Urban Informatics

Fall 2018

dr. federica bianco fbianco@nyu.edu

@fedhere

Recap:

- Good practices with data: falsifiability, reproducibility
- Basic data retrieving and munging: APIs, Data formats
- Basic statistics: distributions and their moments
- Hypothesis testing: *p*-value, statistical significance
- Geospatial analysis

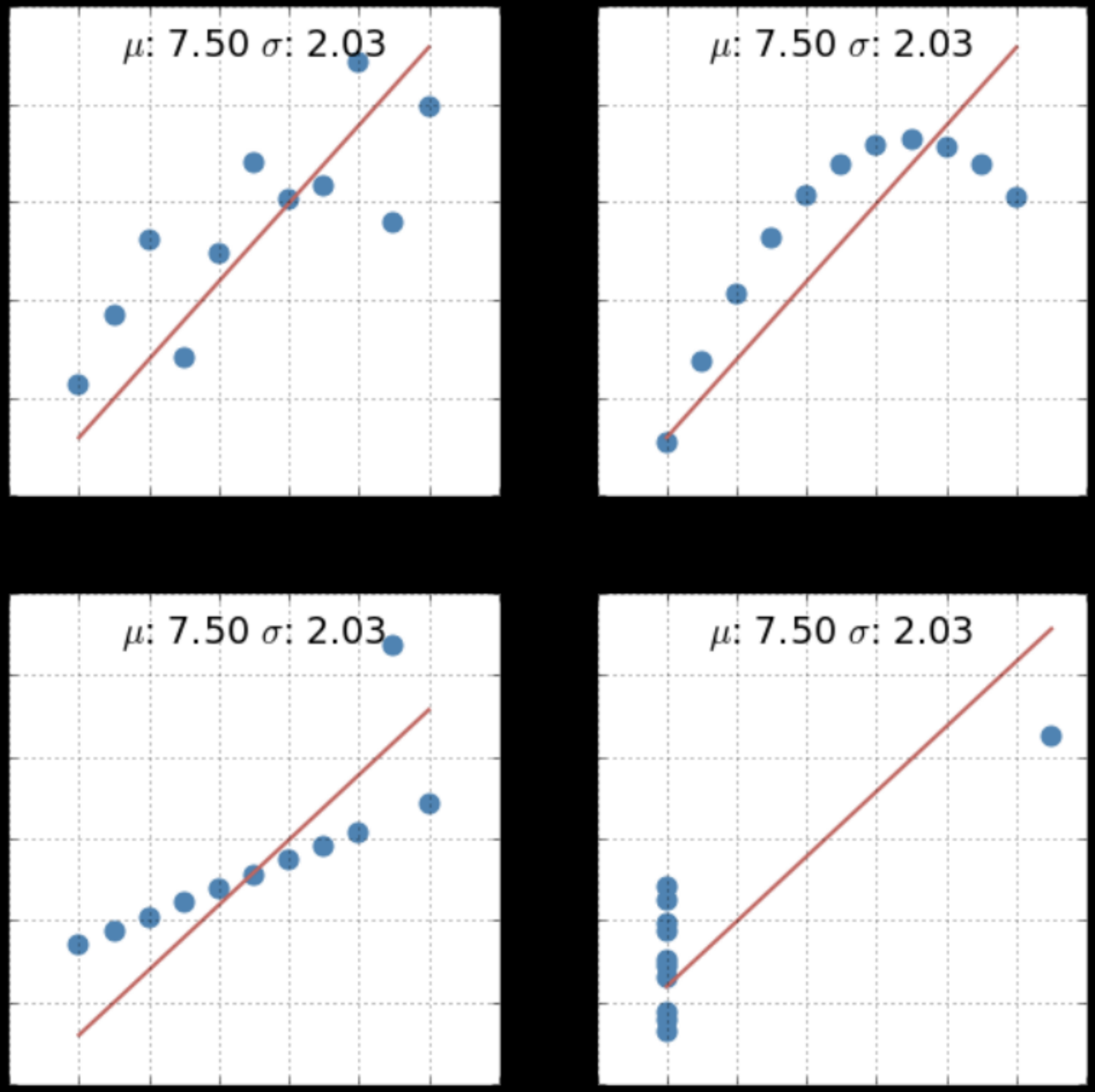
Today:

- Visualizations

Why visualize?

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

What’s this??


μ: 7.50 σ: 2.03
μ: 7.50 σ: 2.03
μ: 7.50 σ: 2.03
μ: 7.50 σ: 2.03

```python
from PIL import Image
im = Image.open('sectret.png')
pixels = list(im.getdata())
import numpy as np
data = np.asarray(im)

data[:,:,0]
```

```
array([[38, 35, 39, ..., 25, 33, 39],
       [44, 41, 42, ..., 25, 31, 37],
       [47, 45, 42, ..., 31, 35, 39],
       ...,
       [47, 55, 69, ..., 63, 66, 69],
       [33, 38, 51, ..., 56, 63, 69],
       [26, 20, 28, ..., 39, 44, 48]], dtype=uint8)
```

```
pl.figure(figsize=(10,10))
pl.imshow(data[500:1500,500:1500,0], cmap=plt.get_cmap('bone'))
pl.xticks([])
pl.yticks([])
```

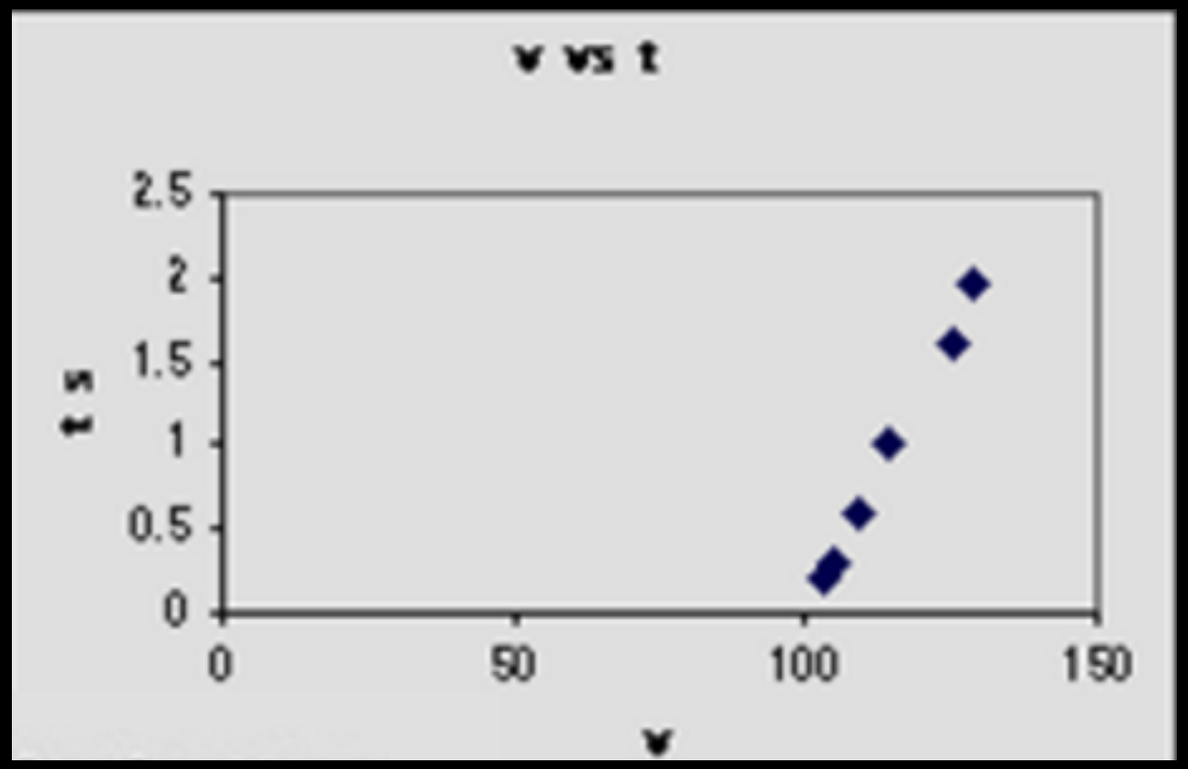
```
([], <a list of 0 Text yticklabel objects>)
```

we visualize to
communicate (Tufte)
and
explore (Thorp)

what makes a bad visualization?

Ambiguity, distortion (misleading), distraction.

http://slideplayer.com/slide/4956536/

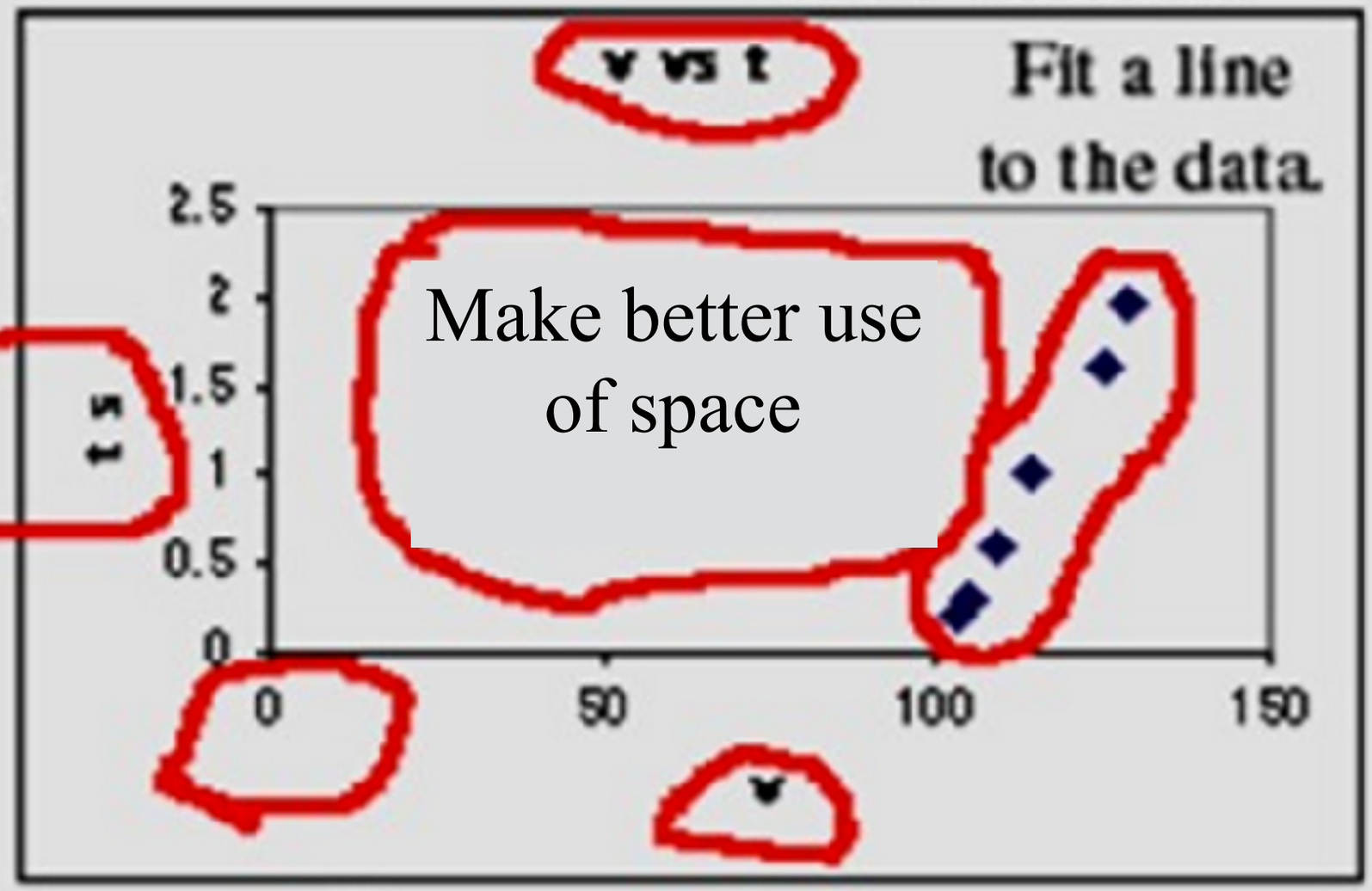


7 wrong things with this plot…

The entire graph is too small.

The title should be better. This graph is t versus v, not v versus t.

The axis label should have words, and the units should be in parentheses.

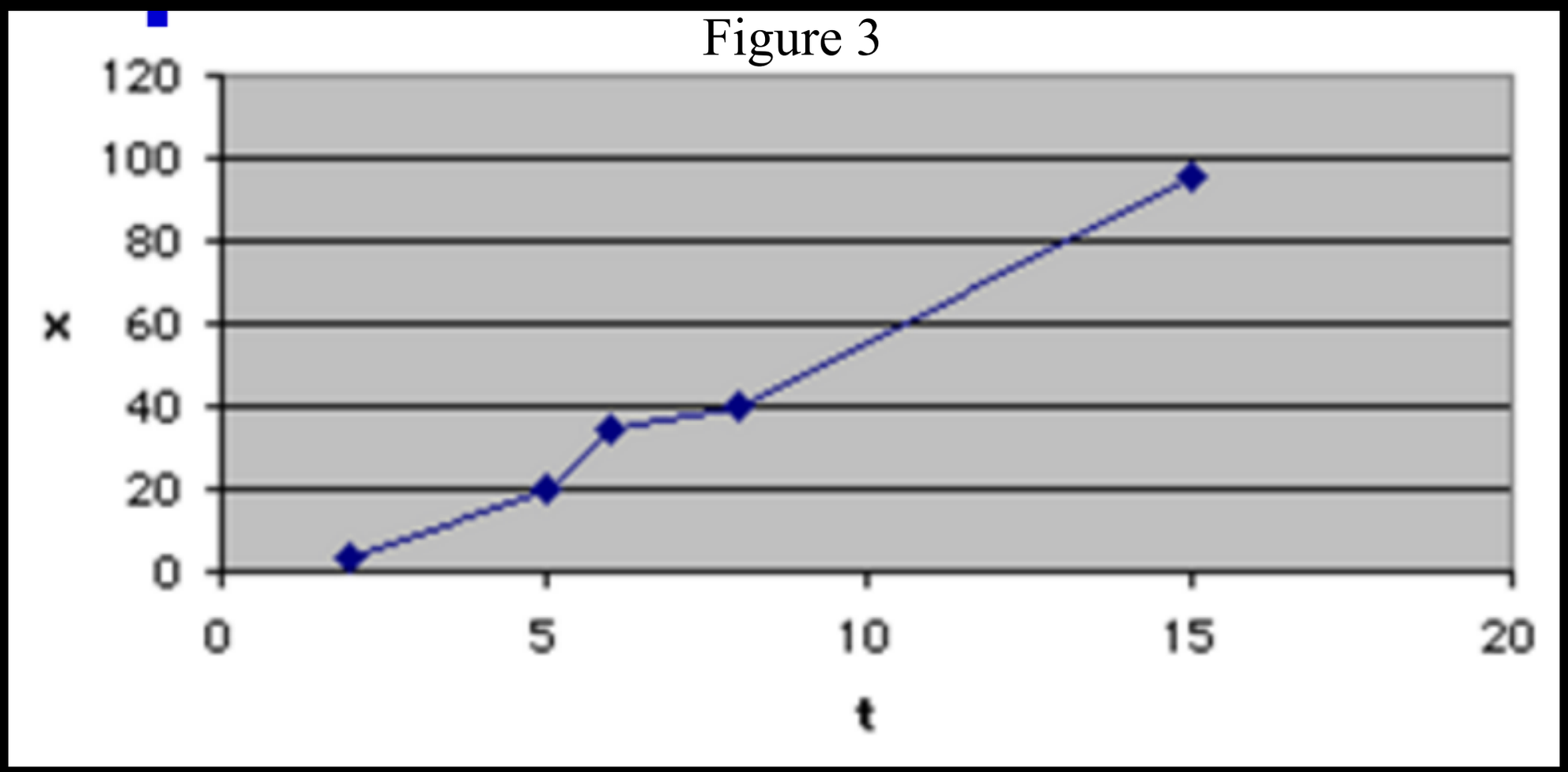


The first data occurs at v = 100, so the scale can begin at 100.

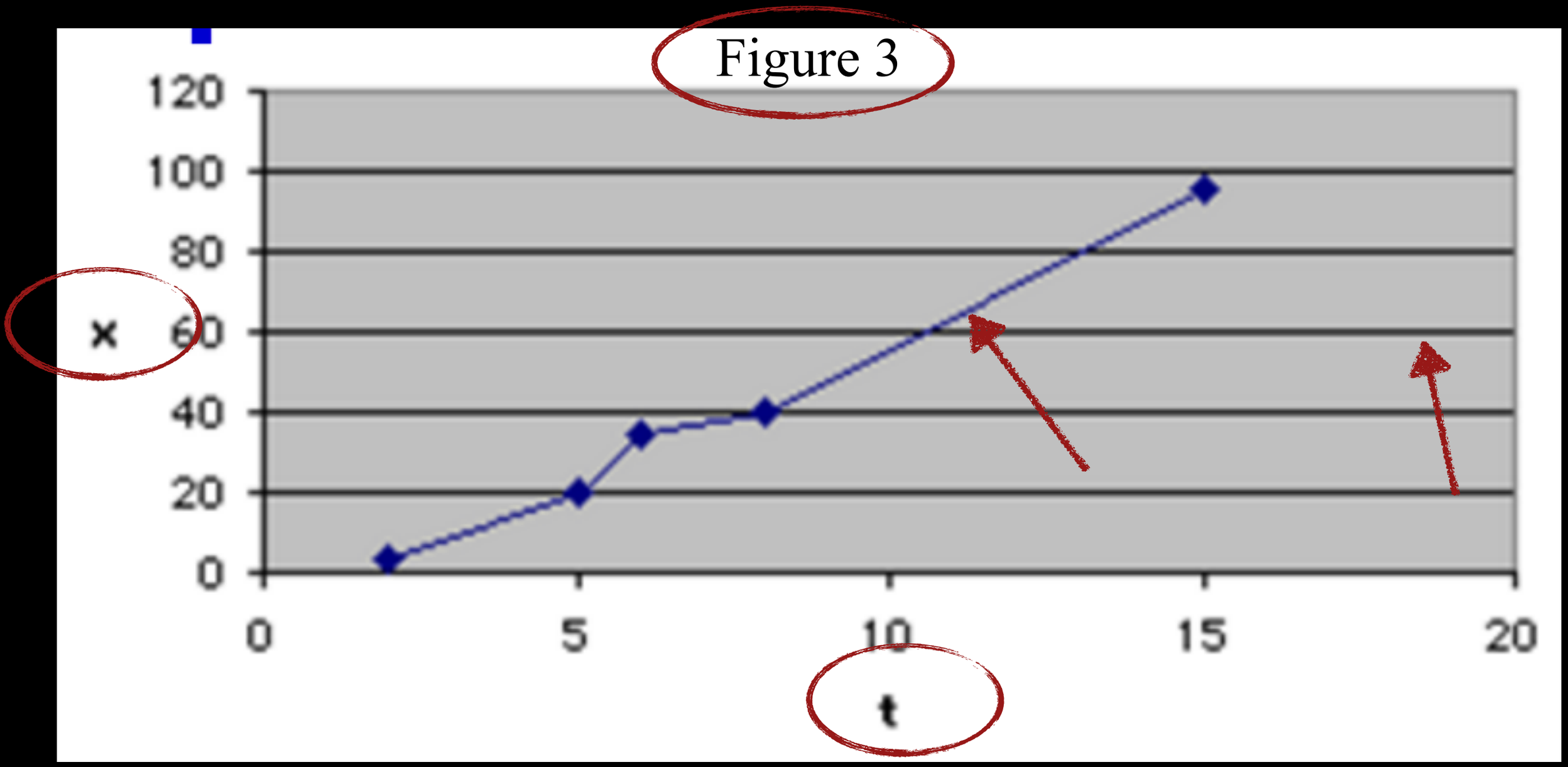
The axis label should have words and units in parentheses.

http://slideplayer.com/slide/4956536/

Figure 3



7 wrong things with this plot…

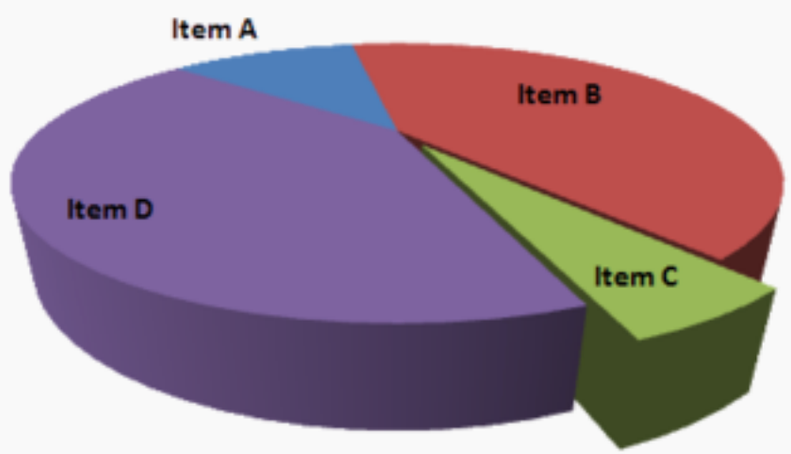
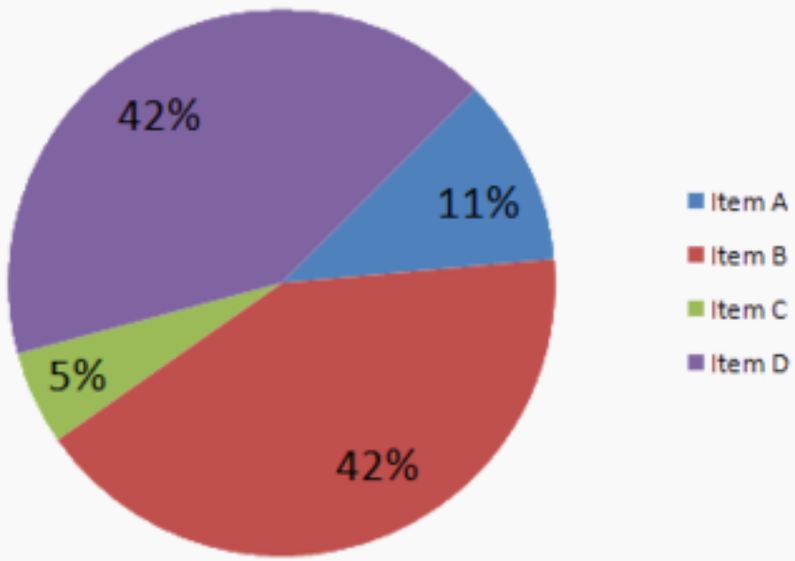


low res and small size!

http://slideplayer.com/slide/4956536/

Ambiguity, distortion (misleading), distraction.

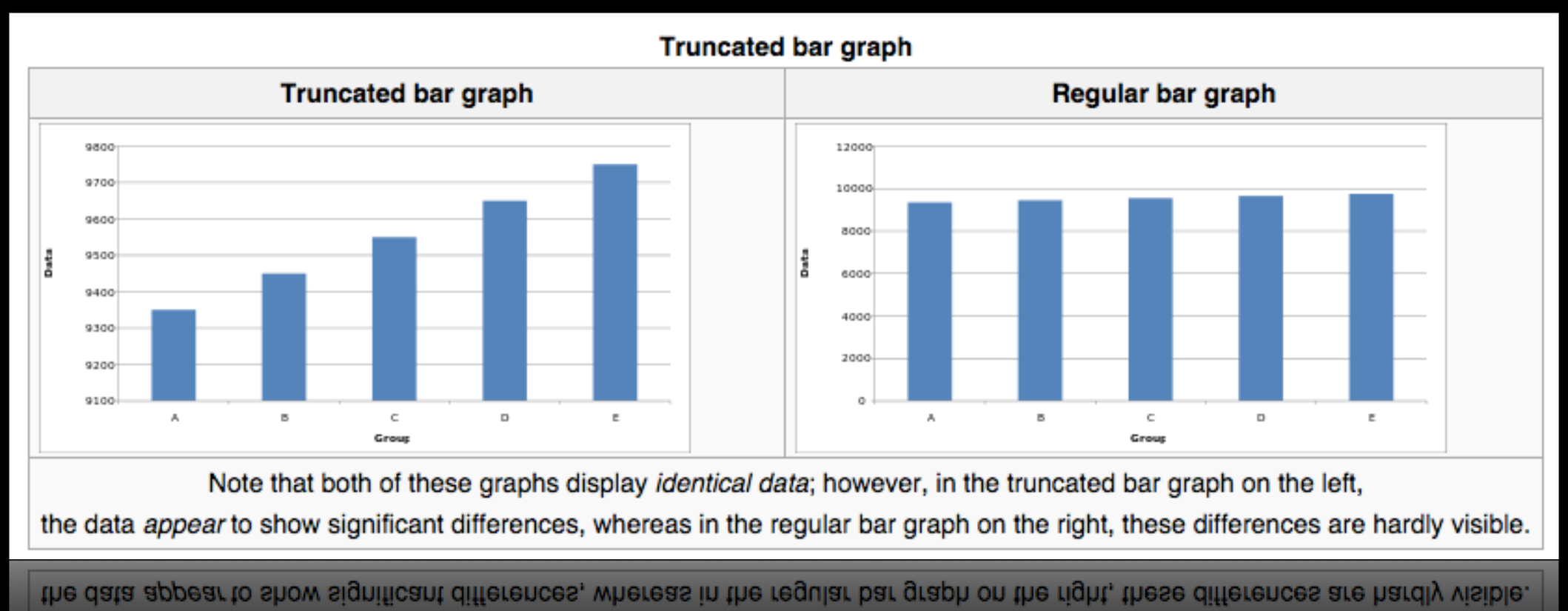
Comparison of pie charts

Misleading pie chart	Regular pie chart
	

In the misleading pie chart, Item C appears to be at least as large as Item A, whereas in actuality, it is less than half as large.

It is less than half as large.

http://gizmodo.com/how-to-lie-with-data-visualization-1563576606

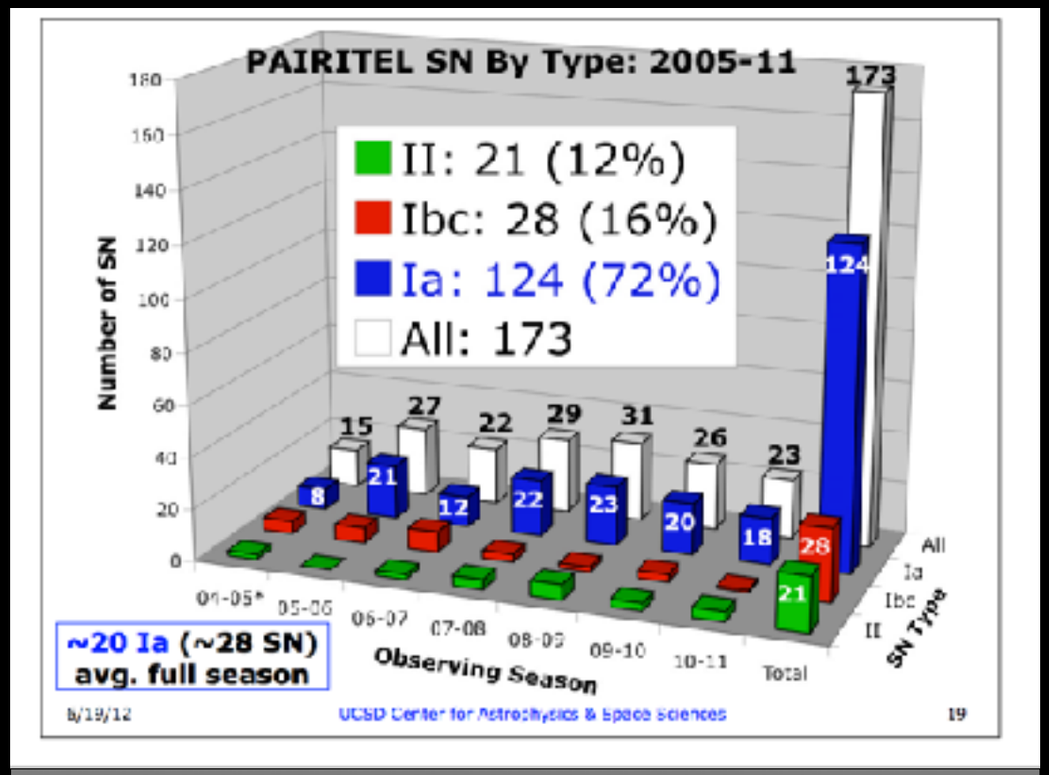


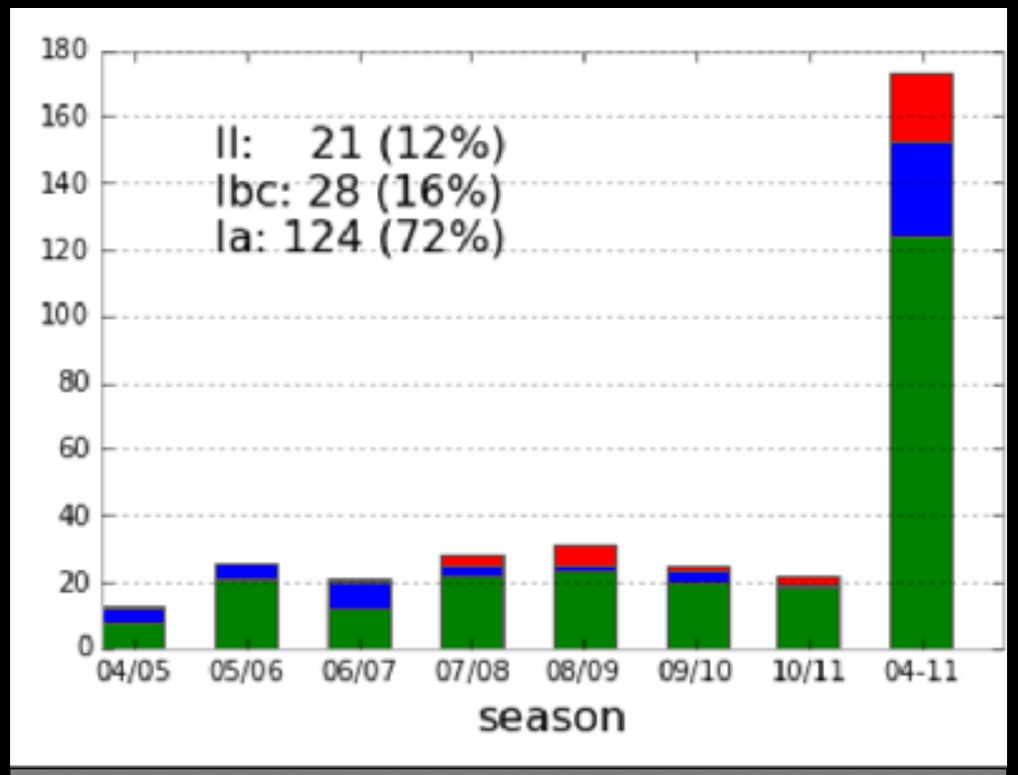
Ambiguity, distortion (misleading), distraction.

http://www.nodexlgraphgallery.org/Pages/Graph.aspx?graphID=56967

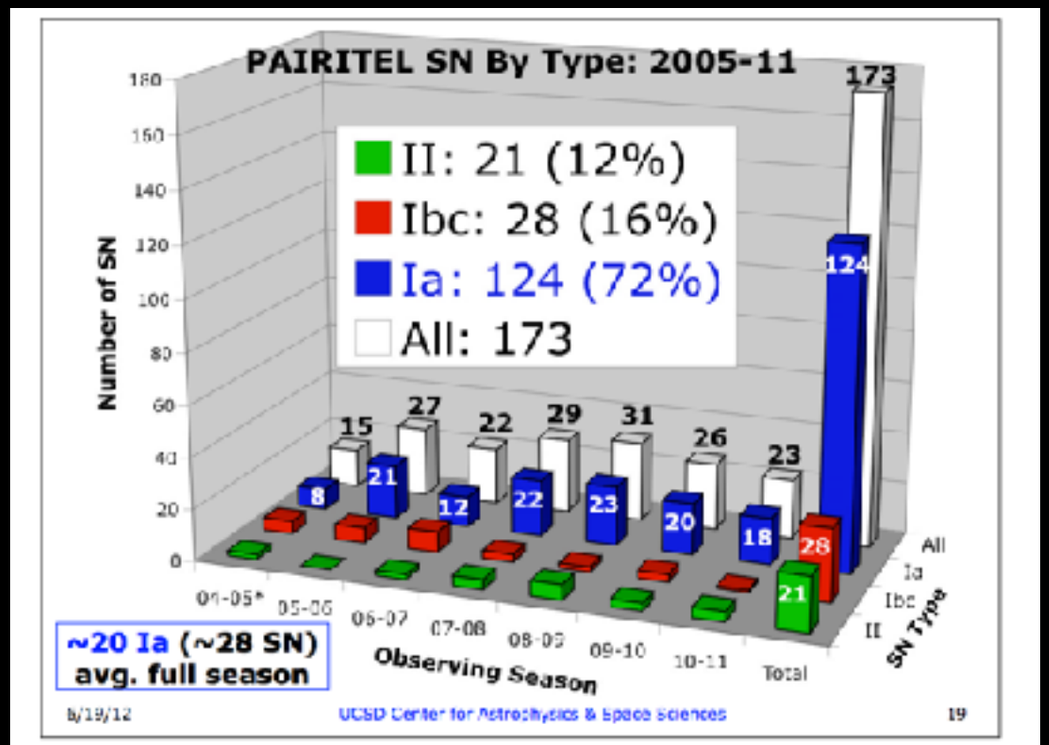
Violent crime rates
1990=100
Philadelphia
Houston
Phoenix
San Diego
Dallas
Los Angeles
New York
120
100
80
60
40
20
0
1990
95
2000
05
08
Sources: Federal Bureau of Investigation; The Economist

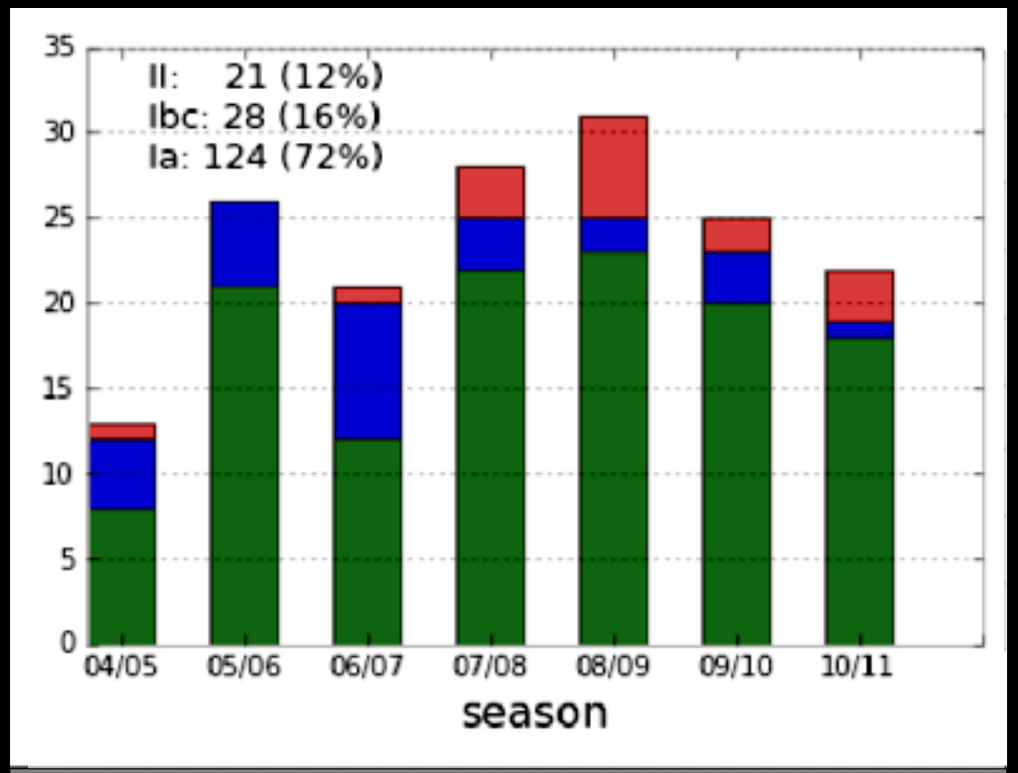
Ambiguity, distortion (misleading), distraction.

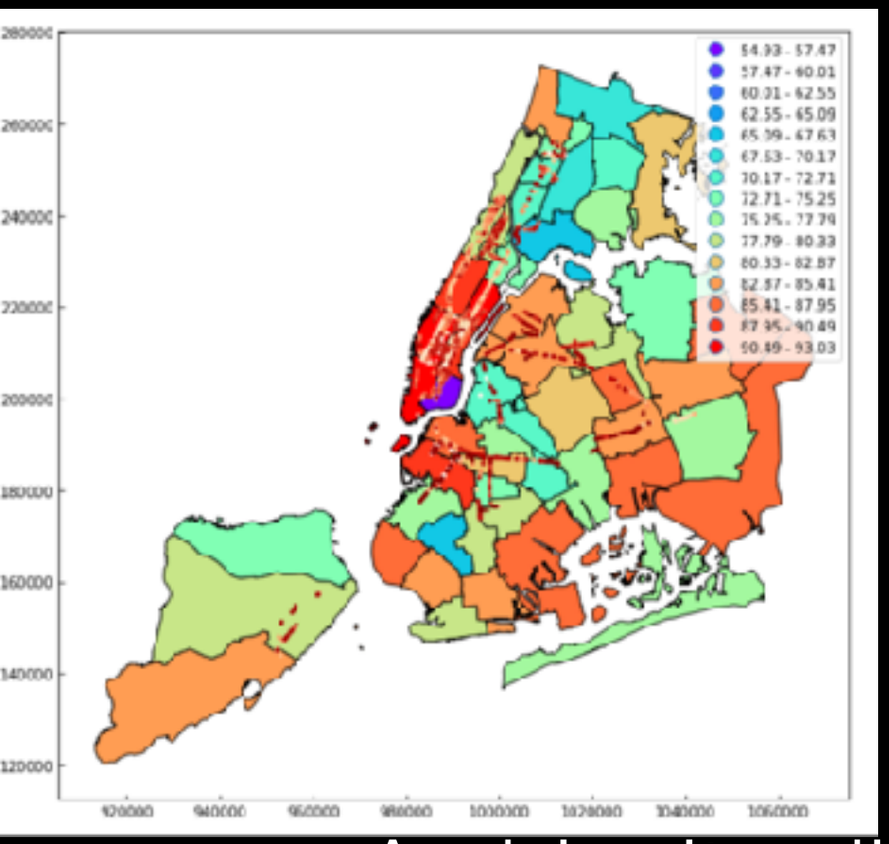




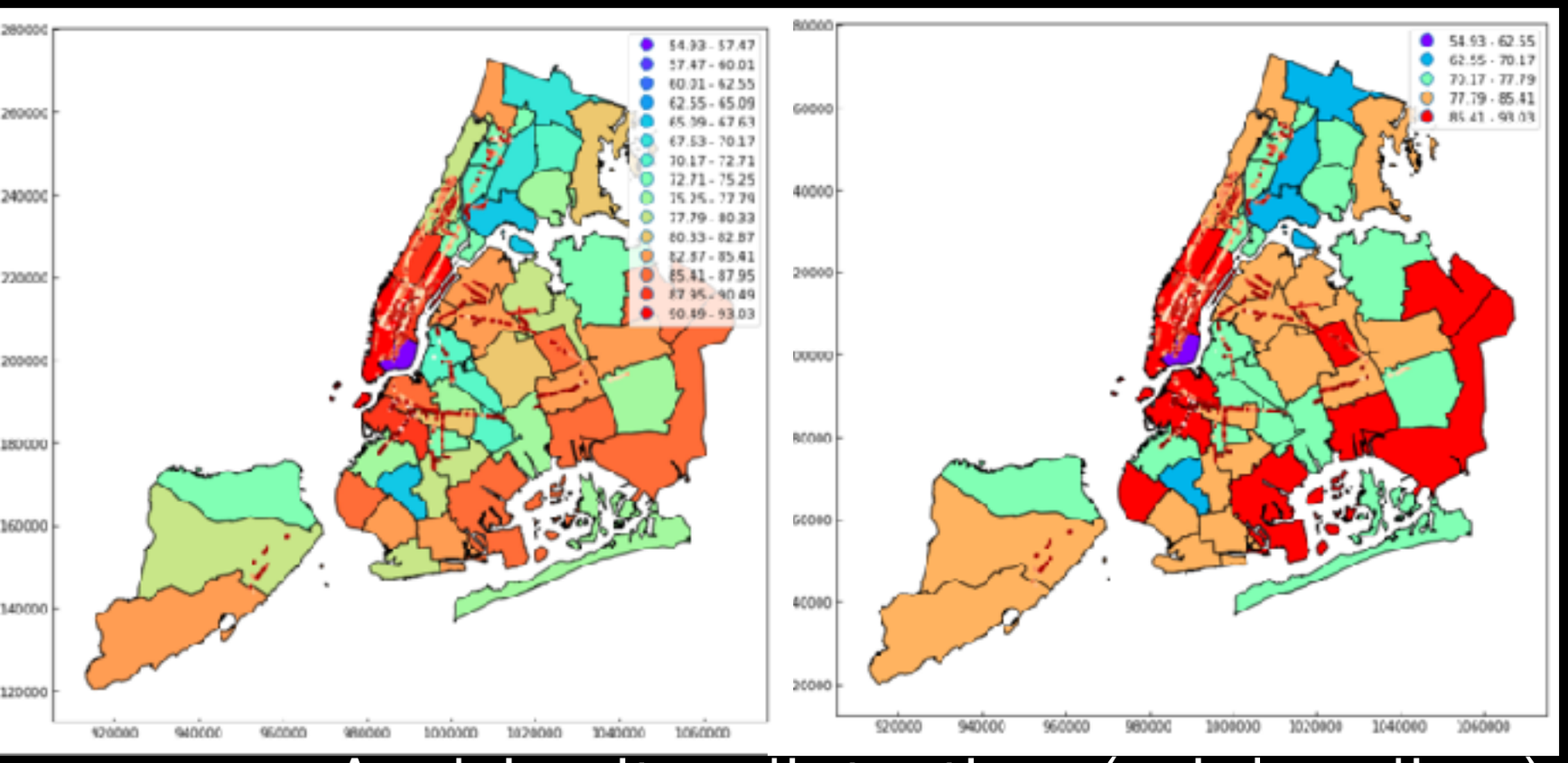
Ambiguity, distortion (misleading), distraction.



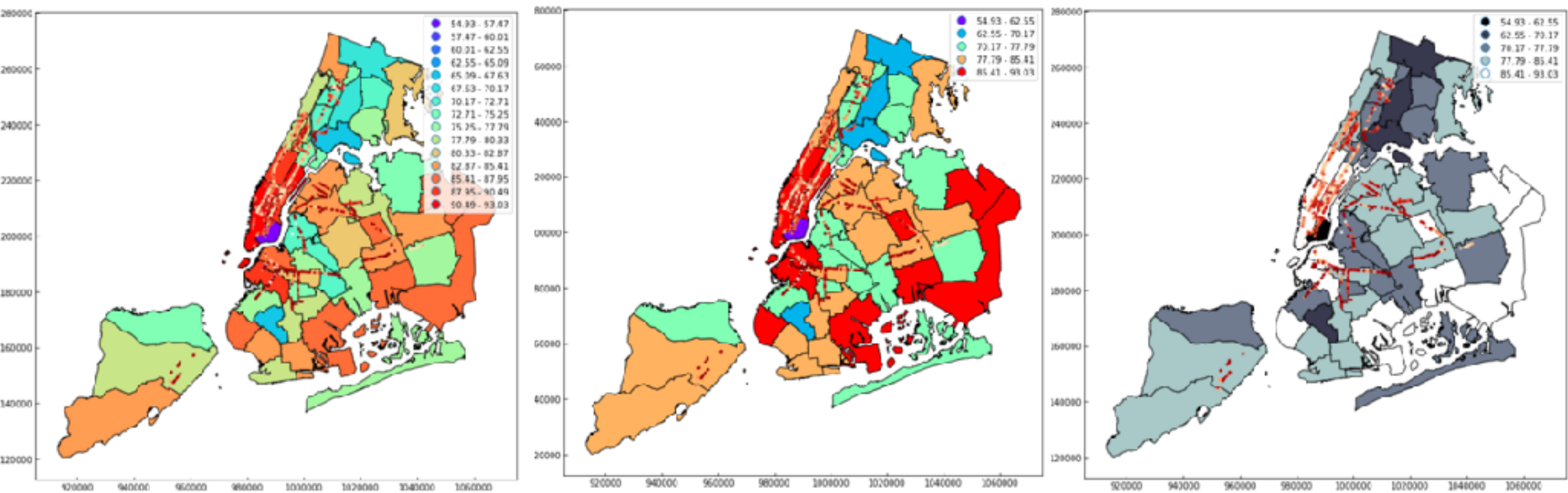




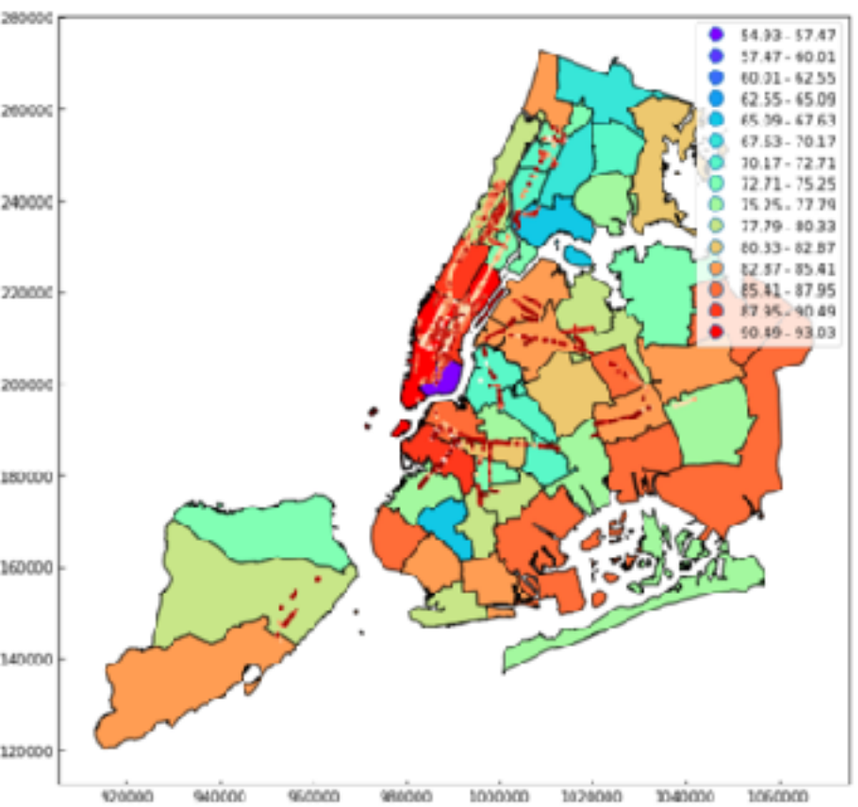
Ambiguity, distortion (misleading), distraction.

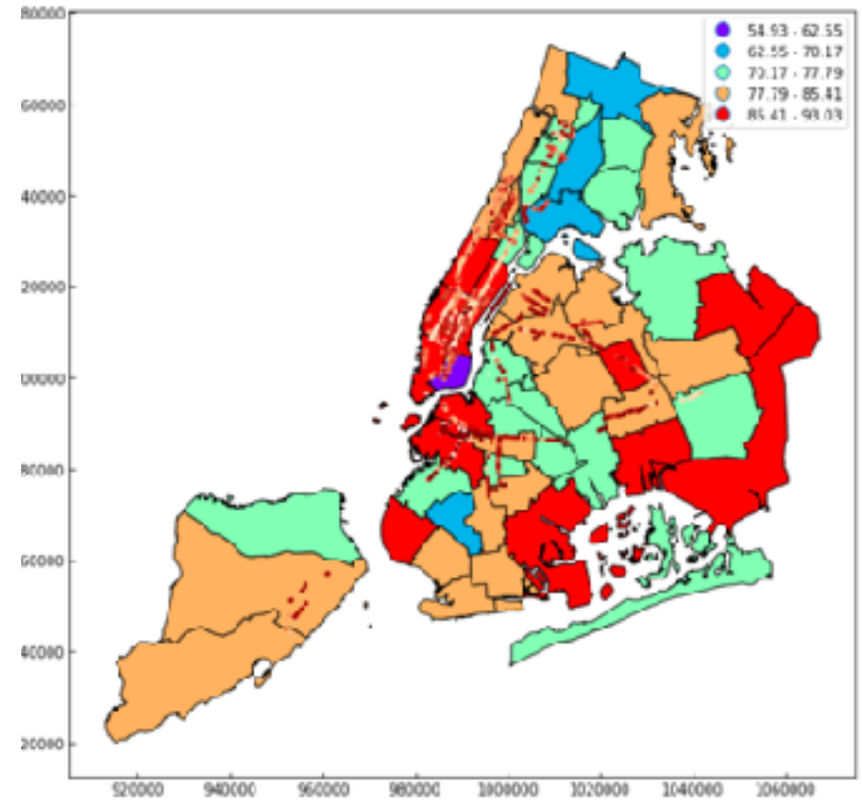


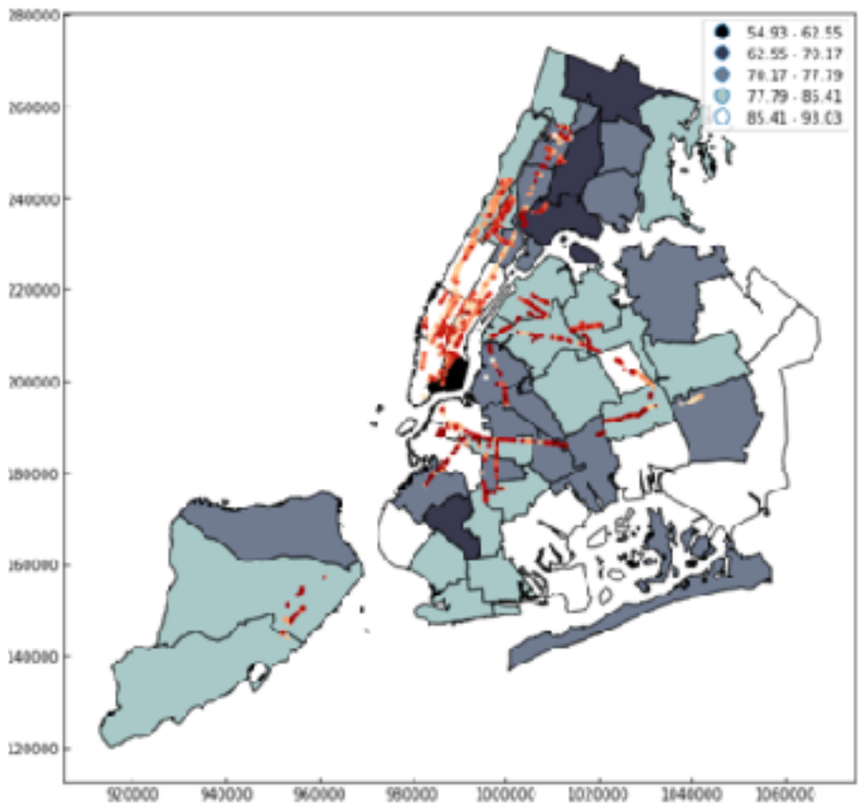
Ambiguity, distortion (misleading), distraction.



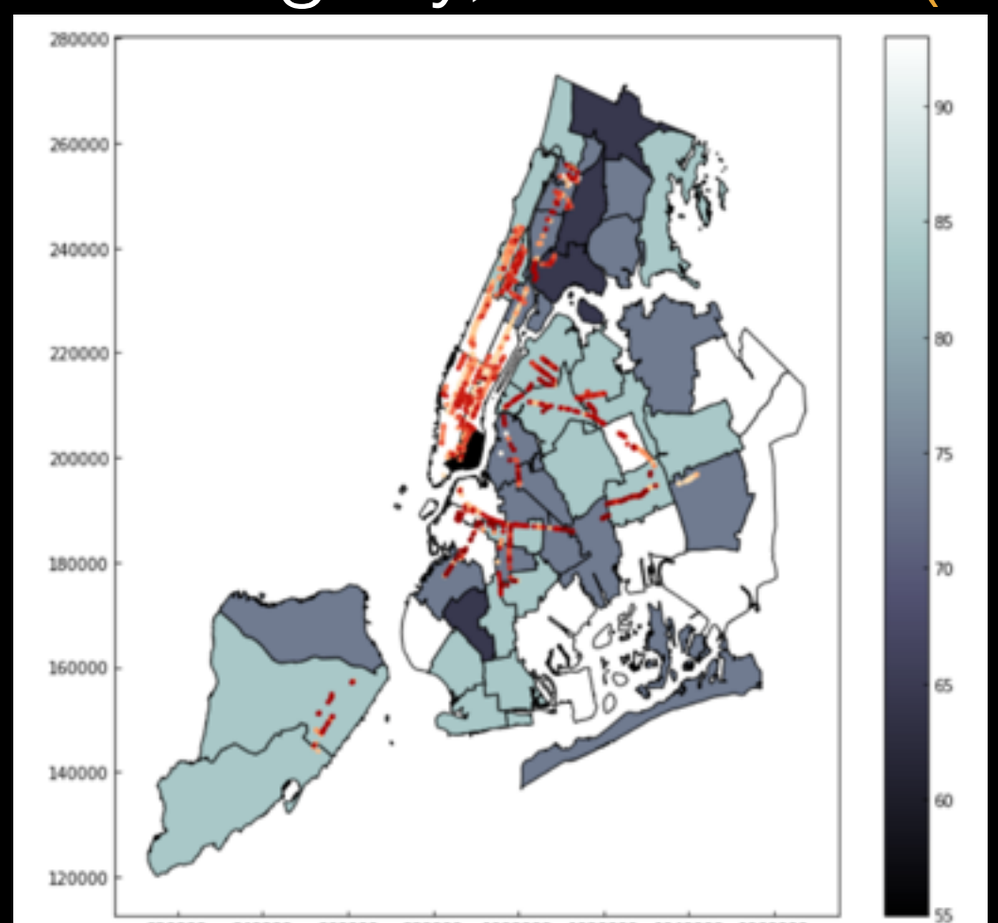
Ambiguity, distortion (misleading), distraction.

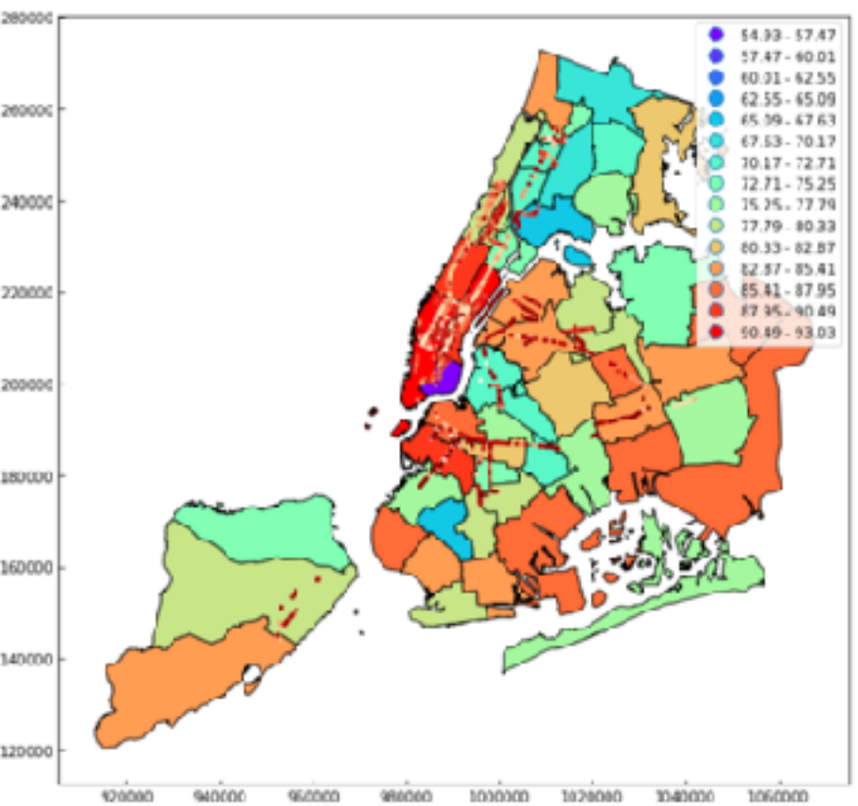


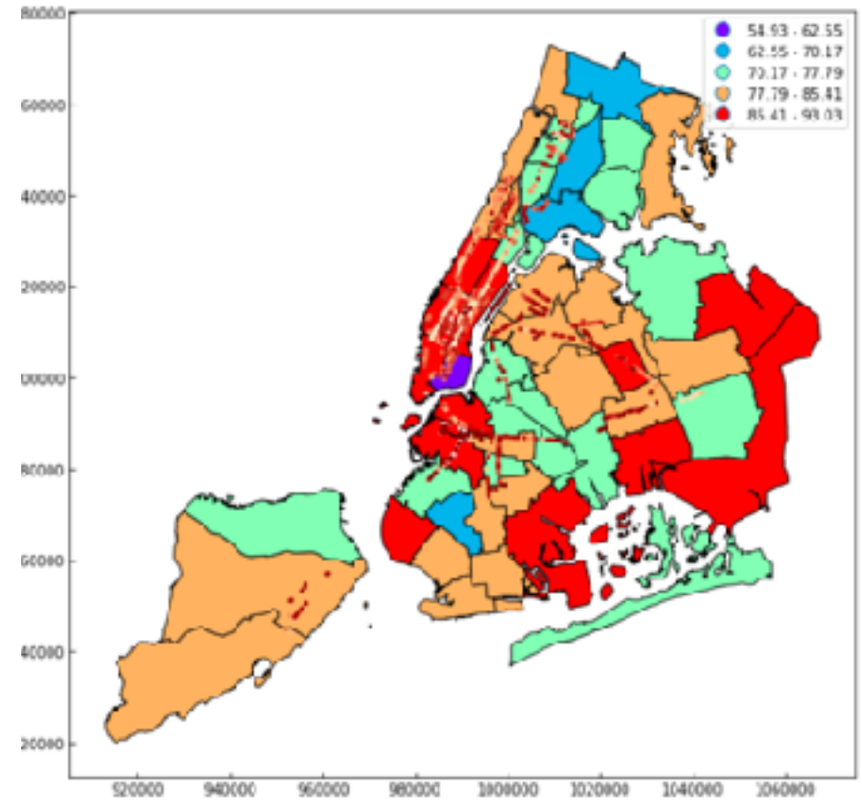


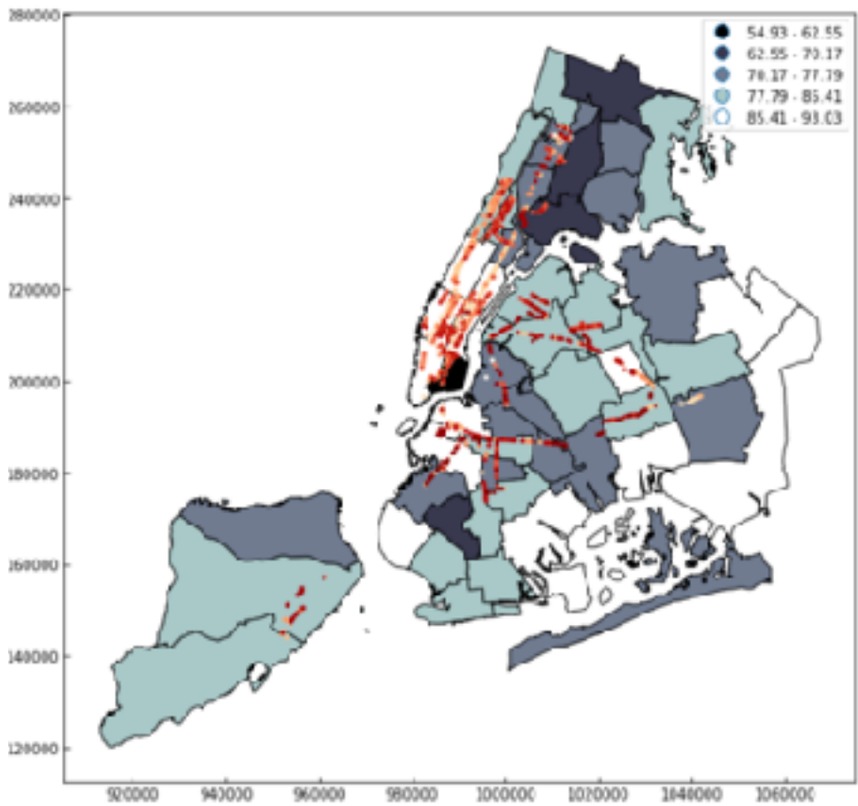


Ambiguity, distortion (misleading), distraction.

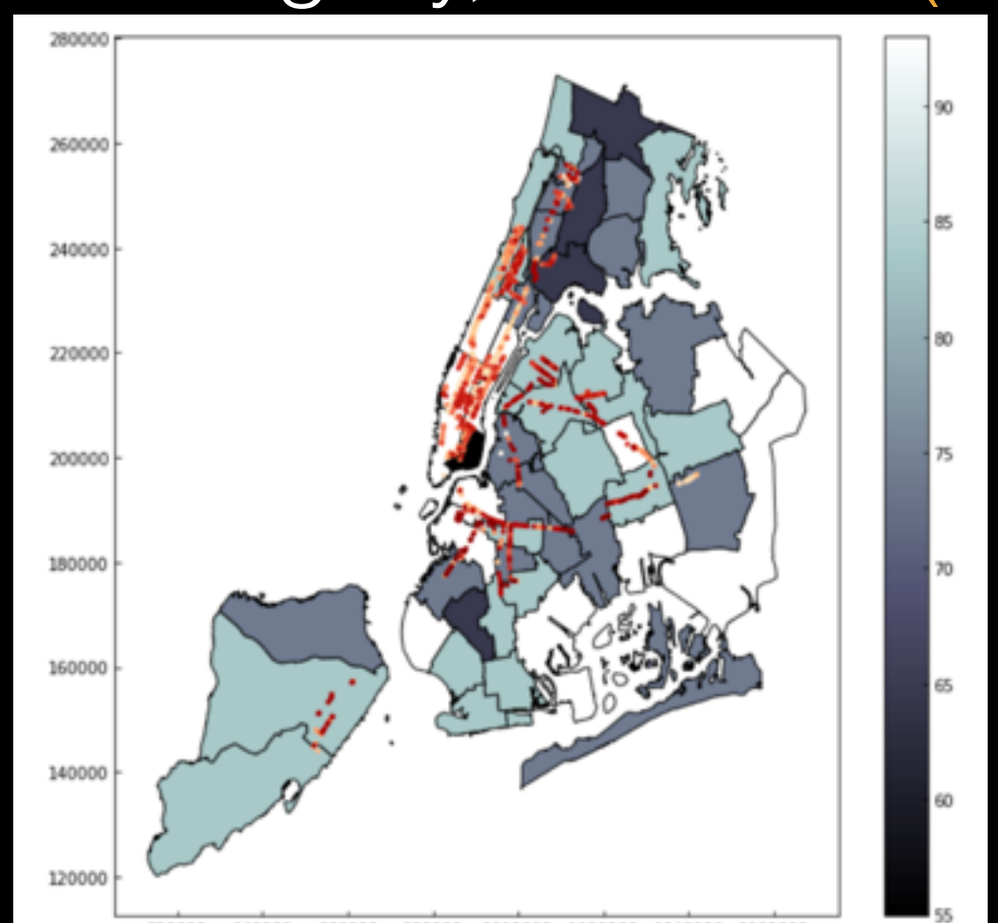


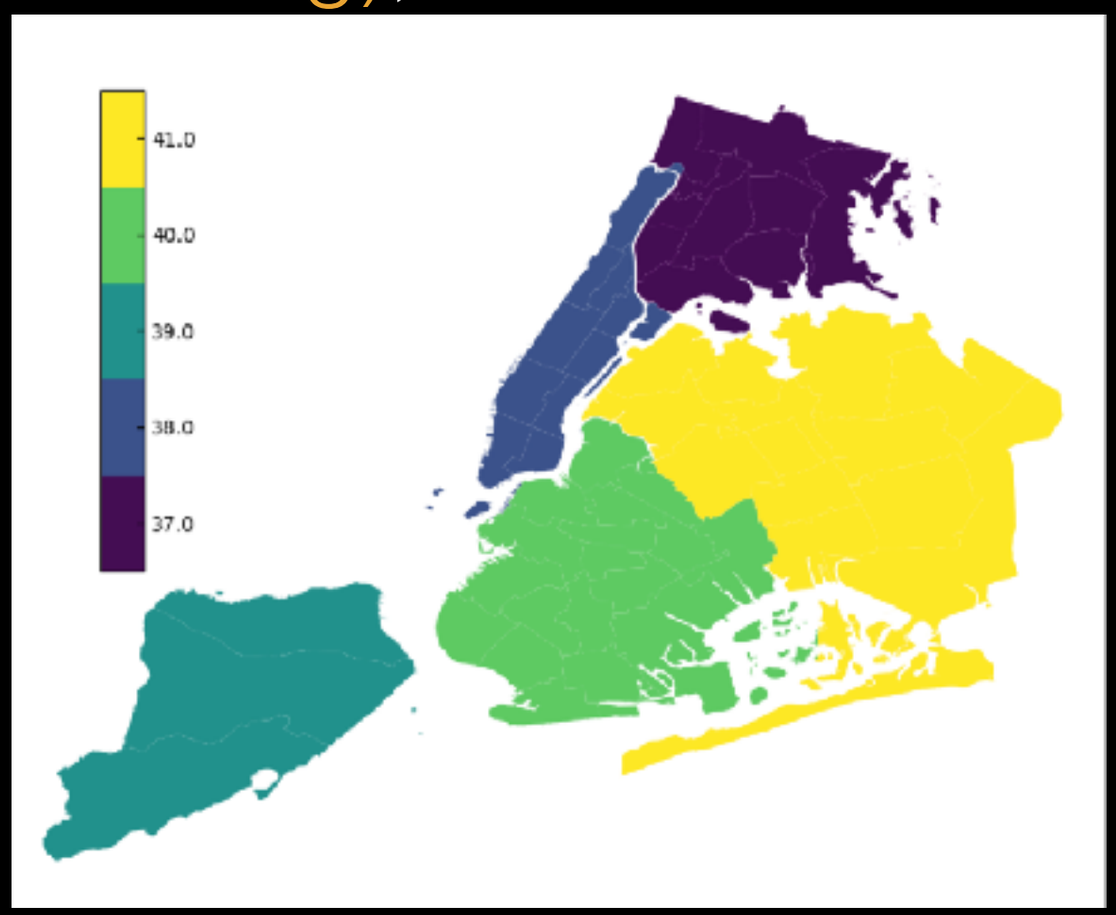






Ambiguity, distortion (misleading), distraction.





- Descriptive data viz
  - Lie with statistics
  - Tufte's rules
- Exploratory data viz
  - Jer Thorp
- Psychophysics
- Esthetics vs(??) functionality
  - color blindness
  - the third dimension
- Interactivity

	1999.1.1	65 months	2004.4.28	low	high		2003.4.28	12 months	2004.4.28	low	high
Euro foreign exchange $	1.1608		1.1907	.8252	1.2858	$	1.1025		1.1907	1.0783	1.2858
Euro foreign exchange ¥	121.32		130.17	89.30	140.31	¥	132.54		130.17	124.80	140.31
Euro foreign exchange £	0.7111		0.6665	.5711	0.7235	£	0.6914		0.6665	0.6556	0.7235

Edward Tufte

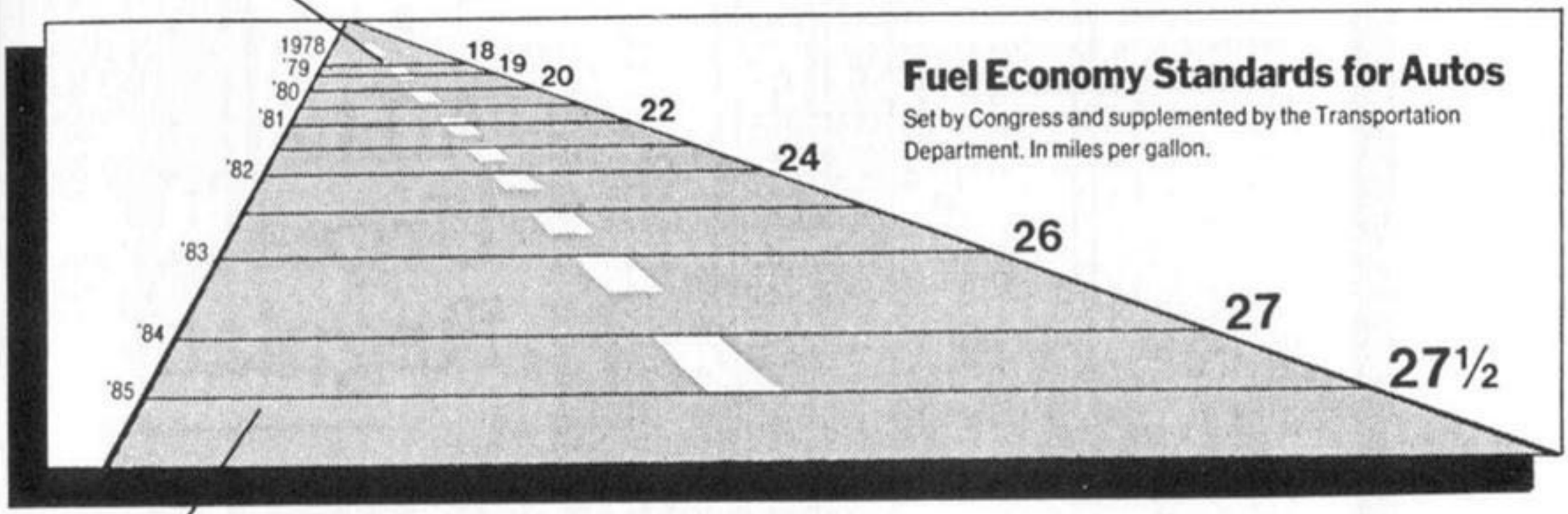
Tufte's rules:

$$\textbf{Lie factor} = \frac{\text{size of the effect in the graphic}}{\text{size of the effect in the data}}$$

Tufte's rules:

$$\text{Lie factor} = \frac{\text{size of the effect in the graphic}}{\text{size of the effect in the data}}$$

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

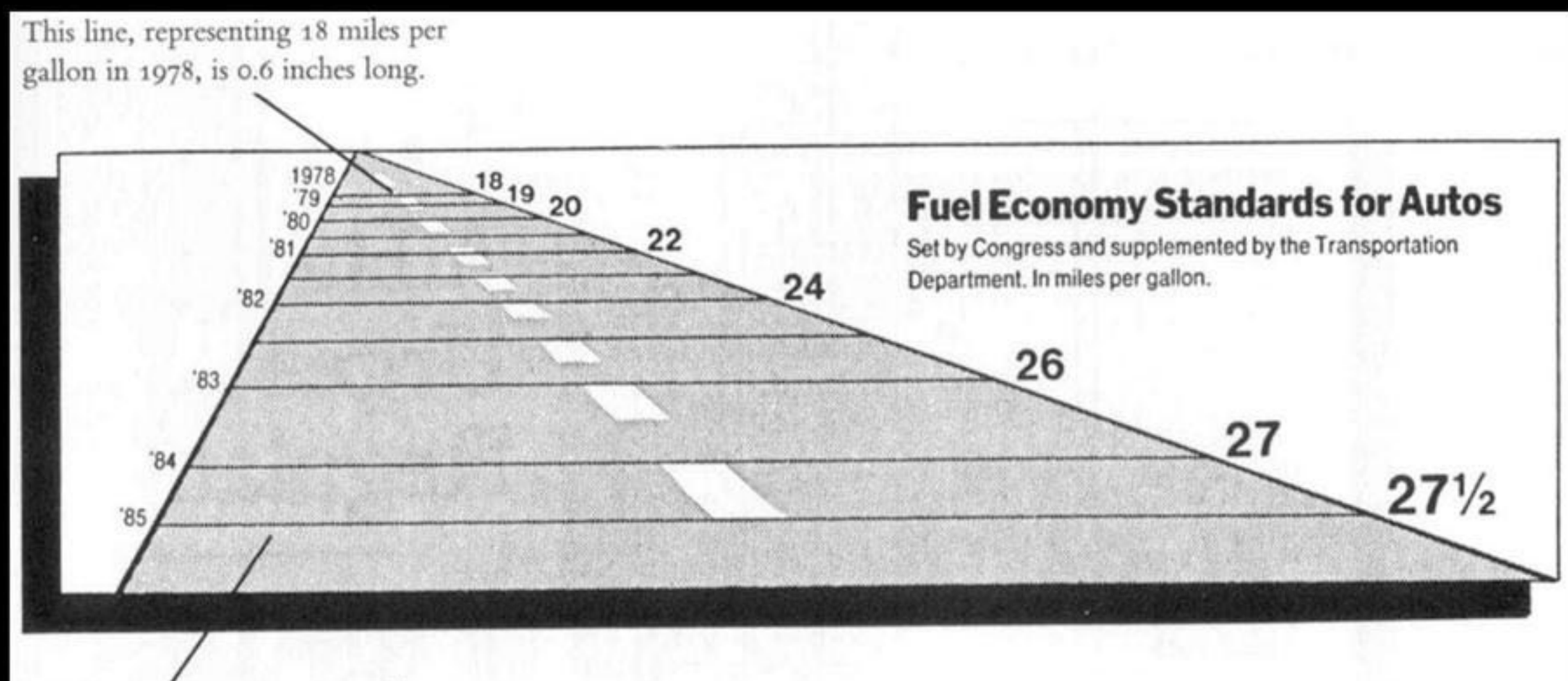


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

Tufte's rules:

$$\text{Lie factor} = \frac{\text{size of the effect in the graphic}}{\text{size of the effect in the data}} = \frac{5.3 / 0.6}{27.5/18} = 5.78$$



New York Times, August 9, 1978, p. D-2.

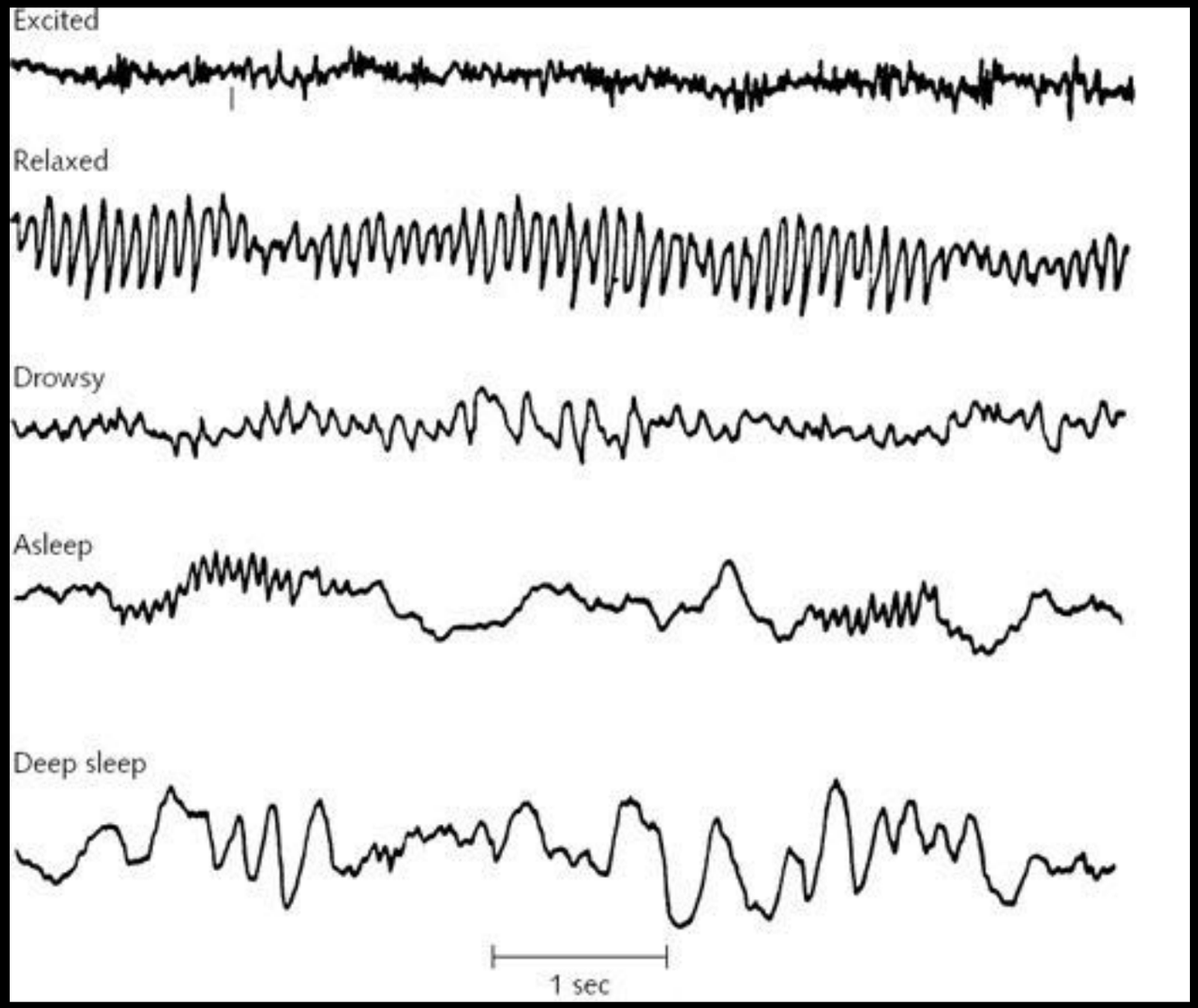
Lie factor = 6

57%
of Europeans are worried their personal information is not safe.
Symantec.

Tufte’s rules:

Data-ink ratio

$$\frac{\text{amount of data}}{\text{amount of ink}}$$



STATE OF **CREATIVITY IN EDUCATION:** AN ADOBE SURVEY

Sample Size: 1014 educators representing 13 countries across Asia Pacific

Attitudes & Aspirations

Current and Desired State of Creativity

62% Educators strongly felt that they should be creative regardless of the subjects they teach

35% Strongly felt fostering creativity is educators' primary responsibility

21% Strongly felt that parents are primarily responsible for creativity

58% Ideally wanted to spend time fostering creativity in the classroom

45% Educators spent their time fostering creative skills in their classroom in the last one year

Barriers & Enablers

What Hinders and Helps Creativity

57% Are hamstrung by an education system that is not geared towards creativity

49% Blame it on lack of resources

29% Blame it on lack of time

36% Believe the biggest barrier to creativity is a system that is heavily reliant on testing and assessment

43% Felt the current education system was either outdated or restrictive or both

41% Educators believe tools and training for educators are the most critical need to promote creativity

Tools & Techniques

Methods to Foster Creativity

85% Respondents felt that technology and digital tools play an important role

34% Indicated that they were very prepared to adapt their teaching methodology to leverage digital tools

Innovation & Growth

Impact on Country's Success

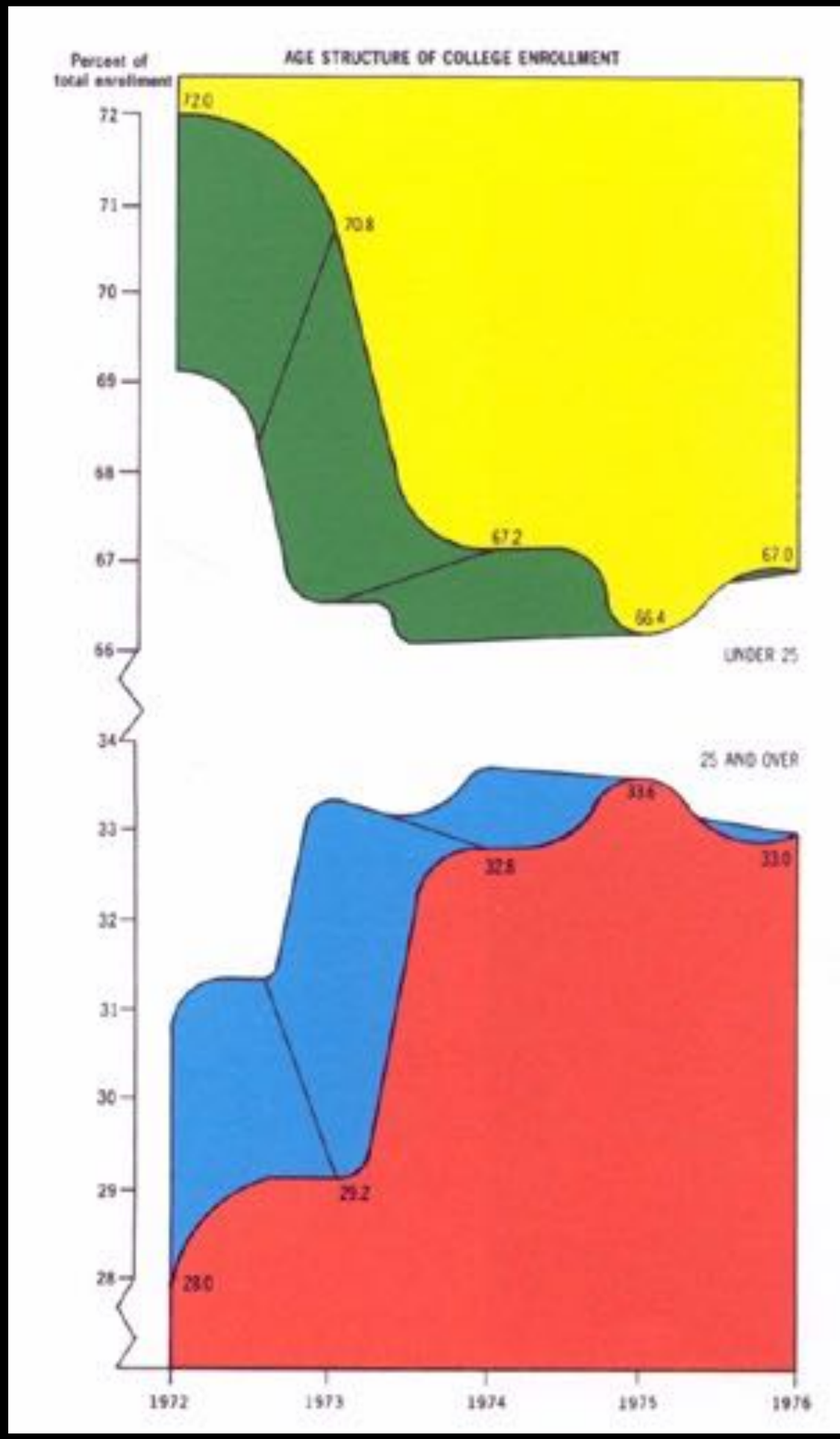
8.4 Rating on a scale of 1 to 10; on how important it was to infuse creativity in education to ensure their country's long-term success

5.0 Rating on a scale of 1 to 10; on the efficiency of the current education system in developing a new generation of innovators

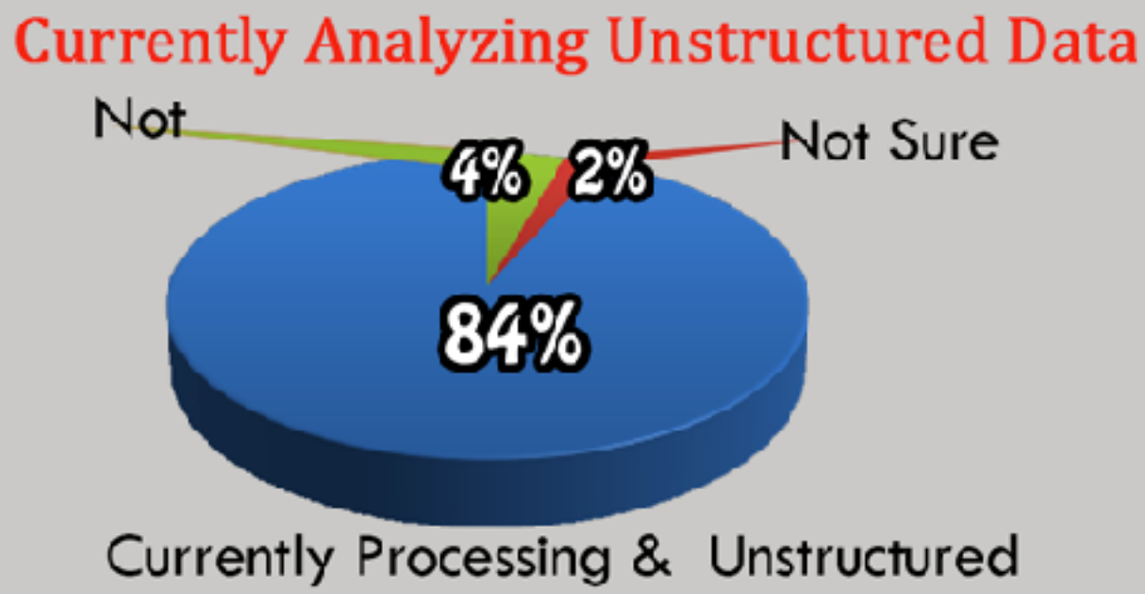
Tufte's rules:

Chart Junk

the excessive and unnecessary use of graphical effects



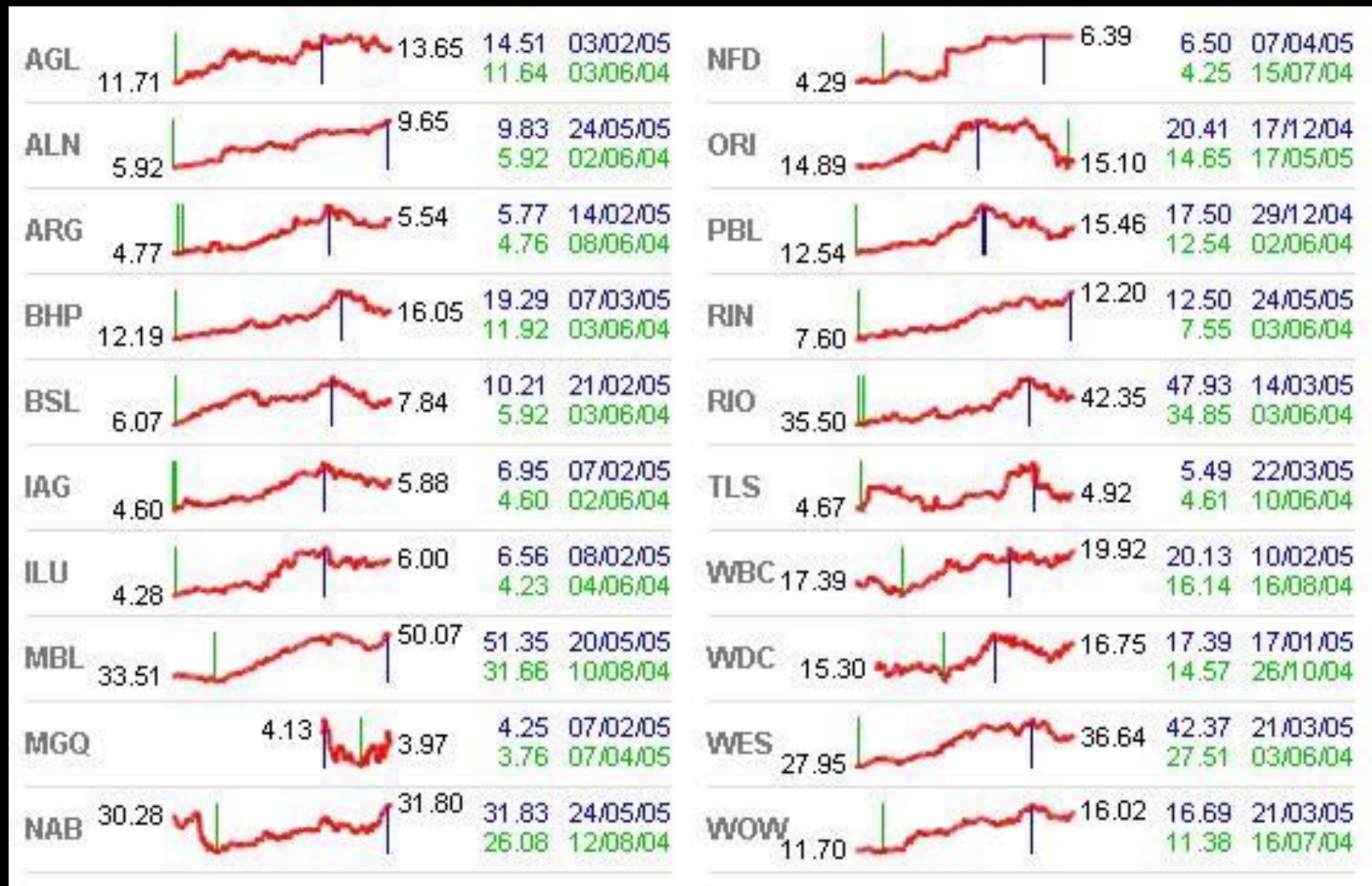
4 The Rise of Unstructured and Semistructured Data Analytics



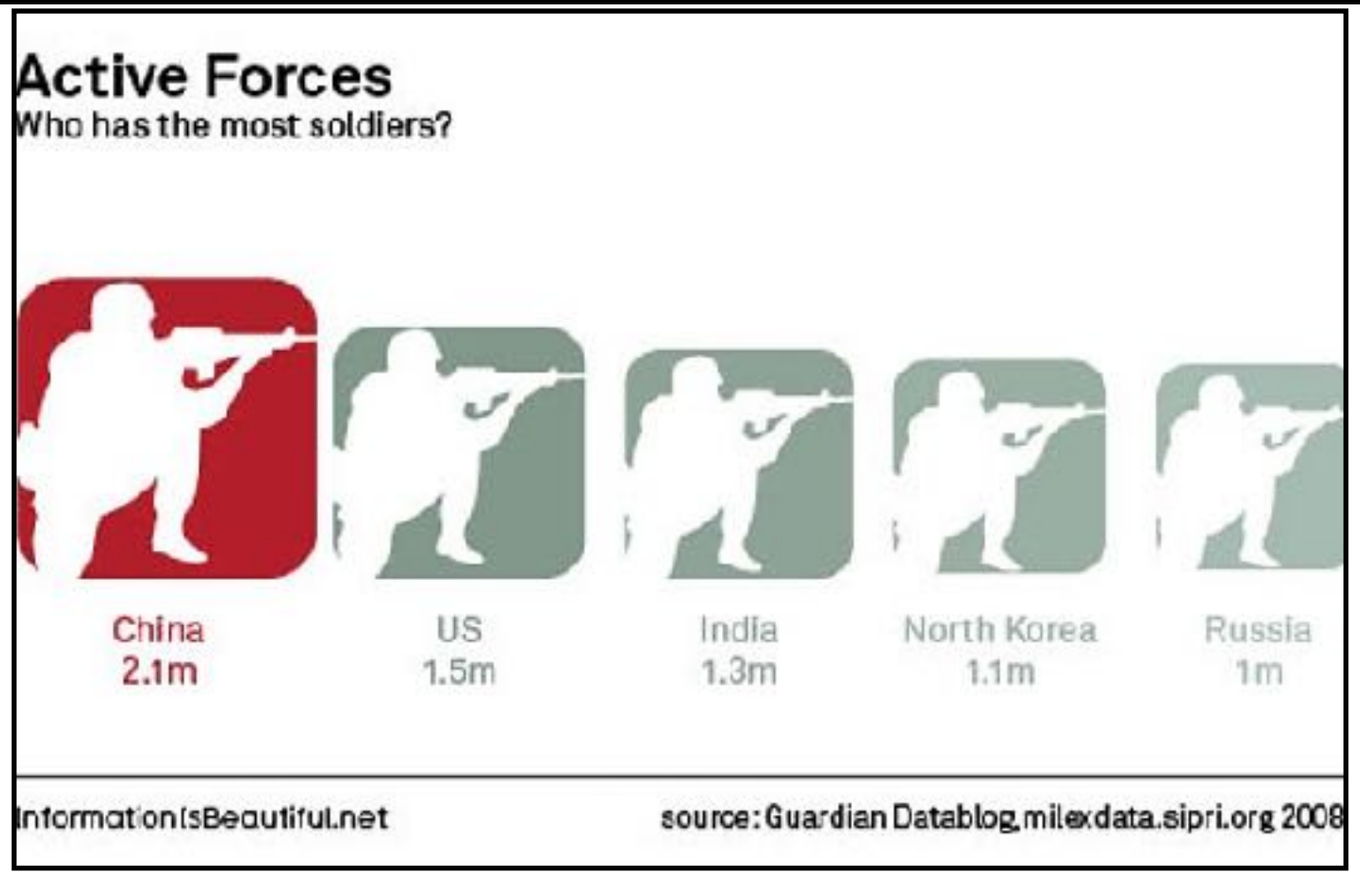
The 'Peer Research – Big Data Analytics' survey clearly reports that there is a huge growth when it comes to unstructured & semistructured data analytics.

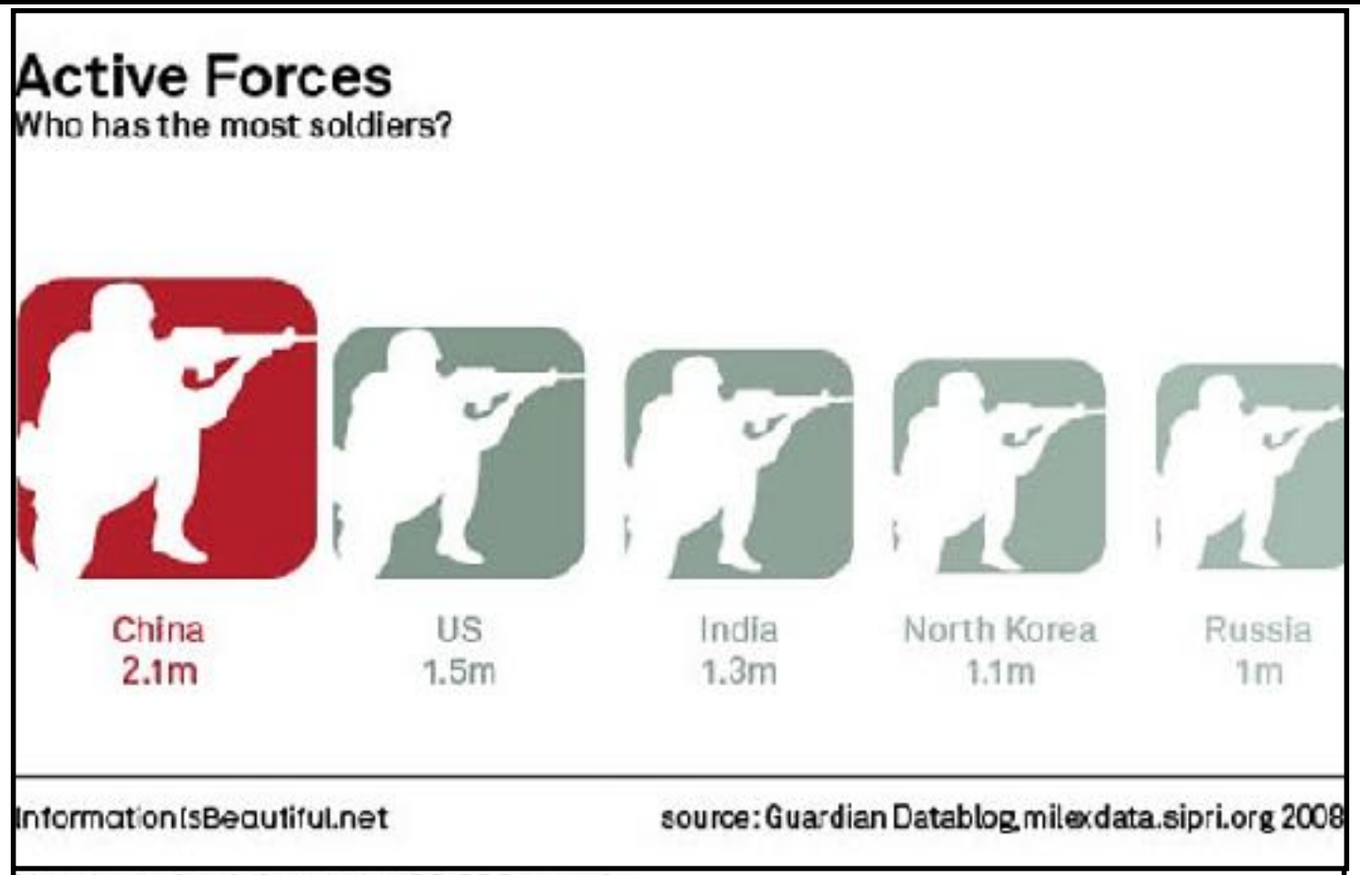
Tufte's rules:

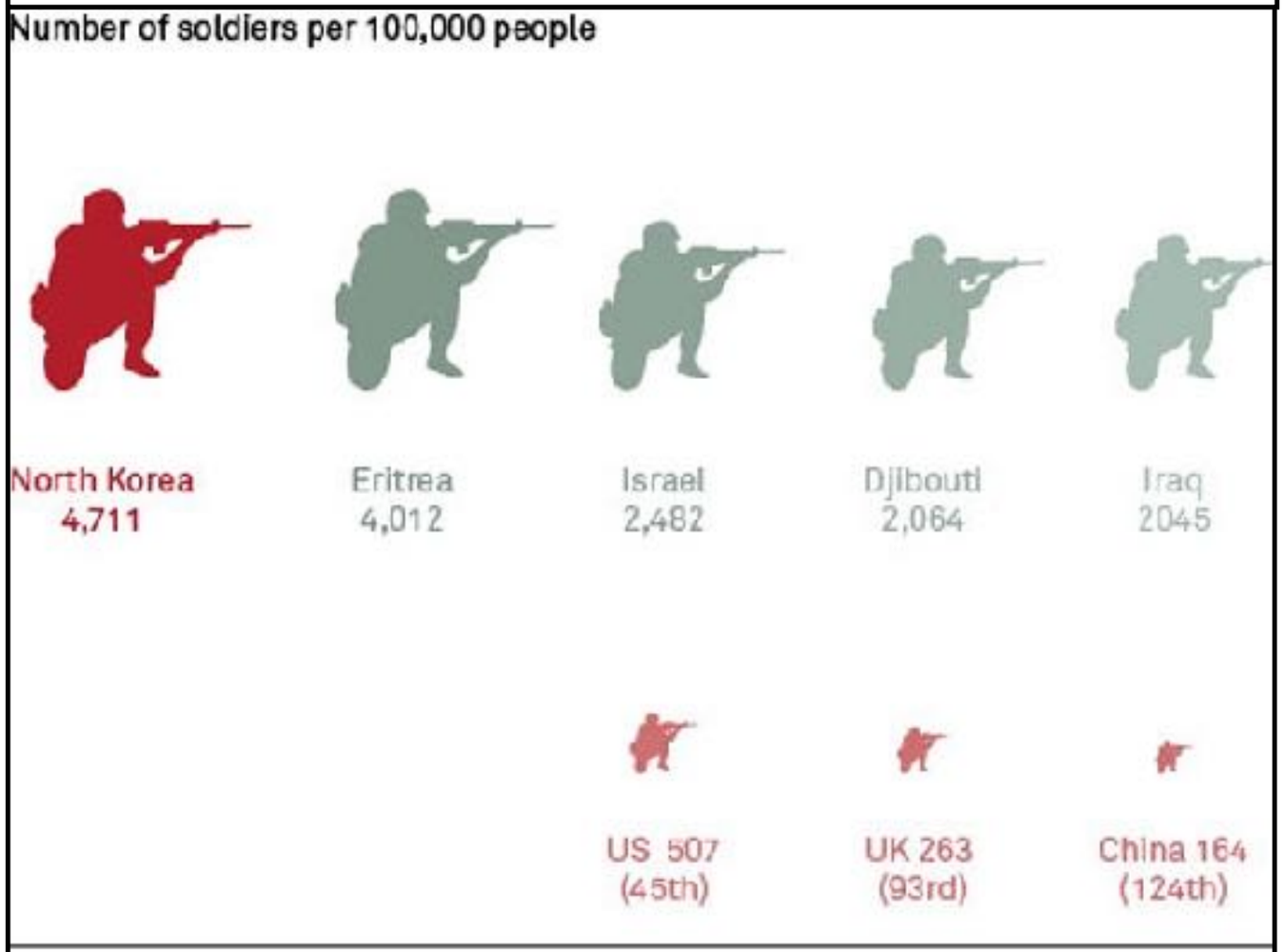
Small Multiples encourage comparison



sparkline graph

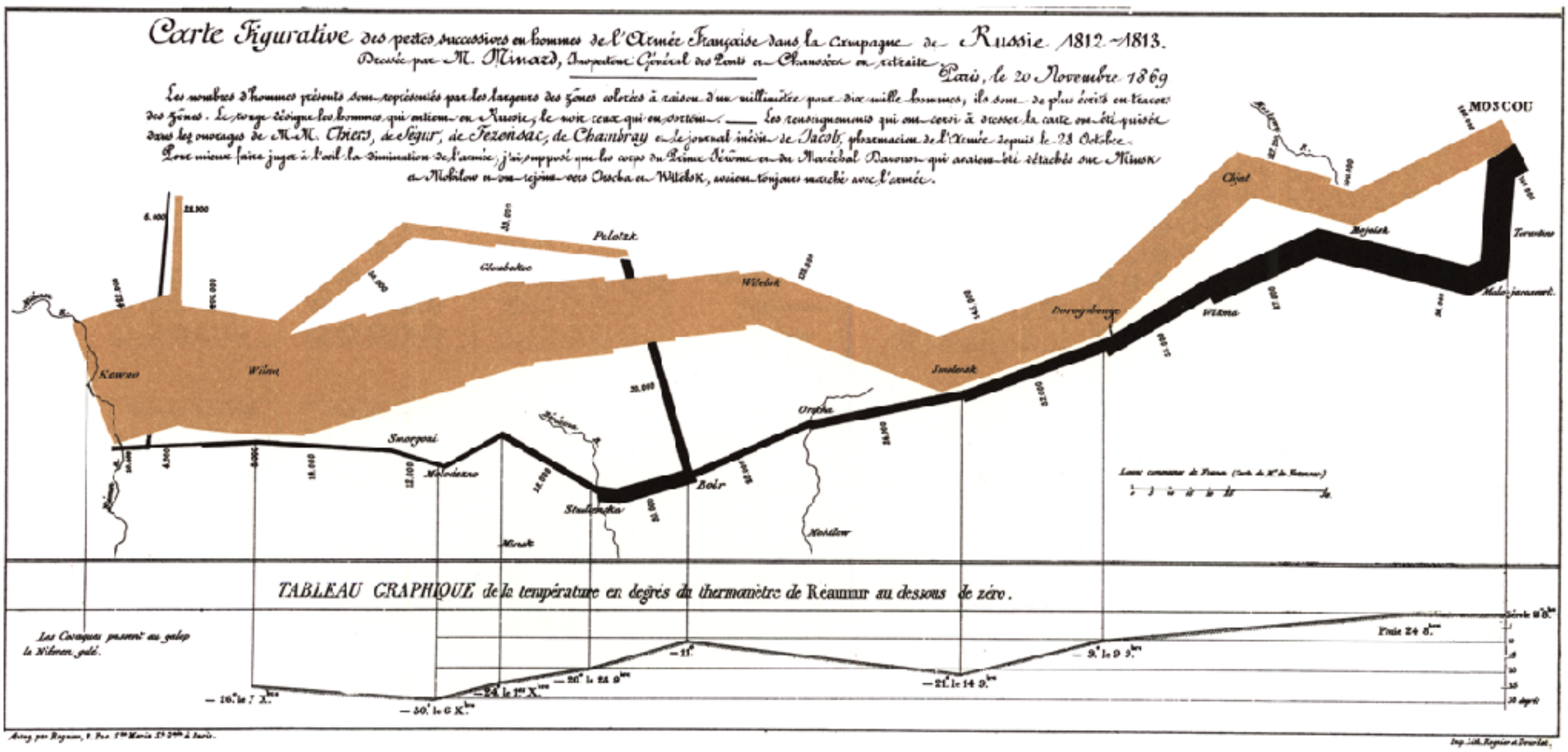

Active Forces
Who has the most soldiers?
China
2.1m
US
1.5m
India
1.3m
North Korea
1.1m
Russia
1m
InformationIsBeautiful.net
source: Guardian Datablog, milexdata.sipri.org 2008


Active Forces
Who has the most soldiers?
China
2.1m
US
1.5m
India
1.3m
North Korea
1.1m
Russia
1m
InformationIsBeautiful.net
source: Guardian Datablog, milexdata.sipri.org 2008


Number of soldiers per 100,000 people
North Korea
4,711
Eritrea
4,012
Israel
2,482
Djibouti
2,064
Iraq
2045
US 507
(45th)
UK 263
(93rd)
China 164
(124th)
InformationIsBeautiful.net
source: Guardian Datablog, milexdata.sipri.org 2008

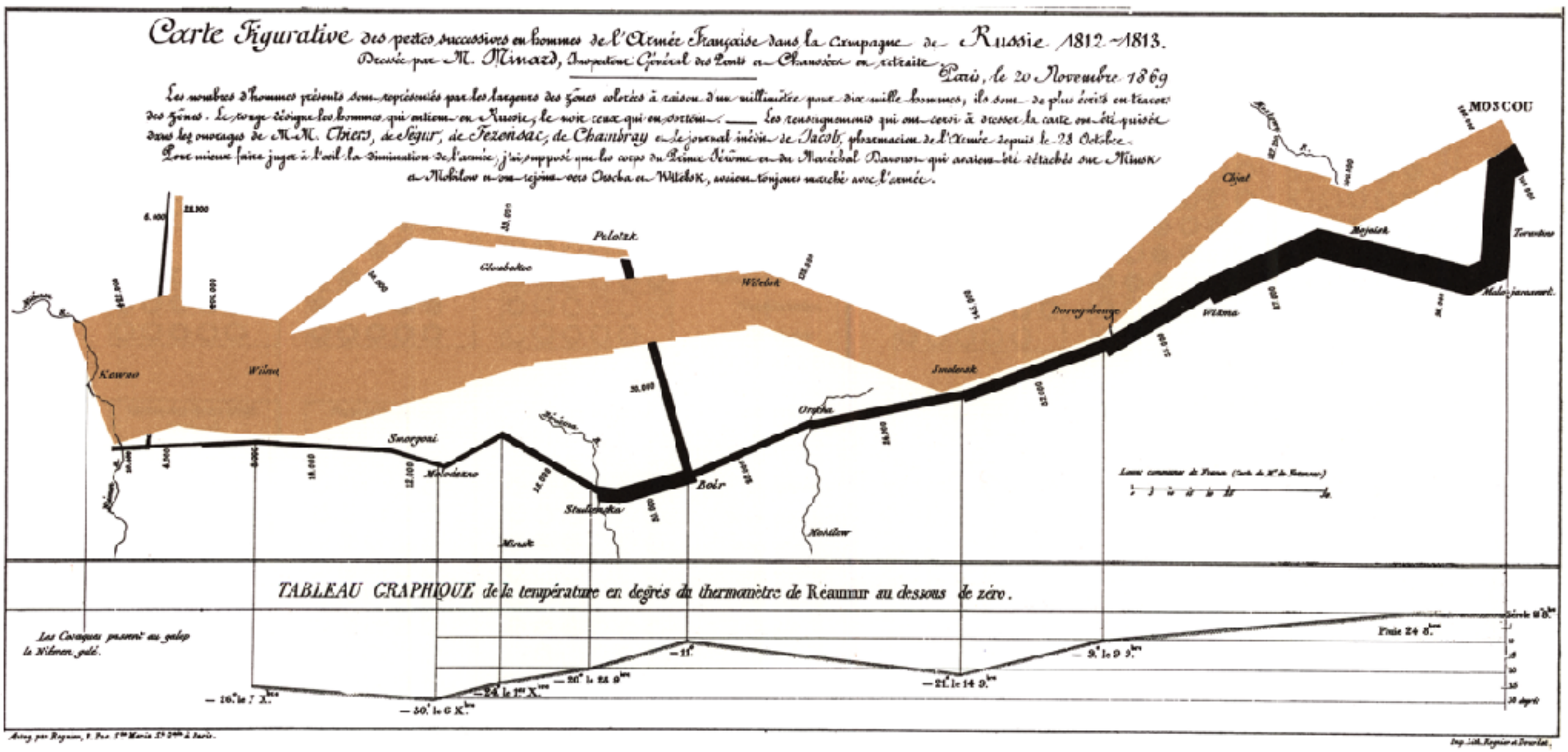
Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869.
The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October.
In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jèrôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.



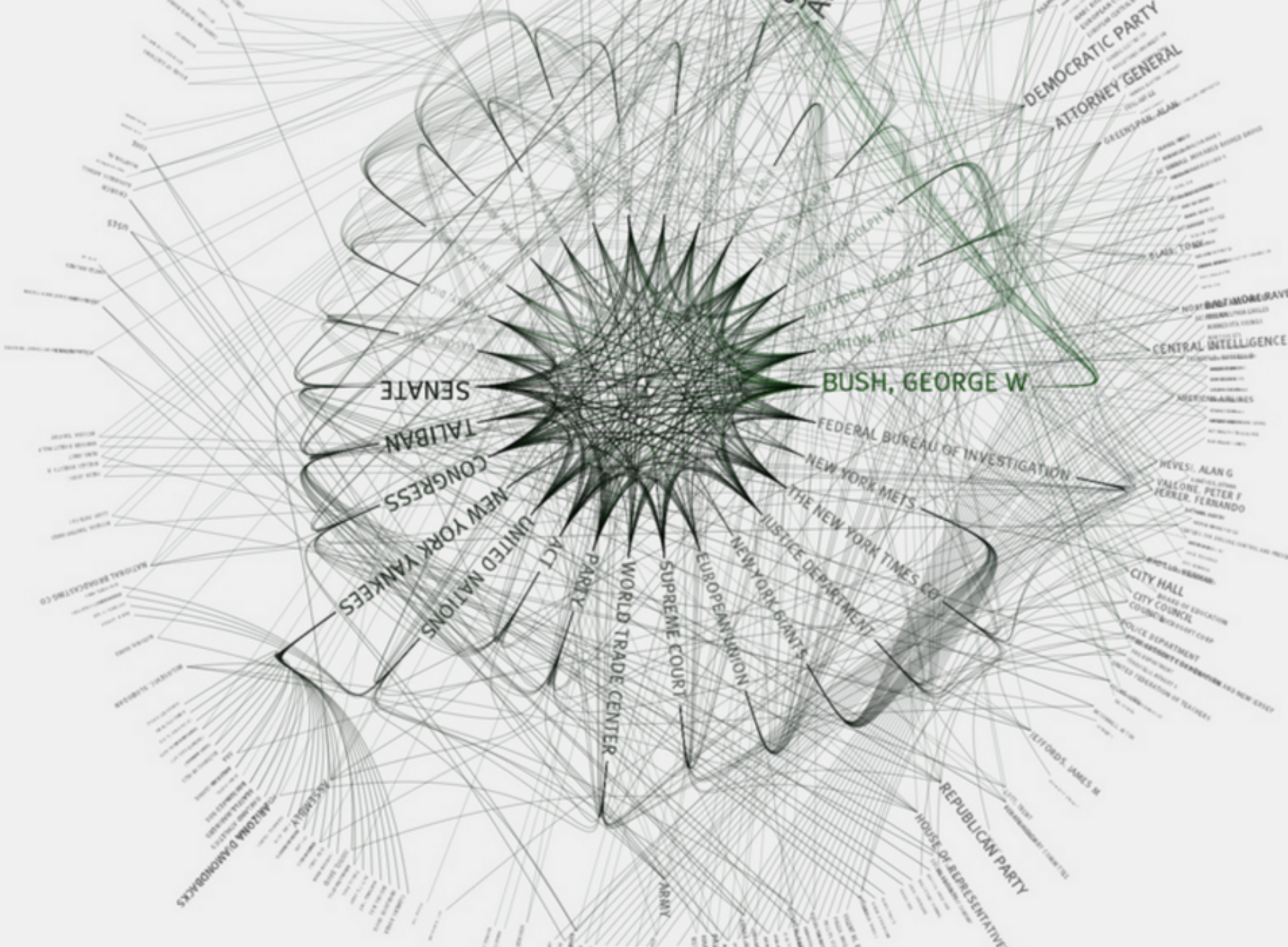
Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869.
The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October.
In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jèrôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.



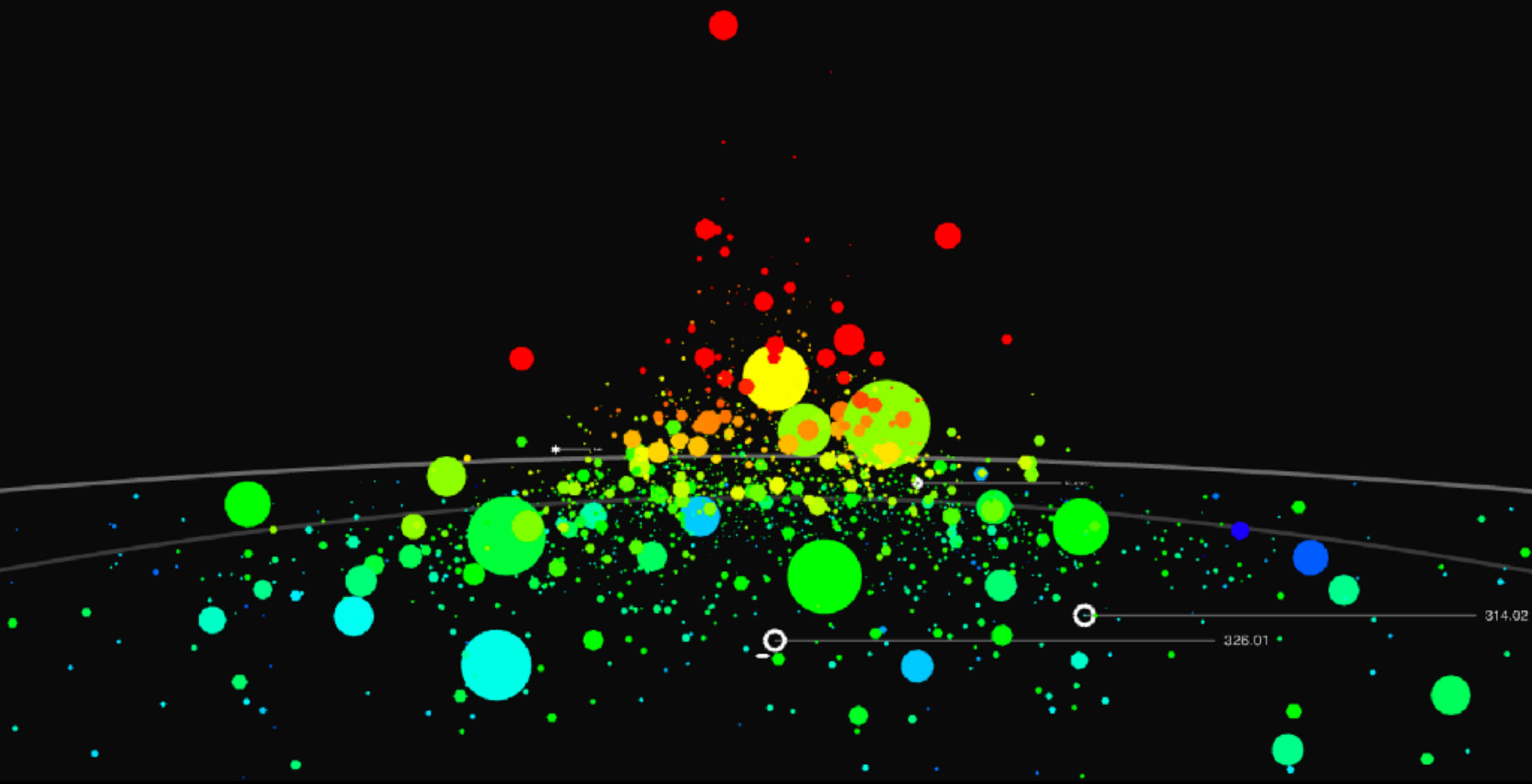
1. The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented
2. Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. *Write out explanations of the data on the graph itself.* Label important events in the data.
3. Show data variation, not design variation.
4. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
5. The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data ($ND_{graph} <= ND_{data}$).
6. Graphics must not quote data out of context.

Jer Thorp



https://www.flickr.com/photos/blprnt/3289727424/in/album-72157614008027965/

http://nytlabs.com/projects/cascade.html



http://blog.blprnt.com/blog/blprnt/data-in-an-alien-context-kepler-visualization-source-code

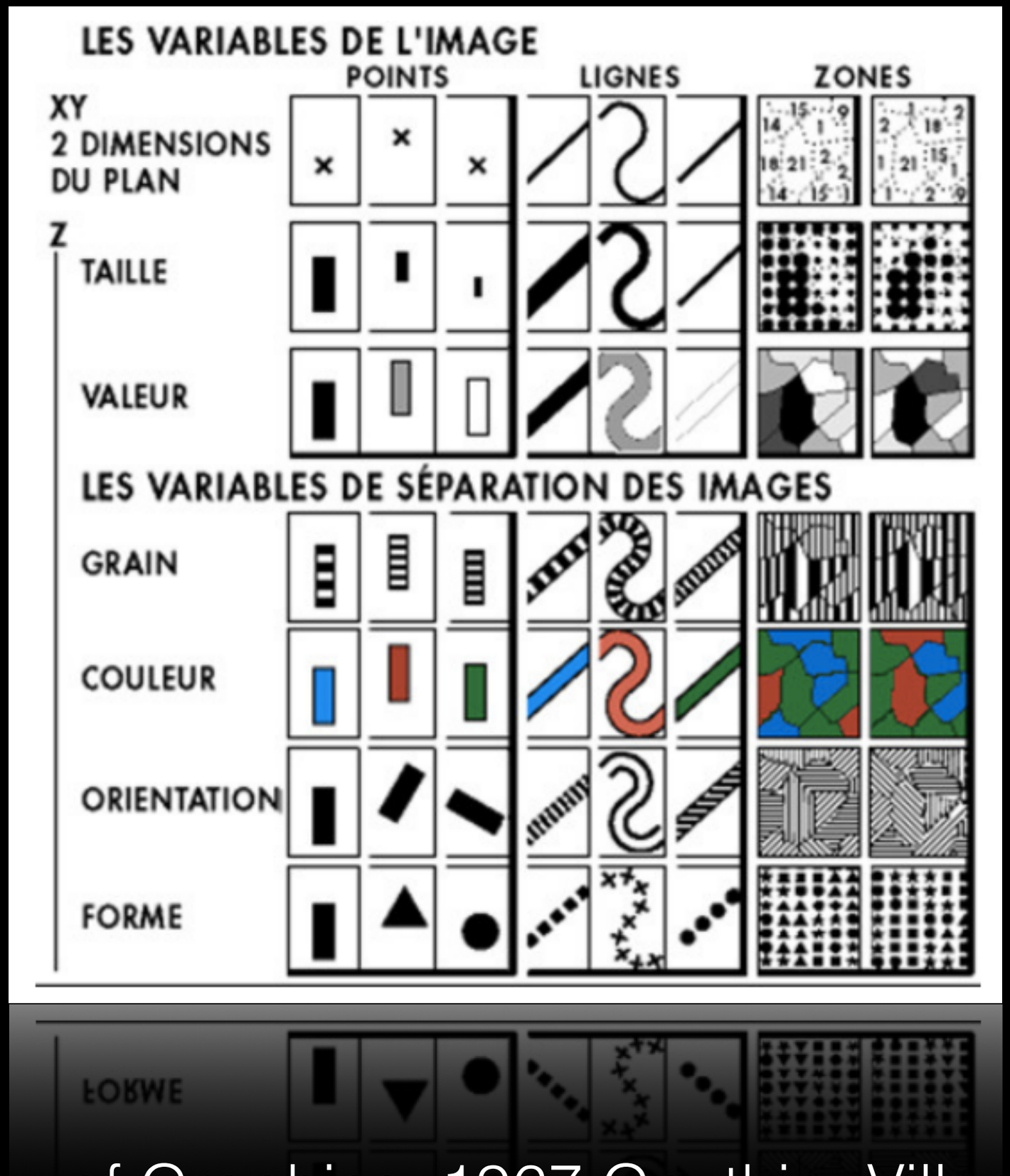
making sparklines!

jupyter

Remove to improve

http://i.imgur.com/RzYaLZg.gif

Graphical Vocabulary



Jacques Bertin: Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS

point line area

position

size

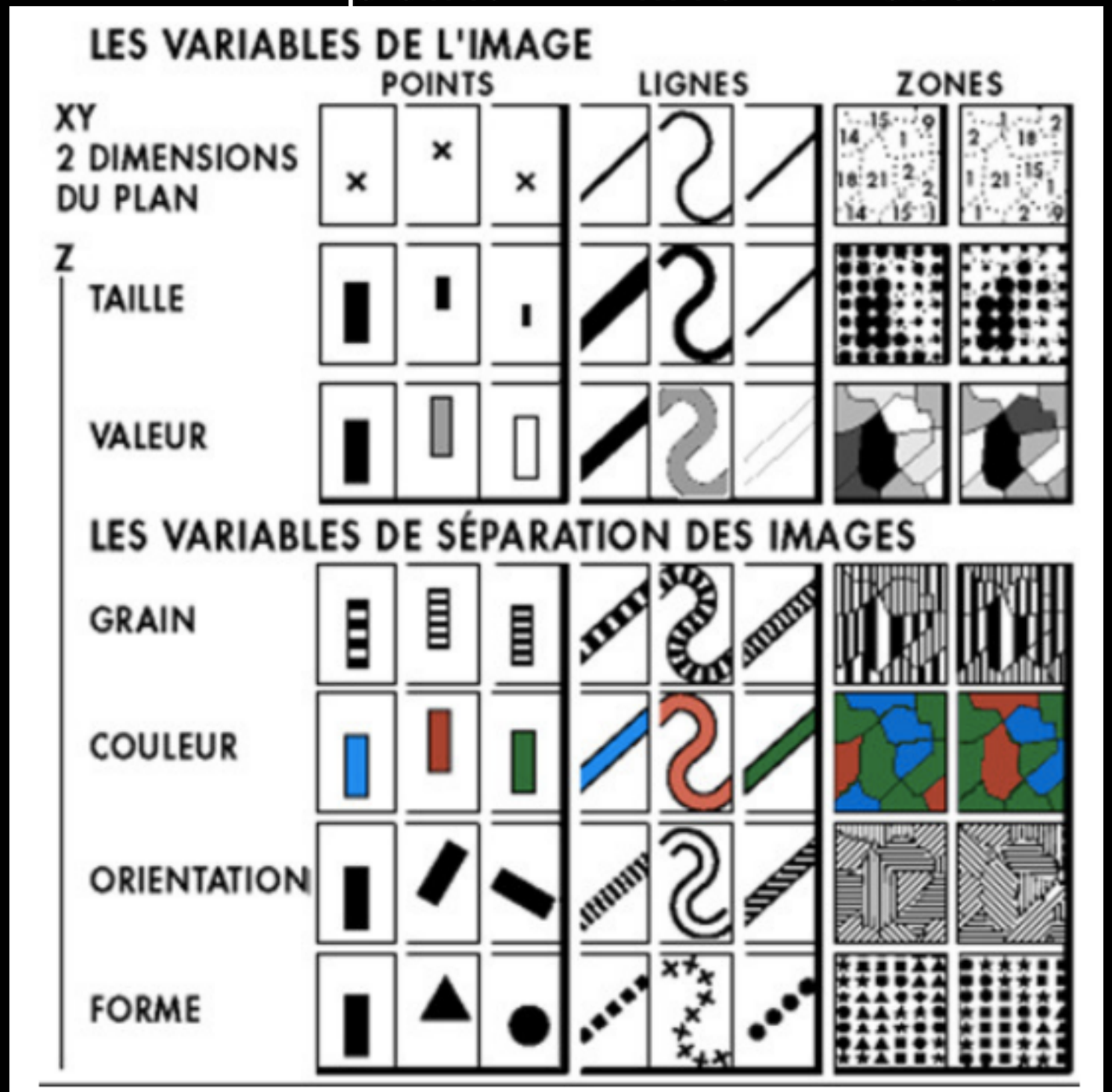
intensity

texture

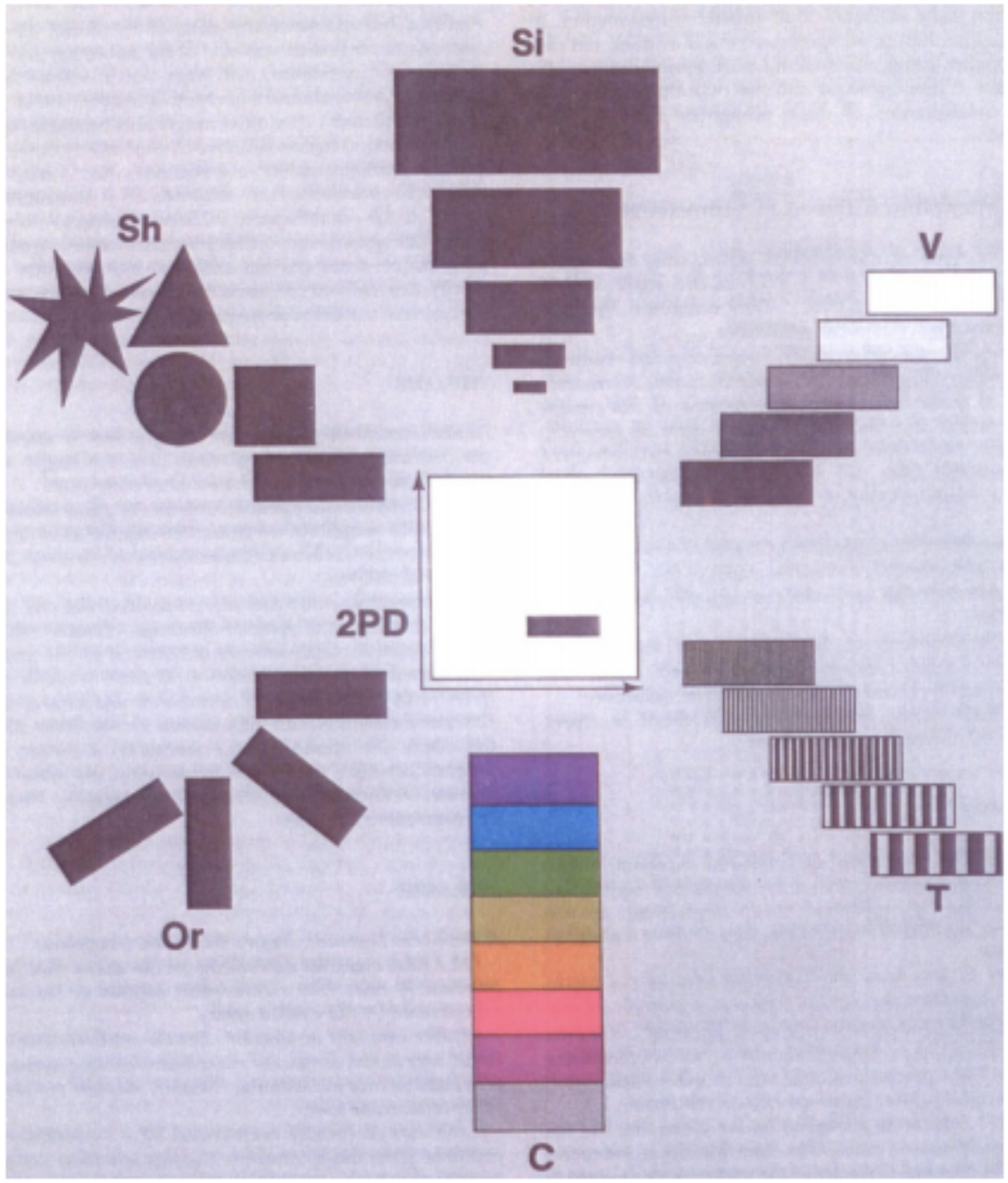
color

orientation

shape



Jacques Bertin: Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS



- Size
- Value (Density)
- Texture
- Color
- Orientation
- Shape

- 3D
- Animation/Time

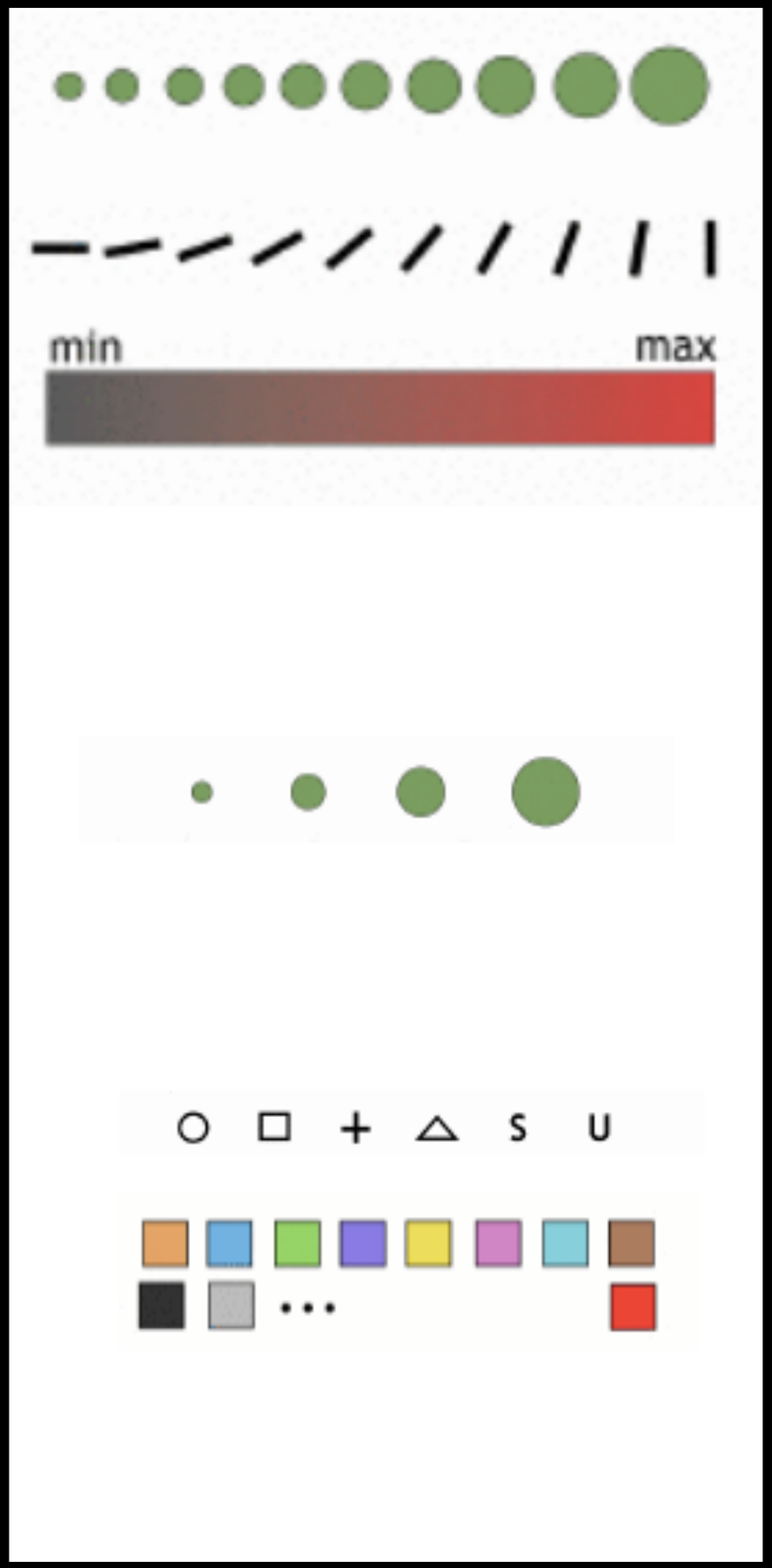
Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1983 UW Press

Data types

continuous

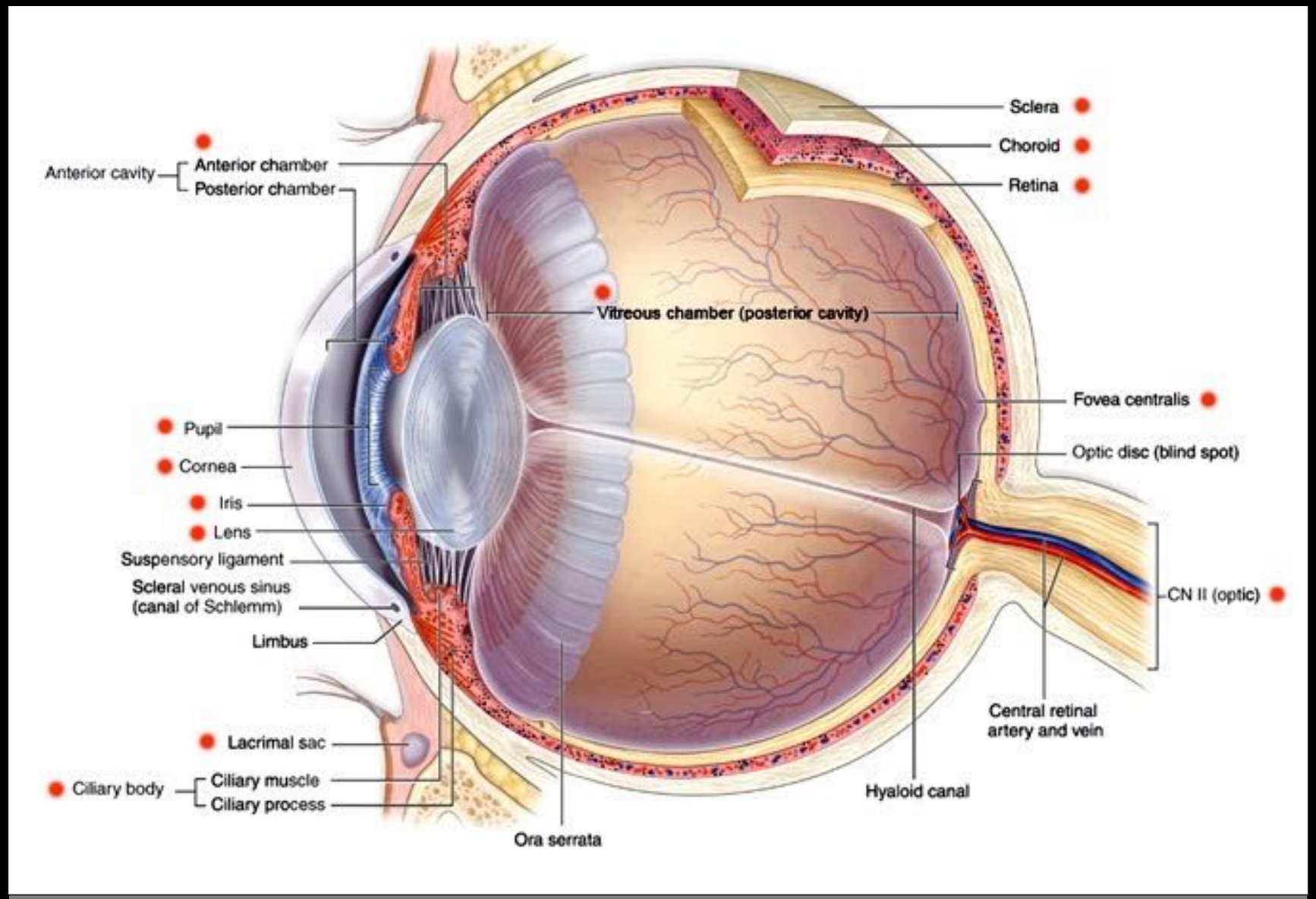
ordered

categorical



Psychophysics

The study of human perception



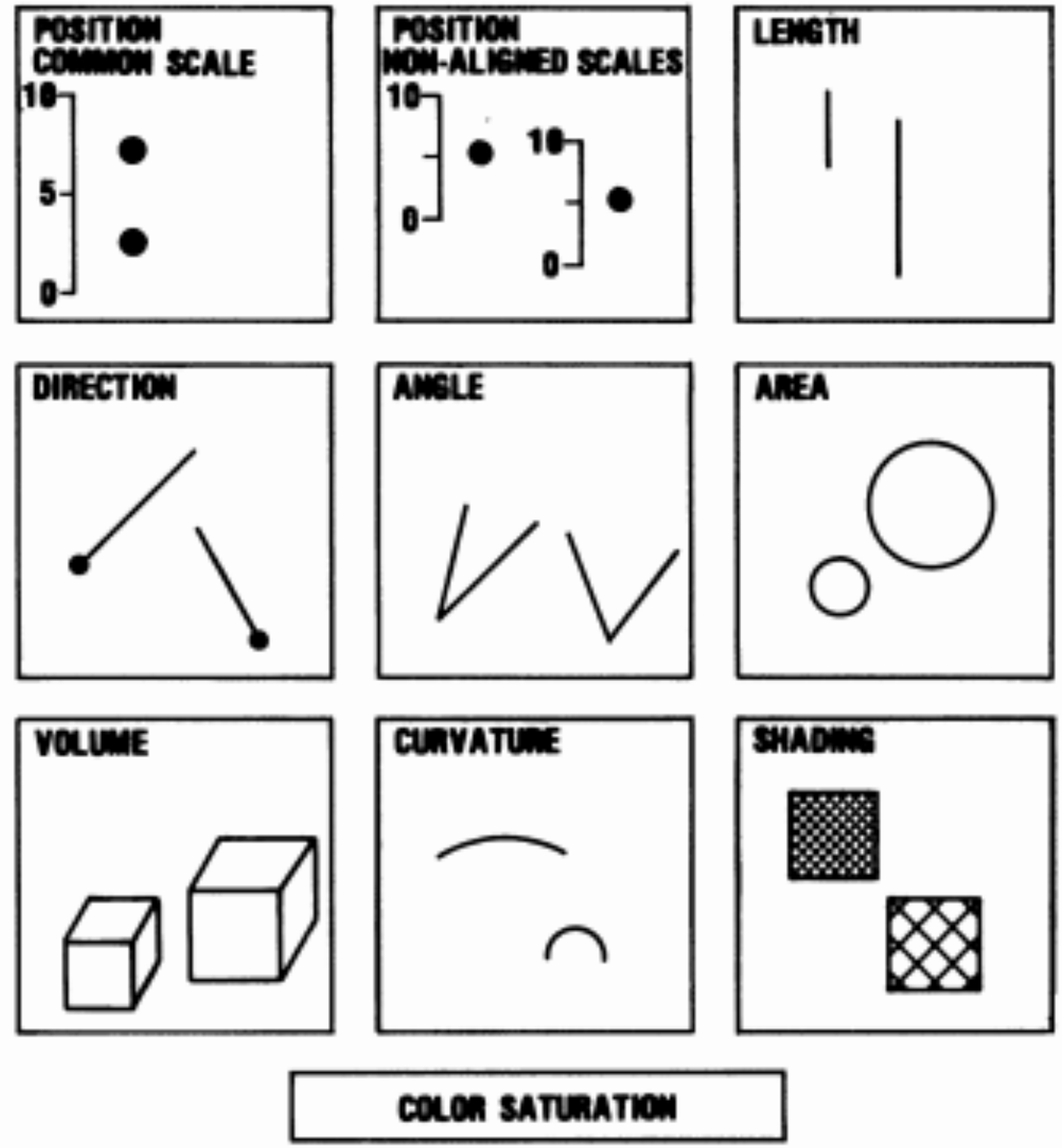


Figure 1. Elementary perceptual tasks.

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

Stevens 1975

Psychophysical power law

The apparent magnitude of all sensory channels follows a power law based on the stimulus intensity

$$S = I^n$$

S sensation, I intensity

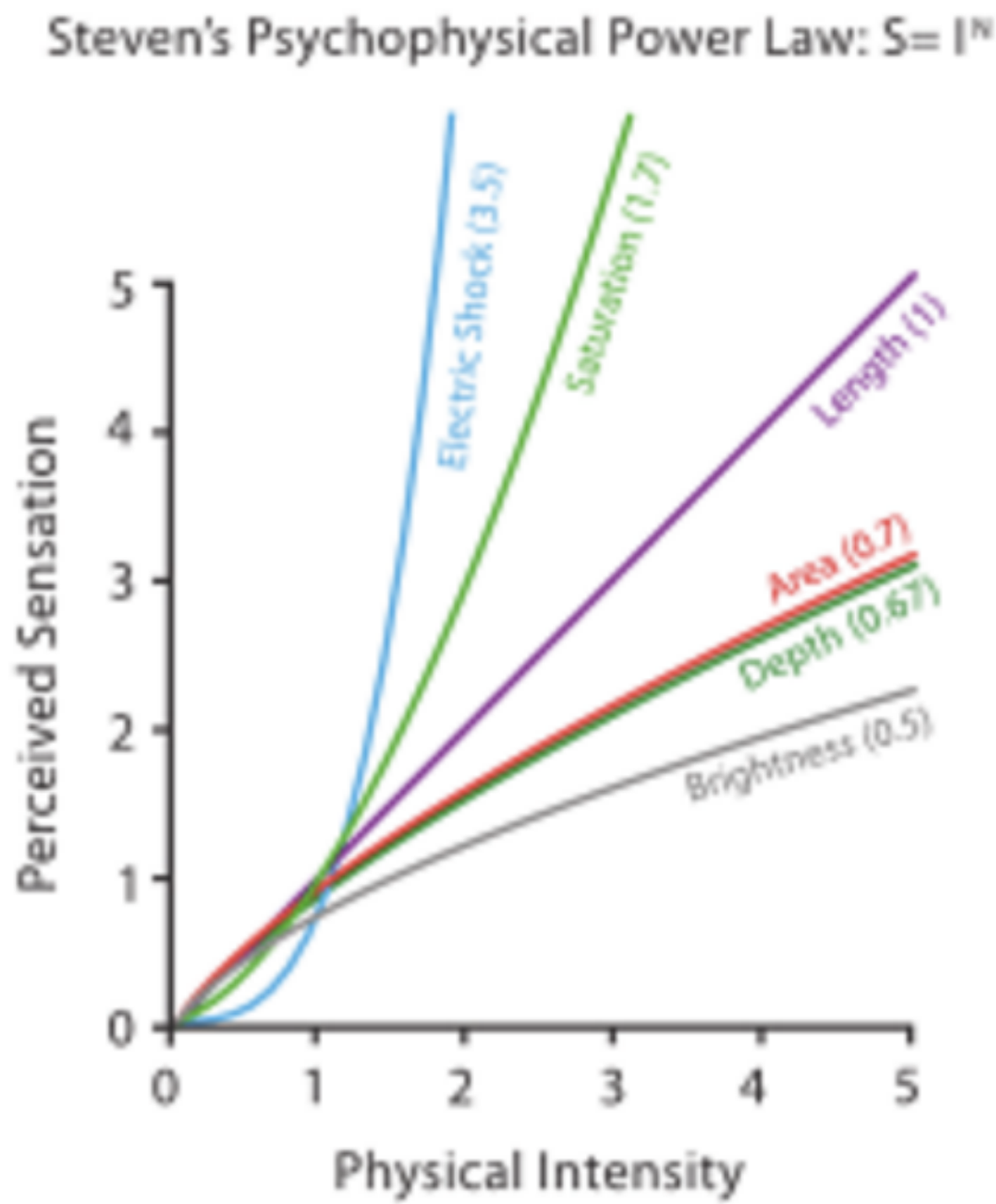
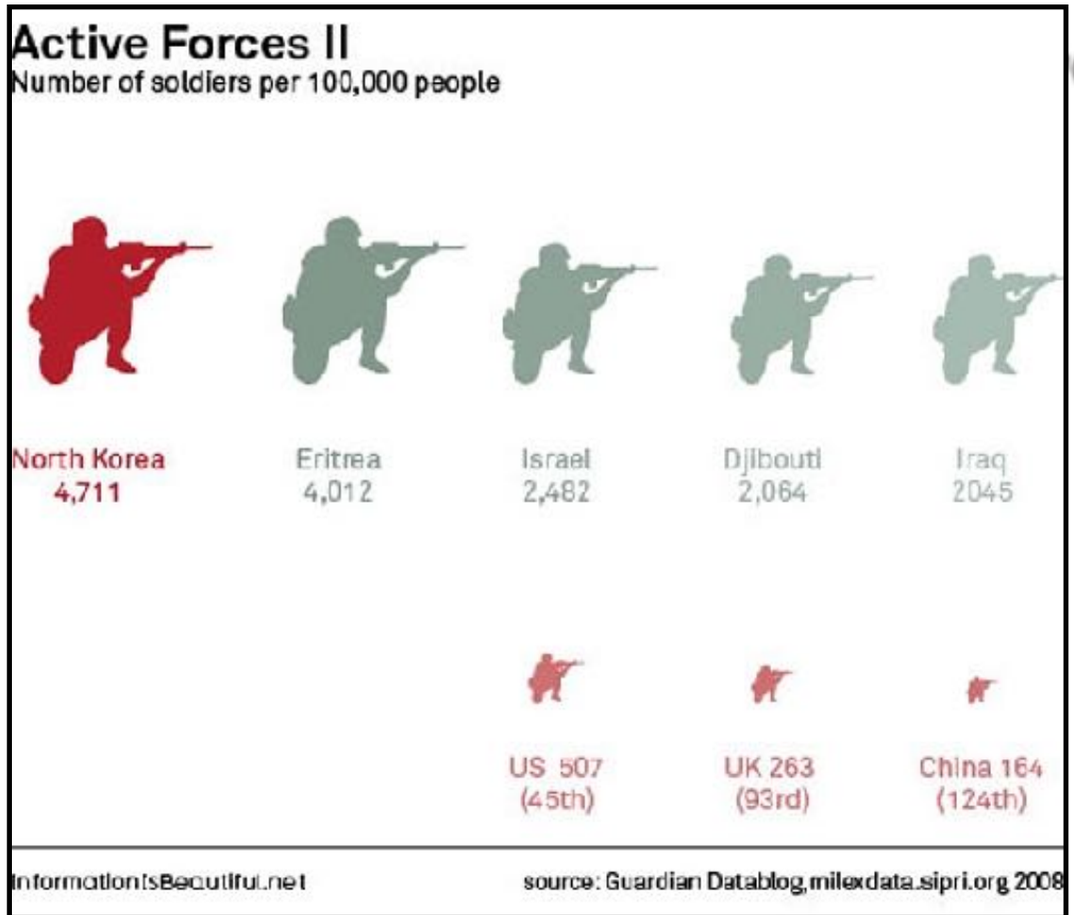


Figure 5.7. Stevens showed that the apparent magnitude of all sensory channels follows a power law $S = I^n$, where some sensations are perceptually magnified compared with their objective intensity (when $n > 1$) and some compressed (when $n < 1$). Length perception is completely accurate, whereas area is compressed and saturation is magnified. Data from Stevens [Stevens 75, p. 15].



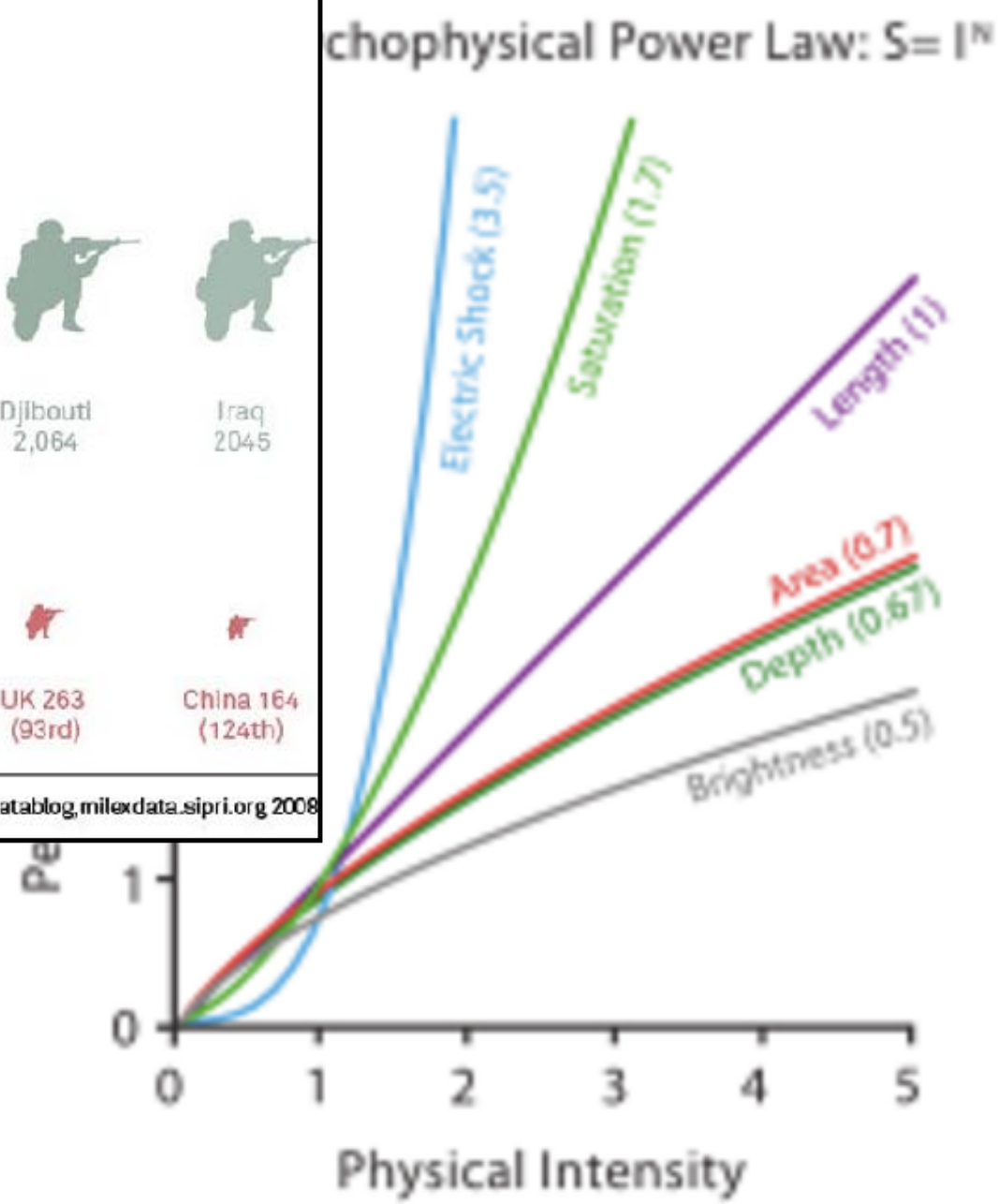
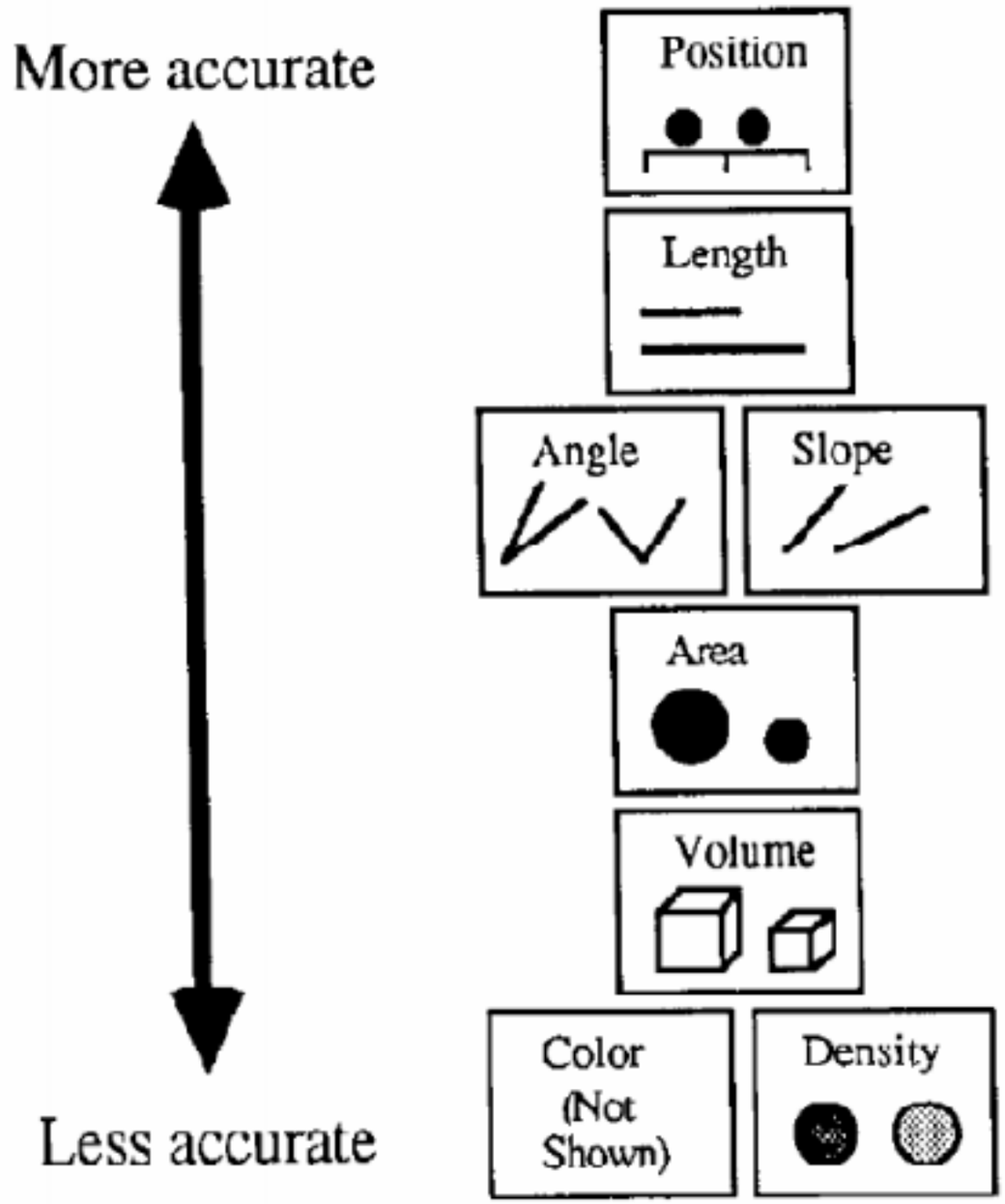
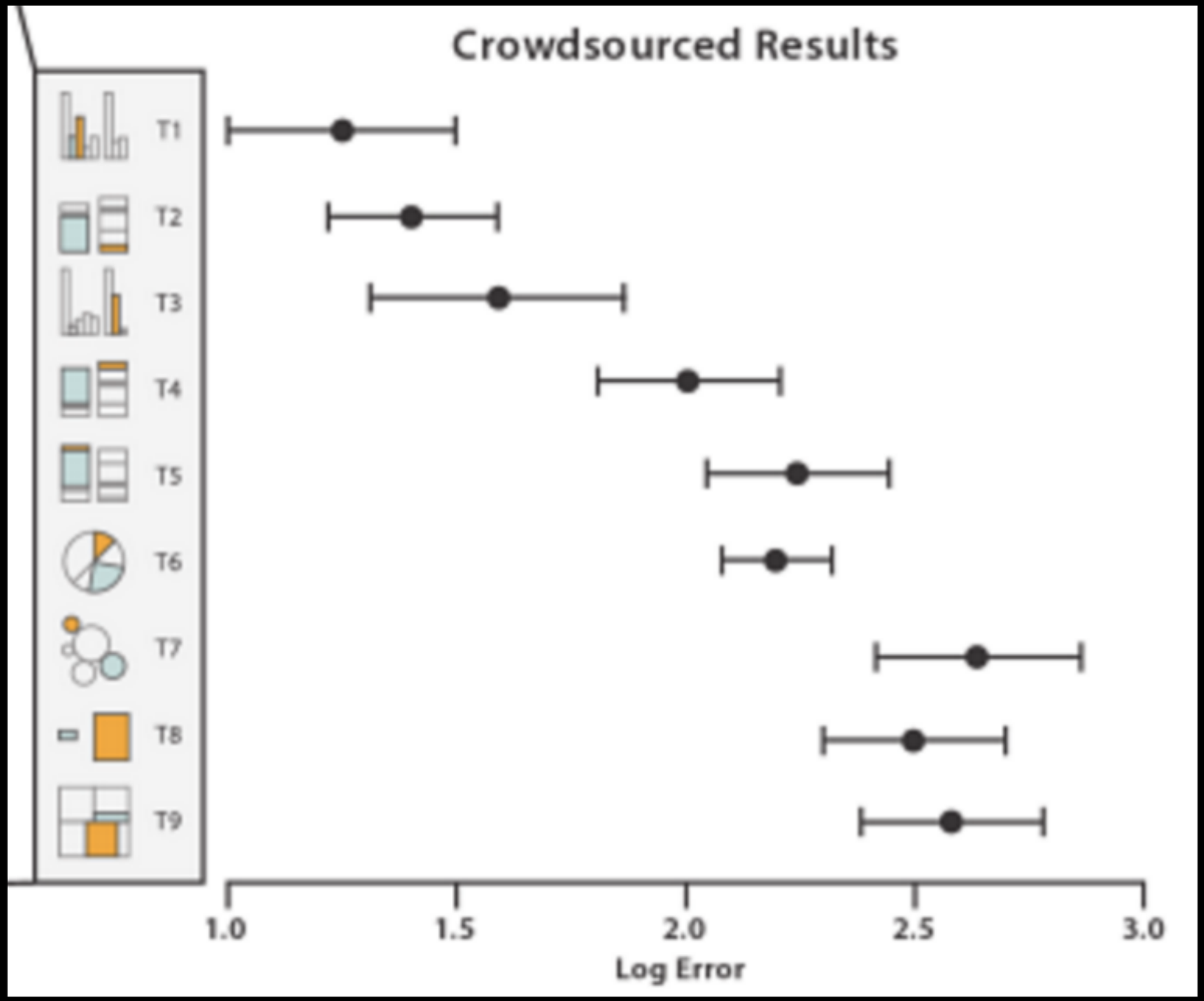


Figure 5.7. Stevens showed that the apparent magnitude of all sensory channels follows a power law $S = I^n$, where some sensations are perceptually magnified compared with their objective intensity (when $n > 1$) and some compressed (when $n < 1$). Length perception is completely accurate, whereas area is compressed and saturation is magnified. Data from Stevens [Stevens 75, p. 15].

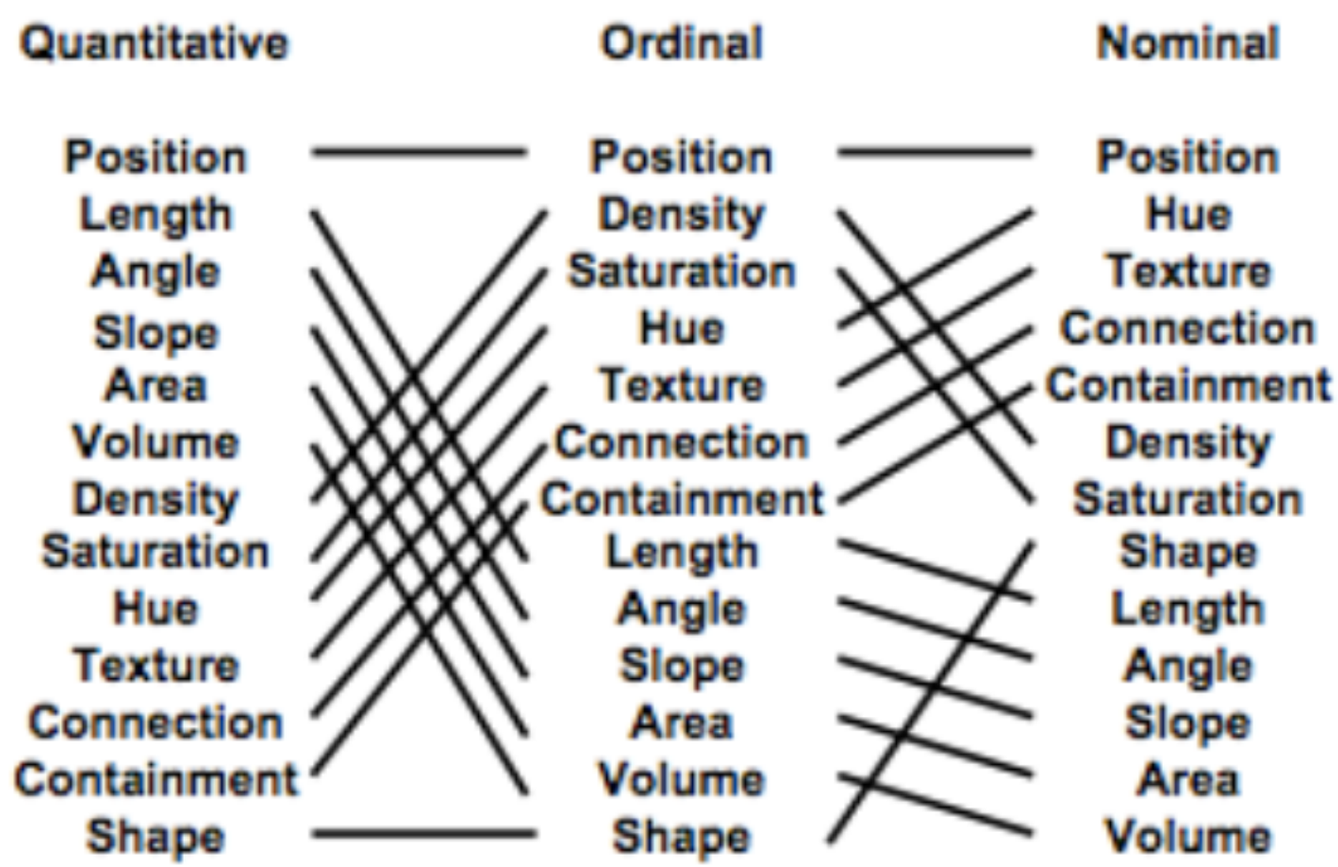


Automating the Design of Graphical Presentations of Relational Information ,Jock Mackinlay, ACM Transaction on Graphics, 1986

Cleveland & McGill, Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods 1984



Heer and Bostock 2010



[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]

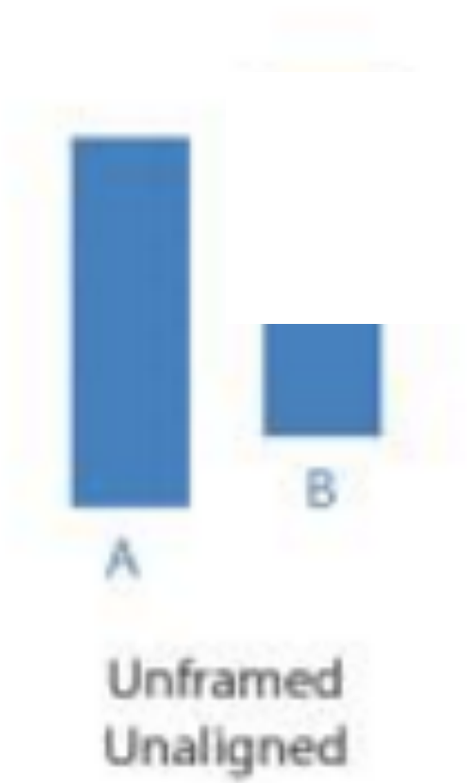
Weber Law

We judge based on relative differences

The detectable difference in stimulus intensity is a fixed percentage of the object magnitude

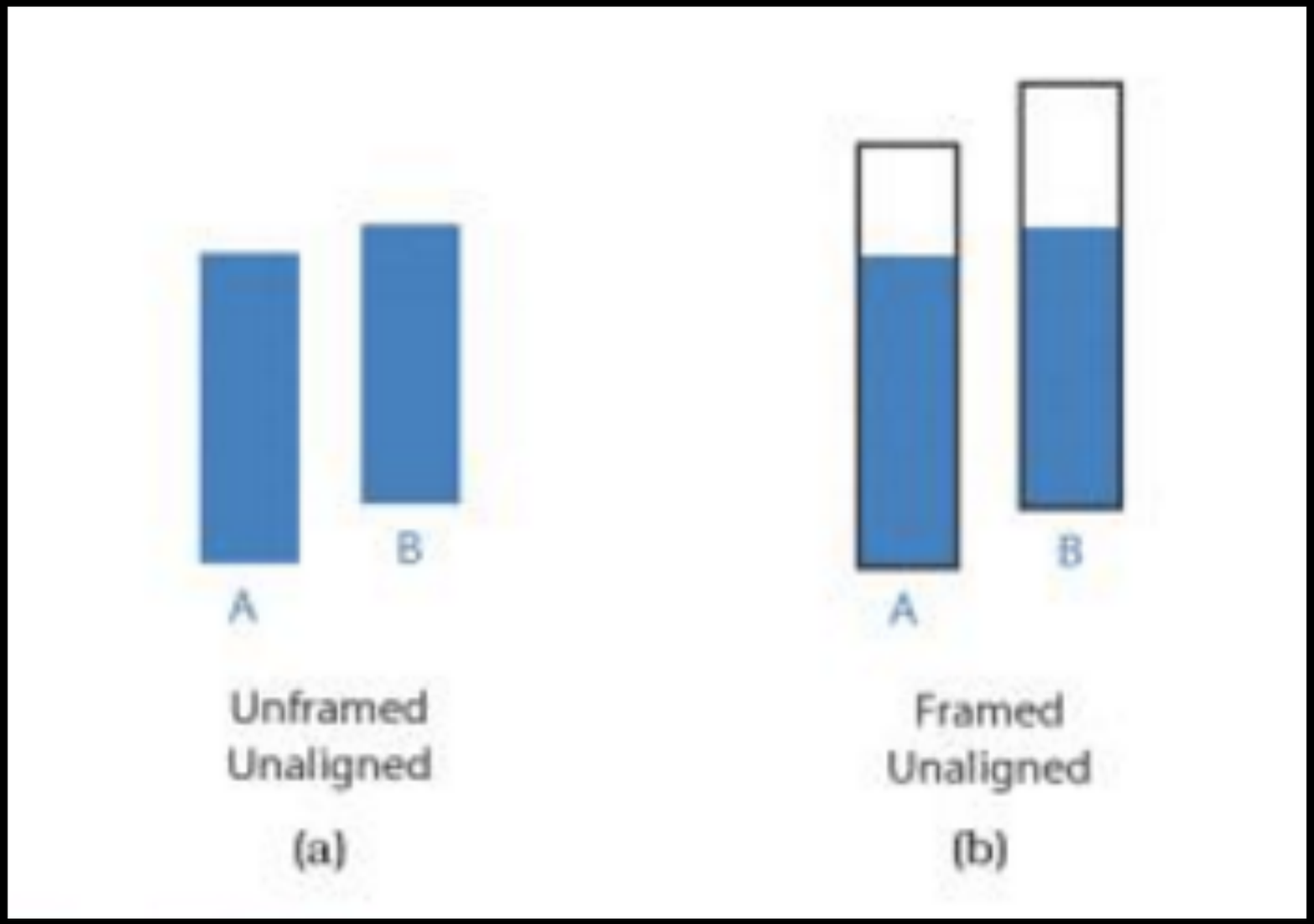
$$\delta I / I = K$$

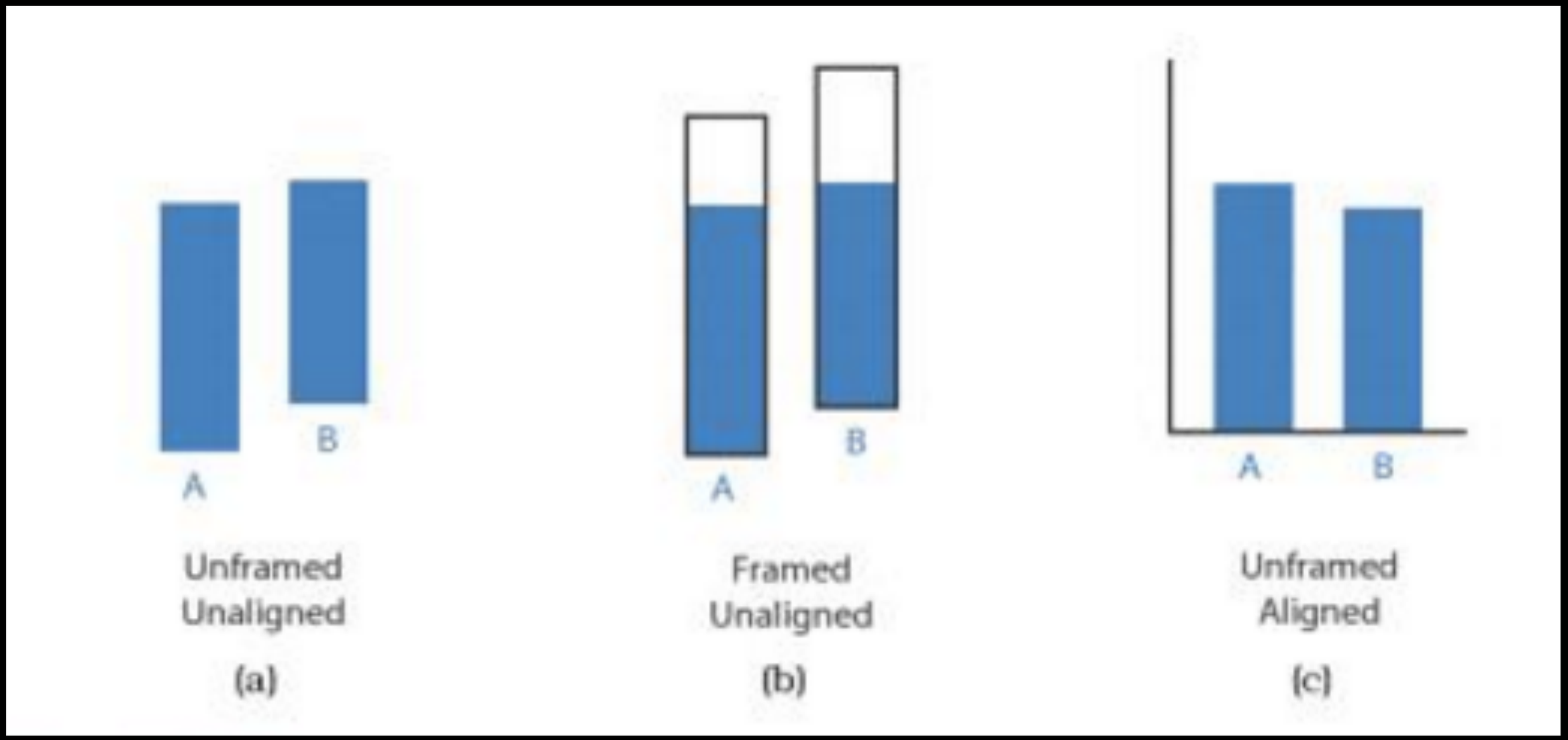
I intensity, K constant



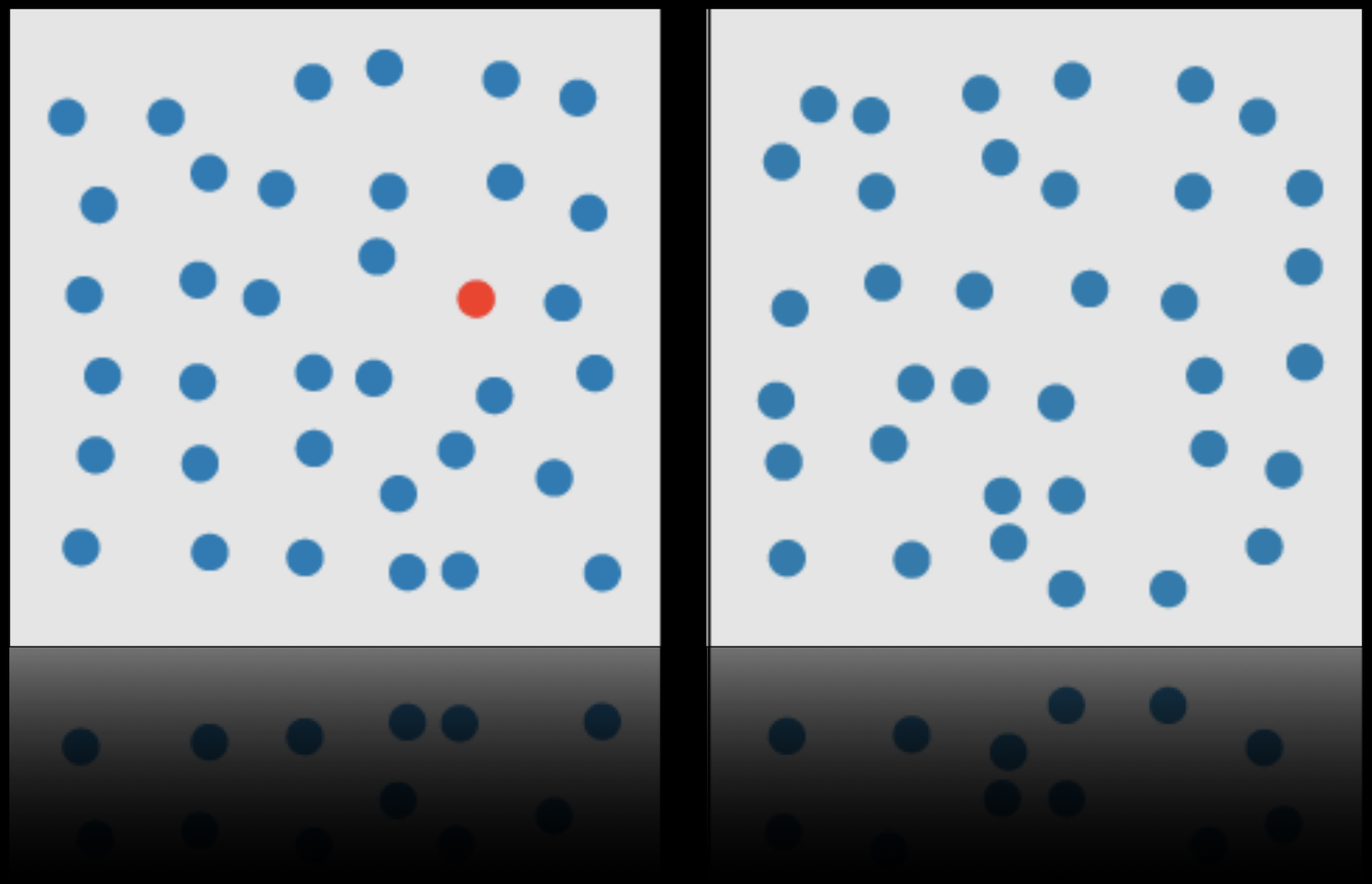
(a)

(a)

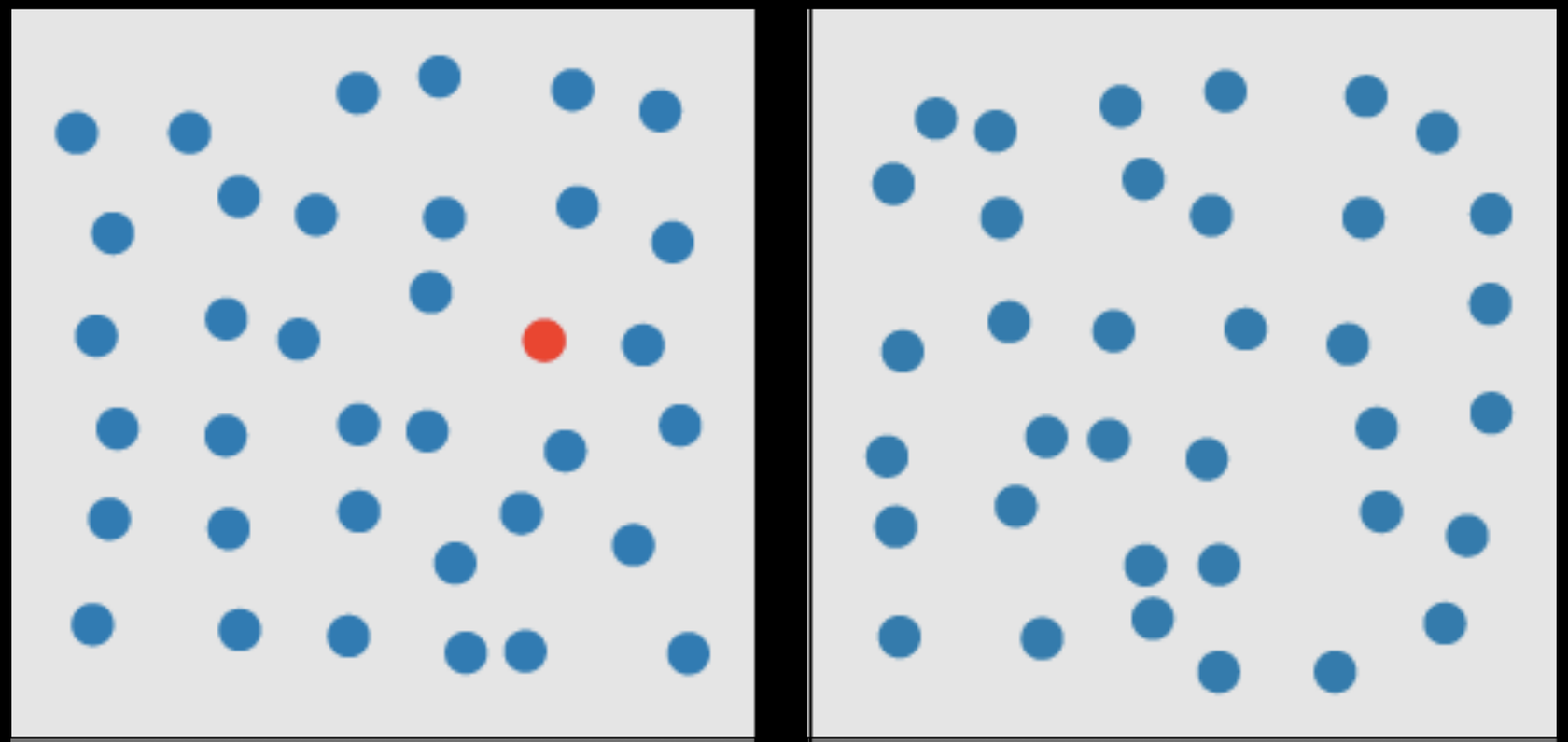

B
A
Unframed
Unaligned
(a)
B
A
Framed
Unaligned
(b)


A
B
Unframed
Unaligned
(a)
A
B
Framed
Unaligned
(b)
A
B
Unframed
Aligned
(c)

Preattentive tasks



Preattentive tasks

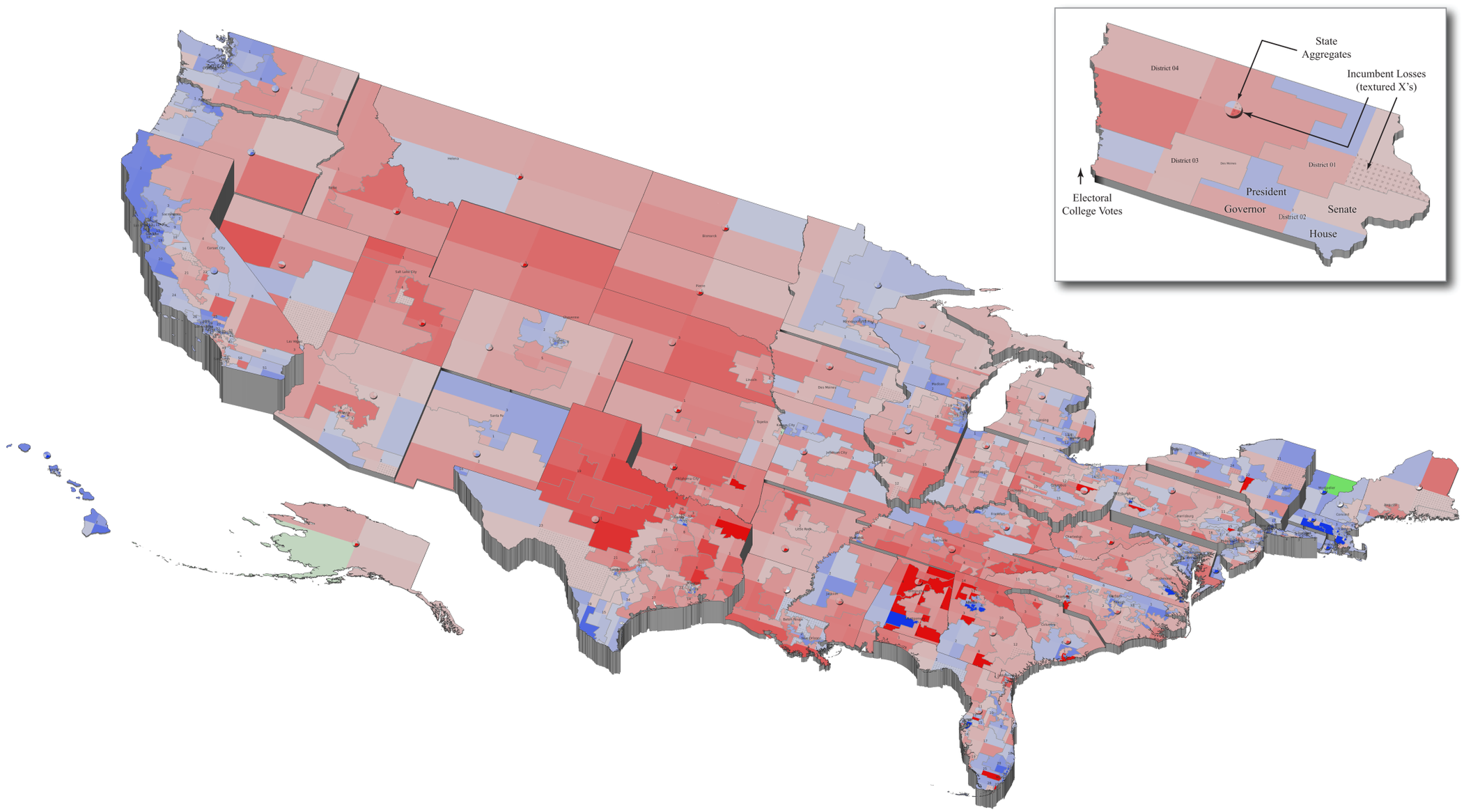


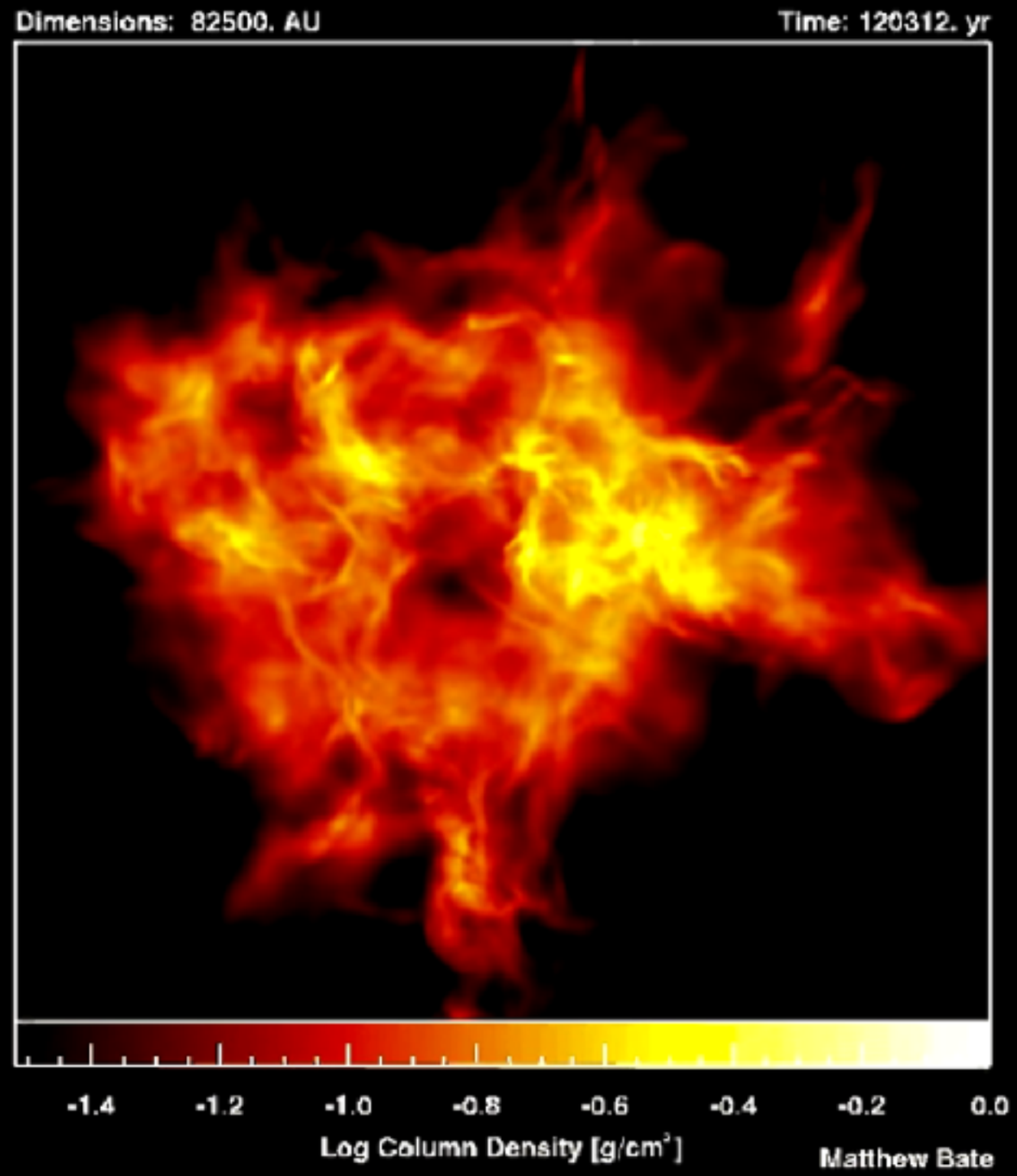
a limited set of visual properties that are detected very rapidly and accurately by the low-level visual system. (tasks that can be performed on large multi-element displays in less than 200 to 250 milliseconds)

http://www.csc.ncsu.edu/faculty/healey/PP/index.html#jscript_search

Christopher G. Healey

perceptually-motivated multidimensional visualization of recent U.S. election result





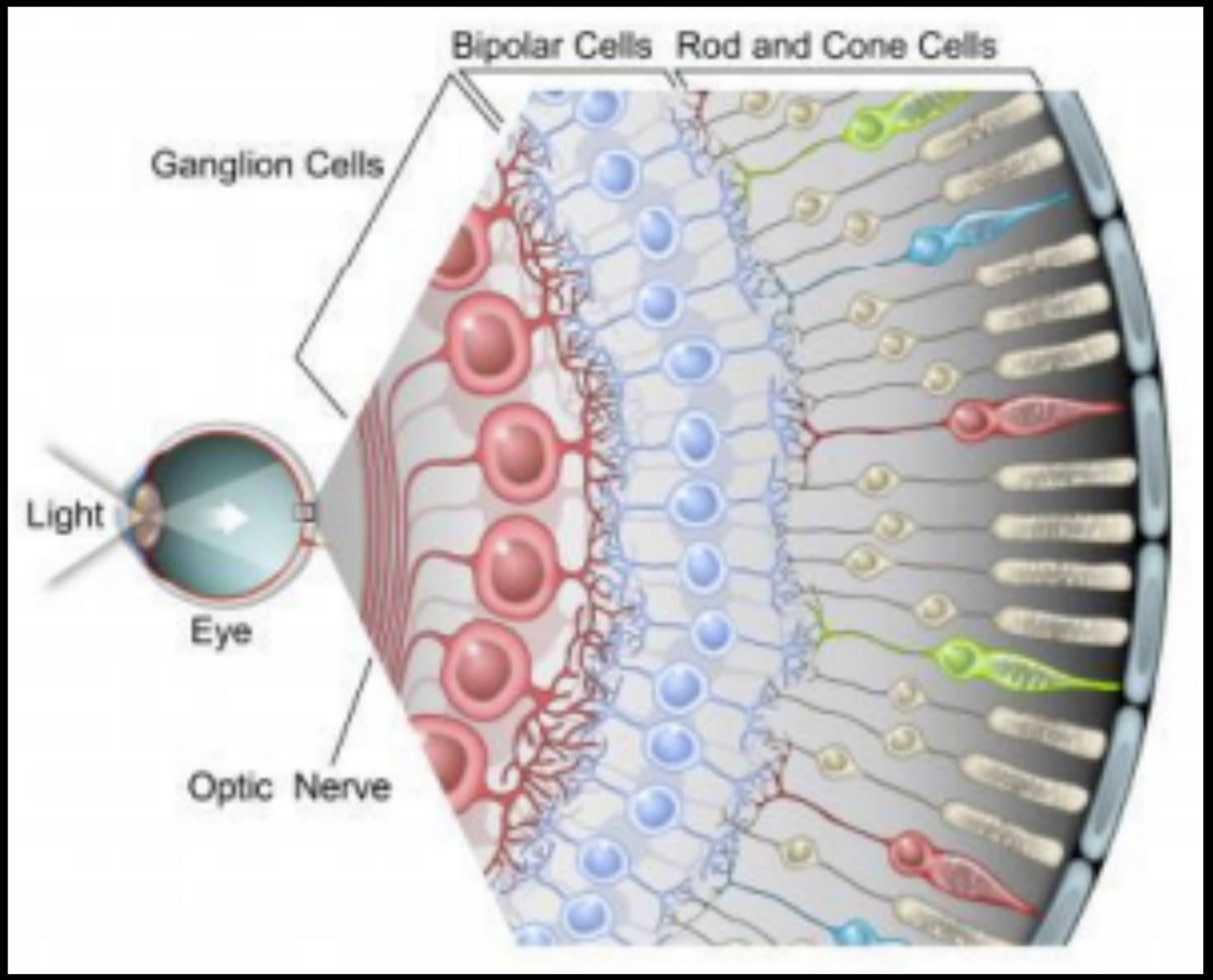
https://www.youtube.com/watch?v=YbdwTwB8jtc

http://morphocode.com/data-city-urban-visualizations/

http://www.theinvisiblegorilla.com/videos.html

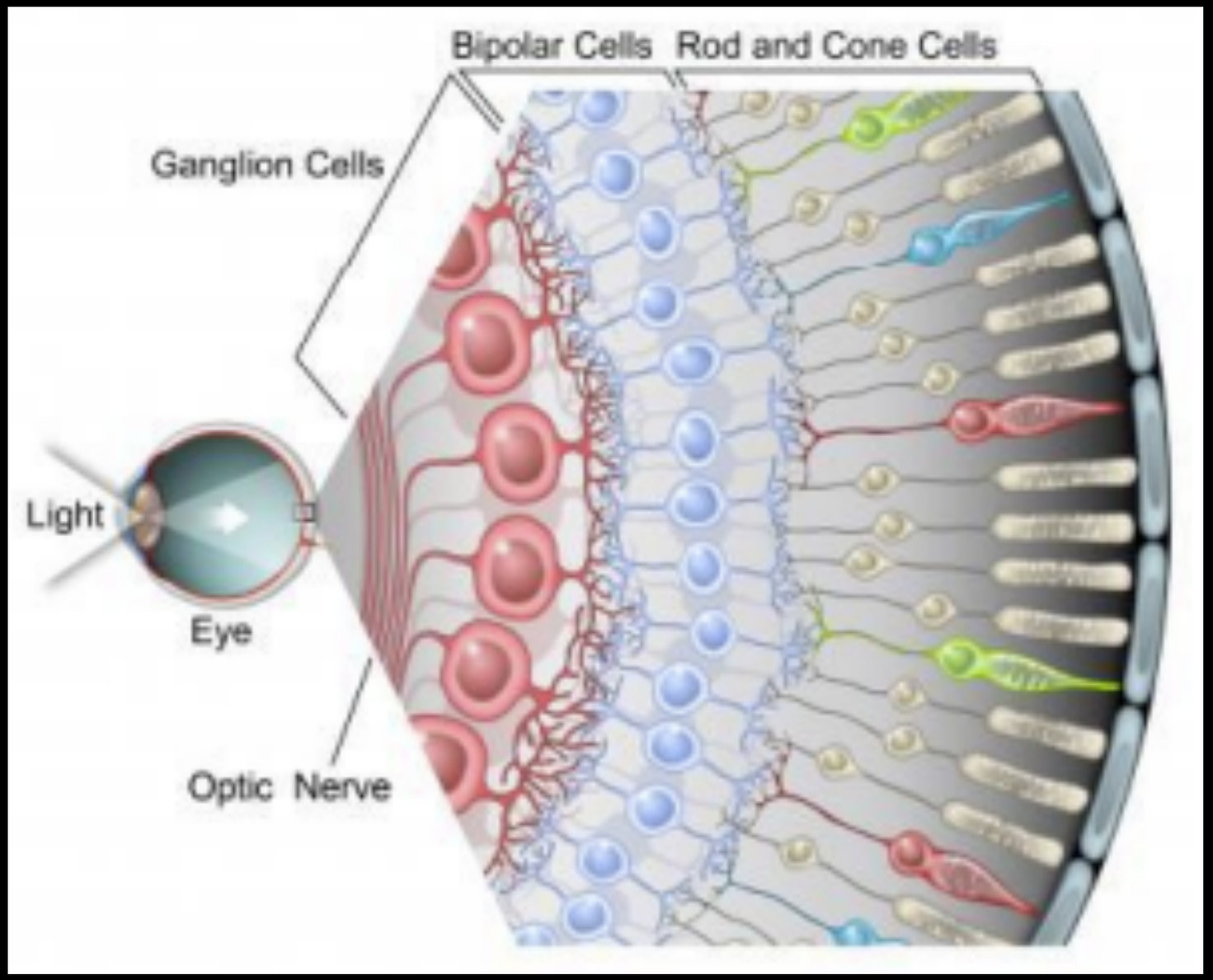
color theory
(and good practice)

Rods & Cones


Bipolar Cells
Rod and Cone Cells
Ganglion Cells
Light
Eye
Optic Nerve

Rods & Cones

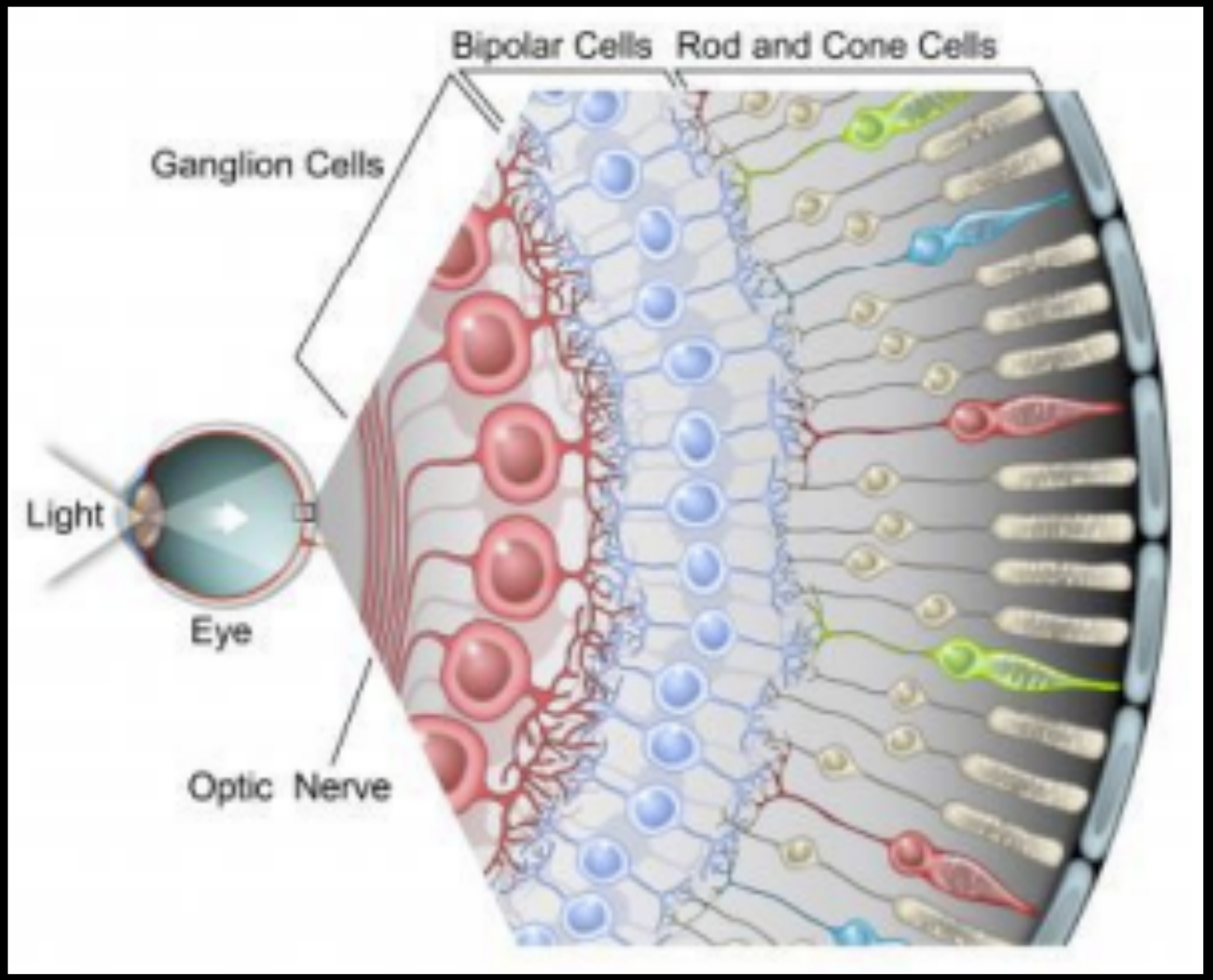
Brightness & Color

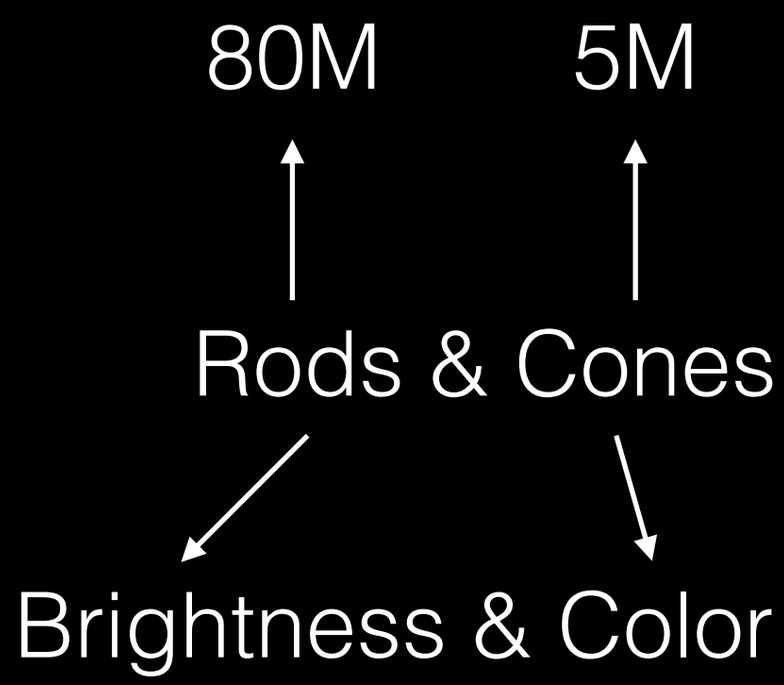


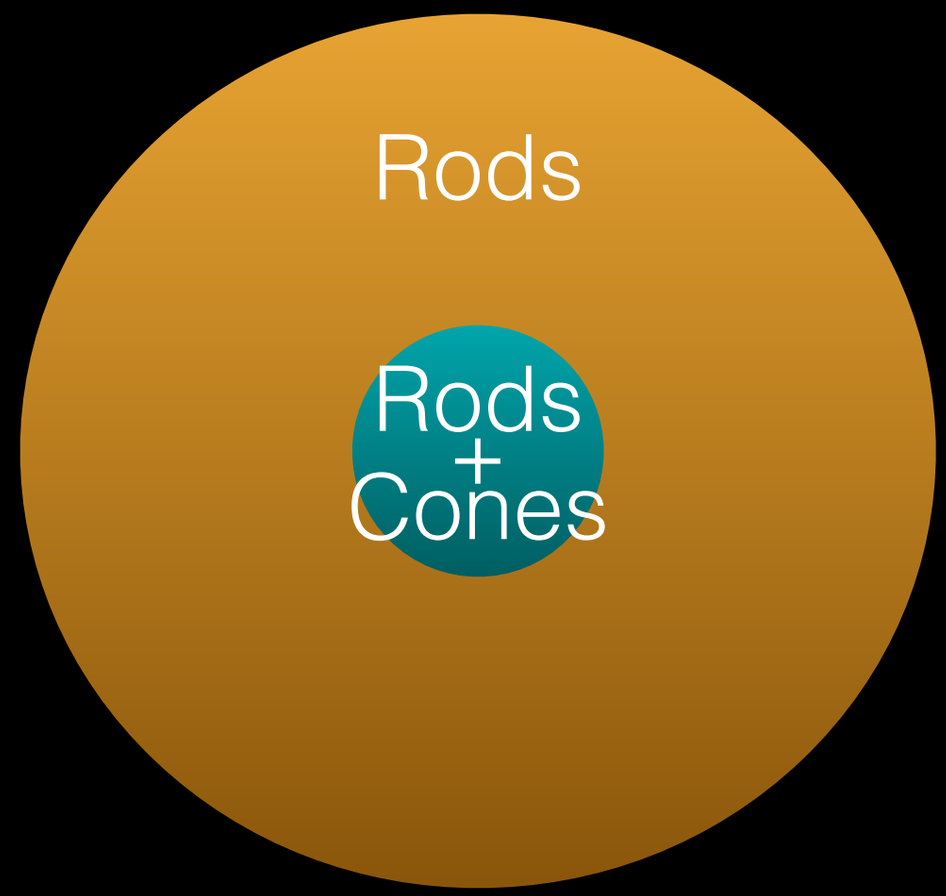
80M 5M

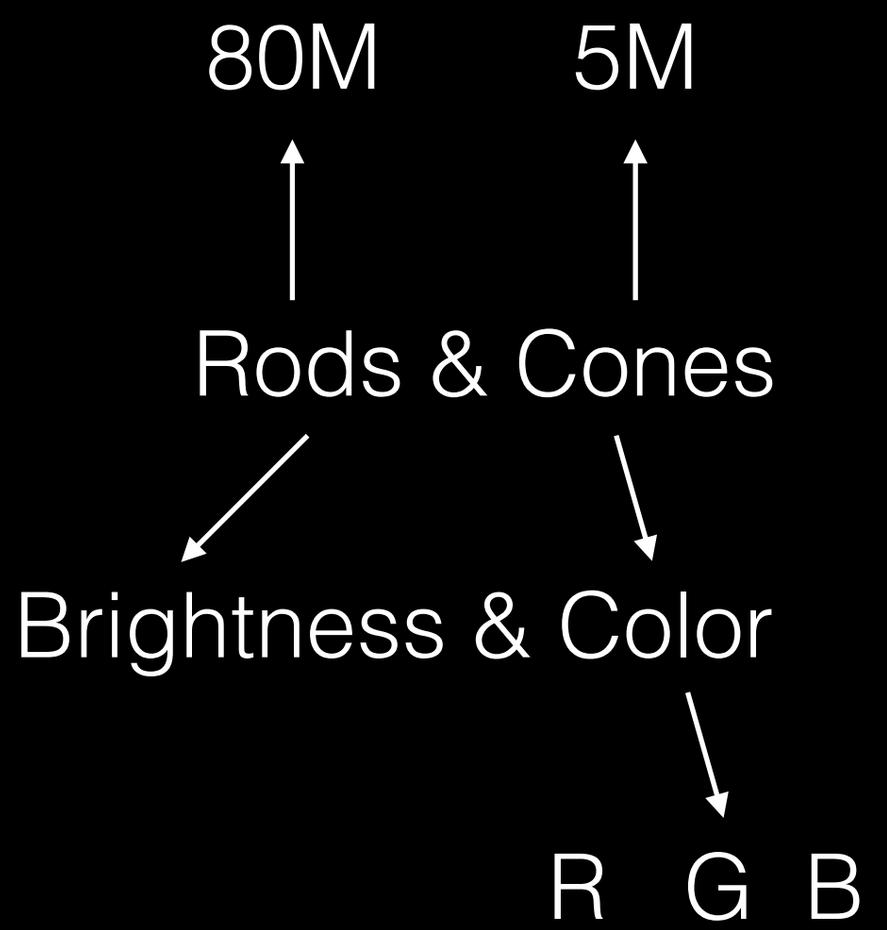
Rods & Cones

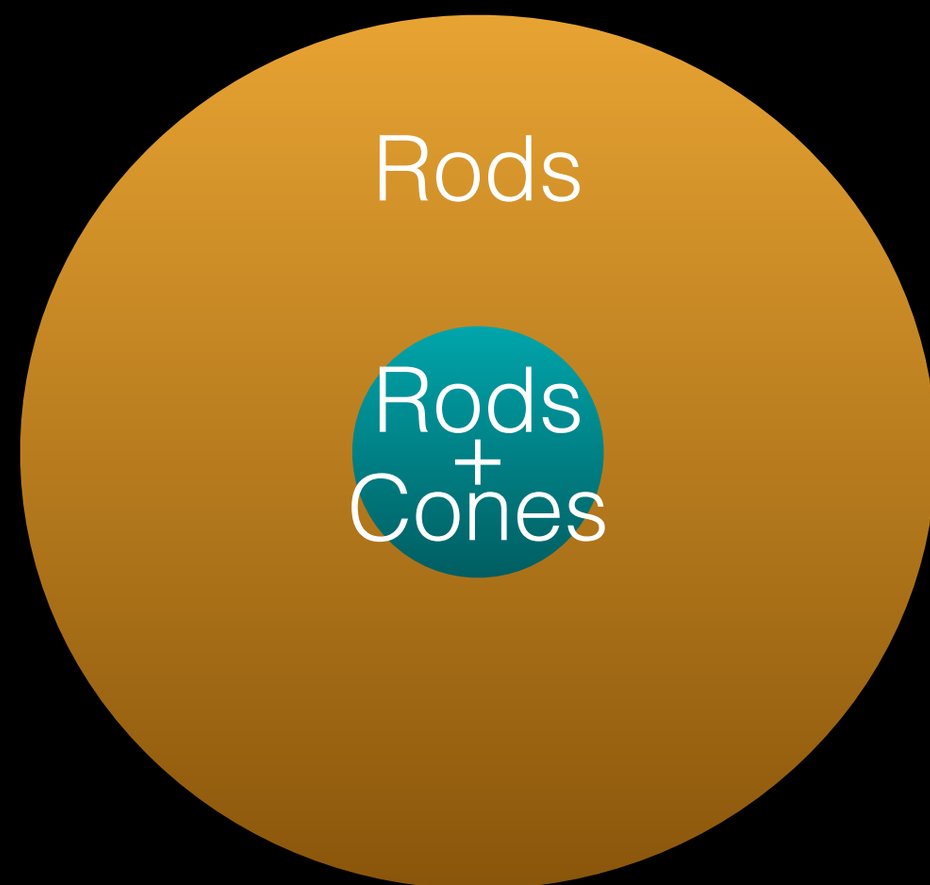
Brightness & Color




80M
5M
Rods & Cones
Brightness & Color


Rods
Rods
+
Cones





the message to the brain is:

black ⟷ white
yellow ⟷ blue
red ⟷ green

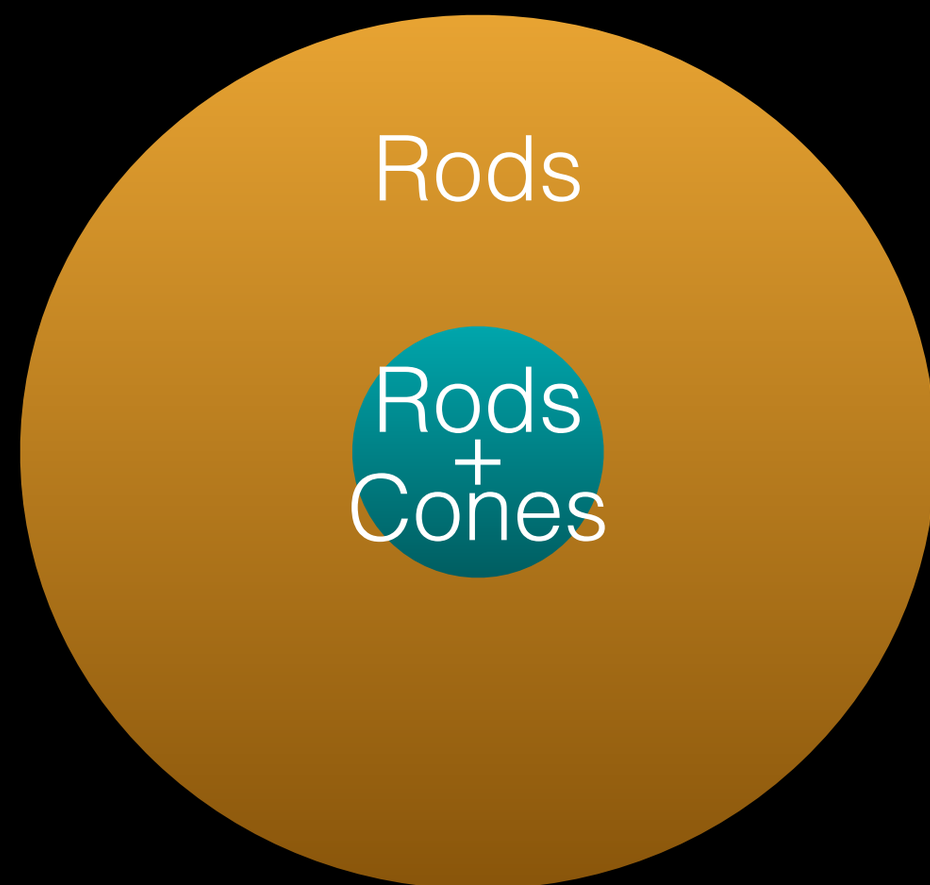
COLOR BLINDNESS

80M 5M

Rods & Cones

Brightness & Color

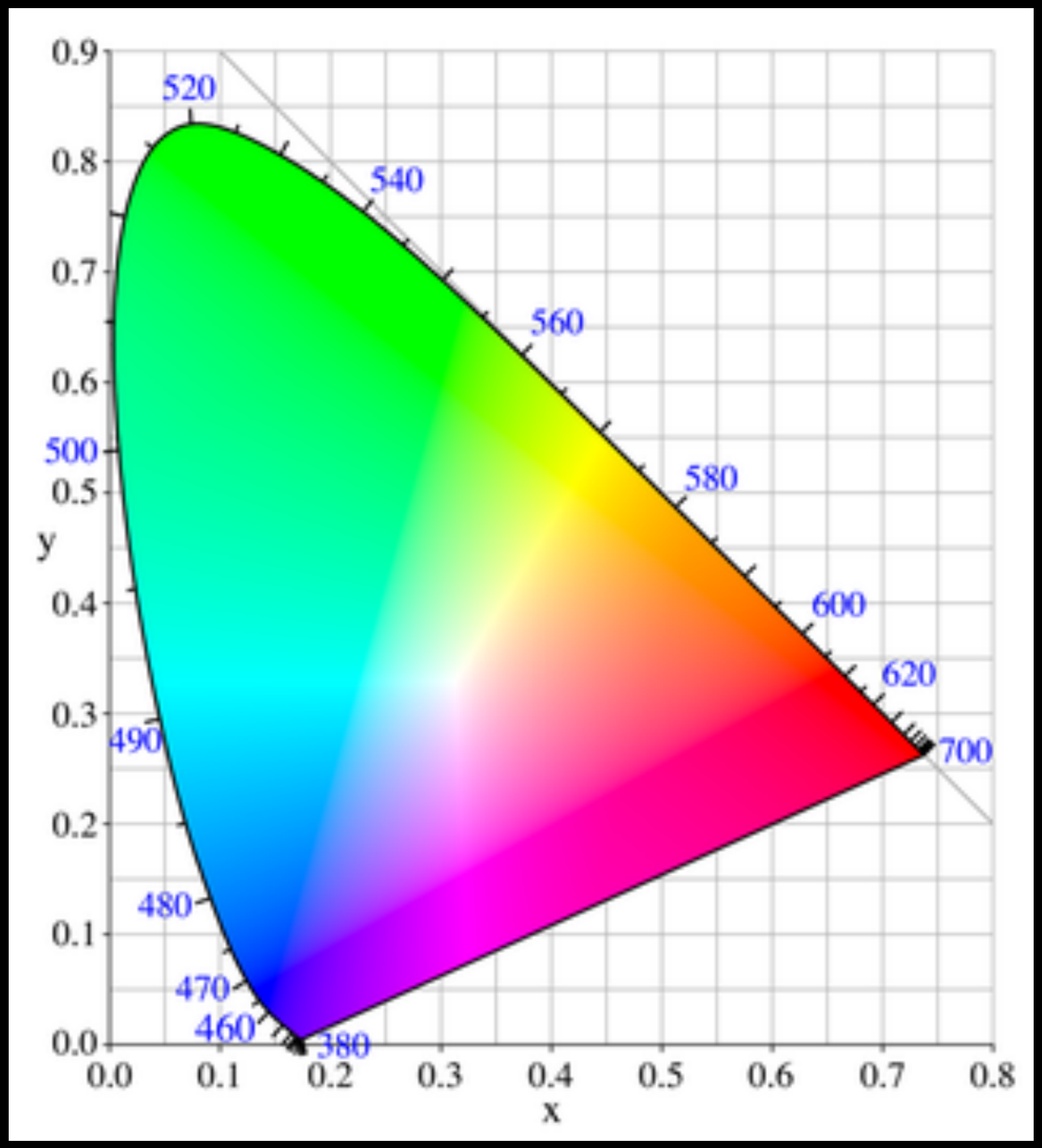
R G B



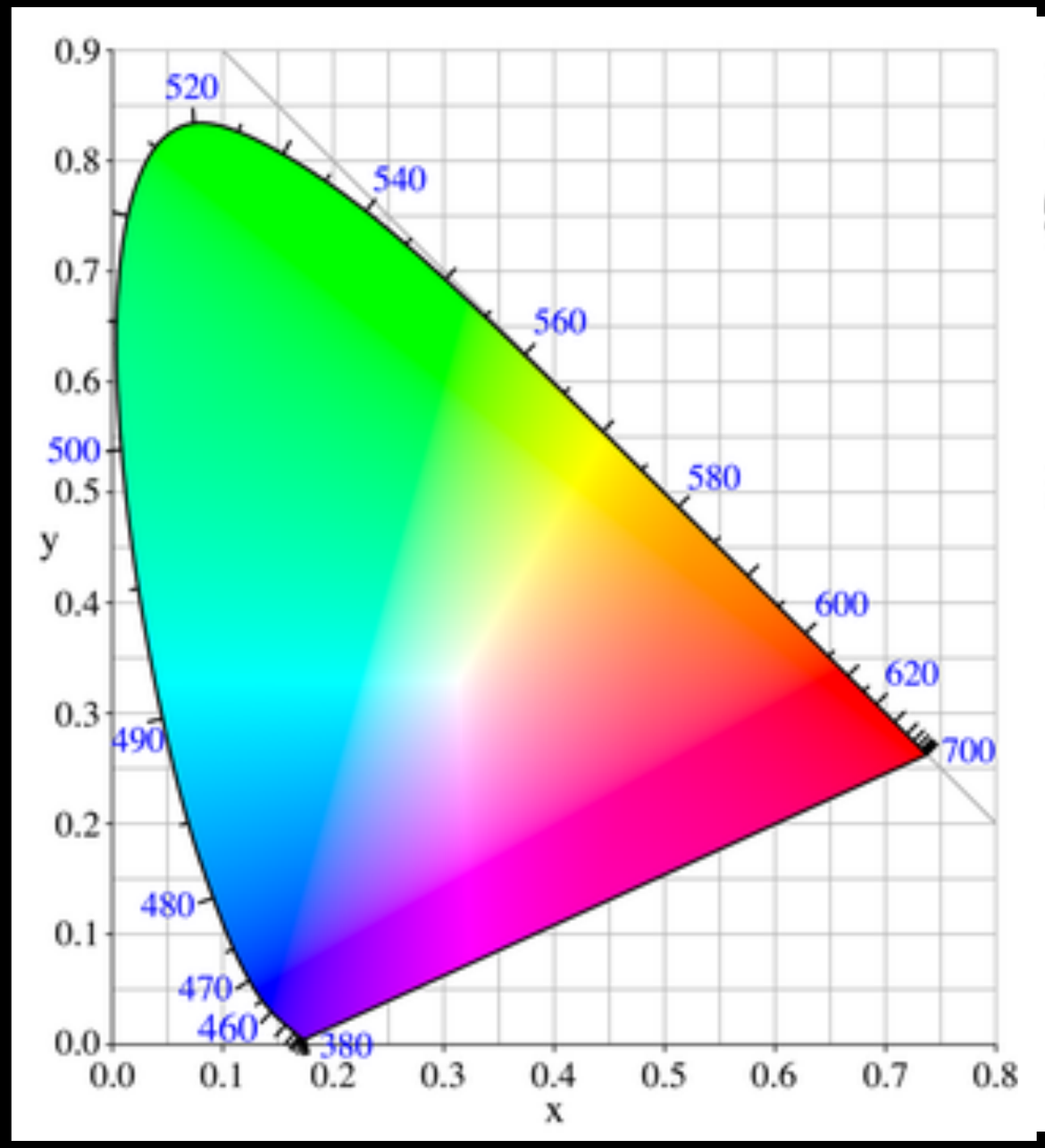
the message to the brain is:

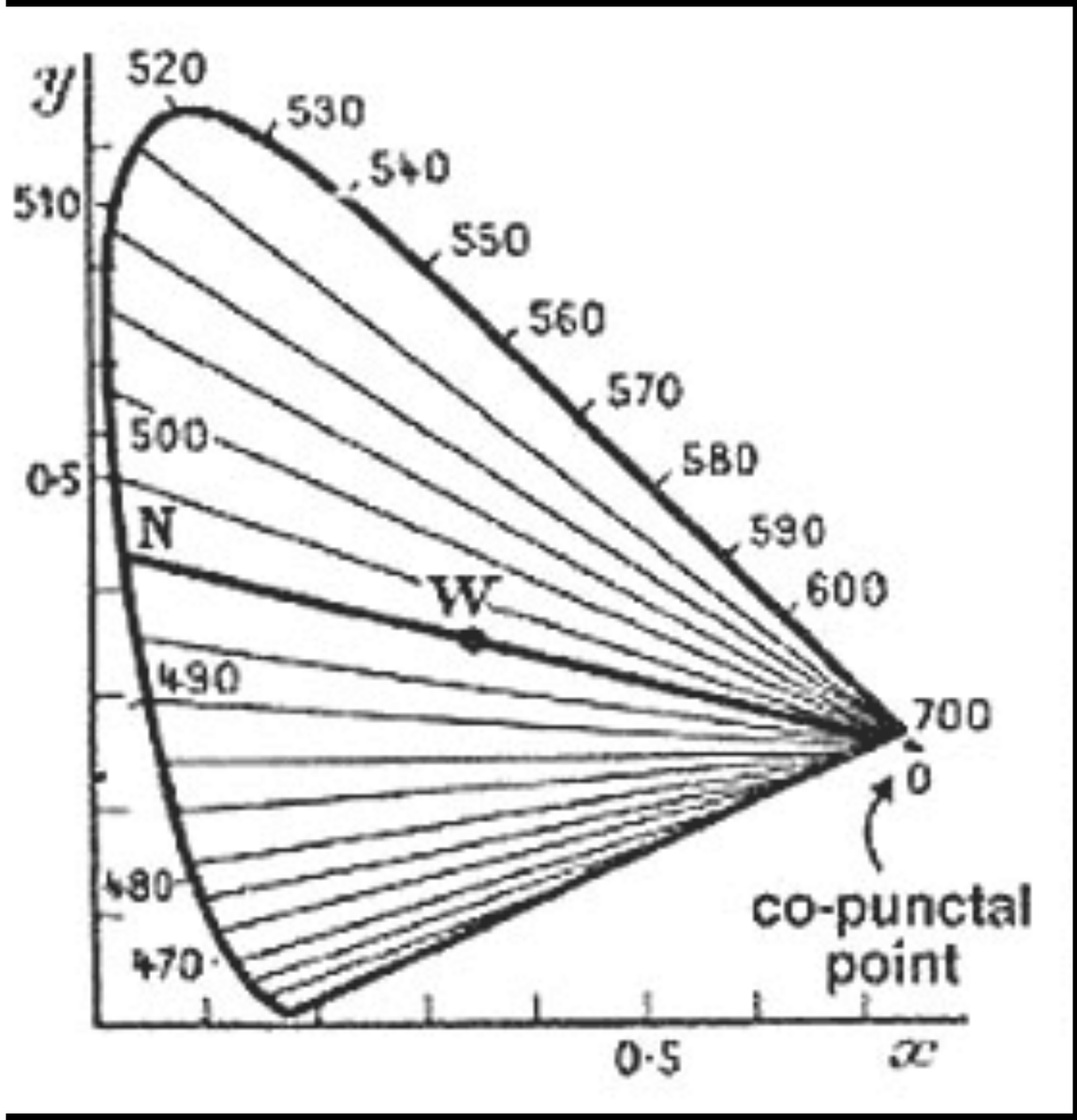
black ⟷ white
yellow ⟷ blue
red ⟷ green

COLOR BLINDNESS



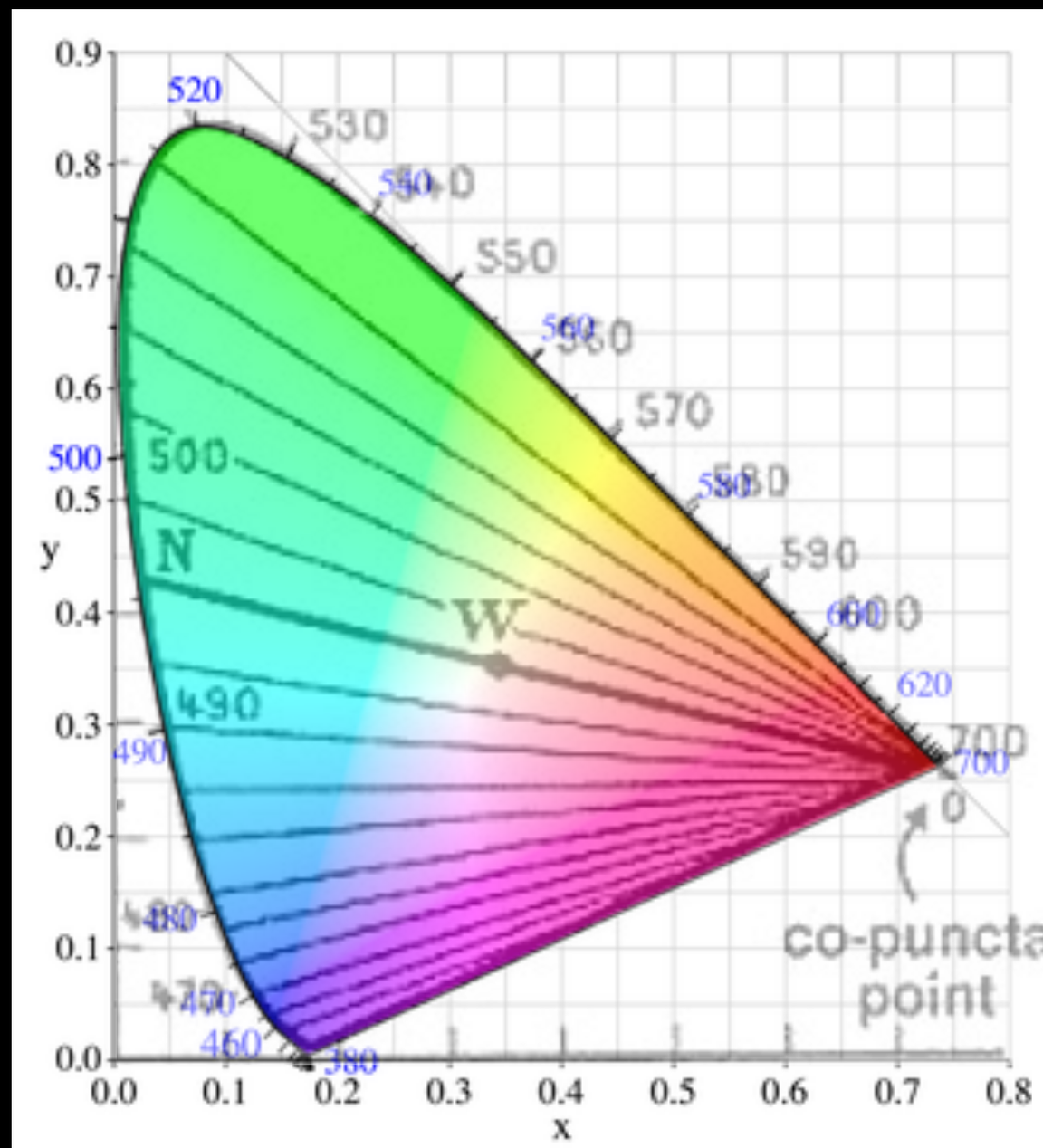
COLOR BLINDNESS

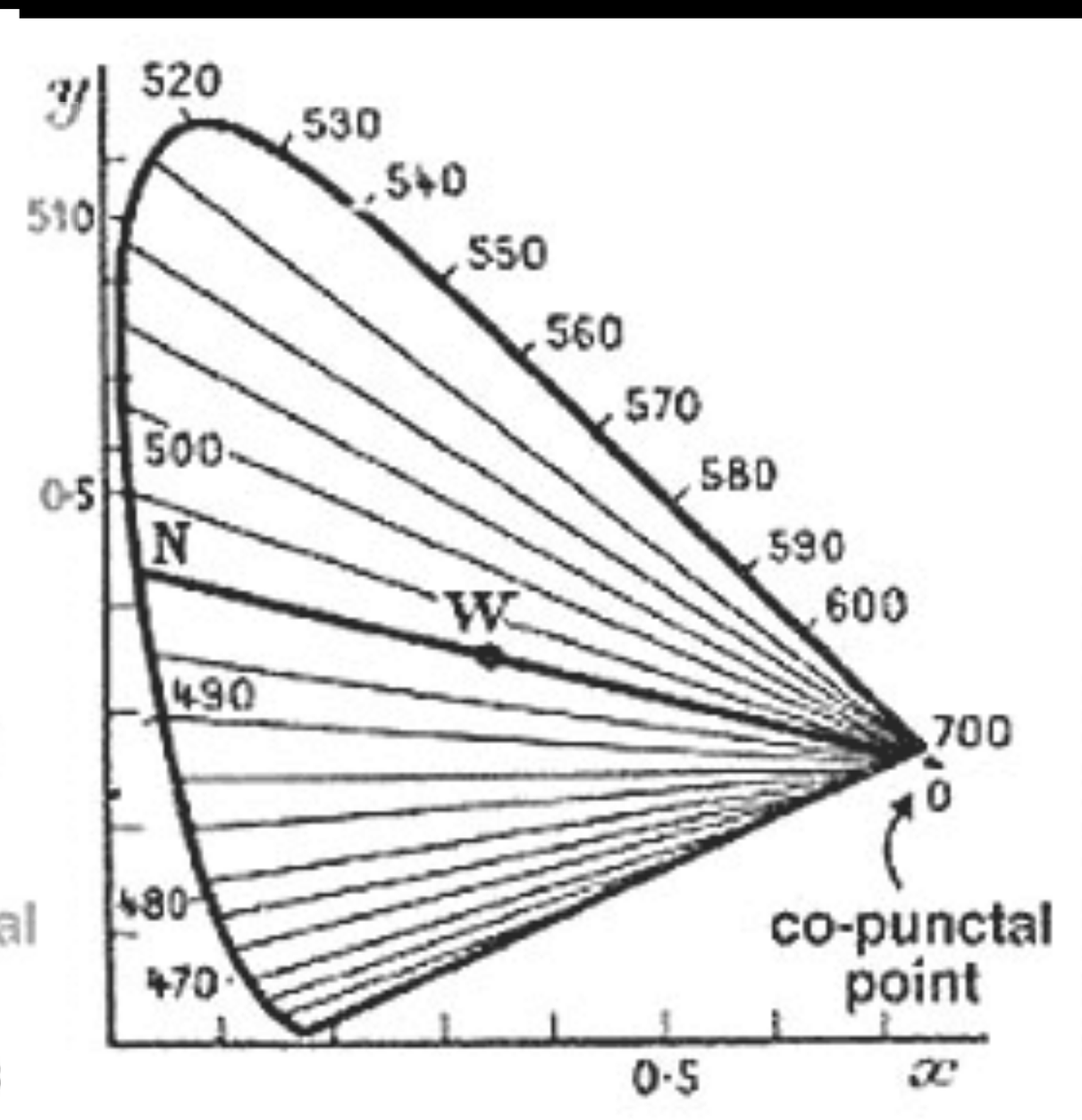




Protanopia

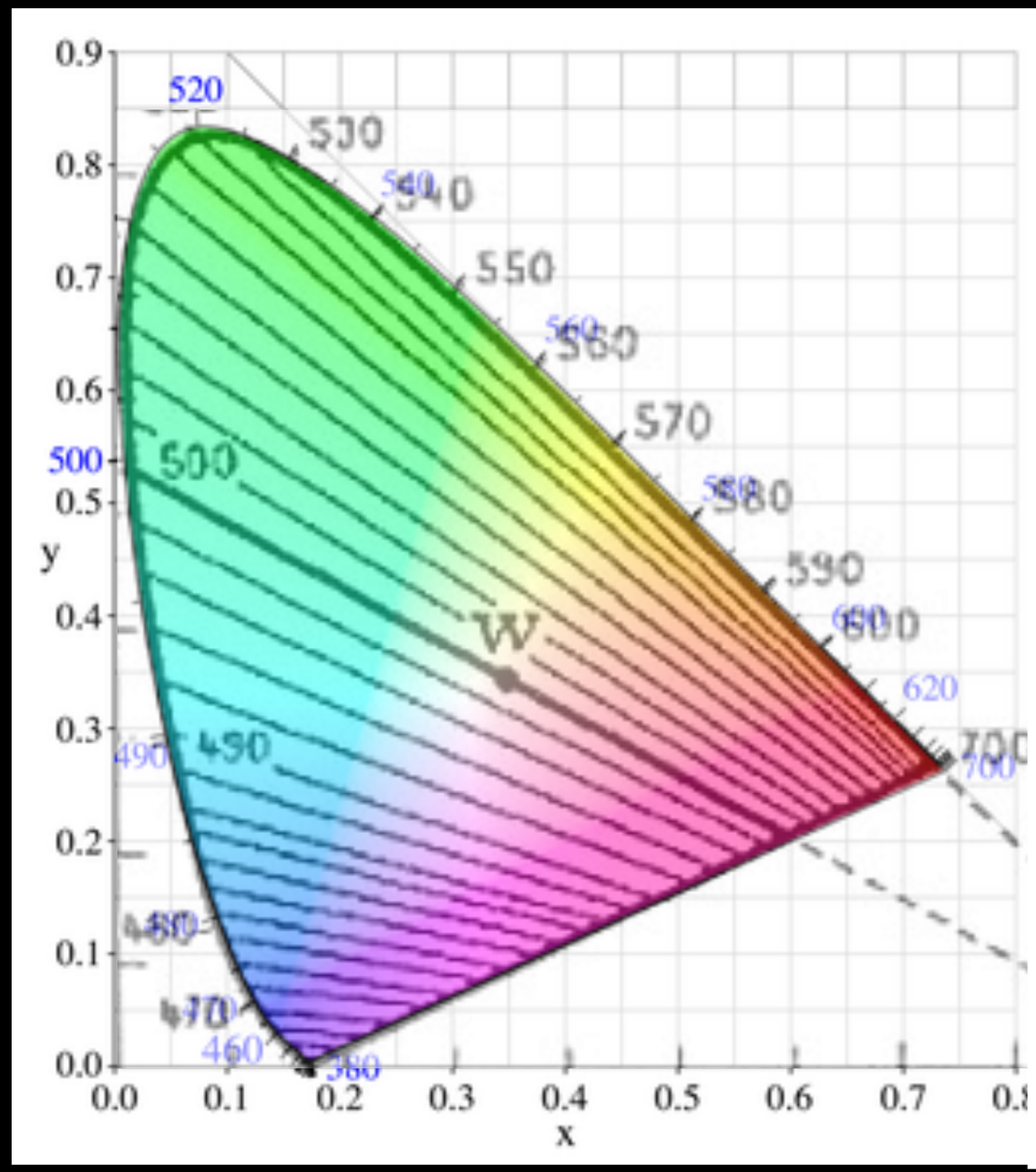
COLOR BLINDNESS

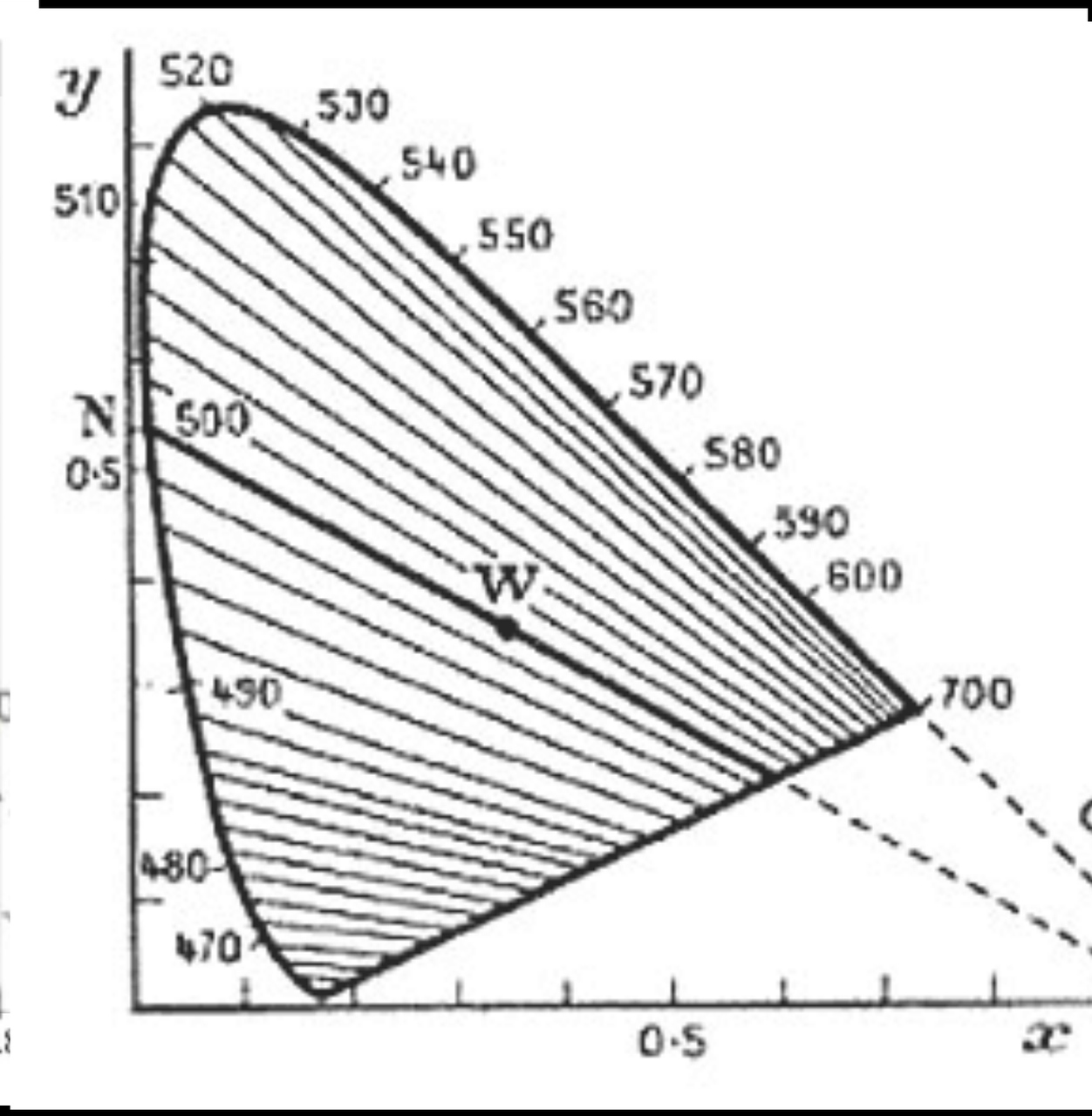




Protanopia
(red-blind)

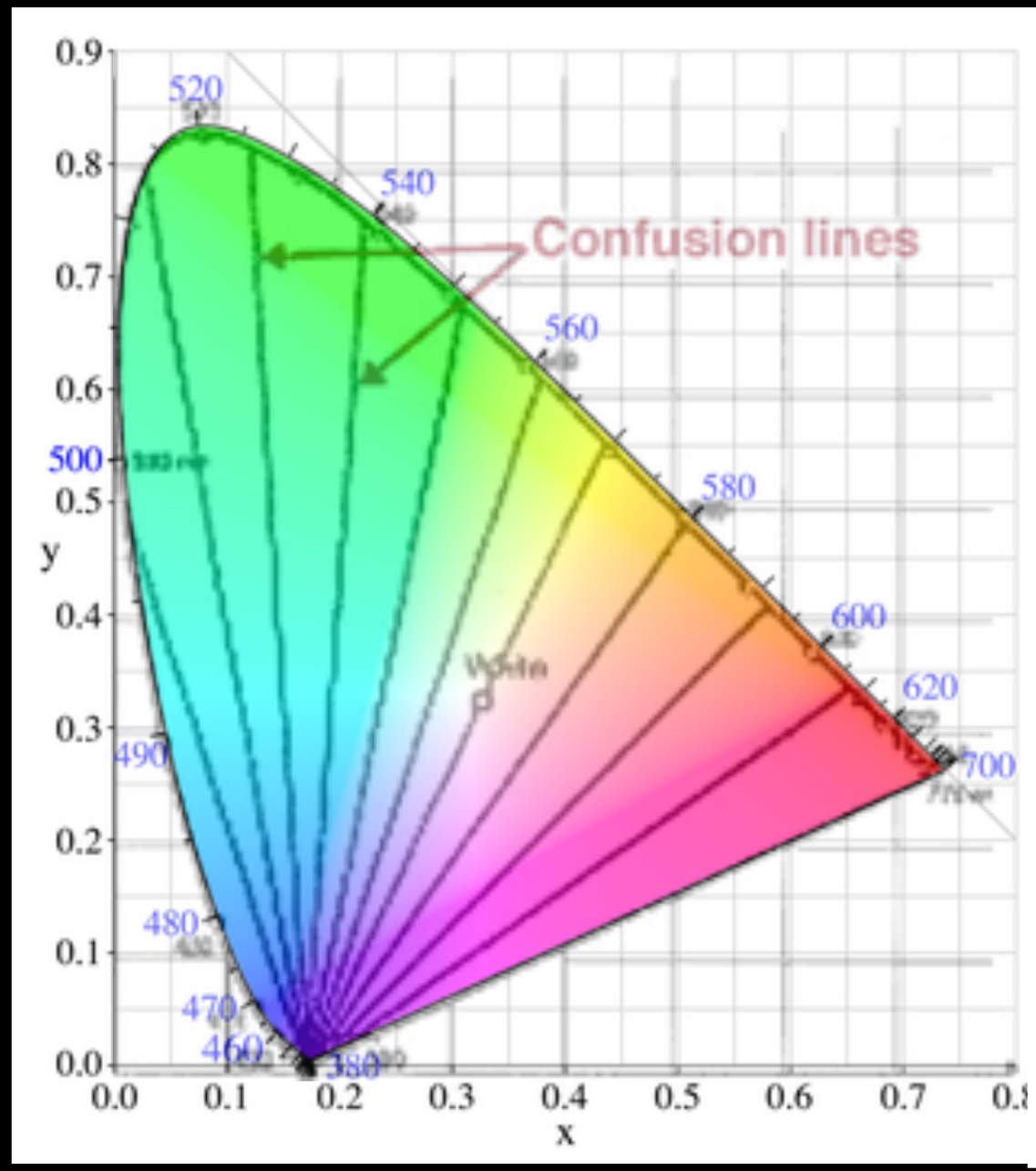
COLOR BLINDNESS

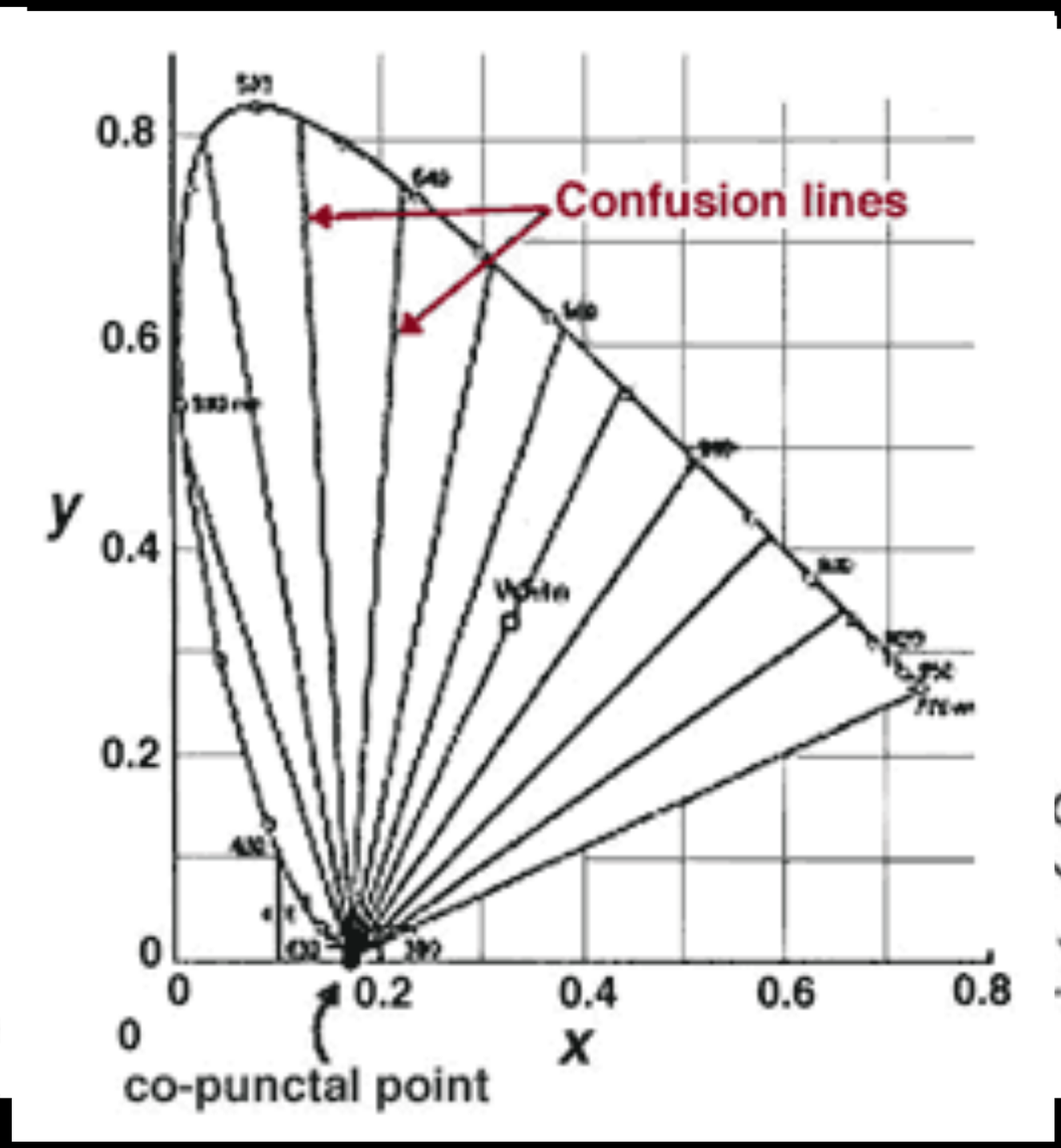




Deuteranopia
(green-blind)

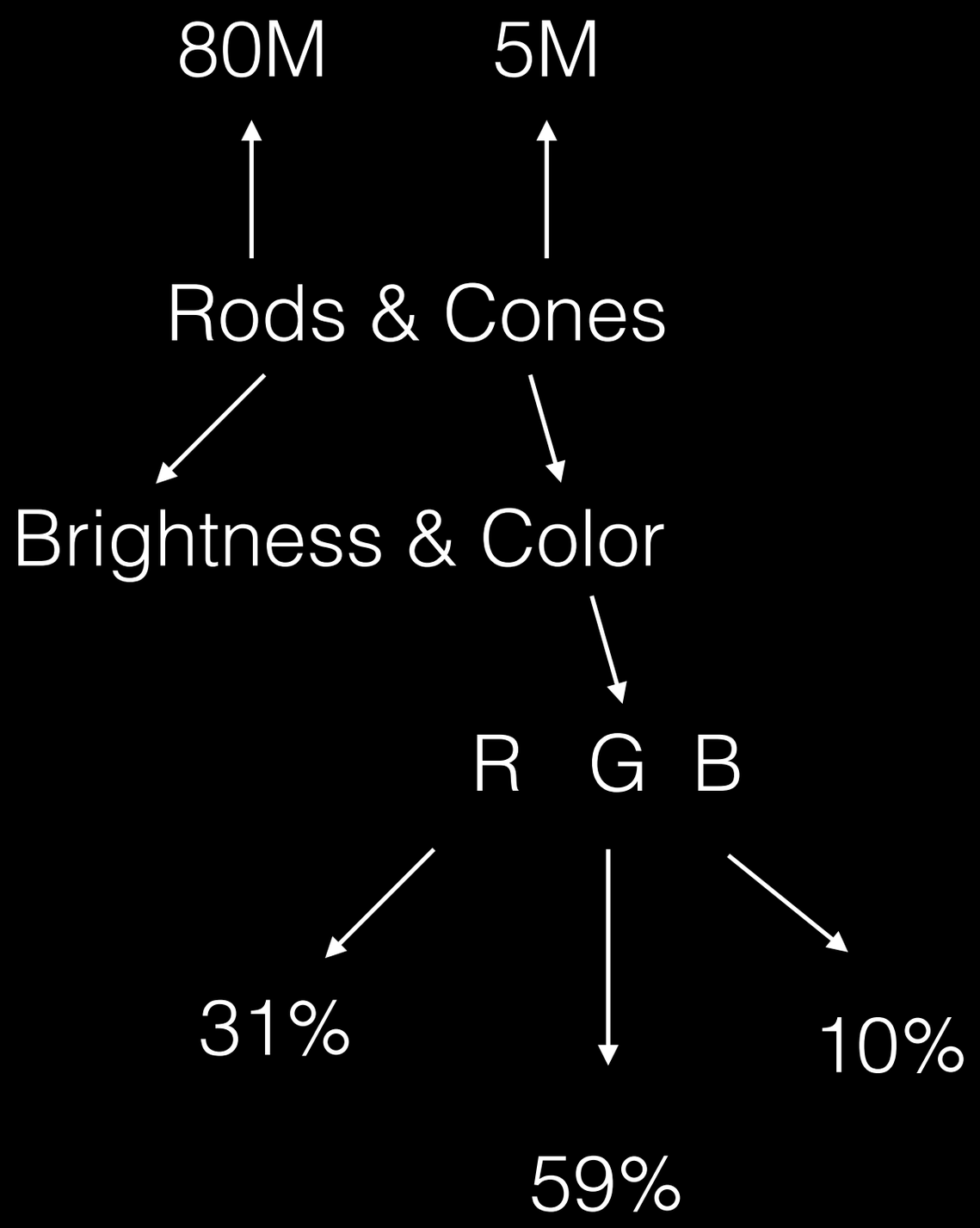
COLOR BLINDNESS

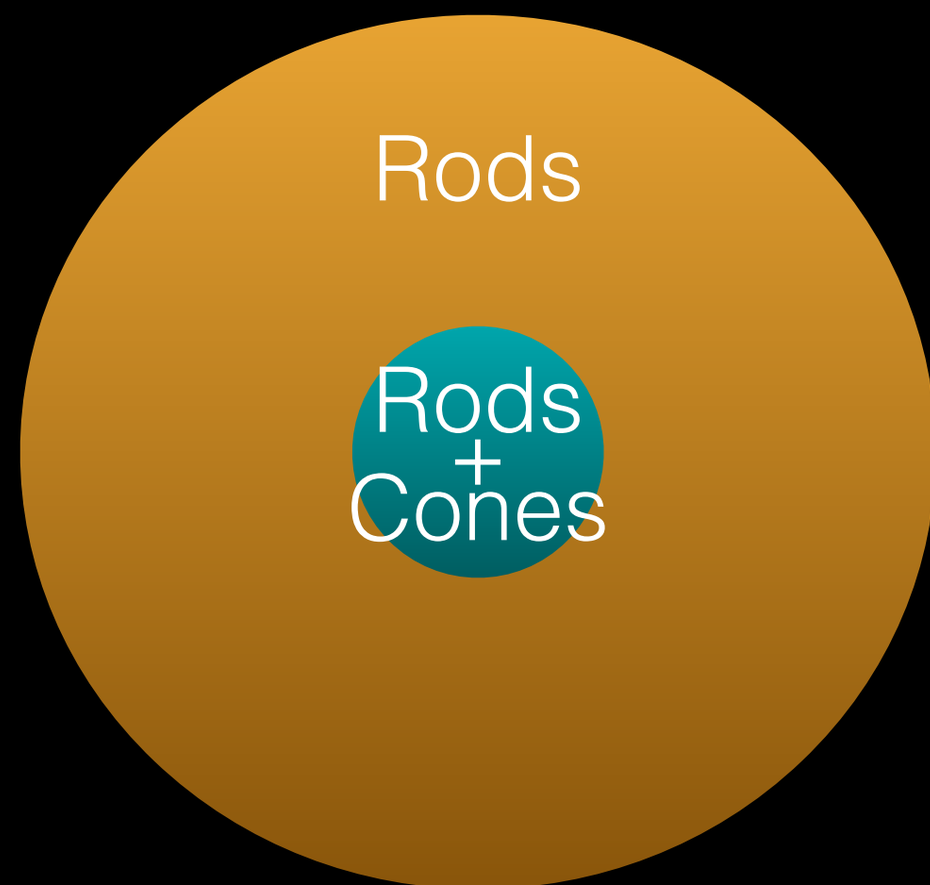




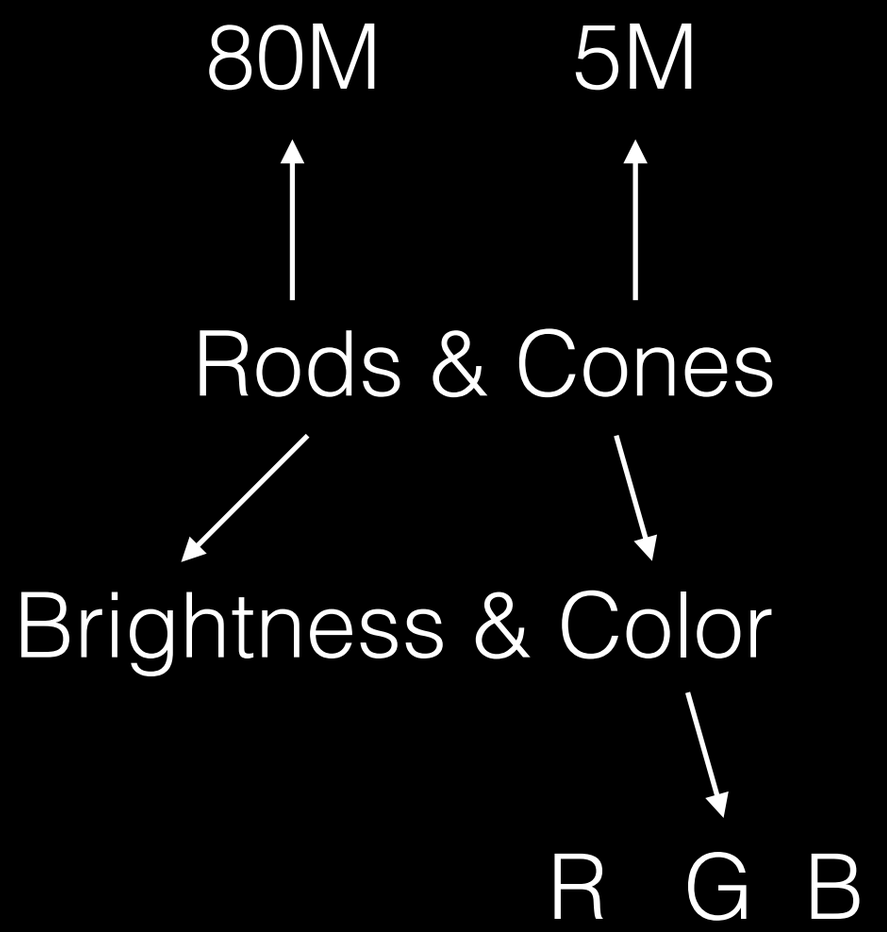
Tritanopia
(blue-blind)

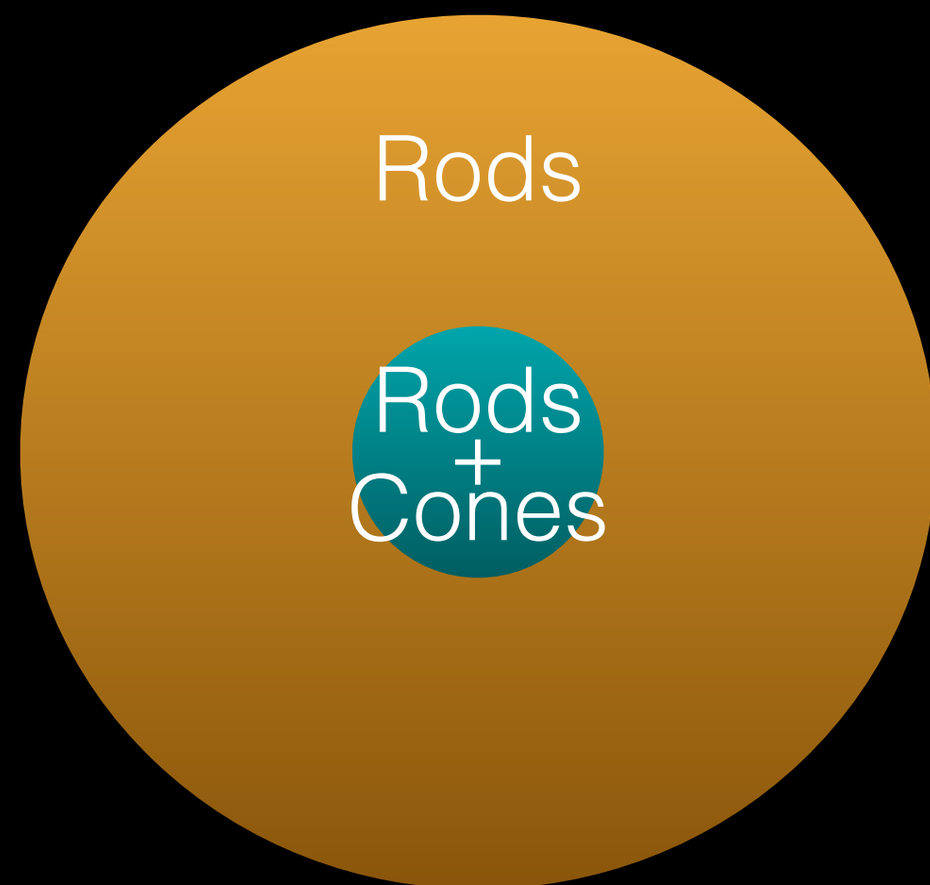
COLOR BLINDNESS





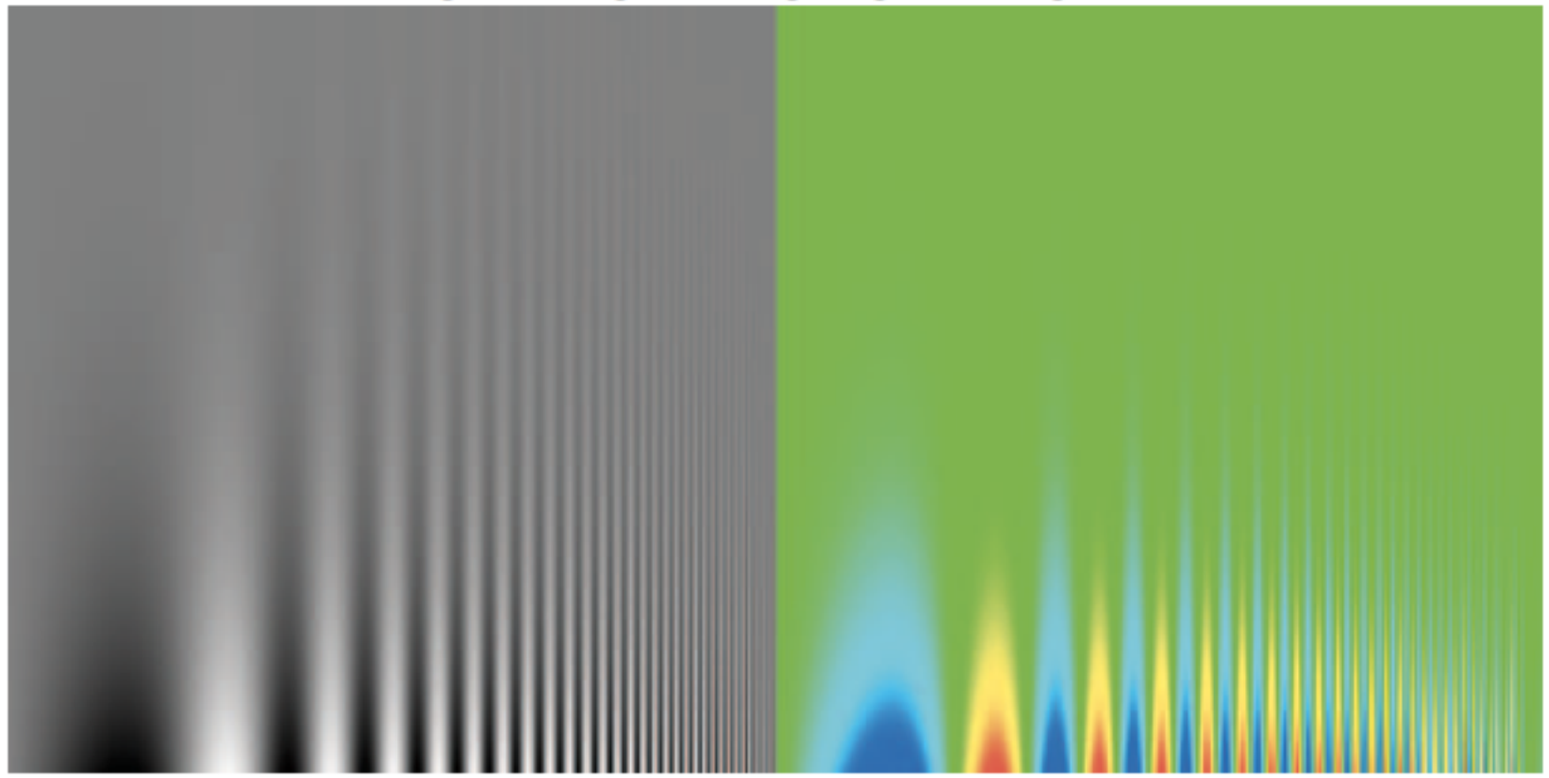
COLOR BLINDNESS





http://colororacle.org/

5. **Detail is actually harder to see in a rainbow.**
   The logic that it is easier to see detail in a range when you add colors seems to make sense, but in reality, more detail can be seen in a single hue image with a high brightness range.



(source)

http://blog.visual.ly/rainbow-color-scales/

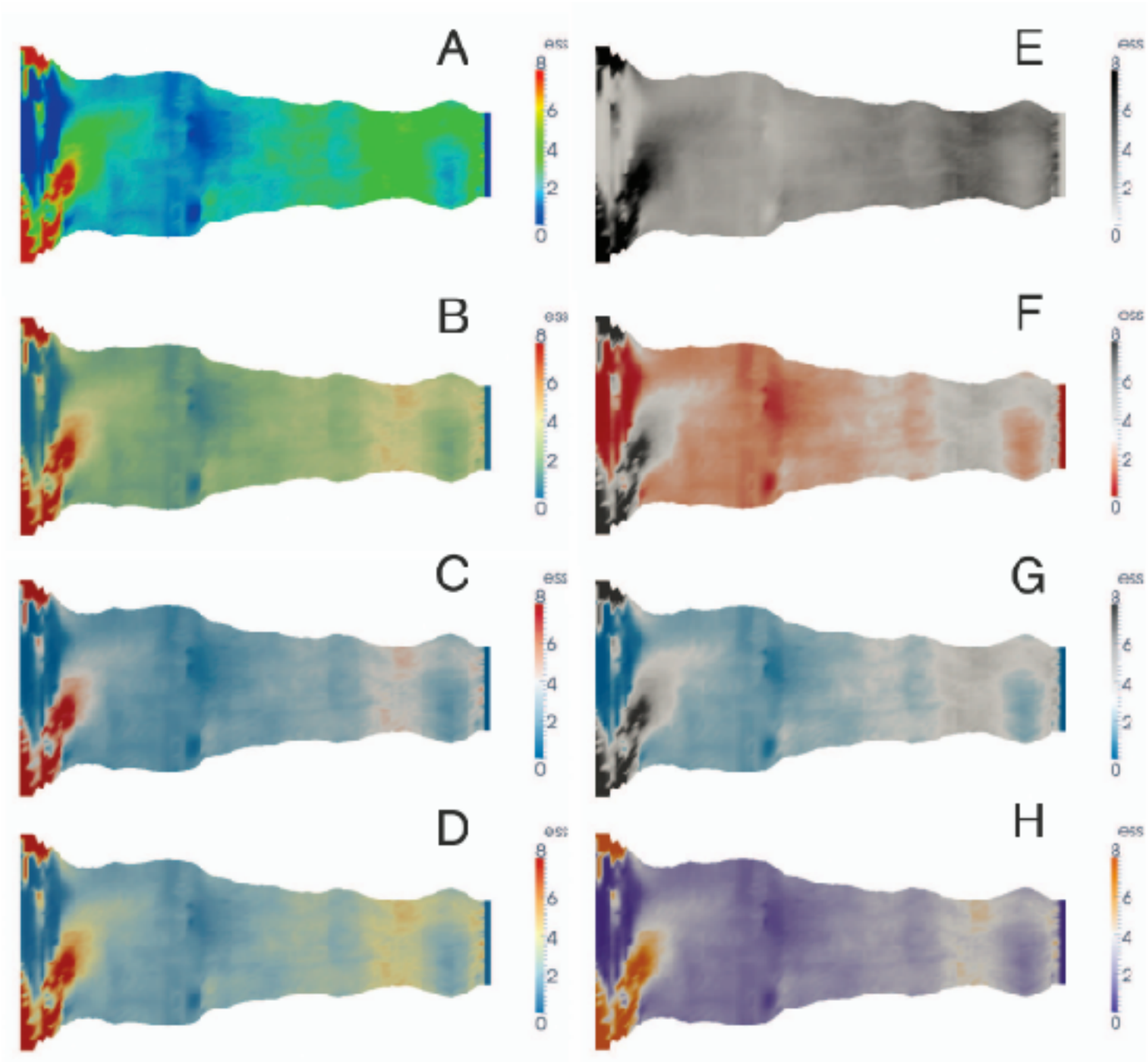


Fig. 4. Color schemes presented during the qualitative user study. The rainbow scheme (A) was preferred by most since it is what they are accustomed to viewing. The next most popular scheme was the red-black diverging scale (F). The grayscale image (E) was unanimously disliked since participants assume black-and-white images to be raw radiological data, while color indicates that the data has been processed or simulated.

http://www.eecs.harvard.edu/~kgajos/papers/2011/borkin11-infoviz.pdf

very real consequences of the choice of color maps (Borkin et al. 2011)

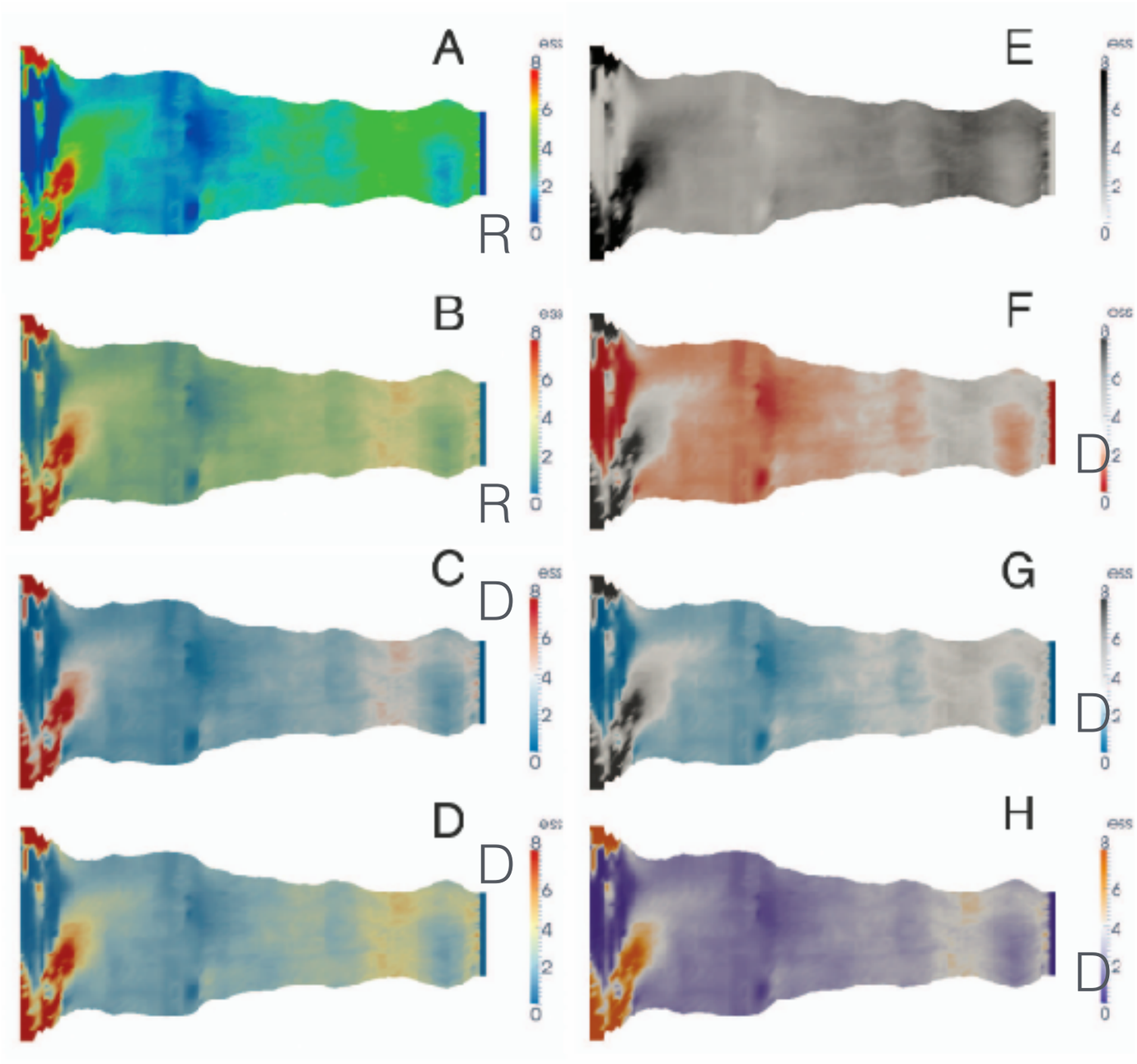


Fig. 4. Color schemes presented during the qualitative user study. The rainbow scheme (A) was preferred by most since it is what they are accustomed to viewing. The next most popular scheme was the red-black diverging scale (F). The grayscale image (E) was unanimously disliked since participants assume black-and-white images to be raw radiological data, while color indicates that the data has been processed or simulated.

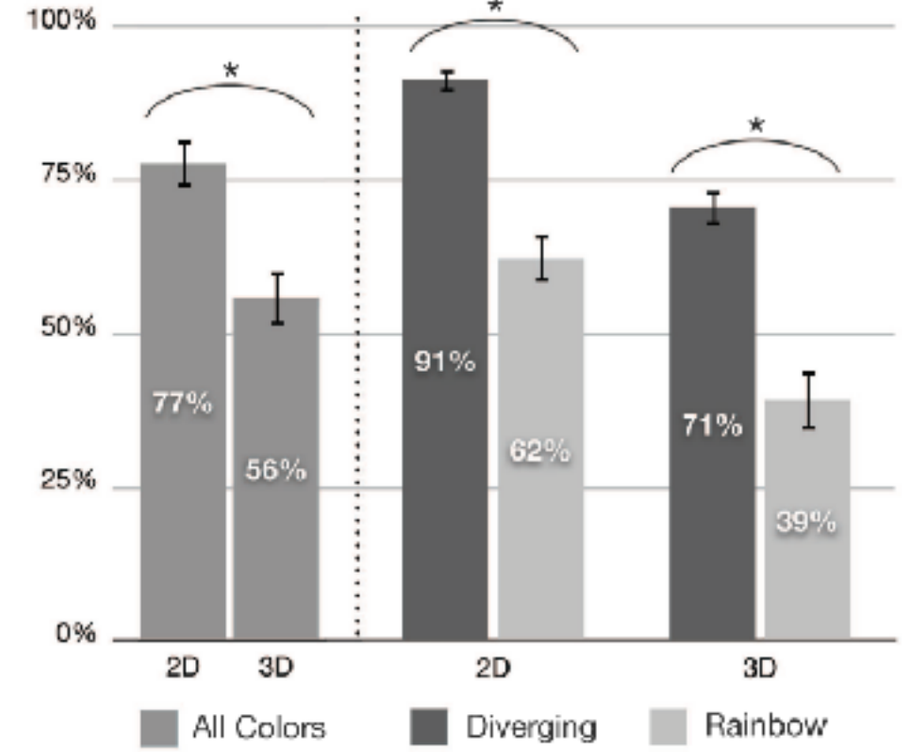


Fig. 7. **Average percent of low ESS regions identified** broken down by 2D and 3D representation, and color. Error bars correspond to the standard error and the asterisks indicate results of statistical significance. Participants were more accurate in 2D and when using the diverging color map.

http://www.eecs.harvard.edu/~kgajos/papers/2011/borkin11-infoviz.pdf

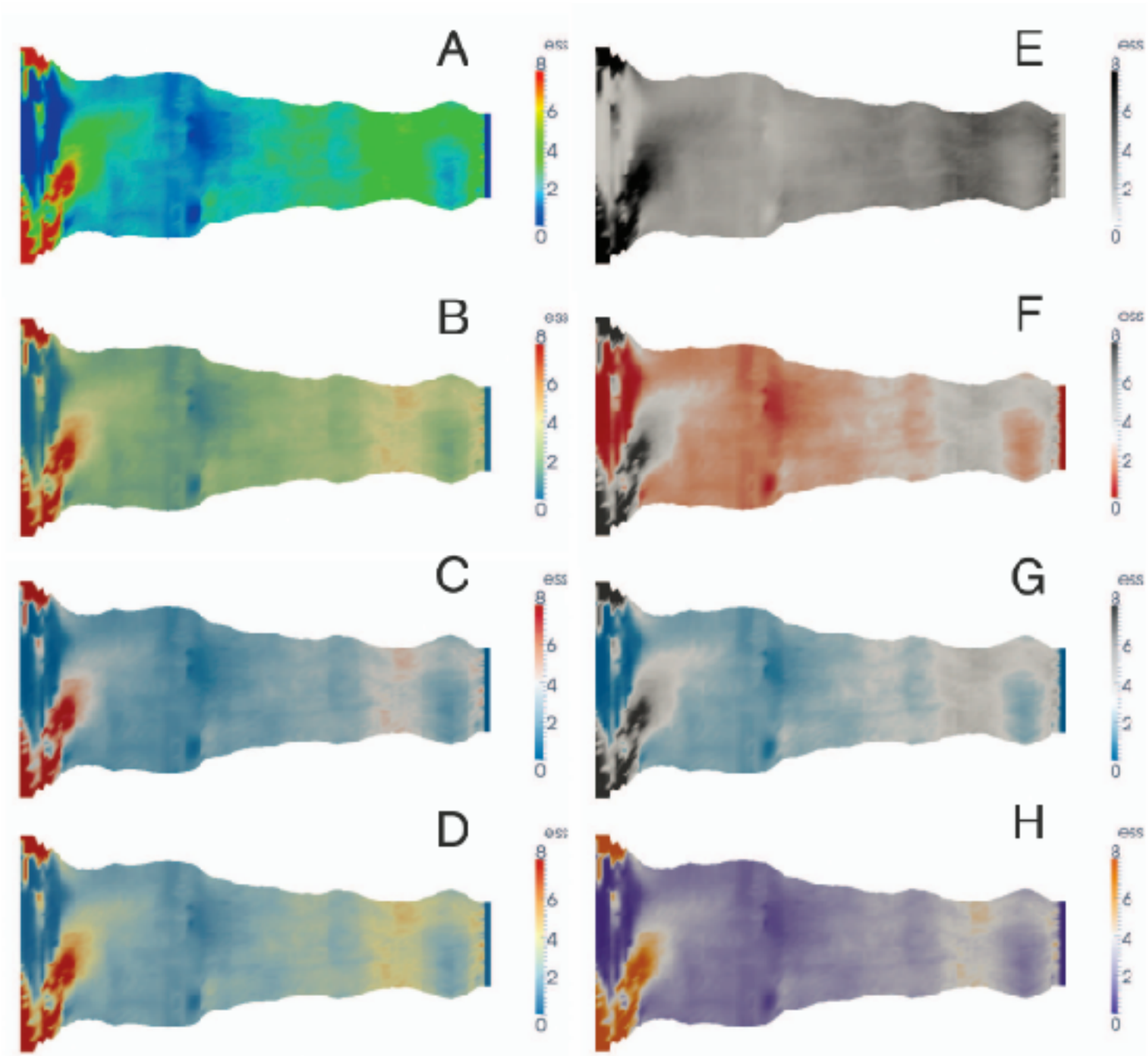


Fig. 4. Color schemes presented during the qualitative user study. The rainbow scheme (A) was preferred by most since it is what they are accustomed to viewing. The next most popular scheme was the red-black diverging scale (F). The grayscale image (E) was unanimously disliked since participants assume black-and-white images to be raw radiological data, while color indicates that the data has been processed or simulated.

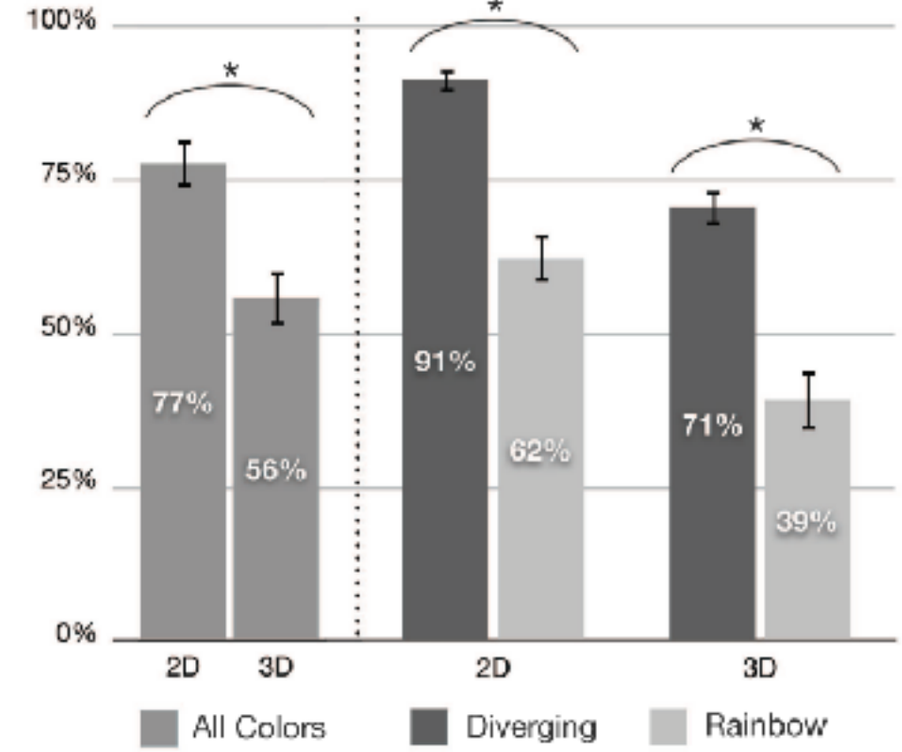


Fig. 7. **Average percent of low ESS regions identified** broken down by 2D and 3D representation, and color. Error bars correspond to the standard error and the asterisks indicate results of statistical significance. Participants were more accurate in 2D and when using the diverging color map.

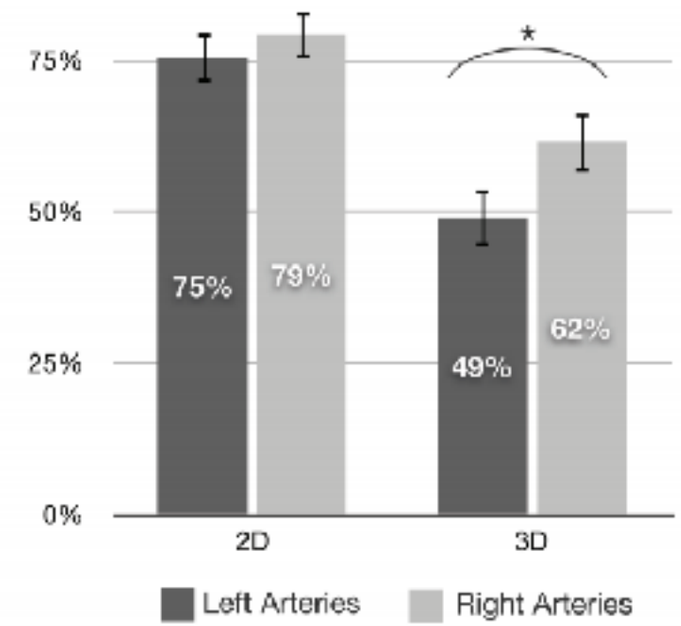
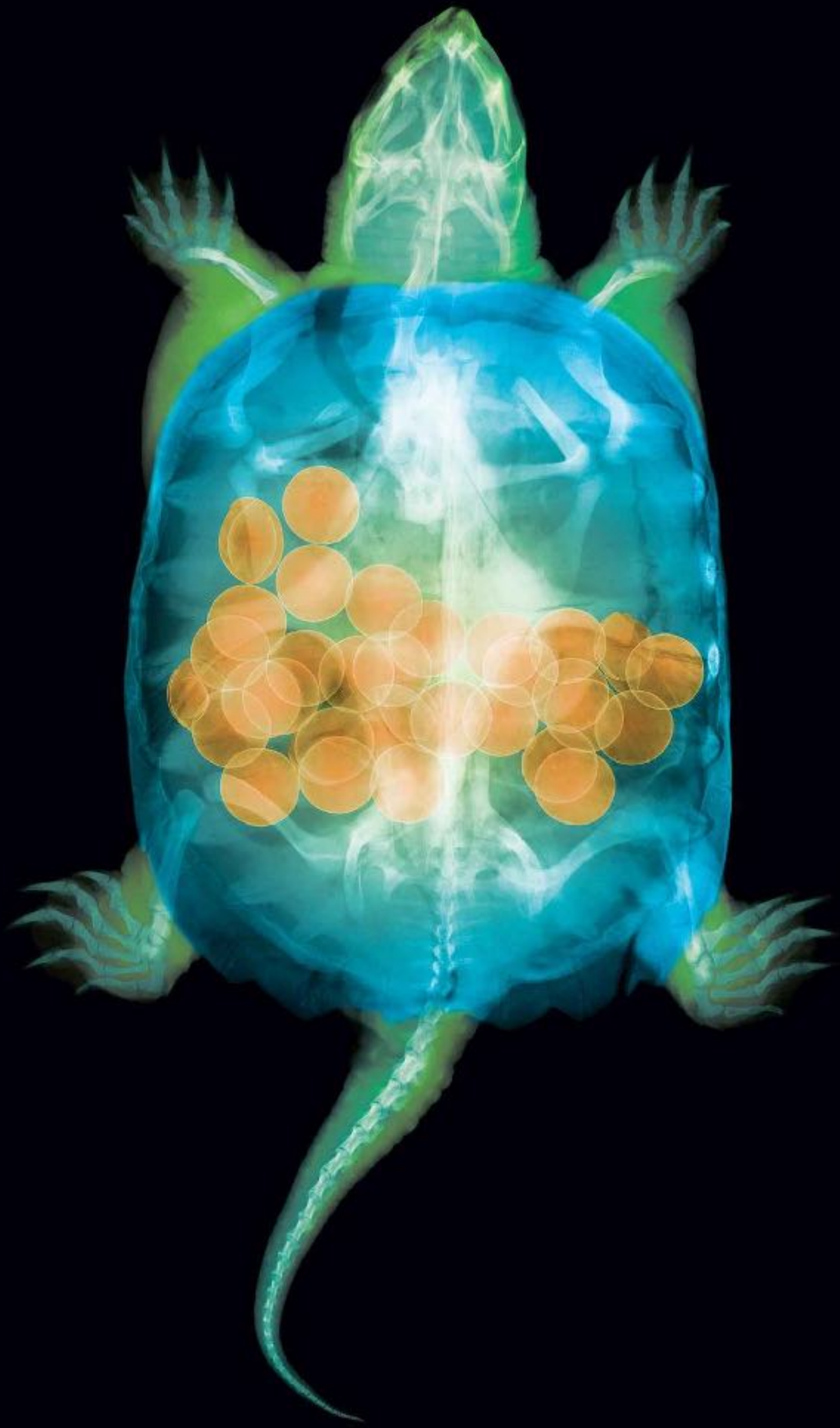
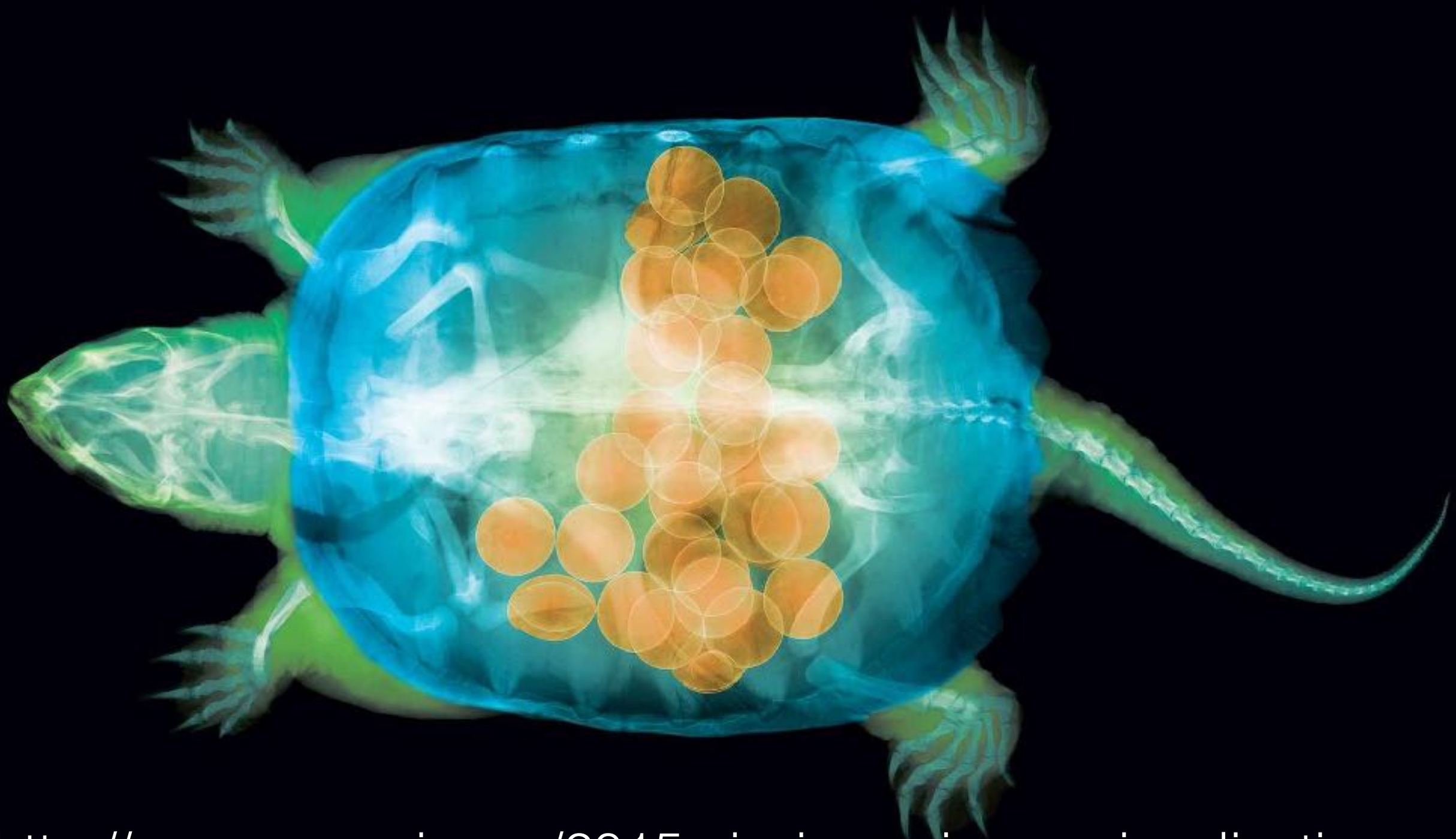


Fig. 8. **Average percent of low ESS regions identified** broken down by 2D and 3D representation, and left and right artery systems. Error bars correspond to the standard error and the asterisks indicate results of statistical significance. In 3D, users were less accurate identifying regions in the most complex data sets (i.e., left artery systems). Whereas in 2D, performance was the same regardless of task complexity.

http://www.eecs.harvard.edu/~kgajos/papers/2011/borkin11-infoviz.pdf





http://www.popsci.com/2015-vizzies-science-visualizations-video-images?image=0

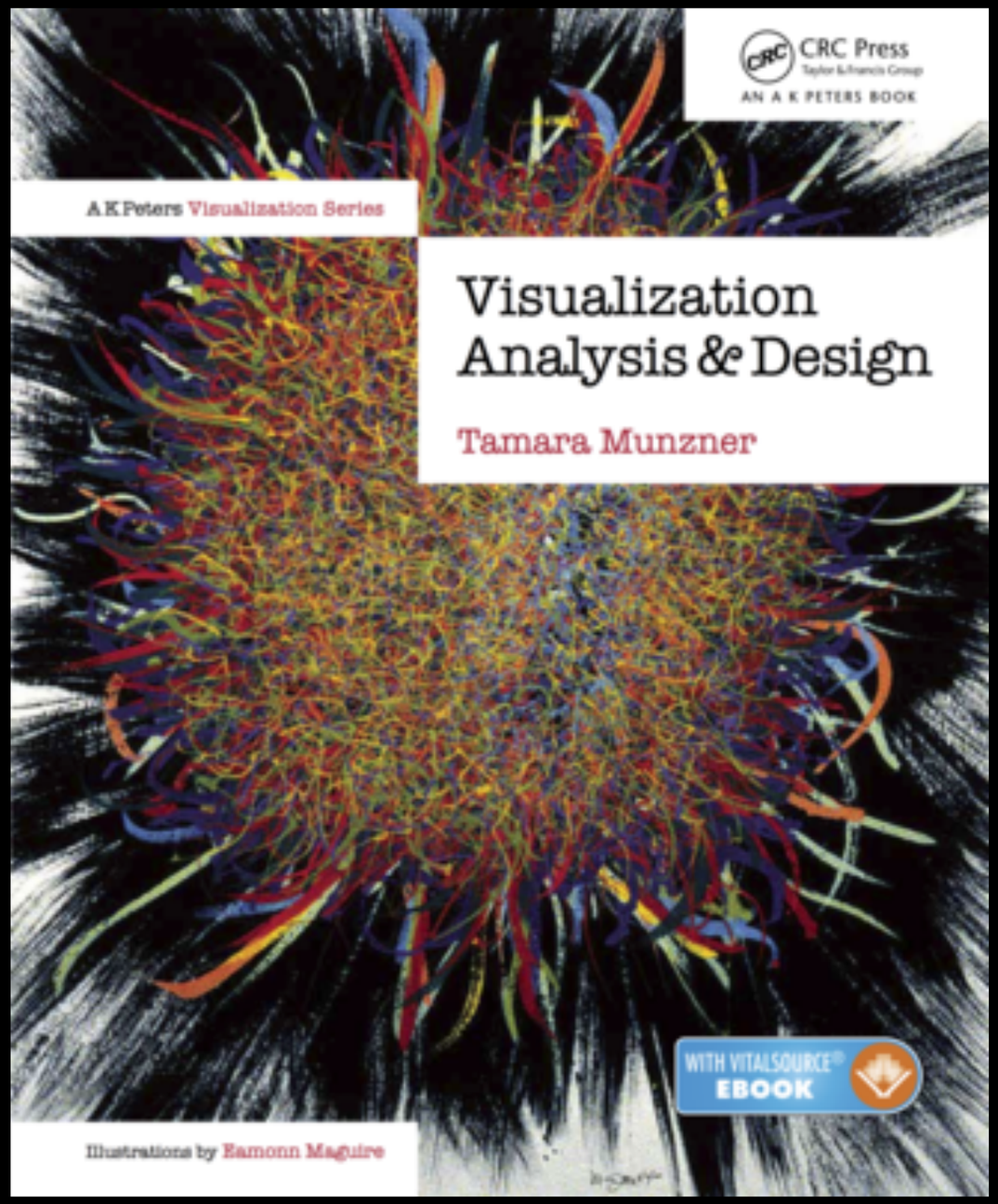
Matplotlib 2.0 color map change

http://matplotlib.org/style_changes.html

Tamara Munzner

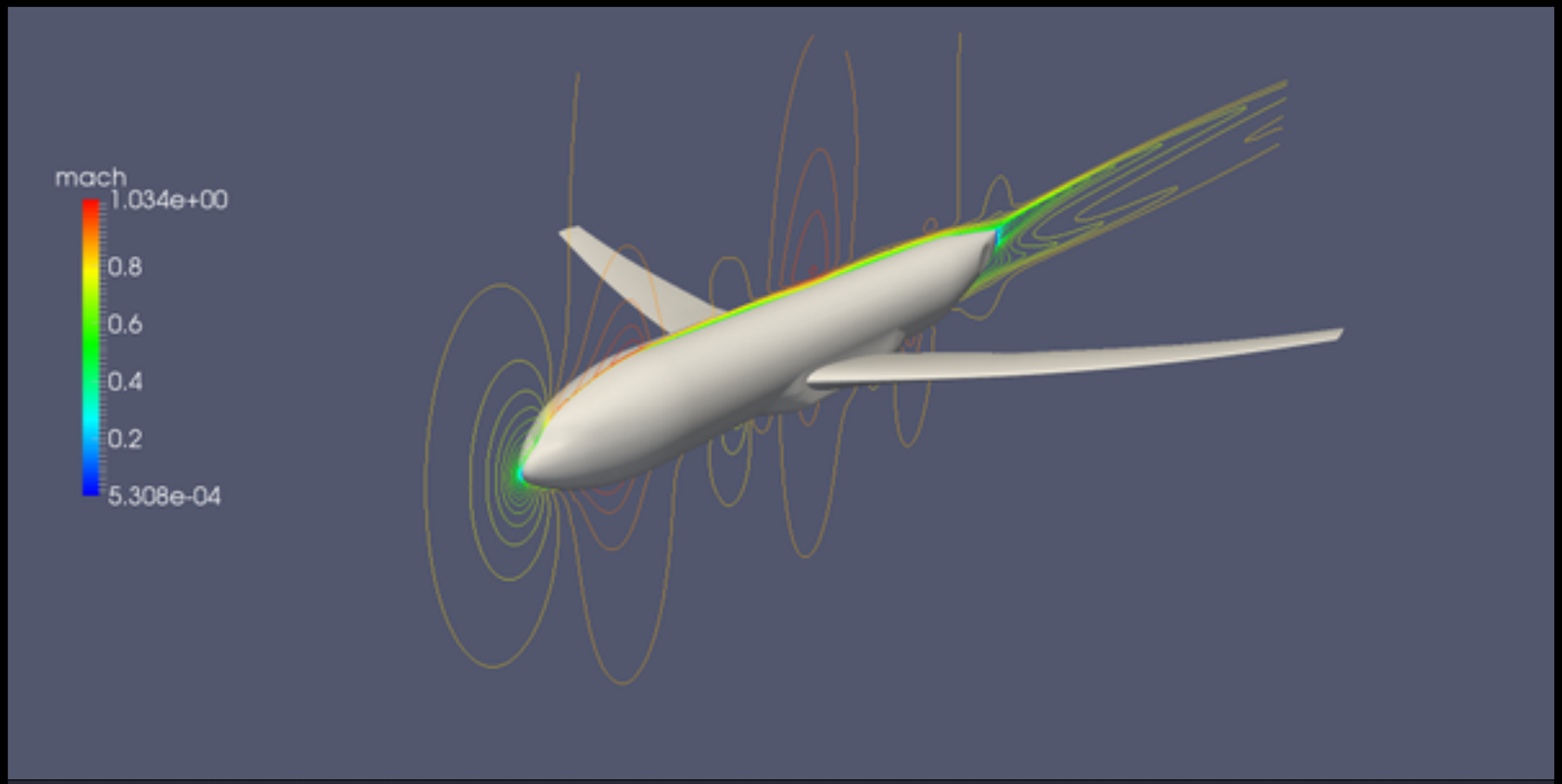
Rules of thumb

Chap 6



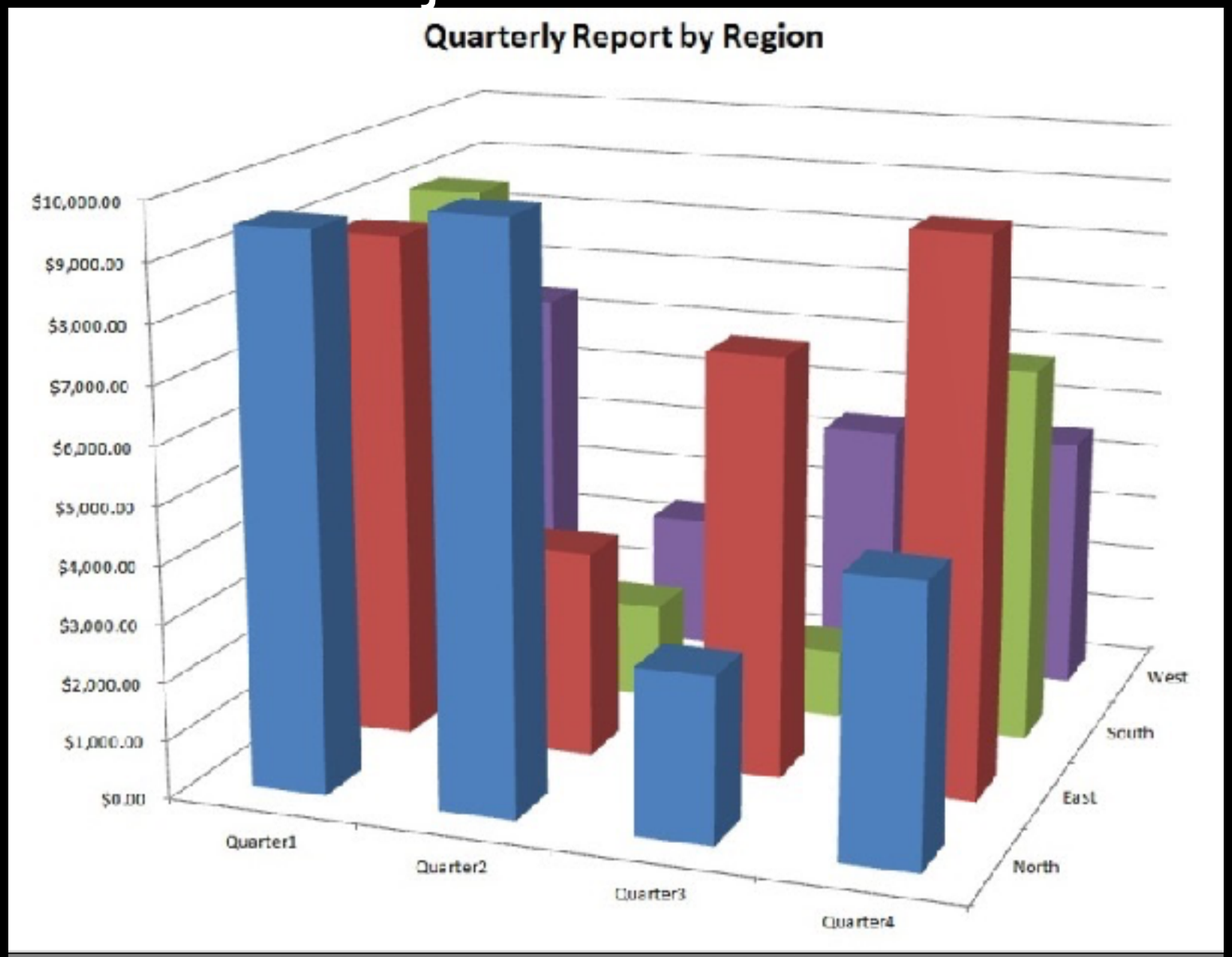
let’s improve a plot

No Unjustified 3D

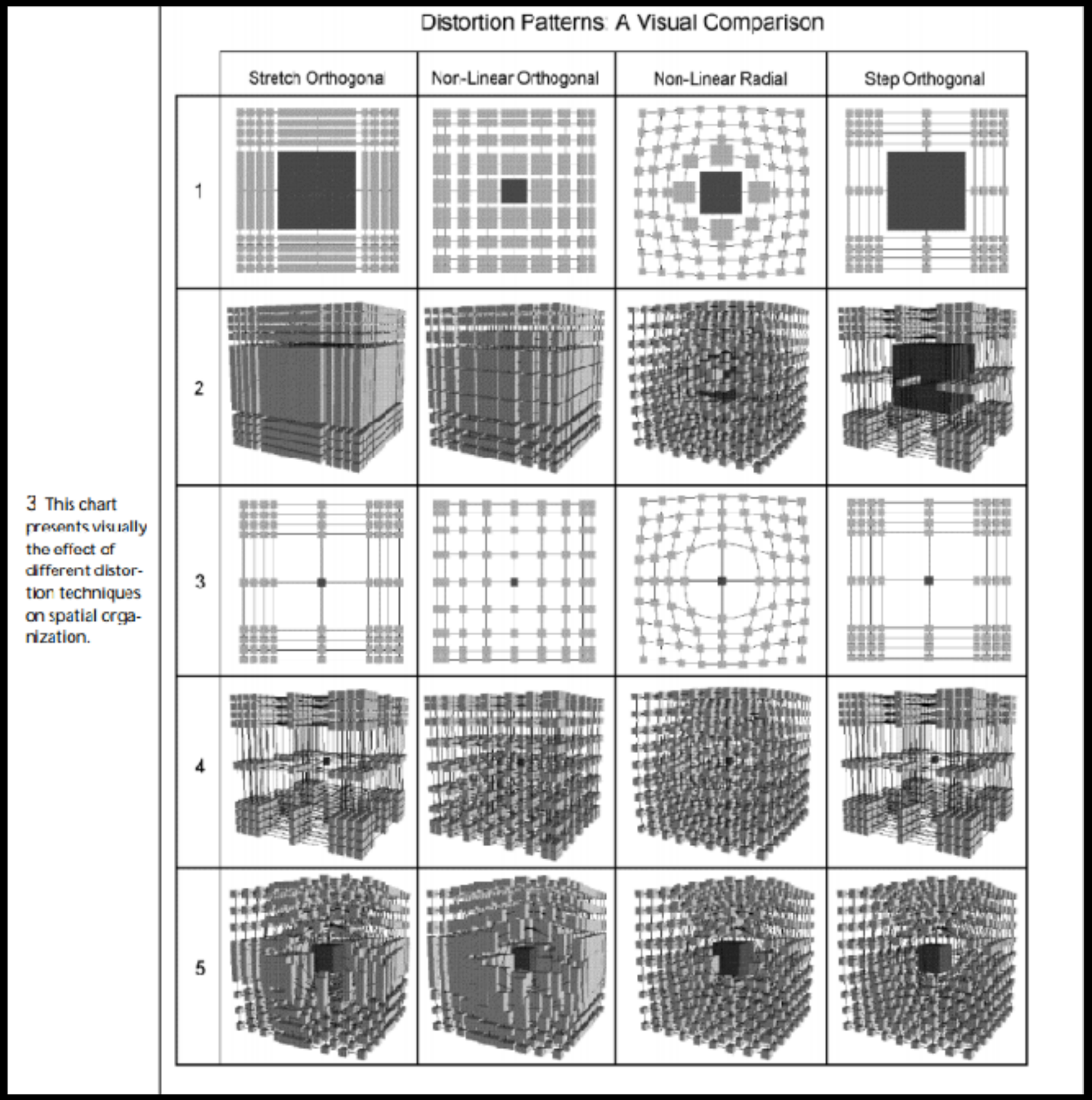


No Unjustified 3D

unjustified 3D

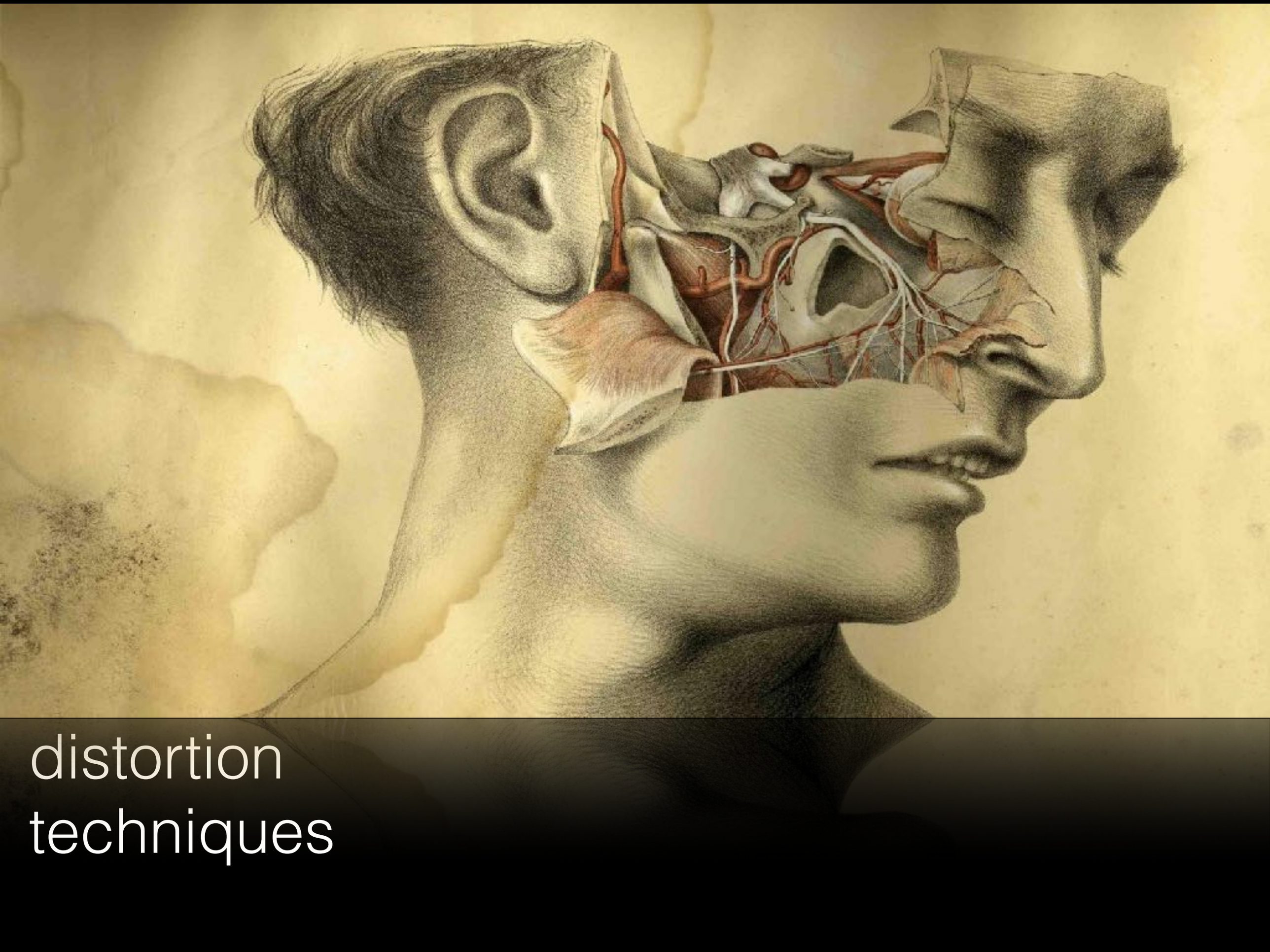


distortion techniques



3 This chart presents visually the effect of different distortion techniques on spatial organization.

Extending Distortion Viewing Techniques from 2D to 3D Data.
Carpendale et al. CG&A 17(4):42-51, July 1997



distortion techniques

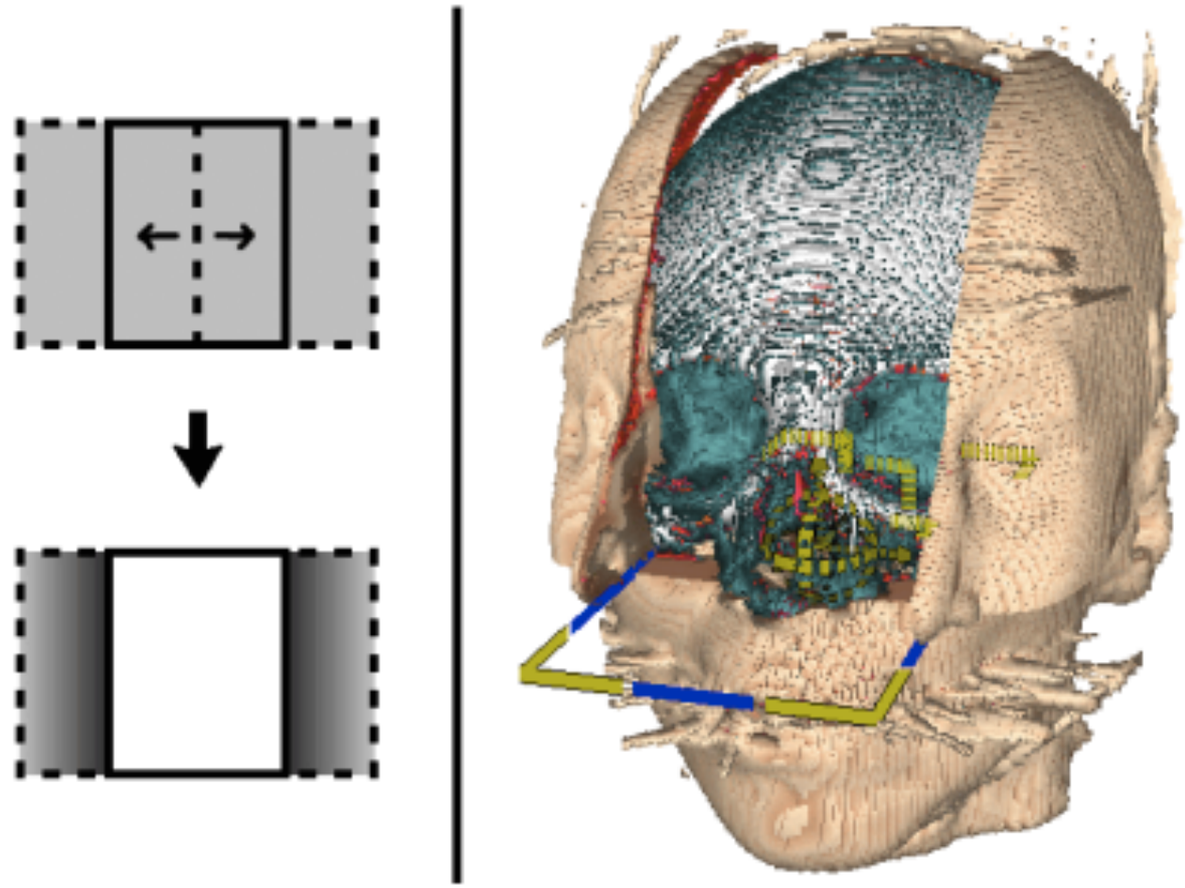


Figure 3.11: The Box Spreader. Left: voxels before and after deformation. All voxels contained in the box are pushed sideways, compressing surrounding voxels that are within twice the box's width. Right: the Box Spreader, acting only on layers outside the skull, cuts the upper face in half and pushes each half off the skull.

distortion techniques

An Investigation of Issues and Techniques in Highly Interactive Computational Visualization
Michael John MCGuffin

Mercator

https://en.wikipedia.org/wiki/List_of_map_projections

Mollweide

distortion
techniques

3D view

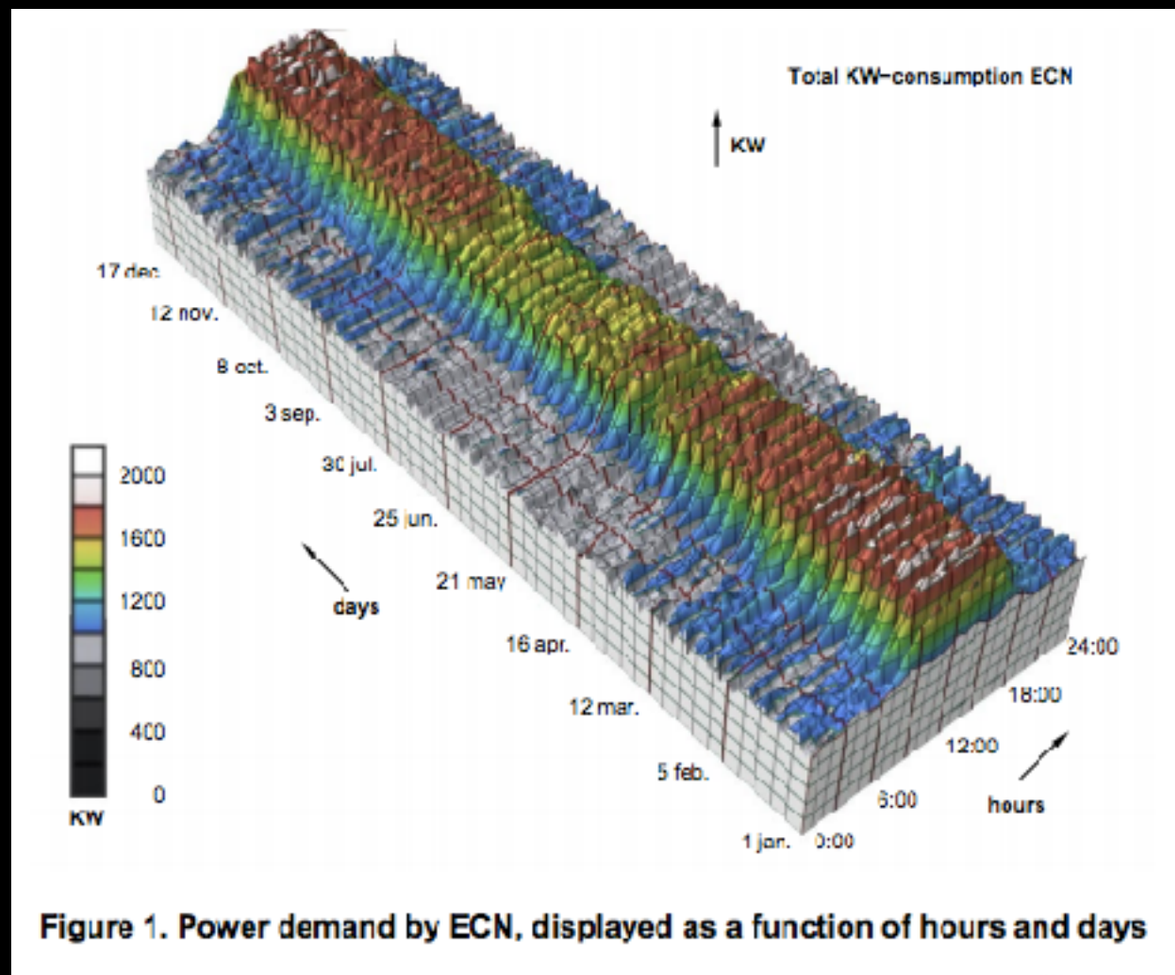


Figure 1. Power demand by ECN, displayed as a function of hours and days

Jarke J. van Wijk & Edward R. van Selow1999

Cluster View

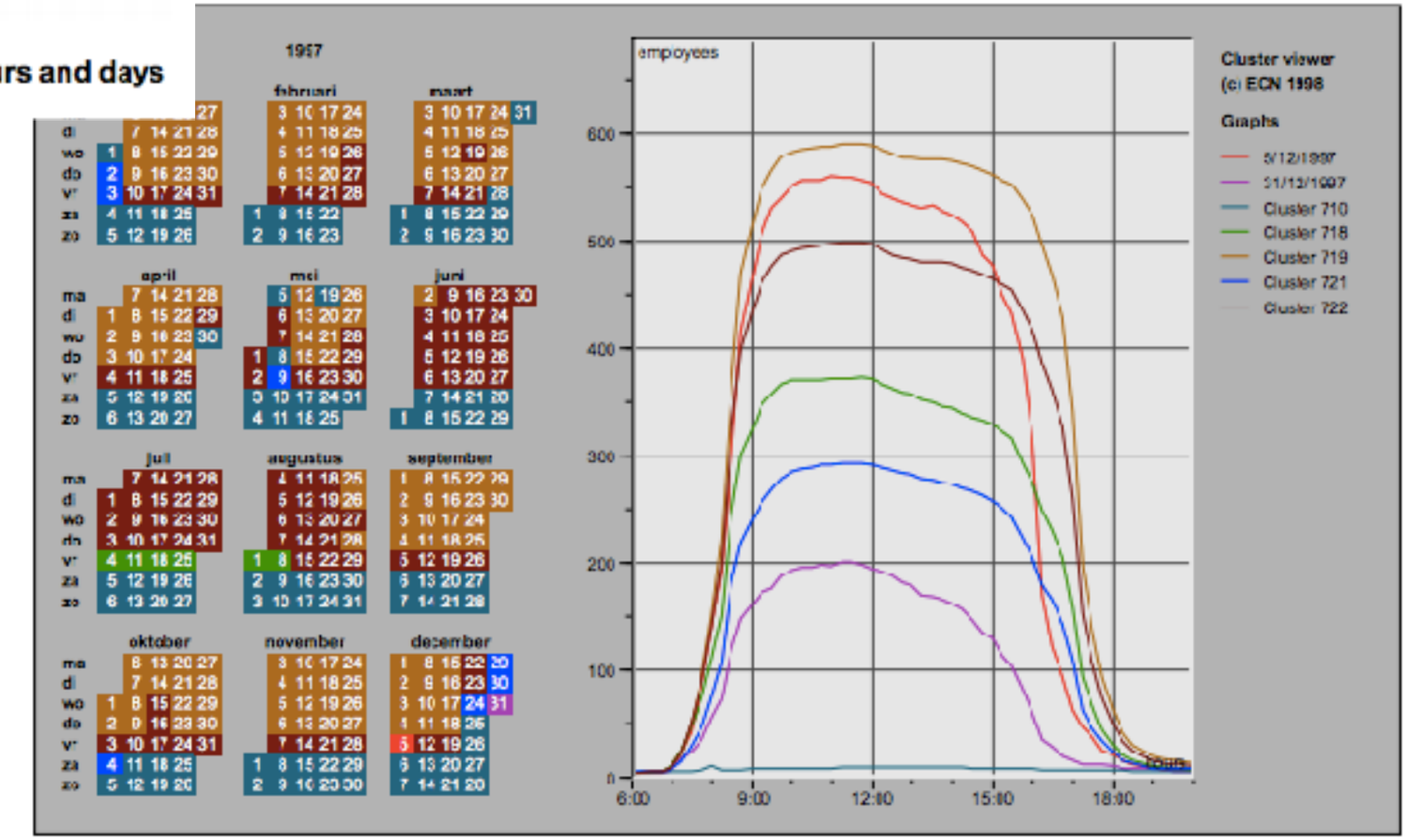
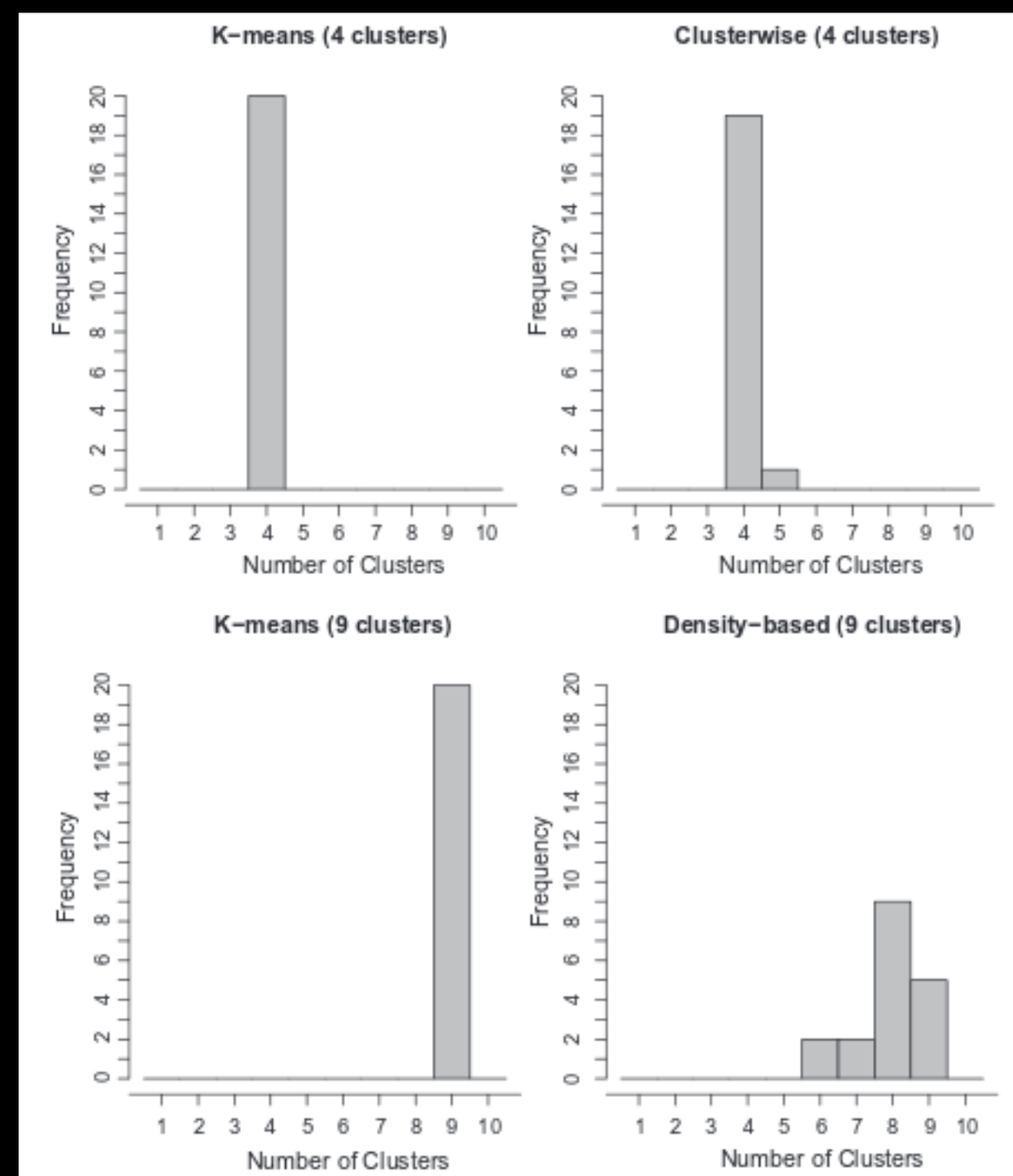


Figure 4. Calendar view of the number of employees

Also:

No Unjustified 2D!

consider lists when appropriate



Eyes over Memory

Eyes over Memory

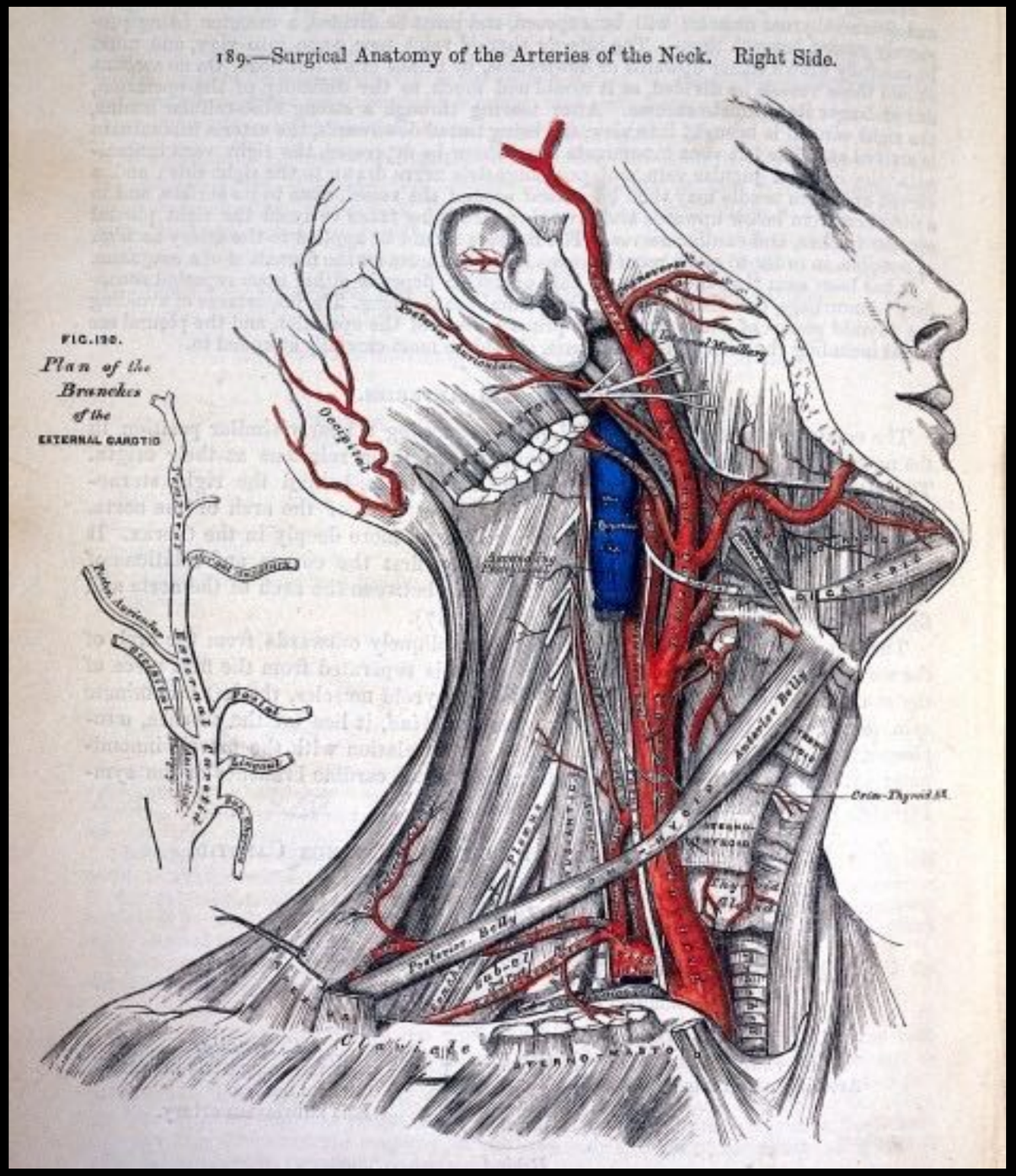
Eyes over Memory

no unjustified animation

Get it right in Black & White

no unjustified color

Maureen Stone



Maureen Stone

http://www.columbia.edu/~brennan/subway/SubDia.pdf

Function first,
Form next

no unjustified beauty

interactive visualization rules of thumb:

Resolution over immersion

interactive visualization rules of thumb:

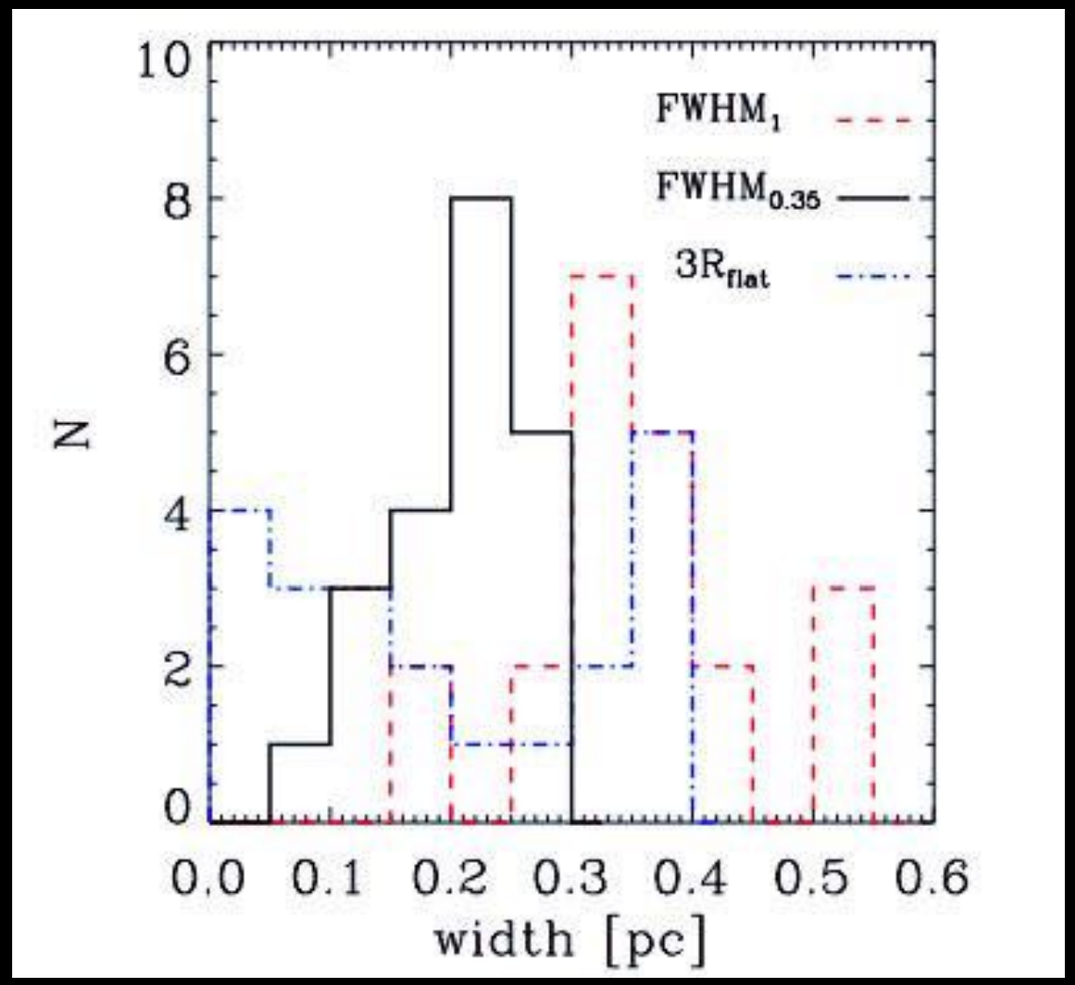
Details on demand

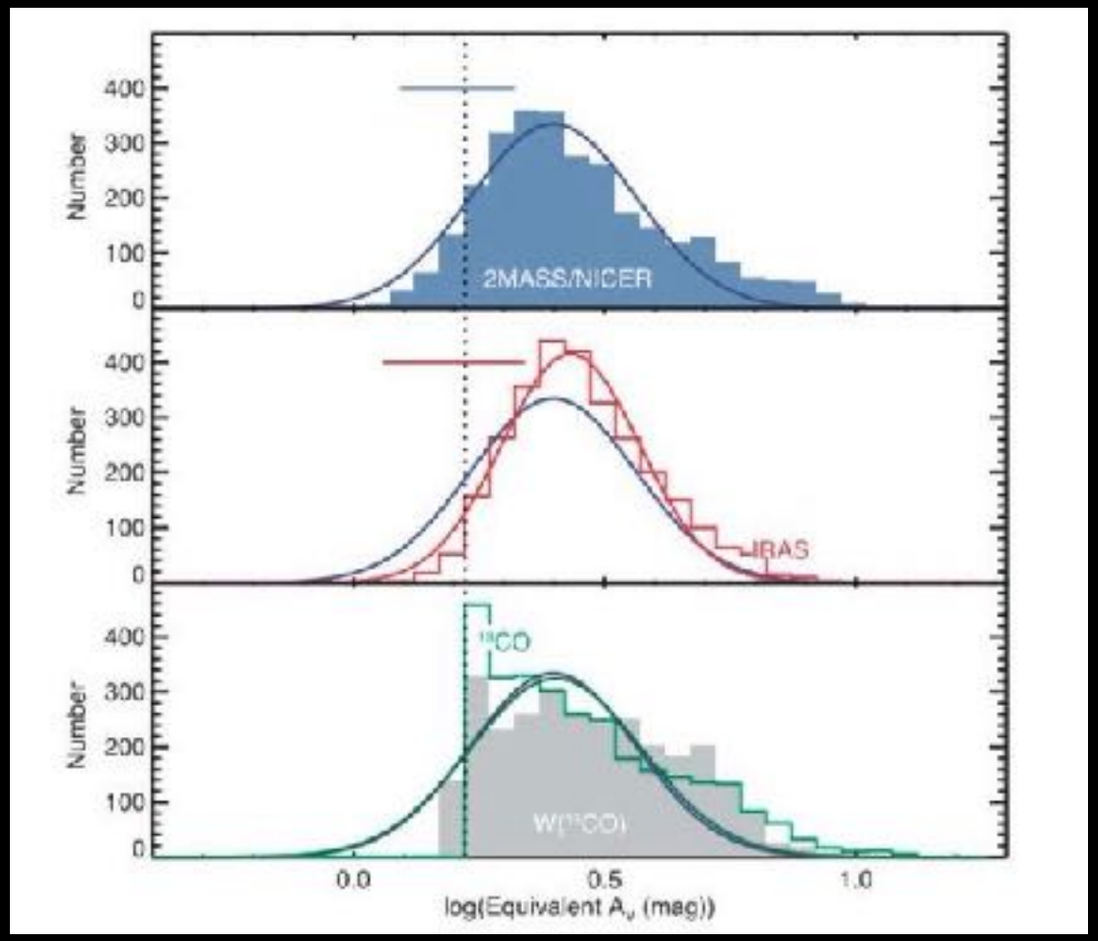
interactive visualization rules of thumb:

Avoid latency

let me add:

AVOID CLUTTER


FWHM$_1$
FWHM$_{0.35}$
3R$_{flat}$
N
width [pc]


Number
2MASS/NICER
IRAS
^{13}CO
W(^{13}CO)
log(Equivalent A$_v$ (mag))

3D view

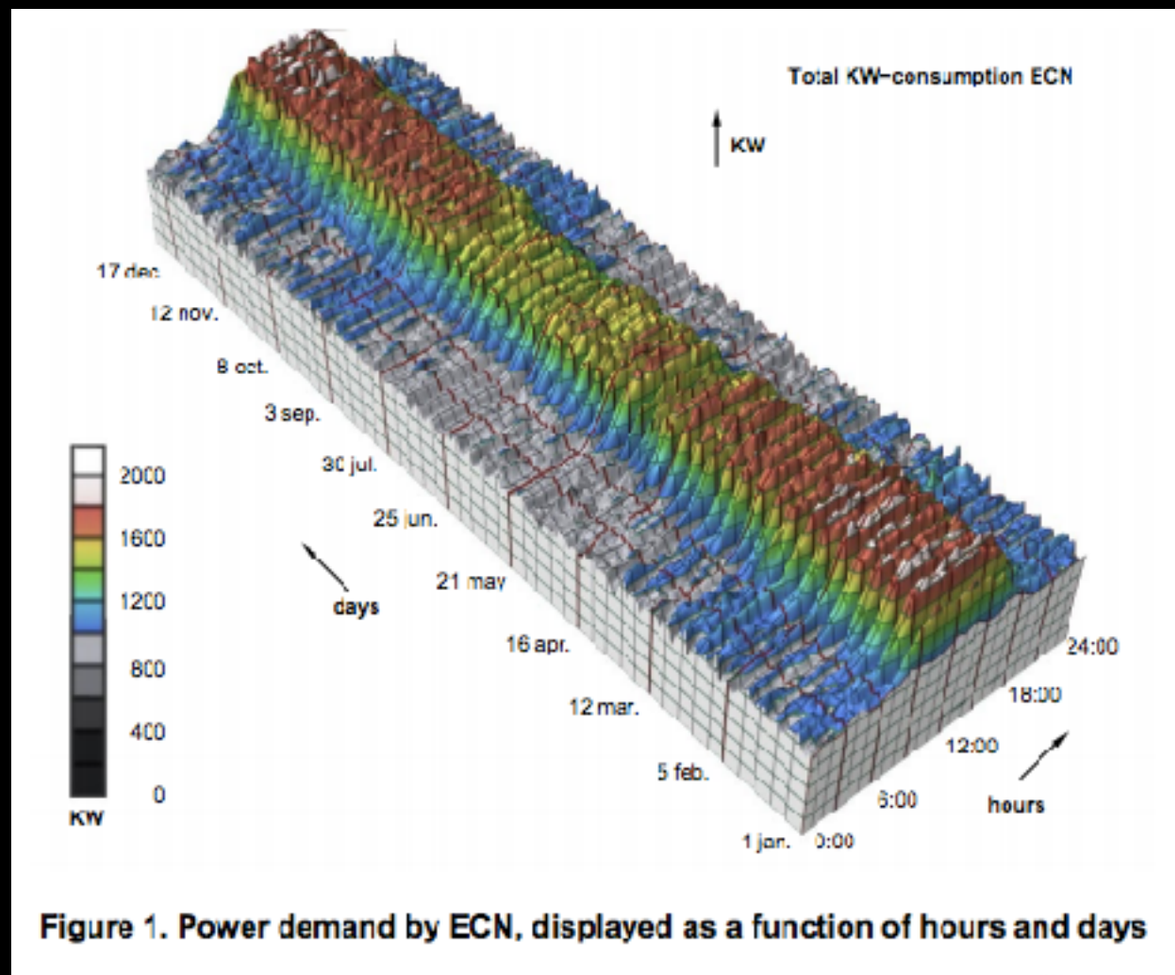


Figure 1. Power demand by ECN, displayed as a function of hours and days

Jarke J. van Wijk & Edward R. van Selow1999

Cluster View

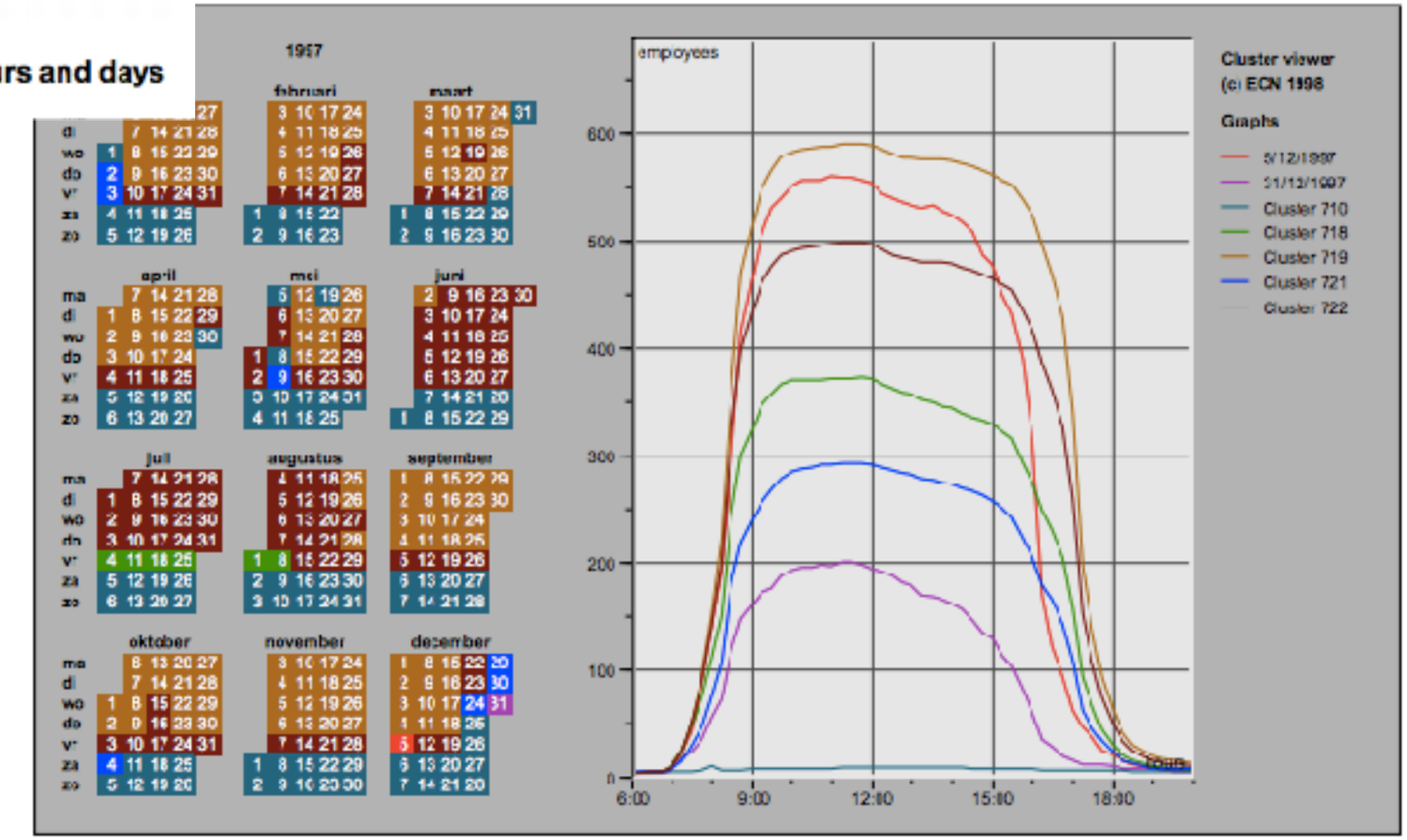
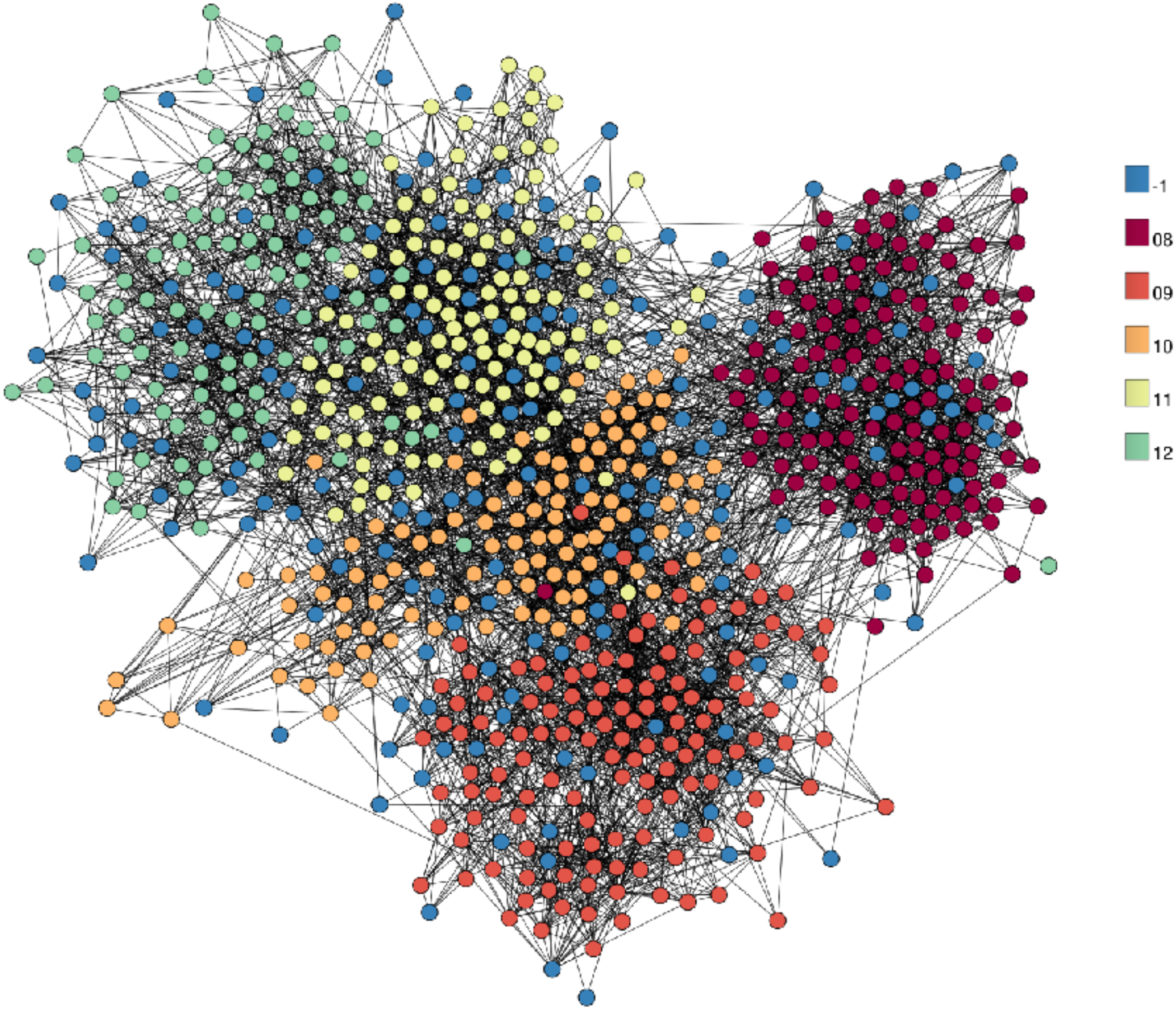


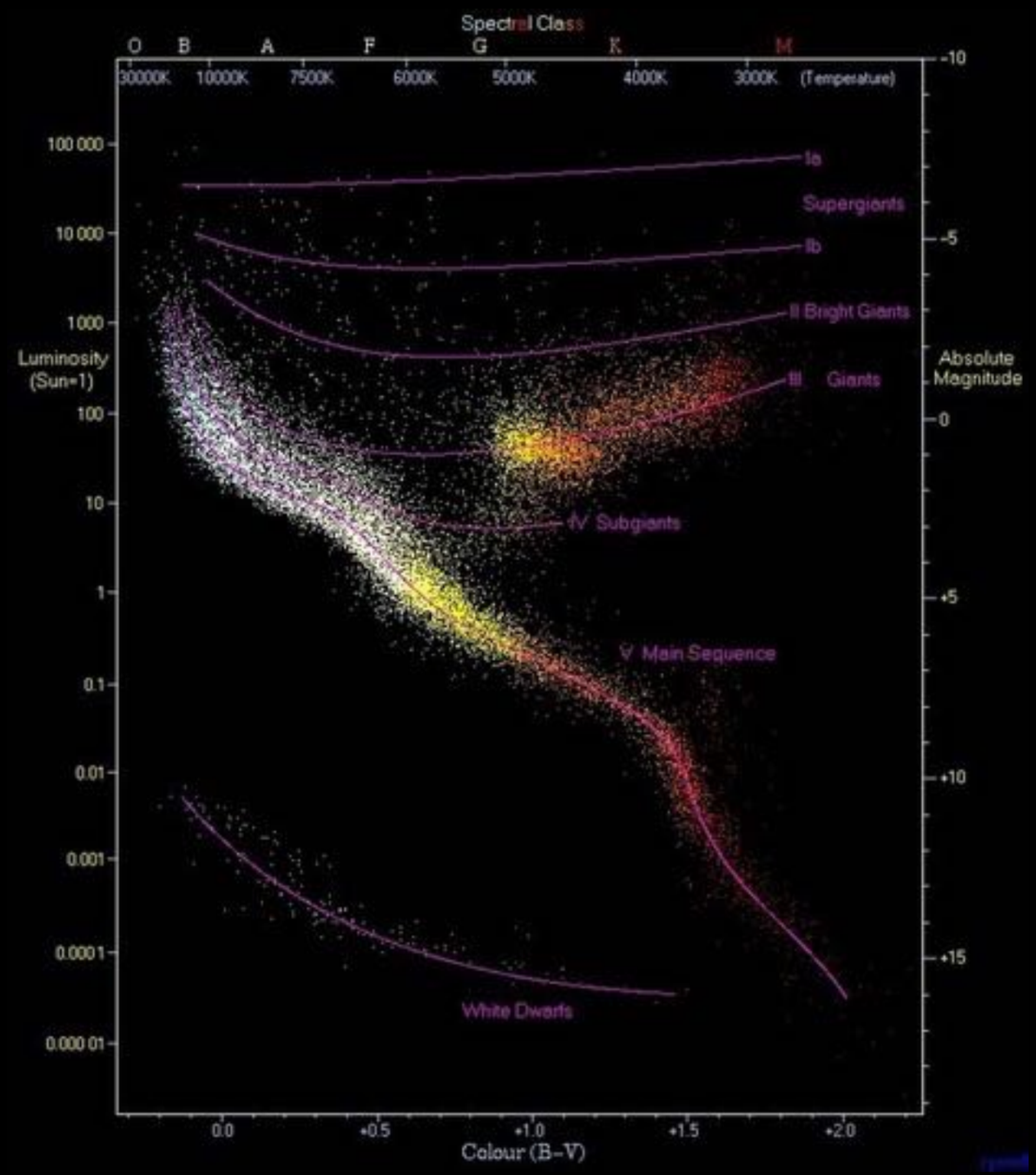
Figure 4. Calendar view of the number of employees

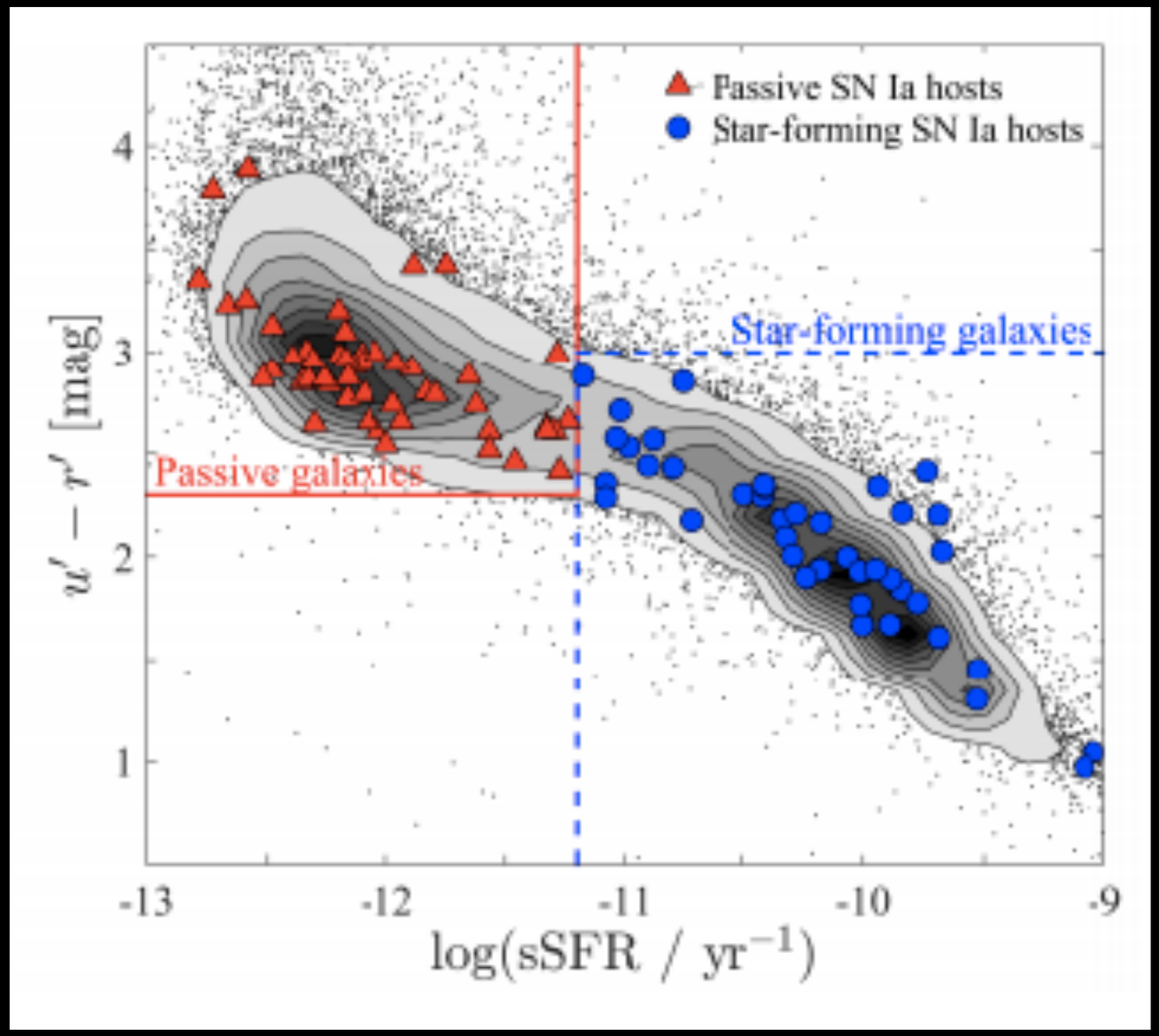

-1
08
09
10
11
12

Reduction

http://cosmo.nyu.edu/~fb55/vizs/astrotrend/arxiv2.html

Life of a Star





Or, Bianco, Modjaz, ApJ 2014

cornerplot.ipynb

scatter_contour.ipynb

The infamous rule of 7

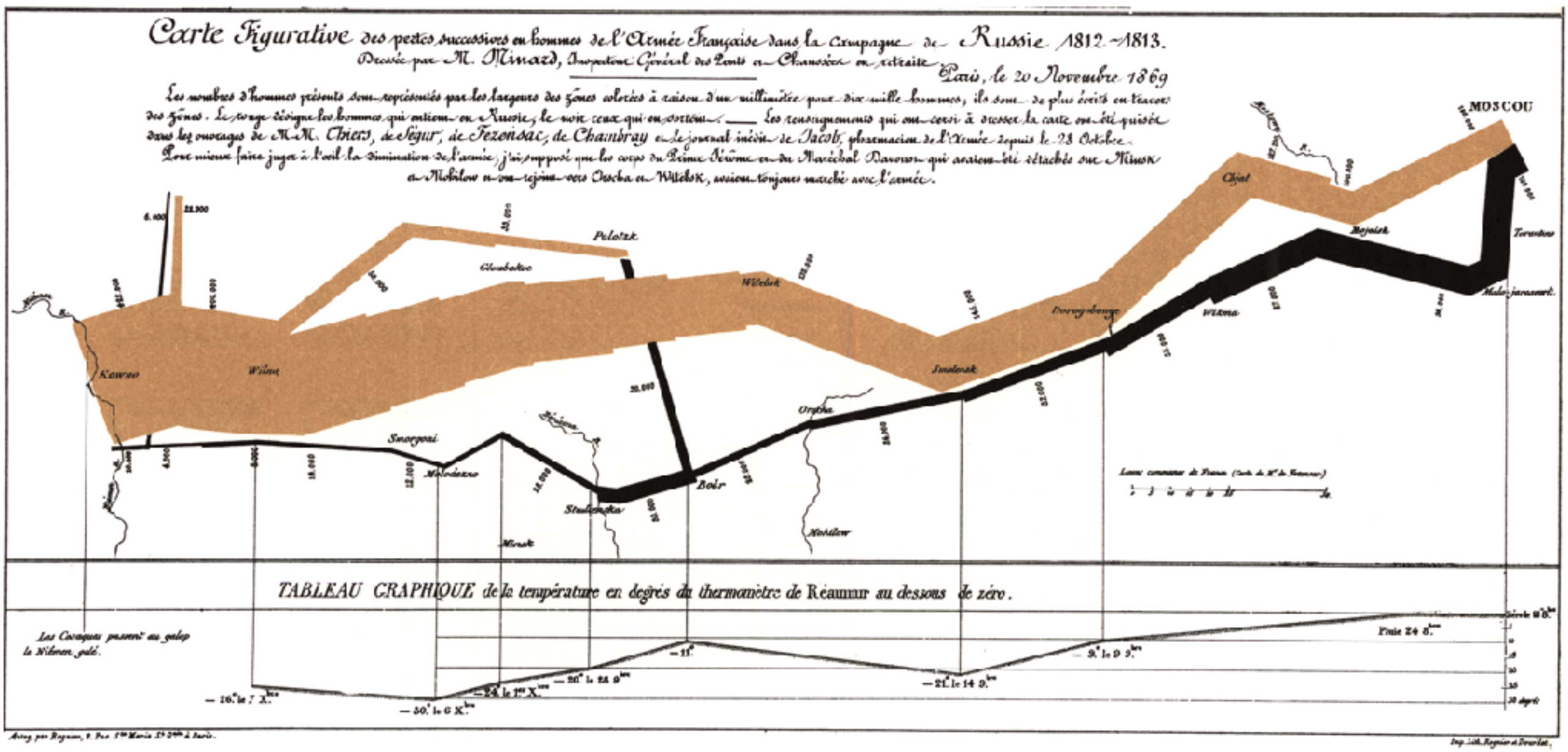
The fact that some undergrads payed $20 cannot repeat more than 7 random words does not mean you should limit your graphic to 7 elements:

to learn how to plot visualize at Nature and Science

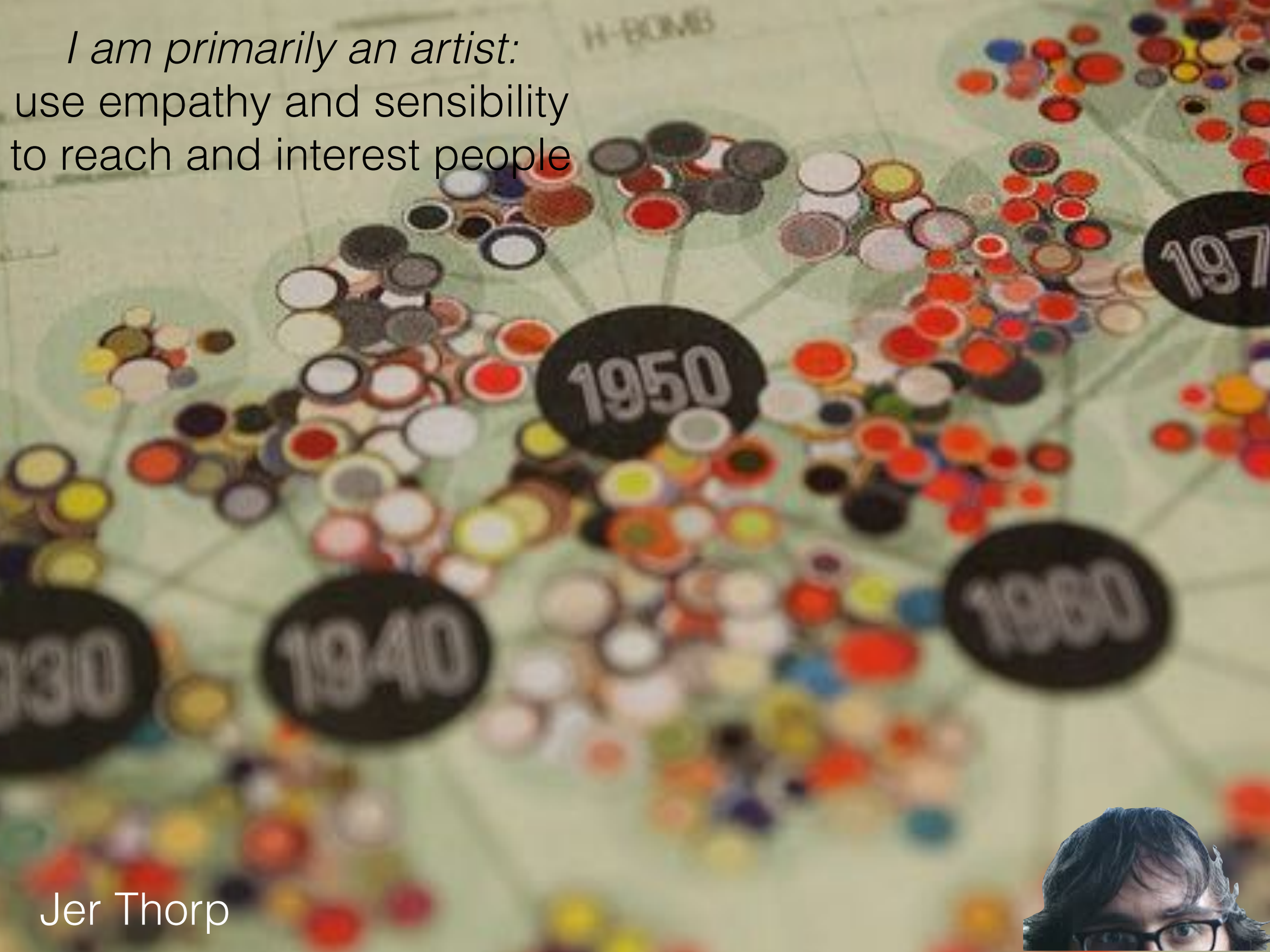
Edward Tufte

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869.
The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October.
In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jèrôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.



I am primarily an artist:
use empathy and sensibility
to reach and interest people



Jer Thorp

https://vimeo.com/4587178

Key Concepts:

Be thoughtful and make sure your visualizations are (in this order):

honest

clear

convincing

beautiful

Homework:

PHASE 1: Just one plot this week! Choose to plot anything you wish, from the data we used so far. Please make it a good plot.
Put though into the esthetic of the plot, how clearly the content is communicated, avoid clutter, avoid misleading elements, mind your choice of colors accordingly to what was discussed in class.

PHASE 2: The plot is due on next class. Turning in the plot will automatically award you the full 10 points. At that time you will receive 3 plots from your class mates, and you have to evaluate their strengths and weaknesses. Think about it as a pier review, which is something that is normally done on scientific papers in order to accept them or reject them for publication. The revision of the plots is due 4 days later and will be evaluated on a scale (10 points).

Upload your plots to github (before next class) in a new folder HW7.
Upload the plots revisions as a README file in the same directory, including the plots as images in the README (4 days later by the end of the day).

Homework:

READING: this reading will help you review your peer's plots

7 Great Visualizations from History
http://data-informed.com/7-great-visualizations-history/

Six Lessons from the Bauhaus: Masters of the Persuasive Graphic
http://blog.visual.ly/six-lessons-from-the-bauhaus-masters-of-the-persuasive-graphic/

Resources:

Tamara Munzner
Visualization Analysis & Design, 2014
Chap 6 (photocopies will be in the library)
http://www.cs.ubc.ca/~tmm/talks/minicourse14/vad15london.pdf

7 classical vis papers
http://fellinlovewithdata.com/guides/7-classic-foundational-vis-papers