Equity of Public Transit Access to Hiking Trails in San Francisco, CA

Will McAllister and Ian Beilfuss

Geography 578 Applications of GIS
12/11/2025

Research Statement

The San Francisco Bay Area is known for its hiking trails, but public transportation inequities are a potential barrier for many communities to access them. We seek to evaluate access to unpaved hiking trails in the San Francisco Bay Area by creating a distance matrix of the duration of travel by public transportation from census block group centroids in the City of San Francisco to trailheads. We will then overlay demographic data to understand how public transit access to trails varies across different socioeconomic and racial groups.

Table of Contents

Introduction

The importance of easy access to the outdoors has become an essential element in urban planning conversations. Spending time outside has been proven to improve mental and physical health. Initiatives like the Transit to Trails Act, introduced by California Representative Gomez to the U.S. Congress in 2021, highlight the importance of equity in nature accessibility. While cities like San Francisco have made strides in reducing walking time to city parks, more research is needed into the accessibility of larger natural areas via public transit, especially for underserved communities. In this paper, we propose to evaluate the following statement: The San Francisco Bay Area is known for its hiking trails, but public transportation inequities are a potential barrier for many communities to access them. We seek to evaluate access to unpaved hiking trails in the San Francisco Bay Area by creating a distance matrix of the duration of travel by public transportation from census block group centroids in the City of San Francisco to trailheads. We will then overlay demographic data to understand how public transit access to trails varies across different socioeconomic and racial groups.

Background

San Francisco is located on the San Francisco Peninsula in central western California. The larger metropolitan area, referred to as the Bay Area, is composed of the nine counties surrounding San Francisco, which is the site of our analysis[1].

[1] See Appendix Figure 4 for map of study area

Home to 873,965 residents, San Francisco is the fourth-largest city in California and the 17th-largest city in the US (“San Francisco”, 2020). However, primarily due to rising housing prices, San Francisco's population has declined since 2020; in contrast, other major California cities have seen population increases (Bandlamudi, 2025). Racially and ethnically, San Francisco is 39% White, 34% Asian, 16% Hispanic, and 5% Black, according to the 2020 Census. Thanks to its proximity to Silicon Valley, San Francisco is one of the wealthiest cities in the US, with a 2023 median household income of $141,446 and a median property value of $1.38 million (“San Francisco, CA”, n.d.). High household incomes and property values lead to a very high cost of living. In 2018, according to the federal poverty threshold, 10% of the San Francisco population was living in poverty, however, the actual number is likely much higher, as the poverty line is fixed at a national level. In San Francisco, the low-income limit in 2018, according to the Department of Housing and Urban Development, was $82,200 for an individual and $117,400 for a family of four. These figures are more than 4.6 and 6.8 times higher than the national low-income limit, respectively. Additionally, Black and female residents experienced higher-than-average rates of poverty (“City performance scorecards”, n.d.).

According to a 2019 study by the San Francisco Municipal Transportation Agency (SFMTA), 34.5% of households don’t have access to a personal vehicle. This rate increases to 53% for households with an annual income of less than $100,000. By race/ethnicity, 61% black residents have no vehicle, 32% of Asians, and 35% of Hispanics. This is likely due to the high cost of vehicle ownership in California. California drivers spend an average of $1,199 on car ownership expenses per month (Micalizio, 2025). Instead, many San Francisco residents are priced out of car ownership, instead opting to use the public transit network. According to the SFMTA in

2019, residents making less than $100,000 per year use public transit for 47.5% of their commutes. A Muni pass (access to all public transportation, including Bay Area Rapid Transit (BART), within San Francisco) costs $104 per month for most residents, and $43 a month for low-income residents. While lacking the flexibility provided by car transportation, SFMTA and BART offer a much cheaper option for lower-income residents who cannot afford a car.

Spending time in natural environments, such as going for a hike, has been shown to have numerous health benefits. According to a paper by White (et al., 2019), the likelihood of reporting good health or high well-being was significantly higher for people who spent 120 minutes or more in nature per week. Hiking benefits physical health through improving cardiovascular health, building stronger muscles, and decreasing the risk of certain respiratory problems. Additionally, a Stanford study found that hiking for just 90 minutes in natural environments can reduce stress, anxiety, and depression (Bratman et al., 2015). However, this begs the question of whether hiking trails are accessible to those without a car, especially for underserved communities, due to their relative distance from urban areas.

The Bay Area is home to more than 3,000 registered hiking trails on AllTrails ("AllTrails", 2025). This is largely due to California's numerous local, regional, and national natural areas. More than 25% of the roughly 4.5 million acres of total land in the Bay Area is held in public trust (Breting-Garcia, 2022). This helped San Francisco to become the first city in the United States where all residents live within a ten-minute walk to a park ("Trust for public land", 2017).

In 2021, California Representative Gomez introduced the Transit to Trails Act to the US Congress. The Act would work to combat the issue of public land inaccessibility for underserved communities through investments in transportation.

> *"Getting out in nature is essential for our mental health and physical wellbeing, especially for our children. But many cities are built in a way that makes America's national parks and public lands inaccessible for underserved communities. Everyone deserves access to green spaces, not just those who can afford to drive to them" ("S.1440 - 118th Congress (2023-2024)", 2023).*

San Francisco already meets some of the Act's goals by having green spaces within a ten-minute walk of all residents; however, ease and equity of access from San Francisco to public lands with longer hiking trails remain largely unresearched. In this paper, we will outline our methodology for constructing our GIS analysis to evaluate public transit access to hiking trails and determine if access varies across different socioeconomic and racial groups. From our analysis, we will conclude whether using Bay Area public transportation to access hiking trails is feasible and if that access is equitable.

Literature Review

We used three studies to help guide our research project. O'Connell (2024) explored how the Trailhead Direct in King County, WA, and the Muir Woods Shuttle in Marin County, CA, made trailheads more accessible. Although this paper uses a different type of analysis, it still provides useful information on why group transportation services matter for accessibility. Bernard and Minehart (2024) used geospatial analysis to explore accessibility to nature-based recreation opportunities by demographically diverse communities in the San Gabriel Mountains National

Monument, Mt. Hood, and the Columbia River Gorge Recreation Area. They based their analysis on travel distance via all modes of transportation, but the paper still provides a good framework for how a network analysis could be accomplished. Sriraj (2017) used geospatial analysis to calculate travel times in Cook County, IL, via public transit to visit one of the six nature centers owned by the Forest Preserves of Cook County. This paper provides a useful framework for how to calculate travel times and how to approach transit equity. Additionally, the authors designed hypothetical future transit routes that would increase accessibility, which is something we considered for our project, but fell out of scope.

Methodology

Diagrams

Conceptualization and implementation diagrams for our study are presented in the Appendix as Figures 1, 2, and 3, respectively. A detailed description of each section of the diagrams, including the rationale for operationalizing each variable and the specific GIS steps taken in our analysis, is provided in the remainder of this section.

Data Sources

Data	Source	Data Model	Data Type
bay_area_gtfs.zip	Metropolitan Transportation Commission	Vector	Points, Lines, Polygons
Census.gpkg	NHGIS (2020 census block groups)	Vector	Polygons
norcal.osm.pbf	OSM (using Geofabrik/Overpass Turbo)	Vector	Points and Lines*
bay_area_counties.gpkg	California State Portal	Vector	Polygons
Trails.gpkg	EBDRP, CA State Parks, USGS	Vector	Lines

*Input Data Source Table, * = Nodes and Edges*

Key Concepts

SoVI Equitability

Operationalization

The social vulnerability index quantifies a community's ability to respond to hazards based on demographic and socioeconomic factors. We chose to use 3 variables to comprise the SoVI for our analysis: income, vehicle, and race. We are using block groups from the 2019-2023 American Community Survey (ACS), with data provided by NHGIS, as our spatial units. We chose to use census block groups to form the basis of our analysis because we are interested in very localized areas for calculating distance to transit stops and to provide useful SoVI information. Census tracts did not provide detailed enough data for our analysis. While census blocks would provide the highest level of detail, the time constraints of our project make this option unfeasible due to the large number of blocks, not to mention the computational toll. Census block groups provide the best balance between level of detail and manageable data.

The tables used for this analysis are B19001: Household Income in the Past 12 Months (in 2023 Inflation-Adjusted Dollars), B25044: Tenure by Vehicles Available, and B02001: Race Total population.

Income was operationalized as the percent of households in each block group earning less than $125,000 annually. The threshold to be considered low-income, according to the State of California Department of Housing and Community Development in 2023, is $119,300 annually for a family of two and $134,200 for a family of three. The average household size in San

Francisco is 2.23, according to the 2020 Census. The ACS table for income provides discrete income cutoffs, instead of continuous, with the nearest discrete income field being $100,000 - $124,999. Therefore, we are using less than $125,000 to operationalize income, which falls between the low-income cutoff for two and three-person households.

Vehicle was operationalized as the percentage of households in each block group without access to a personal vehicle. Because our study is concerned with using public transportation instead of personal vehicles, it is important to include no vehicle as a variable in our SoVI.

Race was operationalized as the percent of residents in a block group who identify as non-white or multiracial. We chose to include these variables because of a 2021 study by Xiao et al., which found that low-income, Black, and Hispanic persons are less likely than white persons to visit National Parks. However, the ACS doesn't supply a specific field for Latinx/Hispanic. Instead, we chose to use all non-white persons, which is also beneficial because it accounts for multiracial persons.

GIS Analysis[2]

To analyze the SoVI, we joined the ACS tables to the block groups by the GISJOIN field. Next, we calculated the percentage of households with income below $125,000, the percentage of households with no vehicle, and the percentage of non-white households and appended them to new fields: % below 125,000, % No Car, and % Non-white, respectively[3]. We then calculated the median and third quartile for each of the fields[4]. Using the median and quartiles, we created new

---

2 Detailed diagrams and code can be found in the Appendix for the processes described in this section.
3 See Appendix Figure 3 Specific Implementation for details
4 See Appendix Figure 28 for code

fields of Income, Race, and Vehicle, which were appended for each category with the following scoring:

Value	Score
Above the third quartile	2
Above the median, less than or equal to the third quartile	1
Equal to/below median	0

SoVI Variable Scoring Table, see Appendix Figure 28 for code.

Trail Suitability

Operationalization

Trail suitability consists of two variables: trail length and trail surface type. We gathered trail data from three sources: East Bay Regional Park District (EBRPD), the State of California, and the United States Geological Survey (USGS) National Map.[5]

To operationalize trail length, we used the recommended 120 minutes in natural environments (White et. al., 2019) and Naismith's Rule for estimating hike times (Naismith, 1892). Although other tools are more accurate than Naismith's Rule for specific fitness levels, it still provides a good estimate of the hiking time for an average person. Using Naismith's Rule:

hiking time in hours = (miles / 3) + (feet of ascent / 2000)

Assuming 30 minutes of break time for water and snacks, the hike distance should be at least five miles. It's important to note that Naismith's Rule assumes an average hiking speed of three miles per hour. Some may deem this pace to be too fast, but we are comfortable with this, particularly

[5] See Appendix Figure 5 for a map of input trail data sources.

in thinking about those who are able to go for a 5-mile hike. To determine suitably long trails, we dissolved all the trails in a given area to one feature and calculated the total trail length. Notably, this length is from the trailhead and does not include the walking distance from the nearest transit stop.

To operationalize trail surface type, we chose to only include gravel, dirt, or grass trails. We chose to exclude paved trails from our analysis because hikers prefer a natural surface type over asphalt (Molokáč et al., 2022). Additionally, we chose to exclude sand trails on beaches because they lack designated routes. Further, this filter helps in our goal of targeting trails that are more natural and avoiding urban trails such as rail-to-trail conversions.

GIS Analysis[6]

To calculate suitable trails, we used three vector data sources: East Bay Regional Park District (EBRPD) Trails, California State Park (CSP) Trails, and United States Geological Survey (USGS) Trails, which were clipped to our 9-county study area and reprojected to NAD83 / California zone 3 (ESPG:2227). Some sources included surface type attribution, while others did not. This necessitated a different approach for handling the removal of paved trails. For EPRBD trails, we removed trails with a paved surface type[7]. For CSP trails, we removed trails with surface types of asphalt, chip seal, and concrete[8]. Additionally, we removed trails designated for motorcycle and ATV use[7]. For USGS trails, we queried and then manually removed any undesignated beach routes as well as paved trails using aerial imagery[9]. The three data sources had some spatial overlap, so we manually removed overlapping trails, giving priority to the

[6] Detailed diagrams and code can be found in the Appendix for the processes described in this section.
[7] See Appendix Figure 22 for code
[8] See Appendix Figure 23 for code
[9] See Appendix Figure 24 for code

EBRPD trails due to their higher spatial accuracy, followed by the CSP trails, and then the USGS trails. Due to ingesting trail data from three disparate sources, snapping quickly became a large issue. Due to the scale involved, as well as the complexity of the input data, we decided to manually snap trail segments together where appropriate. After individually editing the trail data sources, we merged them into a single vector layer.

Additional manual editing of trail geometries was performed in locations where trail features did not extend to the road or where two trail segments were bisected by a road. While this does involve crossing a paved surface, which violates our natural surface type requirements, we are justifying this because a hiker will likely have to cross a road anyway to reach the trailhead from the transit stop, so crossing an additional road for the purpose of extending a hike shouldn't pose an issue. Because we manually edited the geometries of trails to create a vector layer of suitable trails in the Bay Area, this constitutes our unique data layer for this project.

After our trail network was finalized, we dissolved the trails based on connected segments. The tools available in QGIS and ArcGIS Pro were insufficient for the processing we needed, as they only offered a single-feature fully dissolved output instead of dissolved groupings of trails. Thus, we used ChatGPT to help write Python code to dissolve the trails using geopandas[10]. This code builds a network to find interconnected trails, assigns a unique identifier to each grouping, and then dissolves based on the unique identifier attribute.

Using the dissolved trails layer, we removed any trails with a length less than 5 miles, yielding all suitable trails in the Bay Area.

[10] See Appendix Figure 25 for details

Public Transportation Suitability

Operationalization

Public transportation suitability consists of ten variables: trip duration, trip day, trip time, number of transit rides, walking speed, public transit stops, public transit routes, trip origin, walking time (from origin to transit stop and from transit stop to destination), and trip destination.

To operationalize trip duration, we subtracted the estimated two hours of hiking time from an eight-hour day, then factored in an hour buffer for any unexpected transit or hiking issues. This yields a maximum travel time of two and a half hours from origin to destination, each way. While we acknowledge this is a somewhat arbitrary choice, it is difficult to set a cap for recreational travel, as most literature on travel choice centers around travel to work or essential services.

To operationalize the trip day, we chose to use Saturday transit routes and schedules. This is working under the assumption that people generally work 9-5 during the week, and thus would only have sufficient free time on a weekend to hike. Vulnerable populations may be working unconventional schedules or working in the gig economy, but this is a limitation we are comfortable with, especially considering transit schedules are often more frequent on weekdays, making the Saturday limitation more restrictive.

To operationalize trip time, we chose 9:00 am as the trip start time. The idea behind this time is that transit trips are often long to relatively remote hiking trails, so we need to allow enough time for transit in both directions, as well as the hike itself, before sunset. Trip time begins when a person leaves the point of origin, not when the first bus departs.

To operationalize trip day, we chose Saturday, November 1, 2025. This date was chosen because the data from the transit feeds does not have historical data and doesn't project far into the future. When we started this analysis and downloaded the transit feed data, November 1, 2025, was the nearest Saturday, so it provided the most accurate timetables.

To operationalize public transit stops and routes, we used all stops and routes from Bay Area transit operators. This data comes from the Metropolitan Transportation Commission as a consolidated General Transit Feed Specification (GTFS) regional transit feed, which includes all 39 Bay Area transit operators. We did not include any private transit operators in our analysis.

To operationalize trip origin, we chose to use census block group centroids. Census block groups are the basis for our Social Vulnerability Index (SoVI), as detailed later in the paper, matching geographies significantly simplifies the analysis. While centroids are an imperfect measure, particularly considering varying shapes and sizes of block groups, routing from individual homes is beyond the scope of this project. Block group centroids were generated using QGIS. r5r handles snapping by searching for routable segments within a 300-meter radius of input points. This occurs for both origin and destination points.

To operationalize walking time to the transit stop and from the transit stop to the trailhead, we set the maximum walking time to 10 minutes for each leg. The generally accepted value used in city planning is that citizens are willing to walk 0.25-0.33 miles to reach a transit stop (Weinstein et al., 2008). A 2008 study found that the median distance transit users walked to rail stops in Oregon and California was 0.47 miles (Weinstein et al., 2008). While this study only examined distance to rail stops, in conjunction with the generally accepted value, a maximum walking distance to transit stops of 0.5 miles will be adequate. However, r5r expects time instead of distance package we are using to calculate transit routes uses a time metric instead of distance to determine which bus stops are accessible from block group centroids. Therefore, we are using the average time to walk a half mile, 10 minutes, to operationalize walking time. Due to the limited time frame of our analysis, we chose to ignore the topography of San Francisco. Additionally, we are operating under the assumption that if a resident is going for a longer hike, a hilly walk to a transit stop will not act as a deterrent.

To operationalize the trip destination, we used custom-created trailhead data. It became clear early on that there was not a sufficient source of data for trailheads, particularly given we sourced trails from various sources. We decided to generate the trailhead data ourselves by overlaying trails and the OSM street network on top of aerial imagery and snapping trailheads to the ends of trails or where trails intersected roads. This process resulted in 152 trailheads, which serve as our trip destinations, and is an additional original data layer.

We operationalized the number of transit rides as less than or equal to six. When calculating the travel matrix later in our analysis, the maximum number of transit rides used to reach trails within 2.5 hours was 6. We only added this as a variable because it is an expected parameter in generating a travel time matrix in r5r, and as it was useful in weighting scoring of hiking accessibility.

GIS Analysis[11]

To analyze public transport suitability, we extracted 2020 census block groups within the City of San Francisco from NHGIS. Block groups on islands or without any physical connection to the San Francisco Peninsula were removed. We then generated centroids for each block group.

For the bulk of our public transportation suitability analysis, we used the R^5 routing engine within the r5r R package to generate a travel time matrix from the block group centroids to each trailhead. This outputs information about trips to each reachable trailhead within our parameters (duration, walking time, etc.). We deemed a distance matrix to be more robust than simply finding the quickest reachable hiking trail from each centroid because it better reflects real-world nuances, such as a case where the quickest reachable trail is relatively close, but there are many other trails that take a long time to reach. r5r expects a routable street network; for this, we chose to use OpenStreetMap (OSM), a community-sourced map of the world, as our data provider. OSM data was extracted using Geofabrik as a .pbf file, then clipped to a bounding box of our nine-county boundary using osmconvert on the command line[12]. It also needs a General Transportation Feed Specification (GTFS) layer for transit routing. This feed was downloaded from the Metropolitan Transportation Commission (MTC) as a consolidated regional feed,

[11] Detailed diagrams and code can be found in the Appendix for the processes described in this section.

[12] See Appendix Figure 21 for code.

including all 39 Bay Area transit operators. Origins and destinations had to be reprojected to WGS 84 and include specific fields (id, lat, lon).

While using r5r, it is important to be attentive to where origins and destinations are relative to the routable network. We ran into issues where block groups on nearby islands had been included, which were far beyond the snapping tolerance and caused problems in generating the travel time matrix[13].

After generating the travel time matrix[14], we removed three block groups which were extreme outliers in the number of accessible trails: two with zero accessible trails and one with one. We then assigned each centroid a Trail Accessibility Score to balance the total travel time and the number of transit rides on a trip. For each centroid-to-trailhead route, we calculated Duration Score and Rides Score using the following scoring:

Duration Score:

Minutes	0-15	15-30	30-45	45-60	60-75	75-90	90-105	105-120	120-135	135-150
Score	10	9	8	7	6	5	4	3	2	1

Rides Score:

Number of Rides	1	2	3	4	5	6
Score	6	5	4	3	2	1

[13] See Appendix Figure 27 for code to check point snapping.
[14] See Appendix Figure 29 for code to generate a travel time matrix

We summed the Duration Score and Rides Score to produce a Trail Accessibility Score, which was then averaged for each centroid to produce an Average Trail Accessibility Score[15].

Methodological Choices

We chose to work in the vector world as we are dealing with distinct origin and destination pairs in census block group centroids and trailheads, respectively. Furthermore, our SoVI index has discrete boundaries due to relying on census block group geographies, so the fuzziness of a raster would not be appropriate.

Our approach to the technology stack in this project is purposely platform agnostic and makes every attempt to make use of free and open source software (FOSS). QGIS was used for the vast majority of data processing. However, for creating SoVI, ArcGIS Pro was used due to errors in QGIS while joining census tables to block group polygons.

The r5r package in R formed the backbone of our routing analysis. We chose this because of it being FOSS, avoiding ArcGIS credits, and because of our relative familiarity with R.

Discussion

For each combination of our SoVI variables, income, race, and vehicle, we calculated the R-squared value and created bivariate choropleth maps. Income and vehicle had a correlation of 0.3, income and race had a correlation of 0.14, race and vehicle did not correlate[16]. Additionally,

[15] See Appendix Figure 30 for code and Appendix Figure 15 for map
[16] See Appendix Figures 11-13 for maps

we created choropleth maps for each conditional combination of the three variables, and calculated the linear regression between the variables, but didn't find any notable results[17]. Therefore, for our analysis of equity in reaching trails via public transportation, we decided to analyze each of the SoVI variables individually.

Before calculating our travel-time matrix, we had 81 trails and 152 trailheads. Of these trailheads, 43 were reachable under our conditions, whereas 109 were not[18]. The minimum number of reachable trailheads from a centroid was 20, while the maximum was 43[19]. Northern areas of San Francisco were able to access the most trailheads, while northeastern areas were able to access the second most. Only a few centroids were able to access fewer than 24 trails, and were scattered across the city with no discernible pattern, indicating that the placement of the centroid within each black group may have played a larger role in the inaccessibility than the general location of the block group. We decided that the ability to access 20 trails was adequate to say every census block centroid has good access to trails under our criteria. Therefore, we focused our analysis on comparing social vulnerability variables to the average Trail Accessibility Score.

After calculating and mapping the average Trail Accessibility Score for each centroid, an interesting pattern emerged. Most of the centroids with the highest average Trail Accessibility Score were clustered around the BART route. This makes intuitive sense because the BART provides high-speed access to the eastern Bay Area, where many trails are concentrated. The

[17] See Appendix Figures 14-14.4 for maps and charts. Conditional Combination Maps: Map 1: Race or Income or Vehicle, Map 2: Race and Vehicle or Income, Map 3: Income and Vehicle or Race, Map 4: Income and Race or Vehicle, Map 5: Income and Race and Vehicle
[18] See Appendix Figure 7 for map
[19] See Appendix Figure 16 for map

areas with the lowest average Trail Accessibility Scores were concentrated in the southeast, central, and western areas of San Francisco. These areas are likely serviced by slower buses, making the overall trip time longer, which decreases the Trail Accessibility Scores[20].

For each of our vulnerability variables, we determined the correlation with the average Trail Accessibility Score. Income vulnerability and Trail Accessibility had an R-squared of 0.02. Race vulnerability and Trail Accessibility had an R-squared of -0.02. Vehicle vulnerability and Trail Accessibility had an R-squared of 0.28[21].

Limitations

One limitation of our study is the use of ACS data. While this data is widely accepted as being suitable for demographic analysis, it uses a more limited sample size than the Census, leading to high margins of error. However, we wanted more recent data than the 2020 Census, and we chose to use the 5-year ACS to reduce some of the error. Additionally, we are looking at general ranges of vulnerability, not specific values, so we are willing to accept slightly higher margins of error. Additionally, three of the census block groups were made up mostly of parks and golf courses, so the centroids were placed further than a 10-minute walk away from a transit stop, meaning no trailheads were reachable under our criteria. Because we were not able to determine the location of the population center, we chose to exclude these block groups from our analysis.

Another limitation of our analysis is determining trial length. For the scope of our analysis, it was not feasible to manually measure the length of a trail from each trailhead. Additionally, it

[20] See Appendix Figure 17 for map
[21] See Appendix Figures 18-20 for maps

would be difficult to determine exact hiking routes because a trailhead could provide access to a variety of trails and route options. Therefore, we decided to use any dissolved trail groupings longer than five miles for our analysis.

Conclusion

In this paper, we examined the accessibility of hiking trails via public transportation from San Francisco. From our analysis, we found that every block group centroid has access to at least 20 trailheads, with the maximum being 43. We decided that the ability for each block group to access at least 20 trailheads constitutes good access under our criteria. Although this may seem low compared to the 152 trailheads used in our analysis, due to how large our study area was, many of the periphery trailheads are barely accessible in 2.5 hours with a car. We also assessed whether this accessibility is equitable for communities with high income, race, and vehicle ownership vulnerabilities. We found that having a household income below $125,000 or being non-white does not significantly affect the ease of accessing hiking trails. This is likely because San Francisco has a robust public transit network, providing all communities with accessible transit. However, we found that not having access to a household vehicle has some correlation with better trail accessibility. This makes some intuitive sense because the majority of the block groups with high vehicle vulnerability are located near the BART line. For people living in these block groups, a car may not be as necessary because there is a good public transportation system.

Future Research

Our analysis serves as a starting point for assessing the equity of trail access in San Francisco, but much more research is needed. We would like to create an interactive map allowing users to explore our data and findings, as our variables are complex and result in a large quantity of data. Because our analysis relies on input data from various sources and spatial scales, it is only as robust as those inputs. It would be beneficial to ground-truth our analyses to determine how practical each of our transit routes are. Some potential sources could be Strava, which provides an interesting community-sourced angle allowing for a weighting of how well used a trail is, and AllTrails, which provides difficulty and popularity, as well as more accurate distance tied to specific trail routes.

Expanding our analysis to include block groups in the larger Bay Area and conducting analyses in other cities could provide valuable insight into the equity of trail access. From our research and assumptions, San Francisco is an anomaly in that it is proximate to many hiking trails and has a robust public transportation system, so it would be valuable to draw comparisons between other U.S. cities.

Bibliography

Bandlamudi, A. (2025, May 7). *California's population is rebounding. In San Francisco, it's a different story*. KQED. https://www.kqed.org/news/12038968/californias-population-rebounding-san-francisco-different-story

Bernard, K., & Minehart, K. (2024). *A Geospatial Analysis of Equitable Recreation Access on Public Lands: A Case Study in the American West*. https://doi.org/10.31235/osf.io/s96bd

Bratman, G. N., Hamilton, J. P., Hahn, K. S., Daily, G. C., & Gross, J. J. (2015). Nature experience reduces rumination and subgenual prefrontal cortex activation. *Proceedings of the National Academy of Sciences*, *112*(28), 8567–8572. https://doi.org/10.1073/pnas.1510459112

Breting-Garcia, V. (2022). *San Francisco Bay: Research starters: EBSCO research*. EBSCO. https://www.ebsco.com/research-starters/science/san-francisco-bay

City performance scorecards. Poverty in San Francisco | City Performance Scorecards. (n.d.). https://www.sfgov.org/scorecards//safety-net/poverty-san-francisco

Fares. SFMTA. (2025). https://www.sfmta.com/getting-around/muni/fares

Kirkeby, M. (2023, June 6). *2023 state income limits*. State Of California - Business, Consumer Services And Housing Agency. https://www.hcd.ca.gov/sites/default/files/docs/grants-and-funding/income-limits-2023.pdf

Micalizio, P. (2025, June 6). *The true cost of owning a car in California*. SmartFinancial. https://smartfinancial.com/how-much-does-it-cost-to-own-a-car-in-california

Molokáč, M., Hlaváčová, J., Tometzová, D., & Liptáková, E. (2022). The preference analysis for hikers' choice of Hiking Trail. *Sustainability*, *14*(11), 6795. https://doi.org/10.3390/su14116795

Mulley, C., Ho, C., Ho, L., Hensher, D., & Rose, J. (2018). Will bus travellers walk further for a more frequent service? an international study using a stated preference approach. *Transport Policy*, *69*, 88–97. https://doi.org/10.1016/j.tranpol.2018.06.002

Naismith, W. (1892). Notes and queries. *Scottish Mountaineering Club Journal, vol. 2, p. 133*. https://simrich.ams3.digitaloceanspaces.com/SMCJ/Number%207%20to%2012%20-%20Vol%20%202%20-%201892-1893.pdf

O'Connell, N. (2024). *Wildly Inaccessible: Reaching Public Lands via Public Transit* (Order No. 31851329). Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (3171647625). https://ezproxy.library.wisc.edu/login?url=https://www.proquest.com/dissertations-theses/wildly-inaccessible-reaching-public-lands-via/docview/3171647625/se-2

Rep. Gomez and senator Booker introduce Bill to promote access to public lands and green spaces through public transit. U.S. Representative Jimmy Gomez. (2023, May 4). https://gomez.house.gov/news/documentsingle.aspx?DocumentID=2793

San Francisco. Population | Explore census data for over 7 million residents of the Bay Area | Bay Area Census. (2020). https://census.bayareametro.gov/population?location=san_francisco

San Francisco, CA. Data USA. (n.d.). https://datausa.io/profile/geo/san-francisco-ca/

San Francisco is first U.S. city where all residents live within a 10-minute walk to a park - trust for public land. (2017, May 16). https://www.tpl.org/media-room/san-francisco-first-us-city-where-all-residents-live-within-10-minute-walk-park

Sriraj, P. S. (2017). *Public Transit to Public Lands: The Nature Express*. https://doi.org/10.5038/cutr-nctr-rr-2017-04

Survey findings: Demographics - San Francisco. San Francisco Municipal Transportation Agency. (2019). https://www.sfmta.com/sites/default/files/reports-and-documents/2021/02/survey_findings_demographics.pdf

S.1440 - 118th Congress (2023-2024): Transit to Trails Act | congress.gov | library of Congress. Transit to Trails Act. (2023, May 4). https://www.congress.gov/bill/118th-congress/senate-bill/1440

Weinstein Agrawal, A., Schlossberg, M., & Irvin, K. (2008). How far, by which route and why? A spatial analysis of pedestrian preference. Journal of Urban Design, 13(1), 81–98. https://doi.org/10.1080/13574800701804074

White, M. P., Alcock, I., Grellier, J., Wheeler, B. W., Hartig, T., Warber, S. L., Bone, A., Depledge, M. H., & Fleming, L. E. (2019). Spending at least 120 minutes a week in nature is associated with good health and Wellbeing. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-44097-3

Xiao, X., Lee, K. J., & Larson, L. R. (2021). Who visits U.S. national Parks (and who doesn't)? A national study of perceived constraints and vacation preferences across diverse populations. *Journal of Leisure Research*, *53*(3), 404–425. https://doi.org/10.1080/00222216.2021.1899776

O'Connell, N. (2024). *Wildly Inaccessible: Reaching Public Lands via Public Transit* (Order No. 31851329). Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (3171647625). https://ezproxy.library.wisc.edu/login?url=https://www.proquest.com/dissertations-theses/wildly-inaccessible-reaching-public-lands-via/docview/3171647625/se-2

Bernard, K., & Minehart, K. (2024). *A Geospatial Analysis of Equitable Recreation Access on Public Lands: A Case Study in the American West*. https://doi.org/10.31235/osf.io/s96bd

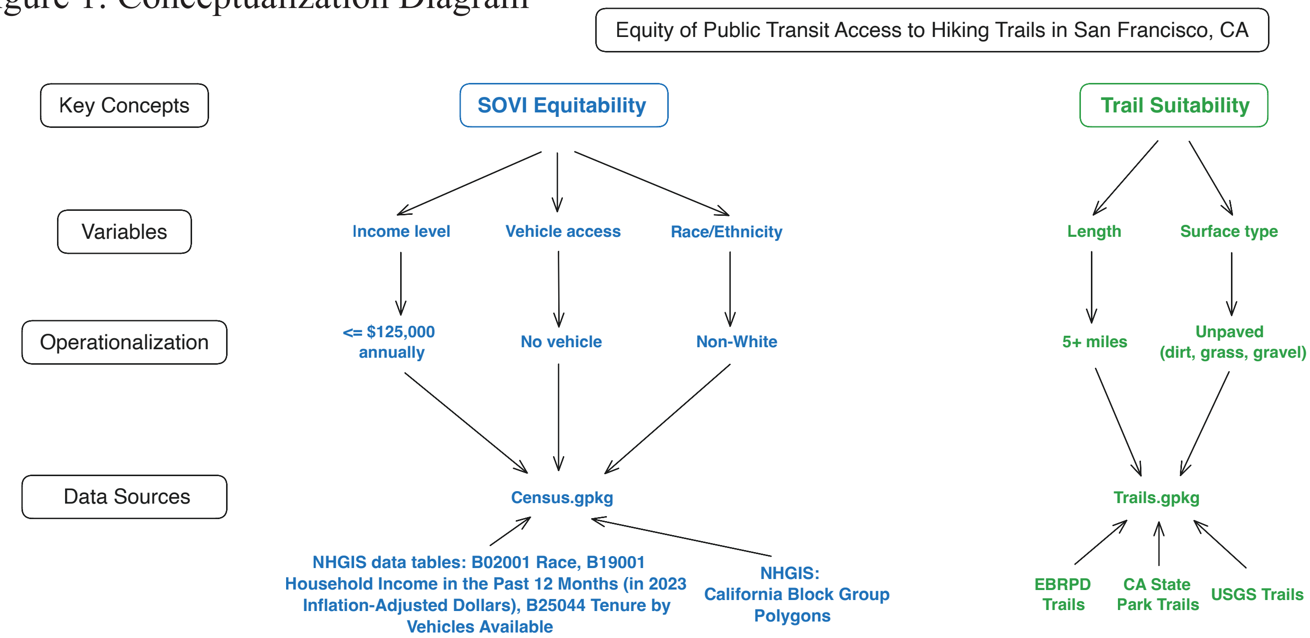
Sriraj, P. S. (2017). *Public Transit to Public Lands: The Nature Express*. https://doi.org/10.5038/cutr-nctr-rr-2017-04

Appendix

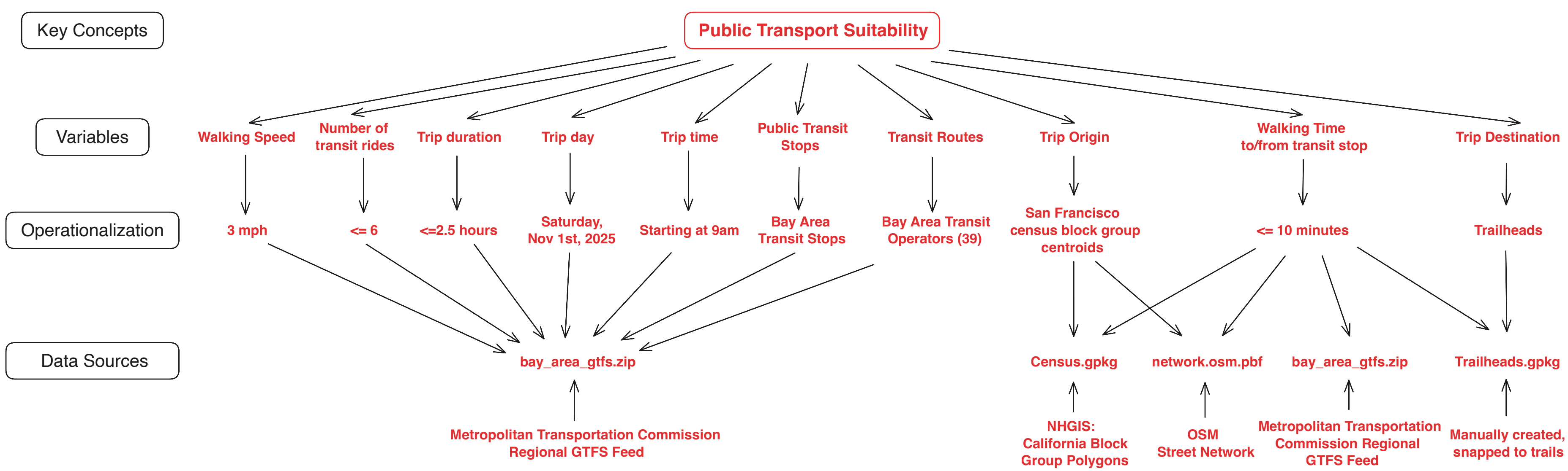
Diagrams

(Diagrams begin on next page)

Figure 1: Conceptualization Diagram

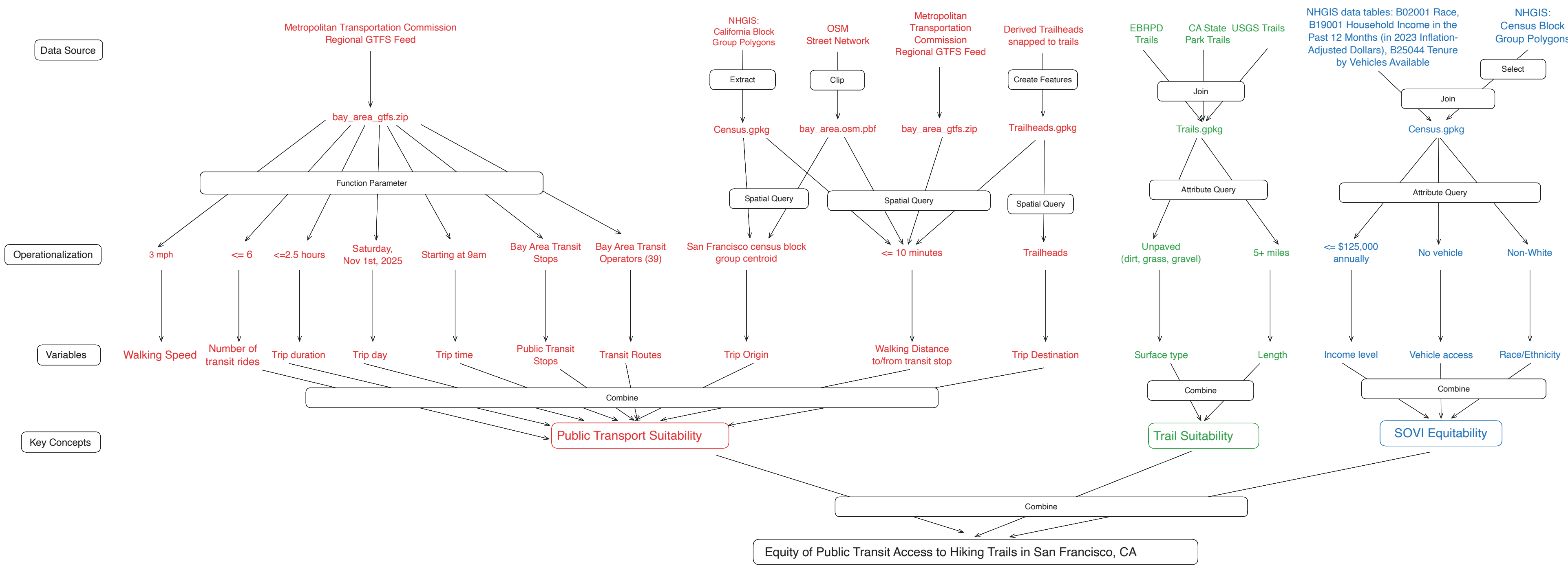


Part 1, SoVI Equitability and Trail Suitability



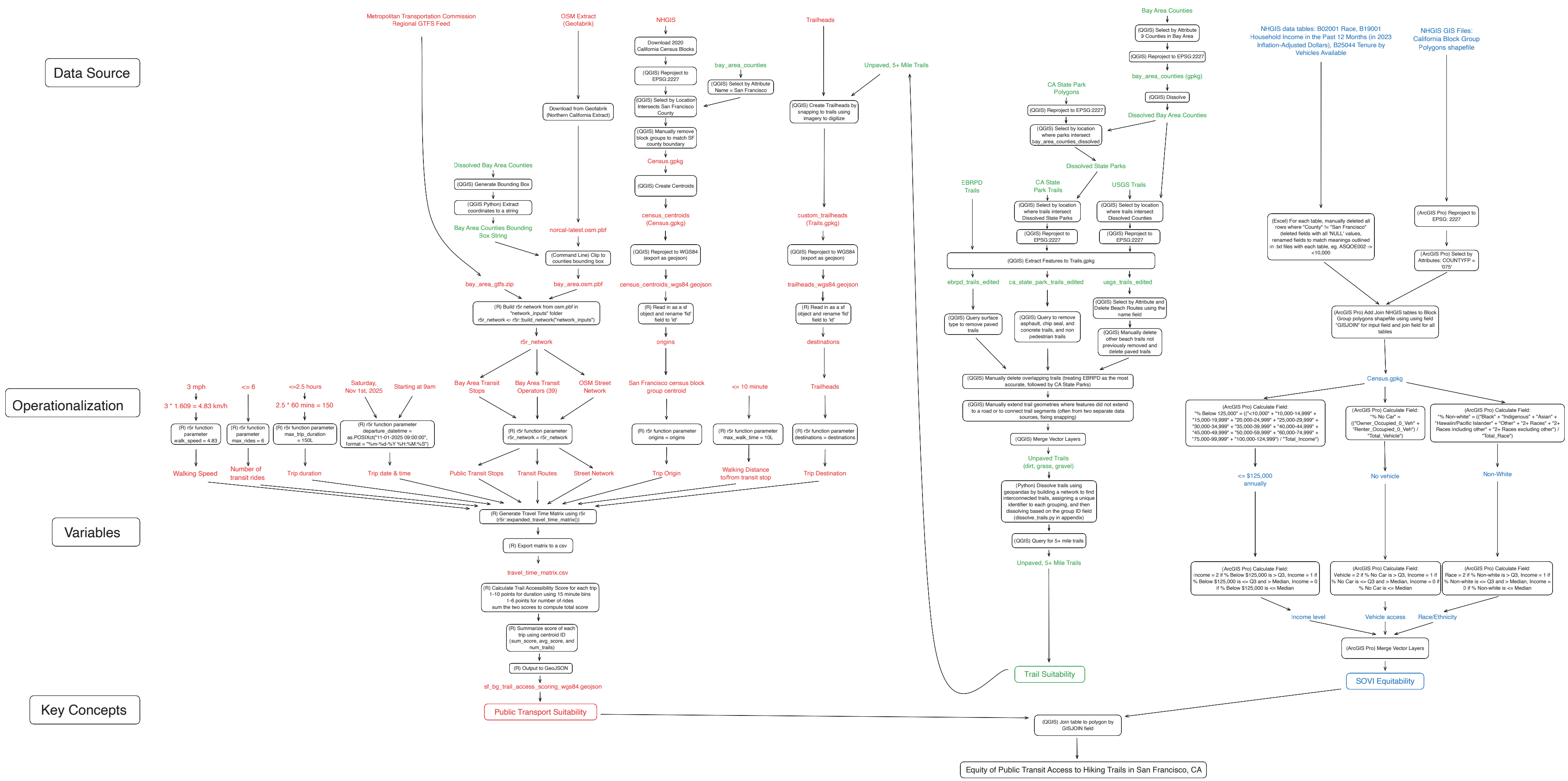
Part 2, Public Transport Suitability

Figure 2: General Implementation Diagram



Full General Implementation Diagram

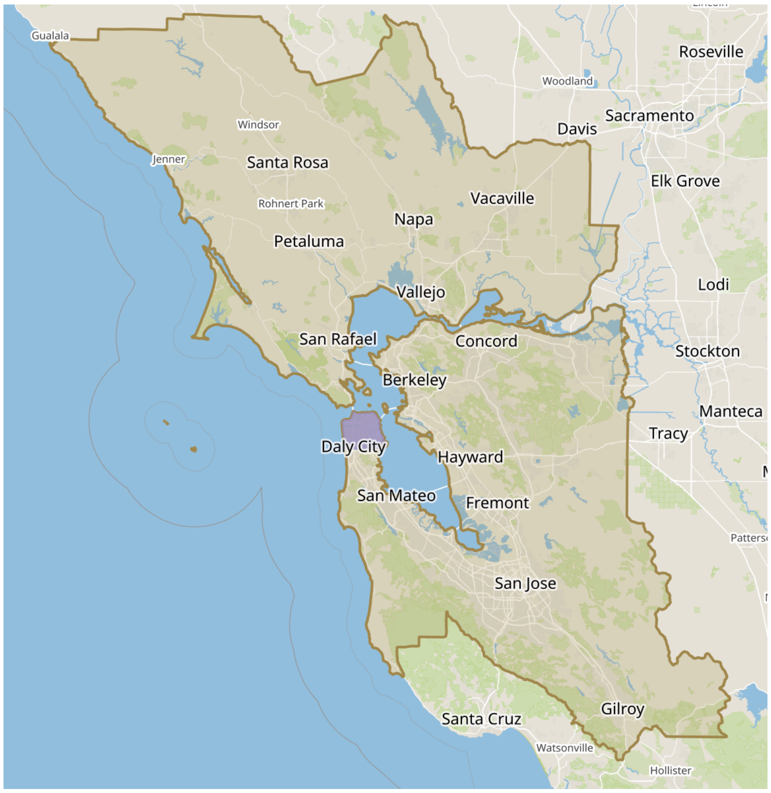
Figure 3: Specific Implementation Diagram



Full Specific Implementation Diagram

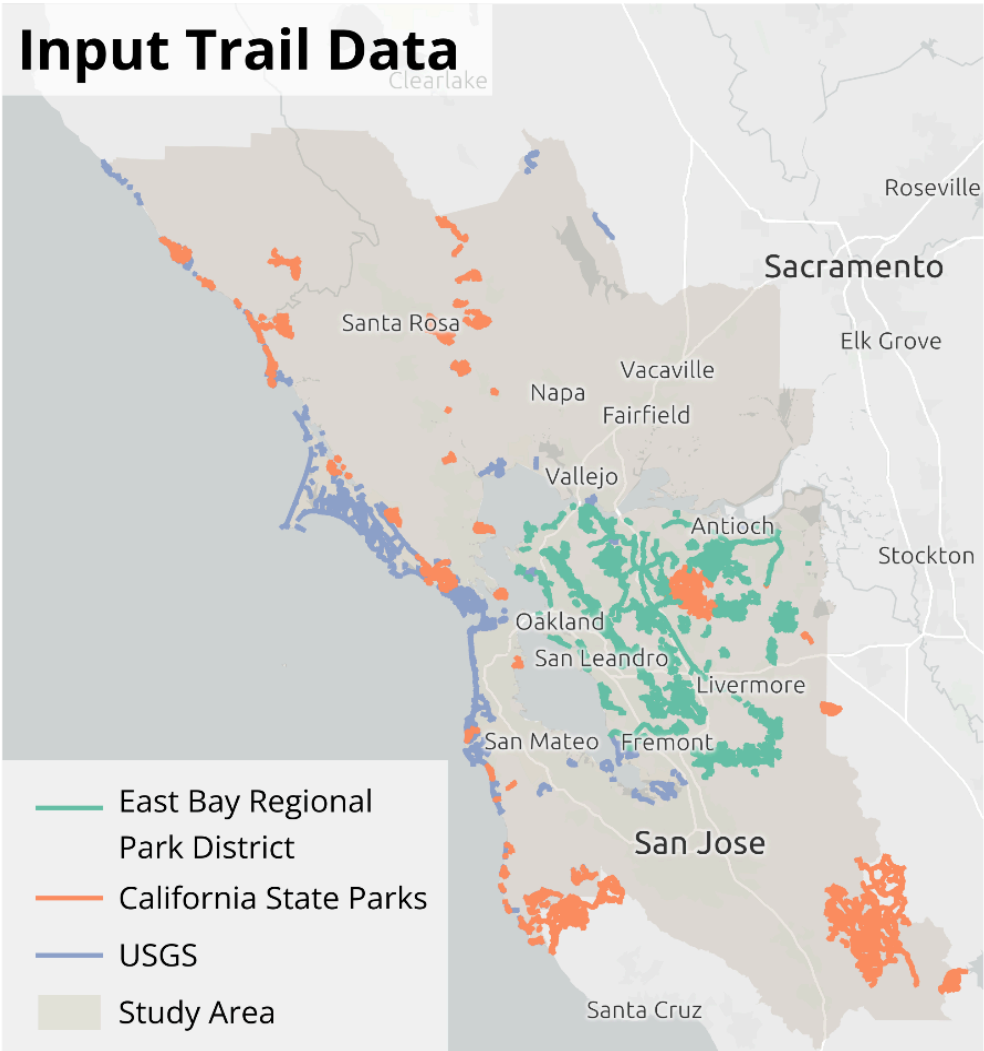
Maps

Figure 4: Map of Study Area



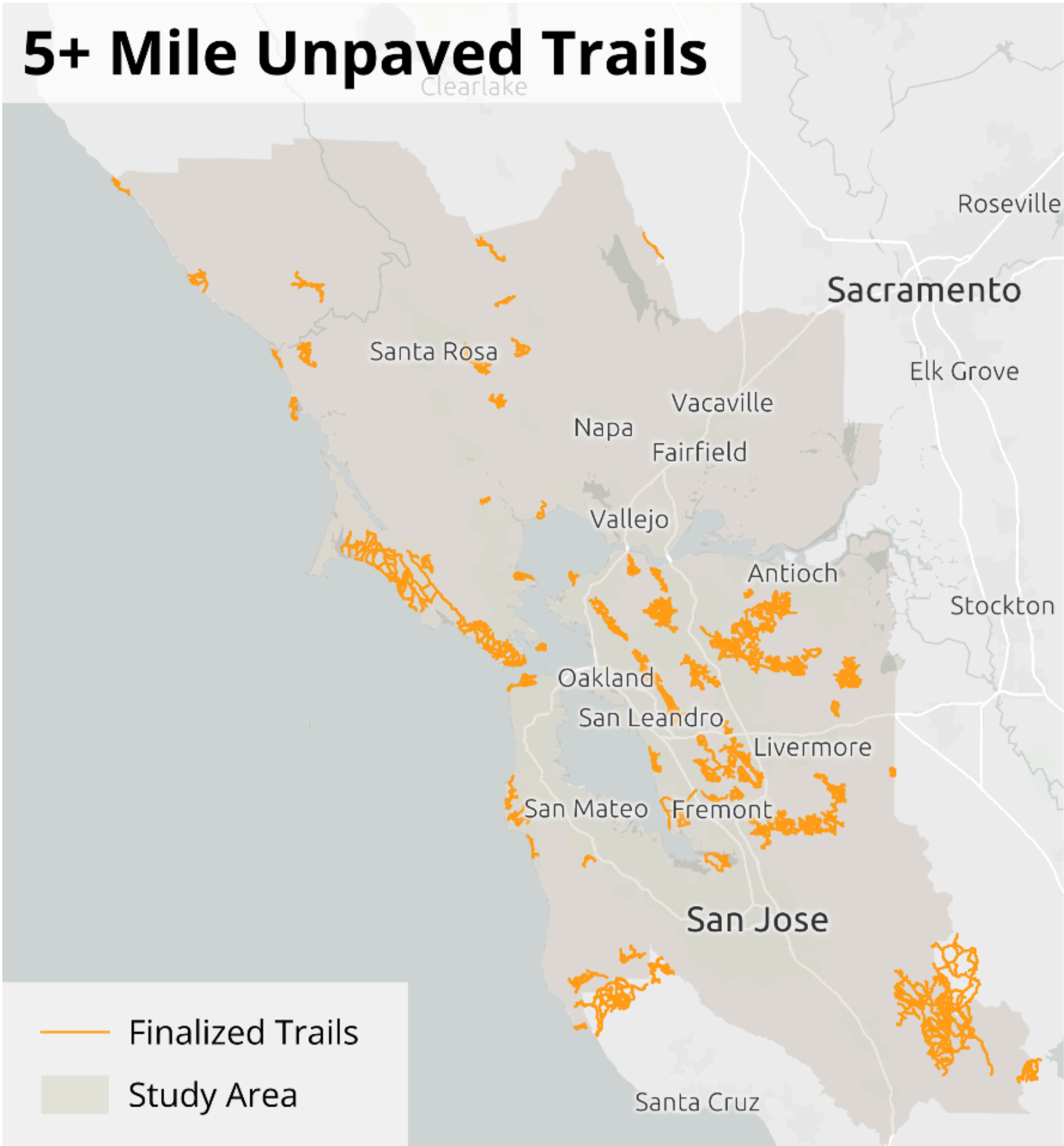
Map of Study Area, data from NHGIS and State of California.

Figure 5: Input Trail Sources



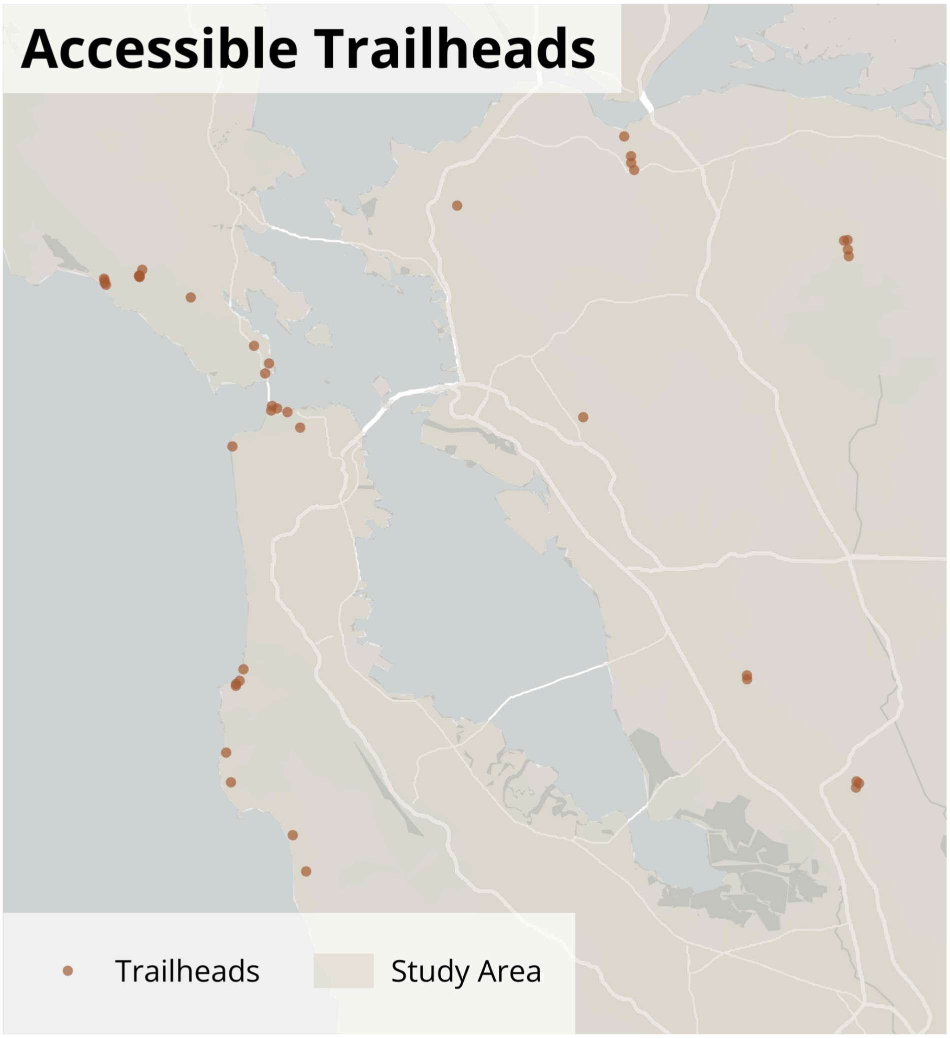
Map of Input Trail Source, data from EBRPD, CA State Parks, and USGS

Figure 6: Suitable Trails



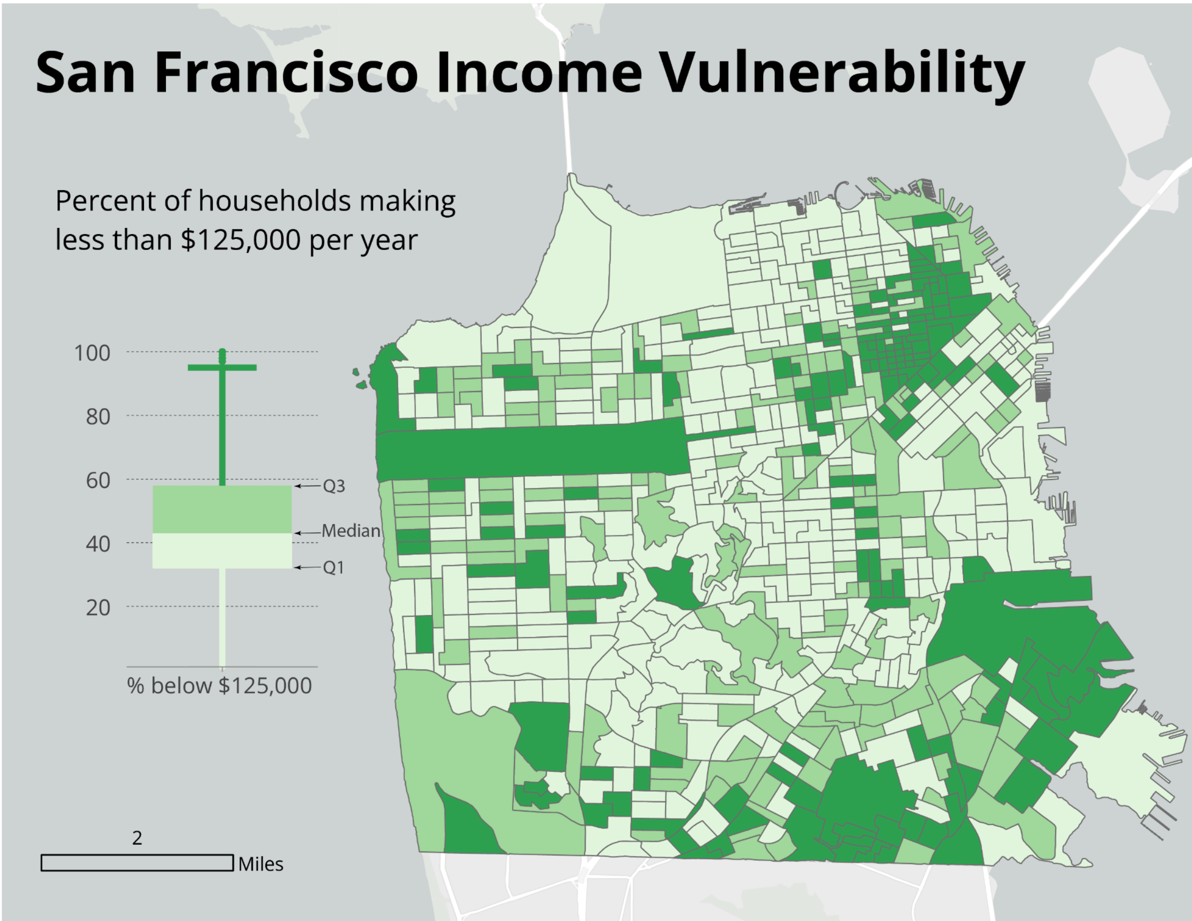
Map of Suitable Trails (5+ Mile, Natural Surface Trails). Data generated from EBRPD, CA State Parks, and USGS inputs.

Figure 7: Accessible Trailheads



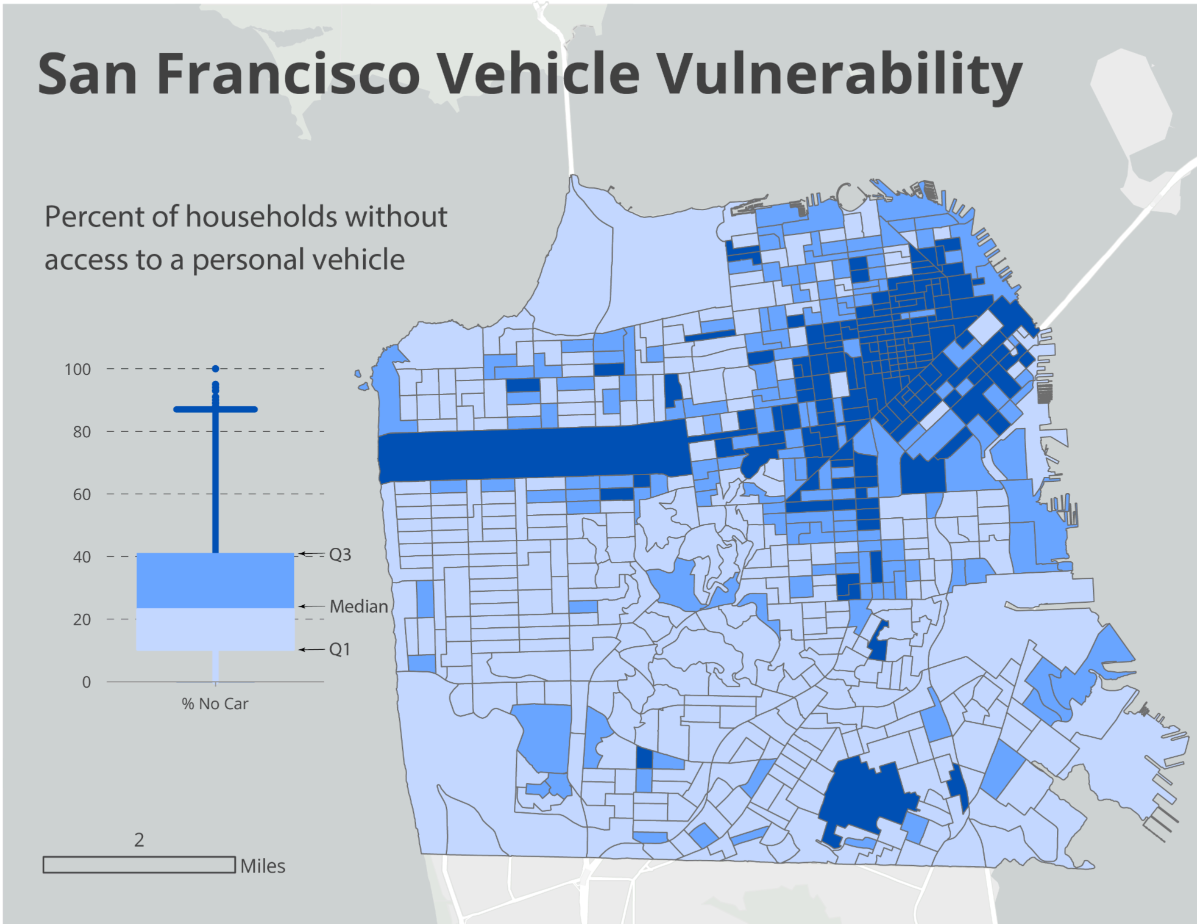
Map of accessible trailheads

Figure 8: Income Vulnerability



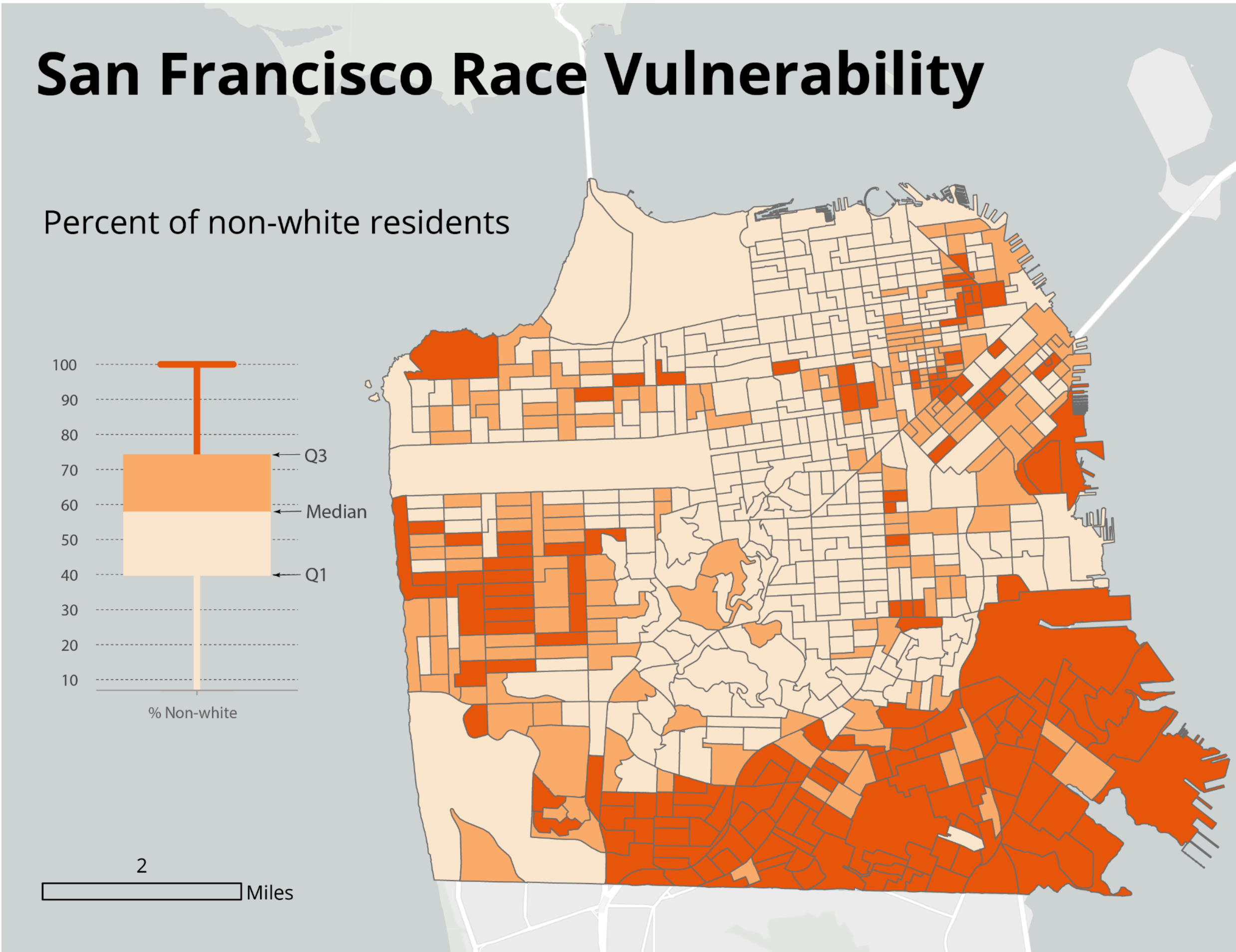
Map of San Francisco income vulnerability. Data from NHGIS

Figure 9: Vehicle Vulnerability



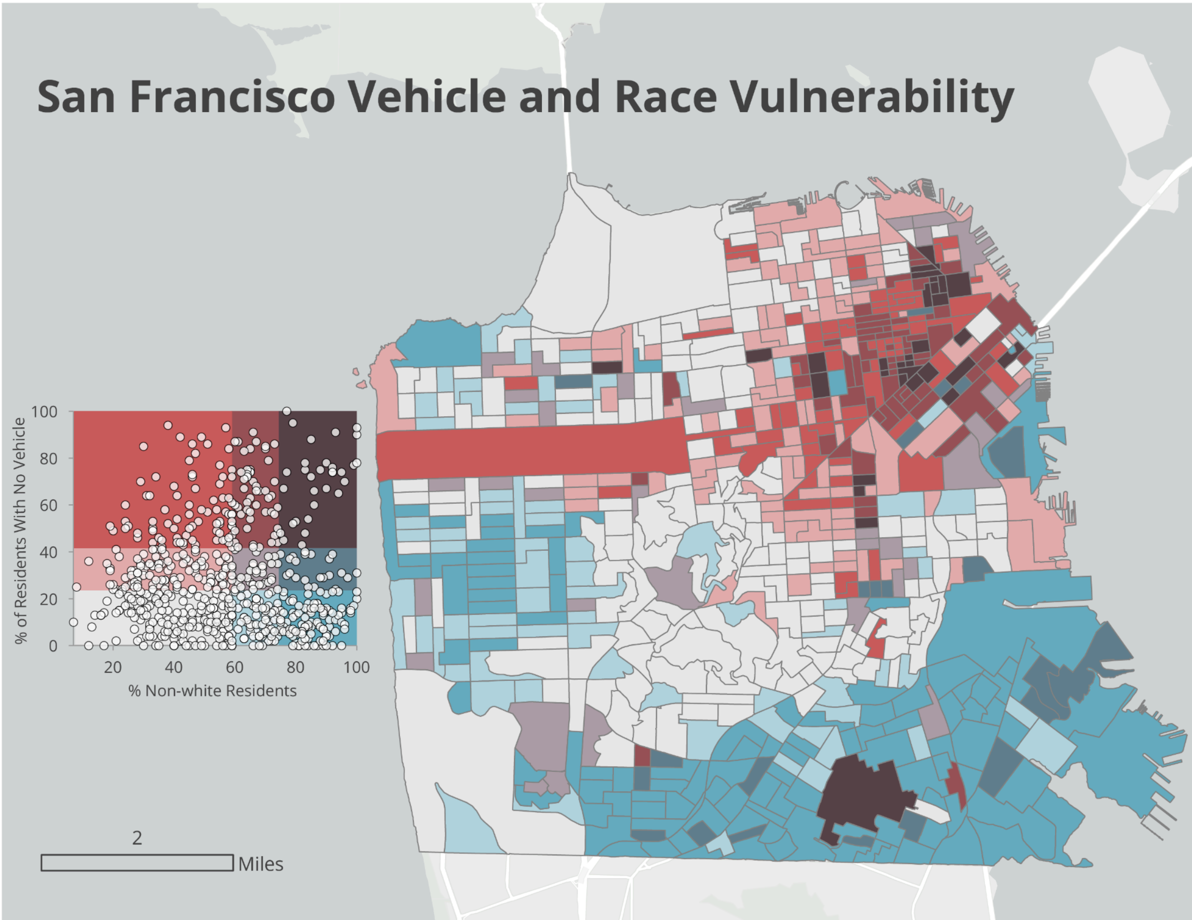
Map of San Francisco vehicle vulnerability. Data from NHGIS

Figure 10: Race Vulnerability



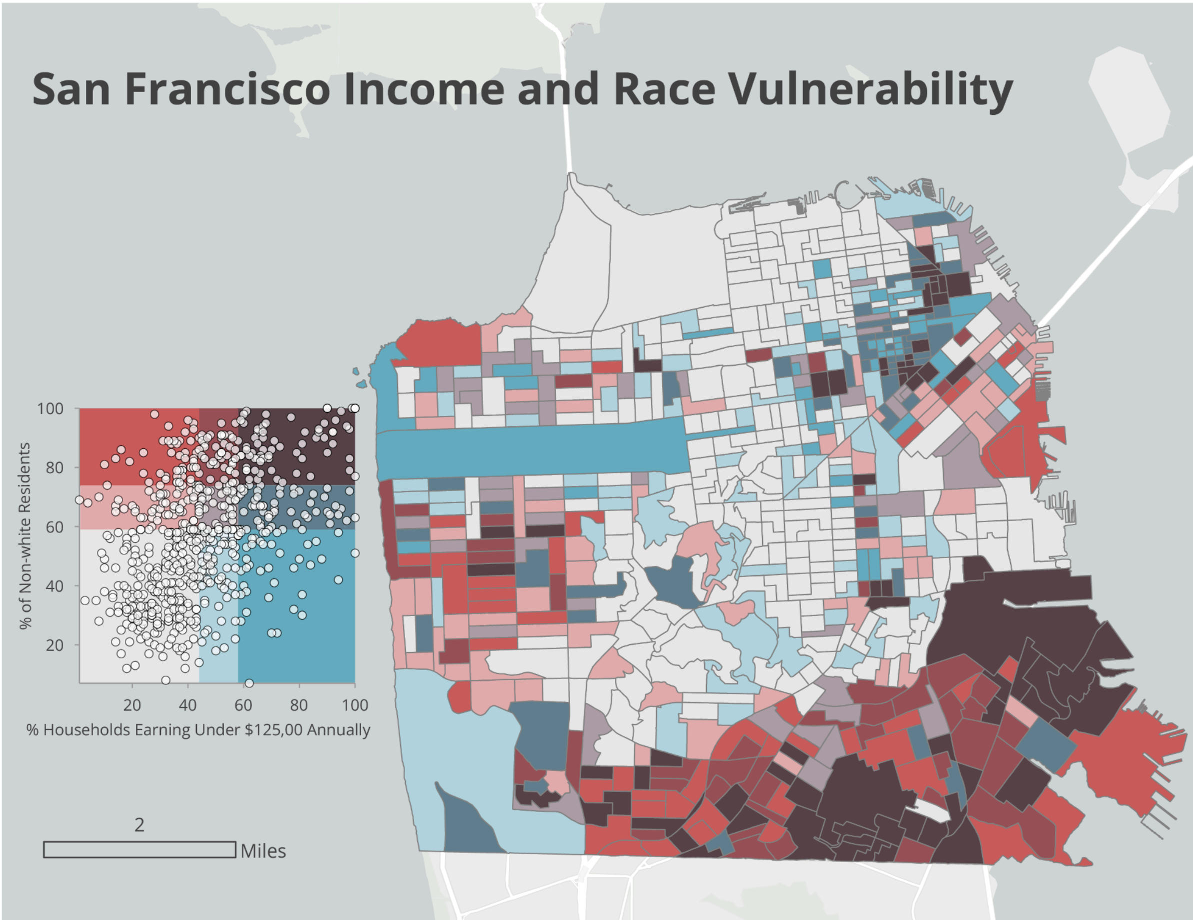
Map of San Francisco race vulnerability. Data from NHGIS

Figure 11: Vehicle and Race Vulnerability



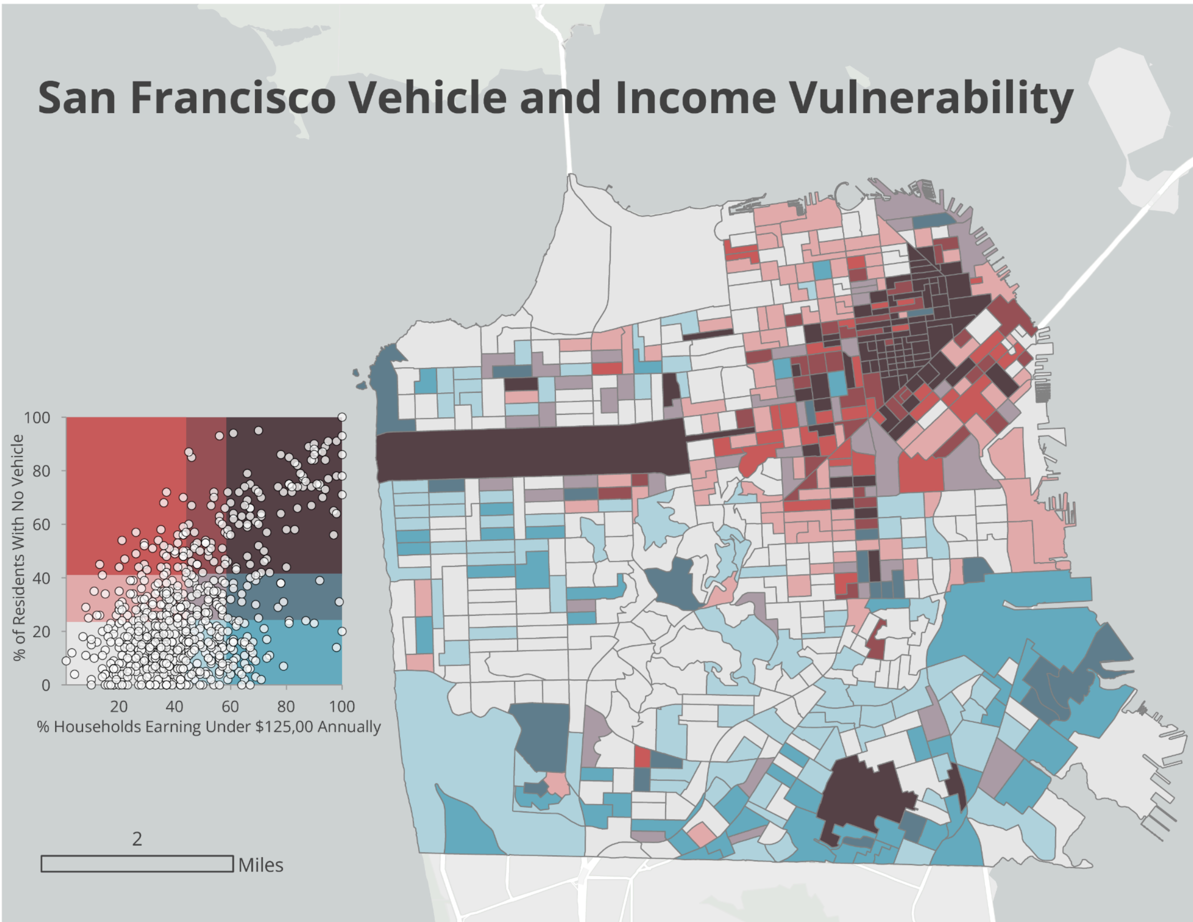
Bivariate choropleth of San Francisco vehicle and race vulnerability. Data from NHGIS

Figure 12: Income and Race Vulnerability



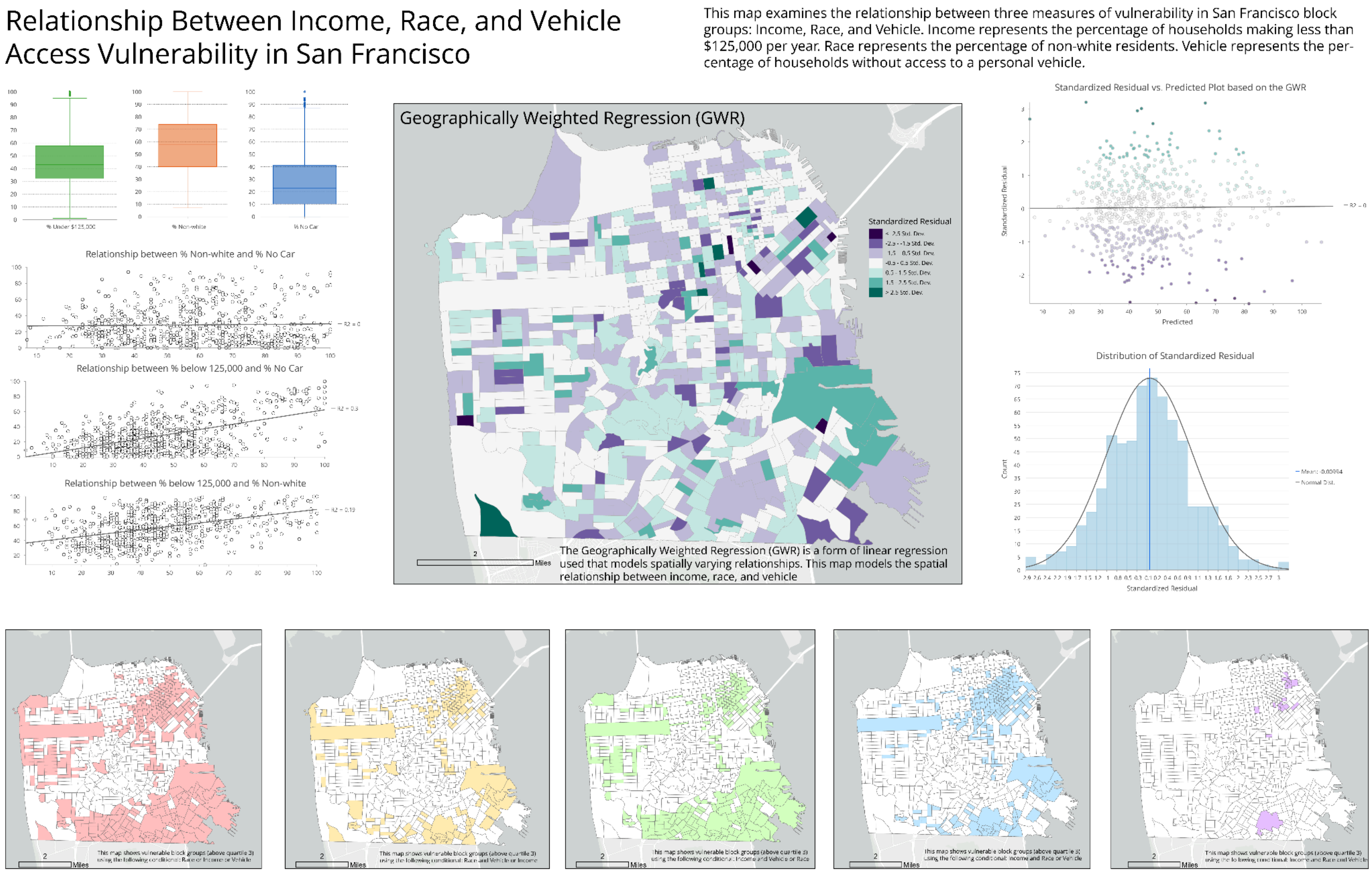
Bivariate choropleth of San Francisco income and race vulnerability. Data from NHGIS

Figure 13: Vehicle and Income Vulnerability



Bivariate choropleth of San Francisco vehicle and income vulnerability. Data from NHGIS

Figure 14: Relationship between Income, Race, and Vehicle Access Vulnerability



Relationship Between Income, Race, and Vehicle Access Vulnerability in San Francisco. Detailed sections in the following figures

Figure 14.1: Box Plots and R^2 Values

Figure 14.2: Geographically Weighted Regression of Income, Race, and Vehicle

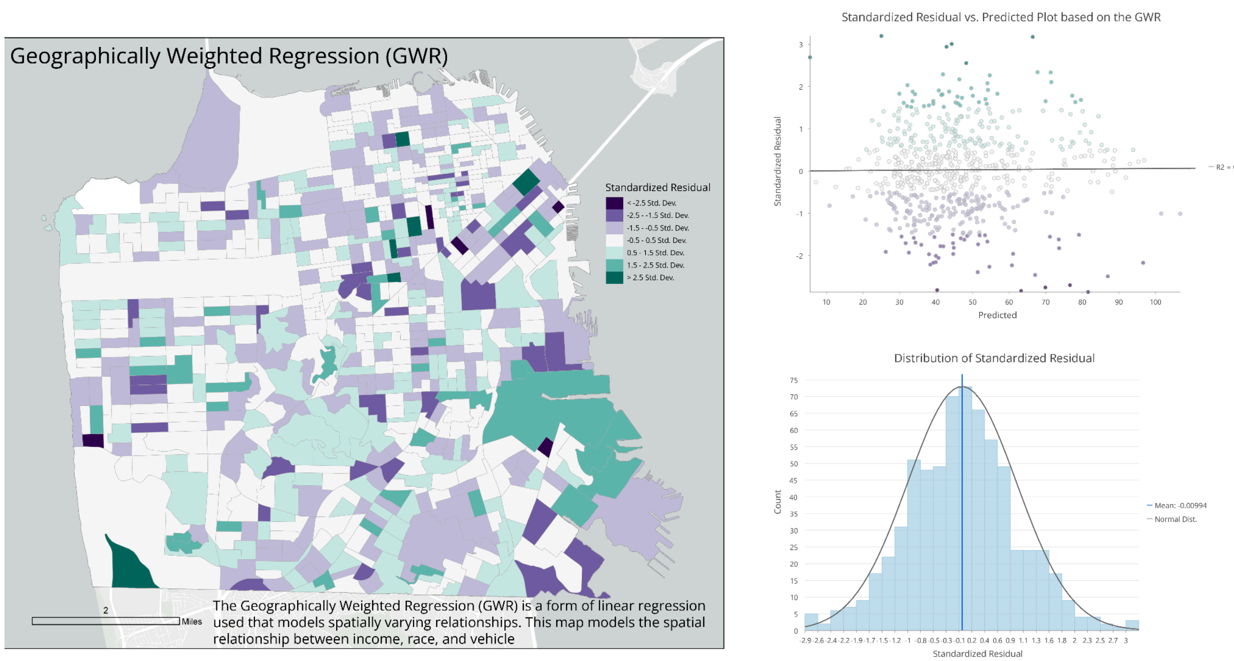


Figure 14.3: First 3 conditional combinations of Race, Vehicle, and Income

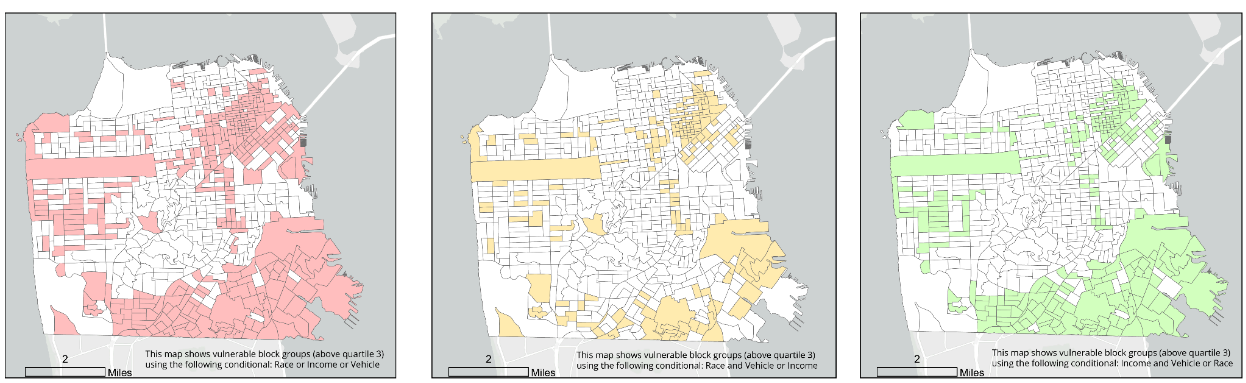
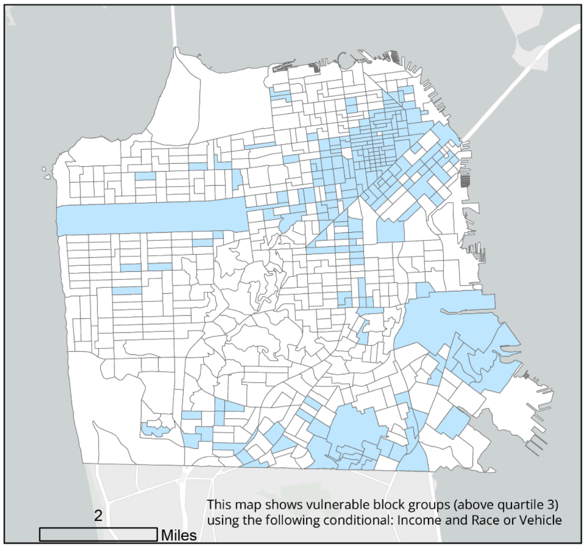


Figure 14.4: Final 2 conditional combinations of Race, Vehicle, and Income



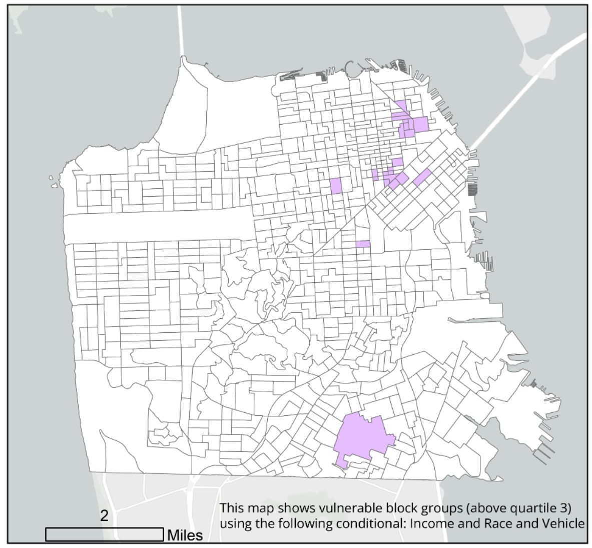
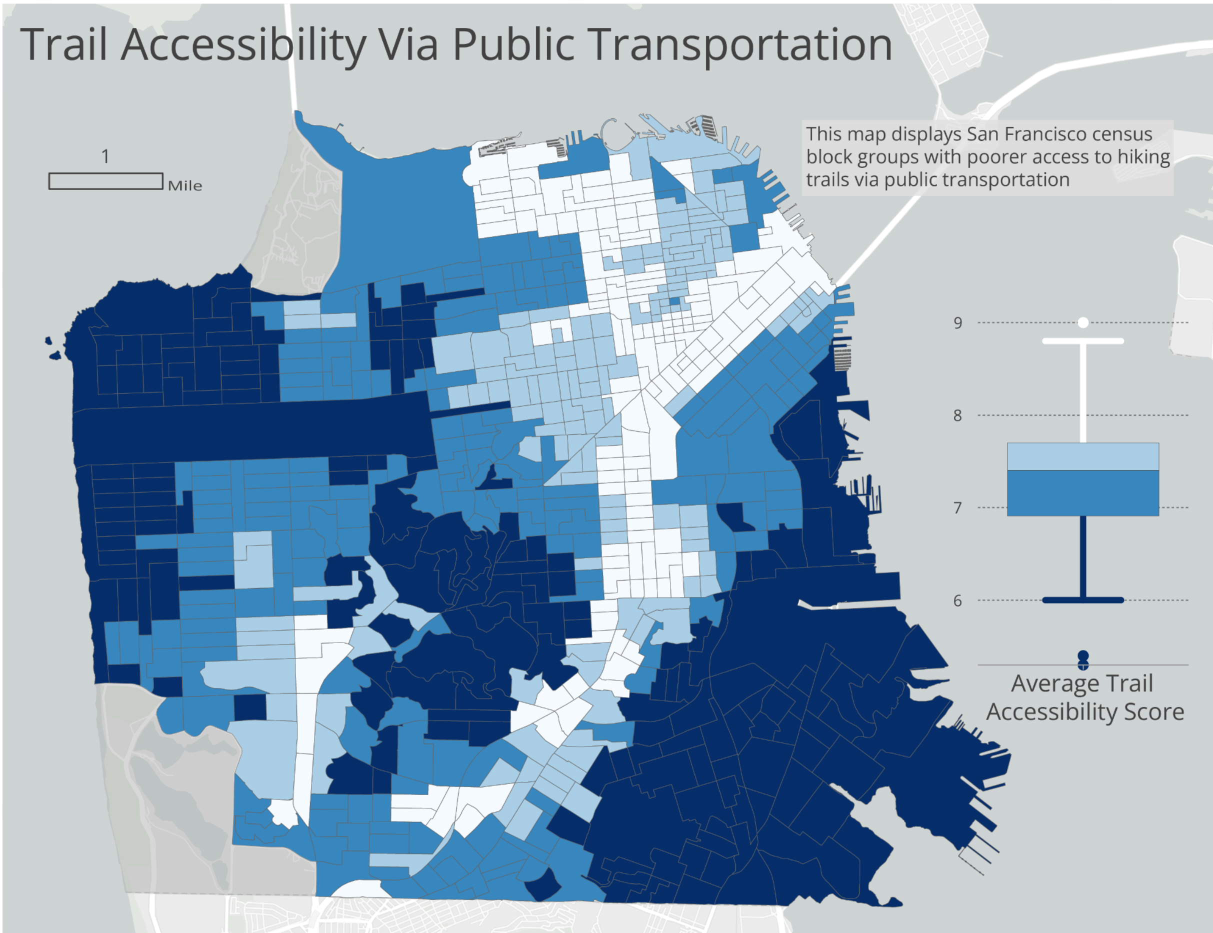
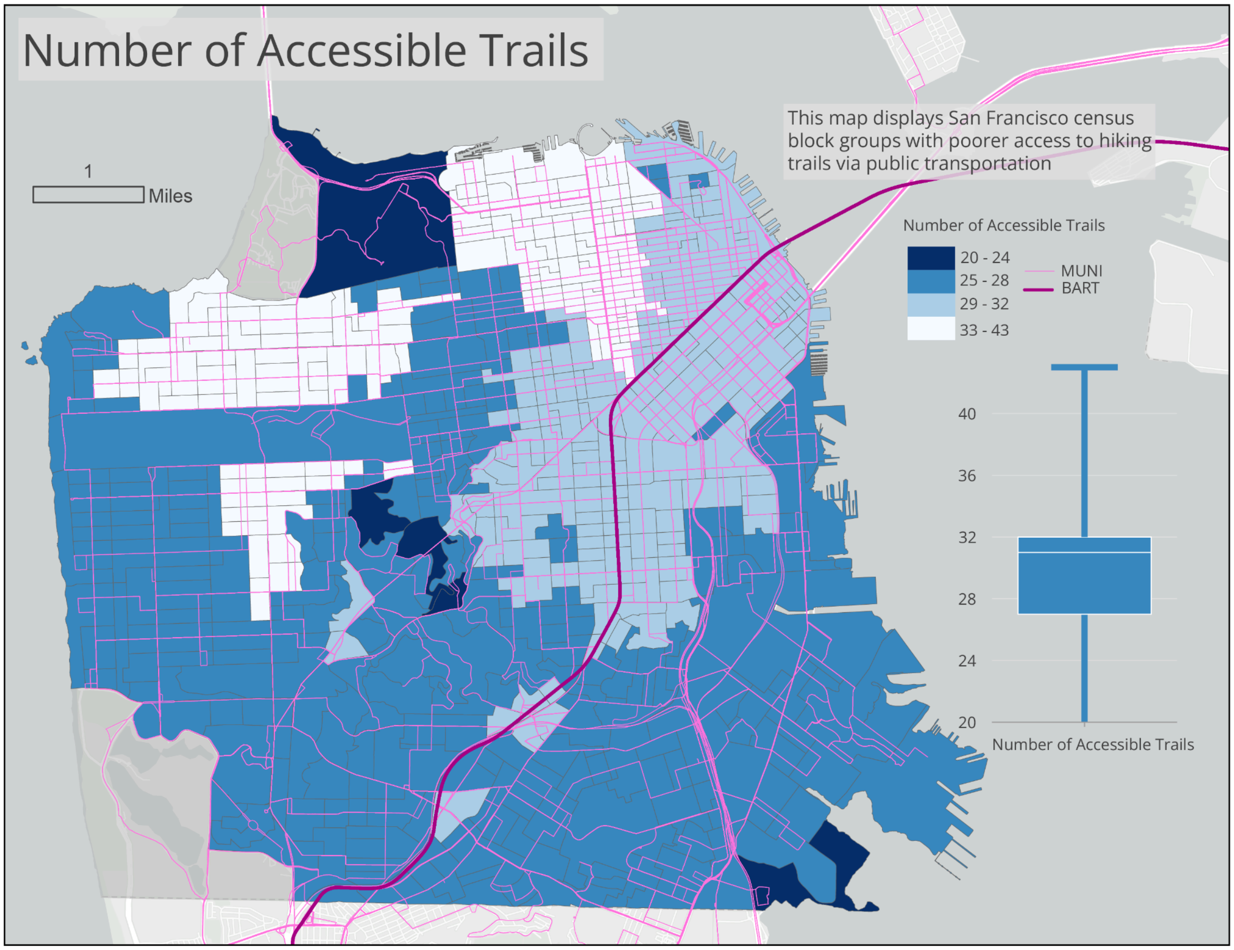


Figure 15: Average Trail Accessibility



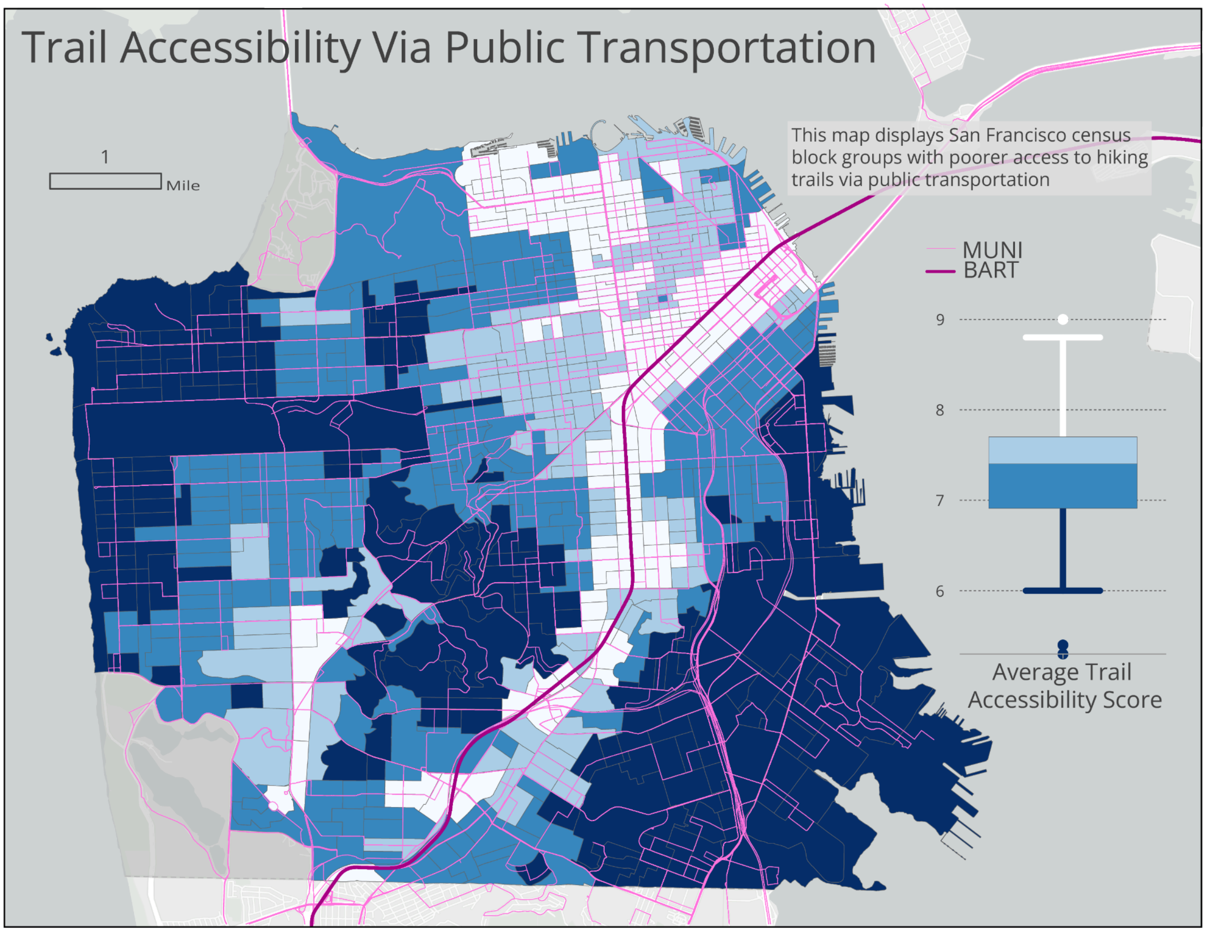
Map of the average Trail Accessibility Score for each San Francisco block group.

Figure 16: Number of Accessible Trails with Transit Overlay



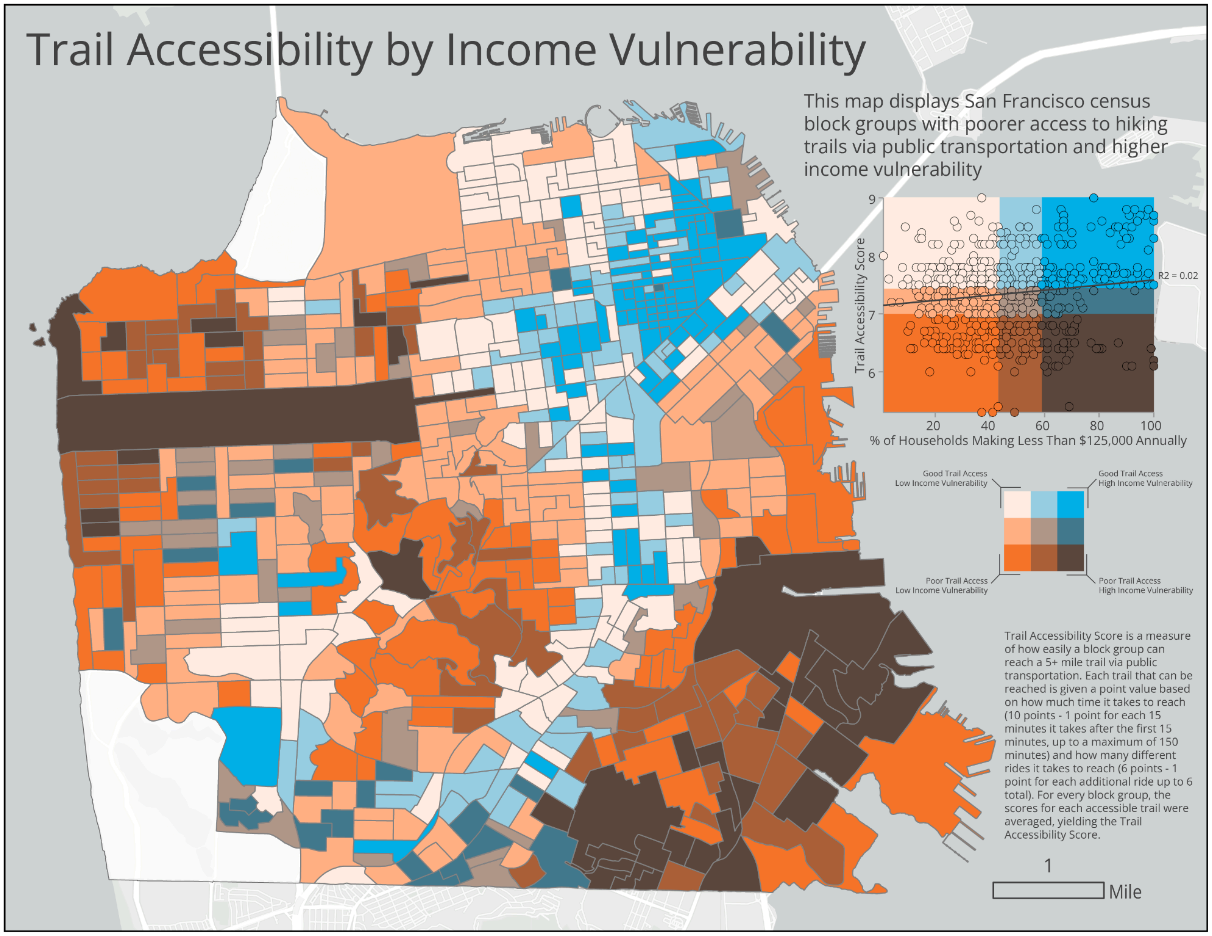
Map of how many trails each block group can reach based on our criteria. Classes created using Jenks Natural Breaks

Figure 17: Average Trail Accessibility Score with Transit Overlay



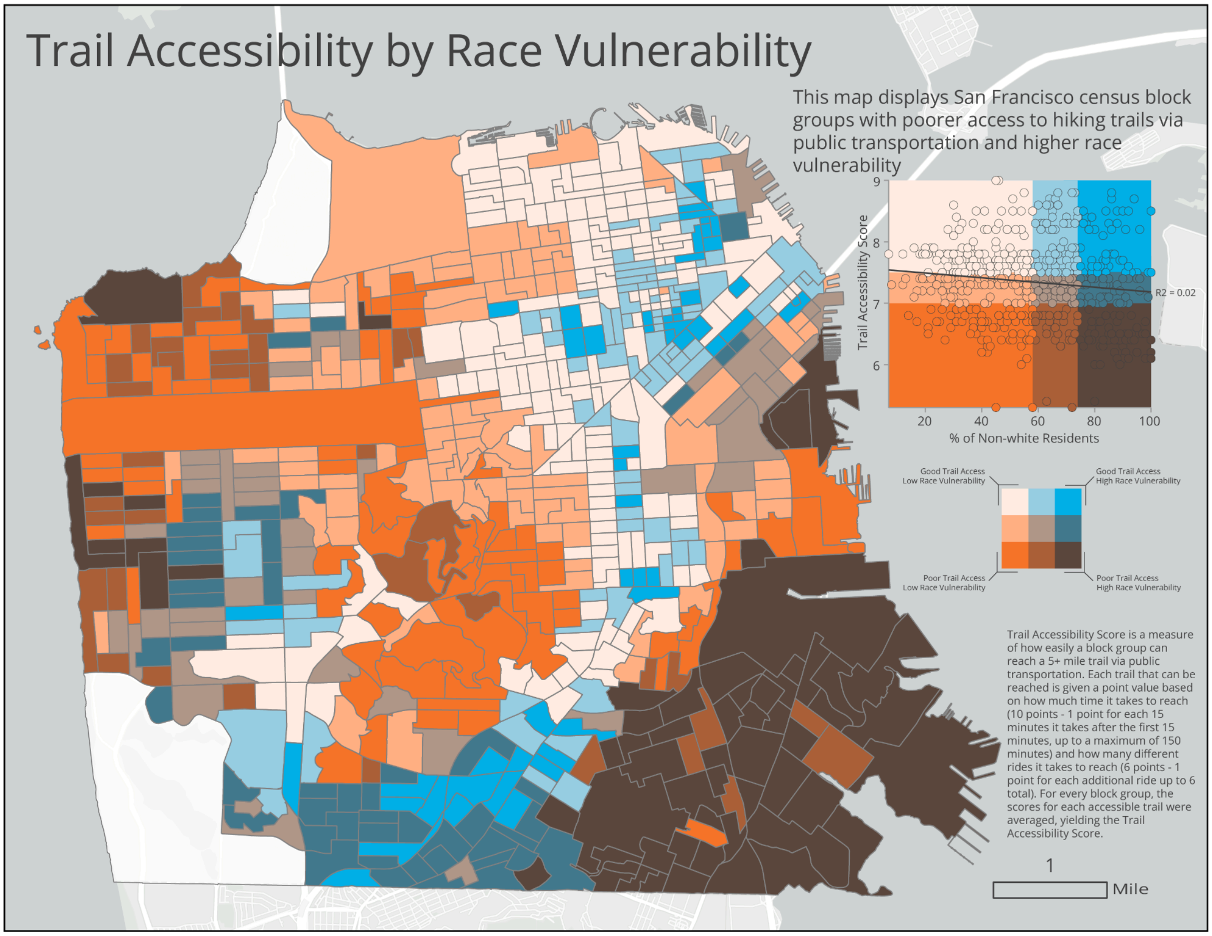
Map of the average Trail Accessibility Score for each San Francisco block group with an overlay of public transit routes.

Figure 18: Trail Accessibility by Income Vulnerability



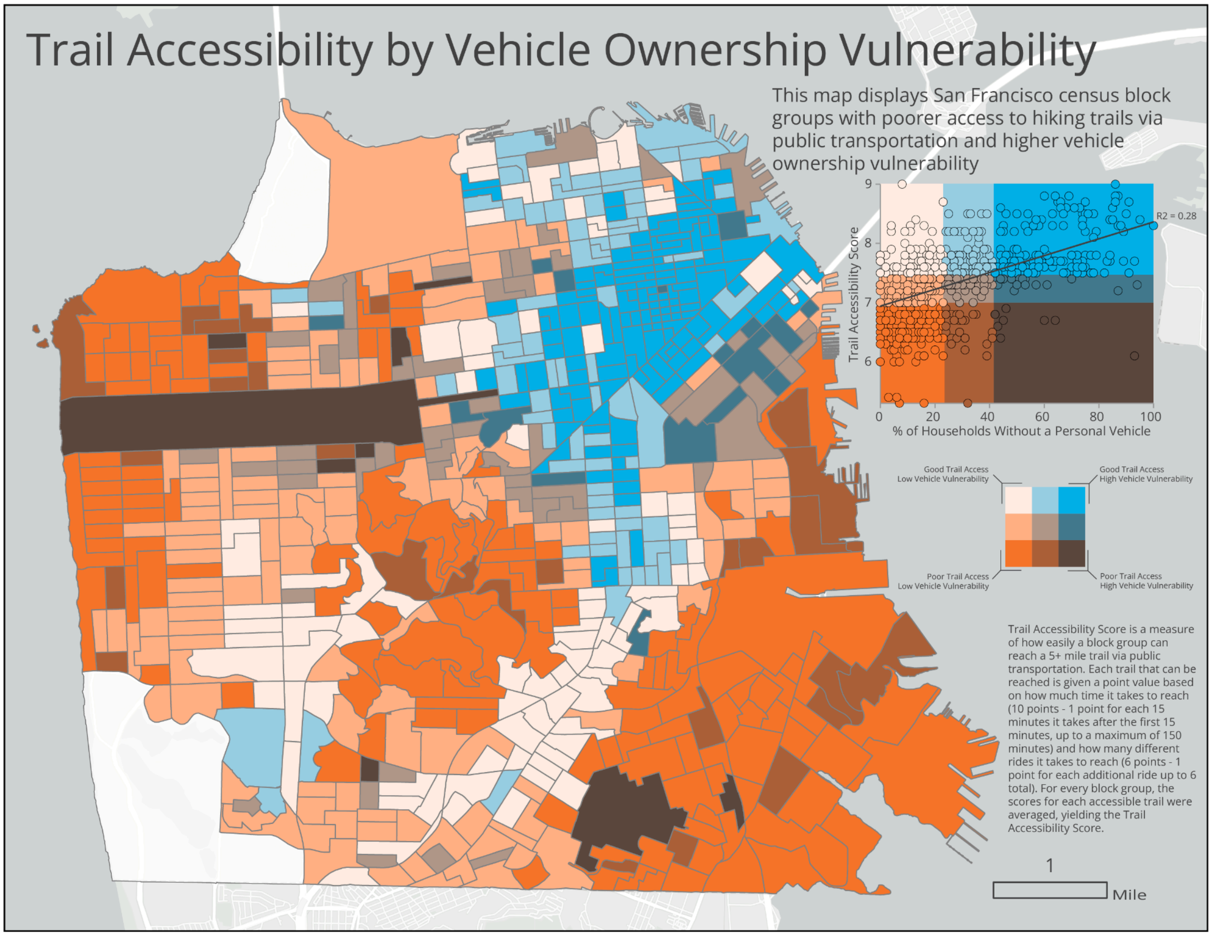
Map of Trail Accessibility by Income Vulnerability.

Figure 19: Trail Accessibility by Race Vulnerability



Map of Trail Accessibility by Race Vulnerability.

Figure 20: Trail Accessibility by Vehicle Ownership Vulnerability



Map of Trail Accessibility by Vehicle Vulnerability.

Code

Figure 21: Clipping OSM Street Network to Bounding Box

```
# requires osmfilter CLI
osmconvert norcal.osm.pbf -b=-123.13,36.89,-121.21,38.51
-o=bay_area.osm.pbf
```

Clipping OSM pbf street network to Bay Area counties bounding box.

Figure 22: EBRPD Trails Surface Type Query

```
-- trail query to remove paved trails
COALESCE("SURFACE", '') != 'Paved'
```

Querying out paved trails from EBRPD Trails.

Figure 23: California State Park Trails Surface Type and Use Query

```
-- remove asphalt, chip seal, and concrete by route type
-- also remove non pedestrian trails
COALESCE("ROUTETYPE", '') NOT IN ('Asphalt', 'Chip Seal', 'Concrete')
AND "TRAILDES" NOT IN ('4X4 / ATV / Motorcycle / ROV', 'ATV', 'ATV /
Motorcycle', 'ATV / Motorcycle / ROV', 'ATV / ROV', 'Motorcycle',
'Prohibited')
```

Querying out paved trails and non pedestrian trails from California State Park Trails

Figure 24: USGS Beach Route Query

```
"name" like '%Beach Route%'
```

Querying out beach routes by name from USGS Trails.

Figure 25: Dissolving Trails by Grouping in Python

```
# Dissolve trails using geopandas with a multipart dissolve
import geopandas as gpd
import networkx as nx
import os

# read in trails
merged_trails = gpd.read_file("sf_trail_access_qgis/data/Trails.gpkg",
layer='merged_trails')

# Output NULL geometries for inspection/troubleshooting
print("Null Geometries")
print(merged_trails[merged_trails.geometry.isnull()])
print('\n\n')

# ---- FIX: Remove invalid geometries ----
merged_trails = merged_trails[merged_trails.geometry.notnull() &
~merged_trails.geometry.is_empty].copy()
merged_trails = merged_trails.reset_index(drop=True)

# Build empty graph
G = nx.Graph()
G.add_nodes_from(merged_trails.index)

# Spatial index for fast lookup
sindex = merged_trails.sindex

# Build edges (intersecting lines → edge in graph)
for i, geom in merged_trails.geometry.items():
    # Candidate matches by bounding box
    candidates = list(sindex.intersection(geom.bounds))

    for j in candidates:
        if i >= j:     # avoid duplicate checks and self-check
            continue

        if geom.intersects(merged_trails.geometry[j]):
            G.add_edge(i, j)

# Find connected components
components = list(nx.connected_components(G))

# Build a mapping from original index to group id
group_id = {}
for cid, comp in enumerate(components):
```

```
    for idx in comp:
        group_id[idx] = cid

merged_trails["group_id"] = merged_trails.index.map(group_id)

# Final dissolve
print("Dissolving by group id")
result = merged_trails.dissolve(by="group_id")

# Add length field and calculate length of each trail
print("Adding length field and calculating trail length")
result["length"] = result.geometry.length

# Save result to gpkg
print("Outputting dissolved trails to Trails.gpkg")
result.to_file('sf_trail_access_qgis/data/Trails.gpkg', driver='GPKG',
layer='dissolved_trails')
```

Dissolving trails by grouping in Python. Code written with assistance from ChatGPT

Figure 26: Extracting Transit Network from r5r to geopackage

```
# extract transit network from r5r_network

transit_network <- transit_network_to_sf(r5r_network)



# Write both layers to the same GeoPackage

st_write(transit_network$stops, "network_output/transit_network.gpkg", layer =

"stops")

st_write(transit_network$routes, "network_output/transit_network.gpkg", layer =

"routes", append = TRUE)
```

Extracting Transit Network from r5r to geopackage

Figure 27: Checking for Origin/Destination Snapping Issues to Network in R

```
# check for snapping issues with origins and destinations for routing
# check snapping using walk mode
# generates a dataframe, found == FALSE indicates snapping issues
snapped_origins <- find_snap(r5r_network = r5r_network, points =
origins, mode = "WALK")
snapped_destinations <- find_snap(r5r_network = r5r_network, points =
destinations, mode = "WALK")

# filter to only origins and destinations that were unsuccessfully
snapped
nonsnapped_origins <- snapped_origins |>
  filter(found == FALSE)
nonsnapped_destinations <- snapped_destinations |>
  filter(found == FALSE)

print(paste0("nonsnapped origins: ", nrow(nonsnapped_origins),
             " nonsnapped destinations: ",
nrow(nonsnapped_destinations)))
```

Checking for Snapping Issues in R.

Figure 28: Creating Fields for SoVI Scoring

```
import arcpy

# --- User inputs ---
gdb_path = r"C:\GEOG 578 Projects\Final Project\Final Project.gdb"
fc_name = "SF_BG_SoVI"
income_field = "F__No_Car"         # existing numeric field
new_field = "Vehicle"        # new field to store the result

# --- Build full path to the feature class ---
fc = f"{gdb_path}\\{fc_name}"

# --- Add new field if it doesn't exist ---
fields = [f.name for f in arcpy.ListFields(fc)]
if new_field not in fields:
    arcpy.AddField_management(fc, new_field, "SHORT")
```

```python
# --- Update rows ---
with arcpy.da.UpdateCursor(fc, [income_field, new_field]) as cursor:
    for row in cursor:
        income_value = row[0]

        # Handle NULL values
        if income_value is None:
            row[1] = 0
        elif income_value > 41: #Q3
            row[1] = 2
        elif income_value > 23: #Median
            row[1] = 1
        else:
            row[1] = 0

        cursor.updateRow(row)

print("Done!")
```

Example code for creating a new field called Income and populating it with 2 for values > the third quartile, 1 for values > the median, and 0 for values <= the median or NULL. This code was created with the help of ChatGPT. Was also used for Race and Vehicle variables with small changes to variable names and median and Q3 values.

Figure 29: Creating the Travel Time Matrix

```r
# setup/libraries
rJavaEnv::use_java("21")
options(java.parameters = "-Xmx6G")
library(r5r)
library(sf)
library(dplyr)

# build r5r network, uses cached 'network.dat' if previously built
r5r_network <- r5r::build_network("network_inputs")

# Routing setup
trip_start_time <- as.POSIXct("11-01-2025 09:00:00", format = "%m-%d-%Y %H:%M:%S")
origins <- sf::st_read("data/blck_grp_centroid_wgs84.geojson")
destinations <- sf::st_read("data/custom_trailheads_wgs84.geojson")

# rename fid to id for origins and destinations, requirement of r5r
origins <- origins |>
```

```r
  rename(id = fid)
destinations <- destinations |>
  rename(id = fid)

# create expanded travel time matrix
travel_matrix <- r5r::expanded_travel_time_matrix(
  r5r_network = r5r_network,
  origins = origins,
  destinations = destinations,
  mode = "TRANSIT",
  mode_egress = "WALK",
  departure_datetime = trip_start_time,
  time_window = 1L, # only start at 9:00
  breakdown = TRUE,
  max_walk_time = 10L, #max 10 minute walk per leg
  max_trip_duration = 150L, #150 minutes (2.5 hours)
  walk_speed = 4.83,
  max_rides = 6,
)

# export matrix to csv
write.csv(travel_matrix, "expanded_travel_time_matrix.csv", row.names = FALSE)

# Cleanup
r5r::stop_r5(r5r_network)
rJava::.jgc(R.gc = TRUE)
```

R script to generate an expanded travel time matrix using operationalized variables as detailed in the Public Transportation Suitability section.

Figure 30: Encoding Trail Accessibility Scoring to Block Groups

```r
# library
library(sf)

# Open csv of travel time matrix
travel_matrix <- read.csv("expanded_travel_time_matrix.csv")

# build scoring
block_scores <- travel_matrix |>
  rename(block_id = from_id) |>
  filter(n_rides > 0) |>
  mutate(
    # use 15 minute bins, with 0-15 receiving a score of 10, 15-30 of 9,...
    duration_score = pmax(1, 10 - pmin(floor(total_time / 15), 9)),
    # 6 points for 1 ride, 5 points for 2, ...
```

```r
    rides_score = pmax(0, 6 - n_rides),
    total_score = duration_score + rides_score
  ) |>
  summarise(
    sum_score = sum(total_score),
    avg_score = round(mean(total_score), 1),
    num_trails = n(),
    .by = block_id
  )

# create table of all origin points
origins <- sf::st_read("data/blck_grp_centroid_wgs84.geojson")
origins <- origins |>
  rename(block_id = fid) |>
  select(block_id, GISJOIN)


# join scores to all origins and fill in nulls with 0s
complete_scores_sf <- origins |>
  left_join(block_scores) |>
  mutate(
    # Fill nulls with 0, otherwise retain values
    sum_score = ifelse(is.na(sum_score), 0, sum_score),
    avg_score = ifelse(is.na(avg_score), 0, avg_score),
    num_trails = ifelse(is.na(num_trails), 0, num_trails)
  ) |>
  select(-block_id)

# export sf object to geojson
sf::st_write(complete_scores_sf, "data/sf_bg_trail_access_scoring_wgs84.geojson")
```

R script to encode trail accessibility scoring.

Figure 31: Cast Trail Scoring to 0-3 Values in QGIS Field Calculator

```sql
-- used in QGIS Field Calculator to cast trail scoring to a new scoring field
CASE
    WHEN "trail_avg_score" <= q1("trail_avg_score") THEN 3
    WHEN "trail_avg_score" > q1("trail_avg_score")
        AND "trail_avg_score" <= median("trail_avg_score") THEN 2
    WHEN "trail_avg_score" > median("trail_avg_score")
        AND "trail_avg_score" <= q3("trail_avg_score") THEN 1
    WHEN "trail_avg_score" > q3("trail_avg_score") THEN 0
END
```

SQL Expression used in QGIS Field Calculator to cast trail scoring to 0-3 values.