Computer-Based Visualization

Data Mining and Visualization

Université de Rennes 1, M2 SIF
Alexandre Termier
Ferran Argelaguet
Web site of the course (Remember)

• http://people.irisa.fr/Alexandre.Termier/dmv/

• Web site contains:
  • General information
  • Up-to-date schedule (will be the reference)
  • Links to documents
Definition

Computer-based visualization systems provide **visual representations of datasets** designed to help people carry out tasks **more effectively**.
Context

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- Need to present results to users -> data visualization
  - Huge lack of communication between data mining / data vis community
  - This course: a small step to improve this communication
Where?

Data -> Target Data -> Preprocessed Data -> Preprocessing -> Transformation -> Data Mining -> Interpretation/ Evaluation -> Knowledge

What? Why? How?
Where?

http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
Simple Example

1.1.1: MY, NP; MY, MB
1.1.2: MY, ME; ME, MB
1.1.3: MY
1.1.4: MY, ME; MY, CL; MY, GE; MY, MC; MY, MB
1.1.5: MY, MB, ME
1.1.6: ME, MY
1.1.7: MY, CV; MY, MB, ME
1.1.8: SN, MY
1.1.9: MB
1.1.10: MY, GG
1.1.11: MY
1.1.12: MY
1.1.13: MY
1.1.14: MY, SN
1.2.1: JL, JV; JV, MT; MR, JV

https://bost.ocks.org/mike/miserables/
Context

• Scientific visualization is mostly concerned with:
  • 2, 3, 4 dimensional, spatial or spatio-temporal data
  • Discretized data

• Information visualization focuses on:
  • High-dimensional, abstract data
  • Discrete data
  • Financial, statistical, etc
  • Visualization of large trees, networks, graphs
  • Data mining: finding patterns, clusters, voids, outliers
Content

• Introduction
• Data abstraction
• Task abstraction
• Visualization Design
• Validation
Introduction

Why data vis (in general)?

Overview
System Overview

1. Raw Data
2. Algorithms
3. Processed Data
4. User Interface
5. Visualization
6. User
7. Interaction
8. Perception
Bring the computer in the loop

- Enable the exploration of large datasets
  - Temporal dimension

Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation.
Bring the computer in the loop

• Computational capacity
  • Scalability: How to deal with increasingly data sizes?

• Display capacity
  • Information density: limit of the displayed information
Bring the computer in the loop
Bring the human in the loop

• The question is not known in advance
  • The analysis process is driven by the user
  • Take advantage of the ability of humans to find patterns

• Augment the human decision process
  • Improve decision making
Provide external representations

• Offload internal cognition and memory usage to the perceptual system
  • Data organized by spatial location
  • Support perceptual inference
  • Select just the relevant information

The Wealth & Health of Nations

https://bost.ocks.org/mike/nations/
Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td></td>
<td>x</td>
<td>v</td>
<td>x</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>10,00</td>
<td></td>
<td>8,04</td>
<td></td>
<td>10,00</td>
<td>9,14</td>
<td></td>
<td>10,00</td>
</tr>
<tr>
<td>8,00</td>
<td></td>
<td>6,95</td>
<td></td>
<td>8,00</td>
<td>8,14</td>
<td></td>
<td>8,00</td>
</tr>
<tr>
<td>13,00</td>
<td></td>
<td>7,58</td>
<td></td>
<td>13,00</td>
<td>8,74</td>
<td></td>
<td>13,00</td>
</tr>
<tr>
<td>9,00</td>
<td></td>
<td>8,81</td>
<td></td>
<td>9,00</td>
<td>8,77</td>
<td></td>
<td>9,00</td>
</tr>
<tr>
<td>11,00</td>
<td></td>
<td>8,33</td>
<td></td>
<td>11,00</td>
<td>9,26</td>
<td></td>
<td>11,00</td>
</tr>
<tr>
<td>14,00</td>
<td></td>
<td>9,96</td>
<td></td>
<td>14,00</td>
<td>8,10</td>
<td></td>
<td>14,00</td>
</tr>
<tr>
<td>6,00</td>
<td></td>
<td>7,24</td>
<td></td>
<td>6,00</td>
<td>6,13</td>
<td></td>
<td>6,00</td>
</tr>
<tr>
<td>4,00</td>
<td></td>
<td>4,26</td>
<td></td>
<td>4,00</td>
<td>3,10</td>
<td></td>
<td>4,00</td>
</tr>
<tr>
<td>12,00</td>
<td></td>
<td>10,84</td>
<td></td>
<td>12,00</td>
<td>9,13</td>
<td></td>
<td>12,00</td>
</tr>
<tr>
<td>7,00</td>
<td></td>
<td>4,82</td>
<td></td>
<td>7,00</td>
<td>7,26</td>
<td></td>
<td>7,00</td>
</tr>
<tr>
<td>5,00</td>
<td></td>
<td>5,68</td>
<td></td>
<td>5,00</td>
<td>4,74</td>
<td></td>
<td>5,00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9,0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linear Regression</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>y=3 + 0.5x</td>
<td></td>
</tr>
</tbody>
</table>

[Anscombe’s Quartet]
Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns

https://matplotlib.org/gallery.html
Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns


https://matplotlib.org/gallery.html
Depend on vision...

• The Visual system is a high-bandwidth channel of information
  • A significant amount of visual information is done in parallel at the preconscious level
... but be careful

• Need to know its strengths and limitations
... but be careful

• Need to know its strengths and limitations
• Human perceptual and cognitive capacity limits?
... but be careful

• Need to know its strengths and limitations
• Human perceptual and cognitive capacity limits?
  • Change blindness: the phenomenon where even very large changes are not noticed if we are attending to something else in our view
Interactivity

• For big datasets, the user nor the display can enable the visualization of the entire dataset at once

https://bl.ocks.org/mbostock/5944371
Interactivity

• A static view only allows for a single view of a limited number of dimensions

• The same dataset can be sorted using a number of different criteria

https://bost.ocks.org/mike/nations/
Introduction

Why data vis?

Overview
System Overview

Interaction → User Interface → Algorithms

User Interface → Visualization → Processed Data

Raw Data → Algorithms

User → Perception
Visualization Design
Data and Dataset Types

→ Items

→ Attributes

→ Links

→ Positions

→ Grids

→ Tables
→ Networks
→ Fields (Continuous)

→ Multidimensional Table
→ Trees

→ Geometry (Spatial)

→ Dataset Availability
→ Static
→ Dynamic
Attributes

• Definition:
  An attribute is some specific property that can be measured, observed or logged

• Categorical
  • Non-numerical, limited number of values, usually fixed
  • Non-ordered, unless explicitly stated

• Ordinal
  • Non-numerical, limited number of values, usually fixed
  • Ordered

• Numerical
  • Quantitative, Infinite number of values
  • Ordered
Bar Charts vs Line Charts

Scatterplot

Multi dimensional table: N countries * T time
Attributes: 3 Qualitative

https://bost.ocks.org/mike/nations/
Chord Diagram

Network
Dependences between classes in a software class hierarchy.

https://bl.ocks.org/mbostock/1046712
Time Varying Data

Table: T time
Attributes: 1 Categorical, 1 Quantitative

http://www.babynamewizard.com/voyager
Tensor Visualization

Comparison: Ellipsoids vs. superquadrics (Kindlmann)

Image credit: W. Shen
Task Abstraction

• Reframe tasks from domain-specific form into abstract form
  • Ease the design process using references

• The focus on tasks enables the improvement of them
  • Making more effective, user-friendly...

• Tasks can be classified in:
  • Actions – Verb - High-level task
  • Targets – Noun - Precise goal
Action Classification

- Analyze
  - Consume

- Produce

- Search

<table>
<thead>
<tr>
<th></th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>Lookup</td>
<td>Browse</td>
</tr>
<tr>
<td>Location unknown</td>
<td>Locate</td>
<td>Explore</td>
</tr>
</tbody>
</table>

- Query

- Identify
- Compare
- Summarize
Target Classification

• All Data
  → Trends
  → Outliers
  → Features

• Attributes
  • One
    → Distribution
    → Extremes
  • Many
    → Dependency
    → Correlation
    → Similarity

• Network Data
  → Topology
  → Paths

• Spatial Data
  → Shape

Why?
Network Visualization

Visualization

Data Abstraction

Task Abstraction

Design

Why?

What?

How?
How to...

• ... arrange the data in the view spatially?
Arrange / Encode

How?

Express  
Separate

Order  
Align

Use

Scatterplot

Pie Chart  
Bullseye chart  
Bar Plot  
Coxcomb plot

How to...

• ... **arrange** the data in the view spatially?
• ... **map** the data with all non-spatial channels?
Display

• What is the best way to visually encode the data?
  • How the human visual system discriminate visual stimuli?

• Two major dimensions
  • Color
  • Shape
Visual Encoding

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

Effectiveness

How?
Color Encoding

• Color Components

• Color sensitivity

How?

https://commons.wikimedia.org/wiki/File:LuminosityCurve2.svg
https://commons.wikimedia.org/wiki/File:Hsl-hsv_models_b.svg
Color Encoding

http://colorbrewer2.org/

Visual discrimination

How?

How to...

• ... **arrange** the data in the view spatially?
• ... **map** the data with all non-spatial channels?
• ... **manipulate** the data in the view?
Manipulate

Titanic Survivors

Survived
Perished
Perished → Male
1,364 (62%)

Sex
Male

Female

Age
Adult

Child

Class
Second Class
First Class
Third Class
Crew

Curves?

Data: Robert J. MacG. Dawson

https://www.jasondavies.com/parallel-sets/
Manipulate

https://bl.ocks.org/mbostock/5944371

How?

- Change
- Select
- Navigate
How to...

• ... arrange the data in the view spatially?
• ... map the data with all non-spatial channels?
• ... manipulate the data in the view?
• ... facet data between views?
Facet (multi-view)

Weaver. Building Highly-Coordinated Visualizations In Improvise. Proc. InfoVis 2004, p. 159-166
How to…

• ... **arrange** the data in the view spatially?
• ... **map** the data with all non-spatial channels?
• ... **manipulate** the data in the view?
• ... **facet** data between views?
• ... **reduce** the data in the view?
Reduce

Reduce


https://matplotlib.org/gallery.html
Reduce

http://www.babynamewizard.com/voyager

Toolglass and magic lenses: the see-through interface. ACM SIGGRAPH '93
Validation

• How do you know if it works?

• How do you measure it?
  • How do you decide the benchmark data?

• How do you argue that one design is better or worse than another?

• What does it better mean?
  • Do user get something done faster?
Why it can go wrong?

- I do not understand
  - Poor match with the properties of the human perceptual and cognitive system

- Lack of context
  - Comprehensible by a human in some other setting

- Scalability problems
Why it can go wrong?

http://yifanhu.net/GALLERY/GRAPHS
Good Practices

• Do not optimize, satisfy
  • The perfect visualization does not exist!!

• Need to know a wide variety of vis techniques
  • Ease the design process using references

• Justification and Alternatives
  • 3D vs 2D?

• Eyes Beat Memory
  • Comparing views better than switch views

• Interactivity and Responsiveness
Life Cycle

• Follow the **life cycle** of a visualization tool
  • Profile of the user (expert, novice,...)

Wrap Up?

http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
Further reading

• Remy Dautriche. Multi-scale Interaction Techniques for the Interactive Visualization of Execution Traces. Chapter 3.3, Pattern Visualization, 2016


  • http://ieeexplore.ieee.org/abstract/document/5290695/


  • http://ieeexplore.ieee.org/abstract/document/841119/

Further reading

• Visualization Analysis and Design
• Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations
• The Functional Art: An introduction to information graphics and visualization