Data Visualization

- Imdadullah Khan

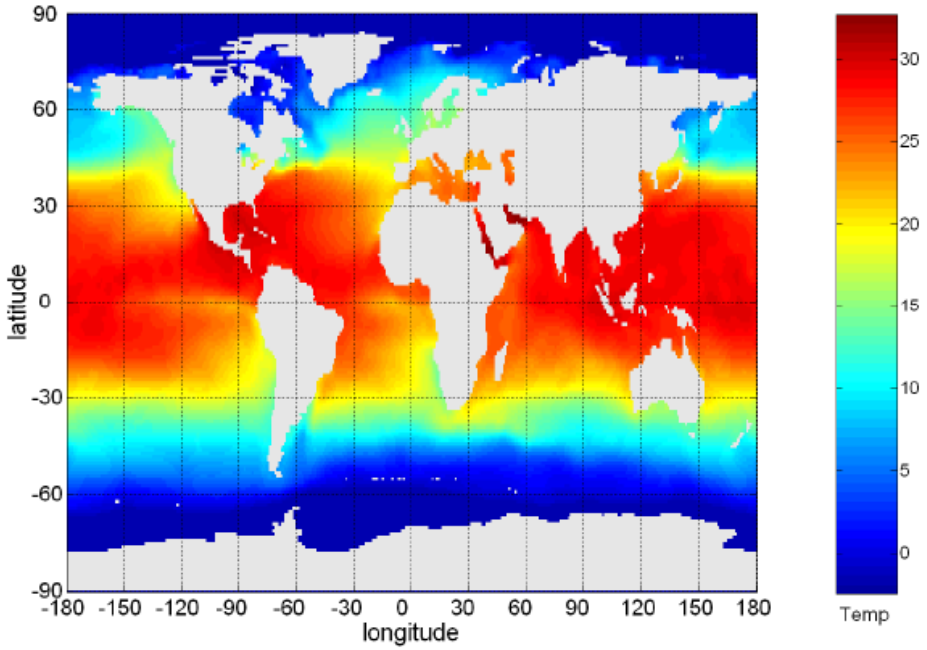
Data Visualization

> What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.
>
> Herb Simon, Scientific American, 1995

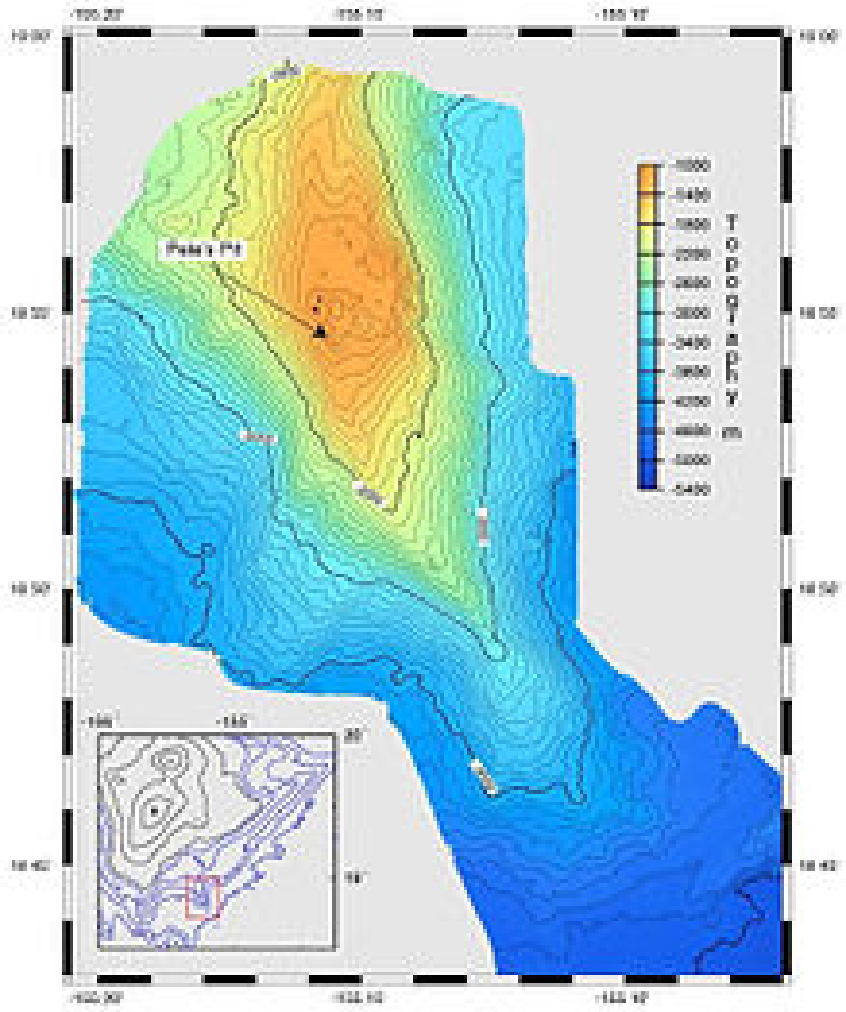
- Data volume and velocity is much higher than our ability to draw knowledge from it
- Visualization helps draw knowledge form data (beyond statistical inference)
- Visualization reveal information that statistics may not
- Visualization of scientific data magnifies the capabilities of science to understand our universe

Example: Sea Surface Temperature

- The following shows the Sea Surface Temperature (SST) for July 1982
  - Tens of thousands of data points are summarized in a single figure



Data Visualization: Why?



- Bathymetric map
- Nautical Chart
- Reveal Hidden dangers
- Help marine navigation

Data Visualization: Why?

- Popular belief in 1850: *cholera spreads via airborne transmission*
- Dr. John Snow plotted each death on a London map
- Noticed clusters around a certain pump
- Turned out that well was contaminated with sewage

Data Visualization: Why?

- Popular belief in 1850: *cholera spreads via airborne transmission*
- Dr. John Snow plotted each death on a London map
- Noticed clusters around a certain pump
- Turned out that the well was contaminated with sewage

images source: https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map

Data Visualization: Why?

States mean income and fraction of college degree holders

Table - StateData ()

Load Snap

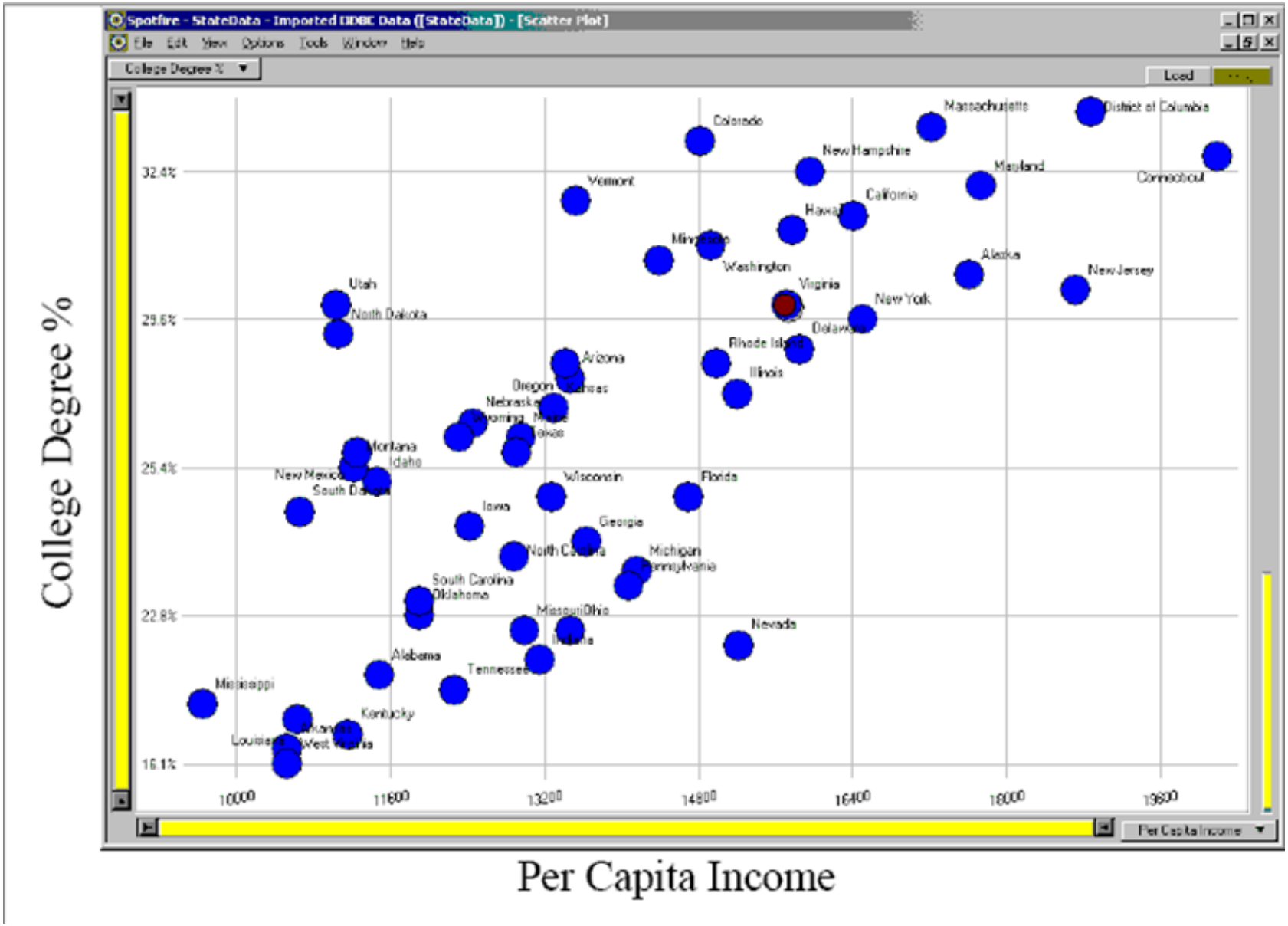
State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311

source: Bradley Hemminger, Uni. of North Carolina

- Which state has the largest and the smallest —?
- Which states are outliers if any?
- How is income related to college degree?

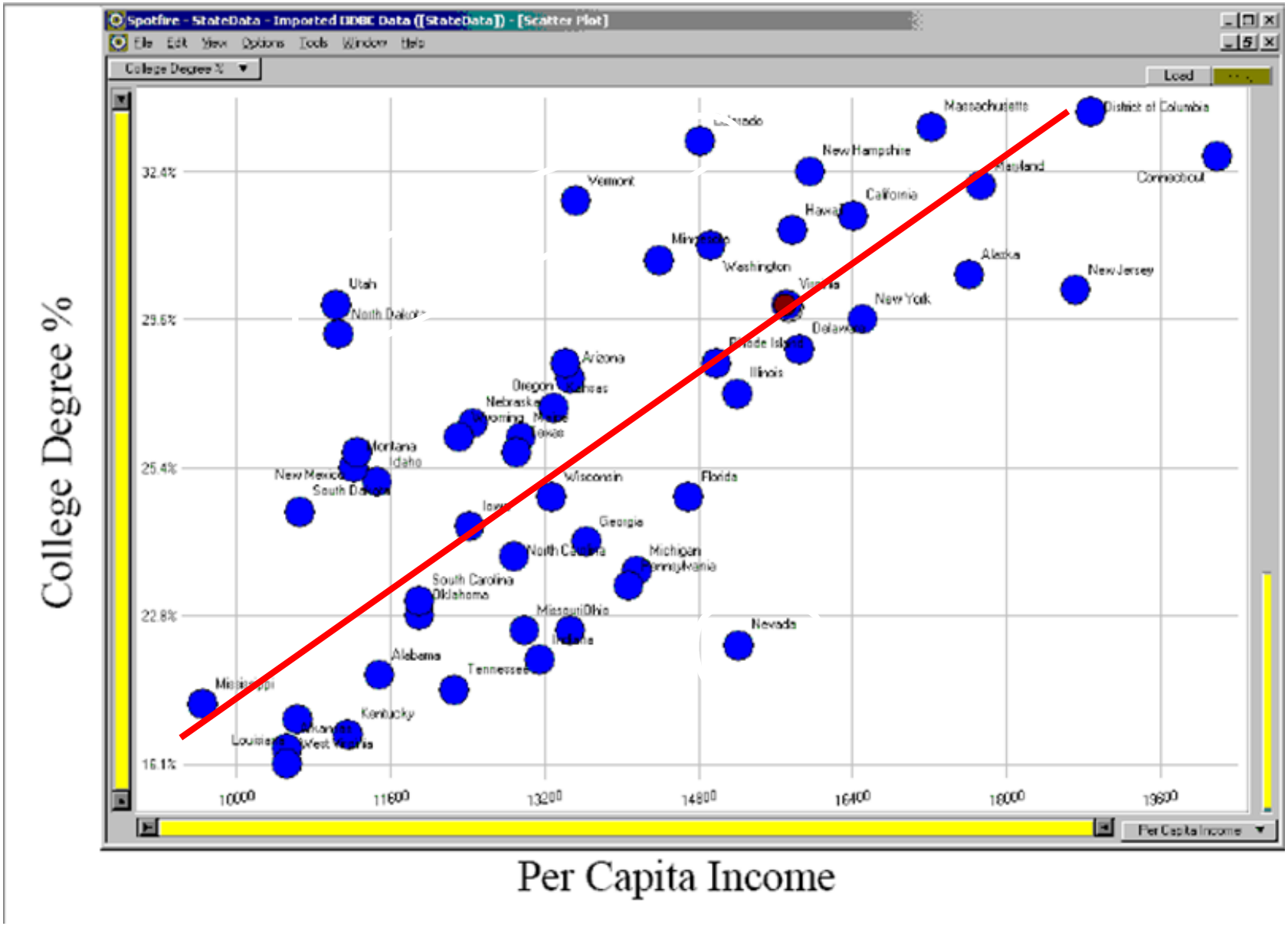
Data Visualization: Why?

Can easily tell what is largest/smallest in every dimension



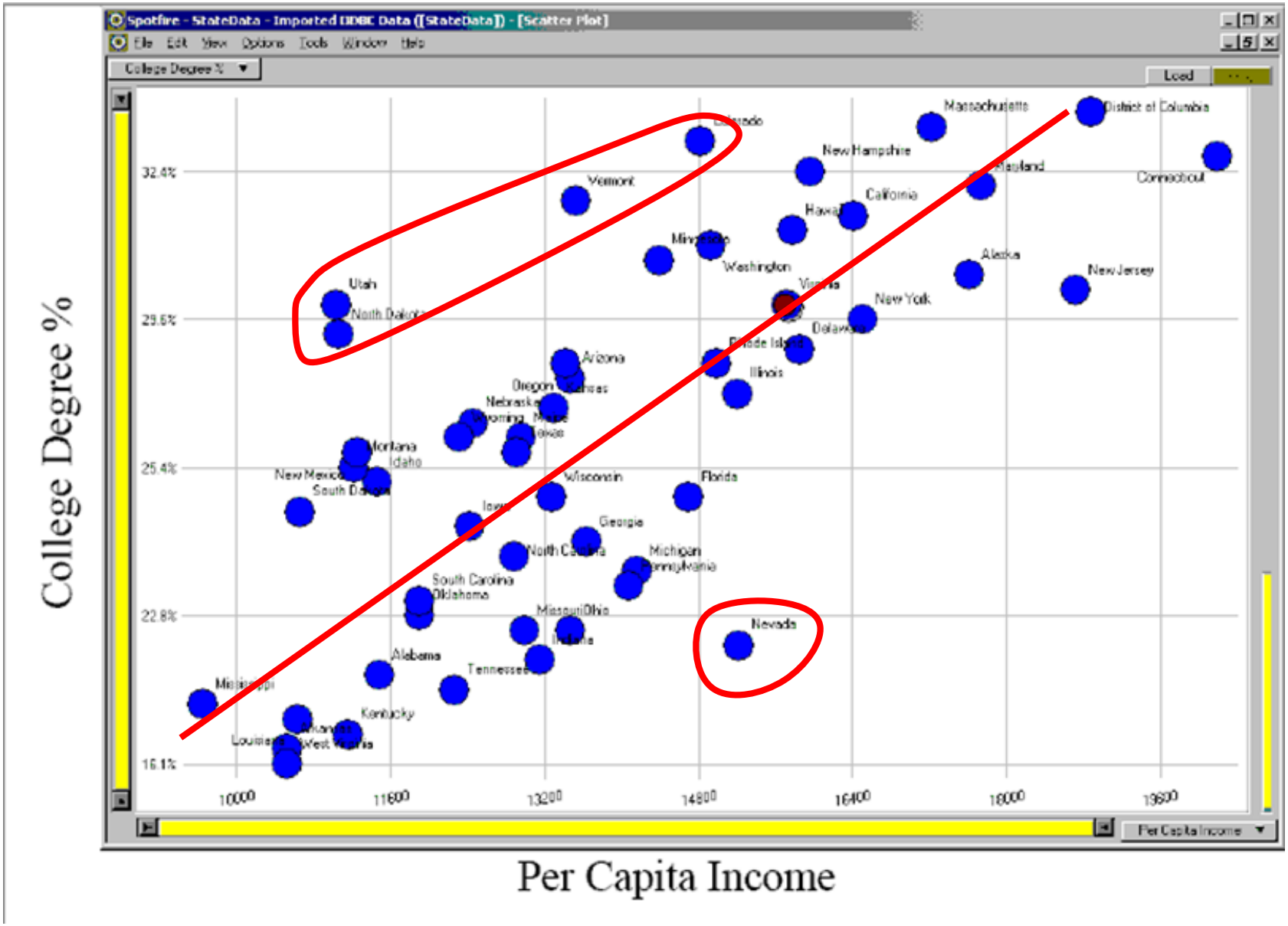
Data Visualization: Why?

Visualization helps to identify relationship easily as compared to raw data



Data Visualization: Why?

Outliers stand out and get identified easily



Data Visualization: Why?

Edward Tufte, *The visual display of quantitative information,*

Anscombe's Quartet: Four datasets with identical statistics

x	4	5	6	7	8	9	10	11	12	13	14
y	4.26	5.68	7.24	4.82	6.95	8.81	8.04	8.33	10.84	7.58	9.96

x	4	5	6	7	8	9	10	11	12	13	14
y	3.1	4.74	6.13	7.26	8.14	8.77	9.14	9.26	9.13	8.74	8.1

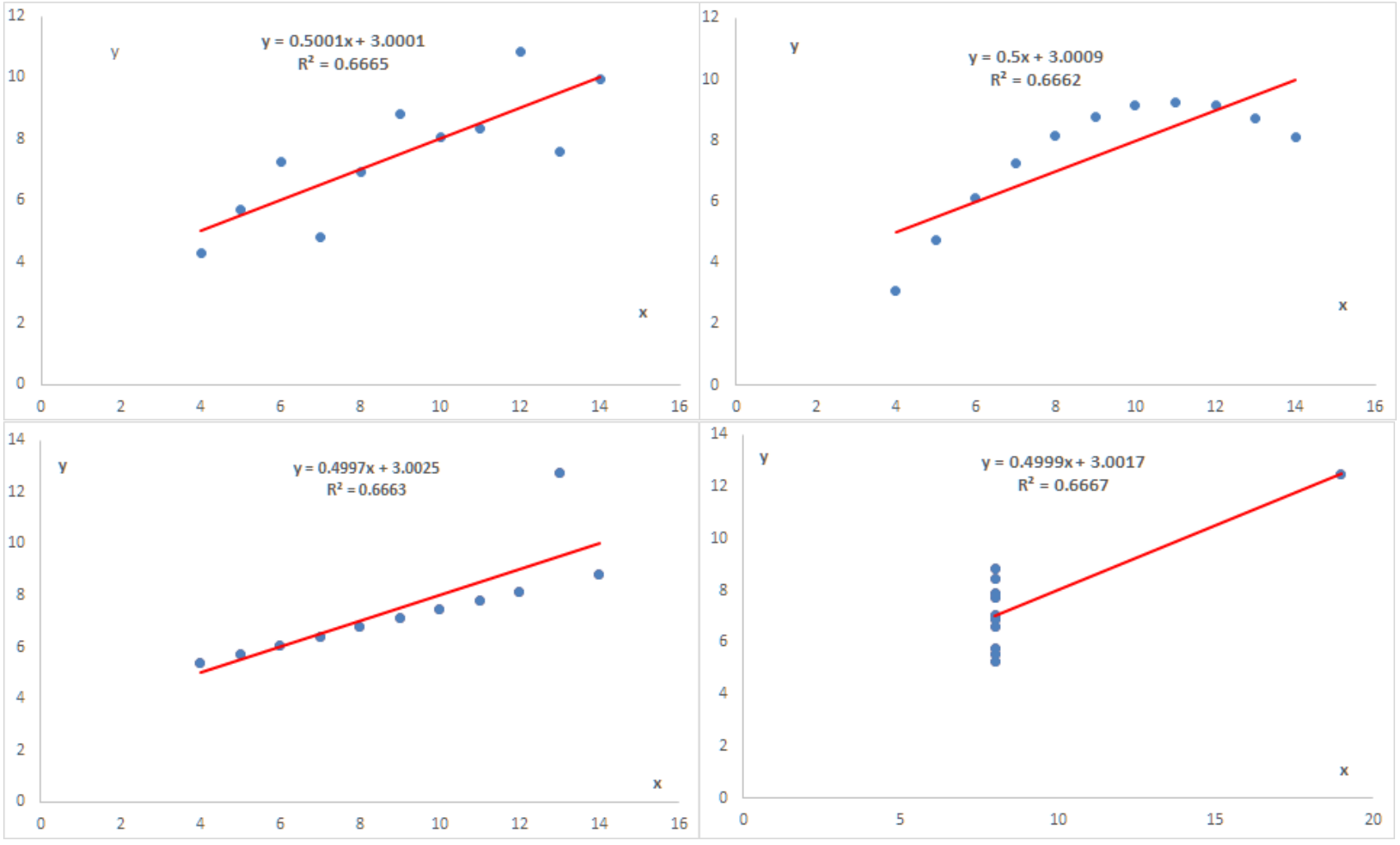
x	10	8	13	9	11	14	6	4	12	7	5
y	5.39	5.73	6.08	6.42	6.77	7.11	7.46	7.81	8.15	12.74	8.84

x	8	8	8	8	8	8	8	8	8	8	19
y	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.5

$\mu_x = 9$ $\sigma_x = 3.316$ $\mu_y = 7.500$ $\sigma_y = 2.031$

Data Visualization: Why?

Anscombe's Quartet: 4 datasets with identical regression line



Data Visualization: Why?

> 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

- Visual system is the highest bandwidth channel to the brain
- 70% of body's sense receptors reside in our eyes
- Metaphors to describe understanding often refer to vision ("I see," "insight," "illumination") Thinking with our Eyes
- Need an efficient way to understand Big Data

Data Visualization: Why?

- Reveals invisible parts that we don't have access to otherwise
- Capture Events
- Analyze things that are otherwise difficult
- Help us tell a story
- See things at a level that is not available at our own perception
- Magnifies our ability to understand things better

Visual Perception

- Important to understand how visual perception works in order to effectively design visualizations
- Understanding mechanisms of the visual processing system and using that knowledge can result in improved displays
- Having some idea of human perception and psychology helps in optimal visual mapping and developing effective and meaningful visualization
- Knowing how the brain would read visualization helps design effective visualizations

Eye vs Camera

Camera:

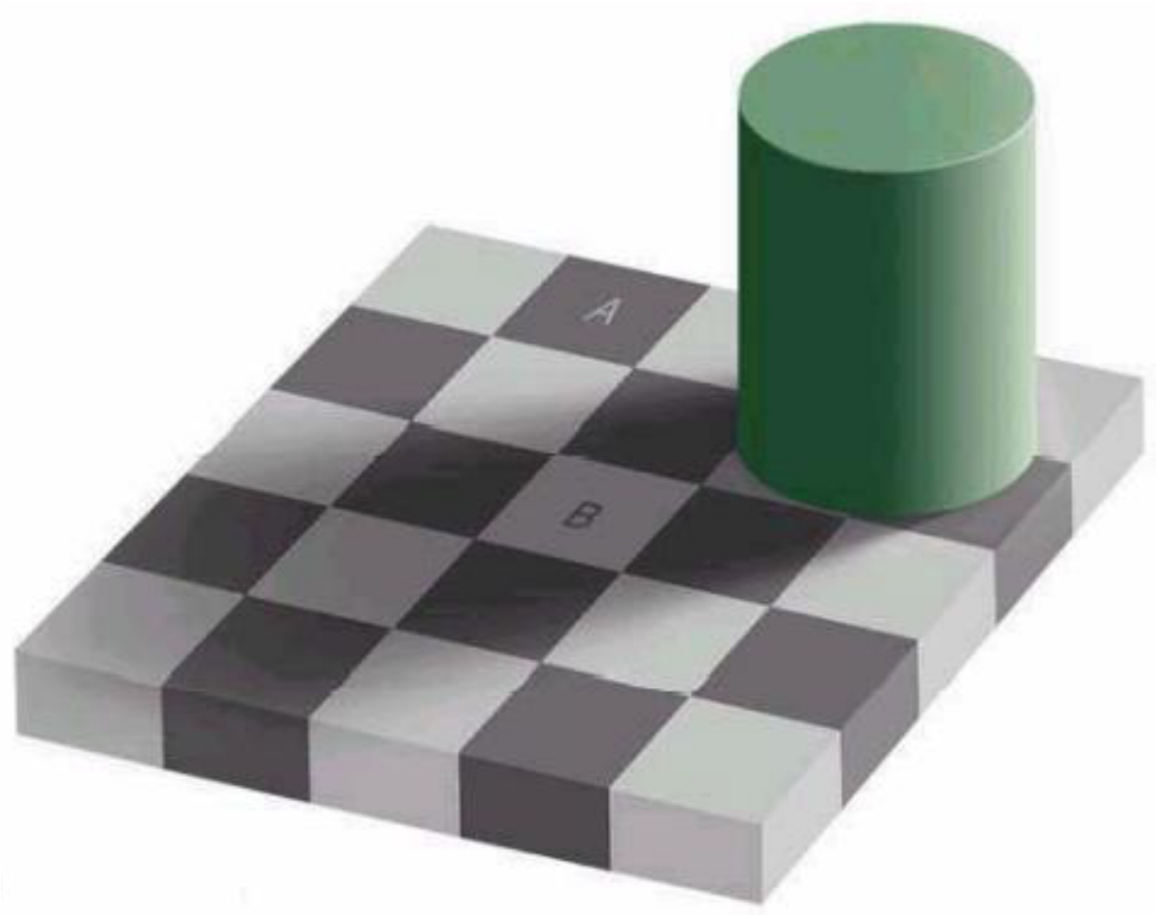
- Good optics
- Single focus, white balance, exposure
- Full image capture

Eye:

- Poor optics
- Constantly scanning (saccades)
- Constantly adjusting focus
- Constantly adapting (white balance, exposure)
- Mental reconstruction of image (sort of)

Eye vs Camera

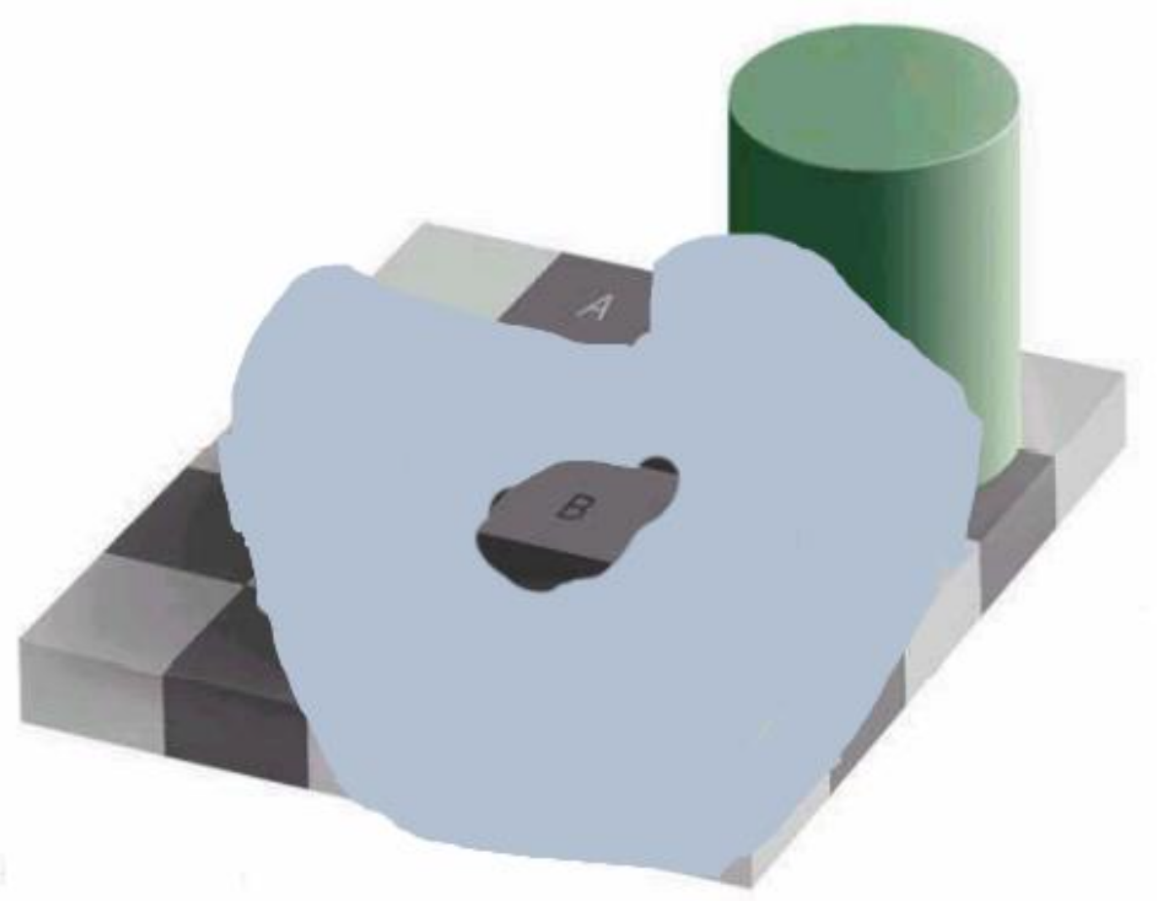
Visual Perception is not just camera work



Which square between A and B is darker?

Eye vs Camera

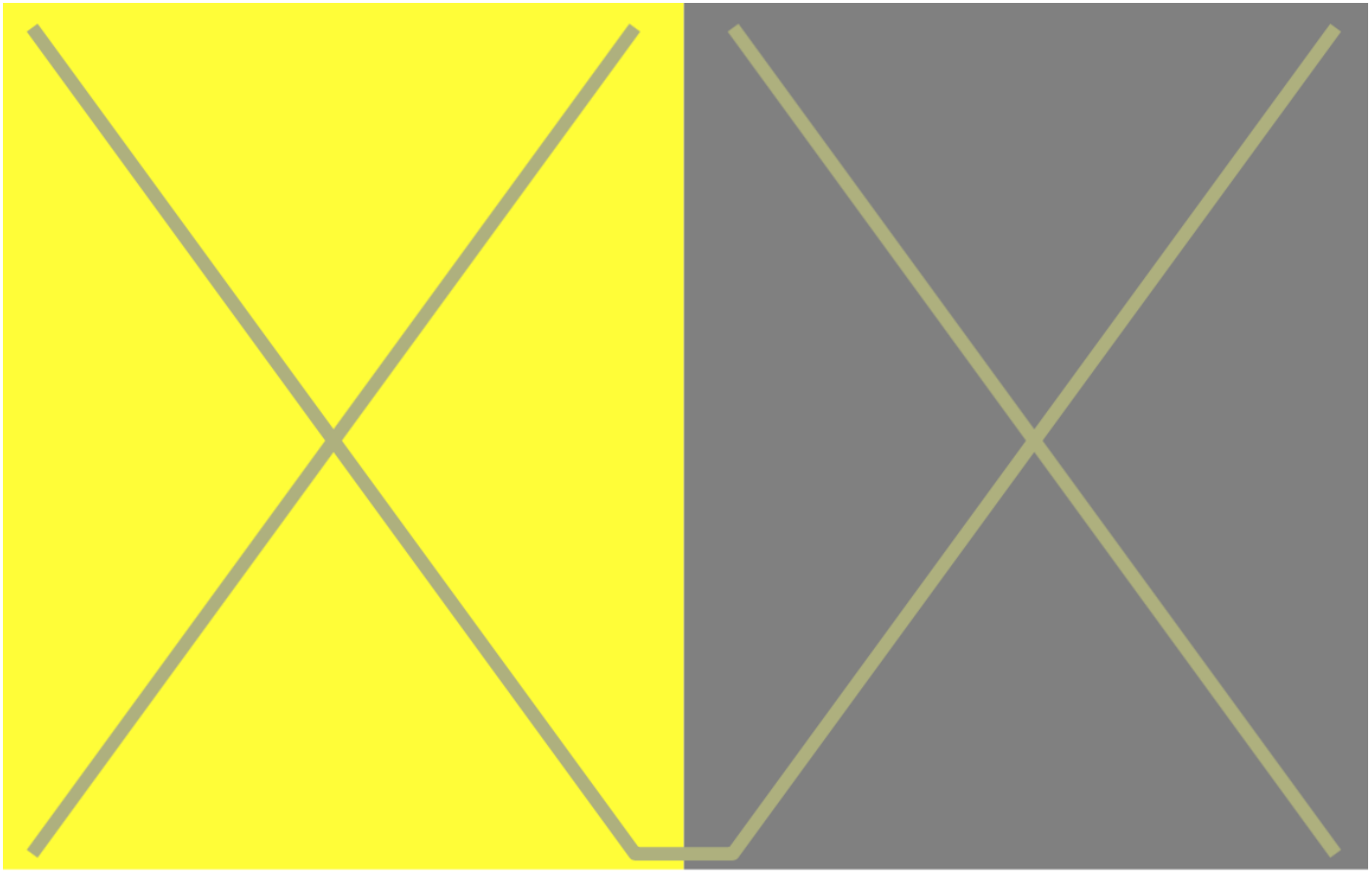
Visual Perception is not just camera work



Both have the same darkness!

Eye vs Camera

Visual Perception is not just camera work



Color is relative

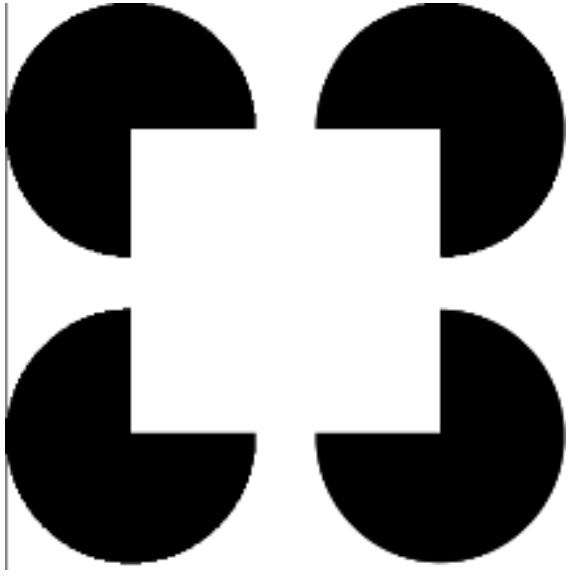
Visual Perception

Gestalt Psychology

> The human mind considers objects in their entirety before, or in parallel with, perception of their individual parts; suggesting the whole is other than the sum of its parts.
>
> Theory of Perception - wikipedia

Visual Perception: Gestalt Principles

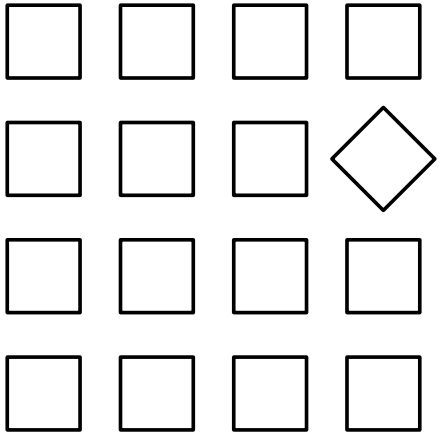
Similarity: The mind perceives similar shapes in a relationship and bring them together to form larger shapes



How many circles and squares are there?

Visual Perception: Gestalt Principles

Anomaly: The mind is very good at identifying outliers



Which piece stands out?

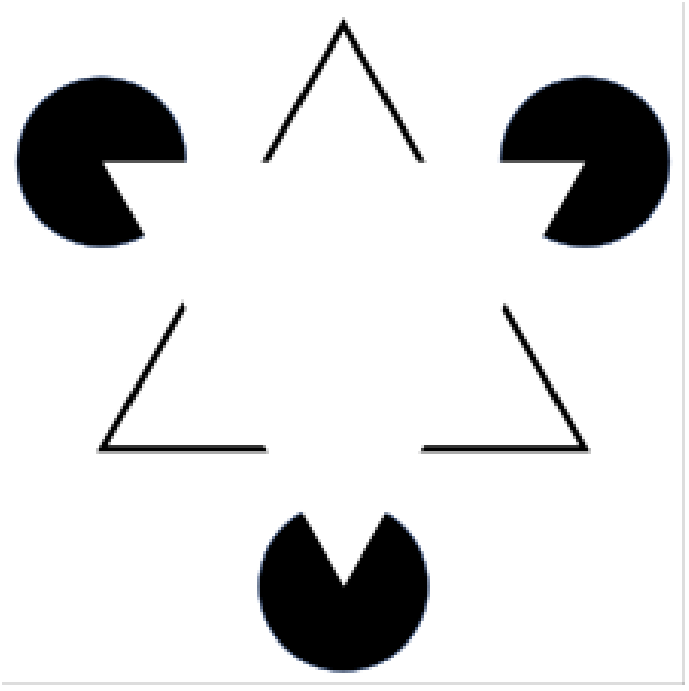
Visual Perception: Gestalt Principles

Continuation: The mind finds meaning in continuation in shapes that are next to each other

Did the leaf come out of the H, Did the lion scare the birds?

Visual Perception: Gestalt Principles

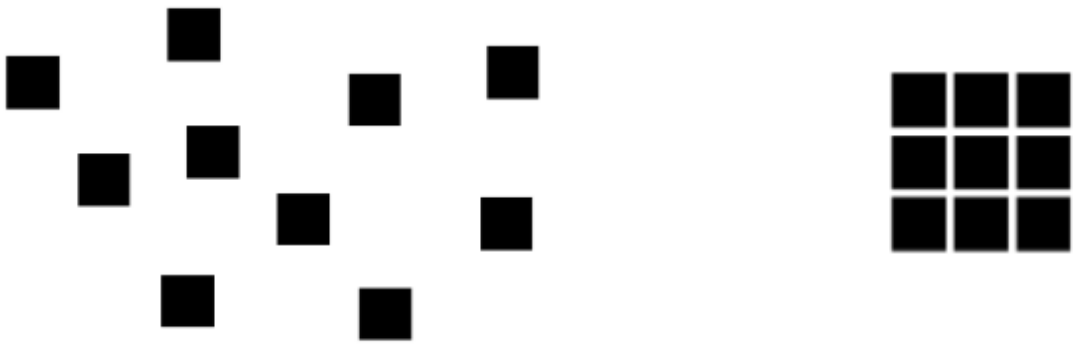
Closure: The mind makes shapes contiguous



How many triangles? Where is the top of the panda?

Visual Perception: Gestalt Principles

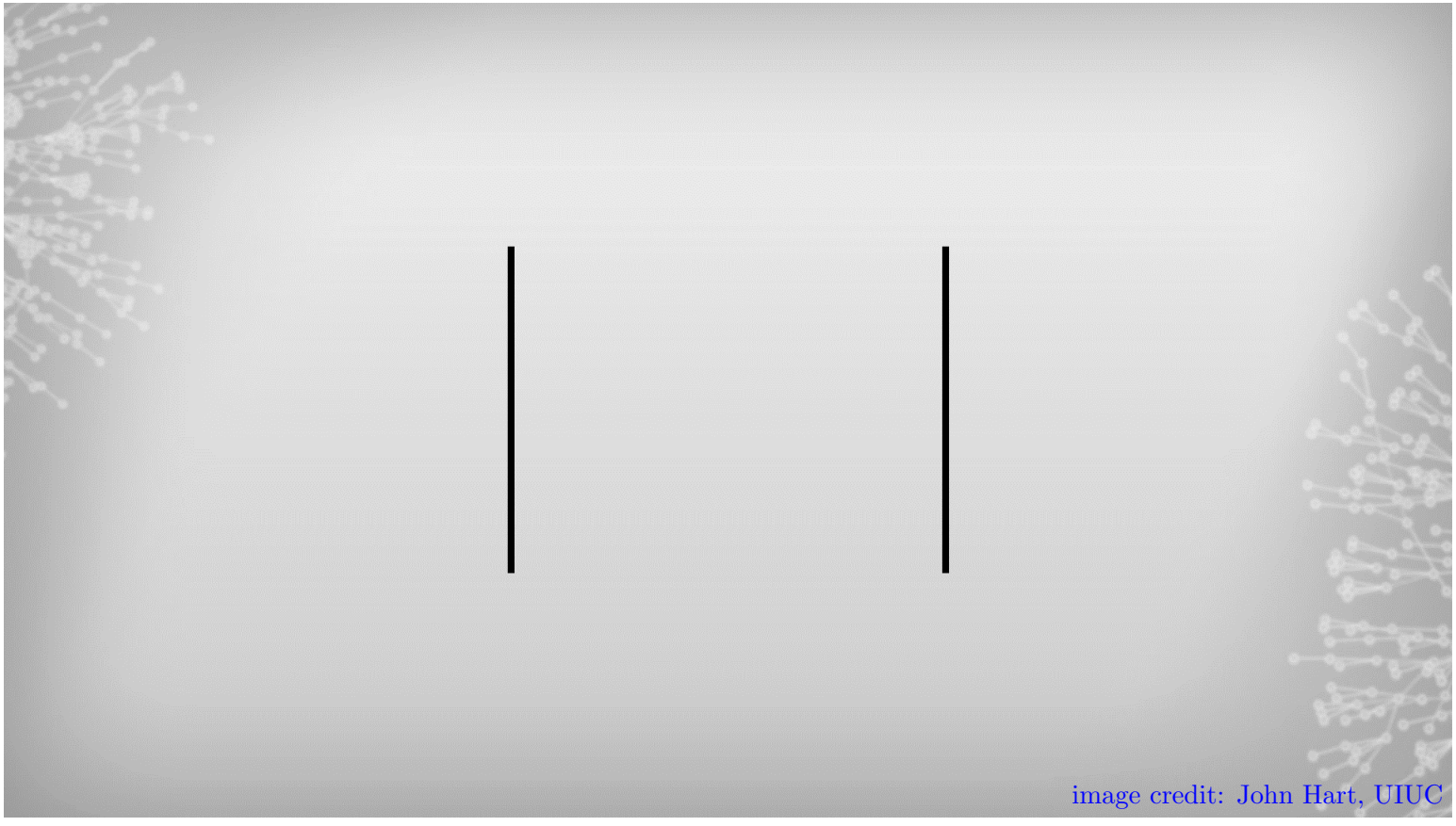
Proximity: The mind perceive closer things as related



Is there any square?

Visual Perception: Context

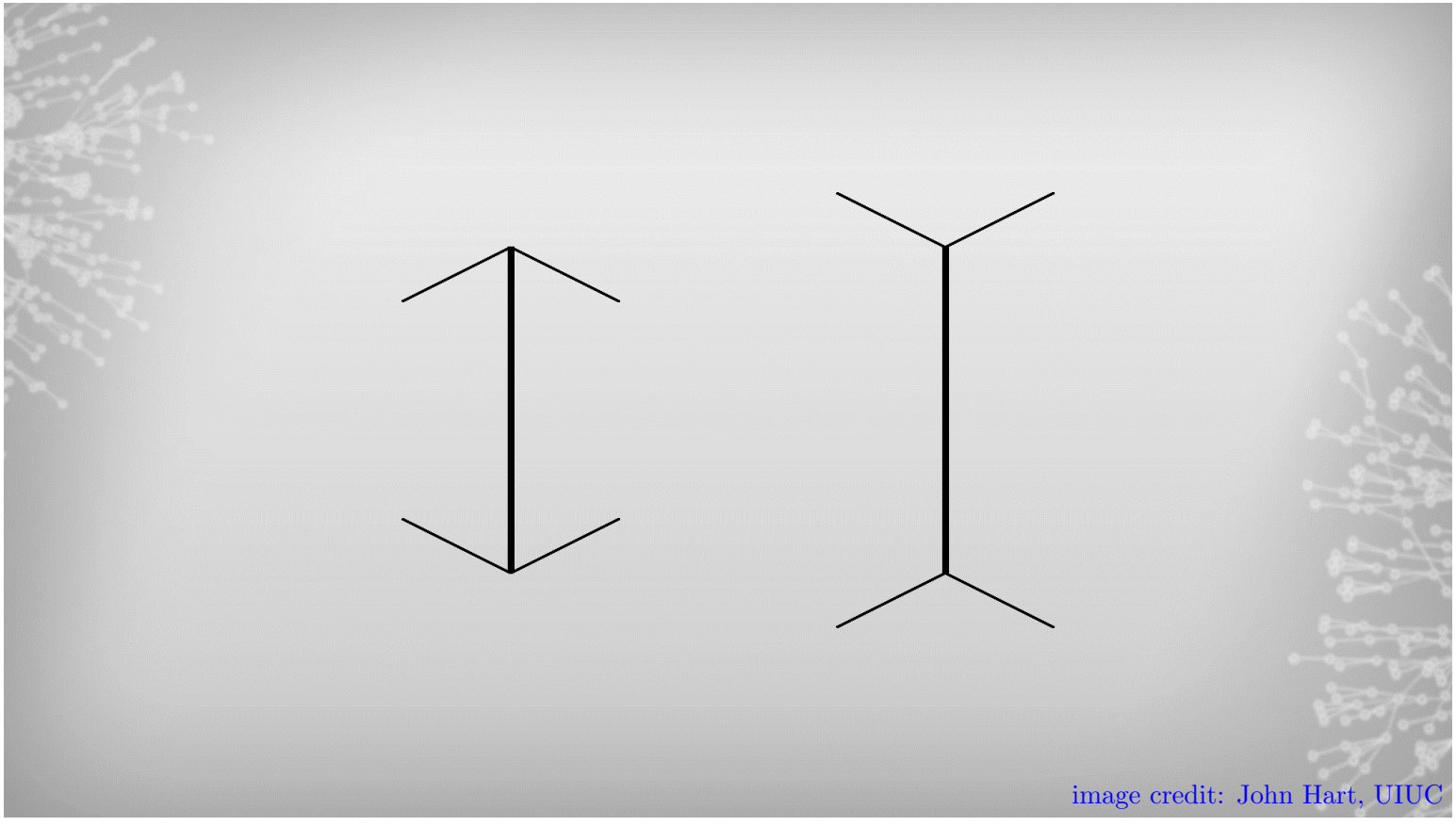
Context can change the appearance of same object



Both lines are equal?

Visual Perception: Context

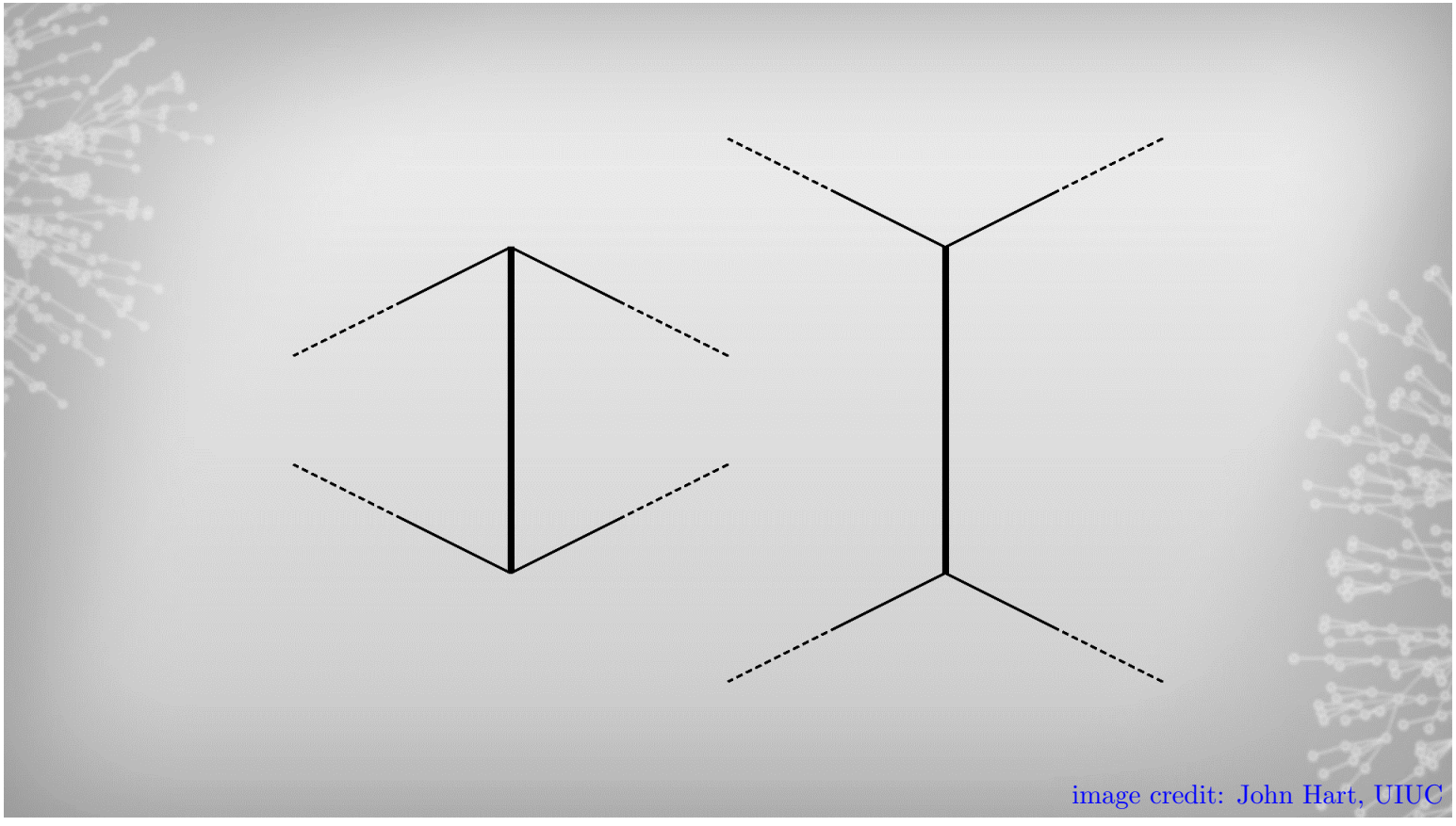
Context can change the appearance of same object



Which line looks longer?

Visual Perception: Context

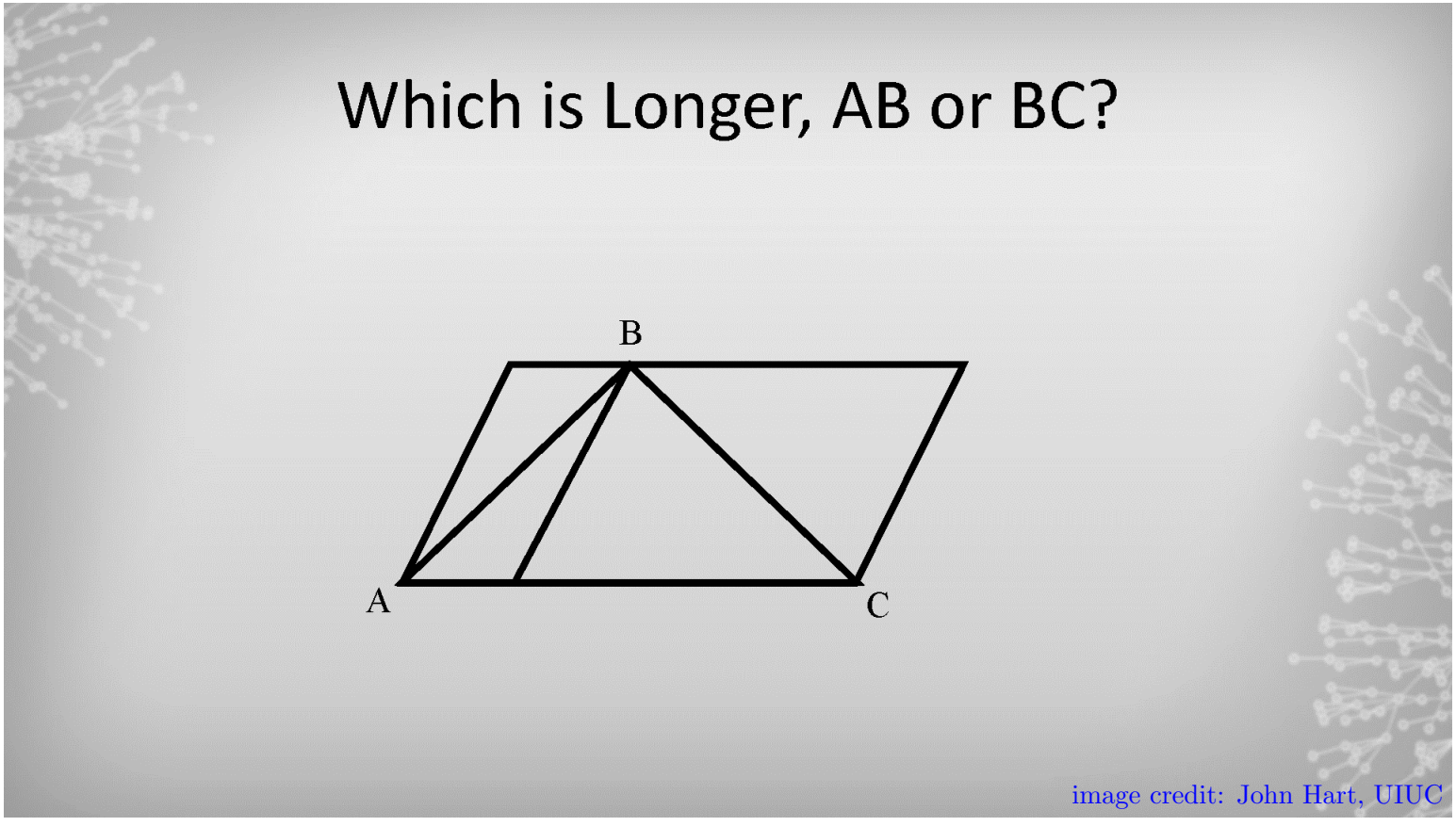
Context can change the appearance of same object



Is the difference more significant?

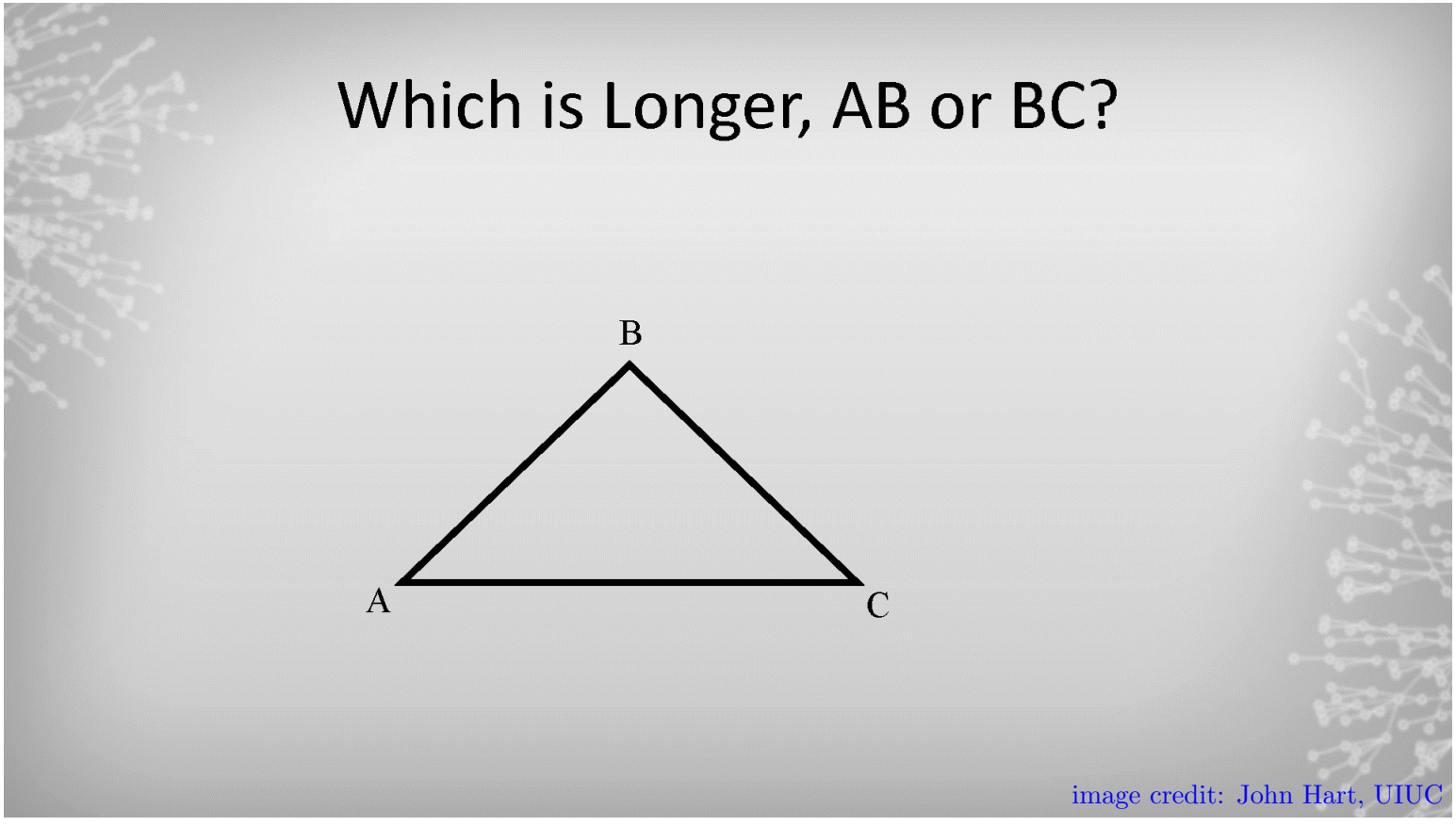
Visual Perception: Context

Context can change the appearance of same object



Visual Perception: Context

Context can change the appearance of same object



Visual Perception: Preattention

Preattention:

- Some visual features are detected immediately
- Pop-out vs. Serial Search
- If recognition takes $200 - 250$ms, then it qualifies as preattentive
- eye movements takes > 200ms, yet some processing can be done quickly
- *If a decision takes a fixed amount regardless of the number of distraction, it is considered to be preattentive.*
- It is important for effective visualization to use better discrimination and avoid misleading viewers

Slided adatped from Cecilia Aragon, Uni. of Washington

Visual Perception: Preattention

How many 5's are there?

$$3857209398237281968372938 27$$
$$3829123583834927301228948 39$$
$$9090201020328937592730914 28$$
$$9383097629658174318692410 24$$

Visual Perception: Preattention

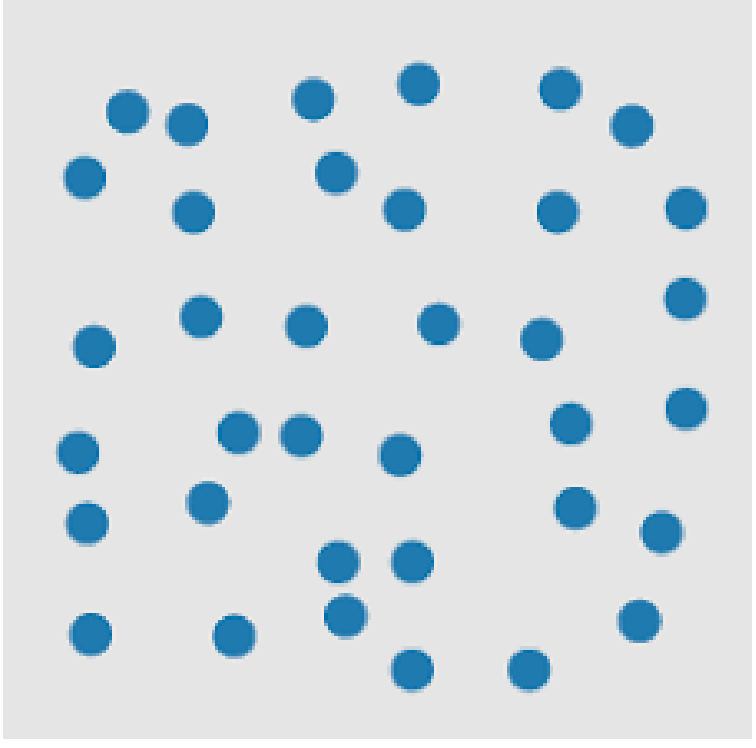
How many 5's are there?

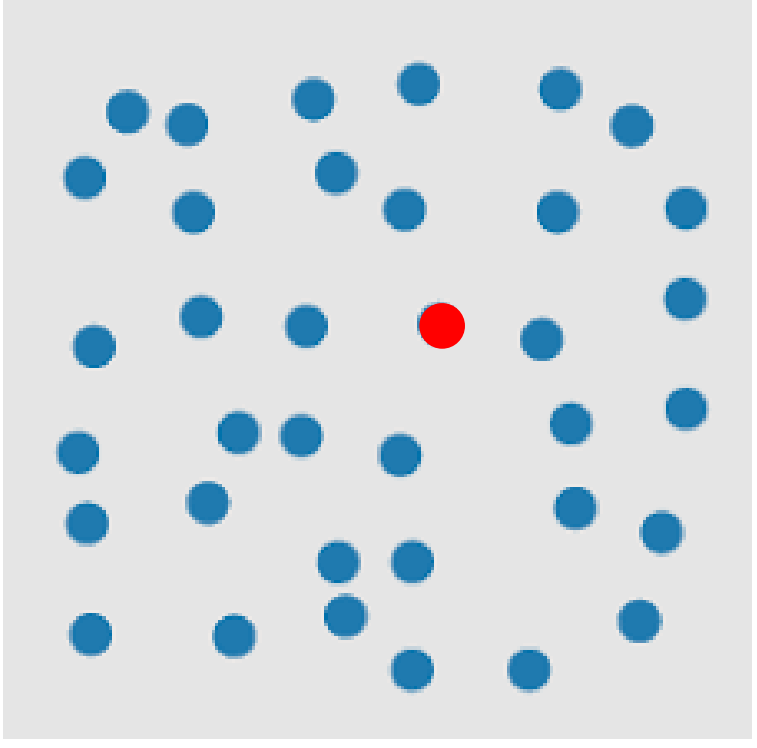
$$385720939823728196837293827$$
$$382912358383492730122894839$$
$$909020102032893759273091428$$
$$938309762965817431869241024$$

Visual Perception: Preattention

Color (hue) is preattentive

Detect red circle among these circles

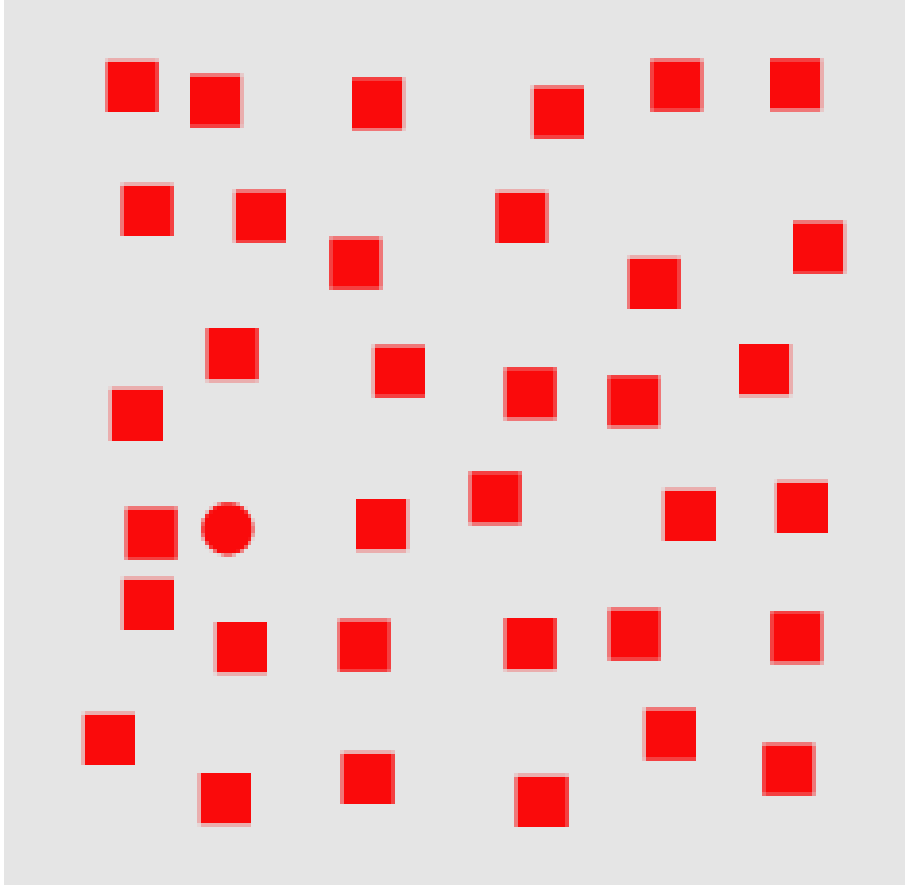




Visual Perception: Preattention

Form (curvature) is (somewhat) preattentive

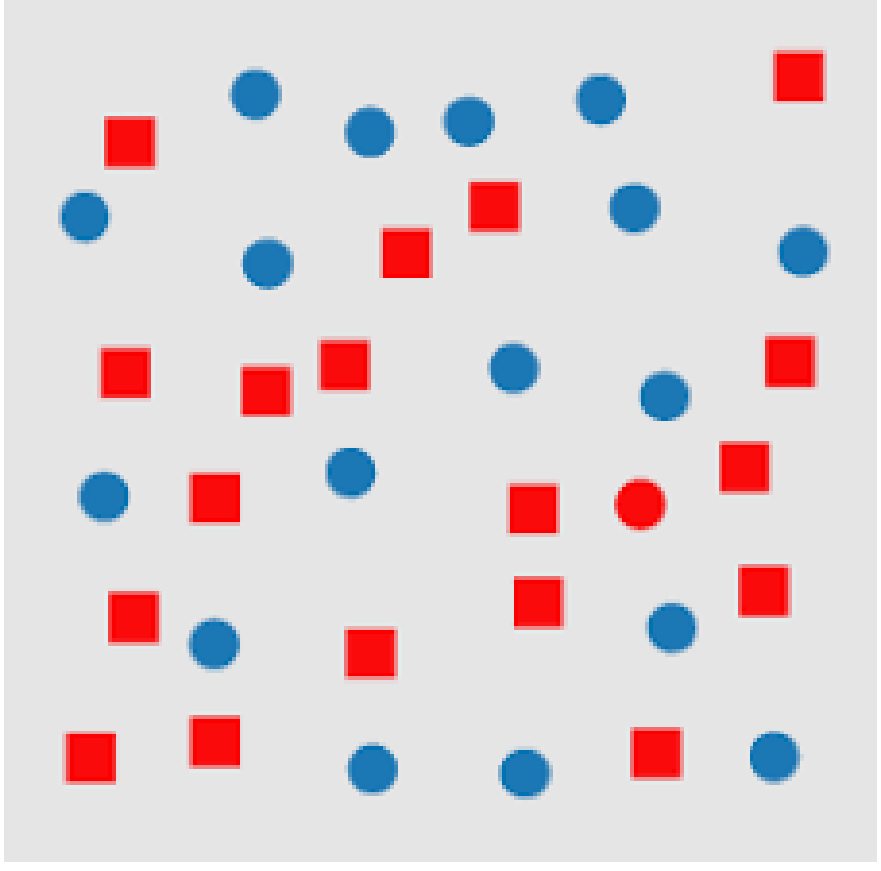
Detect red circle among the following objects



Visual Perception: Preattention

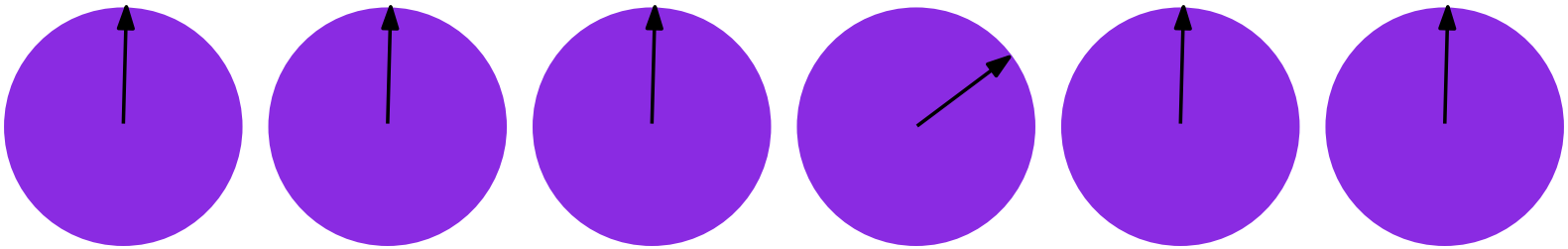
Conjunction of attributes is generally not preattentive

Detect red circle among blue circles and red squares



Visual Perception: Preattention

Detecting slanted line among vertical lines is preattentive

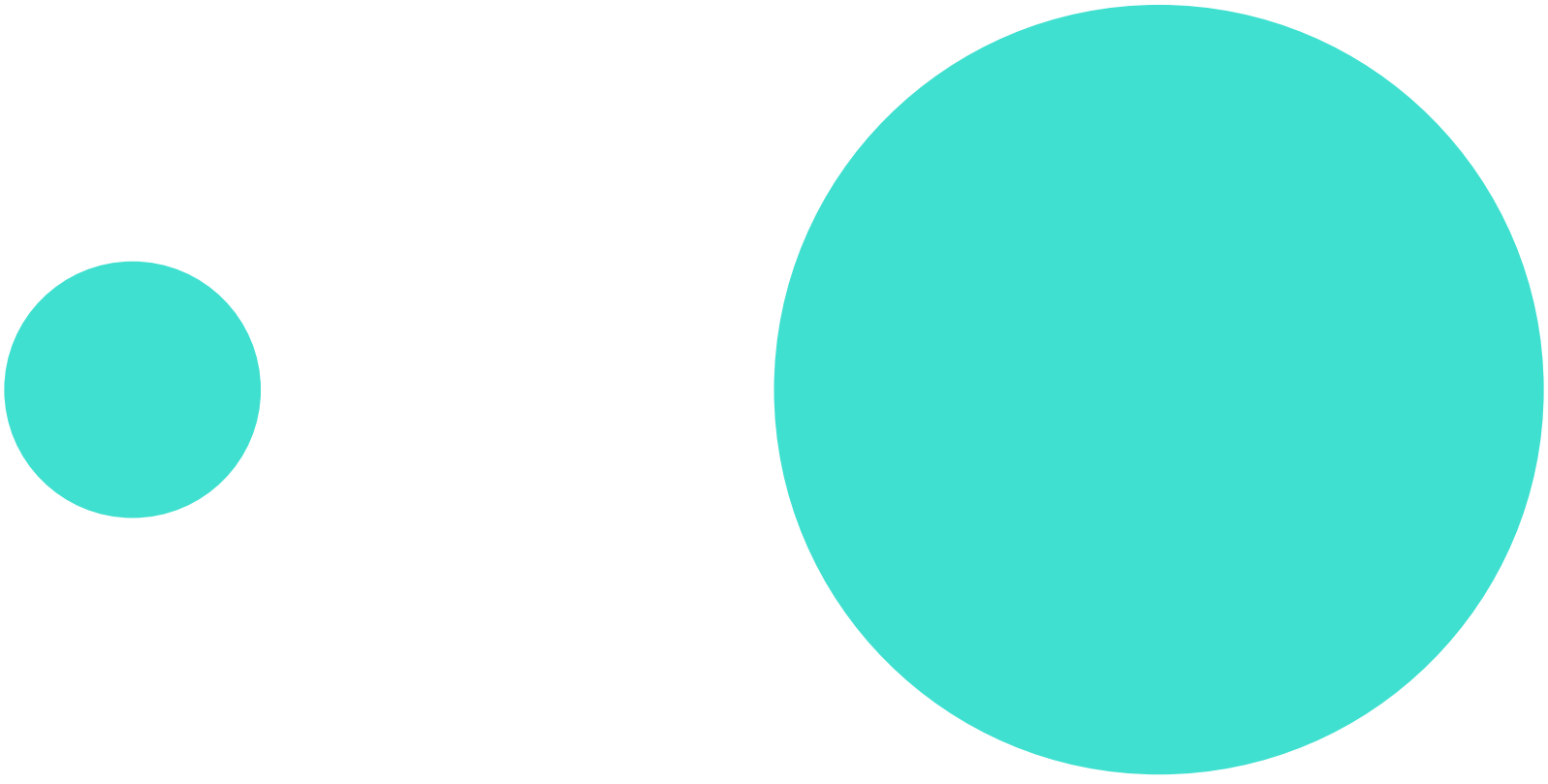


Visual Perception: Selective Attention

- Video of 6 players passing basketballs among themselves
- 3 players wearing black and 3 white shirts
- Count the number of aerial and bounced passes between white shirted players
- You should answer two integers
- `http://viscog.beckman.uiuc.edu/grafs/demos/15.html`

Visual Perception: Magnitude Estimation

How much bigger is the bigger circle?

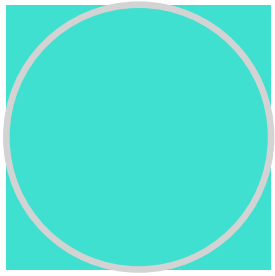


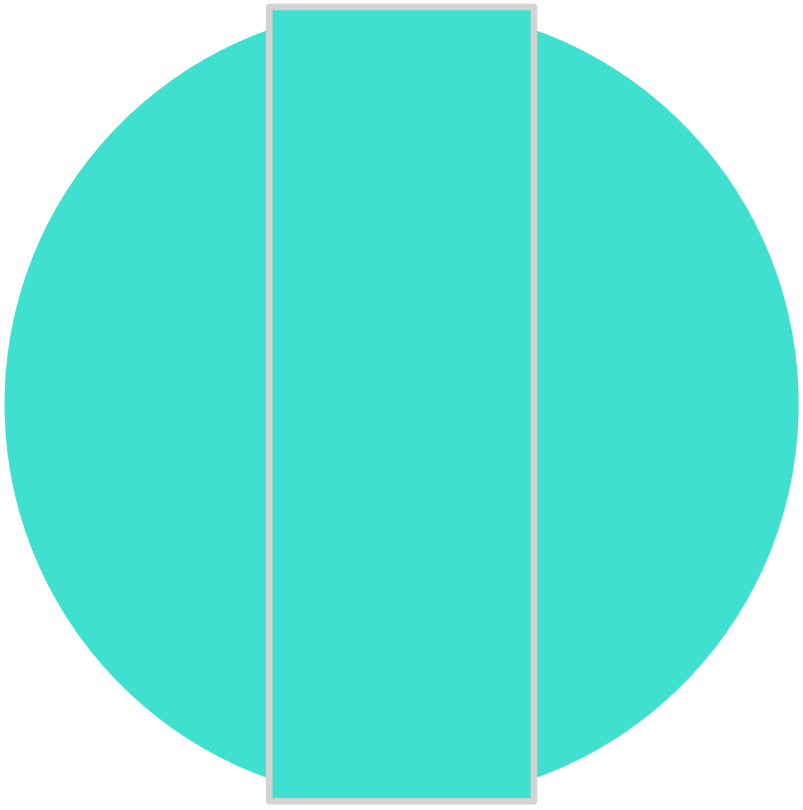
Visual Perception: Magnitude Estimation

How much bigger is the bigger bar?

Visual Perception: Magnitude Estimation

How much bigger?





Visual Perception: Magnitude Estimation

Steven's Power Law

Heuristics for perceptual estimation

- Length is estimated within factors of $[.9 - 1.1]$
- Area is estimated within factors of $[.6 - .9]$
- Volume is estimated within factors of $[.5 - .8]$

Evaluating Visualization

The primary goal of data visualization is to communicate information clearly and efficiently to users via statistical graphics, plots, information graphics tables and charts.

These are the three criteria to evaluate visualizations

- Effectiveness
- Expressiveness
- Consistency
- Integrity

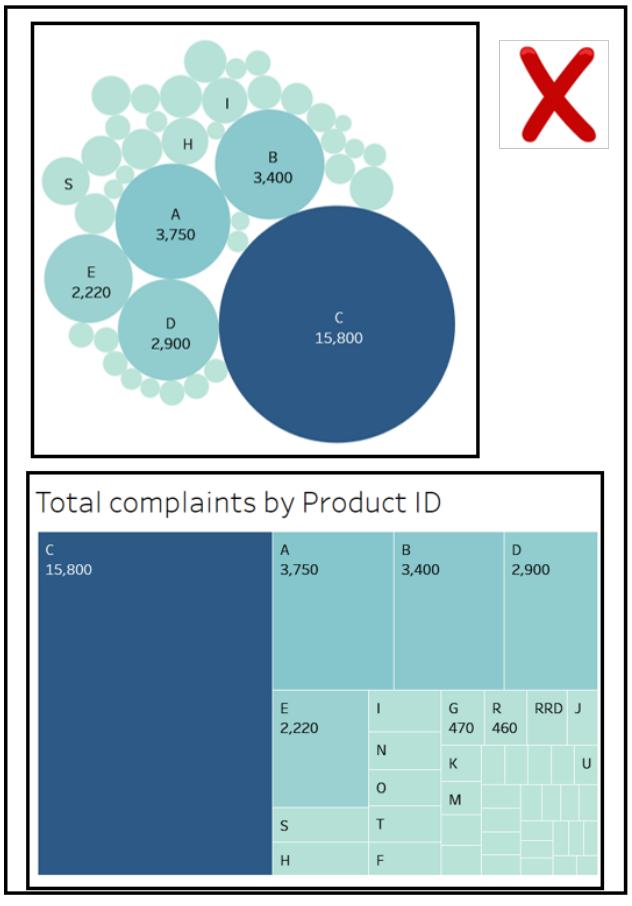
Evaluating Visualization: Effectiveness

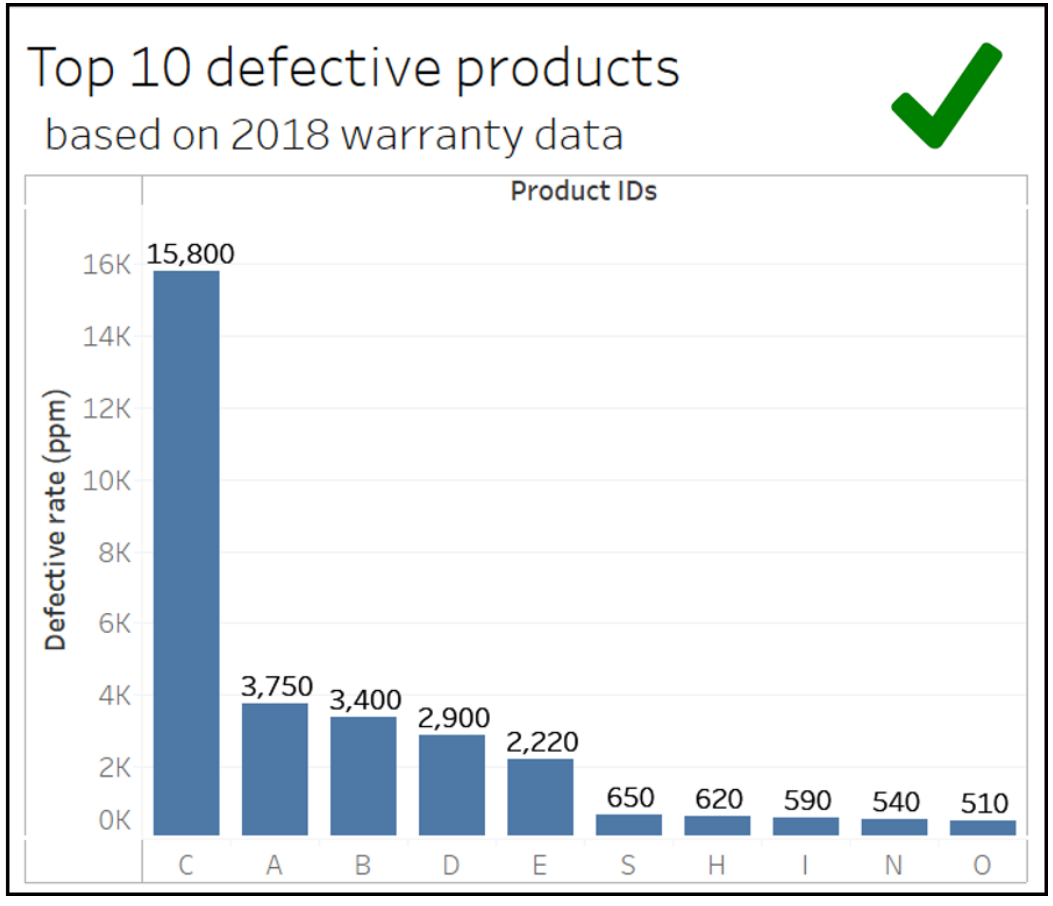
Effectiveness
A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Mackinlay, 1986

Effectiveness - Purpose of Visual

Identify purpose of visual - to compare values, show trends, explore distribution or relationship between variables - choose visual accordingly

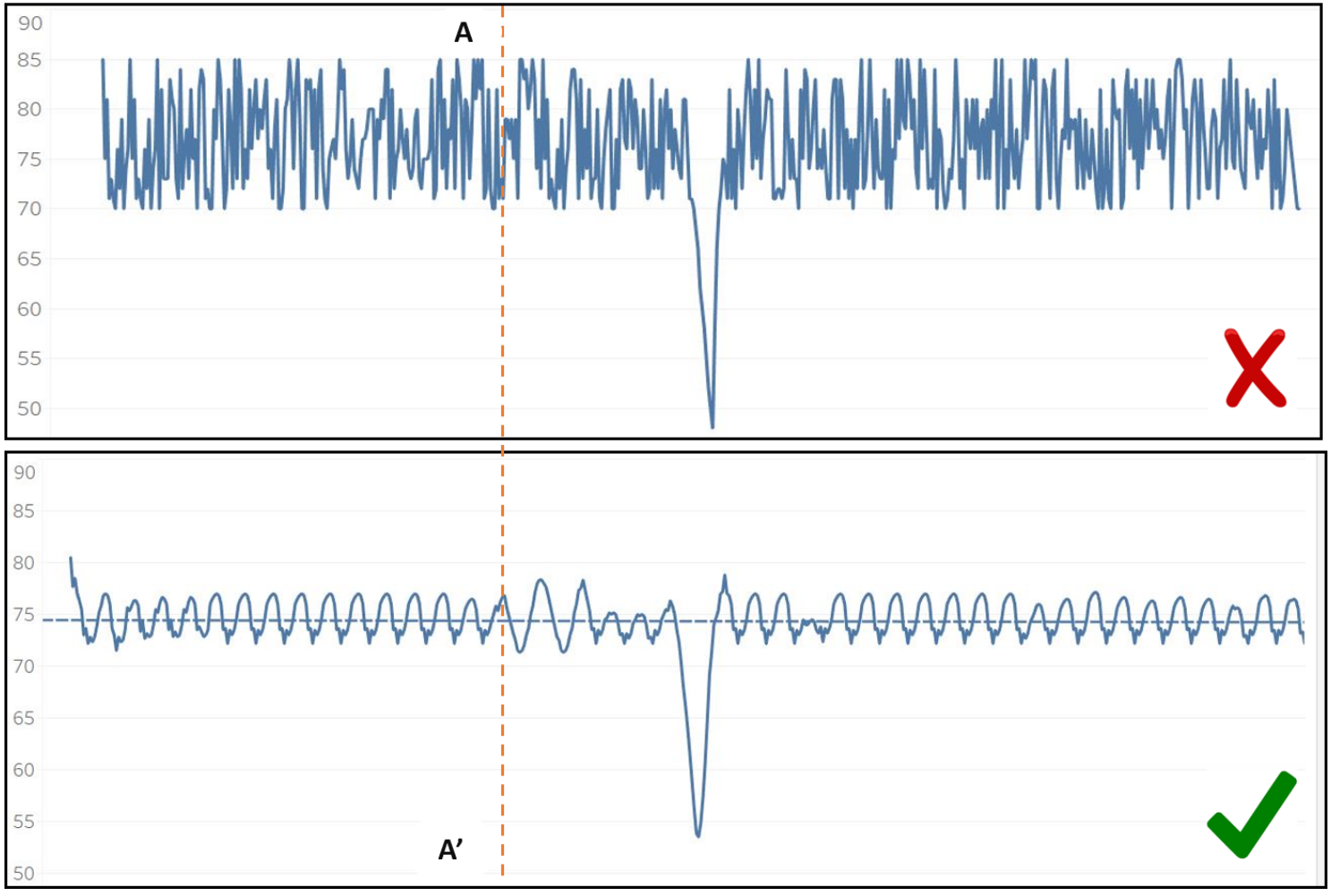




https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

Effectiveness - Focus on Vital Data Points

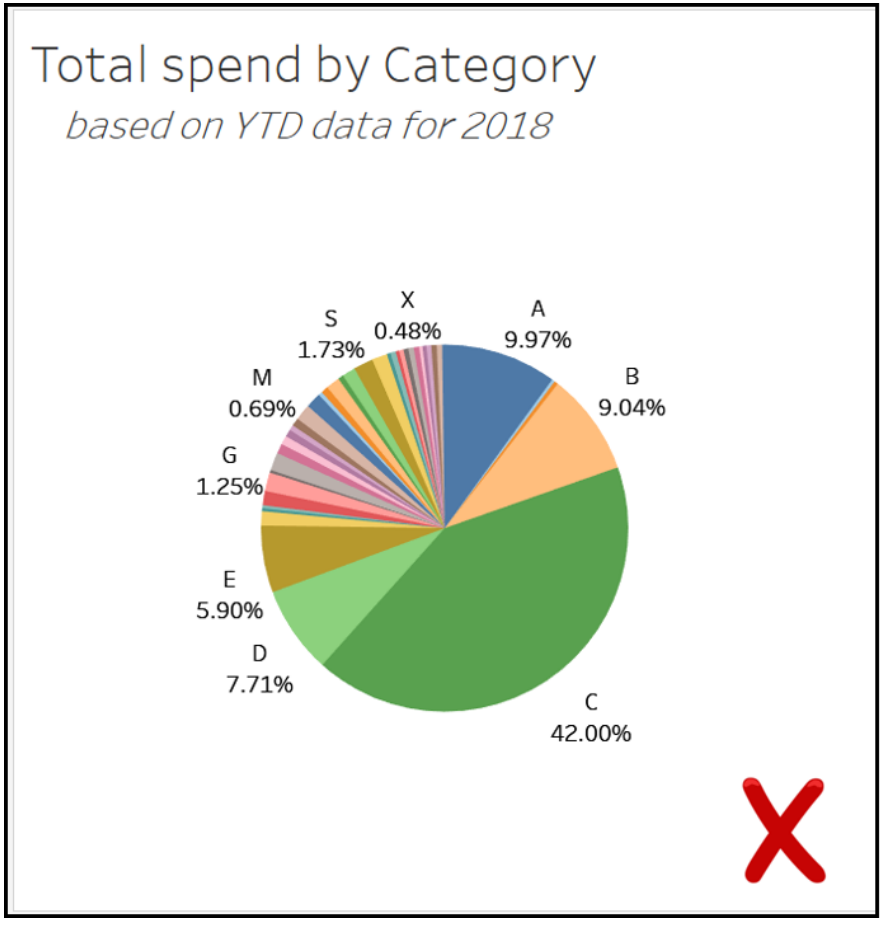
Vital Data Points are few: Which visual gives better insight of sudden dip?

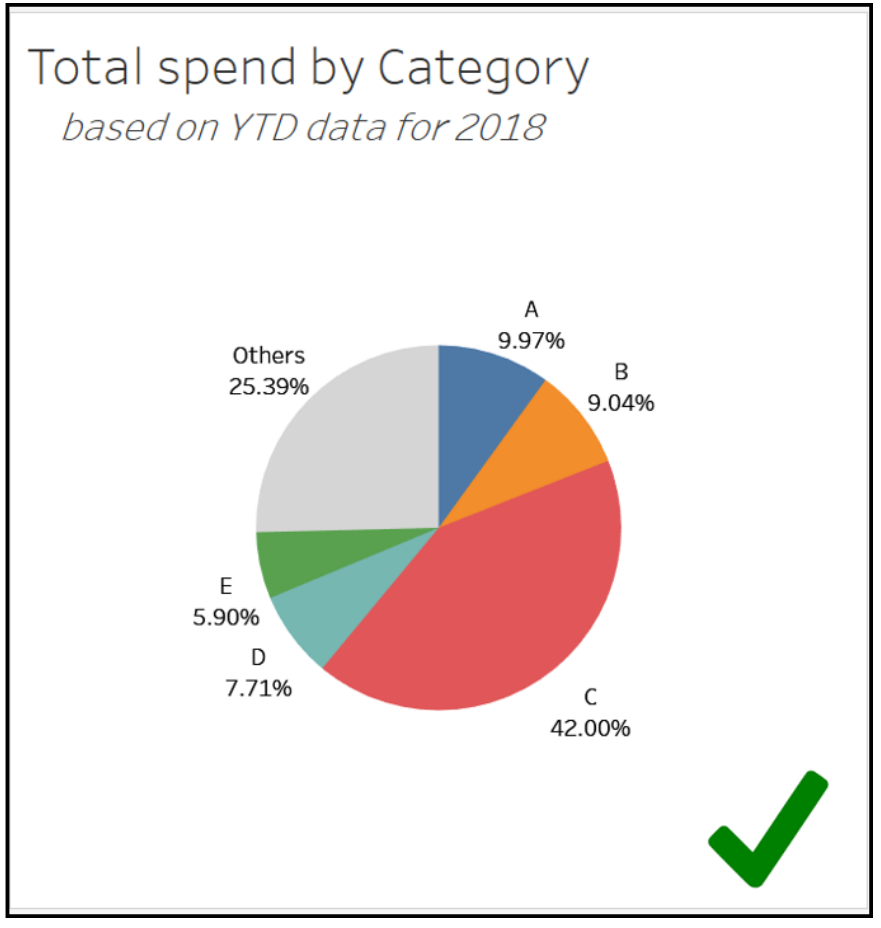


https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

Effectiveness - Suppress the Noise

Make the noise less pronounced

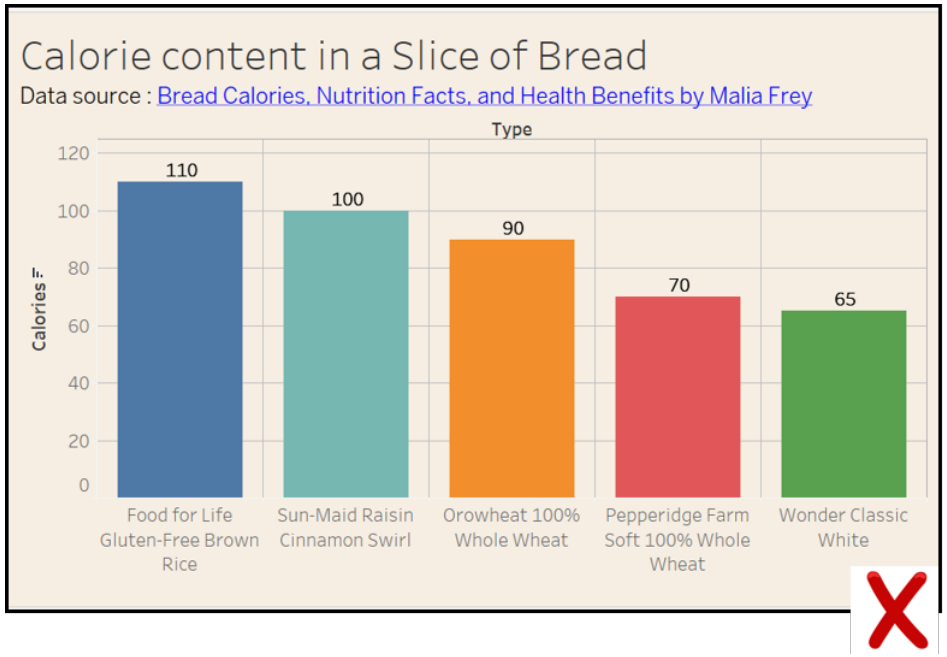


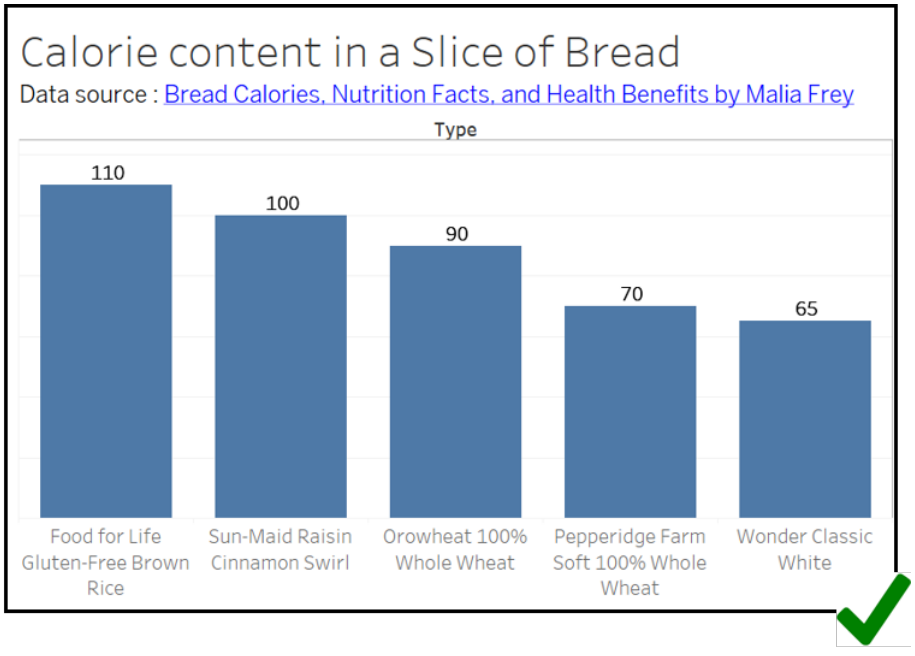


https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

Effectiveness - Use Colors Wisely

Should the same thing be represented with different colors?

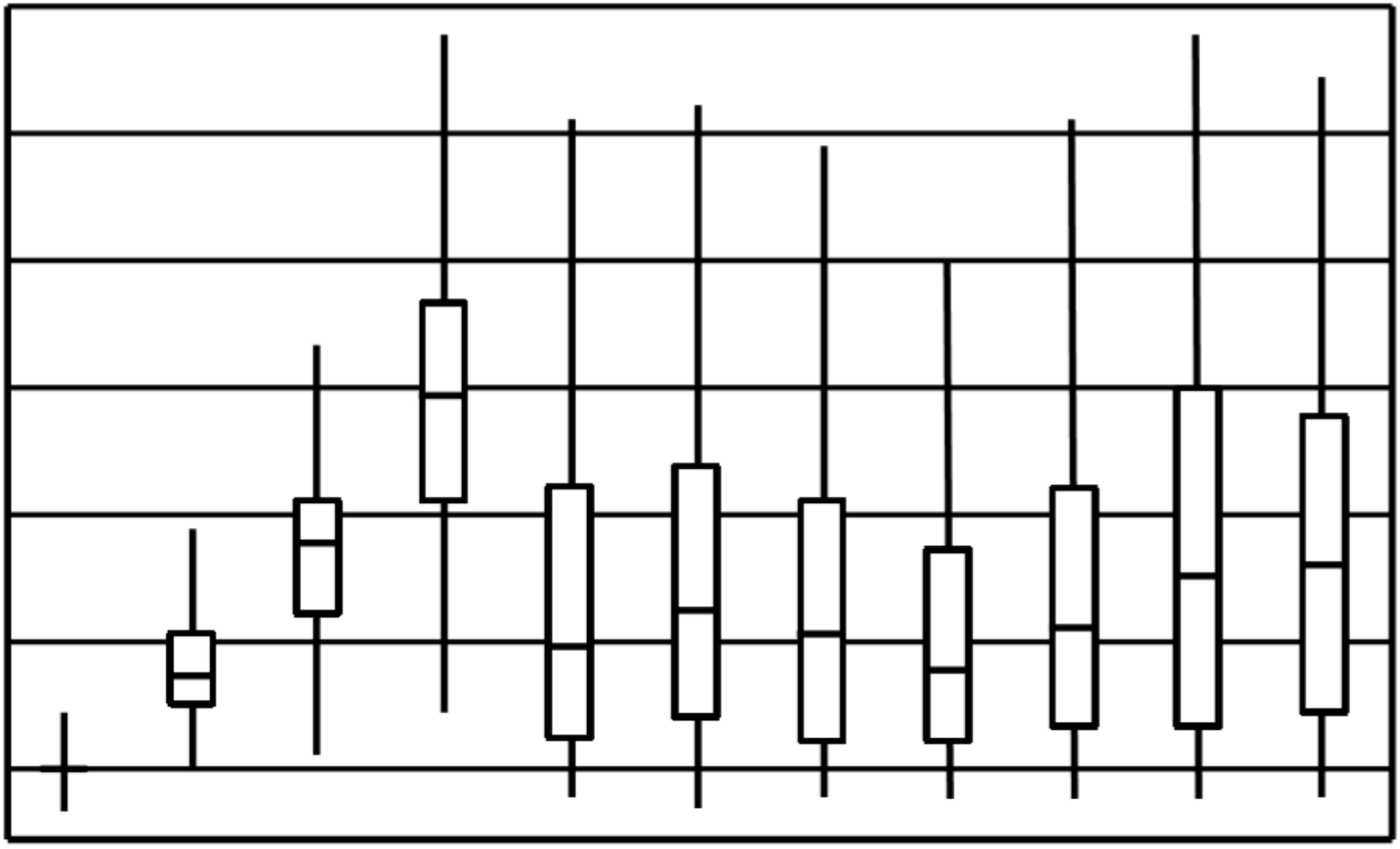




https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

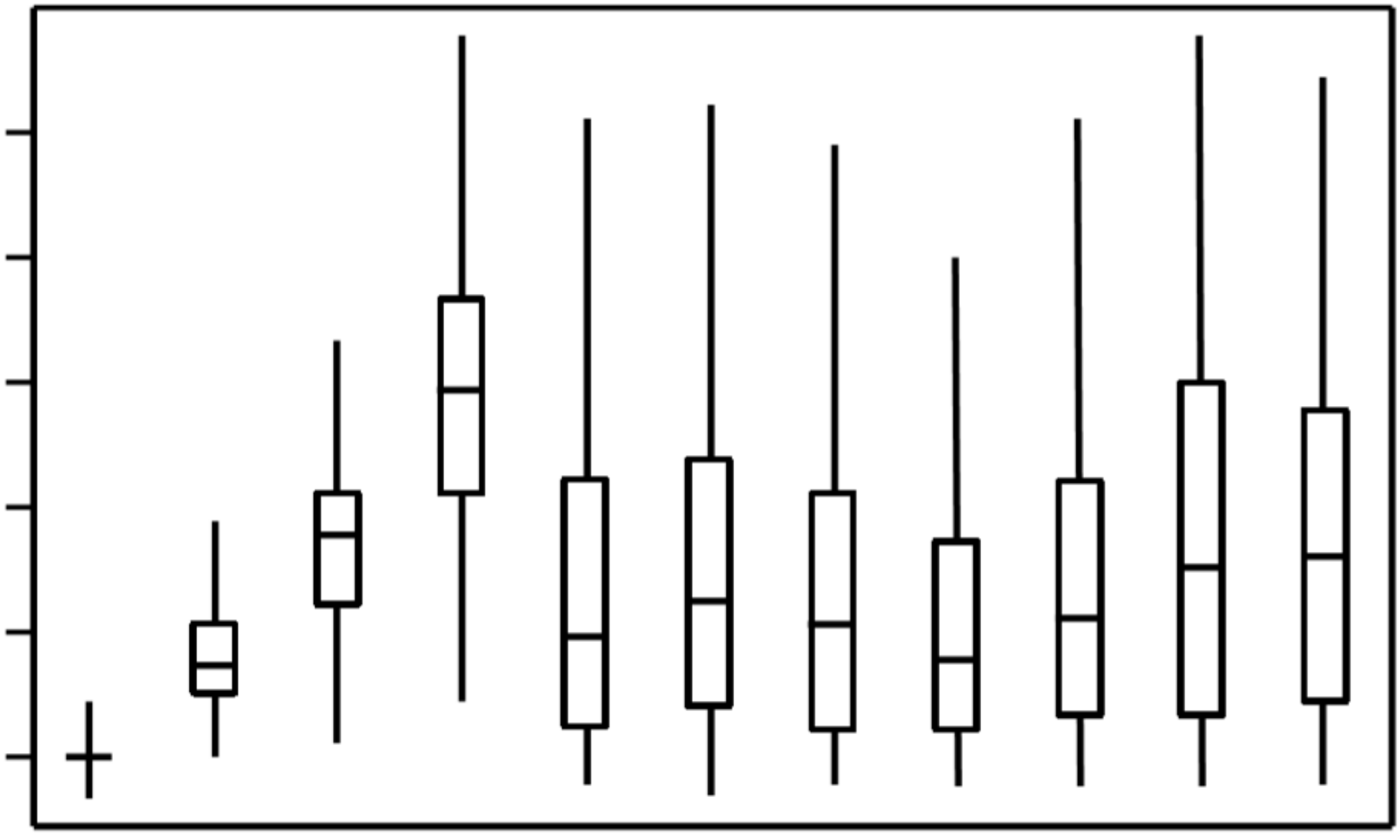
Effectiveness - Avoid Unnecessary Aesthetic Sense

Box Plot with too much aesthetics sense (using too much ink)



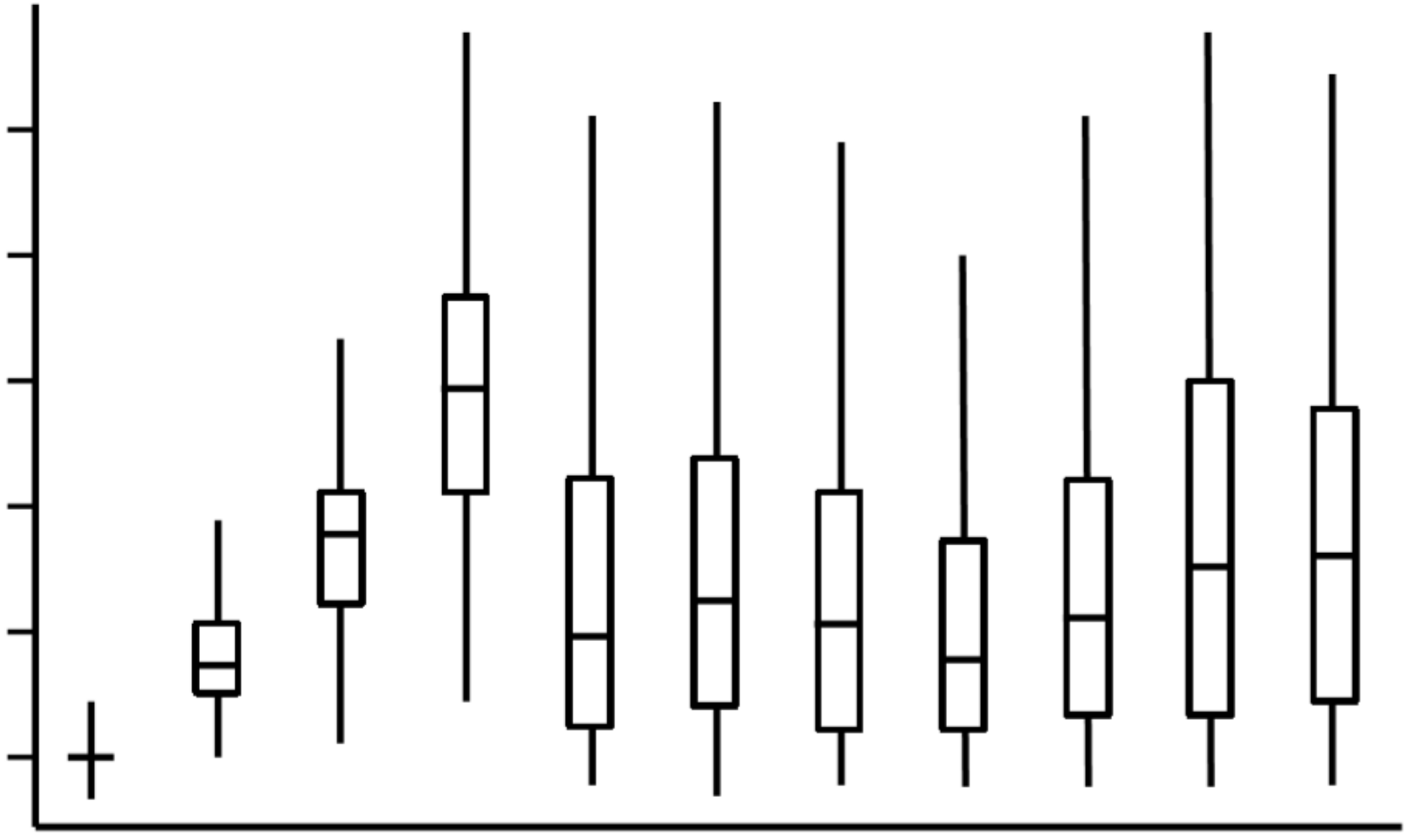
Effectiveness - Avoid Unnecessary Aesthetic Sense

Scale shifted to side



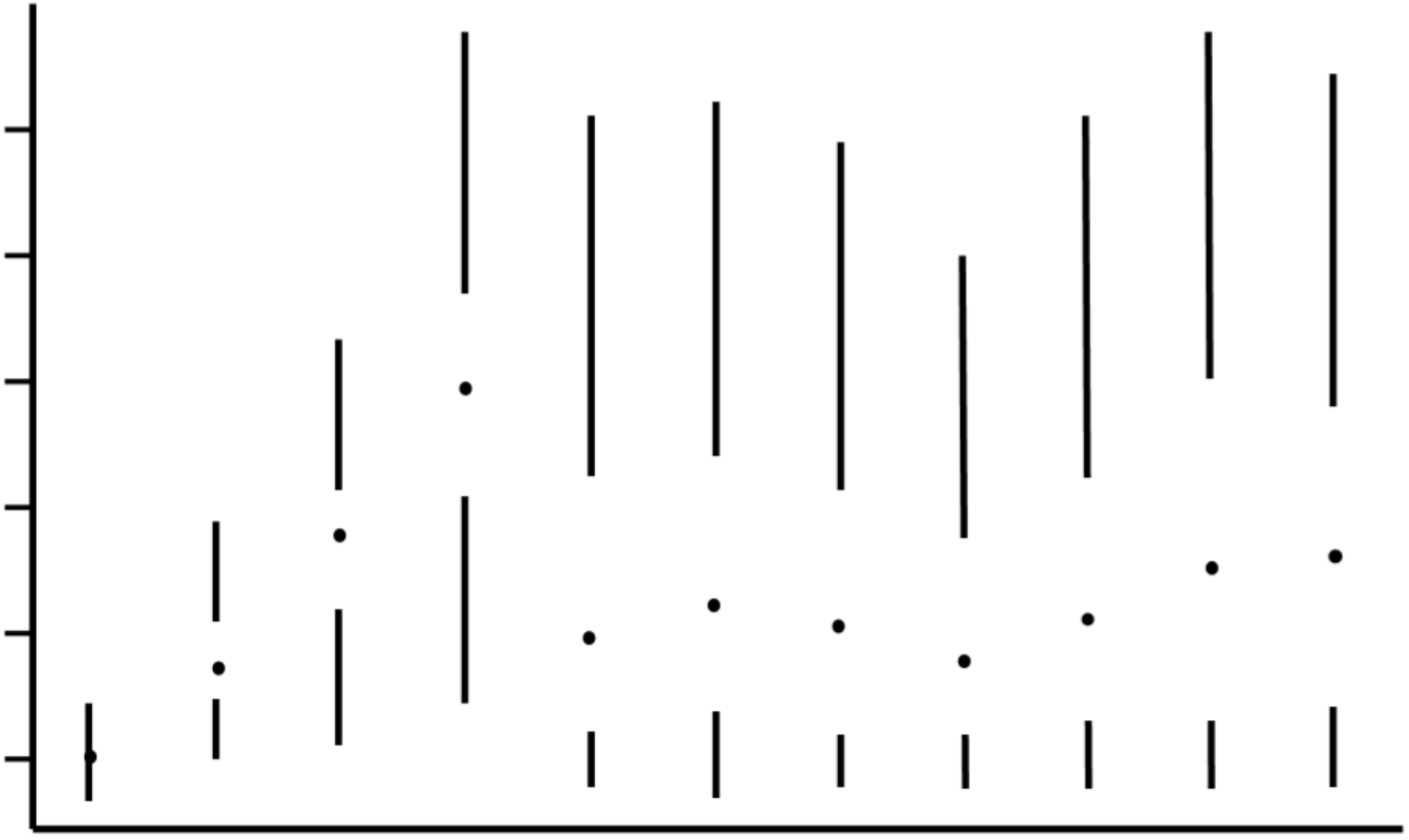
Effectiveness - Avoid Unnecessary Aesthetic Sense

Upper boundaries removed



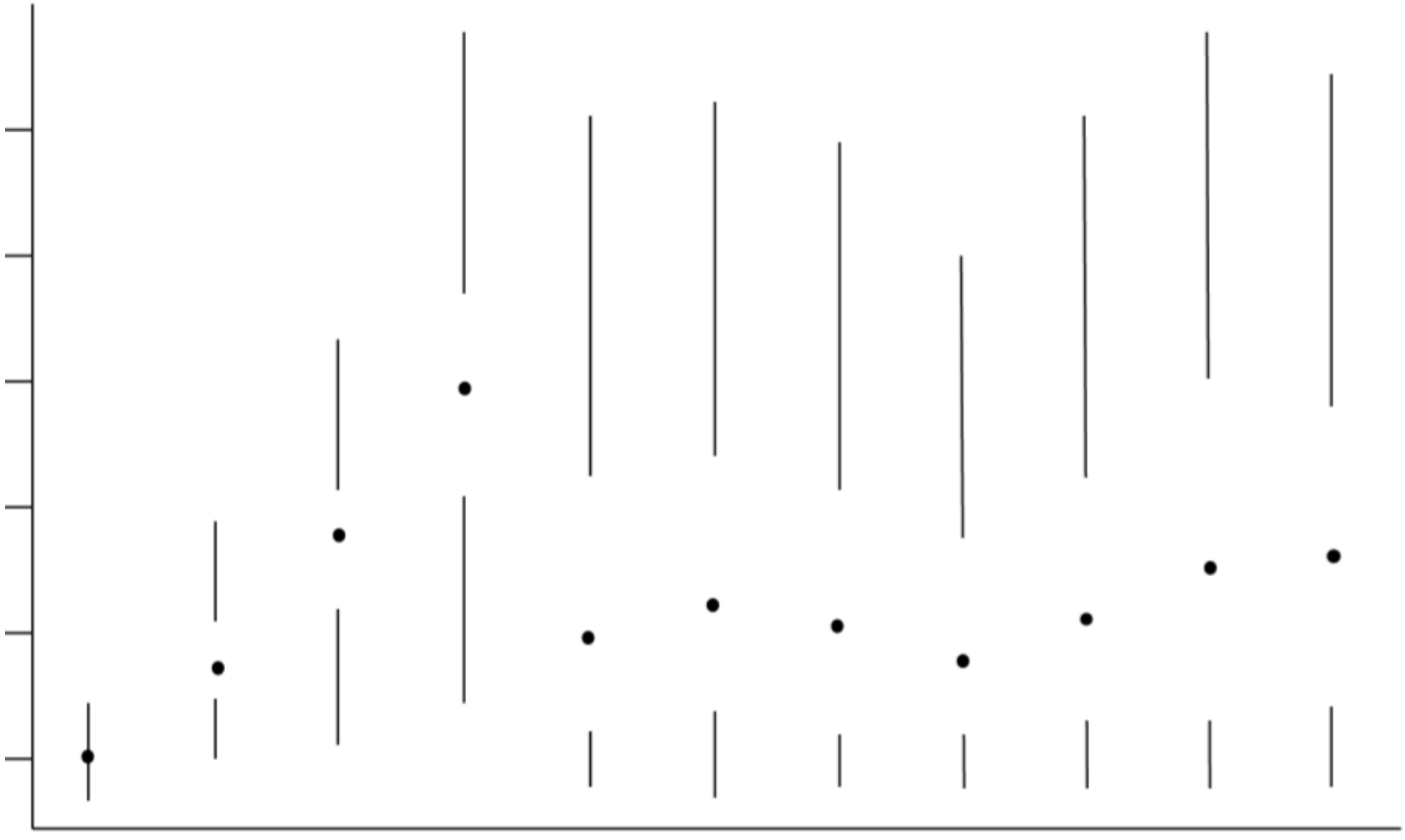
Effectiveness - Avoid Unnecessary Aesthetic Sense

More effective representation



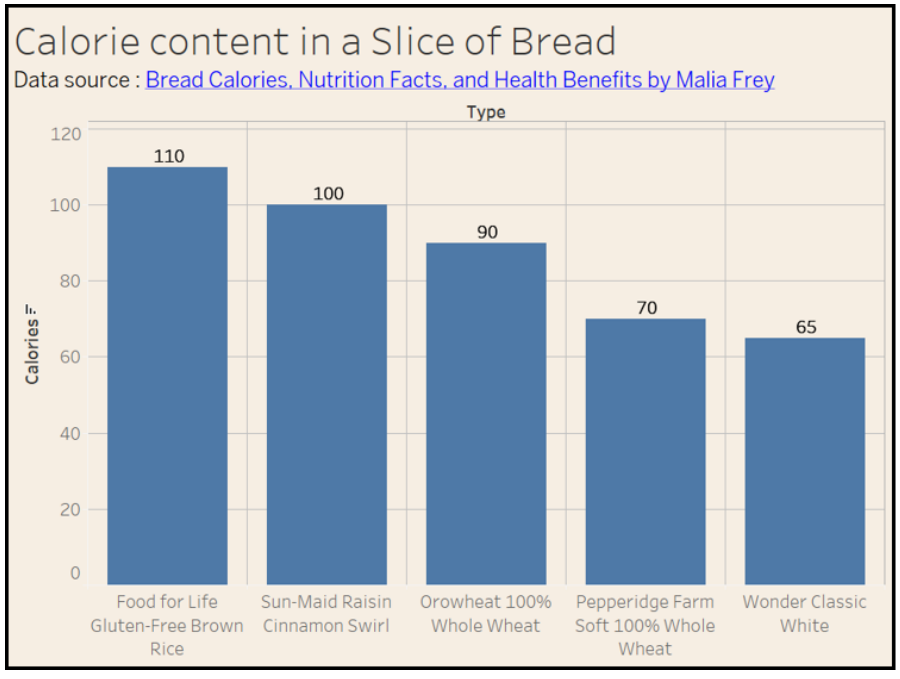
Effectiveness - Avoid Unnecessary Aesthetic Sense

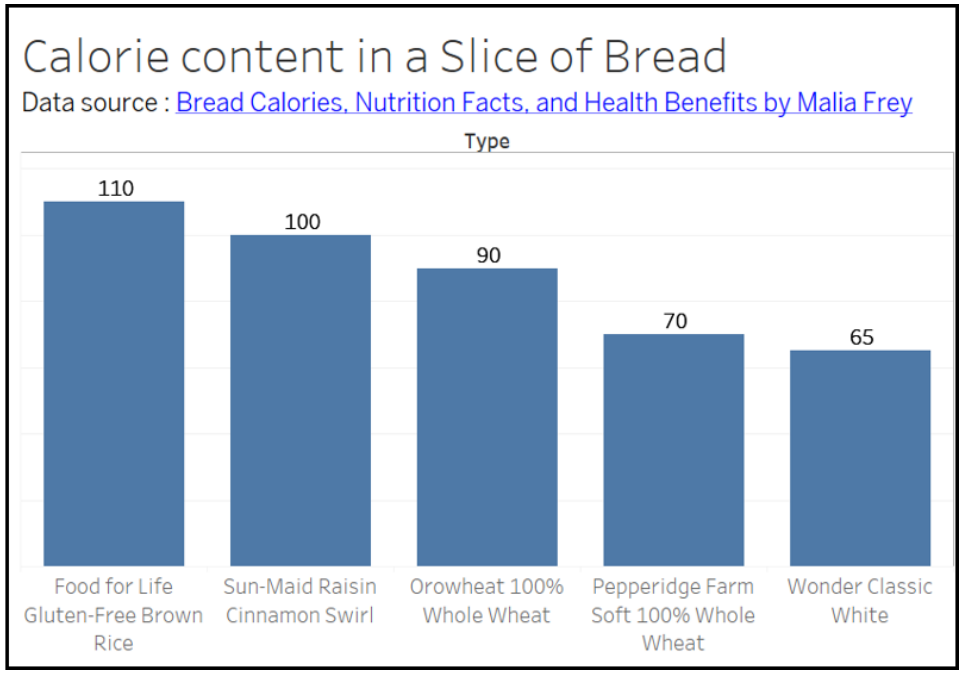
Right brightness



Effectiveness - Avoid Unnecessary Aesthetic Sense

The following plots have exactly the same information but huge difference in ink use

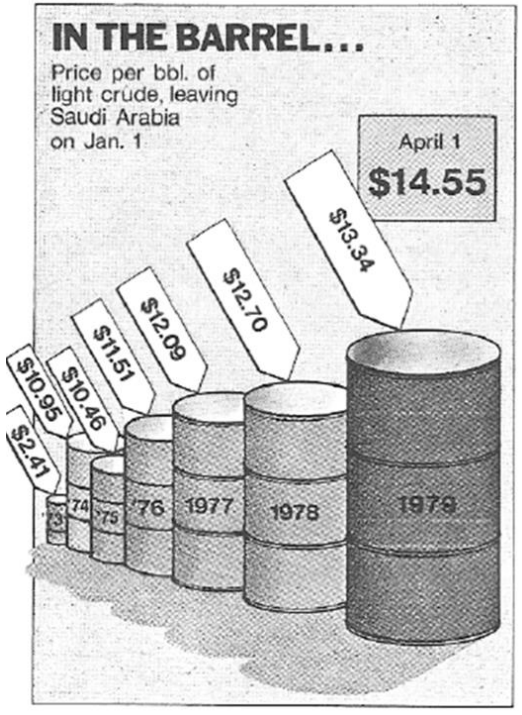




https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

Evaluating Visualization: Consistency

Properties of visualization should match the properties of data



Two-dimensional data mapped with three-dimensional representation

Data Types

Classification of data types: Nominal, ordinal and quantitative

- **N** – Nominal (labels)
  - Fruits: apples, oranges, ...
- **O** – Ordered
  - Quality of meat: Grade A, AA, AAA
- **Q** – Interval (location of zero arbitrary)
  - Dates: Jan 5, 2012; location: (LAT 47 LONG 122)
  - Like a geometric point. Cannot compare directly.
  - Only differences (i.e. intervals) may be compared.
- **Q** – Ratio (zero fixed)
  - Physical measurement: length, mass...
  - Counts and amounts
  - Like a geometric vector, origin is meaningful

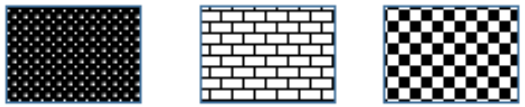
[S. S. Stevens, on the theory of scales of measurements, 1946]

Visual Mapping

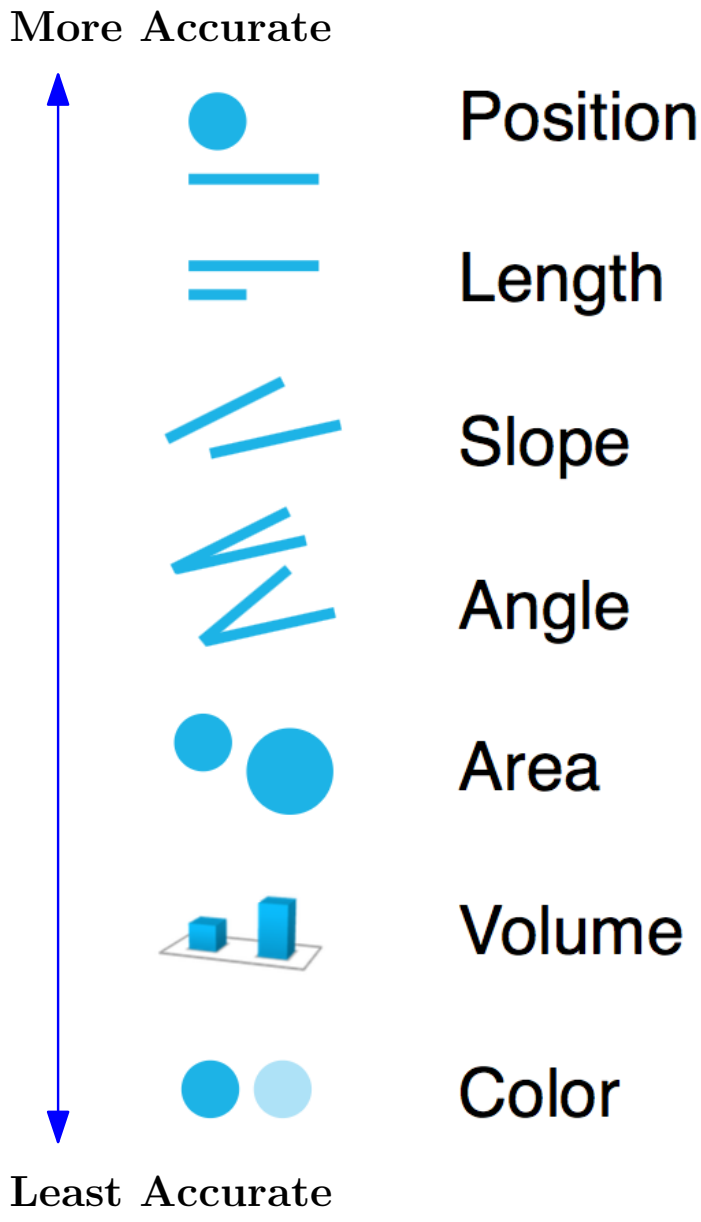
- Mapping data attributes to visual attributes
- Pick the best mapping
- Consider importance Ordering
  - Encode the most important information in the most perceptually accurate way

Visual Mapping: Visual Attributes or Visual Variables

- Position
- Length
- Slope
- Angle
- Area
- Volume
- Texture
- Color
- Shape

Position: changes in the x,y location	
Size: change in length, area or repetition	
Shape: infinite number of shapes	
Value: changes from light to dark	
Colour : changes in hue at a given value	
Orientation: changes in alignment	
Texture: variation in `grain`	

Relative Magnitude Estimation of Visual Variables



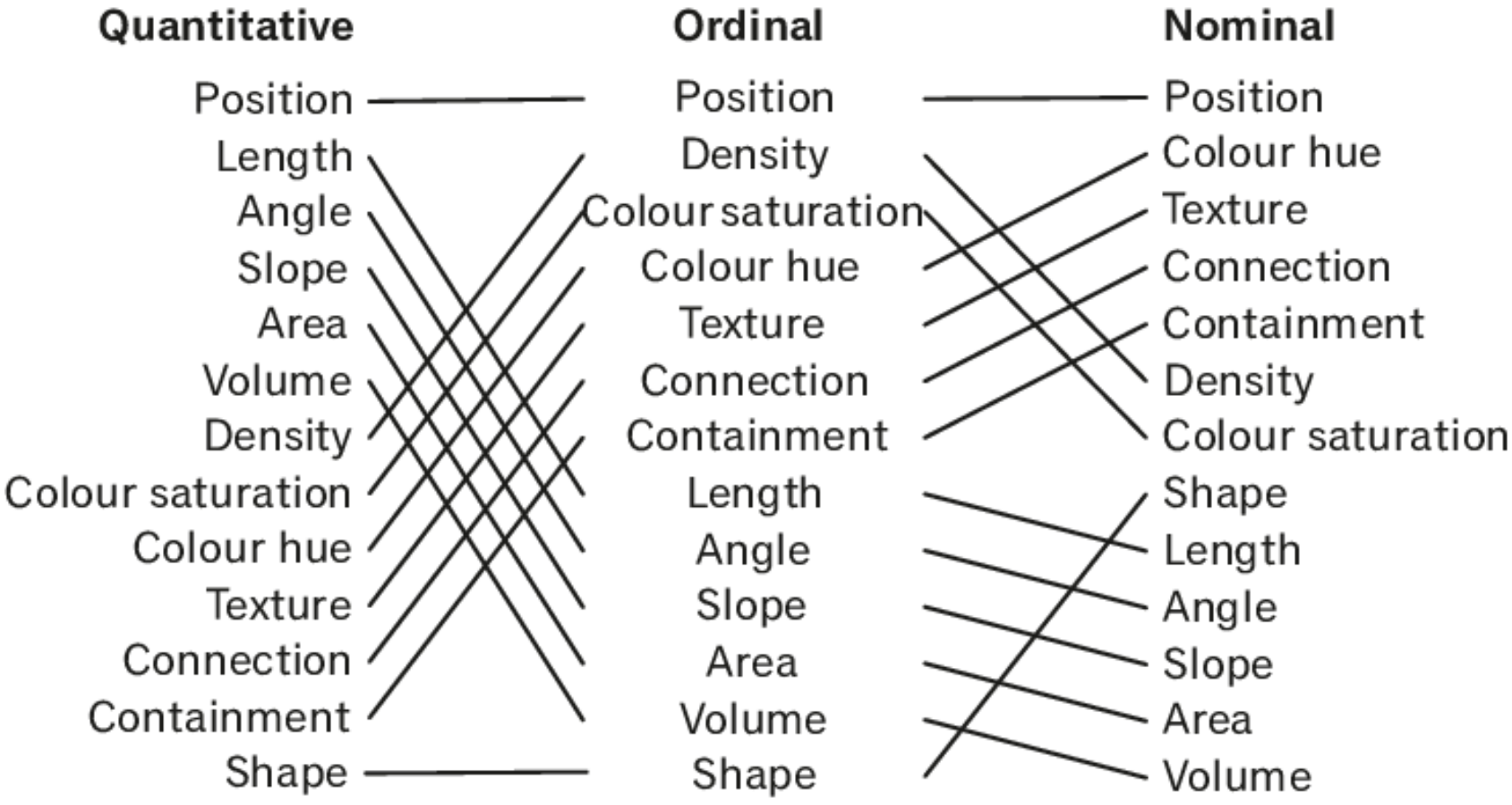
Visual Mapping for Data Types

Bertin's Visual Mapping, Level of Organization

Visual attribute	Suitable target data attributes
Position	**N** **O** **Q**
Size	**N** **O** **Q**
Value	**N** **O** **Q**
Texture	**N** **O**
Color	**N**
Orientation	**N**
Shape	**N**

Visual Mapping for Data Types

Mackinlay ranking of attributes by visualization efficacy



Visual Mapping: Color Encoding

Information in color and value

Value is perceived as ordered

∴ Encode ordinal variables (O)

∴ Encode continuous variables (Q) [not as well]

Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color

May 2013 Cecilia Aragon, HCDE, UW [slide from Maneesh Agrawala]

Guidelines for colors

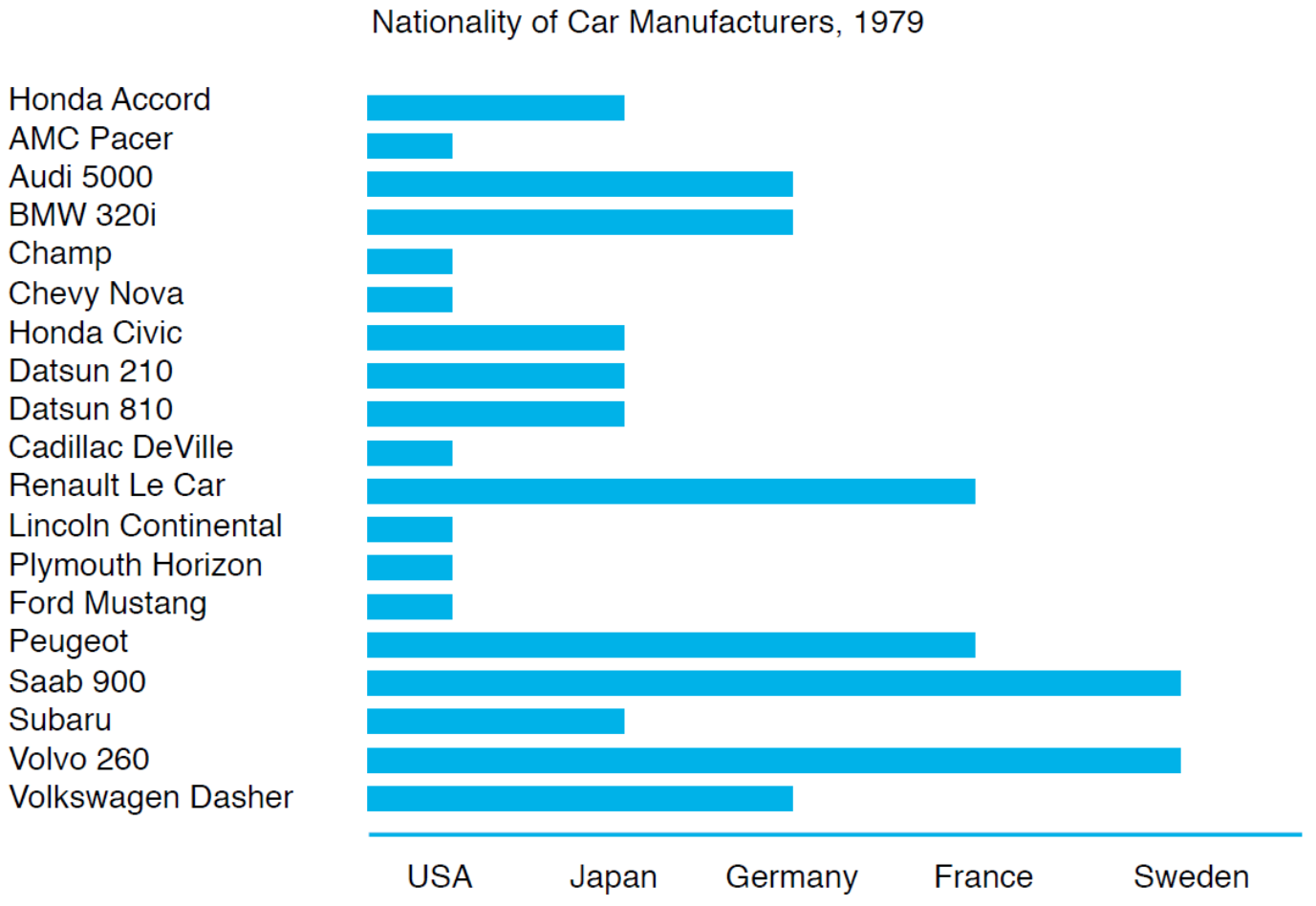
- Use only a few colors
- Colors should be distinctive and named
- Strive for color harmony
- Be aware of cultural conventions
- Beware bad interactions
- Get it right in black and white

Evaluating Visualization: Expressiveness

Expressiveness
A set of facts is expressible in a visual language if the sentences (i.e. the visualization) in the language express all the facts in the set of data and only the facts in the data.

Mackinlay, 1986

Evaluating Visualization: Expressiveness

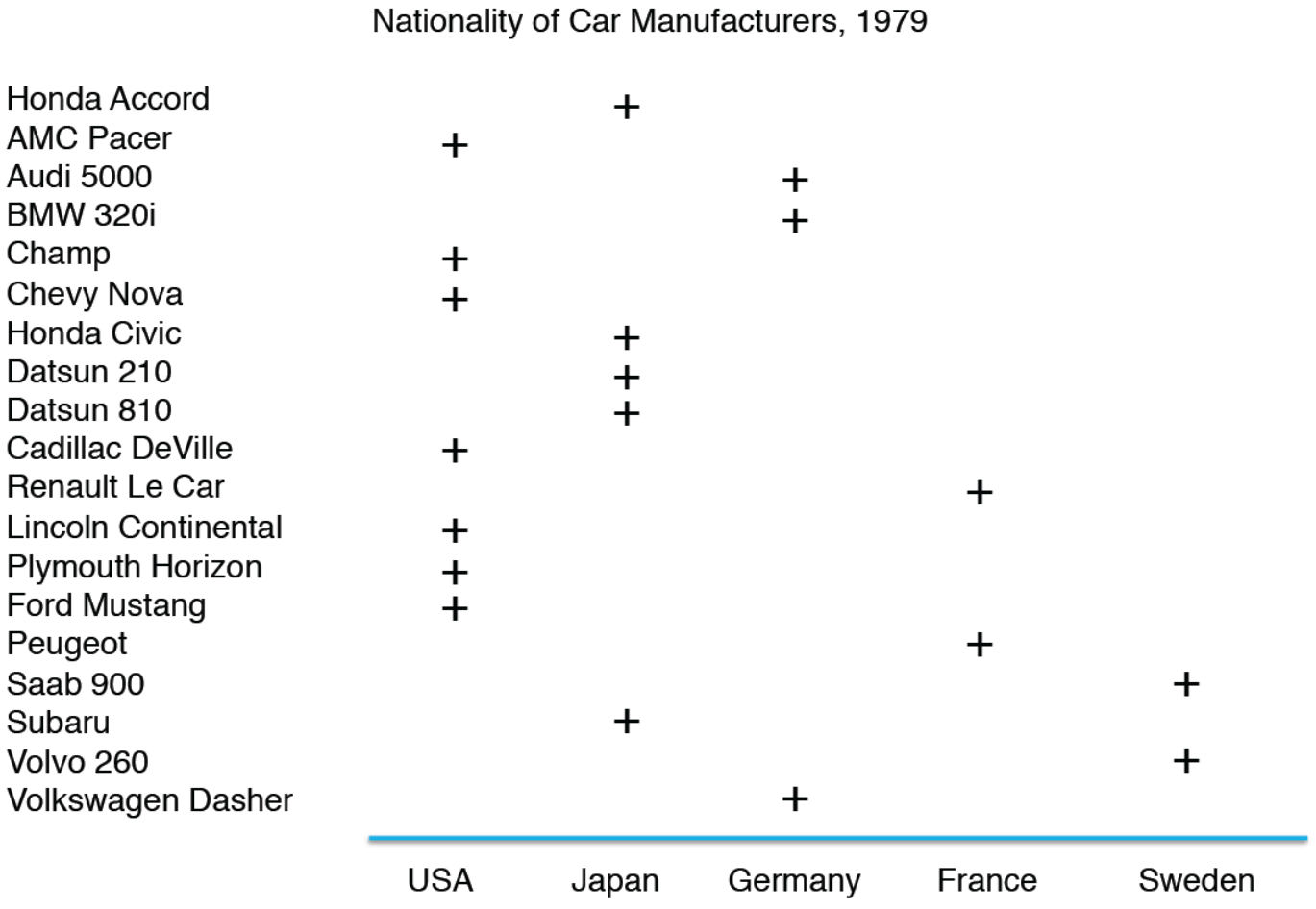


Expressive: Not expressive of the data because faithful is not faithful

Mackinlay, 1986

Lengths (interpreted as quantitative values) express non-facts

Evaluating Visualization: Expressiveness



Expressive: Not expressive of the data because faithful is not faithful

Mackinlay, 1986

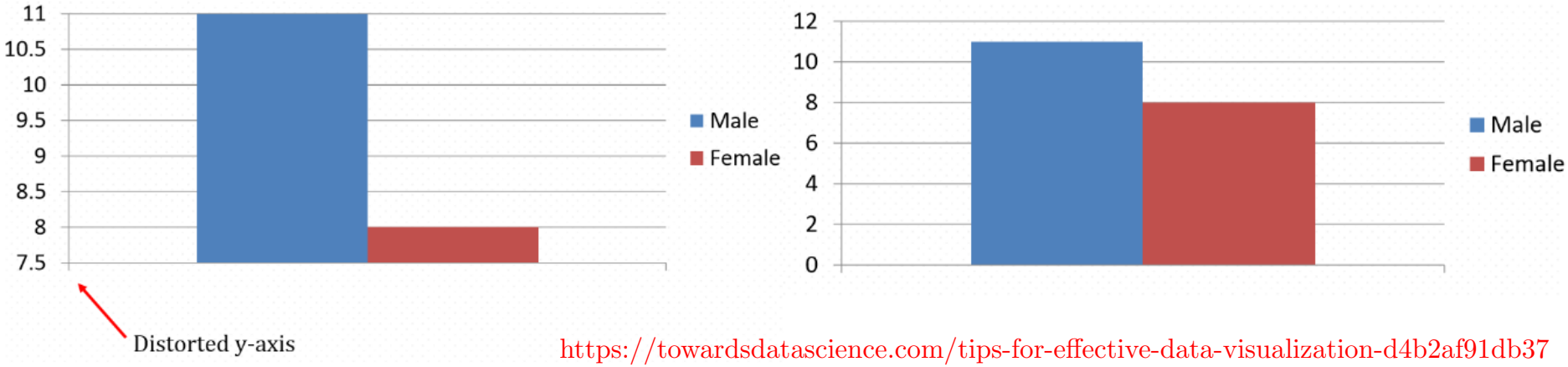
Lengths (interpreted as quantitative values) express non-facts

Evaluating Visualization: Integrity

Integrity
What is presented should accurately represents what is in the data being visualized, and that no design choices should distort or obfuscate the facts and analytical findings

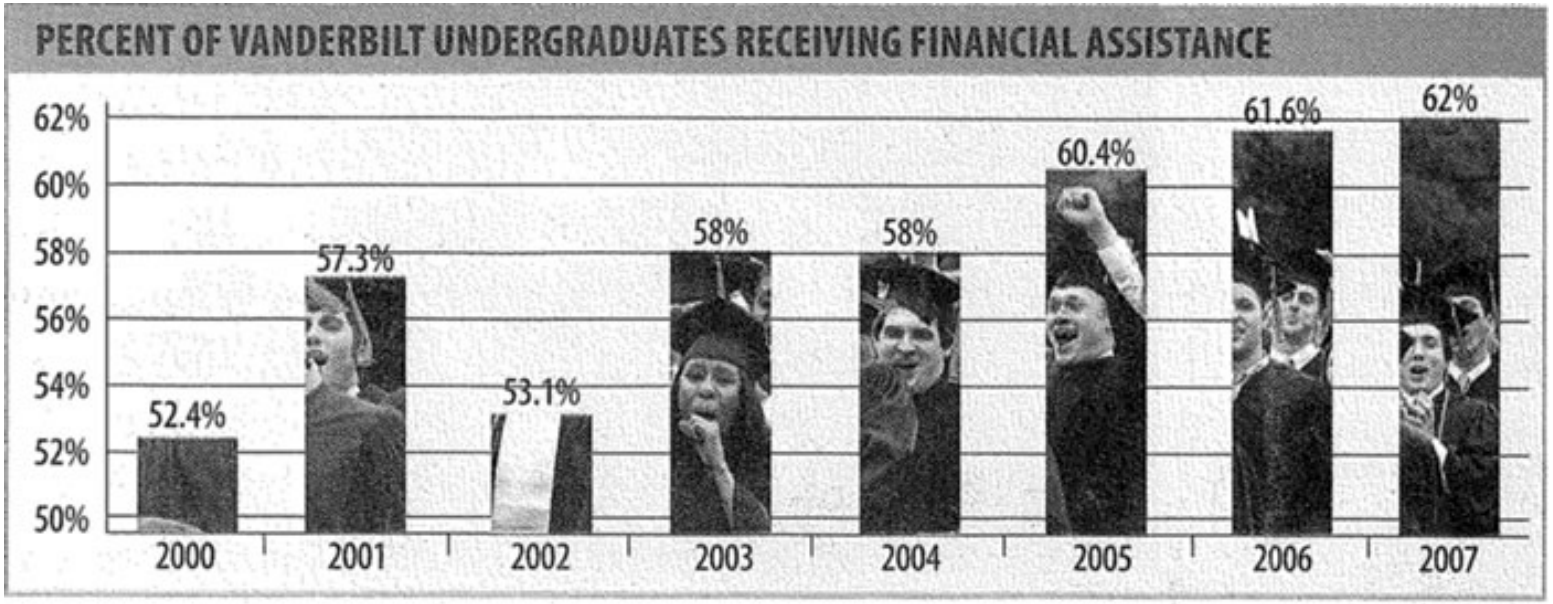
Evaluating Visualization: Integrity

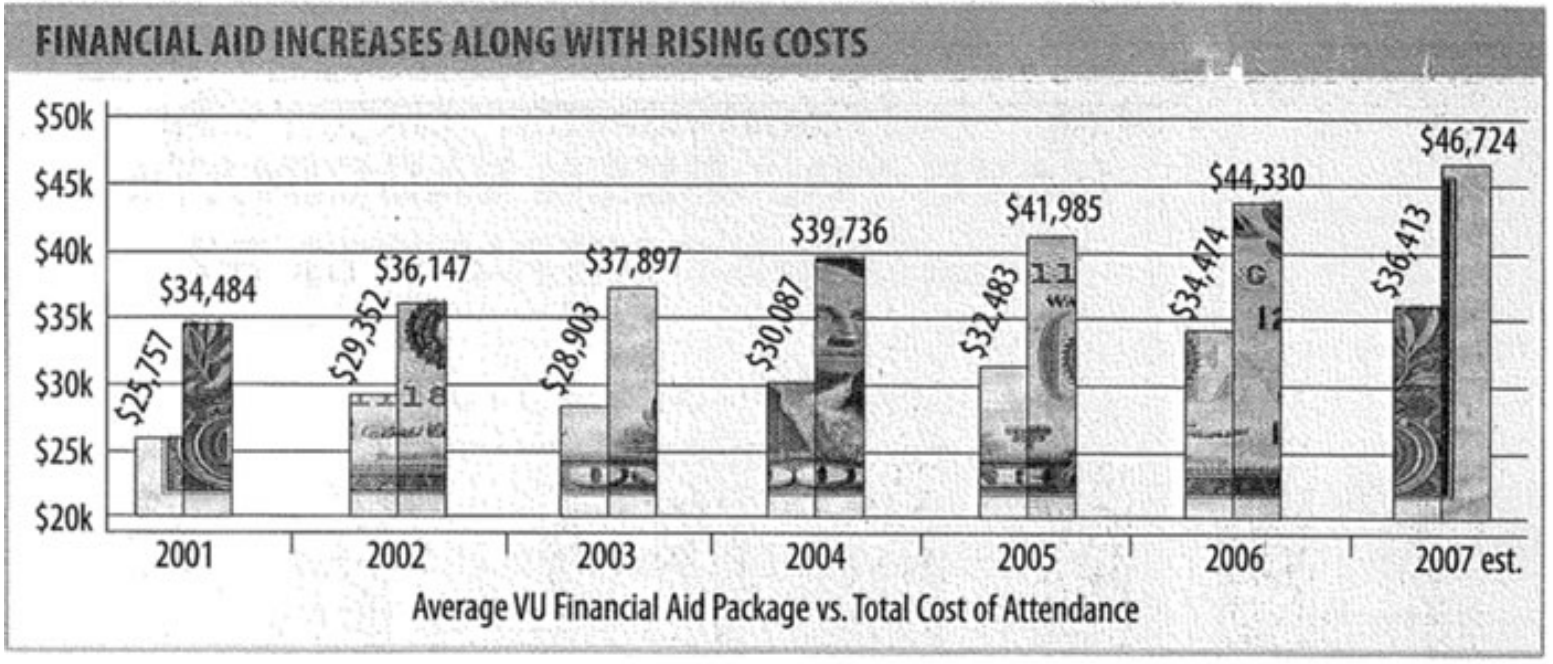
Student Gender Distribution



https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37

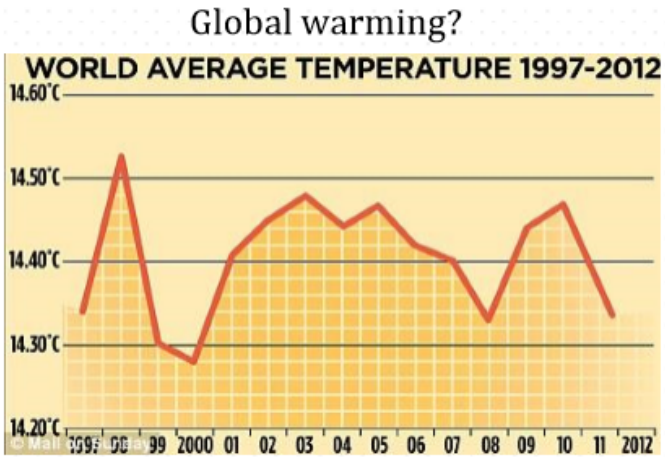
Evaluating Visualization: Integrity

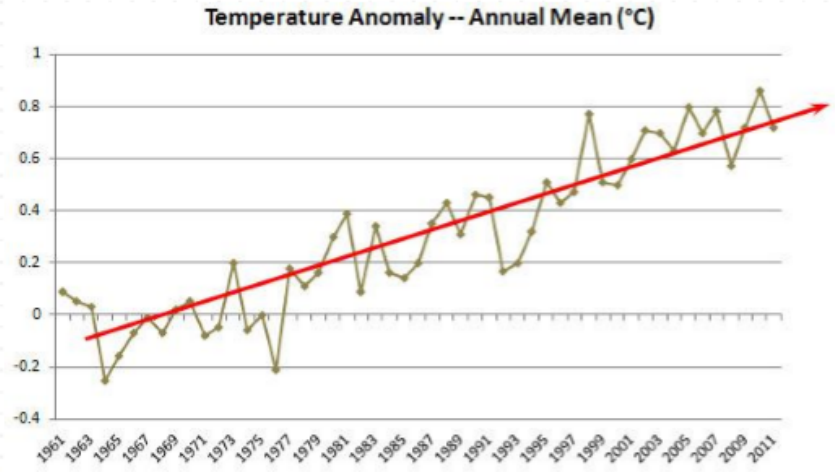


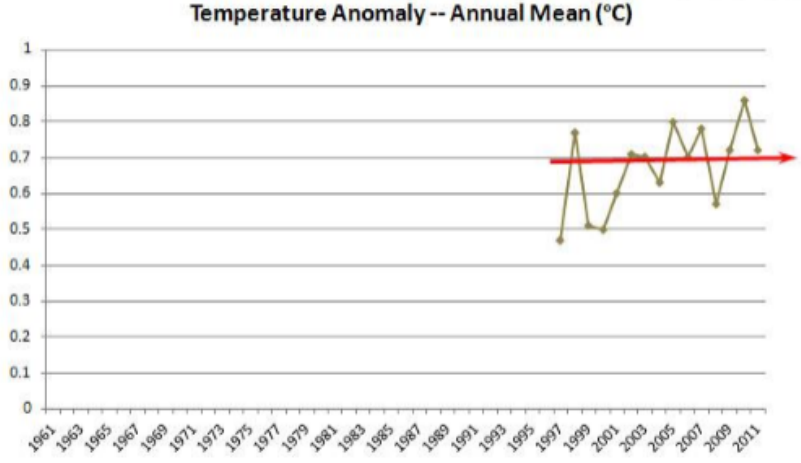


Evaluating Visualization: Integrity

Evaluating Visualization: Integrity







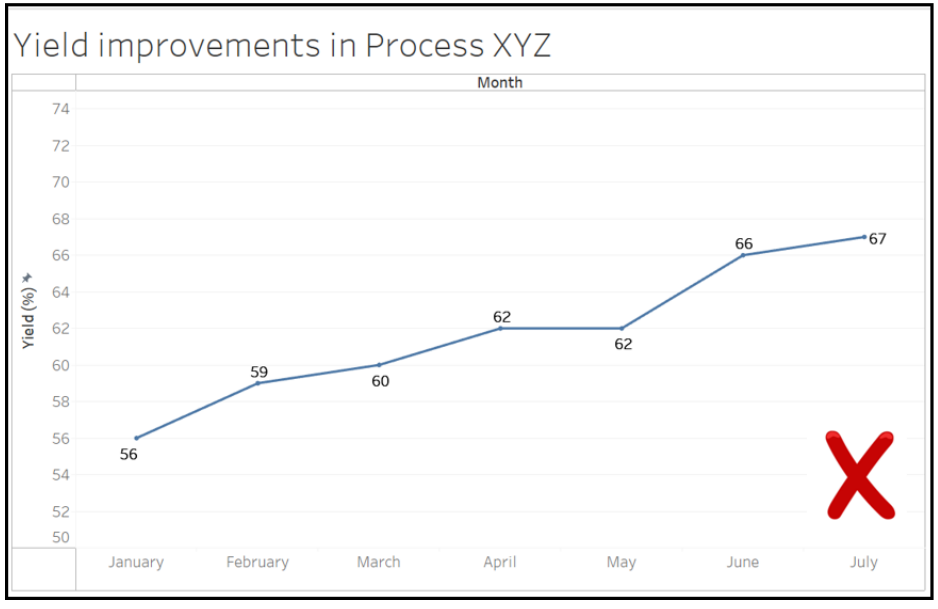
Global warming! ☹

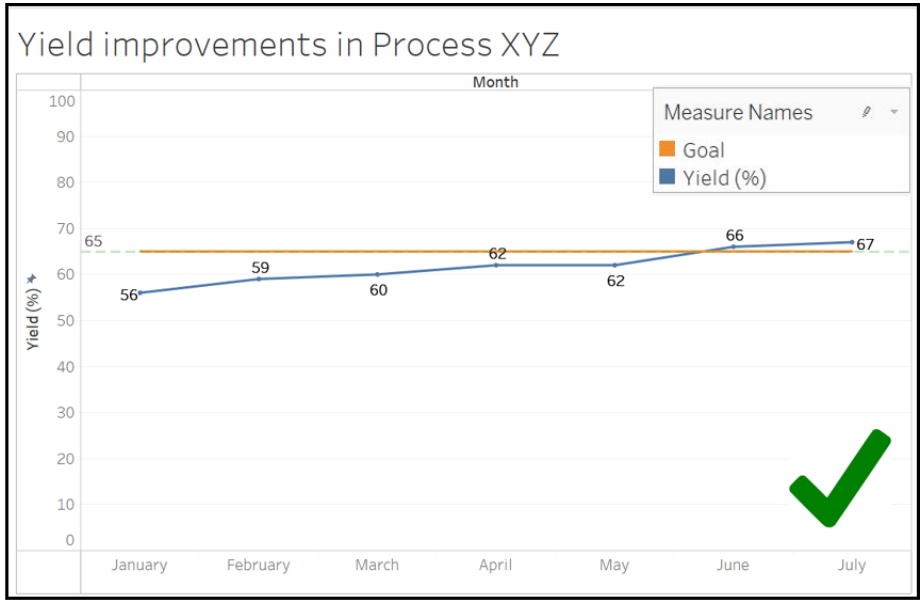
Distorted x-axis for rise in global warming data

Evaluating Visualization: Integrity

Yield of a process increased from 56% to 67% over a period of 6 months

Which visual is exaggerating the increase?

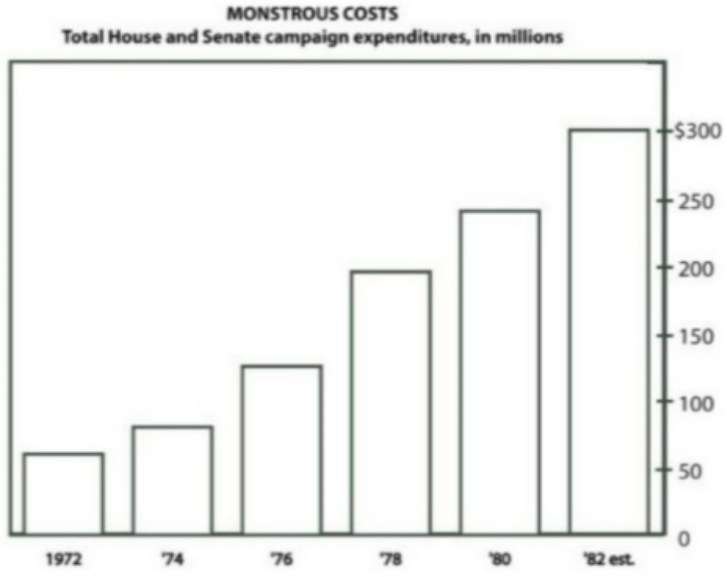




Evaluating Visualization: Integrity

American election expenditures

Which visual is exaggerating the increase in expenses from 1972 to 1982?



http://www.astro.caltech.edu/ay119/bdass/davidoff-3-viscommfund.pdf

Evaluating Visualization: Chart Junk

Maximal Data:Ink Rato
A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts.

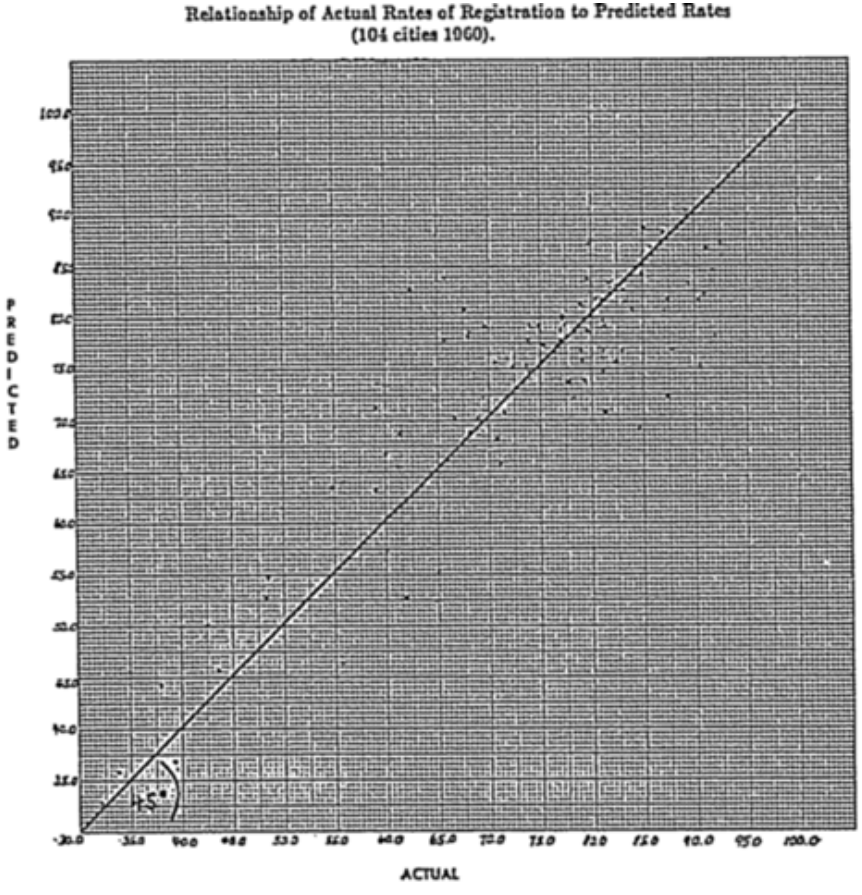
William Strunk, Jr.

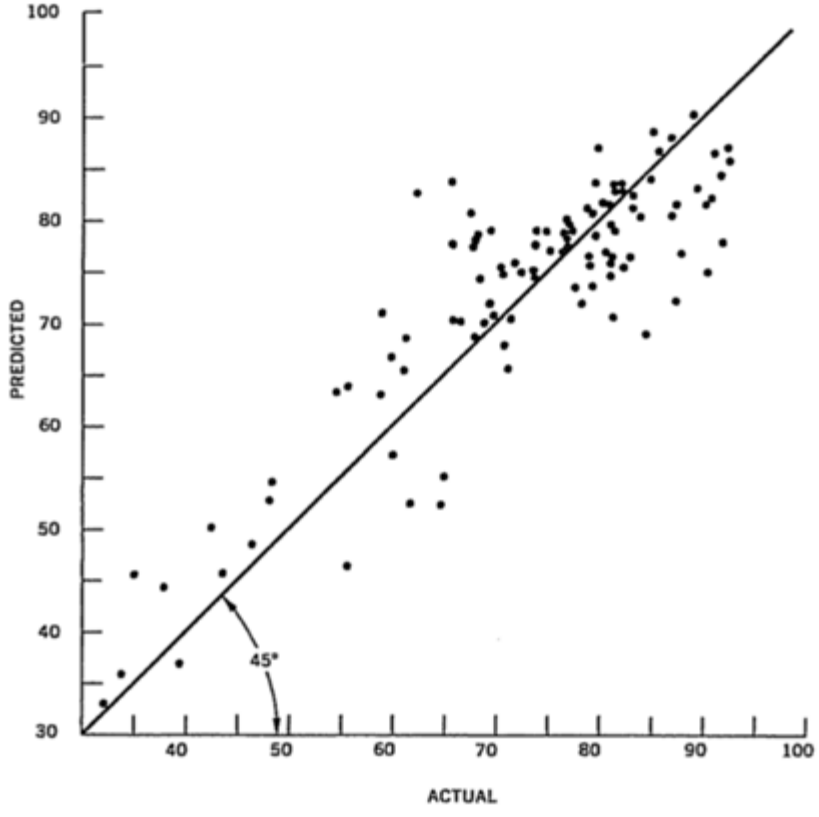
- Do not try to deceive the audience
- Avoid 3D
- Keep chart junk to minimum to prevent distractions
- Minimal usage of Ink
- Some chart junk helps in remembering though

Evaluating Visualization: Chart Junk

Avoid chart junk, if it does not add any value

http://jcsites.juniata.edu/faculty/rhodes/ida/graphicalIntRedes.html

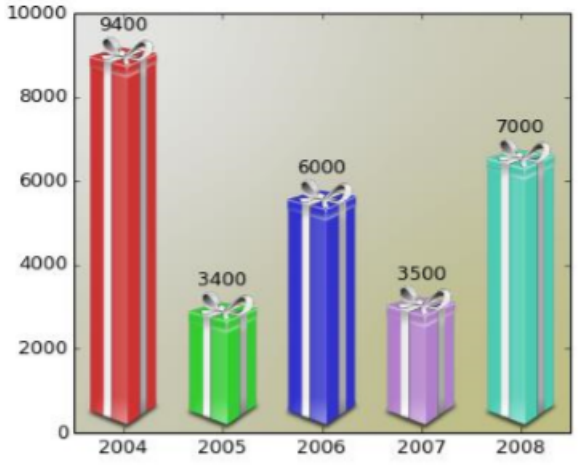


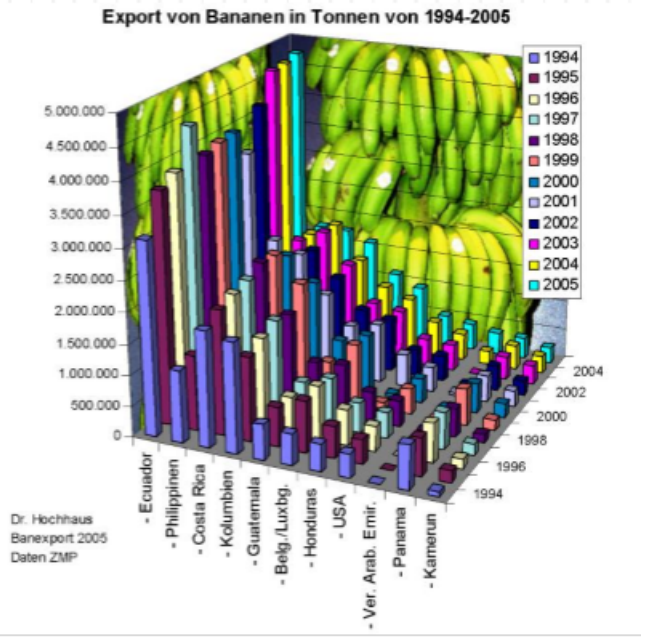


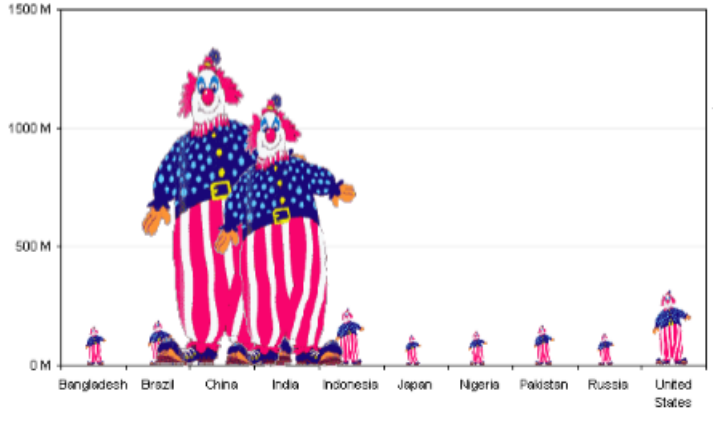
Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

Evaluating Visualization: Chart Junk

Avoid chart junk, if it does not add any value

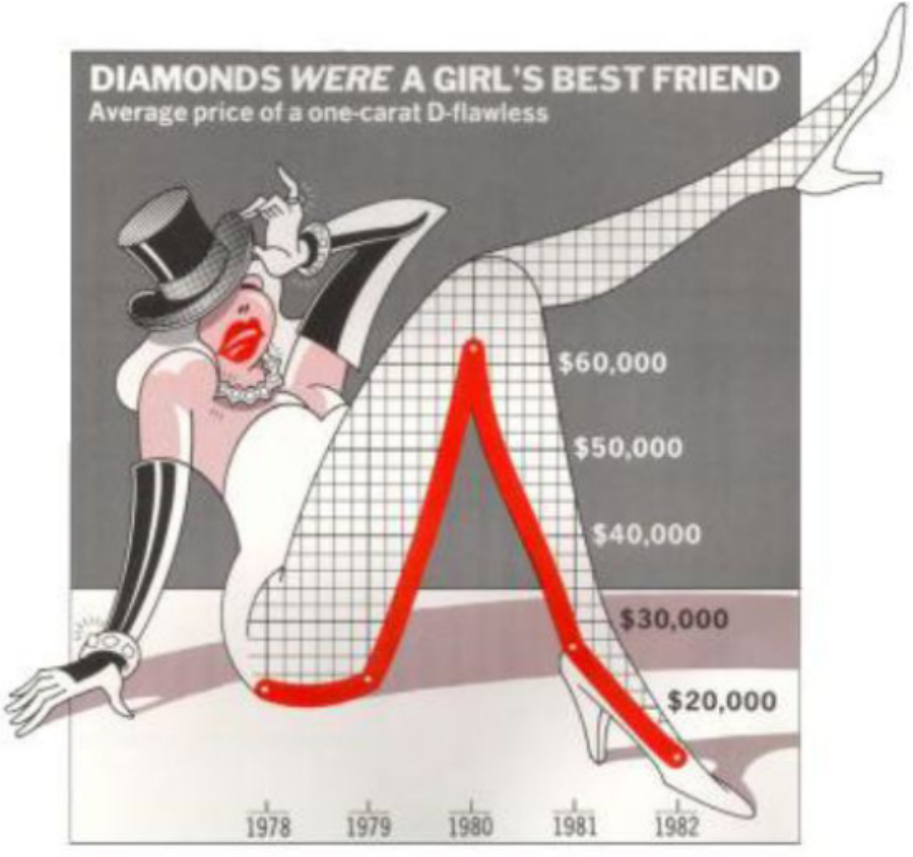


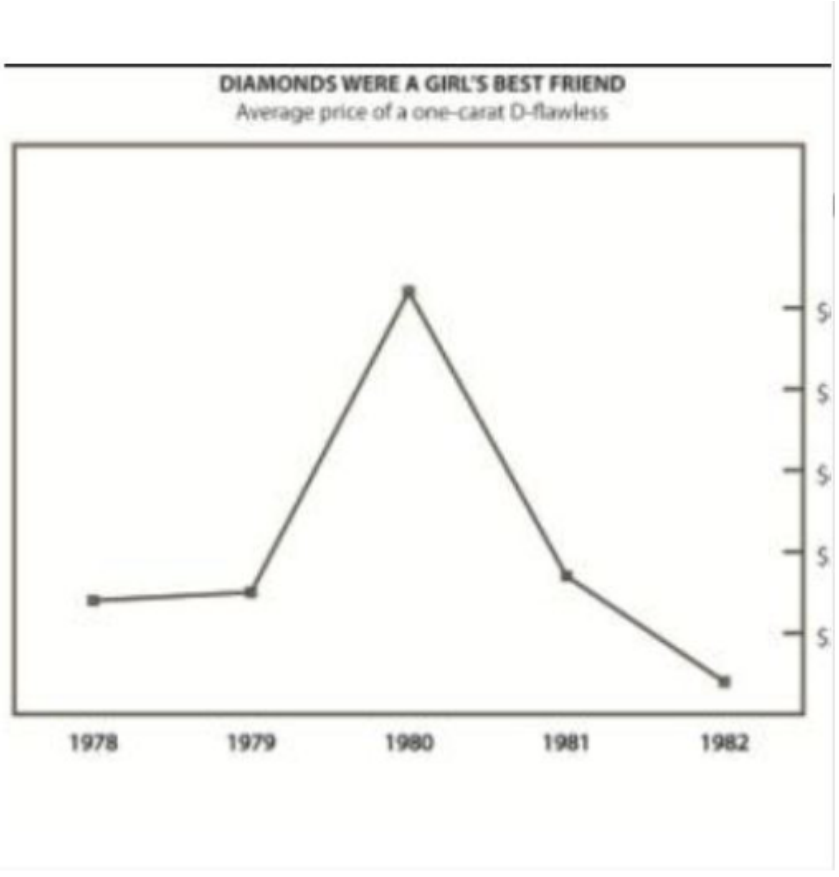




Evaluating Visualization: Chart Junk

Avoid chart junk, if it does not add any value





Historical diamond price
Edward Tufte

The Grid System

Grid system naturally organizes data to give it more meaning

The Grid System

The ultimate resource in grid systems.

"The grid system is an aid, not a guarantee. It permits a number of possible uses and each designer can look for a solution appropriate to his personal style. But one must learn how to use the grid; it is an art that requires practice."
Josef Müller-Brockmann

Hide Grid

Search

Articles

Compose to a Vertical Rhythm
On the Web, vertical rhythm is contributed to by three factors: font size, line height and margin or padding. All of these factors must calculated with care in order that the rhythm is maintained.
04.Dec.2008

Incremental leading
In editorial design, there is a technique used for sidenotes and boxouts that aligns to the baseline grid, or vertical rhythm. It's called incremental leading.
03.Dec.2008

Applying Divine Proportion to Your Web Designs
This article explains what is the Divine proportion and what is the Rule of Thirds and describes how you can apply both of them effectively to your designs.
01.Dec.2008

View All Articles →

Tools

960 Grid System
An effort to streamline web development workflow by providing commonly used dimensions, based on a width of 960 pixels. There are two variants: 12 and 16 columns, which can be used separately or in tandem.
04.Dec.2008

Graph Paper by Konigi
This graph paper is made for visual designers, interaction designers, and information architects. You'll find styles for wireframing, storyboarding, plotting values and for drafting sitemaps.
03.Dec.2008

Syncotype
Syncotype is a simple tool to help align your text to a baseline grid. Enter your line height and offset in pixels in the Syncotype control box and click "Syncotype it" to overlay a baseline grid in red.
01.Dec.2008

View All Tools →

Books

Geometry of Design
The book focuses on the classic systems of proportioning, such as the golden section and root rectangles, as well as systems such as the Fibonacci Series.
04.Dec.2008

The Typographic Grid
We consider this to be the academic part two to "Grid Systems." Hans Rudolf Bosshard tackles a deeper understanding of the complex grid.
30.Nov.2008

Grid Systems
Grid Systems provides a rich, easy-to-understand overview and demonstrates a step-by-step approach to typographic composition.
21.Nov.2008

View All Books →

Templates

InDesign 8.5x11 Grid System (12)
Adobe InDesign file with a grid system for an 8.5"x11" page that is divided into 12 columns and rows using the Rule of Thirds (Golden Ratio). Includes a 12pt baseline grid.
29.Nov.2008

InDesign 11x17 Grid System (12)
Adobe InDesign file with a grid system for an 11"x17" page that is divided into 12 columns and rows using the Rule of Thirds (Golden Ratio). Includes a 12pt baseline grid.
29.Nov.2008

Photoshop 975px Grid System (12)
Adobe Photoshop file with a grid system for a 975px wide page that is divided into 12 columns and rows using the Rule of Thirds (Golden Ratio). Includes a 16px baseline grid.
29.Nov.2008

View All Templates →

Blog

UX Magazine
A well designed collaborative site, with a very nice grid structure, that focuses on user experience.
02.Dec.2008

Doane Paper Utility Notebook
A portable notebook featuring a patent pending Grid+Lines stationery design that combines the benefits of grid and ruled lines onto a single sheet of paper.
28.Nov.2008

Replica Typeface
Replica is a new typeface by Norm that was designed on a strict grid system. Available in the following weights: Regular, Italic, Light, Light Italic, Bold and Bold Italic.
21.Nov.2008

View All Blog Posts →

Inspiration

Ace Jet 170
AisleOne
Athletics
BBDK
Blanka
Build
Corporate Risk Watch
David Airey
Dirty Mouse
Experimenta
Experimental Jetset
Form Fifty Five
Grafik Magazine
Grain Edit
Graphic Hug
Helvetica Film
I Love Typography
Lamosca
magCulture
Mark Boulton
Minimal Sites
Monocle
Neubau
NewWork
OK-RM
Original Linkage
Robin Uleman
SampsonMay
Schmid Today
September Industry
Sonifyer
Soulellis
Subtraction
Swiss Legacy
Thinking for a Living
This Studio
Toko
Visuelle
Xavier Encinas
Year of the Sheep

Grid System classic text
Josef Muller-Brockmann

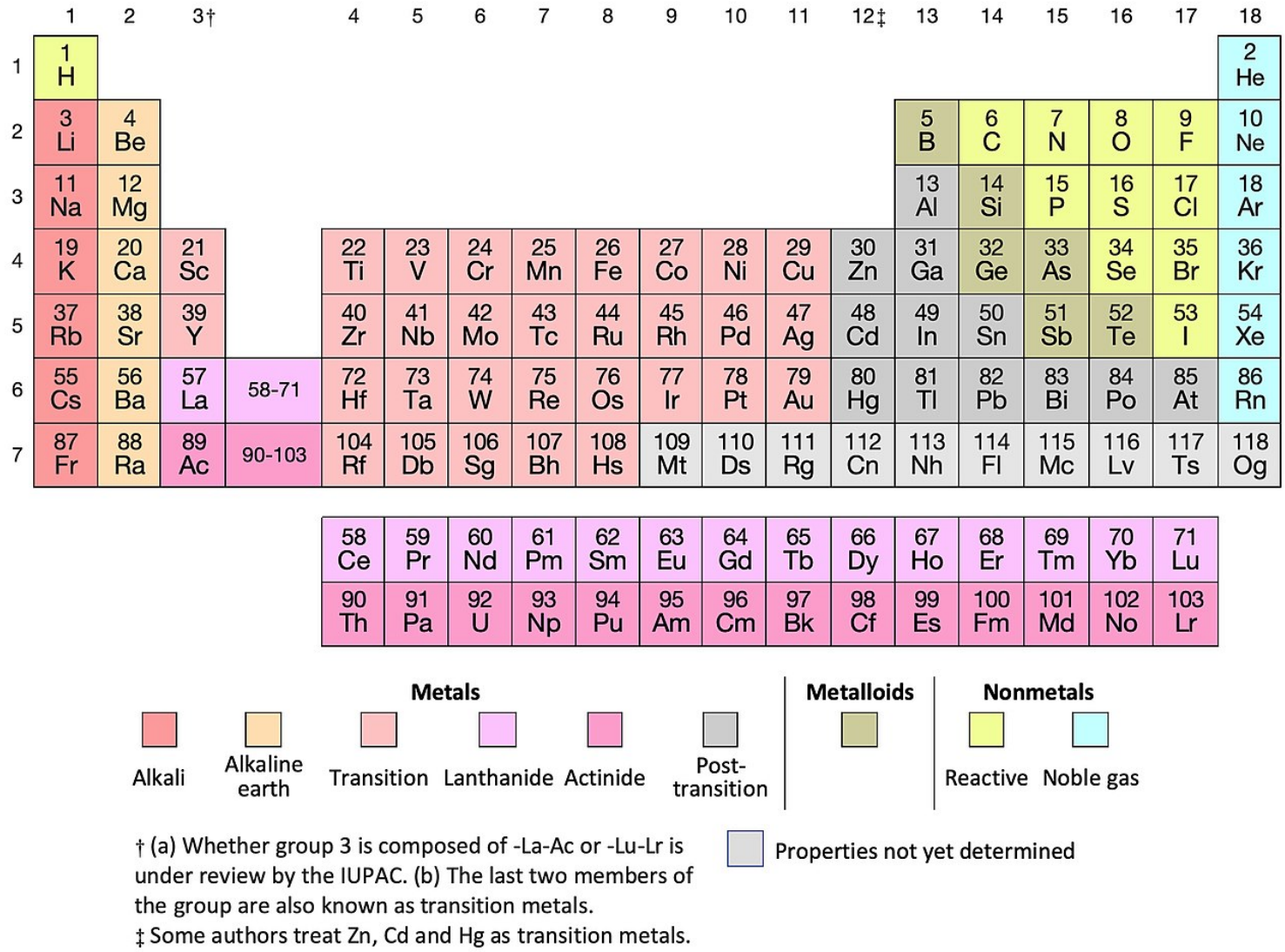
The Grid System

Which news is more important? Which is more visible?

New York Times Cover Page
Grid Fox Firefox extension

The Grid System

Grouping of elements in columns has a certain meaning



Data Visualization Process

1. Purpose of your visualization
   - Are you exploring the data?
   - Are your formatting it for decision making?
   - Or are you telling a story?
2. Eight Principles of communicating through data
   - Define what questions are you answering
   - Use accurate data
   - Experiment with ways to answer
   - Go with cognitive research (go with the rules defined through previous research for data visualization)
   - Faithfully represent your data
   - Tailor it to your audience
   - Make it as simple as possible
   - Remove everything that you can

Data Visualization Process

1. Choosing the visualization for your purpose
   - Simple numbers? pie charts? bar charts? Tables? plots? maps?
2. Choosing right tool and coding language
   - excel, tableau, Microsoft power BI, illustration software
   - R, Python etc.