Spatial Data Analysis of Metro Stations and Green Spaces in Los Angeles

Chenyi Weng

SSCI586 - GIS Programming and Customization

Professor Jennifer N. Swift

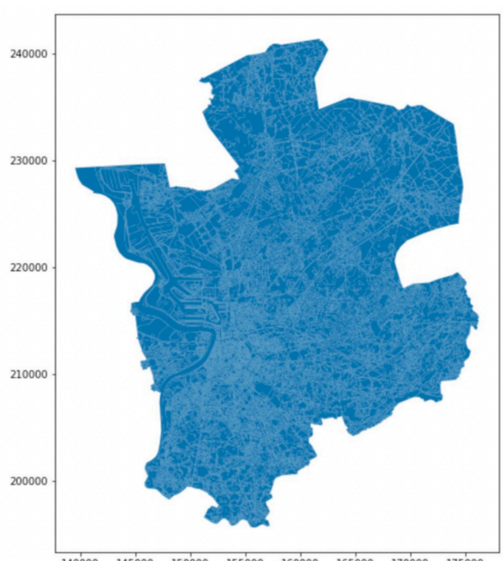
October 7th, 2024

1. Introduction

The aim of this project is to build upon the geospatial coding skills acquired through the initial Jupyter Notebook tutorial, which focused on spatial analysis in Antwerp, Belgium. After completing this first analysis, the techniques learned were applied to a new study area: Los Angeles. The focus of this second analysis is to explore the spatial relationship between metro rail stations and green spaces in Los Angeles.

2. Study Area

The first study area is Antwerp, a port city in Belgium, known for its well-defined urban and industrial spaces. The second study area is Los Angeles, California, USA, characterized by sprawling urbanization and a focus on transportation infrastructure. Los Angeles presents a unique case for analyzing the accessibility of green spaces, given the expanse of its metro system and the growing importance of sustainable, walkable urban environments. However, the spatial extent of both areas is illustrated in Figure 1 and Figure 2, which display basic maps of each study area, generated using GeoPandas.



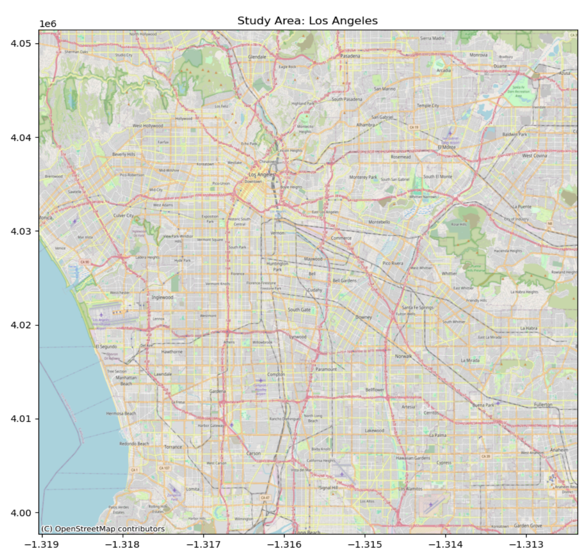


Figure 1. Study Area in Antwerp / Figure 2. Study Area in Los Angeles

3. Data and Data Processing

Name	**Spatial Coordinate System**	**Description**
landuse.shp	ESPG: 31370	Polygon shapefile representing land use coverage in the city of Antwerp.
layer_streets.shp	ESPG: 31370	Street line shapefile representing road networks in the city of Antwerp.
CRAB_subset.shp	ESPG: 31370	Address point shapefile showing building locations within Antwerp.
metro_rail_stops.csv	EPSG:4326 (WGS84)	CSV file containing metro station locations and details.

Table 1. Spatial Data and Coordinate Systems Overview

3.1. Metro Stations (Los Angeles)

Figure 3 illustrates the process of handling a .csv file containing geographic coordinates and metadata for metro stations in Los Angeles, sourced from the Los Angeles GeoHub. The file is first read using the pandas library, which includes station names, stop numbers, and geographic coordinates (X, Y values). These coordinates are then converted into geometrical points using the Shapely library, creating a new geometry column in a geopandas GeoDataFrame, with the coordinate reference system set to EPSG:2877. After processing, the data is previewed to ensure accuracy and saved as a .csv file, including the station names, stop numbers, and coordinates, as shown in the table at the bottom of the Figure 3.

[1]:
```
import pandas as pd
import geopandas as gpd
from shapely.geometry import Point

# Read CSV file
stops = pd.read_csv('/Users/mona/Downloads/Project3_1007_file/metro_rail_stops.
↪csv')

# Display the first few rows to check the content
print(stops.head())
```

```
                      Station  StopNumber            X            Y
0  Downtown Long Beach Station       80101 -13157175.78  3997701.981
1          Pacific Ave Station       80102 -13157262.50  3998262.680
2       Anaheim Street Station       80105 -13156782.04  3999544.611
3    Pacific Coast Hwy Station       80106 -13156781.82  4000517.002
4        Willow Street Station       80107 -13156832.14  4002926.773
```

[2]:
```
# Create geometry using X and Y coordinates
geometry = [Point(xy) for xy in zip(stops['X'], stops['Y'])]

# Create GeoDataFrame and set the coordinate system to EPSG:3857
gdf_stops = gpd.GeoDataFrame(stops, geometry=geometry, crs="EPSG:3857")

# Checkthe GeoDataFrame content
print(gdf_stops.head())
```

```
                      Station  StopNumber            X            Y  \
0  Downtown Long Beach Station       80101 -13157175.78  3997701.981
1          Pacific Ave Station       80102 -13157262.50  3998262.680
2       Anaheim Street Station       80105 -13156782.04  3999544.611
3    Pacific Coast Hwy Station       80106 -13156781.82  4000517.002
4        Willow Street Station       80107 -13156832.14  4002926.773

                          geometry
0  POINT (-13157175.780 3997701.981)
1  POINT (-13157262.500 3998262.680)
2  POINT (-13156782.040 3999544.611)
3  POINT (-13156781.820 4000517.002)
```

Station	StopNumber	geometry
Downtown Long Beach Station	80101	POINT (-118.1929209902119 33.7680710003546)
Pacific Ave Station	80102	POINT (-118.1937000092263 33.77225799866652)
Anaheim Street Station	80105	POINT (-118.1893839636121 33.78183000064353)
Pacific Coast Hwy Station	80106	POINT (-118.1893819873185 33.78908999671687)
Willow Street Station	80107	POINT (-118.1898340195695 33.80707900350561)
Wardlow Station	80108	POINT (-118.1960899770397 33.81986499791522)
Del Amo Station	80109	POINT (-118.2110170126213 33.848221999816)
Artesia Station	80110	POINT (-118.2225029616756 33.87608199789106)
Compton Station	80111	POINT (-118.2242490170933 33.8974899989506)
Willowbrook - Rosa Parks Station - Metro A-Line	80112	POINT (-118.2375550427447 33.92804799654965)
103rd Street / Watts Towers Station	80113	POINT (-118.2431590029817 33.94222000002048)

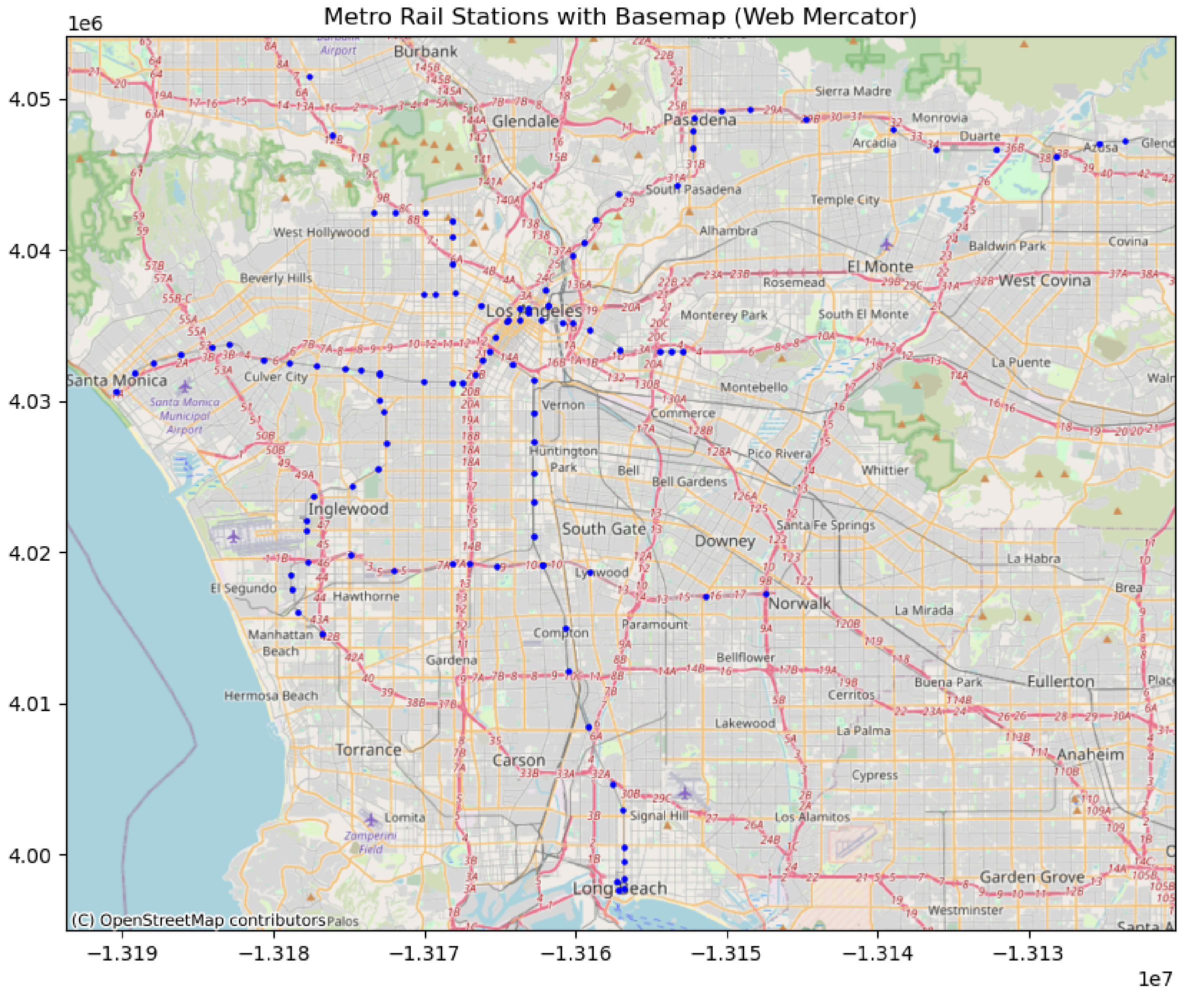


Figure 3. Metro Station Coordinates and Metadata Processing: metro_rail_stops.csv

3.2. Green Spaces

Data from OSMnx on parks, gardens, and natural green areas within Los Angeles. The analysis also uses land-use shapefiles for Antwerp to provide a comparison of urban fabric and green spaces. Additionally, the projection systems used include EPSG:3857 (Web Mercator) for visualization, later transformed into EPSG:4326 (WGS84) for geospatial analysis, as seen in Figure 4.

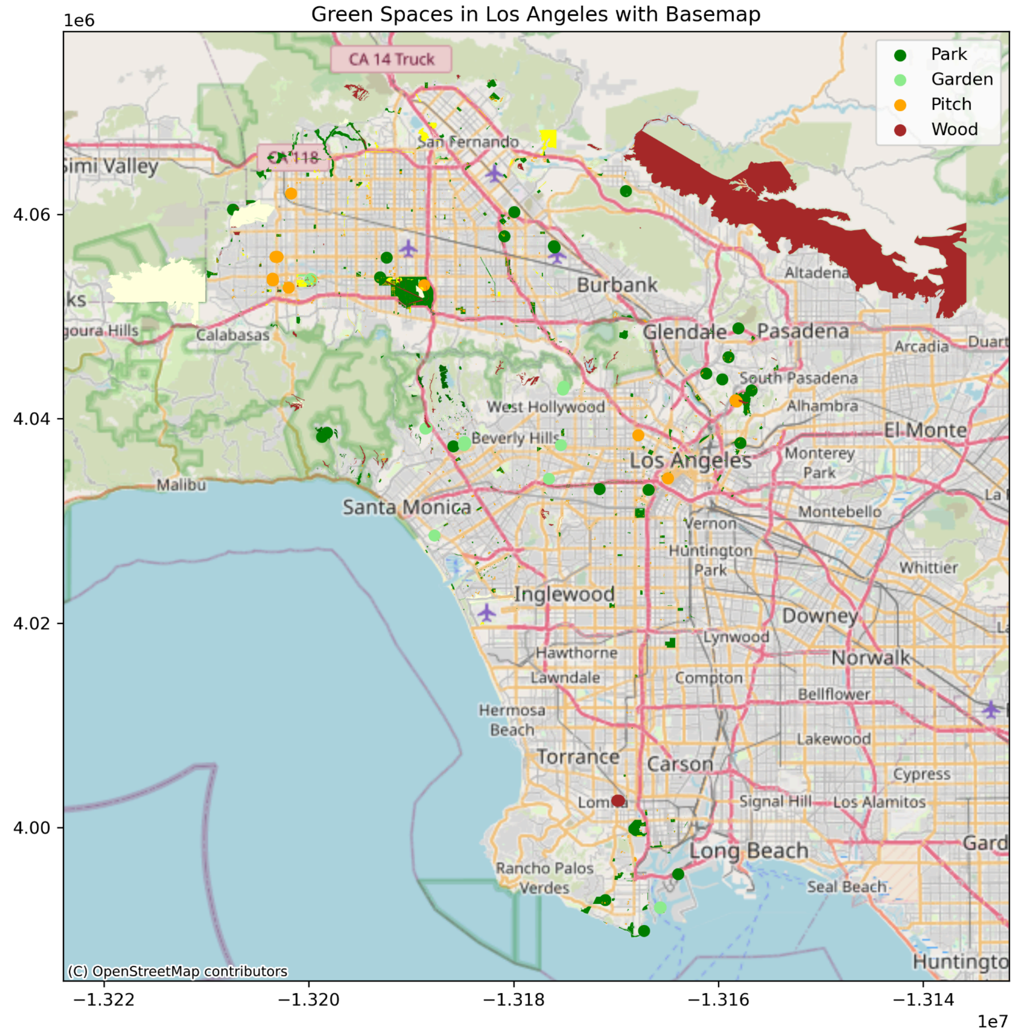


Figure 4. Projection system transformation of spatial data (from EPSG:3857 to EPSG:4326)

4. Methods

4.1. Data Preparation

The metro station data was imported and reprojected into WGS84 using GeoPandas, allowing for accurate distance measurements and spatial operations. The green spaces were extracted from OSMnx using specific tags such as 'park' and 'forest'. Additionally, this process is illustrated in Figure 4, which shows the spatial distribution of green spaces in Los Angeles. Additionally, Figure 5 displays the metro stations, plotted with their precise geographic locations.

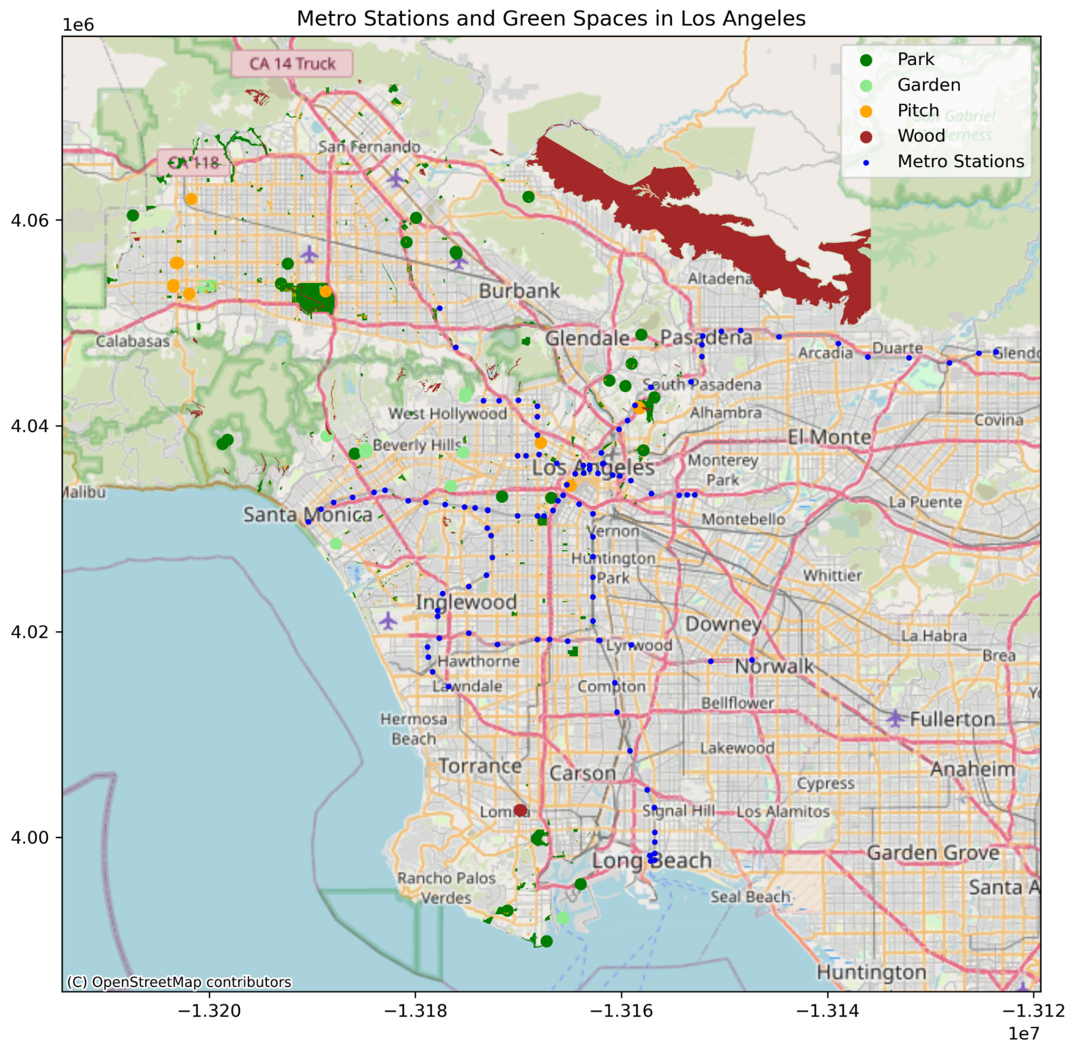


Figure 5. Geographic locations of metro stations in Los Angeles

4.2. Buffer Analysis

A buffer of 300 meters was created around each metro station to evaluate the proximity of green spaces to public transportation. The buffer zones were computed using the projected CRS EPSG:32611 for accurate distance calculations. In Figure 6, the map shows the buffer zones overlaid on the green spaces, providing a clear visual of coverage.

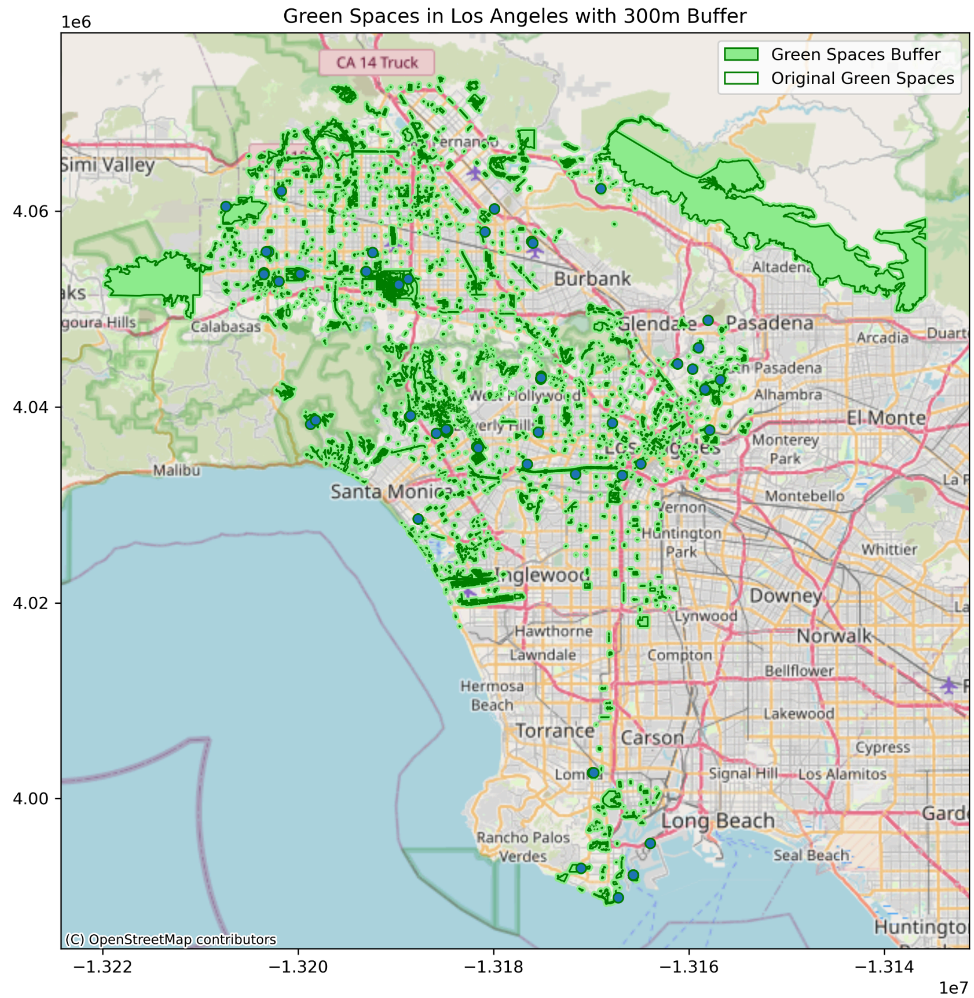


Figure 6. Buffer zones (300 meters) around metro stations and their overlap with green spaces.

4.3. Distance Calculation

The minimum distance between each metro station and the nearest green space was calculated using GeoPandas' spatial operations. The results were visualized with color-coded points, indicating distance categories from the nearest green space (as shown in Figure 7).

[11]:
```
import numpy as np

# Calculate the nearest distance between metro stops and green spaces
gdf_stops['nearest_distance'] = gdf_stops.geometry.apply(
    lambda station: gdf_green_spaces.distance(station).min()
)
```

[12]:
```
from matplotlib.colors import PowerNorm

# Create a PowerNorm object with gamma scaling
norm = PowerNorm(gamma=0.09, vmin=0, vmax=gdf_stops['nearest_distance'].
 ↪quantile(0.95))  #   gamma
cmap = plt.get_cmap('Blues')  # Get the color map

# Apply the color based on the nearest distance
```

```
gdf_stops['color'] = gdf_stops['nearest_distance'].apply(lambda dist:␣
 ↪cmap(norm(dist)))
```

[13]:
```
from matplotlib.lines import Line2D

# Plot metro stops and green spaces with distance coloring
fig, ax = plt.subplots(figsize=(12, 10))

# Plot metro stops and green spaces with distance coloring
gdf_green_spaces[gdf_green_spaces['leisure'] == 'park'].plot(ax=ax,␣
 ↪color='green', label='Park')
gdf_green_spaces[gdf_green_spaces['leisure'] == 'garden'].plot(ax=ax,␣
 ↪color='lightgreen', label='Garden')
gdf_green_spaces[gdf_green_spaces['leisure'] == 'pitch'].plot(ax=ax,␣
 ↪color='orange', label='Pitch')
gdf_green_spaces[gdf_green_spaces['landuse'] == 'forest'].plot(ax=ax,␣
 ↪color='darkgreen', label='Forest')
gdf_green_spaces[gdf_green_spaces['landuse'] == 'meadow'].plot(ax=ax,␣
 ↪color='yellow', label='Meadow')
gdf_green_spaces[gdf_green_spaces['natural'] == 'wood'].plot(ax=ax,␣
 ↪color='brown', label='Wood')
gdf_green_spaces[gdf_green_spaces['natural'] == 'grassland'].plot(ax=ax,␣
 ↪color='lightyellow', label='Grassland')

# Plot the metro stops with colors representing the distance to green spaces
gdf_stops.plot(ax=ax, marker='o', color=gdf_stops['color'], markersize=15,␣
 ↪label='Metro Stations')

# Add a basemap
ctx.add_basemap(ax, source=ctx.providers.OpenStreetMap.Mapnik)

# Create a custom legend
legend_elements = [
    Line2D([0], [0], marker='o', color='w', label='Metro Stations',␣
 ↪markerfacecolor='blue', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Park',␣
 ↪markerfacecolor='green', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Garden',␣
 ↪markerfacecolor='lightgreen', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Pitch',␣
 ↪markerfacecolor='orange', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Forest',␣
 ↪markerfacecolor='darkgreen', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Meadow',␣
 ↪markerfacecolor='yellow', markersize=10),
```

```
    Line2D([0], [0], marker='o', color='w', label='Wood',␣
 ↪markerfacecolor='brown', markersize=10),
    Line2D([0], [0], marker='o', color='w', label='Grassland',␣
 ↪markerfacecolor='lightyellow', markersize=10)
]

# Add the legend to the plot
ax.legend(handles=legend_elements, loc='upper right')

# Set the title and save the plot
plt.title('Metro Stations and Distance to Green Spaces in Los Angeles')

plt.savefig('Metro Stations and Distance to Green Spaces in Los Angeles',␣
 ↪dpi=300)

plt.show()
```

Figure 7. Distance categories of metro stations to the nearest green spaces, color-coded

5. Results Visualization

The buffer analysis reveals that a significant portion of metro stations are located within close proximity to green spaces, especially in central Los Angeles. The distribution of stations relative to green spaces is shown in Figure 8, where metro stations in dense urban areas demonstrate shorter distances to parks, while those in suburban regions exhibit longer distances.

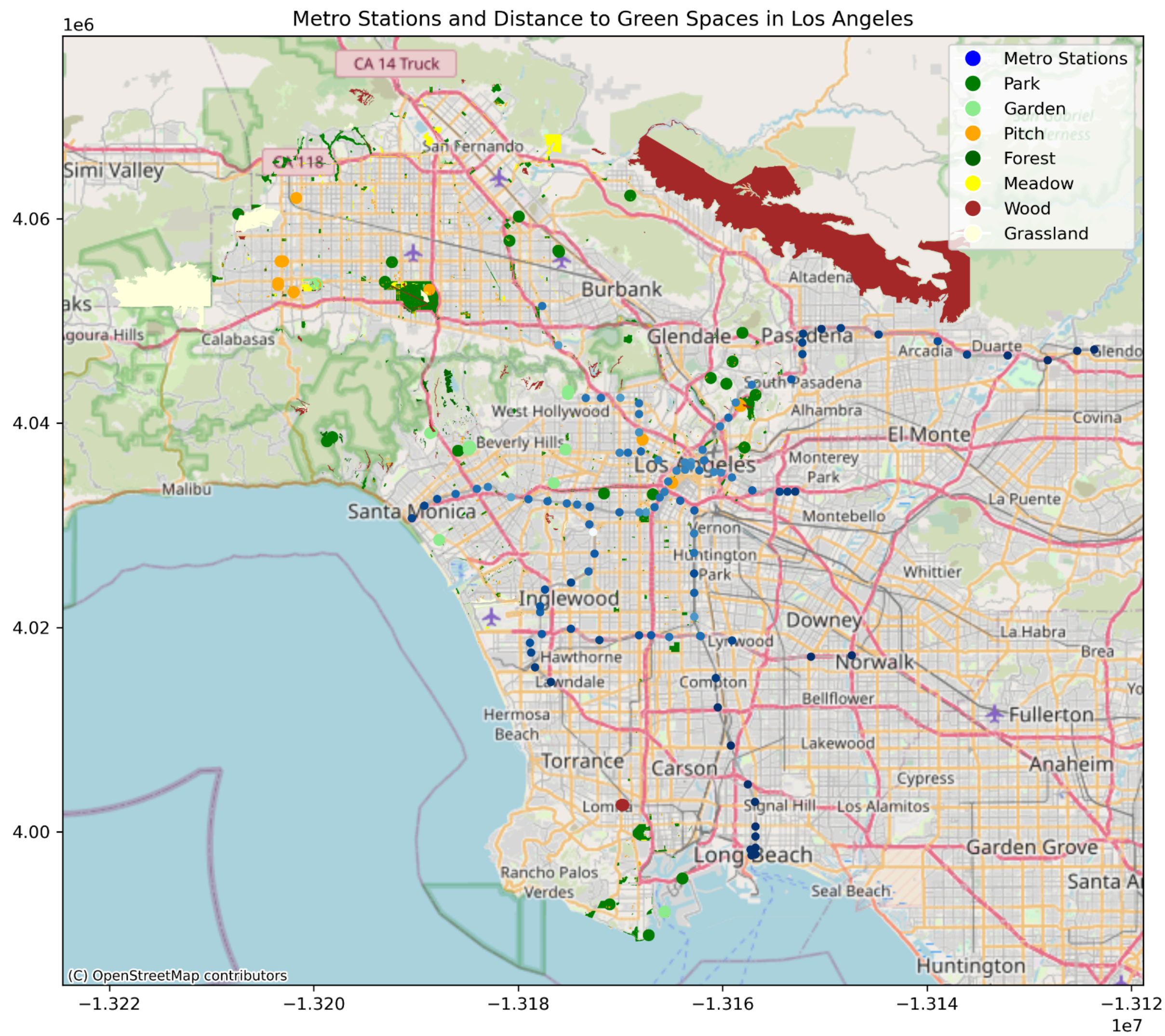


Figure 8. Metro station accessibility to green spaces in Los Angeles

6. Discussion

This analysis underscores the accessibility of green spaces in urban Los Angeles, offering valuable insights into how transportation planning can enhance public access to these areas. One limitation of the study is the reliance on publicly available datasets, which may not always reflect the most up-to-date infrastructure changes. Future research could benefit from integrating real-time transportation data or extending this analysis to cities with different urban designs, such as New York or San Francisco.

6.1 Potential Applications

6.1.1. Transit-Oriented Development

City planners can utilize this methodology to identify areas with limited access to green spaces and prioritize development around existing transit lines, improving overall accessibility.

6.1.2. Public Health

Researchers can apply this analysis to study the correlation between proximity to green spaces and public health outcomes, using metrics such as physical activity or air quality.

References

GeoPandas and OSMnx Python libraries for spatial analysis

https://spatial-analytics.readthedocs.io/en/latest/lessons/L1/intro-to-python-geostack.html

Los Angeles Metro Rail Station dataset from GeoHub: Metro Rail Lines Stops

https://geohub.lacity.org/datasets/6679d1ccc3744a7f87f7855e7ce33395_1/explore?location=33.967338%2C-118.235209%2C10.83&showTable=true

OpenStreetMap data using OSMnx for green spaces

https://autogis-site.readthedocs.io/en/latest/lessons/lesson-6/retrieve-data-from-openstreetmap.html