Looking for an optimal place for relocation

Introduction

Relocation to another country or state is a relatively common situation. For many people the reason for relocation is a new job (often some advancement in their career) or studies. In any case, it can be a wonderful experience for the person traveling alone or for the whole family to experience living in another place and to get familiar with a different culture. I personally plan a relocation to Waltham, Massachusetts for one year, to work in a research laboratory. I hope that the code presented here may help other people to find an optimal place for relocation.

In order to find an appropriate place to live with the family and to work in Waltham, we have set several priorities:

Distance from workplace

Safety

A quiet environment with basic supply stores and restaurants

In order to satisfy these requirements, I have investigated geospatial data of Massachusetts municipalities (unique postal code zones), an FBI data on crimes in Massachusetts and Foursquare data information on venues in each zone.

Data

A database of Massachusetts municipalities by postal codes was publicly available on url: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=MA The data was available for download as a csv file, including postal code zone numbers, municipality names and latitude and longitude of each zone.

Using geospatial location of the laboratory, I was able to calculate the distance from the laboratory to each postal code zone and to sort out all zones that were more than 20 km far from it.

Safety data was obtained from an open FBI dataset of number of crimes by town in Massachusetts in 2013. Url is https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/tables/table-8/table-8-state-cuts/table_8_offenses_known_to_law_enforcement_massachusetts_by_city_2013.xls File name is 'table_8_offenses_known_to_law_enforcement_massachusetts_by_city_2013.xls' The file contains each town's population and number of violent crime and property crime cases during 2013. Based on these numbers we will be able to calculate rates of both types of crime in each city.

Foursquare account was opened in order to query venue data in each postal code zones. The venues were then sorted and 10 most frequent venue types were presented.

The postal code zones were then clustered based on the Foursquare and criminal data. Optimal number of clusters was applied using an "elbow" plot of error for each number of clusters.
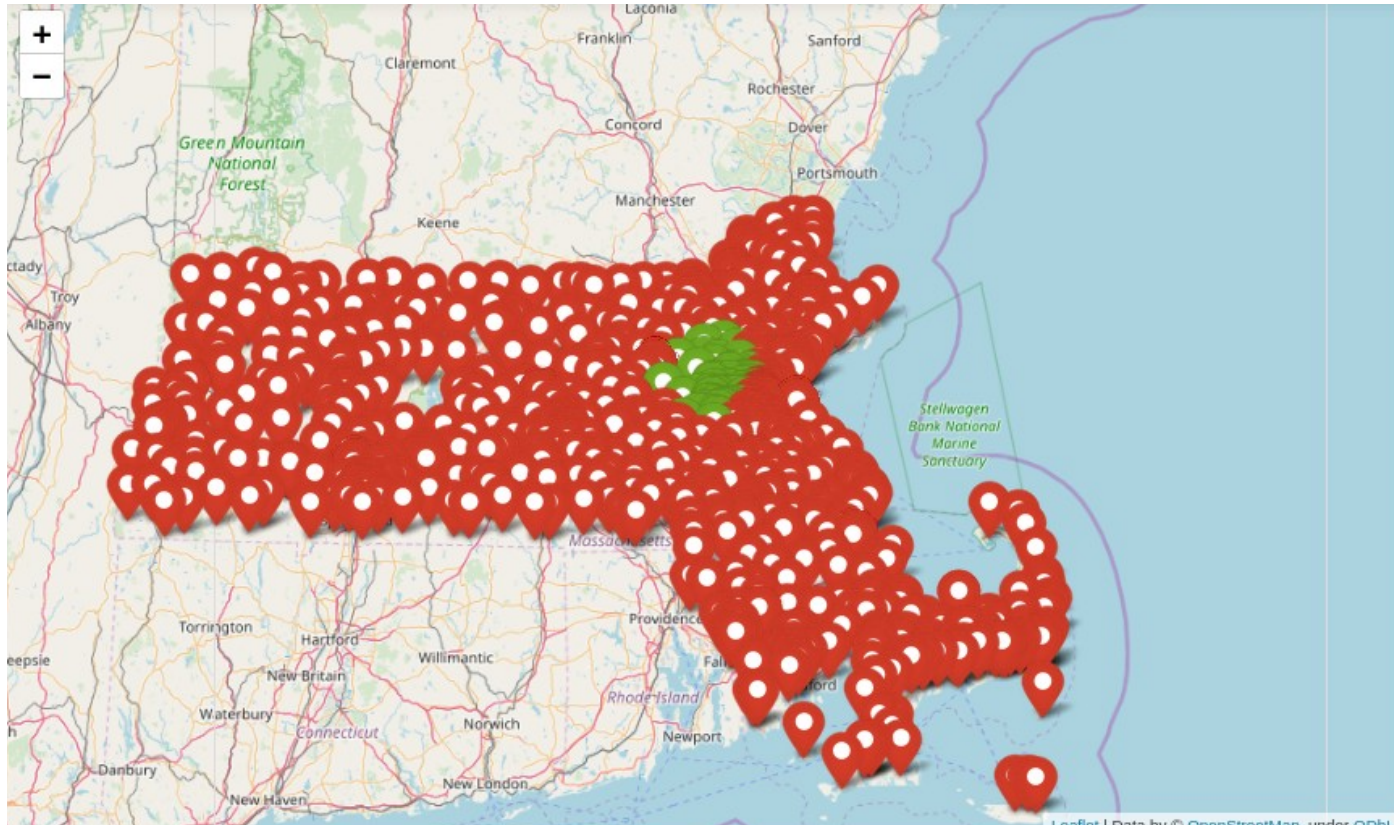
Next, the rates of property and violent crime were summarized and visualized for each cluster using barplots.

Clusters with lowest number of crimes were inspected for most frequent venues. Based on this data, an optimal place for relocation was selected.
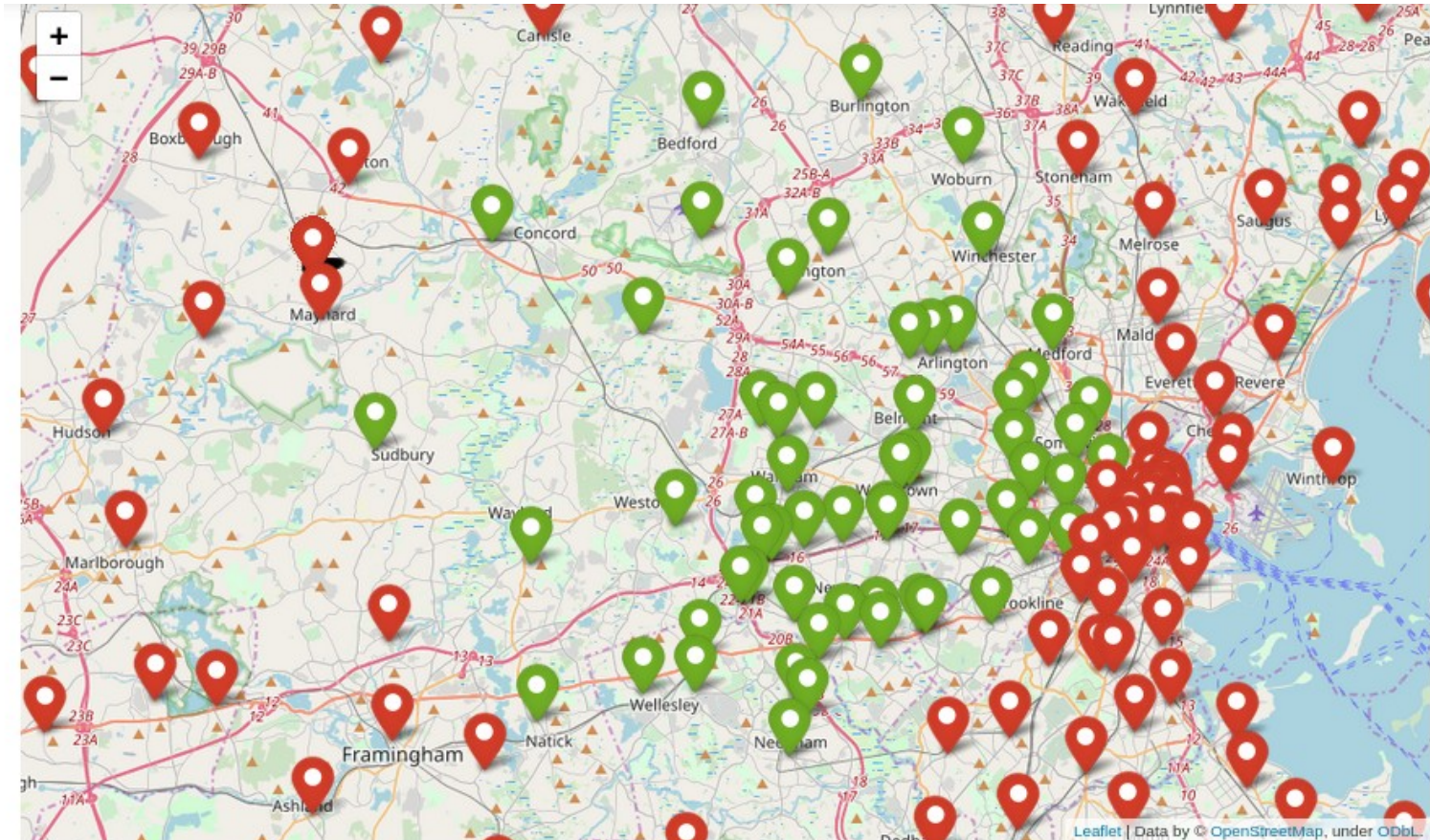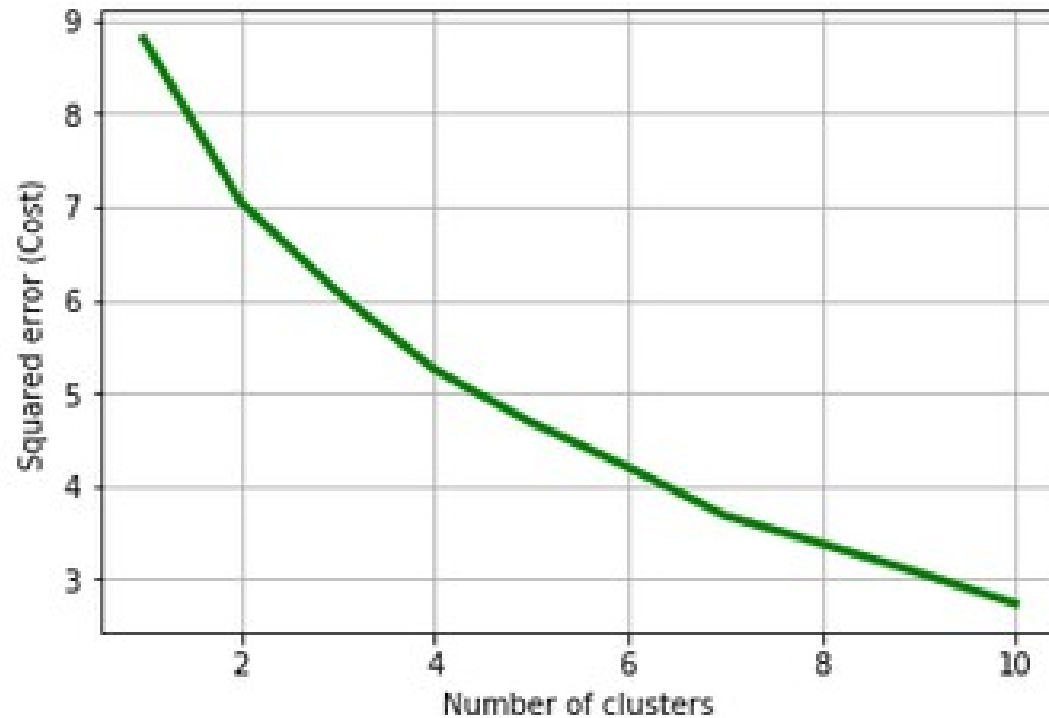
Massachusetts postal code zones
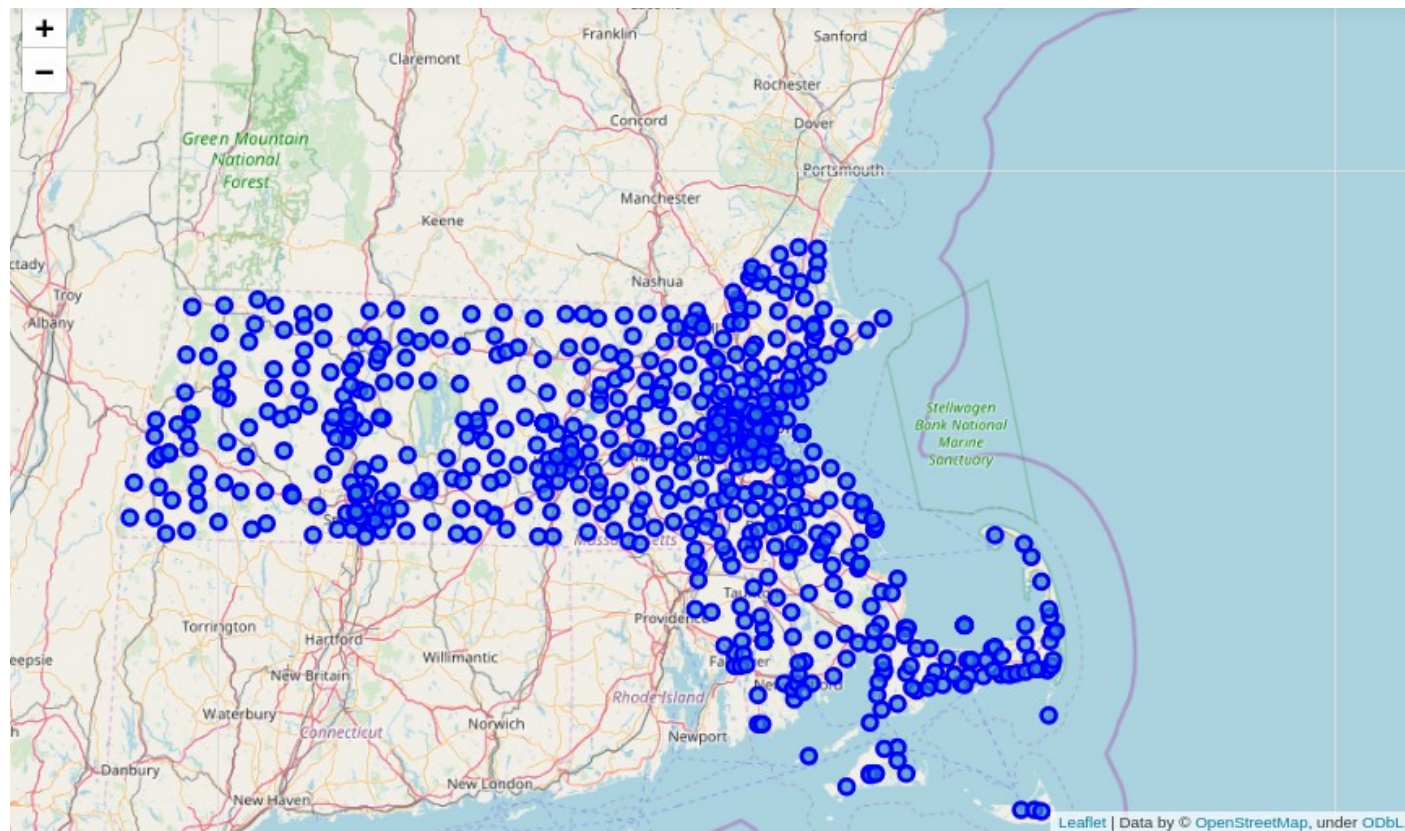
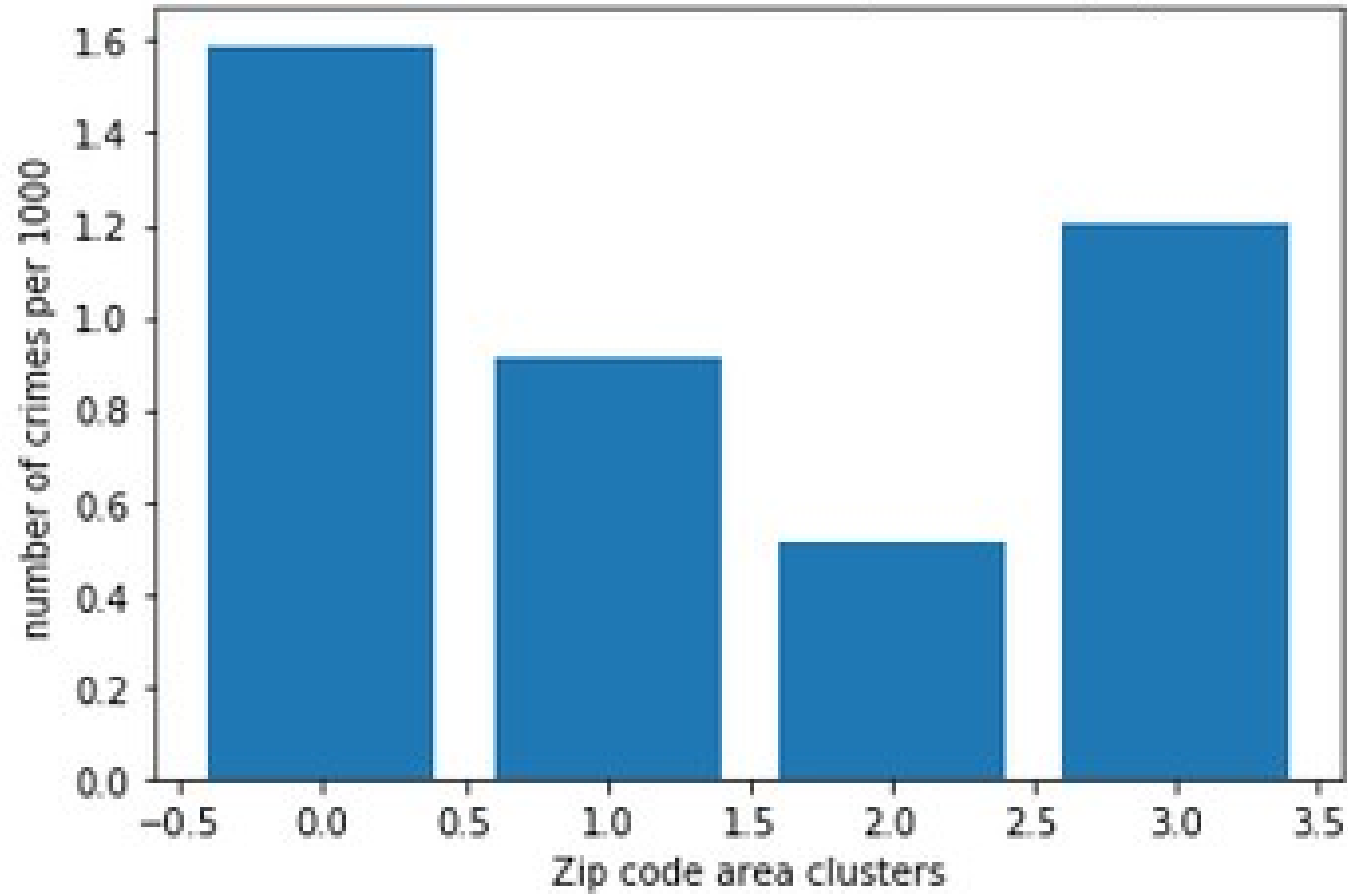# Postal code zones within 15 km of the workplace (green)

Postal code zones within 15 km of the workplace (green)
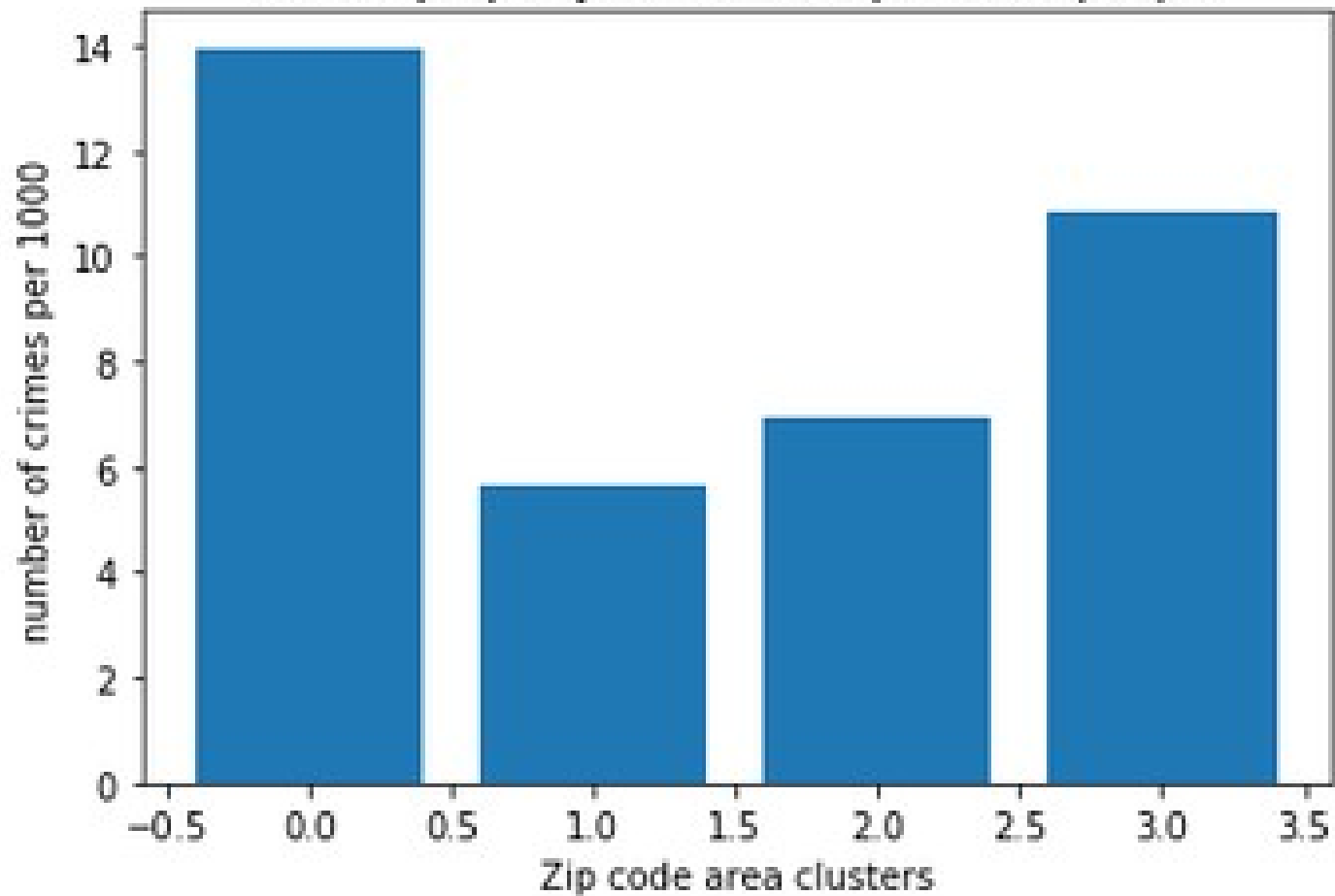
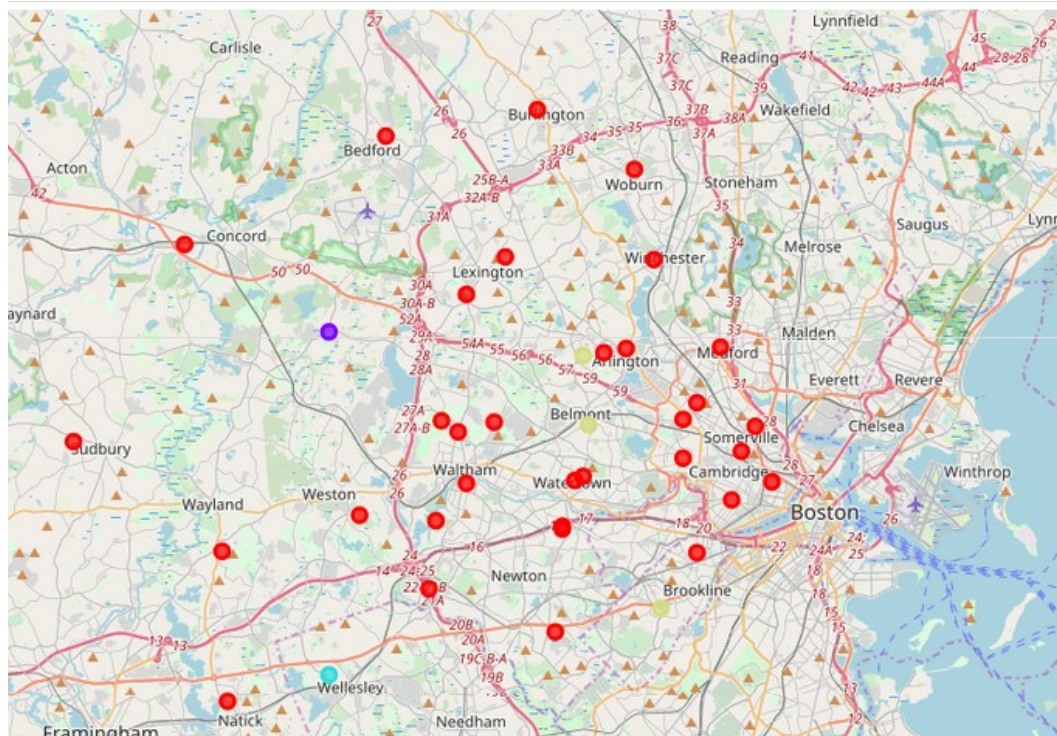# Error plot – for decision on number of clusters

Annual violent crime rates per 1000 people

Annual property crime rates per 1000 people

4 clusters of postal code zones marked on the map

Final selection

	City	Timezone	Daylight savings time flag	Distance	Range	Color	Violent rate	Property rate	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
21	Lincoln	-5	1	4.671661	Yes	green	0.000914216	0.00563767	Museum	Yoga Studio	French Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop
24	Wellesley	-5	1	11.0962	Yes	green	0.000516298	0.00695281	Construction & Landscaping	Yoga Studio	Ethiopian Restaurant	Food Truck	Food & Drink Shop	Food	Flo... Sh...

There were 766 initial postal code zones. After elimination of zones too distant from the workplace, 57 remained. Rates of criminal offenses were available for 39 zones. These were clustered into 4 clusters. Two clusters (one postal code zone in each) were chosen for final selection of place for relocation, after dismissing other two clusters for high crime rates.

Discussion

In this project, I have applied the skills of management of geospatial data, data manipulation and clustering in order to choose the optimal place for living. The resulting number of places to choose from was small, making the choice easier. This was probably possible due to clear setting of priorities, which may vary from person to person. Some may choose a place which is predominantly populated by people of their ethnic community. Others may wish to live very close to their workplace in order to get there on foot and avoid buying a car. Some people may need a more detailed elimination protocol that would include rent rates and proximity to public transport.

The results were surprizing and did not resemble the conventional search that I have performed previously, by asking people about Boston area neighborhoods. I was then recommended a few rich neighborhoods. It is possible that people do not take into account safety data when making a decision on relocation. On the other hand, this data elimination protocol may err due to incomplete or outdated crime data (2013). Additionally, it is possible that despite some differences in crime rates between the postal code zones, the overall crime rates are relatively low, so that people feel equally safe in all zones.

Conclusions

Application of Data Science skills enabled to choose by elimination an optimal place to live. It can be advantageous if we choose to rely on the algorithm in our choice. However, there is a danger in coming to a small number of options if we suspect that some basic assumptions are wrong. In this case it may be better to have a larger output and continue manual search.