6.859: Interactive Data Visualization

Data & Image Models

Arvind Satyanarayan

Have some paper + pens/pencils handy!
<table>
<thead>
<tr>
<th>Course Grading</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class Participation</strong></td>
<td>5%</td>
</tr>
<tr>
<td>Reading Commentaries</td>
<td>5%</td>
</tr>
<tr>
<td>Ao: Sketching Visualizations</td>
<td>2%</td>
</tr>
<tr>
<td>A1: Visualization Design</td>
<td>3%</td>
</tr>
<tr>
<td>A2: Exploratory Data Analysis</td>
<td>10%</td>
</tr>
<tr>
<td>A3: White/Black Hat Visualization</td>
<td>15%</td>
</tr>
<tr>
<td>A4: Interactive Narratives</td>
<td>20%</td>
</tr>
<tr>
<td>Final Project</td>
<td>40%</td>
</tr>
</tbody>
</table>

Lectures will be recorded and posted to Canvas.

You may attend asynchronously but we encourage synchronous attendance if you're able to.

Class Participation grade will be primarily determined by activity on Slack:

- Introduce yourself in #introductions
- Ask and answer questions
- Post links to + critique interesting visualizations you find online.

Share your work!!
Course Grading

Class Participation  5%
Reading Commentaries  5%
A0: Sketching Visualizations  2%  Due 2/22
A1: Visualization Design  3%  Due 3/1
A2: Exploratory Data Analysis  10%  Due 3/9
A3: White/Black Hat Visualization  15%  Due 3/24
A4: Interactive Narratives  20%  Due 4/5, 4/12
Final Project  40%
  Proposal  Due 4/16
  MVP + Presentations  Due 5/3
  Poster Session + Final Deliverables  Due 5/11
Activity!

In **1 minute**, sketch as many visualizations as possible of these two numbers:

75  37
Most Likely Results

Pie Charts

Bar Charts

Scatterplot

Arabic Numbers

75
37

75
37
Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

To overcome fixation:

**sketch**: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas [Buxton 2007]

**consult examples**
This creature is very friendly and has a retractable neck which helps it to eat off of trees.

This creature walks the planet eating all sorts of things like rocks and dirt.

A very fluffy creature that hops from one place to the next using its very strong legs.
Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

To overcome fixation:

**sketch**: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas [Buxton 2007]

**consult examples**: early and repeated exposure to examples improves creativity [Kulkarni 2012]
mVVVo
Examples from Jon Schwabish.
Examples from Jon Schwabish.
Examples from Jon Schwabish.
Examples from Jon Schwabish.
Examples from Jon Schwabish.

1937 Plymouth

1975 Plymouth
Design Fixation

“A blind adherence to a set of ideas or concepts limiting the output of conceptual design” [Jansson & Smith 1991]

To overcome fixation:

**sketch**: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas [Buxton 2007]

**consult examples**: early and repeated exposure to examples improves creativity [Kulkarni 2012]
Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

To overcome fixation:

**sketch**: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas [Buxton 2007]

**consult examples**: early and repeated exposure to examples improves creativity [Kulkarni 2012]

**introduce a constraint**: impose new structures to the problem to spur creativity [Stokes 2006]
Activity!

Share your work: www.yellkey.com/three

In **3 minutes**, sketch as many **new visualizations** as possible that are different from your previous ideas. If you're stuck, introduce a constraint -- e.g., one line, only black/white, only round objects, etc.

75 37
6.894: Interactive Data Visualization

Data & Image Models

Arvind Satyanarayan
Data Visualization
Data

**Physical** Data Types
int, float, string

**Conceptual** Data Types
temperature, location

Mapping or Visual Encoding

Visual

**Visual Channels**
x, y, color, opacity

**Graphical Marks**
rect, line, point, area
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.

[Mackinlay 1986]
Expressiveness

Cannot express the facts

A multivariate dataset may be \textit{inexpressive} in a single horizontal dot plot because multiple records are mapped to the same position.
What's wrong with this visualization?
Expressiveness

Express facts not in the data

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.
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.

[Mackinlay 1986]
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*.

[Mackinlay 1986]

Data models give us a way of talking about this.
Data Models
Conceptual Models vs Data Models

By "default", data is described in terms of a specific domain.  
E.g., The average amount of rain or snow in different towns, cities, countries.  
E.g., friends, followers, connections depending on the social network (or citations in academia!).

To effectively map data to visuals, we need a level of abstraction.  
Data abstraction allows us to consistently encode the same "types" of data, even if different domains use different terminology to describe it.
1. Tabular


https://gender-pay-gap.service.gov.uk
Dataset Types

1. **Tabular**
   A collection of records with named attributes.

2. **Networks**
   Nodes and links can also have attributes (e.g., size of nodes, thickness/directionality of links).
   Trees are special networks where each node has only one parent.

## Dataset Types

1. **Tabular**
   A collection of records with named attributes.

2. **Networks**
   Nodes and links can also have attributes (e.g., size of nodes, thickness/directionality of links).
   
   Trees are special networks where each node has only one parent.

3. **Spatial**
   Continuous "fields" vs discrete "positions"

---


**Attribute Types**

**Dimensions**

~ Independent variables.

Ways of describing the data, often discrete.

E.g., categories, dates, binned quantities.

Can include numerical data, but doesn't make sense to aggregate.

**Measures**

~ Dependent variables (i.e., their value is a function of one or more dimensions).

Numerical data that can be analyzed and aggregated.

Aggregations including sum, count, avg, std. dev, etc.
### Health Indicators

<table>
<thead>
<tr>
<th>Country</th>
<th>Birth Rate</th>
<th>Infant Mortality</th>
<th>Health Exp % GDP</th>
<th>Life Expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>3.8%</td>
<td>0.07</td>
<td>4.4%</td>
<td>58</td>
</tr>
<tr>
<td>Malawi</td>
<td>4.2%</td>
<td>0.07</td>
<td>7.4%</td>
<td>50</td>
</tr>
<tr>
<td>Eritrea</td>
<td>3.9%</td>
<td>0.05</td>
<td>3.9%</td>
<td>48</td>
</tr>
<tr>
<td>Congo, Dem. Rep.</td>
<td>4.6%</td>
<td>0.06</td>
<td>5.8%</td>
<td>48</td>
</tr>
<tr>
<td>Niger</td>
<td>5.1%</td>
<td>0.08</td>
<td>7.2%</td>
<td>48</td>
</tr>
<tr>
<td>Madagascar</td>
<td>3.7%</td>
<td>0.05</td>
<td>4.7%</td>
<td>62</td>
</tr>
<tr>
<td>Mozambique</td>
<td>4.3%</td>
<td>0.05</td>
<td>6.1%</td>
<td>48</td>
</tr>
<tr>
<td>Rwanda</td>
<td>3.8%</td>
<td>0.07</td>
<td>8.0%</td>
<td>54</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.6%</td>
<td>0.07</td>
<td>8.4%</td>
<td>54</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>4.3%</td>
<td>0.08</td>
<td>9.2%</td>
<td>50</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>4.1%</td>
<td>0.13</td>
<td>15.0%</td>
<td>42</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>3.7%</td>
<td>0.11</td>
<td>4.1%</td>
<td>46</td>
</tr>
<tr>
<td>Guinea</td>
<td>4.0%</td>
<td>0.08</td>
<td>5.8%</td>
<td>53</td>
</tr>
<tr>
<td>Tanzania</td>
<td>4.1%</td>
<td>0.08</td>
<td>5.1%</td>
<td>53</td>
</tr>
<tr>
<td>Nepal</td>
<td>2.7%</td>
<td>0.07</td>
<td>5.7%</td>
<td>65</td>
</tr>
<tr>
<td>Togo</td>
<td>3.8%</td>
<td>0.07</td>
<td>6.5%</td>
<td>55</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>4.0%</td>
<td>0.09</td>
<td>5.9%</td>
<td>55</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>4.4%</td>
<td>0.08</td>
<td>5.8%</td>
<td>53</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>3.5%</td>
<td>0.07</td>
<td>5.7%</td>
<td>65</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2.3%</td>
<td>0.05</td>
<td>3.3%</td>
<td>60</td>
</tr>
<tr>
<td>Mali</td>
<td>4.8%</td>
<td>0.10</td>
<td>6.5%</td>
<td>52</td>
</tr>
<tr>
<td>Gambia, The</td>
<td>4.4%</td>
<td>0.06</td>
<td>4.4%</td>
<td>57</td>
</tr>
<tr>
<td>Haiti</td>
<td>2.8%</td>
<td>0.07</td>
<td>4.4%</td>
<td>57</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>3.3%</td>
<td>0.05</td>
<td>4.5%</td>
<td>49</td>
</tr>
<tr>
<td>Benin</td>
<td>4.0%</td>
<td>0.07</td>
<td>4.5%</td>
<td>49</td>
</tr>
<tr>
<td>Cambodia</td>
<td>2.0%</td>
<td>0.05</td>
<td>6.1%</td>
<td>68</td>
</tr>
<tr>
<td>Kenya</td>
<td>3.8%</td>
<td>0.07</td>
<td>6.1%</td>
<td>68</td>
</tr>
<tr>
<td>Chad</td>
<td>4.9%</td>
<td>0.08</td>
<td>4.0%</td>
<td>49</td>
</tr>
<tr>
<td>Timor-Leste</td>
<td>3.8%</td>
<td>0.06</td>
<td>6.9%</td>
<td>64</td>
</tr>
<tr>
<td>Kyrgyz Republic</td>
<td>2.3%</td>
<td>0.03</td>
<td>6.0%</td>
<td>69</td>
</tr>
<tr>
<td>Comoros</td>
<td>3.0%</td>
<td>0.06</td>
<td>4.0%</td>
<td>57</td>
</tr>
</tbody>
</table>

**Measures**

- Birth Rate
- Infant Mortality Rate
- Health Exp % GDP
- Life Expectancy
Attribute Types
## Attribute Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td>Labels or categories. E.g., Fruits: apples, bananas, cantaloupes, ...</td>
</tr>
<tr>
<td><strong>Ordinal</strong></td>
<td>Ordered. E.g., Quality of meat: Grade A, AA, AAA</td>
</tr>
<tr>
<td><strong>Quantitative (Interval)</strong></td>
<td>Interval (zero can be arbitrarily located). E.g., Dates: Jan 1, 2018. Location: (Lat 42.36°, -71.09°). Only differences can be calculated (e.g., distances or spans).</td>
</tr>
<tr>
<td><strong>Quantitative (Ratio)</strong></td>
<td>Ratio (fixed zero). E.g., Physical measurement: length, mass, temperature. Counts and amounts. Can measure ratios or proportions.</td>
</tr>
</tbody>
</table>
Attribute Types

Nominal

Labels or categories.
*E.g.*, Fruits: apples, bananas, cantaloupe, ...

Ordinal

Ordered.
*E.g.*, Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).
*E.g.*, Dates of year: 2018 location (Lat 42.36, -71.09).
Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

Ratio (fixed zero).
*E.g.*, Physical measurements: length, mass, temperature.
Counts and amounts. Can measure ratios or proportions.
### Attribute Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td>Labels or categories. E.g., Fruits: apples, bananas, cantaloupes, ...</td>
</tr>
<tr>
<td><strong>Ordinal</strong></td>
<td>Ordered. E.g., Quality of meat: Grade A, AA, AAA</td>
</tr>
<tr>
<td><strong>Quantitative (Interval)</strong></td>
<td>Interval (zero can be arbitrarily located). E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09) Only differences can be calculated (e.g., distances or spans).</td>
</tr>
<tr>
<td><strong>Quantitative (Ratio)</strong></td>
<td>Ratios meaning Counts and amounts. Can measure ratios or proportions. For Physical measurement: length, mass, temperature.</td>
</tr>
<tr>
<td>Attribute Types</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Nominal</strong></td>
<td>Labels or categories. E.g., Fruits: apples, bananas, cantaloupes, ...</td>
</tr>
<tr>
<td><strong>Ordinal</strong></td>
<td>Ordered. E.g., Quality of meat: Grade A, AA, AAA</td>
</tr>
<tr>
<td><strong>Quantitative (Interval)</strong></td>
<td>Interval (zero can be arbitrarily located). E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09) Only differences can be calculated (e.g., distances or spans).</td>
</tr>
<tr>
<td><strong>Quantitative (Ratio)</strong></td>
<td>Ratio (fixed zero). E.g., Physical measurement: length, mass, temperature Counts and amounts. Can measure ratios or proportions.</td>
</tr>
</tbody>
</table>
### Attribute Types

- **Nominal**
  - Symbols: $=, \neq$
  - Labels or categories.
  - *E.g.*, Fruits: apples, bananas, cantaloupes, ...

- **Ordinal**
  - Symbols: $=, \neq, <, >$
  - Ordered.
  - *E.g.*, Quality of meat: Grade A, AA, AAA

- **Quantitative**
  - **(Interval)**
    - Symbols: $=, \neq, <, >, -$.
    - Interval (zero can be arbitrarily located).
    - *E.g.*, Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)
    - Only differences can be calculated (e.g., distances or spans).

- **Quantitative**
  - **(Ratio)**
    - Symbols: $=, \neq, <, >, -, \%$
    - Ratio (fixed zero).
    - *E.g.*, Physical measurement: length, mass, temperature
    - Counts and amounts. Can measure ratios or proportions.
<table>
<thead>
<tr>
<th>Physical Model</th>
<th>Conceptual Model</th>
<th>Attribute Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.5, 54.0, −17.3, ...</td>
<td>Temperature (°C)</td>
<td>Burned vs. Not-Burned (N)</td>
</tr>
<tr>
<td>Floating point numbers</td>
<td></td>
<td>Hot, Warm, Cold (O)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Temperature Value (Q)</td>
</tr>
</tbody>
</table>
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

**People Count**: # of people in group

**Year**: 1850 – 2000 (every decade)

**Age**: 0 – 90+

**Sex**: Male, Female

**Marital Status**: Single, Married, Divorced, ...
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count:
Year:
Age:
Sex:
Marital Status:
### Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

- **People Count:** Q-Ratio
- **Year:**
- **Age:**
- **Sex:**
- **Marital Status:**

<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>0</td>
<td>1483789</td>
</tr>
<tr>
<td>1850</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1450370</td>
</tr>
<tr>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1411067</td>
</tr>
<tr>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1359668</td>
</tr>
<tr>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1260099</td>
</tr>
<tr>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1216114</td>
</tr>
<tr>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1077133</td>
</tr>
<tr>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1110619</td>
</tr>
<tr>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1017281</td>
</tr>
<tr>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1003841</td>
</tr>
<tr>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>862547</td>
</tr>
<tr>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>799482</td>
</tr>
<tr>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>730638</td>
</tr>
<tr>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>639636</td>
</tr>
<tr>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>588487</td>
</tr>
<tr>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>505012</td>
</tr>
<tr>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>475911</td>
</tr>
<tr>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>428185</td>
</tr>
<tr>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>384211</td>
</tr>
<tr>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>341254</td>
</tr>
<tr>
<td>1850</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>321334</td>
</tr>
</tbody>
</table>
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count: Q-Ratio

Year:

Age:

Sex: Nominal

Marital Status:
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count: Q-Ratio

Year:

Age:

Sex: Nominal

Marital Status: Nominal
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

**People Count:** Q-Ratio

**Year:** Ordinal or Q-Interval

**Age:**

**Sex:** Nominal

**Marital Status:** Nominal
**Activity: U.S. Census**

What are the types of these attributes (N/O/Q and dimension/measure)?

**People Count**: Q-Ratio

**Year**: Ordinal or Q-Interval

**Age**: Ordinal or Q-Interval

**Sex**: Nominal

**Marital Status**: Nominal

<table>
<thead>
<tr>
<th></th>
<th>year</th>
<th>age</th>
<th>marst</th>
<th>sex</th>
<th>people</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1850</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1483789</td>
</tr>
<tr>
<td>3</td>
<td>1850</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1450376</td>
</tr>
<tr>
<td>4</td>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1411067</td>
</tr>
<tr>
<td>5</td>
<td>1850</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1359668</td>
</tr>
<tr>
<td>6</td>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1260099</td>
</tr>
<tr>
<td>7</td>
<td>1850</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>1216114</td>
</tr>
<tr>
<td>8</td>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1077135</td>
</tr>
<tr>
<td>9</td>
<td>1850</td>
<td>15</td>
<td>0</td>
<td>2</td>
<td>1110619</td>
</tr>
<tr>
<td>10</td>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1017281</td>
</tr>
<tr>
<td>11</td>
<td>1850</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>1003841</td>
</tr>
<tr>
<td>12</td>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>862547</td>
</tr>
<tr>
<td>13</td>
<td>1850</td>
<td>25</td>
<td>0</td>
<td>2</td>
<td>799482</td>
</tr>
<tr>
<td>14</td>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>730638</td>
</tr>
<tr>
<td>15</td>
<td>1850</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>639636</td>
</tr>
<tr>
<td>16</td>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>588487</td>
</tr>
<tr>
<td>17</td>
<td>1850</td>
<td>35</td>
<td>0</td>
<td>2</td>
<td>505012</td>
</tr>
<tr>
<td>18</td>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>475911</td>
</tr>
<tr>
<td>19</td>
<td>1850</td>
<td>40</td>
<td>0</td>
<td>2</td>
<td>428185</td>
</tr>
<tr>
<td>20</td>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>384211</td>
</tr>
<tr>
<td>21</td>
<td>1850</td>
<td>45</td>
<td>0</td>
<td>2</td>
<td>341254</td>
</tr>
<tr>
<td>22</td>
<td>1850</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>321344</td>
</tr>
</tbody>
</table>
### Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

<table>
<thead>
<tr>
<th>People Count:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year:</td>
</tr>
<tr>
<td>Age:</td>
</tr>
<tr>
<td>Sex:</td>
</tr>
<tr>
<td>Marital Status:</td>
</tr>
</tbody>
</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>639636</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>321334</td>
</tr>
</tbody>
</table>
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count: Measure

Year:

Age:

Sex:

Marital Status:
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

**People Count:** Measure

**Year:** Dimension

**Age:**

**Sex:**

**Marital Status:**
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

**People Count**: Measure

**Year**: Dimension

**Age**: Dimensions

**Sex**: Dimensions

**Marital Status**: Dimension
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count: Measure

Year: Dimension

Age:

Sex: Dimension

Marital Status: Dimension
Activity: U.S. Census

What are the types of these attributes (N/O/Q and dimension/measure)?

People Count: Measure

Year: Dimension

Age: Depends!

Sex: Dimension

Marital Status: Dimension
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*.

[MacKinlay 1986]
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.

[McKainlay 1986]

Data models give us a way of talking about this.

Questions?

Raise your hand

Post in the chat
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*.

[McCallinlay 1986]
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 if the information it conveys is more readily perceived than the information in the other visualization.

[MaCKinlay 1986]
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 if the information it conveys *is more readily perceived* than the information in the other visualization.

[Data models give us a way of talking about this.]

[Image models give us a way of talking about this.]

[Mackinlay 1986]
Image Models
The Semiology of Graphics (1967)

Jacques Bertin (1918 – 2010)
French cartographer
The **Semiology** of Graphics (1967)

*Study of signs and how cultures use them.*

Jacques Bertin (1918 – 2010)
French cartographer
The **Semiology** of Graphics (1967)

Study of **signs** and how cultures use them.

Anything that stands for something other than itself.

Images are perceived as a set of signs. Sender encodes information in signs. Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?

Jacques Bertin (1918 – 2010)
French cartographer
What do these signs signify?

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

Sender encodes information in signs. Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?

"Resemblance, order, and proportional are the three signfields in graphics."

–Bertin
Visual Variables

Also called visual *channels*. Used to encode data values as characteristics of marks.

*From 1967, so Bertin only accounted for visualizations that were printable, white paper.*
**Channels**: Expressiveness Types and Effectiveness Ranks

- **Magnitude Channels**
  - Position on common scale
  - Position on unaligned scale
  - Length (1D size)
  - Tilt/angle
  - Area (2D size)
  - Depth (3D position)
  - Color luminance
  - Color saturation
  - Curvature
  - Volume (3D size)

- **Identity Channels**
  - Spatial region
  - Color hue
  - Motion
  - Shape

---

**Magnitude Channels**: O or Q attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels**: N attributes

- Spatial region
- Color hue
- Motion
- Shape

---

Channels: Expressiveness Types and Effectiveness Ranks

**Magnitude Channels:** O or Q attributes
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels:** N attributes
- Spatial region
- Color hue
- Motion
- Shape

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 if the information it conveys is *more readily perceived* than the information in the other visualization.

Data models give us a way of talking about this.

Image models give us a way of talking about this.

[Mackinlay 1986]