CMPSC382
Lecture 31: Introduction to Visualization

Prof. John Wenskovitch
11/11/2015
Last Time

• Choosing colors is difficult
  – Preattentiveness can be used to draw the eye so certain features immediately, without brain processing time involved.
  – Colors can be chosen in complementary or opposing schemes to encode information.
  – Change blindness makes it difficult for humans to detect differences in similar scenes.
What is Visualization?

• Definition #1: A technique for turning data into information.
  – **Data** – Individual and independent numbers, characters, and/or symbols
  – **Information** – Knowledge derived from data
## Anscombe’s Quartet

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th></th>
<th>2</th>
<th></th>
<th>3</th>
<th></th>
<th>4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
</tr>
<tr>
<td></td>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>7.91</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
<td>6.89</td>
</tr>
<tr>
<td>Mean</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Variance</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.816</td>
<td>0.816</td>
<td>0.816</td>
<td>0.816</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Anscombe’s Quartet
Visualizations

- Line chart showing trends for Bears, Dolphins, and Whales from January to June.
- Pie chart indicating the percentage distribution of different categories with labels.
- Bar chart illustrating sales data from 2008 to 2012 with categories: Classics, Mystery, Romance, Sci-Fi & Fantasy, Young Adult.
- Donut charts with various segment distributions in A, B, and C.
Parallel Coordinate Plot

Fig. 13. Spectral features of the Water class and their uncertainty characteristics.
Medical Visualizations
Spatial Visualizations
Spatial Visualizations
Spatial Visualizations
Bad Visualizations
Complex Visualizations
What is Visualization?

• Definition #2: Enhancing the visual system to solve problems that algorithms can’t solve efficiently.
Soft Tissue Modeling
Soft Tissue Modeling

Fusion 4.2±0.8mm

Control 6.3±1.1mm
Astronomical Modeling

2dF Galaxy Redshift Survey
Astronomical Modeling
Statistics can be Misleading

- The probability that a woman has breast cancer at age 40 is 1%.
- The probability that the disease can be detected by mammography is 80%.
- The probability that a test misdetects the disease is 9.6%.

- A test comes back positive. What is the probability that she indeed has breast cancer?
- **7.8%**
Statistics can be Misleading

• Assume a population of 1000 women.
  – 1% have breast cancer, so 10
  – 80% of those are correctly diagnosed (8), 20% are missed (2)
  – 99% don’t have breast cancer (990)
  – 90.4% of those are correctly skipped (895), 9.6% are incorrectly diagnosed (95)

• If a test comes back positive (8+95 cases), only $\frac{8}{8+95} = 7.8\%$ are correct diagnoses
Statistics can be Misleading
Genomic Modeling
Urban Planning Modeling

Radius 38.2 m
What is Visualization?

• Definition #3: An iterative visual process used to make hypothesis about data, and then to gain insight/knowledge about that data by exploring what the data is telling us.
Insight

Domain-Specific Hypotheses and Data

Simulate Behavior

Model

Process Data

Evaluate

Explore

Validate

Visually Encode
Any Questions?