DSA5101 Lecture 4
Visual Display of Quant. information

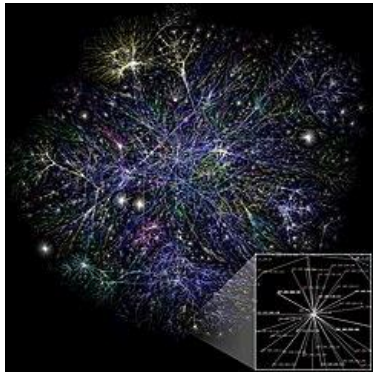
LX Zhang

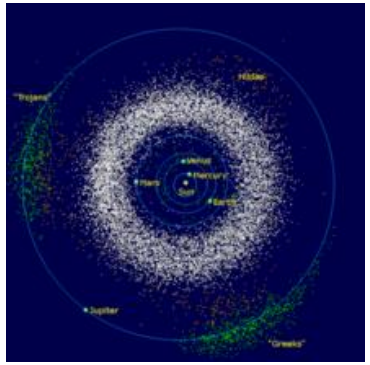
Department of Mathematics

National University of Singapore

Part I: What is Information Visualization

- Information visualization
  -- Visual representations of abstract data to reinforce human cognition
  -- Data can be either numerical or non-numerical
- Scientific visualization
  -- Visual display of spatial data (associated with scientific processes), such as solar system, weather forecast
- (Tamara Munzner) It's information visualization if the spatial representation is chosen, whereas it's sci. visualization if the spatial representation is given.



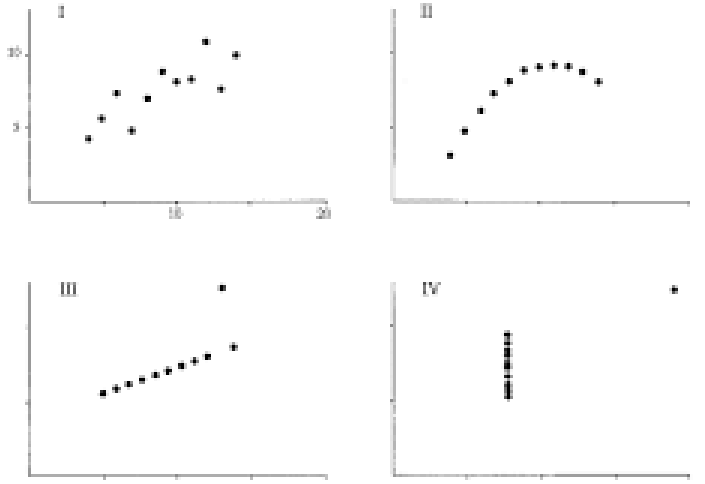


- Why is visualization important?

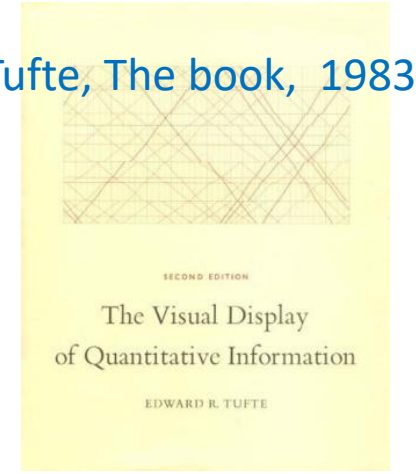
  -- Need an good way to understand a big data

  -- The human visual system is the highest bandwidth channel to the human brain.

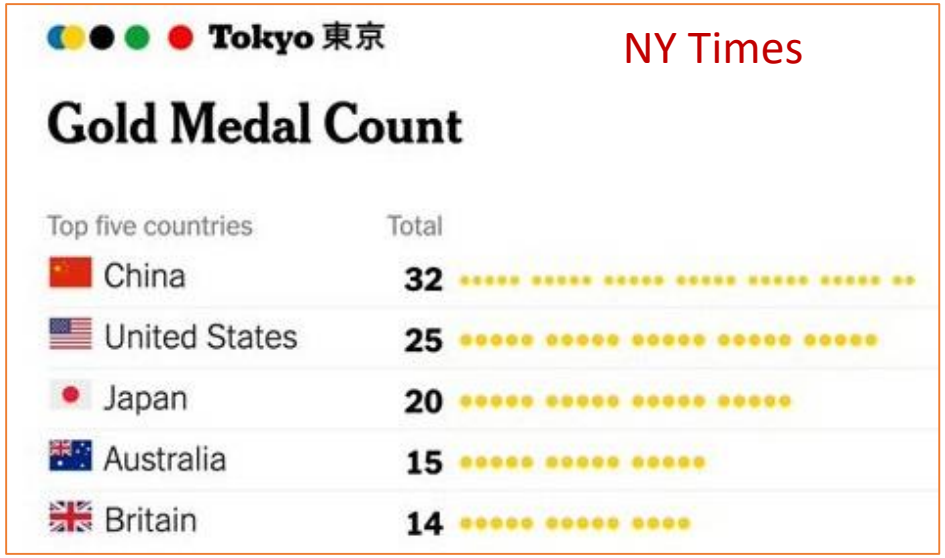
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

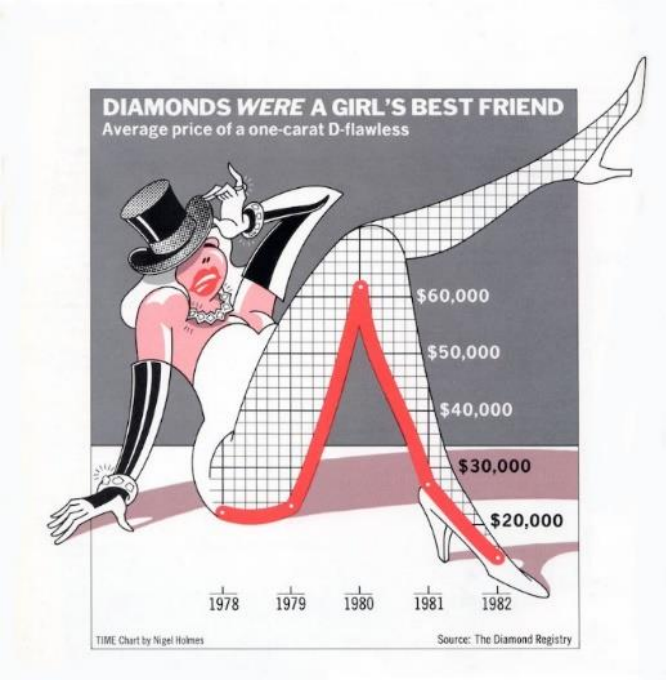


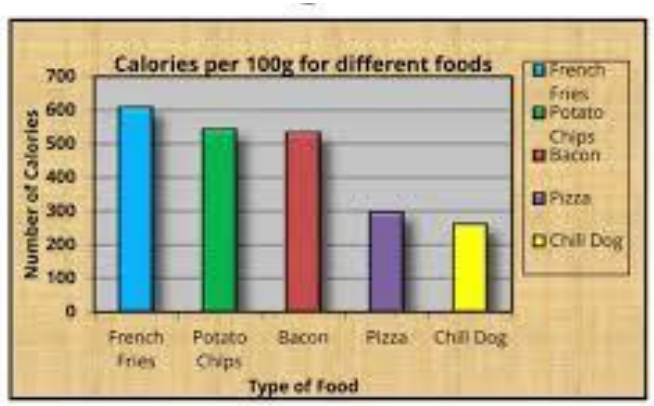
E Tufte, The book, 1983

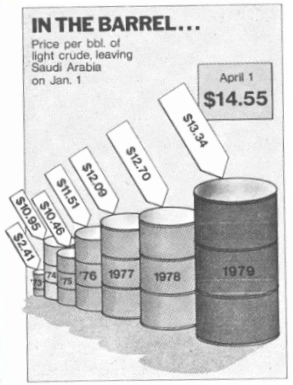


- Goals of information visualization
  -- Effective
  -- Clarity
  -- Integrity

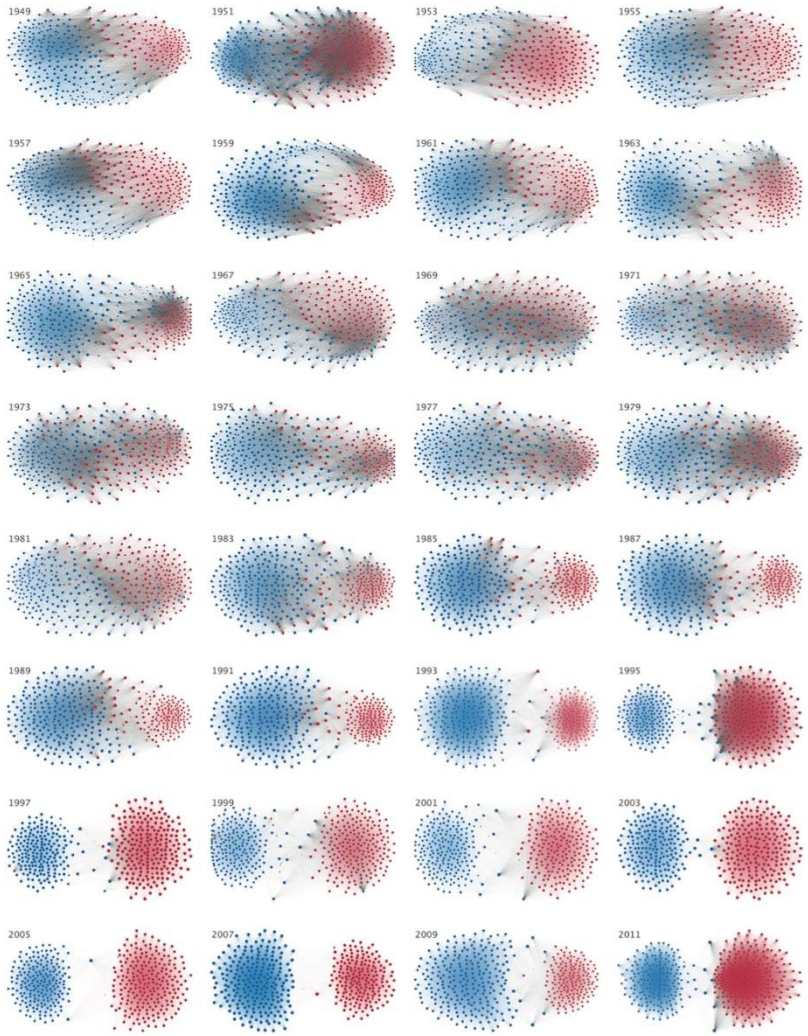




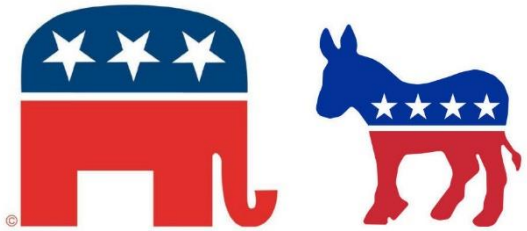




Political polarization



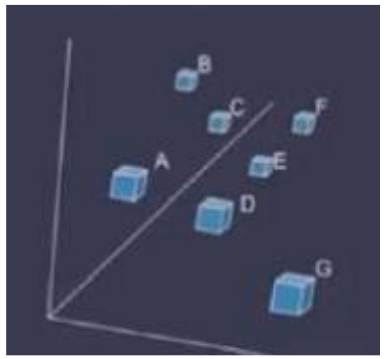
- Dots for each representative,
- Edges connecting pairs of representatives who vote together a given number of times



Demonstration of bipartisanship in USA parliament (Clio Andris et al. 2015)

1. Approach to visualization

- Identify the types of data attributes
- (Data encoding) Map data attributes to visual attributes
  -- Which visual attribute is the most effective for each data type?



X → x axis
Y → y axis
Z → z-axis

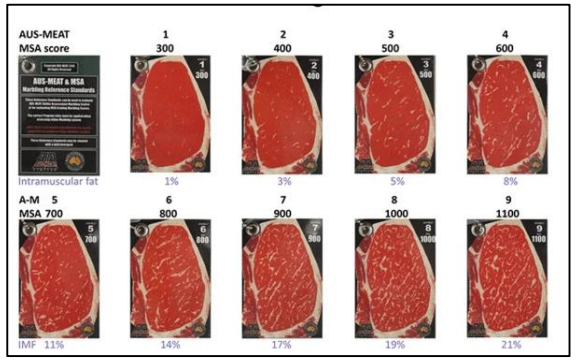
	X	Y	Z
A			
B			
C			
D			
E			
F			
G			

Numeric data

X → x axis
Y → y axis
Z → point size

Data Types: nominal, ordinal and quantitative

- Nominal (labels) (**N**)
  - -- Names, such as index, fruits' names
  - -- eq, neq
- Ordinal (**O**)
  - -- Quality (ranks, grades)
  - -- eq, neq, >



- Quantitative (**Q**): interval (zero arbitrary)
  -- Date, geographical location (latitude, longitude & elevation)
  -- eq, neq, >, – (difference)
  -- span, range
- Quantitative: ratio (zero fixed)

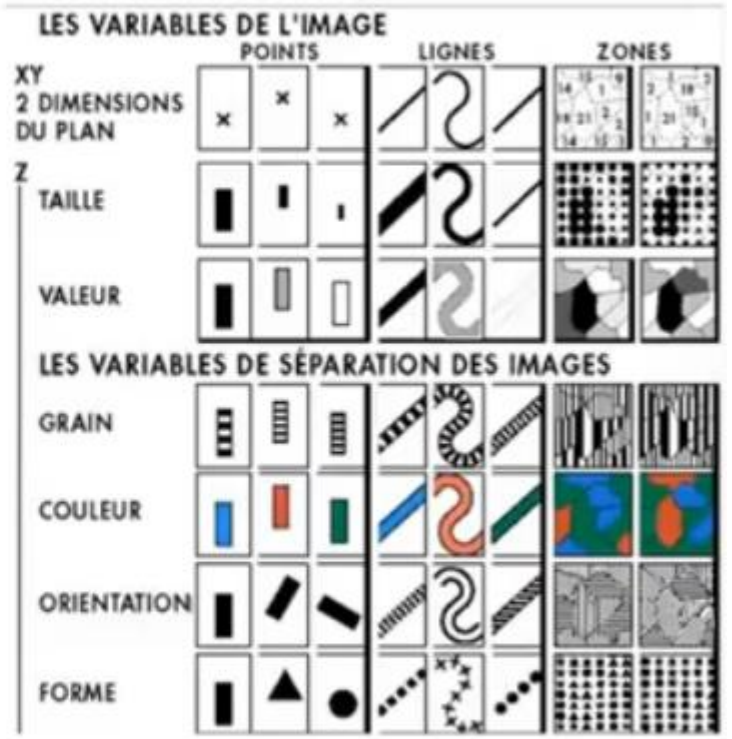
  -- Physical measurement: length, mass, temp, time (hrs).
  -- Counts, frequency
  -- eq, neq, >, -, / (ratio)
  -- span, range, ratios, proportions

- Qualitative
  - Nominal
  - Ordinal
- Quantitative
  - interval
  - ratio

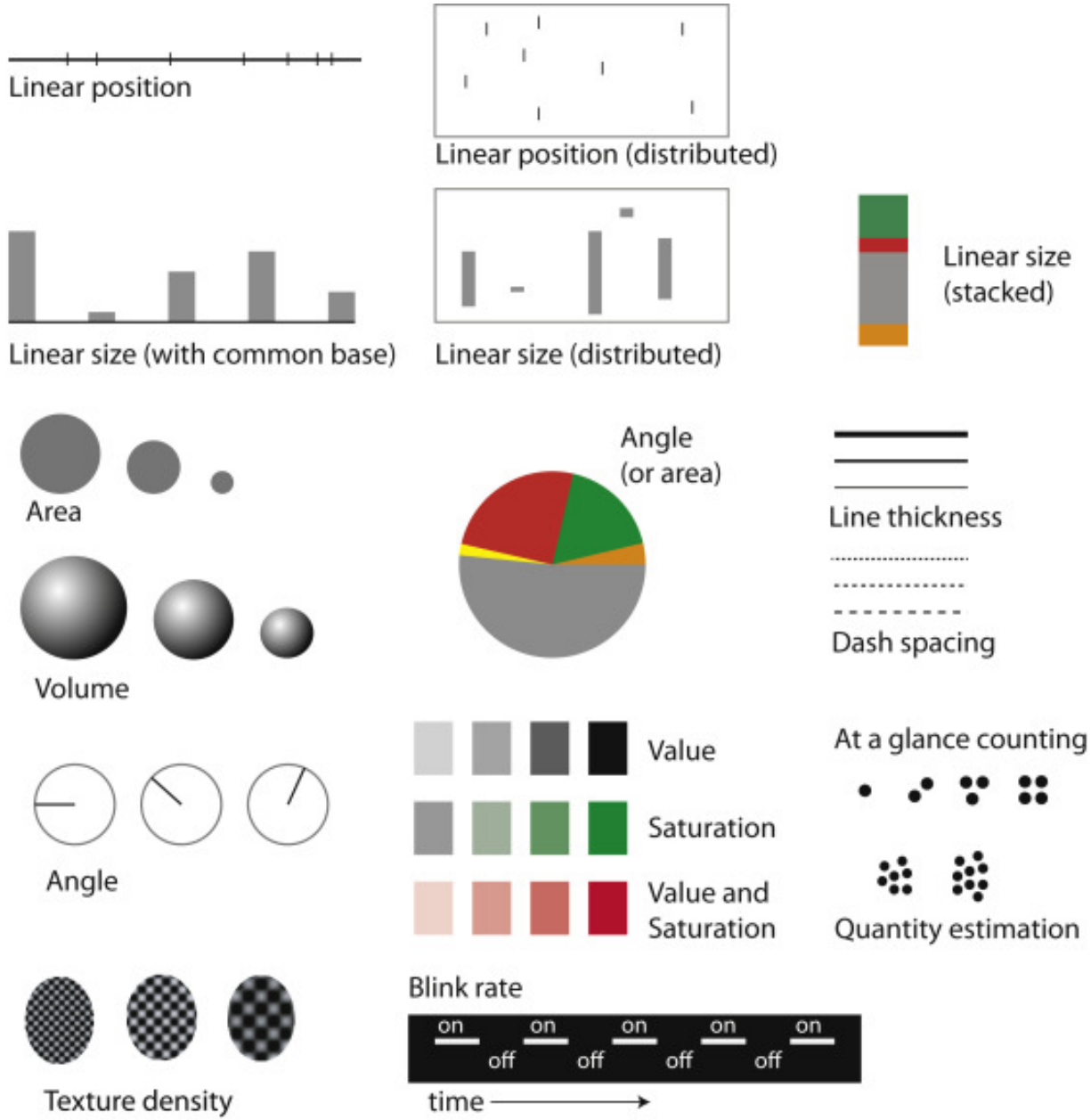
Bertin's Visual Attributes

- Position
- Size
- Value
- Texture
- Color
- Orientation
- Shape

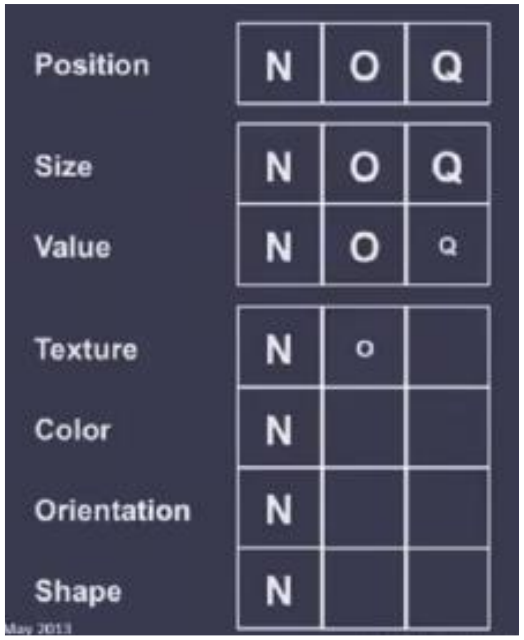
Texture: the feel, appearance, or consistency of a surface or substance

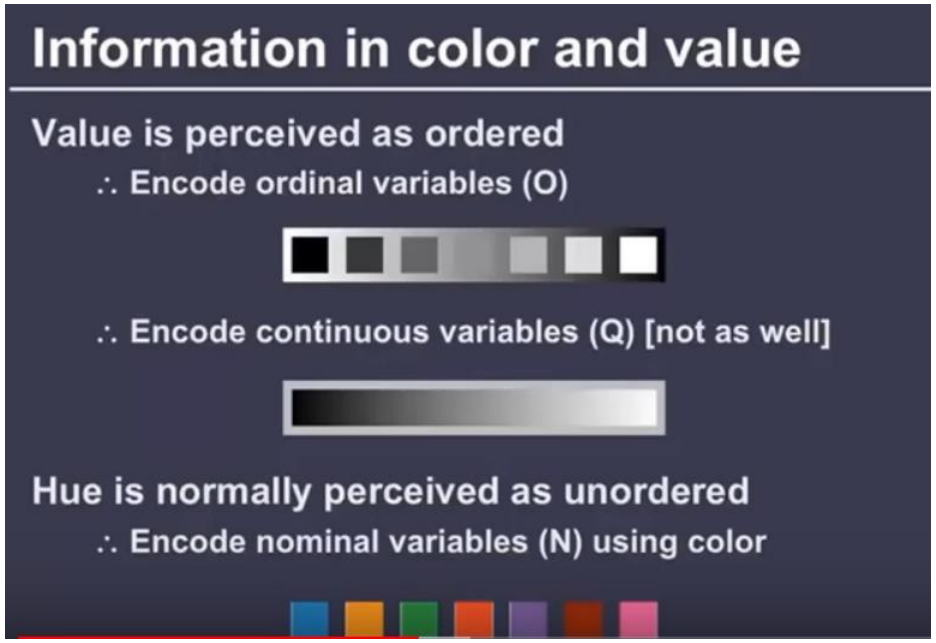


Bertin, *Semiology of Graphics*, 1967, 1983


Linear position
Linear position (distributed)
Linear size (with common base)
Linear size (distributed)
Linear size
(stacked)
Area
Angle
(or area)
Line thickness
Dash spacing
Volume
Value
Saturation
Value and
Saturation
At a glance counting
Angle
Quantity estimation
Blink rate
on
off
time
Texture density

Bertins' recommendation for mapping data types to visual attributes





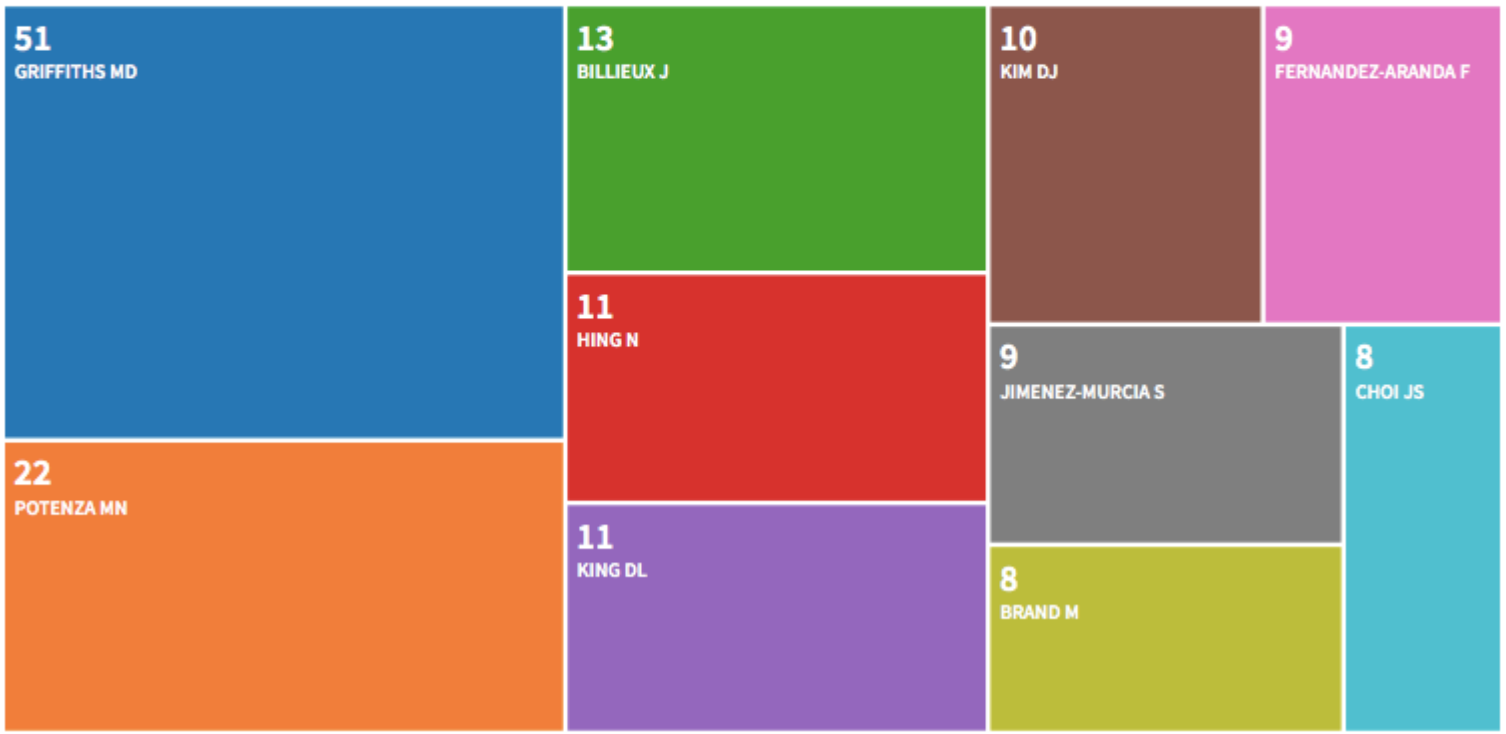
Value is the degree of a color's lightness and describes the luminous intensity of a color

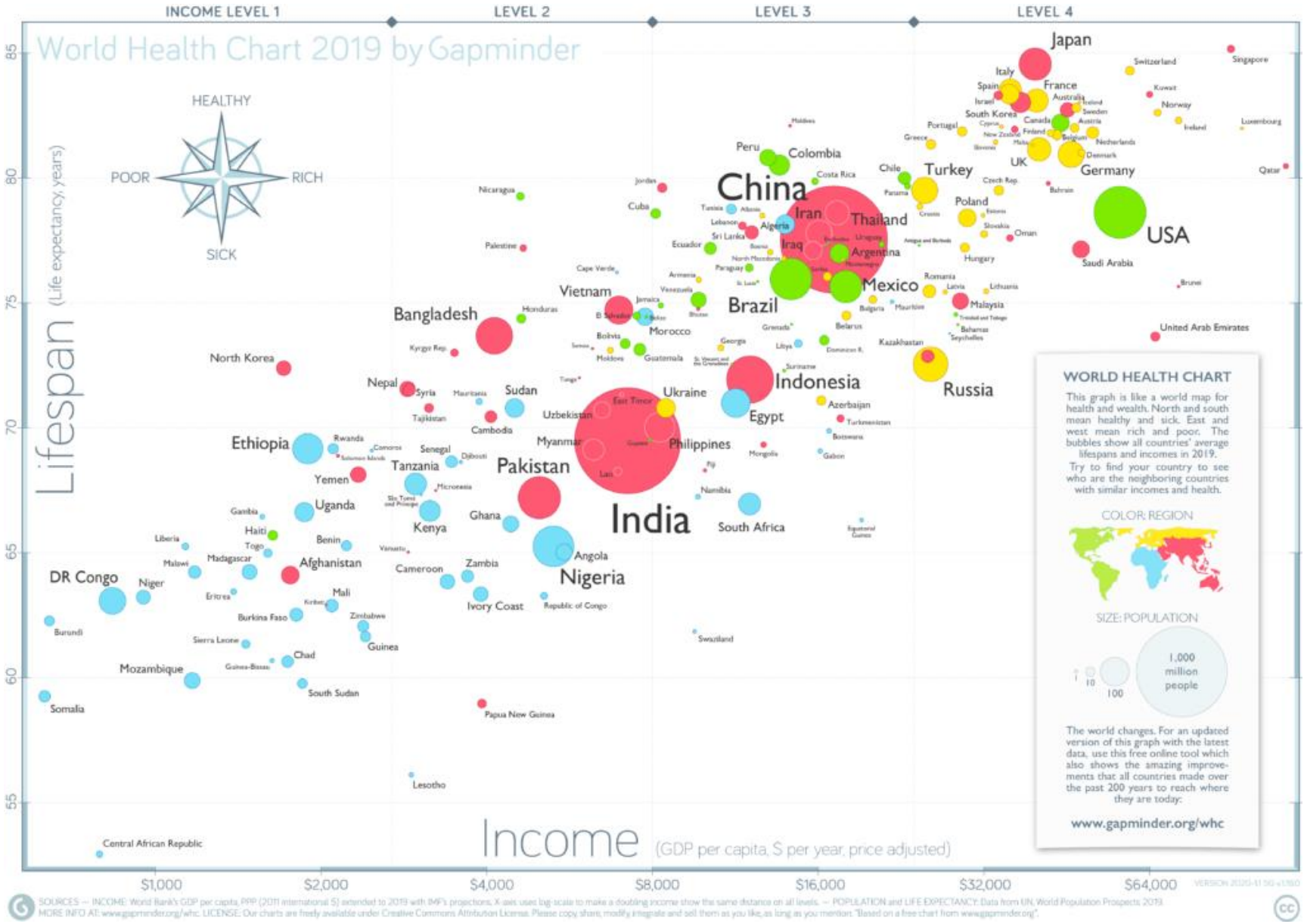
Quiz Identify what types of data and mapping in the following visualizations

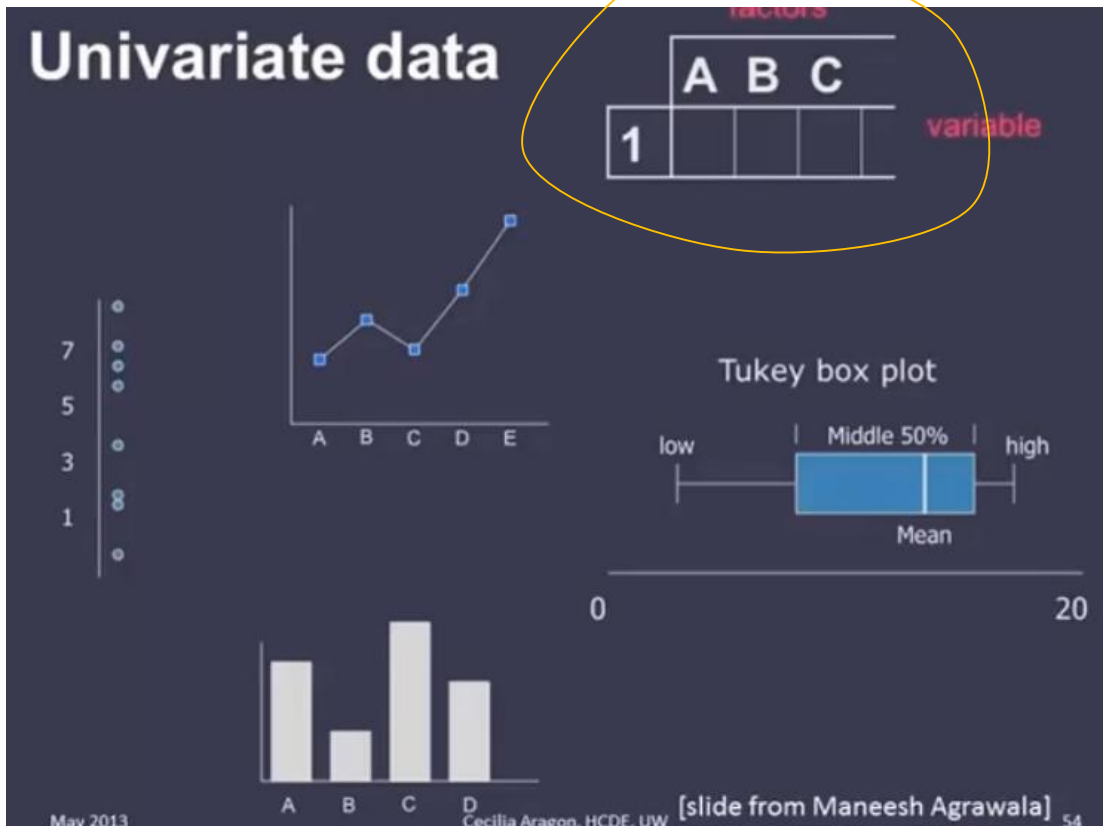
Showing 384 records for **PUBLICATION NAME:** (journal of behavioral addictions) *AND* **DOCUMENT TYPES:** (Article OR Review)

Create Citation Rep

Visualization **Treemap** | Number of results **10** | Download | Hide

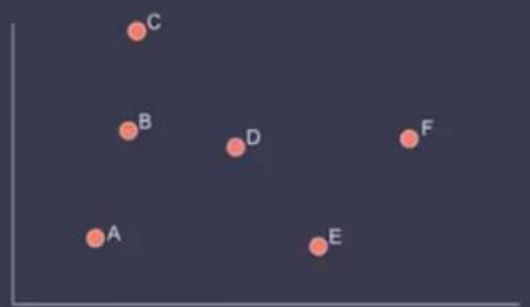




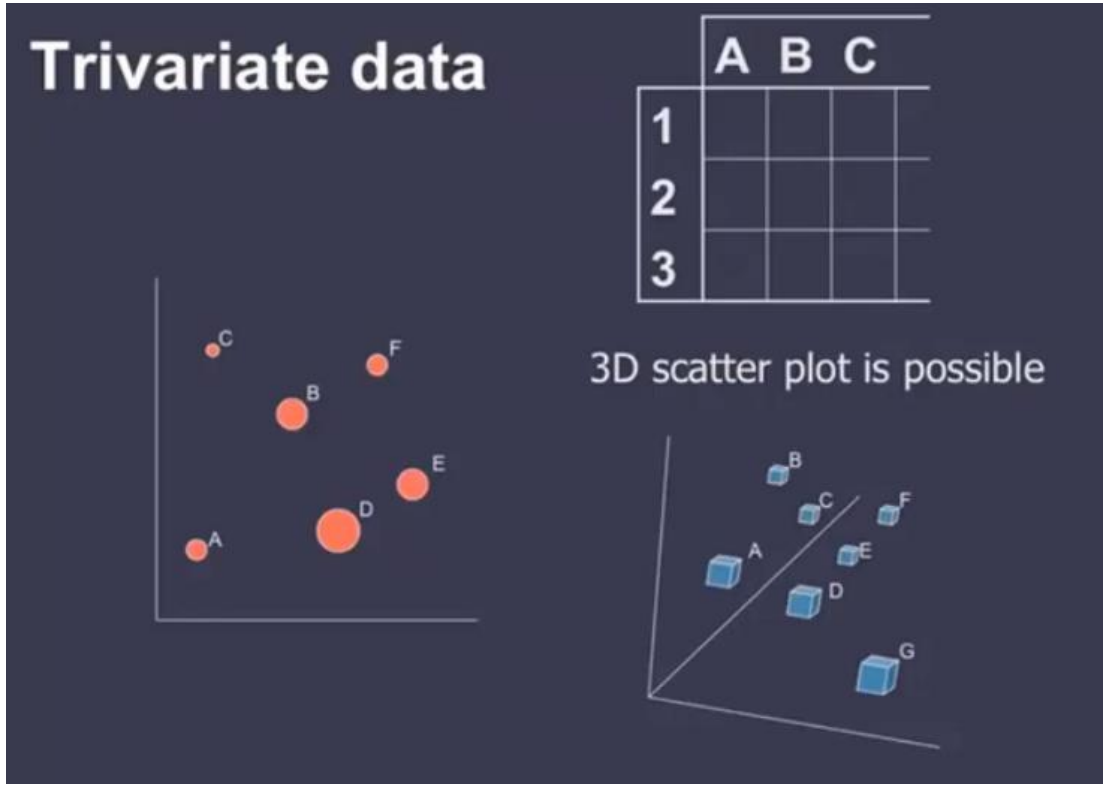

Univariate data
A B C
1
variable
7
5
3
1
A B C D E
Tukey box plot
low
Middle 50%
high
Mean
0
20
A B C D
May 2013
Cecilia Aragon, HCDE, UW
[slide from Maneesh Agrawala]
54

Bivariate data

	A	B	C	
1				
2				



Scatter plot is common



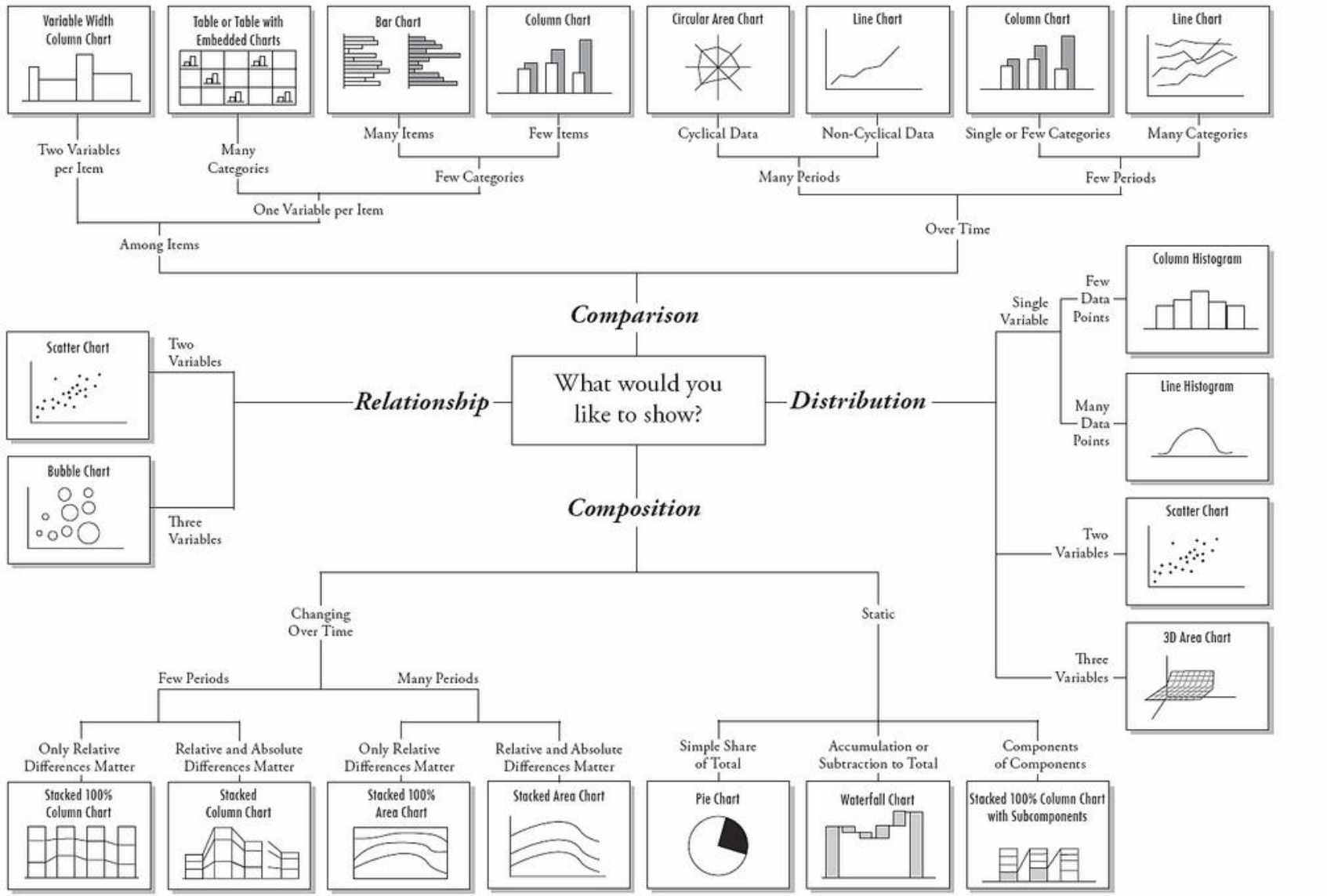
Two variables can map to points

-- Scatterplots, heat maps,

Third variable must use

-- Color, size, shape, …

Chart Suggestions—A Thought-Starter



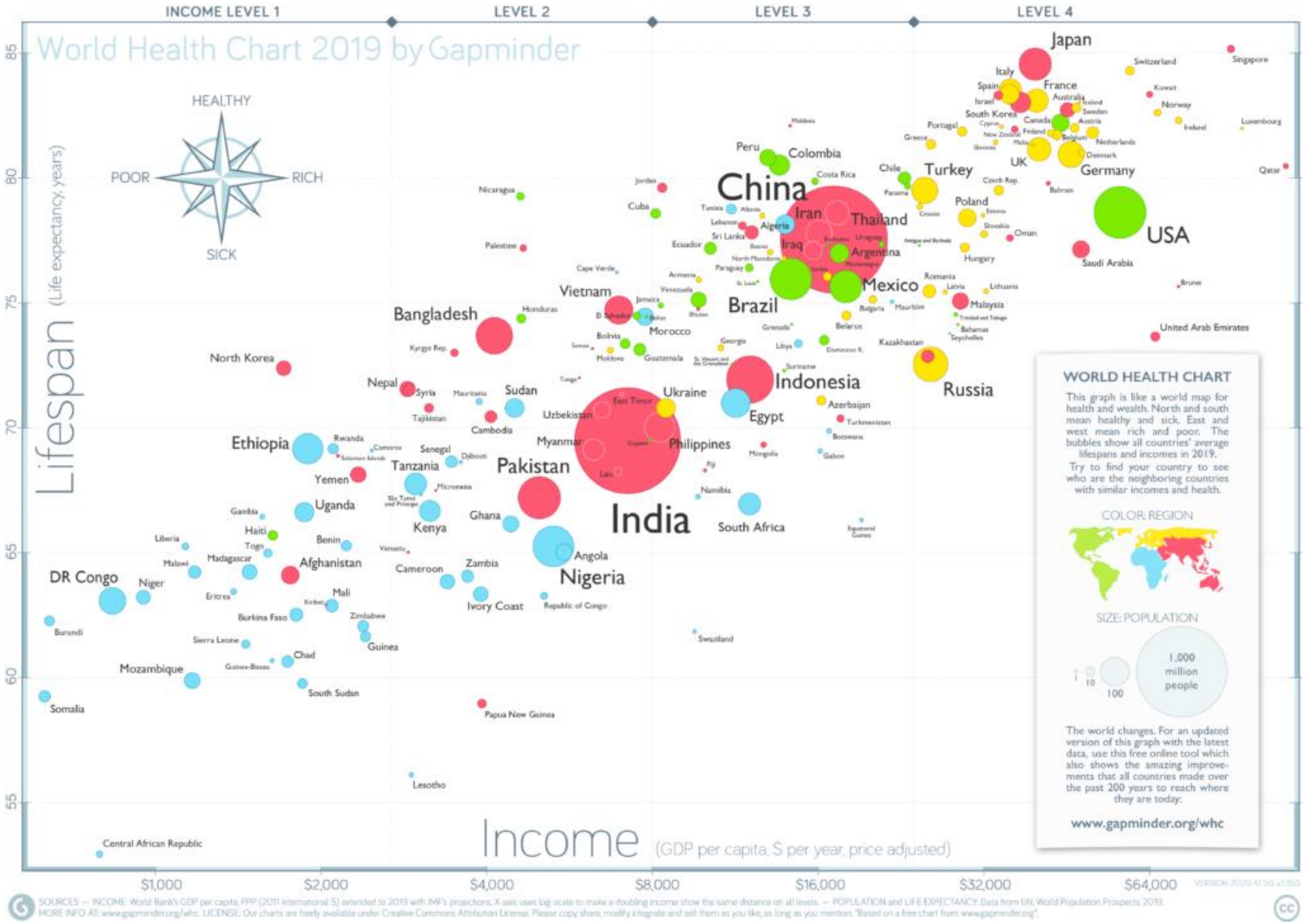
Multidimensional data

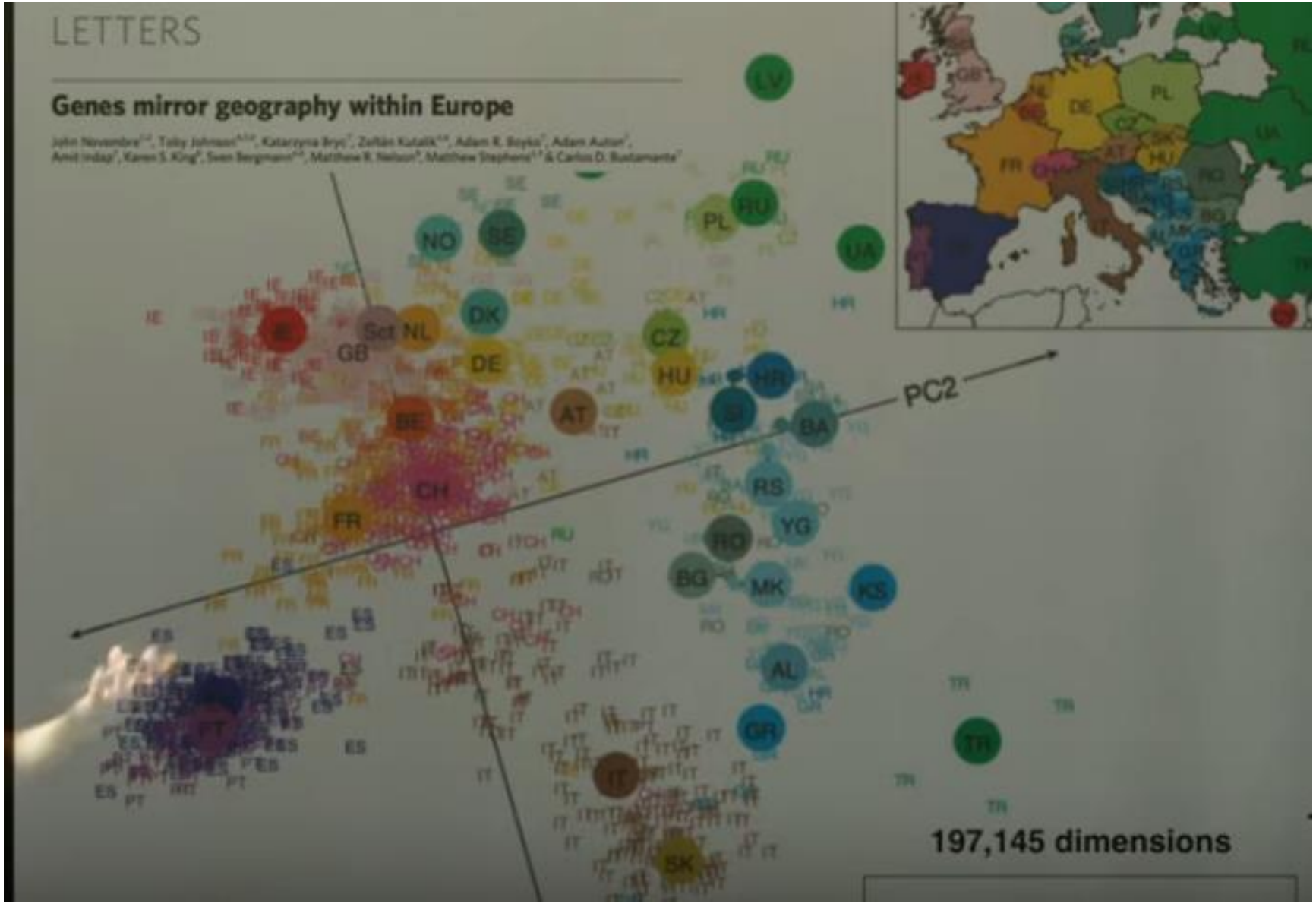
How many variables can be depicted in an image?

	A	B	C	
1				
2				
3				
4				
5				
6				
7				
8				

"With up to three rows, a data table can be constructed directly as a single image … However, an image has only three dimensions. And this barrier is impassible."

Bertin




LETTERS
Genes mirror geography within Europe
PC2
197,145 dimensions

Part 2. Effective Visualization

- 70% of body's sense receptors reside in our eyes
- Metaphors to describe understanding often refer to vision ("I see," "insight," "illumination")
- "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

[slide adapted from Marti Hearst]

Eyes vs. Cameras

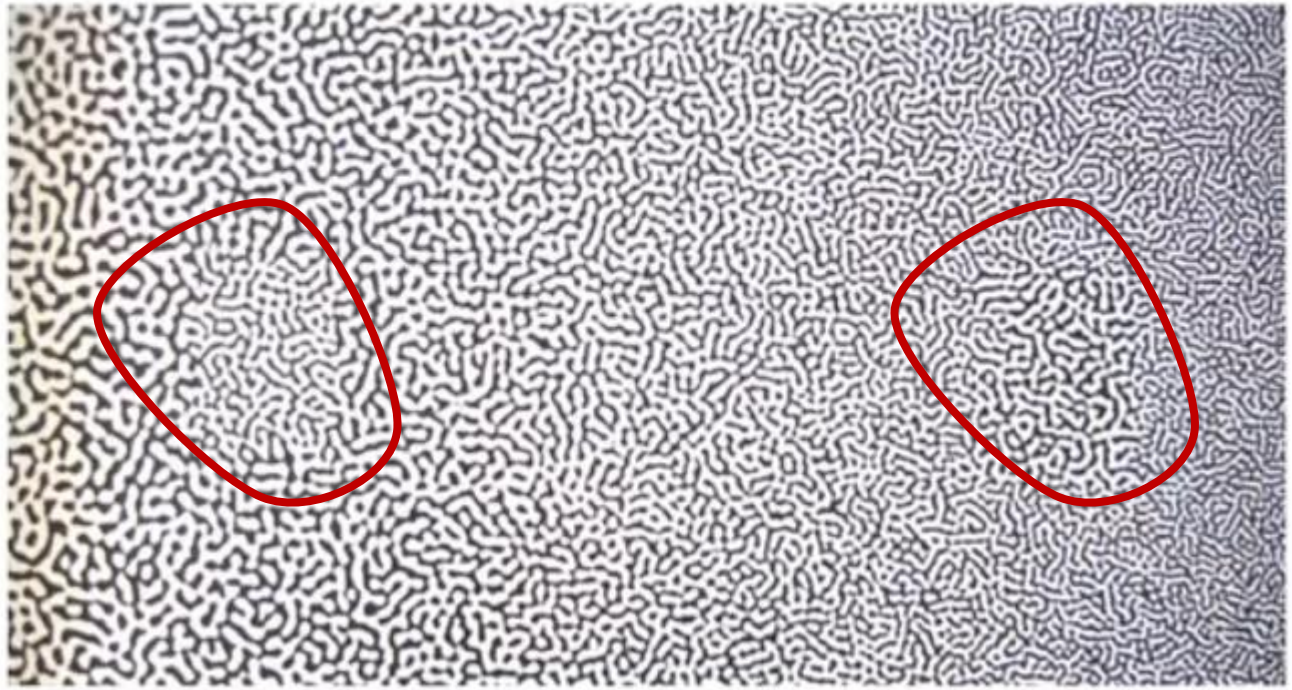
- Cameras
  - Good optics
  - Single focus, white balance, exposure
  - "Full image capture"

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

[http://www.usd.edu/psyc301/ChangeBlindness.htm]

How to Use Perceptual Properties

- Information visualization should cause what is meaningful to stand out



Where are distinguished regions?

How many 5's?

385720939823728196837293827

382912358383492730122894839

909020102032893759273091428

938309762965817431869241024

How many 5's?

38572093982372819683729382 7

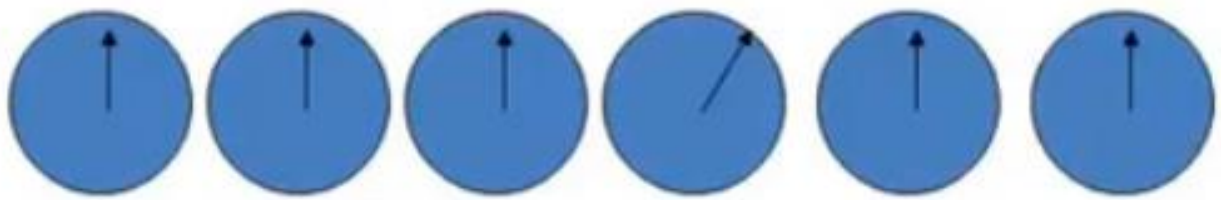
3829123583834927301228948 39

9090201020328937592730914 28

9383097629658174318692410 24

Cockpit dials

Detection of a slanted line in a sea of vertical lines is preattentive



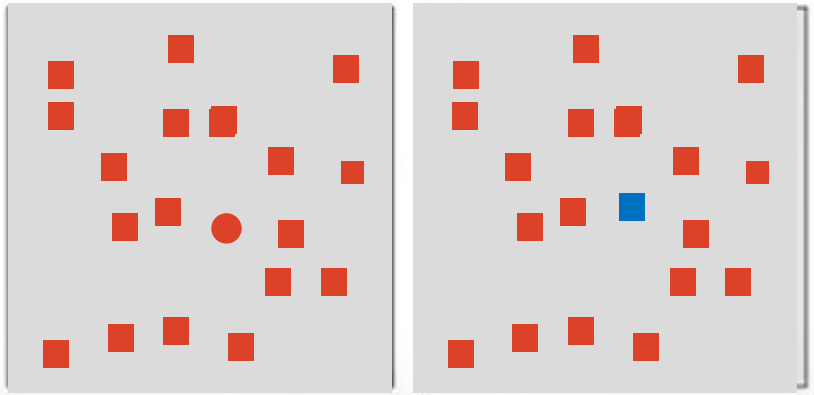
Preattentive Processing

- A limited set of visual properties are processed preattentively
  - (without need for focusing attention).

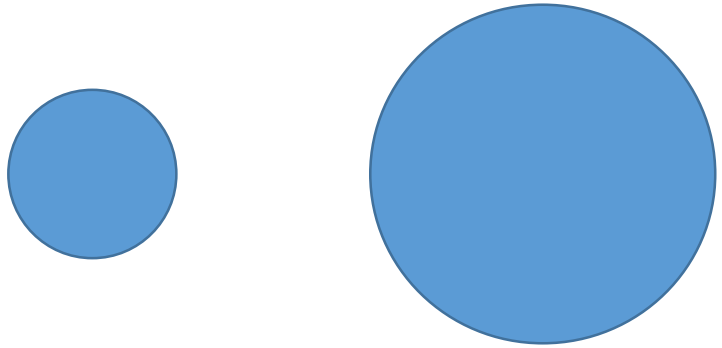
- This is important for design of visualizations
  - What can be perceived immediately?
  - Which properties are good discriminators?
  - What can mislead viewers?

Conjunction of attributes

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



- Estimating length is easy
- Estimating area and volume is not that easy



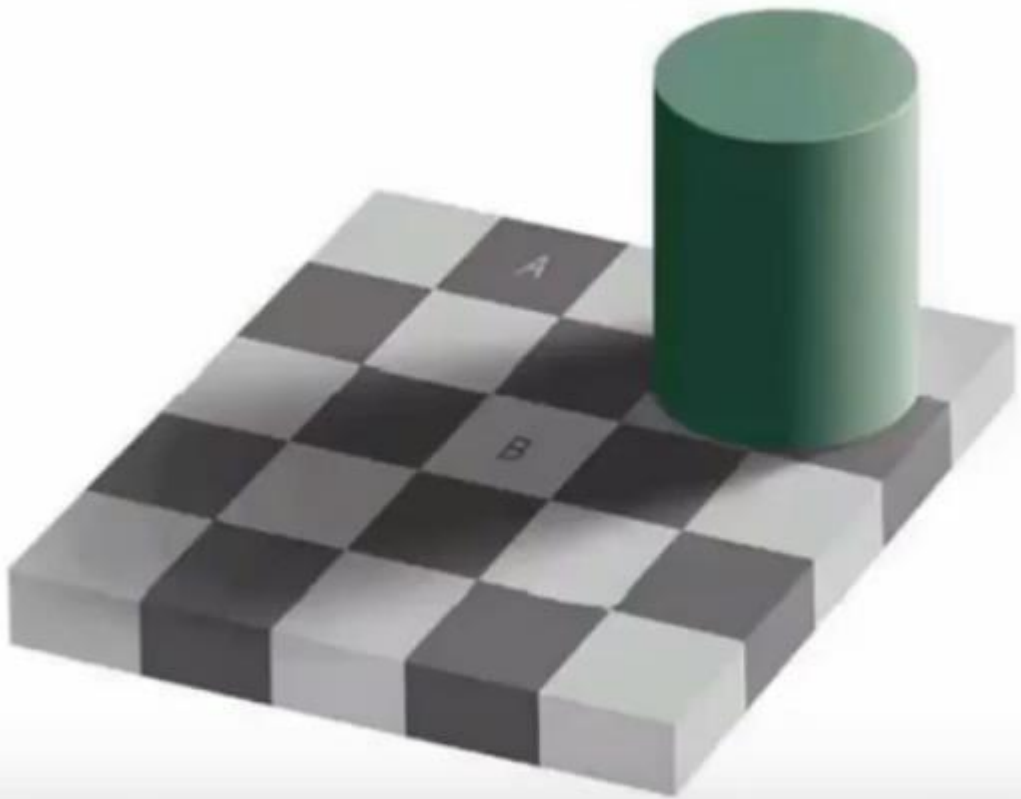
Stevens' Power Law

Experimental results for perceptual estimation:

Length	.9 to 1.1
Area	.6 to .9 (underestimation)
Volume	.5 to .8 (even more underestimation)

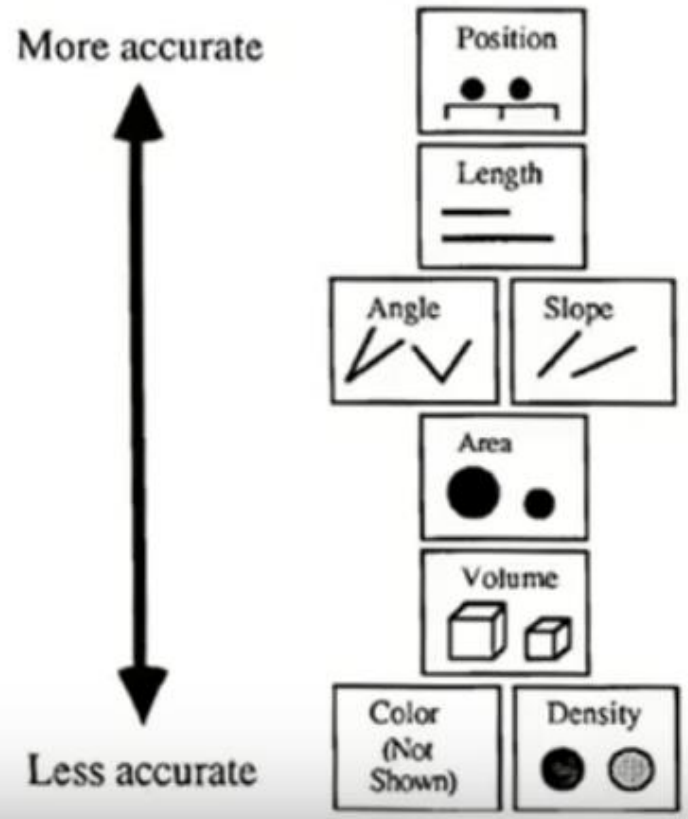
Color is relative

Visual perception is not just camera work



Square A is darker than B, right?

Perceptual properties



Mackinlay, A Presention Tool, 1986

Effective Visual Encoding

- Mapping data to visual attributes:
- 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

Adapted from Mackinlay, APT (A Presentation Tool), 1986

3. Are your figures good?

- 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 the following.

Marti Hearst

11 Overarching Questions (M. Hearst)

A. Is the design visually appealing/pleasing?

B. Is it immediately understandable?
If not, how about after a short period of exam.?

C. Does it provide insight or understanding that was not obtainable with the original text/table?

D. 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?

E. Does the visualization reveal trends, patterns, gaps, and/or outliers?
Can the viewer make effective comparisons?

F. Does the visualization successfully highlight important information, while providing context for that information?

G. Does it distort the information? If it transforms the information in some way, is this misleading or helpfully simplifying?

H. Does it omit important information?

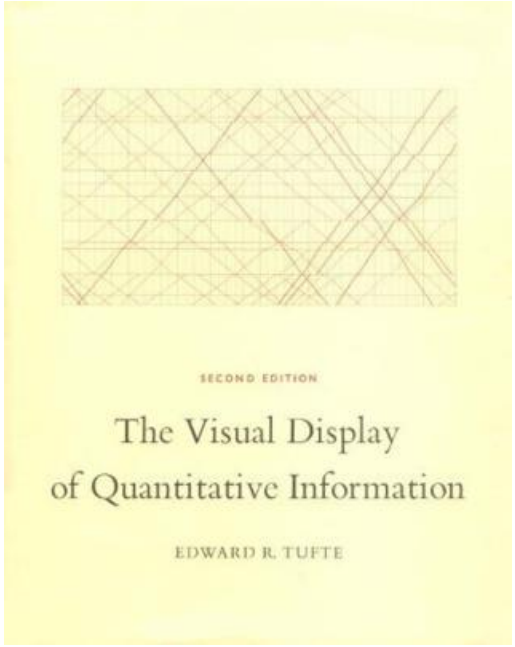
I. Is it memorable?

J. Does it use visual components properly?

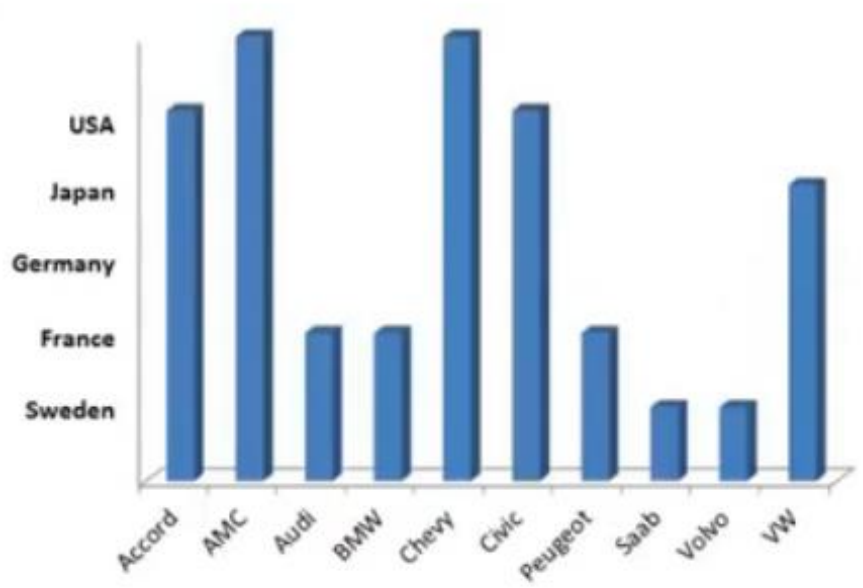
K. Does it use labels and legends appropriately?

Edward Tufte’s Criteria for Good Visualizations

- Maximize data-ink ratio
- Minimize lie factor
- Minimize chartjunk
- Use proper scales and clear labelling



Is this an effective visual representation?



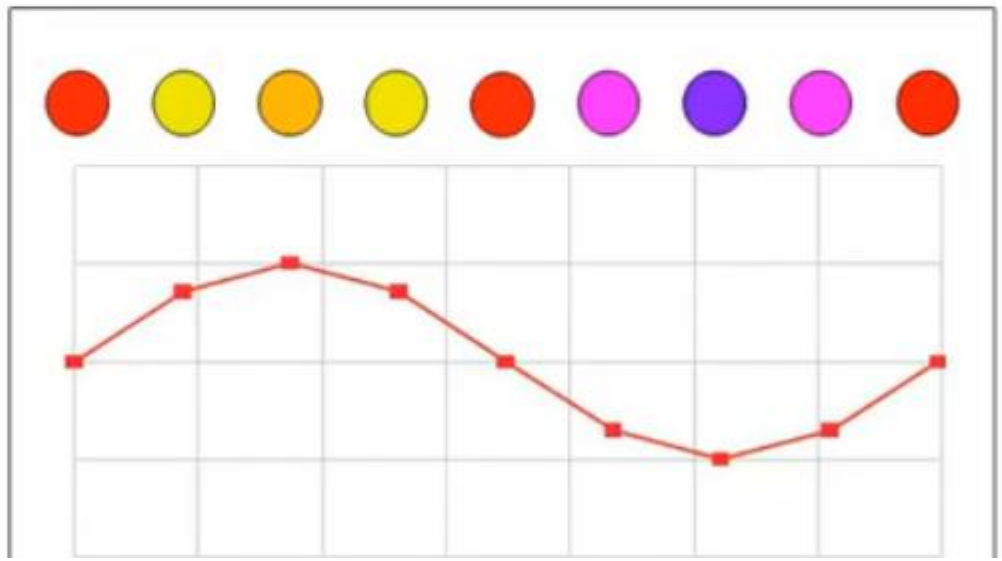
- Shade ink is redundant
- Labelling is not proper
- Bar length for countries
  indicating an ordering among
  countries.

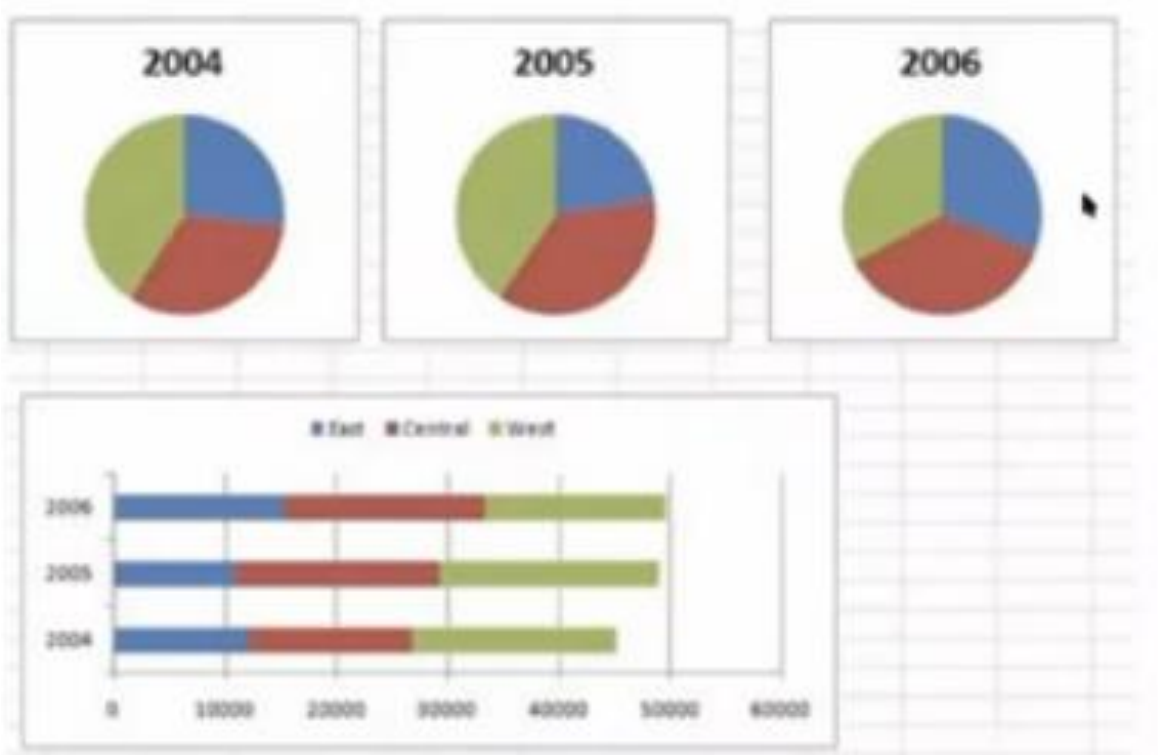
Design criteria: Effectiveness

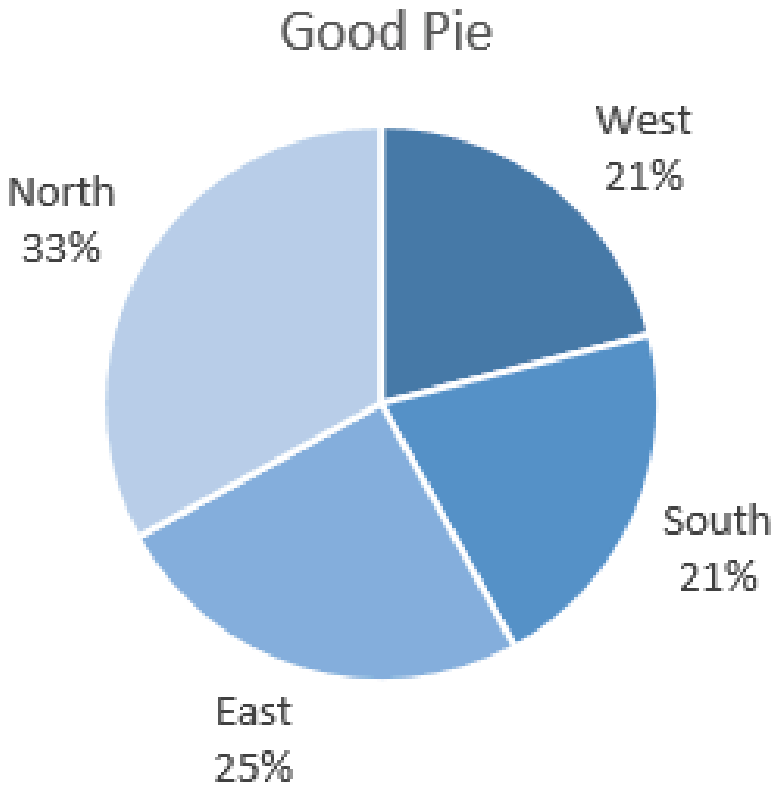
- Faster to interpret
- More distinctions
- Fewer errors

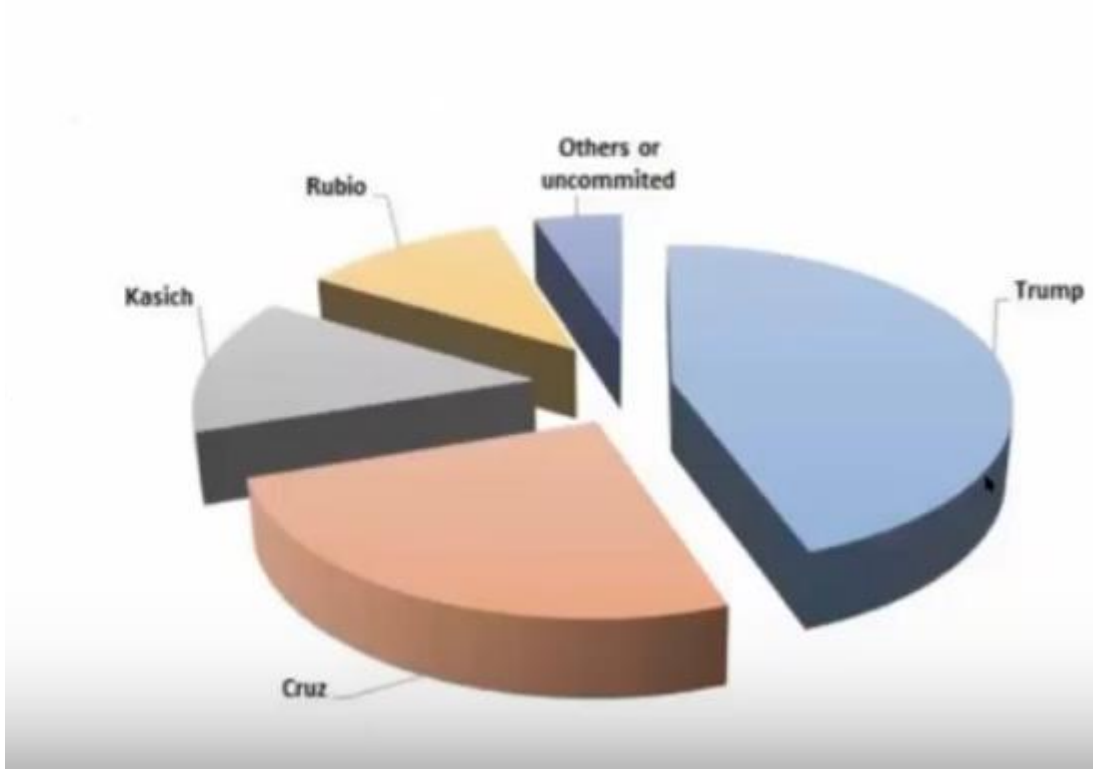
This?

Or this?

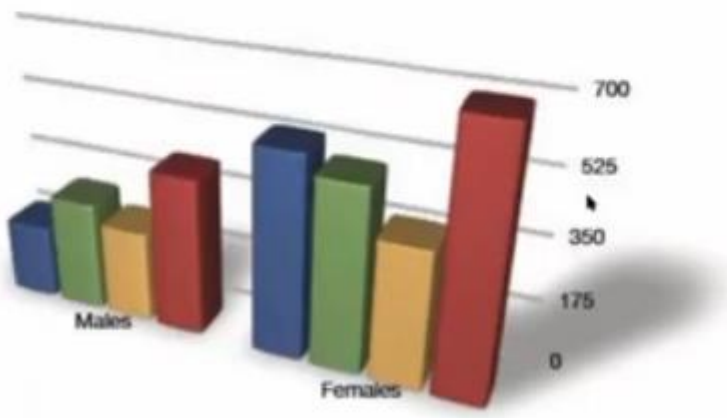


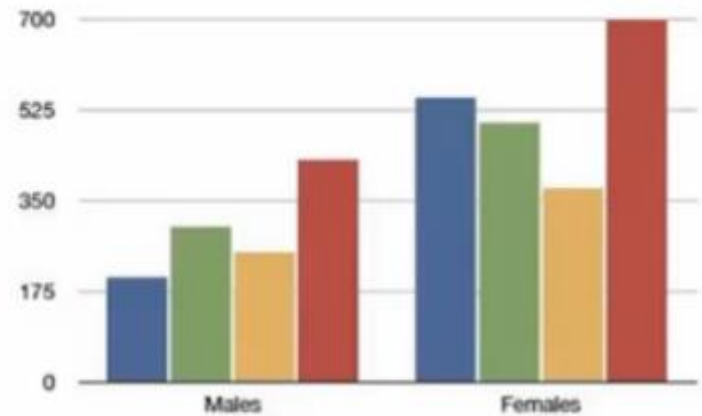

2004
2005
2006
East
Central
West
2006
2005
2004
0
10000
20000
30000
40000
50000
60000


Good Pie
West
21%
North
33%
South
21%
East
25%


Others or
uncommited
Rubio
Kasich
Trump
Cruz

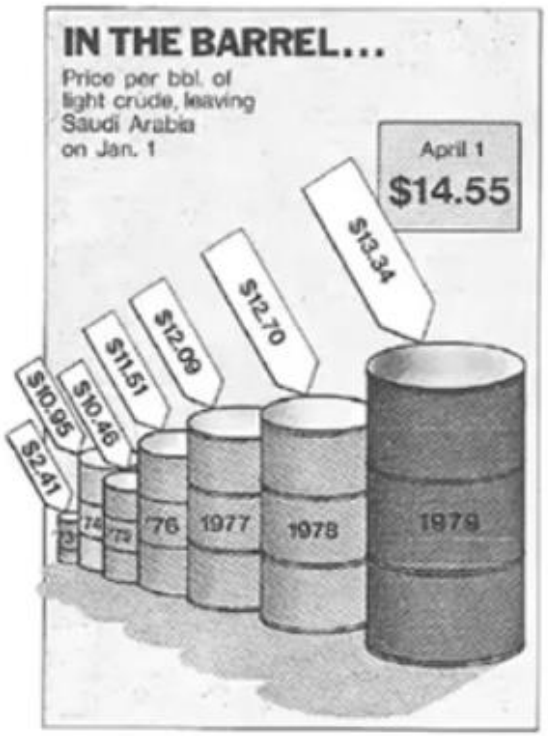
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$





(size of effect in graphic) / (size of effect in data)

- 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!

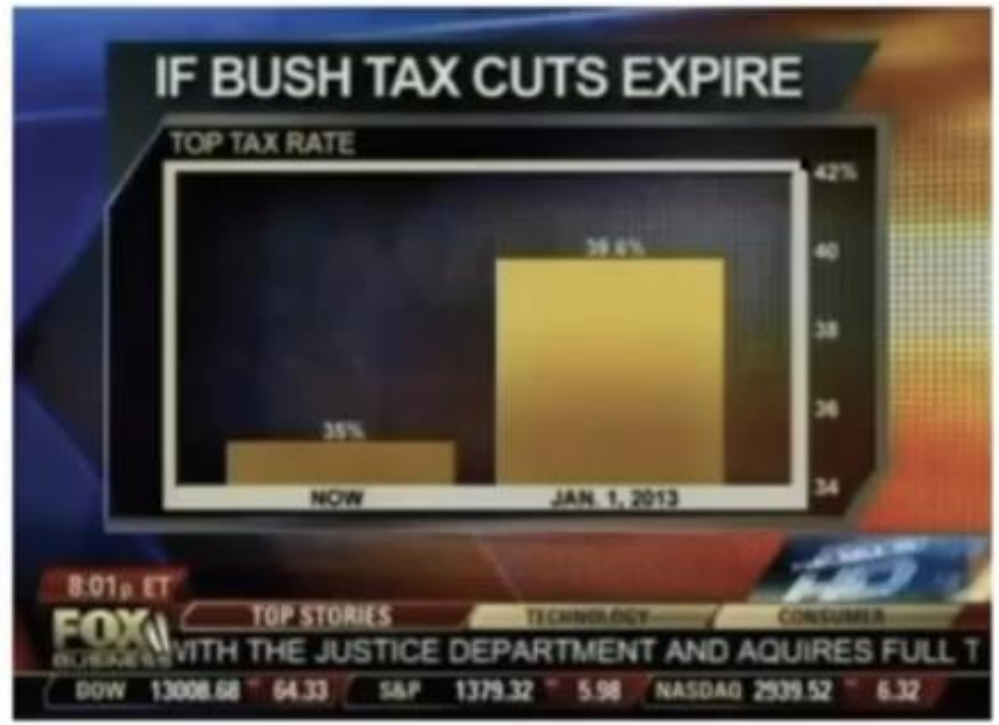


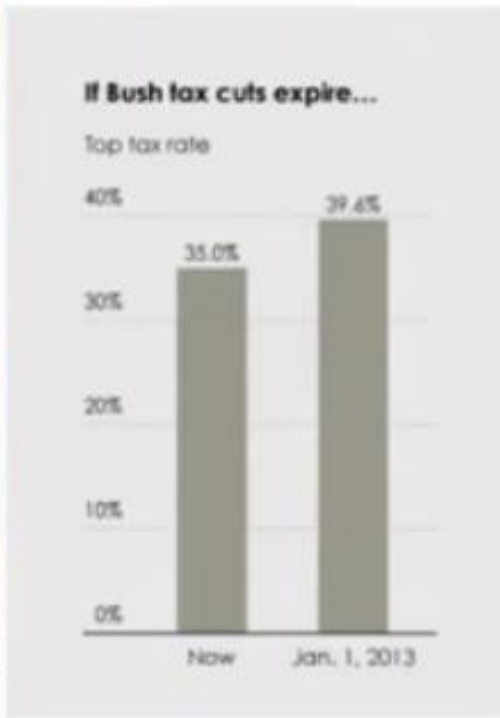
[Tufte, Edward R (1983), *The Visual Display of Quantitative Information*, Graphics Press,

Always start bar graphs at zero.

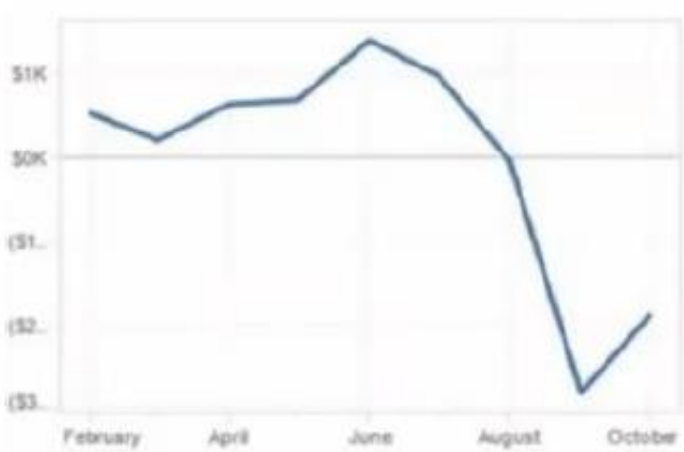
Always properly label your axes.

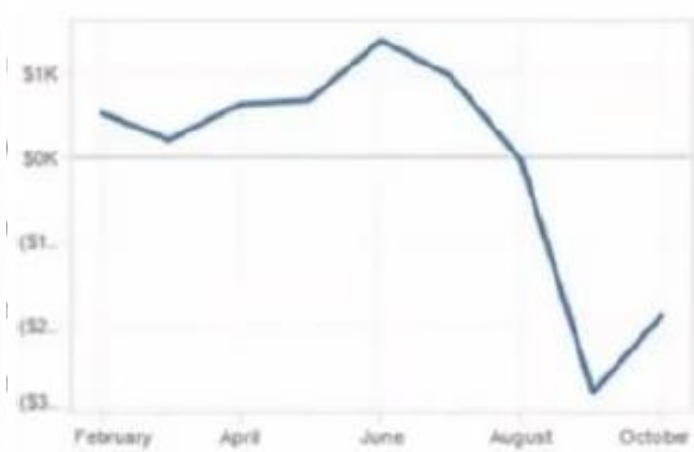
Use continuous scales: linear or labelled!

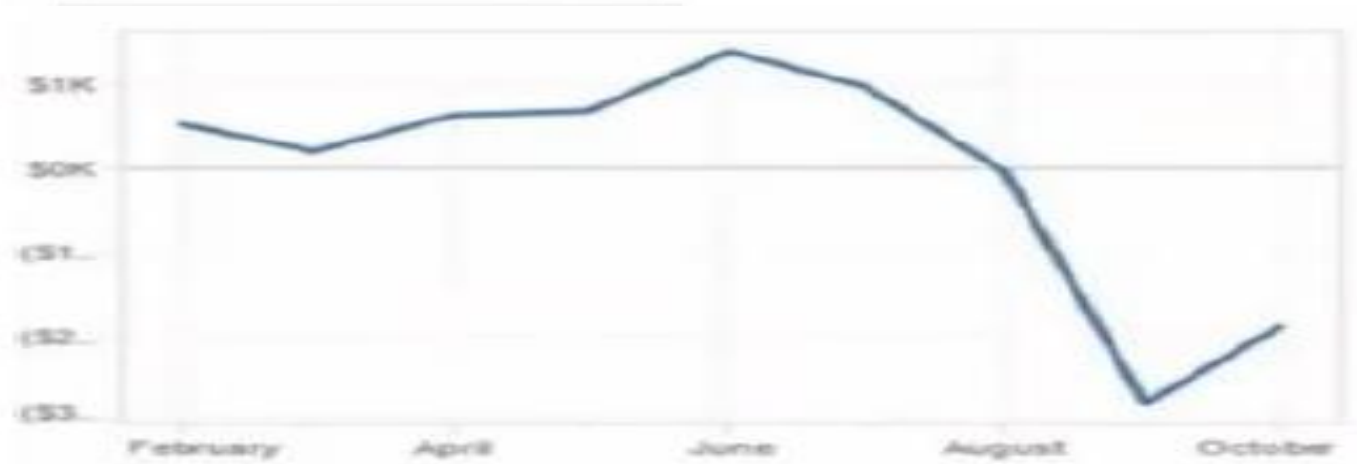




Aspect Ratios and Lie Factors



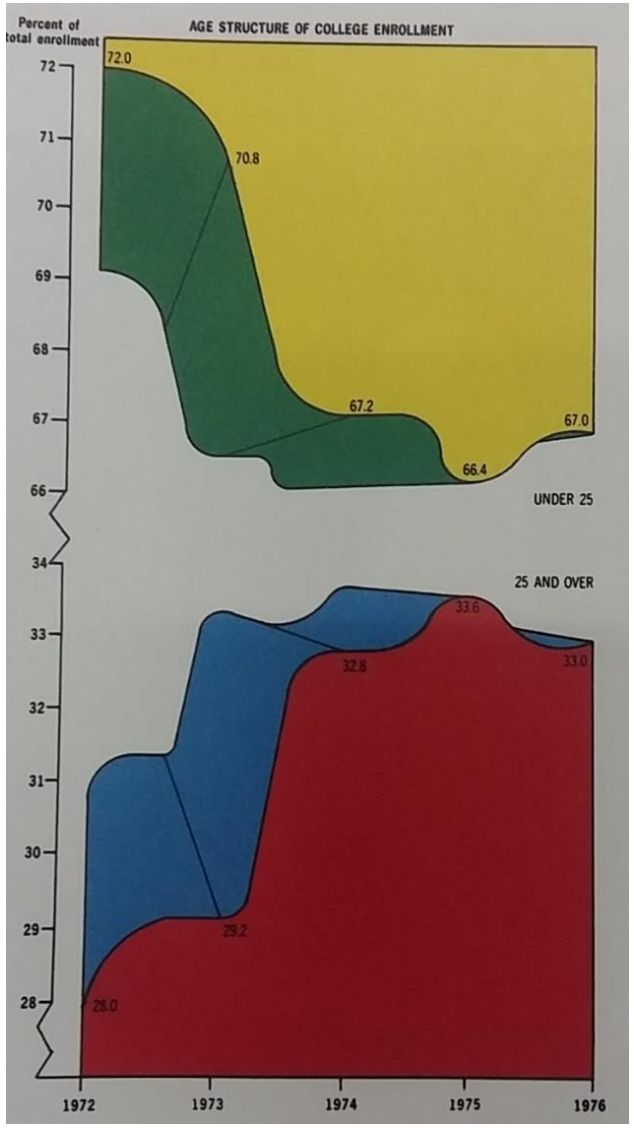




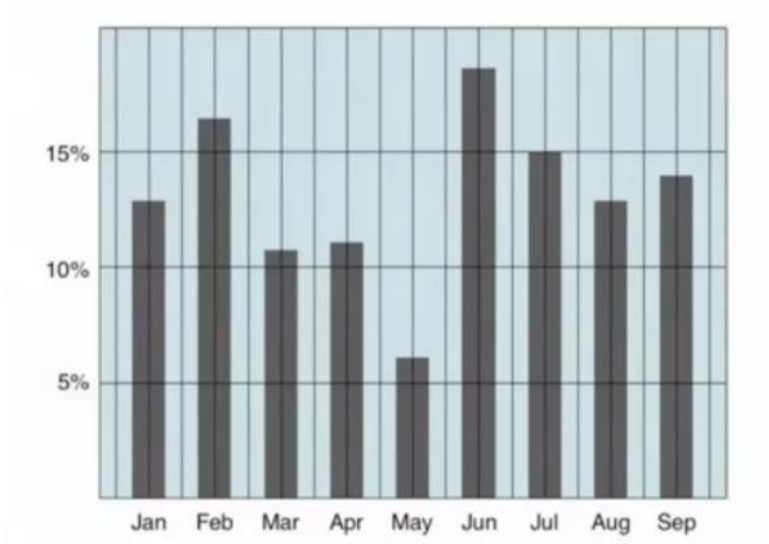

POOL DROWNING
66.6% CORRELATION
NICOLAS CAGE
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
6 films
4 films
2 films
0 films

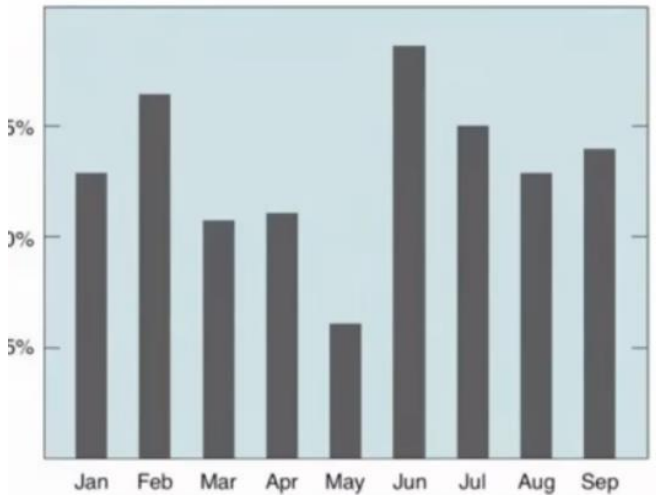
Chartjunk

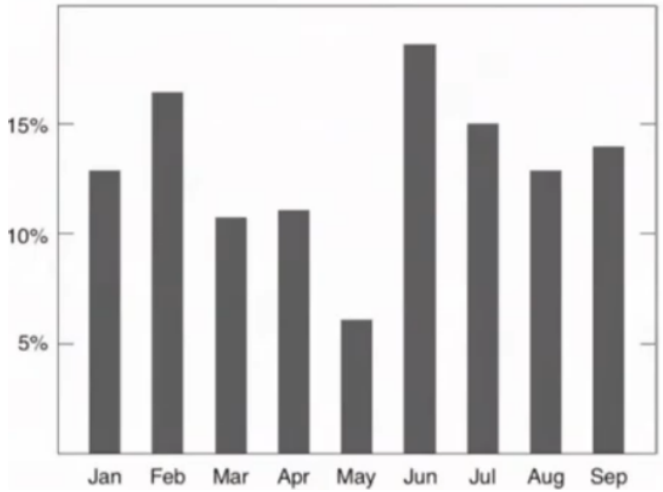
	Under 25 (%)	25 or over (%)
1972	72	28
1973	70.8	29.2
1974	67.2	32.8
1975	66.4	33.6
1976	67	33

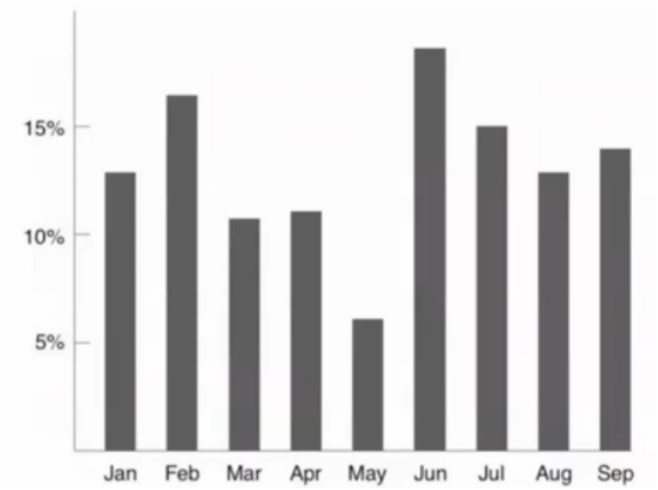


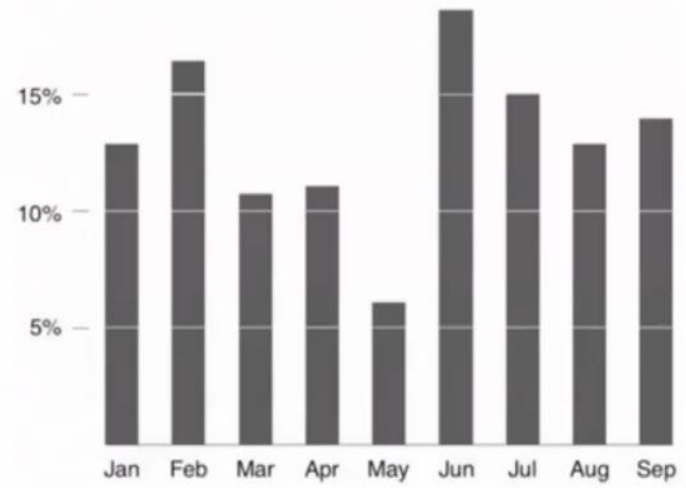
Can you Simplify this Plot?

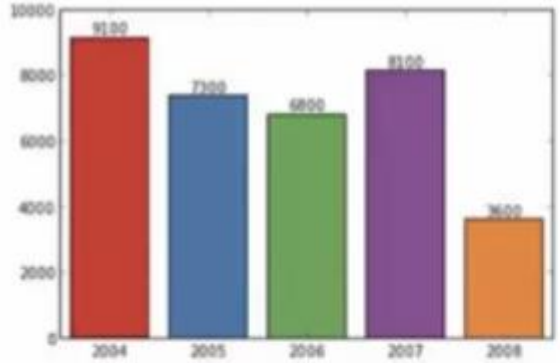


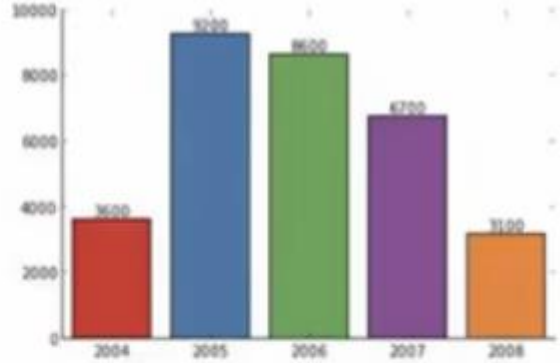


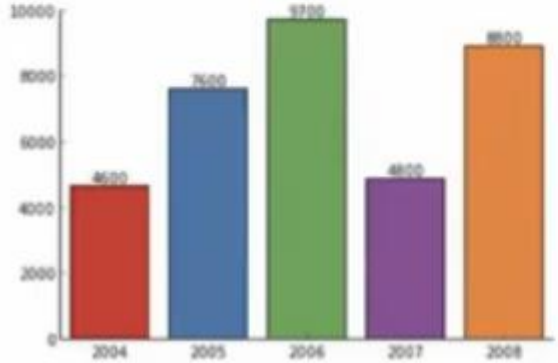


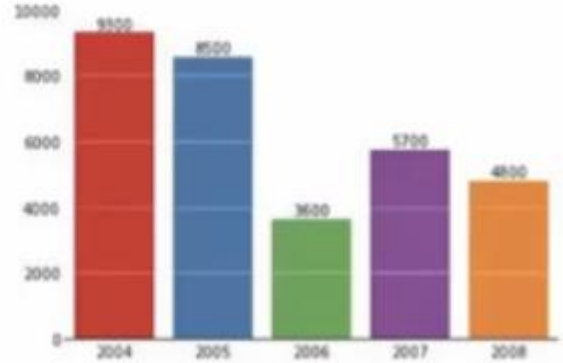





10000
8000
6000
4000
2000
0
9100
7300
6800
8100
3600
2004
2005
2006
2007
2008


10000
8000
6000
4000
2000
0
3600
9200
8600
6700
3100
2004
2005
2006
2007
2008

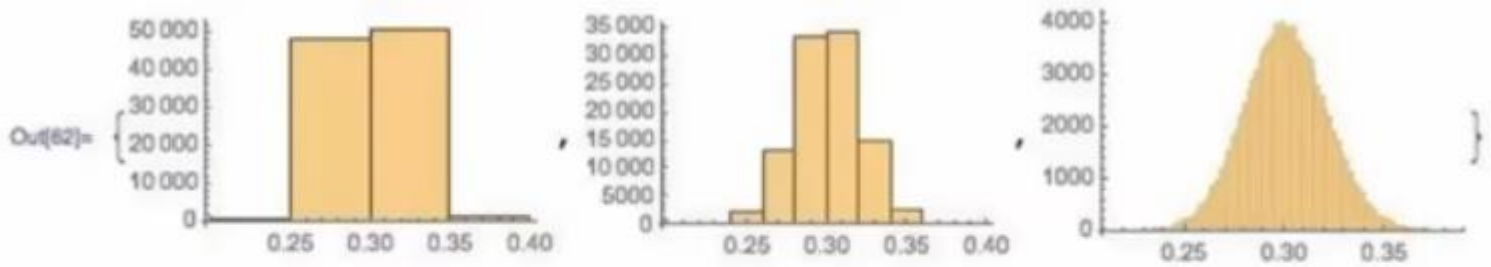

10000
8000
6000
4000
2000
0
4600
7600
9700
4800
8800
2004
2005
2006
2007
2008


10000
8000
6000
4000
2000
0
9500
8500
3600
5700
4800
2004
2005
2006
2007
2008

Histograms: Bin Size

```
In[62]:= {Histogram[d, 5], Histogram[d, 10], Histogram[d, 100]}
```

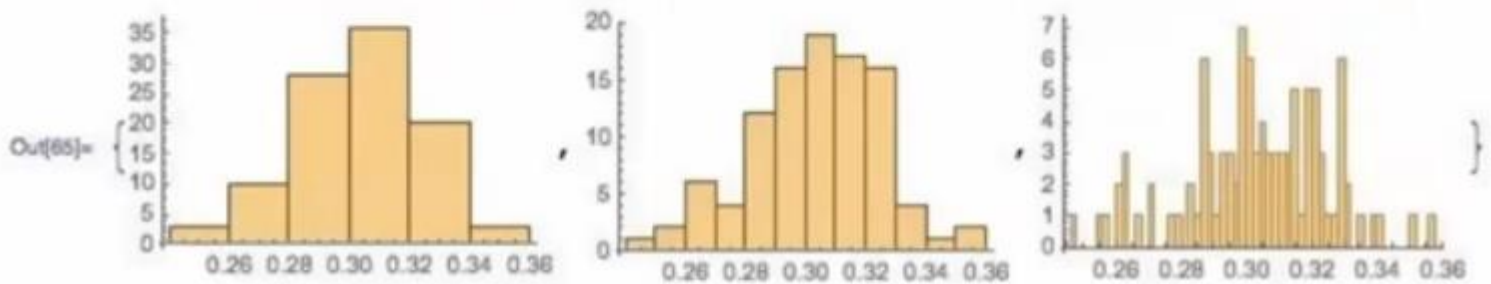
Out[62]=




```
d100 = Take[d, 100];

In[65]:= {Histogram[d100, 5], Histogram[d100, 10], Histogram[d100, 100]}
```

Out[65]=



Summary

- The knowledge of visual perception can aid the design process
- Understanding low-level mechanisms of the visual processing systems can result in improved displays.

Stephen Few http://www.perceptualedge.com/