**Information Visualization Midterm Review**

**Tamara Munzner**
Department of Computer Science
University of British Columbia

Last updated: Mar 5, 2020

https://www.cs.ubc.ca/~tamara/essays/IV/SV-38

---

### Data and Dataset Types

**Dataset Types**
- Attributes (rows)
- Items

**Multidimensional Table**
- Cell containing value

**Networks & Datasets**
- Positions (item)
- Geometry

---

### Why?

**Grid of positions**
- Interaction
- Foundations (table, network, spatial)

---

### How?

**Items & Attributes**
- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - "dependent variable"

**Subtopics**
- Nested model
  - four levels: domain, abstraction, idiom, algorithm

---

### What?

**Assignments**
- Topics
  - Data & Task Abstractions
  - Marics & Channels
  - Tables
  - Interactive Views
  - Maps
  - Color

**Midterm logistics**
- time: 75 min
- materials allowed: one-sided "cheat sheet"
  - one side of 8.5"x11" paper
  - no other materials
- bags under desk, phones off and in bag
- do not open exam until told to do so

---

### Milestone 2

- 80% Programming Achievement
- 5% Project Management
- (see update 3/4)
- 15% Writereap

---

### Milestone 3

- Programming Achievement 40%
- Foundations 40%
- Writereap 20%

---

### Milestone 4: four levels of visualization design

- domain situation
- who are the targets users?
- abstraction
  - translate from specific domain to vocabulary of visualization
- what is shown?
- data abstraction
  - how is the user looking at it?
  - task abstraction
  - often must transform data, guided by task
- idiom
  - how is it shown?
- visual encoding idiom how to draw
  - interaction idiom: how to manipulate
  - algorithm
  - efficient computation

---

### Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
- how many items in the dataset?
- what is cardinality of each attribute?
- number of levels for categorical data
- range for quantitative data
- consider whether to transform data
  - guided by understanding of task
Channel effectiveness

- expressiveness
  - match channel and data characteristics
  - channels differ in accuracy of perception
  - distinguishability
  - match available levels in channel w/ data

Separability vs. Integrity

Position
- constraint view: mark type constrains what else can be encoded
  - points: 0 constraints on size, can encode more attributes w/ size & shape
  - areas: 1 constraint on size (length), can still size code other way (width)
  - rectilinear: 2 constraints on size (length/width), cannot size code or shape code
  - quick check: can you size-code another attribute, or is size/shape in use?

Grouping
- containment
  - same spatial region
- similarity
  - same values as other categorical channels

Manipulate

How to handle complexity: 4 families of strategies

- derive new data to show within view
- change view over time
- facet across multiple views

Coordinate views: Design choice interaction

Decomposing color

- first rule of color: do not talk about color!
  - color: confusing if treated as monolithic
  - decompose into three channels
    - ordered can show magnitude
    - luminance: how bright
    - separation: how colorful
    - categorical can show identity
    - hue, what color
  - channels have different properties
    - how much they convey directly to perceptual system
    - how much they convey how many discriminable bins can we use?