M259 Visualizing Information

Elings 2611 Tues-Thurs 3:30-5:20pm

George Legrady, legrady@mat.ucsb.edu
Weidi Zhang, weidizhang@ucsb.edu

Course Overview

A ten-week comprehensive course focused on techniques for data visualization for Data Science

- Data analytics for data mining and knowledge discovery
- Multivariate dataset
- Frequency mapping, pattern
- Exploration of algorithms for data clustering
- 3D interactive visualization projects in java-based Processing and p5.js

Course Knowledge Acquisition

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

Course Schedule

Wk 1-2 Data Analytics, Knowledge Discovery, Content Analysis with MySQL

Discover unexpected, interesting patterns, anomalies in a large dataset

Wk 3-6 3D Interactive Visualization, Frequency Data Mapping - Visual Language Basics

Acquire visual language basics, visualize data in java-based Processing and p5, design in 3D-interactive space/time, implement associative rule-mining and other algorithms

Wk 7-10 Student Defined Project

Student defines a project with their own data source, a project that builds on skills acquired through the two previous assignments

Course Assignments

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

- **MySQL Data discovery**

  (What is an interesting query?)

- **3D Spatialization**

  (Interactive, spatial visualization)

- **Correlation and Student Defined Project**

  (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*)

Course Format

Tues: Lectures, visualization analysis

Thur: Technical lab

Course Resources

WEBSITE: Course syllabus, code samples, references, and project results

BLOG: Post your concepts, sketches, share algorithms, visualization techniques here

Previous Student Reports: Projects with documentation

Software Resources

- **MySQL** open-source relational database management system
- **MySQL Workbench** to access the data
- **Processing** Java-based scripting language used by graphic designers
- **JSON** a text-based data-interchange format for data correlation
- **P5** a JavaScript library for posting Processing-based on the internet

Wk1-6, 7-10 Data Source

Everyone uses the same data for the MySQL, & 3D assignments so that

- Results can be compared and shared
- Learn quickly from previous projects

Wk7-10 Each student chooses their data source for the final project

Data Science - Work Effort

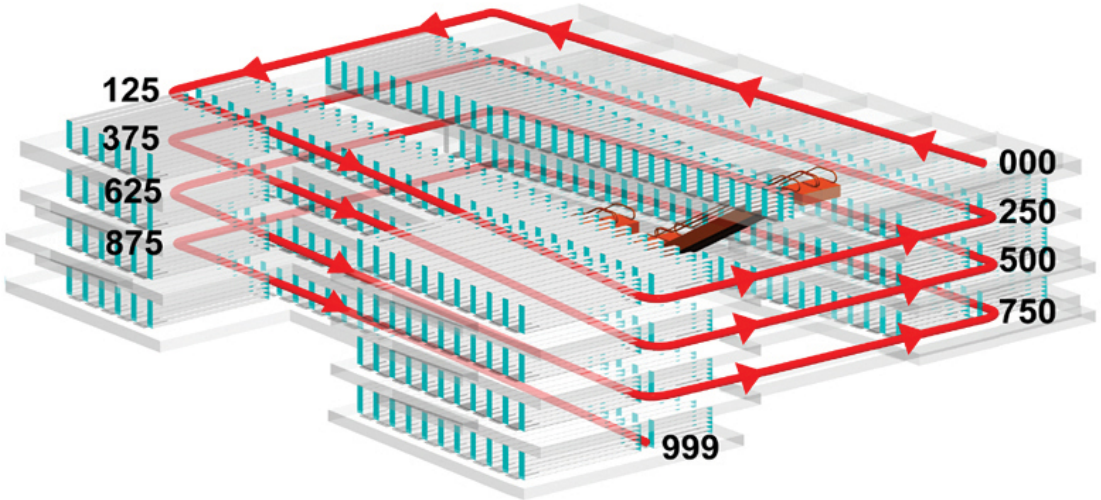
- Collecting the data (**19%**)
- Cleaning and organizing data (**60%**)
  - Most data sets contain noise!!

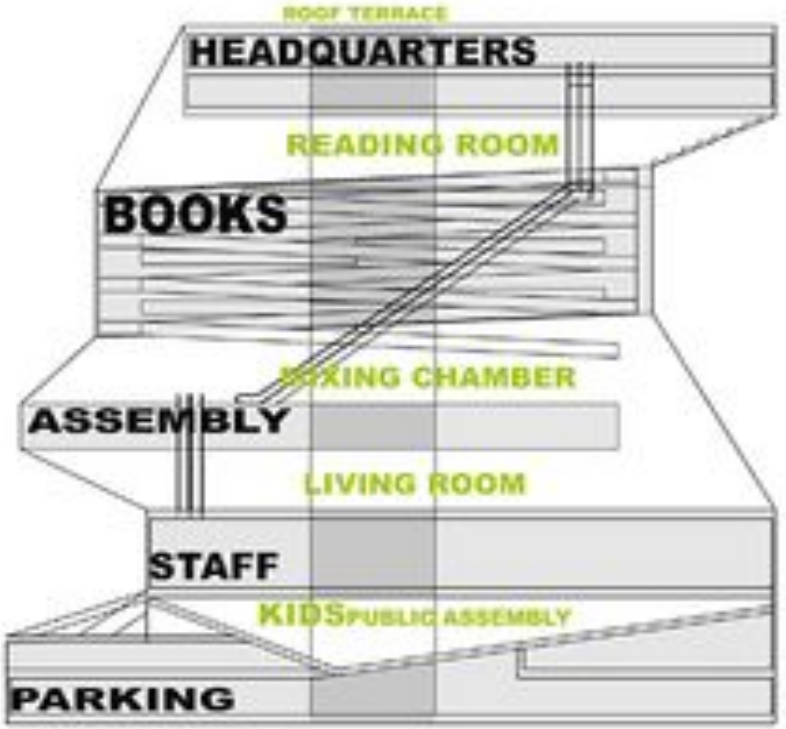
- Building training sets (**3%**)
- Mining data for patterns (**9%**)
- Refining algorithms (**4%**)

Source: *Data Science*, Kelleher, Tierney, MIT Press

Visualization Follows

Architectural Design





Data Source: Seattle Central Library

- Approx. 20,000 – 36,000 checkouts per day
- Over 90 million transactions to-date
- Multi-variate data

- 4 animations, visualizing checkout activities of the past hour
- Located behind Information Desk in the Mixing Chamber
- Possibly the longest running, dynamic data visualization artwork (2005-2019)
- Project collects data over time. Used in the M259 course


librarian

Total Items
2005.09.12 16:43:40

12433 Total Items in circulation today — Past hour

2088

Dewey Items
2005.09.12 16:43:40

5116 Dewey Items in circulation today — Past hour

887

Non-Dewey Items
2005.09.12 16:43:40

7317 Non-Dewey Items in circulation today — Past hour

1201

Books
2005.09.12 16:43:40

6881 Books in circulation today — Past hour

1152

DVDs
2005.09.12 16:43:40

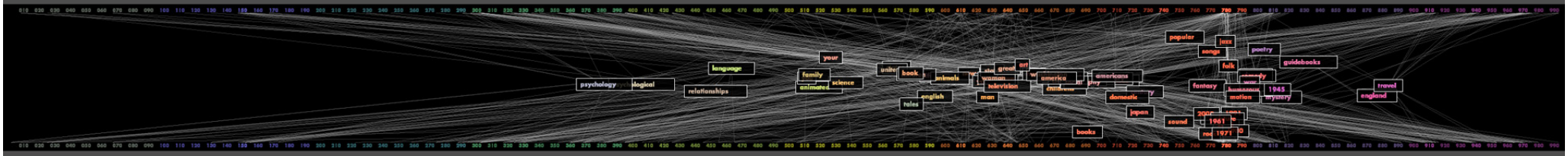
1932 DVDs in circulation today — Past hour

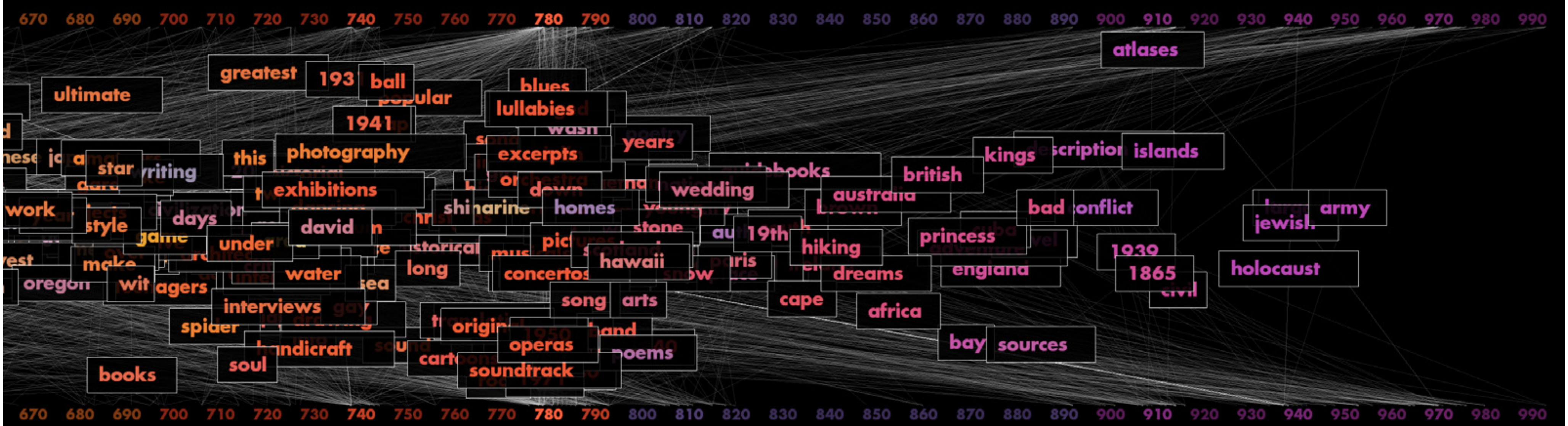
353

Other Media (cd, video, etc.)
2005.09.12 16:43:40

1226 Other Media in circulation today — Past hour

201




670 680 690 700 710 720 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970 980 990
atlases
ultimate
greatest
193
ball
blues
lullabies
1941
years
excerpts
photography
this
star
kings
islands
british
exhibitions
wedding
work
homes
david
days
army
bad
princess
hiking
hawaii
under
long
water
concertos
dreams
england
holocaust
1865
oregon
agers
interviews
song
arts
cape
africa
books
soul
operas
soundtrack
bay
sources

Data as Cultural Analysis

- Patrons **check-out** books, cds, dvds from the Seattle Public Library
- Each time someone checks out a movie, book, cd, a cultural choice is made
- Servers receives data by the hour
- Appx 20000 per day; 10 million annually
- Over 90 million datasets since September 2005
- Data a significant resource: Can be mined for a broad range of cultural trends or research
- Can be correlated with other data sources: NYTimes, Apple iTunes, Amazon, etc.


librarian

Integrate Your Expertise

Computer Science: Integrate complex algorithms for parsing and visualizing data

Statistics: Implement statistical probability problems to data analysis and visualization

Social Science: Identify cultural patterns, changes, transformations

Geography: Explore spatial mapping

Cinematic/Literary: Explore data pattern as narrative development

Science: Visually map data collected through sensors, of phenomena, etc.

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 also shapes the overall content

VISWEEK 2012, Seattle

Copyright & Research Use of Seattle Library Data

You can use the Seattle data for research and publication.

You can circulate your visualization but because of proprietary reasons you cannot circulate the data itself.

If you publish a paper, make sure to credit the course and the project.

Also sign the intellectual property agreement form at the course website

NEXT: Seattle Public Library Data Description