Hood Hunter: A House Hunter's Guide to Narrowing Neighbourhoods

Abiramy Kuganesan*
akuganes@cs.ubc.ca
University of British Columbia
Vancouver, Canada

Ivan Song*
ivansong3@alumni.ubc.ca
University of British Columbia
Vancouver, Canada

LuFei Liu*
liulufei@student.ubc.ca
University of British Columbia
Vancouver, Canada

Abstract

This proposal provides an overview of an explorer visualization tool to support prospective home buyers in narrowing down neighbourhoods for a potential home purchase. A survey was conducted to understand the core criteria that BC residents evaluate when picking the perfect neighbourhood for their residential housing purchase. Grounded by these considerations, and prospective home buyers' current task flow, three solution mock-ups and scenarios are provided. Upon synthesis of the pros and cons of each proposed solution, and analysis of the data and task abstractions presented in this proposal, a mockup of the final solution is proposed. This mockup is currently being implemented using D3 Observable. It will be presented as web application for users to explore.

Keywords: house affordability, neighbourhood search

1 Introduction

In recent years, several urban centres across Canada have faced challenges with affordable housing. These trends have been exacerbated over the course of the pandemic with home buyers looking to move out of densely populated regions to nearby sub-urban neighbourhoods to support newly emerging remote work cultures.

Several factors have contributed to the unafforadable conditions of the Canadian residential real-estate market. Notably, the presence of foreign buyers in the Toronto and Vancouver real estate markets has contributed significantly to the current housing crisis incurred by local residents [1].

Although the house search process is unique to every buyer, certain recurring classes of requirements have been shown to be persistent across buyers and varying demographics. Olanrewaju and Wong conducted a study to determine key criteria that home buyers considered when purchasing a home [2]. Through an exploratory factor analysis, they were able to identify and rank clusters of requirements according to their impact on buying decision. These clusters include price configuration, accessibility, transportation, and sustainability.

Prospective home buyers spend a large amount of time and resources researching neighbourhoods for proximity to city centres, low crime rates, quality schools, and more.

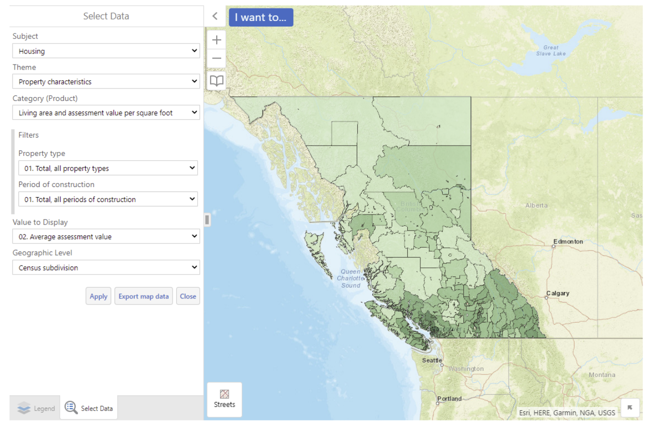

Figure 1. The Canadian Statistical Geo-spatial Explorer [3].

While each home buyer's criteria and ranking of criterion importance are different, identifying neighbourhoods that meet their criteria, while also falling within their budget, can be a cumbersome task.

The primary objective of this project is to support prospective home buyers in narrowing down potential neighbourhoods for their residential property purchase. To narrow the scope of this project, we focus on the residential real-estate market in British Columbia. Seeing that the Toronto and Vancouver housing crises are coupled [1], this project aims to initially address the region where the housing crisis has had the most impact. As such, we propose an explorer tool to support home buyers in identifying neighbourhoods of interest that meet their specific search criteria in accordance to personal importance. An ideal task following an interaction with the proposed tool would be for the user to begin searching for homes in the neighbourhood of interest.

2 Related Works

Several housing affordability visualization tools already exist for house buyers to explore. Many are provided by government agencies such as *BC Assessment Maps* [4] and the *Canadian Statistical Geo-spatial Explorer* [3] (Figure 1). Both of these tools show regional residential housing prices. There are also non-government visualizations, such as a visualizer in a UBC blog [5] (shown in Figure 2), which do not allow for interaction. Other visualizations for neighbourhood characteristics such as safety and education are also

*All authors contributed equally to this research.

provided separately by local governments such as *VPD GeoDash* [6]. These tools locate recent crimes and the Foundation Skills Assessment scores (FSA) for each school district in the province of British Columbia [7]. Another attribute of interest relates to transportation options, such as those included in *WalkScore* [8]. However, each of these tools focus on a specific attribute and do not consider overall neighbourhood livability. Other existing tools weigh a combination of these attributes to help potential home buyers research neighbourhoods in Metro Vancouver, including *Find Your Neighbourhood* by Vancouver Magazine [9] (shown in Figure 3) and *Find a Hood* [10]. Both of these tools require a user survey and offer less room for exploration.

In academic literature, Rinner describes a pilot study of the usefulness of geographical visualization in urban quality of life evaluations based in Toronto [11]. Liu et al. [12] and Balsas [13] show example visualizations and considerations for regional livability. Shabanzadeh et al. visualized livability in Tehran's metropolitan districts using several choropleth maps [14]. Other works such as a Malaysian study on neighbourhood evaluation [2] clearly highlight requirements that impact home-buying decisions. These considerations contribute to our proposed visualization design.

Geographical visualizations are outlined in works such as those by Cartwright et. al [15] and MacEachren et. al [16]. We rely on past works such as map visualizations of spatial and spatiotemporal data [17], cartograms [18], and Hotmaps [19] to explore trade-offs between various types of maps, task abstraction taxonomy, and color, respectively. Work by van Kreveld et al. helps us understand implications of diagram placement on maps [20] and we leverage ideas proposed in Lineup [21], necklace maps [22], and data stripes [23] for our potential solutions. In our final solution, we mimic the use of stacked bars for ranking neighbourhoods from Lineup [21] and distribution based filtering from Crossfilters [24] to filter our data. We also observe insights from Latif et al. which explore the relationship between text and geographical visualizations in data-driven stories and their influence on the reader's understanding [25].

3 Data and Task Abstractions

3.1 Identifying Criteria of Importance

The grounding premise of this project relies on potential home buyers finding dwelling type, budget range, community safety, quality of education, proximity to amenities, and commuter friendliness to be the core criteria when searching for a neighbourhood. Although this is validated and derived from a Malaysian study [2], we verify that this trend applies to a North American context through the distribution of the survey presented in Figures 12, 13, and 14. This survey was distributed through Reddit in the following threads:

- r/britishcolumbia
- r/vancouver

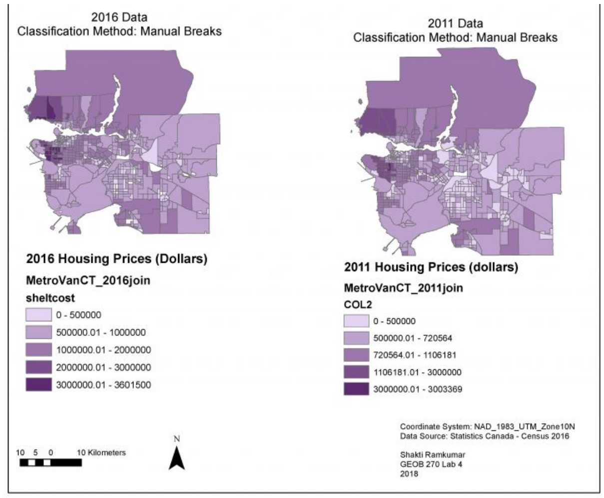

Figure 2. A static visualization of housing affordability in Metro Vancouver by Ramkumar [5].

- r/CanadaHousing
- r/RealEstate
- r/SampleSize

A total of 56 respondents between the ages of 18-54 contributed to the following findings. 25 respondents were British Columbia residents and 31 residents were North Americans that reside outside of British Columbia. Responses from British Columbia residents directly inform our project while the responses from the secondary participants, belonging to the broader North American demographic, inform the scalability of our assumptions. Figure 4 shows respondents' rankings for criteria in accordance to their importance. Cost of purchase price takes precedence over any other criterion. Neighbourhood safety was second most important criterion, followed by access to parks and recreation. Additionally, survey respondents indicated that proximity to employment, grocery stores, and schools were also important.

3.2 Dataset Information

We obtain our data from multiple sources in order to consider different attributes that relate to our analysis. These datasets include the Canadian Census Criminal Code Violations from 2020 [3], BC Foundational Skills Assessment data (FSA) from the 2020 to 2021 school year [7], average home purchase prices according to home type from 2010 to 2020 [26], and proximity measure data provided by Statistics Canada [27]. These datasets are intended to support the following dimensions of neighbourhood search:

3.2.1 House Preference and Affordability. The Canadian Mortgage and Housing Corporation offers a dataset of averages for absorbed homeowner and condominium units in Canada, by dwelling type and municipality for urban centres with more than 50,000 residents [28]. These dwelling

Dataset Source	Attribute Name	Attribute Description	Attribute Type
Census	Census subdivision	Geographical region	Categorical
	All Criminal Code Violations Excluding Traffic	Safety of the region, normalized per 100,000 capita	Ordinal (Quantitative)
Quality of School Scores	District Name	Name of school district	Categorical
	Grade	Level of study in educational year of schooling	Categorical
	FSA Skill Code	Type of assessment issued	Categorical
	Score	Average assessment score for assessment issued	Ordinal (Quantitative)
Canadian Mortgage and Housing Absorbed Homeowner and Condominium Units	Census subdivision	Geographical region	Categorical
	Dwelling type	Type of home	Categorical
	Year	Year the sales data was aggregated for	Ordinal (Categorical)
	Average home price	Average home price per dwelling type within a specific region for specified year.	Ordinal (Quantitative)
Proximity Measures Database	Census subdivision	Geographical region	Categorical
	Longitude and latitude	Coordinates of a location	Ordinal (Categorical)
	Proximity to employment	Closeness to any dissemination block with a source of employment.	Ordinal (Quantitative)
	Proximity to grocery stores	Closeness to any dissemination block with a grocery store.	Ordinal (Quantitative)
	Proximity to health care	Closeness to any dissemination block with a health care facility.	Ordinal (Quantitative)
	Proximity to primary education	Closeness to any dissemination block with a primary school.	Ordinal (Quantitative)
	Proximity to secondary education	Closeness to any dissemination block with a secondary school.	Ordinal (Quantitative)
	Proximity to public transit	Closeness to any source of public transportation.	Ordinal (Quantitative)
	Proximity to neighbourhood parks	Closeness to any dissemination block with a neighborhood park.	Ordinal (Quantitative)

Table 1. Summary of data attributes.

types include single detached homes, semi-detached homes, row homes or townhouses, and apartments. We use annual data from 2010 to 2020 to provide the most current averages with trend information from the recent decade.

3.2.2 Safety. Statistics Canada provides a dataset on incident based crime statistics across Canada [3]. The most recent report was generated from all incidents in 2020. For relevance, we select the crime rates normalized by 100,000 population for all criminal code violations excluding traffic violations to use in our tool. The geographic level of analysis chosen was police service and detachment for the richest data available on this subject.

3.2.3 Quality of Education. The BC Foundational Skills Assessment from 2020 to 2021 school year provides an overview of literacy (reading and writing) and numeracy in grades 4

Attribute Name	Range (Quantitative)	Ordering Direction	Number of Items
All Criminal Code Violations Excluding Traffic	Min: 2,159.72, Max: 24,793.39	Diverging	735
Score	Min: 1, Max: 612.3617	Sequential	360
Average home price	Min: 401,665 Max: 5,533,311	Sequential	520
Proximity to employment	Min: 0 Max: 0.6633	Sequential	35,345
Proximity to grocery stores	Min: 0.0001 Max: 0.833	Sequential	35,345
Proximity to health care	Min: 0 Max: 0.5934	Sequential	35,345
Proximity to primary education	Min: 0.0004 Max: 0.6614	Sequential	35,345
Proximity to secondary education	Min: 0.0005 Max: 0.8396	Sequential	35,345
Proximity to public transit	Min: 0 Max: 0.53	Sequential	35,345
Proximity to neighbourhood parks	Min: 0.0001 Max: 0.462	Sequential	35,345

Table 2. Quantitative attribute details.

Attribute Name	Number of Categories (Categorical)	Number of Items
District Name	60 categories	360
Grade	2 categories (4, 7)	360
FSA Skill Code	3 categories (Writing, Reading, Numeracy)	360
Census subdivision	735	735
Dwelling type	4	224
Year	10	224
Longitude and latitude	35,345	35,345

Table 3. Categorical attribute details.

and 7 students by school district. This dataset contains 60 unique districts with 360 items [7].

3.2.4 Proximity to Amenities.

According to our survey described in Section 3.1, home buyers often prioritize proximity to amenities such as neighbourhood parks or transportation. Statistics Canada offers a dataset of proximity measurements [27] for several coordinates in each census subdivision. Proximity measures are based on a gravity model for the distance between a reference coordinate within a census subdivision and a service. Multiple reference coordinates are provided per census subdivision. There are 10 proximity measures and each is included as a normalized index value.

3.3 Data Abstraction

All of our datasets are organized by census subdivisions (such as cities, villages, towns, etc.) with the exception of the FSA dataset, which is organized by school districts. We convert the FSA school districts into equivalent census subdivisions using administrative boundaries data from the government of British Columbia [29] and combine our dataset into a single table as the source of our visualization. Since census subdivisions also include regions, which encompasses other subdivisions such as cities, we extract only non-overlapping subdivisions from the dataset. These datasets also have a temporal element, each associated to a year between 2016 to 2021. We select the most recent data available to us, some stemming from the 2016 census and others acquired more recently. Therefore, we believe each attribute is the best representation of the present state and do not adjust for time differences between attributes. For house preference and affordability, we optionally consider a separate representation to communicate any time series data available to us.

Table 1, Table 2, and Table 3 outline the detailed data abstraction of these datasets, evaluated based on the raw data. These tables exclude columns that we do not plan to include in our visualization. Our final dataset combines these attributes into a single table, organized by census subdivisions. Since average home price data is only available for regions with greater than 50,000 population, we reduce our final dataset to match this subset of census subdivisions. Our final dataset has 52 items, reflecting the 52 census subdivisions we consider in our visualization. These locations are more relevant since they are more popular to live in and the reduced set still covers the majority of the province. The data abstraction for this updated dataset is included as Table 4.

Some abstractions are decided based on our currently planned solution. We choose *diverging* for the criminal code violation attribute considering that it is important for users to know if cities are a lot safer than the mean or a lot more dangerous than the mean when the safeness of each city is relative to one another. For FSA score, a rank will be assigned

Attribute Name	Attribute Description	Attribute Type	Ordering Direction	Range
Census subdivision	Geographical region	Categorical	–	–
Dwelling type	Type of home	Categorical	–	–
Normalized safety score	Safety of the region, normalized per 100,000 capita	Ordinal (Quantitative)	Diverging	Min: 0, Max: 1
Normalized education score	Average FSA score, normalized to highest	Ordinal (Quantitative)	Sequential	Min: 0, Max: 1
Normalized housing affordability	Most recent average home price, normalized to highest	Ordinal (Quantitative)	Sequential	Min: 0, Max: 1
Average home price (2020)	Average home prices for 2020	Ordinal (Quantitative)	Sequential	Min: 401,665 Max: 5,533,311
Average home price (2019)	Average home prices for 2019	Ordinal (Quantitative)	Sequential	Min: 378,231 Max: 5,881,634
Average home price (2018)	Average home prices for 2018	Ordinal (Quantitative)	Sequential	Min: 531,250 Max: 6,531,910
Average home price (2017)	Average home prices for 2017	Ordinal (Quantitative)	Sequential	Min: 500,240 Max: 5,651,571
Average home price (2016)	Average home prices for 2016	Ordinal (Quantitative)	Sequential	Min: 465,855 Max: 5,380,366
Average home price (2015)	Average home prices for 2015	Ordinal (Quantitative)	Sequential	Min: 364,746 Max: 3,810,023
Average home price (2014)	Average home prices for 2014	Ordinal (Quantitative)	Sequential	Min: 367,990 Max: 3,947,987
Average home price (2013)	Average home prices for 2013	Ordinal (Quantitative)	Sequential	Min: 385,730 Max: 3,842,258
Average home price (2012)	Average home prices for 2012	Ordinal (Quantitative)	Sequential	Min: 366,242 Max: 3,518,374
Average home price (2011)	Average home prices for 2011	Ordinal (Quantitative)	Sequential	Min: 375,961 Max: 3,263,195
Average home price (2010)	Average home prices for 2010	Ordinal (Quantitative)	Sequential	Min: 382,568 Max: 3,647,379
Proximity to employment	Closeness to source of employment, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to grocery stores	Closeness to grocery store, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to health care	Closeness to health care facility, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to primary education	Closeness to primary school, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to secondary education	Closeness to secondary score, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to public transit	Closeness to source of public transportation, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1
Proximity to neighbourhood parks	Closeness to neighbourhood parks, normalized to closest	Ordinal (Quantitative)	Sequential	Min: 0 Max: 1

Table 4. Data abstraction of updated combined dataset.

Figure 3. Find Your Neighbourhood: a survey based interactive visualization tool by Vancouver Magazine [9].

to each census subdivision depending on the average school performance.

3.4 Task Abstraction

3.4.1 Who. The intended users of our visualization tool consist of potential home buyers who are trying to pinpoint which neighbourhood to purchase residential housing in. The visualization tool should help users answer questions of where to look for housing prior to house hunting. Secondary users may also include investors, realtors, and renters. For these users, certain attributes may be less relevant than others. For instance, proximity scores may be more relevant to renters than mean purchase price. This specific iteration of the Hood Hunter tool will target home buyers interested in residential housing within British Columbia. One additional question we hope to help users answer is when to look for housing in a particular neighbourhood. We will optionally include this in our visualization based on time and resource constraints.

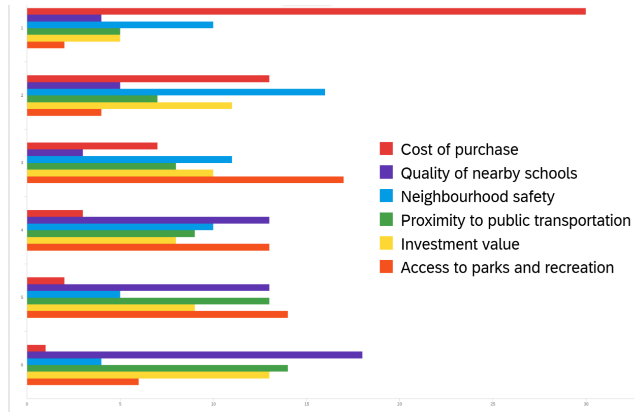

Figure 4. Survey respondents ranked the importance of the criteria cost of purchase, quality of nearby schools, neighbourhood safety, proximity to public transportation, investment value, and access to parks. The vertical axis indicates how important each criterion is where 1 is more important and 5 is less important. The horizontal axis represents the number of participants that are the same opinion.

3.4.2 Actions. From our survey responses, 3 primary actions were identified as common across many prospective users' current neighbourhood search task flows.

1. Filter neighbourhoods that meet certain criteria such as budget range for a specific type of home.
2. Compare a narrow list of neighbourhoods according to user-specific dimensions of interest.
3. Dive deeper into a specific neighbourhood to understand specific assets and limitations.

At the *search* abstraction level, we expect the users to perform either *lookup* or *browse*. In the most common case, users will have a budget and a few criteria in mind when searching for places to live in. This would then fall into the category of browsing. In other cases, a user might already have a neighbourhood in mind and want to look up certain attributes regarding that neighbourhood. *Lookup* could also follow after *browse* as users start to gain a better understanding of their neighbourhoods of interest.

For *query*, we expect users to *identify*, *compare*, and *summarize*. Users can use our visualization to identify attributes regarding a particular neighbourhood after *lookup* or identify a particular neighbourhood with certain attributes after browsing. One example of this is to identify the neighbourhood with the cheapest housing. Another use of the visualization is help users make comparisons between multiple neighbourhoods to find the most suitable one. Lastly, the ability to summarize is not only useful for users to have a high-level overview of the attributes for all neighbourhoods in BC, it also provides them with opportunities to identify trends in certain attributes for particular neighbourhoods.

In terms of the *analyze* abstraction level, the main use for our visualization tool is to consume existing data regarding neighbourhoods in BC. In particular, users will use our visualization to discover new insights regarding which neighbourhoods are more suitable for their needs. Another use of the visualization tool is to produce new information about neighbourhoods. In some cases, we will be transforming raw quantitative data to ordered ranged data to improve the legibility of the information presented to them.

3.4.3 Targets. The main high-level targets are *trends* and *distributions* which provide users with rich insights about the various attributes across neighbourhoods for filtering, comparison, and detailed inspection. Seeing the distribution of the data according to a specific criteria when filtering can be more informative than specifying values without much context. An example of how *trends* might be useful for detailed inspections is for a user who cares about the investment value of housing. They can compare how fast the housing price within a particular neighbourhood is growing compared to another. For the specific attribute of housing price, *distribution* and *extremes* are also targets of interest. The distribution of the price for houses sold in a neighbourhood will give a more accurate representation of pricing than solely average price. Looking for extreme minima and maxima in pricing data over time can also enable users to identify when the most appropriate time is to purchase houses in a particular neighbourhood. A secondary target of interest might be *outliers* when users are trying to find the most suitable neighbourhood for specific criteria. One example of this is that of a user who prioritizes safety; they might want to find the neighbourhood with the lowest crime rate in BC.

4 Solution

4.1 Proposed Solutions

This section outlines three potential scenarios and the possible solutions associated with them.

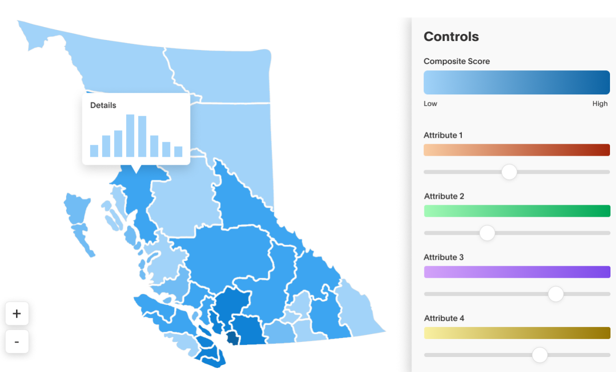

Figure 5. Solution A mockup

4.1.1 Solution A & Scenario. Figure 5 shows a possible visualization design. In this example, a user may be interested in finding a potential neighbourhood in British Columbia that is most suitable for them to live in. They are in an exploration phase and are open to the possibility of living anywhere in the province. The map provides the user with a generalized view. The main map is divided into Census subdivisions and the luminance channel encodes a composite score based on the user's priorities. These priorities are defined by the position of the slider inputs. The hue channel encodes different regional characteristics and the luminance of each of these indicators encode their value. To explore different attributes, the user can also select a specific attribute to colour the map in replacement of the composite score.

Once the user has selected a suitable region, they may zoom into the map further to view smaller Census subdivisions on the map. When they have narrowed down their search to specific subdivisions, the user can select the subdivision and view further details. For instance, a histogram encoding the distribution of housing prices in the subdivision with corresponding lengths.

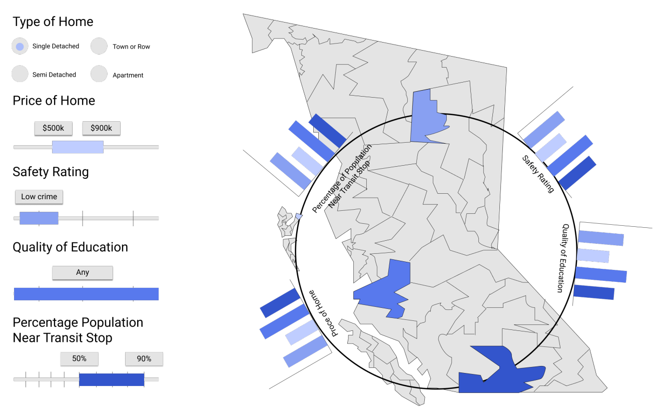

Figure 6. Solution B mockup

4.1.2 Solution B & Scenario. Loosely inspired by the concept of Necklace Maps [22], Figure 6 presents an interactive explorer tool to support neighbourhood search.

The mark of type containment encodes census subdivisions. The fit of neighbourhood is a cumulative score of each attribute. The visual channels saturation and luminance redundantly encode a unique neighbourhood and the fit of the neighbourhood. The top four results are shown to assist the user in the task of filtering. Each neighbourhood's attribute value is represented by a mark of type line. The hue and luminance of the mark on the histogram denote the neighbourhood on the map. Additional labelling may be used to display neighbourhood names.

In order to manipulate this tool, the user is required to input their desired criteria to filter the neighbourhoods for best fit. These inputs will be user defined ranges of tolerance for each attribute. The user may drag the line mark and expand it to indicate a wider tolerance and shrink the length to indicate a narrower tolerance. The user is asked to input type of home preference through a multiple choice selector. As the user toggles these controls, a cumulative score will be computed according to the ratings inputted by the user. The top four neighbourhoods will be assigned saturation and luminance values indicating best fit. Darker and more saturated regions will indicate better fit than lighter and less saturated regions. The histograms along the necklace may be used to easily understand and compare each attribute for the top four neighbourhood recommendations.

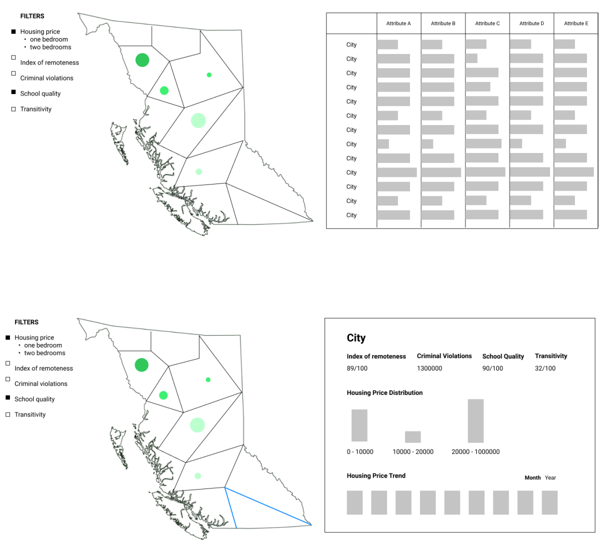

Figure 7. Solution C mockup

4.1.3 Solution C & Scenario. Figure 7 showcases another possible visualization design which allows users to visualize particular attributes of interest to support the task of neighbourhood comparison. On the left hand side, a user can filter which attribute to visualize on the map. Each neighbourhood is represented with a point mark where size and saturation channels are used to encode two attributes. A user can choose up to two attributes to visualize with the default attribute being housing price encoded by the size channel. As this attribute's assumed significance is high, it would be justified to encode it using a channel with relatively higher effectiveness. Users can then choose which other attribute they want to encode as well on the map. On the right, there is a table with embedded bars showcasing attributes for every neighbourhood. Users can use this table to easily rank and compare neighbourhoods by attributes of interest.

The lower mock-up on Figure 7 shows how the display changes once a neighbourhood is selected. The selected area will be highlighted as the table with embedded bars is replaced by a summary of key information regarding the selected neighbourhood. This will support the task of lookup, offering an overview of a particular neighbourhood. The overview will showcase the values for attributes such as index of remoteness, criminal violations, school quality, and transitivity. It will also showcase the neighbourhood's housing price distribution and housing price trend overtime. This can be particular useful if users want to learn more about a neighbourhood they are further interested in.

4.1.4 Final Solution. Figure 8 presents our final iteration of the visualization tool which combines aspects from all previous solutions. As previously mentioned, from the survey, we identified respondents' common neighbourhood search task flows. Similarly, we designed our final solution to support these core tasks of filtering, comparing, and looking up details.

On the left pane of the mock-up is the filter view which allows users to filter neighbourhoods in BC by attribute values. Each attribute is paired with a a histogram which simultaneously shows users the attribute's distribution and allows users to select their range of interest. The filtered neighbourhoods will then be highlighted on the map on the right. The histogram is meant to give users more context on what is considered a good or bad attribute value. This is intended to help clarify more ambiguous attributes such as crime score. A composite score generated by this filtering task will highlight appropriate census subdivisions on the map view.

On the right pane of the mock-up is the ranking view which facilitates comparison between filtered neighbourhoods. Taking inspiration from Line-up, the attribute values are represented by stacked horizontal bars with color encoding each attribute value. The filtered neighbourhoods will be ranked based on the selected attributes. Users can also choose to change the weight of each attributes which will effect the multi-attribute ranking of the neighbourhoods. Lastly, users can star the neighbourhoods of interest which will bring them to the top of the bar graph for easy comparison.

At the bottom of the mock-up is the detail view which displays detailed information of a single selected neighbourhood triggered by a mouse click. The detail view will first show one short text description of the neighbourhood. Then, the average housing price in the last 1o years in the neighbourhood is encoded by a line graph allowing users to estimate the neighbourhood's investment value. The other attributes are then shown with the value on top and the distribution encoded by a histogram below. The bin that the attribute value falls under will be highlighted.

All three views are juxtaposed on the screen to better facilitate interactions between them without requiring users to recall from memory. For example, after a user is interested in a neighbourhood after checking its details, they can directly

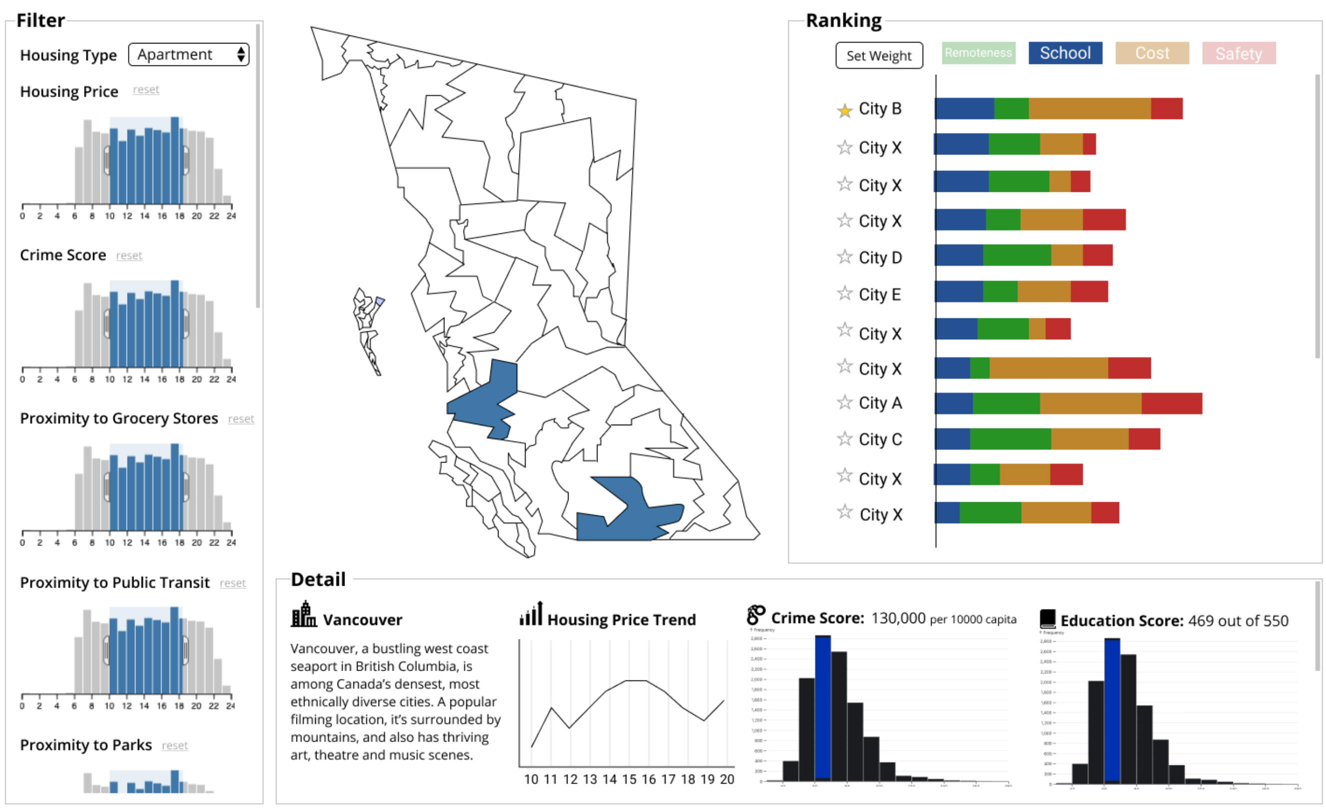

Figure 8. Final solution mock-up

star the neighbourhood in ranking view and make comparisons. Clicking on the census subdivision on the map will bring up the detail pane for that census subdivision. Clicking the census subdivision title on the ranking view has the same effect.

4.2 Tools

In order to implement our interactive explorer visualization tool, we have been leveraging Observable D3 due to ease of prototyping and integration with web. We aim to host our visualization tool on a website which can be easily accessed by prospective users. To perform any background processing or algorithmic computation, we may leverage Python.

5 Implementation

5.1 Map View

QGIS Geographic Information System [30] was used to convert Census Subdivision shapefiles to GeoJSON format and to transform the coordinate system of these shapefiles to latitude and longitude values. The Canadian Census Subdivisions were filtered for entries containing British Columbia as the province name. The Leaflet.js library [31] was then used to create the map component of the proposed visualization tool. The current progress with this implementation is pictured in Figure 9.

5.2 Ranking View

Figure 10 demonstrates our current progress in creating the ranking view described in Section 4.1.4 (right). We generate a stacked bar chart in an Observable notebook, which presents a composite score and normalized components from each attribute from the dataset. The bar is currently responsive to a drop-down selection of the attribute to sort by and a choice to rank from best to worst or worst to best.

5.3 Detail View

Figure 11 demonstrates our current progress in creating the detail view described in Section 4.1.4 (bottom). Based on the city selected, represented by the string typed in the text box in the prototype, detail information such as housing price trend and proximity to employment score of the selected neighbourhood are shown. In the histogram, the bin where the neighbourhood falls under is also highlighted.

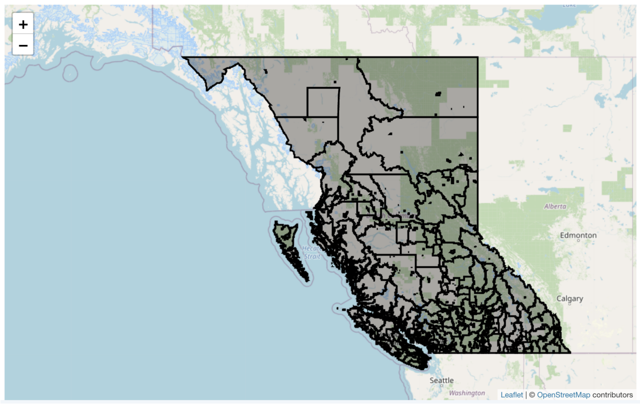

Figure 9. Work in progress implementation of the map component of the visualization tool in Observable D3 using Leaflet.js.

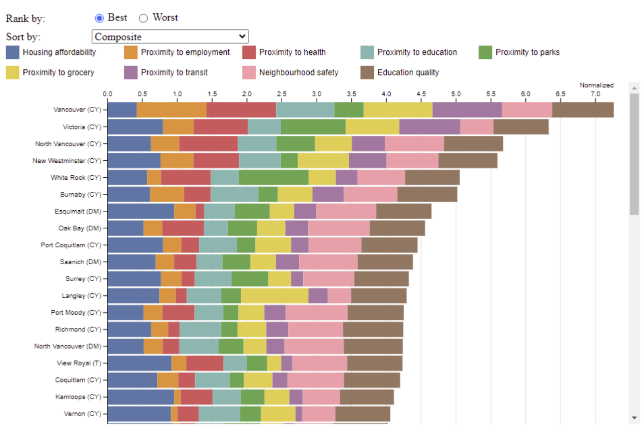

Figure 10. Work in progress implementation of the ranking view in Observable D3.

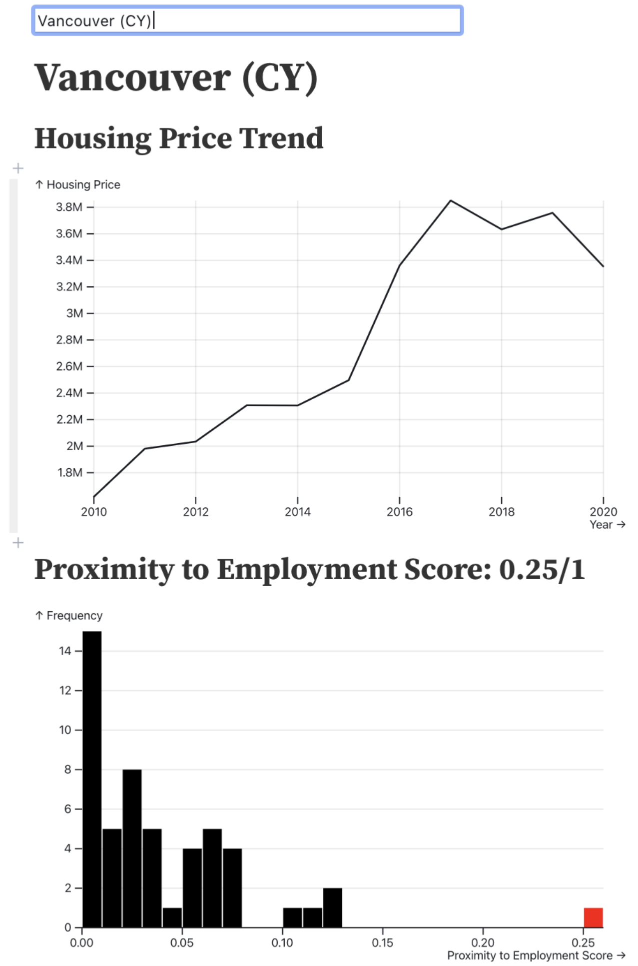

Figure 11. Work in progress implementation of the detail view in Observable D3.

6 Milestones

Table 5 outlines our proposed milestones, their associated deadlines and the team members assigned to the task. The total amount of hours estimated is 80 hours per group member.

References

[1] Joshua Gordon. Reconnecting the housing market to the labour market: Foreign ownership and housing affordability in urban canada. *Canadian Public Policy*, 46:e2019009, 02 2020.

[2] AbdulLateef Olanrewaju and Chew Wong. Evaluation of the requirements of first time buyers in the purchase of affordable housing in malaysia. *Journal of Housing and the Built Environment*, 35:309 – 333, 2020/03/01 2020.

[3] Statistics Canada. Canadian statistical geospatial explorer, 2021.

[4] BC Assessment. Bc assessment maps, 2021.

[5] Shakti Ramkumar. Visualizing housing affordability in metro vancouver, 2018.

[6] Vancouver Police Department. Geodash crime map, 2021.

[7] BC Data Catalogue. Foundation skills assessment 2017/18-2020/21, 2021.

[8] Walk Score. Cities & neighbourhoods, 2020.

[9] Stacey McLachlan. Find your neighbourhood: The vanmag neighbourhood ranking tool, 2018.

[10] Clockworkmice Ltd. Find your perfect neighbourhood, 2021.

[11] Claus Rinner. A geographic visualization approach to multi-criteria evaluation of urban quality of life. *International Journal of Geographical Information Science*, 21(8):907–919, 2007.

[12] Liu Jianxiao, Bi Han, and Meilian Wang. Using multi-source data to assess livability in hong kong at the community-based level: A combined subjective-objective approach. *Geography and Sustainability*, 1(4):284–294, 2020.

[13] Carlos JL Balsas. Measuring the livability of an urban centre: an exploratory study of key performance indicators. *Planning, Practice & Research*, 19(1):101–110, 2004.

[14] Reza Shabanzadeh Namini, Mirella Loda, Abolfazl Meshkini, and Abdolreza Roknedineftekhari. Comparative evaluation of livability indicators of the metropolitan tehran's districts. *International Journal of Urban Sustainable Development*, 11(1):48–67, 2019.

[15] William Cartwright, Suzette Miller, and Christopher Pettit. Geographical visualization: past, present and future development. *Journal of*

Spatial Science, 49(1):25–36, 2004.
[16] Alan M MacEachren, Monica Wachowicz, Robert Edsall, Daniel Haug, and Raymon Masters. Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science*, 13(4):311–334, 1999.
[17] Mohammad Shaito and Ramez Elmasri. Map visualization using spatial and spatio-temporal data: Application to covid-19 data. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*, pages 284–291, 2021.
[18] Sabrina Nusrat and Stephen Kobourov. Visualizing cartograms: Goals and task taxonomy. *arXiv preprint arXiv:1502.07792*, 2015.
[19] Danyel Fisher. Hotmap: Looking at geographic attention. *IEEE transactions on visualization and computer graphics*, 13(6):1184–1191, 2007.
[20] Marc van Kreveld, Étienne Schramm, and Alexander Wolff. Algorithms for the placement of diagrams on maps. In *Proceedings of the 12th annual ACM international workshop on Geographic information systems*, pages 222–231, 2004.
[21] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics*, 19(12):2277–2286, 2013.
[22] authro Speckmann and Kevin Verbeek. Necklace maps. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):881–889, 2010.
[23] Ramana Rao and Stuart K Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 318–322, 1994.
[24] Jason Davies and Tom Carden. Crossfilter: Fast multidimensional filtering for coordinated views, 2021.
[25] Shahid Latif, Siming Chen, and Fabian Beck. A deeper understanding of visualization-text interplay in geographic data-driven stories. In *Computer Graphics Forum*, volume 40, pages 311–322. Wiley Online Library, 2021.
[26] Canada Mortgage and Housing Corporation. Price quartiles and averages: Absorbed homeowner and condo units, 2021.
[27] Statistics Canada. Proximity measures database – early release, 2021.
[28] Canada Mortgage and Housing Corporation. Absorbed units by price range, 2021.
[29] Government of British Columbia. Administrative boundaries, 2021.
[30] Qgis: A free and open source geographic information system.
[31] Leaflet: An open-source javascript library for mobile-friendly interactive maps.

7 Appendix

7.1 Attribute of Interest Survey

The following survey was used to understand the most important criteria to prospective home buyers when searching for a neighbourhood in which to purchase residential property. This survey focuses on BC home buyers specifically, but includes the wider North American channel as a secondary target. Synchrony between results gathered from both types of respondents indicate that a solution developed to visualize these attributes is scalable to a country-wide visualization tool.

7.1.1 Survey Results. Figure 15 represents the age distribution of survey respondents. To gauge whether respondents possessed any interest in purchasing residential property, respondents were asked if they had any interest in purchasing

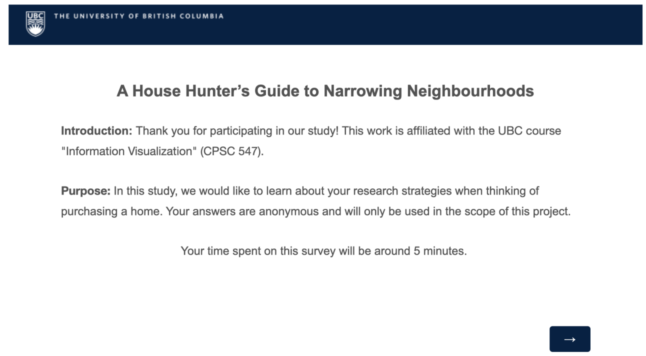

Figure 12. General description of survey objective

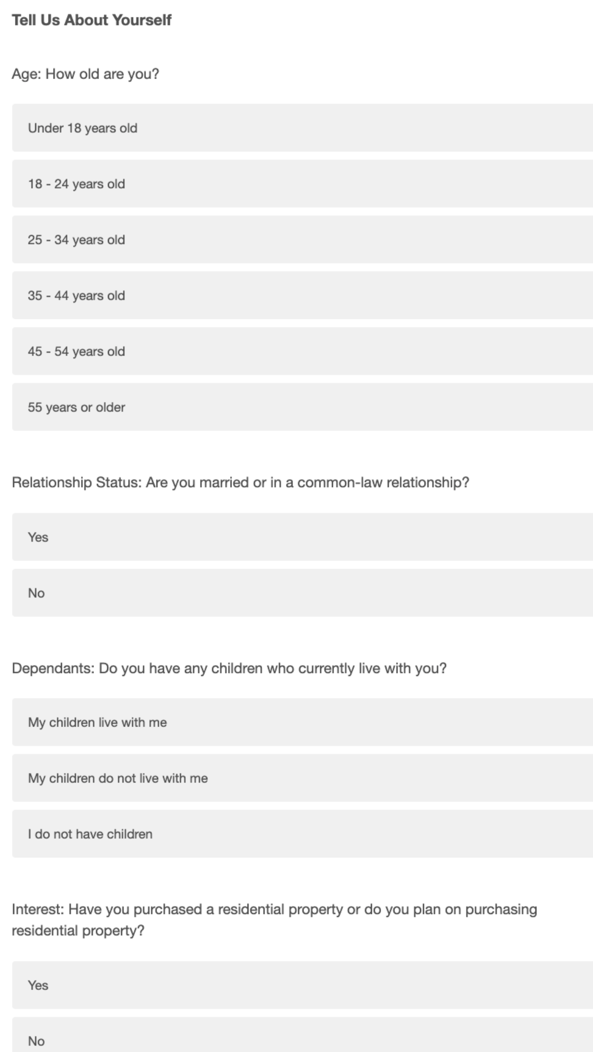

Figure 13. Demographic information about survey respondent

residential property or had purchased residential property already. Figure 16 shows that 84% of survey respondents fit our primary target demographic.

Respondents were asked to rank the following criteria from 1 to 6 in descending order of their importance: cost of purchase, quality of schools, neighbourhood safety, proximity to public transportation, investment value, and access to parks and recreation. The results of this task are shown in Figure 17. As indicated in Figure 18, it is noteworthy that 73% of respondents did not have children. This may explain the low importance ranking for the quality of schools criterion. To ensure that our criteria classes were not solely biased by prior works in the space, survey respondents were asked to share any other criteria that were relevant to the residential property search. Figure 19 depicts a word cloud encompassing these responses. As proximity to work and grocery stores were identified as recurring themes, we amended our original data to include a dataset that encompasses proximity to grocery stores, employment, health services, and more.

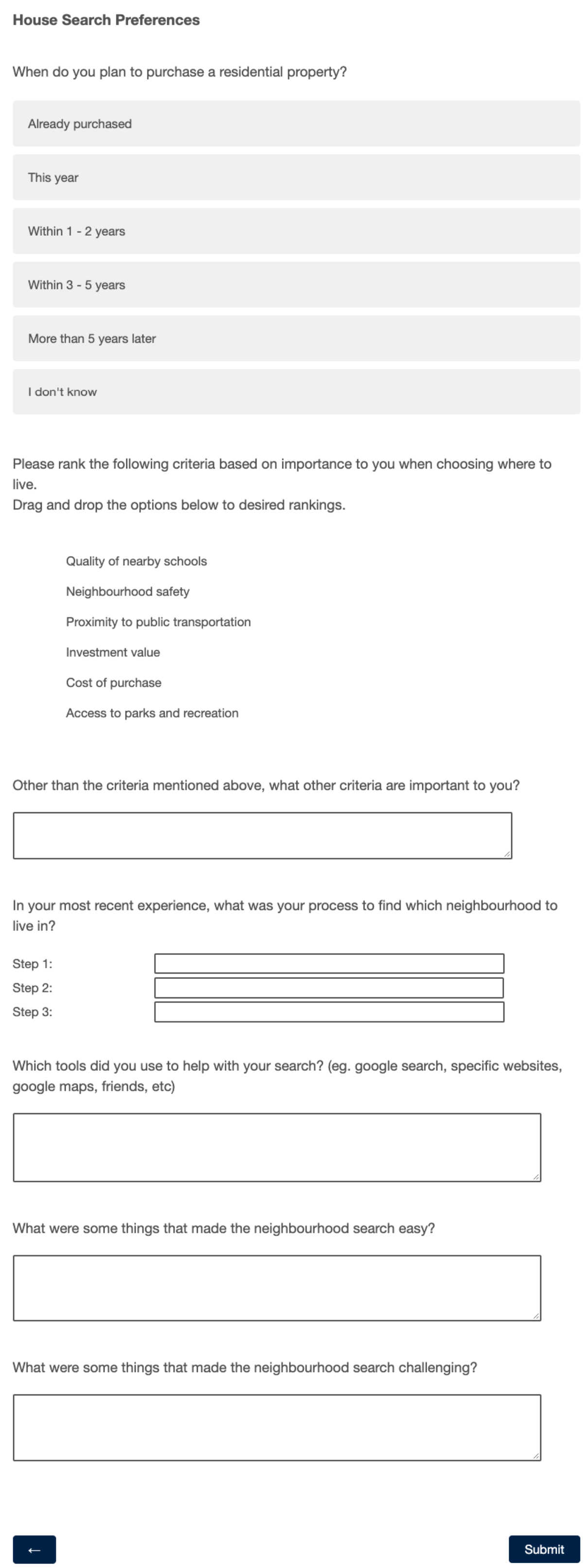

Figure 14. House search preferences of survey respondent

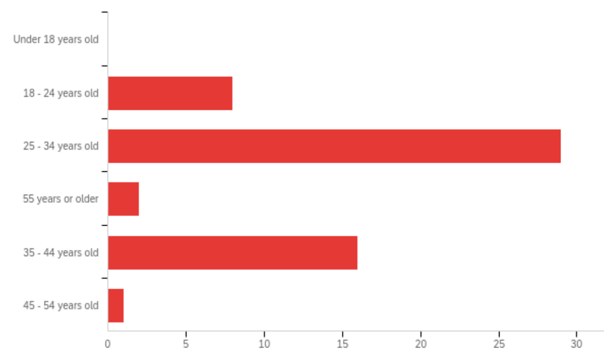

Figure 15. Age distribution of survey respondents.

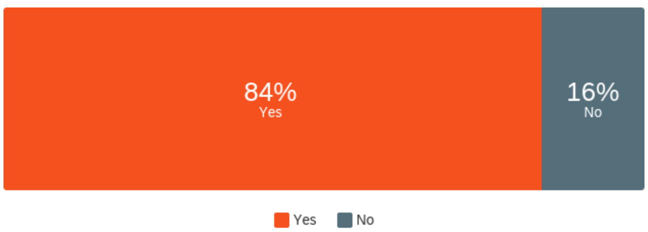

Figure 16. Assessment of past, present or future intent to purchase residential property.

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Cost of purchase	1.00	6.00	1.88	1.21	1.47	56
2	Quality of nearby schools	1.00	6.00	4.43	1.55	2.39	56
3	Neighbourhood safety	1.00	6.00	2.93	1.47	2.17	56
4	Proximity to public transportation	1.00	6.00	4.07	1.64	2.67	56
5	Investment value	1.00	6.00	3.79	1.68	2.81	56
6	Access to parks and recreation	1.00	6.00	3.91	1.26	1.58	56

Figure 17. Respondents were asked to rank the importance of each criterion when home hunting.

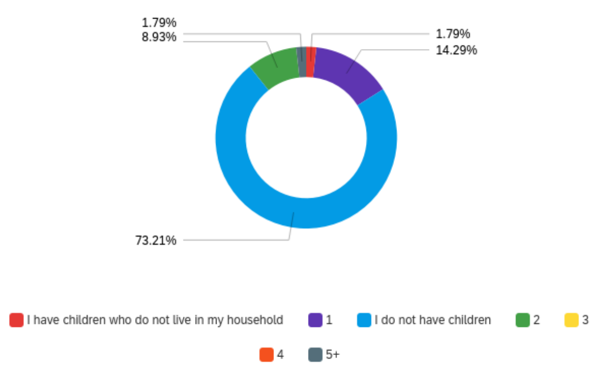

Figure 18. Demographic data about respondents' number of children.

Figure 19. Word cloud of responses from survey respondents about criteria not mentioned in Figure 4 that are important to them when choosing where to live.

Table 5. Milestone timeline

Milestone	Task	Hours per Person	Deadline	Task Description	Team Member	Status
Pitch	Individual pitches	2h	Sept 29	Research projects of interests, develop presentation, present to class or generate video	Everyone	Completed
Proposal	Idea selection meeting	1h	Oct 7	Decide project of interest, research relevant datasets, initial task split for proposal	Everyone	Completed
	Survey development and existing solution research	1.5h	Oct 12	Created a survey to gather initial user preference, researched existing solutions	Ivan	Completed
	Pre-proposal report writing	1.5h	Oct 12th	Set up overleaf, write introduction and literature review	Abi, Lucy	Completed
	Pre-proposal Meeting	3h	Oct 13	Meeting with Tamara, initial brainstorming, iterate on survey	Everyone	Completed
	Collect datasets	1h	Oct 14	Search for datasets that can be used for the project	Everyone	Completed
	Define data and task abstraction	2h	Oct 14	Meeting to discuss data and task abstraction	Everyone	Completed
	Proposal Write-up	3h	Oct 21	Finish proposal writing	Everyone	Completed
Update Report	Evaluate pros and cons of proposed solutions	1h	Oct 23	Discuss each solution ideas and decide on an unified design	Everyone	Completed
	Tool familiarization	10h	Oct 26	Learn D3, Observable, Leaflet	Everyone	Completed
	Analyze survey results	1h	Oct 26	Distribute survey and analyze results	Everyone	Completed
	Proposal Feedback Meeting	1h	Nov 2	Meeting with Tamara to discuss proposal feedback	Everyone	Completed
	Data acquisition, cleaning, filtering, and normalization	4h	Oct 30	Clean up data for implementation, discard irrelevant fields, transform data where relevant	Everyone	Completed
	Create initial implementation	10h	Nov 14	Create the MVP of the solution	Everyone	In progress
	Update writeup	2h	Nov 16	Finish writeup for updates	Everyone	Completed
Implementation Deadline	Implement filter view with map	10h	Nov 21		Abi	In progress
	Implement ranking view	10h	Nov 21		Lucy	In progress

	Implement detail view	10h	Nov 21		Ivan	In progress
	Integrate views and implement interactions	10h	Dec 3		Everyone	Not started
Validate Implementation (optional: time can be allocated to implementation if needed)	Create evaluation plan	1h	Dec 3	Create an evaluation plan for validating the solution	Everyone	Not started
	Execute evaluation study	2h	Dec 6	Recruit participants and conduct the evaluation studies from the plan	Everyone	Not started
	Summarize results	1.5h	Dec 8	Merge, discuss, and analyze evaluation results	Everyone	Not started
Final Presentation	Finish presentation	3h	Dec 14	Prepare and rehearse presentation	Everyone	Not started
Final Report	Finalize report	6h	Dec 17	Update report from update write-up, and finalize all other sections	Everyone	Not started