

## Data, Tasks, and Clients

CMSC 436/636

Data Visualization

**Penny Rheingans**

University of Maryland Baltimore County



## Announcements

- Project teams by Thurs
- Paper reflection due by Thurs
- Quiz schedule on website; also laptop exercise dates
- Questions?



## Data Taxonomy

- Can characterize data by its characteristics
- Can generalize about data with similar characteristics
  - potential problems
  - natural visualization techniques
  - ease of implementation



## Data Items

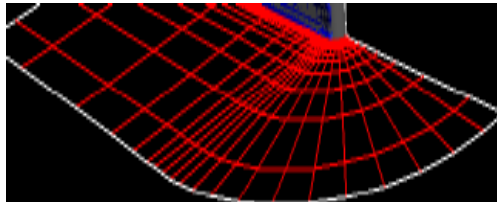
- Entities
- Attributes
- Relations

## Data Characteristics: Continuity

- Continuity
  - discrete: anything sampled or stored
    - ex: computational model, CT scan
    - issues:
      - representation error
      - possible aliasing
      - artifacts of sampling
  - continuous: only implicitly defined
    - ex: mathematical functions, predictive model

## Data Characteristics: Structure

- Geometry vs Topology
- Topological Structure
  - Structured
    - Inherent spatial relationship among points (gridded)
    - Common grid types
      - Regular
      - Rectilinear
      - Curvilinear
    - Advantages
      - easy computation
      - possibly efficient storage (for densely populated grids)

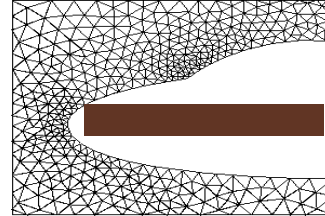


## Data Characteristics: Structure

- Structure

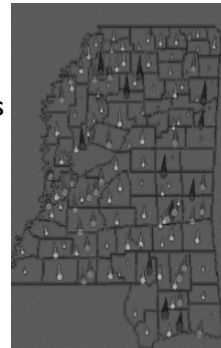
- Irregularly structured

- non-grid connectivity
    - ex: FEM results, surface meshes
    - advantages:
      - flexibility



- Completely unstructured

- no known spatial relationship among points
    - ex: pollution monitors, documents, atoms
    - advantages:
      - flexibility
      - efficient storage (for sparsely populated grids)



## Data Characteristics: Dimension

- Dimensionality

- # independent variables (usually # spatial/temporal variables)
  - commonly:
    - 2D
      - ex: weather info at ground, xray
    - 3D
      - ex: weather info in atmosphere, CT/MRI scan
    - $n$  D
      - ex: census info, stock market conditions, document word frequency
  - Grid dimensions may differ from spatial dimensions

## Data Characteristics: Multiple

- Number of variables per position
  - scalar
    - one value
      - ex: temperature, rainfall, or wind speed
  - multivariate:
    - multiple scalars
      - ex: temperature, rainfall, and wind speed
    - vector
      - ex: wind direction
    - tensor
      - ex: stress and strain forces
- Multivariate vs multidimensional

## Data Characteristics: Scale

- Types
  - nominal
    - categories or identifiers
    - ex: county, land use, ethnicity, tissue type
  - ordinal
    - ordered values
    - ex: preference, ranking
  - integer
    - constant step size
    - ex: test scores, degrees Fahrenheit
  - ratio
    - meaningful zero
    - ex: degrees Kelvin, income, wind speed

## Criteria for Internal Representation

- Compact
  - efficient memory use
  - ex: unstructured schemes, sparse matrices, shared verts
- Efficient
  - computationally accessible
  - retrieve and store in constant time

## Criteria for Internal Representation

- Mappable
  - straight-forward conversions
    - native --> rep: simple conversion, no info lost
    - rep --> graphics prim: fast for interactive display
- Minimal coverage
  - manageable # options
  - few variants which work for wide variety of data
- Simple
  - easier to use
  - easier to optimize
  - errors less likely

## Visualization Tasks

- See values
  - extrema
  - anomalies
  - boundaries/thresholds
  - distribution / structure
- See multiple variables
  - relationships
- See flow/change
- Understand process

## Data Posters Exercise

- Break into five groups, discuss a poster
  - What are data items displayed?
    - Item, type, number
  - What elements of the visual vocabulary are used to display data items?
  - What design choices have been made?
    - Location, color, glyph
  - What makes the strongest impression?
  - What information is obscured?
  - What discoveries about the data can you make from the visualization?
- Report back to class

## Toolsmith Paper Highlights

- Computer scientist as toolsmith -- our success is in the success of our users
- Intelligence amplification -- a machine and a mind together can beat a mind-imitating machine
- Collaboration
  - Driving problem approach: scale, honest, whole problem, new challenges, fun
  - Costs of collaboration

## Interview Types

- Informal
  - Casual conversation without obvious information seeking
- Unstructured
  - Interviewer has plan for topics, but lets interviewee lead direction
- Semistructured
  - Questions from interview guide with followup
- Structured
  - Set questions in scripted order



## General Interview Guide

- What are the questions you are trying to answer with this data? What are your goals?
- Where did the data come from? What is the structure of the data?
- What do you expect to find in the data?
- What are your current methods for analyzing the data?
- What do you want to do with the data that isn't currently possible?

## Client Exercise

- Divide into groups of four
- Role play scenarios with following roles
  - Client
  - Interviewer
  - Coach
  - Evaluator
- Share observations

