The Science of Information Visualization

Houston Data Visualization Meetup

October 2015

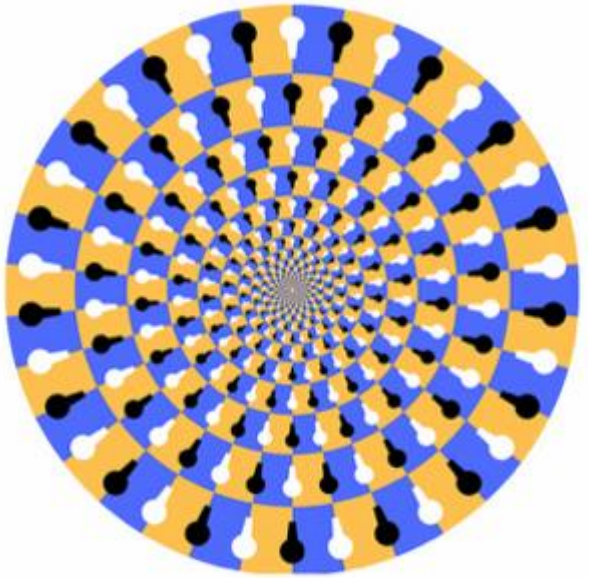
Overview

Visual Perception

- 70% of the body's sense receptors reside in our eyes
- "The eye and 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

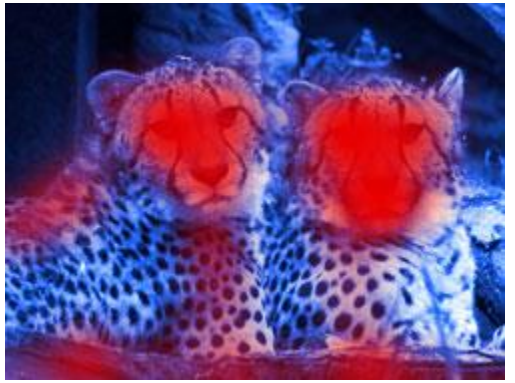
- The eye is not a camera!
- Better metaphor for vision: "dynamic and ongoing construction project" – Healey, 95
- Attention is selective (filtering)



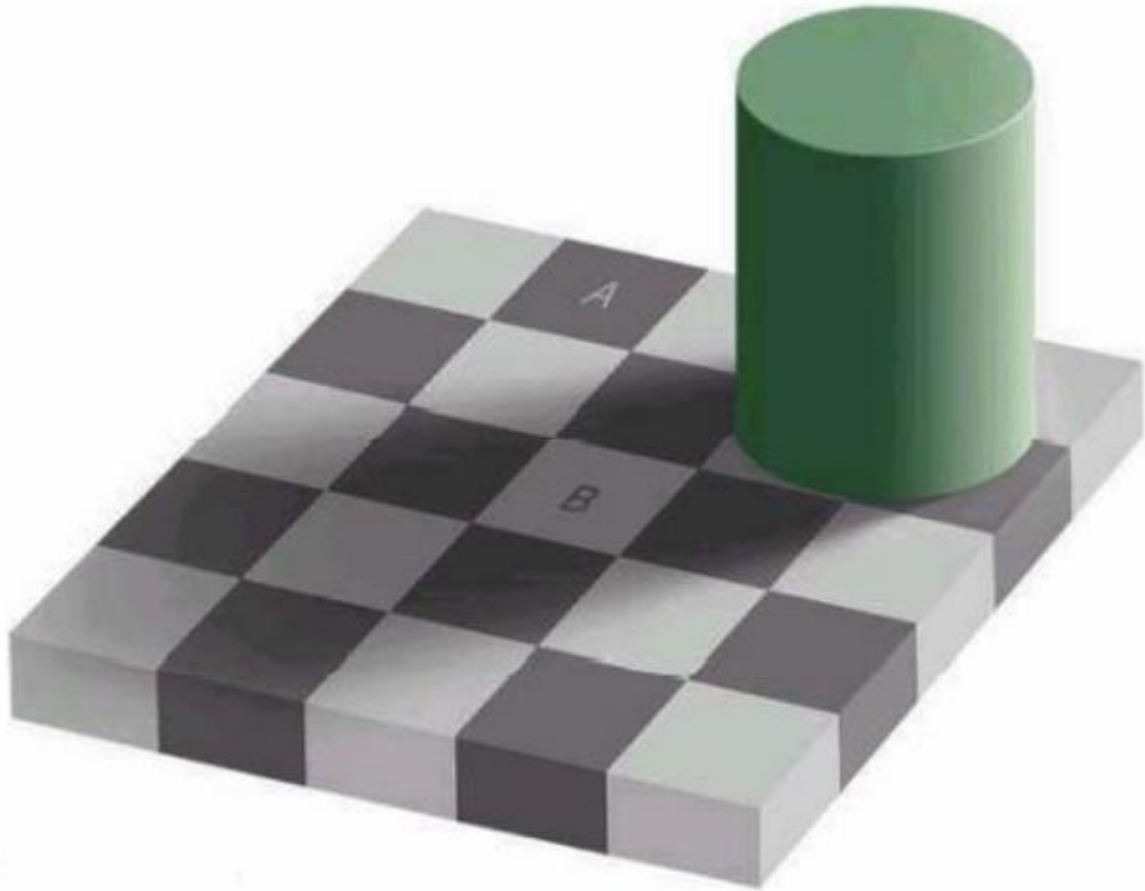
Eyes vs. Cameras

- Cameras
  - Good optics
  - Single focus, white balance, exposure
  - “Full image capture”

- Eyes
  - Relatively poor optics
  - Constantly scanning
  - Constantly adjusting focus
  - Constantly adapting (white balance, exposure)
  - Mental reconstruction of image (sort of)

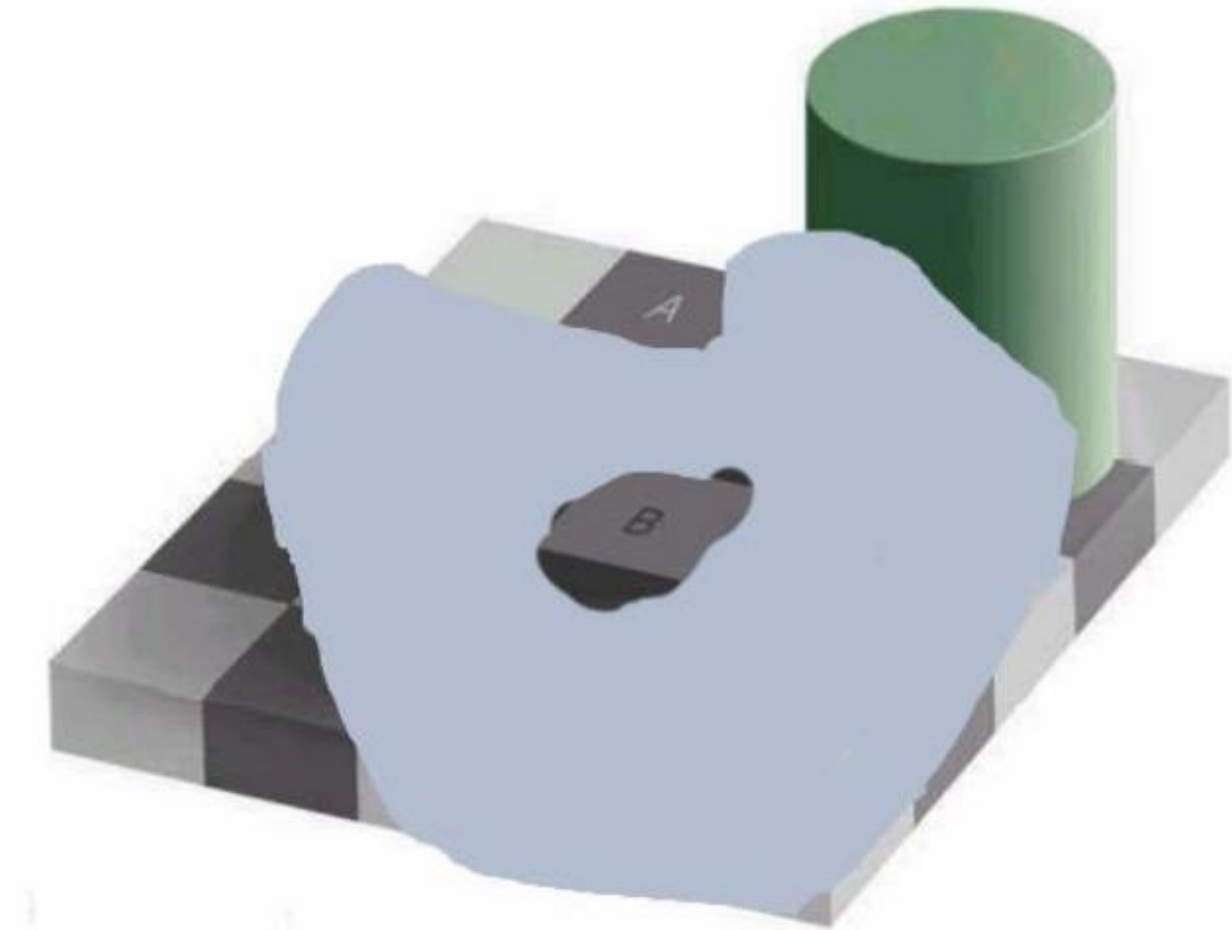


Visual perception is not just camera work



Square A is darker than B, right?

Visual perception is not just camera work



Square A is darker than B, right?

How many 5's

38572093982372819683729382 7
38291235838349273012289483 9
90902010203289375927309142 8
93830976296581743186924102 4

How many 5's

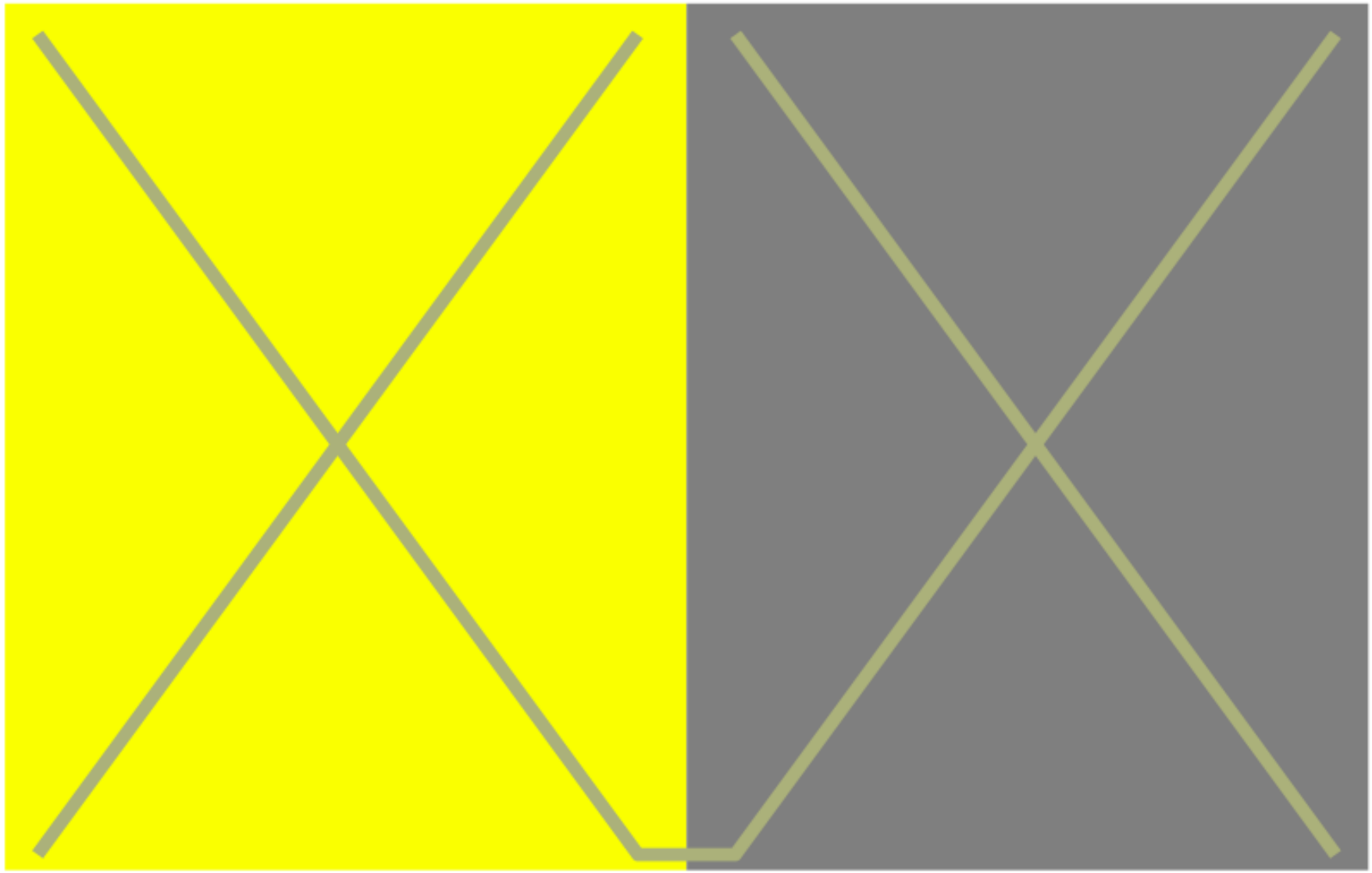
38572093982372819683729382 7

38291235838349273012289483 9

90902010203289375927309142 8

93830976296581743186924102 4

Color is relative



Visual Marks

Basic geometric elements

Points

0D

Lines

1D

Areas

2D

Visual Variables (aka Channels)

- **Position**
  - Horizontal
  - Vertical
  - Both

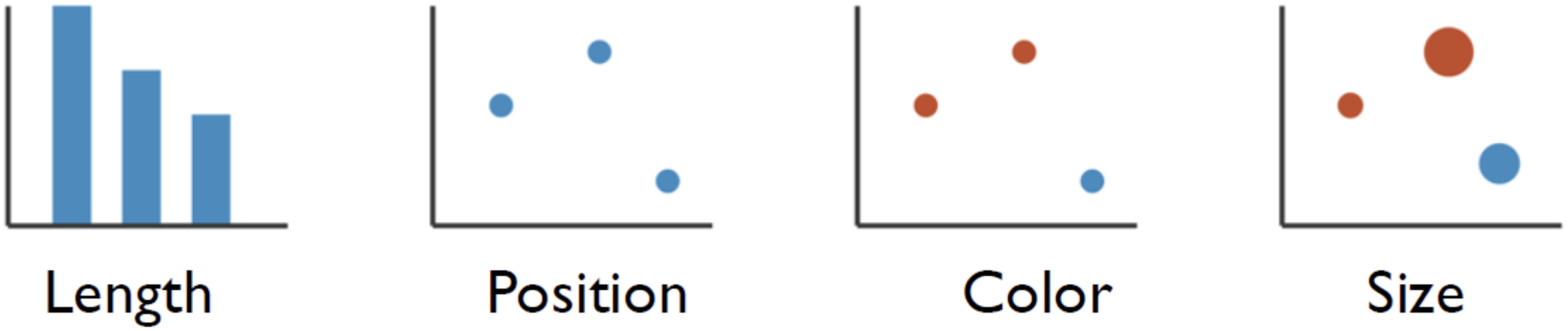
- **Color**

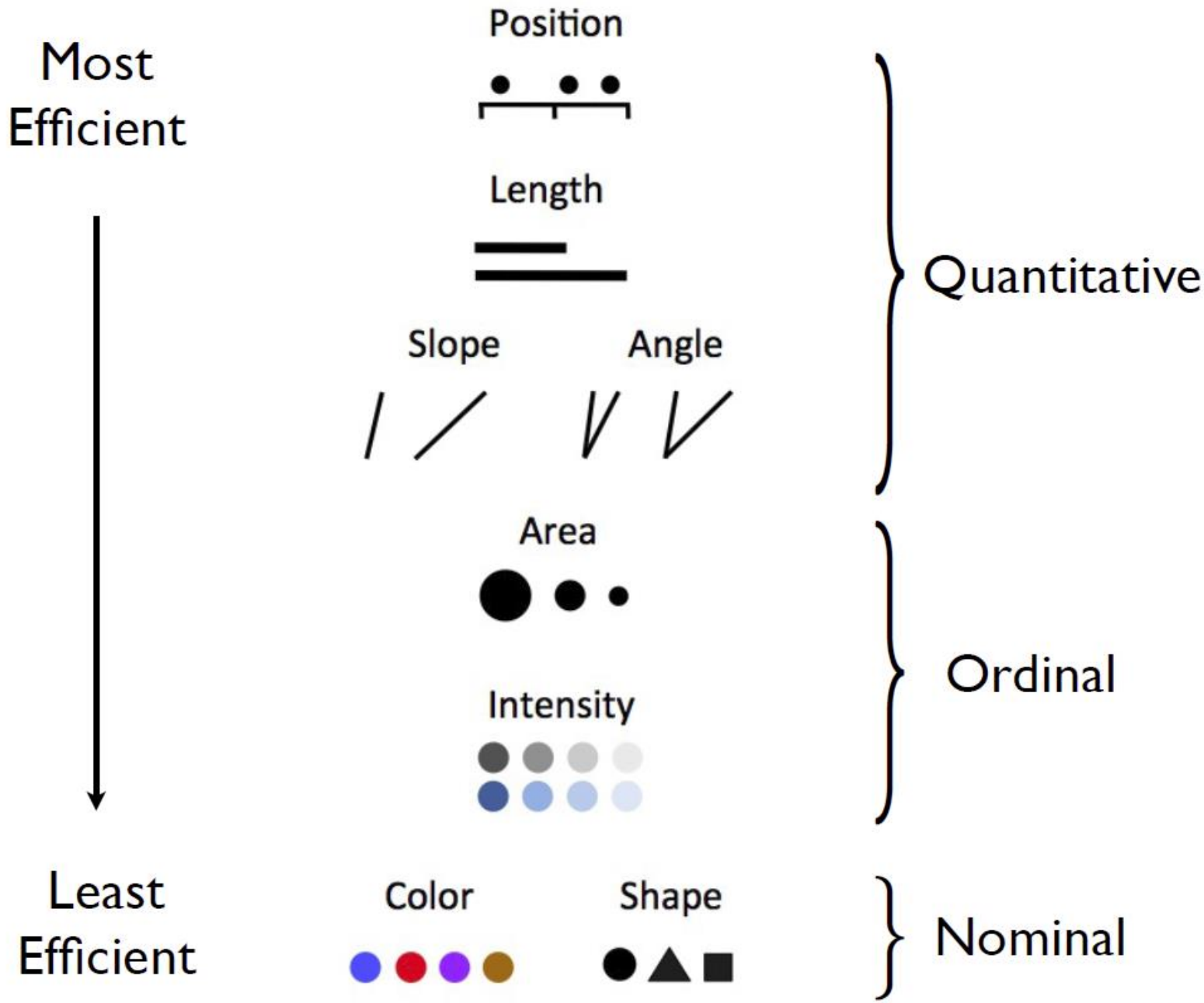
- **Shape**

- **Tilt**

- **Size**
  - Length
  - Area
  - Volume

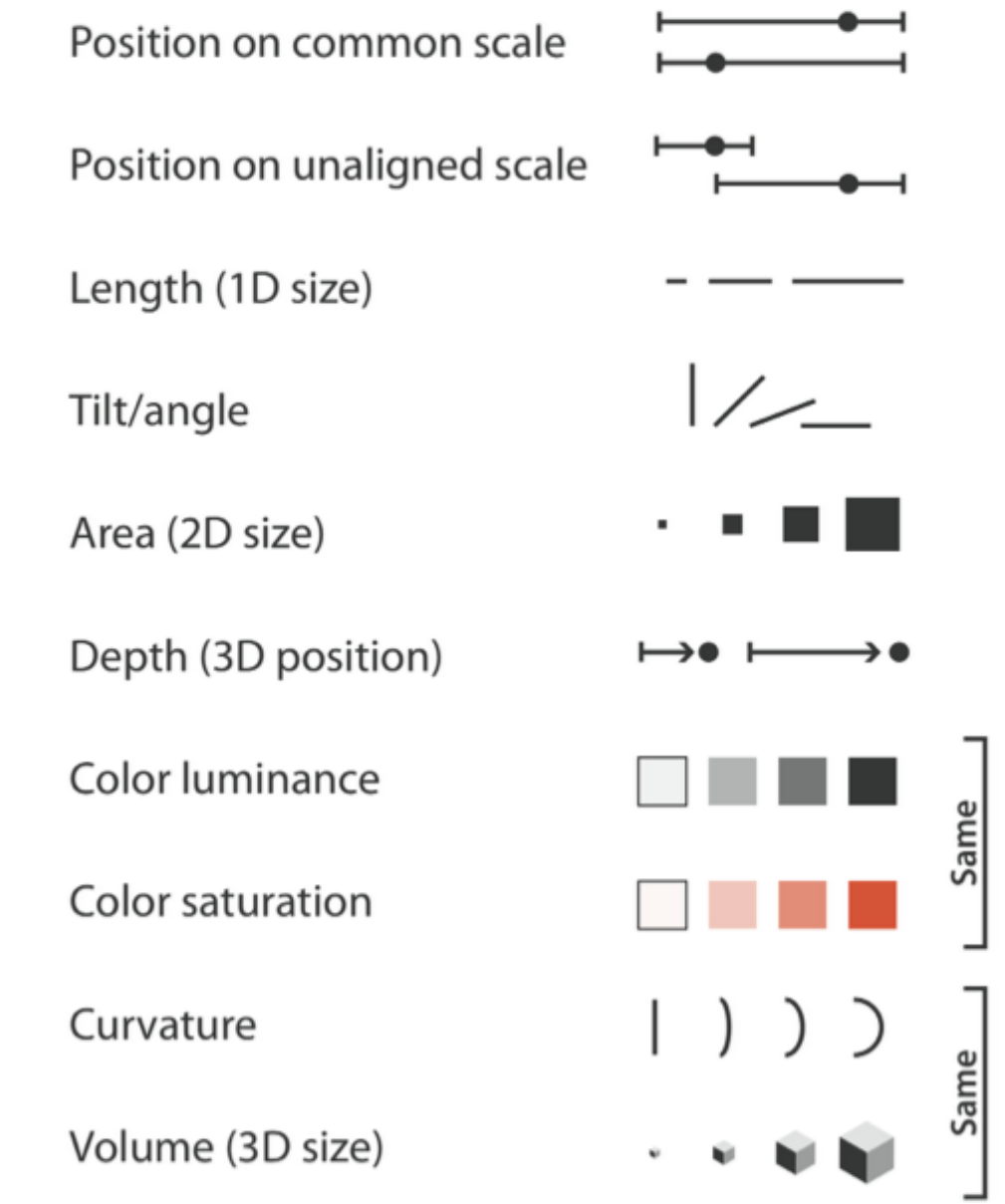
Using Marks and Attributes





C. Mulbrandon
VisualizingEconomics.com

Magnitude Channels: Ordered Attributes

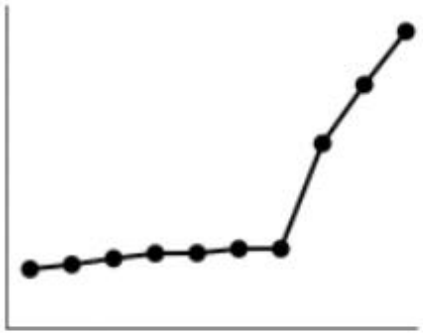


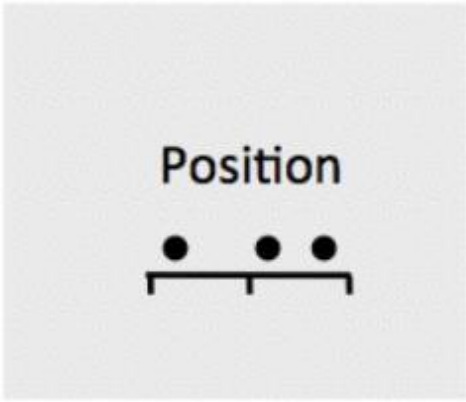
Identity Channels: Categorical Attributes

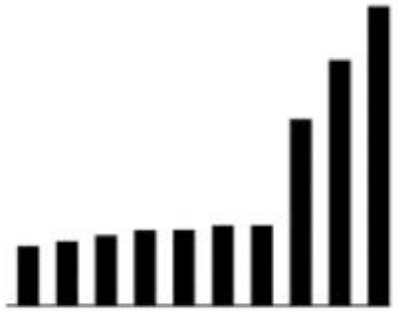
T. Munzner, Visualzation Analysis and Design, 2014

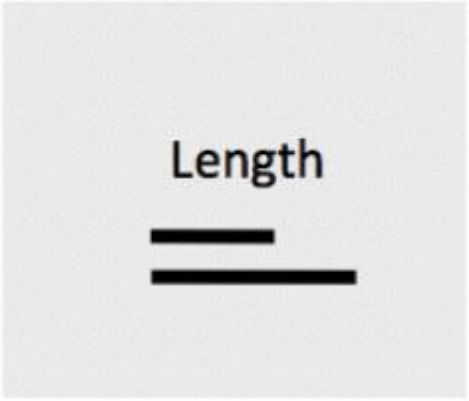
Most Effective

For Quantitative/Ordinal Data



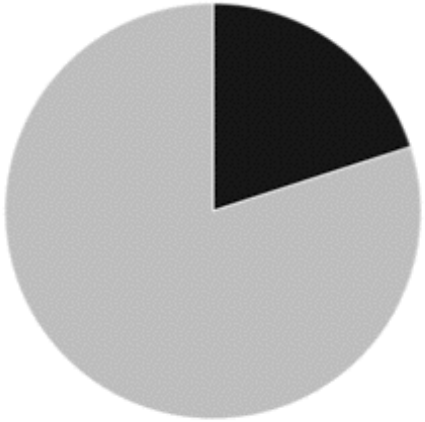


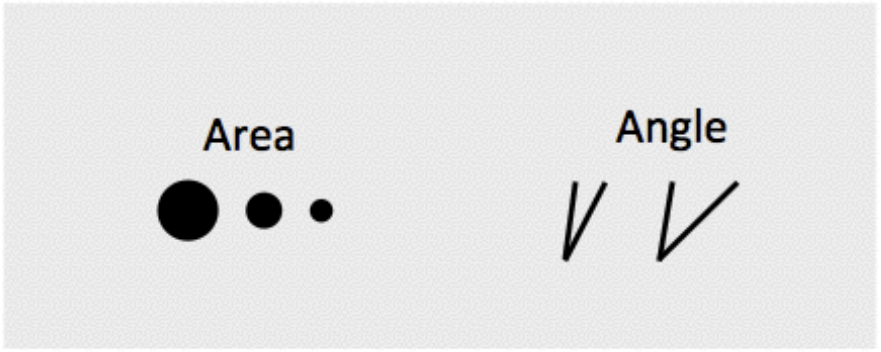


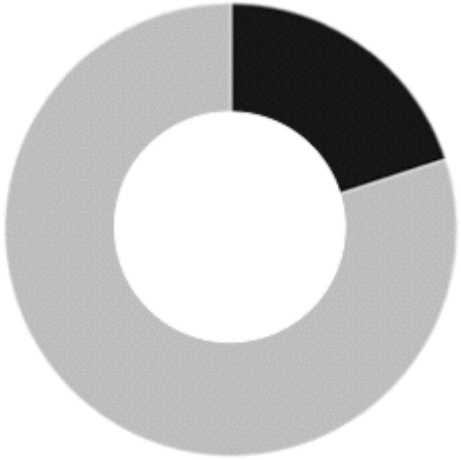


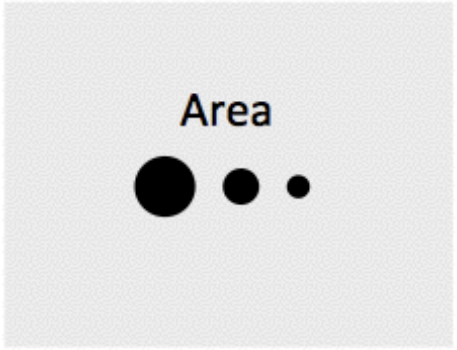
Less Effective

For Quantitative/Ordinal Data









Least Effective: Color

For Quantitative/Ordinal Data

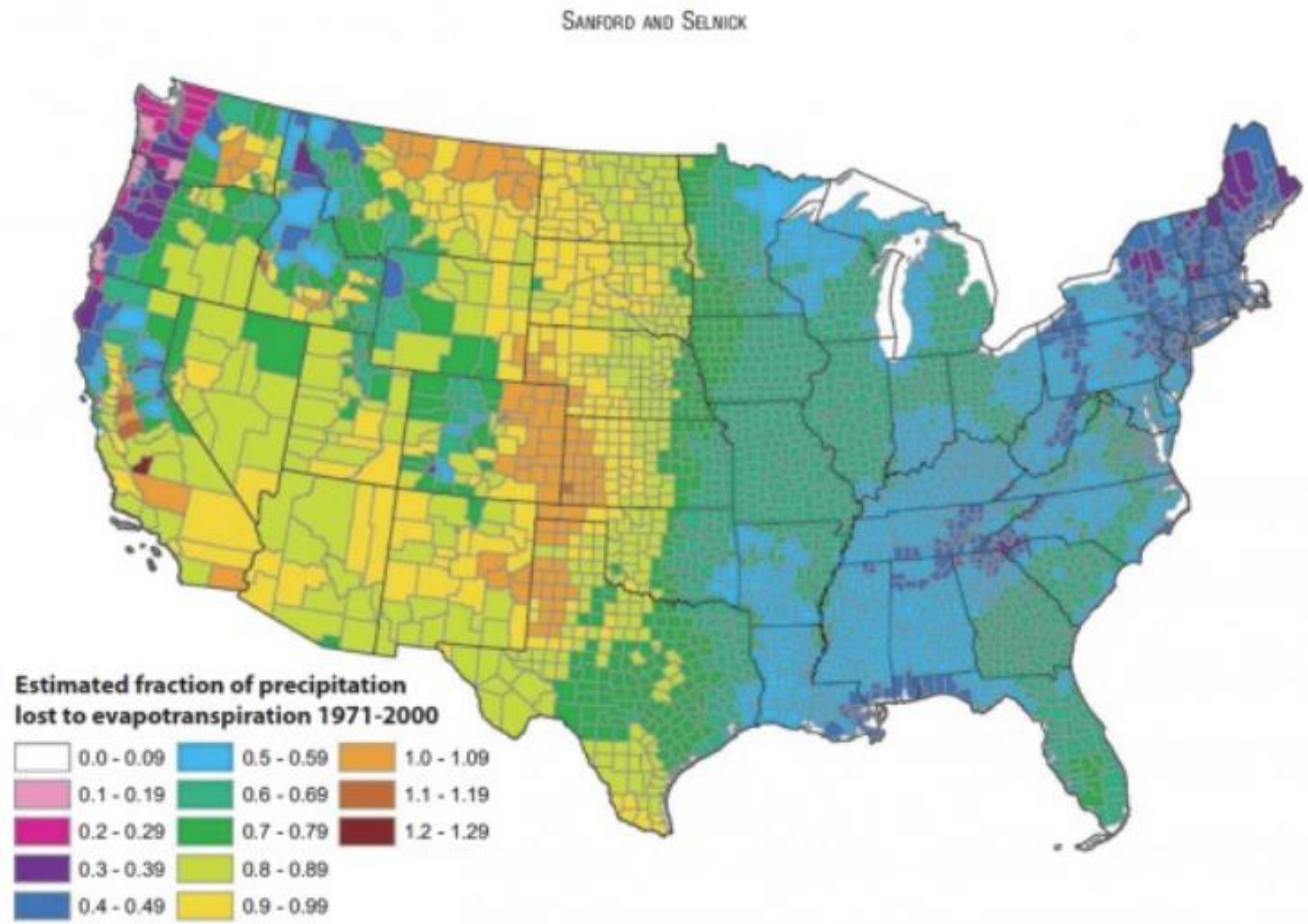


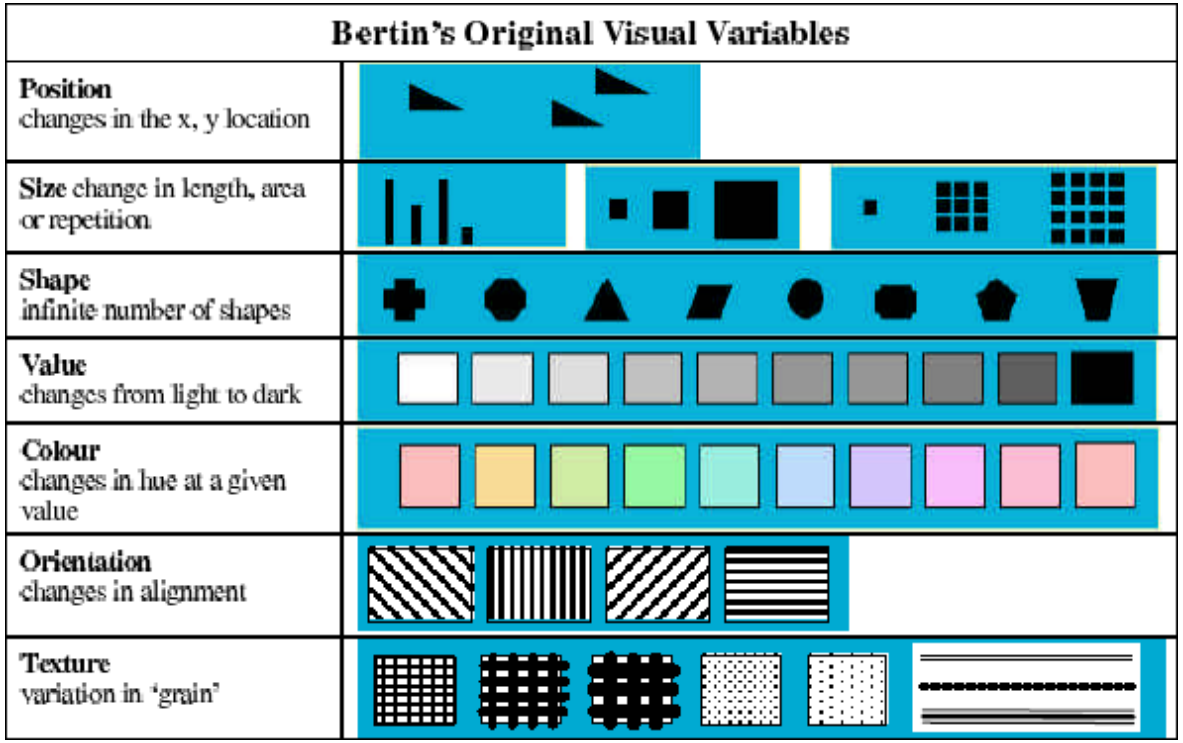
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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.

Bertin's Visual Attributes

Jacques Bertin

- French cartographer (1918 – 2010)
- Semiology of Graphics (1967)
- Theoretical principles for visual encodings

Bertin's Visual Variables



Bertin's Visual Attributes

Marks Points Lines Areas

Channels

Position

Size

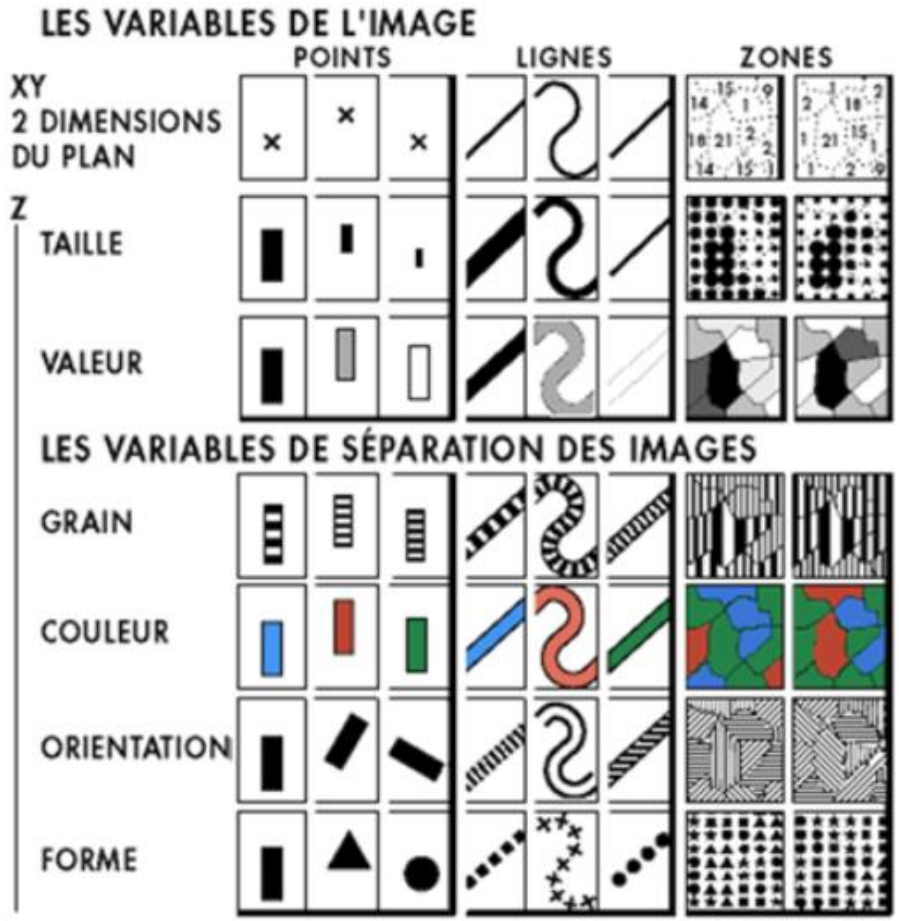
(Grey)Value

Texture

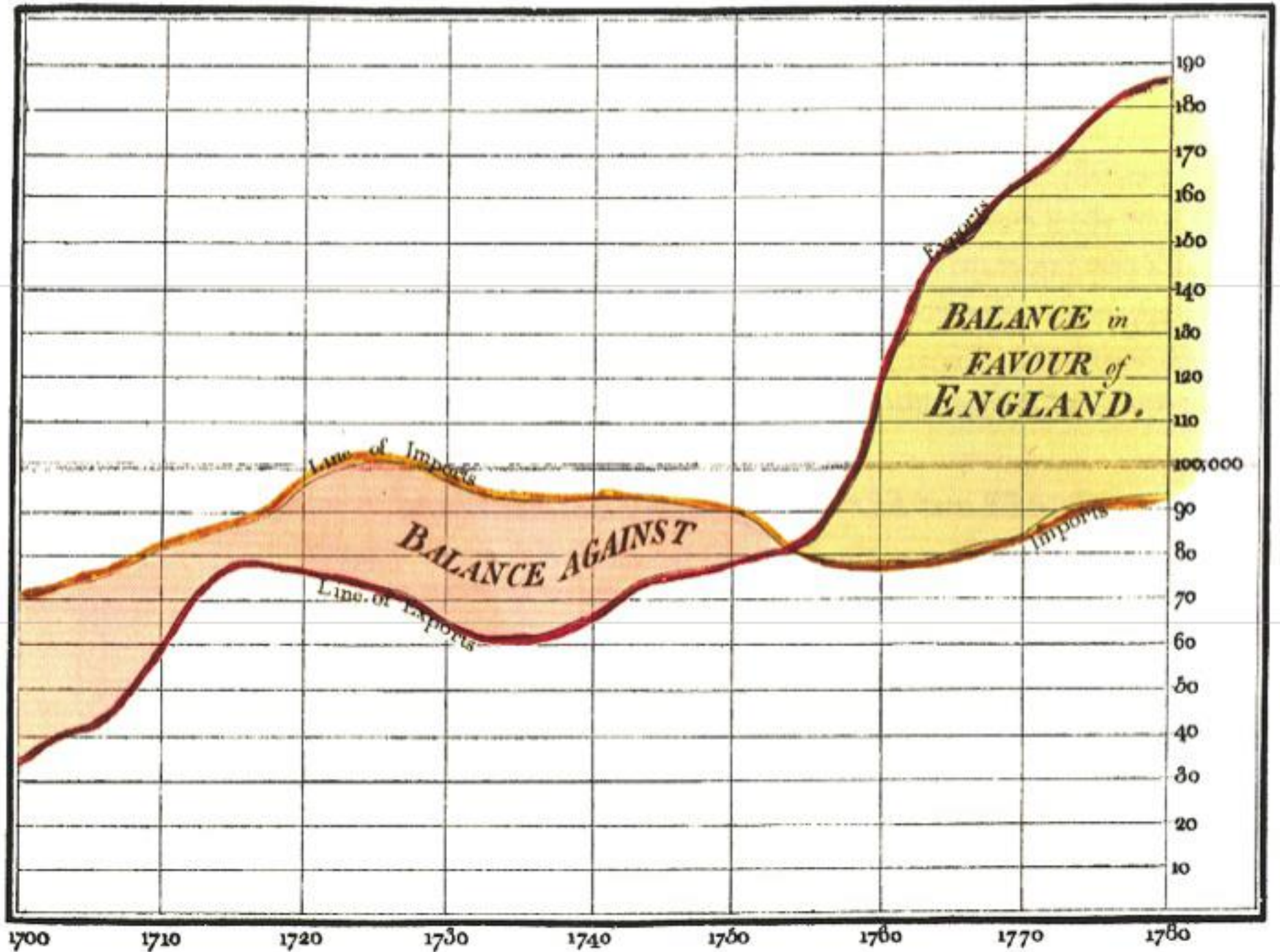
Color

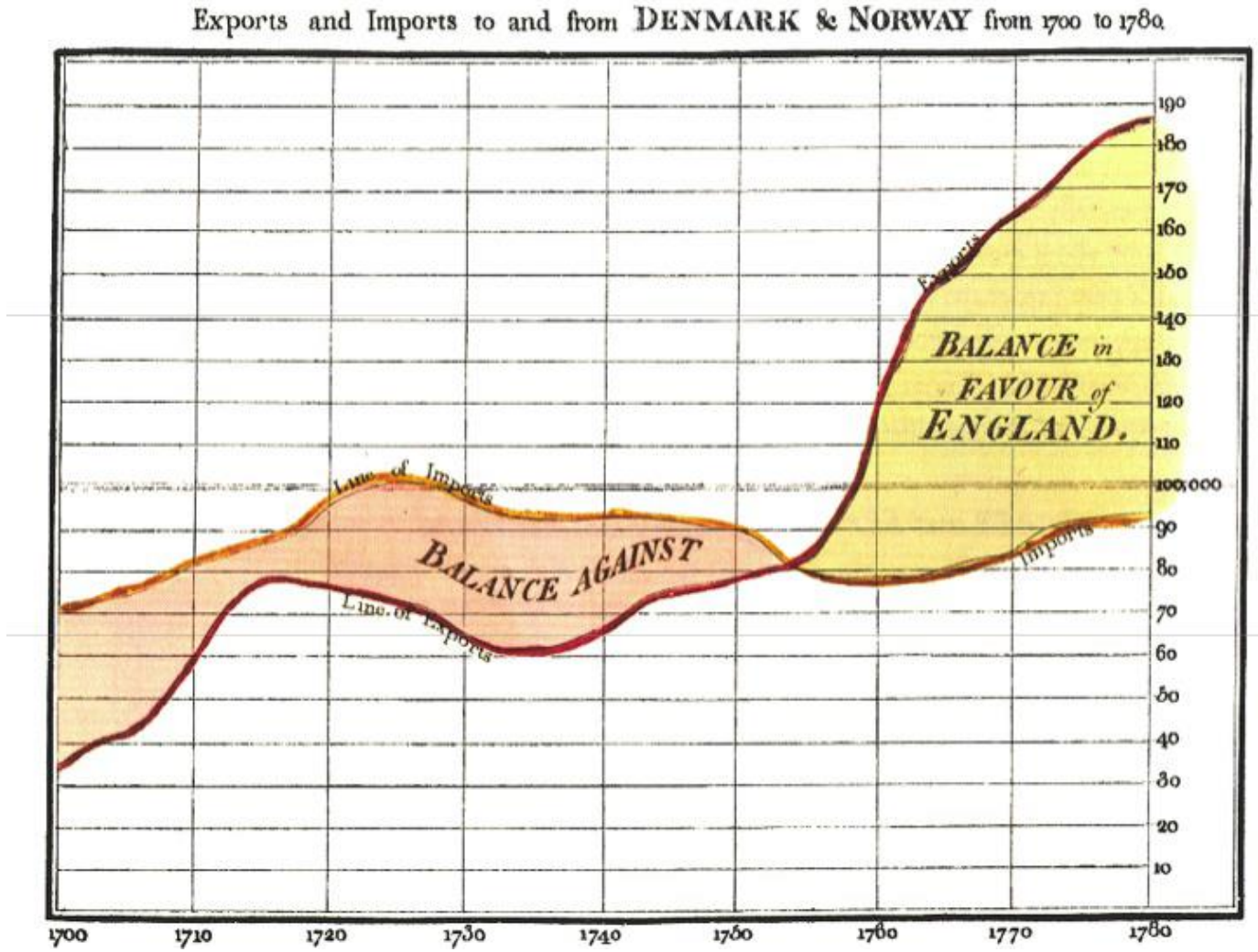
Orientation

Shape



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.





x-axis: Year (Q)uantity
y-axis: Currency (Q)uantity
Color: Imports / Exports (N)ominal
Color: Positive / Negative (O)rdered

Bertin's Visual Attributes (1967)

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

✔ = Good
~ = OK
✖ = Bad

Gestalt's Laws of Perceptual Organization

Gestalt's Laws of Perceptual Organization

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

Gestalt's Laws of Perceptual Organization

Law of Similarity:

Items that are similar tend to be grouped together.

In the image above, most people see vertical columns of circles and squares.

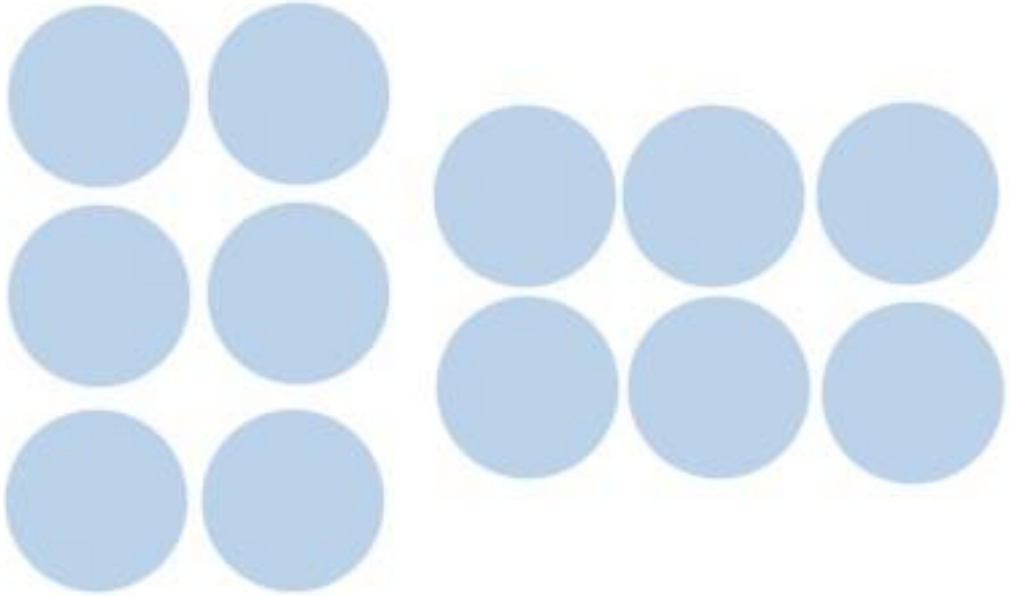
Gestalt's Laws of Perceptual Organization

Law of Pragnanz:

Reality is organized or reduced to the simplest form possible.

For example, we see the image above as a series of circles rather than as many much more complicated shapes.

Gestalt's Laws of Perceptual Organization

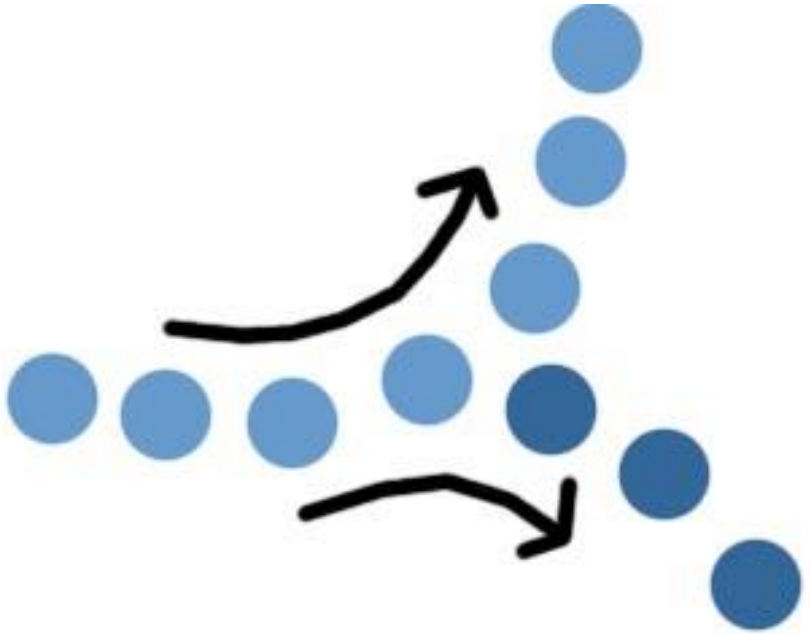


Law of Proximity:

Objects near each other tend to be grouped together.

The circles on the left appear to be grouped in vertical columns, while those on the right appear to be grouped in horizontal rows.

Gestalt's Laws of Perceptual Organization

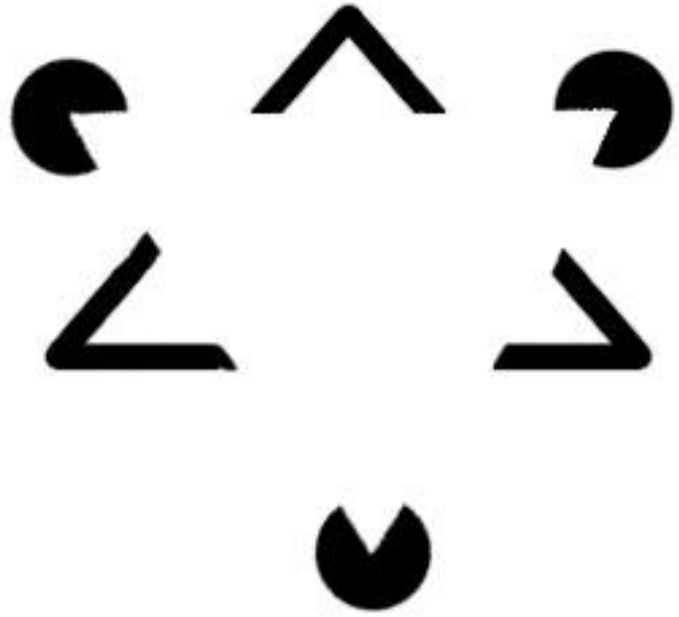


Law of Continuity:

Lines are seen as following the smoothest path.

In the image above, the top branch is seen as continuing the first segment of the line. This allows us to see things as flowing smoothly without breaking lines up into multiple parts.

Gestalt's Laws of Perceptual Organization



Law of Closure:

Objects grouped together are seen as a whole.

We tend to ignore gaps and complete contour lines. In the image above, there are no triangles or circles, but our minds fill in the missing information to create familiar shapes and images.

Graphical Perception

Graphical Perception

- Some visual processing takes place without any conscious effort on our part
- This is termed p*reattentive vision*
- Graphical perception derives from preattentive vision to extract information from graphs
- Graphs which convey their information at this unconscious level allow us to extract the information without any considered effort on our part

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?

Preattentive Attributes (Ranked)

1. Position on common scale
2. Position on non-aligned scale
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading

Journal of the American Statistical Association, September 1984

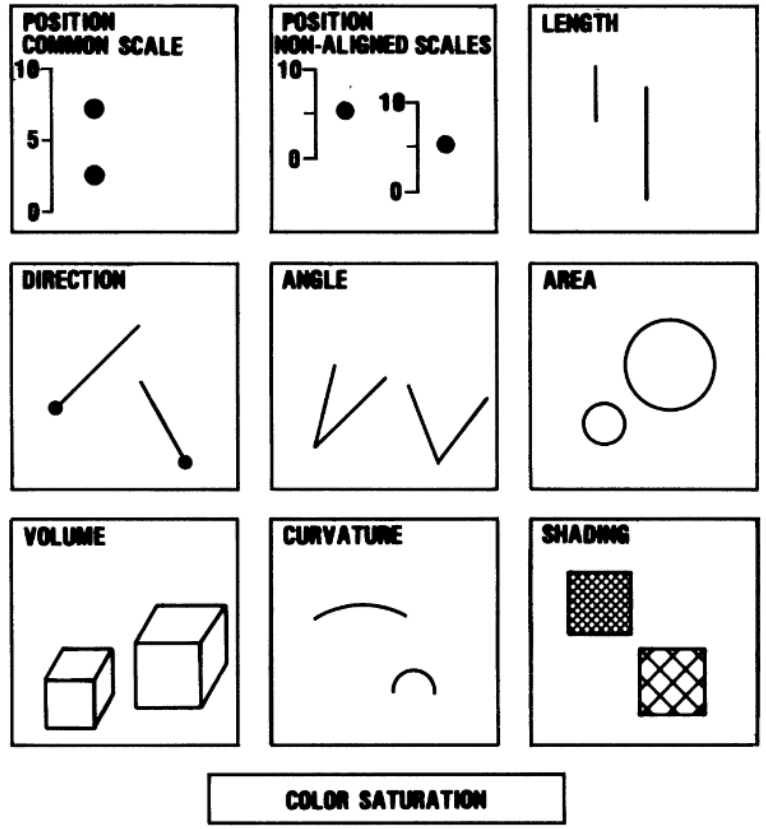
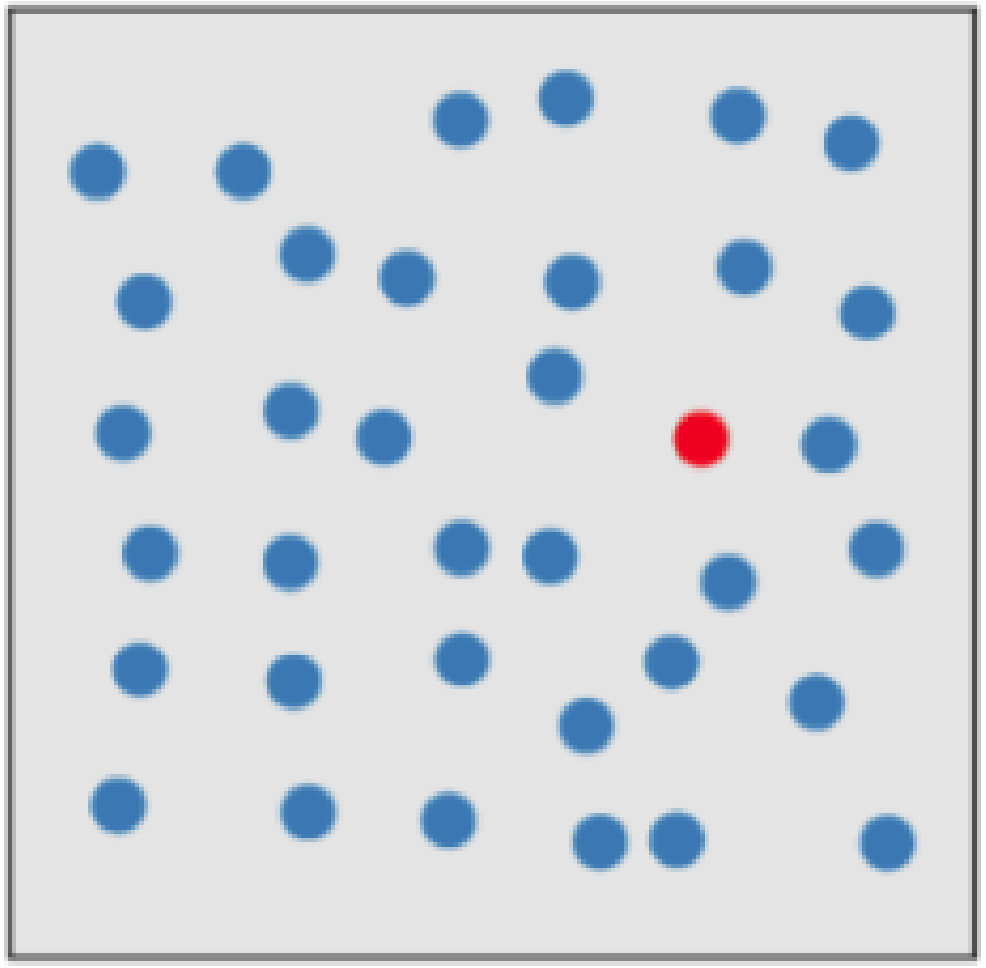
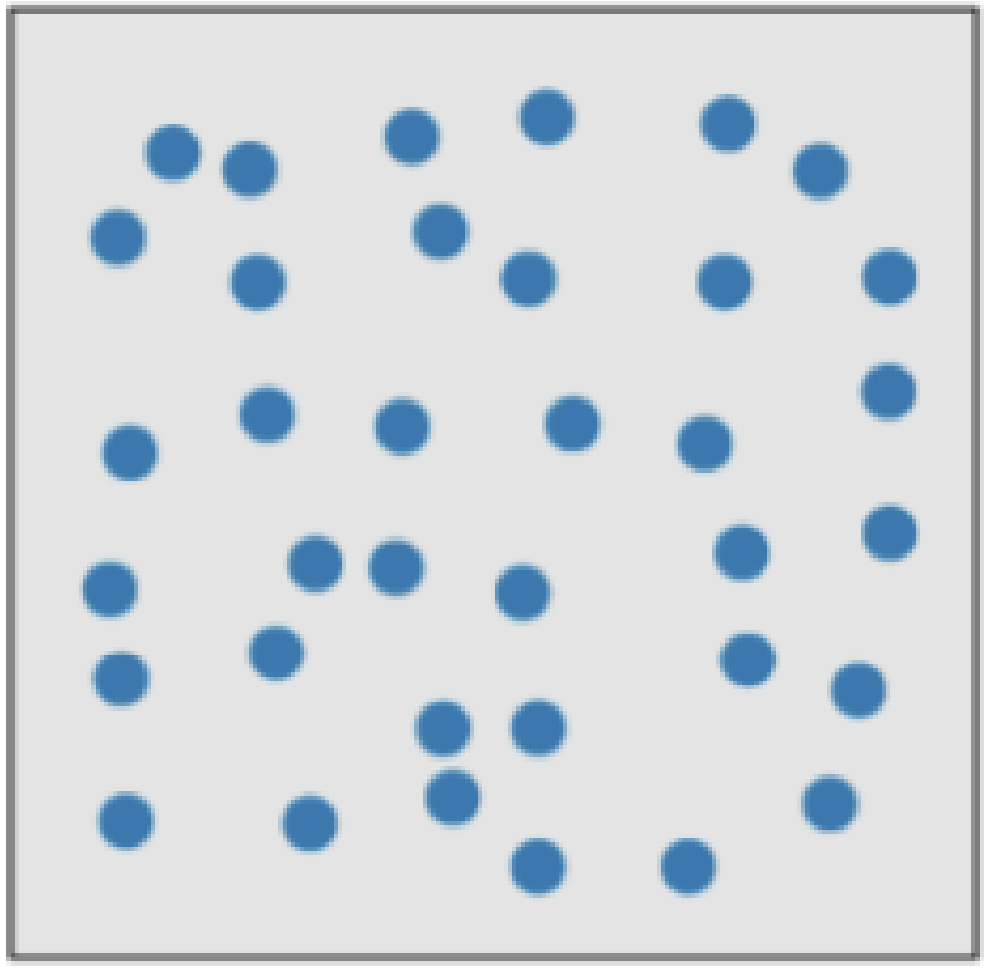


Figure 1. Elementary perceptual tasks.

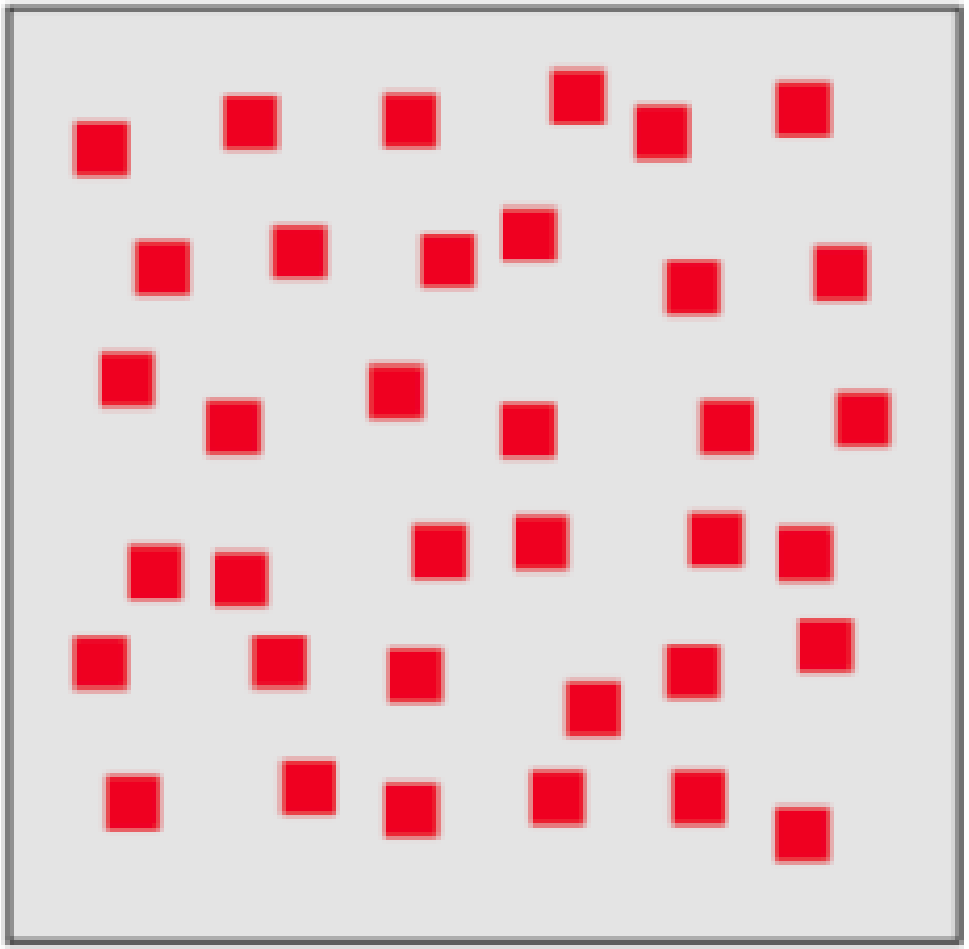
https://www.cs.ubc.ca/~tmm/courses/cpsc533c-04-spr/readings/cleveland.pdf

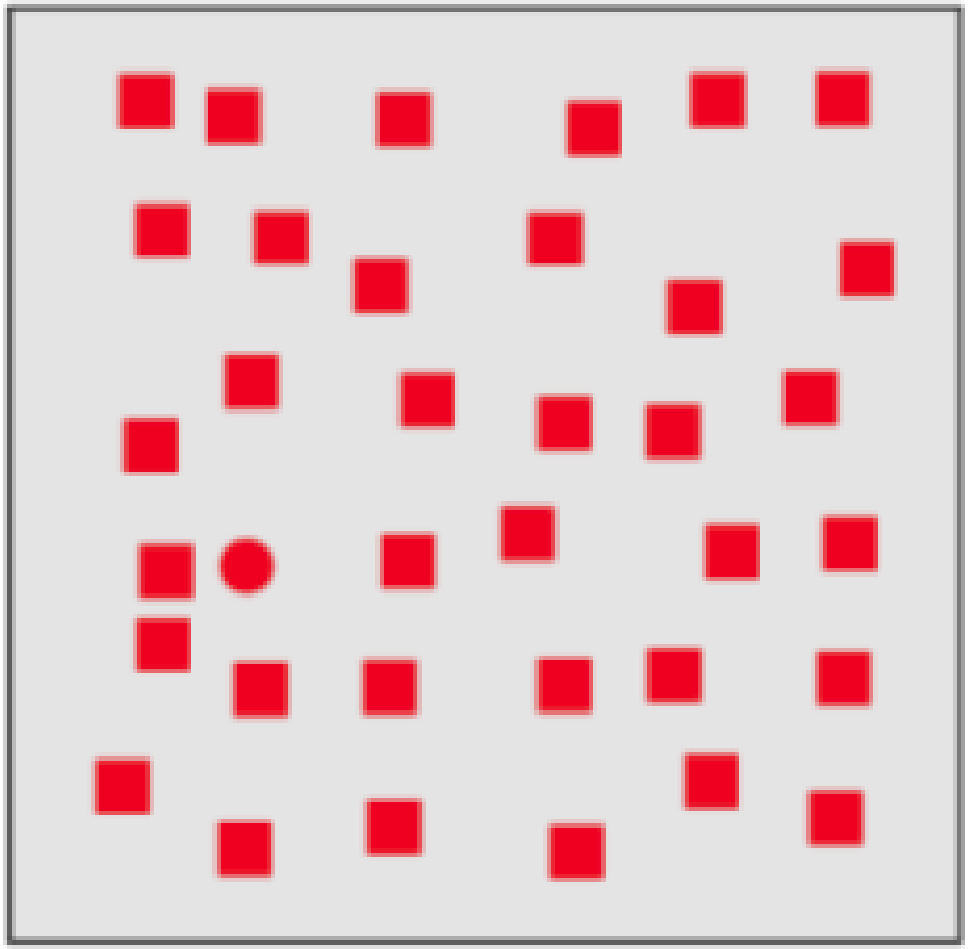
Preattentive Processing (Hue)



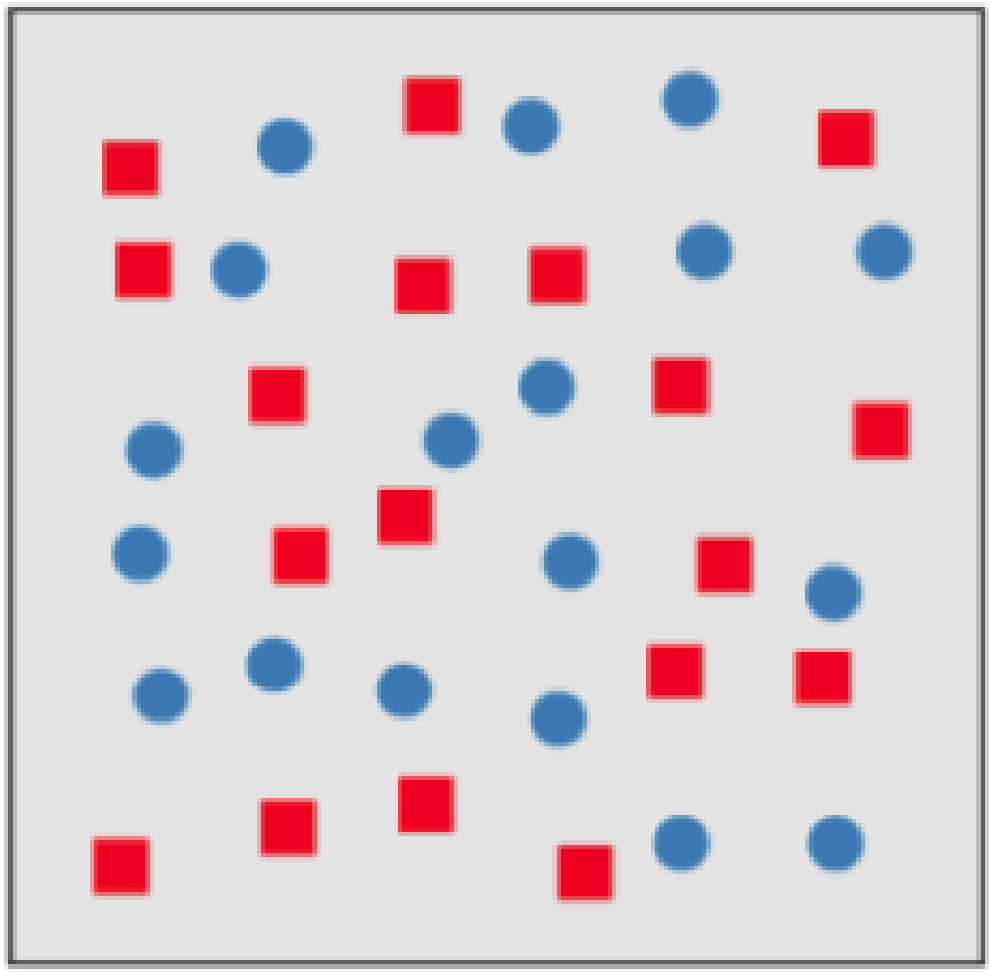


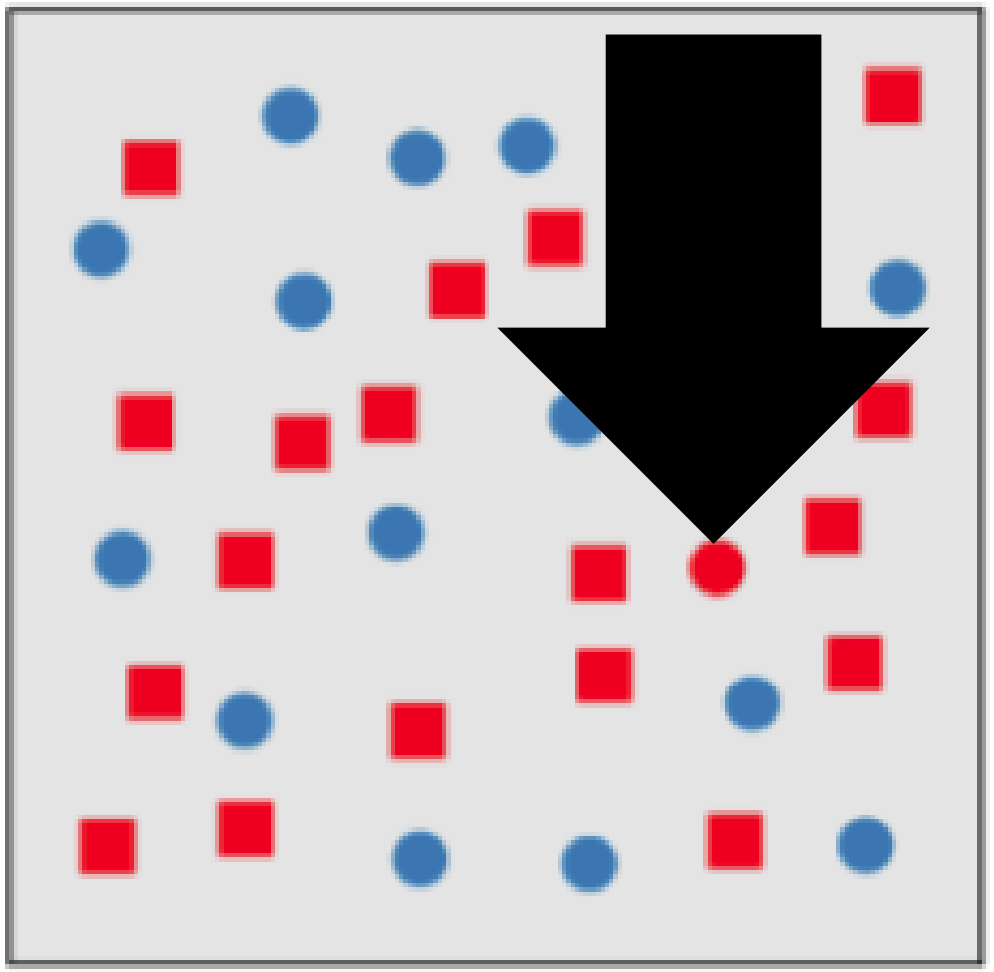
Preattentive Processing (Curvature)





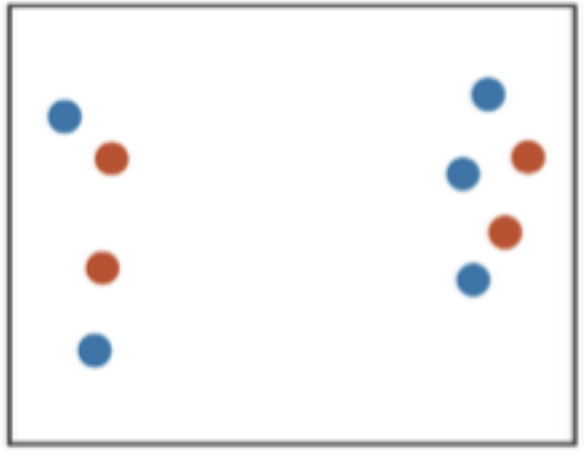
Preattentive Processing Limits (Conjunction of Attributes)





Separability of Attributes

Position
+ Hue (Color)



Fully separable

Size
+ Hue (Color)

Some interference

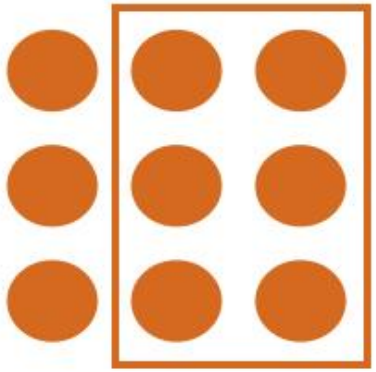
Width
+ Height

Some/significant interference

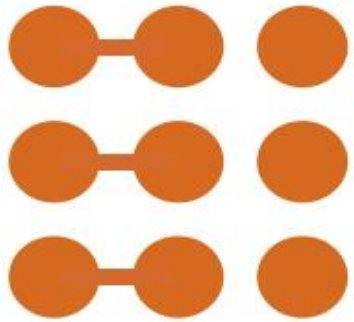
Red
+ Green

Major interference

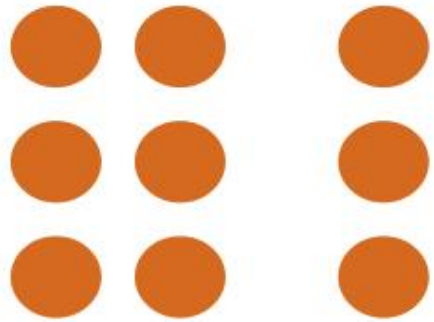
Grouping Principles



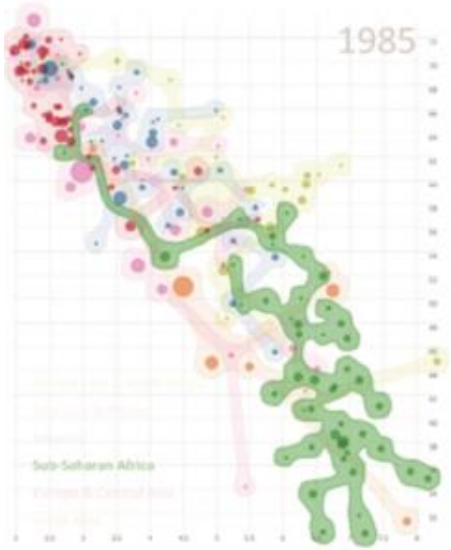
Containment



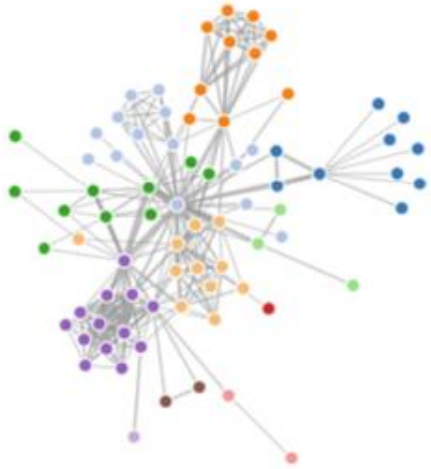
Connection



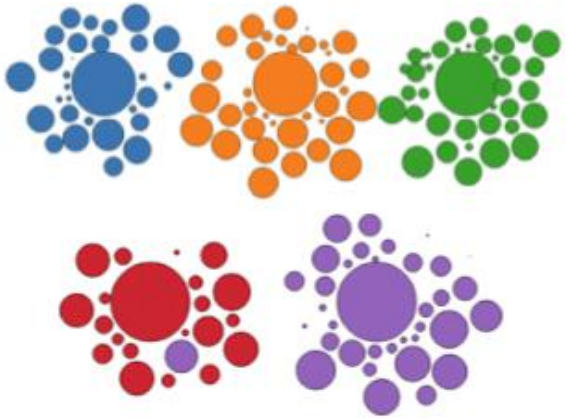
Proximity



Collins et al. 2009

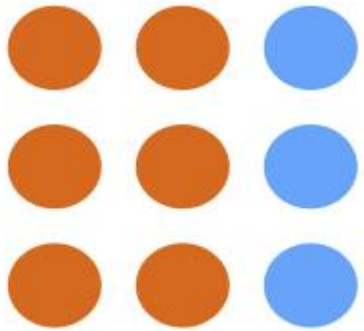


D3.js Example

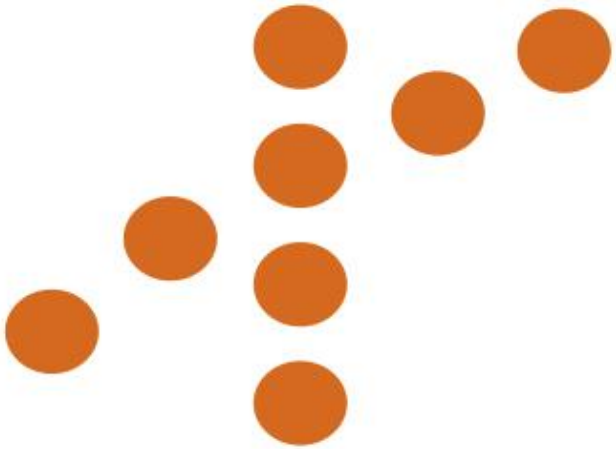


D3.js Example

Grouping Principles

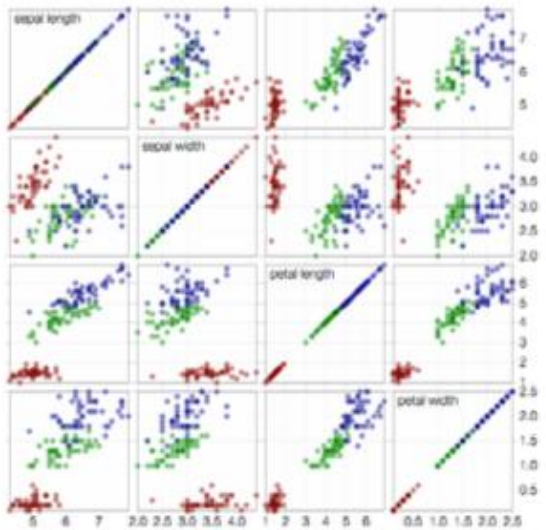


Similarity

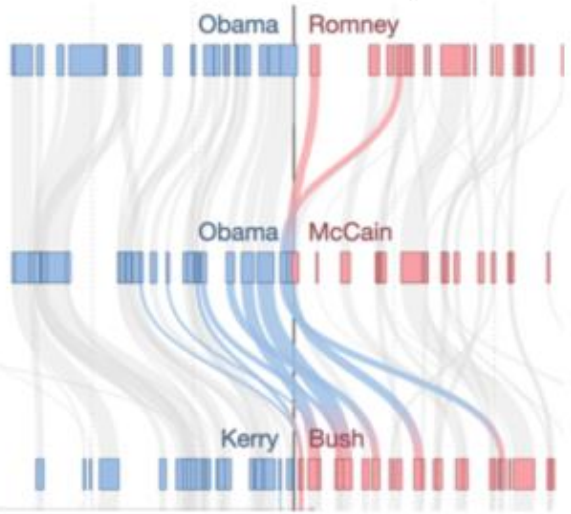


Continuity

Common Fate



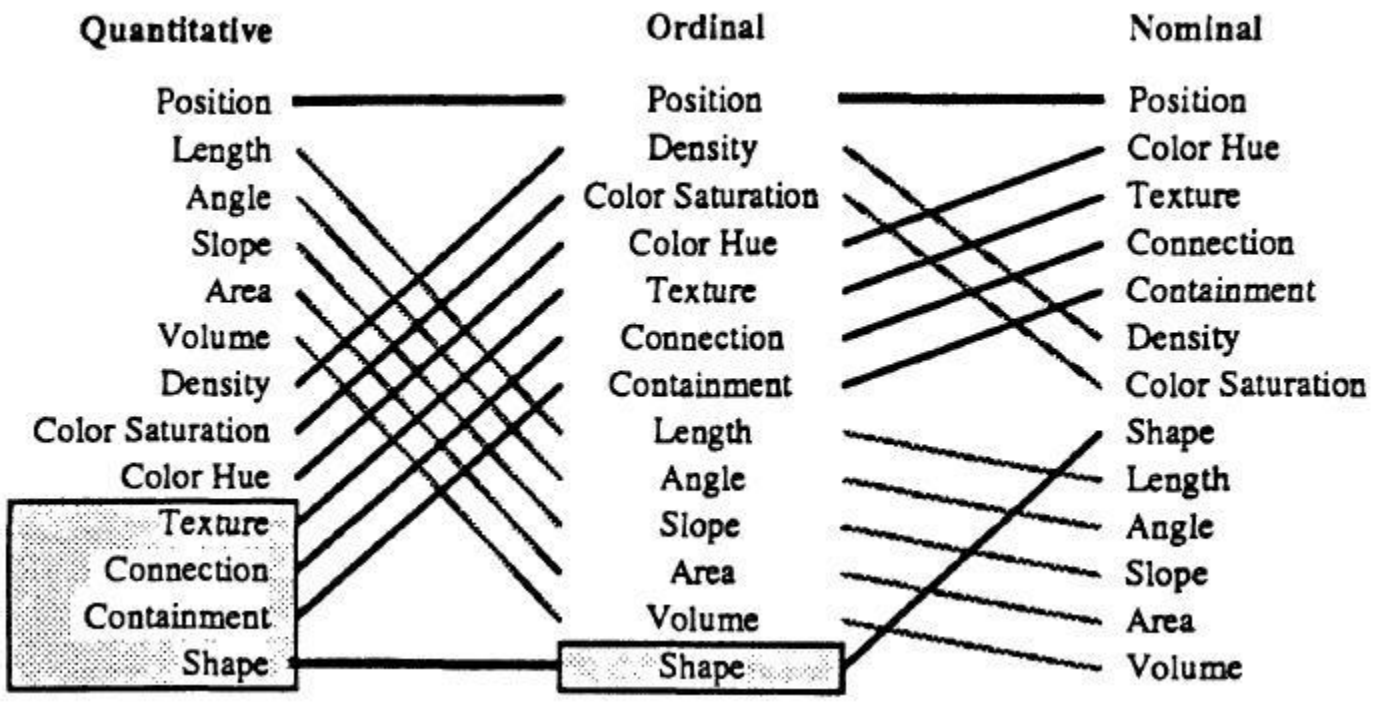
D3.js Example



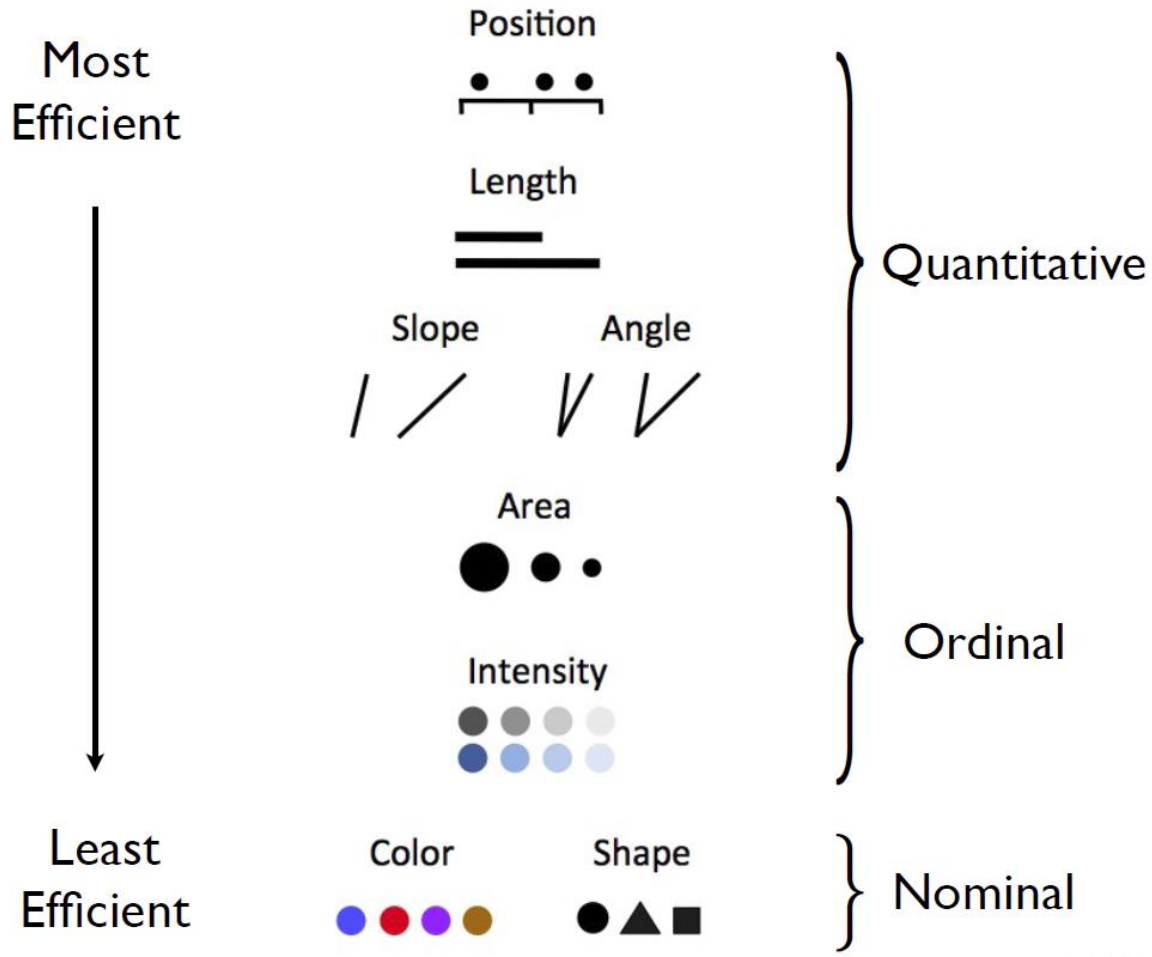
NYT Swing States

Closure

Perceptual Task (Mackinlay, 1986)



Perceptual Task (Mackinlay, 1986)

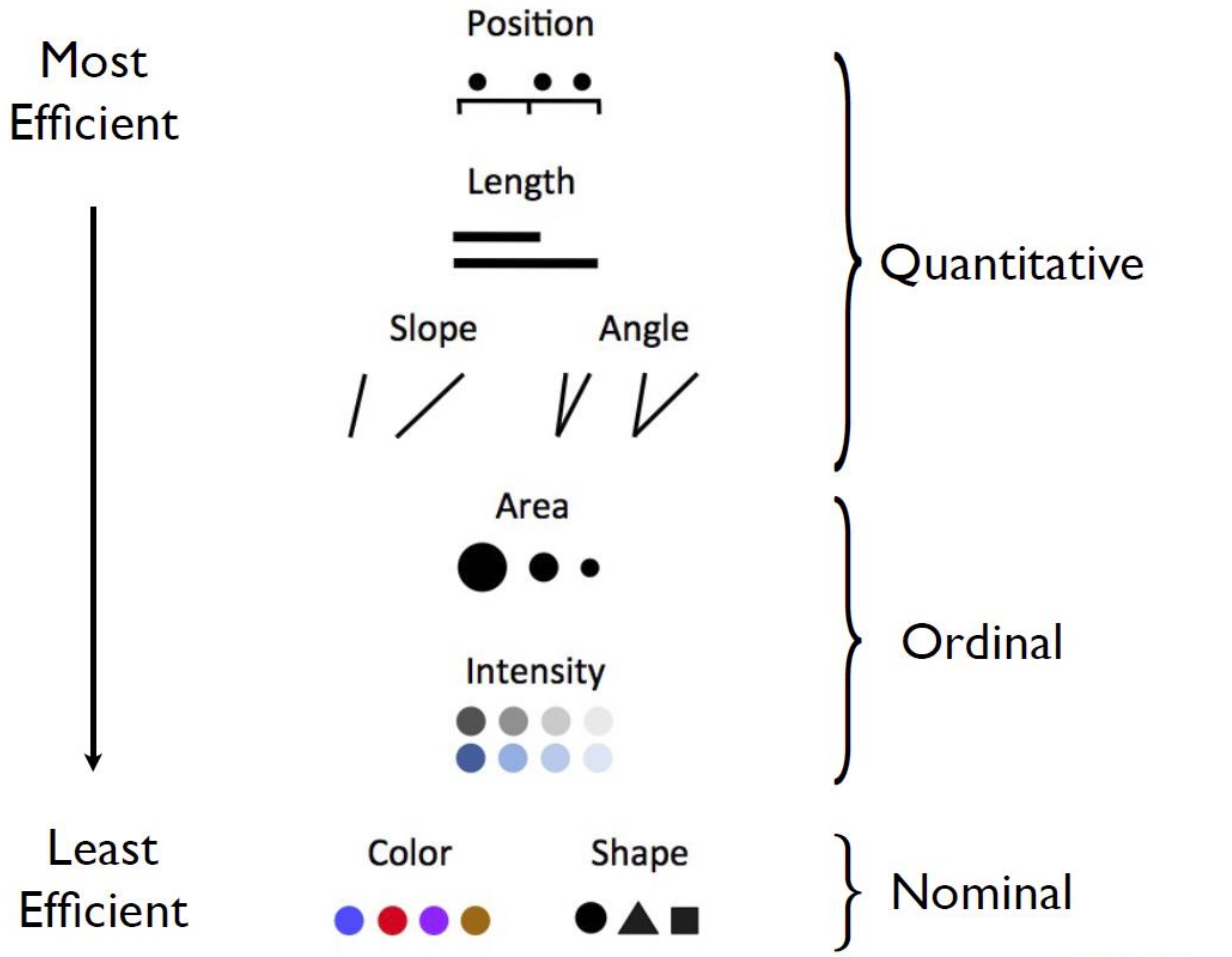


C. Mulbrandon
VisualizingEconomics.com

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



C. Mulbrandon
VisualizingEconomics.com

Expressiveness

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

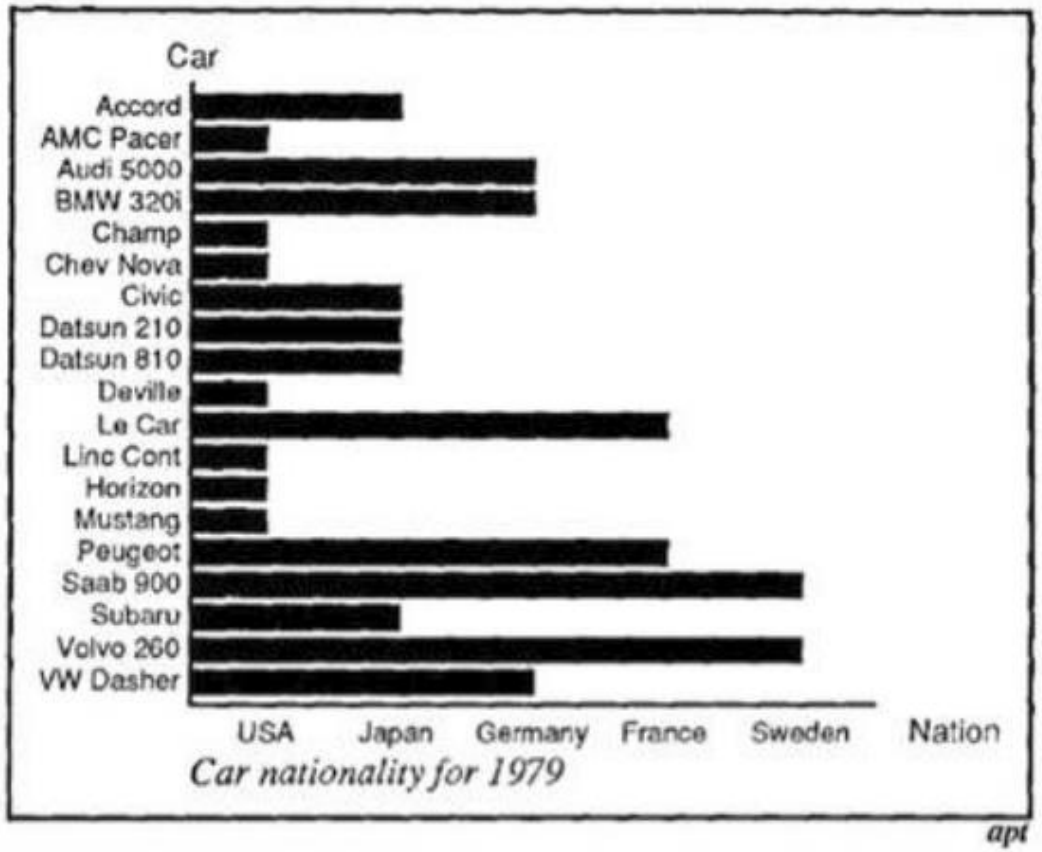
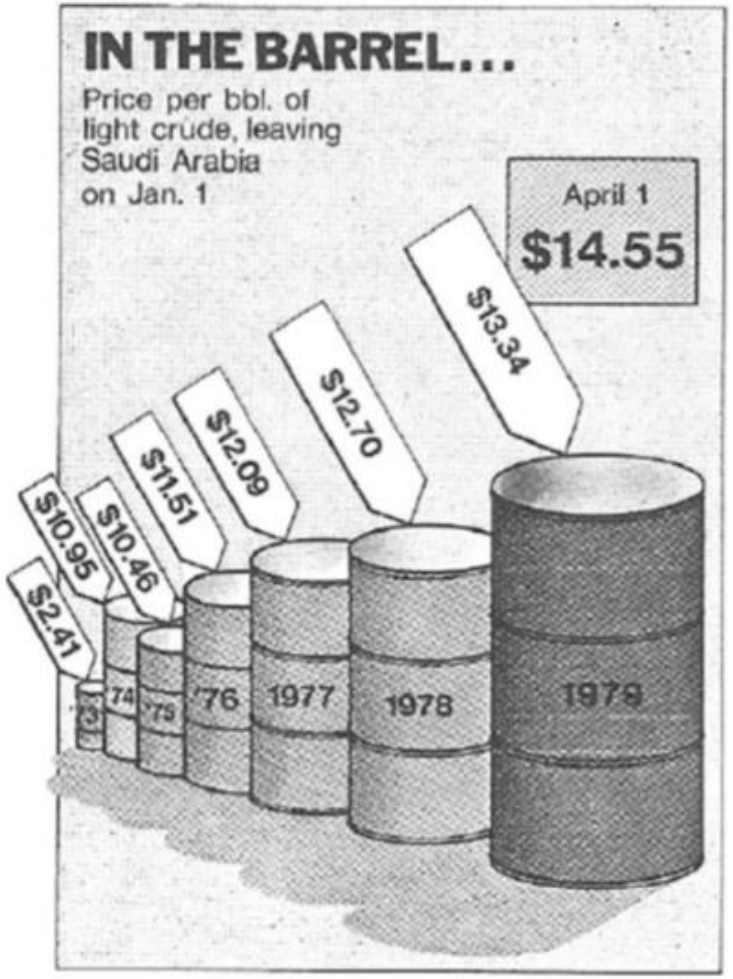


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

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!

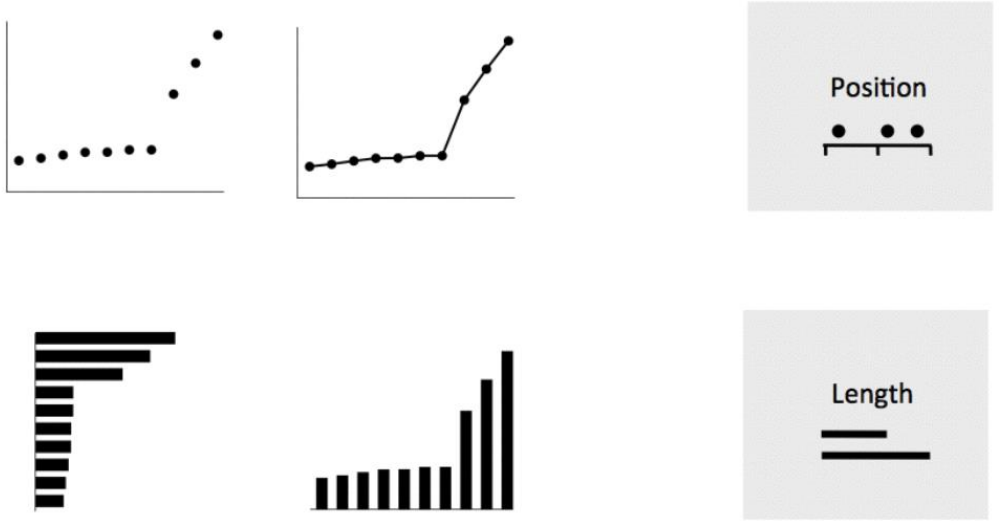


[Tufte, Edward R (1983), *The Visual Display of Quantitative Information,* Graphics Press, from *Time Magazine*, April 9, 1979, p. 57.]

Effectiveness of Attributes

Most Effective

For Quantitative/Ordinal Data



VisualizingEconomics.com

Less Effective

For Quantitative/Ordinal Data

VisualizingEconomics.com

Least Effective: Color

For Quantitative/Ordinal Data

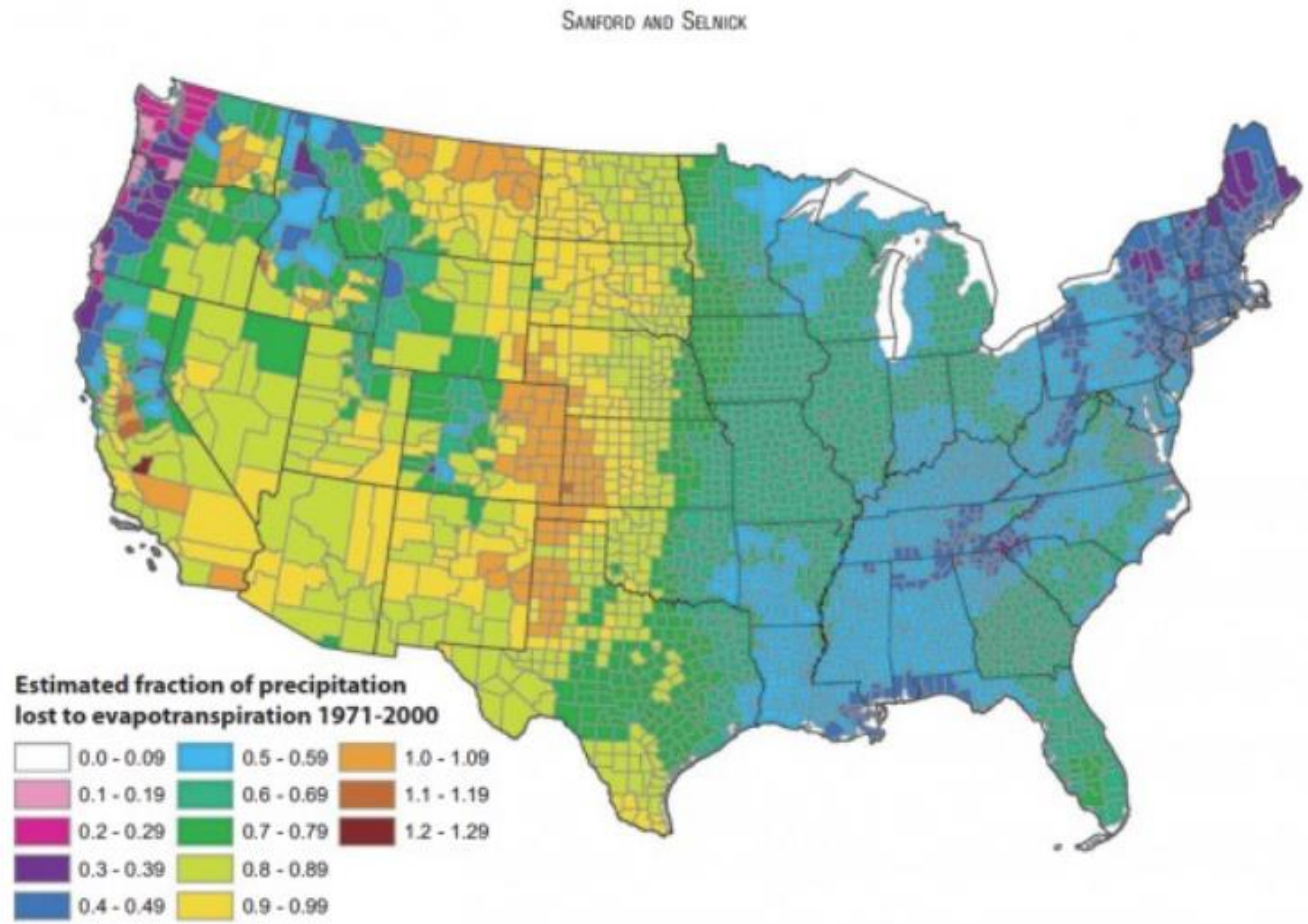
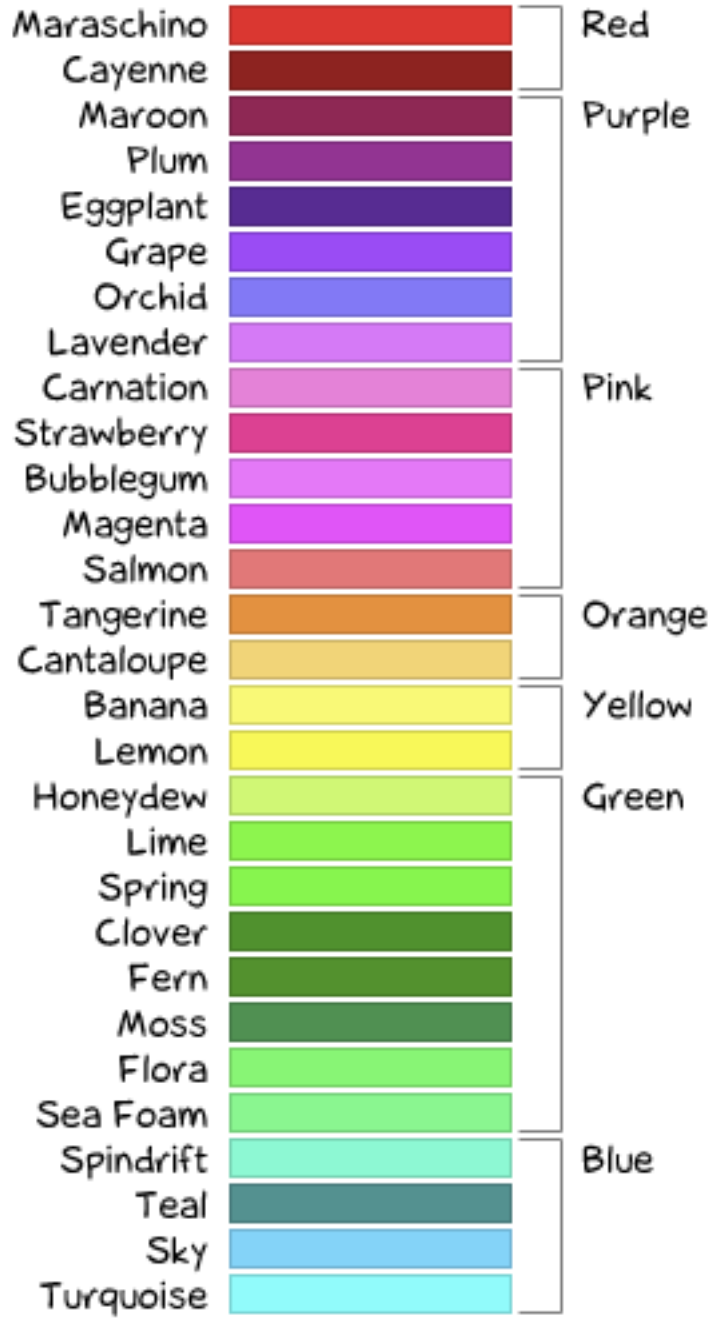


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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.

Color

Color names if you're a girl...

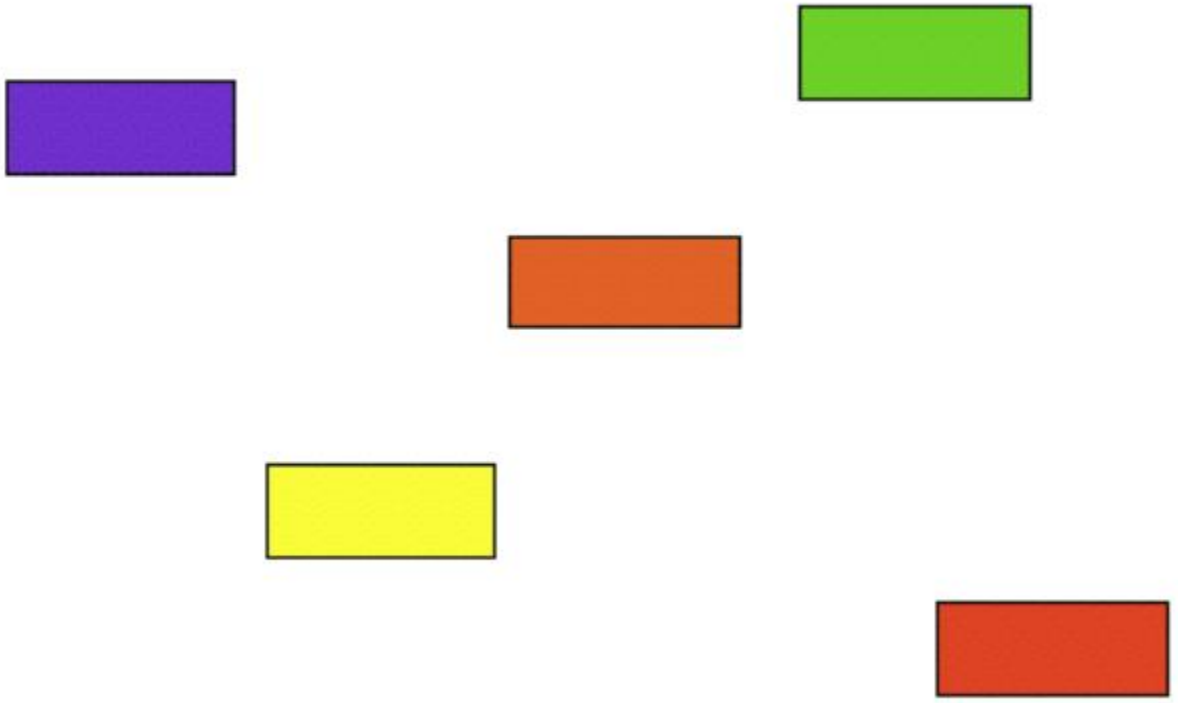


Color names if you're a guy...

Doghouse Diaries
"We take no as an answer."

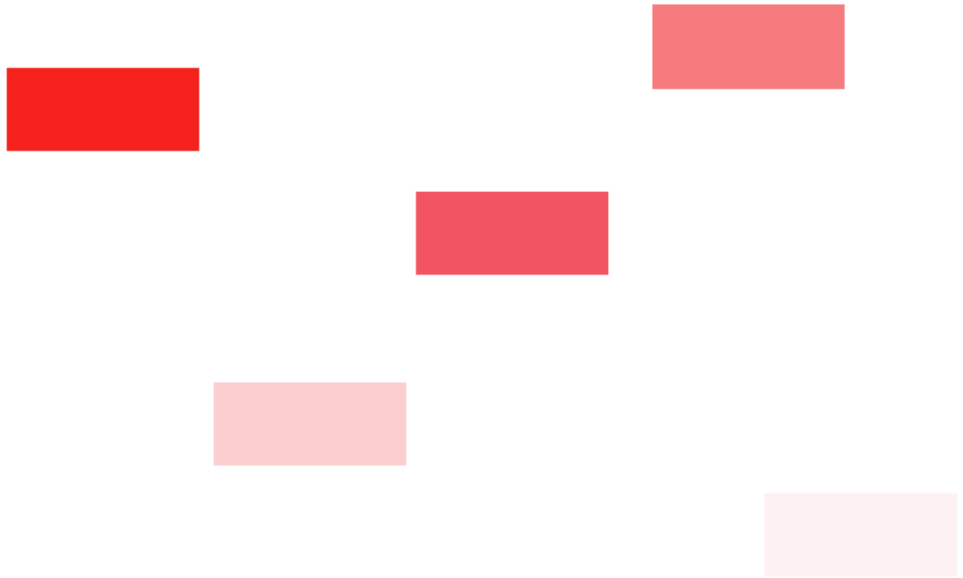
http://blog.xkcd.com/2010/05/03/color-survey-results/

Order These Colors



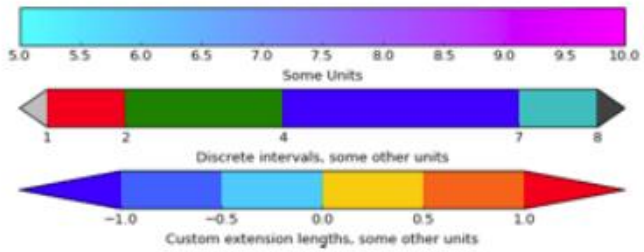
Order These Colors

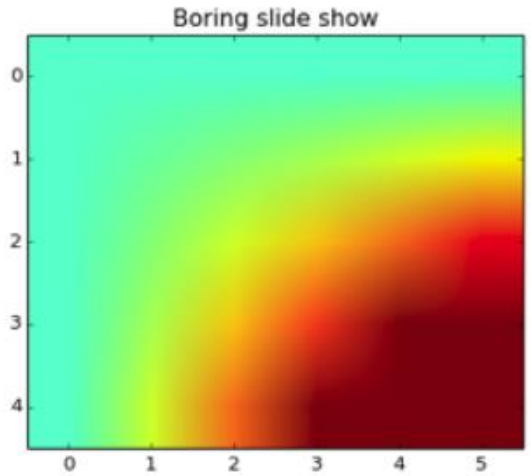
Order These Colors

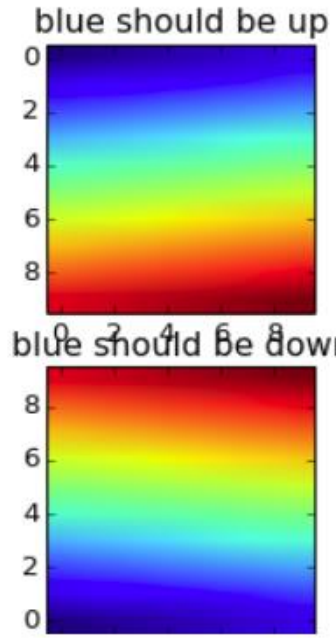


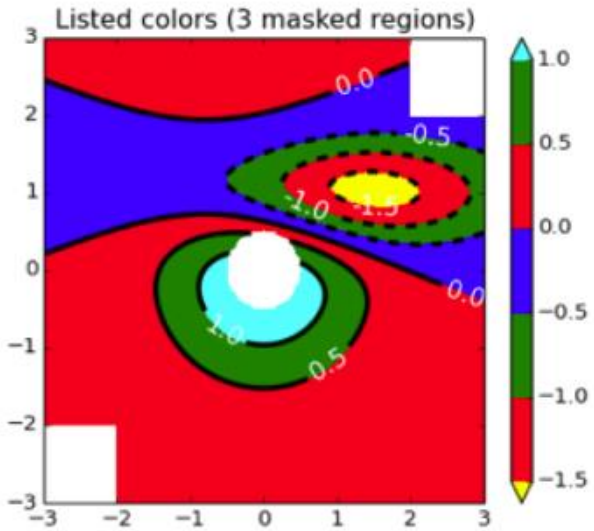

Brightness
Saturation
Hue
Perceived as Ordered
Not as much

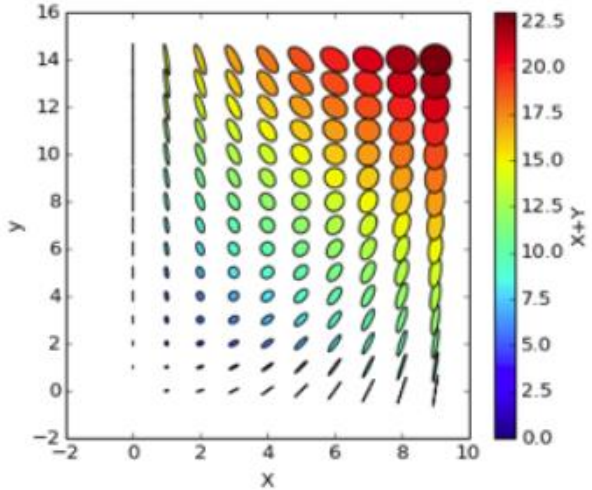
Rainbow Colors

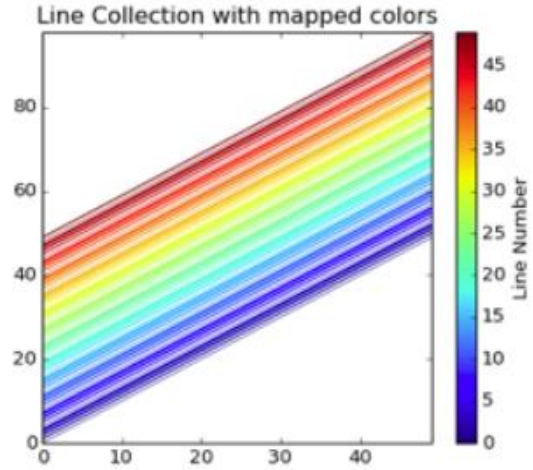












matplotlib gallery

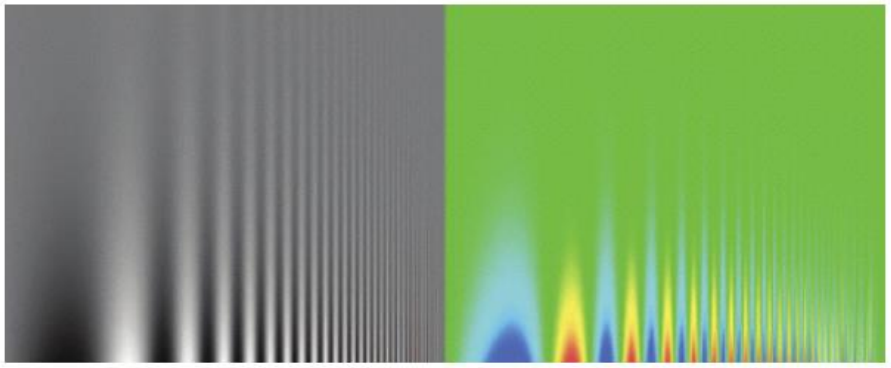
Rainbow Colormap

hard to order

easy to order

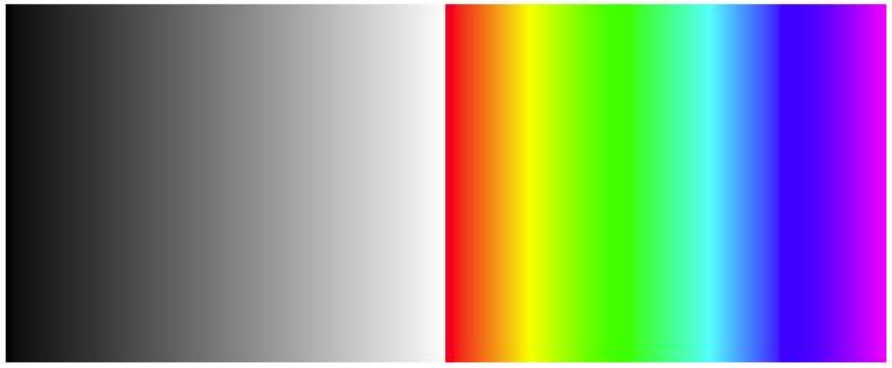
lower resolution

contrast

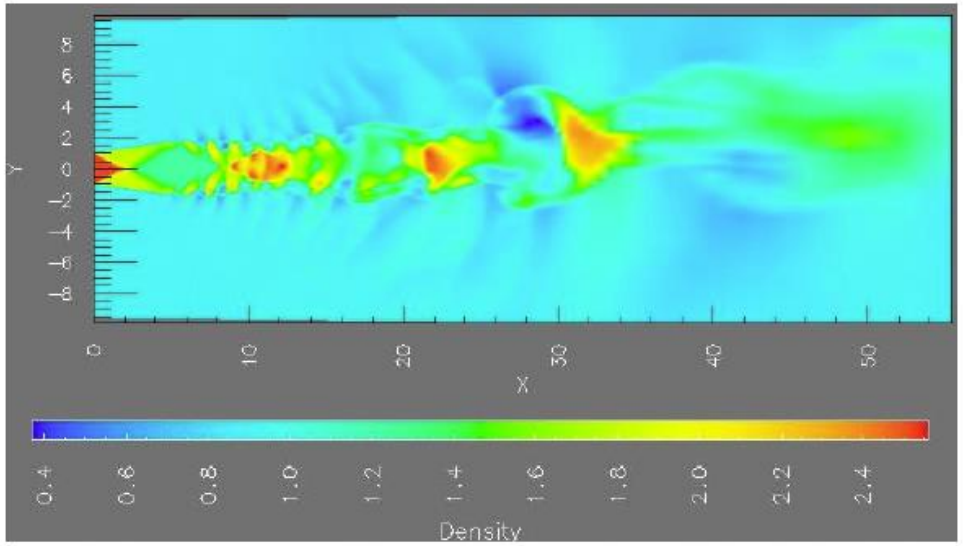


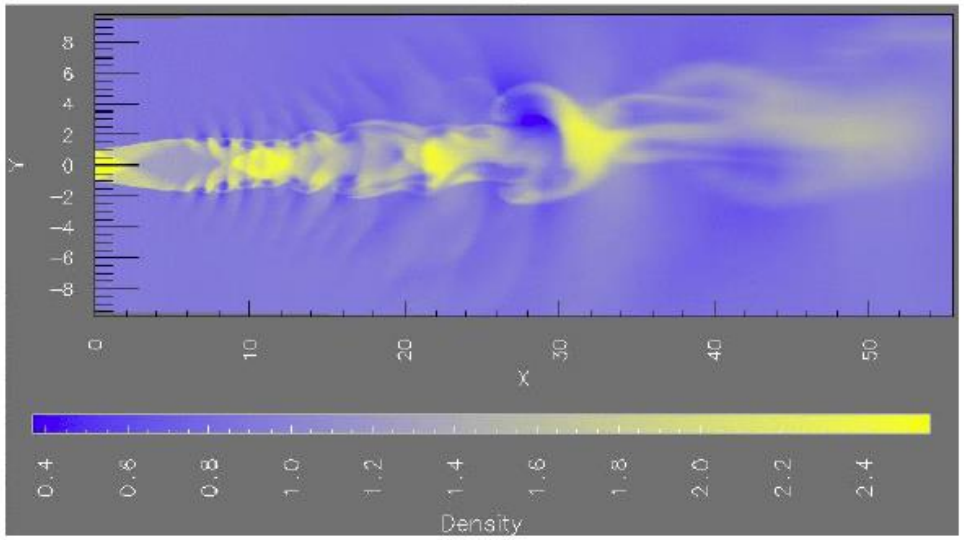
spatial frequency

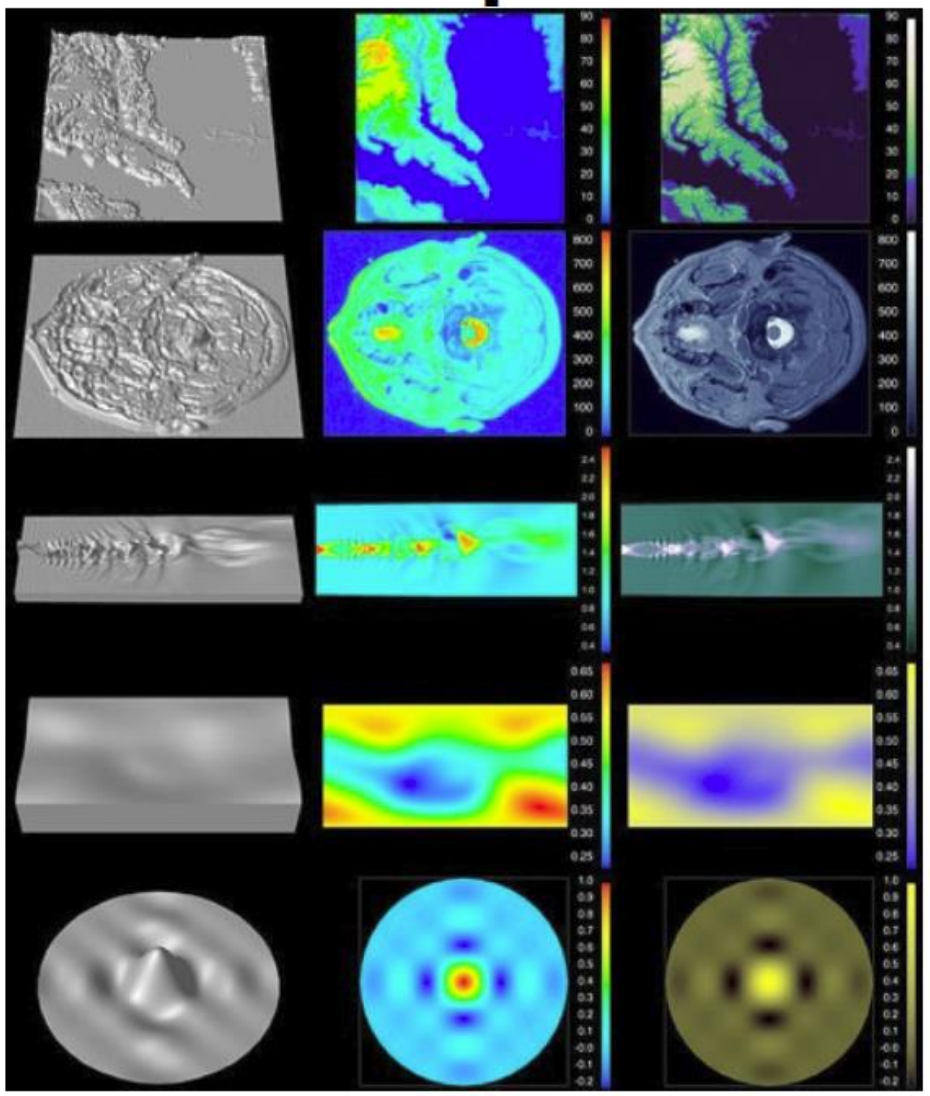
creates artifacts



Rainbow Colormap





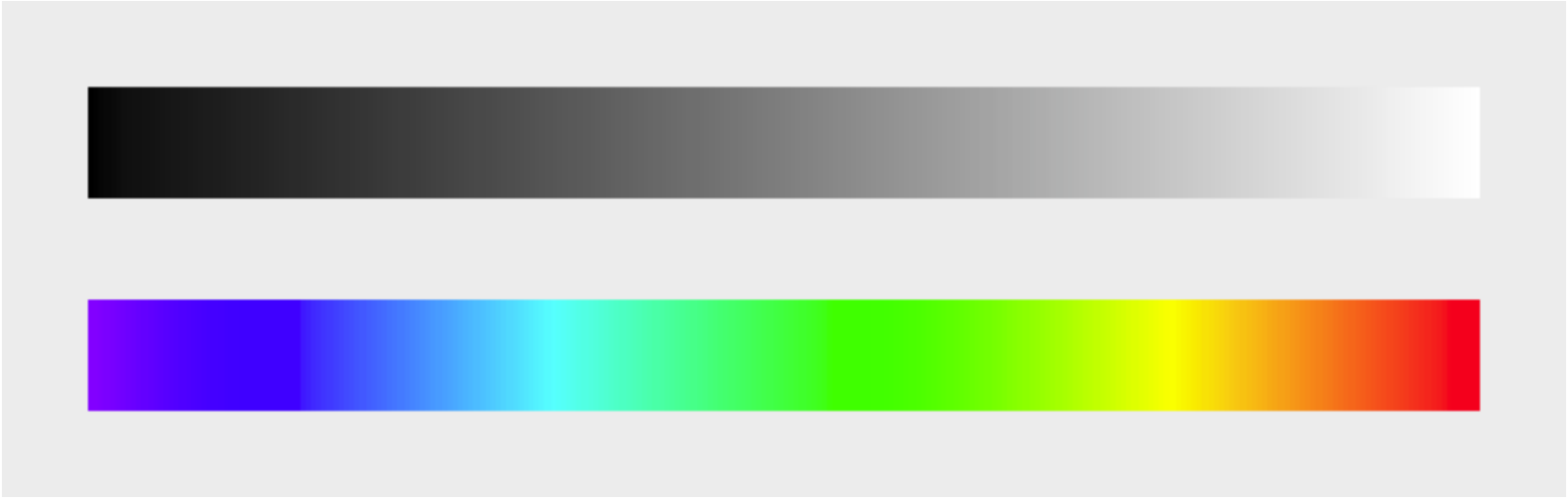


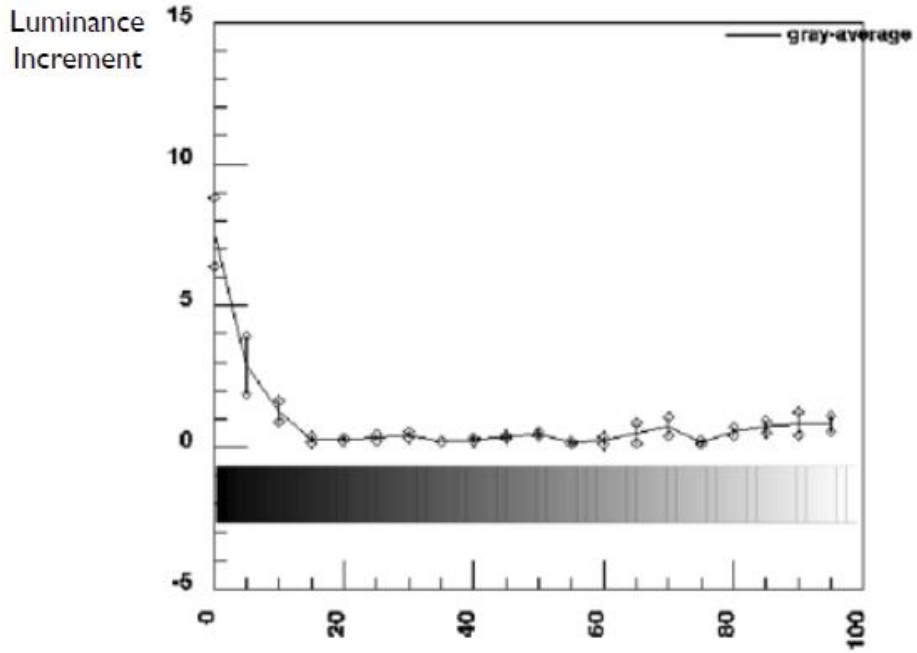
http://www.research.ibm.com/people/l/lloydt/color/color.HTM

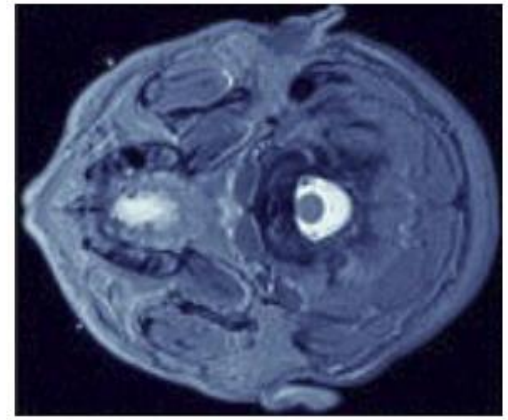
B. Rogowitz & L. Treinish

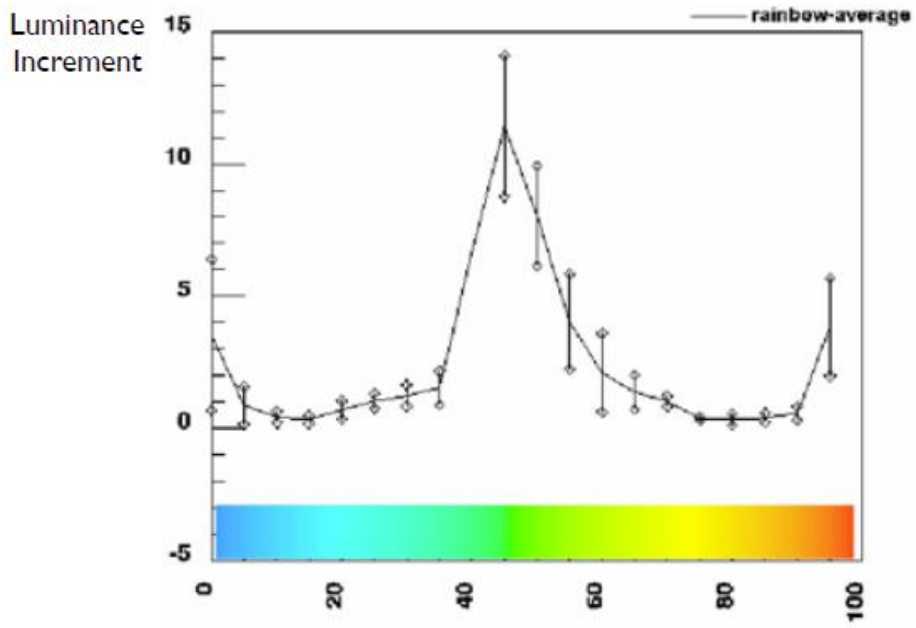
Rainbow Colormap

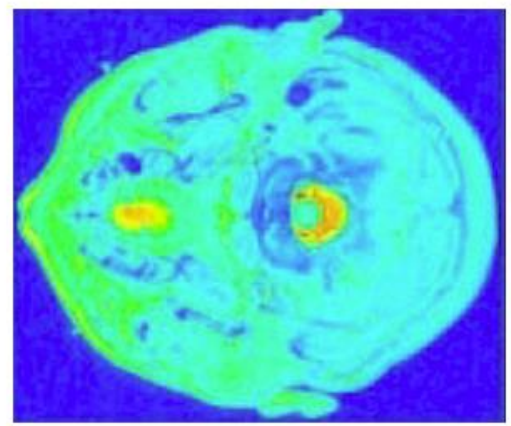
Rainbow map is perceptually nonlinear



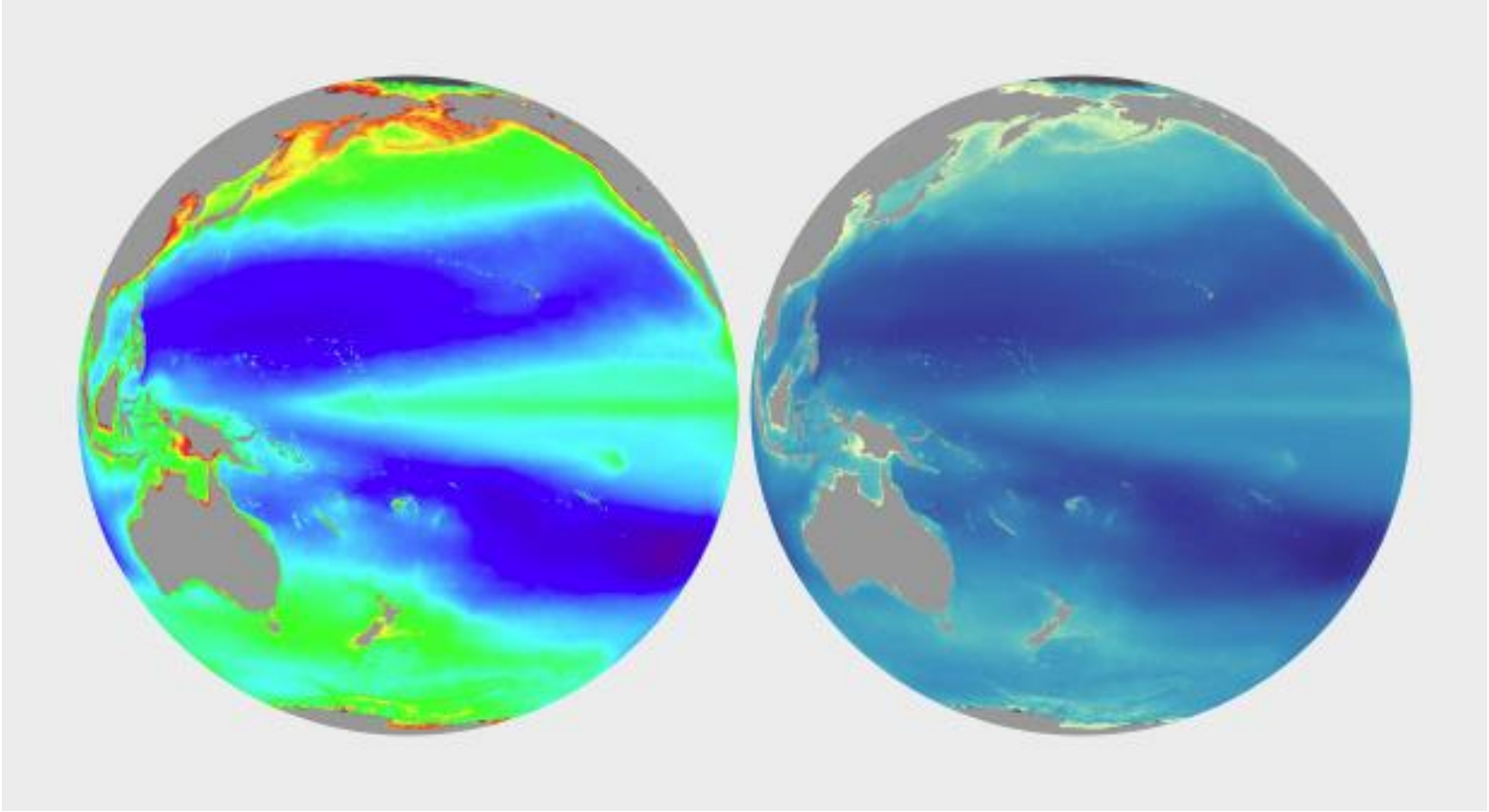

Luminance
Increment
gray-average
15
10
5
0
-5
0
20
40
60
80
100




Luminance
Increment
rainbow-average
15
10
5
0
-5
0
20
40
60
80
100



Rainbow Colormap



Map Example Revisited

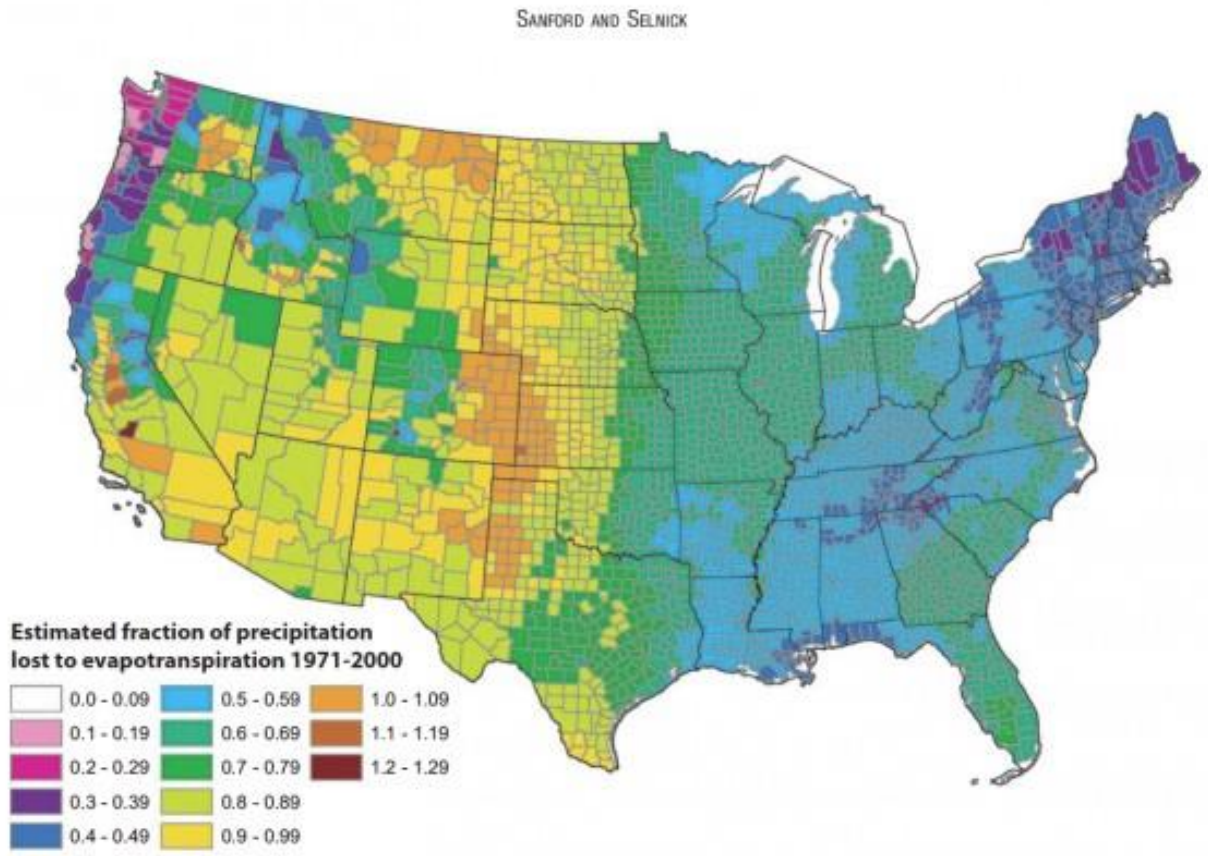
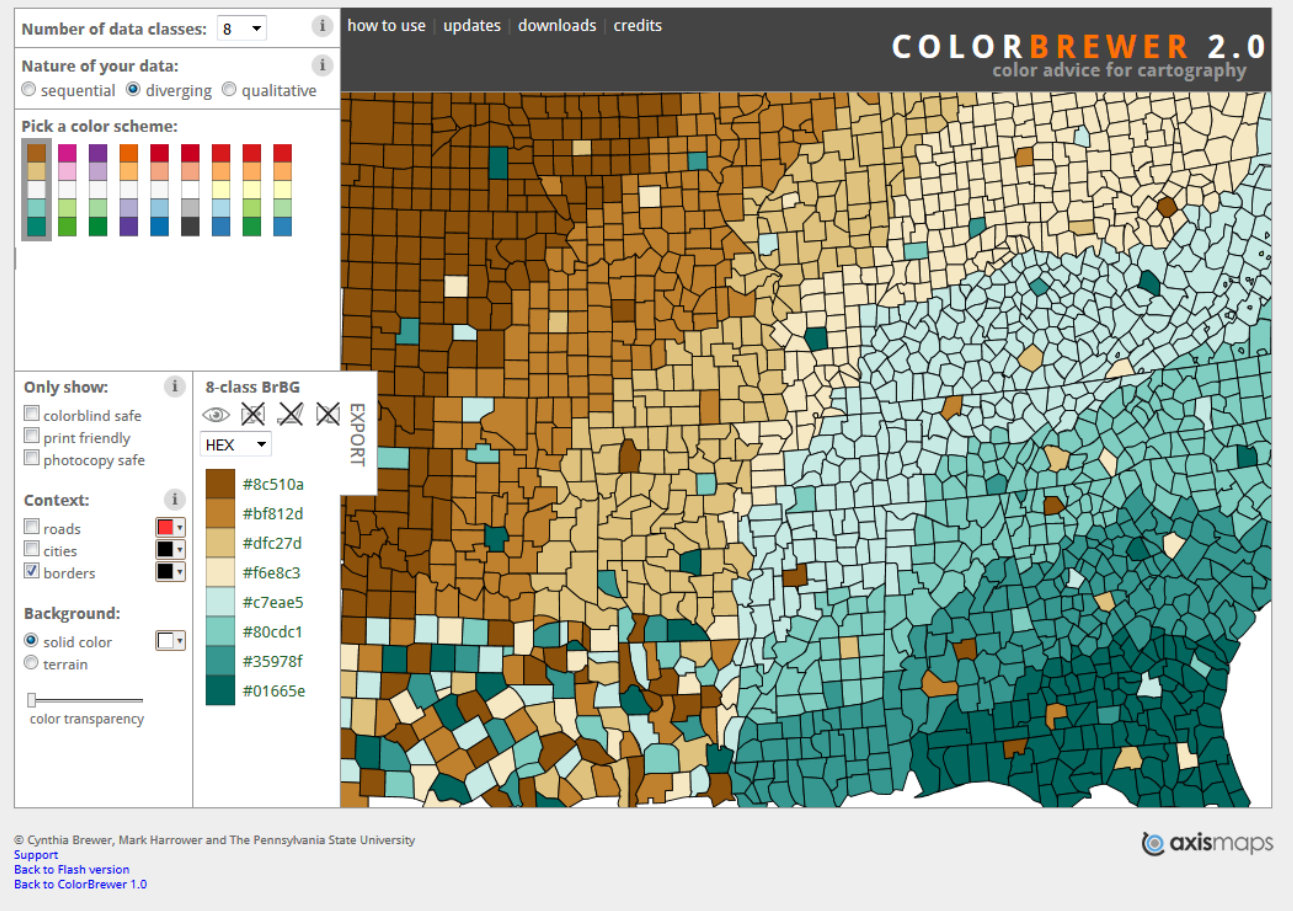


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (*P*) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/*P* were made first at the 800-m resolution of the 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.

0.0 - 0.09
0.1 - 0.19
0.2 - 0.29
0.3 - 0.39
0.4 - 0.49
0.5 - 0.59
0.6 - 0.69
0.7 - 0.79
0.8 - 0.89
0.9 - 0.99
1.0 - 1.09
1.1 - 1.19
1.2 - 1.29

0.0 - 0.09
0.1 - 0.19
0.2 - 0.29
0.3 - 0.39
0.4 - 0.49
0.5 - 0.59
0.6 - 0.69
0.7 - 0.79
0.8 - 0.89
0.9 - 0.99
1.0 - 1.09
1.1 - 1.19
1.2 - 1.29

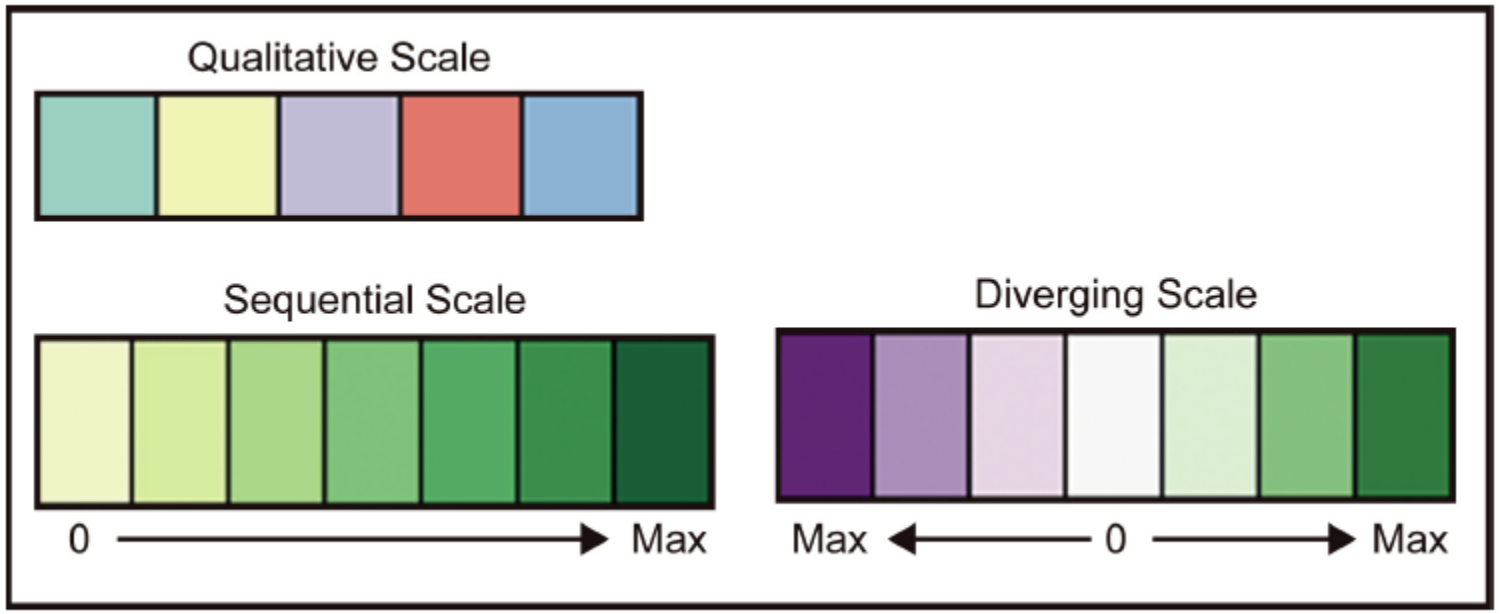
Use ColorBrewer! - http://colorbrewer2.org/

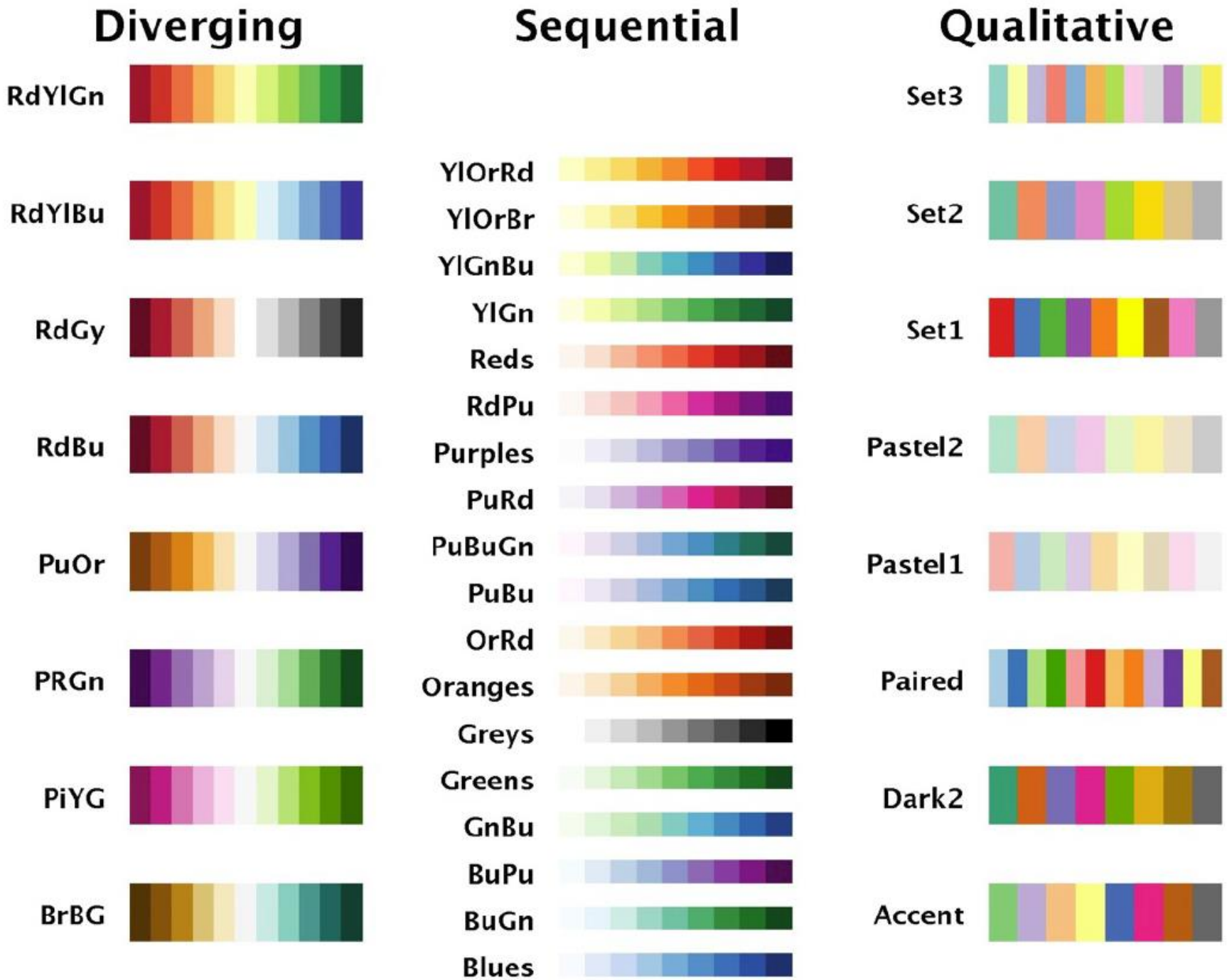


Brewer Scales (Cynthia Brewer)

Nominal

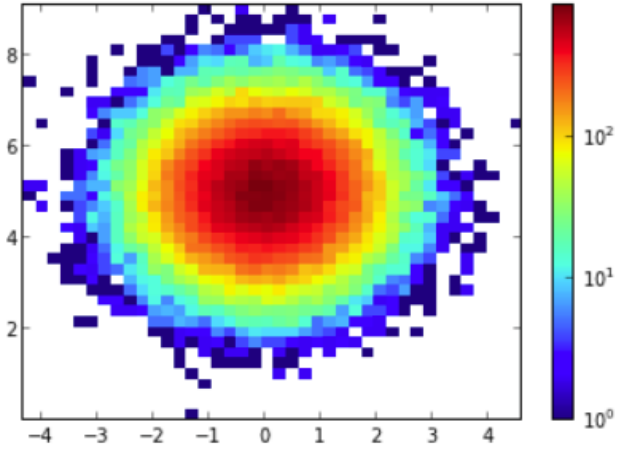
Ordinal

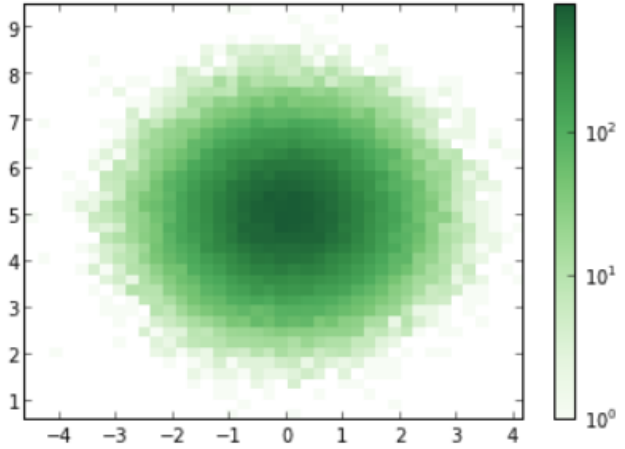


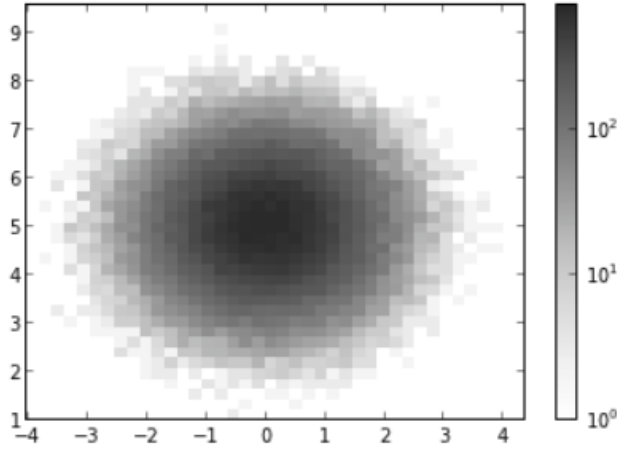

Diverging
RdYlGn
RdYlBu
RdGy
RdBu
PuOr
PRGn
PiYG
BrBG
Sequential
YlOrRd
YlOrBr
YlGnBu
YlGn
Reds
RdPu
Purples
PuRd
PuBuGn
PuBu
OrRd
Oranges
Greys
Greens
GnBu
BuPu
BuGn
Blues
Qualitative
Set3
Set2
Set1
Pastel2
Pastel1
Paired
Dark2
Accent

Sequential Brewer Scales

No!

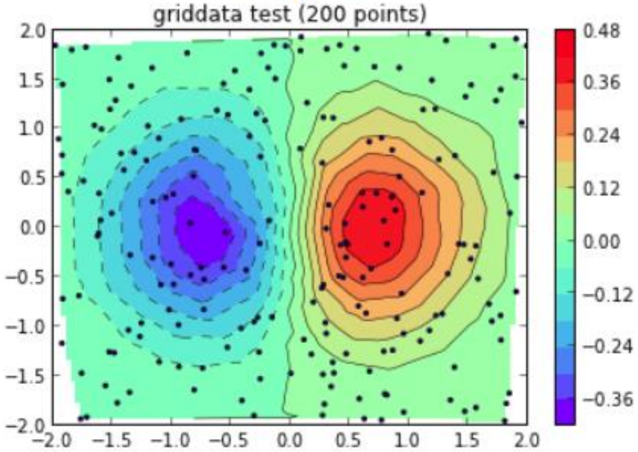


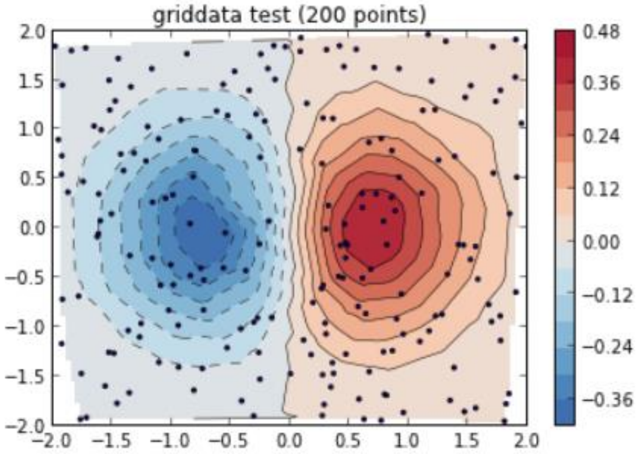


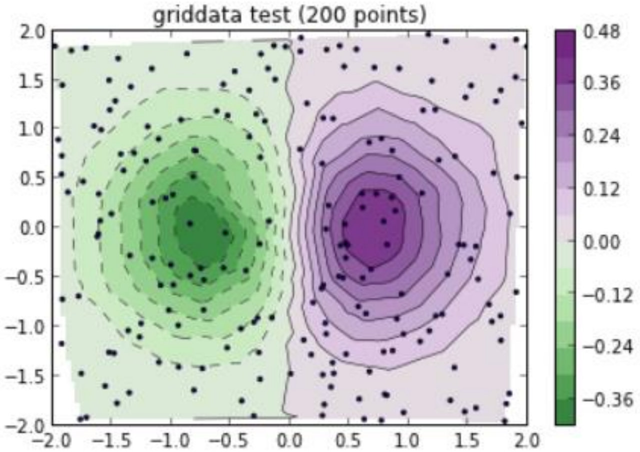


Divergent Brewer Scales

Not great

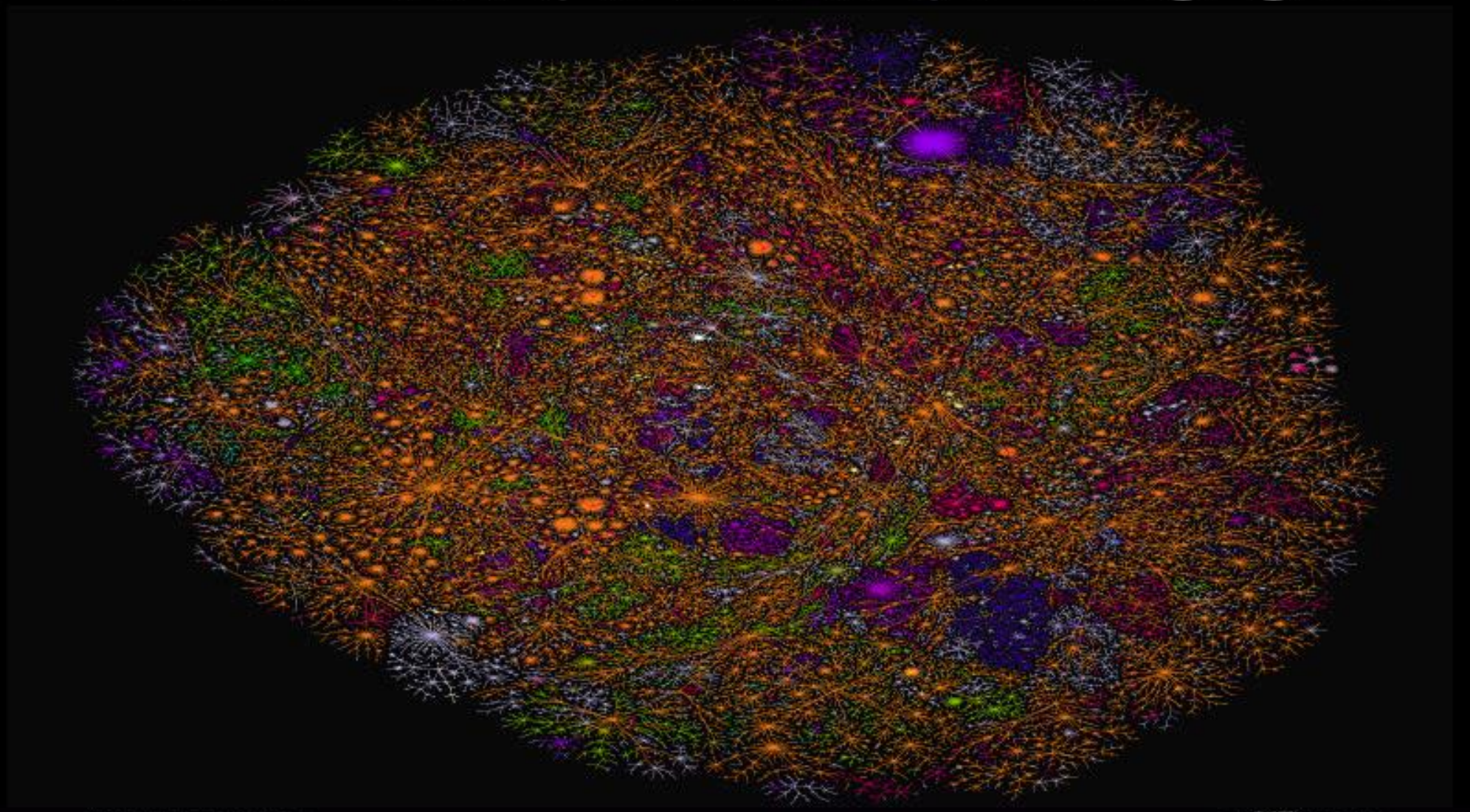


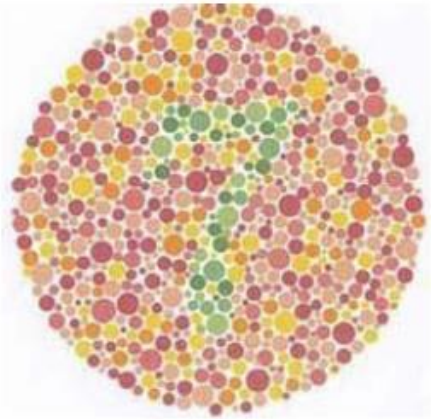


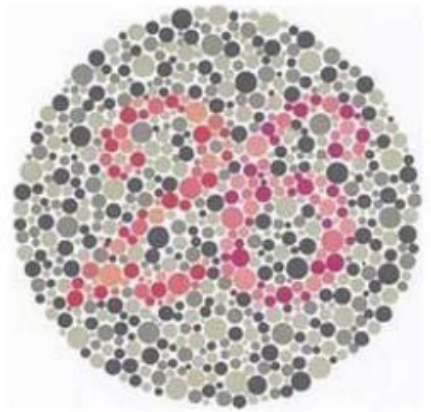


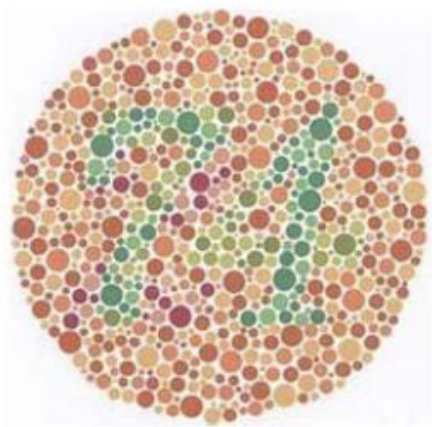
http://nbviewer.ipython.org/5357268

The Internet: 2002









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

Fonts

A helpful guide

Title 28 bold monospace:

Consolas, Lucida Console, Courier New

Title 28 bold serif:

Palatino Linotype, Times New Roman, Georgia, Baskerville, Garamond, Cambria

Title 28 bold san-serif:

Century Gothic, Verdana, Tahoma, Arial, Arial Narrow, Franklin Gothic Medium, Lucida Sans, Lucida Sans Unicode, GillSans, Calibri, Trebuchet

Body Copy 11 monospace:

Consolas 1234567890, Lucida Console 1234567890, Courier New 1234567890

Body Copy 11 serif:

Palatino Linotype 1234567890,Times New Roman 1234567890, Baskerville 1234567890, Georgia 1234567890, Cambria 1234567890

Body Copy 11 san-serif:

Century Gothic 1234567890, Lucida Sans 1234567890, Lucida Sans Unicode 1234567890, Verdana 1234567890, GillSans 1234567890, Calibri 1234567890, Arial 1234567890, Arial Narrow 1234567890, Verdana 1234567890, Tahoma 123456789..

Numbers 8 monospace:

Consolas1234567890, Lucida Console1234567890, Courier New1234567890

Numbers 8 serif:

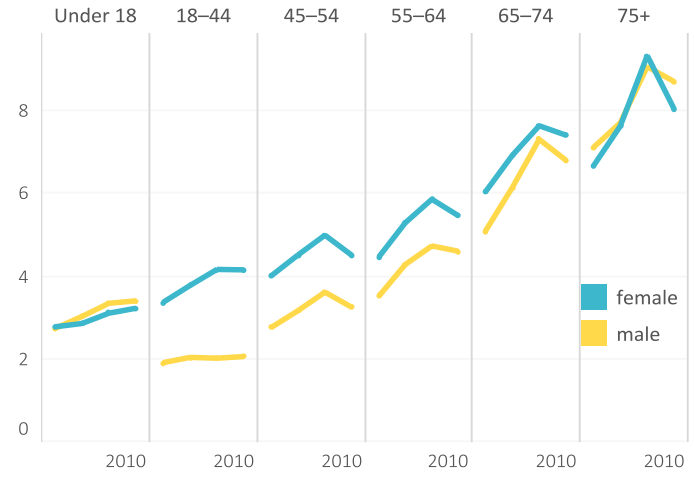
Palatino Linotype1234567890,Times New Roman1234567890, Baskerville1234567890, Georgia1234567890, Cambria1234567890

Numbers 8 san-serif:

Century Gothic,1234567890, Lucida Sans1234567890, Lucida Sans Unicode1234567890, Verdana1234567890, GillSans1234567890, Calibri1234567890, CalibriLight1234567890, Arial 1234567890, Arial Narrow1234567890, Verdana1234567890, Tahoma1234567890, Trebuchet1234567890

Health Care Visits

Average visits per person for males and females in different age categories for the years 1995-2010



Title:Consolas 28 bold

Body text: Consolas 11

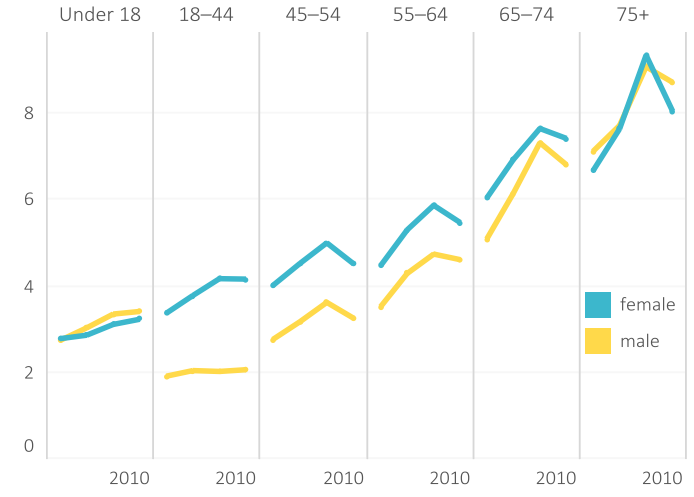
Headers: Calibiri 9

Legends: Calibiri 8

Numbers: Calibiri 8

Health Care Visits

Average visits per person for males and females in different age categories for the years 1995-2010



Title: GillSans MT 28 bold

Body text: GillSans MT 11

Headers: Calibiri Light 9

Legends: Calibiri Light 8

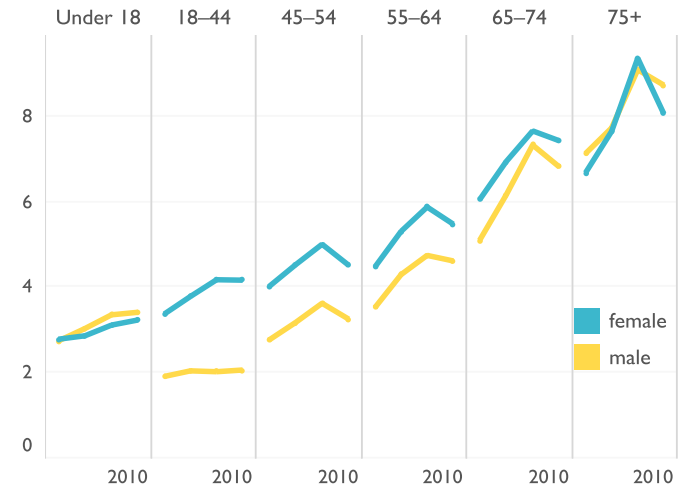
Numbers: Calibiri Light 8

Title: Century Gothic 28 bold

Body text: Palatino Linotype 11
Headers: GillSans MT 9
Legends: GillSans MT 8
Numbers: GillSans MT 8

Health Care Visits

Average visits per person for males and females in different age categories for the years 1995-2010

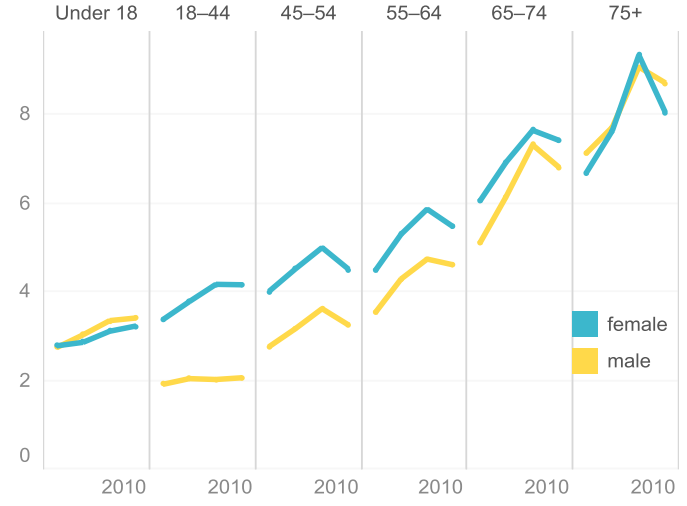


Title: Georgia 28 bold

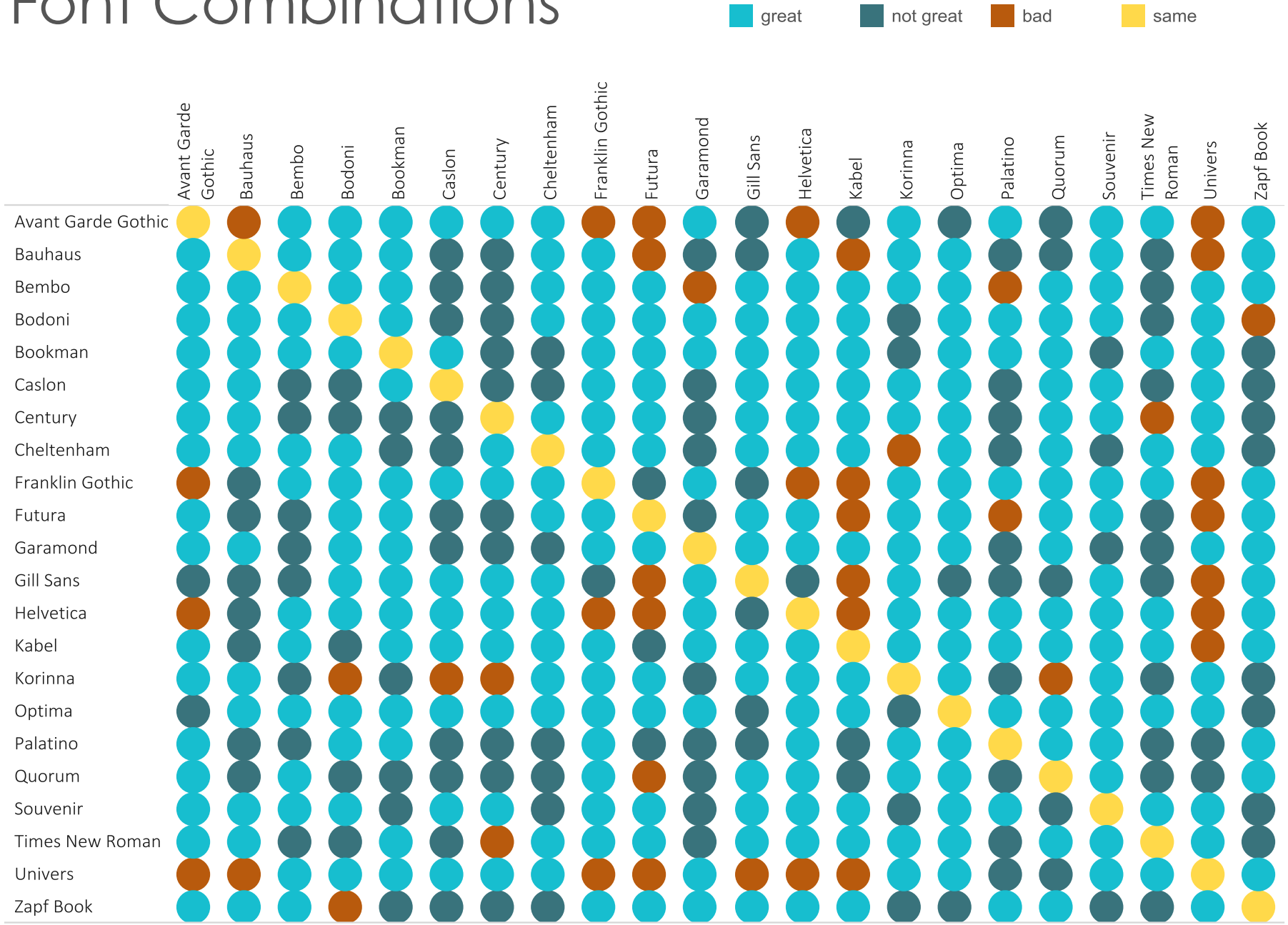
Body text: Franklin Gothic Medium 11
Headers: Arial 9
Legends: Arial 8
Numbers: Arial 8

Health Care Visits

Average visits per person for males and females in different age categories for the years 1995-2010



Font Combinations



Graphical Excellence & Honesty in Data Visualization

Edward Tufte

Graphical Excellence (Tufte)

Well-designed presentation of interesting data – a matter of

- substance
- statistics
- design

That which gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space

Nearly always multivariate

…and requires telling the truth about the data

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. ___ Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'oeil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

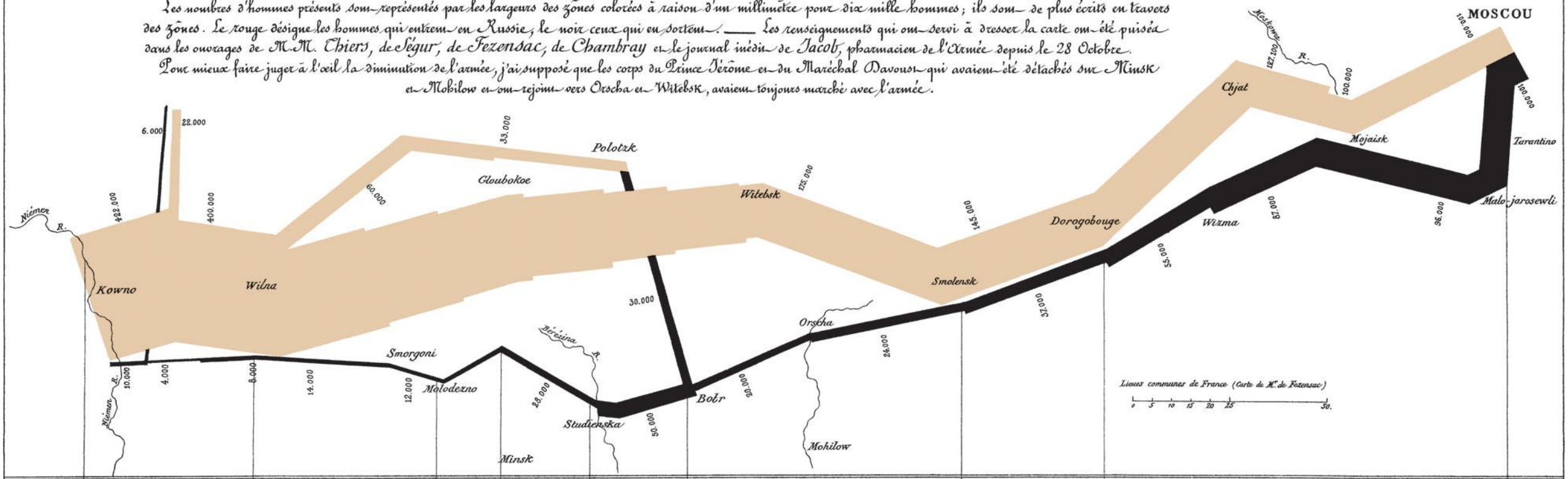
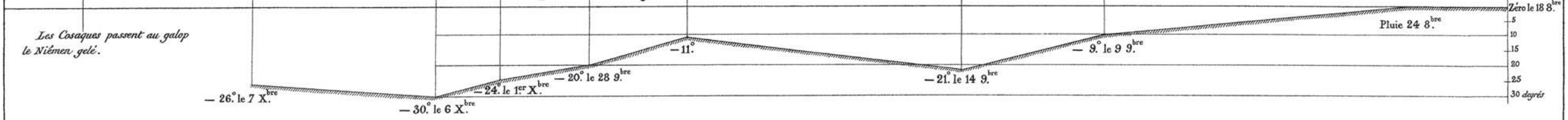


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Autog. par Regnier, 8. Pas. Ste. Marie St. Gein à Paris.

Imp. Lith. Regnier et Dourdet.

Timing cycle for ignition of engines to make orbit corrections: "Zone of corrective engines."

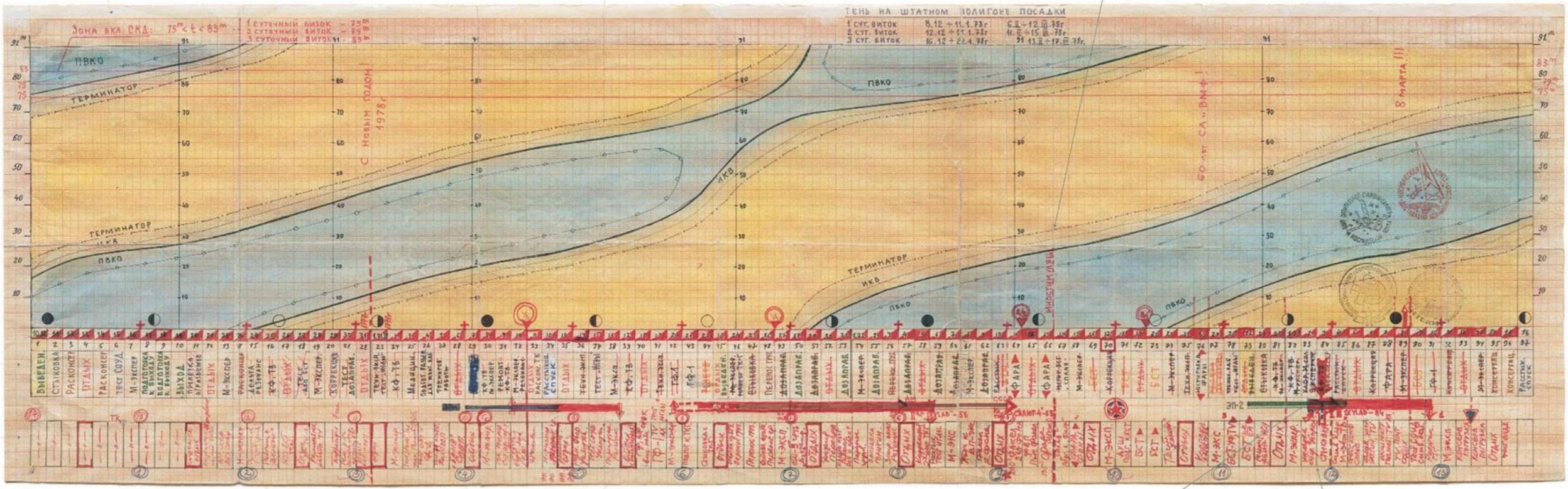
Happy New Year! С Новым годом! 1978г

Visit by *Soyuz 27*, January 10–16. Grechko caught a cold, probably from one of the guests. Visits are marked by solid red lines between the planned and actual schedules.

Shadow on site of landing.

A cargo ship brought up an electric muscle-stimulator (миостимулят) to supplement the daily exercise regime, which the cosmonauts were neglecting.

Markings for space-mail delivery.



Grechko and Romanenko arrive at space station *Salyut 6*, via *Soyuz 26*.

Actual activity for each day, recorded in red pen, while in orbit. As the flight goes on, the pen tip squashes and broadens.

Planned activity for each day, schedule prepared prior to flight, with variously colored pens. Note the many deletions and changes (e.g., ~~ТЕСТ~~ ~~430 БСТ~~).

Red boxes (ОТДЫХ ОТДЫХ, rest) indicate every 6th day of rest (revised schedule).

Spacewalk (ВЫХОД ВЫХОД, exit) by Grechko on December 20. When Grechko returned, Romanenko decided on the spur of the moment to look around outside. While pushing himself through the tight airlock, he lost his grip and began to drift away from the spacecraft! Romanenko's line was not secure; at the last second, Grechko caught hold of the line and pulled his floating colleague to safety. The cosmonauts waited until months after their return to Earth before saying anything to authorities about their near disaster. Romanenko went on to spend 430 days in space on this and later flights.

Visit by *Progress 1*, January 20 to February 6. This automated cargo ship, without a crew, brought equipment, fresh fruit, mustard, horseradish, bread, music tapes and a cassette recorder, clothes, linen, air filters, an atlas, newspapers, and mail.

"БАНЯ" СЛАВП (БАНЯ) "Steamroom," a traditional Russian bath. Surrounded by quotation marks, the word is used here sarcastically, mocking the engineering jargon (the acronym СЛАВП) that calls a space-bath "A System of Taking Water Procedures."

On a television broadcast of February 10, Grechko noted a birthday: "Hello everyone. Today marks the one hundred and fiftieth birthday of Jules Verne, the remarkable French writer. There's hardly a person who hasn't read his books, at any rate not among the cosmonauts, because Jules Verne was a dreamer, a visionary who saw flights in space. I'd say this flight too was predicted by Jules Verne."

A visiting cosmonaut quietly mentioned to Romanenko that Grechko's father had just died. Romanenko decided to tell Grechko only after they were safely back to earth.

This red arrow celebrates the 84th day in orbit when *Salyut 6* equaled the space endurance record set five years earlier by America's *Skylab 4*. A thick red line extends from the 84th to the 91st day, when Grechko and Romanenko exceeded the previous endurance record by the officially necessary ten percent.

February 23, the 60th anniversary of the Soviet Army and Navy. The cosmonauts observed notable dates to mark the progress on their very long stay in space.

8 МАРТА !!! March 8, International Women's Day. Grechko and Romanenko made a television broadcast and prepared mail to send back on *Soyuz 28*.

Visit by *Soyuz 28*, March 2–10. In addition to a Czech cosmonaut (who became the first non-Soviet, non-American in space), the visitors brought a package with fresh onions, garlic, Bulgarian peppers, lemons, apples, milk, gingerbread, and honey.


The Broad Street Well
GREAT MARLBOROUGH STREET
WORK HOUSE
SOHO SQUARE
GOLDEN SQUARE

Principles of Graphical Integrity

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

Lie Factor = (size of effect shown in graphic) / (size of effect in data)

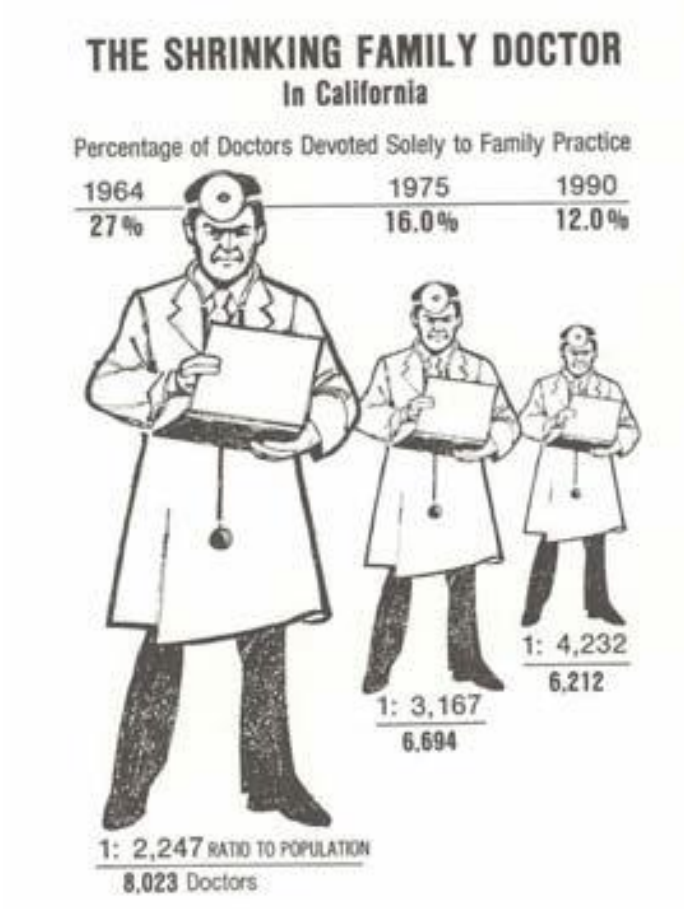
Clear, detailed, and thorough labeling should be used.

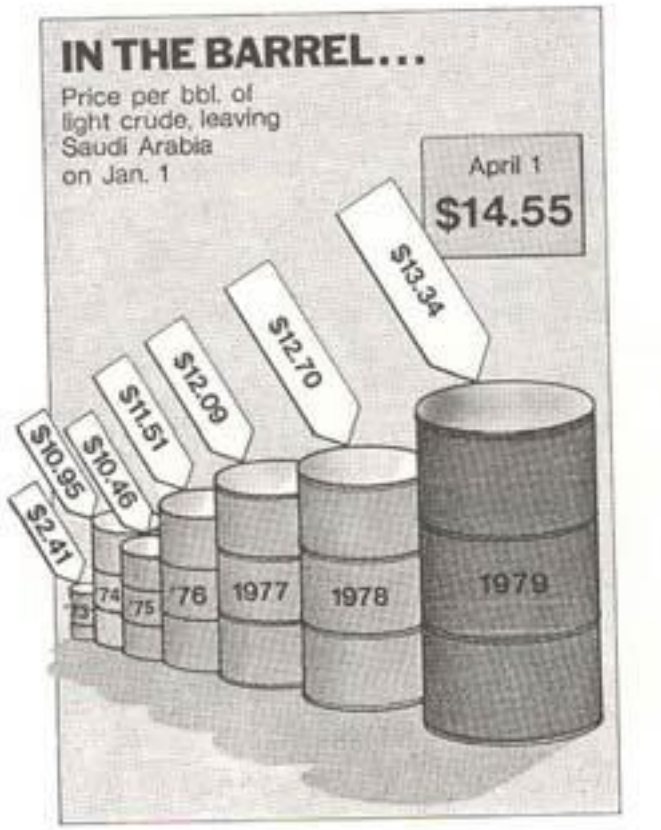
Show data variation, not design variation.

Graphics must not quote data out of context.

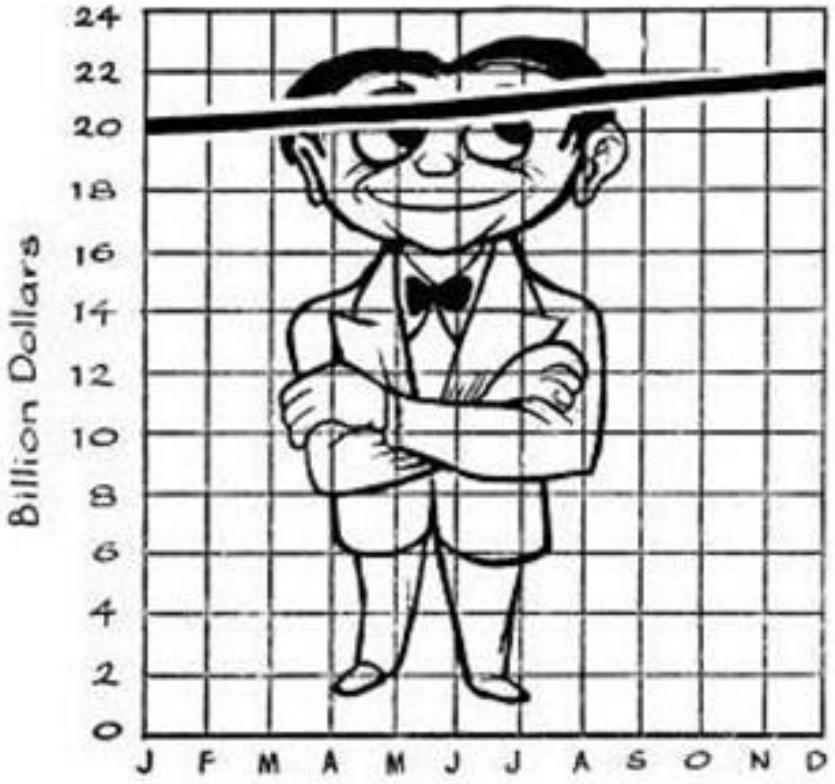
The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data

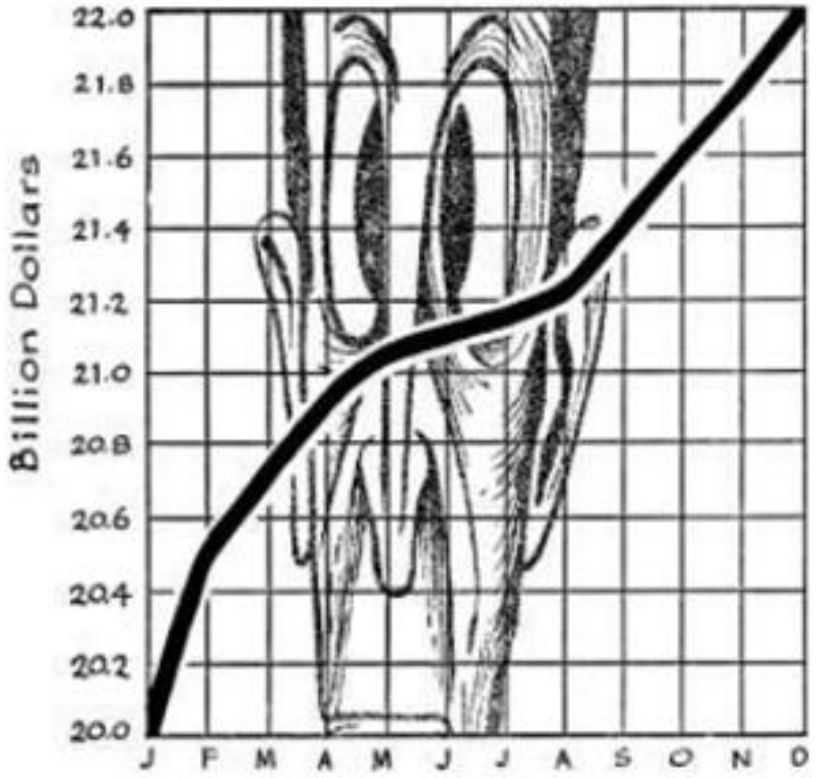
The Lie Factor = $\frac{\text{(size of effect shown in graphic)}}{\text{(size of effect in data)}}$

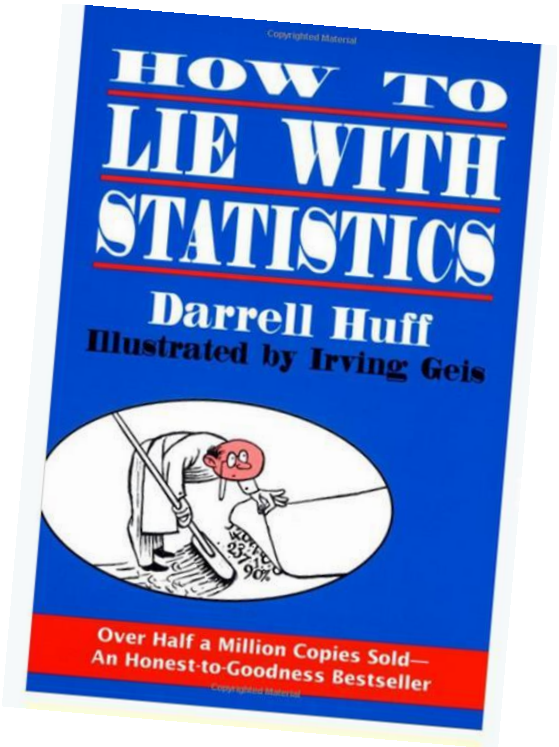




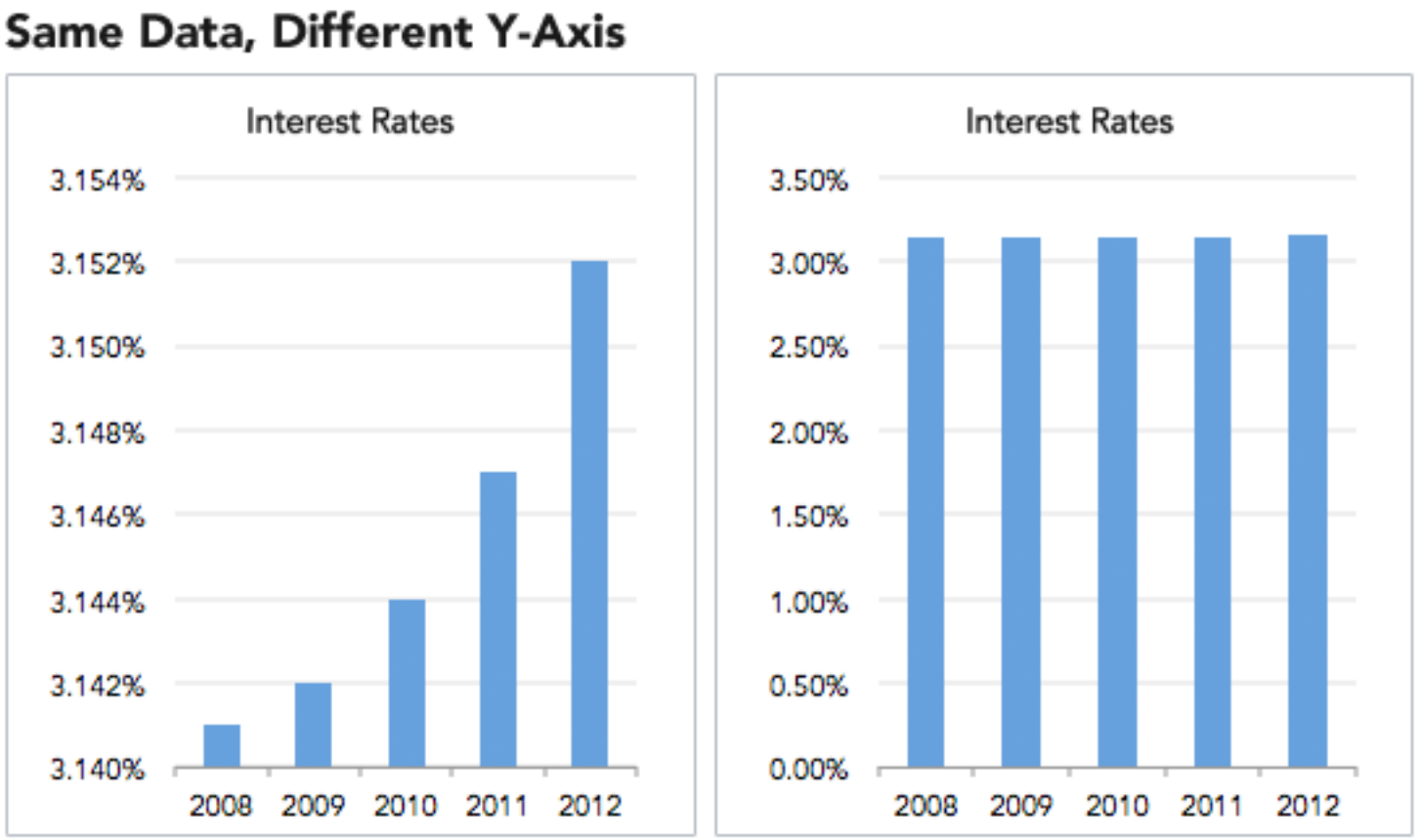
Honesty in Data Visualization



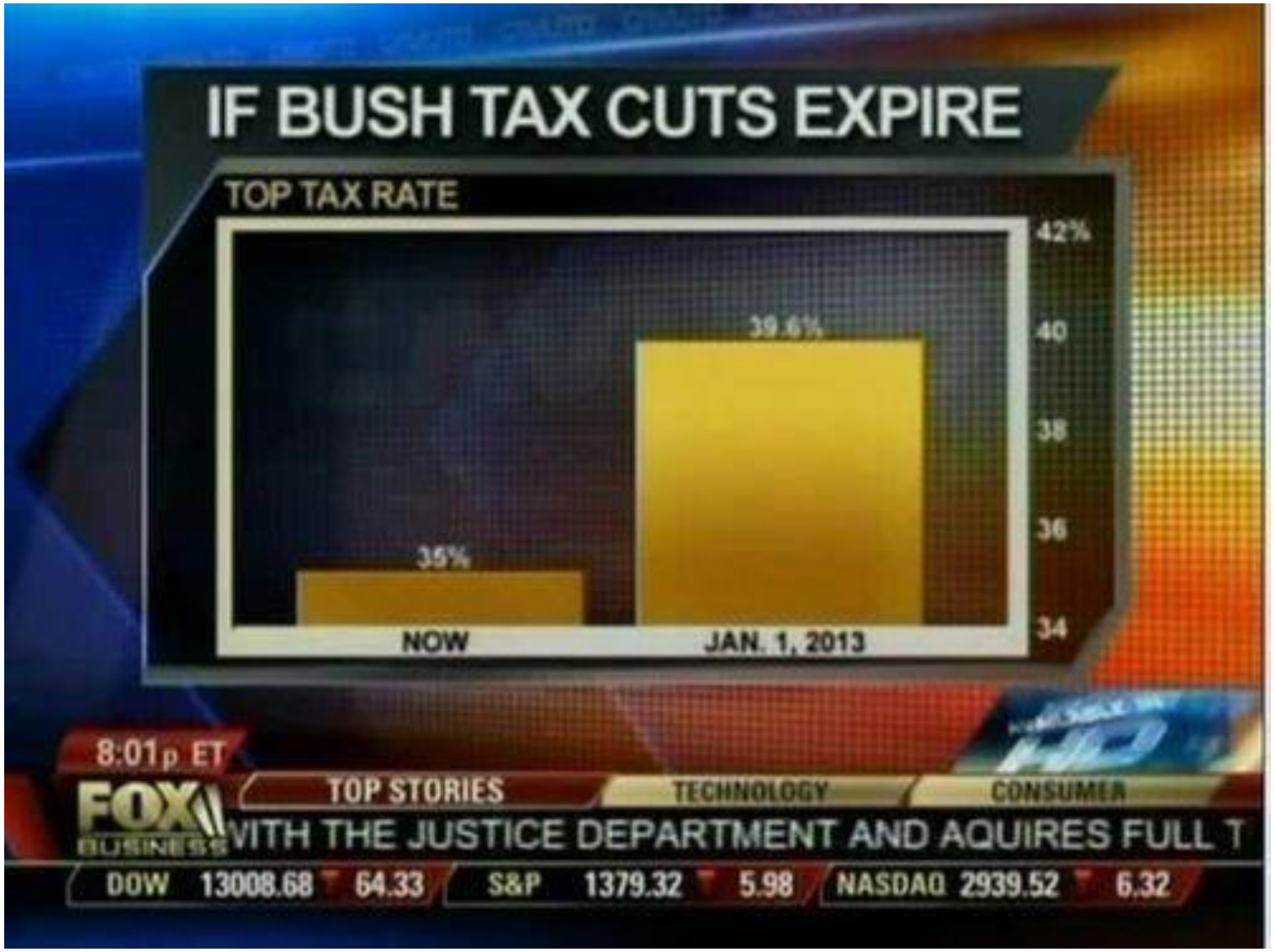




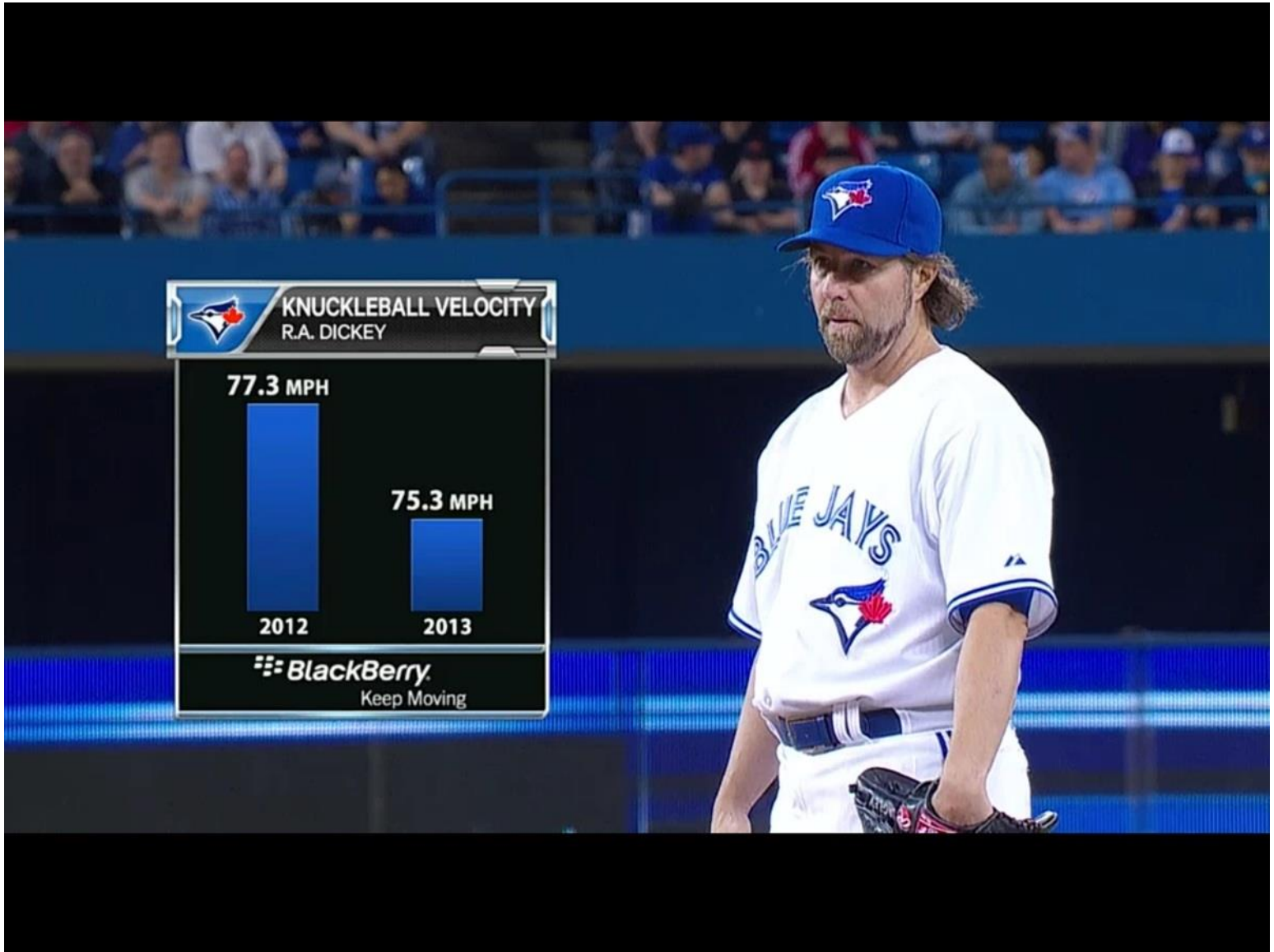
Lies - Truncated Y-Axis



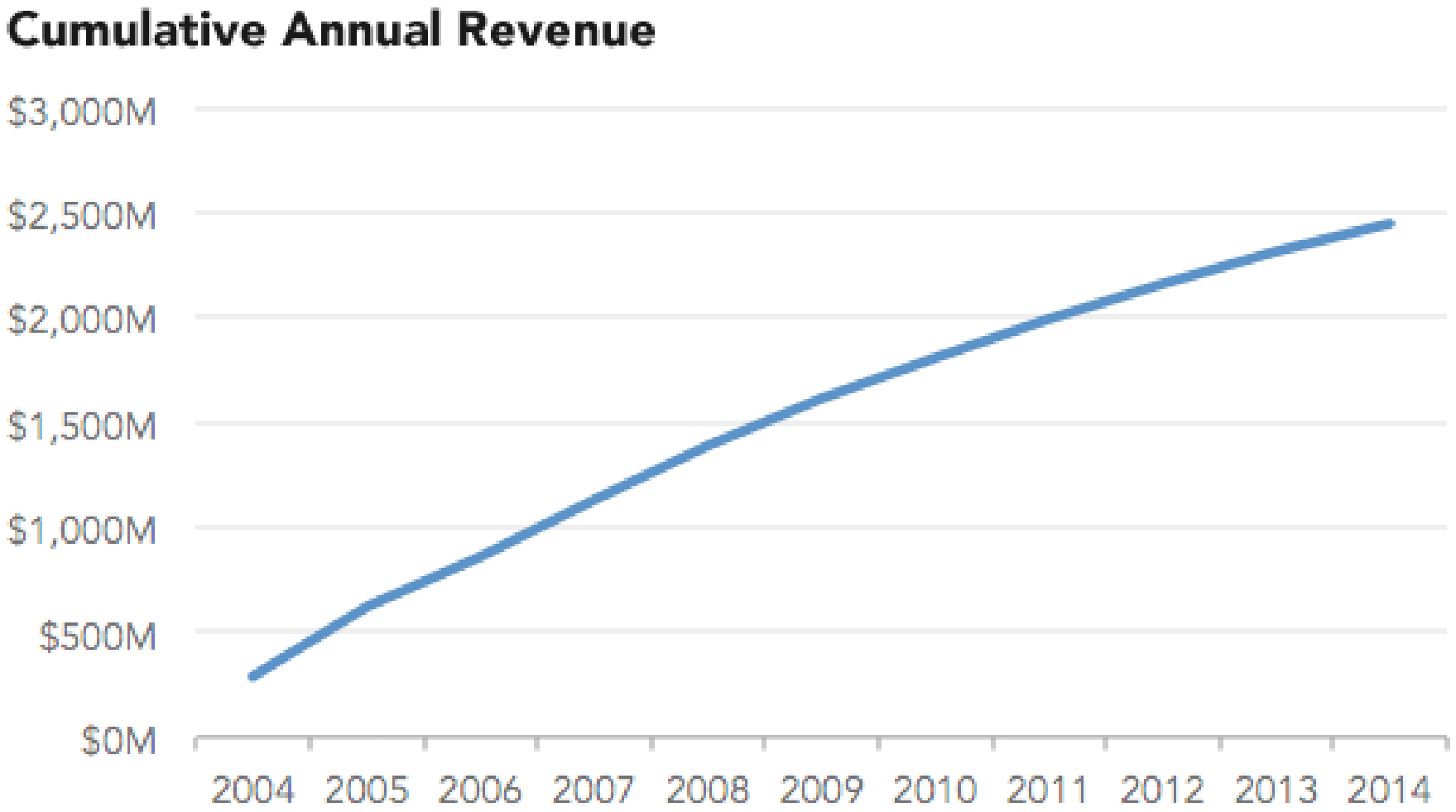
Lies - Truncated Y-Axis



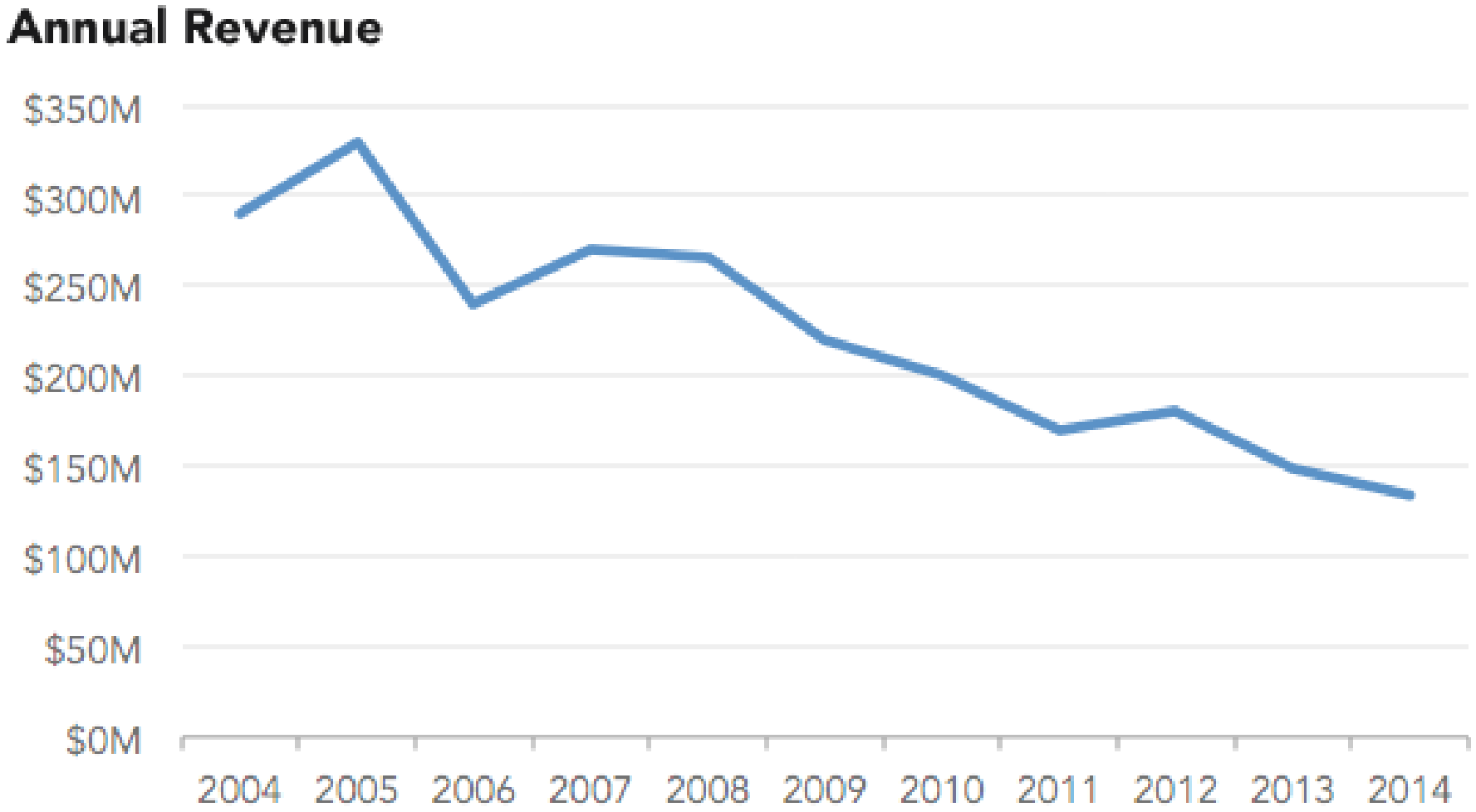
Lies - Truncated Y-Axis



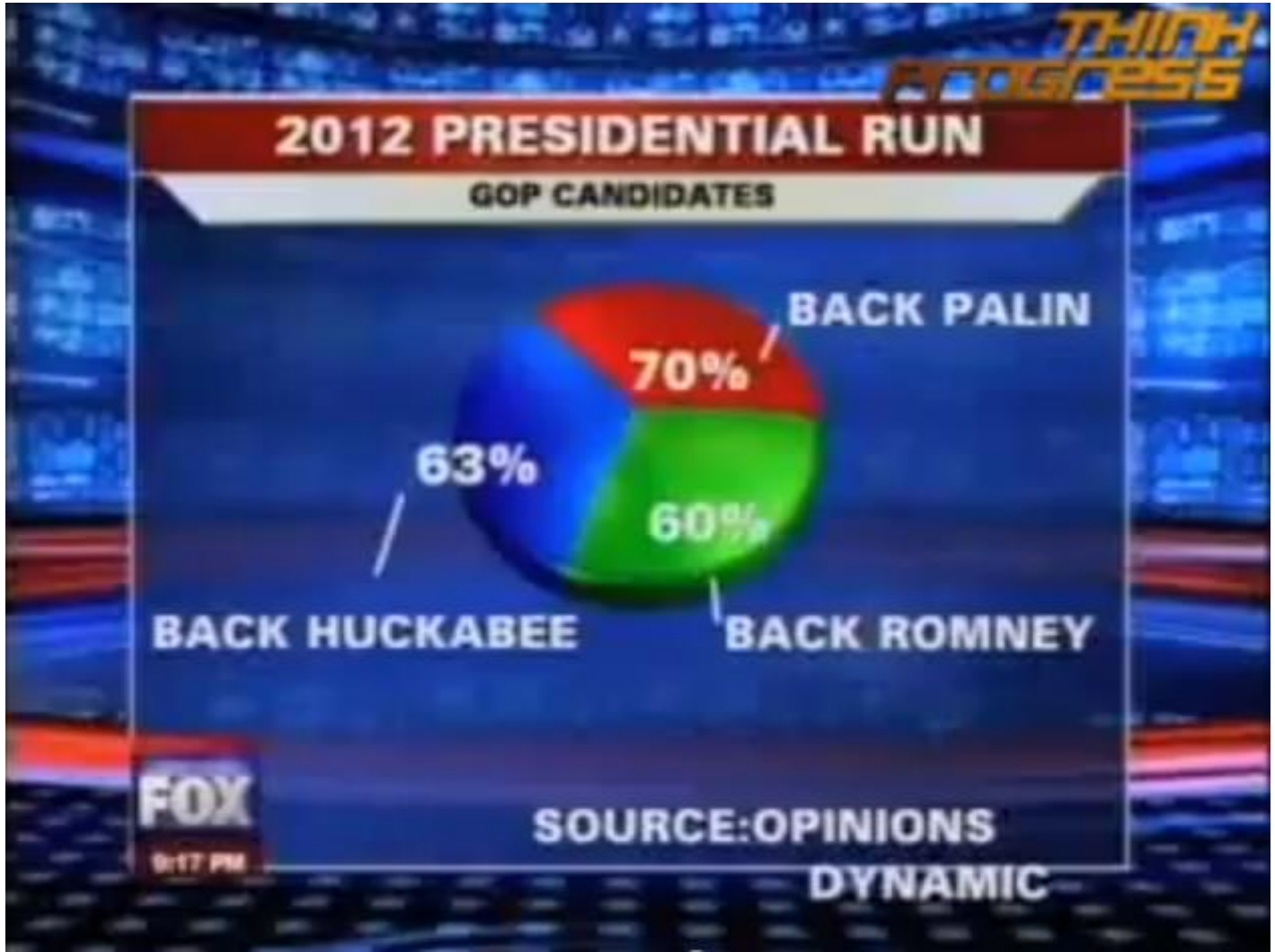
Lies - Cumulative Graphs



Lies - Cumulative Graphs



Lies - Ignoring Conventions



Why Do People Lie?

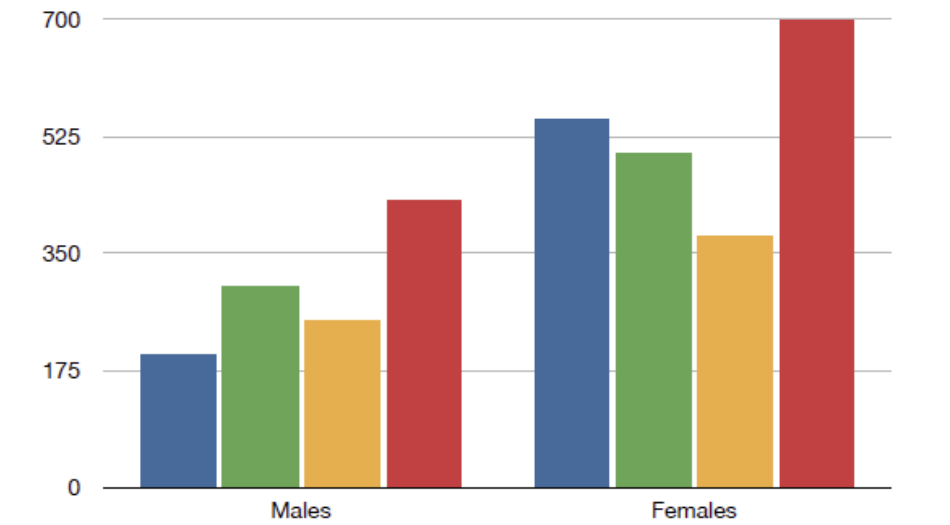
- Lack of quantitative skills
- Desire to “jazz up” the data
- Desire to simplify the data

Maximize Data-Ink Ratio

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

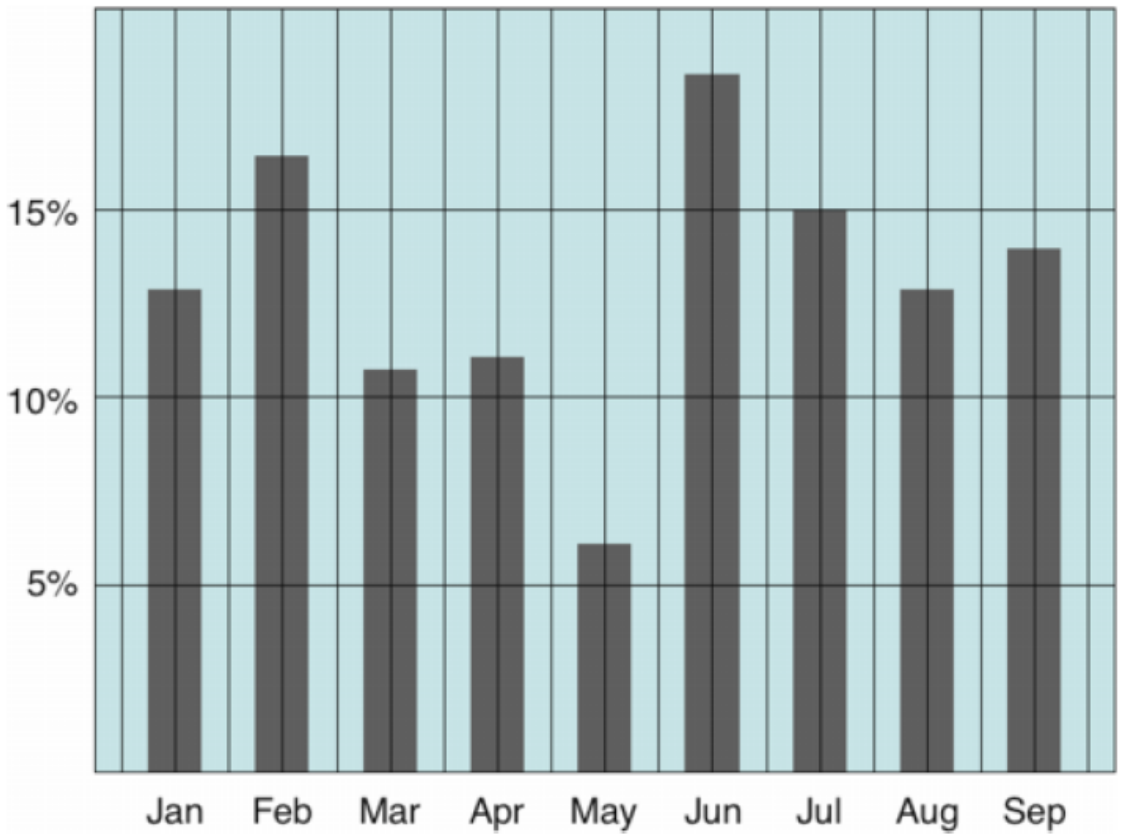
Maximize Data-Ink Ratio

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



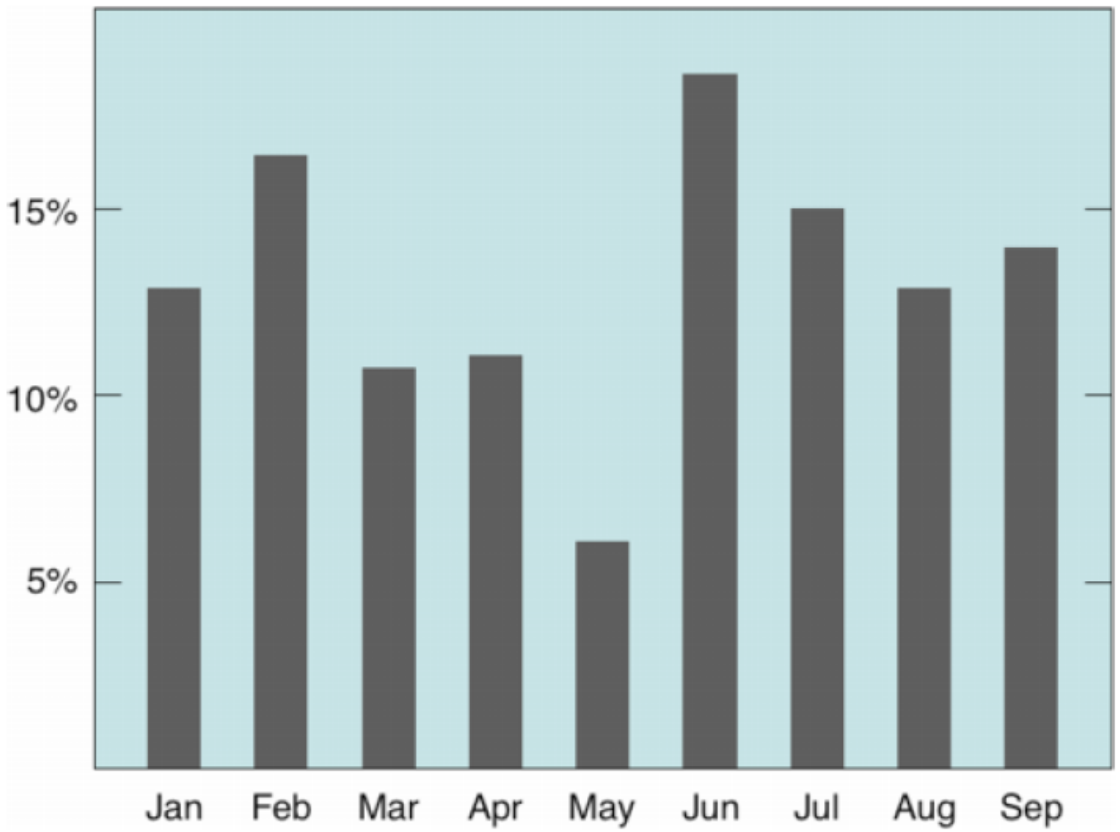
Avoid Chart Junk

Extraneous visual elements that distract from the message



Avoid Chart Junk

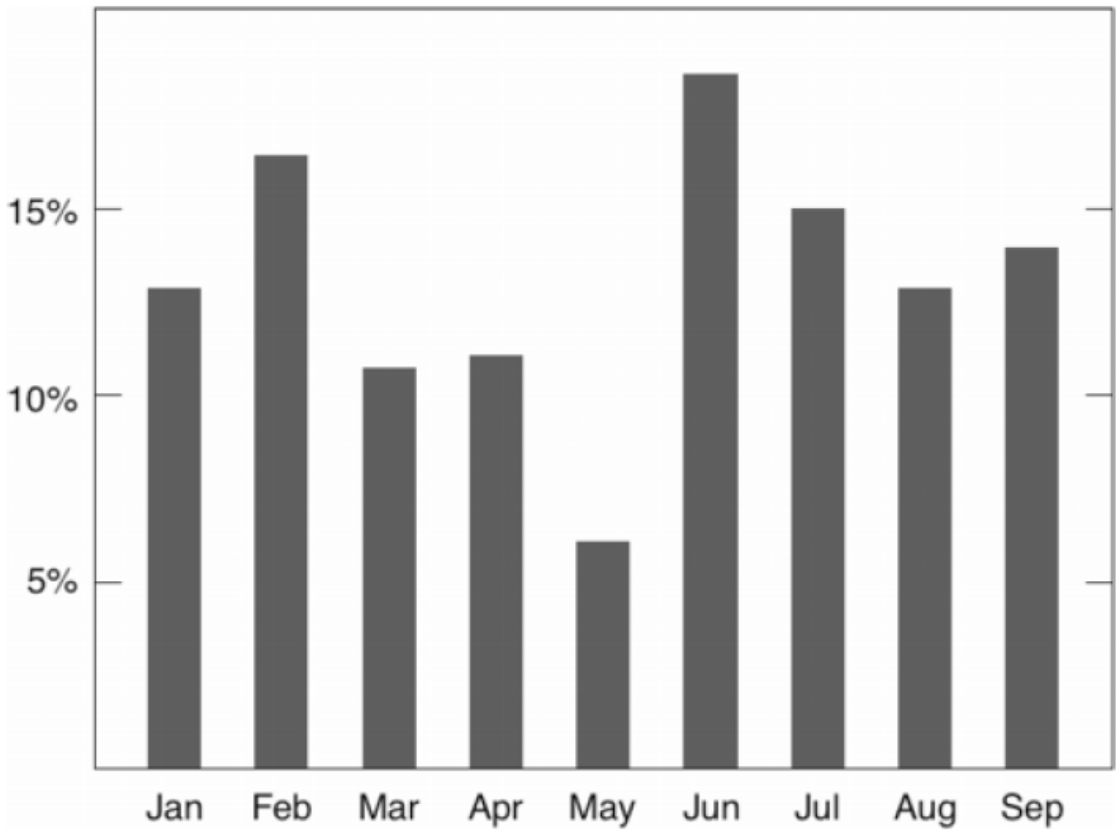
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

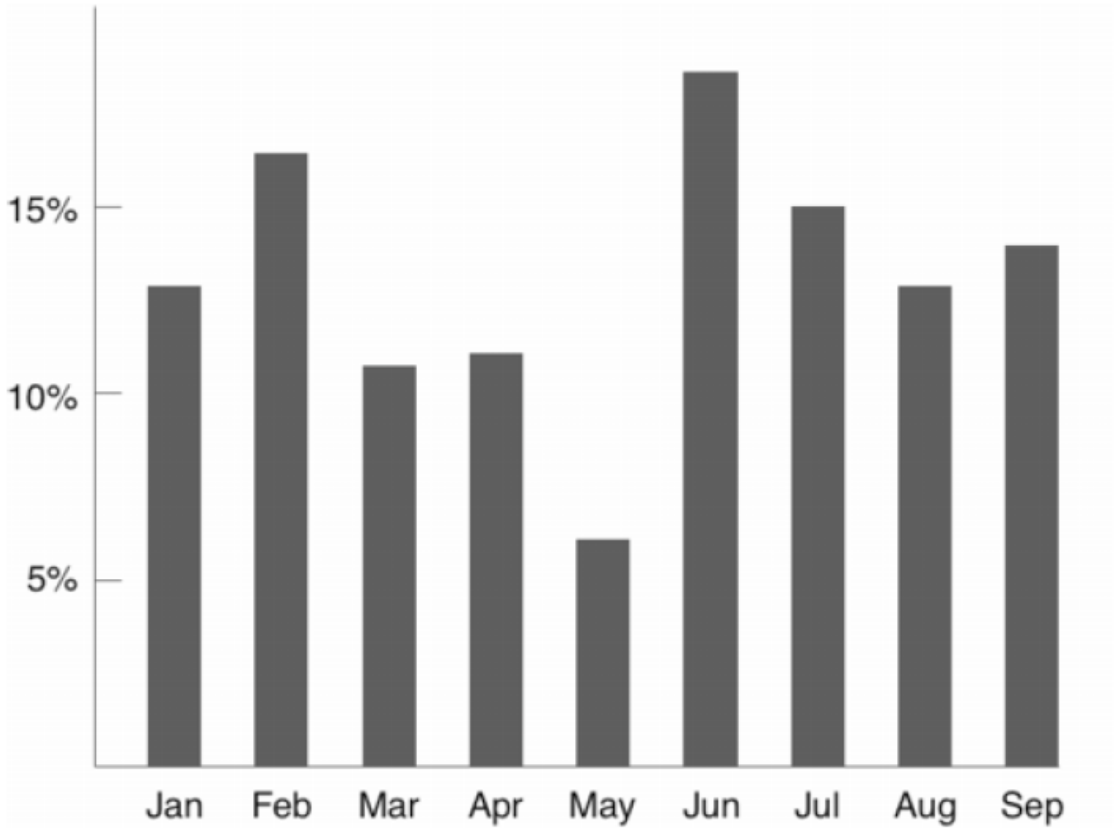
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

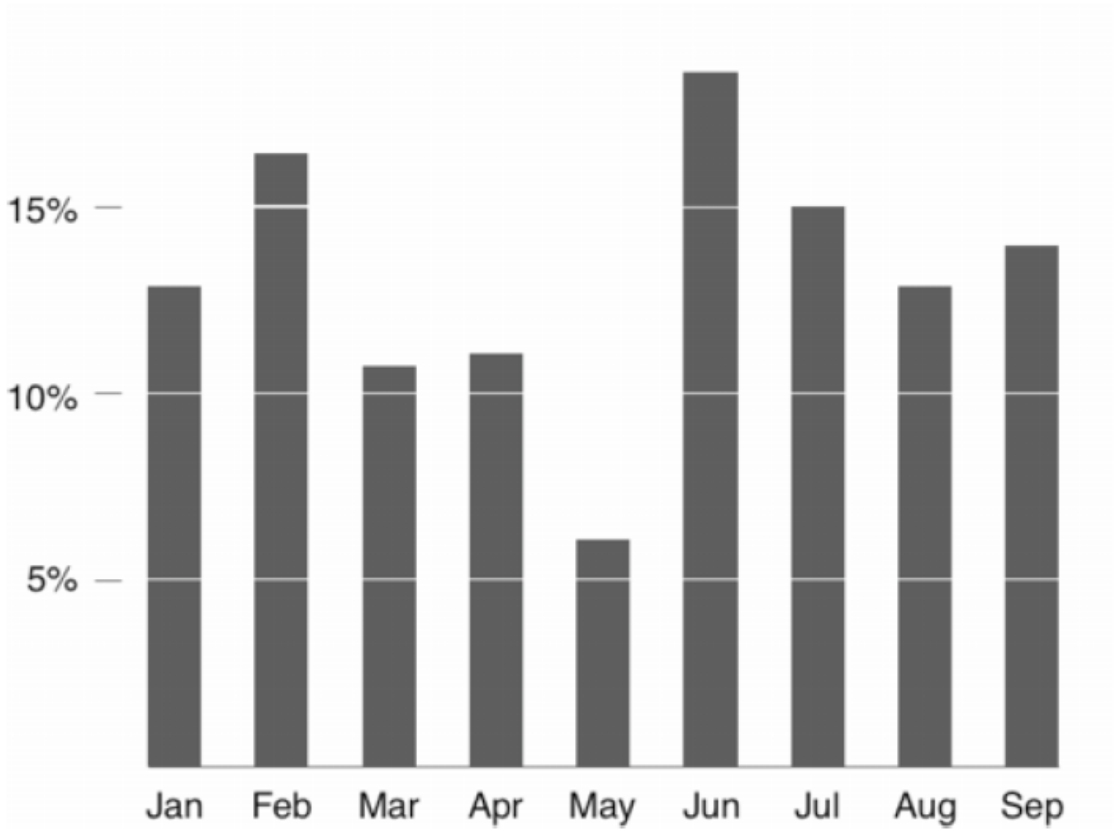
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

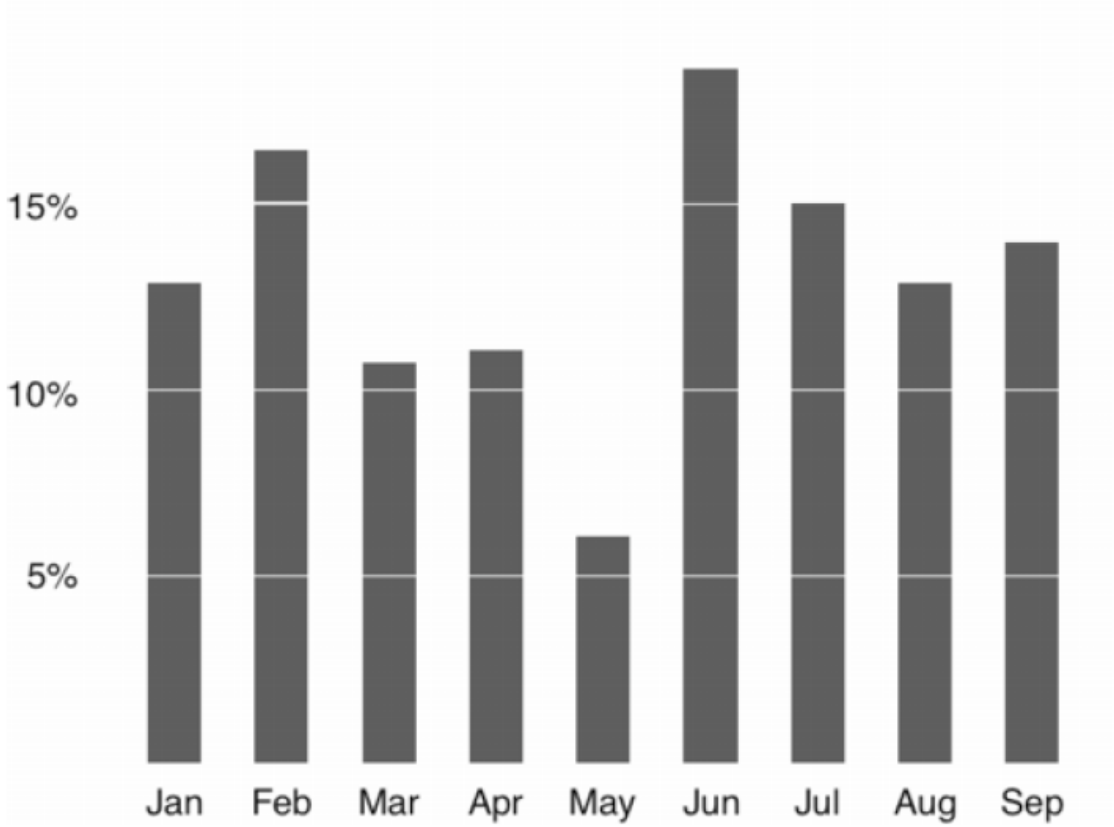
Extraneous visual elements that distract from the message



ongoing, Tim Brey

Avoid Chart Junk

Extraneous visual elements that distract from the message



ongoing, Tim Brey

Critiquing a Visualization

Critiquing a Visualization

- First, consider the purpose of the visualization and who the intended audience is.
- Then, ascertain your initial reaction.
- Then, examine the visualization in detail.
- Then, answer questions like…

Critiquing a Visualization

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?

Critiquing a 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?

Critiquing a Visualization

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?