Why Visual Analytics?

Big Data

12+ TBs of tweet data every day

30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones worldwide

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

Big Data → Actionable Intelligence → Profitability
PROBLEMS WITH SCALABILITY

Must be scalable to

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

*Visual Analytics comes to the rescue...*
The Goal of Visualization

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

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)

Interaction

VA

Analytics

Visualization
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
EXAMPLE #2

In this video you will do some counting.

It is very important that you get the right number!

Ready?

watch video (YouTube)

Video by Dan Simons (U Illinois)
Another distraction experiment

watch video (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
Another deficiency of humans

- humans tend to stick with an “opinion” for a long time
- how long does it take you to switch?
Human Limitations

The Magic Number Seven

- $\pm 2$: the number of things most people can keep in working memory at one time
- causes problems for complicated analysis

An excellent book that has more on this topic
- discussed next
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 one’s head and down on paper or on a computer screen in some simplified form
- shows 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,...
Support visualization with computations for data processing
Form a loop: visualize - refine
Gather (forage) information
Re-represent
  ▪ choose form that aids analysis
Develop insight
  ▪ through manipulation of representation
Produce results
  ▪ “product”
Nominal Sense-Making Process
Elemental artifacts
  - source intelligence, evidence, assumptions

Pattern artifacts
  - relationships, temporal and spatial structure

Higher-order knowledge constructs
  - arguments, causality, models

Complex reasoning constructs
  - hypotheses, scenarios

*All these become part of the Visual Analytics sense-making (reasoning) process*
**Standard Information Displays**

Showing about 50 - 300 data values; 10-20 dimensions

Examples from the VisualInsights WebPage
Drill down into certain aspects of the data
  - essentially isolates a subset of the data
  - for multivariate data with coordinated views (dashboards)

Crossfilter
  - see github https://square.github.io/crossfilter/
  - fast multidimensional filtering for coordinated views
  - even with datasets containing a million or more records;
Cross-Filter Example

See here for the animation: https://discourse.metabase.com/t/cross-filtering/3466
Use Visualizations to Evoke The Right Thoughts
How Many 9s Do You See?

| 3 | 3 | 0 | 3 | 0 | 1 | 8 | 7 | 6 | 8 | 2 | 1 | 4 | 0 | 3 | 8 | 3 | 7 | 7 | 2 | 0 | 5 | 2 | 3 | 2 | 7 | 0 | 2 | 0 |
| 7 | 1 | 4 | 6 | 0 | 2 | 1 | 3 | 2 | 7 | 6 | 0 | 2 | 5 | 3 | 6 | 2 | 5 | 7 | 6 | 3 | 3 | 0 | 2 | 0 | 3 | 0 | 7 | 2 |
| 8 | 7 | 5 | 7 | 2 | 8 | 3 | 8 | 7 | 7 | 8 | 2 | 0 | 7 | 7 | 5 | 2 | 3 | 1 | 1 | 5 | 6 | 3 | 8 | 4 | 7 | 8 | 2 | 0 |
| 0 | 5 | 0 | 5 | 1 | 6 | 1 | 7 | 5 | 6 | 8 | 0 | 4 | 4 | 6 | 7 | 4 | 7 | 1 | 4 | 0 | 0 | 8 | 4 | 4 | 3 | 0 | 3 | 2 |
| 2 | 4 | 3 | 1 | 3 | 5 | 4 | 9 | 5 | 0 | 7 | 6 | 0 | 7 | 4 | 3 | 1 | 8 | 2 | 7 | 3 | 4 | 6 | 0 | 2 | 4 | 8 | 2 | 3 |
| 8 | 7 | 7 | 2 | 6 | 5 | 4 | 6 | 7 | 0 | 7 | 6 | 0 | 0 | 3 | 9 | 0 | 2 | 4 | 7 | 1 | 7 | 2 | 3 | 3 | 5 | 8 | 7 | 0 |
| 0 | 9 | 8 | 4 | 1 | 3 | 1 | 7 | 6 | 4 | 5 | 4 | 1 | 2 | 4 | 5 | 3 | 3 | 5 | 4 | 9 | 6 | 7 | 7 | 6 | 3 | 4 | 2 | 5 |
| 4 | 7 | 7 | 0 | 2 | 2 | 0 | 1 | 1 | 7 | 7 | 7 | 0 | 2 | 6 | 6 | 4 | 7 | 5 | 8 | 6 | 1 | 4 | 3 | 7 | 8 | 5 | 4 | 6 |
| 4 | 3 | 6 | 6 | 4 | 6 | 6 | 2 | 8 | 4 | 8 | 5 | 3 | 7 | 8 | 8 | 1 | 3 | 8 | 5 | 4 | 5 | 7 | 4 | 0 | 3 | 2 | 8 | 4 |
| 5 | 5 | 0 | 3 | 5 | 3 | 5 | 3 | 8 | 3 | 2 | 3 | 8 | 2 | 3 | 1 | 6 | 2 | 7 | 2 | 4 | 6 | 3 | 6 | 4 | 4 | 3 | 2 | 5 |
| 4 | 4 | 0 | 2 | 1 | 7 | 2 | 4 | 4 | 7 | 4 | 1 | 9 | 2 | 4 | 5 | 2 | 5 | 0 | 4 | 0 | 0 | 5 | 3 | 6 | 3 | 3 | 6 | 7 |
| 7 | 4 | 6 | 6 | 8 | 7 | 5 | 7 | 9 | 2 | 0 | 2 | 8 | 8 | 8 | 3 | 2 | 4 | 2 | 6 | 4 | 0 | 4 | 6 | 3 | 7 | 2 | 1 |
| 0 | 1 | 7 | 1 | 5 | 9 | 1 | 4 | 2 | 8 | 7 | 3 | 7 | 1 | 4 | 5 | 1 | 8 | 7 | 8 | 0 | 5 | 1 | 7 | 0 | 5 | 8 | 8 | 1 |
| 2 | 8 | 5 | 2 | 1 | 2 | 8 | 7 | 7 | 6 | 2 | 5 | 6 | 2 | 6 | 4 | 1 | 5 | 1 | 6 | 1 | 2 | 1 | 1 | 0 | 5 | 6 | 4 | 0 |
| 2 | 1 | 1 | 7 | 2 | 0 | 0 | 1 | 8 | 7 | 0 | 2 | 9 | 0 | 2 | 8 | 5 | 7 | 8 | 4 | 6 | 0 | 6 | 5 | 0 | 7 | 1 | 2 |
| 0 | 5 | 2 | 4 | 1 | 5 | 3 | 3 | 1 | 5 | 5 | 1 | 4 | 0 | 1 | 6 | 4 | 3 | 3 | 9 | 8 | 8 | 8 | 3 | 4 | 6 | 8 | 4 | 8 | 6 |
| 7 | 3 | 7 | 5 | 2 | 4 | 0 | 2 | 7 | 6 | 3 | 8 | 5 | 5 | 4 | 5 | 8 | 8 | 7 | 5 | 5 | 6 | 5 | 6 | 7 | 9 | 7 | 7 | 4 |
| 0 | 3 | 2 | 8 | 1 | 4 | 4 | 6 | 0 | 8 | 2 | 3 | 0 | 1 | 3 | 4 | 6 | 2 | 0 | 5 | 7 | 3 | 6 | 1 | 8 | 7 | 3 | 5 |
| 4 | 4 | 8 | 3 | 3 | 3 | 5 | 0 | 1 | 0 | 3 | 8 | 6 | 3 | 2 | 0 | 5 | 0 | 6 | 1 | 3 | 3 | 4 | 3 | 6 | 1 | 5 | 8 | 6 |
| 1 | 0 | 2 | 2 | 7 | 6 | 3 | 3 | 0 | 8 | 8 | 0 | 3 | 1 | 8 | 8 | 1 | 2 | 1 | 7 | 5 | 2 | 9 | 3 | 5 | 8 | 3 | 2 | 5 |
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</th>
<th></th>
<th>East</th>
<th></th>
<th>South</th>
<th></th>
<th>West</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sum of Profit</td>
<td>Sum of Sales</td>
<td>Sum of Profit</td>
<td>Sum of Sales</td>
<td>Sum of Profit</td>
<td>Sum of Sales</td>
<td>Sum of Profit</td>
<td>Sum of Sales</td>
</tr>
<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>$11,253</td>
<td>$30,357</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>$5,201</td>
<td>$14,163</td>
<td>$4,064</td>
<td>$18,876</td>
</tr>
<tr>
<td></td>
<td>Caffe Mocha</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>Decaf Espresso</td>
<td>$8,660</td>
<td>$24,185</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>$10,062</td>
<td>$24,036</td>
<td>$10,062</td>
<td>$24,036</td>
<td>$10,062</td>
<td>$24,036</td>
</tr>
<tr>
<td>Total</td>
<td></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>$4,330</td>
<td>$14,380</td>
<td>$4,330</td>
<td>$14,380</td>
</tr>
<tr>
<td>Total</td>
<td></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>$11,780</td>
<td>$28,769</td>
<td>$10,425</td>
<td>$27,387</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>$10,425</td>
<td>$27,387</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>($7,109)</td>
<td>$16,063</td>
<td>$15,097</td>
<td>$72,220</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$22,330</td>
<td>$68,380</td>
<td>$15,557</td>
<td>$32,172</td>
<td>$15,097</td>
<td>$72,220</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</th>
<th>East</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0K</td>
<td>20K</td>
<td>40K</td>
<td>60K</td>
</tr>
<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>

Sum of Sales:
- Central: 0K, 20K, 40K, 60K
- East: 0K, 20K, 40K, 60K
- South: 0K, 20K, 40K, 60K
- West: 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 spend during the upcoming campaign?

Based upon sales growth potential...
2011 sales by state

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

Top sales states are quite low in sales per million people! Great potential!
The top 3 states generate 30% of sales. The top 8 states generate 50%. And the top 21 states generate 80%.
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*
What are the top states based on sales per capita?

2011 sales per state (top/bottom 3 labeled)

ID at $697 0.0%

Top- TX at $467,644 11.6%
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 +
Dashboards should pass the 5-second test
Important rules:

- most important view goes on top or top left
- legends go near their views
- avoid using multiple color schemes
- use 5 views or fewer
- provide interactivity