IoT Platforms

- A trend in IoT industry. 450 active IoT platforms are available.
- Managing things and users.
- Data Visualization: a responsibility.

Why IoT

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

IoT Domain

- Massive data: 20.4 Billion connected things 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 Data Characteristics

- Massive data: 20.4 Billion connected things 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

A typical IoT environment

- Things' Time/location Issue: In some cases, how to visually classify them?
- 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.

Our Scope

- IoT is a vast scope.

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 cases, how to visually classify them?
- 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

- Visualizing the hierarchical Map of Things:
- Time when teams are most active
- To Be Done...
- Finding ways to solve time/location issue
- 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.

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

InsightVis

- By Lucas Zamprogno and Syed Ishtiaque Ahmad
- For CPSC 310

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)
- Roughly 180 or 360 students per term
- We have records of test results for all the students commits (100MB for one term)

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
- These will both take a lot of preprocessing to get out only data need, and to derive new data by combining sources

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

To Be Done...

- Finding ways to solve time/location issue
- Visualizing the hierarchical Map of Things:
  - /agent/)(thing) : 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.
Background: PaintDB
- DENetwork: contains attributes and methods to combine DE gene lists and TnSeq gene lists.
- Combined network: additional attributes and methods to combine DE gene lists and TnSeq gene lists.
PaIntDB pipeline:
1. Run experiment (gene knockouts, antibiotic treatment, temperature...)
2. Perform RNASeq/TnSeq.
3. Perform statistical analyses to determine genes of interest.
4. Cluster by network topology or by expression values? Both?
5. Upload list to PaIntDB and generate a network of interactions between these genes.
PaIntDB:
Input:
List of genes with optional expression data.
Output:
Network showing interactions between these genes.
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:
- Node degree (quantitative)
- Log2FoldChange (quantitative, divergent)
- P-value (quantitative, sequential)
- Source of interest

Networks & Tables:
What:
- Networks:
  - Present inequality over generations.
  - Discover other socioeconomic patterns.
- Tables:
  - China Multi-Generational Panel Dataset, Shuangcheng, 1866-1913
  - Margot Chen

Why:
- Filters, aggregation, and navigation for network views.
- Streamgraph to show trends.

How:
- Python back-end for generating networks and statistical analyses.
- Dash front-end for GUI.
- Dash-Cytoscape library for interactive node-link network visualization.
- PaIntDB pipeline:
  - Run experiment (gene knockouts, antibiotic treatment, temperature...)
  - Perform RNASeq/TnSeq.
  - Perform statistical analyses to determine genes of interest.
  - Cluster by network topology or by expression values? Both?
  - Upload list to PaIntDB and generate a network of interactions between these genes.
  - Dash.Cytoscape library for interactive node-link network visualization.
  - D3.js for matrix view?
  - Systems-level understanding of biological function (looking at groups of genes instead of individual genes).
Now recruiting!

Time-based Restaurant Map
Kevin Chow
CPSC 54

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

Tech:
- Leaflet
- Polymaps
- ...

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 Systems are everywhere
- Graph processing
- Stream processing
- Distributed databases
- Failure detectors
- Cluster schedulers
- Version control
- ML frameworks
- Blockchains
- KV stores
- ...

Datasets
- 2 Trace Datasets & respective source code
  - DeathStarBench: https://github.com/delimitrou/DeathStarBench
    - Modified Version: https://gitlab.mpi-sws.org/cld/systems/deathstarbench
  - Hadoop: https://gitlab.mpi-sws.org/cld/systems/hadoop
- DSB: 22390 traces
- Hadoop: 72030 traces

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

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.

Distributed tracing can answer these questions

TreViz: Visualization of Distributed Traces

- Matthew Stolet
- Vaastav Anand
Rumour evaluation and Sentiment Analysis of the tweets

Mona Fadaviardakani

October 2019
Department of Computer Science
University of British Columbia

Tasks
- Want to support 3 different classes of tasks
  - Overview Tasks
  - Individual Trace Tasks
  - Comparison Tasks

Overview Tasks
- Overview of the different tasks
- Sitemap
- Functionalities
- Features

Individual Trace Tasks
- Visualisation of the tasks of interest
- Filtering of the tasks
- Sitemap

Comparison Tasks
- Main aspects comparison
- Filters
- Example comparison
- Sitemap

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 300 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.

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.

How will the visualization solution be implemented?
- 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 categorized
  - by topic: the we can have topic clusters 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.

Mockups
- Search by attack types or category
- Overview
- Individual Trace
- Comparison

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.
- 50% of motives are due to financial gains and espionage.
- Cyber security is about understanding network vulnerabilities and protecting them from cyber attacks.

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.
My Previous Research II

1. What happened to Dunkin’ Donuts?

2. Holy crap does this place suck.

3. The donuts are stale and taste weirdly like chemicals.


5. Great donuts

6. and locally owned.

7. What might be useful:

   - Visualizing transformation actions
   - Profiling underlying data

8. Patrick Huber (UBC)

   - Journalists spend lots of time combining tables
   - Schema drift: Periodically published or journalism where the data schema slightly change, or journalism analyses like this notebook.

   - What might be useful:
     - Visualizing transformation actions
     - Profiling underlying data

   - Patrick Huber (UBC)
What is Sunset Explorer?

Tasks
▶ To summarise the common colour patterns of sunsets.
▶ for example, some sunsets may be red, while others may be golden.
▶ To derive statistics about sunset colours.
▶ for example, it is all grey and gloomy on 80% of the days.
▶ Webcams produce consistent, unprocessed images taken from the same location.
▶ In the case of Kat Kam, past images are openly available online.
▶ Add weather data to the mix!
▶ Or trends on social media?

Data
▶ Webcam images.
▶ One example is Kat Kam (www.katkam.ca).

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

Inspiration
And, of course, the beautiful sunsets of Vancouver.
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.

In the case of Kat Kam, past images are openly available online.
To explore webcam images. To derive statistics about sunset colours.
for example, it is all grey and gloomy on 80% of the days.
for example, some sunsets may be red, while others may be golden.
for example, the purple sunset on the 29th of September.
for example, the common colour patterns of sunsets.
the common colour patterns of sunsets.

Statistics about sunset colours.
Sunsets with rare colour patterns.
The common colour patterns of sunsets.

Days.
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

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

BIM - AR - Data vis.

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

Why BIM data vis in AR?

- Why working on this dataset is worthwhile?
- What questions do I ask from this dataset?
- What’s next?

Why working on this dataset is worthwhile?

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

олько на русском

What’s next?

- Analyze the content of the talk
- Analyze the audio

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

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

John-Jose Nunez
Julia Zhu
Tiffany Quon

WHAT'S NEXT?

Why some TED talks become more popular than others?
- Why working on this dataset is worthwhile?
- What questions do I ask from this dataset?
- What’s next?
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 subtexts
- 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?)

Larger Project Members

- Dr. Elodie Portales-Casamar, PhD | BCCH Clinical Informatics, PI larger data project
- Dr. Giuseppe Carenini | CS Professor
- Dr. Ali Eslami, MD | BCCH Child Psychiatrist
- Dr Raymond Ng | CS Professor
- Sinead Nugent | BCCH Research Coordinator
- Julia Zhu | Eng-phys undergrad developing the NLP pipeline

Motivation/Who

- 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 realise 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

Data/What

- Attributes: (Categorical, Ordinal, Quantitative)
  - Demographics (gender, age, ethnicity, postal code)
  - Date and reason for admission
  - History:
    - Psychiatric history (diagnosis, previous admissions)
    - Medical history (diagnoses, surgeries)
    - Social history (family structure, foster care)
  - Symptoms: an ad-hoc list of various symptoms

Visualizing Medical Data

- Julia Zhu

Idioms and Channels

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

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 a visualization of intersections.

- Solo project
  - Dance with me
    - Three shapes are "stamped" onto an image and can be moved
    - These shapes are "happening" over time and are visualized in final image
**Project Status**
- Still working towards MVP
- Able to map data to shape, hue, brightness
- Able to generate and gather data

**Overall Idea**
- Visualize intersections of body movement data
- Visualize intersections of body movement data and how these connect us and make us feel.

**Our daily tasks**
We all have one specific application for our daily tasks or tasks we do more often:
- Transportation
  - Google maps
  - Waze
- Listening to music
  - Spotify
- Restaurant
  - ? (There is a bunch of applications but none of them cover the whole experience)

**Appetize**
Appetize is an application that covers your whole experience when you want to eat out.
- You can search for restaurants
- You can order
- You can pay
- You can collect points

**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

**CPSC 547 Information Visualization**
Instructor: Tamara Munzner
Arya Rashtchian
My email is: aryara@cs.ubc.ca
I would be more than happy to talk to you about this after class.
Thank you very much!

Visualizing Student Team Sentiment Reports

CPSC 547 Course Project Pitch
Nico Ritschel

Current Visualization

Team Sentiment Reports

... and other questions in the same style:
- Who spoke the most?
- Who steered the team?

Visualizing Student Team Sentiment Reports

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

Current Visualization

Proposed Design Study

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Current Visualization

Media Conglomerates

- Who owns our media?

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: Who owns our media?

Information on company acquisition is publicly available (Crunchbase)

How can this information be visualized in an informative 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

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.

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”

Provides: context, verification

**Digital Provenance**

**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?

**Related Work**

- Program comprehension
- Provenance visualization
- Workflow and trace visualization

Defining Task Requirements

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

**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?

**Requirements Analysis**

Oct 5
Data and Task Abstraction
Oct 9-15
Proposed Visualization Design
Oct 18-Nov 15
Maybe user study?
Nov 15 - Early Dec
Paper writing & Drafts
During the process, but in Dec

**VisTrails — Workflow/Analysis Visualization**

**Orbiter — Visualization of System-Provenance**

**InProv — A response to orbiter**

We’re still looking!
We have more related work to look at to explore the full design space.

**CS 547 · Project Pitch**

**Interactive Explainers for Geometric Processing Algorithms**

**Jerry Yin**
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?

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?
DNA sequencing ideally

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

DNA sequencing ideally

- Generate sequencing "reads"

Data (reads raw)

- Few millions of reads
- Few hundreds/thousands characters

How?

- Select 2 or 3 visualization tools
- A unified dataset

TBD

What?

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

Reference


Thank you
**Factors**

- The area of knowledge: Some areas of knowledge benefit more than others.
- Whether exploration is constrained or not: Constrained exploration was 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.
- Perform a controlled experiment on few visualization examples before and after adding interactivity and check whether learning effectiveness improves.

**Examples**

- MenuVis: Menu creation visualization tool for chefs.
- Silver Burla: Downstream bioinformatics.

**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!
(let's make money)