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 URL

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.

```sql
-- Exploring the data to query for parks only within the city of Newark, NJ
SELECT osm_id, place, name as newark_parks, ST_Area(way) as area, leisure, way as city_boundary
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')
ORDER BY area DESC;
```

Data Output | Messages | Notifications

	osm_id bigint	place text	newark_parks text	area double precision	leisure text	city_boundary geometry
1	-170486	city	Newark	118051002.99685349	[null]	0103000020110F000001000000C801000058E1FC27E5875FC14ADE54131FF95
2	411817733	[null]	Branch Brook Park	3099945.772320144	park	0103000020110F000001000000DB00000019A57396B5805FC11A4D0D8FEAFB
3	40136555	[null]	West Side Park	222717.70670769303	park	0103000020110F00000100000007000000E40ABE4969835FC15D2CE88386F85
4	54672577	[null]	Vailsburg Park	213500.564951774	park	0103000020110F0000010000000270000008AEDEAE56D845FC1759C8E389DF9
5	37688357	[null]	Ivy Hill Park	129453.61150074906	park	0103000020110F00000100000017000000DC7D75D88D875FC14A2DAB7F6FF9
6	39680961	[null]	Independence Park	83111.03193104148	park	0103000020110F0000010000001500000092A5BF362F7E5FC1EC110F4812F75
7	75689636	[null]	Riverbank Park	72954.49816473982	park	0103000020110F000001000000050000002BA9E713557D5FC12B67CE1A4DF8
8	76554678	[null]	Military Park	41988.80805344251	park	0103000020110F000001000000E00000078B086B7227F5FC101EDAE2F26F95
9	39680168	[null]	Lincoln Park	31362.300792993603	park	0103000020110F00000100000009000000895674072 4805FC1089BBBF17FF75
10	259617397	[null]	Harriet Tubman Square	23085.82756654335	park	0103000020110F000001000000120000002A89F3A1297F5FC16EB72A4BE7F95
11	259492782	[null]	Peter Francisco Park	2478.724098040465	park	0103000020110F000001000000070000019252DE17B7E5FC1A3F206BC80F85

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

```sql
-- 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.

```sql
-- 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.

```sql
--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

```sql
-- 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.

```sql
-- Spatial Analysis
-- View all data in the table
SELECT *
FROM green_spaces
ORDER BY area_sq_m DESC;
```

Data Output | Messages | Notifications

	id [PK] integer	name character varying (255)	location geometry	area_sq_m numeric
1	2	Newark	0101000020110F0000E5AF04F64E7F5FC16C22CED238F752...	118051002.996853
2	11	Branch Brook Park	0101000020110F0000DD030797987F5FC1A22B7CE139FE52...	3099945.77232014
3	1	West Side Park	0101000020110F0000E473CF7126835FC1E20A3352BBF852...	222717.706707693
4	4	Vailsburg Park	0101000020110F000077736F7D2B845FC1A8AF47BDE8F952...	213500.564951774
5	3	Ivy Hill Park	0101000020110F00006DAC851A49875FC1FA2BC32E60F952...	129453.611500749
6	5	Independence Park	0101000020110F0000C2D88F09067E5FC196114B7142F75241	83111.0319310415
7	8	Riverbank Park	0101000020110F00009322BACD1D7D5FC15EBD88F254F852...	72954.4981647398
8	9	Military Park	0101000020110F00006FBAC941FA7E5FC1630727AB64F952...	41988.8080534425
9	6	Lincoln Park	0101000020110F000036D1A432FB7F5FC193EE23E185F75241	31362.3007929936
10	10	Harriet Tubman Square	0101000020110F000066641F0A137F5FC16ED5409CF3F952...	23085.8275665434
11	7	Peter Francisco Park	0101000020110F00009577C189717E5FC1F247BE9A82F85241	2478.72409804046

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

```sql
-- Calculate the total number of parks, total area (converted to sq km), and average area.
-- Exlcuded one entry for the entire city of Newark from the analysis.
-- Solution for exclusion found here: https://stackoverflow.com/questions/20075910/where-column-is-not-value
SELECT COUNT(*) AS total_parks, ROUND((SUM(area_sq_m)/1e6), 2) AS total_area_sq_km, ROUND((AVG(area_sq_m)/1e6), 2) AS average_area_sq_km
FROM green_spaces
WHERE name <> 'Newark';
```

Data Output | Messages | Notifications

	total_parks bigint	total_area_sq_km numeric	average_area_sq_km numeric
1	10	3.92	0.39

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

```sql
-- Select the top 5 largest green spaces
SELECT name, ROUND((area_sq_m/1e6), 2) as area_sq_km
FROM green_spaces
WHERE name <> 'Newark'
ORDER BY area_sq_m DESC
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