visualizing crime in vancouver

alex kim & amon ge
oct 17 2017

dataset

Type	Year	Month	DAY	Hour	Minute	Hundred Block	Neighbourhood	X	Y
Vehicle Collision or Pe...	2003	2	15	12	43	SE MARINE DR / NAN...	Victoria-Fraserview	495,662.00	5,450,521.00
Theft from Vehicle	2003	9	10	2	0	9XX E CORDOVA ST	Strathcona	493,844.21	5,458,804.80
Break and Enter Resi...	2003	4	29	11	20	5XX CASSIAR ST	Hastings-Sunrise	497,761.47	5,458,596.15
Break and Enter Resi...	2003	5	18	21	16	5XX CASSIAR ST	Hastings-Sunrise	497,761.47	5,458,596.15
Break and Enter Resi...	2003	10	8	7	0	5XX CENTENNIAL RD	Strathcona	493,462.55	5,459,359.85
Offence Against a Per...	2003	8	16	*null*	*null*	OFFSET TO PROTECT ...	*null*	0.00	0.00
Mischief	2003	5	31	19	11	X W 23RD AVE	Riley Park	492,348.61	5,455,239.48
Mischief	2003	9	8	7	0	X W 20TH AVE	Riley Park	492,360.08	5,455,555.67
Theft from Vehicle	2003	8	20	14	0	19XX PENDRELL ST	West End	489,642.41	5,459,595.07
Break and Enter Resi...	2003	2	15	16	0	5XX COMMODORE RD	Fairview	491,555.74	5,457,210.53
Break and Enter Resi...	2003	8	14	9	20	5XX COMMODORE RD	Fairview	491,555.74	5,457,210.53
Break and Enter Resi...	2003	11	9	6	0	5XX COMMODORE RD	Fairview	491,555.74	5,457,210.53
Break and Enter Resi...	2003	11	19	11	30	5XX COMMODORE RD	Fairview	491,555.74	5,457,210.53
Theft from Vehicle	2003	4	6	2	0	9XX E CORDOVA ST	Strathcona	493,844.21	5,458,804.80
Break and Enter Resi...	2003	11	20	23	0	5XX COMMODORE RD	Fairview	491,555.74	5,457,210.53
Break and Enter Com...	2003	9	26	2	30	10XX ALBERNI ST	West End	491,067.65	5,459,114.22
Theft from Vehicle	2003	9	13	23	0	9XX E BROADWAY AVE	Mount Pleasant	493,930.70	5,456,638.79
Theft from Vehicle	2003	8	29	13	0	E 23RD AVE / MAIN ST	Riley Park	492,648.77	5,455,280.57
Break and Enter Resi...	2003	6	11	22	20	14XX E 4TH AVE	Grandview-Woodland	494,551.61	5,457,133.66
Break and Enter Resi...	2003	6	14	16	15	14XX E 4TH AVE	Grandview-Woodland	494,551.61	5,457,133.66
Mischief	2003	5	15	0	0	21XX COLUMBIA ST	Mount Pleasant	492,032.63	5,457,089.39
Break and Enter Resi...	2003	6	16	1	15	14XX E 4TH AVE	Grandview-Woodland	494,551.61	5,457,133.66
Break and Enter Resi...	2003	7	22	16	55	14XX E 4TH AVE	Grandview-Woodland	494,551.61	5,457,133.66

data.vancouver.ca/datacatalogue/crime-data.htm

Open Data Catalogue

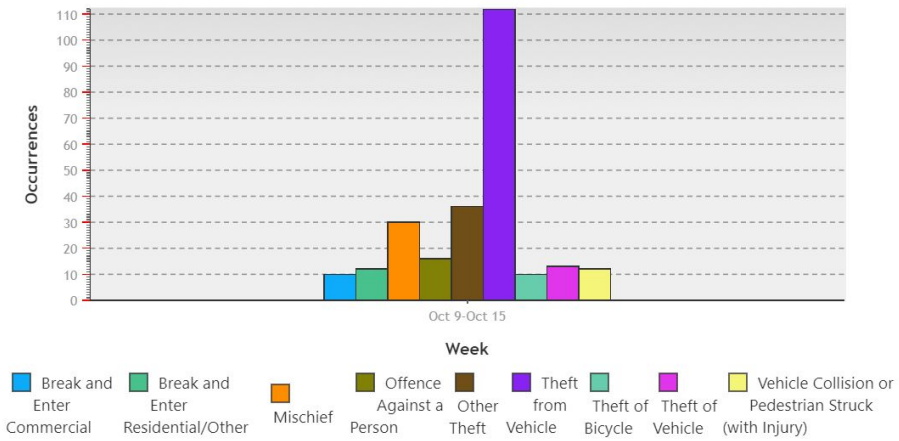
Crime

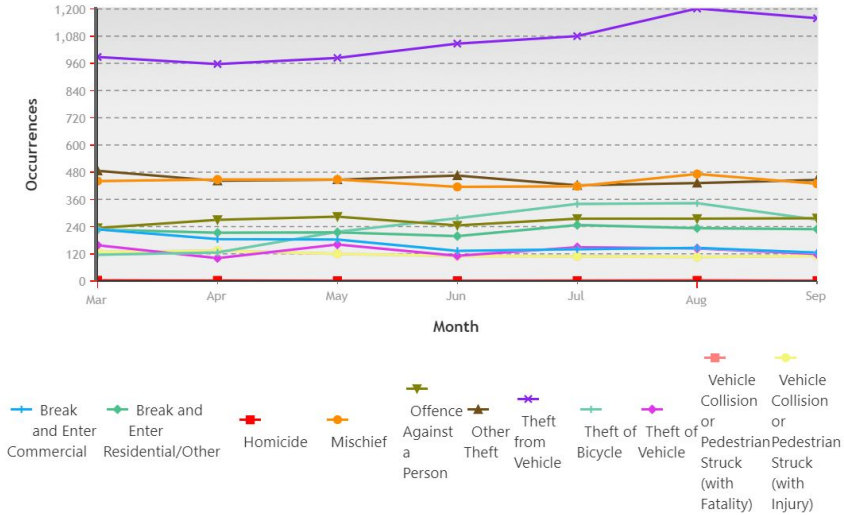
Data custodian	Vancouver Police Department
Data currency comments	The data on this site is scheduled to be updated every Sunday morning. Note: there can be a delay of up to a week between when data is updated in its home system and the publication to the Open Data feed.
Data set description	This is a dataset of crime data on a year-by-year basis beginning in 2003. **Legal Disclaimer from Vancouver Police Department** The release of Vancouver Police Department (VPD) crime data is intended to enhance community awareness of policing activity in Vancouver. Users are cautioned not to rely on the information provided to make decisions about the specific safety level of a specific location or area. By using this data the user agrees and understands that neither the Vancouver Police Department, Vancouver Police Board nor the City of Vancouver assumes liability for any decisions made or actions taken or not taken by the user in reliance upon any information or data provided. While every effort has been made to be transparent in this process, users should be aware that this data is designed to provide individuals with a general overview of incidents falling into several crime categories. The information provided therefore does not reflect the total number of calls or complaints made to the VPD. Please refer to the FAQ for further details. The data provided is based upon information contained in the VPD Records Management System. The crime classification and file status may change at any time based on the dynamic nature of police investigations. The VPD has taken great care to protect the privacy of all parties involved in the incidents reported. No personal or identifying information has been provided in the data. Locations for reported incidents involving Offences Against a Person have been deliberately randomized to several blocks and offset to an intersection. No time or street location name will be provided for these offences. For property related offences, the VPD has provided the location to the hundred block of these incidents within the general area of the block. All data must be considered offset and users should not interpret any locations as related to a specific person or specific property.
Data accuracy comments	The Vancouver Police Department's GeoDASH Crime Map remains the authoritative source. Note: GeoDASH stands for Geographic Data Analysis and Statistics Hub. It is a crime mapping tool used by Vancouver Police Department (VPD) to inform residents on the crime activities happening in Vancouver.
Attributes	▪ TYPE ▪ YEAR ▪ MONTH ▪ DAY ▪ HOUR ▪ MINUTE ▪ HUNDRED_BLOCK ▪ NEIGHBOURHOOD ▪ X ▪ Y

current visualization

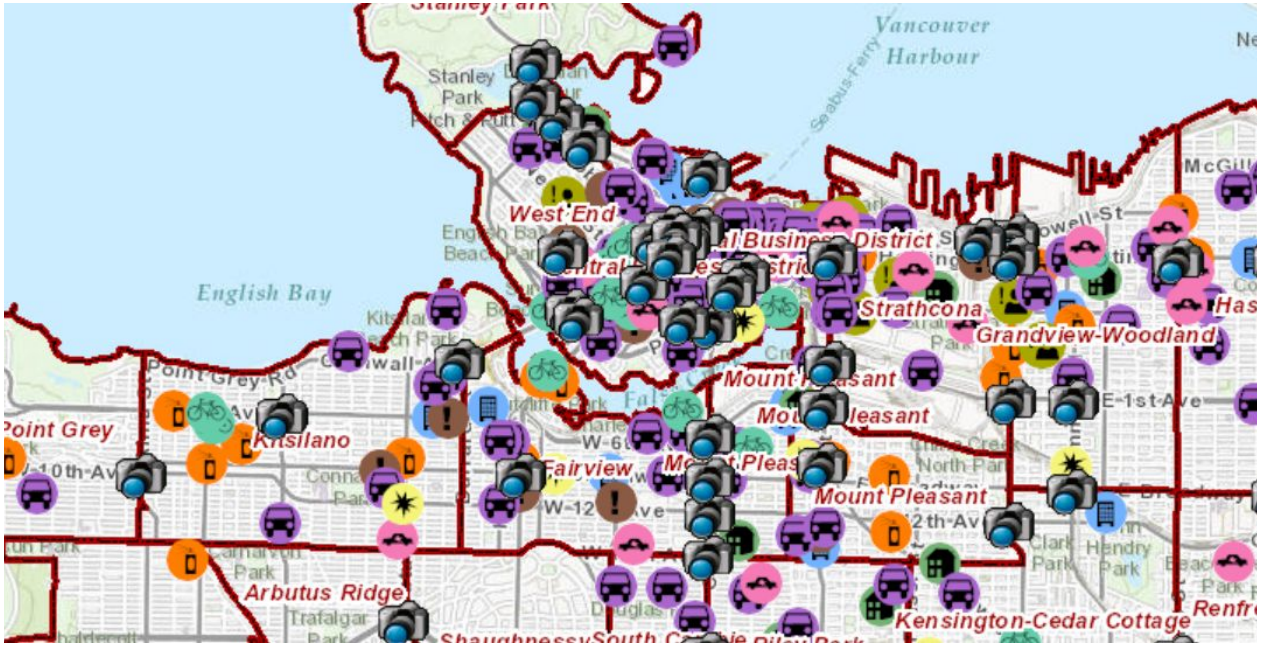
drawbacks:

- impossible to see the past trends, beyond 2 years in the past
- doesn't allow choosing a period of time of interest
- can't view hourly/daily trends
- can't look at other context (neighbourhoods)
- doesn't look visually appealing





current visualization

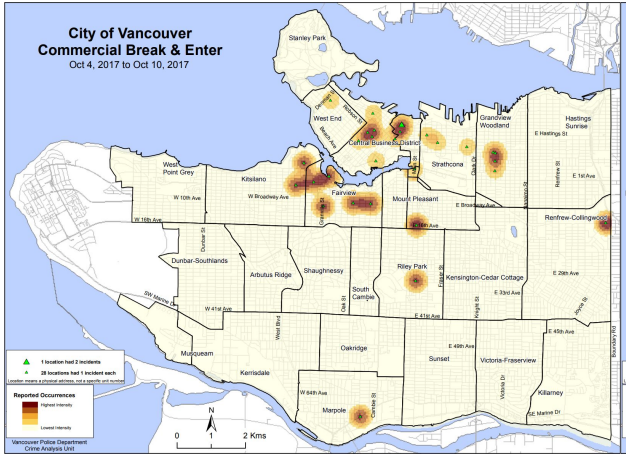


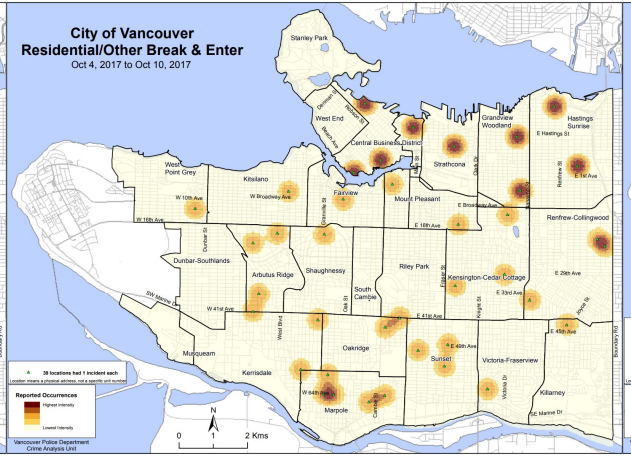
drawbacks:

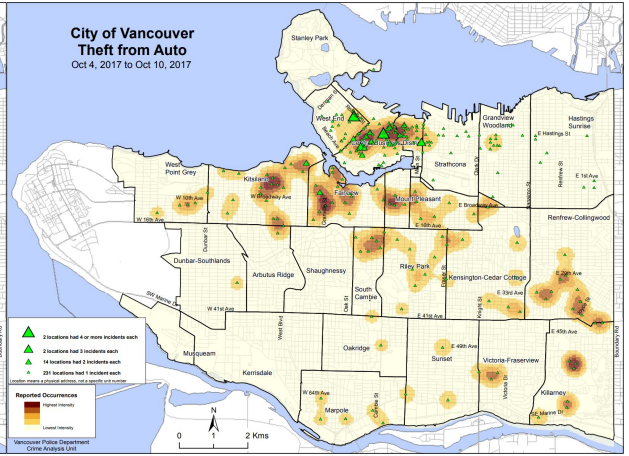
- cluttered when zoomed out
- shows all crimes at the same time
- only displays data for the past week

geodash.vpd.ca

current visualization



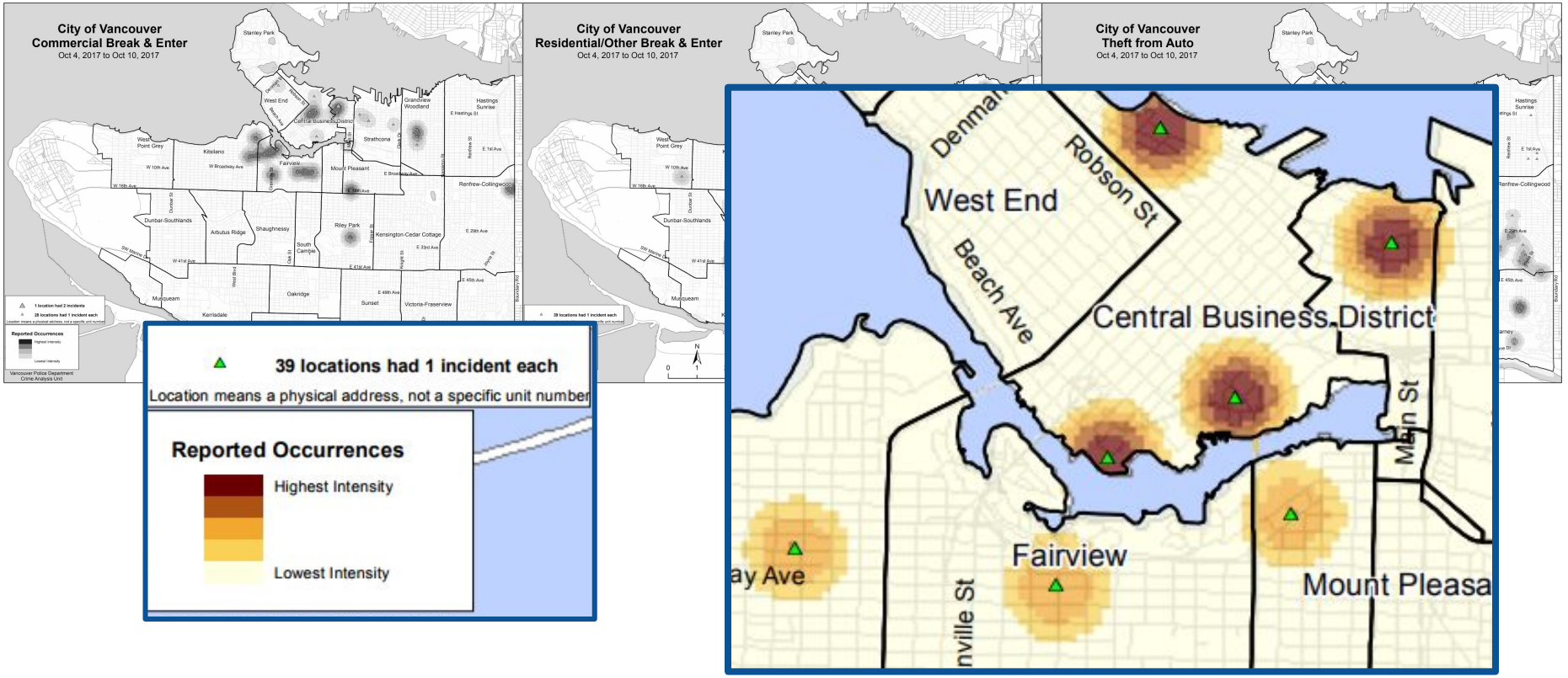




vancouver.ca/police/crimemaps

only current week available, exists only in pdf(!) format

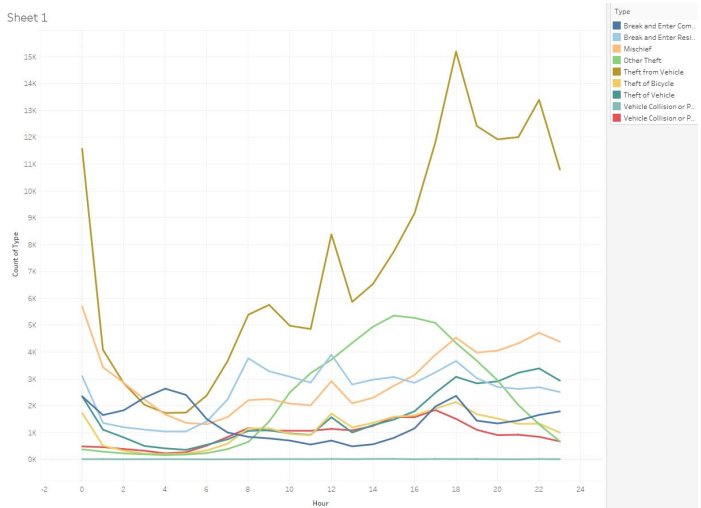
current visualization

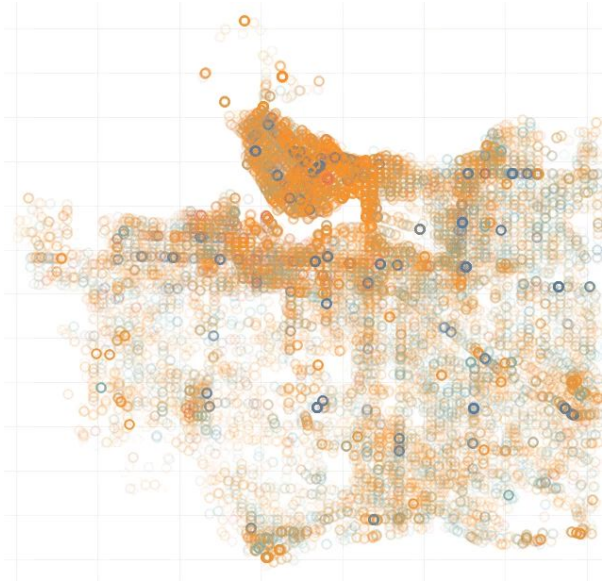


proposal

tackle the mentioned drawbacks:

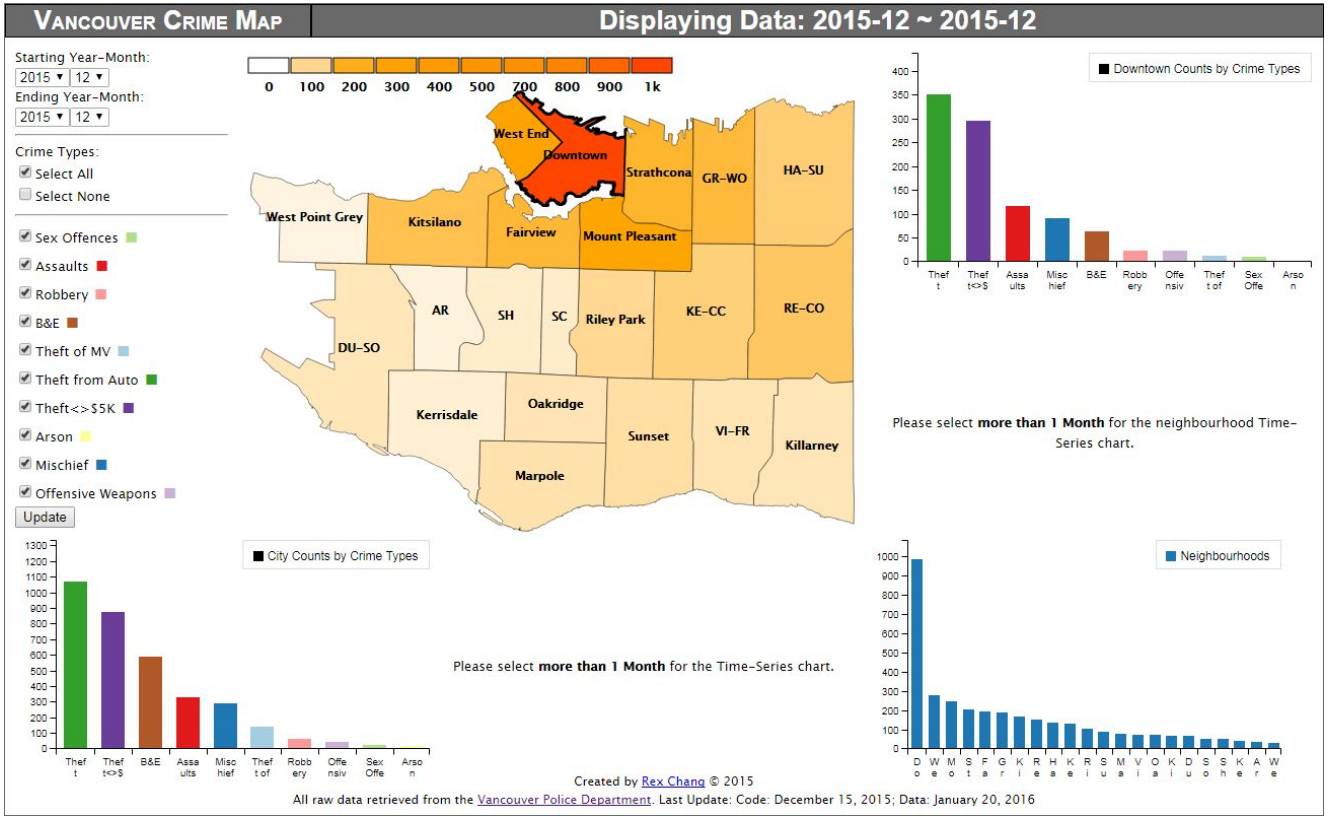
- interactivity: selecting crime type, time range, region, etc.
- animate trends over time
- cleaner





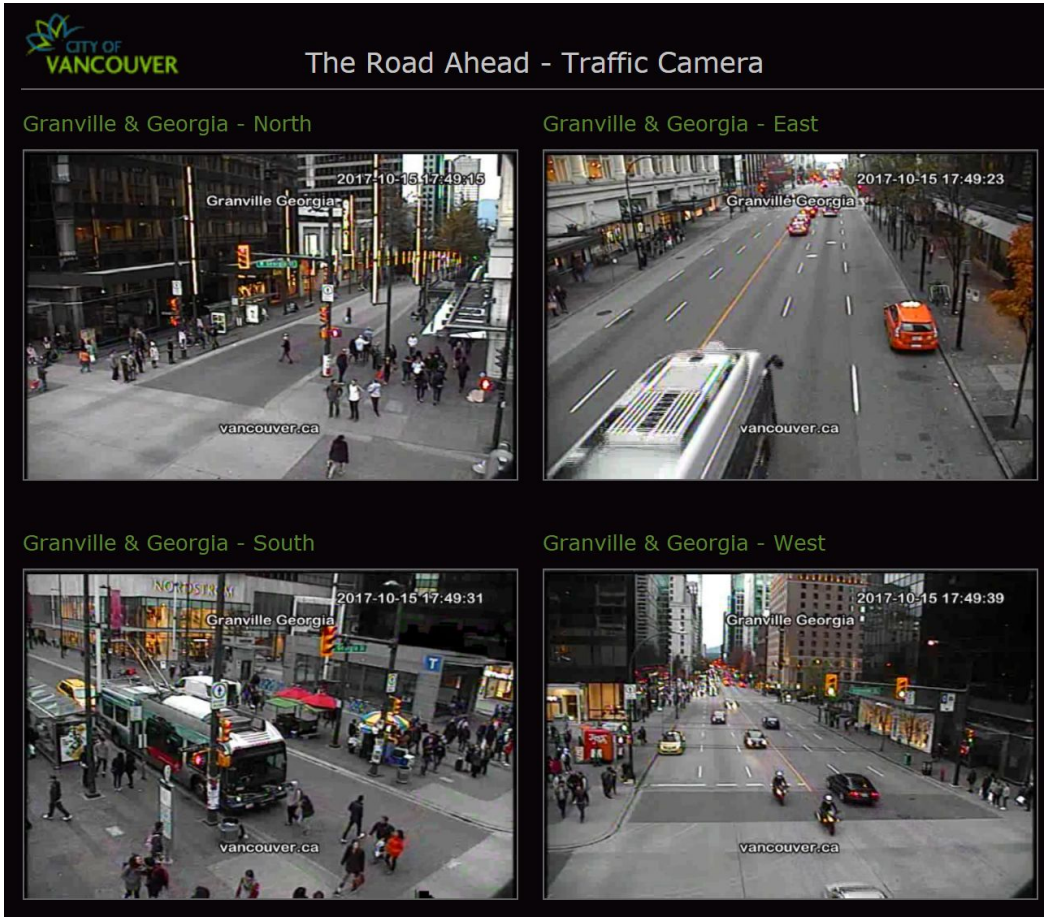
2015 project

rexchang.com/vancouver-crimemap



tangent: traffic cams

update every 2~15 min



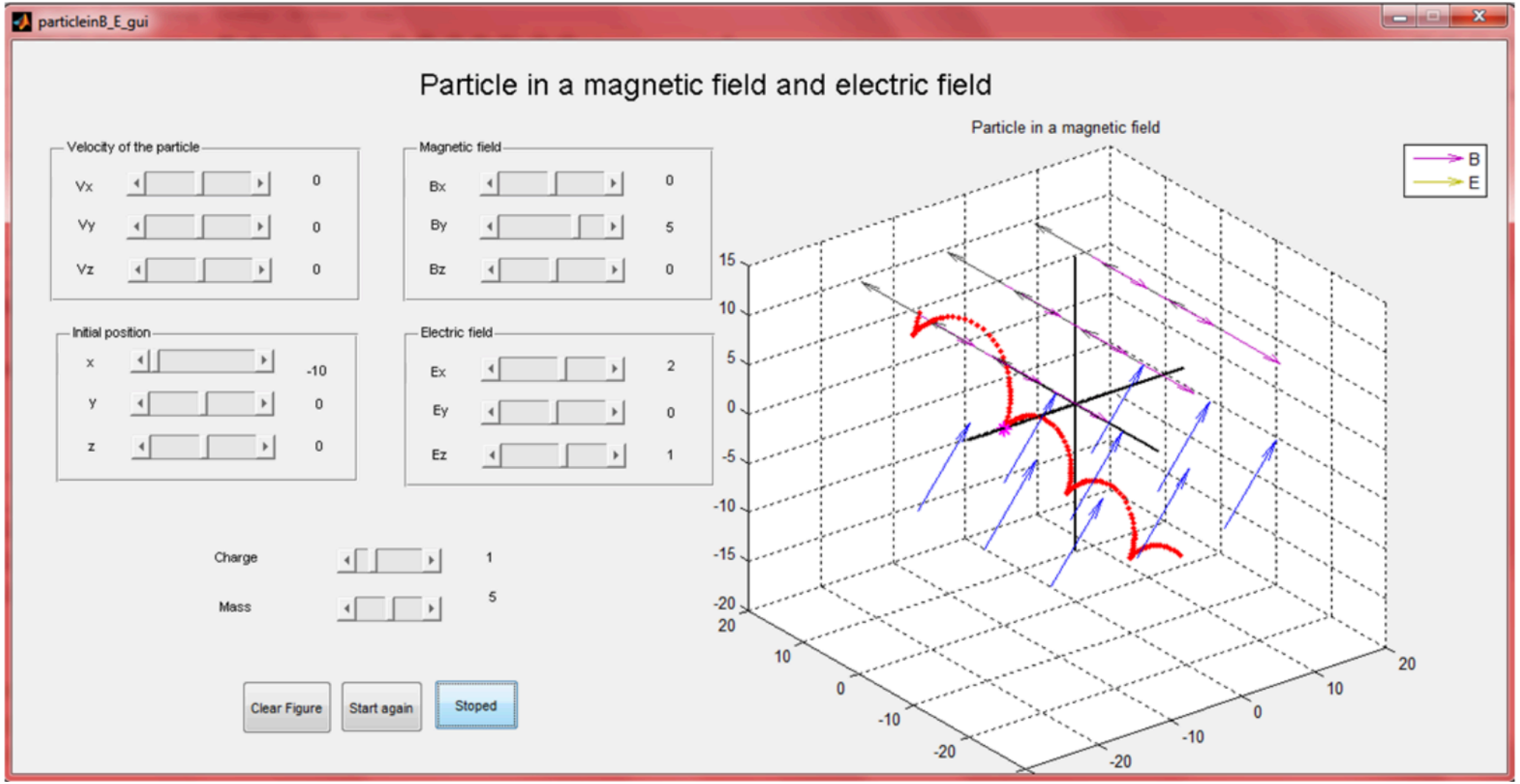
Visualizing algorithms

Gursimran

Big picture idea

- Pedagogical focus
  - Convergence of optimization functions
    - Simple – Netwon raphson method
    - How does the PSO converge?
  - Movement of particle in some electric and magnetic field
    - How do we represent electric and magnetic field
    - How do we show the particle moving
    - How do we show all forces on the particle at any time?
    - What happens when we have multiple particles.

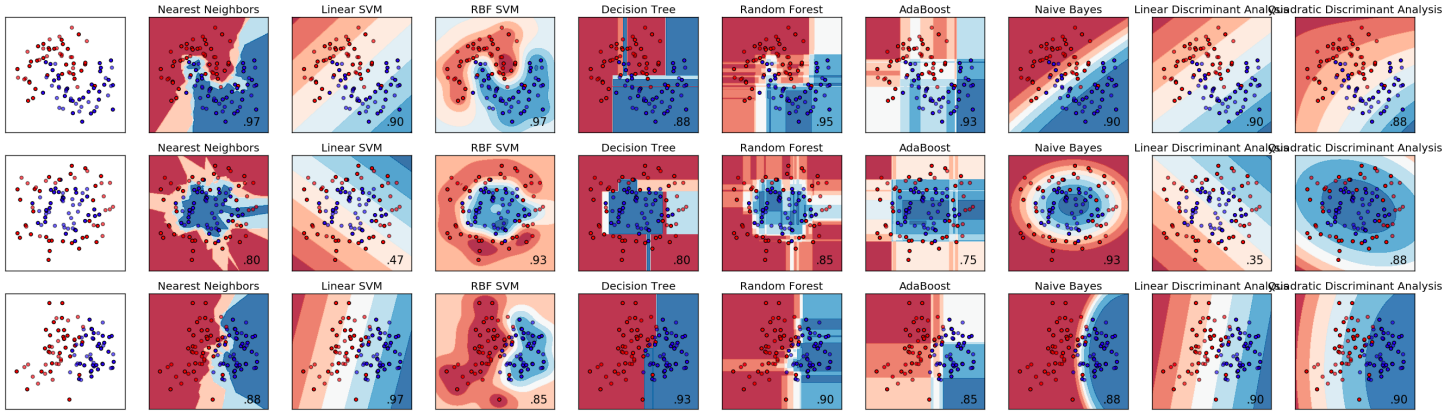
Example – particle in E and B



Some examples on ML

- http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
- http://playground.tensorflow.org/

- Or what if just use 2D figures; when people click then can interact with these as well



Why visualize algorithms

- Very rich from IV perspective
  - We will have to work in very high dimentions
    - Really have to make sure we use our channels appropriately
    - How to represent complex fields/ data – say elec and mag field together?
  - Will have to care about principal of expressiveness
    - As we are making it for pedagogical purposes
  - When do we use 3D? When to use interactivity?

- Impact
  - Useful and publishable material
  - Pedagogical significance so someone will use it at the end
  - We learn about cool algorithms

- Tools
  - D3 – explanatory analysis
  - May be we can try some python tools as well

Thanks – any questions

- Call for project partners who have background in
  - Computer algorithms (or ML algorithms)
  - Coding (cos we will do stuff in d3)

- **Motivation taken from**
  - https://distill.pub/about/
  - Distill Prize for Clarity in Machine Learning
  - http://rawgraphs.io/

Another idea

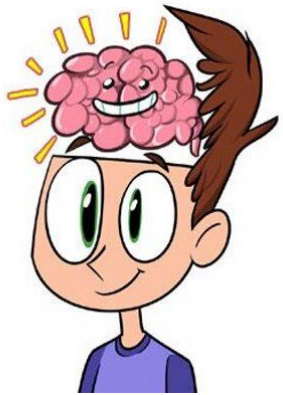
- ML based viz system which suggest viz based on data attributes
- 2D representations of algorithms which can explain how it works
  - Or possibly simple gifs and a framework to make these gifs

- People
  - http://cs.stanford.edu/people/karpathy/

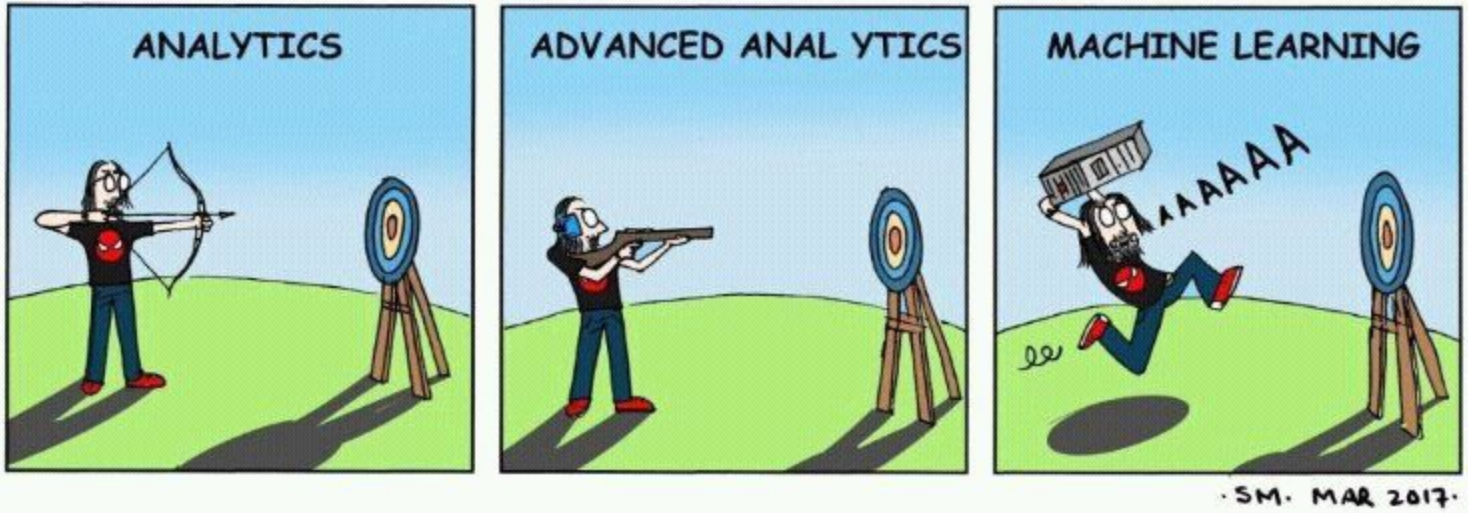
- https://www.quora.com/What-are-the-best-visualizations-of-machine-learning-algorithms

Intuitive explanations

Halldor Thorhallsson



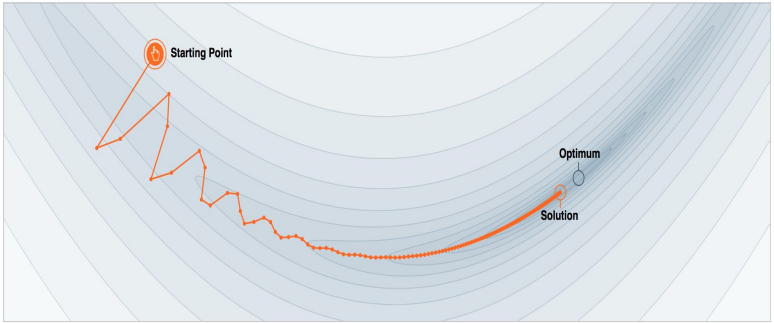

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS


ANALYTICS
ADVANCED ANAL YTICS
MACHINE LEARNING
AAAAAA
·SM· MAR 2017·


Machine Learning
Linear Algebra
Statistics

Distill.pub

Why Momentum Really Works



Step-size α = 0.02

0 0.003 0.006

Momentum β = 0.99

0.00 0.500 0.990

We often think of Momentum as a means of dampening oscillations and speeding up the iterations, leading to faster convergence. But it has other interesting behavior. It allows a larger range of step-sizes to be used, and creates its own oscillations. What is going on?

Step
1,960

Random (Gaussian)
Jump

Share this view

Number Of Points 100

Dimension 100

Perplexity 10

Epsilon 5

MARTIN WATTENBERG
Google Brain

FERNANDA VIÉGAS
Google Brain

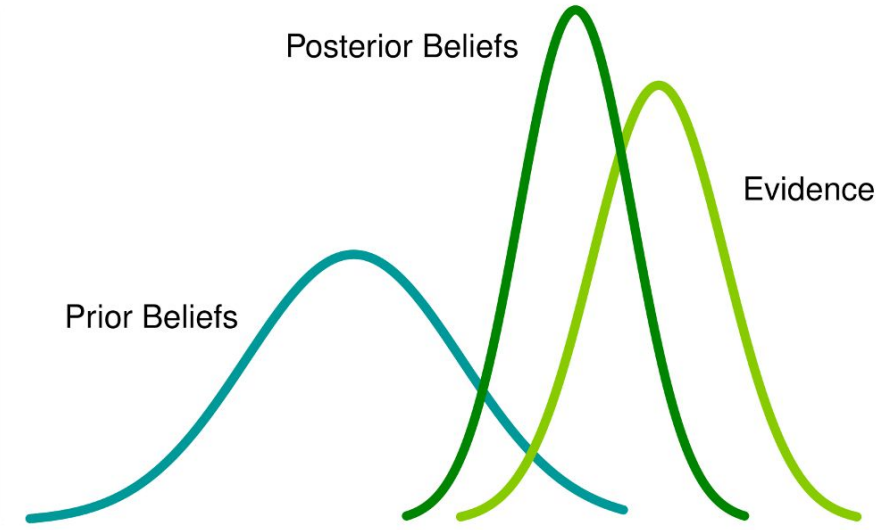
IAN JOHNSON
Google Cloud

Oct. 13
2016

Citation:
Wattenberg, et al., 2016

Sample topics

- Covariance matrix
- CLT
- Bayes rule
- PCA



Storytelling

“Maybe stories are just data with a soul.” - *Brené Brown*

```
Call:
lm(formula = iris)

Residuals:
     Min       1Q   Median       3Q      Max
-0.79424 -0.21874  0.00899  0.20255  0.73103

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.17127    0.27979   7.760 1.43e-12 ***
Sepal.Width        0.49589    0.08607   5.761 4.87e-08 ***
Petal.Length       0.82924    0.06853  12.101  < 2e-16 ***
Petal.Width       -0.31516    0.15120  -2.084  0.03889 *
Speciesversicolor -0.72356    0.24017  -3.013  0.00306 **
Speciesvirginica  -1.02350    0.33373  -3.067  0.00258 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3068 on 144 degrees of freedom
Multiple R-squared:  0.8673,    Adjusted R-squared:  0.8627
F-statistic: 188.3 on 5 and 144 DF,  p-value: < 2.2e-16
```

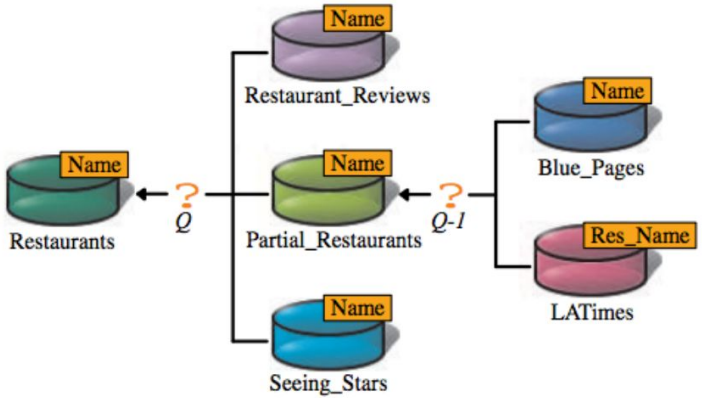
CPSC547 Pitch

What is Data Integration

- Data Integration is the process of combining data from different data sources.
  - Example:
  - Dataset 1 contains all human genes available since 1975,
  - Dataset 2 contains all primate genes discovered using the Next Generation Sequencing method.
  - We want to integrate them to create a more complete dataset for the human genome.
- What problems does it have? Data might be stored in different formats.
  - Example:
  - Dataset 1 stores date in the format of 2017/10/16, and
  - Dataset 2 stores in the format of October 16, 2017.
- What solutions are out there? Apply transformations to each dataset to convert values in each dataset to a conventional form, and then integrate.
  - Example: convert both 2017/10/16 and October 16, 2017 to 20171016

Visualization

- Task: visualize the process of integration between 2 or more datasets
- Dataset: multiple datasets taken from the Bioinformatics domain.
  - Example: Reactome, Ensembl, Chembl, BioModels
  - **All these datasets are already stored in a common format: RDF**
  - Data are tabular, well-curated, and cleaned
- Idiom: encode a number of attributes as node-link diagrams
  - Example:

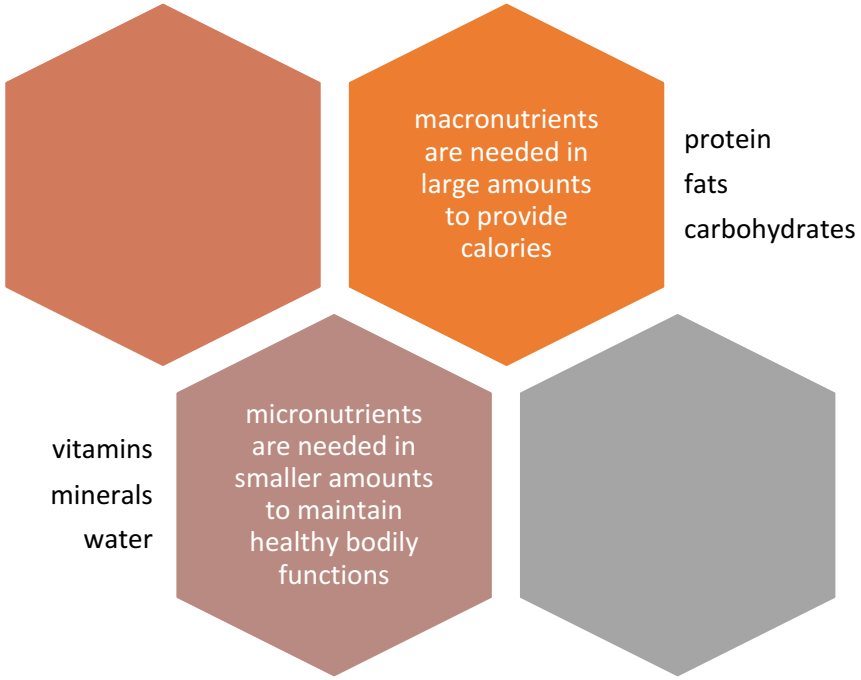


What you will learn

- Data Integration research domain
- Bioinformatics: learn what data do systems biologists use in their research.
- A variant of SQL: SPARQL. This is the language used to generate integrated data from multiple data sources

Meal Planning by Macronutrients

Hayley Guillou



what are macronutrients?

how are macronutrients measured?

macronutrients have a consistent amount of calories per gram

- 1 gram protein = 4 calories
- 1 gram carbohydrate = 4 calories
- 1 gram fat = 9 calories

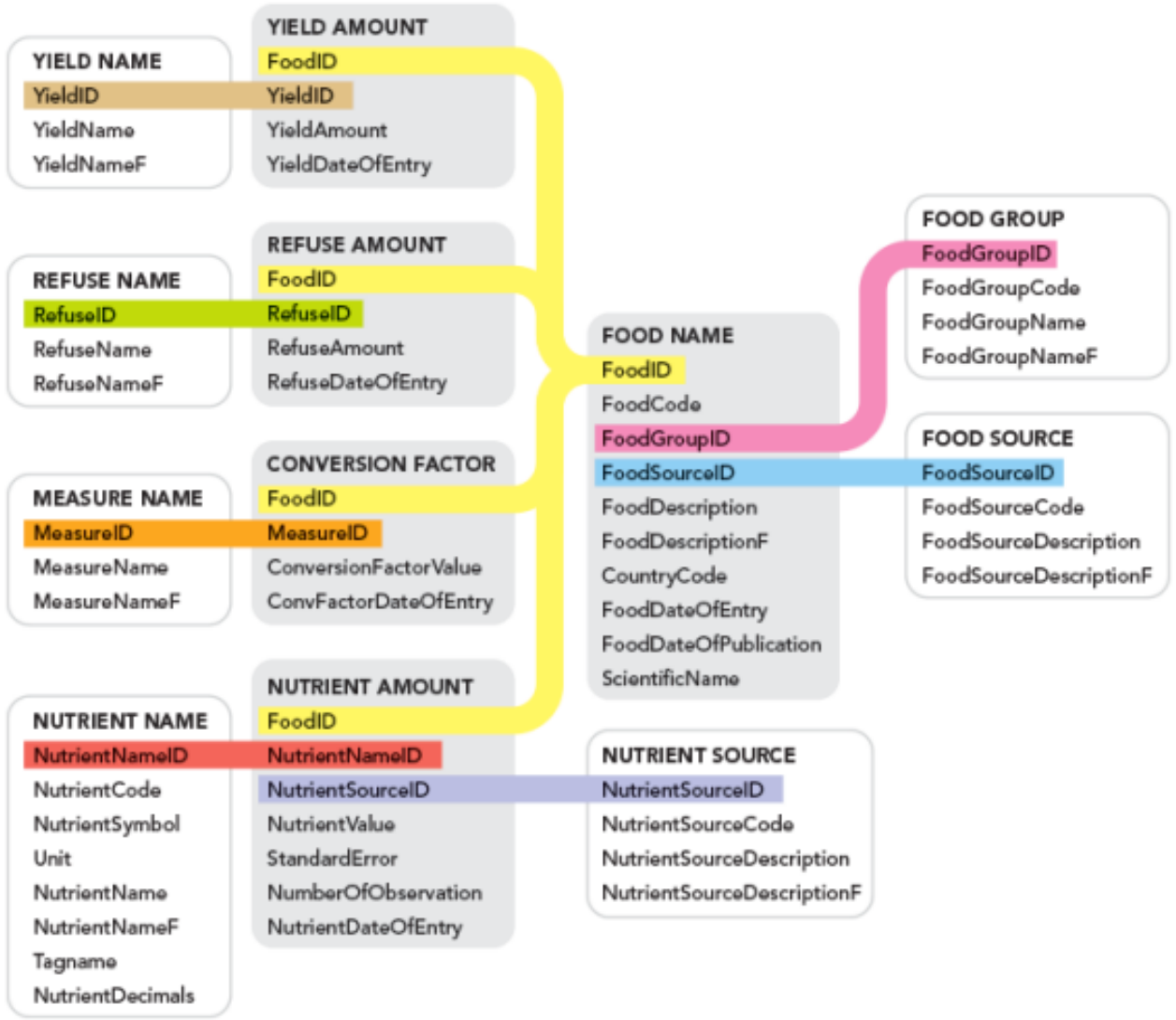
calculate calorie intake based on total daily energy expenditure

calculate the grams of each macronutrient based on ratios of calories

- ex. ketogenic diet (5% carb, 20% protein, 70% fat)

Canadian Nutrient File (CNF)

- over 5600 foods
- over 150 nutrients
- nutrient values per 100 g of food



possible research questions

what kind of visualization would be best suited for daily meal planning based on macronutrients?

what filtering, sorting, and visual features can be added to speed up meal planning?

what trends in personal nutrition can be mapped over time?

Visualizing Eye-tracking data from reading tasks

Jan Pilzer

Motivation and Data Source

Course Project for 539 (with Xinhong Liu):
Detection of future self-distractions during reading using gaze patterns

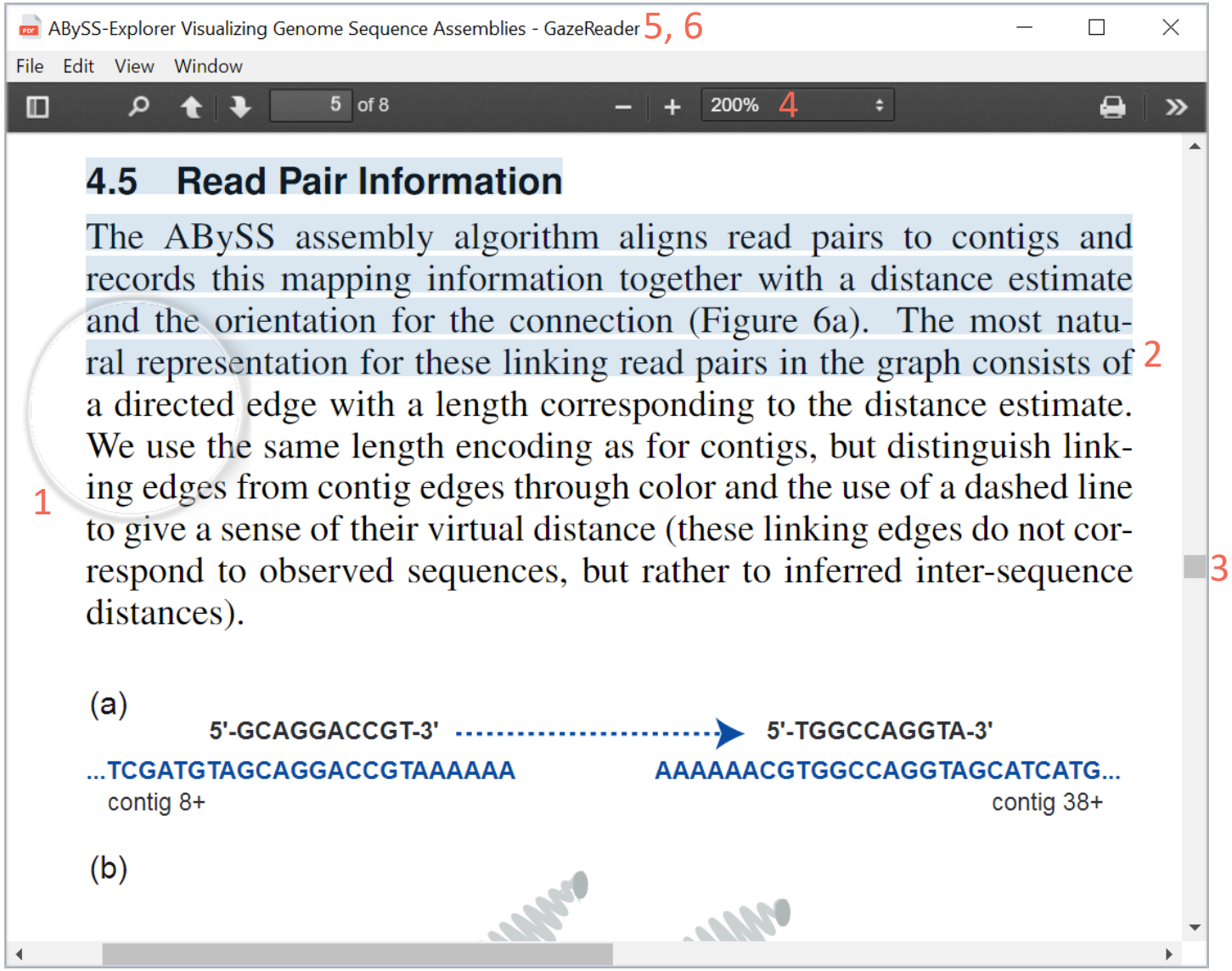
Custom built application that collects information about the document, active windows, and eye tracking data during reading of PDF documents.

Application exists in beta, and is actively being developed. Changes possible.

Data collection

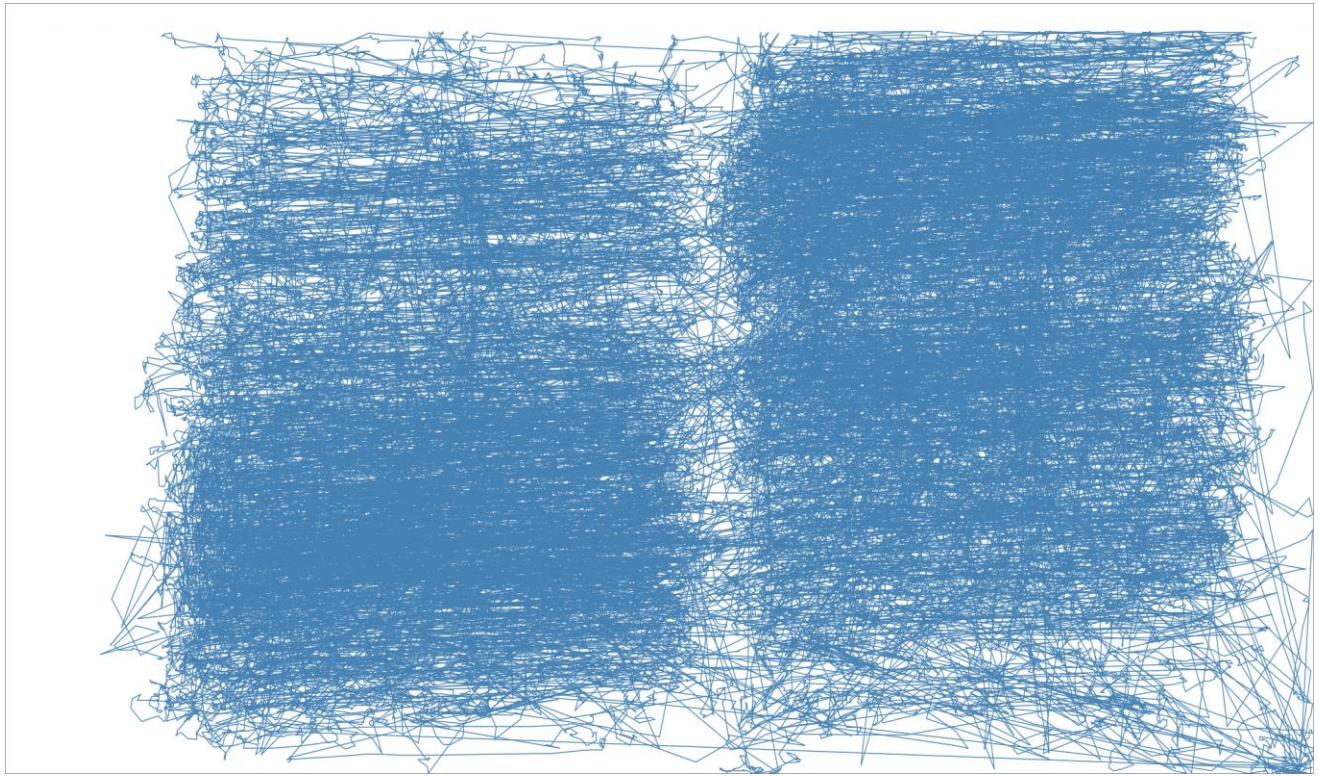
1. Gaze location (in pixel)
2. Target sentence
3. Scroll level
4. Zoom level
5. App focus / blur
6. Active window

Further collection or refinement possible if necessary.

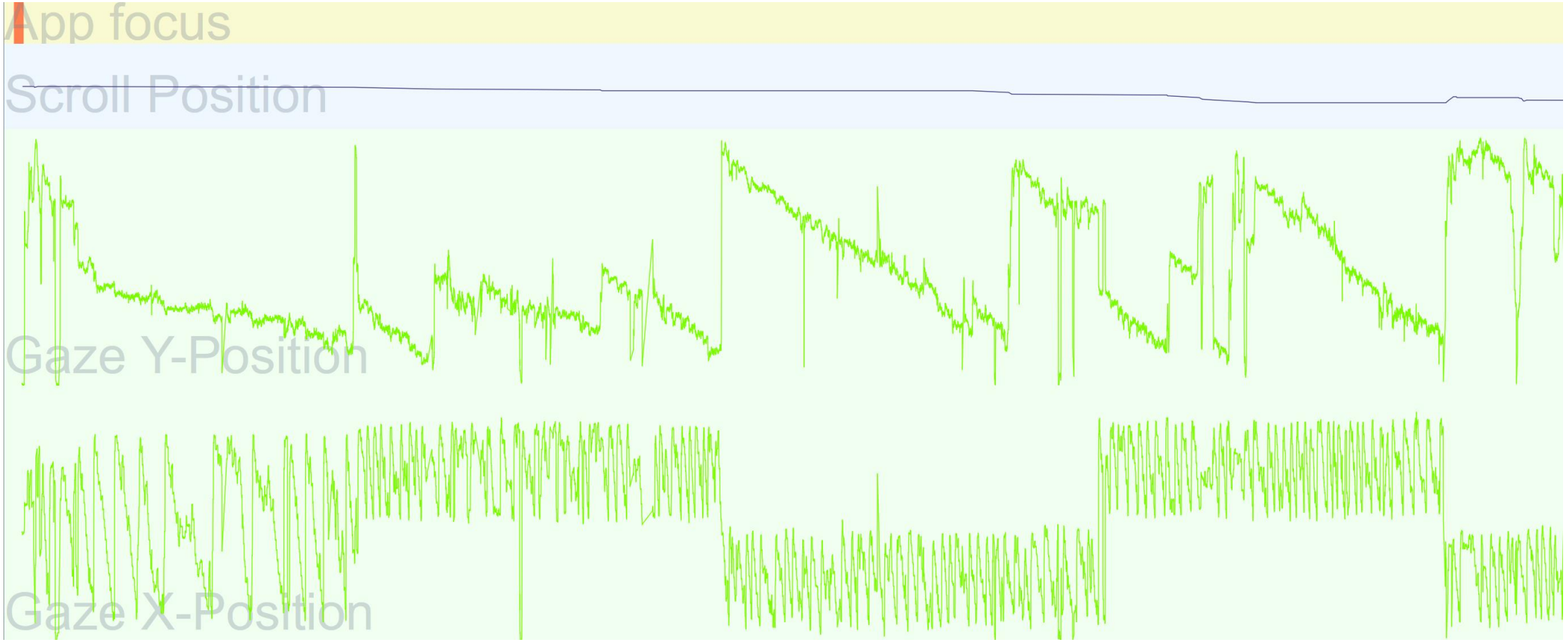


Data Sample

```
2017-10-15T21:35:31.327Z | FOCUS
2017-10-15T21:35:32.183Z | ACTIVE | electron.exe | PDF.js viewer
2017-10-15T21:35:34.894Z | BLUR
2017-10-15T21:35:35.186Z | ACTIVE | explorer.exe | 547
2017-10-15T21:35:37.080Z | OPEN | C:/Users/Jan/Documents/Documents/547/4.3 - ABySS-Exp
2017-10-15T21:35:38.429Z | FOCUS
2017-10-15T21:35:38.929Z | GAZE | (738.32|384.69) | (30.34%|5.91%) | —One bottleneck i
2017-10-15T21:35:38.940Z | GAZE | (738.47|387.90) | (30.36%|31.30%) | —One bottleneck
2017-10-15T21:35:38.950Z | GAZE | (739.56|395.37) | (30.51%|90.26%) | —One bottleneck
2017-10-15T21:35:38.962Z | GAZE | (740.71|399.90) | (35.48%|1.49%) | sequences produce
2017-10-15T21:35:38.974Z | GAZE | (739.89|400.73) | (35.37%|8.08%) | sequences produce
2017-10-15T21:35:38.984Z | GAZE | (740.16|399.03) | (35.41%|-5.39%) | sequences produc
2017-10-15T21:35:38.995Z | GAZE | (740.72|400.02) | (35.48%|2.47%) | sequences produce
2017-10-15T21:35:39.007Z | GAZE | (740.71|399.54) | (35.48%|-1.33%) | sequences produc
2017-10-15T21:35:39.018Z | GAZE | (741.34|400.31) | (35.56%|4.72%) | sequences produce
2017-10-15T21:35:39.029Z | GAZE | (740.84|399.66) | (35.50%|-0.40%) | sequences produc
2017-10-15T21:35:39.051Z | GAZE | (732.52|396.97) | (29.53%|102.86%) | —One bottleneck
2017-10-15T21:35:39.062Z | GAZE | (730.71|396.86) | (29.28%|102.01%) | —One bottleneck
2017-10-15T21:35:39.084Z | GAZE | (731.10|400.39) | (34.24%|5.37%) | sequences produce
2017-10-15T21:35:39.095Z | GAZE | (731.15|401.79) | (34.25%|16.43%) | sequences produc
2017-10-15T21:35:39.106Z | GAZE | (730.55|400.99) | (34.17%|10.15%) | sequences produc
2017-10-15T21:35:39.117Z | GAZE | (730.73|400.61) | (34.19%|7.10%) | sequences produce
2017-10-15T21:35:39.129Z | GAZE | (732.49|400.52) | (34.42%|6.39%) | sequences produce
2017-10-15T21:35:39.140Z | GAZE | (732.21|402.08) | (34.38%|18.73%) | sequences produc
2017-10-15T21:35:39.151Z | GAZE | (732.44|402.99) | (34.41%|25.95%) | sequences produc
2017-10-15T21:35:39.162Z | GAZE | (732.61|402.34) | (34.44%|20.77%) | sequences produc
2017-10-15T21:35:39.174Z | GAZE | (733.91|402.12) | (34.60%|19.01%) | sequences produc
2017-10-15T21:35:39.184Z | GAZE | (734.64|402.28) | (34.70%|20.28%) | sequences produc
2017-10-15T21:35:39.197Z | GAZE | (737.32|403.80) | (35.04%|32.29%) | sequences produc
2017-10-15T21:35:39.207Z | GAZE | (737.72|403.98) | (35.09%|33.69%) | sequences produc
2017-10-15T21:35:39.217Z | GAZE | (737.46|403.99) | (35.06%|33.82%) | sequences produc
2017-10-15T21:35:39.229Z | GAZE | (736.93|403.41) | (34.99%|29.24%) | sequences produc
2017-10-15T21:35:39.240Z | GAZE | (736.61|404.13) | (34.95%|34.92%) | sequences produc
2017-10-15T21:35:39.251Z | GAZE | (736.77|403.75) | (34.97%|31.87%) | sequences produc
2017-10-15T21:35:39.262Z | GAZE | (736.82|403.74) | (34.98%|31.80%) | sequences produc
2017-10-15T21:35:39.273Z | GAZE | (736.70|403.45) | (34.96%|29.56%) | sequences produc
2017-10-15T21:35:39.285Z | GAZE | (735.35|402.44) | (34.79%|21.60%) | sequences produc
2017-10-15T21:35:39.296Z | GAZE | (734.92|401.95) | (34.73%|17.70%) | sequences produc
2017-10-15T21:35:39.308Z | GAZE | (733.76|401.46) | (34.58%|13.87%) | sequences produc
2017-10-15T21:35:39.317Z | GAZE | (732.45|400.29) | (34.42%|4.60%) | sequences produce
```



Initial Analysis



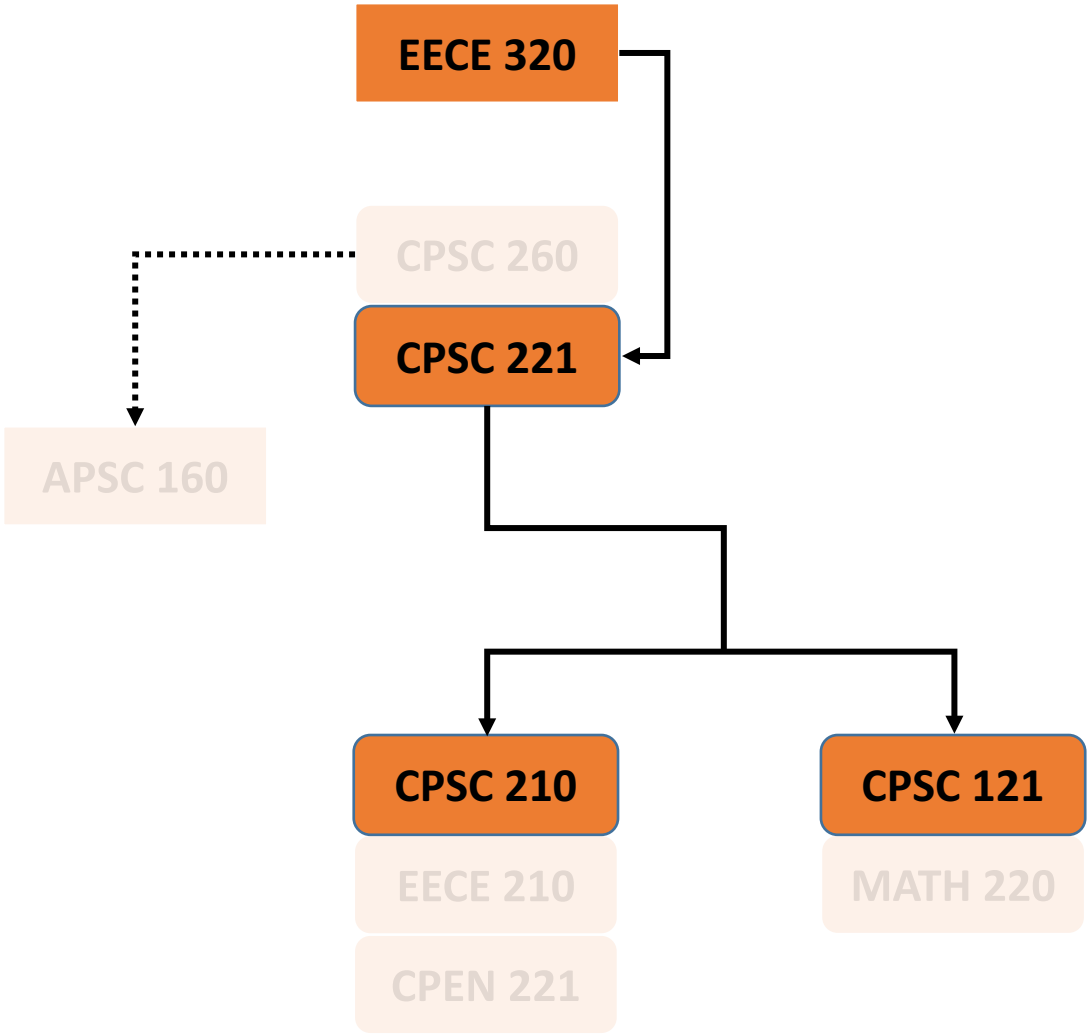
https://cs.ubc.ca/~pilzer/projects/547

Visualization of UBC Courses

Jiahong Chen (Department of Mechanical Engineering)

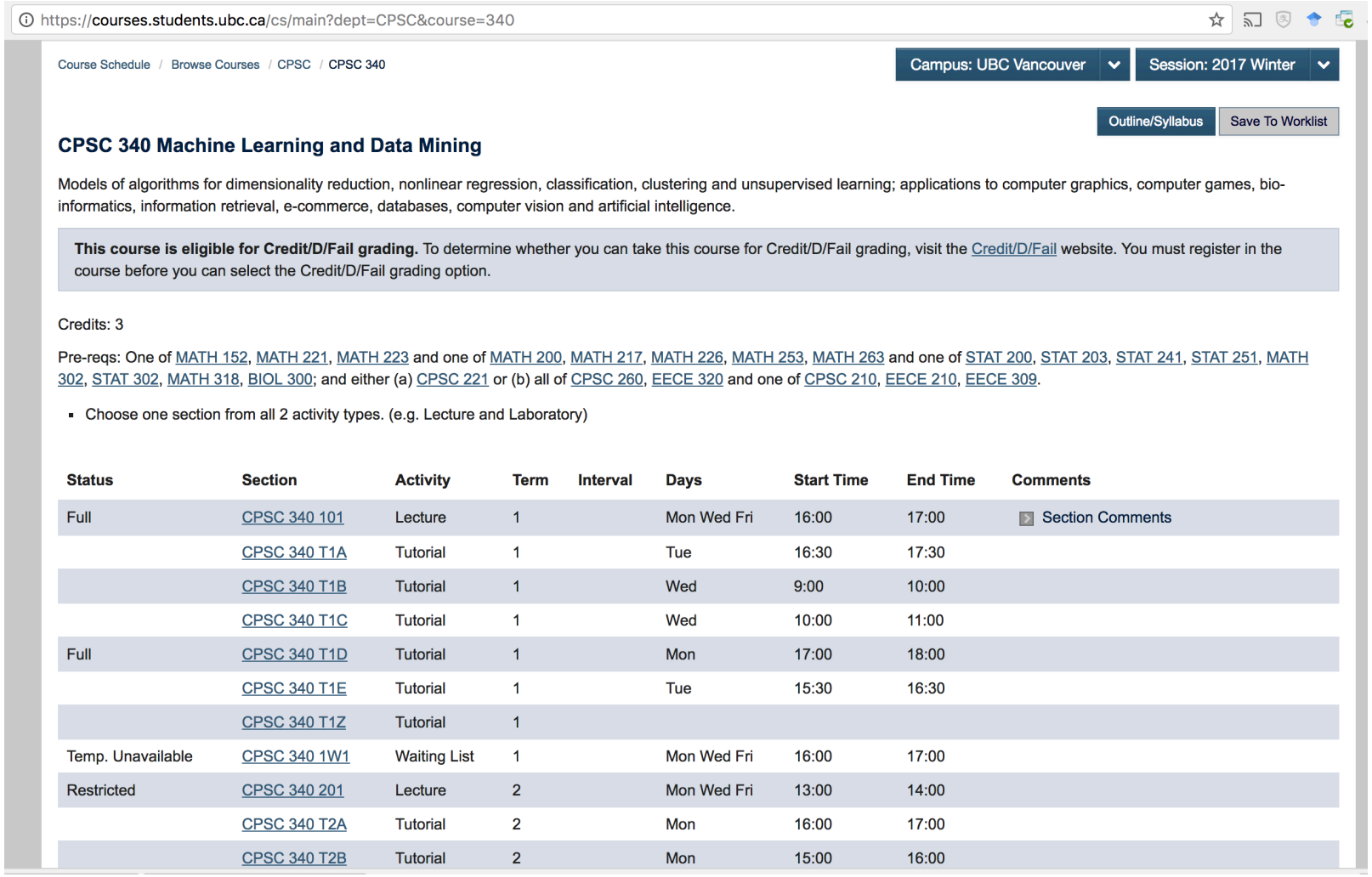
Siyuan He (Department of Computer Science)

Dozens of Pre-reqs



- Many pre-reqs (especially in undergrad course)
- Pre-reqs of pre-reqs
- All of / one of relationship
- Overlap-pre-reqs

Dozens of Pre-reqs



https://courses.students.ubc.ca/cs/main?dept=CPSC&course=340

Course Schedule / Browse Courses / CPSC / CPSC 340

Campus: UBC Vancouver | Session: 2017 Winter

Outline/Syllabus | Save To Worklist

CPSC 340 Machine Learning and Data Mining

Models of algorithms for dimensionality reduction, nonlinear regression, classification, clustering and unsupervised learning; applications to computer graphics, computer games, bio-informatics, information retrieval, e-commerce, databases, computer vision and artificial intelligence.

This course is eligible for Credit/D/Fail grading. To determine whether you can take this course for Credit/D/Fail grading, visit the Credit/D/Fail website. You must register in the course before you can select the Credit/D/Fail grading option.

Credits: 3

Pre-reqs: One of MATH 152, MATH 221, MATH 223 and one of MATH 200, MATH 217, MATH 226, MATH 253, MATH 263 and one of STAT 200, STAT 203, STAT 241, STAT 251, MATH 302, STAT 302, MATH 318, BIOL 300; and either (a) CPSC 221 or (b) all of CPSC 260, EECE 320 and one of CPSC 210, EECE 210, EECE 309.

- Choose one section from all 2 activity types. (e.g. Lecture and Laboratory)

Status	Section	Activity	Term	Interval	Days	Start Time	End Time	Comments
Full	CPSC 340 101	Lecture	1		Mon Wed Fri	16:00	17:00	Section Comments
	CPSC 340 T1A	Tutorial	1		Tue	16:30	17:30	
	CPSC 340 T1B	Tutorial	1		Wed	9:00	10:00	
	CPSC 340 T1C	Tutorial	1		Wed	10:00	11:00	
Full	CPSC 340 T1D	Tutorial	1		Mon	17:00	18:00	
	CPSC 340 T1E	Tutorial	1		Tue	15:30	16:30	
	CPSC 340 T1Z	Tutorial	1					
Temp. Unavailable	CPSC 340 1W1	Waiting List	1		Mon Wed Fri	16:00	17:00	
Restricted	CPSC 340 201	Lecture	2		Mon Wed Fri	13:00	14:00	
	CPSC 340 T2A	Tutorial	2		Mon	16:00	17:00	
	CPSC 340 T2B	Tutorial	2		Mon	15:00	16:00	

Where to get data?

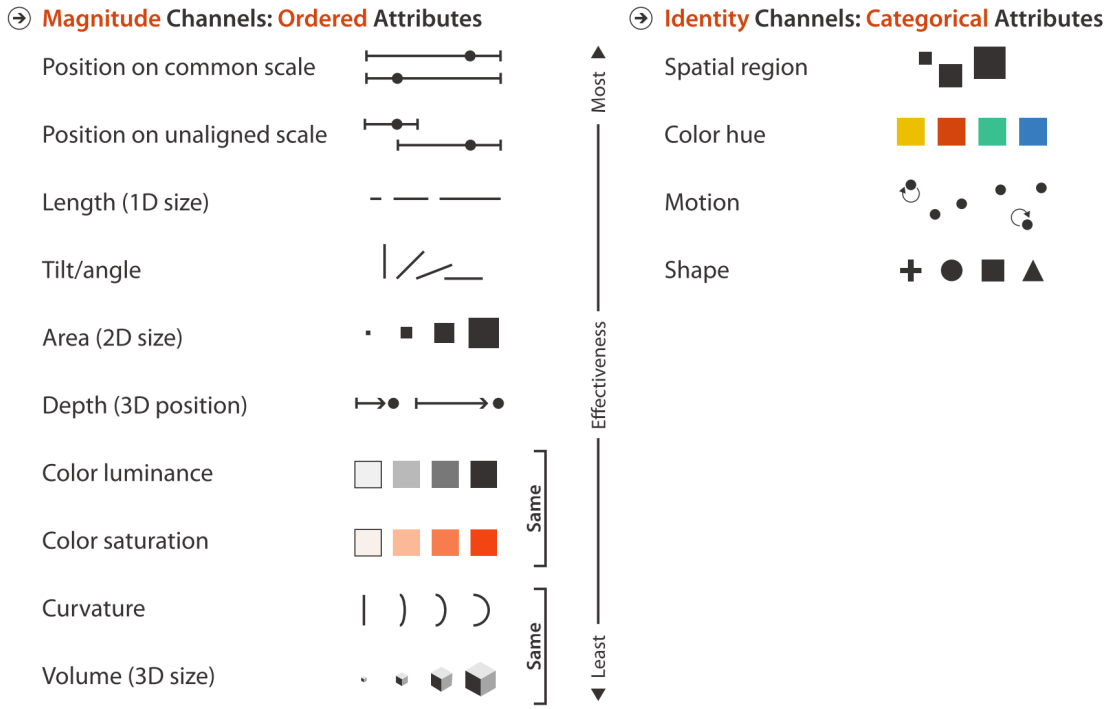
Web Crawling!

view-source:https://courses.students.ubc.ca/cs/main?dept=CPSC&course=340

```
</p>
<p>Pre-reqs:     One of <a href = "/cs/main?
pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=152">MATH 152</a>, <a href = "/cs/main?
pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=221">MATH 221</a>, <a href = "/cs/main?
pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=223">MATH 223</a> and one of <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=200">MATH 200</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=217">MATH 217</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=226">MATH 226</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=253">MATH 253</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=263">MATH 263</a> and one of <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=STAT&amp;course=200">STAT 200</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=STAT&amp;course=203">STAT 203</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=STAT&amp;course=241">STAT 241</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=STAT&amp;course=251">STAT 251</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=302">MATH 302</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=STAT&amp;course=302">STAT 302</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=MATH&amp;course=318">MATH 318</a>, <a
href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=BIOL&amp;course=300">BIOL 300</a>;  and
either (a) <a href = "/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=CPSC&amp;course=221">CPSC
221</a> or (b) all of <a href = "/cs/main?
pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=CPSC&amp;course=260">CPSC 260</a>, <a href = "/cs/main?
pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=EECE&amp;course=320">EECE 320</a> and one of <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=CPSC&amp;course=210">CPSC 210</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=EECE&amp;course=210">EECE 210</a>, <a href =
"/cs/main?pname=subjarea&amp;tname=subjareas&amp;req=3&amp;dept=EECE&amp;course=309">EECE 309</a>. </p>
```

HTML source page of the course page

Vis Techs



Channel

- Size: credits
- Saturation: level of course
- Color: different faculty

Marks

- Points: courses
- Lines: links between courses

Why is useful?

- Curriculum Overview
- Determine which path you want to go
- Determine if you have a breadth of knowledge
- Some other interesting questions such as
  - Determine fundamental courses that applies to all disciplines.
  - Determine which course combines most of the knowledge
  - Clustering all courses.

Thank you

Survey: Data mining and information visualization

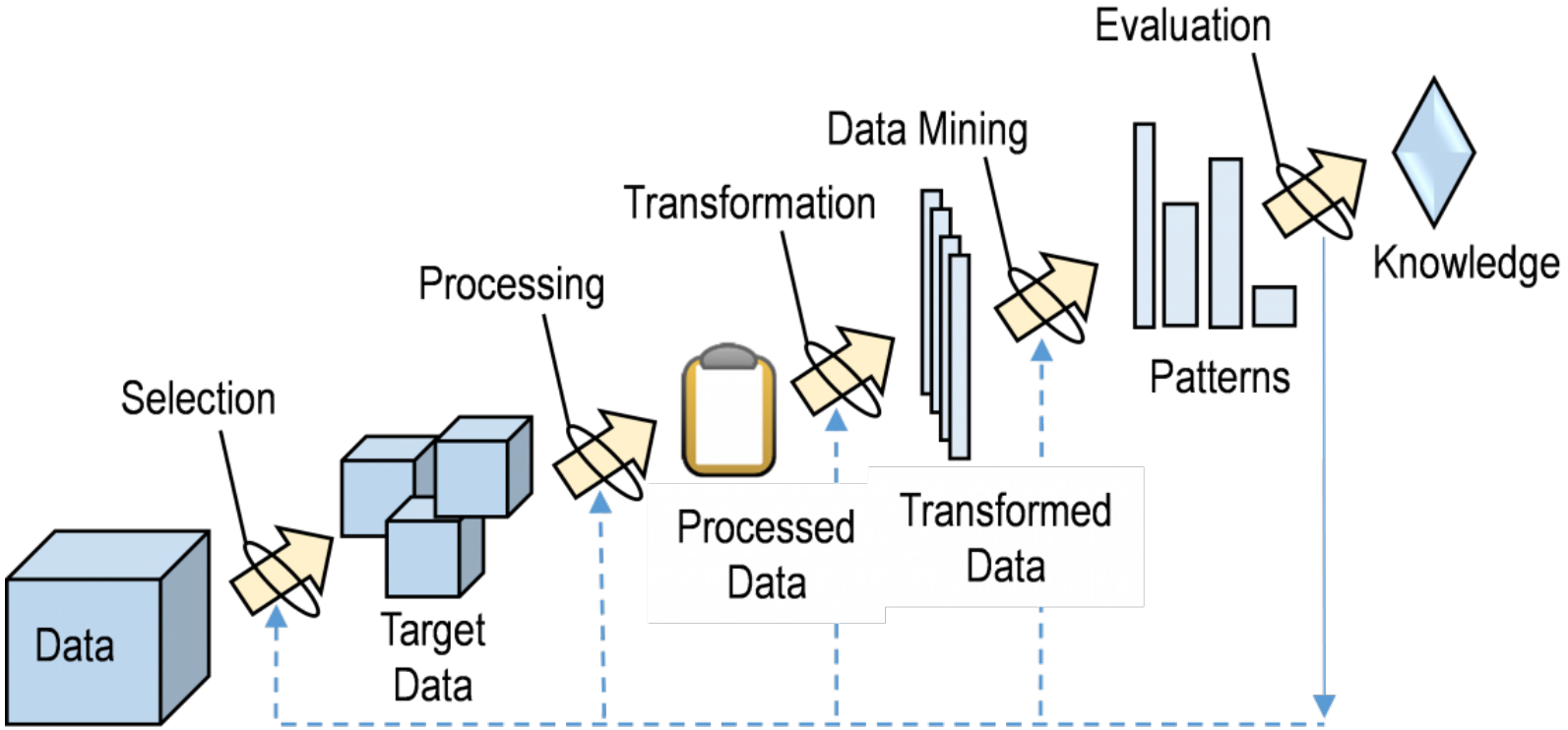
CPSC-547

KAIYUAN LI

Motivations

- Development of IoT (internet of things) and Big data system
- Higher requirement for visualization of different types of data
- The interrelationship between applications and information visualization technology

Figure data mining definition [1]



Outline

- Explain the relationship between information visualization and real-world application
- Categorize different types of data from Big-data system
- List Current Vis-infor technology/tools and commends on each of them

Expectation

- Provides an insights for future Vis-infor technique and overview for current state of art
- Make contributions on awareness of importance of Vis-technique, data mining and big data period
- Be familiar with current technology

Reference

[1] “data mining definition”, no author, [online access] https://www.dragon1.com/terms/data-mining-definition

[2] “Information visualization and visual data mining”, D.A. Keim, IEEE Transactions on Visualization and Computer Graphics ,Vol: 8, Issue 1, aug.07.2002

[3] E.Achtert , H.P.Kriegel , E.Schubert , A.Zimek, Interactive data mining with 3D-parallel-coordinate-trees, Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, June 22-27, 2013, New York, New York, USA

[4] S. Liu, W. Cui, Y. Wu, and M. Liu. A survey on information visualization: Recent advances and challenges. The Visual Computer, To appear, 2014.

The State of the Salmon:
Visualizing salmon population trends

Michael Barrus

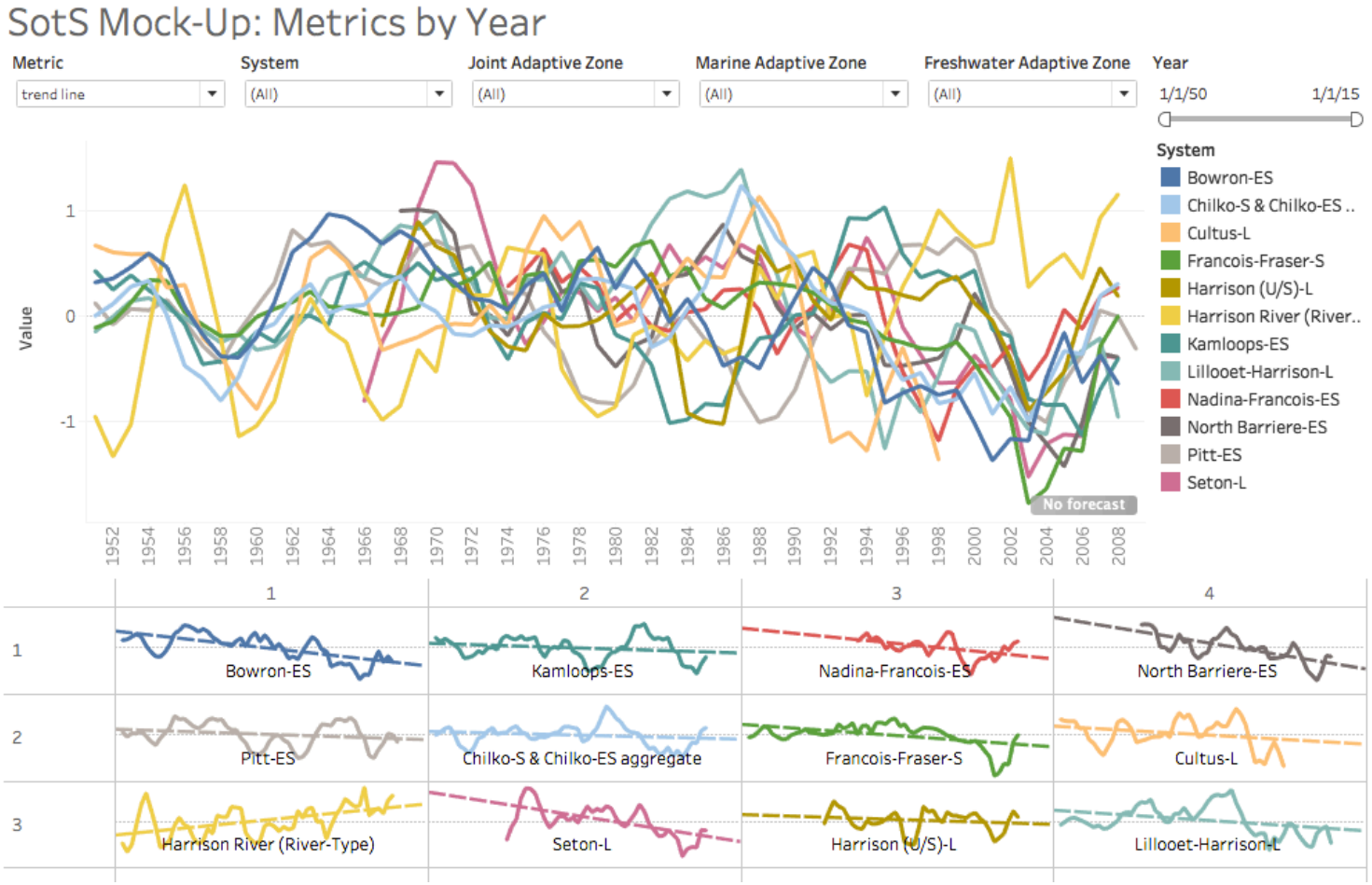
Overview

- Many salmon populations in BC are in decline but the causes are unclear
- Federal Department of Fisheries is tasked with understanding these trends, but has not analyzed these holistically
- Appropriate visualizations could aid data exploration and improve insight, management

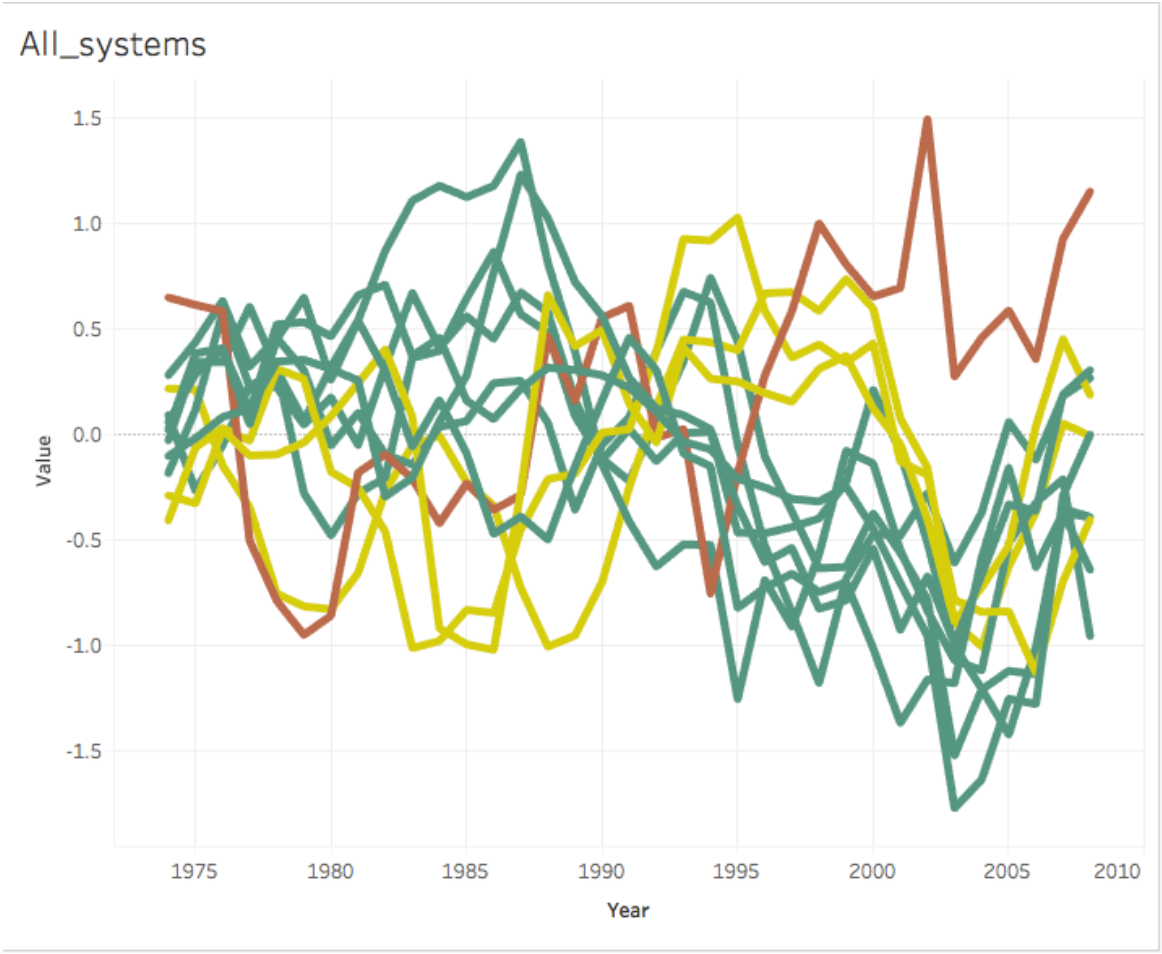
Objectives

1. Define tasks and needs of salmon researchers within Department of Fisheries
2. Build a series of visualizations to facilitate exploration
3. Conduct user studies to evaluate ability of vis to promote understanding

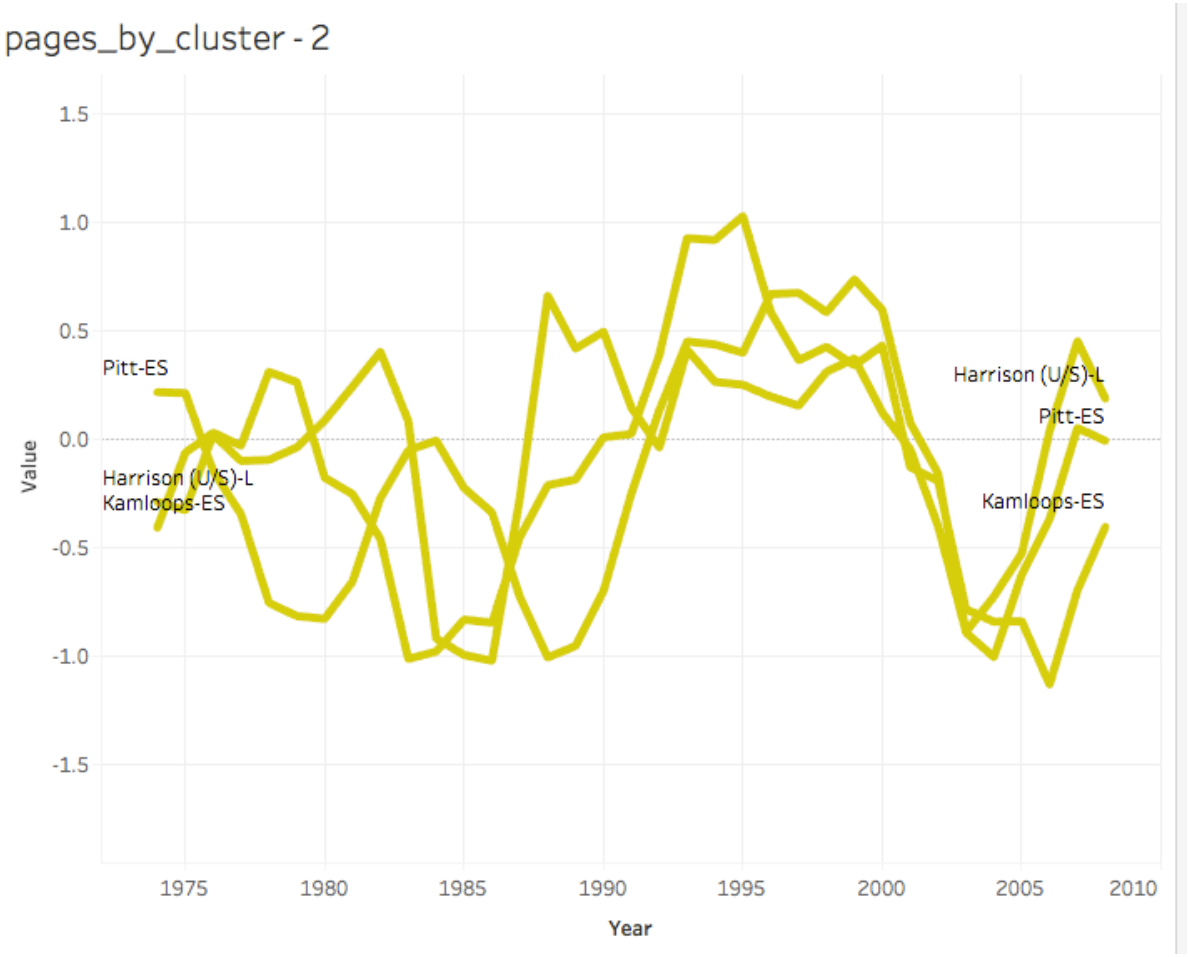
Mock ups: Population trends



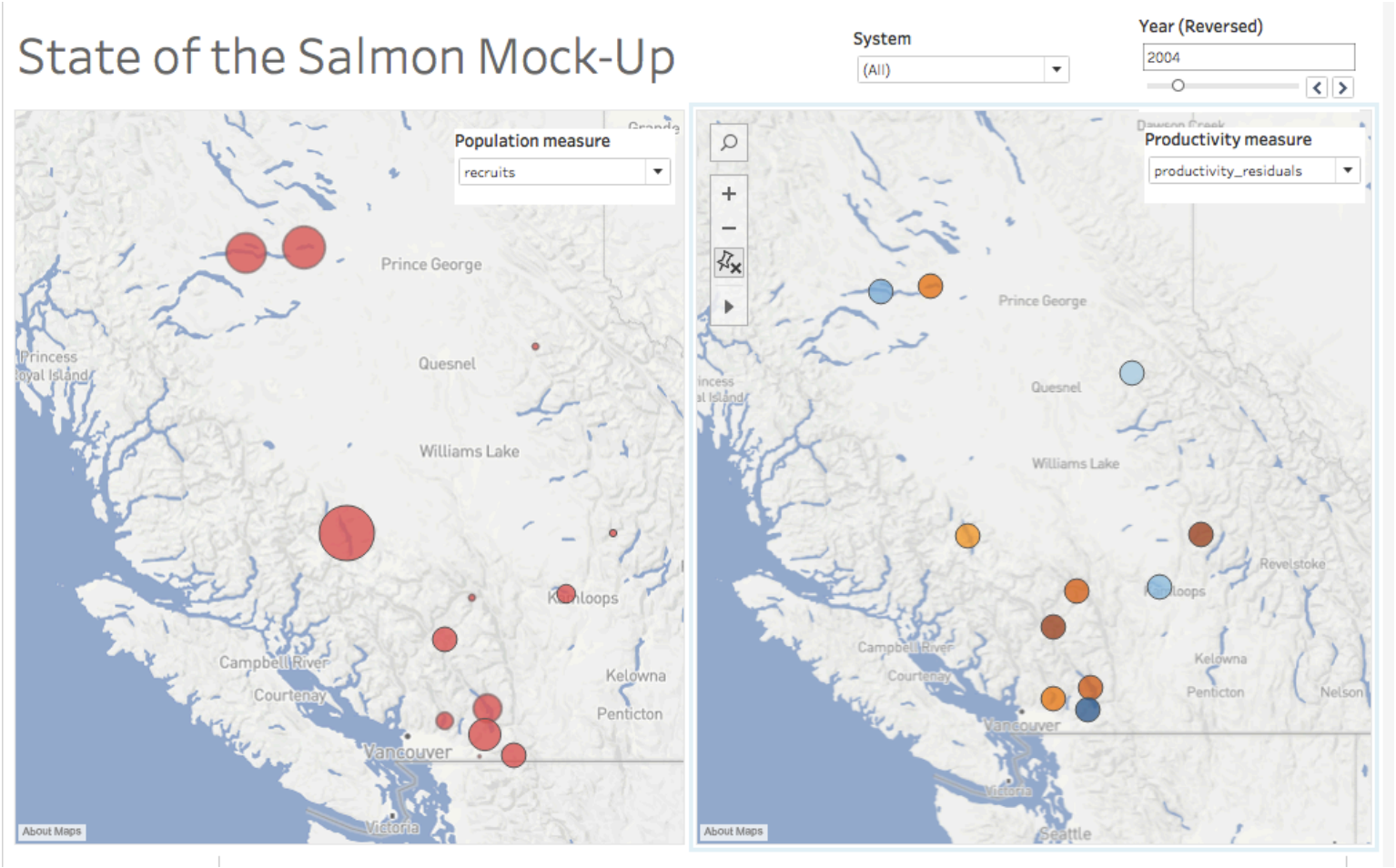
Mock ups: Population trends



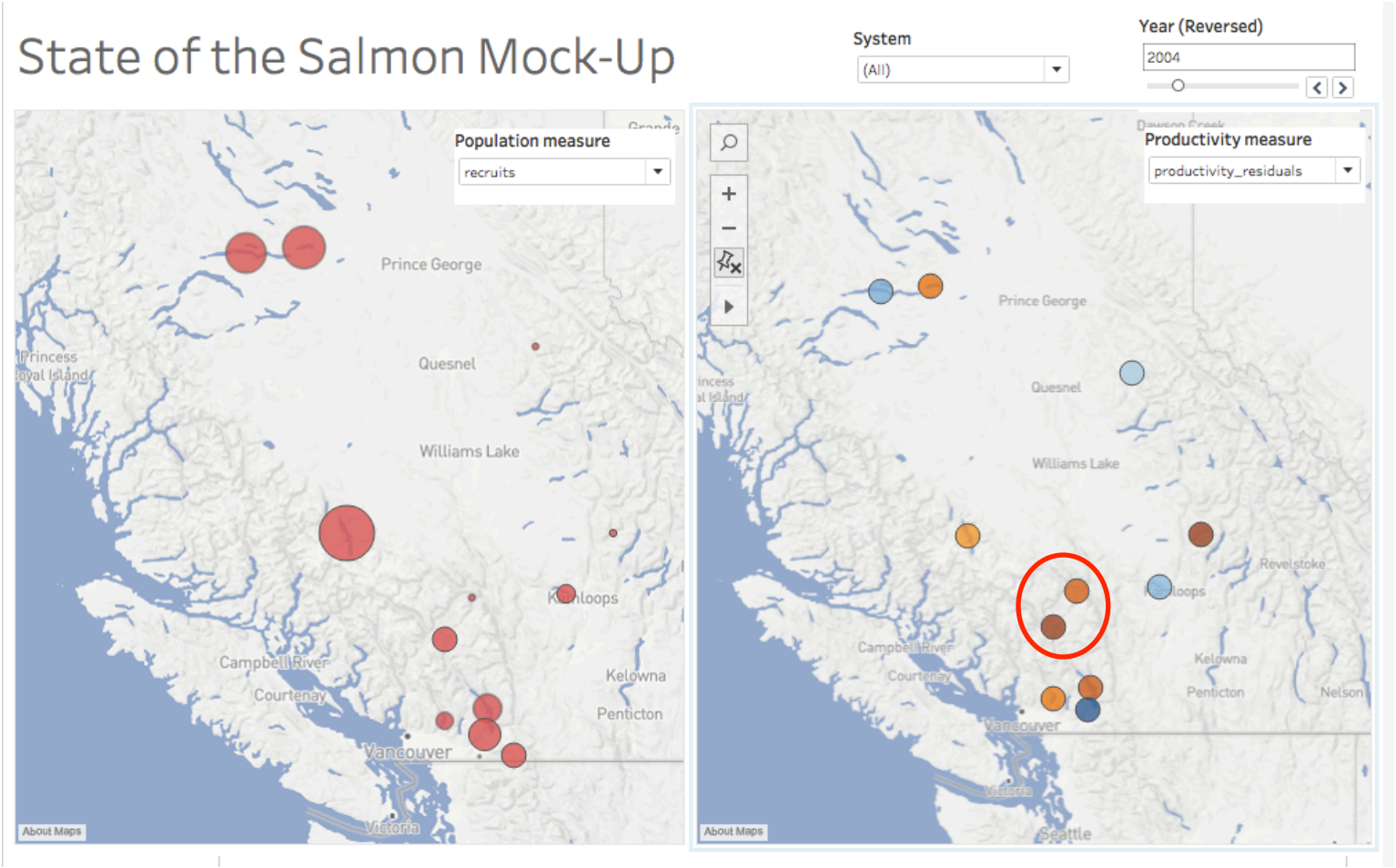
Mock ups: Population trends



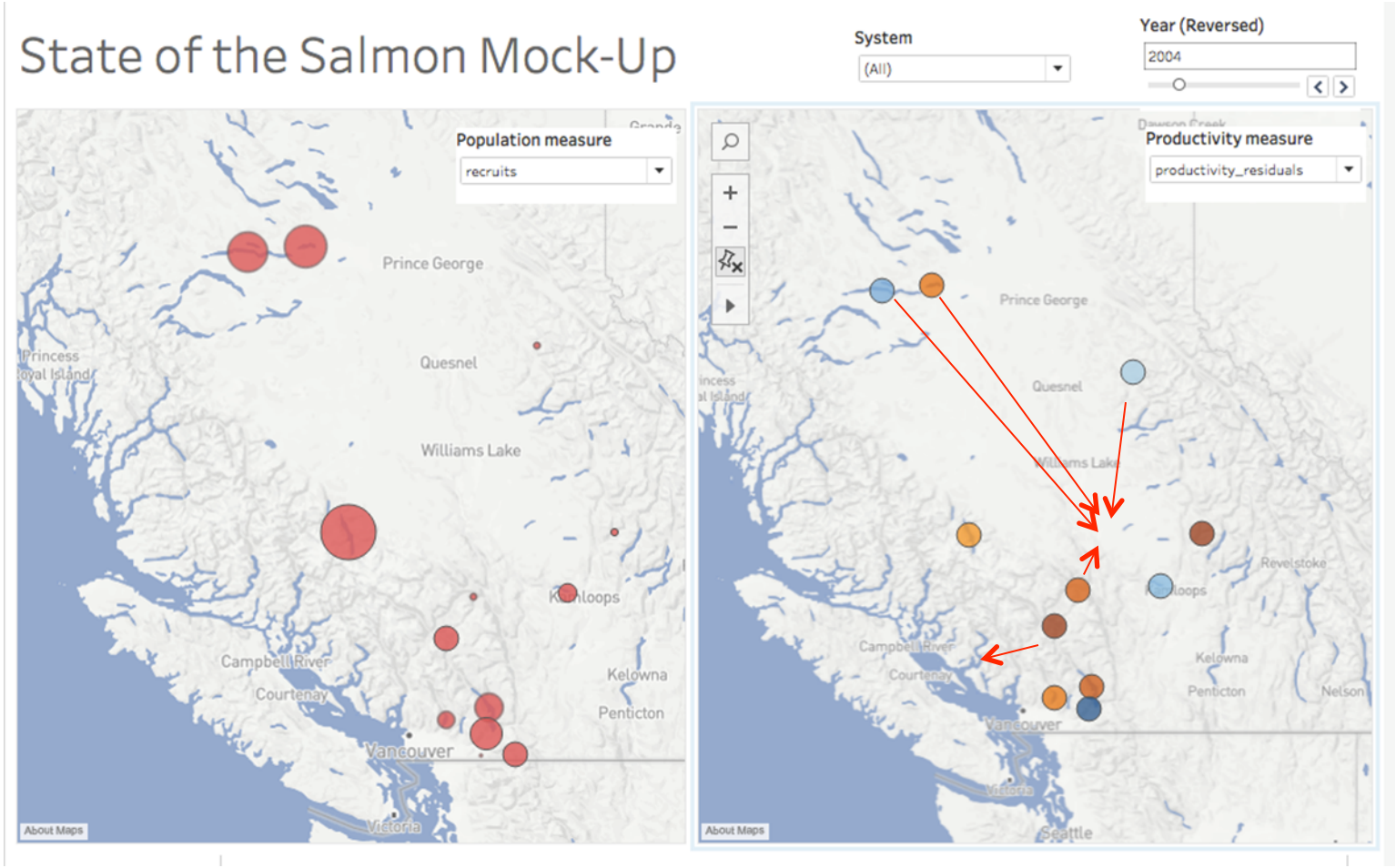
Mock up: Geographical data



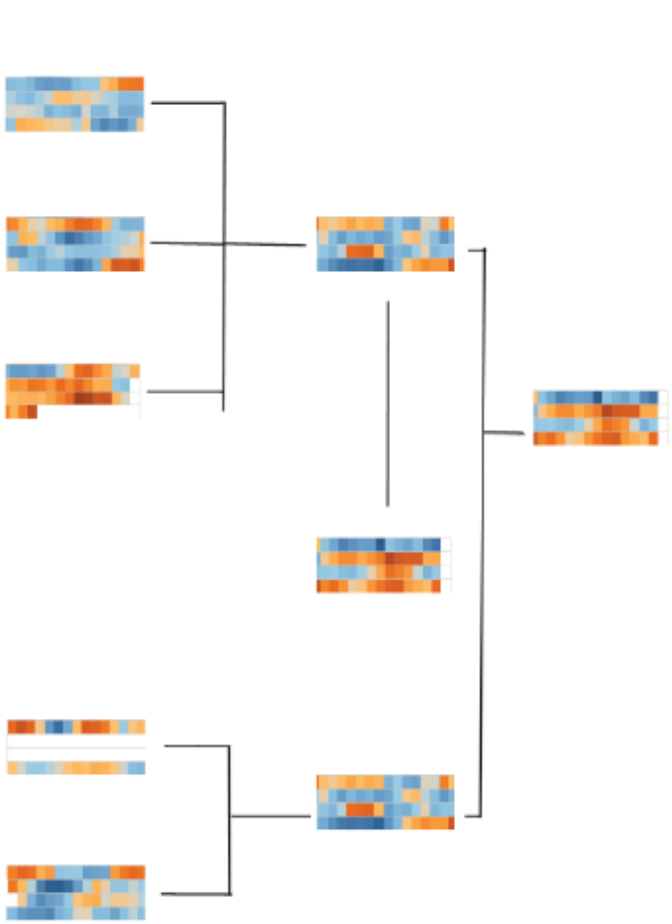
Mock up: Geographical data

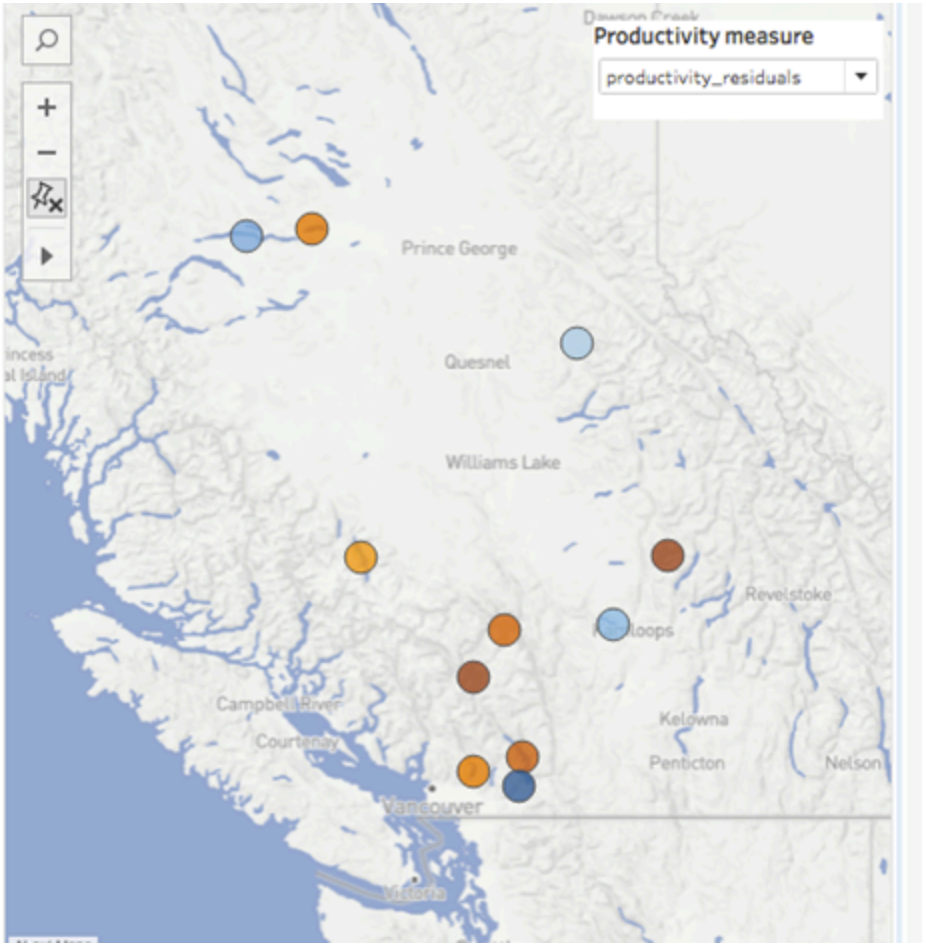


Mock up: Geographical data



Mock up: Geographical data





Outcomes and significance

1. Development of specific tool that improves understanding of highly significant salmon population
2. Development of methodology that quantifies how well vis helps understanding in general
3. Development of list of hypotheses that are prioritized by panel of experts

PROJECT PITCH

Peyvand Forouzandeh

SCOPE

- Municipal data in accessible formats with open licence
- City of New Westminster in Metro Vancouver, BC

PROJECT PURPOSE

- Improve visualizations in open data platform to have noticeable impacts on increasing efficiency, transparency, easiness of navigation and develop a mechanism measure the impact of open data program
- *Performance dashboard*: Produce understandable metrics, inspire thinking and allow monitoring
- Possibility of application program interface to build live communication channel between applications and datasets
- An attempt to increase citizen engagement and city operations with providing more organized, visually easy-to-read open data platform design and suggesting that can suggest in depth analysis of data and meaningful information
- Within Intelligent Cities Forum (ICF) framework. ICF indicators:
  - Broadband Connectivity
  - Innovation
  - Digital Equity
  - Knowledge Workforce
  - Sustainability
  - Advocacy

CURRENT STATE

- About 160 categories of tabular datasets in open data portal

Datasets

When you download datasets, you are agreeing to our licence.

Accessible Public Washrooms

Listing of all the accessible washrooms that are available within the City.

Metadata | CSV | DWG | JSON | KMZ | SHP | XLSX

Addresses

A list of addresses for the City of New Westminster.

Metadata | CSV | JSON | JSON | KMZ | SHP

Alternative Fuels and Electric Charging Stations

Electric vehicles are an environmentally friendly mode of transportation. As cleaner emission vehicles gain momentum across the lower mainland, the City of New Westminster is…

Metadata | CSV | DWG | JSON | KMZ | SHP | XLSX

Artists

Learn about New Westminster's vibrant cultural community. You can search out and connect with the people, places, businesses and organizations that bring culture to our city every…

Metadata | CSV | DWG | JSON | KMZ | SHP | XLSX

Bike Routes

This dataset contains bike routes including planned and current bikeways, on-street and off-street, as well as dedicated lanes.

Metadata | DWG | JSON | KMZ | SHP | SHP

Block Reference File

The blocks correspond to a division of the City into about 400 blocks, set up by the City Planner in about 1970. The purpose of these geographic descriptions was to enable more…

Metadata | SHP

Building Age

Building Attributes

CURRENT STATE

About

Listing of all the accessible washrooms that are available within the City.

Metadata

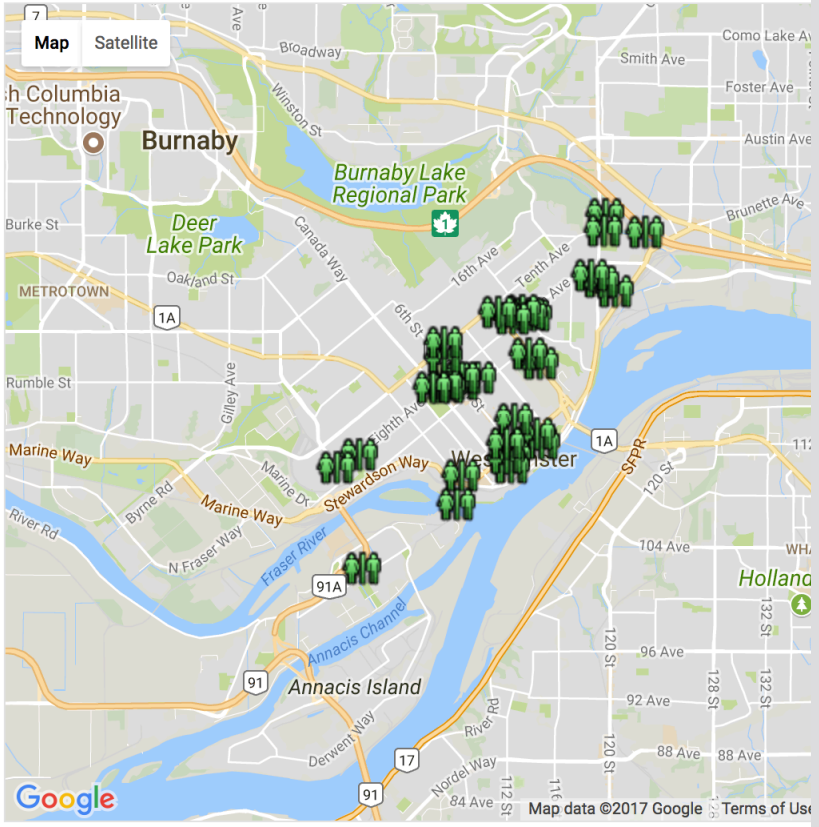
Update Frequency	Ad-hoc
Contributor	Parks and Recreation
Coordinate System	UTM10 NAD83
Geographic Coverage	City Wide
Use Limitation	The City of New Westminster publishes open data under the terms of the Open Government Licence - City of New Westminster. You are encouraged to use the Information that is available under this licence with only a few conditions.
Fields	NWID, Name, Address, Category, Neighbourhood, Hours, Source, Accessible, X, Y
Last Updated	Mar 27, 2017
Categories	Parks and Recreation

Downloads

When you download datasets, you are agreeing to our licence.

CSV (6 KB) | DWG (19 KB) | JSON (14 KB) | KMZ (8 KB) | SHP (3 KB) | XLSX (53 KB)

Preview



DATA FORMATS

- CSV: These files are used for tabular data, and can be opened in software like Excel or Numbers. It can also be viewed as plain text in applications like Notepad.
- KMZ / KML: These files are used for mapping data, and can be opened in Google Earth. It is also used for data previews on the website.
- SHP: A shape file contains geographical reference data as individual objects such as a street, a river, a landmark or a zip code area. Features exist as objects and their attributes within the SHP file. Shapefiles can be viewed using a application: ArcGIS and most GIS software applications.

CPSC 547 – Project Pitch

Eye Movement to Evaluate User Experience

BY

SHAREEN MAHMUD

Motivation

- Imagine a usability test in which the user attempts to buy a laptop online. On the homepage, he quickly finds the “laptop” link, but on the next page he hesitates. “I wasn’t sure where to click! There were a lot of options.”

- What if we (designers) could see **what he saw**

Information

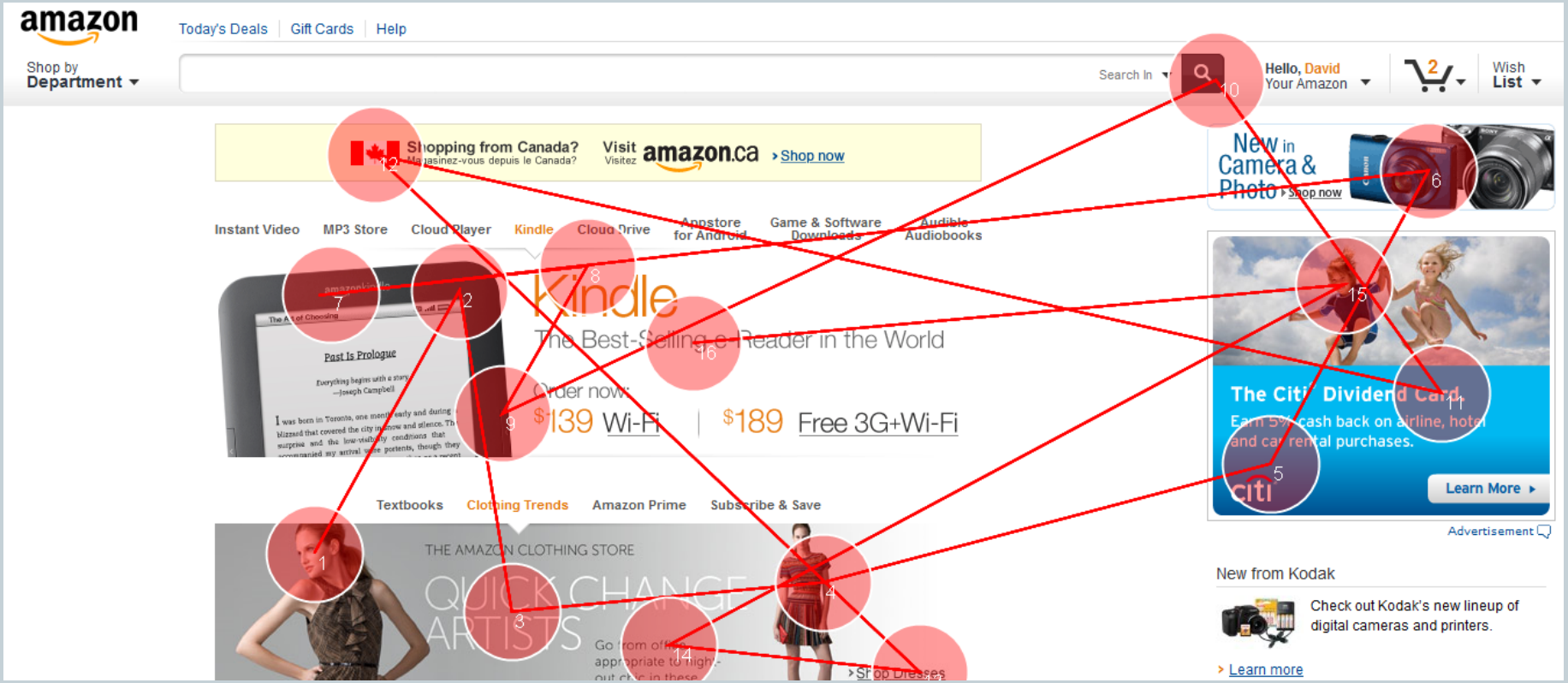
- Eye movements data can identify fixation points- where the user's gaze lingered for some time.
- It can also identify the point at which the user's gaze rapidly move to another position.

Visualization

- Heat Maps can be used to reveal the focus of visual attention.

Visualization

- Gaze Plots can be used to reveal the order in which users moved their gaze. Size encoding can be done to represent time duration.



Data Sets

- The Massvis MIT group has publicly available eye movement data of a number of participants looking at different visualizations.

- I am looking for other possible data sets that require visualization to evaluate user's experience in interacting with a system.

Thank you

Visual Exploration into the Factors Leading to Absenteeism in the Canadian Workplace

A Problem-Driven Design Study

CPSC 547:The Pitch

Shirlett Hall

Introduction

- Background
  - Absenteeism is the absence with or without pay for at least half a day but less than 52 weeks from work
  - In 2011, the estimated cost of absenteeism was over $16 billion
  - Less than half of Canadian employers track employee absences
- Source: Conference Board of Canada

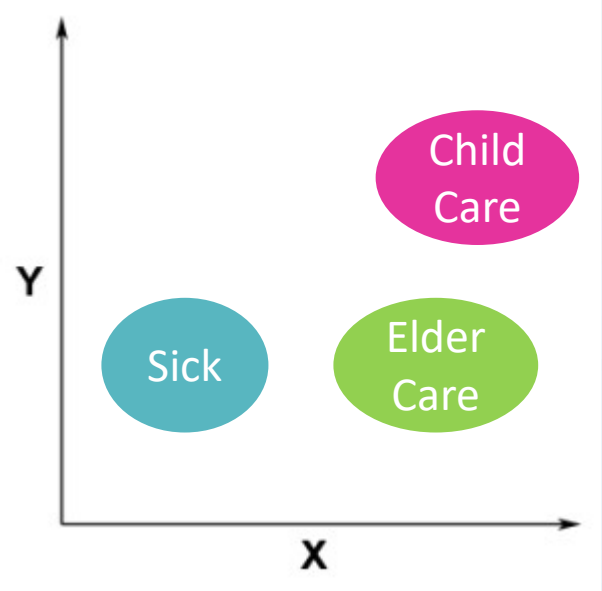
- Motivation
  - The per capita productivity of Canada lags behind many of its counterparts like the US and Australia
  - Absenteeism plays a role in the overall productivity of the country
  - Employers must not only have the ability to track absenteeism but also identify the factors so there is a chance for corrective action

Process

- Source Data – Monthly Labor Sample Survey report from StatCan on UBC DataVerse

SEX	Highest educatn attainment(1990 onward)	Class of Worker - Main Job	Industry of main job	Tenure of current job in months	Usual Hourly Earnings	Union Member
SEX	EDUC	COWMAIN	NAICS_21	TENURE	HRLYEARN	UNION
2	6	2	17	94	15.4	3
2	4	1	17	204	22	1
2	2	2	17	114	12.4	3
2	4	1	17	159	27	1
1	4	1	17	107	25.53	2
2	0	2	17	4	17	3
2	3	1	17	240	26	1
2	4	2	17	233	27	3

- Visualization Tool – R with ggplot layers



Automated Image Feature Quantification

Theodore Smith

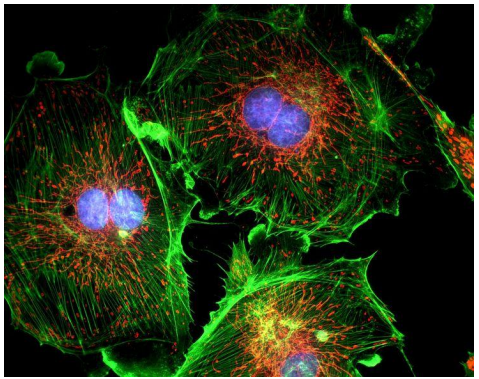
CPSC 547

October 17, 2017

Concept

- Images frequently contain a large number of target features
  - Analysis by hand is time consuming and prone to error and bias
- These features can be extracted using automated processes
- Transformation of the original image based on automated feature identification simplifies and accelerates human analysis
- Generation of secondary, descriptive statistics guides inference
  - Number of features
  - Density of features
  - Spatial variation of feature distribution
  - Quantitative likelihood of feature identity

Applications



Goals

- Coarse-grained quantification of features of interest
  - Initially, no attempt will be made to apply sophisticated annotations to identified features
  - Intended to augment, rather than replace human interpretation of output
- Rendering of reduced-form image
  - Isolate features of interest from background
  - Represent features with simple, distinct area marks
- Generation of descriptive statistics
  - Number of target features in frame
  - Region-based density of target features
  - Confidence metric

Implementation

- Pre-processing
  - Contrast enhancement
  - Grey-scale conversion (depending on input and statistical method)
- Possible feature identification methods
  - Independent Component Analysis (ICA)
  - 2-D Fourier Transformation
  - Artificial Neural Network (with sufficiently large training set)
  - Brute-force edge detection
- Outputs
  - Reduced-form image generation
  - Descriptives

Visualization of Eye Tracking Data

Vanessa Putnam

Why Eye Tracking?

- Eyetracking is important for evaluating user behaviour.
- Analysing eye tracking data is used in many fields for research such as: Psychology, Medicine, Usability, HCI, and Information Visualization. Just to name a few!
- Usually done quantitatively, but recently a more qualitative approach is being explored based on visualization techniques.

MetroQuest

- MetroQuest is an interface used to address the problem of building a new transportation system on the UBC campus.
- This study investigated the impact of individual differences on user experience and gaze behavior with MetroQuest.
- Gaze, Pupil, and Head Distance features were collected to predict user characteristics during interaction with MetroQuest.
- The study explores how some user cognitive abilities relevant for processing information visualizations can be predicted from eye tracking data.

Prior Work

- Eye Tracking device collects raw data of recorded gaze points
- These gaze points can be aggregated into *fixations* and *saccades* for measuring which areas on the stimulus have been focused on.
- Areas of interest (AOIs) also identified to concentrate the analysis to specific regions.

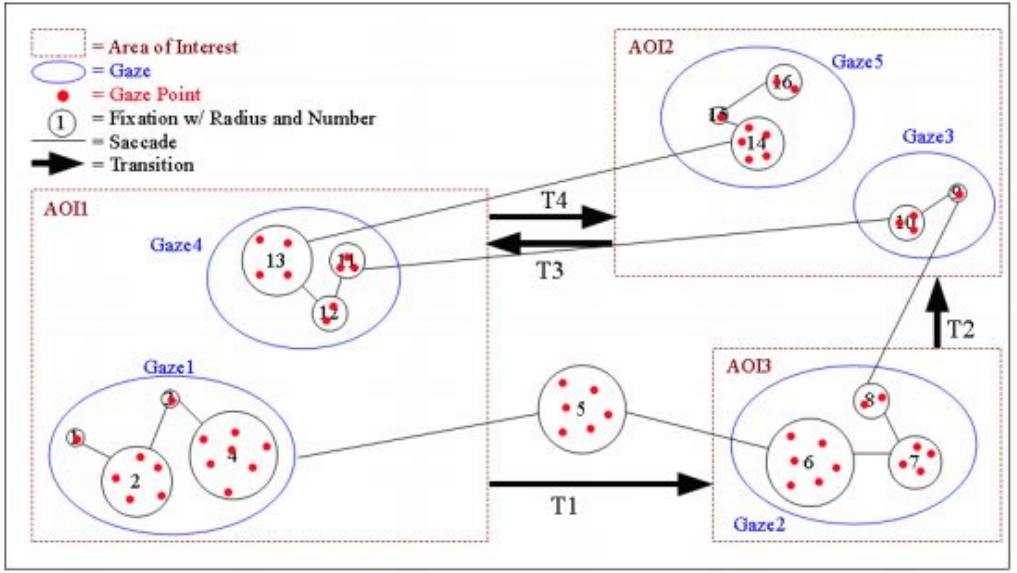


Figure 2.
State-of-the-Art of Visualization for Eye Tracking Data
T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf & T. Ertl

Works Cited

[1] Cristina Conati, Sébastien Lallé, Md. Abed Rahman, Dereck Toker, 2017. **Further Results on Predicting Cognitive Abilities for Adaptive Visualizations** Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main track. Pages 1568-1574. https://doi.org/10.24963/ijcai.2017/217

[2] T. Blascheck , K. Kurzhals , M. Raschke , M. Burch , D. Weiskopf and T. Ertl, 2014. **State-of-the-Art of Visualization for Eye Tracking Data. Eurographics Conference on Visualization** (EuroVis) (2014).

[3] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf and T. Ertl, 2017. **Visualization of Eye Tracking Data: A Taxonomy and Survey**. COMPUTER GRAPHICS forum Volume 00 (2017), number 0 pp. 1–25.

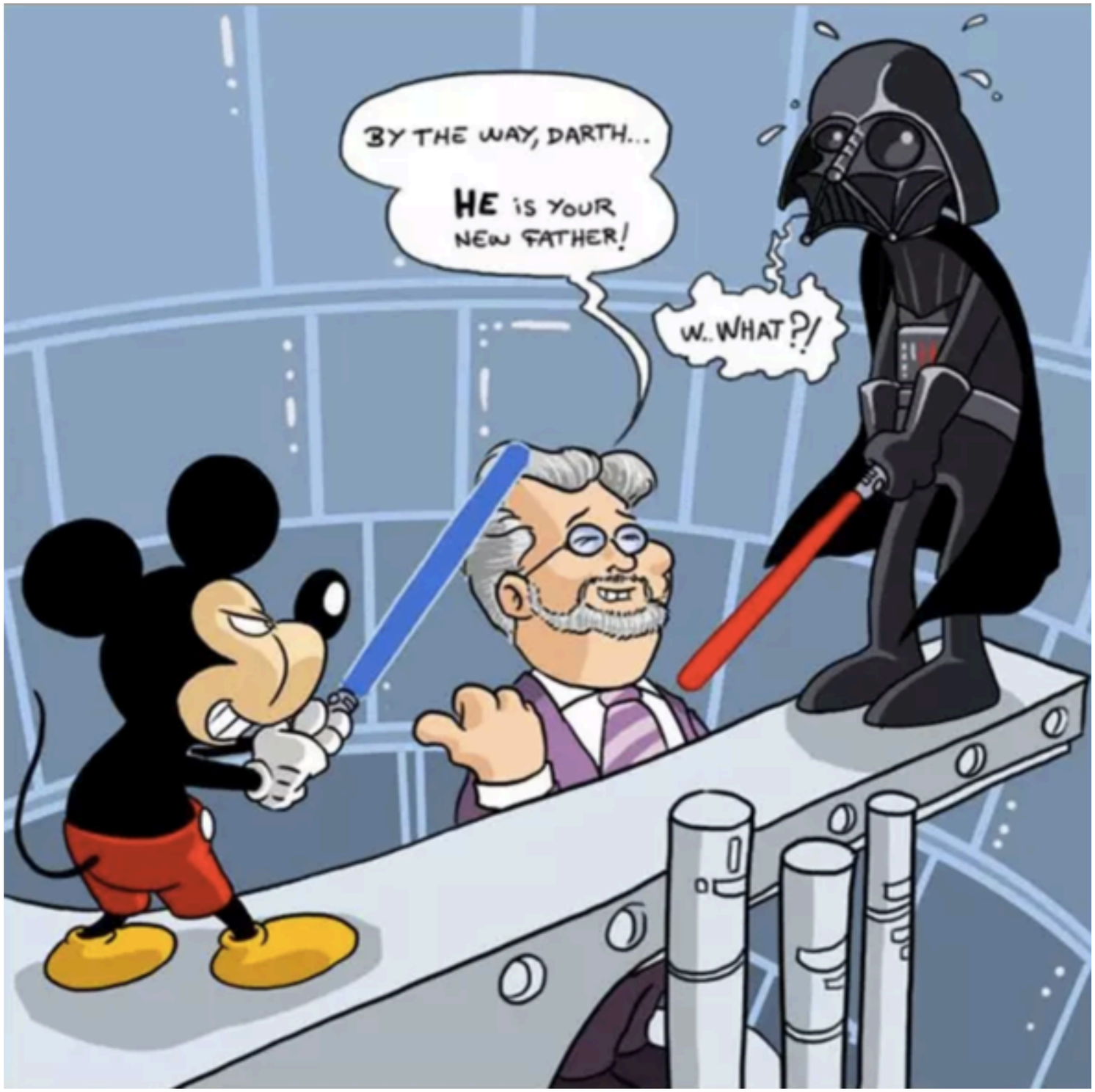
Visualization of Marvel Films Data

Zixiao ZHANG

10.17

Background

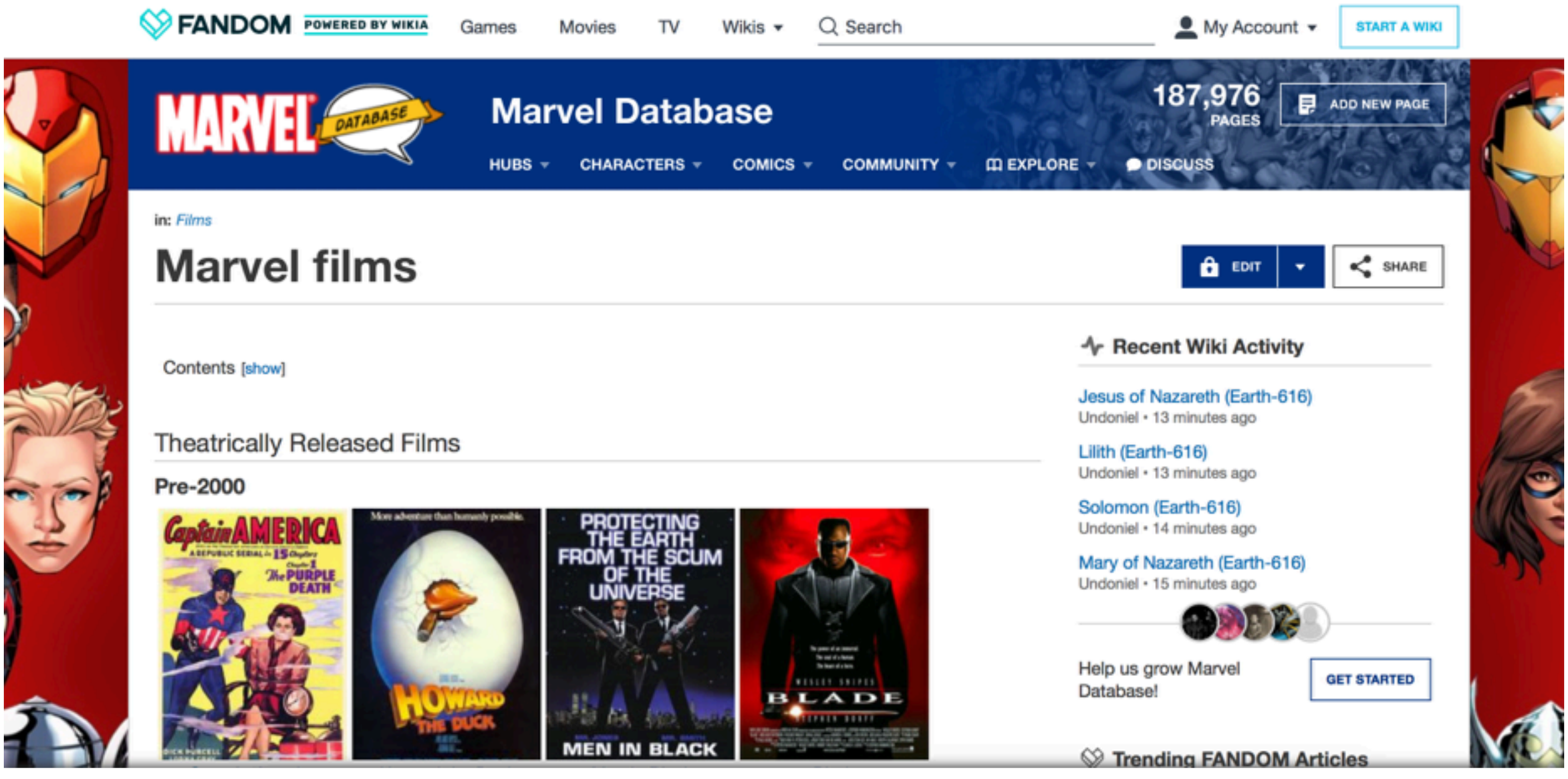
- When people watch the movies like I*ron Man* or *Star War Series*, they may feel confused without making enough preparations.
- Some characters appear in multiple films.
- Most audience will get a better experience by simply getting some general ideas but not digging into the information.



Main Design Task

Present more details based on the characters and their relationships

Data Repository

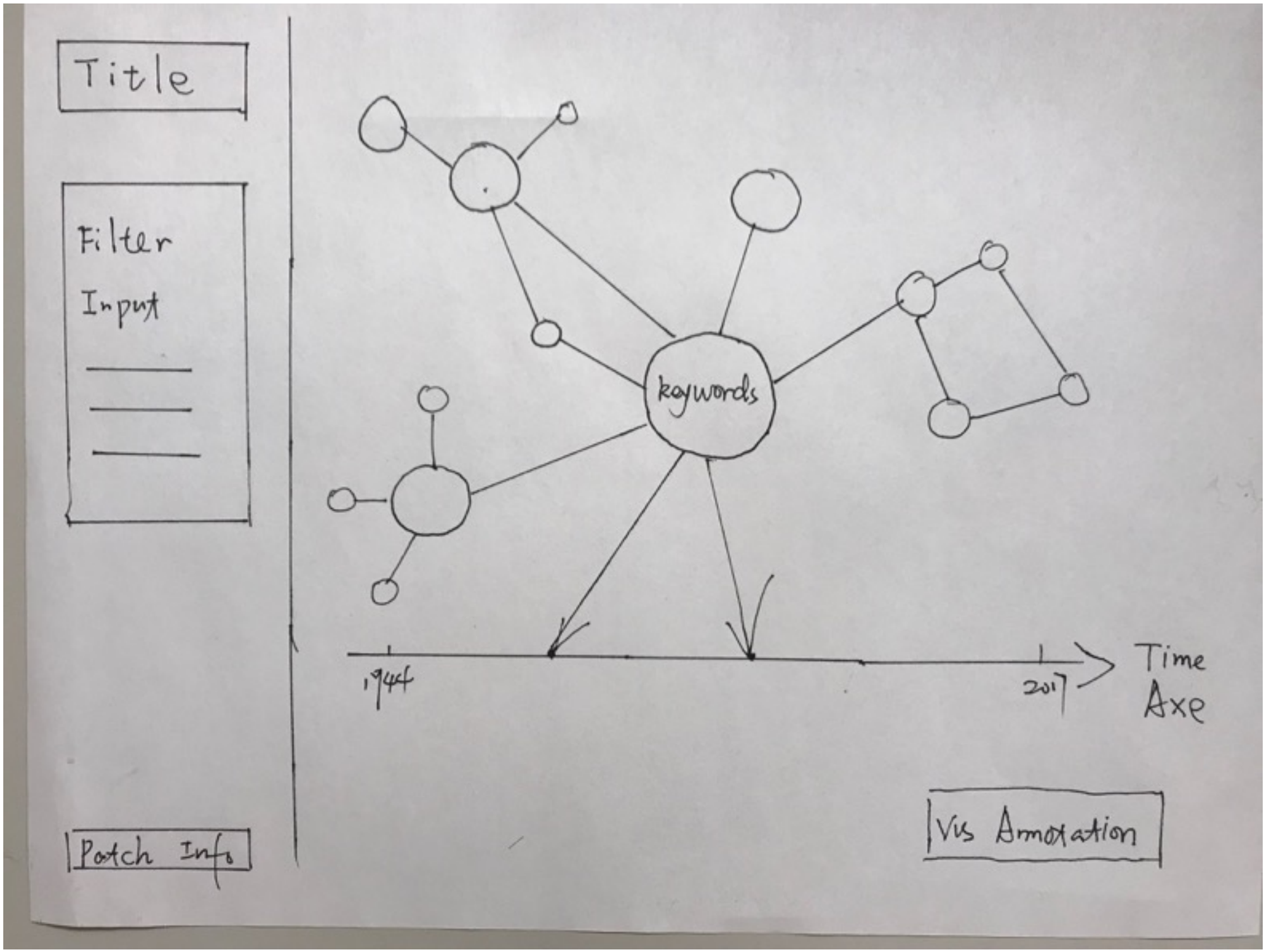


http://marvel.wikia.com/wiki/Marvel_films

Prototype

- Networks is used to interpret the relationships.
- Time (year) is considered as a crucial key.
- A widget for the user to filter the result by entering key words.
- More information such as directors can be shown by clicking the nodes.
- Algorithm needs to be designed to arrange the network structure.

Sketch


Title
Filter
Input
keywords
1944
2017
Time
Axe
Patch Info
Vis Annotation

Issues for consideration

- What kind of the information do the common audience look for?
- Will the movie fans have special needs than others?
- How can we present the details of actors (actresses) and characters simultaneously?
- What standard must be set up for filter?
- How to make the interaction naturally?

Steps

- Collect and analyze the user’s requirement
- Determine the details to be shown
- Encode the data format
- UI Design
- Primary Visualization
- Interaction design and Optimization

Thanks!