Case Studies

CS 4460/7450 - Information Visualization
Jan. 15, 2009
John Stasko

Agenda

• Survey some example projects to get a flavor for the kinds of work done in infovis
• Just some favorites of mine from over the years
For Each Study...

- **Who** – Who are the target users of the tool/technique?
- **Problem** – What is the main problem or challenge being addressed?
- **Data** – What data are being used in the project
- **Solution** – What was done and what is interesting and unique about it?

---

**Study 1**

- Understand patterns of presence/events over time
- Focus: People’s presence/movements in some space
- Situation:
  - Workers punch in and punch out of a factory
  - Want to understand the presence patterns over a calendar year
- Alternate: Power plant electricity usage over a year
**Particulars**

- Who – Factory boss/manager
- Problem – Show this large amount of data in an easily understandable and query-able manner
- Data – Punch in/out times for workers

**Ideas**

- Any ideas on what we could do here?
One Idea

Good
- Typical daily pattern
- Seasonal trends

Bad
- Weekly pattern
- Details

Approach Taken

- Cluster analysis
  - Find two most similar days, make into one new composite
  - Keep repeating until some preset number left or some condition met

- How can this be visualized?
  - Ideas?
Display

Characteristics

- Unique types of days (individual or cluster) get their own color
- Contextually placed in calendar and line graph for it is shown
- Stop clustering when a threshold met or at a predetermined number of clusters
**Interaction**

- Click on day, see its graph
- Select a day, see similar ones
- Add/remove clusters

**Insights**

- Traditional office hours followed
- Most employees present in late morning
- Fewer people are present on summer Fridays
- Just a few people work holidays
- When the holidays occurred
  - School vacations occurred May 3-11, Oct 11-19, Dec 21-31
- Many people take off day after holiday
- Many people leave at 4pm on December 5
**Study 2**

Weaver et al

*Information Visualization '07*

- Understand spatial and temporal patterns of movements of people and groups of people, understand commerce and travel patterns
- Focus: Social networks of people and historical data
- Alternate: Understand actions and movements of people in terrorist groups

---

**Particulars**

- **Who** – Historical geographers, genealogists
- **Problem** – Provide platform where these people can do research and more easily understand the movement patterns
- **Data** – Old hotel records & registers
**Data Collection**

- Two hotel registers in PA
  - ~10k guest entries
  - Hundreds of place of origin (mostly NE)
- Transcribed into spreadsheets (400 hours)
- Many repeat visitors
- 6% of dataset not legible/usable

**Preliminary Data Analysis**

- Simple tables generated
- GIS software for mapping geographic locations
**Improvise**

- System used to build visualization tool
- Modular library of visualization components
- Multiple views key
- Coordination model coupled with data processing language at its heart

- More later in term...

---

**Visualization**

- How?
  - Consult with historical geographers
  - Iterative design

- Multiple, coordinated views
User interface

Looking for patterns of visitation
Showing time of year of visits

Video
Data Analysis

• Through interaction with views, analysts can pose queries and look for answers
• “For these people on this date, did they ever meet prior to that?”
• Can also ask questions for unknown people (match a ?)

Data Attributes Explored

• Weekly circuits
• Weather effects
• Biweekly circuits
• Particular hotel characteristics
• Groups & gatherings
• Non-periodic regular visits
Study 3

- DateLens
- Helping people better manage their calendars and appointments on a handheld display
- Uses technique called a “fisheye view” we will learn about later in term

Particulars

- Who – Everyday people
- Problem – How to show a potentially large amount of appointment information in a small number of screen pixels (and allow flexibility for different tasks)
- Data – Set of appointments
**Premise**

- At different points in time, you want different perspective on your appts.
  - See how my month looks
  - What’s happening later this week
  - Am I double-booked this afternoon

**Technique**

- Adopts fisheye view technique
  - Focus item(s) shown in more detail while context still visible, but simplified
- Interaction is key with smooth transitions
Different Perspectives

Month view  Zooming to a week  Zooming to a day

Video
Reactions?

• Thoughts and impressions?

Study 4

• WireVis
• Helping Bank of America examine wire transfers of money
• Want to detect fraud and illegal actions
Particulars

- **Who** – Bank analysts
- **Problem** – Detect money laundering and fraud in wire transfers of money
- **Data** – Electronic records of wire transactions and information associated with each

---

Background

- **Wire transfers of money can be complex**
  - Have a “from” and “to” but often many “middlemen”
  - May not know who intermediaries are
- **Millions of transfers per day occur**
  - Vast majority are legal
- **Bank has legal responsibility to report suspicious activities**
Data

- Each transaction:
  - Money amount
  - Payer (could be third party)
  - Payee (could be an agent)
  - Potential intermediaries
  - Addresses of payer and payee, instructions, additional comments are optional

Challenges

No Standard Form...
When a wire leaves Bank of America in Charlotte...
The recipient can appear as if receiving at London, Indonesia or Singapore
Vice versa, if receiving from Indonesia to Charlotte
The sender can appear as if originating from London, Singapore, or Indonesia
Challenges

- Scale: BoA may do 200k transfer per day
- No international standard: loosely structured data
- Bad guys are smart and one step ahead
  - Detection tools are always reactive

Existing Detection

- Examine for certain temporal patterns of activity
- Look for keywords in free text
  - Filter transactions based on these highly secretive words
  - Typically a few hundred
  - Updated based on intelligence reports
Current Practices

- Load transactions into large relational DB
- Download some amount to spreadsheets via filtering based on keywords, amounts, dates, ...
- Can only look at a week or two this way
- Difficult to notice temporal patterns

System Overview

- Heatmap View (Accounts to Keywords Relationship)
- Search by Example (Find Similar Accounts)
- Keyword Network (Keyword Relationships)
- Strings and Beads (Relationships over Time)
**Heatmap View**

- List of Keywords
- Sorted by frequency from high to low (left to right)
- Hierarchical clusters of accounts
- Sorted by activities from big companies to individuals (top to bottom)
- Fast “binning” that takes $O(3n)$
- Number of occurrences of keywords
- Light color indicates few occurrences

**Strings and Beads**

- Each string corresponds to a cluster of accounts in the Heatmap view
- Each bead represents a day
- Y-axis can be amounts, number of transactions, etc.
- Fixed or logarithmic scale
- Time
Keyword Network

- Each dot is a keyword
- Position of the keyword is based on their relationships
  - Keywords close to each other appear together more frequently
  - Using a spring network, keywords in the center are the most frequently occurring keyword
- Link between keywords denote co-occurrence

Search By Example

- Target Account
- Histogram depicts the occurrences of keywords
- User interactively selects features within the histogram used in comparison
- Accounts that are within the similarity threshold appear ranked (most similar on top)
- Similarity threshold slider
Study 5

- PeopleGarden
- Visualization technique for portraying online interaction environments (Virtual Communities)
- Provides both individual and societal views
- Utilizes garden and flower metaphors
**Particulars**

- **Who** – Anyone visiting online community
- **Problem** – Help someone gain a more rapid understanding of the community as a whole and the individual participants
- **Data** – Postings from past

**Data Portrait: Petals**

Fundamental view of an individual

His/Her postings are represented as petals of the flower, arranged by time in a clockwise direction.
**Data Portrait: Postings**

**Time of Posting**

New posts are added to the right
Slide everything back so it stays symmetric
Each petal fades over time showing time since posting
A marked difference in saturation of adjacent petals denotes a gap in posting

**Data Portrait: Responses**

**Response to posting**

Small circle drawn on top of a posting to represent each follow-up response
**Data Portrait: Color**

Initial post vs. reply

Color can represent original/reply
Here magenta is original post, blue is reply

---

**Garden**

Combine many portraits to make a garden

Message board with 1200 postings over 2 months

Each flower is a different user
Height indicates length of time at the board
Alternate Garden View

Sorted by number of postings

Interpreting Displays

Group with one dominating person

More democratic group
Thoughts

• Is it an effective visualization technique?
• Likes/dislikes?

Patterns & Themes

• Any recurring patterns or themes evident from these five case studies?
**Administrativa**

- HW 2 out today, due Tuesday 27th

**Upcoming**

- **Visual Perception**
  - Reading:
    Stone, web article

- **Cognitive Issues**
  - Readings:
    Amar et al
    Liu et al
Bookkeeping

- Survey forms
- Pictures!