**Course:** Information Visualization

**Lecturer:** Peter Bak

**Department**: Industrial Engineering & Management  
**Submission Date:** 19.9.2015

Correlation between Education Levels and Real Estate Prices in Israel

Shanee Honig & Gal Shukry

Ben-Gurion University of the Negev

Table of Contents

[Introduction 3](#_Toc429491471)

[Users 4](#_Toc429491472)

[Task 5](#_Toc429491473)

[Data Acquisition and Understanding 5](#_Toc429491474)

[Visualization 7](#_Toc429491475)

[Iteration 1: Scatterplot 7](#_Toc429491476)

[Iteration 2: Treemap 12](#_Toc429491479)

[Iteration 3: Treemap - Labels + Decluttering 15](#_Toc429491482)

[Iteration 4: Treemap - Color Palette + Normalizing Data 18](#_Toc429491485)

[Iteration 5: Treemap - Adding Filters, Legend, & Title 22](#_Toc429491489)

[Discussion 24](#_Toc429491490)

[Value of visualization 24](#_Toc429491491)

[Self Reflection 25](#_Toc429491497)

[Conclusion 27](#_Toc429491498)

[Appendices 28](#_Toc429491499)

[Appendix 1: Original Dataset 28](#_Toc429491500)

[Appendix 2: Data Analysis 29](#_Toc429491501)

[Appendix 3: Storyboard of visualization process 30](#_Toc429491502)

[References 36](#_Toc429491511)

# Introduction

Across communities in Israel, there is a positive correlation between educational attainment and income, and between income and real estate prices (Ben David, 2014). This suggests a possible positive correlation between educational attainment and real estate prices. Research supports this relationship, demonstrating that housing prices in neighborhoods rise when they attract educated owners (Bayer, Ferreira & McMillan, 2007; Kane, Riegg & Staiger, 2006).

One possible explanation for the relationship is that as a person’s educational level increases, so does their income, allowing them to purchase houses with higher real estate value. Neighborhoods with a negative correlation between educational attainment and real estate value can then potentially be used to indicate various social and demographic issues that need to be addressed. For example, a neighborhood in which the educational attainment rate is low but the housing prices are high indicates social imbalance which can lead to higher crime rates and civil unrest (Bursik, 1988). Consequently, understanding the relationship between educational attainment and real estate prices in different geographical regions across Israel is an important task.

In this paper, we aim to provide an interface that is robust and insightful enough to support such analysis. Because Israel has 76 city municipalities, each with multiple neighborhoods, identifying trends and meaningful patterns is nearly impossible to do using the raw data alone. For this reason, we use *“The Grammer of Graphics”* (Wilkenson, 2006) to display the relationship between educational attainment and real estate value across different cities and neighbourhoods in Israel in a way that enables its users to quickly detect and extract meaningful insights. We begin by examining the original data used for the visualization, the users and the tasks the visualization is targeted for. We then discuss our methodology in detail using images and visual mapping summaries. We conclude by presenting and assessing our final visualization of the data.

# Users

The relationship between educational attainment and real estate prices in different geographical regions in Israel can be useful for a variety of users:

1. **Government, Municipalities, and Police Force:** a city or neighborhood in which the educational attainment rate is low but the housing prices are high indicates social imbalance which can lead to higher crime rates and civil unrest (Bursik, 1988). Politicians and police force could use this information in order to target unbalanced cities and neighborhoods and begin corrective actions, like building low-income housing in those areas and allocating crime fighting efforts more effectively.
2. **Apartment Seekers:** Israel has one of the highest real estate prices in the world relative to resident income (Shafer, 2014). This makes searching for an affordable apartment an extremely difficult task for many young couples and families. When educated people move into non-expensive neighborhoods, it often raises the value of real estate in that neighborhood (Bayer, Ferreira & McMillan, 2007). Visualizing the relationship between educational attainment and real estate values could thus provide an efficient way for both regular homebuyers and real estate investors to quickly seek out affordable, up-and-coming neighborhoods worth investing in.
3. **Educators:** Many studies have been done on the relationship between the educational attainment of parents and the educational attainment of their children. These studies have shown that there is a direct relationship between parent’s educational attainment and their children’s success in school (Behrman, 1997). Children with parents who did not received an undergraduate education have a much lower chance of receiving a degree themselves than children with parents who did receive an undergraduate education (Horn & Bobbitt, 2000; Nunez & Cuccaro-Alamin, 1998). This relationship holds true even when you take into consideration additional related variables, like family income and family structure (single-parent or two-parent families) (Berkner & Chavez, 1997). Moreover, there is a correlation between low real estate prices and the quality of education available in the specified region (Chiodo, Hernández-Murillo & Owyang, 2010; Li, 2012). Educators could therefor use the relationship between educational attainment and real estate prices in order to better allocate resources for special education programs to areas in which they’re needed the most.

# Task

Summarizing the needs of all the various potential users listed above, it is clear that their primary tasks are to (1) compare apartment prices and educational attainment rates between cities and between neighborhoods across Israel and (2) identify outliers, that is, neighborhoods or cities with:

* high apartment prices
* low apartment prices
* low education attainment rates
* high education attainment rates
* high apartment prices and low educational attainment rates
* low apartment prices and high educational attainment rates

Values that are considered “high” and “low” are likely to change depending on the user of the visualization and their task. For example, educators searching for neighborhoods with low educational attainment rates in order to target them for special education programs might have a lower threshold than policemen using the information in order to allocate police force more effectively. For this reason, it is fundamental that the visualization does not bias the user towards any particular conclusion, but rather provides all the information necessary to the users in order to reach an educated conclusion.

# Data Acquisition and Understanding

The data used to visualize the relationship between educational attainment and real estate prices in different geographical regions in Israel was collected from Israel’s Central Bureau of Statistics website (<http://www.cbs.gov.il/reader>). The data itself was collected by the Israeli government in 2012, and published online in 2015.

Israel contains 14 major cities with populations over 100,000. Each of these cities is listed in the dataset, and divided further into their respective statistical areas. There is a grand total of 943 statistical areas in the dataset, in accordance with the geographical division done by Bureau of Statistics of Israel.

For almost all statistical areas in the dataset, the following information is provided: (1) the average apartment value in Shekels for one square meter, and (2) the percentage of residents ages 20-55 with an undergraduate degree. The first page of the dataset used can be viewed in [Appendix 1](#_Appendix_1:_Original)[[1]](#footnote-1). The average apartment value in Shekels for one square meter will be used as a measure of apartment prices, and the percentage of residents aged 20-55 with an undergraduate degree will be used as a measure of educational attainment rate. Statistical areas that did not contain both data points were removed from the dataset, since both data points are required in order to answer our primary user tasks.

In order to better understand the data obtained, for each city we calculated the mean and standard deviation of the percentage of people with undergraduate degree and the average apartment value (see [Appendix 2](#_Appendix_2:_Data)). A number of interesting insights can be made from this initial analysis. First, it is evident that Tel Aviv, Ramat Gan, and Rehovot have the highest average percentage of residents with a bachelor’s degree, whereas Bene Beraq and Bat Yam have the lowest. Bene Beraq also has the smallest variation in educational attainment, whereas the largest variation is located in Tel Aviv, Be’er Sheva, and Netanya. In terms of average apartment value, Tel Aviv, Jerusalem, and Ramat Gan are quite high (as expected), whereas Be’er Sheva and Ashquelon are extremely low. The variation in average apartment value is very difficult to detect from the raw data without additional calculations since the values of both the standard deviation and the mean are quite high.

Since both the average apartment value and the percentage of residents with an undergraduate degree are consequences of social phenomena, we assume them both to be normally distributed.

# Visualization

In this section, we describe the methodology used to visualize the relationship between educational attainment and real estate prices in Israel. We initially wanted to map the relationship geographically since the data is inherently spatial, however, the data set did not come with (x,y) coordinates, and we could not find any coordinates online representing statistical areas. For this reason, we decided to explore the use of non-geographic visualizations.

## Iteration 1: Scatterplot

In order to obtain a general understanding of the relationship between the two variables and see if there are any significant outliers, we plotted each statistical area on a scatterplot based on the percentage of residents with an undergraduate education (X-axis) and average apartment value per meter squared (y-axis).

We chose to begin with a scatterplot because scatterplots can be used to display trends in the relationship between two continuous variables, highlight outliers and identify minimum or maximum values. These features could potentially address a significant portion of our user tasks. We implemented the scatterplot in R-Studio in order to find out.

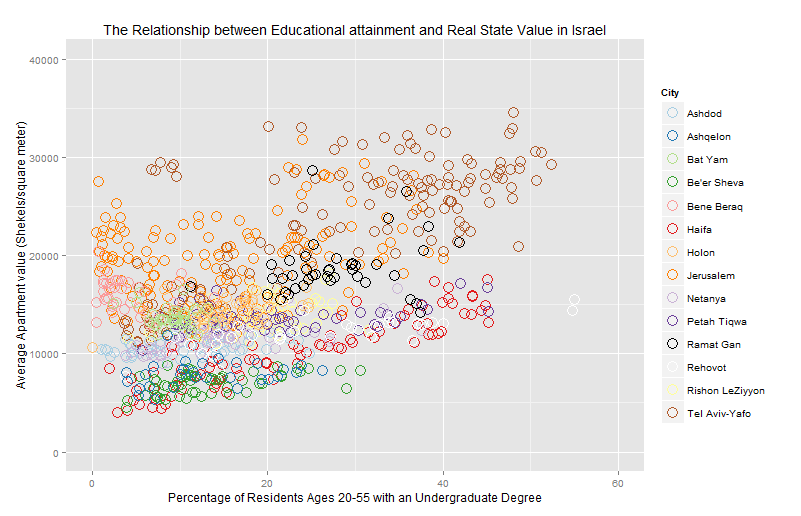
We decided to map statistical areas to location (X,Y), where X represents the percentage of adults (ages 22-50) with Bachelor’s degree, and Y represents the average apartment value per squared meter. We started both axes at 0 in order to ensure that differences in location between statistical areas are easily understood and correctly interpreted. Since the data set is large, we anticipated a significant amount of overlap between points, and decided to indicate neighborhoods using circle stroke marks with no fill in order to maximize visibility.

The color of the stroke signifies the city to which the neighborhood belongs to. The color palette is categorical (qualitative) and was taken from colorBrewer[[2]](#footnote-2). Since the maximum number of data classes colorBrewer supports is 12 but there are 14 cities that need to be identified, we manually added black and white to the color palette. A legend was then added to allow the users to identify the meaning of each color. The visual mapping for iteration 1 is summarized in Table 1. The resulting visualization can be seen in Figure 1.

*Table 1: Scatterplot - Visual Mapping*

|  |  |  |
| --- | --- | --- |
| **Data Attribute** | **Visual Attribute** | **Additional Information** |
| Neighborhood | Hollow Circle Mark |  |
| City | Stroke Color |  |
| Apartment Value | Y-value location | Low apartment value = low Y value  High apartment value = high Y value |
| Percentage of Residents with Undergraduate education | X-value location | High educational attainment rate = high X value Low educational attainment rate = low X value. |

*Figure 1: Scatterplot*



### Insights

From this scatterplot, we can identify a linear relationship between the average apartment value and the percentage of residents with an undergraduate degree. Moreover, there is a positive correlation between the two variables: neighborhoods with a high percentage of people with an undergraduate degree tend to also have a high average apartment value. Neighborhoods with a lower percentage of people with an undergraduate degree tend to also have a lower average apartment value. This is true in all cities. A significant overlap in data points at the bottom left corner of the graph indicates that many neighborhoods have an average apartment value between 5000-15000 shekels/square meter with 7-20% of residents with an undergraduate education. There are two outliers in Rehovot with relatively low apartment values and high education rates.

|  |  |
| --- | --- |
| **Neighborhoods** | |
| **Task** | **Visualization insights** |
| Identifying neighborhoods with high apartment value | Found fairly easily by looking at the highest points in the graph, but the neighborhoods cannot be identified by name. |
| Identifying neighborhoods with low apartment value | Found fairly easily by looking at the lowest points in the graph, but the neighborhoods cannot be identified by name. Since many neighborhoods fall in this region, some are hidden by multiple layers of data points. |
| Identifying Neighborhoods with high education attainment rates | Found fairly easily by looking at the points that are furthest to the right in the graph, but the neighborhoods cannot be identified by name. |
| Identifying Neighborhoods with low education attainment rates: | Found fairly easily by looking at the points that are furthest to the left in the graph, but the neighborhoods cannot be identified by name. Since many neighborhoods fall in this region, some are hidden by multiple layers of data points. |
| Identifying neighborhoods with high education attainment rates and low apartment value | Found fairly easily by looking for points that are closest to the top right corner of the graph, but these neighborhoods cannot be identified by name. |
| Identifying neighborhoods with low education attainment rates and high apartment value | Found fairly easily by looking for points that are closest to the bottom left corner of the graph, but these neighborhoods cannot be identified by name. |
| **Cities** | |
| Identifying cities with high apartment value | Tel Aviv and Holon seem to have the highest overall apartment values. Finding them was quite easy by finding colors that are consistently at the top of the graph. |
| Identifying cities with low apartment value | Be’er Sheva, Ashkelon and Haifa seem to have the lowest average apartment values. This was more difficult than finding cities with high apartment values since there was high density in the data points in this region. Identifying Ashkelon (blue) for instance was extremely hard since it was mostly masked by Be’er Sheva (green) and Haifa (red) points. |
| Identifying cities with high education attainment rates | Tel Aviv has the most neighborhoods with high education rates. Haifa has several as well, however Haifa has a much higher variance than Tel Aviv. |
| Identifying cities with low education attainment rates: | Bat Yam and Jerusalem appear to have low education attainment rates, although it is very hard to spot due to the large amount of visual clutter in the bottom left region of the visualization |
| Identifying cities with high education attainment rates and low apartment value | Haifa seems to have the most neighborhoods with high education rates and relatively low average apartment values |
| Identifying cities with low education attainment rates and high apartment value | Jerusalem seems to have the most neighborhoods that meet this criteria more than other cities |

### Problems with Visualization

In this visualization, we are not able to identify any specific neighborhoods - we cannot match a particular circle to the neighborhood it represents. Such an understanding is important so that the users will be able to make actionable decisions based on the visualization. Adding labels for each neighborhood might solve this problem but will result in a significant amount of clutter and will be hard to implement in a way that ensures each label is both legible and easily attributed to the correct circle.

Secondly, when many neighborhoods share similar (X, Y) values, their representations are placed one on top of another in a way that adds significant clutter makes comparing between neighborhoods and cities nearly impossible. Comparing between specific neighborhoods and cities is a fundamental task for all users. Adding transparency might help detect additional points, however it would warp the distinction of the colors and thus hurt our ability to detect city-wide trends.

Third, the relationship between neighborhoods and the cities they are in is detectable but not intuitively or easily understood through this visualization, since color is not a hierarchical attribute. Having such an understanding could drastically improve the speed in which users interpret the data.

Finally, despite using a categorical color palette from colorBrewer that is designed to be easily distinguishable, it is hard to distinguish between several colors, for example between Jerusalem and Tel Aviv. Moreover, the white and yellow markings are barely visible above the light grey background of the plot.

## Iteration 2: Treemap

Since the scatterplot was not suited for identifying and comparing specific neighborhoods, we decided to switch visualization types. Many of the scatterplot’s disadvantages resulted from the fact that the data is hierarchical, but the visualization does not support this internal hierarchy. Treemaps, on the other hand, are designed for hierarchical data.

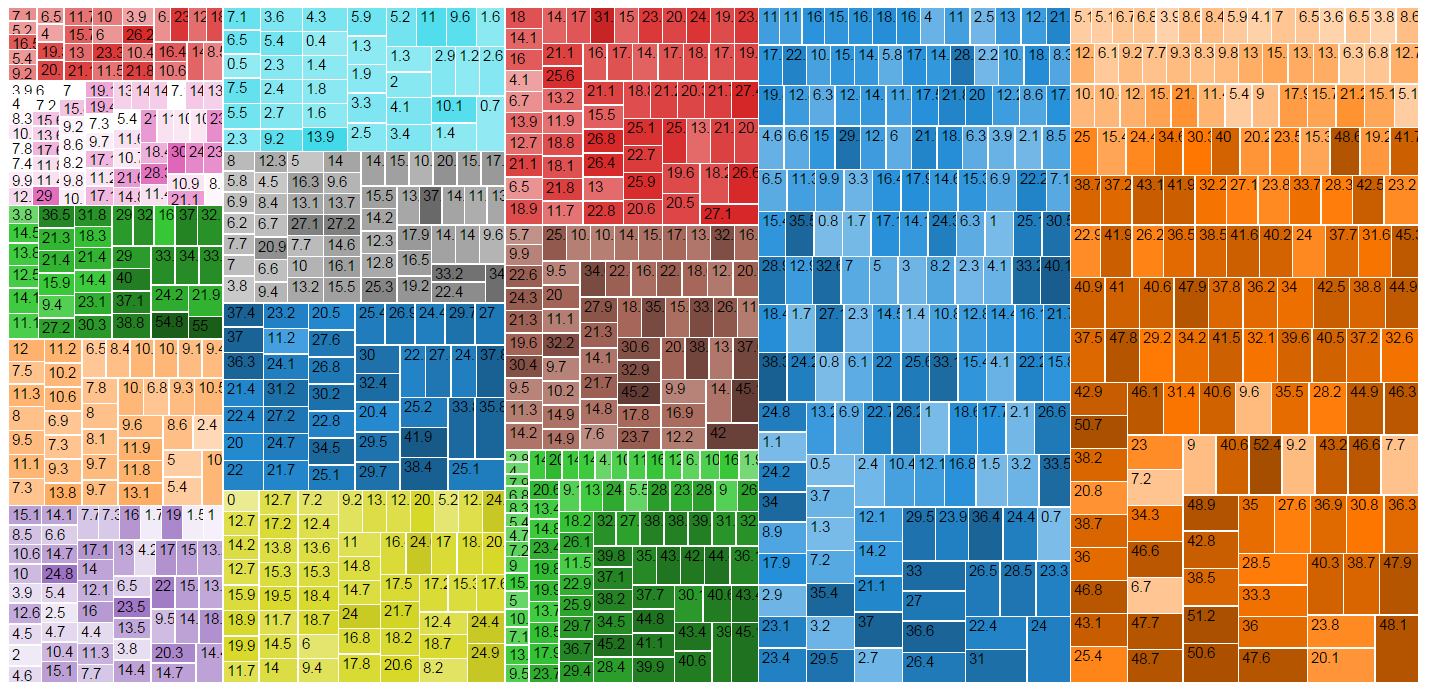
Treemaps display hierarchical data using nested rectangles. Each branch of the tree is given a rectangle, which is then filled with smaller rectangles representing sub-branches. Sub-branch rectangles have an area proportional to a specified dimension of the data, and are often colored to represent an additional dimension. When the color and size dimensions are correlated in some way, one can often easily see patterns that would be difficult to spot in other ways. In addition, by construction, they can legibly display thousands of items on the screen simultaneously – fixing the scatterplot’s problem of not being able to view all the statistical regions at once and decreasing clutter.

Since we have a large number of data points, our data is hierarchical and we wish to map two primary dimensions, we believe the treemap could provide a good visualization solution to support our user tasks. We implemented the treemap using D3. The visual mapping for iteration 2 is summarized in Table 2. The resulting visualization can be seen in Figure 2.

*Table 2: Treemap – Visual Mapping*

|  |  |  |
| --- | --- | --- |
| **Data Attribute** | **Visual Attribute** | **Additional Information** |
| Neighborhood | Rectangle mark |  |
| City | Color |  |
| Apartment Value | Size | High apartment value = large rectangle  Low apartment value = small rectangle |
| Percentage of Residents with Undergraduate education | Brightness + Exact Value displayed on each rectangle. | High percentage = low brightness value  Low percentage = high brightness value  We decided to add the exact value in order to make it easier for the user to compare between percentages of two non-adjacent neighborhoods. |

*Figure 2: Treemap*



### Insights

As expected, the city-neighborhood hierarchy is clearly visible in this visualization. It is easy to distinguish between the different cities, and each neighborhood is represented legibly on the visualization. The color palette selected is a qualitative color palette selected from colorBrewer. By mapping brightness to educational attainment rate, it is relatively easy to see that the cities represented by yellow and red are fairly homogeneous (low variance) whereas the cities represented by blue and brown have more variance in educational attainment. This insight was much harder to detect in the scatterplot due to clutter. Moreover, it is quite easy to see that in several cities, like orange and green, there is a linear positive relationship between education and apartment value since the smaller rectangles tend to be bright and larger rectangles tend to be dark.

Differences in apartment value, represented by rectangle size, are harder to see, since the differences in area between neighborhoods are fairly subtle. However, it is pretty clear that the city represented in red at the top left corner of the visualization and the city represented by pink beneath it have the smallest average apartment value, whereas the cities in orange and blue have the highest values. Within-city comparison is slightly easier, although not by much. For example, you can see that neighborhoods within the city represented by blue that are located at the top have smaller apartment values than the neighborhoods located at the bottom. However, it is hard to compare the neighborhoods with smaller apartment value in the blue city and the neighborhoods with the neighborhoods with smaller apartment values in the orange city.

|  |  |
| --- | --- |
| **Neighborhoods** | |
| **Task** | **Visualization Performance** |
| Identifying neighborhoods with high apartment value | Sometimes difficult to do since the rectangle sizes seem very similar between neighborhoods and there is a lot of clutter. However, organizing neighborhoods within each city by size helps. Individual neighborhoods still can’t be identified by name. |
| Identifying neighborhoods with low apartment value |
| Identifying Neighborhoods with high education attainment rates | Relatively easy. However, the values of education attainment rate displayed in each neighborhood is not always legible and adds significant amount of clutter to the visualization. Individual neighborhoods still can’t be identified. The mapping of high education rates to low brightness is a bit confusing. |
| Identifying Neighborhoods with low education attainment rates: |
| Identifying neighborhoods with high education attainment rates and low apartment value | Finding small, dark rectangles in the visualization is quite easy. However, comparing between neighborhoods from different cities is hard since the sizes are very homogeneous. |
| Identifying neighborhoods with low education attainment rates and high apartment value | Finding bright, large rectangles in the visualization is also relatively easy. |
| **Cities** | |
| Identifying cities with high apartment value | The cities in blue and orange overall have high average apartment values, whereas the cities in pink and red (the small one) have low average apartment values. Individual cities can’t be identified by name. |
| Identifying cities with low apartment value |
| Identifying cities with high education attainment rates | You can easily see that the city in pink has relatively low education attainment rates, whereas the city in green has a lot of regions with high education attainment rates. Individual cities can’t be identified by name |
| Identifying cities with low education attainment rates |
| Identifying cities with high education attainment rates and low apartment value | Difficult to do due to the large amount of visual clutter. |
| Identifying cities with low education attainment rates and high apartment value |

### Problems with Visualization

This visualization still doesn’t identify individual neighborhoods or individual cities, which is extremely important to the end user. A legend or labels need to be added. In addition, it is still relatively hard to compare apartment values of different neighborhoods and different cities, since the variation in rectangle sizes is quite small. The educational attainment rates are not legible and they add a lot of clutter to the visualization. There aren’t enough colors in the color palette so it appears as if multiple cities might be connected or related somehow, even though they are not. It is possible to identify some patterns between educational attainment rate and average apartment value across different neighborhoods, although they are not immediately evident due to the large amount of clutter.

## Iteration 3: Treemap - Labels + Decluttering

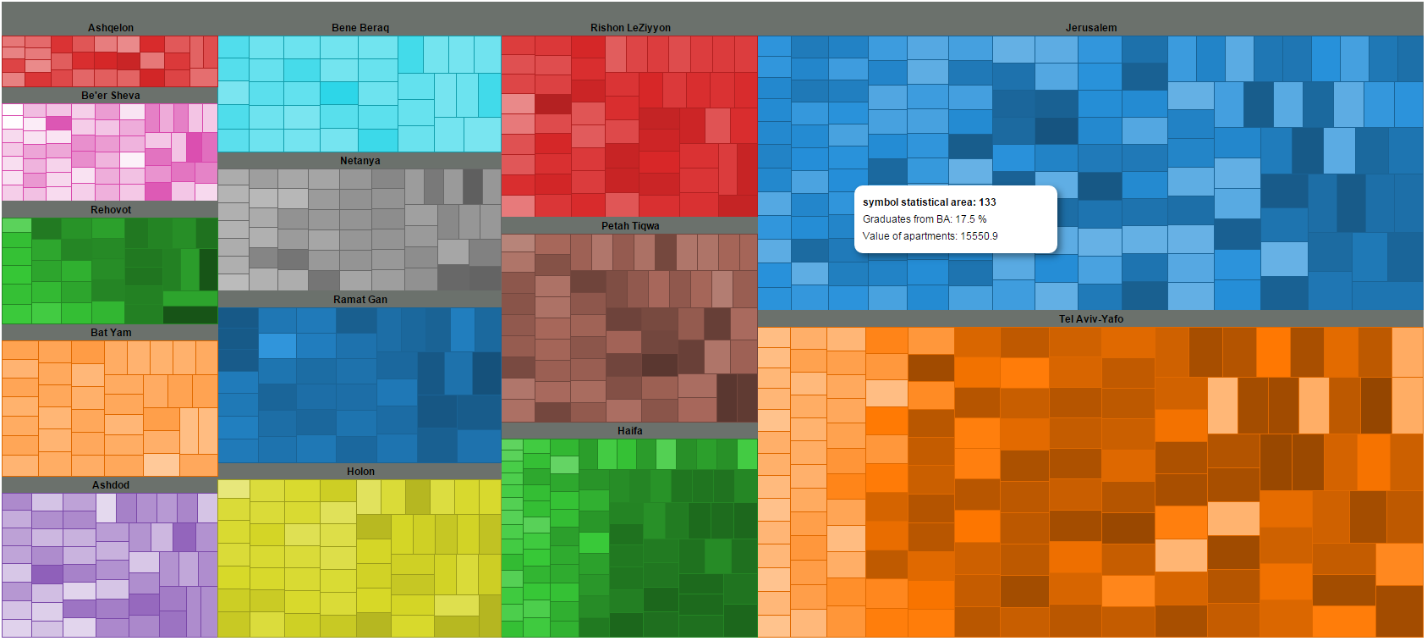
In this iteration, our aim was to address the cluttering introduced in Iteration 2, and allow the user the ability to distinguish by name statistical areas and the cities they are part of. In order to declutter, we eliminated the numerical values that were displayed in each statistical area, and removed the white stroke that surrounded the different statistical areas (rectangles).

In order to allow the user the ability to distinguish by name the different cities, we added labels above each city cluster with the name of the city the color represents. In order to allow the user the ability to distinguish by name and value between statistical areas, we decided to add a tooltip that appears when the user hovers over a particular statistical area. This tooltip shows the statistical area number, the percentage of graduates with a BA within that statistical area, and the average value of apartments within that statistical area. The visual mapping for iteration 3 is summarized in Table 3. The resulting visualization can be seen in Figure 3.

*Table 3: Treemap with Labels + Decluttering - Visual Mapping*

|  |  |  |
| --- | --- | --- |
| **Data Attribute** | **Visual Attribute** | **Additional Information** |
| Neighborhood | Rectangle mark | Hovering over a particular rectangle displays a tooltip with the statistical area number, the percentage of graduates with a BA within that statistical area, and the average value of apartments within that statistical area |
| City | Color + Label |  |
| Apartment Value | Size | High apartment value = large rectangle  Low apartment value = small rectangle |
| Percentage of Residents with Undergraduate education | Brightness | High percentage = low brightness value  Low percentage = high brightness value |

*Figure 3: Treemap with Labels + Decluttering*



### Insights

The amount of clutter has been substantially reduced, and the user is now able to identify each statistical area and each city cluster by name. By doing so, it is now much easier to detect trends in the data and be able to utilize it to create real change. For example, it is much easier now to identify neighborhoods and cities with low education rates and high apartment prices. Jerusalem, for instance, has many of them. The most educated cities overall appear to be Haifa, Tel Aviv, and Rehovot. The relationship between education rate and apartment value is much more linear in Haifa than in most cities. It becomes much clearer now that Tel Aviv and Jerusalem have the highest apartment values. Bene Baraq is clearly the most evenly distributed both in terms of education and apartment values.

|  |  |
| --- | --- |
| **Neighborhoods** | |
| **Task** | **Visualization Performance** |
| Identifying neighborhoods with high apartment value | The differences between neighborhoods are still difficult to see at times, but in those cases the user can use the tooltip to compare numerical values. Statistical areas with high and low apartment values can now be identified by number. |
| Identifying neighborhoods with low apartment value |
| Identifying Neighborhoods with high education attainment rates | Much easier than in the previous iteration. However, the mapping of high education rates to low brightness is still a bit confusing. A legend is required in order to ensure the users reach the correct conclusions. |
| Identifying Neighborhoods with low education attainment rates: |
| Identifying neighborhoods with high education attainment rates and low apartment value | It is easy to see that there are multiple neighborhoods in Be’er Sheva, Ashkelon, and Rehovot that meet this criteria (small, dark rectangles). |
| Identifying neighborhoods with low education attainment rates and high apartment value | Jerusalem seems to have the largest number of neighborhoods with low education attainment rates and high apartment values. There are a few neighborhoods in Tel Aviv as well, but not nearly as many. |
| **Cities** | |
| Identifying cities with high apartment value | Jerusalem and Tel aviv overall have high average apartment values, whereas Ashqelon and Be’er Sheva have low average apartment values. |
| Identifying cities with low apartment value |
| Identifying cities with high education attainment rates | It is a bit difficult to make comparisons between cities, since each city is represented by a different color palette. For example, it is hard to compare Bene Barak to Holon in terms of education attainment rates without reading the raw data using the tooltip. This is because it is difficult to compare brightness levels across colors. |
| Identifying cities with low education attainment rates |
| Identifying cities with high education attainment rates and low apartment value | Because it is still difficult to do a between-city comparison of rectangle sizes and brightness level, it is hard to identify these trends. |
| Identifying cities with low education attainment rates and high apartment value |

### 

### Problems with the Visualization

It is still relatively hard to compare apartment values of different neighborhoods and different cities, since the variation in rectangle sizes is quite small. It is possible to identify patterns between educational attainment rate and average apartment value across different neighborhoods, although they are not immediately evident due to the small visual variation in apartment values. City-wide comparisons are very difficult to do because brightness comparisons between different colors is hard and size comparisons are difficult when the cities are not adjacent. The maximum number of categorical colors that colorBrewer suggests is 12, and we have 14 cities. This is the reason that blue orange, and red are displayed twice. The problem with repeating colors for different cities is that it suggests a visual connection between these cities when in fact there is none.

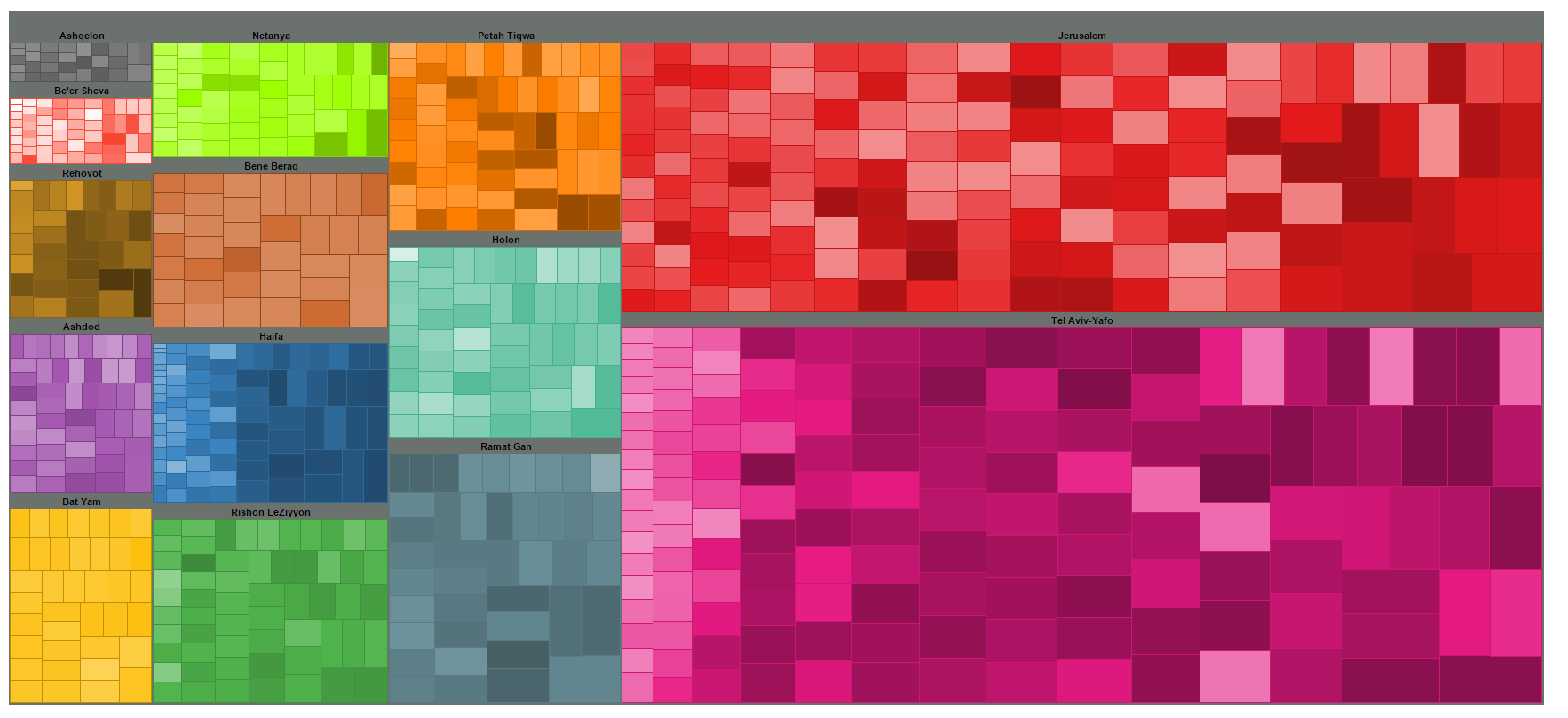
An additional problem in this iteration is that hovering over one of the neighborhoods displays the tooltip in the same exact location as the mouse cursor. Thus, you aren’t sure exactly which rectangle the tooltip represents. In order for the tooltip to be useful, it is crucial to be able to both view the tooltip and identify visually the rectangle it represents.

## Iteration 4: Treemap - Color Palette + Normalizing Data

The previous iterations identified a difficulty in comparing the average apartment value between cities and neighborhoods. In order to try and address this, we decided to normalize the rectangle sizes by raising the average apartment value by the power of 2. Doing so increased the relative size of neighborhoods with high apartment values exponentially, thus helping the user distinguish them more easily from neighborhoods with low apartment values.

Secondly, we decided to address the issues experienced in Iteration 3 as a result of our limited color palette. We first tried addressing the problem by attempting to select 2 additional colors that are different enough from the other colors chosen. It was extremely difficult finding colors that were not already represented in the visualization. Eventually, after a significant search, we added a florescent green and orange to the color palette, as can be seen in Figure 4.1 below. Although they are easily distinguishable from the other colors in the palette, they significantly distort our perception of brightness for between-city comparisons. The florescent green makes it appear as if Netanya has the highest percentage of residents with a BA in the entire country, when in fact this is not the case.

### *Figure 4.1 - Selecting 2 additional colors*

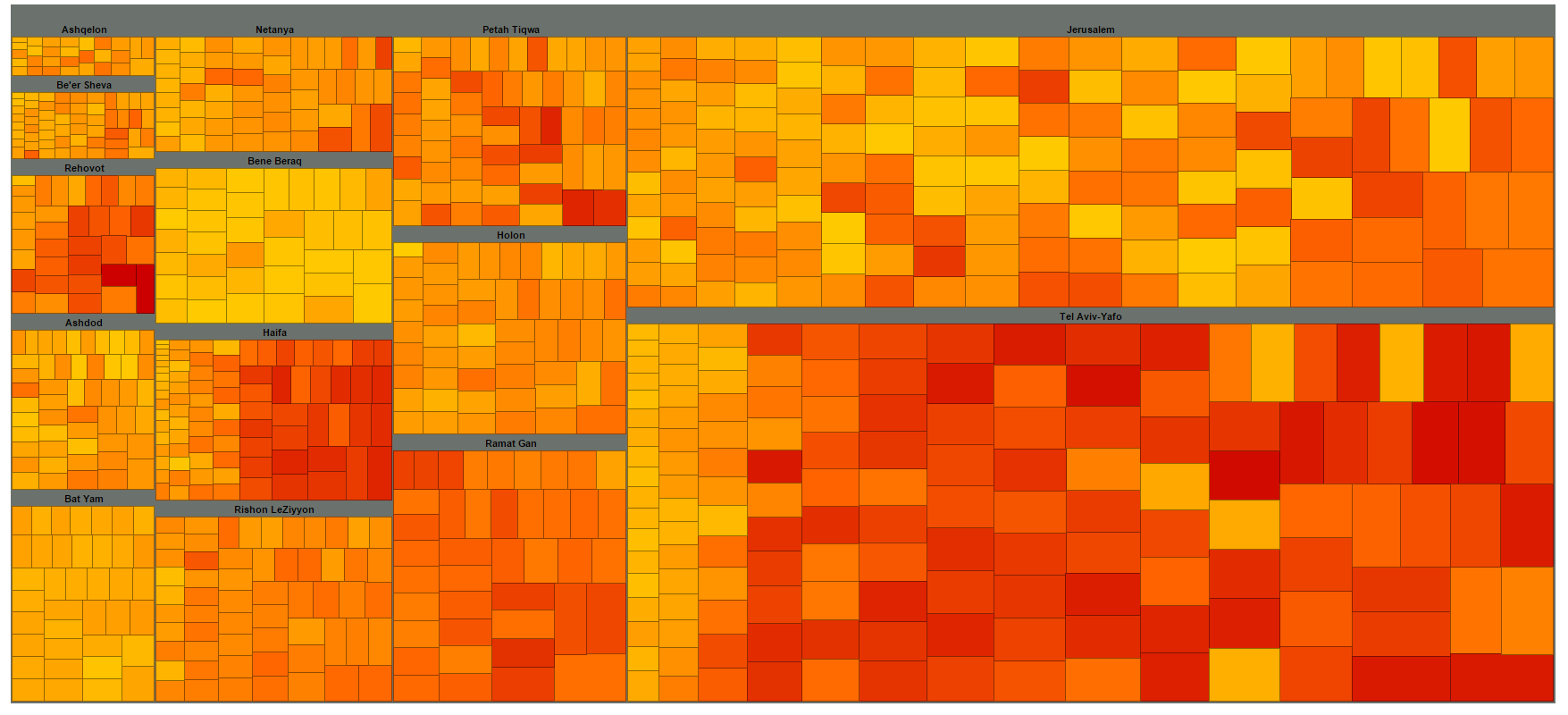


Since variation in color interferes with our ability to compare between cities, we decided to change the visual mapping so that color is not used as an indicator for city. Instead, we decided to map a sequential color palette to education rate. The complete visual mapping is summarized in Table 4.2. The resulting visualization can be seen in Figure 4.2.

*Table 4.2: Treemap – Color Palette + Normalizing Data – Visual Mapping*

|  |  |  |
| --- | --- | --- |
| **Data Attribute** | **Visual Attribute** | **Additional Information** |
| Neighborhood | Rectangle mark | Hovering over a particular rectangle displays a tooltip with the statistical area number, the percentage of graduates with a BA within that statistical area, and the average value of apartments within that statistical area |
| City | Label |  |
| Apartment Value | Size2 | High apartment value = large rectangle  Low apartment value = small rectangle |
| Percentage of Residents with Undergraduate education | Color | High percentage = Red  Low percentage = Yellow |

*Figure 4.2: Treemap - Color Palette + Normalizing Data*



### Insights

First, the differences between cities and neighborhoods in terms of average apartment value are much more apparent now. You can see much quicker that Ashquelon and Be’er Sheva have the smallest values, whereas Tel Aviv and Jerusalem have the largest values. It is also a lot easier to distinguish now within cities between neighborhoods with smaller values and neighborhoods with larger values. The variance in terms of average apartment value is more apparent as well. Haifa and Tel Aviv have more variance than Bene Barak and Bat Yam.

In terms of education rate, it is much easier now to compare differences between cities and neighborhoods now that they are all in the same color scale. For example, you can suddenly see that the neighborhoods with the lowest education rates in Tel Aviv in most cases are still better than neighorhoods in Bene Barak, where there are the lowest overall education rates. Moreover, relative to most cities, Ramat Gan, Tel Aviv, and Haifa have a large number of neighborhoods with high education rates. The variance in education in Jerusalem is clearly the highest among all major cities.

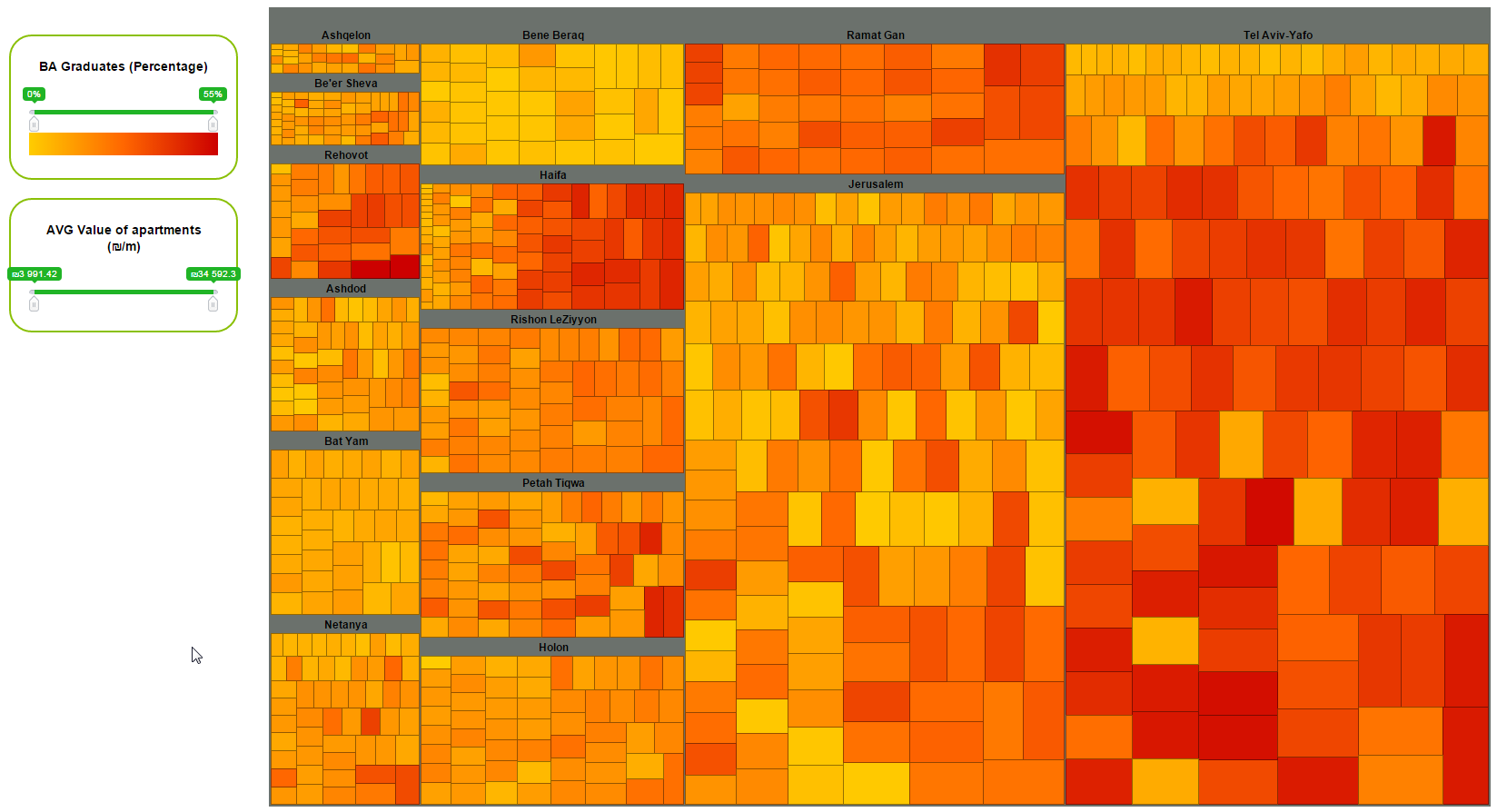
|  |  |
| --- | --- |
| **Neighborhoods** | |
| **Task** | **Visualization Performance** |
| Identifying neighborhoods with high apartment value | The differences between neighborhoods are much clearer after the normalization process. Comparing between specific neighborhoods can still be difficult when the neighborhoods aren’t adjacent, but the tooltip can be used as support for such cases |
| Identifying neighborhoods with low apartment value |
| Identifying Neighborhoods with high education attainment rates | The ability to identify and compare neighborhoods with high or low education attainment rates is easier than before. A legend is still required in order to ensure the users understand the significance of the colors, however |
| Identifying Neighborhoods with low education attainment rates: |
| Identifying neighborhoods with high education attainment rates and low apartment value | Easily done regardless if you’re looking at a particular city or trying to find specific neighborhoods within a city. |
| Identifying neighborhoods with low education attainment rates and high apartment value | Easily done regardless if you’re looking at a particular city or trying to find specific neighborhoods within a city. |
| **Cities** | |
| Identifying cities with high apartment value | More easily identifiable. Jerusalem and Tel aviv overall have high average apartment values, whereas Ashkelon and Be’er Sheva have low average apartment values. |
| Identifying cities with low apartment value |
| Identifying cities with high education attainment rates | Easily compared, unlike in the previous iteration. Tel Aviv, Haifa, and Rehovot have the highest, and Bene Barak and Bat Yam have the lowest. |
| Identifying cities with low education attainment rates |
| Identifying cities with high education attainment rates and low apartment value | These relationships seem easily identifiable now. Haifa and Rehovot seem to have the most neighborhoods with high education attainment rates and low apartment values. Jerusalem and Bene Barak seems to be the cities with the most neighborhoods with a high apartment value nd low education attainment rates. |
| Identifying cities with low education attainment rates and high apartment value |

### Problems with visualization

First, a legend and title are still required in order for the user to understand the visual mapping in front of them. Secondly, users of this visualization will often likely have different values in mind regarding what they consider “low” and “high”. The current visualization, does not support a user’s need to filter between regions with specific values. If, for instance, an educator identified that the most problematic neighborhoods are those where less than 10% of their residents have a BA and with apartment values above 10,000 Shekels per meter square and he would like to find those neighborhoods, it would be almost impossible to do using this visualization.

## Iteration 5: Treemap - Adding Filters, Legend, & Title

In this iteration, we decided to address the need for a legend and a title. We also added interactive filters that allow the users to filter neighborhoods based on both the percentage of residents with a BA and the average apartment value per meter square. The visual mapping remained the same as in Iteration 4. The resultant visualization can be seen in Figure 5 below.

****

Adding the filters allows users not only to identify neighborhoods with specific criteria but it also facilitates better understanding of the information by decluttering the visualization from information that is not relevant to their primary task. With these new additions this visualization seems to answer all our user needs adequately.

# Discussion

After many iterations in which we continuously improved upon our visualization, we resulted in a visualization we believe effectively displays the relationship between educational attainment and real estate value across different cities and neighborhoods in Israel. In this section, we assess the value of our visualization and reflect upon the various design decisions made.

## Value of visualization

In order to assess our visualization, we will use John Statsko’s value of visualization model that includes four assessment criteria: time, Insights, Essence & Confidence.

### Time

This criteria refers to the ability of the visualization to minimize total time needed to answer a wide variety of questions. As summarized in Table 4, our visualization allows users to quickly receive answers to a wide variety of questions, from “which city has the largest portion of neighborhoods with apartment values above 10,000?” to “which neighborhoods in Haifa have a low percentage of BA graduates and a high average apartment value per square meter?”. The visualization does so by simultaneously presenting multiple data dimensions (location, education, and apartment value) on multiple levels (neighborhood specific, city-wide, inter-city, and country-wide) in a way that is easy to comprehend. Using a spreadsheet to answer such questions would likely take significantly more time. Furthermore, the filtering ability helps the user save additional time by allowing them to identify the most relevant neighborhoods to their question quickly.

Insights

This criteria refers to the user’s ability to discover insights or insightful questions about the data through the visualization. Many insights can be made through our visualization. For example, the relationship between education rate and apartment value in Tel Aviv-Yafo is mostly positive and linear, except for 5 neighborhoods in which the education rate is extremely low and the average apartment prices are very high. Finding these outliers would have been nearly impossible by looking at a spreadsheet with ~1000 data points, however it is immediately detectable using our visualization. That said, because the majority of users do not remember the geographical locations of each statistical area number, it is hard to identify possible explanations for these outliers.

Essence

Essence is a visualization’s ability to convey an overall take-away sense of the data. One essence that is successfully conveyed through our visualization is the fact that educated people in Israel tend to live in particular cities, specifically Tel Aviv, Ramat Gan, Haifa, and Rehovot.

Confidence

The final criteria is the visualization’s ability to generate confidence and trust about the data. On the one hand, we did not visualize aggregated data and we presented the original data in a tooltip so the user always has full access it. On the other hand, by setting the size of each neighborhood according to the squared value of the average apartment and not the original number, we reduced the user’s ability to accurately compare and assess different neighborhoods according to that value. Because of the transformation, a rectangle could appear to be over double the size of another rectangle when in fact the increase in value is much less than that.

Self Reflection

Overall, we believe our final visualization answers a good portion of our user’s needs. With that, there are several improvements that could have been made.

For one, our visualization does not allow users to begin their analysis with a specific neighborhood in mind. If, for instance, an apartment seeker would like to compare a specific neighborhood to other neighborhoods in the city in terms of education rate and apartment price, they have no means of doing so. Adding a search function which highlights the relevant neighborhood could have addressed this issue.

Secondly, comparing the average value of an apartment in two neighborhoods is not always easy even after we normalized the data. This is partially because many neighborhoods share similar values, and partially because people are not very good at assessing size. The user may use the tooltip to compare exact values, however there is no way for the user to visibly see the values of two neighborhoods at once. Turning each neighborhood into a toggle which can anchor or hide a tooltip on click might have helped address this issue. Moreover, by setting the size of each neighborhood according to the squared value of the average apartment and not the original value, we reduced the user’s ability to accurately assess the difference between the average apartment value of different neighborhoods. One rectangle could appear to be over double the size of another rectangle when in fact the increase in value is much less than that.

In addition, the connection between the city labels and the rectangles relevant to that city might not be obvious enough. Many visualizations and images contain labels beneath the image, so it is possible the users will get confused regarding which label refers to which cluster. For example, the “Haifa” label could refer to the very-red neighborhoods beneath it or to the very-yellow neighborhoods above it. In order to ensure that there is no confusion amongst users, it might be beneficial to create a stronger visual connection between the city label and the cluster of neighborhoods it represents.

Finally, the visualization might help its users quickly identify trends and outliers, however it does not and cannot display additional variables, like population age, religion, and geography, which could help explain the findings. Legible treemaps are normally bound to 3 primary variables that need to be mapped, one of which has a hierarchical structure. Treemaps that try displaying more than that mostly end up being extraordinarily confusing and unreadable.

## Conclusion

In this work, we present a treemap visualization for the relationship between educational attainment and real estate value across different cities and neighborhoods in Israel. The visualizationemerges from the need of educators, apartment seekers, government and police workers to identify regions in Israel based on their educational attainment rate, average apartment value, and the ratio between the two. In our visualization, we utilized interactive elements in order to broaden the analysis potential and compensate for limitations. A detailed storyboard of the visualization process can be found in [Appendix 3](#_Appendix_3:_Storyboard).

From the visualization, we gained many important insights. For example, Haifa and Rehovot have many neighborhoods with high education attainment rates and low apartment values. These areas might be worth investing in in terms of real estate. Jerusalem and Bene Barak, on the other hand, have many neighborhoods with a high apartment value and low education attainment rates, which could potentially explain or be a result of the social unrest that is often exhibited in both cities. Moreover, we discovered that Tel Aviv contains 5 neighborhoods which do not follow the same linear relationship as the rest of the city. Unlike the majority of neighborhoods in Tel Aviv-Yafo, these neighborhoods have very high apartment values and very low education rates, making them relevant points of interest for our users.

Even though the visualization successfully highlighted these important insights, it is by no means perfect: users cannot gain neighborhood-driven insights, comparing neighborhoods according to average apartment value is sometimes difficult and misleading, and the visualization cannot be easily mended to display additional confounding variables. Despite these limitations, the visualization supports all the primary user tasks and thus ensures that many important insights can be found accordingly. In order to further improve our understanding of the relationship between educational attainment and apartment values, future work could involve manually mapping statistical areas to geographic coordinates in order to display the data geographically. Doing so might bring to light additional insights that cannot be obtained by the current visualization.

# Appendices

## Appendix 1: Original Dataset

|  |  |  |  |
| --- | --- | --- | --- |
|  | **בעלי תואר ראשון** | **שווי דירת מגורים** | סמל |
| Locality | אחוז מכלל בני 20–55 בא"ס | ממוצע למטר מרובע, בש"ח | אזור |
| Jerusalem | 2.5 | 13,598 | 111 |
| Jerusalem | 4.0 | 13,305 | 112 |
| Jerusalem | 6.3 | 14,869 | 113 |
| Jerusalem | 6.0 | 16,348 | 114 |
| Jerusalem | 5.8 | 14,398 | 115 |
| Jerusalem | 8.3 | 14,598 | 116 |
| Jerusalem | 14.1 | 15,096 | 121 |
| Jerusalem | 13.0 | 13,795 | 122 |
| Jerusalem | 15.4 | 14,155 | 123 |
| Jerusalem | 15.4 | 12,873 | 124 |
| Jerusalem | 16.1 | 12,873 | 125 |
| Jerusalem | 21.5 | 13,823 | 131 |
| Jerusalem | 18.3 | 13,270 | 132 |
| Jerusalem | 17.5 | 15,551 | 133 |
| Jerusalem | 16.4 | 12,036 | 134 |
| Jerusalem | 16.5 | 13,287 | 135 |
| Jerusalem | 17.1 | 13,877 | 136 |
| Jerusalem | 3.9 | 16,636 | 411 |
| Jerusalem | 2.1 | 16,794 | 412 |
| Jerusalem | 4.6 | 15,897 | 413 |
| Jerusalem | 15.3 | 17,591 | 421 |
| Jerusalem | 14.1 | 18,701 | 422 |

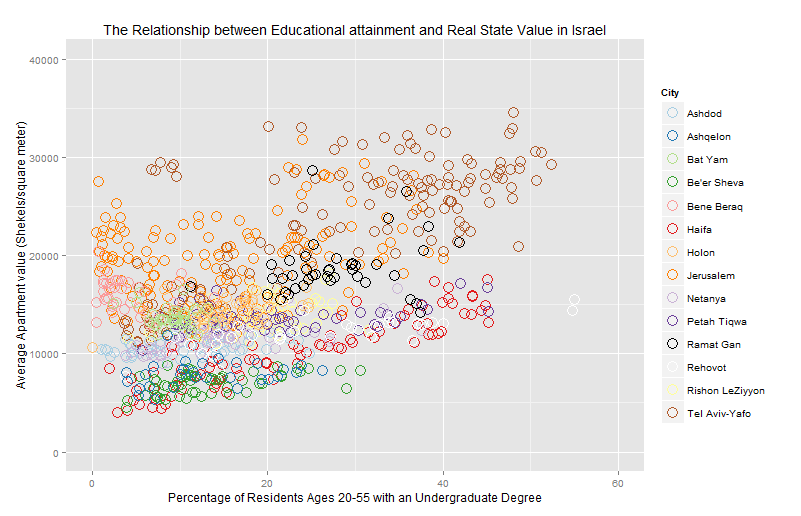
## Appendix 2: Data Analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| City | Number of statistical areas | Percentage With Bachelor’s Degree \* | | Apartment Worth (NIS per Sq. Meter) | |
|  | Mean | Standard Deviation | Mean | Standard Deviation |
| Jerusalem | 157 | 15.5 | 10.13333 | 19463.1 | 4241.577 |
| Tel Aviv | 147 | 27.8 | 14.9 | 23288.7 | 6393.5 |
| Haifa | 86 | 23.1 | 13.2 | 10155.8 | 3476.1 |
| Rishon LeZiyyon | 58 | 19.1 | 5.4 | 13791.9 | 1632.1 |
| Ashdod | 51 | 11.3 | 6.1 | 10727.1 | 105.2 |
| Petah Tiqwa | 61 | 20.4 | 9.7 | 13542.7 | 1176.8 |
| Be’er Sheva | 55 | 13.5 | 6.3 | 7032.4 | 1042.6 |
| Netanya | 53 | 14.4 | 7.4 | 11682.3 | 1449.9 |
| Holon | 54 | 15.5 | 5.1 | 14507.1 | 1203.9 |
| Bene Beraq | 37 | 4.1 | 3.3 | 16004.4 | 1656.9 |
| Ramat Gan | 33 | 27.1 | 6.1 | 18282.7 | 2596.5 |
| Bat Yam | 38 | 9.2 | 2.3 | 13641.7 | 1015.3 |
| Rehovot | 34 | 26.2 | 11.9 | 12241.8 | 1660.0 |
| Ashquelon | 29 | 13.1 | 6.5 | 7943.1 | 1266.2 |
|  |  |  |  |  |  |
| \* Percentage of people ages 20-55 living within the statistical area | | | |  |  |
|  |  |  |  |  |  |

# 

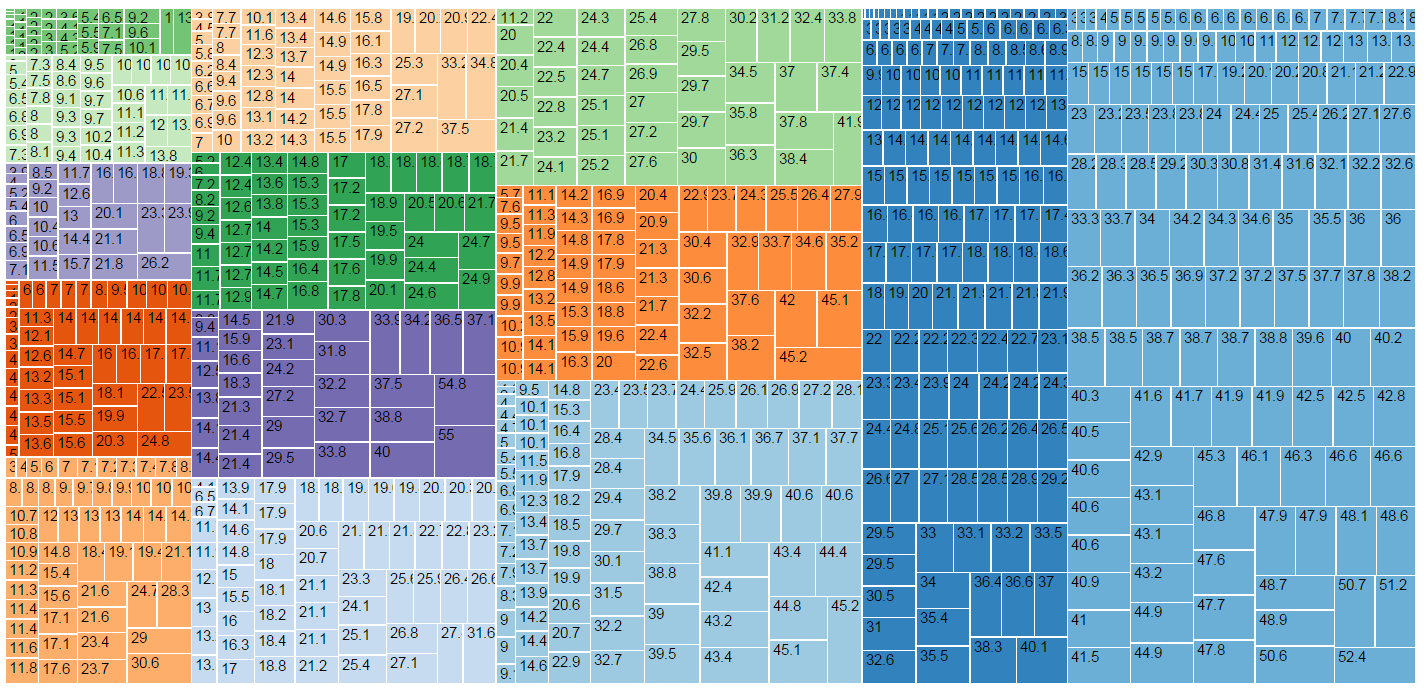
## Appendix 3: Storyboard of visualization process

### Iteration 1: Scatterplot

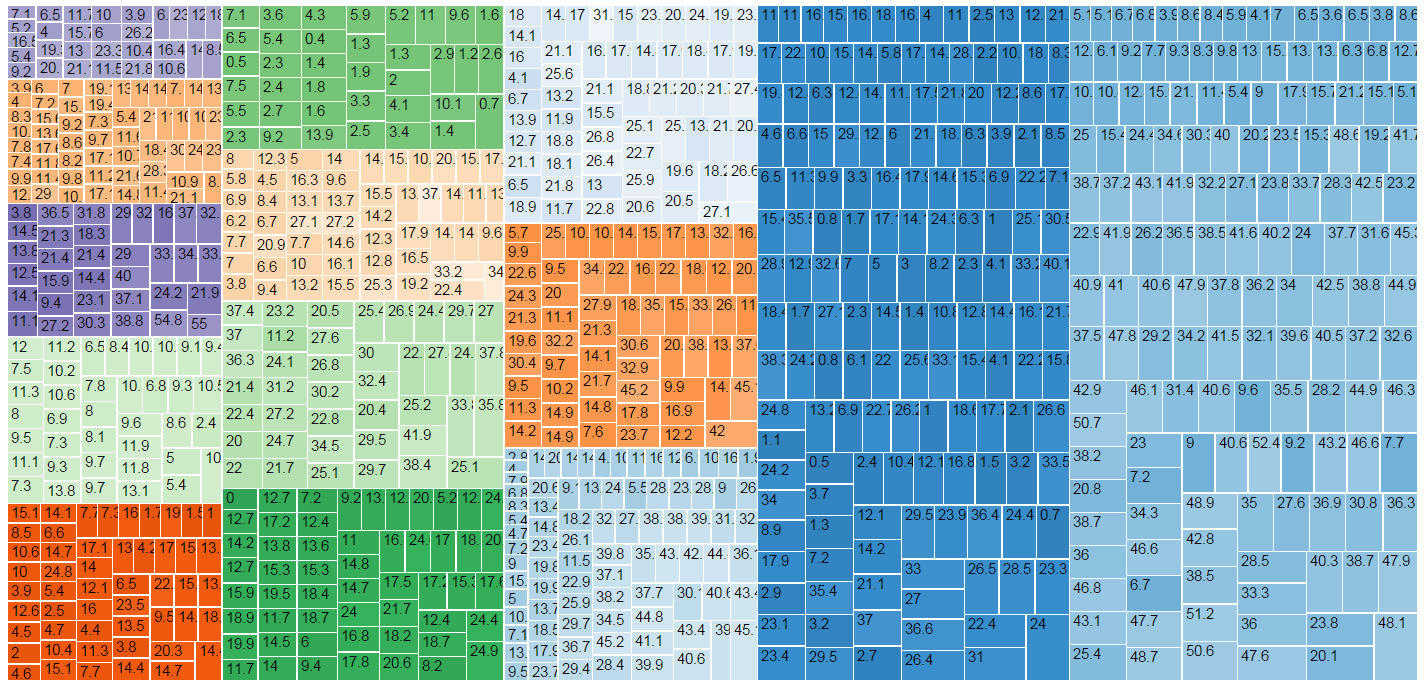


### Iteration 2: Treemap

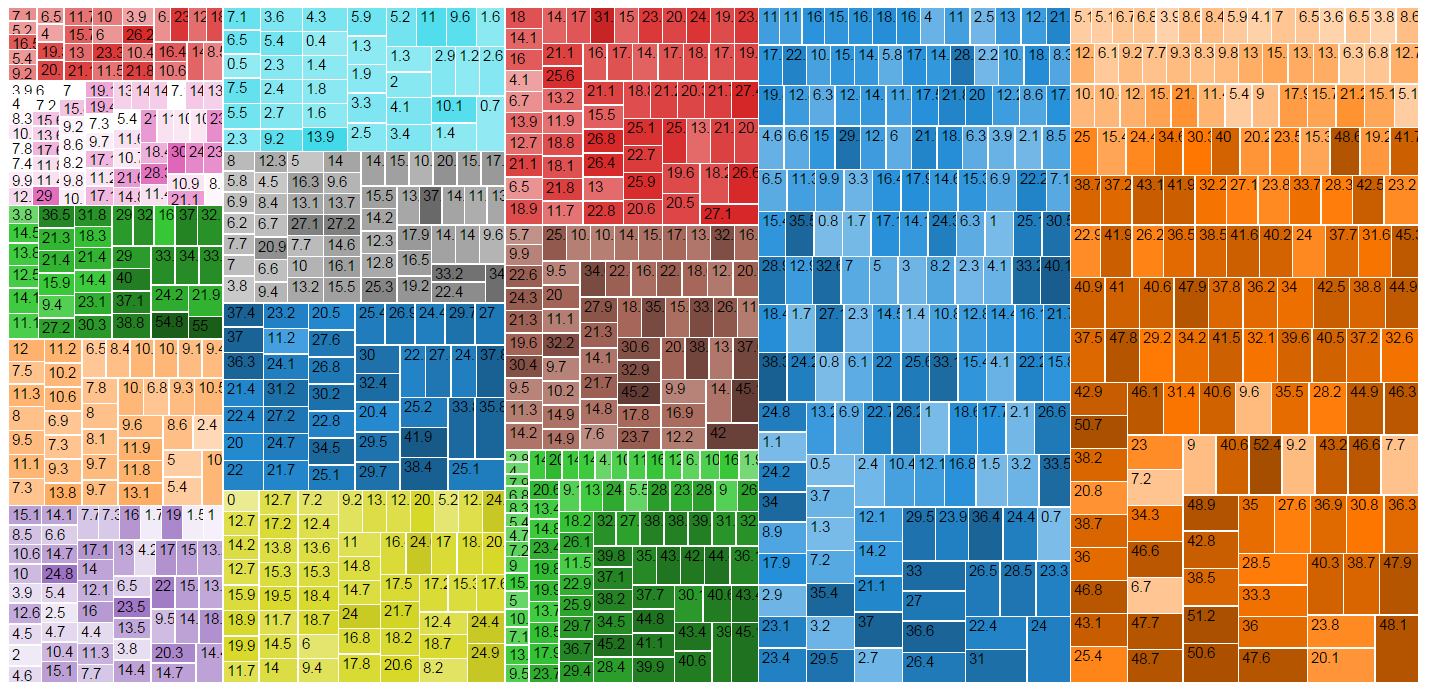
### Step 2.1 Creating the initial treemap in D3. We mapped apartment value to size and displayed education rate in the form of a label in each rectangle.



### Step 2.2 Mapped brightness to education rate. We noticed that the brightness spectrum wasn’t wide enough to visually see a difference.

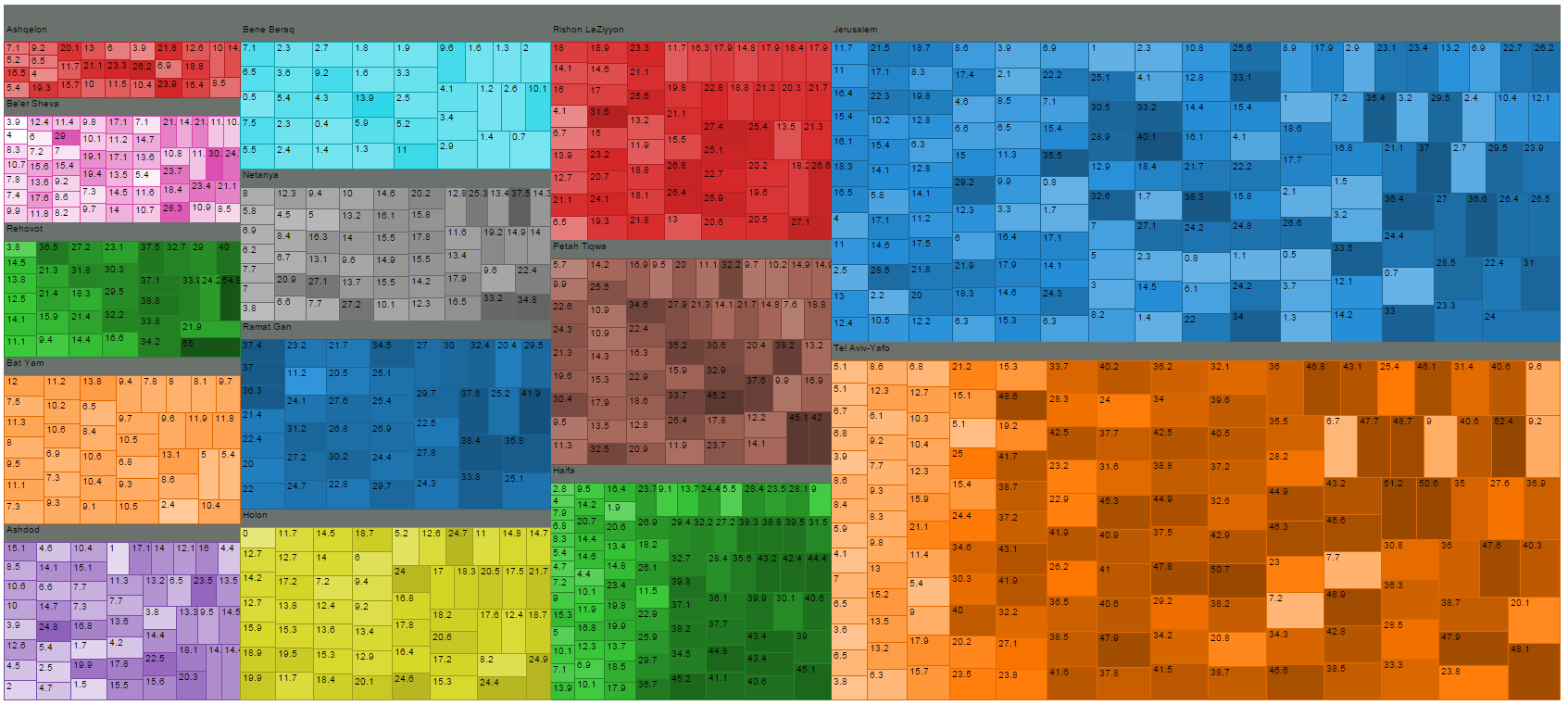


### Step 2.3 We widened the brightness spectrum so that it is more visible.

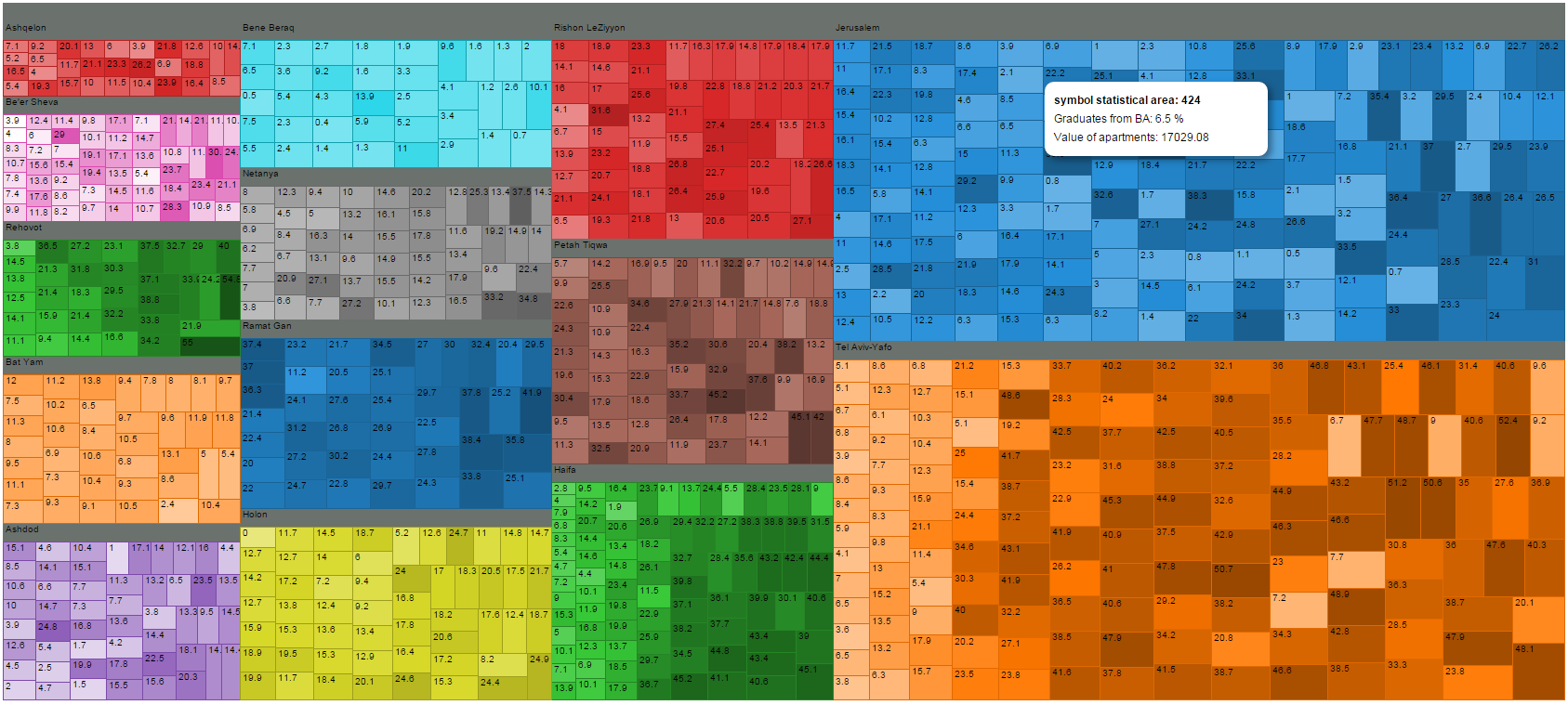


### Iteration 3: Treemap – Labels + Decluttering

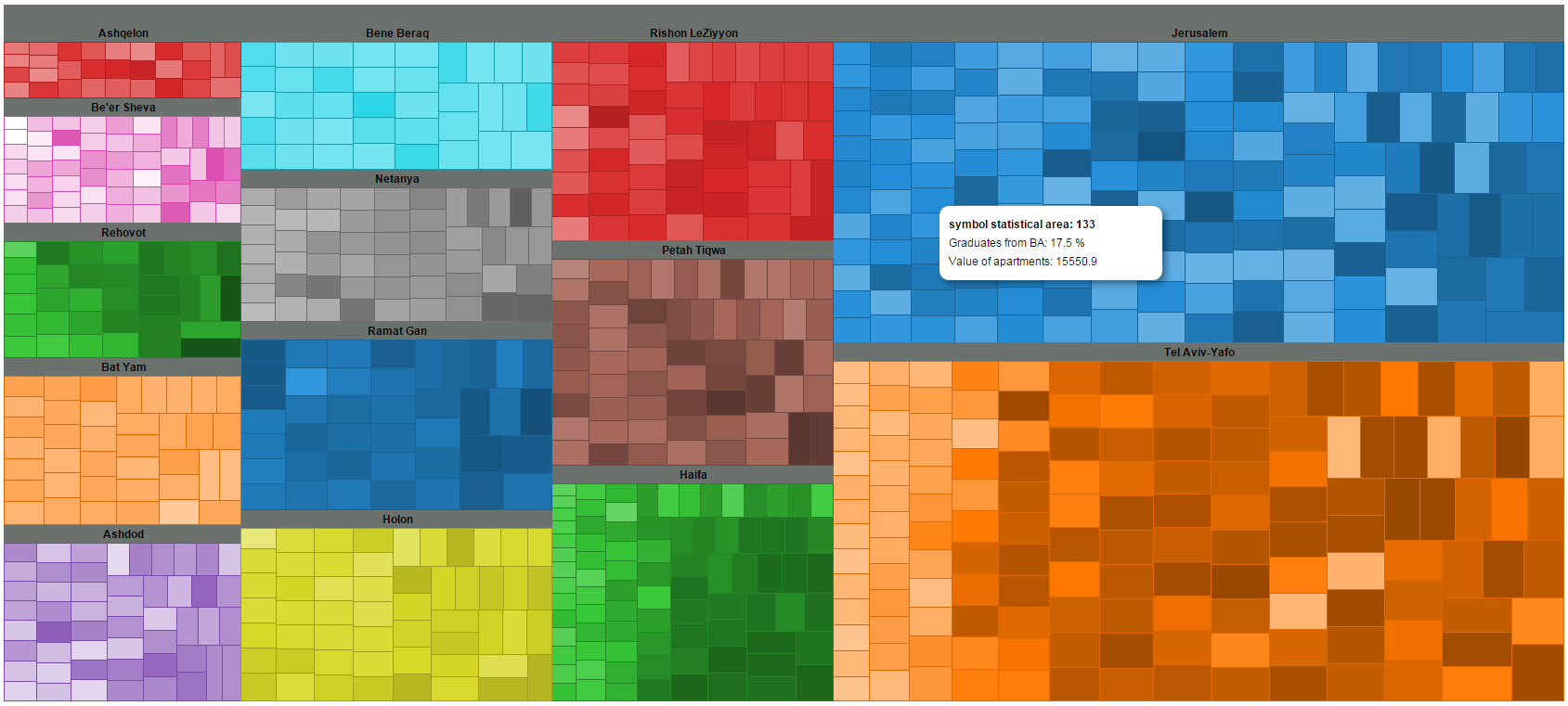
Step 3.1 Adding city labels, changing stroke colors



Step 3.2 Adding Tooltip

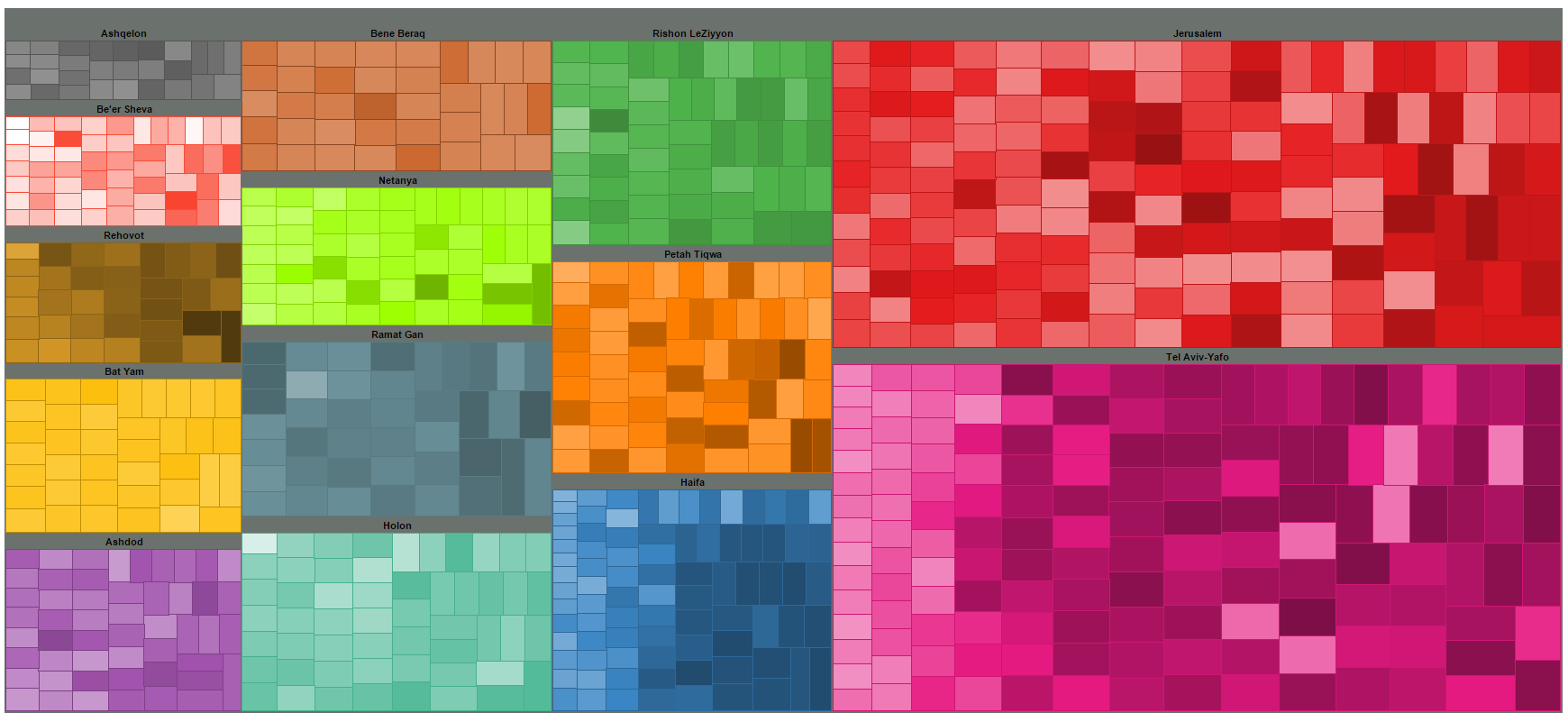


Step 3.2 Removing education rate values

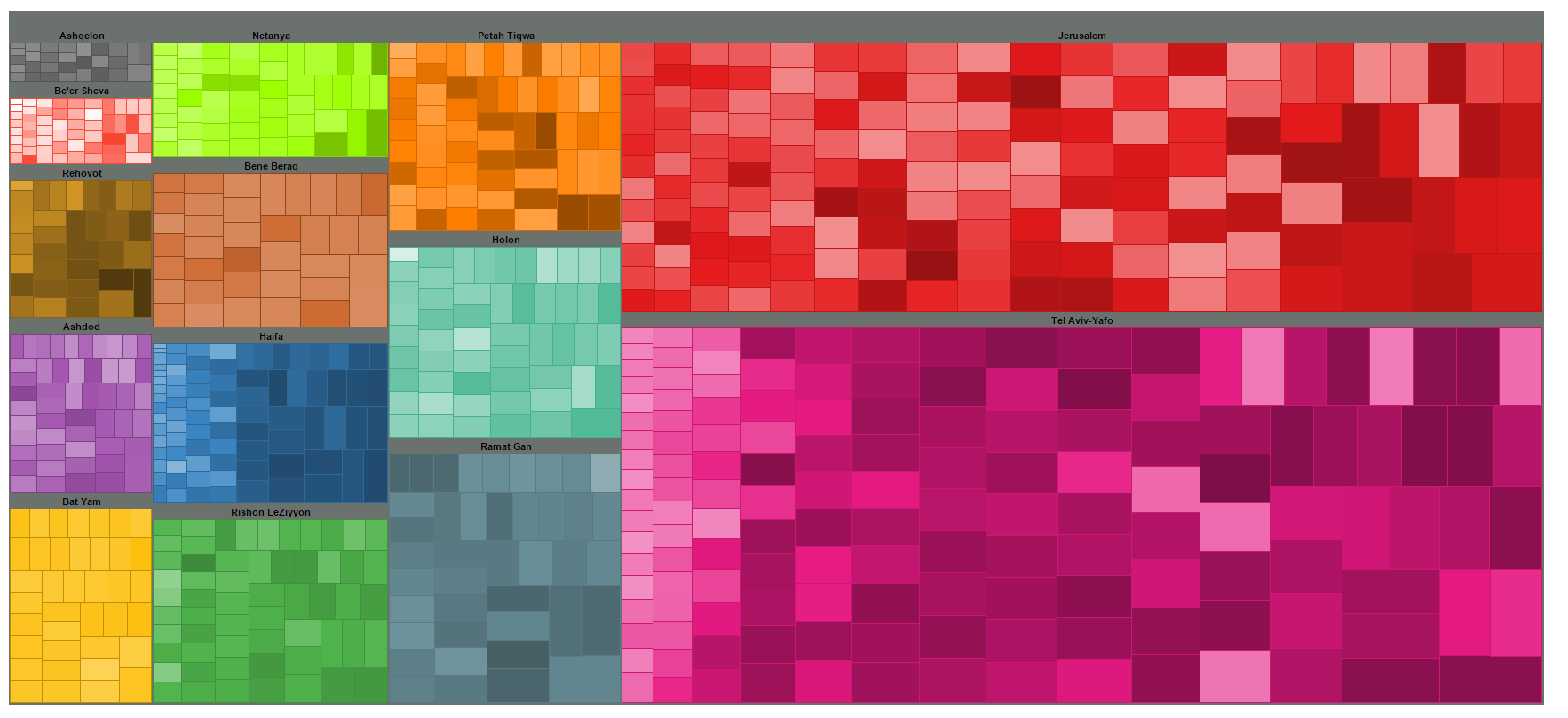


### Iteration 4: Treemap – Color Palette + Normalizing Data

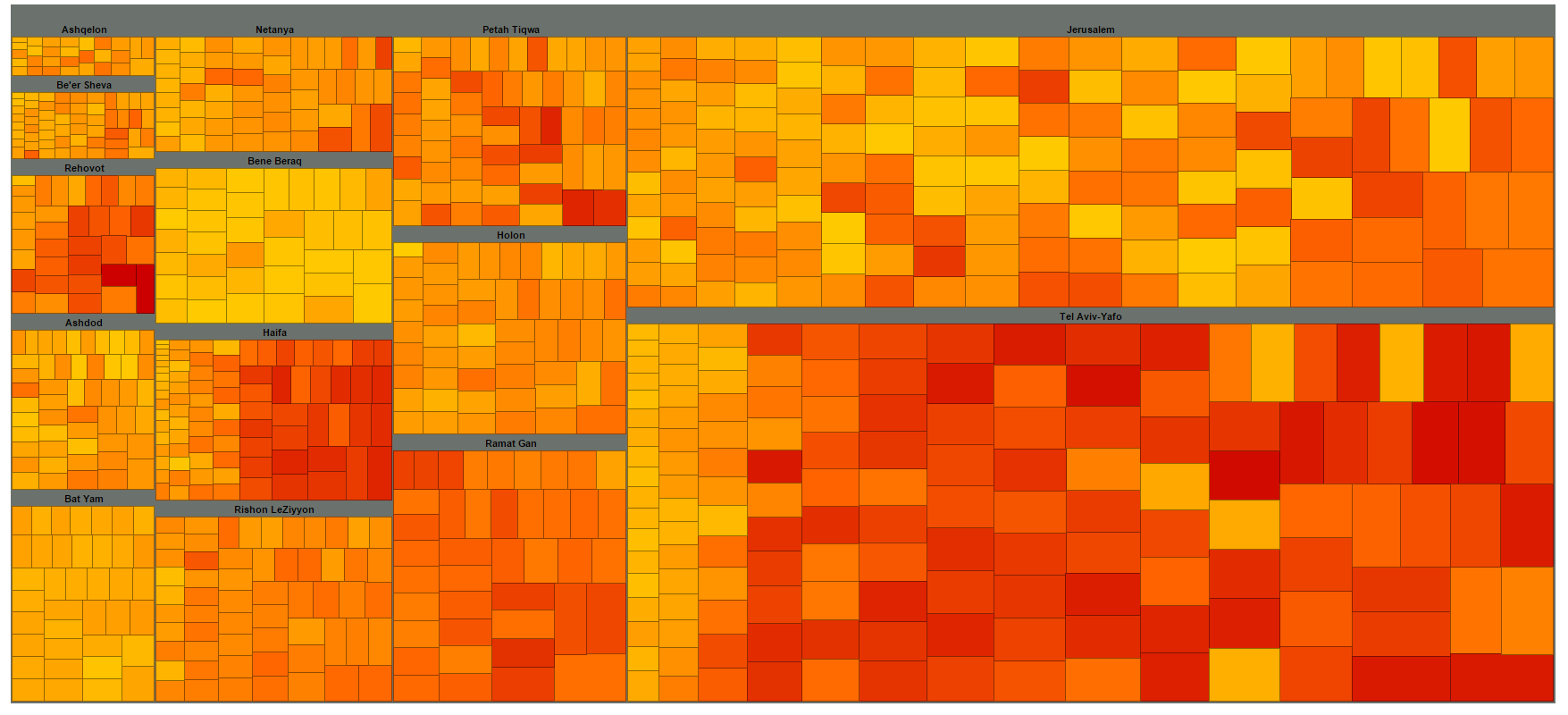
Step 4.1 Adding two new colors to color palette



Step 4.2 Raising apartment value to the power of 2

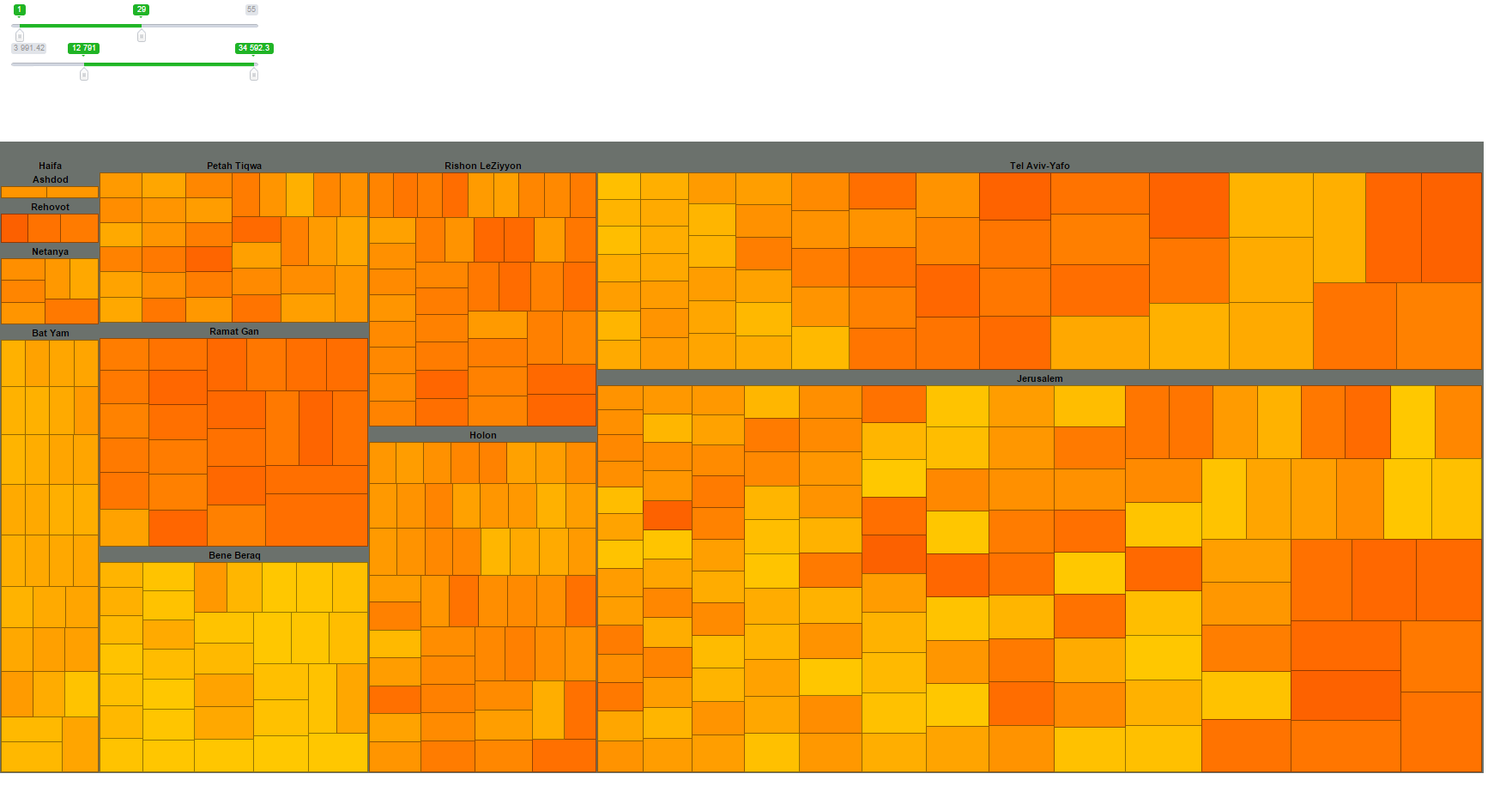


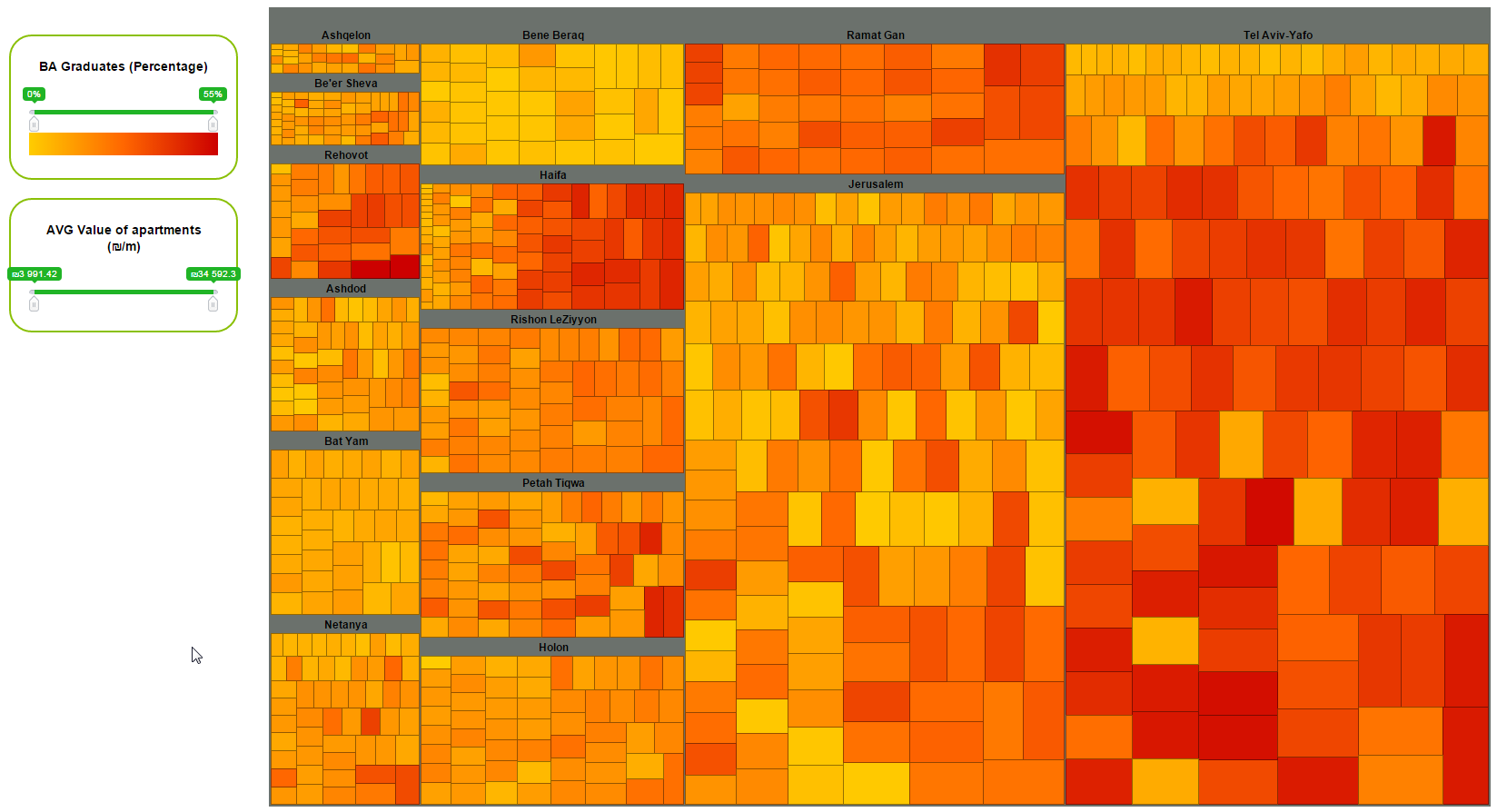
Step 4.1 Mapping color to education rate rather than to city



### Iteration 5: Treemap – Adding Filters, Legend & Title

Step 5.1 Adding filters



Step 5.2 Adding a range of colors and ordering filters placement****

# References

Bayer, P., Ferreira, F., & McMillan, R. (2007). *A unified framework for measuring preferences for schools and neighborhoods* (No. w13236). National Bureau of Economic Research.

Behrman, J. R. (1997). *Mother's schooling and child education: A survey*. Philadelphia: University of Pennsylvania.

Ben-David, D. (Ed.). (2014). *State of the Nation Report: Society, Economy, and Policy in Israel*. Taub Center for Social Poilicy Study in Israel.

Berkner, L. B. & Chavez, L. (1997). Access to postsecondary education for the 1992 high school graduates. (NCES Publication No. 98-105). Washington, DC: U.S. Government Printing Office

Bursik, R. J. (1988). Social disorganization and theories of crime and delinquency: Problems and prospects\*. *Criminology*, *26*(4), 519-552.

Chiodo, A., Hernández-Murillo, R., & Owyang, M. T. (2010). Nonlinear effects of school quality on house prices. *Federal Reserve Bank of St. Louis Review*,*92*(May/June 2010).

Horn, L. & Bobbitt, L. (2000). Mapping the road to college: First generation students 'math track, planning strategies, and context of support (NCES Publication No. 2000-153). Washington, DC: U.S. Government Printing Office.

Kane, T. J., Riegg, S. K., & Staiger, D. O. (2006). School quality, neighborhoods, and housing prices. *American Law and Economics Review*,*8*(2), 183-212.

Li, W. (2012). Does School Quality Affect Real Estate Prices? The Effect of Top-Tier Elementary Schools on Property Prices in Shanghai.

Nunez, A. M. & Cuccaro-Alamin, S. (1998). First-generation students: Undergraduates whose parents never enrolled in postsecondary education (NCES Publication Number 1999- 082). Washington DC: U.S. Government Printing Office.

Schroeder, W. J., Lorensen, B., & Martin, K. (2004). *The visualization toolkit*. Kitware.

Shafer, H. (2013, August 14). International comparison: 126 paychecks for an apartment in Israel, twice more than in Britain. *Calcalist*. Retrieved September 4, 2015, from <http://www.calcalist.co.il/real_estate/articles/0,7340,L-3609995,00.html>

1. The full dataset can be found in the shared GitHub folder [↑](#footnote-ref-1)
2. A widely accepted online tool designed to help people select good color schemes for visualizations. <http://colorbrewer2.org/> [↑](#footnote-ref-2)