Grad student life: Cost of Living @UBC

Jingxuan (Carol) Huang, Devyani McLaren, Tommy Nguyen

Index Terms—Explainer visualization, Comparison visuzation

1 Introduction

It is safe to say that finance has become an important topic for everyone in recent days—especially for students [10]. Inflation in Vancouver reached a recent high of 8.8% [4] and the average rent for a 1-bedroom apartment is at CA $2,500[1]. In January 2022, a small collective of Computer Science Graduate students at the University of British Columbia came together to compile a cost of living report [15]. The results were an anonymized rich dataset that provide quantitative and qualitative insight into these students' finances and life contexts.

Our aim is to take this rich data and provide more robust and interactive visualization solutions beyond the simple charts and graphs outlined in the report. We hope to create an interface where users can learn about the income and expenses graduate students have and understand how their current financial situation compares to current graduate students in the UBC Computer Science Department.

Our team consists of graduate students sharing experiences in research, development, and design. Although we have little experience in this topic, we are actively affected by it. Given the rising prices and our graduate student stipend, we are constantly making decisions in our life that stem from our finances. Our team takes a lot of interest in this topic and is passionate about delivering a strong visualization of a graduate's cost of living to help inform current students, prospective students, and administrators to make informed decisions.

2 Related Work

Gleicher et al. [11] highlight the difficulty and complexity involved when making comparison-type visualizations. They define an abstraction taxonomy for such visualizations which consists of "*juxtaposition* - showing different objects separately, *superposition* - overlaying objects in the same space, and *explicit encoding of relationships* - directly encoding connections between objects visually." Types of these visualizations can be combined to enhance a viewer's understanding of the data. For example, when juxtaposition is combined with explicit encoding, users can see different views of the data while also understanding the connections and relationships of the data. We will use this data abstraction taxonomy to better plan and understand our proposed project.

Explainer articles are often created to promote active reading and a greater understanding of data presented in an article [12, 18]. These types of articles contain written contextual information but also allow readers to understand what is written with interactive visualized examples. "A visual introduction to machine learning" [14] is an explainer article example that uses the action of a user scrolling to understand complex machine learning information (see Figure 1).

To the best of the author's knowledge, no specific visualization-focused work has been done on student expenses/income [2]. However, there is a past example of a short explainer-type article for the cost of living in the United States, organized by spatial (region) data [1]. Other articles available typically list cost of living data in written or tabular format only [5, 6]. Figure 2 provides an example of how financial data can be represented by region [3]. As such, there is a large opportunity to explore how the cost of living data can be visualized, particularly if the data is focused only on one region and spatial data is not required.

The Cost of Living Estimator, created by Raj Tiwari [16, 17], is a comparator-type interactive visualization that takes user-inputted expenses/income data and shows users how their income compares to others living in large cities in the United States. This interface uses colour hue identity channels and length and position on common scale-type channels. It is unclear why a bar graph and scatter plot idioms were chosen to represent the data. Figure 3 shows the visualization used in this interface [4].

3 Abstraction

3.1 Data Abstraction

Our visualization data consists of a qualtrics survey that was sent out to all current UBC graduate students [5], which consisted of questions about student income and expenses [15]. An anonymized version of the survey data was released in June 2022 [3]. In total, there were 94 responses from a mix of Ph.D., MSc and Ph.D. track students. The dataset (139 items x 26 attributes) is tabular and mainly contains ordered quantitative and sequential data. These attributes (columns) are where students listed their income and expenses for different aspects of their life (e.g. food, housing, travel etc.). The other attributes consist of categorical data (e.g. are they the primary care provider?). The items of this dataset are individual students that responded to the survey. Please refer to our dataset (Dataset on google sheets) to see our data abstraction for the attributes.

An additional dataset available to us, also created by the same group of CS graduate students compares graduate student funding at different universities (available at link). The data is also tabular and attributes consist of ordered, quantitative, sequential data. Currently, data is listed for 12 universities (UBC included) and only compares PhD. level stipend funding from those universities.

As described in section 2 previous cost of living visualizations have used geospatial datasets and/or reported additional aggregate data about the larger population living in a certain area. Numbeo [6], is a large cost-of-living database that uses crowd-sourced data to gather information about housing cost, healthcare quality, crime rates etc [2]. This database is organized by region and could be useful for our comparative purposes should we decide to show our target users how current student finances compare to those living in Vancouver. Figure 4 is an example of the kind of data we could use from Numbeo [7].

- *Jingxuan is with the DMM Lab, UBC Dept of Computer Science.*
- *Devyani McLaren is with the SPIN Lab, UBC Dept of Computer Science.*
- *Tommy Nguyen is with the SPIN Lab, UBC Dept of Computer Science.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

[1] https://www.zumper.com/rent-research/vancouver-bc

[2] Although we are still actively looking

[3] Evolytics link

[4] Cost of Living Interface Link

[5] as of May 2022

[6] Numbeo website link

[7] Numbeo website link

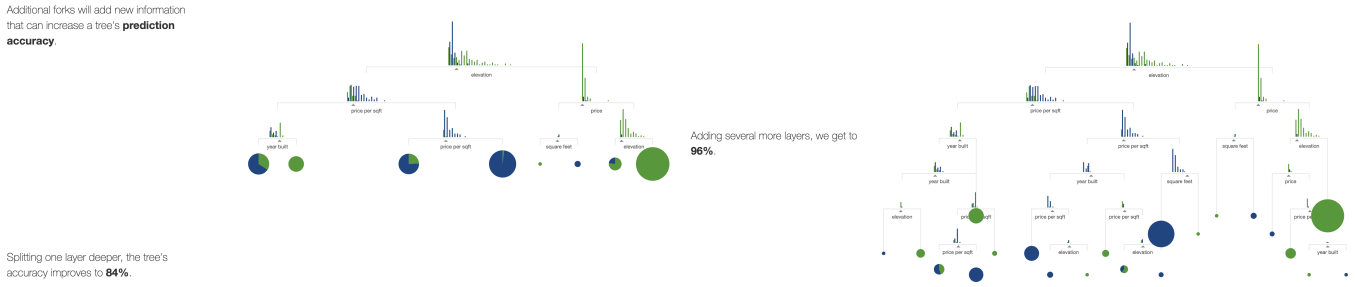

Figure 1: Left: example of contextual information with visualization, Right: example of how the visualization changes after completing the action of scrolling

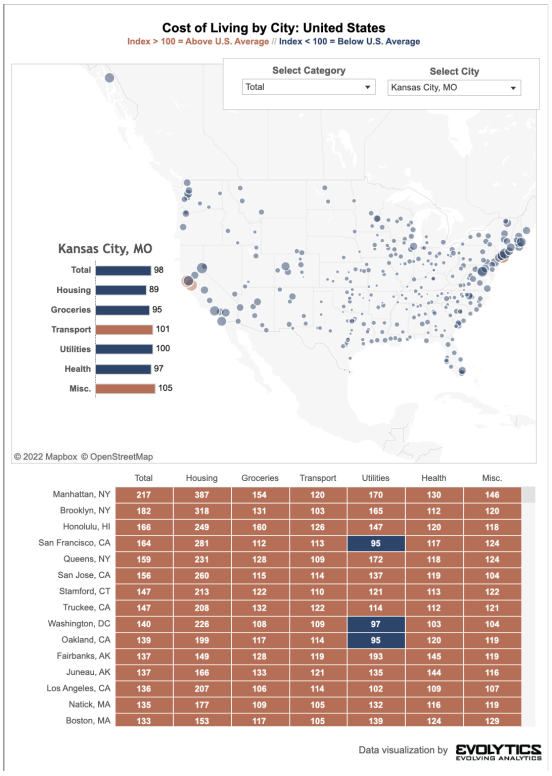

Figure 2: Example of cost of living data spatial and tabular data.

3.2 Task Abstraction

In this section, we identify our end users and list the tasks our work aims to resolve. We also abstract these tasks with visualization terminology.

3.2.1 Who: Identify end users

Our work is designed for a student base population and university staff. Specifically, we would like to focus on designing for computer science (CS) graduate students and prospective CS graduate students. We hope our work can:

- educate and assist graduate students in the UBC CS department,
- persuade decision-makers in UBC to make changes, and
- inform the prospective students.

Therefore, our work should be informative, intuitive and engaging. In that sense, the explainer is the most suitable way to achieve our goals.

3.2.2 Tentative task list

Based on the end users we identified and the datasets available to us, we summarize some tasks that may attract our users and are solvable with existing data. For each task, we also put the domain-specific description into abstract terms provided in [13]. We believe this abstraction will make our work more generalizable and comparable with similar visualizations in different domains.

1. **Composition of expenses**. As graduate students, they need to spend on lots of things every day. In our work, we focus on seven major categories of basic expenses: food, social, health, housing, utilities, commute, and others. The first thing our users might be interested in is the overall structure of expenses. This overview can answer some basic questions such as ”which category accounts for the most expenses?”.
    - Actions: Summarize
    - Targets: Part-whole relationship
2. **The trend of cost of living in Vancouver**. One thing that stresses graduate students out is the ever-increasing cost of living in Vancouver. However, most of them only have a rough idea about it. To help them clearly understand how fast it grows, our work plans to visualize this trend with additional data from NUMBEO.

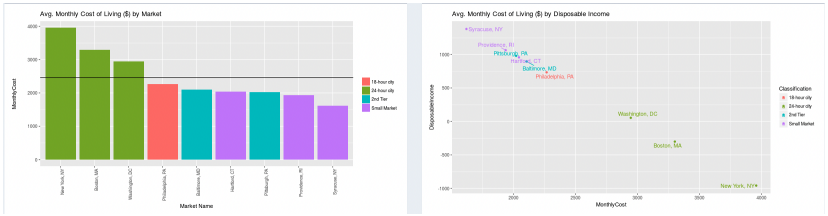

Figure 3: Idioms used in a cost of living comparator interface.

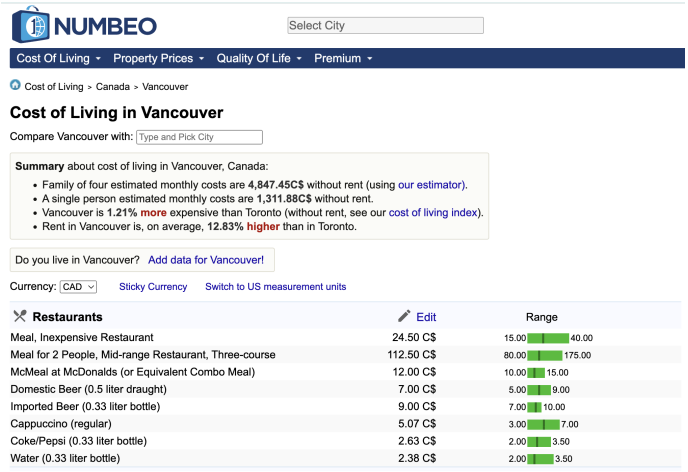

Figure 4: Example data that Numbeo contains for the city of Vancouver.

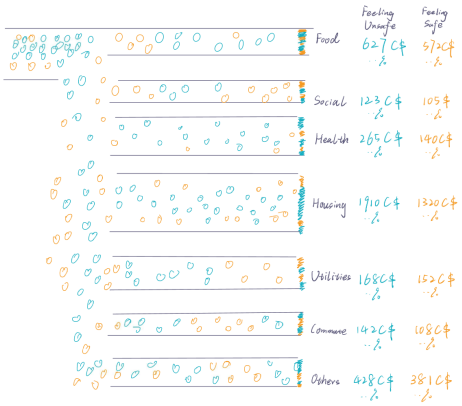

Figure 5: A unit visualization to show the composition of expenses. The categorical attribute "Safety" is encoded in hue. The animation will be similar to [8].

- Actions: Discover
- Targets: Trends

3. **The gap between funding and expenses**. While the cost of living is growing all the time, the funding of CS graduate students is not matching up. So what's the gap, and should students feel worried about it? Our work tries to answer this question by showing an overview and helping users find their own position in the graph.

   - Actions: Summarize, Compare, Locate, Annotate
   - Targets: Distribution

4. **Competitiveness of UBC CS department in terms of funding**. Though students usually choose graduate schools based on multiple criteria, it is true that funding plays a critical role in their decision-making process. Given the increase in the cost of living in Vancouver, it is concerning whether the UBC CS department will be less competitive due to insufficient funding. We thus use data from similarly ranked universities [9] around the world and compare their funding-to-cost ratio with visualization.

   - Actions: Compare
   - Targets: Distribution

4 Solution

4.1 Visualizations

The final artifact of our work will be an explainer, an article with interactive components embedded in it. The explainer is a good way to serve our target users, as it will boost learning and engagement for readers compared to its static counterparts [12]. We also considered several visualizations that may be included in our explainer. Note that the figures used in the visualizations below are fabricated and do not represent our datasets.

Figure 5 is the sketch of an animated unit visualization to display the composition of expenses. It is based on the insights from an interactive article in *New York Times* [8], where a similar visualization is used to highlight the inequality between white and black boys in rich families. In our context, we preserve the marks and channels they use in this idiom but encode different attributes. Each unit will denote 1 Canadian dollar, and the animation encodes the spending. We also use volume to encode the expenses in each category. Since we have some categorical attributes in our datasets and each only has 2-3 levels, it is appropriate to use hue to encode them and hence make comparisons among groups. Figure 5 shows an example of encoding Safety (i.e., "Do you feel safe in your current grad stipend amount") in hue.

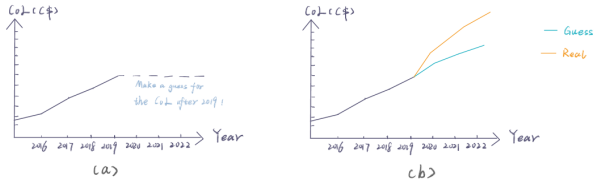

Figure 6: "You Draw It" visualization to show the trend of cost of living in Vancouver. (a) It first only shows part of the line and encourages readers to complete it. (b) After users draw down their own answers, the plot will be updated and show the real trend. Users can have a deeper impression of data by comparing two answers.

Our second visualization is inspired by the "You Draw It" visualization to show the trend of CO2 emission [12], an idiom first proposed in *New York Times* [7]. It is a line chart embedded with an interactive component, which can better engage readers and leave a deep impres-

sion.

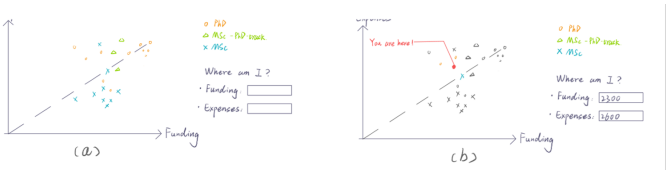

Figure 7: Scatterplot of funding versus expenses. (a) Original scatterplot. (b) User can input their own data and locate their position through popout. The plot will grey out other points except for those in the neighborhood.

Figure 7 is the scatterplot we will use to show the distribution and relation between funding and expenses. We use both colour and shape to encode the program that a student is in to reinforce this information. An auxiliary line is added to conveniently observe if funding is enough for expenses. We also encourage users to input their own data so that they can locate their position in the graph. Figure 7b shows how we "pop out" the point that denotes the user and grey out unrelated points except for its neighbours. An alternative way is to preserve the colour for students in the same program and grey out the others.

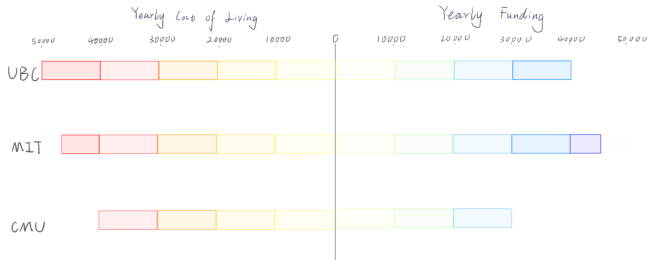

Figure 8: Diverging bar chart showing the funding and expenses of different universities.

Figure 8 shows the diverging bar chart we use to compare UBC with other universities in terms of funding. Luminance is used to encode amounts of money, where warm colours denote the cost of living and cold colours denote funding. We will choose colormaps more carefully in our real design process. In addition, note that we mimic the stacked bar chart with binned amounts of money. We hope this will make it more intuitive for readers to understand and make comparisons. The plot will also support manual reordering and filtering to customize the view.

Finally, we are considering adding an interactive budget tracker in our explainer to make it more engaging. It is based on a stacked bar chart where we use colour to encode the category of expenses. Users will first input their budget (funding), and we denote it with an auxiliary line showing the maximum volume of the bar (Figure 9a). Then we invite the users to answer some questions to decide on the average expenses in each category (Figure 9b). Each answer will add a block in the bar chart, whose size is determined by the number of expenses. After going through all the questions, we can conclude if users can afford to study CS at UBC or not (Figure 9c and 9d). Compared to the UBC official budget calculator[8], our datasets are more up-to-date, and our approach is more intuitive.

4.2 Implementation (*Proposed*)

To execute our potential solution, we are looking at leveraging a few tools to bring this project to fruition. The original plan was to do a robust front-end React project paired with D3. However, given that most of our team is not familiar with both tools and our limited time,

[8]UBC budget calculator for graduate students.

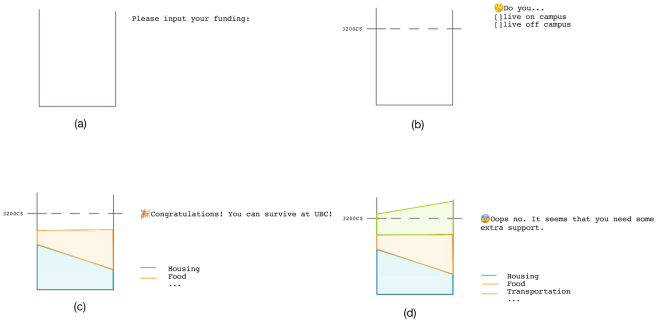

Figure 9: Interactive budget planner based on a bar chart. (a) It first requires users to input their budget (funding). (b) The budget will be shown in the auxiliary line. It then asks users several questions to decide on expenses in different categories. The bar will be filled up based on their answers. (c) An example is when the overall expenses do not exceed the budget. (d) An example is where the overall expenses exceed the budget.

we scaled down to something simpler. Minimizing the amount of new complicated tools to learn, we decided to primarily focus our attention on D3. D3 is known to have a steep learning curve. However, our team has chosen to move ahead with it since it demands us to understand our data on a deeper level to visualize our data appropriately and adequately. D3's complexity comes from having to design the world of one's data; designers are required to have high attention to detail and make clear choices concerning the marks, channels, axis, scales, and labels that they will utilize. This means our team will have to do our due diligence to study the data and make meaningful informed decisions to visualize our data best properly when using D3.

Since we are looking at doing an explainer and interactive visual website, we scoped out what tools are out there for the various levels of expertise when it comes to designing your own. Exploreables [9] is a hub for learning through interactivity. Their tools page pointed us towards two highlighted resources to build our own explainer: Idyll [10] and Observable [11]. Both tools were designed to simplify the process of creating explorable explanations. Observable offers more functionality, like team collaboration, even in its free version—allowing us to take advantage of it with little compromise. Furthermore, Observable is significantly more popular than Idyll, with over 2,000 weekly downloads [12] compared to Idyll's 85 weekly downloads [13]. This gives us confidence that there will be more support and resources for Observable. When it comes to interacting with D3, Observable also has a comprehensive robust set of tutorials that interconnects these two [14], especially since Observable was created by the same developer as D3.

Observable is a browser-hosted notebook that allows us to write markdown, write code, as well as import and embed D3's library directly into its cells. The cells accommodate live interactions and support animations. Using Observable allows us to bootstrap our work and immediately start working with data visualization. One potential concern is that our interactive explainer will look fairly similar in structure to other explainers. We believe this is okay for our minimum viable product; however, with time available, we can extract our work from Observable into a vanilla JavaScript, HTML, and CSS project.

We plan to further optimize our time by taking advantage of existing libraries, plugins, and utilities that are already out there supporting D3. Awesome D3 [15] and The D3.js Graph Gallery [16] are things we

[9]Explorables link
[10]Idyll link
[11]Obserable link
[12]Observable downloads link
[13]Idyll Downloads link
[14]d3-observable tutorials link
[15]Awesome D3 link
[16]D3 graph gallery

will be looking at, and iterating over to refrain from having to start our visualization from scratch. For instance, The D3.js Graph Gallery provides an interactivity section with sampled codes and outputs; these can be refactored to suit our needs.

5 Results (Scenario of use)

Since we identify three types of end users, we will give usage scenarios for each of them.

1. **Current graduate students in UBC CS departments.** Jenny, a 1st year CS graduate student, feels herself living on a tight budget. She wants to figure out what is going on. Using the "you-draw-it" visualization (Figure 6), she may be astonished by the inflation rate in Vancouver, which is way higher than she thought. Then the scatterplot (Figure 7) helps her know the overall financial situation of graduate students in the department. By putting her dot in the graph, she finds herself actually having higher funding than her fellow graduate students, even though she still cannot offset the skyrocketing cost of living in Vancouver. With little hope of increasing funding, she sets out to cut down her daily expenses. She first learns the composition of expenses using Figure 5. After playing with our budget planner, she decides to cook more frequently to cut down on the expenses of food.

2. **Decision makers in UBC.** Laks, our department associate head of grad affairs, receives our Cost of Living Survey [3] and wants to know how serious the issues are. He opens our explainer and sees that the funding of most students cannot offset the basic expenses (Figure 7). What really concerns him is that our department seems to be less competitive in terms of financial support (Figure 8). In order to improve the well-being of current students and avoid losing potential candidates, he begins to seriously consider taking action.

3. **Prospective students.** Allen, an excellent CS undergrad who is applying for graduate schools, recently received offers from UBC and the University of Waterloo (UWaterloo). He hasn't paid off his student loans from his undergrad, so he really does not want to increase his financial burden. To make a decision, he does some googling and finds our explainer. He is first stunned by the rapid increase in the cost of living in Vancouver (Figure 6) and then finds that with the funding at UBC he cannot afford his basic expenses. The diverging bar chart (Figure 8) is also very convenient for him as he can filter out other schools to compare only UBC and UWaterloo. After further consideration, he decides to accept the offer from UWaterloo even though he really likes the research groups at UBC.

References

[1] Data viz exploration: Cost of living in the united state. *Evolytics*. link.
[2] Numbeo: Cost of living. link.
[3] Ubc cs graduate cost of living survey data. link.
[4] Annual inflation index for dcls, cac targets, and density bonus contributions. *City of Vancouver*, 2022. link.
[5] The cost of living in canada. *canadianvisa.org*, 2022. link.
[6] cost of living in canada by province (2022). *Canada Crossroads*, 2022. link.
[7] G. Aisch, A. Cox, and K. Quealy. You draw it: How family income predicts children's college chances. *The New York Times*, 28, 2015.
[8] E. Badger, C. C. Miller, A. Pearce, and K. Quealy. Extensive data shows punishing reach of racism for black boys. *The New York Times*, 19, 2018.
[9] concerned UBC CS Grad Students. Funding at competing unis, 2022. link.
[10] L. Culbert. B.c.'s university, college students scrounging to pay for it all. *Vancouver Sun*, 2022. link.
[11] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi: 10.1177/1473871611416549
[12] F. Hohman, M. Conlen, J. Heer, and D. H. P. Chau. Communicating with interactive articles. *Distill*, 2020. https://distill.pub/2020/communicating-with-interactive-articles. doi: 10.23915/distill.00028
[13] T. Munzner. *Visualization analysis and design*. CRC press, 2014.
[14] Y. Stephanie and C. Tony. A visual introduction to machine learning. *R2D3*. link.
[15] C. U. G. Students. Ubc cs graduate students cost of living survey. *UBC*, 2022. link.
[16] R. Tiwari. Cost of living estimator. *NY Data Science Academy*. link.
[17] R. Tiwari. Data visualization on cost of living. *NY Data Science Academy*. link.
[18] B. Victor. Explorable explanations. *WorryDream*. link.

6 Appendices

See our milestone timeline on the next page.

Table 1: Projected milestones for this project.

Milestone	Hours/person	Deadline	Task Breakdown	Team member
Pitch	3	Sep 28	Cost of Living video pitch	Everyone
Pre-proposal meeting	2	Oct 12	Meet with Tamara	Everyone
Proposal	3	Oct 21	Meeting Time	Everyone
	8 (JH) 6 (DM, TH)		Report Writing	Everyone
Implementation	10	Ongoing	Learn selected tools (D3, observable)	Everyone
	1	Oct 30	Data exploration, simple visualization (e.g. histograms), and data cleaning	Everyone
	5	Nov 6	First draft of written information (DM) reviewed by everyone else (TN, JH)	Everyone
	5	Nov 6	First draft of final visualizations	Everyone
	10	Nov 12th	First draft of implementation of comparator & visualisations	TN, JH
	3	Nov 14	Review of initial comparator combined with explainer written content	Everyone
Update report	3	Nov 15	Submit update report	Everyone
Peer Project Review	2	Nov 15	Prepare demo	Everyone
Refine implementation	20	Dec 5th	Refine implementation & prepare demo	Everyone
	5	Dec 11th	Review everything in preparation for presentation	Everyone
Final Presentation	2	Dec 14th	Prepare & Rehearse	Everyone
Final Report	7	Dec 16h	Submit Final Report	Everyone