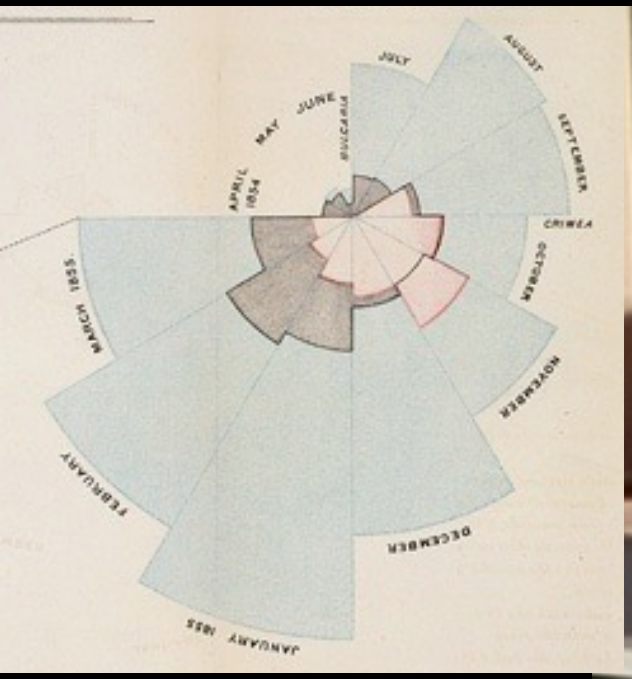
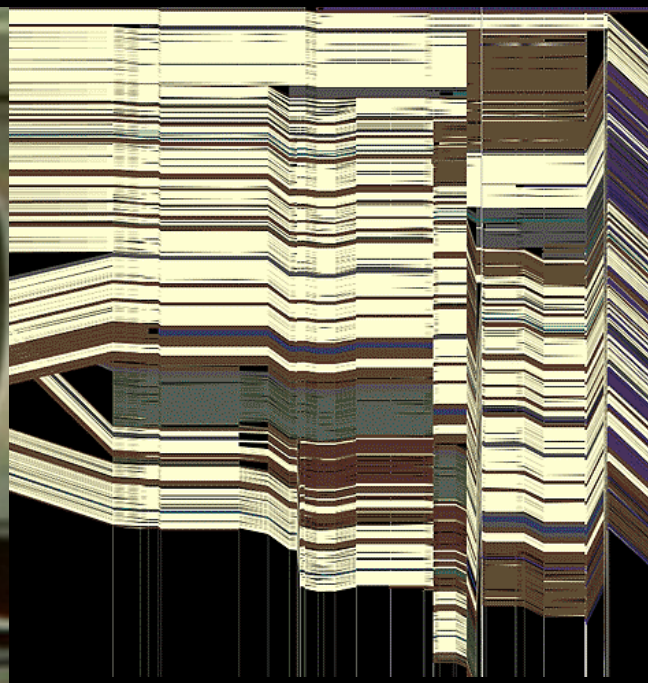
CSE 512 - Data Visualization

Progress Presentations





Spring 2016 University of Washington

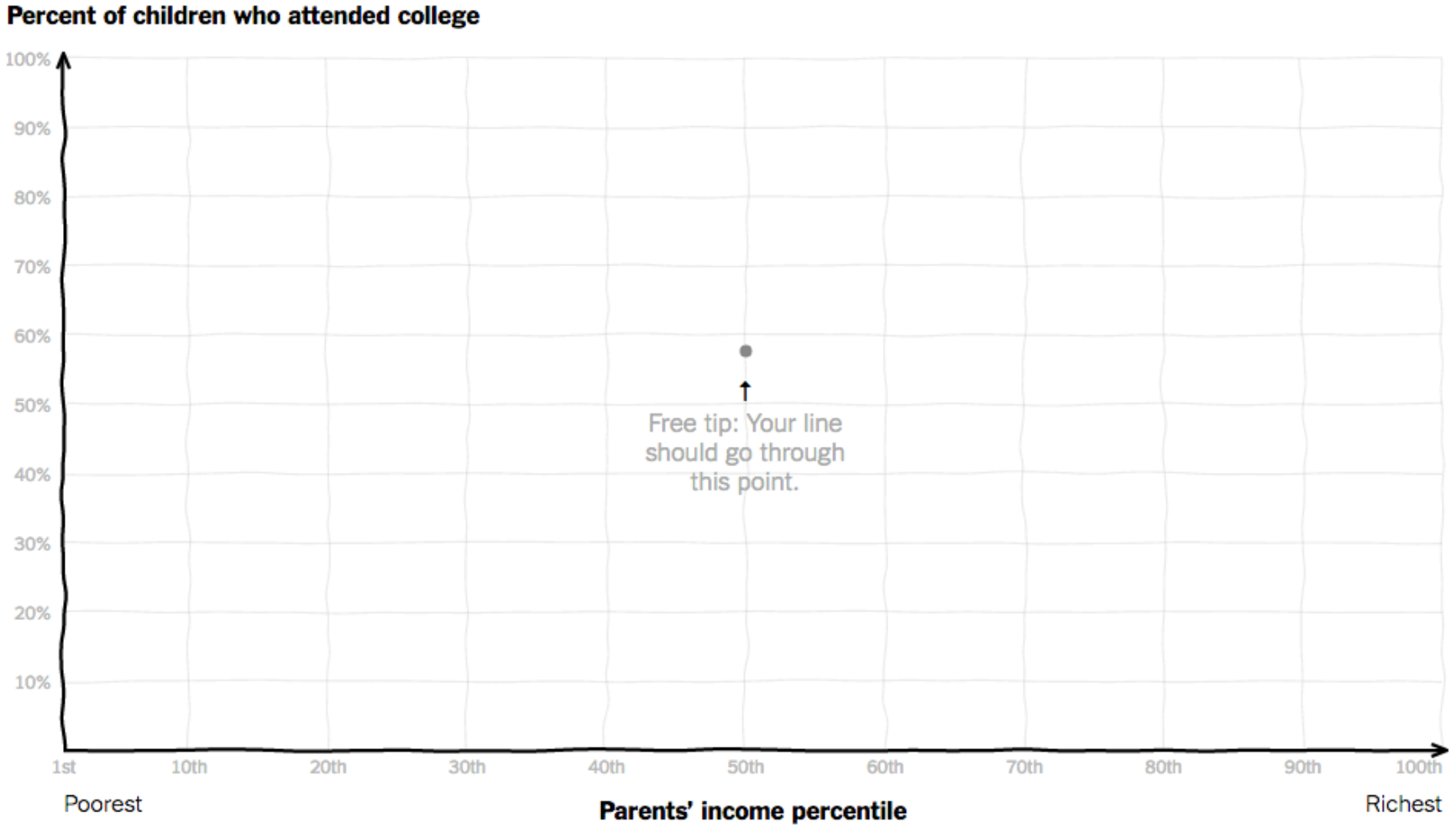
Presentation Order (EEB 105)

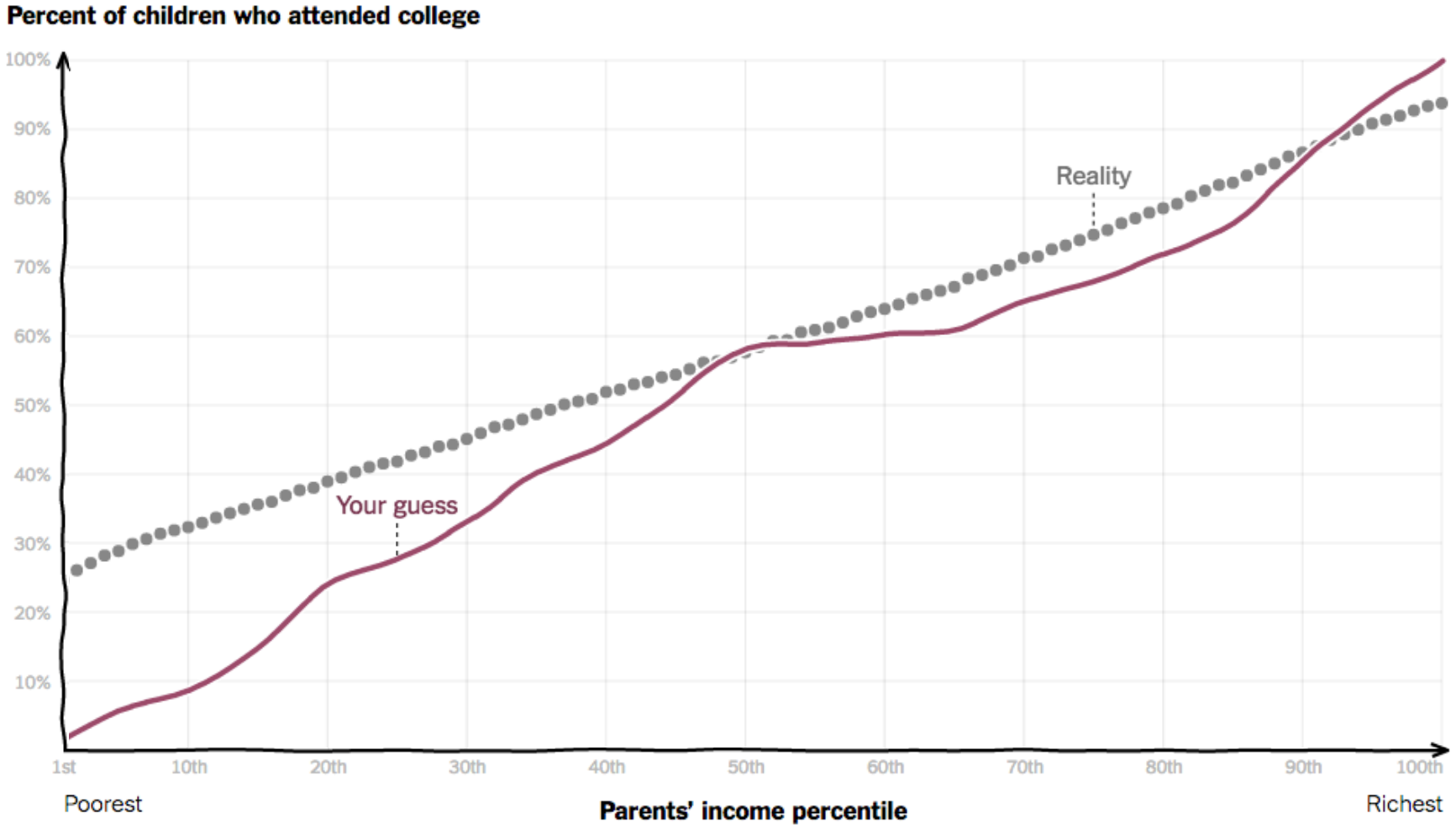
1. Lovenoor Aulck, Yea-Seul Kim, Ian Wesley-Smith
2. Fida Alsughayer
3. Elizabeth Clark, Lucy Lin, George Mulcaire, Maarten Sap
4. Gagan Bansal, Christopher Clark, Mohit Jain, Jin Qu
5. Wesley Beckner, Janet Matsen, Matthew Murbach
6. Adwin Jahn, Ryan Maas, Harley Montgomery
7. Shirley Leung, Philippe Vaillant, Helena van Tol, Michelle Weirathmueller
8. Rachel Li, Guanming Wang, Xiaojing Zhu
9. Peiran Liu, Mengjie Pan, Alexander Tank, Yali Wan
10. Chris Chung, Hyun Kim, Lyle Klyne, Ankit Potdar
11. Benjamin Jones, Julie Newcomb
12. Helga Gudmundsdottir
13. Ryan McGee
14. Arushi Prakash

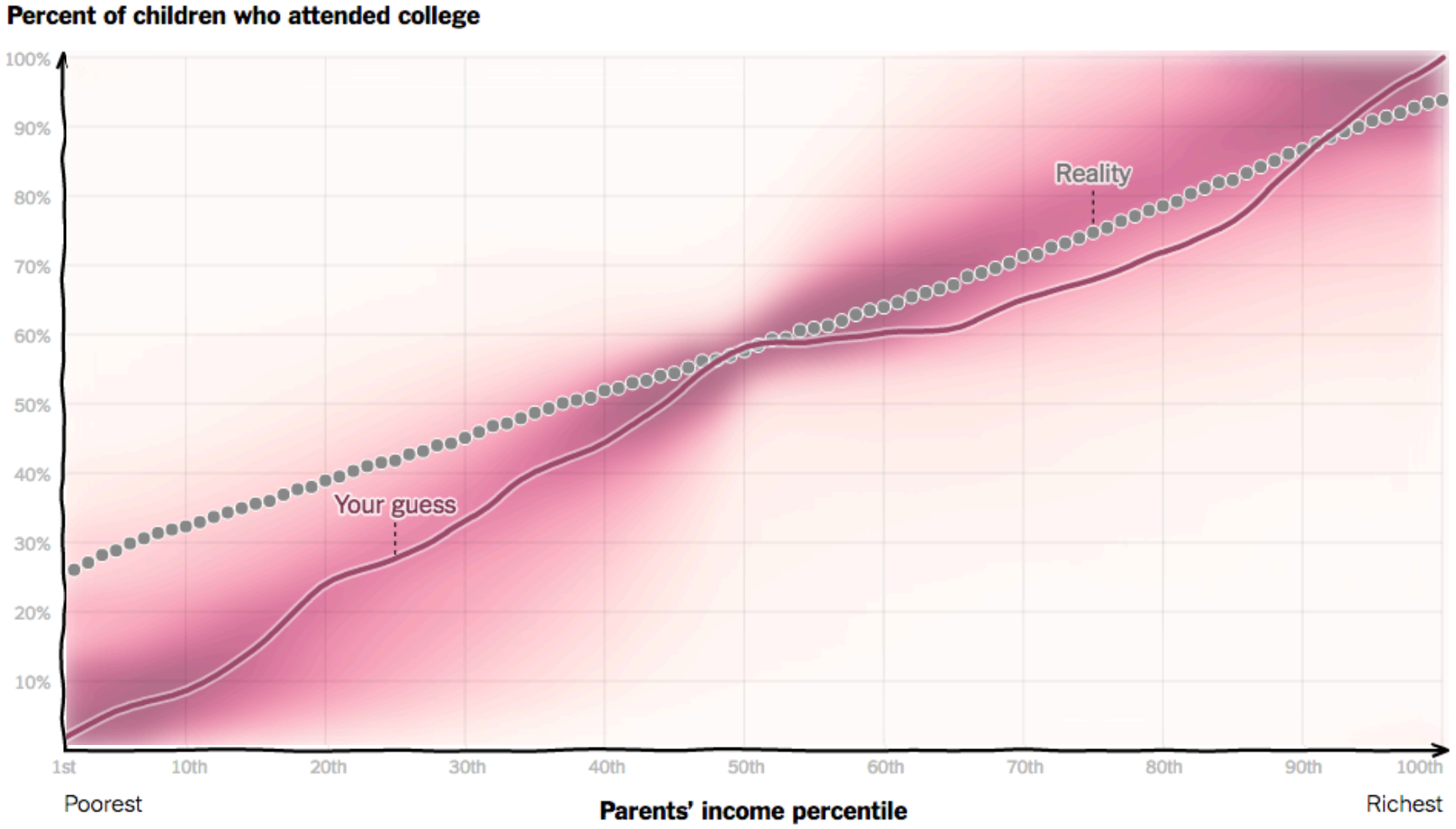
Expectation Visualization

Ian Wesley-Smith, Lovenoor Aulck, Yea-Seul Kim

“What if visualizations allowed people to draw their expectations of the data prior to viewing?”


Percent of children who attended college
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
Free tip: Your line should go through this point.
1st
10th
20th
30th
40th
50th
60th
70th
80th
90th
100th
Poorest
Parents' income percentile
Richest


Percent of children who attended college
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
Reality
Your guess
1st
10th
20th
30th
40th
50th
60th
70th
80th
90th
100th
Poorest
Parents' income percentile
Richest


Percent of children who attended college
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
Reality
Your guess
1st
10th
20th
30th
40th
50th
60th
70th
80th
90th
100th
Poorest
Parents' income percentile
Richest

Process

Define design space

Identify application

Building Tools

Validate effects

User-driven Expectation Visualization: Opportunities for Personalized Feedback

Yea-Seul Kim and Jessica Hullman
University of Washington

Abstract

In this paper, we define and motivate Expectation Visualization, an interactive technique for soliciting, and presenting personalized feedback on, a user's expectation of the data. Expectation Visualization (EV) addresses the common challenge faced by designers of how to engage users with visualized data on a deeper level. We describe the design space of EV, including how it can be used for data encoded in marks and mark attributes, and describing forms of training and personalized feedback. We propose three specific applications where the benefits of EV may be particularly useful. We conclude with ideas for future research.

Keywords: Human-centered computing, Visualization, Information visualization.

1 Introduction

Traditionally interaction in visualization has been understood as user-driven manipulation of views of an existing data set: for example a user might select, sort, or filter a view [3]. Such interactions allow a user to generate views that are particularly relevant to him or her. As rich interactive visualizations become easier to author and more ubiquitous in media and other outlets, novel interactions, particularly those that make the data personally relevant to a user, are desired.

One form of "personalized" interaction that researchers have argued for is the manipulation by the user of internal (mental) representations as he or she makes sense of visualized data [10]. Relating an external visualization to one's internal representation, or mental model, can lead to better comprehension of gaps in one's knowledge [4][13]. For example mentally animating a set of small multiples showing a physics process, or comparing the small multiples to an internal representation, can lead to deeper understanding of the process [6].

Prompting users to engage in self-explanation, the process by which users explain its concept or example to themselves, may be useful as a means of guiding a user to compare internal representations to an external visualization [1]. As a form of prediction, self-explaining stimulates greater engagement with the topic. For example, it has been proposed that a system might prompt more active processing of data by asking a user to guess the direction of the trend prior to viewing the data [7].

In most examples studied in psychology [4][6][13], the benefits of self-explaining and internal visualizations come only after considerable cognitive work on the part of the user, who must mentally imagine the difference between their internal representation and the external visualization. In this work, we consider a new possibility: What if visualizations allowed people to draw their expectations of the data prior to viewing? Seeing the data along with the expectation provides a form of personalized feedback, as it renders this gap between expectation and fact explicit. The act of drawing maintains a user's engagement, while the visual representation of the expectation against the result allows more detailed observations than may be possible through mental visualization alone.

In this paper, we define Expectation Visualization (EV), an interactive technique for soliciting and presenting personalized feedback on a user's expectations of the data. Expectation Visualization (EV) addresses the common challenge faced by designers of how to engage users with visualized data on a deeper level. In the rest of the paper, we describe the design space of EV including how EV could be applied to common visualization tasks and visual encodings. We propose specific types of applications in which the cognitive benefits of EV could be useful. We conclude by offering ideas for future work.

2 Expectation Visualization Design Space

We use a recent New York Times interactive graphic to motivate the design space of EV. We define the design space according to tasks and visual encodings. We discuss forms of training and presentation of personalized feedback.

2.1 Predicting Marks

We begin our characterization of the design space for EV by considering a recent New York Times graphic of this technique (Fig. 1). The interface presents a XY plot without any data shown, however the axes are labeled as *Parent's income percentile* and *Percent of children who attend college*. The user is encouraged to draw their guess for each income level in the chart (Fig. 1a). In the accompanying text, definitions of various visual trends are provided to help users to relate the graphical representation to their expectations (e.g.,  or ). After the user is done drawing the line, the true trend is presented as an overlay on the chart showing the user's expectation (Fig. 1c). Statistics describing how the user's guess compares to those of other users are presented in text below the chart.

In this example, the user's mental model is represented by a trend line. Two continuous variables are being considered, so a simple labeled XY plot is sufficient for the drawing interface. However other visual encodings and data types require different drawing interactions and interfaces. We envision how EV can be applied to other tasks and encodings in Table 1.

2.2 Extension to Mark Attributes

Visualizations are composed of marks (e.g., a bar in a bar chart, a circle in a scatterplot) and mark attributes (e.g., size, shape, color). While the above example provides interactive support for visualizing one's expectations by adding marks, it may also be possible to support interactive prediction of a variable, which will be encoded as a mark attribute. For example in a choropleth map, a user brush on color to predict the trend in a region for a continuous variable. In these examples, marks are presented and the user interacts to add value expectations (Table 2).

* yeaseul1@uw.edu, jhullman@uw.edu

Electronic proceedings of the IEEE VIS 2015 workshop
Personal Visualization: Exploring Data in Everyday Life

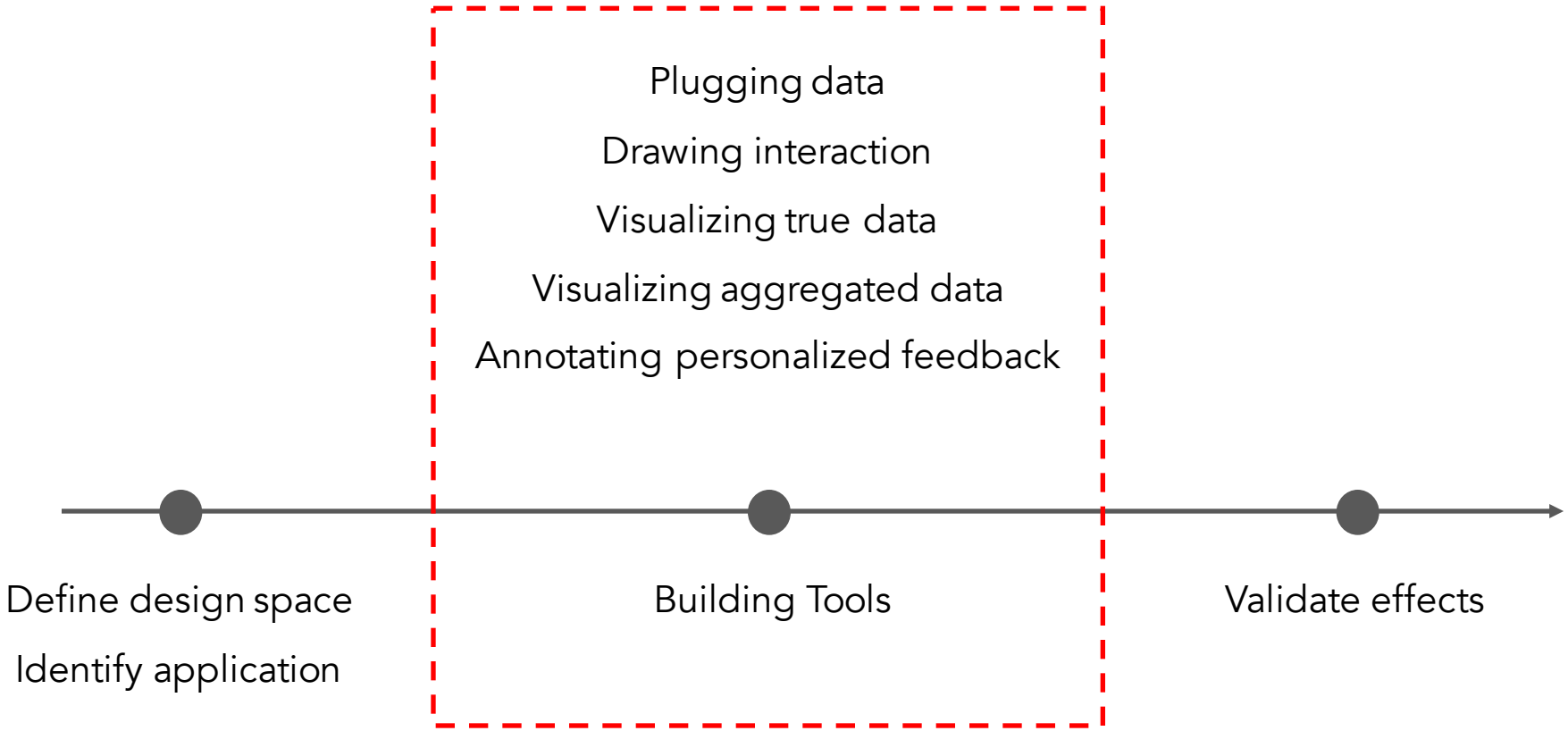
INFOVIS 15 Workshop Paper

Described the design space of EV

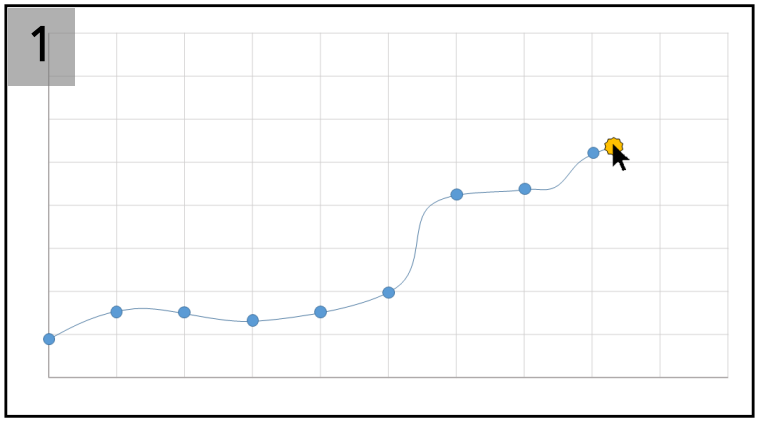
Proposed applications

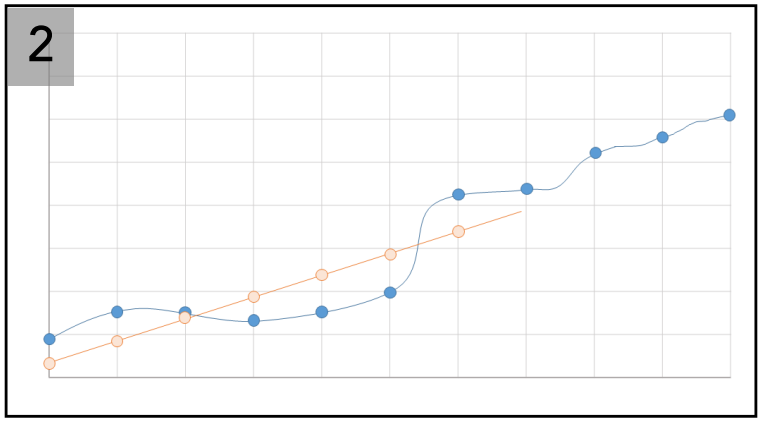
Suggested ideas for future research.

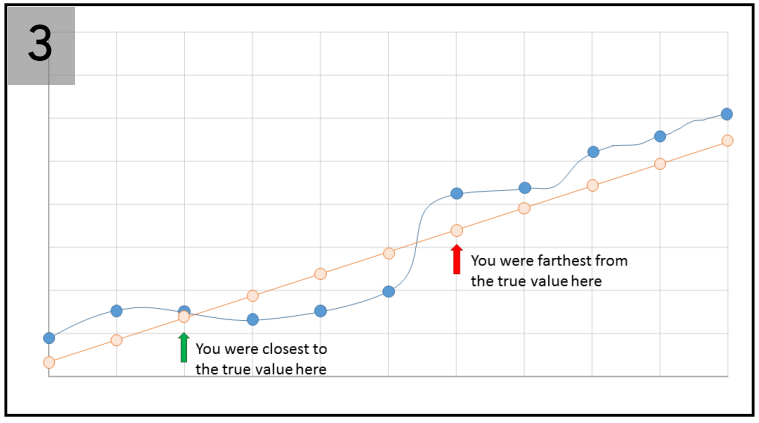
Scope

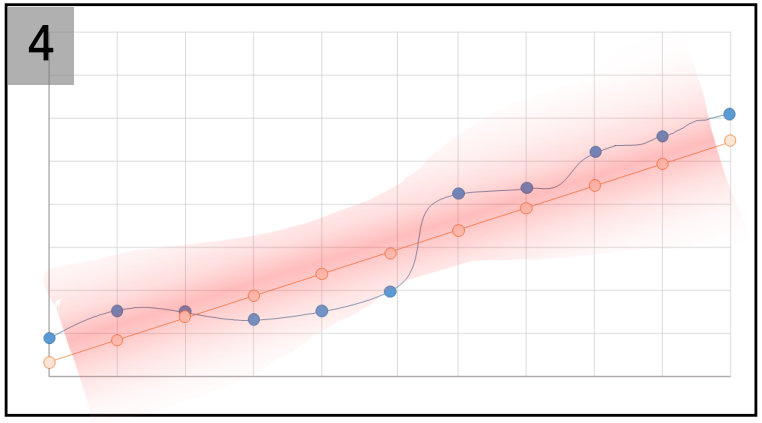

Plugging data
Drawing interaction
Visualizing true data
Visualizing aggregated data
Annotating personalized feedback
Define design space
Identify application
Building Tools
Validate effects

Storyboard – Line Chart

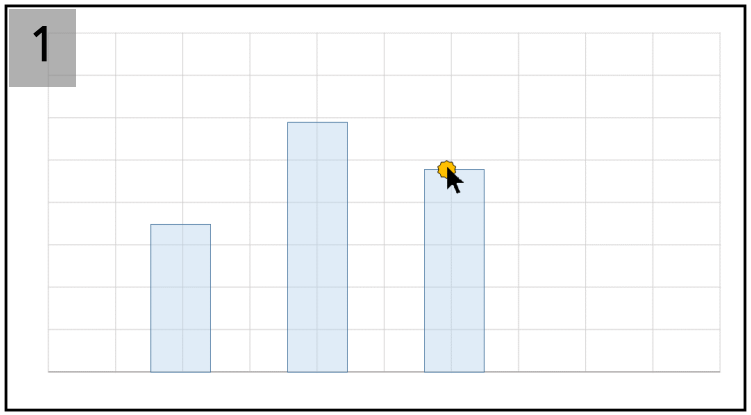


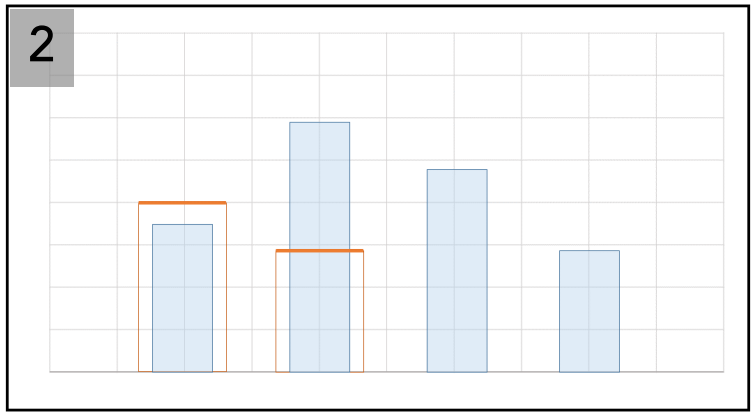


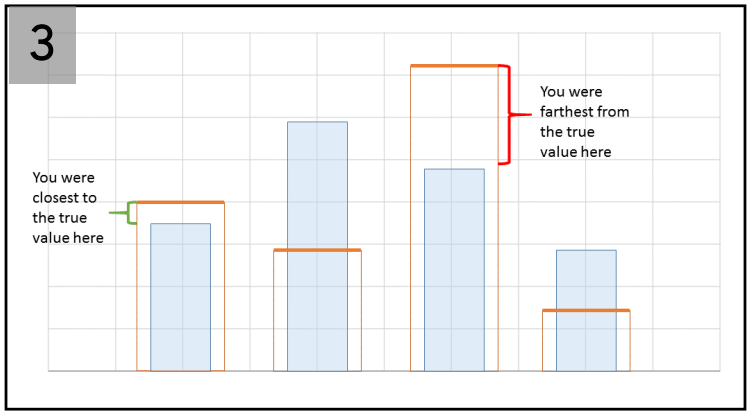


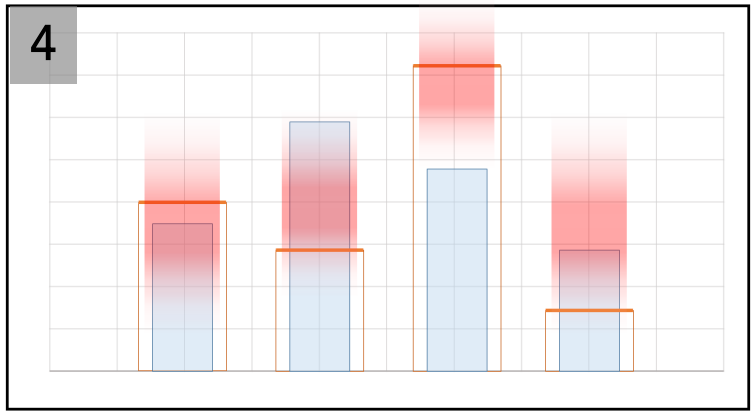


Storyboard – Bar Chart

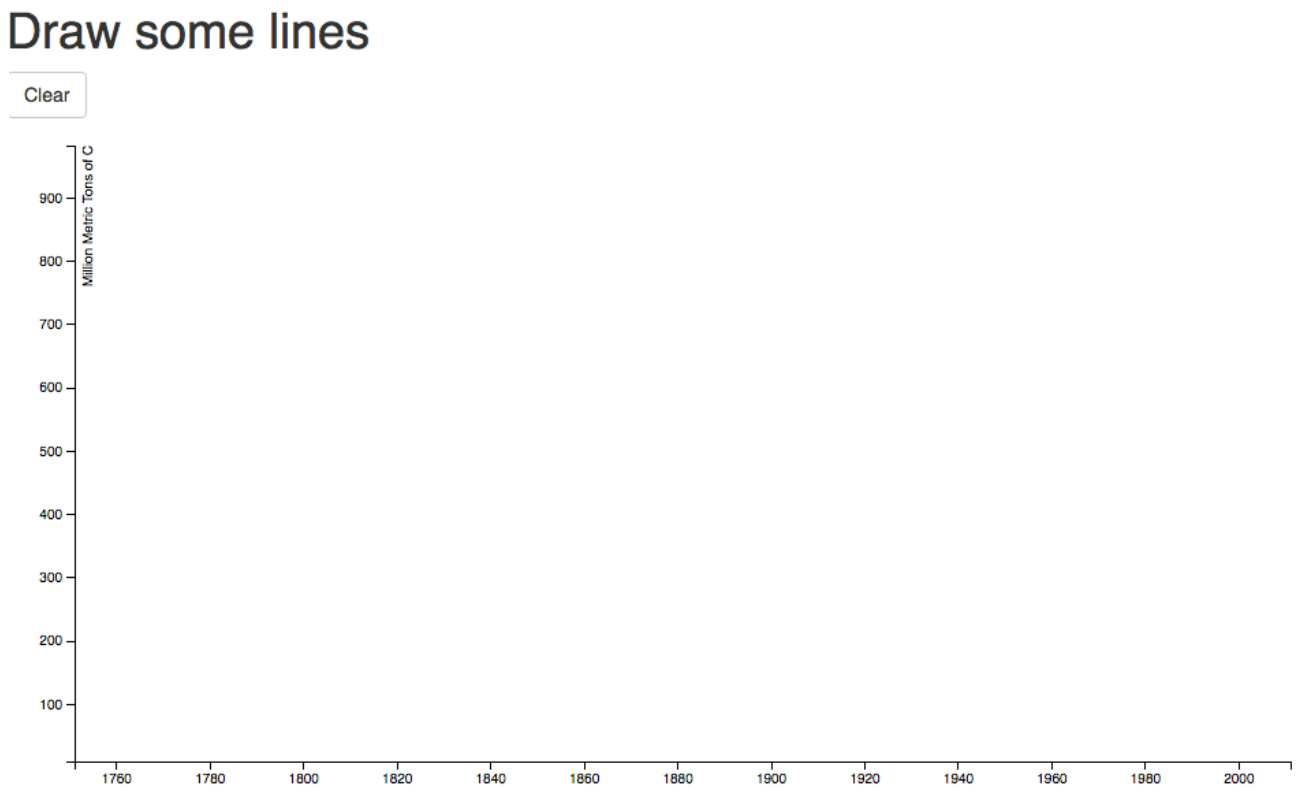




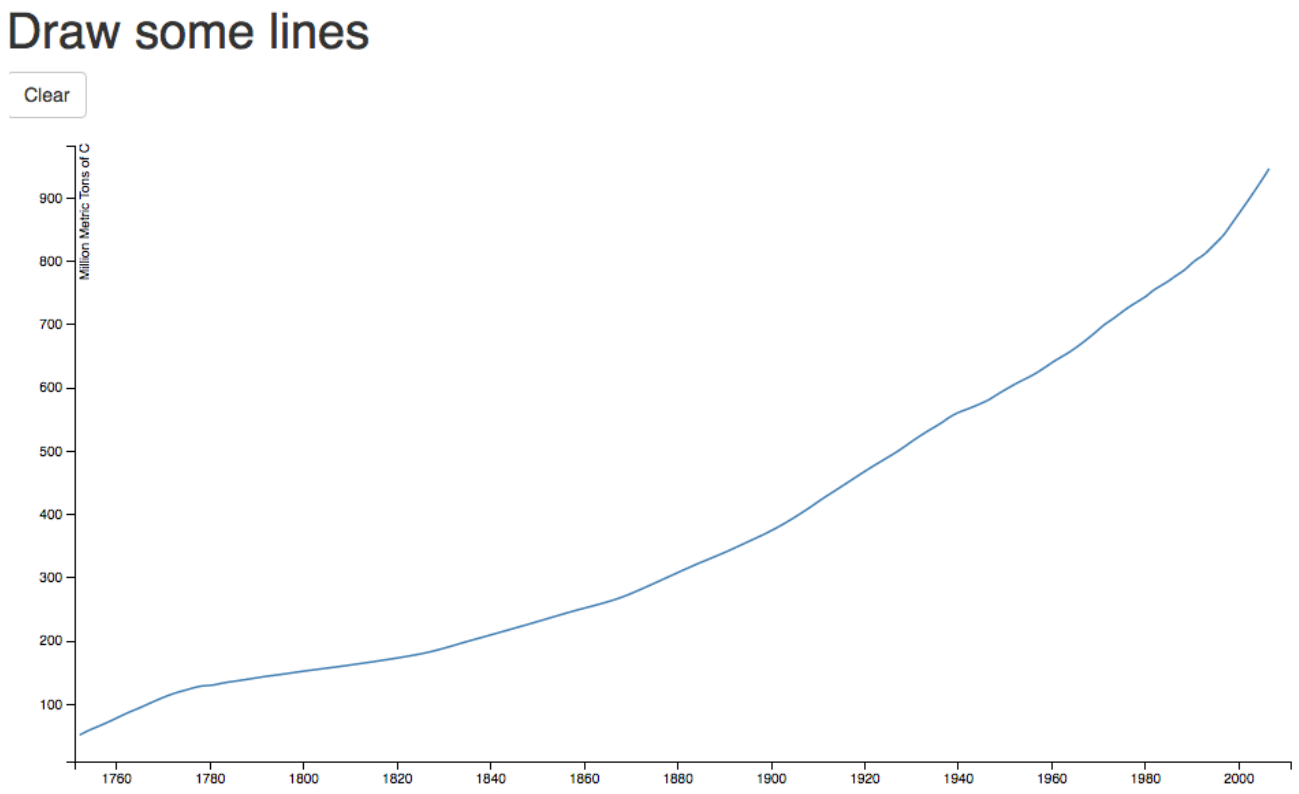




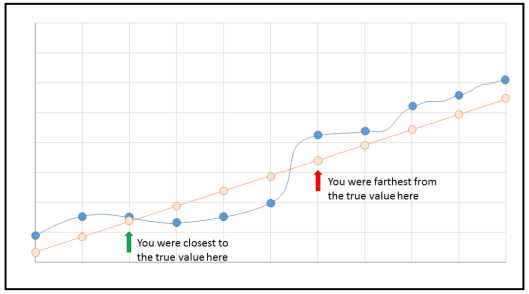
Prototype

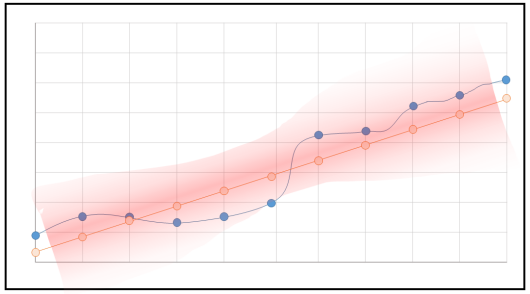

Draw some lines
Clear
Million Metric Tons of C
900
800
700
600
500
400
300
200
100
1760
1780
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000

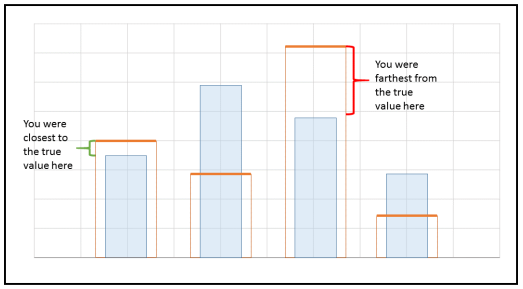
Prototype

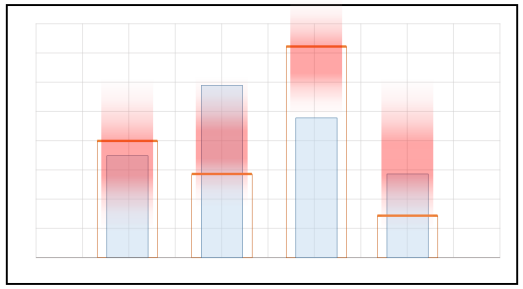

Draw some lines
Clear
Million Metric Tons of C
900
800
700
600
500
400
300
200
100
1760
1780
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000

Questions









How to visualize aggregated data?

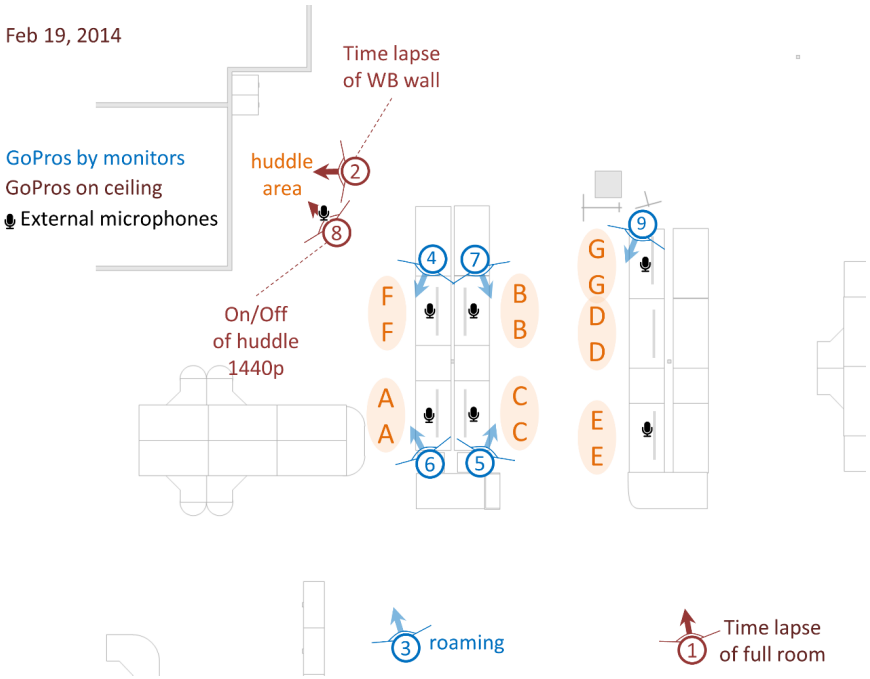
How to annotate personalized feedback?

Spatial-Temporal Visualization of Interactions for Qualitative Analysis

Fida AlSughayer

CSE512 – Data Visualization

Final Project



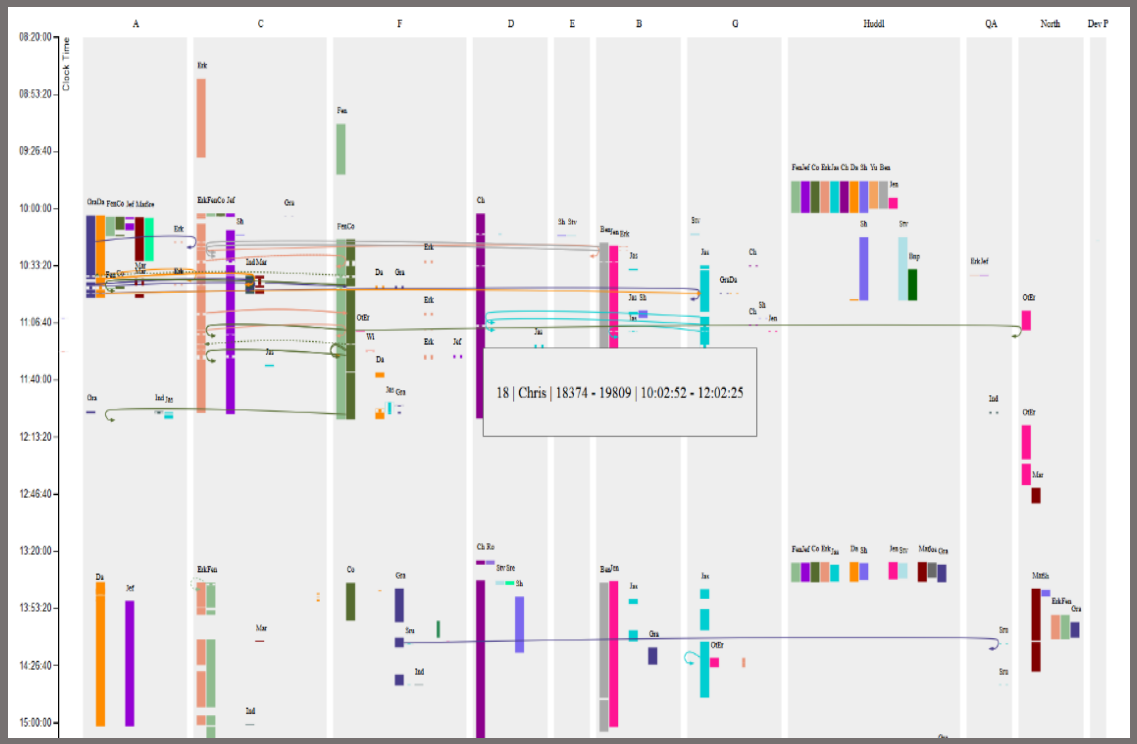
Dataset

- Wide Field Ethnography (WFE) Research team.
- 6 terabyte dataset of video, audio, photos, and screen capture, transcripts and logs.
- Software Engineers at work.

Problem

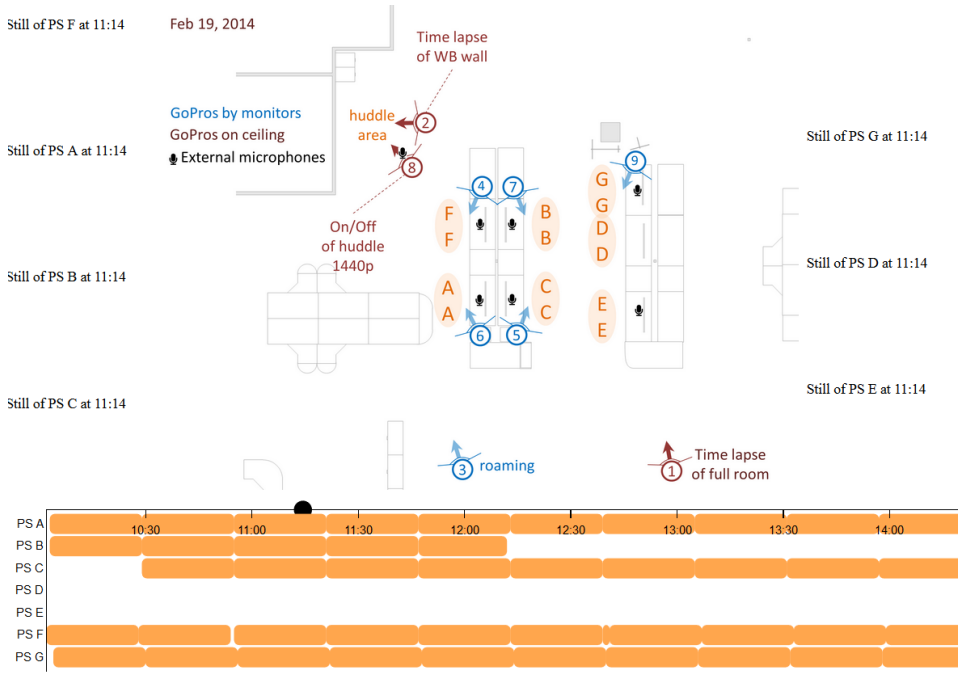
- Pair-programming **interactions** and **interruptions** [based on data logs extracted from videos].
- The **spatial and temporal** distribution of cameras and video feeds [based on video files metadata].

Past Solutions



Scope & Progress

- Motivated by loops of **exploration** (search focused on discovery) and **investigation** (search focused on a targeted query).
- It has a great need for **context** of data, rather than isolated statistics.



For your feedback:

1- Comments on the effectiveness of heatmaps with connection to a timeline.
2- Suggestions on the transition between different views of interaction in a Focus+Content manner.
3- References to relevant visualizations in the realm of: workplace interaction, multimedia data viewing, temporal view of heatmaps,
4- References to research groups on campus who are using visualization for qualitative analysis.

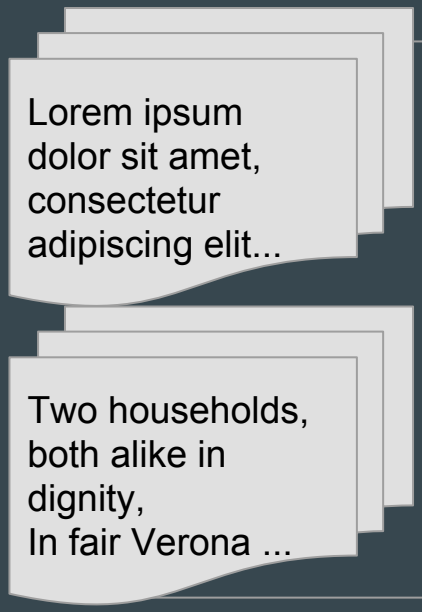
Exploring Frame-Annotated Documents

...

Elizabeth, Lucy, George & Maarten

Motivation

- NLP for social science applications
- Text documents annotated with various labels
- Documents span various other dimensions



	Latin or English?	Year	*More labels* ...
Lorem ipsum dolor sit amet, consectetur adipiscing elit...	Latin	1500s	...
Two households, both alike in dignity, In fair Verona ...	English	1590s	...

The Media Frames Corpus (Card et al. 2015)

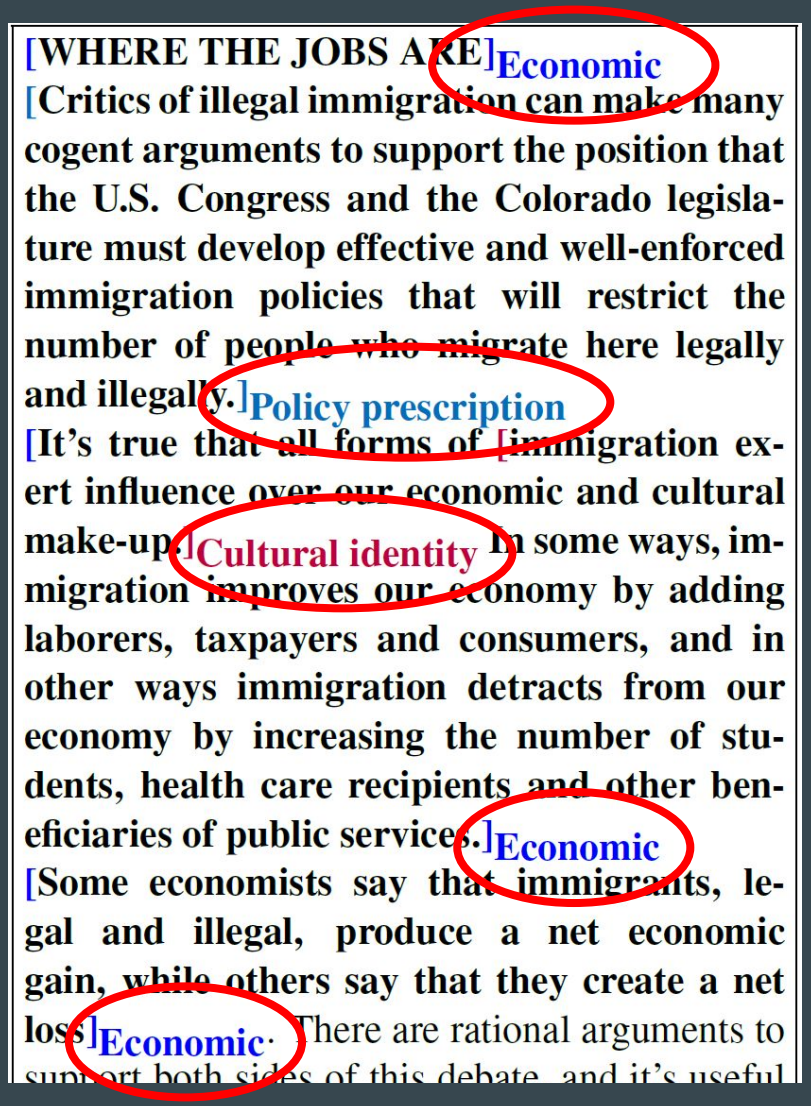
- News articles about
  - Marriage Equality
  - Smoking
  - Immigration
- 15 “framing dimensions”
  - Economic, moral, quality of life, etc.

The way issues are framed influences the reader’s opinion

Economic: costs, benefits, or other financial implications
Capacity and resources: availability of physical, human or financial resources, and capacity of current systems
Morality: religious or ethical implications
Fairness and equality: balance or distribution of rights, responsibilities, and resources
Legality, constitutionality and jurisprudence: rights, freedoms, and authority of individuals, corporations, and government
Policy prescription and evaluation: discussion of specific policies aimed at addressing problems
Crime and punishment: effectiveness and implications of laws and their enforcement
Security and defense: threats to welfare of the individual, community, or nation
Health and safety: health care, sanitation, public safety
Quality of life: threats and opportunities for the individual’s wealth, happiness, and well-being
Cultural identity: traditions, customs, or values of a social group in relation to a policy issue
Public opinion: attitudes and opinions of the general public, including polling and demographics
Political: considerations related to politics and politicians, including lobbying, elections, and attempts to sway voters
External regulation and reputation: international reputation or foreign policy of the U.S.
Other: any coherent group of frames not covered by the above categories

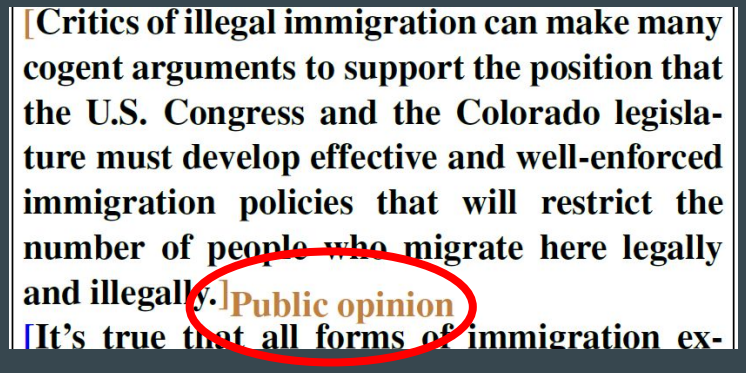
What the data looks like

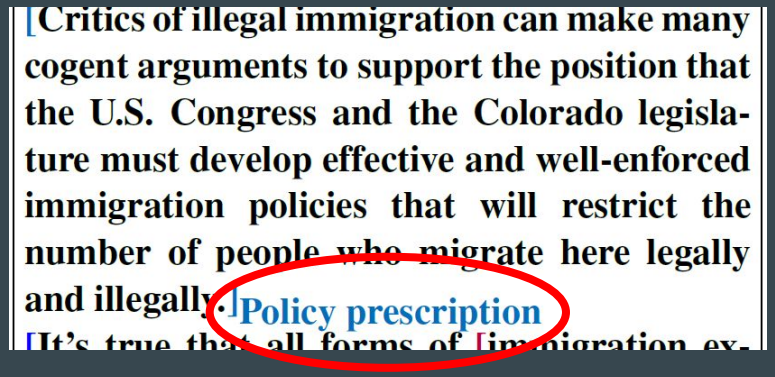
- Text spans are annotated



What the data looks like

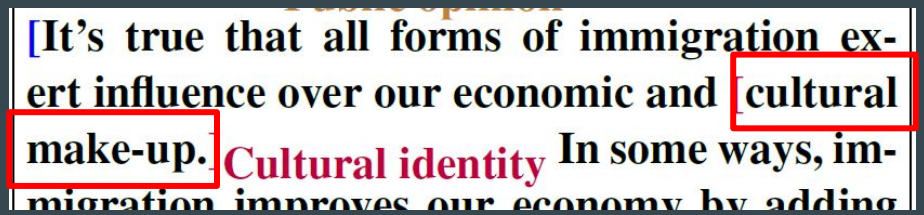
- Text spans are annotated
- Annotators don't always agree

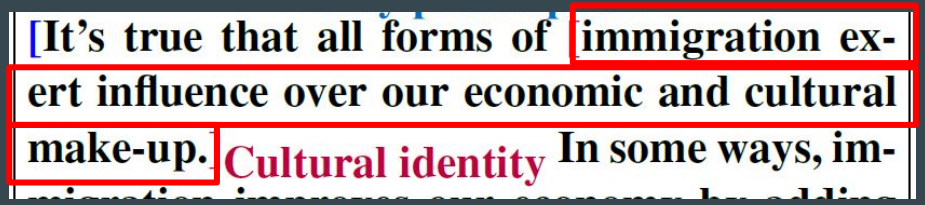




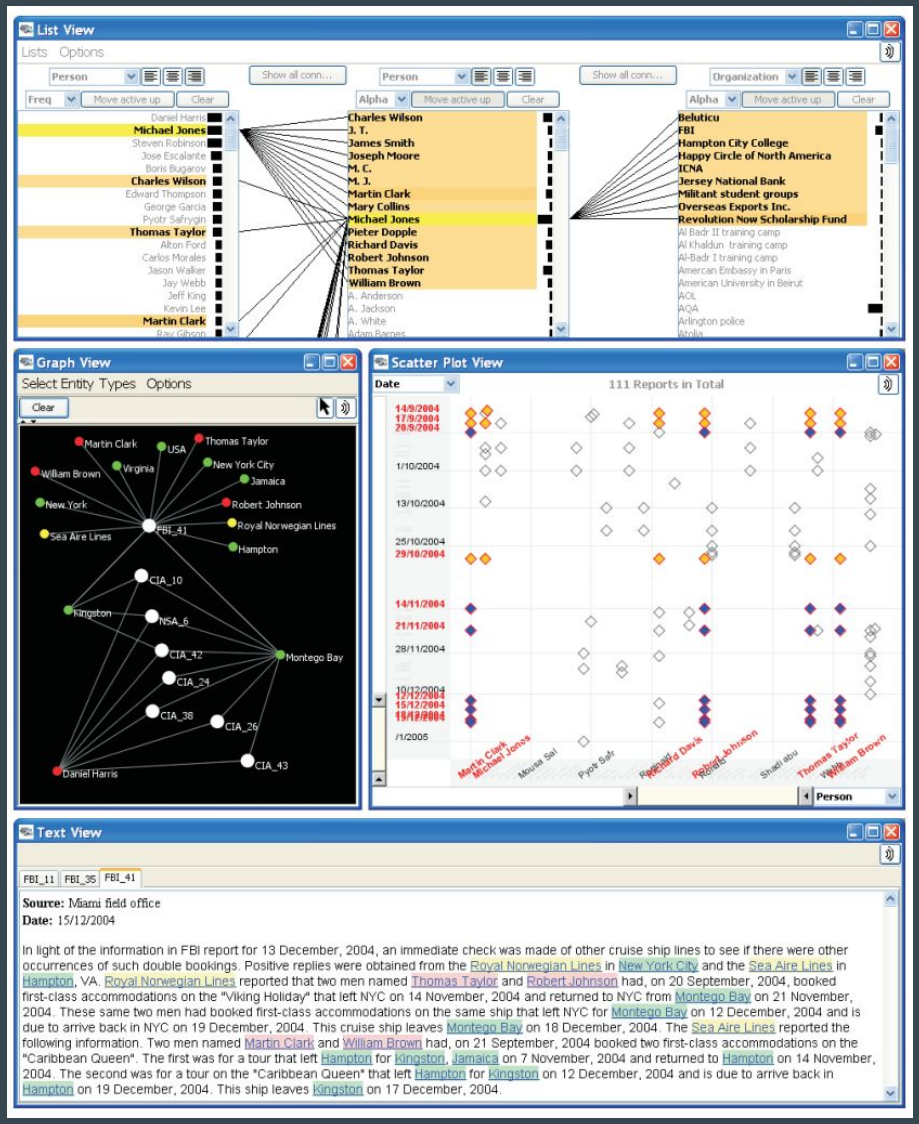
What the data looks like

- Text spans are annotated
- Annotators don't always agree
- Spans aren't predefined



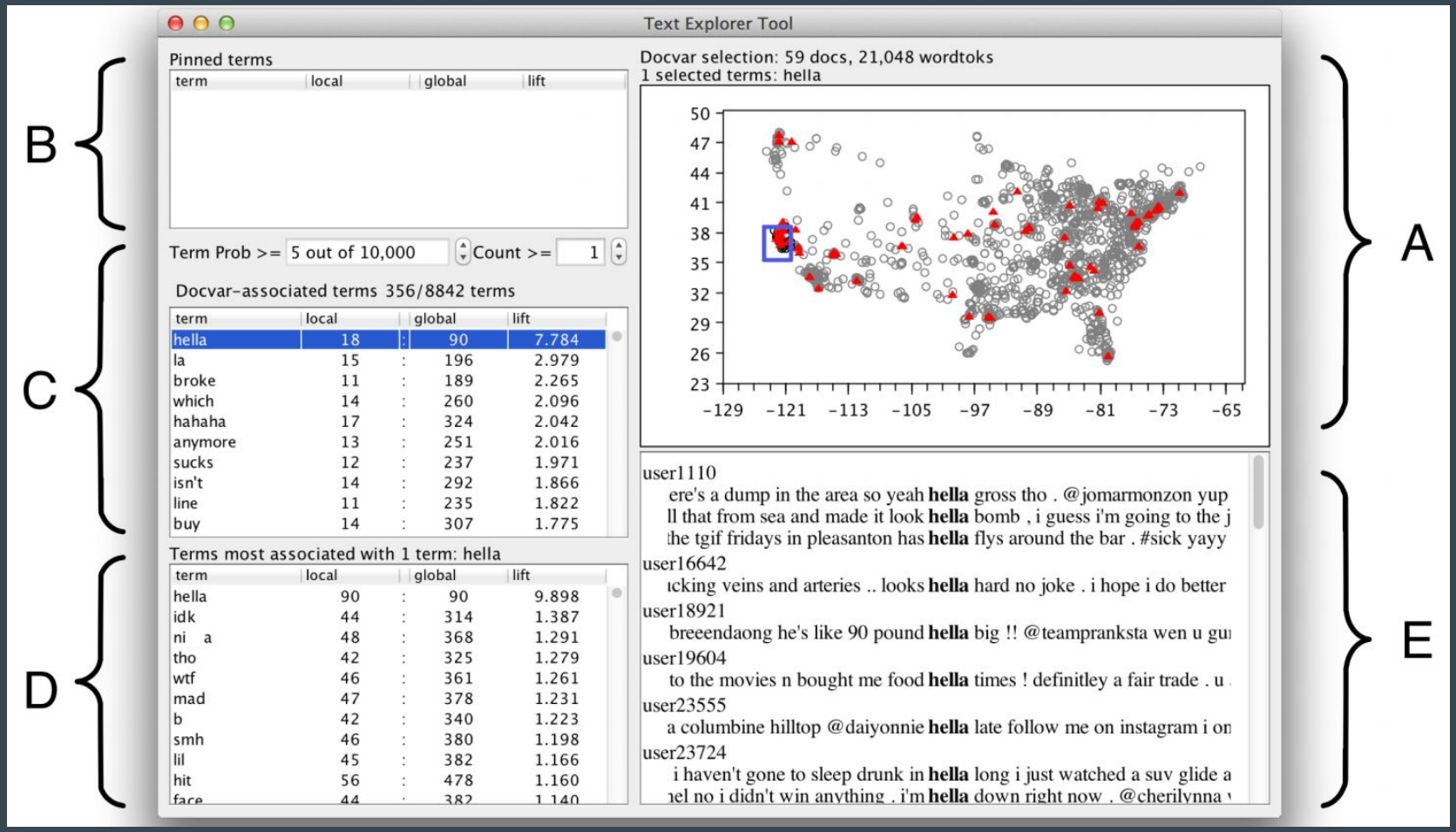


Related Work: Jigsaw



http://www.cc.gatech.edu/~stasko/papers/vast07-jigsaw.pdf

Related Work: MITextExplorer



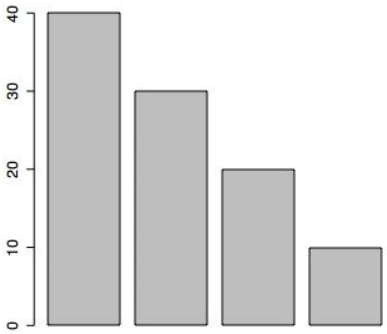
http://nlp.stanford.edu/events/illvi2014/papers/oconnor-illvi2014.pdf

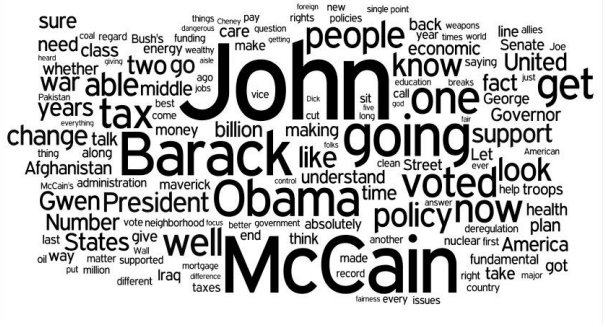
Related Work - What we want to do differently

- Streamlined version of some tools
- Find interesting overall patterns and still have the raw data
- Target audience: social science collaborators, somewhere in between a lay person and an analyst

Our proposed tool

Explore | About: corpus | About: app



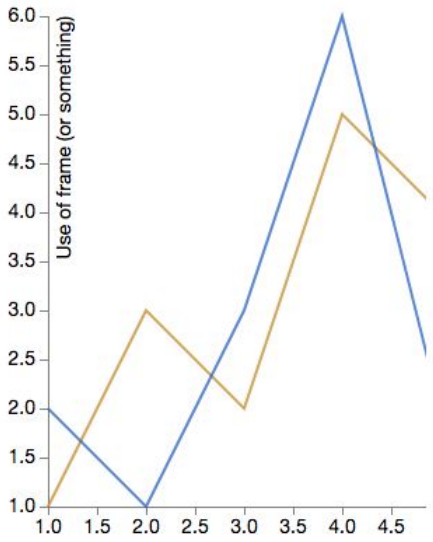


[WHERE THE JOBS ARE]Economic
[Critics of illegal immigration can make many cogent arguments to support the position that the U.S. Congress and the Colorado legislature must develop effective and well-enforced immigration policies that will restrict the number of people who migrate here legally and illegally.]Policy prescription
[It's true that all forms of [immigration exert influence over our economic and cultural make-up.]Cultural identity In some ways, immigration improves our economy by adding laborers, taxpayers and consumers, and in other ways immigration detracts from our economy by increasing the number of students, health care recipients and other beneficiaries of public services.]Economic
[Some economists say that immigrants, legal and illegal, produce a net economic gain, while others say that they create a net loss]Economic. There are rational arguments to support both sides of this debate, and it's useful and educational to hear the varying positions.

Prototype

Explore | About this corpus | About this app

Interactive things here! (part 1)



Interactive things here! (part 2)

[WHERE THE JOBS ARE]

[Critics of illegal immigration can make many cogent arguments to support the position that the U.S. Congress and the Colorado legislature must develop effective and well-enforced immigration policies that will restrict the number of people who migrate here legally and illegally.]

[It's true that all forms of [immigration exert influence over our economic and cultural make-up.] In some ways, immigration improves our economy by adding laborers, taxpayers and consumers, and in other ways immigration detracts from our economy by increasing the number of students, health care recipients and other beneficiaries of public services.]

[Some economists say that immigrants, legal and illegal, produce a net economic gain, while others say that they create a net loss. There are rational arguments to support both sides of this debate, and it's useful and educational to hear the varying positions] .

Annotator:

(•) A ( ) B

We learned from A3 that...

- Exploratory tools are good, but ...
- Better when it's guided exploration
- Our solution:
  - A focused landing page (e.g. one result)
  - "About this corpus" tab
  - "About this tool" tab

Questions

- Would this tool be something you could use?
- Do you think this tool would be helpful for solving the issue of having to make sense of an abundance of text data?
- What aggregate statistics over the entire corpus would be helpful for exploration?
- Any of you who worked with text in A2/A3, do you have any words of wisdom?

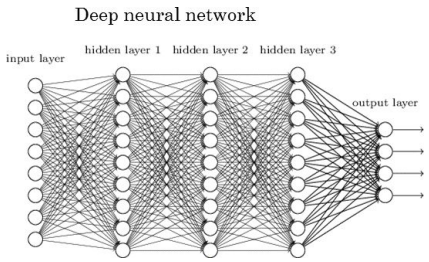
Data Exploration for Feature Engineering

Jin Qu, Christopher Clark, Mohit Jain, Gagan Bansal
University of Washington

Feature Engineering

Transforming ‘raw’ data (such as documents, click logs, raw sensor data) into feature vectors that can be used in machine learning algorithms

Often of the main focus of machine learning practitioners in industry



Visualization for Featuring Engineering

One the feature vectors have been built, they can be loaded into most existing data visualization tools

But current tools provide no way to ‘look back’ to see the sources that were used to build the feature vectors

Makes these tools practically useless for feature engineering

We need a way to integrate source documents into data exploration tools

Prior Work

Several Machine Learning + visualization research papers:

Human-guided ML, Labeling data, Interactive ML, ML performance analysis tools, etc.

Feature engineering + Visualization paper:

Only for text classification [1]

Does not link back to the raw data [2]

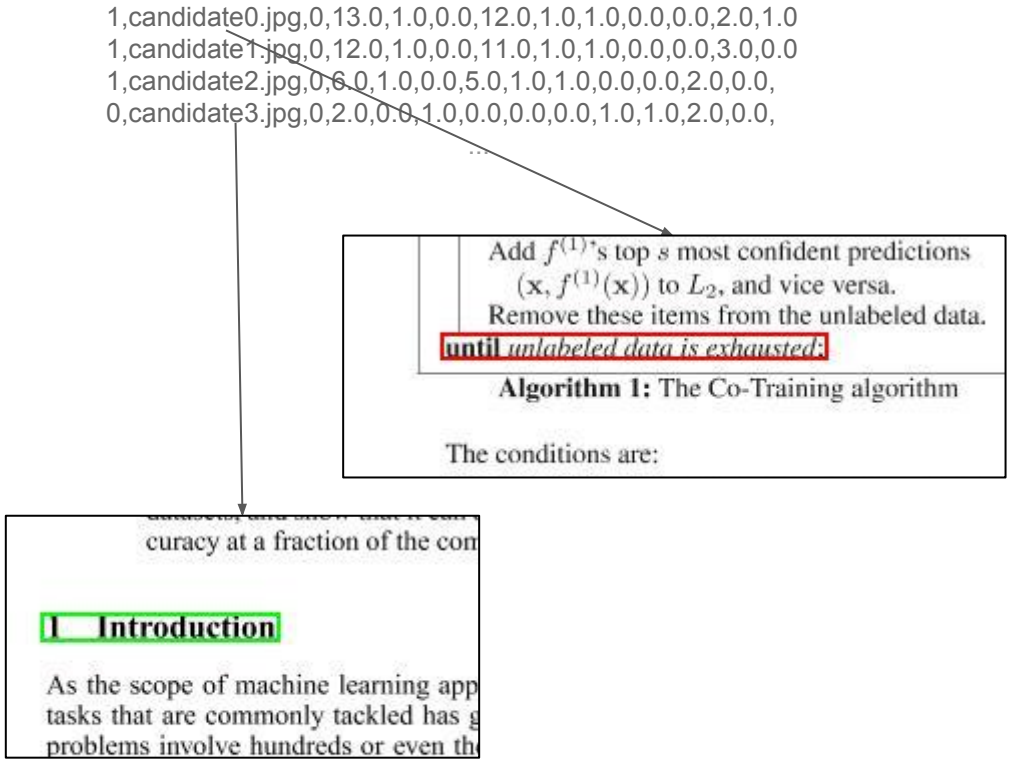
[1] FeatureInsight: Visual Support for Error-Driven Feature Ideation in Text Classification. Michael Brooks, Saleema Amershi, Bongshin Lee, Steven M. Drucker, Ashish Kapoor, Patrice Simard. IEEE VAST 2015.
[2] FeatureForge: A Novel Tool for Visually Supported Feature Engineering and Corpus Revision. Florian Heimerl, Charles Jochim, Steffen Koch, Thomas Ertl. COLING 2012.

Example: Document Layout Analysis

Attempt to identify section titles within scholarly articles

Images of candidate section titles can be rendered offline

Rendered images and features can be loaded into our tool (through a webpage) and explored

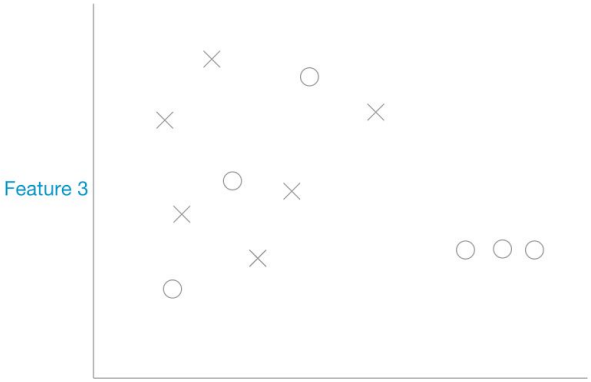


/Users/bob/Desktop/features.csv | UPLOAD

FEATURE SELECTOR

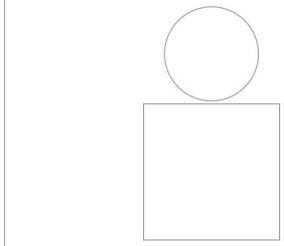
- [x] Feature 1
- [ ] Feature 2
- [x] Feature 3
- [ ] Feature 4
- [ ] Feature 5
- [ ] Feature 6
- [ ] Feature 7
- [ ] Feature 8
- [ ] Feature 9

ERROR VIEWER

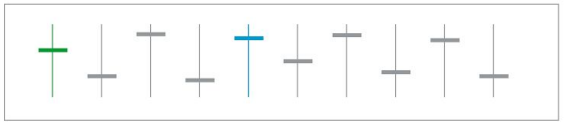


SOURCE VIEWER

/Users/bob/Desktop/data/sample1.png



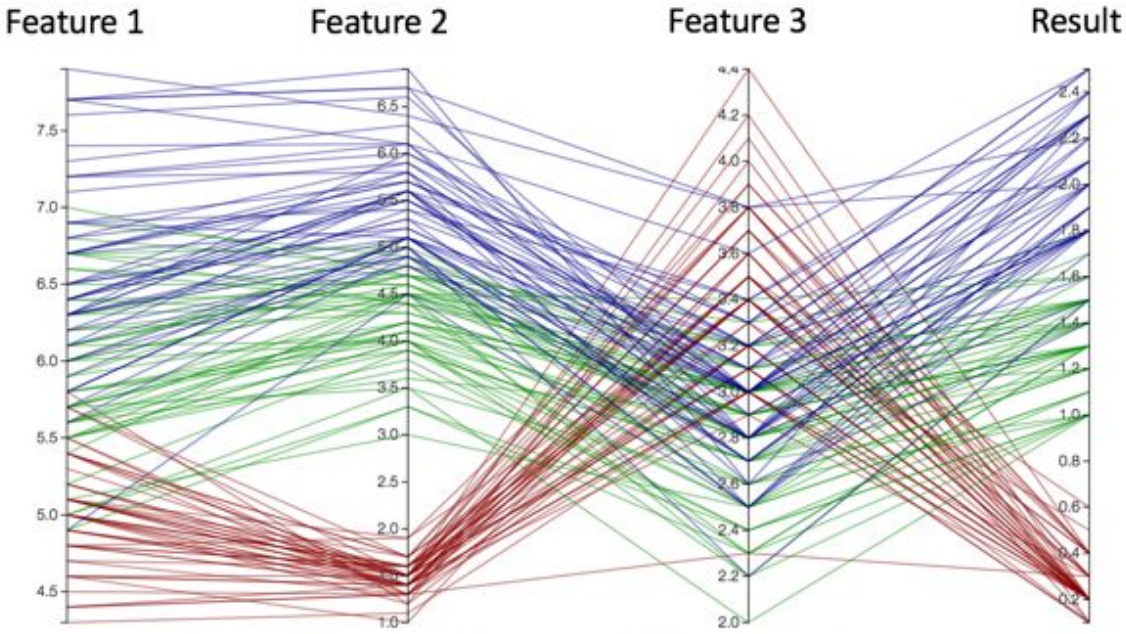
FEATURE SUMMARIZER



New Tab

Feature List

- Feature 1 ☑
- Feature 2 ☑
- Feature 3 ☑
- Feature 4 ☑
- Feature 5 ☐
- Feature 6 ☑
- Feature 7 ☑
- Feature 8 ☐



Questions

Feature engineering visualization: Scatter plot? Parallel coordinates? Etc?

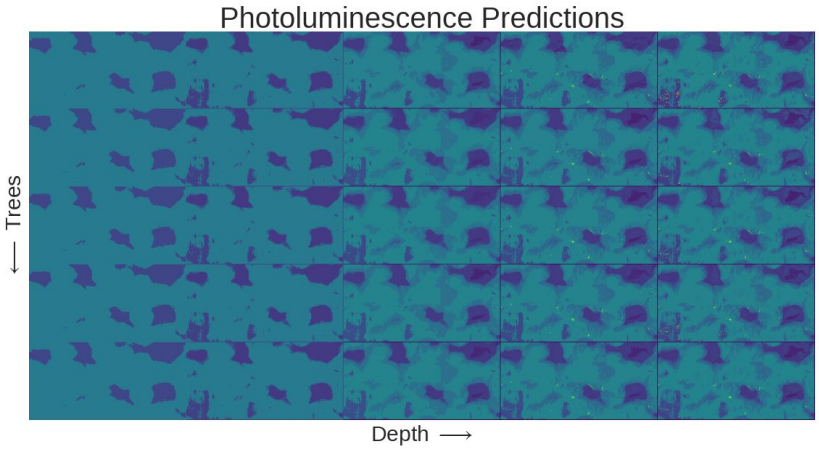
Different source viewer: Image? Pdf document?

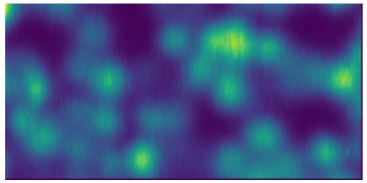
Other kinds of sources that would be useful: Json? Tables?

Teaching ML for Image Generation

Goal: Educational tool for machine learning

Example case: prediction of photoluminescence data from atomic force microscopy data using decision trees.



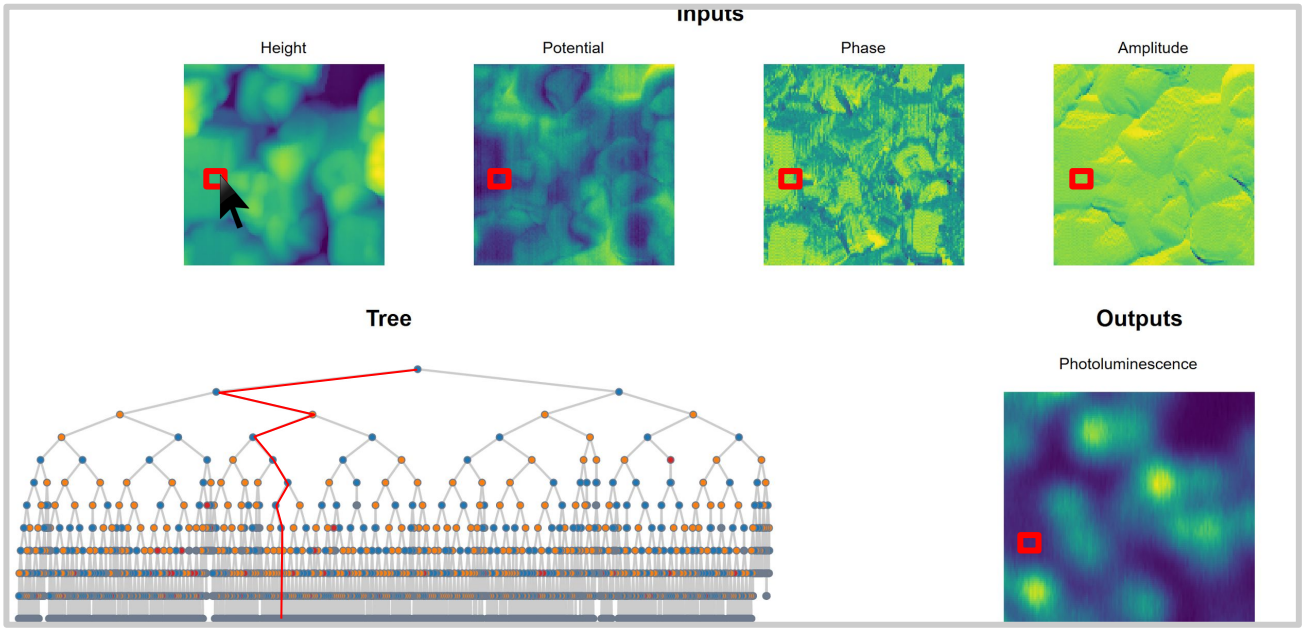


Potential tuning knobs to provide users:

- Tree depth
- # of decision trees
- # of surrounding pixels
- train/test split

Simplest Version of Design

Users can select a pixel and see how the decision tree predicted the output value



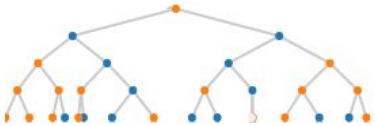
To add: user controls for model hyperparameters

Tree used to predict resulting pixel value

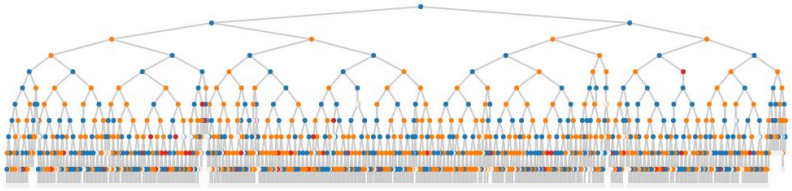
Prediction of pixel and surrounding pixels

Potential User Controls

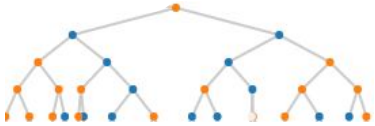
Tree depth



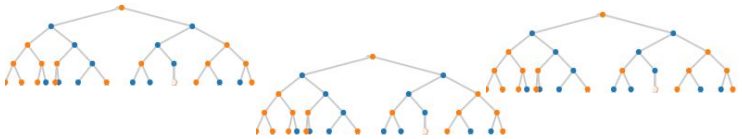
vs



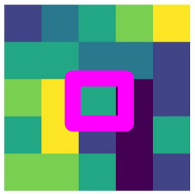
Number of trees



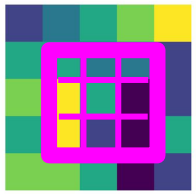
vs



# of surrounding pixels



vs

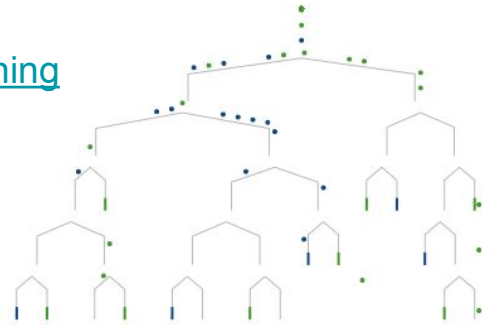


Train/validation split (under/over-fitting)

Prior (similar) work

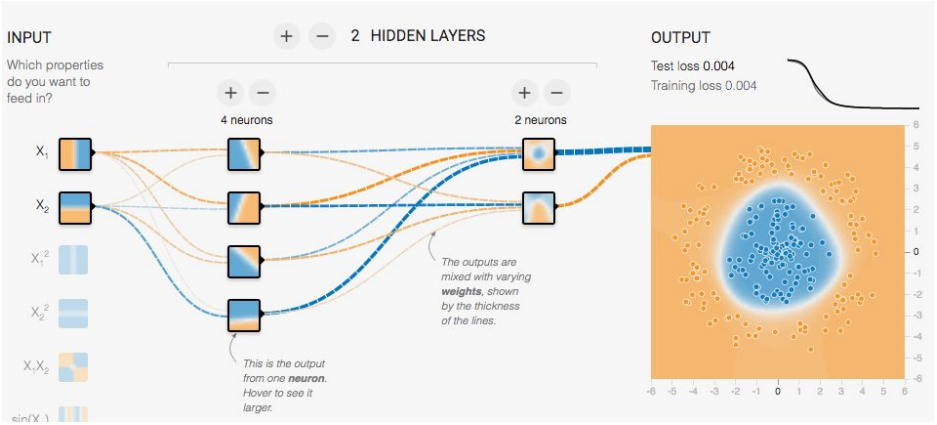
[Visual Introduction to Machine Learning]

(classification w/ decision tree)



[Tensorflow]

(neural networks)

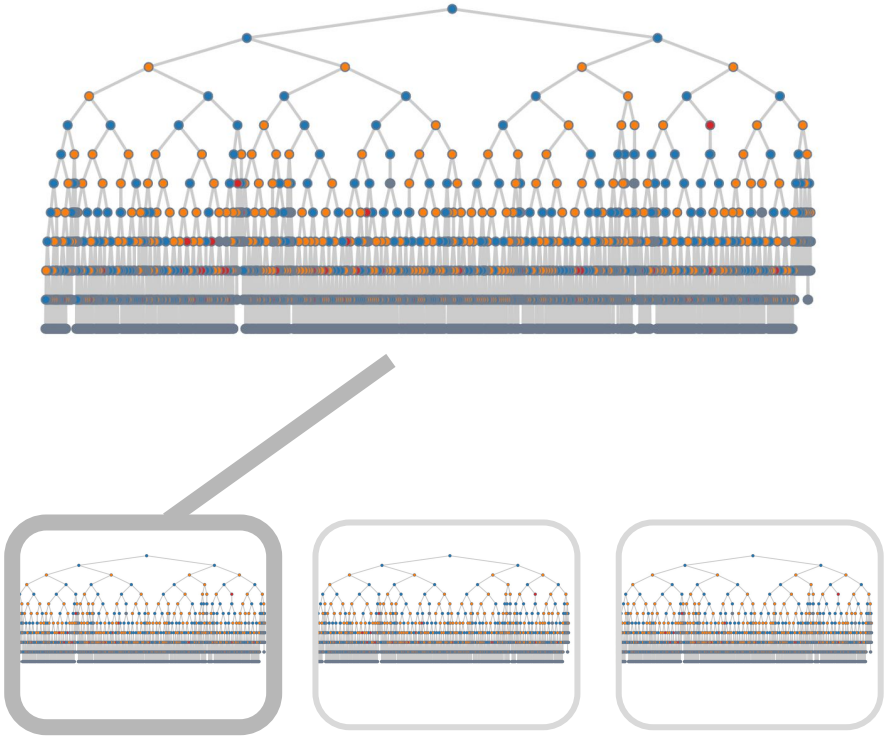


Our tool's differences:

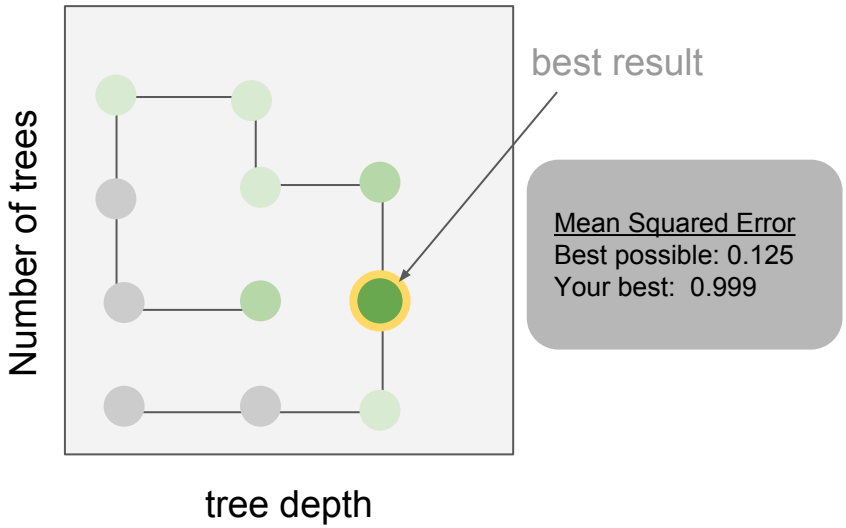
- continuous output
- hyperparameters' effect on prediction quality
- generation of an image, rather than classes

Design Addition Ideas

"carousel" of trees in forest



graphical history of explored space



Questions

- How to visualize trees with depth > 5 (crowding)
- Focus options:
  - decisions made at each tree node, given a pixel
  - prediction quality
  - "Forest" of trees
- How many tuning knobs to provide:
  - Tree depth
  - Number of trees
  - Number of surrounding pixels
- How to display "best score" and space explored.
  - If 2 nobs, can explore a 2D space
- Convey relative importance of each image?
  - What metric of importance?

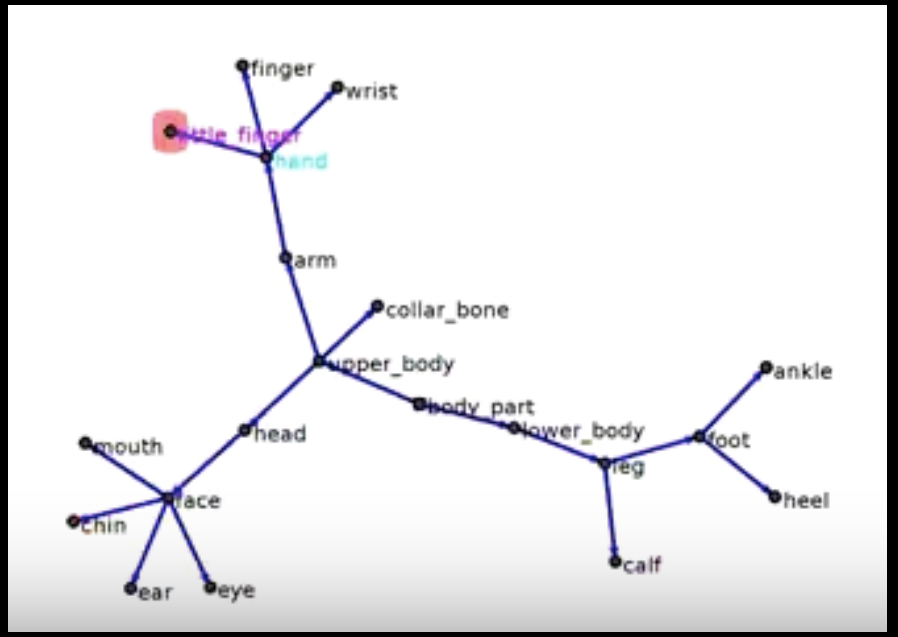
Interactive learning for hierarchy of concepts

Adwin Jahn, Ryan Maas, Harley Montgomery

Concepts

- System : Question hand -> little finger
- User : Yes
- N^2 /2 questions

MAP solution



Relevant Work and Difference

Crowd Sourcing Version

- questionnaire
- video

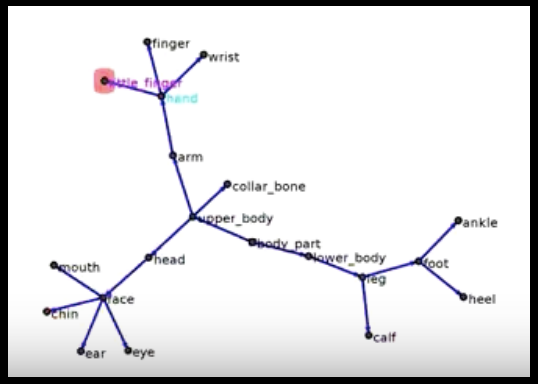
Interactive Learning Version

- visualize the current state to the user
- show the hidden edges
- minimize the difference between sequence

Learn How the Machine Learns

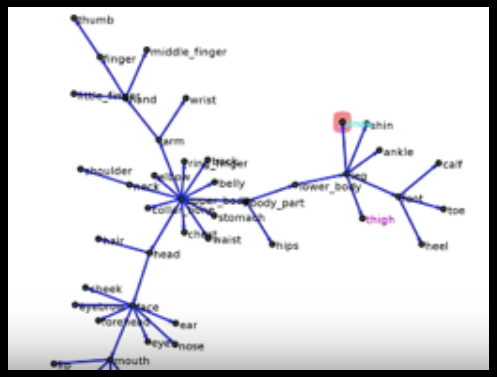
Step 3

- System : Question hand -> little finger
- User : Yes



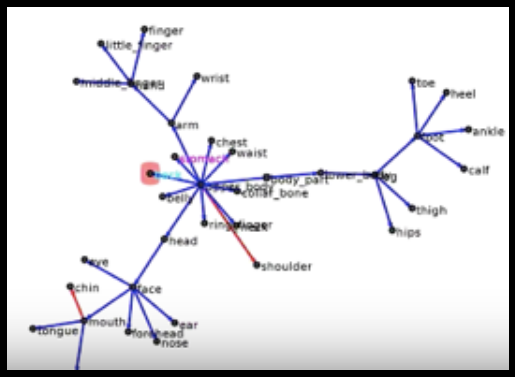
Step 5

- System : Question back -> Stomach
- User : No

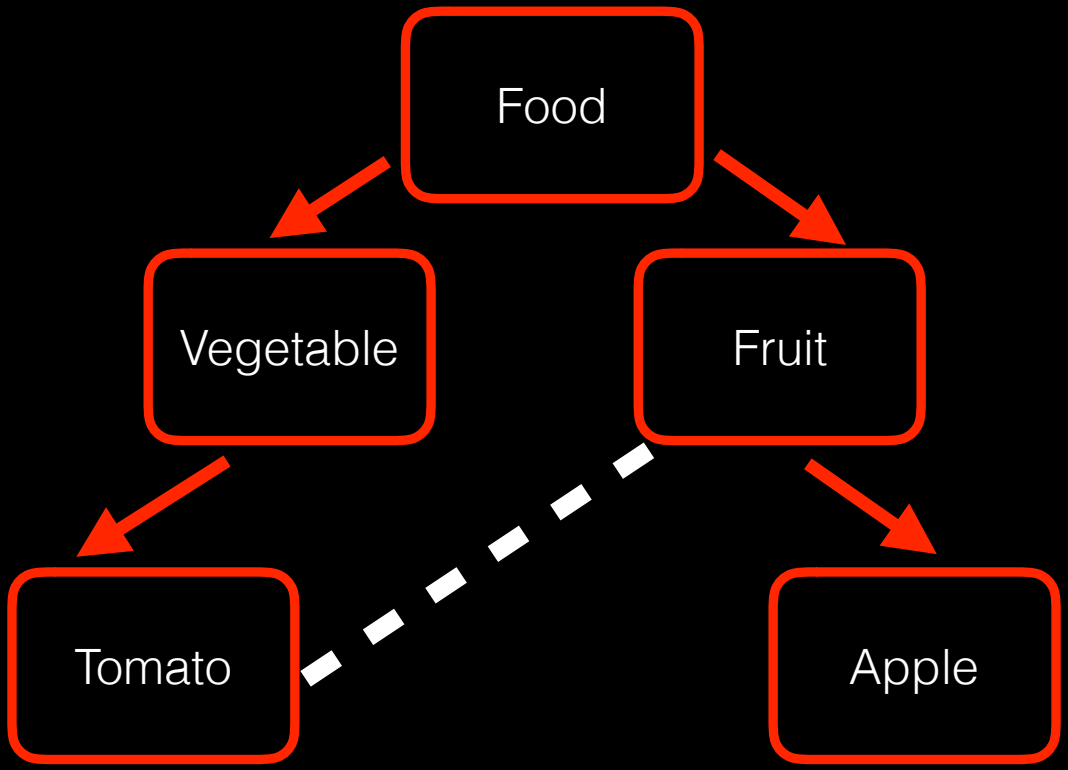


Step 10

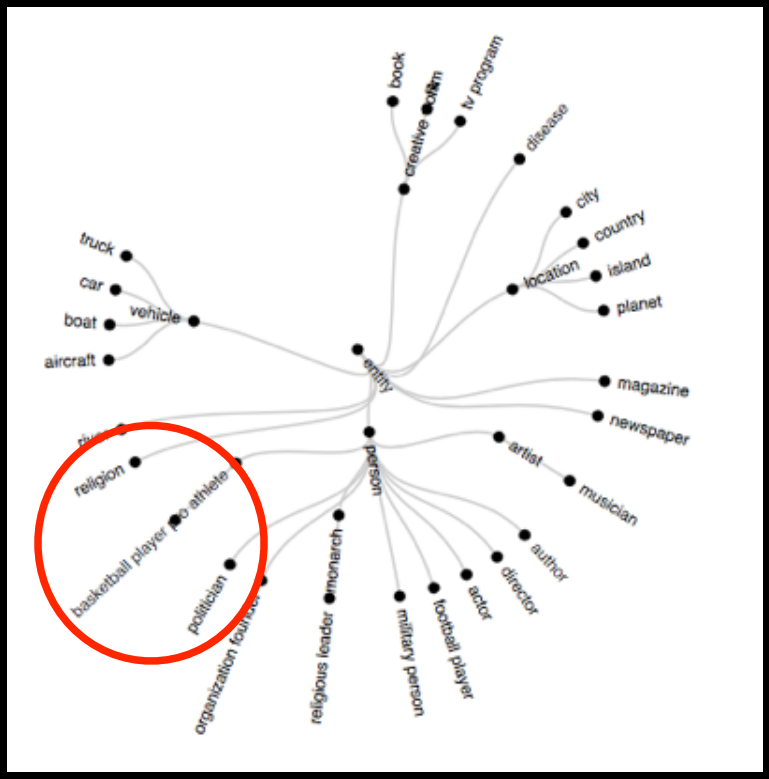
- System : Question knee->heel
- User : No



Hidden Edges



Current Progress and Minimize the difference between sequence



CSE 512 Final Project
Visualizing Climate Change

Group members: Helena van Tol, Shirley Leung, Michelle Weirathmueller, Philippe Vaillant
In collaboration with Dargan Frierson

Description of the problem

- Climate change is one the greatest environmental and communication challenges of our time
  - Multiple sources of variable data
  - Small changes in data (eg. degrees Celsius) constitute major change on a global scale
  - Audience with varying levels of expertise and ability

- Target audience is consumer online content (potentially via Inside Climate News)

- Goal: Create a visualization that truthfully displays multiple sources of variable data in a way that is convincing to the general public

Relevant prior works

- http://climate.nasa.gov/interactives/climate-time-machine
  - Displays gradual changes in global variables over time
- http://climate.nasa.gov/images-of-change
  - Shows before and after satellite imagery using a simple interface, visually impactful

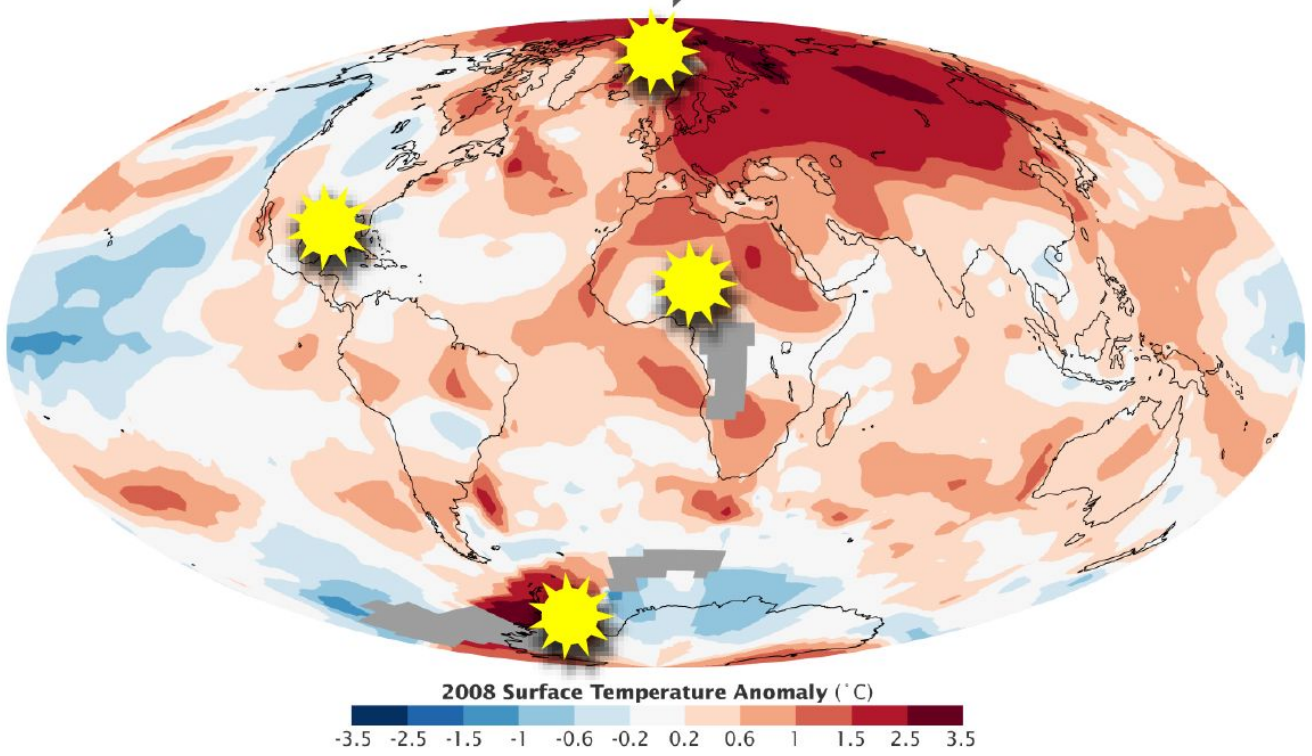
How our work is different

- Past communication efforts have focused on gradual changes
  - **We want to emphasize how dramatic these changes have really been by showing only 2 data points using before and after photographs and data**
- Past communication efforts have tended to obscure HOW we know what we know
  - **We want to emphasize that there is a scientific basis behind the data we show**
- Past communication efforts have focused on pre-satellite era data
  - **We want to emphasize how much climate change has occurred since 1979**

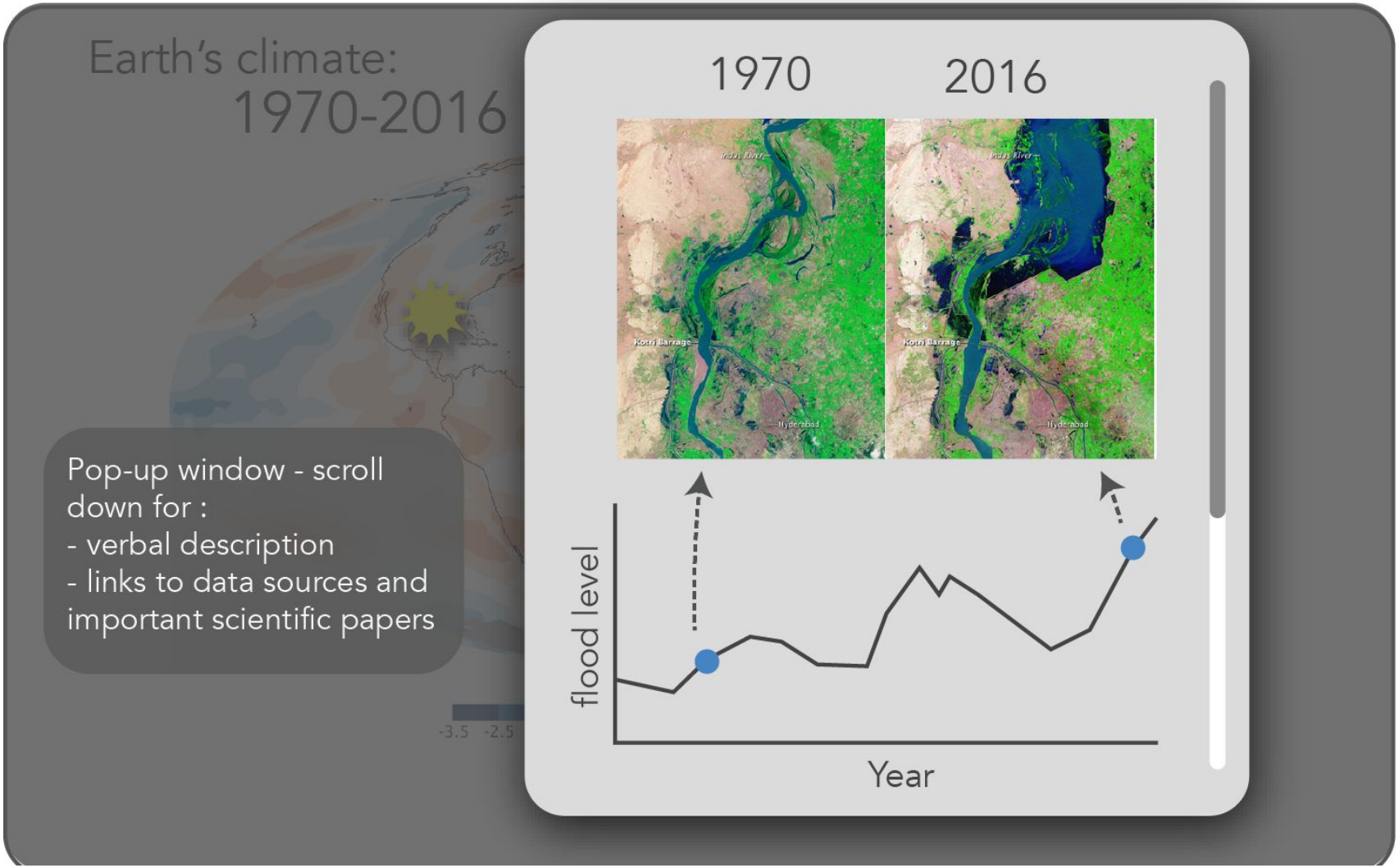
Storyboard

Earth's climate:
1970-2016

click on location icons to pull up before and after images + data



Storyboard


Earth's climate:
1970-2016
-3.5
-2.5
Pop-up window - scroll down for :
- verbal description
- links to data sources and important scientific papers
1970
2016
Indus River
Kotri Barrage
Hyderabad
flood level
Year

Impacts of climate change on peak wildflowers

Rachel Li, Xiaojing Zhu, Guanming Wang

University of Washington

May 18, 2016

Problem Description

- Janneke Hille Ris Lambers from Department of Biology, who is interested in forecasting the impacts of climate change on species distributions, population dynamics, and community structure of wild flowers
- Our project: Impacts of climate change on peak wildflowers
- The goal is to develop a visualization that displays the dynamic change of snowmelt and blooming for each species and locations over the summer season for each year

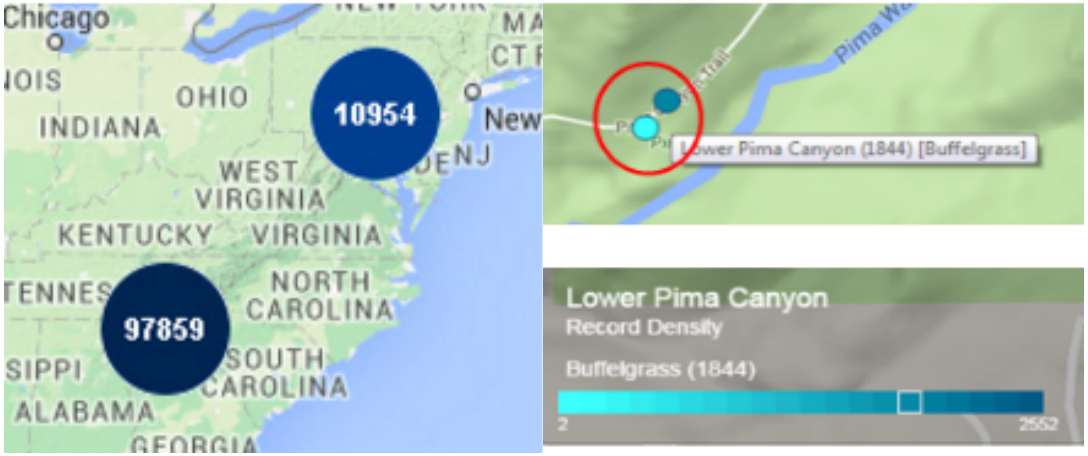
Data Description

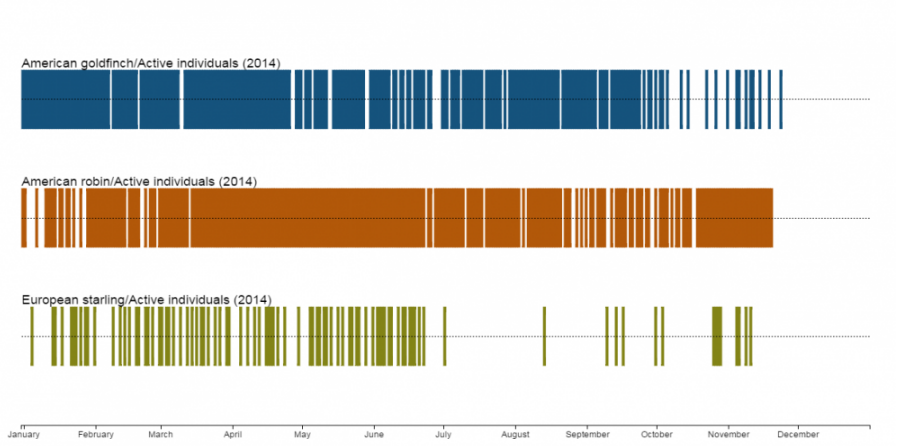
- Data is collected between 2013-2015 by 70+ volunteers along one of the hiking trails in Mt. Rainier.
- The data set contains

Worksheets & Column Definitions		
SiteData: includes information on each site (year and site specific)		
	Year	year of data collection
	Site	site (plot number)
	Elev	elevation (in meters above sea level); doesn't change from year to year
	SnowmeltD	date that snow melted at that site (based on microclimate sensors, personal observation, and models)
	JulDay	Julian day that snow melted (days since January 1 of that year)
PhenoData: data collected by volunteers on phenophases		
	Date	Date data was collected
	Year	year of observation
	Month	month of observation
	Day	day of observation
	JulianDay	Julian day of observation (days since January 1 of that year)
	Site	site (plot number). Note each site is ~ 0.5 by 1 meter large
	Species	species (see below for key to 4 letter codes). Note not all species occur in each plot.
	Flower	1 or 0 - 1 means a flower phenophase of that species was observed, 0 means it was not. It could be many
Species codes		
	ANOC	Anemone occidentalis - western anemona
	CAPA	Castilleija parviflora - magenta paintbrush
	ERMO	Erythronium montanum - avalanche lily
	ERPE	Erigeron peregrinus - mountain daisy
	LIGR	Ligusticum grayi - grays lovage
	LUAR	Lupinus arcticus - subalpine lupine
	MIAL	Microseris alpestis - north microseris
	PEBR	Pedicularis bracteosa - bracted lousewort
	POBI	Polygonum bistortoides - American bistort
	VASI	Valeriana sitchensis - Sitka valerian

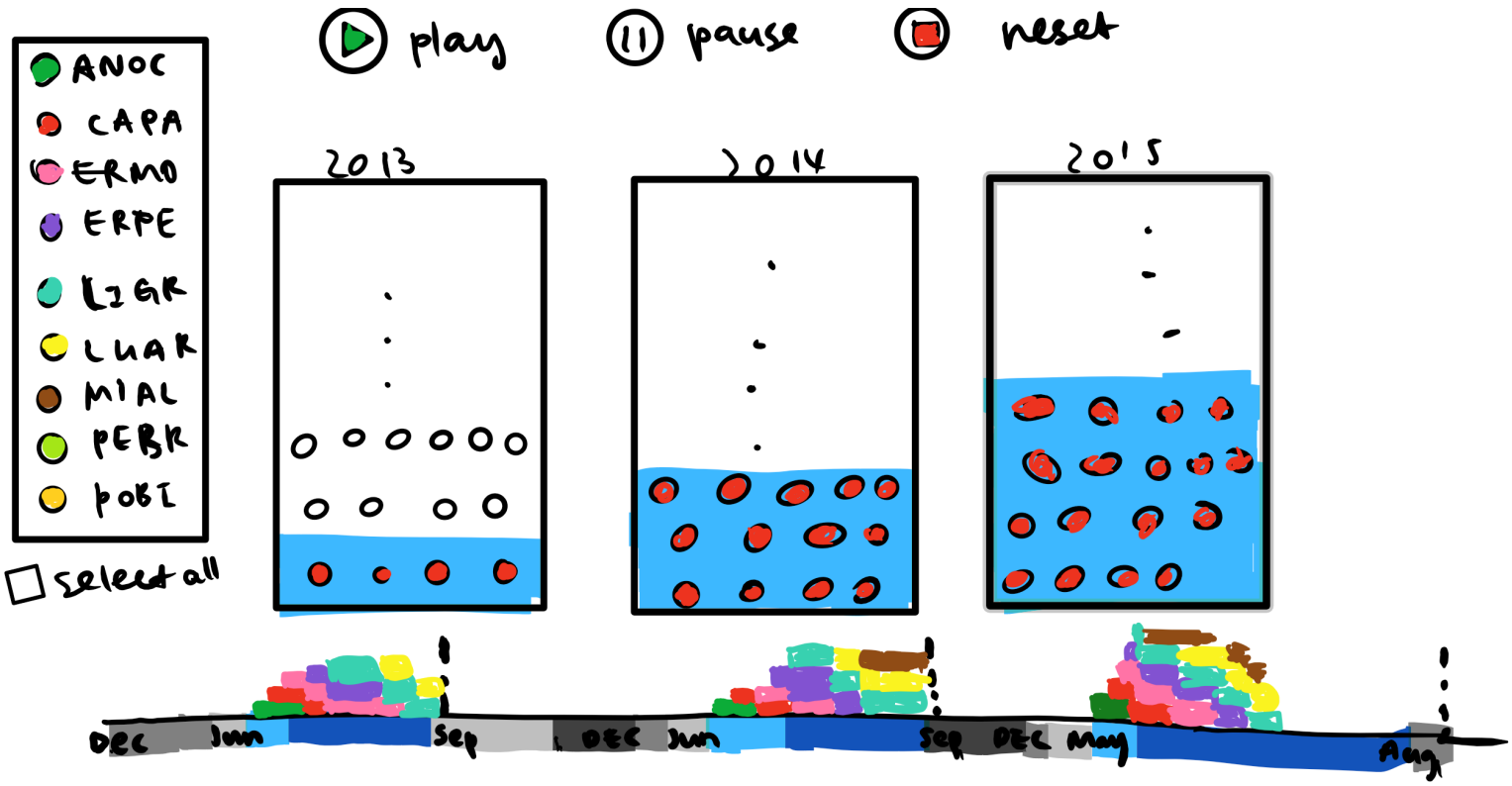
Relevant Prior Work

1. The USA National Phenology Network
2. Project BudBurst





Storyboard


play
pause
reset
ANOC
CAPA
ERMO
ERPE
LIGR
LHAR
MIAL
PEBR
POBI
select all
2013
2014
2015
DEC
Jun
Sep
DEC
Jun
Sep
DEC
May
Aug

Questions for You

- What is your least favorite part of the design?
- Is there anything you find confusing?
- How difficult is it to do a visual search of the flowers? Any alternative approach you'd like to suggest?

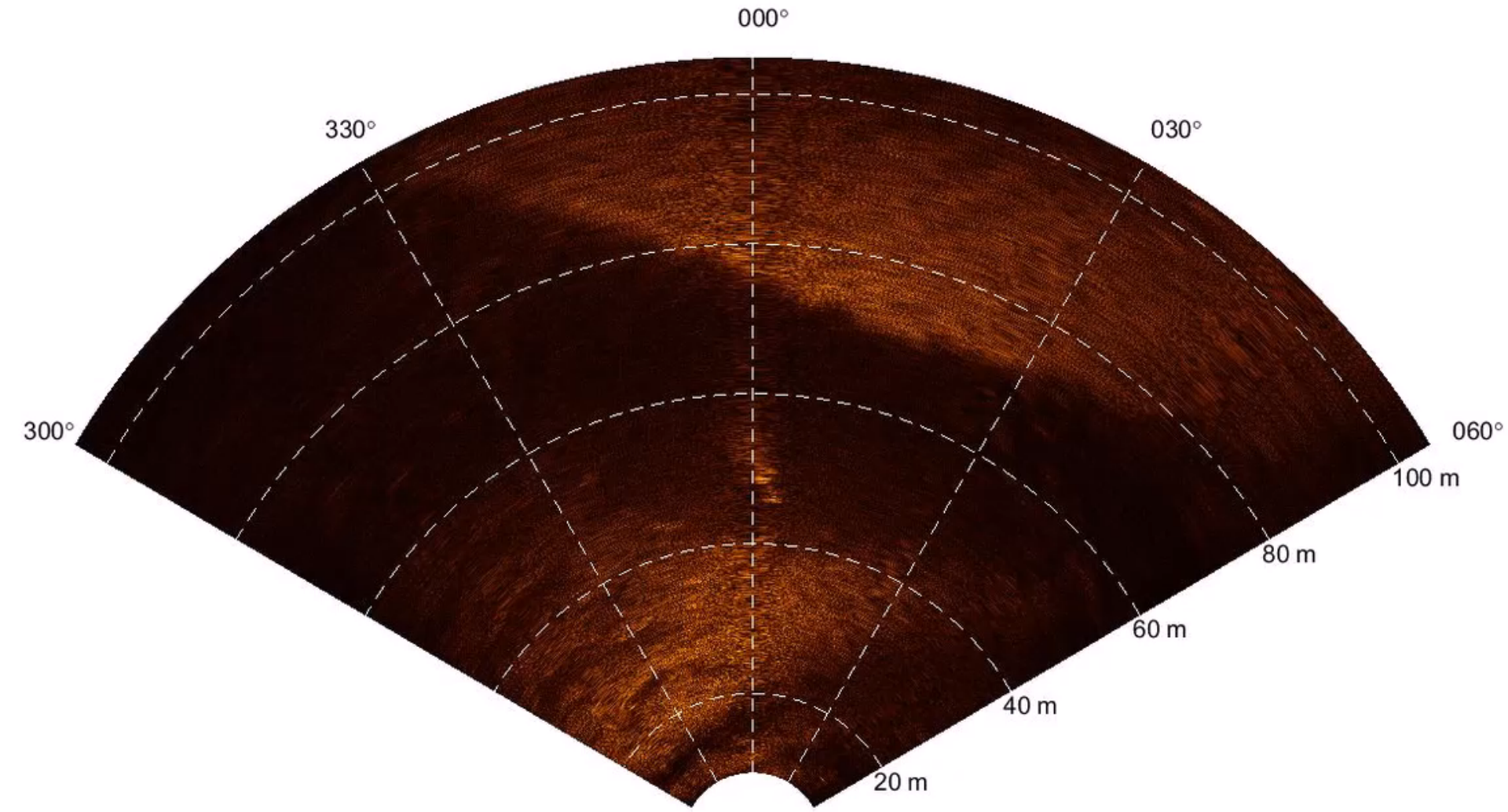
Multimodal Oceanographic Data Visualization

Peiran Liu, Mengjie Pan,
Alex Tank, Yali Wan

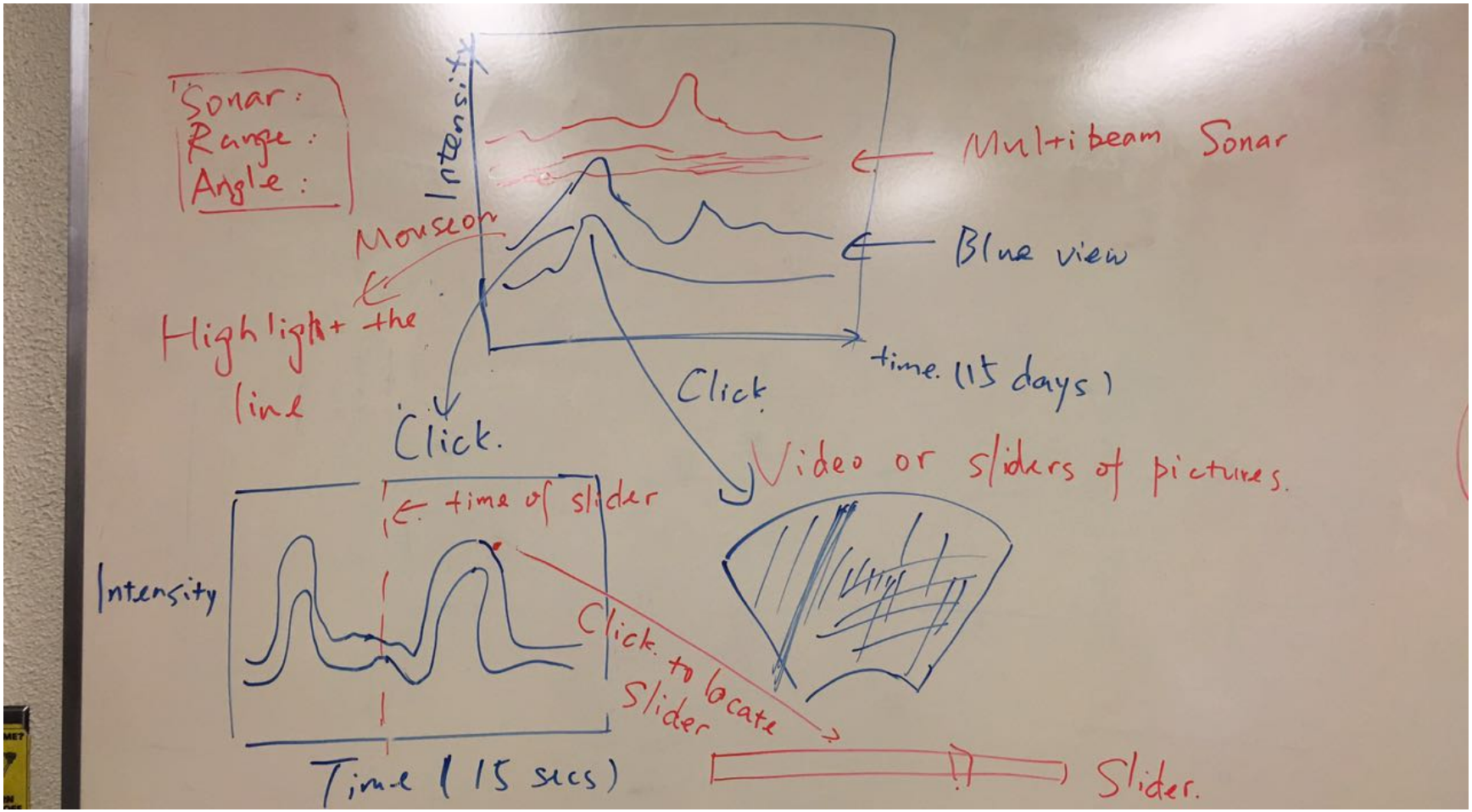
The problem

- Goal: simultaneously interpret data from Multibeam sonar and Blueview sonar and identify interesting cases (e.g. school of fish)
- Dataset: matrices of intensity values
- Data size: 15 days, 96 records a day, 150 pictures per record

Multibeam Sonar



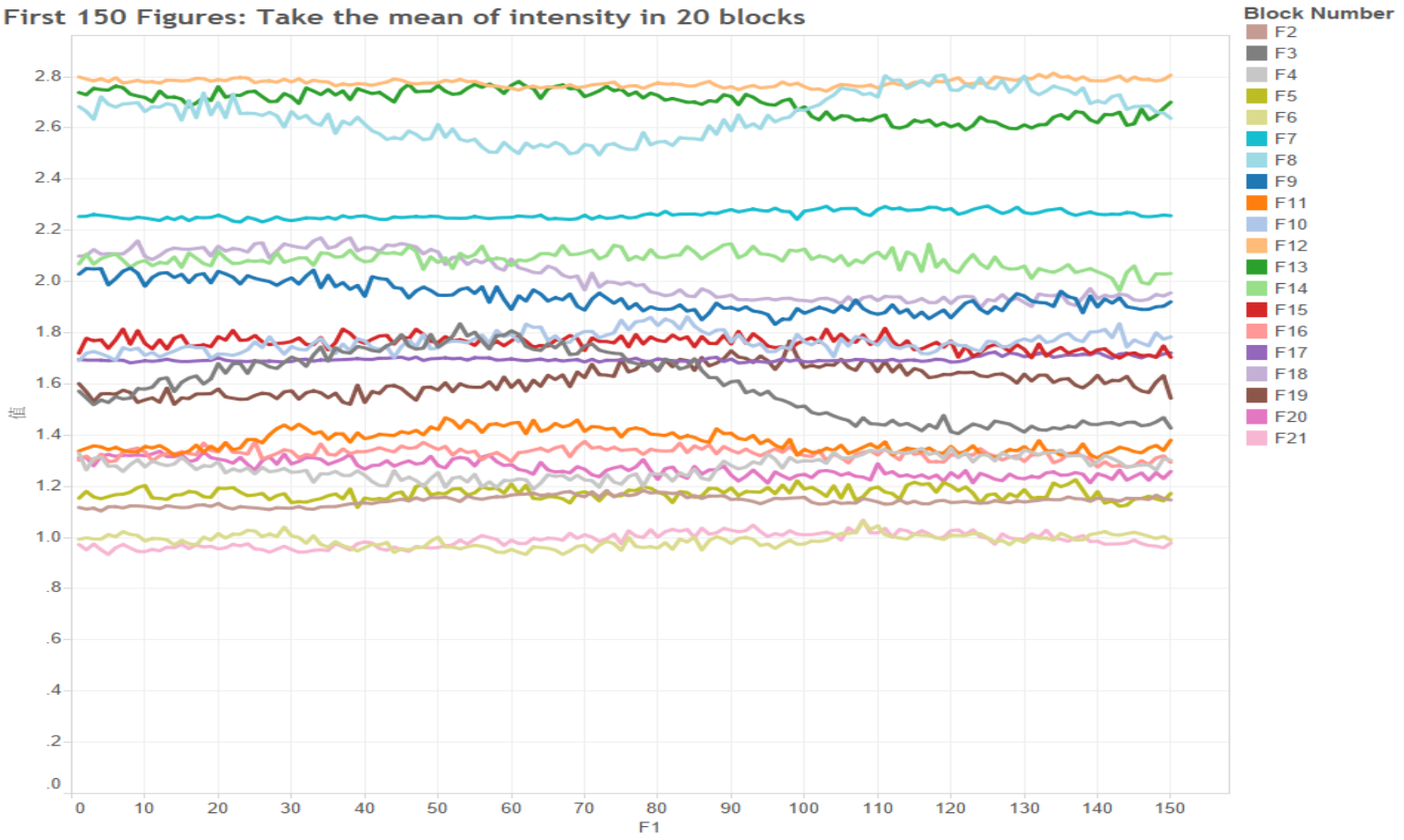
Design and Storyboard



Relevant Work

- Scripps Whale Acoustic Lab at USCD uses long-term spectral average plot to visualize time, frequency, and spectral level simultaneously and identify acoustically significant events, like whale calls.

Current Progress



Issues

- Big data preprocessing
- Use of mean or max or other statistical summary to identify significant changes

River Plume Salinity

Chung, Kim, Klyne, Potdar

Sam Kastner, Ph.D. Candidate

Civic and Environmental Engineering
and Applied Physics Lab

Where does the water go?

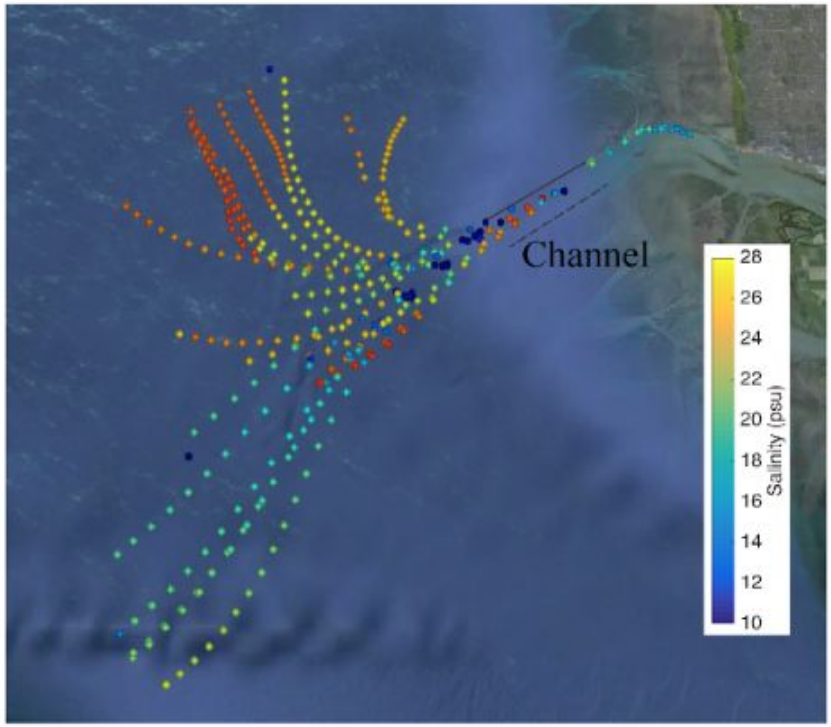
Can fresh water mixing with ocean water be modelled by wave height and salinity?

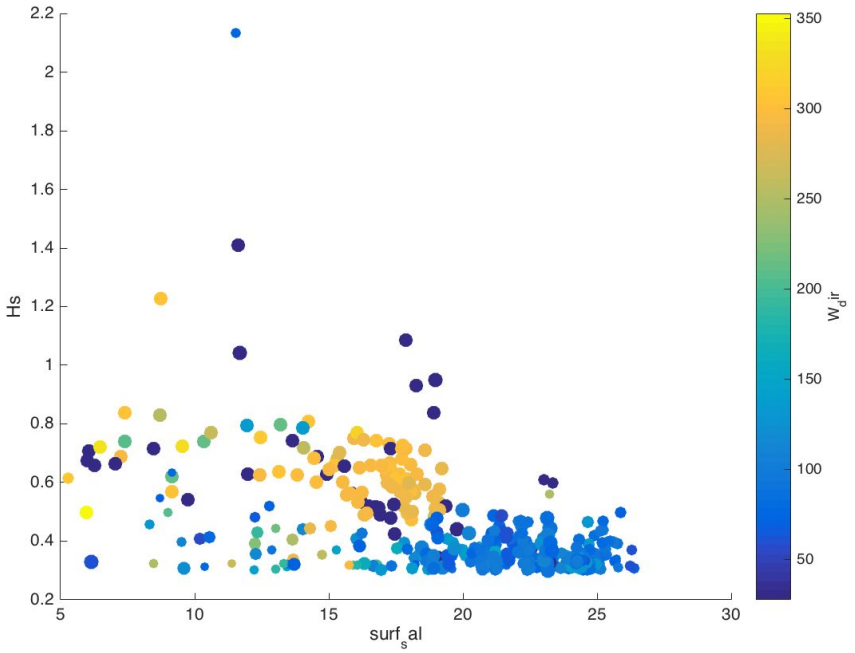
About the Data

- Fraser River Delta
- 6 buoys
- Collecting multiple dimensions of data (i.e. wind height, wind direction, etc.)
- Deployed over 10 days
- Sampled every 12 minutes
- Collected for 4 Hours each day

Current Visualization Techniques

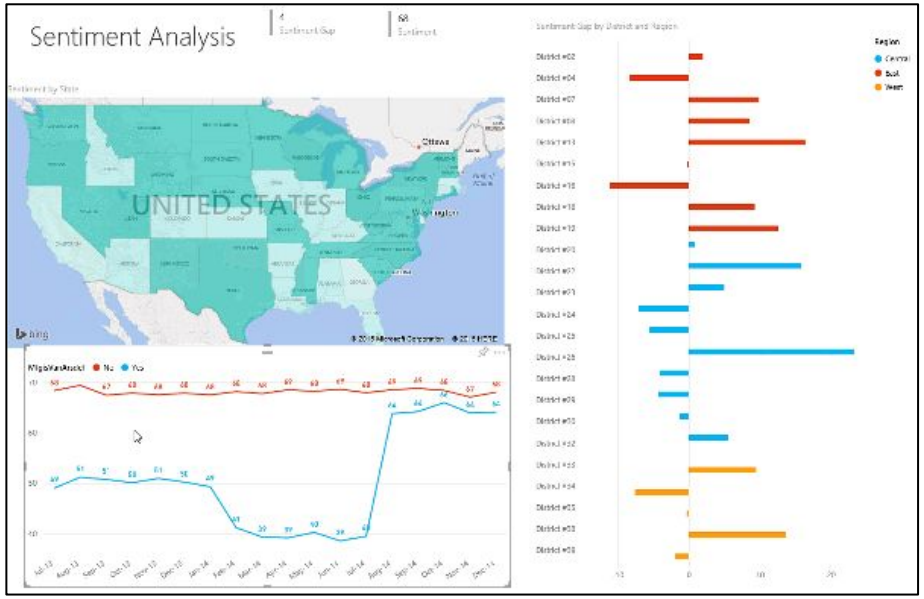
a) Drift track salinity





Potential Improvements

- Provide ability to compare data
- Allow filtering of data
- Improve data encodings
- Improve workflow

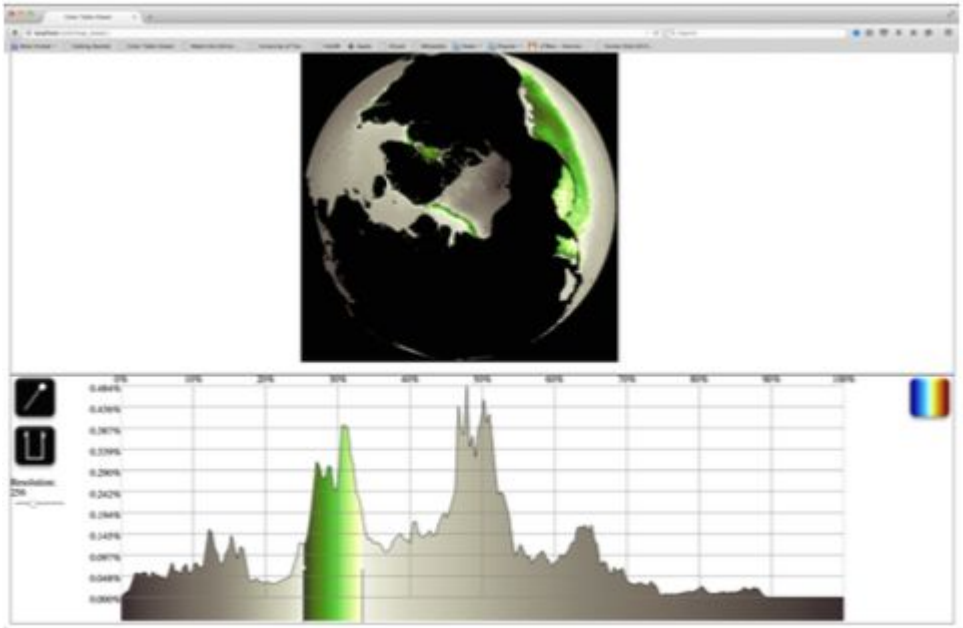


https://powerbi.microsoft.com/en-us/documentation/powerbi-service-tutorial-filled-maps-choropleths/

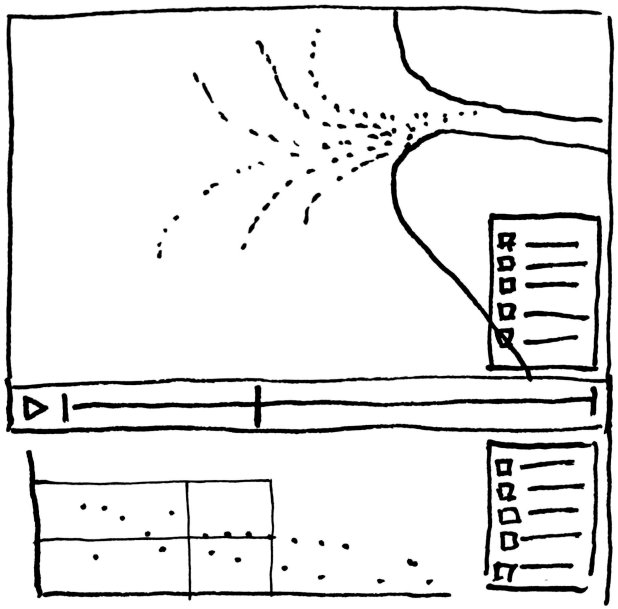
Relevant Work

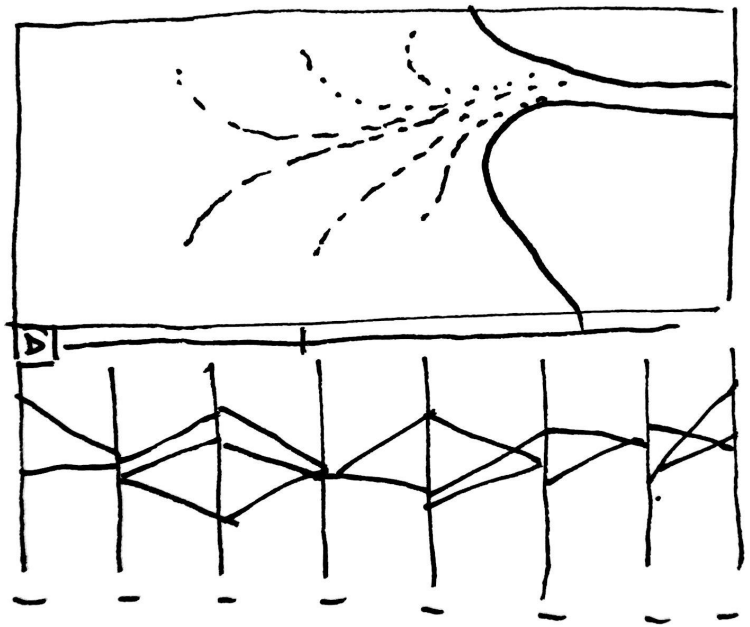
ColorMoves

Samsel et. al.



Ideation Sketches





Questions

- When applying cross-filtering, how many dimensions can you include before it causes too much cognitive overload?
- How can parallel coordinates and cartographic visualization be cross-brushed?
- How can data of varying ranges be compared easily? (i.e. wave height ranges vs. wind speed)

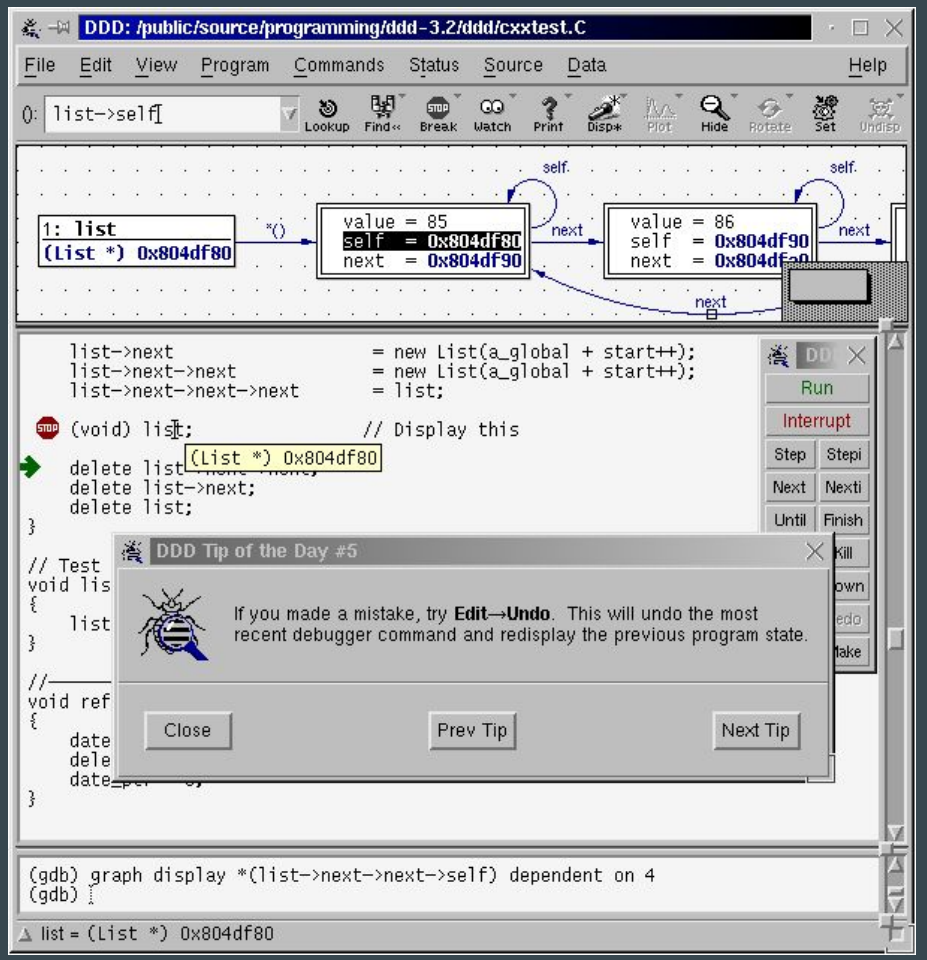
Celicia (See-C)

...

A Visual Debugger that draws memory the way you think about memory.

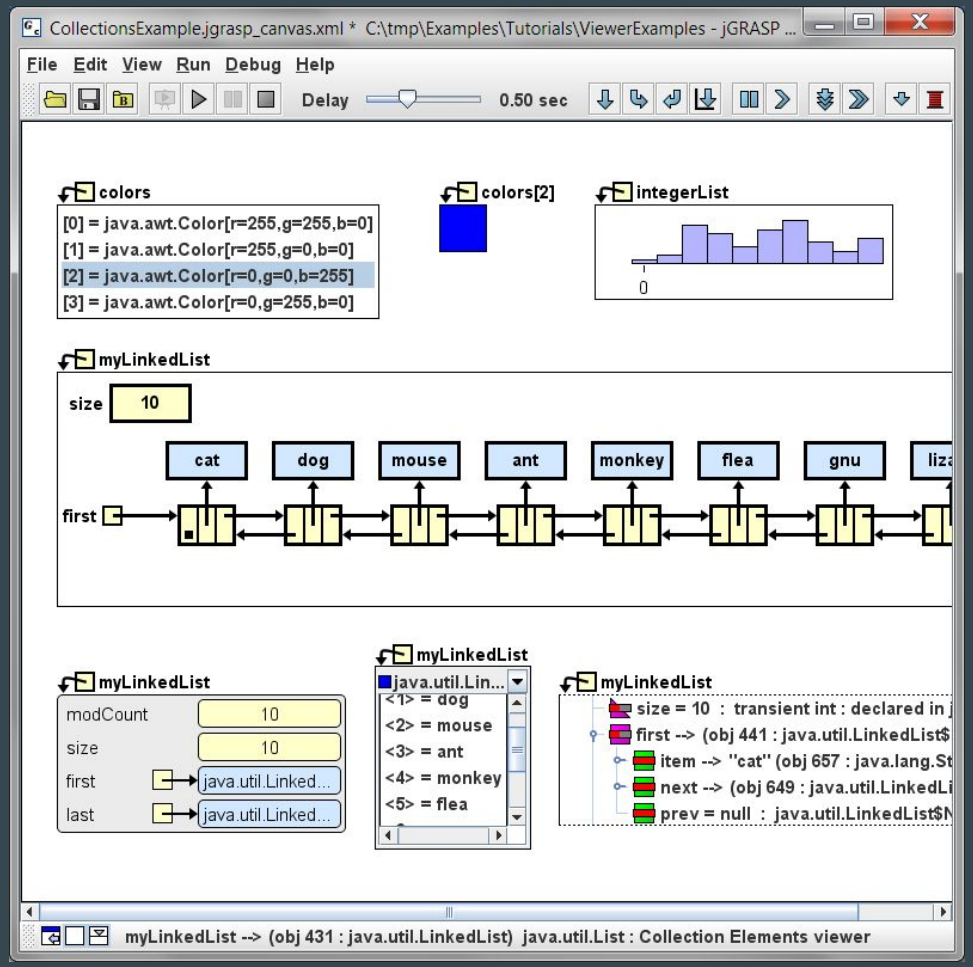
Prior Work

- DDD (Data Display Debugger)
- JGrasp
- Python Tutor



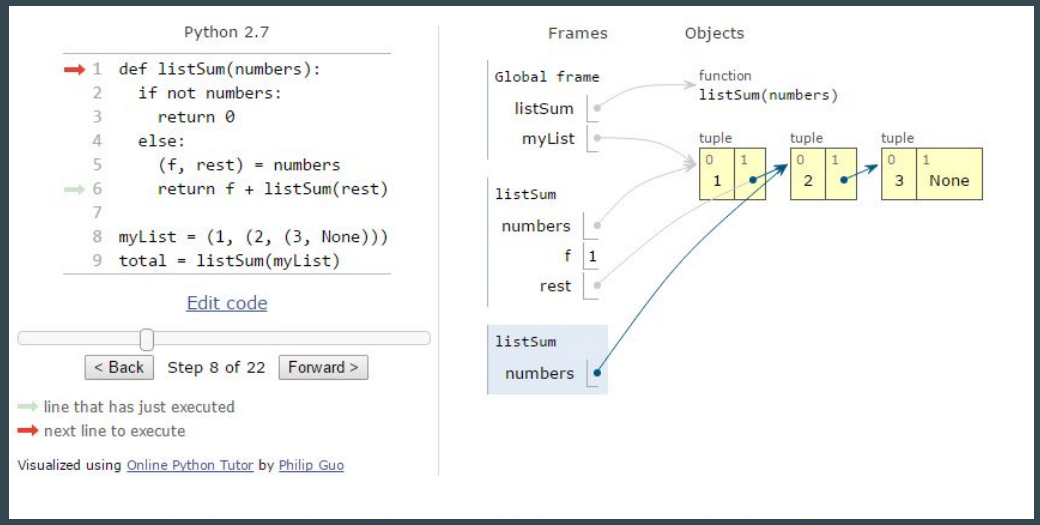
Prior Work

- DDD (Data Display Debugger)
- JGrasp
- Python Tutor



Prior Work

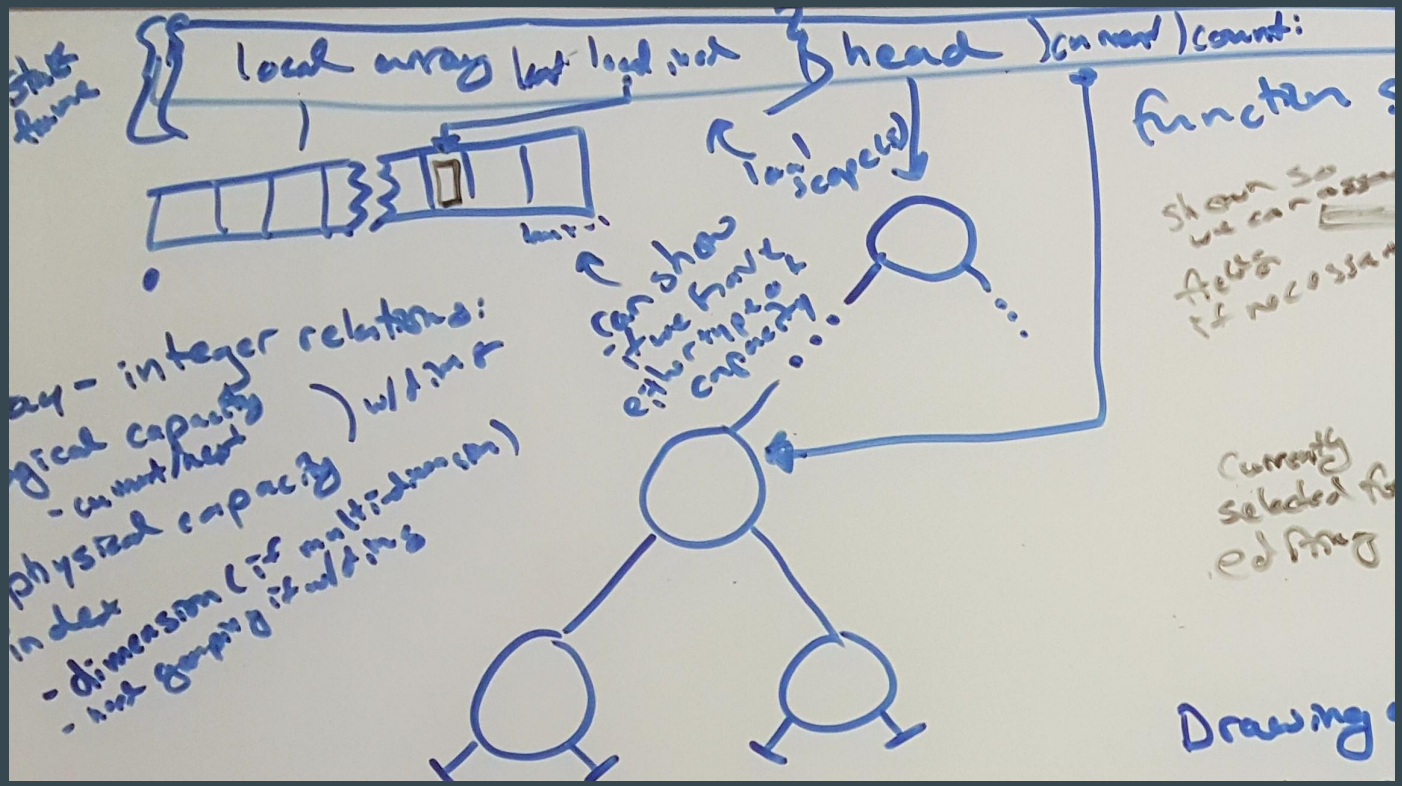
- DDD (Data Display Debugger)
- JGrasp
- Python Tutor



Cecilia is Different

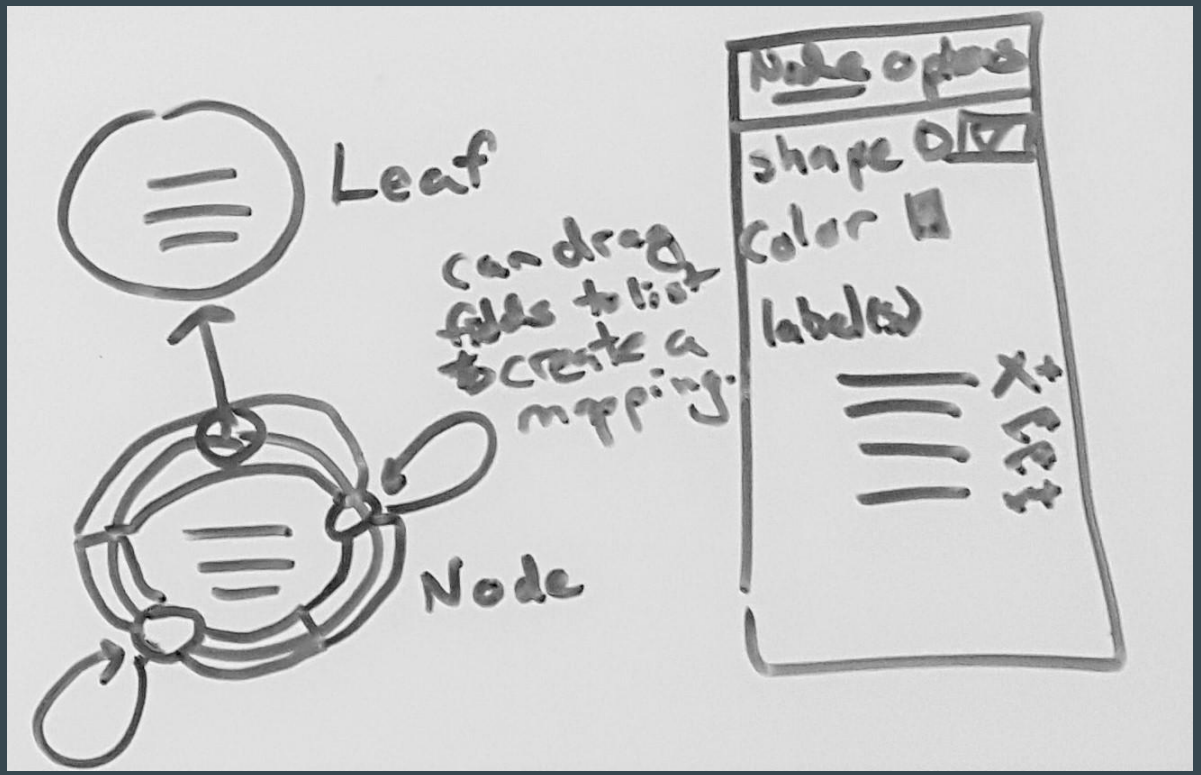
- Visually distinguish types: programmer can alter defaults dynamically (“Tableau for debugging”)
- Visually indicate scopes
- Separate stack and heap data
- JS instrumentation lets us step forward and backward through time
- Minimize context switching by visualizing program state alongside code, even while editing
- Programmer can provide semantic information about their program to associate variables

Separate Stack, Scopes, and the Heap



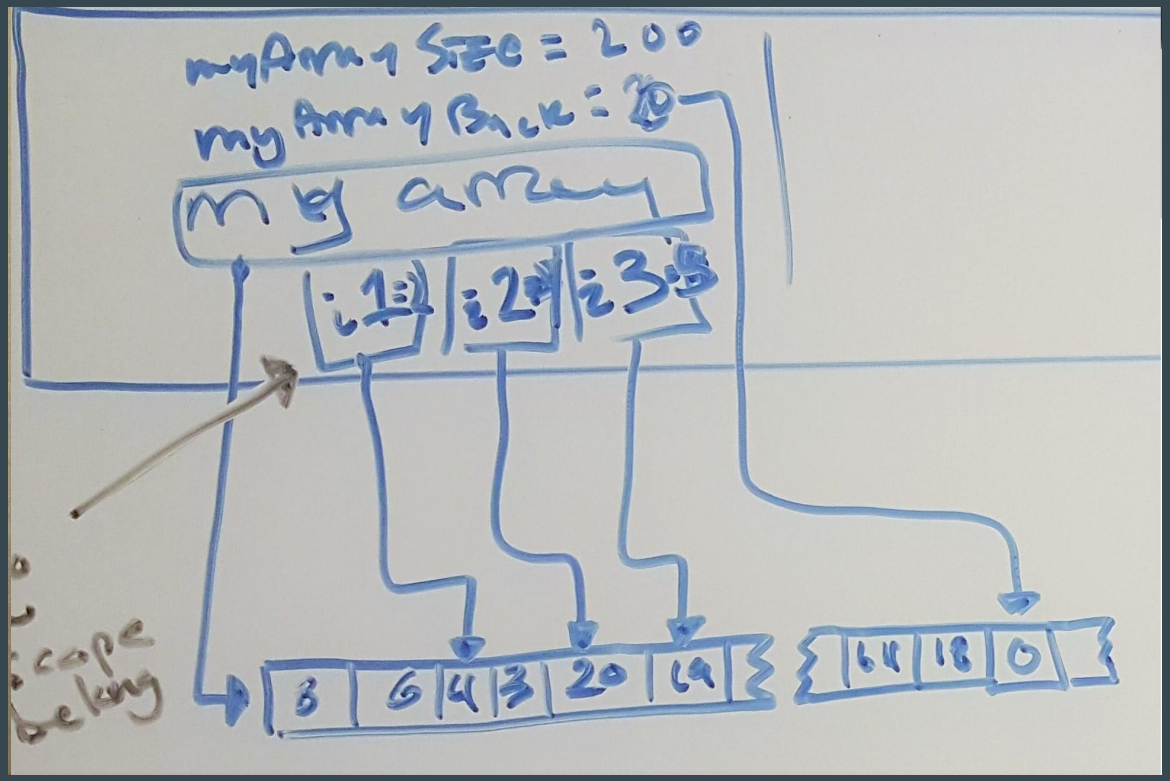
Editing View

```c
#include <stdio.h>
#include <stdlib.h>

typedef struct node {
    int value;
    struct node *next;
    bool color;
    char *name;
} node;

void addNode(node **head, int
    node *next = malloc(sizeof
    next->next = *head;
    next->value = value;
    *head = next;
```

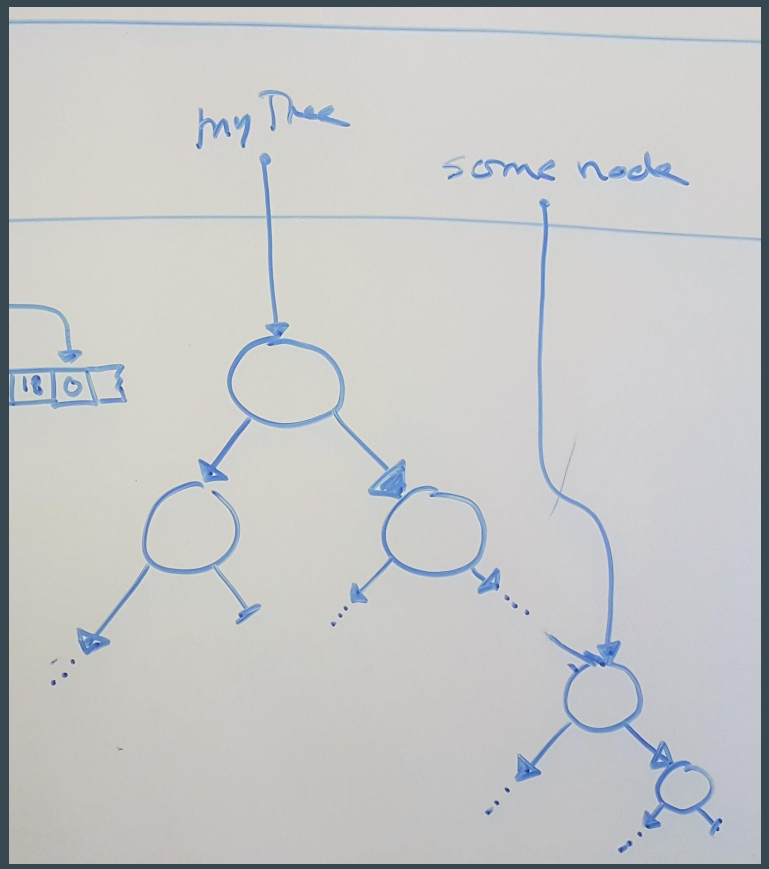


Variable Associations

- Array Bounds
  - Logical
    - Current
    - Next
  - Physical
- Array Indices



Data Hiding



Project Plan

- Design and Storyboarding (100% - Pending Feedback) (Ben and Julie)
- GDB Integration (80% - In Progress) (Ben)
- IDE Integration (20% - In Progress) (Julie)
  - Embed Visualization in IDE Window
  - Parse C
- Customizing Visualization (Ben)
  - Code Annotation
- Visualization of Debugging (Ben and Julie)
  - Interactive Elements

Questions

- What do YOU draw when you are writing C?
- How do we balance grouping scopes and grouping variable associations?
- How do we balance grouping variable associations with one memory location <-> one display location?
- How much of a data structure should be shown by default?

Exploratory Performance Analysis of Query Execution Engines

Helga Gudmundsdottir

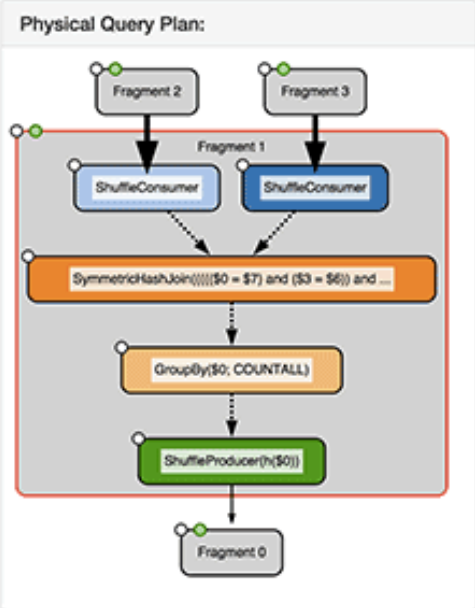
-- CSE512: Final Project Progress Presentation --

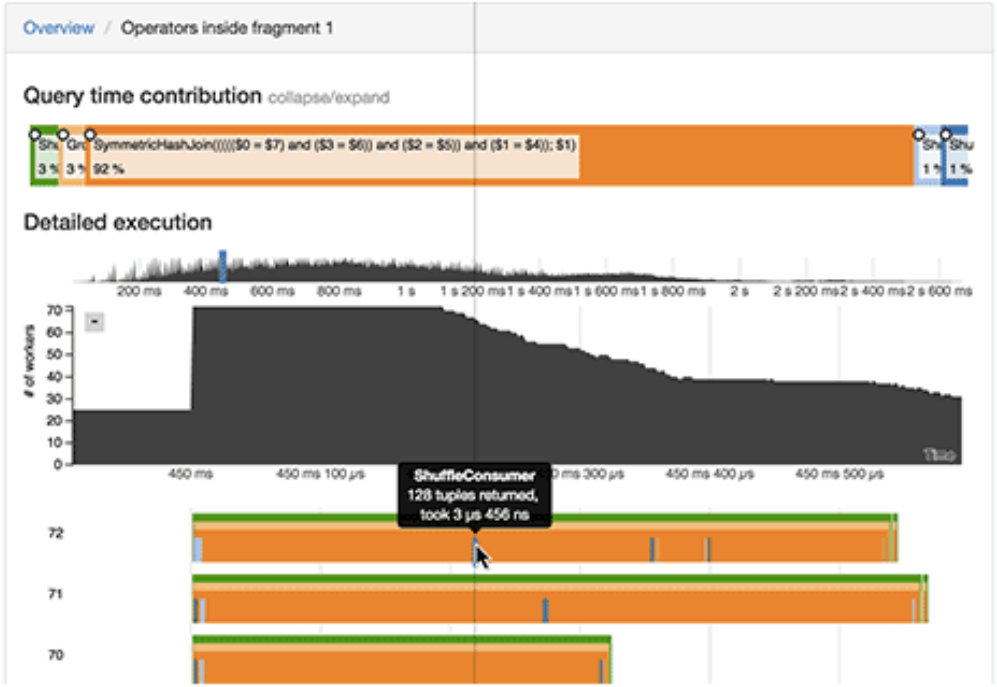

The Datafloq Open Source Landscape 2.0
Data Analysis & Platforms
HPCC Systems
Storm
Dremel
Spark
Myria
SAMOA
APACHE DRILL
IKANOW | BRILLIANT DECISIONS
Hortonworks
Databases / Data warehousing
bigdata
INFOBRIGHT
Cassandra
4store
H2
GlobalsDB
InfiniDB
riak
Infinispan
HYPERTABLE
MariaDB
Drizzle
SQLite
RethinkDB
terrastore
HIVE
Firebird
ORACLE
BERKELEY DB
HyperSQL
monetdb
In-Memory Computing
GridGain
hazelcast
TERRACOTTA
GORA
NMemory
ERP BI Solutions
Business Intelligence
openi.org
talend
Jaspersoft
Palo
spagobi
pentaho
jedox.
BIRT
Data Mining
orange
rapidminer
KNIME
mahout
WEKA
KEEL
togaware
SPMF
Big Data search
lucene
Apache Solr
elasticsearch.
Multivalue database
Rocket
U2
REVELATION
QM
northgate
jBASE INTERNATIONAL
ScarletDME
Programming
julia
Data aggregation
cloudera
hortonworks
KeyValue
EROSPIKE
leveldb
redis
Chordless
Tokyo Cabinet
MEMCACHED
SCALIEN
Project Voldemort
A distributed database.
hamsterdb
RAPTORDB
FairCom
STSDB
HyperDex
OpenLDAP
IQLECT
ioremap.net
STORAGE AND BEYOND
Scalaris
Document Store
mongoDB
CouchBase
Raven DB
CLUSTERPOINT
Tokutek
RaptorDB
EJDB
djon DB
JasDB
SchemafreeDB
sisodb
CouchDB
relax
Graph databases
Gephi
Gremlin
FRANZ INC.
GraphBuilder
*Sparksee
InfiniteGraph
INFO GRID
HYPERGRAPHDB
Neo4j
FlockDB
GraphBase
BrightstarDB
Operational
VOLTDB
Social
Apache Kafka
ThinkUp
Corona
Multidimensional
FIS
SciDB
rasdaman
db4objects
Object databases
NEOPPOD
ZOPE
mobject
Magma
Picolisp
siaqodb
FRAMER D
PERSEVERE
EyeDB
STARCOUNTER
Sterling
NDatabase
C# Lightweight Object Database
Multimodel
ArangoDB
alchemydatabase
XML Databases
existdb
BASE
Qizx
sedna
LIQUIBASE

Related Work

Perfopticon: Visual Query Analysis for Distributed Databases

Dominik Moritz, Daniel Halperin, Bill Howe, Jeffrey Heer

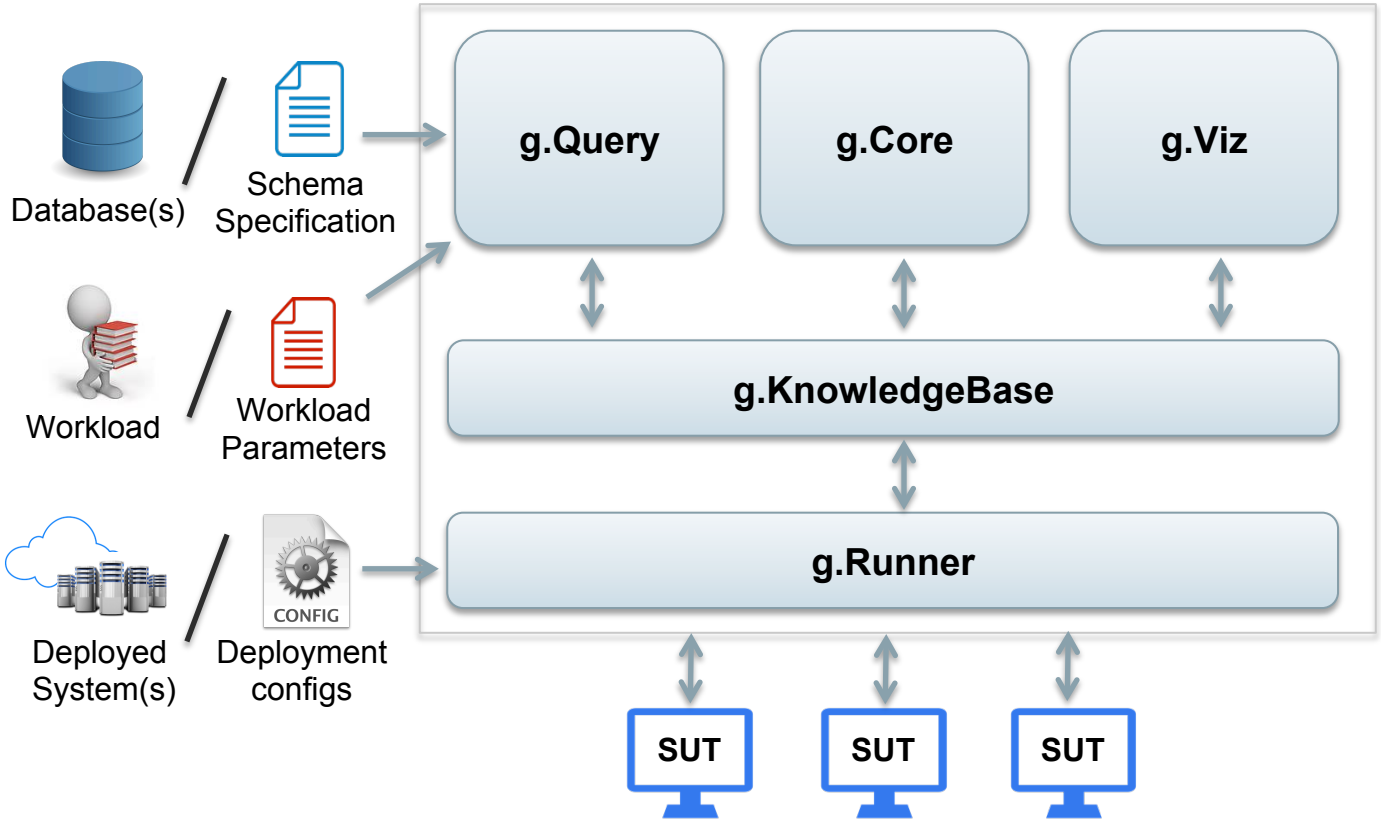




3

Greinir:

A toolkit for analyzing performance characteristics of data analytics systems



g.KnowledgeBase:

- Simple queries with extracted features
- Performance metrics (query runtimes, per-node resource utilization)

g.Viz:

- Interactive performance exploration!
- Support in-depth analysis

Initial prototype -- http://cse512-16s.github.io/a3-helgag/

How to use color?

How to deal with domains of axes/ zooming?

Many more features…

Dynamic number of systems?

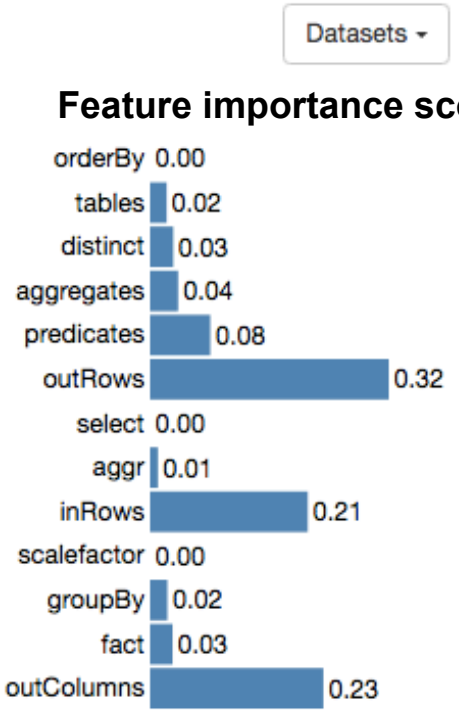
Different level of detail?

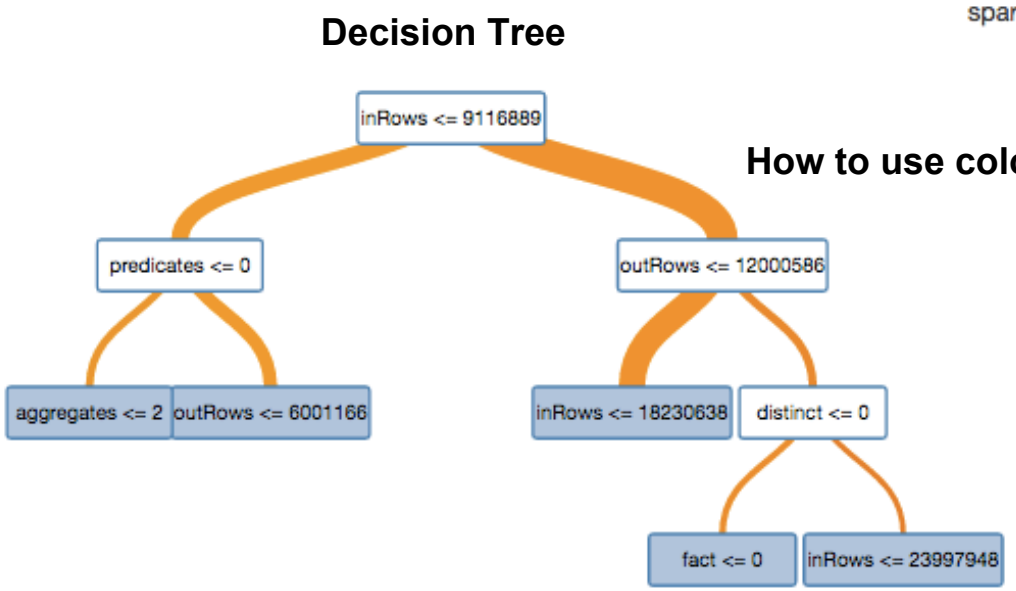
When to zoom/ what to zoom?

Different performance metrics?

5

Clustering and classification





How to use color?

How visualize branches/leaves?

Questions & Challenges

- **Goal:** aid exploration, comparison between systems, grouping queries
- High dimensionality – avoid clutter but not hide potentially interesting information
- How to deal with coordinated views?
  - Should axes be static/elastic? Zoom one, zoom all?
- How to deal with outliers?
- How to incorporate output from machine learning algorithms?

Final Project Progress

Ryan McGee

Problem

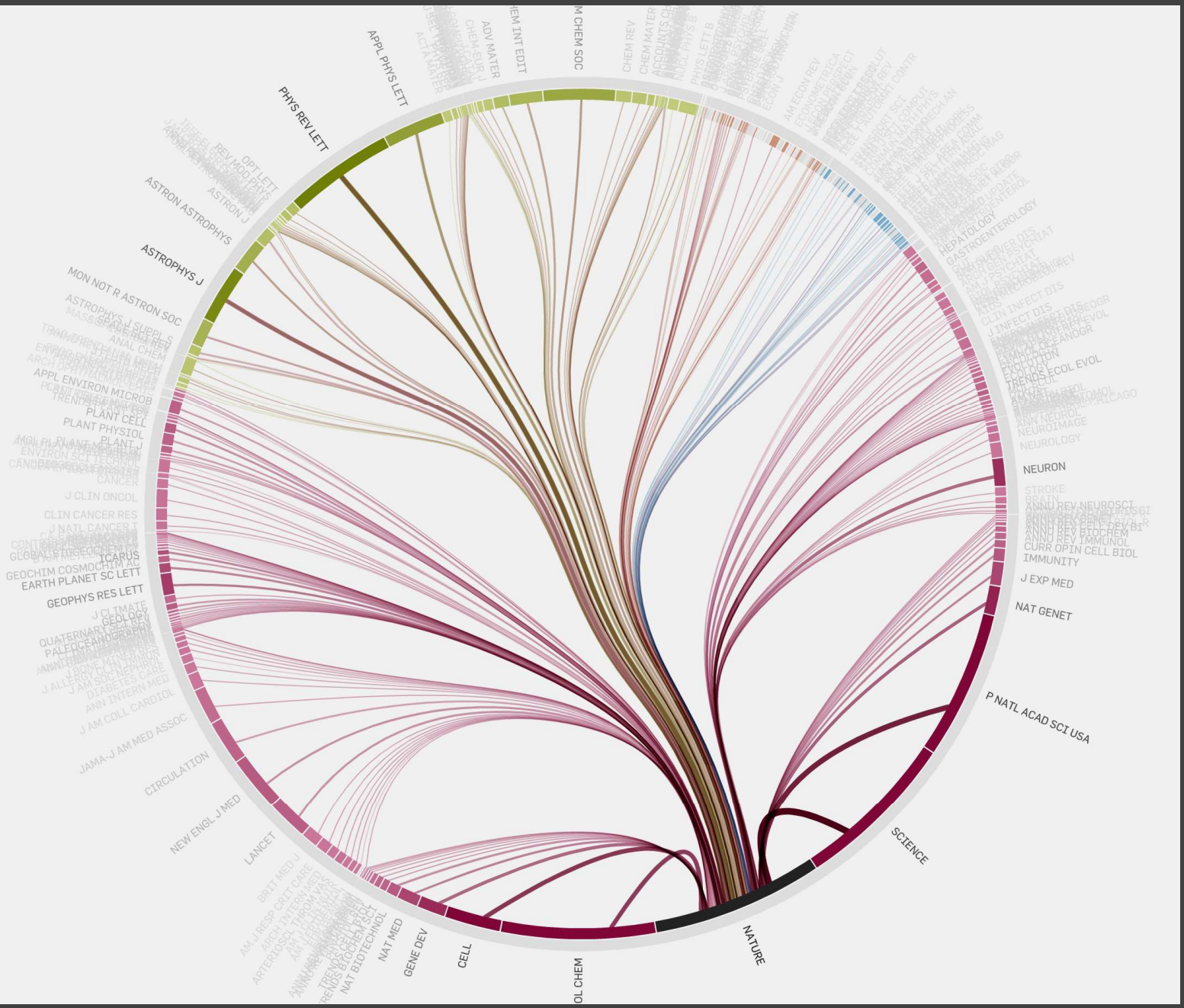
Make sense of large interaction networks

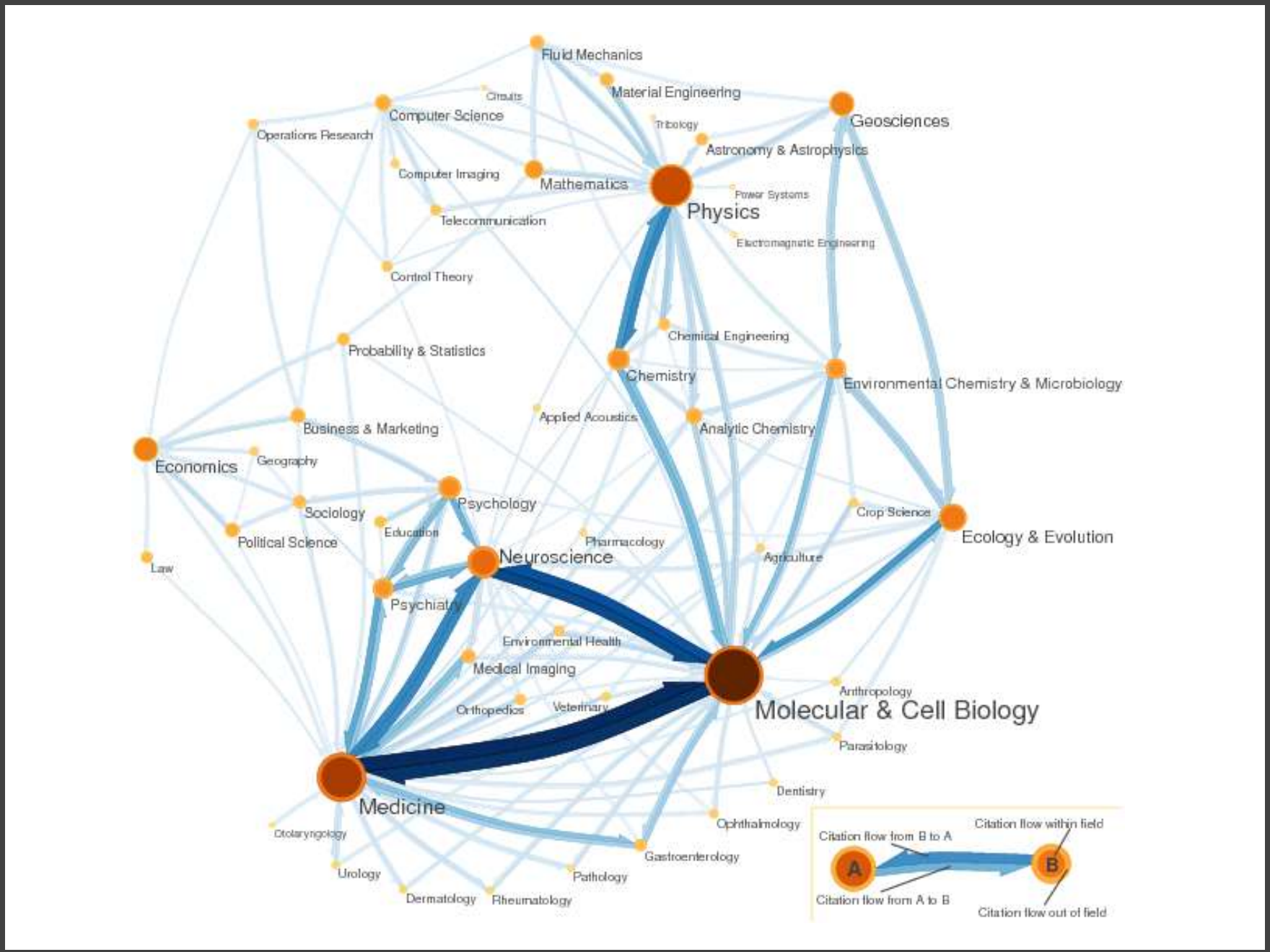
- Understand community structure
- Understand the influential/defining players in communities

In Practice: Using citation networks to understand the structure of science

- How are fields and subfields organized?
- How do implicit fields apparent in citation networks relate to explicit institutional fields (departments)
- What are the most important works and authors in given fields?

NATURE


PHYS REV LETT
APPL PHYS LETT
ADV MATER
M CHEM SOC
ASTRON ASTROPHYS
ASTROPHYS J
MON NOT R ASTRON SOC
APPL ENVIRON MICROB
PLANT CELL
PLANT PHYSIOL
ICARUS
EARTH PLANET SC LETT
GEOPHYS RES LETT
CLIN CANCER RES
J CLIN ONCOL
JAMA-J AM MED ASSOC
CIRCULATION
NEW ENGL J MED
LANCET
NAT MED
GENE DEV
CELL
NATURE
SCIENCE
P NATL ACAD SCI USA
NAT GENET
J EXP MED
IMMUNITY
CURR OPIN CELL BIOL
ANNU REV IMMUNOL
NEURON
NEUROLOGY
GASTROENTEROLOGY
HEPATOLOGY


Fluid Mechanics
Material Engineering
Circuits
Computer Science
Geosciences
Operations Research
Tribology
Astronomy & Astrophysics
Computer Imaging
Mathematics
Power Systems
Physics
Telecommunication
Electromagnetic Engineering
Control Theory
Chemical Engineering
Probability & Statistics
Chemistry
Environmental Chemistry & Microbiology
Applied Acoustics
Business & Marketing
Analytic Chemistry
Economics
Geography
Psychology
Crop Science
Sociology
Education
Ecology & Evolution
Political Science
Pharmacology
Neuroscience
Agriculture
Law
Psychiatry
Environmental Health
Medical Imaging
Anthropology
Orthopedics
Veterinary
Molecular & Cell Biology
Parasitology
Dentistry
Medicine
Ophthalmology
Otolaryngology
Gastroenterology
Urology
Pathology
Dermatology
Rheumatology
Citation flow within field
Citation flow from B to A
A
B
Citation flow from A to B
Citation flow out of field

The Data

Citation network

JSTOR citation data for 1.7 million papers

Scientific Fields & Subfields

Identified using Map Equation on the citation network

“Impact Score” for each paper

Calculated using Eigenfactor metric

Motivation for Visualization

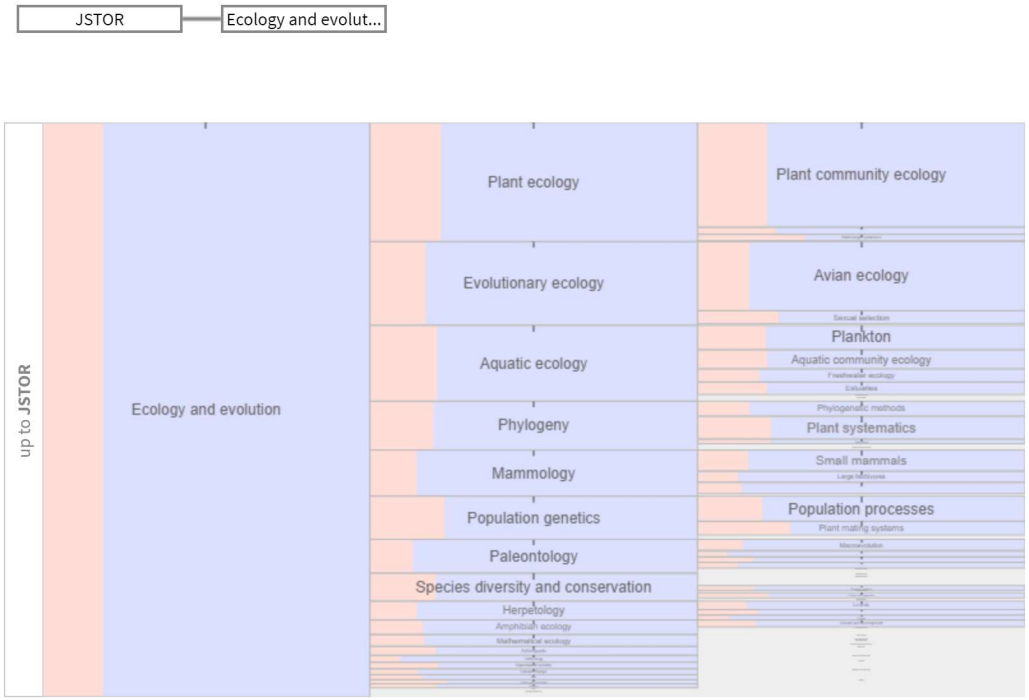
Understand structure of scientific fields/subfields

Identify the top papers in various fields/subfields

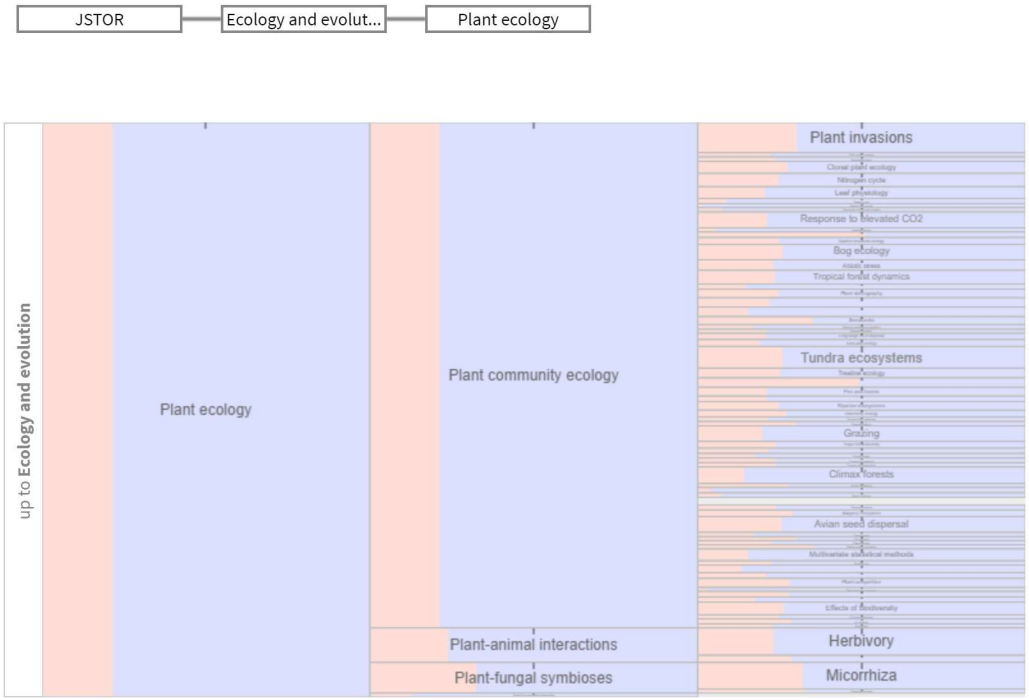
Identify where subfields/papers/authors of interest lie in the overall structure of science

Gain insight about interdisciplinarity of papers/authors

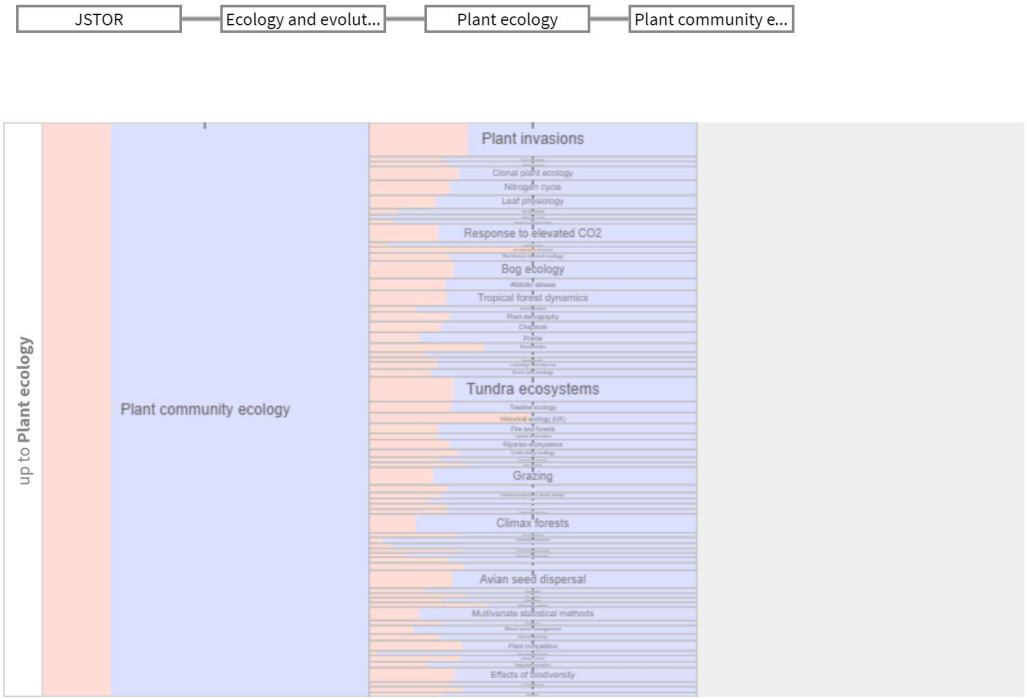
Previous Visualization



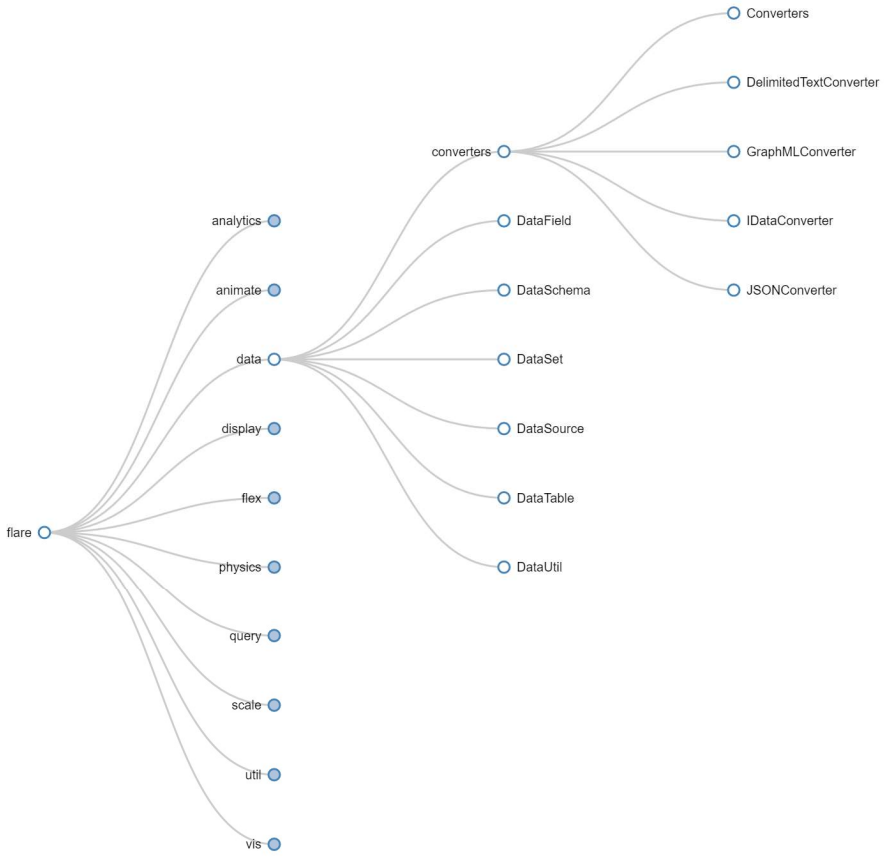
Previous Visualization



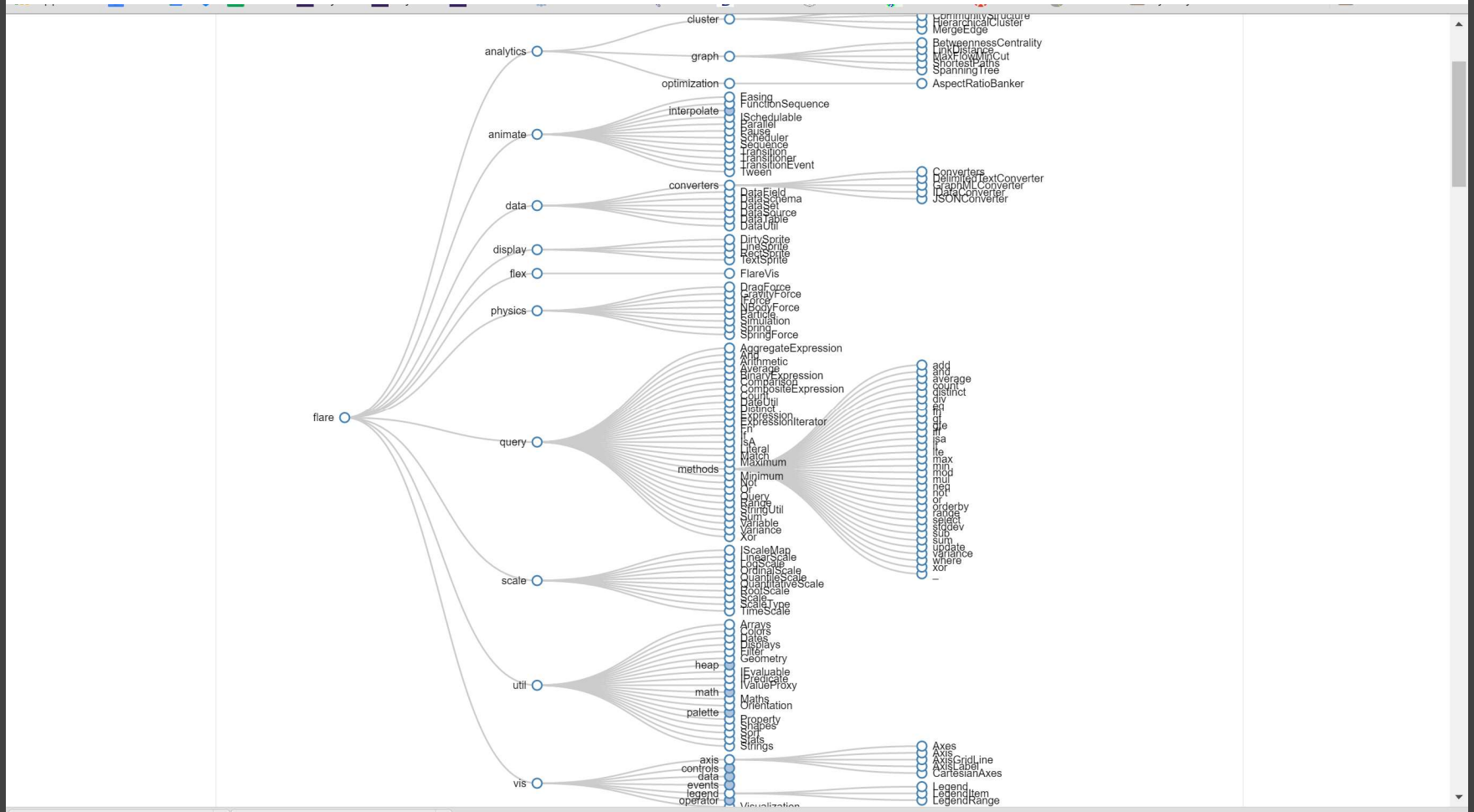
Previous Visualization



Collapsible Tree



Collapsible Trees



Degree-of-Interest Trees

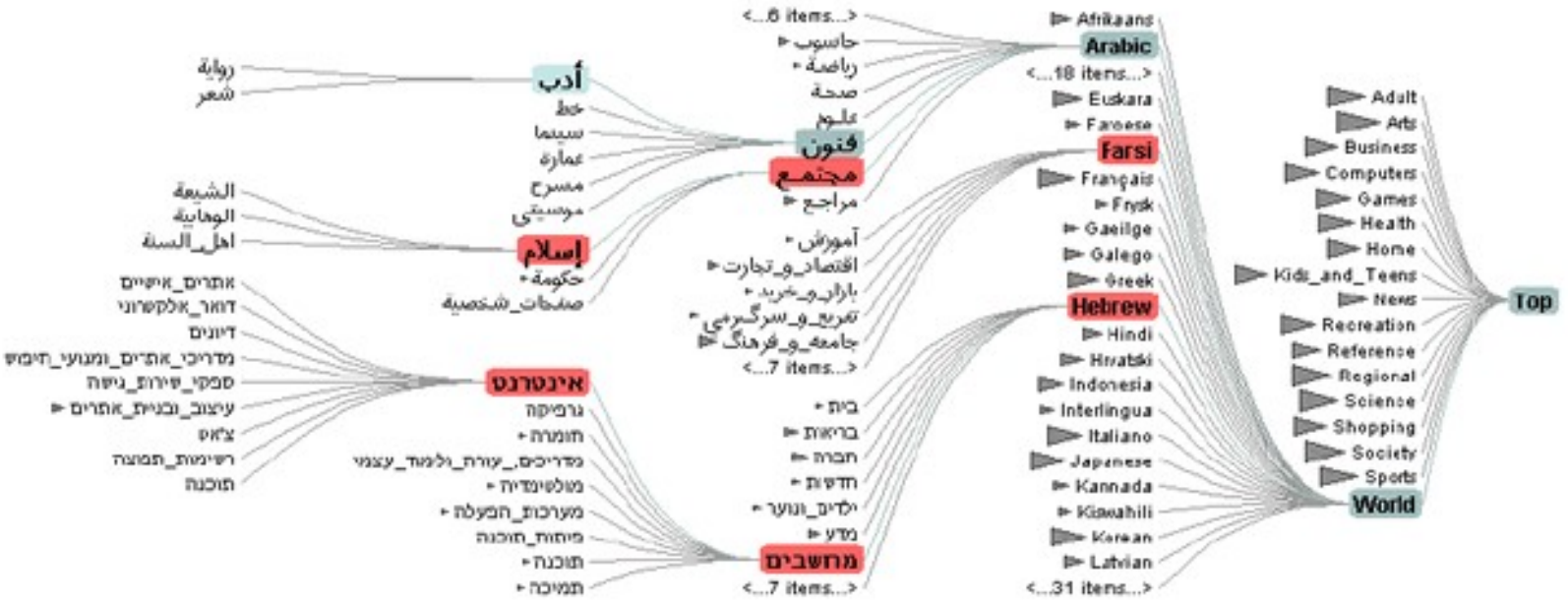
Degree-of-Interest Trees:
A Component of an Attention-Reactive User Interface

Stuart K. Card, David Nation
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304 USA
card@parc.com, dnation@acm.org

DOITrees Revisited: Scalable, Space-Constrained Visualization of Hierarchical Data

Jeffrey Heer [1,2]
[1]Group for User Interface Research
University of California, Berkeley
Berkeley, CA 94720-1776 USA
jheer@cs.berkeley.edu

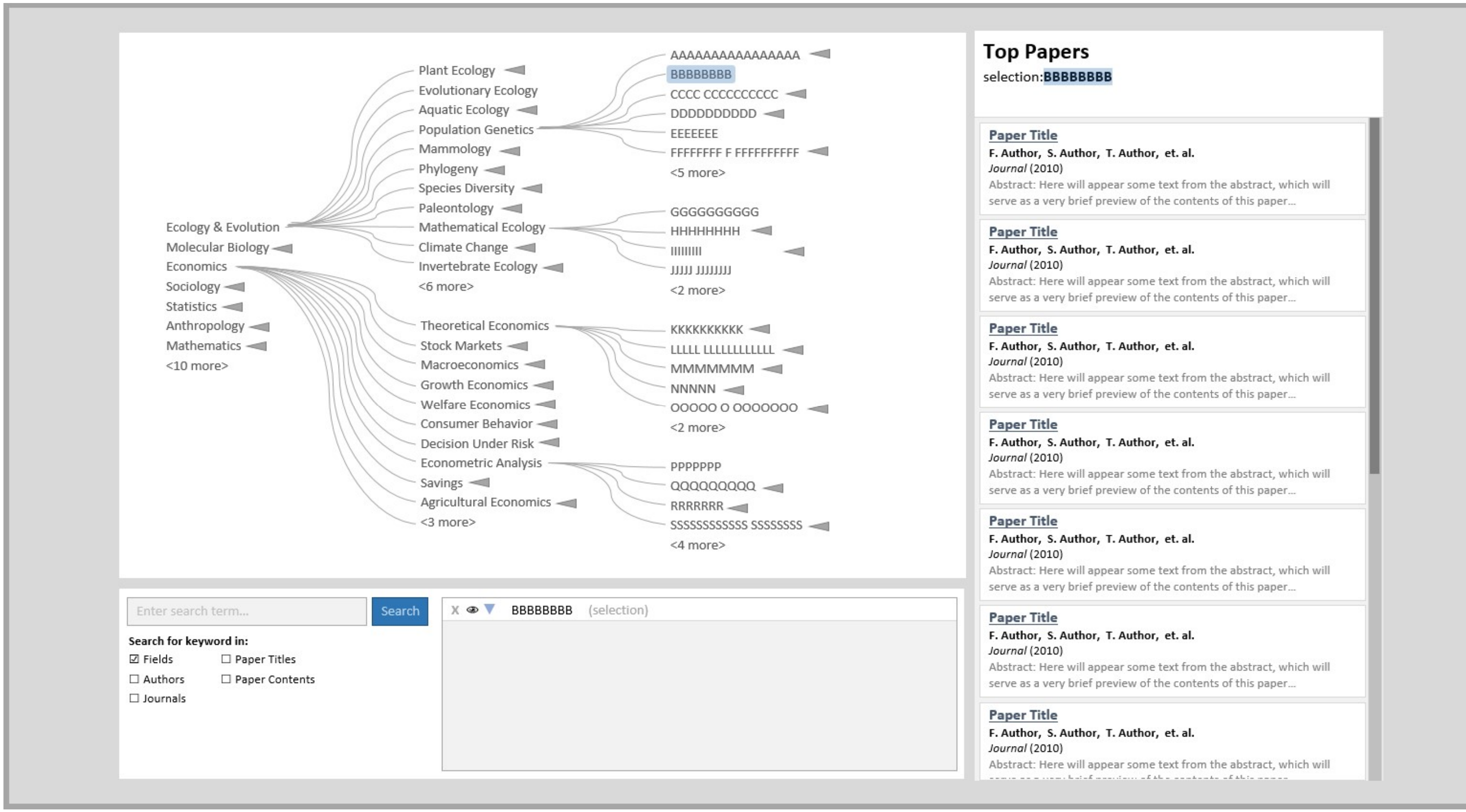
Stuart K. Card [2]
[2]Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94301 USA
card@parc.com

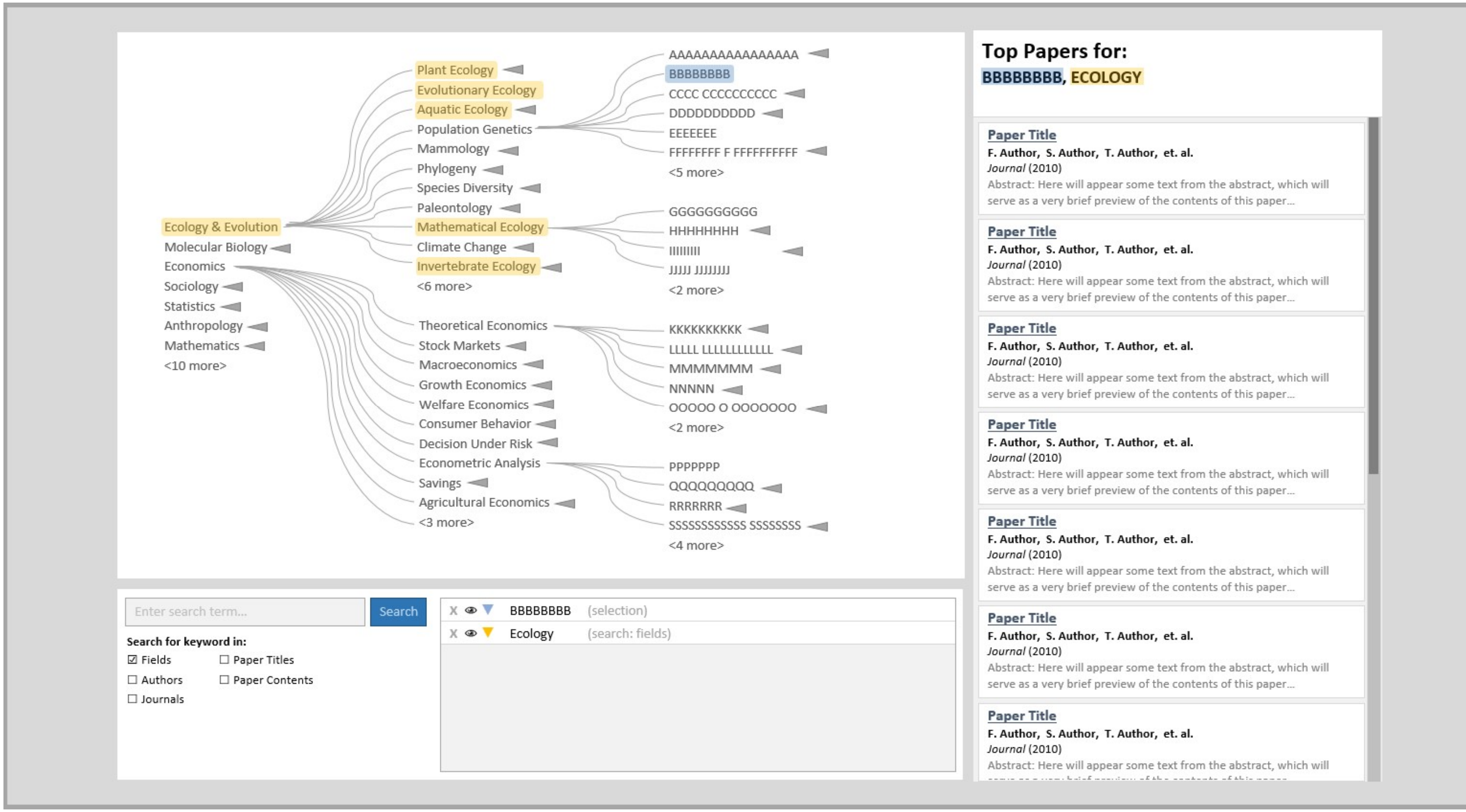


Contributions of this work

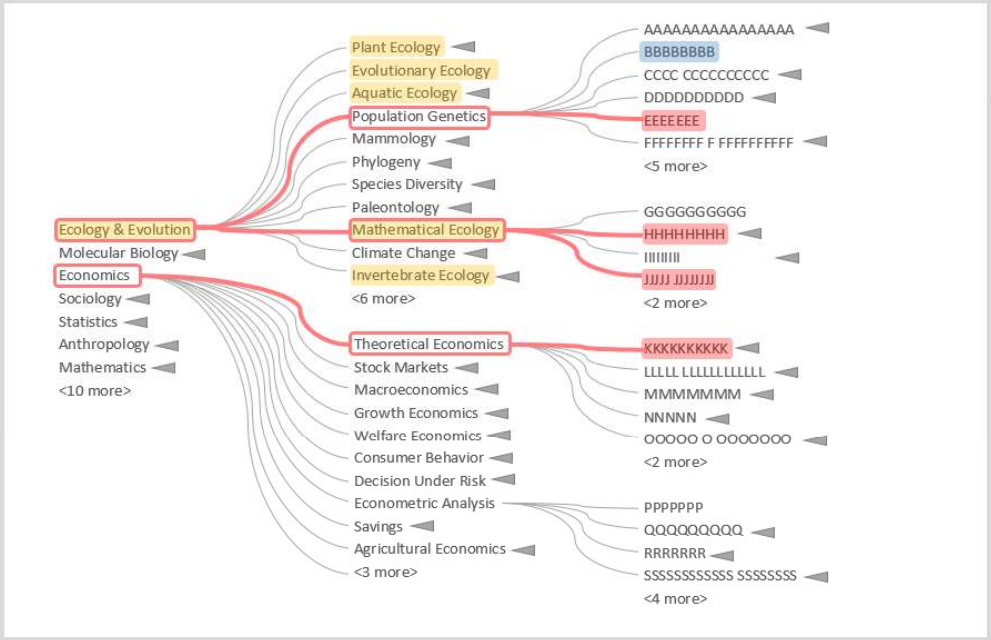
- Implement a fully-featured DOI tree for web with D3
  - Optimize focus-context tree layout using interest scores
  - Search the tree for fields, paper content, authors, etc
    - Apply interest, highlighting, other viz to search results
  - General purpose (not data specific)

- Option to use Eigenfactor impact scores in interest calculation for DOI layout
- View content for selected/searched nodes at any level (e.g., top papers)
- Visualize interdisciplinary connections between nodes


Ecology & Evolution
Molecular Biology
Economics
Sociology
Statistics
Anthropology
Mathematics
<10 more>
Plant Ecology
Evolutionary Ecology
Aquatic Ecology
Population Genetics
Mammology
Phylogeny
Species Diversity
Paleontology
Mathematical Ecology
Climate Change
Invertebrate Ecology
<6 more>
Theoretical Economics
Stock Markets
Macroeconomics
Growth Economics
Welfare Economics
Consumer Behavior
Decision Under Risk
Econometric Analysis
Savings
Agricultural Economics
<3 more>
AAAAAAAAAAAAAAAA
BBBBBBBB
CCCC CCCCCCCCCC
DDDDDDDDDD
EEEEEEE
FFFFFFFF F FFFFFFFFFF
<5 more>
GGGGGGGGGG
HHHHHHHH
IIIIIIII
JJJJJ JJJJJJJJ
<2 more>
KKKKKKKKKK
LLLLL LLLLLLLLLLL
MMMMMMM
NNNNN
OOOOO O OOOOOOO
<2 more>
PPPPPPP
QQQQQQQQQ
RRRRRRR
SSSSSSSSSSS SSSSSSSS
<4 more>
Enter search term...
Search
Search for keyword in:
Fields
Paper Titles
Authors
Paper Contents
Journals
X BBBBBBBB (selection)
Top Papers
selection:BBBBBBBB
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...



Ecology & Evolution
Molecular Biology
Economics
Sociology
Statistics
Anthropology
Mathematics
<10 more>
Plant Ecology
Evolutionary Ecology
Aquatic Ecology
Population Genetics
Mammology
Phylogeny
Species Diversity
Paleontology
Mathematical Ecology
Climate Change
Invertebrate Ecology
<6 more>
AAAAAAAAAAAAAAA
BBBBBBBB
CCCC CCCCCCCCCC
DDDDDDDDDD
EEEEEEE
FFFFFFFF F FFFFFFFFF
<5 more>
GGGGGGGGGG
HHHHHHHH
IIIIIIII
JJJJJ JJJJJJJJ
<2 more>
Theoretical Economics
Stock Markets
Macroeconomics
Growth Economics
Welfare Economics
Consumer Behavior
Decision Under Risk
Econometric Analysis
Savings
Agricultural Economics
<3 more>
KKKKKKKKK
LLLLL LLLLLLLLLLL
MMMMMMM
NNNNN
OOOOO O OOOOOOO
<2 more>
PPPPPPP
QQQQQQQQQ
RRRRRRR
SSSSSSSSSSS SSSSSSSS
<4 more>
Enter search term...
Search
Search for keyword in:
Fields
Paper Titles
Authors
Paper Contents
Journals
X BBBBBBBB (selection)
X Ecology (search: fields)
Top Papers for:
BBBBBBBB, ECOLOGY
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...


Ecology & Evolution
Molecular Biology
Economics
Sociology
Statistics
Anthropology
Mathematics
<10 more>
Plant Ecology
Evolutionary Ecology
Aquatic Ecology
Population Genetics
Mammology
Phylogeny
Species Diversity
Paleontology
Mathematical Ecology
Climate Change
Invertebrate Ecology
<6 more>
Theoretical Economics
Stock Markets
Macroeconomics
Growth Economics
Welfare Economics
Consumer Behavior
Decision Under Risk
Econometric Analysis
Savings
Agricultural Economics
<3 more>
AAAAAAAAAAAAAAA
BBBBBBBB
CCCC CCCCCCCCC
DDDDDDDDD
EEEEEEE
FFFFFFF F FFFFFFFFF
<5 more>
GGGGGGGGG
HHHHHHHH
IIIIIIII
JJJJJ JJJJJJJ
<2 more>
XKKKKKKKKK
LLLLL LLLLLLLLLLL
MMMMMMM
NNNNN
OOOOO O OOOOOOO
<2 more>
PPPPPPP
QQQQQQQQQ
RRRRRRR
SSSSSSSSSSS SSSSSSS
<4 more>
Enter search term...
Search
Search for keyword in:
☐ Fields
☐ Paper Titles
☑ Authors
☐ Paper Contents
☐ Journals
X BBBBBBBB (selection)
X Ecology (search: fields)
X Bergstrom (search: authors)
Top Papers for:
BBBBBBBB, ECOLOGY, BERGSTROM
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
C. Bergstrom, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
C. Bergstrom, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
C. Bergstrom, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will serve as a very brief preview of the contents of this paper...
Paper Title
F. Author, S. Author, T. Author, et. al.
Journal (2010)
Abstract: Here will appear some text from the abstract, which will

Feedback

General comments and ideas

How to handle nodes being highlighted by multiple selections/search terms?

- Also nodes being highlighted by multiple ancestor paths

Layout of page reasonable?

- Concerned about small screen sizes

Recommendations for search approaches?

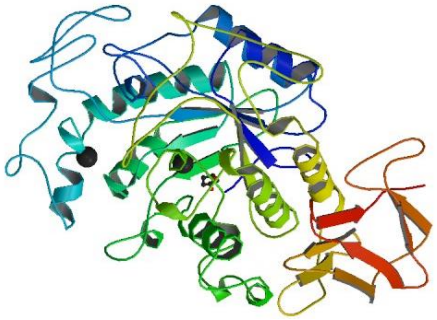
Visualizing Protein Folding Data Using Time Curves

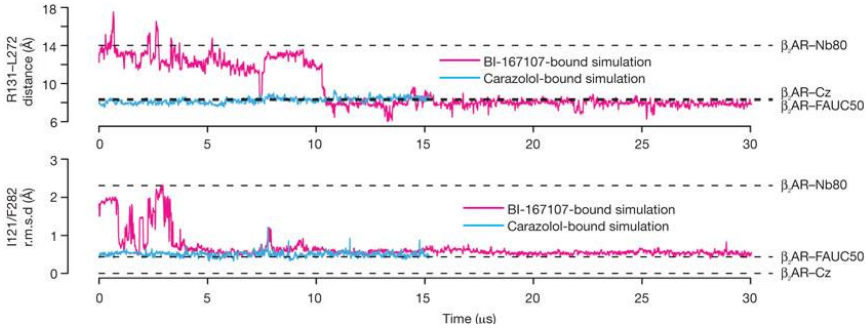
Arushi Prakash

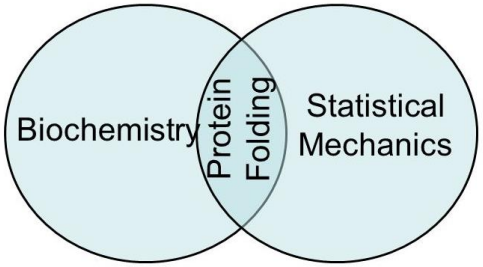
Department of Chemical Engineering,

University of Washington, Seattle

Problem Statement







Data properties

- LOTS
- Time series
- Fluctuating

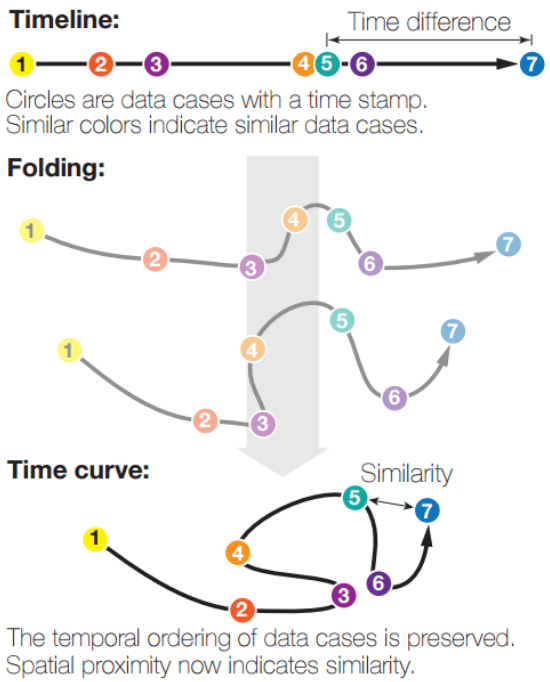
Standard visualization options

- See the video
- Plot a line curve

Problems

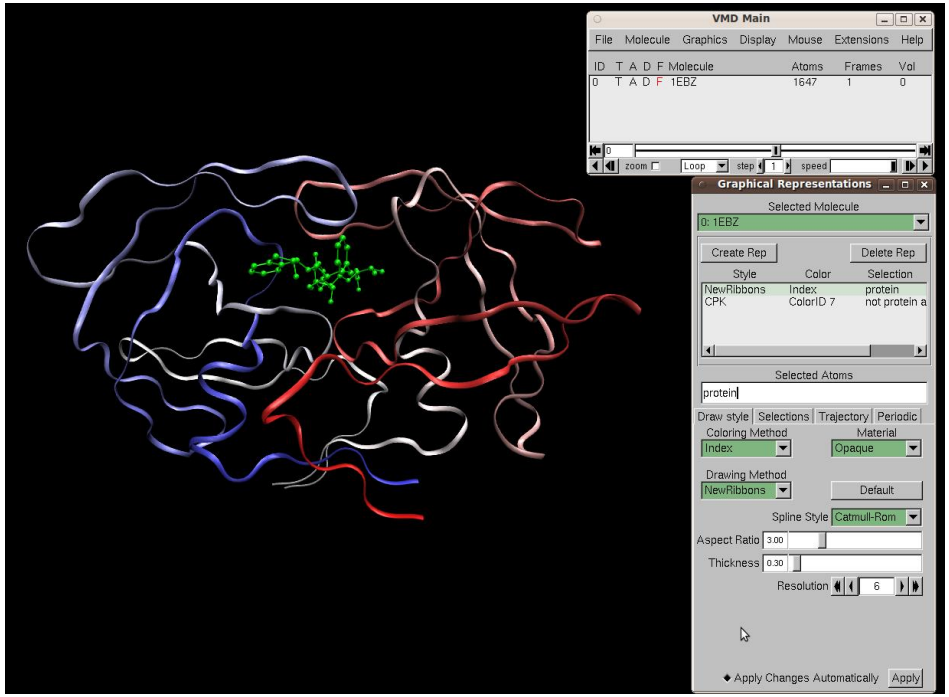
- Difficult to track by eye
- Association between variables not apparent
- Back and forth between tools

Literature Review



*Bach et al. 2015, IEEE Transactions on Visualization and Computer Graphics

Visual Molecular Dynamics



Wikipedia lists ~60 software for macromolecular visualization, including 5 web-based engines

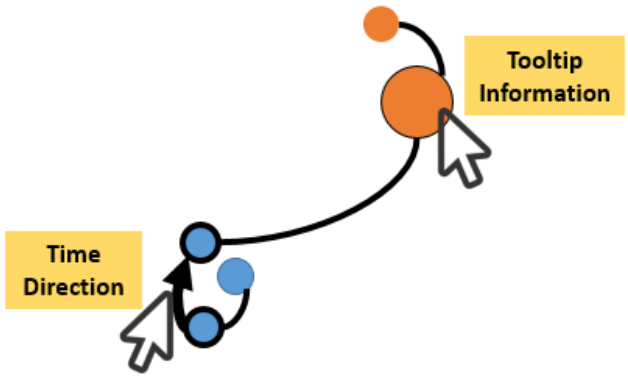
Storyboard

Load CSV

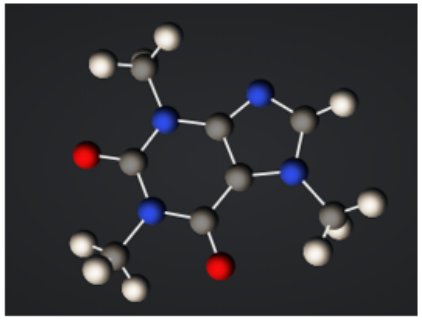
Load Protein Trajectory

Filters
Time
Variables
Range

Time Curve



3D Simulation Model



Filters
Atom
XYZ

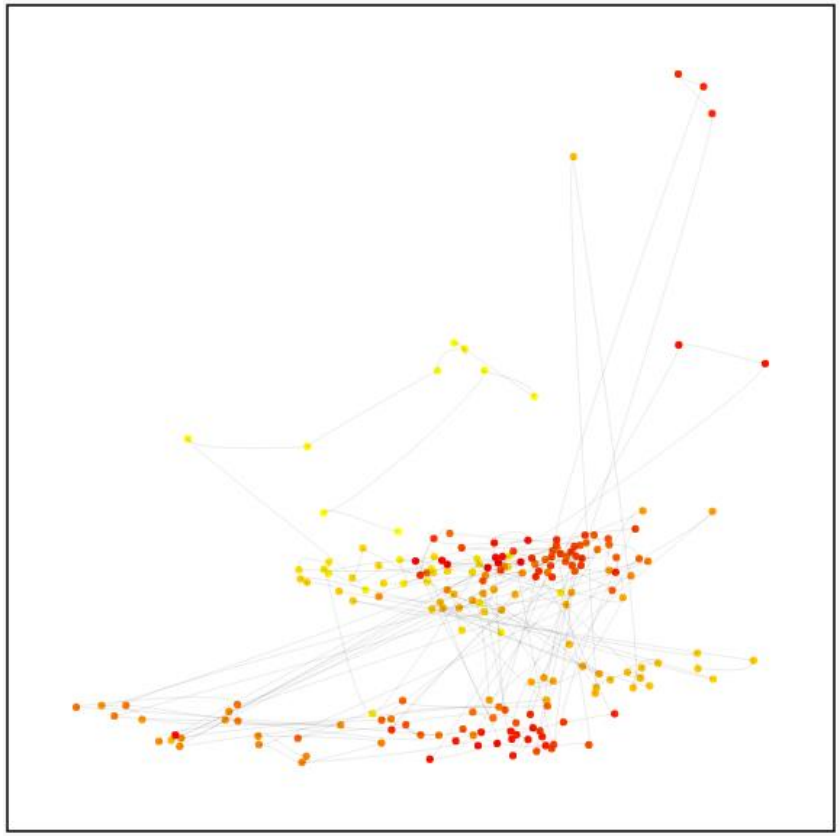
Graphics
Colors
Shape

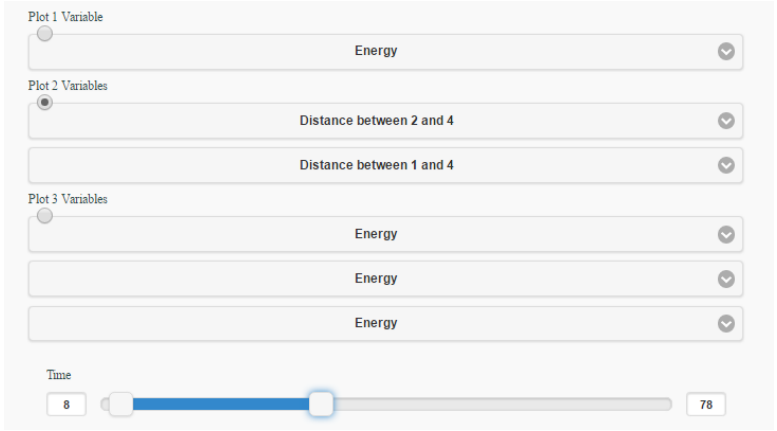
Play Movie

Feedback

- Clarity of Visualization
  - Clusters, Points, Curves
  - Contrast
- Additional interaction options
- Ways to show underlying data

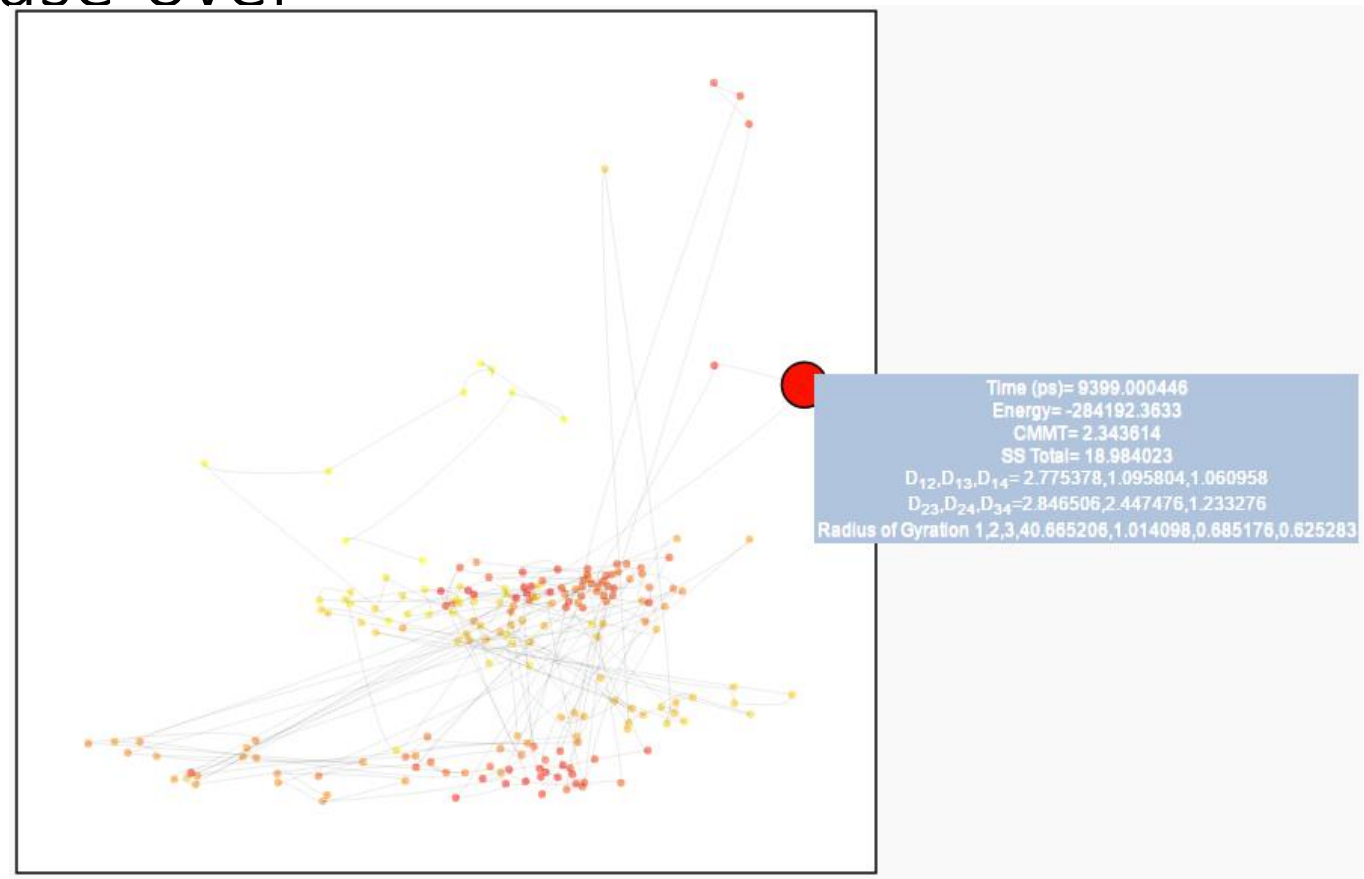
Progress





- Sum of squares distance algorithm vs Manhattan/Taxi-cab distance metric
- Multi dimensional scaling
- Points colored according to time
- Time curve faded
- Sliding bar updates the curve like a “movie”
- Option to see clustering between 1-3 variables

Mouse-over


Time (ps)= 9399.000446
Energy= -284192.3633
CMMT= 2.343614
SS Total= 18.984023
D_{12},D_{13},D_{14}= 2.775378,1.095804,1.060958
D_{23},D_{24},D_{34}=2.846506,2.447476,1.233276
Radius of Gyration 1,2,3,40.665206,1.014098,0.685176,0.625283

Feedback

- Clarity of Visualization
  - Clusters, Points, Curves
  - Contrast
- Additional interaction options
- Ways to show underlying data