CSE 512 - Data Visualization

Data and Image Models

Jeffrey Heer  University of Washington
Last Time:
Value of Visualization
The Value of Visualization

Record information
  Blueprints, photographs, seismographs, ...

Analyze data to support reasoning
  Develop and assess hypotheses
  Find patterns / Discover errors in data
  Expand memory

Communicate information to others
  Share and persuade
  Collaborate and revise
E.J. Marey’s sphygmograph [from Braun 83]
Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
“to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears”

1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale
InfoVis vs. SciVis?
Informative vs. Aesthetic?

wind map

April 1, 2015
11:35 pm EST
(time of forecast download)

top speed: 30.5 mph
average: 10.2 mph

1 mph
3 mph
5 mph
10 mph
15 mph
30 mph
Data & Image Models
The Big Picture

**Task**
- questions, goals
- assumptions

**Data**
- physical data type
- conceptual data type

**Domain**
- metadata
- semantics
- conventions

**Processing**
- algorithms

**Mapping**
- visual encoding

**Image**
- visual channel
- graphical marks
Topics

Properties of Data
Properties of Images
Mapping Data to Images
Data
Data Models / Conceptual Models

**Data models** are low-level descriptions
Math: sets with operations on them
Example: integers with + and \(x\) operators

**Conceptual models** are mental constructions
Include semantics and support reasoning

**Examples** *(data vs. conceptual)*
1D floats vs. temperatures
3D vector of floats vs. spatial location
Taxonomy of Data Types (?)

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 [Shneiderman 96]
Nominal, Ordinal & Quantitative
Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)
  · Fruits: apples, oranges, ...
Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)
  • Fruits: apples, oranges, ...

O - Ordered
  • Quality of meat: Grade A, AA, AAA
Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)
  - 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)
  - Only differences (i.e. intervals) may be compared
Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)
  - 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)
  - Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)
  - Physical measurement: Length, Mass, Temp, ...
  - Counts and amounts
Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)
- Operations: =, ≠

O - Ordered
- Operations: =, ≠, <, >

Q - Interval (location of zero arbitrary)
- Operations: =, ≠, <, >, -
- Can measure distances or spans

Q - Ratio (zero fixed)
- Operations: =, ≠, <, >, -, %
- Can measure ratios or proportions
From Data Model to N, O, Q

Data Model
32.5, 54.0, -17.3, ...
Floating point numbers

Conceptual Model
Temperature (°C)

Data Type
Burned vs. Not-Burned (N)
Hot, Warm, Cold (O)
Temperature Value (Q)
Sepal and petal lengths and widths for three species of iris [Fisher 1936].

<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>1</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>I. Versicina</td>
<td>Petal</td>
<td>24</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>31</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>28</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>23</td>
<td>51</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>23</td>
<td>51</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>36</td>
<td>46</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>31</td>
<td>69</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>33</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>20</td>
<td>52</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>19</td>
<td>51</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>17</td>
<td>45</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>26</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>25</td>
<td>49</td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>ID</td>
<td>Case</td>
<td>Species No</td>
<td>Species</td>
<td>Organ</td>
<td>Width</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------------</td>
<td>---------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>33</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>20</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>30</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>32</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>19</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>12</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>3</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>36</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>27</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>26</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>17</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>10</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>3</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>32</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>23</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>3</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
</tr>
</tbody>
</table>
Dimensions & Measures

**Dimensions** (~ independent variables)
Discrete variables describing data (N, O)
Categories, dates, binned quantities

**Measures** (~ dependent variables)
Data values that can be aggregated (Q)
Numbers to be analyzed
Aggregate as sum, count, avg, std. dev…
Example: U.S. Census Data
Example: U.S. Census Data

People Count: # of people in group
Year: 1850 - 2000 (every decade)
Age: 0 - 90+
Sex: Male, Female
Marital Status: Single, Married, Divorced, ...
Example: U.S. Census

People Count
Year
Age
Sex
Marital Status

2,348 data points
### Census: N, O, Q?

<table>
<thead>
<tr>
<th>People Count</th>
<th>Q-Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Q-Interval (O)</td>
</tr>
<tr>
<td>Age</td>
<td>Q-Ratio (O)</td>
</tr>
<tr>
<td>Sex</td>
<td>N</td>
</tr>
<tr>
<td>Marital Status</td>
<td>N</td>
</tr>
<tr>
<td>Census: Dimension or Measure?</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td></td>
</tr>
<tr>
<td>People Count</td>
<td>Measure</td>
</tr>
<tr>
<td>Year</td>
<td>Dimension</td>
</tr>
<tr>
<td>Age</td>
<td>Depends!</td>
</tr>
<tr>
<td>Sex</td>
<td>Dimension</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Dimension</td>
</tr>
</tbody>
</table>
Data Transformation
Relational Data Model

Represent data as a **table** *(relation)*

Each **row** *(tuple)* represents a record
  Each record is a fixed-length tuple

Each **column** *(attribute)* represents a variable
  Each attribute has a **name** and a **data type**

A table’s **schema** is the set of names and types

A **database** is a collection of tables *(relations)*
Relational Algebra [Codd ’70]

Data Transformations (sql)
Projection (select) - selects columns
Selection (where) - filters rows
Sorting (order by)
Aggregation (group by, sum, min, max, ...)
Combine relations (union, join, ...)
Roll-Up and Drill-Down

Want to examine marital status in each decade?
**Roll-up** the data along the desired dimensions

```
SELECT year, marst, sum(people)
FROM census
GROUP BY year, marst;
```
Roll-Up and Drill-Down

Need more detailed information?
**Drill-down** into additional dimensions

```
SELECT year, age, marst, sum(people)  
FROM census  
GROUP BY year, age, marst;
```
A two-dimensional projection.

Sum along Marital Status

Sum along Age

Sum along Year

All Marital Status

All Ages

All Years
<table>
<thead>
<tr>
<th>YEAR</th>
<th>AGE</th>
<th>MARST</th>
<th>SEX</th>
<th>PEOPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1,483,789</td>
</tr>
<tr>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1,411,067</td>
</tr>
<tr>
<td>1860</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2,120,846</td>
</tr>
<tr>
<td>1860</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1,804,467</td>
</tr>
</tbody>
</table>

Which format might we prefer?
Administrivia
A1: Visualization Design

Design a static visualization for a data set.

After the World War II, antibiotics were considered “wonder drugs.” To learn which drug is most effective for which bacterial infection, performance of the three most popular antibiotics were gathered.

You must choose the message you want to convey. What task do you want to support? What insight do you want to communicate?
A1: Visualization Design

Design a **static visualization** for the data set. You are free to **use any tools** (inc. pen & paper).

**Deliverables** (upload via Canvas; see A1 page)
- Image of your visualization (PNG or JPG format)
- Short description + design rationale (≤ 4 paragraphs)

Due by **5:00 pm, Monday April 6**.
Next Tuesday: Design Exercise

We will review A1 submissions
So be sure to turn yours in on time!

We will then have a redesign exercise
Please bring paper, pens, etc for sketching

Prof. Heer will be out attending OpenVisConf
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
Sémiologie Graphique, 1967
Bertin’s Semiology of Graphics

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

∴ Encode quantitative variables

“Resemblance, order and proportion are the three signfields in graphics.” - Bertin
Visual Encoding Variables

- Position (x 2)
- Size
- Value
- Texture
- Color
- Orientation
- Shape
Visual Encoding Variables

Position
Length
Area
Volume
Value
Texture
Color
Orientation
Shape
Transparency
Blur / Focus …
Information in Hue 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
Bertin’s “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>N</td>
<td></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>

Nominal
Ordinal
Quantitative
Note: Q ⊂ O ⊂ N
Deconstructions
Exports and Imports to and from Denmark & Norway from 1700 to 1780.
William Playfair, 1786

X-axis: year (Q)
Y-axis: currency (Q)
Color: imports/exports (N, O)
Wattenberg’s Map of the Market

Rectangle Area: market cap (Q)
Rectangle Position: market sector (N), market cap (Q)
Color Hue: loss vs. gain (N, O)
Color Value: magnitude of loss or gain (Q)
Minard 1869: Napoleon’s March
Single-Axis Composition
Mark Composition

Y-axis: temperature (Q)

X-axis: longitude (Q) / time (O)

Temp over space/time (Q x Q)
Mark Composition

Y-axis: longitude (Q)

+ X-axis: latitude (Q)

+ Width: army size (Q)

= Army position (Q x Q) and army size (Q)
Depicts at least 5 quantitative variables. Any others?
Formalizing Design
Choosing Visual Encodings

Assume $k$ visual encodings and $n$ data attributes. We would like to pick the “best” encoding among a combinatorial set of possibilities of size $(n+1)^k$.

Principle of Consistency
The properties of the image (visual variables) should match the properties of the data.

Principle of Importance Ordering
Encode the most important information in the most effective way.
Design Criteria [Mackinlay 86]

Expressiveness
A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness
A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.
Design Criteria  [Mackinlay 86]

Expressiveness
A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness
A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.
A multivariate relation may be *inexpressive* in a single horizontal dot plot because multiple records are mapped to the same position.

Can not express the facts

A multivariate relation may be *inexpressive* in a single horizontal dot plot because multiple records are mapped to the same position.
Expresses facts not in the data

A length is interpreted as a quantitative value.

Fig. 11. Incorrect use of a bar chart for the Nation relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the Nation relation.
Design Criteria  [Mackinlay 86]

Expressiveness
A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness
A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.
Design Criteria [Mackinlay 86]

Expressiveness
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness
A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.
Design Criteria  [Tversky 02]

Congruence
The structure and content of the external representation should correspond to the desired structure and content of the internal representation.

Apprehension
The structure and content of the external representation should be readily and accurately perceived and comprehended.
Design Criteria *Translated*

Tell the truth and nothing but the truth
(don’t lie, and don’t lie by omission)

Use encodings that people decode better
(where better = faster and/or more accurate)
Mackinlay’s Ranking

Conjectured effectiveness of encodings by data type
Mackinlay’s Design Algorithm

APT - “A Presentation Tool”, 1986

User formally specifies data model and type
Input: ordered list of data variables to show

APT searches over design space
Test expressiveness of each visual encoding
Generate encodings that pass test
Rank by perceptual effectiveness criteria

Output the “most effective” visualization
APT

Automatically generate chart for car data

Input variables:
1. Price
2. Mileage
3. Repair
4. Weight
Limitations of APT?
Limitations of APT

Does not cover many visualization techniques
Networks, hierarchies, maps, diagrams
Also: 3D structure, animation, illustration, …

Does not consider interaction

Does not consider semantics / conventions

Assumes single visualization as output
Summary: Data & Image Models

Formal specification
Data model: relational data; N,O,Q types
Image model: visual encoding channels
Encodings map data to visual variables

Choose expressive and effective encodings
Rule-based tests of expressiveness
Perceptual effectiveness rankings

Question: how do we establish effectiveness criteria? Subject of perception lectures...
A1: Visualization Design

Design a **static visualization** for the data set. You are free to **use any tools** (inc. pen & paper).

**Deliverables** (upload via Canvas; see A1 page)
- Image of your visualization (PNG or JPG format)
- Short description + design rationale (≤ 4 paragraphs)

Due by **5:00 pm, Monday April 6**.