Week 1: Intro, Marks and Channels

Tamara Munzner
Department of Computer Science
University of British Columbia

JRNL 520M, Special Topics in Contemporary Journalism: Visualization for Journalists
Week 1: 15 September 2015

http://www.cs.ubc.ca/~tmm/courses/journ15
Who’s who

• Instructor: Tamara Munzner
  – UBC Computer Science

• Journalistic kibitzer: Alfred Hermida
  – UBC Journalism

• Guest lecturer and significant labs help: Robert Kosara
  – Research Scientist, Tableau Software
  – previously UNC Charlotte Computer Science
Class time

• 6 weeks, Sep 15 - Oct 20
  – 1 3-hr session per week

• standard week
  – foundations lecture/discussion: 90 min
  – break: 15 min
  – demos: 30 min
  – lab: 45 min

• demo-intensive weeks
  – Week 1 & Week 4: longer demo from guest lecturer Robert Kosara
    – foundations 60 min, break 15 min, demos 60 min, lab 45 min
Structure

• participation
  – attendance and discussion in class, 16%
    • tell me in advance if you’ll miss class (and why)
    • tell when you recover if you were ill

• homework, 84%
  – 6 assignments, 14% each
    • start in lab
    • finish over one week
    • due at start of next class session
  – some solo, some in groups of 2
    • gradual transition from structured to open-ended
    • final assignment: find your own interesting data and design your own visualization for it

• draft plan, may change as pilot continues!
Further reading

• optional textbook for following up on lecture topics
    • http://www.cs.ubc.ca/~tmm/vadbook/
      – library has multiple ebook copies
      – to buy yourself, see course page

• optional papers/books
  – links and references posted on course page
  – if DL links, use library EZproxy from off campus
Finding me

• email is the best way to reach me: tmm@cs.ubc.ca
• office hours by appointment
  – X661 (X-Wing of ICICS/CS bldg)

• course page is font of all information
  – don’t forget to refresh, frequent updates
  – http://www.cs.ubc.ca/~tmm/courses/journ15
Topics

• Week 1
  – Intro
  – Marks and Channels
  – Demo: Tableau I, Kosara

• Week 2
  – Task and Data Abstractions
  – Arrange Tables
  – Demo: TBD

• Week 3
  – Color
  – Arrange Spatial Data
  – Demo: Text Tools & Resources, Brehmer

• Week 4
  – Arrange Networks
  – Demo: Tableau II, Kosara

• Week 5
  – Facet Into Multiple Views
  – Reduce Items and Attributes
  – Demo: TBD

• Week 6
  – Rules of Thumb
  – Putting It All Together
  – Demo: TBD
VAD Ch 1: What’s Vis and Why Do It?

• Why have a human in the decision-making loop?
• Why have a computer in the loop?
• Why use an external representation?
• Why depend on vision?
• Why show the data in detail?
• Why is the vis idiom design space so huge?
• Why focus on tasks and effectiveness?
• Why are there resource limitations?
• Why analyze vis?
Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?...
Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

• don’t need vis when fully automatic solution exists and is trusted
• many analysis problems ill-specified
  – don’t know exactly what questions to ask in advance
• possibilities
  – long-term use for end users (e.g. exploratory analysis of scientific data)
  – presentation of known results
  – stepping stone to better understanding of requirements before developing models
  – help developers of automatic solution refine/debug, determine parameters
  – help end users of automatic solutions verify, build trust
Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception

Why have a computer in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- beyond human patience: scale to large datasets, support interactivity
  - consider: what aspects of hand-drawn diagrams are important?

Why depend on vision?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• human visual system is high-bandwidth channel to brain
  – overview possible due to background processing
    • subjective experience of seeing everything simultaneously
    • significant processing occurs in parallel and pre-attentively

• sound: lower bandwidth and different semantics
  – overview not supported
    • subjective experience of sequential stream

• touch/haptics: impoverished record/replay capacity
  – only very low-bandwidth communication thus far

• taste, smell: no viable record/replay devices
Why show the data in detail?

- summaries lose information
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

### Anscombe’s Quartet

<table>
<thead>
<tr>
<th>Identical statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x mean</td>
<td>9</td>
</tr>
<tr>
<td>x variance</td>
<td>10</td>
</tr>
<tr>
<td>y mean</td>
<td>8</td>
</tr>
<tr>
<td>y variance</td>
<td>4</td>
</tr>
<tr>
<td>x/y correlation</td>
<td>1</td>
</tr>
</tbody>
</table>
Why analyze?

• huge design space
  – visual encoding: combinatorial explosion of choices
  – add interaction: even bigger
  – add data abstraction transformation: truly enormous

• most possibilities ineffective for particular task/data combination
  – implication: avoid random walk, be guided by principles

• analysis framework: scaffold to think systematically about design space
  – ensure that consideration space encompasses full scope of possibilities
  – improve chances that selected solution is good not mediocre
  – next week’s focus: abstractions and idioms, what-why-how
Analysis framework: Four levels, three questions

• **domain** situation
  – who are the target users?

• **abstraction**
  – translate from specifics of domain to vocabulary of vis
  • **what** is shown? **data abstraction**
  • **why** is the user looking at it? **task abstraction**

• **idiom**
  • **how** is it shown?
    • **visual encoding idiom**: how to draw
    • **interaction idiom**: how to manipulate

• **algorithm**
  – efficient computation
Why is validation difficult?

• different ways to get it wrong at each level

- **Domain situation**
  - You misunderstood their needs

- **Data/task abstraction**
  - You’re showing them the wrong thing

- **Visual encoding/interaction idiom**
  - The way you show it doesn’t work

- **Algorithm**
  - Your code is too slow
Why is validation difficult?

- solution: use methods from different fields at each level

- **Domain situation**
  Observe target users using existing tools

- **Data/task abstraction**
  - **Visual encoding/interaction idiom**
    Justify design with respect to alternatives
  - **Algorithm**
    Measure system time/memory
    Analyze computational complexity
    Analyze results qualitatively
    Measure human time with lab experiment (*lab study*)

- **Domain situation**
  Observe target users after deployment (*field study*)

- **Data/task abstraction**
  Measure adoption

---

Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks and effectively.

• tasks serve as constraint on design (as does data)
  – idioms do not serve all tasks equally!
  – challenge: recast tasks from domain-specific vocabulary to abstract forms

• most possibilities ineffective
  – validation is necessary, but tricky
  – increases chance of finding good solutions if you understand full space of possibilities

• what counts as effective?
  – novel: enable entirely new kinds of analysis
  – faster: speed up existing workflows
Why are there resource limitations?

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

• computational limits
  – processing time
  – system memory

• human limits
  – human attention and memory

• display limits
  – pixels are precious resource, the most constrained resource
  – information density: ratio of space used to encode info vs unused whitespace
    • tradeoff between clutter and wasting space, find sweet spot between dense and sparse
VAD Ch 5: Marks and Channels

Channels: Expressiveness Types and Effectiveness Ranks

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape

[VAD Fig 5.1]
Encoding visually

- analyze idiom structure
Definitions: Marks and channels

• marks
  – geometric primitives

• channels
  – control appearance of marks
Encoding visually with marks and channels

- analyze idiom structure
  - as combination of marks and channels

1: vertical position  
mark: line

2: vertical position  
horizontal position  
mark: point

3: vertical position  
horizontal position  
color hue  
mark: point

4: vertical position  
horizontal position  
color hue  
size (area)  
mark: point
Channels

Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Spatial region
Color hue
Motion
Shape
### Channels: Rankings

#### Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

#### Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

- **Effectiveness principle**
  - Encode most important attributes with highest ranked channels

- **Expressiveness principle**
  - Match channel and data characteristics
Accuracy: Fundamental Theory

Steven’s Psychophysical Power Law: $S = I^N$
Accuracy: Vis experiments

Discriminability: How many usable steps?

• must be sufficient for number of attribute levels to show
  – linewidth: few bins

[mappa.mundi.net/maps/maps_014/telegeography.html]
Separability vs. Integrality

Position
+ Hue (Color)

- Fully separable
  - 2 groups each

Size
+ Hue (Color)

- Some interference
  - 2 groups each

Width
+ Height

- Some/significant interference
  - 3 groups total: integral area

Red
+ Green

- Major interference
  - 4 groups total: integral hue
Popout

• find the red dot
  – how long does it take?

• parallel processing on many individual channels
  – speed independent of distractor count
  – speed depends on channel and amount of difference from distractors

• serial search for (almost all) combinations
  – speed depends on number of distractors
Popout

- many channels: tilt, size, shape, proximity, shadow direction, ...
- but not all! parallel line pairs do not pop out from tilted pairs
Grouping

• containment
• connection

• proximity
  – same spatial region
• similarity
  – same values as other categorical channels

Marks as Links

- Containment
- Connection

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape
Relative vs. absolute judgements

• perceptual system mostly operates with relative judgements, not absolute
  – that’s why accuracy increases with common frame/scale and alignment
  – Weber’s Law: ratio of increment to background is constant
    • filled rectangles differ in length by 1:9, difficult judgement
    • white rectangles differ in length by 1:2, easy judgement

Relative luminance judgements

- perception of luminance is contextual based on contrast with surroundings

http://persci.mit.edu/gallery/checkershadow
Relative color judgements

- color constancy across broad range of illumination conditions
Further reading

  – Chap 1: What’s Vis, and Why Do It?
  – Chap 5: Marks and Channels

• Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Jeffrey Heer and Michael Bostock. Proc. CHI 2010

• Perception in Vision web page with demos, Christopher Healey.

Now

• Break (15 min)
• Demo: Guest lecture/demo from Robert Kosara on Tableau
• Lab: you’ll try it!
Lab/Assignment (Updated after class)

• install Tableau on your own laptop
  – using course key from me or individual license key that you request personally

• work through Vienna tutorial (data: Chicago crime 2015, US forest fires)

• work through intro tutorial (data: music sales)

• download 1033 dataset from Tableau Public
  – play with it based on what you learned from Robert’s demo

• pick three datasets from Tableau public
  – visualize them with Tableau with what you learned from demo and tutorials, also try at least two new features for each

• submit next week
  – by 9am Tue, email tmm@cs.ubc.ca with subject JOURN Week 1
  – reflections on what you’ve found in the 7 datasets
    • text illustrated by screenshots of what you’ve created, in PDF format
  – what did you find in the vis?
    • could you tell a story to others? could you get a sense of the story for yourself? did you find nothing useful?