**The spatial distribution and the spatial auto correlation of the house price (buying price) in the city of Buffalo**

Introduction:

1.Package: ggmap, rvest, ape

2.Data source:

1) [www.trulia.com](http://www.trulia.com) (one of housing price website in the United States)

2) Google Map API

3.The method of getting data:

1) using the package(rvest) to grab the house data from the [www.trulia.com,](http://www.trulia.com,) because it will take a long time if you get the record from the website by inputting to the table by yourself. So the package(rvest) is a good choice to get the data in the HTML file by identifying the CSS selectors where the data stores in. From the [www.trulia.com,](http://www.trulia.com,) I want to get the attributes about the address, price, and the coordinates. The address and the price can get from the HTML page directly by the package. But the coordinates in the website are also from the Google map API, in order to get the coordinates of houses, I use the function geocode()\* from the package(ggmap). Inputting the address name into the function geocode(), and the function will send the requires to the Google map API, and then download the coordinates(lon,lat) of the address automatically. Eventually, joining the columns of address, price, lon and lat together to build the final data table(Buffalo\_House\_Price).

2) Keeping the same coordinate system between Buffalo\_House\_Price table and the boundary of Buffalo city. Plotting the house price distribution result(Not done yet).

4. Analyzing the spatial auto correlation of the house price by the package(ape) and calculate the Local Moran’s I by R(Not done yet):

wps_clip_image-6157

Plotting the LISA plot to evaluate the cluster of the houses price in Buffalo city.

Geocode()\*: the function can download the coordinates 2500 times a day.