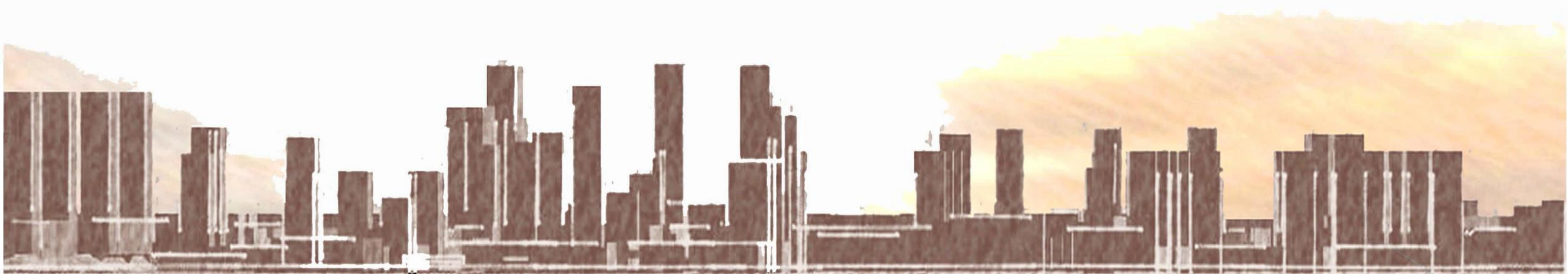
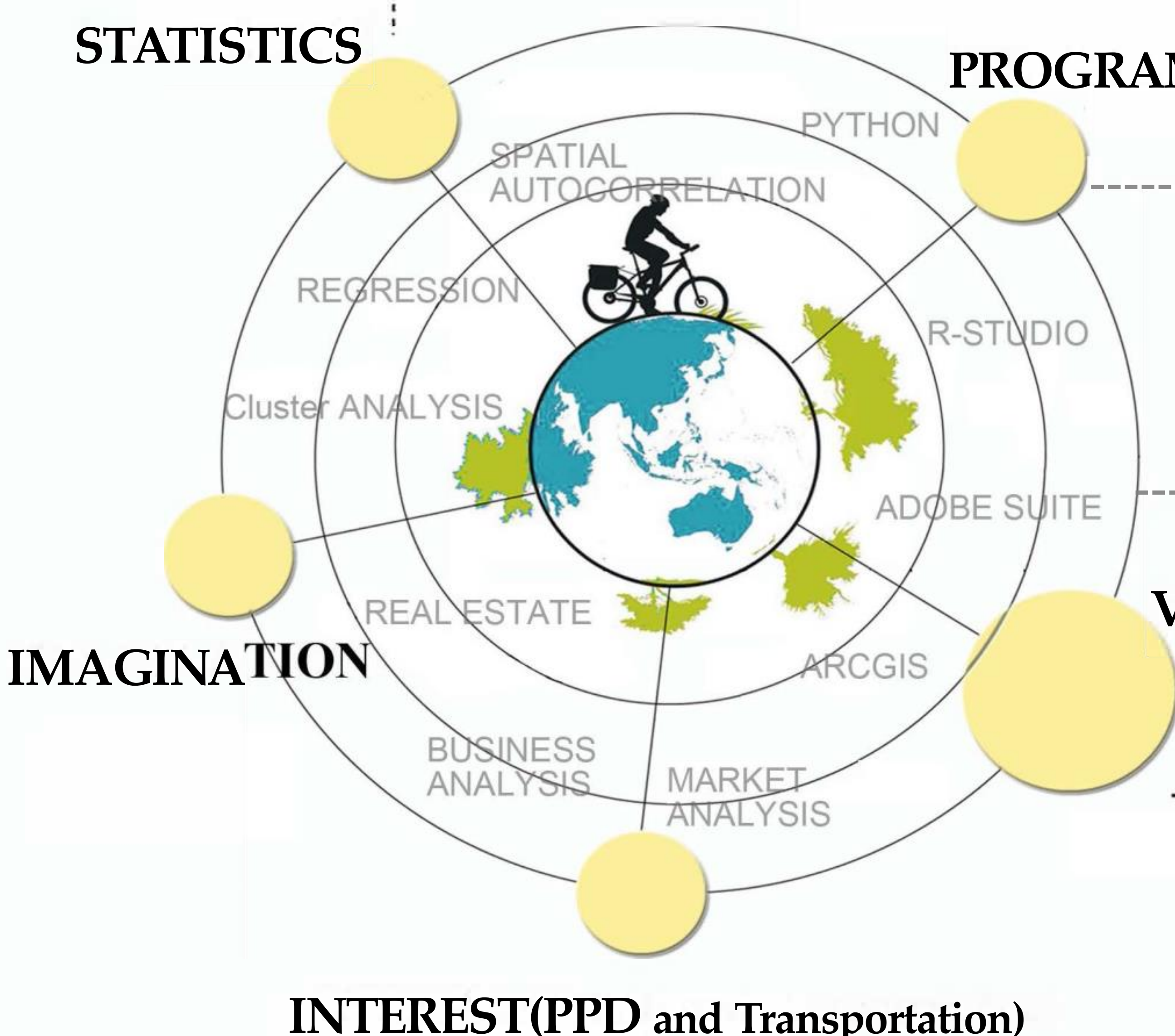
PORTFOLIO OF XIE JIE JUN

MASTER OF CITY PLANNING@UNIVERSITY OF PENNSYLVANIA, USA

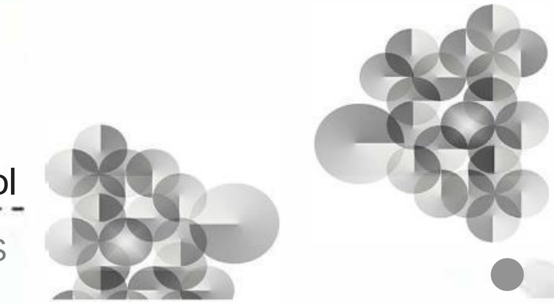


SELECTED WORKS OF XIE JIE JUN: **From a Designer to a Data Analyst**



Geospatial Software Design: Point Pattern Spatial Independece Analysis Tool

Techniques: Python programming for ArcGIS



Spatial Autocorrelation Test of Median Income in Philadelphia

Techniques: R studio, Geoda

Calculation of Retail Site Suitability and Trade Area Analysis

Techniques: ArcGIS. Methods: Huff Model

Hedonic Regression and Willingness to Pay For Transit Access in Philly

Techniques: R studio, ArcGIS

Other works

Techniques: Python, Java, Flex

LARP 743 Geospatial Software Design
Final Project: Point pattern Spatial Independence Analysis Tool
Instructor: Professor Dana Tomlin
Student : Jiejun Xie

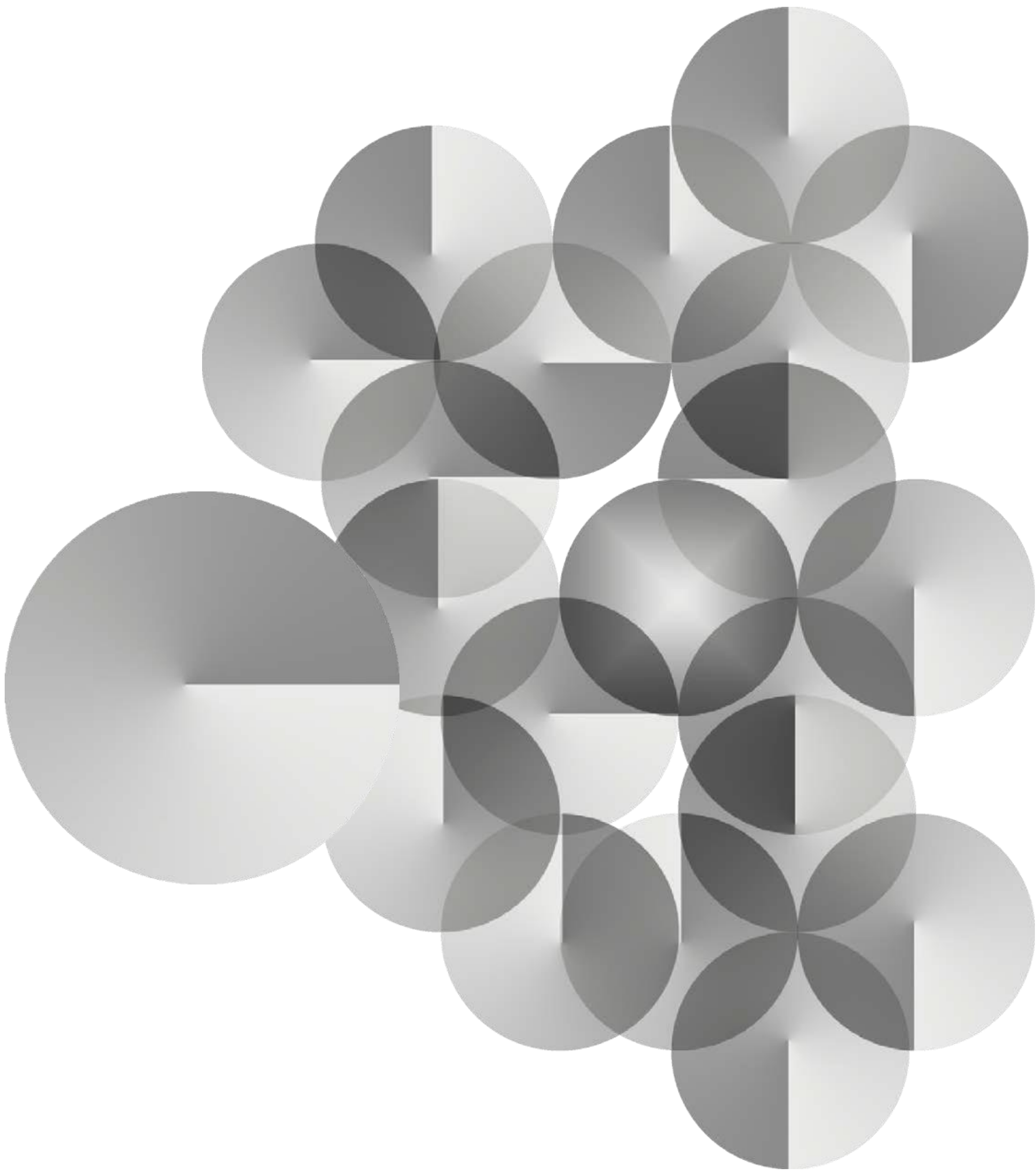


Image: abstract point data clustering pattern

1. Purpose

Some business tend to locate near each other (like Starbucks and Dunkin' Donuts), but others tend to disperse from each other(like Walmart will have a repulsive effect on other supermarkets). In this project, we can use this tool to do point pattern analysis to test whether two point patterns are randomly distributed, clustered or dispersed.

Disperse

Cluster

Graffiti and Crime

2. Assumption

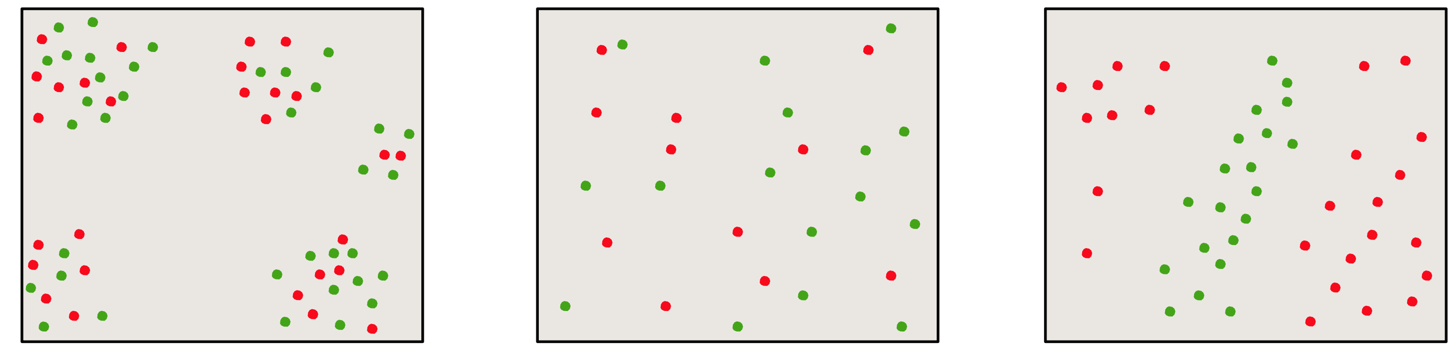
In order to conduct the analysis, we have several assumption:
1.In order to test whether there is a clustering point pattern for two point types, we should at least have 15 points for each type, which make us have at least 30 observations in total. (made by myself)
2.The analysis can just output global results. If there is any local clustering pattern, we can not find it by this script. This is also one of the most important limitations of our tool.

4. Definition of Spatial Independence for two point patterns

CASE	HYPOTHESIS FRAMEWORK
One Pop	Clustering ← Spatial Randomness → Dispersion
Two Pops	Attraction ← Spatial Independence → Repulsion

LARP 743 Geospatial Software Design
Final Project: Point pattern Spatial Independence Analysis Tool
Instructor: Professor Dana Tomlin
Student : Jiejun Xie

Spatial Relationship of Two Point Patterns



Attraction Spatial Independence Repulsion

MY Definition of Spatial Independence:

First, Red and Green points can be anything represents their location.They can be crime locations, business locations, public facility locations or even plants locations .

Attraction: Red and Green points would like to locate near each other.

Spatial Independence: Red and Green points would locate randomly and do not seem to have any relationship with each other.

Repulsion: Red and Green points would like to locate futher away from each other.

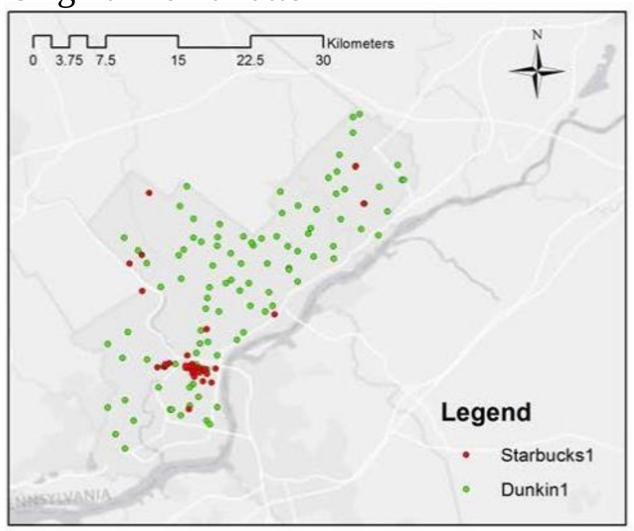
5. Methodology

In this script, I want to learn from the ArcGIS NNI method, but my script is to run for two point pattern. I will first ask each red points to calculate the distance of their nearest green neighbor(ArcGIS “Near” method), then for all the red points, calculate their average distance to their nearest neighbor(set as x). Then, merge the red and green points and run permutations for the following steps: For each permutation, random select certain number of points as red and others as green.Use “Near” to calculate their average nearest distance and then calculate the mean, create a list and append the mean to list. After permutation, by checking where x stands in the list, we can see whether the original point pattern is spatially independent or not.

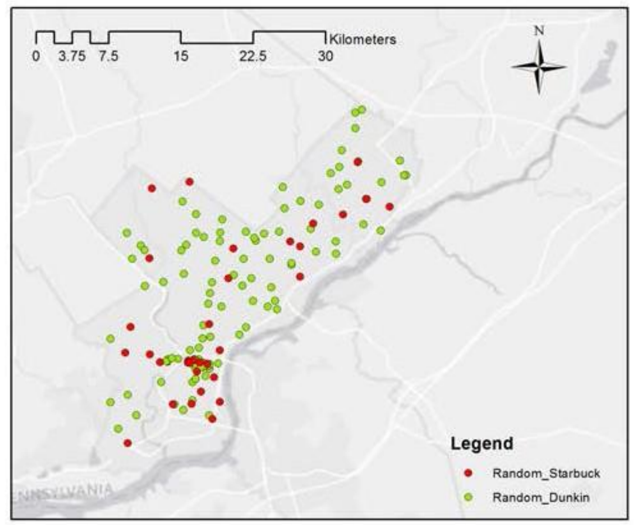
Example

1.Starbucks and Dunkin Donuts in Philadelphia

Original Point Pattern



Randomized Point Pattern



Print Result

```
Completed                                              Close
                                                       << Details
Close this dialog when completed successfully

1275.1740652275366, 908.99790290702435, 1167.8453680500033, 1169.4224385284242]

19
have created the field RndValue with random numbers

The average distance for the random pattern is 1094.59370263
The average distance for the random pattern is [1031.8611751585058, 1091.2885619991971,
1147.3255829860425, 1020.4672241591273, 1340.6137652372877, 1018.7843424099364,
983.35358597383947, 1109.5401698055211, 1149.6728242007271, 902.57383742458489,
1240.1508759266578, 1082.5835640200576, 1080.5436462549151, 1259.7631703384793,
1275.1740652275366, 908.99790290702435, 1167.8453680500033, 1169.4224385284242,
1094.5937026267572]

20
have created the field RndValue with random numbers

The average distance for the random pattern is 1269.6029354
The average distance for the random pattern is [1031.8611751585058, 1091.2885619991971,
1147.3255829860425, 1020.4672241591273, 1340.6137652372877, 1018.7843424099364,
983.35358597383947, 1109.5401698055211, 1149.6728242007271, 902.57383742458489,
1240.1508759266578, 1082.5835640200576, 1080.5436462549151, 1259.7631703384793,
1275.1740652275366, 908.99790290702435, 1167.8453680500033, 1169.4224385284242,
1094.5937026267572, 1269.6029354037094]
```

Compare original average distance with randomized ones

original	570.627
1	1031.861
2	1091
3	1147
4	1020
5	1340
6	1018
7	983
8	1109
9	1149
10	902
11	1240
12	1082
13	1080
14	1259
15	1275
16	908
17	1167
18	1169
19	1094
20	1269

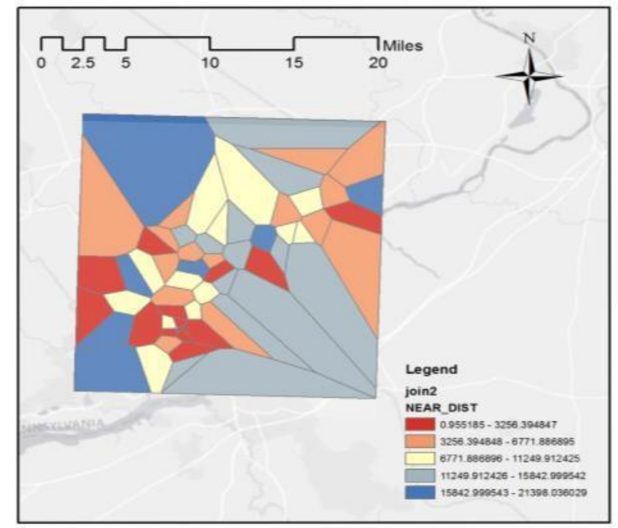
Conclusion: P=0. Thus, we have a clustering pattern for starbucks and Dunkin Donuts in Philadelphia.

Their spatial relationship in Philadelphia is : Attraction.

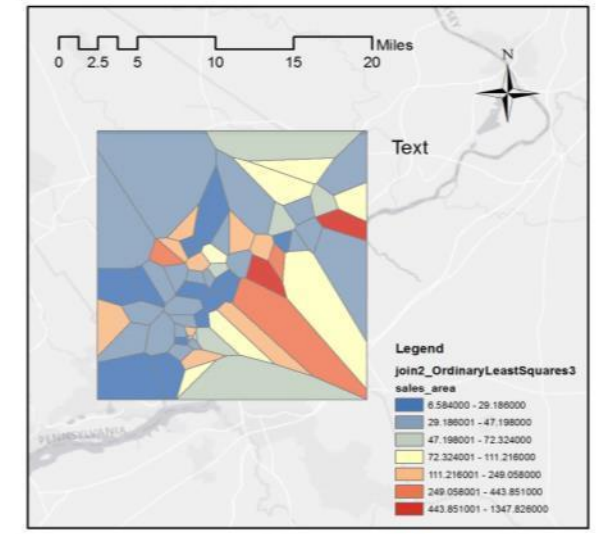
LARP 743 Geospatial Software Design
Final Project: Point pattern Spatial Independence Analysis Tool
Instructor: Professor Dana Tomlin
Student : Jiejun Xie

Testing the benefit of Attraction

Choropleth Map of distance to nearest Dunkin Donuts in each Starbucks PTA



Choropleth Map of in each sales per square foot in each Starbucks PTA



The Map is the primary Trade Area for Starbucks created by ArcGIS tool "create thiessen polygon " . We can see that there are some similarities with the two choropleth maps.
Then, we do OLS regression to test whether the Near_Dist value is significant for sales of each Starbuck.The result is showed below.

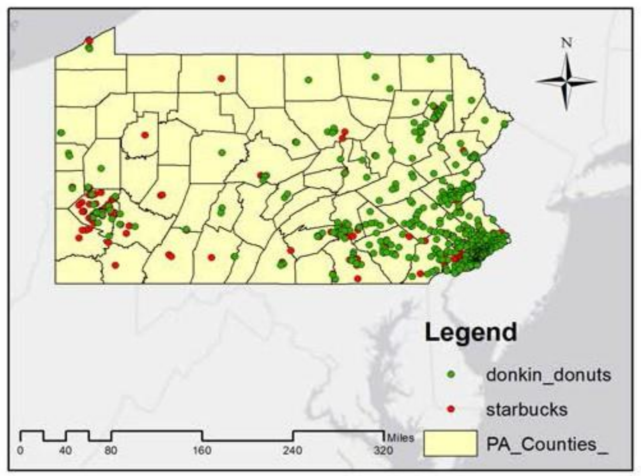
```
Start Time: Wed Dec 17 17:00:58 2014
Running script OrdinaryLeastSquares...
                                    Summary of OLS Results
                 Variable    Coefficient       StdError t-Statistic Probability      Robust_SE  Robust_t Robust_Pr
VIF [1]
                Intercept -535914.270399 3110319.386977   -0.172302   0.863810  1229143.411555 -0.436006 0.664482
-----
             JOIN2.POP2010N       17.677619       11.373525     1.554278    0.125657        13.793147  1.281623 0.205165
1.404464
            JOIN2.AVG_MEDIAN   -41073.632841    56785.063254    -0.723318    0.472440     42618.661340 -0.963748 0.339240
2.510474
            JOIN2.AVG_PCTWHI    12520.186110    14345.196372     0.872779    0.386442      8013.139608  1.562457 0.123719
4.557702
            JOIN2.AVG_MDHHIN       39.644808       34.905005     1.135792    0.260796        22.373060  1.771989 0.081744
5.092588
            JOIN2.AVG_PCTBAC   -26932.699095    21904.675672    -1.229541    0.223921     18229.427845 -1.477430 0.145068
5.111583
            JOIN2.AVG_PCT_PU 4076960.084221 4144909.200197       0.983607   0.329463  2221810.721286  1.834972 0.071732
5.770464
            JOIN2.SUM_EMPLOY        0.124065        7.760325     0.015987    0.987300         4.029794  0.030787 0.975546
3.250504
DUNKIN_GENERATENEARTABLE.NEAR_DIST     -63.258500       33.562554   -1.884794    0.064561        47.799764 -1.323406 0.190986
1.213295                                    OLS Diagnostics

Input Features:                    join2   Dependent Variable:                              JOIN2.SALES_VOL
Number of Observations:               66   Akaike's Information Criterion (AICc) [2]:        2074.248405
Multiple R-Squared [2]:         0.164164   Adjusted R-Squared [2]:                              0.046854
Joint F-Statistic [3]:          1.399404   Prob(>F), (8,57) degrees of freedom:                 0.216637
Joint Wald Statistic [4]:       5.872416   Prob(>chi-squared), (8) degrees of freedom:          0.661521
Koenker (BP) Statistic [5]:     8.550162   Prob(>chi-squared), (8) degrees of freedom:          0.381651
Jarque-Bera Statistic [6]:   4340.753620   Prob(>chi-squared), (2) degrees of freedom:          0.000000*
```

The p value for the Near_Dist to Donkin Donuts is 0.06, which is the highest of all the predictors. Thus, futher proof that there is attraction between the two business types. However, this step is just for testing the clustering benefit, but not generated from the script.

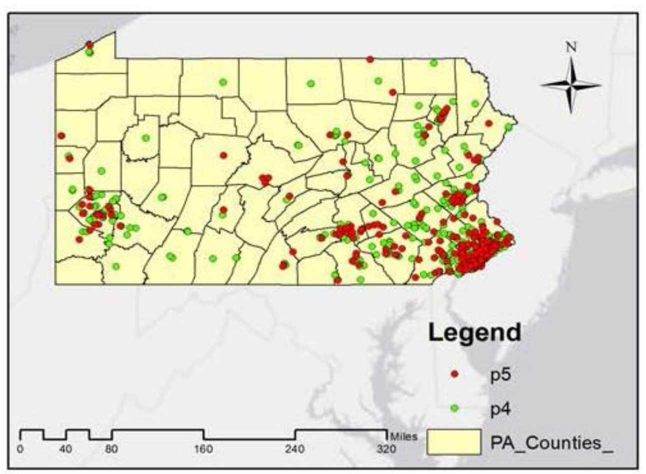
LARP 743 Geospatial Software Design
Final Project: Point pattern Spatial Independence Analysis Tool
Instructor: Professor Dana Tomlin
Student : Jiejun Xie

Example 2 : Starbucks and Dunkin Donuts in Pennsylvania

Original Point Pattern



Randomized Point Pattern



Compare original average distance with randomized ones

Original	5056.445
1	3764.92
2	4063.66
3	4368.51
4	4384.57
5	4423.06
6	4429.44
7	4556.93
8	4560.56
9	4569.46
10	4691.26
11	4699.87
12	4785.25
13	4791.82
14	4865.19
15	5097.64
16	5342.79
17	5625
18	5683.57
19	6096.45
20	6323.64

```
The average distance for the random pattern is 5624.99590933
The average distance for the random pattern is [5097.6382527741189,
4423.0578392622583,      5683.5661698803215,      4569.4593147100468,
4560.5619033064249,      4384.5661619123575,      4865.1926762544626,
4791.8227112052718,      5342.7924594249716,      6323.6374166163105,
4699.8654619192566,      4063.6596751490983,      4368.5076057833885,
3764.9246783646818,      4785.2459216835305,      4556.9288962151077,
4429.4389522523861,      4691.2593611899702,      6096.4464236140311,
5624.9959093308935]
 The value is 21
 The value is 14
The p-value is 0.666666666667
```

Conclusion: P=0.66. Thus, we have a quite random pattern. Their spatial relationship in Philadelphia is :Independence.
When we change the scale, the result is quite different. In state level, there is no spatial attration for the two point patterns. Thus, it is one of this tests limitation: we can not find there is local cluster under a global randomness.

Code

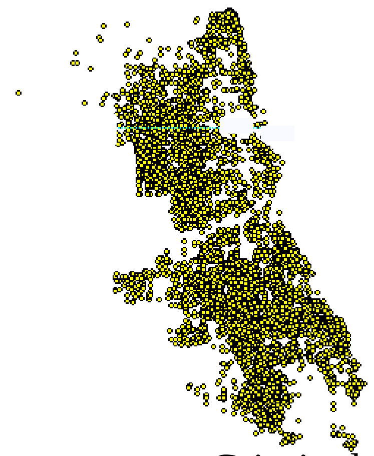
```
# Import external modules
import sys, os, string, math, arcpy, traceback
from arcpy import env
from arcpy.sa import *
try:
   # Read user-specified name of input shapefile and echo it
   input1 = arcpy.GetParameterAsText(0)  # the first input point file
   input2 = arcpy.GetParameterAsText(1)  # the other input point file
   output  =  arcpy.GetParameterAsText(2) # merged point file
   relabel  =  arcpy.GetParameterAsText(3) # randomlized point file
   outputl  =  arcpy.GetParameterAsText(4)     # relabled selected point file
   output2 = arcpy.GetParameterAsText(5)     # relabled second point file
   output3 = arcpy.GetParameterAsText(6)     # relabled point file with the near dist data

   # Step 1:  Process Near analysis for the two input point file
   arcpy.Near_analysis(input1, input2, "", "NO_LOCATION", "NO_ANGLE")

   # update the near result table, loop through "Near_DIST" table and calculate their mean
   distance. enumerationOfRecords = arcpy.UpdateCursor(input1)

   Firstdist = 0
   n = 0
   for nextRecord in enumerationOfRecords:
      Thisdist = nextRecord.getValue("NEAR_DIST")
      Firstdist = Firstdist + Thisdist
      n = n+1
   # calcuate their average distance and echo
   it mean = Firstdist/n
   arcpy.AddMessage(" The average distance for the random pattern is " + str(mean))

   # Step 2: Merge the two point files and get the output
   arcpy.Merge_management([input1,input2], output)

   # count features in input1, input2, output, and echo
   it ft = arcpy.GetCount_management(input1)
   ft2 = arcpy.GetCount_management(input2)
   total = arcpy.GetCount_management(output)
   arcpy.AddMessage( '\n' + "number of features in input1 is " + str(ft) )
   arcpy.AddMessage( '\n' + "number of features in input2 is " + str(ft2) )
   arcpy.AddMessage( '\n' + "number of features for the two files is " + str(total) )

   # Create a empty list to store the result we get
   below mean_list = []

   # Step 3: iniatiate a permutation for the merged
   file permutation=0
   iteration = 2
   for i in range(iteration):
```

LARP 743 Geospatial Software Design
Final Project: Point pattern Spatial Independence Analysis Tool
Instructor: Professor Dana Tomlin
Student : Jiejun Xie

Graffiti and Crime in Chicago

Graffiti

Narcotics



Criminal

Print Result

```
The average distance for the random pattern is 201.205813218
The average distance for the random pattern is [200.70672430724991,
197.8724644711796, 197.62843805527532, 195.38205770395524,
196.86373348903601, 198.37402029973944, 199.65491488025455,
201.427640948507981, 198.13150665443851, 200.55846668741066,
197.02897493209264, 198.08962232625732, 197.98699124865669,
195.7569262032622, 196.78199168054729, 202.06019998500619,
201.61949045336257, 200.75713964418512, 201.01328115137227,
201.2058132176515]
The value is 21
The average distance for the original pattern is 200.706724307
```

Conclusion:

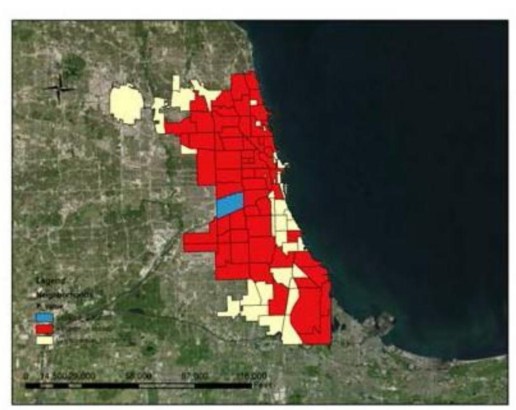
The P value for graffiti and criminal spatial randomness is 0.6195>0.05. Meaning we the two point pattern do not have spatial Attraction in Chicago.

We have the similar results for graffiti and Narcotics, there is no global spatial attraction for the two point patterns.
However, we can see the graph on the right, which is a local spatial clustering p-value result calculated from R-studio(in MUSA 507 course). The blue area all have p value less than 0.05, which means attration. Thus, further indicates that for our script, there is a lot of limitation in order to find a how the spatial attraction behave differently in different areas.

original	201.206
1	195.382
2	195.757
3	196.782
4	196.864
5	197.029
6	197.628
7	197.628
8	197.987
9	198.09
10	198.132
11	198.374
12	199.655
13	200.558
14	200.707
15	200.757
16	201.013
17	201.206
18	201.428
19	201.619
20	202.06

Graffiti and Narcotics relabeling result
(from MUSA 507 homework assignment)

Graffiti and Criminal clustering neighborhood
(from MUSA 507 homework assignment)



Code

```
for i in range(iteration):
    permutation = permutation +1
    arcpy.AddMessage('\n'+ str(permutation))

    # we need to create a field called RndValue, and put random numbers in the field
    fieldName = "RndValue"
    expression = "arcgis.rand('Integer 0 1000')" # is it better to set the maximum value?

    # Execute AddField
    arcpy.AddField_management(output, fieldName, "LONG")

    # Execute CalculateField
    arcpy.CalculateField_management(output, fieldName, expression, "PYTHON_9.3")

    # echo that we have already created a field with random numbers
    arcpy.AddMessage("have created the field " + fieldName + " with random numbers \n")

    # Sort the input shapefile's records in descending order of value in the specified
    field arcpy.Sort_management(output,relabel,[[fieldName,"DESCENDING"]])

    arcpy.MakeFeatureLayer_management(relabel,"MyLayerObject") #

    # Process: Select Layer By Attribute
    # select the first "ft" number of figures from the table and export.(the same amount as input1)
    arcpy.SelectLayerByAttribute_management("MyLayerObject", "NEW_SELECTION", "\"FID\"<"+str(ft))
    arcpy.CopyFeatures_management("MyLayerObject",outputl)

    # select the left number of features from the table and export as output2( the relable the point file)
    arcpy.SelectLayerByAttribute_management("MyLayerObject", "NEW_SELECTION", "\"FID\">="+str(ft))
    arcpy.CopyFeatures_management("MyLayerObject",output2)

    # Similar to step 1, Process: Near, table to use
    # what is different is that we can generate many results as the permutation goes
    on arcpy.Near_analysis(outputl, output2, "", "NO_LOCATION", "NO_ANGLE")
    arcpy.Copy_management(outputl, output3)

    # Then as the first step above, update cursor for the near result
    table enumerationOfRecords = arcpy.UpdateCursor(output3)

    # Loop through that enumeration, calculating the mean of the
    Near_Dist sumdist = 0
    m = 0
    for nextRecord in enumerationOfRecords:
        Thisdist = nextRecord.getValue("NEAR_DIST")
        sumdist = sumdist + Thisdist
        m = m+1

    meandist = sumdist/m
```

8. Discussion and Conclusion

Although the script worked well to test local spatial independence, it did have some limitations:

1. Firstly, it can not test the local spatial independence if we choose to test on a broader region. Most of the time, we will meet with a situation that points might cluster in a special region, and when they behave different within regions, problem would arise. Thus, it is very important to find this regional difference.

2. The enumeration time could be adjusted in the script, but could not be adjusted on the parameter table. Thus, I will try to set it as parameter next time.

3. The list of the calculated distance is just printed on the ArcGIS dialog, I would like to create a table next time and put those numbers in the table.

Conclusion:

Analyzing point pattern Independence is very interesting and meaningful thing for us to do. Since location matters . if we want to make an investment, either on housing or a business, location matters. If we want to choose a place that has lower crime rates, the neighborhood environment matters. If we want to find a place to work for a longer time, location is also important. Thus, it is very important for us to learn how the surrounding environment affect this location in order to make a good decision. That is related to our report this time.

There is still a lot to do in order to better analyze spatial independence for two point patterns. We should keep on working and update with the nearest data and technology in order to make better analysis, creating better tools to facilitate our analysis. Overall, I think it will be an interesting as well as meaningful area to learn in the future.

Also,thanks for these brand showed below who has given me a lot of ideas, as well as the knowledge I learned from MUSA 501 and MUSA 507 combined with this course, which is very interesting and enlightening!


```
        # Loop through that enumeration, calculating the mean of the
        Near_Dist sumdist = 0
        m = 0
        for nextRecord in enumerationOfRecords:
            Thisdist = nextRecord.getValue("NEAR_DIST")
            sumdist = sumdist + Thisdist
            m = m+1

        meandist = sumdist/m
        arcpy.AddMessage("The average distance for the random pattern is " + str(meandist))


        # Append the average distance result to list and echo it
        mean_list.append(meandist)
        arcpy.AddMessage("The average distance for the random pattern is " + str(mean_list))

    j=mean_list
    # count the number of figures in mean-list and plus 1(1 as the original average distance calculated
    above) a=len(j)+1
    arcpy.AddMessage(" The value is " + str(a))

    # In order to find how many figures are smaller than "mean"(average distance for the orignnal point
    data),
    # delete the figure that is greater than "mean" in
    list w = [x for x in j if x < mean]
    # count the number of figures left in mean-list and plus
    1 g = len(w)  # the problem is here
    arcpy.AddMessage(" The value is " + str(g))
    p = g *1.00 /a
    arcpy.AddMessage(" The p-value is " + str(p))

except Exception as e:
    # If unsuccessful, end gracefully by indicating why
    arcpy.AddError('\n' + "Script failed because: \t\t" + e.message )
    # ... and where
    exceptionreport = sys.exc_info()[2]
    fullermessage   = traceback.format_tb(exceptionreport)[0]
    arcpy.AddError("at this location: \n\n" + fullermessage + "\n")
```

"Everything is related to everything else, but near things are more related than distant things."

---------Waldo Tobier

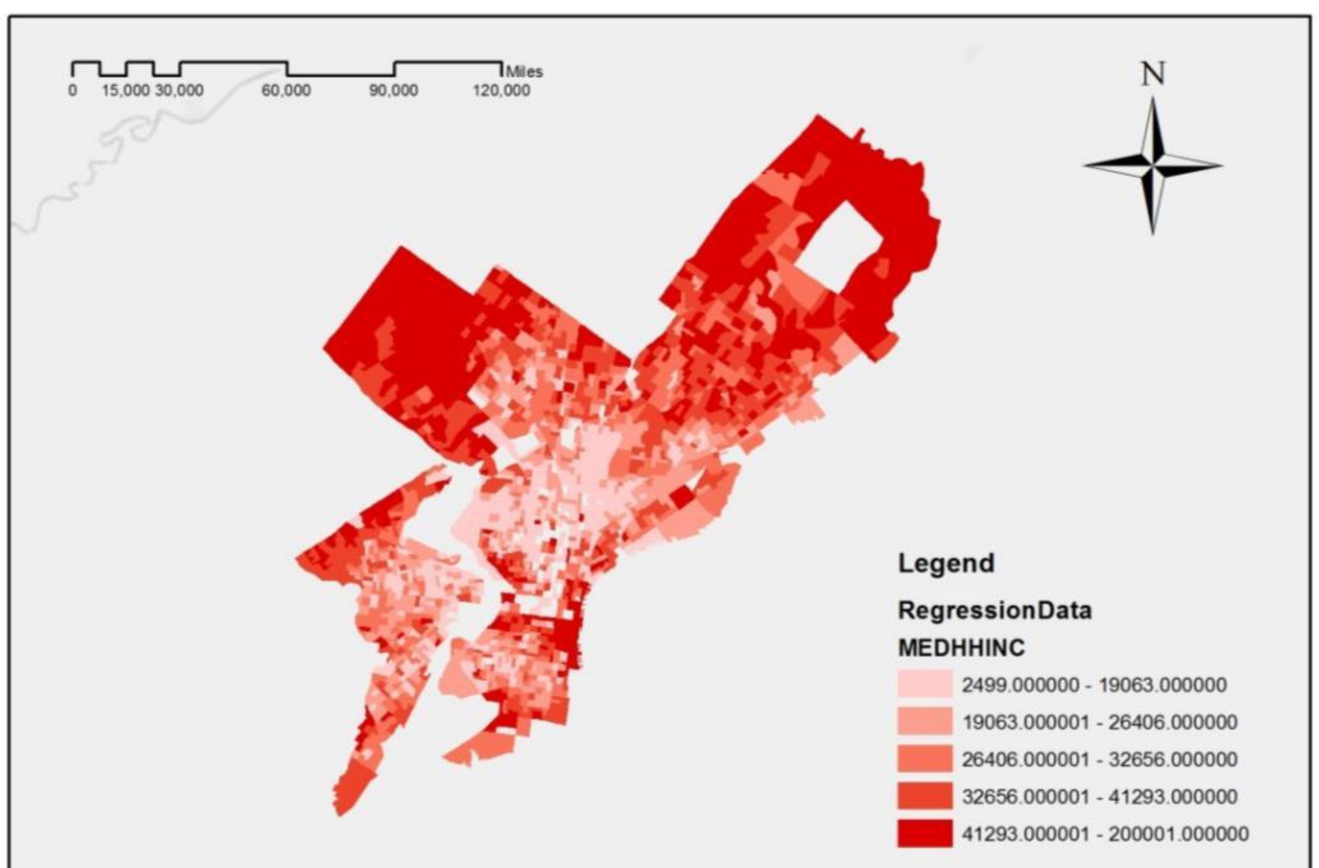
MUSA 501 Spatial Statistics and Data Analysis
Project : Spatial Autocorrelation Test of Median Income in Philadelphia
Instructor: Professor Eugene Brusilovskiy
Student : Jiejun Xie

1. Purpose

We are very interested in whether there is spatial autocorrelation in terms of median income in Philadelphia. By checking the pove1iy cluster areas we can see if there is a strong trend for clustering for different median income levels in Philadelphia.

Theoretically speaking, if there is no government support, rich people will choose to live in a more affluent and well facilitated neighborhood. However, poor people have to choose a less convenient but a more affordable neighborhood. However, this will result in a much vicious circle that the rich who have access to good education as well as other facilities will attract more business to be invested in this area, thus creating more employment. However, the poor who not only do not have good access to education, but might also suffer from the problem of unemployment. This will cause a lot of social problems, thus we have to see whether the problem exists in Philadelphia now and to see how we can change the situation.

Map l. Median income spatial distribution by census tract, Philadelphia (classification method: quantile)



2. Methods

According to Waldo Tobler, the first law of geography is expressed as: "Everything is related to everything else, but near things are more related than distant things." This means, there is always a tendency for the near things to be similar with each other (positive correlation). For example, neighborhoods have higher education level tend to live near each other, and higher housing prices would also tend to cluster. The l st Jaw of geography is the basic premise behind all of spatial statistics (source from slides P 2), also the basic for the spatial autocor relation method.

Moran' s I is an old (1950) but probably the most conu11only used method to test for spatial auto correlation, or spatial dependencies (source from slide Page 14). The formula to calculate Mo ran' s I is showed as below:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(X_i-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

In this formula, X is the mean of the variable;

X_i is the variable value at a specific location,

Xj is the variable value at another location ,
Wij is a weight indexing location of i relative to
j N is the number of observations

Map 2 example for the Moran' s I Formula

a	b	c
d	e	f
g	h	i

Thus, for the nominators, we can see that Wij is the binary weight matrix of the general cross-product statistic (source from http://www.lpc.uottawa.ca/publications/moransi/moran.htm). And when i andj are neighbors with each other, Wij=l; when i and j are not neighbors, Wij=O.

3. Hypotheses Testing Process

In this case, the null hypotheses is: there is no spatial autocorrelation for Median income in Philadelphia. However, the alternative Hypotheses is: There is significant spatial autocorrelation for census tract in terms of median income in Philadelphia.

In order to test whether there is spatial autocorrelation for census tract in terms of median income in Philadelphia, we need to test whether the result of Moran's I for our original Median Income is significantly different from the result of Moran's I when there is no spatial autocorrelation. Thus, we need to randomly shuffle the values of the Median Income variable (make them randomly distributed), then calculate Moran's I for each shuffled map. Then, we need to repeat the shuffling process for 999 times, so we can have 999 observations for a randomly distributed result of median income, and calculate Moran's I for each new permutation. With the Moran's I result we calculated above, we need to put all the 1000 results in a descending order, and to see where the Moran's I value for the observed Median Income variable stands, compared to the other 999 random shuffled ones. Finally, we can get the p-value with the formula: (rank of the Moran's I for our original pattern) / 1000.

Overall, if the result is smaller than 0.05, we can reject the null hypothesis of No spatial autocorrelation, which means there is significant spatial autocorrelation

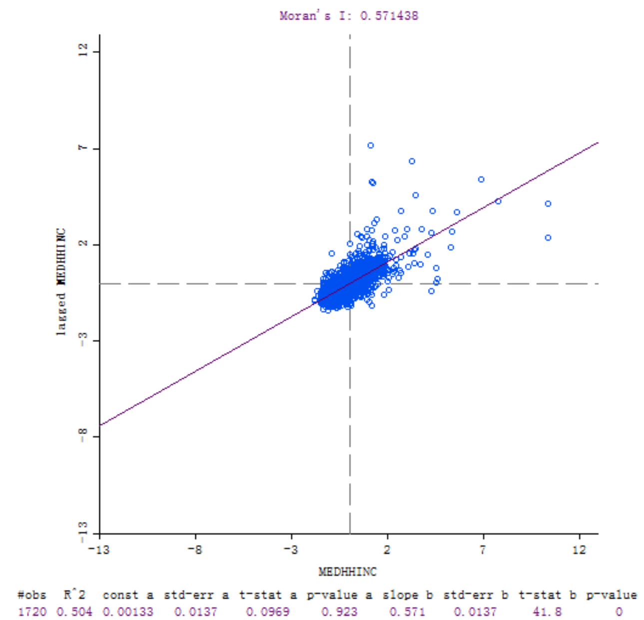
4. Results

Table 4. Moran's I result.

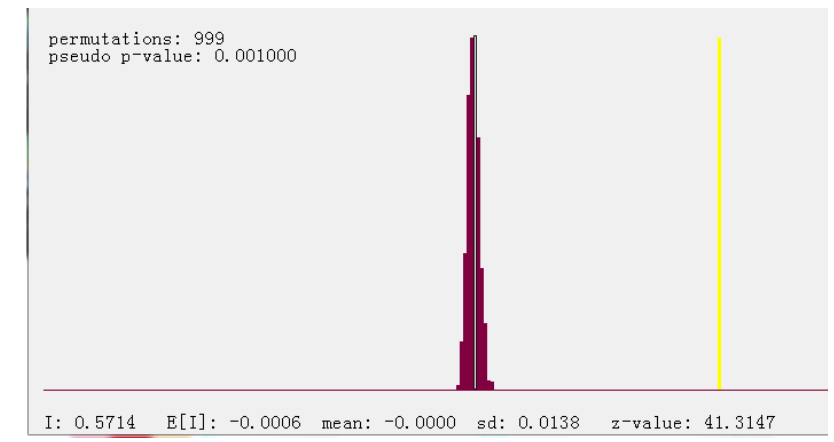
	Moran's I	random permutations	P-value
Queen Neighbor	0.571438	999	0.001
5 Nearest Neighbor	0.545142	999	0.001

MUSA 501 Spatial Statistics and Data Analysis
Project : Spatial Autocorrelation Test of Median Income in Philadelphia
Instructor: Professor Eugene Brusilovskiy
Student : Jiejun Xie

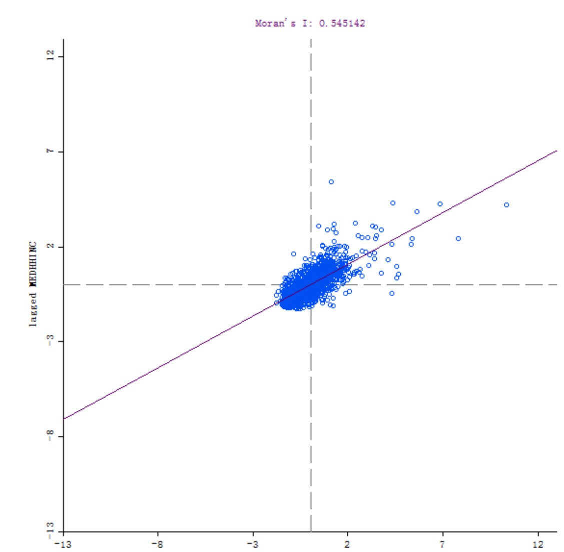
Map 4. Moran's I results using Queen Neighbor weight matrices.



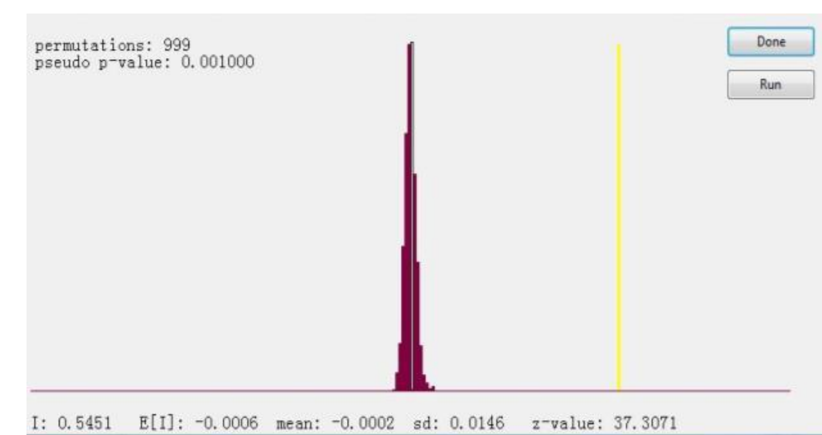
Map 5. Significance Testing for the Moran's I (Using Queen Neighbor weight matrices)



Map 6. Moran's I results using 5 nearest neighbor weight matrices.



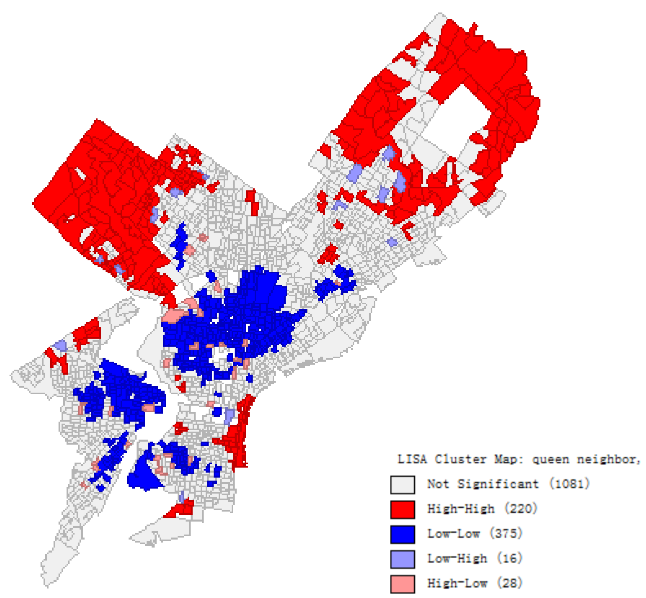
Map 7. Significance Tests for the Moran's I (Using 5 Nearest Neighbor weight matrices)



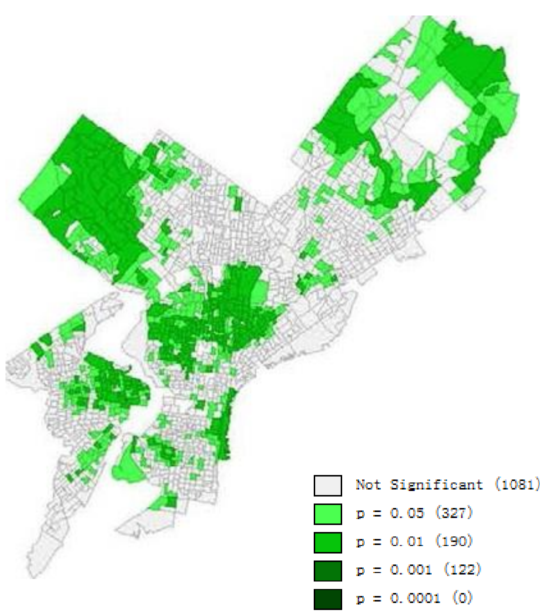
The result of the Moran's I and P-value of the random tests are shown above. We can see that both of the Queen neighbor weight matrices and the 5 nearest neighbor results of Moran's I are quite high, 0.571438 and 0.545142 respectively. The p-values for both of the two methods are 0.001<0.05, which both reject the null hypothesis of no spatial autocorrelation, meaning there is strong spatial autocorrelation of median income in Philadelphia census tract. The two results are quite consistent with each other

Local Moran' s I results:

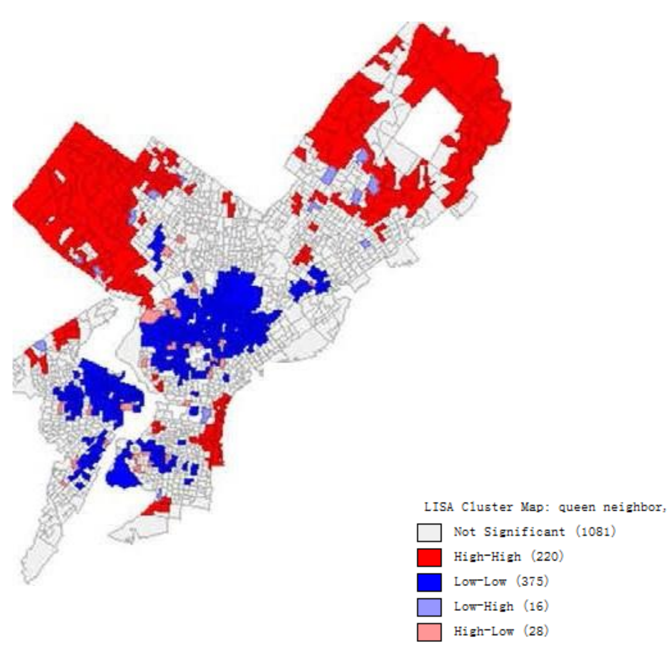
Map 8. LISA Cluster Map (using queen neighbor weight matrices)



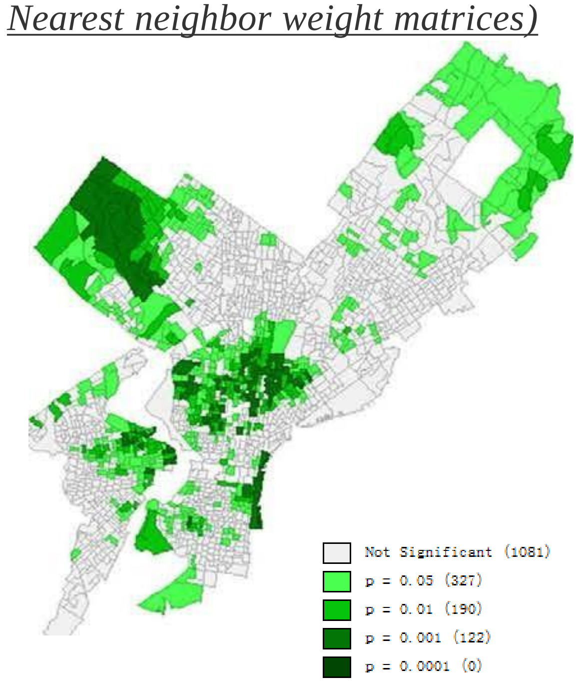
Map 9. LISA Significance Map (queen neighbor weigbt matrices)



Map 10. LISA Cluster Map (5 Nearest neighbor weight matrices)



Map 11. LISA Significance Map (5 Nearest neighbor weight matrices)



5. Conclusion:

For the 5 Nearest Neighbors weight matrices, the LISA results are almost the same with the first one. As the map showed above, the not significant areas are mostly in the Olney and Near North east Philadelphia, meaning there is no strong spatial autocorrelation in terms of Median Income in these census tract. However, there are also two big High-High clusters which are located in the northwest and Far Northeast Philadelphia, Especially for the Chestnut Hill area. The Low-Low clusters are also in the Lower North and Upper North part of Philadelphia, as well as the west Philadelphia. In the south Philadelphia, there are also two Low-Low clusters areas.

What is more, the Low-High and High-Low areas are also not clustered in a certain region, they are quite randomly distributed on the edge areas of the high-high or Low-Low clustering.

As a planner, from the analysis above, the two large portions of poverty clusters are in the Lower North and Upper North pa1t of Philadelphia, as well as west Philadelphia. What is interesting is that poverty areas are quite clustered near the center city areas and rich neighborhood are much further away from the center city (which is on the no1theast and northwest of Philadelphia). Thus, I think it very important to revitalize these areas and to change the current poverty status. Since poverty are usually associated with poor health as well as high crime rates, thus I think we should think of some other methods to upgrade these area, especially they are the areas that are so close to the center city, which should have benefitted a lot from the resources of the center city, and showed a city image of status. Thus, it is very important to upgrade these poverty areas and to in crease social equality.

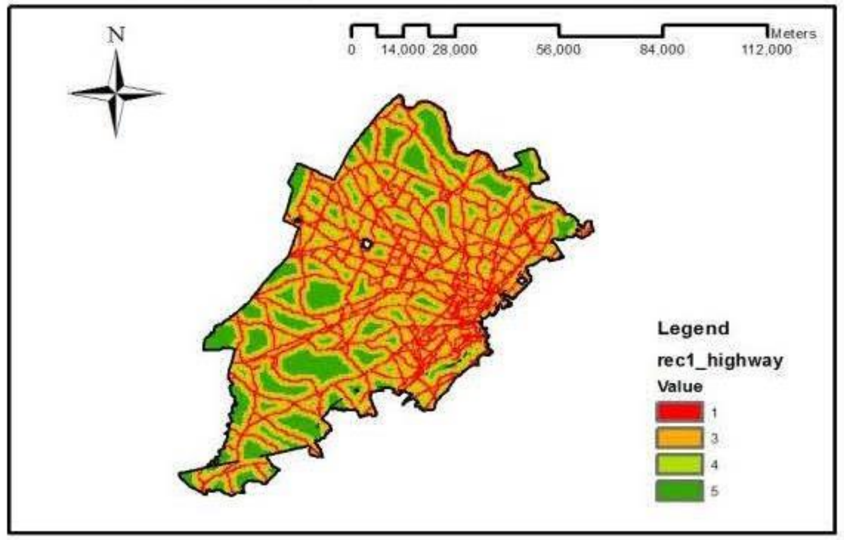
MUSA 507 Applications of Urban Spatial Analysis
Project :Calculation of Retail Site Suitability And Trade Areas Analysis
Instructor: Professor Ken Steif
Student: Jiejun Xie

Image: Retail Site Selectian(from google)

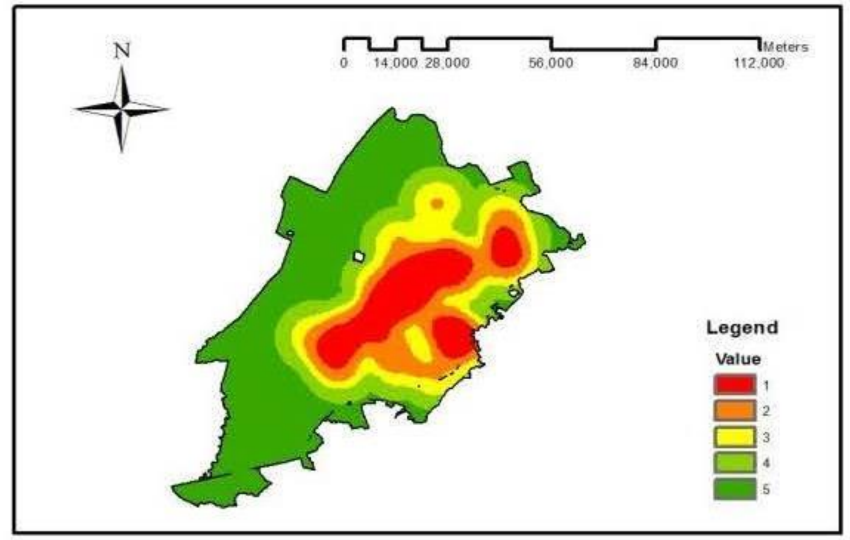
Analysis 1

Supply Side Decision Factors

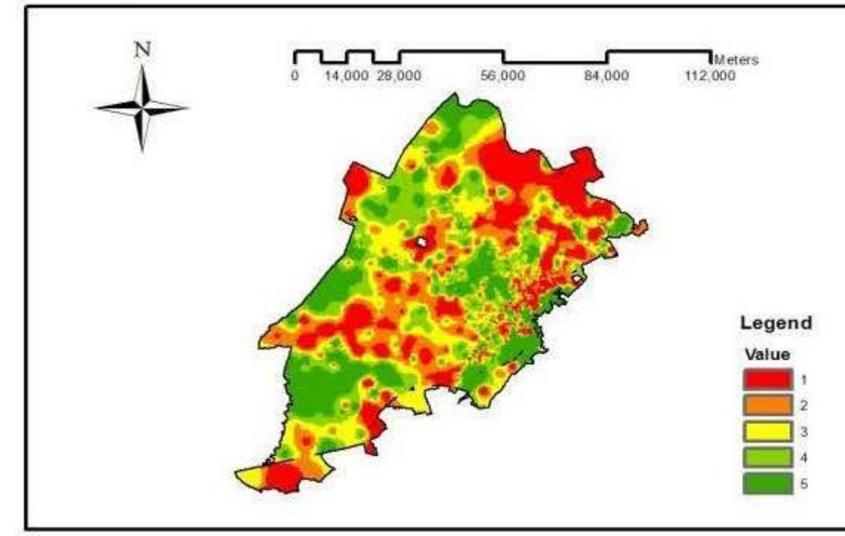
Distance to Highway



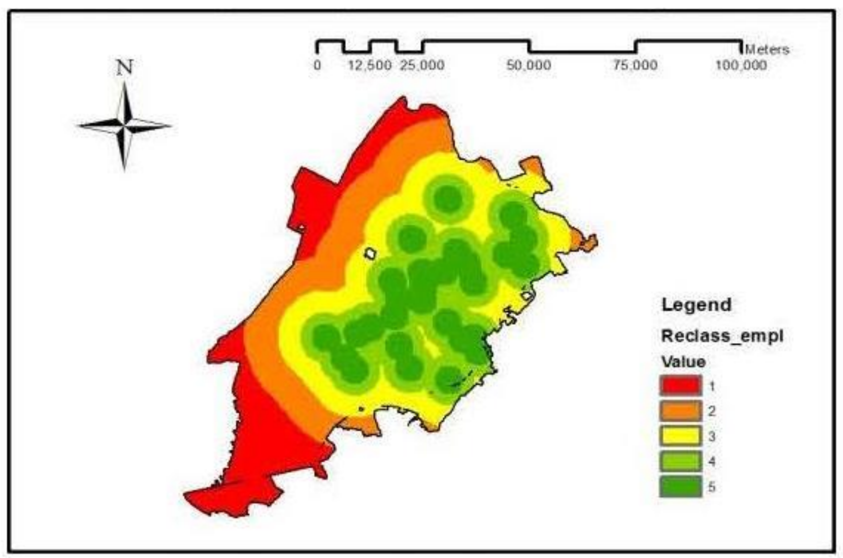
Kemal Density Map of Existing Employment Center
(Total employmenbt as a weight factor)



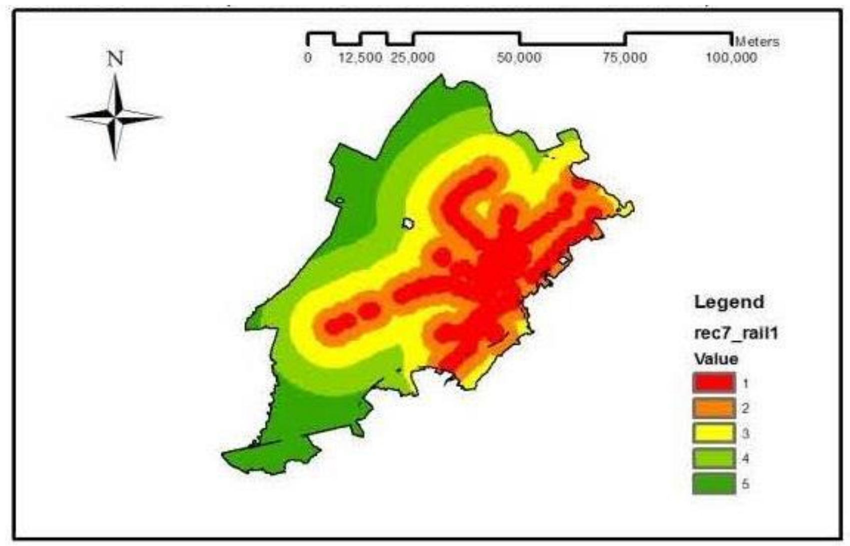
PA employment Density Map(IDW map of PA employment)



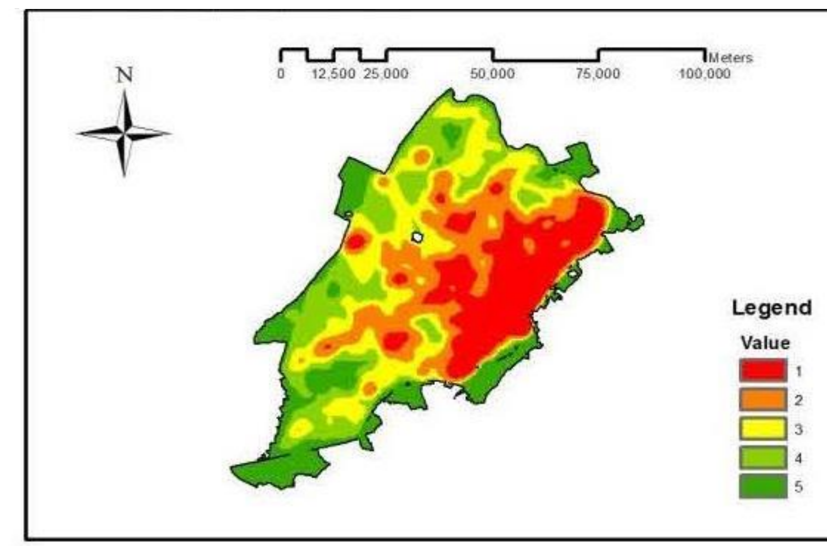
Distance to Competitors(lower score means big distance)



Distance to rail(larger score means bigger distance)

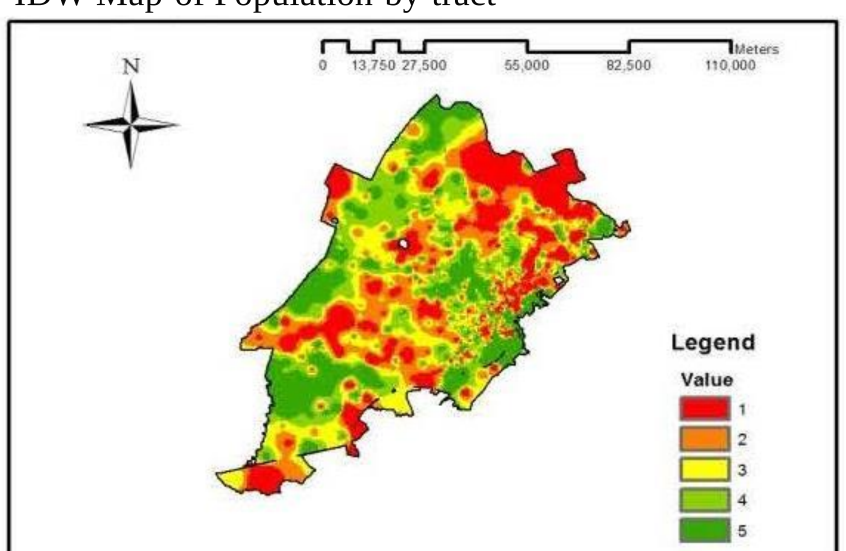


Kernal Density Map of road systems

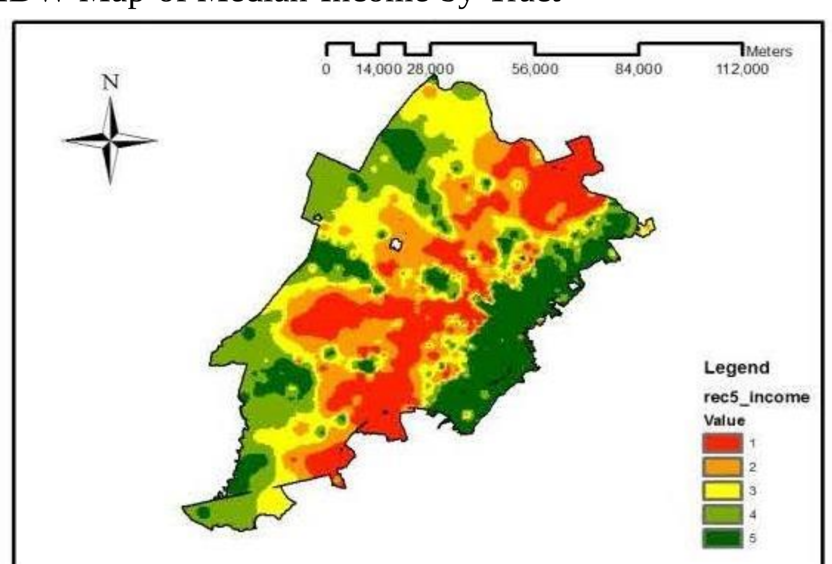


Demand Side Decision Factors

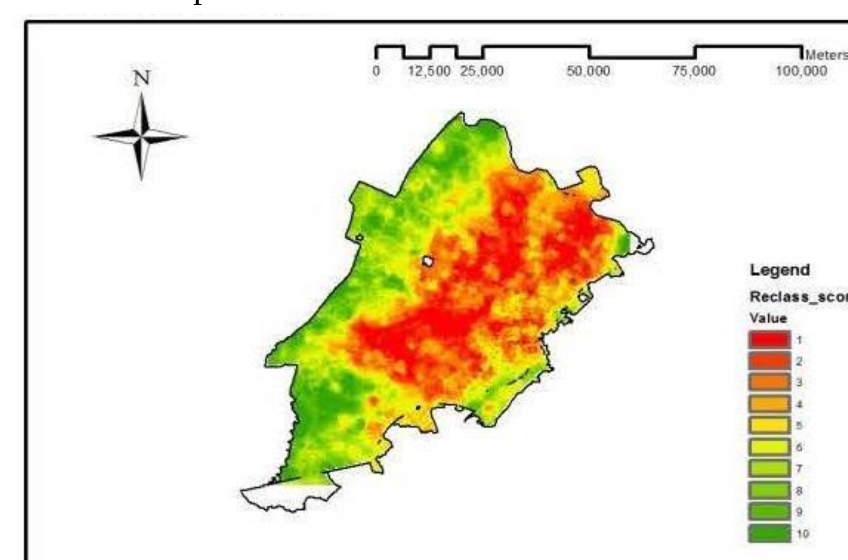
IDW Map of Population by tract



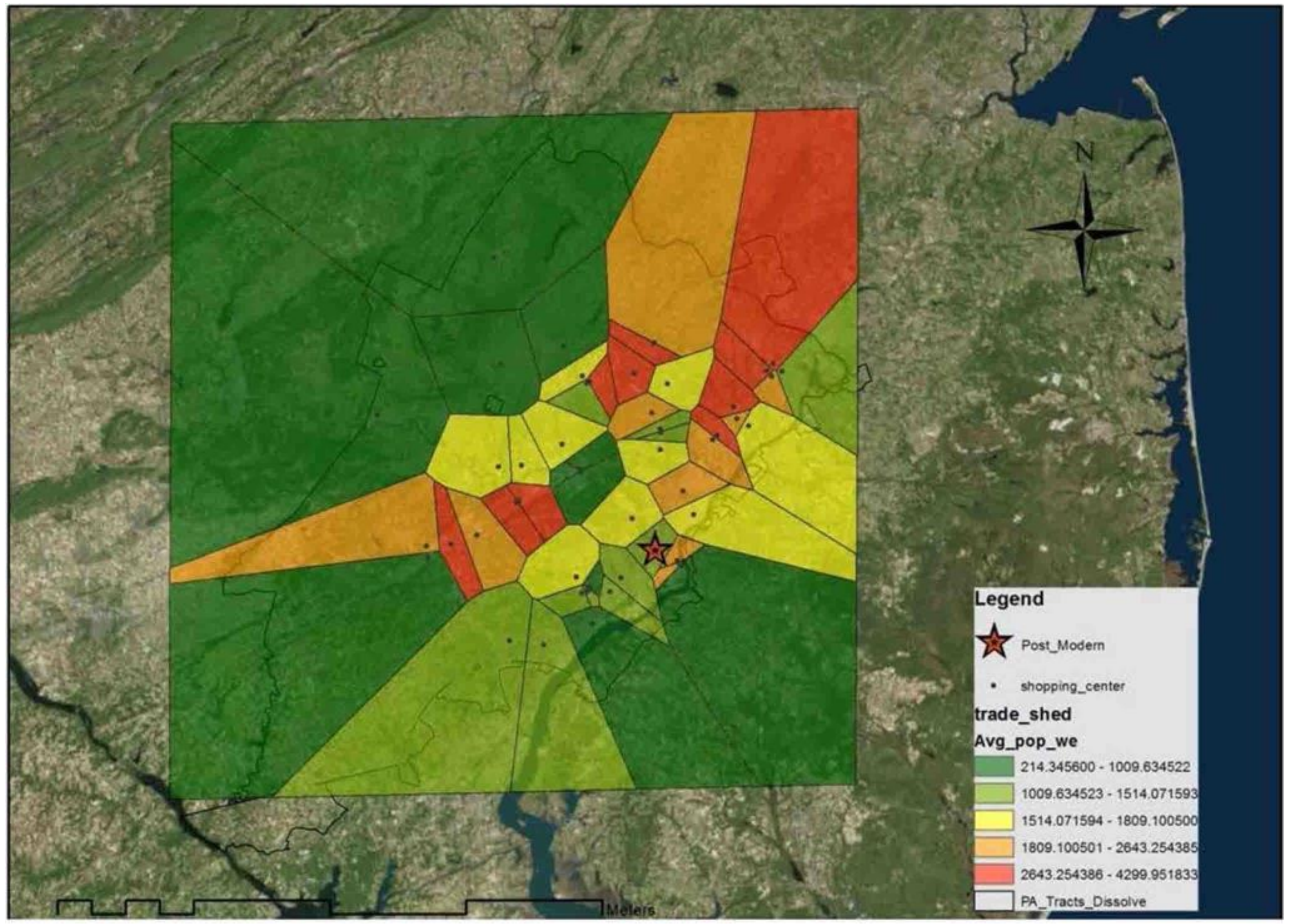
IDW Map of Median Income by Tract



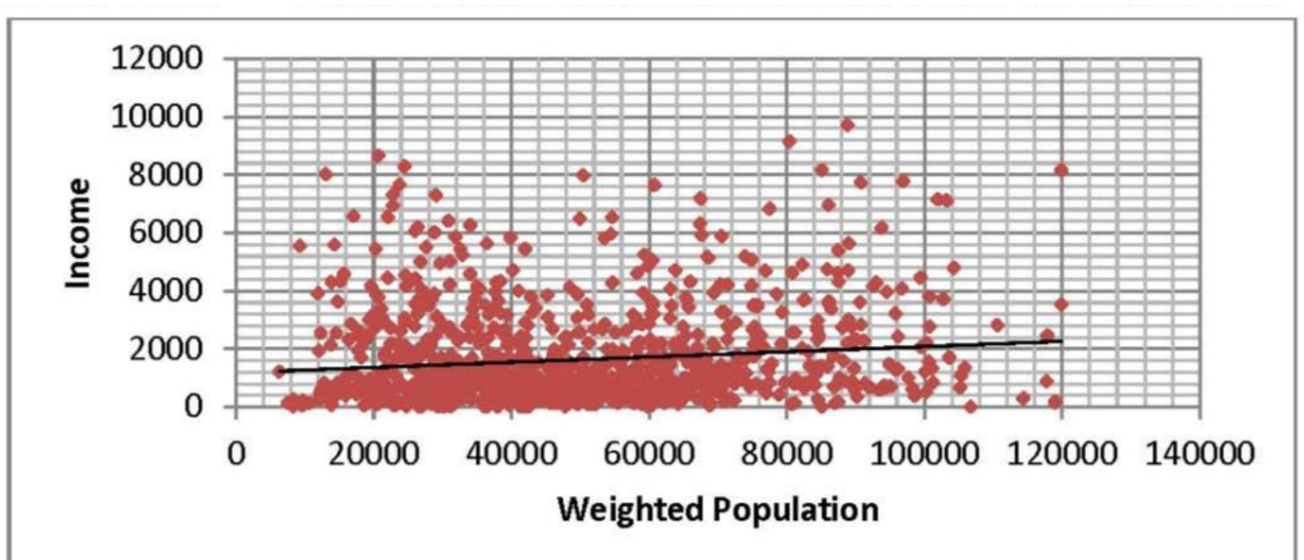
Final development Potential Score



Trade sheds Map for Major Shopping Centers and Post Modem(displayed by weighted Population)



Median Income versus weighted population Map(weighted population get from Huff model)



MUSA 507 Applications of Urban Spatial Analysis
Porject 1: Calculation of Retail Site Suitability And Trade Areas
Analysis Instructor:Professor Ken Steif
Student: Jiejun Xie

Shopping centers' trade shed area data

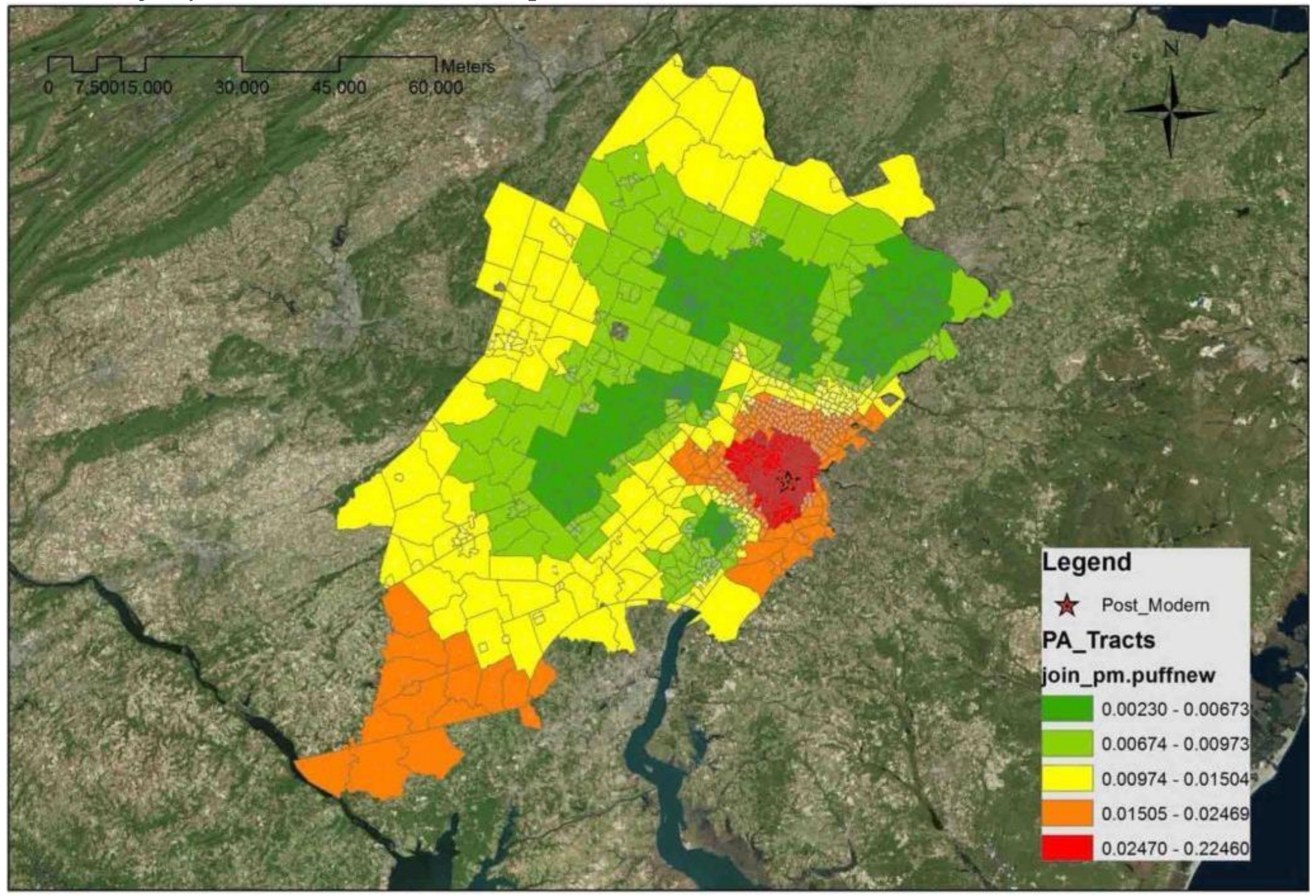
			income	population		Units
CON AME	SALES VOLUME	GLA	MEOIAN INCOME	WEIGHTED POPULATION	POPULATION	HOUSING UNITE
WAL-MART	40500	86432	56420	989	1744n	60432
TARGET	32400	76532	65121	1738	110071	41102
TARGET	25920	79432	25759	214	16906	6306
WAL-MART	35640	84312	44390	1719	67360	25932
OUTMO< STEAKHOUSE	2600	4534	54808	1610	151585	60413
BEST BUY	40375	43134	59118	1643	71305	27129
OUTBAO< STEAKHOUSE	3000	5632	49214	454	93893	35752
OUTBAO< STEAKHOUSE	2000	5323	56839	834	72874	27002
OUTBAO< STEAKHOUSE	3600	4897	68720	2334	60330	23003
W.Al-MART	38718	88532	88244	4208	63124	22761
WAI..-MART	48600	82134	52694	452	68535	25582
W.Al MART	81000	88322	74181	1837	33200	12642
TARGET	2916	80123	66847	3639	35288	14317
OUTBACK STEAKHOUSE	3200	4524	74767	2076	55961	19371
TARGET	48600	74359	54216	896	24468	9508
W.Al.-MART	32400	84509	42441	1090	73187	28926
BEST BUY	40375	44342	54408	1315	12540	4952
W.Al·MART	35640	83900	70047	1288	28373	10990
TARGET	25920	76346	37328	801	89478	33616
WAL-MART	35640	87403	45562	1104	68536	25280
TARGET	25920	78743	80505	1268	107776	36386
BEST BUY	40375	42867	22207	1343	293989	120970
W.Al·MART	113400	79120	33998	2005	78447	36513
TARGET	32400	74031	29849	2643	48952	19808
W.Al MART	43254	88766	32774	1514	198343	74385
OUTBAO< STEAKHOUSE	2800	5432	45516	2200	10965	3929
BEST BUY	40375	45221	63921	3798	61994	22596
TARGET	48600	72314	53321	2828	52509	19542
WAL-MART	35640	84323	42212	2126	132661	53173
WAL -MART	39690	87433	65165	1149	22864	8087
WAIMART	40500	83450	62143	2536	46578	18548
OUTBAO< STEAKHOUSE	3000	5012	61224	1135	20861	8184
TARGET	1296	69998	28327	1697	143263	48745
BEST BUY	40375	40123	32402	1934	388961	140688
TARGET	40500	75689	47776	1724	283681	115834
OUTBAO< STEAKHOUSE	3200	4532	70696	469	5023	1921
OUTBAO< STEAKHOUSE	2800	4121	42990	774	12004	5164
WAl·MART	1620	89321	75468	1010	59932	22002
TARGET	32400	81343	64265	1809	49671	18744
OUTBAO< STEAKHOUSE	3000	4932	83250	3128	42213	17137
BEST BUY	40375	46435	99812	4277	27521	10756
BEST BUY	6460	47523	63547	1591	89118	32829
BEST SUV	40375	42869	60447	1648	49461	19232
TARGET	36450	67430	n393	1356	26935	10339
TARGET	25920	72987	79041	3291	13568	4891
OUTBAO< STEAKHOUSE	4800	5067	63462	915	71442	25797
W.Al·MART	35640	82134	62983	1697	65680	23601
TARGET	32400	75349	86044	4259	18178	5733
WAL-MART	486	84532	66358	4300	33600	11548
WAL-MART	51840	89539	50263	1036	65968	24389
BEST BUY	40375	43890	49856	1942	48042	17508
Post_Modern	0	0	76120	2850	44692	15540

Using Huff model to calculate the probability people travel in each store, and then we get the number of People choose to travel in each store(weighted population).

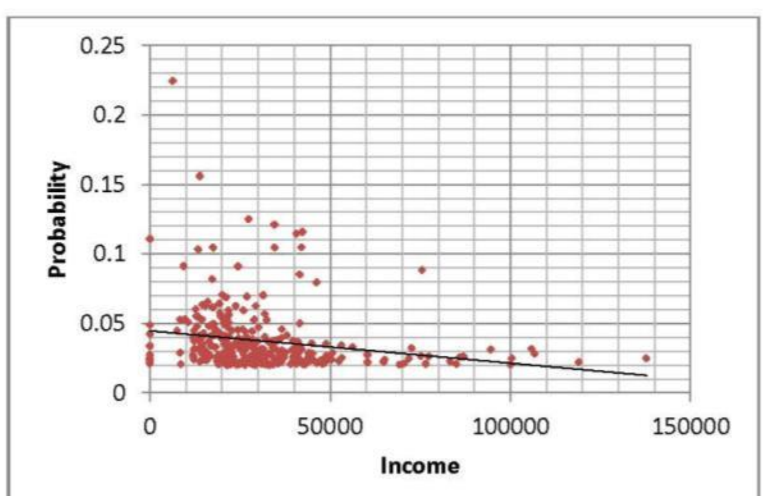
The result of the weighted population by tract was quite in consistent with the original population distribution. We can see the potential of people shopping in each trade area is not just decided by the proximity. Trade areas which have larger size might have higher score. have more competitors but is more accessible might have a lower score.

However, we do not consider the busines quality effect. If the store has better quality, the weighted result might be higher. However, we can see from the scatterplot that there is no strong relationship between median income and the weighted population. One reason might be, we do not set a high score for income preferences, the other might be, people do not have a strong preference for Post Modern nomatter how much they earned.

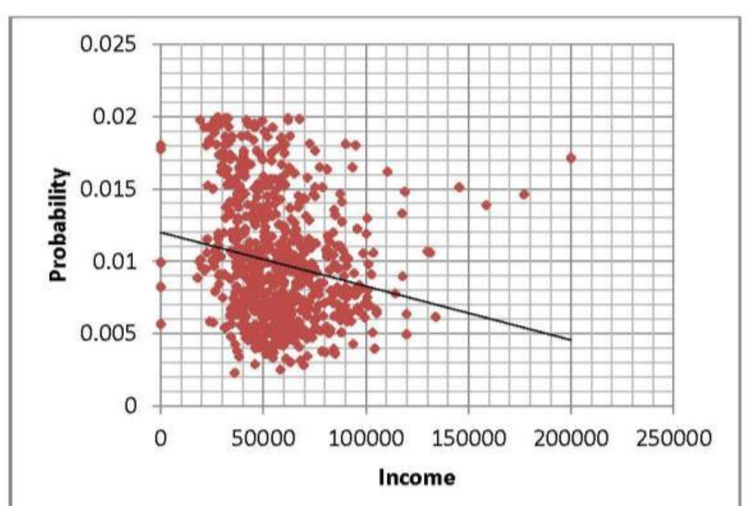
Probability of Consumer choose to shop at Post@Modern



Probability of Consumer choose to shop at Post@M odern



Probability of Consumer choose to shop at Post@Modern (p<0.025)



MUSA 507 Applications of Urban Spatial Analysis
Porject 1: Calculation of Retail Site Suitability And Trade Areas
Analysis Instructor:Professor Ken Steif
Student: Jiejun Xi e

Probability of Consumer choose to shop at Post@Modern

	Shopping times/year	Shop times @post modern/year	Shop times @ post modern/year	post modern
shopping frequency/year	156	2	3	3.12
probability		0.012821	0.019231	0.02

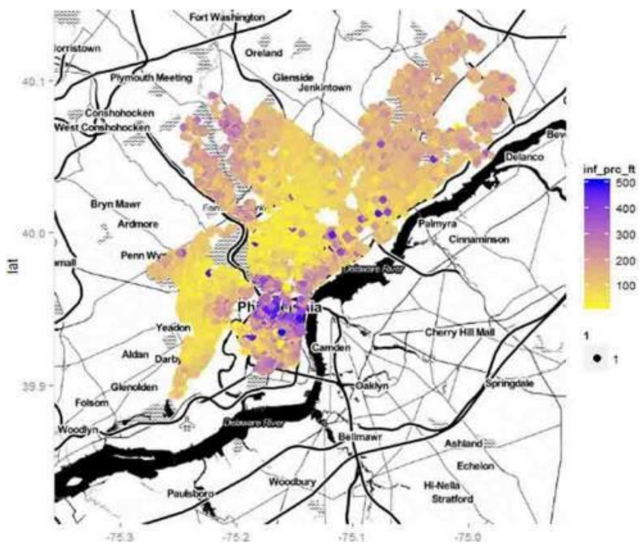
I choose 0.02 as the gap probability, since we assume people shop 3 times a week and each would shop 156 times in a year. If we go to post modem just 3 times a ear, that means we have a 0.019 probability to shop there.The same with the situation when probability is below 0.02, we can see the point is quite randomly distributed and there is no strong linear relationship between probability and median income.

However, we do not consider the consumer preference for different types of people. If Post Modem has a certain type of consumer to attract who lives in certain parts of the city, the analysis will not be so reliable. However, we can see the result of Huff model is quite in consistent with the distance score result. What is more, since the first analy sis has considered the income factor, there might be no strong relationship between median income and market share in comparison to other factors like distance.

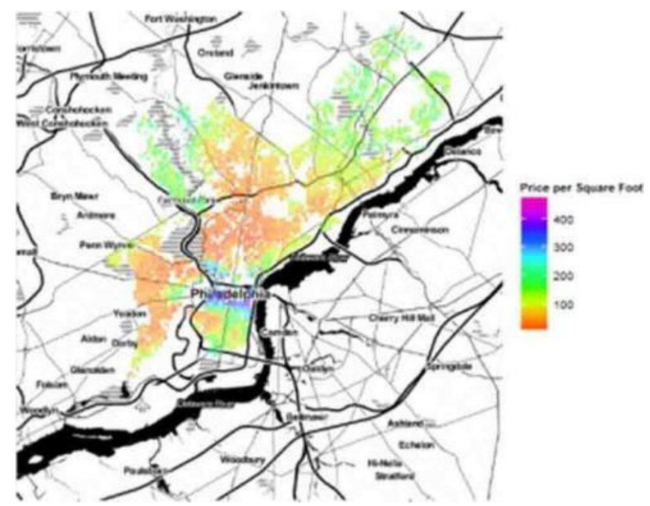
Overall, I strongly recomend the developer go ahead with Post@Modem, since I think it not only has a reaonable price but could also serve most of the consumer needs. What is more, it could supplement a gap in the market that there is no Post@Modem in Philadelphia.

Test On The Data

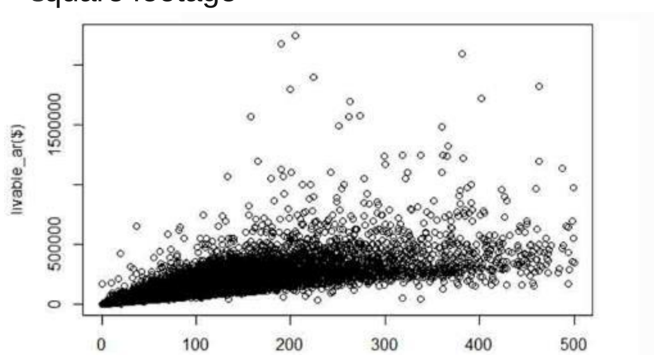
Map 2: Sale price per square foot in Philadelphia



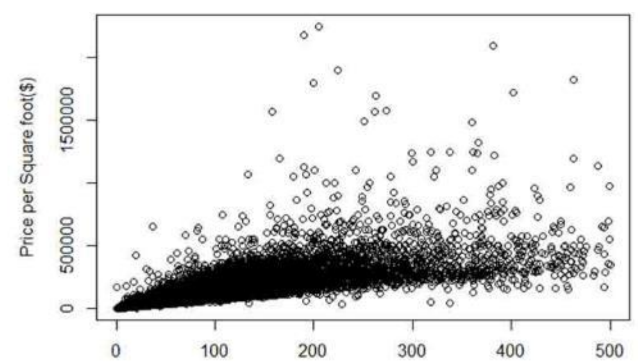
Map 3: interpolation map of mean price per square foot in Philadelphia



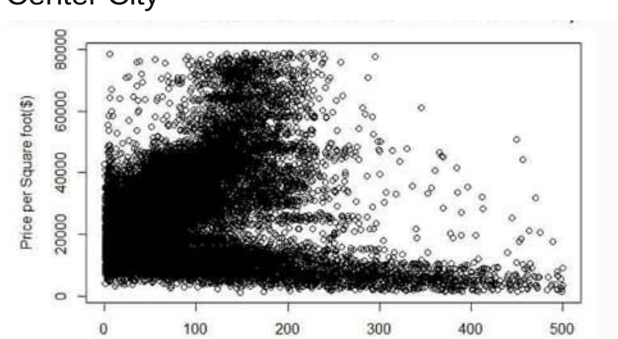
PlotI :Price as a function of internal home square footage



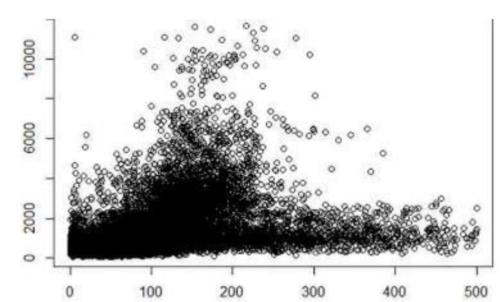
Plot2: House Price as a function of Sale price



Plot 3: Price as a function of distance to Center City



Plot 4: Price as a function of distance to crime



As we can see from the plot map above, Plot 1 and Plot 2 both have a trend to cluster on a specific line. That is we could use "Sale Price" and "internal home square footage" to esti mate "price". However, for the "distance to center city" and "distance to crime variable", the two plots seem to have a trend line but at some point, they are quite random.

MUSA 507 Applications of Urban Spatial Analysis
Porject : Hedonic regression and willingness to pay for transit access in Philadelphia
Instructor:Professor Ken Steif
Student: Jiejun Xie

Kitchen Sink Model to identify the willingness to pay for transit in Philadelphia

```
Call:
lm(formula = log(hed$inf_prc_ft)   d_septa + d_crime + pct_non_wh,
   data = hed)

Residuals:
    Min 1Q  Median      3Q     Max
-6.6061 -0.3247  0.1235  0.5405  2.7029

Coefficients:
             Estimate Std.Error  t value  Pr(>|t|)
(Intercept)  4.535e+00  1.623e-02  279.441  < 2e-16 ***
d_septa      4.290e-06  .227e-06   3.497    0.000471 ***
d_crime      1.502e-04  8.571e-06  17.526   < 2e-16 ***
pct_non_wh  -1.220e+00  .905e-02  -64.001   < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9323 on 23254 degrees of freedom
Multiple R-squared: 0.2841, Adjusted R-squared: 0.284
F-statistic: 3076 on 3 and 23254 DF, p-value: < 2.2e-16
```

(1). How do I choose my model

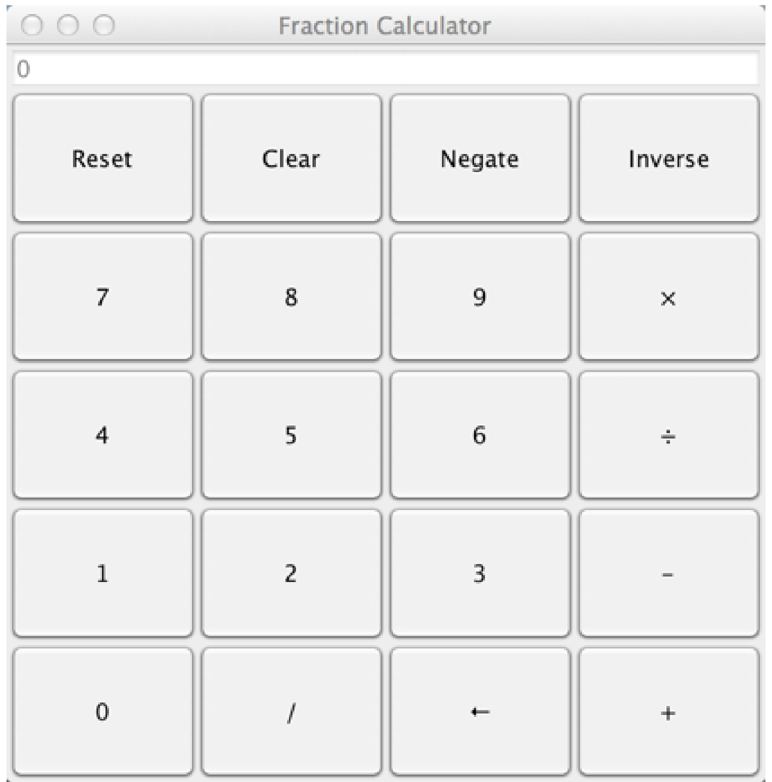
First, the independent variable is not random, so I use " log" function for "inf_prc_ft" to make regression. Then, I just use "distance to septa" to make regression. The result is quite well, P value is very high and r square is higher than 0.5, the correlation test is higher than 0.3. However, the residual (Plot five) is not randomly distributed along the fitted value, there is a trend line for the result.

What is more, when I add sale price as an estimator , the residual becomes less random. I thought the sale price might be not proper to estimate this result because it might be too strong and it might be not a decision factor for price. Then I dropped it and added "distance to crime" and "percent non-white" to estimate. The result of r square is only 0.28, but the re sidual is quite randomly distributed now.

(2). Interpret your distance to transit variable

Distance to transit variable is "distance to septa". It' s P-value is 0.000471,less than 0.05, so it is statistically significant. It' s coefficient value is (4.290e-0.6), that means, when dis tance to septa increase 1, the house price per foot will increase 0.0004290009.

Other works: Java/Python project



Java project to create a calculator to do Fraction calculations.
You can do normal arithmetic, using +, -, x, /, you can also invert and negate the value, by using the Inverse and Negate button. The calculator can catch any exceptions that do not follow the required routine.

```
Enter a startpoint for shoot a ray:1T

         1   2   3   4   5   6   7   8

      -  H   -   -   -   -   -   -   -   -
1     |  -   -   -   -   -   -   -   -   |     1
2     |  -   -   -   -   -   -   -   -   |     2
3     |  -   -   -   -   -   -   -   -   |     3
4     |  -   -   -   -   -   -   -   -   |     4
5     |  -   -   -   -   -   -   -   -   |     5
6     |  -   -   -   -   -   -   -   -   |     6
7     |  -   -   -   -   -   -   -   -   |     7
8     |  -   -   -   -   -   -   -   -   |     8
      -  -   -   -   -   -   -   -   -   -

         1   2   3   4   5   6   7   8

None
S(shoot) or M(mark) or E(end):
```

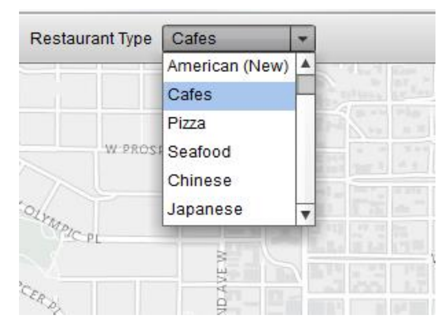
Black Box game(Python project):
The basic idea is: The computer "hides" four atoms somewhere on an 8x8 grid. The player, try to figure out where the balls are by shooting in "rays" from any of the 32 positions along the sides of the grid (8 on each side, including the top and the bottom), and seeing where the ray comes out, or if it comes out at all.

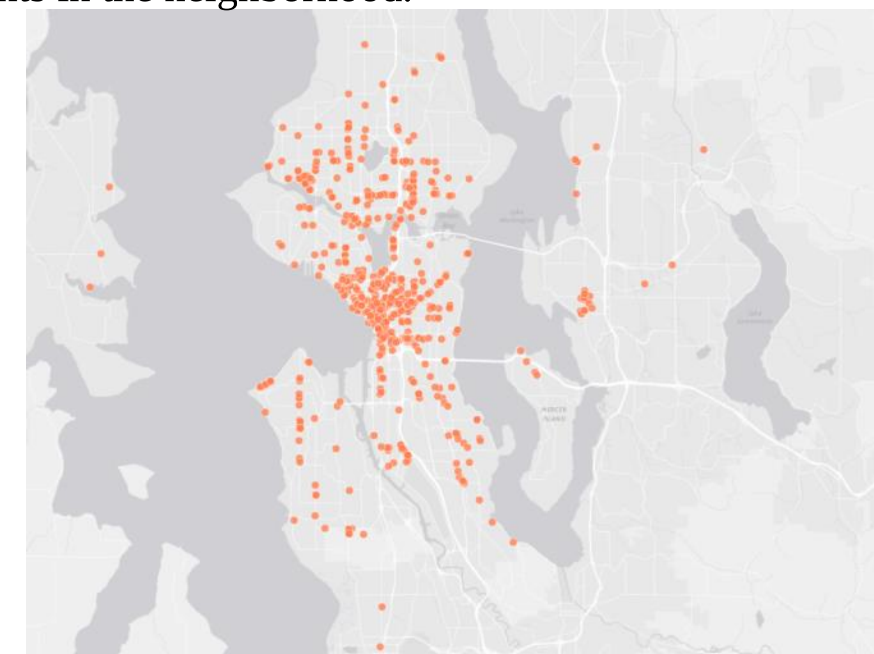
Other works: Flex Project(Seattle Dining Guide)

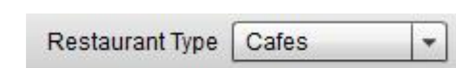
Use the Yelp API(and python) to get hotel data from yelp, use the data
To create a small flex application for Seattle areas, by which people can find where to eat in their neighborhood. Researchers or business person also can count the total number of certain type of restaurants in the neighborhood.

Main Function Of the Application

Search by distance
Zoom in to Search Areas
Search by category
Render by rating score
Get Business Information







Alert Total number of restaurants