What does data mean?

• What does this sequence of six numbers mean?
  – two points far from each other in 3D space
  – two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link

Basil, 7, S, Pear

• What about this data?
  – food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
  – Basil Park neighborhood of city had 7 inches of snow

• lab rat Basil made 7 attempts to find way through south section of maze, these trials used pear as reward food

Now what?

• semantics: real-world meaning

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Shirt Size</th>
<th>Favorite Fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben</td>
<td>18</td>
<td>M</td>
<td>Apple</td>
</tr>
<tr>
<td>Alex</td>
<td>19</td>
<td>S</td>
<td>Pear</td>
</tr>
<tr>
<td>John</td>
<td>20</td>
<td>L</td>
<td>Orange</td>
</tr>
<tr>
<td>Mary</td>
<td>21</td>
<td>M</td>
<td>Blueberry</td>
</tr>
<tr>
<td>Lisa</td>
<td>22</td>
<td>L</td>
<td>Peach</td>
</tr>
<tr>
<td>Sam</td>
<td>23</td>
<td>S</td>
<td>Orange</td>
</tr>
<tr>
<td>Mark</td>
<td>24</td>
<td>M</td>
<td>Apple</td>
</tr>
</tbody>
</table>

Other data types

• links
  – express relationship between two items
  – eg friendship on facebook, interaction between proteins
• positions
  – spatial data: location in 2D or 3D
  – items in photos, rows in MRI scan, latitude/longitude
  – (grids) – sampling strategy for continuous data

Information Visualization
Data Abstraction

Items & Attributes

• item: individual entity, discrete
  – eg patient, car, stock, city
  – "independent variable"
• attributes: property that is measured, observed, logged...
  – eg height, blood pressure for patient
  – eg horsepower, make for car
  – "dependent variable"

Mismatches: Common problem

• flat table
  – one item per row
  – each column is attribute
  – cell holds value

Dataset types

• flat table
  – one item per row
  – each column is attribute
  – cell holds value for item-attribute pair
  – unique key (could be implicit)

What: Data Abstraction

How to evaluate a visualization: So many methods, how to pick?

• Computational benchmarks?
  – quant: system performance, memory
• User study in lab setting?
  – quant: (human) time and error rates, preferences
• Analysis of results?
  – quant: metrics computed on result images
  – qual: interviews with users, case studies, observations

Nested model: Four levels of visualization design

• domain situation
  – who are the target users?

interdisciplinary: need methods from different fields at each level

Data Abstraction

Trees and Dendrograms

• trees
  – express relationship between two items
  – eg family tree, genealogy

Attributes (columns)

• flat table
  – one item per row
  – each column is attribute
  – cell holds value

Dataset types

• flat table
  – one item per row
  – each column is attribute
  – cell holds value for item-attribute pair
  – unique key (could be implicit)
Data and Dataset Types

Dataset Types

- Dataset Availability
  - Static
  - Dynamic

- Other data concerns
  - derived attribute: compute from originals
    - simple change of type
    - acquire additional data
    - complex transformation
  - reordered data"