Lecture 3: Principles of Data Visualization

Instructor: Saravanan Thirumuruganathan

Outline

1. Bertin's Visual Attributes
2. Tufte's Principles
3. Effective Visualization
4. Intro to PsychoPhysics

In-Class Quizzes

- **URL:** `http://m.socrative.com/`
- **Room Name: 4f2bb99e**

Announcements

- One-time attendance recording
- Form teams soon!
- Programming Assignment 1 will be released this weekend (due in 3 weeks)

Bertin's Visual Attributes

Jacques Bertin

- French cartographer [1918-2010]
- Semiology of Graphics [1967]
- Theoretical principles for visual encodings

Bertin's Visual Attributes

Marks Points Lines Areas

Channels

Position

Size

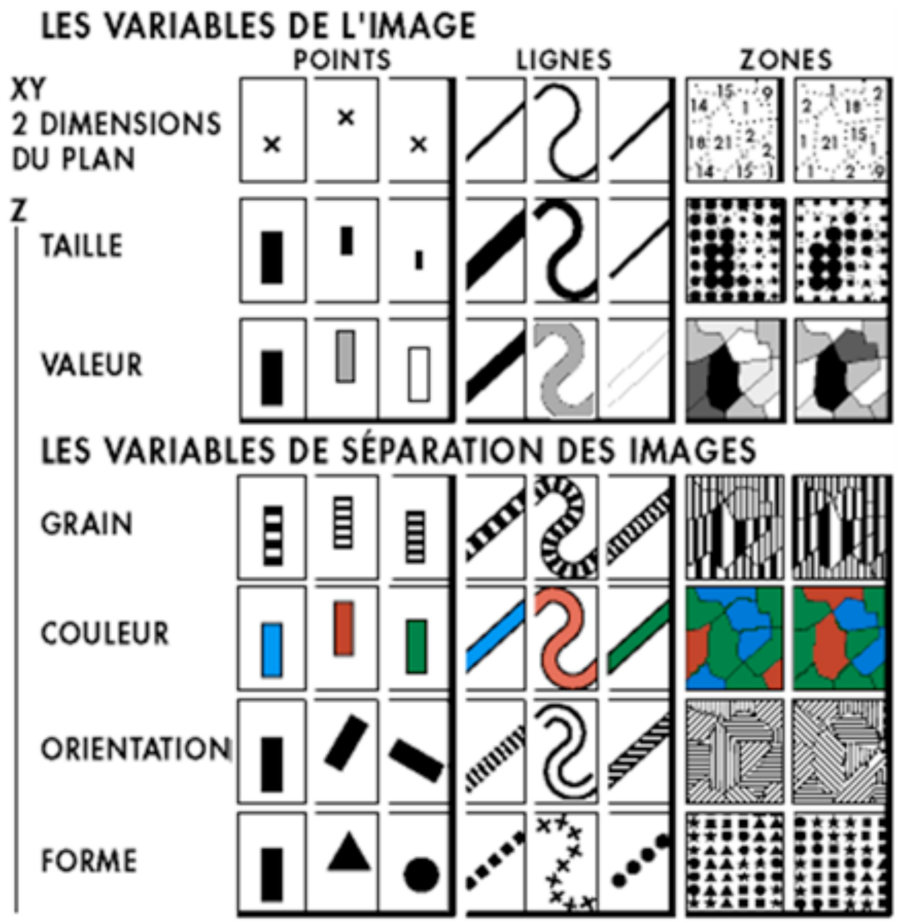
(Grey)Value

Texture

Color

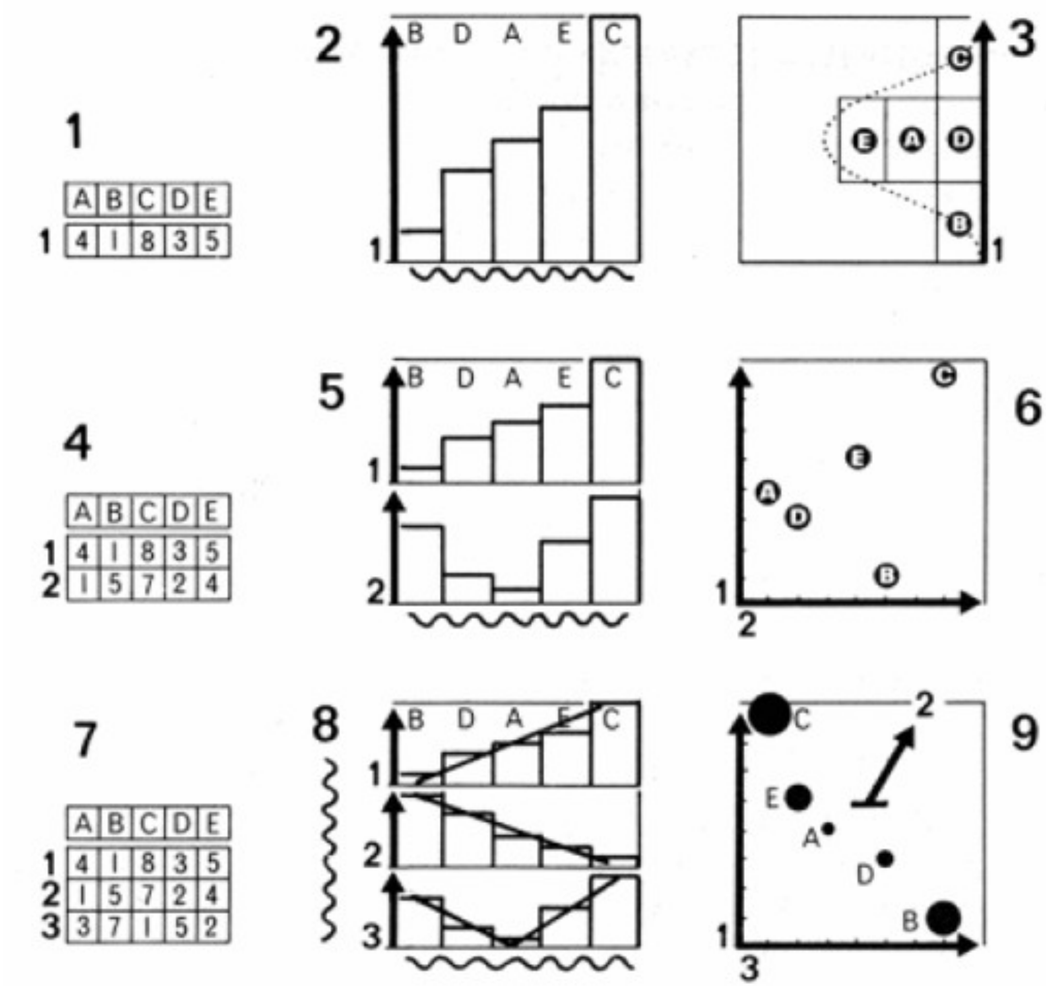
Orientation

Shape



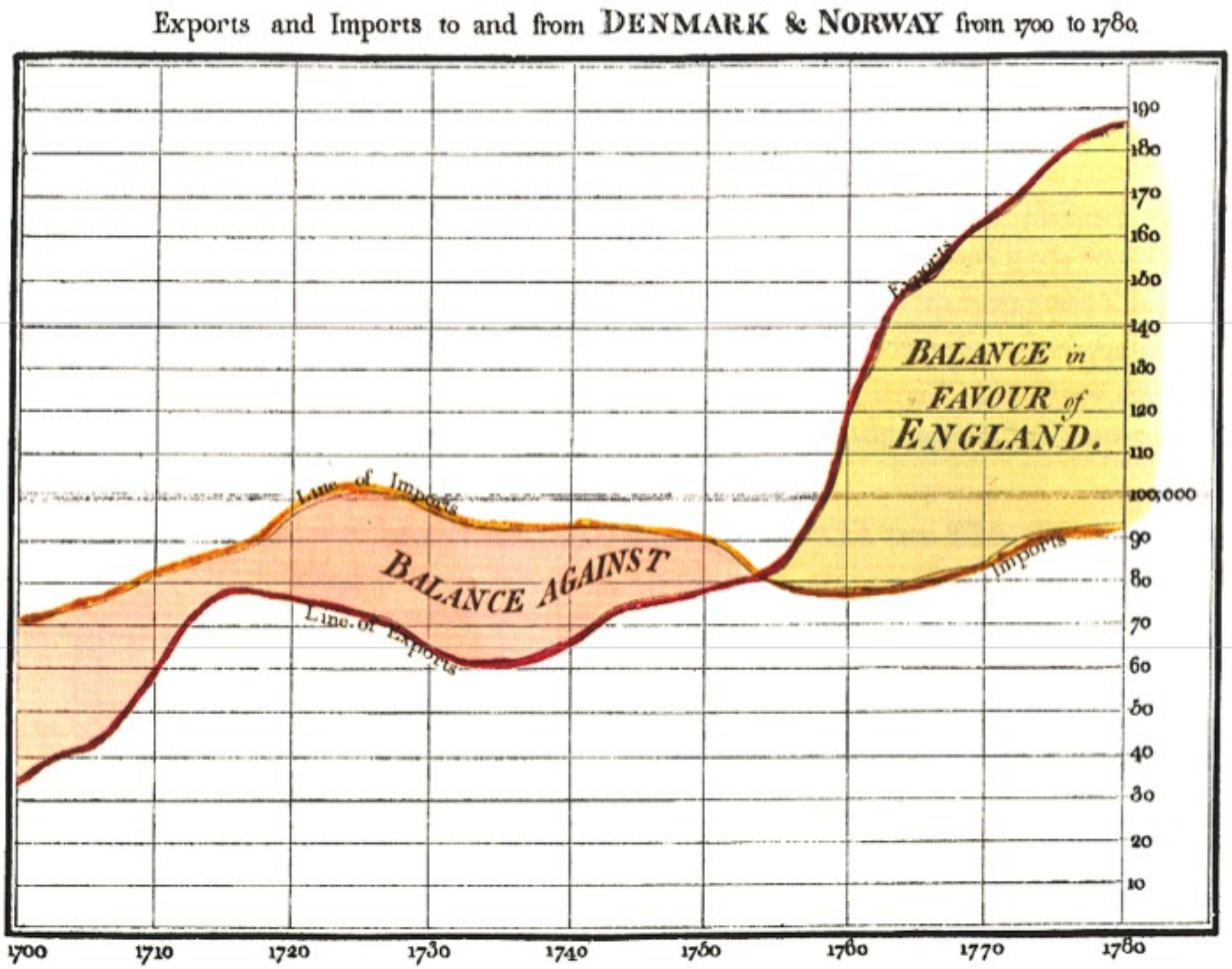
Bertin, Semiology of Graphics, 1967

Large Design Space (Visual Metaphors)



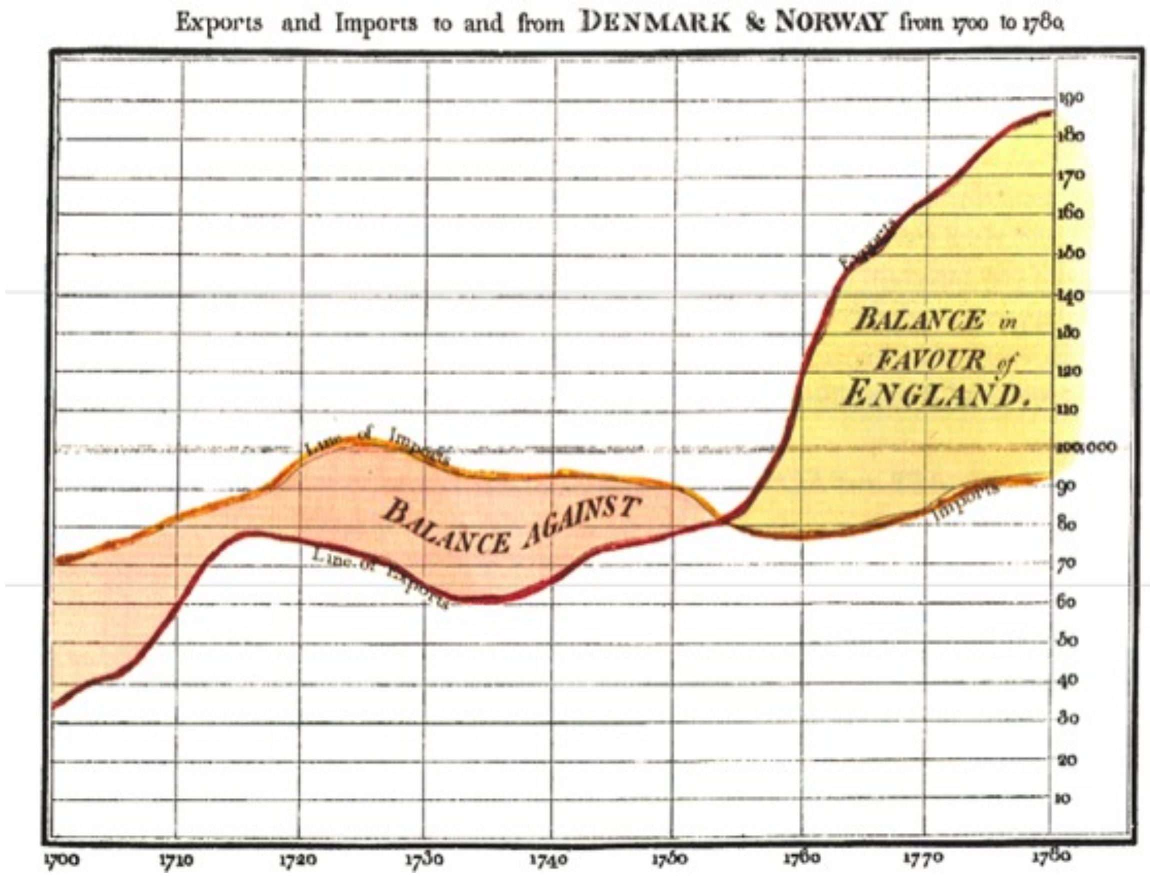
J. Bertin, "Graphics and Graphic Information Processing", 1981

Example



W. Playfair, 1786

Example



x-axis: Year (Q)
y-axis: Currency (Q)
Color: Imports / Exports (N, O)

W. Playfair, 1786

Visual Attributes per Data Type

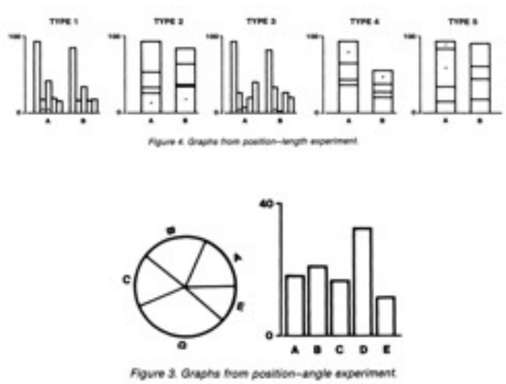
Bertin, 1967

	Categories	Ordinal	Quantitative
Position	✔	✔	✔
Length	✔	✔	✔
Brightness	✔	✔	~
Texture	✔	~	✖
Color	✔	~	✖
Angle	✔	✖	✖
Shape	✔	✖	✖

✔ = Good
~ = OK
✖ = Bad

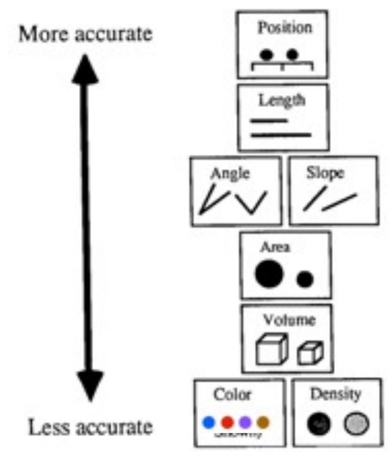
Bertin, Semiology of Graphics, 1967

Cleveland / McGill, 1984

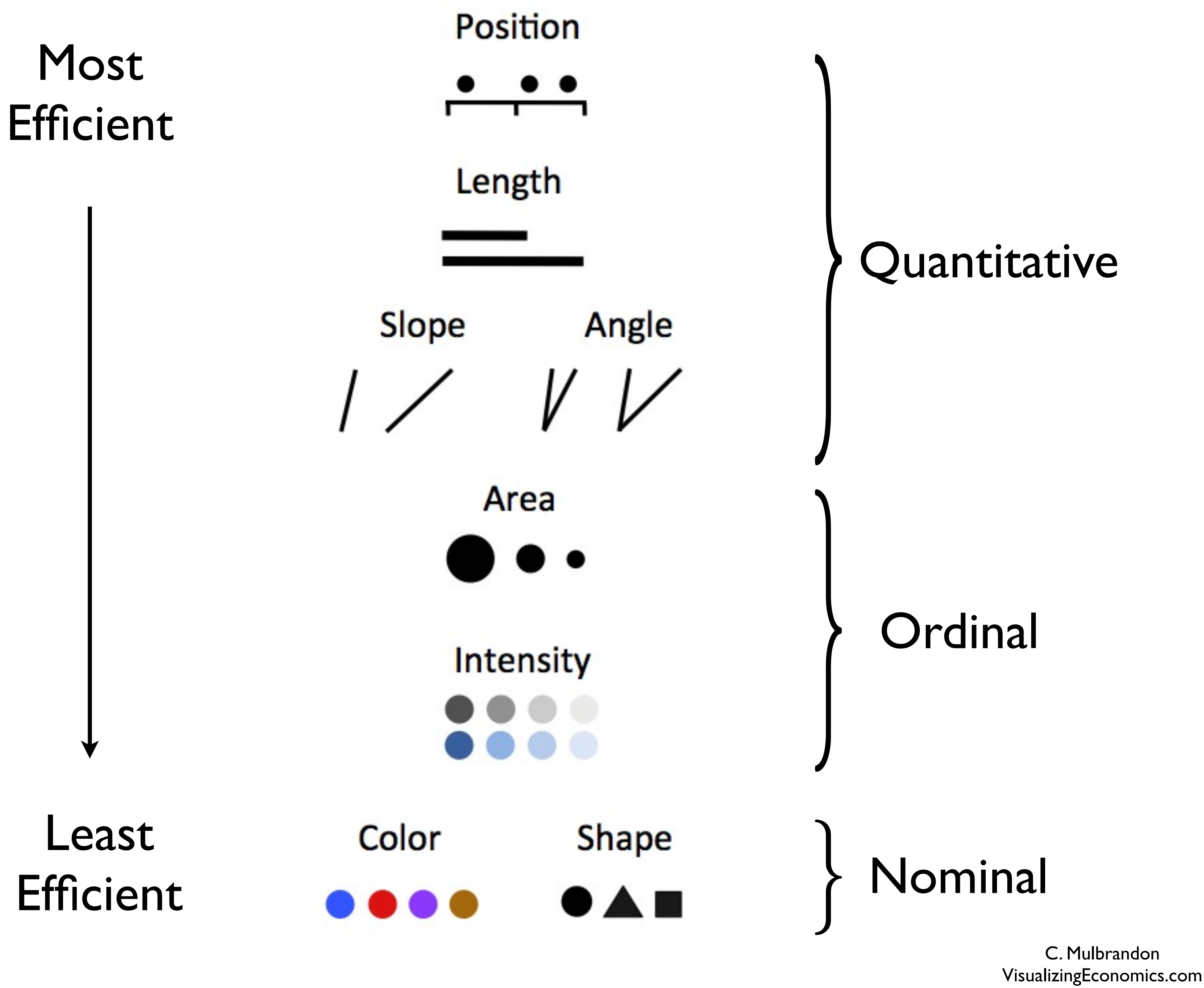


William S. Cleveland; Robert McGill , "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." 1984

Mackinlay, 1986



Jock Mackinlay "Automating The Design of Graphical Presentations." 1986


Most
Efficient
Least
Efficient
Position
Length
Slope
Angle
Area
Intensity
Color
Shape
Quantitative
Ordinal
Nominal
C. Mulbrandon
VisualizingEconomics.com

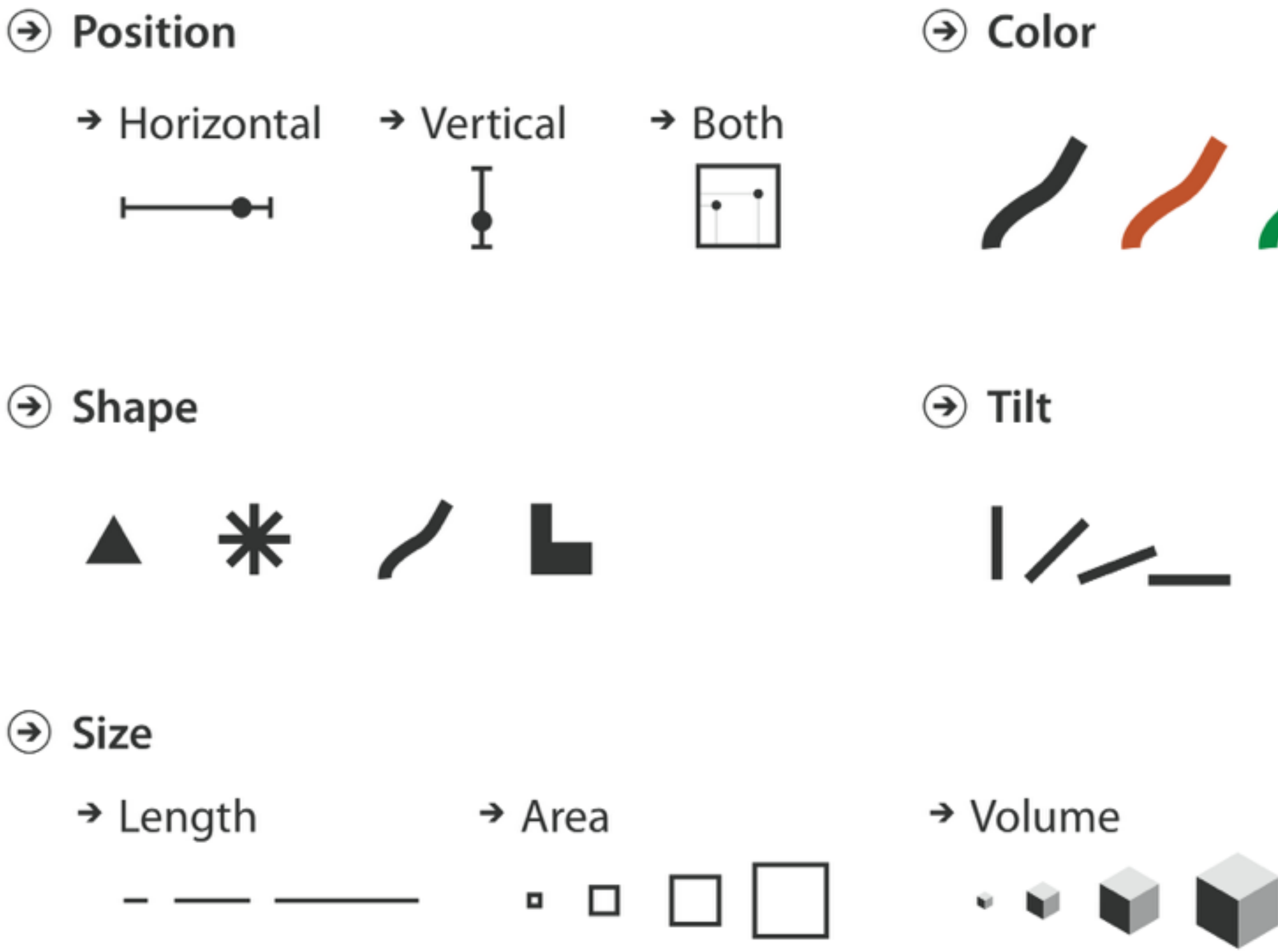
Visual Marks

Basic geometric elements

0D 1D 2D

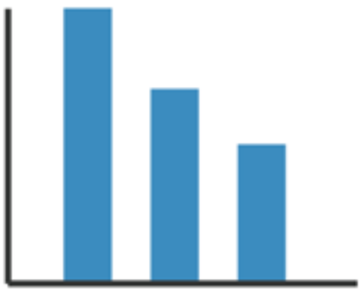
T. Munzner, Visualization Analysis and Design, 2014

Visual Variables (aka Channels)



T. Munzner, Visualization Analysis and Design, 2014

Using Marks and Attributes



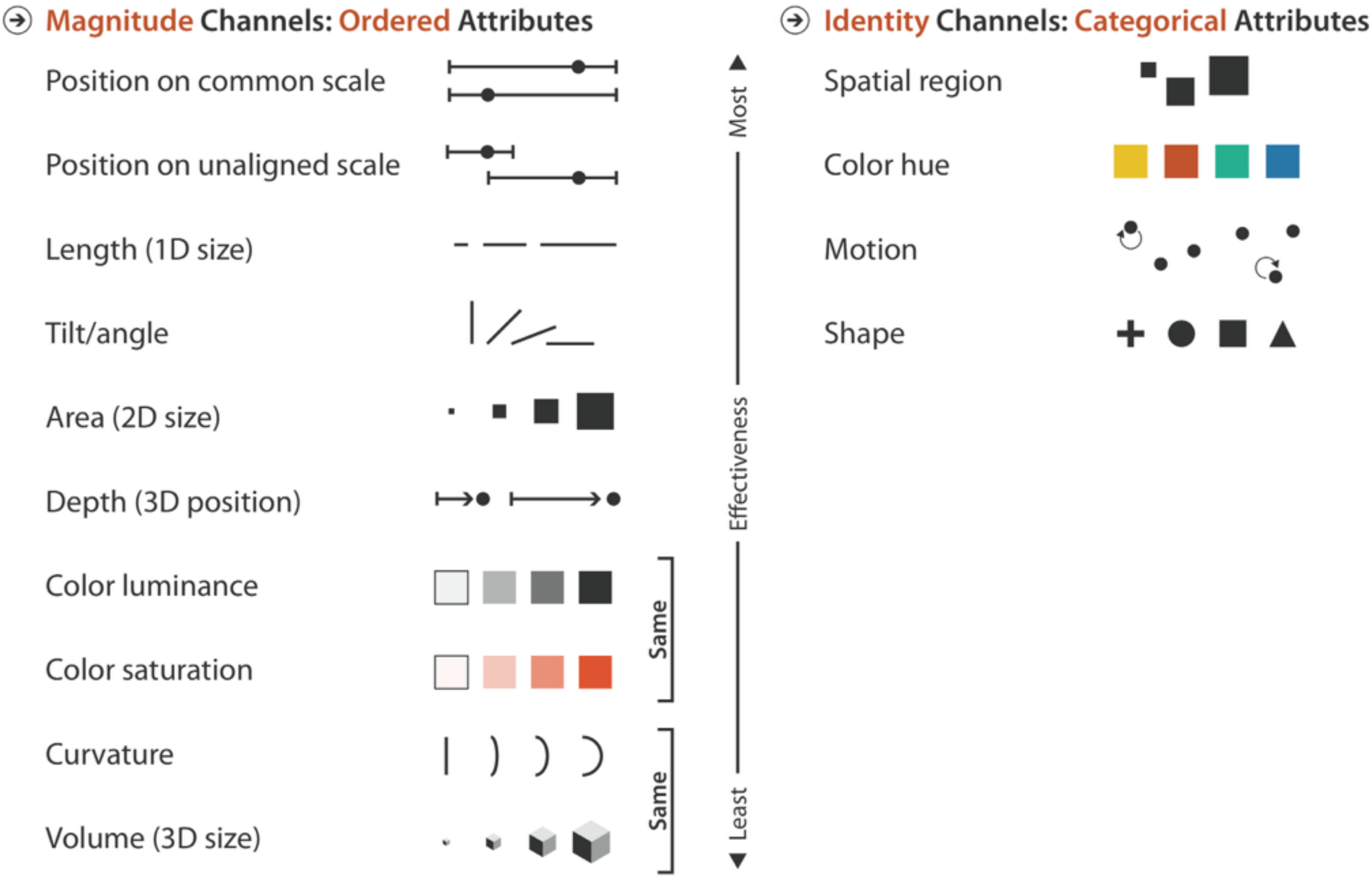
Length

Position

Color

Size

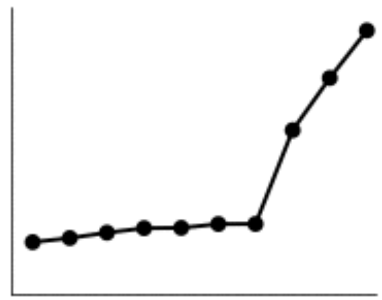
T. Munzner, Visualization Analysis and Design, 2014

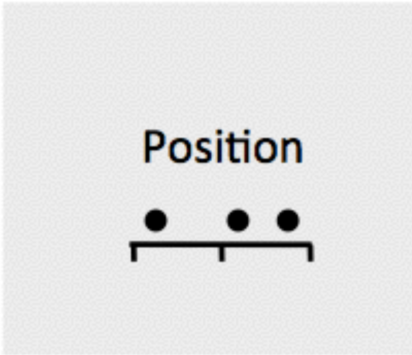


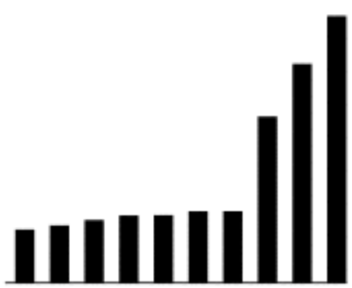
T. Munzner, Visualzation Analysis and Design, 2014

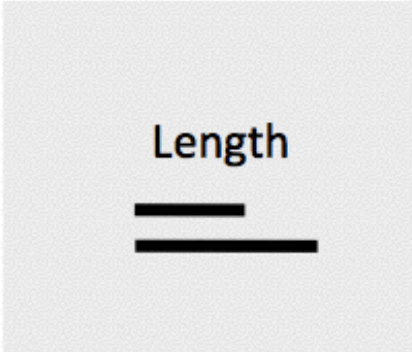
Most Effective

For Quantitative/Ordinal Data



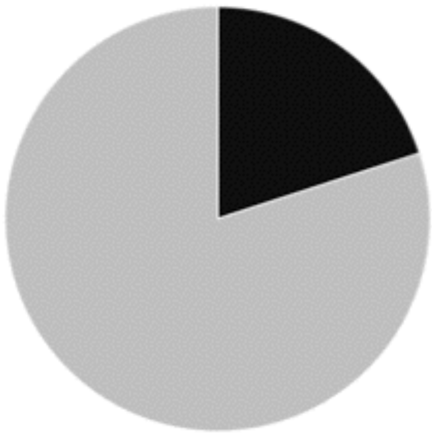


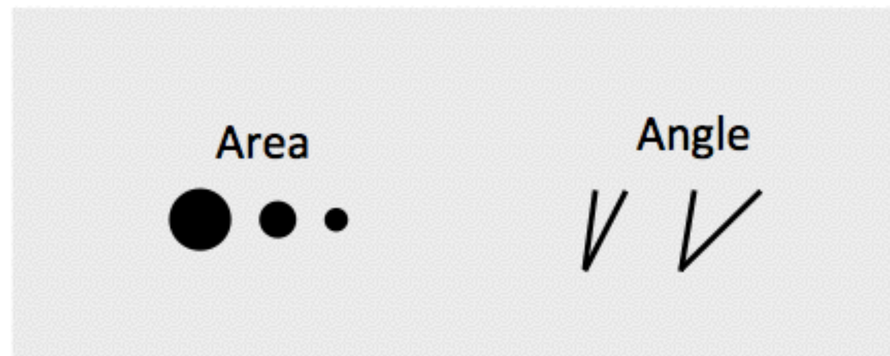


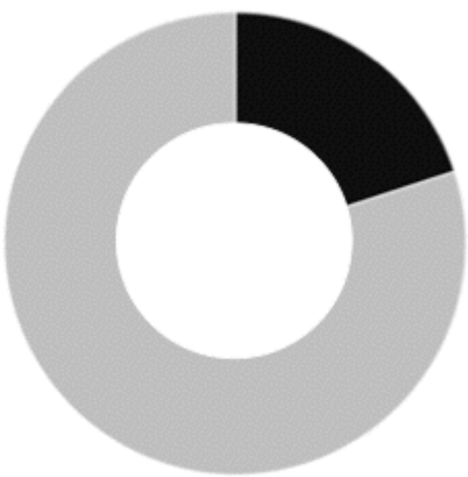


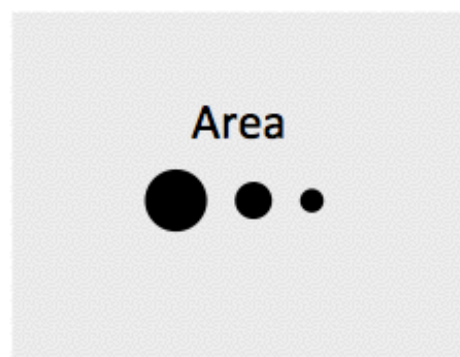
Less Effective

For Quantitative/Ordinal Data









Least Effective: Color

For Quantitative/Ordinal Data

SANFORD AND SELNICK

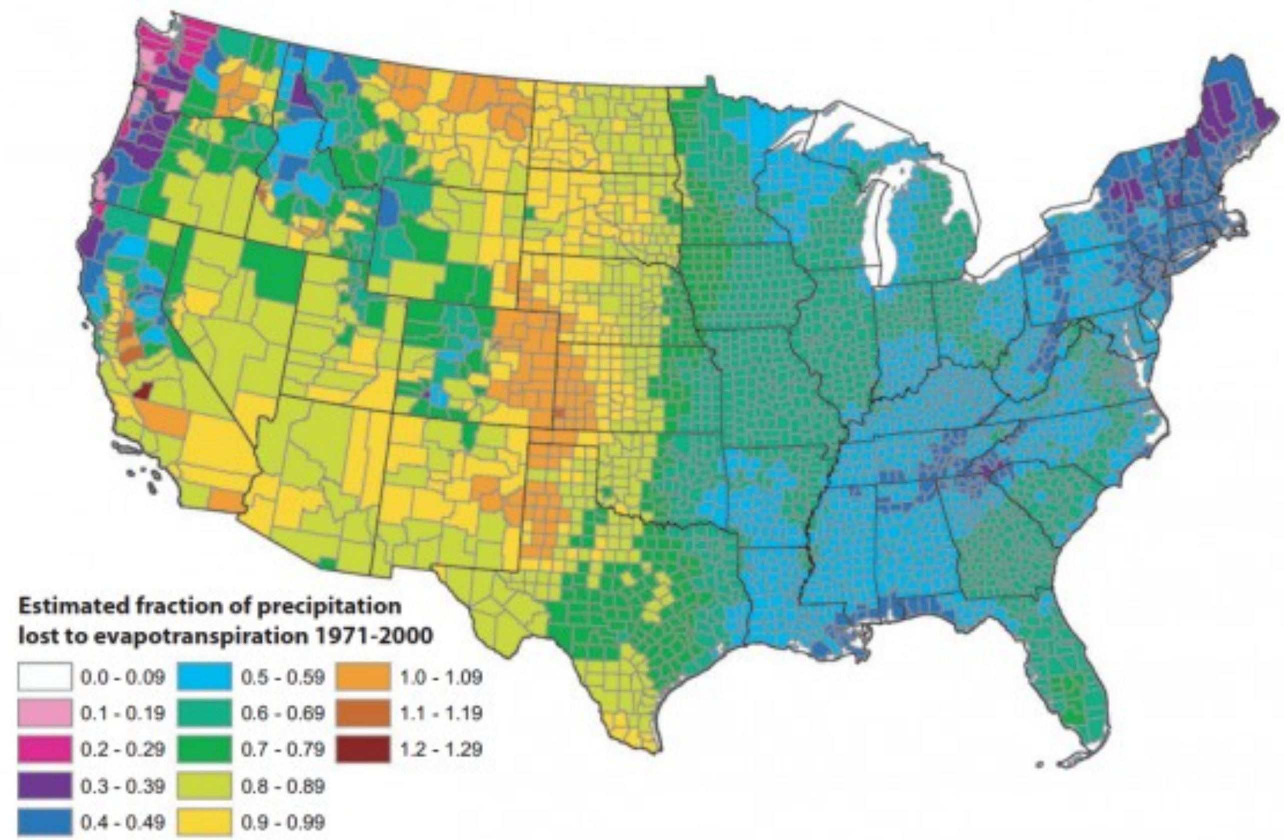
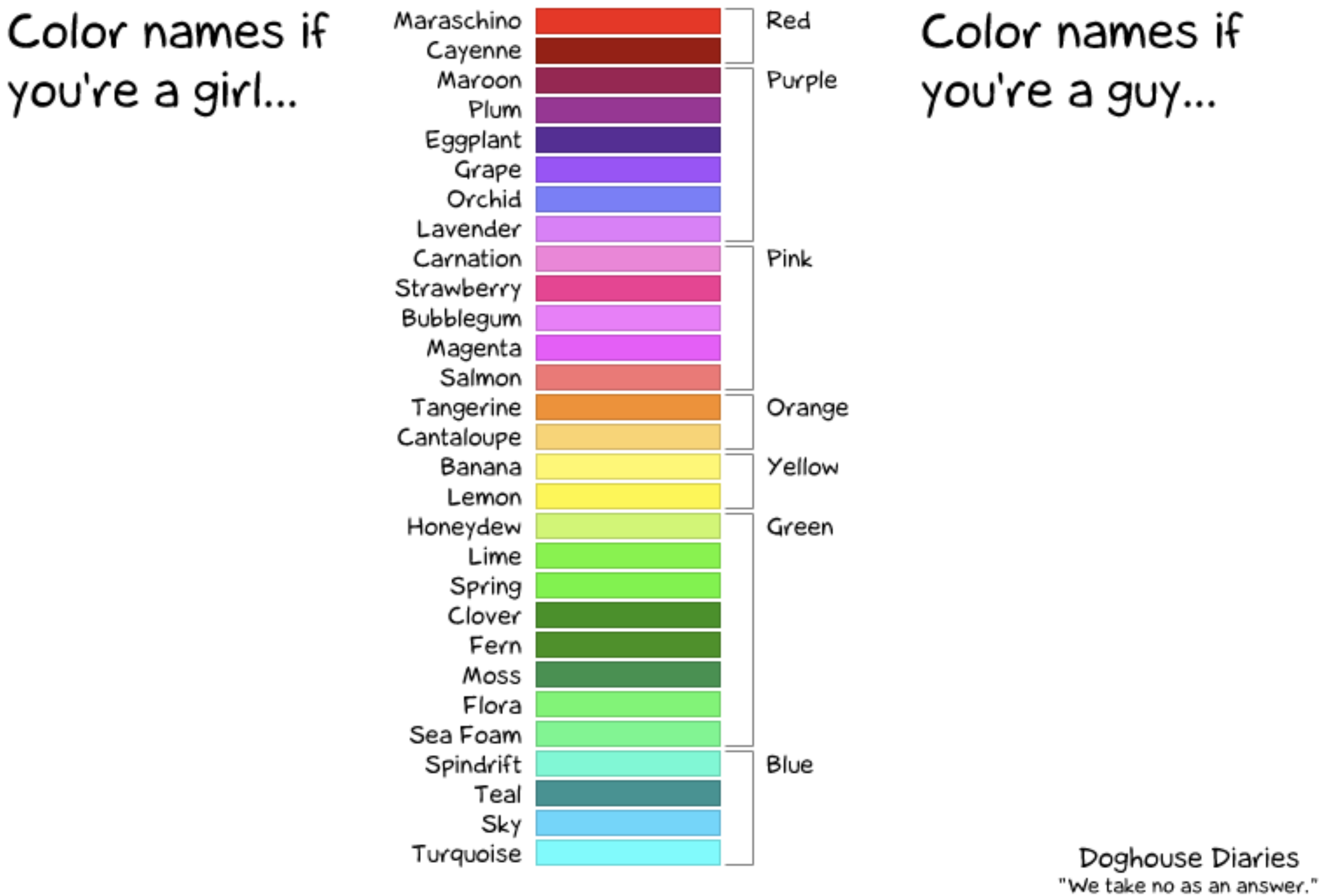


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (*P*) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/*P* were made first at the 800-m resolution of the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

Cliff Mass

Color

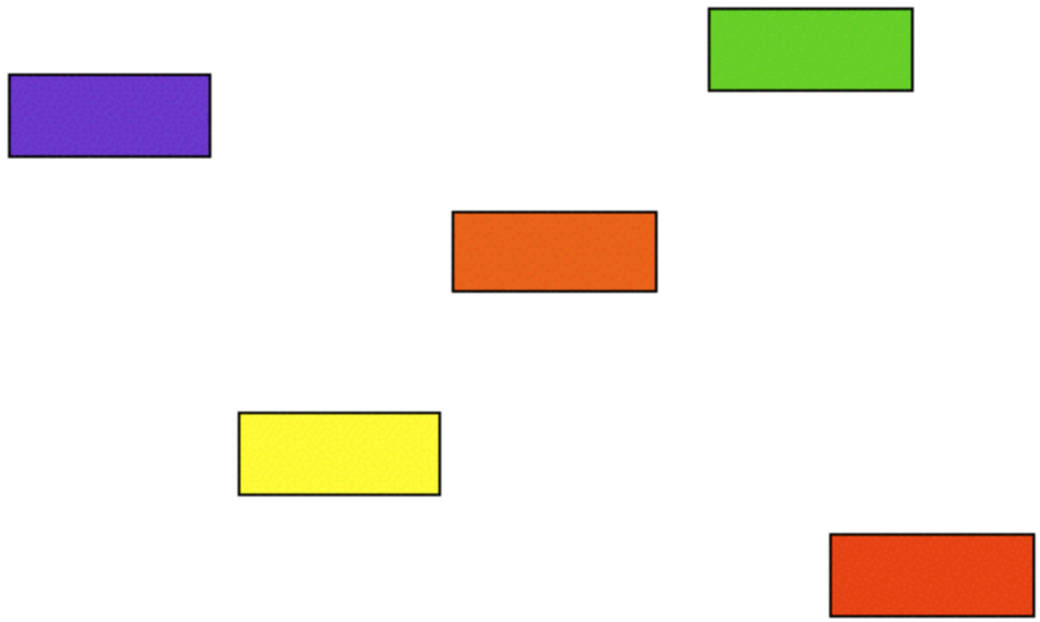


http://blog.xkcd.com/2010/05/03/color-survey-results/

Actual color names if you're a girl ... *Actual* color names if you're a guy ...

http://blog.xkcd.com/2010/05/03/color-survey-results/

Order These Colors

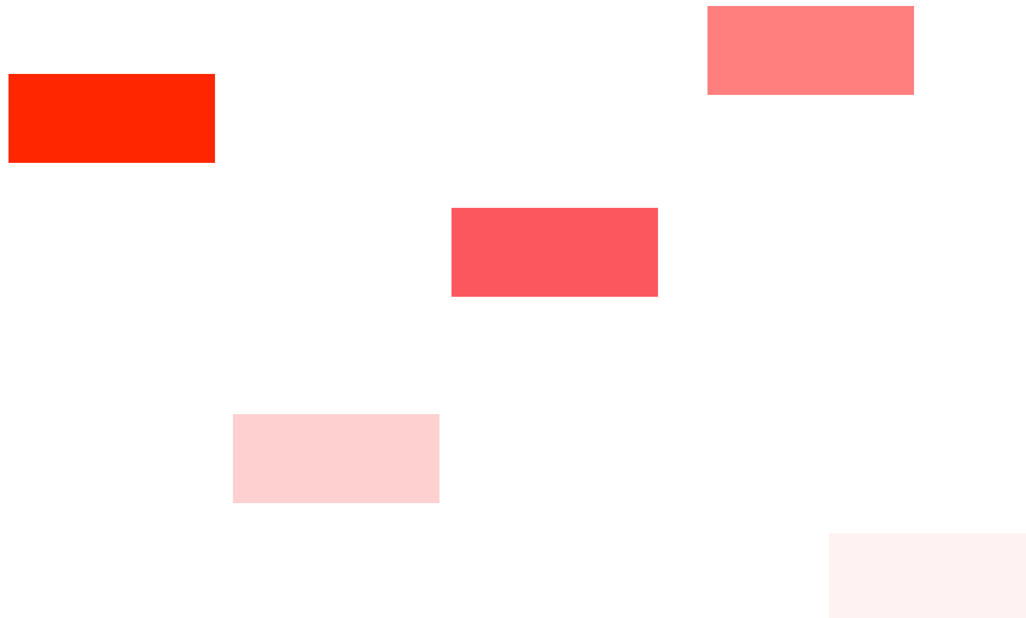


Based on slide from Stasko

Order These Colors

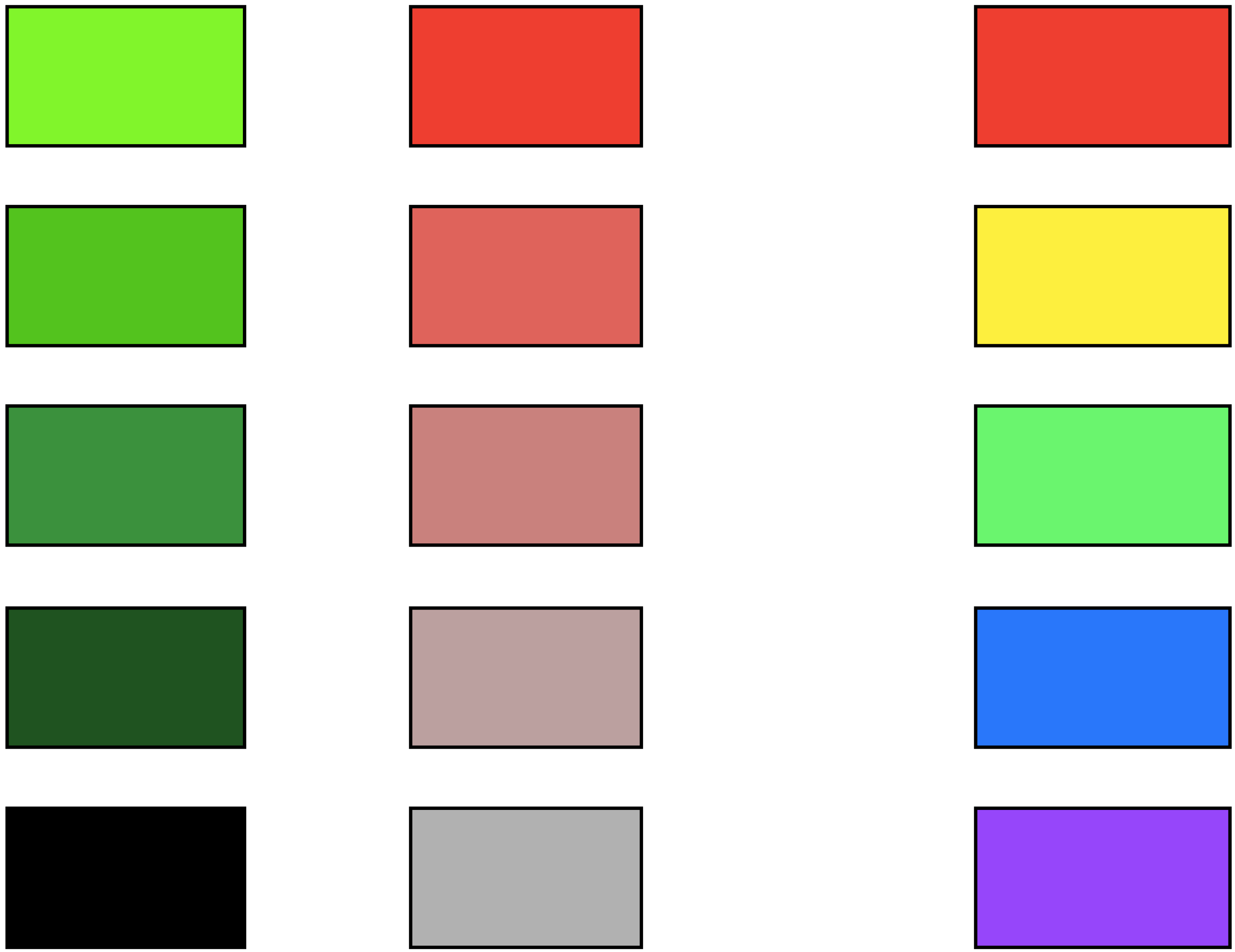
Based on slide from Stasko

Order These Colors



Based on slide from Stasko

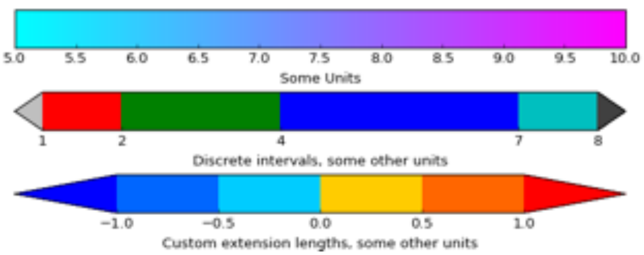
Brightness **Saturation** **Hue**

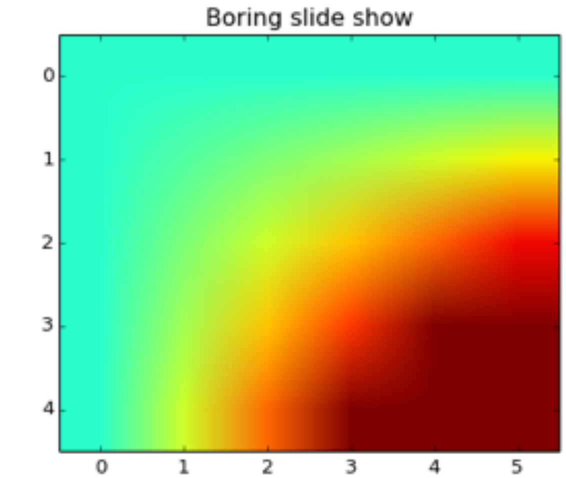


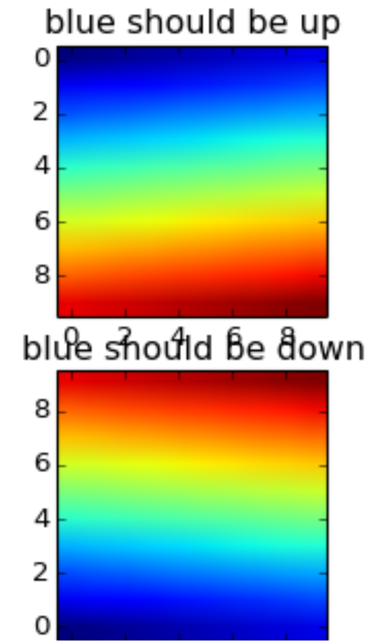
Perceived as Ordered

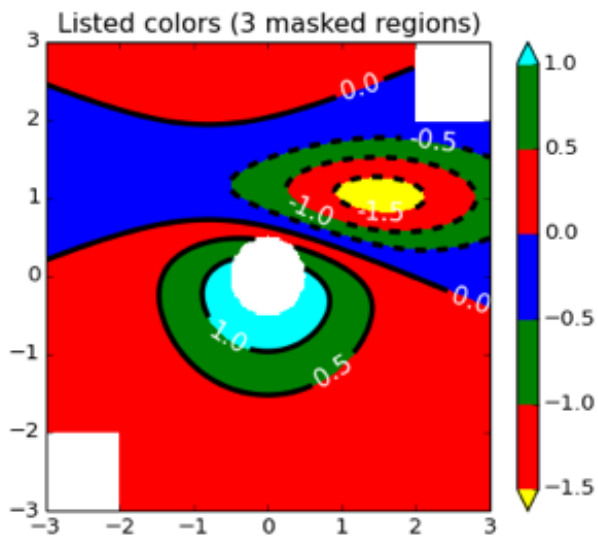
Not as much

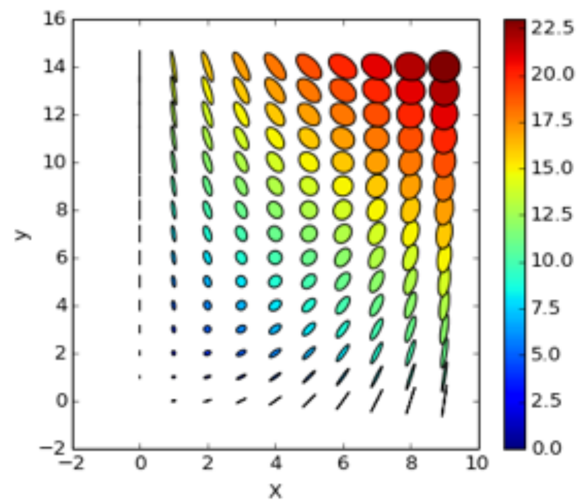
Rainbow Colors

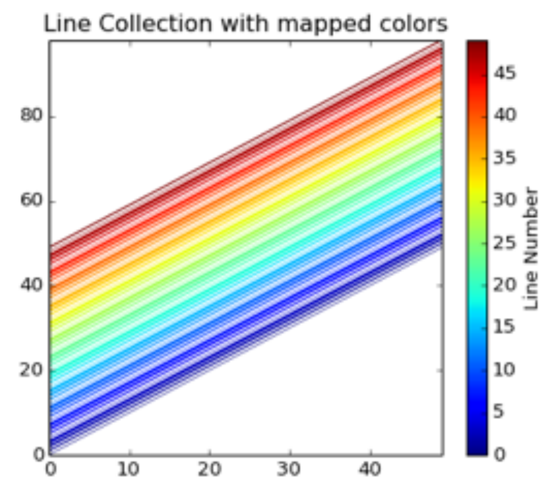






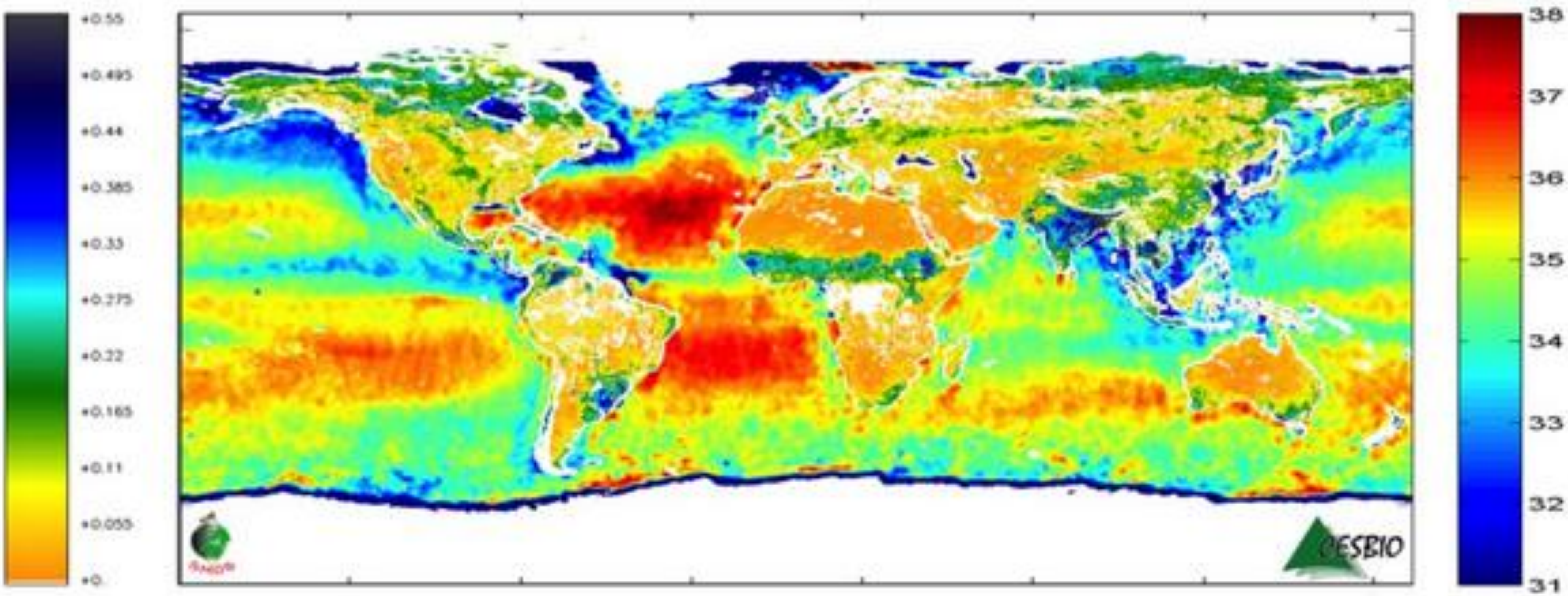






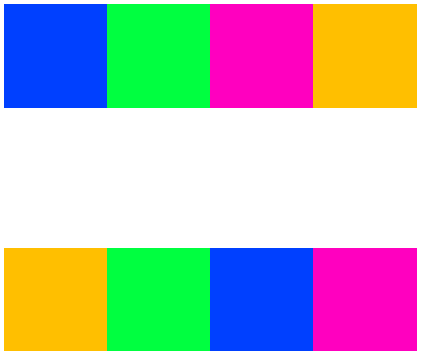
matplotlib gallery

Rainbow Colormap

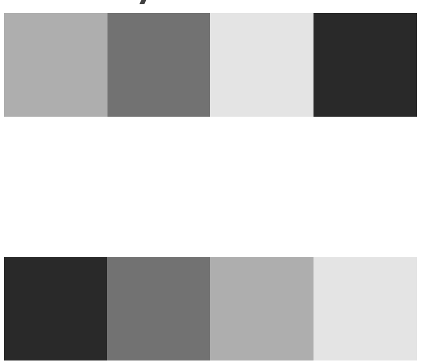


Rainbow Colormap

hard to order

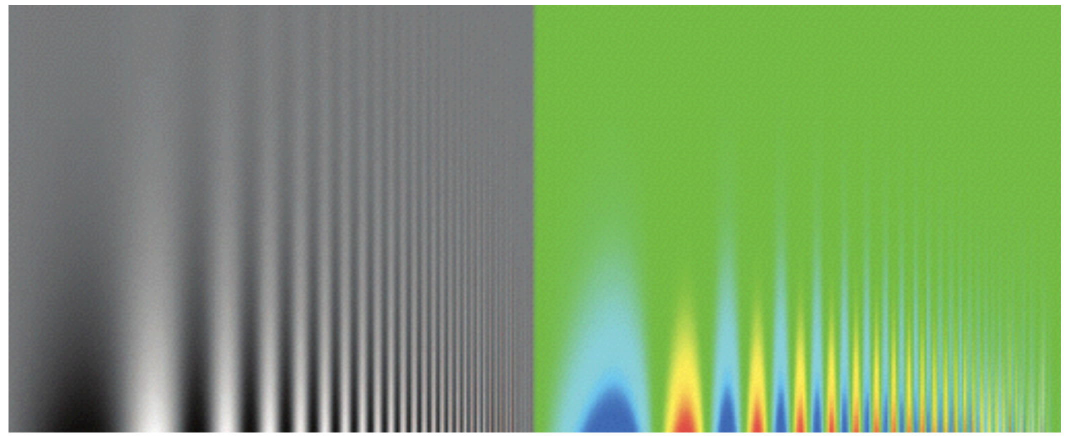


easy to order



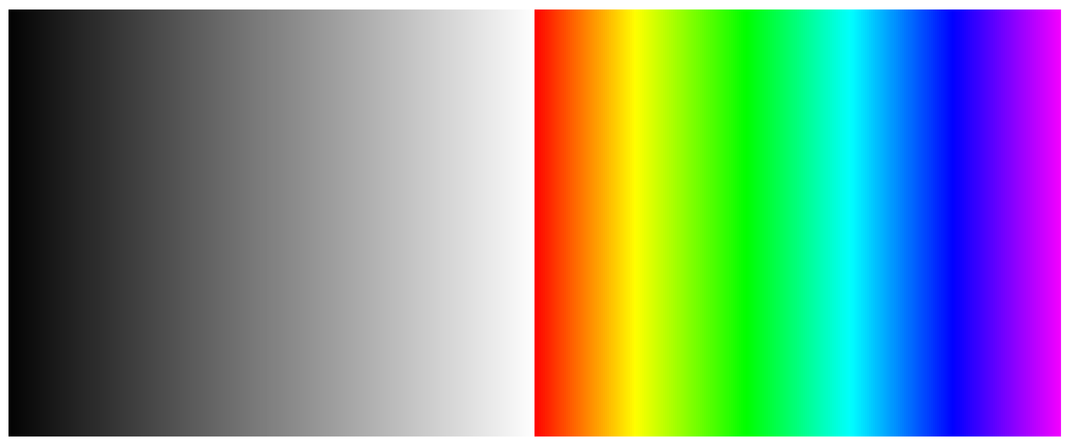
lower resolution

contrast



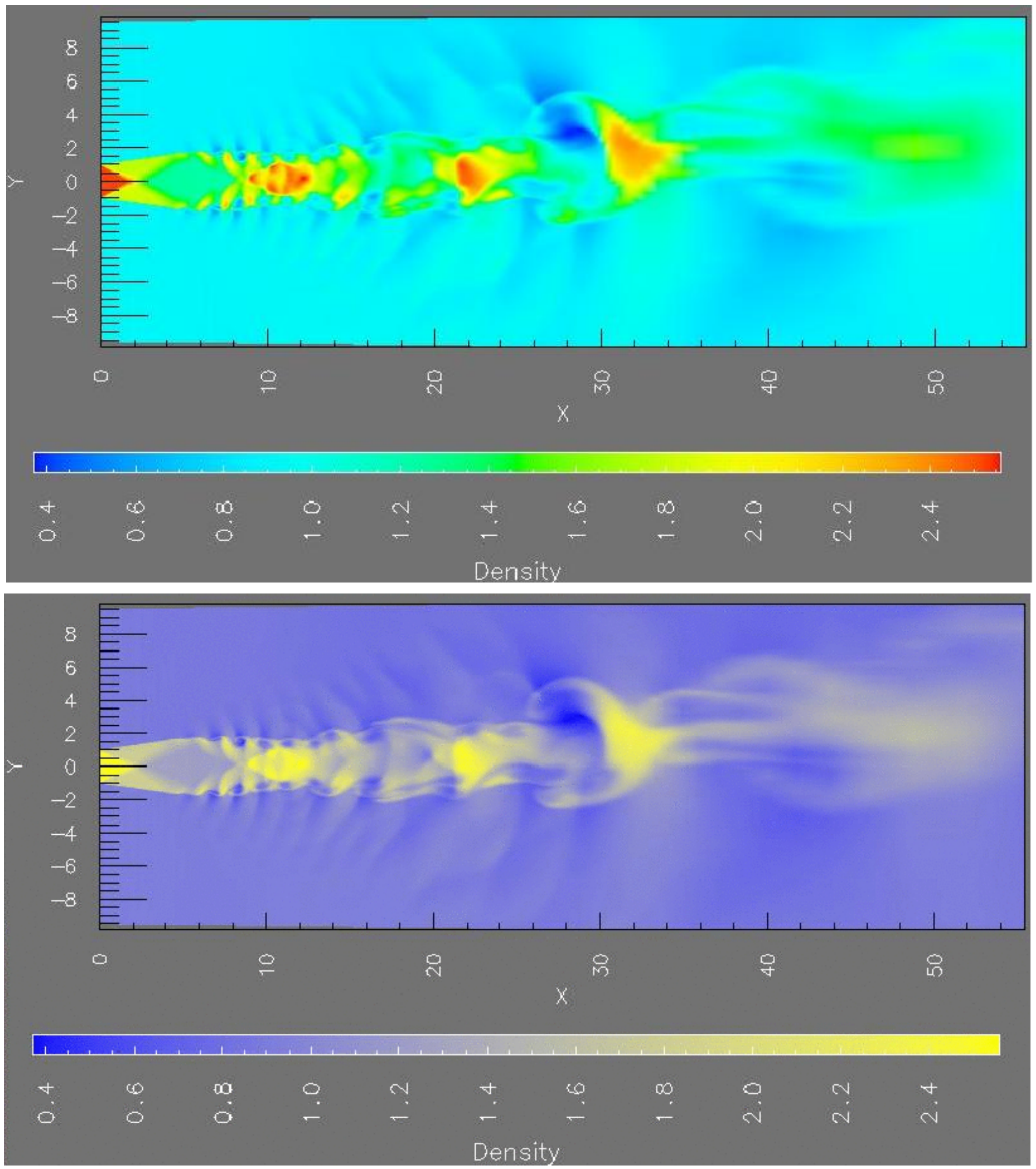
spatial frequency

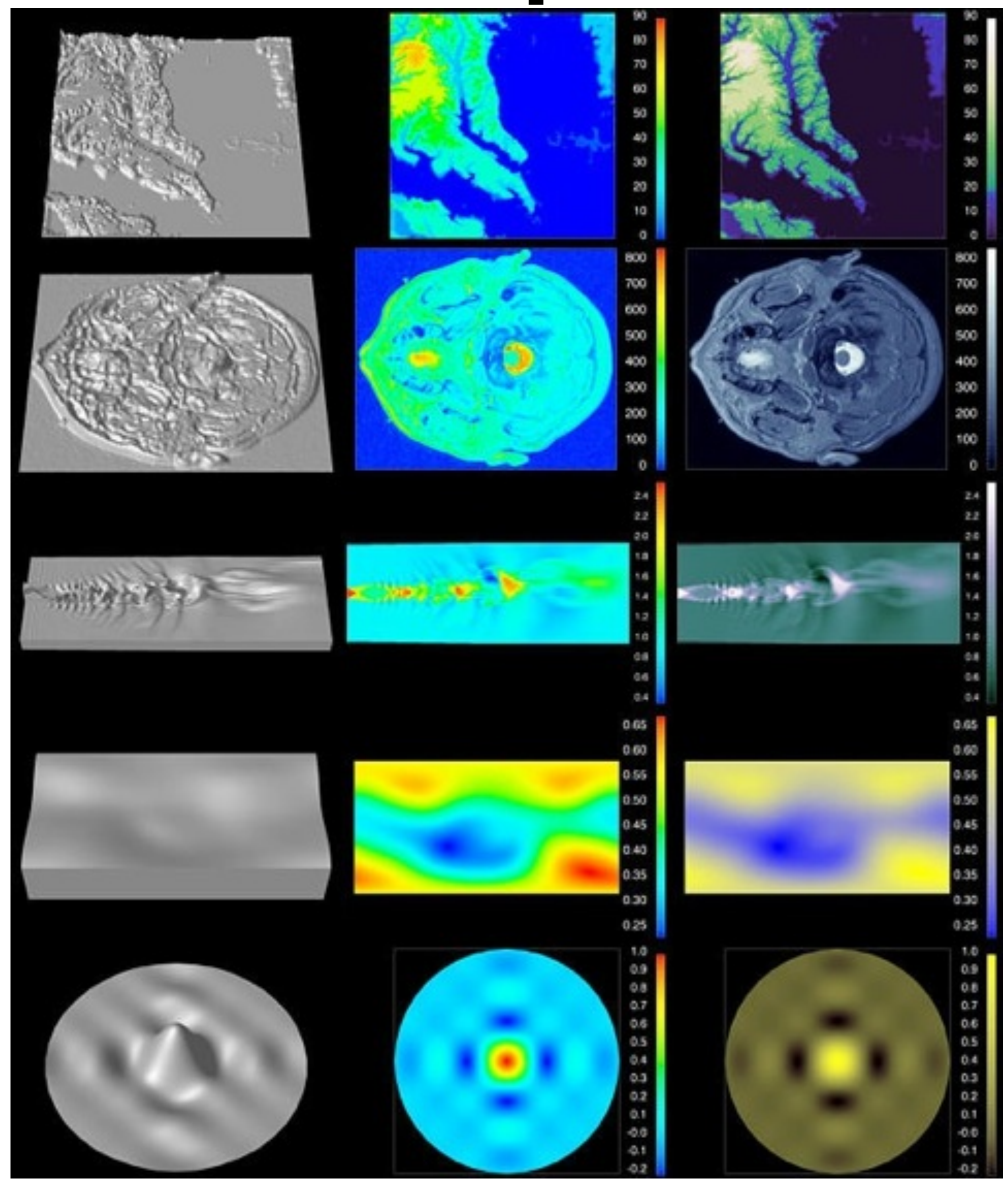
creates artifacts



Borland 2007

Rainbow Colormap



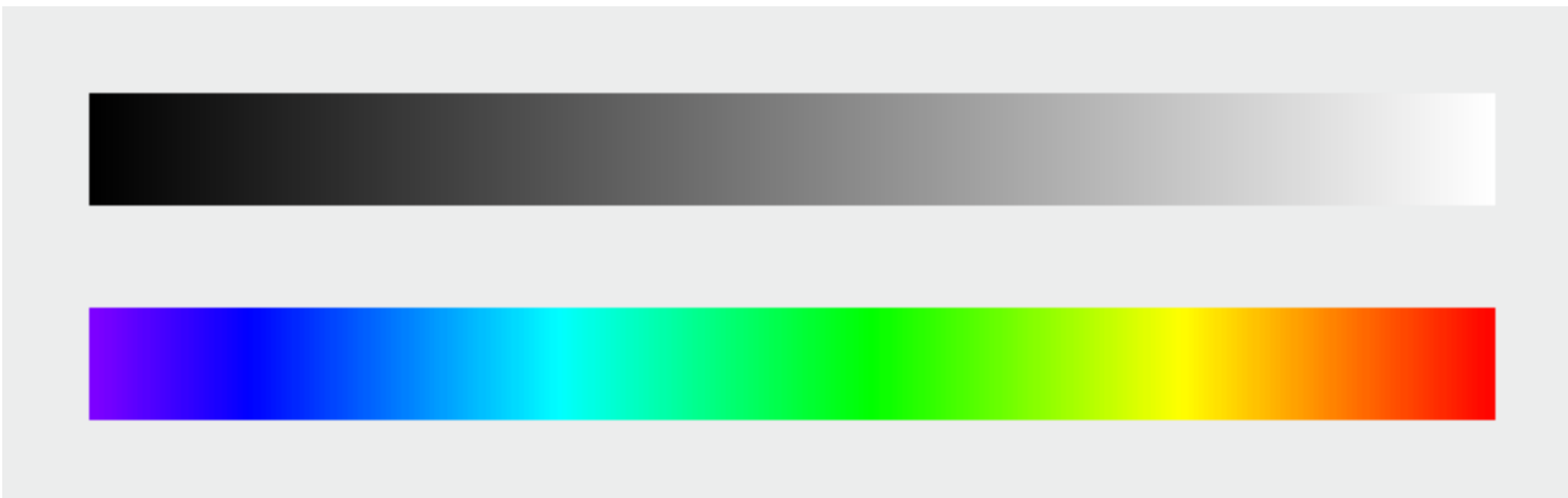


http://www.research.ibm.com/people/l/lloydt/color/color.HTM

B. Rogowitz & L. Treinish

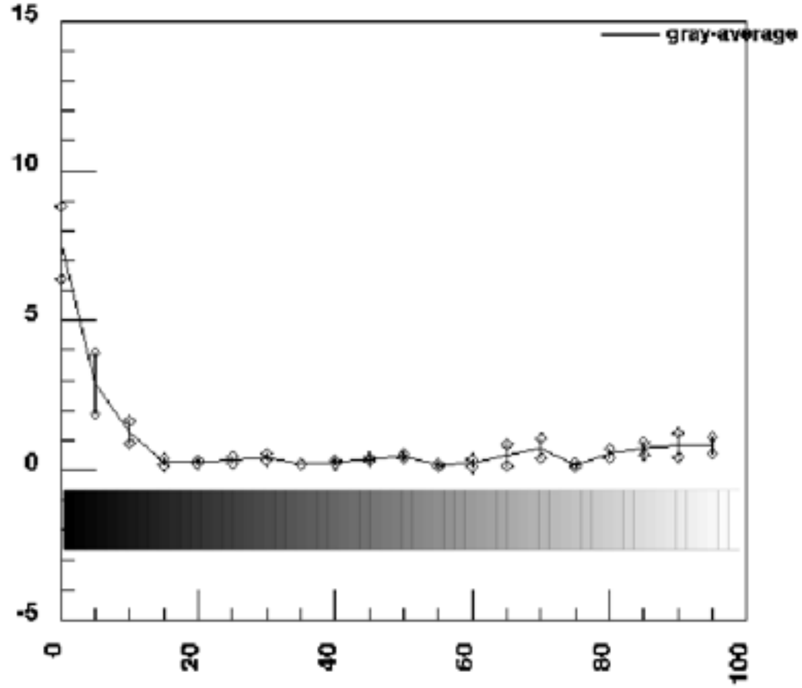
Rainbow Colormap

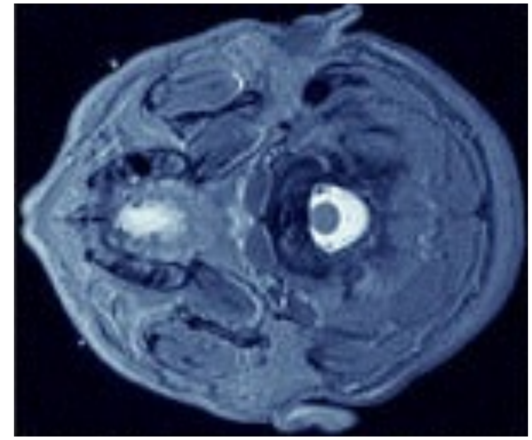
Rainbow colormap is perceptually nonlinear



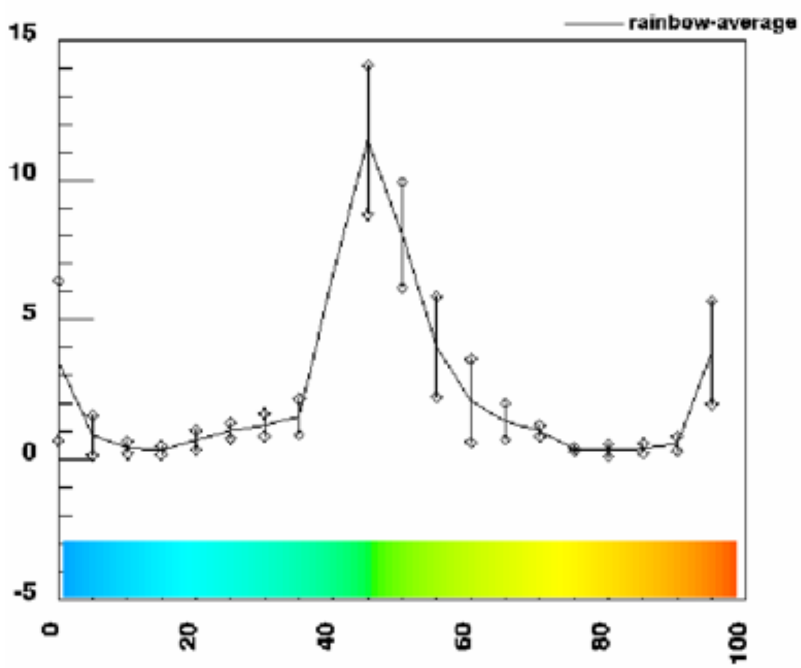
R. Simmon

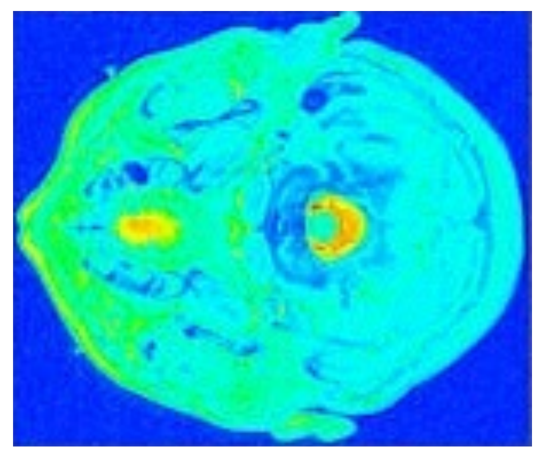
Luminance Increment



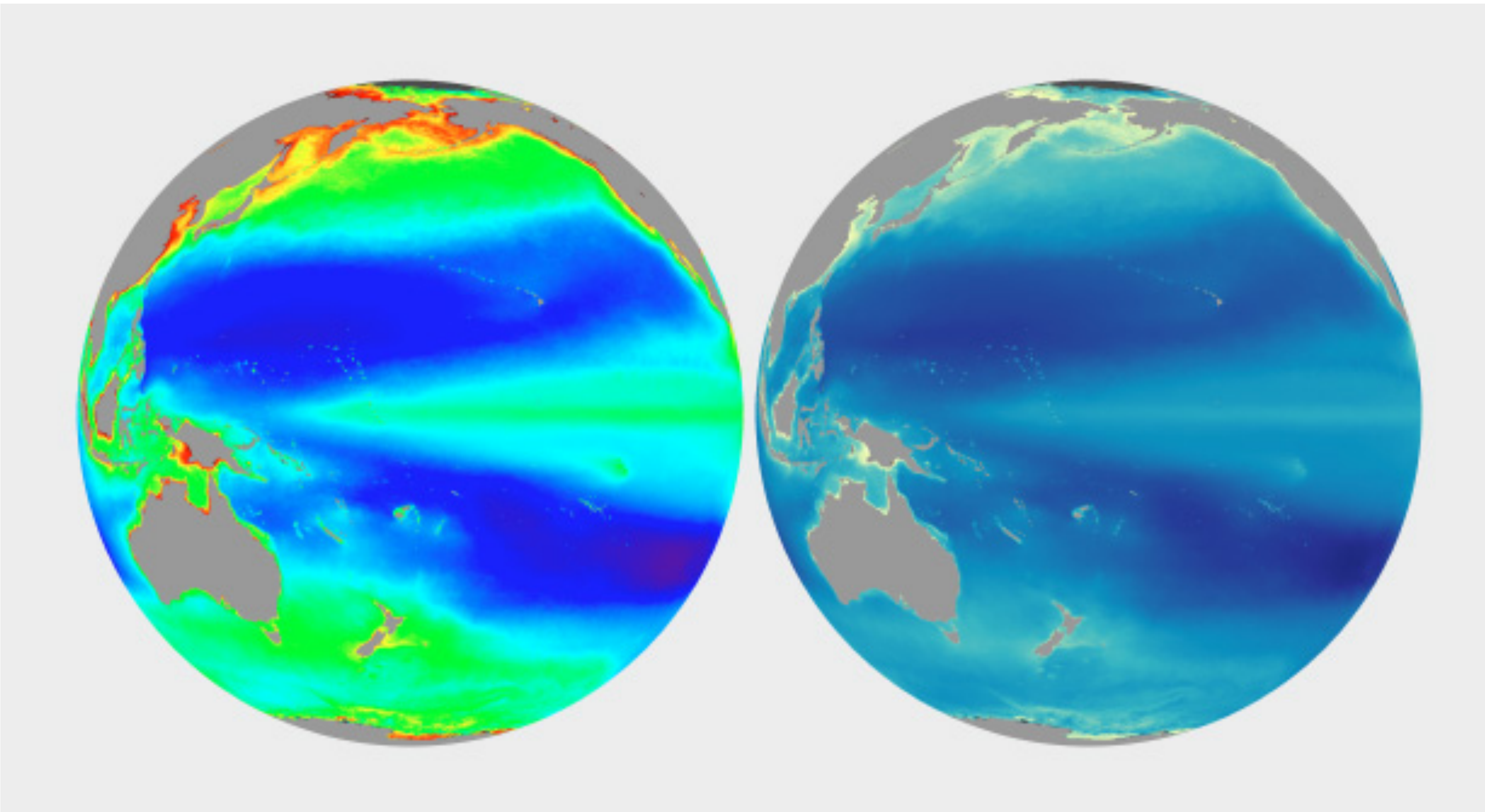


Luminance Increment





Rainbow Colormap



R. Simmon

Map Example Revisited

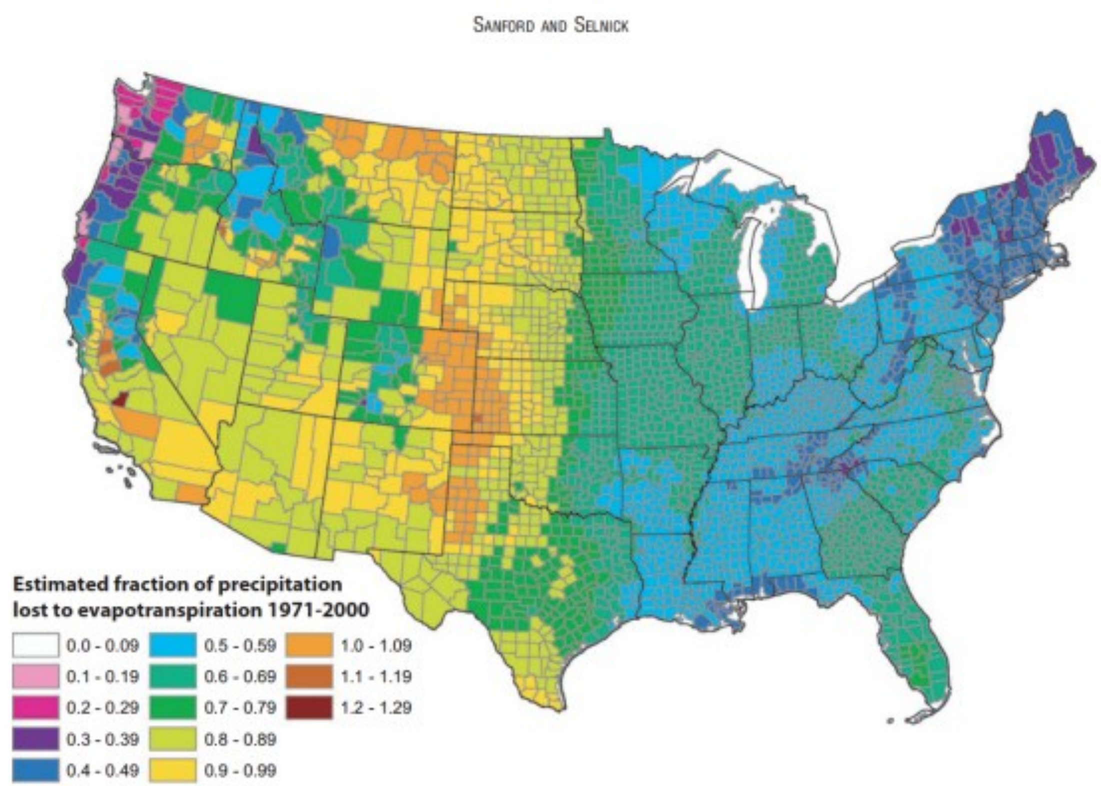
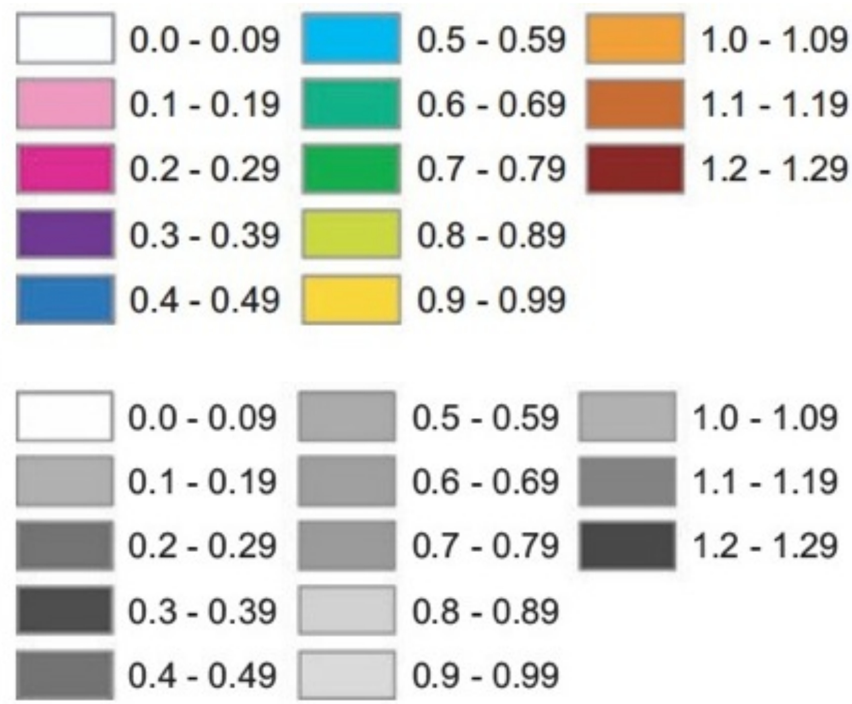


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (*P*) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/*P* were made first at the 800-m resolution of the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

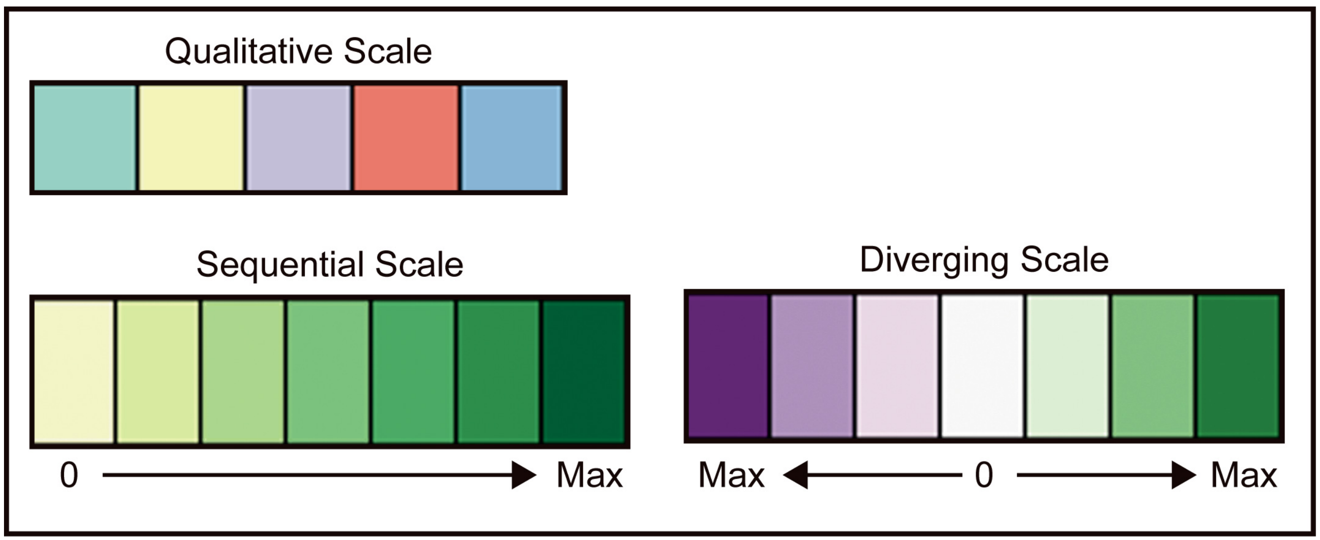


http://eagereyes.org/basics/rainbow-color-map

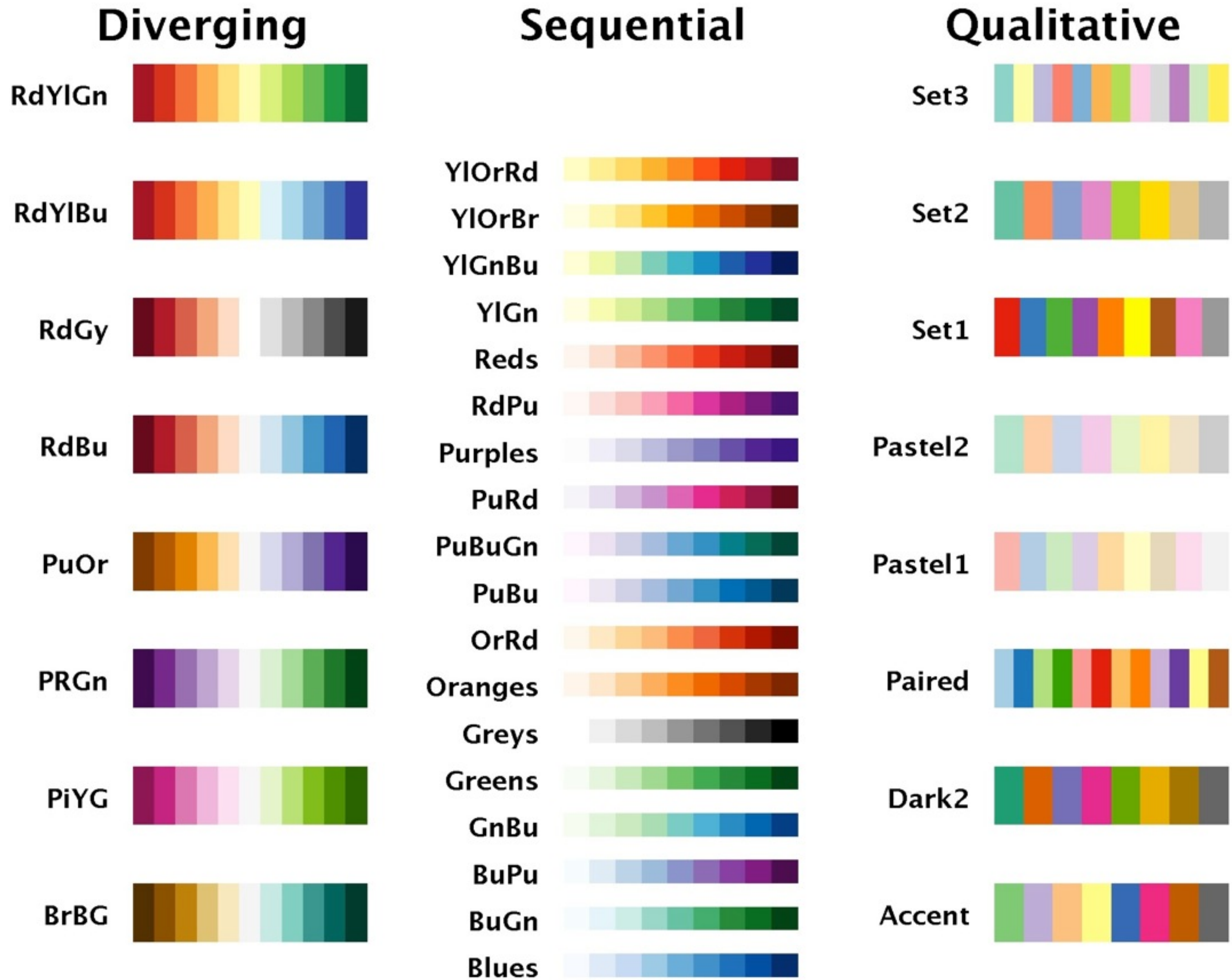
Brewer Scales

Nominal

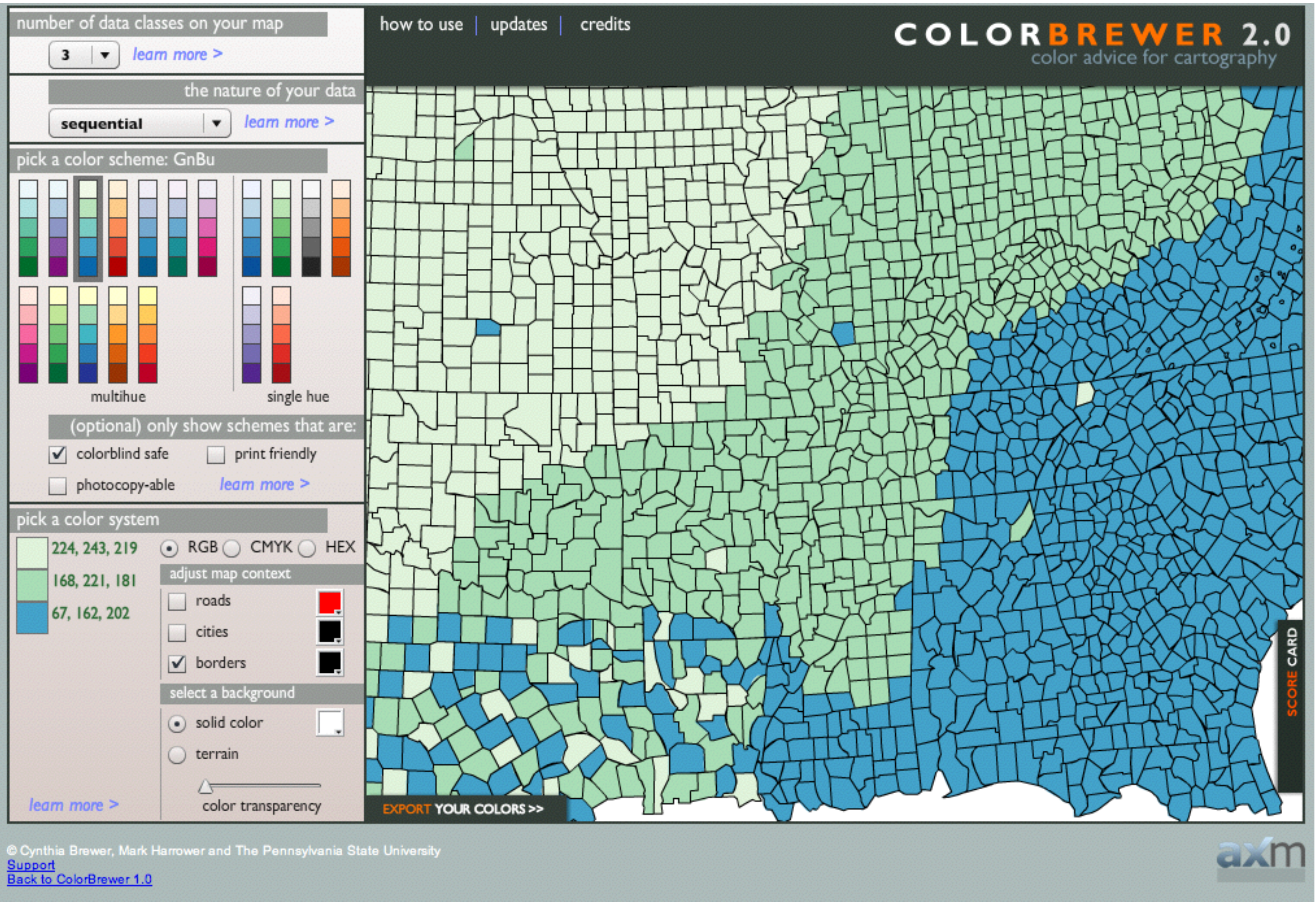
Ordinal



Cynthia Brewer, Color Use Guidelines for Data Representation



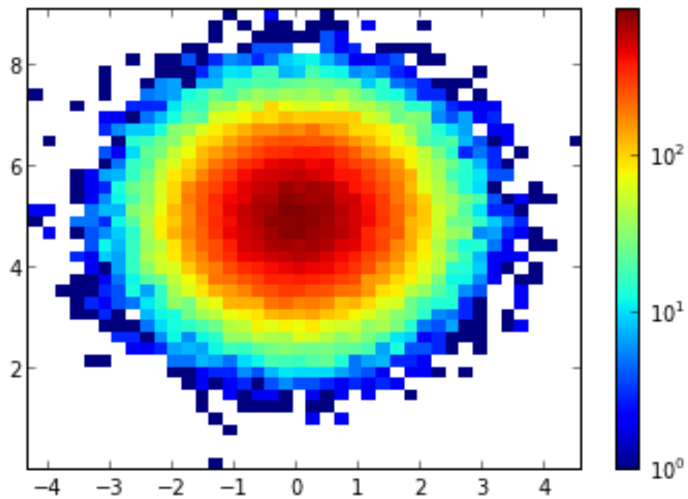
http://www.mathworks.com/matlabcentral/fileexchange/screenshots/6223/original.jpg

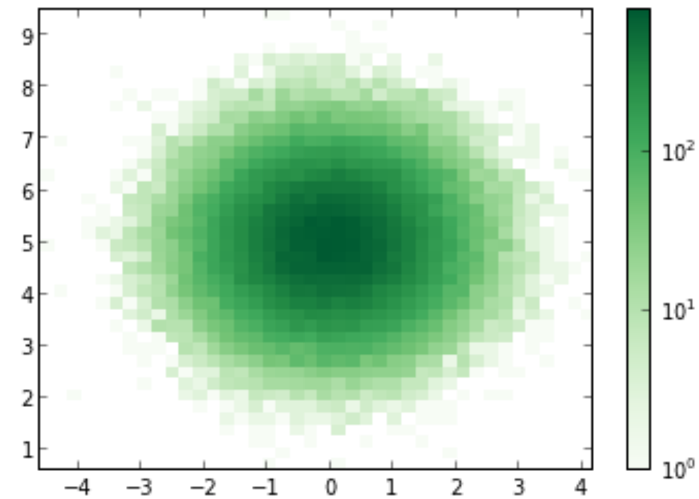


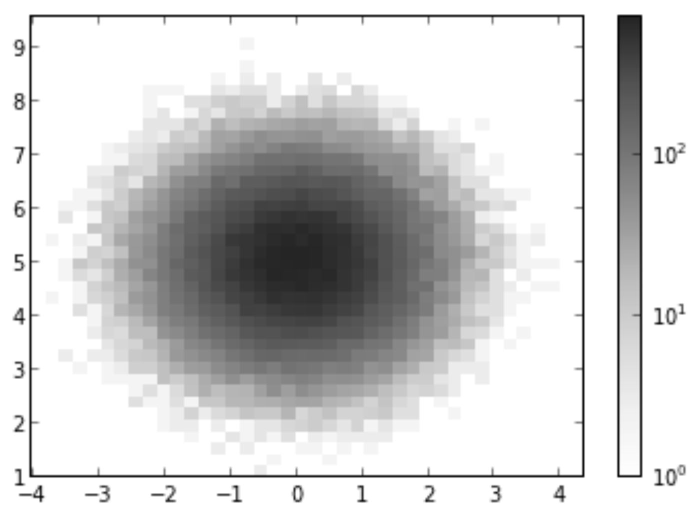
http://colorbrewer2.org

Sequential Brewer Scales

No!



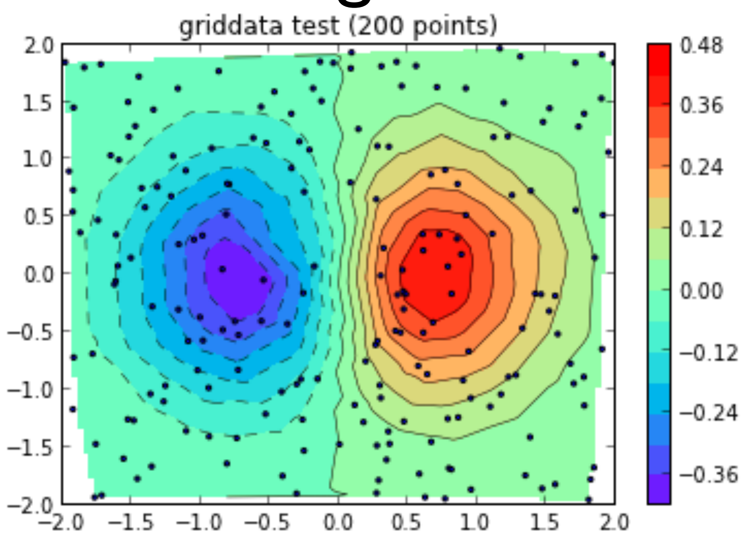


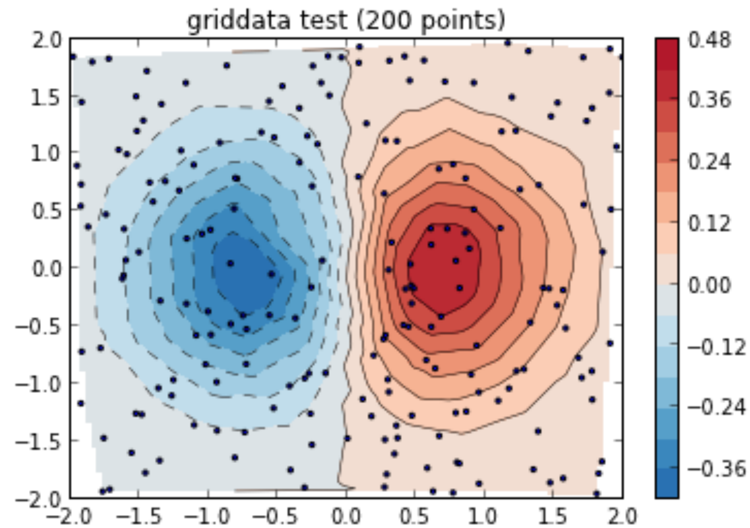


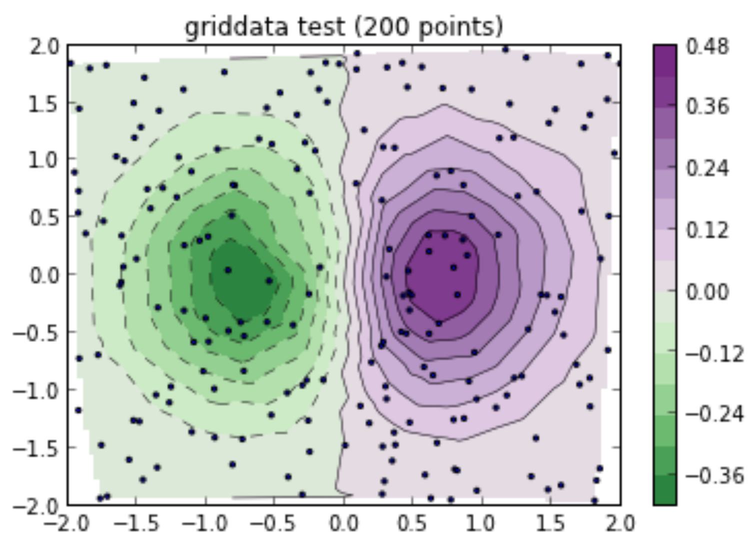
http://nbviewer.ipython.org/5357268

Divergent Brewer Scales

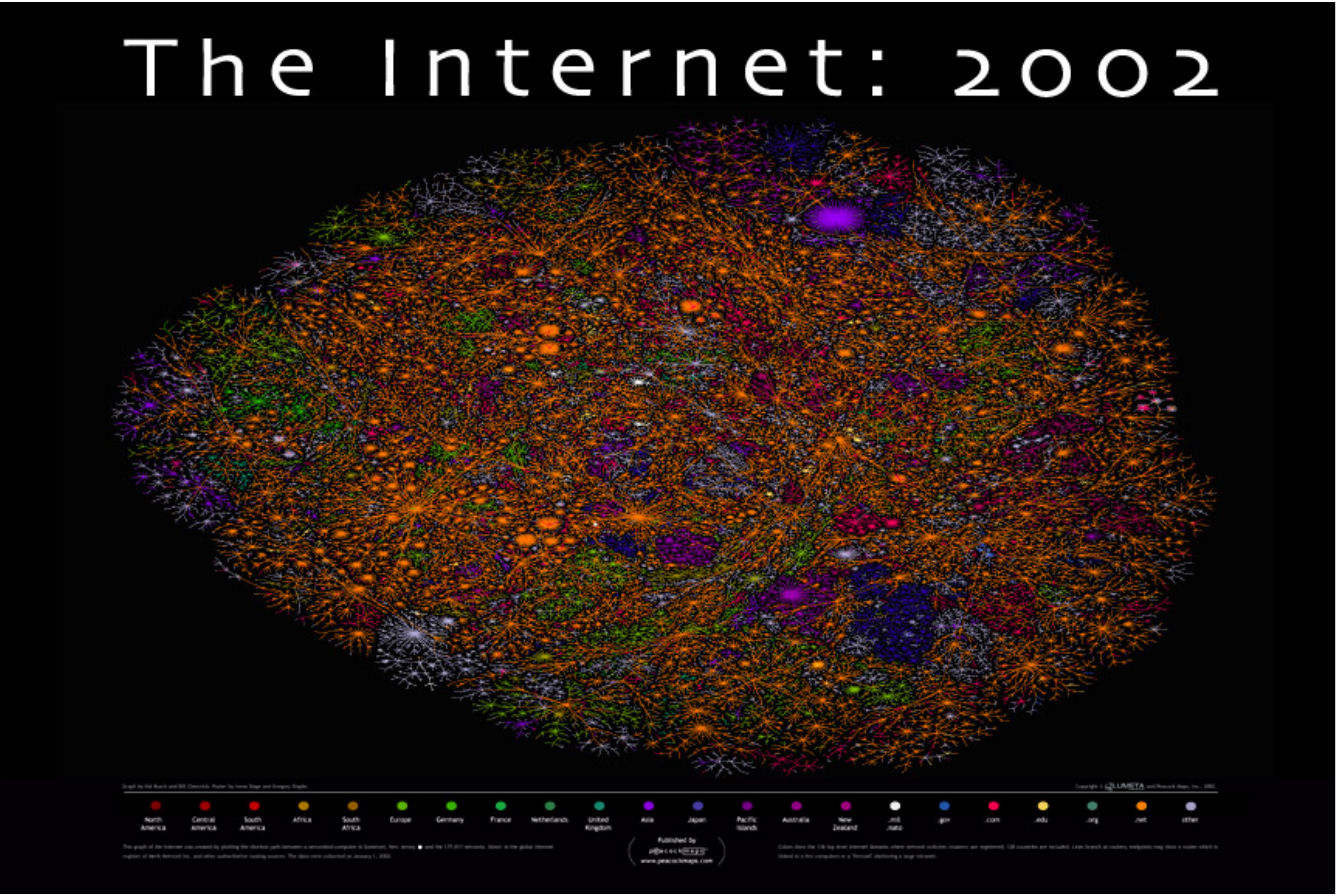
Not great





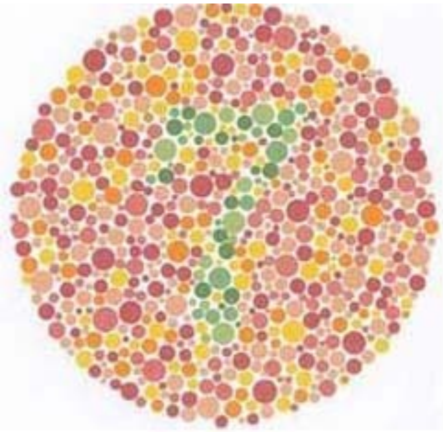


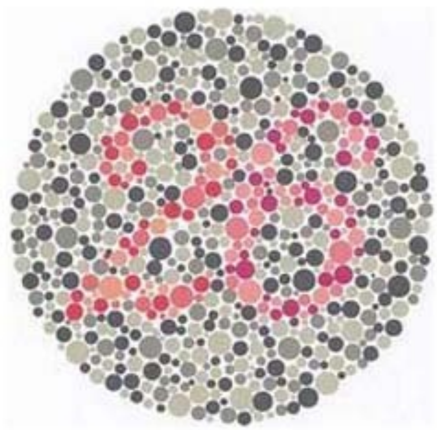
http://nbviewer.ipython.org/5357268

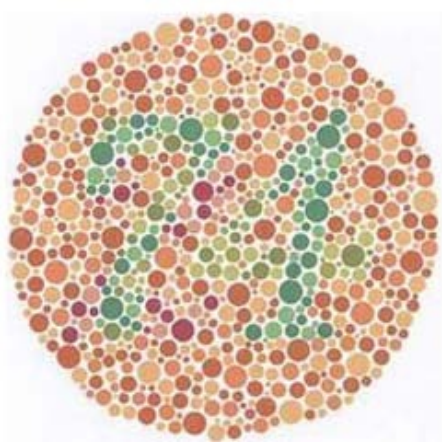


Peacock Maps, 2002

Nominal Data: Do not use more than 6-10 colors!







Color Blindness

8% of males, 1% of females

Most common is red-green weakness / blindness

Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

Bang Wong, PoV, Nature Methods, 2011

Color is Relative

Tufte's Design Principles

Clear, detailed, and thorough labeling and appropriate scales

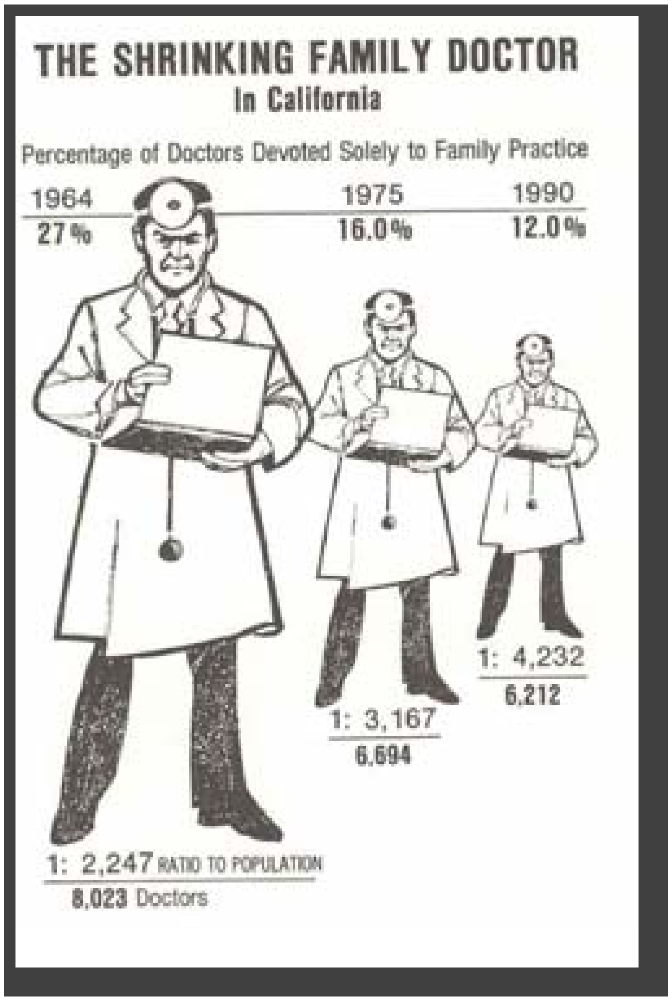
Size of the graphic effect should be directly proportional to the numerical quantities ("lie factor")

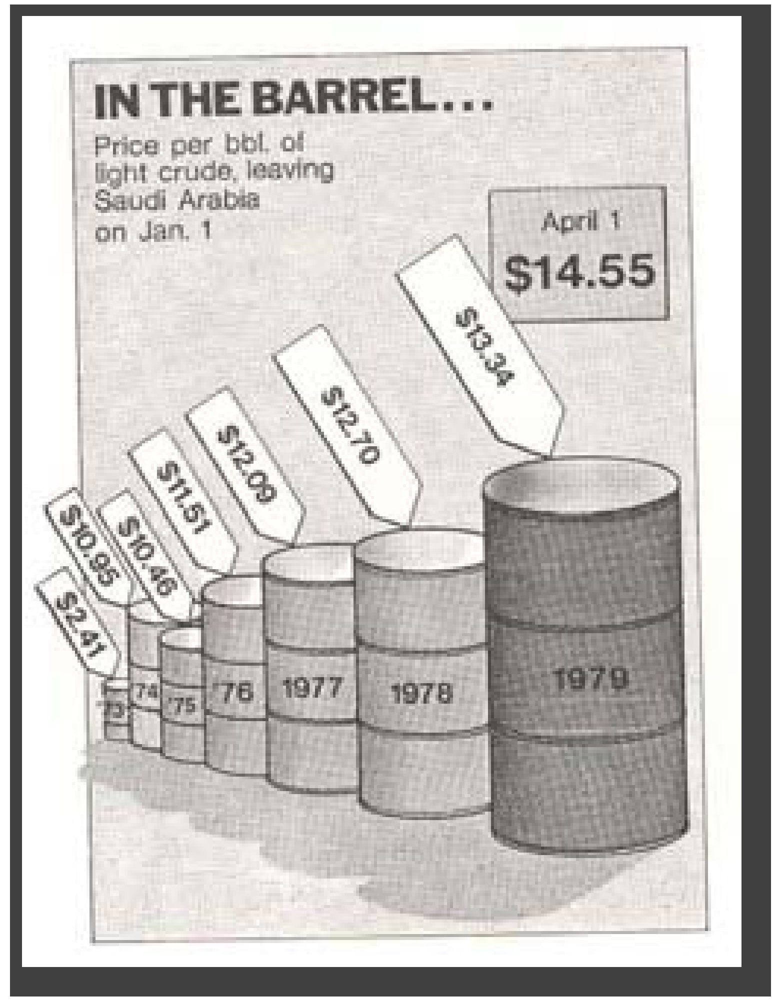
Maximize data-ink ratio

Avoid chart junk

The Lie Factor

$$\frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

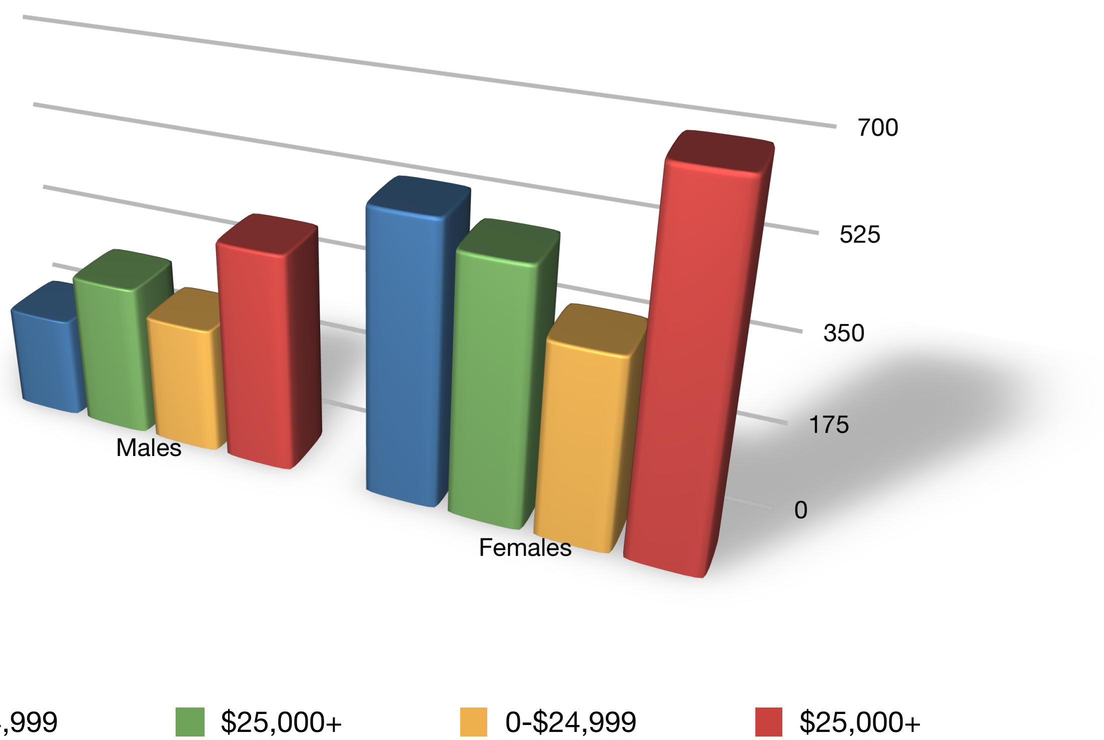




Tufte, VDQI

Maximize Data-Ink Ratio

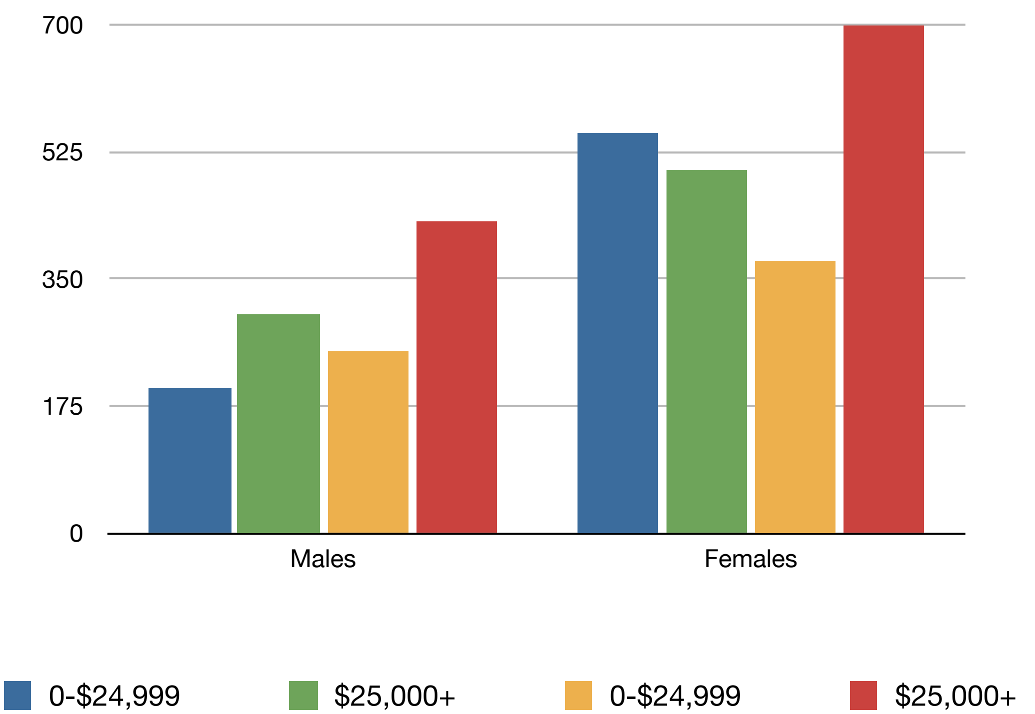
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



Edward Tufte

Maximize Data-Ink Ratio

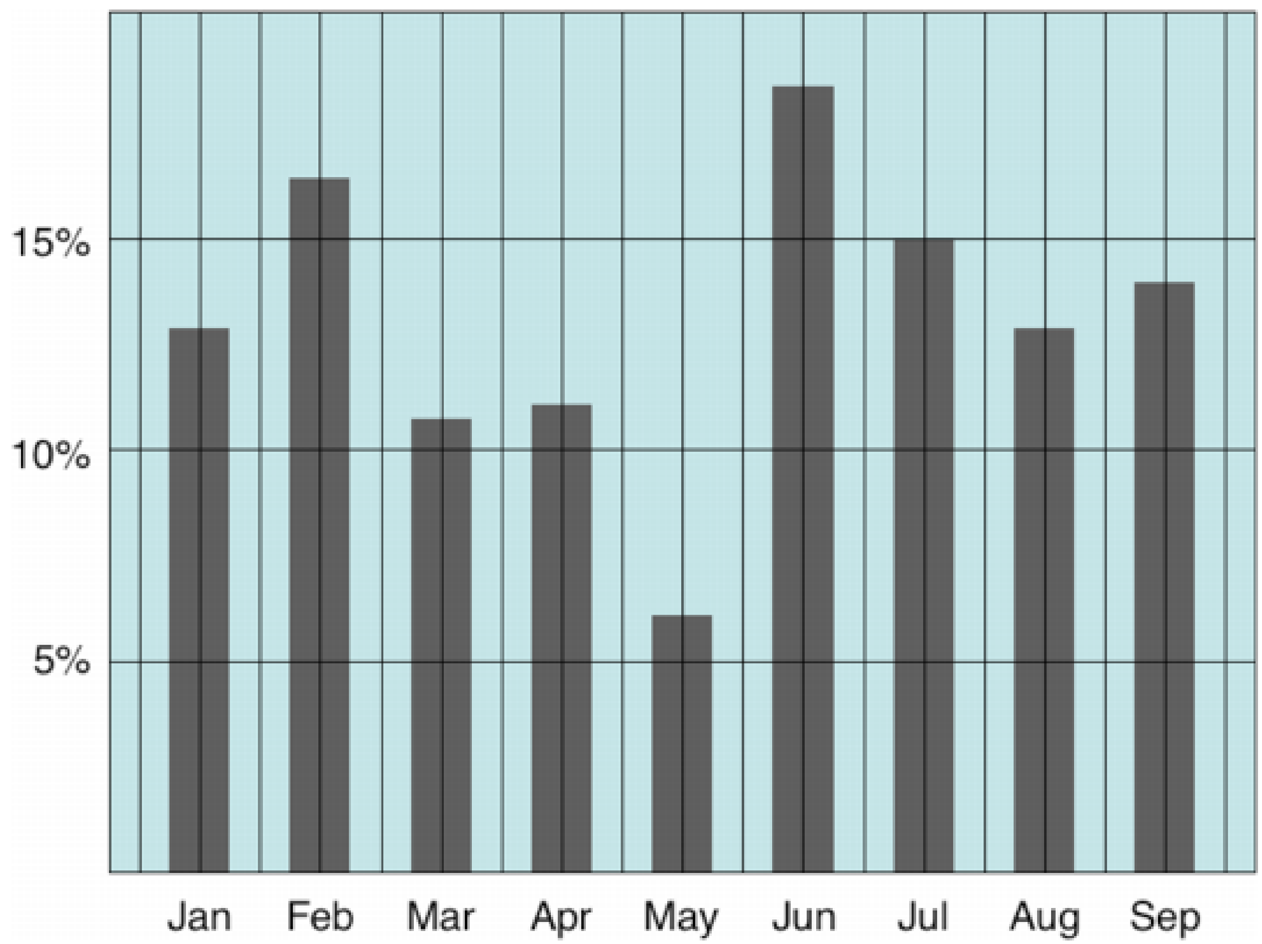
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



Edward Tufte

Avoid Chart Junk

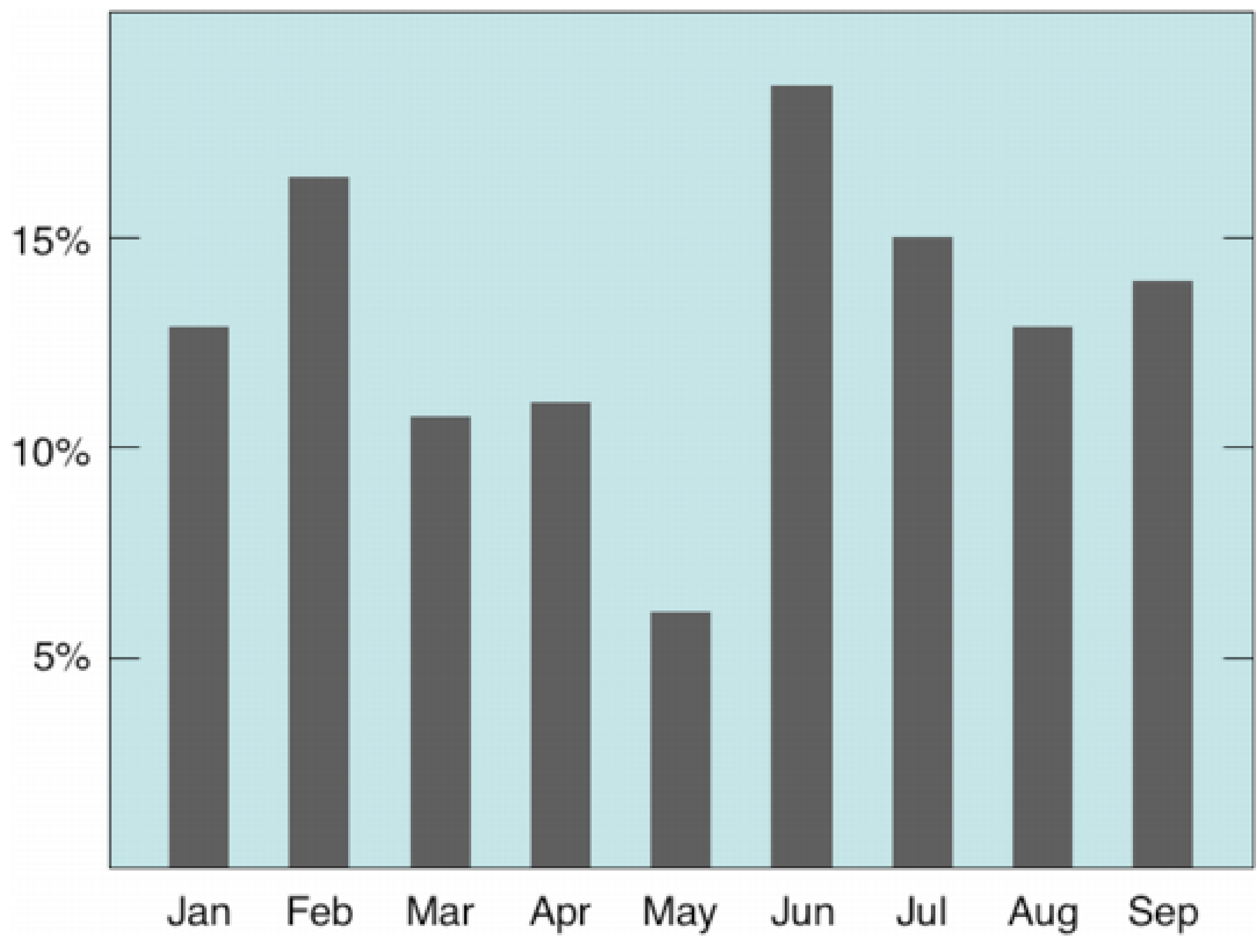
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

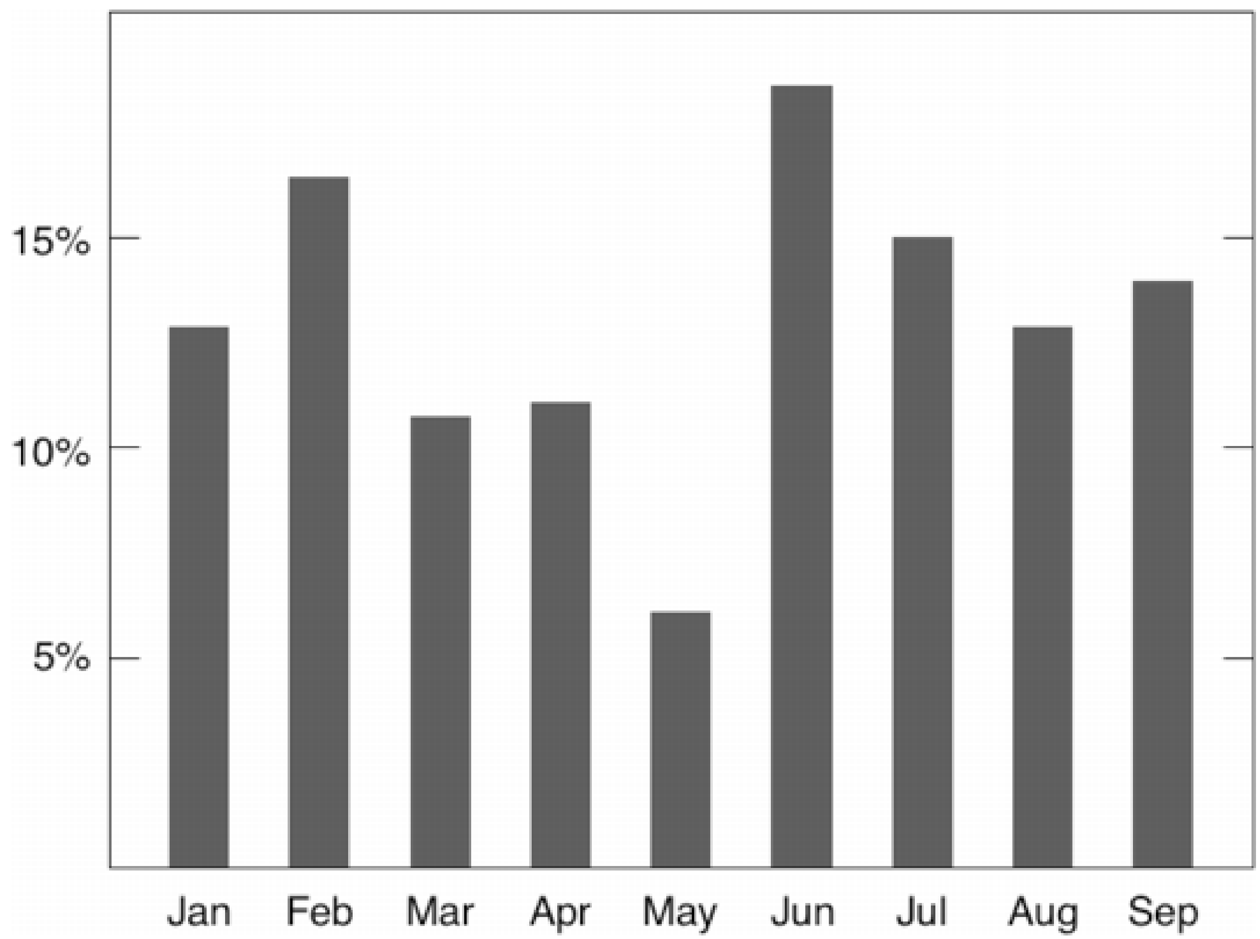
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

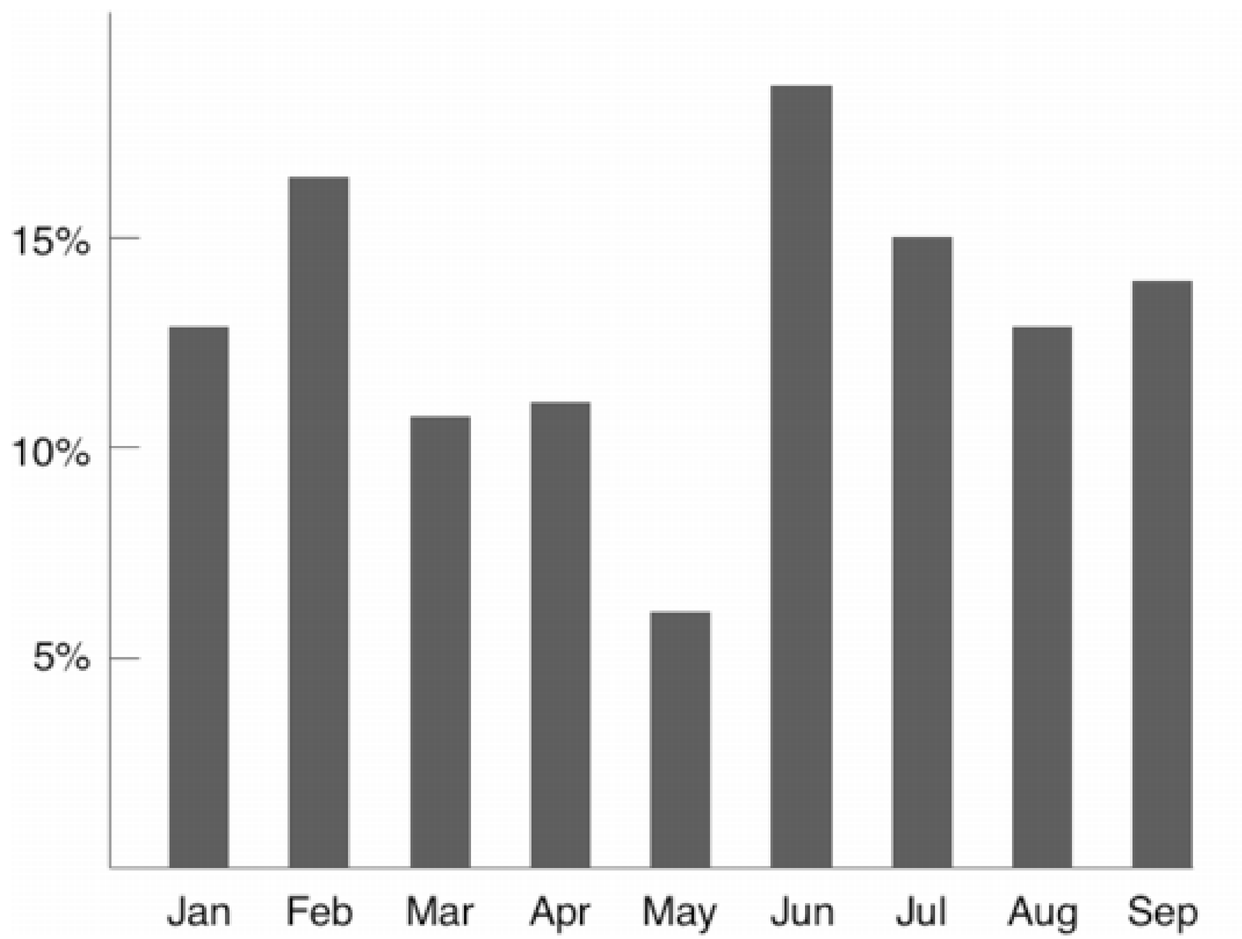
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

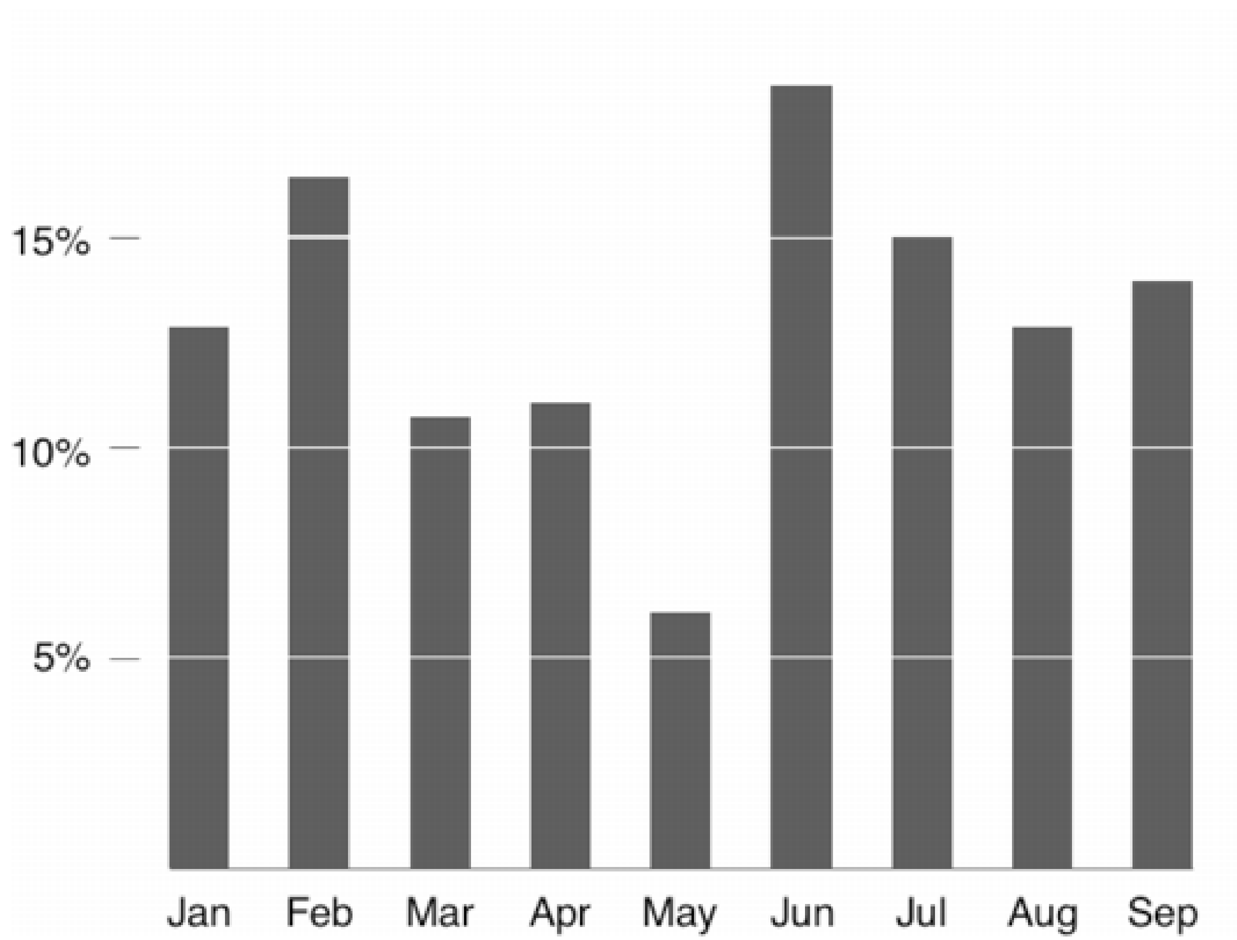
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

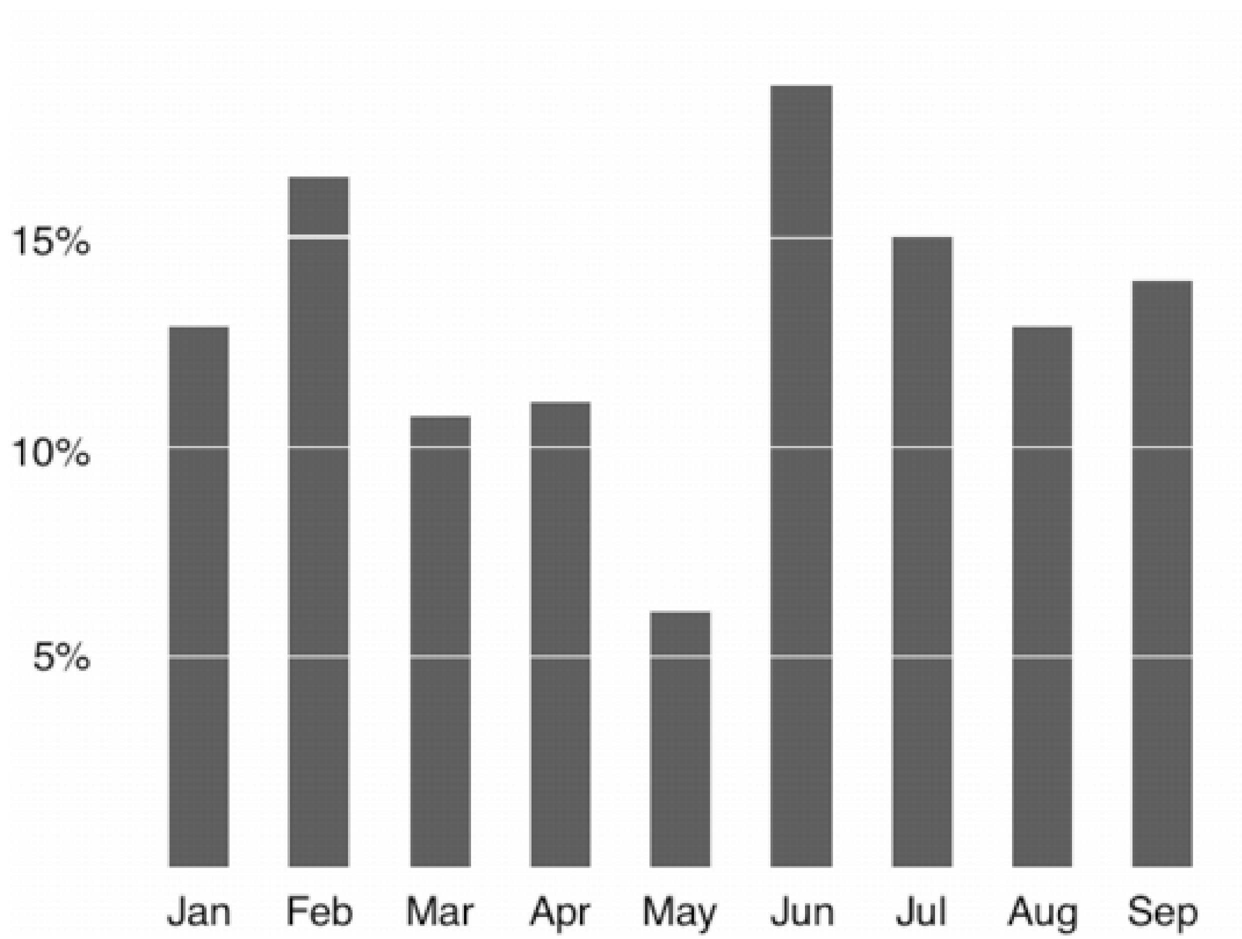
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

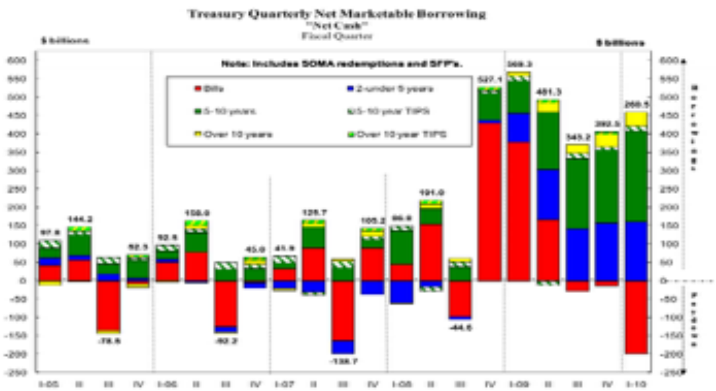
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Effective Visualizations

Not Effective...



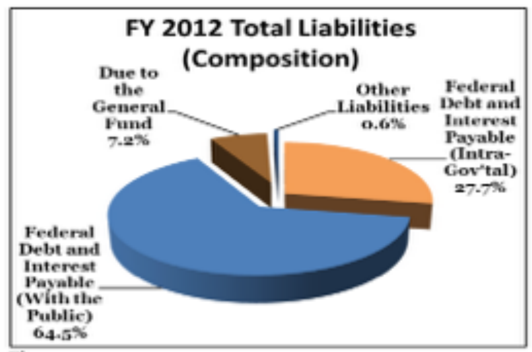
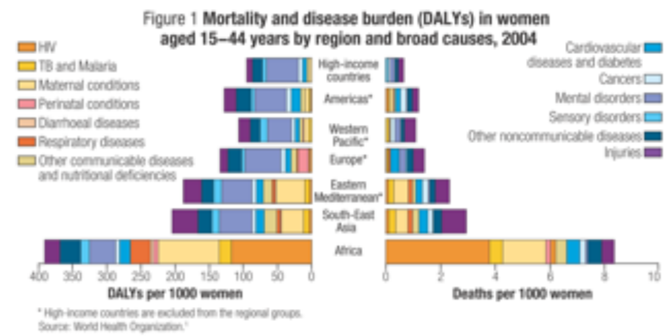
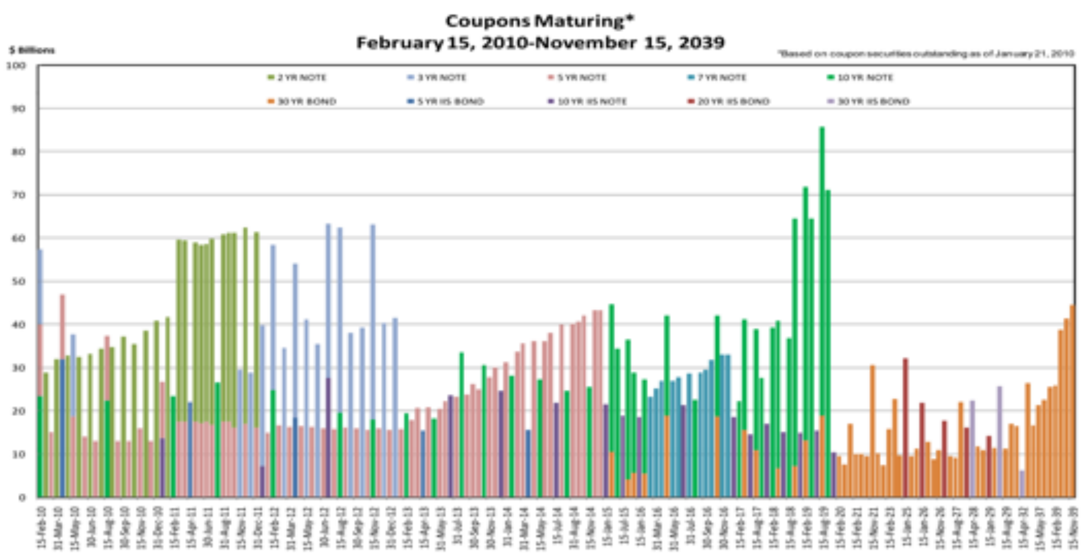


Figure 10





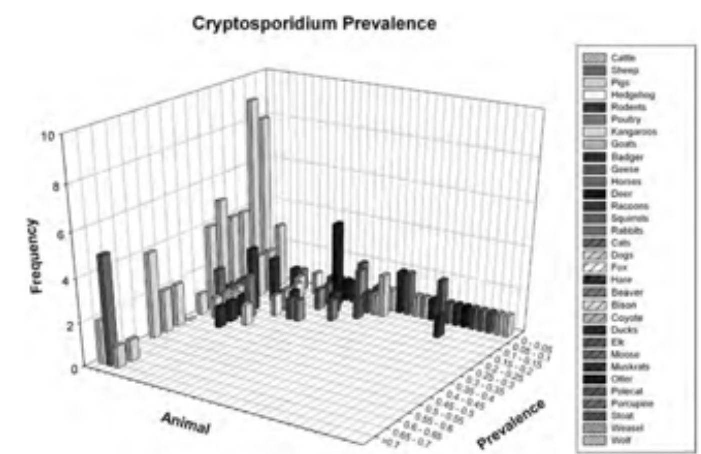
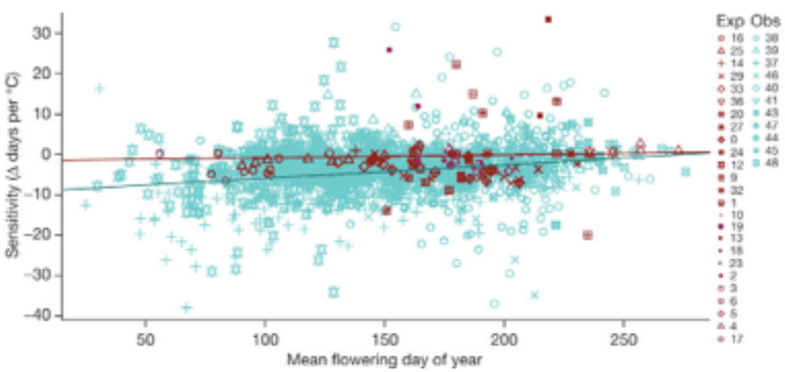
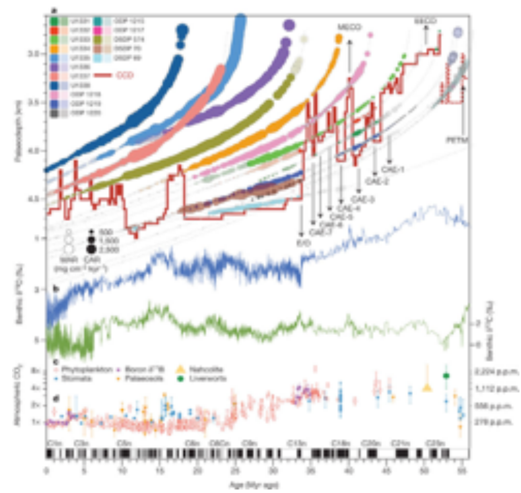


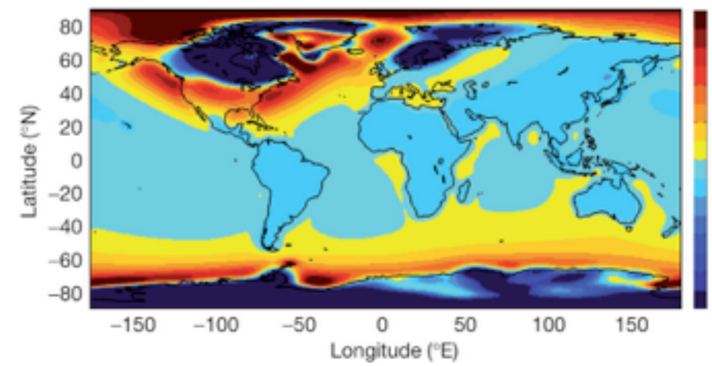
Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

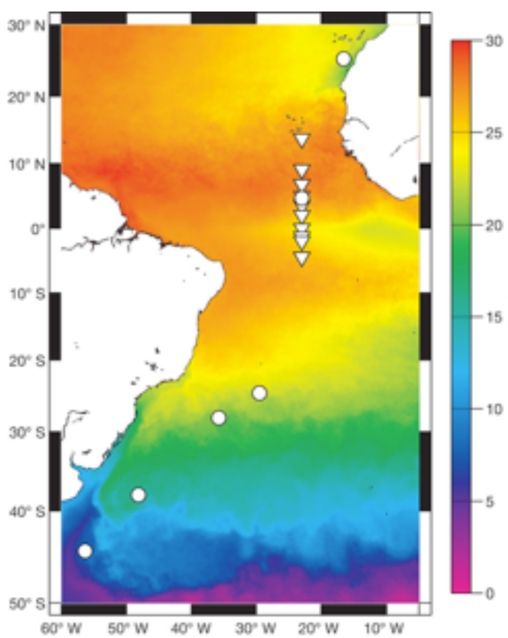
Sources: US Treasury and WHO reports

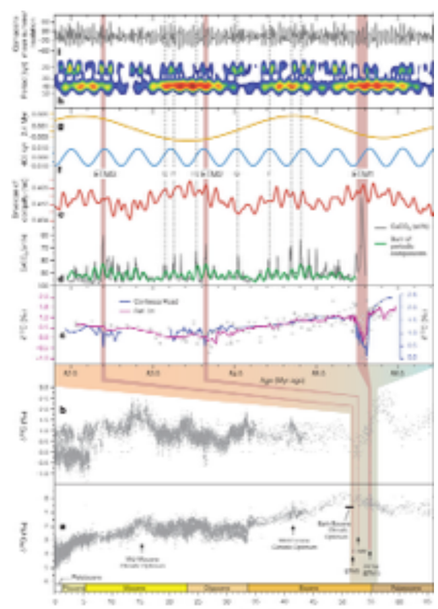
Also *not* effective...

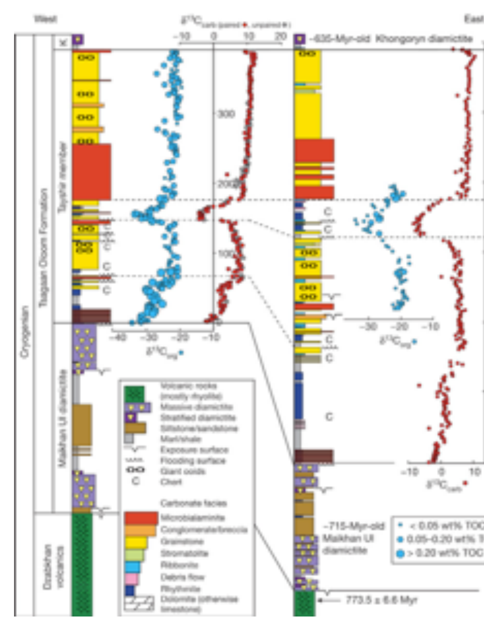


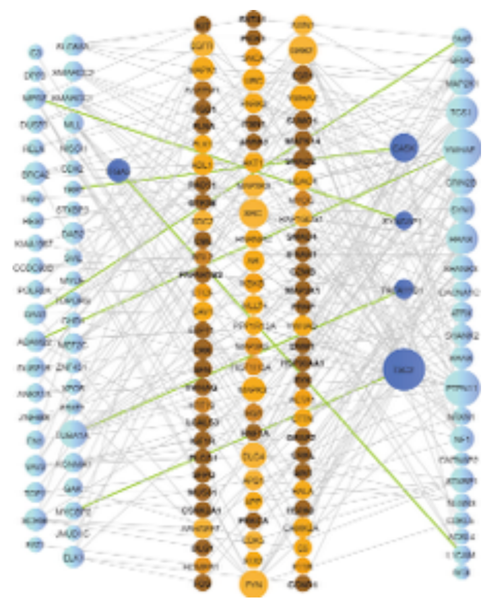








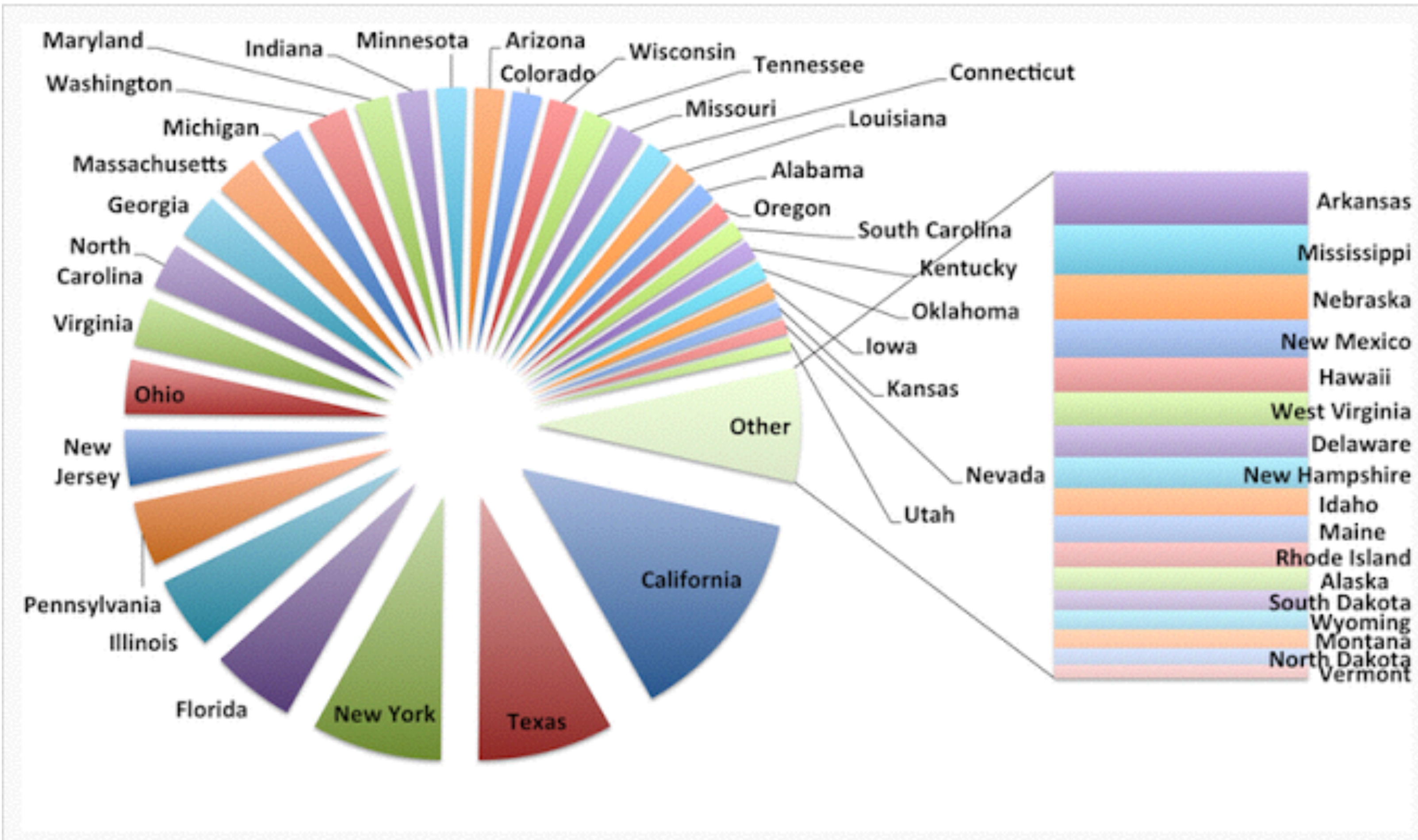




Source: Nature

WTF Visualizations

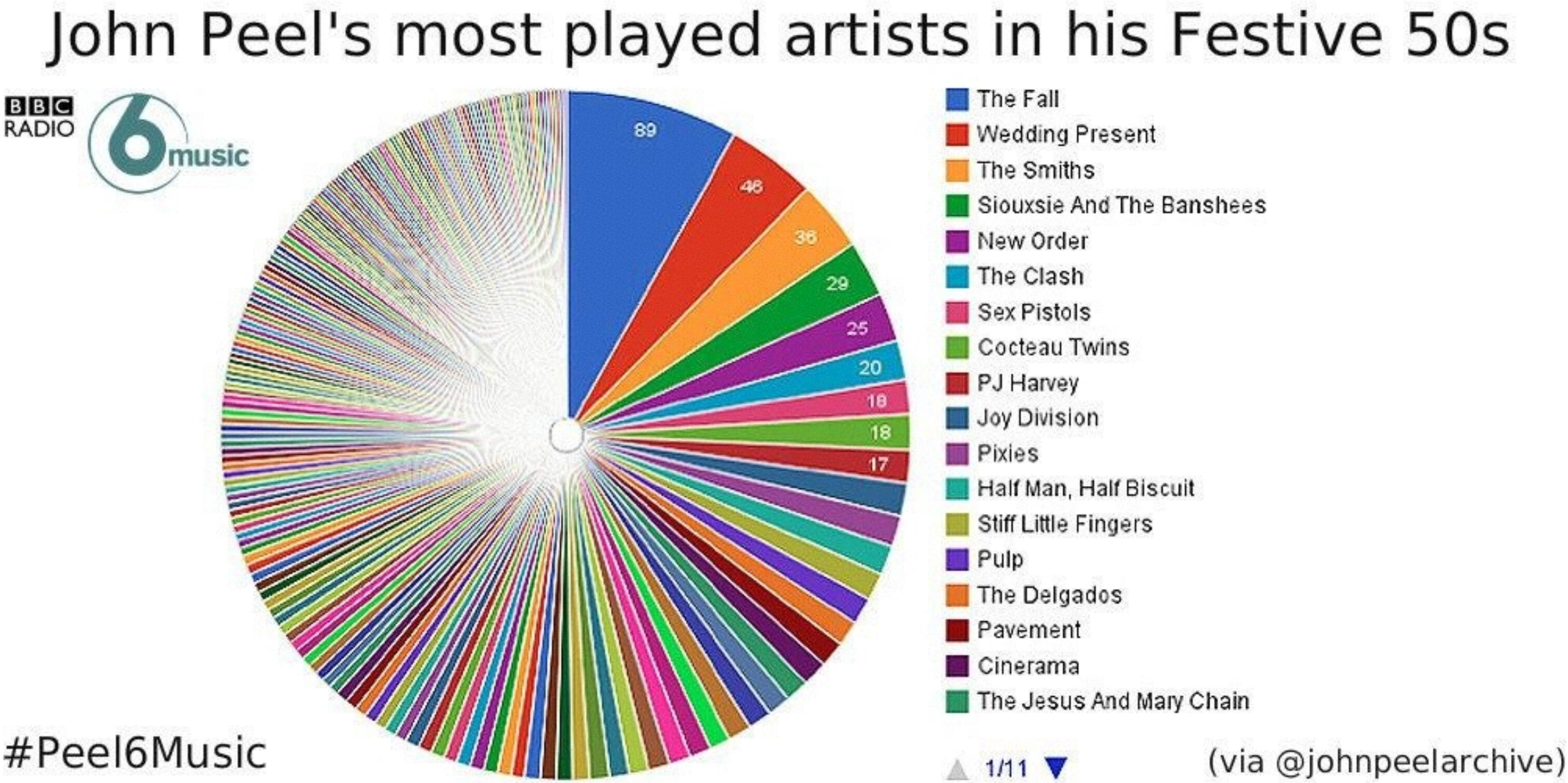
http://wtfviz.net



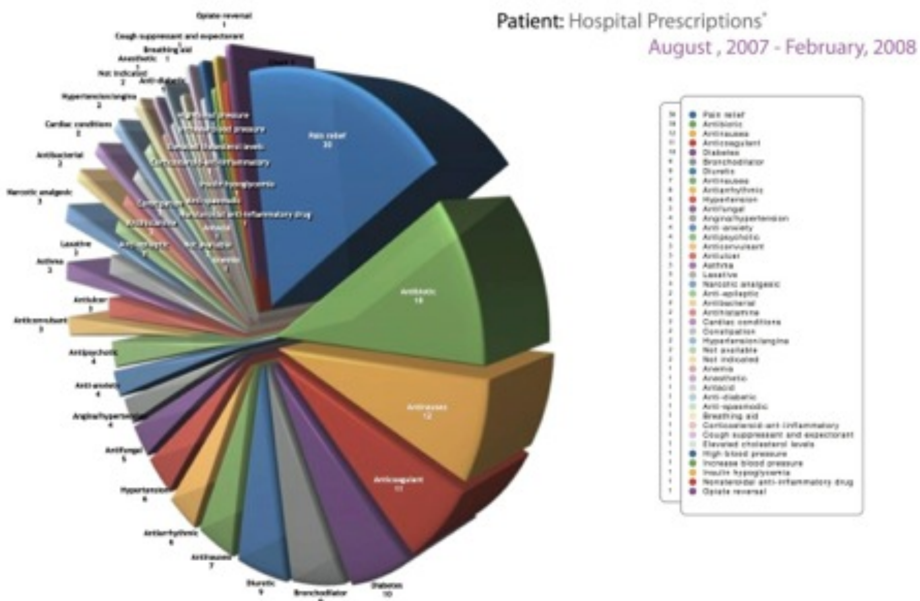
The Unites States GDP for individual states as a contribution to the total US GDP. Fraction of the total US GDP per state was taken from Wikipedia and refers to 2010.

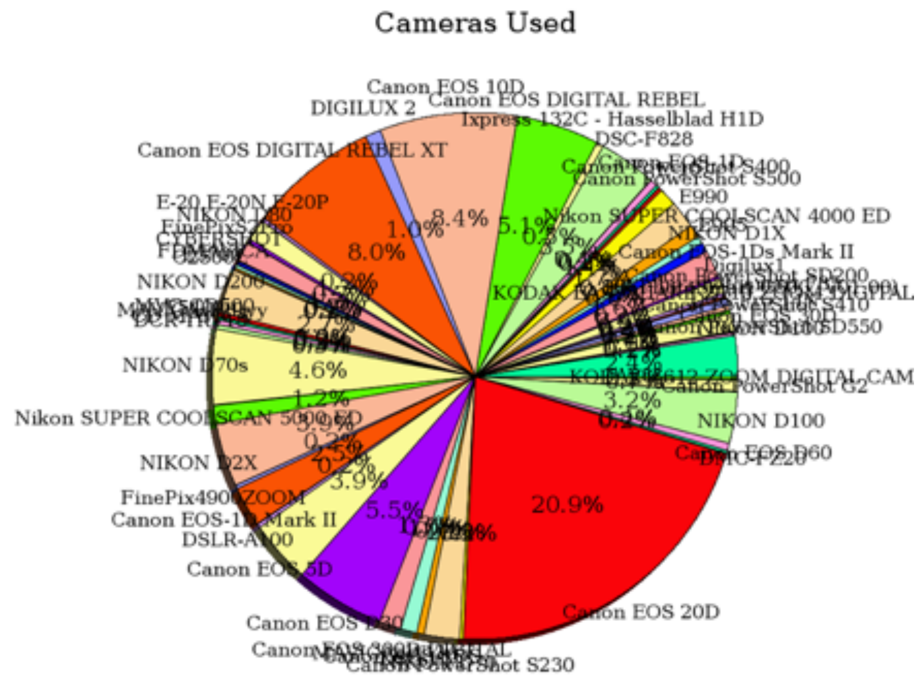
WTF Visualizations

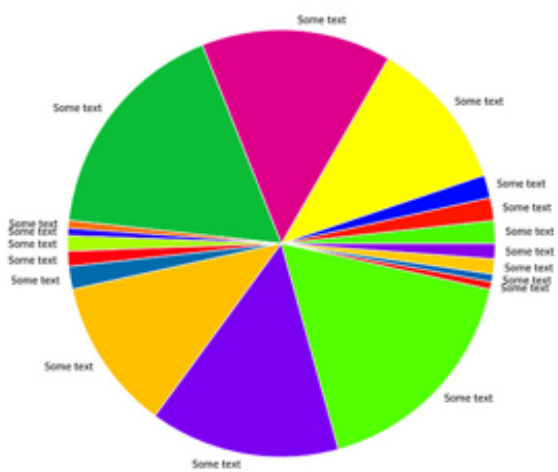
http://wtfviz.net

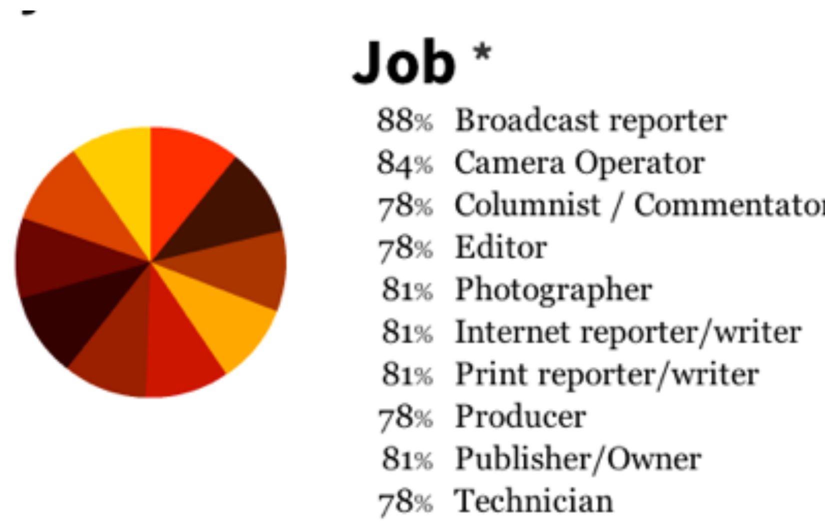


Don’t


Patient: Hospital Prescriptions
August , 2007 - February, 2008

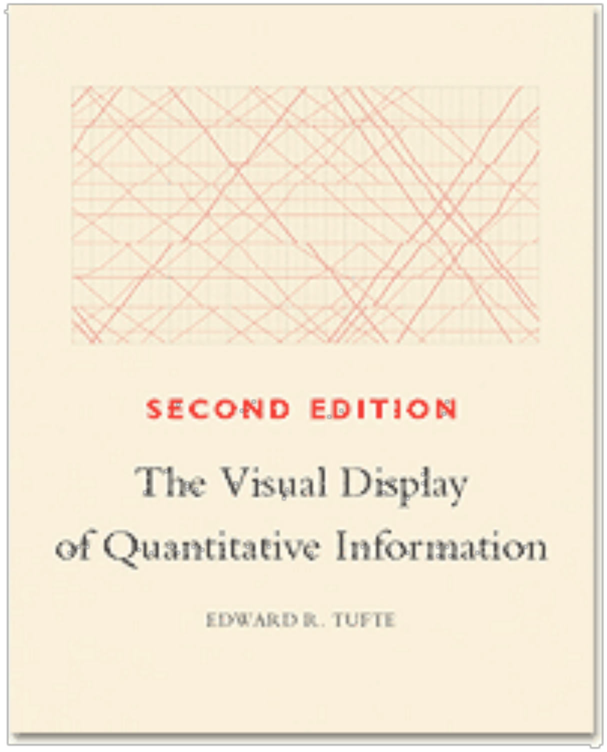

Cameras Used
Canon EOS 10D
DIGILUX 2
Canon EOS DIGITAL REBEL
Canon EOS DIGITAL REBEL XT
DSC-F828
E990
NIKON D1X
NIKON D200
NIKON D70s
Nikon SUPER COOLSCAN 4000 ED
NIKON D100
NIKON D2X
FinePix4900ZOOM
DSLR-A100
Canon EOS 5D
Canon EOS 20D
20.9%
8.4%
8.0%
5.5%
4.6%
3.9%
3.2%

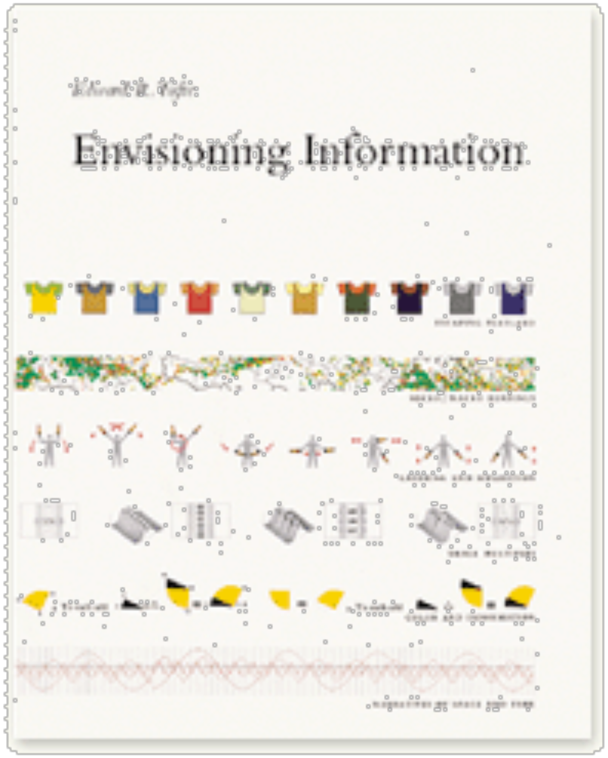

Some text

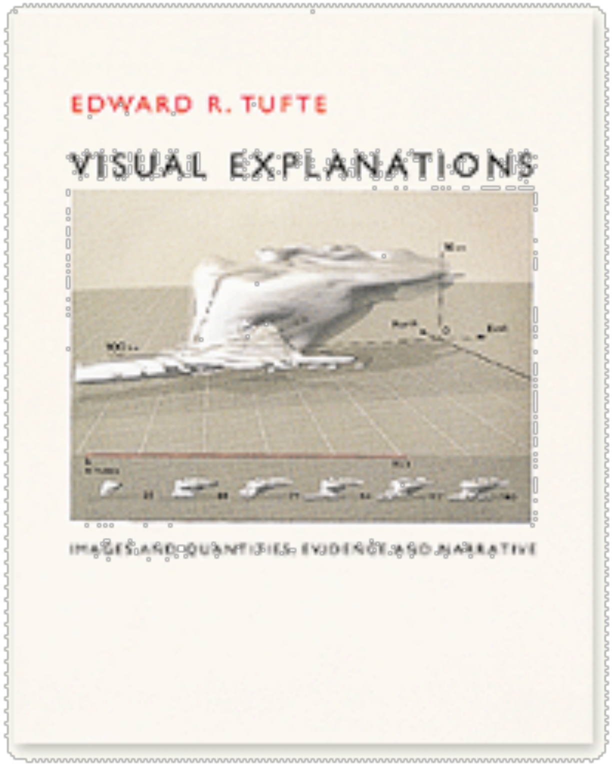

Job *
88% Broadcast reporter
84% Camera Operator
78% Columnist / Commentator
78% Editor
81% Photographer
81% Internet reporter/writer
81% Print reporter/writer
78% Producer
81% Publisher/Owner
78% Technician

Further Reading

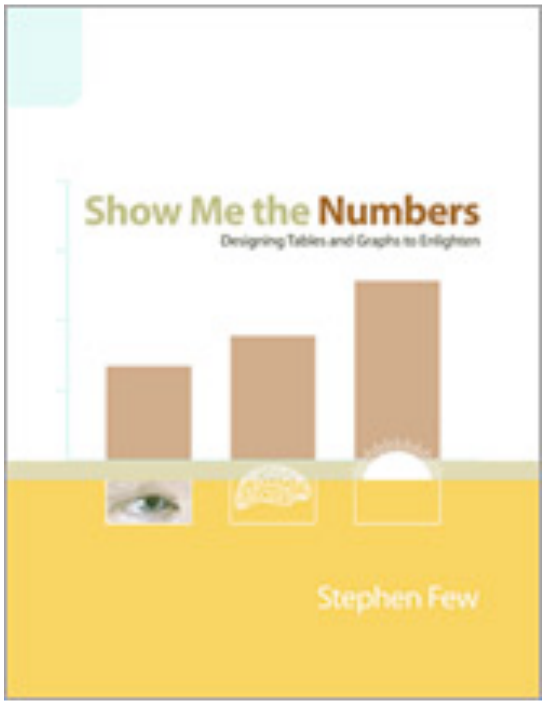
Edward Tufte

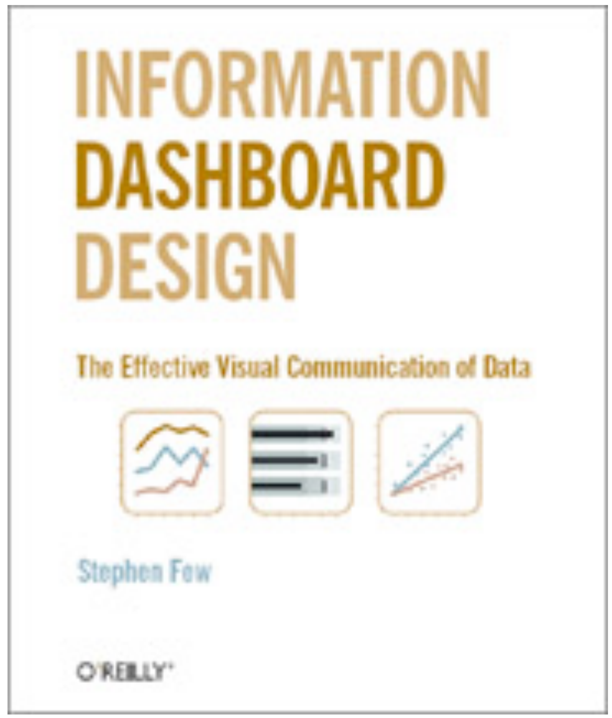


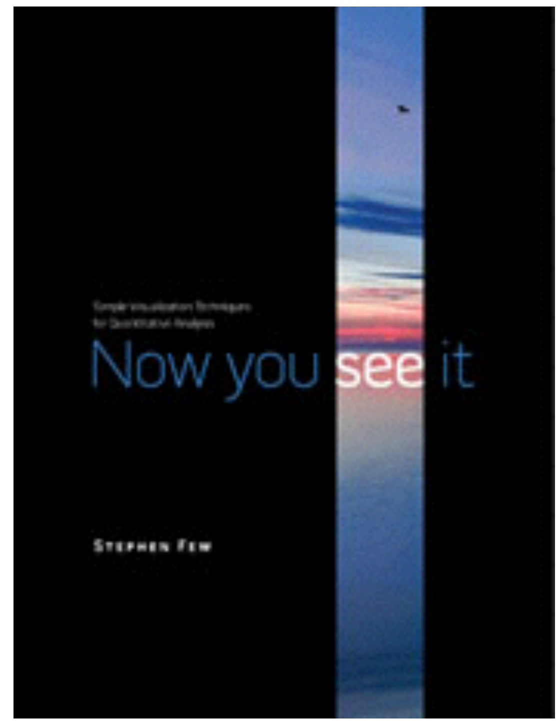




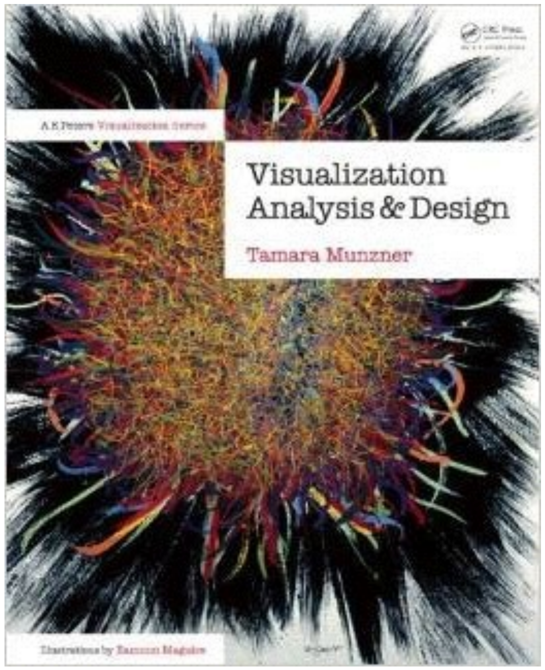
Stephen Few


Show Me the Numbers
Designing Tables and Graphs to Enlighten
Stephen Few


INFORMATION DASHBOARD DESIGN
The Effective Visual Communication of Data
Stephen Few
O'REILLY


Now you see it
Stephen Few

Tamara Munzner



Visualization Analysis and Design

Introduction to PsychoPhysics

PsychoPhysics[1]

- "the scientific study of the relation between stimulus and sensation"
- "the analysis of perceptual processes by studying the effect on a subject's experience or behaviour of systematically varying the properties of a stimulus along one or more physical dimensions"

[1] From Wikipedia

PsychoPhysics[2]

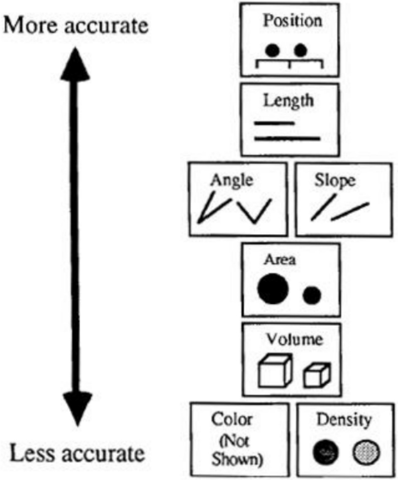
- **Visual Encoding**: the way in which data is mapped into visual structures, upon which we build the images on a screen.
- **Visual Perception:** ability to interpret the surrounding environment by processing information that is contained in visible light.

[2] From Wikipedia

Effective Visual Encoding

- **Challenge:** Pick the best encoding (or mapping) from many possibilities. Consider:
  - **Importance Ordering**: Encode the most important information in the most perceptually accurate way
  - **Expressiveness**: Depict all the data, and only the data
  - **Consistency**: The properties of the image (visual attributes) should match the properties of the data

Importance Ordering: Perceptual Properties



Mackinlay, APT (A Presentation Tool), 1986

Expressiveness

- A length is interpreted as a quantitative value
- Length of bar says something untrue about data

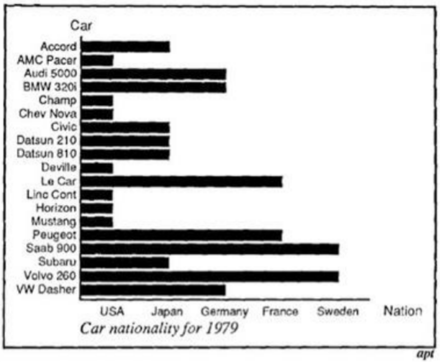
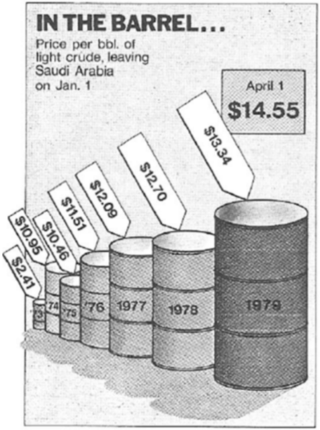


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

Mackinlay, APT (A Presentation Tool), 1986

Consistency

- The properties of the image (visual attributes) should match the properties of the data
- E.g. don't map one-dimensional data to two-or three- dimensional representations!



[Tufte, Edward R (1983), *The Visual Display of Quantitative Information,* Graphics Press, from *Time Magazine*, April 9, 1979, p. 57.]

Visual Perception

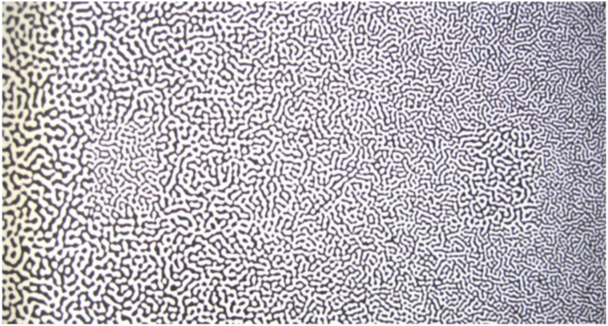
- 70% of body's sense receptors reside in our eyes
- "The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers." Colin Ware, Information Visualization, 2004
- Important to understand how visual perception works in order to effectively design visualizations

How the Eye Works

- The eye is not a camera!
- Better metaphor for vision: "dynamic and ongoing construction project" - Healey, 95
- Attention is selective (filtering)

How to Use Perceptual Properties

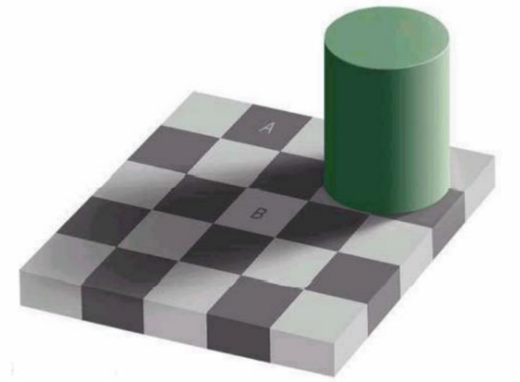
- Information visualization should cause what is meaningful to stand out



Eyes vs. Cameras

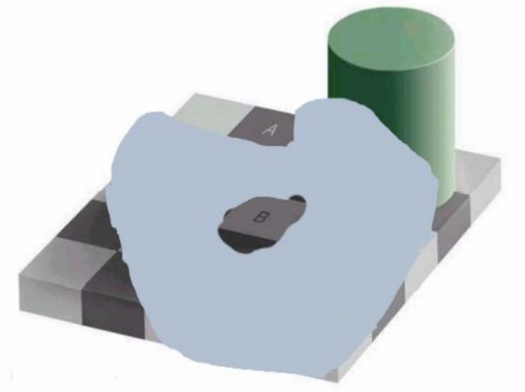
- Cameras
  - Good optics
  - Single focus, white balance, exposure
  - "Full image capture"
- Eyes
  - Relatively poor optics
  - Constantly scanning (saccades)
  - Constantly adjusting focus
  - Constantly adapting (white balance, exposure)
  - Mental reconstruction of image (sort of)

Visual perception is not just camera work



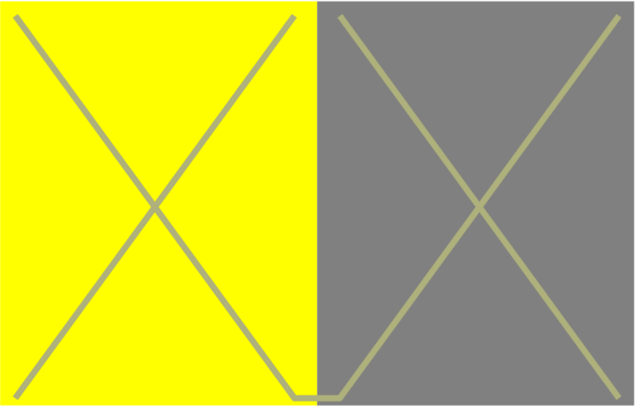
Square A is darker than B, right?

Visual perception is not just camera work



Square A is darker than B, right?

Color is relative



How many 5's

385720939823728196837293827
3829123583834927301228948 39
9090201020328937592730914 28
9383097629658174318692410 24

How many 5's

385720939823728196837293827
382912358383492730122894839
909020102032893759273091428
938309762965817431869241024

Stroop Effect

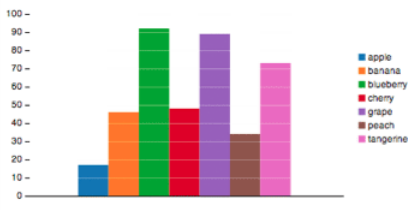
- **Stroop Effect:** interference in the reaction time of a task.
  1. Green Red Blue
     Purple Blue Purple
  2. Blue Purple Red
     Green Purple Green

Stroop Effect Theories[3]

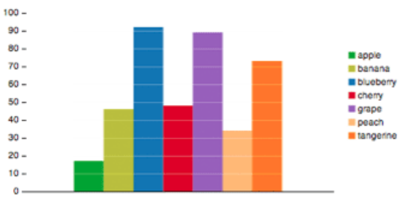
1. brain's ability to recognize the color of the word since the brain reads words faster than it recognizes colors.
2. color recognition as opposed to reading a word, requires more attention.
3. recognizing colors is not an "automatic process" there is hesitancy to respond; whereas, the brain automatically understands the meaning of words as a result of habitual reading.
4. brain analyzes information, different and specific pathways are developed for different tasks. Some pathways, such as reading, are stronger than others

[3] From Wikipedia

Semantically Resonant Color Assignments



Default color assignment



Semantically resonant color assignment

Preattentive Processing

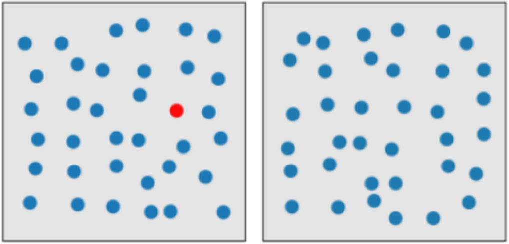
- Certain basic visual properties are detected immediately by low-level visual system
- "Pop-out" vs. serial search
- $<$ 200 - 250ms qualifies as preattentive
    - eye movements take at least 200ms
    - yet certain processing can be done very quickly, implying low-level processing in parallel
- If a decision takes a fixed amount of time regardless of the number of distractors, it is considered to be **preattentive**.

Preattentive Processing

- A limited set of visual properties are processed preattentively
  - (without need for focusing attention).
- This is important for design of visualizations
  - What can be perceived immediately?
  - Which properties are good discriminators?
  - What can mislead viewers?

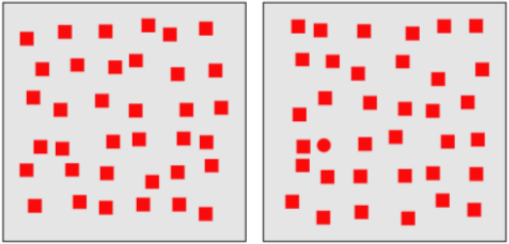
Color (Hue) is Preattentive

- Detection of red circle in group of blue circles is Preattentive



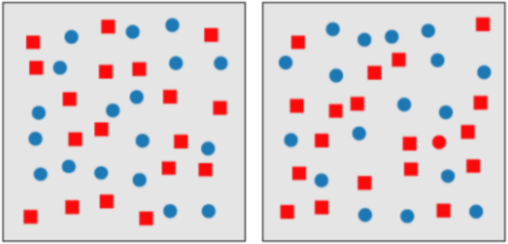
Form (curvature) is preattentive

- Curved form “pops out” of display

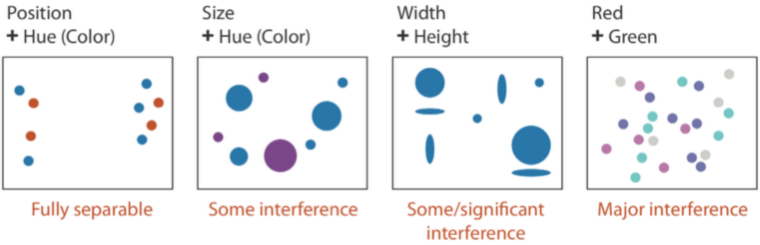


Conjunction of Attributes

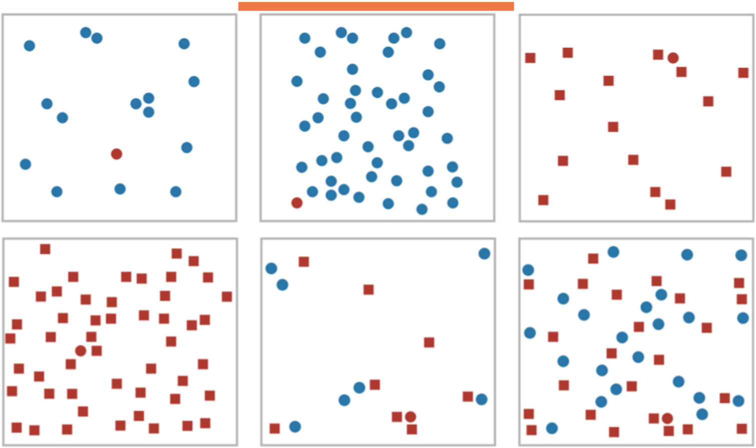
- Conjunction target generally cannot be detected preattentively (red circle in sea of red square and blue circle distractors)



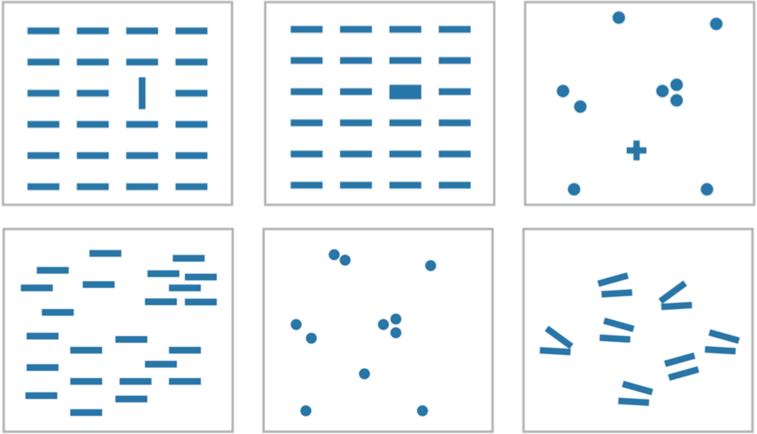
Separability of Attributes



Visual Popout (Preattentive Features) - I

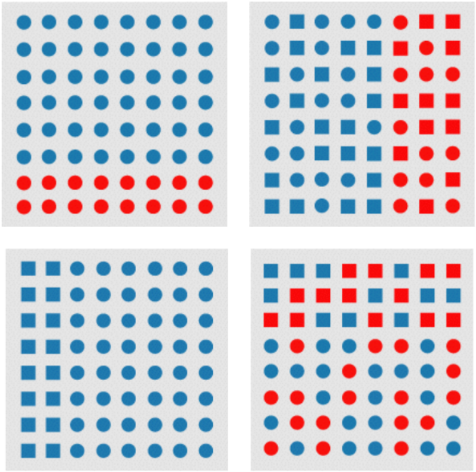


Visual Popout (Preattentive Features) - II

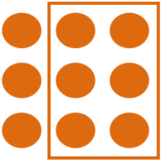


http:/www.csc.ncsu.edu/faculty/healey/PP

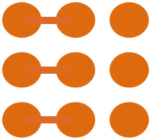
Feature Hierarchy



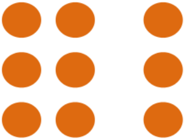
Grouping Principles



Containment



Connection



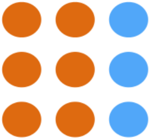
Proximity

Collins et al. 2009

D3.js Example

D3.js Example

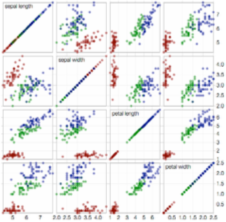
Grouping Principles



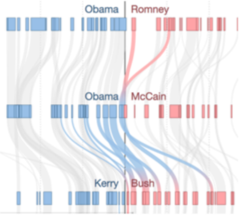
Similarity

Continuity

Common Fate



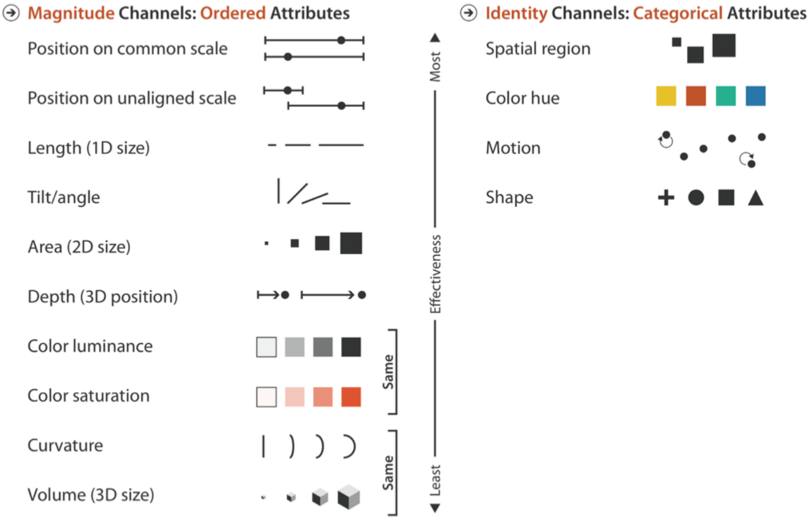
D3.js Example



NYT Swing States

Closure

Munzner Hierarchy



T. Munzner, Visualzation Analysis and Design, 2014

Preattentive Visual Properties (Healey 97)

length	Triesman & Gormican [1988]
width	Julesz [1985]
size	Triesman & Gelade [1980]
curvature	Triesman & Gormican [1988]
number	Julesz [1985]; Trick & Pylyshyn [1994]
terminators	Julesz & Bergen [1983]
intersection	Julesz & Bergen [1983]
closure	Enns [1986]; Triesman & Souther [1985]
color (hue)	Nagy & Sanchez [1990, 1992]; D'Zmura [1991] Kawai et al. [1995]; Bauer et al. [1996]
intensity	Beck et al. [1983]; Triesman & Gormican [1988]
flicker	Julesz [1971]
direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
binocular luster	Wolfe & Franzel [1988]
stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
lighting direction	Enns [1990]

Critiquing a Visualization

1. First, consider the purpose of the visualization and who the intended audience is.
2. Then, ascertain your initial reaction.
3. Then, examine the visualization in detail.
4. Then, answer questions like the following.

Over-Arching Questions

1. Is the design visually appealing/aesthetically pleasing?
2. Is it immediately understandable? If not, is it understandable after a short period of study?
3. Does it provide insight or understanding that was not obtainable with the original representation (text, table, etc)?
4. Does it provide insight or understanding better than some alternative visualization would? Or does it require excessive cognitive effort? What kind of visualization might have been better?

How Successful is the Visualization?

5. Does the visualization reveal trends, patterns, gaps, and/or outliers? Can the viewer make effective comparisons?
6. Does the visualization successfully highlight important information, while providing context for that information?
7. Does it distort the information? If it transforms it in some way, is this misleading or helpfully simplifying?
8. Does it omit important information?
9. Is it memorable?

Questions about the Visual Transformation

10. Does it use visual components properly?
    - Does it properly represent the data using lines, color, position, etc?
    - Does it transform nominal, ordinal, and quantitative information properly?
11. Does it use labels and legends appropriately?

Summary

Major Concepts:

- Visual Attributes
- Principles of effective visualization
- Visual encoding and perception

Slide Material References

- Slides from Harvard CS 109 (2013 and 2014)
- Slides by Cecilia Aragon