Visualizing real-estate data of the Netherlands

Rick Groeneweg Dongqi Pu Chang Liu Longxiang Wang

Liselore Borel Rinkes

April 26, 2020

Abstract

The report demonstrates our group work on Information Visualisation. It starts by motivating the added value of data visualisation to our target audience, namely, house hunters and real-estate investors in the Netherlands. Then some of the related work is briefly explained. In Section 3, the main data sources used in this project are described as well as the necessary data cleaning and aggregating work done to facilitate data visualisation. Afterwards, all main visualisation component used are described in details. We finish by summarizing our findings and comment on possible improvements.

1 Introduction

Finding new real-estate that meets all your desires is not an easy job. Whether or not a neighborhood suits you depends on many aspects, such as the housing prices, energy expenses, how crowded it is or what the facilities are in that area. Additionally, getting a good overview on all these different aspects is a challenge on its own. Alongside these points in question, the demand for housing in the Netherlands has been exceeding the offer. The housing shortage further complicates the search for house hunters in finding their perfect new home.

Taking all this under consideration, we decided to create an application to facilitate the housing search in the Netherlands. The goal of our work is to assist possible real-estate buyers in finding a good location for their new acquisition. We want to achieve this by adding multiple data sets together in one visualization, on topics ought to be relevant to potential buyers. This way, rather than searching several websites to obtain the same amount of knowledge, users can ease their search by using of just one application. Clear visualizations of these different data sets must be build to facilitate the goal. Our target group consists of house hunters, real-estate investors or anyone with a particular interest in the information visualized.

Section 2 elaborates on some related works. All data sets used are addressed in section 3. The fourth section focuses on the implementation of the application, where each individual component is covered. The discussion and conclusion can be found in section 5.

2 Related Work

House hunters resort to many different places for information. One of the most prominent sources in the Netherlands is Funda [1], which has 43 million visits per month on average according to [2]. Funda provides users with detailed information about a specific property. However, it lacks information on municipality and provincial level. It is far from effortless for house hunters to make easy choices between different areas nor gain insights about the neighborhood.

Other related work includes publications from CBS, the Dutch Central Bureau of Statistics. For example, [3] visualizes the housing price differences in different regions using a heat-map overlain on a geographical map of the Netherlands. In [4], it uses a tabular view to show the annual energy and gas consumption of the Netherlands. The drawbacks of these visualisations are: 1) they are mostly static and isolated, in

other words, it is not interactive and not connected with other related data sources; 2) they are not visually appealing enough for users to stay engaged; 3) they are not providing insights directly.
In this work, we try to use web-based technology to provide an interactive, insightful and user friendly experience for our end users. Thus, they can extract insights from data in an easy and pleasant way.

3 Data

Multiple data sets are combined to provide information and insights for our main target users: real estate investors, house hunters or people who are interested in the Dutch demographics and housing market in general. We have chosen to focus on 5 different components: electricity and gas consumption, housing prices, public transportation facilities, and population density. We based the choices of data sets on our own estimations of factors that would be of importance.

3.1 Energy Consumption Data

The main data set [5] contains the amounts of electricity and gas consumption of almost the entire country. The set contains data from 3 major energy providers: Liander [6], Enexis [7] and Stedin [8] and is publicly available on kaggle.com. These three companies yearly release a table of energy consummation in the respective areas where they are the provider. The data we used, accessible as CSV files, were from the year of 2019. For the visualization components only two columns were relevant, namely the annual consumption and the zip code. The annual consumption is measured in kilowatt/hour (KWh) for electricity and m^3 for gas. The data was processed to obtain consumption on municipality and province level. In order to achieve this, a TXT file was created containing all zip codes and corresponding provinces and municipalities. By using this file, the zip codes present in the consumption data sets could be easily linked to corresponding municipalities and provinces. The data of all three providers were combined, only lacking data on the province of Zeeland and hence corresponding municipalities.

3.2 Housing Price Data

Secondly, a data set containing the average sale-price of existing houses of 2018 in the Netherlands was retrieved from the Dutch Central Bureau of Statistics (CBS) [3]. The CBS collects data about the Dutch society. They are a government-sponsored institution. The data was aggregated on municipality and province level. On provincial level, the average price is calculated by dividing by the total number of municipalities in that province.

3.3 Public Transportation Data

Data about the public transport network in the Netherlands was made freely available[1] by the *Geodienst* of the University of Groningen[9]. The data set includes all public transport stations, including buss stops, metro stations and train stations. The data was in agreement with the situation of March 2018. For every municipality in the Netherlands, we summed the number of public transport facilities, and divided this number by the land area of the municipality in km^2. This way we construct the number of public transport facilities in every municipality per km^2, which serves as a good estimate of how accessible the real estate inside the municipality is by public transport.

3.4 Population Data

Like the housing price data, population data was also made available by the CBS [10]. From their data set we used the number of people per area, provided per postal-code region in the Netherlands and agreeing with the state of the country on 1 January 2018. We aggregated the data by summing up the total number of people per municipality and province.

[1] Under creative commons-licence CC-BY-ND NDOV

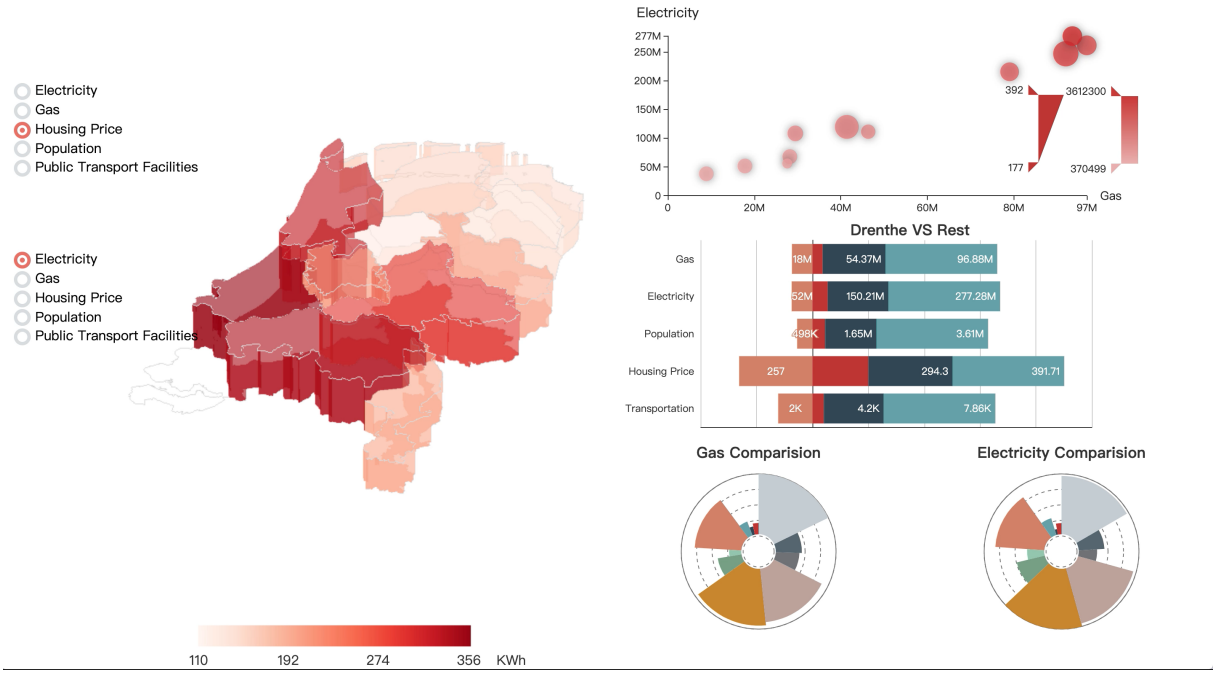

Figure 1: The visualization app with all its components. Both housing prices (height) and electricity consumption rates (color) can be read from the main 3D geographical plot. Additional visualizations are given on the right side.

4 Implementation

In this section we will first provide you with an overview of our application, after which all individual components will be discussed.

4.1 Design Overview

Figure 1 shows the overview of how all components have been put together. The 3D map of the Netherlands has been placed as centre piece of the visualization as this is our key element. Due to our instinct to scan a page from left to right, we have situated the radio buttons on the left as they can be considered as being the starting point. The remaining three components are grouped on the right side of the map, where they have been sorted from most informative and insightful to least from top to bottom. To design the radiobuttons, the IChart library was used [11] and the map was created by incorporating Three.js, which is based on WebGL. The Chart, Stacked Bar Chart, and Nightingale Rose Chart were implement by using Echart.js [12], which is a JavaScript open source library. We use it to make diagrams and interact with web application.

4.2 Components

We will now explain the separate components of the visualization.

4.2.1 Radio Buttons

Radio buttons were created to enable a user to focus on specific parts of the data, according to their preferences. Radio buttons are very effective for this specific application as we provide a selected set of options and allow just one value to be selected. We chose radio buttons over a drop-down selection mechanism as radio buttons are quicker and easier to use and we only have 5 items to show.

The radio buttons designed can be found in Figure 1 on the left. As can be seen, all 5 distinct data types are included. The design of the buttons is kept simple, as the options provided do not really lend

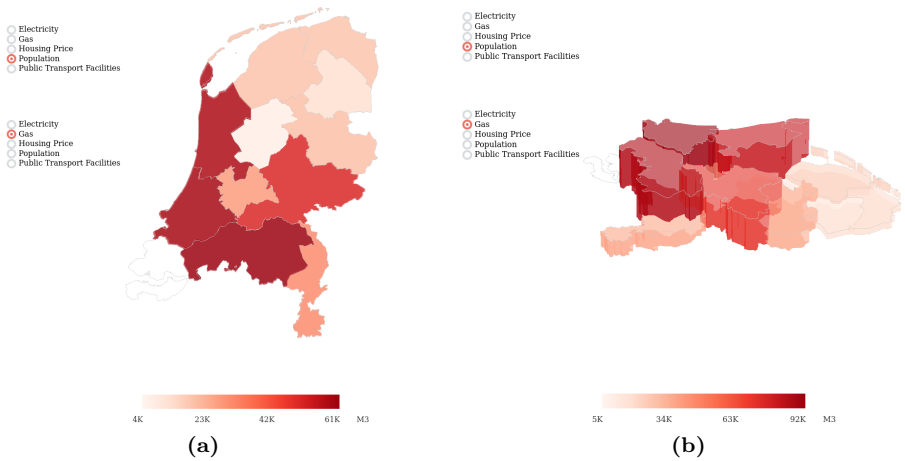

Figure 2: On the left (a) a classical heat-map visualization of the gas consumption in the Netherlands can be seen. On the right (b) the same map is rotated, showing additional information on the population sizes per province.

themselves to be displayed by icons. The radio buttons are linked to a 3D-map, where they control both the color of the map as well as the height of the bars (see section 4.1.2 for more information). Hence, a user can decide which data to show in the 3D-map. Therefore, not only people interested in electricity and gas consumption but also those more concerned with housing prices and population density can benefit from the application. As the main focus is on electricity and gas consumption, the default values for the color and height buttons are electricity consumption and gas consumption respectively.

4.2.2 3D-map

We figured to use a map of the Netherlands to immediately make the association between the data (on consumption, housing prices, public transport and population density) and the area apparent, which is ought to be of importance to the users. Instead of a regular heat-map, a 3D-map was designed. By adding a third dimension more information can be visualized. The result is a combination of a geographical heat-map and a bar-chart. This way the visualization maintains the recognizable shape of the country, that makes the spacial relationships clear. The third dimension creates a clear numerical comparability that bar-charts provide. By rotating the visualization, the user can find the optimal display of the information they are interested in.

In Figure 2 (a), a top view of the map can be found. The map divides the Netherlands into the corresponding twelve provinces. As data for the province Zeeland was lacking, that area is left blank. The data has been communicated in the form of coloring, where the darker colors represent higher values and lighter colors lower ones. The heat-map distribution depends on the selected value of the Color radio button. For this specific snapshot, the values for Gas consumption were applied.

In Figure 2 (b), the same map is displayed but has been rotated. This new angle provides additional data, as now the country is represented as a heat-map combined with a bar-chart. The height of these bars depend on any selected value of the Height radio buttons. In Figure 2, the height describes the population size per province.

As our target group covers people interested in buying real-estate, it could be useful to not only provide an overview on province level, but also present a more zoomed-in view. Therefore, it has been made possible to click on a specific province. The map changes from province level to municipality level, where all municipalities in the clicked-on province are colored and have a height based on to the underlying data. It is possible the increase the size map by using the mouse pad. This more detailed view can be found in Figure 3.

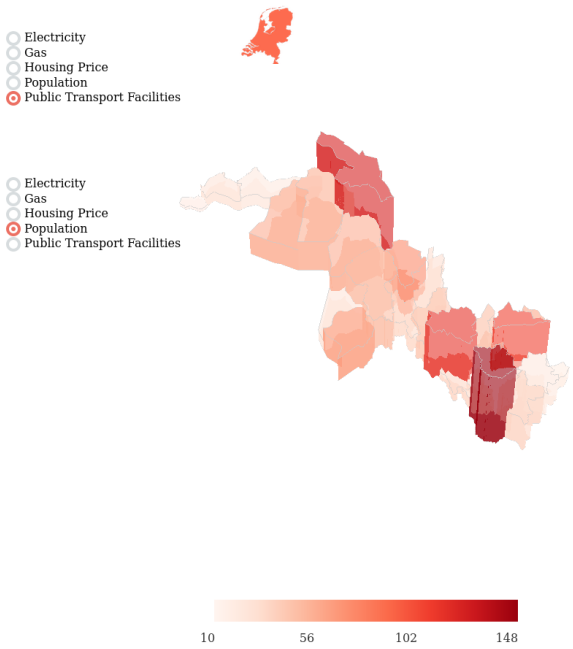

Figure 3: A 3D plot of the province of Limburg. Colors are used to visualize the population sizes of different municipalities, with a legend on the bottom. Bar heights are used to visualize how well connected each municipality is to public transport.

4.2.3 Bubble Chart

The geographical map shows colors and heights but omits the corresponding numbers. Therefore, to make the information on all data gathered more accessible, a bubble chart was added to the application. The bubble chart complements the geographical map since the two axis of the graph do not need to give geographical information. Rather than displaying three dimensional data, now four dimensions can be visualized. This way relationships in the data that are independent of geography can be studied.

Figure 4 shows the bubble chart, where each province is represented by a bubble. The chart interacts with the map by changing from province to municipality level based on mouse-movements and clicks respectively. The x-axis and y-axis represent gas and electricity consumption respectively. The gas consumption is expressed in millions of cubic metres, while the electricity consumption is expressed in KWh. The size of each bubble is dependent on the average of the housing prices in that area, which can be filtered by using the left bar on the bottom right of the chart. The higher the price, the greater the size of the bubble. Additionally, the color illustrates the population size, where a deeper color corresponds to a larger population. The color slicer is also situated at the bottom right. By enabling the function of setting upper and lower boundaries for both housing prices and population size, a user can see which provinces are most suited according to the one's preference. Each bubble is associated with a tooltip containing all corresponding values.

4.2.4 Stacked Bar Chart

While the bubble chart does provide some comparison and filter options, it is lacking more in depth comparison functionality and the transportation data is not included. Therefore, in addition to the bubble chart, a stacked bar chart was created. When a potential real-estate buyer is searching for a location, it could be interesting to see how a province or municipality scores compared to other areas. The stacked bar chart provides such information as values for a specific area will be compared to the overall minimum, mean and maximum value. For province level values will be compared to those of other provinces, while on municipality level the comparison is made for the accompanying municipalities in the specified province.

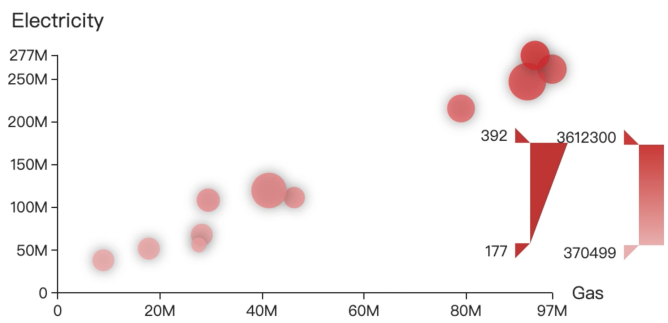

Figure 4: A bubble chart as explained in section 4.2.3. In this plot's upper right corner the most populous provinces, that also have the highest gas and electricity consumption are located. Close to the origin are the provinces that that consume much less electricity and gas. Two sliders in the figure can be used to filter out provinces based on population size and average housing prices.

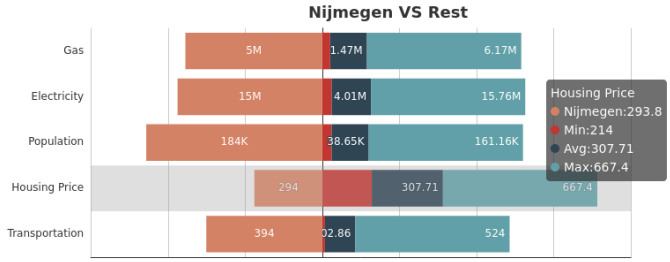

Figure 5: Stacked bar chart showing the data of the municipality of Nijmegen against the rest of the municipalities of Gelderland. Hovering over a bar provides additional clarifying information.

The stacked bar chart is shown in Figure 5. The values of the chart depend on your mouse-movements on the geographical map. Whenever a zone is hovered over, the bars change accordingly. For all five data information types a bar is created. The left side of each bar displays the value of the hovered-over area, starting at 0 and pointing to the left. That way all five bars are aligned ensuring readability. The three bar segments starting at 0 and pointing to the right side of the graph correspond to the minimum, average and maximum values of all other remaining areas. Again, tooltips have been created for each bar to present a clear list of all values to the user.

4.2.5 Nightingale Rose Chart

The final visualization components are two nightingale rose charts. The goal of the application is to give potential real-estate buyers insight in relevant data, with a main focus on electricity and gas consumption. Therefore, a clear overview of these consumption levels is essential. The 3D-map already contributes to this goal, but as can be seen in Figure 2 , the deviation in color can be very minimal while the values are considerably different. The nightingale rose charts ensure that the differences in both gas and electricity consumption between the provinces is always apparent. This type of chart was chosen over a regular pie chart because for a nightingale rose chart the area of the circle does not necessarily fill up. That makes the distinction between values close to each other straightforward.

In Figure 6 the charts can be found, one for gas and one for electricity consumption. The larger the area, the higher the consumption for that province. When a part of the chart is hovered over, a tooltip shows which province/municipality it represents and what the consumption for that area is. Additionally, to support the sizes of the chart elements, the tooltip displays the percentage of that province/municipality compared to the entire consumption in the country/province. Whenever one performs mouse-movements over the 3D-map, the nightingale rose charts will react accordingly by pointing out the area in the charts.

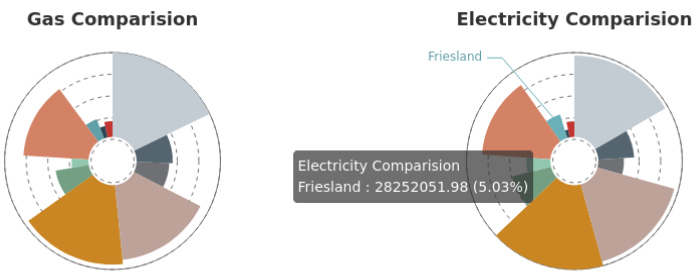

Figure 6: Nightingale rose charts showing comparisons in Gas and Electricity consummations. When hovering over a province or municipality, extra clarifying information appears.

5 Discussion and Conclusion

The visualization makes the data on real-estate more approachable to the end user - people who are interested in buying real estate in the Netherlands. The combination of a geographical heat-map and a bar-plot using a 3D plot gives the user extra insight, which could be really helpful. The ease at which information appears by hovering over the visualization makes the data more approachable. Getting extra information by hovering over a visualization element also reduces the need to explain every element in the visualization with text, reducing clutter and making everything more streamlined. The right balance, however, is subjective and rather difficult to find. We found that a strong relationship between the main visualization of the Netherlands and the other visualizations makes it easier to get the information the user wants. For example, when the user hovers over the municipality of Nijmegen, the other visualizations change and clarify how their information content is connected to the data about Nijmegen.

In order to investigate whether or not the application truly assists possible real-estate buyers in finding a good location for new acquisition could be further invested by analysing real user behavior. In addition, experiments with users could show how effective the designed components are by for example providing people with certain tasks to be completed. Recording the number of mouse-clicks or time spend could give insight into which parts of the application need improvement by for instance providing more explanation, and which are used efficiently.

Furthermore, the application could benefit from adding more data sets. For example data on number of households, number of houses available per area could be included. Questioning real-life house-hunters could assist in determining the relevance of possible new data sets. Additionally, an even more zoomed-in view, for example a city level, could be interesting for potential users. The data sets used limited us to stop at municipalities.

Overall, we believe we have created an application that makes real-estate data on distinct topics more accessible than already existing platforms do. However, experiments should be conducted to validate this belief and to further improve the application.

References

[1] Funda. *funda.com*. URL: `https://www.funda.nl/`.

[2] Funda. *funda.com*. URL: `https://content.funda.nl/over/funda/`.

[3] Centraal Bureau voor de Statistiek. *Regionale verschillen in huizenprijzen groter [from Dutch: Regional differences in house prices increased]*. URL: `https://www.cbs.nl/nl-nl/nieuws/2019/12/regionale-verschillen-in-huizenprijzen-groter`.

[4] Centraal Bureau voor de Statistiek. *Energy consumption private dwellings; type of dwelling and regions*. URL: https://opendata.cbs.nl/statline/#/CBS/en/dataset/81528ENG/table?ts=1584633156551.

[5] kaggleweb. *datadiensten*. URL: https://www.kaggle.com/lucabasa/dutch-energy.

[6] lianderweb. *datadiensten*. URL: https://www.liander.nl/partners/datadiensten/open-data/data.

[7] enexisweb. *datadiensten*. URL: https://www.enexis.nl/over-ons/wat-bieden-we/andere-diensten/open-data.

[8] stedinweb. *datadiensten*. URL: https://www.stedin.net/zakelijk/open-data/verbruiksgegevens.

[9] University of Groningen. *Openbaar Vervoer Haltes Nederland [from Dutch: Public Transport Stops Netherlands]*. URL: https://www.arcgis.com/home/item.html?id=68d84ef791594f329dc4b557645090a9#overview.

[10] Centraal Bureau voor de Statistiek. *Bevolking en huishoudens; 4-cijferige postcode 1-1-2018 [from Dutch: Population and households; 4-number postal codes 1-12018]*. URL: https://www.cbs.nl/nl-nl/maatwerk/2018/49/bevolking-en-huishoudens-4-cijferige-postcode-1-1-2018.

[11] Radio Buttons Customization Plugin.

[12] The Apache Software Foundation. URL: https://www.echartsjs.com/.