Visualization for Transparency
Jack van Wijk
Responsible Data Science Seminar
Amsterdam
December 14th, 2016

The human visual system

http://eofdreams.com

The human visual system

http://vinceantonucci.com

## Anscombe’s quartet

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$y_1$</td>
<td>8.04</td>
<td>9.14</td>
<td>7.46</td>
<td>6.89</td>
</tr>
<tr>
<td>$x_2$</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>$y_2$</td>
<td>6.95</td>
<td>6.77</td>
<td>7.76</td>
<td>5.76</td>
</tr>
<tr>
<td>$x_3$</td>
<td>7.58</td>
<td>7.11</td>
<td>12.74</td>
<td>7.71</td>
</tr>
<tr>
<td>$y_3$</td>
<td>5.68</td>
<td>8.0</td>
<td>19.0</td>
<td>7.04</td>
</tr>
<tr>
<td>$x_4$</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$y_4$</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>7.04</td>
</tr>
</tbody>
</table>

Four data-sets: same average x and y, same variance, same correlation, ...

Francis Anscombe, 1973
Data produce and collect

**BIG MESSY**

Visualization

People want insights

**Standard diagrams don’t scale**

- Multivariate visualization: scatterplot
- Tree visualization: tree diagram
- Graph visualization: node link diagram

**SequoiaView**

Van Wijk and Van de Wetering, 1999

**Botanically inspired treevis**

Kleiberg, Van de Wetering, van Wijk, 2001

**One picture is worth a lot of pixels**

- Focus on visual presentation
- Show everything in one picture?

**Visualization**

Van Wijk and Van de Wetering, 1999

Kleiberg, Van de Wetering, van Wijk, 2001
Information Visualization

- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (Card et al., 1999)

Visualization: exploration process, not a picture.

Statistics, machine learning, data mining, …
Visual Analytics

**Computers**
- fast
- precise
- computation
- search and store data
- graphics

**Humans**
- Flexible, inventive
- Solve problems
- Handle new situations
- Handle incomplete and/or inconsistent information
- Have domain knowledge and experience
- Can see things that are hard to compute

Visual Analytics: Synergy between computers and humans

Data size

<table>
<thead>
<tr>
<th>Business graphics</th>
<th>Infovis</th>
<th>Visual analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>small (1-10)</td>
<td>medium (1000)</td>
<td>huge (&gt; 10^6)</td>
</tr>
</tbody>
</table>

Try to move to the left:
- Filter, aggregate, statistics, machine learning, …
- without loosing essential information

Visual Analytics Time Series Data

![Time series data visualization](image)

Figure 1: Power demand by ECN, displayed as a function of hours and days.

Challenges

- Handling big, messy data
- Dealing with models

- Examples from Eindhoven
- Lessons learned, challenges ahead

The pattern

- combination of simple presentations,
- use interaction and statistics
Dealing with big, messy data

- How to scale up?
- How to generalize?

Dealing with models

- How to define ML/DM models?
  - What model, what parameter settings?
  - Which features of the data?
- How to understand the result?
- How to make the black box transparent?
Context

Evidence → Reasoner → "75% Dangerous!" → Decision Maker

AIS-data, radar data, web data, reports... on vessels → Probabilistic first order logic inference engine → Coast guard

Problem

Evidence → Reasoner → "75% Dangerous!" → Decision Maker

AIS-data, radar data, web data, reports... on vessels → Probabilistic first order logic inference engine → Coast guard

Problem

Evidence → Reasoner → "75% Dangerous!" → Decision Maker

Rationale → Rationale Visualization

Example
How to understand a model?

Custom cases can be solved, using a combination of simple presentations, interaction and statistics

Challenges:
- How to scale?
- How to generalize?
- Can we understand neural networks…?

Conclusions

Data and models can be made transparent, using a combination of simple presentations, interaction and statistics and with smart students and a lot of effort!

Challenges:
- How to scale?
- How to generalize?
- How to develop custom solutions efficiently?

Thank you!