CS109A Milestone 3 **Nov 5th**

Predicting Housing Price in City of Boston

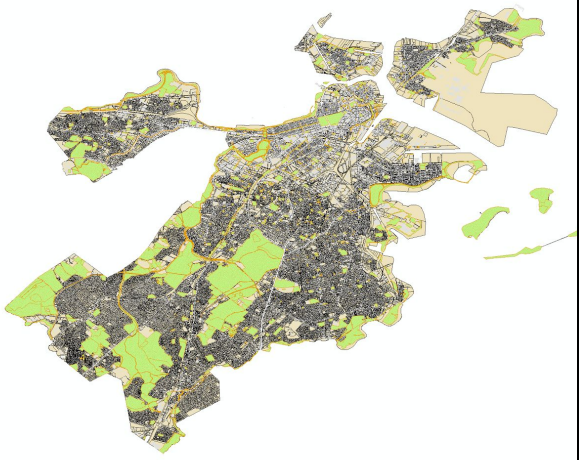
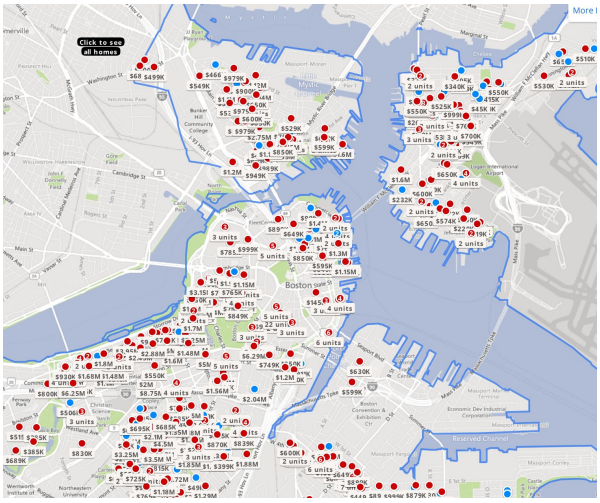
Jia Gu, jgu@gsd.harvard.edu
Ellie Jungmin Han, jhan2@gsd.harvard.edu
Namju Lee, nj.namju@gmail.com

Abstract:

This project aims to ask how what factors contribute to predict the housing price in the City of Boston. Traditionally, statisticians or data scientists has explored top-down(population income, accessibility to transportation,school districts,crime rates etc.) on various instances. However, we think that there is value to test how much bottom-up data such as visual environment can explain housing price and we are curious to what extent it can replace some of the top down data. The data we obtained are housing price, housing types, and year of built using Zillow API; crime rates and school districts from the official Boston government website that have geolocation information.

Data Exploration:

The political boundary of City of Boston is shown in the graph below. It is interesting geographically that there are three separated areas by the ocean.The North west land includes the Logan airport which might be a crucial factor that affects housing price. The southern part includes a diverse neighborhoods of suburban and urban housing. The west land is adjacent Harvard University.

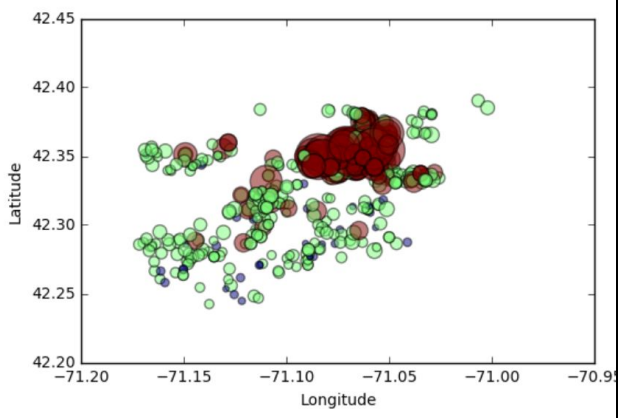
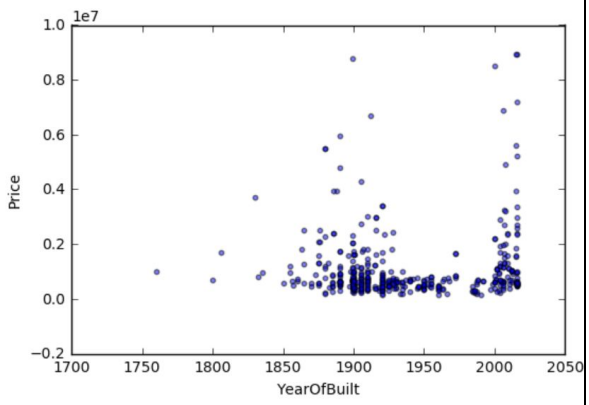
Boston area basemap in ArcGIS (a visualization in map)	*Database for housing price prediction (a visualization in map)*
	

```
data = GetPandasFromFileCSV('bostonHousing.csv')

data.head()
```

Unnamed: 0	fsz	lat	lng	rm	sz	url	use	val	yr	price
0	1288.0	42.286256	-71.120207	NaN	21969.0	http://www.zillow.com/webservice/GetDeepSearch...	Condominium	360741.0	2011.0	$399,000
1	2040.0	42.248300	-71.115951	7.0	8276.0	http://www.zillow.com/webservice/GetDeepSearch...	SingleFamily	629907.0	2000.0	$599,000
2	NaN	42.307030	-71.115430	NaN	NaN	http://www.zillow.com/webservice/GetDeepSearch...	Apartment	NaN	NaN	$644,900
3	3132.0	42.244716	-71.121903	13.0	7046.0	http://www.zillow.com/webservice/GetDeepSearch...	SingleFamily	510706.0	1880.0	$218,000
4	1421.0	42.380374	-71.033872	7.0	1750.0	http://www.zillow.com/webservice/GetDeepSearch...	SingleFamily	339559.0	1900.0	$389,900

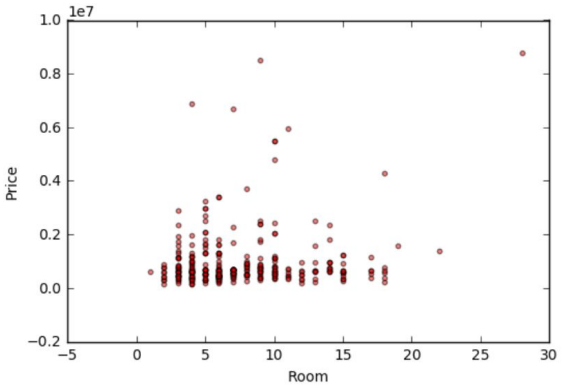
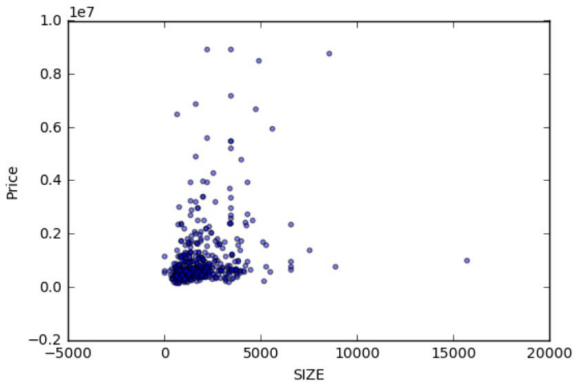
We used Zillow API to obtain housing prices of the greater boston area. The dataset contains longitudinal and latitudinal coordinates , house values and price using Zestimates methodology defined by Zillow. The Zestimate home valuation is Zillow's estimated market value, computed using a proprietary formula. The Zestimate is calculated from public and user-submitted data, taking into account special features, location, and market conditions. We had to keep this in mind for future exploration.

Housing price of different location	*Year of Built and Housing Price*
	

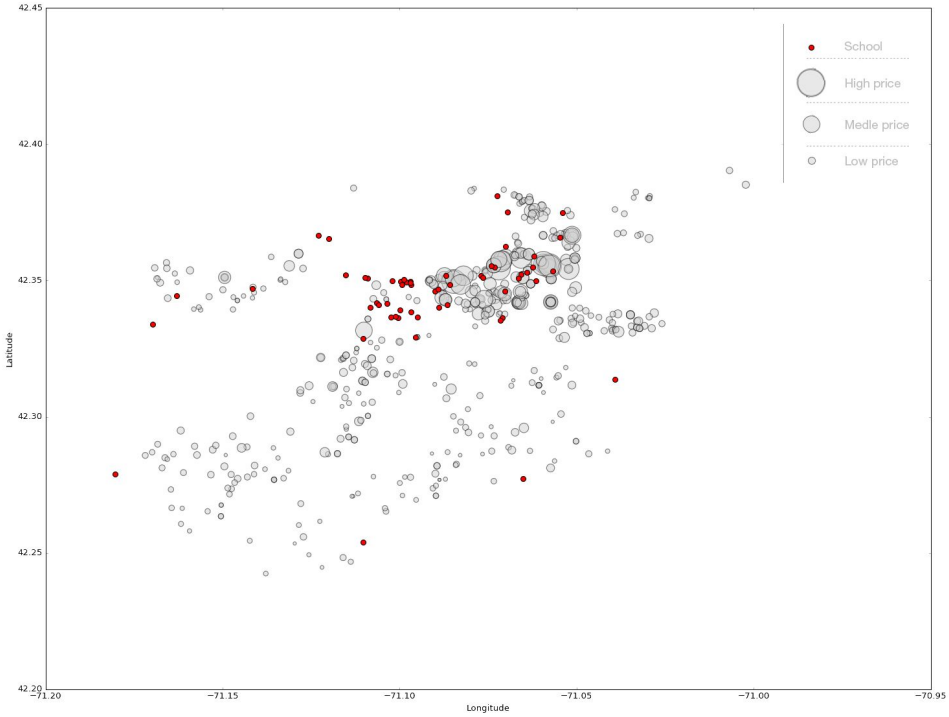
We observe that in Boston downtown and back bay area, the housing price is high as well as in areas that are adjacent to cambridge and brookline neighborhoods. There are sporadic red spots on the graph indicating some clusters of high housing price, which might be affected by local school districts and greenery development. We identified those areas with their coordinates. They are Roslindale Village, Dorchester Centers, Bellevue Hill(adjacent to Roxbury Latin School).

We also noticed that there are interesting patterns compare housing price with their build years. Around 1900 and 2010 there are two periods of expensive development. We hope to dig deeper into this and look at historical events that caused this housing pattern.

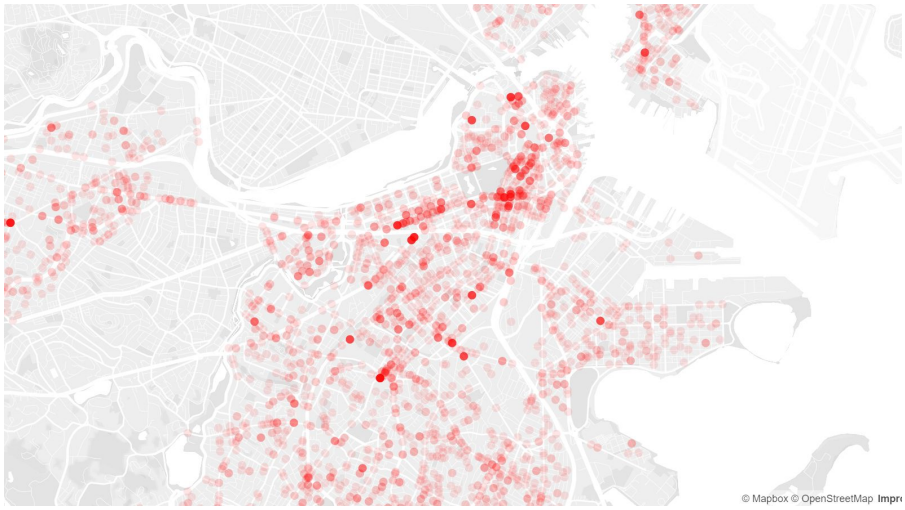
We also explored many other features with housing price, but they are not super useful at this point.

Number of rooms/ Housing Price	Size of House/ Housing Price
	

Housing Price and Schools and Universities data



We also want to incorporate crime data and other data such as closeness to school, green spaces, and closeness to transportation facilities. Below is a visualization of crime rates in the City of Boston. We still needs to post process the closeness to public facilities based on raw data.



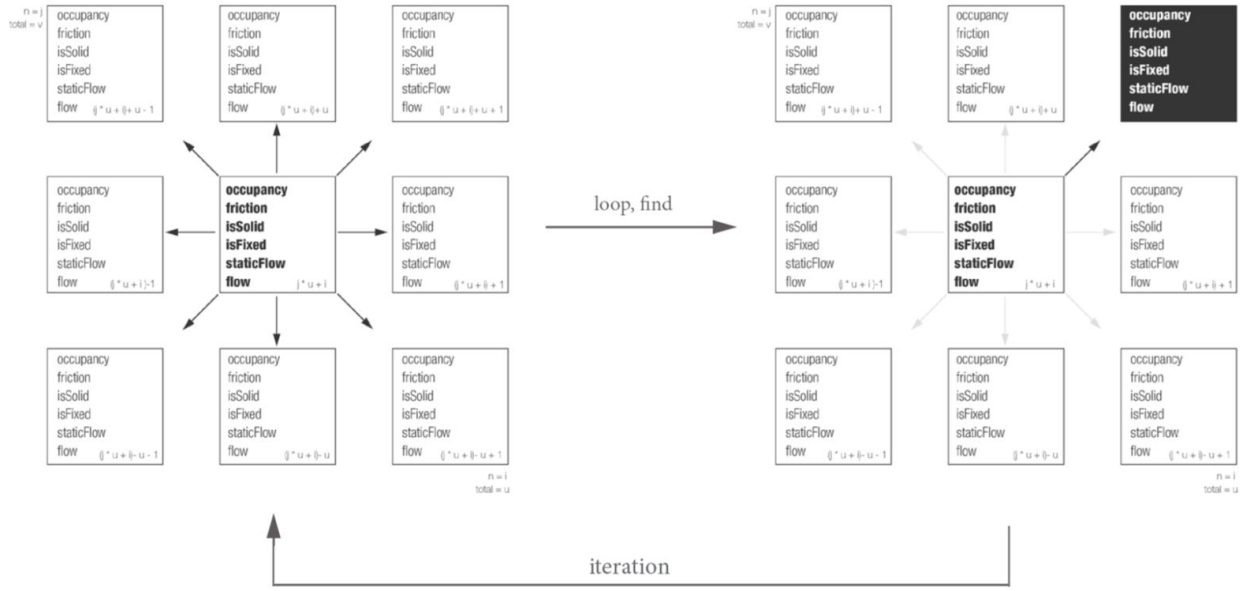
Appendix

Data Processing :

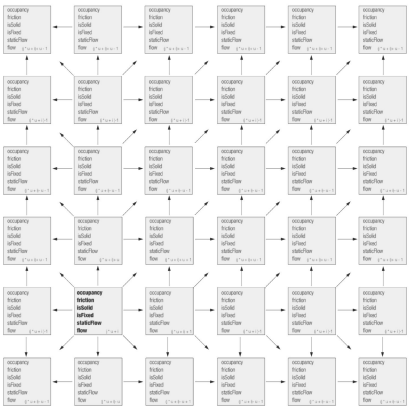
On top of the top down data, the bottom-up data(post processing of google street views) can be map for predicting housing price. This is currently being processed and haven't being analyzed yet.

To process the google street view data, there are two data structures(pixel and graph data structure) where individual data are populated and calculated.Pixel data structure is a matrix, discretizing a urban or district into a finite setting for analysis, in which each pixel has the relationship with its neighbors, and each one computes its own data on the basis of neighbors' settings, so that urban data can be naturally addressed and computed in spatial context.

Pixel data



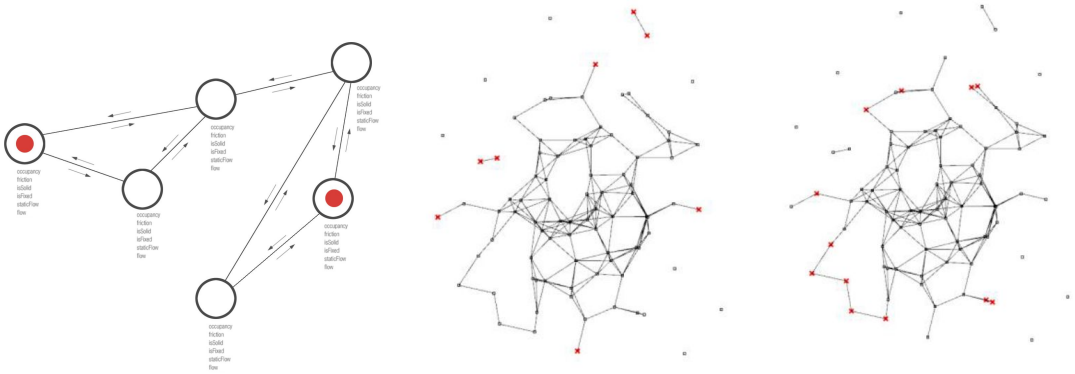
Propagation in pixel data structure



Graph structure is mathematical objects that consist of nodes and edges, and are widely used to represent relational data structures. The street network of urban, street, highway or the subway map are examples of objects whose graphs closely resemble their physical form. Thus, The structure will deploy to process urban data in spatial relationships in order to produce features for the house prediction model.

Data management (Graph structure)

CS109A Milestone 3 **Nov 5th**



Data maganement (location manipulation of given data)