A spatial analysis approach to characterizing commercial and industrial building GHG emissions across minority and low-income communities – developed by Yasmin Romitti as part of an internship with EGE [2](last updated 11/30/21)

For questions, contact Yasmin at yromitti@bu.edu

Methodology

The following provides a methodological overview of a spatial analysis approach that can be used to ascertain opportunities and potential policy recommendations for advancing energy efficiency and clean energy improvements for minority businesses across the United States, in addition to characterizing the commercial and industrial greenhouse gas (GHG) reduction potential in minority and low-income communities. This approach was developed as part of a project exploring the mitigation opportunities and policies needed to advance energy efficiency and clean energy improvements in minority-owned non-residential buildings in across six case study cities (Chicago, IL; New York, NY; Los Angeles and San Francisco, CA; Miami and Orlando, FL). We make use of several publicly available datasets that include building energy performance measures, greenhouse gas inventories, minority business certification directories, and fine resolution community socioeconomic and demographic information. In the following documentation, Los Angeles, CA and New York, NY are used to illustrate the description of what is a two-part methodology that primarily relies on the spatial joining of building benchmarking data with additional datasets in order to identify current GHG emissions and other characteristics of buildings and industries for minority-owned businesses, low-income communities, and communities of color in these two cities.

I. Data

Building Benchmarking

The key data used for this analysis are publicly available municipal building benchmarking data (as it pertains to building energy use, benchmarking is a method of measuring and tracking building energy performance over time).[1] Cities and municipalities that have implemented benchmarking measures do so through the mechanism of a benchmarking ordinance that is typically part of a broader energy (and water use) efficiency strategy. While the individual details and conditions of benchmarking ordinances vary across cities and municipalities, the main feature is the requirement of certain building property types (e.g. commercial, institutional, residential) of a designated threshold size to report their energy performance to the city on an annual basis (or some other designated interval). Details regarding the specific ordinance in place in both Los Angeles and New York and associated benchmarking data sources are included in Table 1 below. We obtain building benchmarking data for each city for data year 2019 and filter the primary property type to exclude residential properties (e.g. senior living/care community, other - lodging/residential, multifamily housing, and residence hall/dormitory). Benchmarking data was also filtered to remove missing geographic coordinate information.

Table 1 Building benchmarking data ordinance and data source information

City	Highlighted ordinance details	Data source
Los Angeles, CA	"The benchmarking requirements of the EBEWE Program for Benchmark Year 2018 and beyond apply to buildings in the City of Los Angeles that meet one of the following criteria: 1. City-owned buildings with a gross floor area of 7,500 square feet or more; or 2.	[3]

[1] https://www.energy.gov/eere/slsc/building-energy-use-benchmarking
[3] https://data.lacity.org/City-Infrastructure-Service-Requests/Existing-Buildings-Energy-Water-Efficiency-EBEWE-P/9yda-i4ya

	Privately owned building or a building owned by a local agency of the State with a gross floor area of 20,000 square feet or more."[2]	
New York, NY	"COVERED BUILDING. As it appears in the records of the department of finance (i) a building that exceeds [50,000] 25,000 gross square feet, (ii) two or more buildings on the same tax lot that together exceed 100,000 gross square feet, [or] (iii) two or more buildings held in the condominium form of ownership that are governed by the same board of managers and that together exceed 100,000 gross square feet, or (iv) a city building. City building: A building that is more than 10,000 gross square feet, as it appears in the records of the department of finance, that is owned by the city or for which the city regularly directly pays all [or part] of the annual energy bills, provided that two or more buildings on the same tax lot shall be deemed to be one building."[4]	[5]

Large Facility GHG Emissions

The U.S. Environmental Protection Agency (EPA) collects information about GHG emissions in the U.S. through both the Inventory of U.S. Greenhouse Gas Emissions and Sinks and the Greenhouse Gas Reporting Program (GHGRP). The GHGRP in particular covers large emission sources (defined as sources emitting greater than 25,000 $MTCO_2e$ per year) analogous to around 8000 facilities across various sectors.[6] We obtain information on emissions from large facilities using the EPA's Facility Level Information on GreenHouse gases Tool (FLIGHT).[7] 2019 FLIGHT data are downloaded for the metropolitan area corresponding to each city and geospatially subset to city boundaries using R software's sf (simple features) package, which supports encoding of spatial vector data.[8]

LEED Buildings

LEED certification provides a measure of the planning, design, construction, and operations of buildings with the consideration of energy use, water use, indoor environmental quality, materials, and the building's environmental impact. Together, these elements provide the basis for different credit categories within the LEED rating system.[9] We utilize information on LEED certification status as an indicator of building energy efficiency and implied opportunities for further energy management. LEED project directories were downloaded from the US Green Building Council for California and New York state and manually filtered for the city of interest.[10] Building entries for each directory were subsequently geocoded using R software's ggmap package[11] and Google's geocoding API[12]. Entries where geocoding was unsuccessful, and therefore had missing geographic coordinates, were removed.

[2] https://www.ladbs.org/docs/default-source/forms/EBEWEComplianceInstructions2020.pdf
[4] https://www1.nyc.gov/site/buildings/codes/benchmarking.page
[5] https://data.cityofnewyork.us/Environment/Energy-and-Water-Data-Disclosure-for-Local-Law-84-/qb3v-bbre
[6] https://www.epa.gov/ghgreporting
[7] https://ghgdata.epa.gov/ghgp/main.do
[8] https://cran.r-project.org/web/packages/sf/sf.pdf
[9] https://www.usgbc.org/help/what-green-building
[10] https://www.usgbc.org/projects
[11] https://cran.r-project.org/web/packages/ggmap/ggmap.pdf
[12] https://developers.google.com/maps/documentation/geocoding/overview

Minority Business Enterprises (MBEs)

MBE certification is an enterprise level ownership and diversity certification available to any minority-owned business in the U.S. Qualification as a minority-owned business requires that a business bye physically located in the U.S. or U.S.-owned territories and be owned, operated, and controlled by minority group members.[13] We gather directories of MBEs for both Los Angeles and New York by downloading publicly available certified business directories from the state, county, and/or municipality, as available in each area. Directories are manually filtered to search for MBEs in the city of interest. Details of the directory data source(s) are shown in Table 2 below. MBE entries for each city were geocoded using R software's ggmap package and Google's geocoding API and entries where geocoding was unsuccessful, and therefore had missing geographic coordinates, were removed.

Table 2 MBE directory data source information

City	**Data source(s)**
Los Angeles, CA	State[14]
New York, NY	City[15], State[16]

Demography and Income

Demographic and income variables for 'Hispanic or Latino Origin by Race' and 'Median Household Income in the Past 12 Months (in 2010 Inflation-Adjusted Dollars)' were extracted from the 2019 5-year American Community Survey (ACS)[17] for census tracts in each study city using R's tidycensus package. Minority majority tracts were identified as those with > 50% non-white Hispanic or African American race.[18] Following the US Department of Housing and Urban Development (HUD) definition of low income[19], low income tracts were identified as tracts with a median area income less than or equal to 80% of the median family income for the study city.

City Geography

The legal geographic boundaries for both Los Angeles and New York, and therefore the associated boundaries used for our spatial data analysis, were obtained from the U.S. Census Bureau TIGRIS/line shapefiles for 'Census Designated Places'.[20] [21] Although low-income communities and communities of color may be located across the metropolitan area of a city in its surrounding towns and cities, CDP

---

[13] https://nmsdc.org/mbes/what-is-an-mbe/
[14] https://californiaucp.dbesystem.com/
[15] https://data.cityofnewyork.us/Business/M-WBE-LBE-and-EBE-Certified-Business-List/ci93-uc8s
[16] https://www.osc.state.ny.us/state-vendors/resources/minority-and-women-owned-business-enterprises-mwbes
[17] https://www.census.gov/programs-surveys/acs
[18] This definition of minority focuses exclusively on black and brown census tracts according to the statement of work between the Institute for Sustainable Communities, Institute for Market Transformation, and EGE[2] Empowering a Green Environment and Economy. The same ACS variables can be used to define racial minority as needed by the research analysis.
[19] https://www.huduser.gov/portal/datasets/il/fmr98/sect8.html
[20] https://www.census.gov/content/dam/Census/data/developers/understandingplace.pdf
[21] https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

boundaries were selected for this analysis in order to most closely align with the data provided by municipal benchmarking data.

II. Methods

This spatial analysis approach makes use of a two-part methodology in order to identify 1) the current GHG emissions of buildings and large facilities in each study city 2) whether a MBE is located in these buildings 3) the community context of these buildings: low-income or minority majority, and 4) the LEED status of these buildings.

Part I

We use building benchmarking data to identify which buildings are also likely MBE locations, as well as the GHG emissions context of these buildings. Due to the lack of granular information on minority owned buildings, we use MBEs as a loose proxy to identify potential GHG emissions reductions related to minority owned buildings. We spatially merge city building benchmarking data with geocoded MBE addresses using the geojoin function from R's fuzzyjoin package.[22] Geojoin performs a spatial join based upon distance combinations of longitudes and latitudes. Distance is computed using the haversine method and data sets are merged using a maximum join distance parameter of 0.05 miles. This maximum joining distance can be chosen based upon the median building size (square footage) of buildings in the benchmarking data set for each city. The merged data are filtered for duplicate building ID and/or MBEs (to avoid double-counting of emissions and/or MBE information). The merged MBE and building benchmarking data are then spatially linked to identify the LEED status of buildings. Geocoded LEED data are joined to this new merged MBE/building data using the same geojoin function and parameters. We describe the overall energy performance of buildings where MBEs are located by performing summary statistics on variables—e.g. benchmarking reporting compliance, total GHG emissions (metric tons CO_2e), Energy Star rating, and LEED status.

Part II

Next, we use demographic and socioeconomic information in conjunction with building benchmarking data to characterize the GHG emissions from buildings and large facilities across low-income and majority minority census tracts in both study cities. Using American Community Survey demographic and income variables (described in "Data – Demography and Income" above) census tracts are categorized as either 'low income' or 'all other income' and 'majority minority' or 'non-minority majority'. We then spatially merge this information with previously geojoined LEED/city building benchmarking data to identify the demography and income level of each building's location. The overall energy performance of buildings located in low income and majority minority census tracts is described by performing identical summary statistics as above ("Method I"). In addition, we characterize the GHG emissions from large facilities obtained from the EPA FLIGHT database across these disadvantaged communities by identifying the census tract designation for each large facility's location. We summarize the energy performance of large facilities in each city by performing summary statistics on the following variables: sector (e.g. metals, power plant, chemicals) and total emissions (metric tons CO_2e).

[22] https://cran.r-project.org/web/packages/fuzzyjoin/fuzzyjoin.pdf

A spatial analysis approach to characterizing commercial and industrial building GHG emissions across minority and low-income communities

To provide further context on the most relevant areas and sectors of interest for GHG emissions reductions opportunity and policy recommendations in Los Angeles and New York, we identify the top 10 emitting buildings and the top 5 emitting large facility in each city and their associated property types, size, GHG emissions quantity, Energy Star rating, and demographic/income information. We additionally visually identify 1-2 dense 'clusters' of buildings in each city for which similar information on density, benchmarking compliance, primary property type, total GHG emissions, Energy Star rating, and LEED status is extracted. All analyses and mapping are done in R version 4.0.2.

Appendix – Example Figures

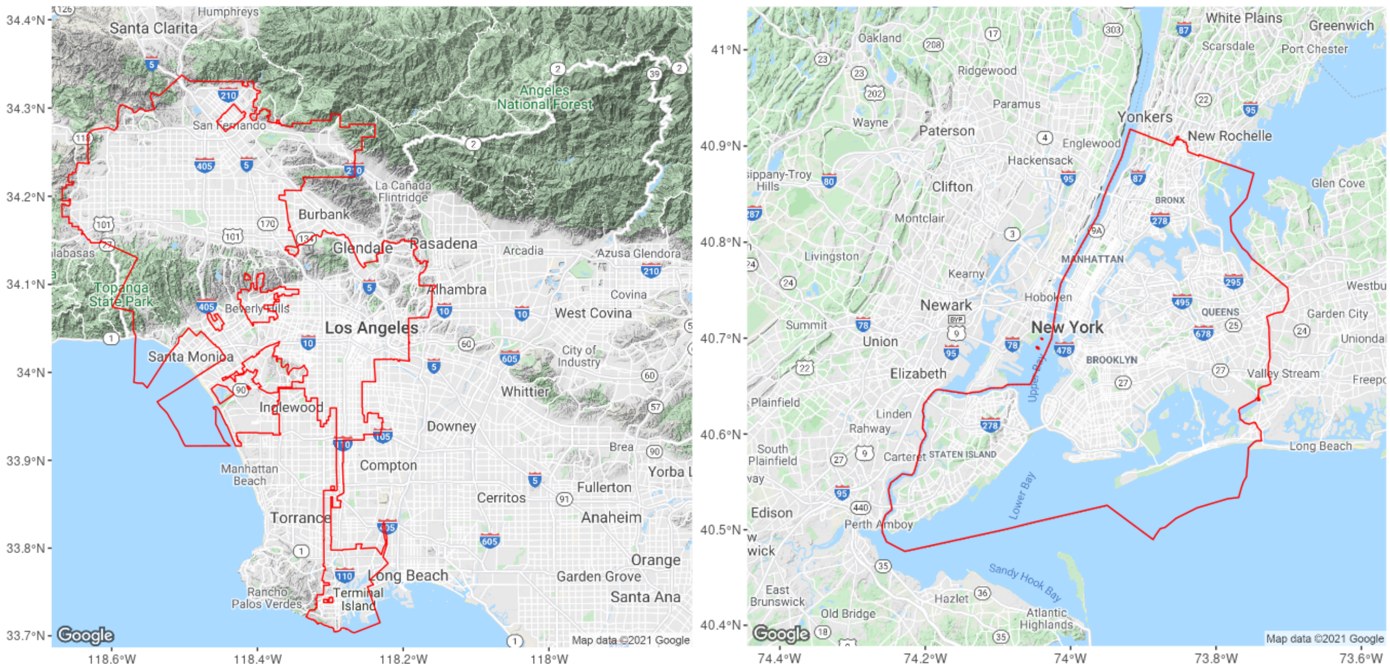

Figure 1 U.S. Census Bureau 'Census Designated Places' geographic boundaries for Los Angeles, CA and New York, New York

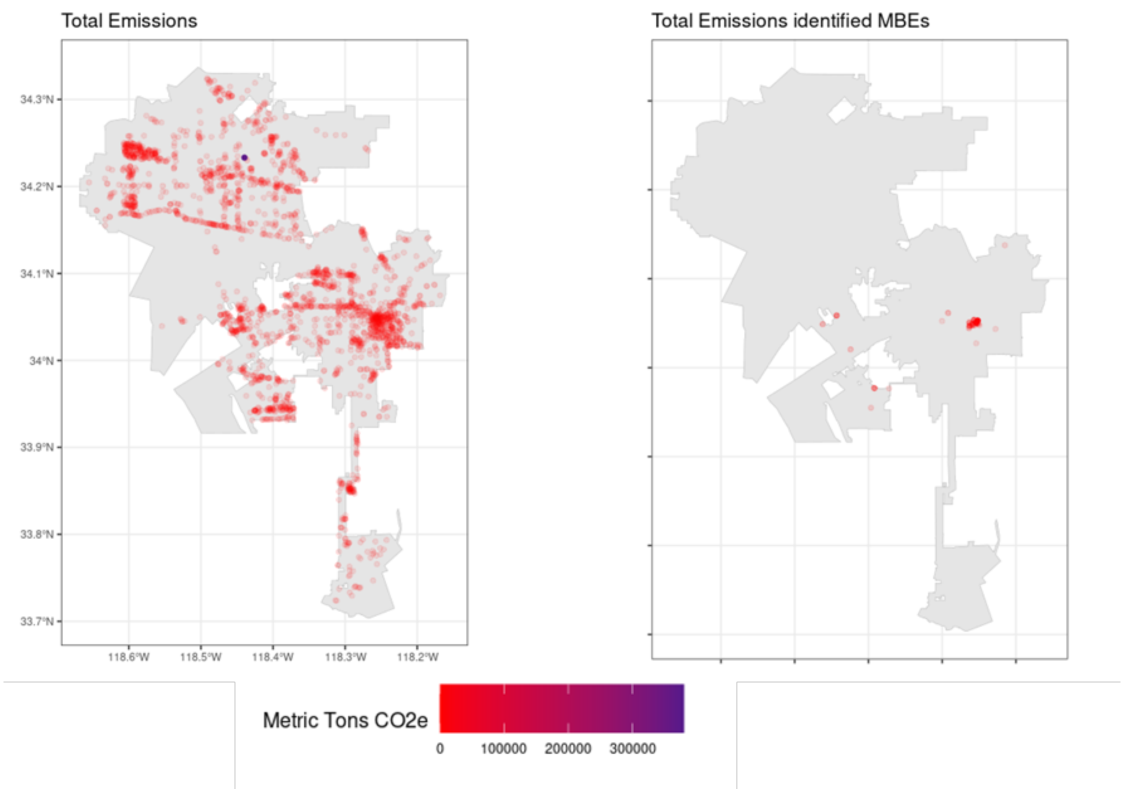

Figure 2 Los Angeles total emissions (metric tons CO2e) all buildings (left) and all buildings identified as MBE locations (right)

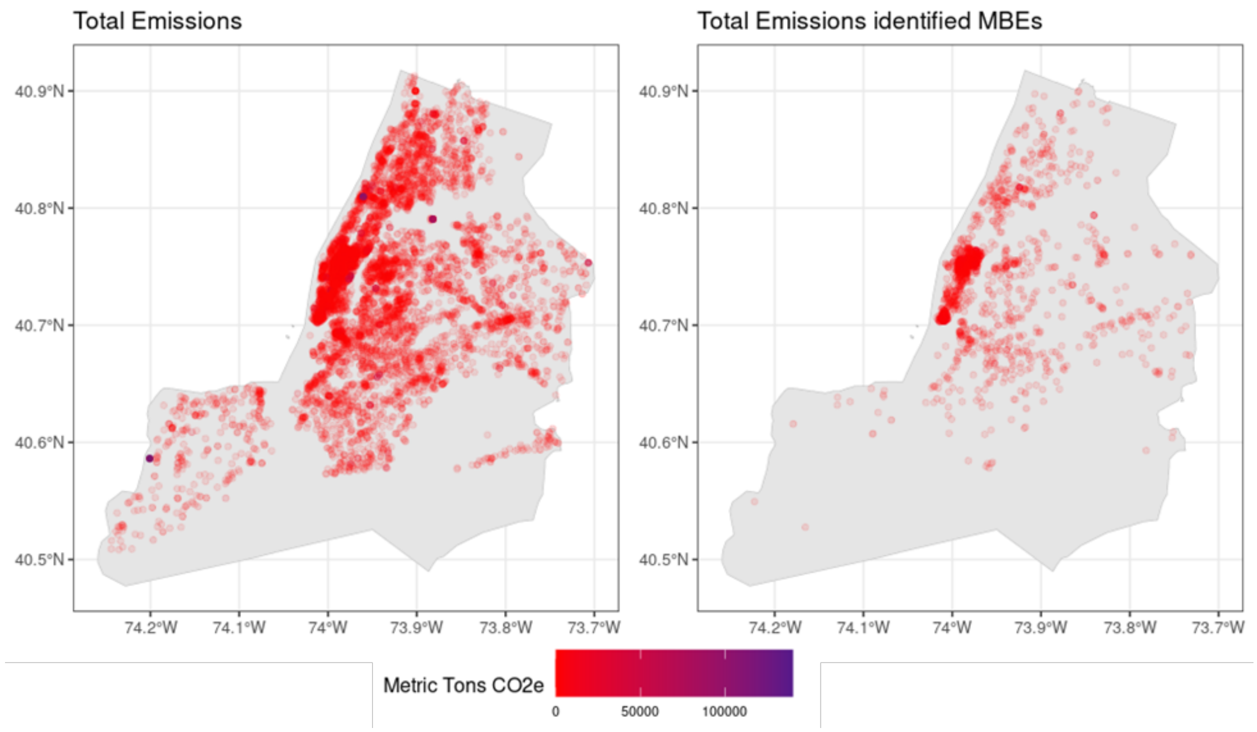

Figure 3 New York total emissions (metric tons CO2e) all buildings (left) and all buildings identified as MBE locations (right)

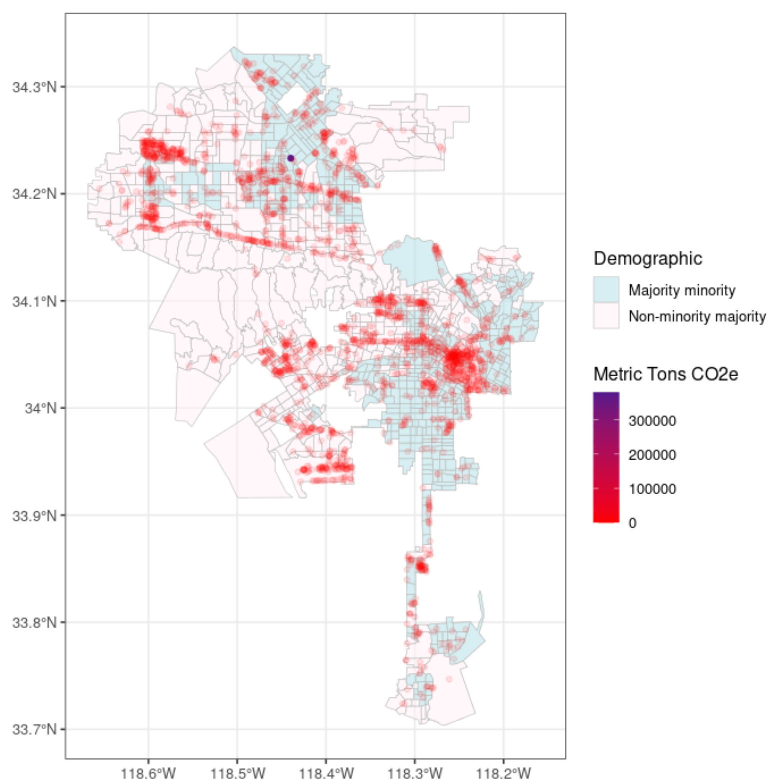

Figure 4 Los Angeles total emissions (metric tons CO2e) all buildings stratified by demographic

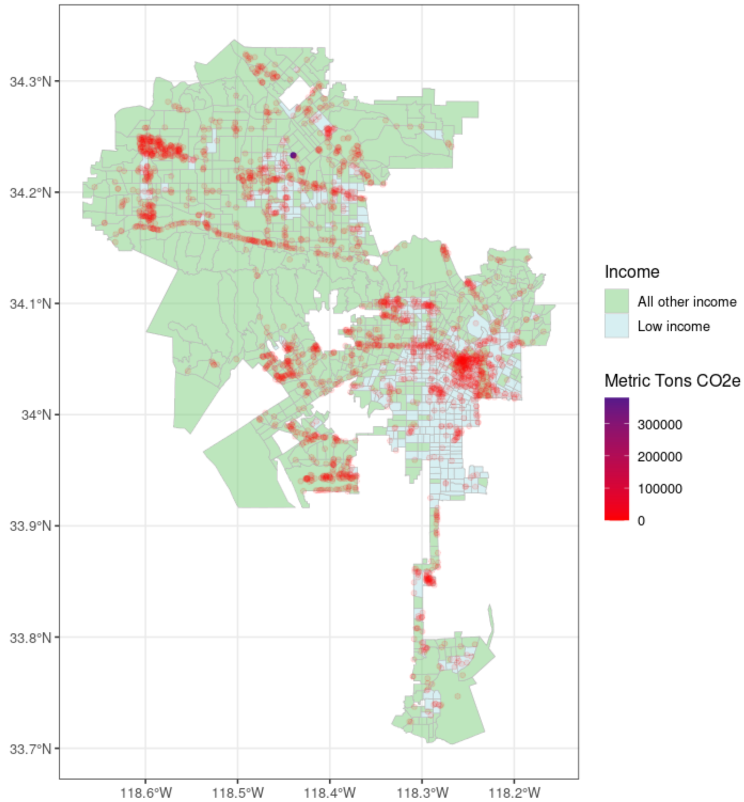

Figure 5 Los Angeles total emissions (metric tons CO2e) all buildings stratified by income

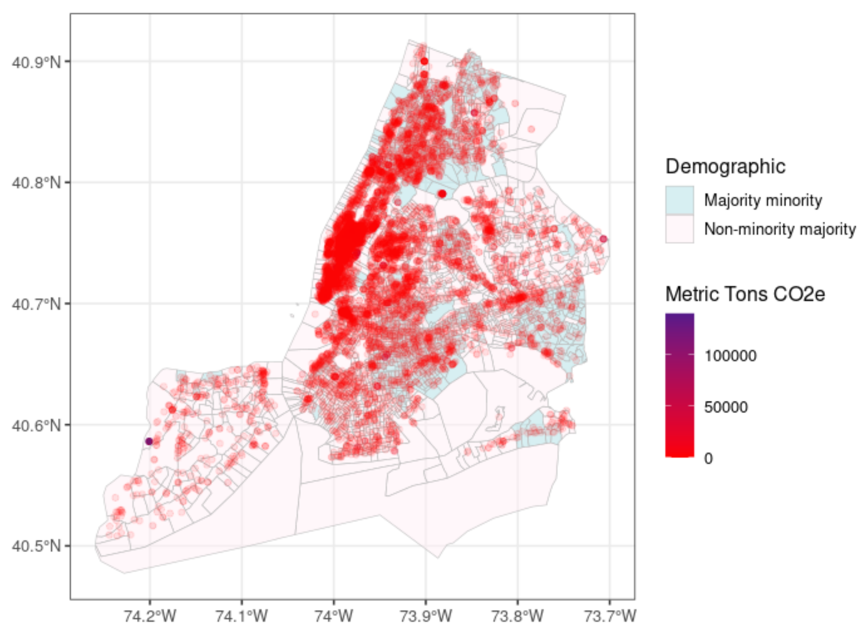

Figure 6 New York total emissions (metric tons CO2e) all buildings stratified by demographic

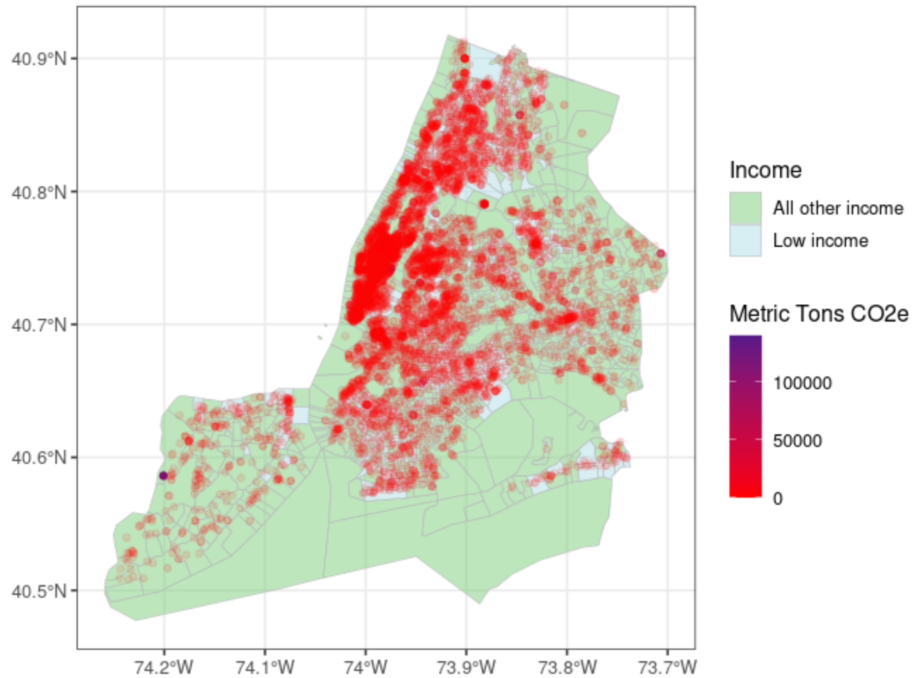

Figure 7 New York total emissions (metric tons CO2e) all buildings stratified by income