Ben Gaskill

Spatial Database Development

Assignment 1

### **Newark Green Space Analysis**

Data:

GitHub Repository

**new-jersey-latest.osm.pbf** The data was obtained from Geofabrik from the following <u>URL</u>

# Objectives:

The objective of this lab is to work with OpenStreetMap data for a certain city and utilize the PostgreSQL database to explore the data and run queries for analysis.

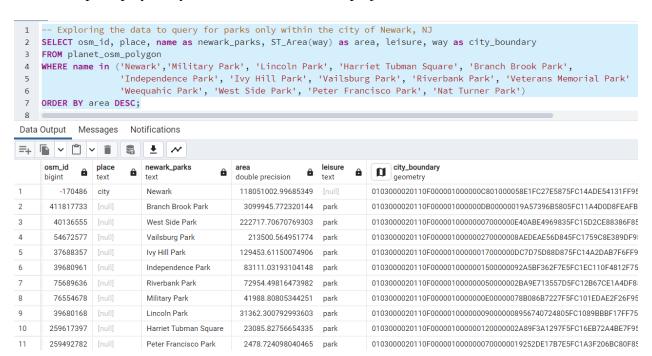
#### Methods

- 1. Explore the New Jersey OSM dataset for green spaces in Newark, New Jersey
- 2. Run Various SQL scripts to clean the data and perform analysis
- 3. Create and populate a new green\_spaces database for Newark, New Jersey
- 4. Run additional analysis on the isolated Newark park green\_spaces database
- 5. Import the data and visualize in QGIS

## **Findings**

Each screenshot below includes a commented SQL script with the corresponding output (if applicable). Note: when attempting to re-create the query, it is necessary to execute each query individually as opposed to running the entire script at once.

1. This script displays only the relevant fields to the project.



2. This script is used for data cleaning to delete entries of Newark parks in which the leisure and place fields are NULL.

```
-- Cleaning the data
-- Deleting entries where the OSM data held a NULL value for leisure and place.

DELETE FROM planet_osm_polygon

WHERE name in ('Newark','Military Park', 'Lincoln Park', 'Harriet Tubman Square', 'Branch Brook Park',

'Independence Park', 'Ivy Hill Park', 'Vailsburg Park', 'Riverbank Park', 'Veterans Memorial Park'

'Weequahic Park', 'West Side Park', 'Peter Francisco Park', 'Nat Turner Park') AND leisure is NULL AND place is NULL;
```

3. This script is used for data cleaning to delete entries that have the same name as a park in Newark but are in another county.

Notes: There is no attribute data in the OSM layers that attributes the name of the park to the city where it is located. Thus, duplicate parks with the same name in other NJ counties show up. To fix this, I visualized the data in QGIS then recorded the name and area of the parks that are not within the boundaries of the Newark shapefile.

```
-- After visualizing entries in QGIS, I identified duplicate names of parks that are in different cities
-- Deleting these entries
-- Exploring the data to query for parks only within the city of Newark, NJ

DELETE FROM planet_osm_polygon

WHERE ST_Area(way) = 31569.52439827191

OR ST_Area(way) = 1448.7891236321857

OR ST_Area(way) = 16564.699621195792

OR ST_Area(way) = 203921.83202167338

OR ST_Area(way) = 7602.520831999341

OR ST_Area(way) = 15037.090347355754

OR ST_Area(way) = 1862464.633198247

OR ST_Area(way) = 47306.26771780349;
```

4. This script is used for creating a new table for Newark Parks.

```
--Create a new table for greenspaces (parks) in Newark, NJ
CREATE TABLE green_spaces (
   id SERIAL PRIMARY KEY,
   name VARCHAR(255),
   location GEOMETRY(Point, 3857),
   area_sq_m NUMERIC
);
```

5. This script is used for insertion of the data into the new table

```
-- Populating the new green_spaces table

INSERT INTO green_spaces (name, location, area_sq_m)

SELECT name, ST_Centroid(way), ST_Area(way)

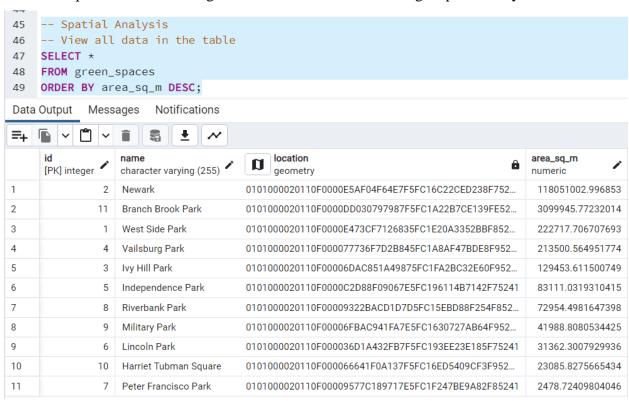
FROM planet_osm_polygon

WHERE name in ('Newark','Military Park', 'Lincoln Park', 'Harriet Tubman Square', 'Branch Brook Park',

'Independence Park', 'Ivy Hill Park', 'Vailsburg Park', 'Riverbank Park', 'Veterans Memorial Park'

'Weequahic Park', 'West Side Park', 'Peter Francisco Park', 'Nat Turner Park');
```

6. This script is used for viewing all fields in the new table to begin spatial analysis.



7. This script is used for the spatial analysis of total parks, total area, and average area of Newark Parks.



8. This script is used for the spatial analysis of the top 5 largest parks in Newark.

- 58 -- Select the top 5 largest green spaces
- 59 **SELECT name**, ROUND((area\_sq\_m/**1e6**), **2**) **as** area\_sq\_km
- 60 FROM green\_spaces
- 61 WHERE name <> 'Newark'
- 62 ORDER BY area\_sq\_m DESC
- 63 **LIMIT 5**;

Data Output Messages Notifications

=+		• ~
	name character varying (255)	area_sq_km numeric
1	Branch Brook Park	3.10
2	West Side Park	0.22
3	Vailsburg Park	0.21
4	Ivy Hill Park	0.13
5	Independence Park	0.08

# **QGIS** Visualization

1. The top 5 largest parks in Newark, New Jersey

Relevant Data Layers:

**top5:** The top 5 largest parks are symbolized with larger red points.

**green\_spaces:** The additional parks within the administrative boundaries are symbolized with smaller green points.

newark\_boundary: Administrative boundaries defined in the OSM data.

NJ 2015 Aerial Imagery: NAIP Imagery added to QGIS using the QuickMap Services Plugin

