CSE 332
Introduction to Visualization

Visual Analytics & The Visual Sense Making Process

Klaus Mueller

Computer Science Department
Stony Brook University
<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topic</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intro, schedule, and logistics</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Basic tasks and data types</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Data sources and what to look out for</td>
<td>Project 1 out</td>
</tr>
<tr>
<td>4</td>
<td>Data preparation and representation</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Data reduction, notion of similarity and distance</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Dimension reduction</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Bias, introduction to plotly and plotly Dash</td>
<td>Project 2 out</td>
</tr>
<tr>
<td>8</td>
<td>Visual perception and cognition</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Visual design and aesthetic</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Introduction to D3</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cluster analysis</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>High-dimensional data VIS: numerical data (linear schemes)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>High-D data VIS: non-linear embeddings, categorical data</td>
<td>Project 3 out</td>
</tr>
<tr>
<td>14</td>
<td>Visualization of spatial data: applications and data origins</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Visualization of spatial data: volume rendering</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Visualization of spatial data: raycasting, transfer functions</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Scientific visualization</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Non-photorealistic and illustrative rendering</td>
<td>Project 4 out</td>
</tr>
<tr>
<td>19</td>
<td>Midterm review</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Midterm discussion</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Principles of interaction</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Visual analytics and the visual sense making process</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Visualization of graphs and hierarchies</td>
<td>Project 5 out</td>
</tr>
<tr>
<td>25</td>
<td>Visualization of time-varying and streaming data</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Maps</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Memorable visualizations, visual embellishments</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Evaluation and user studies</td>
<td></td>
</tr>
</tbody>
</table>
Why Visual Analytics?

Big Data

12+ TBs of tweet data every day

25+ TBs of log data every day

30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones world wide

100s of millions of GPS enabled devices sold annually

76 million smart meters in 2009... 200M by 2014

2+ billion people on the Web by end 2011
Visual Analytics
Problems With Scalability

Must be scalable to

- number of data points
- number of dimensions
- data sources
- diversity of data sources (heterogeneity)
- number of users
- diversity of users and tasks
- quality of the data

Visual Analytics comes to the rescue...
Ease understanding of the data by providing an effective visual representation

**Amplify Perception**

*Detect the Expected, Discover the Unexpected™*
WHAT IS VISUAL ANALYTICS

Visualization plus...

- interaction (HCI)
- data processing (analytics)
- story telling
- scientific approach

but also...

- intelligent computing (AI, machine learning)
- behavioral psychology (cognitive science, human factors)

Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface

Agenda setting book:
http://nvac.pnl.gov/agenda.stm
The Daniel Keim Mantra of Visual Analytics

"Analyze First - Show the Important – Zoom, Filter and Analyze Further - Details on Demand"

The triangle of Visual Analytics (VA)
Intelligence analysis is challenging
Huge amounts of data
Low signal vs. noise (SNR)
Many data types
  - text, images, video, sensor data, etc.
Uncertainty
Contradictions
Omissions
Use of Visualization

Visual perception
- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction
- direct manipulation
- two-way communication

Recall intro lecture on the human visual system...
Use of Visualization

Visual perception
- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction
- direct manipulation
- two-way communication

*Recall intro lecture on the human visual system...*  
*But... humans are imperfect*
Humans tend to overlook/ignore non-focus (and unexpected) objects even when very close and obvious

- note the Visual Analytics slogan: *Detect the Unexpected*

Humans also have limited working memory

- fine details are quickly forgotten when focus changes
- big effect in animated or interactive visualizations
- need to preserve temporal context
EXAMPLE #1

Spot a difference?

This is called change blindness
In this video you will do some counting.

It is very important that you get the right number!

Ready?

YouTube

Video by Dan Simons (U Illinois)
Another distraction experiment

[YouTube]

Video by Dan Simons (U Illinois)
Thoroughly studied by Dan Simons (U Illinois)
  ▪ see http://www.dansimons.com/index.html

Visual Analytics tools
  ▪ help human analysts cope with insufficient memory
    → visualizations externalize memory
    → allow humans to perform visual queries (see C. Ware book)
  ▪ help human analysts deal with change blindness
    → analytics can detect changes
    → visualization can highlight/emphasize these changes
  ▪ we have seen many visual tools this semester
    → this lecture is more about strategy building
The Magic Number **Seven (7)**

- ± 2: the number of things most people can keep in working memory at one time
- causes problems for complicated analysis
  - not more than 7 segments in a pie chart
  - not more than 7 colors in a line chart
  - and so on
do I need these grid lines? maybe not

“Perfection is achieved not when there is nothing more to add, but when there is nothing left to take away.”
— Antoine de Saint-Exupery.
Strategies for Dealing with Complexity

Decomposition

- decompose a complex problem into simpler problems
- get one’s thinking straight in these simpler problems

Externalization

- get the decomposed problem out of your head and down on paper or on a computer screen in some simplified form
- show the main variables, parameters, or elements of the problem and how they relate to each other

Recall principles of information visualization

- overview and detail
- focus and context
- analyze, filter, zoom,...
Mentioned his method of solving decision problems

Why is the decision problem so difficult?
- folks cannot keep all pros and cons in mind at the same time

Solution?
- write down all the pros and cons onto paper in some visible, shorthand form
- allows you make a global judgment effectively
Support visualization with computations for data processing
Form a loop: visualize - refine
Gather (forage) information
Re-represent
  ▪ choose a form that aids analysis
Develop insight
  ▪ through manipulation of representations
Produce results
  ▪ “product”
Nominal Sense-Making Process
Use Visualizations to Evoke The Right Thoughts
How Many 9s Do You See?
How Many 9s Do You See?
Who has the best profit and who has the worst sales?

<table>
<thead>
<tr>
<th>Product Type</th>
<th>Product</th>
<th>Central Sum of Profit</th>
<th>Central Sum of Sales</th>
<th>East Sum of Profit</th>
<th>East Sum of Sales</th>
<th>South Sum of Profit</th>
<th>South Sum of Sales</th>
<th>West Sum of Profit</th>
<th>West Sum of Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>Amaretto</td>
<td>$5,105</td>
<td>$14,011</td>
<td>$1,009</td>
<td>$2,993</td>
<td>($1,225)</td>
<td>$9,265</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Columbian</td>
<td>$8,528</td>
<td>$28,913</td>
<td>$27,253</td>
<td>$47,386</td>
<td>$8,767</td>
<td>$21,664</td>
<td>$11,253</td>
<td>$30,357</td>
</tr>
<tr>
<td></td>
<td>Decaf Irish Cream</td>
<td>$9,632</td>
<td>$26,155</td>
<td>$2,727</td>
<td>$6,261</td>
<td>$2,933</td>
<td>$11,592</td>
<td>($1,305)</td>
<td>$18,235</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>$23,265</td>
<td>$69,080</td>
<td>$30,989</td>
<td>$56,640</td>
<td>$11,700</td>
<td>$33,256</td>
<td>$8,724</td>
<td>$57,856</td>
</tr>
<tr>
<td>Espresso</td>
<td>Caffe Latte</td>
<td>$14,640</td>
<td>$35,218</td>
<td>($6,230)</td>
<td>$16,646</td>
<td>$3,872</td>
<td>$15,442</td>
<td>$7,502</td>
<td>$20,458</td>
</tr>
<tr>
<td></td>
<td>Caffe Mocha</td>
<td>$14,640</td>
<td>$35,218</td>
<td></td>
<td></td>
<td>$5,201</td>
<td>$14,163</td>
<td>$4,064</td>
<td>$18,876</td>
</tr>
<tr>
<td></td>
<td>Decaf Espresso</td>
<td>$8,860</td>
<td>$24,485</td>
<td>$2,410</td>
<td>$7,722</td>
<td>$5,930</td>
<td>$15,384</td>
<td>$12,302</td>
<td>$30,578</td>
</tr>
<tr>
<td></td>
<td>Regular Espresso</td>
<td>$10,062</td>
<td>$24,036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>$23,500</td>
<td>$59,703</td>
<td>$6,242</td>
<td>$48,405</td>
<td>$15,003</td>
<td>$44,989</td>
<td>$23,868</td>
<td>$69,911</td>
</tr>
<tr>
<td>Herbal Tea</td>
<td>Chamomile</td>
<td>$14,434</td>
<td>$36,570</td>
<td>$765</td>
<td>$2,194</td>
<td>$3,180</td>
<td>$11,186</td>
<td>$8,852</td>
<td>$25,632</td>
</tr>
<tr>
<td></td>
<td>Lemon</td>
<td>$6,251</td>
<td>$21,978</td>
<td>$7,901</td>
<td>$27,176</td>
<td>$2,593</td>
<td>$14,497</td>
<td>$13,120</td>
<td>$32,274</td>
</tr>
<tr>
<td></td>
<td>Mint</td>
<td>$4,069</td>
<td>$9,337</td>
<td>($2,242)</td>
<td>$11,992</td>
<td></td>
<td></td>
<td>$4,330</td>
<td>$14,380</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>$24,754</td>
<td>$67,885</td>
<td>$6,424</td>
<td>$41,362</td>
<td>$5,774</td>
<td>$25,683</td>
<td>$26,301</td>
<td>$72,285</td>
</tr>
<tr>
<td>Tea</td>
<td>Darjeeling</td>
<td>$10,772</td>
<td>$30,289</td>
<td>$6,497</td>
<td>$14,096</td>
<td></td>
<td></td>
<td>$11,780</td>
<td>$28,769</td>
</tr>
<tr>
<td></td>
<td>Earl Grey</td>
<td>$10,331</td>
<td>$32,881</td>
<td>$3,405</td>
<td>$6,505</td>
<td></td>
<td></td>
<td>$10,425</td>
<td>$27,387</td>
</tr>
<tr>
<td></td>
<td>Green Tea</td>
<td>$1,227</td>
<td>$5,211</td>
<td>$5,654</td>
<td>$11,571</td>
<td></td>
<td></td>
<td>($7,109)</td>
<td>$16,063</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>$22,330</td>
<td>$68,380</td>
<td>$15,557</td>
<td>$32,172</td>
<td></td>
<td></td>
<td>$15,097</td>
<td>$72,220</td>
</tr>
</tbody>
</table>
### Who has the best profit and who has the worst sales?

<table>
<thead>
<tr>
<th>Product Type</th>
<th>Product</th>
<th>Central Sum of Sales</th>
<th>East Sum of Sales</th>
<th>South Sum of Sales</th>
<th>West Sum of Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>Amaretto</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Columbian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decaf Irish Cream</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Espresso</td>
<td>Caffe Latte</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caffe Mocha</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decaf Espresso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regular Espresso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herbal Tea</td>
<td>Chamomile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mint</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tea</td>
<td>Darjeeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Earl Grey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Green Tea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The sum of sales is categorized into ranges: 0K, 20K, 40K, 60K.
Do The Right Analytics, Don’t Just Visualize Data
Doubling down on states for strong growth

Maria
Senior Sales Analyst
March 15th, 2012
Today’s question

In which states should we invest additional marketing dollars during the upcoming campaign?

Based upon sales growth potential...
Sales per State/Capita

2011 sales per million residents by state (top/bottom 3 labeled)

Top sales states are quite low in sales per million people! Great potential!
Potential sales by state???

+ Is there a better metric?
+ The emphasis is on potential

Average sale per capita for top states multiplied by Current population of top sales states
Sales per State/Capita x Capita

2011 sales per state (top/bottom 3 labeled)

Top - TX at $467,644 11.6%

ID at $697 0.0%
Highest growth potential in top 8

If we were to pick just one state, California has the greatest potential.

The next tier is Texas, New York & Florida.
Useful metrics

1. Total sales per state was OK

2. Better: Total sales per million residents per capita is better than looking at existing customers, because we want new customers

3. Top five states to target: 90\textsuperscript{th} percentile +