Information Visualization
Midterm Review - Updates

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 11 async lecture, Jan 2021

https://www.students.cs.ubc.ca/~cs-436v/21Jan/
Midterm logistics UPDATED

• time: 75 min content, 5 min image upload
  – 75 points (to help you budget your time, 1 pt ~= 1 minute)
  – 5 extra minutes for image upload
  – materials allowed: open book/notes, but no communication with other people

• you will take exam entirely within Canvas
  – multiple choice & true/false questions (randomized)
  – also short answer enter & image upload within Canvas
    • we export to Gradescope, so it's not your problem to switch platforms in the middle
    • short answer: much much better to type, avoid hand writing if at all possible
    • sketches (can include annotations): sketch on paper and take picture, or sketch on computer
      – hardcopy: use blank paper, one sheet per question, take picture & upload
      – softcopy: use drawing tools and text editing tools of your choice, & copy/paste
Midterm material covered

• Topics
  – Intro
  – Data & Task Abstractions
  – Marks & Channels
  – Multivariate Tables
  – Interactive Views
  – Maps
  – Color
  – Aggregation
  – Networks & Trees
  – Rules of Thumb

• Assignments
  – F1
  – F2
  – F3
Subtopics UPDATED

– Multivariate Tables
  • how many keys?
  • express values vs separate, order, align regions
  • rectilinear vs radial vs parallel axes
  • information-dense layouts

– Maps
  • thematic map types: choropleth, symbol, cartogram, dot density
Multivariate Tables
Arrange tables

- Express Values
- Separate, Order, Align Regions
  - Separate
  - Order
  - Align
- Axis Orientation
  - Rectilinear
  - Parallel
  - Radial
- Layout Density
  - Dense
- 1 Key
  - List
- 2 Keys
  - Matrix
Keys and values

- **key**
  - independent attribute
  - used as unique index to look up items
  - simple tables: 1 key
  - multidimensional tables: multiple keys

- **value**
  - dependent attribute, value of cell

- classify arrangements by key count
  - 0, 1, 2, ...

- **List**
- **Matrix**

- **Tables**

- **Multidimensional Table**

- **Networks**
  - Link
  - Node (item)

- **Trees**
  - Fields (Continuous)
  - Attributes (columns)
  - Value in cell

- **Geometry** (Spatial)
  - Position

- **Dataset Types**

- **Arrange Tables**
- **Express Values**
- **Separate, Order, Align Regions**
- **Axis Orientation**
- **Layout Density**
- **Dense**
- **Space-Filling**
Some keys: Categorical regions

- **Separate**: contiguous bounded areas distinct from each other
  - using space to *separate* (proximity)
  - following expressiveness principle for categorical attributes

- **Order** and **Align** regions

- **regions**: contiguous bounded areas distinct from each other
  - using space to *separate* (proximity)
  - following expressiveness principle for categorical attributes

- **use ordered attribute to order and align regions**
Orientation limitations

• rectilinear: scalability wrt #axes
  • 2 axes best, 3 problematic, 4+ impossible

• parallel: unfamiliarity, training time

• radial: perceptual limits
  – polar coordinate asymmetry
    • angles lower precision than length
    • nonuniform sector width/size depending on radial distance
  – frequently problematic
    • but sometimes can be deliberately exploited!
      – for 2 attrs of very unequal importance

Maps
Key question:

Does the given spatial/geographic data matter for my task?

» *Position* is the most effective visual channel
  
  » we don’t want to waste it for non-relevant geographic spatial information

» A geo map is not always the best or only solution.
Geographic Map

Interlocking marks

- shape coded
- area coded
- position coded

• cannot encode another attribute with these channels, they're "taken"
Thematic maps

• show spatial variability of attribute ("theme")
  – combine geographic / reference map with (simple, flat) tabular data
  – join together
    • region: interlocking area marks (provinces, countries with outline shapes)
      – also could have point marks (cities, locations with 2D lat/lon coords)
    • region: categorical key attribute in table
      – use to look up value attributes

• major idioms
  – choropleth
  – symbol maps
  – cartograms
  – dot density maps
Choropleth map: Pros & cons

• pros
  – easy to read and understand
  – well established visualization (no learning curve)
  – data is often collected and aggregated by geographical regions

• cons
  – most effective visual variable used for geographic location
  – visual salience depends on region size, not true importance wrt attribute value
    • large regions appear more important than small ones
  – color palette choice has a huge influence on the result
Symbol map: Pros & cons

• pros
  – somewhat intuitive to read and understand
  – mitigate problems with region size vs data salience
    • marks: symbol size follows attribute value
    • glyphs: symbol size can be uniform

• cons
  – possible occlusion / overlap
    • symbols could overlap each other
    • symbols could occlude region boundaries
  – complex glyphs may require explanation / training
Cartogram: Pros & cons

• pros
  – can be intriguing and engaging
  – best case: strong and surprising size disparities

• cons
  – require substantial familiarity with original dataset & use of memory
    • compare distorted marks to memory of original marks
    • mitigation strategies: transitions or side by side views
  – major distortion is problematic
    • may be aesthetically displeasing
    • may result in unrecognizable marks
  – difficult to extract exact quantities
Dot density maps: Pros and cons

• pros
  – straightforward to understand
  – avoids choropleth non-uniform region size problems

• cons
  – challenge: normalization
    • many dot maps primarily show population density
      (with which the target variable is correlated)
      instead of the effect of interest
    • same challenge as choropleths
  – perceptual disadvantage:
    difficult to extract quantities
  – performance disadvantage:
    rendering many dots can be slow