Multimedia: Visualisation of student residences in Leuven

Simon Chuptys, Niels Claeys, Jeroen de Coninck, Stijn de Haes
KU Leuven
Belgium
[simon.chuptys;niels.claeys;jeroen.d3.coninck;stijn.de.haes]@gmail.com

ABSTRACT
In this paper we explain the different stages in the development of an information visualization application. The application presents information about the student housing situation in Leuven. The paper elaborates on the process of creating this visualization and has the following structure: 1) A description of use-case scenarios for the visualization, 2) The selection of data we chose to visualize and how they were gathered 3) The progression of the visualization through different versions 4) Related work that proved useful in the development of the visualization. 5) The tools and packages that were used and 6) Evaluation and a conclusion of the used visualization.

Author Keywords
HCI; information visualization; data analysis;

ACM Classification Keywords
H.5.m. **TODO:** Information Interfaces and Presentation (e.g. HCI): Miscellaneous

BASIC IDEA

The visualisation described in this paper presents the residences of the students in the city of Leuven. The reason for choosing this visualisation was that, being students ourselves, we could relate to the topic: Most of us had to search for a student residence and therefore it would be interesting to work on this visualisation. Furthermore we are convinced that it is a common concern with our fellow students, who will also be interested in the created visualization.

The purpose of the visualisation is not to show individual houses and their prices but rather to aid the student to identify the best locations to start his search for a new residence. We offer them an interactive visualisation through which they can find pattern, insights in the data. Possible patterns could be: Only residences around the Oude Markt fit my budget range, all my fellow students have houses near their Campus, all engineering students have houses outside the center of Leuven, ...

For every residence, the price, surface area, length of the contract, overall satisfaction, address, distance to the campus and the field of study is recorded (see section 'Housing Data' for more info). The student housing data will be shown on a map of Leuven. In order to aid the students in exploring this data, they can pick a preferred price and area range. Only corresponding residences will be shown on the map. This way, the user can see where most of best houses fitting his criteria are located. In order to further narrow the search it is necessary that the user interacts closely with the visualization.

The visualisation can be used by different users, all with their own vision on the data, as explained below.

New student

The first (and main) usage scenario is for a student who is looking for a new residence in Leuven. The person doesn't know Leuven nor the typical properties of student residences or locations of residences in Leuven. This user wants to get an overview of: 1) The prizes and surface areas of student residences, 2) Which locations are good for a student residence and which places are to be avoided (for instance, too expensive), 3) Where students from a similar field of study are housed, 4) What current residents think about their home (Are they satisfied with it? Would they rather have another residence?), 5) How far the residence is located from the campus.

We used these properties because we found them most important when selecting a residence. In the evaluation part we also present the opinions of other students on the chosen properties.

The purpose of this visualisation is to answer all these questions and to guide the student towards a suitable location for his new residence.

Settled student

This student already has a residence and he is curious to see if he pays a correct price. Maybe there are cheaper residences in the same area? Since this person has a better view on the residence-case, he will mainly focus on comparing his current residence with other residences. The student might want to give a review (or more data) of his current residence.

For this scenario it is important that our visualisation makes it easy to compare all residences in Leuven rather than just using the user input to favour certain houses.

Residence provider

This person is interested in the general distribution of the residence characteristics in Leuven. He wants to know which price he should ask for the residence in order to stay competitive, while not asking a price that is too low. The following information would be useful for him. 1) Prices of residences close by and more generally in Leuven. 2) Surface areas of residences close by / general for Leuven. 3) Other accommodations offered for these prices.

We used these questions to distinguish what is important or not in the visualization. For this visualization to help these people it is important to have as much recent data as possible about the residences. The data used will be discussed in the following section.

HOUSING DATA

This section elaborates on what data was used, the reasoning behind the selection and how it was gathered.

The Data

The data received can be split into 2 groups based on the way they were gathered. The KU Leuven data has the following information: 1) Address 2) Price (monthly) 3) Contract duration 4) Surface area 5) Field of study of the current inhabitant. The data we got from crowdsourcing has some additional data entries, being the mean travel duration to classes and an overall satisfaction rating (on a scale from 1 to 5).

Why these data?

Since we envisioned our visualization from the beginning to utilize a map, gathering the addresses of the different residences was necessary. The main goal of the visualization was to help students in finding suitable residences. Conventional methods (most often used by "immo" offices) to compare different residences typically utilize the price and surface area, hence this data was also included in our visualization. Furthermore, we wanted to provide some guidance for new students who typically have little knowledge of the distribution of residences and learning facilities in Leuven. In order to do so, we opted to include the fields of study of the current inhabitants and a mean travel distance to classes. In this way, new students who plan to start their academic career in a specific field of study can identify the residences where their classmates most likely will be housed. We also hoped to discover some interesting dependencies between the fields of study on one hand and the location, price and surface area on the other hand. the overall rating of a residence seemed interesting to us since it can both help students in selecting a residence and residence owners in improving their facilities.

Gathering the data

One of the difficult things was the gathering of the necessary information. Following steps were taken:

Firstly, there exist many "immo" offices that rent hundreds of student houses in Leuven. After the first contact, it was obvious that they did not want to help with this visualisation because they did not trust us with their data. Also because we had no prototype we had nothing to gain their trust. This was a shame because they could also benefit from our visualization.

Next, there is the university of Leuven which has a lot of information about the students and their residences. They were a lot more cooperative and it was possible to get data from them. The only problem was that this data was not available during the first month of implementation. Therefore we used another approach to gather some initial data and start our first visualisation.

The last option we considered was crowdsourcing. The data was gathered based on a simple form and then distributed trough social media to get enough responses. This approach worked very well: within a week our data almost reached 200 entries which was more than enough to get started with.

DESIGN OF VISUALISATION & INTERACTION

The following paragraphs discuss the different stages in the development of the visualisation. During the process, weekly feedback was provided by means of a blog where people could experiment with and comment on the latest version of the visualization. Furthermore, co-students could evaluate the visualization during presentations that were held during the multimedia course.

Mock-up

We started by creating a static mock-up of our visualization (see figure 1). Our initial ideas were inspired by a similar visualisation about the housing prices in London [1]. From the start, it was clear that the map of Leuven should play a major role. Therefore, this map is the most dominant visual element in our mock-up. The user should be able to zoom and pan the map, allowing to switch between a more global or detailed view. The location of student residences is indicated by small circles on the map. Three buttons on the map allow the user to change the color of the circles to denote either price, surface area or price per square meter. The circle radius is proportional to the number of residences that are available at the same address. To allow the user to filter the displayed residences, sliders are added that control the price range and surface area range. Furthermore, the travel distance to a specified location on the map can be selected. This results in an overlay mask on the map, indicating which residences lie within specified travel distance. Initially the travel distance was delivered by the crowdsourcing data, afterwards we used Google maps API to calculate it accurately (this is discussed in section).

Our main concern at this point is what should happen when a large number of residences is displayed on the map: Will the view be too cluttered? Do we need to aggregate residences that are close to each other? What color would this aggregation have? Etc. An implemented prototype version which uses actual data will give us more insight in these matters.

V0.1

Based on the mock-up, an initial version was implemented. Figure 2 shows this visualization. We used the data gathered from our crowd sourcing efforts, which contained about 250 entries at the time. For this first version, we focused on the map display, the color coding of residences and the filtering using sliders. For now, it seems like our concerns from the mock-up aren't an issue: the individual residences remain distinguishable, even when showing the full Leuven area. A problem that arose was the scale used for the color coding. The colors scale linearly in the range specified by the sliders. However, our data contains a few outliers in price and area, which results in a color coding that gives almost all residences the same color when these outliers are included in the specified range. One could reason that, if the user wants

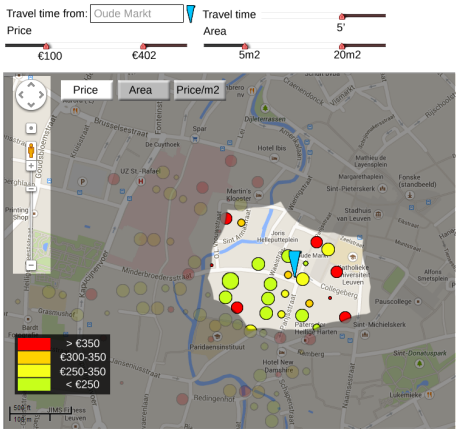

Figure 1. The static mockup of the visualisation.

a more interesting color coding, she should change the range sliders to exclude the outliers. However, we would like to present the user with an overview, displaying all residences, that still provides a useful color coding. Therefore, we will change this linear scaling in future versions. We gave a presentation demonstrating this version of the visualization to our co-students. The comments we received after this presentation confirmed our findings.

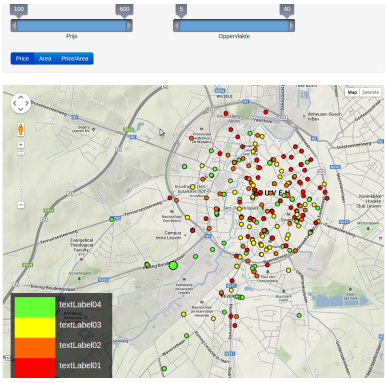

Figure 2. Version 0.1

V0.2

In the second version of our visualisation a couple of subtle adjustments have been added which are shown in Figure 3. First of all, pop-ups are introduced to give extra information about a residence when hovering over it. This way the user can see the price of the residence together with the surface area and the number of residences located there. Furthermore, we hooked up the color legend to the rest of the GUI, so that it displays the range of values corresponding to each color. We also experimented with the color-scale: Instead of using a linear scale, a quantile scaling was implemented, meaning that each color contains about the same number of residences. Two different opinions exist on the result: On one hand, the ranges on the color legend now give a better understanding of the distribution of residences in the selected price and area ranges. On the other hand, the map now always shows an equal number of circles of every color, losing this same insight in distribution on the map. We found that the linear scaling gave you better information so linear scaling was used. However, it also proved difficult to get a more detailed impression of the distribution this way. A related issue that arose during a discussion with our co-students was that, when dragging the sliders, suddenly a lot of residences appeared or disappeared because of the distribution of the residences. The user should be presented with a way to tell if a lot of residences lie in the specified ranges. These issues are handled in the next version.

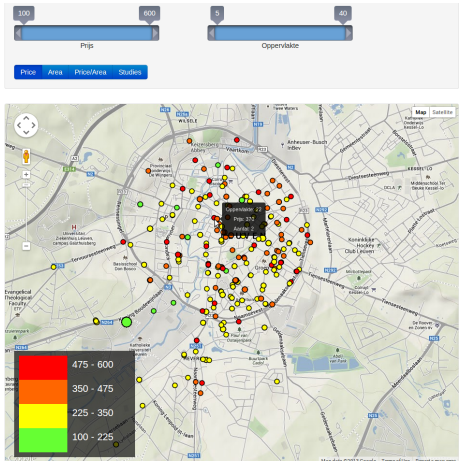

Figure 3. Version 0.2

V0.3

In the third iteration some extra functionality is added in the form of bar graphs, and distance calculation, as can be seen in figure 4. The bar graphs above the range sliders show the distribution of the prices and surface areas. With this information, the user gets a better notion of how many residences exist with the specific price/surface area, solving the issues that arose during the previous version.

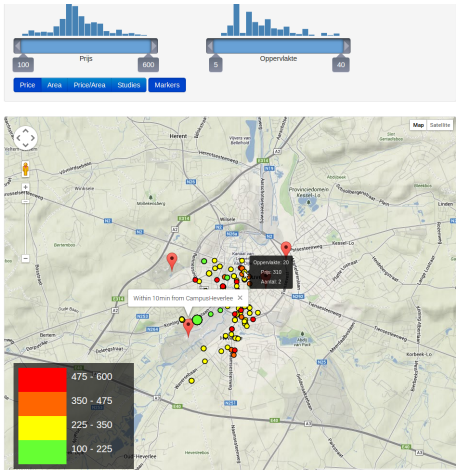

Figure 4. Version 0.3

V1.0

TODO: **figure** In the final iteration we integrated the data received from the KU Leuven into our visualisation. This

means that we show a lot more data and this also means that we use Json instead of a google spreadsheet to display the data. Another improvement made in this iteration is the additional information of a specific location. You get to see 3 bargraphs, representing the price, area and study distribution of this location. This is visible when a user hovers over a circle. This was needed because with the increase of data, one circle now often represents multiple residences. On the map we only show the averages over all residences, but the extra view also shows the distribution on the location.

RELATED WORK

When developing an information visualization, it is important to understand what the focus of the result should be. We should consider a *data focus* and a *mapping focus*. Of course, a visualization can be at any point between these two extrema. [7] defines and discusses these different focus groups: The author presents a two dimensional area where each visualization focus can be mapped upon (see figure 5). The vertical axis defines how the information behind the visualization is shown and focusses on the functionality and usefulness of the visualization. The horizontal axis represents the artistic influence that exists in the visualization. Based on the techniques that are used, different kinds of visualizations will be created. On this representation, our visualization is situated in the bottom left corner. The reason for being situated here is due to the goal of our visualization. The idea is to help users to get insights into our data and to display this data as useful as possible. This defines the positioning of our visualization.

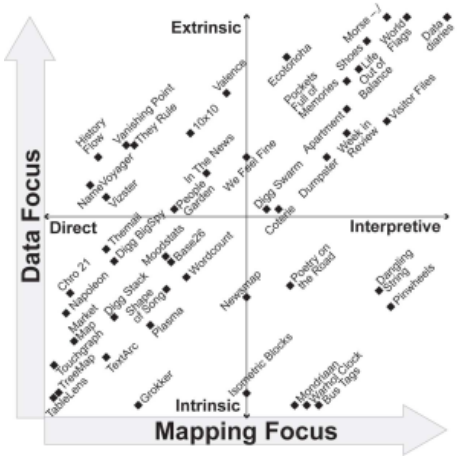

Figure 5. This illustrates information aesthetics of different visualizations.

Relevance to our visualization

Data Focus

On the point of data focus we have primarily an intrinsic data focus. We want to help people find an interesting region to live and thus want them to get an insight of where those interesting regions are.

Direct mapping

We mainly use a direct mapping approach. It is not possible to see the exact underlying data but it is inferred by the color of the houses. This gives the user the necessary tools to find an interesting region where the housing is to his liking.

Extra

Additionally, we want to promote interactivity. By making our visualization interactive, we hope that people get more insight and are able to narrow down the interesting regions.

Human perception

When designing a good information visualization, one needs to understand how humans perceive what they see. For example, our vision is sensitive to changes in color and is trained to identify patterns in visual data. [5] discusses these characteristics and we will apply them to our visualization. A good visualization should: 1) indicate the relation between objects: All the circles represent different houses. Based on the highlighted button users can see which property is shown. 2) represent quantities accurately. The colors show an approximation of the values but when you hover over the circles it is possible to see the exact values. 3) make it easy to compare quantities. The different colors make it easy to compare the values. 4) make it easy to rank the values. The ranking of values can also easily be performed based on the colors. It is intuitive to use red circles for expensive houses and green for cheap ones which makes the visualization easy to understand. 5) make it clear how the information should be used. The sliders, buttons and the ability to zoom make it obvious that the main idea is to play with the visualization. Only through using it, the patterns can be seen.

Empirical studies in information visualization

Paper [6] proposes 7 scenario's to evaluate a visualization, only the ones related to the visualization will be discussed. The focus for our visualization lies on getting additional insights and not on user experience or usability.

Evaluating Visual Data Analysis and Reasoning

This scenario assesses the visualization's ability to support visual analysis and reasoning about the data. The following questions should be answered:

1)Data exploration? How does it support processes aimed at seeking information, searching, filtering, reading and extracting information? Our visualization provides sliders to set price and area ranges. Also, the user can choose to either use color coding for price, area or price/area. While the sliders have the purpose of narrowing down the dataset, the colors are meant to quickly get an idea about the selected parameter. The actual data (the numbers) can be accessed on a per-residence basis by hovering over the residence. **2)Knowledge discovery?** How does it support the schematization of information or the (re-)analysis of theories? In our case, theories like residences in noisy areas are cheaper can easily be verified by having a look at the color-coded map (provided the user knows which areas are noisy). **3) Hypothesis generation?** How does it support hypothesis generation and interactive examination? By examining the map of all residences, one might try to find patterns in the colors that are displayed to form hypotheses. **4) Decision making?** How does it support the communication and application of analysis results? Residences that do not specify the users needs are filtered

from the map by means of the sliders. Which of the remaining residences are to be inspected in more detail is up to the user, who can hover over them to acquire specific data.

Evaluating communication through visualization
This scenario tries to reveal if the visualization helps users to get an understanding of the data. The main question that should be answered by our visualization is the following:

Do people learn better/faster using the visualization tool? Our tool will help students with the hardest part in finding a residence, which is getting to know what the options are. Using our tool, students will get an idea about typical price and area ranges. Furthermore, they get an overview of which residences are worth a visit, which is important to know since the time span that residences are open to public is typically very limited. An evaluation of our visualization with actual test persons (see section 'Evaluation') will assess if our visualization succeeds in the goals we have in mind.

Inspiring visualizations

One of the most inspiring visualization was [4], which describes the liveability of Melbourne. This is because the visualization uses similar data and because it illustrates the data in an interactive manner which leads to a better understanding for the user. One remark on the visualization is that it is difficult to understand the scoring scheme, this is something that we must avoid. We found it not clear how the scores were defined which leads to a less effective visualization because you have no idea what the color proposes.

Another inspiring visualization is one about the Irish population in the US[3]. This visualization is nice due to its simplicity of only using one color. It also is a great example of a visualization that displays data on a large scale. One minor issue of the visualization is that it only uses one property which is easier than showing more properties on the same map.

The next visualization that inspired us displays information about meteorites[8]. The good things about this visualization are the way in which the circles change when zooming in and the presentation of the bar chart. The bar chart describes nicely when a meteorite was found or hit the ground.

Lastly we were also inspired by a visualization about the housing prices in Slovenia [2]. The data used is similar to ours. Furthermore zooming in is nicely handled in the visualization. A minor remark is that the visualization uses to much colors for specifying the prices of the different residences.

Theory on creating a visualization

In this section we will discuss how well our visualization follows the theory behind visualizing data.

Figure 6 shows what technique you should use for displaying different sorts of data. In our visualization we use both quantitative data (prices, areas, number of residences) as well as categorical data (field of studies). For the number of residences we use area. Based on the figure this are not the best choice but it's not a terrible choice because position was already used. In the different visualizations of area, price, price

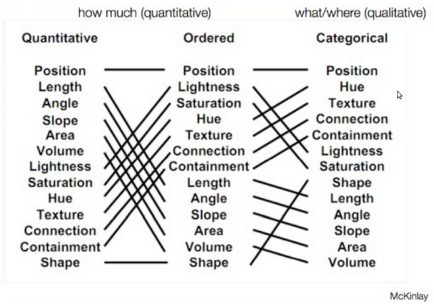

Figure 6. Techniques to display properties in a visualization.

over area and studies, we placed those in categories. For studies this is obvious, the others were placed in different price, area and area over price categories. These categories are represented by hue so this was an excellent choice.

Another guideline in information visualization is:

> Overview first, zoom and filter, then details-on-demand.

which was stated by Ben Shneiderman. Our visualization follows this principle perfectly. First you receive an overview of all residences. Afterwards you can filter the residences in which you are interested in and when you hover over a particular residence, you get in-depth details.

Lastly there also exist the Gestalt principles which describe some patterns that are good to use in your visualization. Our visualization shows primarily the law of similarity and the law of isomorphism. Similarity is illustrated by the different circles, that represent the different houses. Isomorphism is used for the color coding, for example when talking about prices red is expensive and green is cheap.

SOFTWARE DESIGN

TODO: zegt weinig over design eerder tools The software that displays our visualization and makes it interactive is written in JavaScript. With the help of some different libraries (D3.js, jquery), our code is restricted to 2 files: gui.js and map.js. As the names imply, they respectively handle our controls (buttons and sliders) that are able to update our visualization. The visualization code itself resides only in gui.js. This made it easy to decouple both.

1) **Data importing** is done through Miso[1]. This is used because it allows us to import our crowd-sourced data directly from Google Docs with just a few lines of code. When we started using the data from the KULeuven housing administration the data was moved to a JSON file. It was possible to keep using MISO for reading this JSON file. 2) **Data handling** is done with D3.js. D3.js ties our data to DOM elements and updates them. This framework was a logical choice since every member of the team has familiarized himself with D3 at the beginning of the course. 3) As a **geospatial API**, the Google Maps API is used. Google Maps has a well documented

[1] `http://misoproject.com/dataset/`

and extensive geospatial API. Thanks to Google maps it is easy to render a map with zoom & pan controls, and draw markers on it. Furthermore Google maps also provides support for the calculation of travel distances and the conversion between addresses and coordinates. 4) JavaScript is simplified using **jQuery**. jQuerys tagline is Write less, do more which is certainly correct from our point of view. To write even less, we have also used some jQuery plugins: Tipsy for the tooltips and JQRangeSlider for our slider controls

Opinion of d3.js

After working with d3.js for a semester, we have a well formed idea about d3. We think it is very useful and convenient to work with when you are over the initial learning curve. Certainly for people who are new to Javascript it is a bit of a hurdle. The power of d3 lies primarily in the easy visualisation of data, certainly when this data changes.

In the beginning d3 seemed a bit like a burden because we only wanted to display the circles on a map and this functionality is also offered by Google. The placement of the circles was not so easy in the beginning, also interacting with the map and de circles was cumbersome at first.

Afterwards when we thought that it has a lot of advantages, especially when implementing the different bar charts. The individual location bar charts can be nicely displayed with d3. When you make a distinction between entering data, leaving data and changed data, displaying a changing bar chart becomes very easy.

One of the important advantages of d3 is that it has a good user community, you find a lot of examples for everything you need to do which is always a good starting point.

IMPLEMENTATION ISSUES

Housing department data

One of the issues was incorporating the data we got from the housing department of the KULeuven into our visualization. We were provided with two files, respectively containing addresses and studies or prices. The mapping of these two files was done on latitude and longitude coordinates because this is relatively simple. The google maps API was used for doing this conversion. The problem here was that the coordinates were sometimes not accurate enough, which meant that we lost a lot of the initial data. Because the most significant loss of data was in the file that linked the studies to the addresses, this is less critical to our visualisation, we kept the data the way it was. From the file that had the prices and areas we kept around 42% which are around 1800 residences. Given the time constraint we will not try to improve this but rather try to incorporate the data into our existing application.

Distance calculation

For distance calculation, we first opted to let the user place markers on the map, after which he could specify a travel time. This would result in filtering all residences that lie further away. However, as our dataset keeps increasing (about 400 entries from crowd sourcing, about 4000 from the KULeuven database, to be included in the next version) it becomes too slow to calculate these travel distances at real time. The Google maps API also limits the number of simultaneous distance calculation requests allowed, so our initial idea became infeasible. We therefore opted for a different approach. We selected a number of locations that are important for students (like the campuses of the different fields of study) and pre-calculate the travel time to each of these location for each residence. For this version, 4 important locations were selected: the station, campus Heverlee, gasthuisberg and groupT (all locations frequently visited by students). These location are displayed by markers on the map. The user can select one of these markers and choose the maximal time he wants to travel in order to reach the location. Residences outside this time range are filtered. A button to reset/disable/enable the markers was also added.

Working in group

During the implementation there were also some issues with assigning different jobs to everyone. Because only 2 files are used in our representation of the data, it was difficult in the beginning to work on the visualization independently. We split the task in 2: one group worked on the visualization and on on the user interface. Later when more functions were added this became less of a problem. Most of the functions were independent and only required the data already available in the base visualization or data independent of the rest of the functions, making it easier to work independently of each other.

EVALUATION

When evaluating the visualization, our main focus lies on the *added value* for the intended group of users. However, at the time the evaluation took place (December) , it was difficult to find students that are looking for a new residence. Therefore, during the evaluation we will ask test subjects to impersonate a person searching for a new residence. Furthermore, our evaluation does not assess the *usability* of the developed application (although this is an important aspect to evaluate), because it is not the focus of this project.

TODO: add explanation of face-to-face evaluation approach

Afterwards, we present the test subjects with a questionnaire, which is answered anonymously without interference form the test coordinators. The questionnaire tries to assess the added value of our visualization through following questions:

1. The visualization helped me to choose a suitable student housing,

2. The visualization gives me an insight in the characteristics of the student housings in Leuven,
3. The visualization empowers me to verify hypothesis about the student housings in Leuven,
4. I thinks its fun to use this visualization and to learn more about the student housings in Leuven,
5. I would be willing to pay the following amount to use this visualization to find a student house if the complete list of student housing was available
6. Choose 3 parameters that are of the most important for you when searching for a student housing: price, the other residents, number of other residents, area, distance to class, price over area, accommodations, fellow student of your discipline

TODO: **results**

CONCLUSION

REFERENCES

1. Housing prices in london. `http://www.mysociety.org/2007/more-travel-maps/morehousing`. Accessed: 2013-12-10.
2. Housing prices in slovenia. `http://visual.ly/map-real-estate-prices-slovenia-m2?view=true`. Accessed: 2013-12-10.
3. Where irish-americans live.
4. Datalicious. Melbourne liveability. `http://www.theage.com.au/national/greater-melbourne/our-liveable-city/map`. Accessed: 2013-12-10.
5. Few, S. *Data Visualization for Human Perception*. The Interaction Design Foundation, Aarhus, Denmark, 2013.
6. Lam, H., Bertini, E., Isenberg, P., Plaisant, C., and Carpendale, S. Empirical studies in information visualization: Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on 18*, 9 (2012), 1520–1536.
7. Lau, A., and Vande Moere, A. Towards a model of information aesthetics in information visualization. In *Information Visualization, 2007. IV '07. 11th International Conference* (2007), 87–92.
8. Torre, R. Macrometeorites. `http://visual.ly/macrometeorites?view=true`. Accessed: 2013-12-10.