M259 Visualizing Information

Elings 2611 Tues-Thurs 12:00-1:50pm

M259 Visualizing Information

George Legrady, Instructor
Yoon Chung Han, Teaching Assistant

Special Presentations:
Qian Liu, Data Visualization consultant
Karl Yerkes, Knowledge Discovery consultant

Course Focus

"*Visualizations Give Meaning to Information*"

- A practice-based course includes lectures, some research and visualization proejcts
- Promote an experimental approach
- Goals:
  - Advance graphical skills in translating abstract data into visualization
  - Finetune skills in data search, algorithmic data processing, and visual language syntax.

Course Format

TUES: Lectures, visualization analysis

THUR: Technical lab

BLOG: Post your concepts, sketches, share algorithms, visualization techniques. Also faculty and TA feedback

WEBSITE: Course syllabus, references, and project results

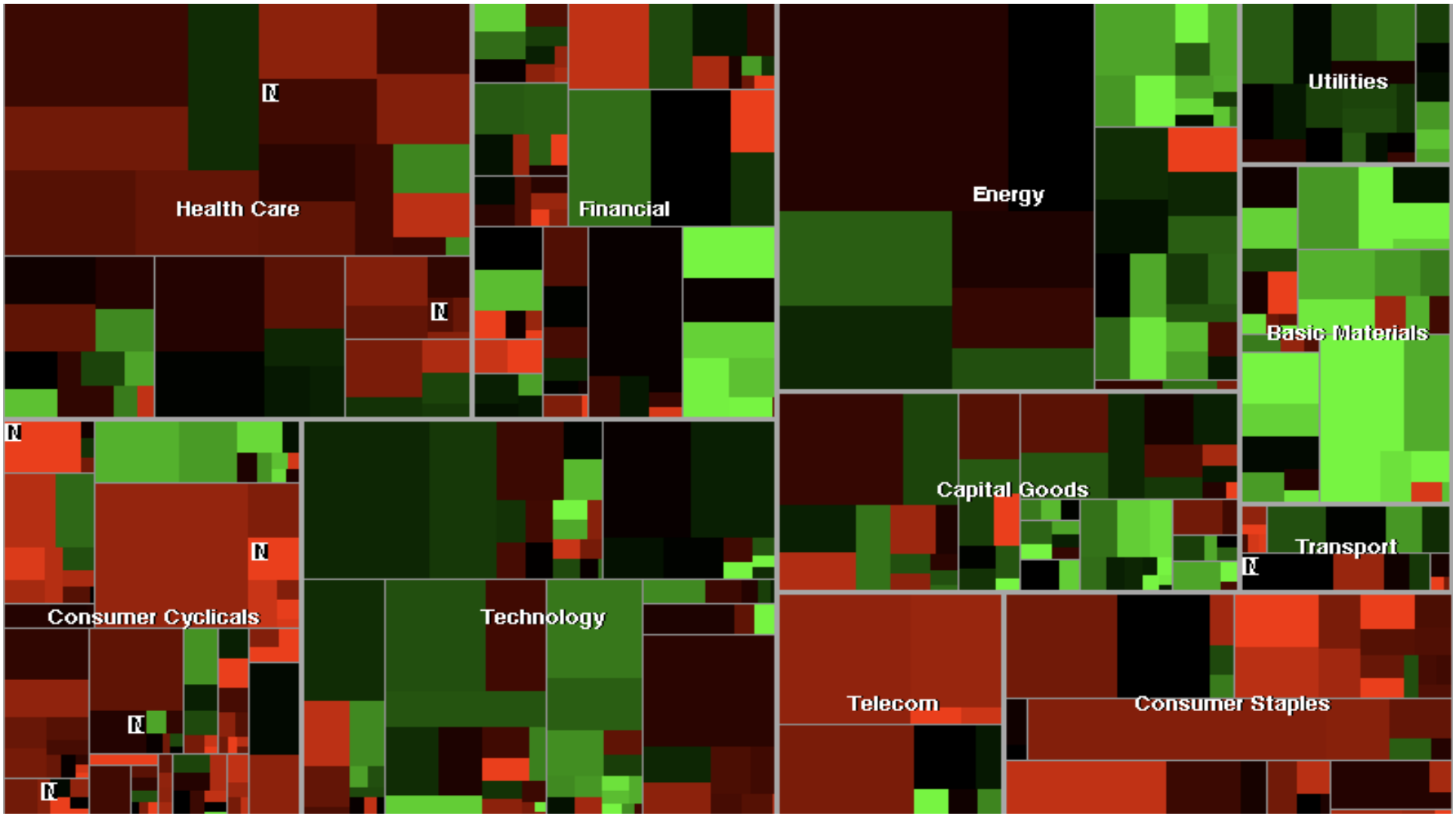
Student project presentations may be on Tues or Thurs

Meaning in Data is Through Representation

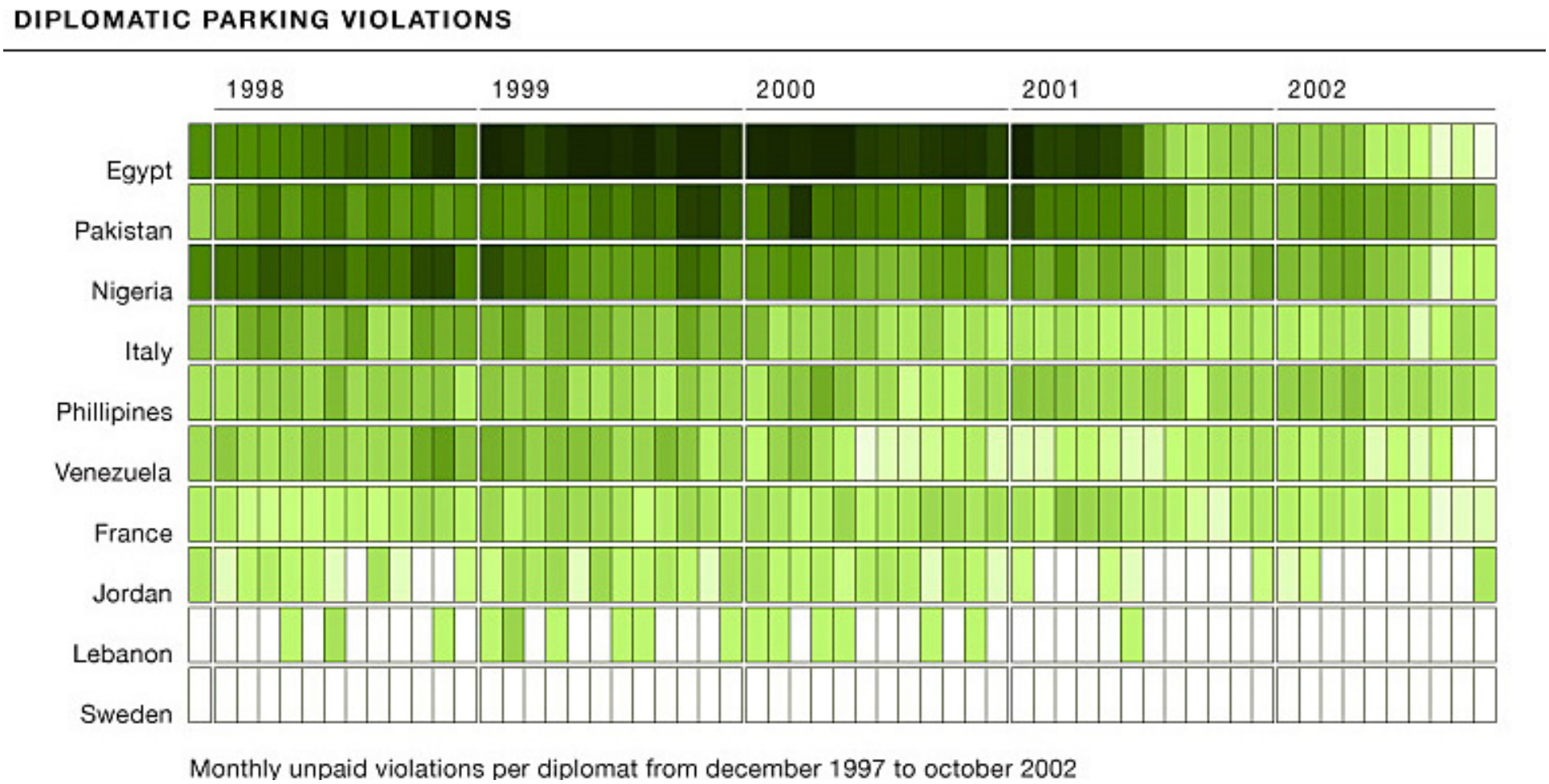
Data is neutral but representation is a cultural process. It is how you represent data that creates meaning!

- The focus of the course is learning how to work with visual language.
- To make this happen in our very short time we will all use the same dataset.
- You will be able to take the knowledge picked up through the use of the course dataset and apply it afterwards to your own specific data and needs.
- Using the same dataset will allow us to 1) use existing modules developed in earlier courses, 2) compare results and learn from each other

TreeMap: SmartMoney (Schneiderman/Wattenberg)



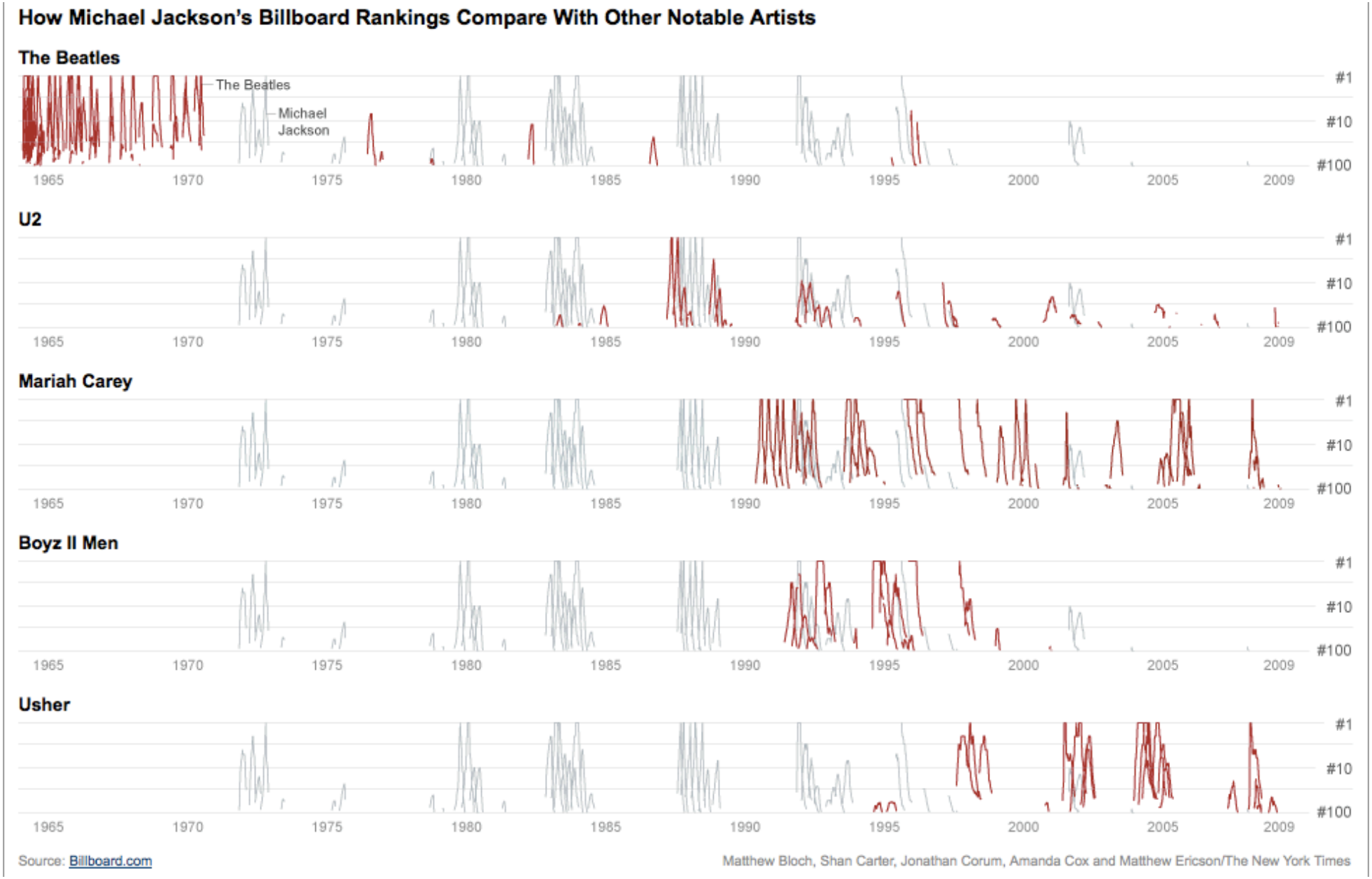
Frequency Maps Can Tell a Story (catalogtree)



Meaning in Data is Through Representation

Data is neutral but representation is a cultural process. It is how you represent data that creates meaning!

- The focus of the course is learning how to work with visual language.
- To make this happen in our very short time we will all use the same dataset.
- You will be able to take the knowledge picked up through the use of the course dataset and apply it afterwards to your own specific data and needs.
- Using the same dataset will allow us to 1) use existing modules developed in earlier courses, 2) compare results and learn from each other


How Michael Jackson's Billboard Rankings Compare With Other Notable Artists
The Beatles
The Beatles
Michael Jackson
U2
Mariah Carey
Boyz II Men
Usher
#1
#10
#100
1965
1970
1975
1980
1985
1990
1995
2000
2005
2009
Source: Billboard.com
Matthew Bloch, Shan Carter, Jonathan Corum, Amanda Cox and Matthew Ericson/The New York Times

Skills you will acquire

Knowledge Discovery through Data Search: Exploration of data searching for interesting results (MySQL)

Data Formulation: Data processing and algorithmic implementation for innovating data results (what patterns will emerge)

Data Visualization: Java-based programming (Processing)

Visual Language: How form, color, space, timing, movement, etc. impact on content (this is the primary goal of the course)

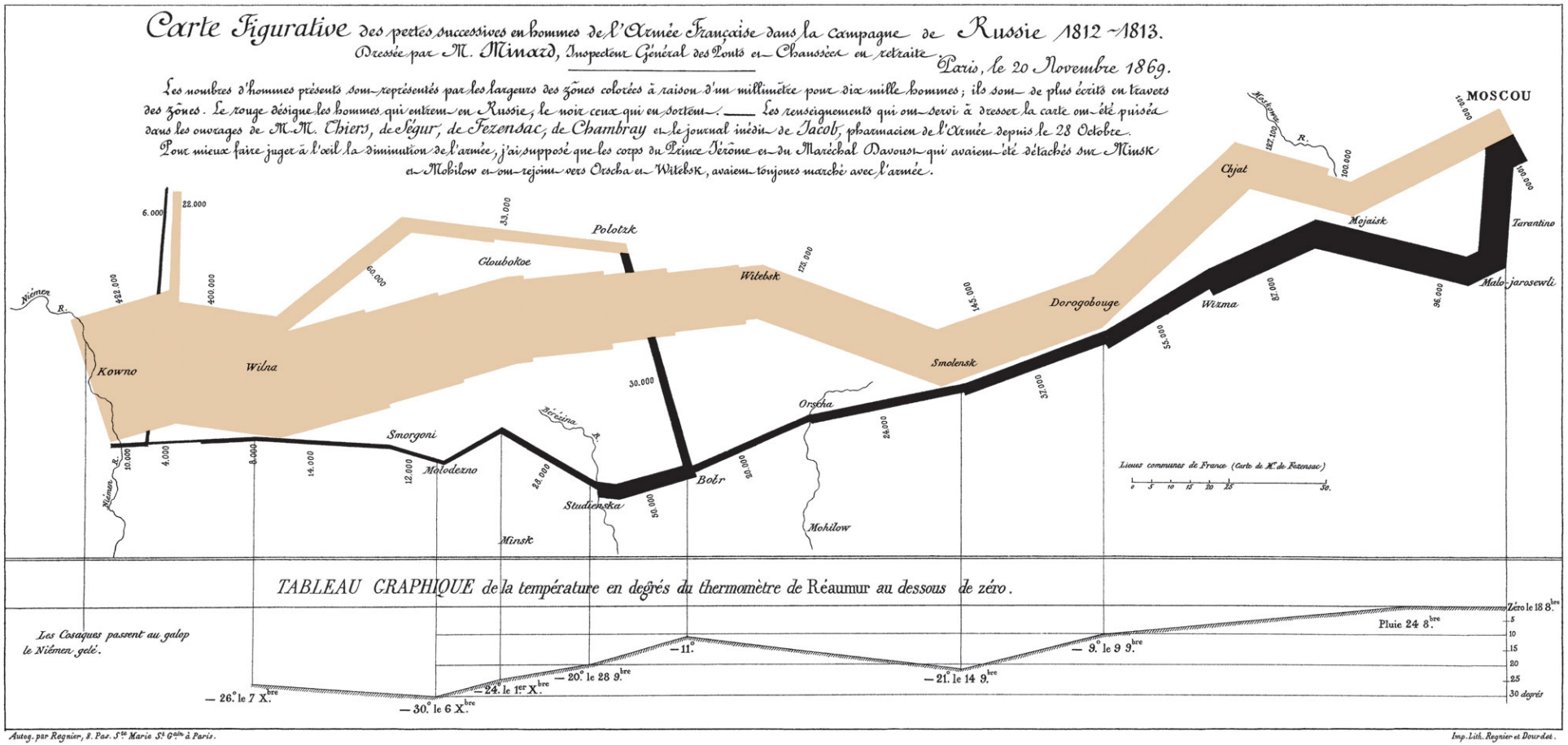
Your Contribution

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

Charles Joseph Minard (1781-1870)

- French civil engineer and pioneer in the field of information graphics



Integrate Your Expertise

CS: Integrate complex algorithms to visualization

Statistics: Implement statistical probability problems to data analysis and visualization

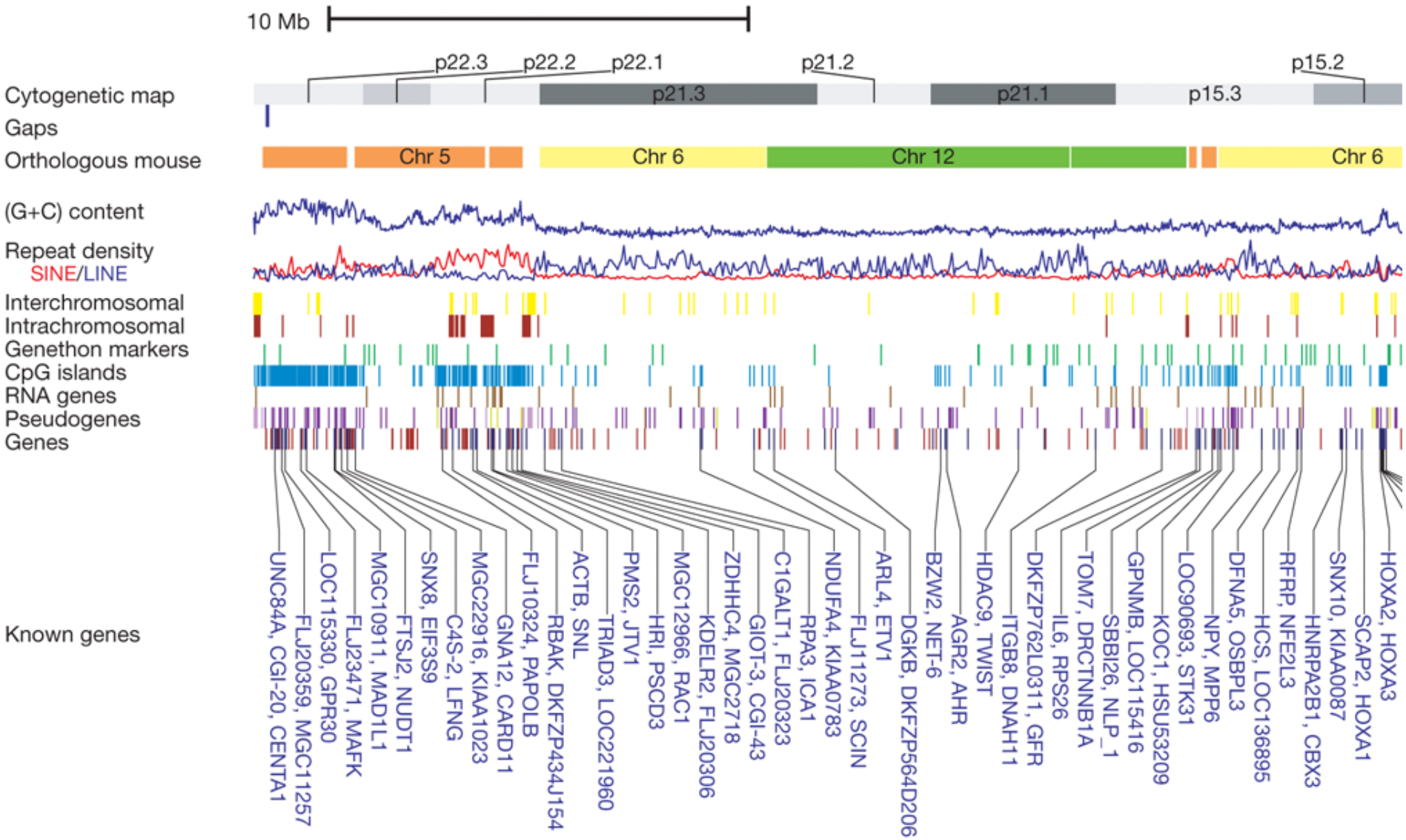
Sound/Signal processing: Consider data as signal and explore translation between sonic, signal and visual patterns

Cinematic/Literary: Explore data pattern as narrative development

Social Science: Identify cultural patterns, changes, transformations

Geography: Explore spatial mapping

DNA Sequential Map



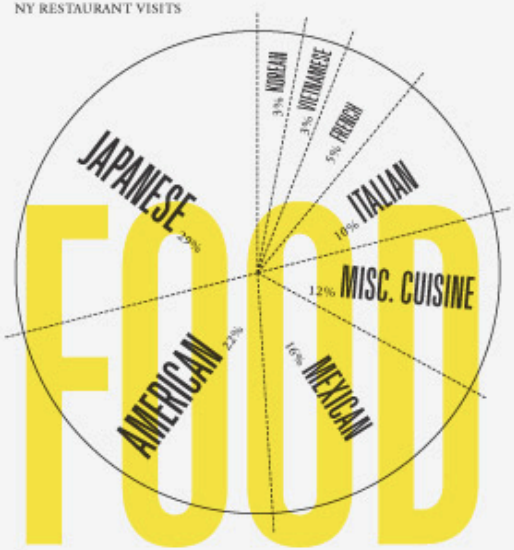
BEST MEAL

PER SE

5 MOST FREQUENTED RESTAURANTS

01. TAKAHACHI AVENUE A
02. TAKAHACHI DUANE STREET
03. CHIPOTLE 8TH STREET
04. EL PORTAL ELIZABETH STREET
05. EDWARD'S WEST BROADWAY

NY RESTAURANT VISITS



MISC. CUISINE TYPES

GREEK
SPANISH
GERMAN
IRISH
BRAZILIAN
THAI
CUBAN
MOROCCAN
SYRIAN
CHINESE
INDIAN

SOCIAL DINNERS : SOLO DINNERS

1.15:1

ANIMALS EATEN

LEGS	WINGS	FINS	SHELLS	NO SHELLS
COW	CHICKEN	COD	CLAM	JELLYFISH
DEER	DUCK	EEL	CRAB	OCTOPUS
HORSE	TURKEY	FLUKE	LOBSTER	SQUID
KANGAROO		HERRING	MUSSEL	
LAMB		MACKEREL	OYSTER	
PIG		MANTA RAY	SCALLOP	
RABBIT		POMPANO	SEA URCHIN	
		RED SNAPPER	SHRIMP	
		SALMON	SNAIL	
		TILAPIA		
		TUNA		
		YELLOWTAIL		

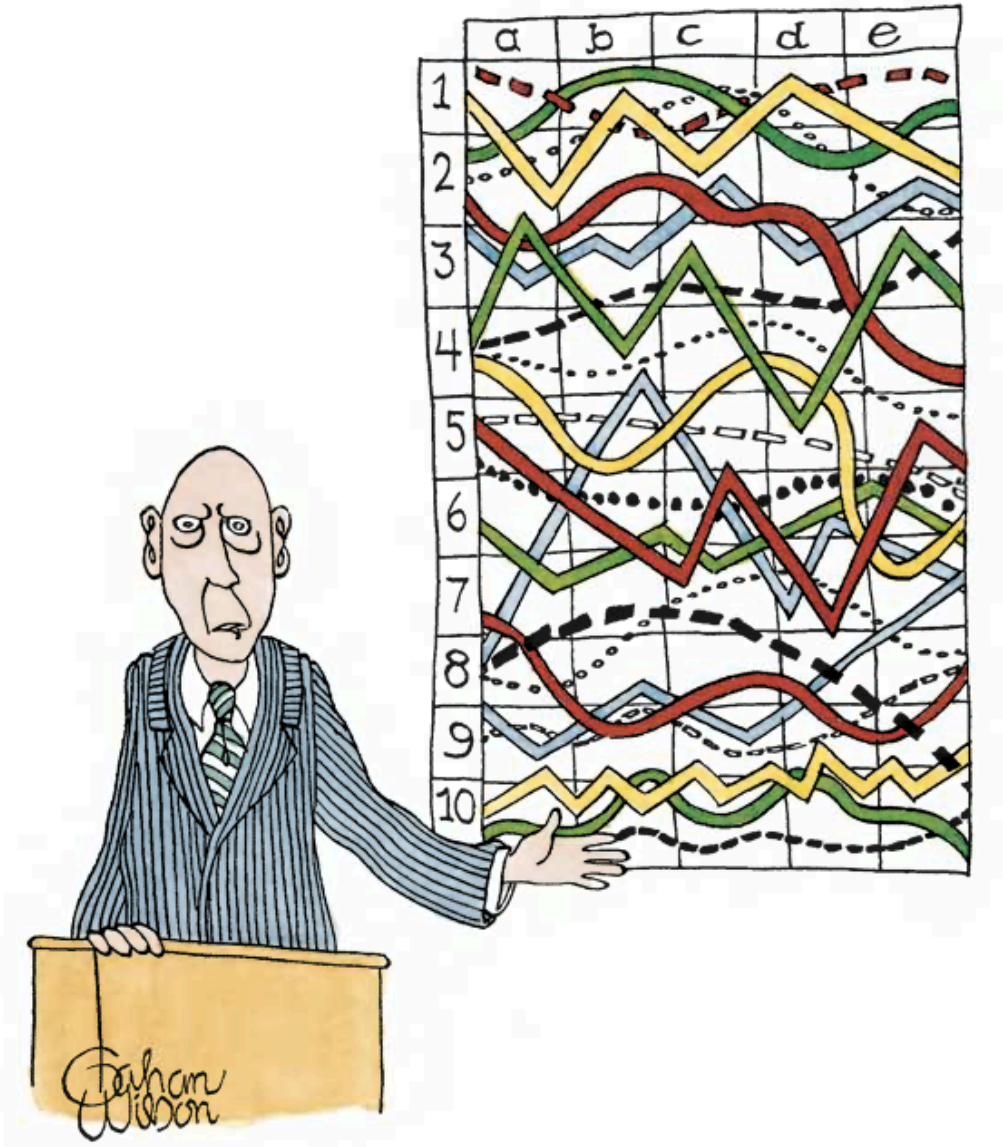
An Interdisciplinary Process

Expertise Differentations:

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

Course focus: Integrated approach

Break– Questions?



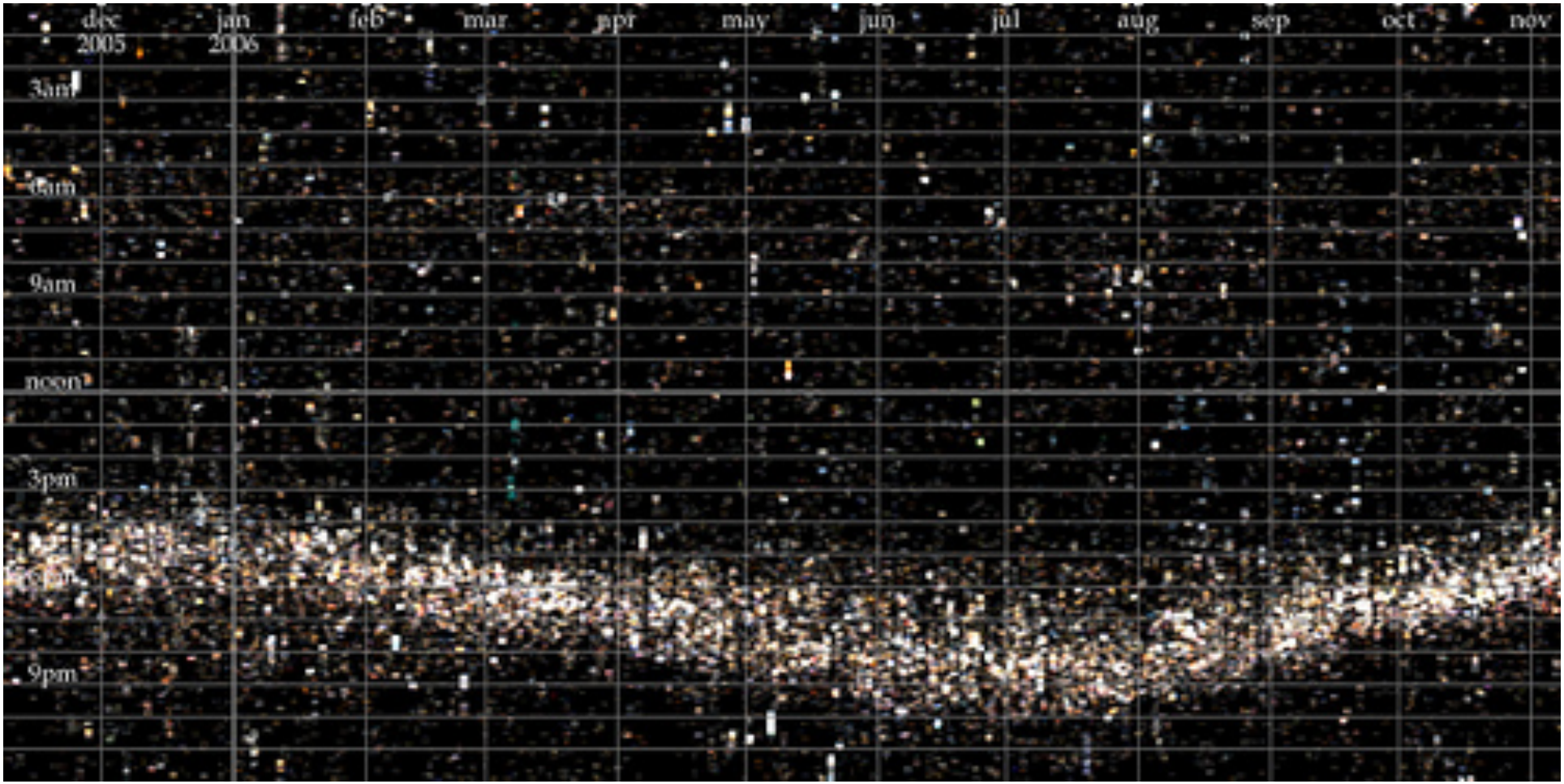
"I'll pause for a moment so you can let this information sink in."

Course Metadata

Metadata is data that describes data

- Examples:
  - **Personal Statistics**: Your name, age, gender, height, soc sec number, etc.
  - **Dewey Decimal System** (DDC)**:** To organize books according to subject
  - **EXIF** (Exchangeable image file format): Standard that specifies the format for images, sound, in digital devices: Type of camera, date/time, compression, image resolution, exposure program, focus, metering, flash, colorspace,

Linear Time Graphs



Course Metadata Format

- **Data Source**: Patrons check out books, cds, dvds from the Seattle Public Library
- Appx 30000 per day; 10 million annual; 70 million datasets since September 2005
- 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, etc.

Course Schedule

Wk1: Introduction and software setup

Wk2,3: Data Query, data analysis (MySQL)

Wk4: 1D Linear Frequency project (Processing)

Wk5,6: 2D Spatial Mapping

Wk7: Correlating 2 datasets

Wk8,9: 3D Interactive visualization

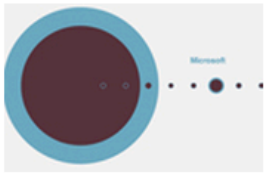
Wk10: Finalize Project(s)

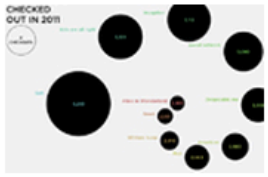
Wk11: Presentation

M259 Data Visualization - 2012

One-Dimensional Frequency Graph Assignment

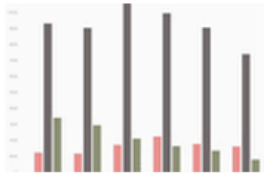

Anis Haron


HanYoon Jung



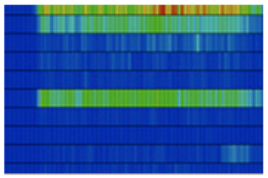
RJ Duran


Ankit Srivastava

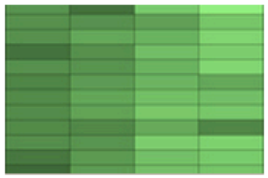

David Gordon

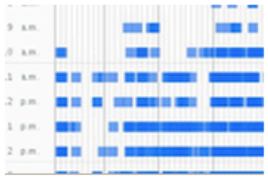
Two-Dimensional Spatial Map Assignment


Ankit Srivastava


HanYoon Jung


RJ Duran

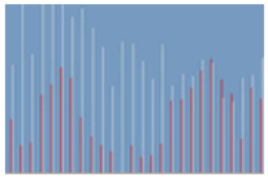

Dallas Mercer

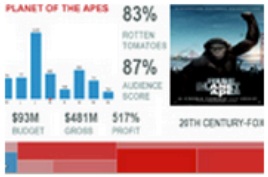

David Gordon

External Correlation or FP-Growth Algorithm Assignment


RJ Duran

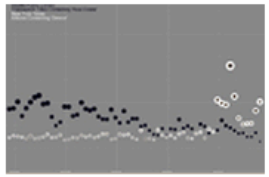

Anis Haron


Dallas Mercer

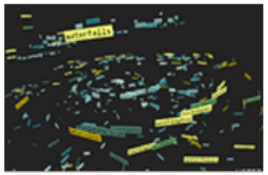


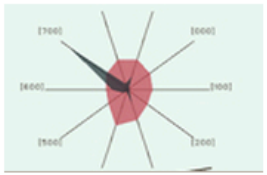
Ankit Srivastava


HanYoon Jung

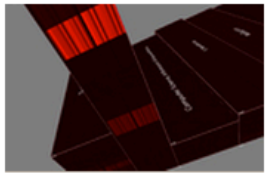

David Gordon

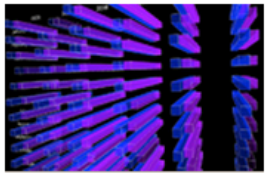
Three-Dimensional Interactive Assignment

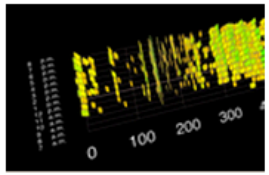

RJ Duran


Anis Haron


HanYoon Jung


Ankit Srivastava


Dallas Mercer



David Gordon

Wk1-3: Data Query, Knowledge Discovery, MySQL

Question: How has the popularity of JD Salinger's book "Catcher in the Rye" or the lead character "Holden Caulfield" changed during the period 2005 to 2011, separated by year.

Process: A simple single query visualization. Minimal design, and use of color to fully utilize available space.

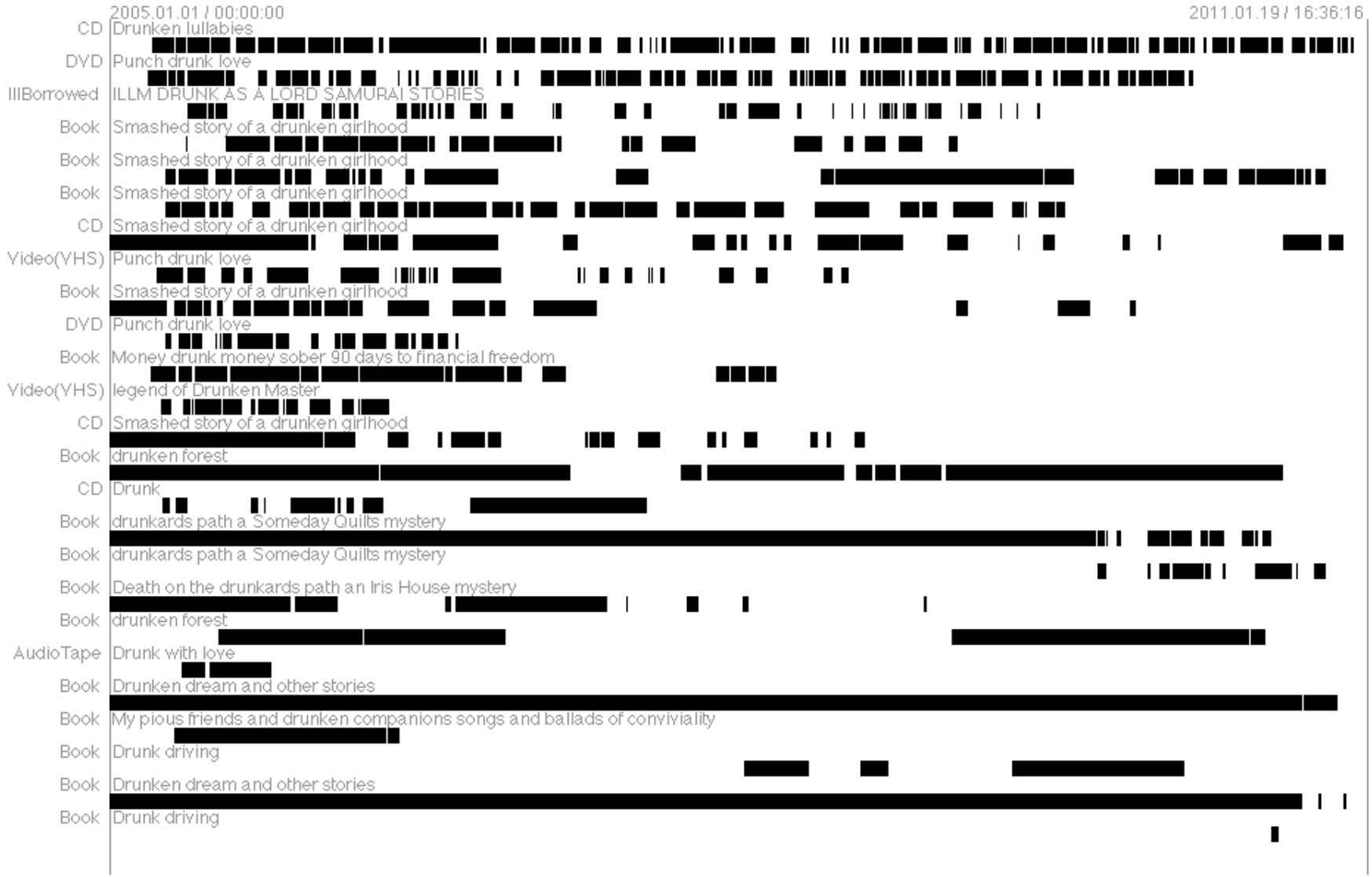
Query: select year(ckoutDateTime) as year,count(*) from trans_2005_thru_2011 where year(ckoutDateTime)>2004 AND (title like "%the catcher in the rye%" OR title like "%Holden Caulfield%") group by year order by year;

2005
11
2006
20
2007
16
2008
62
2009
75
2010
89
2011
12

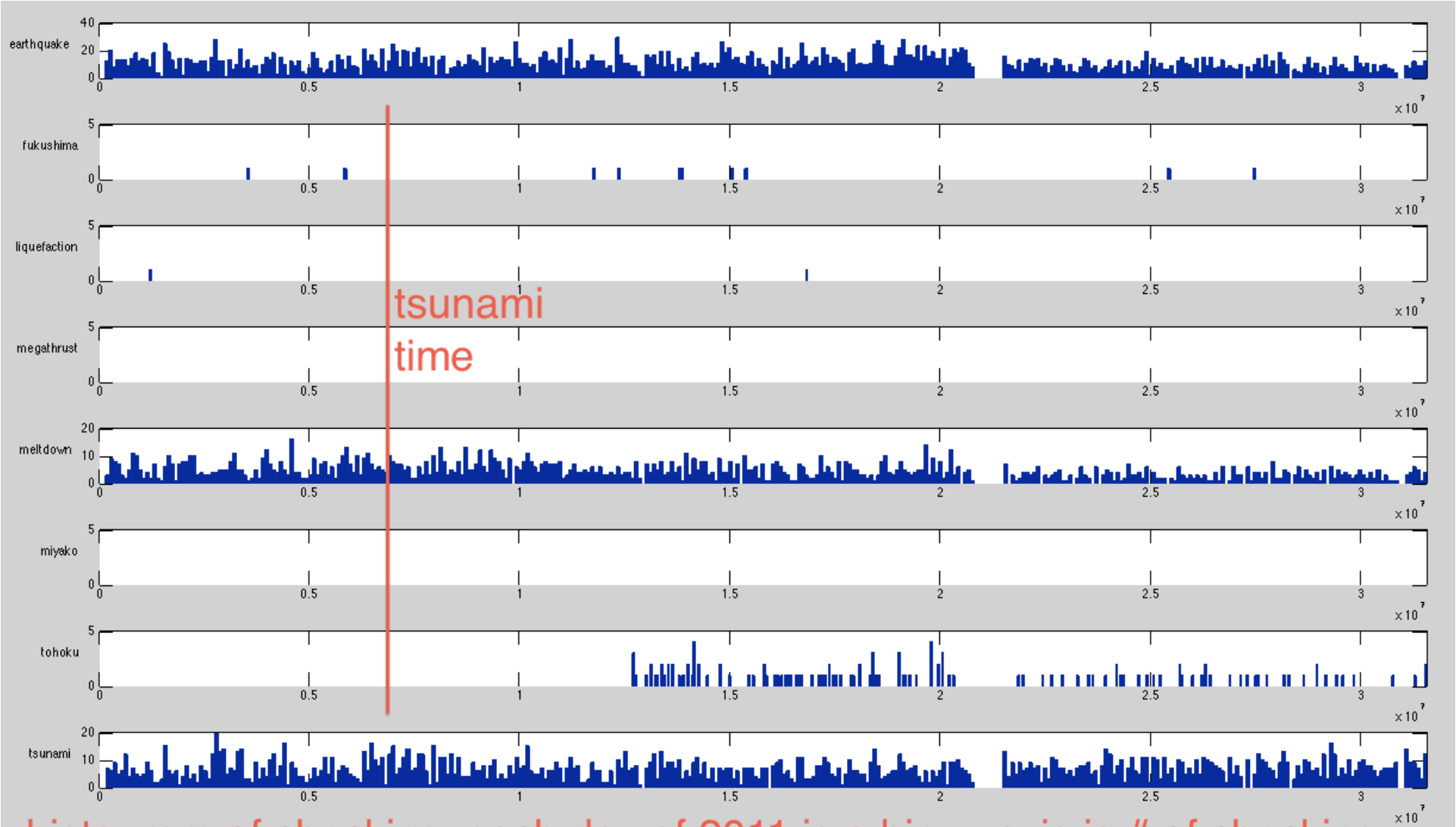
Wk4: 1D linear Frequency Project

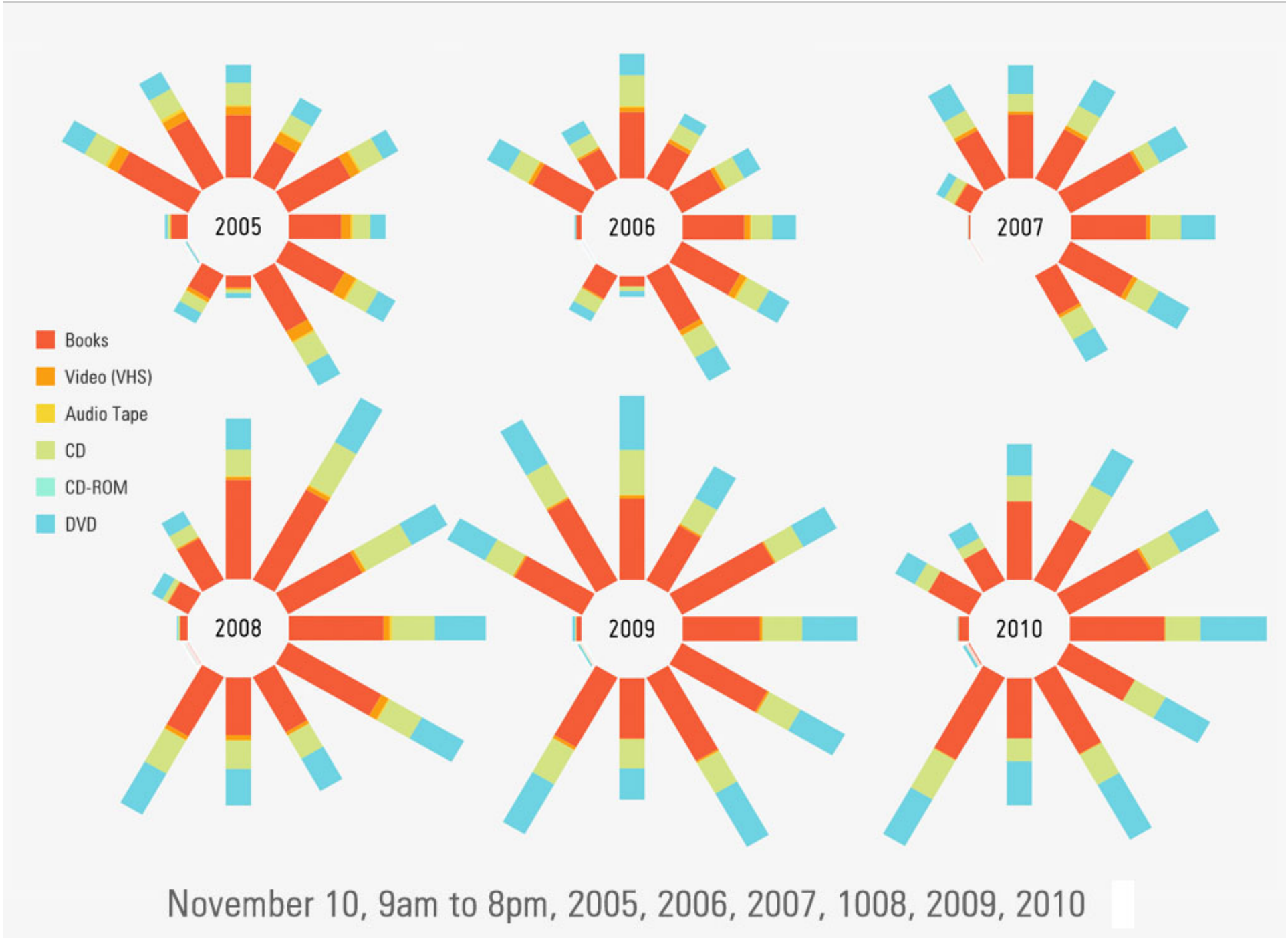
Explores how to map patterns and data relationships

title like '%drunk%' group by collcode;



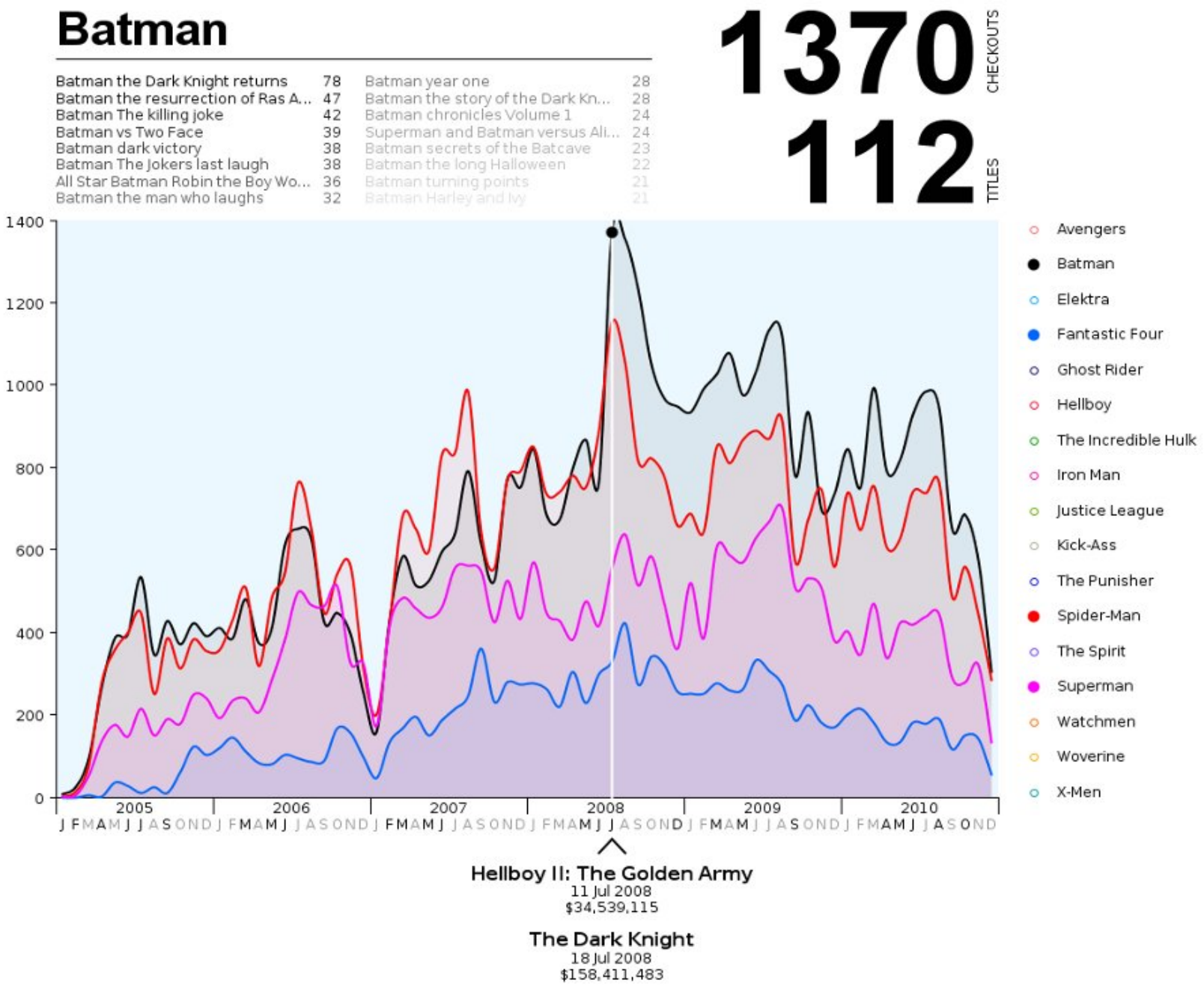
Karl Yerkes: DataMining Tsunami 2011

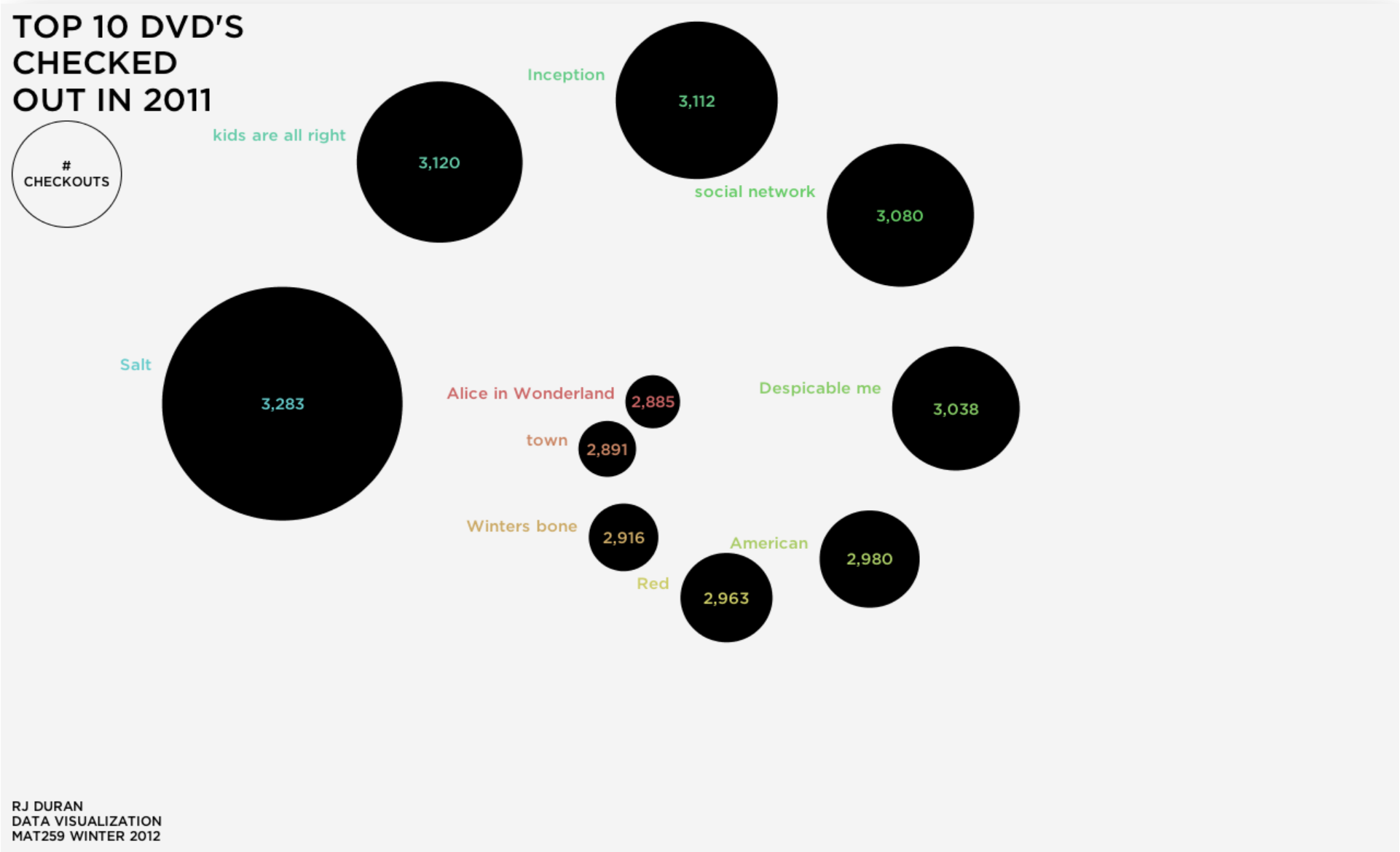


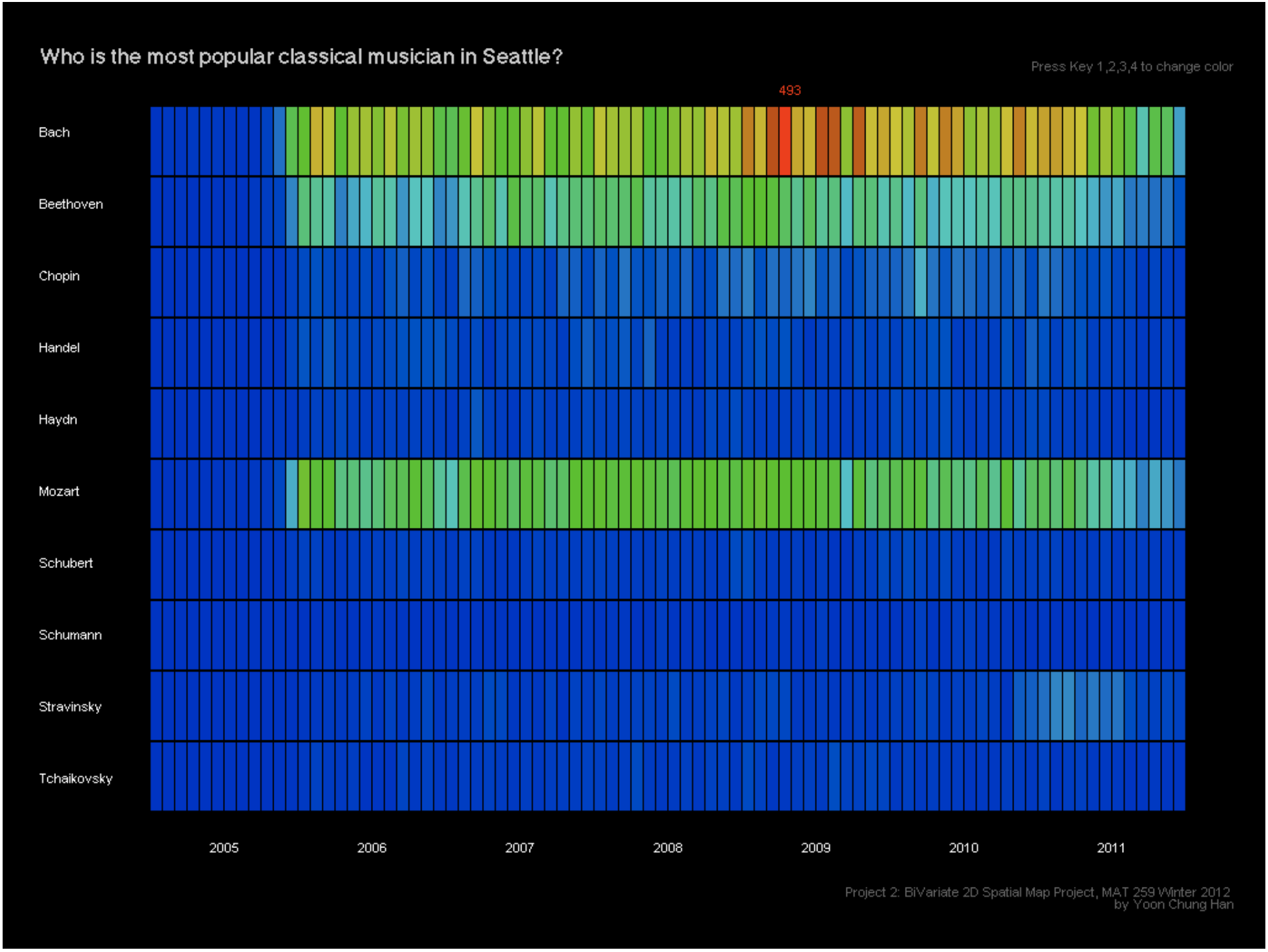

2005
2006
2007
2008
2009
2010
Books
Video (VHS)
Audio Tape
CD
CD-ROM
DVD
November 10, 9am to 8pm, 2005, 2006, 2007, 1008, 2009, 2010

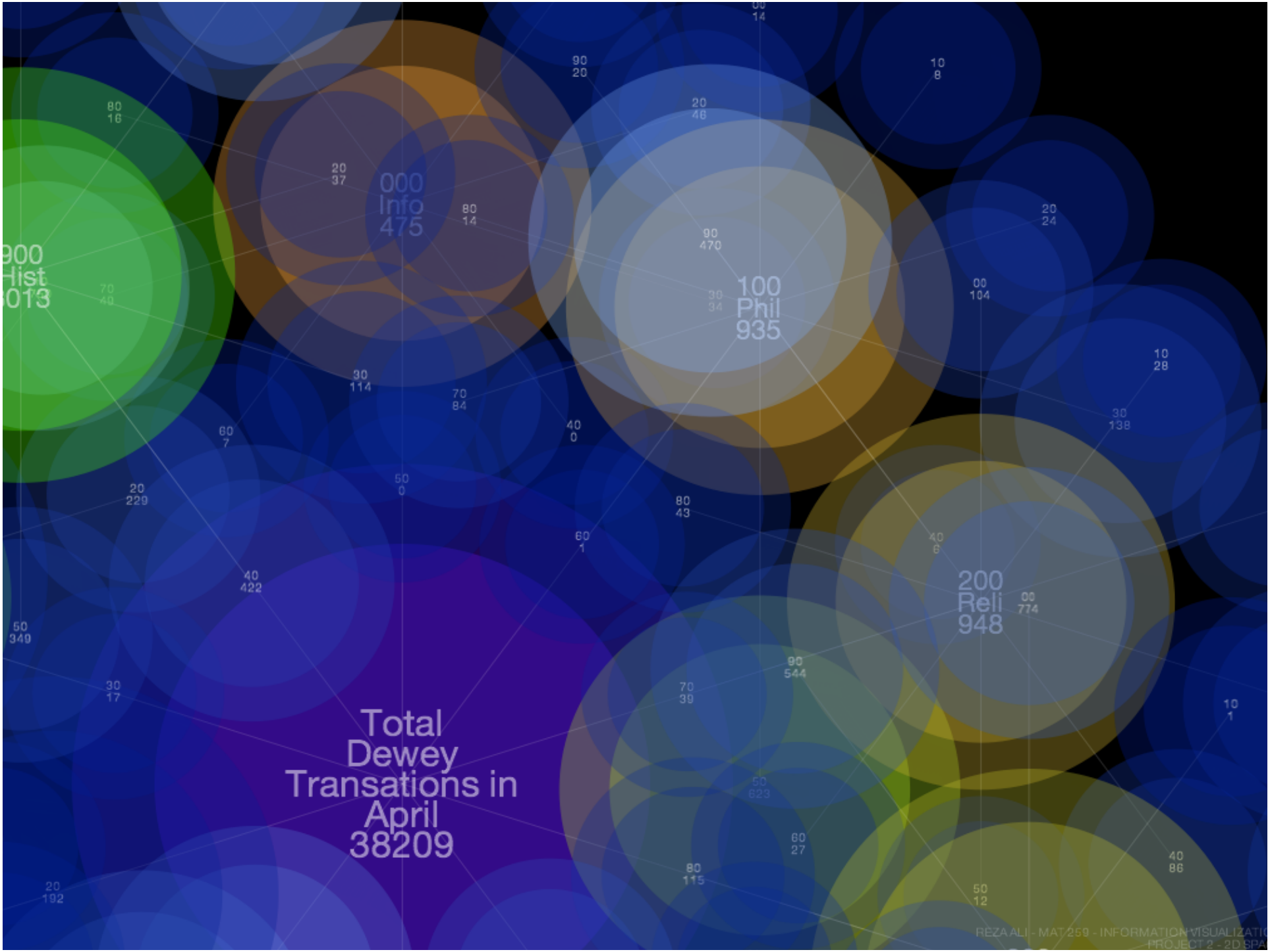
Wk5,6: Spatial Mapping

Explores how to map data in a 2D space where Horizontal, Vertical, and color or size relationship at specific locations can have meaning


Batman
Batman the Dark Knight returns 78
Batman the resurrection of Ras A... 47
Batman The killing joke 42
Batman vs Two Face 39
Batman dark victory 38
Batman The Jokers last laugh 38
All Star Batman Robin the Boy Wo... 36
Batman the man who laughs 32
Batman year one 28
Batman the story of the Dark Kn... 28
Batman chronicles Volume 1 24
Superman and Batman versus Ali... 24
Batman secrets of the Batcave 23
Batman the long Halloween 22
Batman turning points 21
Batman Harley and Ivy 21
1370 CHECKOUTS
112 TITLES
1400
1200
1000
800
600
400
200
0
2005
2006
2007
2008
2009
2010
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D
Avengers
Batman
Elektra
Fantastic Four
Ghost Rider
Hellboy
The Incredible Hulk
Iron Man
Justice League
Kick-Ass
The Punisher
Spider-Man
The Spirit
Superman
Watchmen
Woverine
X-Men
Hellboy II: The Golden Army
11 Jul 2008
$34,539,115
The Dark Knight
18 Jul 2008
$158,411,483


TOP 10 DVD'S CHECKED OUT IN 2011
CHECKOUTS
Inception
3,112
kids are all right
3,120
social network
3,080
Salt
3,283
Alice in Wonderland
2,885
town
2,891
Despicable me
3,038
Winters bone
2,916
American
2,980
Red
2,963
RJ DURAN
DATA VISUALIZATION
MAT259 WINTER 2012


Who is the most popular classical musician in Seattle?
Press Key 1,2,3,4 to change color
493
Bach
Beethoven
Chopin
Handel
Haydn
Mozart
Schubert
Schumann
Stravinsky
Tchaikovsky
2005
2006
2007
2008
2009
2010
2011
Project 2: BiVariate 2D Spatial Map Project, MAT 259 Winter 2012
by Yoon Chung Han

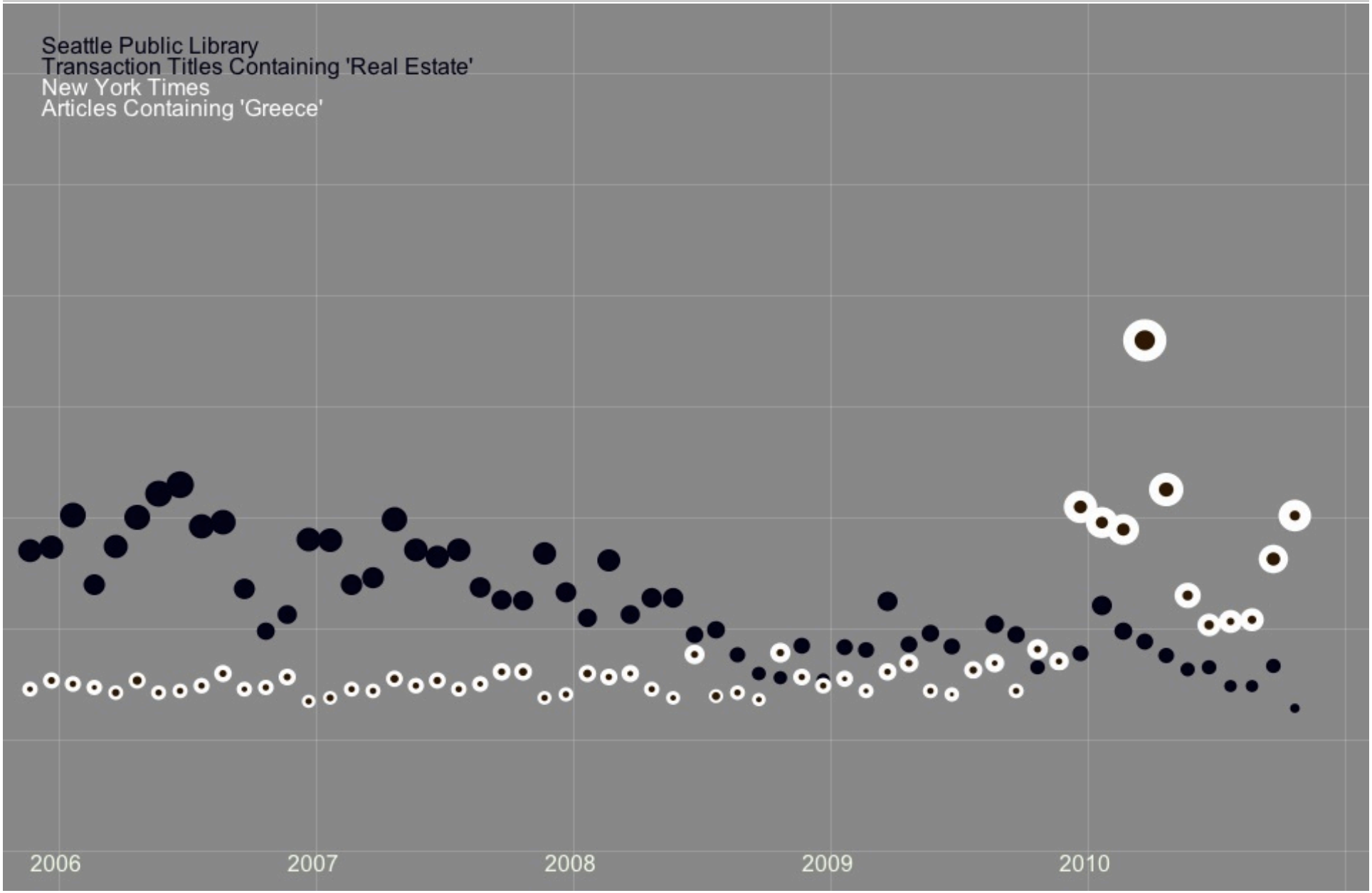

000
Info
475
100
Phil
935
200
Reli
948
900
Hist
3013
Total
Dewey
Transations in
April
38209
REZA ALI - MAT 259 - INFORMATION VISUALIZATIO
PROJECT 2 - 2D SPA

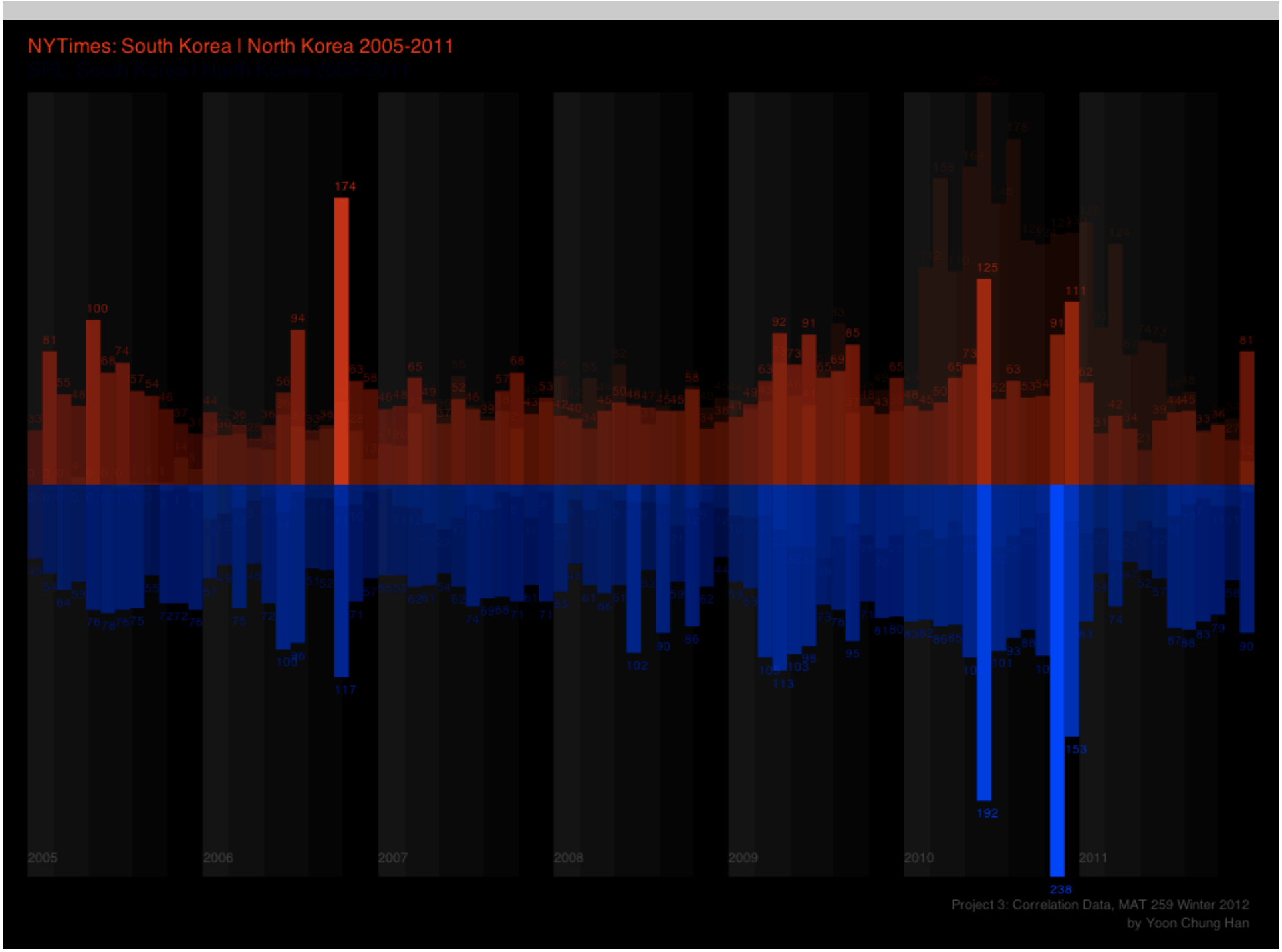
Wk7: Correlate Data with an External Source

Explores how to correlate the Seattle data with another

- NY Times
- Facebook
- Amazon
- Itunes
- Etc.

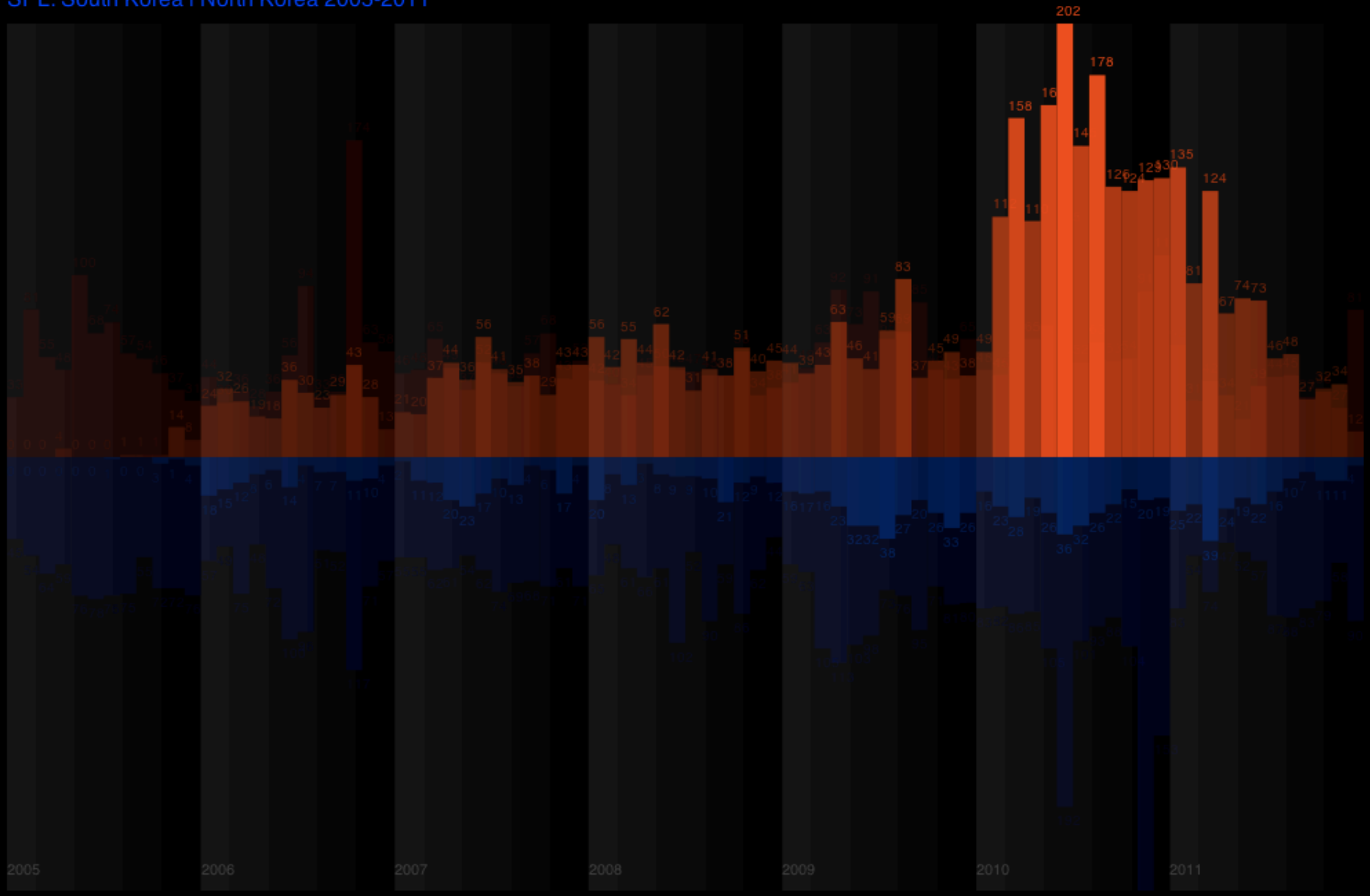
Also introduce interactivity


Seattle Public Library
Transaction Titles Containing 'Real Estate'
New York Times
Articles Containing 'Greece'
2006
2007
2008
2009
2010


NYTimes: South Korea | North Korea 2005-2011
SPL: South Korea | North Korea 2005-2011
2005
2006
2007
2008
2009
2010
2011
Project 3: Correlation Data, MAT 259 Winter 2012
by Yoon Chung Han

NYTimes: South Korea | North Korea 2005-2011

SPL: South Korea | North Korea 2005-2011



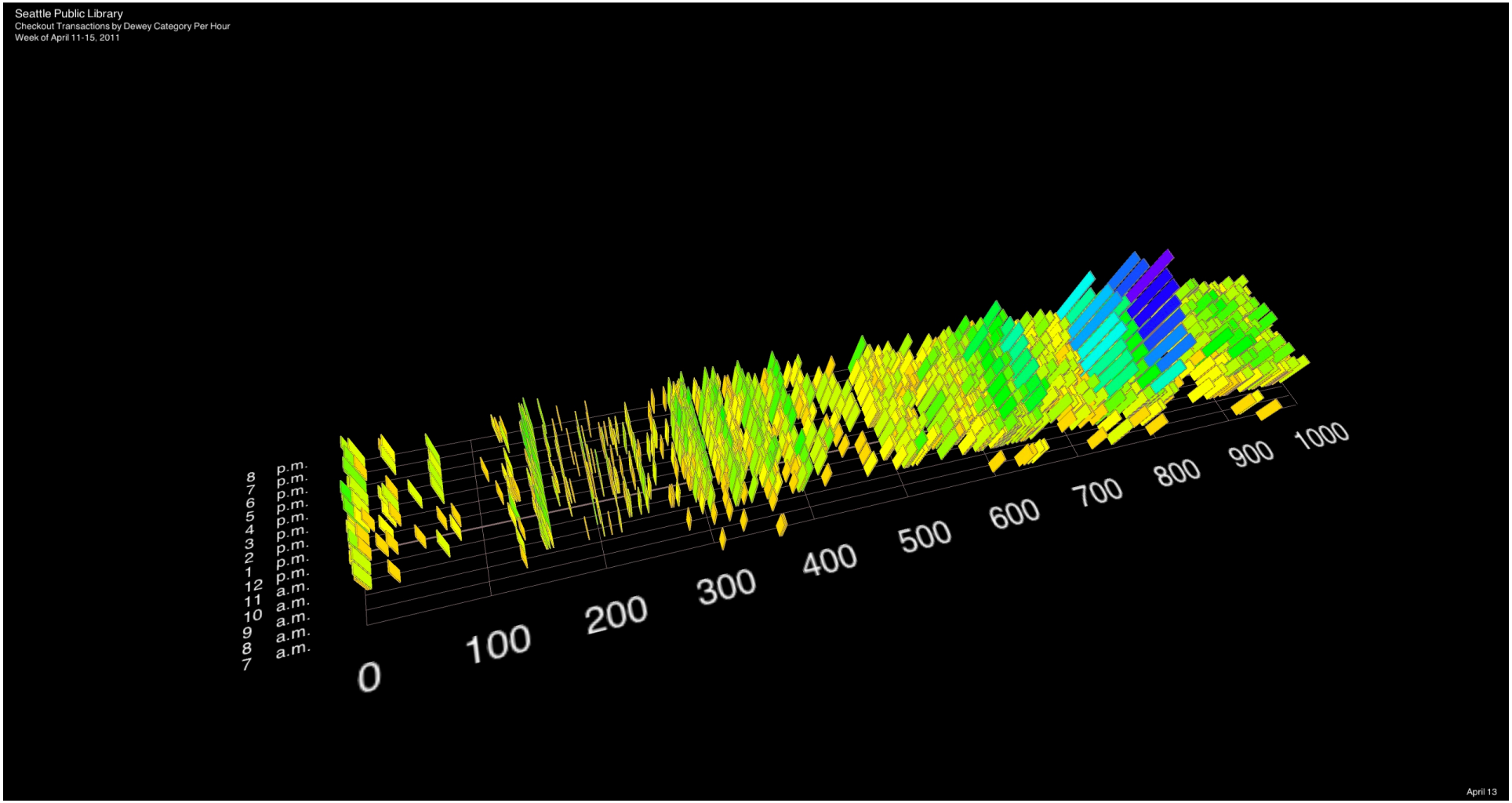
Project 3: Correlation Data, MAT 259 Winter 2012

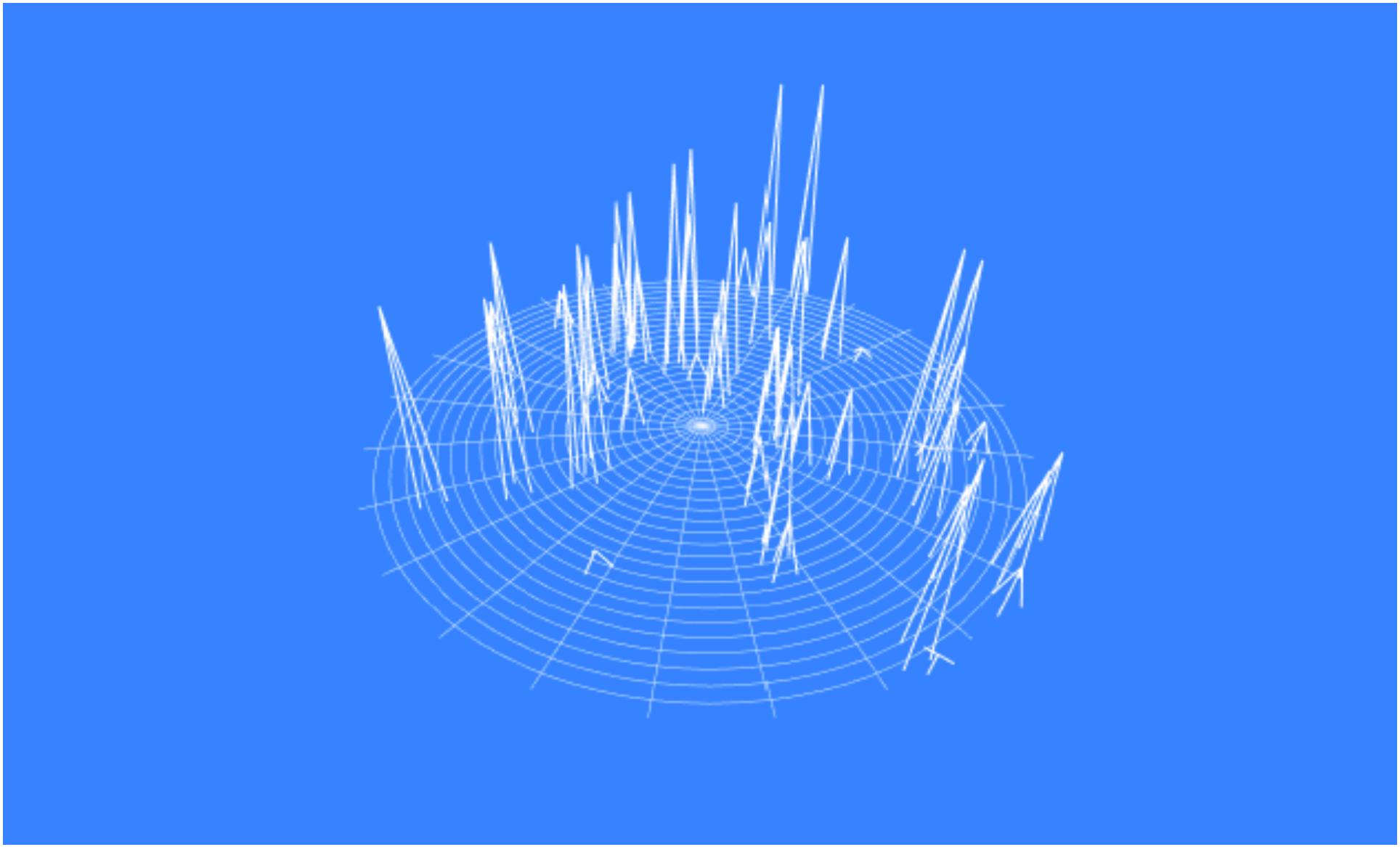
by Yoon Chung Han

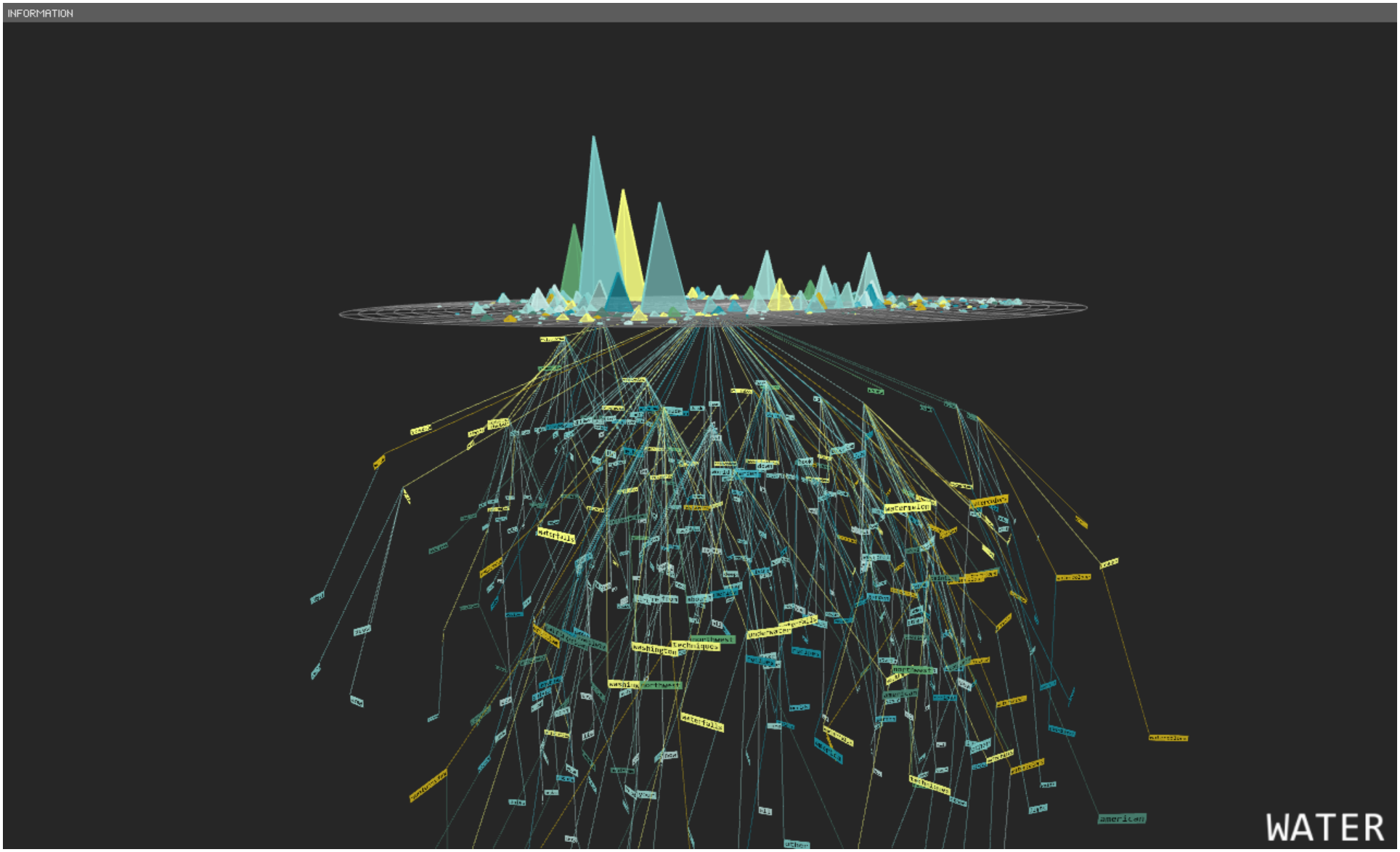
Wk8,9: 3D Interactive

Explores how to map data in

- 3D space
- Introduce interactivity
- And time


Seattle Public Library
Checkout Transactions by Dewey Category Per Hour
Week of April 11-15, 2011
8 p.m.
7 p.m.
6 p.m.
5 p.m.
4 p.m.
3 p.m.
2 p.m.
1 p.m.
12 a.m.
11 a.m.
10 a.m.
9 a.m.
8 a.m.
7 a.m.
0
100
200
300
400
500
600
700
800
900
1000
April 13




INFORMATION
WATER

Earth

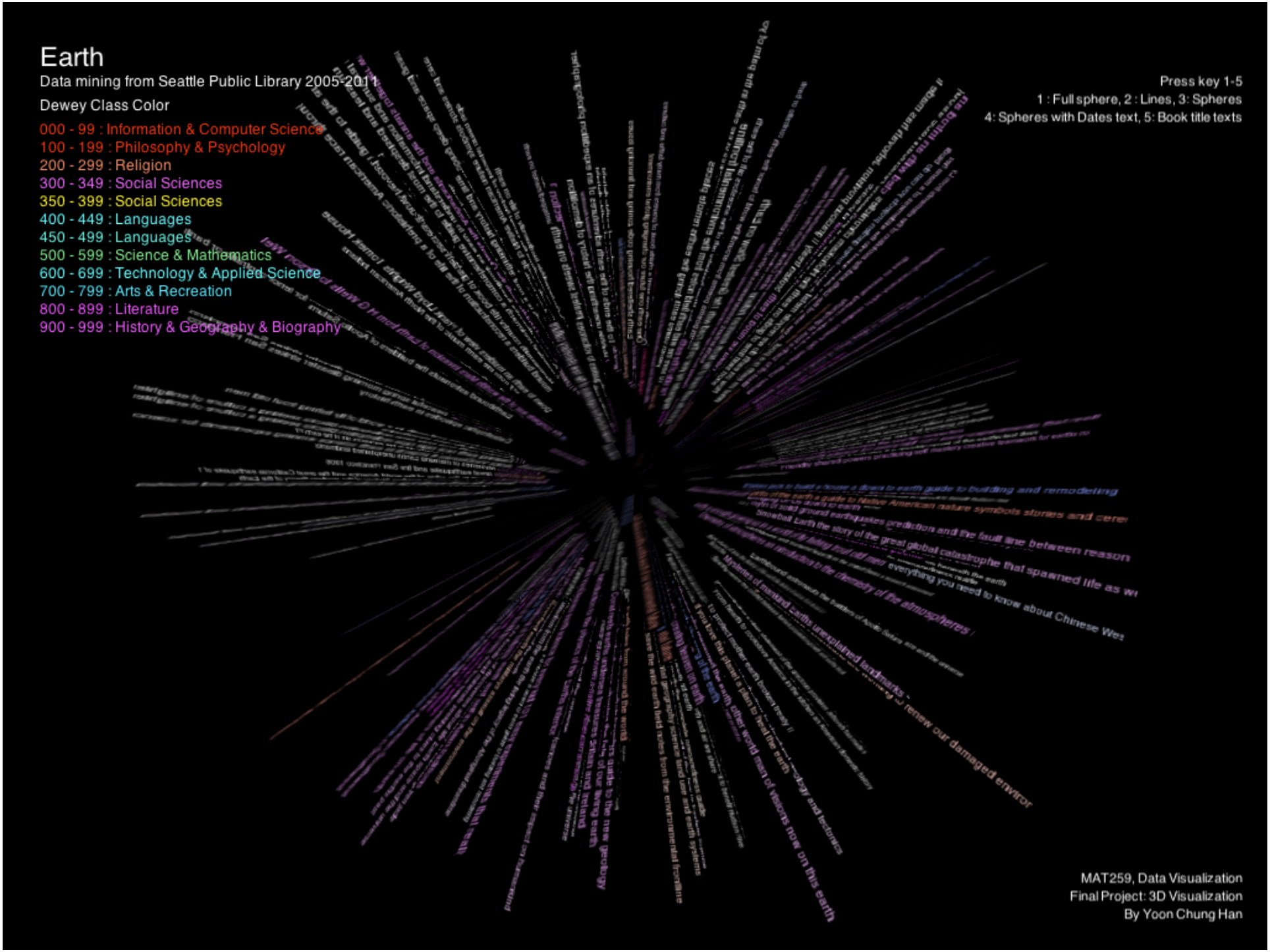
Data mining from Seattle Public Library 2005-2011

Dewey Class Color

000 - 99 : Information & Computer Science
100 - 199 : Philosophy & Psychology
200 - 299 : Religion
300 - 349 : Social Sciences
350 - 399 : Social Sciences
400 - 449 : Languages
450 - 499 : Languages
500 - 599 : Science & Mathematics
600 - 699 : Technology & Applied Science
700 - 799 : Arts & Recreation
800 - 899 : Literature
900 - 999 : History & Geography & Biography

Press key 1-5
1 : Full sphere, 2 : Lines, 3: Spheres
4: Spheres with Dates text, 5: Book title texts

MAT259, Data Visualization
Final Project: 3D Visualization
By Yoon Chung Han


Earth
Data mining from Seattle Public Library 2005-2011
Dewey Class Color
000 - 99 : Information & Computer Science
100 - 199 : Philosophy & Psychology
200 - 299 : Religion
300 - 349 : Social Sciences
350 - 399 : Social Sciences
400 - 449 : Languages
450 - 499 : Languages
500 - 599 : Science & Mathematics
600 - 699 : Technology & Applied Science
700 - 799 : Arts & Recreation
800 - 899 : Literature
900 - 999 : History & Geography & Biography
Press key 1-5
1 : Full sphere, 2 : Lines, 3: Spheres
4: Spheres with Dates text, 5: Book title texts
MAT259, Data Visualization
Final Project: 3D Visualization
By Yoon Chung Han

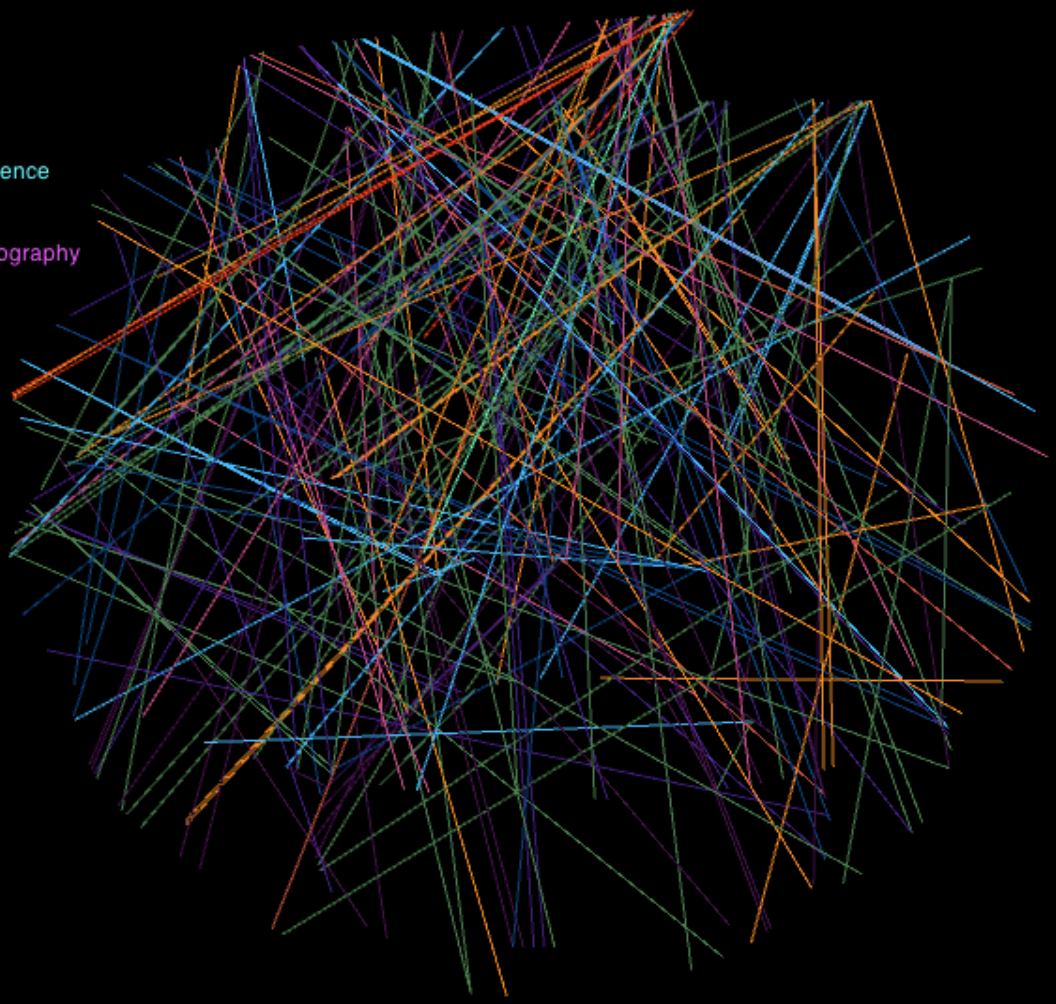
Earth

Data mining from Seattle Public Library 2005-2011

Dewey Class Color

000 - 99 : Information & Computer Science
100 - 199 : Philosophy & Psychology
200 - 299 : Religion
300 - 349 : Social Sciences
350 - 399 : Social Sciences
400 - 449 : Languages
450 - 499 : Languages
500 - 599 : Science & Mathematics
600 - 699 : Technology & Applied Science
700 - 799 : Arts & Recreation
800 - 899 : Literature
900 - 999 : History & Geography & Biography

Press key 1-5
1 : Full sphere, 2 : Lines, 3: Spheres
4: Spheres with Dates text, 5: Book title texts



MAT259, Data Visualization
Final Project: 3D Visualization
By Yoon Chung Han

Project Development Process

- **Data search:** Knowledge discovery
- **Data analysis:** What patterns emerge
- **Data formulation through algorithmic processing**
- **Translation into visualization**
- **Iterative process** < > feedback until results are as needed
- **Publication**

Breadth of Visualization: From Pragmatic to Poetic

- Vis. requires dual approaches:
  - Computation is utilitarian and visualization requires aesthetic skillsets

- Experience in both give best results
- Both require extensive practice

What Do We Represent? (AJ Ayer's Propositions)

- **Synthetic:** We synthesize some knowledge about the world
- **Subjective:** We make a statement that is evaluative, and expressive of a feeling (but not about the world)
- **Syntactic:** We test the communication process, re-affirm that communication is proceeding as planned. We explore the syntax of the language

Software

- MySQL (the database)
- MySQL Workbench (access to data)
- Processing (Java-based scripting language)

Some References

- Atlas of Science, Katy Borner
- Graphics of Large Datasets, Unwin, Theus, Hofmann
- Visualizing Data, Ben Fry
- Robert Kosara, social analysis
- Visual Display of Quantitative Information, Edward Tufte (historical representation)


VisWeek Art Show
WATER
RECYCLE
intel