Science of Visualization pt. 1

Summarized from Coursera Course:
Data Visualization by Professor John C. Hart,
University of Illinois at Urbana-Champaign

What we are talking about

Topics covered:

- Overview of Visualization
- How the Mind Processes Images
- Visualization Guidelines

Modes of Visualization

Interactive Visualization

- Used for discovery
- Intended for a single investigator or collaborators
- Rerenders based on input
- Prototype quality

Presentation Visualization

- Used for communication
- Intended for large group or mass audience
- Does not support user input
- Highly polished

Interactive Storytelling

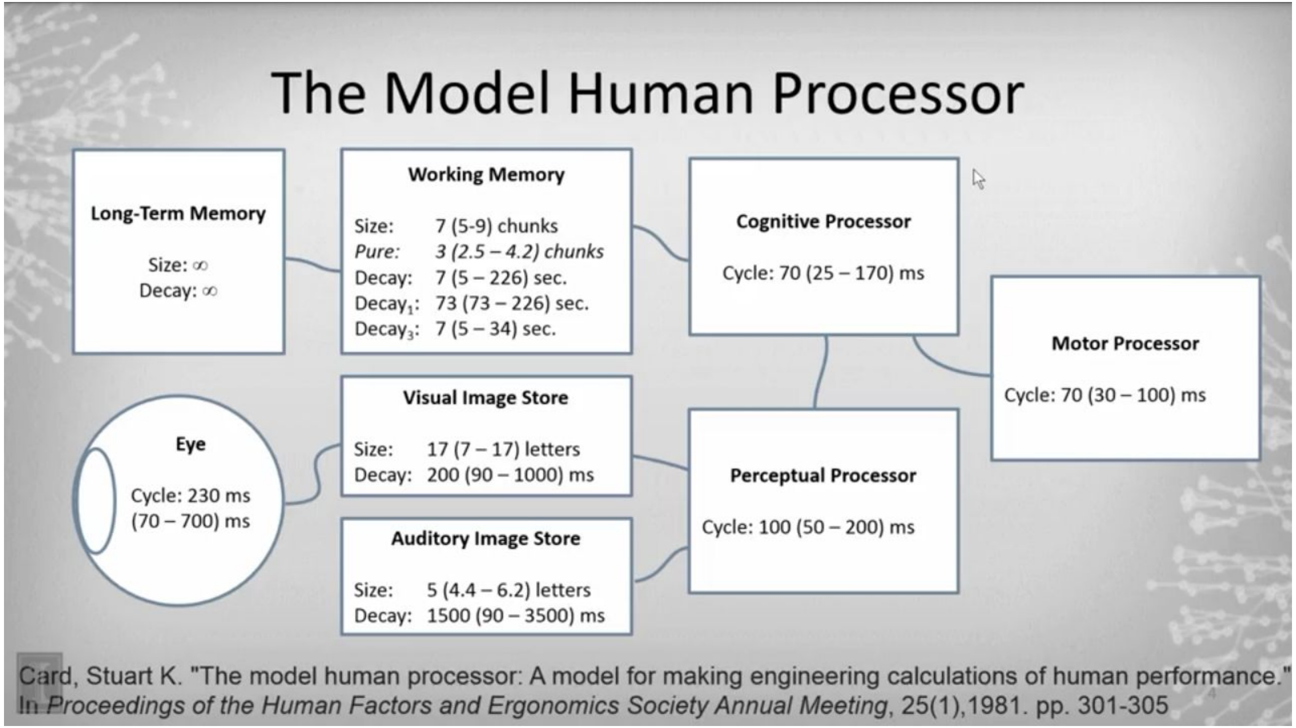
- Presentations via interactive webpages

Modes of Visualization

Visualization Mode	User Interaction	Graphics Rendering	Target	Medium
Interactive Visualization	User controls everything, including dataset	Real-time rendering	Individual or collaborators	Software or internet
Interactive Storytelling	User can filter or inspect details of preset datasets	Real-time rendering	Mass audience	Internet or kiosk
Presentation Visualization	User only observes	Precomputed rendering	Colleagues, mass audience	Slide shows, video

Model for Perception

- Cognitive Processor
  - Brain
- Perceptual Processor
  - Media Processor
- Long-Term Memory
  - Like Disk Store
- Working Memory
  - Like Ram
- Visual Image Store
  - Image Buffer from Eye
  - 7-17 letters/numbers
  - Decay: 90-1000ms
- Motor Processor
  - Movement Coordinator
- Eye
  - Camera
  - 70-700ms



Reading

- *Saccades* - eye scans forward
- *Fixations* - eye is still
  - Perception happens
  - 94% of the time
- We do not perceive saccades
- 9pt,12pt equally legible
- We still read books faster than computers

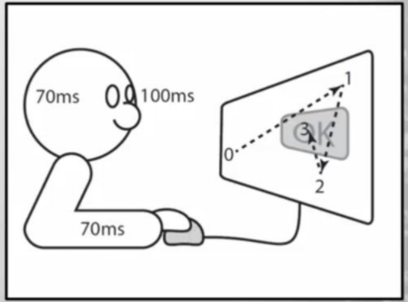
Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

Fitt's Law

- Kinesthesis: We know where our limbs are
- Larger movements are faster but less accurate than smaller ones

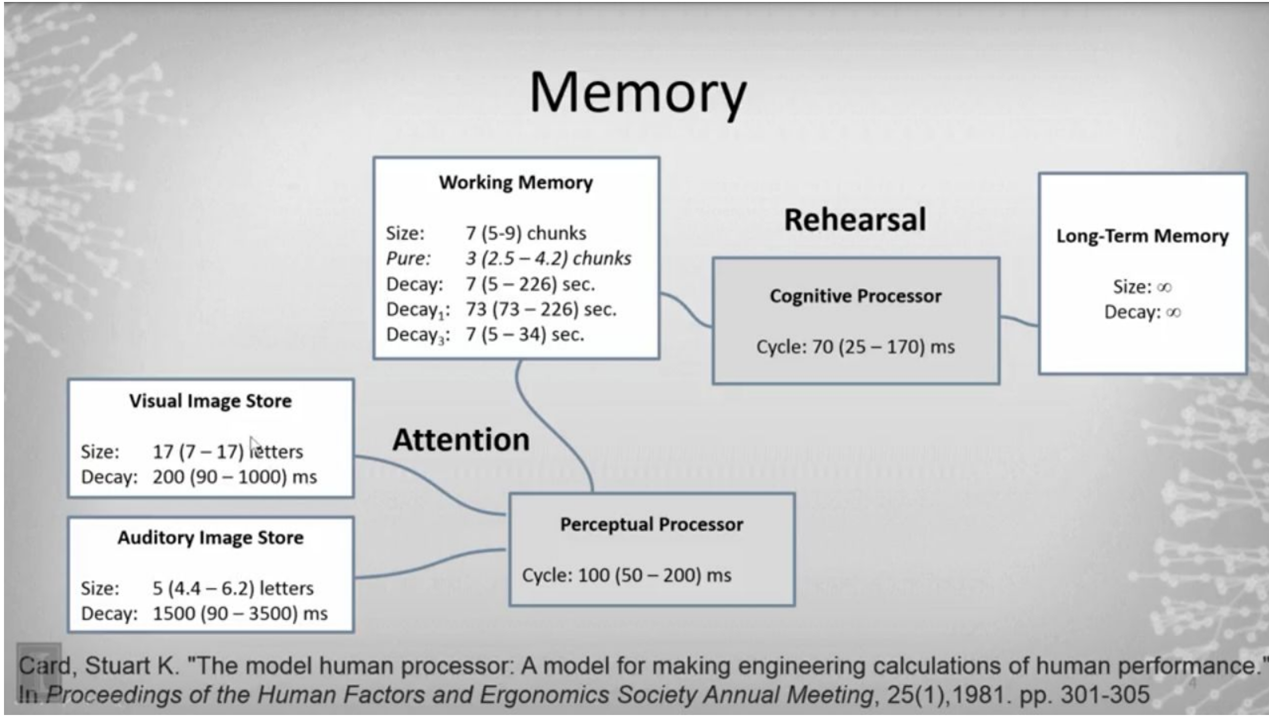
$$T \approx 600 \text{ ms} + 240 \text{ ms} \lg(1+D/S)$$

- D = distance to target
- S = size of target
- 240 ms = 70 ms to move your hand +
  100 ms to see the result +
  70 ms to decide how to correct it



Memory

- Visual Image Store
  - 7-17 items
  - 90-1000ms
  - Buffers Graphical Tokens
- Attention
  - Taking from Visual and Auditory and placing in Working Memory
  - 5-9 chunks
- Rehearsal
  - Remembering - taking from working memory to long-term memory



Sensory Memory

Human Device Memory

- *Iconic* memory - visual
  - Persistence of vision
  - .5 seconds
- *Echoic* memory - aural
- *Haptic* memory - touch
- *Arousal* - level of interest or need

- Iconic - brief - remember what was on the screen
- Arousal
  - shortens acquisition times
  - extends decay times

- 5-9 Chunks (digits / letters, words)
- grouped numbers / letters increase capacity

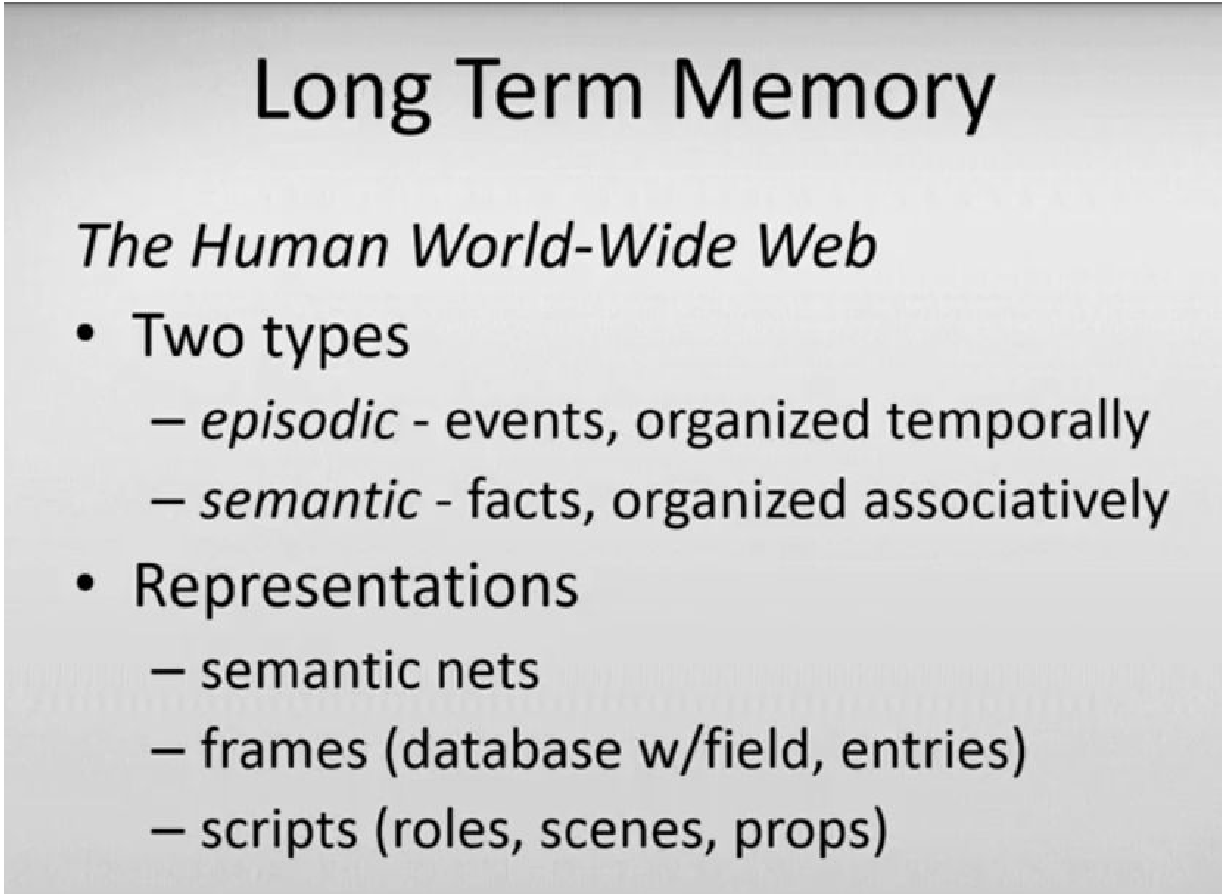
Working Memory

Human DRAM

- 70ms access time
- 200ms *refresh* time
- Size: 7 +/- 2 items (digits, chunks, words)
- *Recency* effect – last is best

Long-Term Memory

- Episodic
  - Temporally Organized
  - Events
- Semantic
  - Associatively Organized
  - Facts
- Representations
  - Scripts
  - Frames
  - Semantic Nets



Reasoning - Cognitive Processor

Deductive

- If A then B
- A, therefore B
- Not B, therefore not A
- When eliminated the impossible, the improbable must be true

CORRELATION IS NOT CAUSATION !!!!!!

Inductive

- If true for x, then also x+1

- Generalizations
- Extrapolation
- Interpolation

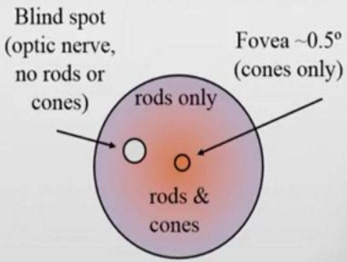
- Allows us / Leads us to infer or interpolate or makeup for missing data

Abductive

- Asking Why
- Abstracting from Concrete
- Looking for Simplest Explanation
- Cognitive Dissonance
  - Entertaining simultaneous contradictory opinions
  - What happens when evidence disagrees with model?
- Can lead to bad assumptions / over simplifications
- Discounting exceptions because they don't match our model

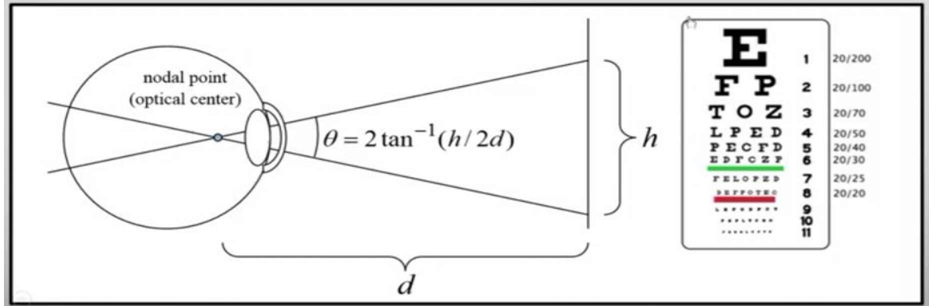
Rods & Cones

- Rods measure intensity
  - 80 million
  - Denser away from fovea
  - Astronomers learn to glance off to the side of what they are studying
  - sensitive, shut down in daylight
- Cones (sensitive to "red", "green" & "blue")
  - 5 million total
  - 100K – 325K cones/mm^2 in fovea
  - 150 hues
- Combined: 7 million shades



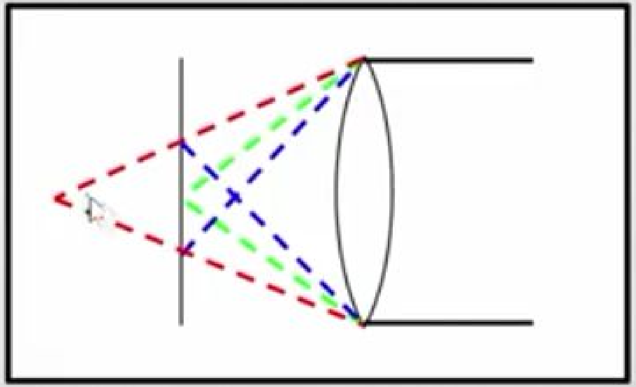
Acuity

- Angular resolution of retina
- Snellen ratio: 20/X means you distinguish at 20 feet what the average person distinguishes at X feet.
- 20/20 = distinguish two points 1 arc minute apart



Chromatic Aberration

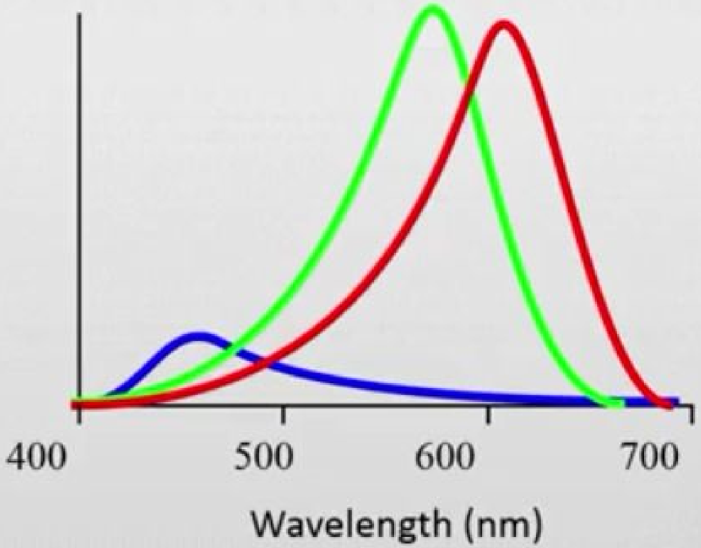
- Refractive index of lens material varies by wavelength
- Resulting dispersion causes focal plane to vary by color



- 1.5 diopters between focus of red and blue
- This is why amber sunglasses aid vision
- Never use pure blue (add at least a bit of red or green to aid in focusing on edges)

Color Perception

- L = 31% R + 59% G + 10% B
- 10% of males are color blind
- Pay attention to contrast!



- Eye color space
  Y = R + G, Y – B, R – G
- Color space is black⇔white, yellow⇔blue, red⇔green

Retinal Processing

Cornea, lens focus light onto *Retina*

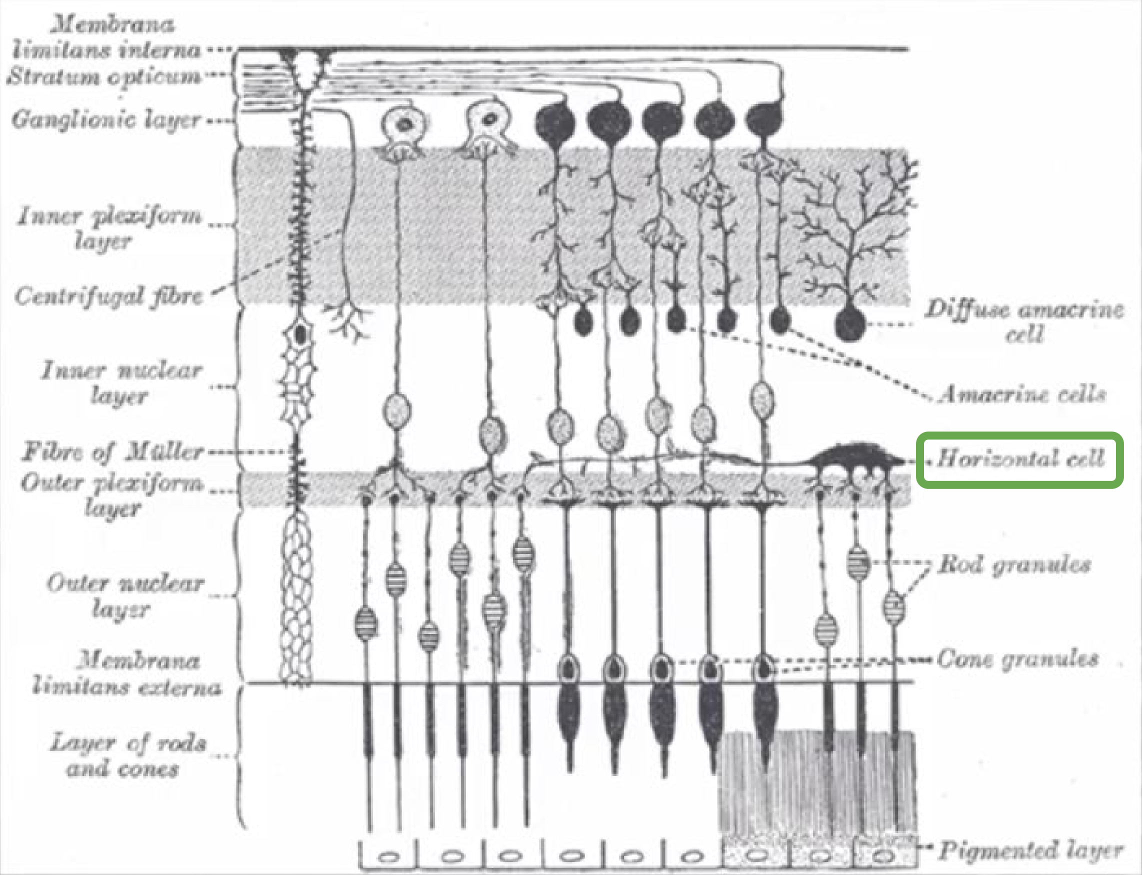
Photoreceptors

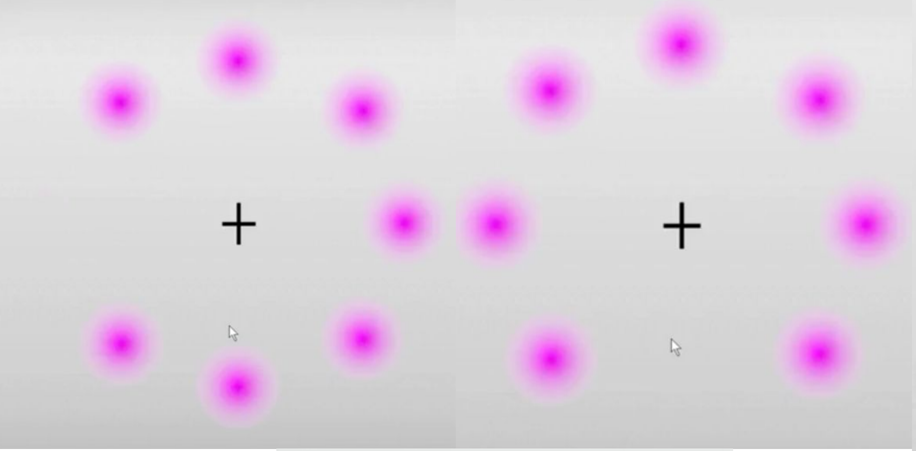
- *rods* - brightness
- *cones* – color (red, green, blue)

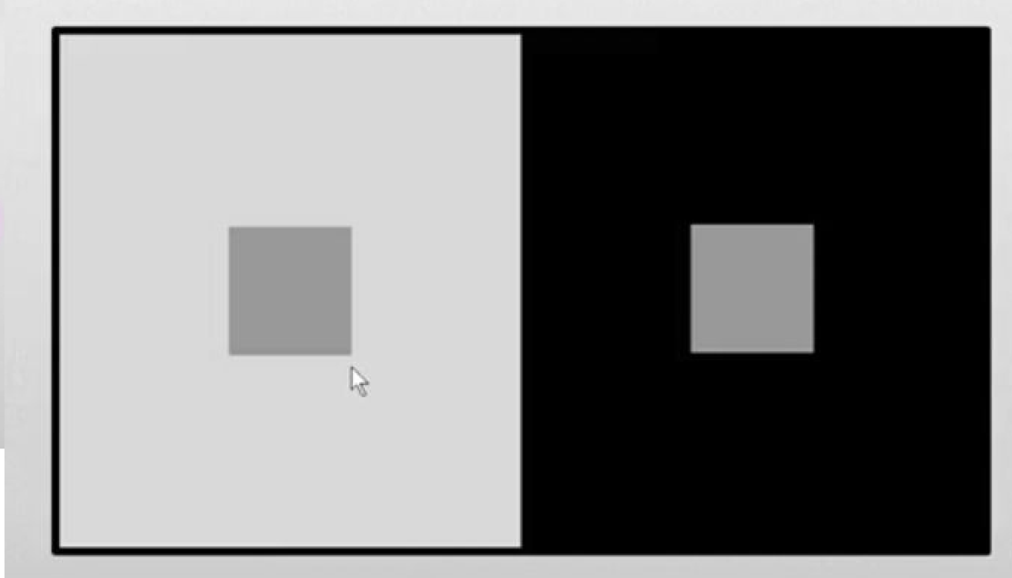
Ganglions – nerve cells

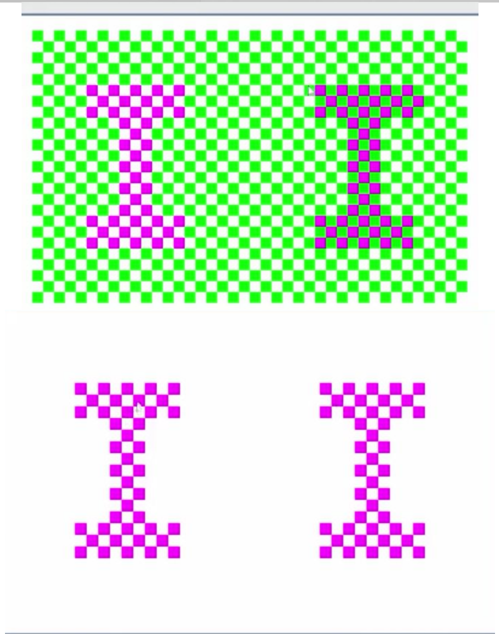
- (*X-cells*) detect pattern
- (*Y-cells*) detect movement

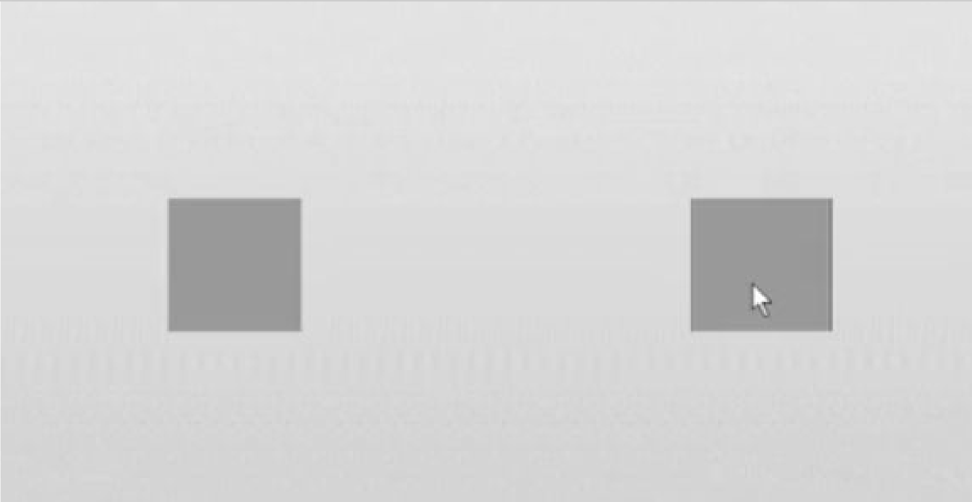
from Gray's Anatomy

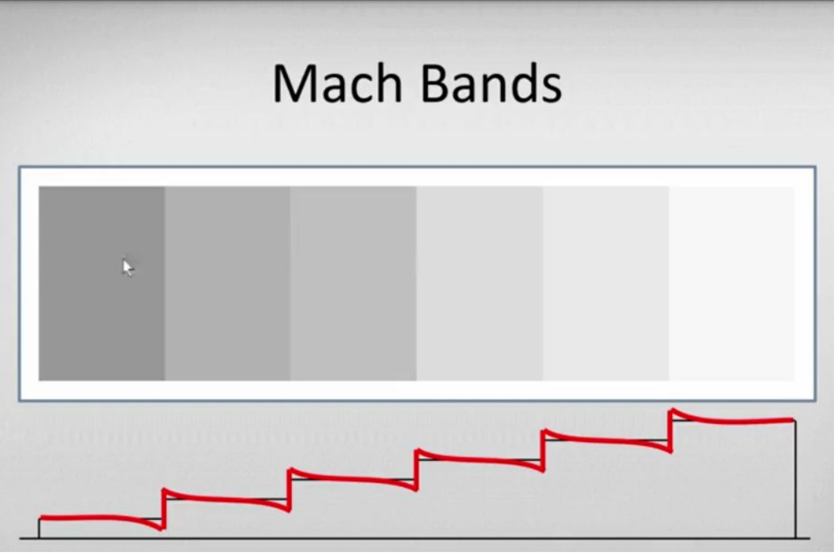


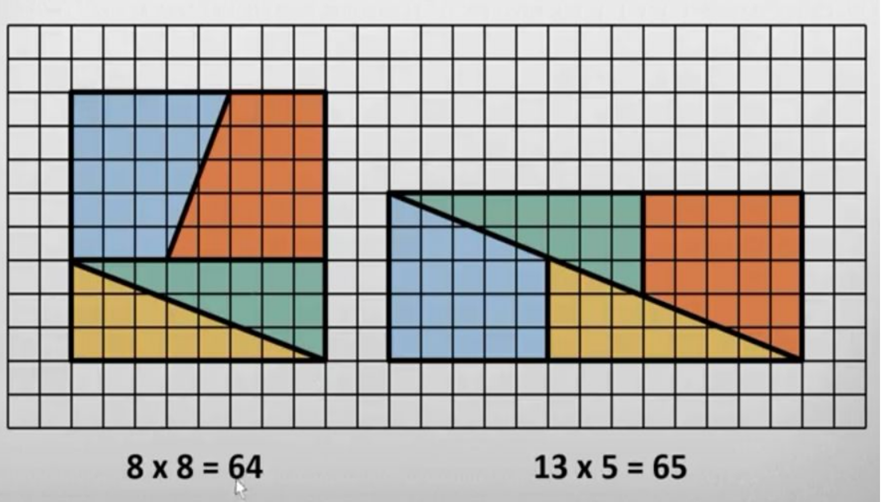


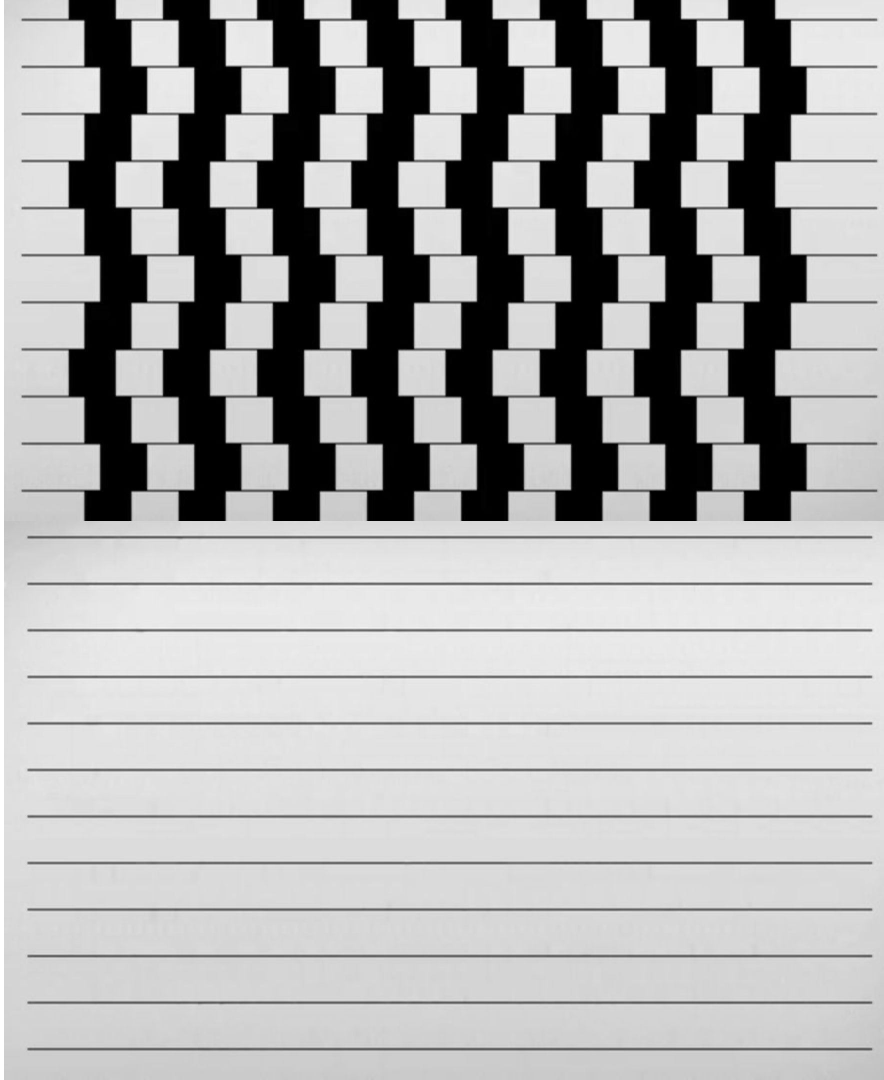


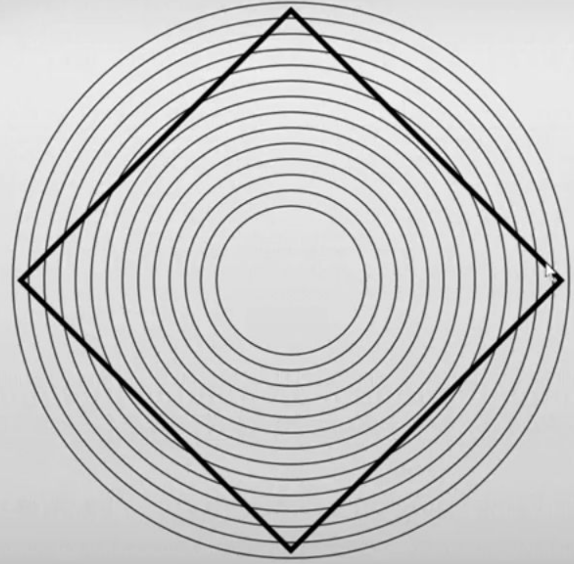


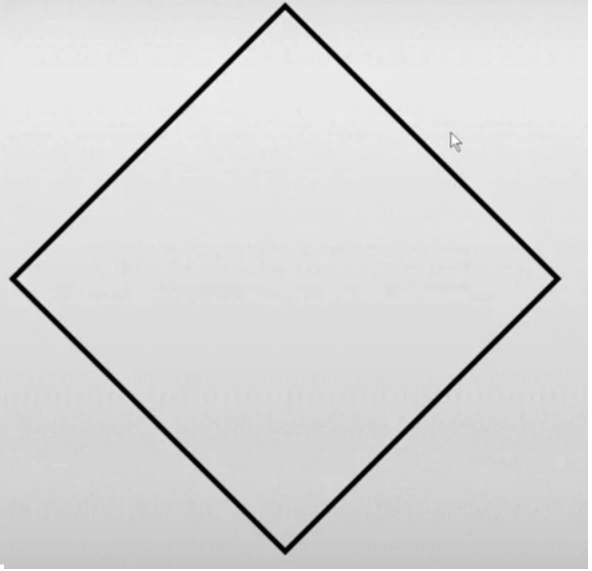


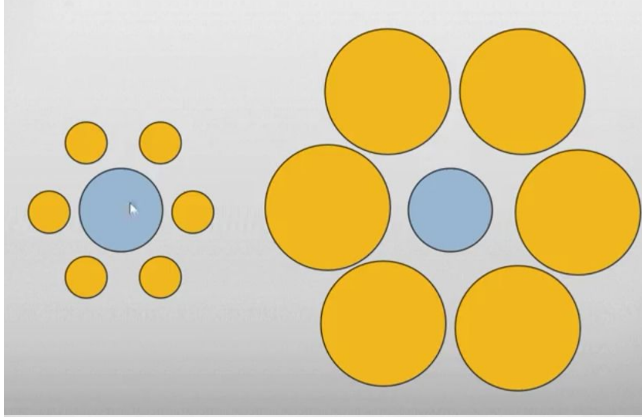

Mach Bands

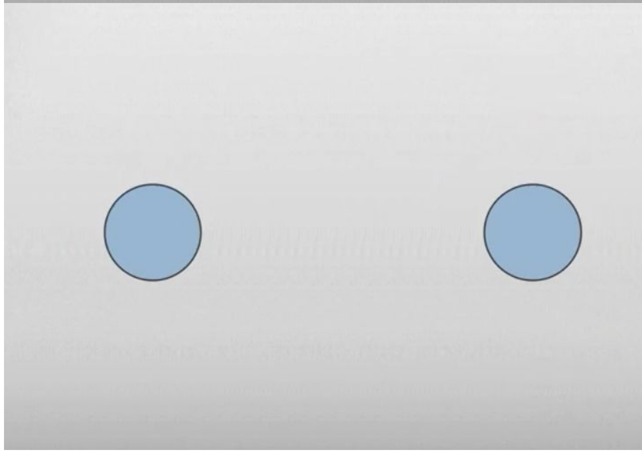

8 x 8 = 64
13 x 5 = 65









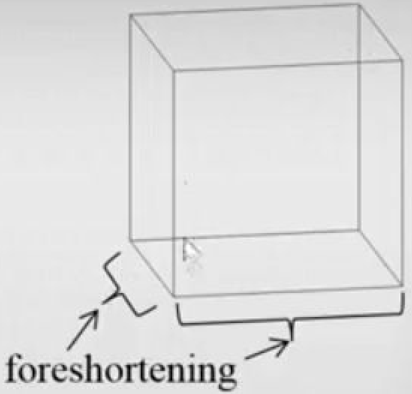


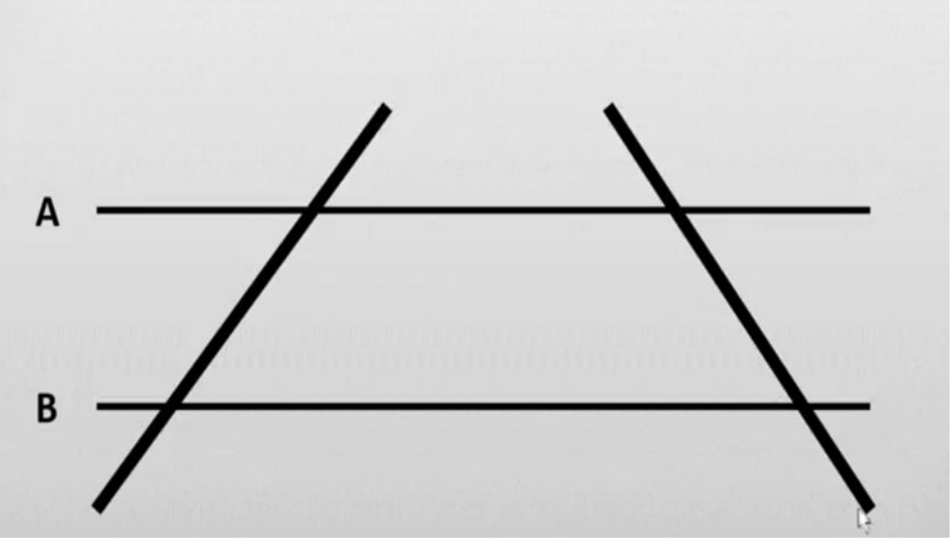
What Did We Learn?

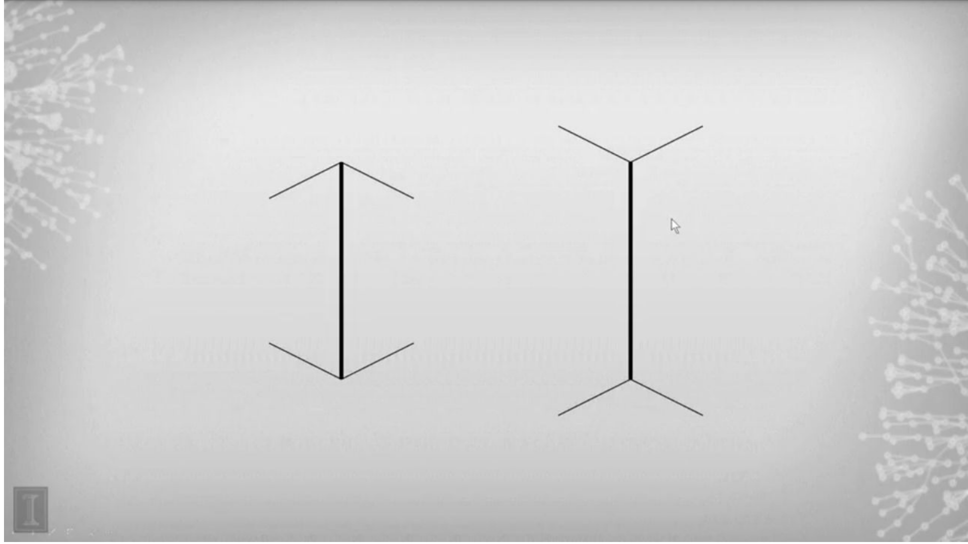
- Various forms of lateral inhibition help our visual system see and accentuate shapes in context of neighboring shapes
- This lateral inhibition can also interfere with the proper perception of visual data
- Always use consistent contexts for visual comparisons

Perspective

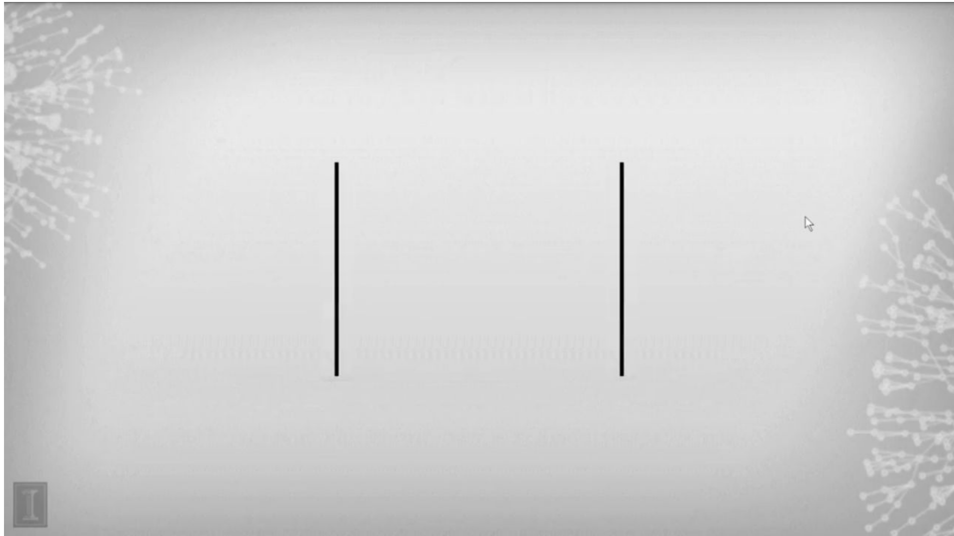
- **Foreshortening**: Objects at different depth along a similar line of sight project to nearby locations on the image plane
- **Linear Perspective**: Objects farther away appear smaller
- **Size Constancy**: Objects do not change size, so smaller objects must be farther away than larger objects



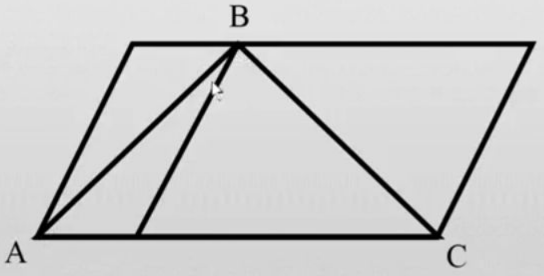

A
B

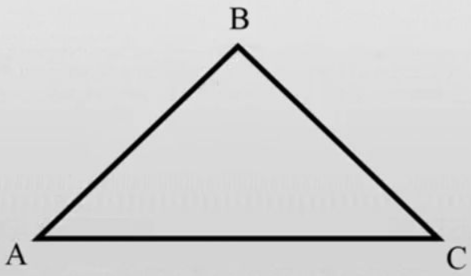



A
B



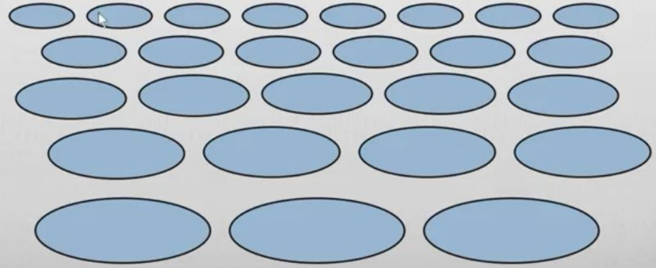
Which is Longer, AB or BC?





Texture Constancy

- Expect objects to be same size in 3-D
- Differences must be due to perspective



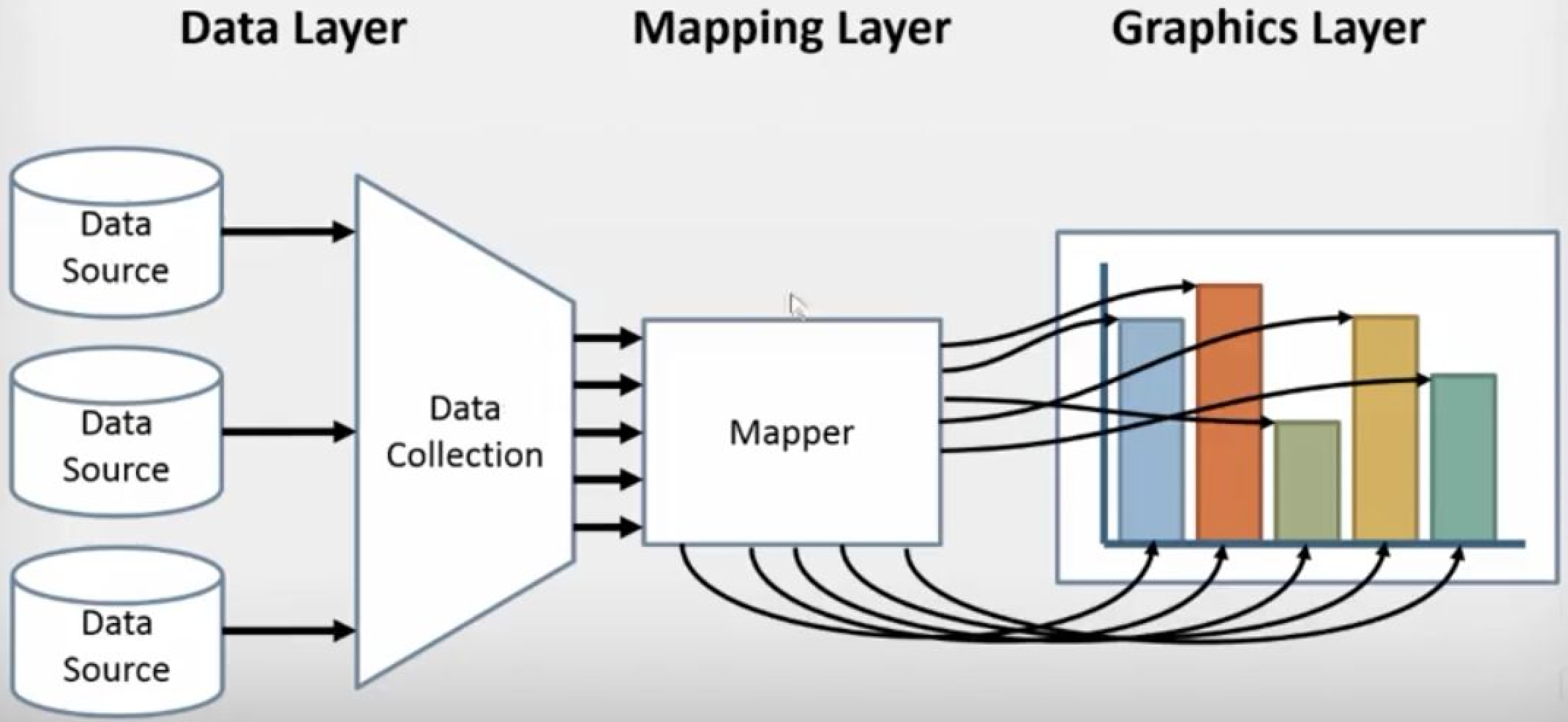
Lighting Assumptions

What Did We Learn

- Our perception of size of an object is influenced by our perception of the distance to the object
- Avoid the incorporation of artificial 3-D elements in the presentation of 2-D data

Actual Data Visualization

Data Visualization Framework



Data Layer	Mapping Layer	Graphics Layer
• Locating and obtaining data • Importing data in proper format • Relating data for proper correspondence • Data analysis and aggregation	• Associating appropriate geometry with corresponding data channels • Data analysis and algorithms (e.g. contouring)	• Conversion of geometry into displayable image • Decorations • Managing interaction

Data Types

	Discrete (no between values)	**Continuous** (values between)
Ordered (values are comparable)	**Ordinal,** e.g. size: S,M,L,XL,… **Quantitative,** e.g. counts: 1,2,3,…	**Fields,** e.g. altitude, temperature
Unordered (values not comparable)	**Nominal,** e.g. shape: □○△ **Categories,** e.g. nationality	**Cyclic values,** e.g. directions, hues

Data as Variables

Science	Databases	Data Warehouses
Independent Variable	Key	Dimension
Dependent Variable	Value	Measure

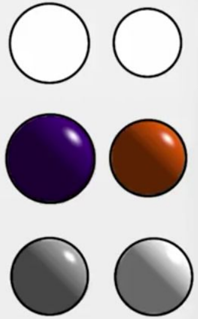
Mapping Quantitative Values

- Position
- Length
- Angle/Slope
- Area
- Volume
- Color/Density

CLEVELAND, W. S., AND MCGILL, R. Graphical perception: Theory, experimentation and application to the development of graphical methods. Journal of the American Statistical Association, 79(387) 1984

Mapping Quantitative Values

- Position
- Length
- Angle/Slope
- Area
- Volume
- Color/Density



CLEVELAND, W. S., AND MCGILL, R. Graphical perception: Theory, experimentation and application to the development of graphical methods. Journal of the American Statistical Association, 79(387) 1984

Quantitative	Ordinal
Position	Position
Length	Density
Angle	Saturation
Slope	Hue
Area	
Volume	
Density	
Saturation	
Hue	

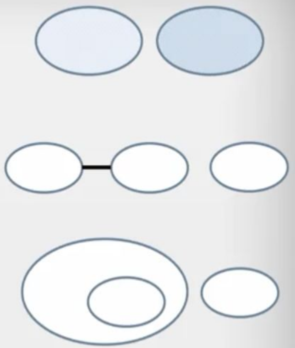
Mapping Quantitative Values

- Position
- Length
- Angle/Slope
- Area
- Volume
- Color/Density

1-D (Position, Length, Angle/Slope)
2-D (Area)
3-D (Volume)

CLEVELAND, W. S., AND MCGILL, R. Graphical perception: Theory, experimentation and application to the development of graphical methods. Journal of the American Statistical Association, 79(387) 1984

Quantitative	Ordinal
Position	Position
Length	Density
Angle	Saturation
Slope	Hue
Area	**Texture**
Volume	**Connection**
Density	**Containment**
Saturation	Length
Hue	Angle
	Slope
	Area
	Volume



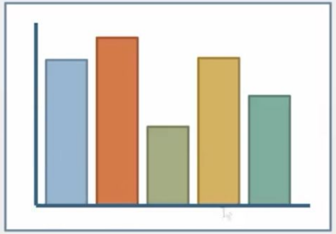
J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986

Bar Chart

Quantitative dependent variable

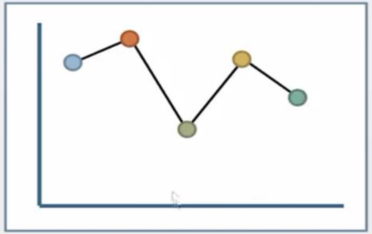


Benefits from both position (top of bar) and length (size of bar)

Discrete/nominal independent variable

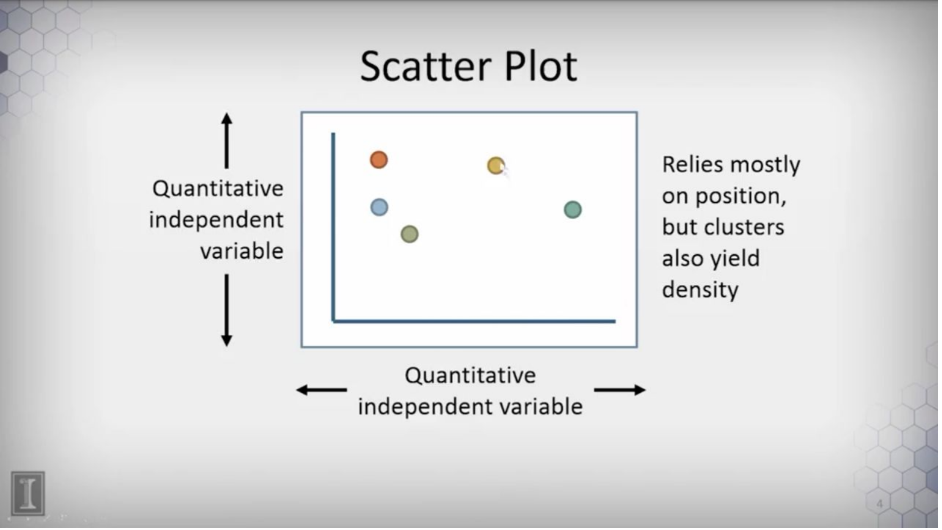
Line Chart

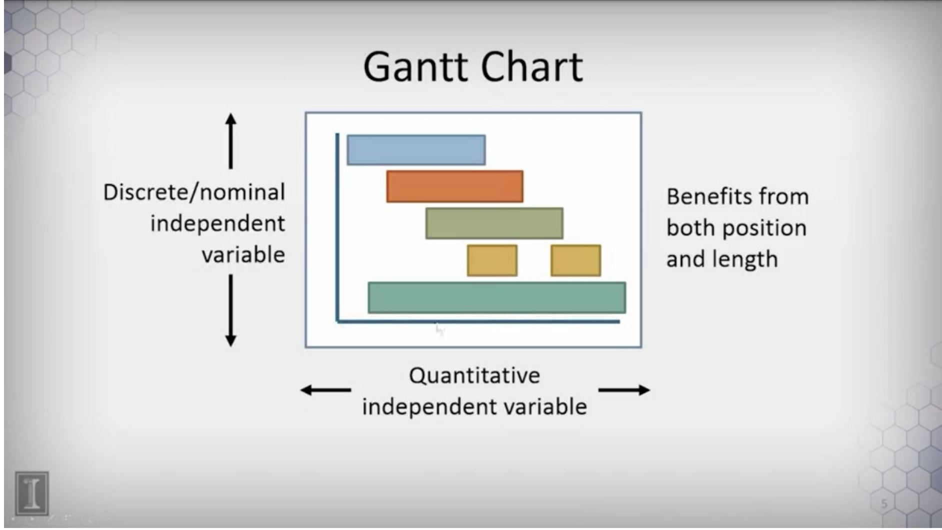
Quantitative continuous dependent variable



Benefits from position but not length

Quantitative continuous independent variable


Scatter Plot
Quantitative independent variable
Relies mostly on position, but clusters also yield density
Quantitative independent variable


Gantt Chart
Discrete/nominal independent variable
Benefits from both position and length
Quantitative independent variable

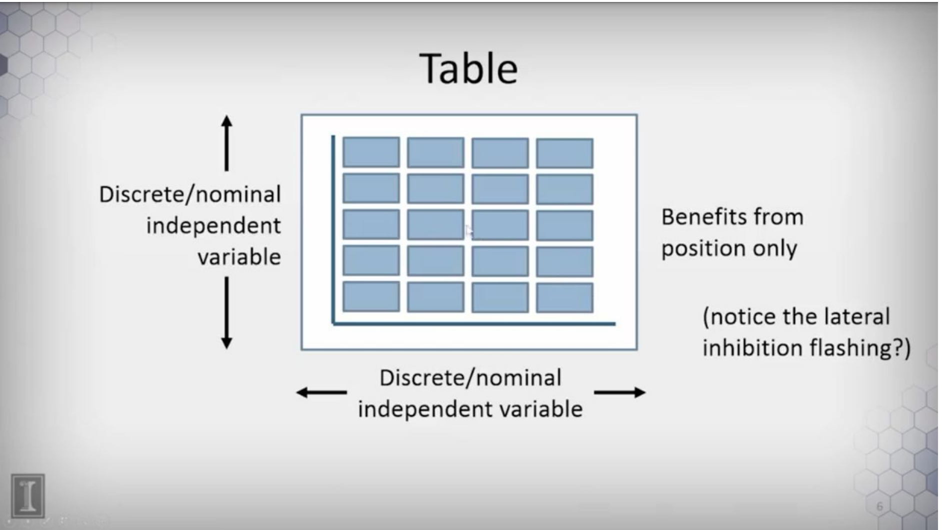


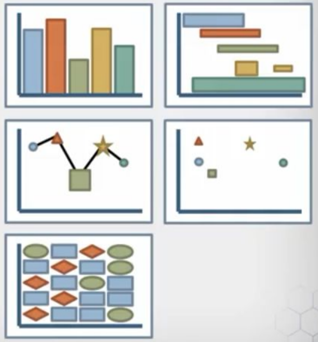
Table
Discrete/nominal independent variable
Benefits from position only
(notice the lateral inhibition flashing?)
Discrete/nominal independent variable

What to Use?

Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Independent	

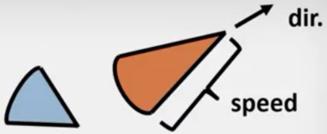
Glyphs in Charts

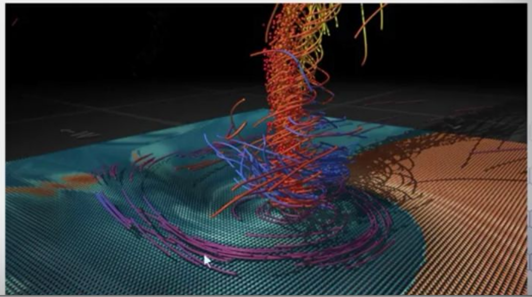
- Shape at top of bar charts, start/end of Gantt chart, and bar shape
- Shape of points in line charts and scatter plots
- Table is a scatter plot of regular variables
- Can vary shape, color, size, orientation

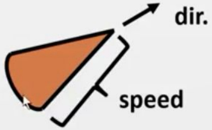


Glyphs

- Cones shape indicates wind speed (size) and direction (orientation)
- Color also used: orange → rising blue → falling
- Adds extra dimensions of data to visualization



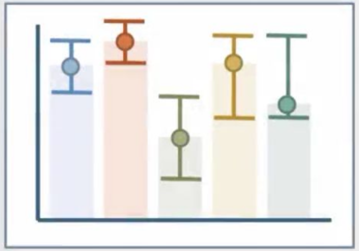


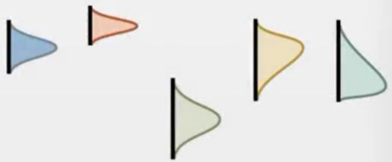


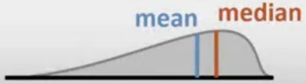
Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986

Error Bars

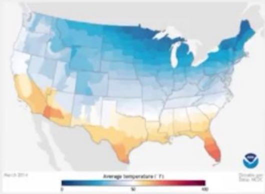






Heatmap

- Table where entries are displayed as a color
- Weather maps are heat maps on a table with columns = latitude and rows = longitude
- Familiar method for visualizing other data too



Quantitative

Position
Length
Angle
Slope
Area
Volume
Density
Saturation
Hue

parameter 2

311	304	298	293	289	286	284	281	283	281
276	270	265	262	258	255	253	254	253	251
260	252	245	240	237	235	233	234	241	260
257	249	246	244	255	303	383	456	531	551
264	273	330	418	553	609	621	623	626	625
363	470	633	661	675	678	680	681	683	680

parameter 1

Microsoft Excel conditional formatting

program optimization

Worlds within Worlds

- Each glyph is itself a plot
- E.g. a table of tables
- Different scales for major axis and minor axis for both horizontal and vertical axes
- Can work in 3-D or even deeper nesting (worlds within worlds within worlds), but less effectively

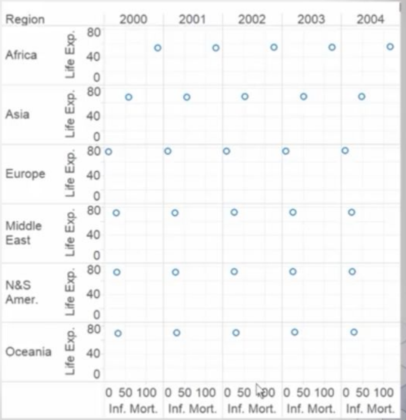
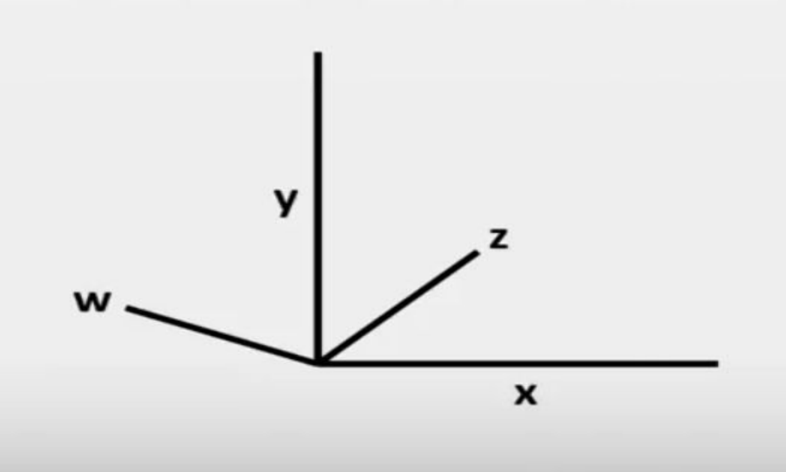


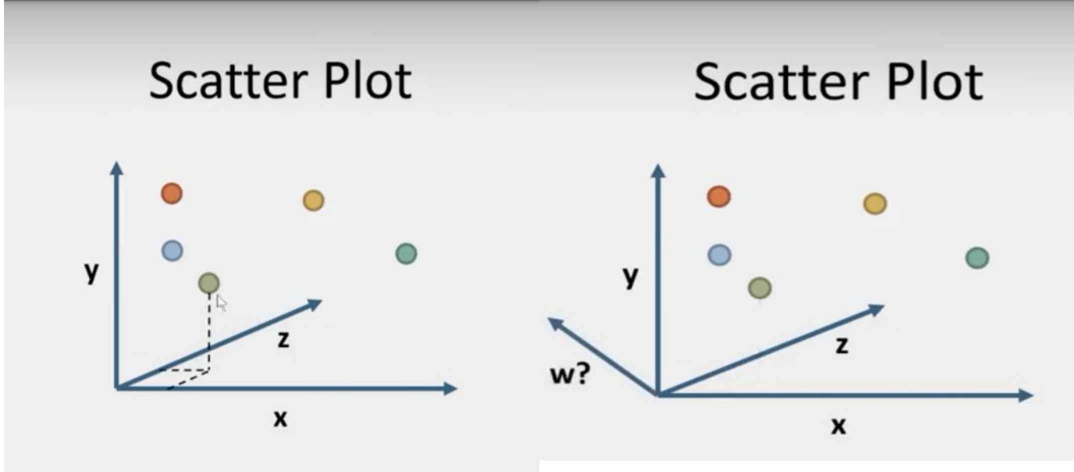
Tableau layout of World Bank Indicator Data

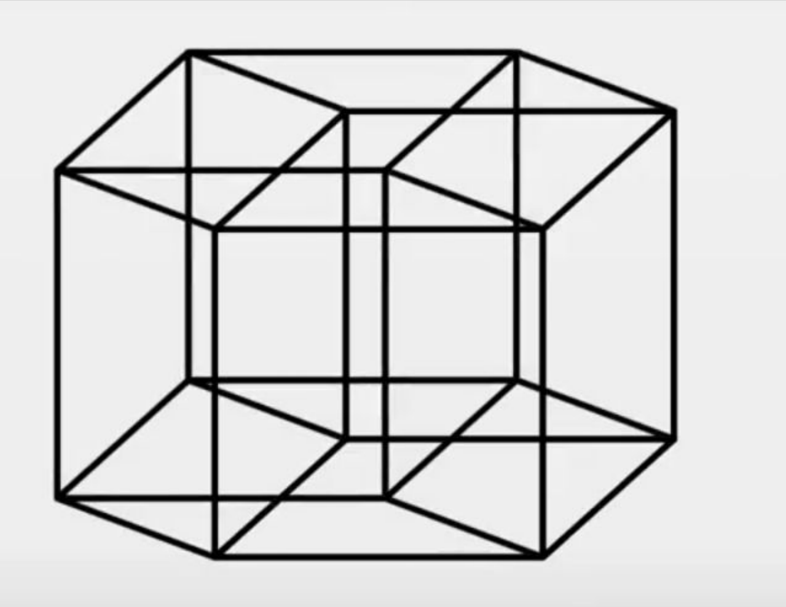
Chernoff Faces

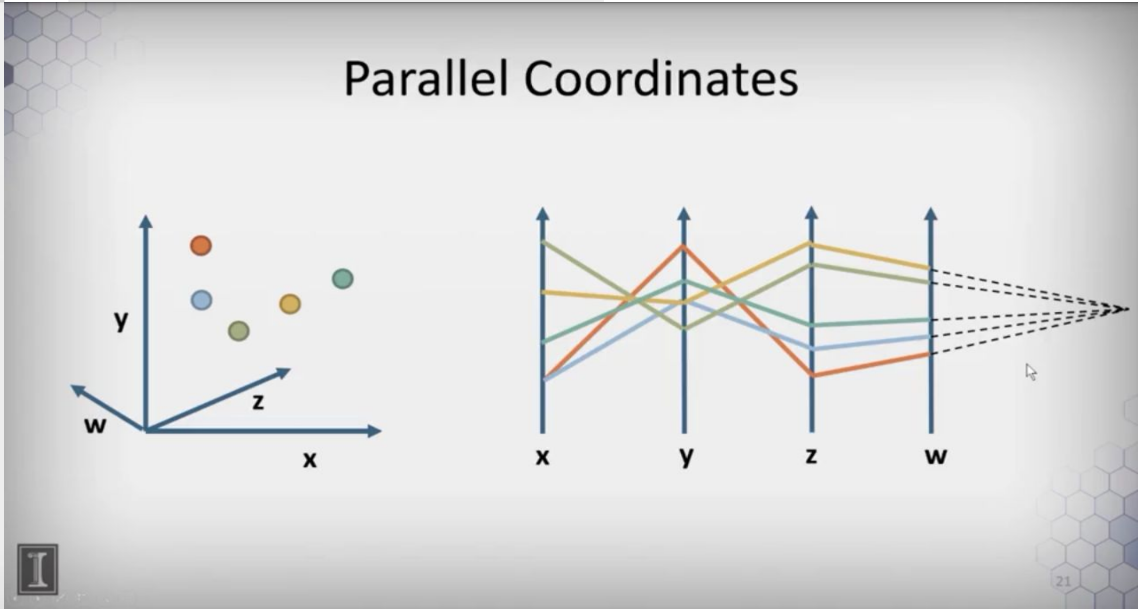
- Glyphs in the form of a cartoon human face
- Maps data to facial features (eyebrows, face shape, expression, etc.)
- Perception and memory designed to detect and recall facial features

12 sample state judges as rated by lawyers and plotted in R by Wikipedia user "Avenue"


y
z
w
x

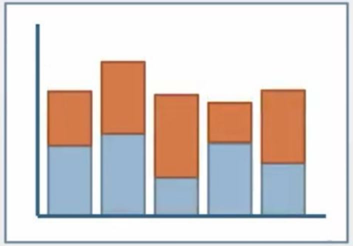

Scatter Plot
y
z
x
Scatter Plot
y
z
w?
x




Parallel Coordinates
y
z
w
x
x
y
z
w
21

Stacked Bar Chart

Two (accumulating) q. dep. variables

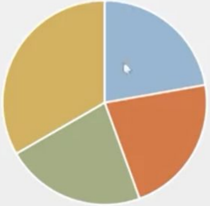


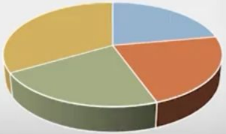
Central limit theorem → as more bars are added, sums will vary less

← Independent variable →

Pie Chart

- Used to indicate relative portions of a quantitative dependent variable of a single dimension
- Maps percentage of total to angle of wedge arc
- Perspective (both distortion and foreshortening) confounds perception of angle



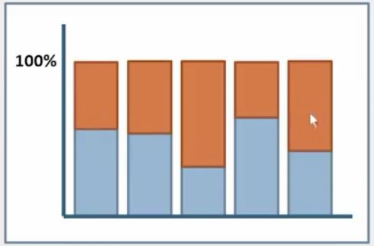


Position
Length
Angle
Area
Volume
Color

Cleveland & McGill, 1984

Relative Stacked Bar Chart

Two quantitative dependent variables

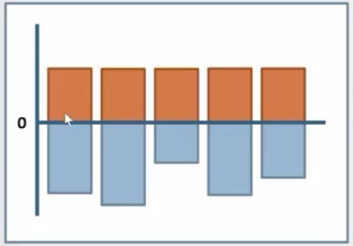


← Independent variable →

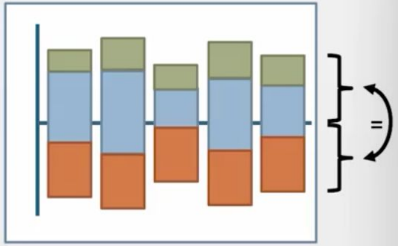
Position
Length
Angle
Area
Volume
Color

Cleveland & McGill, 1984

Diverging Stacked Bar Charts

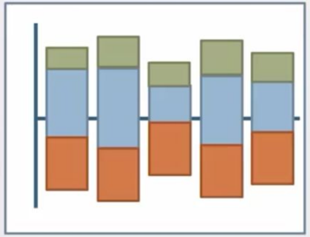


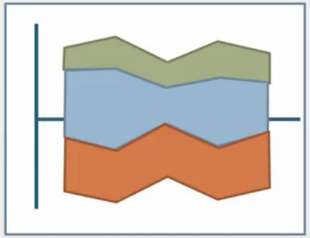
- Benefits from pos. & length
- Only works for two variables
- Negative connotation for lower bars



- Only indicates length
- Works for many variables
- Bar trends can still be obscured by neighboring bar variance

Stacked Bar Charts v. Stacked Line Graphs

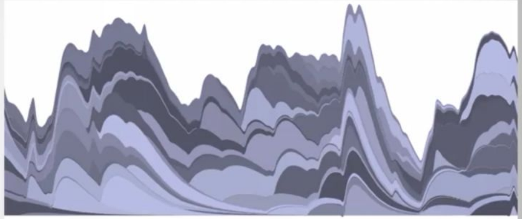




- Appropriate for continuous data over a continuous independent variable
- Can smooth regions using curves instead of line segments

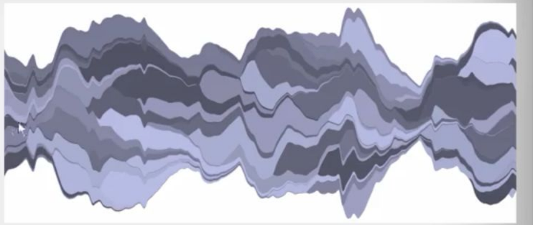
Stacked Graph Layout

- Let g_i be the position of the top of the i'th stacked bar
  $g_i = g_0 + f_1 + f_2 + \ldots + f_i$
- Setting $g_0 = 0$ results in an ordinary bar chart that distorts data when stacked on varying data underneath



ThemeRiver Layout

- Let g_i be the position of the top of the i'th stacked bar
  $g_i = g_0 + f_1 + f_2 + \ldots + f_i$
- ThemeRiver centers the bar chart on the horizontal axis by setting
  $g_0 = -½ (f_1 + f_2 + \ldots + f_n)$



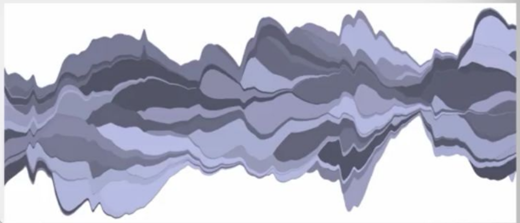
- Minimizes the girth of the chart $(g_0^2 + g_n^2)$ and the top and bottom slopes $(g_0'^2 + g_n'^2)$
- Havre, S., Hetzler, B., Nowell, L. ThemeRiver: Visualizing Theme Changes over Time. *Proceedings of the IEEE Symposium on Information Visualization, 2000*

Images © 2008 Byron & Wattenberg, usage granted, understanding that usage does not imply that authors endorse or promote any products from Coursera, this course, its instructor or the University of Illinois.

Streamgraph Layout

- Let g_i be the position of the top of the i'th stacked bar
  $g_i = g_0 + f_1 + f_2 + \ldots + f_i$
- Streamgraph sets the base at

$$g_0 = -\frac{1}{n+1}\sum_{i=1}^{n}(n-i+1)f_i$$



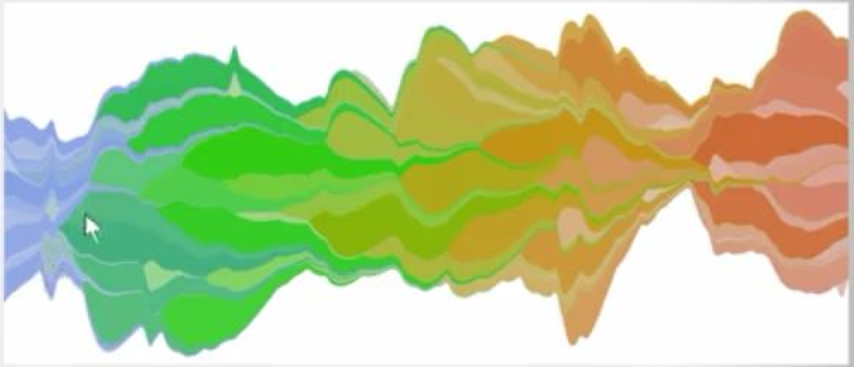
(actually uses a weighted version, but harder to evaluate)

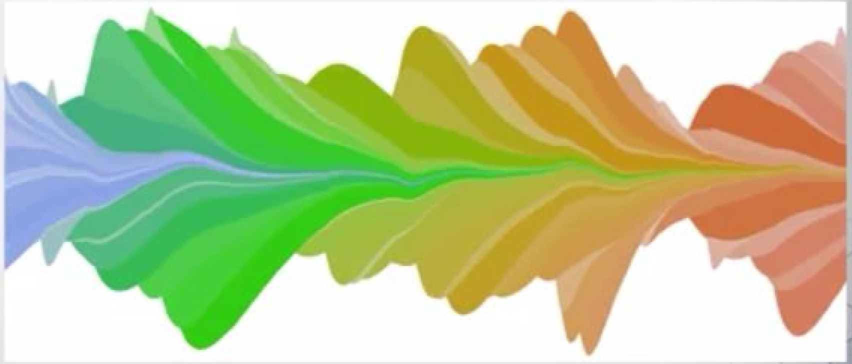
- Minimizes the "deviation" $(\Sigma\, g_i^2)$ and the "wiggle" $(\Sigma\, g_i'^2)$
- Byron, Lee, and Martin Wattenberg. "Stacked Graphs – Geometry & Aesthetics." IEEE Trans. On *Visualization and Computer Graphics* 14(6), 2008, pp. 1245-1252.

Images © 2008 Byron & Wattenberg, usage granted, understanding that usage does not imply that authors endorse or promote any products from Coursera, this course, its instructor or the University of Illinois.

Streamgraph Ordering

- Compute total weight w_i of each series i (sum of values of each datapoint)
- If $(w_1 + \ldots + w_{n/2}) > (w_{n/2+1} + \ldots + w_n)$, then add next series to bottom, otherwise add next series to the top
- By adding new series at bottom (f_1) or top (f_n), new data is introduced near high-contrast silhouette where it is better noticed, and fades toward middle

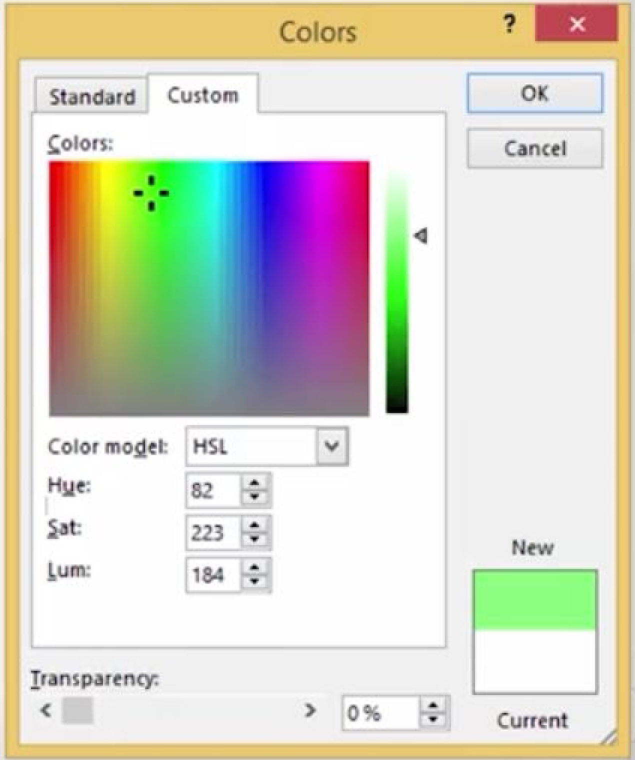




Hue, Saturation and Value

- Hue – angle around the color wheel
  0° = red, 60° = yellow, 120° = green,
  180° = cyan, 240° = blue, 300° = magenta
- Saturation – distance from gray
- Value – distance from black

```
//Convert R,G,B to H,S,V
V = max(R,G,B)
D = V - min(R,G,B)
S = D/V
if (V == R) then H = (G-B)/D
else if (V == G) then H = (B-R)/D
else H = (R-G)/D
H = (60*H) mod 360
```



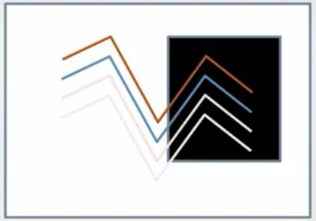
Hues

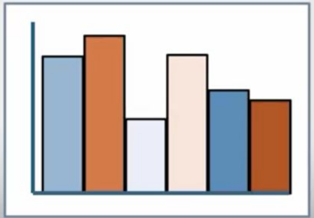
- Observers can rapidly differentiate between only five to ten hues [Healy, "Choosing effective colors for data visualization" Proc. Visualization, 1996]
- Twelve colors (6 + 6) recommended by Ward's "Information Visualization"
- Based on Berlin & Kay, "Basic Color Terms" (plus cyan)

White	Black	Red	Yellow	Green	Blue
Gray	Pink	Orange	Brown	Purple	Cyan

Saturation

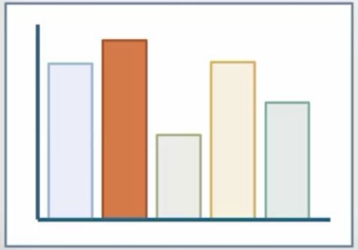
- Use saturated colors for points, strokes and symbols
- Use desaturated colors for fills and larger areas
- Desaturation blends with white, increases luminance
- Perceptual issues with color constancy and lateral inhibition





Contrast

- Use higher luminance contrast to gain attention
- Make sure text has sufficient luminance contrast



Here is some sample text to demonstrate the need for luminance contrast instead of color contrast. The hue of the text is complementary to the hue of the background, but as the background changes its luminance from less than the text to greater than text, the text becomes significantly harder to read.

Usage

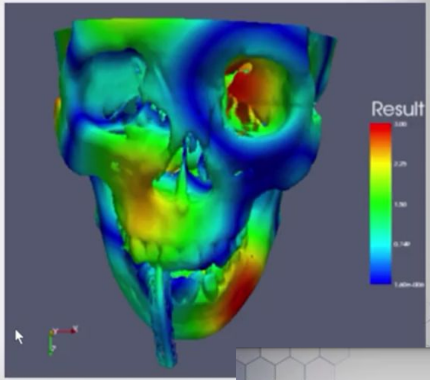
- Density equivalent to value or brightness
- Use different hues for categories
  - Easier to make a hue reference
  - Brightness & saturation more susceptible to color constancy issues
- Can tell brighter, more saturated colors from darker, grayer colors
- Cannot really tell how much brighter or how much more saturated

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986

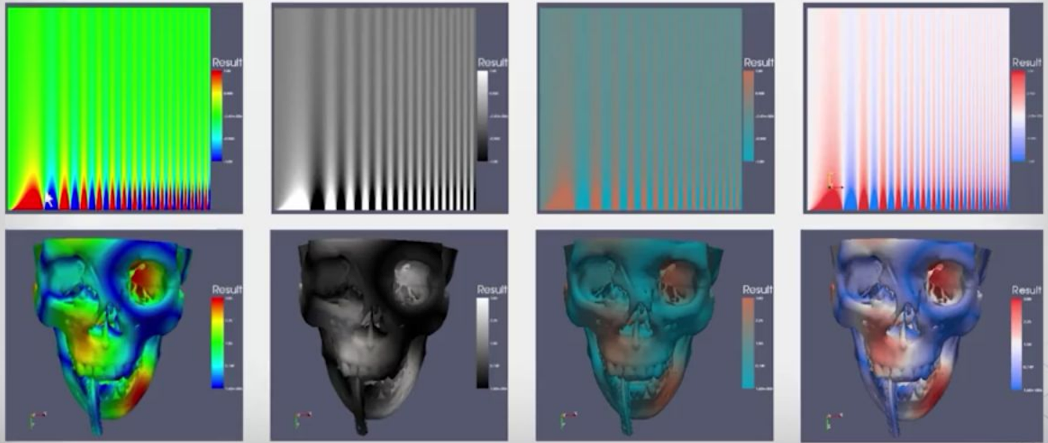
Quantitative Colormaps

- Colormap is a mapping between a quantitative variable and an array of corresponding colors
- Frequent luminance variation in color map helps with perception of detail in the data
- Avoid brightness and saturation mapping on illuminated 3-D surface renderings
- Brightness and saturation maps more error prone than hue maps



© 2007 Moreland & Taylor fr
default color map documenta

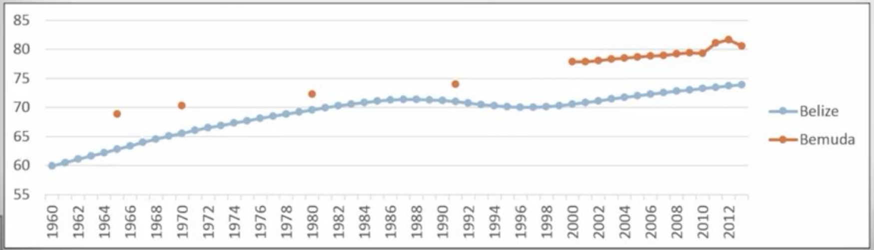
Some ParaView Color Maps



© 2007 Moreland & Taylor from ParaView default color map documentation

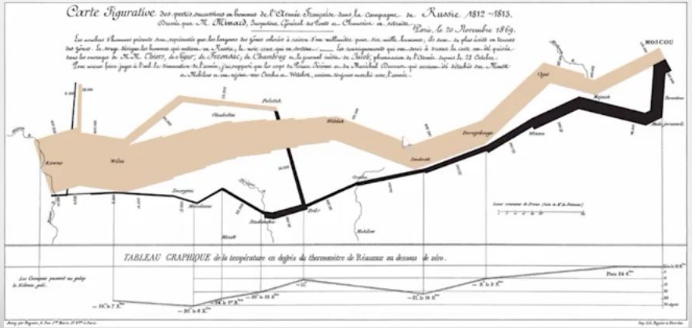
Let the Data Speak

- Avoid summaries and aggregations
- Show where data is missing but don't let it distract the viewer
- Rely on the deductive, inductive and abductive reasoning of the viewer



A Picture is Worth a Thousand Words

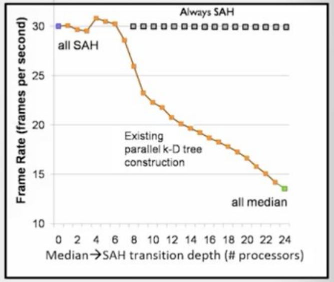
- Consider using pictures/icons/glyphs in place of words
- Tufte: "Only a picture can carry such a volume of data in such a small space"



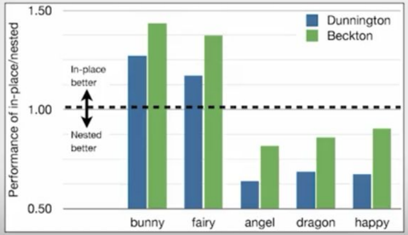
Charles Minard, 1869

Annotation

- Label your axes!
- Pictures still need words
- Label should stand out from data

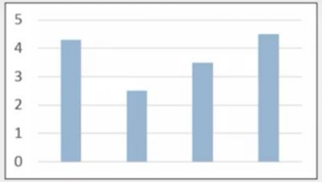


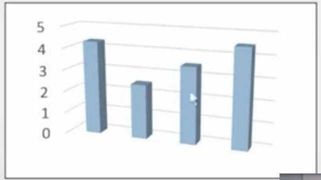
Choi et al. Parallel SAH k-D Tree Construction. Proc. High Performance Graphics, 2010



Chartjunk

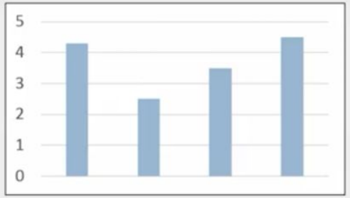
- Making a visualization look prettier often makes it less effective at communicating its data
- Using 3-D can make a 2-D boring chart more engaging
- Using 3-D can often lead to erroneous interpretations





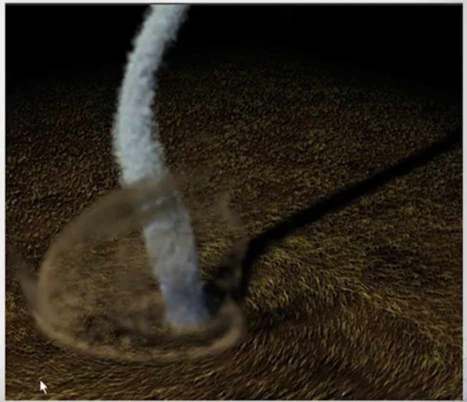
The Data-Ink Ratio

- Maximize the ratio of data to ink in your visualization
- Don’t waste ink on elements of the visualization not associated with data
- Tufte’s minimalism



Micro/Macro

- Fine micro-level details become texture when viewed at the macro level
- Create interactive zoomable interfaces when possible
- Leads to part of Schneiderman's mantra: overview first, then details on demand



NCSA

Information Layers

- Different elements of a visualization should have different appearance
- Use multiple, redundant visual differences

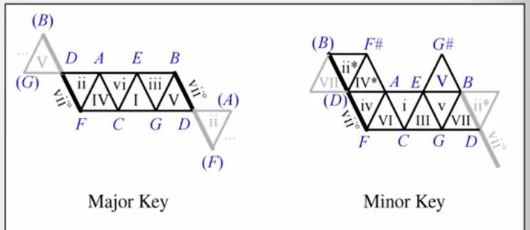
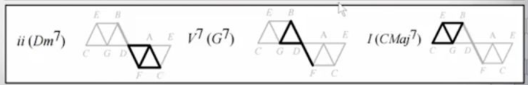


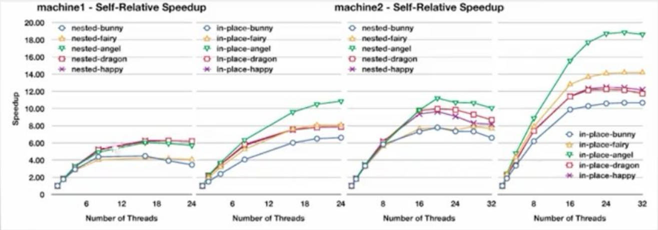
Figure 7: Isochords organization of common major and minor chord structures, demonstrated (blue) in C major and A minor.



Bergstrom, Karahalios, Hart. Isochords: Visualizing Structure in Music. Proc. Graphics Interface, 2007

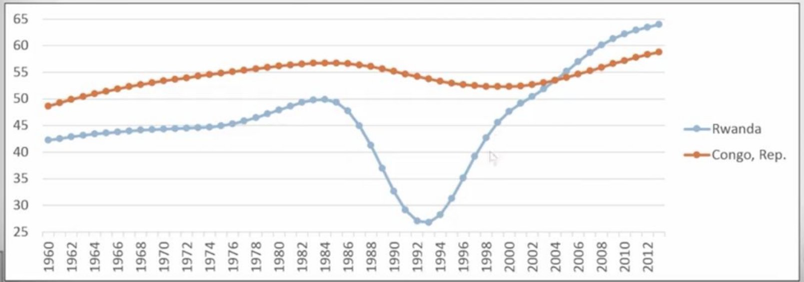
Multiples

- Maintain a consistent design
- Do not change appearance for the sake of change only
- Consistent appearance puts emphasis on data, not the visual design
- Changes in design can distract from irregularities in the data



Narrative

- What story is your data visualization telling?
- What is happening, over time, across space?



Final Conclusions

Data Visualization is about communicating a story told through the content.

The format should NOT be the focus or distract from the main goal or the viewer Reasoning on Data and its Relationships.

The Viewer sees the image through the physical constraints of the physical biological system of the Eye, Visual Image Memory, Perceptive Processing, Working Memory, Cognitive Processing, Long-Term Memory.

Practical Considerations

- Blue needs added colors
- Perception can lead to false observations and conclusions so careful with angles, adjacent lines, gaps, etc.
- Use the appropriate type of representation for the type of data. Maximize correct reasoning by the viewer.
- Present the data in the clearest most direct way possible to guide the viewers reasoning
  - Minimized Formatting, Maximize Content
  - Subtle over Drastic - Eyes are sensitive
  - Annotate and Mix Text, Shapes, Lines, Arrows
  - Browse over Navigate

Next Time (pt.2)

Working with Coordinates

Practical Examples in R

Practical Examples in D3

Practical Examples in Python

External Examples (Good and Bad)

What Else Do You Want To See?

- ____________________
- ____________________
- ____________________
- ____________________
- ____________________

Thank you for coming!

Sites of value:

http://www.edwardtufte.com/tufte/index

Books of value:

- Anything by Edward Tufte
  - Link to 4 best books in a set - what is handed out at his 1 day course
- The Visualization Handbook (Chuck Hansen, Chris Johnson)
  - Scientific Visualization - Curated chapters

Books cont.

- Readings in Information Visualization (Stuart Card, others from Xerox PARC)
  - Curated collection of research papers
- Information Visualization: Perception for Design (Colin Ware, U. New Hampshire)
  - Perceptual Principles applied to Data Visualization
  - Focus on user interaction
- Visualization Analysis & Design (Tamara Munzer, U. British Columbia)
  - Covers both scientific and information visualization
  - Includes task analysis