Seattle Public Library Data

George Legrady © 2021
Experimental Visualization Lab
Media Arts & Technology
UC Santa Barbara


HTTP://WWW.SPL.COM/~MIXING CHAMBER
SPL SERVICE BAR


librarian

Data Flow of Check-Outs

- Database began January 1, 2006
- Project server receives data by the hour
- We download it every night around 11:30pm
- Currently approximately 98 million checkouts
- We also get the return time-stamp (check-in)
- Data is multivariate (numeric, string, etc.)
- Over 50% fiction, non-fiction labeld according to the Dewey Decimal Classification System

Dewey / Non-Dewey

year	nonDewey	Dewey			year	nonDewey	Dewey	
2006	3424214	2653745	6077959		2006	3572597	3378622	6951219
2007	3362577	2683372	6045949		2007	3928484	3795238	7723722
2008	4713593	3585947	8299540		2008	5256145	4872025	10128170
2009	5141529	3814351	8955880		2009	5725926	5131491	10857417
2010	4979017	3442582	8421599		2010	5539309	4574848	10114157
2011	4655379	3216899	7872278		2011	4609492	3545685	8155177
2012	4361197	2900748	7261945		2012	3441180	2266234	5707414
2013	4750906	3135548	7886454		2013	3664329	2389017	6053346
2014	4515221	2872244	7387465		2014	3500947	2192083	5693030
2015	4298895	2677565	6976460		2015	3296759	2012686	5309445
2016	4007669	2466122	6473791		2016	3087764	1857528	4945292
2017	3648182	2260682	5908864		2017	2681466	1610939	4292405
2018	2576158	1549714	4125872		2018	2094239	1245136	3339375
2019	3694800	1984203	5679003		2019	2838526	1507385	4345911
2020	1037786	573562	1611348		2020	903036	489560	1392596
OutGoing	59167123	39817284	98984407		InComing	54140199	40868477	95008676

Monthly, Hourly in 2020

MONTH(cout)	BOOK	DVD	CD	MISC	ILL
1	322923	108208	33062	1989	199
2	225982	79283	24581	1515	250
3	206172	60038	16560	1034	123
4	66	4	4	0	0
5	87	3	0	2	0
6	42	4	0	0	0
7	50	12	1	0	0
8	8182	3328	1147	21	4
9	60264	13404	5878	235	16
10	87805	18128	8250	312	5
11	87232	20395	7877	345	18
12	82493	24449	8392	415	124

HOUR(cout)	BOOK	DVD	CD	MISC	ILL
0	1	0	0	0	0
6	8	4	0	4	0
7	59	96	10	1	0
8	888	392	102	8	5
9	8465	2788	1184	104	69
10	69683	19926	6323	346	96
11	94144	25385	7764	519	121
12	129797	39578	13228	733	88
13	145452	46284	14603	802	85
14	147667	47225	15152	751	81
15	158719	45613	15248	804	77
16	164767	44947	15186	810	50
17	118230	36894	11763	677	50
18	25388	10201	2980	154	11
19	17963	7910	2188	153	6
20	64	13	21	2	0
21	1	0	0	0	0
22	2	0	0	0	0

The SPL Multivariate Data & its Metadata

Ordinal (*In a numeric sequence*)

- **ID:** Each database entry has a unique ID number
- **ItemNumber**: Assigned when object enters system
- **Dewey Classification** (Dewey numeric)

Interval Scale (*Time-Stamp*)

- **Check-out/check-in hour, day, month, year**

Categorical (*Not necessarily numerically orderable*)

- **BibNumber**: Each title has a specific number, copies of titles all have same number
- **Barcode**: Each item has a unique number on RFID sticker
- **CallNumber**: by which to locate items on shelves - Ordinal if Dewey, otherwise categorical
- **Collection Code**: What the item is and where its located

Semantic (*Text-based*)

- **Title:** Each Item has a title
- **ItemType**: books, cds, dvds, music sheets, etc.
- **Subjects**: Keywords (arbitrary labeling)

MySQL: Select * from spl_2016.inraw where year(cout) = 2007 and month(cout) >= 1 and month(cout) <= 4 limit 50;

itemNumber	cout	collcode	itemtype	barcode	title	callNumber	deweyClass	subj
2124890	2007-01-02 09:20:00	nacd	accd	0010053294418	tigers have spoken	CD 782.421642 C2665T	782.421642	Rock music 2001 2010^Country music 2001 2010
1963406	2007-01-02 09:12:00	cafic	acbk	0010046075908	Heart on the line	FIC ARNOLD	NULL	Love stories^Women television producers and directors New York State N...
1348064	2007-01-02 09:03:00	nalpfic	acbk	0010047150288	Too hot to handle	FIC LOWELL	NULL	Love stories^Large type books^Ranch life Fiction
264450	2007-01-02 09:13:00	canf	acbk	0010027180271	Governing public schools new times new requirements	371.2 DANZBER 1992	371.2	School boards Rating of
436074	2007-01-02 09:13:00	cs9	acbk	0010002326972	great deception the inside story of how the Kremlin took over Cuba	972.91 M741G	972.91	Cuba History 1959^Communism Cuba^Soviet Union Foreign relations Cu...
660741	2007-01-02 09:13:00	canf	acbk	0010041599258	Logic or The right use of reason in the inquiry after truth with a va...	160 L793W 1996	160	Locke John 1632 1704^Logic Early works to 1800^Conduct of life Early w...
2225714	2007-01-02 09:12:00	cafic	acbk	0010050230746	Something about Emmaline	FIC BOYLE2005	NULL	Love stories^Historical fiction^Regency fiction^London England Fiction
860548	2007-01-02 09:14:00	canf	acbk	0010048238389	Understanding installation art from Duchamp to Holzer	709.04 R7277U 2003	709.04	Installations Art^Art Modern 20th century
15316	2007-01-02 09:12:00	cs9	acbk	0000103502175	Frank B Kellogg a biography	B K292 B	NULL	Kellogg Frank B Frank Billings 1856 1937^Statesmen United States Biography
418329	2007-01-02 09:13:00	cs9	acbk	0010002324977	German raider Atlantis	940.953 F8517G	940.953	World War 1939 1945 Naval operations German^Atlantis Ship
745716	2007-01-02 09:13:00	cafic	acbk	0010033394031	Zabelle	FIC KRICORI1998	NULL	Armenian American women Massachusetts Boston Fiction
1529347	2007-01-02 09:14:00	canf	acbk	0000102981248	Reservation narrow gauge Omak Creek railroad Bow Arrow Short Li...	385.52097 L587R	385.52097	Biles Coleman Lumber Company History^Logging railroads Washington St...
1558007	2007-01-02 09:13:00	canf	acbk	0010004731641	place no one knew Glen Canyon on the Colorado	779 P833P2a	779	Glen Canyon Utah and Ariz^Glen Canyon Utah and Ariz Pictorial works
1681487	2007-01-02 09:12:00	canf	acbk	0010002691342	Buddhism in translations passages selected from the Buddhist sacr...	294 W252B	294	Buddhism Sacred books
1688675	2007-01-02 09:14:00	canf	acbk	0010048100100	death in Washington Walter G Krivitsky and the Stalin terror	327.1247 K4592D 2003	327.1247	Krivitsky W G Walter G 1899 1941^Espionage American^Espionage Soviet...
2424499	2007-01-02 09:13:00	canf	acbk	0010054269096	101 diseases you dont want to get	614.5 P8715o 2005	614.5	Epidemiology Popular works^Diseases Popular works
27939	2007-01-02 09:09:00	nafic	acbk	0010040162363	Twilight in Texas	FIC THOMAS	NULL	Love stories^Texas Fiction^Texas Rangers Fiction
1534336	2007-01-02 09:00:00	nanf	acbk	0010041237933	Boundless healing meditation exercises to enlighten the mind and...	294.34435 T387B 2000	294.34435	Mind and body^Meditation Buddhism^Healing Religious aspects Buddhism
2205408	2007-01-02 09:14:00	canf	acbk	0010050859288	fragrance of faith the enlightened heart of Islam	297 R1294F 2004	297	Islam
2631813	2007-01-02 09:14:00	nanf	acbk	0010053789672	How to reduce workplace conflict and stress how leaders and their...	658.1053 M327H 2005	658.1053	Teams in the workplace^Interpersonal conflict^Mediation^Negotiation^O...
702785	2007-01-02 09:13:00	canf	acbk	0010034758416	home of the blizzard a true story of Antarctic survival	919.8904 MAWSON 1998	919.8904	Antarctica Discovery and exploration^Australasian Antarctic Expedition 1...
1168659	2007-01-02 08:58:00	nab	acbk	0010045761730	Lives of mothers daughters growing up with Alice Munro	B M9265M 2001	NULL	Novelists Canadian 20th century Family relationships^Novelists Canadian...
2498965	2007-01-02 09:14:00	canf	acbk	0010054107684	Foxes in the henhouse how the Republicans stole the South and th...	324.70973 R29996J 2006	324.70973	Democratic Party U S^Republican Party U S 1854^Politics Practical United...
1076021	2007-01-02 09:13:00	canf	acbk	0010045523353	False intimacy understanding the struggle of sexual addiction	241.66 Sch196F 1997	241.66	Sex addicts Rehabilitation^Intimacy Psychology Religious aspects Christia...
1703576	2007-01-02 09:12:00	canf	acbk	0010025511717	Lakota recollections of the Custer fight new sources of Indian milit...	973.82 LAKOTA 1991	973.82	Dakota Indians Wars 1876^Little Bighorn Battle of the Mont 1876 Person...
2033121	2007-01-02 09:13:00	cafic	acbk	0010020070768	Furors die a novel	FIC HOFFMAN	NULL	NULL
2633117	2007-01-02 09:06:00	nanf	acbk	0010042818186	devils dictionary of business monkey business high finance and lo...	330.0207 V895D 2005	330.0207	Finance Dictionaries^Business Dictionaries
295289	2007-01-02 09:13:00	canf	acbk	0010041330571	Running with the Buffaloes a season inside with Mark Wetmore Ad...	796.42809 C7193L 2000	796.42809	Colorado Buffaloes Cross country team^Cross country running Colorado...
2684484	2007-01-02 09:13:00	canew	acbk	0010054076608	opened grave Sherlock Holmes investigates his ultimate case	FIC JAMES2006	NULL	Mystery fiction^Holmes Sherlock Fictitious character Fiction^Missing pers...
756568	2007-01-02 09:14:00	canf	acbk	0010028375219	Transforming vision writers on art	810.80357 TRANSFO 1994	810.80357	Art^American poetry 20th century^Art Poetry
1437750	2007-01-02 09:13:00	cs9	acbk	0000102286903	De Shazer the Doolittle raider who turned missionary a true and th...	B D459W	NULL	Missions Japan^De Shazer Jacob 1912
67201	2007-01-02 09:45:00	nchol	jcbk	0010045871109	Jothams journey a storybook for Advent	J YTREEID	NULL	Family Prayer books and devotions English^Advent Prayer books and dev...
1336766	2007-01-02 09:44:00	cs9o	acbk	0010001512887	Burri	B B942B	NULL	Burri Alberto 1915
2582486	2007-01-02 09:44:00	ncnew	jcbk	0010050556819	Lady in the water a bedtime story	E SHYAMAL	NULL	Imaginary creatures Fiction
1978356	2007-01-02 09:45:00	cs6ro	arbk	0010019836690	Bullard Arms	338.76834 J241B	338.76834	Bullard James Herbert 1842 1914^Bullard Repeating Arms Company Hist...
1404338	2007-01-02 09:44:00	canf	acbk	0010038257472	Perfect bones a six point plan to promote healthy bones	616.716 L5785P 2000	616.716	Osteoporosis Popular works^Women Nutrition
2347671	2007-01-02 09:44:00	nanf	acbk	0010050592780	Satisfaction the science of finding true fulfillment	155.9 B4581S 2005	155.9	Satisfaction
468341	2007-01-02 09:41:00	nafic	acbk	0010022410574	Fly fishing tales literary bait by angling authors	FIC FLY FIS1994	NULL	Fishing stories American^Fly fishing Fiction
1558499	2007-01-02 10:02:00	nanf	acbk	0010037361580	Chariots of the gods unsolved mysteries of the past	001.94 DANIKEN 1999	001.94	Archaeology^Life on other planets^Civilization Ancient Extraterrestrial in...
2064174	2007-01-02 09:41:00	canf	acbk	0010036046778	way of agape	241.4 MISSLER 1999	241.4	Agape^Love Religious aspects Christianity
18430	2007-01-02 09:41:00	canf	acbk	0010026014141	Future of medicine toward a science of prevention based on ancie...	613 DUGLISS 1993	613	Medical care United States^Medicine Preventive United States
614085	2007-01-02 09:41:00	canf	acbk	0010045701462	Tanker operations a handbook for the person in charge PIC	623.88245 H8627T 2001	623.88245	Tankers Handbooks manuals etc
1591175	2007-01-02 09:39:00	caesl	bcbk	0010034398809	Angliiskii iazyk prosto o slozhnom prakticheskii kurs	RUSSIAN 428.24917 L576A	428.24917	English language Textbooks for foreign speakers Russian
2670845	2007-01-02 08:48:00	canf	acbk	0010051680022	Gorgeous disaster the tragic story of Debra LaFave	364.153 L131L 2006	364.153	Sexual abuse victims United States Case studies^Female sex offenders U...
1064531	2007-01-02 10:16:00	nacd	accd	0010040958539	Spirituals in concert	CD784.73 B322S	784.73	Spirituals Songs^Songs High voice with instrumental ensemble^Songs Hi...
965750	2007-01-02 09:38:00	capf	acbk	0010025431536	Kathy and Mo show parallel lives	812.54 GAFFNEY 1992	812.54	Women Drama
1970799	2007-01-02 09:39:00	nafic	acbk	0010046761242	revelation	FIC LITTLE	NULL	Horror fiction
1006825	2007-01-02 10:17:00	cacd	accd	0010046200415	peaceful Christmas	CD 782.21723 P3133	782.21723	Christmas music^New Age music
1831232	2007-01-02 10:17:00	canf	acbk	0010048571573	Tauntons family home idea book	728.37 St546T 2003	728.37	Architecture Domestic United States^Interior architecture United States^R...

```
1 • select * from spl_2016.outraw limit 10
2
```

130% 1:2

Result Grid Filter Rows: Search Edit: Export/Import: Fetch rows:

id	itemNumber	bibNumber	cout	cin	collcode	itemtype	barcode	title	callNumber	deweyClass	subj
1	2365014	2303430	2006-01-01 15:59:00	NULL	cafic	acbk	0010054310593	On beauty a novel	FIC SMITH2005		NULL
2	1662460	1792579	2006-01-01 16:00:00	NULL	nafic	acbk	0010034862846	blood artists a novel	FIC HOGAN1998		NULL
3	1800999	2055046	2006-01-02 08:48:00	NULL	cacd	accd	0010041494690	Same ol timeously vocalisthenics stu...	CD 782.421643 B119S	782.421643	NULL
4	1975813	2105769	2006-01-02 08:50:00	NULL	cyfic	acbk	0010042882703	Hard love	YA WITTLIN		NULL
5	2215126	2288710	2006-01-02 09:30:00	NULL	canew	acbk	0010051150489	Always time to die	MYSTERY LOWELL2005		NULL
6	2347693	2298921	2006-01-02 09:31:00	NULL	nanf	acbk	0010050668051	Garbage land on the secret trail of tra...	362.7285 R816G 2005	362.7285	NULL
7	686211	1881970	2006-01-02 09:58:00	NULL	nacd	accd	0010043883858	singles 8185	CD 782.42166 D4411S	782.42166	NULL
8	737154	1331901	2006-01-02 09:58:00	NULL	canf	acbk	0010026319490	Istanbul and northwest Turkey	914.9618 ISTANBU 1993	914.9618	NULL
9	1732724	2097769	2006-01-02 09:58:00	NULL	cacd	accd	0010043997435	Beautiful chaos greatest hits live	CD 782.42166 P9592B	782.42166	NULL
10	2260994	2304170	2006-01-02 09:58:00	NULL	nacd	accd	0010053546353	Napoleon Dynamite original motion p...	CD 781.542 Sw57N	781.542	NULL
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

```
1 • select * from spl_2016.outraw order by ID DESC limit 10
2
```

130% 43:1

Result Grid Filter Rows: Search Edit: Export/Import: Fetch rows:

id	itemNumber	bibNumber	cout	cin	collcode	itemtype	barcode	title	callNumber	deweyClass	subj
98988737	7243036	3518699	2021-01-02 18:00:00	NULL	nadvd	acdvd	0010101529542	Uncut gems	DVD UNCUT GEMS		NULL
98988736	6589136	3369151	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010093736543	Minecraft Guide to PVP minigames	J793.932 M643Mp 2018	793.932	NULL
98988735	7225041	3516259	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010101093473	Minecraft the Survivors book of secrets	J793.932 M643M 2016	793.932	NULL
98988734	7016850	3461884	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010096462279	Hacks for Minecrafters Command blo...	J793.932 M6165Hcb 2019	793.932	NULL
98988733	7265080	3102505	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010101252327	ultimate unofficial encyclopedia for M...	J793.932 M6165U 2015	793.932	NULL
98988732	7121621	3487323	2021-01-02 17:57:00	NULL	ncser	jcbk	0010099687849	Night of the bats	J/E ELIOPULOS		NULL
98988731	7148860	3493199	2021-01-02 17:57:00	NULL	ncnew	jcbk	0010101902616	Minecraft maps	J793.932 M643Mm 2019	793.932	NULL
98988730	6814993	3422411	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010094697835	Minecraft Guide to farming	J793.932 W713M 2018	793.932	NULL
98988729	7389644	3618209	2021-01-02 17:57:00	NULL	namys	acbk	0010101310182	Murder take two	MYSTERY PERRY 2020		NULL
98988728	6677338	3369152	2021-01-02 17:57:00	NULL	ncnf	jcbk	0010092147890	Minecraft guide to enchantments poti...	J794.8 M643M 2018	794.8	NULL
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Dewey Decimal Classification

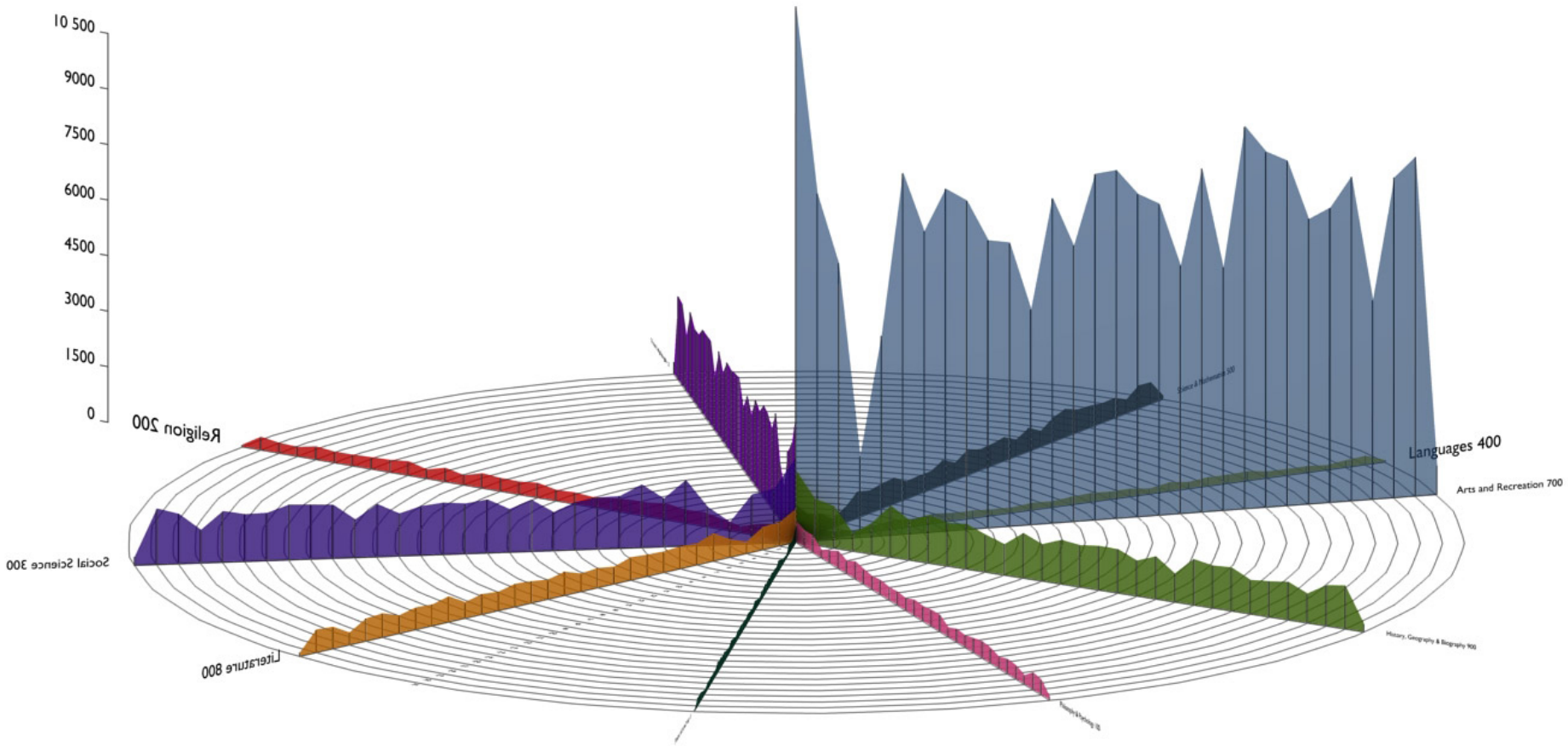
Ten topics each subdivided into 100 subclasses:

000-099 - Generalities
100-199 - Philosophy & Psychology
200-299 - Religion
300-399 - Social Science
400-499 - Language
500-599 - Natural Science & Mathematics
600-699 - Technology & Applied Sciences
700-799 - Arts
800-899 - Literature
900-999 - Geography & History

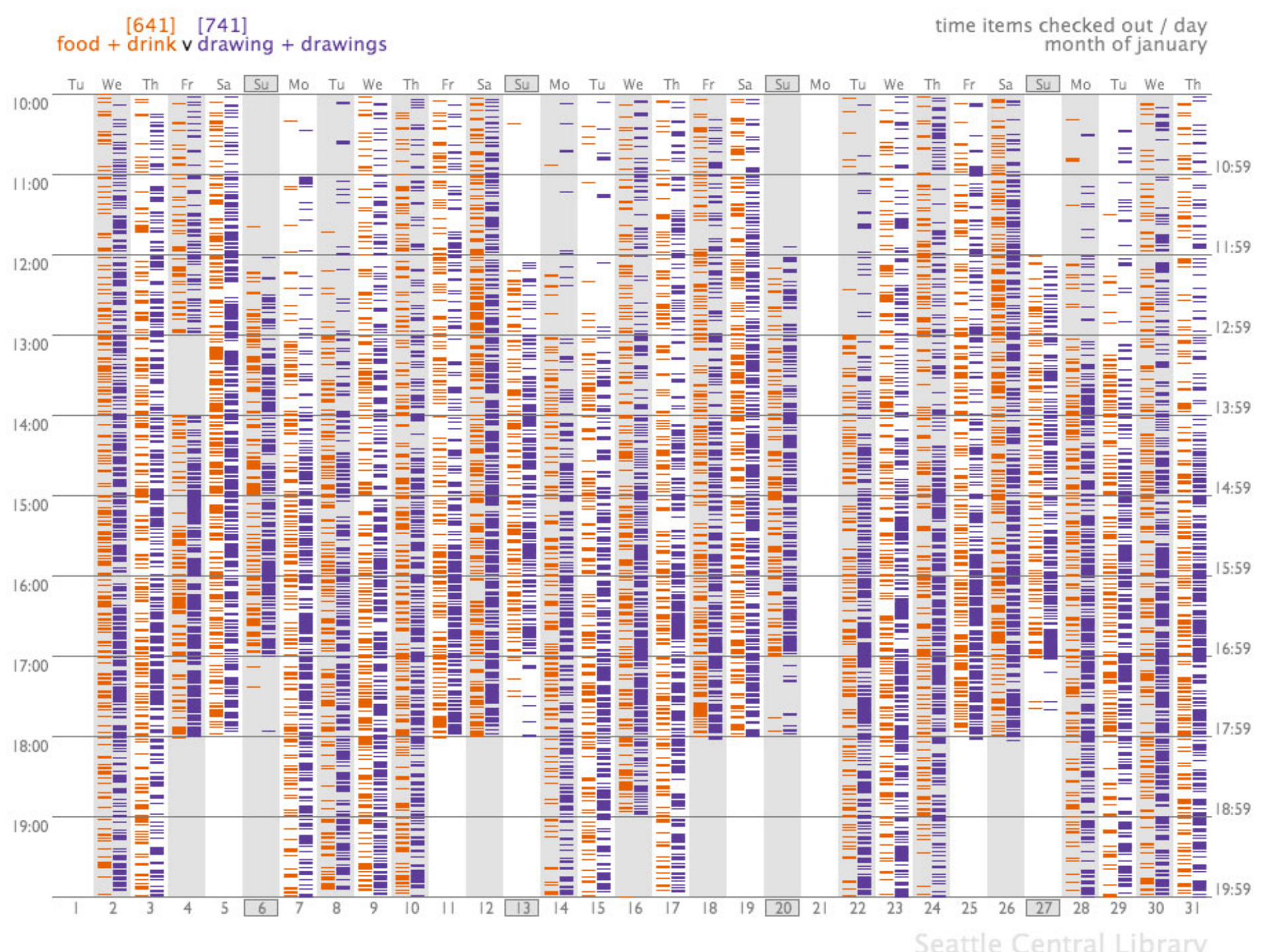
Typical daily Dewey activity - 20 most active

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	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
2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	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
6	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
7	1	2	0	5	0	1	1	0	0	1	0	0	0	2	1	0	1	0	0	2
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	70	78	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	1417	1444	1360	1388	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	1381	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	6	0	0	0	0	0	0	0	0	0	0	0
23	6	1	2	5	1	0	6	0	0	0	4	2	2	0	2	7	1	1	0	0

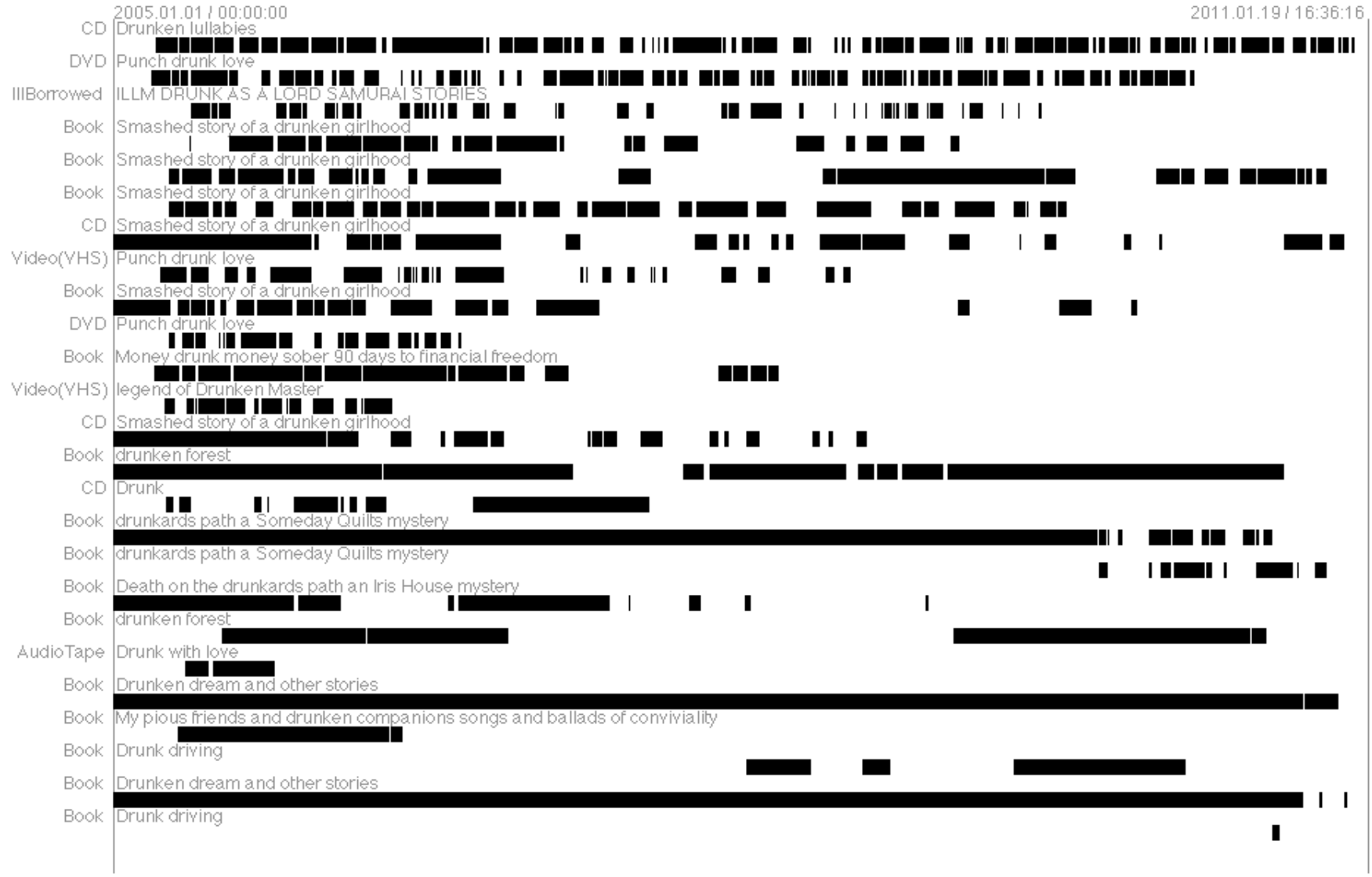
Volume of Dewey Activities



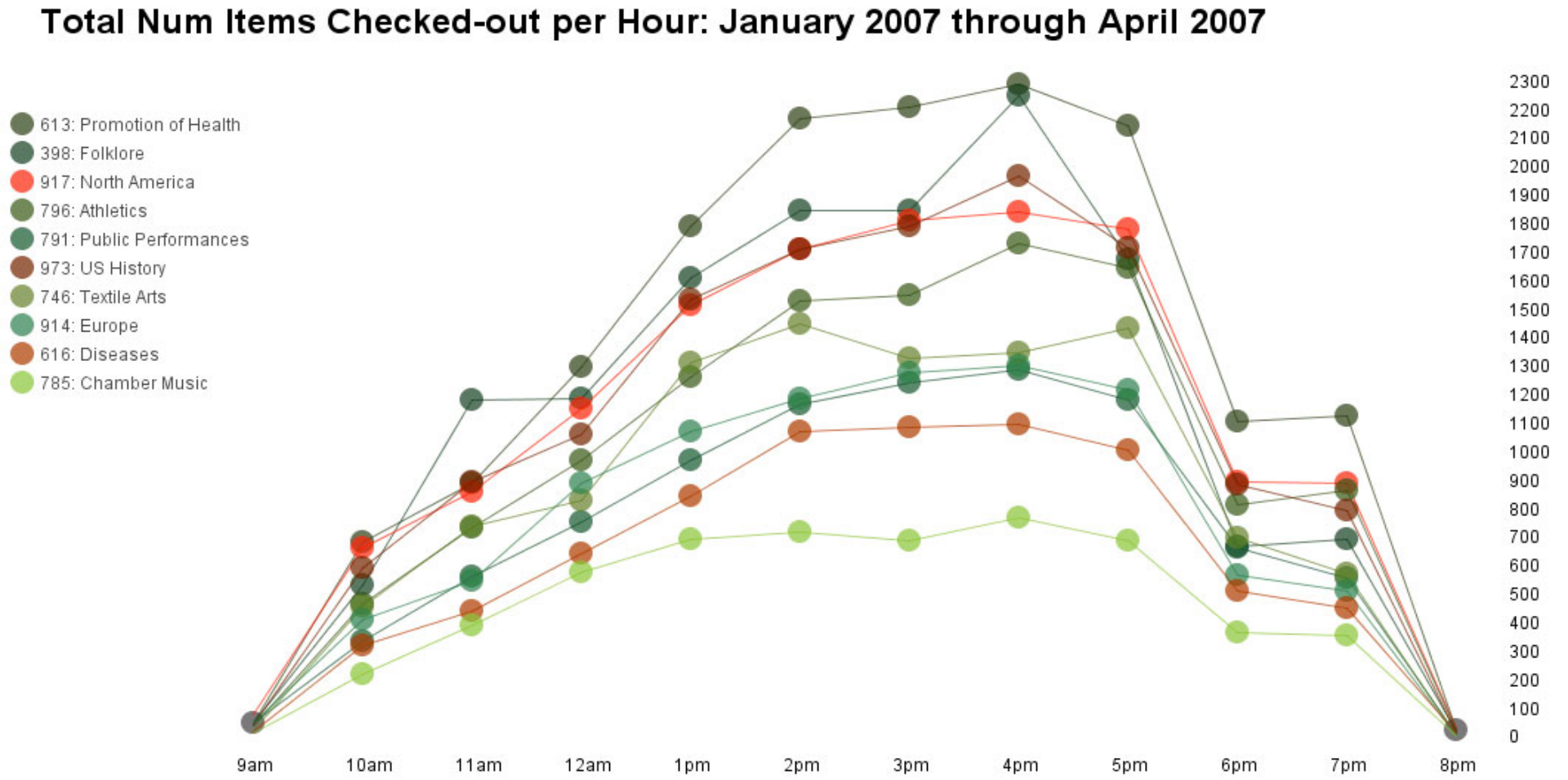
Stanislas Bernatt

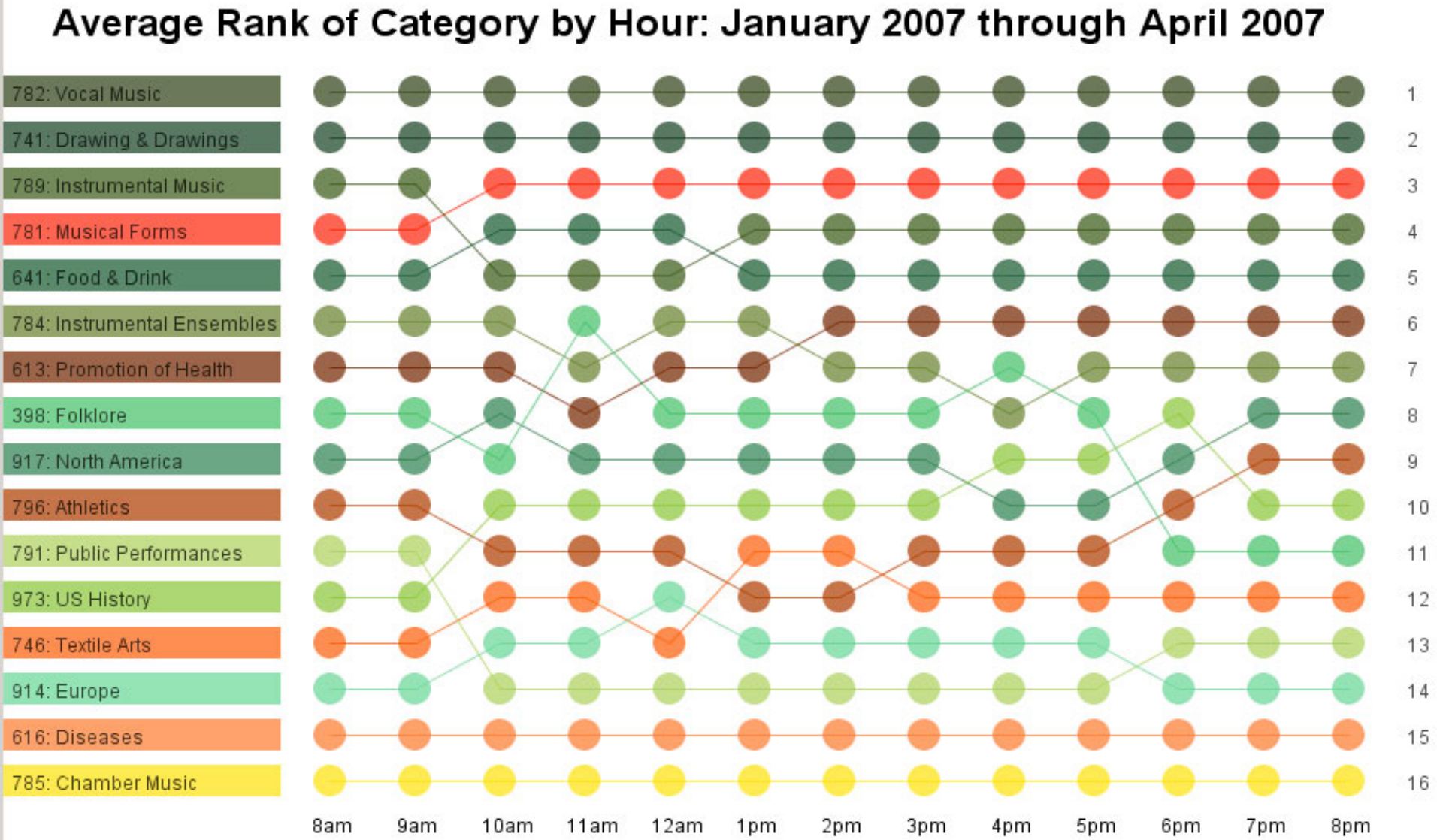

[641] [741]
food + drink v drawing + drawings
time items checked out / day
month of january
Tu We Th Fr Sa Su Mo Tu We Th Fr Sa Su Mo Tu We Th Fr Sa Su Mo Tu We Th Fr Sa Su Mo Tu We Th
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
10:59
11:59
12:59
13:59
14:59
15:59
16:59
17:59
18:59
19:59
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Seattle Central Library

title like '%drunk%' group by collcode;



Typical daily Dewey activity

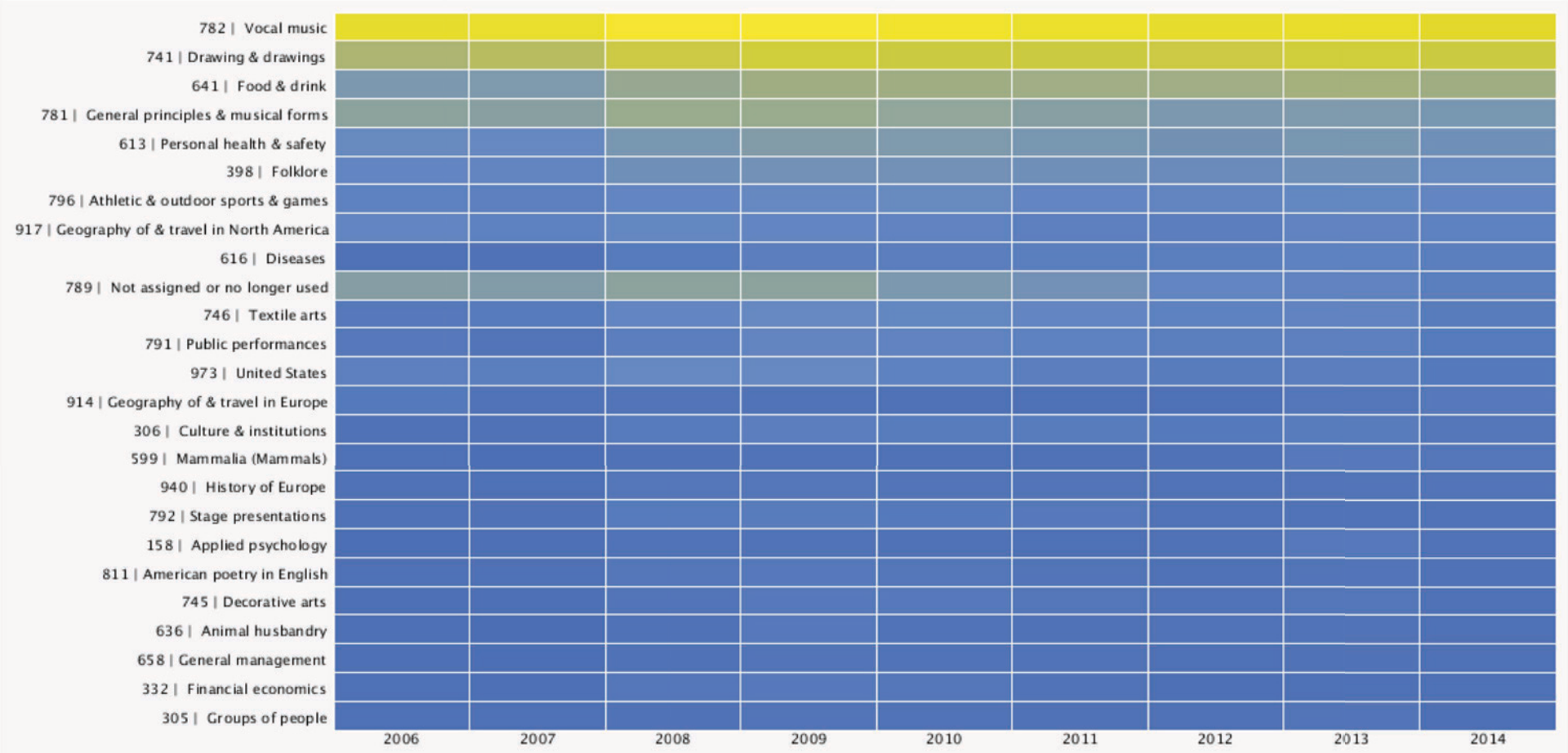


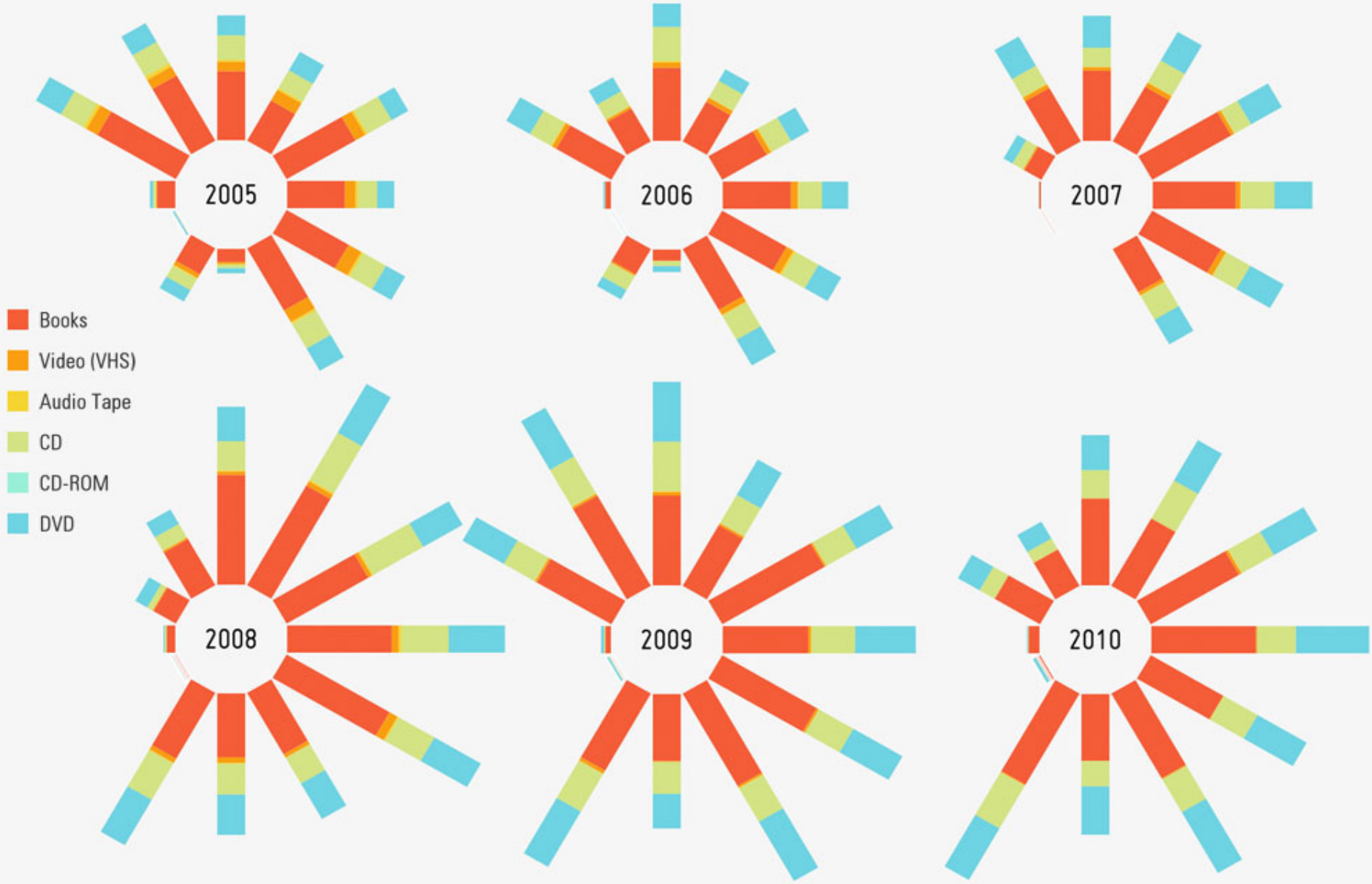

Average Rank of Category by Hour: January 2007 through April 2007
782: Vocal Music
741: Drawing & Drawings
789: Instrumental Music
781: Musical Forms
641: Food & Drink
784: Instrumental Ensembles
613: Promotion of Health
398: Folklore
917: North America
796: Athletics
791: Public Performances
973: US History
746: Textile Arts
914: Europe
616: Diseases
785: Chamber Music
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
8am
9am
10am
11am
12am
1pm
2pm
3pm
4pm
5pm
6pm
7pm
8pm

Top 25 Dewey categories 2006-2014

		2006	2007	2008	2009	2010	2011	2012	2013	2014
782	Vocal music	657750	721853	1006571	1060888	868834	774273	665173	700986	593881
741	Drawing & drawings	158295	201324	312239	354992	323449	330355	315883	366327	334448
641	Food & drink	65057	67695	99965	121586	125332	121088	116785	137652	125462
781	General principles & musical forms	82066	79529	107747	109614	90869	75959	62922	66173	61207
613	Personal health & safety	45227	43632	61698	69461	65593	61935	56773	61824	50671
398	Folklore	39845	38035	50875	52707	53403	50482	46519	51337	45293
796	Athletic & outdoor sports & games	34380	32561	38817	42247	44010	39552	37113	40243	39833
917	Geography of South America	39684	35096	40760	39546	35287	32660	30101	35325	34345
616	Diseases	20419	21025	27711	29572	30783	31993	30817	34247	31410
789	Not assigned or no longer used	73223	68907	81785	82129	65317	53936	39824	36572	30845
746	Textile arts	25431	27902	39846	42885	40551	39058	34315	36683	29033
791	Public performances	25355	23240	30165	37187	35966	34799	31029	32736	28085
973	United States	32823	31155	42087	42253	33891	29942	28671	30198	25967
914	Geography of & travel in Europe	26029	21671	22858	22489	20251	18843	18709	24124	25444
306	Culture & institutions	20867	20265	27332	29848	29218	28702	25989	28742	24006
599	Mammalia (Mammals)	18354	16843	21175	22274	20500	20242	21384	25018	23832
940	History of Europe	19949	18692	23822	24143	23512	22826	21790	22787	22698
792	Stage presentations	20546	21101	27299	28246	26509	25376	22394	22098	22229
158	Applied psychology	15175	17365	21914	22247	21172	19775	20593	25702	22020
811	American poetry in English	19182	19333	24485	25675	23911	22671	20281	23207	21159
745	Decorative arts	17982	17718	23134	25661	25695	24278	22567	23684	20007
636	Animal husbandry	17494	16494	21404	24755	21680	21436	21963	22178	19135
658	General management	17487	15602	20868	23521	22493	21660	18913	20892	19050
332	Financial economics	18573	17640	24030	25262	23902	22333	17976	18527	18254
305	Groups of people	19305	19014	21563	22664	22339	21106	19896	22861	18054

Top 25 Dewey categories 2006-2014



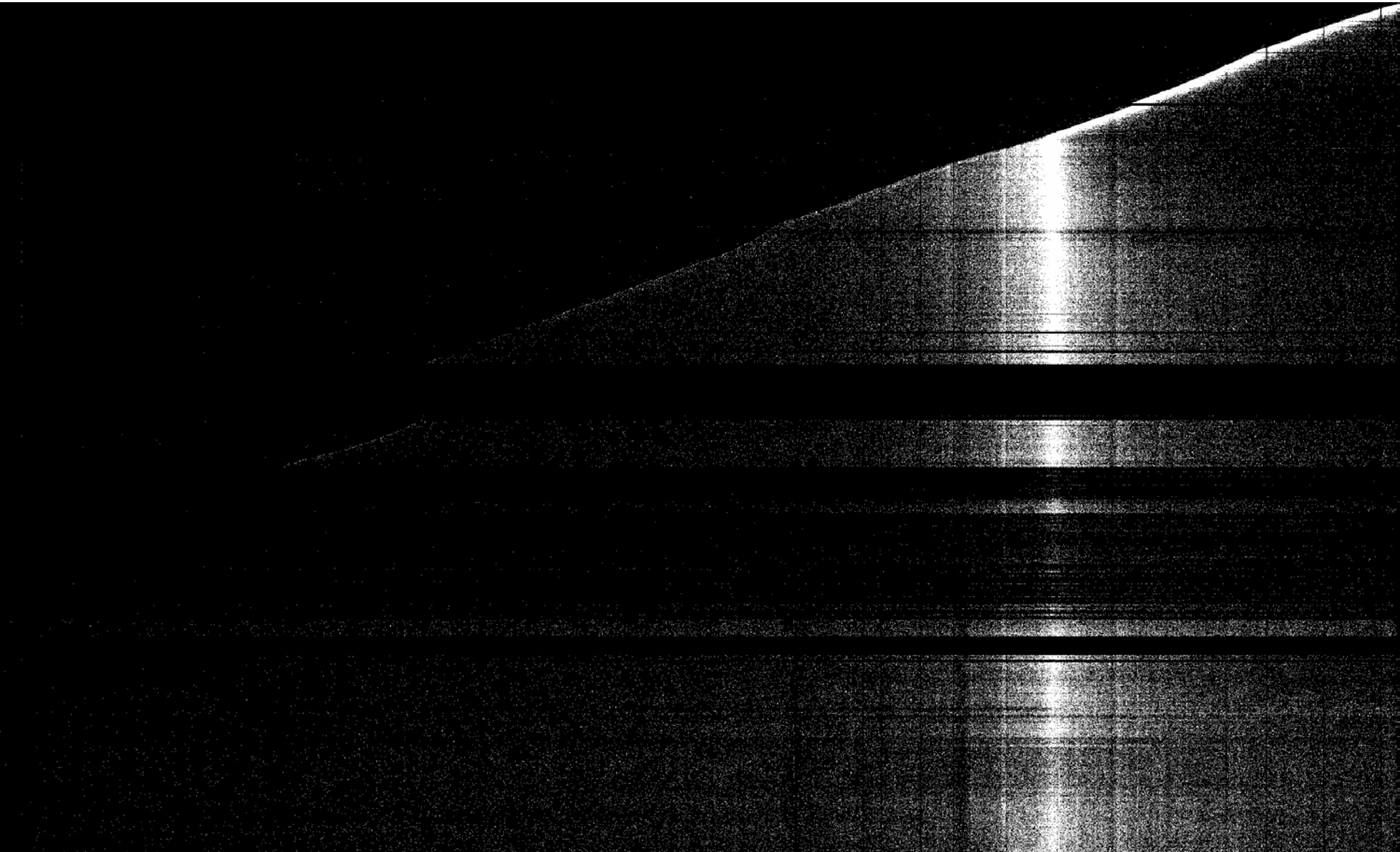


November 10, 9am to 8pm, 2005, 2006, 2007, 1008, 2009, 2010

Anomaly Detection

- Outlier/change/deviation detection
- The identification of unusual data records, that might be interesting or
- data errors that require further investigation

ItemNumber (*scalar numeric code maps acquisition history*)

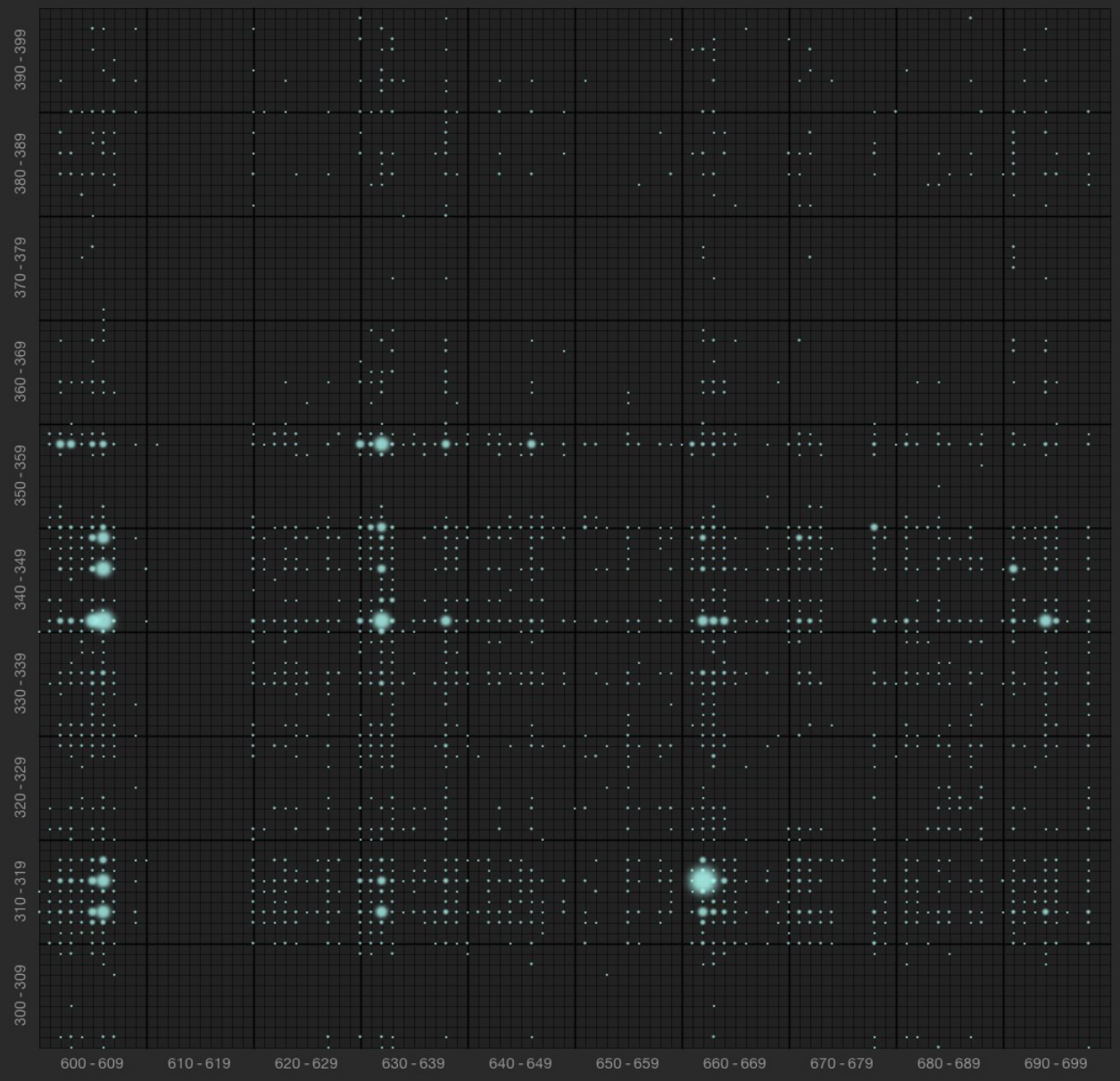


Correlate data with another

JSON: Language independent format to transmit data

Projects correlate APIS from :

- NY Times
- Instagram
- iTunes & Apple
- Twitter
- Google Books



Frequently Borrowed Together

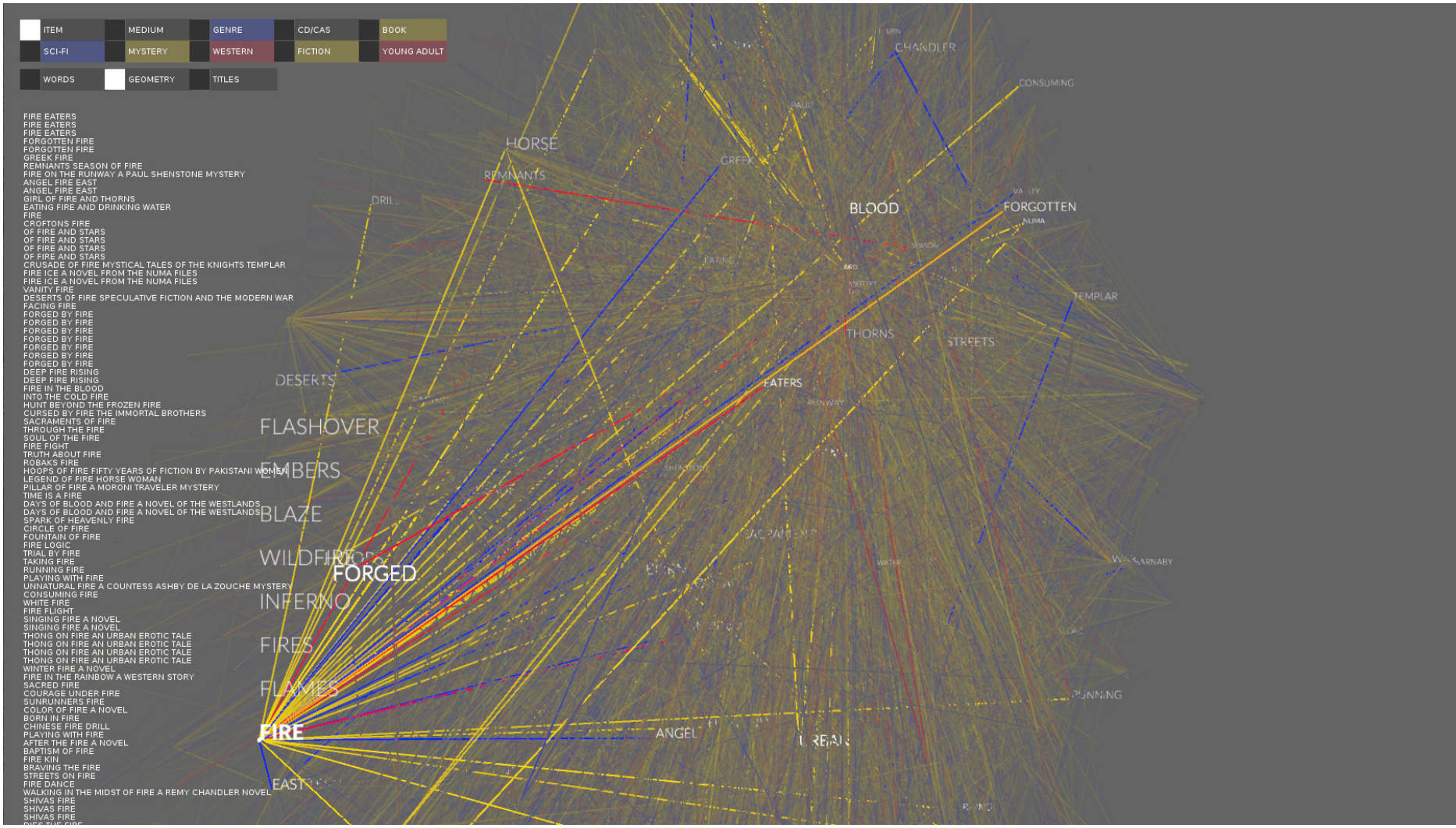
use the radio buttons to zoom

000 - 099	Computer science, information & general works
100 - 199	Philosophy and psychology
200 - 299	Religion
300 - 399	Social sciences
400 - 499	Language
500 - 599	Science
600 - 699	Technology
700 - 799	Arts & recreation
800 - 899	Literature
900 - 999	History & geography

300 - 309	Social sciences, sociology & anthropology
310 - 319	Statistics
320 - 329	Political science
330 - 339	Economics
340 - 349	Law
350 - 359	Public administration & military science
360 - 369	Social problems & social services
370 - 379	Education
380 - 389	Commerce, communications, & transportation
390 - 399	Customs, etiquette, & folklore

600 - 609	Technology
610 - 619	Medicine & health
620 - 629	Engineering
630 - 639	Agriculture
640 - 649	Home & family management
650 - 659	Management & public relations
660 - 669	Chemical engineering
670 - 679	Manufacturing
680 - 689	Manufacture for specific uses
690 - 699	Construction of buildings

Topic Modeling using word2vec



Hannah Wolfe, http://vislab.mat.ucsb.edu/2017/p3/HannahWolfe/index.html

Associative Relationships

- Association rule learning (Dependency modeling)
- Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits - which products are frequently bought together.
- **FP-Growth Algorithm**: Frequency-Pattern uses recursive-built tree structure to show paired occurrences

Predictions of Numbers of Checkouts in Seattle Public Library

MAT 259, 2020 **Guanyu Chen**

Concept

The primary interest of this project is forecasting monthly amounts of dewey books using historical data. The dataset used here, which contain 168-month chekc-out records for different sectors from January 2006 to December 2019, is provided by Seattle Public Library database and are alll time-correlated. It is possible for us to summary statistics and graphical representations of check-out records and conduct further predictions.

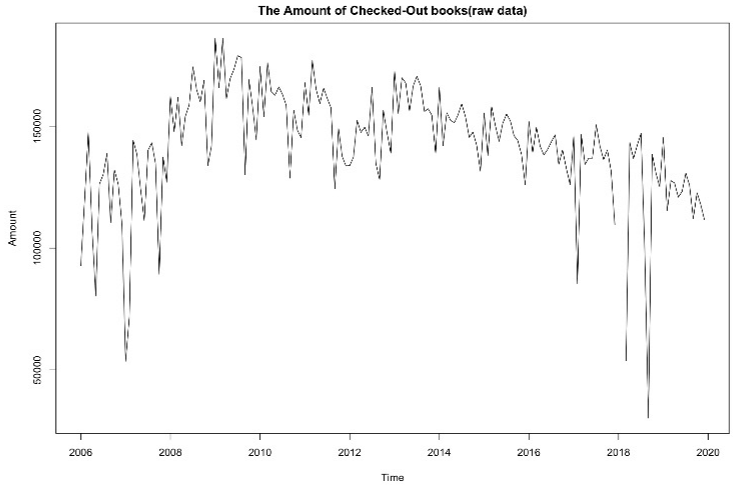
Query

```
SELECT YEAR(cout) AS Years, MONTH(cout) AS Months, SUM(CASE WHEN deweyClass != '' THEN
1 ELSE 0 END) AS Dewey FROM spl_2016.outraw WHERE itemtype LIKE '%bk' AND YEAR(cout) >=
'2006' AND YEAR(cout) <= '2019' GROUP BY YEAR(cout) , MONTH(cout); SELECT YEAR(cout) AS
Years, MONTH(cout) AS Months, SUM(CASE WHEN deweyClass > 000 AND deweyClass < 100
THEN 1 ELSE 0 END) AS D000_099, SUM(CASE WHEN deweyClass > 100 AND deweyClass < 200
THEN 1 ELSE 0 END) AS D100_199, SUM(CASE WHEN deweyClass > 200 AND deweyClass < 300
THEN 1 ELSE 0 END) AS D200_299, SUM(CASE WHEN deweyClass > 300 AND deweyClass < 400
THEN 1 ELSE 0 END) AS D300_399, SUM(CASE WHEN deweyClass > 400 AND deweyClass < 500
THEN 1 ELSE 0 END) AS D400_499, SUM(CASE WHEN deweyClass > 500 AND deweyClass < 600
THEN 1 ELSE 0 END) AS D500_599, SUM(CASE WHEN deweyClass > 600 AND deweyClass < 700
THEN 1 ELSE 0 END) AS D600_699, SUM(CASE WHEN deweyClass > 700 AND deweyClass < 800
THEN 1 ELSE 0 END) AS D700_799, SUM(CASE WHEN deweyClass > 800 AND deweyClass < 900
THEN 1 ELSE 0 END) AS D800_899, SUM(CASE WHEN deweyClass > 900 AND deweyClass < 1000
THEN 1 ELSE 0 END) AS D900_999 FROM spl_2016.outraw WHERE itemtype LIKE '%bk' AND
YEAR(cout) >= '2006' AND YEAR(cout) <= '2019' GROUP BY YEAR(cout) , MONTH(cout);
```

Preliminary sketches

We can easily check that numbers of checkout books in Seattle Public Library increased from 2006 to 2009 and decrease from 2010 to 2020. The fluctuations of numbers of checkout between each year look kinds of similar, which indicates similar yearly or monthly variance over time. It is possible to decompose time-series data into three components to investigate more information of data.

For the decline of numbers of books' checkouts, a main supposed reason is that with more mature of digital reading and free online-books, readers did not have to borrow books from libraries.



Also, when checking numbers of checkouts in different dewey classes, we find that each checkouts kept similar volumes in each year and there was a hugh drop in 2018.

Visualizing the Items Acquired by Seattle Public Library over Time

MAT 259, 2019 **Jiaheng Tang**

Concept

I'm interested in one particular metadata - **itemNumber**. It is a unique and incrementing id that gets assigned and incremented each time a new item gets acquired. I want to plot the **itemNumber** as well as the **deweyClass** of each item over time to see if there are any interesting patterns.

Query

```
SELECT DISTINCT
    i.itemNumber,
    i.bibNumber,
    it.itemType,
    t.title,
    b.deweyClass
FROM
    spl_2016.itemToBib AS i
```

This query returns **itemNumber**, **bibNumber**, **itemType**, **title** and **deweyClass** of all the items sorted by itemNumber. It takes about 29 seconds and returns 2,391,281 results.

It's worth noting that there are duplicates with the same item number, and only differs slightly in their titles. I removed them later when loading into processing. After removing the duplicates, there remain 2,326,621 records.

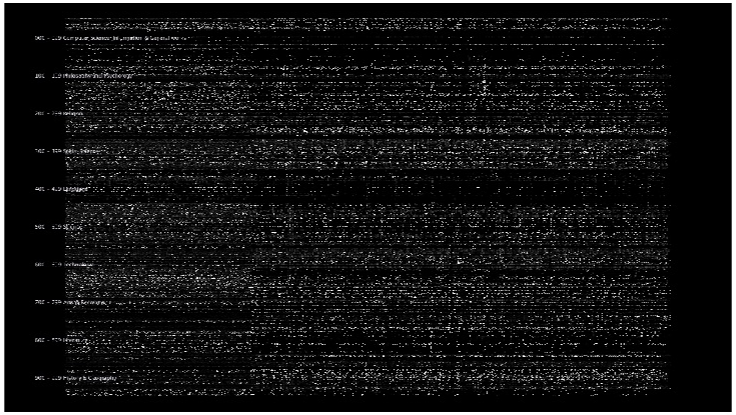
Preliminary sketches and process

I then used a python script to substitute the itemType of each record to book or media or misc based on their original itemType. Here is the python script I wrote.

```
import csv

MEDIA = {'accas', 'arcas', 'bccas', 'jccas', 'jrcas', 'accd',
'arcd', 'jccd', 'jrcd', 'accdrom', 'arcdrom',
         'bccdrom', 'drcdrom', 'jccdrom', 'acdisk', 'ardisk',
'jrdisk', 'acdvd', 'ardvd', 'bcdvd', 'jcdvd', 'jrdvd', 'xrcdrom',
'acrec', 'arrec', 'jcrec', 'jrrec', 'ucflpdr', 'acvhs', 'alvhs',
'arvhs', 'bcvhs', 'blvhs', 'jcvhs', 'jlvhs', 'jrvhs', 'xrvhs',
```

I started with a simple black background and plotting on x-y plane with deweyClass on the y-axis and item number on the x-axis. Since the item number is merely a number that gets incremented over time, it's hard to get the exact time that the item was bought.



I then changed the background color to white and decided to use three different colors to represent three categories, book, media or misc. I used the color from Maroon 5's album cover "Red Pill Blues".

I was interested in the anomaly of repeat titles in the database of items. My first query revealed that the most commonly repeated titles with unique bibNumbers were simply “ “ and “test”.

```
SELECT spl_2016.subject.subject, COUNT(spl_2016.subject.subject) as count
FROM spl_2016.title
INNER JOIN spl_2016.transactions ON spl_2016.title.bibNumber = spl_2016.transactions.bibNumber
AND title = 'test'
INNER JOIN spl_2016.subject ON spl_2016.subject.bibNumber = spl_2016.transactions.bibNumber
GROUP BY spl_2016.subject.subject
ORDER BY count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

subject	count
Historical films	574
San Francisco Calif Drama	574
Video recordings for the hearing impaired	574
Dancers Drama	574
Feature films	574
Fiction films	574
Gay men Drama	574
Science fiction	204
Conspiracies Fiction	203
Education Fiction	203

Result 19 | Read Only

I became curious about the nature of these ‘test’ entries. My next query was into all subjects of entries titled “test”. It appeared to me that subject lines seemed only partially random. They appeared arbitrary and telling.

```
SELECT spl_2016.subject.subject, spl_2016.transactions.checkOut, spl_2016.transactions.checkIn
FROM spl_2016.transactions
INNER JOIN spl_2016.subject ON spl_2016.subject.bibNumber = spl_2016.transactions.bibNumber
INNER JOIN spl_2016.title ON spl_2016.transactions.bibNumber = spl_2016.title.bibNumber
WHERE spl_2016.title.title = 'test'
ORDER BY CONCAT(IF(spl_2016.transactions.checkIn != NULL, spl_2016.transactions.checkIn, 10000), ' ', spl_2016.transactions.checkOut) ASC;
```

Limit to 1000 rows

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

subject	checkOut	checkIn
Immigrants Fiction	1970-01-01 00:00:00	2008-05-07 12:05:00
Education Fiction	1970-01-01 00:00:00	2008-05-10 17:30:00
Educational tests and measurements Fiction	1970-01-01 00:00:00	2008-05-14 17:12:00
Political corruption Fiction	1970-01-01 00:00:00	2008-06-06 16:18:00
Conspiracies Fiction	1970-01-01 00:00:00	2008-05-16 12:26:00
Political corruption Fiction	1970-01-01 00:00:00	2008-05-17 12:00:00
Education Fiction	1970-01-01 00:00:00	2008-05-14 17:12:00
Immigrants Fiction	1970-01-01 00:00:00	2008-06-06 16:18:00
Immigrants Fiction	1970-01-01 00:00:00	2008-05-17 12:00:00
Political corruption Fiction	1970-01-01 00:00:00	2008-05-03 10:42:00

Result 16 | Read Only

My question became: “Why would someone put a particular subject line for test titles”. So my next step was to query the timestamps of check ins and check outs associated with each subject line. I realized that there were only a few subjects, although there were nearly 5000

MAT259 - Assignment 1

Dongyu Meng
Jan 15 - 2020

In this assignment I tried to find out if there is a correlation between the theme of books people borrow and the time in the day these books get borrowed. Like, is it true that people tend to borrow technical books in the morning and novels at night? Inquiries like this give insight to the activity patterns of different readers or even professions.

The basic SQL query is the following:

```sql
select deweyClass, count(*), hour(cout) from spl_2016.outraw
where cout Between '2015-01-01' AND '2016-01-01'
and deweyClass != ""
and itemtype = "acbk"
group by deweyClass, hour(cout)
```

In the records in the year of 2015, I counted the number of books borrowed group by the book's Dewey class and the time in the day (discretized to hours) the book got borrowed. This query gave me data like the following:

	deweyClass	count(*)	hour(cout)
1	001	5	10
2	001	9	11
3	001	9	12
4	001	15	13
5	001	17	14
6	001	10	15
7	001	11	16
8	001	11	17
9	001	7	18
10	001	4	19
11	001.01	1	12
12	001.01	2	13

I further processed the data with Python.

First, I unified the Dewey class number as shown above to the major class number only (001.01 -> 001) and aggregated the count aligning to the (deweyClass, hour) tuple.

QUERY METHODS

QUERY 1: *Find objects that contain titles like "feminism" or "feminist"*

In this query, my objective is to pull items that contain phrases similar to "feminism" and "feminist" in the title. I obtained the data from the spl_2016.outraw and printed the title, bibNumber, counts, and itemtype. I assigned the variable counts to aggregate an item's frequency based on the bibNumber. Overall, 557 rows returned.

```
SELECT
      title, bibNumber, COUNT(bibNumber) AS counts, itemtype
FROM
      spl_2016.outraw
WHERE
      (title LIKE "%feminism%")
      OR (title LIKE "%feminist%")
GROUP BY
      bibNumber, itemtype, title
ORDER BY
      counts DESC
```

QUERY 2: *Create a monthly subset for the top 10 items from 2006-1 to 2017-12*

In this second query, my objective is to further examine the top 10 results and explore checked-out dates by months from 2006-1 to 2017-12 (year-month). The SQL input for this query is found at the end of the report.

title	bibNumber	counts	2006-1	2006-2	2006-3	2006-4	2006-5	2006-6	2006-7
Bad feminist essays	3008726	943	0	0	0	0	0	0	0
We should all be feminists	3077814	629	0	0	0	0	0	0	0
Feminist sweepstakes	2290181	409	3	7	10	9	4	7	6
Mysteries of the Middle Ages the rise of feminism...	2368106	402	0	0	0	0	0	0	0
Feminist fight club an office survival manual for a...	3201929	246	0	0	0	0	0	0	0
Dear Ijeawele or A feminist manifesto in fifteen su...	3255843	239	0	0	0	0	0	0	0
Feminism is for everybody passionate politics	1969986	224	1	2	4	0	2	0	2
Full frontal feminism a young womens guide to w...	2430480	204	0	0	0	0	0	0	0
Feminism is for everybody passionate politics	3122735	163	0	0	0	0	0	0	0
Enlightened sexism the seductive message that f...	2627029	143	0	0	0	0	0	0	0

MAT 259 > Project 1: Data Query

Grant McKenzie (grant.mckenzie@geog.ucsb.edu)

QUESTION:

One of the most interesting aspects of a dataset such as the one provided by the *Seattle Public Library*, is its ability (when asked the proper questions) to speak on user behavior . Given the completeness of the dataset, one can explore *trends* in the data and arguably infer user-interests based on these trends.

For this assignment I chose to explore how interest in subject matter changes over time. To do this I decided to investigate the variance in the items that users checked-out based on the Dewey Decimal Class of the medium. *Variance* is computed as it allows one to compare values between temporal units (e.g., month or day). A large variance indicates that media related to a wide range of classes were checked-out of the library, while a small (relative) variance indicates that the checkout history during the specified time period did not vary much in terms of subject (Dewey Class). It is important to note here that the Dewey Decimal Class system is not assigned to all media in the Library dataset. Given this, the results of the queries, as well as any inferences, are based on a subset of the library data.

Since Variance alone can often be misleading, it was important to also list the total number of items checked out as well as the count for each Dewey Class.

The two queries below show classification Variance by Month and by Day of the Week.

SQL QUERIES:

MONTH:

```
SELECT month, monthnum, VARIANCE(count) AS variance, count(*) as cnt, sum(count) as
sum
FROM
  (SELECT monthname(cin) as month, month(cin) as monthnum,
substring(deweyClass,1,3) as subdew, count(*) as count
  FROM
      (SELECT cin, deweyClass
      FROM spl2.inraw
      WHERE deweyClass <> '') as a1
  GROUP by subdew, month) as a2
GROUP BY month, monthnum
ORDER BY monthnum;
```

DAY OF WEEK:

```
SELECT day, daynum, VARIANCE(count) AS variance, count(*) as cnt, sum(count) as sum
FROM
  (SELECT dayname(cin) as day, dayofweek(cin) as daynum, substring(deweyClass,1,3)
as subdew, count(*) as count
  FROM
      (SELECT cin, deweyClass
      FROM spl2.inraw
      WHERE deweyClass <> '') as a1
  GROUP by subdew, day) as a2
GROUP BY day, daynum
ORDER BY daynum;
```

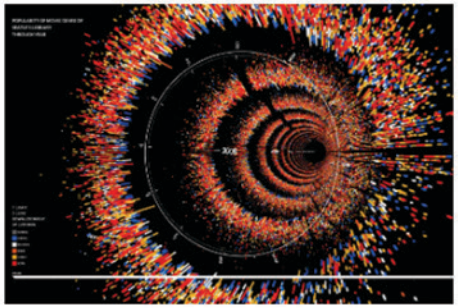
QUERY EXPLANATION:

The question I intended to answer required a two level hierarchical query (sub queries). The initial query simply returns the *Check-in Timestamp* as well as the *DeweyClass* for all records in the table *inraw* where a *DeweyClass* exists. From this query the rows are aggregated by the day of the week (using *dayname and dayofweek* functions) and the 3 digits before the decimal (class) are taken from the *DeweyClass* column using a *substring* function. Additionally, the count for all records of the specified *DeweyClass* are computed. Lastly, the day, the numerical representation of the day, the variance (of the count), the total number of unique DeweyClasses per day and the total number of records per day are reported. The above steps were repeated for month.

RESULTS:

MONTH	MONTH_NUM	VARIANCE	DDC_COUNT	TOTAL_ITEMS
January	1	625628160.2322	872	2482939
February	2	584416553.6046	886	2441234
March	3	811697936.6273	879	2918522
April	4	703141854.3939	875	2684755
May	5	666337555.9436	874	2580837
June	6	707490932.3043	875	2681675
July	7	726666001.7056	878	2711706
August	8	724558276.6155	879	2670405
September	9	496804534.169	871	2241631
October	10	604580899.4472	868	2460373
November	11	554948159.1354	879	2376709

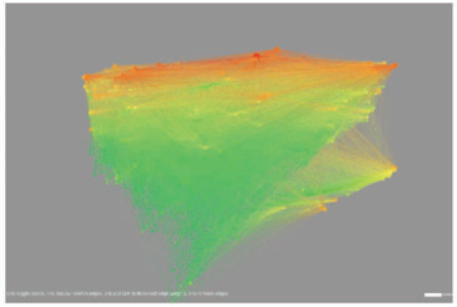
DAY	DAY_NUM	VARIANCE	DDC_COUNT	TOTAL_ITEMS
Sunday	1	397993479.8416	874	1969910
Monday	2	2660434015.5491	897	5269587
Tuesday	3	2588466694.0293	883	5128994
Wednesday	4	2527227004.2695	893	5101534
Thursday	5	2218511032.4472	888	4747085
Friday	6	1514679123.5407	890	3929554
Saturday	7	1993389576.0715	888	4525020



M259 Projects
Data Visualization

2020

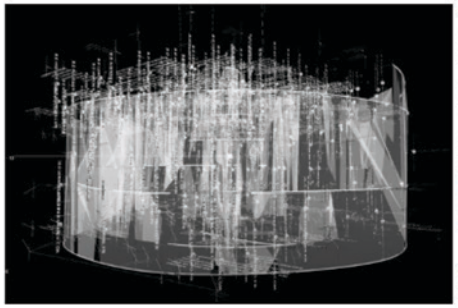
SPL data, predictive analysis, COVID-19, spatial memory in virtual maze, Volga Germans, SF Eviction Rates and gentrification, popularity of BiliBili, game sales, Air BNB, 2016 presidential election data, etc.



M259 Projects
Data Visualization

2019

SPL data, Fen Shui prevalance, statistics in cooking, query variances, travel segregation, Behance, Foreign news at BPI, US lighting strikes, Hate groups, Health in counties, etc.



M259 Projects
Data Visualization

2018

SPL data, particle visualization of movies, travel books, supple & demand, spherical harmonics, StarWars, Indi Rock Vs Soul, global hazards, NYC taxi data, TED text mining, endangered species trade, EUR/US rate, etc.

Experimental Visualization Lab

The Experimental Visualization Lab is one of 8 dedicated research labs in the Media Arts & Technology arts-engineering program located on the 2nd floor of Elings Hall (California Nanosystems Institute) at the University of California, Santa Barbara.

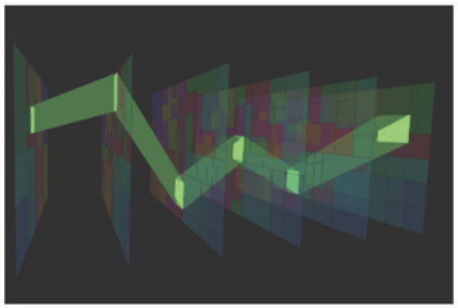
The lab focuses on creative explorations in the fields of data visualization, visual language, machine vision, computational photography, interactive digital installations and related directions to explore the intersections of computation and visualization in both the arts and engineering.

The lab is directed by Professor George Legrady, an internationally exhibited, multi-disciplinary artist and scholar with projects realized in interactive digital media installations, and computationally generated data and photographic-based visualizations. He is former chair (2013-2017) of the **Media Arts & Technology** program, and is affiliated on campus with the **Department of Art**, the **Data Science Initiative**, the **Center for Digital Games Research**, and the **Center for Information Technology and Society**. Legrady is a Guggenheim Fellow with research funded by the Creative Capital Foundation, National Science Foundation, Robert. W Deutsch Foundation, the Center for Nanotechnology in Society at UCSB, and others.

Contact

George Legrady
Director
lab 2611, Elings Hall
glegrady [a] ucsb.edu

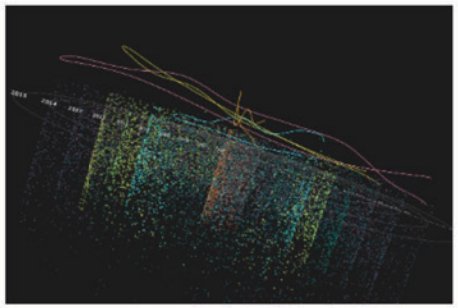
Laura Cheung
MAT, managing officer
Pheips Hall 3309
e.mat-info [a] mat.ucsb.edu
(805) 893-5244



M259 Projects
Data Visualization

2017

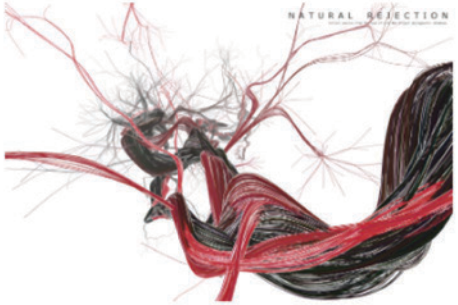
SPL data, lost and forgotten books, high-school classics, Michael Jackson, title-length to checkout-length, correlate weather data with library checkouts, MALLET topic modeling, Crunchbase trends, Wikigraph, space-time path, Washington DC bike circulation, Reddit image data, European Power Exchange, etc.



M259 Projects
Data Visualization

2016

SPL data, acquisition frequency, David Bowie over time, banned titles, film adaptation, shared bib-number, bigram analysis, seven deadly sins, travel, wars, global warming, occurrence of plant bugs, Instagram data, S&P 500, aggregate herd movement, EXIF global tracking, etc.



M259 Projects
Data Visualization

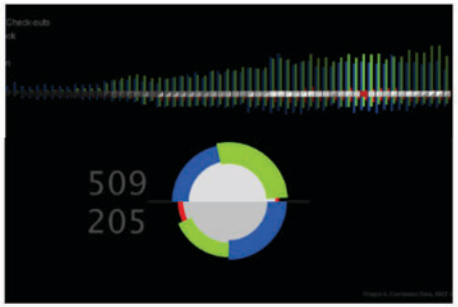
2015

SPL data, urban agriculture, Restricted Boltzmann Machine, mapping controversy, diurnal and seasonal cycles, outliers, checkouts & weather, financial crisis, reorder matrix, Gothic topic treemap, spatial-temporal, k-means clustering, autobiography, spider diagram, financial market chaos, etc.

M259 Projects
Data Visualization

2014

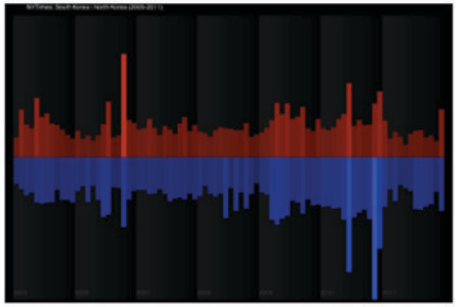
SPL data, Ukraine conflict, barcode anomalies, multi-dimentional scaling, 2D spatial treemap, checkout variance, Manga culture, etc.



M259 Projects
Data Visualization

2013

SPL data, Twitter, Facebook correlation, top six Seattle foods, NYTimes/SPL data, Jane Austin statistics, global travel, etc.



M259 Projects
Data Visualization

2012

SPL data, Catcher in the Rye, linear frequency, Last.fm correlation, geo-tagged Flickr data, North/South Korea, mortage collapse, FP-Tree association, etc.

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

Visual Design: Explore the full potential of visual language beyond functionality

Cinematic/Literary: Explore data pattern as narrative development

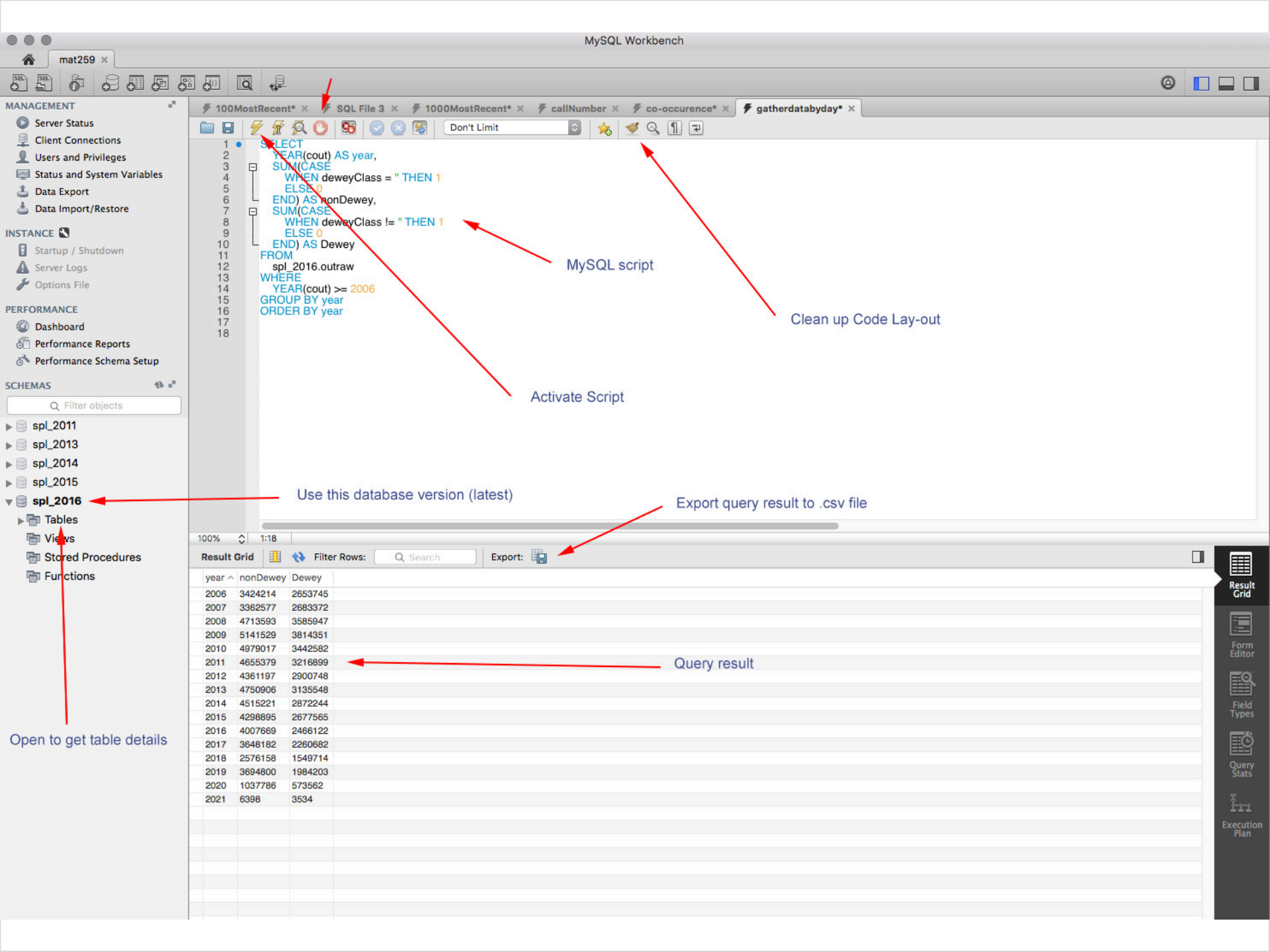
Data Analysis & Application

*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

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


MySQL Workbench
mat259
MANAGEMENT
Server Status
Client Connections
Users and Privileges
Status and System Variables
Data Export
Data Import/Restore
INSTANCE
Startup / Shutdown
Server Logs
Options File
PERFORMANCE
Dashboard
Performance Reports
Performance Schema Setup
SCHEMAS
Filter objects
spl_2011
spl_2013
spl_2014
spl_2015
spl_2016
Tables
Views
Stored Procedures
Functions
100MostRecent*
SQL File 3
1000MostRecent*
callNumber
co-occurence*
gatherdatabyday*
Don't Limit
SELECT
YEAR(cout) AS year,
SUM(CASE
WHEN deweyClass = '' THEN 1
ELSE 0
END) AS nonDewey,
SUM(CASE
WHEN deweyClass != '' THEN 1
ELSE 0
END) AS Dewey
FROM
spl_2016.outraw
WHERE
YEAR(cout) >= 2006
GROUP BY year
ORDER BY year
MySQL script
Clean up Code Lay-out
Activate Script
Use this database version (latest)
Export query result to .csv file
Query result
Open to get table details
100%
1:18
Result Grid
Filter Rows:
Search
Export:
year
nonDewey
Dewey
2006 3424214 2653745
2007 3362577 2683372
2008 4713593 3585947
2009 5141529 3814351
2010 4979017 3442582
2011 4655379 3216899
2012 4361197 2900748
2013 4750906 3135548
2014 4515221 2872244
2015 4298895 2677565
2016 4007669 2466122
2017 3648182 2260682
2018 2576158 1549714
2019 3694800 1984203
2020 1037786 573562
2021 6398 3534
Result Grid
Form Editor
Field Types
Query Stats
Execution Plan

Seattle Public Library Data

George Legrady © 2021
Experimental Visualization Lab
Media Arts & Technology
UC Santa Barbara