# Research Methods

## Scope

We conducted this research by collecting secondary geospatial data from public sources at the census tract level and investigating correlation and other geospatial relationships between them. This general investigation intended to answer two questions, the first general and the second specific. First, is there a relationship between socioeconomic indicators of vitality and environmental quality indicators? This question was investigated at a county-wide scope and broken down by development age through mapping and correlational analysis. Second, what political or environmental forces shaped the boundaries of these areas? Through the previous analysis we were able to identify neighborhoods that were similar in development age but contrasted along environmental quality. We used these as case studies, investigating the environmental, socioeconomic and public policy histories that might have impacted these “sibling neighborhoods” in their development.

## County Univariate Analysis

We first conducted simple mapping and graphing of single variables to determine the context of the county’s environmental and socioeconomic status. Box plots with scatter plot values were also created for each variable, using the plot.ly web app, to show simple summary statistics over the county. These univariate analyses were used to support further investigation, and to begin the search for areas of case study interest. Choropleth maps with 7 Jenck’s “natural breaks” value bins were created for each variable, using QGIS, to show similar amounts of variation across the variables, which all use different units.

Finally, maps were created using the Getis-Ord Gi\* Cluster Analysis tool in Esri ArcMap, which identify clusters of low and high values as well as outliers, for all main variables. This analysis was used to determine whether the individual variables showed similar clustering behavior over the county’s tracts. This tool was useful, but could be extended. We found no quantitative spatial analysis tool that could compare two variables meaningfully. This will be outlined further in Limitations below.

## Bivariate & Temporal Analyses

After the univariate mapping and graphing procedures we conducted bivariate mapping and graphing across all the variable pairs. The shapefile data was exported to Excel spreadsheets, through which we applied Pearson’s *r* column correlation for all pairs. This was our main analysis to establish an initial quantitative relationship between environmental and socioeconomic quality. The strength of these correlations suggested that we investigate further into the temporal relationship of these phenomena.

By aggregating Franklin County Auditor tract-level build dates to the a county tract median, we were able to proxy “development ages” for 234 of Franklin County’s 274 tracts. Using these, we broke the set of tracts into 5 *age cohorts* (using Jenck’s natural breaks). We then conducted the same Pearson’s *r* column correlation using Excel, and mapped these on a multi-line graph. We believe this shows meaningful changes in outcome between different eras of sociopolitical and environmental housing.

Finally, we created bivariate maps of key variable pairs to visually assess the spatial overlap of low and high areas, as well as identify case study candidates. Our most common method for this procedure was to overlay a choropleth with scaled circles representing the second variable. This again was in lieu of a more sophisticated bivariate spatial analysis tool.

## Case Study

Our final investigation was into a case study area representing two adjacent neighborhoods, Bexley and Near East/South Side. This was meant to be a contrast case study. By identifying one of the stronger boundaries of difference in the county and investigating the area’s background, we hope to bring to light some of the strongest natural, political, and cultural forces that can shape the socioeconomic and environmental quality of a place in tandem.

Our case study first involved a review of variable values in relation to their means, as well a a general background in demographics and development for both neighborhoods. We then reviewed historical policies of economic and environmental discrimination that might have contributed to the current state of the neighborhoods, as well as natural circumstances of site which may play a more foundational role. While the main aim of this research is to begin a discussion and investigation of the quantifiable relationship between socioeconomic and environmental quality, our case study adds suggestions for changes in future policy to correct environmental inequity and segregation.

## Variables & Data Sources

The following are descriptions and supporting explanation for the indicators used during this research. Versions of our dataset can be found at [LINK NEEDED].

*Franklin County Census Tract Shapefile*

A census tract shapefile extracted from the Ohio dataset from Census.gov (2014). This served as our template with which to link all other data. After initial investigation available census variable within the shapefile, we edited it down to include only white population % (WP) and housing vacancy % (HV), explained below.

*Opportunity Index (OI)*

Created by Reece et al. at the Kirwan Institute for Race & Ethnicity in OSU, this composite metric measures expected life outcomes and neighborhood vitality more robustly than a simple poverty index (2013). It has been calculated at the census tract level. We took it as a simple measure on its own, but there is much more investigation that can be done by investigating the relationship between environmental indicators and the OI’s component variables individually. The OI approximated the balance of socioeconomic assets and liabilities for our research.

*Mean Canopy Coverage % (MCC)*

We used MCC as a proxy for environmental assets. Trees, while being one of the simplest ecological systems to remotely sense, are one the most impactful aspects on air quality, psychological comfort, and property value within the urban built environment. MCC behaved as an environmental asset in our research.

It is represented in a national raster dataset from USGS National Land Cover Database 2011 showing % of green tree coverage, taken at a 30 meter resolution. After clipping to Franklin county, census tract mean percentages were calculated in ArcMap using the Zonal Statistics tool. We used mean instead of median because it proved to show more nuance in value, especially in rural tracts where coverage can vary drastically with an unstable median.

*Mean % Impervious Surface Coverage (MISC)*

MISC was used to proxy environmental stressors. More robust proxies of this exist that measure pollution, air quality, and other stressors as well, but we sought to focus on environmental assets’ relationship with the OI in this exploratory inquiry. MISC behaved as an environmental liability in our research.

It also represented in a national raster dataset from USGS National Land Cover Database 2011 showing % of green tree coverage, taken at a 30 meter resolution. After clipping to Franklin county, census tract mean percentages were calculated in ArcMap using the Zonal Statistics tool. We used mean instead of median because it proved to show more nuance in value, especially in rural tracts where coverage can vary drastically with an unstable median.

*White Population % (WP)*

WP was used to investigate the environmental dimension of a known high correlation between WP and the OI. Furthermore, the case studies of this research explored historical discriminatory housing policy, along lines of wealth and race, often favored predominantly white neighborhoods over other ethnicities. In our research it behaved as a socioeconomic asset, however we believe that this is due to racial biases in housing and economic policy that the presence of high white population accompanies socioeconomic vitality, not due to any inherent superiority or tenacity of the population itself. This variable was within the Census Bureau shapefile, mentioned above.

*Housing Vacancy % (HV)*

This indicator was used to explore housing’s relationship to both the OI and environmental variables. It stood in as a measure of neighborhood disinvestment, which we suspected might indicate disinvestment in environmental quality as well.In our research it behaved as a socioeconomic liability. This variable was within the Census Bureau shapefile, mentioned above.

*Median Built Date (MBD)*

This indicator was used to divide up Franklin County’s census tracts into development eras. These MBD *cohorts* were used to both track trends in the correlation between environmental and socioeconomic variables, and to help identify the sibling neighborhoods to be explored further through case study.

MBD was calculated using zonal statistics at the census tract level for the build date variable in the Franklin County Auditor’s housing data, recorded at the plot level. Median build date showed a truer central character of each tract than the mean. It controlled for an old tract being redeveloped much later, which under a mean would return an era in which neither period of development is represented.

## Limitations

This research was highly exploratory in nature, and as such met several sizeable limitations to conclusions and recommendations. First with regards to data, was the resolution and size of the data sets. Census tracts are often misaligned with neighborhood borders, and don’t reflect the functioning semantic divisions of the city that often guide discriminatory policy. Furthermore, these semantic divisions shift continuously throughout history, so place over time might not best track the varying environmental segregation.

Sample size was another issue: there are only 274 census tracts in Franklin County, and each age cohort had only 20-60 tracts within it. Correlation values were used to demonstrate the hint of a powerful and measurable relationship between economic, ethnic, and environmental inequality, and should be taken primarily as motivational evidence for further and more robust lines of inquiry..

Multivariate spatial analysis is a young field, and we encountered difficulty in finding a tool to quantitatively compare the geographic distribution of values for one variable with another. Further development of such robust geographic tools will be necessary for future investigation along the intersection of environmental and socioeconomic assessments.

Finally, the issue of causation posed an impenetrable web of connections for us. Determining whether a land use policy or existing site conditions contributed more to an area’s poor environmental quality is outside the scope of this research. Our case study investigates a highly contrasting area for this reason; to suggest what might influence the strongest differences in the county. Further research controlling for pre-development site conditions or land value would be beneficial to understand different policy impacts.

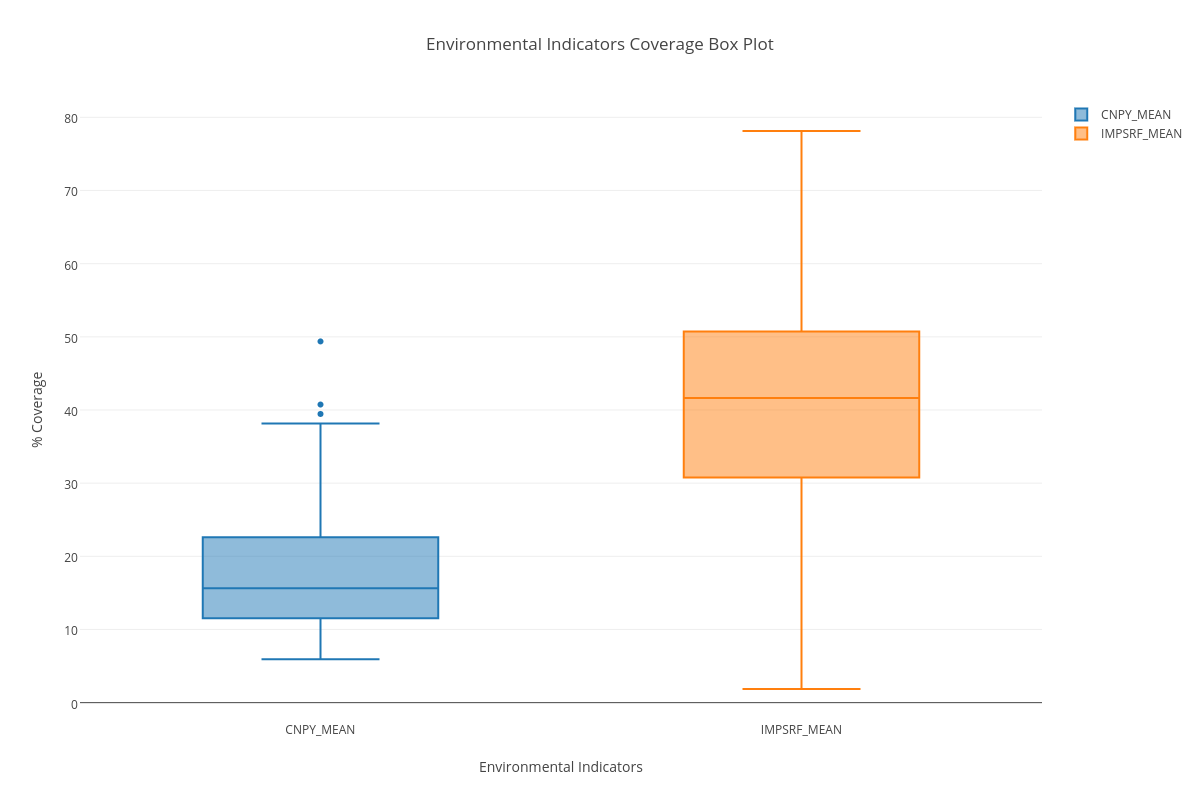
# Data Analysis

Our analysis shows a cursory picture of the landscape of environmental assets and stressors in Columbus, as well as their relationship with socioeconomic variables such as the OI. As stated above, we moved from general univariate analysis of the county to bivariate and temporal analyses, then to a case study exploring historical inequitable policy.

## County Univariate Analysis

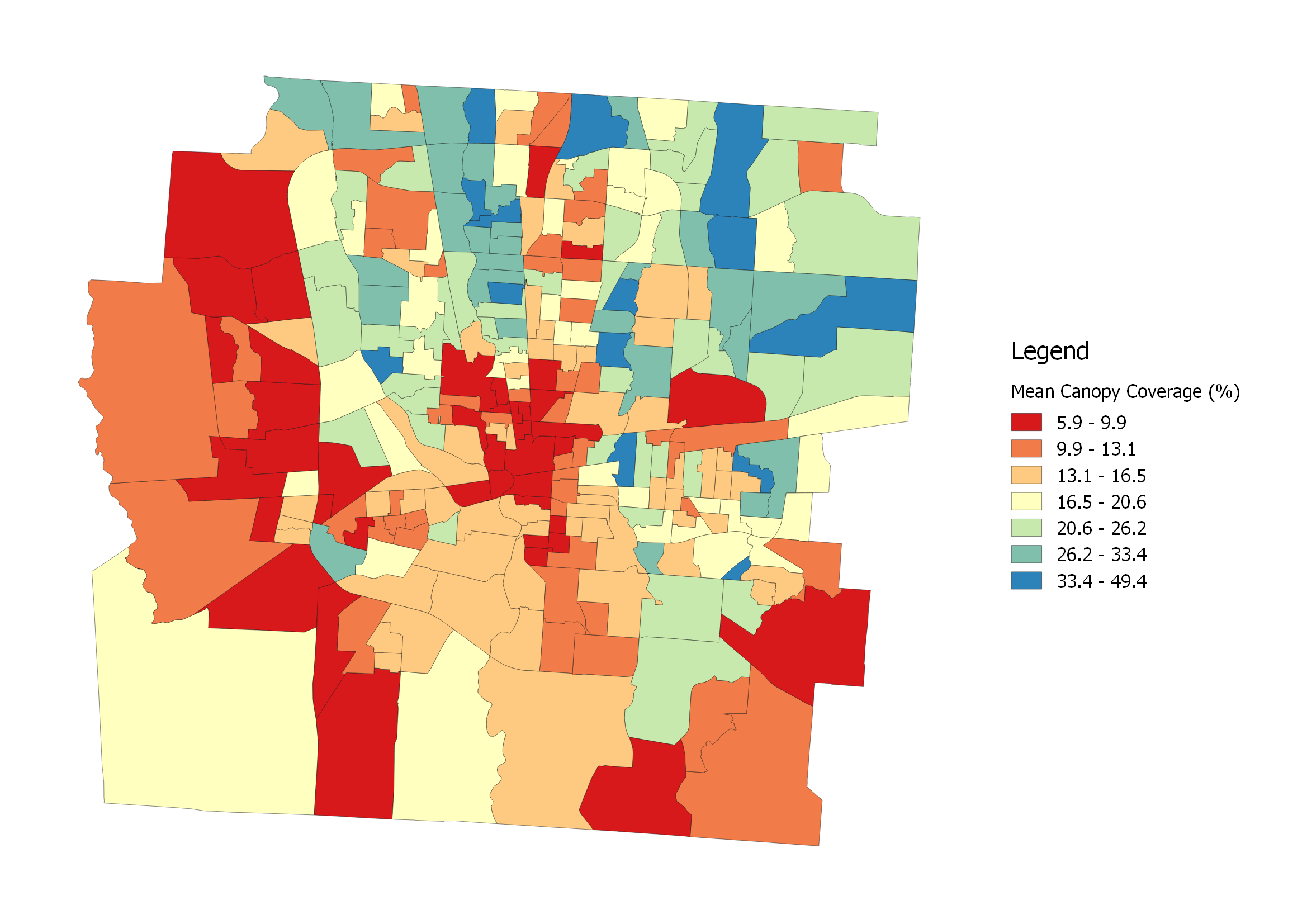
### Environmental Indicators

Franklin County shows moderate to high levels of MISC compared to MCC on the whole, with a mean of 41.6% coverage, well above MCC’s 15.6% mean (Figure X). The variation was also much greater with MISC, which had a standard deviation of 15.2%, while MCC’s was 7.9%. In general, Franklin County has much more impervious surface coverage than tree canopy.



*Figure 1: Boxplot of environmental variables over Franklin County*

The choropleth maps below show the geographical layout of MCC and MISC, respectively. MCC is lowest at the Columbus urban center, in the large airport census tract east of the core, and out around the southern ring suburbs, which are dotted with industrial and agricultural land uses. The areas of high canopy coverage occur along a green corridor running NNW from the city center, and in the first and second ring suburbs in the NE corner of the county. MISC levels follow a roughly inverted geography, highest in the Columbus downtown and reducing quite smoothly toward the suburban and rural borders of the county.

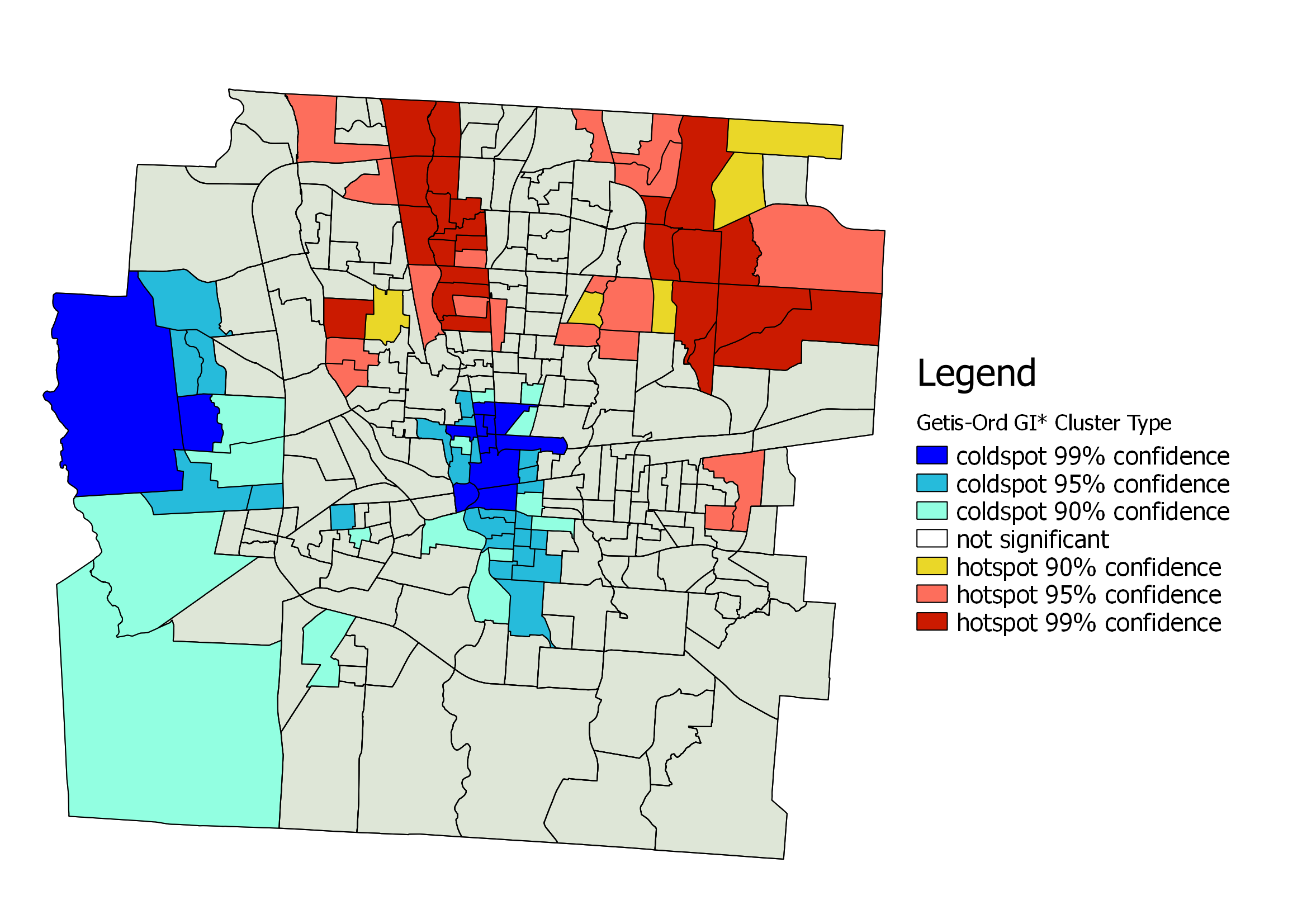


*Map 1: Choropleth map of Franklin County mean canopy coverage (MCC)*

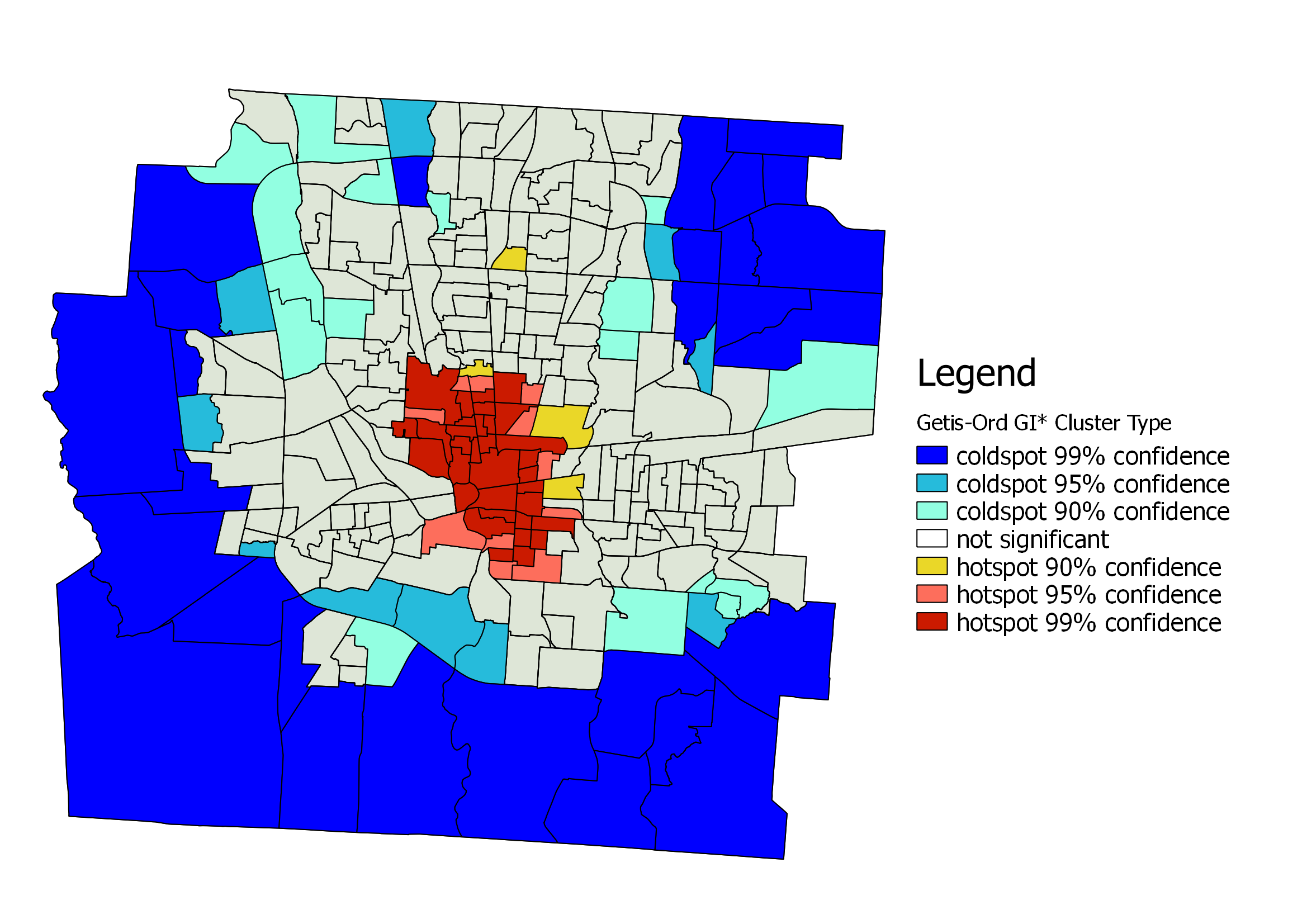


*Map 2: Choropleth map of Franklin County mean impervious surface coverage (MISC)*

Our application of Getis-Ord Gi\* Hotspot Analysis through QGIS revealed likely clusters of high and low values for our indicator variables. MCC was shown in Map 3 to have likely low clustering downtown and toward the agrarian west end of the county, with high value clustering in the NE suburbs and NNW from the first-ring suburbs out, which is likely due to a wealth of park space along the Scioto riverfront. MISC further demonstrated in Map 4 that ring-like gradient from downtown out to the rural suburbs. Both have noticeable non-clustered areas extending out from downtown to the north and east all the way to the county boundary.



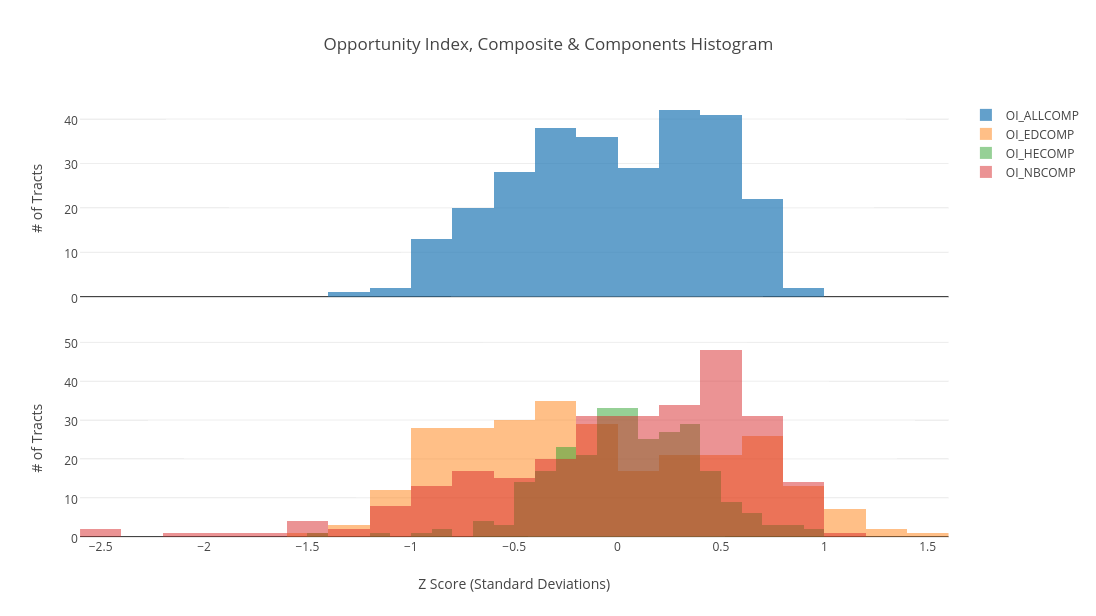
*Map 3: Getis-Ord Gi\* hot and cold spots of Franklin County mean canopy coverage (MCC)*



*Map 4: Getis-Ord Gi\* hot and cold spots of Franklin County mean impervious surface coverage (MISC)*

### Opportunity Index

As stated in Methods, the OI is a composite of three component indexes: Education, Job Access & Mobility, and Environmental Hazards (Kirwan, 2013). These component indicators are worth surveying individually for now, as they reveal variation between the different dimensions of opportunity. As seen in Figure X, the OI itself has a roughly bimodal distribution, with peaks located at +-0.5 of the mean. A similar pattern can be seen in OI-Ed below, with more of a plateau than a peak at <-0.5 z-score. The OI-NB has a skewed-right normal distribution, with a pronounced peak above the mean. OI-HE has the most compact and normal distribution of the three components.



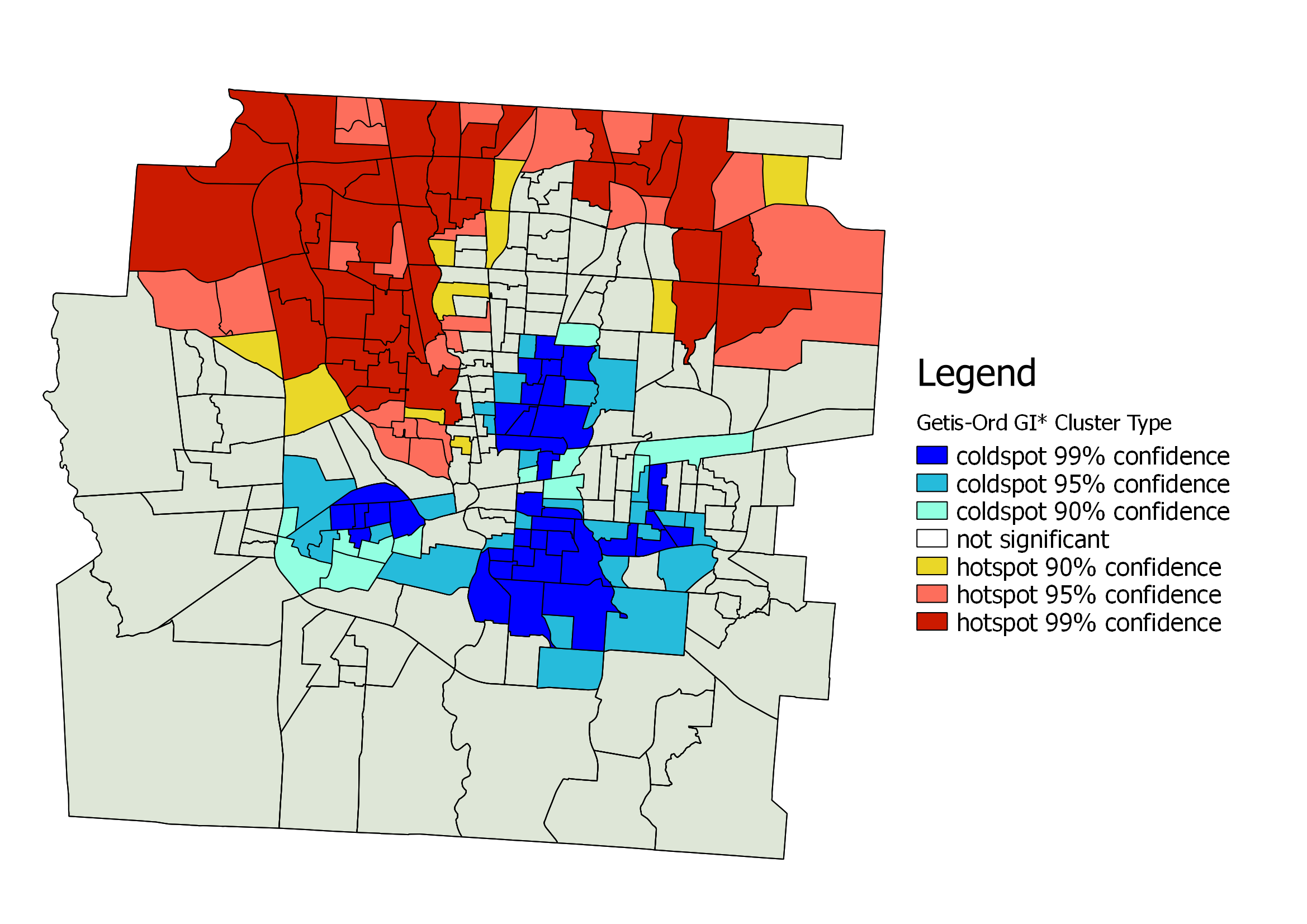
*Figure 2: Histogram of Opportunity Index and component indexes*

The county’s lowest OI scores occur in a swath close on the east side of the I-71 corridor, as well as the near south and southwest neighborhoods outside of Columbus’s downtown. It’s highest can be seen extending NW from downtown out to Dublin at the corner of the count, as well as a NE swath past the first ring suburbs toward Gahanna and New Albany. A notable high outlier is the Bexley, just to the east of downtown, which sits in the middle of the large sector of low-scoring neighborhoods.



*Map 5: Choropleth of Franklin County Opportunity Index (OI) scores*

Geits-Ord Gi\* Hotspot Analysis of OI, shown in Map 6 below, shows a large outer ring of socioeconomic vitality to the north of the Columbus center, with a complimentarily-shaped area of low OI centered just east of the downtown core. This clear divide of clustering shows the results of the suburban exodus that created the wealthy towns of Dublin, Powell, Gahanna, and most recently New Albany to the north of the city.

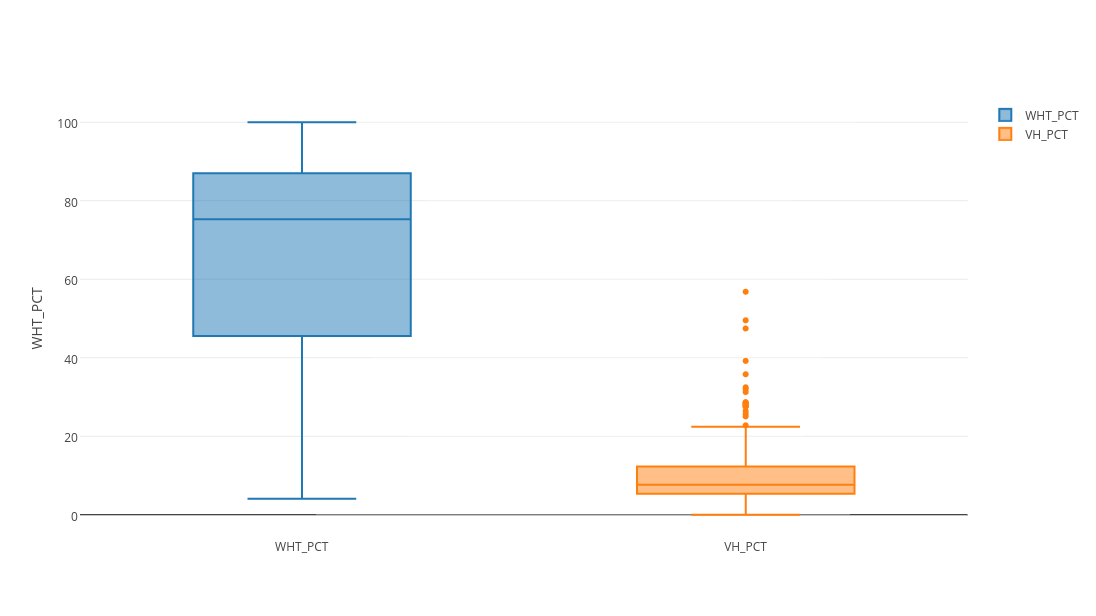


*Map 6: Getis-Ord Gi\* hot and cold spots of Franklin County Opportunity Index (OI)*

Other Socioeconomic Variables

White population percentage (WP) and vacant housing percentage (VH) are the two other socioeconomic indicators we investigated in our spatial analysis. Columbus generally has a high white population as seen in the box plot below, with a median of 75.28% per tract. However, the variation in per tract percentage is nearly 100%, with a minimum WP of 4.09%. This wide of a variation already suggests a segregated racial landscape before mapping values.

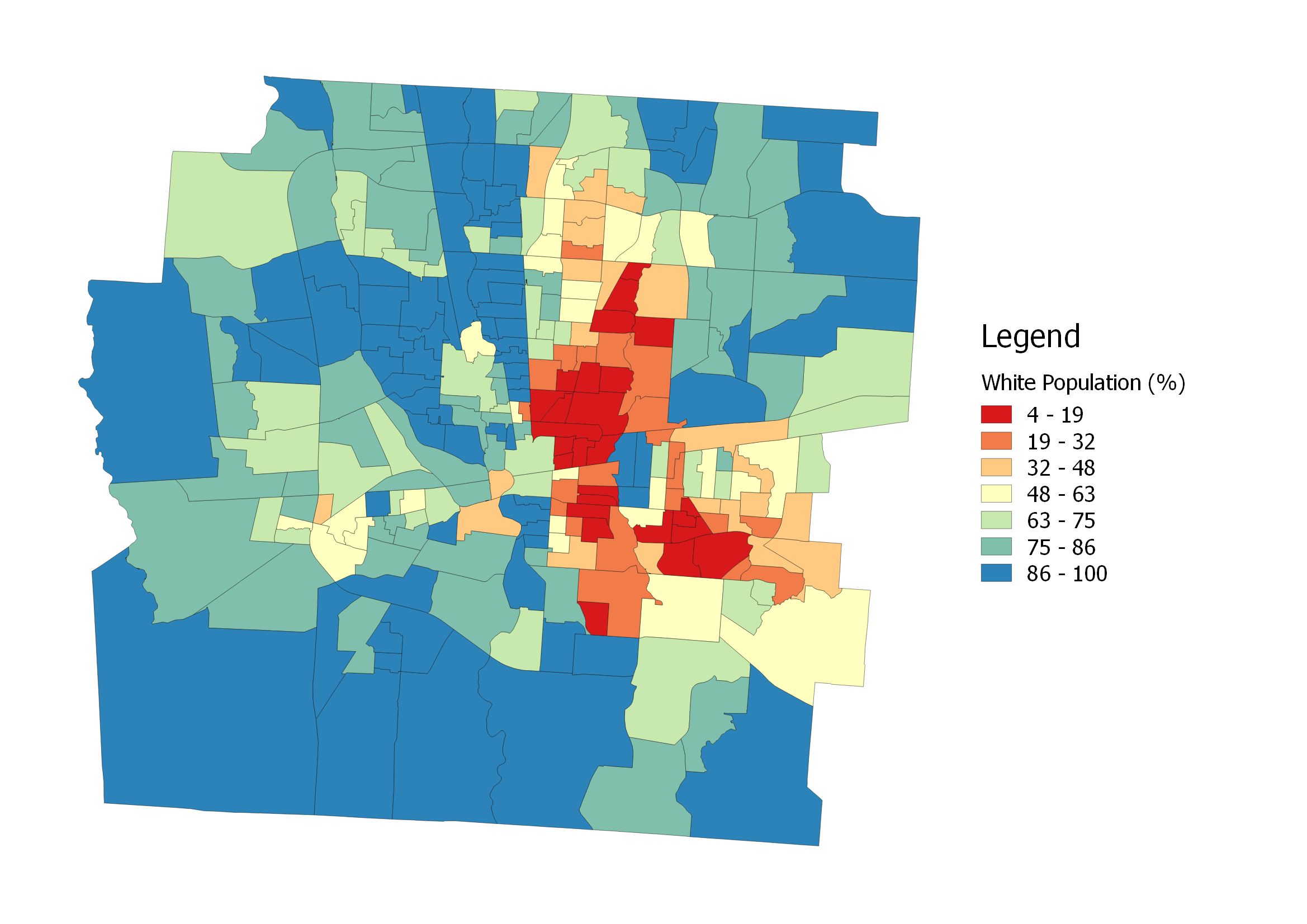
Housing vacancy has a much tighter spread of values, with a variation of 8%, less than a third of WP (26%). It’s median is also low at 7.6%. There are a notable number of high outliers, however, with the maximum being 56.8%.



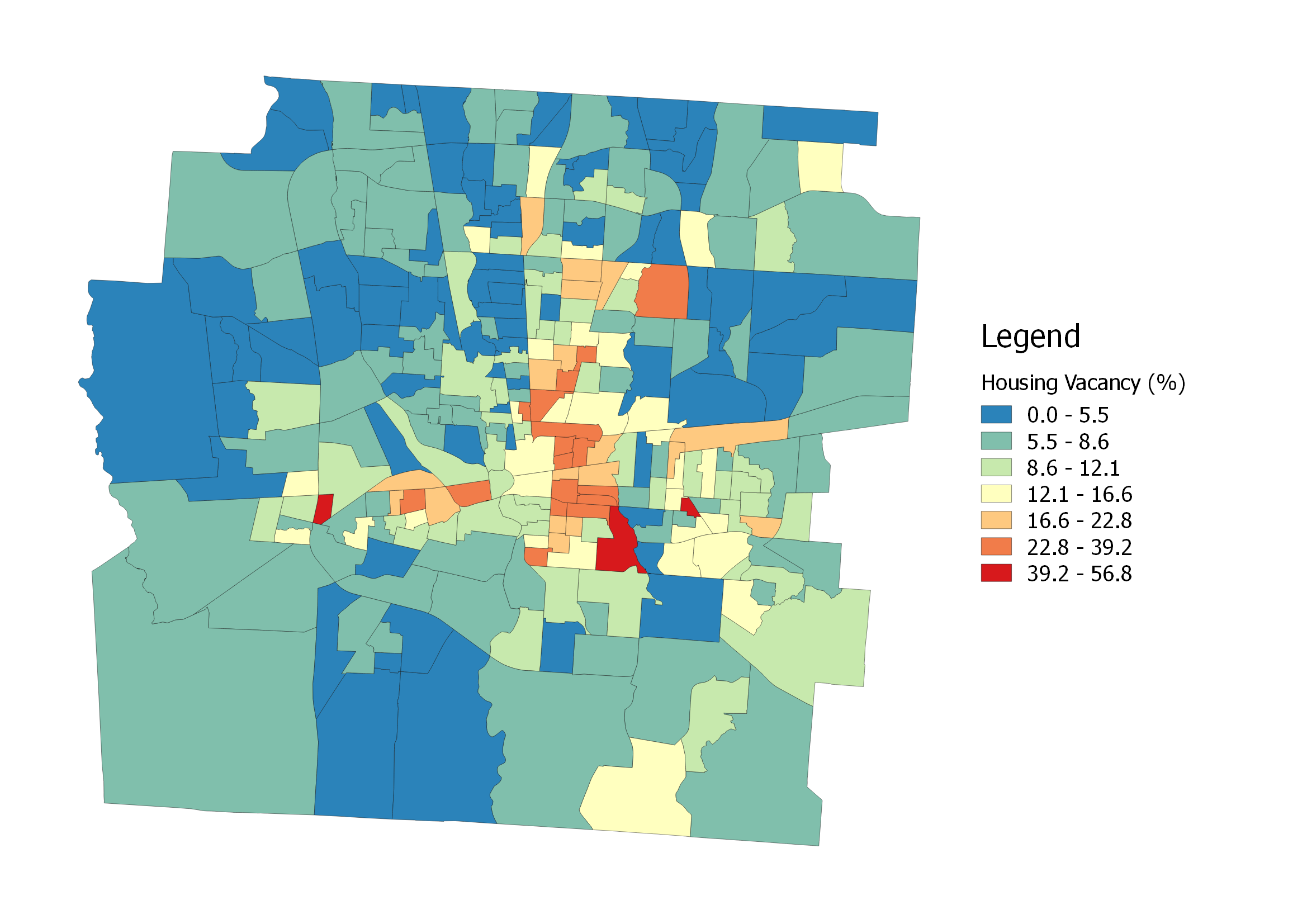
*Figure 3: Boxplot of Franklin County census tract white population (WP) and vacant housing (VH)*

Mapping WP reveals clear lines of division in racial makeup. WP follows a generally increasing trend going out toward the rural boundaries of the county, with pockets of high outliers within the urban center, Again, Bexley can be seen here as an outlier of high WP among a swath of low values just to the east of downtown.

Housing vacancy follows a roughly similar though inverted trend of higher frequency downtown and decreasing out toward the rural areas. Interestingly, the high values downtown seem to extend out into the first ring suburbs in a cross formation in the N-S and W-E directions. This may be due to the extensive land taken up by the I-71 and I-70 highways.



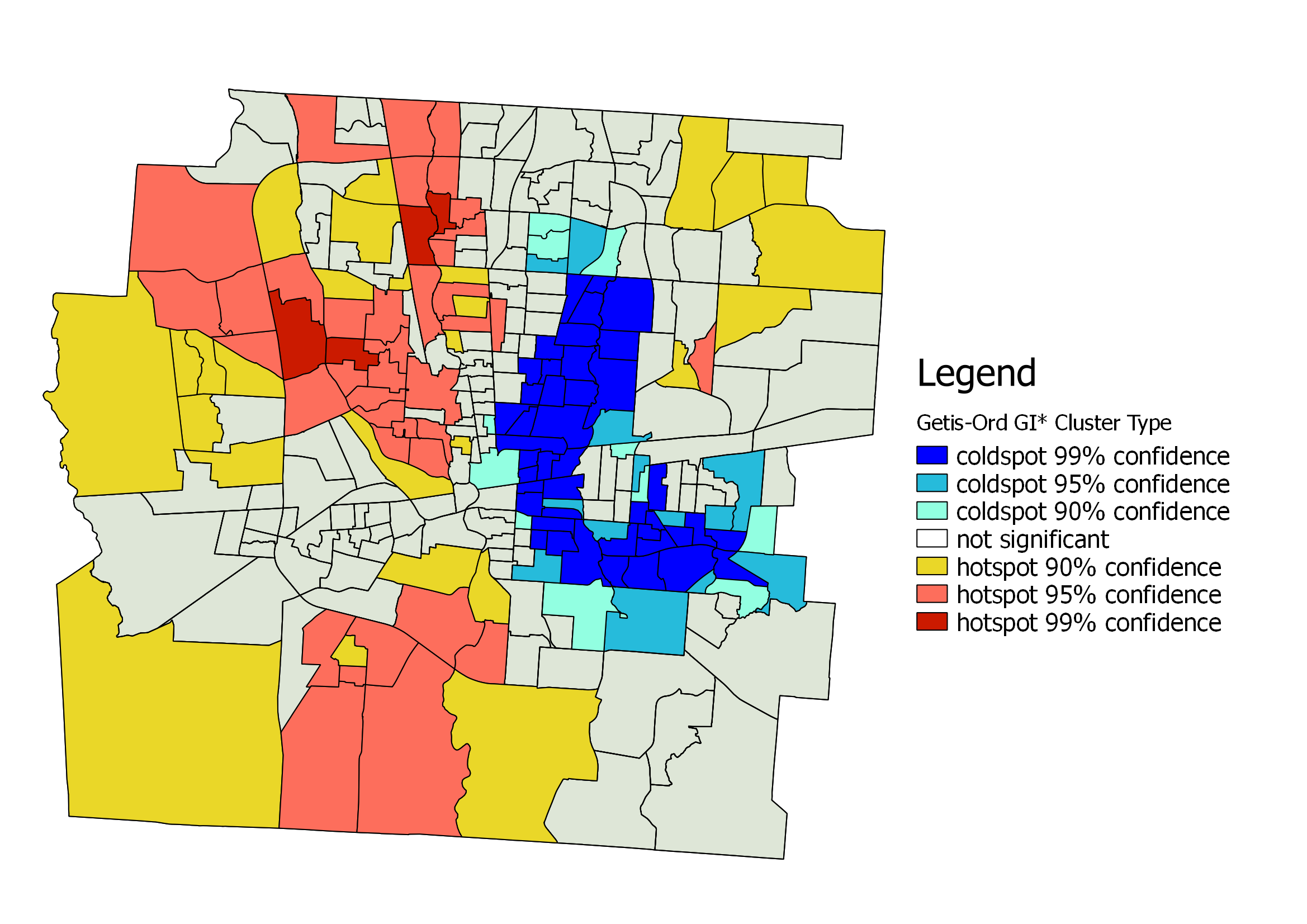
*Map 7: Choropleth of Franklin County white population (WP)*



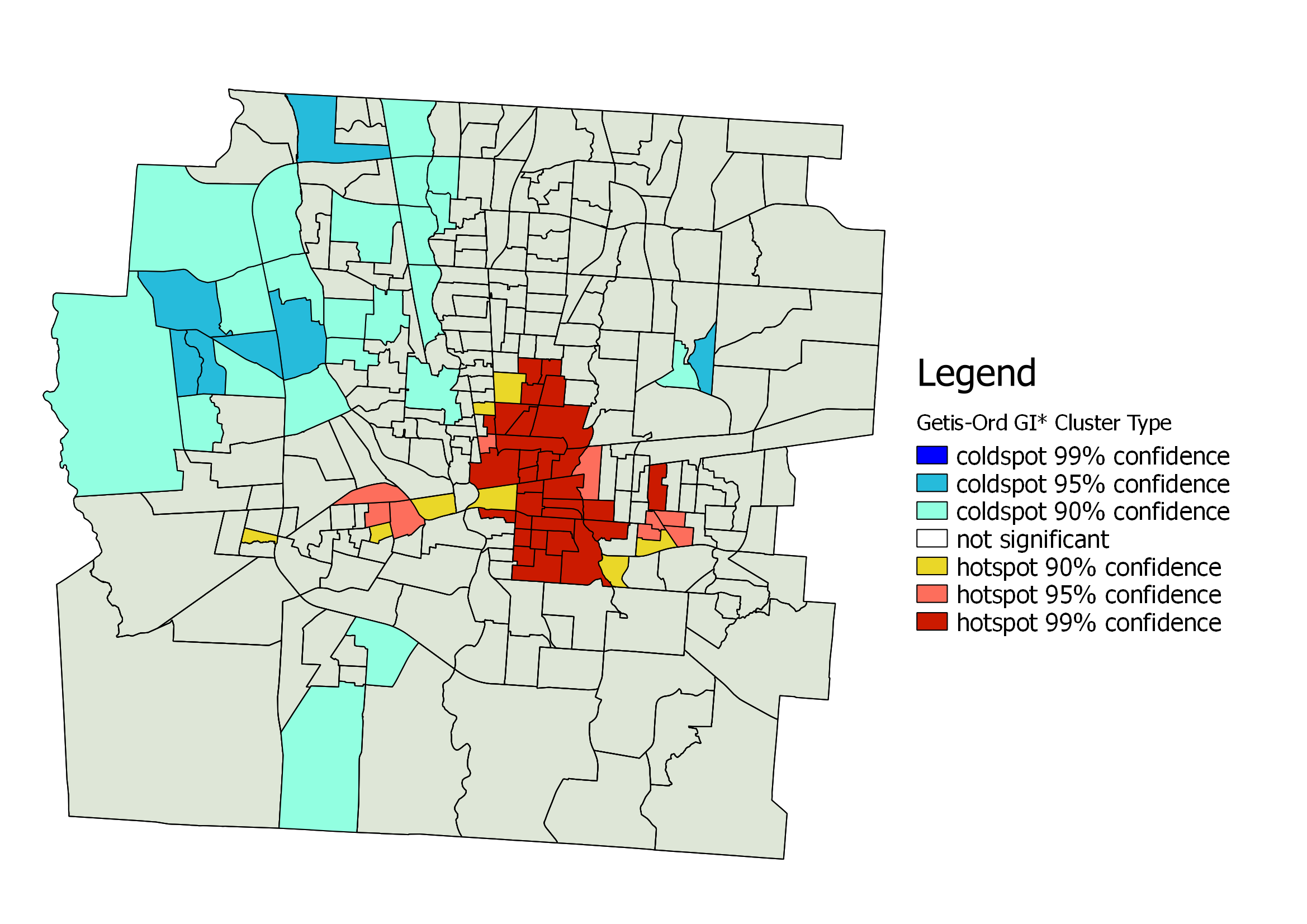
*Map 8: Choropleth of Franklin County vacant housing (VH)*

The Getis-Ord Gi\* Hotspot Analysis of WP shown in Map 8 reveals a varied landscape of high WP areas and a tightly-packed area of low WP. This is indicated by the prevalence of yellow-colored or uncertain high values, which suggest the the boundaries of these high-WP areas along the outer and western areas of the county are inconclusive. Meanwhile, areas of high minority population have one large swath along the eastern part of Columbus’s first ring suburbs, with much more clearly-defined borders. The gray area at the center of that swath of blue is Bexley, the location of our case study.

Getis-Ord Gi\* Hotspot Analysis of VH rates across the county are shown in Map 9. The NW exurban areas show a lightly-clustered area of low values, which high vacancy rates are tightly clustered just to the east of the Columbus downtown core.



*Map 8: Getis-Ord Gi\* hot and cold spots of Franklin County white population (WP)*



*Map 9: Getis-Ord Gi\* hot and cold spots of Franklin County vacant housing (VH)*

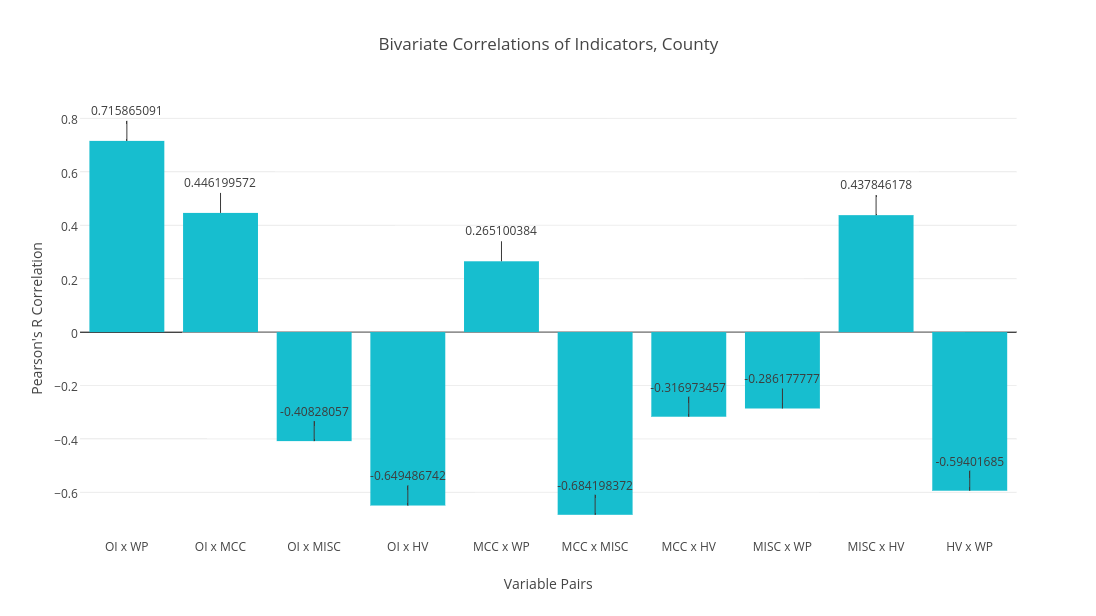
## Correlation & Temporal Analysis

### County-wide Correlations

The bar graph below shows Pearson’s *r* calculations for all possible pairings of our five indicator variables. The values represent the relationship between the two indicators’ variations over the county, with stronger relationships being further from 0. A very negative number indicates a strong inverse relationship, such as between the environmental asset canopy coverage (MCC) and the environmental stressor impervious surface coverage (MISC).

The striking relationship between WP and OI that serves as a foundation of this research is seen here, were *r* = 0.716. The powerful inverse relationships between the other socioeconomic pairings can be seen as well, where OI x HV = -0.649 and HV x WP = -0.59. In fact, all of the asset-liability indicator pairings below show an inverse relationship.

Our main target pairings for this research are those of MCC and, to a lesser extent, MISC. OI x MCC showed a moderately strong relationship at *r* = 0.446, and it was the second strongest positively-correlated pairing behind OI x WP. What was surprising for us was the much weaker relationship of white population (WP) to our environmental indicators: MCC x WP = 0.265 and MISC x WP = -0.286. We believe this suggests that race plays a secondary or further-removed role in the demonstrated inequity of access to environmental assets in Columbus, but that there is a substantial relationship between that access disparity and other socioeconomic forces of inequality.

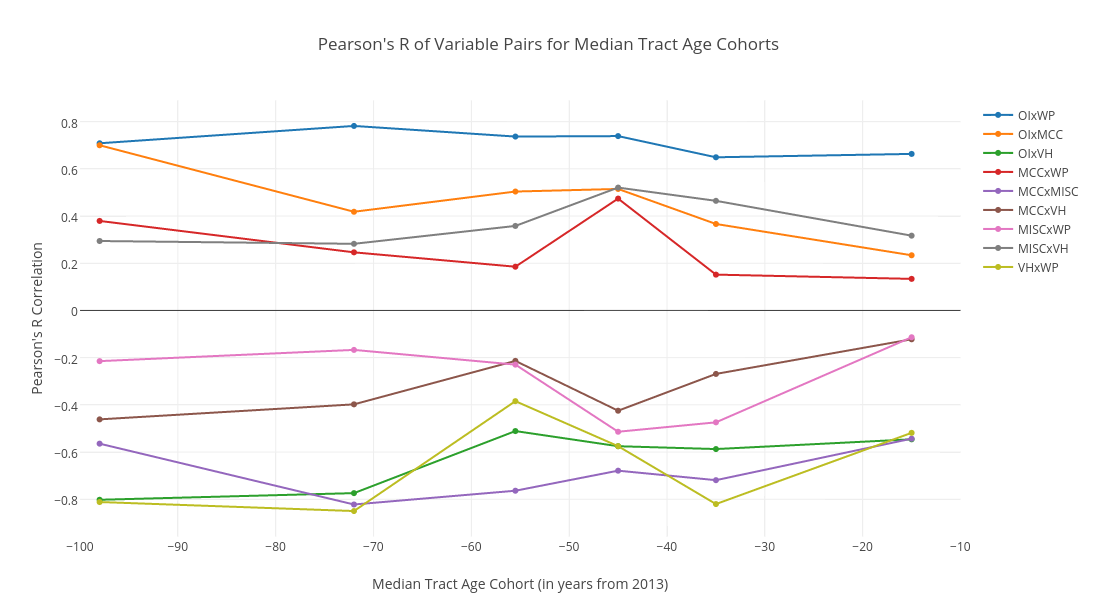


*Figure 4: Pearson’s r correlation of indicator pair values in Franklin County*

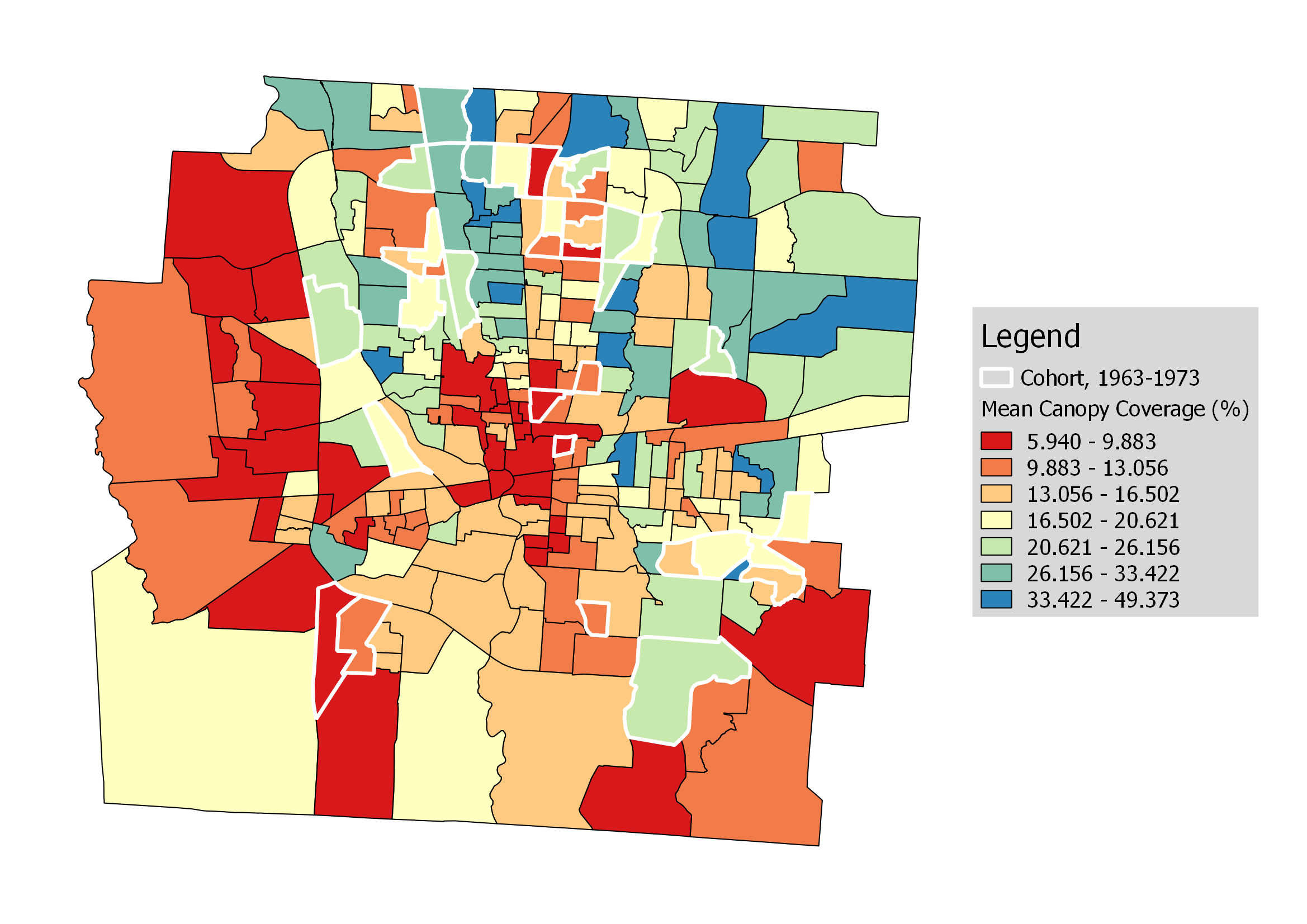
### Age Cohort Correlations

We aggregated tract-level data from the Franklin County Auditor’s public data set to calculate the *median build date* (MBD) of each census tract. We then grouped these 153 tracts into six *age cohorts* by their MBD. This data set was collected in 2013, so the cohorts represent a median age in years before that year, which can be seen in the x-axis of Figure 5 below.

The resulting graph shows a compelling, though heavily approximated, possible view of historical trends between the “relationship proximity” between socioeconomic and environmental indicators. Two trends were of enough interest to us to include this graph. First, there is a general downward trend in correlation across all variables over the last 110 years. This could either mean that socioeconomic and environmental policy is less and less closely related, or that impacts of the relationship between the two takes a very long time to manifest in the built environment. The second area of interest was in the bubble at -45 years, representing the age cohort from 1963-1973, where all indicator pairs experienced a sharp bump in significance except MCC x MISC. We believe this represents the close tie between economic vitality, racial segregation, and greenery that are the foundations of the suburban era, which occurred during this part of the 20th Century. Map 10 shows the census tracts composing this cohort highlighted on the MCC choropleth.



*Figure 5: Line graph of Pearson’s r correlation of indicator pair values across median build date (MBD) age cohorts in Franklin County*

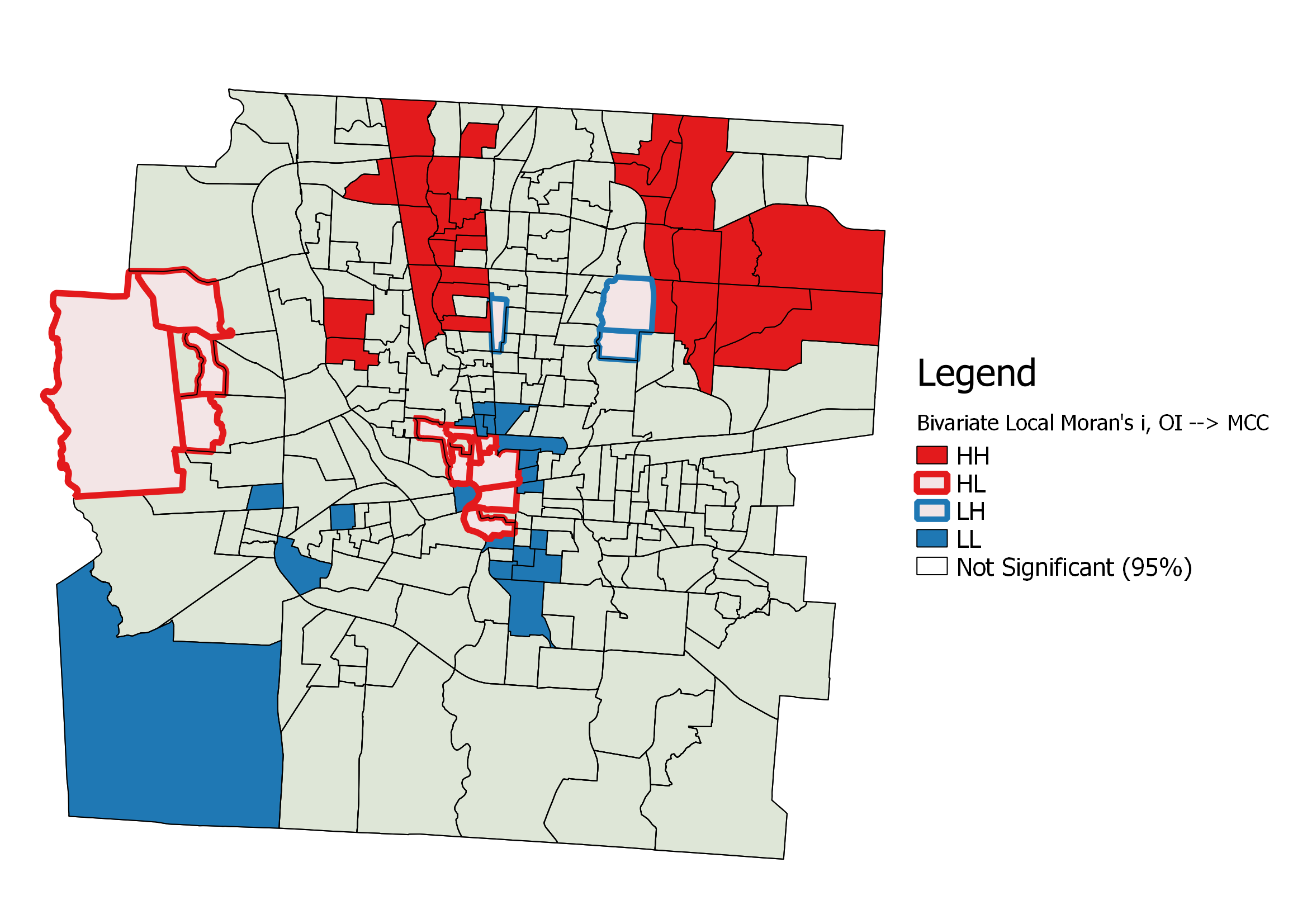


*Map 10: 1963-1973 age cohort highlighted on MCC choropleth*

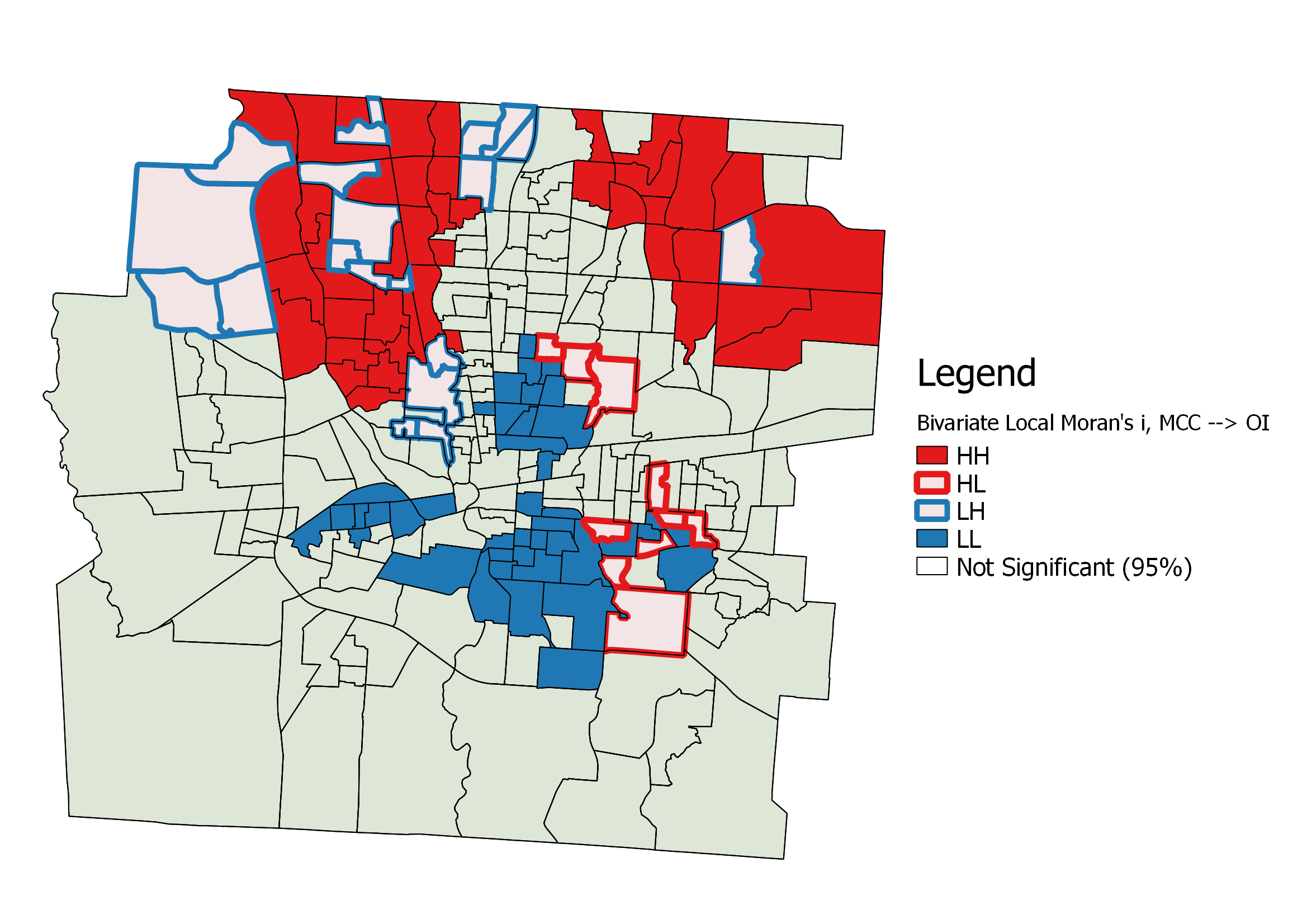
### Bivariate Mapping

Bivariate mapping proved to be a confounding problem in our research. The issue can be broken into two facets: first, quantitative measures are scarce and difficult to glean meaningful relationships from, and second, it is difficult to create clear visual coding of bivariate spatial trends.

An example of the first issue can be seen in Maps 11 and 12 below, where we applied the only available bivariate cluster analysis tool we could find to OI x MCC, called Bivariate Local Moran’s I. This type of cluster analysis shows the relationship between an “ego variable” and a “neighbor variable,” where a High-Low Outlier reflects a high ego variable value surrounded by low neighbor variable values (BioMedware, 2014). This means that there is a *directionality* of relationship between the two variables, which can be seen in Map 12, where MCC is now the ego variable and OI the neighbor variable. This makes interpreting such results even more complicated, and has made it clear to us that new bivariate spatial analysis methods must be developed for future interdisciplinary spatial assessment research.

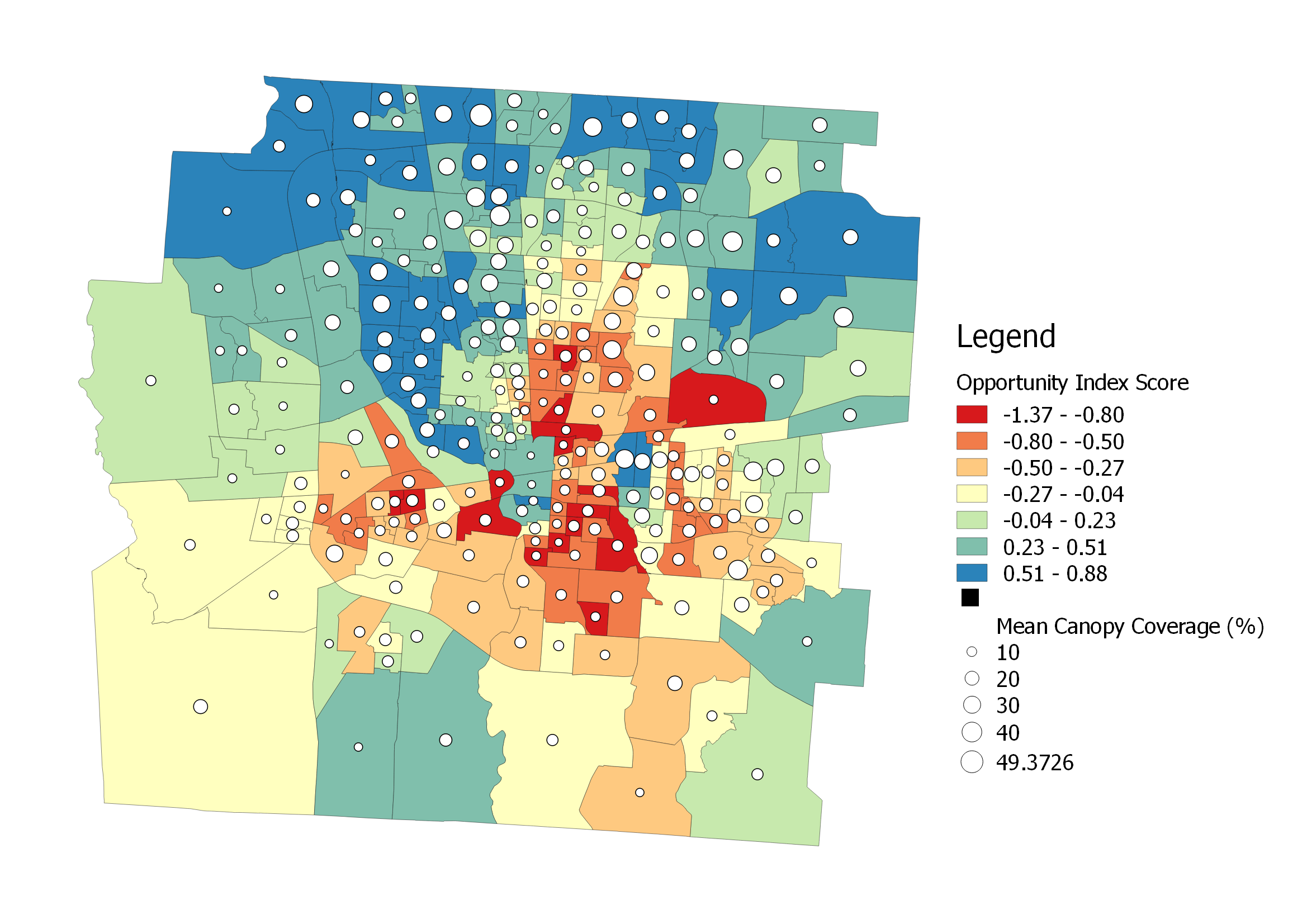


*Map 11: Bivariate local Moran’s I for OI → MCC*

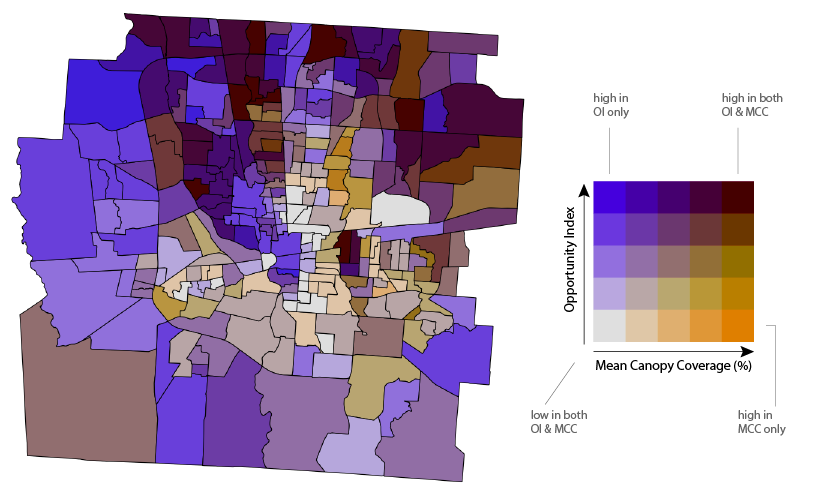


*Map 12: Bivariate local Moran’s I for MCC → OI*

The second issue can be seen in Maps 13 and 14 below. Visual coding of scales is very difficult in bivariate maps. Two contemporary methods are shape overlays (Map 13) and bivariate choropleth maps (Map 14). We present these for reader review, as well as other variable pairings in our Appendix, however we have reservations that these methods of visual communication are too complex and too easily manipulated to be trusted for in-depth reading. With maps as varied and perceptual as these, we run into a Rorschach Test problem of reading our biases toward perceived clustering into these maps.



*Map 13: Bivariate choropleth-shape map showing MCC as circle overlays on census tract OI*



*Map 14: Bivariate choropleth of OI x MCC using a dual color gradient*