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

By Huaying Tian & Arthur Sun

OUTLINE

- What are we going to do?
- What are we going to do with the data?
- Why do we need visualization?
- What data do we abstract?
- How to visualize the data?
- Components of our analysis and function

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 → human-visualized graphs

astract key data types

reduce complexity into simplicity by counting data types

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:

	A	B	C	D	E	F	G	H	I	J	K	
1	_projectid	_teacher_acctid	_schoolid	school_n	school_la	school_l	school_ci	school_state	school_zi	school_me	school_district	school_coun
2	"e565fb42185c6e	"2e17c8c91cb58132d810	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
3	"76108ed46f99f2	"6b3721c9585633fa716e	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
4	"2568882e490684	"ed55d66251be5810b38e	"923a1f4850d4e09ad5b8d	3.60E+11	40.74221	-74.002	New York	NY	10011	urban	New York City Dept	New York (M
5	"e49f6d07f9da70	"ba7ada7b8ef61280b6e8	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
6	"548b161b4e808f1	"58574f5ee5f3d74a4be1	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
7	"dfc5e3a3bd22d8	"fedc7312fac70db43893	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
8	"be3c6a88818ef6	"2e6deb2edcb9b2ef9bf4	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
9	"dd7ed1d74f70fb	"d8f151a88a6d298fe5e6	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
10	"c5d4a7877c463e	"262af4de8528628259ac	"8c9ab8c021074381647582e1115072		40.82696	-73.8922	BRONX	NY	10459			Bronx
11	"8d94b238b50e7f	"f8e7fa9cc031249fa674	"15b661612d21c9dd37de0092ed81e0		40.82696	-73.8922	Bronx	NY	10459			Bronx
12	"8b19da4400111a	"760f8b3ac86e3ac286c6	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
13	"4f93f03be059a8	"e5d28482308e2007992a	"9bc7d4d92193a453669b6	3.60E+11	40.70178	-74.0118	New York	NY	10004	urban	Other	New York (M
14	"eef82859f1dd50	"8fb8a281a46527b9e7ea	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
15	"f38c2d5f55f407	"964aa5f2734897fbd99a	"eab5debe9b88fbb9b15fe	3.60E+11	40.78333	-73.9459	New York	NY	10128	urban	New York City Dept	New York (M
16	"ec6717d760d379	"8fb8a281a46527b9e7ea	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
17	"17466db6d83288	"dab65ac402878a610353	"1ae4695be589a36816188	3.60E+11	40.76552	-73.9601	New York	NY	10065		New York City Dept	New York (M
18	"f4a27a5e375c9a	"5955995b38eba124a51b	"5c5c444dd1bf31c3bdbcf	3.60E+11	40.76825	-73.989	New York	NY	10019	urban	Empowerment Support	New York (M
19	"52b04391f8e028	"ded4b1235dbdaedeb157	"8c9ab8c021074381647582e1115072		40.82696	-73.8922	BRONX	NY	10459			Bronx
20	"c97b3d603e4999	"263a93c35a3a52ea19ab	"cb1ac51c57e5ecb8f9a5a	3.60E+11	40.68683	-73.9808	Brooklyn	NY	11217	urban	New York City Dept	Kings (Broo
21	"0cab719d819193	"2e17c8c91cb58132d810	"45e7ddbdd7023f1eb65a6	3.60E+11	40.84169	-73.8755	Bronx	NY	10460	urban	New York City Dept	Bronx
22	"57d68f35b13a19	"f7fada7c19ac859803c3	"38ada3531c8f57cab78b0418bbc03e		40.64959	-73.9583	Brooklyn	NY	11226		New York City Dept	Kings (Broo
23	"b24064984f054b	"067ca75f07e29bc8daf2	"8c9ab8c021074381647582e1115072		40.82696	-73.8922	BRONX	NY	10459			Bronx
24	"9a3f66830e2112	"d1d684f1549f4e5b9693	"ec26adc6baee71abbcebf	3.60E+11	40.72369	-73.9816	New York	NY	10009	urban	New York City Dept	New York (M
25	"d1cffca7e85ba9	"fff493aaf228cd88040d	"756e5d578f2c6482c287b	3.60E+11	40.68494	-73.9473	Brooklyn	NY	11216	urban	New York City Dept	Kings (Broo
26	"9a6f4fb43bf27e	"5d11379c28c7a4103287	"60d87aa721f3a17c06289	3.60E+11	40.83826	-73.8661	Bronx	NY	10460	urban	New York City Dept	Bronx
27	"8aa59963a2a959	"ded4b1235dbdaedeb157	"8c9ab8c021074381647582e1115072		40.82696	-73.8922	BRONX	NY	10459			Bronx

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 → horizontal bar chart

donation counted by grade → horizontal bar chart

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

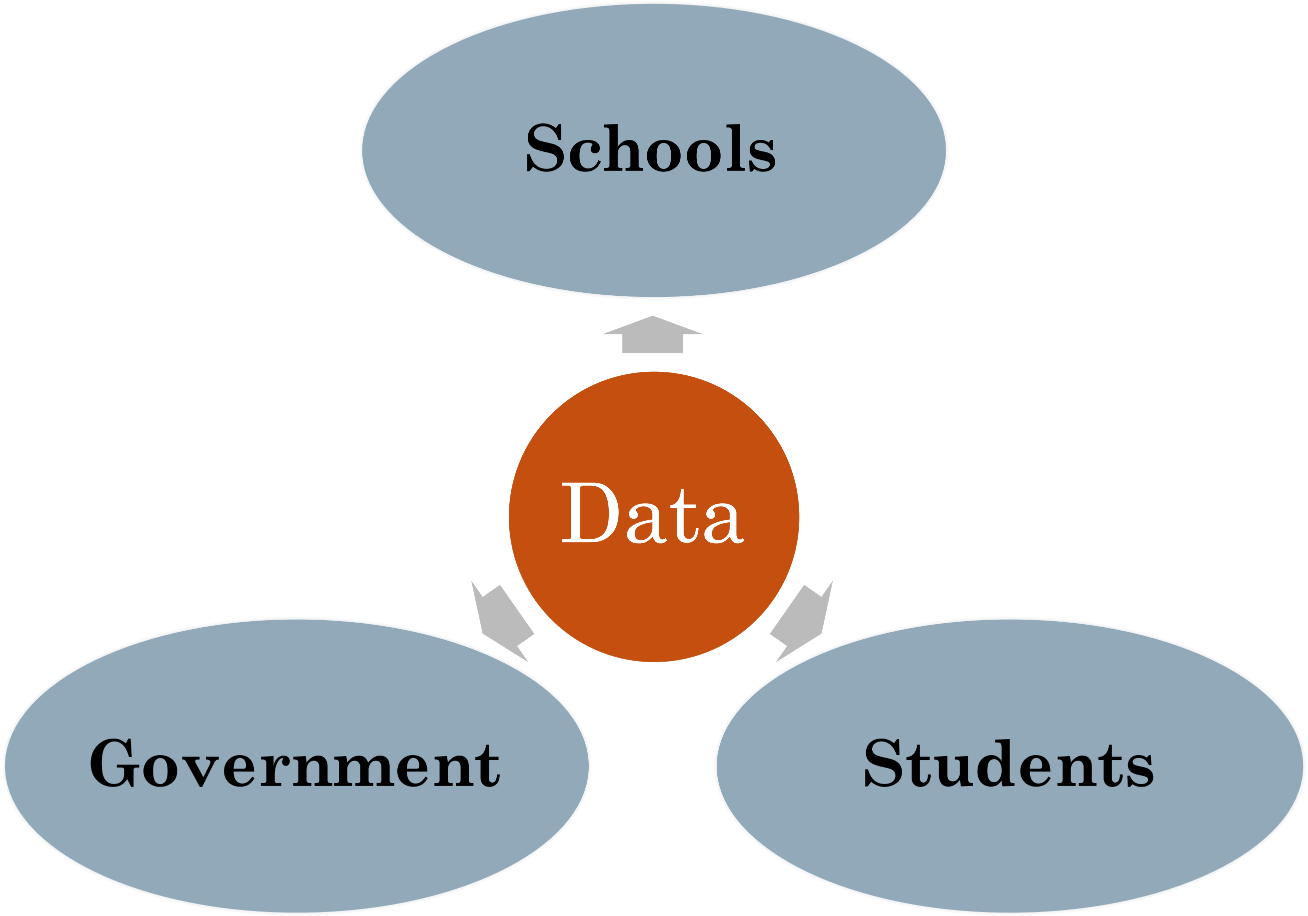
- Educational Census (public available, per year)
  - School code
  - School name
  - School type (private/public)
  - School location (Latitude, Longitude, Postal Code, City, District)
  - Census Year
  - Student Code
  - Student Grade
- Data size (per census year, we need at least two)
  - 7.789.831 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



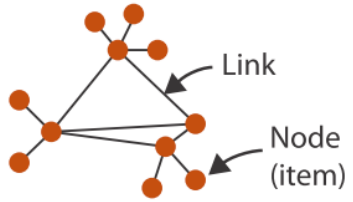
T1 - Tasks for Schools

- 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

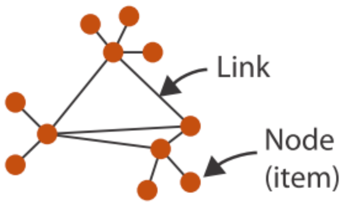
- Are there any areas of the state receiving more students than others?
- Are students migrating from public to private schools?

What?

➔ Dataset Types

➔ Geometry (Spatial)

➔ Networks



T1 - Help schools identify migration pattern of students

- map: school types

↓

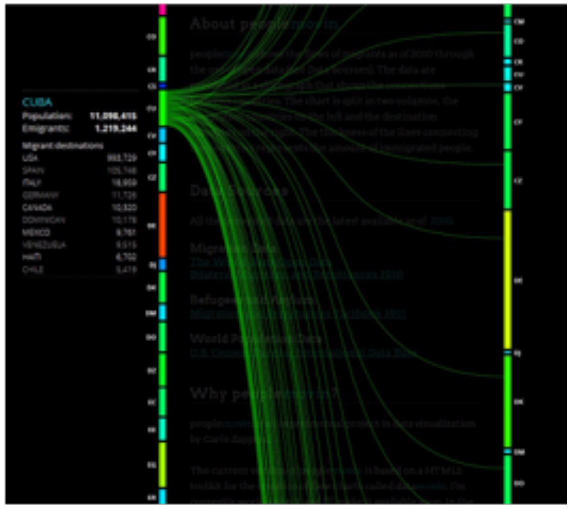
categorical color map

# of incoming & outgoing students

↓

line thickness

- interaction: select
- facet: juxtapose



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

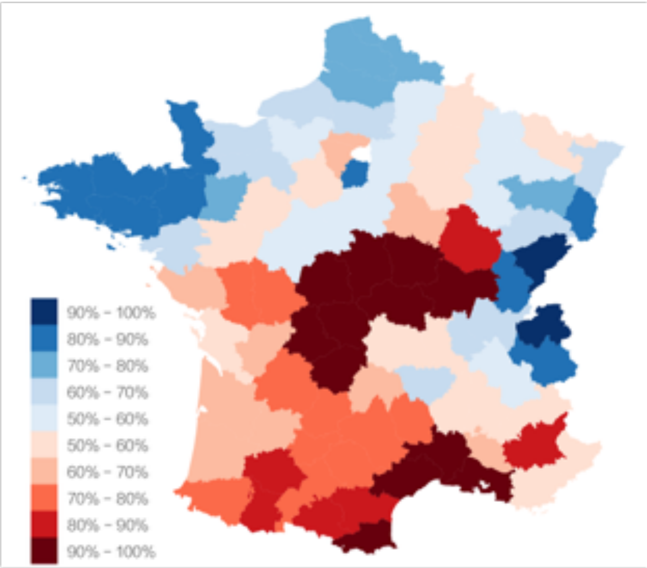
- map: (incoming-outgoing)%

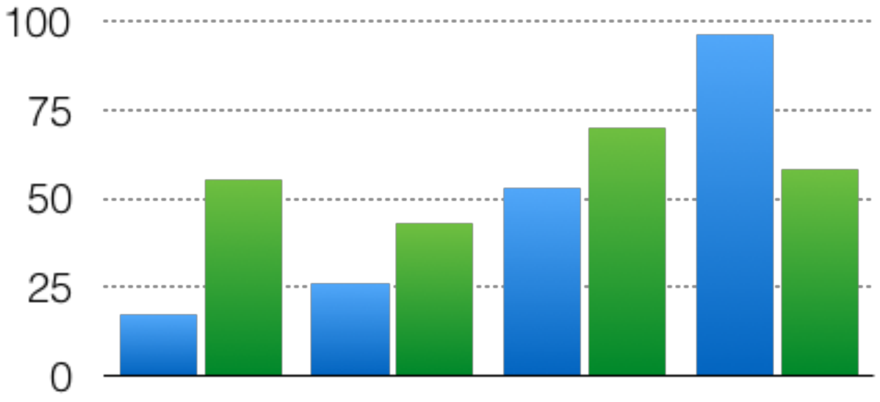
diverging color map

↓

- interaction: select

- facet: juxtapose





Thank you!

Carolina Roman Amigo

carolamigo@gmail.com

Wenqiang (Dylan) Dong

wdong@cs.ubc.ca

VISUALIZATION OF YOUTUBE COMMENTS


Yeah, the posing isn't quite
I'm hosting a seminar where you can
re advanced posing techniques next
just head over to my website at.....
Unlike · 2
iPhone correct? Might want to try
a point and shoot digital. Best Buy has so
or under $300!!!
Blue & Green color com
2 hrs · Unlike · 1
some of the flaws in MS Paint, no ne
Photoshop. You also need to remem
of 3rds when shooting. That would h
for a much better photo. There are
videos to help you out on You-Tube
1 hr · Like
would never
those flowers to shoot in front of. D
eyes. I've definitely done better.
How come its gotta be a white
racist!
2
NICE DOF DUMBASS
2 hrs · Unlike · 1
It would look
the outside black
keep her shoes colored. Also,
that shot with Janel Jackson
behind her holding her breast
neat. But she should smile ar
because your model looks re
2 hrs · Unlike · 2
your watermark is distracting
hite. you should do it like me...
etters in the bottom corner of every
worry... nobody on the itnernet will
2
Omg Nino Bat
dited those limbs from ir
leg. I can't believe you
ion like that
fancy camera makes. You
long way with your hobby.
2 hrs · Unlike · 2
you really just obj
can you live with
hoo haa..... (shar
www.
A little tip on posing.. H
and middle finger are to close together
ems she had webbed alien hands. I have
Get this snapshot outta here
pposed to shoot outsi
I know you are already finishe
with this but....

- Doesn’t support easy finding of entertaining comments.

- Emotionally draining arguments and trolls.

- Doesn't support easy finding of entertaining comments.

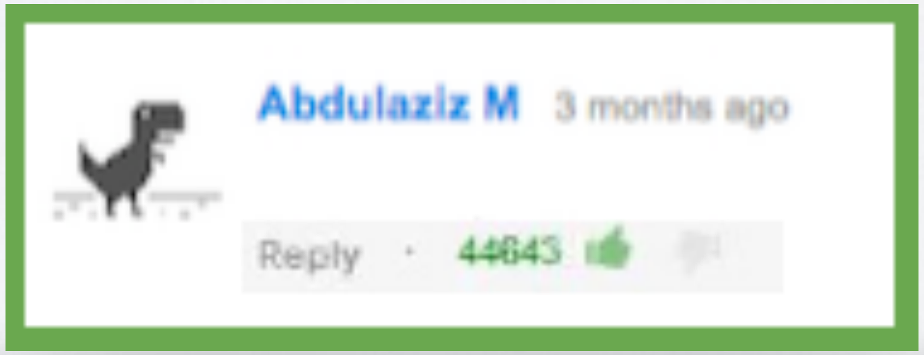
  Task 1: Explore for entertaining comments.

- Emotionally draining arguments and trolls.

  Task 2: Identify arguments.

  Task 3: Identify trolls.

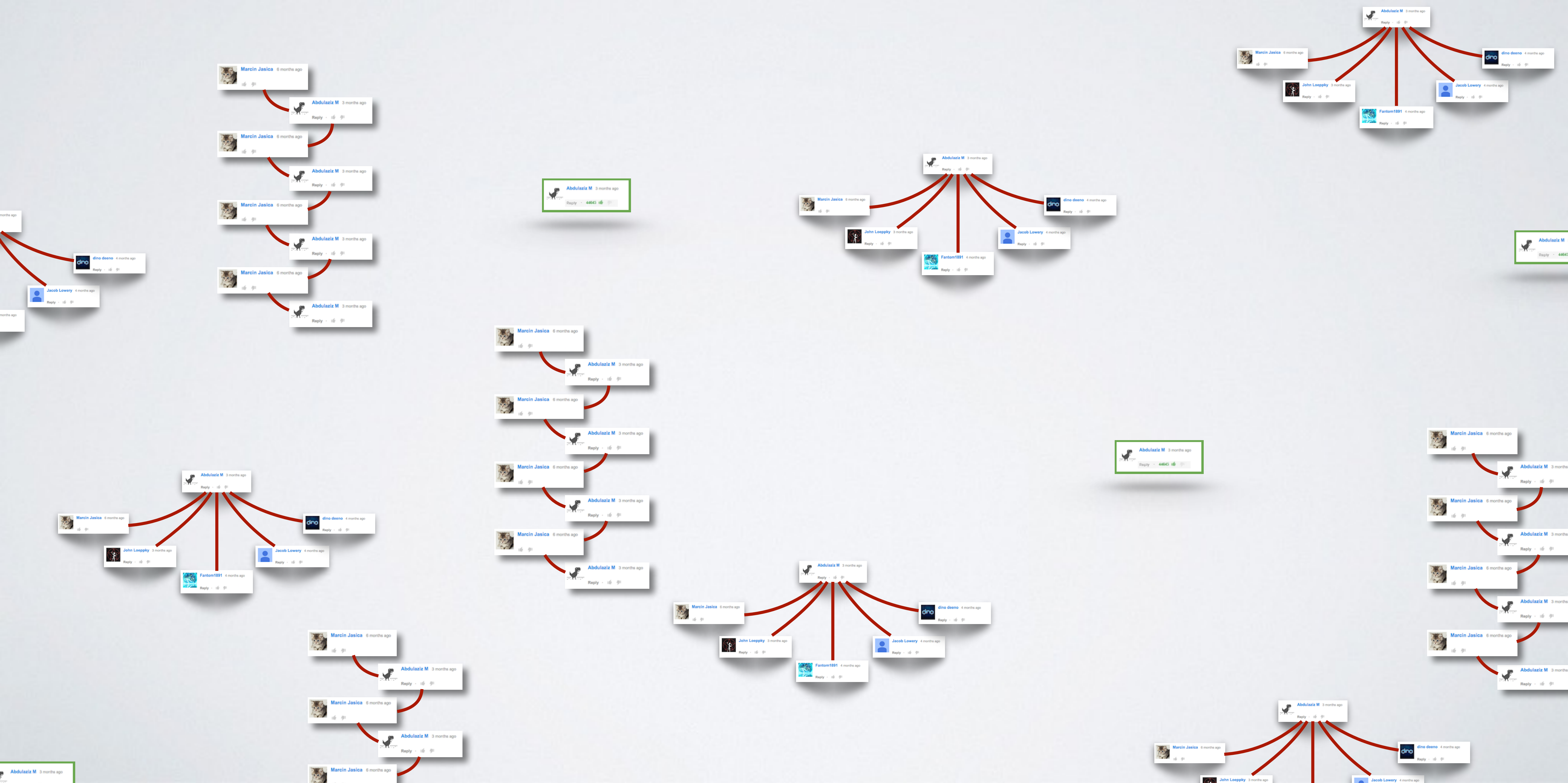
- Entertaining comments = highly liked comments (generally)



- 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



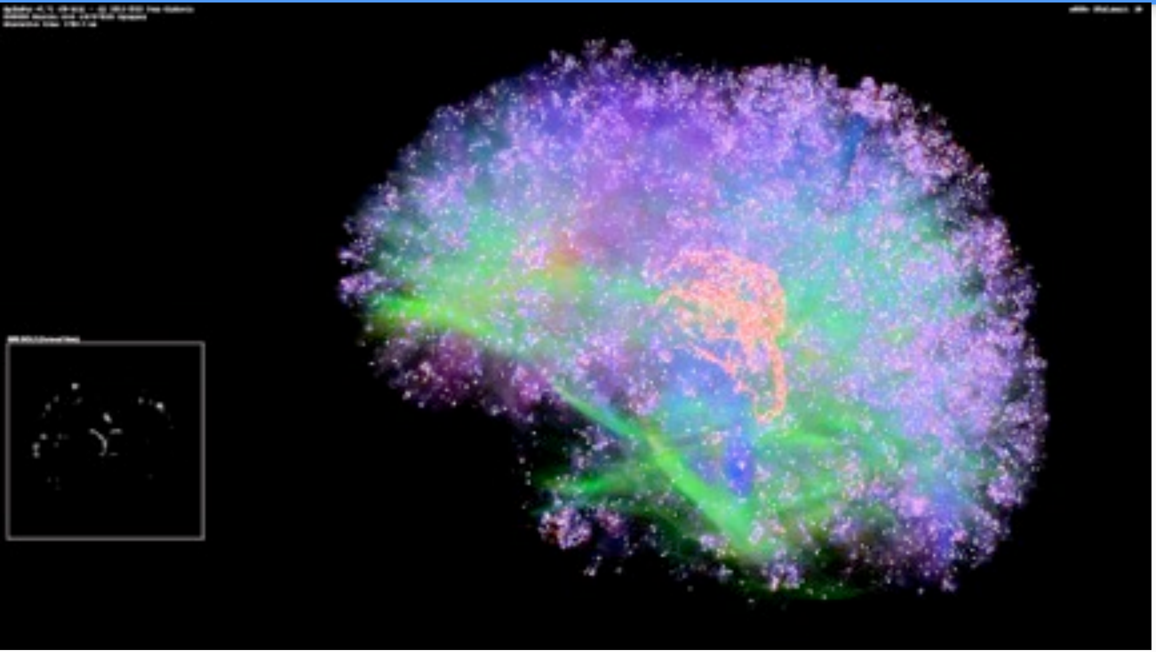
Neuron electrophysiology data visualization (Neuroelectro)

Presented by: Dmitry, Emily and Mike

Introduction

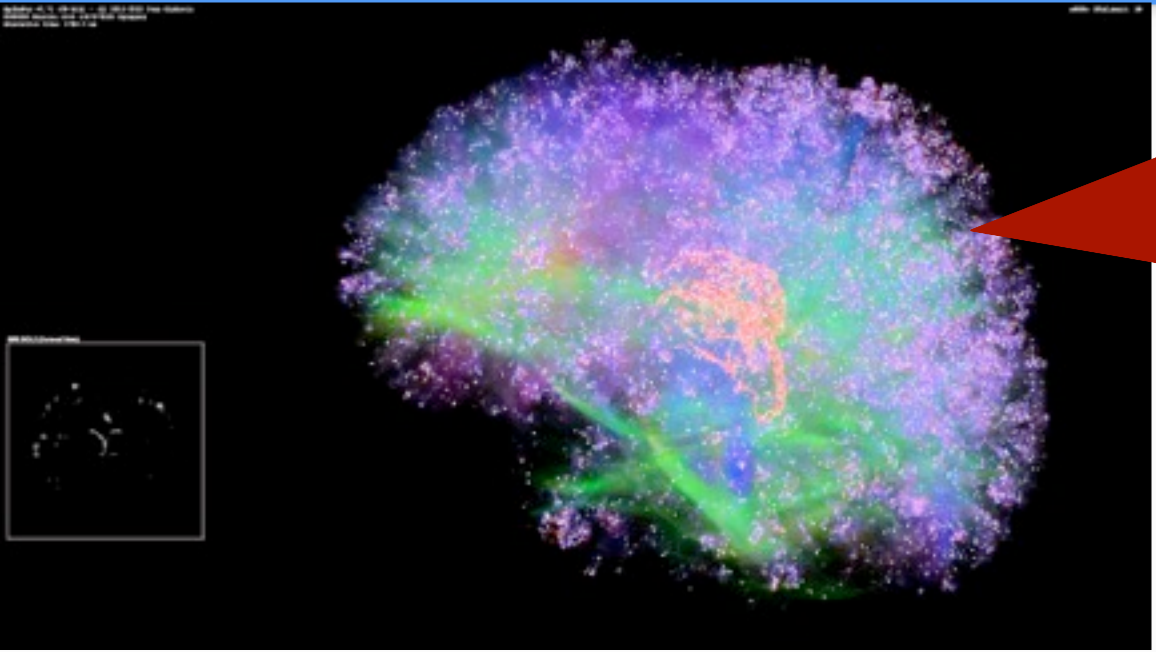
Dmitri

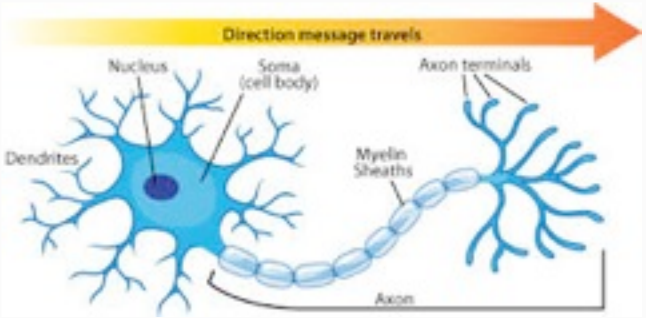
How does your brain work? It’s complicated



https://www.youtube.com/watch?v=u28ijlP6L6M

How does your brain work? It’s complicated





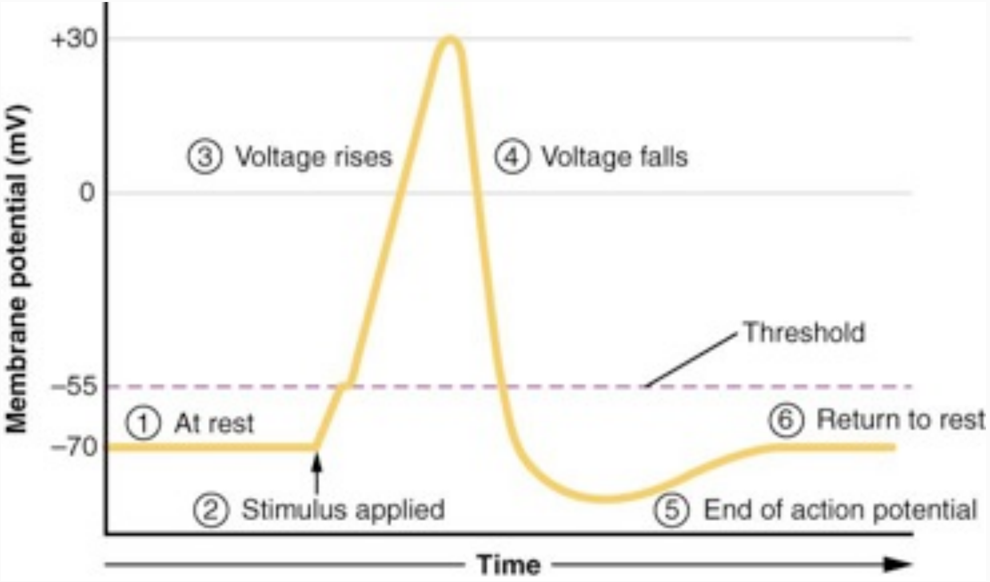
askabiologist.asu.edu

https://www.youtube.com/watch?v=u28ijlP6L6M

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.

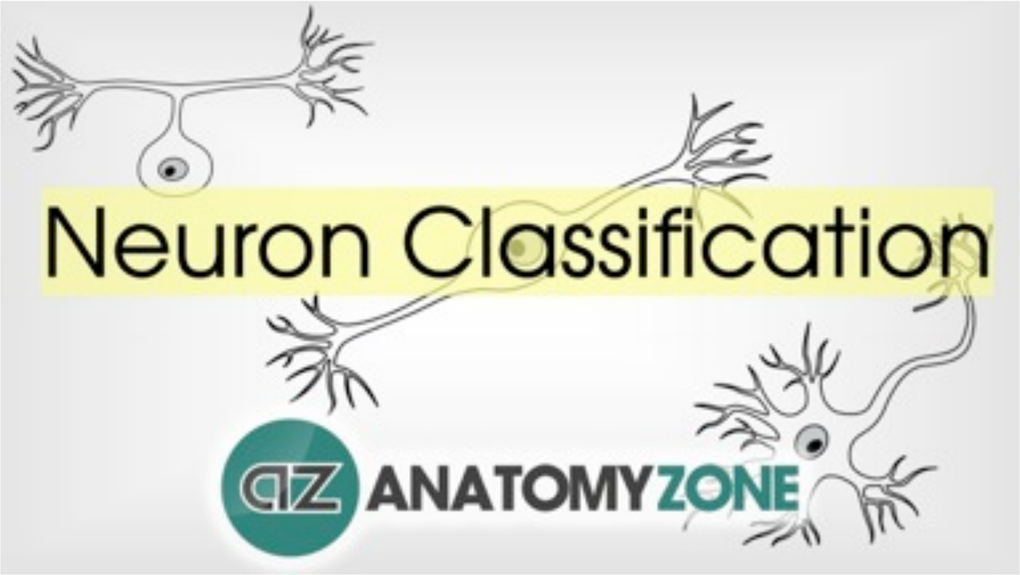
-wikipedia



courses.candelalearning.com

What is our data?

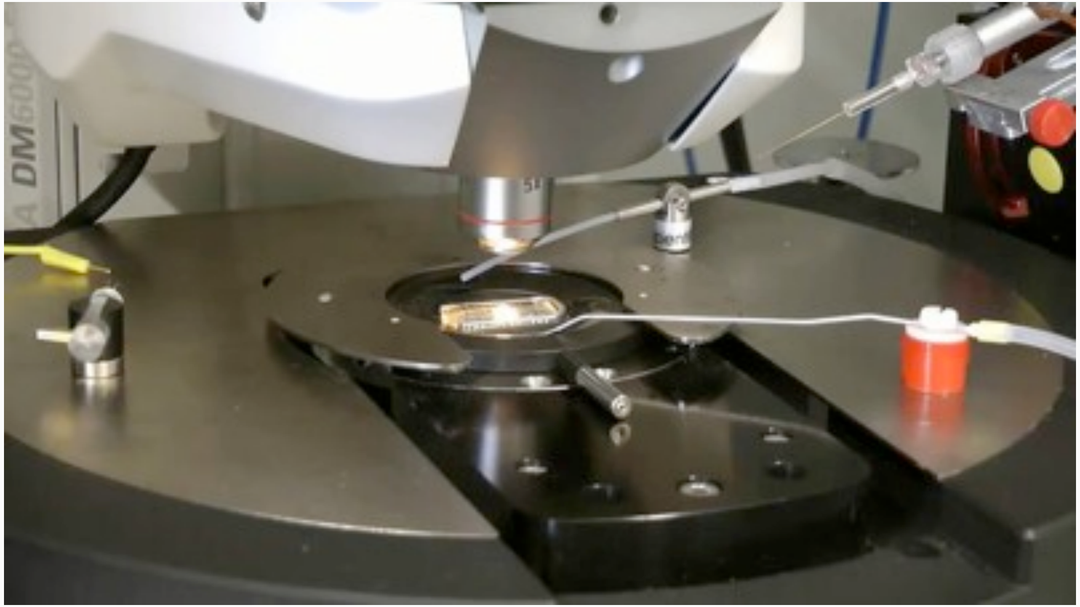
How many neuron types are there? The debate has been ongoing for decades. We use enhanced NeuroLex.org definitions (~100 Neuron types)



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

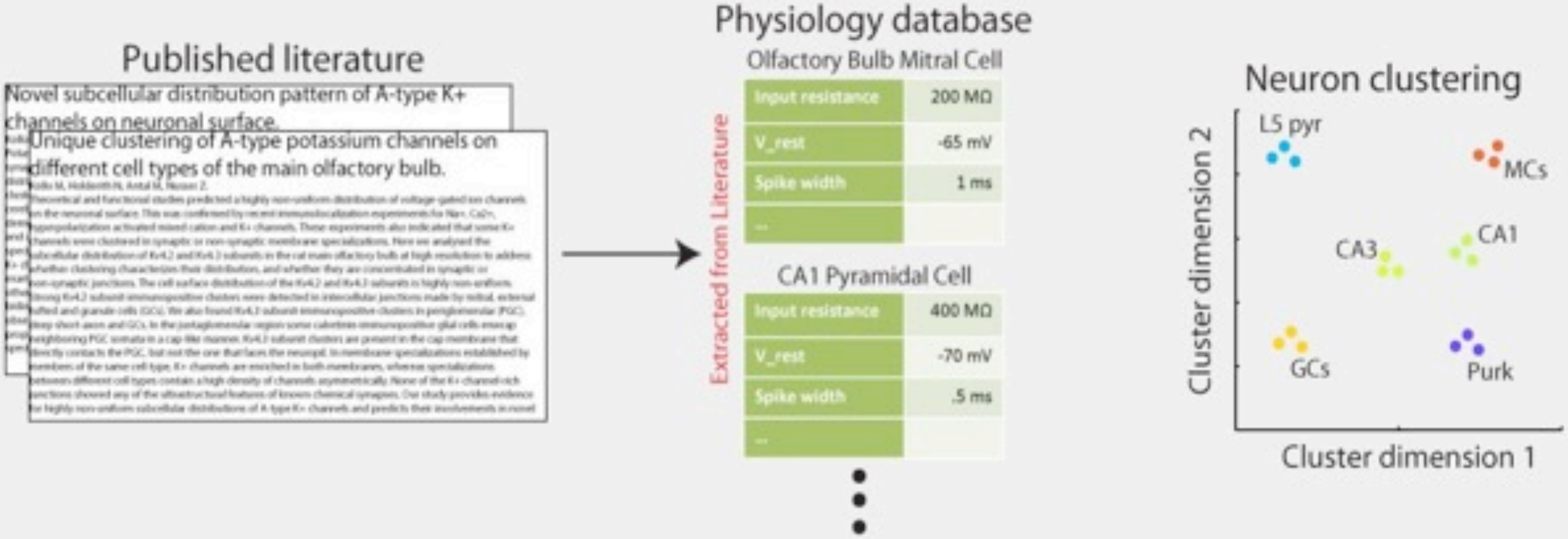
Mike

NeuroElectro Publications

About | Neuron Types | Electrophysiology Properties | Articles | FAQs | Data/API | Contribute

NeuroElectro: *organizing information on cellular neurophysiology.*

The goal of the NeuroElectro Project is to extract information about the electrophysiological properties (e.g. resting membrane potentials and membrane time constants) of diverse neuron types from the existing literature and place it into a centralized database.



Our goal is to facilitate the discovery of neuron-to-neuron relationships and better understand the role of functional diversity across neuron types.

NeuroElectro

NeuroElectro Publications

About | Neuron Types | Electrophysiology Properties | Articles | FAQs | Data/API | Contribute

Listing of neuron types in the database

Electrophysiology values across neuron types obtained are obtained from semi-automated literature text-mining.
(neuron types are mostly from NeuroLex.org)

Show 10 entries

Search:

Neuron type	Number extracted electrophysiology values	Number articles
Other	1119	70
Hippocampus CA1 pyramidal cell	505	66
Dorsal root ganglion cell	353	23
Neocortex pyramidal cell layer 5-6	281	33
Neocortex basket cell	185	25
Neostriatum medium spiny neuron	163	23
Neocortex pyramidal cell layer 2-3	144	19
Neocortex uncharacterized cell	143	12
Dentate gyrus granule cell	107	18
Medial vestibular nucleus neuron	102	7

Showing 1 to 10 of 235 entries

Previous Next

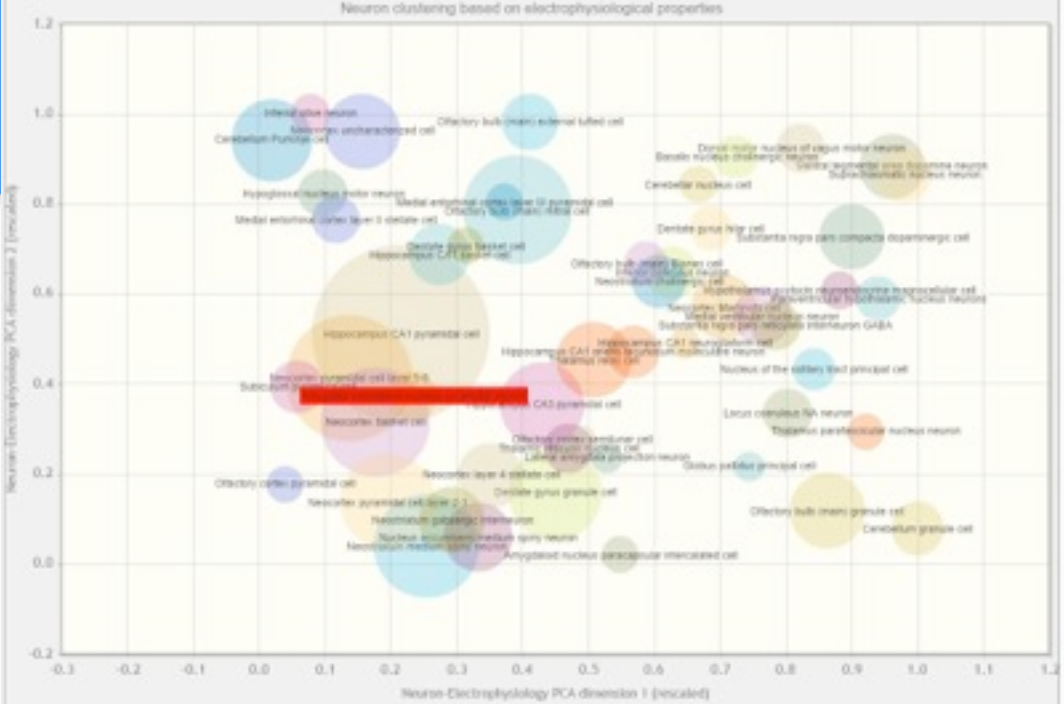
Clustering of neuron types based on similarities in electrophysiological values:

Description:

- Projection of NeuroElectro neuron types onto space defined by the first 2 electrophysiology principal components which collectively represent ~70% of variance. Neuron types are indicated by circles and circle size is proportional to number of corresponding articles indexed in NeuroElectro database.

Interactivity:

- Click on neuron types to go to corresponding neuron page.
- Zoom in on a section of plot by dragging cursor. Zoom out by double clicking on plot.



Show neuron names

Don't show neuron names

Display a specific neuron(s):

None selected
Amygdala basolateral nucle
Amygdaloid nucleus paraca
Basalis nucleus cholinergic
Cerebellar nucleus cell
Cerebellum granule cell

Electrophysiological properties of *Amygdala basolateral nucleus pyramidal neurons* from literature:

Legend:

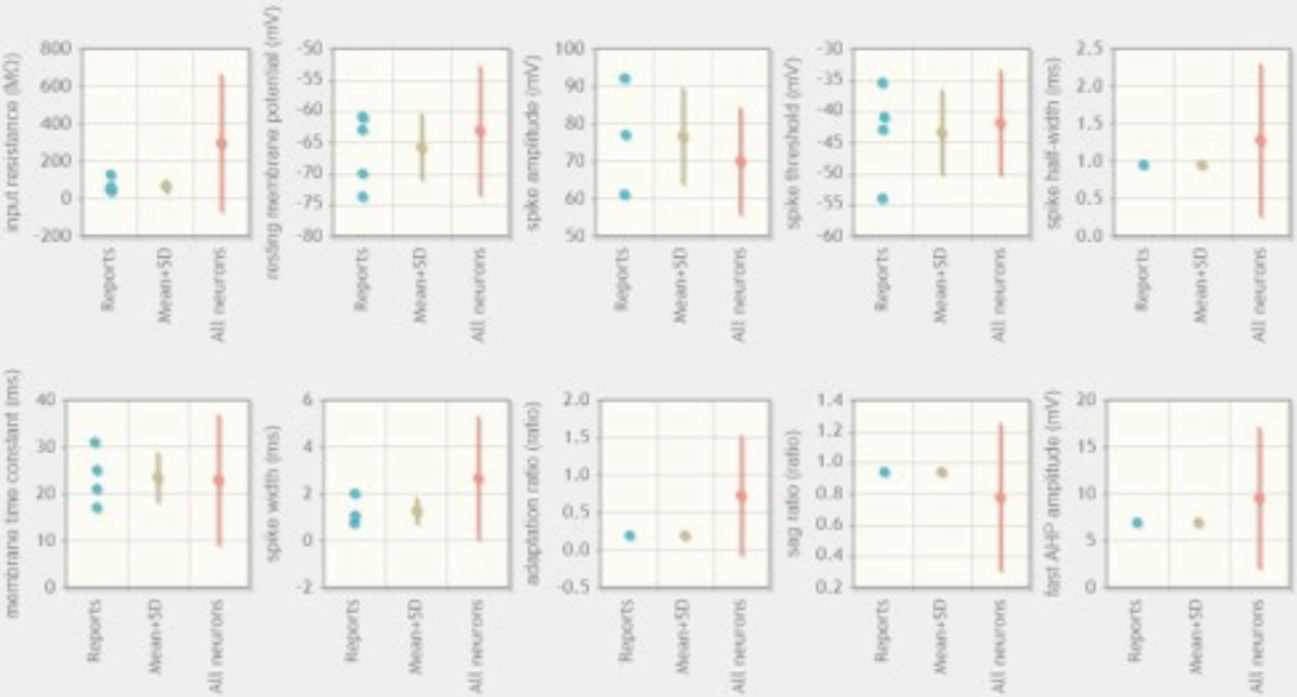
- Reports: Blue dots indicate human-curated values; Orange dots indicate non-human curated values
- Mean+SD: mean and standard deviation of human-curated neuron measurements
- All neurons: mean and standard deviation computed over all neurons in database

Interactivity:

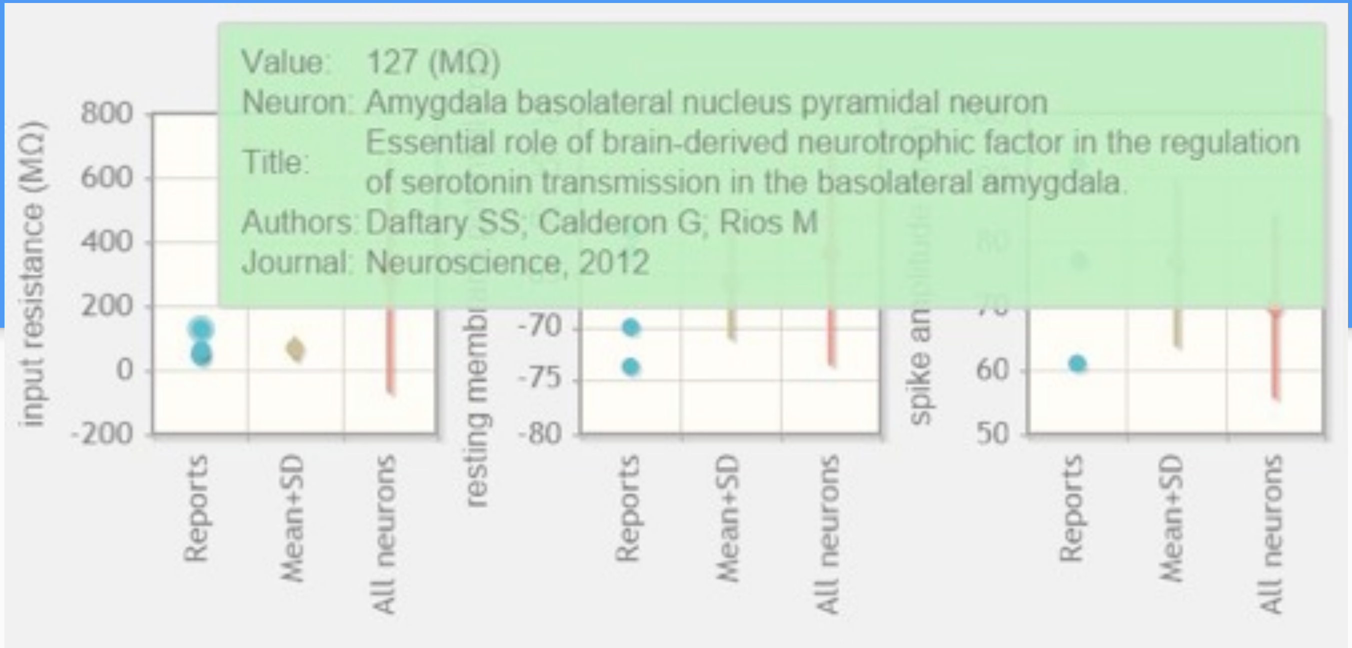
- Mouse over neuron report data points and click to view corresponding publication
- Mouse over y- axis labels to view definition or click to view values across neuron types
- Zoom in on a section of plot by dragging cursor. Zoom out by double clicking on plot.
- Legend: Blue dots = text-mined values human curated; Orange dots = text-mined values not human curated

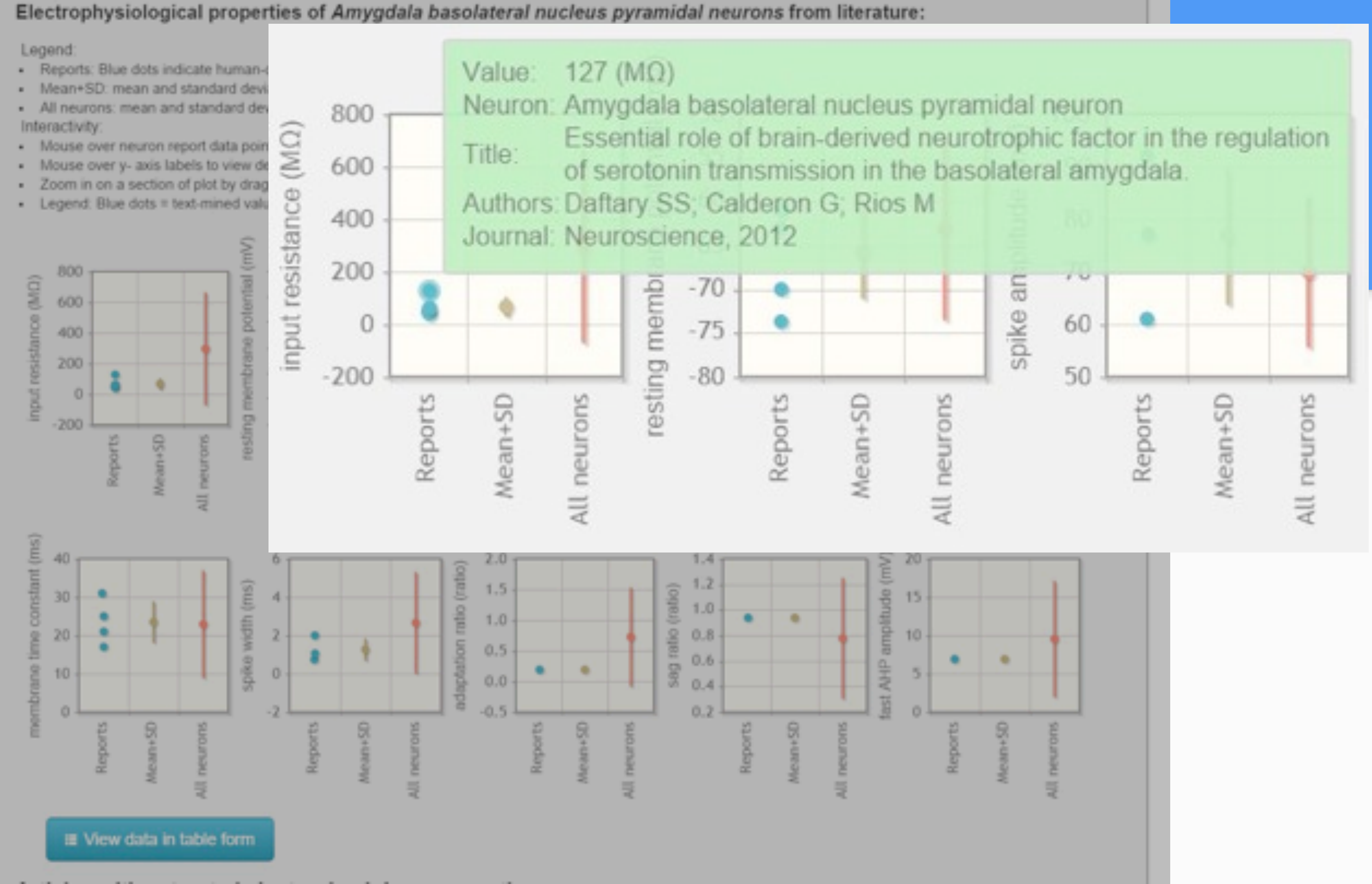
View data in table form

Report miscurated data



View data in table form


Value: 127 (MΩ)
Neuron: Amygdala basolateral nucleus pyramidal neuron
Title: Essential role of brain-derived neurotrophic factor in the regulation of serotonin transmission in the basolateral amygdala.
Authors: Daftary SS; Calderon G; Rios M
Journal: Neuroscience, 2012
input resistance (MΩ)
800
600
400
200
0
-200
-70
-75
-80
60
50
Reports
Mean+SD
All neurons
Reports
Mean+SD
All neurons
Reports
Mean+SD
All neurons


Electrophysiological properties of Amygdala basolateral nucleus pyramidal neurons from literature:
Legend:
Reports: Blue dots indicate human-
Mean+SD: mean and standard devi
All neurons: mean and standard dev
Interactivity:
Mouse over neuron report data poin
Mouse over y- axis labels to view de
Zoom in on a section of plot by drag
Legend: Blue dots = text-mined valu
Value: 127 (MΩ)
Neuron: Amygdala basolateral nucleus pyramidal neuron
Title: Essential role of brain-derived neurotrophic factor in the regulation of serotonin transmission in the basolateral amygdala.
Authors: Daftary SS; Calderon G; Rios M
Journal: Neuroscience, 2012
input resistance (MΩ)
resting membrane potential (mV)
spike amplitude
membrane time constant (ms)
spike width (ms)
adaptation ratio (ratio)
sag ratio (ratio)
fast AHP amplitude (mV)
Reports
Mean+SD
All neurons
View data in table form

Article: *Essential role of brain-derived neurotrophic factor in the regulation of serotonin transmission in the basolateral amygdala.*

Full Text (publisher's website) ; Article Metadata ; Article Data (extracted)

Daftary SS, Calderon G, Rios M

Neuroscience, 2012

Table 1Physiological properties of BLA pyramidal neurons

Parameter	WT *Concept: Amygdala basolateral nucleus pyramidal neuron*	BDNF2L/2LCk-Cre
Vm (mV) Concept: *resting membrane potential*	−61±1.2	−63±1.1
Rm (mΩ) Concept: *input resistance*	127±12	110±10
Tau (ms) Concept: *membrane time constant*	25±2.3	23±1.8
Spike threshold (ms) Concept: *spike threshold*	−41±0.6	−40±2.8
Spike amp. (mV) Concept: *spike amplitude*	77±3.1	77±2.5
Half-spike width (ms) Concept: *spike half width*	0.95±0.04	1.02±0.05
10–90% rise time (ms) Concept: *spike rise time*	0.34±0.01	0.37±0.02

Vm, resting membrane potential; Rm, membrane input resistance. No significant difference was observed in the physiological properties of BLA projection neurons from controls and mutants ($p > 0.05$ for all categories).

Report miscurated data

Inferred neuron-electrophysiology data values

Show 50 entries Search:

Neuron Type	Neuron Description	Ephys Prop	Extracted Value	Standardized Value	Content Source
Amygdala basolateral nucleus pyramidal neuron		membrane time constant	25.0 ± 2.3	25.0 (ms)	Data Table
Amygdala basolateral nucleus pyramidal neuron		spike threshold	-41.0 ± 0.6	-41.0 (mV)	Data Table
Amygdala basolateral nucleus pyramidal neuron		spike amplitude	77.0 ± 3.1	77.0 (mV)	Data Table
Amygdala basolateral nucleus pyramidal neuron		spike half width	0.95 ± 0.04	0.95 (ms)	Data Table
Amygdala basolateral nucleus pyramidal neuron		spike rise time	0.34 ± 0.01	0.34 (ms)	Data Table
Amygdala basolateral nucleus pyramidal neuron		resting membrane potential	-61.0 ± 1.2	-61.0 (mV)	Data Table
Amygdala basolateral nucleus pyramidal neuron		input resistance	127.0 ± 12.0	127.0 (MΩ)	Data Table

Showing 1 to 7 of 7 entries

Previous Next

Input resistance

Common definition: Input resistance measured at steady-state voltage response to current injection

Electrophysiological values of *input resistance* across neuron types from literature:

Standardization criteria:

- Values corrected for differences in units, but are otherwise unchanged. Refer to individual articles for specific definitions and calculation methodologies.

Legend:

- Blue dots = text-mined values human curated; Orange dots = text-mined values not human curated

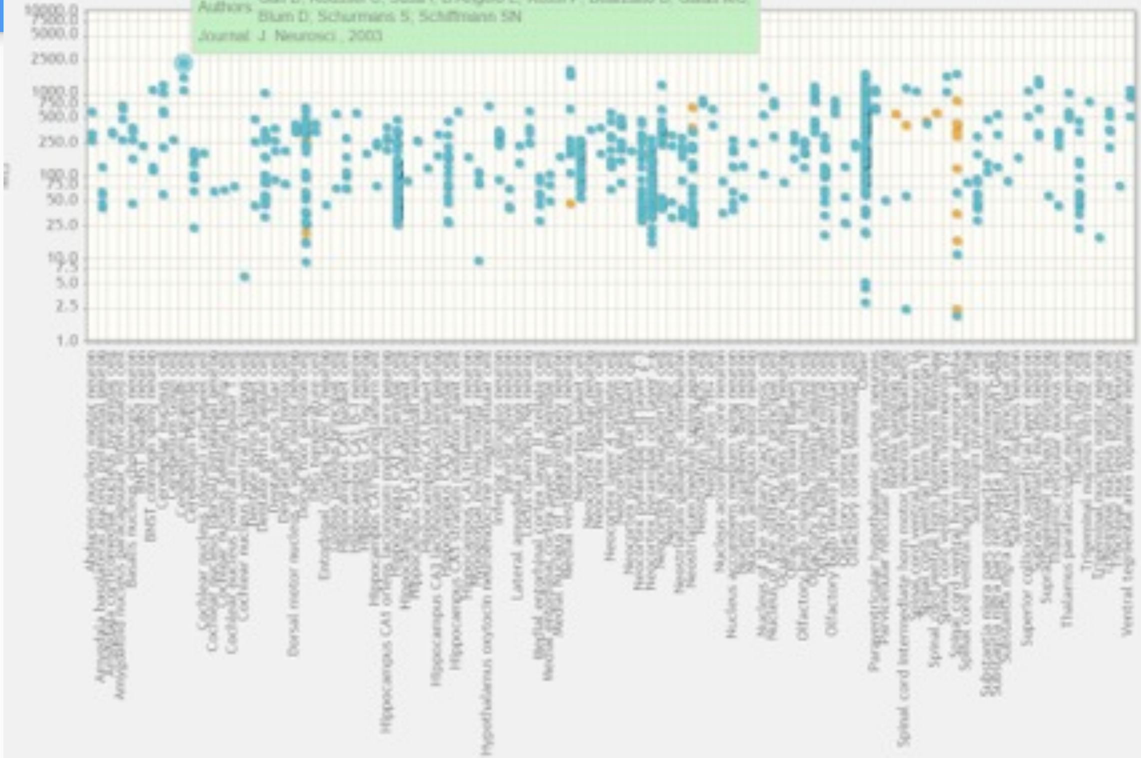
Interactivity:

- Mouse over neuron
- Click on neuron nam
- Zoom in on a section

Value: 2300 (MΩ)
Neuron: Cerebellum granule cell
Title: Altered neuronal excitability in cerebellar granule cells of mice lacking calretinin.
Authors: Gall D; Roussel C; Susa I; D'Angelo E; Rossi P; Bearzatto B; Galas MC; Blum D; Schurmans S; Schiffmann SN
Journal: J. Neurosci., 2003

View data in table form

Report miscurated data



Listing of articles with extracted electrophysiology properties

Article title links out to pubmed abstract

Show 10 entries

Search:

Article Title	Authors	Journal	Year	Electrophys values	Neuron types
Are all spinal segments equal: intrinsic membrane properties of superficial dorsal horn neurons in the developing and mature mouse spinal cord. (NeuroElectro data) (PubMed)	Tadros MA; Harris BM; Anderson WB; Brichta AM; Graham BA; Callister RJ	J. Physiol. (Lond.)	2012	90	1
Morphological and electrophysiological properties of pyramidal-like neurons in the stratum oriens of Cornu ammonis 1 and Cornu ammonis 2 area of Proechimys. (NeuroElectro data) (PubMed)	Scorza CA; Araujo BH; Leite LA; Torres LB; Otalora LF; Oliveira MS; Garrido-Sanabria ER; Cavalheiro EA	Neuroscience	2011	88	3
Target-specific output patterns are predicted by the distribution of regular-spiking and bursting pyramidal neurons in the subiculum. (NeuroElectro data) (PubMed)	Kim Y; Spruston N	Hippocampus	2012	88	2
Hyperexcitability of axotomized and neighboring unaxotomized sensory neurons is reduced days after perineural clonidine at the site of injury. (NeuroElectro data) (PubMed)	Liu B; Eisenach JC	J. Neurophysiol.	2005	78	1
The largest group of superficial neocortical GABAergic interneurons expresses ionotropic serotonin receptors. (NeuroElectro data) (PubMed)	Lee S; Hjerling-Leffler J; Zagha E; Fishell G; Rudy B	J. Neurosci.	2010	69	2
Lateral hypothalamic GAD65 neurons are spontaneously firing and distinct from orexin- and melanin-concentrating hormone neurons. (NeuroElectro data) (PubMed)	Karnani MM; Szabó G; Erdélyi F; Burdakov D	J. Physiol. (Lond.)	2013	56	1
5-HT(3A) receptor-bearing white matter interstitial GABAergic interneurons are functionally integrated into cortical and subcortical networks. (NeuroElectro data) (PubMed)	von Engelhardt J; Khrulev S; Eliava M; Wahlster S; Monyer H	J. Neurosci.	2011	55	1
Cellular neuroanatomy of rat presubiculum. (NeuroElectro data) (PubMed)	Simonnet J; Eugène E; Cohen I; Miles R; Fricker D	Eur. J. Neurosci.	2013	52	1
Inter- and intralaminar subcircuits of excitatory and inhibitory neurons in layer 6a of the rat barrel cortex. (NeuroElectro data) (PubMed)	Kumar P; Ohana O	J. Neurophysiol.	2008	50	2
Synaptic interactions between pyramidal cells and interneurone subtypes during seizure-like activity in the rat hippocampus. (NeuroElectro data) (PubMed)	Fujiwara-Tsukamoto Y; Isomura Y; Kaneda K; Takada M	J. Physiol. (Lond.)	2004	49	5

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...

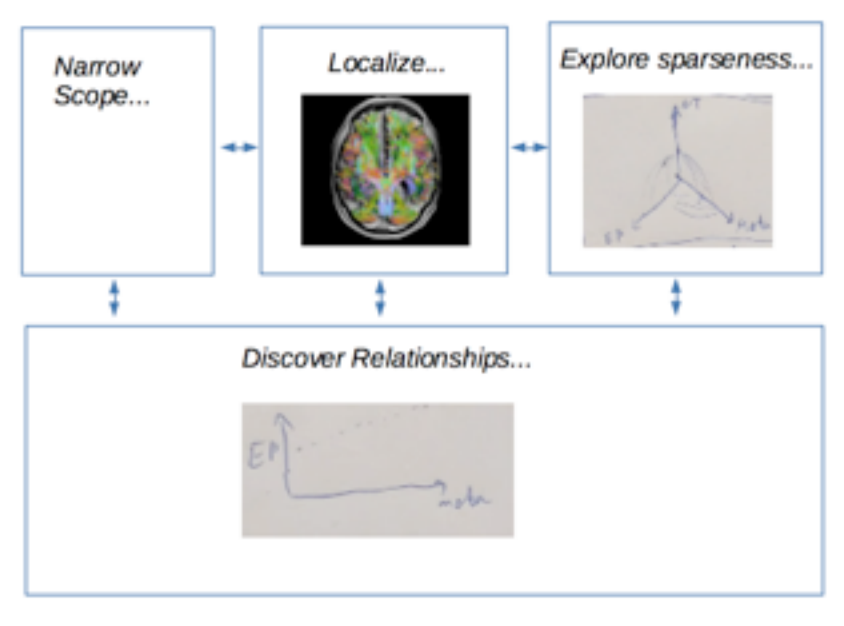
Task Analysis

- *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

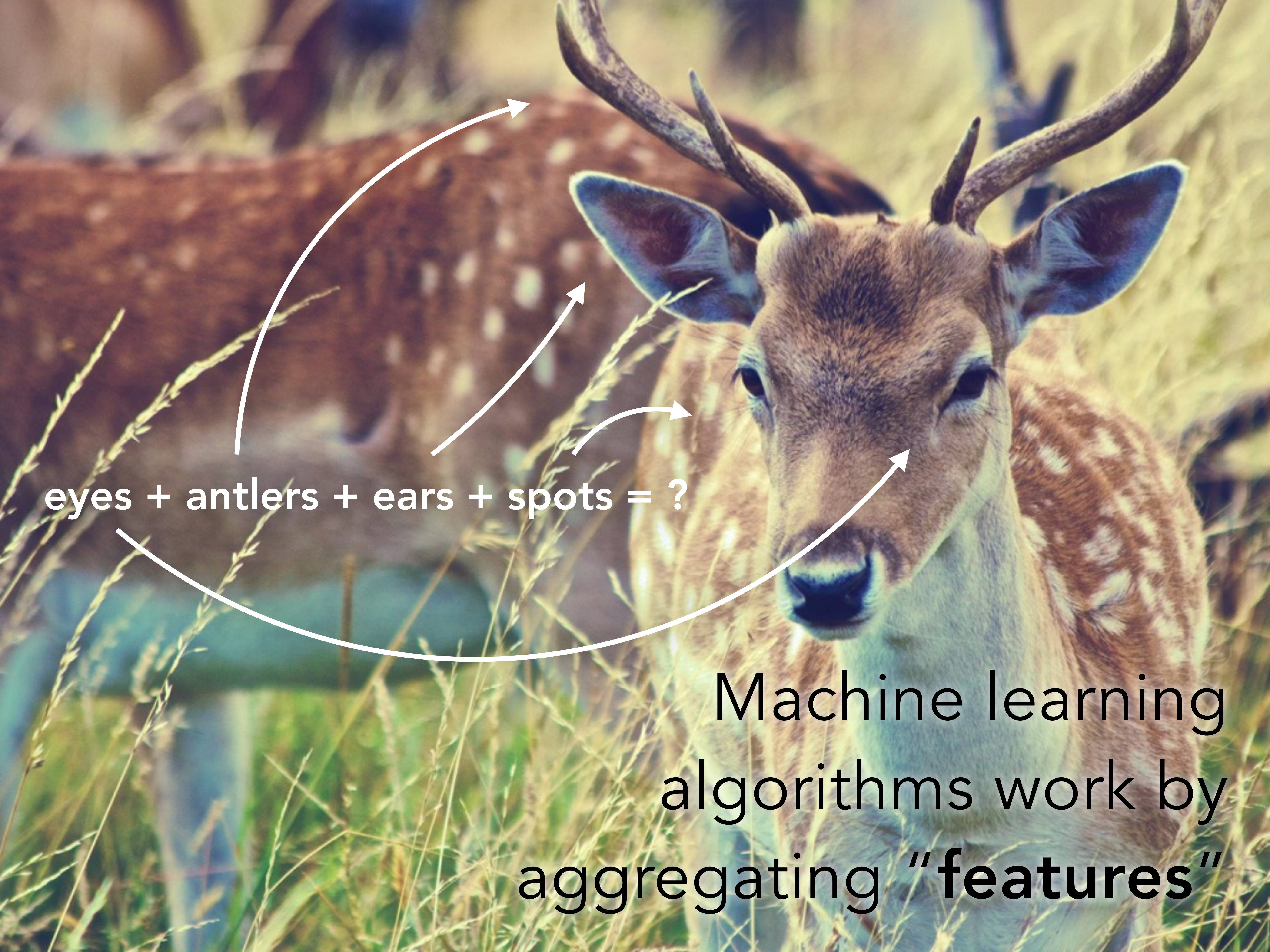
Existing Visualizations

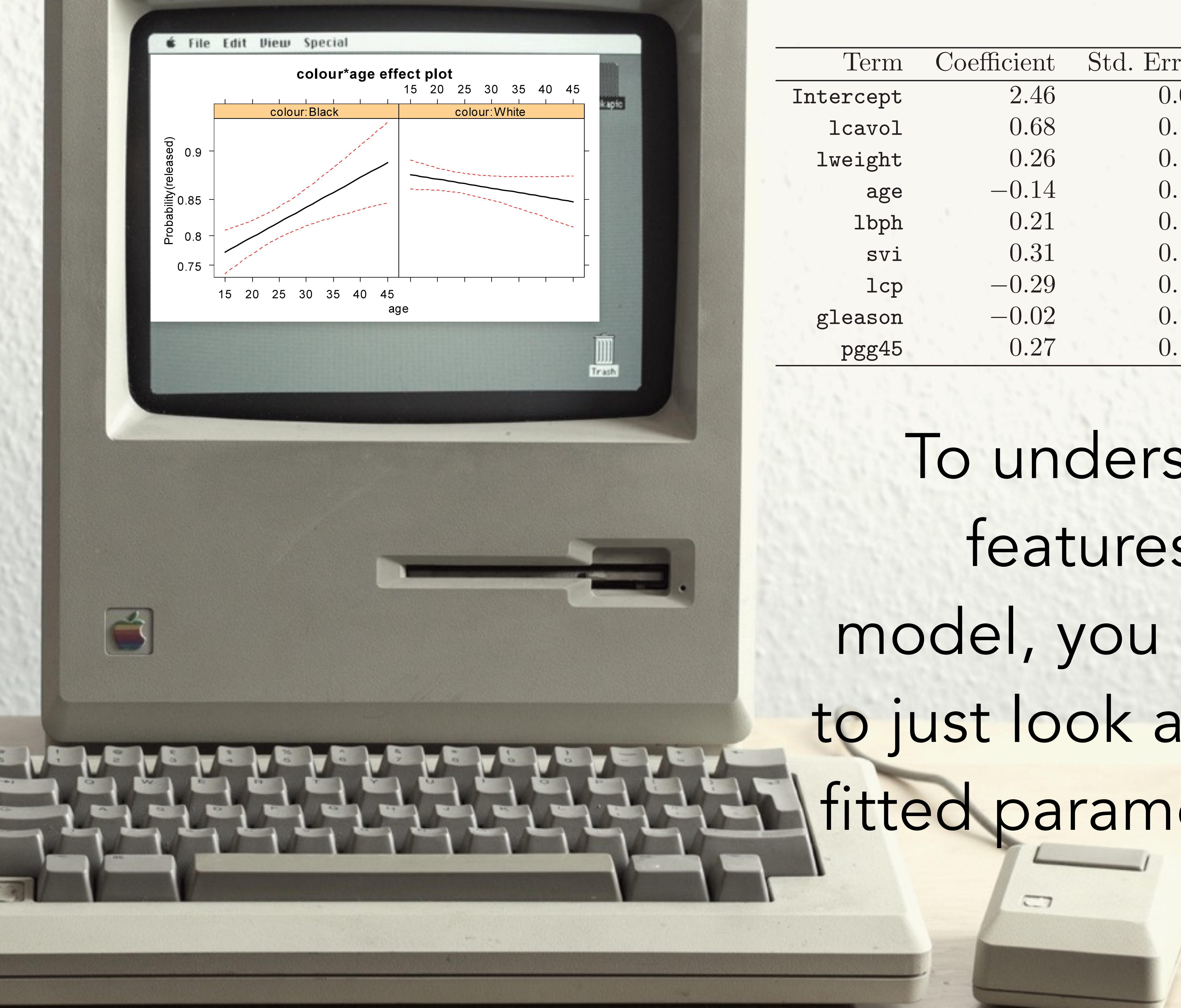
- Calendar
- Pie chart
- Bar chart
- …

Partner welcome
(no pressure)

VISUALISING FEATURE LEARNING

Jason Hartford


eyes + antlers + ears + spots = ?
Machine learning algorithms work by aggregating "**features**"



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

Modern models learn features from data

may have…

100s

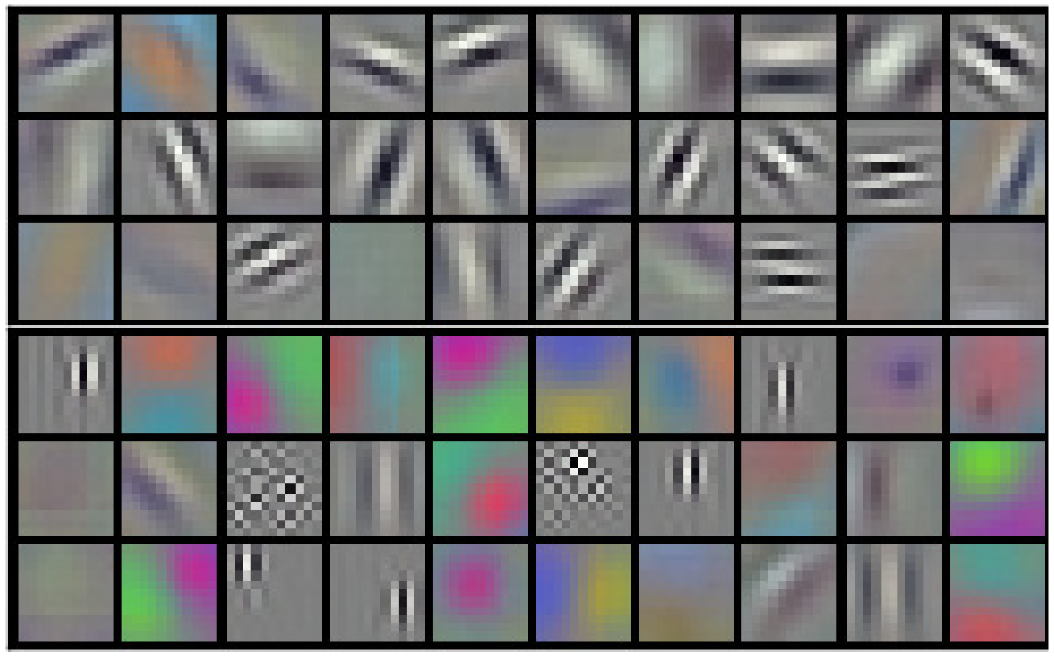
1000s

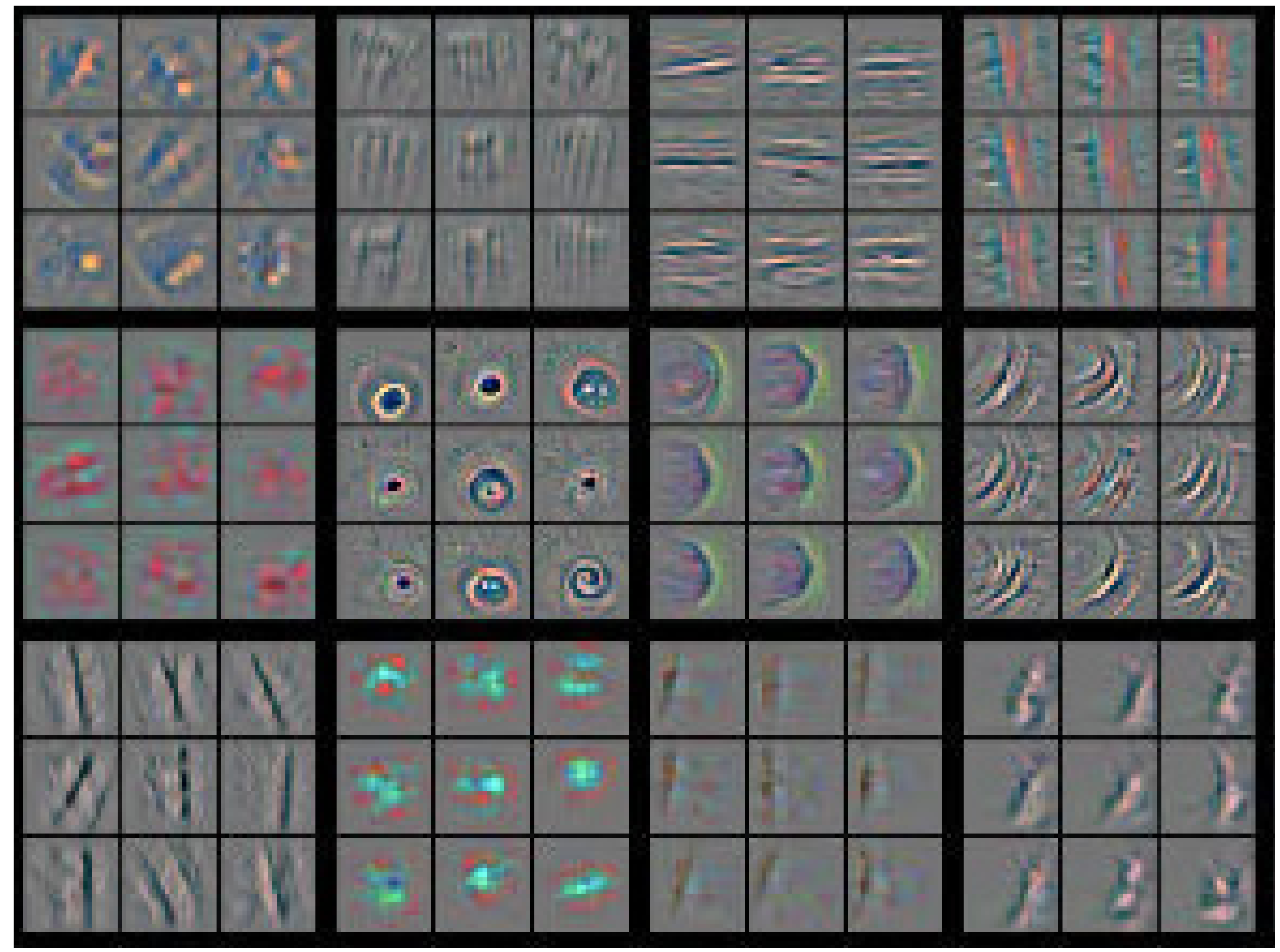
Millions

of parameters

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

In vision you can plot parameters directly.

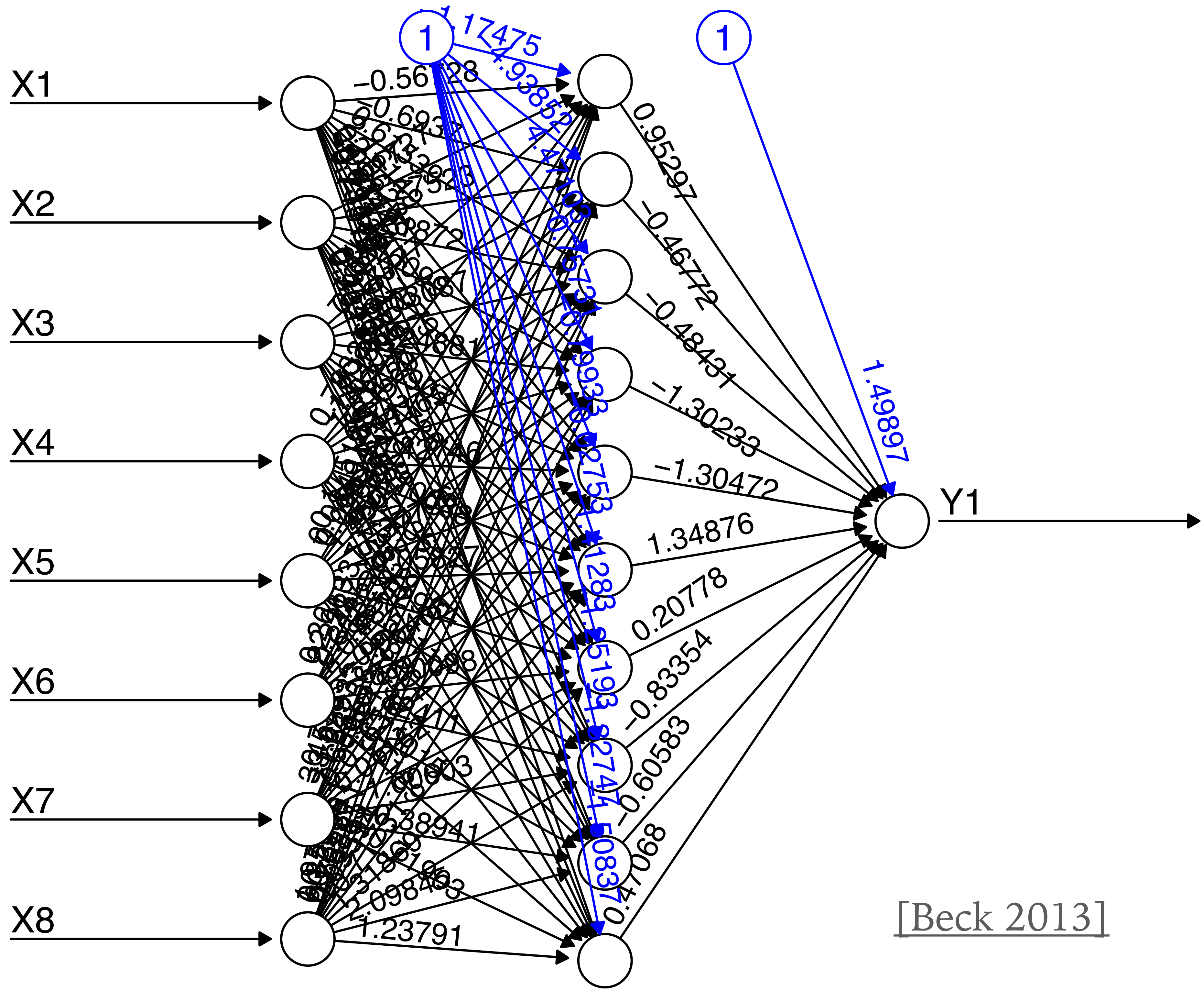




[Krizhevsky et al. 2012] and [Zeiler et. al 2014]

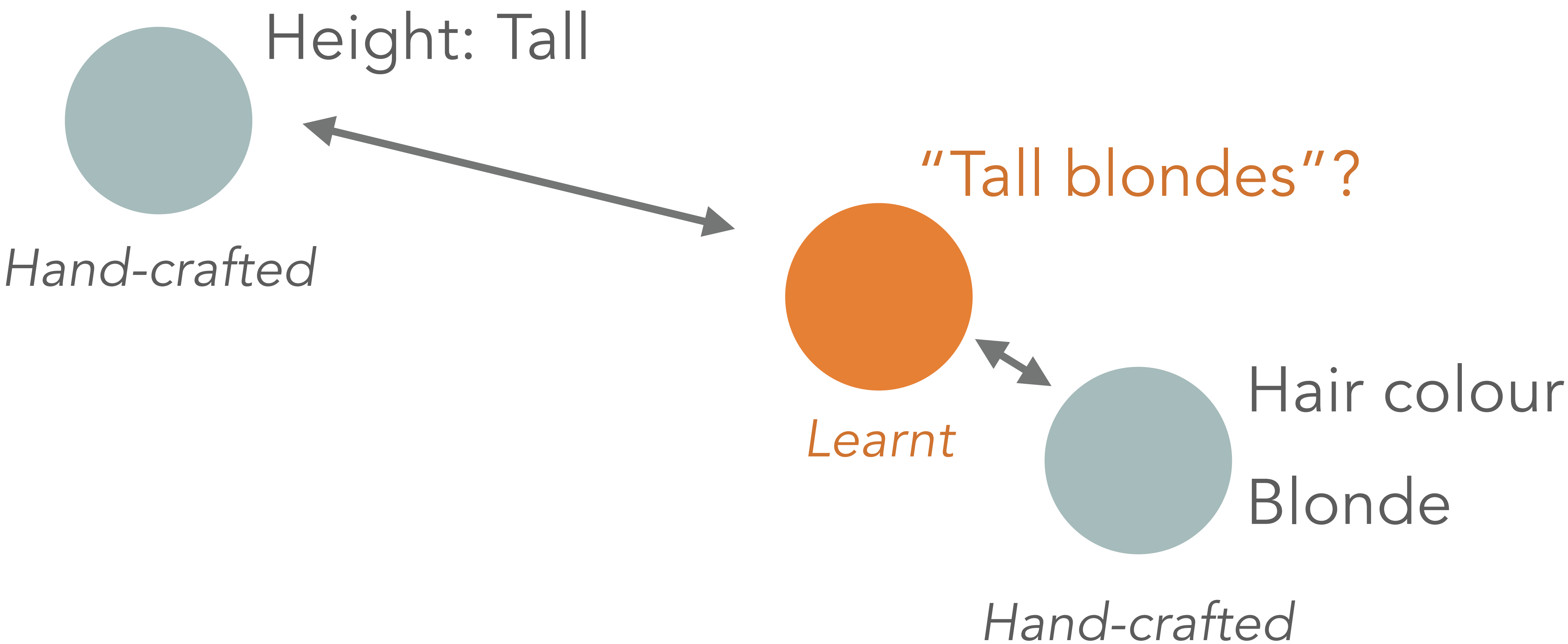
… but this only works because of the visual structure of their models.

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.

My domain: **Behavioural Game Theory**

understanding human behaviour in strategic situations

- 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…

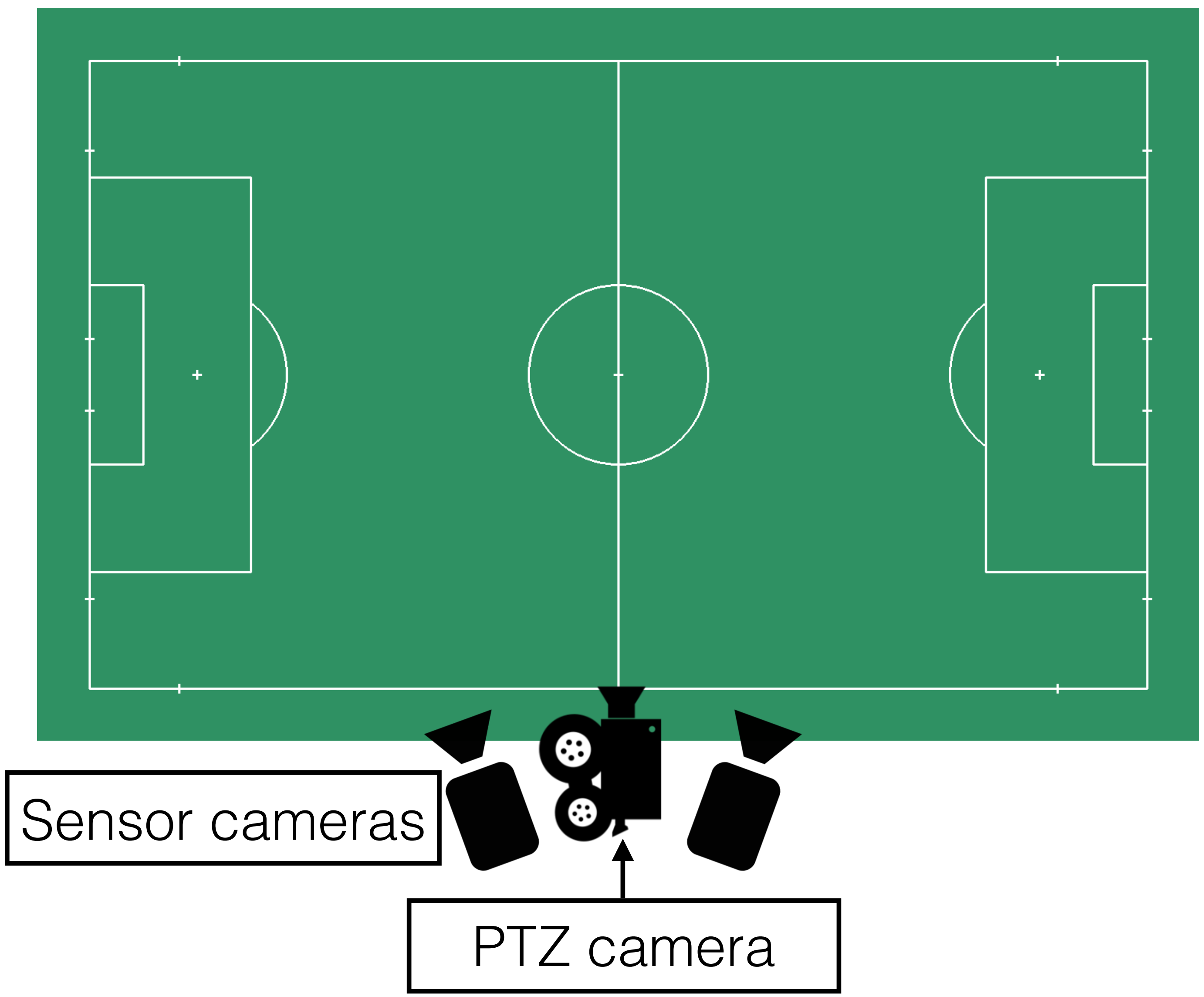
CameramanVis

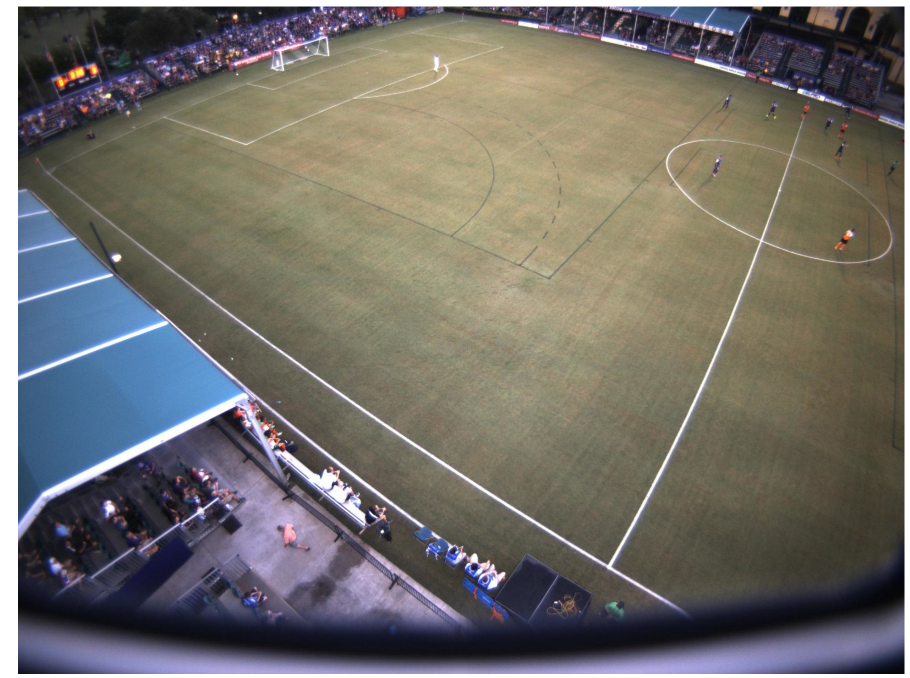
-Where camera should look?

Jianhui (Jimmy) Chen

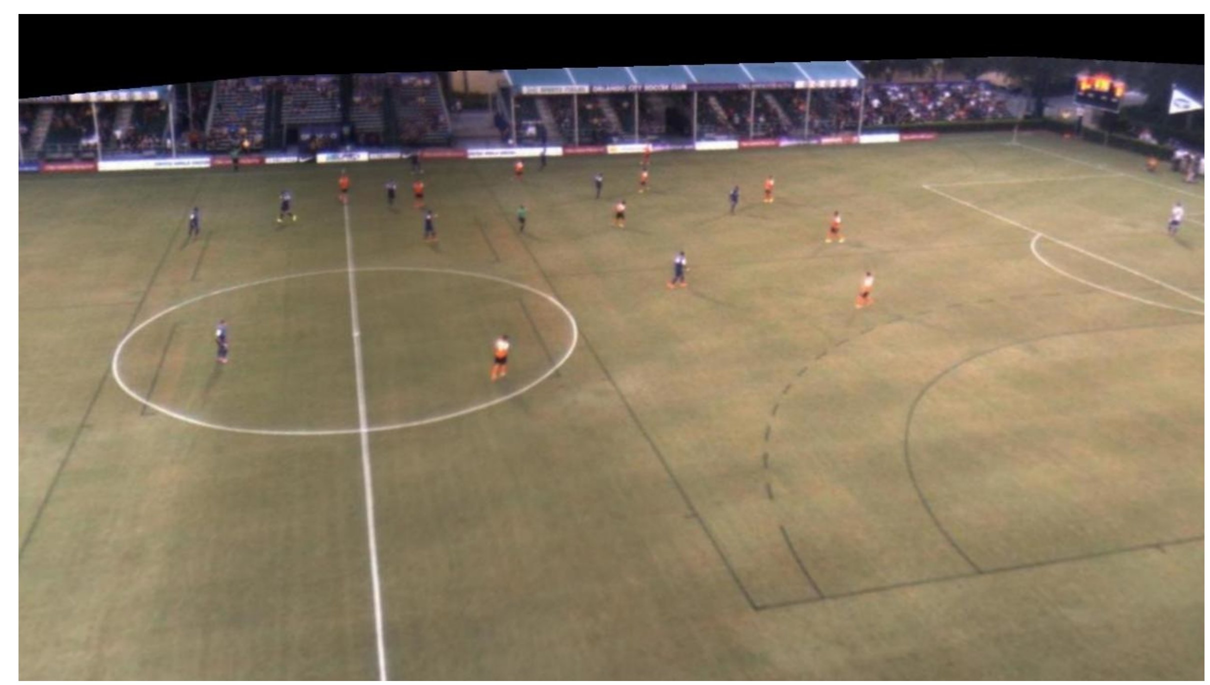
CPSC 547 Pitch

Cameras for soccer games





Sensor



PTZ

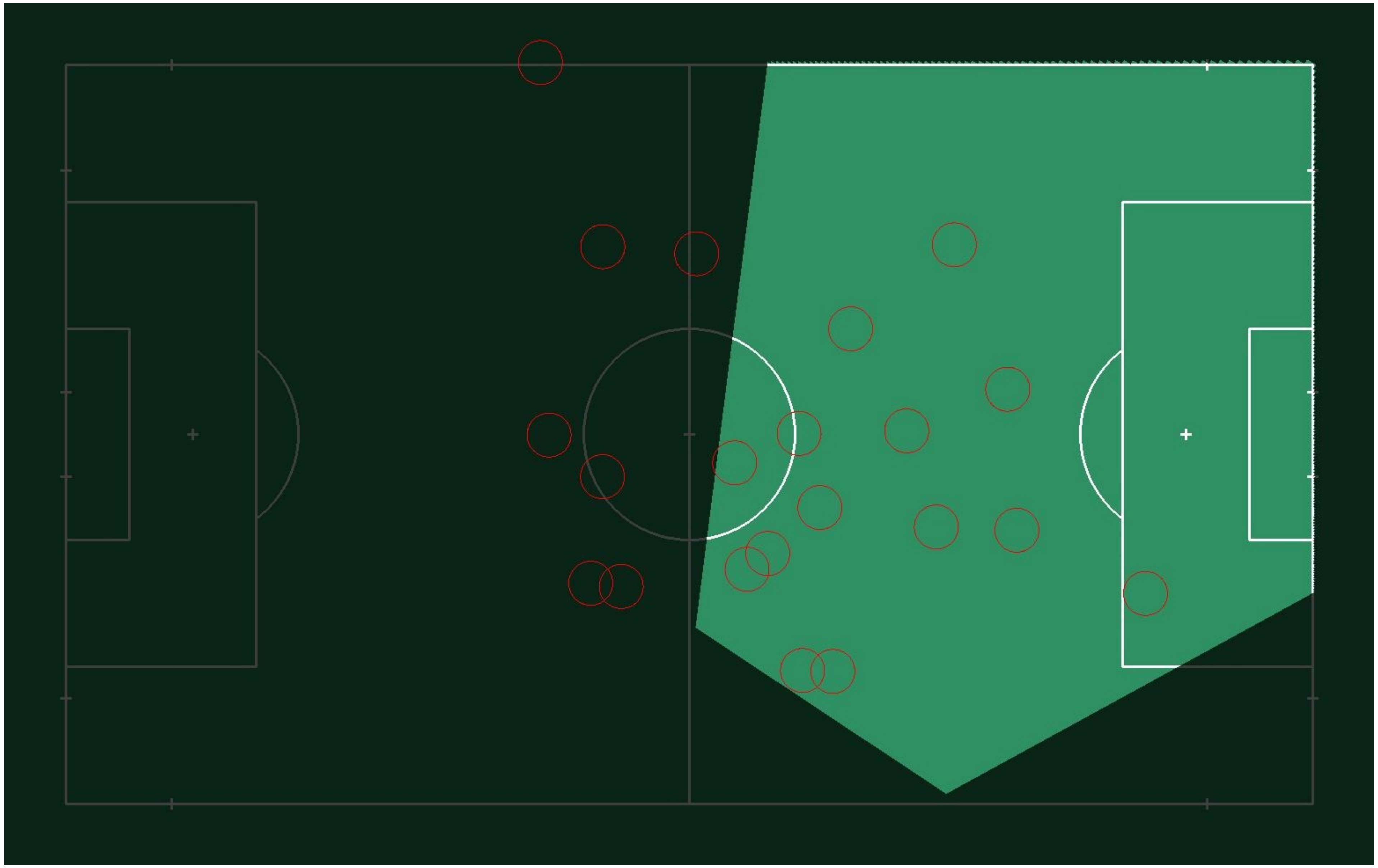
Left or right?

overview

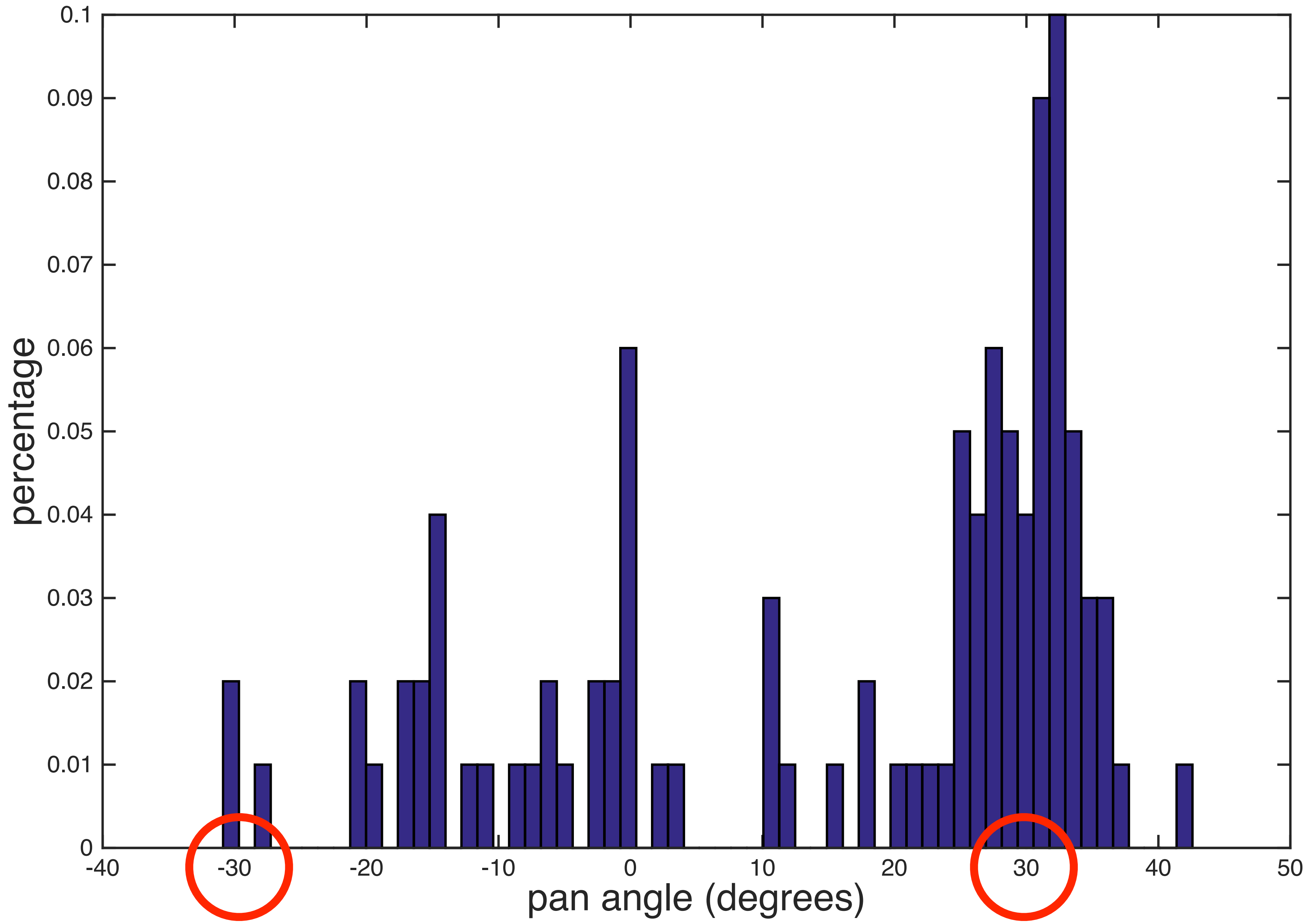
look left

look right

Goal:analyze cameraman’s view



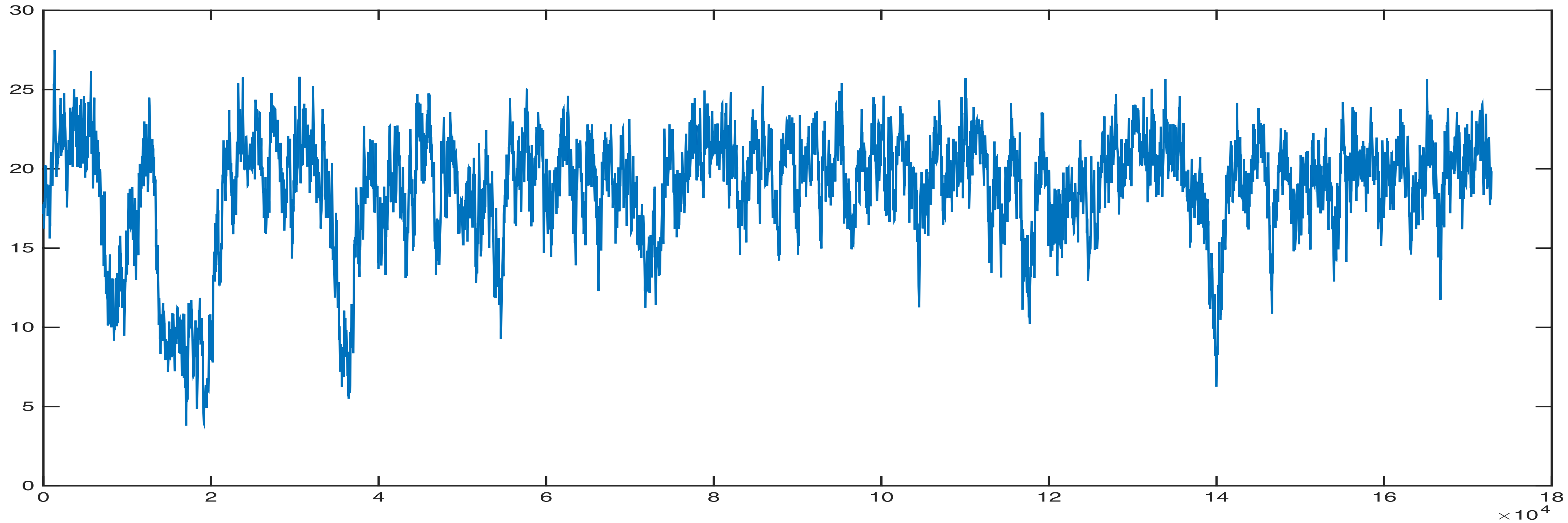
Not always the same angle



-30^o, 30^o are almost opposite.

Data

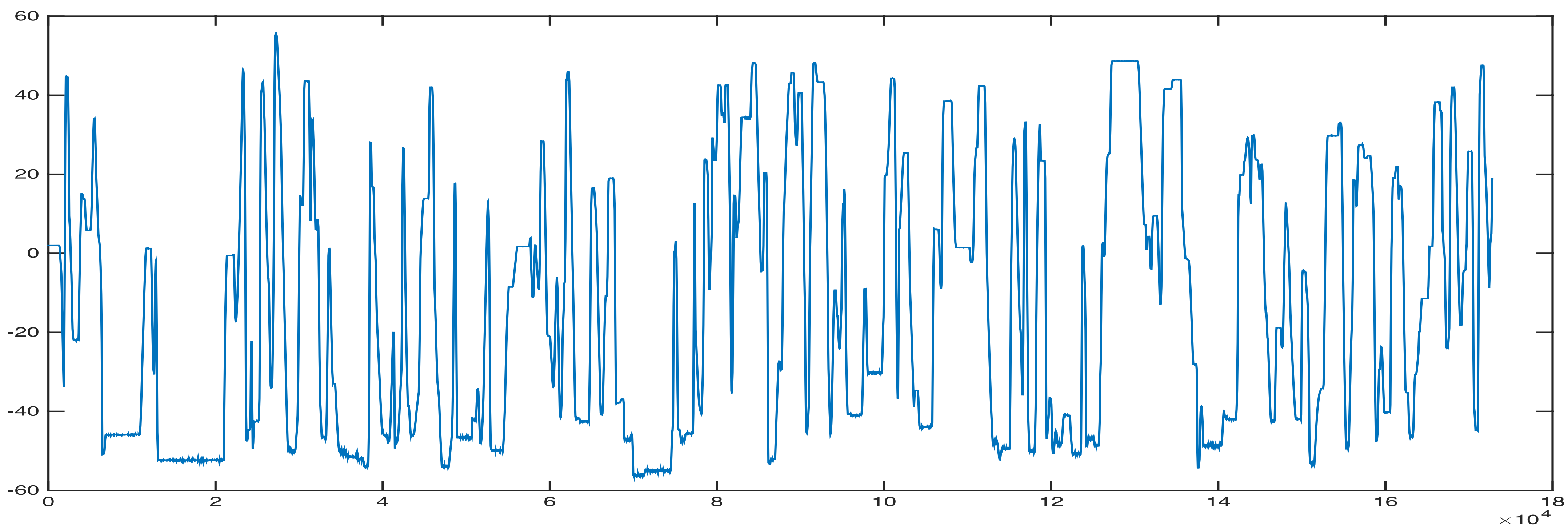
48-min (172, 800 frames) soccer game
Player positions on the playing ground
Camera angles of a PTZ camera
Both of them are noisy



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

Works and plays well with others

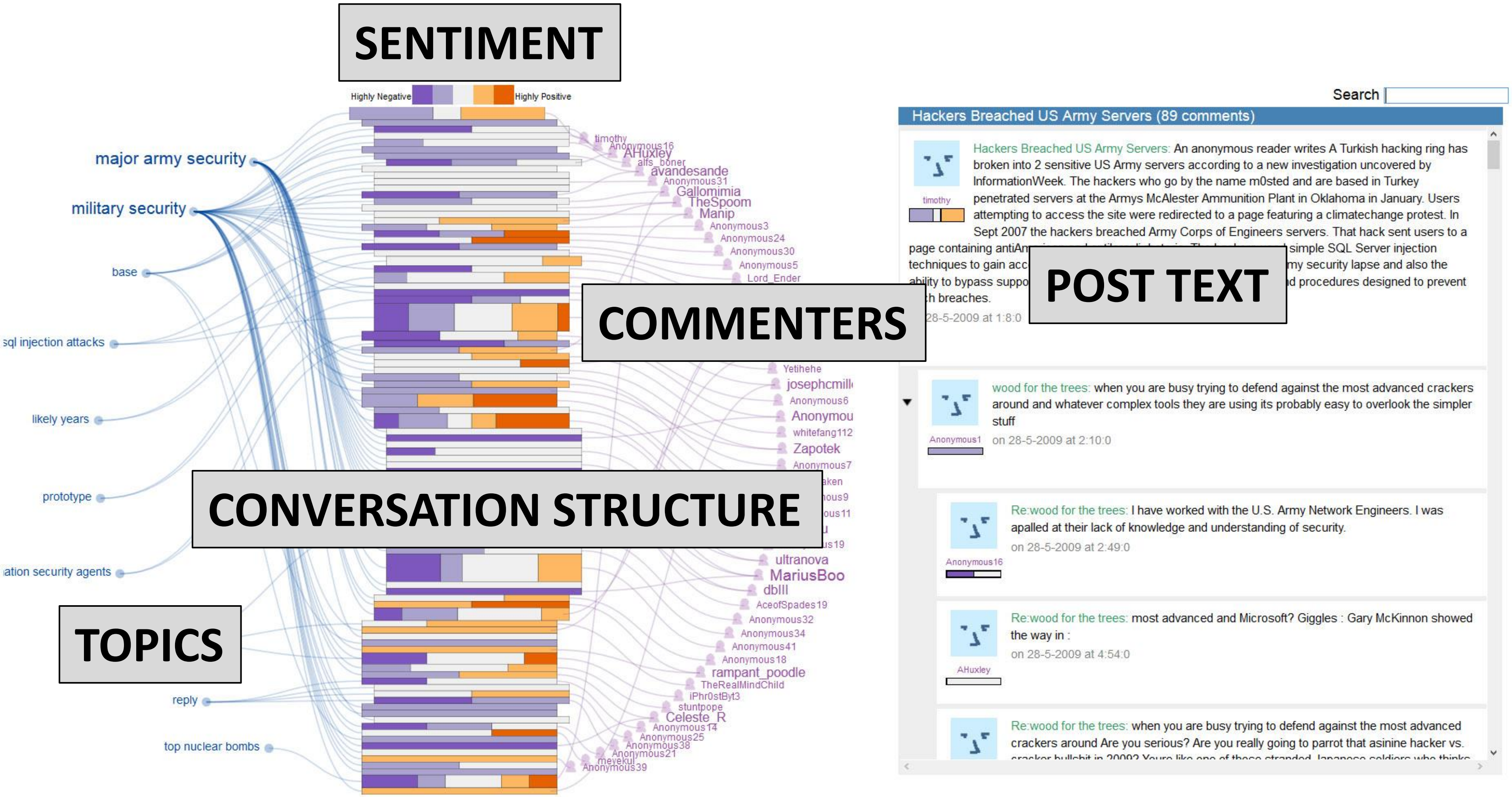
Jordon Johnson

M.Sc. Student

UBC Computer Science

jordon@cs.ubc.ca

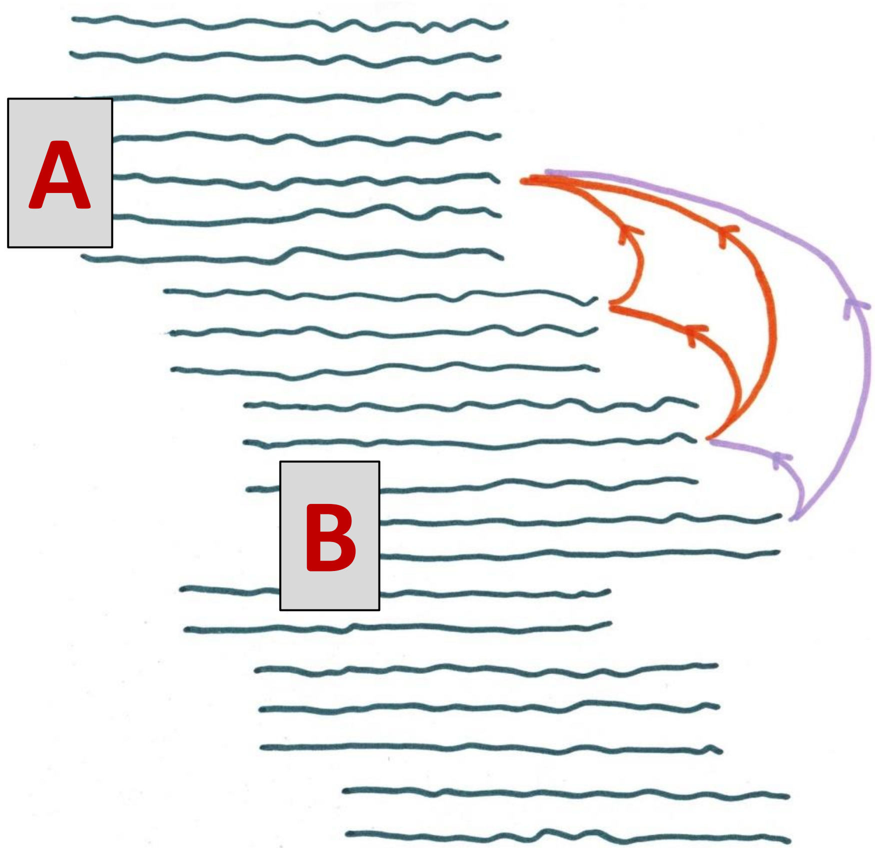
ConVis @ UBC



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

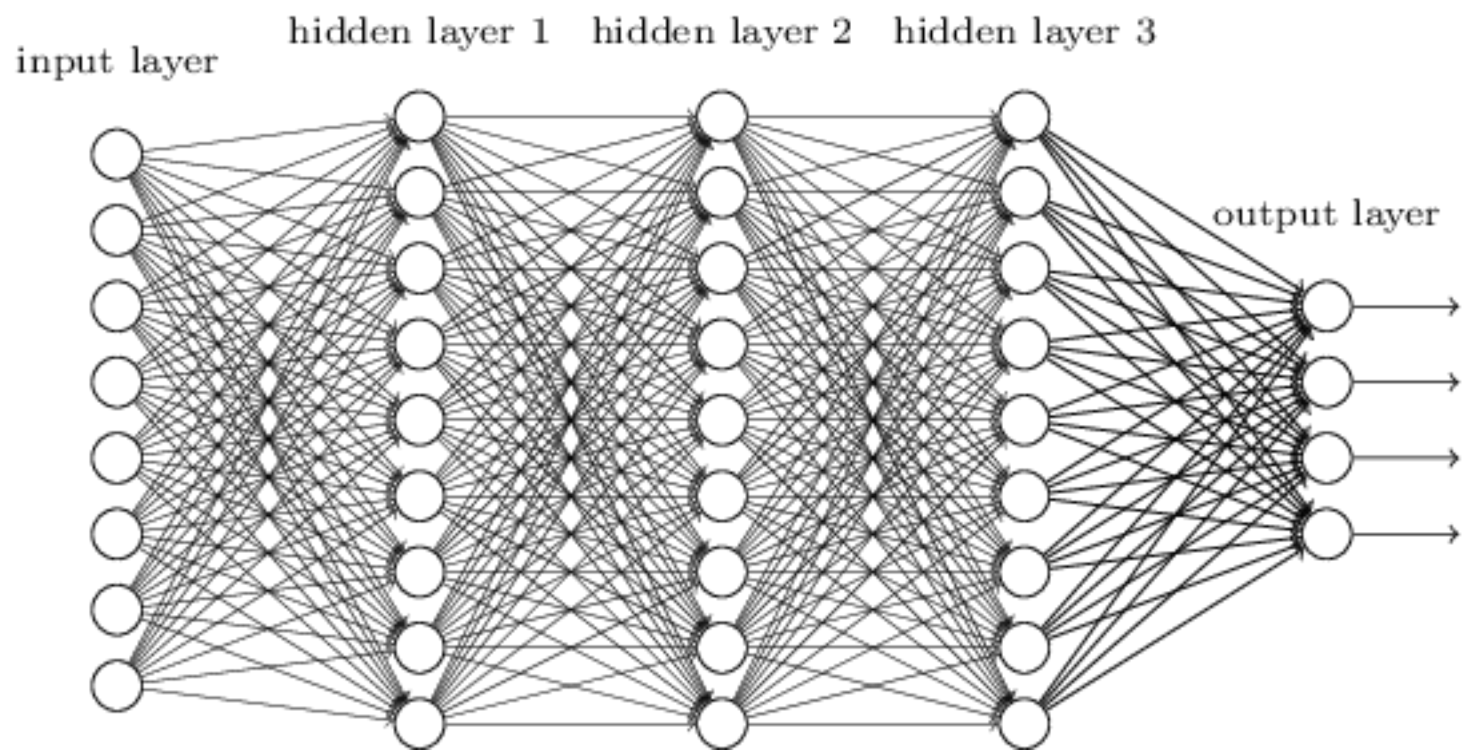
- Study link chains to associate link patterns with conversation types
  - Agreements
  - Debates
  - Off-topic/flamewars
  - Can we profile **trolling**?

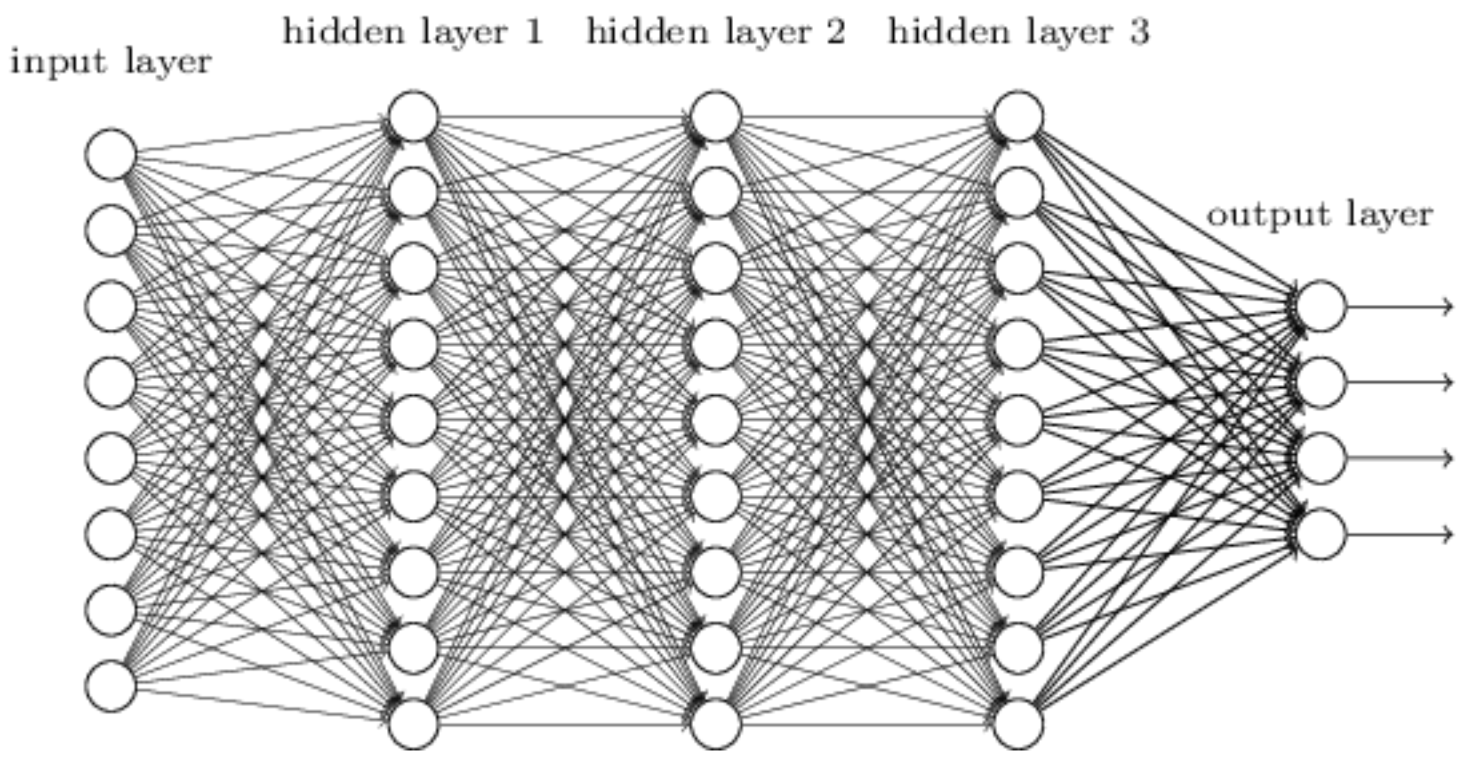
https://eclosure.com.au/wp-content/uploads/2015/04/Internet-Troll.jpg

Thank you!

vidviz: scaling up video annotation

Julieta Martinez


input layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer


input layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer

+ “Plumeria Frangipani”

The Internet

The Internet

The Internet

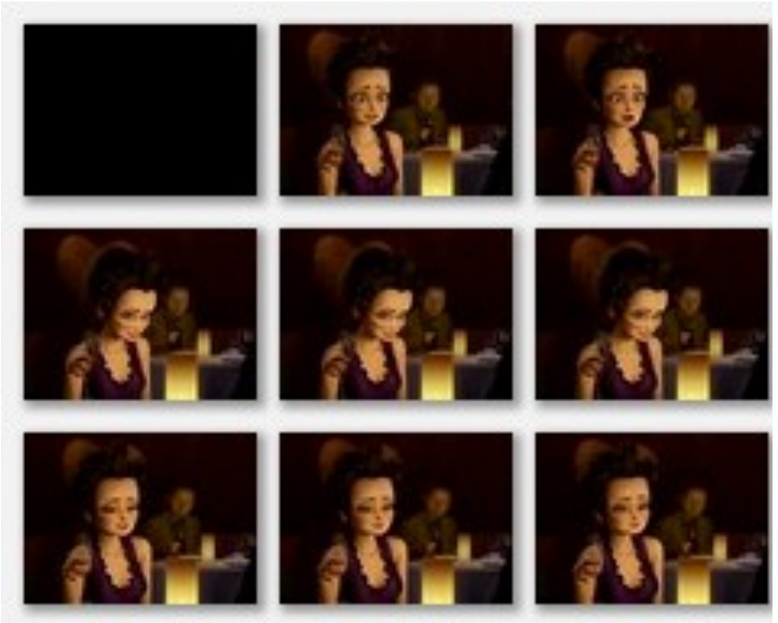
The Internet

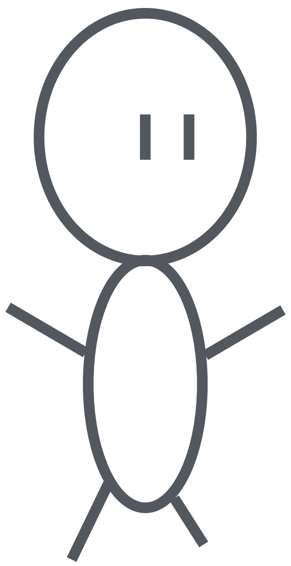
The Internet

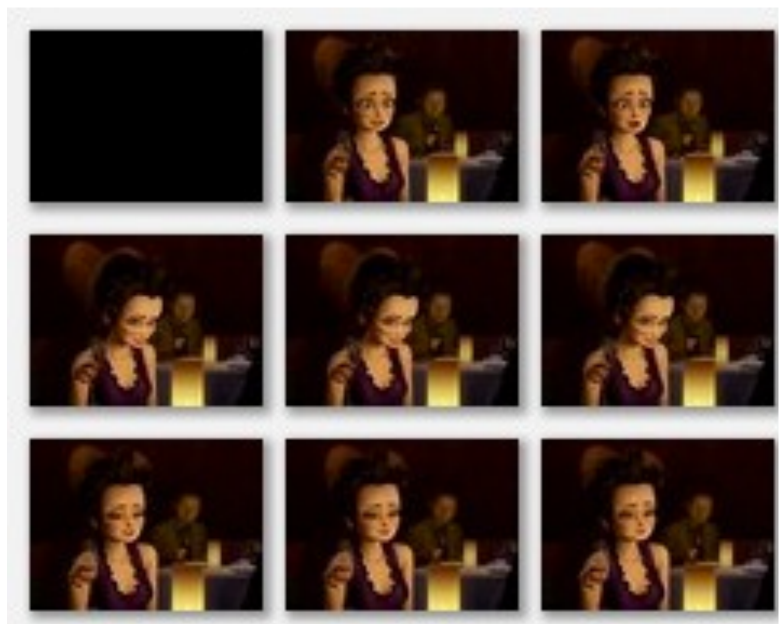
The Internet

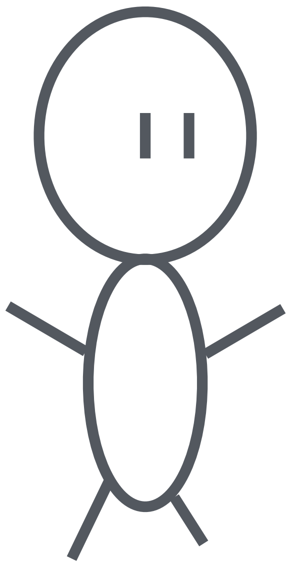
The Internet

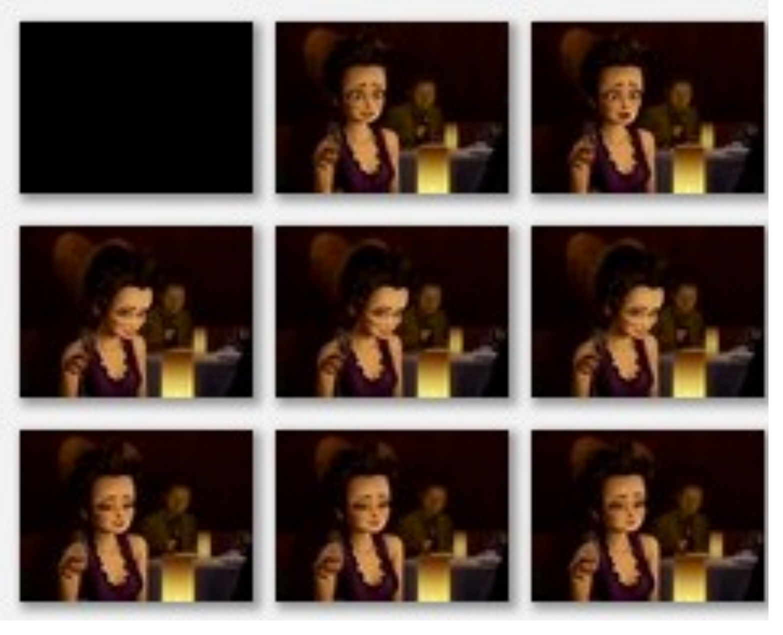
“the dark matter of the Internet” — Fei-Fei Li

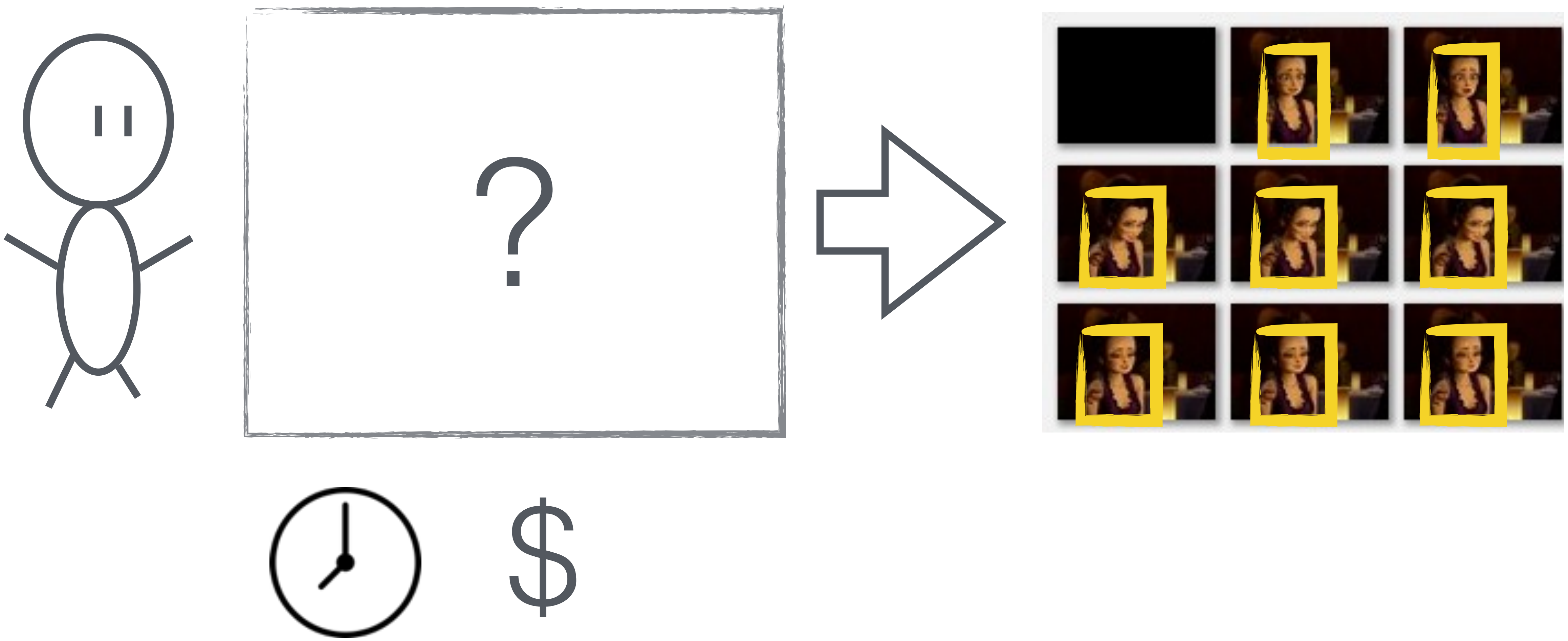


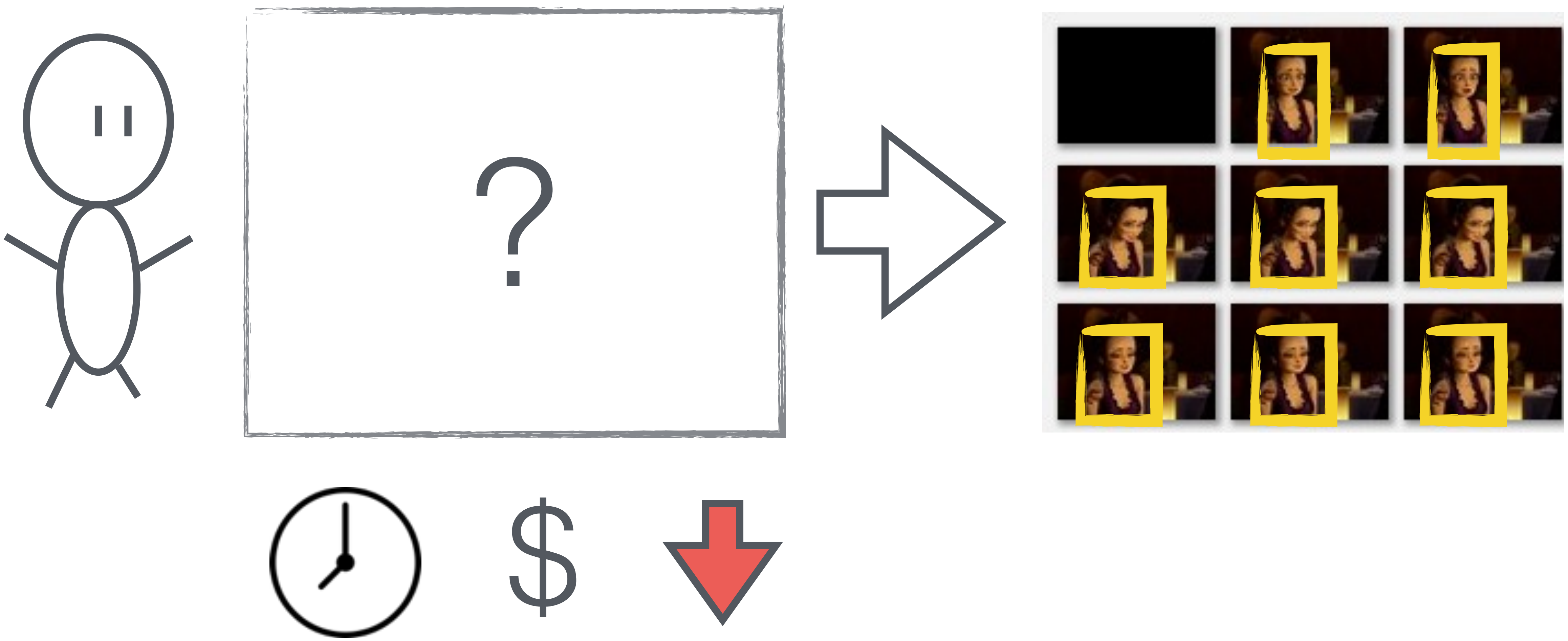


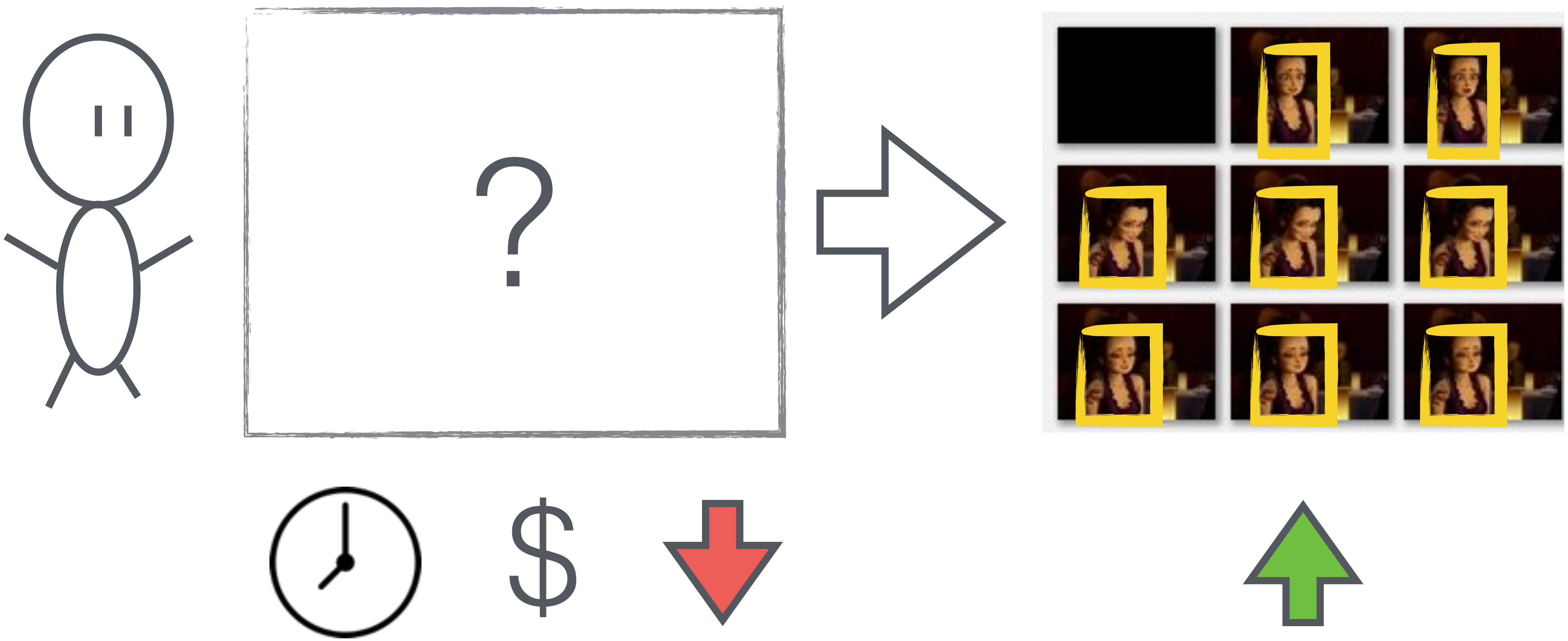













vidviz
$

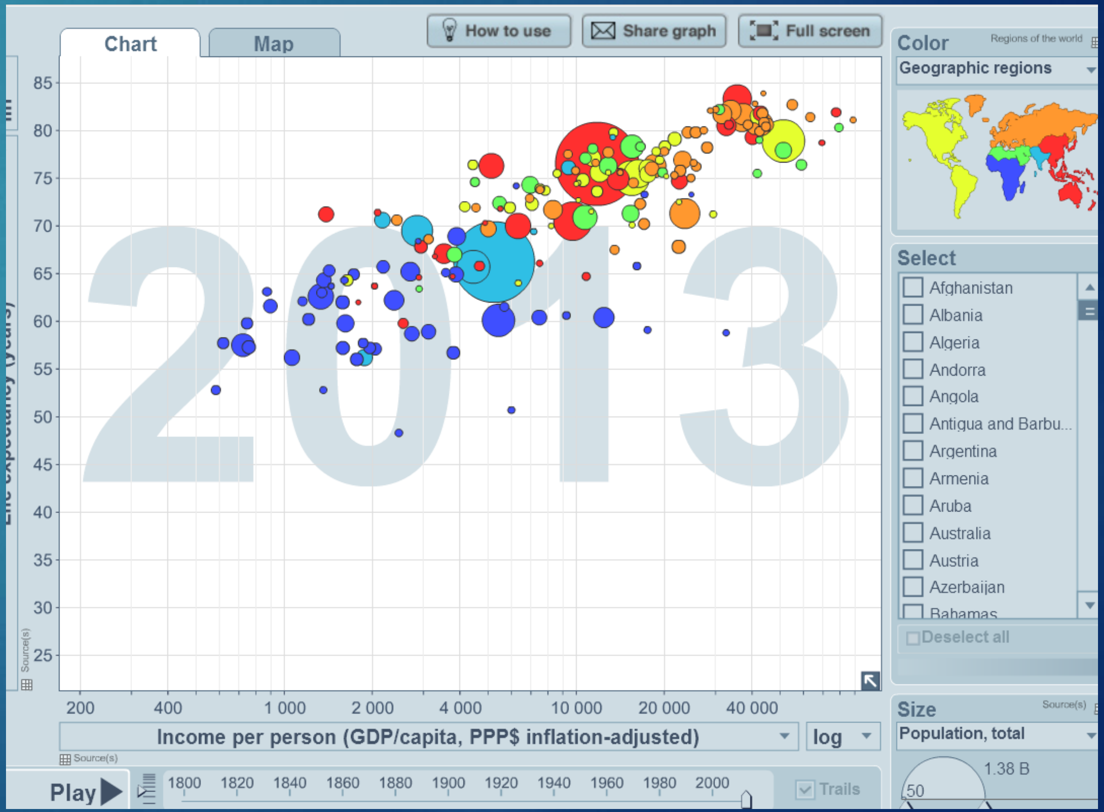
Interactive graphs using R Shiny

And maybe d3 if I have time.

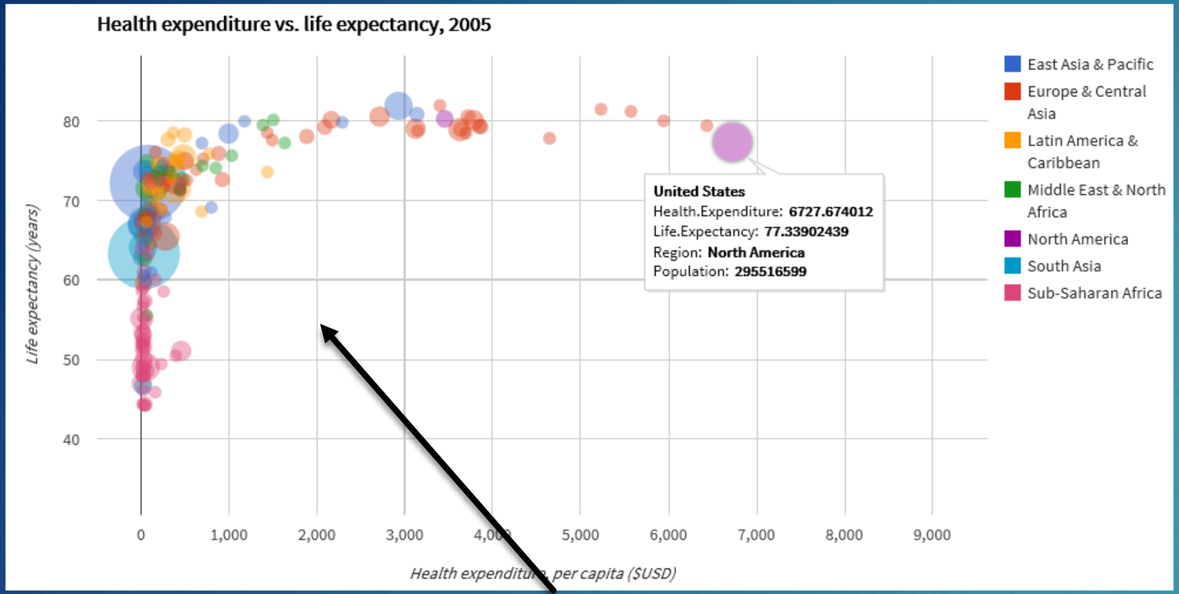
Ken Mansfield, CPSC 547, Oct 22, 2015

- 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.

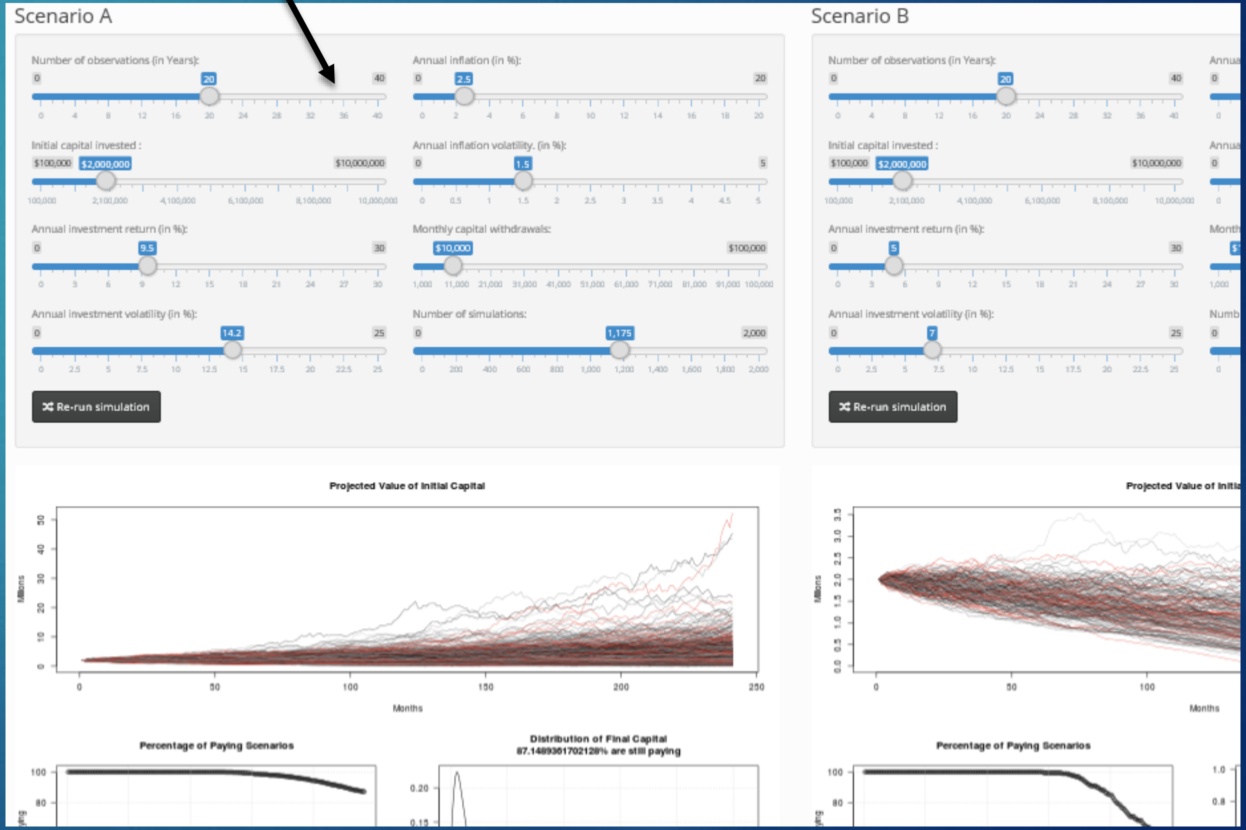
Example: Gapminder world



R Shiny Examples



Interactivity



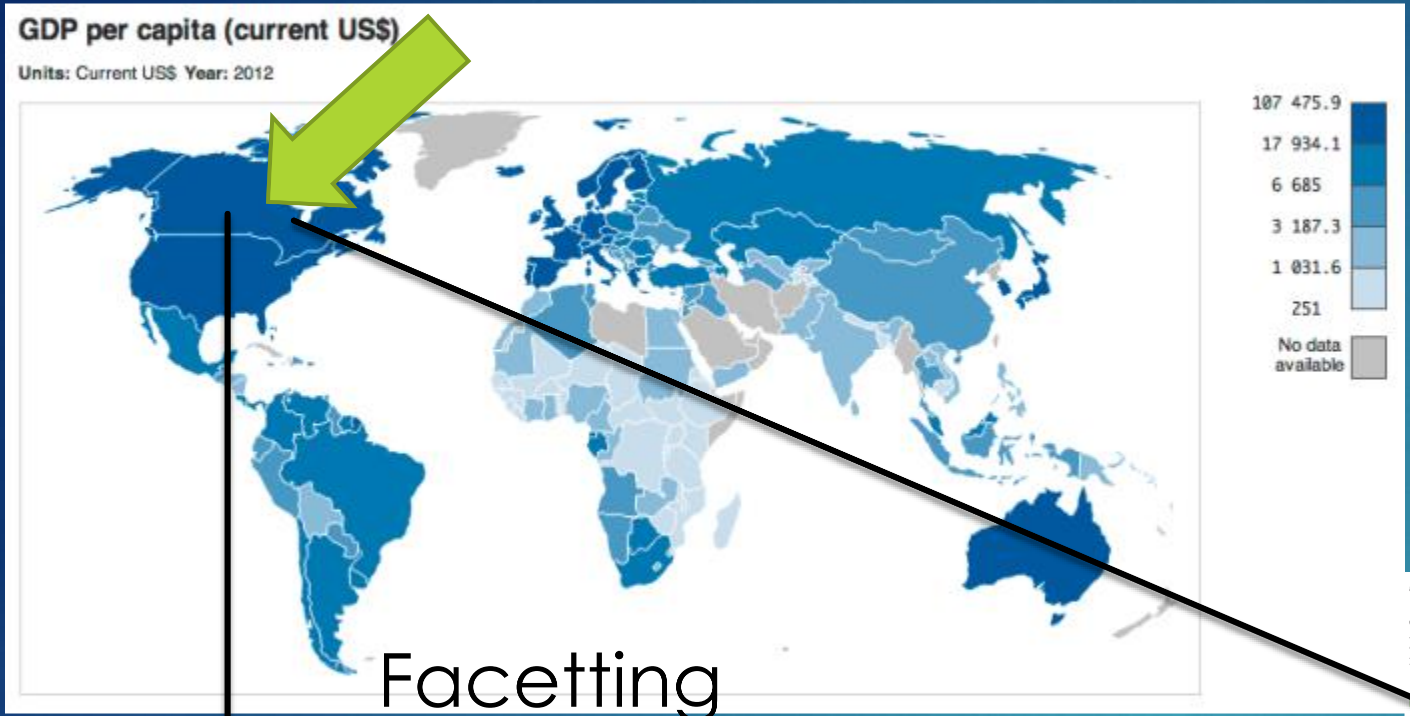
- Animation
- Colours
- Different types of Encoding

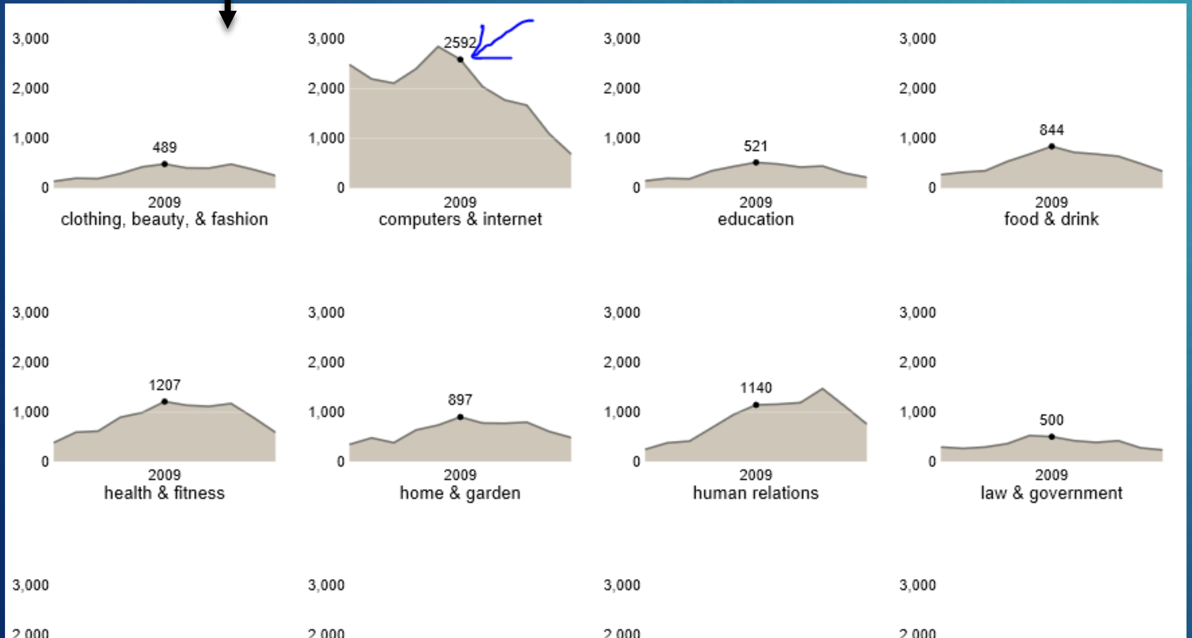
Expand on Gapminder

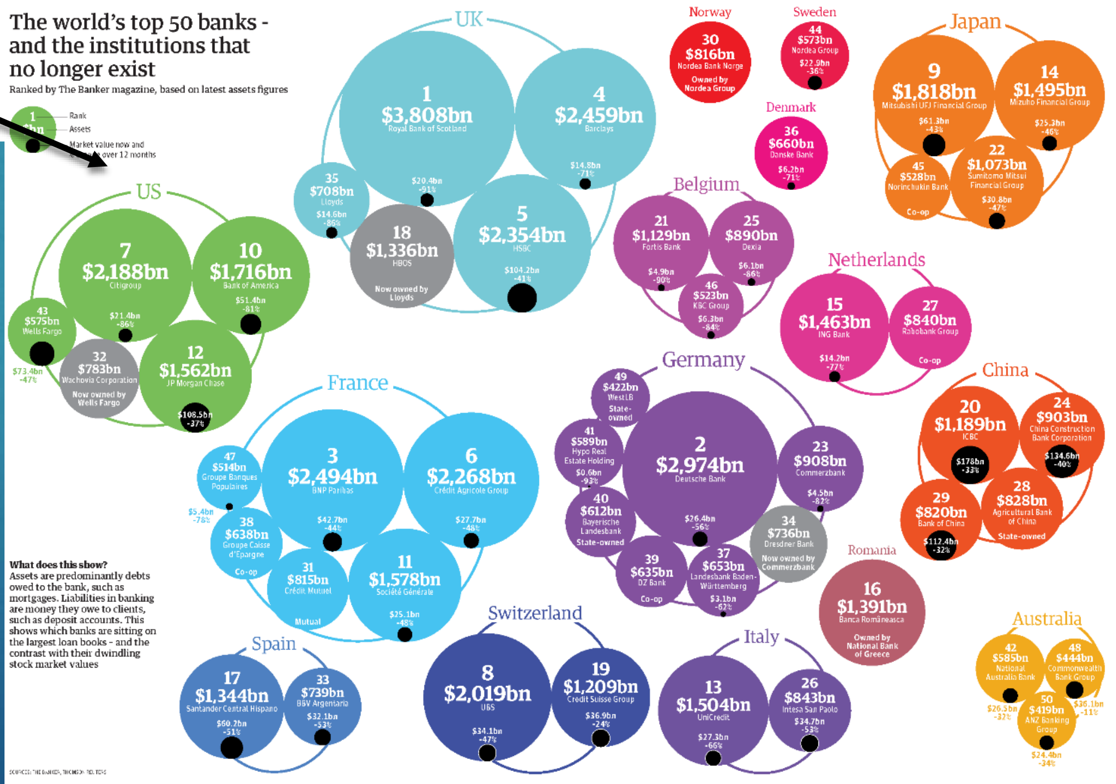
Choropleth

Linked highlighting

Facetting

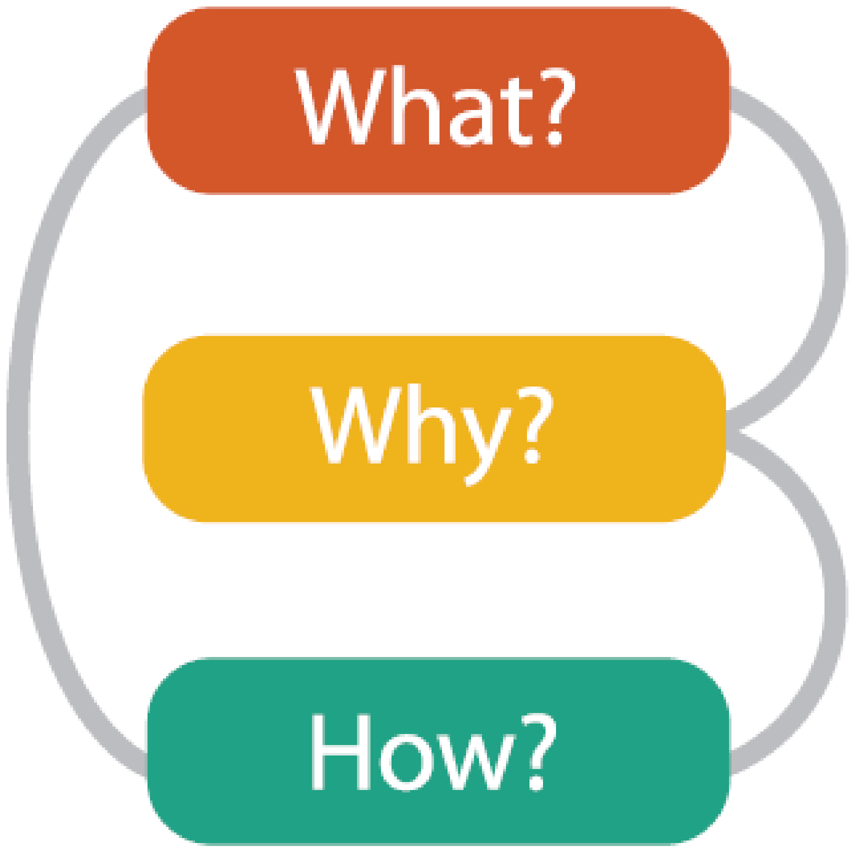






InfoVis Project Pitch

Kimberly Dextras-Romagnino

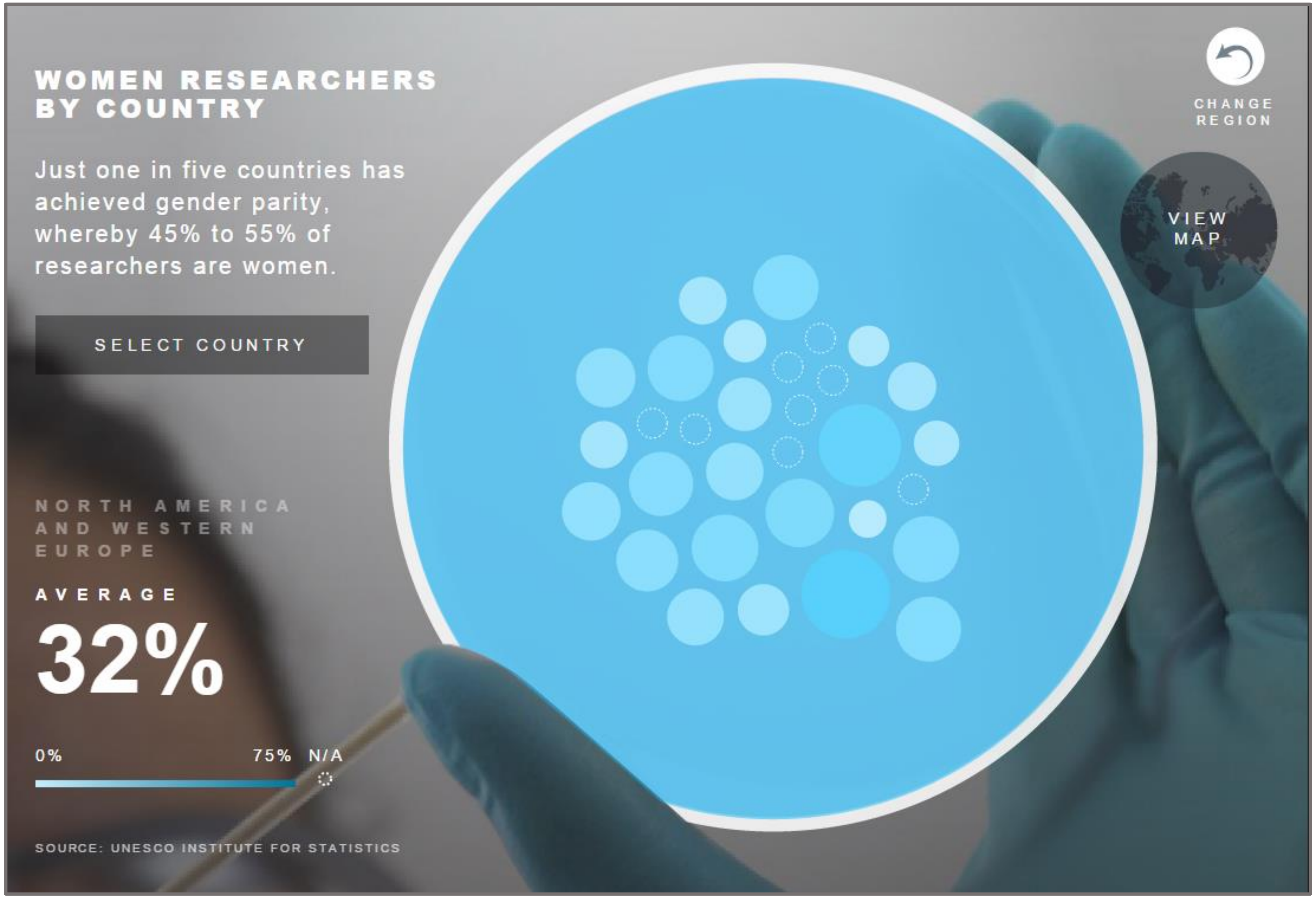

What?
Why?
How?

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

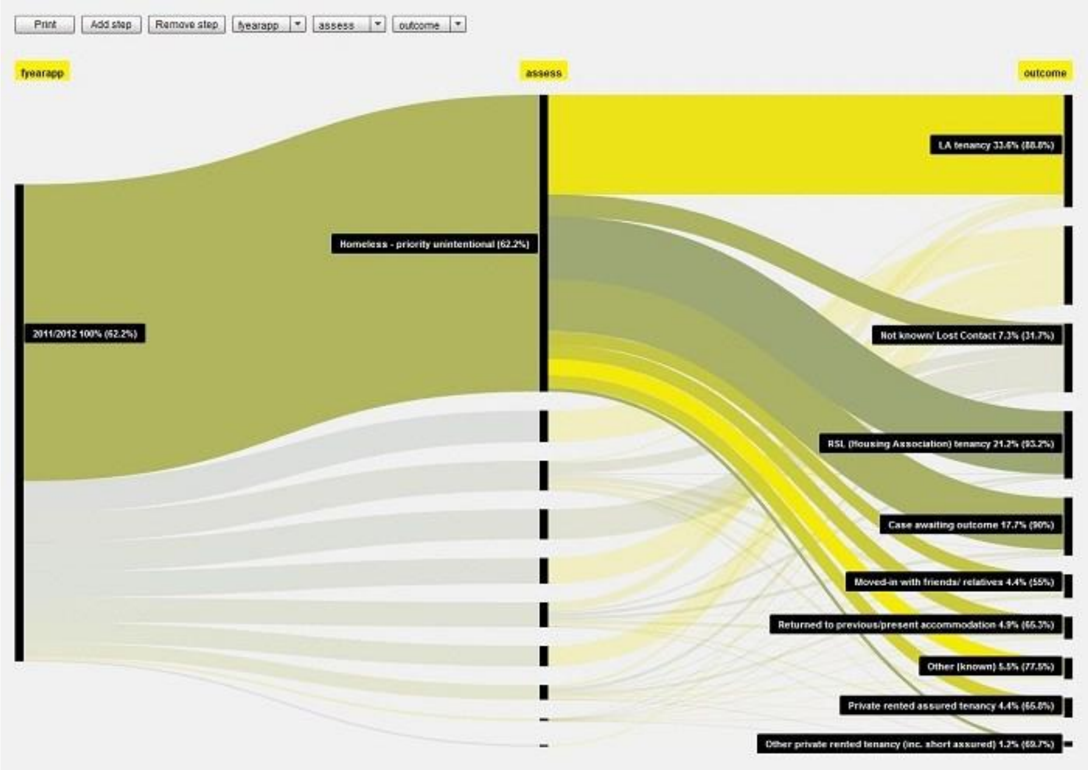

WOMEN RESEARCHERS
BY COUNTRY
Just one in five countries has achieved gender parity, whereby 45% to 55% of researchers are women.
SELECT COUNTRY
NORTH AMERICA AND WESTERN EUROPE
AVERAGE
32%
0%
75%
N/A
SOURCE: UNESCO INSTITUTE FOR STATISTICS
CHANGE REGION
VIEW MAP

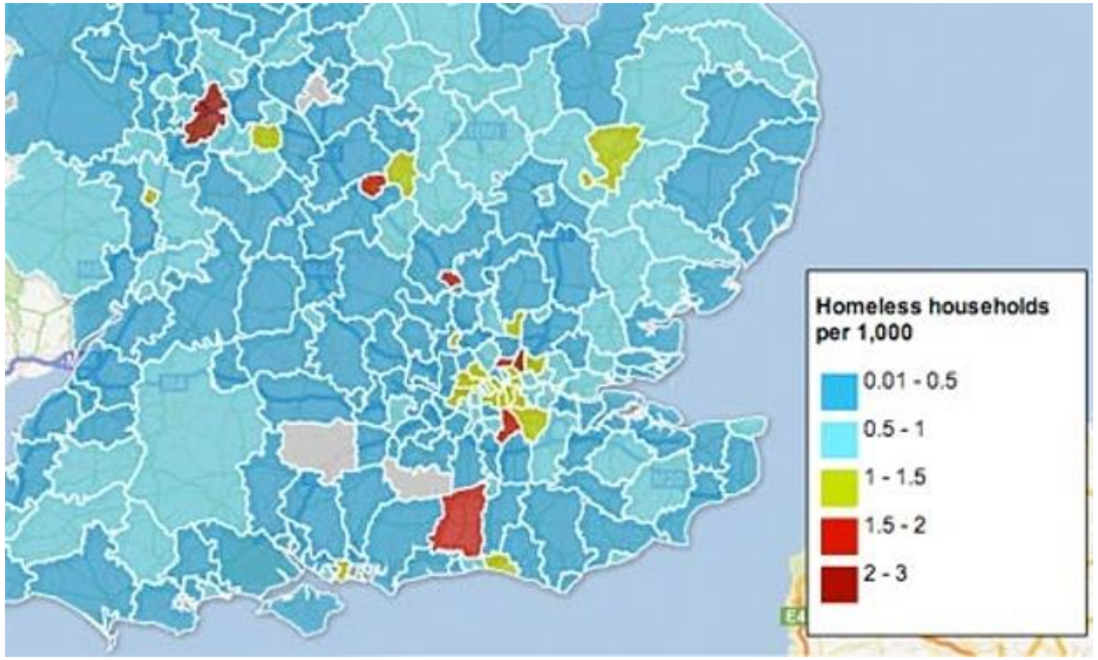
What?

Homelessness in Canada

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

Homeless Vis Examples

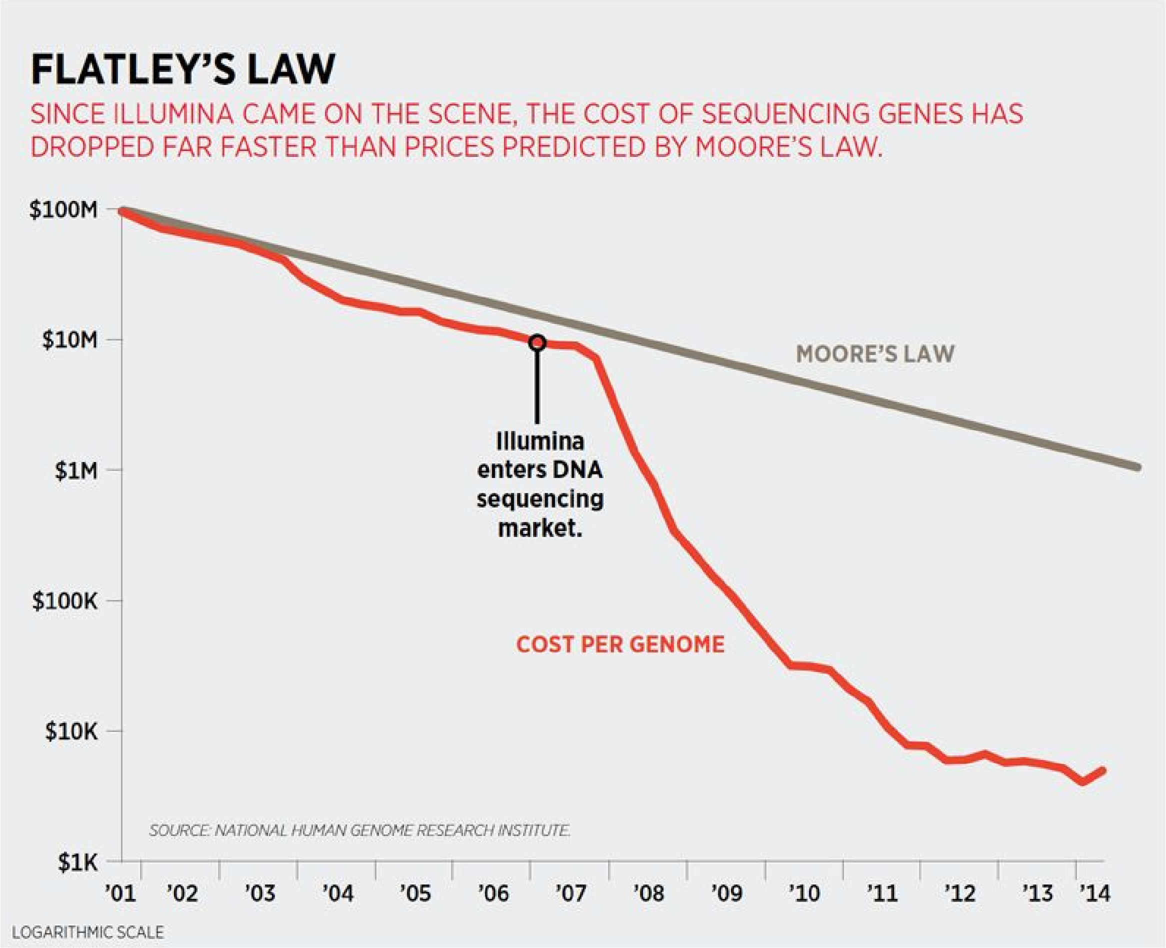




Any other ideas?

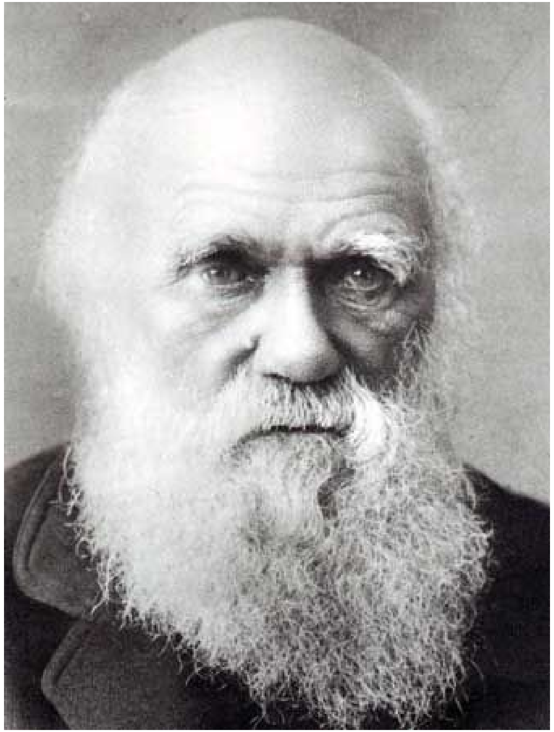
How Does Your Body Manage Its Army?

By: Louie Dinh



FLATLEY'S LAW
SINCE ILLUMINA CAME ON THE SCENE, THE COST OF SEQUENCING GENES HAS DROPPED FAR FASTER THAN PRICES PREDICTED BY MOORE'S LAW.
$100M
$10M
$1M
$100K
$10K
$1K
MOORE'S LAW
Illumina enters DNA sequencing market.
COST PER GENOME
SOURCE: NATIONAL HUMAN GENOME RESEARCH INSTITUTE.
'01
'02
'03
'04
'05
'06
'07
'08
'09
'10
'11
'12
'13
'14
LOGARITHMIC SCALE

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

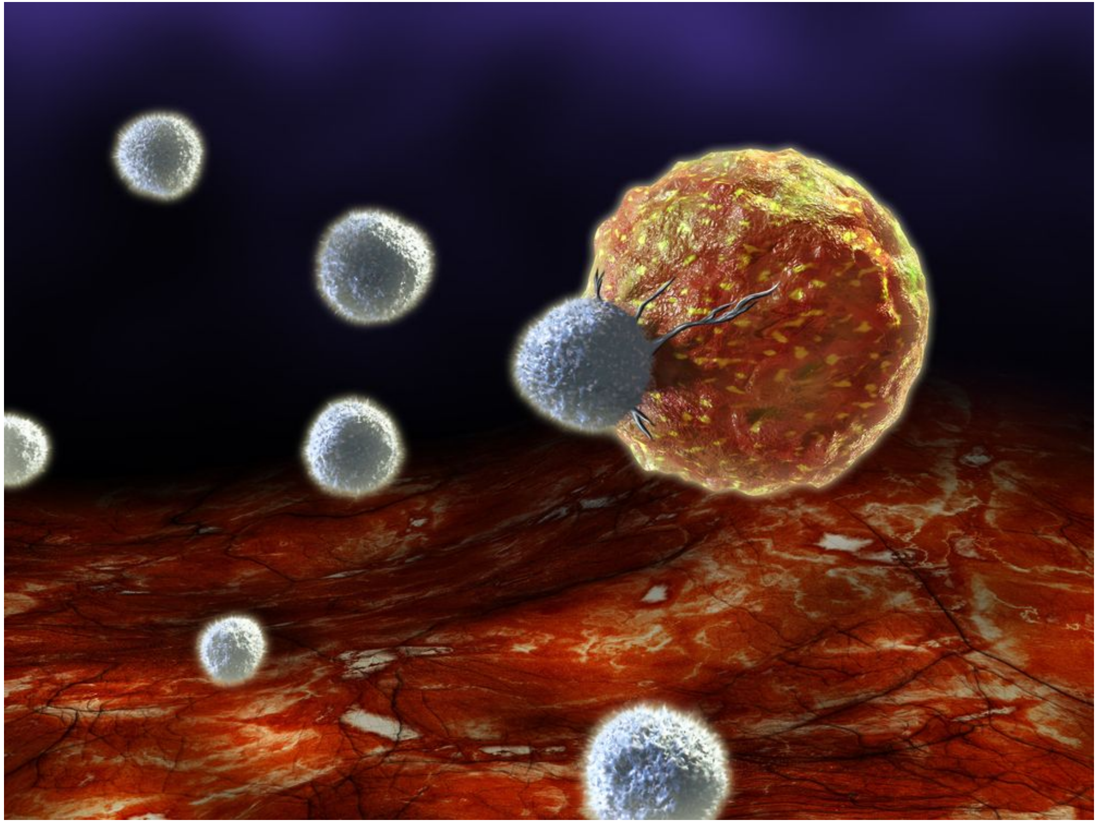


Darwin until now: understanding of biology has been very qualitative

So….

Let’s Quantify! (With Sequencing)

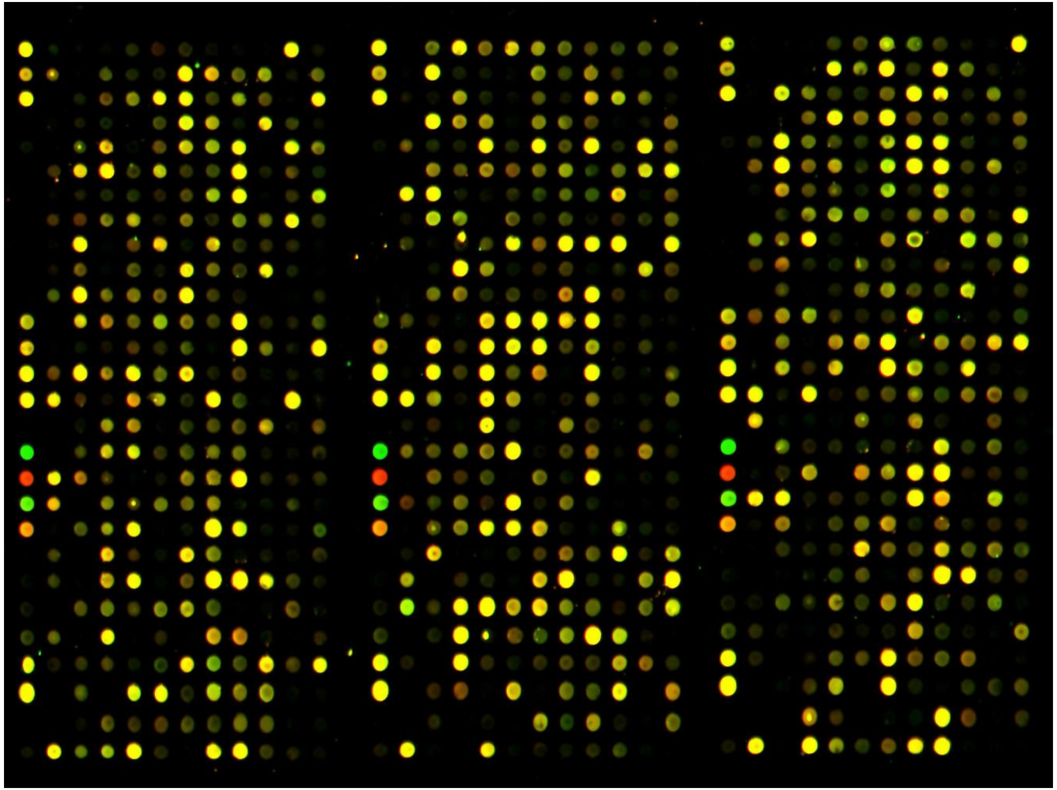
Interesting problem: How does your body manage its army (the immune system)



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!

You are just beginning your research on Cordacae, a subject you know nothing about. Where do you begin?

- ◯ Google
- ◯ Your course syllabus
- ◯ UBC Library's Summon Search

Next

Google	50	68.5%
Your course syllabus	5	6.8%
UBC Library's Summon Search	18	24.7%
	Total: 73	

[summon]

Response	Count	
Google	12	66.7%
Syllabus	6	33.3%
	Total: 18	

[syllabus1]

Response	Count	
The assignment requirements. Why else would I be researching something I know nothing about?	4	36.4%
I think we read something about		

[donereadings]

Response	Count	
Ok, I can suspend my disbelief far enough to pretend to use Google	1	50.0%
Can I use Bing instead?	1	50.0%
	Total: 2	

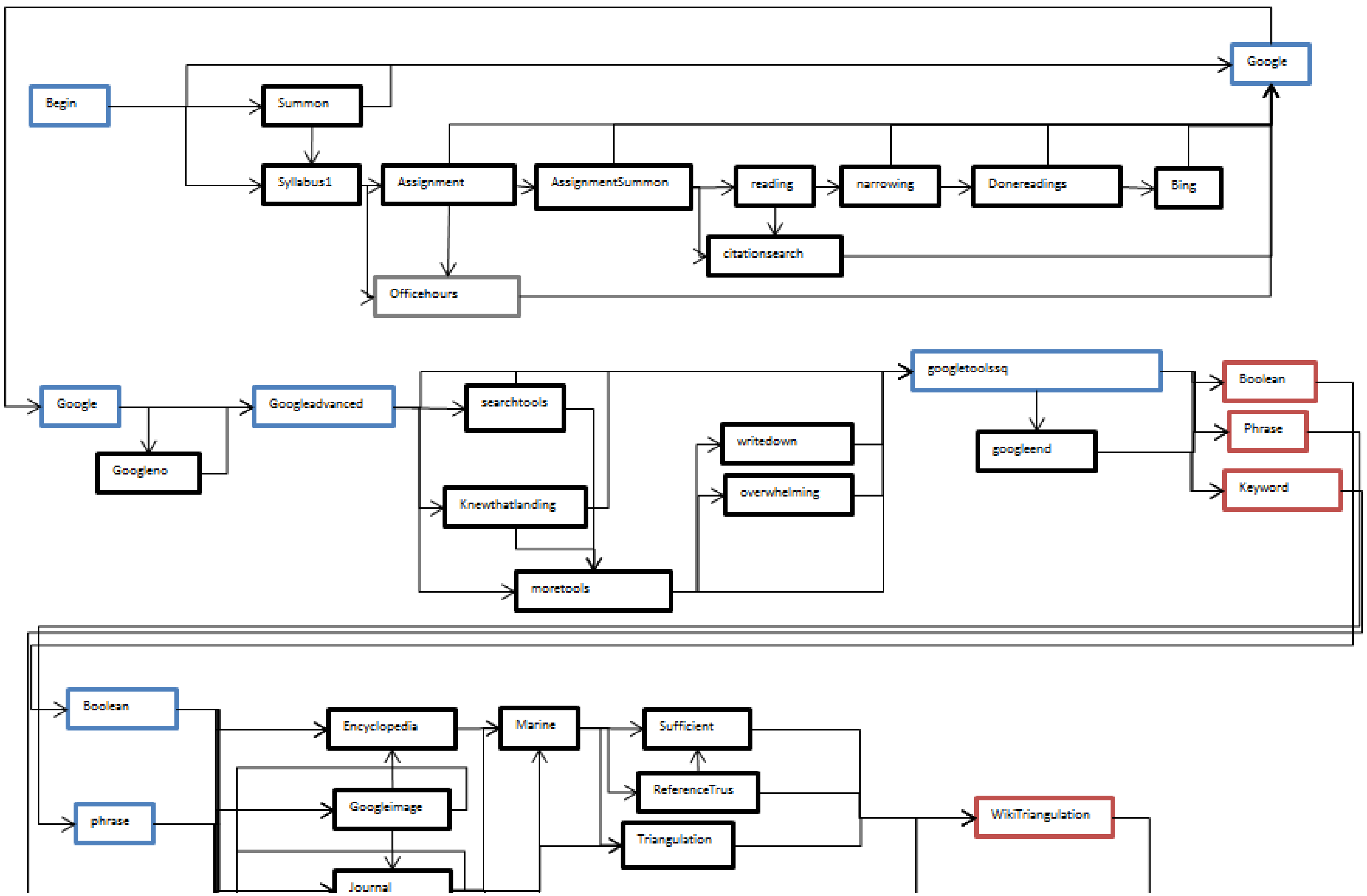
[Bing]

Response	Count	
I'm ready	1	100.0%
	Total: 1	

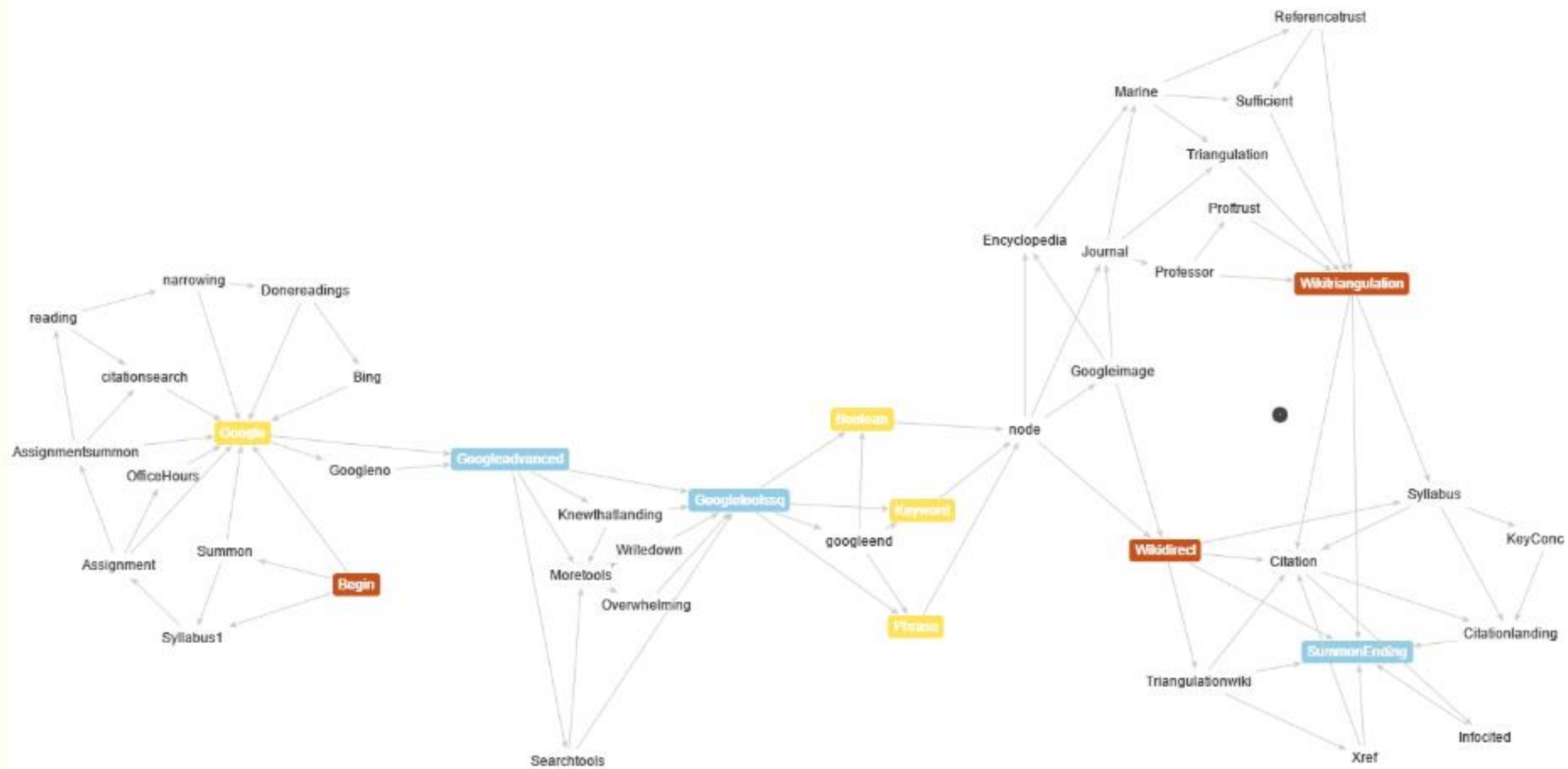
[OfficeHours]

No data available to display

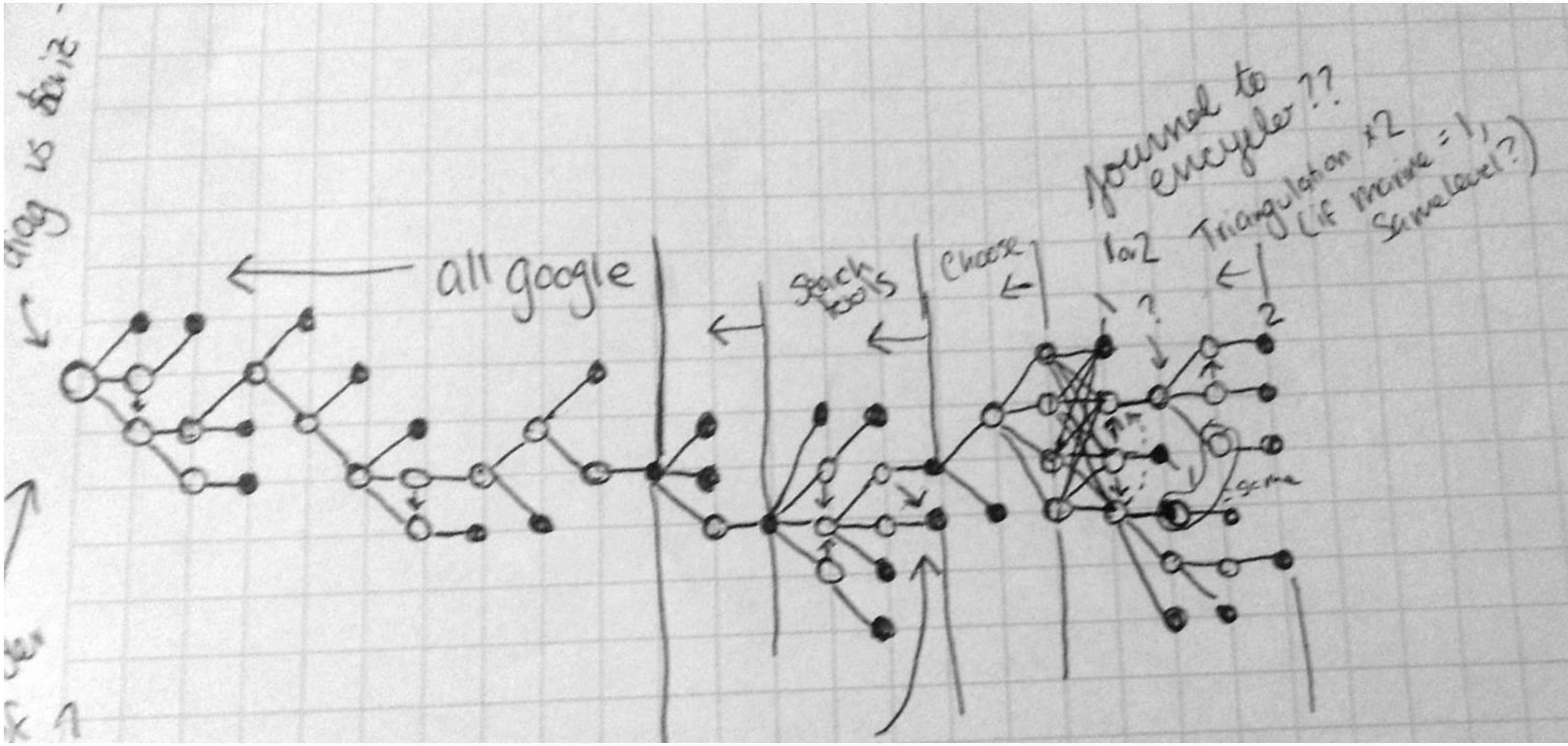
R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	
[begin]	[summon]	[syllabus1	[Assignme	[assignme	[Reading]	[citations	[narrowin	[donereac	[Bing]	[OfficeHo	[GoogleLa	[googlenc	[googlead	[searchto	[knewthat	[mo
Google											...Google has advanc		Are there any more search ope			Let
Google																
Google											No	Sure, I mi	Are there any more search ope			Let
Google																
UBC Librar	Syllabus	I think we read something abo			I want to highlight k		Honestly? I'd start Googling it.				Yes		...I knew all that		I'd love to	I di
Google																
Google											Yes		Are there any more search ope			I di
Google											...Google has advanc		Are there any more s		I'd love to	Let
Google																
UBC Librar	Google															
UBC Librar	Syllabus	I think we read something abo			I want to highlight k		Honestly? I'd start Googling it.									
UBC Librar	Google										Yes		...I knew all that		I'd rather skip	
Google																
Google											Yes		Where are	Can I still see the ext		Ok,
Your course syllabus		I think we read something abo			I want to highlight k		Honestly? I'd start Googling it.				...Google has advanc		Where are	Can I still see the ext		I di
Your course syllabus		I think we read something abo			I want to highlight k		Even in th	Ok, I can suspend my disbelief			Yes		...I knew all that		I'd love to	Let
Google											...Google has advanc		Where are	Thanks!		Ok,
Google											Yes		Where are	Thanks!		I di
Google											No	Yes	...I knew all that		I'd love to	Ok,
Google											No	Sure, I mi	Are there any more search ope			Ok,
Google																
Google											Yes		...I knew all that		I'd love to	I di
Google																



Google
Begin
Summon
Syllabus1
Assignment
AssignmentSummon
reading
narrowing
Donereadings
Bing
citationsearch
Officehours
googletoolssq
Boolean
Google
Googleadvanced
searchtools
Phrase
writedown
googleend
Googleno
overwhelming
Keyword
Knewthatlanding
moretools
Boolean
Encyclopedia
Marine
Sufficient
ReferenceTrus
Googleimage
WikiTriangulation
phrase
Triangulation
Journal

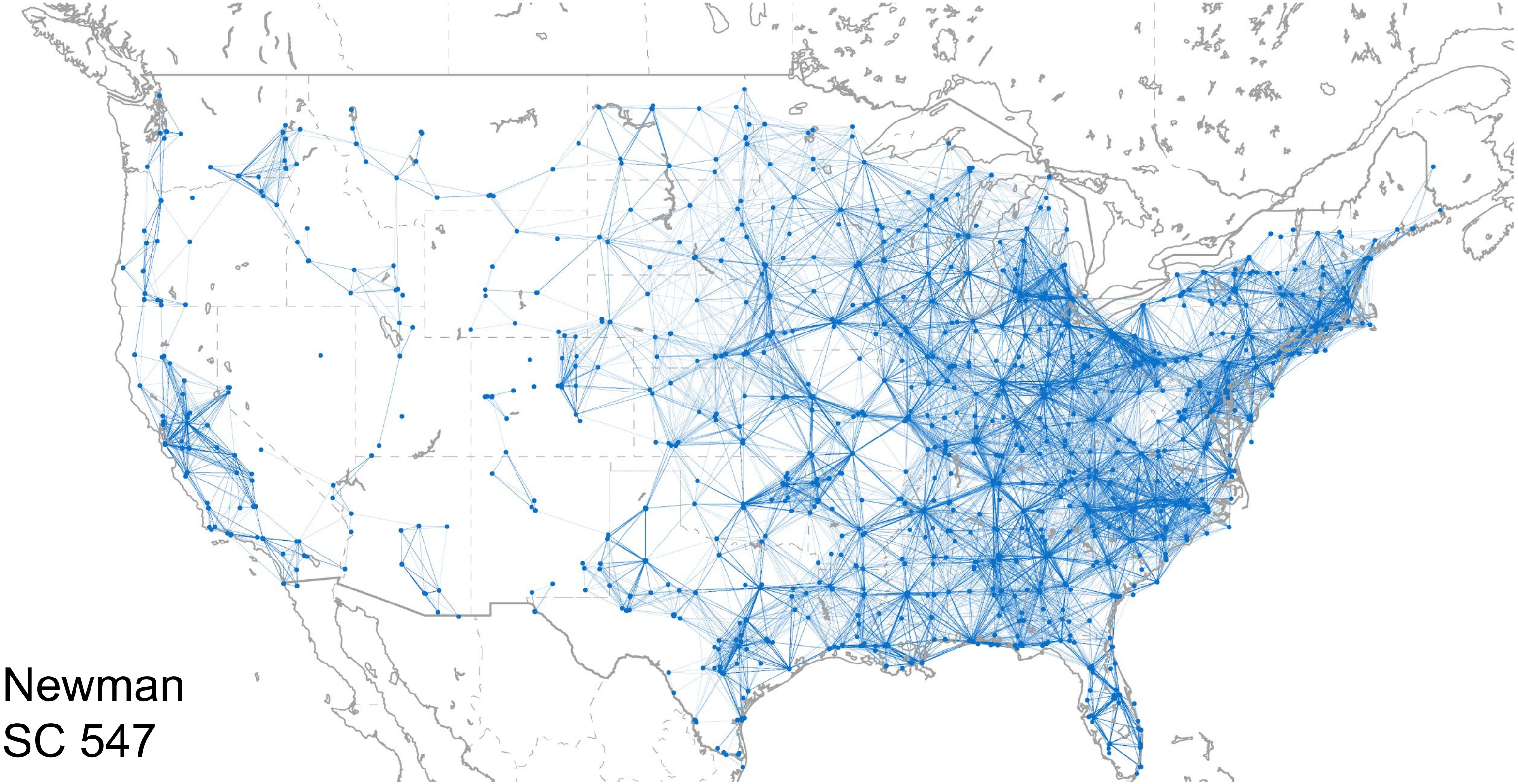


Referencetrust
Marine
Sufficient
Triangulation
Proftrust
Encyclopedia
Journal
Professor
Wikitriangulation
narrowing
Donereadings
reading
citationsearch
Bing
Googleimage
Google
node
Assignmentsummon
Googleno
Googleadvanced
OfficeHours
Keyword
Syllabus
Knewthatlanding
googleend
KeyConc
Wikidirect
Citation
Writedown
Summon
Assignment
Begin
Moretools
Overwhelming
Phrase
Citationlanding
SummonEnding
Syllabus1
Triangulationwiki
Infocited
Xref
Searchtools



all google
Seach tools
Choose
journal to encycler??
1 or 2
Triangulation x2
(if marine = 1, same level?)
?
2

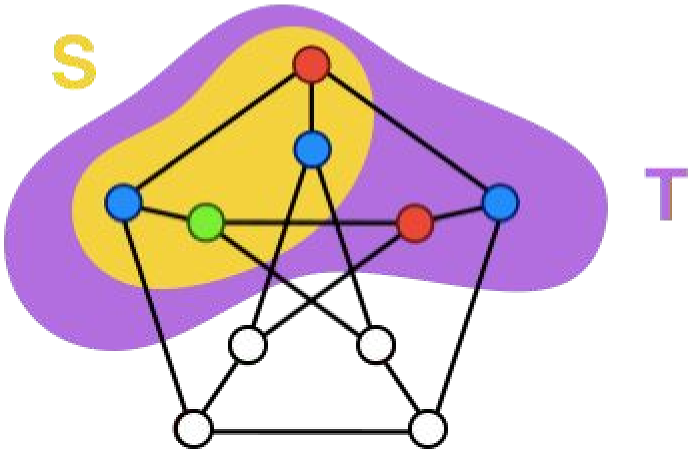
Spectrum Auction -> SAT Problems



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®

Loblaws

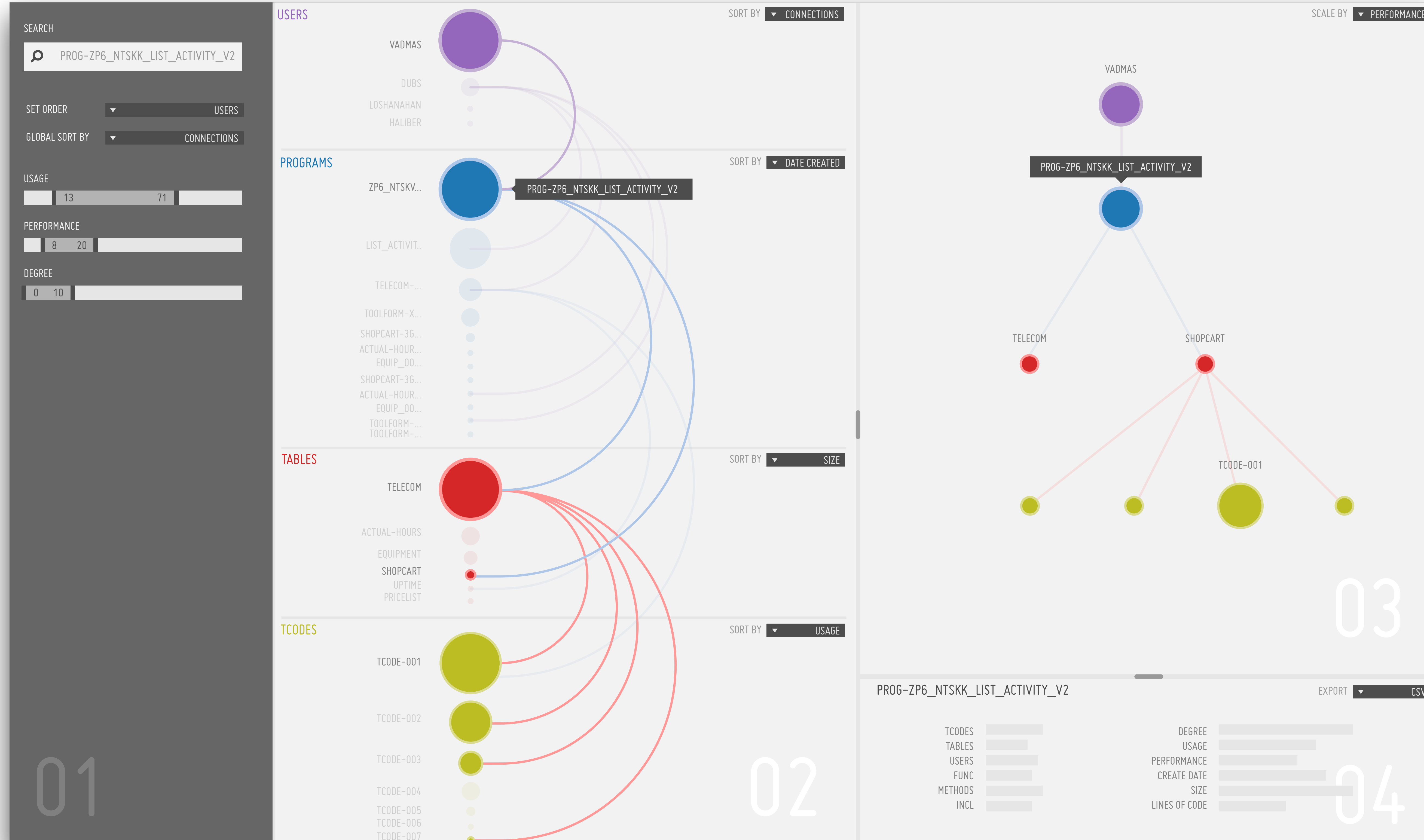
Indigo
Books · Gifts · Kids

Kellogg's

THE HOME DEPOT®

adidas®

Bell


SEARCH
PROG-ZP6_NTSKK_LIST_ACTIVITY_V2
SET ORDER
USERS
GLOBAL SORT BY
CONNECTIONS
USAGE
13
71
PERFORMANCE
8
20
DEGREE
0
10
01
USERS
SORT BY
CONNECTIONS
VADMAS
DUBS
LOSHANAHAN
HALIBER
PROGRAMS
SORT BY
DATE CREATED
ZP6_NTSKV...
PROG-ZP6_NTSKK_LIST_ACTIVITY_V2
LIST_ACTIVIT...
TELECOM-...
TOOLFORM-X...
SHOPCART-3G...
ACTUAL-HOUR...
EQUIP_00...
SHOPCART-3G...
ACTUAL-HOUR...
EQUIP_00...
TOOLFORM-...
TOOLFORM-...
TABLES
SORT BY
SIZE
TELECOM
ACTUAL-HOURS
EQUIPMENT
SHOPCART
UPTIME
PRICELIST
TCODES
SORT BY
USAGE
TCODE-001
TCODE-002
TCODE-003
TCODE-004
TCODE-005
TCODE-006
TCODE-007
02
SCALE BY
PERFORMANCE
VADMAS
PROG-ZP6_NTSKK_LIST_ACTIVITY_V2
TELECOM
SHOPCART
TCODE-001
03
PROG-ZP6_NTSKK_LIST_ACTIVITY_V2
EXPORT
CSV
TCODES
TABLES
USERS
FUNC
METHODS
INCL
DEGREE
USAGE
PERFORMANCE
CREATE DATE
SIZE
LINES OF CODE
04

547 Information Visualization

Pitches - Thu. Oct 22, 2015

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

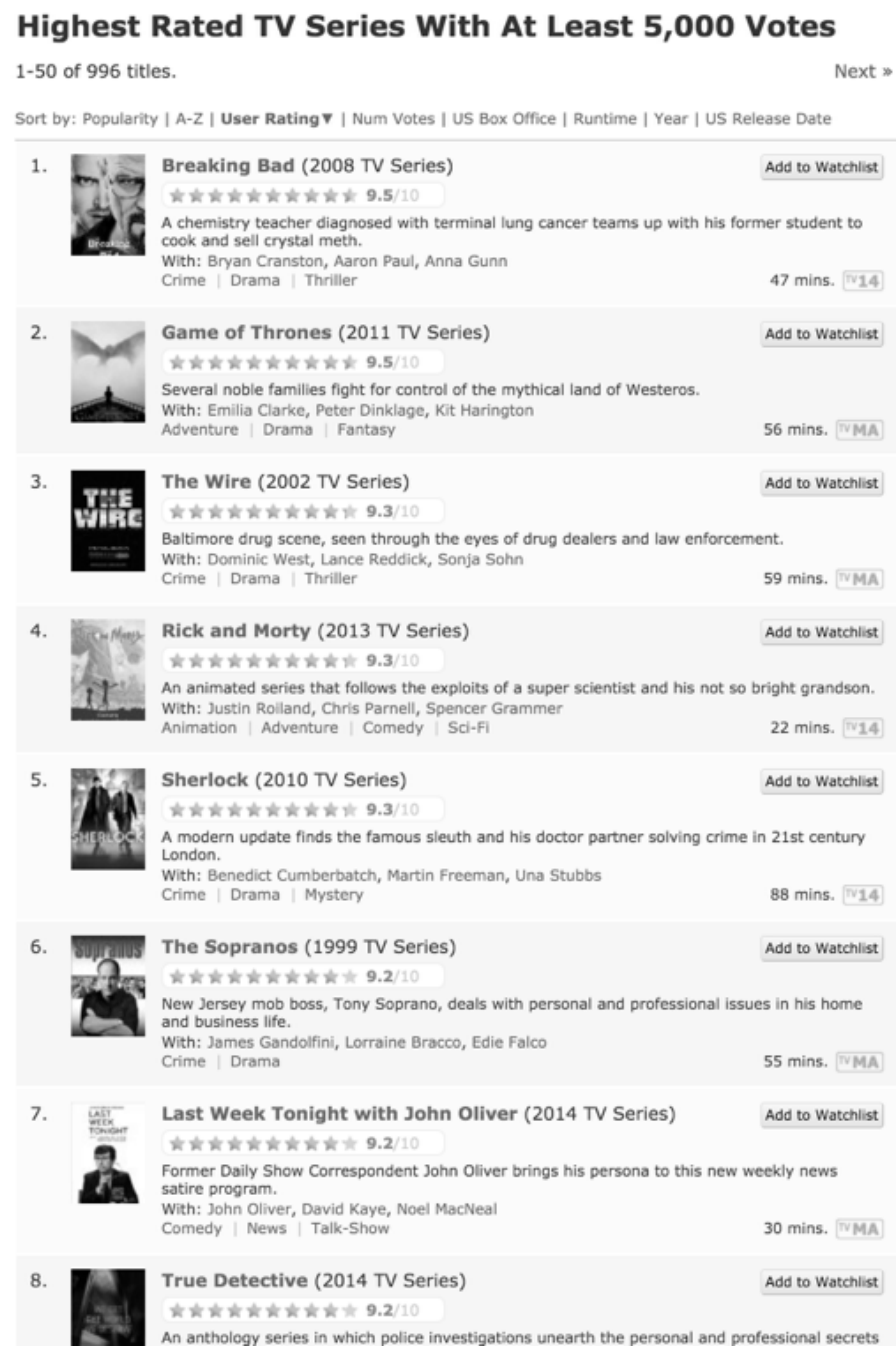
- 230K Users
- 30K Movies
- 21M Ratings

Photo from: http://www.fanpop.com/clubs/suits/images/34619277/title/wallpaper-wallpaper

Data set

Features
+ Titles
+ Genre
+ Rating
+ Tags (Themes)

- IMDB ID!
  + IMDB Rating
  + Metascore
  + Year
  + Director
  + Writer
  + Awards
  ++ Many more…

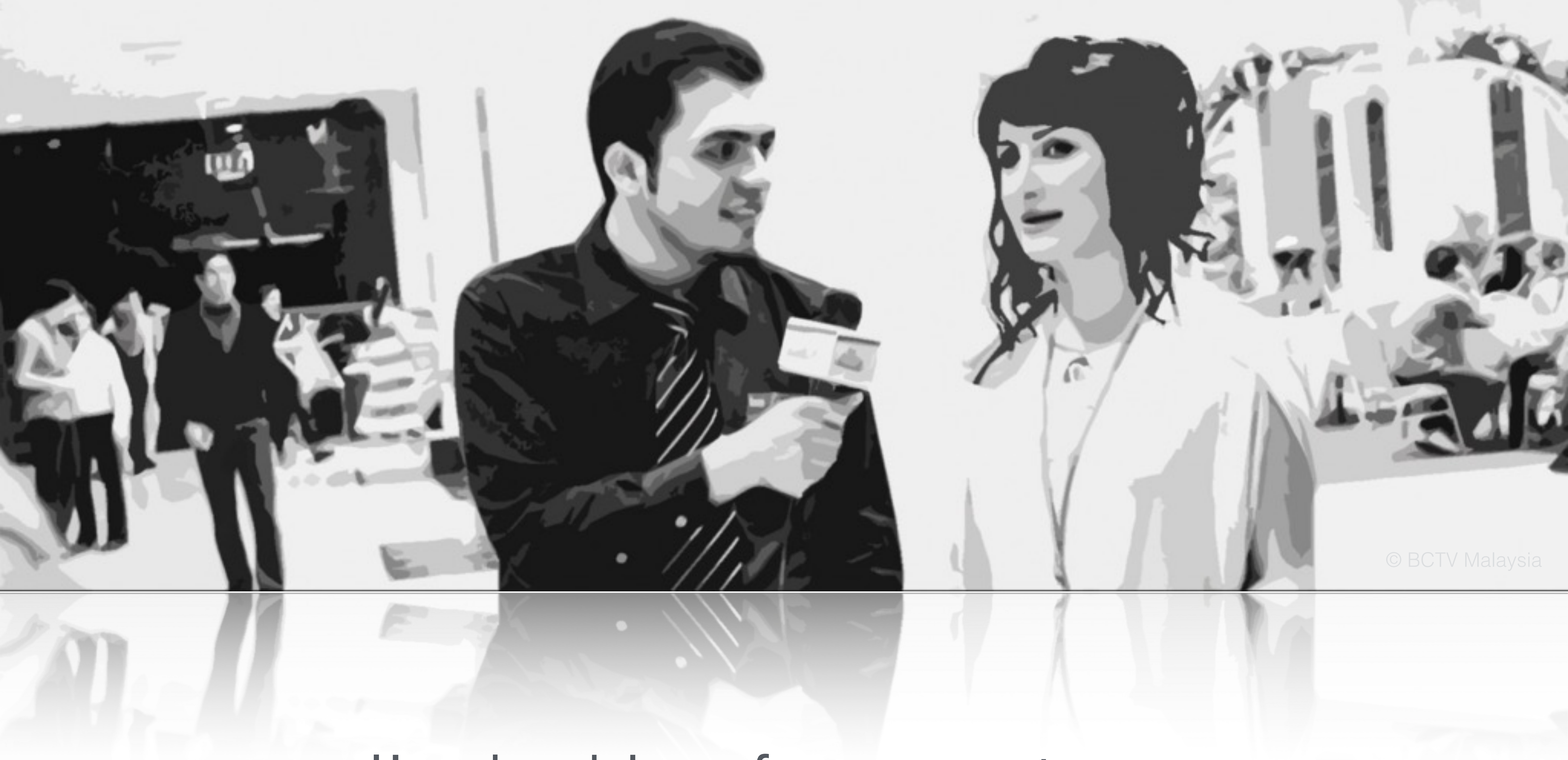


Challenges

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

Photo from: http://upallnightnews.com/top-10-reasons-jon-hamm-may-be-channeling-don-draper/



I'm looking for a partner,

Wanna join me?

Visualizing Uncertainty in Incomplete Election Data

CPSC 547 Project Pitch

Yasha Pushak

Federal election results

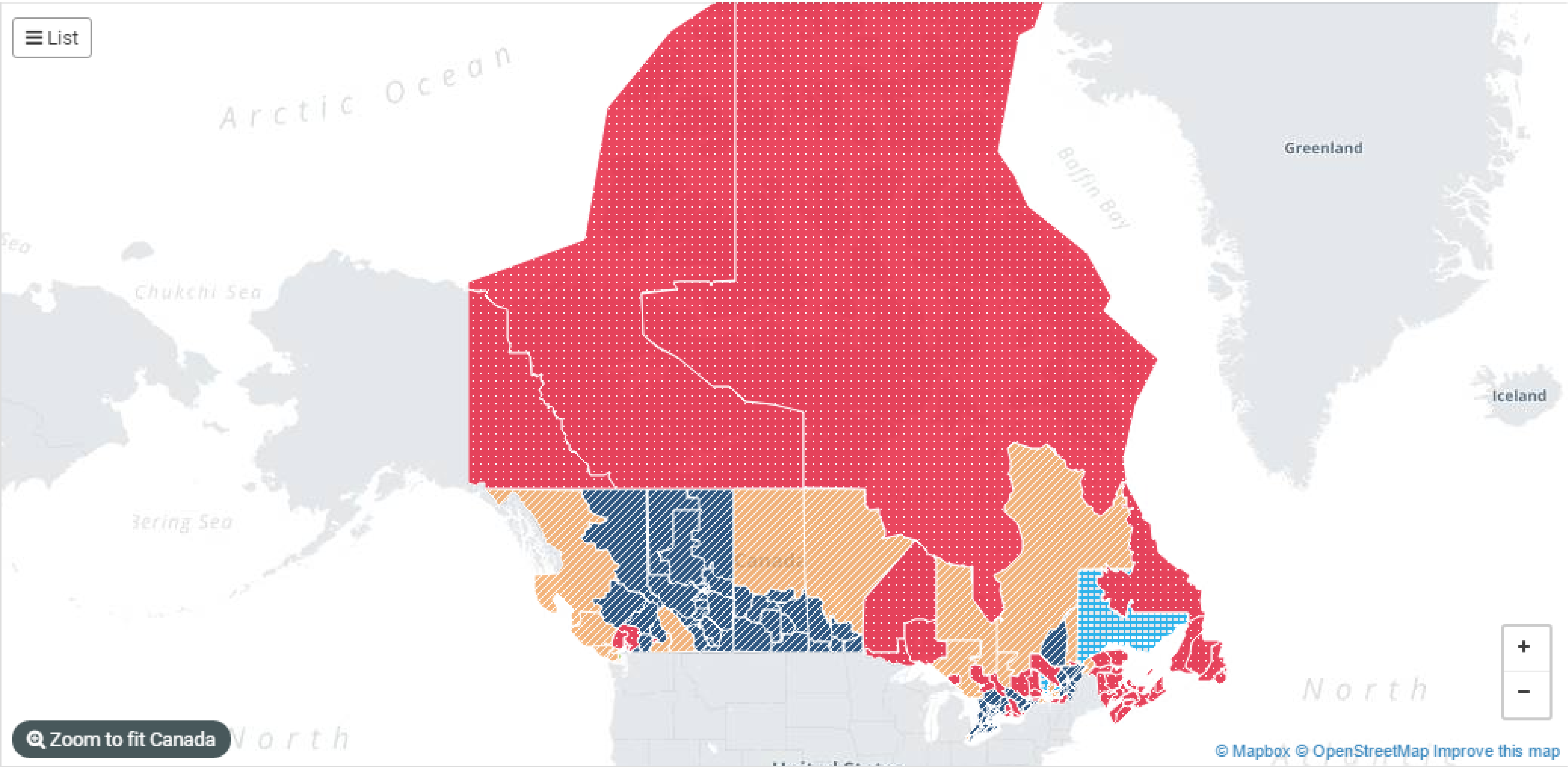
Leading and elected

Majority

Liberal majority

338

LIB	184	Elected: 184	Lead: 0	39.5%
CON	99	Elected: 99	Lead: 0	31.9%
NDP	44	Elected: 44	Lead: 0	19.7%
BQ	10	Elected: 10	Lead: 0	4.7%
GRN	1	Elected: 1	Lead: 0	3.5%
IND	0	Elected: 0	Lead: 0	0.2%
OTH	0	Elected: 0	Lead: 0	0.6%



Federal election results

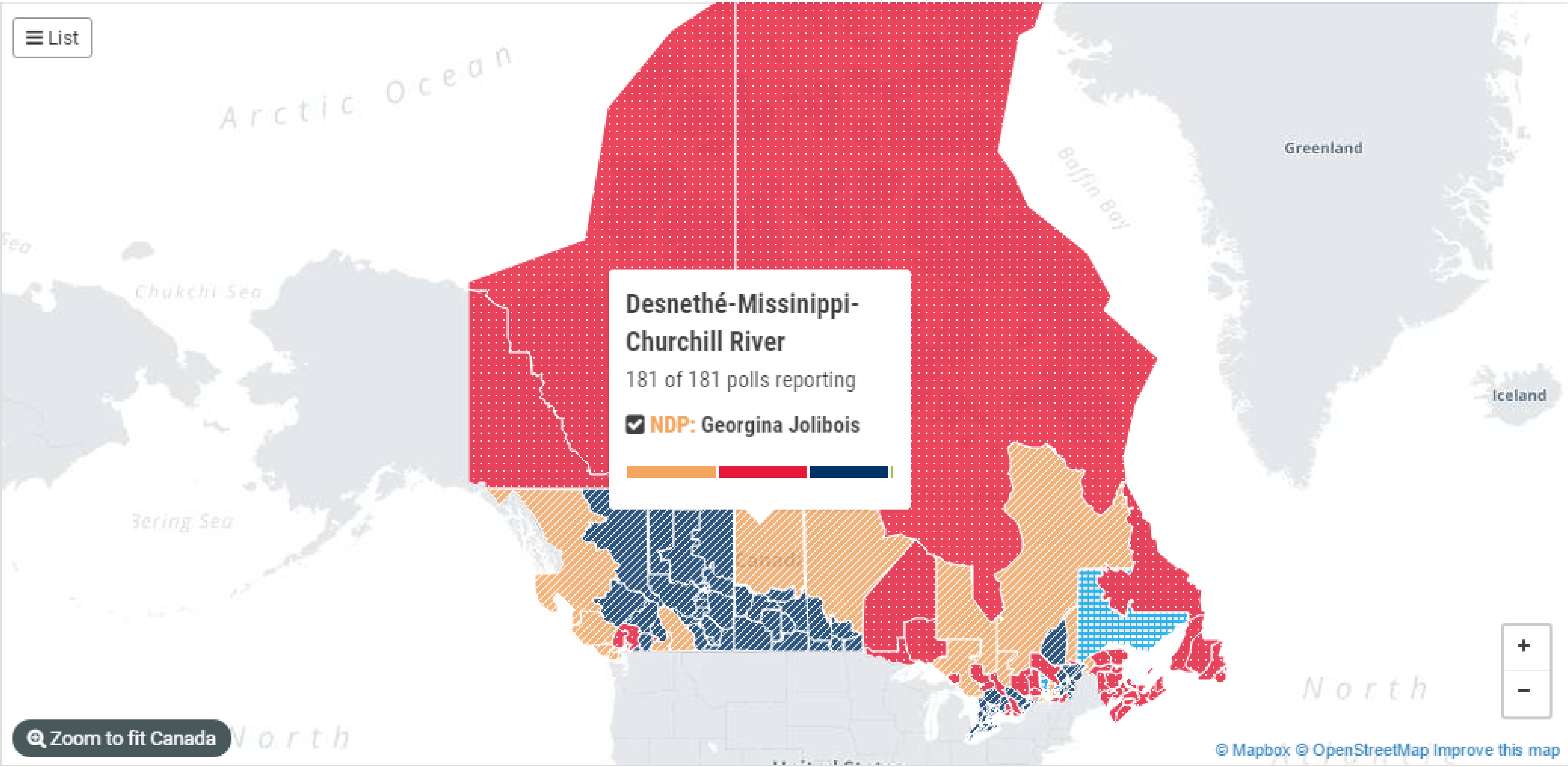
Leading and elected

Majority

Liberal majority

338

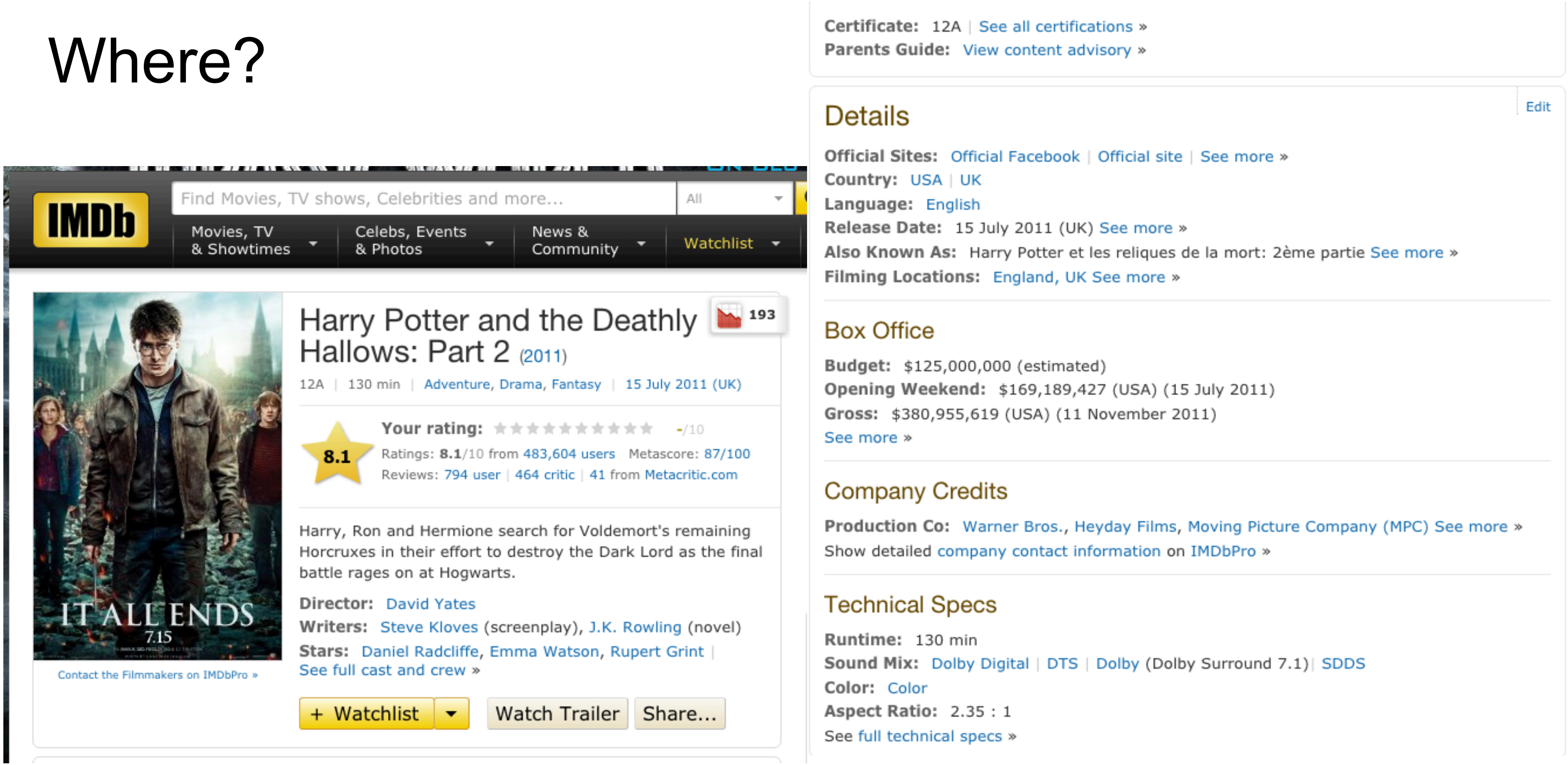
LIB	184	39.5%	Elected: 184	Lead: 0
CON	99	31.9%	Elected: 99	Lead: 0
NDP	44	19.7%	Elected: 44	Lead: 0
BQ	10	4.7%	Elected: 10	Lead: 0
GRN	1	3.5%	Elected: 1	Lead: 0
IND	0	0.2%	Elected: 0	Lead: 0
OTH	0	0.6%	Elected: 0	Lead: 0



Visualizing Movie Data

Yujie Yang, Ye Chen

Where?


IMDb
Find Movies, TV shows, Celebrities and more...
All
Movies, TV & Showtimes
Celebs, Events & Photos
News & Community
Watchlist
Harry Potter and the Deathly Hallows: Part 2 (2011)
193
12A | 130 min | Adventure, Drama, Fantasy | 15 July 2011 (UK)
Your rating: -/10
8.1
Ratings: 8.1/10 from 483,604 users Metascore: 87/100
Reviews: 794 user | 464 critic | 41 from Metacritic.com
Harry, Ron and Hermione search for Voldemort's remaining Horcruxes in their effort to destroy the Dark Lord as the final battle rages on at Hogwarts.
Director: David Yates
Writers: Steve Kloves (screenplay), J.K. Rowling (novel)
Stars: Daniel Radcliffe, Emma Watson, Rupert Grint | See full cast and crew »
IT ALL ENDS
7.15
Contact the Filmmakers on IMDbPro »
+ Watchlist
Watch Trailer
Share...
Certificate: 12A | See all certifications »
Parents Guide: View content advisory »
Details
Edit
Official Sites: Official Facebook | Official site | See more »
Country: USA | UK
Language: English
Release Date: 15 July 2011 (UK) See more »
Also Known As: Harry Potter et les reliques de la mort: 2ème partie See more »
Filming Locations: England, UK See more »
Box Office
Budget: $125,000,000 (estimated)
Opening Weekend: $169,189,427 (USA) (15 July 2011)
Gross: $380,955,619 (USA) (11 November 2011)
See more »
Company Credits
Production Co: Warner Bros., Heyday Films, Moving Picture Company (MPC) See more »
Show detailed company contact information on IMDbPro »
Technical Specs
Runtime: 130 min
Sound Mix: Dolby Digital | DTS | Dolby (Dolby Surround 7.1) | SDDS
Color: Color
Aspect Ratio: 2.35 : 1
See full technical specs »

What?

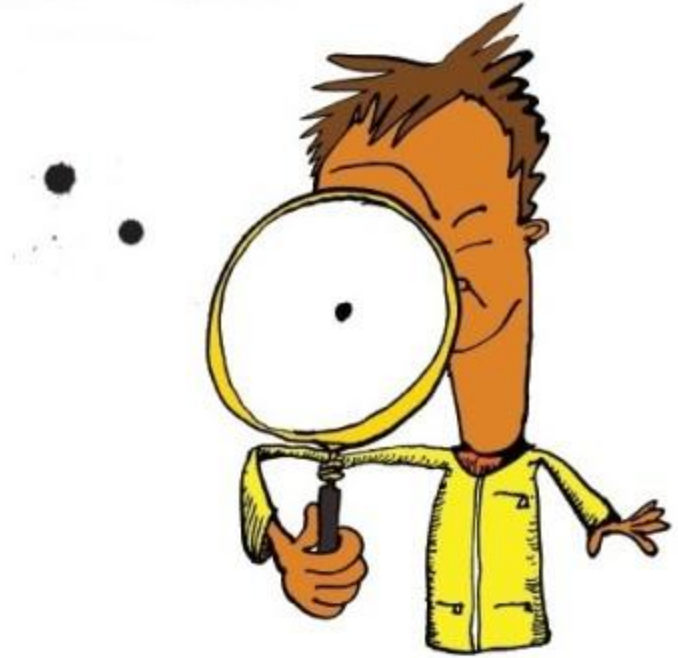
name	year	rate	runtime(mi	gener	certificate	boxoffice	budget	opening_thea	release_date
Avatar (2009)	2009	7.9	162	Action | Adventure | Fantasy | Sci-Fi	PG_13	761000000	$237,000,000 (estimated)	3,452	18 December 2009 (Ca
The Dark Knight (2008)	2008	9	152	Action | Crime | Drama	PG_13	533000000	$185,000,000 (estimated)	4,366	18 July 2008 (Canada
Pirates of the Caribbean: Dead Man's Chest (2006)	2006	7.3	151	Action | Adventure | Fantasy	PG_13	423000000	$225,000,000 (estimated)	4,133	7 July 2006 (USA)
Toy Story 3 (2010)	2010	8.4	103	Animation | Adventure | Comedy | Family | Fa	G	415000000	$200,000,000 (estimated)	4,028	18 June 2010 (Canada
Transformers: Revenge of the Fallen (2009)	2009	6	150	Action | Adventure | Sci-Fi	PG_13	402000000	$200,000,000 (estimated)	4,234	24 June 2009 (Canada
Harry Potter and the Deathly Hallows: Part 2 (2011)	2011	8.1	130	Adventure | Fantasy | Mystery	PG_13	381000000	$125,000,000 (estimated)	4,375	15 July 2011 (Canada
Star Wars: Episode III - Revenge of the Sith (2005)	2005	7.7	140	Action | Adventure | Fantasy | Sci-Fi	PG_13	380000000	$113,000,000 (estimated)	3,661	19 May 2005 (Canada)
Transformers: Dark of the Moon (2011)	2011	6.3	154	Action | Adventure | Sci-Fi	PG_13	352000000	$195,000,000 (estimated)	4,011	29 June 2011 (Canada
Spider-Man 3 (2007)	2007	6.2	139	Action | Adventure	PG_13	337000000	$258,000,000 (estimated)	4,252	4 May 2007 (Canada)
Alice in Wonderland (2010)	2010	6.5	108	Adventure | Family | Fantasy	PG	334000000	$200,000,000 (estimated)	3,728	5 March 2010 (Canada
Shrek the Third (2007)	2007	6	93	Animation | Adventure | Comedy | Family | Fa	PG	321000000	$160,000,000 (estimated)	4,122	18 May 2007 (Canada)
Transformers (2007)	2007	7.2	144	Action | Adventure | Sci-Fi	PG_13	319000000	$150,000,000 (estimated)	4,011	2 July 2007 (Canada)
Iron Man (2008)	2008	7.9	126	Action | Adventure | Sci-Fi	PG_13	318000000	$140,000,000 (estimated)	4,105	2 May 2008 (Canada)
Indiana Jones and the Kingdom of the Crystal Skull (20	2008	6.2	122	Action | Adventure	PG_13	317000000	$185,000,000 (estimated)	4,260	22 May 2008 (Canada)
Iron Man 2 (2010)	2010	7.1	124	Action | Adventure | Sci-Fi	PG_13	312000000	$200,000,000 (estimated)	4,380	7 May 2010 (Canada)
Pirates of the Caribbean: At World's End (2007)	2007	7.1	169	Action | Adventure | Fantasy	PG_13	309000000	$300,000,000 (estimated)	4,362	25 May 2007 (USA)
Harry Potter and the Half-Blood Prince (2009)	2009	7.5	153	Adventure | Family | Fantasy | Mystery	PG	302000000	$250,000,000 (estimated)	4,275	15 July 2009 (Canada
The Twilight Saga: Eclipse (2010)	2010	4.9	124	Adventure | Drama | Fantasy | Romance	PG_13	301000000	$68,000,000 (estimated)	4,416	30 June 2010 (Canada
The Twilight Saga: New Moon (2009)	2009	4.6	130	Adventure | Drama | Fantasy | Romance	PG_13	297000000	$50,000,000 (estimated)	4,024	20 November 2009 (Ca
Harry Potter and the Deathly Hallows: Part 1 (2010)	2010	7.7	146	Adventure | Family | Fantasy | Mystery	PG_13	295000000	Â£150,000,000 (estimated	4,125	19 November 2010 (Ca
Up (2009)	2009	8.3	96	Animation | Adventure | Comedy | Family	PG	293000000	$175,000,000 (estimated)	3,766	29 May 2009 (Canada)
Inception (2010)	2010	8.8	148	Action | Mystery | Sci-Fi | Thriller	PG_13	293000000	$160,000,000 (estimated)	3,792	16 July 2010 (Canada
The Chronicles of Narnia: The Lion, the Witch and the	2005	6.9	143	Adventure | Family | Fantasy	UNRATED	292000000	$180,000,000 (estimated)	3,616	9 December 2005 (Ca
Harry Potter and the Order of the Phoenix (2007)	2007	7.4	138	Adventure | Family | Fantasy | Mystery	PG_13	292000000	$150,000,000 (estimated)	4,181	11 July 2007 (Canada
Harry Potter and the Goblet of Fire (2005)	2005	7.6	157	Adventure | Family | Fantasy | Mystery	PG_13	290000000	$150,000,000 (estimated)	3,858	18 November 2005 (Ca
The Twilight Saga: Breaking Dawn - Part 1 (2011)	2011	4.9	117	Adventure | Drama | Fantasy | Romance	PG_13	281000000	$110,000,000 (estimated)	4,066	18 November 2011 (Ca
The Hangover (2009)	2009	7.8	100	Comedy	UNRATED	277000000	$35,000,000 (estimated)	3,269	5 June 2009 (Canada)
Star Trek (2009)	2009	8	127	Action | Adventure | Sci-Fi	PG_13	258000000	$140,000,000 (estimated)	3,849	8 May 2009 (Canada)
I Am Legend (2007)	2007	7.2	101	Drama | Sci-Fi | Thriller	PG_13	256000000	$150,000,000 (estimated)	3,606	14 December 2007 (Ca
The Blind Side (2009)	2009	7.7	129	Biography | Drama | Sport	PG_13	256000000	$29,000,000 (estimated)	3,110	20 November 2009 (Ca
The Hangover Part II (2011)	2011	6.5	102	Comedy	R	254000000	$80,000,000 (estimated)	3,615	26 May 2011 (Canada)
Despicable Me (2010)	2010	7.7	95	Animation | Comedy | Family	TV_PG	252000000	$69,000,000 (estimated)	3,476	9 July 2010 (Canada)
Night at the Museum (2006)	2006	6.4	108	Action | Adventure | Comedy | Family | Fantas	PG	251000000	$110,000,000 (estimated)	3,768	22 December 2006 (Ca

What?

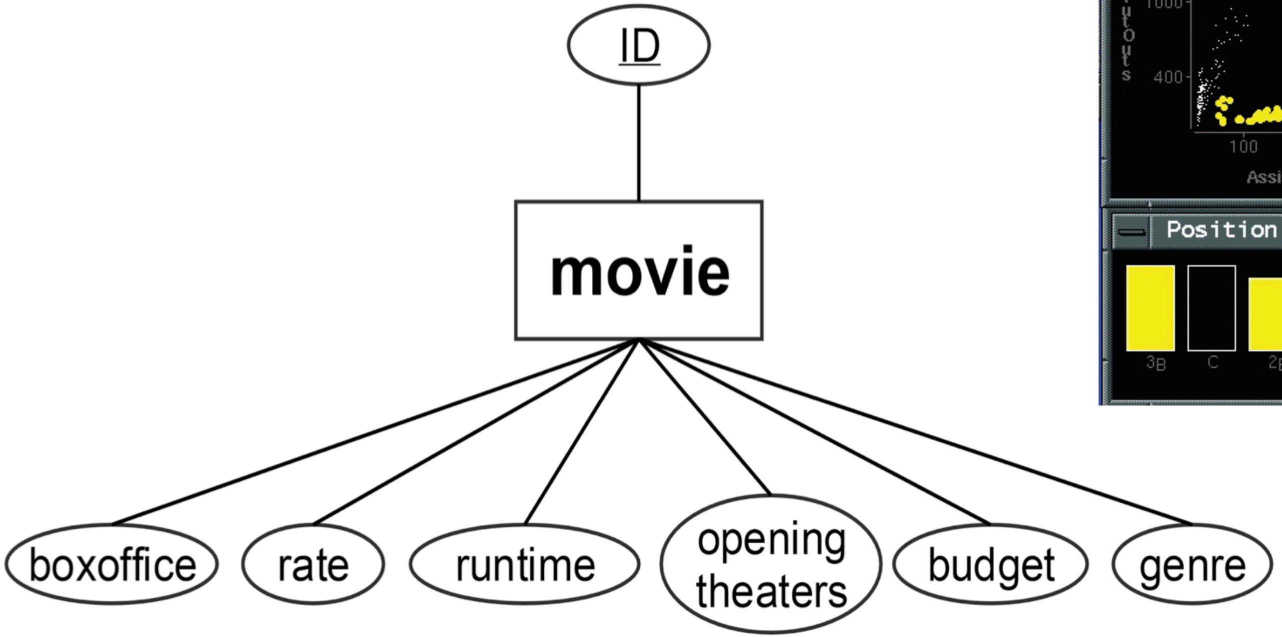
ID	Name	Boxoffice (USD)	Rate	Runtime (mins)	Opening Theaters	Budget	Release Date	Genre
47	Up	293000000	8.3	96	3766	175000000	May 2009	Animation | Adventure | Comedy | Family

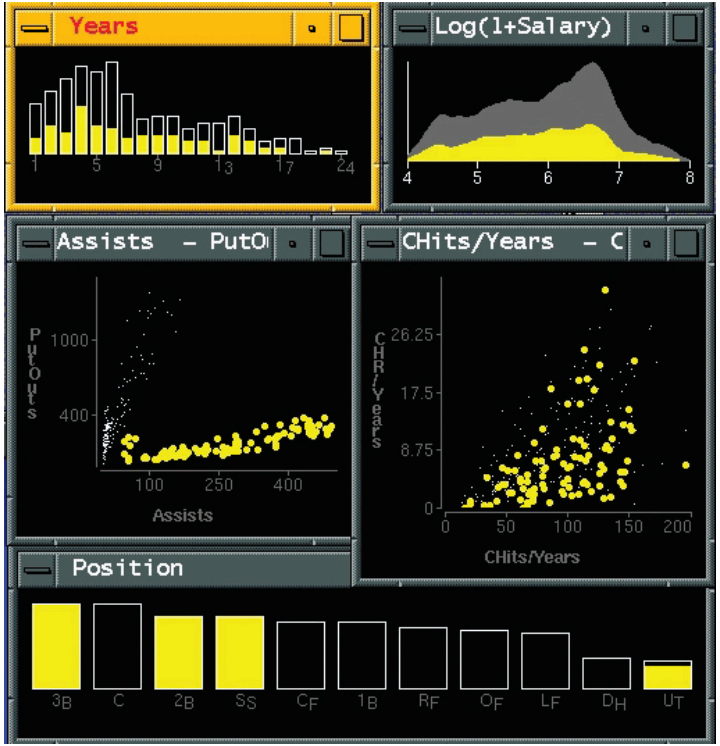
Why?

Explore the distribution of data.



How?

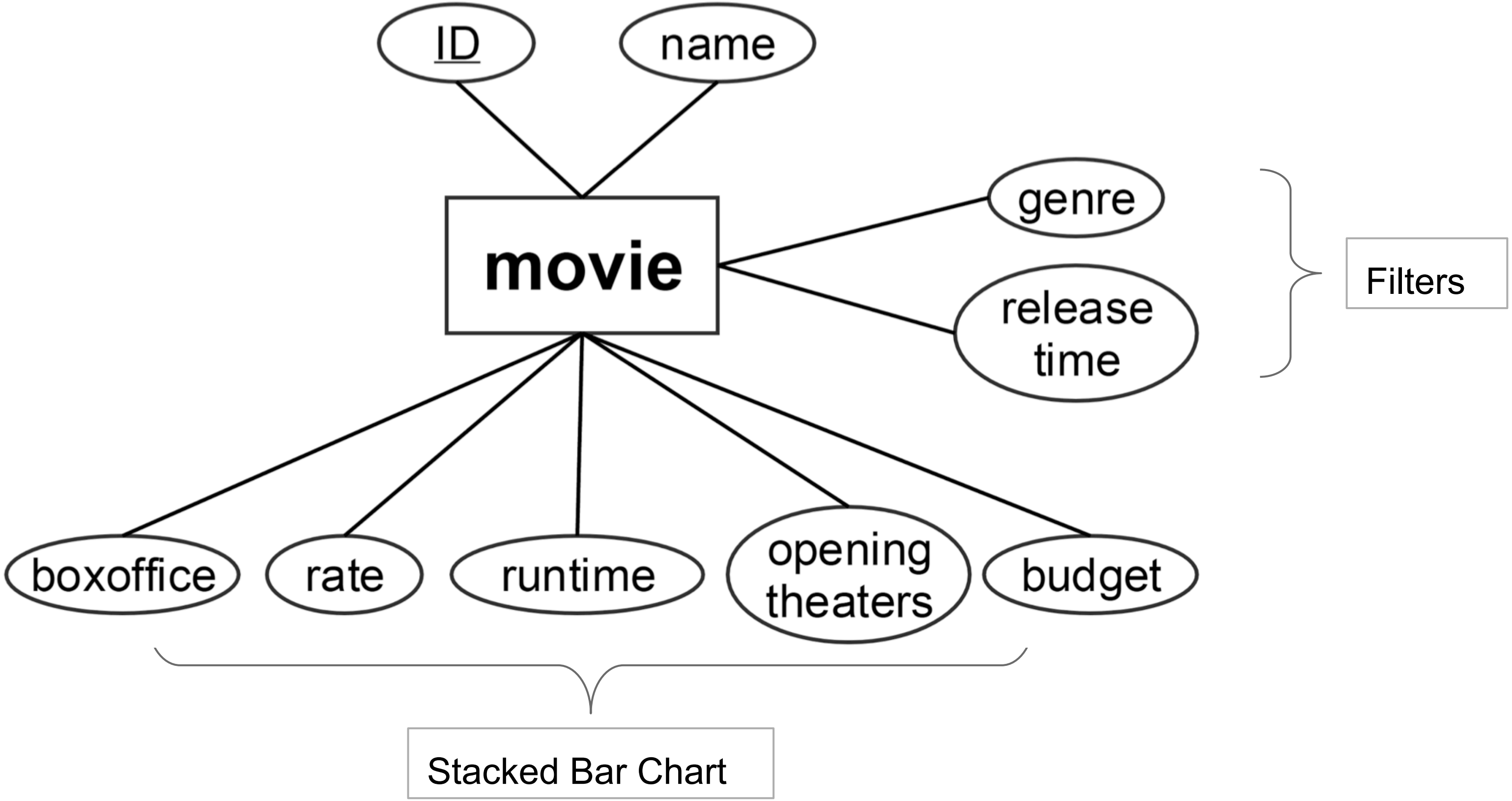

ID
movie
boxoffice
rate
runtime
opening theaters
budget
genre


Years
Log(1+Salary)
Assists – PutO
CHits/Years – C
Position
Assists
CHits/Years
3B
C
2B
SS
CF
1B
RF
OF
LF
DH
UT

Why?

Help people to choose a movie.

How?


ID
name
movie
genre
release time
Filters
boxoffice
rate
runtime
opening theaters
budget
Stacked Bar Chart

How?

Quantity Filter

Genre Filter

Year Filter

Month Filter

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…

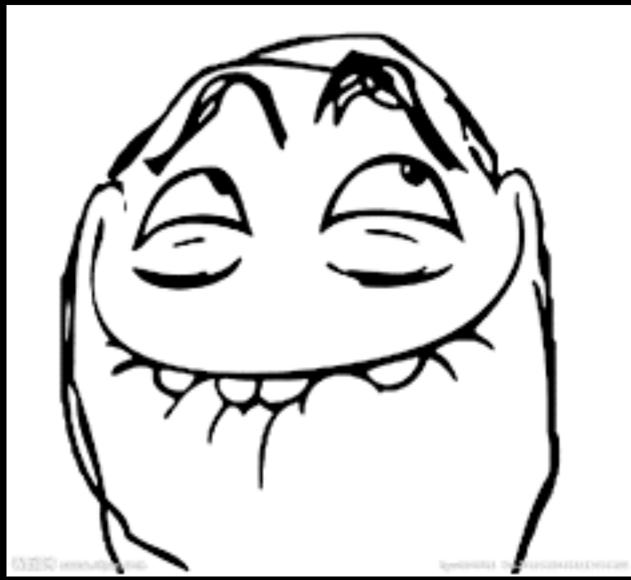
Sets

Dataset

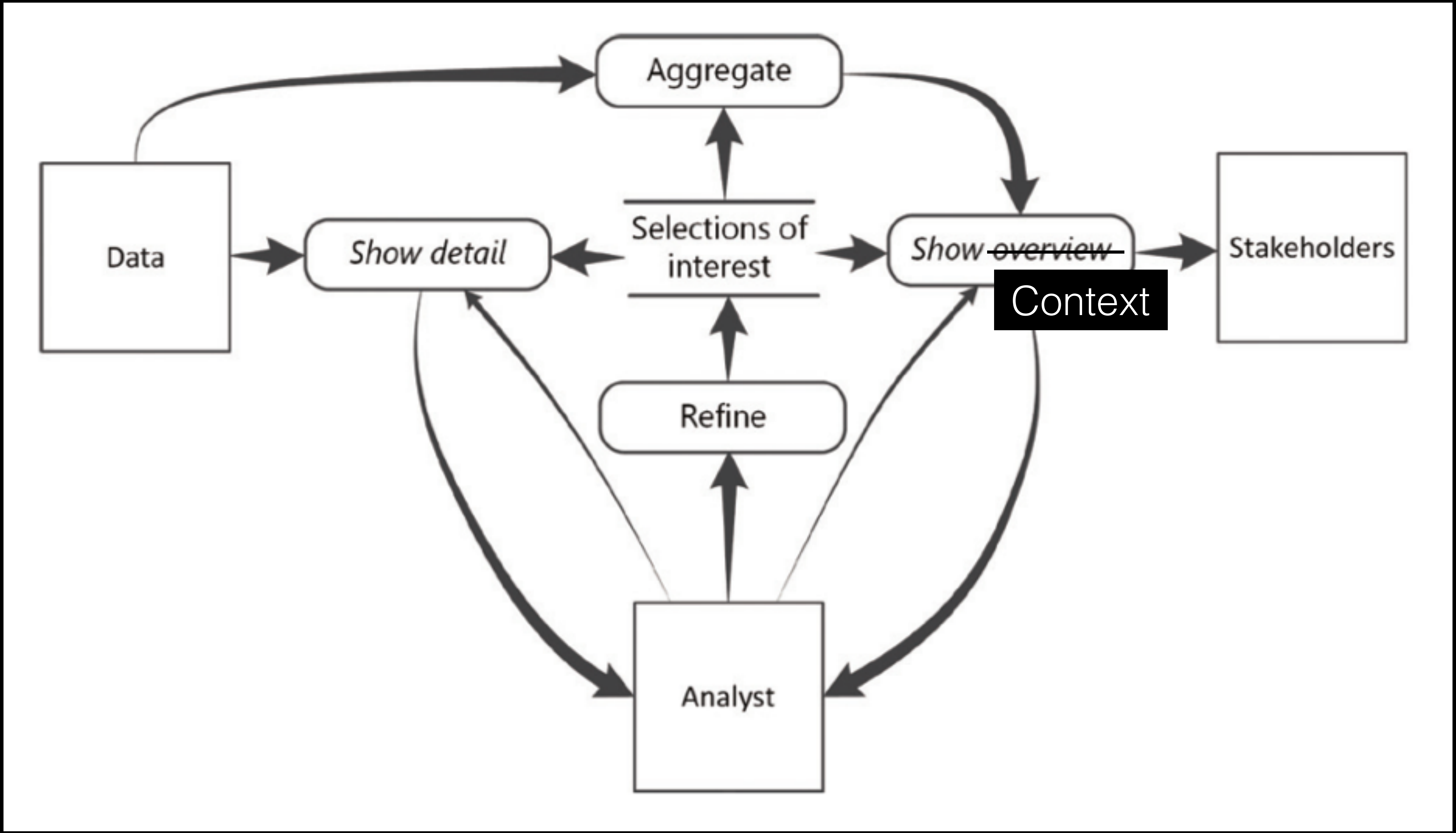
- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms…

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

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

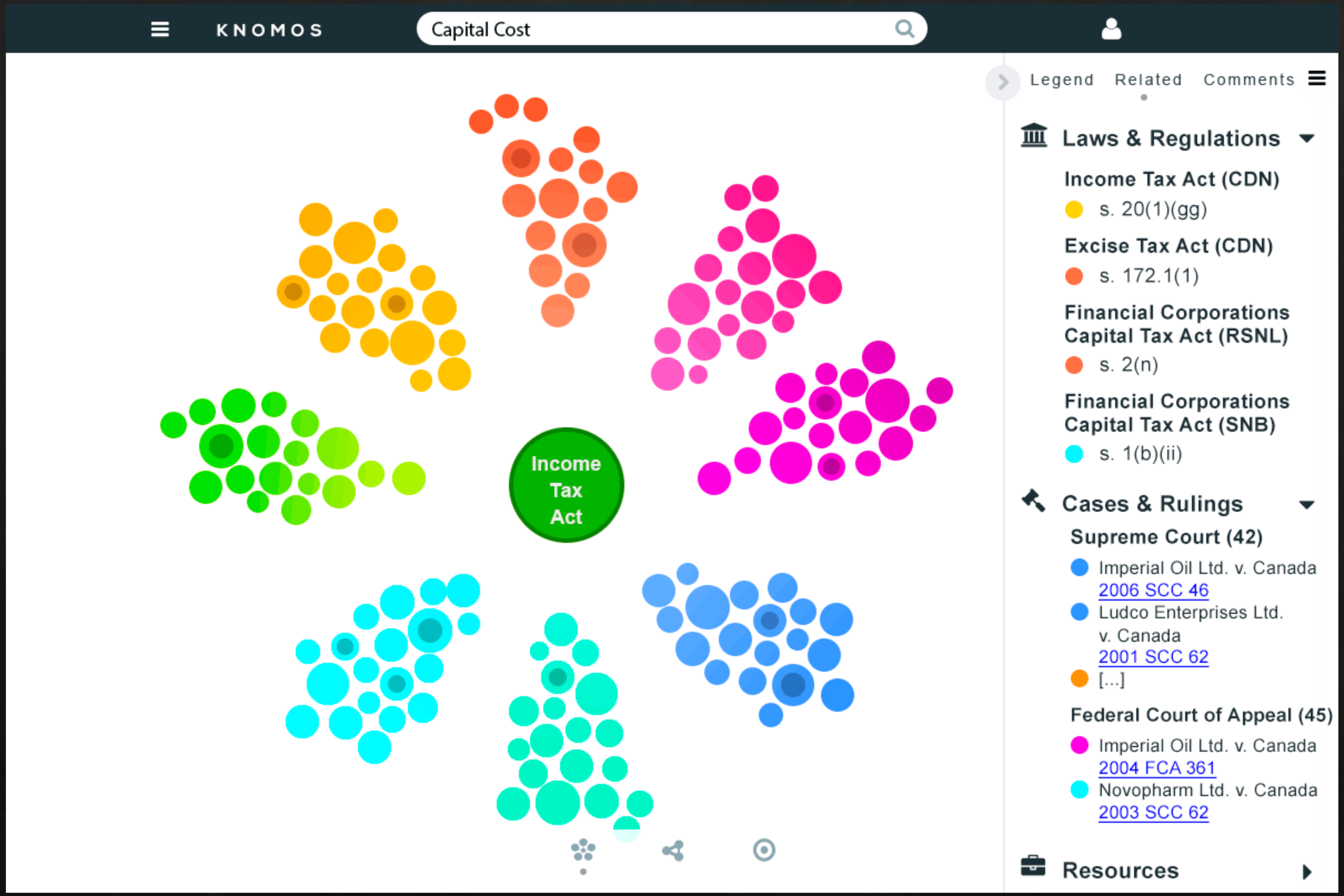
High Client Costs
Limited Access to Justice

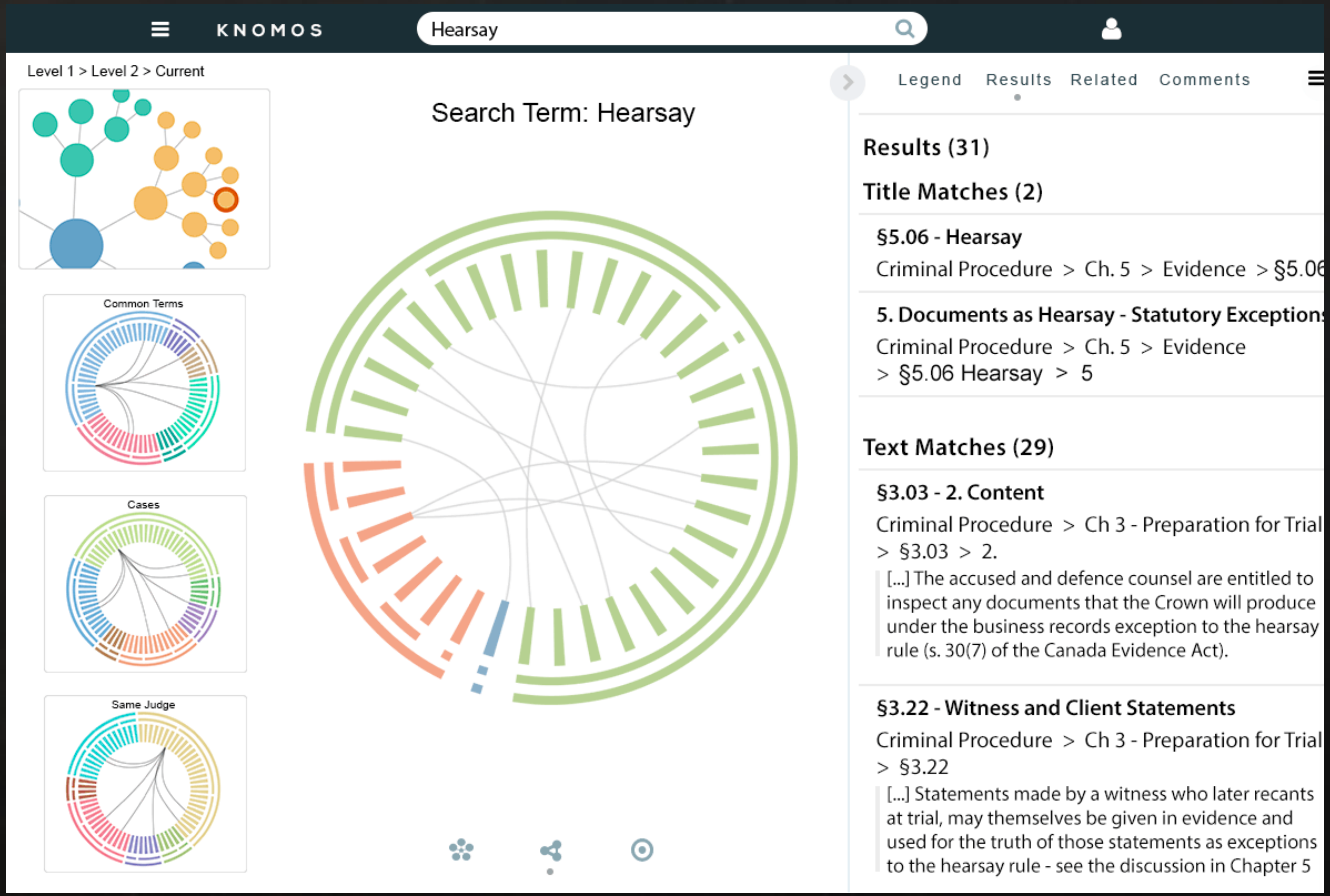
Content Barriers
Private Content Paywall

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

Core Team




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

Key Partners

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