CS 109A
TF: Baijie Lu

How much does visual environment explain the housing price?

Jia Gu, jgu@gsd.harvard.edu
Ellie Han, elliehan07@gmail.com
Namju Lee, nj.namju@gmail.com

Project Proposal:

Housing price is a highly spatial metric that involves multifarious interests including macro economic policies, the development of the area, the local people, and specific houses' features. We define city infrastructural data such as the current housing price, energy consumption of the area, income of neighborhoods, transportation accessibility, green space in the neighborhoods, specific housing features, and crime rates as "top-down data". We also define informal crowd-sourcing data such as, twitter texts, Instagram tags, or yelp reviews as "bottom up data" . There is a third category of data that has not been widely deployed yet which we define as the "holistic visual data". It is the general impression to your visual surroundings when you at a specific spot that we believe can be captured by google street views.

This project aims to ask how do top down, bottom up, and the holistic visual approaches work individually or together to predict the housing price in Massachusetts. Traditionally, statistitians or data scientists has explored top-down and bottom-up data on various instances. However, we think that there is value to test how much visual environment can explain housing price and we are curious to what extend it can replace some of the top down data such as officially registered green space, crime rates record. We think that the visual image some times describes specific location better. For example, a town house is located to a garden a hundred meters away but do not have the sight of the garden because it is blocked by a building versus if it had an open view of the garden will results in different housing price although the distances to the garden are the same.

By testing on the google street views as dependant variables, we hope to capture missing information from the top down data and therefore tell if the visual environment is a significant feature in a housing prediction analytical model. We hope that our model shapes better understanding of how the housing price can be determined by adding more architectural interpretation of different neighborhoods.

Methodology:

We propose to build a multi-variate regression based on the top down and bottom up data first to see how well it predict housing prediction. We use machine learning methodology to pre process Google street views to give certain attributes. This methodology has already been developed by students from Graduate school of Design. We need to obtain the method from them and adapt it for our use. Then we simply use the attributes that we obtained from google street views to build another multi-variety regression and compare the results to the first one.

CS 109A
TF: Baijie Lu

Literature Reviews:

1. Steven C. Bourassa & Eva Cantoni & Martin Hoesli (2007),Spatial Dependence, Housing Submarkets,and House Price Prediction
2. Myungjun Jang a, Chang-Deok Kang(2015), Retail accessibility and proximity effects on housing prices in Seoul,Korea: A retail type and housing submarket approach
3. Peter Sanders a, Mark Zuidgeest b, c, *, Karst Geurs(2015), Liveable streets in Hanoi: A principal component analysis
4. Bin Geng a, Haijun Bao b, *, Ying Liang(2015), A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model

Tentative feature lists:

top-pdown data (government collected)
numerical data

- housing price
- energy consumption
- income
- crime rates
- population demographic information
- transportation accessibility
- green space (park)
- parcel ID or Zipcode
- housing features (types)

bottom-up data (online crowd sourcing)
social data

- twitter text analysis
- instagram

Wholistic visual data (processed by authors)
google street view, observation

- color palette(landscape and enviroment) as data
- living condition

Time Line

GIS , geo-spatial data

- geo data - Eille
- eleveation
- distance to public transportation - NJ

PRE-PROCESSING
numerical model for emerging data

- pixcel data structure - NJ

CS 109A
TF: Baijie Lu

- shape anlytical mode - NJ
-

PRESENTATION

- web --NJ
- data visualization - NJ - Eliie
- video - NJ
- math model

Milestone #3 due Tues, Nov 1:

Data Exploration. Collect the relevant data and submit a document with 10-12 visuals and 1 to 2 page write-up of key findings. Some summary measures should be submitted as well; be smart to only include what helps you explain what has been noticed.