**IoT Visualization**

A typical IoT environment

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

**Why IoT**

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

**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.

**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.

**Location Issue in IoT**

**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 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)

**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**

For CPSC 310

By Lucas Zamprogno and Syed Ishtiaque Ahmad

**The famous "VVV"**

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

**Background - The project**

- Students are tasked to build a simple data storage and query language system
- Project is testing the project’s ability to pass a suite of automated tests (the details of which are not entirely known by the students)

**InsightVis**

By Lucas Zamprogno and Syed Ishtiaque Ahmad

Roughly 180 or 360 students per term

CPSC 310 is a project-heavy course, and a requirement of the Computer Science Major

Bad team dynamics / Unequal contributions

Relationships between test cases

Difficulty of tests

Can we find struggling teams/strong teams

Visualize technical debt

Time when teams are most active
Perform RNASeq/TnSeq. Dash front-end for GUI.

Protein-protein and protein-metabolite interactions in P. aeruginosa. Core P. aeruginosa proteome and interactome datasets are included in the datasets included in the dataset. The dataset is also used for the development matrix view for large networks to complement the node-link view.

Implement node clustering and expand on-demand for node-link views. Dash.Cytoscape library for interactive node-link network visualization. D3.js for matrix view?

Python back-end for generating networks and statistical analyses. Analyze and interpret list of genes of interest. Helps visualize and interpret RNASeq Differentially Expressed genes, TnSeq phenotypically important genes, or any kind of gene list.

Why
Present inequality over generations; Discover other socioeconomic patterns.

How
Filtration, aggregation, and navigation for network views; Streamgraph to show trends.

Background: PaIntDB
- Pseudomonas Interaction Database
- Pseudomonas proteome and protein metabolite interactions in Pseudomonas aeruginosa strains PAO1 and PA14. (157,427 interactions)
- P. aeruginosa is a multi-drug resistance 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.
- Filter functionally related 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 to PaIntDB and generate a network of interactions between these genes.
6. Output:
   - Network showing interactions between these genes.
   - Network showing interactions between these genes.
   - Network showing interactions between these genes.

PaIntDB
Input:
List of genes with optional expression data.
Output:
Network showing interactions between these genes.

PaIntDB
Three network classes:
2. DENetwork: contains attributes and methods to handle differential expression data. (Log2FoldChange, adjusted p-values for every gene)

Attribute types
<table>
<thead>
<tr>
<th>Network Class</th>
<th>Categorical</th>
<th>Ordered</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioNetwork</td>
<td>Location</td>
<td>No</td>
<td>Node degree (quantitative)</td>
</tr>
<tr>
<td>DENetwork</td>
<td>Type</td>
<td>Yes</td>
<td>Log2FoldChange, adjusted p-values for every gene</td>
</tr>
<tr>
<td>Combined network</td>
<td>Source of interest</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Issues
One solution:
- Generate sub-networks out of functional enrichment.
- Cluster by network topology or by expression values in PaIntDB.
- Develop realistic use case for large networks to complement the web interface.
- How to order the nodes in the table?

Project Goals
- Implement node clustering and expand on-demand for node-link views.
- Filter functionally related genes, or any kind of gene list.
- Develop realistic use case for large networks to complement the web interface.
- How to order the nodes in the table?

Implementation
- Python back-end for generating networks and statistical analyses.
- In progress:
  - Dash front-end for GUI
  - Dash.Cytoscape library for interactive node-link network visualization.
  - D3.js for matrix views

What
Networks & Tables
- 1.3 million annual observations of families.
- over 100,000 unique individuals descended from families.
- including ethnicity, life event, occupation, and relationships.
- in Northeastern China, for the period 1865 - 1913

THANK YOU!
Now recruiting!

Time-based Restaurant Map
Kevin Chow
CPSC 547

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

Tech:
- Leaflet
- Polymaps
- ...

What is Distributed Tracing?
- Each trace represents the 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.

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

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

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
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
Pros of a visually interactive and animated visualization of cyber data
- Hardwares
- Softwares
- Time series graphs
- Bar charts
- Maps

Individual Trace Tasks
Pros of a visually interactive and animated visualization of cyber data
- Visualisation of the flow of traces
- Hollywood-style timelines
- Node-link diagrams

Comparison Tasks
Pros of a visually interactive and animated visualization of cyber data
- Step-by-step traces
- Example comparisons
- Formative and summative
- Metrics

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.

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 detailed sentiment or rumour analysis
  - We can zoom to see the whole tweet or move around the tweets which has re-tweeted and retweeted to different replies
- I would like to use D3.js and python for the visualization and NLP approaches for the sentiment and rumour analysis part.

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:
- The author of the response supports the veracity of the rumour they are responding to.
- The author of the response does not support the veracity of the rumour they are responding to. The author of the response does not support the veracity of the rumour they are responding to.
- The author of the response does not support the veracity of the rumour they are responding to.
- The author of the response does not support 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 embedded 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.

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.
The Idea I
- My Previous Research I
- Background on Discourse Parsing

Example:
- Goal: Reveal the underlying structure of coherent text (a discourse)
- What might be useful:
  - Visualizing transformation actions
  - Profiling underlying data

Data Consolidation
- Data consolidation: a subfield of data integration where the user is combining fragmented, multi-year dataset.
- What to do before analysis:
  - Obtaining, restructuring, aligning, reformatting data
  - “Obtaining, restructuring, aligning, reformatting data” from the public records

Data Wrangling
- All the stuff you have to do before analysis.
- Spent the summer thinking about how data journalists wrangle their data.
- Example: US Refugee Analysis from BuzzFeed News
  - Lost Wyoming in this left join
  - Had to issue a correction

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

Supporting data consolidation with visualization
- Steve Kasica

Thank you
- Patrick Huber
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.

And, of course, the beautiful sunsets of Vancouver.

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

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.

To summarise the common colour patterns of sunsets.
For example, the purple sunset on the 29th of September.

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

To explore rare colour patterns of sunsets.
To explore sunsets with rare colour patterns.

Data
Webcam images.

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.

Inspiration
And, of course, the beautiful sunsets of Vancouver.
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 an advice from relevant field participants

Shortcoming
- High Risk, High Return
- 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

Why BIM data vis in AR?
- BIM sensors within UBC
- Future cities (almost today) in Architecture
- Weather, usability, humidity, temperature, lighting, ventilation, eye tracking in AR
- Tension and stress in structure, plumbing fixtures in big projects...

Do we have enough data?

What is the justified task?

How can you help?

What is the main idea?

Why working on this dataset is worthwhile?

What questions do I ask from this dataset?

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
Visualizing Medical Data

**Idioms and Channels**

- **Idioms:**
  - Line graph to show how data changes over time.
  - Pie chart to show summaries of percentages as a whole.

- **Channels:**
  - We visualize areas of a pie chart showing diagnoses.
  - Hue could be used for diagnostic clusters.
  - Luminance for severity.

**Background**

- Child and Adolescent Psychiatry Unit (CAPE) only short-stay psychiatric ward in the province for 17-year-old or younger patients.
- Current-generation visual analytics, depression, suicide prevention.
- Ongoing large multi-disciplinary project to collect data on patients and use for suicide prediction.
- 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 incorporation of data into daily activities.
- 1 patient may have up to hundreds of medical files—now imagine millions of patients and all the data they generate.

**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 admissions 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?)

**Motivation/Who**

- We possess a manually created database covering around 250 patients.
- Would like to visualize their data.
- We would allow exploration to learn about our patients.
- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**Data/What**

- Would allow exploration to learn about our patients.
- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**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.
- Overall, the goal is to provide actionable insights that help drive change.

**Visualization**

- Visualization involves interpreting data visually.
- Types of visualization:
  - 2D graphical images
  - 3D graphical images
  - Maps
  - Interactive visualizations

**Solo project**

- **Dance with me**
  - Tiffany Quon
  - Dance with me
  - Solo project

- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**How might we use our data to connect with others?**

- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**Data/What**

- Would like to visualize their data.
- We would allow exploration to learn about our patients.
- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**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 admissions 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?)

**Motivation/Who**

- We possess a manually created database covering around 250 patients.
- Would like to visualize their data.
- We would allow exploration to learn about our patients.
- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**Data/What**

- Would allow exploration to learn about our patients.
- Little previous work looking at this.
- Users: hospital managers, psychiatrists, researchers.

**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.
- Overall, the goal is to provide actionable insights that help drive change.
**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
  - 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

**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 is our Value Proposition?**
1. To provide a better experience for customers.
2. To provide remarkable insight for restaurant owners about their customers.

**Our daily tasks**
- We all have one specific application for our daily tasks or tasks we do more often
  - Transportation
  - Music
  - Restaurant
  - Social media

- There is a bunch of applications but none of them cover the whole experience!

1. **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 kind of information?**
- Demographic data (Users have to sign up)
- We track whatever they do in the app.

**What is the project for this course**
- How to aggregate these data?
- How to extract useful information?
- How to visualize these information?

An example could be:
- How do I gather data to be used for your item's list, more than other parts, and how do I visualize this information helping the customer to be able to see all of your items. See four different alternatives for info visualization.
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

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

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

InfoVis Interactivity Survey

Learning
How interactivity helps us understand the nature of a topic

Plan

Examples
Link 1
Link 2

Media Conglomerates: Who owns our media?
What is a media conglomerate?
- A large company that owns multiple smaller companies involved in media enterprises
Why is this interesting?
- We consume media everyday
- Concentration of media ownership has been falling into the hands of fewer and fewer corporations

Potential Ideas for Visualization

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!

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

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

Defining Task Requirements
• What tasks are important for users of provenance visualizations?
• What questions do we have to answer for program comprehension?

PROVENANCE
"Chronology of the ownership, custody or location of a historical object"

Provides: context, verification
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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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.
   Limited to things supported by Desmos (points, lines, areas)
   No 3D (second half of course)

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 . . . ?

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

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?

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?

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

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

---

**WHY**

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

---

**WHAT**

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

---

**HOW**

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

---

**TBD**

- Select 2 or 3 visualization tools
- A unified dataset

---

**REFERENCE**


---

THANK YOU!
DNA* Sequencing Vis
Exploring sequencing structural noise

End aim of informing:
- Data simulation
- Downstream bioinformatics
- End aim of informing:
  - Data simulation
  - Downstream bioinformatics

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!

Data (reads mapped on target)

Data (reads raw)

DNA sequencing ideally
- target
- duplicate target
- generate sequencing "reads"

DNA sequencing ideally reality
- target
- duplicate target
- generate sequencing "reads"

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!

Data (reads mapped on target)