Information Visualization
Midterm Review

Tamara Munzner
Department of Computer Science
University of British Columbia

Lect 16, Mar 5 2020

https://www.cs.ubc.ca/~tmm/courses/436V-20
Schedule

• phase change
  – phase 1 done: no more D3 videos, quizzes, programming exercises
  – phase 2 starts: project work
    • Milestone 1 due Friday Saturday (11:59pm)
    • foundations exercises continue in parallel

• schedule shift
  – midterm review & survey today
  – shift to Tuesday:
    • Aggregation 1 lecture
    • Foundations 6 release
Final project marks breakdown

• Final project 30% of total
  – breakdown: M1 15%, M2 35%, M3 50%
  – of total: M1 4.5%, M2 10.5%, M3 15%

• Milestone 1
  – Foundations 60% [Sec 1-5]
  – Project Management 15% [Sec 6]
  – Writeup 25% [overall]

• Milestone 2
  – 80% Programming Achievement
  – 5% Project Management
  • (see update 3/4)
  – 15% Writeup

• Milestone 3
  – Programming Achievement 40%
    • includes demo
  – Foundations 40%
  – Writeup 20%
Survey

• mid-semester survey
• anonymous

https://ubc.ca1.qualtrics.com/jfe/form/SV_50zwSEo5DihPzIV

• on socrative, pick true when done
Midterm Review
Midterm material covered

• Topics
  – Intro
  – Data & Task Abstractions
  – Marks & Channels
  – Tables
  – Interactive Views
  – Maps
  – Color

• Assignments
  – F1
  – F2
  – F3
  – F4 (will be returned Wed)
Midterm logistics

• time: 75 min
• materials allowed: one-sided "cheat sheet"
  – one side of 8.5"x11" paper
  – we'll check it when we check your ids
  – no other materials
• bags under desk, phones off and in bag
• do not open exam until told to do so
Midterm scope

- **scope:** emphasis on foundations material
  - What kind of attribute is X? (categorical, ordinal, quantitative)
  - What kind of dataset is X? (table, network, spatial)
  - What channels are in use in this visual encoding?
  - Map this domain-language description of tasks and data into abstractions
  - Analyze this existing visualizations by breaking down into marks and channels
  - Critique suitability of this existing visual encoding for abstract task+data combination
    - including scalability assessment for #items, #attributes, # levels within an attribute
  - Propose appropriate visual encoding for task+data combination
    - and provide rationale to justify your design choices versus key alternatives
Midterm scope

• scope: emphasis on foundations material
  – How is spatial position being used to arrange data?
    • express values
    • separate, order, align
    • use given spatial data
  – Discuss tradeoffs between major visual encoding choices
    • choropleth vs symbol maps vs cartograms for maps
    • rectilinear vs radial vs parallel layouts
Subtopics

– Nested model
  • four levels: domain, abstraction, idiom, algorithm

– Data
  • items vs attributes
  • attribute types: categorical, ordered, quantitative
  • dataset types: tables, networks, spatial

– Tasks
  • action-target pairs
  • query: one/sum/all

– Marks and Channels
  • channel types (magnitude vs identity)
  • accuracy, discriminability, separability, popout
  • perceptual system mostly operates with relative judgements, not absolute
Subtopics

– Interactive Views
  • selection and highlighting strategies
  • navigation strategies
  • types of multiple views: multiform, overview/detail same encoding, overview/detail multiform, small multiples
  • strengths and weaknesses of juxtapose vs superimpose
  • impact of partitioning strategies

– Color
  • channel characteristics for hue, saturation, value
  • sequential vs diverging for quantitative attributes
  • univariate vs bivariate
  • color deficiency: nature of problem and strategies to address it
Nested model: Four levels of visualization design

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

• **abstraction**
  – translate from specifics of domain to vocabulary of visualization
    • **what** is shown? data abstraction
    • **why** 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 and Dataset Types

- **Data Types**
  - Items
  - Attributes
  - Links
  - Positions
  - Grids

- **Dataset Availability**
  - Static
  - Dynamic

#### Attributes

- **Attribute Types**
  - Categorical
  - Ordered
  - Quantitative

#### Ordering Direction

- Sequential
- Diverging
- Cyclic

#### Dataset Types

- **Tables**
- **Networks & Trees**
- **Fields**
- **Geometry**
- Clusters, Sets, Lists

#### What?

- Datasets
- What?
- Attributes
- Why?
- 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
  – eg horsepower, make for car
  – "dependent variable"

 attributes: name, age, shirt size, fave fruit

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Shirt Size</th>
<th>Favorite Fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amy</td>
<td>8</td>
<td>S</td>
<td>Apple</td>
</tr>
<tr>
<td>Basil</td>
<td>7</td>
<td>S</td>
<td>Pear</td>
</tr>
<tr>
<td>Clara</td>
<td>9</td>
<td>M</td>
<td>Durian</td>
</tr>
<tr>
<td>Desmond</td>
<td>13</td>
<td>L</td>
<td>Elderberry</td>
</tr>
<tr>
<td>Ernest</td>
<td>12</td>
<td>L</td>
<td>Peach</td>
</tr>
<tr>
<td>Fanny</td>
<td>10</td>
<td>S</td>
<td>Lychee</td>
</tr>
<tr>
<td>George</td>
<td>9</td>
<td>M</td>
<td>Orange</td>
</tr>
<tr>
<td>Hector</td>
<td>8</td>
<td>L</td>
<td>Loquat</td>
</tr>
<tr>
<td>Ida</td>
<td>10</td>
<td>M</td>
<td>Pear</td>
</tr>
<tr>
<td>Amy</td>
<td>12</td>
<td>M</td>
<td>Orange</td>
</tr>
</tbody>
</table>

item: person
Attribute types

• which classes of values & measurements?

• categorical (nominal)
  – compare equality
  – no implicit ordering

• ordered
  – ordinal
    • less/greater than defined
  – quantitative
    • meaningful magnitude
    • arithmetic possible
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
• \{action, target\} pairs
  – discover distribution
  – compare trends
  – locate outliers
  – browse topology
Marks: Constrained vs encodable

• math view: geometric primitives have dimensions

  ➔ Points ➔ Lines ➔ Areas

  0D
  ➔  ➔  ➔

  1D
  ➔  ➔  ➔

  2D

• constraint view: mark type constrains what else can be encoded
  – points: 0 constraints on size, can encode more attributes w/ size & shape
  – lines: 1 constraint on size (length), can still size code other way (width)
  – areas: 2 constraints on size (length/width), cannot size code or shape code
    • interlocking: size, shape, position

• quick check: can you size-code another attribute, or is size/shape in use?
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

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

[www.cs.ubc.ca/~tmm/talks.html#vad20alum](http://www.cs.ubc.ca/~tmm/talks.html#vad20alum)
Channel effectiveness

• accuracy: how precisely can we tell the difference between encoded items?
• discriminability: how many unique steps can we perceive?
• separability: is our ability to use this channel affected by another one?
• popout: can things jump out using this channel?
Separability vs. Integrality

Position
+ Hue (Color)

2 groups each
Fully separable

Size
+ Hue (Color)

2 groups each
Some interference

Width
+ Height

3 groups total: integral area
Some/significant interference

Red
+ Green

4 groups total: integral hue
Major interference
Grouping

- containment
- connection

Marks as Links

- **Containment**
- **Connection**

Identity Channels: Categorical Attributes

- **Spatial region**
- **Color hue**
- **Motion**
- **Shape**

- **Proximity**
  - same spatial region

- **Similarity**
  - same values as other categorical channels
### How?

<table>
<thead>
<tr>
<th>Encode</th>
<th>Manipulate</th>
<th>Facet</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arrange</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡ Express</td>
<td>➡ Change</td>
<td>➡ Juxtapose</td>
<td>➡ Filter</td>
</tr>
<tr>
<td>➡ Order</td>
<td>➡ Select</td>
<td>➡ Partition</td>
<td>➡ Aggregate</td>
</tr>
<tr>
<td>➡ Use</td>
<td>➡ Navigate</td>
<td>➡ Superimpose</td>
<td>➡ Embed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Map</th>
<th>Manipulate</th>
<th>Facet</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>➡ from categorical and ordered attributes</td>
<td>➡ Change</td>
<td>➡ Juxtapose</td>
<td>➡ Filter</td>
</tr>
<tr>
<td>➡ Color</td>
<td>➡ Select</td>
<td>➡ Partition</td>
<td>➡ Aggregate</td>
</tr>
<tr>
<td>➡ Hue</td>
<td>➡ Navigate</td>
<td>➡ Superimpose</td>
<td>➡ Embed</td>
</tr>
<tr>
<td>➡ Saturation</td>
<td>➡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡ Luminance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡ Size, Angle, Curvature, ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡ Shape</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➡ Motion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**What?**

**Why?**

**How?**
Arrange tables

Express Values

Separate, Order, Align Regions

Separate, Order, Align

1 Key 2 Keys 3 Keys Many Keys

List Recursive Subdivision

Axis Orientation

Rectilinear Parallel Radial

Layout Density

Dense Space-Filling

List Recursive Subdivision

Rectilinear Parallel Radial

Many Keys

Recursive Subdivision
Manipulate

Change over Time
- Select

Navigate
- Item Reduction
  - Zoom
    - Geometric or Semantic
  - Pan/Translate
  - Constrained

Coordinate views: Design choice interaction

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Subset</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Redundant</td>
<td>Same form, Overview/Detail</td>
</tr>
<tr>
<td></td>
<td>Small Multiples</td>
</tr>
<tr>
<td>Multiform</td>
<td>Multiform, Overview/Detail</td>
</tr>
<tr>
<td></td>
<td>No Linkage</td>
</tr>
<tr>
<td>Different</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Subset</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Small Multiples</td>
</tr>
<tr>
<td></td>
<td>No Linkage</td>
</tr>
</tbody>
</table>
Decomposing color

• first rule of color: do not talk about color!
  – color is confusing if treated as monolithic

• decompose into three channels
  – ordered can show magnitude
    • luminance: how bright
    • saturation: how colorful
  – categorical can show identity
    • hue: what color

• channels have different properties
  – what they convey directly to perceptual system
  – how much they can convey: how many discriminable bins can we use?
Colormaps

- Categorical
  -
- Ordered
  - Sequential
  - Diverging
- Bivariate

How to handle complexity: 4 families of strategies

- **Derive**
  - derive new data to show within view
  - change view over time
  - facet across multiple views
  - reduce items/attributes within single view

- **Manipulate**
  - Change
  - Select
  - Navigate

- **Facet**
  - Juxtapose
  - Partition
  - Superimpose

- **Reduce**
  - Filter
  - Aggregate
  - Embed

• derive new data to show within view
• change view over time
• facet across multiple views
• reduce items/attributes within single view