Announcements

Auditors, *please* enroll in the class (1 unit, P/NP)
- Requirements: Come to class and participate (online as well)
- Requirements: Assignment 1a and 1b

Enrollees, class participation requirements
- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question on wiki within a week of each lecture

All, add yourself to participants page on the wiki

Class wiki
http://vis.berkeley.edu/courses/cs294-10-fa07/wiki/
Assignment 1a: Good and Bad Vis.

Find two visualizations one **good** and one **bad**

Use original sources
- Journals
- Science magazines
- Newspapers
- Textbooks

Make wiki page
- Clearly mark as good or bad
- Provide short explanation
- Be prepared to succinctly describe in class on Wed Sep. 5

Due before class on Sep 5, 2007

---

Data and Image Models
The big picture

- task
- data
  - physical type: int, float, etc.
  - abstract type: nominal, ordinal, etc.
- domain
  - metadata
  - semantics
  - conceptual model

processing algorithms → image
  - visual channel
  - retinal variables

mapping
  - visual encoding
  - visual metaphor

[Based on slide from Munzner]

Topics

Properties of data or information
Properties of the image
Mapping data to images
Data models vs. Conceptual models

Data models are low level descriptions of the data
- Math: Sets with operations on them
- Example: integers with + and × operators

Conceptual models are mental constructions
- Include semantics and support reasoning

Examples (data vs. conceptual)
- (1D floats) vs. Temperature
- (3D vector of floats) vs. Space
Relational data model

- Records are fixed-length tuples
- Each column (attribute) of tuple has a domain (type)
- Relation is schema and a table of tuples
- Database is a collection of relations

Example: Digital cameras
Relational algebra [Codd]

Data transformations (SQL)
- Selection (SELECT)
- Projection (WHERE)
- Sorting (ORDER BY)
- Aggregation (GROUP BY, SUM, MIN, ...)
- Set operations (UNION, ...)
- Join (INNER JOIN)

Statistical data model

Variables or measurements
Categories or factors or dimensions
Observations or cases
Statistical data model

Variables or measurements
Categories or factors or dimensions
Observations or cases

<table>
<thead>
<tr>
<th>Month</th>
<th>Control</th>
<th>Placebo</th>
<th>300 mg</th>
<th>450 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>165</td>
<td>163</td>
<td>166</td>
<td>168</td>
</tr>
<tr>
<td>April</td>
<td>162</td>
<td>159</td>
<td>161</td>
<td>163</td>
</tr>
<tr>
<td>May</td>
<td>164</td>
<td>158</td>
<td>161</td>
<td>153</td>
</tr>
<tr>
<td>June</td>
<td>162</td>
<td>161</td>
<td>158</td>
<td>160</td>
</tr>
<tr>
<td>July</td>
<td>166</td>
<td>158</td>
<td>160</td>
<td>148</td>
</tr>
<tr>
<td>August</td>
<td>163</td>
<td>158</td>
<td>157</td>
<td>150</td>
</tr>
</tbody>
</table>

Blood Pressure Study (4 treatments, 6 months)

Dimensions and measures

Independent vs. dependent variables
- Example: $y = f(x,a)$
- Dimensions: Domain($x$) $\times$ Domain($a$)
- Measures: Range($y$)
<table>
<thead>
<tr>
<th>Case</th>
<th>Species_No</th>
<th>Species</th>
<th>Organ</th>
<th>Width</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>4.3</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>5.8</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>4.6</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>7.2</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>3.6</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>4.4</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>6.9</td>
</tr>
</tbody>
</table>

Sepal and petal lengths and widths for three species of iris (Fisher 1936).

Format of the data in Appendix 14, pp. 365-366
Chambers, Cleveland, Kleiner, Tukey, *Graphical Methods for Data Analysis*
Data cube

Measure

Width
Length

Petal
Sepal

Organ

I. setosa
I. versicolor
I. virginica

Species

Projections summarize data

Multiscale visualization using data cubes [Stolte et al. 02]
Taxonomy

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)

Are there others?

The eyes have it: A task by data type taxonomy for information visualization [Schneiderman 96]

Types of variables

Physical types
- Characterized by storage format
- Characterized by machine operations

Example:
bool, short, int32, float, double, string, ...

Abstract types
- Provide descriptions of the data
- May be characterized by methods/attributes
- May be organized into a hierarchy

Example:
plants, animals, metazoans, ...
Nominal, ordinal and quantitative

N - Nominal (labels)
- Fruits: Apples, oranges, …

O - Ordered
- Quality of meat: Grade A, AA, AAA

Q - Interval (Location of zero arbitrary)
- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Like a geometric point. Cannot compare directly
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)
- Physical measurement: Length, Mass, Temp, …
- Counts and amounts
- Like a geometric vector, origin is meaningful

S. S. Stevens, On the theory of scales of measurements, 1946

From data model to data type

Data model
- 32.5, 54.0, -17.3, …
- floats

Conceptual model
- Temperature

Data type
- Burned vs. Not burned (N)
- Hot, warm, cold (O)
- Continuous range of values (Q)

[based on slide from Munzner]
Sepal and petal lengths and widths for three species of iris (Fisher 1936).
Visual language is a sign system

Images perceived as a set of signs
Sender encodes information in signs
Receiver decodes information from signs

Jacques Bertin
Semiology of Graphics, 1983
Information in position

1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

:. Encode quantitative variables (Q)

"Resemblance, order and proportional are the three signfields in graphics." - Bertin
Information in color and value

Value is perceived as ordered
∴ Encode ordinal variables (O)

Encode continuous variables (Q) [not as well]

Hue is normally perceived as unordered
∴ Encode nominal variables (N) using color

Visual variables

- Position
- Size
- Value
- Texture
- Color
- Orientation
- Shape

Note: Bertin does not consider 3D or time
Note: Card and Mackinlay extend the number of vars.
Bertins’ “Levels of Organization”

<table>
<thead>
<tr>
<th>Position</th>
<th>N</th>
<th>O</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>N</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Value</td>
<td>N</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Texture</td>
<td>N</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Bertin actually breaks visual variables down into differentiating (≠) and associating (≡)

Note: Q < O < N

Encoding rules
Univariate data

A B C D

[based on slide from Stasko]
Bivariate data

Scatter plot is common

[based on slide from Stasko]

Trivariate data

3D scatter plot is possible

[based on slide from Stasko]
Three variables

Two variables [x,y] can map to points
- Scatterplots, maps, …

Third variable [z] must use …
- Color, size, shape, …

Large design space (visual metaphors)

Multidimensional data

How many variables can be depicted in an image?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“With up to three rows, a data table can be constructed directly as a single image ... However, an image has only three dimensions. And this barrier is impassible.”

Bertin
Deconstructions

Stock chart from the late 90s

- March 1998: Bert Ellis beside OIL. It is backed initially with money from Values Co., a New York investment firm, and Ellis.
- Dec. 19, 1998: OIL, shares along the way, where Ellis predicts the stock will hit $100 in 2000. Stock was trading at under $30 a share.
- Jan. 31, 2000: OIL exits biggest tragedy. Ellis is an OIL. Ellis is CEO, Ellis is a CEO. Ellis used to work for OIL.
- Jan. 7, 2000: OIL exits biggest tragedy. Ellis is an OIL. Ellis is CEO, Ellis is a CEO. Ellis used to work for OIL.
- July 1, 1999: OIL exits biggest tragedy. Ellis is an OIL. Ellis is CEO, Ellis is a CEO. Ellis used to work for OIL.
- April 30, 1999: OIL exits biggest tragedy. Ellis is an OIL. Ellis is CEO, Ellis is a CEO. Ellis used to work for OIL.
Stock chart from the late 90s

- x-axis: time (Q)
- y-axis: price (Q)

Playfair 1786
Playfair 1786

- x-axis: year (Q)
- y-axis: currency (Q)
- color: imports/exports (N, O)

Wattenberg 1998

http://www.smartmoney.com/marketmap/