Visualization: The Human Lens to Big Data

Dr. Michael Sedlmaier

Visualization and Data Analysis group (VDA)
Big Data / Data Science

- database
- algorithms

MapReduce
Hadoop
Data Mining
Machine Learning
Visualization

Look at / Interact

- Understand patterns
- Gain insights
- Make decisions
- Communicate
Visualization — Why?
Visualization — Why?

Perception beats cognition
Wait, why humans in the loop? We have great algorithms!

Michael Sedlmair, Miriah Meyer, Tamara Munzner
Design Study Methodology: Reflections from the Trenches and the Stacks (IEEE InfoVis 2012).
Visualization: The human lens to big data

Outline

• Visual Encoding: Principle & Pitfalls
• Visualization of Big Data
• Case Study: Cardiogram
## Visual Encoding

<table>
<thead>
<tr>
<th>dim 1</th>
<th>dim 2</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>3.3</td>
<td>0</td>
</tr>
<tr>
<td>5.4</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>3.6</td>
<td>6.4</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>7.2</td>
<td>0</td>
</tr>
<tr>
<td>4.8</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>7.3</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>5.6</td>
<td>8.3</td>
<td>1</td>
</tr>
<tr>
<td>7.6</td>
<td>2.3</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>7.2</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>5.4</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>4.8</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>7.3</td>
<td>4.8</td>
<td>2</td>
</tr>
<tr>
<td>5.6</td>
<td>7.3</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>5.6</td>
<td>2</td>
</tr>
<tr>
<td>5.4</td>
<td>7.6</td>
<td>2</td>
</tr>
<tr>
<td>3.6</td>
<td>5.4</td>
<td>2</td>
</tr>
</tbody>
</table>

**Data**

**Visualization**
Visual Encoding: 3D?

<table>
<thead>
<tr>
<th>dim 1</th>
<th>dim 2</th>
<th>dim 3</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>3.3</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>5.4</td>
<td>7.5</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>3.6</td>
<td>6.4</td>
<td>4.8</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>7.2</td>
<td>7.3</td>
<td>0</td>
</tr>
<tr>
<td>4.8</td>
<td>5.4</td>
<td>5.6</td>
<td>0</td>
</tr>
<tr>
<td>7.3</td>
<td>1.3</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>5.6</td>
<td>8.3</td>
<td>5.4</td>
<td>1</td>
</tr>
<tr>
<td>7.6</td>
<td>2.3</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>1.1</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>7.2</td>
<td>1.2</td>
<td>7.5</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>5.4</td>
<td>6.4</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>3.6</td>
<td>7.2</td>
<td>1</td>
</tr>
<tr>
<td>4.8</td>
<td>1.4</td>
<td>5.4</td>
<td>2</td>
</tr>
<tr>
<td>7.3</td>
<td>4.8</td>
<td>1.3</td>
<td>2</td>
</tr>
<tr>
<td>5.6</td>
<td>7.3</td>
<td>8.3</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>5.6</td>
<td>3.3</td>
<td>2</td>
</tr>
<tr>
<td>5.4</td>
<td>7.6</td>
<td>7.5</td>
<td>2</td>
</tr>
<tr>
<td>3.6</td>
<td>5.4</td>
<td>1.2</td>
<td>2</td>
</tr>
</tbody>
</table>

...
## Visual Encoding: 3D?

What's the value of this point?

<table>
<thead>
<tr>
<th>dim 1</th>
<th>dim 2</th>
<th>dim 3</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>3.3</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>5.4</td>
<td>7.5</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>3.6</td>
<td>6.4</td>
<td>4.8</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>7.2</td>
<td>7.3</td>
<td>0</td>
</tr>
<tr>
<td>4.8</td>
<td>5.4</td>
<td>5.6</td>
<td>0</td>
</tr>
<tr>
<td>7.3</td>
<td>1.3</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>5.6</td>
<td>8.3</td>
<td>5.4</td>
<td>1</td>
</tr>
<tr>
<td>7.6</td>
<td>2.3</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>1.1</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>7.2</td>
<td>1.2</td>
<td>7.5</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>5.4</td>
<td>6.4</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>3.6</td>
<td>7.2</td>
<td>1</td>
</tr>
<tr>
<td>4.8</td>
<td>1.4</td>
<td>5.4</td>
<td>2</td>
</tr>
<tr>
<td>7.3</td>
<td>4.8</td>
<td>1.3</td>
<td>2</td>
</tr>
<tr>
<td>5.6</td>
<td>7.3</td>
<td>8.3</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>5.6</td>
<td>3.3</td>
<td>2</td>
</tr>
<tr>
<td>5.4</td>
<td>7.6</td>
<td>7.5</td>
<td>2</td>
</tr>
<tr>
<td>3.6</td>
<td>5.4</td>
<td>1.2</td>
<td>2</td>
</tr>
</tbody>
</table>

...
Visual Encoding: 3D?

<table>
<thead>
<tr>
<th>dim 1</th>
<th>dim 2</th>
<th>dim 3</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>3.3</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>5.4</td>
<td>7.5</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>3.6</td>
<td>6.4</td>
<td>4.8</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>7.2</td>
<td>7.3</td>
<td>0</td>
</tr>
<tr>
<td>4.8</td>
<td>5.4</td>
<td>5.6</td>
<td>0</td>
</tr>
<tr>
<td>7.3</td>
<td>1.3</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>5.6</td>
<td>8.3</td>
<td>5.4</td>
<td>1</td>
</tr>
<tr>
<td>7.6</td>
<td>2.3</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>1.1</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>7.2</td>
<td>1.2</td>
<td>7.5</td>
<td>1</td>
</tr>
<tr>
<td>5.4</td>
<td>5.4</td>
<td>6.4</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>3.6</td>
<td>7.2</td>
<td>1</td>
</tr>
<tr>
<td>4.8</td>
<td>1.4</td>
<td>5.4</td>
<td>2</td>
</tr>
<tr>
<td>7.3</td>
<td>4.8</td>
<td>1.3</td>
<td>2</td>
</tr>
<tr>
<td>5.6</td>
<td>7.3</td>
<td>8.3</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>5.6</td>
<td>3.3</td>
<td>2</td>
</tr>
<tr>
<td>5.4</td>
<td>7.6</td>
<td>7.5</td>
<td>2</td>
</tr>
<tr>
<td>3.6</td>
<td>5.4</td>
<td>1.2</td>
<td>2</td>
</tr>
</tbody>
</table>

Scatterplot Matrix

No 3D for abstract data!

Michael Sedlmair, Tamara Munzner, Melanie Tory
Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices (InfoVis 2013).
Visual Encoding: Color?


noncontiguous small regions of color: only 6-12 bins
ColorBrewer
Human Perception

Popout
Human Perception

Popout
Human Perception

Popout: combining channels?
Human Perception

Visual Channels (Jock Mackinlay)
Design

Data-to-Ink Ratio (Edward Tufte)
Design

Data-to-Ink Ratio (Edward Tufte)
What we can do now ...  

Static representations to present information
BUT:
Big Data?
Visualization

Data

volume
velocity
variety
veracity

...
Showing all the data?

https://twitter.com/axelmaireder/media
Interaction

eexample: baby name voyager

Johnny
Rank in 1950s: 54
Guidance

example: finding interesting projections in high-dim data?

3D Scatterplot Matrix

—> 3 Views
Guidance

example: finding interesting projections in high-dim data?

10D $\rightarrow$ 45 Views

200D $\rightarrow$ $\sim$20k Views
Guidance

Visual quality measure, e.g. class separation measures

Guidance

Visual quality measure, e.g. class separation measures

Good!

Abstraction/Aggregation

E.g. dimension reduction

Matthew Brehmer, Michael Sedlmair, Stephen Ingram, Tamara Munzner
Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences (BELIV 2014).
What we can do now...

• **Combined Visual & Algorithmic Analysis**
  - *called:* Visual Analytics, Visual Data Science, ...
  - or just Visualization

• **Build powerful systems for visual analysis to**
  - explore the data
  - gain insights
  - support decision making
  - share and communicate
Case Study: Cardiogram

Visual Analytics for Automotive Engineers

M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, A. Butz

more and more electronics...
... enabled by in-car communication networks

~70 Controllers / Car
Problem Characterization

- **data**: recorded traces (15k messages/sec)
- **task**: finding errors
- **process**: test drives
- **current practices**: mainly textual lists
Problem Characterization

• Main Challenges
  • handling masses of test traces (large and many)
  • understanding correlation between trace and car behavior
  • …
Cardiogram
overview

Trace → State Machine Pre-Analysis → Visualization
Cardiogram

state machine pre-analysis
State Machines: Behavior
State Machines: Error Detection

Correct \quad \rightarrow \quad Error

\[ \text{Diagram showing state transitions from Correct to Error and vice versa.} \]
Cardiogram
create/select state machines

**Multiple State Machines**
- State Machine 1
- State Machine 2
- State Machine 3
  ... (dozens)
Cardiogram

data abstraction/reduction

Trace → State Machine Pre-Analysis

from each state machine

Visualization

Verification Tag
error / warning / ok

Transition List
time: state \( x \rightarrow state \ y \)

...
Hey Gunther,
I think I know the reason for this. Please contact me so we can talk about it.
Hank
Longitudinal field study

1 year, 15 engineers
Longitudinal field study results

- externalization and sharing of expert knowledge
Longitudinal field study results

- externalization and sharing of expert knowledge
- complete coverage of traces vs. sparse sampling
Longitudinal field study results

- externalization and sharing of expert knowledge
- complete coverage of traces vs. sparse sampling
- understand behavioral correlations

Overpressure error:
- ok

Doors:
- Door 1 closed
- Door 2 closed
- Door 3 closed
- Door 4 closed
Could we have done it with a general purpose tool?

• Likely not, not yet …

• Complex problem space (big data)
• Complex solution space (visual encoding, interaction, abstraction)
Visualization: The Human Lens to Big Data

http://www.microscope-antiques.com/grunow.html

http://galileo.rice.edu/sci/instruments/telescope.html
Summary

• Visualization for Human-in-the-Loop tasks

• Visualization is not just about pretty images
  • perception, design
  • interaction, guidance, abstraction

• Case study: Cardiogram
Thanks

VDA Group @ University of Vienna

Many, many other collaborators

...
Thanks to you!

Visualization: The Human Lens to Big Data

Dr. Michael Sedlmair