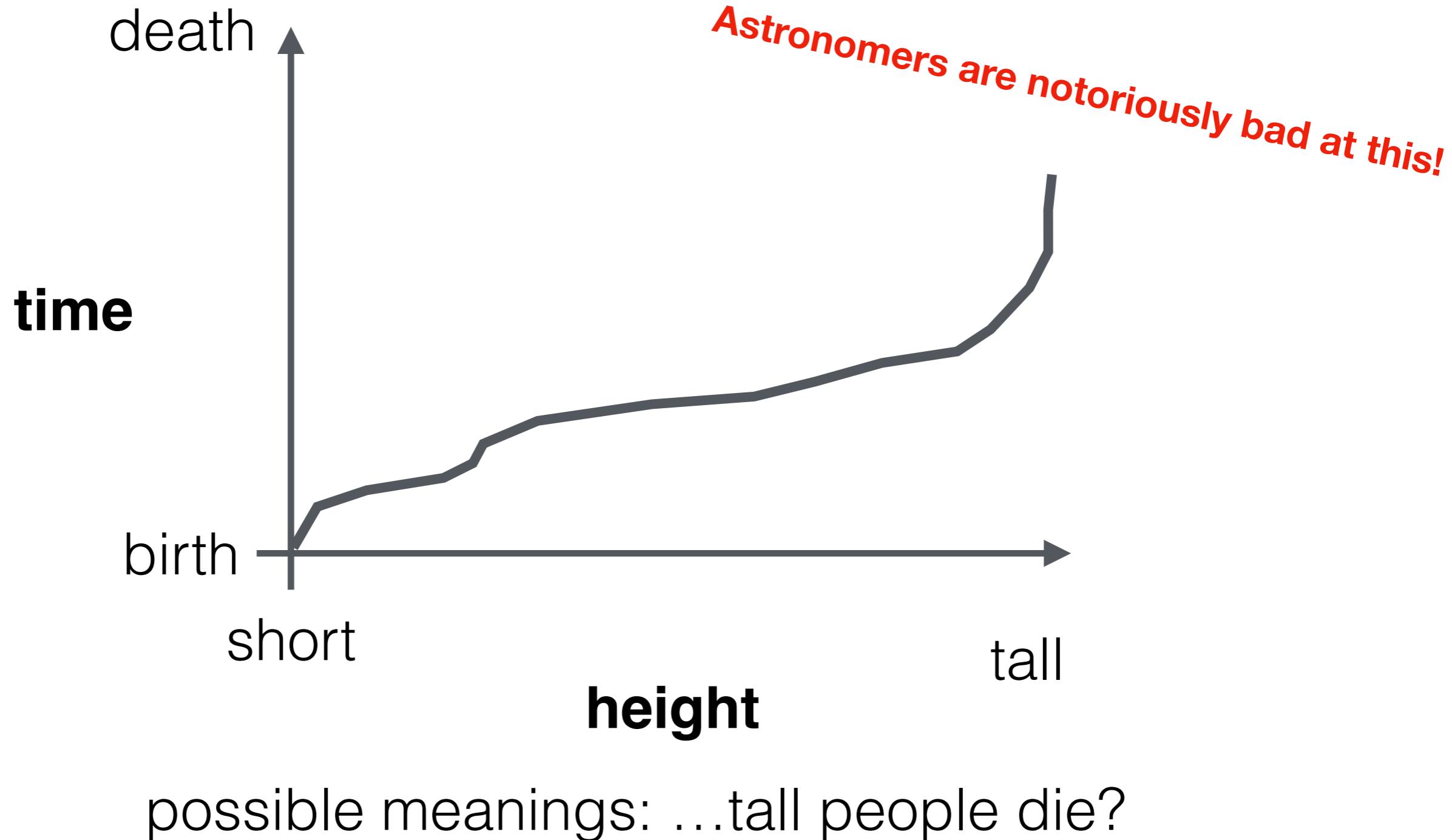


possible meanings: growth over time, growth-spurts,

# Everything must have **meaning**



# Everything must have **meaning**



**Visualization is about managing  
the viewer's attention!**

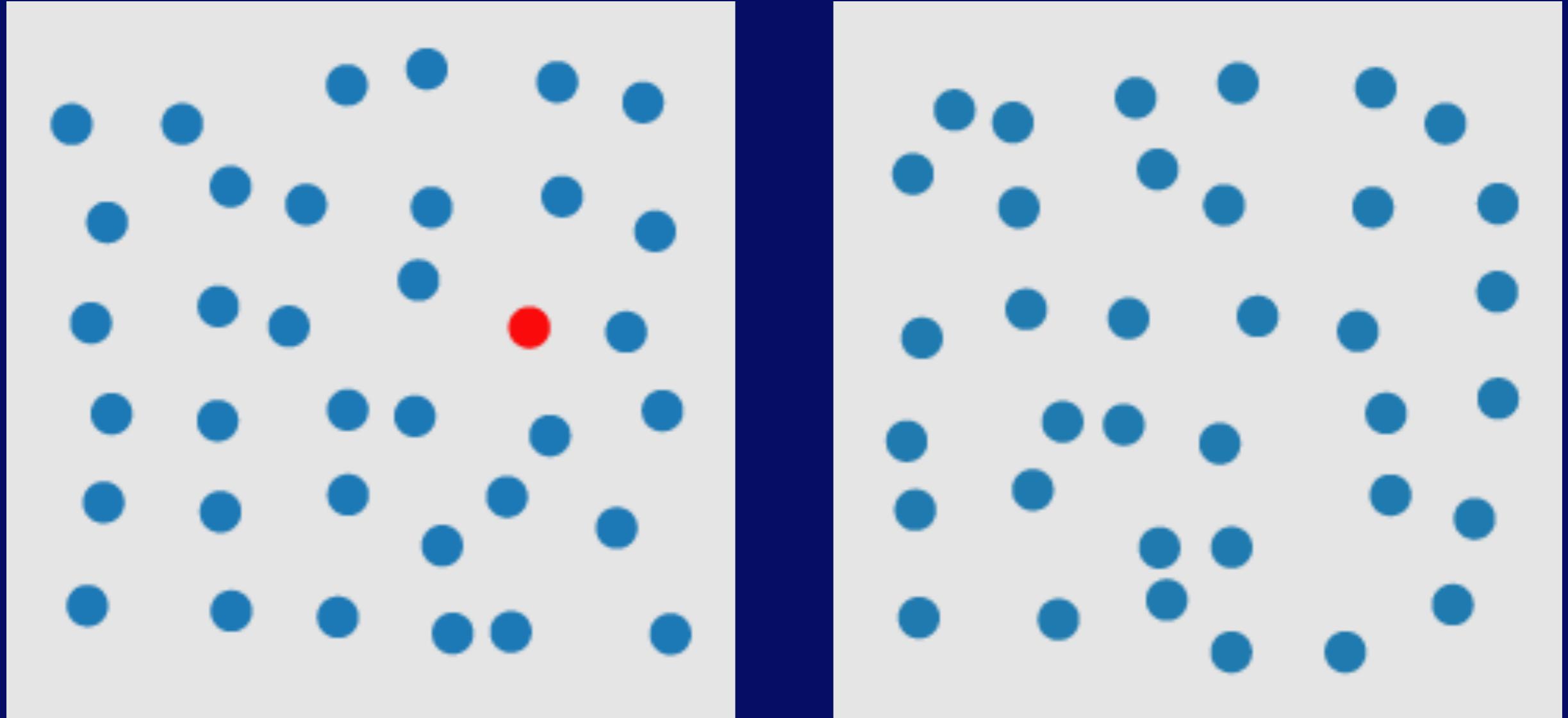


**human vision is not like  
photography!**

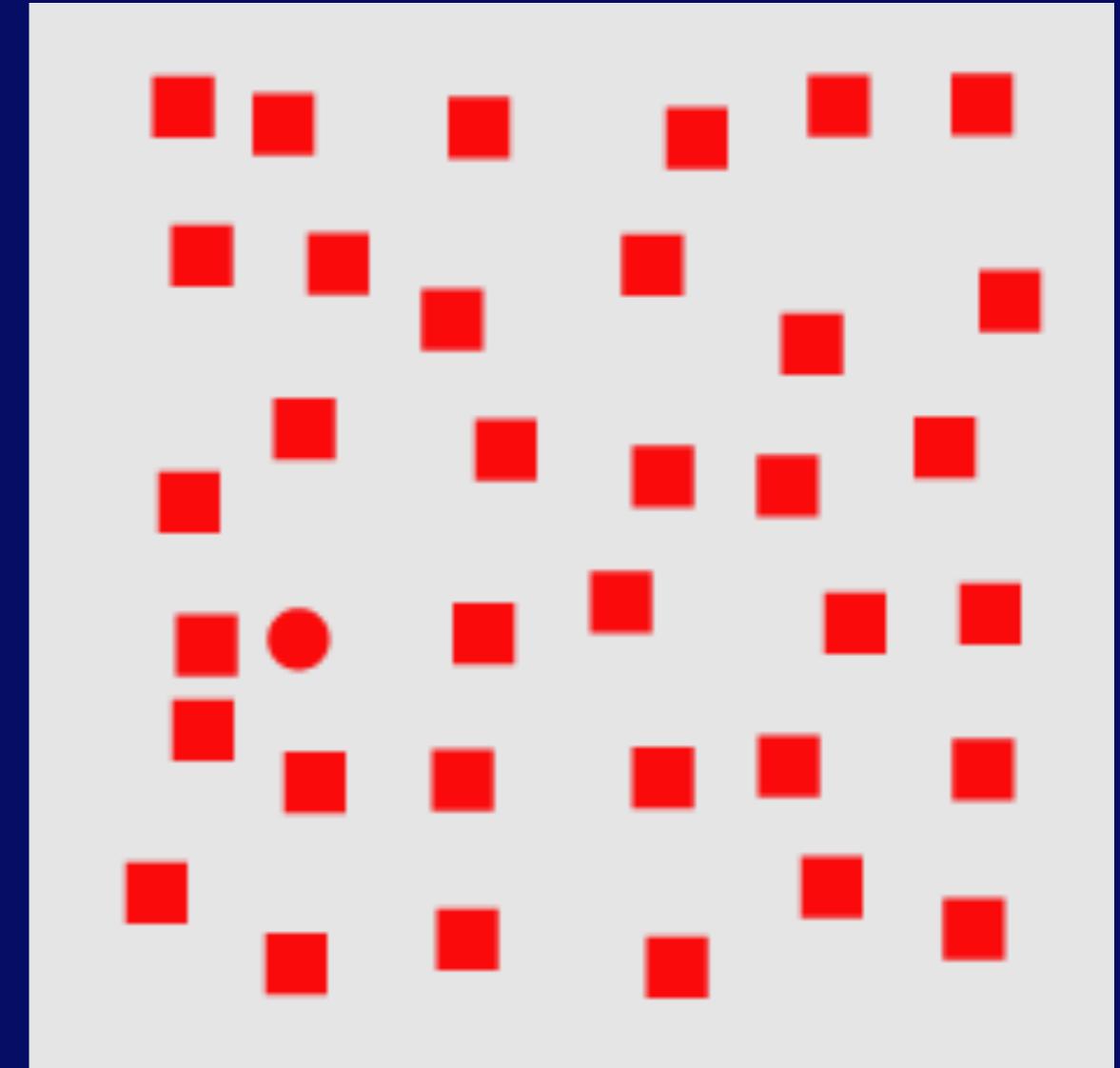
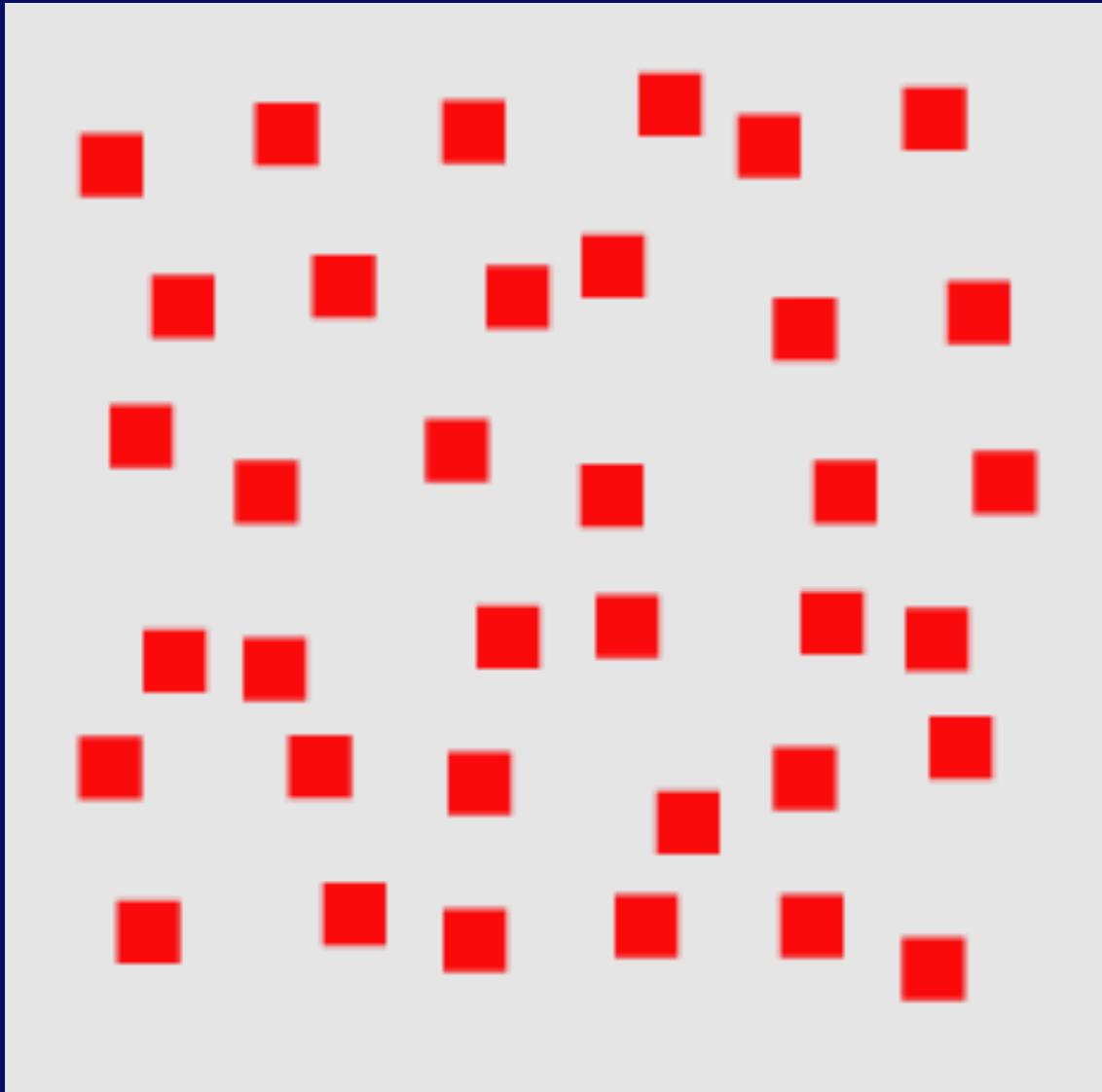
**pre-attentive processing**

# Let's do an experiment!

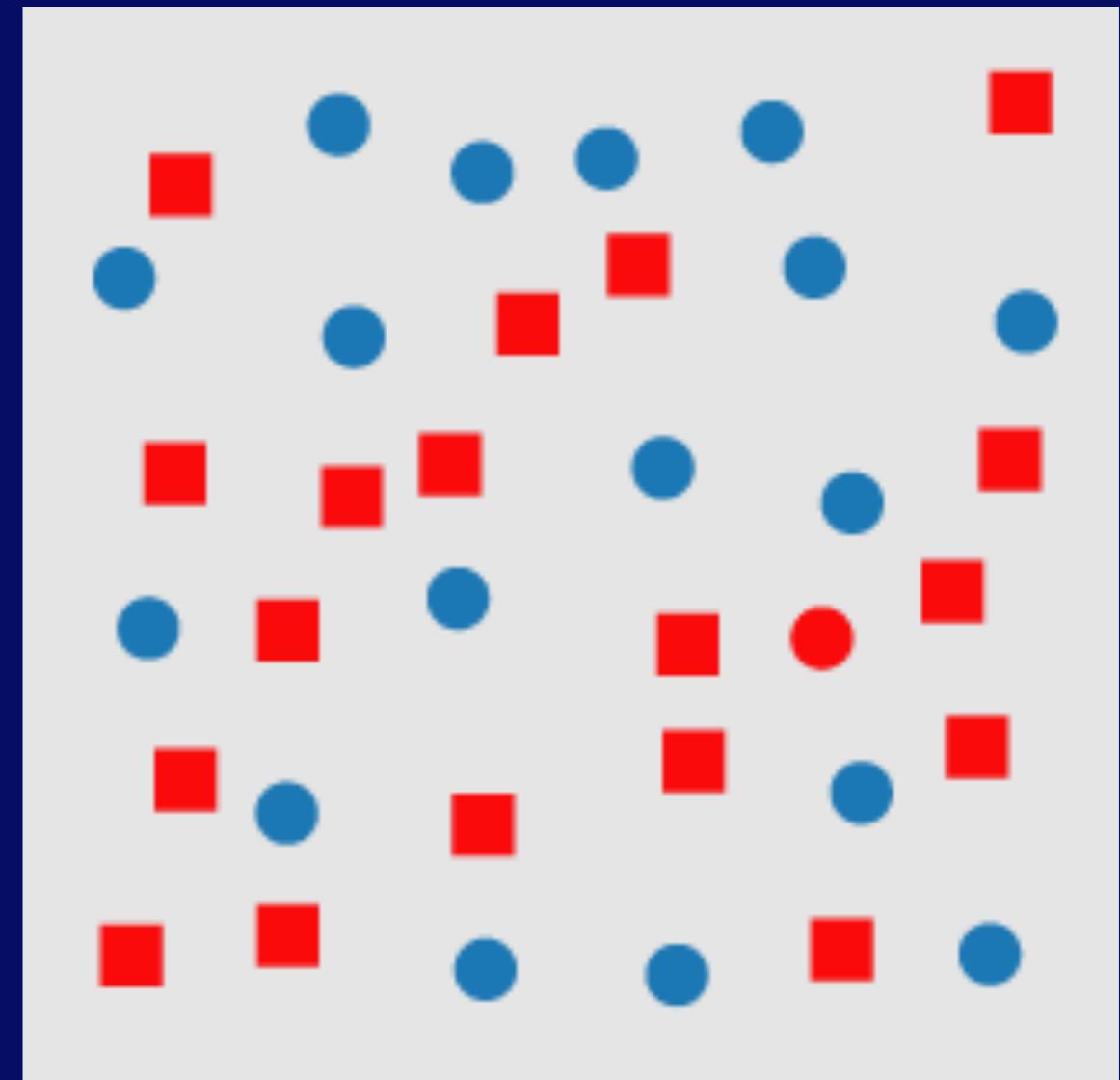
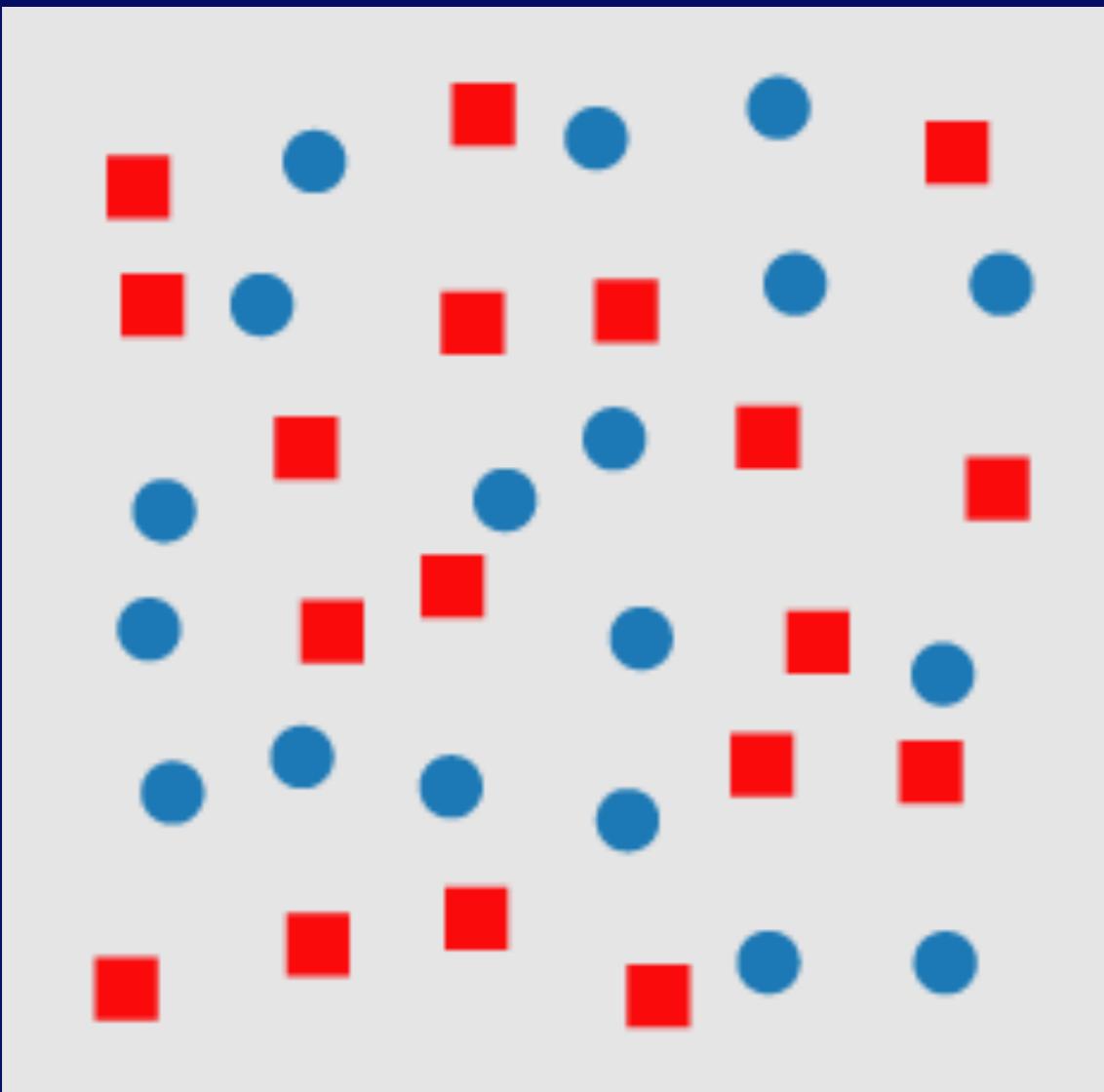
<https://www.csc2.ncsu.edu/faculty/healey/PP/>



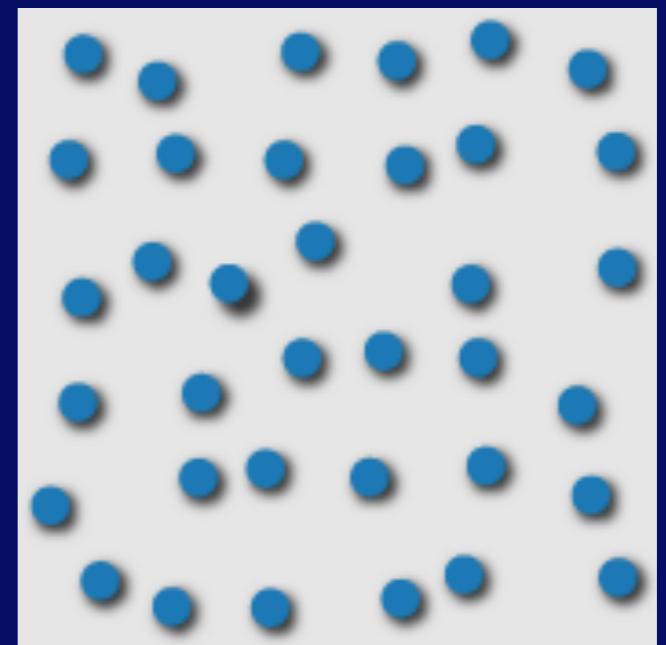
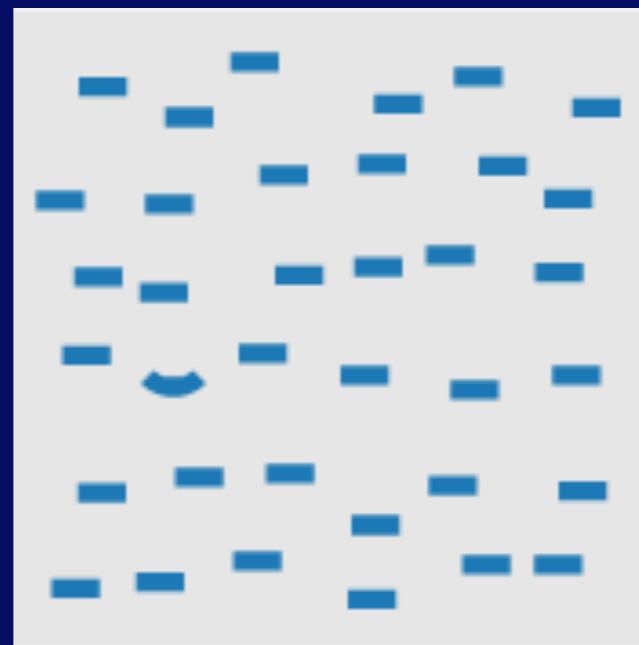
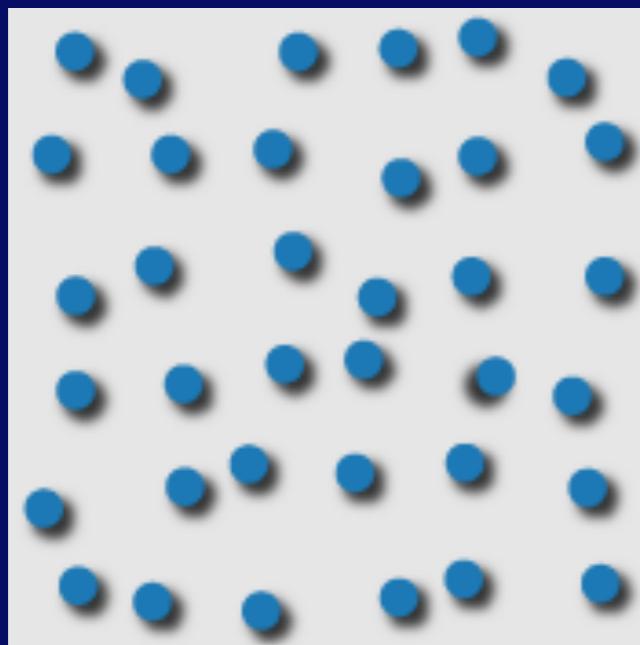
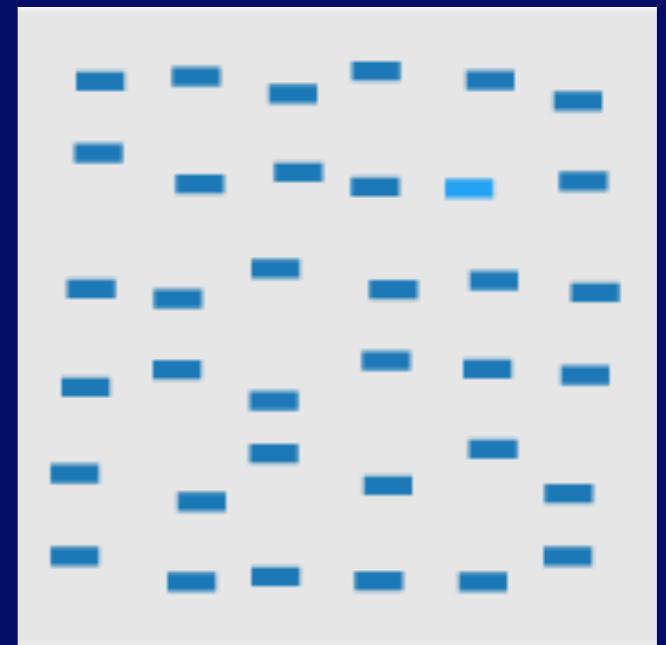
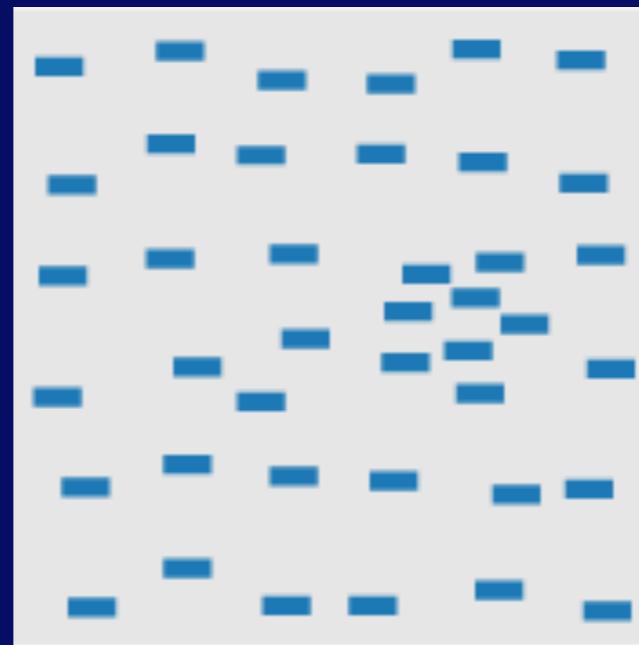
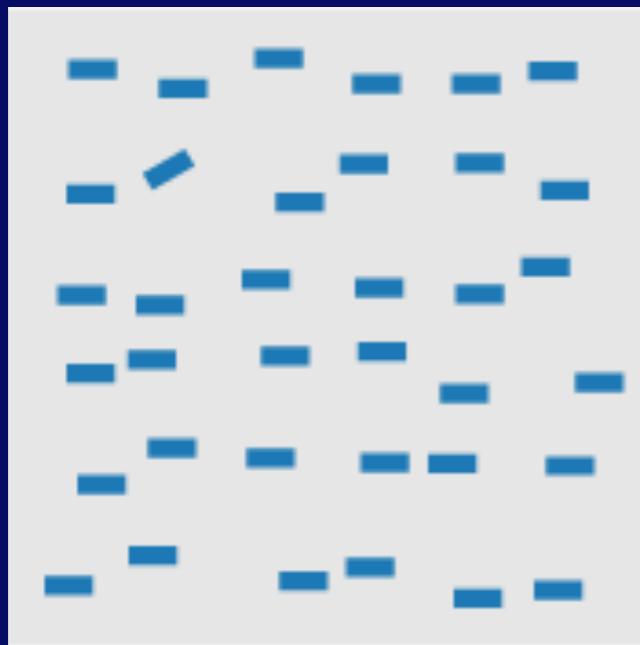
**pre-attentive task**



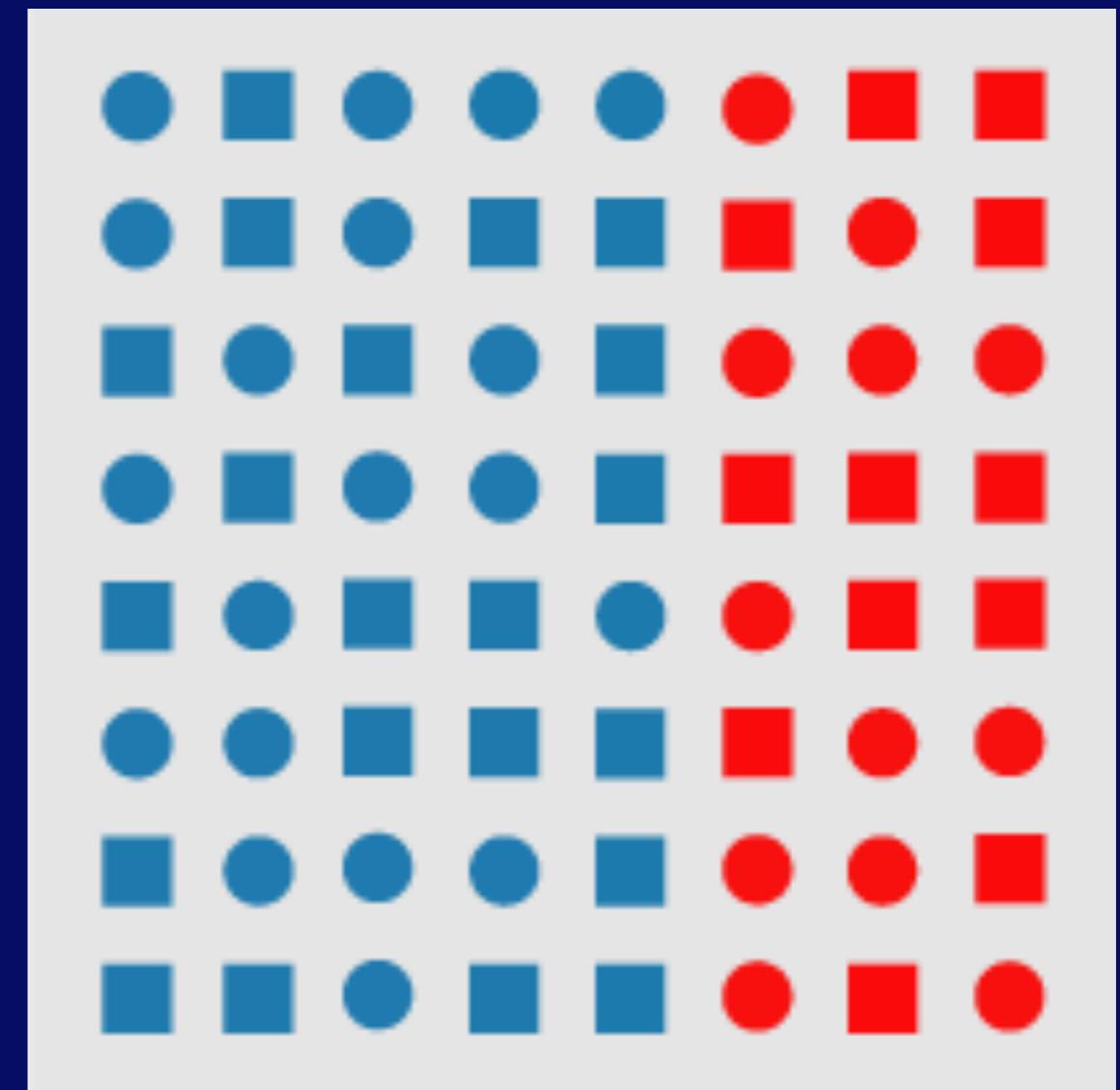
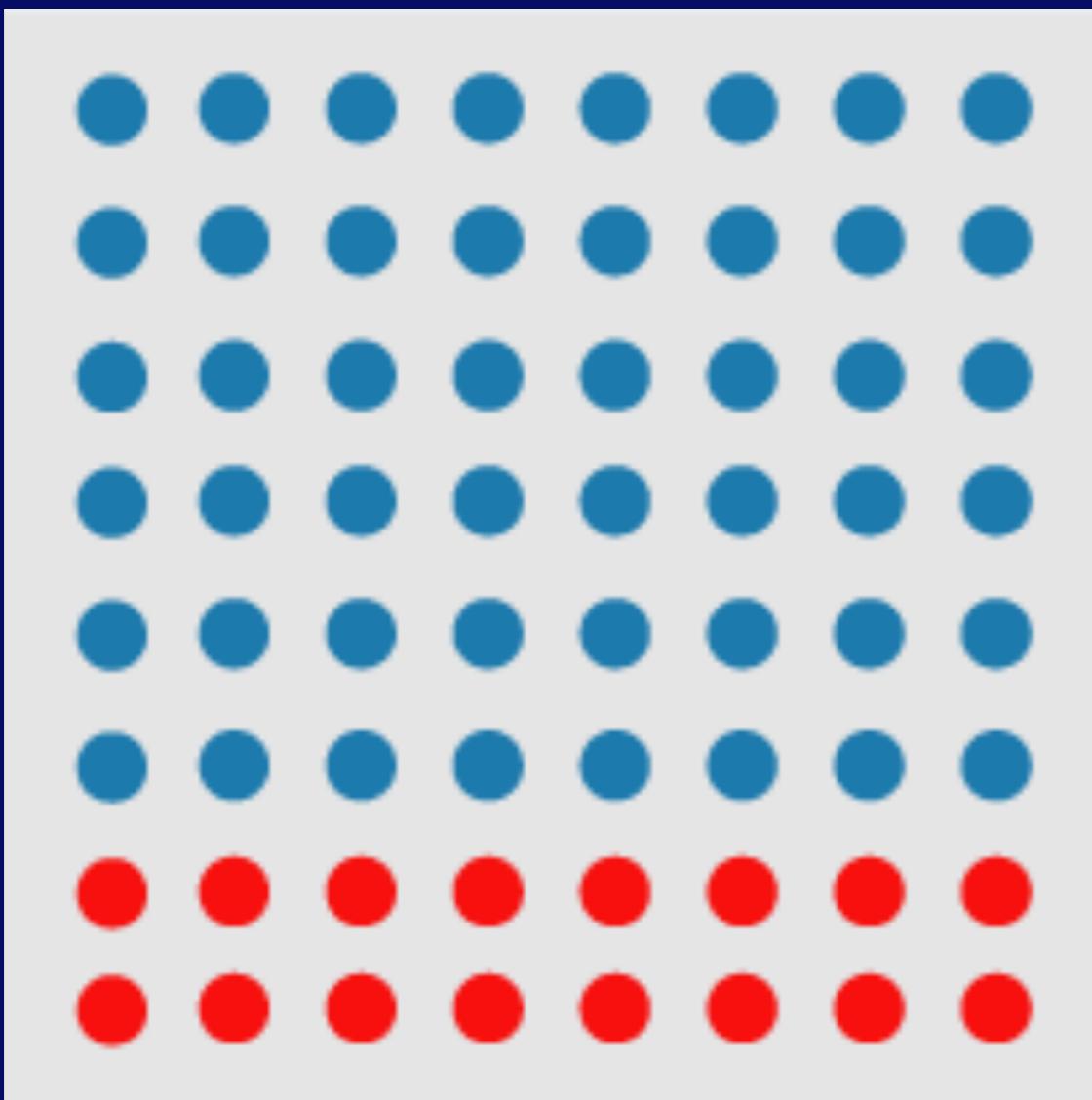
pre-attentive task



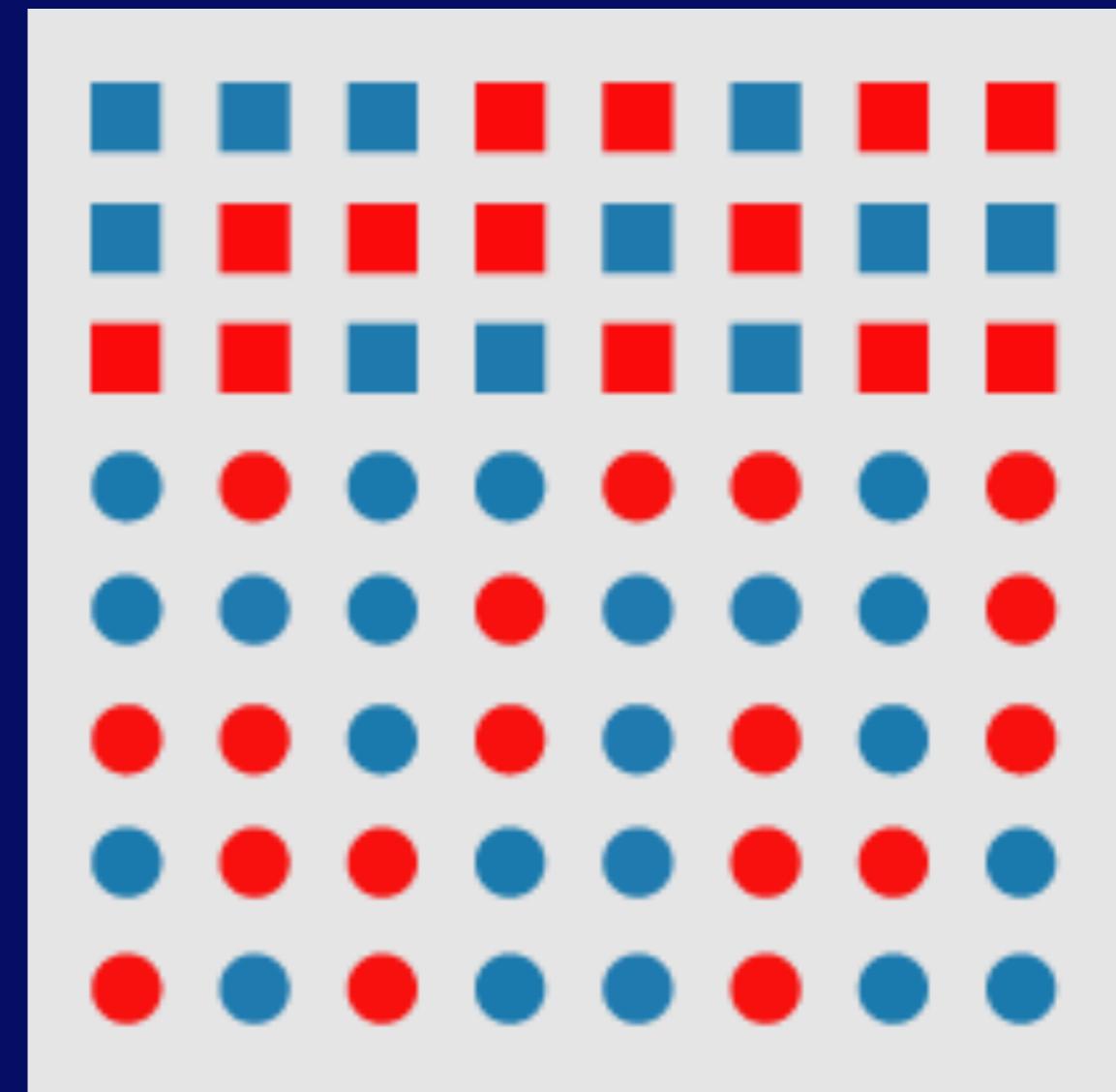
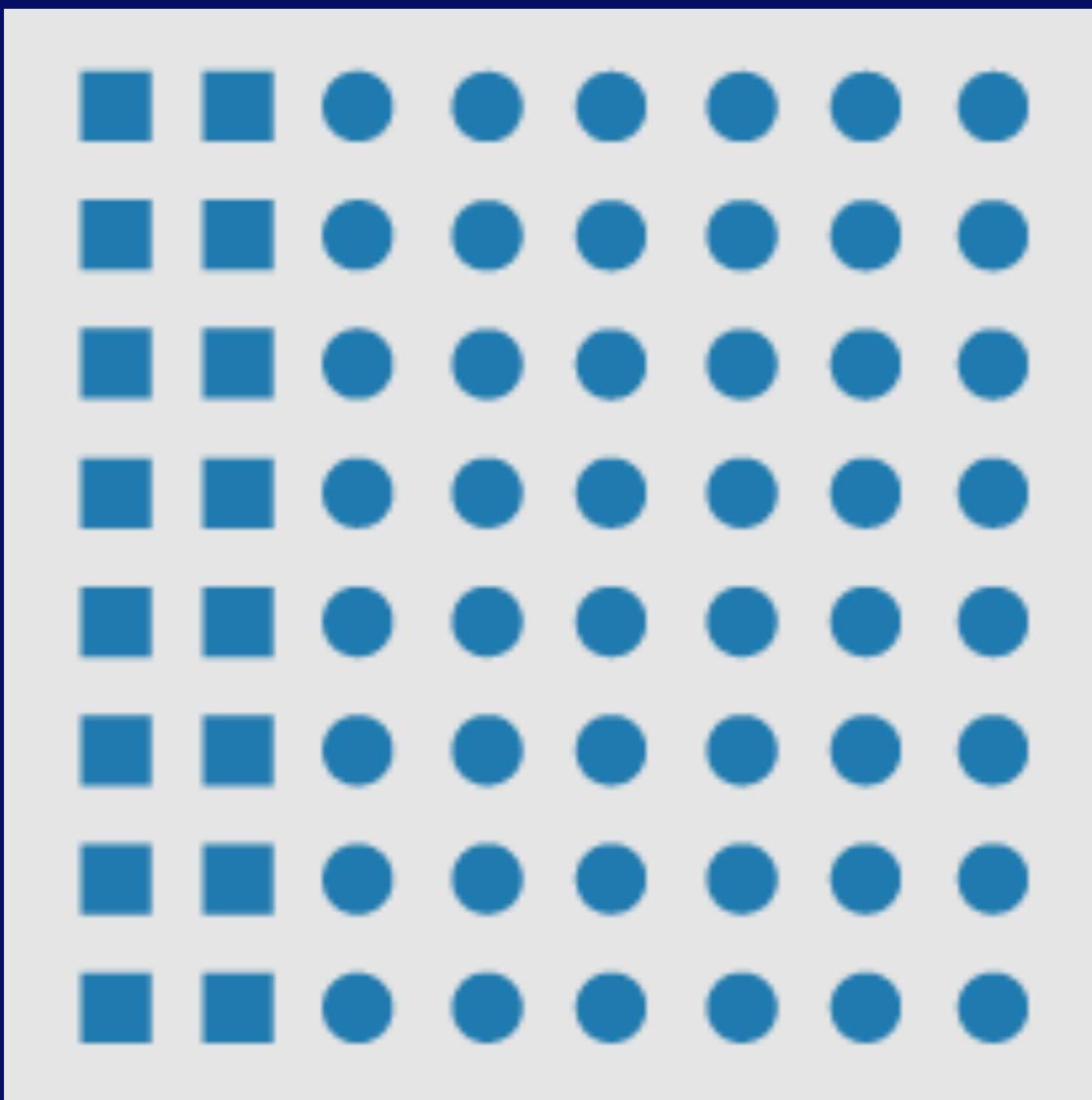
# serial search



# Feature Hierarchy

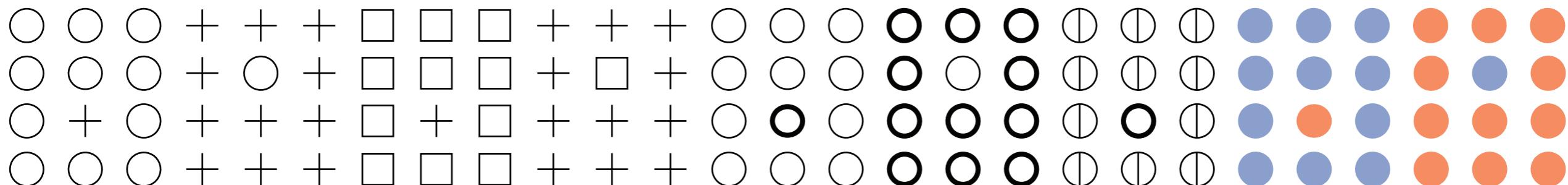


# Feature Hierarchy

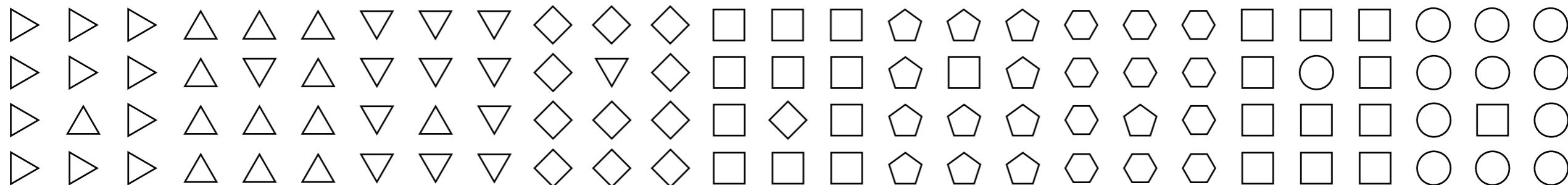


# 1) Choose strong visual boundaries

Strong visual boundaries

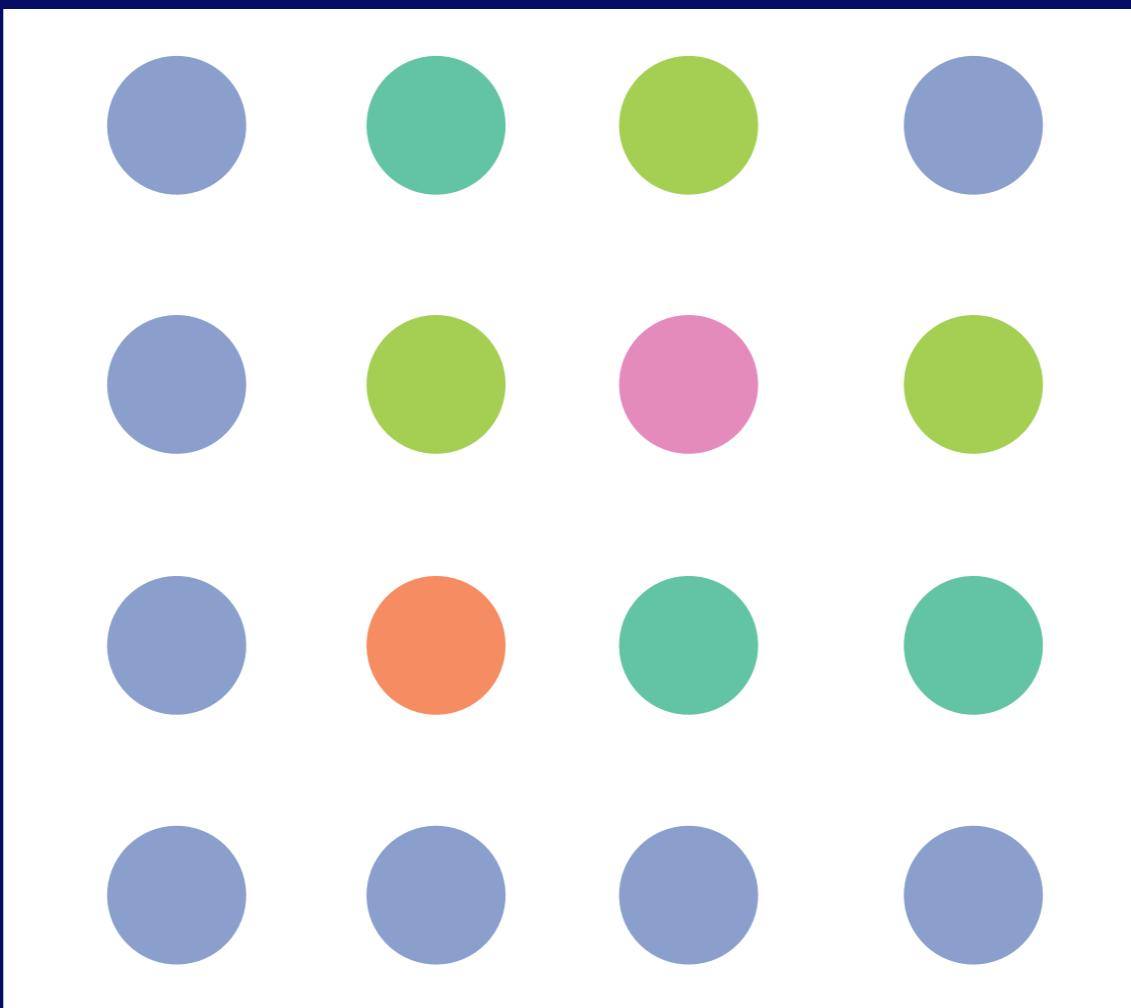


Weak visual boundaries



Krzywinski & Wong, Nature Methods, 2013

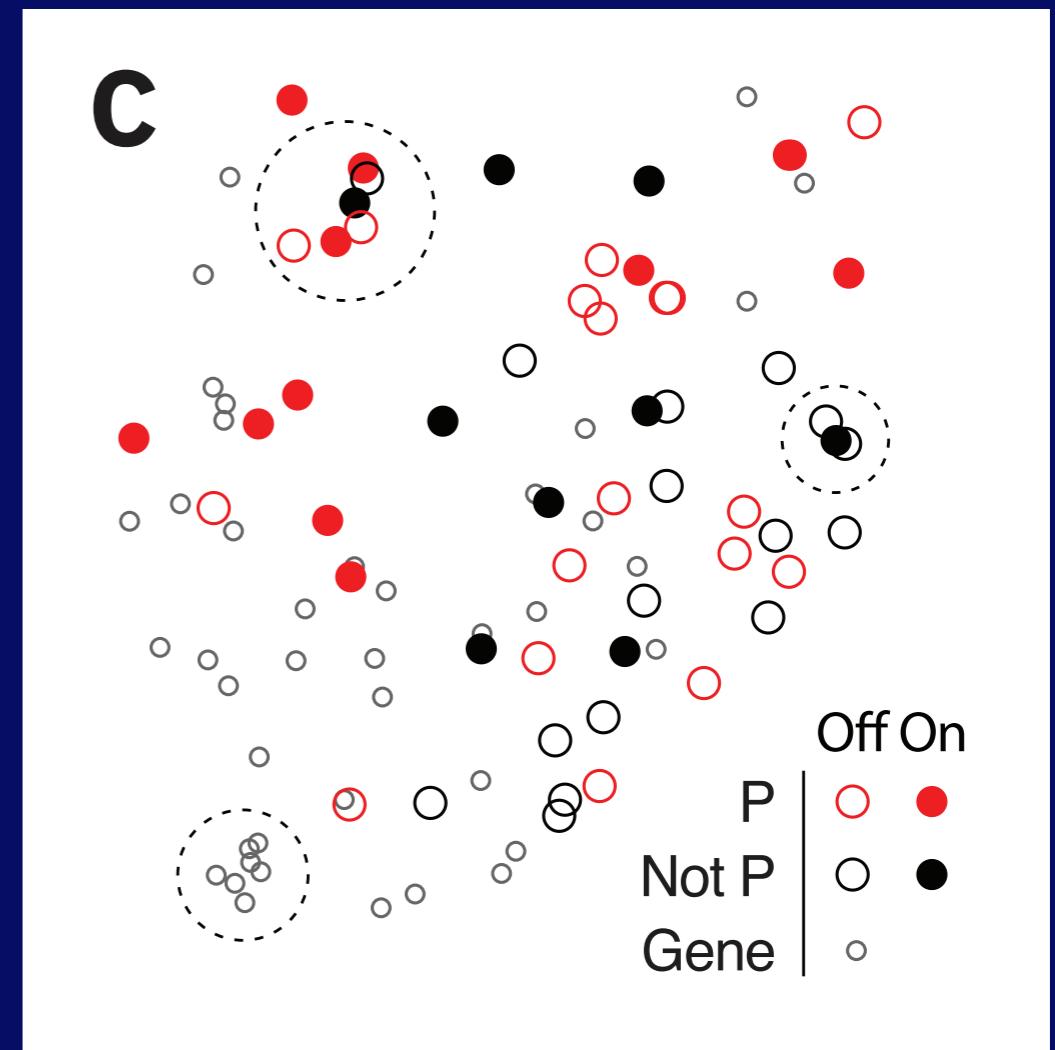
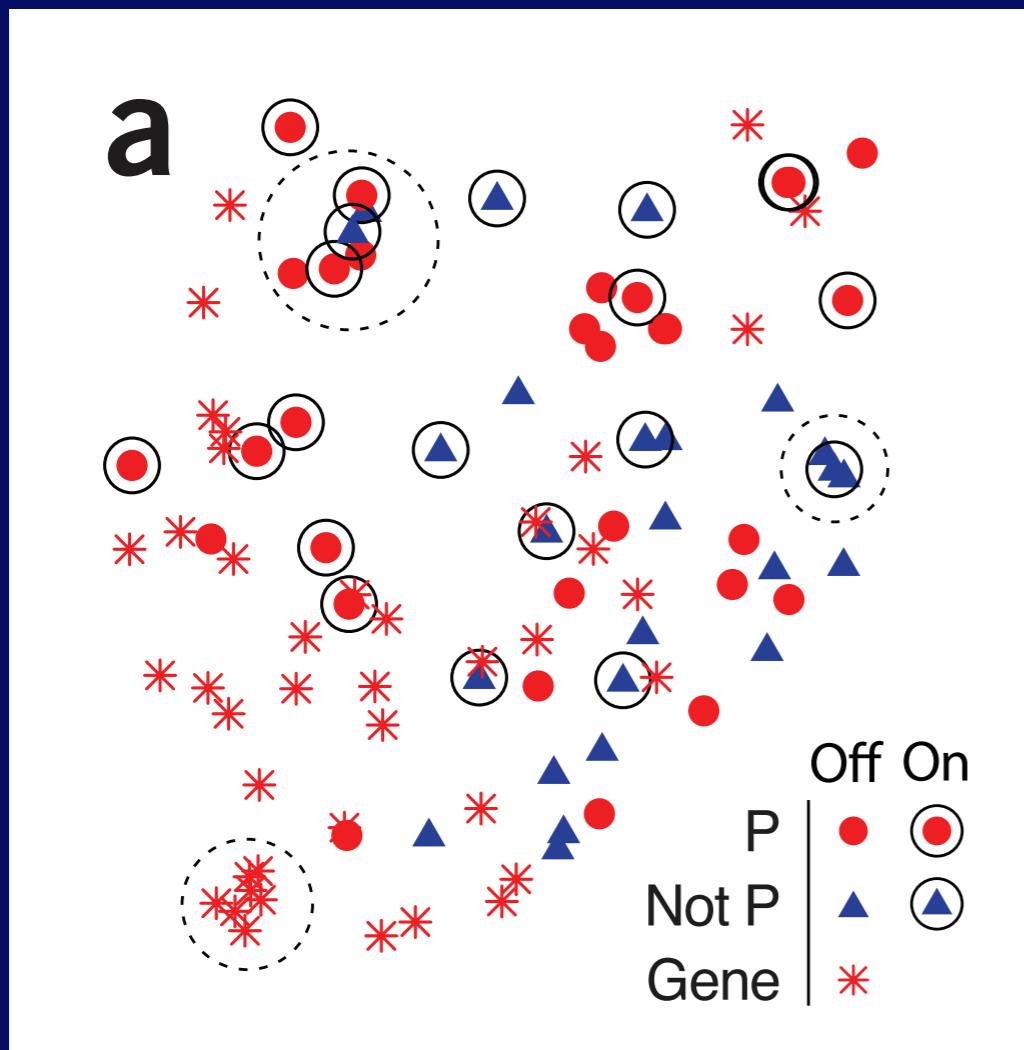
## 2) Use colour judiciously



Krzywinski & Wong, Nature Methods, 2013

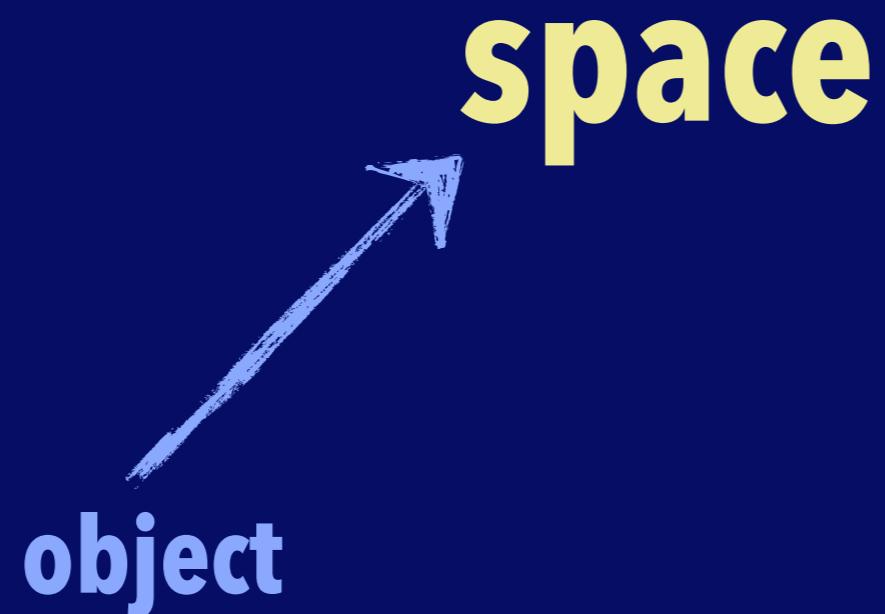
<http://colorbrewer2.org/>

# 3) Represent data hierarchies



Krzywinski & Wong, Nature Methods, 2013

space

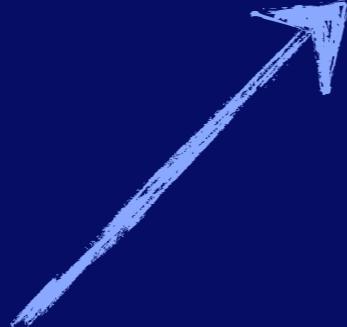


space

object

**negative space**

**space**



**object**

# figure- ground organization

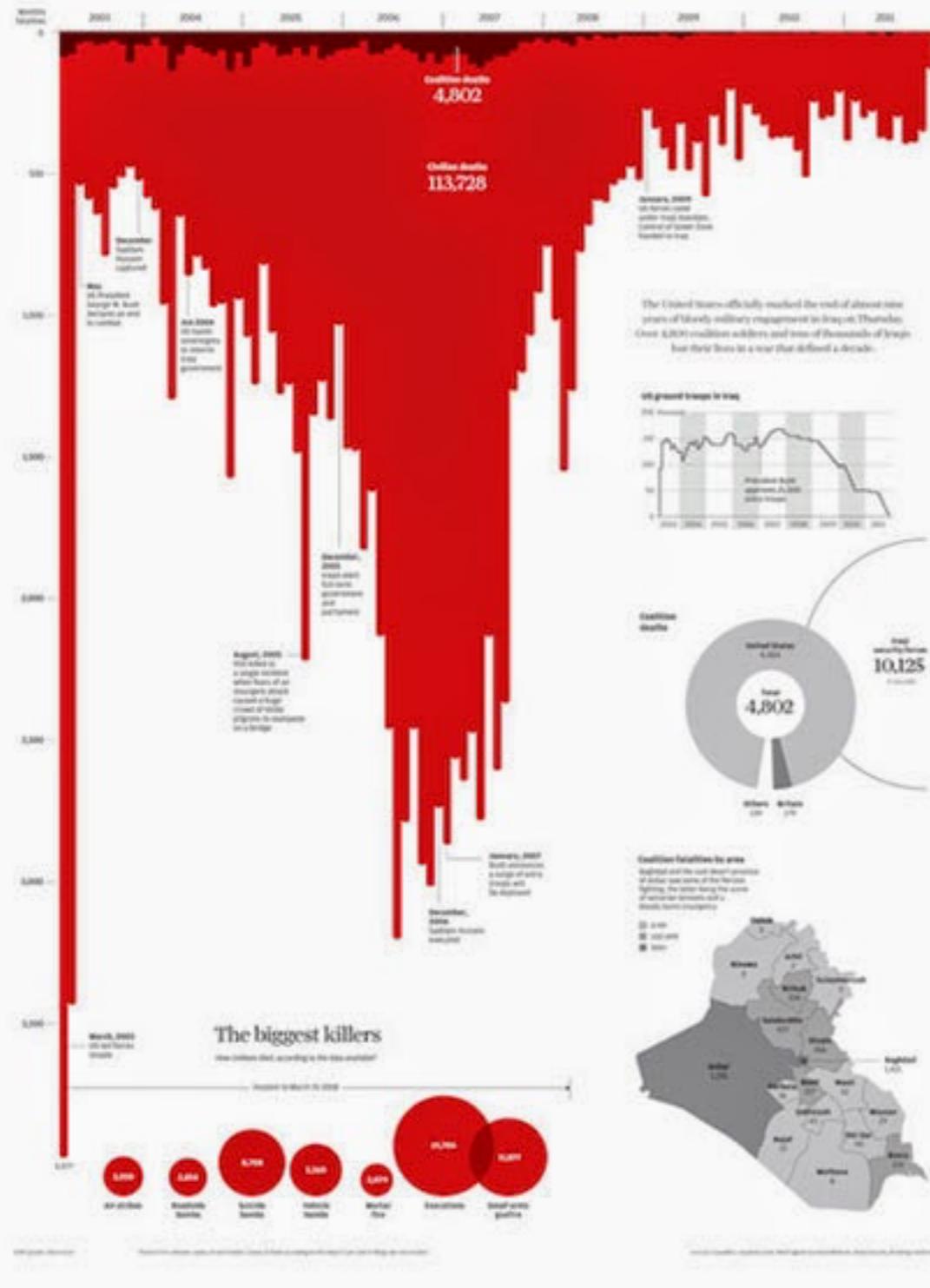
Credit: Volkswagen/Adam & Eve/DDB London



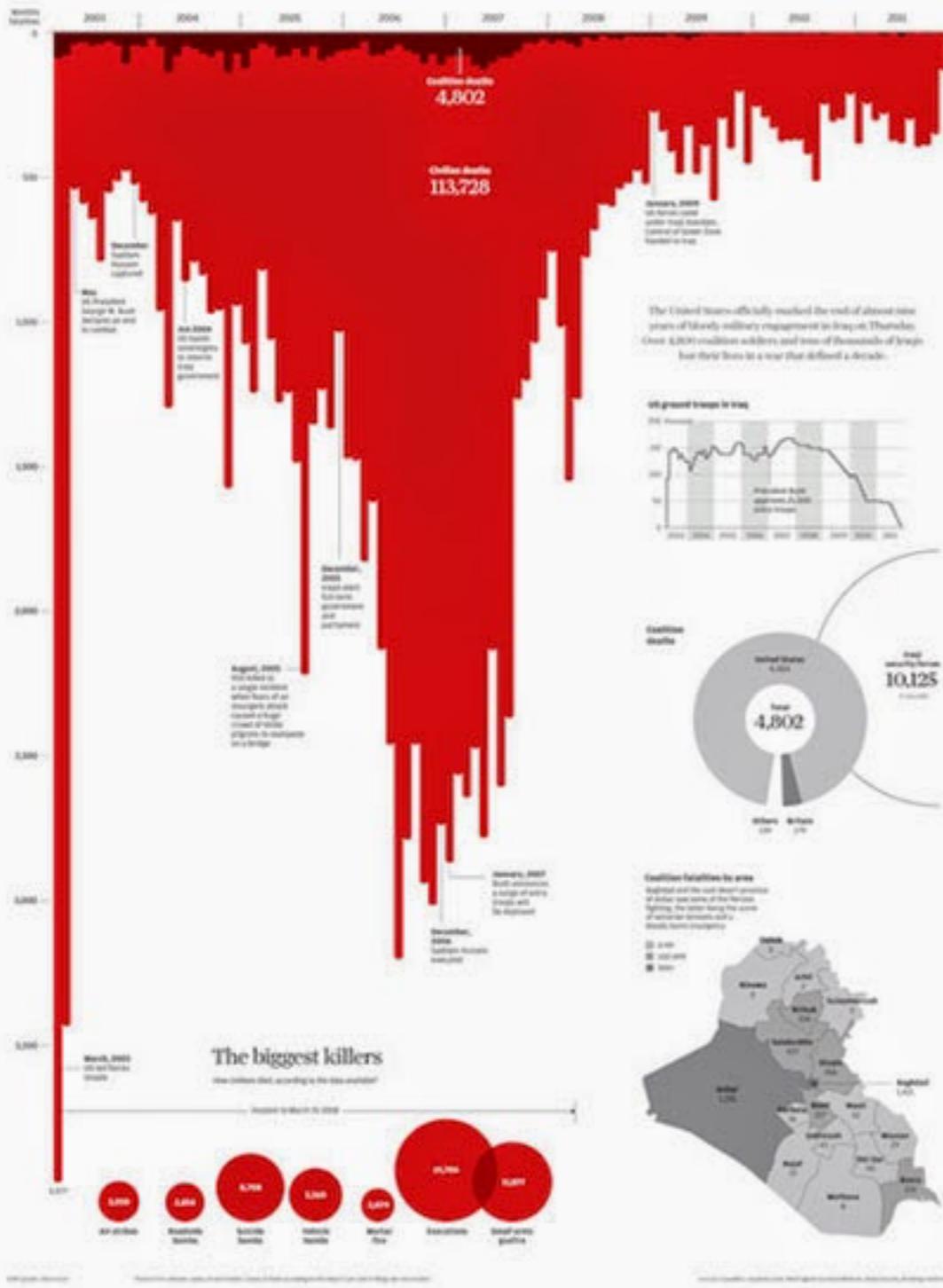
See film differently.  
Volkswagen supports independent cinema.



# Iraq's bloody toll

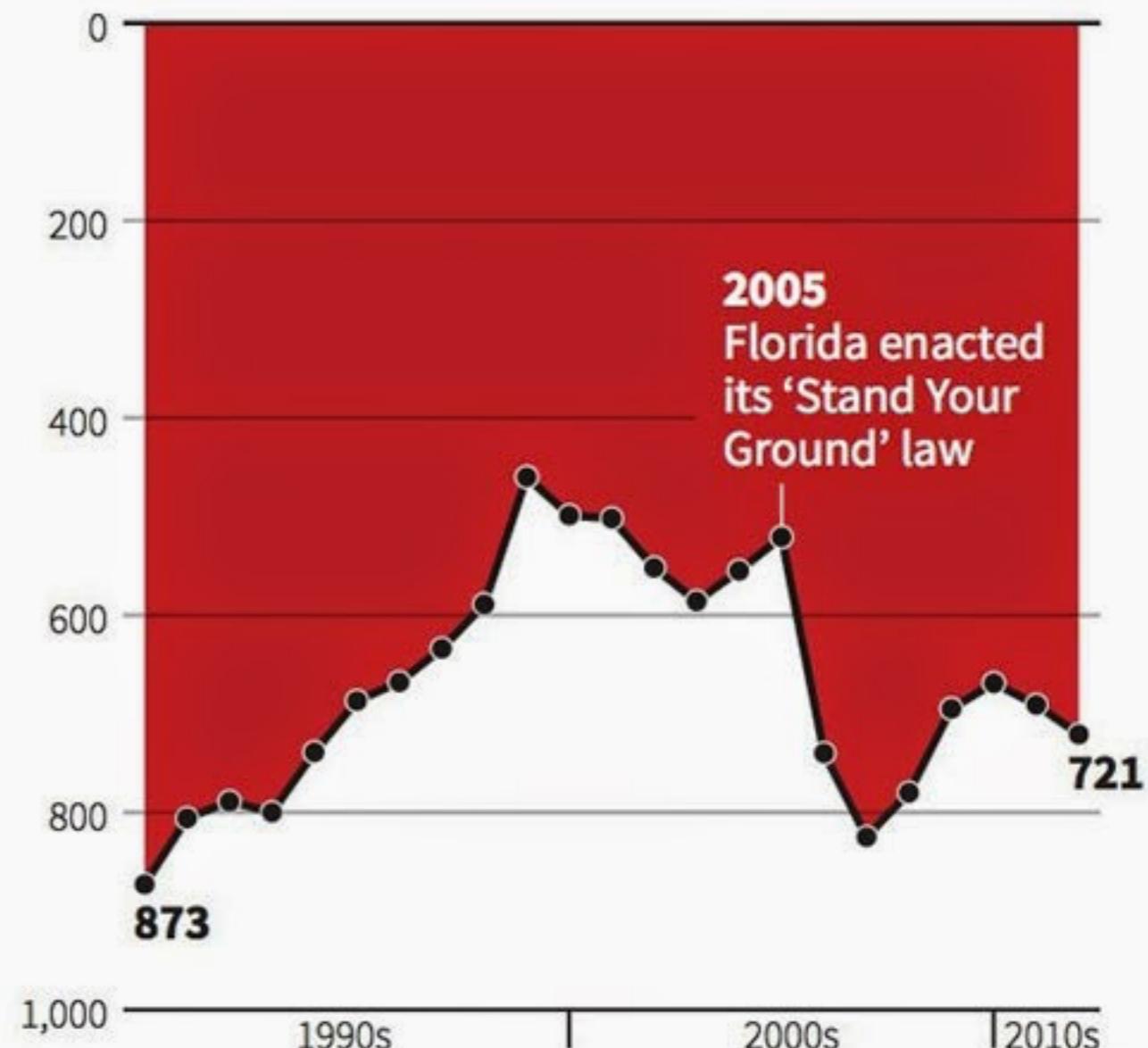


## Iraq's bloody toll



## **Gun deaths in Florida**

## Number of murders committed using firearms

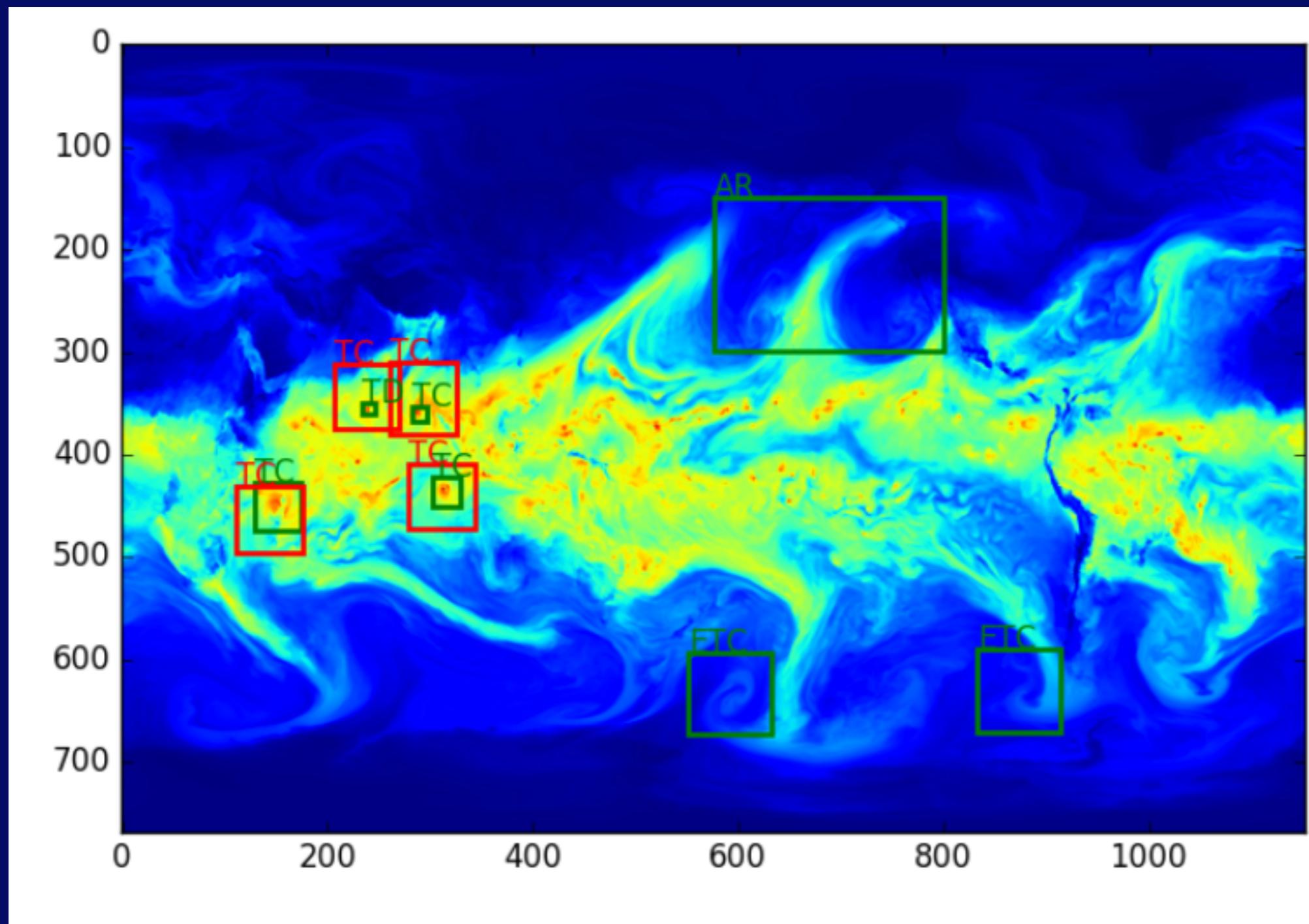


Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

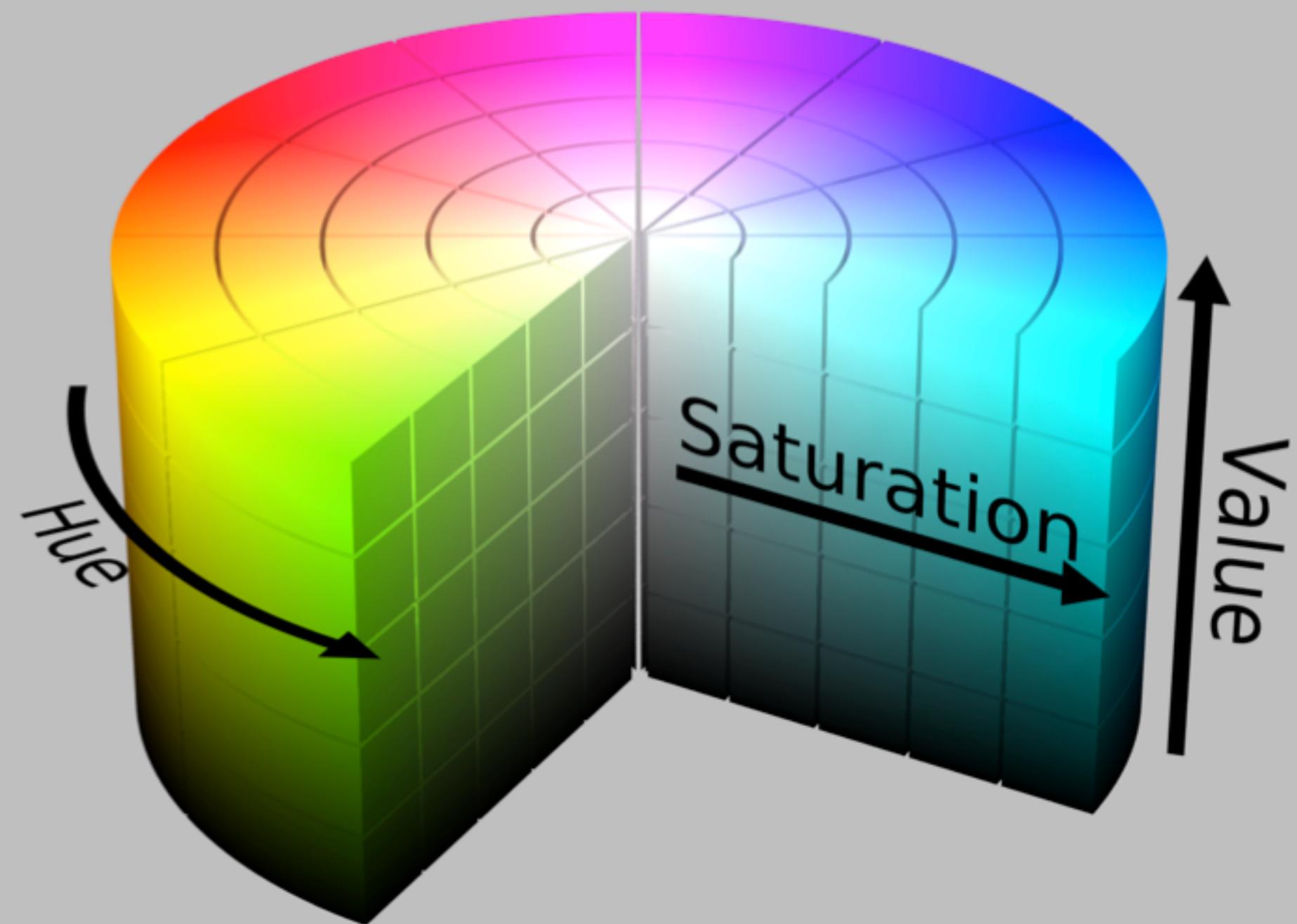


# Colour



# How does it look in colour?

- hue
- luminance
- saturation



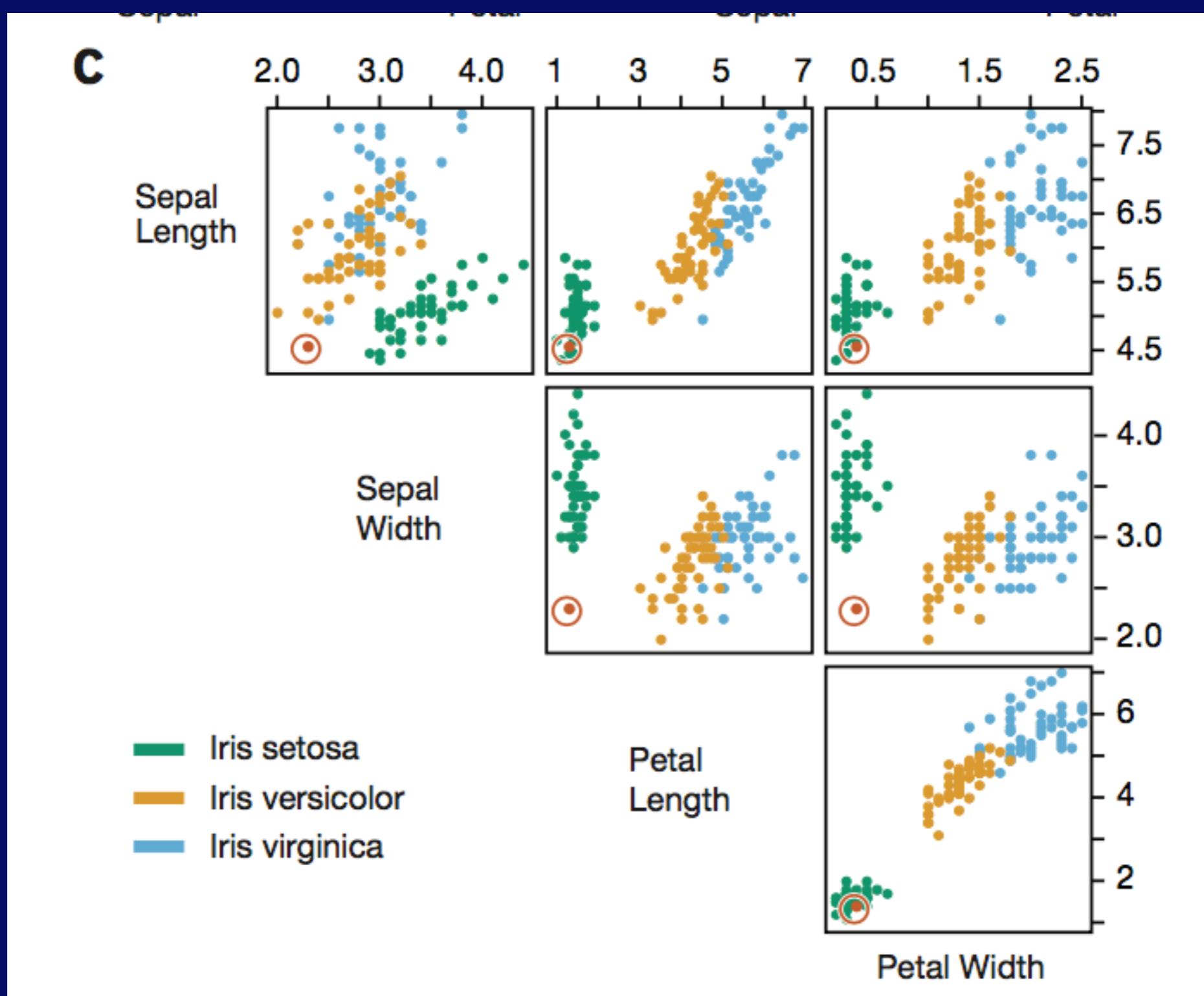
**What do you want to  
use colour for?**

**- colour does not help  
us understand shape  
and layout of objects in  
their environment**

+ colour acts as a visual  
attribute to objects

# Colour for nominal information coding\*

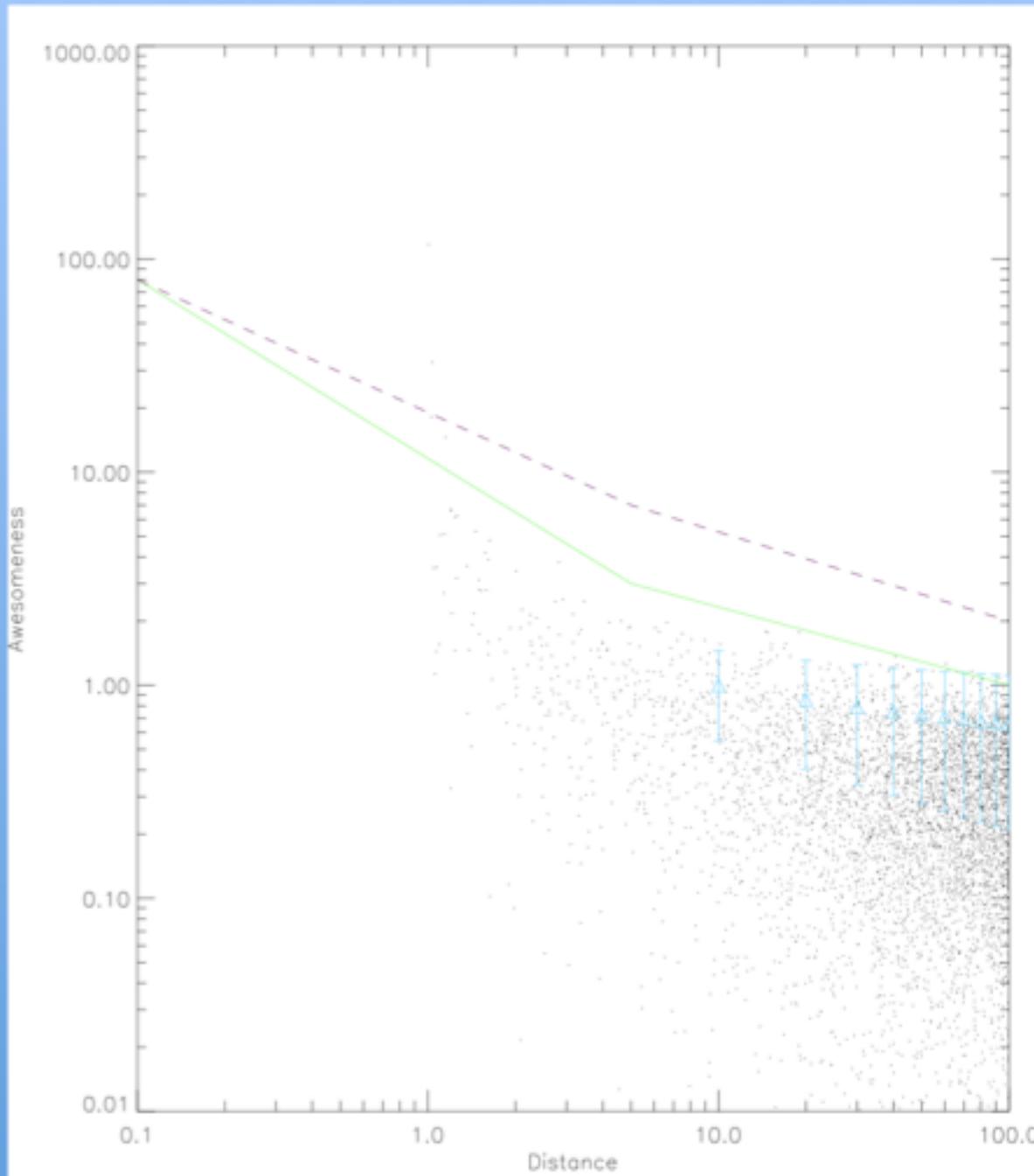
\* = labelling



**12 recommended  
colours: red, green,  
yellow, blue, black,  
white, pink, cyan, grey,  
orange, brown, purple**

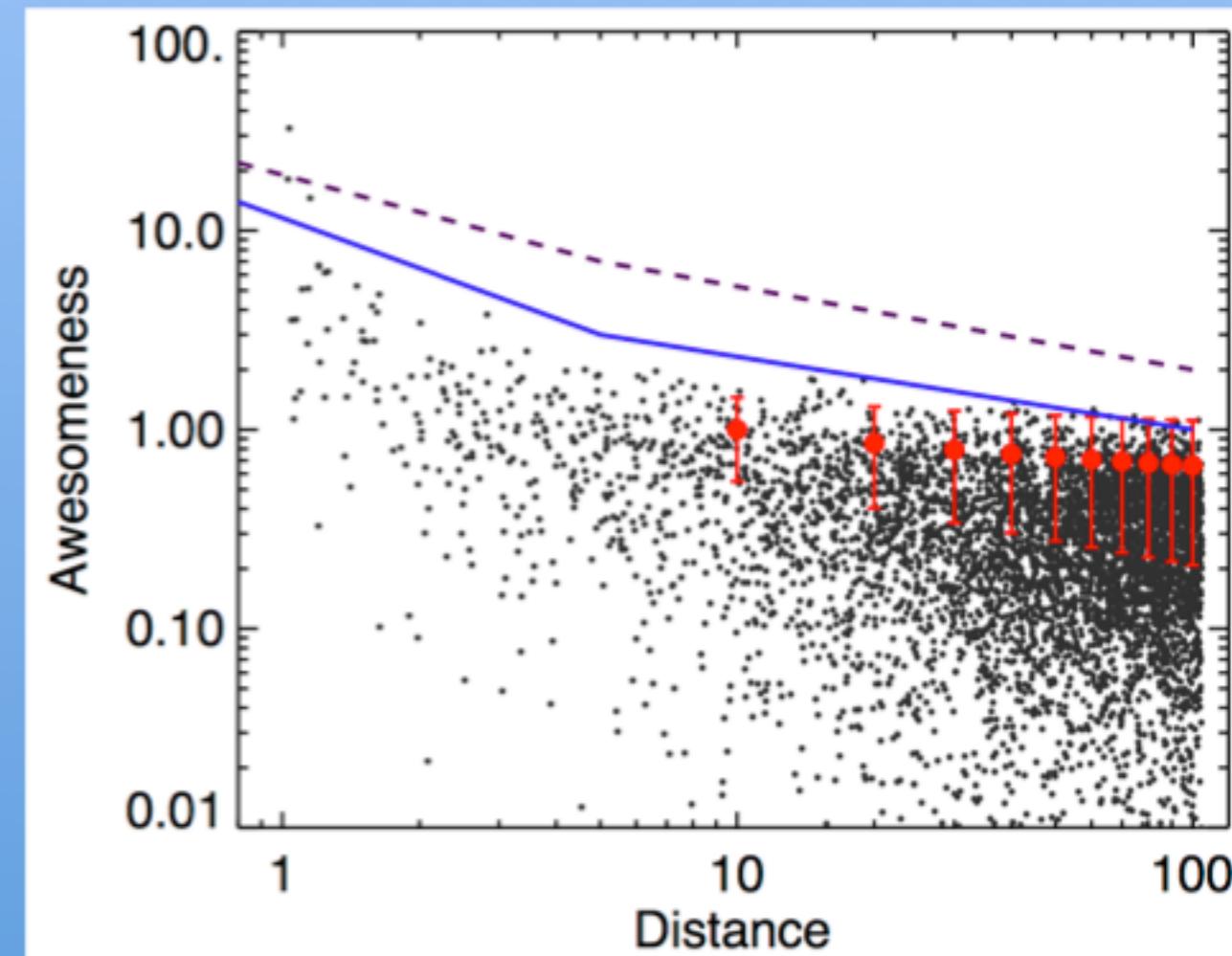
# Contrast is King

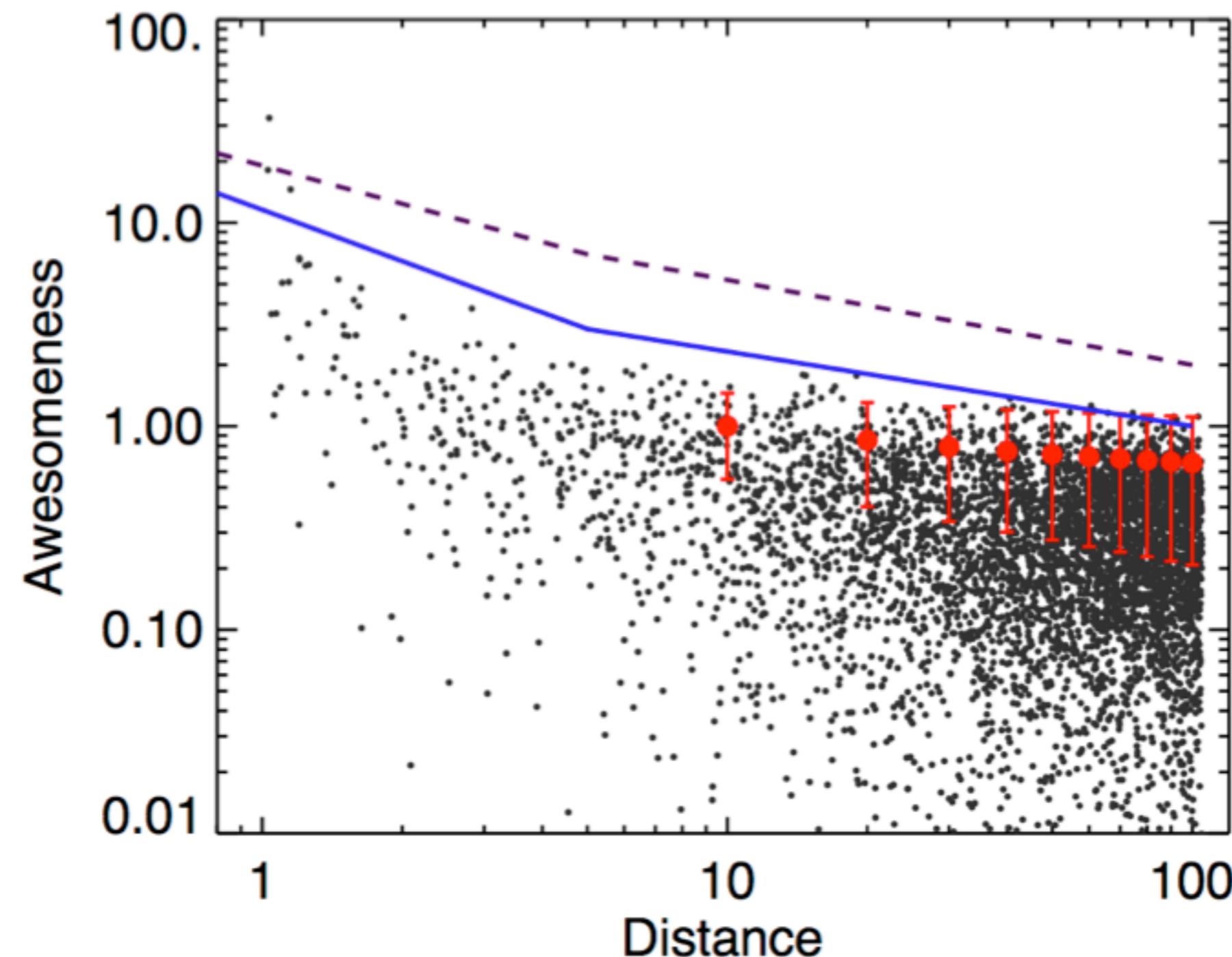
- The data matches the model remarkably well
- This changes everything
- Oh, this looks better on *my* screen...



# Contrast is King

- The data matches the model remarkably well
- This changes everything
- Now *you* can see it
- **Seeing is believing**





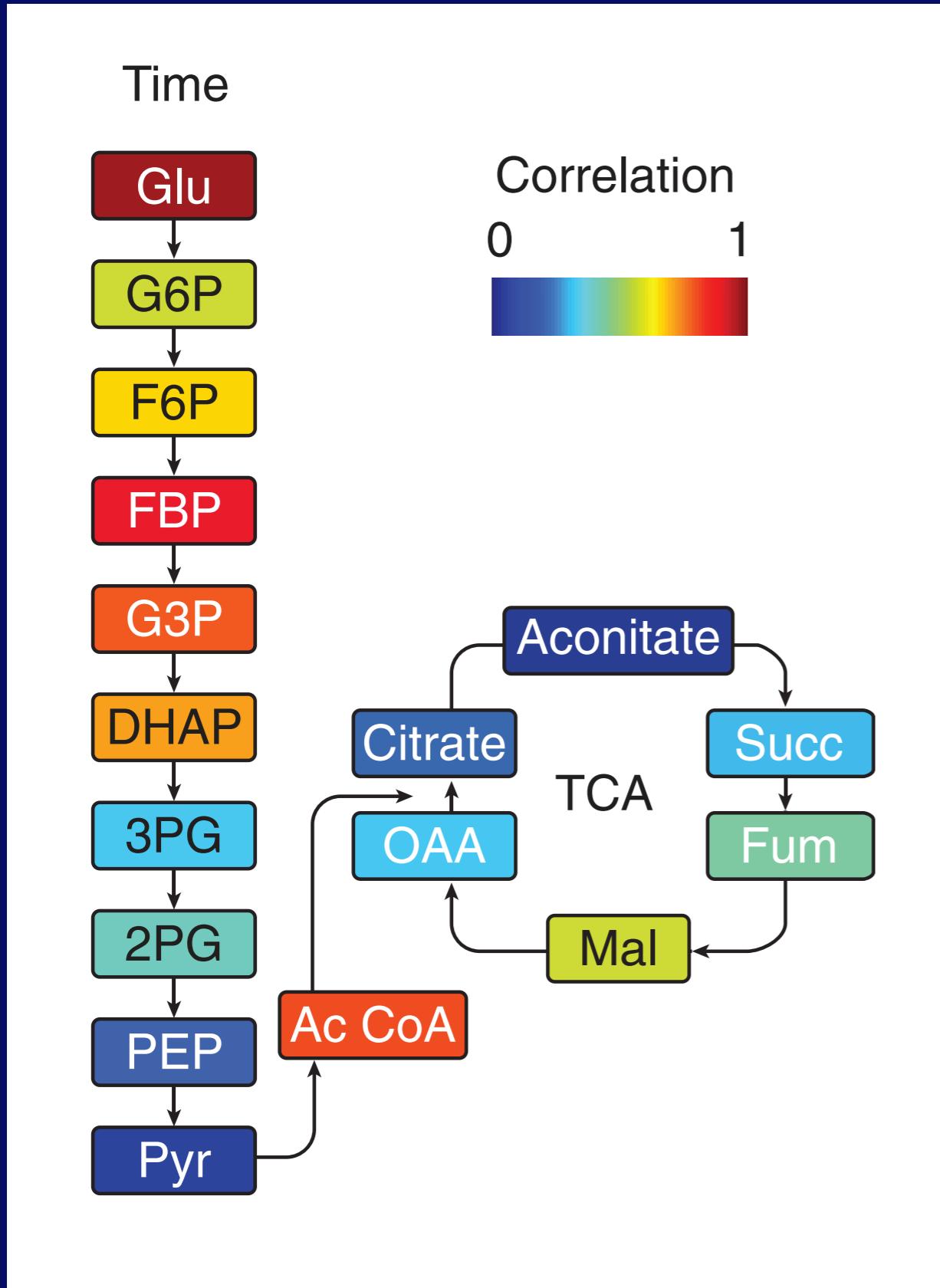
**Ingredients:**

1. X-range
2. Y-range
3. Empty space
4. Log-log space
5. Y-tick labels
6. Axis labels
7. Point colors
8. Point sizes
9. Filled/empty
10. Line styles
11. Line colors
12. Line thickness
13. Plotting orders
14. Font size
15. Font thickness

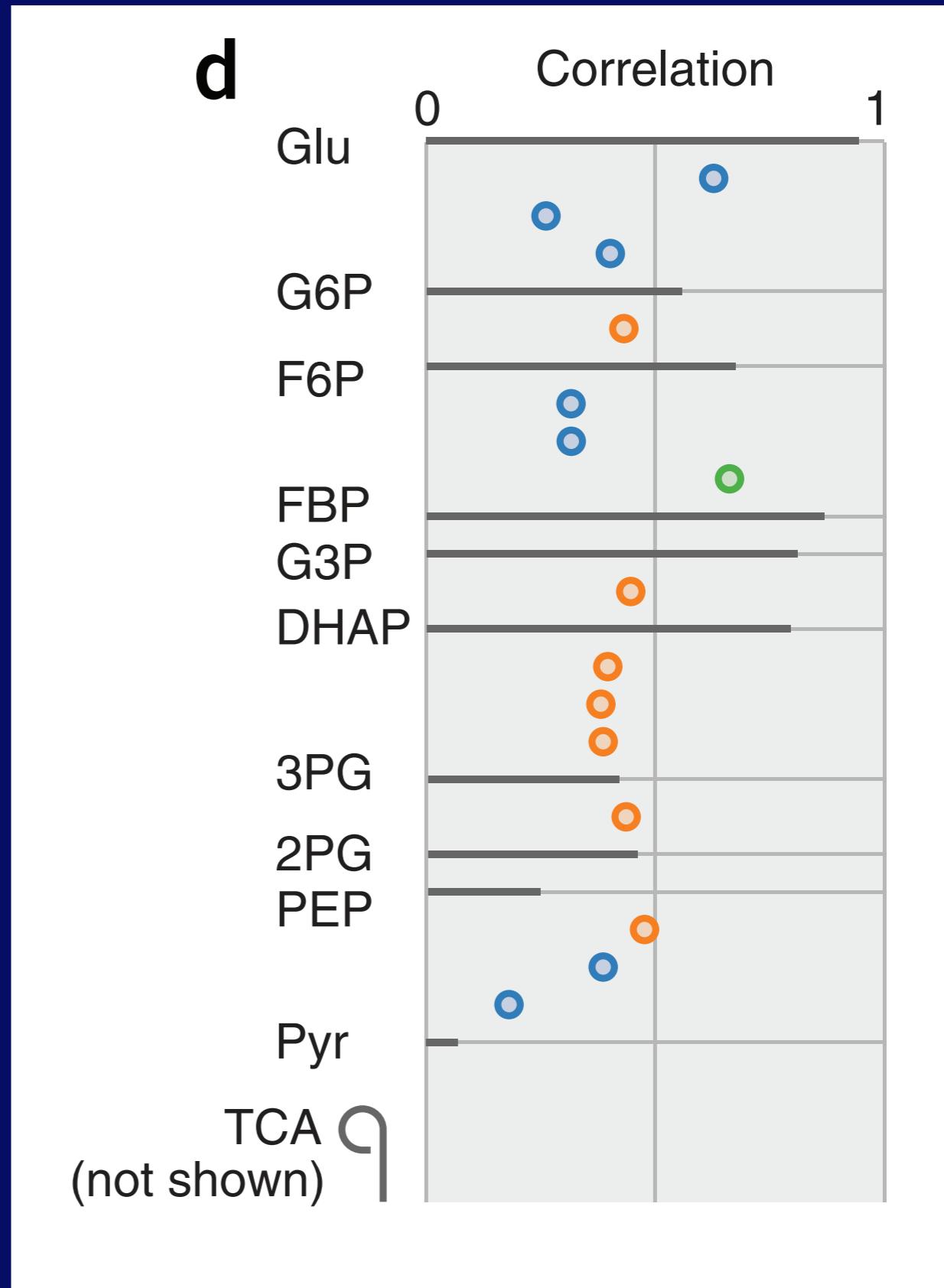
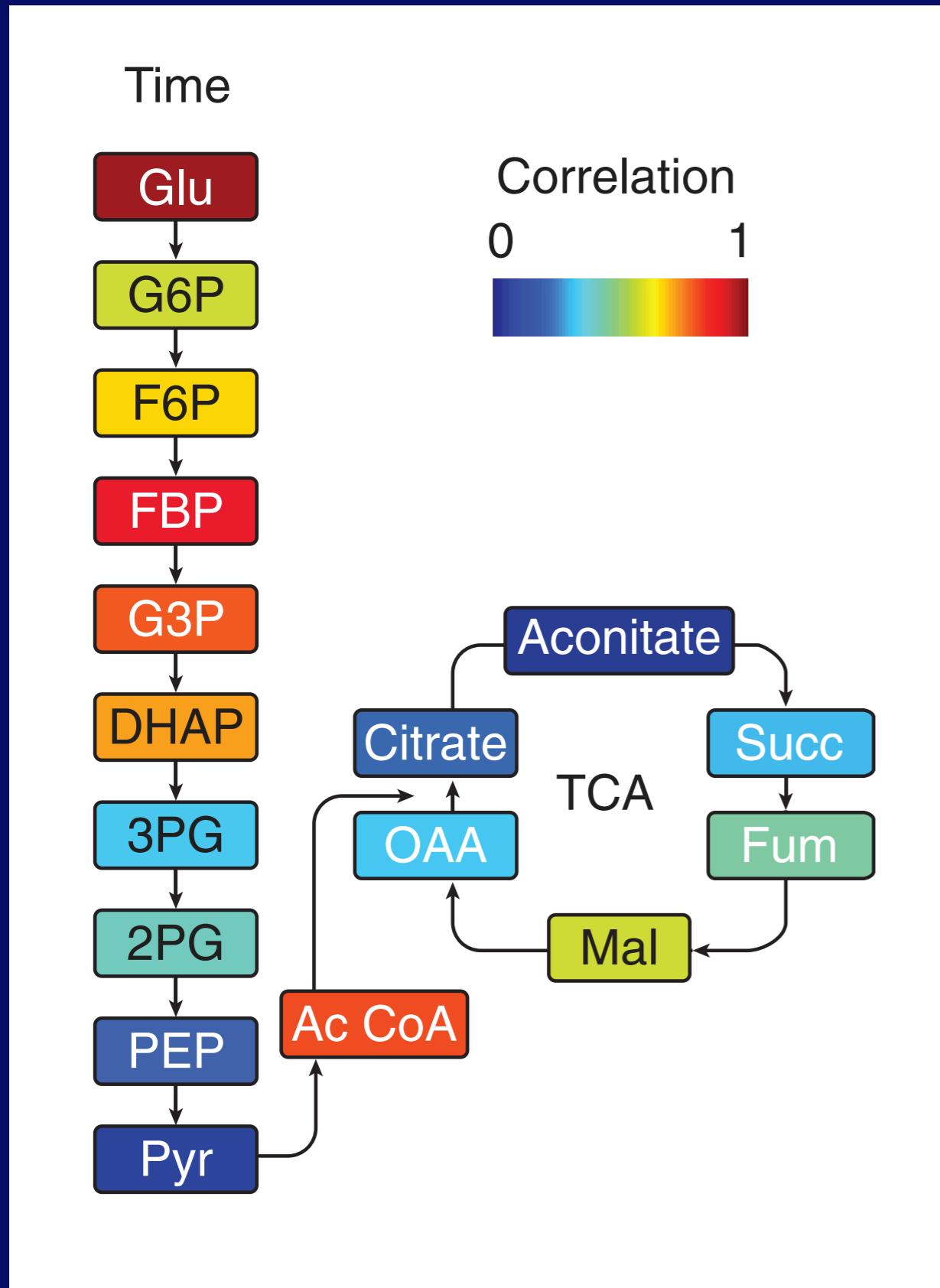
+ effective for  
classifying symbols into  
different categories

**– not great for  
continuous variables**

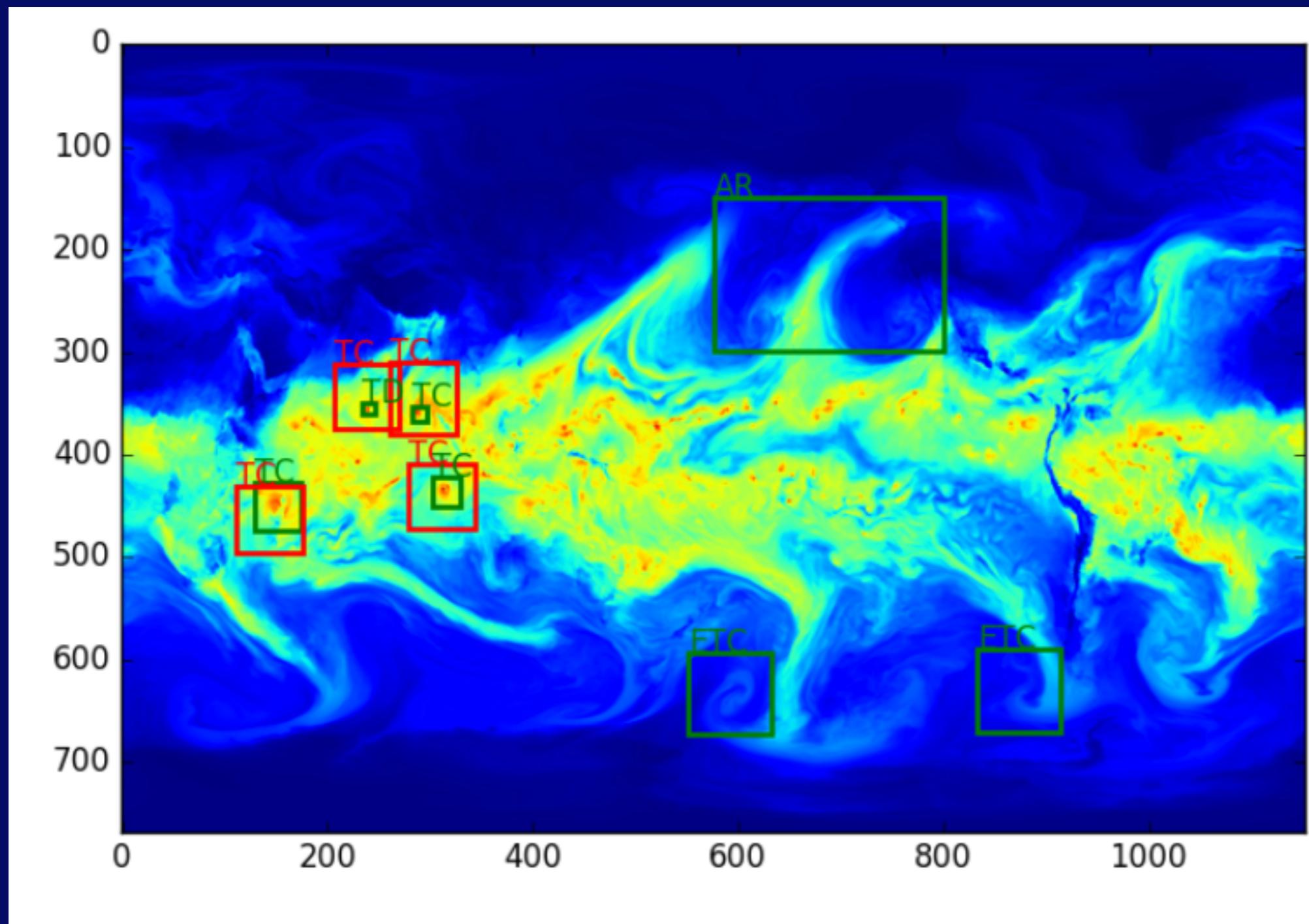
# How does it look in colour?



# How does it look in colour?



# Colour sequences for maps





- + approximates physical spectrum

- + approximates physical spectrum
- + lower errors in reading values from keys

- + approximates physical spectrum
- + lower errors in reading values from keys
- + conveys more colour steps

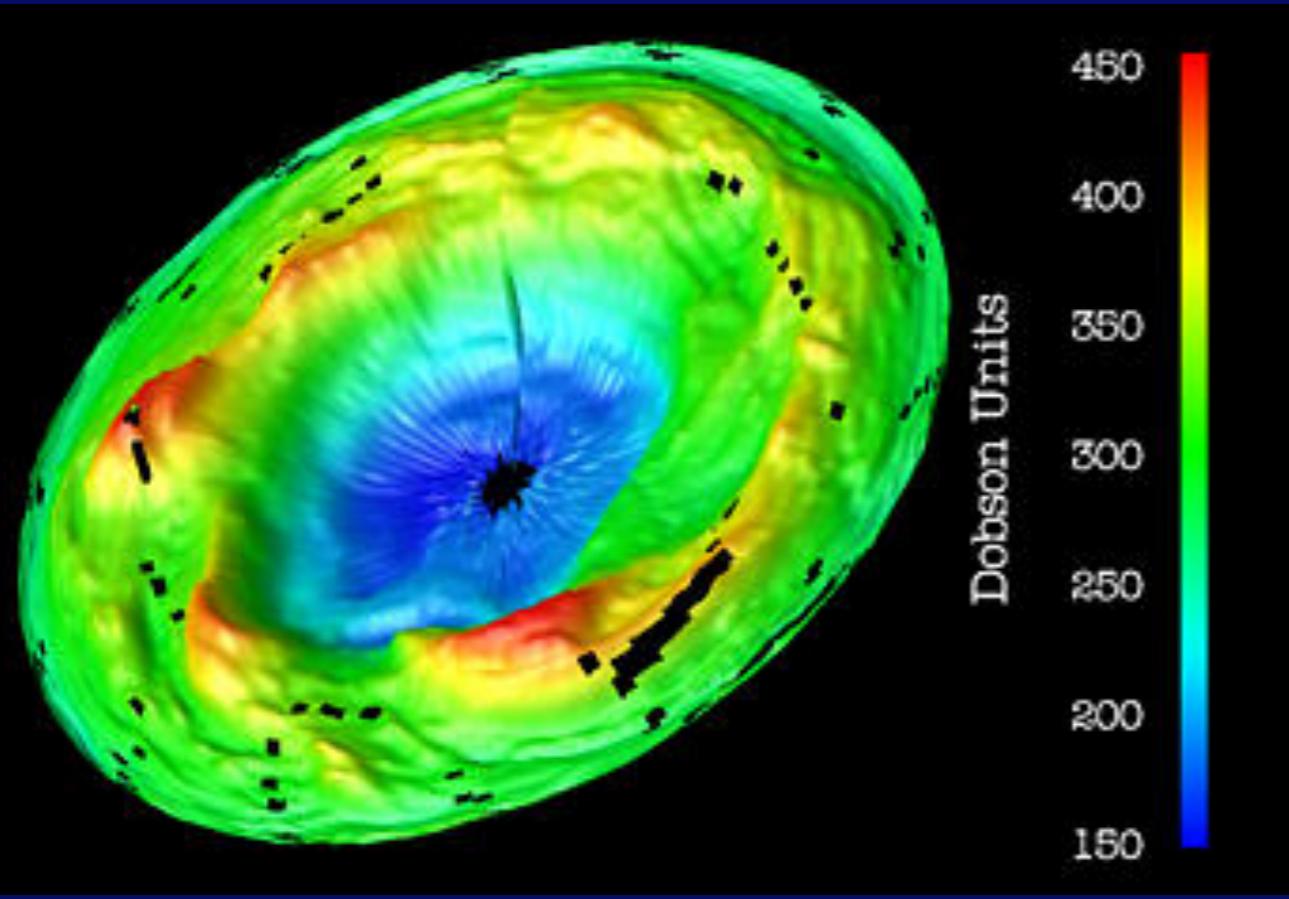
- + approximates physical spectrum
- + lower errors in reading values from keys
- + conveys more colour steps
- + semantic meaning

- + approximates physical spectrum
  - + lower errors in reading values from keys
  - + conveys more colour steps
  - + semantic meaning
- 
- not perceptually ordered

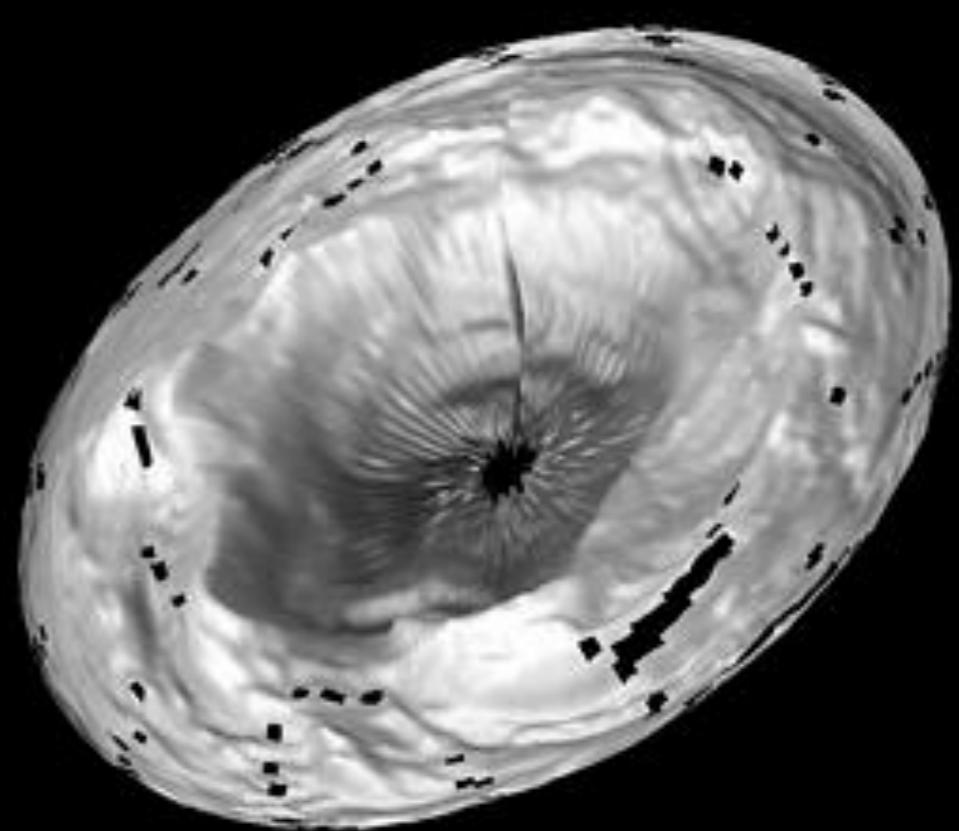
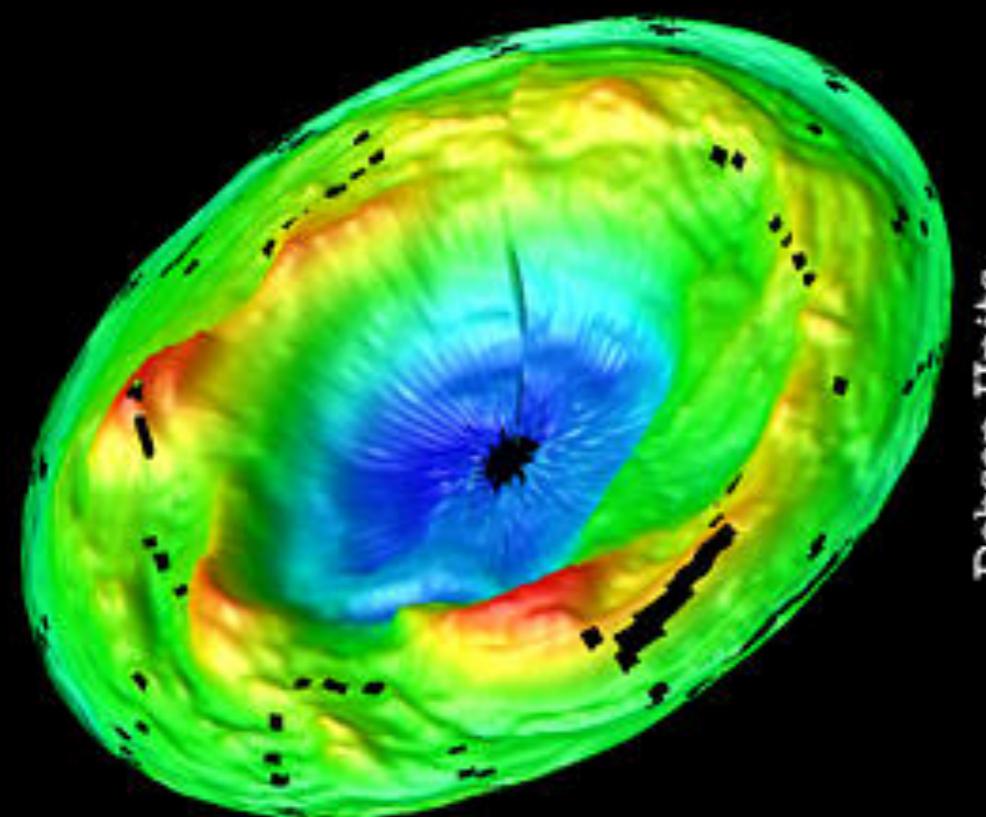
- + approximates physical spectrum
  - + lower errors in reading values from keys
  - + conveys more colour steps
  - + semantic meaning
- 
- not perceptually ordered
  - can be confusing and hinder perception tasks

- + approximates physical spectrum
  - + lower errors in reading values from keys
  - + conveys more colour steps
  - + semantic meaning
- 
- not perceptually ordered
  - can be confusing and hinder perception tasks
  - hard to understand in greyscale

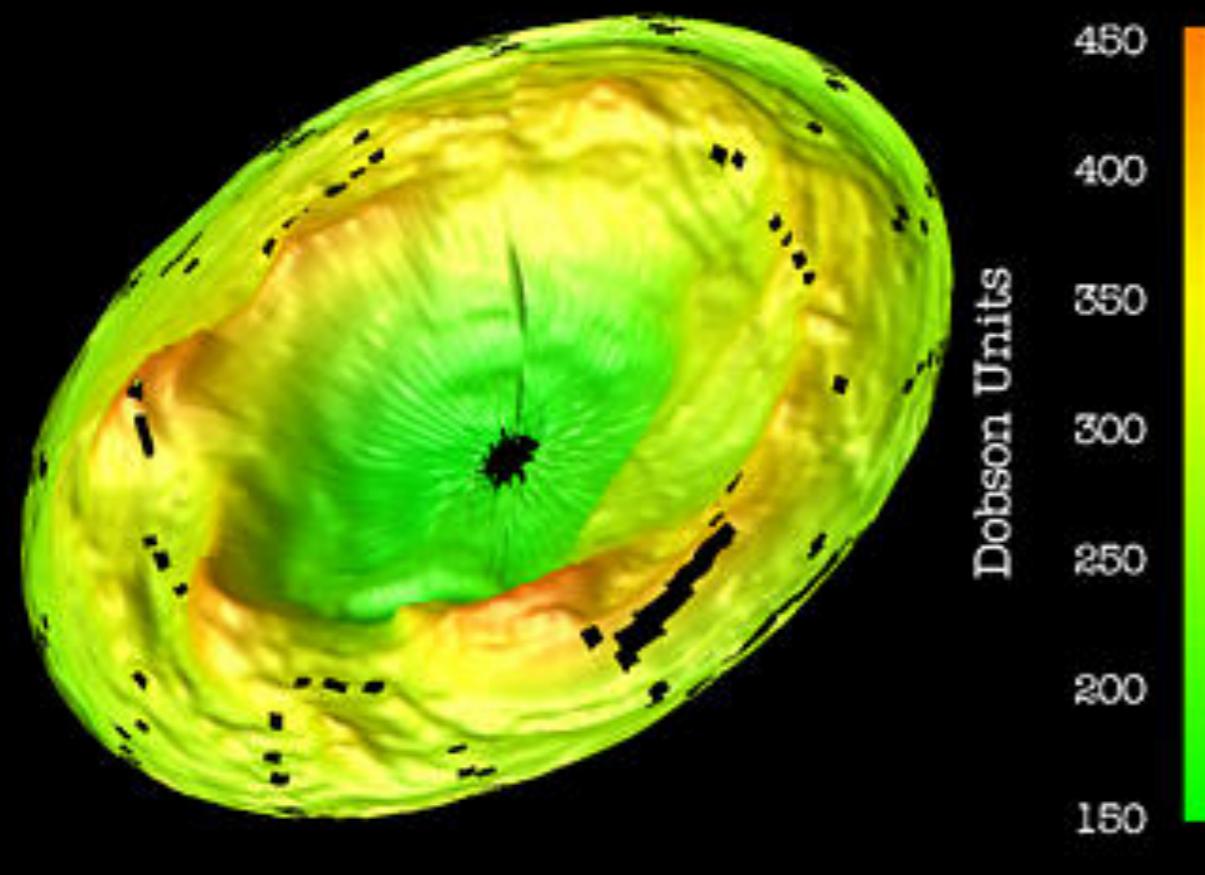
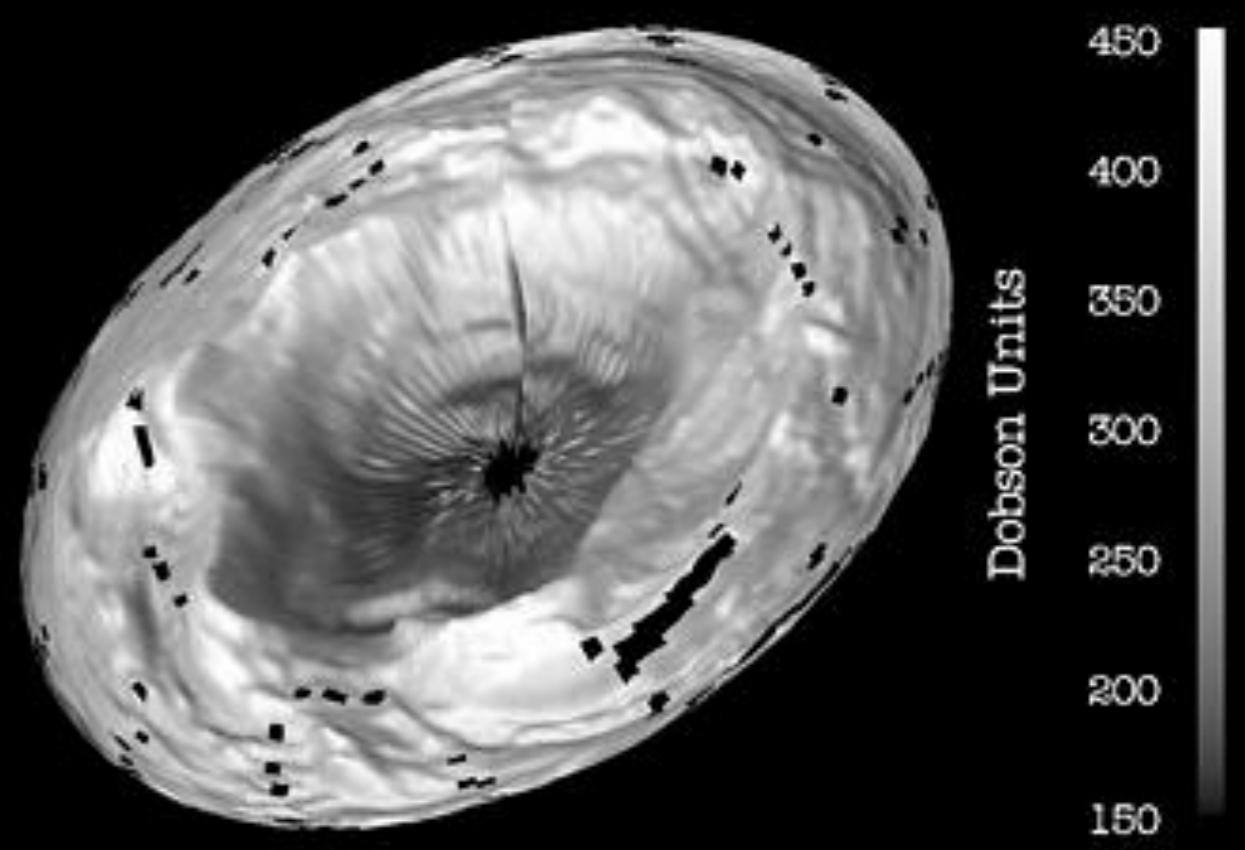
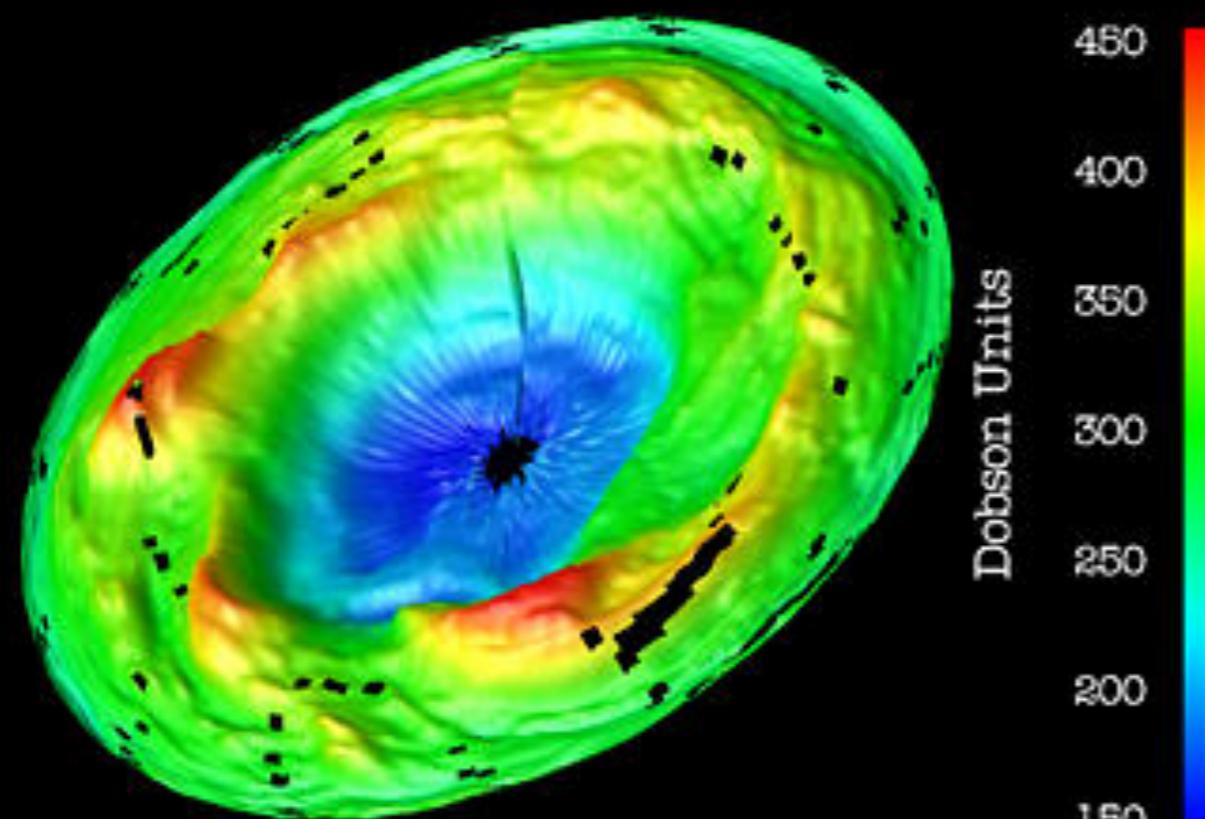
**Hue is not a good dimension for  
encoding magnitude data!**



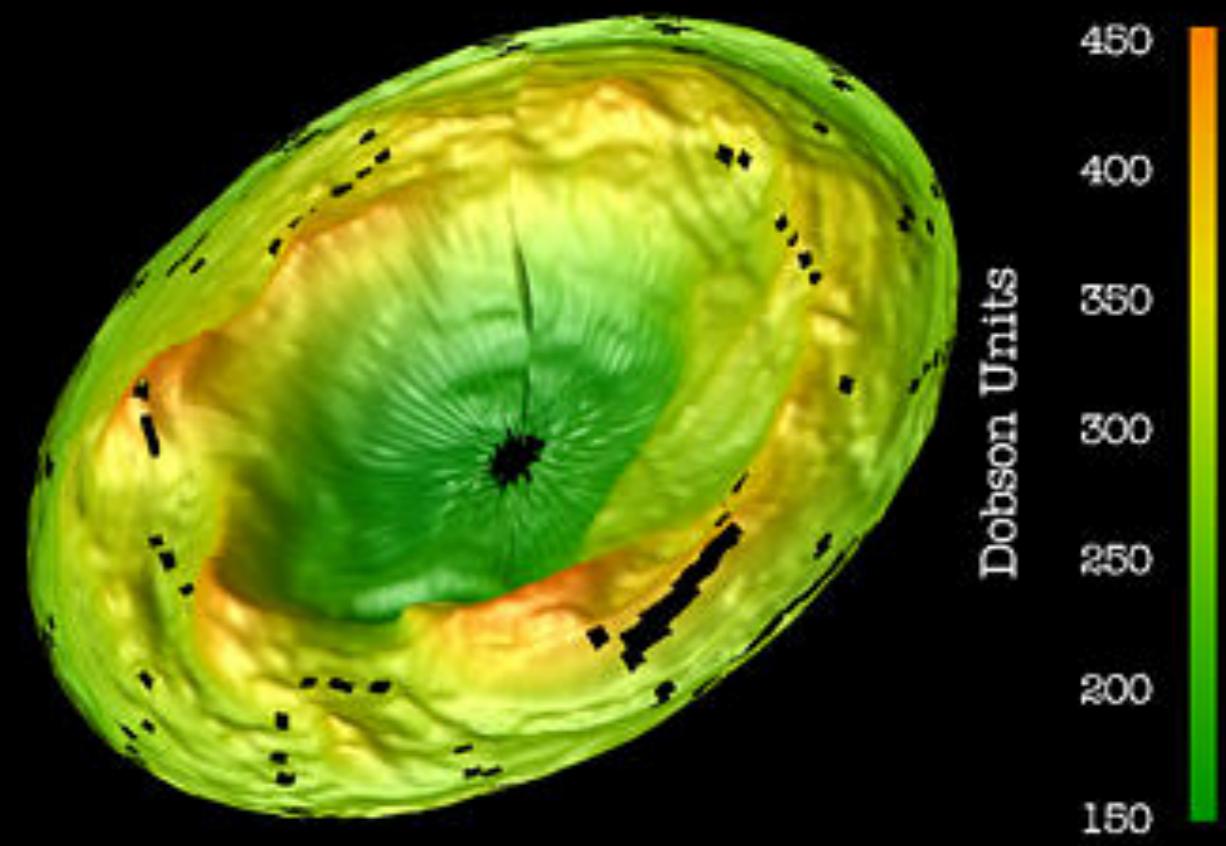
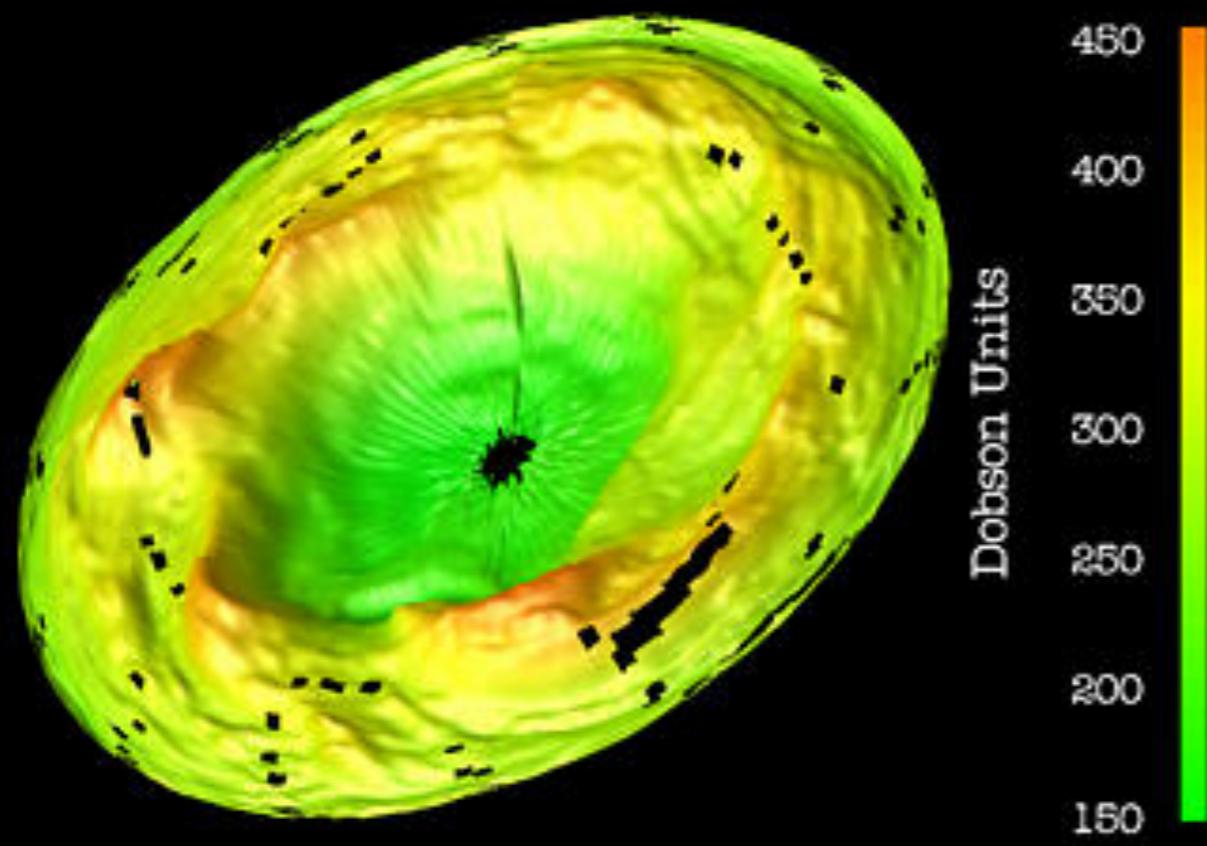
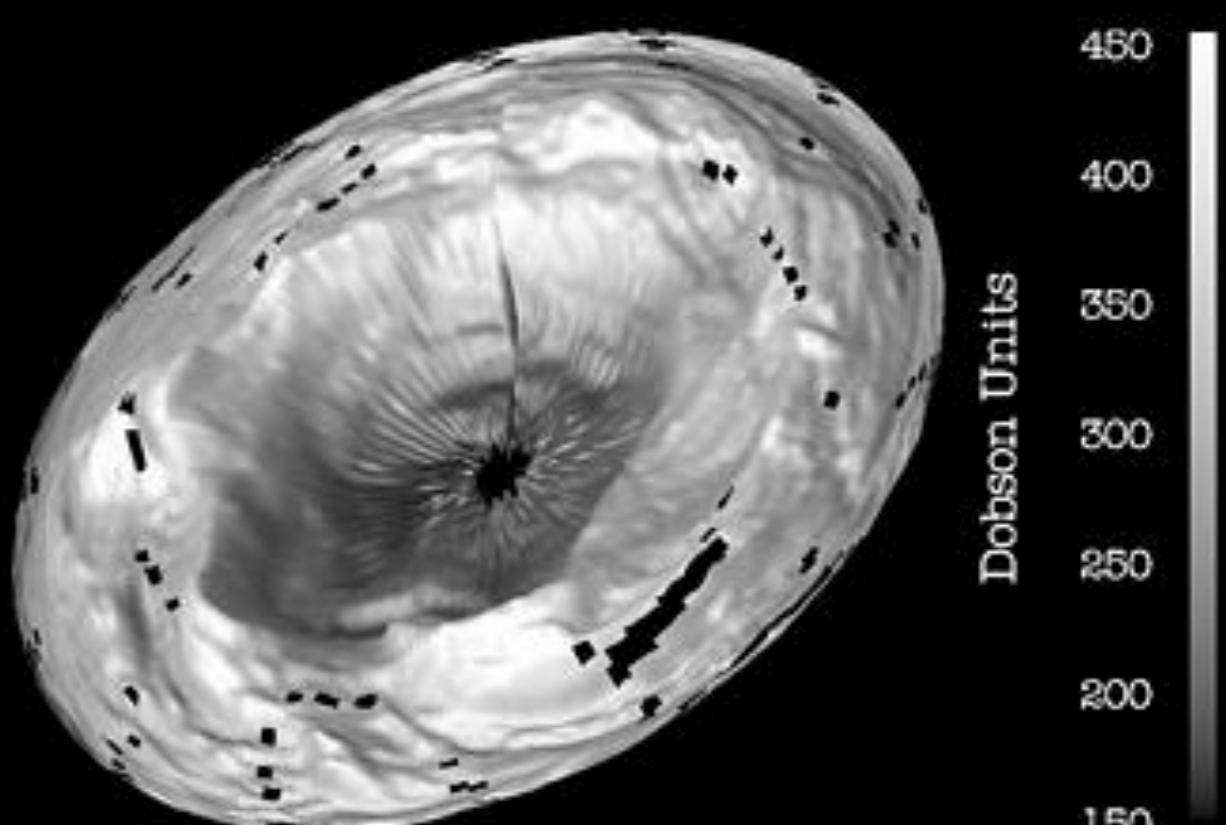
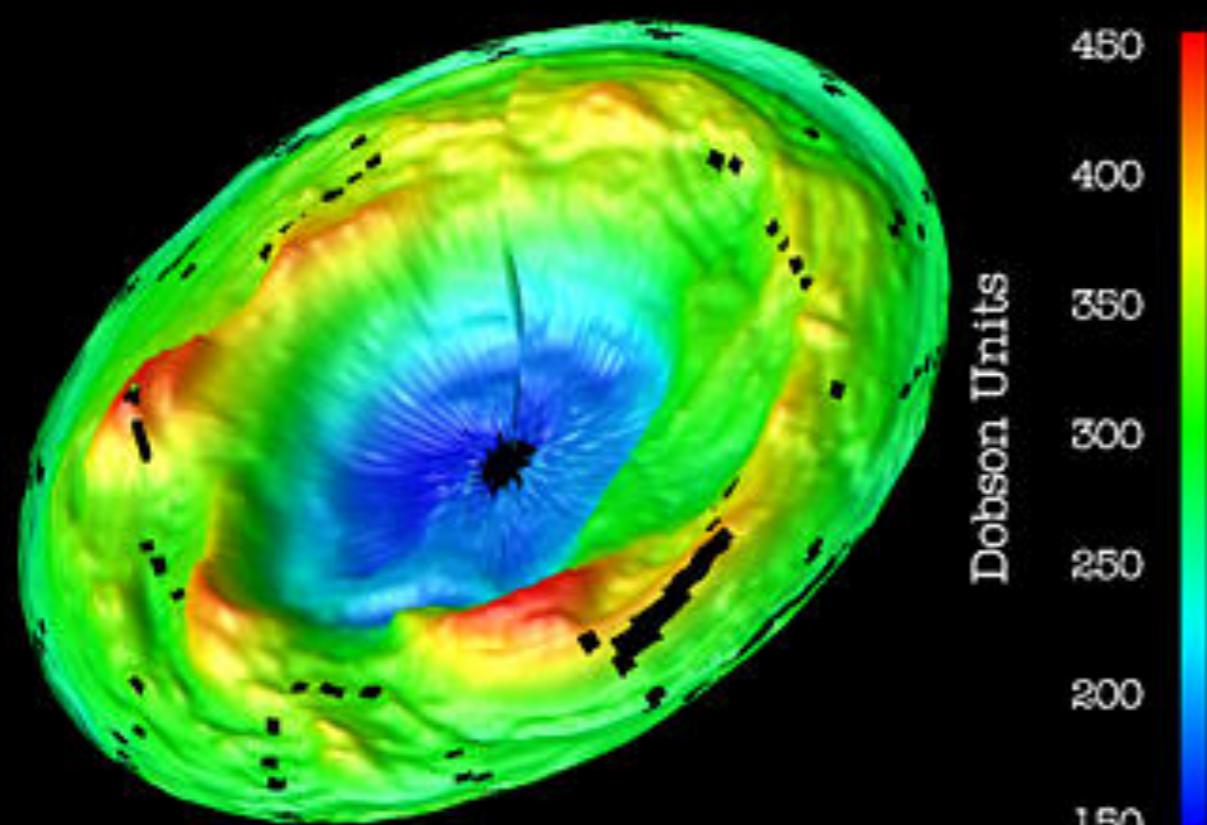
3d dimension for  
encoding magnitude data!

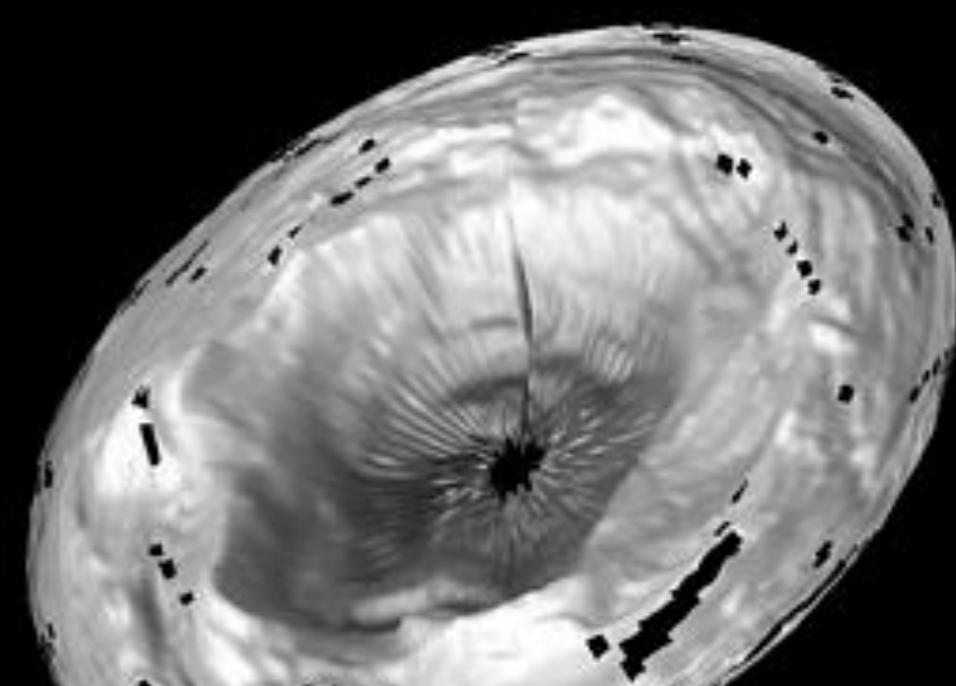
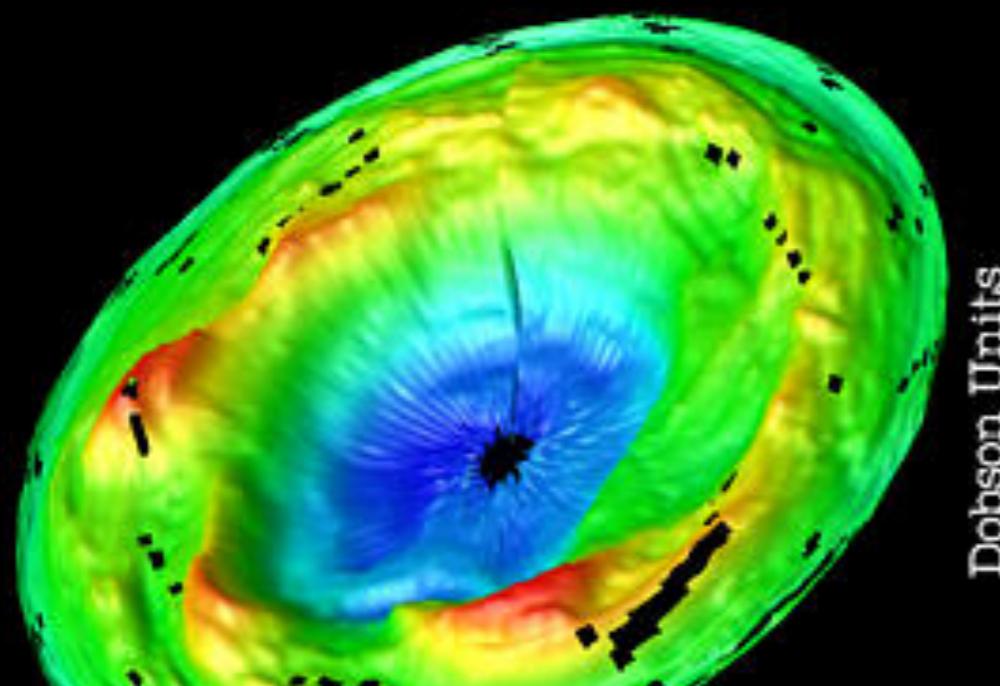


encoding magnitude data!

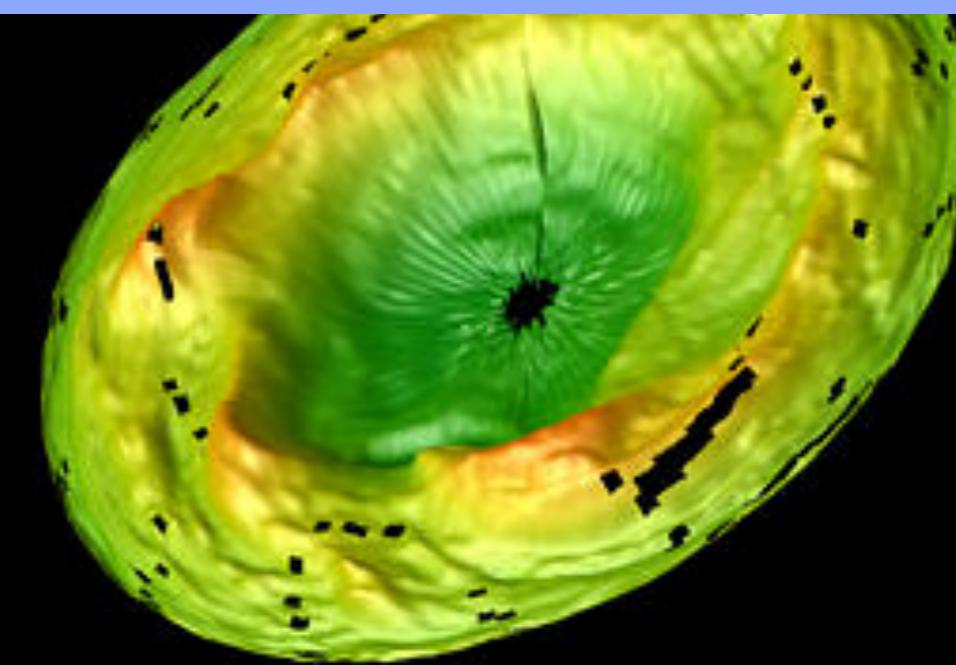
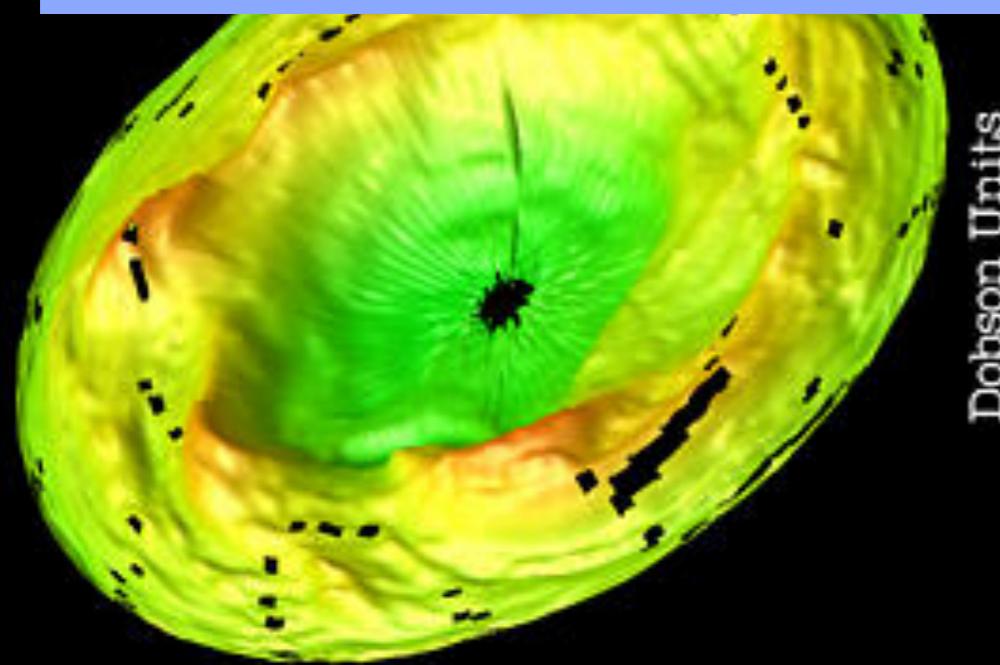


Latitude data!

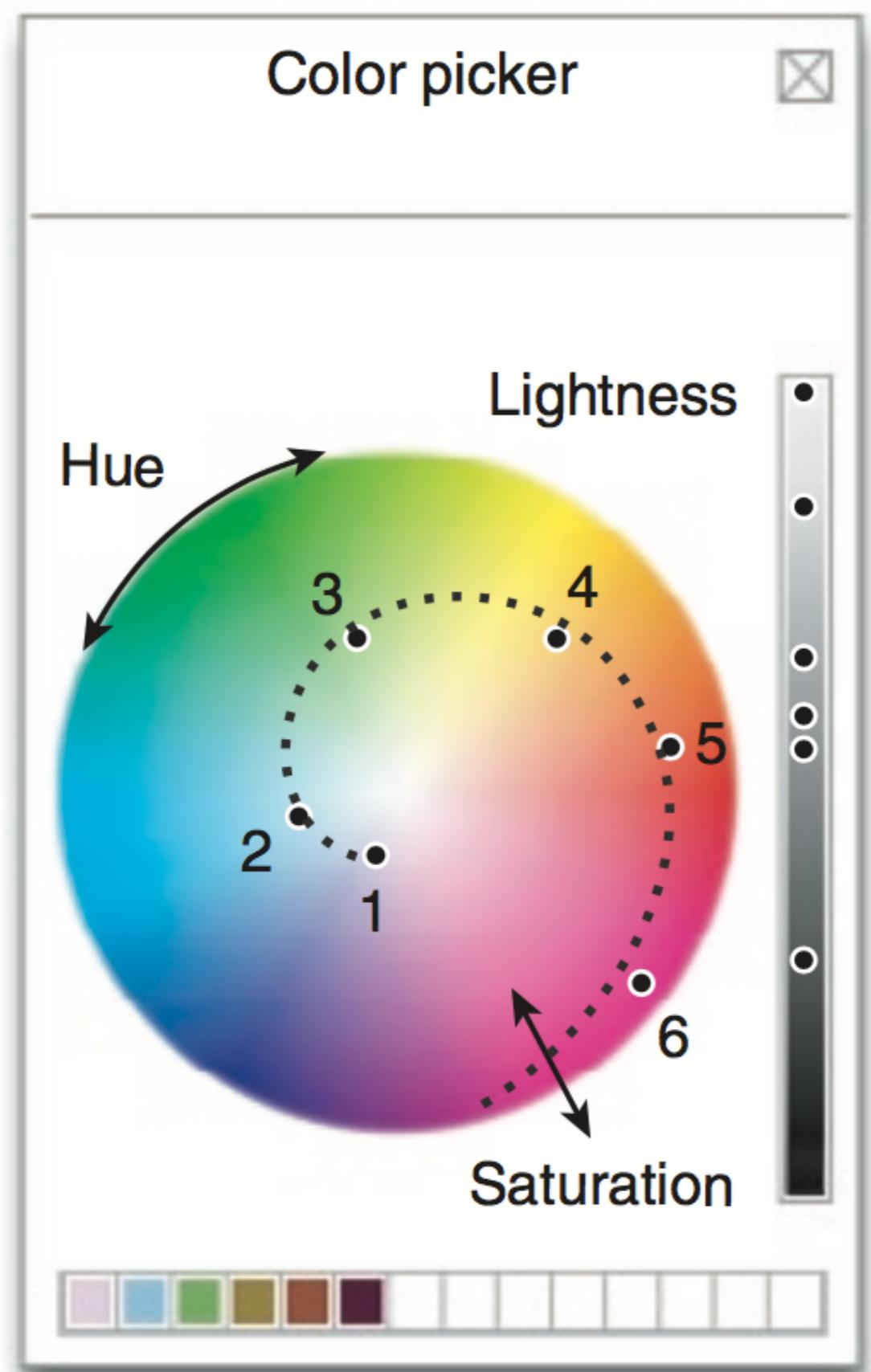
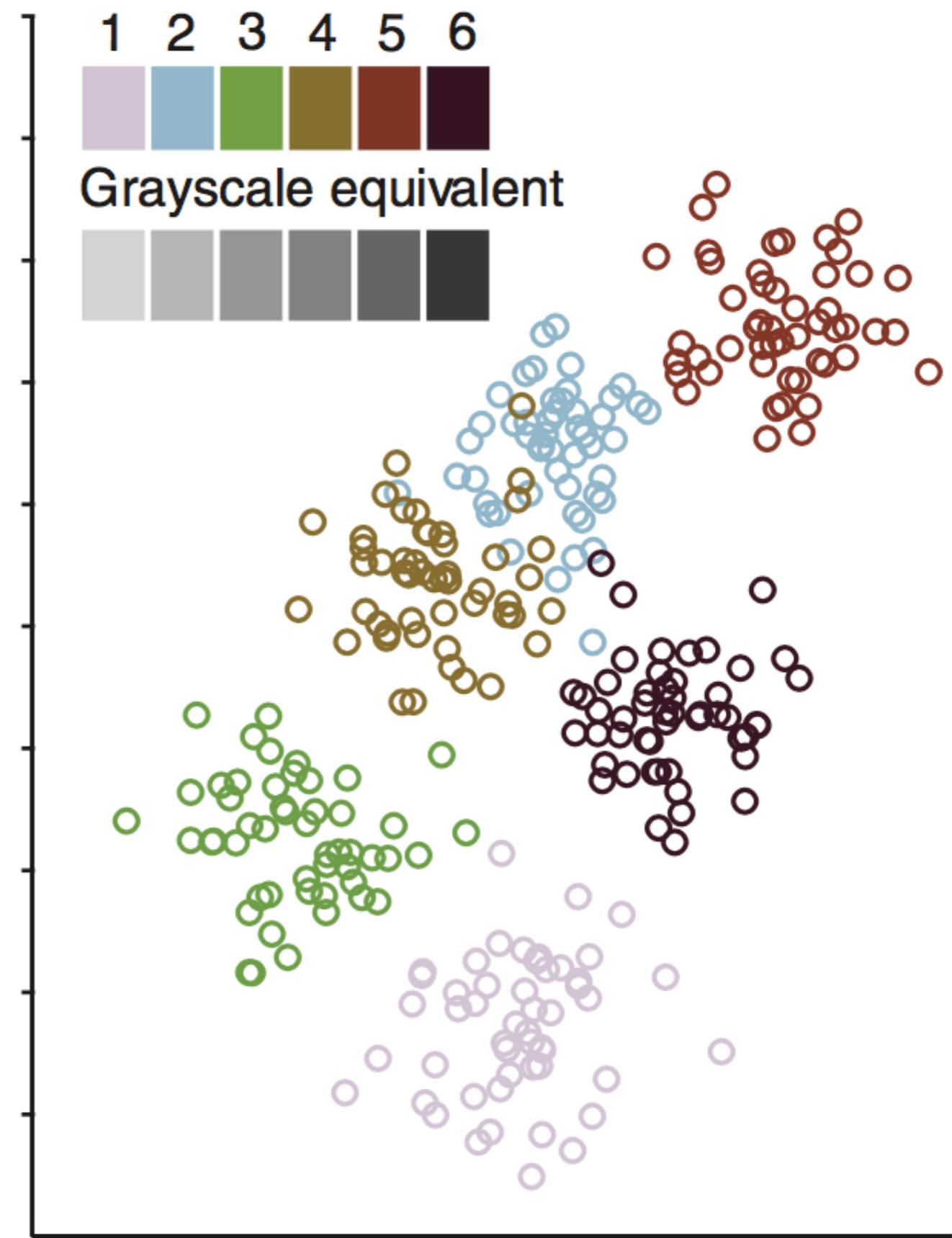




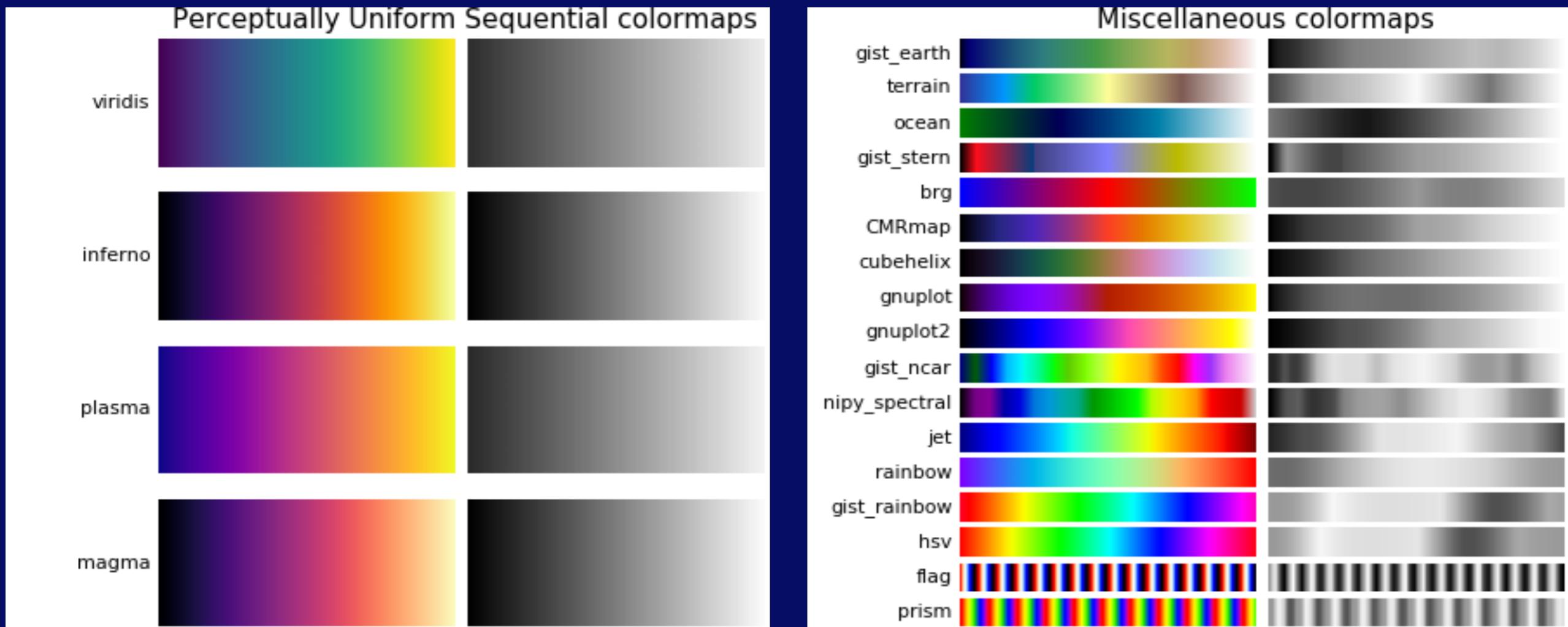
**luminance → high spatial frequency data  
saturation → low spatial frequency data**



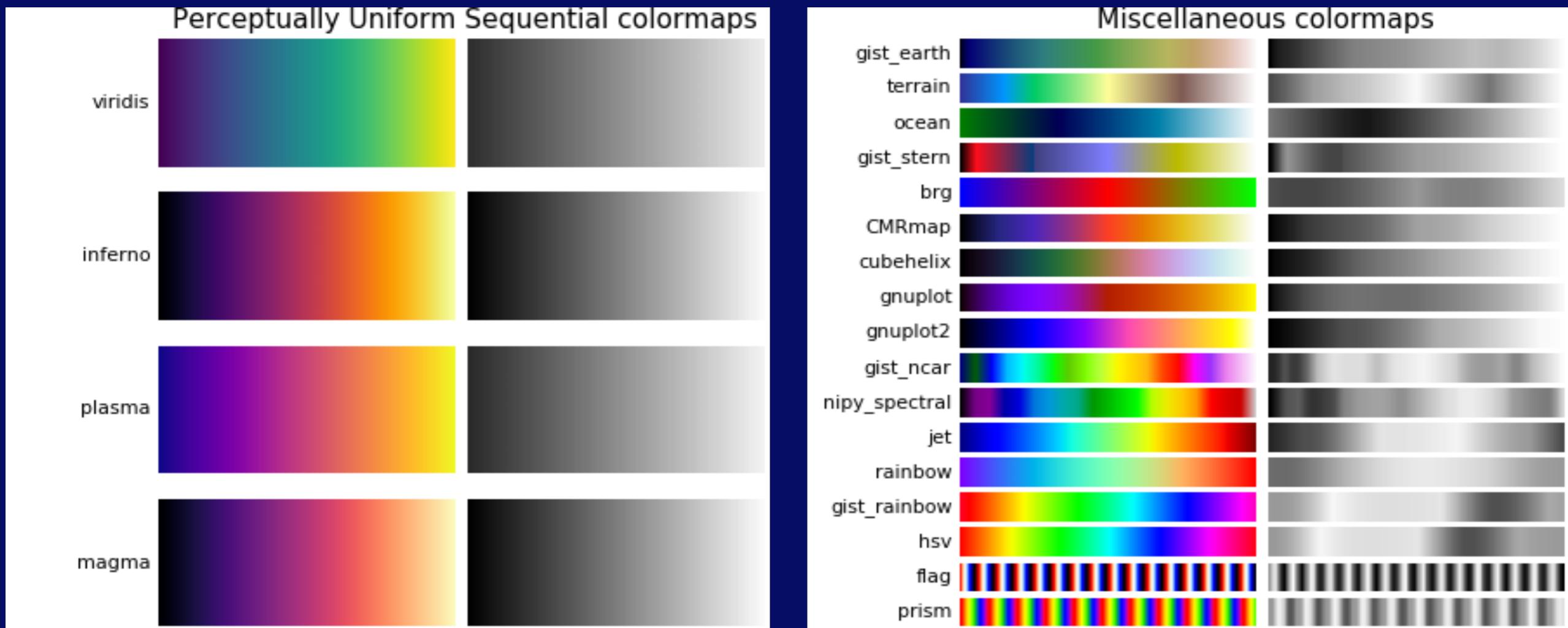
+ use luminance  
contrasts within colours

**a****b**

# How does it look in grayscale?

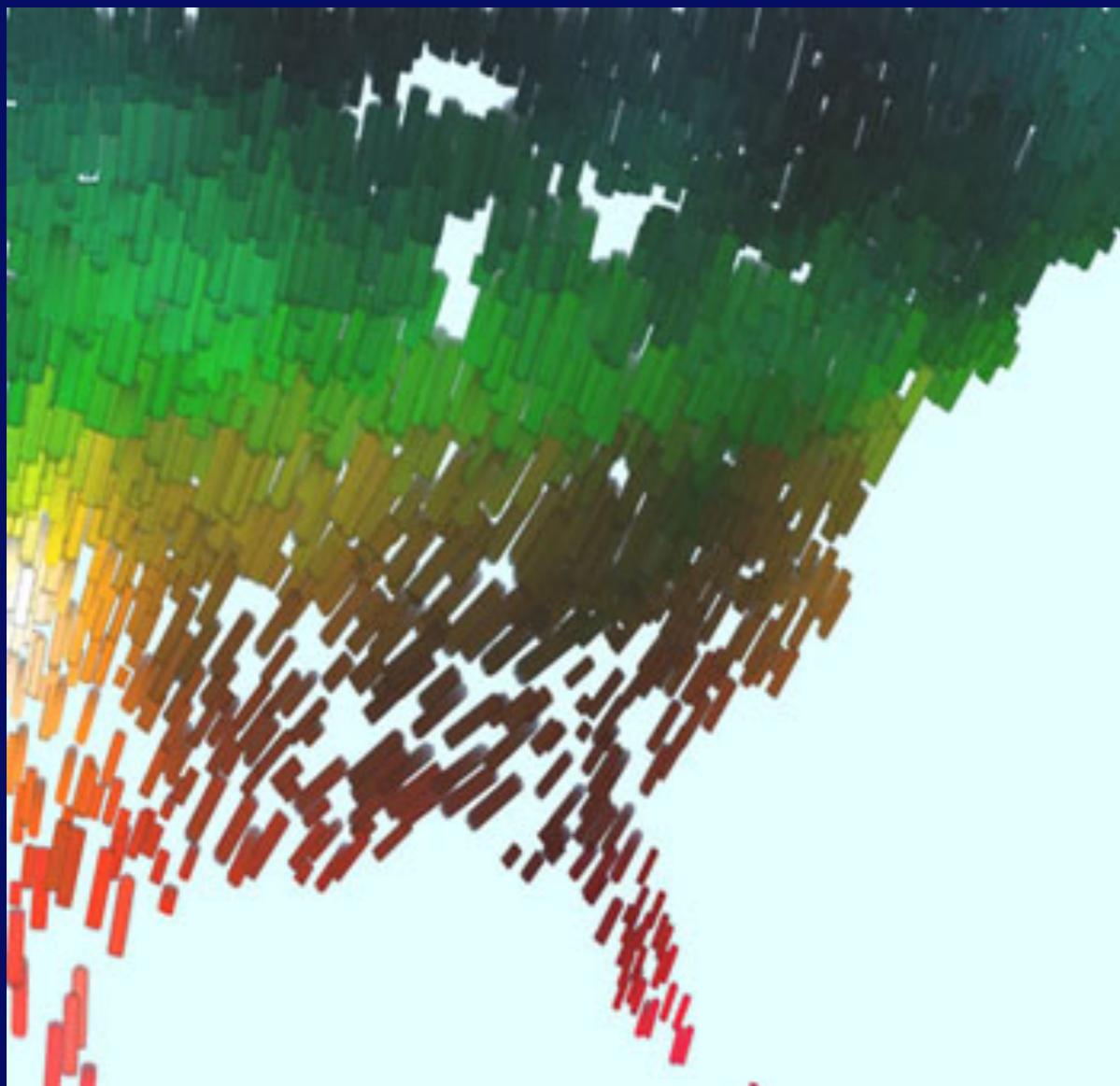


# How does it look in grayscale?

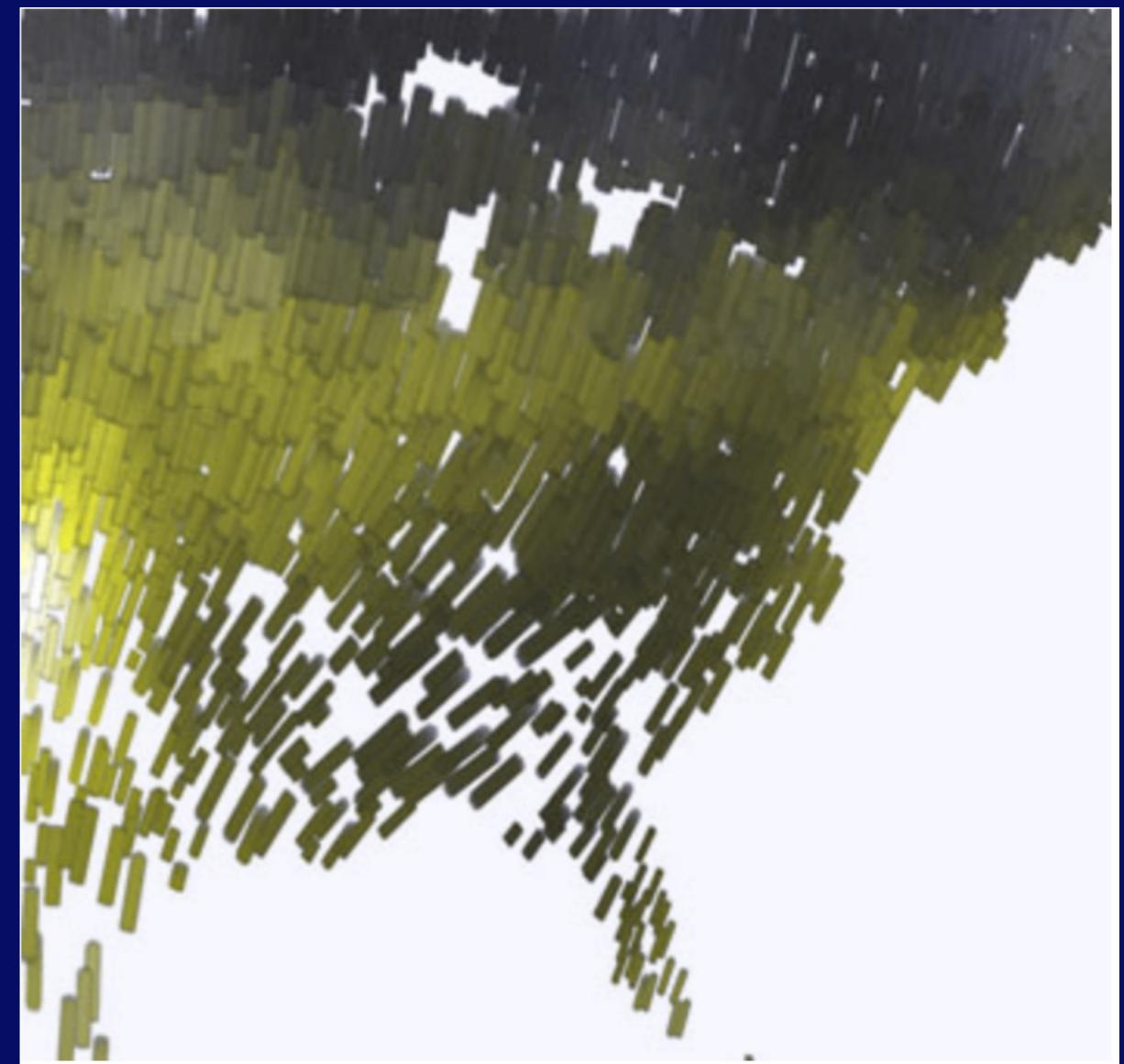
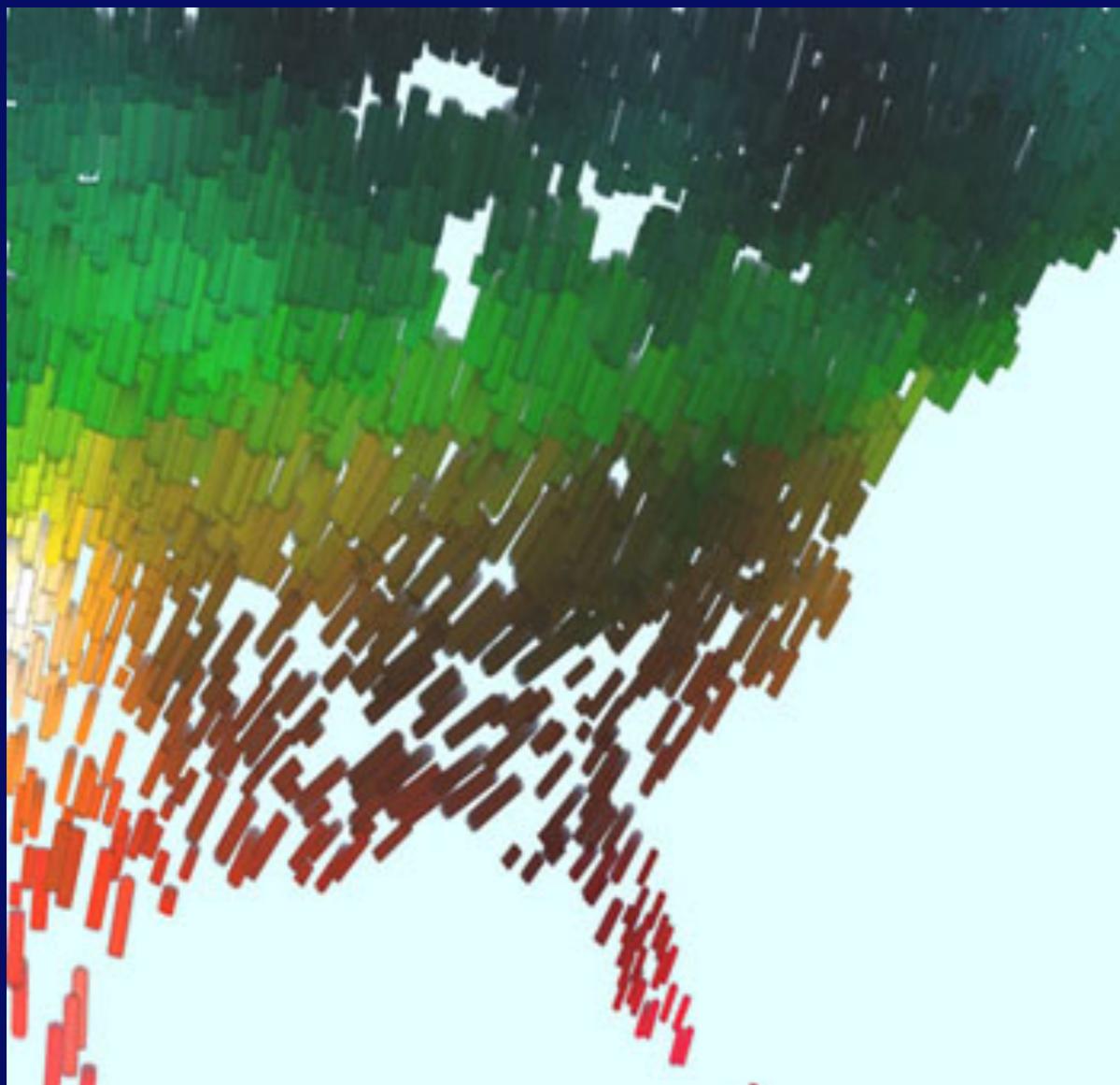


<http://matplotlib.org/users/colormaps.html>

# How does it look for colour blind viewers?



# How does it look for colour blind viewers?



# How does it look for colour blind viewers?



<https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463>

# How to colour-blind-proof your figure:

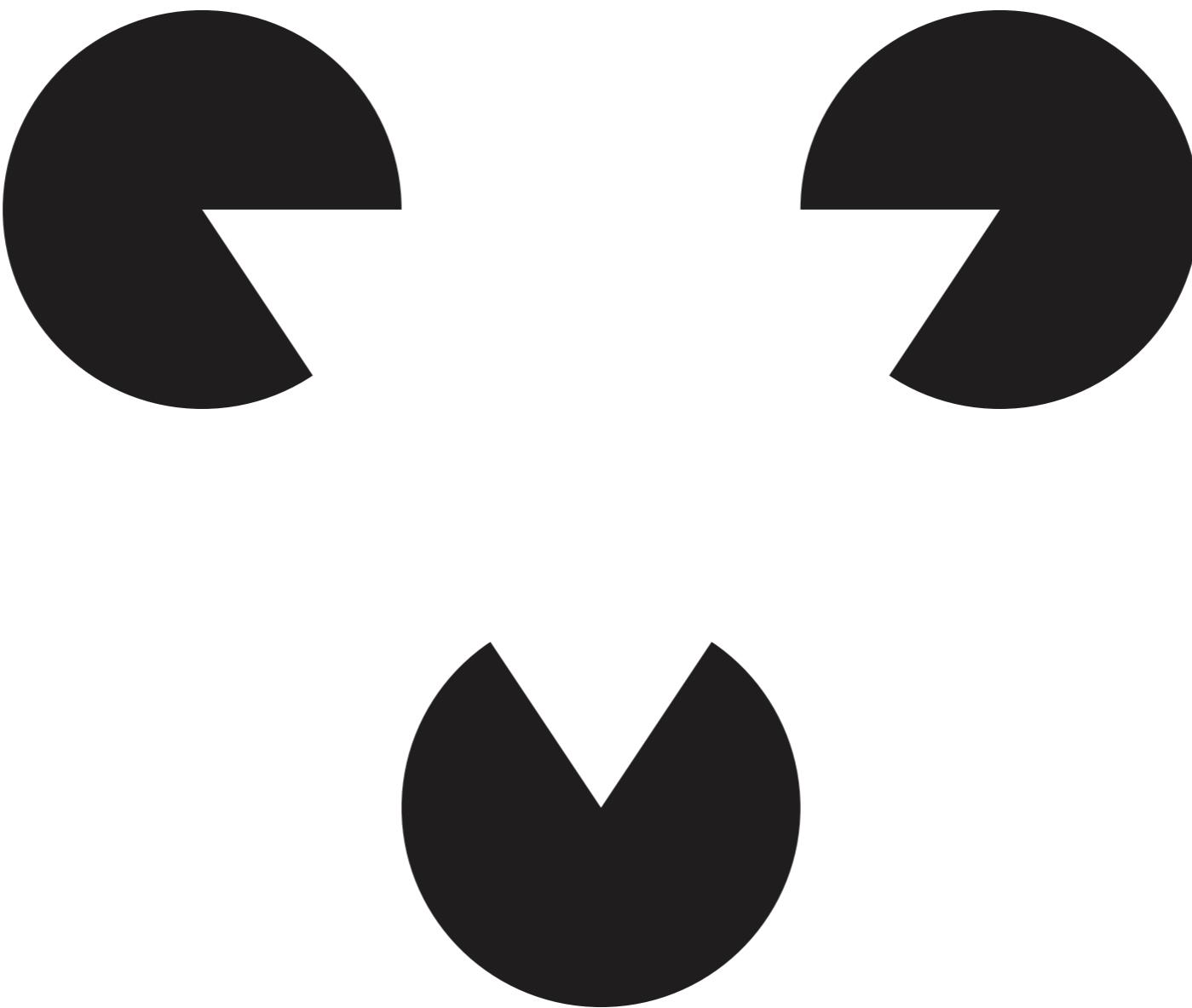
- **matplotlib: e.g. viridis**
- **use luminance to distinguish**
- **use shape to distinguish**
- **check using simulators, plug-ins etc.**

## Colour Resources:

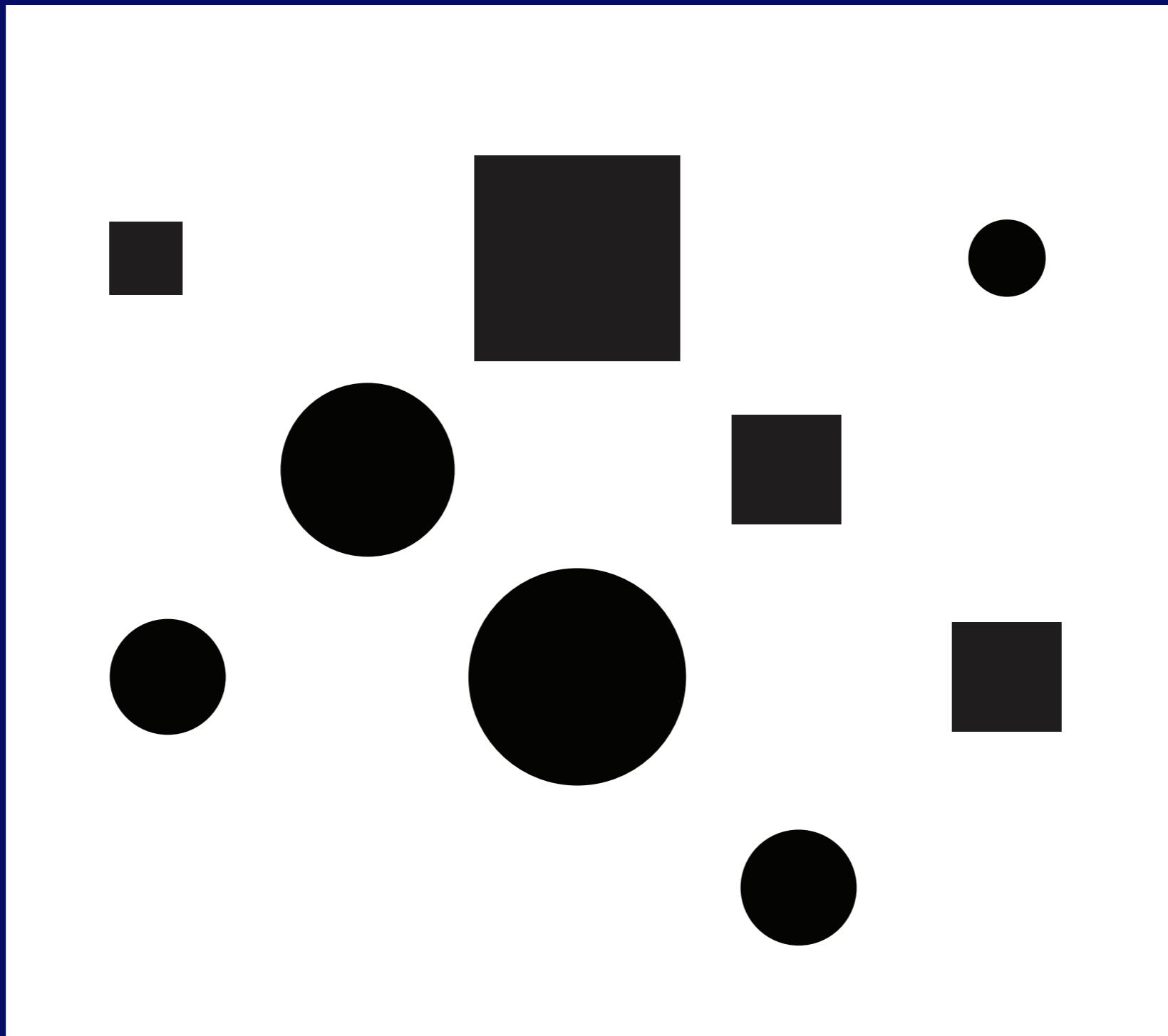
- <http://www.research.ibm.com/people/l/lloydt/color/color.HTM>
- <https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463>
- <http://matplotlib.org/users/colormaps.html>
- <http://www.color-blindness.com/coblis-color-blindness-simulator/>
- <http://www.vischeck.com>
- <http://colororacle.org>
- <http://colorbrewer2.org>
- <http://www.palettable.io>

# visual hierarchy

# Gestalt Principles

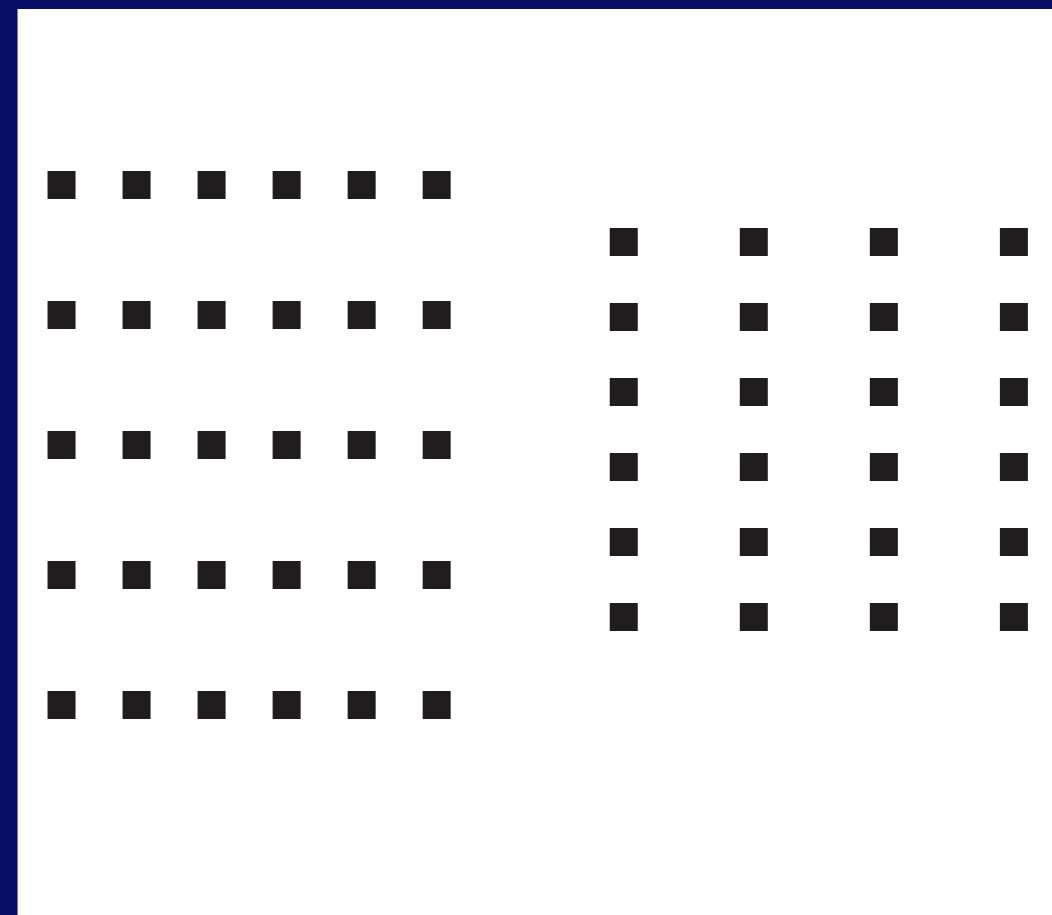
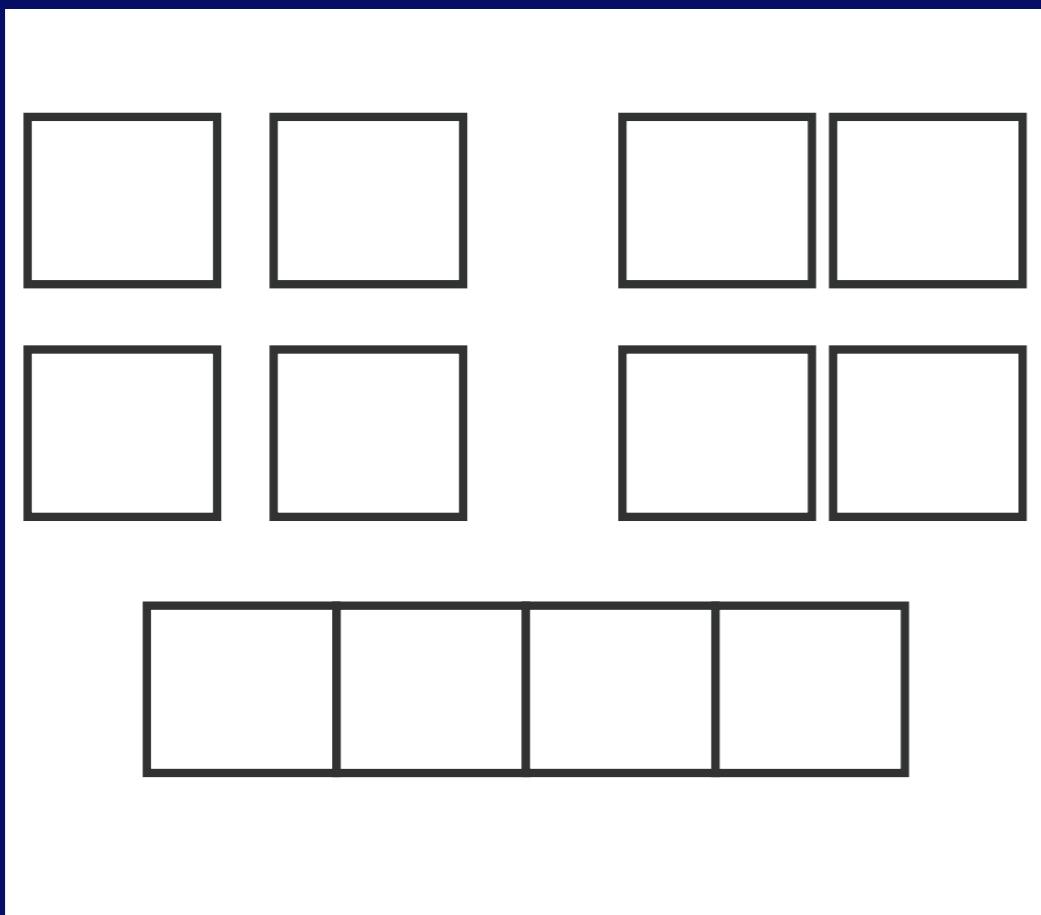


# Similarity

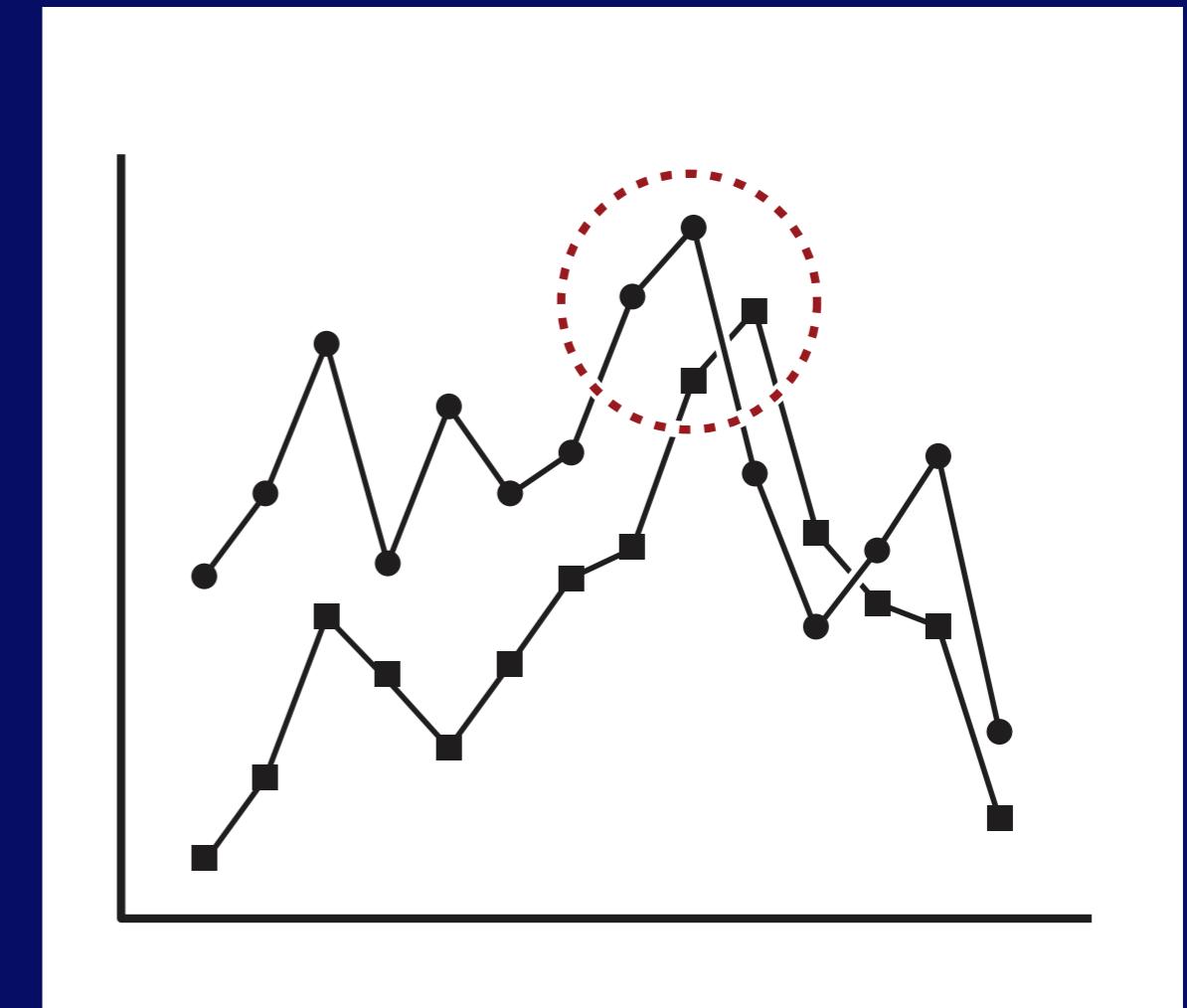
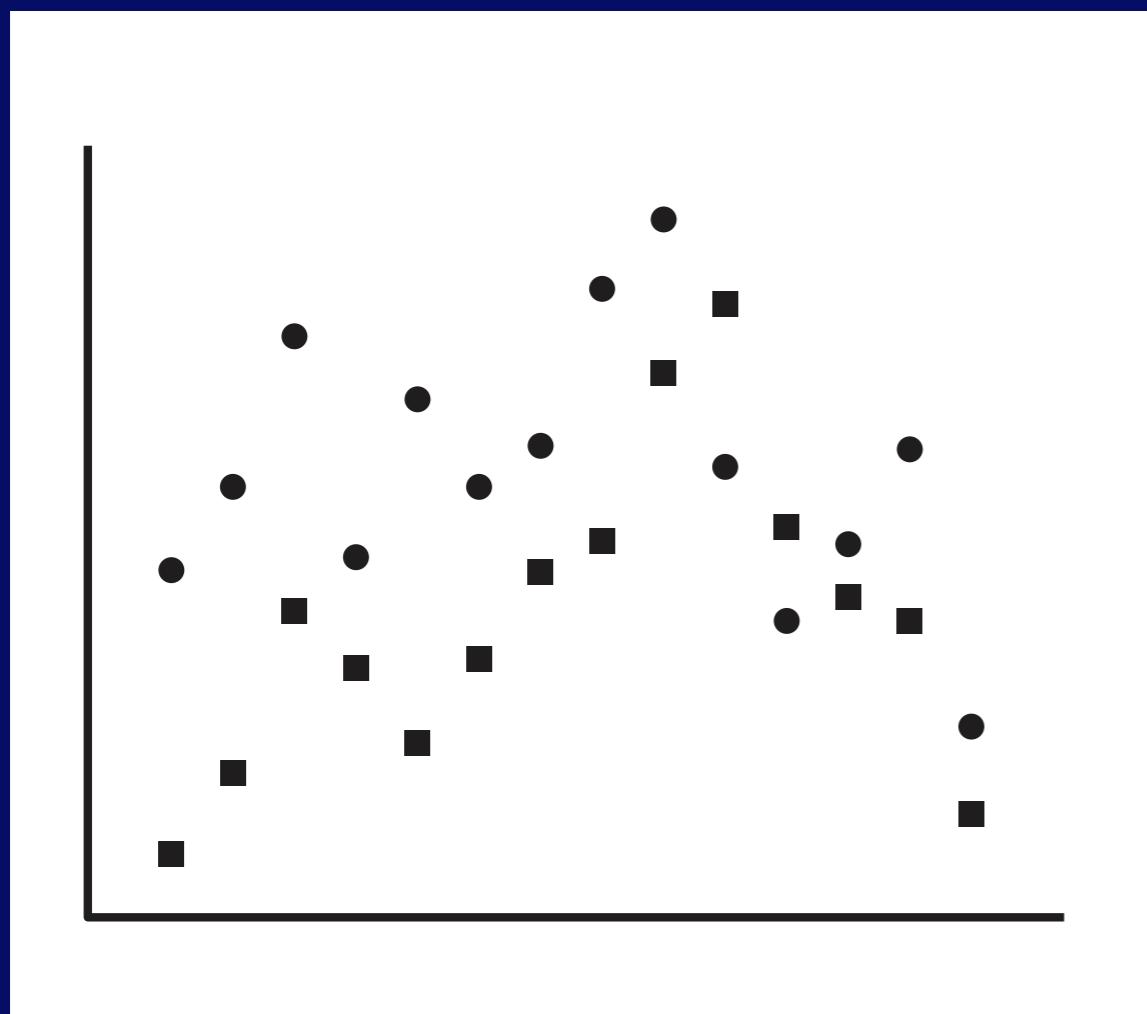


Wong, Nature Methods, 2010

# Proximity



# Connection + Enclosure



# Types of Data:

- ordinal
- categorical
- continuous
- ...

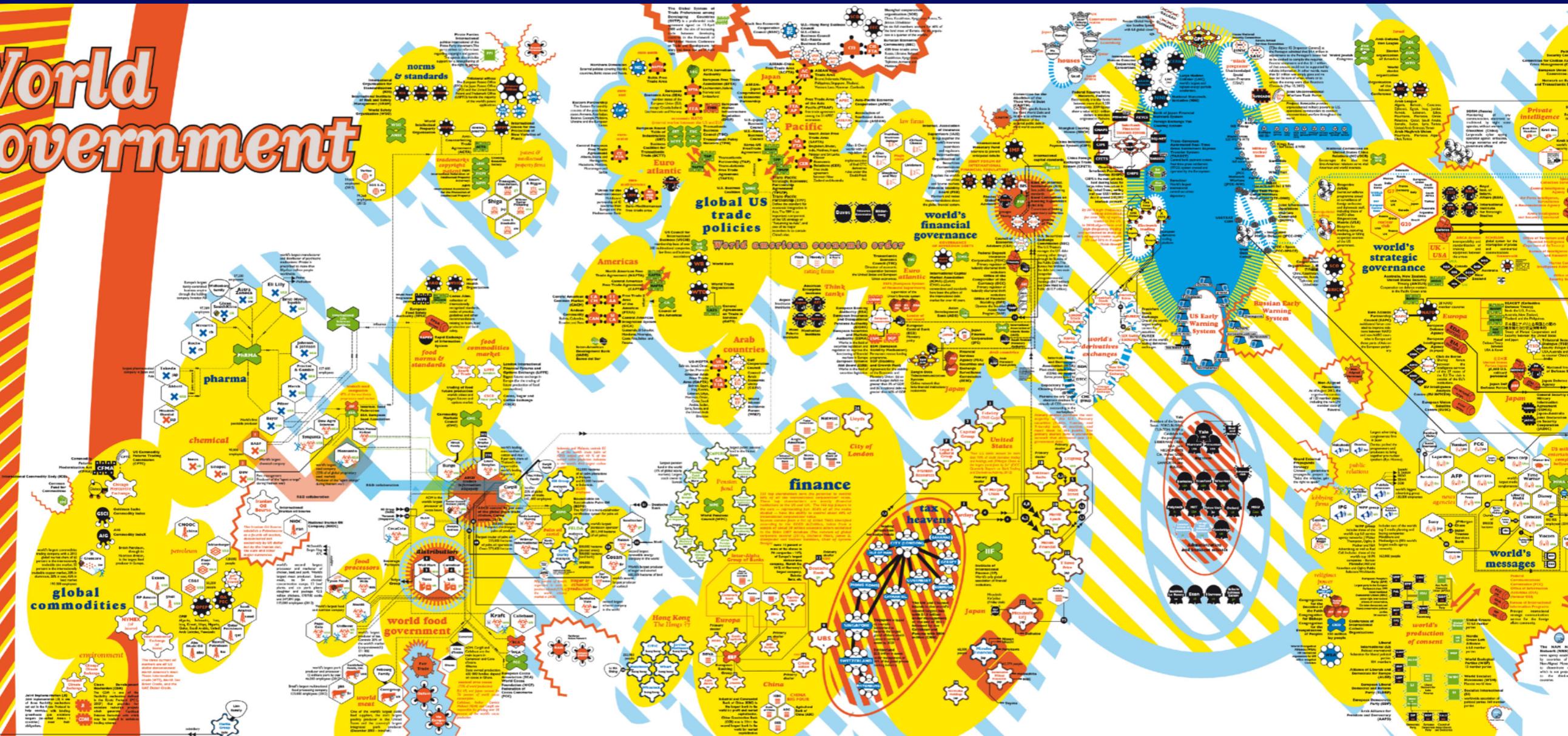
# Types of Relationships:

- proportions
- relationships,
- hierarchy
- distribution
- data over time ...

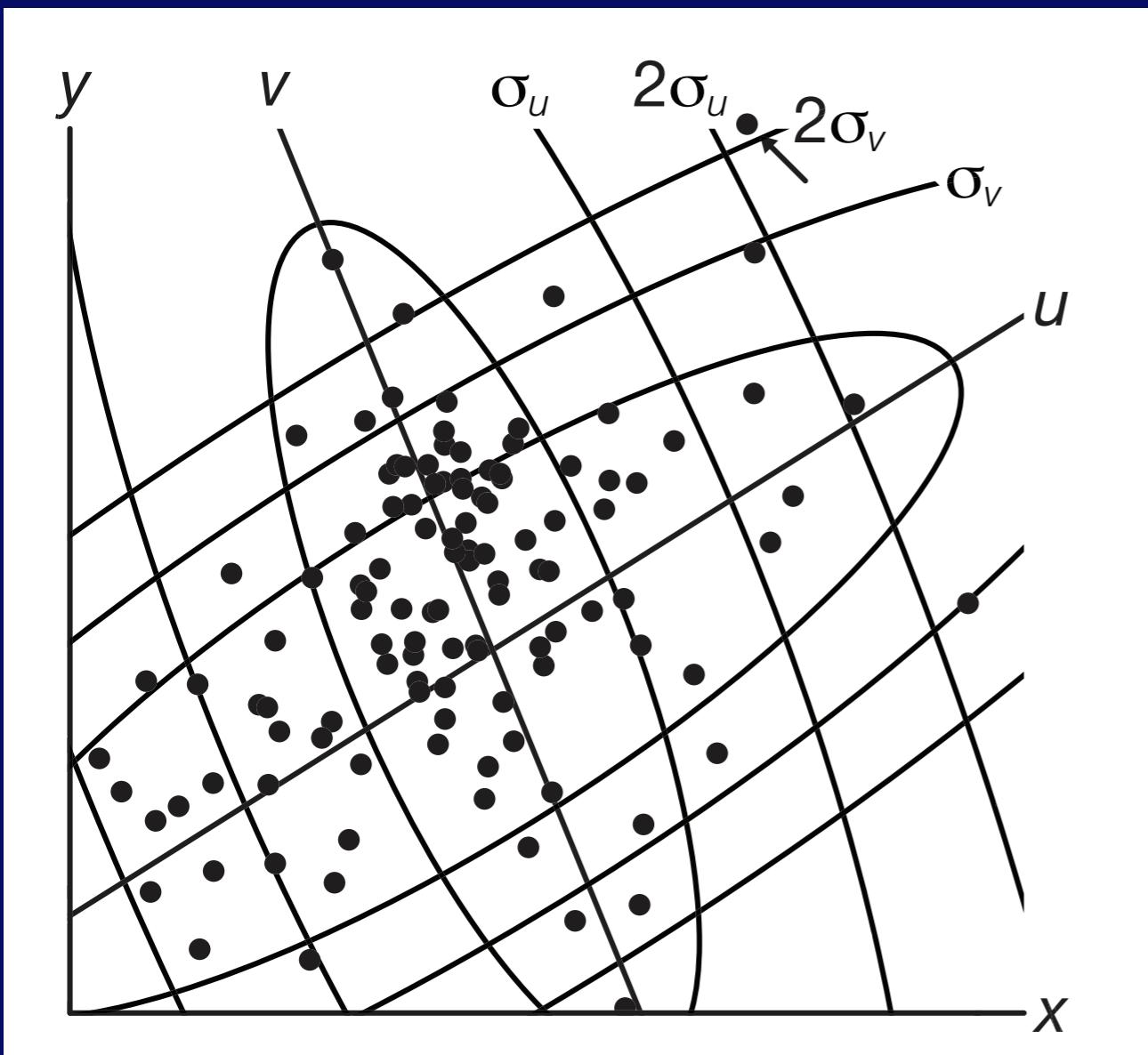
<http://www.datavizcatalogue.com>

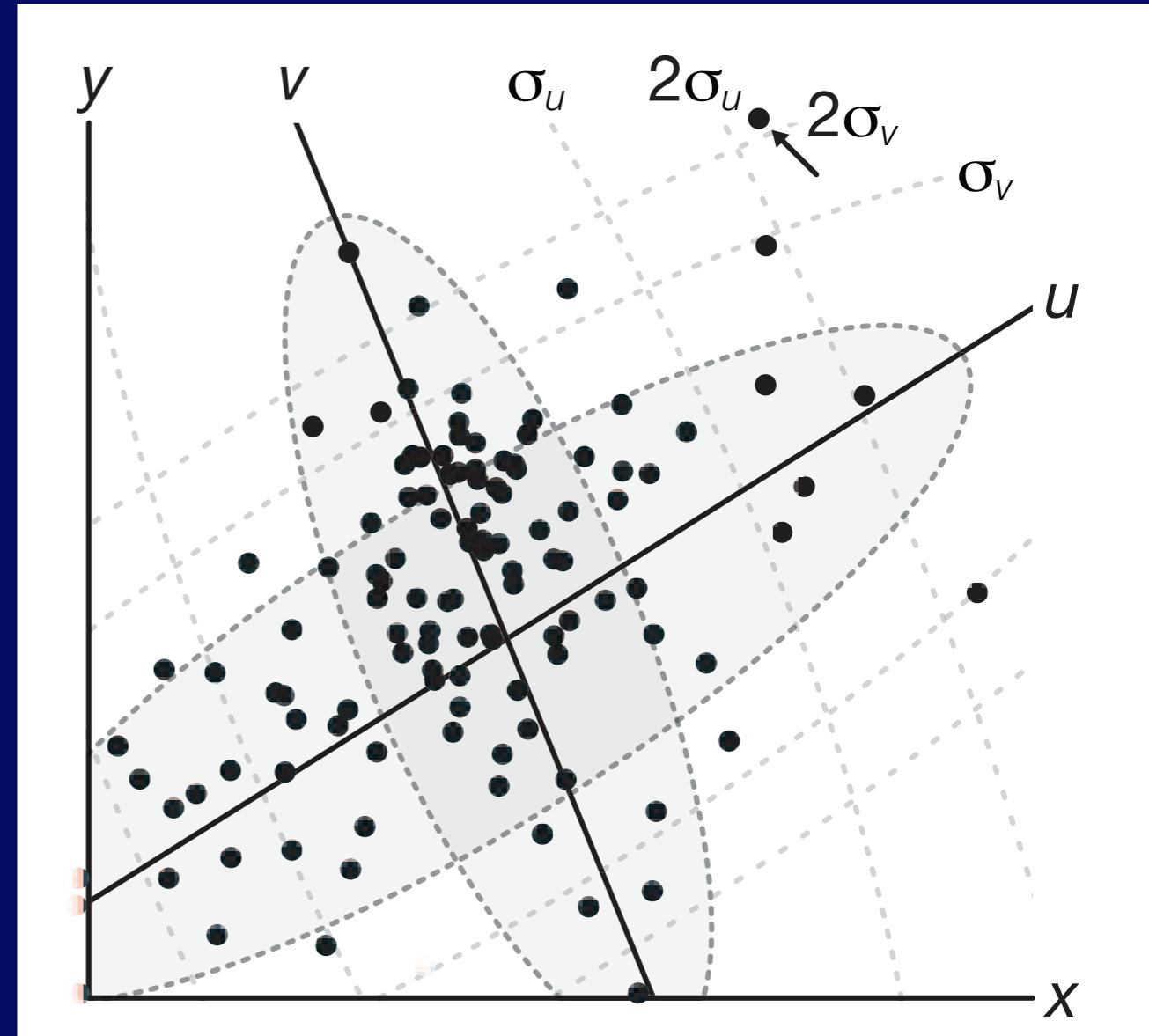
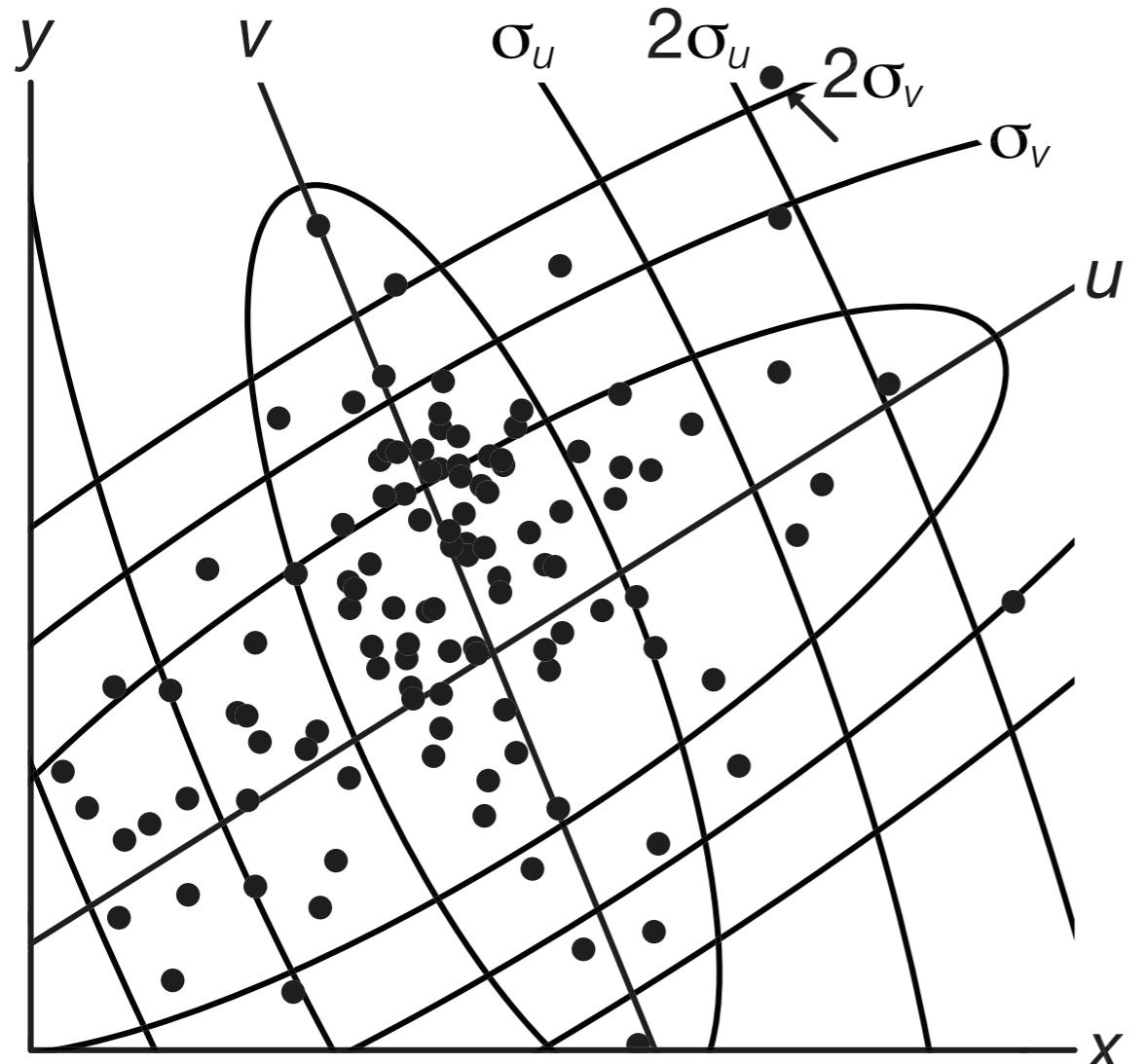
**in practice ...**

# World government



Bureau d'Etudes







**“Let the data speak.”**

**“Let the data speak.”**

**“Let the data speak.”**

**“Let the visualization support  
your hypothesis.”**

## More Resources

- <http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html>
- <http://www.jstor.org/stable/pdf/2683253.pdf>
- <http://mason.gmu.edu/~dcarr/lib/v6n3.pdf>
- <http://www.stat.columbia.edu/~gelman/research/published/dodhia.pdf>
- <https://flowingdata.com/category/visualization/ugly-visualization/>
- <http://www.datavis.ca>
- <https://curriculum.code.org/csp/unit2/10/>
- [http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/save\\_the\\_pies\\_for\\_dessert.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf)
- <http://www.datavizcatalogue.com>
- **Edward Tufte: The Visual Display of Quantitative Information**
- <https://www.csc2.ncsu.edu/faculty/healey/PP>
- <http://www-users.cs.umn.edu/%7einterran/texture/index.html>
- [https://en.wikipedia.org/wiki/Figure-ground\\_\(perception\)](https://en.wikipedia.org/wiki/Figure-ground_(perception))
- [http://www.improving-visualisation.org/case-studies/id=7#jump\\_3](http://www.improving-visualisation.org/case-studies/id=7#jump_3)

# Your turn!

(1) find a figure to improve (10 min)

# Your turn!

**(2) In groups of 3, discuss your figure! (20 min)**

- **In one sentence, what do you think the main point of the figure is?**
- **Do the others agree?**
- **What has the figure done well?**
- **What do you think could be improved?**
- **Consult data visualization checklist!**

# Your turn!

**(3) In your group, draft an improved version of the figure  
(on paper) + discuss draft (20 min)**

**(4) Remake the figure! (40 minutes)**

- **If you're stuck, ask for help!**
- **If you're done, walk around and help others!**

**(5) Figure show-and-tell!**

# Visualizations for Exercise

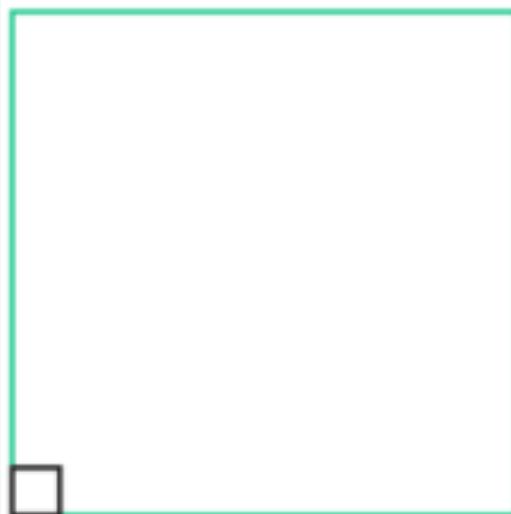
## Fragmented Food

Large number of small producers boosts China's vulnerability to food fraud

**Number of farms with cropland (million)**

China 184

U.S. 1.6



Source: PricewaterhouseCoopers LLC's Food Fraud Vulnerability Assessment and Mitigation report; citing China 2006 Agricultural Census, 2014 China Animal Husbandry Year Book, 2012 USDA Census of Agriculture

Note: Numbers most recent available

**Number of farms that raised pigs (million)**

China 52

U.S. 0.56



**Number of farms that raised broilers (million)**

China 24

U.S. 0.33



Bloomberg

## Anatomy of a Winning TED Talk

1%

### Sophisticated Visual Aids

We're not sure who puts the D in TED—most of the best presentations favor tepid PowerPoint slide shows (sorry, Brené Brown), Pictionary-quality drawings (really, Simon Sinek?), or no props at all.

5%

### Opening Joke

Remember the one about the shoe salesmen who went to Africa in the 1900s? That's how Benjamin Zander opened his talk—which turned out to be about classical music.

5%

### Spontaneous Moment

Don't overprepare. Tease the guy in the front row ("You could light up a village with this guy's eyes"). Command the stagehand who handles the human brain you brought.

5%

### Statement of Utter Certainty

People come for answers—give 'em what they want, as Shawn Achor did: "By training your brain ... we can reverse the formula for happiness and success."

12%

### Snappy Refrain

The TED equivalent of "I have a dream." Example: "People don't buy what you do; they buy why you do it." Repeat 7x.

23%

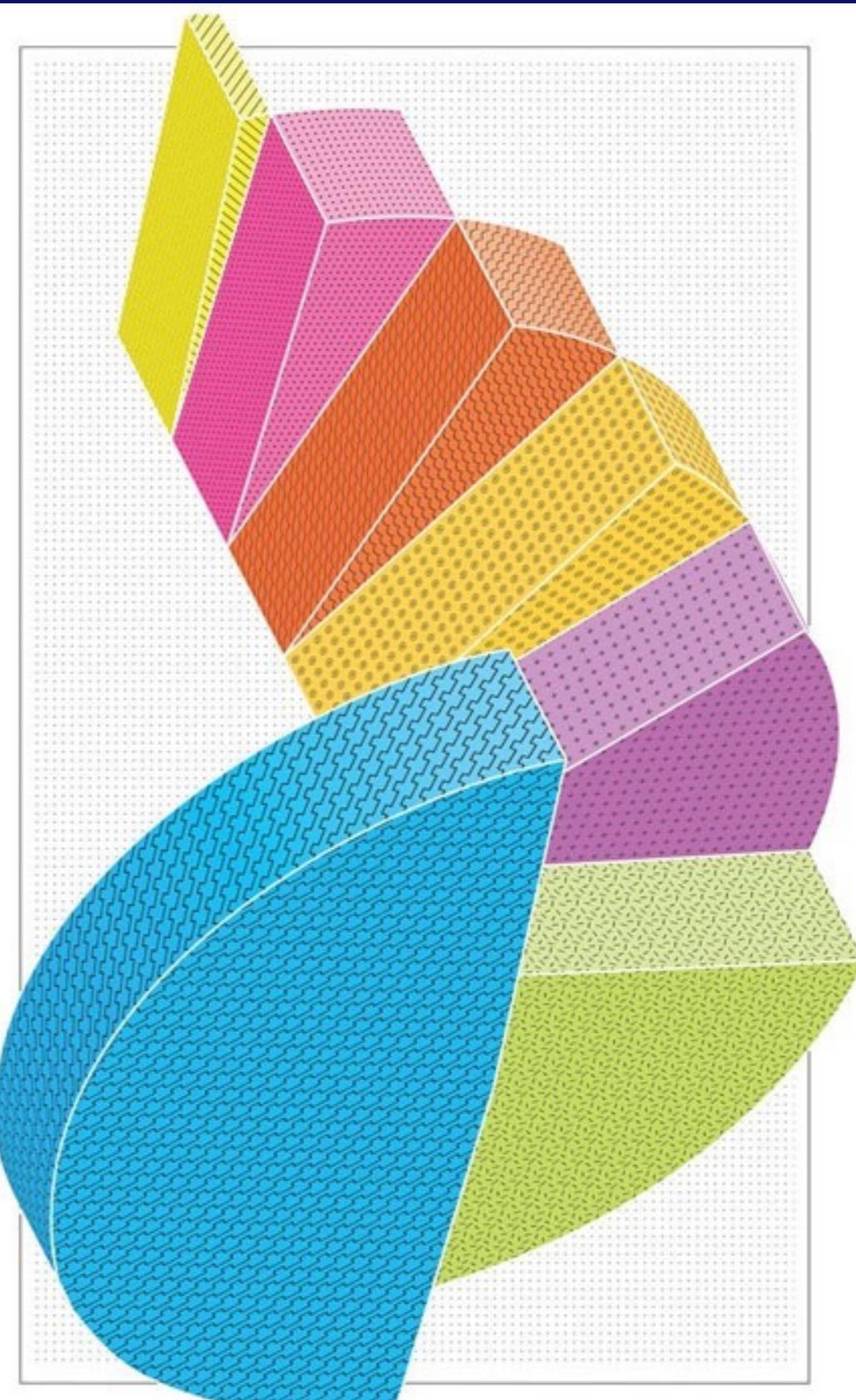
### Personal Failure

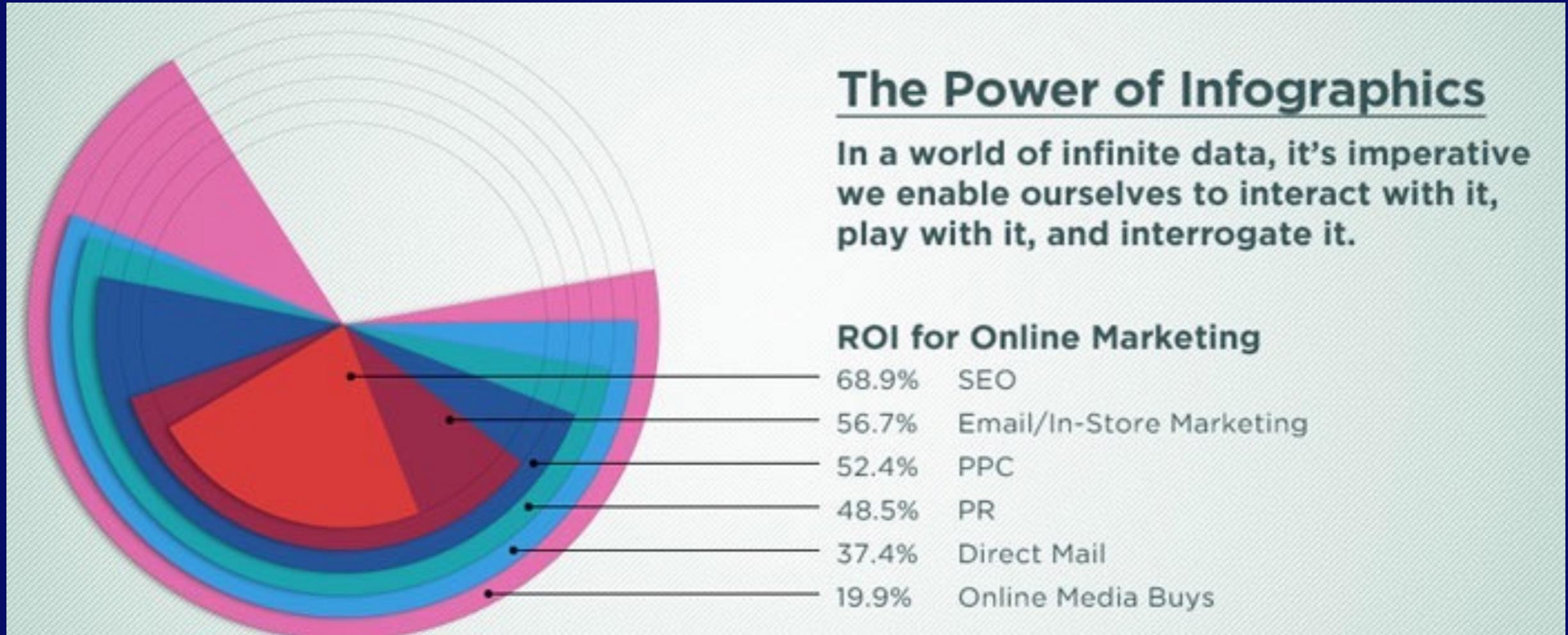
Be relatable. We want to know about that nervous breakdown. Or at least the time you didn't fit in at summer camp.

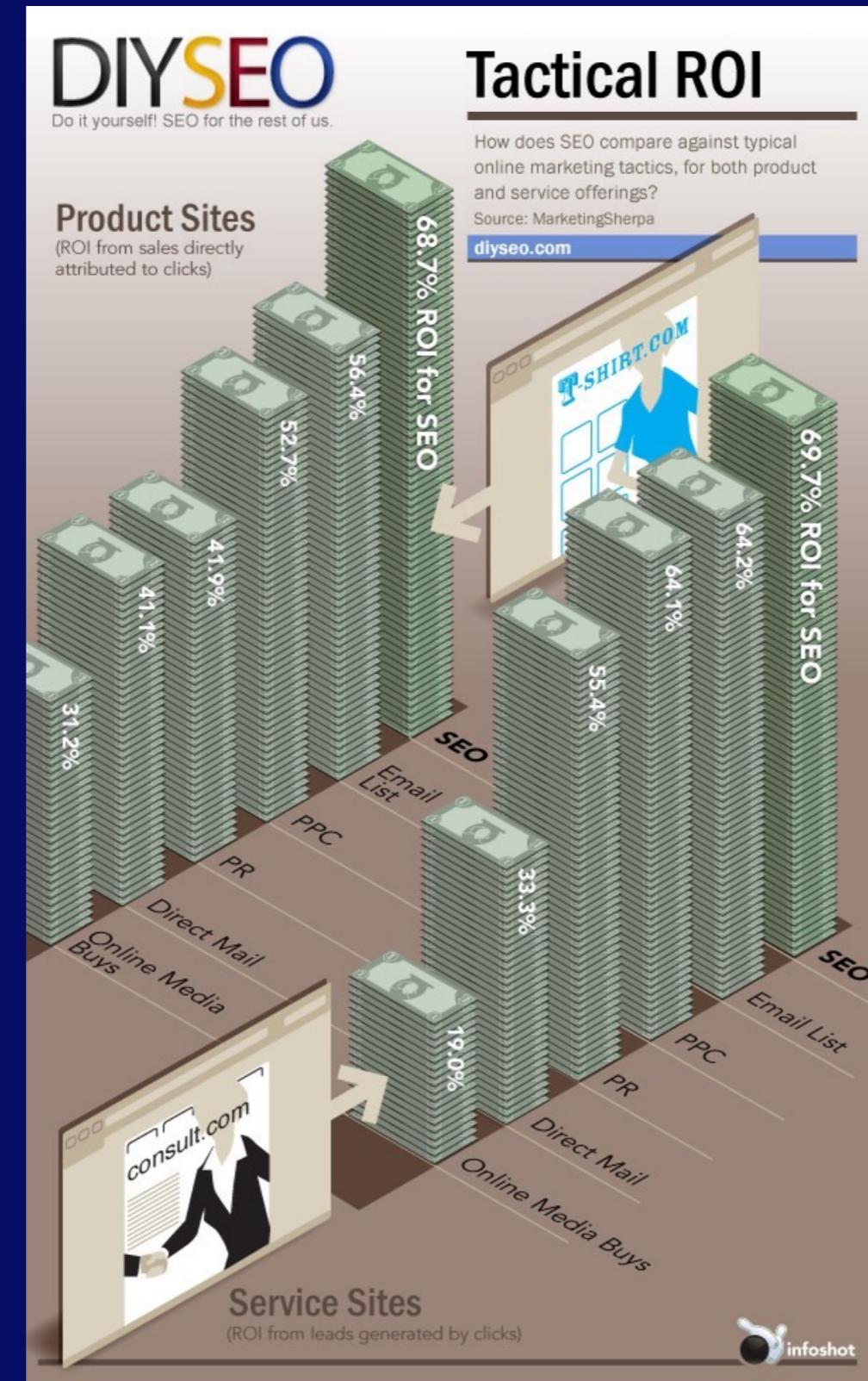
49%

### Contrarian Thesis

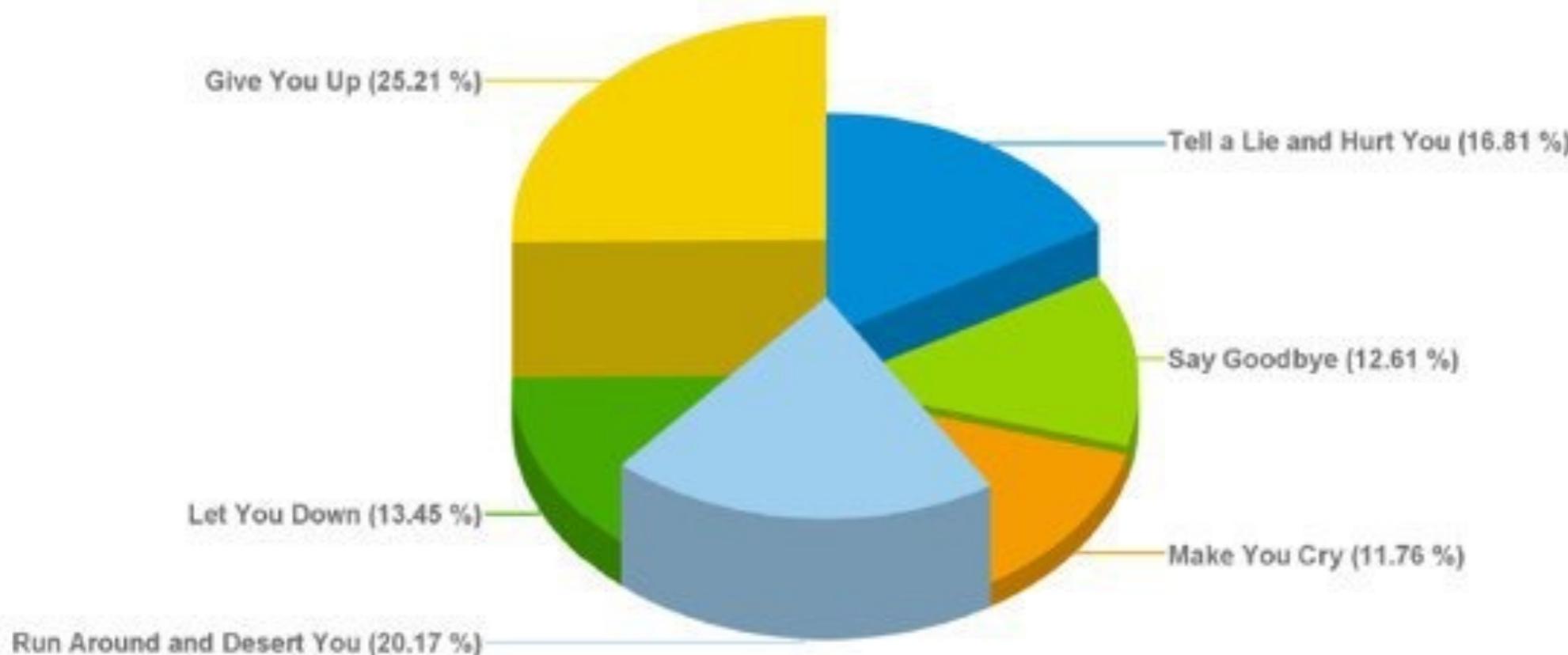
Wait a sec—we should be playing *more* videogames? The more choices we have, the worse off we are? TED is where conventional wisdom goes to die.







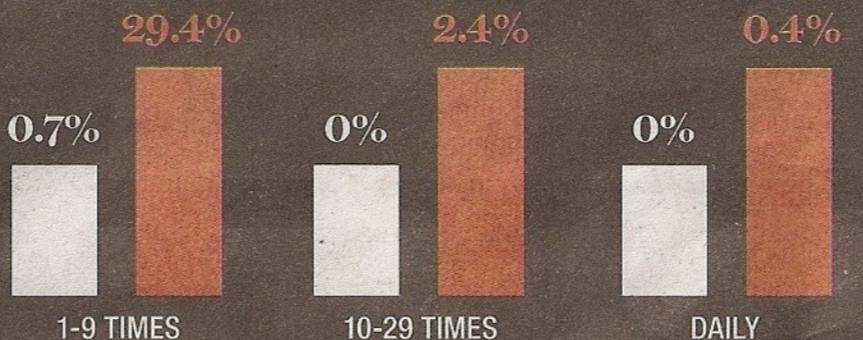
## Things Rick Astley Would Never Do



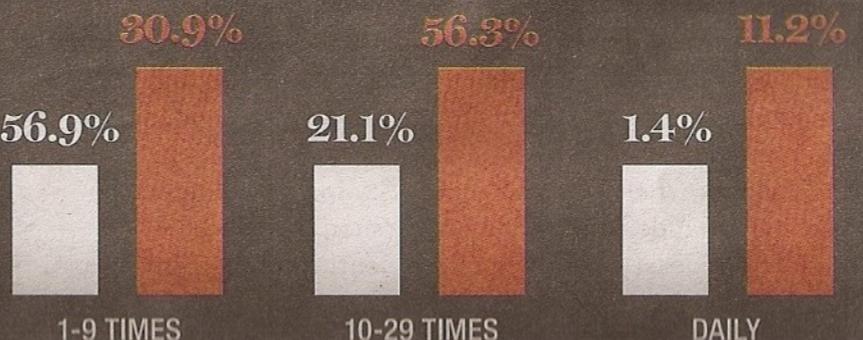
# BY THE NUMBERS

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.

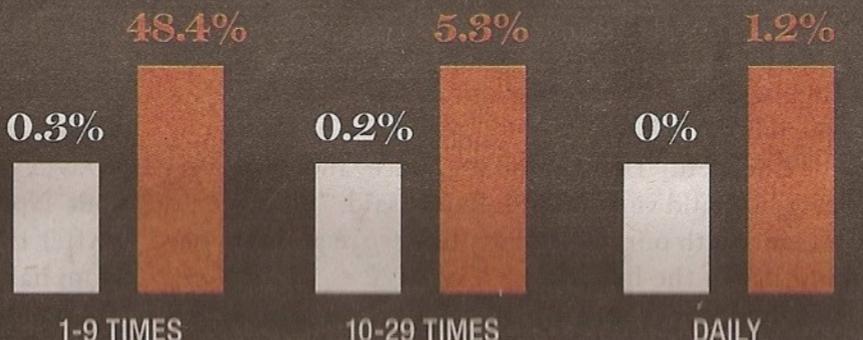
## OPIATES

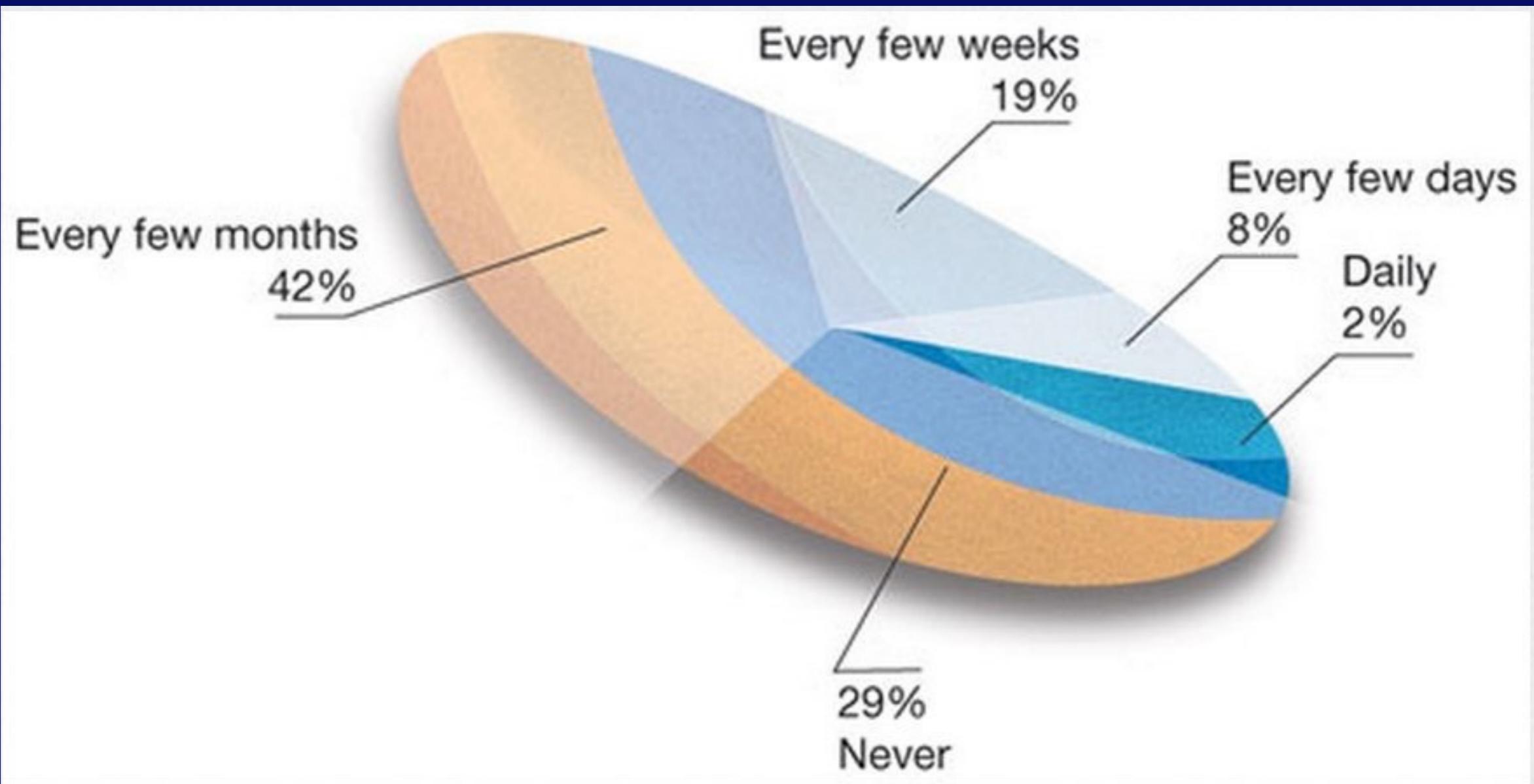


## ALCOHOL



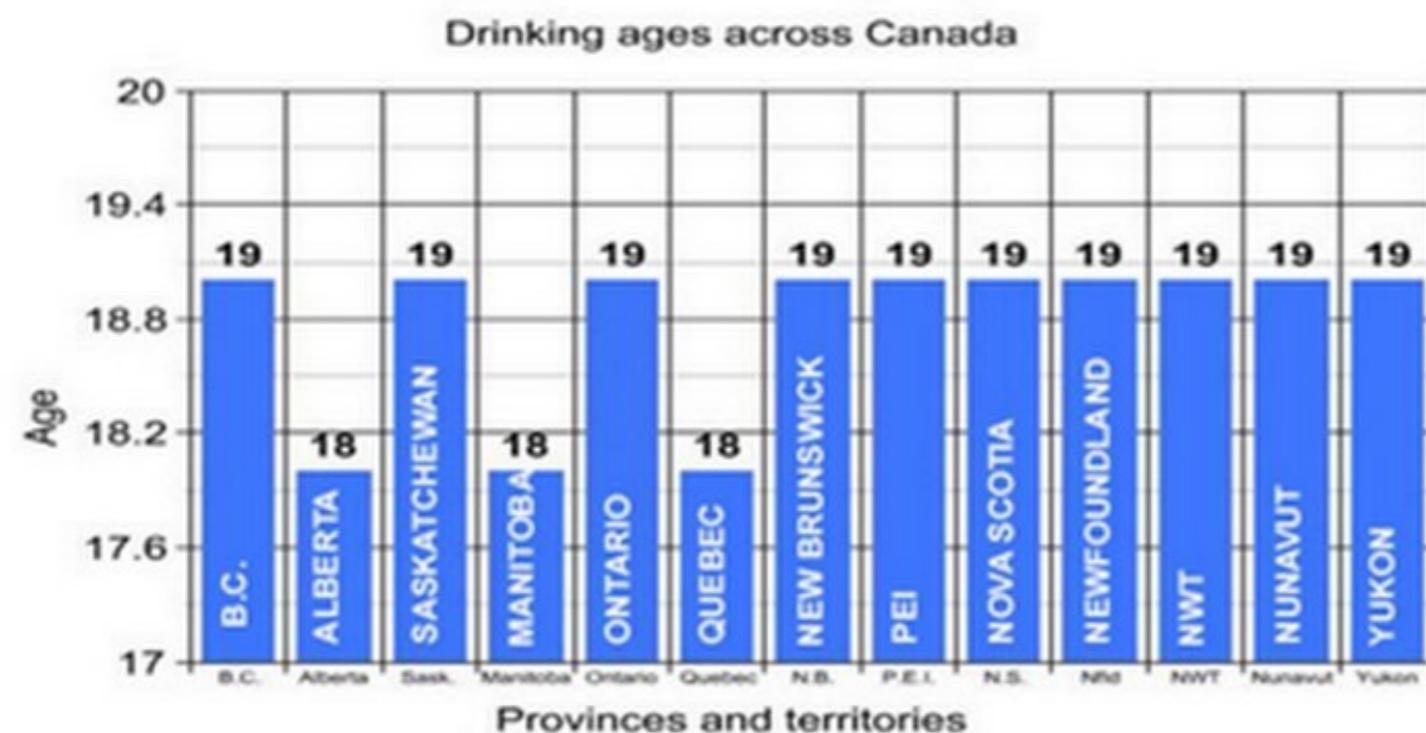
## COCAINE





# Drinking age will remain 19 in Saskatchewan

CBC News Posted: Mar 4, 2013 11:59 AM CST | Last Updated: Mar 4, 2013 11:55 AM CST 25

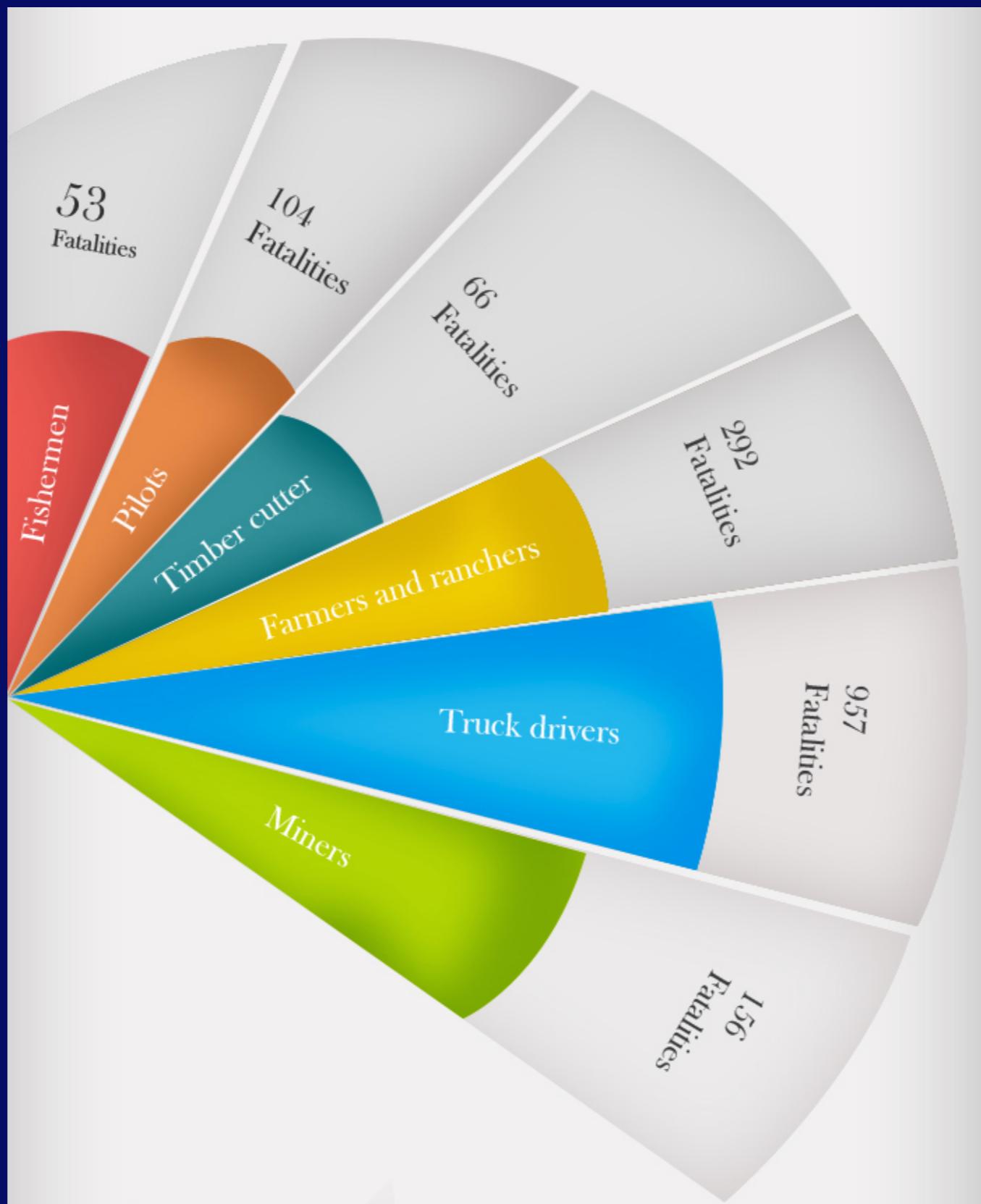


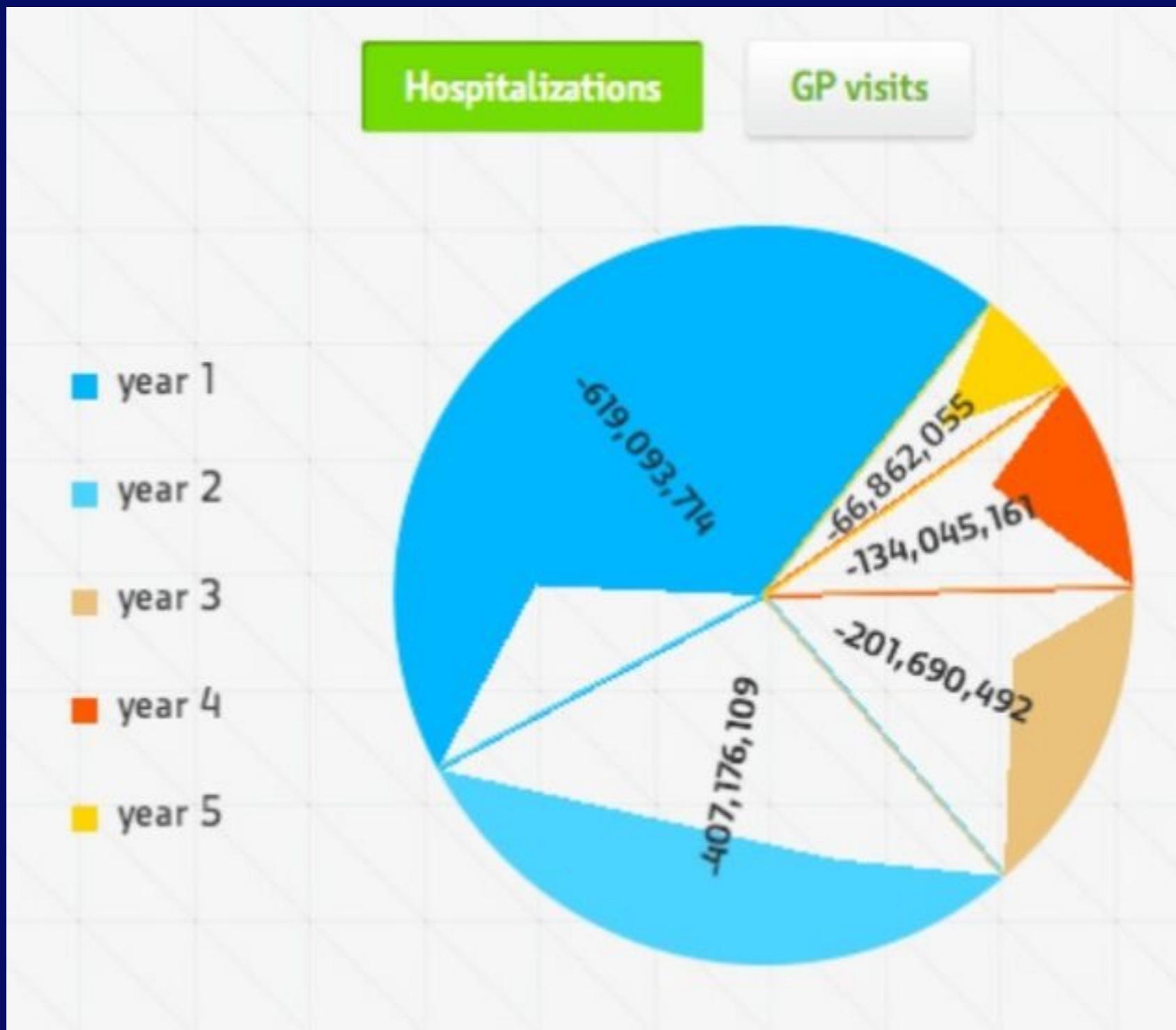
Canadian Centre on Substance Abuse

You have to be 19 in Saskatchewan to have a drink, while in Alberta and Manitoba, the drinking age 18. (CBC)

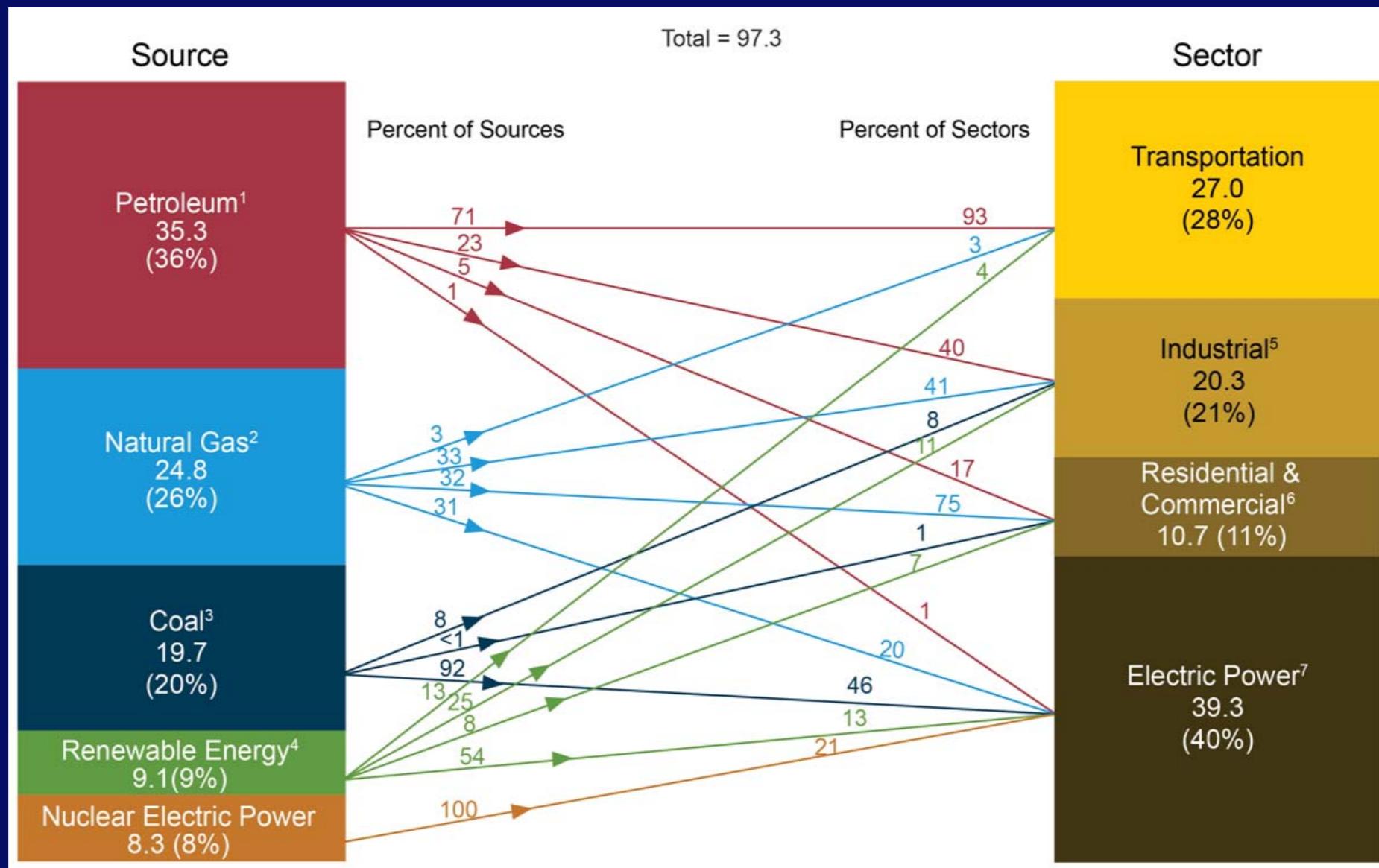
The Saskatchewan Party government has ruled out lowering the drinking age, four months after party members put the issue in the public eye.







**Figure 2.0 Primary Energy Consumption by Source and Sector, 2011**  
 (Quadrillion Btu)



<sup>1</sup> Does not include biofuels that have been blended with petroleum—biofuels are included in “Renewable Energy.”

<sup>2</sup> Excludes supplemental gaseous fuels.

<sup>3</sup> Includes less than 0.1 quadrillion Btu of coal coke net imports.

<sup>4</sup> Conventional hydroelectric power, geothermal, solar/photovoltaic, wind, and biomass.

<sup>5</sup> Includes industrial combined-heat-and-power (CHP) and industrial electricity-only plants.

<sup>6</sup> Includes commercial combined-heat-and-power (CHP) and commercial electricity-only plants.

<sup>7</sup> Electricity-only and combined-heat-and-power (CHP) plants whose primary business is to sell electricity, or electricity and heat, to the public. Includes 0.1 quadrillion Btu of electricity net imports not shown under “Source.”

Notes: Primary energy in the form that it is first accounted for in a statistical energy balance, before any transformation to secondary or tertiary forms of energy (for example, coal is used to generate electricity). • Sum of components may not equal total due to independent rounding.

Sources: U.S. Energy Information Administration, *Annual Energy Review 2011*, Tables 1.3, 2.1b-2.1f, 10.3, and 10.4.

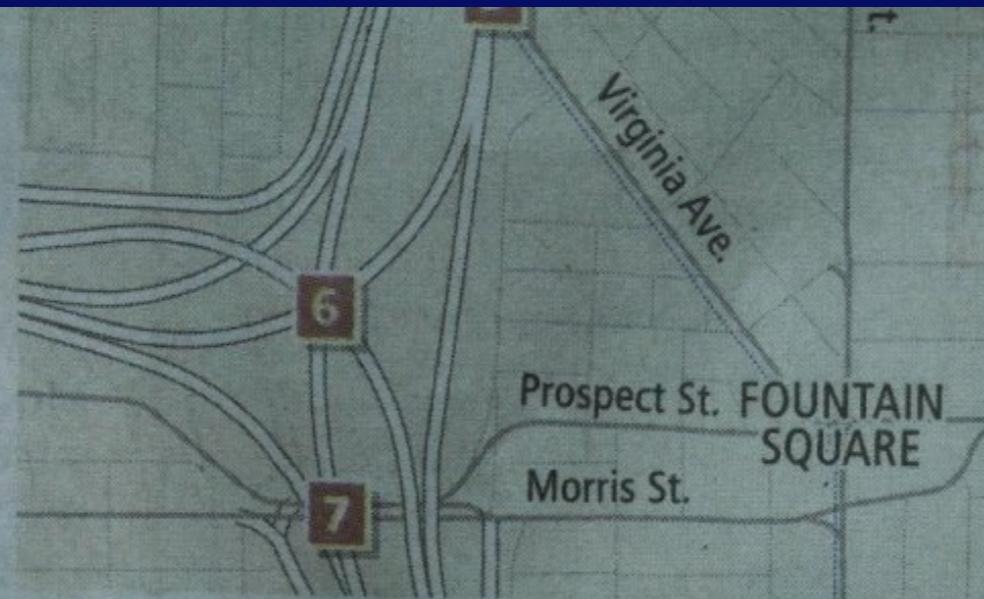
INDOT hopes to begin work in late August. About 9,000 drivers travel through the area daily, according to numbers from 2011, the most recent year available. The announcement follows a report in March 2012, in which an overpass bridge struck the Virginia Avenue bridge, shutting down westbound I-65 and eastbound traffic for emergency repairs over a week.

There have been no recorded more similar incidents since.

"We want to emphasize that the bridge is structurally safe," said Brian Wingfield, INDOT deputy director. "Our goal is to decrease the probability of bridge

closures, Wingfield said. Movement will be low-

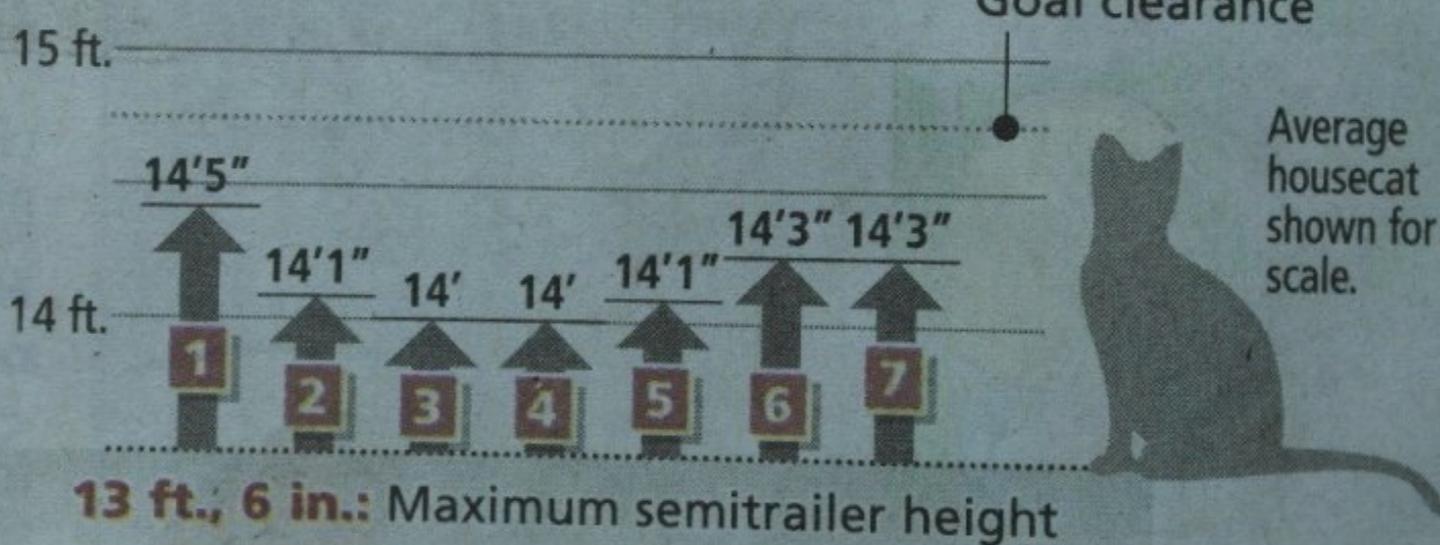
» See SPLIT, Page B3



5. Prospect Street bridge over I-65 northbound  
6. I-65 southbound ramp bridge to I-70 westbound  
7. Morris Street bridge over I-65 southbound

How heights of the bridges above compare with the maximum height for semitrailers:

15 ft.



SOURCE: INDOT

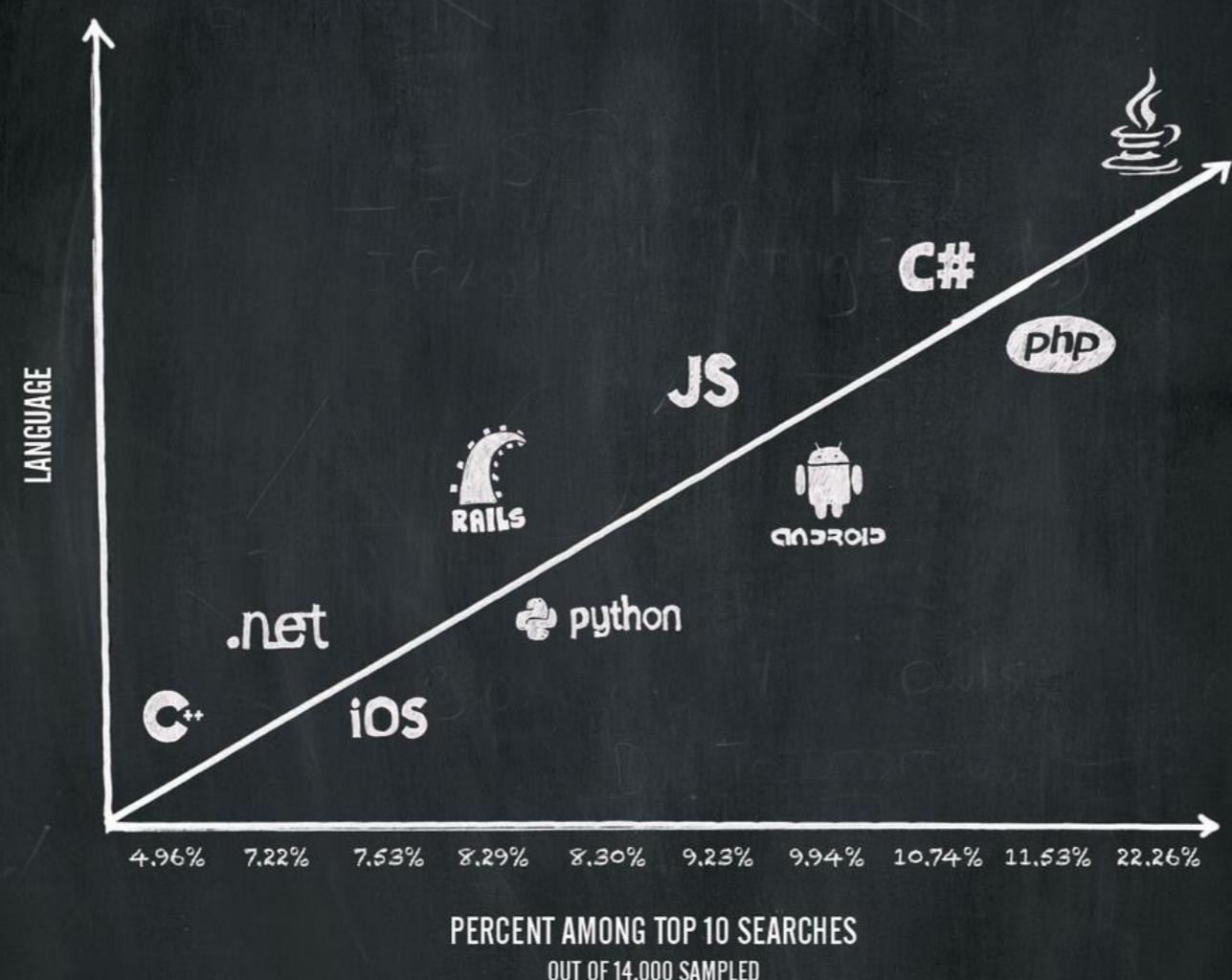
STEPHEN J. BEARD / THE STAR

# This Carmel appearance

readwrite presents

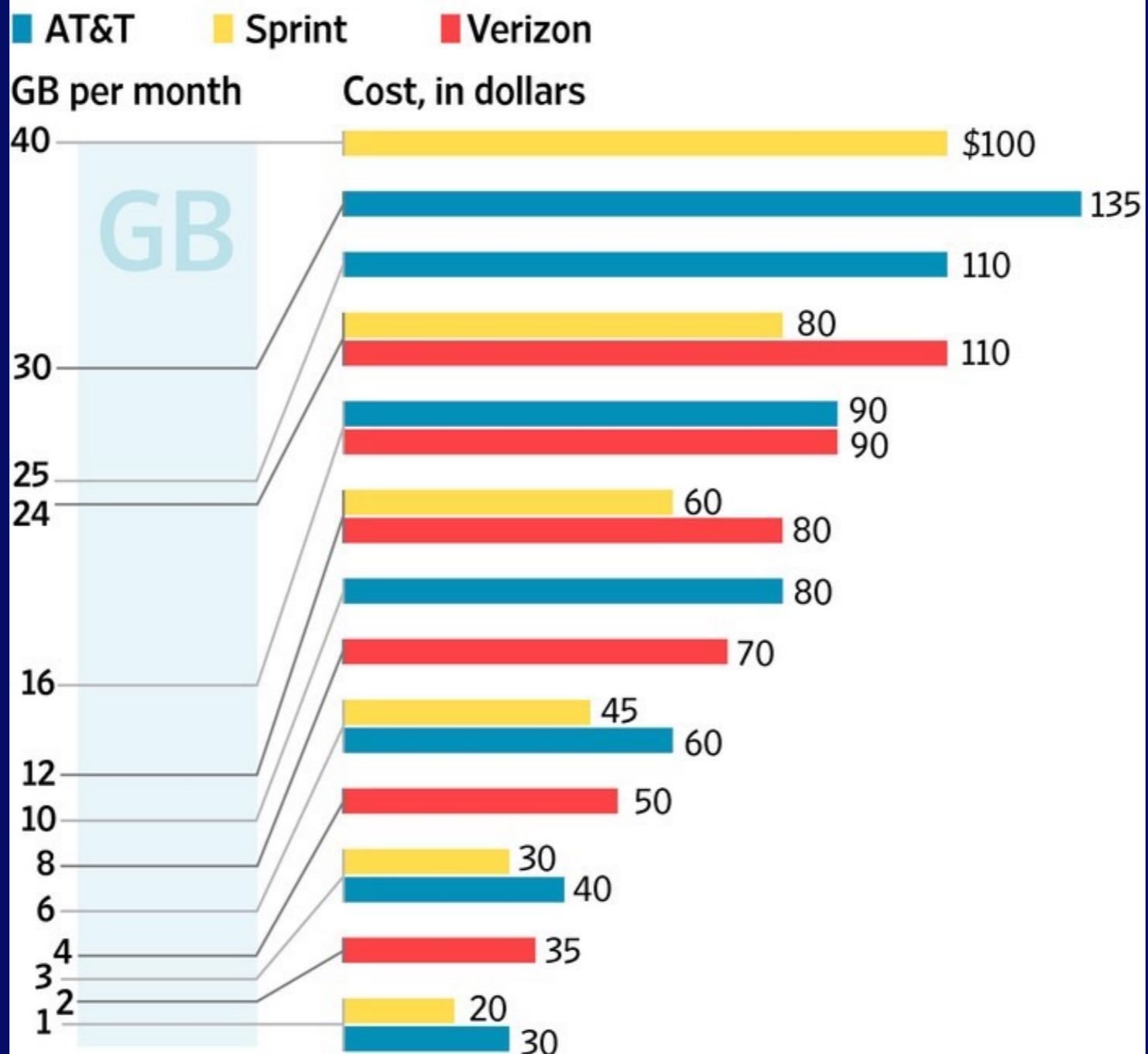
## TOP 10 MOST IN DEMAND DEVELOPER SKILLS OF 2013

INFORMATION COMPILED BY STACK OVERFLOW



## Buying in Buckets

AT&T, Verizon and Sprint charge the same \$20 per phone but have different data allowance levels. Comparison isn't easy.

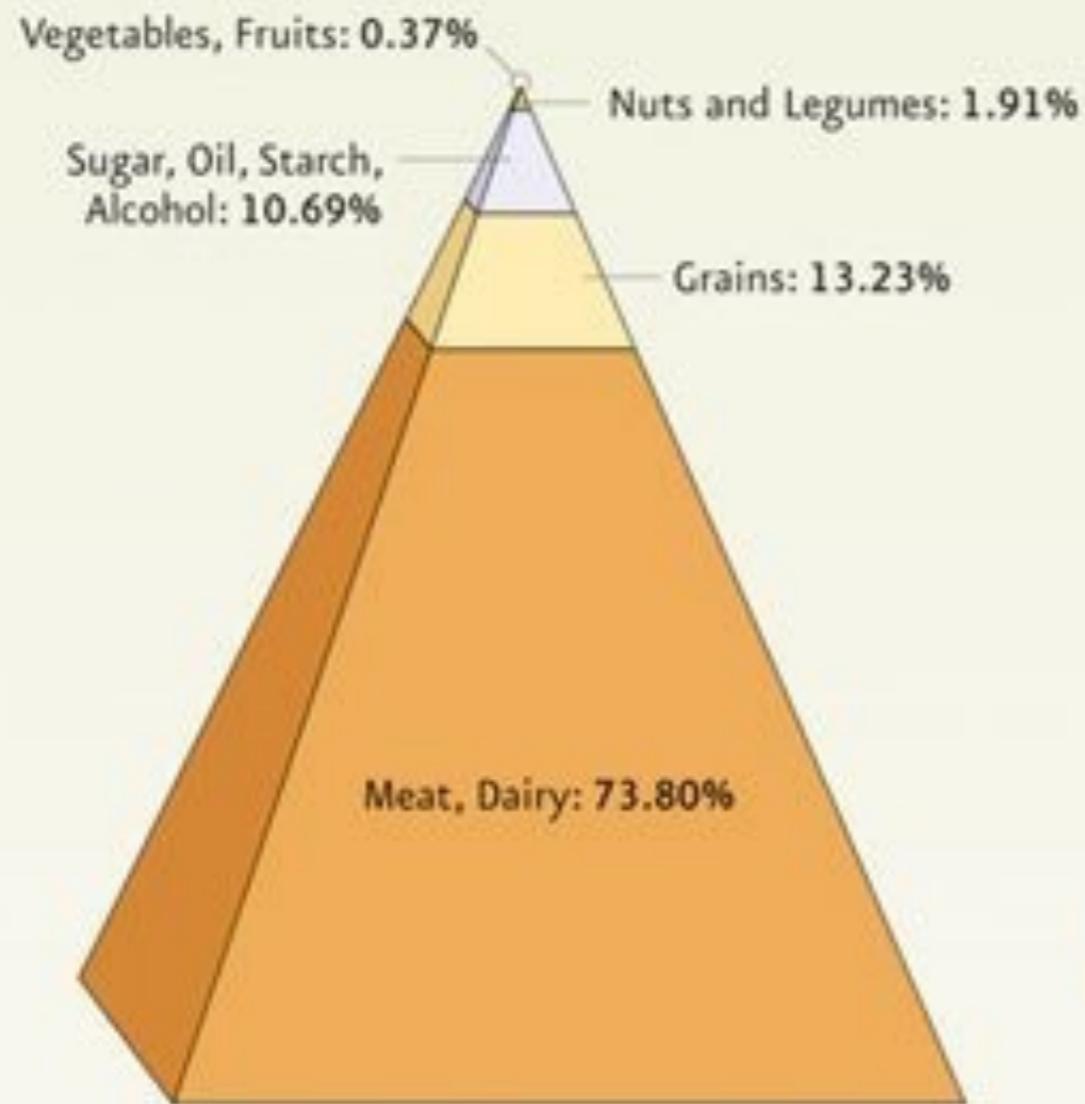


Sources: the companies

THE WALL STREET JOURNAL.

# Why Does a Salad Cost More Than a Big Mac?

Federal Subsidies for Food Production, 1995-2005\*



Federal Nutrition Recommendations

