Course Overview

- **Focus on Innovation in Visualization**
  - Expressing (Abstract) Data visually
  - Awareness of cultural conventions

- **Software** production work in:
  - MySQL for data query
  - java-based Processing

- **6 Projects:**
  - Data Query
  - 2D Matrix
  - 2D Re-orderable Matrix
  - 3D Spatialization
  - 3D Change over Time
  - Correlation between 2 DataSets
Course Goals & Methods

- Advance skills in:
  - Asking **innovative questions** about a dataset thru data queries
  - Exploring Data Aggregation thru algorithms
  - Visual language & syntax

- **Our approach**: Multi-disciplinary, experimental

Course Knowledge Acquisition

1) how to identify and **retrieve significant data** from a dataset with MySQL
2) Develop skills in the **fundamentals of visual language** through programming
3) Visualize abstract data to **reveal patterns** and relationships
4) **Normalize data** to enhance legibility and coherence
5) Implement interactivity within 3D volumetric visualization
6) **Correlate** 2 sets of data from diverse sources
7) **Visual Language**: How form, color, space, timing, movement, etc. impact on content (this is the primary goal of the course)

Course Assignments

Every project has **conceptual, technical** and **aesthetic** challenges:

1. MySQL Data discovery
   (What is an interesting query?)
2. 2D Visualization
   (3 or more data properties)
3. 2D Re-orderable Matrix
   (How can a 2D matrix be re-ordered to create greater clarity?)
4. 3D Spatialization
   (What does 3D contribute)
5. Interactivity
   (What does interactivity contribute)
6. Correlation
   (Correlation between 2 different datasets provide new insight)

An Interdisciplinary Process

We want to integrate diverse Expertises:

- **Analytical**: theoretical, cultural, information research (social scientist)
- **Technical**: Statistical aggregation and computational processes (engineer, scientific)
- **Aesthetic**: Visual design, expression (artist, designer, architect)
Data Visualization Function & Situation

"Visualizations Give Meaning to Information"

- Data Visualization is the **study** and **production** of visually representing data
- An active field of research (IEEE VisWeek and other conferences)
- Visualization is **rule-based**, and **culturally influenced**
- Data may be **abstract** (numeric, symbolic), **textual**, or **iconic**

Course Format

- **TUES**: Lectures, visualization analysis
- **THUR**: Technical lab

- **BLOG**: Post your concepts, sketches, share algorithms, visualization techniques here
- **WEBSITE**: Course syllabus, code samples, references, and project results

Resources: Software

- **MySQL** (the database)
- **MySQL Workbench** (access to data)
- **Processing** (Java-based scripting language used by graphic designers)
- **JSON** for data correlation

All Projects to use the Same Data Source

- To allow us to quickly master examples and share solutions
- Everyone uses the same data so results can be **compared** and **shared**
- Learn quickly from previous projects
Data is Multivariate

- Over 70 million dataset in database, acquired hourly since 2005
- Data is multivariate. Each transaction includes numeric, ordinal, interval scale (time, date), string, and other classification data of objects retrieved from collection:
  - **ItemNumber**: Collection acquisition time-stamp
  - **bibNumber**: Each topic-specific item in collection
  - **Barcode**: Each item has a unique rfid sticker
  - **Check-out/check-in hour/day**: In/out interaction with database
  - **ItemType**: books, cds, dvds, music sheets, etc.
  - **Title**
  - **Dewey Classification**
  - **Subjects**: Keywords

Data as “Cultural Content”

- Patrons check out books, cds, dvds from the Seattle Public Library
- A **public resource** (can be mined for a broad range of cultural trends)
- Each time someone checks out a movie, book, cd, data is received hourly
- Appx 30000 per day; 10 million annual;
- **Over 70 million datasets** since September 2005
- Can be correlated with library site or other sources: NYTimes, etc.

MySQL

- **MySQL**: Open-source relational database (Structured Query Language)
- **Industry standard**
- MySQL exercises to develop skills in retrieving meaningful information
Wk 1, 2, 3: Data Mining OR Knowledge Discovery

- The most time intensive effort in data mining is **understanding your data** before any analysis or visualization can take place.
- Discovery: detecting something new or relevant
- Unusual records (anomaly detection)
- Dependencies (associative relationships)

MySQL: Select itemNumber, cout, collcode, itemtype, barcode, title, callNumber, deweyClass, subj from inraw where year(cout) = 2007 and month(cout) >= 1 and month(cout) <= 4 limit 50;

Your Contribution

- Data is not content. What you do with it is the content!
- You choose what to feature from the dataset based on your interests
- You introduce data processing methods
- You select algorithms to implement
- You make design decisions
- You determine “look and feel” which becomes the content

Wk 4  2D Visualization / Mapping
Frequency Maps Can Tell a Story (catalogtree)

DIPLOMATIC PARKING VIOLATIONS

Bio-Rhythm: Frequency Map reveals pattern
TreeMap: SmartMoney (Schneiderman/Wattenberg)

Wk5: Bertin: Reorderable Matrix
MySQL & Processing allow for a multiplicity and complex modes of data correlation with multiple sources such as:

- Twitter
- NY Times
- Etc.
- JSON: Data format standard
MySQL & Processing allow for a multiplicity and complex modes of data correlation with multiple sources such as:

- Twitter
- NY Times
- Etc.

- JSON: Data format standard
Frequency Map: What is Involved?

- **Data Correlation**: MySQL & Processing allow for a multiplicity and complex modes of data correlation with multiple sources such as:
  - Twitter
  - NY Times
  - Amazon, etc.

Data Processing Functions

- **Validation**: Ensuring that data is “clean, correct and useful”
- **Sorting**: Arranging items in some sequence and/or in different sets
- **Summarization**: Reducing detail data to its main points
- **Aggregation**: Combining multiple pieces of data (possibly from various sources)
- **Analysis**: Collection, organization, analysis, interpretation and presentation of data
- **Reporting**: List detail or summary data or computed information

Integrate Your Expertise

- **Computer Science**: Integrate complex algorithms to visualization
- **Statistics**: Implement statistical probability problems to data analysis and visualization
- **Sound/Signal processing**: Consider data as signal and explore translation between sonic, signal and visual patterns
- **Social Science**: Identify cultural patterns, changes, transformations
- **Geography**: Explore spatial mapping
- **Cinematic/Literary**: Explore data pattern as narrative development
Your Contribution

*Data is not content. What you do with it is the content!*

- You choose what to feature from the dataset based on your interests
- You introduce data processing methods
- You select algorithms to implement
- You make design decisions
- You determine “look and feel” which becomes the content

Additional Directions

- **Time-based animations** (change over time)
- **Scientific Animation** (visual simulation of phenomena)
- **Fluid Animation** (data change simulating natural flow)