SOCIOECONOMIC FACTORS IN RELATION TO PERCENTAGE OF GREEN SPACE IN CALIFORNIA CITIES

Madi Arndt, Thomas Burgess, Brianna Lee

1. INTRODUCTION

- Our project focuses on the relationship between different socioeconomic variables and the percentage of green space in three different cities in California.
- We used Landsat 8 data from between April 1st, 2020 and June 12st, 2020 in order to calculate the percent green space of San Diego, San Francisco, and Los Angeles. We did this by using the "red edge" that vegetation shows in the Near-Infrared wavelength. We chose images taken in the spring time because that is when the red edge is the clearest.
- We first downloaded images of our respective study sites from the USGS website, then used shapefile polygons of our cities in order to crop that data to just the relevant study site. Then, we used the bands to calculate NDVI of the sites.
- We acquired census data for the different socioeconomic variables in the graphs and results
- Our research questions were:
  - How do different socioeconomic variables affect the percentage of green space available within a city?
  - How are vulnerable populations affected by a lack of green space?
  - What is the statistical correlation between median income and percentage of green space available in a city?
  - Future work: How do different cities compare to the findings of these three California cities?

2. METHODS

Pre-Processing (Perform on each study site and for each Landsat image)

1. Download Landsat 8 (OLI) images for each site.
2. Create a RasterStack for each image using bands 2-7.
3. Import shapefile delineating each city's boundary.
4. Assign correct coordinate reference system to city shapefile and Landsat image.
5. Crop and mask Landsat image to extent of study site.
6. Overwrite pixel values outside viable Landsat range (7273-43636).
7. Change scale factor using the equation (Landsat image * 0.0000275 – 0.2) * 100.
8. Plot final product to check accuracy.

Processing in R (Perform on each study site)

1. Create function calculating NDVI for each pixel in a RasterStack and creates new RasterLayer.
2. Apply function to each Landsat image we pre-processed.
3. Create RasterStack of the NDVI RasterLayers that we will use for analysis.
4. Save the new NDVI RasterStack as a GeoTIFF to import to QGIS.
5. Calculate mean NDVI for each pixel and add resulting values to new column in the data frame.
6. Omit NA values from the data frame.
7. Perform Binary Classification on the data frame: if NDVI >= 0.19: print "greenspace" ifelse: print "other."
8. Calculate % green space of study area: number of "green space" pixels / total pixels.

Processing in QGIS (Perform on each study site)

Map 1: NDVI Per Pixel

1. Download the GeoTiff of the NDVI Landsat image we created in R.
2. Create new layer with Raster Calculator using Bands 1 and 2 of the NDVI image to calculate mean NDVI.
3. Display this new image to view NDVI for each pixel in the study site.

Map 2: Binary Classification of Greenspace vs Non-Greenspace

1. Using the layer displaying mean NDVI, use Raster Calculator to create a new layer of values between 0.19 and 1.0 (i.e., areas of green space).
2. Display results on binary classification map – black areas in the image (value = 0) are "non-green space" areas and green areas (value = 1) are "green space" areas.

Post-Processing Analysis

Create 3 linear regressions displaying the following relationships among the three sites:

1. Household Median Income ($) vs. Green Space (%)
2. In Civilian Labor Force (%) vs. Green Space (%)
3. Population Density (per square mile) vs. Green Space (%)

3. MAPS OF MEAN NDVI AND THRESHOLD VALUES

SAN DIEGO NDVI | LOS ANGELES NDVI | SAN FRANCISCO NDVI

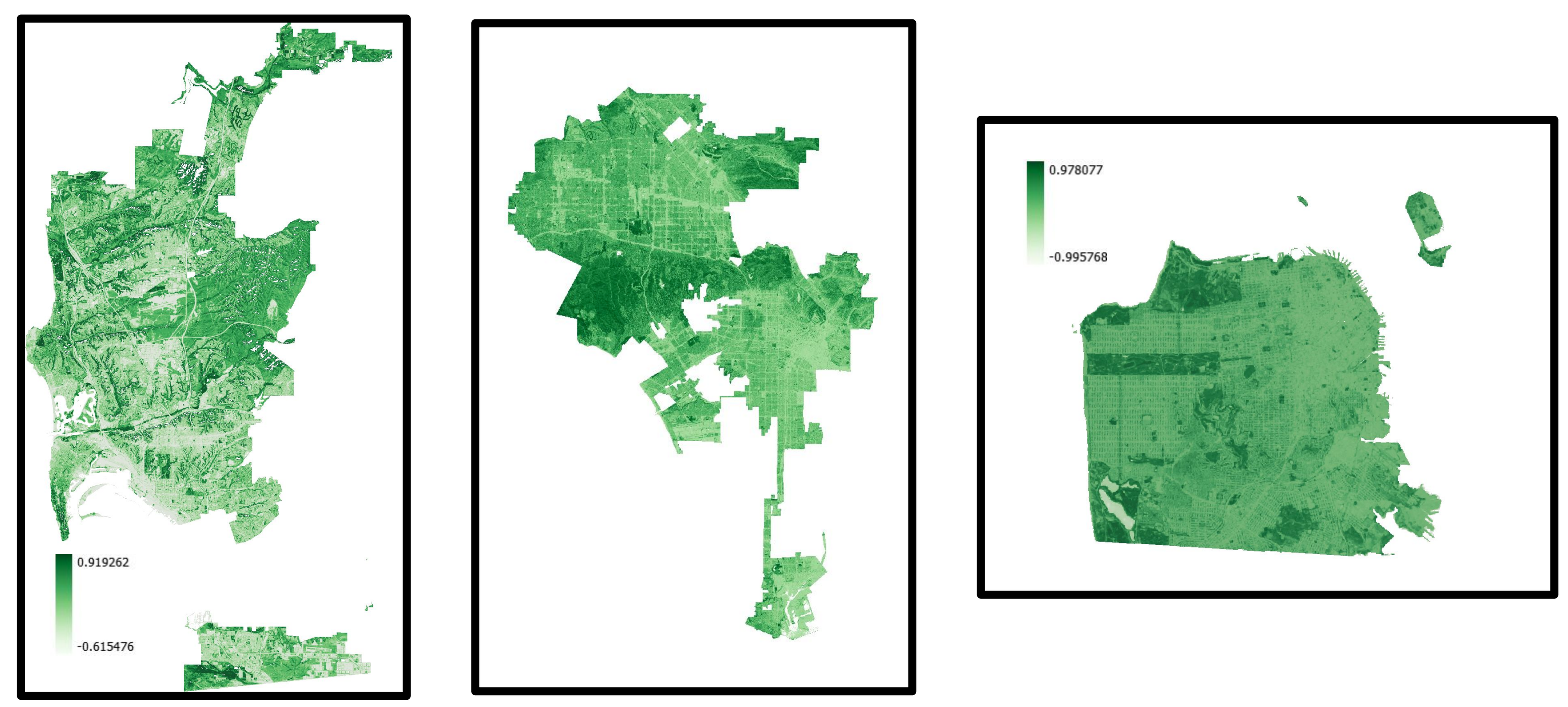

FIGURE 1: These maps display the mean NDVI value for each pixel. The colors correlate with NDVI values that range between 0 and 1, where lighter green characterizes areas of lower NDVI values and darker green characterizes areas of higher NDVI values.

SAN DIEGO THRESHOLD | LOS ANGELES THRESHOLD | SAN FRANCISCO THRESHOLD

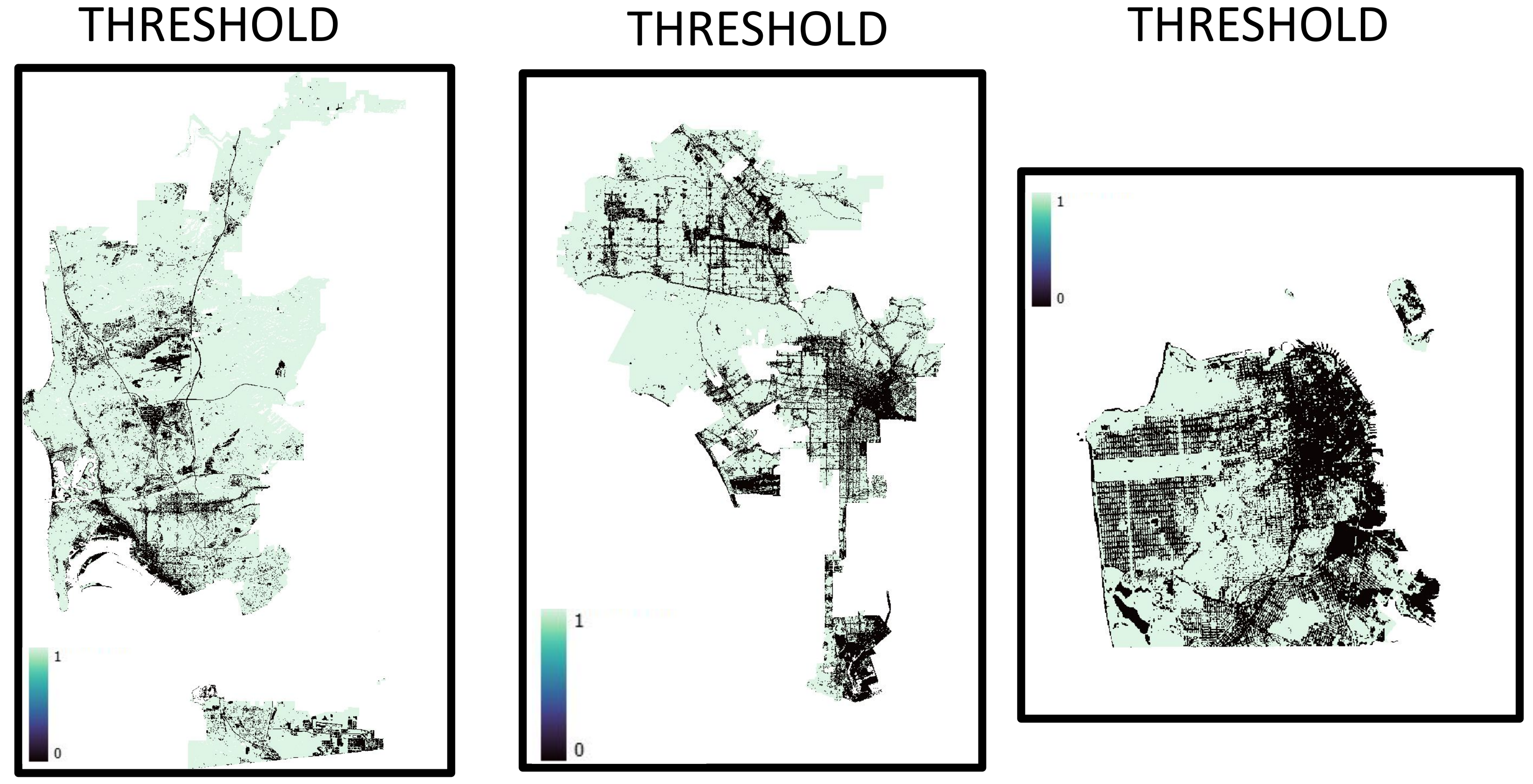

FIGURE 2: These maps show a binary classification of green space versus non-green space. An NDVI threshold was set at 0.19. Pixels with NDVI below this threshold are displayed as black and represent non-green space; pixels with NDVI above this threshold are displayed in a muted light green and represent areas of green space.

4. PLOTS OF SOCIOECONOMIC FACTORS VERSUS PERCENTAGE GREEN SPACE

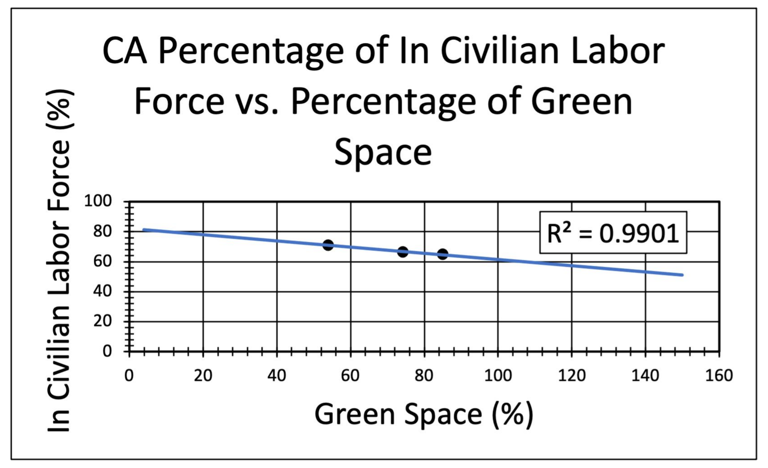

FIGURE 3: The percentage of in civilian labor force in San Diego, San Francisco, and Los Angeles have a strong negative correlation with the percentage of green space in 2020.

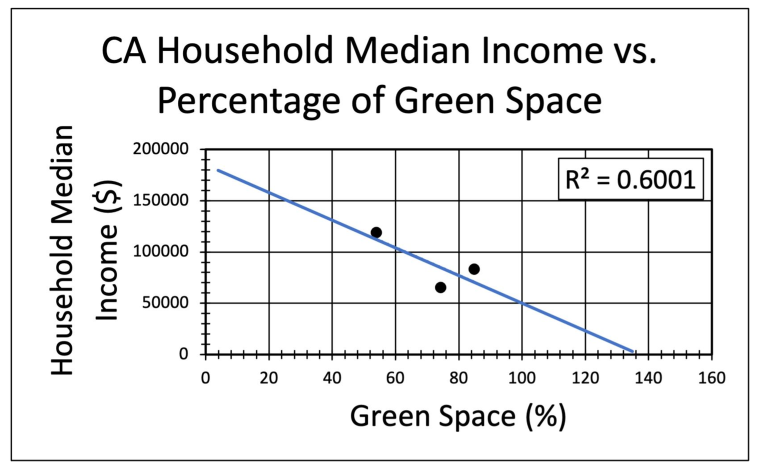

FIGURE 4: The household median income in San Diego, San Francisco, and Los Angeles have a negative correlation with the percentage of green space in 2020.

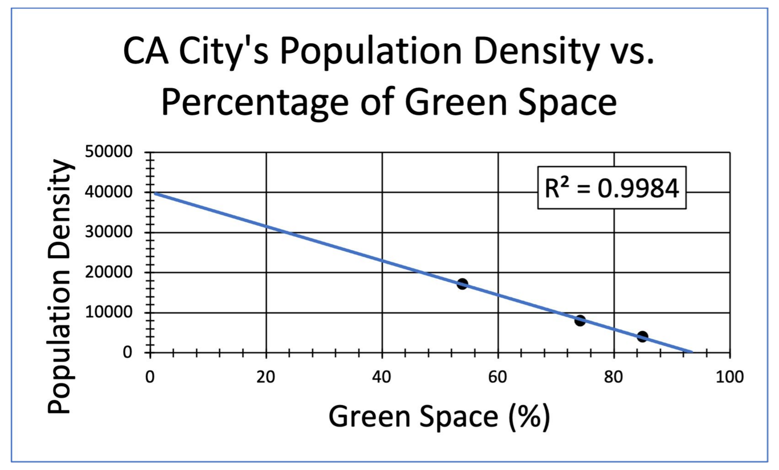

FIGURE 5: The population density in San Diego, San Francisco, and Los Angeles have a strong negative correlation with the percentage of green space in 2020.

5. DISCUSSION

- The percentage of in civilian labor force in San Francisco, Los Angeles, and San Diego has a strong negative correlation with the percentage of green space in those cities as the R^2 value is 0.9901. This strongly suggests that as more people are employed in these cities, there is a smaller percentage of green space in those cities.
- The household median income in San Francisco, Los Angeles, and San Diego has a negative correlation with the percentage of green space in those cities since the R^2 value is 0.6001. This suggests that as the median household income increases the less amount of green space is found in a city.
- The population density of San Francisco, Los Angeles, and San Diego have a strong negative correlation with the percentage of green space in those cities as the R^2 value is 0.9901. This strongly suggests that as population increases the less amount of green space found in that city.
- Binary classification by using a NDVI threshold value of 0.19 to identify green space versus other land cover was more efficient than using a classification tree

6. CONCLUSION

It is clear that population density is the most important factor of the three studied in determining the amount of green space in a city. San Francisco, while comparatively wealthy with the highest median household income of $119,136, still had significantly less green space at 54% than Los Angeles (household median income of $65290, 74% green space), and San Diego (household median income of $83454, 85% green space). San Diego had the highest percentage of green space, which makes sense as it is the least densely populated of the three cities.

There does appear to be a correlation between green space and all three of the socioeconomic variables. Our study took into account the socioeconomic conditions of each city, and thus made sure that all of the variables had significantly different values in each study site. Our figures show that there is a negative correlation between the variables and green space in all the cities, and in a future project we could add more cities and see if the trend continues with more than three data points. This project is a good starting point for many different green space studies, such as how urban green space interacts with different climates or how different ecologies rely on green space to survive in cities.

Future steps:

- Add more study sites to see how the statistical trend continues
- Study how different socioeconomic variables affect the different neighborhoods of cities
- Add different variables such as mental health and race to track how green space affects different social variables
- See how green space is affected by urbanization of cities over time by doing a temporal study of one city over time

7. REFERENCES

Aryal, J., Sitaula, C., & Aryal, S. NDVI Threshold-Based Urban Green Space Mapping from Sentinel-2A at the Local Governmental Area (LGA) Level of Victoria, Australia. *Land*, **11**, 351. (2022)

C. Huang, J. Yang, H. Lu, H. Huang, L. Yu, Green spaces as an indicator of Urban Health: Evaluating its changes in 28 mega-cities. *Remote Sensing* **9**, 1266. (2017)

D.R. Richards, P. Passy, R.R.Y. Oh, Impacts of population density and wealth on the quantity and structure of urban green space in tropical Southeast Asia. *Landscape and Urban Planning* **157**, 553–560 (2017).

Nouri, Hamideh, et al. "Effect of spatial resolution of satellite images on estimating the greenness and evapotranspiration of urban green spaces." *Hydrological Processes* 34.15 (2020): 3183-3199.

Nurdin, E. A., & Wijayanto, Y. (2020). The distribution of green open space in jember city area based on image landsat 8 - OLI. IOP Conference Series.Earth and Environmental Science, 485(1)

Rafiee, R., Mahiny, A. S., & Khorasani, N. (2009). Assessment of changes in urban green spaces of Mashad city using satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *11*(6), 431-438.

8. ACKNOWLEDGEMENTS

This research was supported by Professor Vena Chu of University of California, Santa Barbara. Landsat 8 OLI imagery was taken from the USGS Earth Explorer