

Science of Visualization - Handout

Overview of material from Data Visualization, a Coursera Course

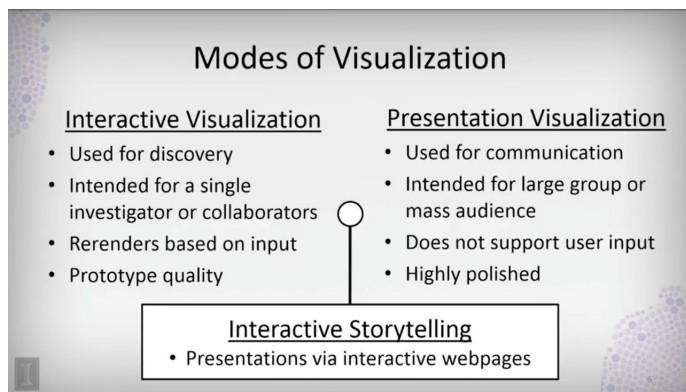
by John C. Hart from University of Illinois

<https://class.coursera.org/datavisualization-001>

What is visualization

Data visualization is a **high bandwidth connection** between **data** on a computer system and a **human brain**, facilitated by visual communication.

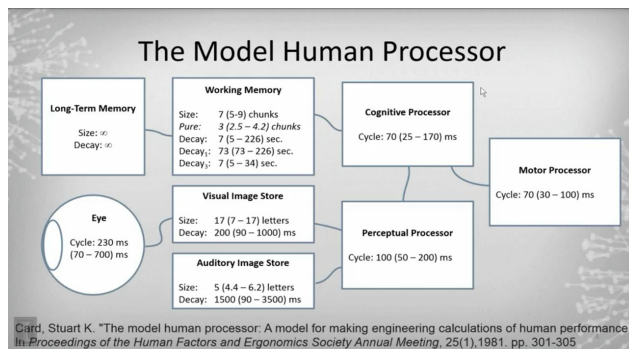
This includes: Data Acquisition, Data Processing, Data Display, Human Perception, Human Memory, and Human Cognition.



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

How the brain processes information / images:



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

According to research at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without a problem. This is because the human mind does not read every letter by itself, but the word as a whole.

Fitt's Law

- Kinesthesia: 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

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

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

The Human World-Wide Web

- Two types
 - episodic** - events, organized temporally
 - semantic** - facts, organized associatively
- Representations
 - semantic nets
 - frames (database w/field, entries)
 - scripts (roles, scenes, props)

How We Remember

How does information get from short term memory into long term memory?

- *Total time hypothesis* – hit the books
- *Distribution of practice effect* - don't cram
- *Meaning* - concrete better than abstract
 - faith age cold tenet quiet logic idea value past
 - boat tree cat child rug plate gun flame head
- Structure, familiarity and concreteness

- 1 bun
- 2 shoe
- 3 tree
- 4 door
- 5 hive
- 6 sticks
- 7 heaven
- 8 gate
- 9 wine
- 10 hen

How We Forget

- **Decay**
 - Logarithmically - forget most early
 - Jost's Law - if two equally strong memories at a given time, then the older is more durable.
- **Interference**
 - proactive inhibition – can't teach an old dog new tricks
 - retroactive interference – mind blown
 - emotion - good old days, forget the mundane

Deductive Reasoning

- **Logic**
 - If A then B
 - A, therefore B
 - Not B, therefore not A
- When you have eliminated the impossible, whatever remains, however improbable, must be the truth
- Correlation is not causation



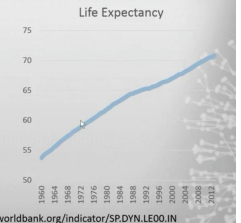
Inductive Reasoning

- If true for x, then true for x+1
- True for x, ...
- Generalizations
- Extrapolation
- Interpolation
- Allows us to infer missing data



Abductive Reasoning

- Human need for meaning
- Asking why?
- Modeling
- Cognitive Dissonance
 - Entertaining simultaneous contradictory opinions
 - What happens when evidence disagrees with model?



The Eye and Visual Perception

Retinal Processing

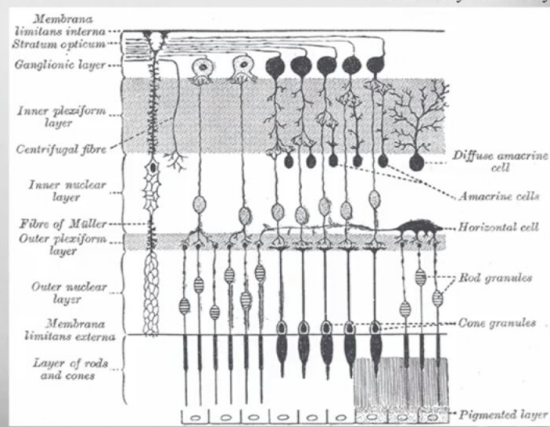
Cornea, lens focus light onto Retina

Photoreceptors

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

Ganglions – nerve cells

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



What challenges exist communicating visually

In neurobiology, lateral inhibition is the capacity of an excited neuron to reduce the activity of its neighbors. Lateral inhibition disables the spreading of action potentials from excited neurons to neighboring neurons in the lateral direction.

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

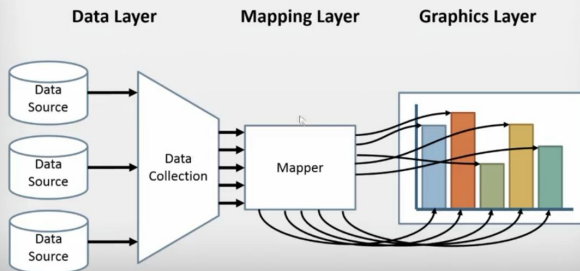


What Did We Learn

- 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

- Locating and obtaining data
- Importing data in proper format
- Relating data for proper correspondence
- Data analysis and aggregation

Mapping Layer

- Associating appropriate geometry with corresponding data channels
- Data analysis and algorithms (e.g. contouring)

Graphics Layer

- Conversion of geometry into displayable image
- Decorations
- Managing interaction

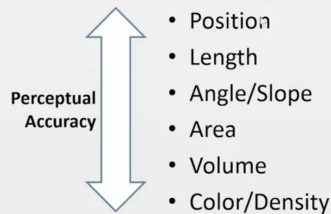
Data Types

	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



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



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

1-D
2-D
3-D

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

- Position
- Length
- Angle
- Slope
- Area
- Volume
- Density
- Saturation
- Hue

Ordinal

- Position
- Density
- Saturation
- Hue

Quantitative

- Position
- Length
- Angle
- Slope
- Area
- Volume
- Density
- Saturation
- Hue

Ordinal

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

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

Quantitative

- Position
- Length
- Angle
- Slope
- Area
- Volume
- Density
- Saturation
- Hue

Ordinal

- Position
- Density
- Saturation
- Hue
- Texture
- Connection
- Containment
- Length
- Angle
- Slope
- Area
- Volume

Nominal

- Position
- Hue
- Texture
- Connection
- Containment
- Density
- Saturation
- Shape
- Length
- Angle
- Slope
- 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

Discrete/nominal independent variable

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

Line Chart

Quantitative continuous dependent variable

Quantitative continuous independent variable

Benefits from position but not length

Scatter Plot

Quantitative independent variable

Quantitative independent variable

Relies mostly on position, but clusters also yield density

Gantt Chart

Discrete/nominal independent variable

Quantitative independent variable

Benefits from both position and length

Table

Discrete/nominal independent variable

Discrete/nominal independent variable

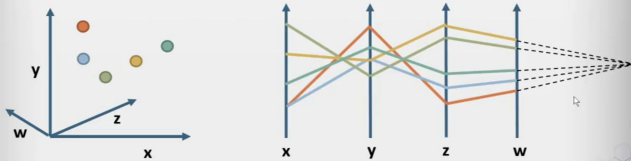
Benefits from position only

(notice the lateral inhibition flashing?)

What to Use?

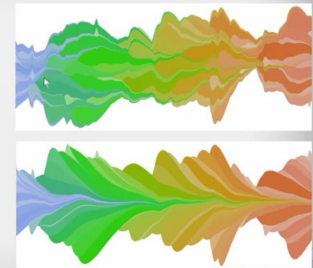
Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
Independent			

Parallel Coordinates



Streamgraph Ordering

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

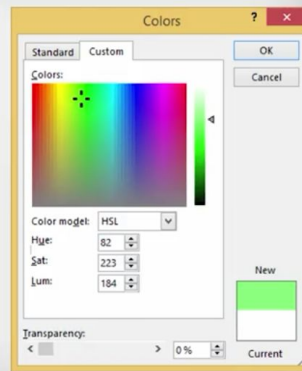


Images © 2008 Byron & Wattenberg, usage granted, understanding that usage does not imply that authors endorse or promote any products from Coursera, this course, its instructor or the University of Illinois.

Hue, Saturation and Value

- Hue – angle around the color wheel
 $0^\circ = \text{red}$, $60^\circ = \text{yellow}$, $120^\circ = \text{green}$,
 $180^\circ = \text{cyan}$, $240^\circ = \text{blue}$, $300^\circ = \text{magenta}$
- 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
```



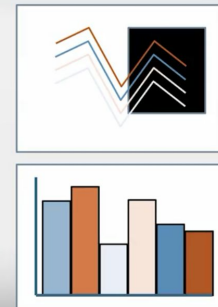
Hues

- 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)



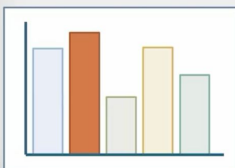
Saturation

- 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

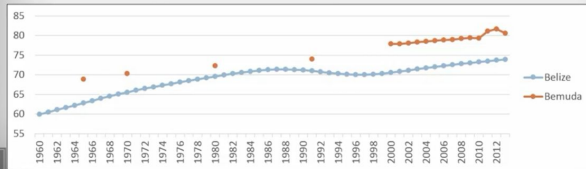
Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

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

Edward Tufte Pointers

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



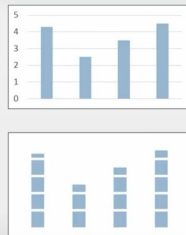
A Picture is Worth a Thousand Words

- 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"



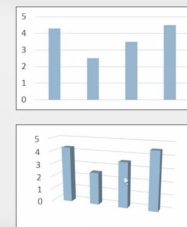
The Data-Ink Ratio

- 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



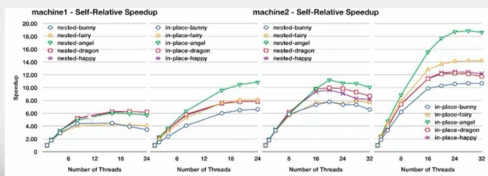
Chartjunk

- 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



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



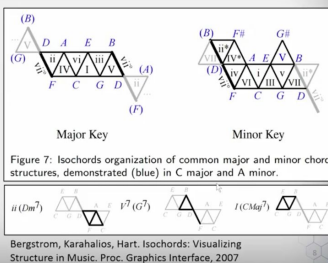
Micro/Macro

- 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



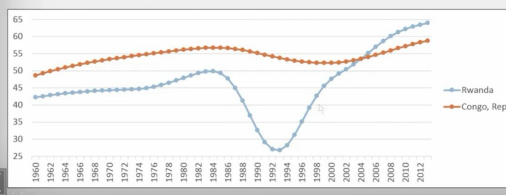
Information Layers

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



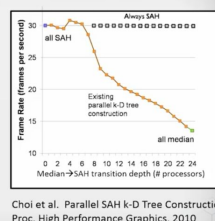
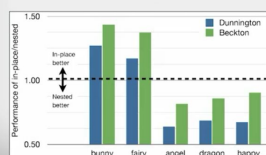
Narrative

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



Annotation

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



- **Focus on Content not Format**
- **Present wider not taller**
- **Repeat differences together with consistent format**
- **Allow Viewer to Browse from Macro to Micro, not navigate down into**
- **Annotate Links and Causal Arrows**
- **Integrate Text, Images, Icons, etc. for a more complete view of the data**