IoT Visualization

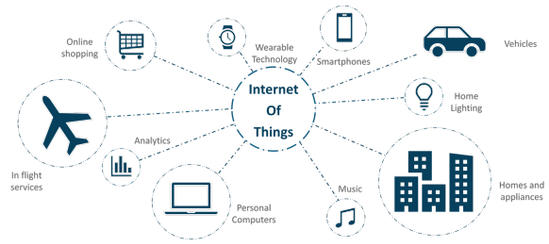
Amirhosein Abbasi
Department of Electrical and Computer Engineering
October 2019

THE UNIVERSITY OF BRITISH COLUMBIA

Why IoT

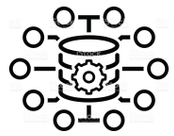
- Growing fast, impact on our life
- Industries are putting effort: Amazon, Microsoft, Intel,…
- 450 IoT platforms, Thousands of individual applications
- Different Criteria: smart home/city/transportation/…
- IoT is not growing as fast as it should be! Users are not convenient yet.

IoT Domain



IoT Data Characteristics

- Massive data: 20.4 Billion connected thing by 2020 (Data Volume)
- Real-time integration of devices (Data Velocity)
- Different Criteria= Different types of Data (Data Variety)
- The Famous “VVV”: Volume,Velocity,Variety



IoT Platforms

- A trend in IoT industry. 450 active IoT platforms are available.

- Managing things and users.
- Data Visualization: a responsibility.

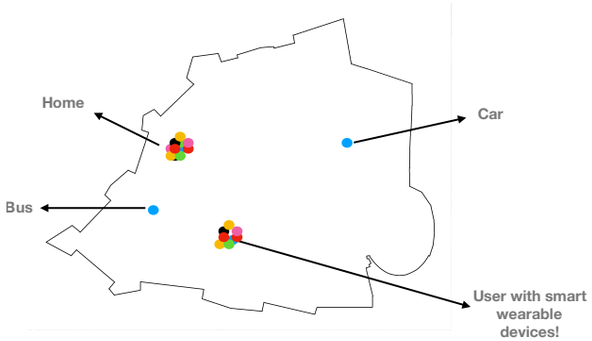
Our Scope

- IoT is a vast scope.
- Visualizing data of a specific IoT application (like visualizing healthcare data)? Good. But not solving the vast issue of IoT today.
- Lots of standards and protocols. (Solution: Using Web of Things)
- Solution: Narrow down the problem to IoT platforms.

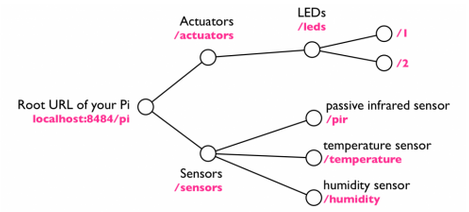
Requirements

- Number of smart things for a single user are increasing: How to keep track of all of them at once?
- Smart things are finding their way through every aspect of our lives, how to visually classify them?
- Things’ Time/location Issue: In Some devices temporal attributes are important while in some others the location is critical.
- For example: location does not make sense for a coffee maker as well as a car. Also time is more valuable for a smart street light rather than a car.

Location Issue in IoT



A typical IoT environment



Source: Building the Web of Things: book.webofthings.io
Creative Commons Attribution 4.0

To Be Done…

- Finding ways to solve time/location issue
- Visualizing the hierarchical Map of Things:
  - /agent(i)/thing(i) : CSdepartment/Room101/light2
- Visualizing smart things of a single user in a way that user can keep track of all of devices while having a sense of devices position on the hierarchy.

InsightVis

For CPSC 310

By Lucas Zamprogno and Syed Ishtiaque Ahmad

Background - The class

- CPSC 310 is a project-heavy course, and a requirement of the Computer Science Major
- Roughly 180 or 360 students per term
- Students work in pairs, meaning we have 90 to 180 teams

Background - The project

- Students are tasked to build a simple data storage and query language system
- Project is divided up into a few segments of related work called deliverables
- Each deliverable is marked by the project’s ability to pass a suite of automated tests (the details of which are not entirely known by the students)

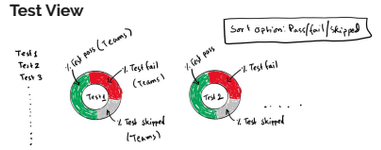
Background - The data

- We have records of test results for all the students commits (100MB for one term)
- We also have their git repositories, which means entire project histories (separately on GitHub)
- These will both take a lot of preprocessing to get out only data need, and to derive new data by combining sources

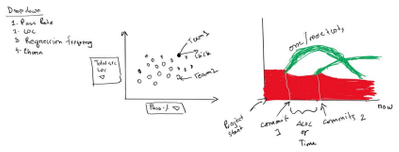
Possible questions we want to answer?

- Relationships between test cases
- Difficulty of tests
- Can we find struggling teams/ strong teams
- Bad team dynamics / Unequal contributions
- Visualize technical debt
- Time when teams are most active

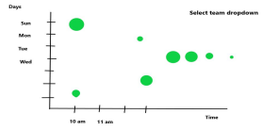
Test View



Team View



Team Activity Vis



THANK YOU!

Visualizing Protein-protein interaction networks in *Pseudomonas Aeruginosa*

CPSC 547 Project Pitch
Javier J. Castillo-Arnemann
October 8, 2019

Background: PaIntDB

- **P**seudomonas **Int**eraction **D**ata**B**ase
- Protein-protein and protein-metabolite interactions in *Pseudomonas aeruginosa* strains PAO1 and PA14. (157,427 interactions)
- *P. aeruginosa* is a multi-drug resistant pathogen involved in cystic fibrosis and other diseases. Antibiotic resistance has gotten worse and will continue to do so.
- Systems-level understanding of biological function (looking at groups of genes instead of individual genes).
- Helps visualize and interpret RNASeq Differentially Expressed genes, TnSeq phenotypically important genes, or any kind of gene list.

PaIntDB pipeline

1. Run experiment (gene knockouts, antibiotic treatment, temperature...)
2. Perform RNASeq/TnSeq.
3. Perform statistical analyses to determine genes of interest.
4. Analyze and interpret list of genes of interest.
5. Upload list to PaIntDB and generate a network of interactions between these genes.

PaIntDB

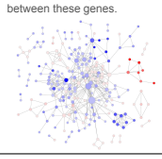
Input:

List of genes with optional expression data.

gene	baseMean	log2FoldChange	pvalue
PA1560	3486.61813214782	2.88944102878766	1.83855837198106E-83
PA1559	6153.24412240954	3.08847395382963	2.28370592812956E-82
PA2358	808.695598527297	3.01960522454999	2.71654441450156E-62
PA4775	1790.7616293644 2	2.19818907841521	2.02984086854849E-60
PA2655	645.450521683579	3.54172280077369	9.42164712779003E-59
PA4774	4271.83860737805	2.65376155361687	3.90484062480077E-57
PA4776	867.685566500117	1.91332193364473	3.54925102992402E-51
PA2357	92.5374717459836	3.13145859125876	1.19422356220102E-40
PA3554	3516.91767956598	2.06754015596055	9.04709996293866E-38
PA3559	2180.85840438509	2.11035091327087	7.90645152368141E-34
PA4773	3537.67629309535	2.28550068614544	1.04702121280 5E-32

Output:

Network showing interactions between these genes.



PaIntDB

Three network classes:

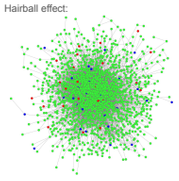
1. **BioNetwork**: basic PPI networks, no experimental data, just database info.
2. **DENetwork**: contains attributes and methods to handle differential expression data. (log2foldchange, adjusted p-values for every gene)
3. **Combined network**: additional attributes and methods to combine DE gene lists and TnSeq gene lists.

Attribute types

Network Class	Categorical	Ordered
BioNetwork	- Location - Type	- Node degree (quantitative)
DENetwork		- Log2FoldChange (quantitative, divergent) - P-value (quantitative, sequential)
Combined network	- Source of interest	

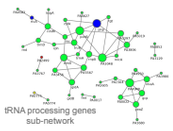
Issues

Hairball effect:



One solution:

Generate sub-networks out of functional enrichment.



tRNA processing genes sub-network

Project Goals

- Implement node clustering and expand on-demand for node-link views.
  - Cluster by network topology or by expression values? Both?
- Develop matrix view for large networks to complement the node-link view?
  - How to order the nodes in the table?

Implementation

Done:

- Python back-end for generating networks and statistical analyses.

In progress:

- Dash front-end for GUI.

For the project:

- Dash.Cytoscape library for interactive node-link network visualization.
- D3.js for matrix view?

China Multi-Generational Panel Dataset, Shuangcheng, 1866-1913

Margot Chen

What

Networks & Tables

- **1.3 million** annual observations of
- over **100,000** unique individuals descended from families,
- including ethnicity, life event, occupation, landholding...
- in Northeastern China, for the period **1866 - 1913**

Why

Present inequality over generations;

Discover other socioeconomic patterns.

How

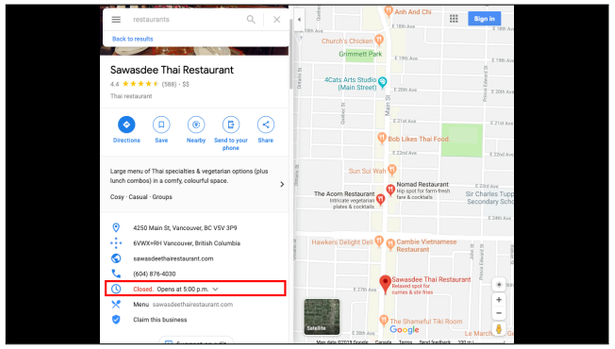
Filtering, aggregation, and navigation for networks;

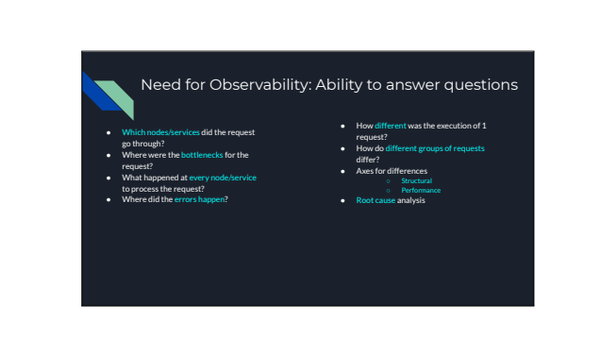
Streamgraph to show trends.

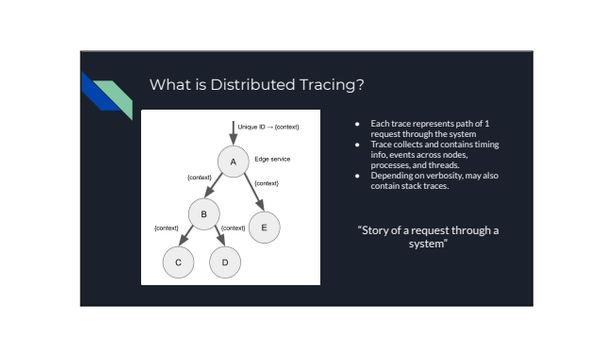
Now recruiting!

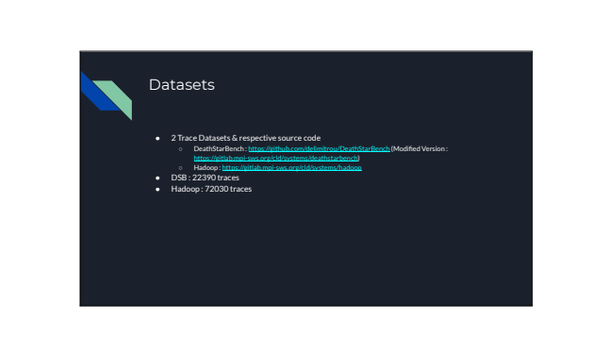
Time-based Restaurant Map

Kevin Chow
CPSC 547









Data:
- Google Maps API
- Yelp Open Dataset/API

Tech:
- Leaflet
- Polymaps
- ....

TraViz: Visualization of Distributed Traces

- Matheus Stolet
- Vaastav Anand

What are Distributed Systems?

> ***"A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable."***
>
> **- Leslie Lamport**

Distributed Systems are everywhere

- Distributed systems are widely deployed [1]
  - Graph processing
  - Stream processing
  - Distributed databases
  - Failure detectors
  - Cluster schedulers
  - Version control
  - ML frameworks
  - Blockchains
  - KV stores
  - ...

[1] Mark **Cavage**. 2013. *There's Just No Getting around It: You're Building a Distributed System*. Queue 11, 4, Pages 30 (April 2013)

Need for Observability: Ability to answer questions

- Which nodes/services did the request go through?
- Where were the bottlenecks for the request?
- What happened at every node/service to process the request?
- Where did the errors happen?
- How different was the execution of 1 request?
- How do different groups of requests differ?
- Axes for differences
  - Structural
  - Performance
- Root cause analysis

Need for Observability: Ability to answer questions

- Which nodes/services did the request go through?
- Where were the bottlenecks for the request?
- What happened at every node/service to process the request?
- Where did the errors happen?
- How different was the execution of 1 request?
- How do different groups of requests differ?
- Axes for differences
  - Structural
  - Performance
- Root cause analysis

Distributed tracing can answer these questions

What is Distributed Tracing?

- Each trace represents path of 1 request through the system
- Trace collects and contains timing info, events across nodes, processes, and threads.
- Depending on verbosity, may also contain stack traces.

"Story of a request through a system"

Datasets

- 2 Trace Datasets & respective source code
  - DeathStarBench : https://github.com/delimitrou/DeathStarBench (Modified Version : https://github.mit.edu/vaastav/DeathStarBench)
  - Hadoop : https://github.mit.edu/vaastav/hadoop
- DSB : 22390 traces
- Hadoop : 72030 traces

Tasks

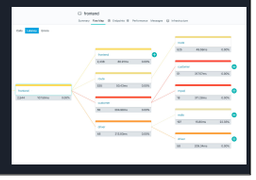
Want to support 3 different classes of tasks

- Overview Tasks
- Individual Trace Tasks
- Comparison Tasks

Overview Tasks

We want to provide general analytics on the workings of a distributed system

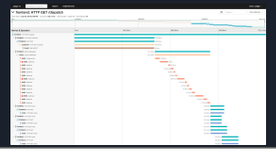
- Overall stats
- Clusters based on request types
- Src code integration
- Outliers



Individual Trace Tasks

Allows users to have a detailed view of a trace.

- Visualization of the flow of the trace
- Highlight critical path in visualization
- Highlight critical path in src code



Comparison Tasks

Want to support 3 comparison tasks

- One vs One
- One vs Many
- Many vs Many

Example comparisons

- Request type
- Day request was made
- Latency

Rumour evaluation and Sentiment Analysis of the tweets

Mona Fadaviardakani

October 2019
Department of Computer Science
University of British Columbia

Introduction and the Dataset

- I want to focus on visualizing the *tweets* posted on Twitter, from both sides of their rumour stance and the sentiment analysis.
- As of March 2011, Twitter was posting an estimated 200 million tweets per day. Tweets are now being archived at the U.S. Library of Congress. I will use the twitter dataset to pull out the tweets.

Tasks- Rumour Analysis

- I want to visualize the type of interaction between a given statement (rumourous tweet) and a reply tweet (the latter can be either direct or nested replies)
- Each tweet in the tree-structured thread will have to be categorised into one of the following four categories:
  - Support: the author of the response supports the veracity of the rumour they are responding to.
  - Deny: the author of the response denies the veracity of the rumour they are responding to.
  - Query: the author of the response asks for additional evidence in relation to the veracity of the rumour they are responding to.
  - Comment: the author of the response makes their own comment without a clear contribution to assessing the veracity of the rumour they are responding to.

Tasks- Sentiment Analysis

- Sentiment is defined as "an attitude, thought, or judgment prompted by feeling."
- My goal is having a visualization that presents basic emotional properties embodied in the text, together with a measure of the confidence in the estimates.
- We can visualize words with different emotional contents in different colours and have a global tweet label regarding its emotion

Tasks- Other Analyses

- We can have the ability to search over tweets with specific words.
- Collections of tweets can be visualized in numerous other ways:
  - by frequent terms: Common words using in the tweets of emotional regions can be cagtegorized.
  - by topic: We can have topic clusteters based on the used keywords
  - And other different ideas.
- We can encode each tweet and its attributes by different visual encodings like colour, brightness, size, and transparency.

How will the visualization solution be implemented ?

- I would like to use MAP, Timeline, Heatmap for my study
  It is useful to include interaction capabilities like zoom in our project:
  - Zoom to see detail sentiment or rumour analysis
  - We can zoom to see the whole tweet or move around the tweets which has relationships with each other to find different replies
- I would like to use D3.js and python for the visualization and NLP approaches for the sentiment and rumour analysis part.

CPSC547 - Pitch

A visualization on cybersecurity attacks & victims.

Jeffrey Goh

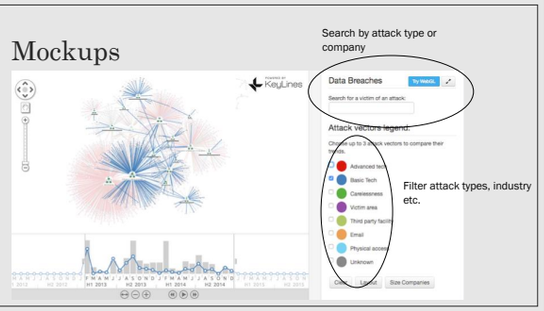
Overview

- Cyber attacks are becoming more sophisticated. New ways and methods are being invented all the time.
- It is estimated that by 2021 the annual cost from cybercrime will cost the world $6 trillion.
- 90% of motives are due to financial gains and espionage.
- Cyber security is about understanding network vulnerabilities and protecting them from cyber attacks.

Objectives

- Spotting anomalies
  - Helps prevent data breaches
  - Identify malware entry points
  - Predict likeliness of future attacks
  - Identify network vulnerabilities of an organization
- Performing forensics/analysis
  - Increase understanding and prevent reoccurrence.
  - Tracking propagation of malware

Mockups



Mockups

Outcomes

- Able to answer questions like:
  - Which industry has been breached the most in the last 5 years?
  - For the food industry, what is the top breach type? Malware? Hacking?
  - For the retail industry, what is the most compromised data? Payment? User info?
  - What are the top data assets involved in breaches? Database? POS terminals?
  - For ABC company, what type of security breach has occurred over the last 5 years.

CPSC547: InfoViz Project Pitch

Patrick Huber

University of British Columbia

huberpat@cs.ubc.ca

October 7, 2019

Background on Discourse Parsing

Goal: Reveal the underlying structure of coherent text (a discourse)

Example:

Figure: Example of a discourse tree

My Previous Research I

Discourse:

[What happened to Dunkin' Donuts?]$_1$ [Holy crap does this place suck.]$_2$ [The donuts are stale and taste weirdly like chemicals.]$_3$ [I can not recommend anything]$_4$ [except that you drive five minutes to Boss Donuts on McDowell.]$_5$ [Great donuts]$_6$ [and locally owned.]$_7$ [Support local.]$_8$

Sentiment:

Document $\equiv$ Very negative

Clauses $\equiv$ [Negative]$_1$, [Very Negative]$_2$, [Very Negative]$_3$ [Very Negative]$_4$, [Neutral]$_5$, [Very Positive]$_6$, [Positive]$_7$, [Neutral]$_8$,

Combine these two information sources using Machine Learning

My Previous Research II

[What happened to Dunkin' Donuts?]$_1$ [Holy crap does this place suck.]$_2$ [The donuts are stale and taste weirdly like chemicals.]$_3$ [I can not recommend anything]$_4$ [except that you drive five minutes to Boss Donuts on McDowell.]$_5$ [Great donuts]$_6$ [and locally owned.]$_7$ [Support local.]$_8$

The Idea I

Figure: Tool to show and compare discourse trees (2015)

The Idea II

- Existing tools only allow to compare discourse trees against a gold-standard (supervised)
- My previous research infers discourse trees from sentiment only (distant supervision)
- **Idea:** Create a visualization system, which generates insights into the alignment of the gold-label sentiment and the created discourse trees

The Idea III

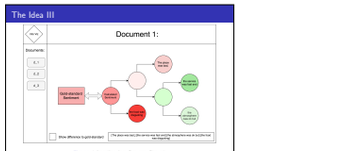


Figure: Visualization System Sketch

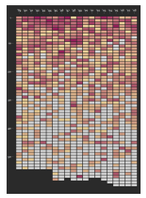
Thank you

Supporting data consolidation with visualization

Steve Kasica

Data Journalism

- **Data Journalism**: "Obtaining, reporting on, curating and publishing data in the public interest." [Stray 2011] or journalism where the data itself is the reporting byproduct.
  - Visualization is a core principle in this field.
- Example NBA Redraft from *The Pudding*



Data Wrangling

- All the stuff you **have** to do before analysis.
- Spent the summer thinking about how data journalists wrangle their data.

```
1  Reported crime in Alabama,
2  ,
3  2004,4029.3
4  2005,3900
5  2006,3937
6  2007,3974.9
7  2008,4081.9
8  ,
9  Reported crime in Alaska,
10 ,
11 2004,3370.9
12 2005,3615
13 2006,3582
14 2007,3373.9
15 2008,2928.3
16 ,
```

Coding Jupyter Notebooks

- Spent the summer doing qualitative research, open coding, on data journalism analyses like this notebook.

Journalists spend lots of time combining tables

- **Schema drift**: Periodically published data schema slightly change, or *drift*, from year to year

```
2014: dict(
      case_number="CASE_NO",
      case_status="CASE_STATUS",
      …
),
2015: dict(
      case_number="CASE_NUMBER",
      case_status="CASE_STATUS",
      …
 ),
```

Welsh, Ben. (2017, May 25). California H-2A visas analysis. *Los Angeles Times*. Retrieved from https://github.com/datadesk/california-h2a-visas-analysis

Journalists lose data in join operations (oops!)

- Example: US Refugee Analysis from *BuzzFeed News*
- Lost Wyoming in this left join
- Had to issued a correction

State	Refugees
Alabama	989
Alaska	1,231
W. Virginia	154

State	Pop
Alabama	4,800,00
Alaska	736,732
W. Virginia	1,850,326
Wyoming	584,153

Data Consolidation

- **Data consolidation**: a subfield of data integration where the user is combining fragmented, multi-year dataset.
  - Such as those periodically published by governments.
- What might be useful:
  - Visualizing transformation actions
  - Profiling underlying data

Sunset Explorer

Junfeng Xu

October 8, 2019

What is Sunset Explorer?

- A visualisation of the colour patterns of sunsets.

Inspiration

I was inspired by this article on Data Sketches:
http://www.datasketch.es/june/

Inspiration

I was inspired by this article on Data Sketches:
http://www.datasketch.es/june/

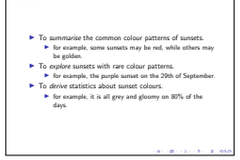
...in which the authors visualised the colour composition of Clow cards, and Taylor Swift music videos.

Inspiration

And, of course, the beautiful sunsets of Vancouver.

Inspiration

And, of course, the beautiful sunsets of Vancouver.



Data

Data

Webcam images.

Data

Webcam images.

One example is Kat Kam (www.katkam.ca).

Data

- Webcams produce consistent, unprocessed images taken from the same location.
- In the case of Kat Kam, past images are openly available online.

Tasks

Tasks

- To *summarise* the common colour patterns of sunsets.
  - for example, some sunsets may be red, while others may be golden.

Tasks

- To *summarise* the common colour patterns of sunsets.
  - for example, some sunsets may be red, while others may be golden.
- To *explore* sunsets with rare colour patterns.
  - for example, the purple sunset on the 29th of September.

Tasks

- To *summarise* the common colour patterns of sunsets.
  - for example, some sunsets may be red, while others may be golden.
- To *explore* sunsets with rare colour patterns.
  - for example, the purple sunset on the 29th of September.
- To *derive* statistics about sunset colours.
  - for example, it is all grey and gloomy on 80% of the days.

Extensions

Extensions

- Add weather data to the mix!

Extensions

- Add weather data to the mix!
- Or trends on social media?

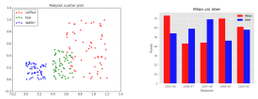
Michael Kim

-1st year master student in CS

-Taking InfoVis to express ideas in image form

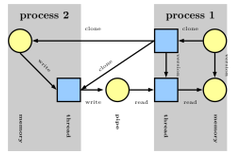
-Team with Junfeng (We have two ideas)

-5 yrs industry experience, Most used language: C, R, Python



-So we can do coding, let's make some cool thing!

CamFlow : Operating System trace log



-It is a raw trace of system in JSON format (300 MB Data per second for whole system) could be used for system statistics or ML
Ex. Pre-fetch for file, Cache management

- Asked for an advice from professor Margo Seltzer

Advantage

-High Risk, High Return
-Good project for who has an interest in OS
-We may ask an advice from relevant field participants

Shortcoming

-High Risk, High Return

Project Pitch
Or
How to prepare a 3min. Pres. in 2 minutes

Vis. BIM Data.

Koosha. M.

BIM - AR -Data vis.

Google trends

Black Mirror

Unity reflect

Microsoft Hololens, oculus, htc vive

Why BIM data vis in AR?

Do we have enough data?

BIM sensors Within UBC

Future cities.(almost today) in Architecture

Weather, usability, humidity, temperature, lighting, ventilation, eye tracking in Arch.

Tension and stress in structure, plumbing fixtures in big projects,...

What is the justified task?

How can you help?

POPULAR TED TALKS

Marjane Namavar
University of British Columbia
Department of Computer Science
Information Visualization
Fall 2019

What is the main idea?

Why working on this dataset is worthwhile?

What questions do I ask from this dataset?

What's next?

What is the main idea?

TED

kaggle

Why working on this dataset is worthwhile?

- First-class speakers and presentations
- Influence on people's lives
- Various topics
- TED prize

Do schools kill creativity? 62M

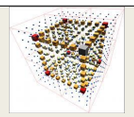
What questions do I ask from this dataset?

TED

- Why some TED talks become more popular than others?
- What are the most common occupations?
- What is the occupation leading to the most popular talks?
- What topics people are most interested in, what topics are overlooked?
- How the pattern has shifted over time?

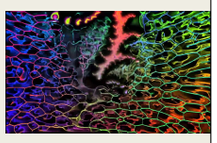
What's next?

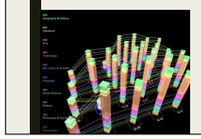
- Analyze the content of the talk
- Analyze the audio

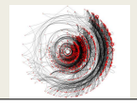


Visualization gives you answers to questions you didn't know you had.

Ben Schneiderman







Visualizing Clinical Data of Patients at the Child and Adolescent Psychiatric Emergency Unit

John-Jose Nunez
Julia Zhu
+/- Tiffany Quon

Background

- Child and Adolescent Psychiatry Unit (CAPE) only short-stay psychiatric ward in the province for 17 year old or younger patients
- Common presentation: suicidality, depression, psychosis
- Ongoing large multi-disciplinary project to collect data on patients and use for suicide prediction

Larger Project Members

- Dr. Elodie Portales-Casamar, PhD | BCCH Clinical Informatics, PI larger data project
- Dr. Ali Eslami, MD | BCCH Child Psychiatrist, PI for some parts of larger project
- Dr. Ali Mussavi Rizi | PHSA Information Technology
- Dr Raymond Ng | CS Professor
- Dr Giuseppe Carenini | CS Professor
- Sinead Nugent |BCCH Research Coordinator
- Esther Lin | Eng-phys undergrad developing the NLP pipeline

Motivation/Who

We posses a manually created database covering around 250 patients

Would like to visualize their data!

Vis would allow exploration to learn about out patients

Little previous work looking at this!

Users: hospital managers, psychiatrists, researchers

Example Questions:

- Do our patients follow expected patterns of illness eg more depression in the fall, mania in the spring?
- Does suicidal ideation/attempts increase at stressful points during the school year?
- Is medication use consistent with evidence-based guidelines?

Data/What

Items = patients = 243

Attributes (Categorical, Ordinal, Quantitative)

- Demographics (gender, age, ethnicity, postal code)
- Date and reason for admission
- Medications and dose
- History:
  - Psychiatric history (diagnoses, previous admissions)
  - Medical history (diagnoses, surgeries)
  - Substance use history
  - Social history (family structure, foster care)
- Symptoms on admission
- Various clinical scale quantifying various symptoms

Actions

- **Consume**
  - Discover- definitely!
  - Present – maybe?
  - Enjoy – no!
- **Produce**
  - Probably not yet, maybe in the future?
- **Search**
  - Explore/browse more than others, but likely all search tasks.
  - We won't be visualizing individual patients, just varying subsets
- **Query**
  - Identify, and summarize will be important. Compare will be too, unsure whether we'll need a specific compare function

Visualizing Medical Data

Julia Zhu

My Interests

- As we know data visualization is a wise investment in our future of big data.
- The nature of the massive data movement has influenced the healthcare industry to realize what a valuable tool data visualization can be when it comes to patient care:
  - Traditionally, doctors would have to sift through patient records, making it very difficult and time-consuming to spot trends
  - Just 1 patient may have up to hundreds of medical files – now imagine millions of patients and all the data they generate

My Interests

- With my background in the life sciences, I am interested in using visualization tools to recognize trends, patterns, and relationships in large volumes of health data that may not be easily seen in raw data or paper reports:
  - Visualization could also be a key process to help better predict trends in the patient's health and to improve a patient's treatment plan
  - This could further be used to identify emerging trends such that safety issues could be addressed before they become bigger problems
- Overall, the goal is to provide actionable insights that help drive change.

Clinical Data of Patients at the Child and Adolescent Psychiatric Emergency Unit (CAPE)

- For this project we wish to identify, summarize, and compare between varying patient sublets
- Eg: If there is a spike in hospital admission from May – Dec, we could focus on how different school grades make up this population and examine the possible reasons for the spike (exam stress?)

Idioms and Channels

- Idioms:
  - Line graph to show how data changes over time
  - Pie chart to show summaries of percentages as a whole
  - Nest tree diagram to toggle the patient subtypes we want to visualize data for or want to be visualizing data of
- Channels:
  - If we visualize areas of a pie chart showing diagnosis:
    - Hue could be used for diagnosis clusters
    - Luminance for severity

Dance with me

Tiffany Quon

Solo project

How might we use our data to connect with others?

Dance with me

- Minute-long experience.
- Person moves around in front of a Kinect sensor.
- Person's movement is compared with previous person's movement to create visualization of intersections.

Visualization

- Movement compared timestamp-for-timestamp
- Types of intersections:
  - 2D physical overlap → shape
  - Joint position similarity → hue
  - Depth similarity → lightness
- These shapes are "stamped" onto an image. The user can keep the final image.

Visualization

- Movement compared timestamp-for-timestamp
- Types of intersections:
  - 2D physical overlap → shape
  - Joint position similarity → hue
  - Depth similarity → lightness
- These shapes are "stamped" onto an image. The user can keep the final image.

this mockup is outdated

Project Status

- Still working towards MVP
- Able to map data to shape, hue, brightness
- Able to generate and gather data

Project Status

- Still working towards MVP
- Able to map data to shape, hue, brightness
- Able to generate and gather data

→ things look promising!

Currently visualizes intersections between current and previous person.

Currently visualizes intersections between current and previous person.

Project addition: also visualize intersections between current person and **all people.**

Overall Idea

- Continue development of vis for current vs. previous person
- Compute "average" movement across all users and visualize the current person's intersections with "overall" person
  - Time-permitting, introduce tuneable scale
    - "Compare with last 10 people"

Visualize intersections of body movement data

Visualize intersections of body movement data and how these connect us and make us feel.

Appetize
CPSC 547

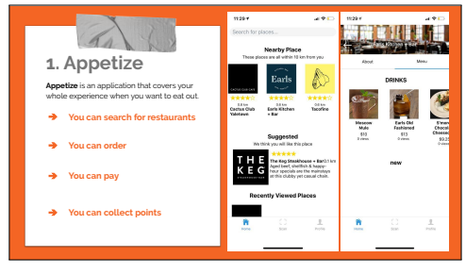
Information Visualization
Instructor: Tamara Munzner

Arya Rashtchian

Our daily tasks

We all have one specific application for our daily tasks or tasks we do more often

- Transportation
  - Google maps
- Listening to music
  - Spotify
- Restaurant
  - ? (There is a bunch of applications but none of them cover the whole experience)



What is our Value Proposition?

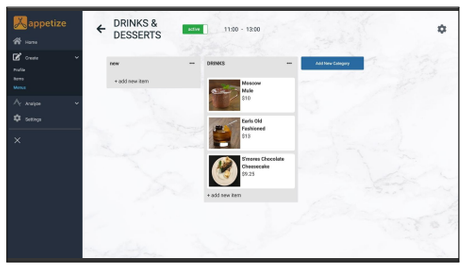
1- To provide a better experience for customers.
2- To provide remarkable insight for restaurant owners about their customers.

Better insight about customers! How?

We provide a platform for restaurant owners which:

1- Allows them to define items and menus

2- Provides helpful information about customers



What kind of information?

- Demographic data (Users have to sign up)
- We track whatever they do in the app.

The project for this course

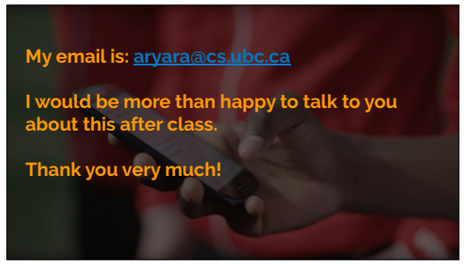
- How to aggregate these data?
- How to extract useful information?
- How to visualize these information?

An example could be:
People tend to click this part of your menu a lot more than other parts and you might as well want to reorder your menu half way into the evening to be able to sell all of your items. (we have different alternatives for info visualization)

Data

We don't have any data yet, since we don't have any customers yet.

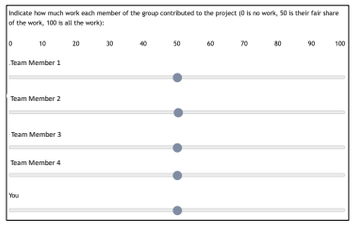
- Synthesize data



Visualizing Student Team Sentiment Reports

CPSC 547 Course Project Pitch
Nico Ritschel

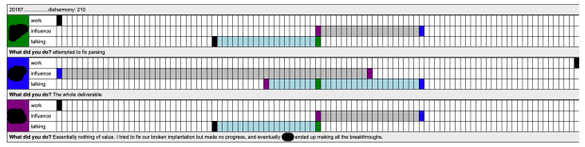
Team Sentiment Reports



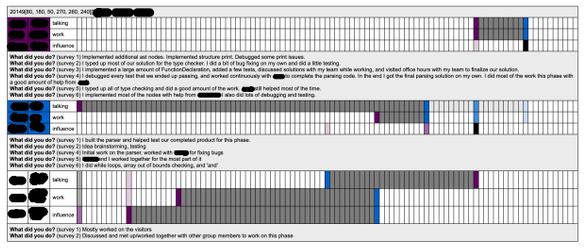
... and other questions in the same style:

- Who spoke the most?
- Who steered the team?

Current Visualization



Current Visualization



Proposed Design Study

Create an improved (or brand-new) vis for team sentiment data

- Course Instructor(s) available for consultation about their needs:
  - Elisa Baniassad (CPSC 410), Reid Holmes (CPSC 310), maybe others
- Existing data from multiple previous courses available
  - May cause privacy issues, will have to discuss this with instructors
  - Suggesting a different style of team report may be part of the results of the project
- Existing vis tool (shown in previous slides) and source code available
  - Resulting vis tool may be deployed more widely for UBC undergrad courses

Media Conglomerates: Who owns our media?

What is a **media conglomerate**?

- A large company that owns multiple smaller companies involved media enterprises

Why is this interesting?

- We consume media everyday
- Concentration of media ownership have been falling into the hands of fewer and fewer corporations

Media Conglomerates

Information on company acquisition is publicly available (Crunchbase)

How can this information be visualized in an informative and and interesting way?

- Acquisition over time
- Areas of investment (i.e. company category)
- Acquisition cost

Potential Ideas for Visualization

Tree diagram

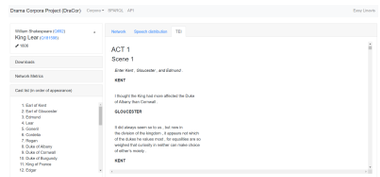
Zoomable circle packing diagram

PROJECT PITCH

Mint Tanprasert

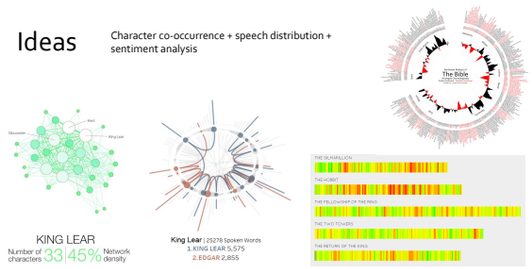
CPSC 547
Winter 2019/20 Term 1
October 8, 2019

Drama Script Visualization

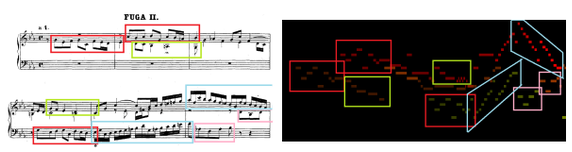


Ideas

Character co-occurrence + speech distribution + sentiment analysis



Bach's Music Visualization



Ideas

Visualize a piece in the way that makes its **structural components** and **component transformations** apparent.

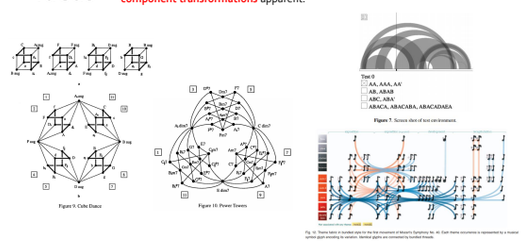


Figure 9: Cube Dance

Figure 10: Power Towers

Figure 7: Screen shot of test environment.

QUESTIONS?

Let me know if any of these ideas interest you!

VISUALIZATION OF PROVENANCE FOR PROGRAM COMPREHENSION

PROVENANCE

"chronology of the ownership, custody or location of a historical object"

PROVENANCE

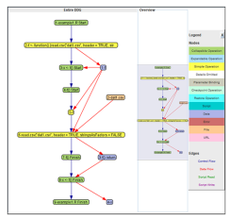
"chronology of the ownership, custody or location of a historical object"

Provides: context, verification

DIGITAL PROVENANCE

Application *Language* *System*

ISSUE: SCALE



PROGRAM COMPREHENSION

Where is a particular subroutine/procedure invoked?
What are the arguments and results of a function?
How does control flow reach a particular location?
Where is a particular variable set, used or queried?
Where is a particular variable declared?
Where is a particular data object accessed?
What are the inputs and outputs of a module?

Challenges

- **Scale of data — programs that are moderately complex can be difficult to comprehend**
- **How do we make informed subsets of data to visualize?**
- **How can this actually be used in program comprehension?**

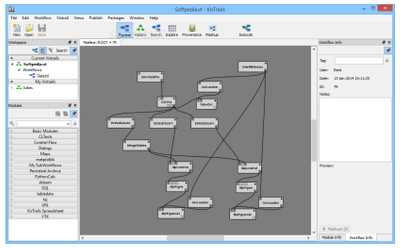
1

Related Work

- Program comprehension
- Provenance visualization
- Workflow and trace visualization

2

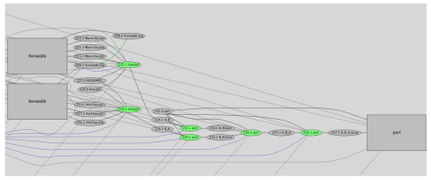
VisTrails — Workflow/Analysis visualization



https://www.earthdata.nasa.gov/esdis/eso/standards-and-practices/provenance

3

Orbiter — Visualization of System-Provenance



https://www.earthdata.nasa.gov/esdis/eso/standards-and-practices/provenance

4

InProv — A response to orbiter

https://www.earthdata.nasa.gov/esdis/eso/standards-and-practices/provenance

5

We're still looking!

We have more related work to look at to explore the full design space.

6

Defining Task Requirements

- **What tasks are important for users of provenance visualizations?**
- **What questions do we have to answer for program comprehension?**

7

Erdos & Sneed 1998

There are probably more recent program comprehension requirements!

1. Where is a particular subroutine/procedure invoked?
2. What are the arguments and results of a function?
3. How does control flow reach a particular location?
4. Where is a particular variable set, used or queried?
5. Where is a particular variable declared?
6. Where is a particular data object accessed?
7. What are the inputs and outputs of a module?

8

Requirements Analysis
Oct 5

Data and Task Abstraction
Oct 9-18

Proposed Visualization Design
Oct 18-Nov 15

Maybe user study?
Nov 15 - Early Dec

Paper writing & Drafts
During the process, but in Dec

9

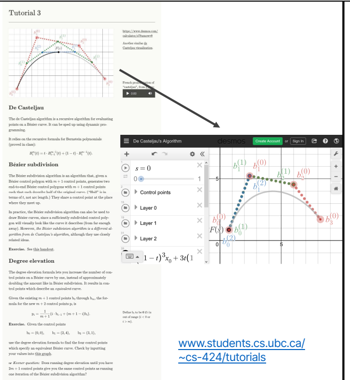
CS 547 · PROJECT PITCH

Interactive Explainers for Geometric Processing Algorithms

JERRY YIN

Current notes

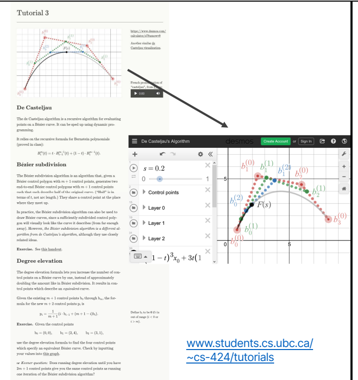
- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)



www.students.cs.ubc.ca/~cs-424/tutorials

Current notes

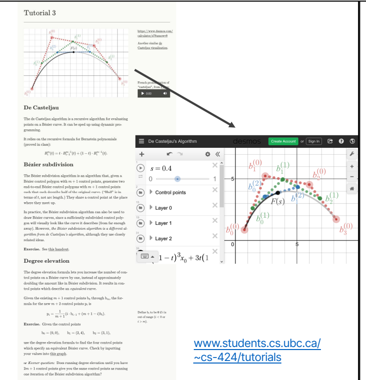
- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)



www.students.cs.ubc.ca/~cs-424/tutorials

Current notes

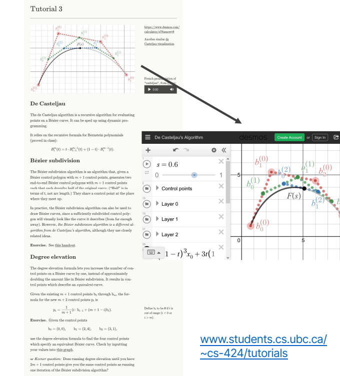
- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)



www.students.cs.ubc.ca/~cs-424/tutorials

Current notes

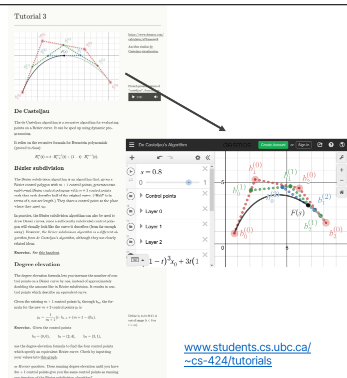
- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)



www.students.cs.ubc.ca/~cs-424/tutorials

Current notes

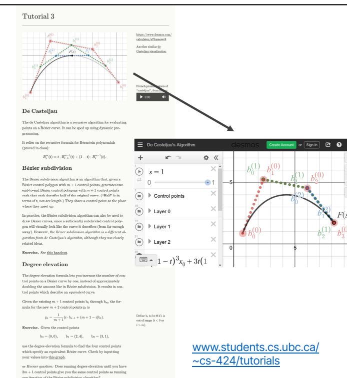
- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)



www.students.cs.ubc.ca/~cs-424/tutorials

Current notes

- Motivation: students in CPSC 424 (geometric modeling) would benefit from nice course notes
- Current notes are mostly static, with some interactive Desmos demos.
- Limitations of Desmos:
  - Poor integration with text
  - Limited to things supported by Desmos (points, lines, areas)
  - No 3D (second half of course)

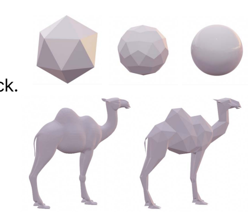


www.students.cs.ubc.ca/~cs-424/tutorials

Possible algorithms to visualize

- Things covered in 424:
  - Half-edge data structures
  - Mesh subdivision
  - Mesh simplification
- If the timing works out, the students can actually benefit from these and we can get some feedback.
- Also possible are things not covered in 424:
  - Mesh deformation
  - Point location
  - Your choice . . . ?

Might do more than one.



Technologies

- Web technologies; final result is one or more web pages
- Libraries:
  - Three.js for 3D vis
  - Possibly D3 for 2D vis?
- Still looking for group members

FINDING PATTERN OF SOCCER WORLD

VIS OF SOCCER

Wei Zheng

WHY

WHICH COUNTRY IS GOOD IN SOCCER?

- Championship does not mean everything! Many countries have high soccer level, such as the Netherlands, who have never won a World Cup.
- For the national team, in addition to the number of champions, is there any other way to see the soccer level of a country?

WHY

WHAT ARE THE KEYS TO BE A SUCCESSFUL TEAM?

- What is the key to the success of the team? Will teams with good players in every position be more successful?
- Is money the key?

WHY

WHAT ARE THE CHARACTERISTICS OF SUCCESSFUL PLAYERS?

- What is the difference between a bad player and a good player?
- Are players with high wages performing better than players with low scores?
- What are the key to their success for players in different positions, such as forward, midfielder, defender and goal keeper?

WHAT

DATASETS

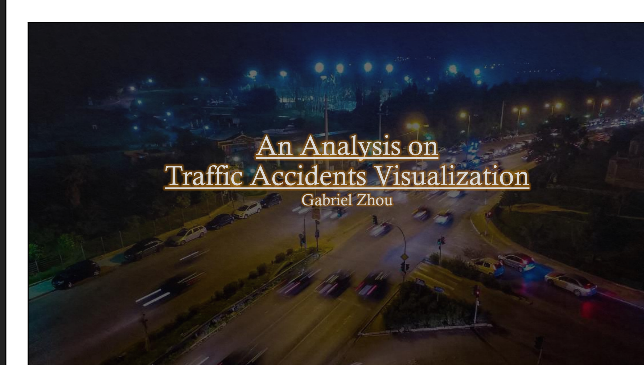
- European Soccer Database: has +25,000 matches, +10,000 players, Players and Teams' attributes, Team line up with squad formation (X, Y coordinates), etc.
- an Excel file of transfer fee among clubs from 2008 to 2017

HOW

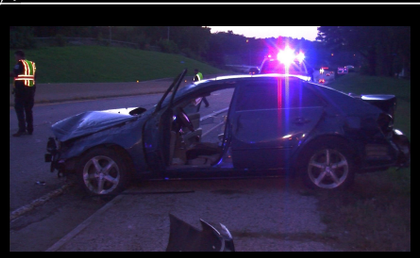
TOOLS

- Python, Pandas, Matplotlib, Seaborn
- may be Tableau

THANK YOU!



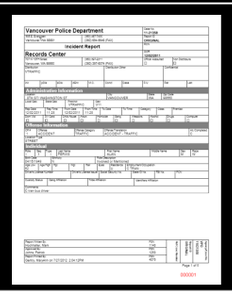
Why?



Why?

- Identify high risk locations
- Identify peak time period of accidents?
- Relationship between accidents and drivers?
- Relationship between accidents and vehicles?

What?



What?

- Location
- Date and Time
- Damage
- Age
- Gender
- Driving Experience, Brand, Model, Year of Make, etc……

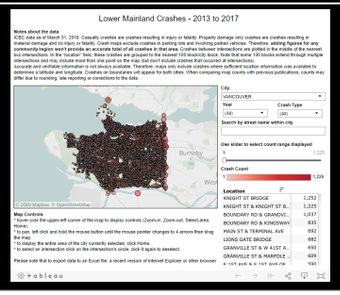
What?

ICBC

NHTSA
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION

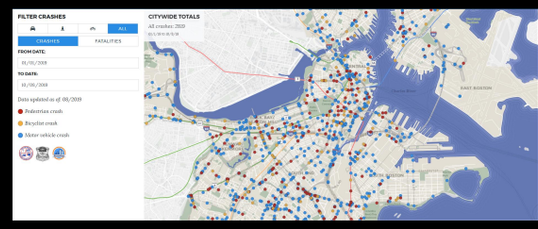
Government of Canada

How?

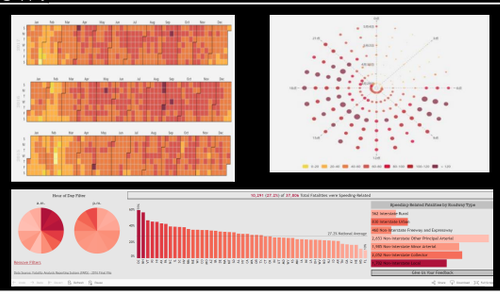


How?

How?



How?



TBD

- Select 2 or 3 visualization tools
- A unified dataset

Reference

1. A. Fang, X. Peng, J. Zhou and L. Tang, "Research on the Map-matching and Spatial-temporal Visualization of Expressway Traffic Accident Information," *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, Singapore, 2018, pp. 23-27. doi: 10.1109/ICITE.2018.8492572
2. ICBC. Lower Mainland Crash. Retrieved from: https://public.tableau.com/profile/icbc#!/vizhome/LowerMainlandCrashes/LMDashboard
3. Data4Democracy. Crash-Model. Retrieved from: https://github.com/Data4Democracy/crash-model
4. City of Boston. Vision Zero. Retrieved from: https://www.boston.gov/transportation/vision-zero

THANK YOU

DNA* Sequencing Vis

Exploring sequencing structural noise

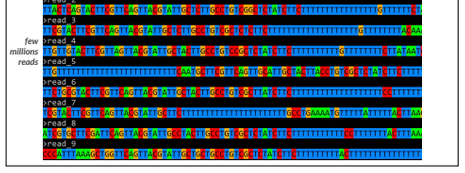
* RNA actually...

Baraa Orabi pitch for CPSC547

DNA sequencing *ideally*

DNA sequencing ~~*ideally*~~ reality

Data (reads raw)



Data (reads mapped on target)

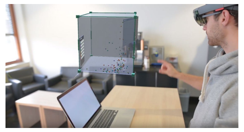
read-id	read-length	start-on-read	end-on-read	target-id	target-length	start-on-target	end-on-target
read_1	1636	77	1004	ENSG00000101608	1265	326	1229
read_1	1636	78	936	ENSG00000101608	935	65	899
read_1	1636	78	936	ENSG00000101608	935	65	899
read_1	1636	88	936	ENSG00000101608	983	121	944
read_1	1636	109	936	ENSG00000101608	1040	194	998
read_2	1321	72	883	ENSG00000137714	3180	151	965
read_3	523	74	299	ENSG00000174669	2298	2036	2262
read_3	523	74	299	ENSG00000174669	2491	2229	2455
read_3	523	81	299	ENSG00000174669	2545	2285	2506
read_3	523	81	299	ENSG00000174669	2416	2155	2376
read_3	523	81	299	ENSG00000174669	2541	2280	2501
read_3	523	81	299	ENSG00000174669	2308	2048	2269
read_3	523	81	299	ENSG00000174669	2510	2249	2470
read_3	523	299	462	ENSG00000174669	2491	2263	2431
read_3	523	299	462	ENSG00000174669	2545	2319	2487
read_3	523	299	462	ENSG00000174669	2416	2189	2357
read_3	523	299	462	ENSG00000174669	2510	2283	2451
read_3	523	299	462	ENSG00000174669	2541	2314	2482
read_3	523	299	462	ENSG00000174669	2298	2070	2238

End aim of informing:

Your quick Q's to me

- Last time I had biology was in 10th grade, is that OK?!
  - Absolutely yes!
- Do you *already* have the data?
  - Yes! I have 2 private and >30 public datasets that I dabbled with for +6 months
- What are you bringing to the table?
  - Data, problem, few years in bioinformatics data experience, and a CS degree worth of programming skills
- What are you looking for in partners (in no particular order)?
  - Decent-ish experience in vis programming and/or design,
  - Some enthusiasm for bioinformatics
  - A dash of awesomeness!

Project Pitch
Or
How to prepare a 3min. Pres. in 2 minutes



Vis. BIM Data. Koosha. M.

BIM - AR -Data vis.

Google trends

Black Mirror

Unity reflect

Microsoft Hololens, oculus, htc vive

Why BIM data vis in AR?

https://www.youtube.com/watch?v=rj-m2SItDl4

https://www.youtube.com/watch?v=muQ_8QyBYVg

https://www.youtube.com/watch?v=TQcyS3BVriq

https://www.youtube.com/watch?v=u76ww3NJFqE

Do we have enough data?

BIM sensors Within UBC

Future cities.(almost today) in Architecture

Weather, usability, humidity, temperature, lighting, ventilation, eye tracking in Arch.

Tension and stress in structure, plumbing fixtures in big projects,...

What is the justified task?

How can you help?

Interactivity and Learning Effectiveness

Youssef Sherif

CPSC 547
October 8, 2019

What does interactivity do for learning effectiveness?

- Triggers readers' intellectual curiosity
- Increases their motivation to learn more

Examples

Link 1

Link 2

Factors

- The area of knowledge. Some areas of knowledge benefit more than others
- Whether exploration is constrained or not. Constrained exploration were found to improve learning effectiveness
- Need to research for more factors

Plan

- Get already existing visualizations from learning blogs that might benefit from adding interactivity based on our previous research
- Try to replicate the visualization but with added interactivity
- Perform a controlled experiment on few visualization examples before and after adding interactivity and check whether learning effectiveness improves

Thank you

Background

- I have extensive connections into numerous restaurants
- The motivations behind menuVis are things I have noticed and spoken about with head chefs
- Currently there is a gap in the market for menu creation support

Motivation

- Head chefs and kitchen managers must juggle cost of ingredients and revenue
- Owners want low cost & high sales
- Chefs want high quality ingredients & successful dishes
- Creating or adjusting existing menus is an iterative time-inefficient process

Current Practices

- There exists support for design and layout of a menu
- Insufficient support in determining cost benefit analyses of potential menu items
- Most chefs keep their own log of ingredient costs, sales (by season), recipes, and ordering schedules

Menu Creation Logistics

- Re-use ingredients across dishes
- Leverage seasonal (cheaper) ingredients
- Remove or adjust poorly selling or high-cost dishes
- Sales are location and season dependant
- Adjust menus twice a year (Fall/Winter & Spring/Summer)

Goals

- Cohesive view of disparate data kept across different files held by chefs
- Efficiently create menus using ingredient costs, recipes, and sales, leveraged against revenue goals
- An app ready for testing/deployment into the wild
  (I have chefs who are willing to try it out!)
- Stretch goal: potentially sell to Sysco
  (North America's #1 food distributor to restaurants)

Thank you!

(lets make money)