Hexagonal Gridded Maps and Information Layers: a Novel Approach for the Exploration and Analysis of Retail Data

Evgheni Polisciuc Catarina Maçãs Filipe Assunção Penousal Machado*

CISUC, Department of Informatics Engineering, University of Coimbra, Portugal

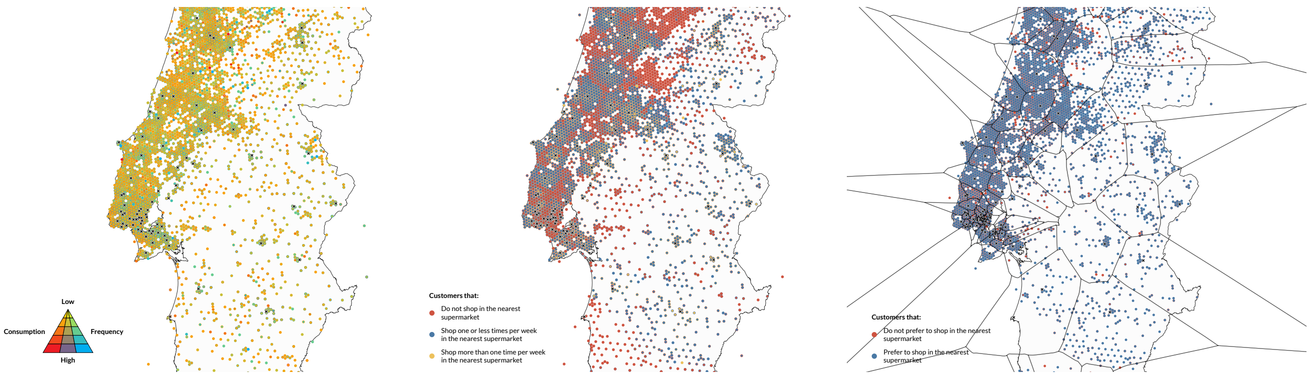

Figure 1: *Diverse combinations of graphical layers along with hexagonal gridded map.*

Abstract

In retail business intelligences, and more specifically in an analysis of customer-supermarket relationship, the factors, such as geographic location of customers, demographic distribution, customers' preferences, accessibility to the store are crucial in decision-making tasks. Visualization is an important tool for analysis and decision-making, which should provide means to make informed business decisions. This article presents a novel approach to the hexagonal gridded maps, which integrates diverse information layers with adaptive zoom. These layers are complementary, providing different points of view over the same dataset and various levels of abstraction. Starting with a dot map, which portrays the impact of supermarket localization on customers choices, up to a choropleth map, which depicts population density in an adaptive form depending on the different granularities of administrative units. Ultimately, the presented visualization provides means to: (i) explore and analyze data regarding customer-supermarket relations; (ii) reveal the impact of supermarkets localization on customer preferences; (iii) suggest areas of low coverage by supermarkets. Additionally, the interplay among the complementary graphical layers provided by the visualization increases its exploratory and analytical power.

Keywords: Geospatial Visualization, Hexagonal Binning, Design Studies, Semantic Zoom, Voronoi Diagram

Concepts: •**Human-centered computing** → **Visualization systems and tools;** *Information visualization;* Visual analytics;

*e-mail: {*evgheni, cmacas, fga, machado*}@dei.uc.pt

SA '16 Symposium on Visualization, December 05-08, 2016, Macao
ISBN: 978-1-4503-4547-7/16/12. . . $15.00

1 Introduction

Visual exploration and graphical analysis tools are essential to reveal benefits from large, complex and diverse data volumes. Nowadays, in the age of big data, the field of information visualization provides valuable tools that enable the users to increase their perception of the information contained in the data. Through visualization, they can discover hidden relationships and patterns within datasets, such as geospatial data, and therefore improve the practices of analysis and decision making.

For retail companies – such as Sonae, one of the biggest retail companies in Portugal – the identification of potential sites to open new supermarkets is an important decision. Factors, such as geographic location, demographic distribution, customers habits, and accessibility, can affect the choice for the potential site. Experts in charge of these decisions need enhance their capabilities with visual tools to interact, explore and analyze the data, in order to find significant patterns, trends and anomalies. In this case, one can rely on visualization techniques to explore and analyze large and complex datasets, and therefore to make properly informed decisions.

We present a novel dot distribution map technique that inherits some of the properties of graduated symbol maps to display georeferenced points in the form of combined circles depicting different data values. Our technique uses an hierarchical hexagonal grid to aggregate the geo-referenced points reducing visual clutter. Additionally, we also employ a Choropleth map, to depict demographic distribution, and a first approach to distorted Voronoi diagram, to define geographically weighted boundaries of supermarkets. These techniques are presented in the form of layers of information, and their combination providing diverse points of view over the dataset. Hence, the visualization presented in this paper provides means to: (i) explore and analyze data regarding customer–supermarket relations in terms of frequency of visits and consumption; (ii) reveal the impact of supermarket locations on customer preferences; (iii) identify areas where the supermarkets coverage is low. Other techniques, such as heatmaps and isopleth maps, and other information about our data, are not included in the scope of

DOI: http://dx.doi.org/10.1145/3002151.3002160

this work, but were previously explored.

This section is followed by the related work, where we give a brief overview of existing works in the field and the used techniques. Then, we describe the application. In the last section we present our findings using the visualization described in the current work.

2 Background and Related Work

This section provides an overview for thematic maps and hexagonal binning. Additionally, we describe zooming and panning techniques that are also of important for the current work.

2.1 Thematic Maps

Thematic maps appeared in the second half of the seventeenth century in the natural sciences and can be defined as the representation of attribute data on a base map. Thematic maps display qualitative or quantitative data. Their purpose is to reveal patterns and frequencies in the geography where they occur [Meirelles 2013], to discover the geographical structure of the data, and to relate its distribution on the map [Robinson 1982]. These maps can be divided in: dot distribution maps [Cable 2013], graduated symbol maps [Young and Bilton 2010], choropleth maps [Brewer and Pickle 2002], isometric and isopleth maps, flow and network maps and area and distance cartograms [Meirelles 2013].

2.1.1 Choropleth Maps

These kind of maps are used to represent data using area symbols. In this technique, each administrative unit is filled with the corresponding color value, saturation or texture, depending on the quantitative data it represents. The color hues are usually used to represent categorical data [Brewer and Pickle 2002]. One classic example of the choropleth map is the map produced by André-Michel Guerry in 1833, which depicts crime in France from 1825 to 1830 [Meirelles 2013]. In modern visualization this techniques are used in the same way with the difference that the maps are generated by computers.

2.1.2 Dot Distribution Maps

Dot distribution maps represent the spatial distribution of geo-referenced discrete data through a basic graphical element – the point [Slocum 2009]. Each point on the map represents either one datum with known geo-location, or the aggregation of values. Furthermore, these maps are used to depict densities in corresponding geographic areas, rather than specific locations.

Dot distribution maps were used as early as 1863 [Hargreaves 1961], but the earliest known use of irregularly spaced dots to depict data was the map produced by Frère de Montizon in 1830 [Meirelles 2013]. A classical example of the use of dot maps is the Disease map, produced by Dr. John Snow [Tufte and Graves-Morris 1983], to depict the distribution of cholera in London.

A more recent example of a dot distribution map is the *Racial Dot Map* by Cable [Cable 2013]. This visualization depicts geographical distribution, population density and racial diversity of people living in the USA. At smaller zoom levels, each dot represents one person, and at national or regional levels each point represents the aggregation of individual dots. The map produced by Ben Fry is another example of the use of dot distribution maps to depict population densities around the world [Fry 2011]. In this work, the size is used in an unusual way – the more dense the region, the smaller the circle size, and vice-versa. Following these ideas, our work uses the dot distribution map technique to depict the distribution of the clients' consuming behaviors.

There are different approaches to tackle the visual clutter commonly generated in this type of visualization. A common approach in popular GIS software is to randomly displace points around their original position or by following a predefined pattern – *jittering* [Hey and Bill 2014]. Other techniques tackle overlapping issue by altering spatial arrangement of dots though means of distortion, and consequently providing additional space for the dots in dense regions [Bak et al. 2009; Keim et al. 2003], or by reducing the number of dots preserving specific characteristics of data – *refinment* [Peters 2013; Yan and Weibel 2008].

2.2 Binning Agregation

Similar to our work there is a technique that addresses the visual clutter by combining adjacent dots in a single graphical element according to the spatial density distribution – *aggregation* [Chua and Vande Moere 2016; Birch et al. 2007]. There are various approaches to aggregate points in geographical space. The work of Slingsby and van Loon uses tiles to aggregate data points, which also enables the representation to show the ratio of different types of points [Slingsby and van Loon 2016]. Additionally, the same work directly links zoom levels to density estimation scale, which provides a coherent transition between zoom layers. Similarly, the work of Zhang et al. looks at various scales of aggregation, providing consistent thematic context while transiting from different zoom levels [Zhang et al. 2016]. Nowadays, it is common to use adaptive zooming techniques [Cecconi and Galanda 2002], this is, the granularity of the representation or the implantation itself changes according to zoom level.

Whilst the majority of the works focus on the use of rectangular grids to subdivide the space and to aggregate points (also known as binning), the are other less popular works that rely on hexagonal binning, i.e., the subdivision of space is done using a hexagonal grid and the data points are aggregated by each hexagon. Hexagons seem to have advantage in what regards spatial coverage efficiency. Their shape is closer to circle than to squares, thus having shorter perimeter, and consequently having more area. Also, the average distance from the center of a hexagon is shorter than of a square. In spite of the previously pointed advantages, it seems that only a small percentage of works in the field use hexagonal grids (2 out of 67 articles [Birch et al. 2007]). For instance, Roth et al. used hexagon binning in their Crime Analysis work [Roth et al. 2015], or hexagon mosaic maps used this technique to display of univariate and bivariate geographical data [Carr et al. 1992]. Ultimately, another common characteristic of these techniques is that they are usually based on hierarchical space subdivision by recurring to Quadtrees [Finkel and Bentley 1974] or similar data structures, starting with a square and recursively subdividing it until a certain condition is met.

3 The Model

This section starts by describing the dataset. Then, we proceed with the detailed description of our approach for hexagonal binning, which is based on dot distribution to depict the geographic distribution of the customers consumption behavior. Along with the adaptive zooming, the map disposes different levels of abstraction. A series of supplementary graphical layers are then presented, allowing the user to explore and analyze the data from different perspectives, such as spatial coverage by supermarkets, population densities with different degrees of granularities, and the potential customers paths to the nearest supermarket.

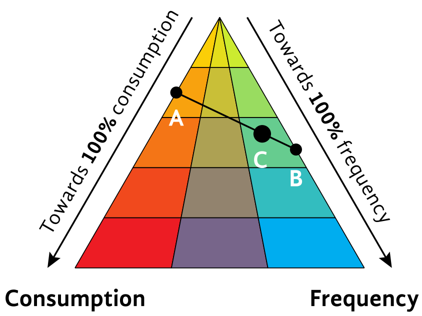

Figure 2: *Discrete triangular color space. Red and cyan colors correspond to the upper bound for consumption and frequency of visits, respectively. Yellow color corresponds to zero consumption and frequency of visits. Other colors are linearly interpolated. Points A, B and C exemplify the projection of values to this color space.*

3.1 Data

The datasets of the supermarket visits and consumption are retrieved from 729 Portuguese supermarkets and hypermarkets of the Sonae chain, which cover the entire country. The consumptions are linked to the client through a client card issued by house hold, which can be shared by the entire family. For this project we used all the transactions made on those supermarkets and hypermarkets from May 2012 to April 2014. Each transaction corresponds to one product bought and it has several properties, such as, spent amount (in euros), date, time, and place of purchase. The data for this project were aggregated by month and were divided in three datasets. The *first* dataset contains the total amounts spent, and the number of visits to the supermarkets for each customer. The *second* dataset contains customer–supermarket pairs with the spent amounts and visits to a particular supermarket. The *third* one is composed by customer–supermarket pairs with the supermarket preference score. The preference score is calculated as follows: $score_{ij} = \frac{F_{ij}+C_{ij}}{D^2}$, where F_{ij} and C_{ij} are the frequency of visits and the spent amount, respectively, by customer i in the supermarket j. D is the number of days in the corresponding period, which, in our case, is one month. In other words, we say that a customer prefers to go to one supermarket if: (i) he spends little, but visits it frequently; or (ii) he visits it once in a while, but spends a lot. We normalize the score by dividing the frequency and consumption by the squared number of days in the corresponding period.

The second and third datasets contain information regarding the nearest supermarket in relation to the customers' residence. The calculation of the nearest supermarket was done in two ways: (i) through geometric distances – the nearest supermarket in straight line; and (ii) computing distances by traversing the road network – the nearest supermarket taking into account the road network topology. A more detailed description is given in section 3.3.1. Finally, the demographic data that consists of population densities by *municipality* and *civil parish* was retrieved from pordata.pt (a Portuguese statistical database).

3.2 Dot Distribution Map of Customers

In our visualizations, we used *combined circles* to represent the residence location of each customer. It is important to note that multiple customers can live in the same location, which is more noticeable in urban areas. To avoid overlapping, the points suffer a small random allocation in space. Our goal was to understand:

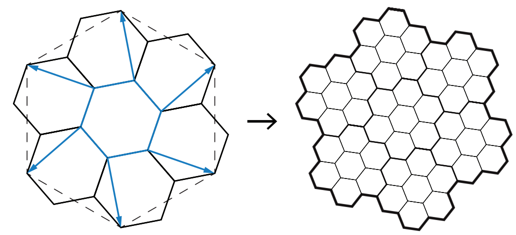

Figure 3: *Illustration of the hexagonal spatial subdivision. Dashed lines show the initial hexagon with seven hexagons that subdivide it (black lines). The blue arrows represent translation vectors. The variation in the line thickness represents different levels of depth.*

(i) the relationship between consumption and frequency of visits; (ii) if the customers consume in the nearest supermarkets and how frequently; (iii) if the customers prefer to visit the nearest supermarkets or not. For that, we converted continuous values in discrete ones by dividing the data in logical groups. This conversion, on the one hand reduces the number of colors to use, simplifying overall visualization, but on the other hand, adds some degree of uncertainty.

3.2.1 Bivariate Color Space

The first dataset contains two variables, the total amount spent in all supermarkets, and the total number of visits to all supermarkets per client during a given period. To encode these values in categories we projected them onto a triangular bivariate color space, where the colors of vertices are predefined and the other ones are interpolated, as illustrated in Figure 2. From the geometric point of view this mapping is done as follows: first, the vector from top vertex to consumption vertex is divided by the normalized consumption value; then, the same is done for the vector towards the frequency vertex; finally, using the ratio of the two values we determine where the point falls between the two vectors. So, let us consider the following example: the consumption and frequency values are 25% and 50%, which are marked with points A and B on the Figure 2, respectively. In this case, C will fall at the point 0.75 on the normalized segment AB, since the ratio of $25/50$ is equal to 0.5, the point will be dislocated 50% from the center of the segment AB towards point B.

The color scheme for the second and third datasets is static. The second dataset is encoded as follows: red represents customers that do not purchase in the nearest supermarket; cyan represents customers that visit the nearest supermarket one or less times per week; and yellow represents customers that visit the nearest supermarket more than once per week. For the third dataset we used cyan and red to depict customers that do and do not prefer consuming in the nearest supermarkets, respectively.

3.2.2 Hexagonal Binning

Hexagonal binning is a process that produces hexagonal grids of variable resolution to subdivide space according to the density of points in geographical space. Similar to quadtrees the hexagonal tree uses a hierarchical structure with hexagons on the leaves and a composition of seven copies at higher levels of the tree. The construction of the hexagonal tree and the insertion of points are more sophisticated when compared to the quadtrees. Since it is impossible to evenly subdivide a hexagon, we employed an algorithm based on the simplified version of Snowflake fractal [Barcellos 1984]. The process proceeds in four successive steps: (i) we

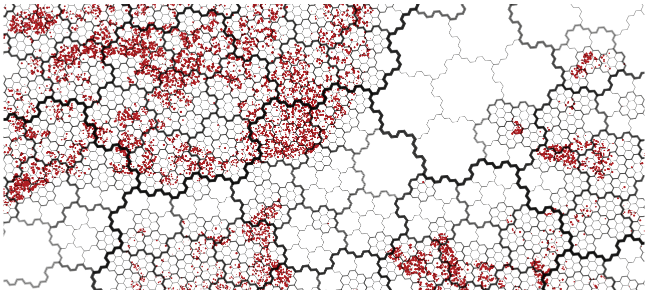

Figure 4: *Spatial subdivision using hexagonal binning. Red dots are geo-referenced points. Line thickness of the boundaries corresponds to the level of depth in the tree.*

start with the hexagon and recursively subdivide it until a certain level of depth; (ii) we compute the geometry of polygons at each level using bottom-up approach; (iii) we insert the points into the tree; (iv) we merge brunches that contains no points or the branches that contains the number of points below a certain threshold.

With that said, the algorithm to construct the hexagonal tree proceeds as follows: start with a hexagon; make seven copies; scale each one by $s = \frac{1}{\sqrt{7}}$; rotate each one by $\alpha = \arcsin(\frac{\sqrt{3}}{2\times\sqrt{7}})$; finally, translate six hexagons so that they exactly surround the central one. The translation vectors for each newly created hexagon, except for the central one, are computed by subtracting appropriate vertex from the corresponding vertex in the original hexagon, as shown in the Figure 3. The process repeats for each newly created hexagon until a certain depth level is achieved (Figure 3). Having generated the tree, the algorithm proceeds with the calculation of the boundaries for each level of depth. We recursively merge polygons coming from $currentlevel - 1$ using a simple union method – the edges that form each polygon are removed if more than one edge coincide. As a consequence, only the edges that form the boundary of the set of polygons remain, as shown in Figure 3.

The insertion of points into the tree proceeds in the same way as in quadtrees – starting from the root and in each level testing a point against the inclusion in the polygons at that level up to the leaves. When all the points are inserted we proceed to the merging stage. The branches that contain no points or those that have the number of points below a certain threshold, are merged. In the end, the space is subdivided as show in Figure 4.

The depth of the tree depends on the minimal allowed size for circles. To be able to distinguish color and size variation of the composite circles a minimal size is predefined. We subdivide the tree until the size of the hexagons at the leaves is close to, but no lower than the minimal predefined size. Then, the size of the composite circles is readjusted according to incircle radius of the hexagons in the leaves.

3.2.3 Composite Circles

The aggregation of points is performed in each cell of the grid and is represented by composite circles. The composite circles, located at the center of each cell, depict the type of points found in the corresponding cell. The number of points of each type is encoded by the area of the corresponding circle, where the maximum allowed size is calculated in the way mentioned earlier. The order in which the circles are drawn depends on the frequency of points of each type – from the inner circle to the outer circle, from high to low numbers, respectively. Finally, we apply a color scheme depending on the dataset being visualized (Figure 5).

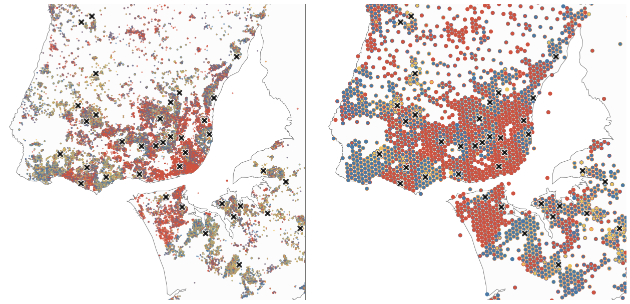

Figure 5: *Comparison of the dot map (left) and our model (right).*

Table 1: *Zoom effect over the graphical error. Zoom level 0 corresponds to the whole country view, and zoom level 5 corresponds to the neighborhood view. Error units are in pixels.*

Zoom Level	Number of Points	Max Error	Mean Error
zoom 0	1785	180.6	6.1
zoom 1	1041	164.2	2.3
zoom 2	542	36.6	0.9
zoom 3	83	17.1	0.4
zoom 4	52	9.59	0.2
zoom 5	45	3.1	0.1

3.2.4 Adaptive Zooming

Adaptive zooming is one of the key features of this work, due to the large amount of represented data. The adaptive zooming enables users to identify areas of interest in higher levels of zoom, and to analyze these areas in more detail at low levels of zoom. Also, due to the restriction for the minimal size of cells, the density estimation can be inaccurate, i.e. the number of points can vary among cells. This leads to some degree of uncertainty in representation. Thus, the map provides different levels of abstraction and granularity depending on the zoom level. In particular, the cells are scaled proportionally to the zoom level, providing different granularities of aggregation. Consequently, the variation in the number of points per cell diminishes as the user zooms in, providing close to optimal representation at the lower levels of zoom (Figure 6).

With that said, adaptive zooming can be seen as a continuous transition from exploratory visualization, where the user searches for areas to analyze, to analytical visualization, where the user analyses the data in close to optimal representation. Regarding the graphical error we considered the following metrics: maximum *number* of points per cell; maximum and mean *displacement* error, which is the distance in pixels from original location to the center of the corresponding cell. For reference, the canvas size defined in our visualization is 2400x1700 pixels. Table 1 summarizes the effect of the zoom over the graphical error.

3.3 Graphical Layers

In this visualization, each layer enriches the model by providing additional graphical tools. The different layers depict different types of information, and have different goals: (i) the identification of territorial domain boundaries and areas with low supermarkets coverage; (ii) the highlight of the customers that reside beyond a certain distance from the nearest supermarket, allowing the identification of areas that are poorly served; and (iii) the representation of the population densities through a Choropleth map, providing the user with additional information about the demographics.

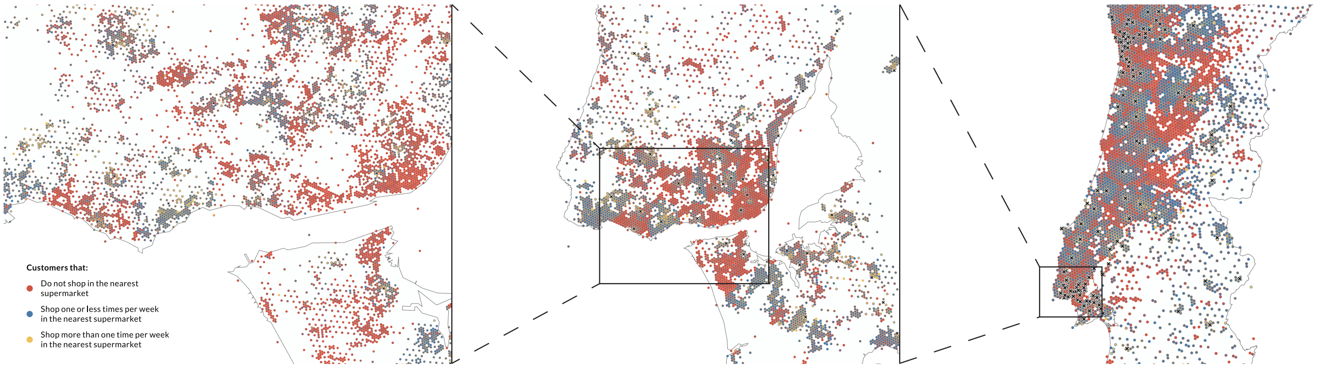

Figure 6: *Illustration of the adaptive zooming behavior. Images from left to right correspond to low, middle and high levels of zoom. The visualization represents the values from the second dataset. Black crosses represent supermarkets.*

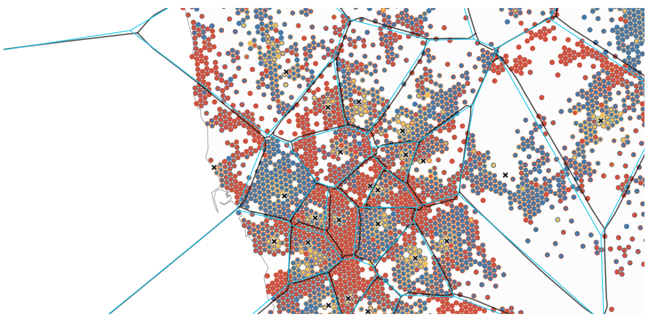

Figure 7: *An illustration of two types of territorial boundaries – geographic (black lines) and geometric (blue lines).*

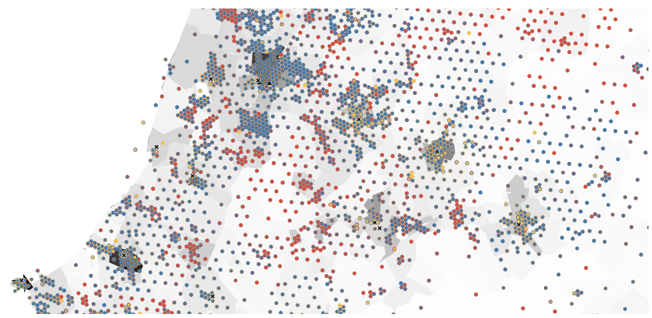

Figure 8: *Second dataset visualization along with the population density by civil parish, using choropleth map technique.*

3.3.1 Geometric and Geographic Boundaries

Aforementioned, the second and third datasets are used to reveal how the customers behave in relation to the nearest supermarket. The nearest supermarkets were calculated using simple euclidean distances, and using the topology of the road network (for the sake of simplicity further referred to as geographic distance). Our study shows that using only the geometric distance is not an optimal choice, due to an associated error. In real life, people are faced with geographical obstacles, such as rivers, mountains, forests, etc. Hence, the nearest supermarket is defined as the one closest to the client, using the road network. Furthermore, the geometric method generates clean and simple forms, which can be beneficial in many cases.

In the case of the euclidean distance, we computed territorial boundaries using the Voronoi diagram technique [Aurenhammer 1991]. By applying this technique with the supermarket sites, we obtained closed areas bounded by the perpendicular bisector of pairs of nearby supermarkets. This bisector separates all the points (customers) closer to one supermarket from the points closer to another. These separators are used to mark territorial boundaries, delimitating the covered areas of each supermarket (Figure 7).

In the case of the geographic distance, we calculated the nearest supermarkets for each client using the road network of Portugal. The road network was retrieved from OpenStreetMap, imported to a PostGIS database, and the topology calculated using the pgRouting extension [Obe 2015]. We found the nearest supermarket for each customer using the Dijkstra algorithm provided by the pgRouting extension. The output of the previous step is a set of customers linked to the corresponding closest supermarket.

To determine the territorial boundaries of each supermarket, we propose a technique for a weighted distortion of the Voronoi boundaries according to geographical properties. Our first approach consists of three successive steps: (i) subdivide edges of the boundaries by smaller segments, with the size that was empirically defined, which depends on the resolution of the image; (ii) each vertex in the diagram is attracted by the points that do not belong to the area being computed, taking into account the geographic proximity between costumer residence and supermarket. So, the attraction force f is computed as: $f = a \times \log\left(\frac{d+1}{b}\right)$, where a and b are the parameters that control the degree and rate of distortion and d is the distance between vertex and the point. The parameters were empirically defined and are equal to 2.6 and 0.3, respectively; (iii) the vertices are pushed back to their original position using attached springs. The push force s is computed as: $s = k * d$, where k is a parameter that controls the original shape, which is equals 0.5 in our case, and d is the distance between the current and original location of the vertex. The force is then applied over a unitary vector pointing towards the original location. We consider the resultant forms as the supermarket territorial boundaries, taking into account the geographical proximity to the customers' residence (Figure 7).

3.3.2 Population Density

The layer below the Dot map is a Choropleth map that depicts demographic data. In particular, we considered population densities by municipalities and civil parishes. Using gray scales, we mapped population densities as follows: the values that are above a certain threshold are represented with black; and the values that are below that threshold are linearly mapped to the shades of gray in the range [0, 95] percent. This mapping ensures that even low values

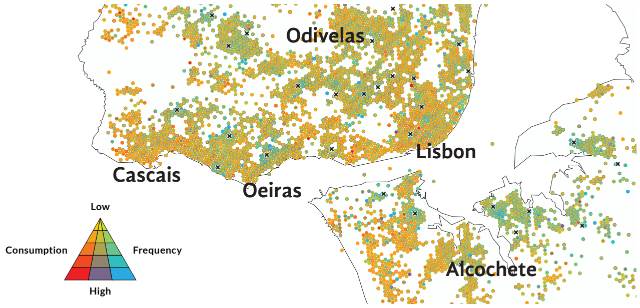

Figure 9: *Close view of Lisbon area. Visualization of the first dataset in a time window of May of 2012.*

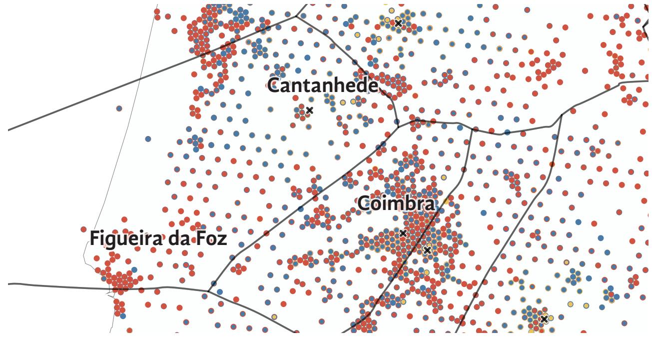

Figure 10: *Coimbra district area. Illustration of the case when customers do not visit the nearest supermarket (May 2012).*

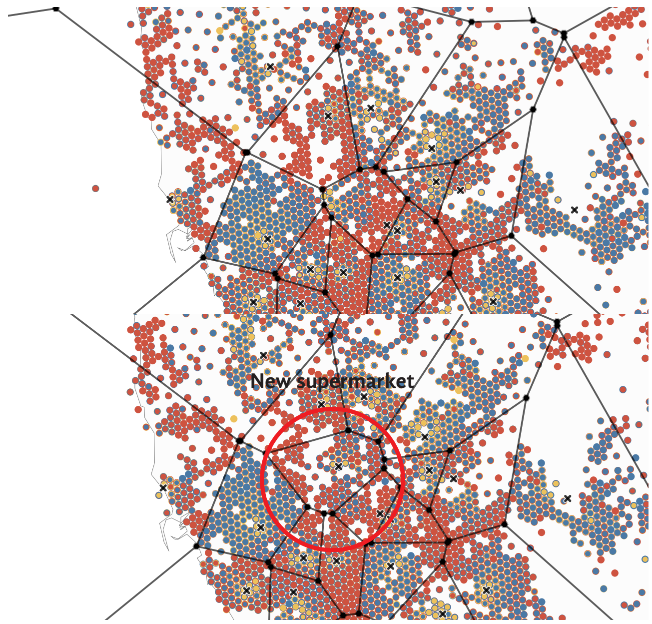

Figure 11: *Visualization of the impact of opening of new supermarket. Second dataset visualization of the time periods May 2012 (image on the top) and June 2012 (image on the bottom).*

are visible on the map (Figure 8). The limitation of this approach is that when the Choropleth map is positioned beneath the Dot map, some of the regions are hidden in the visualization. So, users have to zoom-in to be able to read the map.

4 Findings

In this section we describe four case studies. In particular, the analysis of supermarkets performance in terms of the relationship of customers' consumption and frequency of visits. Then we present one case of site suggestion to open a new supermarket, and also the impact of opening a new supermarket at the suggested area. Additionally, we describe two cases of the impact of the distance to supermarkets on the customers behavior.

4.1 Case study 1

For this case study we focused on the first dataset. Recall that crosses represent supermarket, and color of each dot represents the relationship between frequency of visits and consumption, in accordance with the triangular color scheme described earlier in this paper. The Figure 9 illustrates a close view of the Lisbon area. As can be observed in the selected area there are almost all possible combinations of customer behaviors. There are customers that spend a lot, but visit supermarkets with a low frequency (indicated by the red color). In contrary, there are customers that frequently visit supermarkets, but spent few (indicated by the cyan color). In this case, a clear pattern is perceptible. The customers that reside closer to supermarkets – clusters of cyan points – share this kind of behavior. Also, there is a strong correlation with rich neighborhoods, such as Alcochete, Oeiras, among others. Strangely, Cascais, which is one of the most richest neighborhoods, presents randomness in the customers behavior. In contrast, there are customers, mostly in downtown area of Lisbon, that consume and visit supermarkets at lower rates (indicated by yellowish colors).

4.2 Case study 2

Regarding the potential sites to open new supermarkets let us consider the example of Figure 10. In this case we used the second dataset, which contains information about the frequency of visits. Intuitively users can identify a big cluster of red circles located in Figueira da Foz. The red color indicates that customers do not consume in the nearest supermarket. From the Voronoi diagram one can see that the nearest supermarket is located in Cantanhede. However, it turned out that customers consume mostly in Coimbra city, which is one of the biggest cities in Portugal. Nevertheless, Figueira da Foz seems to be a optimal location to open new supermarket.

As can be observed, this kind of information can also be inferred from the visualization. The the intersection of regions generated by the distorted Voronoi diagram suggests that this is a potential area to open a new supermarket. At least, this is the farthest point from all of the nearest supermarkets. Ultimately, we discovered that, in Figueira da Foz a new supermarket has been opened in October 2014. It is important to note that the available data do not cover this period. So, we were unable to visualize the impact of a new store opening. In the following subsection we consider one of such cases.

4.3 Case study 3

For this case study we used the second dataset, since it better reflects the customers' reaction to the opening of new supermarkets. Figure 11 illustrates the impact of the opening of a new supermarket in the Porto city. As can be observed in the image on the top (Fig. 11) there is a cluster of red points that stands out from the map. That means that there are customers that for some reason did not consume in the nearest supermarkets during the May of 2012. So, in the end of June 2012 Sonae opened a new supermarket precisely in the area indicated by the diagram. The resulting effect

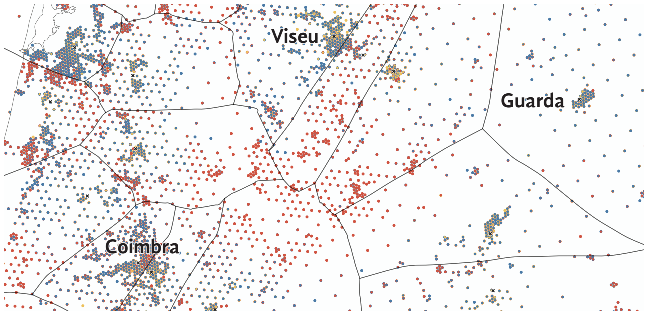

Figure 12: *Visualization of the negative impact of distance.*

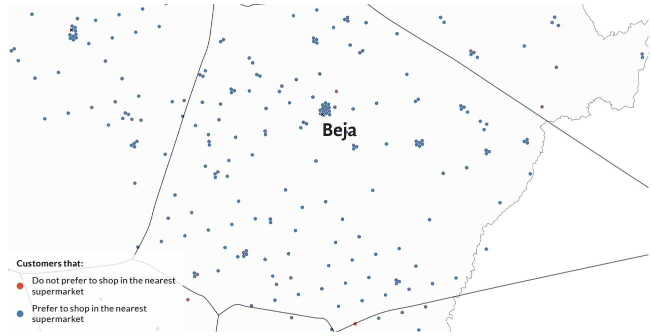

Figure 13: *Visualization of the impact of distance.*

of the customers behavior is illustrated in the image on the bottom (Fig. 11). The remaining red circles may have to do with the accessibility to the supermarket. Nevertheless, there is a large number of points that turned out cyan, which is the indicator that these customers visit the supermarket one or less times per week.

Users can notice the following typical pattern – the customers that reside close to supermarkets are colored with yellow, indicating that these are frequent customers; on the contrary, the customers residing further from supermarket visit them less frequently, which is indicated by the cyan color. So, this pattern did not applied to the newly opened supermarket. Mainly, this is due to the fact that the supermarket opened at the end of the month, and it is natural that during that time period, on average, people have not been able to visit the supermarket more than one time per week. However, it is verified that, in the following month, the pattern changes to a typical one.

4.4 Case study 4

For this case study we used the third dataset, which allow us to understand the customers' preferences. One of such cases is presented in the Figure 12. This image captures the customers that reside near the intersection of three districts – Coimbra, Viseu and Guarda. These areas are mostly covered with mountains. Nevertheless, there seems to live a significant number of customers. As it can be observed, these clients are colored in red, which means that they do not prefer to shop in the nearest supermarket. Ultimately, this can be verified by looking at the Voronoi diagram. As turned out, Sonae has recently opened a new supermarket in that area. Again, we were unable to visualize the impact of the new store opening, since our data do not cover this period.

In some cases, big distances do not have impact over the customers preferences. For instance, consider the example of the Figure 13. This is the south part of Portugal, and most of it is desertified, and it is flat. As can be seen, there live a considerable number of customers and they are spread around a small city, where is a supermarket. Also, it can be seen that the majority of these people reside far away from that supermarket, which is the nearest one. As we saw previously, customers that live that far way, in general prefer not to shop in the nearest supermarkets, and logically they probably shop in the supermarkets of concurrent companies. However, the cyan color indicates that the customers from this area prefer consuming in that specific supermarket.

5 Conclusion

In this paper we presented a unified visualization, which provides different perspectives over the same data. The goal of this visualization is to provide the user with sufficient visual tools to learn from data and to make a more informed decisions. This work started by retrieving data regarding the consumptions in the supermarkets of the biggest retailing company in Portugal. In order to get valuable information the data were aggregated by month, and spent amounts and frequency of visits of each customer to each supermarket was computed. This data were separated in three datasets: (i) total consumption and number of visits; (ii) frequency of visits to the nearest supermarket; (iii) customers' preference in regard to the nearest supermarket. Additionally, the nearest supermarkets were calculated using Euclidean distance and the distance traversing a road network.

All the three datasets are visualized using our approach, where each composite circle represents data using hexagonal binning. The color is used to encode different data values, through the use of the same primary colors. Moreover, the methods to encode each dataset are distinct. In the case of the first dataset values are projected onto the bivariate triangular color scheme, converting continuous values into discrete ones. The second dataset is encoded using three colors, and the third one uses only two colors. Apart from that, we described different graphical layers that enable users to get additional insights, such as the types of consumption behaviors, identification of poorly served areas, impact of distance, among others.

In regard to the visual clarity and efficiency of the visualization we presented a overlap free method that reduces visual clutter and enables the user to easily identify areas of interest. This method relies on adaptive zooming, which consists of different levels of abstraction that facilitate the reading of the map, despite adding some degree of graphical error. Additionally, the presented method provides continuous zooming without abrupt transitions between different levels of abstraction.

The case studies exemplify the usage of the presented visualization in real situation and the kinds of findings that can be discovered and analyzed. As was seen, this visualization helps to understand the relationship between consumption and frequency of visits, and reveals the patterns of consumption behaviors that are closely related with demography. Also, this visualization helps to analyze if the customers consume in the nearest supermarkets and how frequently, or if the customers prefer consuming in the nearest supermarkets or not. Additionally, this visualization enables the user to analyze the impact of opening a new supermarket and the distance from nearby stores. However, the most important point is that this visualization helps users to make a more informative decision of where it is advisable to open new supermarkets.

Acknowledgements

This project is partially funded by Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grant SFRH/BD/109745/2015.

References

AURENHAMMER, F. 1991. Voronoi diagrams - A survey of a fundamental geometric data structure. *ACM Comput. Surv. 23*, 3, 345–405.

BAK, P., SCHAEFER, M., STOFFEL, A., KEIM, D. A., AND OMER, I. 2009. Density equalizing distortion of large geographic point sets. *Cartography and Geographic Information Science 36*, 3, 237–250.

BARCELLOS, A. 1984. The fractal geometry of mandelbrot. *College Mathematics Journal 15*, 2, 98–114.

BERTIN, J. 2011. *Semiology of Graphics: Diagrams, Networks, Maps*. Esri Press, Redlands.

BIRCH, C. P., OOM, S. P., AND BEECHAM, J. A. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling 206*, 34, 347 – 359.

BREWER, C. A., AND PICKLE, L. 2002. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers 92*, 4, 662–681.

CABLE, D. 2013. Racial dot map. *Weldon Cooper Center for Public Service, University of Virginia*.

CARR, D. B., OLSEN, A. R., AND WHITE, D. 1992. Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartography and Geographic Information Systems 19*, 4, 228–236.

CECCONI, A., AND GALANDA, M. 2002. Adaptive zooming in web cartography. In *Computer Graphics Forum*, vol. 21, Wiley Online Library, 787–799.

CHUA, A., AND VANDE MOERE, A. 2016. Binsq: visualizing geographic dot density patterns with gridded maps. *Cartography and Geographic Information Science*, 1–20.

FINKEL, R. A., AND BENTLEY, J. L. 1974. Quad trees: A data structure for retrieval on composite keys. *Acta Inf. 4*, 1–9.

FRY, B., 2011. Dencity. https://fathom.info/notebook/1981/. Accessed: 2016-03-27.

HARGREAVES, R. P. 1961. The first use of the dot technique in cartography. *The Professional Geographer 13*, 5, 37–39.

HARROWER, M., AND SHEESLEY, B. 2005. Designing better map interfaces: A framework for panning and zooming. *Trans. GIS 9*, 2, 77–89.

HEY, A., AND BILL, R. 2014. Placing dots in dot maps. *Int. J. Geogr. Inf. Sci. 28*, 12 (Dec.), 2417–2434.

KEIM, D. A., PANSE, C., SIPS, M., AND NORTH, S. C. 2003. Pixelmaps: A new visual data mining approach for analyzing large spatial data sets. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), 19-22 December 2003, Melbourne, Florida, USA*, 565–568.

MACEACHREN, A. M., AND DIBIASE, D. 1991. Animated maps of aggregate data: Conceptual and practical problems. *Cartography and Geographic Information Systems 18*, 4, 221–229.

MEIRELLES, I. 2013. *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Rockport publishers.

OBE, R. 2015. *PostGIS in action*. Manning Publications, Shelter Island, NY.

PETERS, S. 2013. Quadtree- and octree-based approach for point data selection in 2d or 3d. *Annals of GIS 19*, 1, 37–44.

POLISCIUC, E., ALVES, A., AND MACHADO, P. 2015. Understanding urban land use through the visualization of points of interest. In *Proceedings of the Fourth Workshop on Vision and Language*, Association for Computational Linguistics, Lisbon, Portugal, 51–59.

REITERER, H., AND BÜRING, T. 2009. Zooming techniques. In *Encyclopedia of Database Systems*. 3684–3689.

ROBINSON, A. 1982. *Early thematic mapping in the history of cartography*. University of Chicago Press, Chicago.

ROTH, R. E., ROSS, K., AND MACEACHREN, A. M. 2015. User-centered design for interactive maps: A case study in crime analysis. *ISPRS Int. J. Geo-Information 4*, 1, 262–301.

SLINGSBY, A., AND VAN LOON, E. E. 2016. Exploratory visual analysis for animal movement ecology. *Comput. Graph. Forum 35*, 3, 471–480.

SLOCUM, T. A. 2009. *Thematic cartography and geovisualization*. Prentice hall.

TUFTE, E. R., AND GRAVES-MORRIS, P. 1983. *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT.

TUFTE, E. R. 2006. Beautiful evidence. *New York*.

YAN, H., AND WEIBEL, R. 2008. An algorithm for point cluster generalization based on the voronoi diagram. *Computers & Geosciences 34*, 8, 939–954.

YOUNG, M., AND BILTON, N. 2010. A day in the life of the new york times. In *Beautiful Visualization: Looking at Data Through the Eyes of Experts*, J. Steele and N. Iliinsky, Eds. O'Reilly Media, Inc., ch. 16, 271–290.

ZHANG, J., AHLBRAND, B., MALIK, A., CHAE, J., MIN, Z., KO, S., AND EBERT, D. S. 2016. A visual analytics framework for microblog data analysis at multiple scales of aggregation. *Comput. Graph. Forum 35*, 3, 441–450.