Last Time:
Value of Visualization
The Value of Visualization

**Record** information
- Blueprints, photographs, seismographs, ...

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

**Communicate** information to others
- Share and persuade
- Collaborate and revise
Marey’s sphygmograph [from Braun 83]
Make a decision: Challenger

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
Info-Vis vs. Sci-Vis?
Visualization Reference Model

Data

Raw Data → Data Tables → Visual Structures → Views → Visual Form

Data Transformations

Visual Encodings

View Transformations
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

mapping
visual encoding
visual metaphor

image
visual channel
retinal variables
Topics

Properties of data
Properties of images
Mapping data to images
Data
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
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 [Shneiderman 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
Nominal, Ordinal and Quantitative

N - Nominal (labels)
  · 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

S. S. Stevens, On the theory of scales of measurements, 1946
From data model to N,O,Q data type

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

Conceptual model
- Temperature (°C)

Data type
- Burned vs. Not burned (N)
- Hot, warm, cold (O)
- Continuous range of values (Q)
Sepal and petal lengths and widths for three species of iris [Fisher 1936].
<table>
<thead>
<tr>
<th>ID</th>
<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>1</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>I. Virginica</td>
<td>Petal</td>
<td>24</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>13</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1</td>
<td>I. Virginica</td>
<td>Sepal</td>
<td>31</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>28</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>2</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>2</td>
<td>I. Virginica</td>
<td>Petal</td>
<td>23</td>
<td>51</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>1</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>36</td>
<td>46</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>1</td>
<td>I. Virginica</td>
<td>Sepal</td>
<td>31</td>
<td>69</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>2</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>33</td>
<td>63</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>2</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>3</td>
<td>I. Virginica</td>
<td>Petal</td>
<td>20</td>
<td>52</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>3</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>3</td>
<td>I. Virginica</td>
<td>Sepal</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>3</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>4</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>4</td>
<td>I. Virginica</td>
<td>Petal</td>
<td>19</td>
<td>51</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>4</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>36</td>
<td>49</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>4</td>
<td>I. Virginica</td>
<td>Sepal</td>
<td>27</td>
<td>58</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>4</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>26</td>
<td>58</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>5</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>5</td>
<td>I. Virginica</td>
<td>Petal</td>
<td>17</td>
<td>45</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>5</td>
<td>I. Versicolor</td>
<td>Petal</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>5</td>
<td>I. Setosa</td>
<td>Sepal</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>5</td>
<td>I. Virginica</td>
<td>Sepal</td>
<td>25</td>
<td>49</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>5</td>
<td>I. Versicolor</td>
<td>Sepal</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>6</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>6</td>
<td>I. Setosa</td>
<td>Petal</td>
<td>2</td>
<td>16</td>
</tr>
</tbody>
</table>
Relational data model

Represent data as a **table** *(relation)*
Each **row** *(tuple)* represents a single record
  Each record is a fixed-length tuple
Each **column** *(attribute)* represents a single **variable**
  Each attribute has a **name** and a **data type**
A table’s **schema** is the set of names and data types

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

- Data transformations (sql)
- Projection (select)
- Selection (where)
- Sorting (order by)
- Aggregation (group by, sum, min, ...)
- Set operations (union, ...)
- Combine (inner join, outer 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

**Dimensions:** Discrete variables describing data
Dates, categories of values (independent vars)

**Measures:** Data values that can be aggregated
Numbers to be analyzed (dependent vars)
Aggregate as sum, count, average, std. deviation
Example: U.S. Census Data

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

<table>
<thead>
<tr>
<th>People</th>
<th>Year</th>
<th>Age</th>
<th>Sex</th>
<th>Marital Status</th>
<th>2348 data points</th>
</tr>
</thead>
</table>

<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>1483789</td>
</tr>
<tr>
<td>1850</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1450376</td>
</tr>
<tr>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1411067</td>
</tr>
<tr>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1359668</td>
</tr>
<tr>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1260099</td>
</tr>
<tr>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>1216114</td>
</tr>
<tr>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>1077133</td>
</tr>
<tr>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>2</td>
<td>1110619</td>
</tr>
<tr>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>1017281</td>
</tr>
<tr>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>1003841</td>
</tr>
<tr>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>862547</td>
</tr>
<tr>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>2</td>
<td>799482</td>
</tr>
<tr>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>730638</td>
</tr>
<tr>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>639656</td>
</tr>
<tr>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>1</td>
<td>588487</td>
</tr>
<tr>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>2</td>
<td>505012</td>
</tr>
<tr>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>1</td>
<td>475911</td>
</tr>
<tr>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>2</td>
<td>428185</td>
</tr>
<tr>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>1</td>
<td>384211</td>
</tr>
<tr>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>2</td>
<td>341254</td>
</tr>
<tr>
<td>1850</td>
<td>50</td>
<td>0</td>
<td>1</td>
<td>321343</td>
</tr>
<tr>
<td>1850</td>
<td>50</td>
<td>0</td>
<td>2</td>
<td>286580</td>
</tr>
<tr>
<td>1850</td>
<td>55</td>
<td>0</td>
<td>1</td>
<td>194080</td>
</tr>
<tr>
<td>1850</td>
<td>55</td>
<td>0</td>
<td>2</td>
<td>187208</td>
</tr>
<tr>
<td>1850</td>
<td>60</td>
<td>0</td>
<td>1</td>
<td>174976</td>
</tr>
<tr>
<td>1850</td>
<td>60</td>
<td>0</td>
<td>2</td>
<td>162236</td>
</tr>
<tr>
<td>1850</td>
<td>65</td>
<td>0</td>
<td>1</td>
<td>106827</td>
</tr>
<tr>
<td>1850</td>
<td>65</td>
<td>0</td>
<td>2</td>
<td>105534</td>
</tr>
<tr>
<td>1850</td>
<td>70</td>
<td>0</td>
<td>1</td>
<td>73677</td>
</tr>
<tr>
<td>1850</td>
<td>70</td>
<td>0</td>
<td>2</td>
<td>71762</td>
</tr>
<tr>
<td>1850</td>
<td>75</td>
<td>0</td>
<td>1</td>
<td>40834</td>
</tr>
<tr>
<td>1850</td>
<td>75</td>
<td>0</td>
<td>2</td>
<td>40229</td>
</tr>
<tr>
<td>1850</td>
<td>80</td>
<td>0</td>
<td>1</td>
<td>23449</td>
</tr>
<tr>
<td>1850</td>
<td>80</td>
<td>0</td>
<td>2</td>
<td>22949</td>
</tr>
<tr>
<td>1850</td>
<td>85</td>
<td>0</td>
<td>1</td>
<td>8186</td>
</tr>
<tr>
<td>1850</td>
<td>85</td>
<td>0</td>
<td>2</td>
<td>10511</td>
</tr>
<tr>
<td>1850</td>
<td>90</td>
<td>0</td>
<td>1</td>
<td>5259</td>
</tr>
<tr>
<td>1850</td>
<td>90</td>
<td>0</td>
<td>2</td>
<td>6569</td>
</tr>
<tr>
<td>1860</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2120846</td>
</tr>
<tr>
<td>1860</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2092162</td>
</tr>
</tbody>
</table>
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 (M/F)</td>
<td>N</td>
</tr>
<tr>
<td>Marital Status</td>
<td>N</td>
</tr>
</tbody>
</table>
Census: Dimension or Measure?

<table>
<thead>
<tr>
<th>People Count</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Dimension</td>
</tr>
<tr>
<td>Age</td>
<td>Depends!</td>
</tr>
<tr>
<td>Sex (M/F)</td>
<td>Dimension</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Dimension</td>
</tr>
</tbody>
</table>
Roll-Up and Drill-Down

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

```sql
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

```sql
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 Years

All Ages

All Marital Status

Sum along Age

A one dimensional projection.
A two dimensional projection.

Roll-Up

Drill-Down

Sum along Marital Status

Sum along Age

Sum along Year

All Marital Status

All Ages

All Years

Sum along Age
<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>

... 

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

... 

Which format might we prefer?
Row vs. Column-Oriented Databases
Relational Data Organizations

<table>
<thead>
<tr>
<th>Transactions</th>
<th>vs.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row-oriented</td>
<td></td>
<td>Column-oriented</td>
</tr>
</tbody>
</table>

[Diagram showing comparison between row-oriented and column-oriented data organizations for transactions and analysis.]
Relational Data Organizations

Row-oriented

Column-oriented
Relational Data Organizations

Speed-up Analysis
Reduce data transfer
Improved locality
Data compression

Column-oriented
Administrivia
Announcements

Auditors

• Requirements: Come to class and participate (online as well)

Class participation requirements

• Complete readings before class
• In-class discussion
• Post at least 1 substantive comment/question on Piazza within 24 hours after each lecture (11am next day)
Assignment 1: Visualization Design

Design a static visualization for a given data set.

Deliverables (submit via Catalyst)

- Image of your visualization
- Short description and design rationale (≤ 4 para.)

Due by 5:00pm on Monday 1/13.
Questions?
Visual language is a sign system

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

Sémiologie Graphique, 1967

Jacques Bertin
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 signifieds in graphics." - Bertin
<table>
<thead>
<tr>
<th>Les variables de l'image</th>
</tr>
</thead>
<tbody>
<tr>
<td>XY 2 dimensions</td>
</tr>
<tr>
<td>du plan</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>Taille</td>
</tr>
<tr>
<td>Valeur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Points</th>
<th>Lignes</th>
<th>Zones</th>
</tr>
</thead>
<tbody>
<tr>
<td>🍀</td>
<td>🍀</td>
<td>🍀</td>
</tr>
<tr>
<td>🍀</td>
<td>🍀</td>
<td>🍀</td>
</tr>
<tr>
<td>🍀</td>
<td>🍀</td>
<td>🍀</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Les variables de séparation des images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain</td>
</tr>
<tr>
<td>Couleur</td>
</tr>
<tr>
<td>Orientation</td>
</tr>
<tr>
<td>Forme</td>
</tr>
</tbody>
</table>
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 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
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></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>

Note: Q < O < N
Design Space of Visual Encodings
Univariate data

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

factors

variable
Univariate data

Tukey box plot

factors

A  B  C

variable

1

Middle 50%

Mean

low

high

0

20
Bivariate data

Scatter plot is common
Trivariate data

3D scatter plot is possible
Three variables

Two variables \([x,y]\) can map to points
  - Scatterplots, maps, ...

Third variable \([z]\) must use
  - Color, size, shape, ...

\[ 1 \quad \rightarrow \quad 2 \]

\[ Y \quad \uparrow \quad Z \quad \leftarrow \quad X \]
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>
Multidimensional data

How many variables can be depicted in an image?

“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
Exports and Imports to and from Denmark & Norway from 1700 to 1780.
Playfair 1786

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

[based on slide from Mackinlay]
Mark composition

y-axis: temperature (Q)

+ x-axis: longitude (Q) / time (O)

= temp over space/time (Q x Q)

[based on slide from Mackinlay]
Mark composition

- y-axis: longitude (Q)
- x-axis: latitude (Q)
- width: army size (Q)

 army position (Q x Q) and army size (Q)

[based on slide from Mackinlay]
longitude (Q)

latitude (Q)

army size (Q)

temperature (Q)

latitude (Q) / time (O)

[based on slide from Mackinlay]
Minard 1869: Napoleon’s march

Depicts at least 5 quantitative variables. Any others?
Formalizing Design
(Mackinlay 1986)
Choosing Visual Encodings

Challenge:
Assume 8 visual encodings and \( n \) data attributes. We would like to pick the “best” encoding among a combinatorial set of possibilities with size \((n+1)^8\).

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.
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.
Cannot express the facts

A one-to-many (1 → N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position.
Expresses facts not in the data

A length is interpreted as a quantitative value;
∴ Length of bar says something untrue about N data

[Mackinlay, APT, 1986]
Design Criteria (Mackinlay)

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.
Mackinlay’s Ranking

Conjectured effectiveness of the encoding
Mackinlay’s Design Algorithm

User formally specifies data model and type
• Additional input: ordered list of data variables to show

APT searches over design space
• Tests expressiveness of each visual encoding
• Generates specification for encodings that pass test
• Tests perceptual effectiveness of resulting image

Outputs the “most effective” visualization
Limitations

Does not cover many visualization techniques

- Bertin and others discuss networks, maps, diagrams
- Does not consider 3D, animation, illustration, photography, ...

Does not model interaction

Does not consider semantic data types / conventions
Summary

Formal specification
- Data model
- Image model
- Encodings mapping data to image

Choose expressive and effective encodings
- Formal test of expressiveness
- Experimental tests of perceptual effectiveness
Assignment 1: Visualization Design

Design a static visualization for a given data set.

Deliverables (submit via Catalyst)
• Image of your visualization
• Short description and design rationale ($\leq 4$ para.)

Due by 5:00pm on Monday 1/13.