Spatial Inequalities in Street-Tree Ecological Quality across New York City

Final Report
Urban Analytics | Track B (Advanced)
More Than Green
Di Li, Jialin Sun, Chunman Li, Xiaojing Jin
2026-05-07

1. Problem and Motivation

Urban street trees deliver a well-documented portfolio of ecosystem services: they mitigate the urban heat island effect, attenuate stormwater runoff, improve air quality, and contribute to residents' psychological well-being (Salmond et al., 2016; Berland et al., 2017). Despite growing consensus on their aggregate value, the spatial distribution of street-tree ecological quality is far from uniform. Evidence from North American cities consistently shows that lower-income, higher-density, and historically marginalized neighborhoods bear a disproportionate burden of degraded urban tree canopy (Wolf et al., 2020; Schwarz et al., 2015). New York City, with its extreme socioeconomic heterogeneity and dense built fabric, offers a compelling laboratory to examine these inequalities at fine spatial resolution.

This study pursues three interrelated research questions:

- RQ1: What spatial inequalities and clustering patterns characterize street-tree health, structural attributes, and species diversity across NYC?
- RQ2: How do street-tree ecological quality indicators associate with socioeconomic vulnerability, built-environment intensity (including 2D urban morphology and landscape configuration), and green infrastructure supply?
- RQ3: Where do multiple disadvantages overlap (low tree quality + high social vulnerability + high built density or fragmented greenery), and can we identify priority intervention areas?

Several key analytical constructs require explicit definition. The Tree Health Index is the mean of per-tree health scores (Good=3, Fair=2, Poor=1) aggregated to the Census Tract, providing a continuous proxy for overall structural condition. Shannon diversity (H') quantifies species richness weighted by evenness across the tract-level tree population. PLAND (proportion of landscape) expresses the percentage of a given land-cover class within each tract boundary, derived from the UrbanWatch 1m product. Landscape configuration is captured through a suite of FRAGSTATS metrics: Number of Patches (NP), Patch Density (PD), Largest Patch Index (LPI), Edge Density (ED), Aggregation Index (AI), Shannon Diversity Index (SHDI), and Contagion (CONTAG). Finally, co-burden mapping overlays multiple disadvantage layers to identify tracts simultaneously experiencing low tree quality, high social vulnerability, and high built density or fragmented greenery (Galle et al., 2021).

The primary stakeholders of this research include the NYC Department of Parks and Recreation (which administers the street-tree program), Community Boards seeking evidence for capital allocation, urban forestry practitioners managing planting and maintenance programs, and environmental justice advocates who require spatially explicit co-burden evidence to support policy interventions.

The Census Tract is adopted as the unit of analysis (n = 2,327 tracts covering all five boroughs). This choice is justified on three grounds: (1) it permits direct integration with American Community Survey socioeconomic data; (2) it constitutes the standard spatial unit for environmental justice analyses in U.S. urban research; and (3) it offers sufficient spatial resolution to capture meaningful intra-city variation without the statistical instability associated with smaller block-group geometries.

Variable Domains and Indicators

Domain	**Indicators**
Tree Ecological Quality	Health Index, mean DBH, tree density (trees/km2), Shannon diversity (H'), species count
Built Environment (3D)	Mean/max building height, building coverage ratio, building density
Built Environment (2D morphology)	9-class PLAND from UrbanWatch 1m (building, road, parking, tree canopy, grass/shrub, water, agriculture, barren, other); impervious/pervious ratio
Landscape Configuration	SHDI, CONTAG (landscape-level); NP, PD, LPI, ED, AI for tree canopy and building classes
Green Infrastructure	Park area ratio, tree canopy PLAND
Socioeconomic	Median income, poverty rate, minority %, renter %, SVI composite (RPL_THEMES)

Table 1. Summary of analytical domains and indicators used in this study.

This framework draws on previous empirical work linking tree canopy cover to built-environment indicators (Berland et al., 2017), socioeconomic gradients (Wolf et al., 2020), and landscape configuration (Schwarz et al., 2015), while extending those foundations through the UrbanWatch high-resolution land-cover product (Zhang et al., 2022) and co-burden spatial methods (Galle et al., 2021).

2. Data Sources and Cleaning

Data Sources and Cleaning

Seven distinct datasets were assembled and harmonized to construct the tract-level analytical table. The following subsections describe each source's coverage, granularity, format, and provenance, followed by a record of cleaning and transformation decisions.

2.1 Trees Count! 2015 Street Tree Census

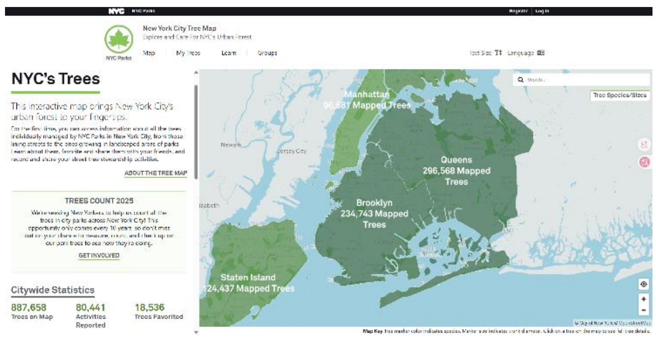
Source: NYC Open Data | Format: CSV with lat/lon | Coverage: City-wide, 2015-2016 |

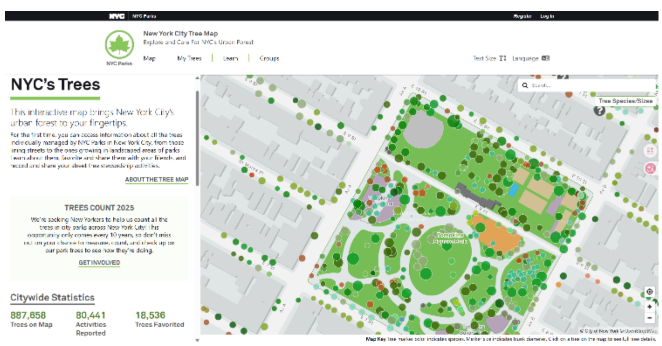
Raw records: 683,788

The NYC Parks Department conducted its third decennial street tree census between 2015 and 2016, recording detailed information for each mapped street tree across the five boroughs. The dataset includes species, diameter at breast height (DBH), health condition, and geographic coordinates, totaling 683,788 records in CSV format.

Data cleaning procedures first removed all records with a status other than "Alive," resulting in 652,173 valid tree observations. Coordinate fields were checked to ensure completeness, with no missing or zero values identified. Implausible DBH values exceeding 100 inches were capped at 100 to mitigate the impact of measurement errors. Tree health conditions were numerically encoded (Good = 3, Fair = 2, Poor = 1) to enable quantitative analysis.

Tree locations were then spatially joined to 2020 Census Tract boundaries (EPSG:4326), successfully matching 652,045 trees across 2,306 tracts. Aggregated tract-level indicators were computed, including total tree count, tree density (trees per square kilometer), mean DBH, a composite Tree Health Index, Shannon diversity index (H'), and species richness.





2.2 UrbanWatch 1m Land Cover

Source: University of Vermont / UNC Charlotte | Format: GeoTIFF (42,000 x 46,674 px) |

CRS: EPSG:26918 | Classes: 9

The UrbanWatch dataset provides a high-resolution (1-meter) land cover classification for the greater NYC area, derived from 2017 aerial imagery and validated at over 90% overall accuracy. The dataset is distributed as a three-band RGB GeoTIFF, where each RGB combination corresponds to one of nine land cover classes.

To enable analysis, the RGB raster was decoded into a single-band classified raster using the provided class-color lookup table, resulting in approximately 749 million valid pixels. Census Tract boundaries were reprojected to UTM Zone 18N (EPSG:26918) to align with the raster data. Zonal statistics were then computed to extract pixel counts per land cover class within each tract.

From these results, PLAND metrics were derived, representing the percentage of each land cover class relative to total tract area, yielding ten indicators per tract. In addition, landscape configuration metrics—including Number of Patches (NP), Patch Density (PD), Largest Patch Index (LPI), Edge Density (ED), Aggregation Index (AI), Shannon Diversity Index (SHDI), and Contagion Index (CONTAG)—were calculated using a FRAGSTATS-equivalent Python workflow at the tract level.


FragStats

Home About us Blog Documentation Tutorial Discussion forum Downloads License Terms of service My page

You are here: Home / Blog / Fragstats 5 and Python

Fragstats 5 and Python

How to use the command line version of Fragstats 5 from Python

For those you you who did not notice the command line version of Fragstats 5, present in the installation but mostly overlooked, this article will try to fill the gap.

Using the command line is very easy and at the same time powerful as it can be invoked from not just the command prompt directly but also from all sorts of scripts. The focus of today's blog post is the simple integration of the command line Fragstats 5 into Python.

Let's start with the basics. To use frg_cmd.exe, we need three things:

- A model to specify which metrics you want to compute. You can create this with the Fragstats 5 GUI application.
- A batch file that specifies all the grids you want to process. You can create one with a text editor or generate it right within your Python script.
- A path to store the result tables if you are doing a standard analysis. The folder must exist before invoking the command.

The command line would look like this:

```
frg_cmd /m "C:\fragstats_work\model\my_model.fca" /b "C:\fragstats_work\model\my_batch.fbt" /o "C:\fragstats_work\out\results"
```

Please note that when specifying the output base path (ex. C:\landscapes\fragstats\out\results), "results" is the common prefix of the output tables and the folder that will hold these tables is "C:\landscapes\fragstats\out\". The folder must exist prior to running frg_cmd otherwise it will not be created.
The output tables will be named: results.patch, results.class, results.land, and results.adj. If a certain analysis level is not selected in the model, the corresponding table will not be created. The same will happen with the adjacency table.

The command line version of Fragstats 5. has become a first class object with its own entry in the start menu from version 5.1. If you have not used it before, you should give it a try. It only needs to be given a model and a user defined batch file and it will do it's job quietly in

P.S. FRAGSTATS: FRAGSTATS is a landscape analysis tool used to quantify spatial patterns in categorical raster data. It identifies contiguous pixels of the same class as patches and computes metrics describing fragmentation, connectivity, edge complexity, and spatial heterogeneity. These metrics are widely applied in landscape ecology and urban studies to understand spatial structure.

In this study, FRAGSTATS-based metrics complement Python-based geospatial analysis. While Python is effective for calculating land-cover composition (e.g., PLAND), it does not provide a standardized framework for analyzing spatial configuration. FRAGSTATS offers a well-established set of patch-based metrics that ensure methodological consistency and comparability.

2.3 NYC Building Footprints

Source: NYC DoITT / NYC Open Data | Format: GeoPackage | Records: 1,082,289 |

CRS: EPSG:4326

The NYC Building Footprints dataset, provided by NYC DoITT, contains polygon geometries for all mapped buildings across the city, along with an attribute for roof height in feet. The dataset includes 1,082,289 records in GeoPackage format.

For processing, building geometries were reprojected to UTM Zone 18N to ensure accurate area calculations. The height_roof attribute was converted from feet to meters using a conversion factor of 0.3048. Representative centroid points were generated for each building and spatially joined to Census Tracts, resulting in a successful match rate of 99.98% (1,082,092 buildings).

Tract-level indicators derived from this dataset include total building count, mean, median, and maximum building height (in meters), building footprint coverage ratio (total footprint area divided by tract area), and building density (buildings per square kilometer). A small number of tracts corresponding to water bodies or major park interiors contained no buildings and were assigned missing values.

2.4 PLUTO v25.4

Source: NYC Department of City Planning | Format: CSV | Records: 858,644 tax lots

MapPLUTO v25.4, published by the NYC Department of City Planning, provides parcel-level land use and building attributes for 858,644 tax lots in CSV format. Key variables include floor area ratio (FAR), number of floors, year built, and land use classification.

To integrate this dataset, a consistent 11-digit Census Tract GEOID was constructed by concatenating the borough code (borocode) and tract identifier (bct2020). Parcel-level attributes were then aggregated to the tract level. Derived indicators include mean FAR, mean number of floors, median year built, and the modal land use class within each tract.

2.5 NYC Parks Properties

Source: NYC Parks / NYC Open Data | Format: CSV with WKT geometry | Records: 2,058

The NYC Parks Properties dataset enumerates all NYC Parks Department-managed properties, represented as multipolygon geometries stored in WKT format. The dataset includes 2,058 records.

WKT geometries were parsed using the Shapely library and intersected with Census Tract boundaries to compute the proportion of each tract occupied by parkland. This resulted in a park area ratio indicator for each tract. Of the 2,327 tracts, 1,354 contained at least partial park coverage, while the remaining 973 tracts were assigned a park_area_ratio of zero. These zero values were treated as structurally missing in terms of park presence rather than absent data.

2.6 American Community Survey (ACS) 5-Year Estimates (2019)

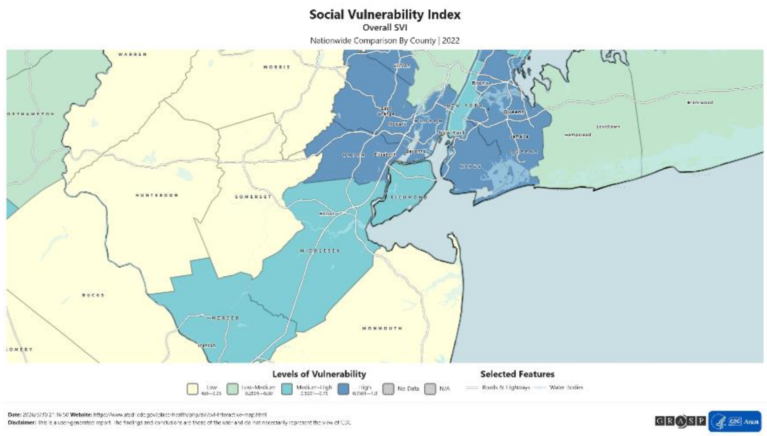
Source: U.S. Census Bureau API | Coverage: 2,167 NYC tracts | Tables: B19013, B17001, B02001, B25003

Socioeconomic variables were obtained from the U.S. Census Bureau's American Community Survey (ACS) 2019 5-year estimates via the Census API, covering 2,167 NYC Census Tracts. Four tables were used: B19013 (median household income), B17001 (poverty status), B02001 (race), and B25003 (housing tenure).

From these tables, key indicators were computed, including poverty rate (percentage of population below the poverty line), minority percentage (non-white population share), and renter percentage (share of renter-occupied housing units). Sentinel values (-666666666) used by the Census Bureau to denote missing or suppressed data were recoded as NaN. Approximately 160 tracts (7%) had suppressed median income estimates, primarily corresponding to small or institutionalized populations.

2.7 CDC/ATSDR Social Vulnerability Index (SVI)

Source: CDC / ATSDR | Format: CSV | Coverage: All U.S. Census Tracts



The CDC/ATSDR Social Vulnerability Index (SVI) dataset provides a composite measure of social vulnerability at the Census Tract level across the United States. The primary indicator, RPL_THEMES, ranges from 0 to 1 and summarizes vulnerability across four dimensions: socioeconomic status, household composition, minority status and language, and housing type and transportation.

SVI data were matched to the analytical dataset using the 11-digit Census Tract GEOID. Sentinel values (-999) were recoded as NaN, affecting approximately 347 tracts (14.9%), which largely correspond to unpopulated areas such as water bodies, parks, or institutional zones identified by the CDC.

2.8 Final Data Integration

All seven datasets were integrated at the Census Tract level using the GEOID as a common key. The final analytical dataset consists of 2,327 Census Tracts and 73 variables, capturing a comprehensive range of environmental, spatial, and socioeconomic characteristics.

To ensure methodological consistency, all spatial operations—including reprojection, area calculation, and spatial joins—were conducted using a unified projected coordinate system, EPSG:26918 (UTM Zone 18N, NAD83). This standardization is critical for maintaining accuracy in distance- and area-based metrics across datasets with differing original coordinate systems.

3. Missing Data Analysis

3.1 Missing Data Identification

To evaluate data integrity, we systematically examined the distribution of missing values across the final tract-level dataset. Missingness was analyzed not only in terms of magnitude but also in terms of its structural and spatial context. We distinguished between structural absence and data gaps. This distinction is critical for applying the Dark Data framework, as it separates meaningful "known missing" values that reflect real-world conditions from missingness arising from data collection or reporting limitations.

3.2 Missing Data Identification

Missing data in the analytical table are structurally informative rather than random. Following the dark data taxonomy (Hand, 2020), each variable group is classified by missingness mechanism and handled accordingly.

Variable Group	Missing N (%)	Dark Data Type	Handling Strategy
Tree indicators (health, diversity)	21 tracts (0.9%)	DD-Type 2 (Known missing)	tree_count = 0 assigned; diversity = NaN for tracts with < 5 trees
Building indicators	5 tracts (0.2%)	DD-Type 2 (Known missing)	Retained as NaN (water / major park tracts)
Park area ratio	973 tracts (41.8%)	DD-Type 2 (Known missing)	Filled with 0.0 (no parkland present)
Median income	~160 tracts (~6.9%)	DD-Type 3 (Choosing not to record)	Retained as NaN; listwise deletion applied in regression models
SVI RPL_THEMES	~347 tracts (~14.9%)	DD-Type 5 (Measurement issues)	Recoded -999 to NaN; excluded from SVI-dependent analyses
Landscape metrics	7 tracts (0.3%)	DD-Type 4 (Self-selection / edge)	Retained as NaN (tracts at raster boundary)

Table 2. Missing data summary by variable group, mechanism classification, and handling strategy. The missing data matrix is shown in Figure 6.

3.3 Spatial and Structural Patterns of Missingness

Missing data are not randomly distributed but instead exhibit clear structural and spatial patterns.

Tree-related missingness is concentrated in tracts with very low tree counts, reflecting insufficient observations rather than data collection failure. Additionally, some apparent absence of tree-related indicators results from differences in data definitions rather than true ecological absence. The TreesCount dataset records only street trees and excludes trees located within parks or other non-street environments. Consequently, areas such as Central Park may appear to have zero tree counts despite substantial tree canopy, indicating a limitation of data scope rather than actual absence.

Park area ratio shows a high proportion of missing values (41.8%); however, this reflects a structural condition, as most tracts do not contain NYC Parks-managed land. These cases are therefore recoded as zero to accurately represent the absence of parkland.

Socioeconomic variables, particularly median income, exhibit missingness due to data suppression in low-population tracts, which are often associated with institutional land uses or minimal residential presence. Similarly, missingness in the Social Vulnerability Index (SVI) is primarily driven by a sentinel value (-999), which systematically excludes non-residential or institutional tracts.

3.4 Implications for Bias and Analysis

The structured nature of missing data has important implications for analysis.

Because SVI and income data are missing primarily in non-residential or low-population tracts, excluding these observations from regression models is analytically appropriate and unlikely to introduce systematic bias into comparisons across residential neighborhoods. However, such exclusion may slightly underrepresent areas with atypical land use characteristics.

The recoding of park area ratio to zero ensures that the absence of parkland is correctly represented rather than misinterpreted as missing data. For tree-related indicators, assigning tree_count = 0 while retaining undefined values (*NaN*) for diversity preserves the distinction between tracts with no trees and those with measurable ecological variation.

To further ensure robustness, a sensitivity analysis will be conducted by excluding tracts with suppressed income data from all income-based regressions.

Overall, although missingness is substantial in certain variables, its structured nature allows for analytically appropriate handling with limited risk of introducing bias into the main findings.

4. Initial Results and Visualizations

4.1 Descriptive Summary

After cleaning and merging, the analytical dataset captures 652,045 alive street trees distributed across 2,306 of 2,327 Census Tracts. Summary statistics for key indicators are presented below.

Indicator	Mean	Std. Dev.	Min	Max
Tree Health Index	2.76	0.12	1.25	3.00
Tree Density (trees/km2)	1,103	712	0	4,890
Mean DBH (inches)	11.4	3.2	1.0	100.0
Shannon Diversity (H')	1.82	0.44	0.00	2.89
Species Count per Tract	28.3	14.1	1	82
Tree Canopy PLAND (%)	27.4	13.5	0.3	71.2
Impervious PLAND (%)	67.0	14.8	11.2	97.8
Mean Building Height (m)	11.6	9.3	0.0	130.4
Park Area Ratio (%)	7.3	14.2	0.0	98.4
Median Household Income ($)	69,420	38,110	9,800	250,001
Poverty Rate (%)	17.3	14.2	0.0	71.8
Minority %	56.4	30.1	1.2	99.8
SVI RPL_THEMES	0.51	0.28	0.00	1.00

Table 3. Descriptive statistics for key tract-level indicators (n = 2,306 tracts with tree data).

Health condition composition across all alive trees is: Good 81.1%, Fair 14.8%, Poor 4.1%. The top three street-tree species by count are London planetree (18.3%), honeylocust (9.9%), and Callery pear (9.0%), the latter raising long-term ecological concerns given its invasive potential. Native species constitute a small minority of the total inventory, concentrated in lower-ranked species by abundance.

4.2 Spatial Distribution of Tree Health Index

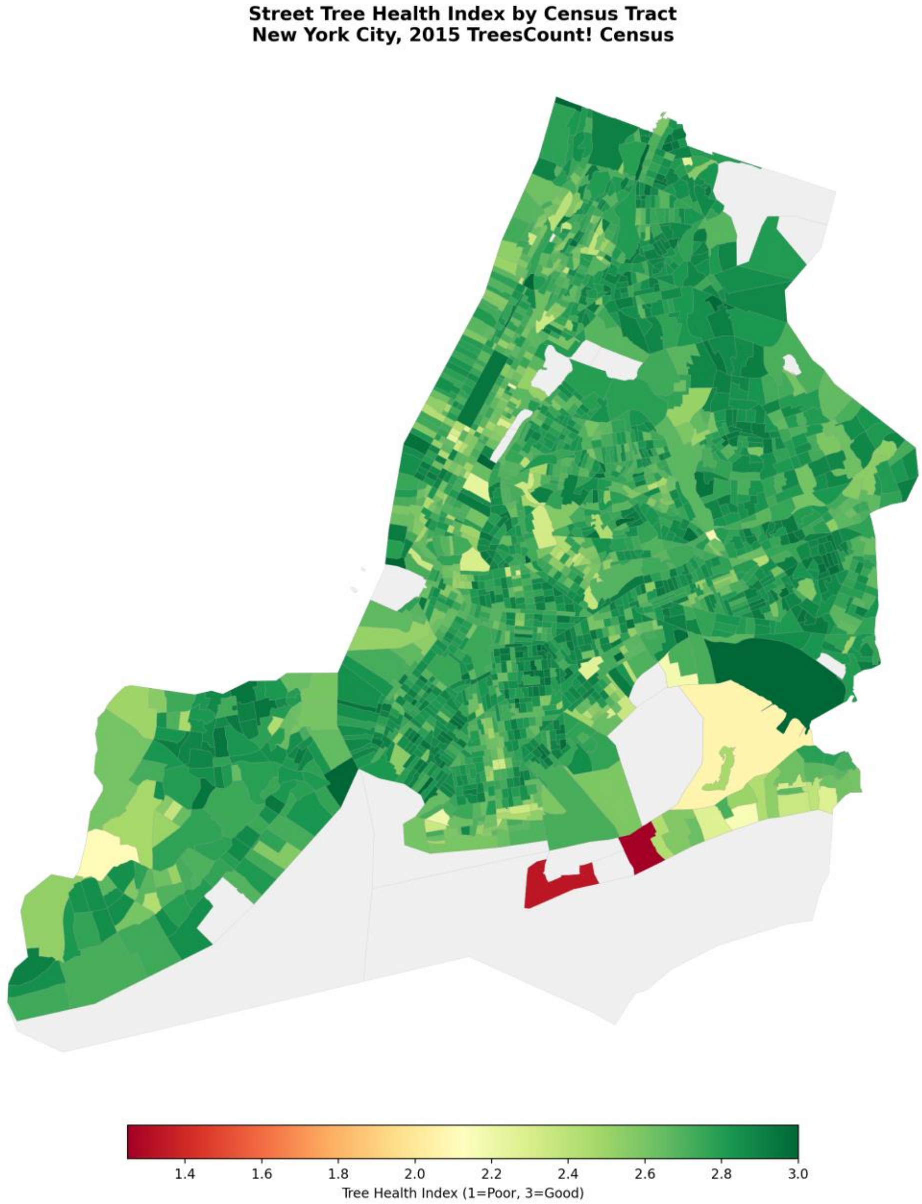


Figure 1. Choropleth Map of Street Tree Health Index by Census Tract New York City, 2015 TreeCount! Census

Figure 1 presents a choropleth map of the Tree Health Index across all 2,306 tracts. The citywide pattern reveals a predominantly green landscape with mean health near 2.76, consistent with a majority-Good distribution. However, pronounced low-health clusters are visible in lower Manhattan (Financial District, Lower East Side) and along the south Brooklyn and south Queens coastlines. These areas combine high building density, elevated impervious surface, and lower canopy coverage, suggesting built-environment constraints on tree vigor. Staten Island exhibits the highest health scores city-wide (see Figure 2).

4.3 Borough-Level Health Composition

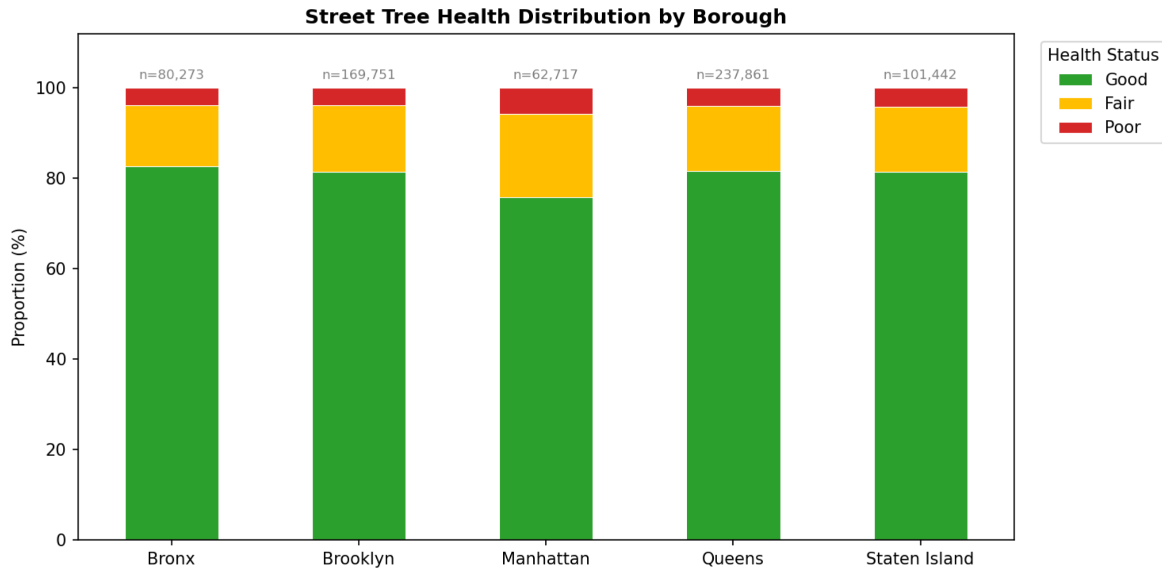


Figure 2. Stacked Bar Chart of Street Tree Health Distribution by Borough

Figure 2 disaggregates health composition by borough. Manhattan shows the lowest proportion of Good-rated trees (~76%) and the highest Poor proportion (~5%), despite having the highest median income among its resident populations. This apparent contradiction is consistent with the built-environment hypothesis: Manhattan's dense, impervious streetscape constrains root zones and moisture retention regardless of neighborhood wealth. Staten Island has the highest Good proportion and lowest Poor proportion, reflecting lower building density and greater canopy coverage.

4.4 Species Composition

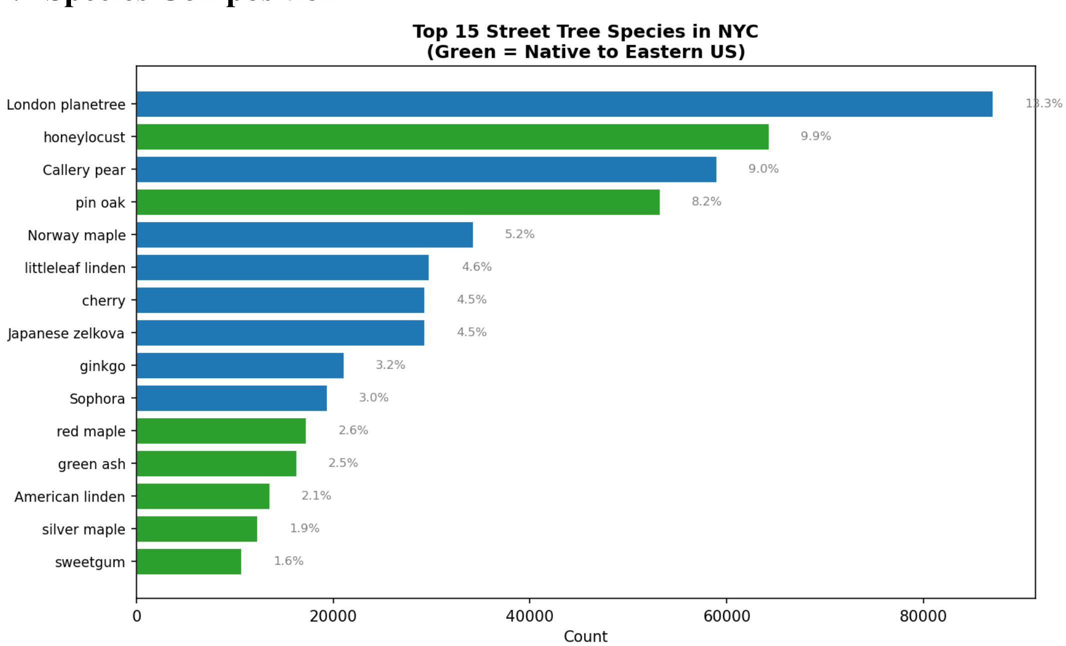


Figure 3. Horizontal Bar Chart of Top 15 Street Tree Species in NYC

Figure 3 displays the top 15 street-tree species by count. London planetree dominates, comprising nearly one-fifth of all alive trees — a legacy of mid-20th century mass planting driven by its pollution tolerance. The predominance of three non-native species in the top ranks (London planetree, honeylocust, Callery pear) underscores ongoing concerns about species homogeneity and its implications for ecological resilience under pest or climate pressures (Berland et al., 2017).

4.5 Socioeconomic Correlates

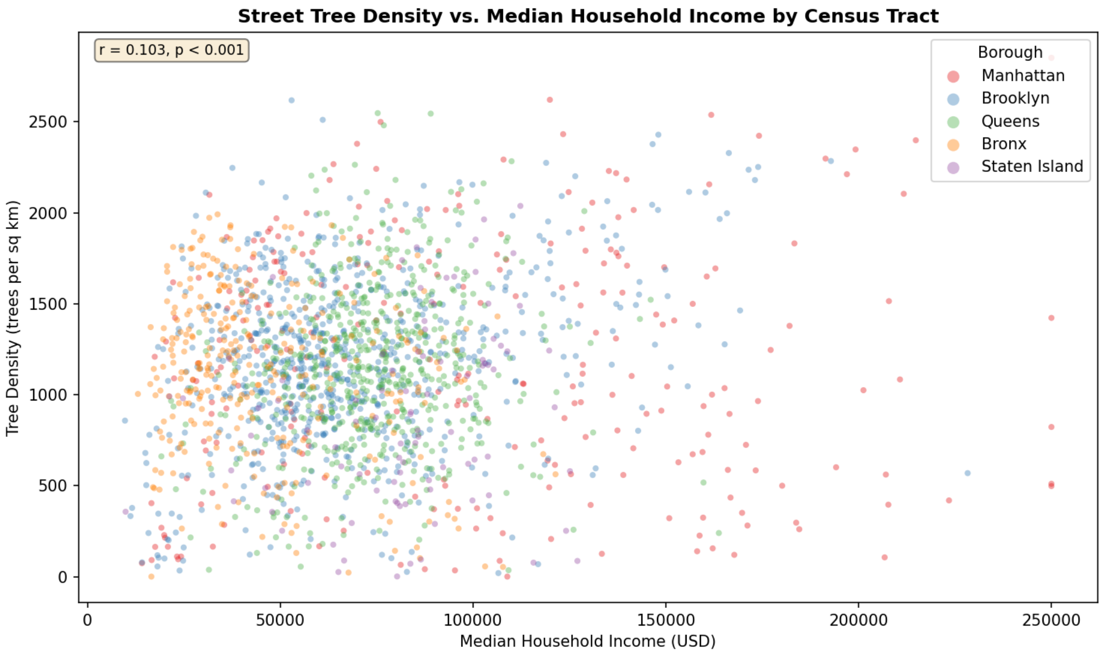


Figure 4. Scatter Plot of Street Tree Density vs. Median Household Income

Figure 4 examines the bivariate relationship between tree density and median household income. A statistically significant but weak positive correlation is observed ($r = 0.102$, $p < 0.001$), indicating that higher-income tracts tend to have marginally higher tree density. Manhattan tracts form a visually distinct cluster characterized by high income but anomalously low tree density, reflecting the borough's extreme built-environment intensity. This pattern suggests that income alone is insufficient to explain tree distribution; built-environment constraints must be considered jointly.

4.6 Land Cover Composition by Borough

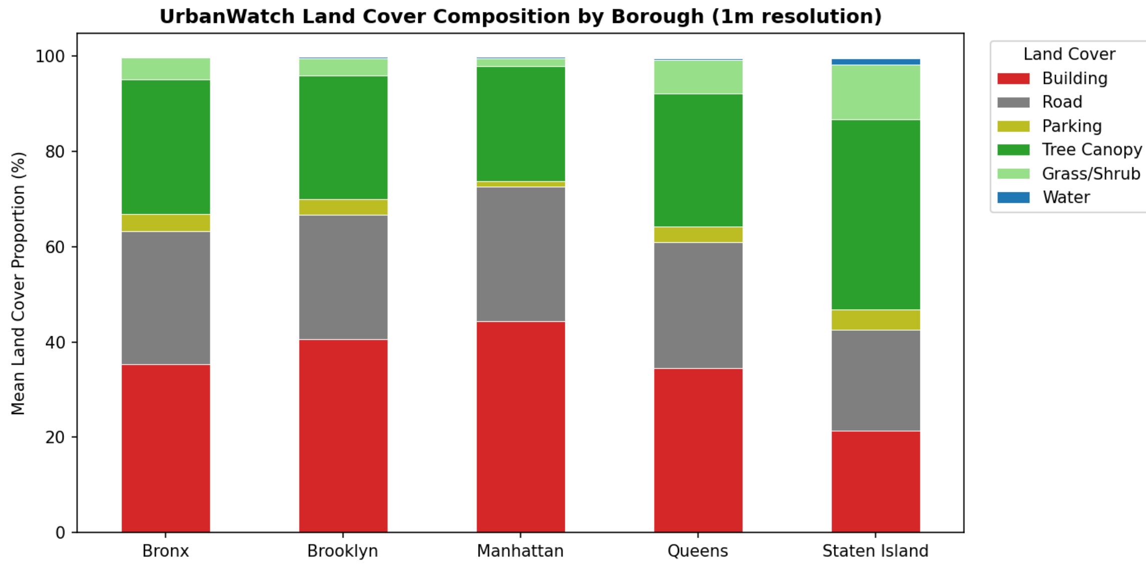


Figure 5. Grouped bar chart of Land Cover Composition by Borough(1m resolution)

Figure 5 presents land cover composition (PLAND by class) stratified by borough. Staten Island has the highest mean tree canopy PLAND (~40%) and lowest impervious surface (~52%), while Manhattan has the lowest tree canopy (~6%) and highest impervious cover (~93%). The Bronx and Brooklyn occupy intermediate positions, with Brooklyn showing notable variability across tracts. These borough-level patterns establish the structural context within which tree health and diversity outcomes are subsequently modeled.

4.7 Built Environment Intensity and Tree Health

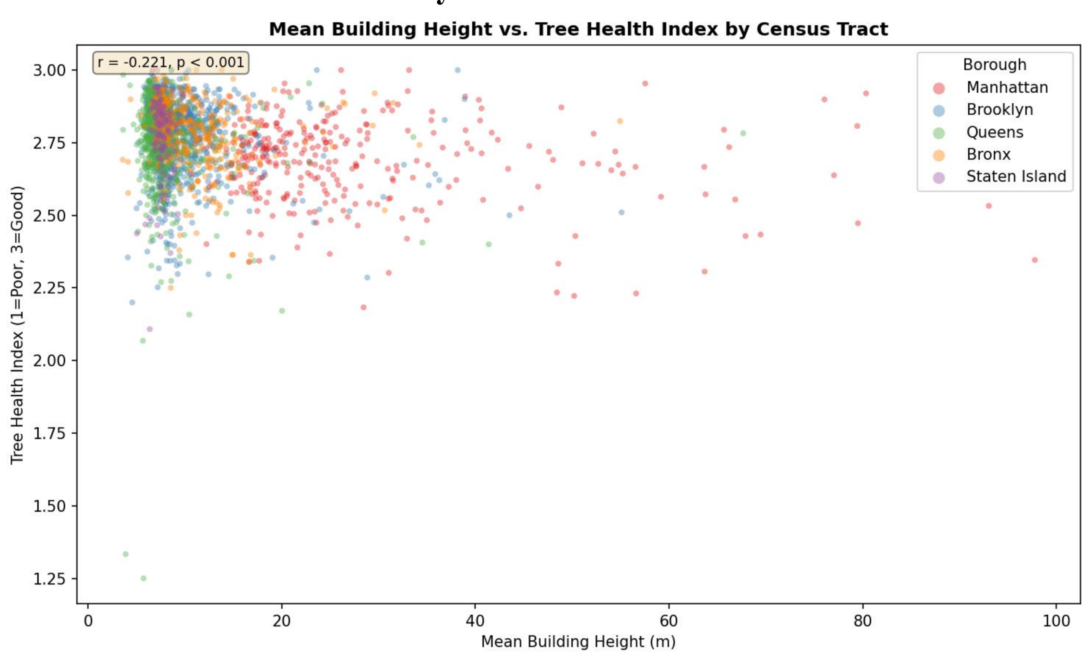


Figure 6. Scatter Plot of Mean Building Height vs. Tree Health Index

Figure 6 examines the relationship between mean building height and Tree Health Index at the tract level. A significant negative correlation is observed ($r = -0.221$, $p < 0.001$), indicating that tracts with taller buildings tend to exhibit lower tree health. This effect is consistent with the hypothesis that vertical densification reduces solar access, increases street-canyon wind effects, and compresses root-zone soil volumes available to street trees. The relationship is notably stronger in the Pearson correlation than the income-tree density relationship, suggesting that physical built-environment factors may be more proximal deter

5. Spatial Autocorrelation and LISA Clusters

After completing tract-level data integration and descriptive mapping, this study further employed spatial autocorrelation analysis to examine whether Tree Health Index, Street Tree Density, and Species Diversity exhibit non-random spatial distributions. Specifically, Global Moran's I was used to test overall spatial clustering patterns across New York City, while Local Moran's I / LISA was used to identify localized spatial typologies, including High-High clusters (high values surrounded by high values), Low-Low clusters (low values surrounded by low values), and two categories of spatial outliers: High-Low and Low-High.

All spatial autocorrelation analyses were conducted using a row-standardized k-nearest-neighbors spatial weights matrix with $k = 8$. This specification helps avoid disconnected-neighbor issues caused by NYC's islands, water boundaries, and discontinuous tract geometries, while also ensuring consistency between the LISA analysis and the subsequent spatial econometric models in terms of spatial neighborhood definition. Statistical significance was assessed using 999 permutation tests, with $p < 0.05$ used as the threshold for identifying significant local spatial clusters.

The results indicate that all three street-tree ecological indicators exhibit significant positive spatial autocorrelation, suggesting that street-tree ecological quality in NYC is not randomly distributed but instead demonstrates stable spatial clustering structures. Tree Health Index has a Moran's I value of 0.313, indicating a moderate level of spatial clustering in tree health conditions. Street Tree Density shows the strongest spatial clustering among the three outcomes, with a Moran's I of 0.361, suggesting that street-tree supply is influenced not only by conditions within individual tracts but also by adjacent planting patterns and continuity in urban form. Species Diversity has a Moran's I of 0.300, indicating that species diversity also exhibits clear spatial organization, potentially reflecting the spatial continuity of historical planting strategies, road hierarchy, neighborhood morphology, and maintenance systems.

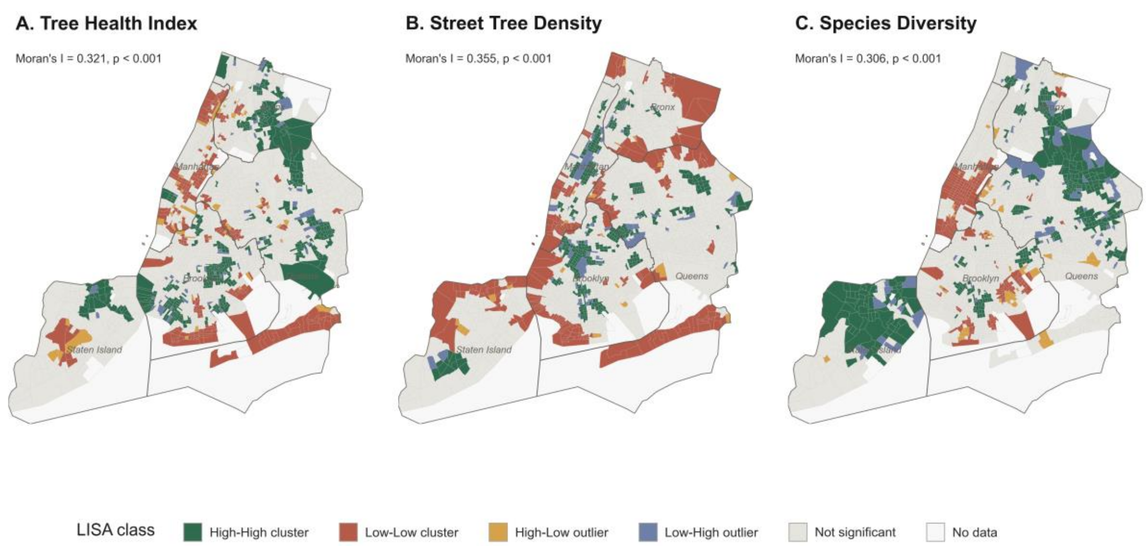
Outcome	N	Moran's I	z-score	p-value
Tree Health Index	2306	0.313	31.667	0.001
Street Tree Density	2327	0.361	35.754	0.001
Species Diversity	2306	0.300	29.498	0.001

Table 4. Global Moran's I results.

From the perspective of LISA interpretation, High-High clusters can be understood as street-tree ecological advantage zones, where high ecological values are surrounded by similarly high neighboring tracts. In contrast, Low-Low clusters represent continuous ecological disadvantage zones and therefore constitute the most critical spatial units in subsequent co-burden mapping analyses. Low-Low clusters of Tree Health Index indicate areas where tree health conditions are persistently weak across neighboring tracts; Low-Low clusters of Street Tree Density point to contiguous areas with insufficient street-tree supply; and Low-Low clusters of Species Diversity suggest areas characterized by low species diversity and weaker ecological resilience.

By comparison, High-Low and Low-High outliers carry stronger diagnostic significance. Rather than representing large-scale continuous problem areas, they reflect localized anomalies associated with specific street environments, park edges, atypical land uses, historical planting projects, or differences in data definition.

Therefore, street-tree inequality in NYC should not be understood simply as differences between individual tracts, but rather as a phenomenon shaped by explicit spatial neighborhood effects. This finding suggests that the co-burden index should not rely solely on tract-level ranking values; it should also consider whether low quality, low supply, or low diversity form continuous spatial clusters, because contiguous Low-Low clusters are more appropriate priority units for street-tree maintenance, replanting, and tree-pit redesign interventions.



Outcome	High-High cluster	Low-Low cluster	High-Low outlier	Low-High outlier	Not significant	No data
Tree Health Index	383 (16.5%)	276 (11.9%)	59 (2.5%)	60 (2.6%)	1520 (65.3%)	29 (1.2%)
Street Tree Density	395 (17.0%)	234 (10.1%)	25 (1.1%)	66 (2.8%)	1595 (68.5%)	12 (0.5%)
Species Diversity	362 (15.6%)	224 (9.6%)	39 (1.7%)	62 (2.7%)	1611 (69.2%)	29 (1.2%)

Figure 7. Local spatial clusters of street-tree ecological outcomes across New York City Census Tracts.

6. Explanatory Baseline and Spatial Econometric Models

6.1 OLS Baseline Models

Building upon the spatial clustering results, this section further examines the explanatory relationships between street-tree ecological quality and socioeconomic variables, built-environment characteristics, land-cover composition, and landscape configuration metrics. Previous analyses demonstrated that Tree Health Index, Street Tree Density, and Species Diversity all exhibit significant positive spatial autocorrelation. Accordingly, ordinary least squares (OLS) models are first introduced as an interpretable baseline framework to evaluate the general explanatory power of different variable groups for the three street-tree ecological outcomes.

Subsequently, spatial econometric models are introduced to assess whether accounting for spatial dependence among neighboring tracts improves model performance and reduces residual spatial autocorrelation.

Three dependent variables were modeled, each representing a distinct dimension of street-tree ecological quality. Tree Health Index captures overall tree health condition, Street Tree Density represents street-tree supply intensity, and Species Diversity reflects species composition diversity and potential ecological resilience.

All predictor variables were standardized prior to modeling, and variance inflation factor (VIF) screening was applied to reduce multicollinearity. The OLS results reveal substantial differences in explanatory power across the three outcomes. Street Tree Density achieves the highest adjusted R^2, approximately 0.609, indicating strong structural associations between street-tree supply and tract-level built environment, land cover, and landscape configuration. Species Diversity shows moderate explanatory power, with an adjusted R^2 of approximately 0.367, suggesting that species diversity is partially shaped by urban morphology and green-space structure, but may also be influenced by unobserved factors such as historical planting strategies, nursery supply systems, street hierarchy, and maintenance practices.

In contrast, Tree Health Index exhibits the weakest explanatory performance, with an adjusted R^2 of only 0.159, suggesting that tree health conditions are comparatively difficult to explain using tract-level variables alone.

Outcome	n	Adjusted R^2	AIC	Residual Moran's I	Residual p	Interpretation
Tree Health Index	1,959	0.159	−2,867.1	0.265	0.001	Weak tract-level explanatory power; strong residual spatial structure
Street Tree Density	1,961	0.609	2,168.7	0.186	0.001	Strongest baseline fit; supply remains spatially structured
Species Diversity	1,959	0.367	1,548.7	0.261	0.001	Moderate fit; diversity retains strong spatial dependence

Table 5. OLS model results.

6.2 Spatial Lag and Spatial Error Models

However, the residuals of the OLS models still exhibit significant spatial autocorrelation. Residual Moran's I values remain significant for all three outcomes, indicating that non-spatial models fail to fully account for neighborhood effects embedded in street-tree ecological outcomes. In other words, street-tree ecological quality is not solely the product of internal tract-level conditions, but also demonstrates strong spatial continuity and neighborhood dependence. This finding directly justifies the need for spatial econometric modeling.

Accordingly, this study further estimates Spatial Lag and Spatial Error models. The Spatial Lag model assumes that the dependent variable in neighboring tracts directly influences the outcome in the focal tract, making it suitable for representing spatial diffusion or coordinated spatial organization in street-tree health, density, and diversity. In contrast, the Spatial Error model assumes that spatial dependence primarily exists in unobserved error terms, representing omitted variables or latent spatial processes.

Together, these two model types help determine whether spatial dependence is better interpreted as "outcome-level neighborhood dependence" or as "unobserved spatial error."

Model comparison results indicate that the Spatial Lag model performs more consistently for Tree Health Index and Species Diversity. For Tree Health Index, the Spatial Lag specification reduces the residual Moran's I from 0.265 in the OLS model to nearly zero and statistically insignificant levels, indicating that the spatial structure of tree health can largely be explained by neighboring tracts' health conditions. Similarly, for Species Diversity, the Spatial Lag model effectively removes residual spatial autocorrelation, suggesting strong spatial continuity in species composition patterns.

By comparison, the Spatial Error fallback models retain significant residual Moran's I values, indicating that they fail to fully eliminate spatial dependence. Therefore, the Spatial Lag specification should be treated as the primary spatial model in the main text, while Spatial Error models are more appropriately positioned as robustness comparisons or supplementary analyses.

The case of Street Tree Density is more complex. Although the Spatial Lag model improves model performance and reduces residual Moran's I, significant residual spatial autocorrelation remains. This suggests that street-tree density is influenced not only by tract-level built environment and land-cover conditions, but also by larger-scale processes such as street-network structure, road hierarchy, planting policy, maintenance districts, and historical greening investments. These broader processes may not be fully captured by Census Tract–level variables.

Outcome	**Preferred spatial model**	**Spatial parameter ρ**	**Pseudo R^2**	**AIC**	**Residual Moran's I**	**Residual p**	**Main finding**
Tree Health Index	Spatial Lag	0.568	0.341	−3,218.8	−0.001	0.476	Spatial dependence largely removed
Street Tree Density	Spatial Lag	0.304	0.638	2,063.2	0.101	0.001	Fit improves, but residual clustering remains
Species Diversity	Spatial Lag	0.557	0.507	1,171.8	0.012	0.144	Spatial dependence largely removed

Table 6. Spatial econometric model comparison.

7. Machine Learning Diagnosis

In this section, machine learning is introduced as a nonlinear diagnostic layer to examine whether socioeconomic, built-environment, land-cover, and landscape-configuration variables can predict street-tree ecological outcomes under conditions of nonlinear relationships, variable interactions, and spatial heterogeneity.

Unlike OLS and spatial econometric models, machine learning models are not primarily used to estimate linear directions or statistical significance. Instead, they are employed to address three diagnostic questions:

1. Whether the three street-tree ecological indicators exhibit stable predictability under multivariate conditions;
2. Whether nonlinear models significantly improve predictive performance;
3. Which variables contribute most strongly to prediction and whether they exhibit threshold effects or spatial heterogeneity.

7.1 Model Training and Validation

Random Forest, CatBoost, and XGBoost models were separately constructed for Tree Health Index, Street Tree Density, and Species Diversity. Model inputs included social vulnerability, built-environment intensity, land-cover composition, green infrastructure supply, and landscape-configuration metrics.

To avoid overly optimistic estimates caused by random train-test splits in spatial data, this study employed not only a standard 80/20 random split, but also a stricter borough-based GroupKFold validation framework. In the latter approach, boroughs were treated as grouping units, allowing the models to be tested on whether relationships learned from some boroughs could generalize to boroughs excluded from training.

7.2 Model Performance Comparison

The results reveal substantial differences in predictability among the three outcomes.

Street Tree Density performs consistently well under both random split and borough-based validation. CatBoost maintains relatively high R^2 values even under borough GroupKFold validation, indicating that street-tree supply can be effectively explained by tract-level land cover, built environment, and landscape configuration variables. This suggests that street-tree density is best understood as a structurally determined supply outcome shaped jointly by planning history, road-space conditions, green-space configuration, and built intensity.

Species Diversity demonstrates moderate predictive performance. While performance is relatively strong under random split validation, it declines noticeably under borough-based validation, indicating spatial heterogeneity in species-diversity prediction. In other words, species-composition patterns learned within certain boroughs may not generalize effectively to others. Therefore, species diversity should not be interpreted solely as a function of land cover or tree quantity, but rather as the combined outcome of historical planning and ecological management processes.

By contrast, Tree Health Index exhibits the weakest predictive performance. Under borough-based validation, R^2 values approach zero, and some models even produce negative values. This does not imply computational failure; rather, it indicates that the current tract-level variables possess very limited ability to predict tree health across boroughs.

Because Tree Health Index itself has a relatively compressed distribution, with most street trees rated as "Good," the amount of learnable variation is inherently limited. Moreover, tree health is more likely determined by fine-scale factors such as tree-pit dimensions, soil volume, root-space availability, irrigation, pruning, pest pressure, road salt exposure, maintenance history, and microclimatic stressors. These variables are not adequately captured within the current Census Tract–level data framework.

Outcome	**Best model**	**Random split R^2**	**Borough GroupKFold R^2**	**RMSE**	**MAE**	**Interpretation**
Tree Health Index	Random Forest	0.140	0.035	0.133	0.100	Weak borough-level transferability
Street Tree Density	CatBoost	0.699	0.675	0.645	0.363	Strong and stable predictability
Species Diversity	CatBoost	0.466	0.266	0.410	0.315	Moderate but spatially sensitive predictability

Table 9. Machine learning model performance.

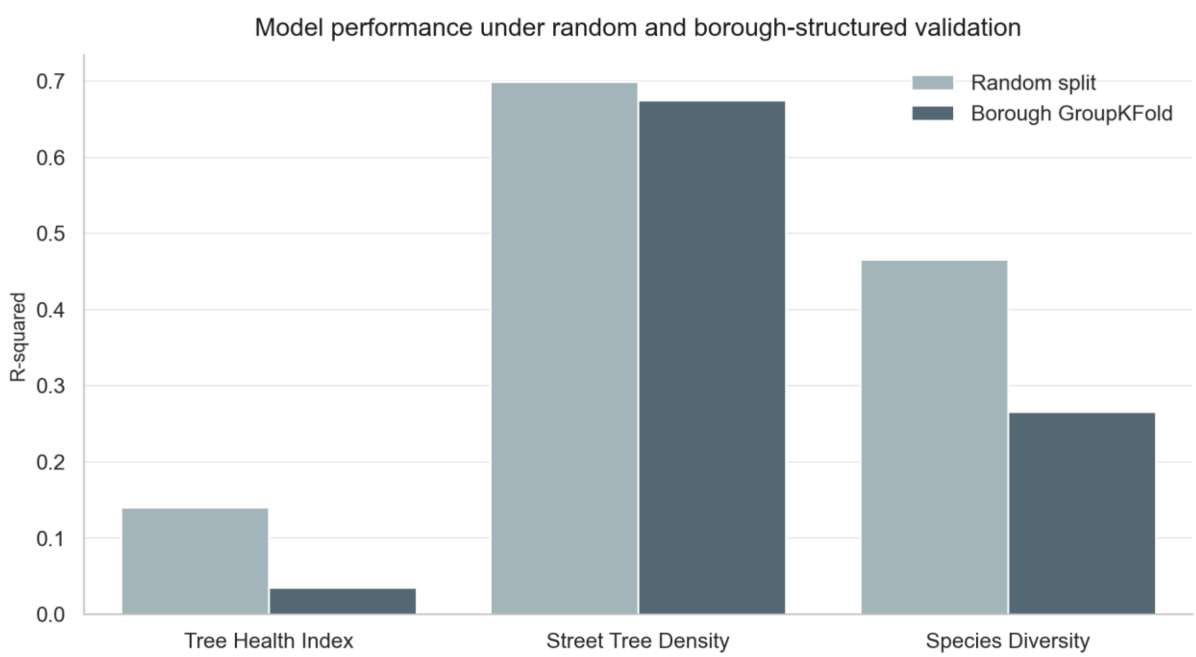


Figure 8. Machine learning performance under random and borough-structured validation.

7.3 Variable Importance and SHAP Interpretation

7.3.1 Permutation Importance Overview

To further understand how machine learning models generate predictions, this study combines permutation importance and SHAP analysis to interpret the best-performing models. It is important to emphasize that these interpretation results reflect predictive contribution rather than strict causal effects. Therefore, they should be understood as tools for mechanism diagnosis and hypothesis generation rather than definitive causal evidence.

Permutation importance reveals clear differences in dominant variable structures across the three outcomes. For Tree Health Index, the most influential variables include mean floors, minority percentage, tree canopy Edge Density, tree canopy PLAND, and landscape SHDI, suggesting that tree-health prediction is not concentrated in a single morphological variable, but is influenced by the combined effects of canopy-edge structure, neighborhood building intensity, and landscape heterogeneity. For Street Tree Density, the most important predictors are concentrated within built-environment and green-structure variables, particularly building density, building coverage, tree canopy ED, and mean FAR, indicating that street-tree supply is mainly shaped by built-space intensity and canopy-structure conditions. For Species Diversity, tree canopy NP, building density, mean floors, building NP, and tree canopy PD show relatively high importance, suggesting that species diversity is related not only to tree quantity itself, but also to the number of green-space patches, built-environment complexity, and neighborhood morphological structure.

SHAP summary plots further reveal the directional and nonlinear relationships embedded in these predictors.

7.3.2 SHAP Interpretation for Tree Health Index

For Tree Health Index, tree canopy ED, tree canopy PLAND, and mean floors are the three most notable variables. SHAP dependence plots show that tree canopy ED exhibits an overall positive relationship with tree health. When tree canopy ED increases from lower levels to approximately 550–600, its contribution to the health index shifts from negative to positive, and the positive effect gradually strengthens thereafter. This suggests that, within NYC's street-tree system, higher canopy-edge density may correspond to more continuous, accessible, or effectively integrated canopy structures embedded within street space.

Meanwhile, tree canopy PLAND exhibits a clear threshold effect. Below approximately 28%–30%, its contribution to the health index is generally negative; above this range, contributions become positive. This suggests that a certain level of surrounding canopy coverage is an important background condition supporting healthier street trees.

In contrast, mean floors displays an inverted-U or staged negative relationship. Within low-rise to mid-low-rise built environments, model predictions are relatively favorable; however, beyond approximately 4–5 floors, contributions to Tree Health Index become increasingly negative. This suggests that greater building intensity may increase street-tree stress through shading, street-canyon effects, and constrained root-zone space.

It should be noted that because Tree Health Index demonstrates relatively weak predictive performance overall, these patterns should be interpreted as exploratory signals rather than strong causal findings.

7.3.3 SHAP Interpretation for Street Tree Density

For Street Tree Density, SHAP dependence plots reveal clearer and more stable structural relationships. The influence of building density follows a nonlinear pattern characterized by rapid increase followed by stabilization. At very low building density, SHAP values are strongly negative; as building density increases, predictive contributions quickly become positive and remain stable across medium-to-high density ranges. This suggests that extremely low-density areas often lack continuous street interfaces and suitable conditions for street-tree configuration, whereas medium- and high-density neighborhoods are more likely to support systematic street-tree supply.

Building coverage exhibits a similar threshold pattern. When coverage remains below approximately 15%–20%, contributions to street-tree density are generally negative. Beyond this threshold, contributions rapidly become positive and stabilize after approximately 30%. This suggests that moderately continuous built street edges may be more conducive to forming regularized street-tree planting systems.

At the same time, tree canopy ED demonstrates a strong monotonic positive relationship, with positive contributions increasing significantly after approximately 800. This further suggests that street-tree supply is closely related to the continuity of canopy-edge structures.

Overall, the SHAP patterns for Street Tree Density reveal the strongest structural regularity and clearest threshold behavior, consistent with its highest and most stable predictive performance in machine learning validation.

7.3.4 SHAP Interpretation for Species Diversity

For Species Diversity, SHAP results reveal a more complex spatial structural logic. First, tree canopy NP is among the most important variables. Its dependence plot exhibits a typical rapid-growth-followed-by-saturation relationship. At low patch counts, SHAP values are strongly negative; as tree canopy NP increases to approximately 400–500, positive contributions increase rapidly before stabilizing. This suggests that too few canopy patches often correspond to simplified, fragmented, or insufficient green-space environments, whereas richer canopy-patch structures are more conducive to supporting higher species diversity.

Second, mean floors exhibits a pronounced inverted-U relationship. Contributions to species diversity are relatively positive within the 2–4 floor range, but become strongly negative after approximately 5–6 floors and continue declining in high-rise environments. This suggests that low- to mid-rise neighborhoods may be most favorable for accommodating diverse species configurations, whereas highly vertical development environments are more likely to compress planting space and reduce compositional complexity.

Third, building density also demonstrates a nonlinear pattern that transitions from negative to positive before stabilizing. Extremely low-density areas are unfavorable for species diversity, medium-density areas are the most favorable, while extremely high-density environments may experience slight declines. This suggests that diversity does not simply increase with density, but is jointly shaped by neighborhood structure, street-space organization, and planting systems.

Taken together, the SHAP results for Species Diversity suggest that species diversity depends not simply on the presence of trees, but also on canopy-patch structure, built-environment organization, and spatial capacity at the low- to mid-rise neighborhood scale.

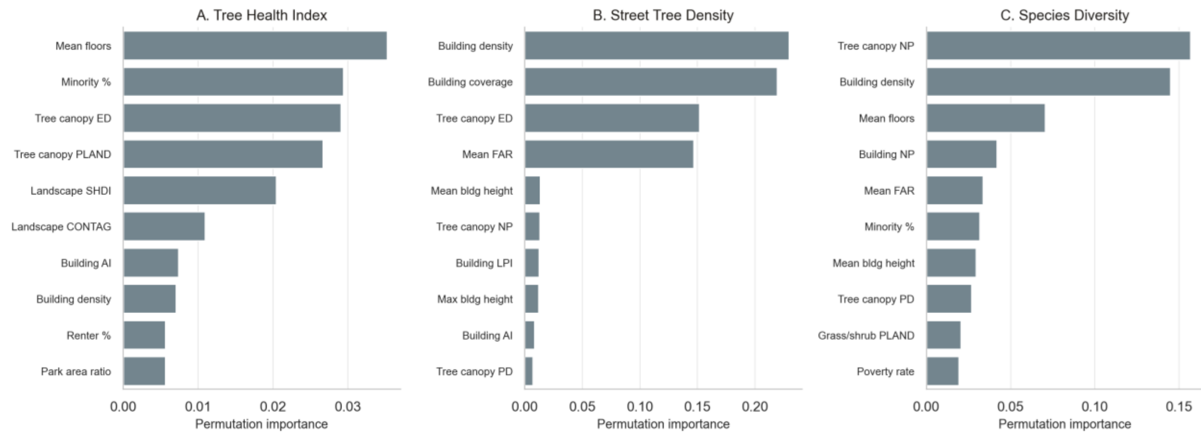


Figure 9. Permutation importance for the best-performing machine learning model for each outcome.

Figure 10. SHAP summary plot for Tree Health Index,Street Tree Density, Species Diversity.

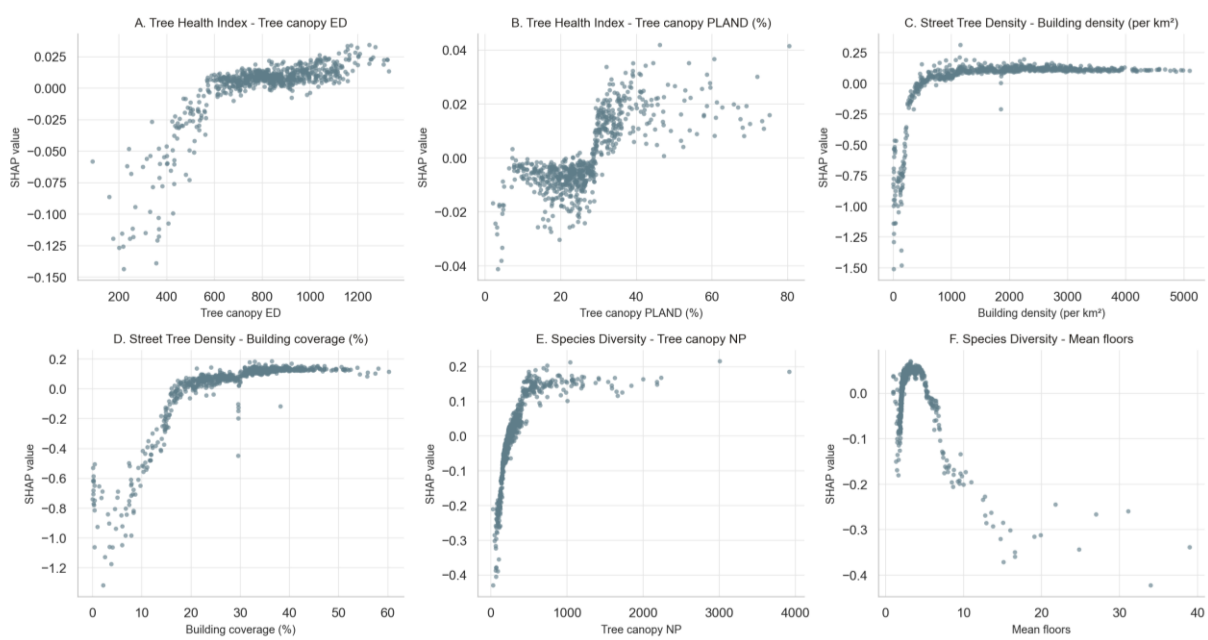


Figure 11. Selected SHAP dependence plots highlight nonlinearities and threshold-like behavior.

7.4 Machine Learning Synthesis

Taken together, the results from model validation, variable importance analysis, and SHAP interpretation reveal substantial differences among the three street-tree ecological indicators in terms of predictability, spatial transferability, and mechanism complexity.

First, Street Tree Density is the most stable and most predictable outcome. Under both random split and borough-based GroupKFold validation, it maintains relatively high R^2 values, indicating that street-tree supply is strongly shaped by tract-level built-environment intensity, land-cover structure, and canopy spatial configuration. This suggests that street-tree density is best understood as a "structural supply outcome," reflecting not only the distribution of green resources but also the combined effects of long-term street planning, planting systems, and roadway-space conditions.

From a planning perspective, this outcome is particularly useful for identifying underserved street-tree areas, comparing supply differences among boroughs, and informing the service-deficit dimension of the subsequent co-burden index.

Second, Species Diversity exhibits moderate predictability, but substantially weaker spatial transferability than street-tree density. Although performance remains relatively strong under random split validation, it declines more noticeably under borough-based validation, indicating that the determinants of species diversity are more localized and historically contingent. Such locality may reflect differences among boroughs in species-selection strategies, street hierarchy, maintenance systems, historical replanting programs, and nursery supply conditions.

Third, the weak predictive performance of Tree Health Index constitutes one of the study's most important findings. Even with nonlinear machine learning models, Tree Health Index remains close to zero under borough-structured validation, indicating that current tract-level socioeconomic, built-environment, land-cover, and landscape-configuration variables are insufficient to consistently explain cross-borough differences in tree health.

This result does not indicate analytical failure. Rather, it demonstrates that "tree health" fundamentally differs from "tree quantity" or "species diversity." Tree health is more strongly controlled by micro-scale conditions such as tree-pit size, soil volume, root-zone space, irrigation and pruning practices, pest pressure, road salt exposure, maintenance frequency, and street-level microclimatic stressors. These variables are not adequately captured within the current dataset.

Therefore, the low predictability of Tree Health Index itself becomes a substantive finding: if future research and policy interventions genuinely aim to improve street-tree health, they must incorporate field surveys, maintenance records, and street-level habitat data, rather than relying solely on tract-level urban-form proxy variables.

8. Co-burden Index and Priority Intervention Areas

This section further examines which Census Tracts simultaneously experience multiple forms of disadvantage and therefore should be identified as priority intervention areas.

To address this question, four single-burden indicators were first constructed and then integrated into a composite co-burden index designed to identify tracts where street-tree ecological disadvantage overlaps most strongly with social and spatial pressures.

8.1 Single-burden Indicators

Methodologically, the four burden indicators each correspond to a different dimension of disadvantage.

First, the low street-tree quality burden identifies tracts with weak street-tree ecological services or poor street-tree conditions. Rather than automatically classifying all treeless tracts as high burden, this indicator incorporates a residential / built-up eligibility mask, identifying street-tree service deficits only within tracts that contain residential populations or street-based built environments. This prevents parks, water bodies, and other non-applicable tracts from being misclassified as high burden areas.

Second, the high social vulnerability burden is derived from SVI percentile scores and reflects the spatial concentration of socioeconomic vulnerability.

Third, the high built-environment intensity burden is constructed from percentile rankings of mean building height, building coverage, and imperviousness, representing the combined physical pressures imposed on street-tree growing environments by tall buildings, dense development, and impermeable surfaces.

Fourth, the high landscape fragmentation burden is based on tree canopy PD, tree canopy ED, tree canopy LPI, and tree canopy AI, identifying areas where canopy structure is more fragmented, edge-dominated, and weakly connected.

To ensure comparability across burdens, all indicators were standardized into normalized percentile scores ranging from 0 to 1, where higher values indicate greater burden.

8.2 Single-burden Maps

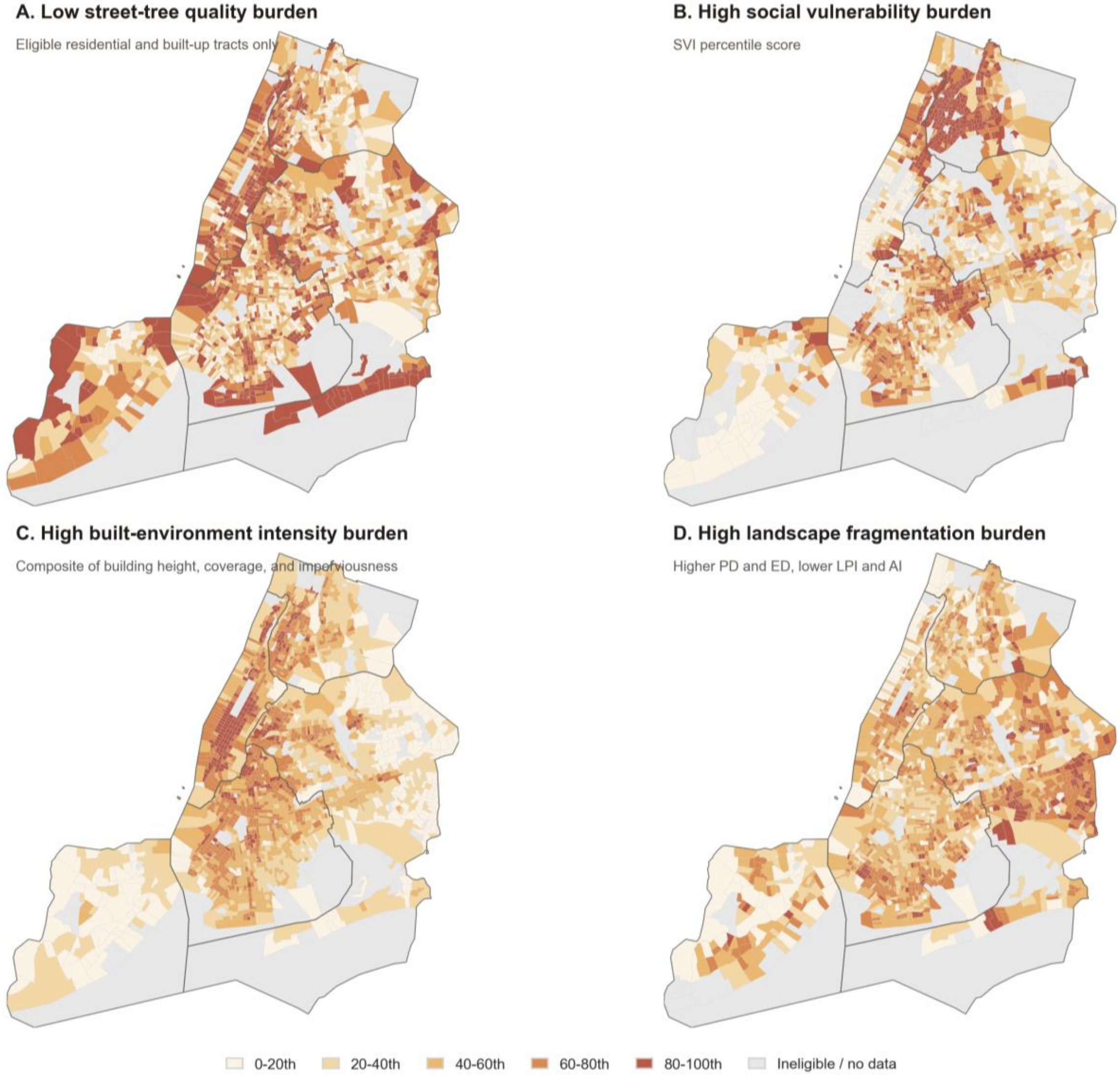


Figure 12. Single-burden maps: low street-tree quality, high social vulnerability, high built-environment intensity, and high landscape fragmentation.

The single-burden maps reveal distinct yet interconnected spatial patterns of disadvantage across New York City.

First, the low street-tree quality burden primarily captures weaknesses within the street-tree system itself, rather than simply reflecting overall urban canopy coverage. Figure A shows that high-burden tracts are not concentrated within a single borough, but instead appear across Manhattan, the Bronx, Brooklyn, and portions of Staten Island. This suggests that deficiencies in street-tree quality or street-tree services are highly localized and may be associated with factors such as street-tree inventory levels, roadway-space conditions, tree-pit environments, maintenance history, and continuity of street interfaces. In other words, low street-tree quality burden should not be interpreted simply as "low canopy cover," but rather as a more direct indication of insufficiencies within the street-tree system embedded in specific built environments.

Second, the high social vulnerability burden exhibits a clearer pattern of socio-spatial concentration. In Figure B, high-value areas are especially prominent in the Bronx, northern Manhattan, central Brooklyn, and portions of Queens, indicating that social vulnerability is not evenly distributed across the city but instead forms continuous high-burden zones within particular communities. This finding provides a critical foundation for environmentally just street-tree interventions: ecological improvement strategies should not be prioritized solely according to ecological indicators, but must also account for which communities simultaneously face higher social vulnerability and lower adaptive capacity.

Third, the high built-environment intensity burden reflects the combined physical pressures created by tall buildings, high building coverage, and highly impervious urban environments. Figure C shows that this burden is particularly concentrated in Manhattan and other high-density urbanized areas, while also appearing across several continuous neighborhoods in Brooklyn and Queens. Unlike social vulnerability burden, built-environment intensity emphasizes the physical constraints imposed on street-tree growing conditions, including limited root-zone space, high levels of impermeable surface cover, compact street-canyon environments, and elevated thermal stress. Thus, it addresses whether street trees possess adequate environmental growing conditions rather than socioeconomic disadvantage itself.

Fourth, the high landscape fragmentation burden captures the degree of fragmentation within canopy spatial structure. High-value areas in Figure D are more spatially dispersed and do not fully overlap with either the most socially vulnerable or the most densely developed areas. This suggests that canopy fragmentation is not determined solely by socioeconomic disadvantage or urban density, but is also related to landscape-configuration characteristics such as canopy patch count, edge density, largest patch proportion, and spatial aggregation. Higher fragmentation burden indicates more isolated and edge-dominated canopy structures with weaker connectivity, potentially reducing continuous shading, ecological connectivity, and heat-mitigation capacity.

Together, the four burdens characterize street-tree ecological inequality from four distinct perspectives: street-tree service deficit, social vulnerability, built-environment pressure, and landscape fragmentation. The differences visible in Figure 12 indicate that some areas primarily experience deficiencies in street-tree quality, while others are more strongly shaped by social vulnerability or built-environment stress. The areas requiring the greatest attention are those where multiple burdens overlap spatially.

Therefore, the role of the single-burden maps is not to directly determine final intervention priorities, but rather to provide an interpretable foundation for the subsequent composite co-burden index by clarifying which forms of disadvantage contribute most strongly to high-priority areas.

8.3 Sensitivity Analysis

Building upon the single-burden indicators, this study constructs a composite co-burden index by equally averaging the four burden dimensions. The index is not intended to function as a causal model, but rather as a decision-support screening tool designed to identify tracts where low street-tree quality, high social vulnerability, high built-environment pressure, and high landscape fragmentation overlap most strongly.

Based on the relative ranking of the composite index, the top 20% of tracts are classified as priority tracts, while the top 10% are classified as high-priority tracts.

To test whether these results are overly dependent on a single burden component or a specific weighting assumption, a sensitivity analysis was conducted. This included an SVI-weighted version of the index, as well as leave-one-out comparisons excluding fragmentation, built intensity, and low street-tree quality burdens individually.

The results indicate that the overall geography of priority areas remains relatively stable across different specifications, suggesting that the composite co-burden index demonstrates strong robustness.

Sensitivity Test	**Top 10% Overlap with Main Index**	**Interpretation**
Equal-weight index	100.000	Main reference index.
SVI-weighted index	68.282	Tests whether priority geography shifts toward socially vulnerable neighborhoods.
Excluding fragmentation	74.449	Tests dependence on landscape-configuration metrics.
Excluding built intensity	73.128	Tests dependence on built-environment pressure.
Excluding low street-tree quality	59.031	Tests whether priority areas are driven mainly by social and built factors.

Table 10. Co-burden sensitivity analysis.

8.4 Priority Tracts

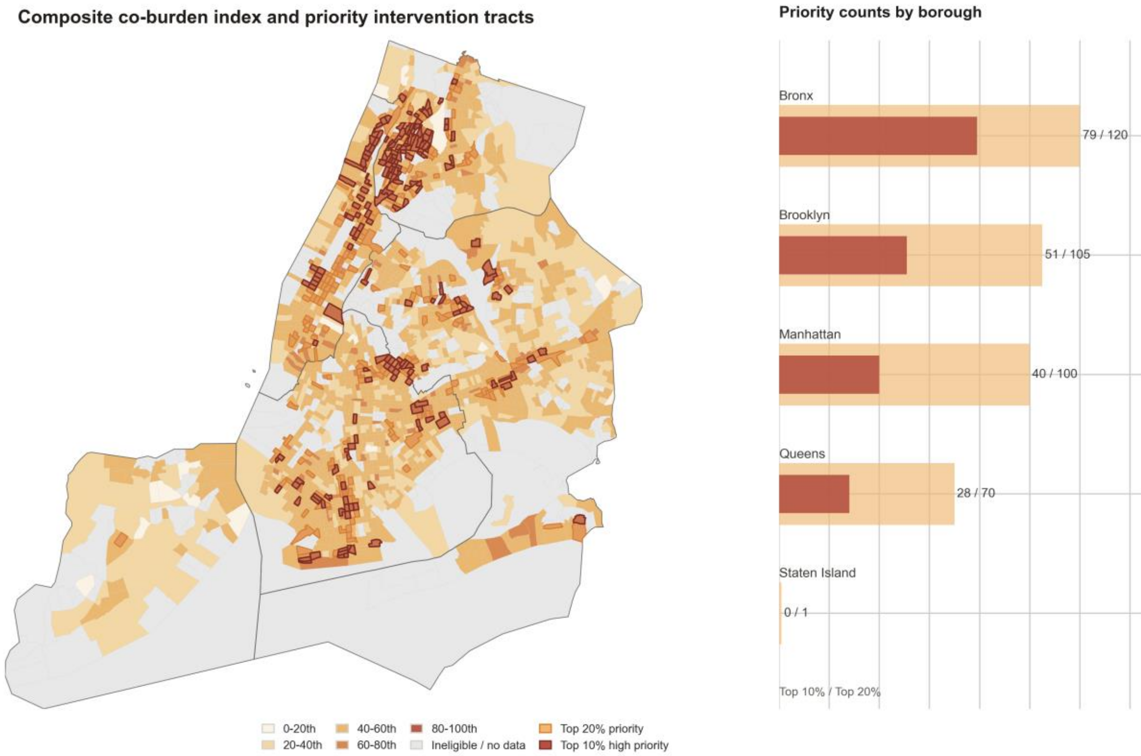


Figure 13. Composite co-burden index and priority tracts.

The composite co-burden index further demonstrates that street-tree ecological inequality in New York City is neither driven by a single indicator nor evenly distributed across urban space. Instead, multiple burdens form pronounced clusters within several continuous urban districts.

In Figure 13, the base map displays percentile rankings of the composite co-burden score. Darker areas indicate stronger overlap among low street-tree quality, high social vulnerability, high built-environment intensity, and high landscape fragmentation. Orange outlines represent the top 20% priority tracts, while dark red outlines indicate the top 10% high-priority tracts.

Compared with the single-burden maps, the significance of the composite map lies in its simultaneous identification of multidimensional spatial disadvantage rather than simply locating tracts with the highest or lowest values for a single indicator.

Spatially, high-priority tracts are strongly concentrated in the Bronx–northern Manhattan corridor, extending through central and northern Manhattan, the South Bronx, and portions of central Brooklyn as continuous high-burden zones. Queens also contains several dispersed priority clusters, although their intensity and continuity are weaker than those observed in the Bronx, Manhattan, and Brooklyn. Staten Island almost entirely falls outside the high-priority category, suggesting that under this study's framework it rarely experiences simultaneous exposure to all four burdens.

The borough summary further reinforces this interpretation. The Bronx contains by far the greatest number of both top 10% and top 20% priority tracts, totaling 79 and 120 respectively. Brooklyn follows with 51 and 105, Manhattan with 40 and 100, Queens with 28 and 70, and Staten Island with only 0 and 1. These results indicate that the Bronx not only contains isolated high-value tracts, but forms the city's most concentrated zone of overlapping ecological and social burdens.

GEOID	**NAMELSAD**	**borough**	**co_burden_equal_weight**	**low_street_tree_quality_burden**	**high_social_vulnerability_burden**	**high_built_environment_intensity_burden**	**high_landscape_fragmentation_burden**
36061026900	Census Tract 269	Manhattan	0.863	0.858	0.984	0.897	0.712
36061010803	Census Tract 108.03	Manhattan	0.860	0.798	NA	0.968	0.814
36047001502	Census Tract 15.02	Brooklyn	0.851	0.954	NA	0.971	0.627
36005038500	Census Tract 385	Bronx	0.846	0.971	0.974	0.691	0.749
36005023702	Census Tract 237.02	Bronx	0.844	0.955	0.956	0.914	0.551
36061012601	Census Tract 126.01	Manhattan	0.840	0.868	NA	0.856	0.797
36005017901	Census Tract 179.01	Bronx	0.839	0.833	0.969	0.854	0.701
36005022102	Census Tract 221.02	Bronx	0.839	0.815	0.989	0.882	0.671
36061013603	Census Tract 136.03	Manhattan	0.839	0.955	NA	0.861	0.700
36005019900	Census Tract 199	Bronx	0.834	0.927	0.948	0.755	0.706
36047044700	Census Tract 447	Brooklyn	0.832	0.970	0.820	0.826	0.711
36005039902	Census Tract 399.02	Bronx	0.832	0.890	0.944	0.877	0.615
36061012602	Census Tract 126.02	Manhattan	0.830	0.881	NA	0.830	0.778
36061013201	Census Tract 132.01	Manhattan	0.827	0.954	NA	0.704	0.822
360610	Census	Manh	0.826	0.907	0.880	0.887	0.632

28500	Tract 285	attan					
36005037300	Census Tract 373	Bronx	0.826	0.960	0.992	0.627	0.725
36005022101	Census Tract 221.01	Bronx	0.826	0.857	0.873	0.900	0.673
36061026100	Census Tract 261	Manhattan	0.823	0.847	0.922	0.895	0.629
36005018102	Census Tract 181.02	Bronx	0.821	0.726	0.978	0.874	0.706
36061024500	Census Tract 245	Manhattan	0.814	0.857	0.894	0.908	0.598

Table 11. Top high-priority intervention tracts.

borough	Other tracts	Top 10% high priority	Top 20% priority
Bronx	227	78	45
Brooklyn	671	52	66
Manhattan	192	54	61
Queens	605	43	54
Staten Island	119	0	1

Table 12. Borough summary of priority tracts.

The ranked tract map further reveals the internal structure of high-priority areas. The top 15 tracts are highly concentrated in the Bronx and Manhattan. The highest-ranked tract is Manhattan Census Tract 269, with a co-burden score of 0.863. This is followed by multiple Bronx tracts, including CT 385, CT 237.02, CT 179.01, CT 221.02, CT 199, CT 399.02, CT 373, CT 221.01, CT 181.02, and CT 233.02. Brooklyn's CT 447 also appears within the top 15.

This ranking pattern demonstrates that high co-burden conditions are not randomly scattered across the city, but instead form stable priority intervention corridors shaped by overlapping social, ecological, and morphological pressures. In particular, the continuous concentration of high values across the Bronx and northern Manhattan suggests that these areas simultaneously experience weak street-tree quality, elevated social vulnerability, strong built-environment pressure, and fragmented canopy structure.

These findings connect directly with the previous analytical sections. Section 5 demonstrated that tree health, density, and species diversity exhibit significant spatial autocorrelation, indicating that street-tree ecological quality forms clustered spatial structures. Section 6 further showed that OLS residuals retain substantial spatial dependence, while Spatial Lag models absorb part of this spatial structure more effectively. Section 7 demonstrated through machine learning that street-tree density and species diversity are strongly associated with built-environment conditions and canopy structure, whereas tree health appears more strongly tied to fine-scale maintenance and habitat conditions.

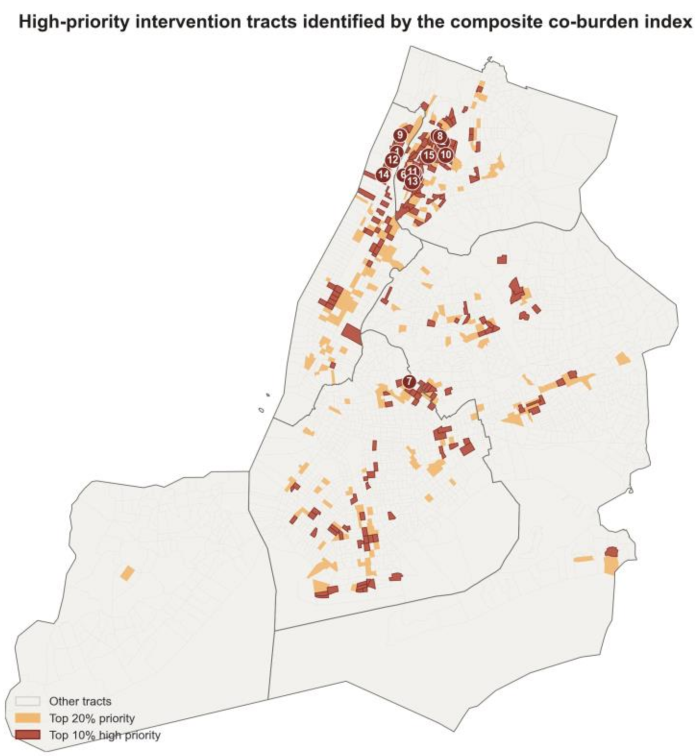
The contribution of Section 8 lies in translating these dispersed statistical and mechanistic findings into a spatial decision-making framework.

From a planning and policy perspective, high-priority tracts should not be interpreted simply as areas requiring more trees. Rather, they should be understood as zones of overlapping pressures requiring differentiated and integrated interventions.

For tracts characterized by both high built-environment intensity and low street-tree quality, intervention priorities should focus on tree-pit expansion, structural soils, permeable pavement, root-zone improvement, and street-section redesign, rather than simple replacement planting alone.

For tracts where high social vulnerability overlaps with low street-tree quality, urban forestry investments should be integrated with heat-risk reduction, public-health objectives, and environmental justice priorities, with maintenance resources and long-term management funding preferentially allocated to these communities.

For tracts with particularly high landscape fragmentation burden, interventions should prioritize improving canopy connectivity, establishing continuous shade corridors, and reducing isolated or edge-dominated canopy structures.



Top ranked tracts

Rank	Tract	Borough	Co-burden score
1	CT 269	Manhattan	0.863
2	CT 385	Bronx	0.846
3	CT 237.02	Bronx	0.844
4	CT 179.01	Bronx	0.839
5	CT 221.02	Bronx	0.839
6	CT 199	Bronx	0.834
7	CT 447	Brooklyn	0.832
8	CT 399.02	Bronx	0.832
9	CT 285	Manhattan	0.826
10	CT 373	Bronx	0.826
11	CT 221.01	Bronx	0.826
12	CT 261	Manhattan	0.823
13	CT 181.02	Bronx	0.821
14	CT 245	Manhattan	0.814
15	CT 233.02	Bronx	0.814

Figure 14. High-priority intervention tracts identified by the composite co-burden index. The map emphasizes the geography of the top 20% and top 10% tracts, while the ranked list supports tract-level planning and field verification.

9. Interactive Decision-Support Map and Policy

9.1 NYC Urban Forestry Context

New York City has a long history of treating street trees as part of its public urban infrastructure rather than as isolated landscape elements. NYC Department of Parks and Recreation is responsible for the planting and care of park and street trees in the public right-of-way, including routine pruning, maintenance, and street-tree planting programs. As a result, street-tree conditions are shaped not only by ecological factors, but also by public investment, maintenance capacity, and long-term governance systems.

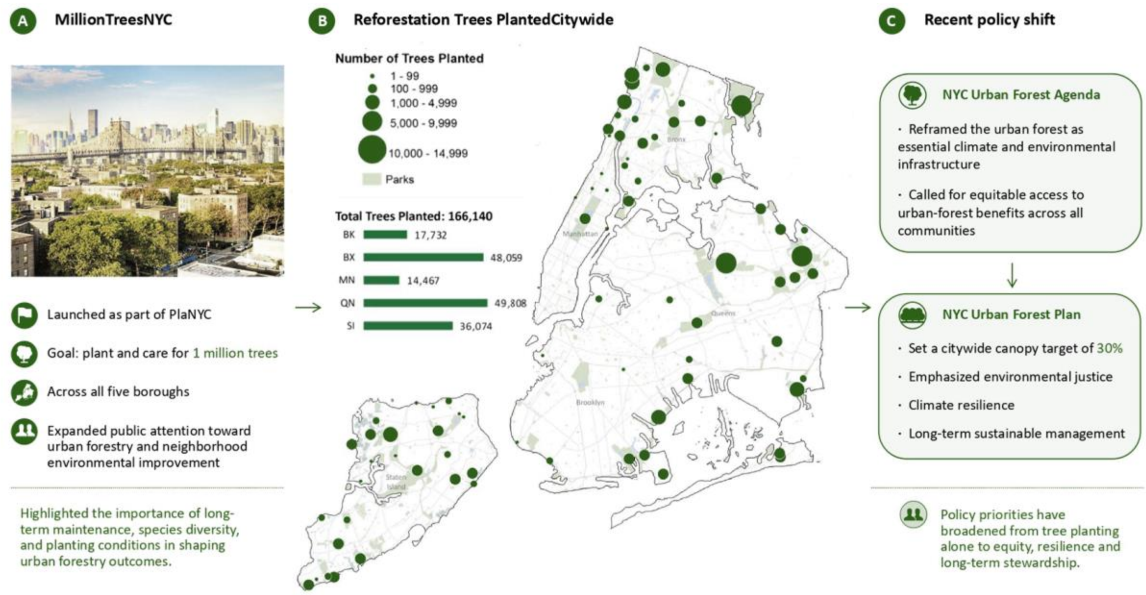


Figure 15. Policy evolution of urban forestry in New York City.

A major precedent is MillionTreesNYC, launched in 2007 as part of PlaNYC. The initiative aimed to plant and care for one million trees across the five boroughs and significantly expanded public attention toward urban forestry, climate resilience, and neighborhood environmental quality (NYC Parks, n.d.). The program also highlighted the importance of long-term maintenance, species selection, and planting distribution in shaping urban forestry outcomes across neighborhoods (Campbell et al., 2014).

More recent policy frameworks have shifted from simple canopy expansion toward broader questions of urban forest equity and resilience. NYC Urban Forest Agenda frames the urban forest as critical climate and environmental infrastructure and calls for equitable access to urban forest benefits across all communities (City of New York, 2021). The related NYC Urban Forest Plan further proposes expanding citywide canopy cover toward 30% while emphasizing environmental justice, climate adaptation, and long-term stewardship (City of New York, 2023).

This policy context directly aligns with the objectives of this study. Rather than asking only where trees exist, the report evaluates where low street-tree quality, social vulnerability, built-environment pressure, and landscape fragmentation overlap spatially. Therefore, the co-burden framework developed in this study can function as a spatial decision-support tool for future urban forestry planning, helping prioritize neighborhoods where ecological disadvantage and social vulnerability are simultaneously concentrated.

9.2 Interactive Map as Planning Interface

The interactive map is not only a technical deliverable, but also a spatial decision-support interface designed to assist urban forestry planning and environmental justice screening. The platform allows users to switch dynamically between single-burden indicators, the composite co-burden index, and LISA cluster outputs, enabling comparison between different dimensions of street-tree ecological inequality.

Built with Deck.gl and MapLibre, the interface supports analysis across multiple spatial scales, from citywide screening and borough-level comparison to tract-level inspection and field verification. Rather than functioning as a static visualization, the map operates as a planning and governance tool that can help identify priority intervention areas, support maintenance prioritization, inform planting prioritization, and guide long-term urban forestry investment.

9.3 Priority Intervention Typologies

The following intervention typologies are synthesized from the co-burden framework, spatial clustering results, and the overlap analysis of ecological, social, and built-environment indicators presented in previous chapters.

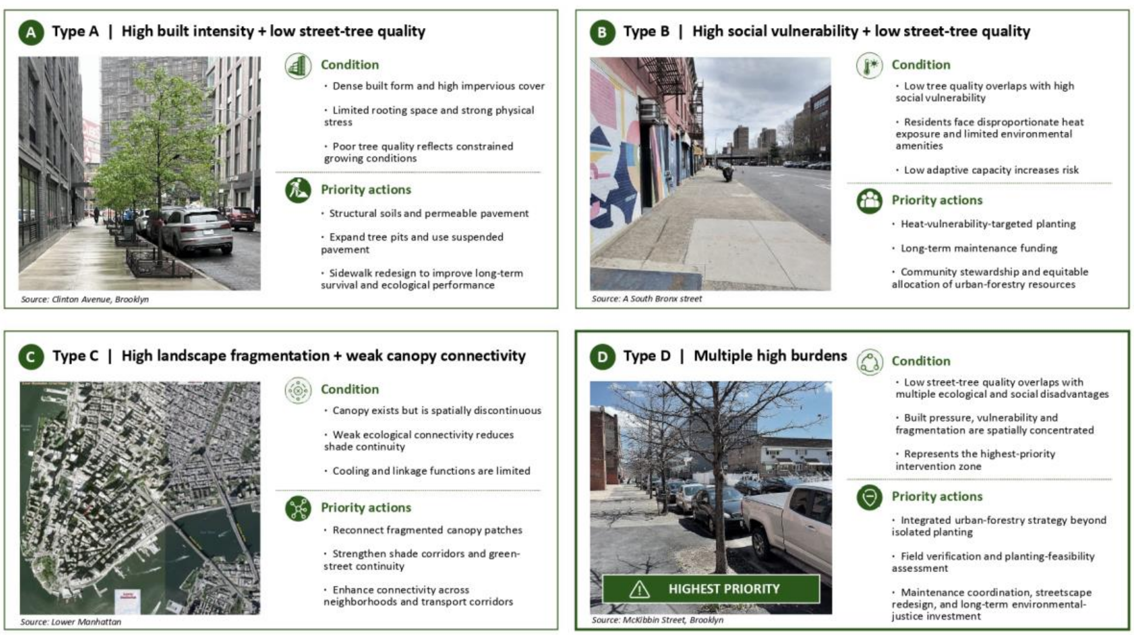


Figure 16. A conceptual framework for place-based urban forestry action in New York City.

Type A: High Built Intensity + Low Street-Tree Quality

Type A tracts are characterized by dense built environments, high impervious surface coverage, and limited rooting space, all of which create significant physical stress for street trees. In these areas, poor tree health is less likely to result from insufficient planting alone and more likely to reflect constrained growing conditions associated with compact urban form. Therefore, interventions should prioritize structural soil retrofits, permeable pavement systems, tree pit expansion, suspended pavement strategies, and sidewalk redesign rather than simple replacement planting. The goal is not only to increase tree numbers, but also to improve long-term survival and ecological performance within high-density streetscapes.

Type B: High Social Vulnerability + Low Street-Tree Quality

Type B tracts combine ecological disadvantage with elevated social vulnerability, indicating neighborhoods where residents may face disproportionate exposure to heat stress, limited environmental amenities, and reduced adaptive capacity. In these areas, street-tree intervention should be framed not only as an ecological improvement strategy, but also as an environmental justice and public-health investment. Priority actions may include heat-vulnerability-targeted planting programs, long-term maintenance funding, community stewardship initiatives, and equitable allocation of urban forestry resources.

Type C: High Landscape Fragmentation + Weak Canopy Connectivity

Type C tracts are characterized by fragmented canopy structure and weak ecological connectivity. Although some tree canopy may exist, the spatial organization of green infrastructure is highly discontinuous, reducing shade continuity, ecological linkage, and cooling performance. Interventions in these areas should focus on reconnecting fragmented canopy patches, strengthening shade corridors, improving green street continuity, and enhancing landscape connectivity across neighborhoods and transportation corridors.

Type D: Multiple High Burdens

Type D tracts experience the simultaneous overlap of low street-tree quality, high social vulnerability, intense built-environment pressure, and fragmented canopy structure. These areas represent the most critical priority zones within the co-burden framework because multiple ecological and social disadvantages are spatially concentrated. As a result, interventions should move beyond isolated planting projects and instead adopt integrated urban forestry strategies that combine field verification, planting feasibility assessment, maintenance coordination, streetscape redesign, and long-term environmental justice-oriented investment.

9.4 Spatial Justice and Priority Areas

The spatial distribution of the composite co-burden index suggests that street-tree ecological inequality in New York City is closely connected to broader patterns of environmental injustice, urban morphology, and long-term governance conditions. High-priority tracts are not randomly distributed, but instead form concentrated clusters in areas such as the South Bronx, northern Manhattan, and parts of central Brooklyn.

9.4.1 South Bronx: Environmental Justice and Concentrated Co-burden

The South Bronx contains some of the highest co-burden scores identified in this study. In these neighborhoods, high social vulnerability overlaps with fragmented canopy structure, intense built-environment pressure, and lower street-tree ecological quality. The area has long been associated with environmental justice concerns, including elevated heat exposure, poor air quality, and limited green infrastructure access.

The results suggest that street-tree inequality in the South Bronx reflects broader patterns of environmental and infrastructural inequality. Rather than representing isolated ecological problems, these burdens form continuous spatial clusters within historically marginalized communities.

9.4.2 Lower East Side: Built Intensity and Morphological Constraints

A different pattern appears in the Lower East Side and other dense parts of Manhattan. Although these areas may not exhibit the city's highest social vulnerability, they often show lower street-tree ecological quality associated with extreme built intensity, high impervious surface coverage, and constrained rooting conditions.

This finding suggests that street-tree inequality is shaped not only by socioeconomic conditions, but also by urban morphology itself. Dense street canyons, narrow sidewalks, and limited soil volume create physical constraints that reduce long-term tree health and ecological performance.

9.4.3 Beyond MillionTreesNYC: From Canopy Expansion to Landscape Justice

The findings also raise important questions regarding the long-term legacy of MillionTreesNYC. While the initiative significantly expanded tree planting across New York City and elevated public attention toward urban forestry, the co-burden results suggest that increasing planting quantity alone is insufficient for achieving equitable ecological outcomes.

The results indicate that street-tree inequality is shaped not only by the number of trees planted, but also by where trees are planted, the environmental conditions in which they grow, and whether long-term maintenance and ecological benefits are distributed equitably across neighborhoods. In many high-burden areas, intense built-environment pressure, fragmented canopy structure, and high social vulnerability continue to constrain long-term ecological performance even where planting efforts exist.

This suggests that the central challenge of urban forestry is not simply canopy expansion, but the uneven spatial distribution of ecological infrastructure and environmental benefits. In this sense, the findings support a broader shift from quantity-oriented planting strategies toward a landscape justice framework that emphasizes long-term stewardship, spatial equity, environmental quality, and differentiated investment in historically underserved communities.

9.5 Limitations and Future Directions

Several limitations remain important. First, this study is cross-sectional and integrates datasets collected between 2015 and 2020 rather than a single synchronized year. As a result, the analysis captures broad spatial relationships rather than temporal ecological change, making it difficult to evaluate long-term street-tree dynamics or post-planting trajectories following initiatives such as MillionTreesNYC. In addition, the analysis is conducted at the Census Tract level and is therefore subject to the modifiable areal unit problem (MAUP). Street-tree ecological processes are often highly localized, while tract-level aggregation may smooth important street-scale variation in canopy condition, maintenance environments, and planting structure.

Second, several key determinants of street-tree health remain unobserved within the current analytical framework. Variables such as soil quality, rooting volume, maintenance schedules, irrigation conditions, pruning frequency, pest pressure, and street-level microclimatic stress were not available in the tract-level datasets used in this study. Future research should therefore integrate field surveys, maintenance records, and street-level environmental measurements to better understand the micro-scale processes shaping long-term urban forest health.

9.6 Conclusion

This study demonstrates that street-tree ecological inequality in New York City is multi-dimensional and spatially uneven. Street Tree Density and Species Diversity are relatively well explained by tract-level land cover, built-environment intensity, and landscape configuration variables, while Tree

Health Index is less effectively explained by these factors. These results indicate that different street-tree ecological indicators respond differently to tract-level urban form and spatial structure.

The findings further suggest that street-tree ecological inequality is associated with broader patterns of social vulnerability and uneven environmental conditions across the city. High-priority areas identified through the co-burden analysis are concentrated in parts of the Bronx, northern Manhattan, and central Brooklyn, where multiple ecological and social burdens overlap.

Finally, the co-burden framework and interactive decision-support map developed in this study provide a planning-oriented approach for identifying priority intervention areas and supporting more equitable urban forestry strategies. Future interventions should therefore focus not only on increasing tree quantity, but also on improving long-term ecological quality and spatial equity.

References

Berland, A., Shiflett, S. A., Shuster, W. D., Gall, A. M., Widder, C. R., & Herrmann, D. L. (2017). The role of trees in urban stormwater management. Landscape and Urban Planning, 162, 167-177. https://doi.org/10.1016/j.landurbplan.2017.02.017

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A social vulnerability index for disaster management. Journal of Homeland Security and Emergency Management, 8(1).

Galle, N. J., Halpern, D., Nitoslawski, S. A., Duarte, F., Ratti, C., & Pilla, F. (2021). Fostering urban forests: A systematic review of expert perceptions, policy instruments, and governance frameworks. Urban Forestry and Urban Greening, 62, 127165. https://doi.org/10.1016/j.ufug.2021.127165

Hand, D. J. (2020). Dark data: Why what you don't know matters. Princeton University Press.

Schwarz, K., Fragkias, M., Boone, C. G., et al. (2015). Trees grow on money: Urban tree canopy cover and environmental justice. PLOS ONE, 10(4), e0122051. https://doi.org/10.1371/journal.pone.0122051

Wolf, K. L., Lam, S. T., McKeen, J. K., Richardson, G. R. A., van den Bosch, M., & Bardekjian, A. C. (2020). Urban trees and human health: A scoping review. International Journal of Environmental Research and Public Health, 17(12), 4371. https://doi.org/10.3390/ijerph17124371

Zhang, Y., Chen, G., Myint, S. W., Zhou, Y., Hay, G. J., Vukomanovic, J., & Meentemeyer, R. K. (2022). UrbanWatch: A 1-meter resolution land use and land cover dataset for 22 major cities in the United States. Remote Sensing of Environment, 278, 113106. https://doi.org/10.1016/j.rse.2022.113106

Campbell, L. K., Monaco, M., Falxa-Raymond, N., Lu, J., Newman, A., Rae, R. A., & Svendsen, E. S. (2014). MillionTreesNYC: The integration of research and practice. New York, NY: New York City Department of Parks and Recreation.

City of New York. (2021). NYC Urban Forest Agenda. City of New York.

City of New York. (2023). NYC Urban Forest Plan. City of New York.

NYC Parks. (n.d.). MillionTreesNYC. NYC Department of Parks and Recreation.