

M259 Visualizing Information

Jan 14: DATA SOURCE

George Legrady, legrady@mat.ucsb.edu
Yoon Chung Han <hanyoonjung@gmail.com>

This Week:

TUES: Jan 14

Discuss Data Source

THUR: Jan 16

- a) Karl Yerkes Presentation on database structure
- b) Yoon will follow with SQL examples

2 Directions for Data Analysis:

1. The study of the **data organization**, its anomalies, history, etc.
2. The **exploration of content** (cultural information)

Data Visualization Focus Areas:

DATA SYSTEMS

- Data retrieval and analysis
- System anomaly analysis

CULTURE

- Cultural patterns:
- Historical Context: How has the library transformed itself

VISUALIZATIONS

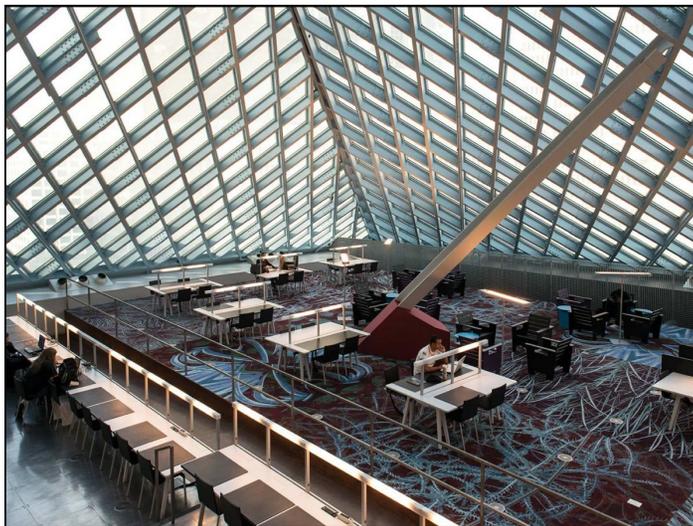
- What are ways to visualize data (content focus)
- Invent new visualizations (exploration of the visual language)

ALGORITHMS

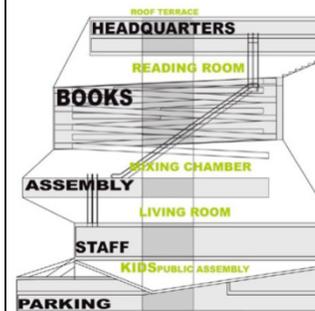
- Apply Algorithms: What are interesting algorithms by which to process data
- Create new algorithms

Data Source

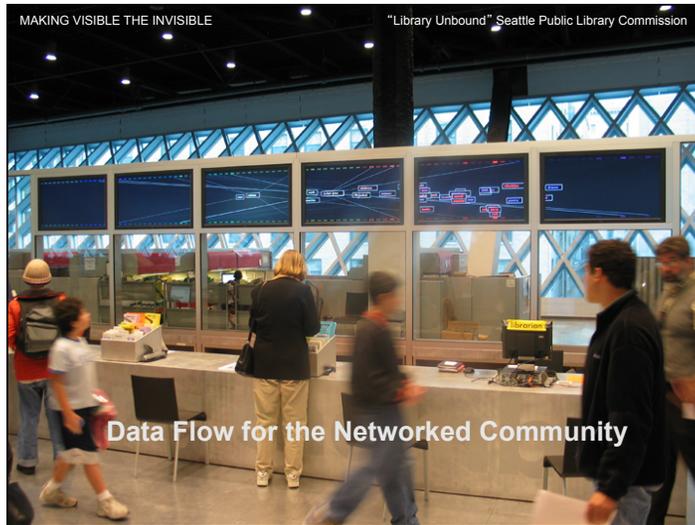
- Patrons check out books, cds, dvds from the Seattle Public Library
- Each time someone checks out a movie, book, cd, data is received by the hour
- Appx 30000 per day; 10 million annual;
- **Over 67 million datasets** since September 2005



Rem Koolhaas Visionary Architecture



- **Seattle Public Library** main collection of books, government publications, periodicals, audio visual materials and the technology to access and distribute information online.
- The building is divided into eight horizontal layers, each varying in size to fit its function.



M259 Visualizing Information George Legrady 2014 Winter

"Making Visible the Invisible" ArtWork

- Seattle Public Library artwork opened Sept. 7, 2005 extended to 2019
- Provides raw data for the study of cultural library patron practices over time
- Possibly the longest running public dynamic data visualization project

M259 Visualizing Information George Legrady 2014 Winter

Project Concept

- Conceptualize the library as a "Data Exchange Center"
- Correlation is made between the flow of data (books, DVD, etc.) leaving the library and
- What patrons considers interesting information at any specific time
- This data is information that can be calculated mathematically and represented visually

M259 Visualizing Information George Legrady 2014 Winter

Rem Koolhaas Concept

[HTTP://WWW.SPL.COM/~MIXING CHAMBER](http://www.spl.com/~mixing-chamber)

Data Source

- Patrons check out books, cds, dvds from the Seattle Public Library
- Each time someone checks out a movie, book, cd, data is received by the hour
- Appx 30000 per day; 10 million annual;
- Over 67 million datasets** since September 2005

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;

itemNumber	year	cout	collcode	itemtype	barcode	title	callNumber	deweyClass	subj
124949	2007-01-02	09:30:00	naft	audi	000000201414	Right to speak	CD 782.41542 C26517	782.41542	Book Music 2007-01-02-Query music 2001-2010
126436	2007-01-02	09:12:00	cafc	audi	000004073908	Heart on the Line	FC ANNO2	761.1	Love stories/Motion picture producers and directors New York State N.
126494	2007-01-02	09:00:00	mgfc	audi	000000720208	You had to be there	FC LONEL	761.1	Love stories/Large tape books/Research His Fiction
264403	2007-01-02	09:10:00	cafr	audi	000000720207	Governing public utilities new times new requirements	371.2 DANBIB 1992	371.2	School boards/Reg of
436024	2007-01-02	09:10:00	cafr	audi	000000220072	Ignatius expounds the inside story of how the Americas took over Cuba	972.5L W514C	972.5L	Cuba/History 1500/Constitution Cuba/Soviet Union Foreign relations C.
660741	2007-01-02	09:10:00	cafr	audi	000000120028	Logic of the right use of reason in the inquiry after truth with a vi.	360 1739B 1995	360	Logic John 1632 1784/Large Early works to 1800/Content of RR Early w.
622274	2007-01-02	09:10:00	cafr	audi	000000020046	Something About Darwin	FC 0102000	761.1	Love stories/Novel fiction/Regime History/Landscape England/Science
605548	2007-01-02	09:14:00	cafr	audi	000004823830	Understanding vaccination and from Duchamp to Hitler	706.04	706.04	Institutions ANNA Modern 20th century
151516	2007-01-02	09:10:00	cafr	audi	000000031075	Fourth Kingdom in Singapore	843.81	761.1	Kingdom Earth 8th Kingdom 1858 1935/Immigrant United States Geography
418239	2007-01-02	09:10:00	cafr	audi	000000224097	German radar alerts	840.953 781.02	840.953	World War 1939 1945/Naval operations German/Alerts Ship
701724	2007-01-02	09:10:00	cafr	audi	000000330021	Zabala	FC 000200099	761.1	American American novel/Realism/Novels Better Fiction
132947	2007-01-02	09:14:00	cafr	audi	000002041248	Restoration narrow gauge Oriskany Creek railroad New Arrow Short L.	385 12097 1878	385 12097	Area Culture/Landmarks/Conifers History/Logging Colorado Washington St.
123987	2007-01-02	09:10:00	cafr	audi	000004073141	How to cure brain Cell Cancer on the Celluloid	779 7812919	779	Cell Culture cell and 4th/Cell Culture/Cell and 4th/Cellular works
148147	2007-01-02	09:10:00	cafr	audi	000000081342	Buddhism in translation passages selected from the Buddhist cano.	294 40228	294	Buddhism Sacred books
148443	2007-01-02	09:10:00	cafr	audi	000004081010	Death in Washington Boston in Kentucky and the Stalin terror	227 227 415333 2003	227 227	Novels 4th of June C. 1861 1865/Epitaphs American/Epitaphs Secret.
264449	2007-01-02	09:10:00	cafr	audi	000005420908	101 diseases you don't want to get	614.3 782156 2005	614.3	Understanding Popular works/Understanding Popular works
27910	2007-01-02	09:00:00	naft	audi	000004020201	Thought in Texas	FC 701066	761.1	Love stories/Novel Fiction/Novel Fiction
151416	2007-01-02	09:10:00	naft	audi	000000120161	Bowling ball production exercises to enlighten the mind and	264 14433 1739 2000	264 14433	Novel and nonfiction/Novel Fiction/Novel Fiction aspects Buddhism
202048	2007-01-02	09:10:00	cafr	audi	000000080208	Regiance of faith the enlightened mind of Islam	297 12129 2004	297	Islam
201123	2007-01-02	09:10:00	naft	audi	000000170802	How to reduce workplace conflict and stress how leaders and their	655.001 92121 2005	655.001	Theory in the workplace/Interpersonal conf/Management/Management/O.
70728	2007-01-02	09:10:00	cafr	audi	000004784610	How the husband a true story of Antarctic survival	915.804 MARCON 1968	915.804	Antarctica Discovery and exploration/Antarctica Antarctic Expedition 1.
158626	2007-01-02	09:10:00	naft	audi	000004011870	Loss of mother daughter growing up with Aunt Muriel	8 84026 2001	8	American Canadian 20th century Female realistic/Novels Canadian.
208989	2007-01-02	09:10:00	cafr	audi	000004517884	Faith in the northwest how the Revolution stole the South and th.	234 70793 2009	234 70793	Democracy Party U/Mexican Party U 1 1814/Politics Practical United.
158626	2007-01-02	09:10:00	naft	audi	000004011870	Loss of mother daughter growing up with Aunt Muriel	241.66 501366 1997	241.66	Novel and nonfiction/Novel Fiction/Novels Religious aspects Canada.
170574	2007-01-02	09:10:00	cafr	audi	000000251127	Labels vocabulary of the Cluster Right now sources of Indian music	973.81 4070A 1991	973.81	Dance Indian Mus 1874-Late 19th/Right Now of the More 1874/Novel.
201212	2007-01-02	09:10:00	naft	audi	000000220048	Paradise on a island	FC 007704A	761.1	Novel Fiction
203312	2007-01-02	09:00:00	naft	audi	000004281830	devil's economy of business monkey business high finance and in.	330 2207 1995	330 2207	Finance Discretionary Business Discretionary
206280	2007-01-02	09:10:00	cafr	audi	000004120021	Building with the Buddha a concise guide with Rick Warren's Ad.	796 43089	796 43089	Calculus/Buddhism/Calculus country/Calculus country/Building/Calculus.
284448	2007-01-02	09:10:00	cafr	audi	000004073908	epined gene Shiroka Horvath investigates his ultimate ca.	FC 0405200	761.1	Novel Fiction/Novel Fiction/Novels Creative Fiction/Meeting pers.
170528	2007-01-02	09:10:00	cafr	audi	000000220048	Paradise on a island	810 80327 786050 1994	810 80327	Novel Fiction/Novel Fiction/Novels Creative Fiction/Meeting pers.
143770	2007-01-02	09:10:00	cafr	audi	000003208003	Do Shaker the Shaker reader who turned missionary a true and th.	8 0490W	8	Missouri Japan Do Shaker Japan 1812
153676	2007-01-02	09:10:00	cafr	audi	000004081010	Death in Washington Boston in Kentucky and the Stalin terror	1750820	761.1	Novel Fiction/Novel Fiction/Novels Creative Fiction/Meeting pers.
153676	2007-01-02	09:10:00	cafr	audi	000000112887	Burn!	8 80428	8	Burn! America 1911
210426	2007-01-02	09:10:00	cafr	audi	000000031075	the water a badtime story	FC 0004044	761.1	Imaginary conditions Fiction
159316	2007-01-02	09:10:00	cafr	audi	000001866049	Bullard Arms	338 7838 1018	338 7838	Bullard James Herbert 1842 31st/Novel Reporting Arms Company m.
210426	2007-01-02	09:10:00	cafr	audi	000000031075	the water a badtime story	810 80327 786050 1994	810 80327	Imaginary conditions Fiction
214741	2007-01-02	09:10:00	naft	audi	000000050780	Satisfaction the art of finding true fulfillment	155 8 845815 2005	155 8	Satisfaction
148414	2007-01-02	09:10:00	naft	audi	000000040778	Fi Future America's Political History	FC 117100A	761.1	Future America's Political History
153449	2007-01-02	09:10:00	naft	audi	000000731190	Chronicle of the gods unraveled mysteries of the past	001 84 DANNO 1999	001 84	Chronology/Myths on other cultures/Chronology Ancient Environmental n.
148414	2007-01-02	09:10:00	naft	audi	000000040778	Fi Future America's Political History	241.	241.	Future America's Political History
14810	2007-01-02	09:10:00	cafr	audi	000000041141	Future of medicine based on science of prevention based on sci.	413 110UGLS 1993	413	Medicine sci United States/Science Prevention United States
41488	2007-01-02	09:10:00	cafr	audi	000000010462	Angels' operation a handbook for the person in charge FC	823 10068 1987	823 10068	Novels Children's Novels etc.
151111	2007-01-02	09:10:00	cafr	audi	000004340800	Angels' operation a handbook for the person in charge FC	823 10068 1987	823 10068	Novels Children's Novels etc.
202048	2007-01-02	09:10:00	cafr	audi	000000080208	Regiance of faith the enlightened mind of Islam	297 12129 2004	297 12129	Islam
206121	2007-01-02	09:10:00	naft	audi	000000010462	Angels' operation a handbook for the person in charge FC	823 10068 1987	823 10068	Novels Children's Novels etc.
160750	2007-01-02	09:10:00	cafr	audi	000004011870	Loss of mother daughter growing up with Aunt Muriel	241.66 501366 1997	241.66	Novel and nonfiction/Novel Fiction/Novels Religious aspects Canada.
159709	2007-01-02	09:10:00	cafr	audi	000004020201	Thought in Texas	FC 007704A	761.1	Novel Fiction
100045	2007-01-02	09:10:00	cafr	audi	000006200103	penalty Christmas	CD 782 2123 73133	782 2123	Christmas Novel/Novel Music
181212	2007-01-02	09:10:00	cafr	audi	000000011871	Architects' dream home like book	728 17 26465 2000	728 17	Architecture/Design/Classical/Style/Novel Architecture United States 4.

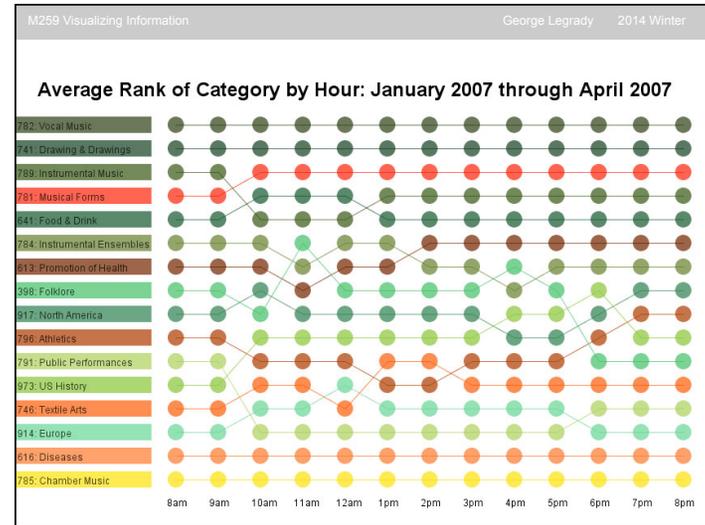
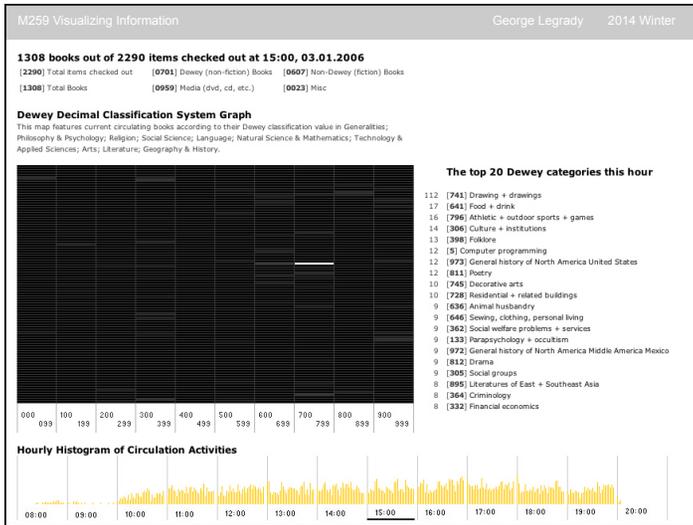
Dewey Classification System

Ten topics each subdivided into 100 subclasses:

- 000 - Generalities
- 100 - Philosophy & Psychology
- 200 - Religion
- 300 - Social Science
- 400 - Language
- 500 - Natural Science & Mathematics
- 600 - Technology & Applied Sciences
- 700 - Arts
- 800 - Literature
- 900 - Geography & History

Daily Volume Activity (see at outlog/inlog)

hour	789	613	784	917	746	973	796	398	616	914	791	306	332	305	158	792	811	658	635	895	
0	0	0	0	0	8	0	4	12	1	0	0	0	1	0	0	1	0	5	0	0	4
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	7	1	2	0	5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	15	58	57	109	38	90	39	73	25	50	41	34	20	31	13	33	33	14	24	42	
7	27	71	48	134	40	87	62	73	41	20	28	36	35	42	12	32	52	14	28	45	
8	15	58	57	109	38	90	39	73	25	50	41	34	20	31	13	33	33	14	24	42	
9	27	71	48	134	40	87	62	73	41	20	28	36	35	42	12	32	52	14	28	45	
10	1192	942	812	777	619	632	582	515	465	472	440	383	466	394	429	362	420	370	402	396	
11	1560	1149	1125	969	880	1006	786	966	702	701	567	610	453	556	542	526	628	488	532	560	
12	2185	1683	1878	1177	1025	1108	1156	889	881	935	782	706	779	700	715	625	788	680	647	960	
13	3266	2334	2085	1769	1717	1615	1462	1324	1181	1275	972	1117	1175	1025	1025	997	968	967	986	987	
14	3336	2698	2227	1811	1890	1744	1685	1487	1421	1271	1144	1146	1059	1044	1062	1041	990	962	1105	1140	
15	3995	2810	2426	1870	1850	1813	1703	1427	1444	1260	1288	1142	1187	1172	1073	1143	1139	1074	1139	1044	
16	3959	2855	2599	2068	1857	1867	1904	1798	1520	1449	1411	1390	1250	1306	1321	1265	1101	1126	1110	1119	
17	4322	2870	2547	2203	1992	1817	1904	1374	1438	1416	1322	1402	1366	1317	1346	1359	1152	1248	1045	798	
18	2398	1574	1181	1012	1024	1034	1059	630	756	615	719	682	678	680	677	685	604	562	532	280	
19	2233	1557	1284	1050	845	942	1027	643	612	524	595	713	706	594	623	645	562	651	497	341	
20	48	18	17	35	9	15	25	18	9	6	15	4	3	3	4	6	11	17	22	4	
21	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	6	1	2	5	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	



M259 Visualizing Information George Legrady 2014 Winter

Data is Multivariate

- Data is multivariate. Each transaction includes numeric, ordinal, interval scale (time, date), string, and other classification data.
- Include:
 - ItemType (bks, cds)
 - Collection date
 - Check-out/check-in hour/day
 - Title
 - Dewey Classification
 - Keywords
 - Unique IDs (barcode, etc.)

M259 Visualizing Information George Legrady 2014 Winter

1st Project Development Process

- **Data search:** Knowledge discovery
- **Data analysis:** What patterns emerge
- **Data formulation through algorithmic processing**
- Translation into visualization
- Correlation with External Data
- Publication

Review Course Links:

- [Dewey Classification System](#)
- [Library itemtypes](#)
- [Seattle Library Item Search](#)
- [Legrady Project: Top 20 Dewey](#)
- [Previous Course Projects](#)
- [MySQL Queries](#)

1st Assignment: A Compelling Data Query

- **State a Question:** Can be cultural, system based, compare data, etc. *Should be exploratory or compelling....*
- **Translate into a Query:**
- **Report Result:** (a Data printout to .csv file)
- **Processing Time:** How long the search
- **Comment & Analysis:** Report what you found through the data

- Assignment is due Tues Jan 21

How to Proceed

- **Ask a general question**
- Change syntax if necessary
- **Review Results**
- Ask question differently
- **Review Results**
- Go into more detail

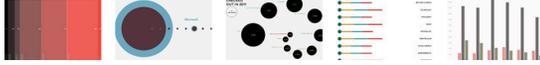
Begin with Inraw – by the week-end switch to more specific tables with MySQL innerjoin function

The Experimental Visualization Lab

Home Projects Courses People Contact

M259 Data Visualization - 2012

One-Dimensional Frequency Graph Assignment



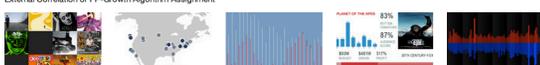
Anis Heron HanYoon Jung RJ Duran Ankit Srivastava David Gordon

Two-Dimensional Spatial Map Assignment



Ankit Srivastava HanYoon Jung RJ Duran Dallas Mercer David Gordon

External Correlation or FP-Growth Algorithm Assignment



RJ Duran Anis Heron Dallas Mercer Ankit Srivastava HanYoon Jung David Gordon

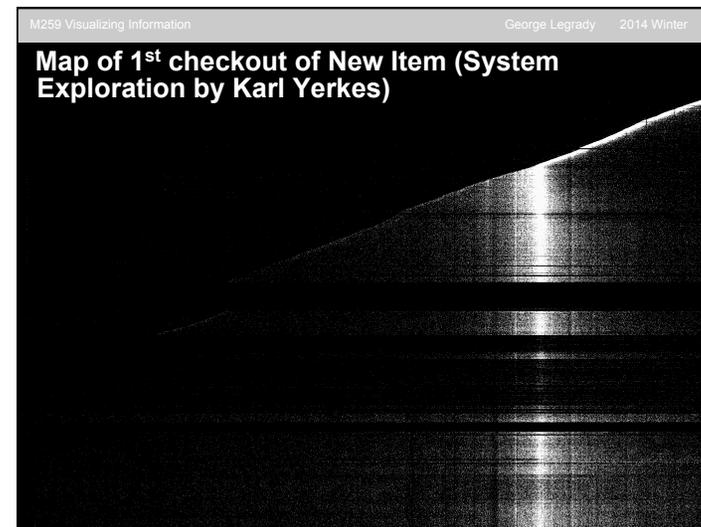
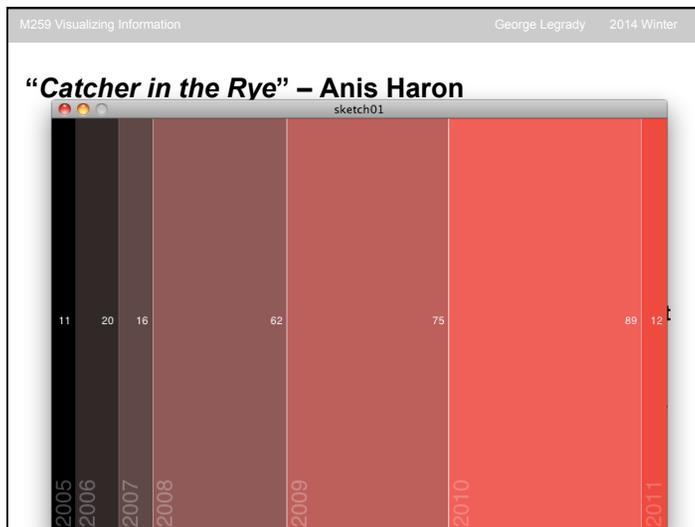
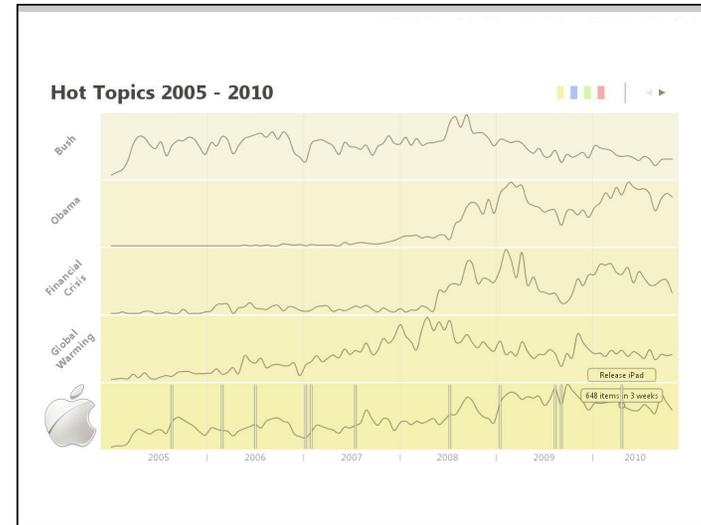
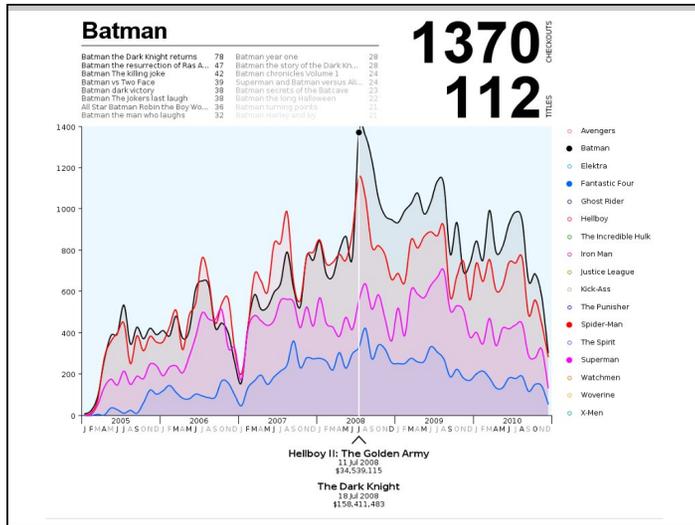
Three-Dimensional Interactive Assignment

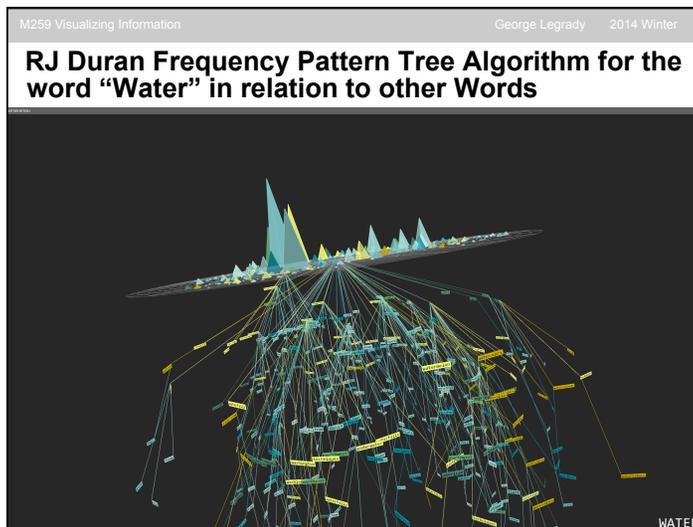
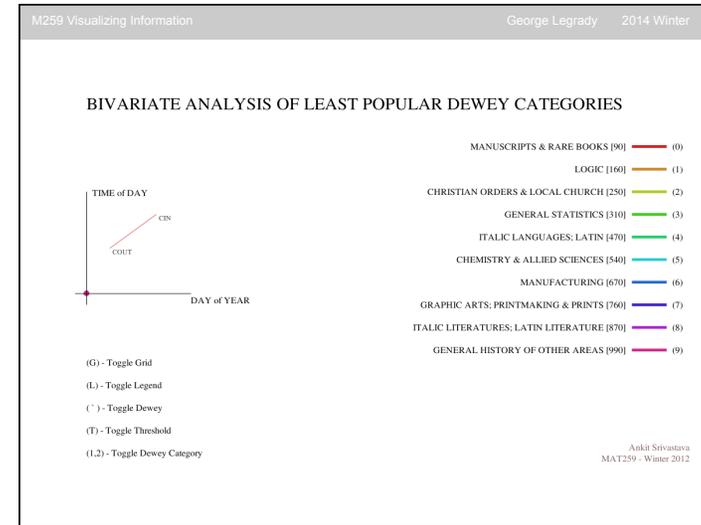
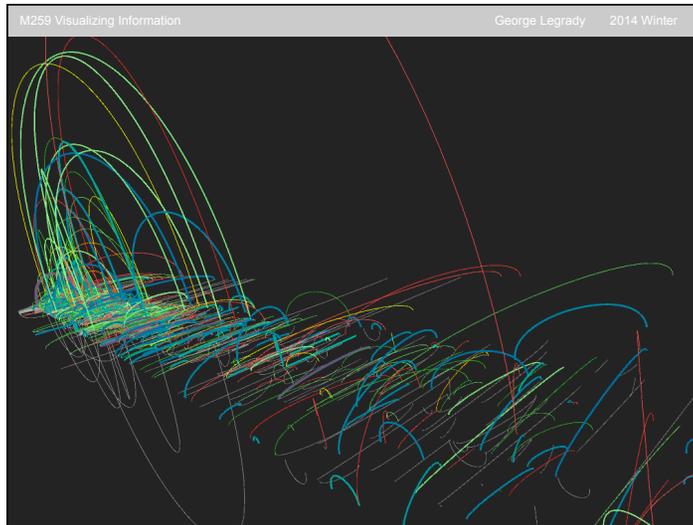


RJ Duran Anis Heron HanYoon Jung Ankit Srivastava Dallas Mercer David Gordon

MAT <http://vislab.mat.ucsb.edu/2012.html>

© MAT University of California Santa Barbara, CA





M259 Visualizing Information George Legrady 2014 Winter

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

Data Processing Functions

- Validation: Ensuring that supplied 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 (from various sources)
- Analysis: Collection, organization, analysis, interpretation and presentation of data
- Reporting: List detail or summary data or computed information

Some References

- Atlas of Science, Katy Borner
- Graphics of Large Datasets, Unwin, Theus, Hofmann
- Visualizing Data, Ben Fry
- Robert Kosara, social analysis

Previous student projects:

<http://vislab.mat.ucsb.edu/courses.html>