Assignment 4 - Spatial Data Science Final

Background

The NYC Education (2000) dataset was selected (https://geodacenter.github.io/data-and-lab/NYC-Census-2000/). This dataset has 2,216 neighborhoods, with 56 variables indicating location and education information.

The dependent variables selected is the percentage of the population aged over 25 that has dropped out of high school (HS_DROP). This was selected as a more representative version of high school drop out than simply the dropout rate between the ages of 16-19, as someone who drops out of high school can always return to high school at a later year. The independent variables are: (1) percentage of students in private school (PER_PRV_SC), and (2) percentage of population that is non-white (PER_MNRTY). These variables were all spatially intensive.

The first research question will be, does the percentage of students in public or private school have an impact on the dropout rate? It should be expected that public schools might have a higher dropout rate, as the parents of private school students may have more resources to ensure that their children will succeed. The second RQ is, does the percentage of minorities predict the dropout rate? It would be expected to be so, but only because it could be possible that higher percentages of minorities are found in lower income communities, possibly leading to less resources, or to them being in public schools.

NYC has 5 boroughs, 59 community districts, and 2216 neighborhoods. The boroughs each have a unique administration. The boroughs are split into districts, which are further split into neighorhoods. The analysis was chosen to be conducted on a neighborhood level, as the neighborhoods each have different community boards and different demographics. Doing the analysis on a smaller scale made more sense as these areas have their own individual administrations. Also, it is known that aggregating groups with different demographics, or strata, within them can lead to unpredictable results. Therefore, the scale of neighborhoods selected is likely correct and MAUP is not a problem in this analysis.

EDA

Firstly we will be doing some EDA on this dataset. A choropleth map of the dropout rates shows that the highest dropout rates are in the Bronx and Brooklyn, while the lowest are in Staten Island and Manhattan. (Fig 1)

The highest percentage of private schools are seen in Manhattan, where the whole of Manhattan is seen with the highest rate of 42 - 100%. Staten Island also has a lot of private schools. There are also some dense scatterings in Brooklyn and Queens. (Fig 2)

The highest percentage of public schools are seen in the lower part of the Bronx, and in between Brooklyn and Queens. (Fig 3)

The highest percentage of minority groups are seen in Brooklyn, Queens, and the Bronx. (Fig 4) A Chow test (Fig 14) on a scatterplot showing the effect of percentage minority on high

school dropout shows that selecting Manhattan has a p-value of 0. This means that the geographic area of Manhattan has an impact on the interaction of minority and dropout rates.

Interestingly, in some of the neighborhoods with the highest percentage of private schools (60-100%), while most of the dropout rates are low (0-10%) there is still a high variance in the drop out rates, showing dropout rates from 0-70%. (Fig 5) Therefore, private schools may not necessarily predict dropout. Looking into this further with a conditional map, we do see an overall trend that private school and dropout are related (Fig 8). Lower percentage of private schools have higher dropout, and high percentage of private schools have lower dropout. However, there do seem to be exceptions to this rule. Some private schools with high dropout can be seen in Brooklyn and the Bronx, and likewise some public schools with low dropout can be seen in Staten Island, Queens, and Brooklyn.

Looking at the minority and dropout relationship by linking between the histograms of minority and dropout, we can also see a clear effect between them: the lower percentage minority groups have lower dropout rates. (Fig 6) We can also use a conditional map to show that as the percentage of minorities increases, the high school dropout increases. (Fig 7) The overall trend is that the map on the left with low percentages of minorities shows low dropout rates, while the image on the right with high percentage of minorities shows higher dropout rates. However, there are still some exceptions to this: there are some higher dropout rates in the map on the left, and low dropout rates in the map on the right, showing this trend does not necessarily hold true for some neighborhoods.

The relationship between percentage of private schools and percentage of minorities on the high school dropout rate was also explored with scatterplots, which are shown in Fig 9 and Fig 10 respectively. There is a negative relationship between private school and dropout rate: as the percentage of private schools increases, the dropout rate decreases (seen in Fig 9). Interestingly, the line appears to be exponential: it drops very quickly at the start and becomes steady at the end. This means that having just a small number of private schools very quickly changes the dropout rate. This could be due to a number of factors. The exponential nature of the relationship between private school and dropout rate may be explained by considering that when private school is an option, students might choose it over public school, thereby accounting for the larger fall in dropout over the incremental increase in percentage of private schools. It could also be explained by the fact that the quality of the education could change as a result of having private schools, or because the increase in private schools highlights a different demographic or geographic area. Exploring the last point further, it is clear that there definitely is a geographic change: highlighting a narrow section on the scatterplot from the least number of private schools to the most shows a change in geographic locations from Queens and the Bronx, into Manhattan. This is demonstrated with a Chow test on the scatterplot (Fig 11), where the highest percentage of private schools are highlighted. This lights up the Manhattan area, as well as some areas in Brooklyn and Staten Island, and gives a p-value of 0, showing that the dropout rate between the area highlighted and not highlighted is significant. Therefore, the percentage of private schools is significant in accounting for the dropout rate.

Moving onto the relationship between the percentage of minorities and the dropout rate, there is conversely a positive relationship between them: when the percentage of minorities increases, the dropout rate increases (as seen in Fig 10). Checking with a Chow test, selecting the highest percentage of minorities has a very significant impact on the high school dropout rate (Fig 13). This means that whether there is a high or low percentage of minorities is significant in predicting the dropout rate.

Interestingly, the highest percentage of private schools previously selected highlights just the lower end of the percentage of minorities (Fig 12). While there are exceptions, this might signify that private schools are more likely to be found in areas with less minorities. The Chow test is significant, showing that the percentage of private schools has a significant effect on the percentage of minorities.

Overall, our EDA shows that the percentage of private schools and minorities is statistically significant on the dropout rate. There is also a statistically significant connection between the percentage of private schools and minorities, where private school predicts minorities.

Considering these initial Exploratory Data Analyses, we will ask further interesting questions to explore the data.

Firstly, we will look at the non-linear trend between private school and dropout rate. Since it was non-linear, we can use a LOWESS smoother to see if it will fit the data better. Using a bandwidth of 0.3 seems to capture the information in the data the best. This brings out the structural breaks in the data in a striking way (Fig 15). While there is an initial upwards tilt to the slope, this is very likely due to the lack of data near the boundaries, where it is known that unsupervised methods don't perform well near boundaries. Therefore, we can likely ignore this upwards tilting section. Other than this, this graph confirms the (negative) exponential nature of the relationship between private schools and high school dropout rate that was previously speculated.

Next, since there were some interesting relationships between the variables suggested by the EDA, a multivariate EDA of PCP will be done in order to further explore these effects. Highlighting the highest rates of high school dropout selected low rates of private schools and high rates of minorities. (Fig 16) The areas of high dropout that were very informative of private schools and minorities were located on the map as belonging mostly to the Bronx and Brooklyn. Conversely, selecting low rates of high school dropout did not give any interesting information, however - this highlighted the whole range of private schools and minorities. (Fig 17) Interestingly enough, it seems like a high dropout rate is more informative of selecting subsets of data than a low dropout rate. No other interesting patterns were found with the PCP analysis.

Spatial Autocorrelation

Firstly we will be making some weights. Distance based weights and Knn weights with six neighbors will be used. These weights were chosen as they are two important weight forms. Queen weights were attempted but this lead to an isolate, so these weights were not preferred.

We also have a high confidence that the weights picked were a good choice, as they both gave very similar results, which gives us a higher degree of confidence in our graphs.

Before using the weights we will explore them. For the distance based weights, the minimum neighbor is 1 and the maximum neighbor is 191. For the Knn weights, both the minimum and maximum neighbors were six. Looking into the distance based weights with a histogram and map, it looks like the locations with the highest number of neighbors are found in the Bronx and Brooklyn. (Fig 18) The ones with the lowest number of neighbors are at Staten Island, and on the very boundaries of the mainland. (Fig 19) Knowing this about the distance weights is important, as this might skew our results.

Next, to assess global spatial autocorrelation, we will be using Moran's I. We will check it for the dependent variable (dropout) and both independent variables (minority and private schools), using both weights. The randomization with the permutations will be set as 999. Fig 20 shows the results for the variable of minority, with both weights. For both weights, the pseudo p-value is 0.001, showing that there is global spatial autocorrelation. The z-scores are 149.73 (distance weights) and 70.96 (knn weights). Fig 21 shows the results for the variable of private schools, for both weights. For both weights, the pseudo p-value is 0.001 again, showing that there is also global spatial autocorrelation. The z-scores are 77.93 (distance weights) and 50.06 (knn weights). Fig 22 shows the results for the variable of dropout, for both weights. For both weights, the pseudo p-value is 0.001 again, showing that there is also global spatial autocorrelation. The z-scores are 96.94 (distance weights) and 52.55 (knn weights). These findings of global spatial autocorrelation are not surprising, given the results from the EDA found in the previous section. Since the distance weights repeatedly have higher z scores, we can tell that they might be more sensitive to picking up meaning than the knn weights.

Moving onto local autocorrelation, univariate local Moran's I will be checked for all three variables, with both weights, with permutations.

Fig 23 shows univariate local Moran's I for the percentage of minority groups for both weights. The p value is less than 0.05 for both weights and is significant. The high-high locations are in the North of Manhattan, the whole of the Bronx, in between Brooklyn and Queens, and the South of Queens. The low-low locations are in lower Manhattan (starting from around South of the Metropolitan), Staten Island, the South of Brooklyn, the very East of Queens, and the very West of Queens. Fig 24 shows univariate local Moran's I for the percentage of private schools for both weights. The p value is less than 0.05 for both weights and is significant. The high-high locations are found in Manhattan, Brooklyn, and Staten Island. The low-low locations are found in the Bronx and Queens. Fig 25 shows univariate local Moran's I for the percentage of dropout for both weights. The p value is less than 0.05 for both weights and is significant. The high-high locations are found in the Bronx and in between Brooklyn and Queens. The low-low locations are in Staten Island, Manhattan, the North-West of Brooklyn (close to Manhattan) and South-East of Brooklyn, as well as the very North-East of Queens.

For all three figures, the maps are similar to each other, so the weights seem to be picking up the same information, however the distance weights have the clusters with an even larger

radius, possibly showing that the distance weights have heigher sensitivity. Additionally, for all of the above, the high-lows are found near the low-lows, and the low-highs are found near the high-highs.

Next, the same will be done for the univariate local Geary test: all three variables, both weights, with permutations. To be noted with local Geary, the low-high and high-lows are classes in the same way, so it is impossible to derive any difference between them.

Fig 26 shows univariate local Geary test for the percentage of minority groups for both weights. The p value is less than 0.05 for both weights and is significant. The high-highs are located in the Bronx, in between Brooklyn and Queens, and in the South of Queens. The low-lows are found in Staten Island, south of Brooklyn, Manhattan, and North of Queens. Fig 27 shows univariate local Geary test for the percentage of private schools for both weights. The p value is less than 0.05 for both weights and is significant. The high-highs are found in Manhattan, Staten Island, central Brooklyn, as well as the Northern edges of the Bronx, and a circular region outside the low-low area in Queens. There is a large low-low area starting in the Bronx and northern Manhattan, going down into Northern Queens, and then into the South of Queens and lightly into Brooklyn. The low-low area starts at the North of the Bronx and follows the outside of the low-lows down South. Fig 28 shows univariate local Geary test for the dropout rate for both weights. The p value is less than 0.05 for both weights and is significant. The high-highs are in the South-West of Brooklyn, in between Brooklyn and Queens, the North of Queens, and the North of Manhattan and the Bronx. The low-lows are found in Staten Island, Manhattan, the South of Brooklyn, and the West of Queens.

Just like in the univariate local Moran's I, both weights give similar results, with the distance-weights being a bit more sensitive.

Importantly, while we may have found some clusters, we can know that they are there, but we cannot explain why they occur. We can only state some interesting patterns.

Firstly, we have seen that the global and local spatial autocorrelation statistics for all three variables are significant. This does signify that there is some clustering occurring.

The local Geary and local Moran's I results agree with each other overall, for all three variables. However, it is possible that the local Geary picked up more information. Fig 29 shows the contrast between the local Geary and local Moran's I for the distance weights for two variables, where the local Geary map has more regions colored in.

In terms of the results, we found that the highest minority groups are found in the Bronx, in between Brooklyn and Queens, and in the South of Queens. The lowest minority groups are found in Staten Island, south of Brooklyn, Manhattan, and North of Queens. The highest percentage of private schools are found in Manhattan, Staten Island, central Brooklyn, as well as the Northern edges of the Bronx. The lowest percentage of private schools are found in the Bronx and Queens. The highest percentage of dropout is found in South-West of Brooklyn, in between Brooklyn and Queens, the North of Queens, and the North of Manhattan and the Bronx. The lowest percentage of dropout is found in Staten Island, Manhattan, the South of Brooklyn, and the West of Queens.

Interestingly, the clusters with the highest percentage of private schools (Staten Island, Manhattan, Central Brooklyn) correspond to the locations with the lowest percentage of dropout (Staten Island, Manhattan, the South of Brooklyn), as well as with the lowest percentage of minority groups (Staten Island, Manhattan, South of Brooklyn). This confirms the previous findings that these variables are statistically significant - and geographically significant.

The locations with highest percentage of minority groups (the Bronx, in between Brooklyn and Queens, and in the South of Queens), corresponds to the locations with the lowest percentage of private schools (in the Bronx and Queens). It also corresponds to the locations with highest dropout (South-West of Brooklyn, in between Brooklyn and Queens, the North of Queens, and the North of Manhattan and the Bronx).

While some boroughs consistently show only one cluster, other boroughs seems to have various demographics and variances of the variables. Brooklyn in particular has a very high variance, with very different communities populating its different sections, in a way that is difficult to summarize and should request further attention. Likewise, the North and South of the Bronx behave differently. Also visible is that the North and South of Manhattan behave differently. The South West of Queens and its other areas also behave differently to each other. On the other hand, Staten Island is on the other end, where it has low variance, likely meaning that there is not that much variability in the demographics populating this borough.

Moving onto a bivariate analysis, the bivariate local Moran's I will be done between minority and dropout, and private school and dropout, for both weights.

Firstly, Fig 30 shows the bivariate analysis for minority and dropout for both weights. The high-highs are found in the Bronx, in between Brooklyn and Queens, and the West of Brooklyn. The low-lows are found in Staten Island, Manhattan, North of the Bronx, SouthWest and South of Brooklyn, and North East of Queens. There is a high proportion of high-highs and low-lows compared to high-lows and low-highs. This has been true of most of the previous graphs made.

However, the next figure shows a divergence in this finding. Fig 31 shows the bivariate analysis for private school and dropout for both weights. The proportion of high-highs and low-lows is very low comparative to the high-lows and low-highs. This means that private school is not as good at predicting dropout, comparative to minority. This is a very interesting finding. Overall though, the low-highs (and a very small amount of high-highs) are found in the Southern Bronx, in between Brooklyn and Queen. The high-lows (and a very small amount of low-lows) are found in Staten Island, Manhattan, South of Brooklyn, and Western Queens.

This bivariate analysis shows that minority can predict dropout much better than private school. This is not consistent with the previous findings, which had found that both variables were statistically effective at predicting dropout. The geographic areas located by the bivariate analysis are still the same though, showing that these geographic areas have their own distinct patterns and behaviors.

Conclusion

Overall, there are clear and distinct locations of spatial heterogeneity for all three variables in this dataset, both between and within the boroughs, which have been described above. There is evidence that both independent variables of minority and private school are statistically significant in predicting dropout rates. From the bivariate analysis, it is possible that minority might be better at predicting dropout than private school.

These results have to be taken with a grain of salt, as it is impossible to say that a higher percentage of minority students is the reason for a higher dropout rate (causation), or that minority students cannot perform as well. This simply shows that higher minority rates correspond with higher dropout rates. There may be many reasons for this, which we cannot know or explain. Some possible reasons, though, could be that minority students might typically come from a lower-income background, meaning that they have less resources to succeed. Also, we have seen that the distributions of minority students have clusters across NYC. The geographic areas that they are located in may contribute to their education, if the districts are typically poorer, have more crime, or less public resources. The areas that they are located in - the Bronx, Brooklyn, and Queens - are traditionally known for their higher crime rates.

Likewise, it is impossible to say that a higher percentage of private schools cause a lower dropout rate. It could be the case that the areas of with a higher percentage of private schools are just more well-off and therefore have more resources, both in the community and in each individual family.

The non-linear relationship between private school and dropout was also interesting: it is possible that when private school is an option, more students will select it, which in turn lowers the dropout in this exponential fashion.

The PCP result was also important. While high dropout rates are usually found in areas with high percentages of minority students and low percentages of private schools, there can be a low dropout rate in a very wide variety of rates of minorities and private schools. Low dropout rates are not only found in private schools with low minority rates - they can be achieved anywhere.

Overall, this was an interesting dataset to go through and a very useful class! I hope to use the tools I learned in this class in the future on other datasets.

Fig 1:

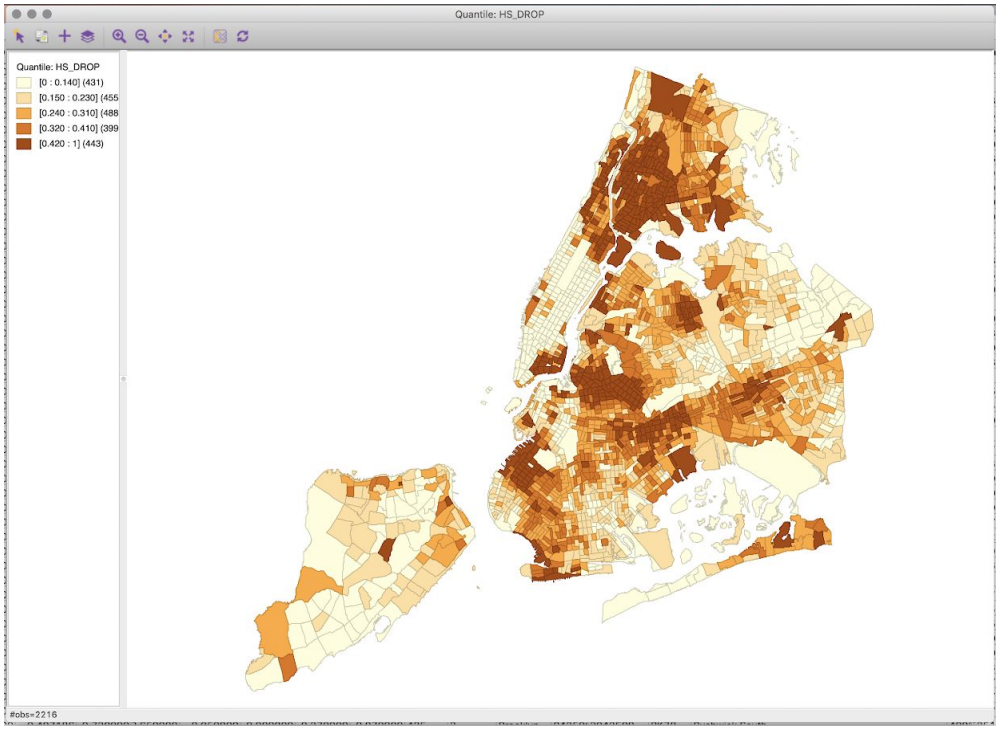

Quantile: HS_DROP
Quantile: HS_DROP
[0 : 0.140] (431)
[0.150 : 0.230] (455
[0.240 : 0.310] (488
[0.320 : 0.410] (399
[0.420 : 1] (443)
#obs=2216

Fig 2:

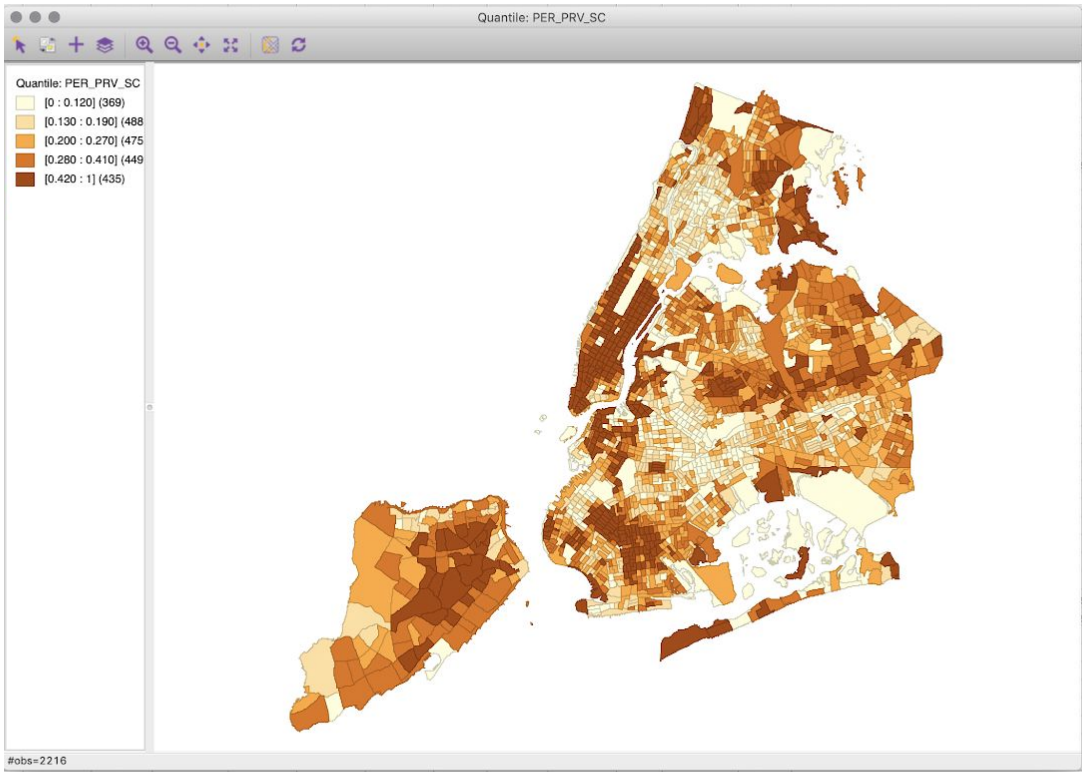

Quantile: PER_PRV_SC
Quantile: PER_PRV_SC
[0 : 0.120] (369)
[0.130 : 0.190] (488
[0.200 : 0.270] (475
[0.280 : 0.410] (449
[0.420 : 1] (435)
#obs=2216

Fig 3:

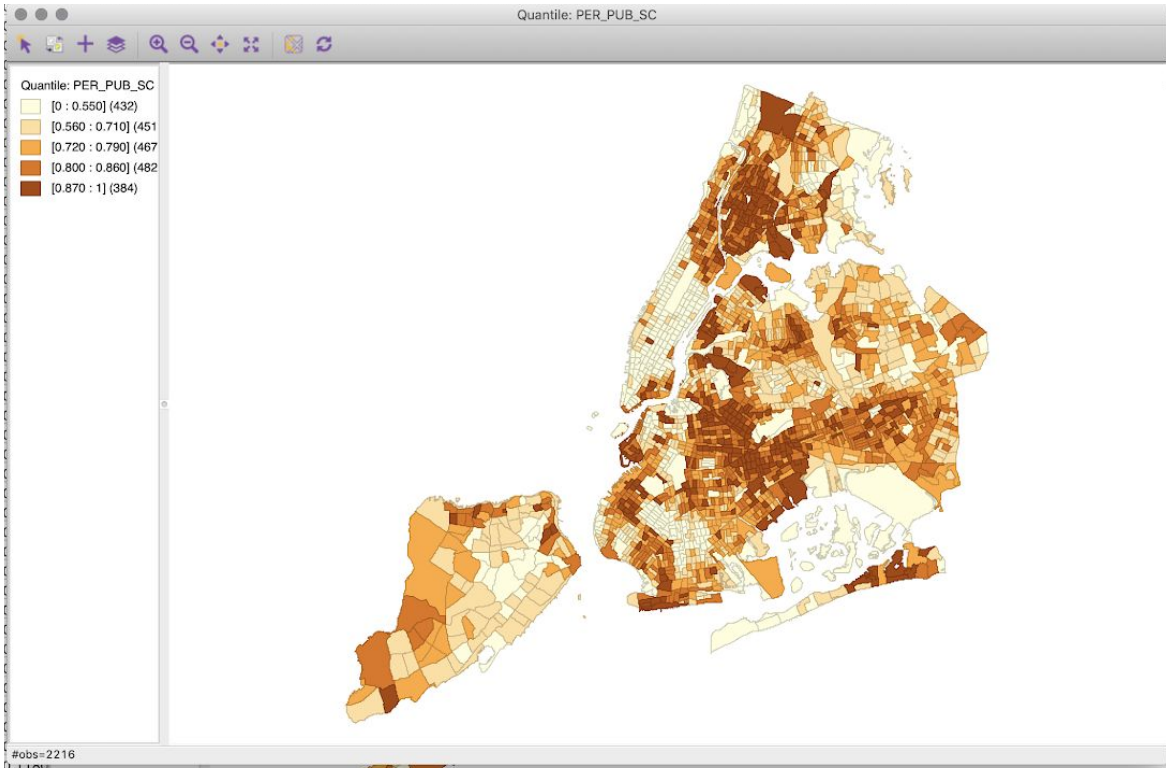

Quantile: PER_PUB_SC
Quantile: PER_PUB_SC
[0 : 0.550] (432)
[0.560 : 0.710] (451
[0.720 : 0.790] (467
[0.800 : 0.860] (482
[0.870 : 1] (384)
#obs=2216

Fig 4:

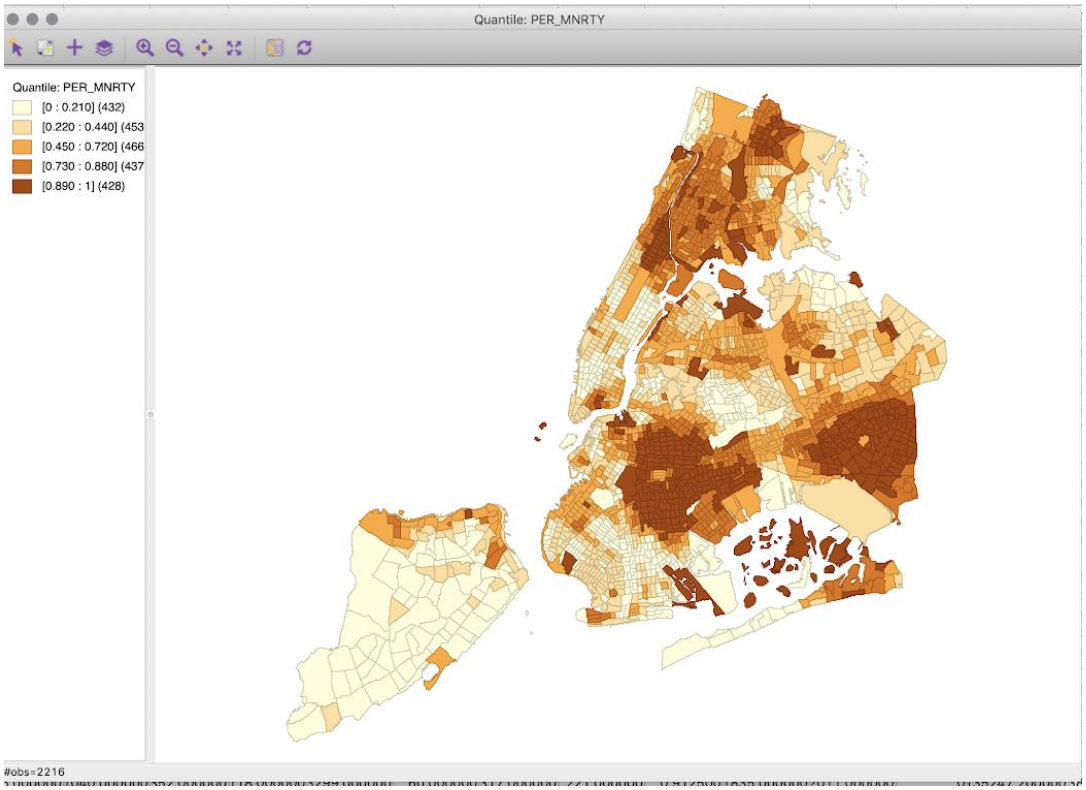

Quantile: PER_MNRTY
Quantile: PER_MNRTY
[0 : 0.210] (432)
[0.220 : 0.440] (453
[0.450 : 0.720] (466
[0.730 : 0.880] (437
[0.890 : 1] (428)
#obs=2216

Fig 5:


Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=217
Histogram: HS_DROP
Frequency
0
150
300
450
600
750
0
0.1
0.3
0.4
0.6
0.7
0.9
1
HS_DROP
#selected=217
Histogram: PER_PRV_SC
Frequency
0
170
340
510
680
850
PER_PRV_SC
#selected=217

Fig 6:

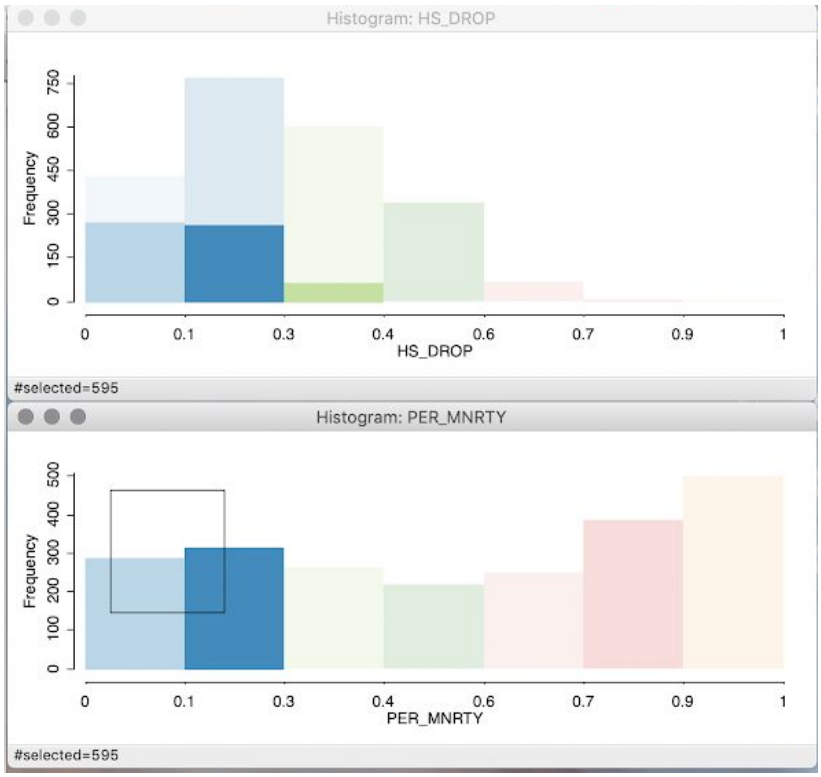

Histogram: HS_DROP
Frequency
0
150
300
450
600
750
0
0.1
0.3
0.4
0.6
0.7
0.9
1
HS_DROP
#selected=595
Histogram: PER_MNRTY
Frequency
0
100
200
300
400
500
PER_MNRTY
#selected=595

Fig 7:

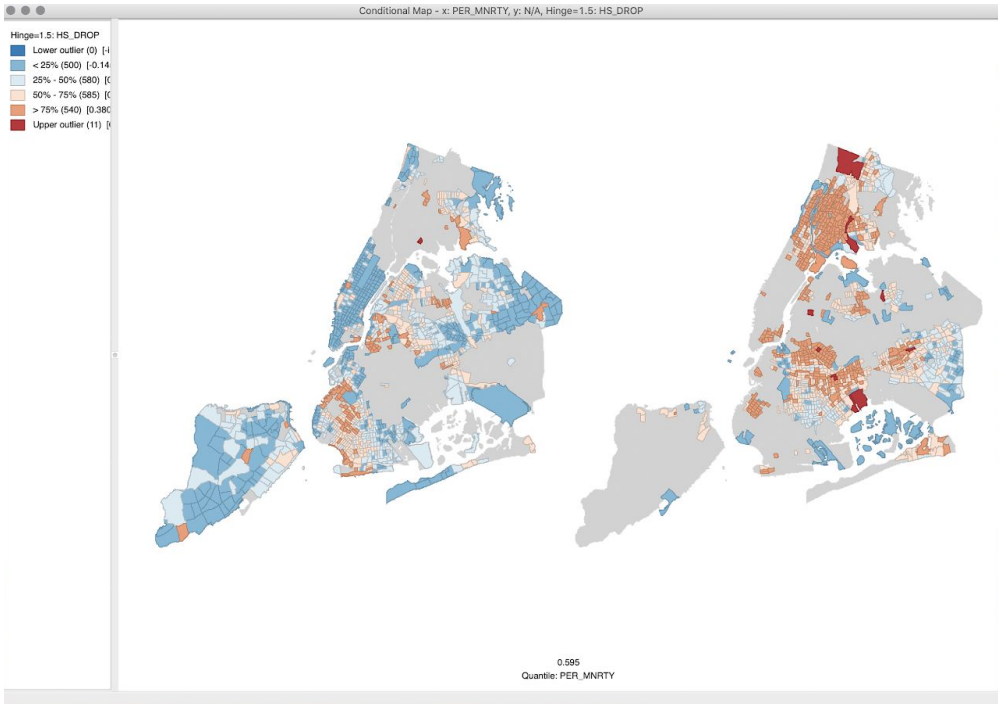

Conditional Map - x: PER_MNRTY, y: N/A, Hinge=1.5: HS_DROP
Hinge=1.5: HS_DROP
Lower outlier (0) [-i
< 25% (500) [-0.14
25% - 50% (580) [0
50% - 75% (585) [0
> 75% (540) [0.380
Upper outlier (11) [
0.595
Quantile: PER_MNRTY

Fig 8:

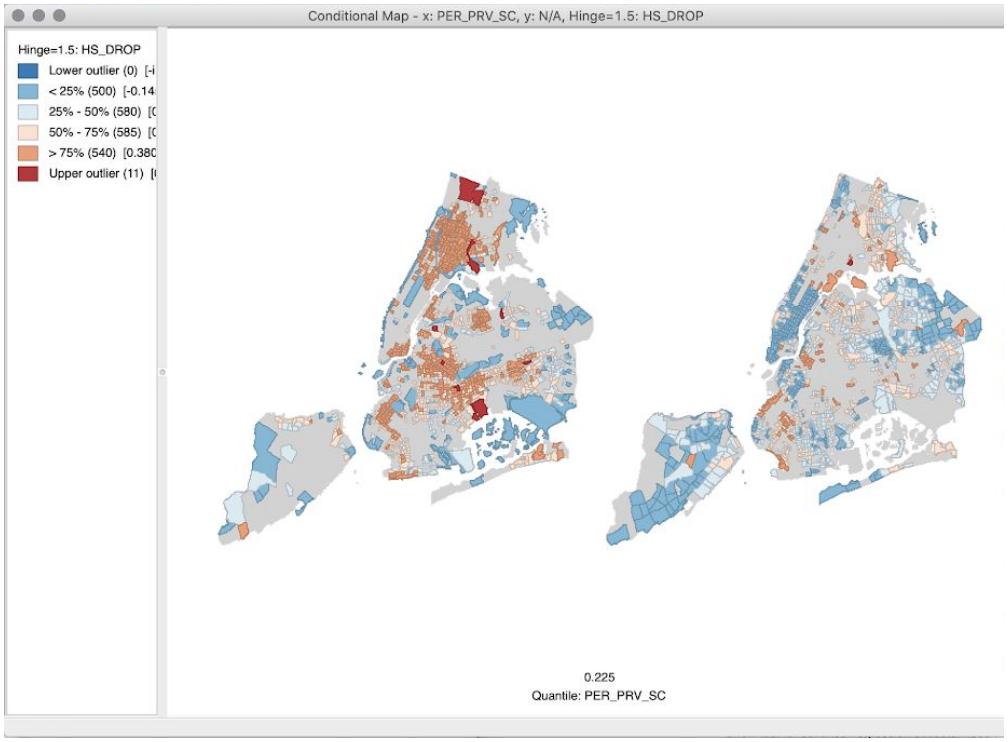

Conditional Map - x: PER_PRV_SC, y: N/A, Hinge=1.5: HS_DROP
Hinge=1.5: HS_DROP
Lower outlier (0) [-i
< 25% (500) [-0.14
25% - 50% (580) [0
50% - 75% (585) [0
> 75% (540) [0.380
Upper outlier (11) [
0.225
Quantile: PER_PRV_SC

Fig 9:

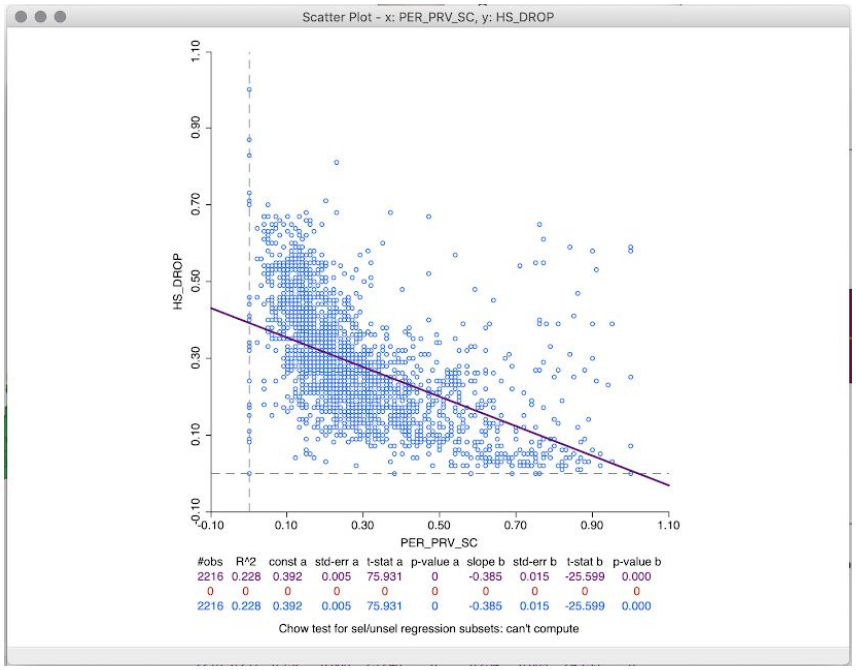

Scatter Plot - x: PER_PRV_SC, y: HS_DROP
HS_DROP
PER_PRV_SC
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.228 0.392 0.005 75.931 0 -0.385 0.015 -25.599 0.000
0 0 0 0 0 0 0 0 0 0
2216 0.228 0.392 0.005 75.931 0 -0.385 0.015 -25.599 0.000
Chow test for sel/unsel regression subsets: can't compute

Fig 10:

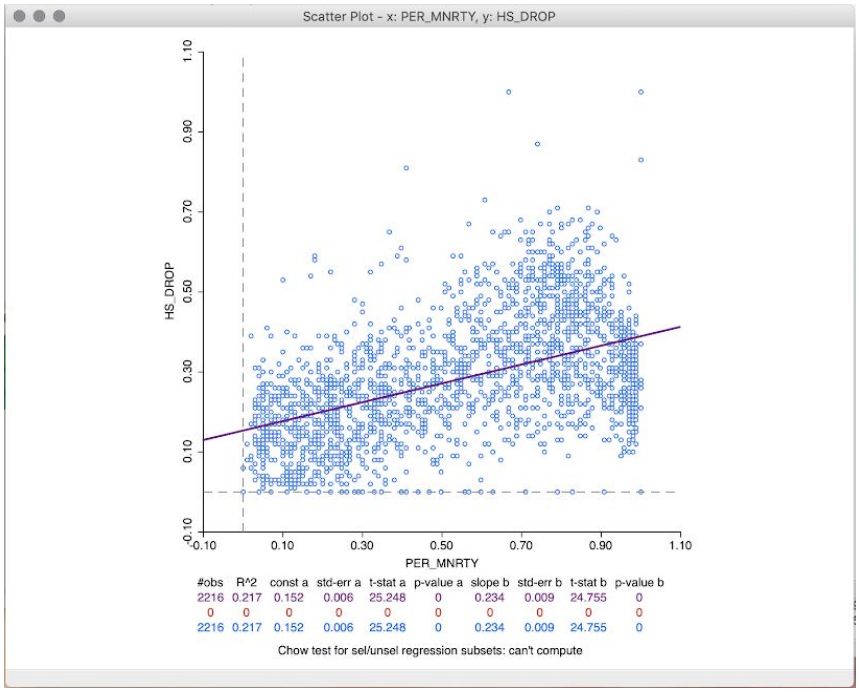

Scatter Plot - x: PER_MNRTY, y: HS_DROP
HS_DROP
PER_MNRTY
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.217 0.152 0.006 25.248 0 0.234 0.009 24.755 0
0 0 0 0 0 0 0 0 0 0
2216 0.217 0.152 0.006 25.248 0 0.234 0.009 24.755 0
Chow test for sel/unsel regression subsets: can't compute

Fig 11:

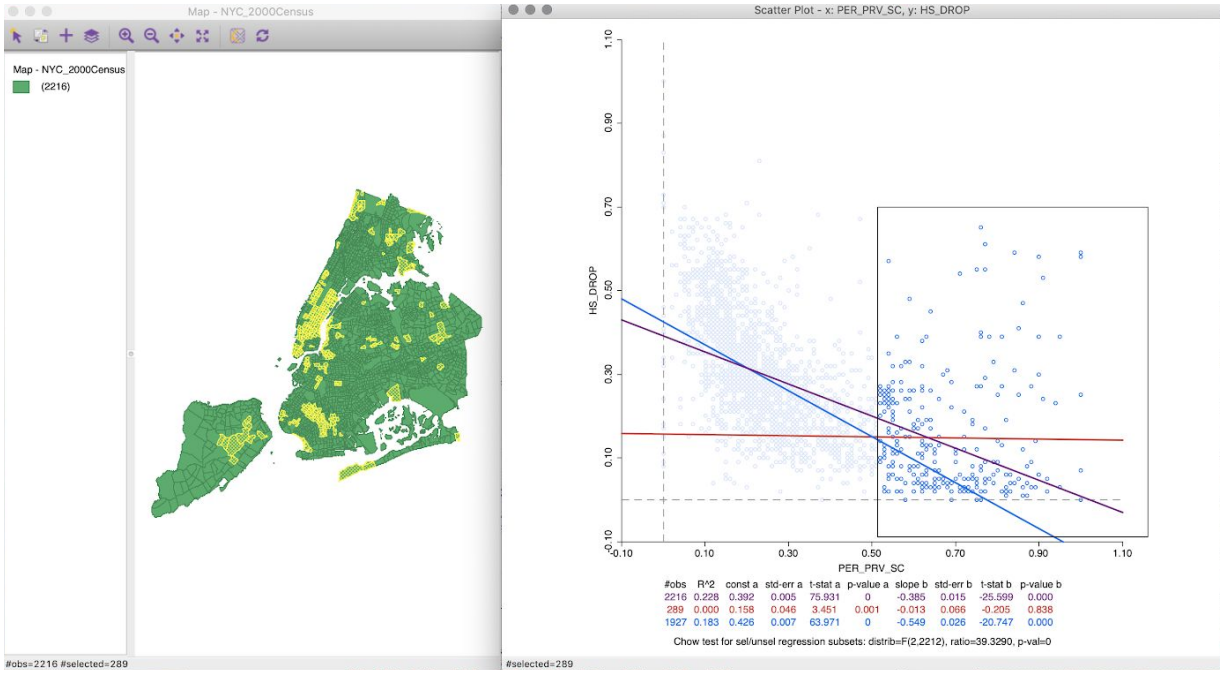

Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=289
Scatter Plot - x: PER_PRV_SC, y: HS_DROP
HS_DROP
PER_PRV_SC
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.228 0.392 0.005 75.931 0 -0.385 0.015 -25.599 0.000
289 0.000 0.158 0.046 3.451 0.001 -0.013 0.066 -0.205 0.838
1927 0.183 0.426 0.007 63.971 0 -0.549 0.026 -20.747 0.000
Chow test for sel/unsel regression subsets: distrib=F(2,2212), ratio=39.3290, p-val=0
#selected=289

Fig 12:

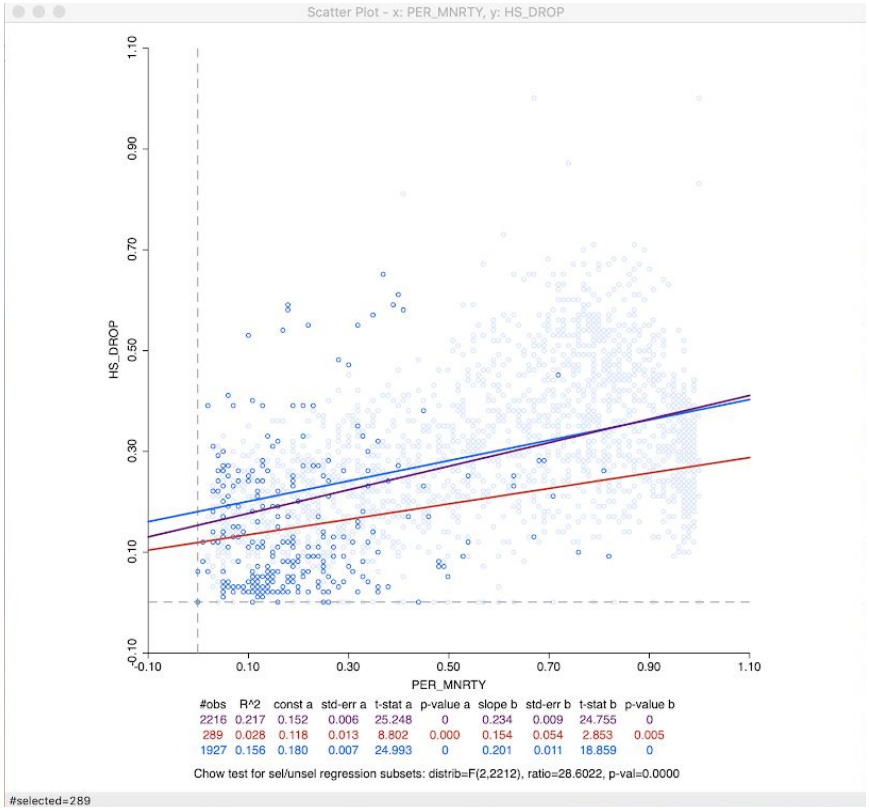

Scatter Plot - x: PER_MNRTY, y: HS_DROP
HS_DROP
PER_MNRTY
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.217 0.152 0.006 25.248 0 0.234 0.009 24.755 0
289 0.028 0.118 0.013 8.802 0.000 0.154 0.054 2.853 0.005
1927 0.156 0.180 0.007 24.993 0 0.201 0.011 18.859 0
Chow test for sel/unsel regression subsets: distrib=F(2,2212), ratio=28.6022, p-val=0.0000
#selected=289

Fig 13:

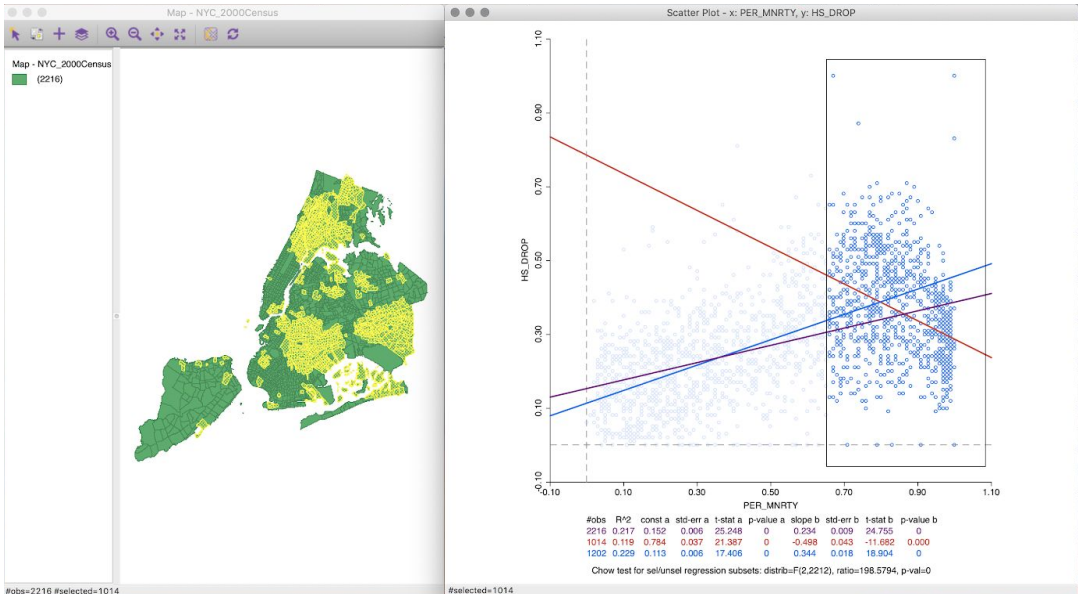

Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=1014
Scatter Plot - x: PER_MNRTY, y: HS_DROP
HS_DROP
PER_MNRTY
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.217 0.152 0.006 25.248 0 0.234 0.009 24.755 0
1014 0.119 0.784 0.037 21.387 0 -0.498 0.043 -11.682 0.000
1202 0.229 0.113 0.006 17.406 0 0.344 0.018 18.904 0
Chow test for sel/unsel regression subsets: distrib=F(2,2212), ratio=198.5794, p-val=0
#selected=1014

Fig 14:

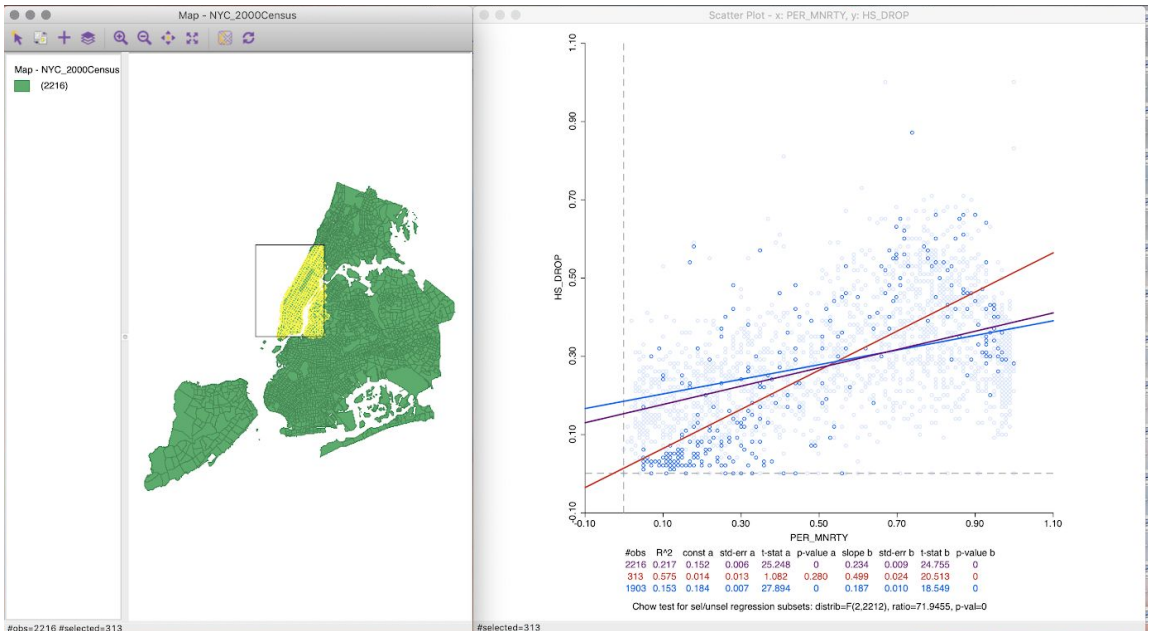

Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=313
Scatter Plot - x: PER_MNRTY, y: HS_DROP
HS_DROP
PER_MNRTY
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.217 0.152 0.006 25.248 0 0.234 0.009 24.755 0
313 0.575 0.014 0.013 1.082 0.280 0.499 0.024 20.513 0
1903 0.153 0.184 0.007 27.894 0 0.187 0.010 18.549 0
Chow test for sel/unsel regression subsets: distrib=F(2,2212), ratio=71.9455, p-val=0
#selected=313

Fig 15:

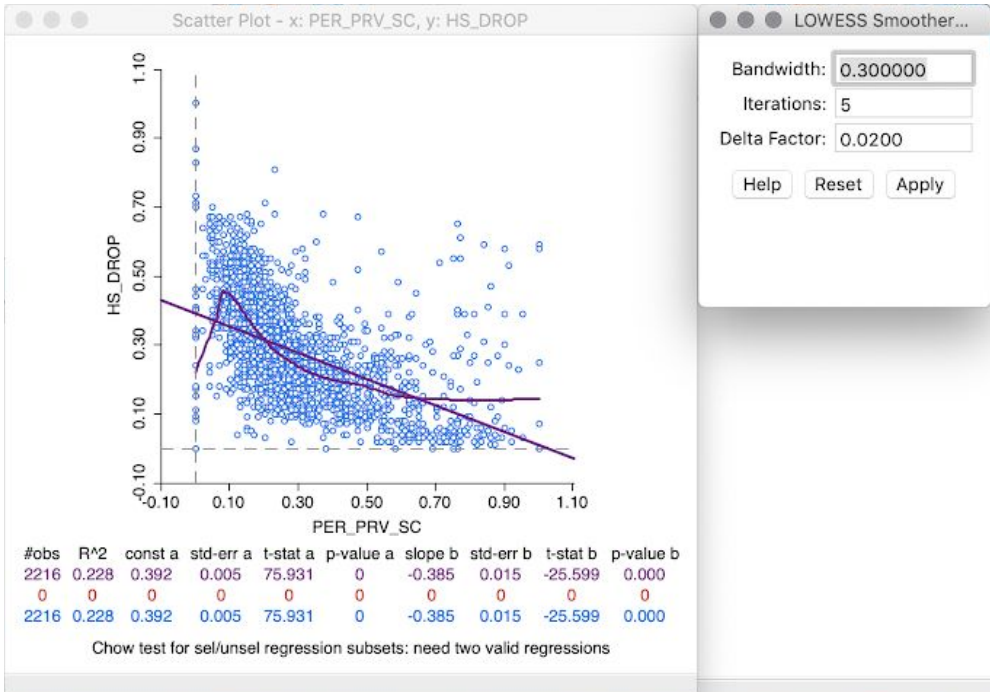

Scatter Plot - x: PER_PRV_SC, y: HS_DROP
HS_DROP
PER_PRV_SC
-0.10
0.10
0.30
0.50
0.70
0.90
1.10
LOWESS Smoother...
Bandwidth: 0.300000
Iterations: 5
Delta Factor: 0.0200
Help
Reset
Apply
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
2216 0.228 0.392 0.005 75.931 0 -0.385 0.015 -25.599 0.000
0 0 0 0 0 0 0 0 0 0
2216 0.228 0.392 0.005 75.931 0 -0.385 0.015 -25.599 0.000
Chow test for sel/unsel regression subsets: need two valid regressions

Fig 16:

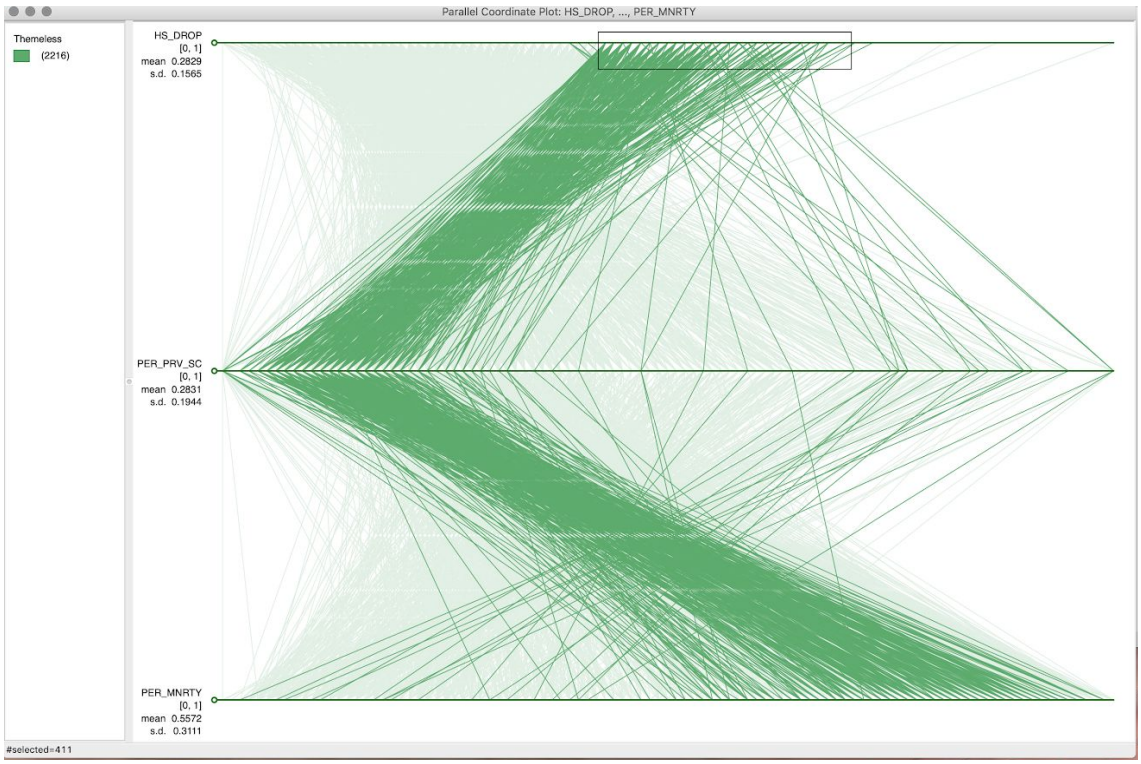

Parallel Coordinate Plot: HS_DROP, ..., PER_MNRTY
Themeless
(2216)
HS_DROP
[0, 1]
mean 0.2829
s.d. 0.1565
PER_PRV_SC
[0, 1]
mean 0.2831
s.d. 0.1944
PER_MNRTY
[0, 1]
mean 0.5572
s.d. 0.3111
#selected=411

Fig 17:

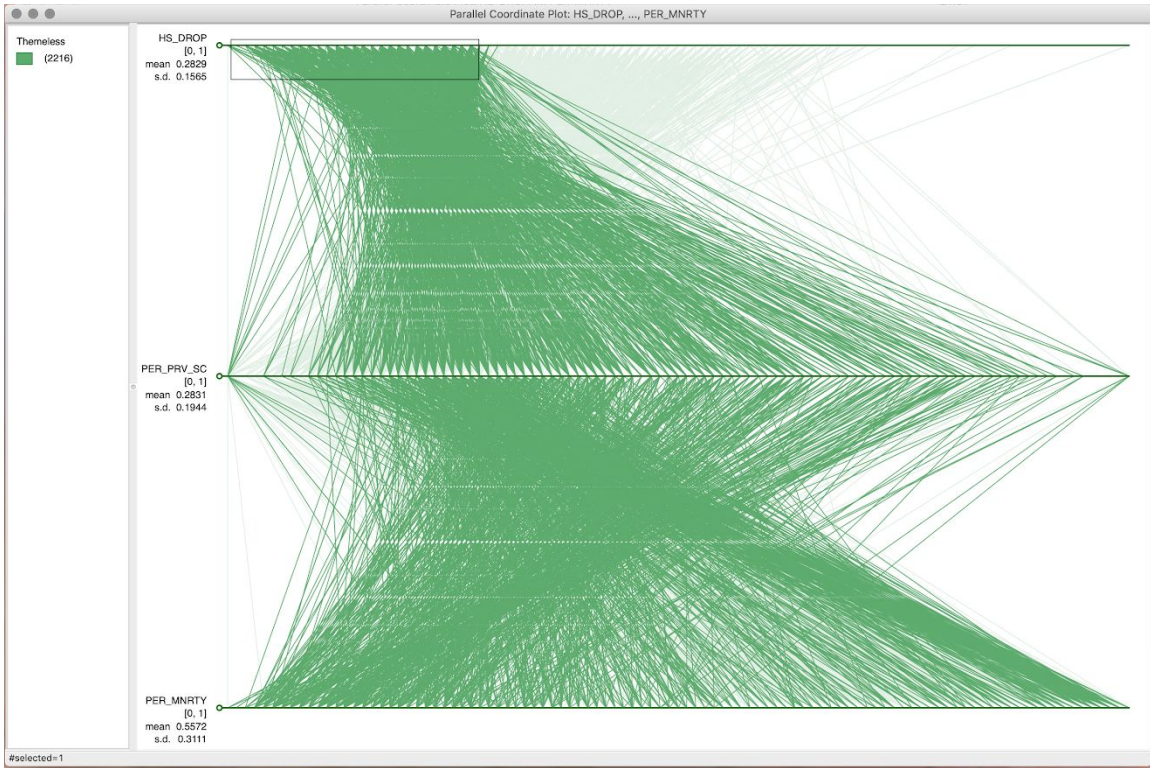

Parallel Coordinate Plot: HS_DROP, ..., PER_MNRTY
Themeless
(2216)
HS_DROP
[0, 1]
mean 0.2829
s.d. 0.1565
PER_PRV_SC
[0, 1]
mean 0.2831
s.d. 0.1944
PER_MNRTY
[0, 1]
mean 0.5572
s.d. 0.3111
#selected=1

Fig 18:

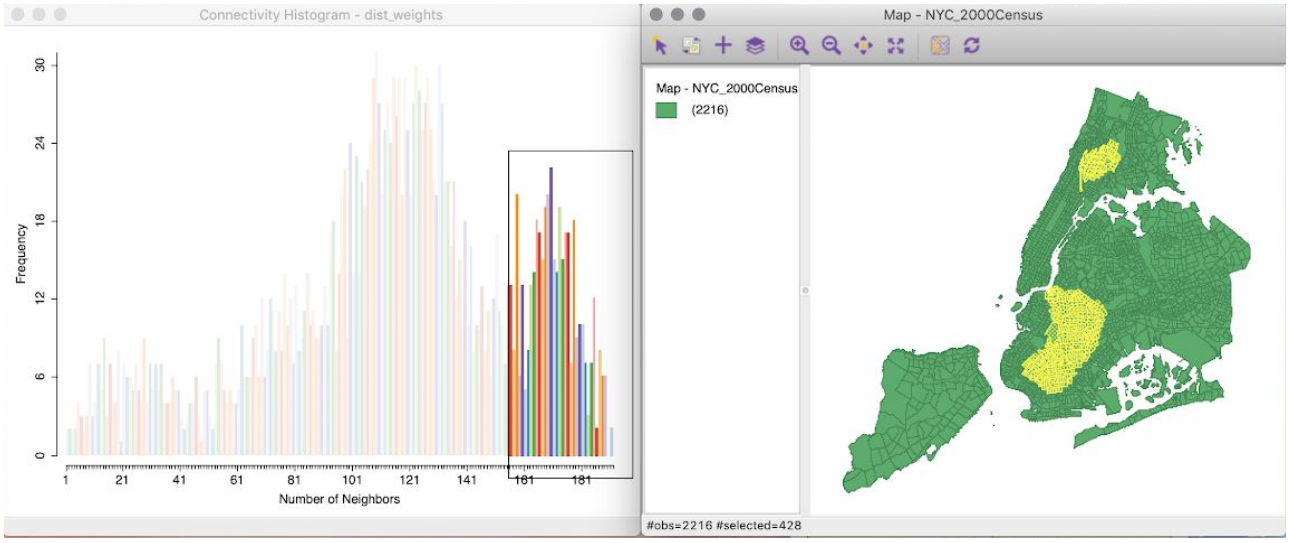

Connectivity Histogram - dist_weights
Frequency
0
6
12
18
24
30
1
21
41
61
81
101
121
141
161
181
Number of Neighbors
Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=428

Fig 19:

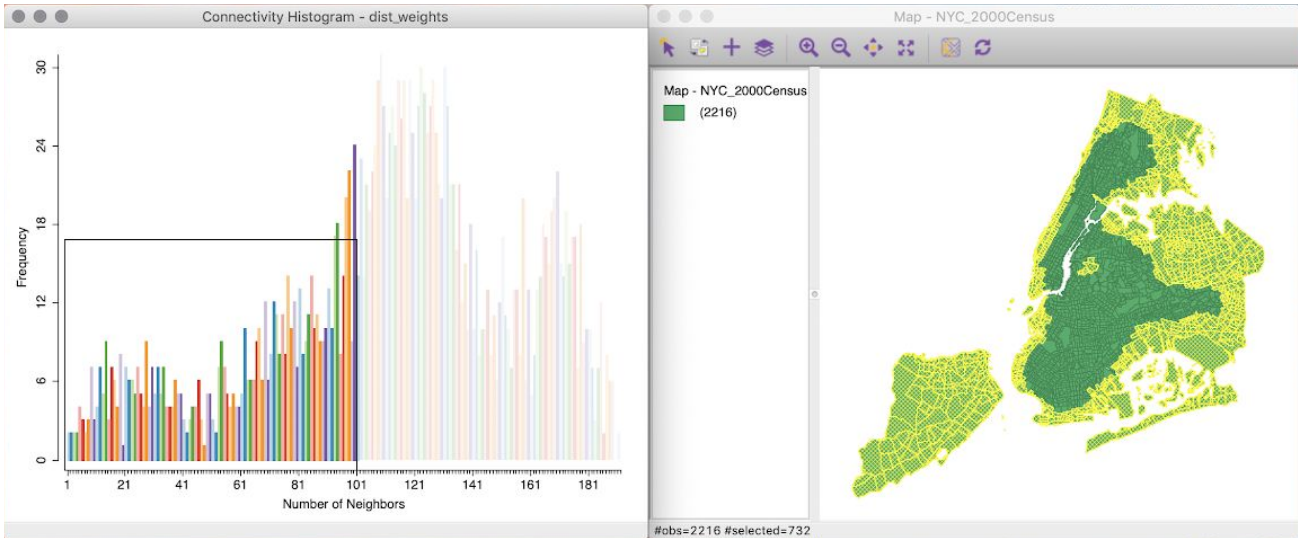

Connectivity Histogram - dist_weights
Frequency
Number of Neighbors
Map - NYC_2000Census
Map - NYC_2000Census
(2216)
#obs=2216 #selected=732

Fig 20:

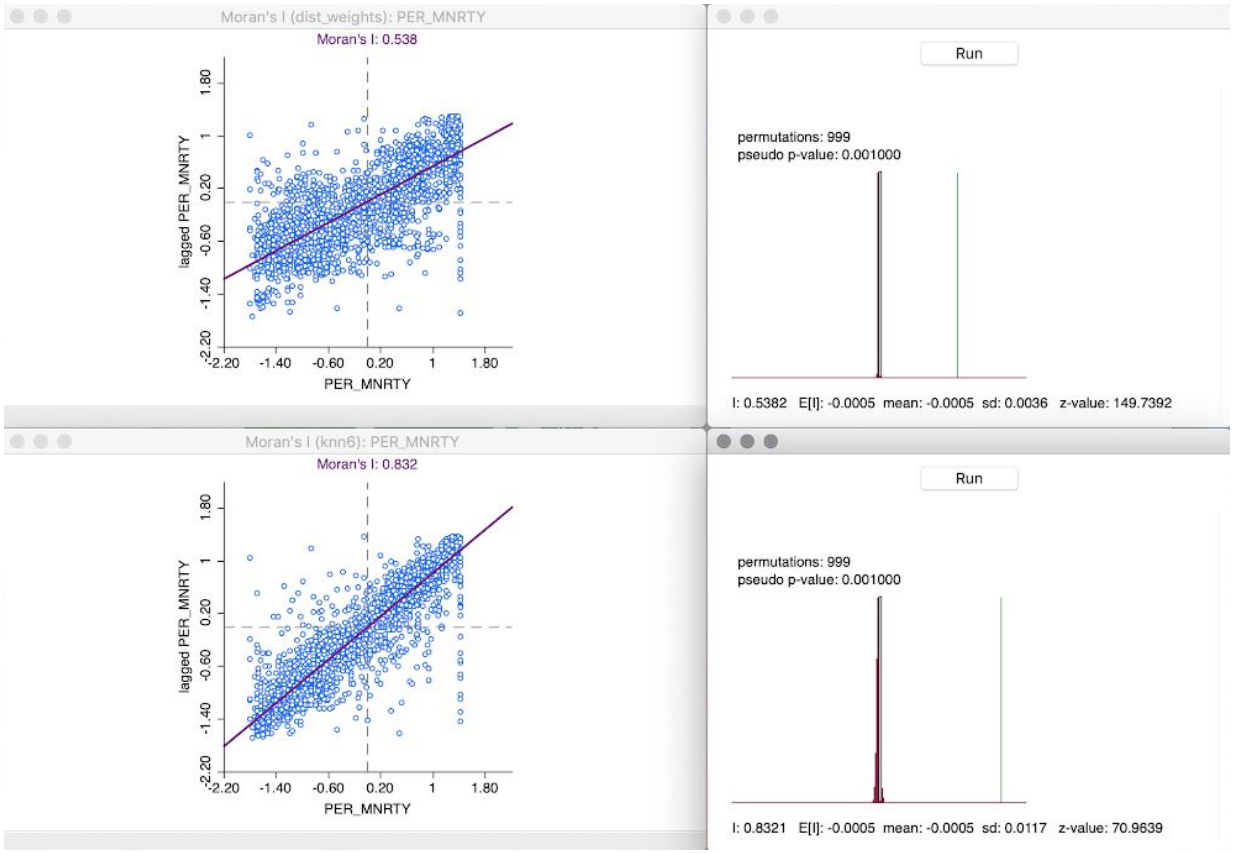

Moran's I (dist_weights): PER_MNRTY
Moran's I: 0.538
lagged PER_MNRTY
PER_MNRTY
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.5382 E[I]: -0.0005 mean: -0.0005 sd: 0.0036 z-value: 149.7392
Moran's I (knn6): PER_MNRTY
Moran's I: 0.832
lagged PER_MNRTY
PER_MNRTY
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.8321 E[I]: -0.0005 mean: -0.0005 sd: 0.0117 z-value: 70.9639

Fig 21:

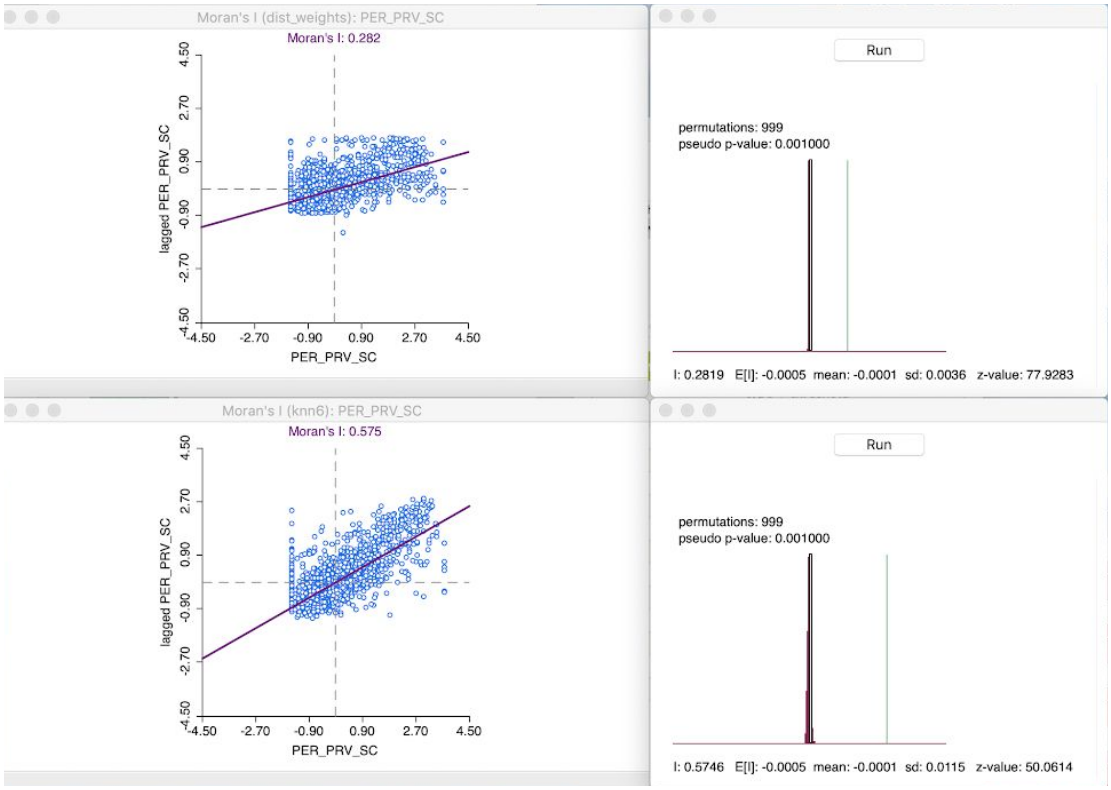

Moran's I (dist_weights): PER_PRV_SC
Moran's I: 0.282
lagged PER_PRV_SC
PER_PRV_SC
4.50
2.70
0.90
-0.90
-2.70
-4.50
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.2819 E[I]: -0.0005 mean: -0.0001 sd: 0.0036 z-value: 77.9283
Moran's I (knn6): PER_PRV_SC
Moran's I: 0.575
lagged PER_PRV_SC
PER_PRV_SC
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.5746 E[I]: -0.0005 mean: -0.0001 sd: 0.0115 z-value: 50.0614

Fig 22:

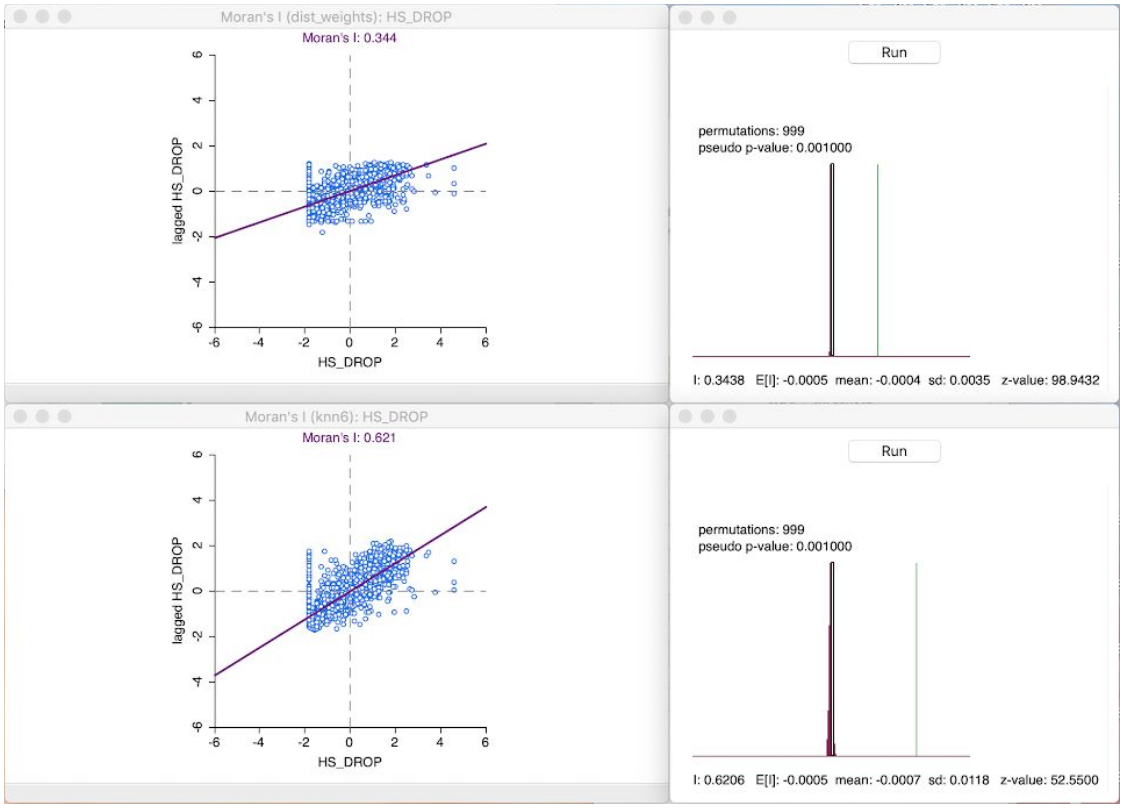

Moran's I (dist_weights): HS_DROP
Moran's I: 0.344
lagged HS_DROP
HS_DROP
6
4
2
0
-2
-4
-6
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.3438 E[I]: -0.0005 mean: -0.0004 sd: 0.0035 z-value: 98.9432
Moran's I (knn6): HS_DROP
Moran's I: 0.621
lagged HS_DROP
HS_DROP
Run
permutations: 999
pseudo p-value: 0.001000
I: 0.6206 E[I]: -0.0005 mean: -0.0007 sd: 0.0118 z-value: 52.5500

Fig 23:

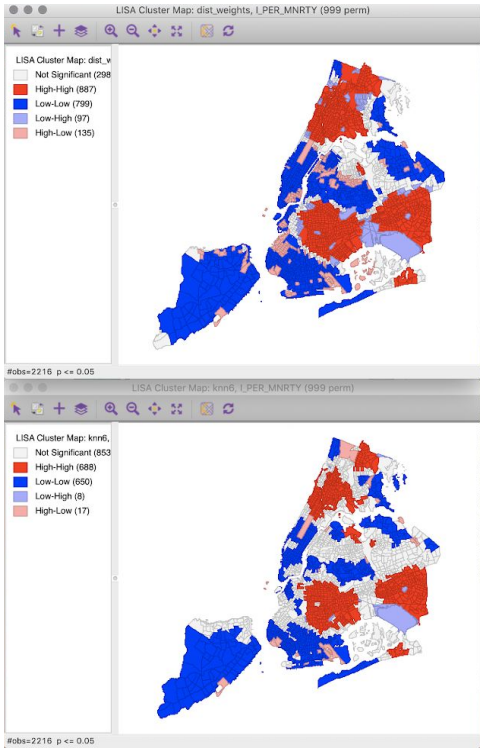

LISA Cluster Map: dist_weights, I_PER_MNRTY (999 perm)
LISA Cluster Map: dist_v
Not Significant (298
High-High (887)
Low-Low (799)
Low-High (97)
High-Low (135)
#obs=2216 p <= 0.05
LISA Cluster Map: knn6, I_PER_MNRTY (999 perm)
LISA Cluster Map: knn6,
Not Significant (853
High-High (688)
Low-Low (650)
Low-High (8)
High-Low (17)
#obs=2216 p <= 0.05

Fig 24:

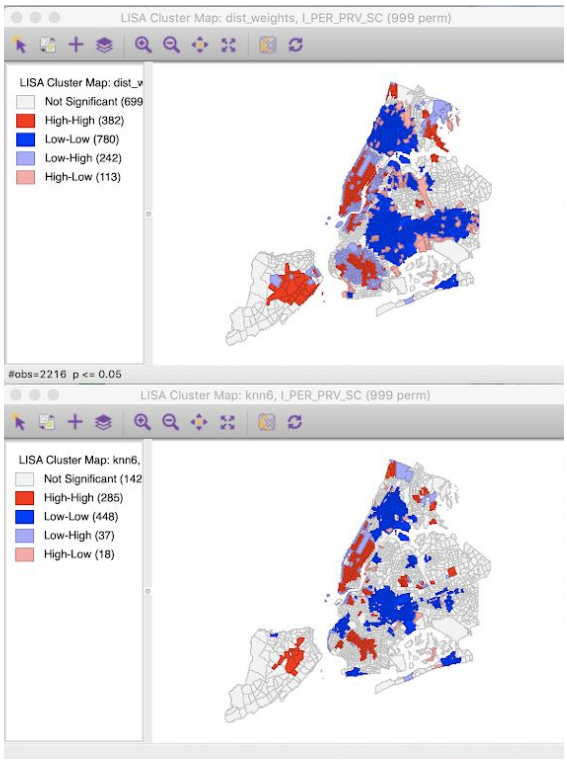

LISA Cluster Map: dist_weights, I_PER_PRV_SC (999 perm)
LISA Cluster Map: dist_v
Not Significant (699
High-High (382)
Low-Low (780)
Low-High (242)
High-Low (113)
#obs=2216 p <= 0.05
LISA Cluster Map: knn6, I_PER_PRV_SC (999 perm)
LISA Cluster Map: knn6,
Not Significant (142
High-High (285)
Low-Low (448)
Low-High (37)
High-Low (18)

Fig 25:

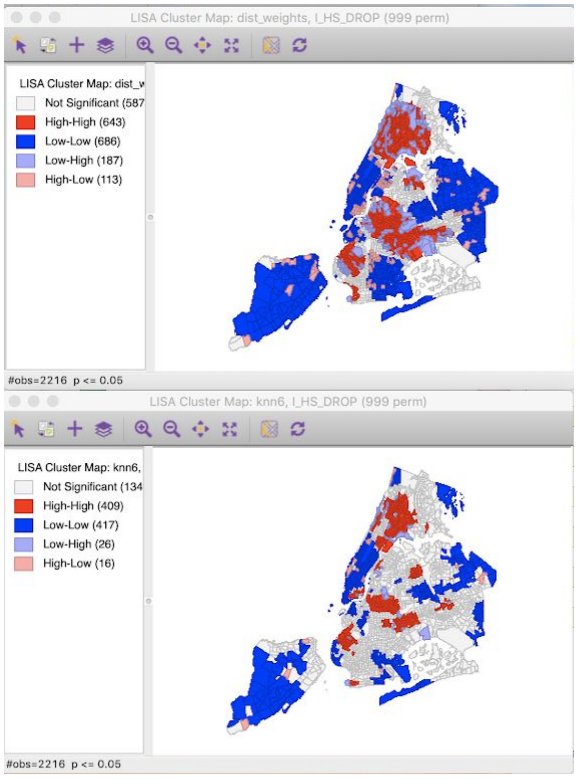

LISA Cluster Map: dist_weights, I_HS_DROP (999 perm)
LISA Cluster Map: dist_w
Not Significant (587
High-High (643)
Low-Low (686)
Low-High (187)
High-Low (113)
#obs=2216 p <= 0.05
LISA Cluster Map: knn6, I_HS_DROP (999 perm)
LISA Cluster Map: knn6,
Not Significant (134
High-High (409)
Low-Low (417)
Low-High (26)
High-Low (16)
#obs=2216 p <= 0.05

Fig 26:

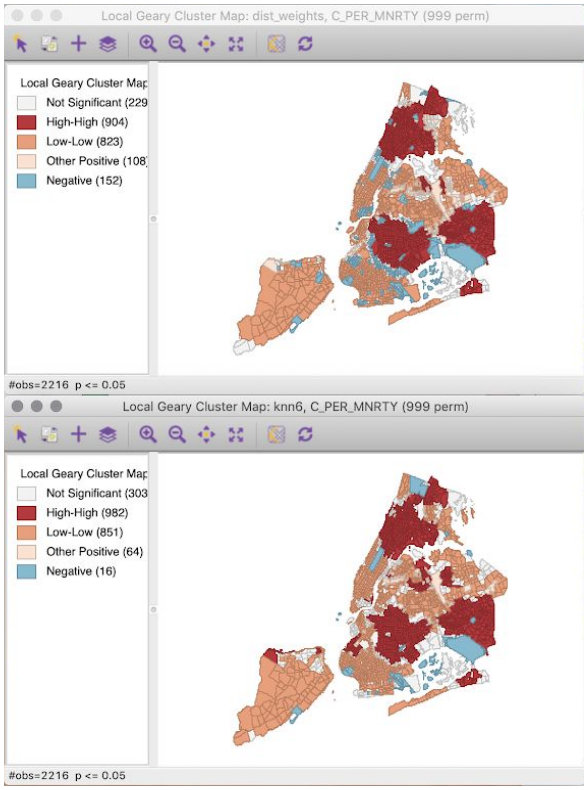

Local Geary Cluster Map: dist_weights, C_PER_MNRTY (999 perm)
Local Geary Cluster Map
Not Significant (229
High-High (904)
Low-Low (823)
Other Positive (108
Negative (152)
#obs=2216 p <= 0.05
Local Geary Cluster Map: knn6, C_PER_MNRTY (999 perm)
Local Geary Cluster Map
Not Significant (303
High-High (982)
Low-Low (851)
Other Positive (64)
Negative (16)
#obs=2216 p <= 0.05

Fig 27:

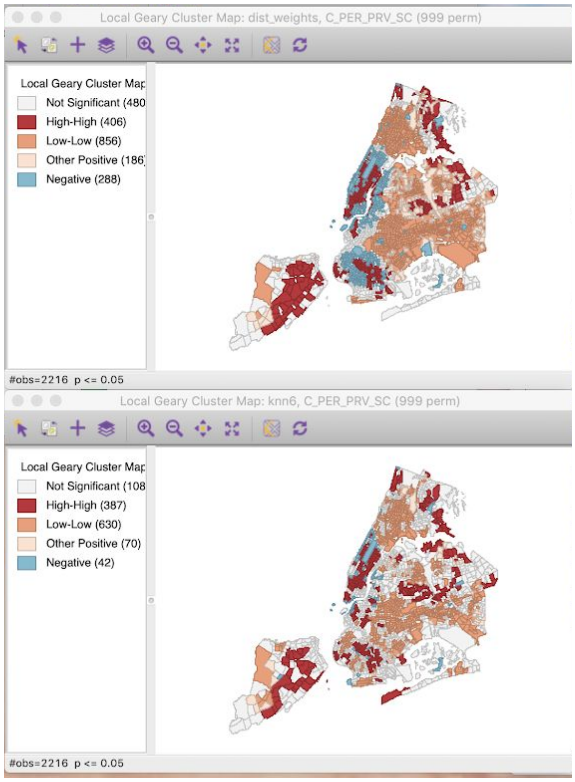

Local Geary Cluster Map: dist_weights, C_PER_PRV_SC (999 perm)
Local Geary Cluster Map
Not Significant (480
High-High (406)
Low-Low (856)
Other Positive (186
Negative (288)
#obs=2216 p <= 0.05
Local Geary Cluster Map: knn6, C_PER_PRV_SC (999 perm)
Local Geary Cluster Map
Not Significant (108
High-High (387)
Low-Low (630)
Other Positive (70)
Negative (42)
#obs=2216 p <= 0.05

Fig 28:

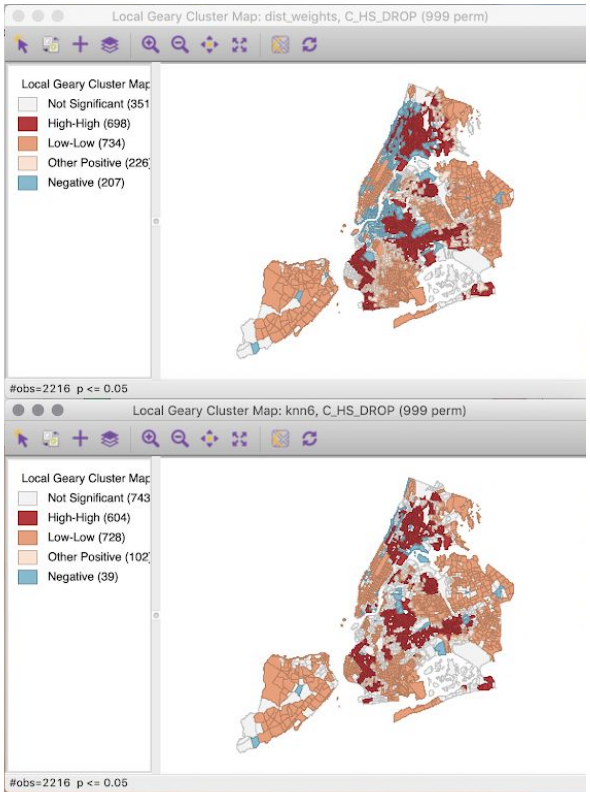

Local Geary Cluster Map: dist_weights, C_HS_DROP (999 perm)
Local Geary Cluster Map
Not Significant (351
High-High (698)
Low-Low (734)
Other Positive (226
Negative (207)
#obs=2216 p <= 0.05
Local Geary Cluster Map: knn6, C_HS_DROP (999 perm)
Local Geary Cluster Map
Not Significant (743
High-High (604)
Low-Low (728)
Other Positive (102
Negative (39)
#obs=2216 p <= 0.05

Fig 29:

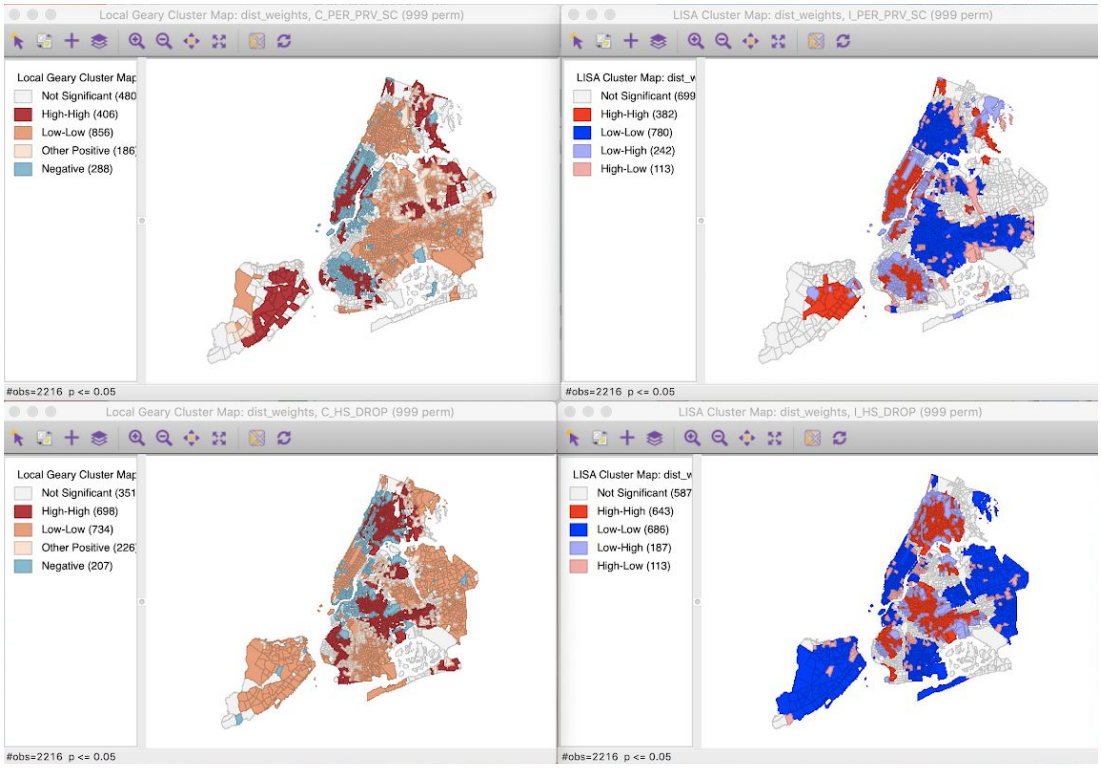

Local Geary Cluster Map: dist_weights, C_PER_PRV_SC (999 perm)
Local Geary Cluster Map
Not Significant (480
High-High (406)
Low-Low (856)
Other Positive (186
Negative (288)
#obs=2216 p <= 0.05
LISA Cluster Map: dist_weights, I_PER_PRV_SC (999 perm)
LISA Cluster Map: dist_v
Not Significant (699
High-High (382)
Low-Low (780)
Low-High (242)
High-Low (113)
#obs=2216 p <= 0.05
Local Geary Cluster Map: dist_weights, C_HS_DROP (999 perm)
Local Geary Cluster Map
Not Significant (351
High-High (698)
Low-Low (734)
Other Positive (226
Negative (207)
#obs=2216 p <= 0.05
LISA Cluster Map: dist_weights, I_HS_DROP (999 perm)
LISA Cluster Map: dist_v
Not Significant (587
High-High (643)
Low-Low (686)
Low-High (187)
High-Low (113)
#obs=2216 p <= 0.05

Fig 30:

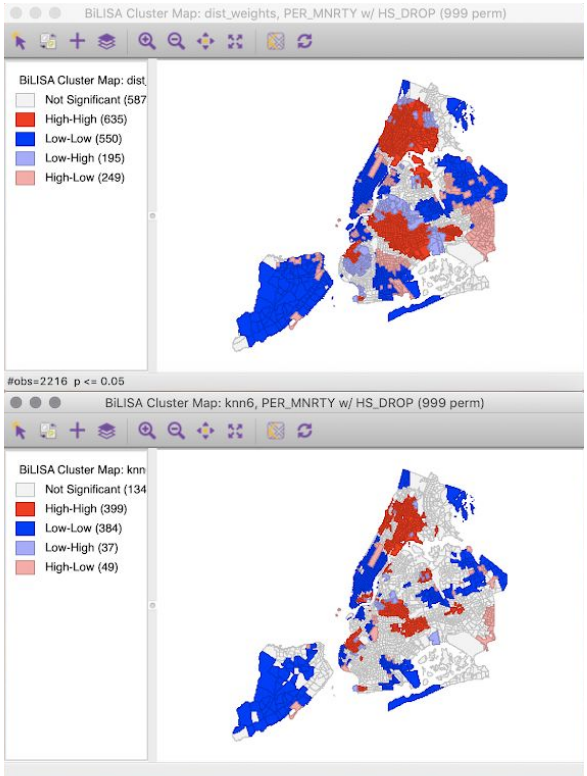
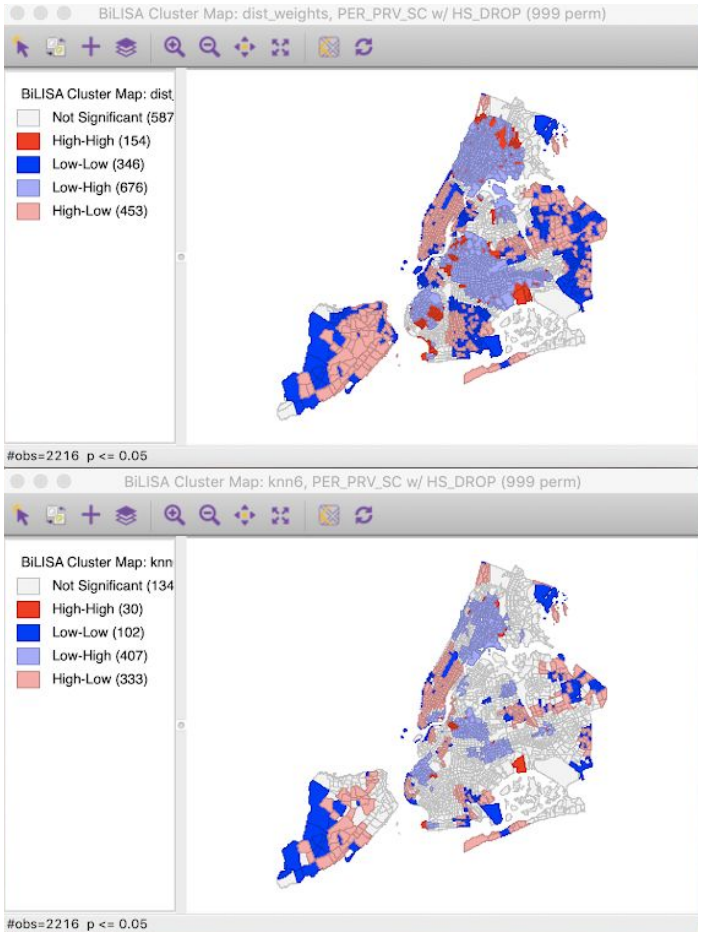

BiLISA Cluster Map: dist_weights, PER_MNRTY w/ HS_DROP (999 perm)
BiLISA Cluster Map: dist_
Not Significant (587
High-High (635)
Low-Low (550)
Low-High (195)
High-Low (249)
#obs=2216 p <= 0.05
BiLISA Cluster Map: knn6, PER_MNRTY w/ HS_DROP (999 perm)
BiLISA Cluster Map: knn
Not Significant (134
High-High (399)
Low-Low (384)
Low-High (37)
High-Low (49)

Fig 31:


BiLISA Cluster Map: dist_weights, PER_PRV_SC w/ HS_DROP (999 perm)
BiLISA Cluster Map: dist_
Not Significant (587
High-High (154)
Low-Low (346)
Low-High (676)
High-Low (453)
#obs=2216 p <= 0.05
BiLISA Cluster Map: knn6, PER_PRV_SC w/ HS_DROP (999 perm)
BiLISA Cluster Map: knn
Not Significant (134
High-High (30)
Low-Low (102)
Low-High (407)
High-Low (333)
#obs=2216 p <= 0.05