## Team: Martin Liu, Giuseppe Peronda

## Descriptive Title: Determining bikeability of census tract regions below 400% poverty line

## Introduction

The premise of the project is based around Assembly Bill 117, which will allocate $10 million dollars in E-bike vouchers to households at 400% of the federal poverty level. That's $51,000 for a single person and $106,000 for a family of four. Purchasable bikes include Class 1 and 2 e-bikes.

The project aims to evaluate the bikeability of census tracts that fall under this income level bracket and determine if the area around them are bike friendly as this would be a factor in determining how effective this bill would be.

We currently have census tract data from the US Census Bureau and bike street data from OSMNX. Census data is used to determine which census tracts the project will be focusing on. OSMNX will used to determine various factors about the bike network as explained in Data and Methods

## Research Papers

*The Missing Link, Bicycle Infrastructure Networks and Ridership in 74 US Cities*
This paper aims to develop a standard methodology for measuring bicycle facility network quality of 74 US cities.
Some things to note about the paper. Firstly, only cities that had a population greater than 100,000 were chosen, second, trips that would involve traveling within the city were selected. We will use two indices that were used as metrics for evaluating bike network in the city, including

$$\beta = edges/vertices \quad (1)$$

Beta index; a low β suggests an increased chance that any given route requires leaving dedicated bike infrastructure to ride with mixed traffic

$$\gamma = edges/ 3 * (vertices-2) \quad (1)$$

Gamma index; edges vs theoretical edges. A Higher gamma values indicate greater internal connectivity and increased redundancy.

*Mapping bikeability: a spatial tool to support sustainable travel*
In this paper, Winters et al. survey Vancouver residents to see what factors are important in determining bikeability. They then collect the objective data behind these factors, and as seen in Figure 3, they weight each of these factors in accordance to its importance as determined by the survey, and assign a bikeability score to the whole city. We plan on using many of the same factors as Winters et al.; all of the factors shown, with the exception of bike route separation can be found via OSMNX. Additionally, the examples provided by Geoff Boeing give the code necessary to extract these statistics.

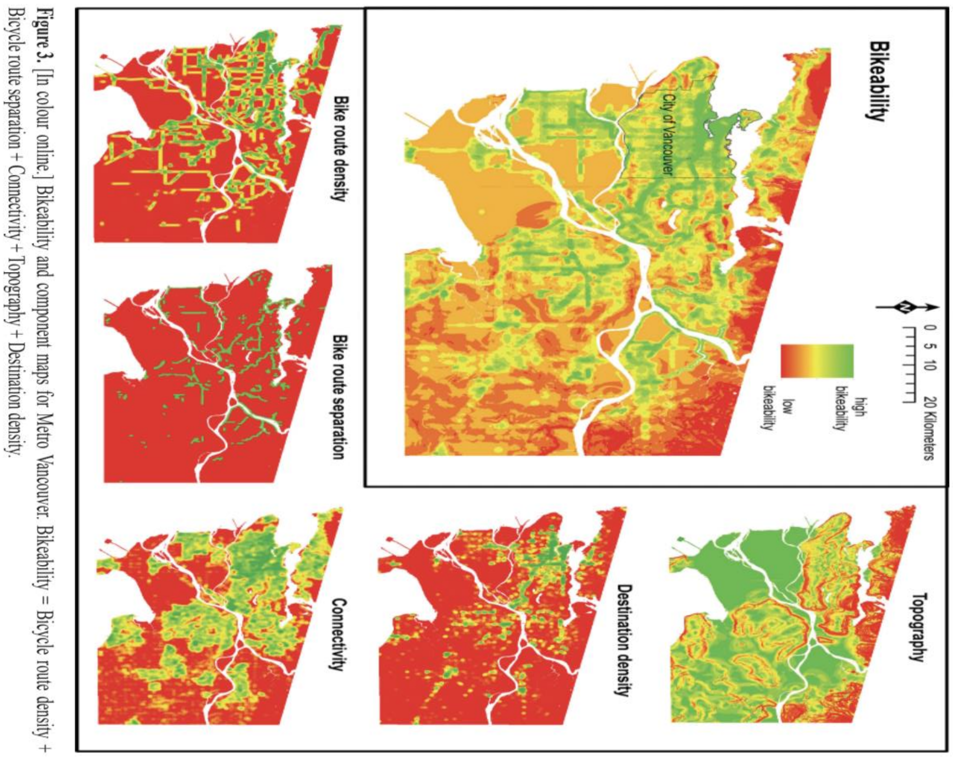



*Figure 1: Map describing bikeability depending on the region of Vancouver*

# Data and Methods (3pts)

## Census Data

Census data are collected using the cenpy package, which allows us to query the census api to find both ACS and decennial census data. We query the 2019 ACS at the census census tract level, fetching the variables corresponding to total population, population in poverty, and number of people without access to a vehicle. Additionally, cenpy gives geometric data, giving us the polygons that define the area covered by each tract. The resulting data structure from the query is a geopandas GeoDataFrame, and some statistics can be called on this geometric data, such as the area, which we use to then calculate the population density of each tract. We can also calculate the poverty rate by dividing the number of people in poverty by the tract population.

We then select the top 25th percentile of the tracts by poverty rate, and filter out tracts in the bottom 25% by population density. This leaves us with poor urban tracts, which would receive the bulk of the money given by AB 117. Plots showing the poverty rates of the census tracts we selected in Los Angeles and the Bay Area are shown in Figure 2 and 3 below, and the one showing every tract in California we selected is in the Appendix.

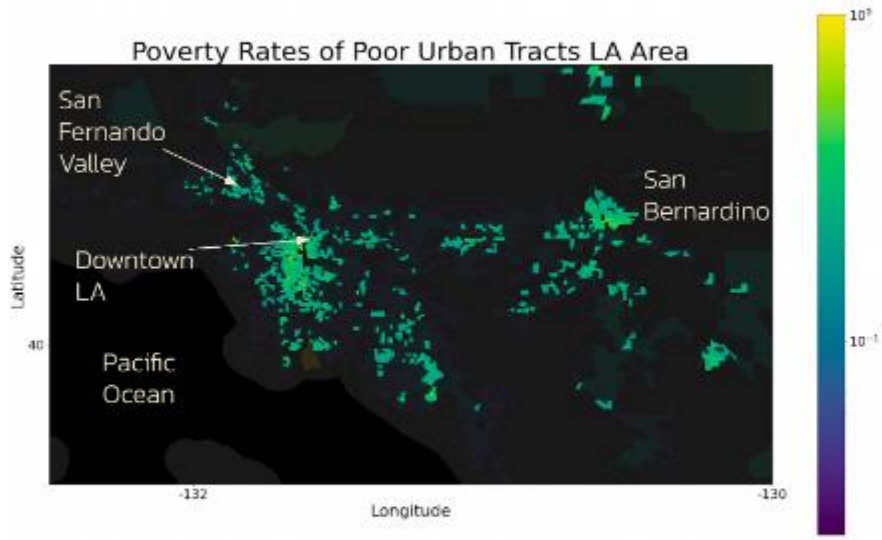



*Figure 2: Map of Southern California counties showing Urban Poverty Tracts*


*Figure 3: Map of San Francisco counties showing Urban Poverty Tracts*

## Points of Interest Data

Another factor we used to evaluate the bikeability of our census tracts was the number of points of interest (POIs) they contained. Via OSMNX, we were able to query the 2km by 2km region around the center of each census tract for buildings and facilities that could serve as destinations for cyclists. These included most commercial, civic, and sports buildings, as well as several amenities, including bicycle parking, and bike share facilities. An example of what this looks like can be seen in Figure 4 below. Following the logic of Winters et al., we believe that neighborhoods with greater POI density are more bikeable.

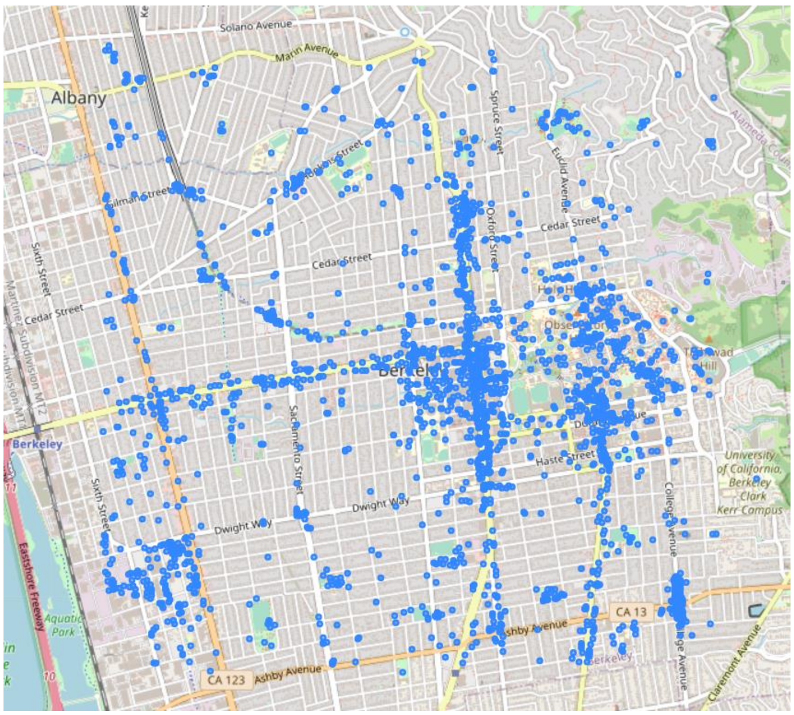



Figure 4: 2km by 2km Map of Berkeley's POI locations

## Bike Street Network

Bike street network data is used for the project. The bike network data comes from OpenStreetMaps and the Python package OSMNX. The nodes are intersections while the edges are the bike paths. The graphs generated are undirected and weighted based on distance. Depending on the network location the graph may or may not have a small-world property.

## Basic Network Statistics

The project aims to use certain network statistics as a means of grading the bikeability of the various census tracts. These statistics are broken down into the following factors

- Number of nodes
- Number of edges
- Average degree
- Beta Index
- Gamma Index
- Highest betweenness
- Average length of edges

Using the data and statistics outlined above, we will adopt an approach similar to that described by Winters et al. We will also incorporate max betweenness centrality of each node as a measure. We argue that tracts with a few nodes of high betweenness (like the one shown to the right) are less bikeable, as this implies a less permeable tract, giving longer travel times, which are especially impactful to cyclists, who tend to travel at lower speeds and over shorter distances than drivers. An example of this can be seen in Figure 5 which shows a map that has a high betweenness and is shown by the bright yellow dot.

*Figure 5: Map of network with a node with high betweenness*

# Results (8 pts)

To begin with, we look at various cities, capturing a point of interest with radius 2km. Cities with higher density as that has been shown to likely lead to a better bike network. (Schoner et.al, 2014). These cities are sampled based on  data from the Decennial Census by Decade.

Table 1 shows basic statistics of various densely populated census tract regions. These statistics will serve as a benchmark when evaluating other census tract regions.

Table 1: Basic Statistics of densely populated census tract region

| | friendship_village | bellerose_terrace | walnut_park | loch_lomond | ewa_gentry | rollingwood | isla_vista | east_los_angeles |
|---|---|---|---|---|---|---|---|---|
| **node** | 1340.000000 | 750.000000 | 1271.000000 | 688.000000 | 470.000000 | 1253.000000 | 2114.000000 | 1113.000000 |
| **edges** | 3587.000000 | 1863.000000 | 3308.000000 | 1690.000000 | 979.000000 | 2920.000000 | 4696.000000 | 2852.000000 |
| **k_avg** | 5.354000 | 4.968000 | 5.205000 | 4.913000 | 4.166000 | 4.661000 | 4.443000 | 5.125000 |
| **edge_length_total** | 244275.200000 | 149287.518000 | 247872.199000 | 110011.806000 | 93250.301000 | 170305.732000 | 167193.499000 | 217801.448000 |
| **edge_length_avg** | 68.100000 | 80.133000 | 74.931000 | 65.096000 | 95.251000 | 58.324000 | 35.603000 | 76.368000 |
| **streets_per_node_avg** | 3.004000 | 3.092000 | 2.910000 | 2.814000 | 2.440000 | 2.808000 | 2.351000 | 2.868000 |
| **intersection_count** | 1198.000000 | 668.000000 | 1116.000000 | 582.000000 | 312.000000 | 1042.000000 | 1359.000000 | 918.000000 |
| **street_length_total** | 129985.544000 | 87713.257000 | 129753.314000 | 62037.753000 | 52712.986000 | 96882.460000 | 89866.441000 | 118016.756000 |
| **street_segment_count** | 1948.000000 | 1101.000000 | 1773.000000 | 939.000000 | 535.000000 | 1715.000000 | 2474.000000 | 1525.000000 |
| **street_length_avg** | 66.728000 | 79.667000 | 73.183000 | 66.068000 | 98.529000 | 56.491000 | 36.324000 | 77.388000 |
| **circuity_avg** | 1.056000 | 1.023000 | 1.033000 | 1.145000 | 1.066000 | 1.111000 | 1.087000 | 1.042000 |
| **self_loop_proportion** | 0.007000 | 0.001000 | 0.003000 | 0.006000 | 0.002000 | 0.008000 | 0.006000 | 0.003000 |
| **beta** | 2.676866 | 2.484000 | 2.602675 | 2.456395 | 2.082979 | 2.330407 | 2.221381 | 2.562444 |
| **gamma** | 0.893622 | 0.830214 | 0.868926 | 0.821186 | 0.697293 | 0.778044 | 0.741162 | 0.855686 |

Filtering out the census tracts using the methodology in Census Tracts, we get around 2000-3000 unique census tracts. As a starting point, we decided to take a sample of 50 census tracts, graph out their network of 2km and collate the following statistics.

Table 2: Basic Statistics of densely populated low income census tracts (8 tracts shown)

| | Census Tract 1895, Los Angeles County | Census Tract 4623.01, Los Angeles County | Census Tract 26.01, Fresno County | Census Tract 414.04, Riverside County | Census Tract 1852.02, Los Angeles County | Census Tract 91.09, Sacramento County | Census Tract 46.01, Sacramento County | Census Tract 62.02, Sacramento County |
|---|---|---|---|---|---|---|---|---|
| node | 960.000000 | 1912.000000 | 1021.000000 | 739.000000 | 620.000000 | 927.000000 | 1451.000000 | 688.000000 |
| edges | 2354.000000 | 4529.000000 | 2956.000000 | 1817.000000 | 1654.000000 | 2433.000000 | 3805.000000 | 1660.000000 |
| k_avg | 4.904000 | 4.737000 | 5.790000 | 4.917000 | 5.335000 | 5.249000 | 5.245000 | 4.826000 |
| edge_length_total | 180358.791000 | 275557.346000 | 255125.728000 | 138125.687000 | 162026.796000 | 192396.524000 | 259102.384000 | 111640.578000 |
| edge_length_avg | 76.618000 | 60.843000 | 86.308000 | 76.019000 | 97.961000 | 79.078000 | 68.095000 | 67.253000 |
| streets_per_node_avg | 2.678000 | 2.612000 | 3.379000 | 2.812000 | 2.834000 | 2.819000 | 2.810000 | 2.709000 |
| intersection_count | 762.000000 | 1416.000000 | 981.000000 | 615.000000 | 531.000000 | 795.000000 | 1230.000000 | 556.000000 |
| street_length_total | 93728.600000 | 148330.780000 | 141198.424000 | 76151.716000 | 82557.344000 | 99405.835000 | 134020.982000 | 58682.389000 |
| street_segment_count | 1232.000000 | 2424.000000 | 1663.000000 | 1004.000000 | 846.000000 | 1254.000000 | 1990.000000 | 889.000000 |
| street_length_avg | 76.078000 | 61.193000 | 84.906000 | 75.848000 | 97.586000 | 79.271000 | 67.347000 | 66.009000 |
| circuity_avg | 1.138000 | 1.062000 | 1.018000 | 1.107000 | 1.105000 | 1.095000 | 1.097000 | 1.099000 |
| self_loop_proportion | 0.007000 | 0.004000 | 0.001000 | 0.002000 | 0.001000 | 0.004000 | 0.008000 | 0.006000 |
| median_income | 54265.000000 | 57898.000000 | 31982.000000 | 84286.000000 | 64623.000000 | 83620.000000 | 41047.000000 | 37031.000000 |
| agg_travel_time_min | 67505.000000 | 75885.000000 | 30110.000000 | 56270.000000 | 60850.000000 | 66240.000000 | 83460.000000 | 34165.000000 |
| pct_workers_who_commute | 0.516000 | 0.502000 | 0.329000 | 0.476000 | 0.497000 | 0.465000 | 0.372000 | 0.369000 |
| beta | 2.452083 | 2.368724 | 2.895201 | 2.458728 | 2.667742 | 2.624595 | 2.622329 | 2.412791 |
| gamma | 0.819068 | 0.790401 | 0.966961 | 0.821800 | 0.892125 | 0.876757 | 0.875316 | 0.806608 |

As can be seen from Table 1 and Table 2. The values for most of the statistics are relatively similar, making it difficult to determine if the network is bike friendly or not.

The main way to counteract this issue was by looking at bike lanes that had cycleways as OSM classified any lane that could be biked on a bike lane. Doing so greatly reduced the number of nodes and edges, as seen in the figures below.

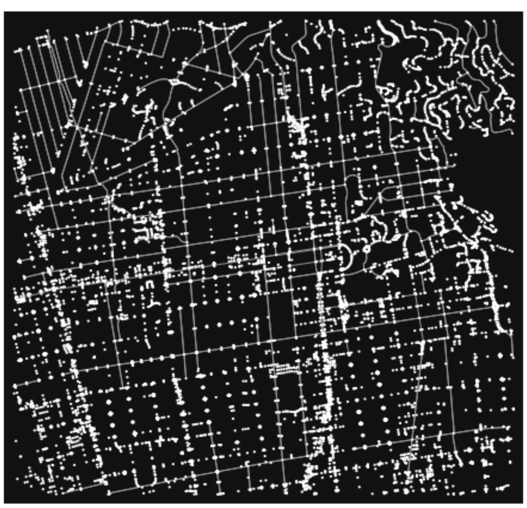


Fig 6: Map of Berkeley, 2km bounded box, all nodes and edges

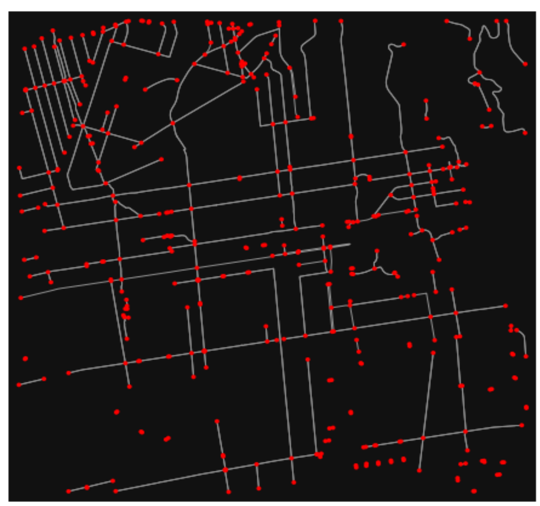


Fig 7: Map of Berkeley, 2km bounded box, cycleway only nodes and edges

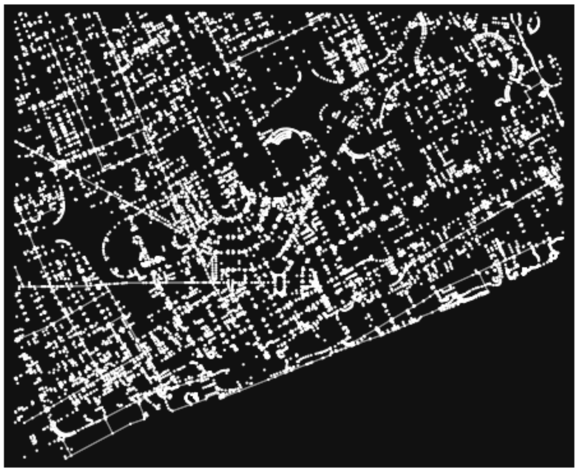


Fig 8: Map of Detroit, 2km bounded box, all nodes and edges

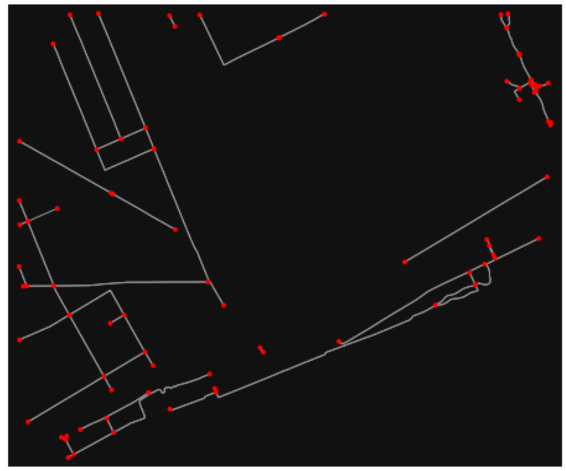


Fig 9: Map of Detroit, 2km bounded box, cycleway only nodes and edges

When looking at Detroit, which was known to not be bike friendly (Candiloro. T, 2022), we noticed that there were still recognizable grid lines and patterns which represent that the city was well connected albeit not for bikes.

Hence moving forward with our analysis we decided to ratio the cycleway network (red nodes) with the bike lane network (white nodes) as this can give us a better idea of how much more connectivity an area should provide to be more bike friendly.

Due to the small nature of cycleway networks, we decided to measure other statistics such as Degree, Beta index, and Gamma index solely based off the cycleway network (red nodes)

Table 3: Table of average statistics for bike friendly cities, non-bike friendly cities and census tracts

| | Bike friendly cities | Non- bike friendly cities | Census tracts (sample size 50) |
|---|---|---|---|
| Bike node ratio | 1.95% | 0.59% | 0.75% |
| Bike edge ratio | 1.7% | 0.47% | 0.58% |
| Total edge length ratio | 13.5% | 6.11% | 5.74% |
| Degree | 3.13 | 2.97 | 2.83 |
| Beta index | 1.56 | 1.48 | 1.41 |
| Gamma index | 0.52 | 0.52 | 0.51 |
| POI | 2287 | 417 | 159 |

From Table 3: we noticed that bike friendly cities on average have three times more cycleway lanes than non bike friendly cities. The degree, beta index and gamma index unfortunately are not distinguishable between the bike friendly and non bike-friendly cities. In terms of the number of relevant POI, bike friendly cities have about 5 times more than non-bike friendly cities.

For the averaged census tract sample, cycleway infrastructure is slightly higher than non-bike friendly cities. Degree, beta and gamma index are relatively similar and POI is significantly lower than the rest.

We noticed that for some census tracts there would be as low as 2 POIs or have barely any cycleway nodes. Therefore we decided to take a look at the top 6 census tracts sorted based on the number of cycleway nodes. In addition, due to the alpha and beta index being indistinguishable between different city types, we decided to exclude them from the analysis and instead used max betweenness. Since betweenness measures how the percentage amount of times a node is present in a shortest path this helps us gauge if there are multiple routes a cyclist can take in that area or just one.

| Census Name | Bike nodes | Bike edges | Average degree | POIs | Betweenness | Median Income |
|---|---|---|---|---|---|---|
| Census Tract 2949, LA | 190 | 456 | 4.8 | 62 | 6% | 42150 |
| Census Tract 2655.10, LA | 175 | 346 | 3.95 | 770 | 8% | 54847 |
| Census Tract 4623.01, LA | 129 | 264 | 4.09 | 416 | 9% | 57898 |
| Census Tract 4020.01, LA | 117 | 185 | 3.162 | 188 | 3.5% | 83409 |
| Census Tract 231.02, SF | 108 | 127 | 2.35 | 239 | .0.5% | 37981 |
| Census Tract 1104, SC | 94 | 181 | 3.85 | 107 | 33% | 59750 |

Nodes, edges and average degree follow similar patterns as their values generally go down with the number of nodes. The goal of this table is to highlight certain anomalies, For instance Census Tract 1104 has a higher max betweenness compared to the other tracts. A closer look at the network can be seen in Figure 10 below.

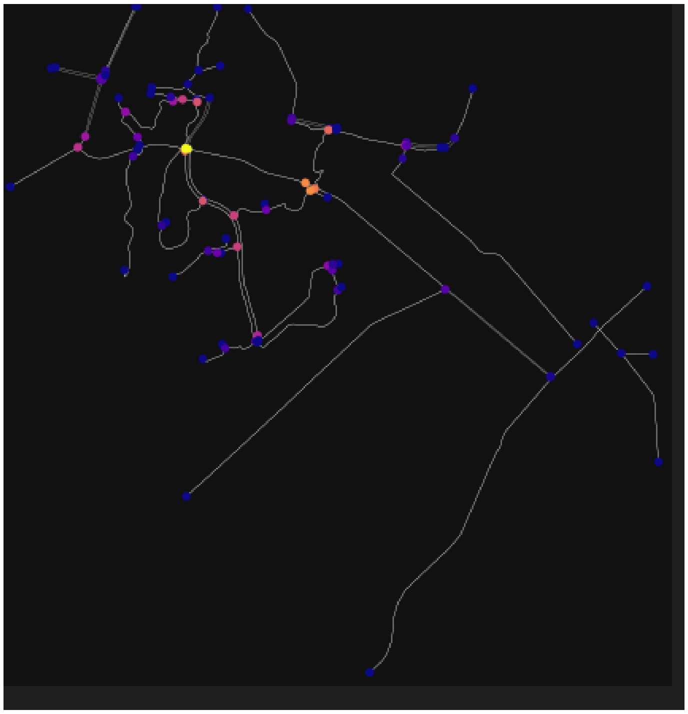


Figure 10: Network of Census Tract 1104 highlighting betweenness of nodes

Figure 10 above shows the max betweenness relative to the other nodes. The yellow node has the highest max betweenness as it connects many bike routes together. Hence a potential fix for this census tract could be to build or retrofit cycleways outside of the yellow node to free up congestion in that node.

## Future Work and Conclusions (2 pts)

Due to computational and time constraints an analysis on all 2000 census tracts proved to be extremely difficult. However a reasonable alternative would be to only select tracts that have a certain number of nodes as this would suggest that there is a reasonable bike network and more importantly there is data on OSM about this census tract. This selection can be further streamlined using other factors like median income. From there further analysis regarding what type of bike infrastructure, how much to build and where to build it can be analyzed. In terms of this paper, the POI count and ratios of cycle lanes to bike lanes for bike friendly cities can serve as a reasonable benchmark that these neighborhoods can aim for.

In addition, it was also difficult to conduct our analysis on larger samples of the data, which would have been necessary to create a bikeability map in the style of Winters et al. The main bottleneck in this regard were the queries to OSMNX, whose speed is unlikely to improve due to the large amount of data we request. A possible workaround would be to query fewer building types, however, this would introduce some bias, as some tracts could have more of one type of building than others, so some deeper analysis would have to be done before this measure was taken. Another idea would be to only query some sample of the buildings matching our filters within a census tract, but we have not found a way to do this via OSMNX.

Lastly, another factor to take into consideration would be the quality of cycling infrastructure to our calculations. As it stands, there is no categorization of bike lanes and cycleways in our project, they are simply included or excluded. But the degree of protection from vehicular traffic likely plays an important role in determining how willing people are to bicycle in these neighborhoods.

## References:

Schoner, J.E., Levinson, D.M. The missing link: bicycle infrastructure networks and ridership in 74 US cities. *Transportation* 41, 1187–1204 (2014). https://doi.org/10.1007/s11116-014-9538-1

T. Candiloro (2022) https://anytimeestimate.com/research/most-bike-friendly-cities-us-2022/

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2013). Mapping Bikeability: A Spatial Tool to Support Sustainable Travel. Environment and Planning B: Planning and Design, 40(5), 865–883. https://doi.org/10.1068/b38185

https://www.census.gov/programs-surveys/decennial-census/decade.2010.html

We would also like to thank Jackson Herron for sharing his cenpy code with us!

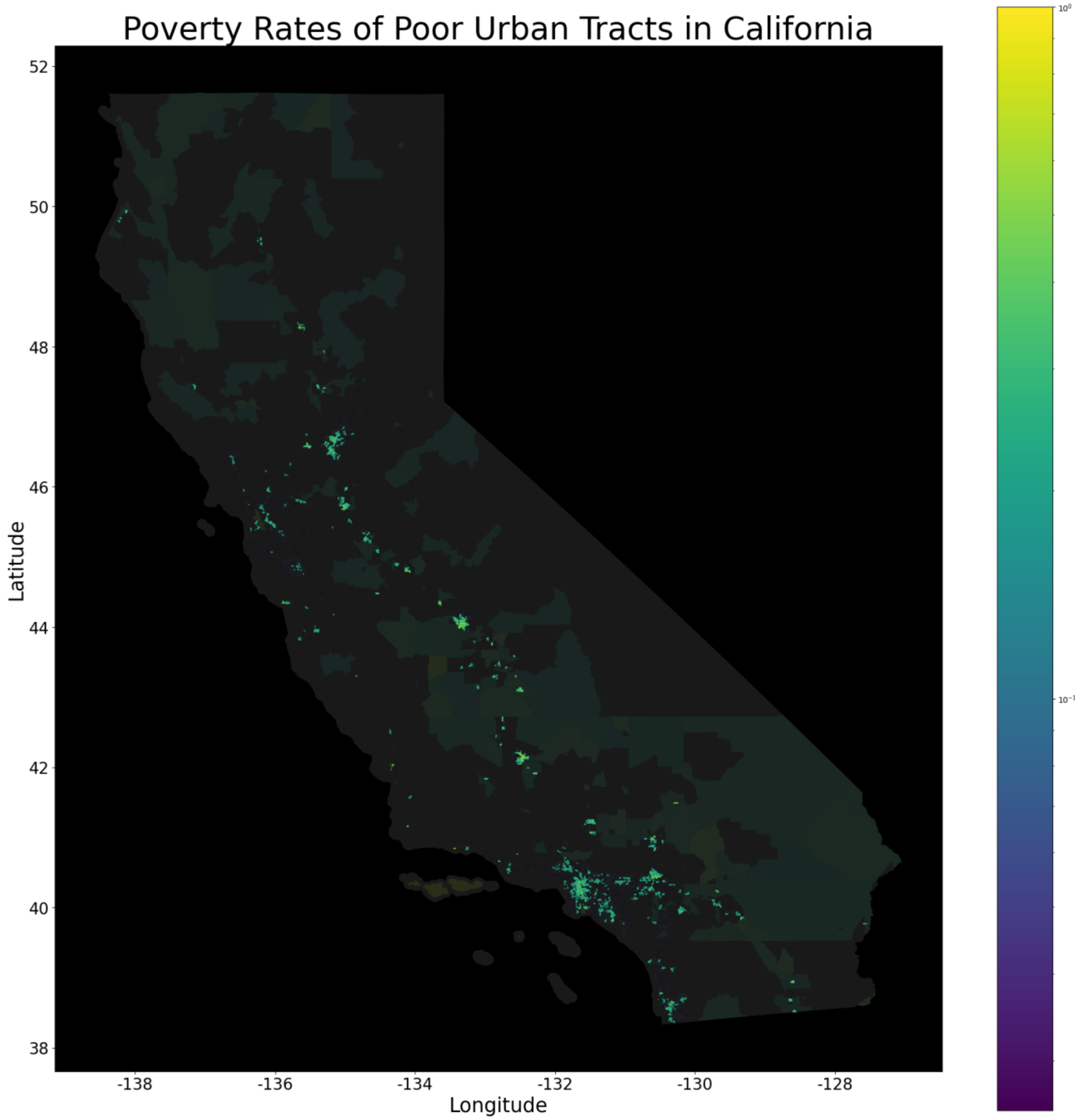
## Github Repository:

Our code can be found here: https://github.com/g-perona/CE-263-Final-Project
For network analysis:
Main -> notebooks -> Project_martin.ipynb

For POI analysis
Main -> Giuseppe Work -> poi_exploration.ipynb

For mapping analysis
Main -> Giuseppe Work -> census_data_collection.ipynb

# Appendix:



Poverty Rates of Poor Urban Tracts in California
Latitude
Longitude
52
50
48
46
44
42
40
38
-138
-136
-134
-132
-130
-128
$10^0$
$10^{-1}$