Visualization
Introduction & Overview

Kristi Potter
kpotter@cas.uoregon.edu
Visualization Specialist, CASSPR, CASIT
University of Oregon

CIS 407/507
November 2013
“Industrial Revolution of Data”
1 Byte = 8 Bits
1 Kilobyte ≈ 10^3 bytes
1 Megabyte ≈ 10^6 bytes
1 Gigabyte ≈ 10^9 bytes
1 Terabyte ≈ 10^{12} bytes
1 Petabyte ≈ 10^{15} bytes
1 Exabyte ≈ 10^{18} bytes
1 Zettabyte ≈ 10^{21} bytes
1 Yottabyte ≈ 10^{24} bytes

2.7 ZB: all the worlds data 2012
5 EB: all words ever spoken
200 PB: all printed material
10 TB: library of congress
1 GB: 7 minutes of HDTV
5 MB: all of shakespeare
2 KB: typewritten page
1 B: computer character
1 Byte = 8 Bits
1 Kilobyte ≈ 10^3 bytes
1 Megabyte ≈ 10^6 bytes
1 Gigabyte ≈ 10^9 bytes
1 Terabyte ≈ 10^{12} bytes
1 Petabyte ≈ 10^{15} bytes
1 Exabyte ≈ 10^{18} bytes
1 Zettabyte ≈ 10^{21} bytes

1,208,925,819,614,629,174,706,176 bytes

2.7 ZB: all the worlds data 2012
5 EB: all words ever spoken
200 PB: all printed material
10 TB: library of congress
1 GB: 7 minutes of HDTV
5 MB: all of shakespeare
2 KB: typewritten page
1 B: computer character
1 byte = 8 bits
1 kilobyte ≈ 10^3 bytes
1 megabyte ≈ 10^6 bytes
1 gigabyte ≈ 10^9 bytes
1 terabyte ≈ 10^12 bytes
1 petabyte ≈ 10^15 bytes
1 exabyte ≈ 10^18 bytes
1 zettabyte ≈ 10^21 bytes

~2 exabytes per day

2.7 ZB: all the world's data 2012
5 EB: all words ever spoken
200 PB: all printed material
10 TB: library of congress
1 GB: 7 minutes of HDTV
5 MB: all of Shakespeare
2 KB: typewritten page
1 B: computer character
1 Byte = 8 Bits
1 Kilobyte ≈ 10^3 bytes
1 Megabyte ≈ 10^6 bytes
1 Gigabyte ≈ 10^9 bytes
1 Terabyte ≈ 10^12 bytes
1 Petabyte ≈ 10^15 bytes
1 Exabyte ≈ 10^18 bytes
1 Zettabyte ≈ 10^21 bytes

~2 exabytes per day

2.7 ZB: all the worlds data 2012
5 EB: all words ever spoken
200 PB: all printed material
10 TB: library of congress
1 GB: 7 minutes of HDTV
5 MB: all of Shakespeare
2 KB: typewritten page
1 B: computer character

data ≠ information
The ability to take data - to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it - that’s going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

visualize
WHAT IS VISUALIZATION?

visualization

noun, plural -s

1) formation of mental visual images

2) the act or process of interpreting in visual terms or of putting into visible form
WHAT IS VISUALIZATION?

“Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.”

- Tamara Munzner
Visualization

is needed because...
- cognition is limited
- memory is limited
Visualization

works because it...
- translates data into a “higher bandwidth” form
- uses perception to free up cognition
- serves as an external aide to working memory
- reinforces cognition
**Why visualize?**

*accelerated understanding*
- generate hypotheses
- answer questions
- make decisions
- see data context
- support analysis
- find patterns
- tell a story
- inspire
Visualization goals

enable users to gain insight into data
- record information
- analyze data to support reasoning
- confirm hypotheses
- communicate ideas to others
visualization process

target → translate → design → implement → evaluate → validate
**Visualization process**

**Target**
- learn about the data
- determine the goals of the user
- identify the intended audience
- acquire and clean the data
translate
- convert user goals into data analysis tasks
- structure and characterize the data
- create abstractions of the problem
Abstraction

“"I want to cure cancer"”

close real-world questions into data analysis tasks
- how much? how many? ⇒ retrieve value
- which subset? ⇒ filter
- what is the average? ⇒ compute derived data
- what is the most? minimum? ⇒ find extremum

Thursday, November 7, 13
ABSTRACTION

→ correlation between x & y

convert real-world questions into data analysis tasks
- how much? how many? ➔ retrieve value
- which subset? ➔ filter
- what is the average? ➔ compute derived data
- what is the most? minimum? ➔ find extremum
**Visualization Process**

**design**
- decide on visual encodings, interaction mechanisms
- transform data with appropriate computational methods
- first solution may not be the best solution
space of possible solutions

- good solution
- best solution
Visualization process

implement
- bring ideas to life
- choose appropriate tools
- optimize algorithms
- polish look & feel
**Visualization process**

*evaluate*

- did we pick the right abstractions?
- are our visual encodings appropriate and good?
- does the software work as expected?
**Visualization Process**

iterate....

- translate
- design
- implement
- evaluate

Thursday, November 7, 13
**Visualization process**

**validate**
- does the system meet the user’s goals?
- does the audience understand the visual encoding?
- are aspects of our approach generalizable?
Visualization process

target → translate

evaluate

implement → validate

design
VISUALIZATION PROCESS
WHAT IS A VIS EXPERT?

- data “counselor”
  translate tasks & data into abstractions
- designer
  convert abstractions into visual metaphors
- computer scientist*
  deploy out-of-the-box or develop custom tools

* in contrast to graphic designers, animators, cartographers, etc.
WHAT IS VISUALIZATION RESEARCH?

- technique
  - novel methods & algorithms
  - significant extensions of existing techniques

  designing a new glyph to show uncertainty
  faster algorithm for generating triangular meshes
WHAT IS VISUALIZATION RESEARCH?

systems
- blend of algorithms, design, technical & user requirements
- novel, implemented system that solves a major problem

visually exploring sensitivity of input parameters on climate models
system targeting the workflow of geneticists
What is visualization research?

design studies
- exploration of the choices made when applying methods
- use of visualization to glean insight into a domain

design refinements for genetics data
scientific discoveries found using visualization
What is visualization research?

theory/model
- interpretation of the foundational theory behind visualization
- how visualization complements & exploits cognition

model describing the process of designing visualizations
what constitutes a good colormap
What is visualization research?

evaluation
- empirically judge the effectiveness of methods
- experimental protocols, statistical analysis of results

user study comparing methods for conveying uncertainty
using mechanical turk to assess design
Scientific vs Information

what’s the difference?
- data with or without inherent spatial dimension
- displayed as 3D or 2D
- toolkits typically specialize on one or the other
Raleigh Taylor Instability
Hank Childs, University of Oregon
Pathline tool for comparative functional genomics

Miriah Meyer, University of Utah
SCIENTIFIC VS INFORMATION

distinction disappearing
- problems becoming more complex
- technologies at a sufficient level of advancement
- multi-window systems becoming common

choose right tool for the problem
DATA
SEMANTICS VS TYPES

semantics
- real world meaning of the data
  *Smith* is a word that represents a last name
metadata

type
- interpretation in terms of scales of measurement
  count i.e. quantity, code i.e. numerical category
what are the meaningful mathematical operations?
DATASET TYPES

data relationships
- networks, tabular, log

computational representation
- how is the data stored in the computer
- int, float, double, char, string, etc
**Attribute Types**

categorical
- no implicit ordering
- possibly hierarchical
- can only distinguish if things are the same or different

ex. *types of fruit, movie genres*
Attribute Types

ordered
- has an implicit ordering

ordinal
well defined ordering
cannot do full-fledged arithmetic
ex. t-shirt size, rankings
ATTRIBUTE TYPES

ordered
- has an implicit ordering

quantitative
- measurement of magnitude
  supports arithmetic
  ex. age, temperature, stock price
Attribute Types

ordered

sequential
homogeneous range from min to max
ex. mountain height

diverging
two sequences pointing in opposite direction
ex. full elevation from sea depth to mountain height
Visual Encodings
**Visual Encodings**

relativity of perception
- we judge based on relative, not absolute differences

unframed, unaligned

framed, unaligned

unframed, aligned
Visual Encodings

marks and channels
- a mark is a basic graphical element
- a channel controls the appearance of marks
MARKS

- points
- lines
- areas
CHANNELS

color
- hue
- saturation
- lightness
CHANNELS

displaying

- size

- shape

- orientation
Channel Effectiveness

Effectiveness principle
visual encoding should express all of and only the information in the data

Attribute importance should match the effectiveness of the channel
**CHANNEL EFFECTIVENESS**

- **what**
  - region
  - color hue
  - shape
  - stipple pattern

---

Thursday, November 7, 13
CHANNEL EFFECTIVENESS

how much
- position in common scale
- length (1D size)
- area (2D size)
- lightness
- saturation

effectiveness
Techniques
Techniques - 2D

direct plotting
- graphs
- bar charts
- scatterplots

abstract plotting
- boxplots
- contours
- parallel coordinates
Techniques - 3D

- isosurfacing
- 3D surfaces
- volume rendering
- 3D volumes
- animation
- 2/3D plus time
3D
Common Issues

- inappropriate display choice
- distortions in reality
- variety for the sake of variety
- too much information
- inaccurate encodings
- inconsistent ordering & placement
- inconsistent scales
Best Practices
**Gestalt Principles**

**proximity**
- objects close together or connected will be perceived as a group
**Gestalt Principles**

**similarity**
- objects that share attributes are perceived as a group
**Gestalt Principles**

**enclosures**

- objects that have a boundary are perceived as a group
Gestalt Principles

closure
- open structure can be perceived as closed
Edward Tufte

The Leonard da Vinci of data

- statistician at Yale
- extensive writings on data visualization
- theories/principles for good presentation

“Above all else show the data”
-Tufte 1983

NY Times
Tufte’s Design Principles

maximize ink-data ratio
- the amount of ink on a page should be limited to the non-erasable core of the graphic
Tufte’s Design Principles

avoid chart junk
- remove all visual elements not necessary
Tufte’s Design Principles

Use multifunctioning elements
- graphical elements that serve several purposes at once
Tufte’s Design Principles

Layer Information
- layers that are separate visually can be overlapped
Tufte’s Design Principles

Maintain graphical integrity
- Numbers, areas, size should reflect the actual data
Tufte’s Design Principles

maximize data density
- shrink the data display as much as possible without loosing legibility
Tufte is nice, but not always right...
General Advice
If you want objects to look the same color, make background colors consistent.
CONTRAST EFFECTS

- juxtaposition of colors effects our perception of them
- complimentary colors often most effected
Simultaneous Contrast
SIMULTANEOUS CONTRAST
Use color only when needed to serve a particular communication goal.
Use different colors only when they correspond to differences of meaning in the data.
Use soft, natural colors to display most information and bright and/or dark colors to highlight.
Look to Nature
WE’RE A TINY COMPANY
BUT OUR WORK
IS ABSOLUTELY HUGE
GENERAL ADVICE

Stick with a monochromatic color scheme when encoding quantitative values.
COLOR SCHEMES

- monochromatic
  variations in lightness & saturation of a single color
- analogous
  colors adjacent on the color wheel
- complementary
  two colors opposite on the color wheel
COLOR BREWER

http://colorbrewer2.org/  
- pre-defined color schemes for maps
  sequential  
  optimized for data ordered low to high
  diverging  
  place equal emphasis on mid-range and extremes
  qualitative  
  does not imply an order (categorical data)
Color Scheme Designer

http://colorschemedesigner.com
- develop scheme based on a color model
- test out scheme on light and dark pages
KUHLER

http://kuler.adobe.com
- pick a previously created “theme”
- create a theme from color model
- create a new theme from an image
General Advice

Non-data components should be displayed just visibly enough to perform their role.
CHART of all the IMPORTS and EXPORTS to and from ENGLAND
From the Year 1700 to 1762 by W. Playfair

The Divisions at the Bottom, express YEARS, & those on the Right hand, MILLIONS of POUNDS
Published as the Act directs, 20th Aug. 1763

[Graph showing the balance of trade trend from 1700 to 1762, with years and millions of pounds on the axes.]
General Advice

To accommodate for the colorblind, avoid using a combination of red and green in the same display.
Color Blindness

deficiency in color vision
- typically caused by faulty cone development
- found more in men than women

photopigment genes carried in X-chromosome
- 5-8% of men and 0.5% of women
**Types:**

**Monochromacy**

- **1 dimensional color vision**
  - 2 or 3 cone pigments are missing
  - total color blindness
  - very rare
- **rod monochromacy**: non-functioning or missing cones
- **cone monochromacy**: multiple deficient cones

*Multiple types of dichromacy*
TYPES: DICHROMACY

2 dimensional color vision
- 1 cone pigment is missing
- protanopia: absence of red receptors
- deuteranopia: absence of green receptors
- tritanopia: absence of blue receptors
**Types: Trichromacy**

3 dimensional color vision
- 1 cone is altered in spectral sensitivity
- impairment rather than loss
- **protanomaly**: shift in red, poor red-green discrimination
- **deutanomaly**: shift in green, poor red-green discrimination
  *most common form of color deficiency*
- **tritanomaly**: poor blue-yellow discrimination
DESIGNING FOR COLORBLINDNESS

- vary in hue or saturation or brightness
- monochrome color schemes
- use queues besides/in addition to color

Vischeck/Daltonize
http://www.vischeck.com/
Next Time
tutorials!
- matplotlib (python)
- processing (java)