Problem-Driven Design Studies --Money Donation to Public School

By Huaying Tian & Arthur Sun

OUTLINE

What are we going to do?

1.Analyze data from a US based non-profit organization website that allows individuals to donate money directly to public school ;

2.Get the dataset and take a 9000-row table subset of the original dataset for our analysis purposes;

3. Create an informative analysis on the basis of the data attributes;

4.Visualize the data in an efficient and expressive way.

A large two-dimension data table → astract key data types → reduce complexity into simplicity by counting data types → human-visualized graphs

What are we going to do with the data?

Two-dimension table → visualize → Graph,i.e line chart,bar chart,pie chart……

Understand the data better

Purpose: Analyze the data better and give appropriate suggestions on public donation

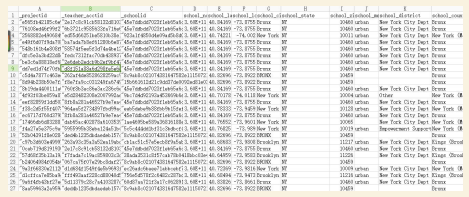
1.What is the trend of number of donations in recent years?Do we need more donations or the status quo is just what we want?

2.Which state shall we pay more attention to?

Why do we need visualization?

Through a problem-driven process, these specialized datasets are often an interesting mix of complex combinations of and special cases of the basic data types,and they also are a mix of original and derived data.

Whitout vis, we may see a table like that:



What a great mass!

But by using visualization,People can have a **clear overview** at first with low-latency page loading of data,and then zoom and filter to check the **details** they demand

What data do we abstract?

Data types:

1. school_state: NY,NC or……
2. resource_type: books ,technologies,or……
3. poverty_level: Highest poverty ,low poverty or……
4. date_posted: day,month,year
5. total_donations: how much donations they've received
6. funding_status: completed or expired
7. grade_level 9-12,5-8 or ……

How to visualize the data?

Data Visualization:

school state → drop-down menu

date_posted range → line chart

resource types → horizontal bar chart

donation counted by poverty level

donation counted by grade

funding status → pie chart

you can choose any range you like to see the attributes you're intersted in

The components of our analysis and their function

1.D3.js: A javascript based visualization engine which will render interactive charts and graphs based on the data.

2.Node JS: Our powerful server which serves data to the visualization engine and also hosts the webpages and javascript libraries.

3.Mongo DB: The resident No-SQL database which will serve as a fantastic data repository for our project.

Thanks

Students Migration

Elementary and Secondary Schools in São Paulo/Brazil

Carolina Roman Amigo & Wenqiang (Dylan) Dong

CPSC 547 – Information Visualization

October 2015

About the Data

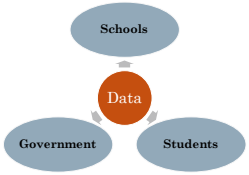
- Educational Census (public available, per year)
  - School code
  - School name
  - School type (private/public)
  - School location (Latitude, Longitude, Postal Code, City, District)
  - Census Year
  - Student Code
  - Student Grade
- Data size (per census year, we need at least two)
  - 7,789,831 Students
  - 20,029 Schools
  - ~ 650 MB

Challenge

Context

- In Brazil, elementary and secondary public education generally has poor quality.
- Every parent that can afford a private school does it, thus we have a huge number of private schools competing for students.
- They run like businesses, so understanding their market share is relevant for them.
- There is an standardized test for being accepted at the best universities, and some private schools specialize in training students for that; so when getting to high school some students opt for migrating to this kind of schools.

Stakeholders



T1 - Tasks for Schools

- Help schools identify migration pattern of students.
  - Are they losing more students than gaining?
  - To which schools are they going?
  - Is there any particular grade in which migration is more intense?
  - How their students migration compares to the other schools?

What?
Dataset Types
Networks

T2 - Tasks for Government

- Are there any areas of the state receiving more students than others?
- Are students migrating from public to private schools?

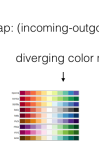
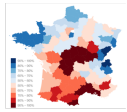
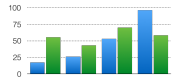
What?
Dataset Types
Geometry (Spatial)
Networks

T1 - Help schools identify migration pattern of students

- map: school types
  - categorical color map
- # of incoming & outgoing students
  - line thickness
- interaction: select
- facet: juxtapose


T2 - Which areas of the state are receiving more students?

- map: (incoming-outgoing)%
  - diverging color map
- interaction: select
- facet: juxtapose

Thank you!

Carolina Roman Amigo
carolamigo@gmail.com

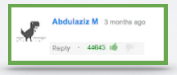
Wenqiang (Dylan) Dong
wdong@cs.ubc.ca

VISUALIZATION OF YOUTUBE COMMENTS

- Doesn't support easy finding of entertaining comments.
- Emotionally draining arguments and trolls.

- Doesn't support easy finding of entertaining comments.
  Task 1: Explore for entertaining comments.
- Emotionally draining arguments and trolls.
  Task 2: Identify arguments.
  Task 3: Identify trolls.

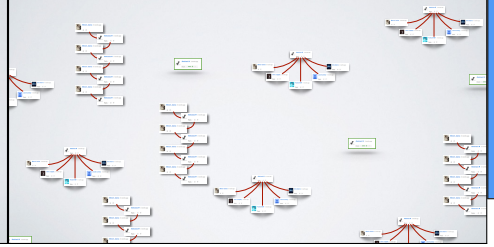
- Entertaining comments = highly liked comments (generally)



- Arguments = Long back-and-forth between two users with little or no likes

- Trolls = A single user (with little or no likes) being bombarded by multiple users

Idea: A Bird's Eye View of the Youtube Comment Section



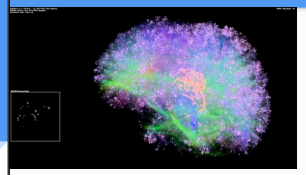
Neuron electrophysiology data visualization (Neuroelectro)

Presented by: Dmitry, Emily and Mike

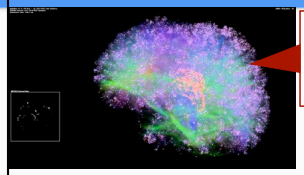
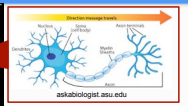
Introduction

Dmitri

How does your brain work? It's complicated



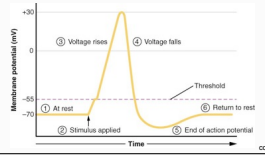
How does your brain work? It's complicated

What is our data?

Electrophysiology is the study of the electrical properties of biological cells and tissues. In neuroscience, it includes measurements of the electrical activity of neurons, and particularly action potential activity.

-wikipedia



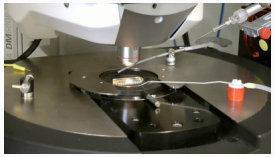
What is our data?

How many neuron types are there? The debate has been ongoing for decades. We use enhanced NeuroLex.org definitions (~100 Neuron types)

http://www.anatomyzone.com

What is our data?

Experimental metadata - solutions used, temperature, electrode types, animal species, strain and age, etc.



www.leica-microsystems.com

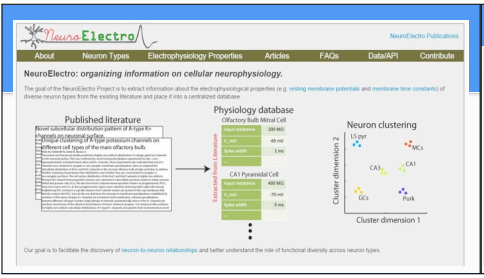
What is our data?

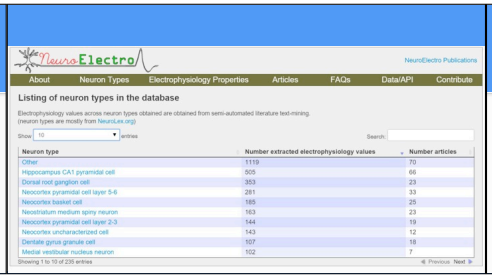
To summarize we have (per article):
1) Electrophysiology properties
2) Neuron types
3) Experimental metadata

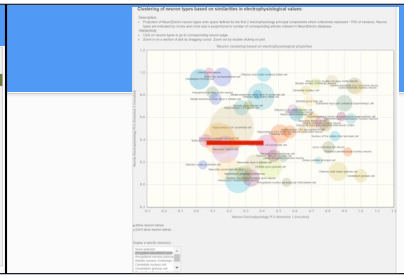
We extract all of the above from published articles through text-mining and curation.

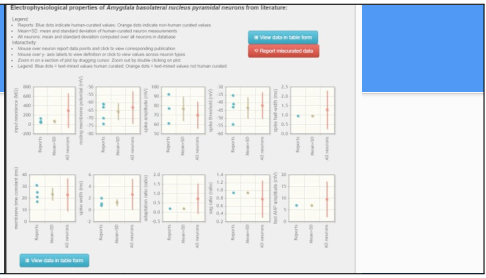
Current State

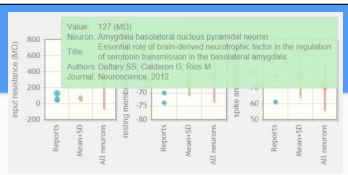
Mike

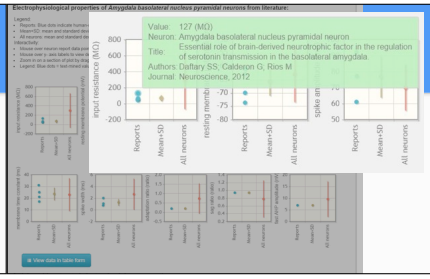


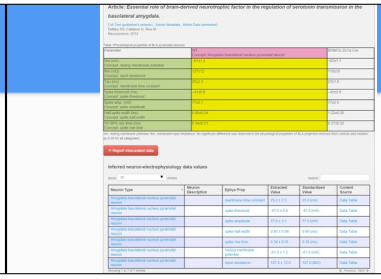


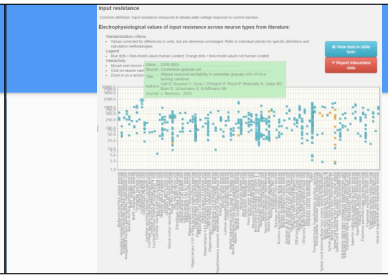


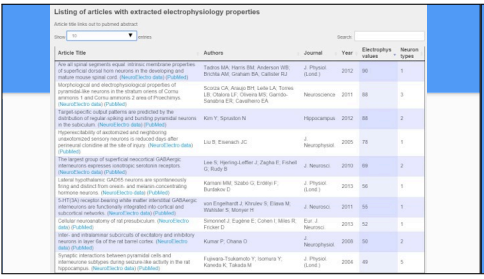












Problem Characterization

Emily

Problem Characterization

- We met with our stakeholder to ascertain high-level questions:
  - What do cells in different parts of the brain do?
  - How do experimental conditions affect electrophysiological measurements?
  - etc.
- We refined these into a few abstract tasks...

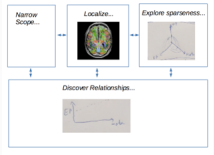
Task Analysis

- *Discover relationships*
  - Neuron types (categorical)
  - Electrophysiological properties (quantitative)
  - Experimental conditions (quantitative and categorical)
- *Narrow scope of analysis*
  - Select experimental conditions and ephys properties to include
  - Filter by neuron type, ephys property, and experimental conditions

Task Analysis

- *Explore* sparseness of data
  - How many data points for each neuron type? property? experimental condition?
- *Localize* neuron types and ephys properties in the brain
- *Lookup* details for individual data points

Tentative Solution



How much time?

Henry

Data Description

- Activity log of my commitments (e.g. CPSC547)
- Only tracks about 40 hours every week
- Categorized by name and project

Task

- Insert text here

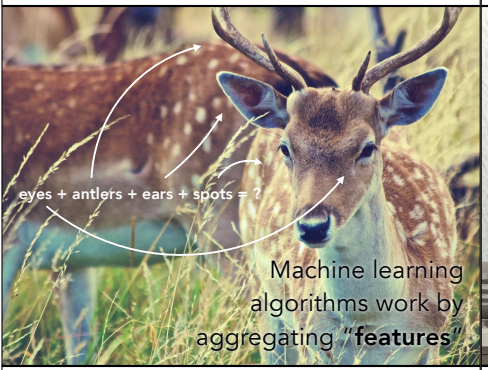
Existing Visualizations

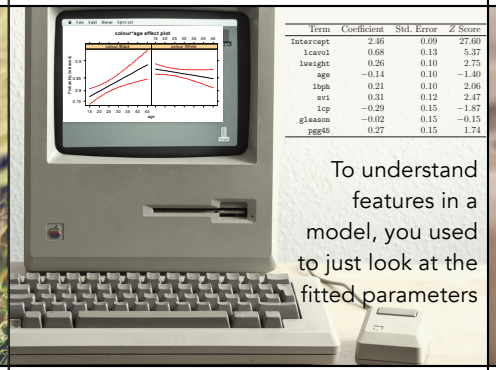
- Calendar
- Pie chart
- Bar chart
- …

Partner welcome
(no pressure)

VISUALISING FEATURE LEARNING

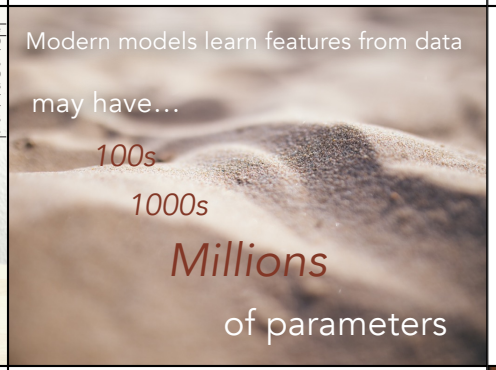
Jason Hartford





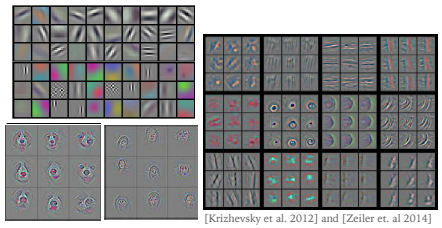
Term	Coefficient	Std. Error	Z Score
Intercept	2.46	0.09	27.60
lcavol	0.68	0.13	5.37
lweight	0.26	0.10	2.75
age	−0.14	0.10	−1.40
lbph	0.21	0.10	2.06
svi	0.31	0.12	2.47
lcp	−0.29	0.15	−1.87
gleason	−0.02	0.15	−0.15
pgg45	0.27	0.15	1.74

To understand features in a model, you used to just look at the fitted parameters



This makes it very difficult to understand what's going on in your model!

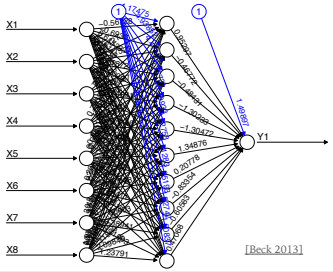
In vision you can plot parameters directly.



[Krizhevsky et al. 2012] and [Zeiler et. al 2014]

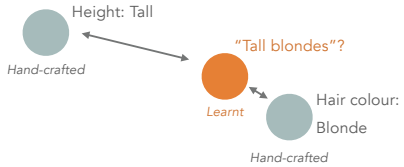
… but this only works because of the visual structure of their models.

Without visual structure, you get plots like this…



[Beck 2013]

Feature discovery "*by analogy*"



Idea: derive ***distance*** between the output of learnt features and hand-crafted features.

- I have data.

- I have data.
- I have hand-crafted features.

- I have data.
- I have hand-crafted features.
- I have a model.

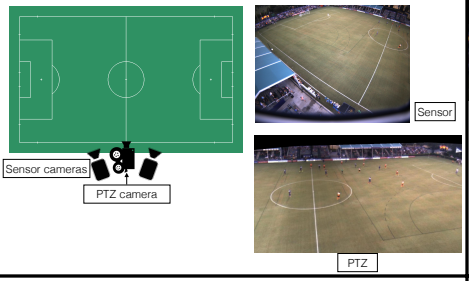
- I have data.
- I have hand-crafted features.
- I have a model.
- I'm not sure what it's learning…

CameramanVis

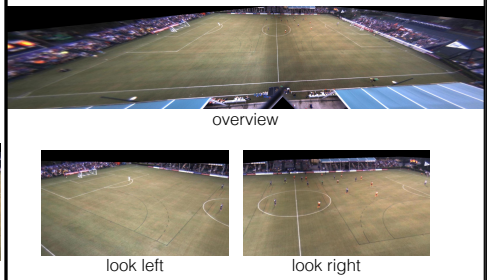
-Where camera should look?

Jianhui (Jimmy) Chen
CPSC 547 Pitch

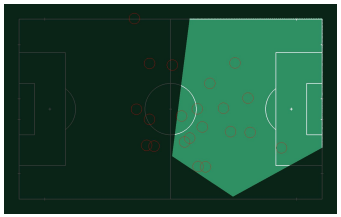
Cameras for soccer games



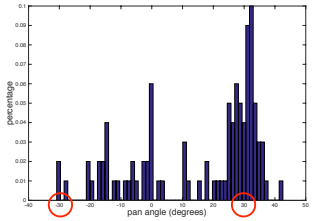
Left or right?



Goal:analyze cameraman's view



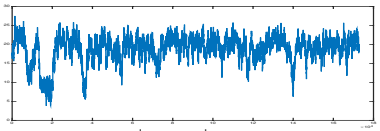
Not always the same angle



-30º, 30º are almost opposite.

Data

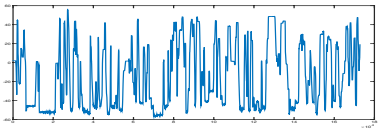
48-min (172, 800 frames) soccer game
Player positions on the playing ground
Camera angles of a PTZ camera
Both of them are noisy



player numbers

Tasks

- Oveview: camera angle, player position in playing ground
- Query1 : given angle get player position distribution
- Query2 : given player position get angle distribution
- Outlier detection: cameraman look at un-normal angle



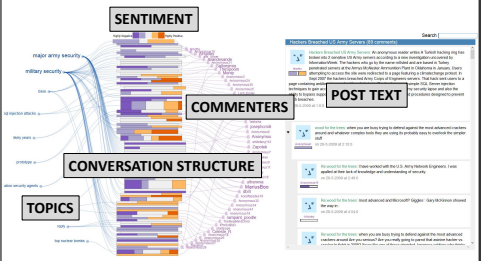
camera pan angle

Linking Sentences in Online Conversations

Jordon Johnson
M.Sc. Student
UBC Computer Science
jordon@cs.ubc.ca

Works and plays well with others

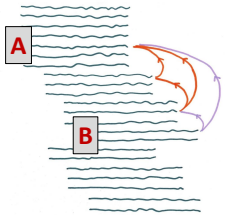
ConVis @ UBC



http://www.cs.ubc.ca/cs-research/lci/research-groups/natural-language-processing/image/convis/3.jpg

Visualizing Links Between Sentences

"Sentence **B** refers to sentence **A**"



GOAL: **visualize/edit links**

Link Attributes:
- Linked sentences
- Agreement value
- Sentiment value

Links may form chains that give additional insight into conversation structure

Data and Current Tasks

- Online conversations, annotated by humans to establish a "gold standard"
  - Not that good…
  - Use vis to **make corrections** and **improve the gold standard**
- (In progress) annotations by candidate linking algorithms
  - Use vis to **evaluate the output**

Future Tasks for NLP Researchers

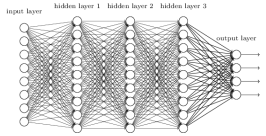
- Study link chains to associate link patterns with conversation types
  - Agreements
  - Debates
  - Off-topic/flamewars
  - Can we profile **trolling**?

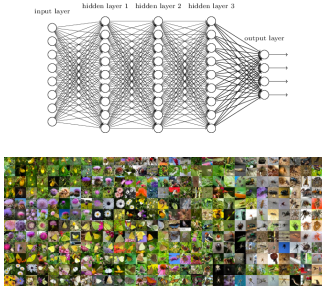
https://eclosure.com.au/wp-content/uploads/2015/04/Internet-Troll.jpg

Thank you!

vidviz: scaling up video annotation

Julieta Martinez





\+ “Plumeria Frangipani”

The Internet

The Internet

The Internet

The Internet

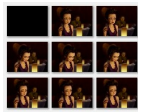
The Internet

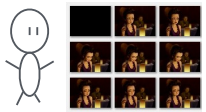
The Internet

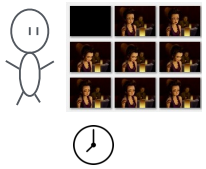
The Internet

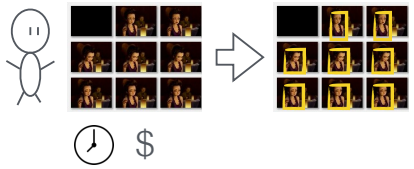
The Internet

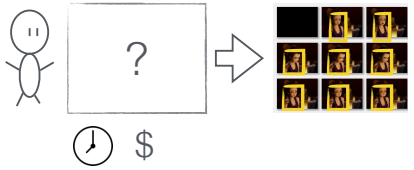
“the dark matter of the Internet” — Fei-Fei Li

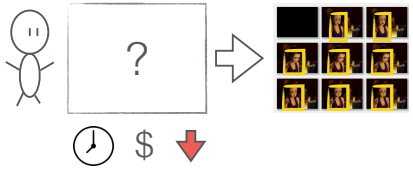


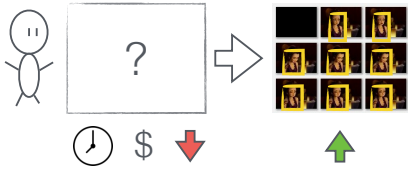












Interactive graphs using R Shiny

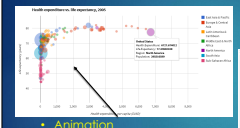
And maybe d3 if I have time.

Ken Mansfield, CPSC 547, Oct 22, 2015

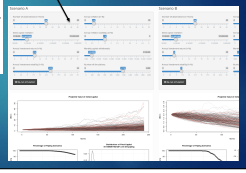
- Goal: Create different types of visualizations for exploration of data.
- Fully Interactive – Linked highlighting, multi-facetted, animations!
- Learn what works, what doesn't.
- Dataset: Gapminder. Wealth of information to explore, fairly easy to use.

Example: Gapminder world

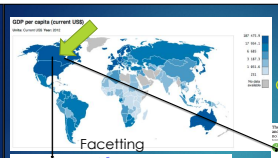
R Shiny Examples



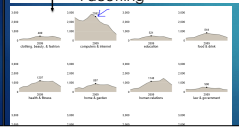
Interactivity



- Animation
- Colours
- Different types of Encoding



Facetting



Expand on Gapminder

Choropleth

Linked highlighting

InfoVis Project Pitch

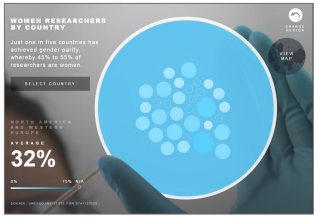
Kimberly Dextras-Romagnino

Why?

Present information to **the masses** in a visually appealing way to both **inform** and **entertain**

Storytelling: The Next Step for Visualization - Robert Kosara

- Presentation and Communication **vs** Exploration and Analysis
- Data Visualization: Medium for Storytelling

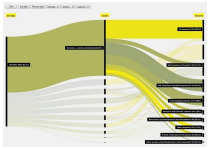


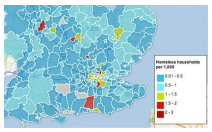
What?

Homelessness in Canada

- Things to compare
  - Areas
  - Sheltered vs Unsheltered
  - Age and Gender Distributions
  - Reasons for being on the streets

Homeless Vis Examples

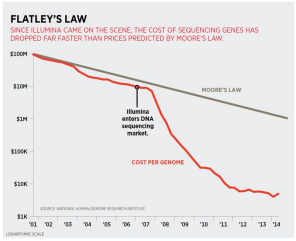




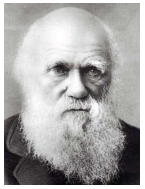
Any other ideas?

How Does Your Body Manage Its Army?

By: Louie Dinh



Anything that can be done via sequencing will be done via sequencing

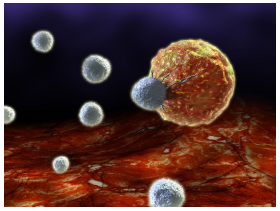


Darwin until now: understanding of biology has been very qualitative

So….

Let's Quantify! (With Sequencing)

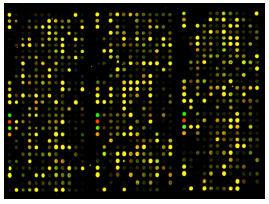
Interesting problem: How does your body manage its army (the immune system)



Which squadron (T-cells, B-Cells, Neutrophils, etc…) does it deploy and how does it coordinate this attack against invaders?

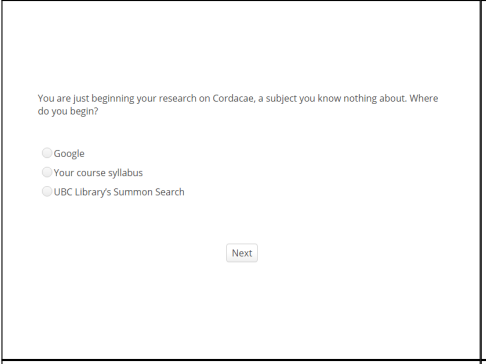
When your body starts a civil war (autoimmune disease), how does that look compared to a normal person?

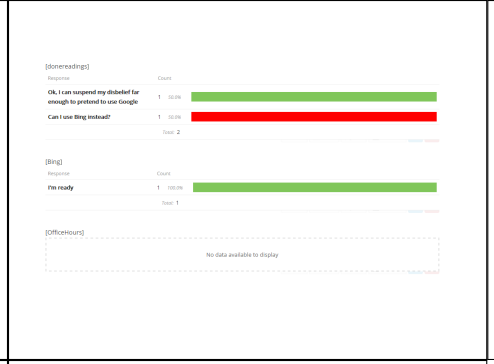
Sequencing data on blood cells is allowing us to visualize the your personal army

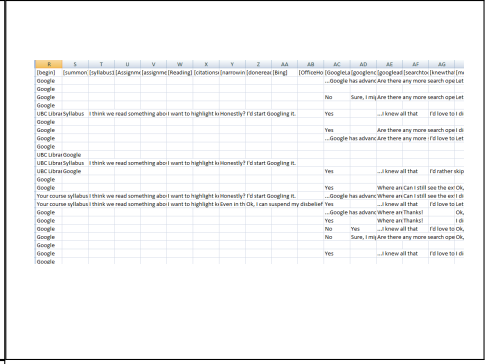


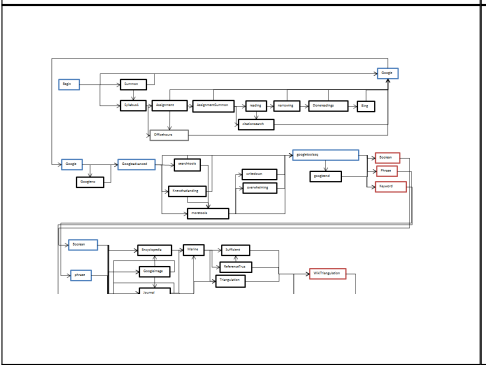
Very high dimensional. Highly correlated. How do we understand it?

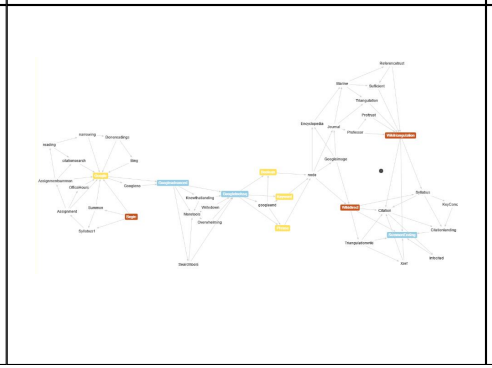
If you "know" biology and find this problem interesting, come talk to me!

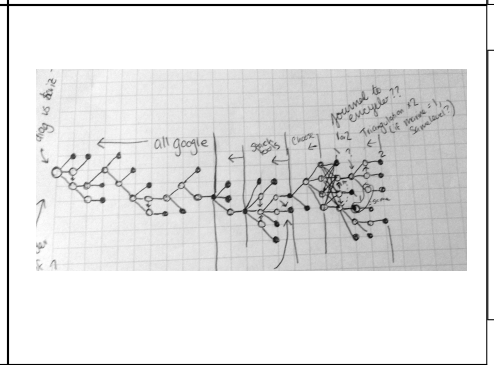




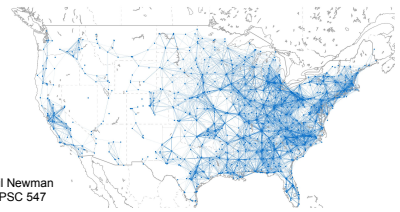








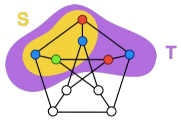
Spectrum Auction -> SAT Problems



Neil Newman
CPSC 547

Containment Caching

- Reuse solutions to already solved problems
- Larger problems are most useful



Visualizing the Cache

- Several million cache entries - need to summarize!
- How much of the multi-dimensional solution space does the cache "cover"?
- How are the stations distributed across cache entries?
- What is the key usage distribution for a given auction trace? Is this similar between traces?
- How "distant" are individual cache entries from each other?
- Potential complication: Data is not public (I've signed an NDA)

Search Trends Visualization

Rex Chang

Search Keywords

- Multiple Related Keywords
- Example
  - HTML5
    - → jQuery
      - → HTML5 jQuery
    - → HTML5 Canvas
    - → SVG
      - → HTML5 SVG
    - → HTML5 SVG Canvas
    - → HTML5 jQuery Canvas

What & Why

- Data: Search Engine Statistics
  - Google AdWords, Trends, Suggests
- Task: Given certain keywords, find related keywords that are:
  - Being searched more: Higher search volume
  - Getting more searched: Trending
- Rationale
  - Search Engine Optimization(SEO)
  - Search Keyword Efficiency

Example

- HTML5: 368000
  - → jQuery: 823000
    - → HTML5 jQuery: 720
  - → HTML5 Canvas: 22200
  - → SVG: 11000
    - → HTML5 SVG: 2400
  - → HTML5 SVG Canvas: 140
  - → HTML5 jQuery Canvas: 90

Example

	(null)	HTML5	HTML5 canvas
(null)		368000	22200
jQuery	823000	720	90
SVG	110000	2400	140

Visualizing a SAP Network

SAP®

SAP

Loblaws Indigo Books · Gifts · Kids Kellogg's

THE HOME DEPOT adidas Bell

547 Information Visualization
Pitches - Thu. Oct 22, 2015

UBC

Yaashaar Hadadian
Fast Track Ph.D. Student in CS Dept
cs.ubc.ca/~hadadian

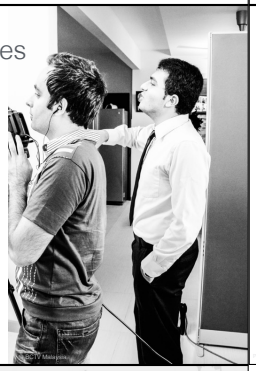
Motivation

Used to be a TV presenter
Had many talk shows
Wrote/Directed ~200 episodes

$o, Viewers' ratings mattered ($$)

- Keeping up w/ trends
- Touching popular genres
- & 100s of more analyses

Movie Databases



Our heaven!

A university to learn:
- what work(ed)/(s)
- what didn't/doesn't

Guess what?

I'm planning to visualize the heaven!

Data set

From movielens.org
Ratings until Aug 2015 !

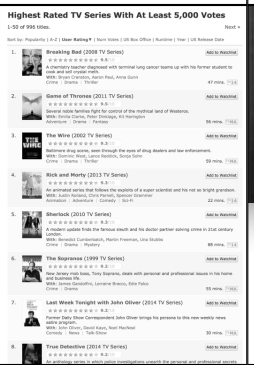
2 versions are available:

Light
- 700 Users
- 9K Movies
- 100K Ratings

Full
- 230K Users
- 30K Movies
- 21M Ratings

Data set

Features
+ Titles
+ Genre
+ Rating
+ Tags (Themes)
- IMDB ID!
  + IMDB Rating
  + Metascore
  + Year
  + Director
  + Writer
  + Awards
  ++ Many more…



Challenges

The DS needs some work
- Variable Recoding
- Data Cleaning
- Remote Data Fetching

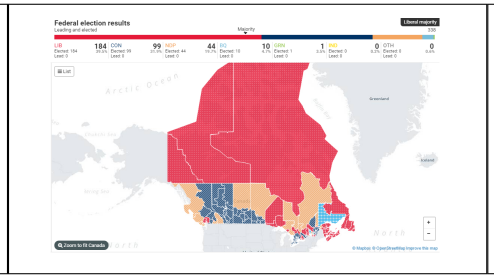
It's a programming project
- HTML/CSS/JavaScript
- JQuery/D3.js
- Java/JSP/JSF

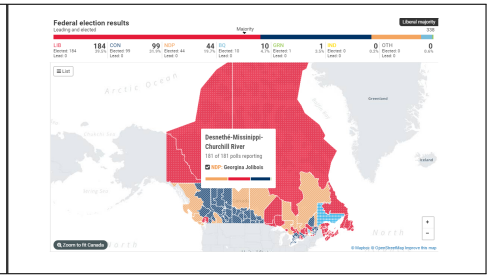
I'm looking for a partner,
Wanna join me?

Visualizing Uncertainty in Incomplete Election Data

CPSC 547 Project Pitch

Yasha Pushak

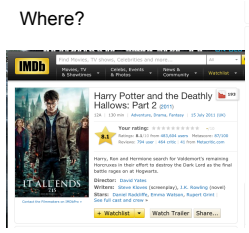




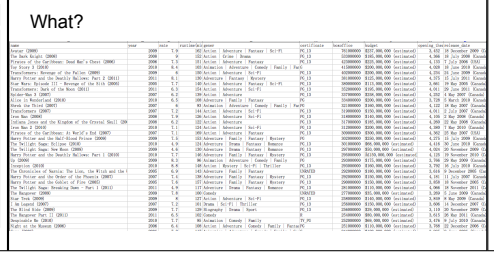
Visualizing Movie Data

Yujie Yang, Ye Chen

Where?



What?

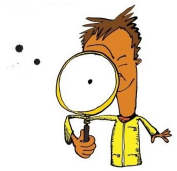


What?

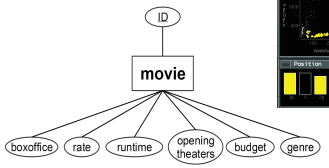
ID	Name	Boxoffice (USD)	Rate	Runtime (mins)	Opening Theaters	Budget	Release Date	Genre
47	Up	293000000	8.3	96	3766	175000000	May 2009	Animation | Adventure | Comedy | Family

Why?

Explore the distribution of data.



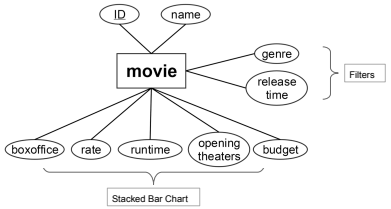
How?



Why?

Help people to choose a movie.

How?



How?

Thanks

First Thought on

Exploratory Scholar Data

Zipeng Liu
Oct 22 2015

Dataset

- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms…

ArnetMiner Scholar Data Challenge. https://aminer.org/big-scholar-challenge/

Dataset

- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms…

Networks

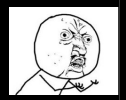
Dataset — Another Angle

- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms…

Sets

Dataset

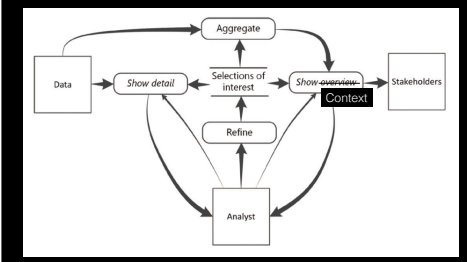
- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms…



Goal?

- ~~Make sense of whole dataset~~
- Explore a paper, a topic, an author, a venue…

Detail to Overview (Context) via Selection and Aggregation



Thank you

Knomos

MAPPING A KNOWLEDGE NETWORK OF LAW

Visual navigation platform for big data research and collaboration in the legal industry

Knomos

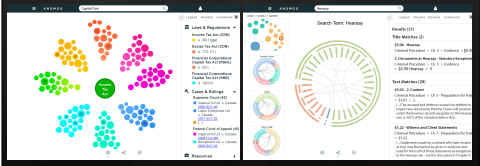
Law is Stuck

Legal research is constrained by:
- High search costs
- Decentralized sources
- Institutional barriers

- Outdated Content Format — Text-heavy, Static, and Linear
- Duplicate Search Costs — Firm Work Product Silo
- High Client Costs — Limited Access to Justice
- Content Barriers — Private Content Paywall
- Solitary Search — No Collective User Activity

Knomos

Enhanced Legal Research Software



Knomos

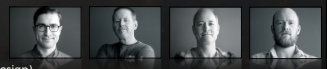
Team

History of industry leading software design applied to a unique visual platform for legal data research & collaboration

PS4 ACTIVISION EA DAVIES

Core Team

Adam La France (CEO)
James Abney (CTO)
Jesse Abney (COO)
Craig McInnes (Systems Design)



Key Partners

Open Data | Data BC | CanLII API

Knomos

Why Knomos?

Legal Research Usability Problems

User Driven Design

Networked Data: BCLaws & CanLII APIs

Choice of D3 Vis Idioms

Established Full Tech Stack w/ Room to Extend

Big Bang | GoT Reference

knomos