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 [20]. Inflation in Vancouver reached a recent high of 8.8% [13] and the average rent for a 1-bedroom apartment is at CA $2,500 [11]. 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 [31]. The results were an anonymized rich dataset that provided 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

Higher cost of living and constant tuition increase has been a burden on university students around the world, preventing them from pursuing higher education [35]. Henderson [22] has found that in general, students in the united kingdom are not financially literate and not always aware of the difference in the cost of living between different geographical regions. This highlights the importance of using visualization to help students understand the cost of living in different geographical regions.

Gleicher et al. [21] 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 [23, 34]. 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" [30] 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. However, Abdullah et al. [17] investigated creating a new visualization for the Malaysian government in 2017. They found that a treemap diagram was useful for as an interactive form of a budget since popup boxes are practical for displaying overlayed information, and the amount of a budget is easily encoded in the size of boxes in the diagram. This supports the findings of the US President Barack Obama administration's attempt to create an easy-to-understand fiscal budget [12]

Additionally, 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 [3]. Other articles available typically list cost of living data in written or tabular format only [14, 15]. Figure 2 provides an example of how financial data can be represented by region [4]. 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 [32, 33], 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 [2].

The UBC Undergraduate programs and admissions page contains a "first-year cost calculator," where prospective undergraduate students can input basic information such as "amount of support from parents" or "savings at the start of the term" [16]. Based on this input the calculator provides a report of the costs associated with living as an undergraduate student at UBC. Notably, this calculator does not say what information it is based on, and it is unclear why this calculator is only for first-year undergraduate students when the majority of UBC is made up of other types of students.

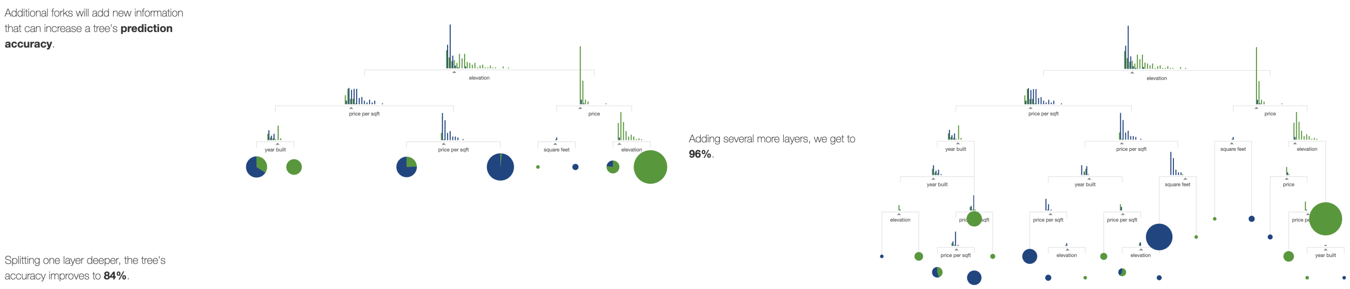
3 Abstraction

3.1 Data Abstraction

Our visualization data consists of a qualtrics survey that was sent out to all current UBC CS graduate students (as of May 2022), which consisted of questions about student income and expenses [31]. An anonymized version of the survey data was released in June 2022 [10]. 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.

- *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.*

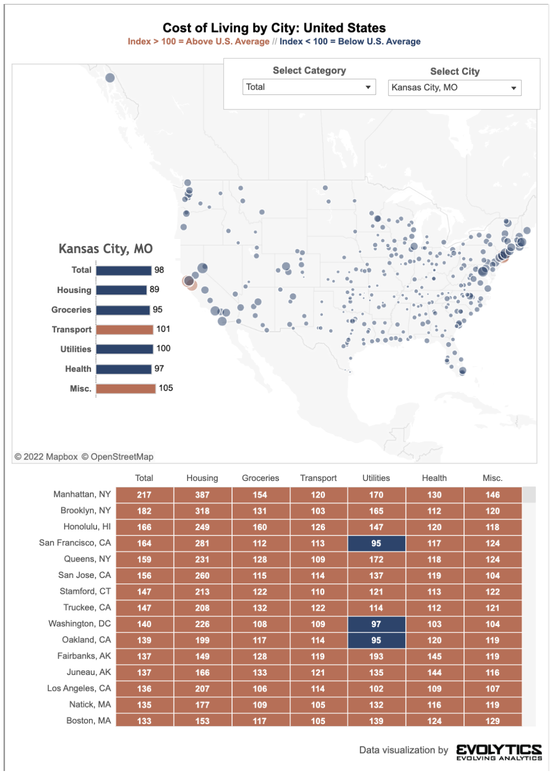
Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx


Additional forks will add new information that can increase a tree's **prediction accuracy**.

Splitting one layer deeper, the tree's accuracy improves to **84%**.

Adding several more layers, we get to **96%**.

Figure 1: Left: example of contextual information with visualization, Right: example of how the visualization changes after completing the action of scrolling. The purpose of this figure is to show that more visualization is added after the interaction of scrolling.



Cost of Living by City: United States

Index > 100 = Above U.S. Average // Index < 100 = Below U.S. Average

Select Category: Total

Select City: Kansas City, MO

Kansas City, MO

Total	98
Housing	89
Groceries	95
Transport	101
Utilities	100
Health	97
Misc.	105

© 2022 Mapbox © OpenStreetMap

	Total	Housing	Groceries	Transport	Utilities	Health	Misc.
Manhattan, NY	217	387	154	120	170	130	146
Brooklyn, NY	182	318	131	103	165	112	120
Honolulu, HI	166	249	160	126	147	120	118
San Francisco, CA	164	281	112	113	95	117	124
Queens, NY	159	231	128	109	172	118	124
San Jose, CA	156	260	115	114	137	119	104
Stamford, CT	147	213	122	110	121	113	122
Truckee, CA	147	208	132	122	114	112	121
Washington, DC	140	226	108	109	97	103	104
Oakland, CA	139	199	117	114	95	120	119
Fairbanks, AK	137	149	128	119	193	145	119
Juneau, AK	137	166	133	121	135	144	116
Los Angeles, CA	136	207	106	114	102	109	107
Natick, MA	135	177	109	105	132	116	119
Boston, MA	133	153	117	105	139	124	129

Data visualization by EVOLYTICS EVOLVING ANALYTICS

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

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 is a large cost-of-living database that uses crowd-sourced data to gather information about housing cost, healthcare quality, crime rates etc [8]. 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.

To visualize the growth of the inflation rate, we also collected the monthly Consumer Price Index (CPI) for the last 6 years (i.e., from Sep-16 to Sep-22) from Statistics Canada [1]. The dataset consists of multiple CPIs of different granularity, such as food CPI and shelter CPI. All attributes are sequential quantitative attributes.

3.2 Task Abstraction

In this section, we identify our end users and abstract our goals using the terms provided in [25]. We believe this abstraction will make our work more generalizable and comparable with similar visualizations in different domains. Note that more detailed task abstraction comes with the "why" analysis of visualizations in Section 4.

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
- informs 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.

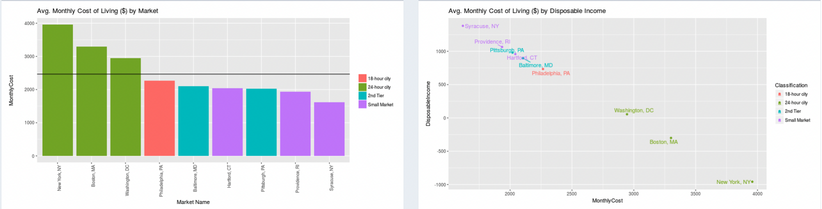
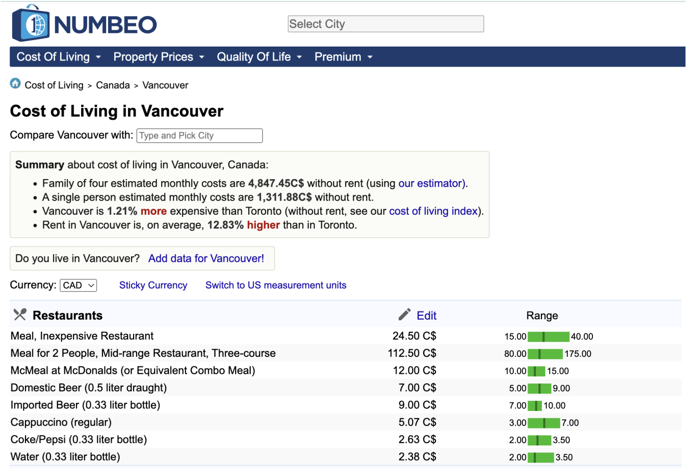


Figure 3: Idioms used in a cost of living comparator interface.



NUMBEO

Select City

Cost Of Living ▾ Property Prices ▾ Quality Of Life ▾ Premium ▾

Cost of Living > Canada > Vancouver

Cost of Living in Vancouver

Compare Vancouver with: Type and Pick City

Summary about cost of living in Vancouver, Canada:

- Family of four estimated monthly costs are **4,847.45C$** without rent (using our estimator).
- A single person estimated monthly costs are **1,311.88C$** without rent.
- Vancouver is **1.21% more** expensive than Toronto (without rent, see our cost of living index).
- Rent in Vancouver is, on average, **12.83% higher** than in Toronto.

Do you live in Vancouver? Add data for Vancouver!

Currency: CAD Sticky Currency Switch to US measurement units

Restaurants	Edit	Range
Meal, Inexpensive Restaurant	24.50 C$	15.00 40.00
Meal for 2 People, Mid-range Restaurant, Three-course	112.50 C$	80.00 175.00
McMeal at McDonalds (or Equivalent Combo Meal)	12.00 C$	10.00 15.00
Domestic Beer (0.5 liter draught)	7.00 C$	5.00 9.00
Imported Beer (0.33 liter bottle)	9.00 C$	7.00 10.00
Cappuccino (regular)	5.07 C$	3.00 7.00
Coke/Pepsi (0.33 liter bottle)	2.63 C$	2.00 3.50
Water (0.33 liter bottle)	2.38 C$	2.00 3.50

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

3.2.2 Actions

Analyze: Our work presents the data of the cost of living (CoL) survey [10] in a more engaging way so that more people will understand the ongoing issue. It can also help users discover the financial problem that current students are facing, which is not that obvious by simply looking at the dataset itself. If the readers are current students, they can also put themselves on the visualization by annotation so that they can compare their situation with other students.

Search: Users can locate themselves in the visualization with ease using the interaction we provided. For example, they can find out if they spend more or less on food than other students by placing themselves on an aligned axis. It is also convenient to browse or explore the visualizations without preset targets to gain some general information about students' financial status.

Query: By default, our work provides summarizations based on categories such as basic expenses and income. The details are offered upon request and are shown by interaction techniques. Some idioms will be designed for comparison tasks, such as contrasting the inflation rate that users assume and the real number.

3.2.3 Targets

Our work will show the growing trend of the cost of living in Vancouver. In addition, users will learn what composes a grad student's income and expenses and what the distribution looks like. They may also learn the features (e.g., students spend most on housing) or correlation (e.g., students receiving less stipend tend to feel unsafe) when exploring the visualizations.

4 Solution

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 [23]. We also considered several visualizations that may be included in our explainer.

4.1 Income vs Expense

To identify the gap between students' income and expenses, we used a scatterplot with an auxiliary line $y = x$ to quickly check if students' stipends can cover their basic expenses (Figure 5). Since many students also receive extra support, we included that information and differentiated it from basic income (i.e. TA, RA, awards) using filled and unfilled circles. Based on the position relative to the auxiliary line, we derived a new attribute, ”financial status,” and encoded it with colour. It can show students' financial status in a more intuitive way. We also plan to include user input so that users can compare his/her own financial situation with other grad students.

4.1.1 What

In this plot, we used a table of 11 quantitative attributes falling into 3 categories:

- basic income: RA, TA, Awards.
- basic expenses: Food, Social, Health, Utilities, Housing, Commute.
- extra support: Other sources, Other support.

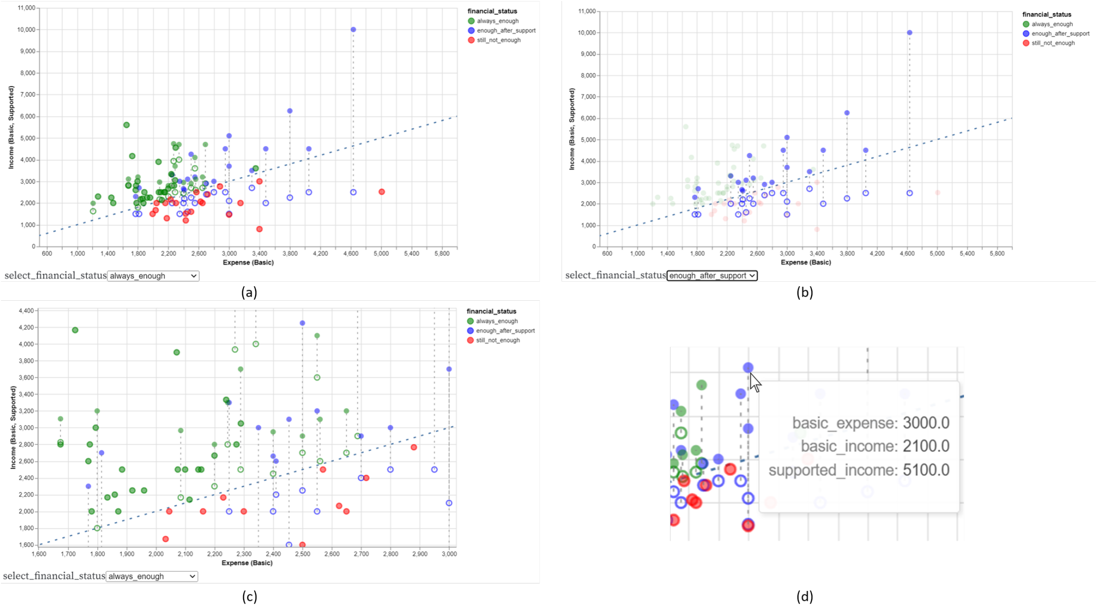
The detailed definition of attributes can be found in the survey dataset [10]. Before visualization, we preprocessed the dataset to tackle missing data and derive new attributes. We first dropped rows that only have null values for all income attributes or expenses attributes. Then we filled the remaining missing value with different strategies: For basic income and extra support, we filled it with 0. This is reasonable as many participants only filled ”RA” and were left empty for ”TA”, and vice versa. However, for basic expenses, we filled it with average instead of zero because one cannot live without spending on these items. We rounded all results to 2 decimal places.

With cleaned data, we derived a new attribute financial status. This new categorical variable divides students into three groups, based on their financial status:

- **Always enough**. The student can afford basic expenses with a basic income.
- **Enough after support**. The student cannot afford basic expenses with a basic income but is able to live after receiving extra financial support.
- **Still not enough**. The student cannot afford basic expenses even after receiving extra financial support.

4.1.2 Why

- **Actions**. At the highest level, this idiom will assist users in discovering the gap between students' income and expenses. By deriving a new attribute, we can help users quickly understand the financial status of students and also focus on some groups of interest by filtering out other groups (Figure 5b). At the middle level, users can explore the plot by zooming and panning the plot using the mouse (Figure 5c). At the lowest level, it can be used to identify the financial status of an individual by hovering on a filled dot, where a tooltip will show the detailed data (Figure 5d). It also supports comparison among students and groups by contrasting their colours and relative positions to the auxiliary line.



(a)

(b)

(c)

(d)

Figure 5: Income vs Expense Visualization. (a) The overview. (b) After selecting one group in the dropdown box, dots in other groups will fade out. (c) Users can zoom in and pan to reduce the number of dots shown in the screen. (d) Tooltip shows the value of attributes when users hover over a filled dot.

- **Targets**. It displays the distribution of income and expenses of CS graduate students at UBC.

4.1.3 How

Marks and Channels

- **Marks**. Point marks encode individual students, while Line marks are used to connect points representing the same individual.

- **Channels**. Position, the most effective channel for ordered attributes [25], is used to encode students' income and expenses, the most important information in the idiom. The color hue encodes students' financial status, where the domain of hue is chosen based on implicit semantics (green: safe, blue: almost safe, red: unsafe). The shape of dots (i.e. unfilled and filled) differentiates a student's income before and after extra support. Note that the length of dashed lines does not encode anything, but it can show the amount of extra financial support that a student receives.

Encode, Manipulate, Facet, and Reduce We used a rectilinear layout to express two quantitative attributes. There are some dense areas where occlusion happens, so we used several approaches to control complexity. In general, it abides by Ben Shneiderman's mantra "Overview First, Zoom and Filter, Details on Demand" [29].

- **interaction**. Users can zoom in by scrolling and pan by dragging the empty space in the plot (Figure 5c). The overlapping points will spread out as users zoom in, and only the points of interest will stay in the plot, reducing users' mental burden. What's more, we added a tooltip for each filled dot (Figure 5d). When users hover over it, an embedded window will show with the accurate number of basic expenses, basic income and extra support of that student. Note that we did not include tooltips for unfilled dots because it may lead to too many pop-out windows when hovering in the dense area.

- **filter out**. The dropdown box below the main plot allows users to select one of the financial statuses. Once selected, points in other groups will fade out to highlight the group of interest. (Figure 5b) If users want to come back to see all groups together, they can do so by double-clicking the plot.

- **derive new data**. The derived attribute "financial status" aggregates original attributes in a semantically meaningful way to help users quickly understand the financial status of current students without diving into detailed categories (e.g., food, commute, etc.)

There are many possible improvements to this plot, for example, limiting the range that users can zoom out and adding an animated transition to show the relation between filled and unfilled dots. We are currently seeking feedback from users to see what features will be useful.

4.2 Compare & Contrast: Your budget vs CS students

To help our audience compare the expenses of an "average" CS student, we used a lollipop chart to display the average expense of a CS student with the interactive self-input that users provide (Figure 6) of their own or expected expense. In this way, our users can see how their current expenses compare (in a visual manner) to that of actual CS students and also use the interactivity to help them budget and "play" around to see what are the maximum and minimum expenses they could afford while living in Vancouver.

4.2.1 What

In this plot, we used a table of 6 quantitative attributes falling under 1 category:

- **basic expenses:** Food, Social, Health, Utilities, Housing, Commute.

The cleaned data described in 4.1.1 was used to populate this visualization.

4.2.2 Why

- **Actions**. There are 2 important aspects to this idiom, (1) display of the average cost of different expense categories for the 94 respondents from [10] (2) user input of their own expense. With regards to (1), users can scan the "lollipops" with blue markers to see what the average cost of certain expenses are. As described in the figure caption, this visualization is still a work in progress. An additional aggregate variable of the "total sum" of expenses can be added to this idiom in the form of a line across the plot showing what the average total sum of expenses is for a CS student. For (2), users can use the sliders at the top of the visualization to enter

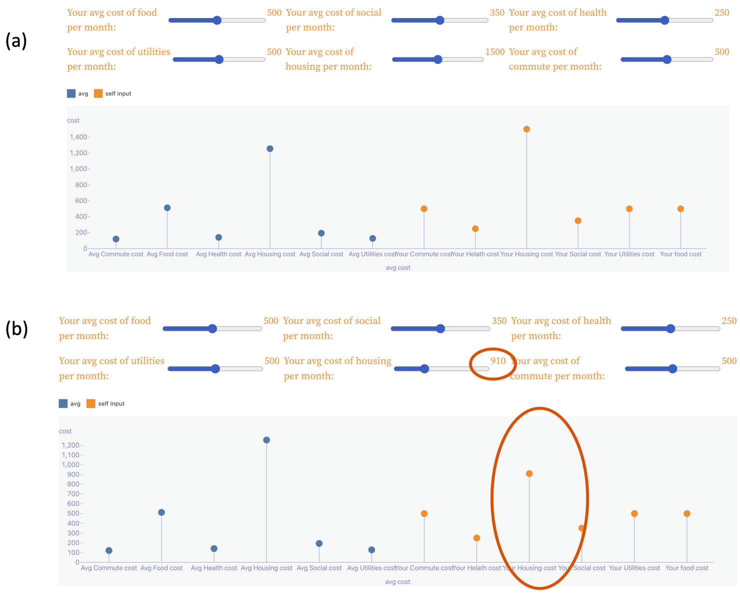


Figure 6: The lollipop chart to display the average expense of a CS student (blue) with the interactive self-input (orange). By changing the slider, the user can see the real-time change of the lollipop bars (orange marks only). The difference between (a) and (b) depicts the if a user was to change their housing cost from $1500 (a) to $900 (b). Note the change in the height of the corresponding lollipop bar. This idiom was built using the observable plot library and is still being updated with additional information and is a work in progress. To utilize the visualization yourself, please visit this observable link: https://observablehq.com/d/b3bb5a56b3d6675e

their own expense data. In real-time, as the slider moves left to right, users can see the corresponding lollipop bar move up and down. Similarly, with the addition of an orange horizontal line, users can see how their total monthly expense sum changes as they input these expenses with the sliders.

- **Targets**. It displays the distribution of expenses of CS graduate students and the user based on their self-input.

4.2.3 How

Marks and Channels

- **Marks**. Point marks encode different costs for different expenses, while Line marks are used to show the continuous nature of the data. Future iterations of this plot will also include the minimum and maximum values as a sort of reduced version of a boxplot. This way, users can compare themselves to the respondents who had the min and max of each expense.

- **Channels**. Position on a common scale is the primary channel represented in this idiom. All expenses are on the same scale. Should the user input data exceed the scale used on the plot, the figure and scale automatically update to represent this, so the position on a common scale is always maintained (interaction - reduce). Colour hue denotes the type of data. Where blue is the average expense for CS students, and orange represents user-inputted data. In future iterations of the plot, the orange lollipop for "health" will be next to the blue lollipop for "health," and the same for the other marks.

Encode, Manipulate, Facet, and Reduce We used a rectilinear layout to express 6 quantitative attributes. As described earlier, there is a partition of the average expense view and user-input view. This is not ideal since the purpose of this idiom is for a user to compare themselves to an average CS graduate student and the current layout of the graph forces them to constantly move their eyes left and right (between blue and orange marks) to make that comparison. By putting single orange marks next to the corresponding blue mark, this complex comparison effort will be reduced.

4.3 You-Draw-It

There are many existing works to visualize inflation [7]. However, most of them are static and complicated, as they are targeted at domain experts instead of laypeople. To give readers a more intuitive understanding of galloping inflation in Vancouver, we designed a ”you-draw-it” visualization (Figure 7). This idiom is first proposed in *New York Times* [18] and later used to show the trend of CO2 emission [23]. It is an interactive line chart where part of the trend is hidden and users are invited to predict how the trend goes. This is the ”place your bet” design pattern for explorable explainers [27] and it is very effective for ”teaching” topics compared to passive reading.

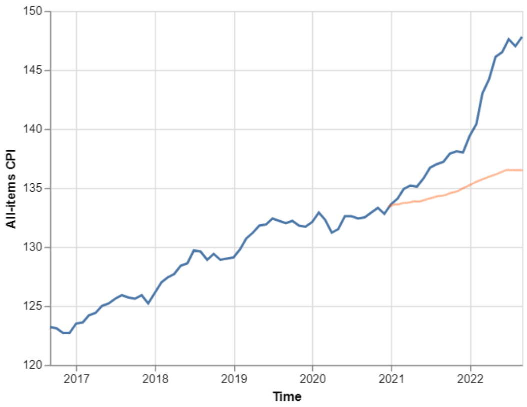


Figure 7: You-draw-it visualization. The blue line stands for the real trend of CPI while the pink line is drawn down by the user.

4.3.1 What

There are different types of inflation [26]. We focused on Consumer Price Inflation as it will be directly reflected in basic expenses. Typically, we will use the time and ”All-items” attribute in the CPI dataset we collected from Statistics Canada. ”All-item” CPI stands for the overarching CPI that consists of eight major goods and services. Both attributes are sequential quantitative attributes.

4.3.2 Why

- **Actions and Targets**. At the highest level, it presents the overall trend of CPI. The interaction feature also allows users to annotate the plot. At the middle level, it adopts an interesting way to help users locate the trend of CPI. At the lowest level, this idiom enables users to compare their hypothesis with the real trend.

4.3.3 How

We used point marks with smoothed connection lines to represent the trend of CPI over time. There will be two layers: one is for the real trend, and another is for the user-drawn trend. Two layers will be superimposed for comparison.

We are still working on this idiom as we came across some technical problems. At first, we tried to implement it with Vega-Lite, but its native APIs for interaction did not support drawing free-form lines. Currently, we switch to D3 where interaction is more customizable.

4.4 Expenses Treemap Diagram

The treemap diagram is a hierarchical visualization [24]. It is rooted in the traditional tree diagram representation where a root parent node is able to have *n* children nodes who may recursively also have children of their own. The treemap diagram is meant to make better use of limited real estate compared to the tree diagram, its traditional predecessor. For instance, trees with over 1000 nodes are able to be visualized on a 13" monitor [24]. We utilize the treemap diagram (Figure 8) to allow users to quickly see at first glance how individual UBC CS Grad students spend their money monthly.

4.4.1 What

In this visualization, we used the clean data, as mentioned in 4.1.1, to populate with some slight modifications to include only the expenses. We have 13 quantitative attributes divided over 3 categories:

- **basic expenses:** Food, Social, Health, Utilities, Housing, Commute.

- **quality expenses:** Travel, Savings.

- **other expenses:** Pet, Childcare, Hobbies, car Loan, Extended Insurance, Other.

The other expense attributes will vary depending on whether or not the participant included any additional other expenses in their survey response.

4.4.2 Why

- **Actions and Targets**. At the highest level, it helps show users, at first glance, how individual UBC CS Grad students may have to spend their money through the area-encoded total expenditure breakdown over the available categories. At the middle level, users can interact with the visualization by clicking in to view how the respective categories are being spent (Figure 8. At the lowest level, users can identify the dollar expenses each individual spends overall monthly and within the respective categories and their attributes. This all works to display the expenditure attributes of an individual UBC CS graduate student in a hierarchical topological manner.

4.4.3 How

For this visualization, we utilized area marks for encoding the respective attribute or category's expenses relative to the total expenses or the parent node if the user had clicked to view a child node through a size channel. A colour hue channel encodes the attributes and categories.

Future iterations will need a way to switch between the different individuals. Due to technical

5 Implementation

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, 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 [5] is a hub for learning through interactivity. Their tools page pointed us towards two highlighted resources to build our own explainer: Idyll [6] and Observable [9]. 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 compared to Idyll's 85 weekly downloads (data retrieved from npmjs package pages). 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, 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 and The D3.js Graph Gallery are things we 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.

As the project goes on, we also decide to include Vega-Lite in our tech stack. Vega-Lite is a high-level grammar that is capable of fast implementation of interactive data visualizations [28]. It supports several common interactions, such as selection, brushing and linking. More importantly, it does not require any knowledge of javascript, and its learning curve is less steep than D3. However, when the visualizations get complicated, the functionality provided by Vega-Lite becomes insufficient. For example, it is hard or even impossible to implement "You-Draw-It" visualization with Vega-Lite (Figure 7). Therefore, we only use it for common idioms such as scatterplots and bar charts and will switch to D3 when requiring sophisticated interaction.

We further decided to explore the full capability of observable and implement a visualization using the observable plot open-source library, which is based on D3 and JavaScript [19]. We anticipate we will move our visualizations onto a website of our own creation, but observable can allow for this migration as any plot made on this software can be embedded in any website.

Thanks to Observable, we are able to utilize various different methods for our visualizations. Observable allows us to lift our visualizations and embed them in our own ways. This could take form in a separate hosted HTML web platform that can hold onto Observable visualization

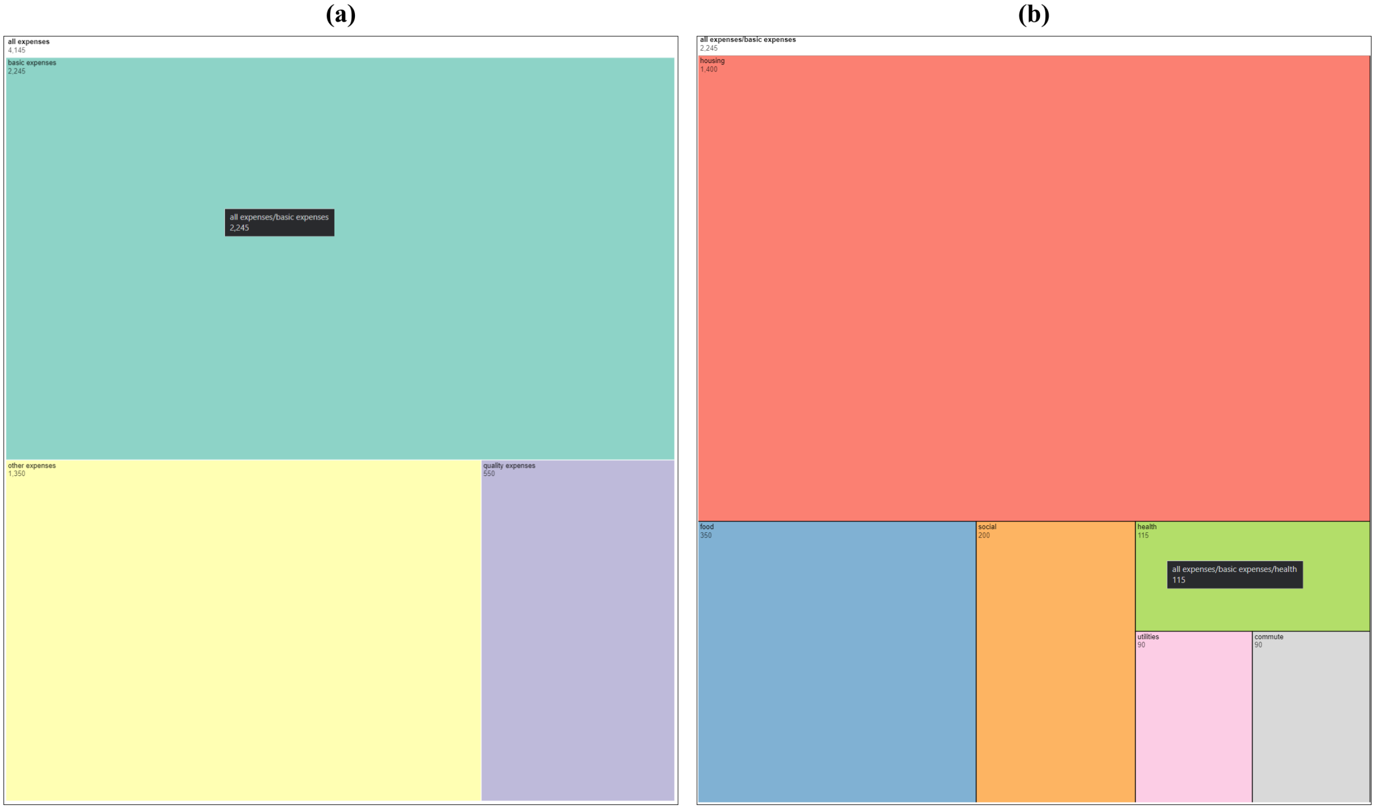


Figure 8: Individual Expenditure. (a) The overview with basic expenses hovered and a tooltip overlay echoing the category's total expenses in dollars along with the category's name. (b) After clicking into an expenditure category (basic expenses)—the user can zoom in and see how that category is being divided for this individual's expenses.

embeds.

6 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 7), she may be astonished by the inflation rate in Vancouver, which is way higher than she thought. Then the scatterplot (Figure 5) helps her know the overall financial situation of graduate students in the department. After exploring the plot with zooming and panning, 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. After playing with the lollipop chart (Figure 6), 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 [10] 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 5). Then he browses the treemap to see how students allocate their funding (Figure 8), and he is shocked by the rent students are paying. 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 7) and then finds that with the funding at UBC, he cannot even afford his basic expenses (Figure 5). After further consideration, he decides to accept the offer from UWaterloo even though he really likes the research groups at UBC.

7 Discussion, Future Work and Conclusions

We found some interesting struggles with the data. Namely, more detail about the attributes would have been helpful. For example, one column in the dataset is labelled ”commute”, however, many people may commute by car, by bus or simply walk (and therefore not spend money on commuting). Thus, it's unclear how this ”commute” attribute is broken down amongst the survey respondents. For now, we must assume that it includes all types of commuters but is not a precise attribute.

See fig 9 for our future milestones. We wish to continue iterating on the idioms presented in this update, as well as incorporate new idioms based on the feedback we receive from course staff and our peer review. We will also consolidate all our visualizations into one website so that we can begin creating the text portion of our proposed explainer article to tie all our visualizations together.

References

[1] Consumer price index, monthly, not seasonally adjusted. https://...endYear=2022referencePeriods=20120901%2C20220901.

[2] Cost of living estimator. https://tiwariraj.shinyapps.io/CostofLivingEstimator/.

[3] Data viz exploration: Cost of living in the united state. *Evolytics*. https://.../cost-living-united-states-visualization/.

[4] Data viz exploration: Cost of living in the united states. https://evolytics.com/blog/cost-living-united-states-visualization/.

[5] Explorable explanations. https://explorabl.es/.

[6] Idyll. https://idyll-lang.org/.

[7] Inflation charts and graphs. https://inflationdata.com/Inflation/images/.

[8] Numbeo: Cost of living. https://www.numbeo.com/cost-of-living/.

[9] Obserable. https://observablehq.com/.

[10] Ubc cs graduate cost of living survey data. https://docs.google.com/spreadsheets/.../editgid=32073719.

[11] Vancouver, bc rent prices. https://www.zumper.com/rent-research/vancouver-bc.

[12] Intereactive budget - office of management and budget. *the White House President Barack Obama*, 2016. https://obamawhitehouse...interactive-budget.

[13] Annual inflation index for dcls, cac targets, and density bonus contributions. *City of Vancouver*, 2022. https://vancouver.ca/.../annual-inflation-index.aspx.

[14] The cost of living in canada. *canadianvisa.org*, 2022. https://canadianvisa.org/.../facts/cost-of-living.

[15] cost of living in canada by province (2022). *Canada Crossroads*, 2022. https://www.canadacrossroads.com/cost-of-living-in-canada-by-province/.

[16] How to plan your first-year costs. *you.ubc.ca*, 2022. https://you.ubc.ca.../cost/.

[17] N. A. S. Abdullah, N. Zulkeply, and Z. Idrus. Malaysian Budget Visualization Using Circle Packing. In M. W. Berry, B. W. Yap, A. Mohamed, and M. Köppen, eds., *Soft Computing in Data Science*, pp. 81–90. Springer Singapore, Singapore, 2019.

[18] G. Aisch, A. Cox, and K. Quealy. You draw it: How family income predicts children's college chances. *The New York Times*, 28, 2015.

[19] M. Bostock. Observable plot. *Observable*, 2022. https://observablehq.com/@observablehq/plot.

[20] L. Culbert. B.c.'s university, college students scrounging to pay for it all. *Vancouver Sun*, 2022. https://.../how-university-college-students-are-scrounging-to-pay-for-food-gas-housing.

[21] 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

[22] H. Henderson. Calculating the cost: place, mobility and price in higher education decision-making for students on small islands around the uk. *Educational Review*, 0(0):1–20, 2021. doi: 10.1080/00131911.2021.1984213

[23] 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

[24] B. Johnson. Treeviz: Treemap visualization of hierarchically structured information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, p. 369–370. Association for Computing Machinery, New York, NY, USA, 1992. doi: 10.1145/142750.142833

[25] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[26] D. Neufeld. Visualizing the three different types of inflation. hrefhttps://advisor.visualcapitalist.com/three-different-types-of-inflation/https://.../three-different-types-of-inflation/.

[27] Nicky. Explorable explanations: 4 more design patterns. https://...4-more-design-patterns/.

[28] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350, 2016.

[29] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pp. 364–371. Elsevier, 2003.

[30] Y. Stephanie and C. Tony. A visual introduction to machine learning. *R2D3*. http://.../visual-intro-to-machine-learning-part-1/.

[31] C. U. G. Students. Ubc cs graduate students cost of living survey. *UBC*, 2022. https://drive.google.com/file/d/.../view?usp=sharings.

[32] R. Tiwari. Cost of living estimator. *NY Data Science Academy*. https://tiwariraj.shinyapps.io/CostofLivingEstimator/.

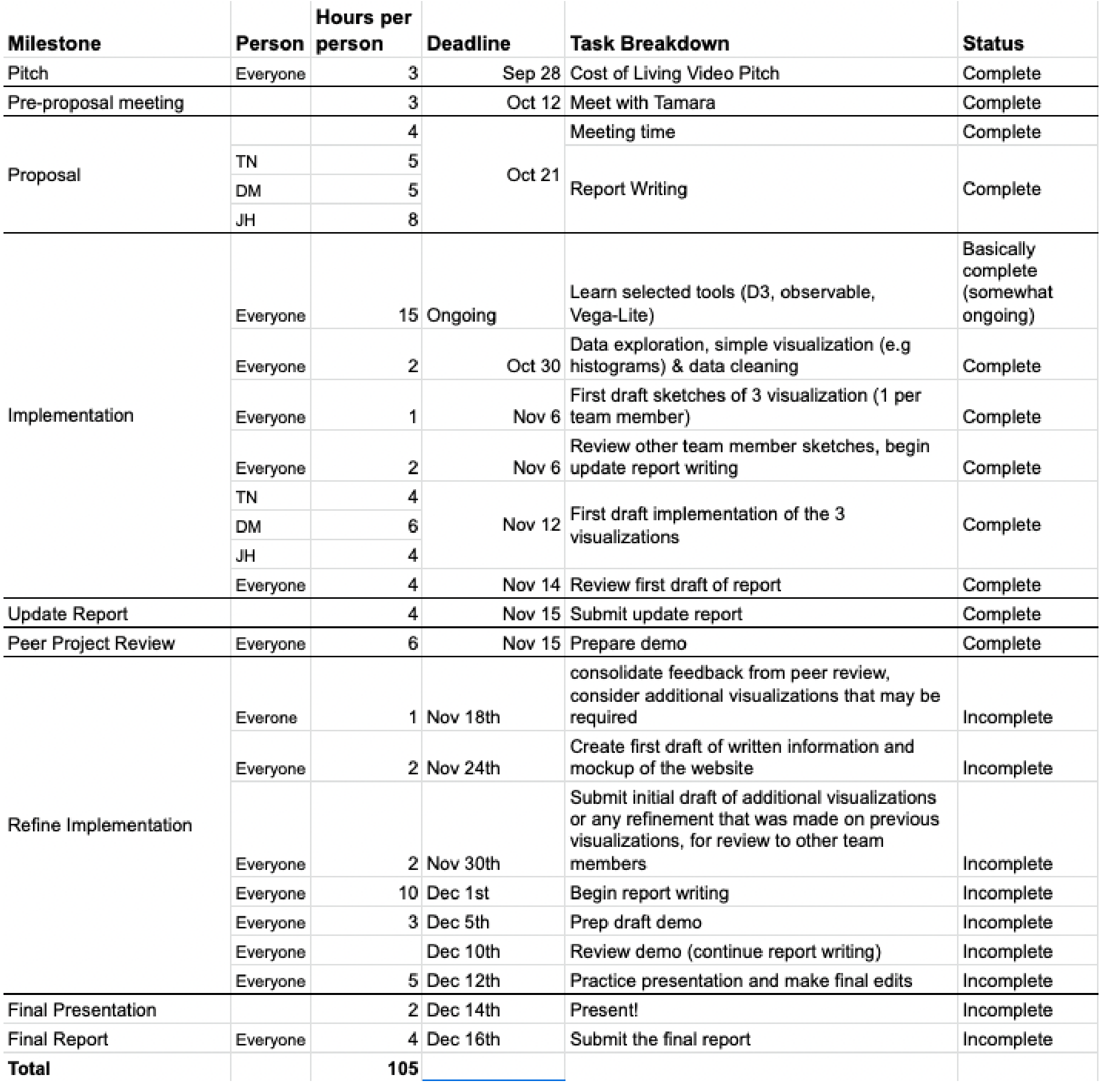
[33] R. Tiwari. Data visualization on cost of living. *NY Data Science Academy*. https://nycdatascience.com/.../data-visualization-on-cost-of-living/.

[34] B. Victor. Explorable explanations. *WorryDream*. http://worrydream.com/ExplorableExplanations/.

[35] S. Wilkins, F. Shams, and J. Huisman. The decision-making and changing behavioural dynamics of potential higher education students: the impacts of increasing tuition fees in england. *Educational Studies*, 39(2):125–141, 2013. doi: 10.1080/03055698.2012.681360

8 Appendices

See our milestone timeline fig 9 (below).



Milestone	Person	Hours per person	Deadline	Task Breakdown	Status
Pitch	Everyone	3	Sep 28	Cost of Living Video Pitch	Complete
Pre-proposal meeting		3	Oct 12	Meet with Tamara	Complete
Proposal		4	Oct 21	Meeting time	Complete
	TN	5		Report Writing	Complete
	DM	5			
	JH	8			
Implementation	Everyone	15	Ongoing	Learn selected tools (D3, observable, Vega-Lite)	Basically complete (somewhat ongoing)
	Everyone	2	Oct 30	Data exploration, simple visualization (e.g histograms) & data cleaning	Complete
	Everyone	1	Nov 6	First draft sketches of 3 visualization (1 per team member)	Complete
	Everyone	2	Nov 6	Review other team member sketches, begin update report writing	Complete
	TN	4	Nov 12	First draft implementation of the 3 visualizations	Complete
	DM	6			
	JH	4			
	Everyone	4	Nov 14	Review first draft of report	Complete
Update Report		4	Nov 15	Submit update report	Complete
Peer Project Review	Everyone	6	Nov 15	Prepare demo	Complete
Refine Implementation	Everone	1	Nov 18th	consolidate feedback from peer review, consider additional visualizations that may be required	Incomplete
	Everyone	2	Nov 24th	Create first draft of written information and mockup of the website	Incomplete
	Everyone	2	Nov 30th	Submit initial draft of additional visualizations or any refinement that was made on previous visualizations, for review to other team members	Incomplete
	Everyone	10	Dec 1st	Begin report writing	Incomplete
	Everyone	3	Dec 5th	Prep draft demo	Incomplete
	Everyone		Dec 10th	Review demo (continue report writing)	Incomplete
	Everyone	5	Dec 12th	Practice presentation and make final edits	Incomplete
Final Presentation		2	Dec 14th	Present!	Incomplete
Final Report	Everyone	4	Dec 16th	Submit the final report	Incomplete
Total		**105**			

Figure 9: Projected (and completed milestones) for this project. An editable version of this table can be found on google sheets at this link