LocationSeeker - A Visualization guide for finding suitable locations for home buyers in the Netherlands

Shreyan Biswas,Wessel van de Brug

Delft University of Technology
Group 4

Abstract

Finding a suitable house to buy or invest in is an intricate affair. But an equally important step that leads to buying a house is identifying a suitable location. Ignoring this step often leads to regrets among home buyers as, even though they are flexible to change the features of the house, changing the location is simply infeasible. We utilize this observation and present LocationSeeker - a web-based platform that helps users explore and compare locations, suitable for finding a property in the Netherlands.

1. Introduction

Considered as one of the biggest financial asset classes, the housing market has seen unbridled growth over the last few decades. Finding a suitable property to own in this proliferated market hence has become an intricate and time-consuming process. In the current age of digitization, people rely mostly on online sites to shortlist their preferred set of properties. These online sites such as Funda[†], Rotsvast[‡] utilize state of the art technologies to provide a concise platform for real estate buying and selling.

Buying a house is a big investment and while the house itself can be modified based on the buyer's preferences, the location of the house is one factor that remains constant throughout. Sure, there can be future developments around the neighbourhood but mostly these are sluggish in nature. Hence deciding on a location becomes one of the first and most critical aspects of buying a house and holds long term consequences for the buyer's decision. In an online survey conducted Bankrate[§] in the US stated that about 10 %of millennials who were polled in the survey indicated their regrets related to the location of their house. But online sites that provide real estate buying services mostly ignore this step and tend to promote the house as its main advert [LBSY16]. This results in a search that begins with evaluating the house without giving much consideration to the location it belongs to. Users can still learn about the location but it changes the natural order of selection of first analysing the location and then the house. Moreover, due to their inherent design, the geographical information provided by these sites are also very limited and fail to include various proximity profiles related to healthcare, education, etc. Their overall design also makes it difficult for novice home buyers to explore the geographical features of a location alien to them, causing them to make sub-optimal decisions. For example, a person (especially expats) with limited knowledge about a country's geography may find it hard to explore and analyze location-specific features and hence make choices based on the pricing only. To summarize, existing commercial systems fail to provide a natural home buying experience and lack proper visualization of geo-date related profiles.

In this paper we propose, LocationSeeker, a web-based platform that enables home buyers to visualize the suitability of a location, based on their preferences and requirements in different geographical layers in The Netherlands - spanning from Municipality level all the way down to Neighbourhood level and help them explore and evaluate their selection of locations for buying real estate.

2. Related Work

Much work has gone into pricing related research in the domain of real estate but very little work has been done in terms of visualization study let alone visualization specific to spatial attributes. [PVS*05] gives an excellent overview of the housing market in the Netherlands and provides insights into spatial variation in availability, different effects of pricing, and provides simplistic visualization of quality and satisfaction. Albeit insightful above-mentioned research is only partially related to our visualization goal of assisting users to make decisions related to the location of their future homes. In this regard, the HomeSeker project by M. Li et. al [LBSY16] is closely related to our research question and project goal. It divides the real estate dataset into four different profiles corresponding to

[†] https://www.funda.nl/en/
[‡] https://www.rotsvast.nl/en/
[§] https://www.cnbc.com/2019/02/28/63-percent-of-millennial-homebuyers-have-regrets-heres-why.html

region, transportation, education, region and a facility profile that includes medical facilities among others. These profiles are utilized to help a user compare between different regions/properties and make an informed choice for buying a real estate. The authors of this visualization study approach the problem in a modular fashion, dividing the entire application in four separate modules, each of which contains a set of distinct idioms and a corresponding visualization and data abstraction that addresses a specific domain task. An interesting characteristic of the paper lies in its representation of the visualization in terms of 3 "exploration queries" that help an uninformed user attain insightful information into a housing market he is unfamiliar with.

3. Suitability Analysis

Based on [SMM12] we first try to analyze whether our visualization goal aligns with the suitability framework provided in Figure 2. We establish that the clarity of our task is more towards the fuzzy side as the scope of our task is not to recommend a house for purchasing/renting based on statistical analysis but rather provide a layered visualization of multidimensional data that helps a home-buyer identify candidate locations for their future house.

In terms of the information location, we establish that it is more towards computers. While a domain expert may have the information regarding the geography of the Netherlands in his head, filtering the map based on different dimensions of data is much more suited for a computer. Since our methodology falls in a suitable region we proceed with our visualization project.

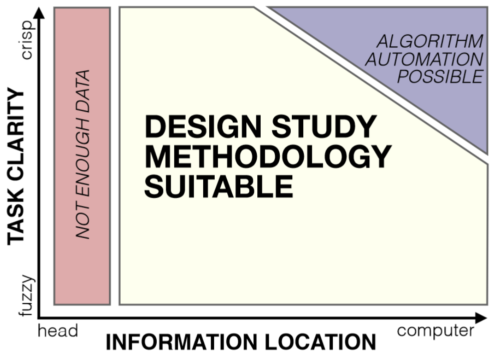


Figure 1: [SMM12] Task suitability

4. Munzners 4-level analysis framework

In this section, we discuss how our design model utilizes Munzers 4-level analysis framework [Mun09] and elucidate our design choices.

4.1. Formulating the Domain Problem

Due to the limited scope of our project, we envisioned ourselves as domain experts. We utilized our experiences with the real estate market, scientific literature in the domain and various surveys to identify the important role that location plays in the process of buying a house. Evaluating different sources we came to conclude that building an assistive visualization model that fulfils the designs gaps that current commercial entail, will be helpful to a majority of the new buyers who wish to invest their hard-earned capital on the housing market.

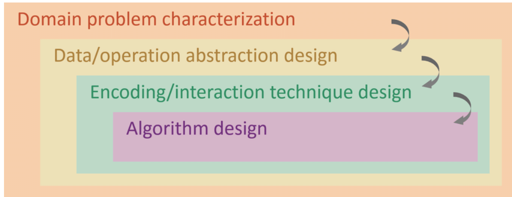


Figure 2: Munzners 4-level analysis framework [Mun09]

4.2. Data Collection and Processing

Collection: The datasets were retrieved from The Central Agency for Statistics (CBS) website using the OData¶ protocol. These datasets provided us with a platform to base our design study. Additionally, data related to The Netherlands geography and regional boundaries were downloaded manually from Cartomap‖.

Cleaning: The raw datasets were not ready for visualisation. First, they needed to be filtered such that only the relevant factors remained, such as the distances to essential facilities and the location of borders between regions. The last step in cleaning the data was translating the Dutch keywords to English.

Composition: The distance measurements and border locations could directly be represented. However, the resulting map would have been unusable due to the overload of information. Instead, we decided to combine a variety of measurements into generalized factors. These factors constituting of preferences and requirements are discussed extensively in section 4.3.

Optimization: The last step of pre-processing is focused on optimization. As the application requires a lot of geographical data to be loaded, we decided to split it up into separate files and only send the required data to the user on demand. This was possible as the map is split into regions and sub-regions for which data could be loaded separately.

4.3. Data and Task Abstraction

After analysing these datasets we identified five main preference and two requirement users may be interested in. Preferences refer to qualities that can be expressed quantitatively, while requirements present the availability of a facility or service.

¶ https://www.odata.org/
‖ https://cartomap.github.io/nl/

4.3.1. Preference Profiles

Price: While many people dream of living in enormous mansions, not all will be able to afford it. Therefore, it is important that users can limit their selection based on their budget. The datasets do not provide pricing for individual houses but do offer an average house worth which could be used as an indication of the cost of buying or renting a house in the associated region.

Urbanity: People may prefer dense cities over the countryside and vice versa. The datasets provide statistics on the urbanity, population density, and size of the population for each reason. These measurements can be combined into a single factor which the users may use to filter their regions of interest.

Healthcare: The elderly generally prefer to have a general practitioner or even a hospital close by. The datasets include distances between each region and their closest general practitioner, general practice, hospital, and pharmacy. A combination of these distances provides a great indicator of the available healthcare in each region.

Education: People with small children may want to live within reach of a primary or secondary school. A combined measurement based on the distance to these facilities is therefore essential for people looking to start or maintain a family.

Public Transport: As not everyone has a driver's license, the ability to drive, or access to a car, it may be important to have direct access to public transport. The datasets provide a measurement between regions and their closest connection to the public transport network. The datasets already distinguish between smaller and larger connections, making the translation from distance and size to a generalized accessibility factor rather intuitive.

4.3.2. Requirement Profiles

Pharmacy: Many people rely on medicines, and may not have the ability to travel far to retrieve them. It would therefore be perfect for them to know whether there is a pharmacy within walking distance from the regions in which they are interested. Again, the datasets provide measurements between each region and the closest facilities such as pharmacies and grocery stores, which often also offer a basic collection of medications. These distances can be combined into a single factor presenting the accessibility to medications.

Daycare: Similarly, it can be important to potential house buyers to have a daycare within walking distance. Especially for single parents, there may not be an alternative.

4.4. Data Encoding/Interaction technique design

In this section, we highlight the different components of our visualization and interaction design based on our data abstraction

4.4.1. Layered Choropleth map

The main component of our visualization is the choropleth map, which contains four different layers[Figure 3]. Each layer defines a specific geographical layer of The Netherlands. The default layer is the municipality layer where the user can see an overview of all the municipalities [Figure 4]. The second layer is visible once the user interacts with the map by clicking a municipality. This introduces the district level drill-down view, here the users can see information about individual districts belonging to the selected municipality. In the third drill down layer, the user can explore the neighbourhoods belonging to their selected district. Finally selecting a specific layer will zoom in on the selected neighbourhood discarding all other neighbourhood shown in the previous layer. Zooming out can be done by right-clicking any of the shown regions, or pressing the 'escape' key on the keyboard.

Justification: A choropleth map is a type of thematic map which can be used to highlight bounded box areas in proportion to their quantitative values. The benefit of this as compared to other types of maps is that users can visualize a clear pattern based on their selected requirements and preference profiles. The map was divided into different layers to provide users with both an overview and a detailed observatory choice. The visual encoding and the data points remain the same but we utilize different levels of zooming to help novice users explore the country's geographical dynamics and understand each location at different levels of granularity.

We utilize the magma colour map to embed comparisons between different regions. This colour map provides a consistent transition from dark to light. Having a clear dark and light side makes it easy for people to understand what is good and bad, and the consistent transition allows us to use it to represent continuous values, instead of discrete values, without causing banding (e.g. in rainbows, it can often look as if there are completely separated groups of values instead of one continuous range of values). It is also among the best colour maps for people with colour deficiencies due to its change in lightness in addition to the change in hue.

4.4.2. Radar Plots

We utilize a Radar chart's ability to display multivariate data in a two-dimensional chart to plot preference attributes for each layer (other than the default layer) in a concise manner. At each layer, the chart adjusts to the combined average value of the visible locations. We have also ensured that the values for all the axes are normalized so that the users do not need to perform additional processing to align those axes.

Justification: This chart adds additional depth to our design. Using the choropleth map users are only able to visualize a combined statistics based on their preferences and check the suitability of the location, however with this chart users can now see more information and compare all the available preference profiles for their selected location. It is important to note that radar charts have limitations in terms of the number or attribute they can plot if a greater number of attributes are present (more than 20) then a parallel coordinate idiom is more suited but since our design consists of only 5 attributes, radar plot is the perfect fit for comparison related task abstraction.

4.4.3. Selection Panel

This panel helps the user select their preference and requirements profiles while also giving a clear indication of the colourmap legend. We utilize five preference profiles, namely price, urbanity, healthcare, education, and public transport. The selection of these profiles is done in the form of a slider with discrete levels. The requirements are made available in the form of checkboxes.

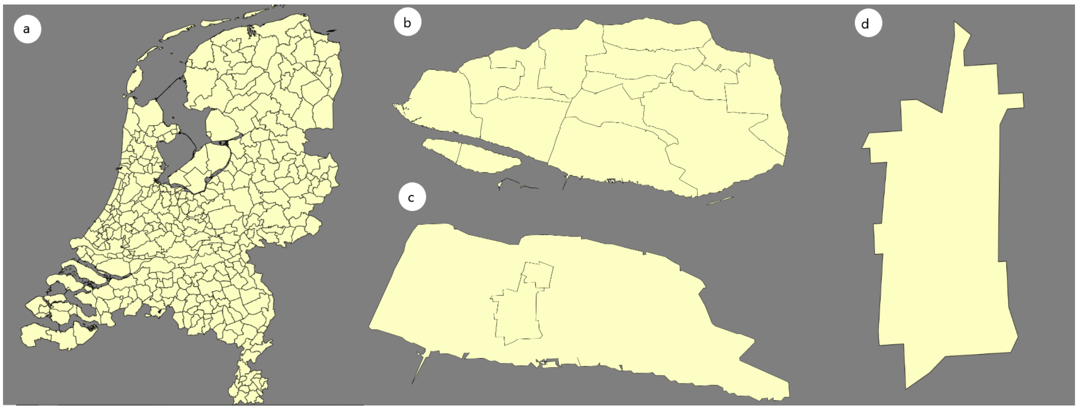


Figure 3: Different Layers - a) Municipality Level b) District Level c) Neighbourhood Level d) Neighbourhood

Justification: The main idea behind making the preference profiles include a slider is to reiterate our intention to provide users with a more exploratory approach. This section also utilizes cross-filtering idiom to coordinate between multi-dimensional data. The reasoning behind using discrete levels for the sliders is to allow users to line up the levels of different preferences and to make it easier for them to repeat a previous selection. The requirements are represented using checkboxes as these do not have a range of values to select from, they are either required or not.

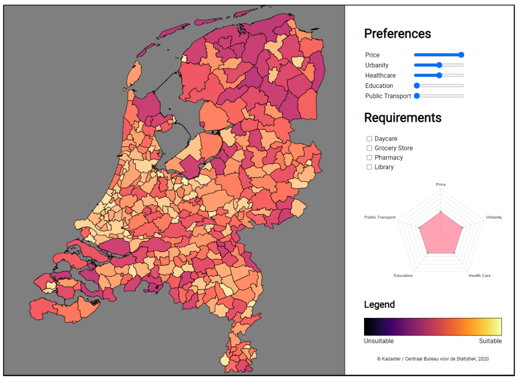


Figure 4: Interface of municipality level visualisation

5. Results and Evaluation

To evaluate our system, we present a set of case studies that in our opinion reflect the different situations a user comes across during buying a home. We evaluate each of these cases and present our observations.

5.1. Case 1: Exploring The Netherlands as an Expat (Novice Buyers)

As a newcomer, the Netherlands's geography can be tricky to understand. There are many regions and each has distinct spatial properties. Hence for this case, we consider a user who recently moved to The Netherlands and wishes to buy a house.

5.1.1. Finding Locations

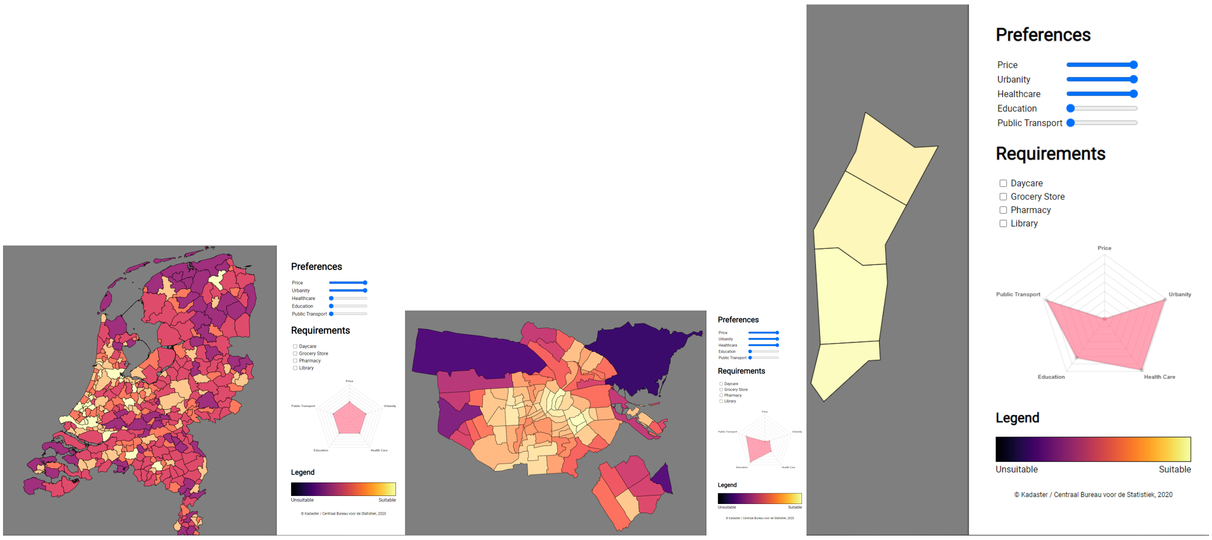
He is liberal with his budget but wishes to live in an urban area. He starts using our platform and uses the sliders to the preferences. He notices that only a few areas appear in light colour in the map[Figure 5 (a)]. He clicks on Amsterdam and notices in the spider chart that this municipality has great educational services as well. Seeing this he thinks it would be good to also have healthcare facilities nearby for the elderly in the family. So he drags the slider for healthcare all the way to left. He sees [Figure 5 (b)]. Now he can still see a few of the districts have all these three properties. He selects Grachtengordel-West and sees [Figure 5 (c)]. He realises that if he buys a property in this district it will satisfy all his preferences as all the neighbourhoods appear suitable.

5.1.2. Finding Properties

After shortlisting his location for buying a house, the user now wishes to look for properties in the Grachtengordel-West district. He is afraid due to heavy demand he may not be able to find a property here. He notices that our platform has no information related to properties. Hence he has to use a different site to explore available properties and information related to them.

5.2. Case 2: Selecting locations for investment (Expert Buyer)

A user seeing the growth of housing markets decides to invest in it. He starts using the platform but immediately notices that there is no information related to historical price data. Which could have given him an indication of how the prices fluctuated in specific locations over the years. However, he notices that he can utilize the price preference option to infer locations for a good investment. His tenet is that places with good educational reach and a large population will increase in value and he also knows for any good investment one needs to buy low and sell high. So he starts with a high preference for education and urbanity and low preference for Price[Figure 6 (a)]. He notices Rotterdam stands out so he clicks on it and sees districts on the eastern side of Rotterdam are more suitable for investment[Figure 6 (b)]. At this point he tries to see if reducing the price preferences further will change the suitability map[Figure 6 (c)]. He sees that indeed a few of the previously suitable places turned a bit darker. But he still notices that the district



(a) Overview of based on user's selection

(b) Map adjusts to the users new selection criteria

(c) Neighbourhoods that satisfy users criteria

Figure 5: Case Study 1: Novice User

Delfshaven still shows a lighter tone. He selects the district to see [Figure 6 (d)] and observes that four neighbourhoods at the centre of Delfshaven are suitable for his investment.

6. Conclusion and open/future work.

As we observe from our case study that the application due to its exploratory design choices can be utilized in many ways. And for novice users, it works as a great bootstrapping tool to get them used to the country's geographical model. But it has still failed to work as a standalone tool for home buyers. Hence for future work, we would like to integrate our application with analytics about properties and their availability in a given location. Furthermore, a historical view of the data in the form of adding a parallel coordinate idiom will also help the users infer the progression of their preference profiles over time, and can potentially suggest future trends. Nevertheless, we believe the direction of our exploratory design study is well suited for any users willing to look for their dream home and provides a base for innovative solutions in this space.

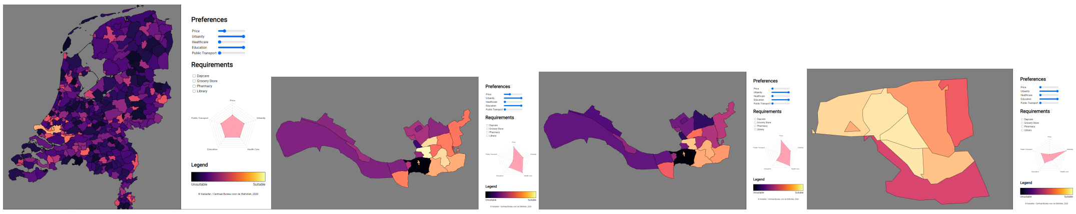
References

[LBSY16] Li M., Bao Z., Sellis T., Yan S.: *Visualization-Aided Exploration of the Real Estate Data*. Distributed Computing and Internet Technology, 2016, p. 435–439. doi:10.1007/978-3-319-46922-5_34. 1

[Mun09] Munzner T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928. doi:10.1109/tvcg.2009.111. 2

[PVS*05] Pellenbarg P. H., van Steen P., et al.: Housing in the netherlands. spatial variations in availability, price, quality and satisfaction. *Tijdschrift voor economische en sociale geografie 96*, 5 (2005), 593. 1

[SMM12] Sedlmair M., Meyer M., Munzner T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics 18*, 12 (2012), 2431–2440. 2



(a) Initial Selection

(b) Overview of Rotterdam based on initial selection

(c) Map adjusts to the users new selection criteria

(d) Neighbourhoods that are suitable for investing

Figure 6: Case Study 2: Expert User

Appendices

A. Individual Contribution

Iterating the fact that this project is an outcome of a intense collaboration we present our individual contributions below:

A.1. Shreyan

Beginning of this project I participated in the brainstorming sessions extensively and helped establish the abstract idea of our project. Next, to understand the domain and come up with a probable research question I explored research papers and user surveys in housing and real estate domain extensively. To understand design studies itself, I also read through different visualization papers by T. Munzner, M. Sedlmair et al and M. Li et al. This gave me insight into turning our abstract idea into a design study. For the implementation part, I created the prototype for the interactive map of The Netherlands consisting of different layers in d3 and JavaScript. I also utilized my knowledge in ChartJs to build the radar chart and integrated it with the main application. I extensively worked on the report to give it a proper structure and content. Writing the introduction, evaluating related work, evaluating our design methodology, justifying the usage of components, and coming up with case studies for evaluating our project. Finally, I also screen cast the presentation video.

A.2. Wessel

At the beginning of the project I participated extensively in the brainstorming sessions. Together, we came up with an idea that unfortunately did not stand up to the requirements of the project. However, by the time we realized, I had already implemented the data collection and pre-processing for that idea. After we had found a new subject, I again set out to collect and pre-process the data, after which I focused on implementing the options panel, composition of variables, and coloring of the interactive map created by Shreyan. I also added an additional layer to this map and reworked much of the code to increase its level of detail and improve its loading times. Finally, I wrote a number of sections of the report and helped in its styling.

A.3. Rebecca

Unfortunately, Rebecca was unable to attribute to the project after the first few brainstorming sessions and the quest for appropriate datasets. She has informed the responsible instructors in week 3 or 4 of the project, at which point she had not yet contributed to the code or report. In the end, she helped a bit on pre-processing, although her contributions were mostly experimental and did not end up being included in the final version of the application.