Information Visualization

Harry Hochheiser
DBMI

harryh@pitt.edu
410 648 9300
Why Visualization?

<table>
<thead>
<tr>
<th>State</th>
<th>College Degree %</th>
<th>Per Capita Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>20.6%</td>
<td>11436</td>
</tr>
<tr>
<td>Alaska</td>
<td>30.3%</td>
<td>17610</td>
</tr>
<tr>
<td>Arizona</td>
<td>27.1%</td>
<td>13461</td>
</tr>
<tr>
<td>Arkansas</td>
<td>17.0%</td>
<td>10520</td>
</tr>
<tr>
<td>California</td>
<td>31.3%</td>
<td>16409</td>
</tr>
<tr>
<td>Colorado</td>
<td>33.9%</td>
<td>14821</td>
</tr>
<tr>
<td>Connecticut</td>
<td>33.8%</td>
<td>20189</td>
</tr>
<tr>
<td>Delaware</td>
<td>27.9%</td>
<td>15854</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>36.4%</td>
<td>18881</td>
</tr>
<tr>
<td>Florida</td>
<td>24.9%</td>
<td>14698</td>
</tr>
<tr>
<td>Georgia</td>
<td>24.3%</td>
<td>13631</td>
</tr>
<tr>
<td>Hawaii</td>
<td>31.2%</td>
<td>15770</td>
</tr>
<tr>
<td>Idaho</td>
<td>25.2%</td>
<td>11457</td>
</tr>
<tr>
<td>Illinois</td>
<td>26.8%</td>
<td>15201</td>
</tr>
<tr>
<td>Indiana</td>
<td>20.9%</td>
<td>13149</td>
</tr>
<tr>
<td>Iowa</td>
<td>24.5%</td>
<td>12422</td>
</tr>
<tr>
<td>Kansas</td>
<td>26.5%</td>
<td>13300</td>
</tr>
<tr>
<td>Kentucky</td>
<td>17.7%</td>
<td>11153</td>
</tr>
<tr>
<td>Louisiana</td>
<td>19.4%</td>
<td>10635</td>
</tr>
<tr>
<td>Maine</td>
<td>25.7%</td>
<td>12957</td>
</tr>
<tr>
<td>Maryland</td>
<td>31.7%</td>
<td>17730</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>34.5%</td>
<td>17224</td>
</tr>
<tr>
<td>Michigan</td>
<td>24.1%</td>
<td>14154</td>
</tr>
<tr>
<td>Minnesota</td>
<td>30.4%</td>
<td>14389</td>
</tr>
<tr>
<td>Minnesota</td>
<td>30.4%</td>
<td>14389</td>
</tr>
<tr>
<td>Mississippi</td>
<td>19.9%</td>
<td>9648</td>
</tr>
<tr>
<td>Missouri</td>
<td>22.3%</td>
<td>12989</td>
</tr>
<tr>
<td>Montana</td>
<td>25.4%</td>
<td>11213</td>
</tr>
<tr>
<td>Nebraska</td>
<td>26.0%</td>
<td>12452</td>
</tr>
<tr>
<td>Nevada</td>
<td>21.5%</td>
<td>15214</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>32.4%</td>
<td>15959</td>
</tr>
<tr>
<td>New Jersey</td>
<td>30.1%</td>
<td>18714</td>
</tr>
<tr>
<td>New Mexico</td>
<td>25.5%</td>
<td>11246</td>
</tr>
<tr>
<td>New York</td>
<td>29.6%</td>
<td>16501</td>
</tr>
<tr>
<td>North Carolina</td>
<td>24.2%</td>
<td>12885</td>
</tr>
<tr>
<td>North Dakota</td>
<td>28.1%</td>
<td>11051</td>
</tr>
<tr>
<td>Ohio</td>
<td>22.3%</td>
<td>13461</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>22.8%</td>
<td>11893</td>
</tr>
<tr>
<td>Oregon</td>
<td>27.5%</td>
<td>13418</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>23.2%</td>
<td>14068</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>27.5%</td>
<td>14981</td>
</tr>
<tr>
<td>South Carolina</td>
<td>23.0%</td>
<td>11897</td>
</tr>
<tr>
<td>South Dakota</td>
<td>24.6%</td>
<td>10661</td>
</tr>
<tr>
<td>Tennessee</td>
<td>20.1%</td>
<td>12255</td>
</tr>
<tr>
<td>Texas</td>
<td>25.5%</td>
<td>12904</td>
</tr>
<tr>
<td>Utah</td>
<td>30.0%</td>
<td>11029</td>
</tr>
<tr>
<td>Vermont</td>
<td>31.5%</td>
<td>13527</td>
</tr>
<tr>
<td>Virginia</td>
<td>30.0%</td>
<td>15713</td>
</tr>
<tr>
<td>Washington</td>
<td>30.9%</td>
<td>14923</td>
</tr>
<tr>
<td>West Virginia</td>
<td>16.1%</td>
<td>10520</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>24.9%</td>
<td>13276</td>
</tr>
<tr>
<td>Wyoming</td>
<td>25.7%</td>
<td>12311</td>
</tr>
</tbody>
</table>

Which state has the highest % college degree?

Highest Income? Relationship between college and income?

Thanks to Chris North for visuals
Electronic Medical Records
Can we do better?
From Data to Knowledge

• “The purpose of Computing is Insight, not Numbers” - Richard Hamming, 1961

• Goal: go from masses of raw data to insight
Amplifying Cognition

• Use tools to help us think
• Papers on desks, blackboards, etc.
• Short-term memory aids
• Visualizations
• Increase memory and processing resources
• Reduce search
• Enhance pattern detection
• Enable perceptual inference
• Support monitoring
• Encode information in manipulable medium
“Knowledge Crystallization”

• Buying a new laptop

From Card, Mackinlay, and Shneiderman, Readings in Information Visualization: Using Vision to Think
Mental Model

- Internal representation(s) of problem/domain of interest
- Often multiple models – cognitive collage
- Which is further west – Reno or Los Angeles?
Information Visualization

• “The use of computer-supported, interactive, visual representations of abstract data to amplify cognition” Card, Mackinlay, Shneiderman 1999

• visualize: to form a mental model or mental image of something

• Information visualization: 'abstract' concepts

• No “given” representation
Three Pillars of Information Visualization

- Visual Representations
  - Exploit the power of human perception
- Interactivity
  - Support direct engagement with data
- Rich data types
  - Heterogenous data for a heterogenous world
Visual Analytics

- Analytical Reasoning
- Visual Representations and Interaction Techniques
- Data Representations and Transformations
- Production, Presentation, and Dissemination
Building Visualizations

- Reference model for visualization – (Card, Mackinlay, Shneiderman 1999)
- 3 stage model (Spence, 2007)
  - Representation
  - Presentation
  - Interaction
Visual Representation: A Picture is worth a thousand words

- Visual bandwidth is huge — more so than for other senses
- Fast, parallel
- Good pattern recognition
- Pre-attentive processing

Appropriate visual representations of data can reveal structure, aid cognition, and facilitate development of understanding.
Pre-attentive processing

• Find the red square
Find the blue circle?
Representation: Visually encode data

- Map data attributes to
  - Spatial Substrate – 2D? 3D? Zoomable?
  - Visual Marks – points, lines, volumes
  - Retinal Properties – position, size, orientation, color, shape, animation, transparency, enclosure

- Representations must be
  - computable
  - invertible
History: Static Graphs and Visualizations

William Playfair
Napoleon's March and Retreat

http://www.napoleonic-literature.com/1812/1812.htm

based on work by Minard, 1861

Harry Hochheiser, harryh@pitt.edu
John Snow's Cholera map (1854)
Florence Nightingale
Crimean War casualties (1855)
An Infamous example: the Challenger Disaster

• 28 January 1986
Alternative View

Edward Tufte: Visual Explanations
Back to EMRs - Visualization
Powsner & Tufte 1994
Abstractions can help

• When appropriate, trade fidelity for understanding.

• Subway maps
  • Don't care about exact shape of routes
  • Care about relative positions and approximate location
London Underground

http://www.kottke.org/plus/misc/images/tubegeo.gif
London Underground: Alternative

Challenges

- Challenge – never enough pixels
- Use
  - Overview + detail - maps, pdf preview
  - Context + focus - hyperbolic browser
  - Suppression
  - Animation
- “Overview First, Zoom and Filter, Details on Demand”
Focus + Context: Hyperbolic Browser

http://www.ibiblio.org/openvideo/video/chi/chi96_02_m1.mpg

Http://www.thejit.org
Beyond static graphs - interaction
http://babynamewizard.com/voyager
Interaction: Static to Dynamic

- Interactive tools add additional value
- Choose which subsets of data to see
- How to present them
- Interactive user controls
- Rapid, Incremental, reversible actions
- Animation
- Greater insight from larger data sets
Tasks

• Navigation
• Browsing
• Searching
• Distributions and global attributes: min/max, average
• Outliers
• Serendipitous discovery
• Identifying spurious data
Types of Data  (Shneiderman, 1996)

- 1-Dimensional: documents, source code, sequential lists
- 2-D: maps, floorplans, grids
- 3-D: physical objects
- Multi-D: multi-attribute data
- Temporal: time-varying data
- Tree/Hierarchical: file systems, organization charts
- Network: arbitrary (possibly typed) relationships between objects
- Arbitrary combinations of the above...
Expression Levels Over Time: Timeboxes and TimeSearcher

- **Timeboxes:** boxes on graph space are queries

  - TimeSearcher
    - InfoVis tool based on timeboxes
    - Overviews of data and result sets
    - Query-by-example
    - Dynamic query (100ms) performance
    - Java: Swing+Piccolo

  - Combine multiple boxes for conjunctive queries.

  - Also used for sequence analysis
Hierarchical Data - Treemap

- http://Newsma.jp
Multi-Dimensional data
Parallel Coordinates

- Infoscope – Macrofocus.com (Brodbeck & Girardin, 2001)
- Coordinate axes as dynamic query sliders
InfoVis & Biomedical Applications

- Bioinformatics
- Patient Records
- Information Seeking
Gene expression microarrays

- How much is each gene expressed, relative to some reference point?
- ~ 20,000 genes, 1-20 conditions

Spellman, et al. 1998
InfoVis & Bioinformatics: Microarrays

- Hierarchical Clustering Explorer - HCE
Networks - Cytoscape
www.cytoscape.org
Genomics

- Comparing sequences - across organisms, conditions, datasets, etc.

- Genome Browsers..
Genomics – VistaChrom
(Kincaid, et al. 2005)

- ArrayCGH – copy number variation
MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data
Related Efforts

AI/NLP

InfoVis

NLP + Vis for Medical Record Review?

Data Mining KDD
FacetAtlas: Multifaceted Visualization for Rich Text Corpora
Cao, et al. 2010
Clinical – Visualizing Patient Records

  • www.cs.umd.edu/hcil/lifelines
  • www.cs.umd.edu/hcil/lifelines2
  • http://www.cs.umd.edu/hcil/lifeflow/
  • http://www.cs.umd.edu/hcil/similan/
EventFlow
http://www.cs.umd.edu/hcil/eventflow/
Midgaard - semantic zoom
Bade, et al. 2004
Genomic research in Alpha1-Antitrypsin Deficiency and Sarcoidosis

- Relate
  - Phenotype – extensive clinical data
  - Microbiome – prevalence of various bacteria in lung samples
  - Expression – gene expression (likely RNA-seq)

- Graduate project opportunities!
Demo

GRADS: Genomic Research in Alpha-1-Antitrypsin Deficiency and Sarcoidosis

Status: Clinical data is loaded. Microbiome data is loaded. Gene Expression data is loaded.

Selected Records: 404 Total Filtered Records: 404 Reset

Clinical Data

Age

Gender

BMI

Diagnosis

Alpha-1(Participant)

Sarcoid(Participant)

Expand all Collapse all

Microbiome

Gene expression

Filters

Downloads

Data Table

Copyright ©2014 DBMI Development

Updated: 01-23-2015
Demo

GRADS: Genomic Research in Alpha-1-Antitrypsin Deficiency and Sarcoidosis

Clinical Data

- Age
- Gender
- BMI
- Diagnosis

Filters

- Save Current Query
- Remove Latest Filter

- Age [53 -> 80]

- Downloads
- Data Table

Microbiome

Gene expression

Copyright ©2014 DBMI Development

Updated: 01-23-2015
Demo
Demo
The Monarch Initiative

www.monarchinitiative.org

- Linking animal models to human disease
  - Human disease
  - Human genes
  - Mouse genes
  - Mouse phenotypes
  - Mouse models
  - Mouse strains
- Ontologically-informed analysis
- Visualize the links and understand the associations?
- Ongoing work with OHSU, LBL, others
The Monarch PhenoGrid
TwinList – Medication Reconciliation


http://www.youtube.com/watch?v=YoSxIKl0pCo
Genome Projector: zoomable genome map with multiple views

Arakawa, et al. BMC Bioinformatics 2009

http://www.g-language.org/GenomeProjector/
The design, development, and deployment of a tabletop interface for collaborative exploration of genomic data


- G-nome Surfer
Caleydo Stratomex

Patient storytelling
McCurdy, 2015
Challenges

- Evaluation – how well does it work? What works and why?
- Moving from exploration and identification to synthesis
  - How can I describe what I've found and use it to build understanding?
  - Post-value recall
- Integration of diverse data sets
  - Genomic, phenotypic, environmental, clinical, image
References

- Information Visualization (R. Spence 2007)
- Infovis wiki - www.infovis-wiki.net
- www.visualcomplexity.org
- www.manyeyes.com
- Processing toolkit - www.processing.org
- Protovis - vis.stanford.edu/protovis
- JavaScript InfoVis toolkit - www.thejit.org