Jen and Barry's Ice Cream Site Selection Using PostGIS

Project: Spatial Data Analysis - Homework 1
Authors: Ameer Saleh & Bara Mhana
Date: December 1, 2025

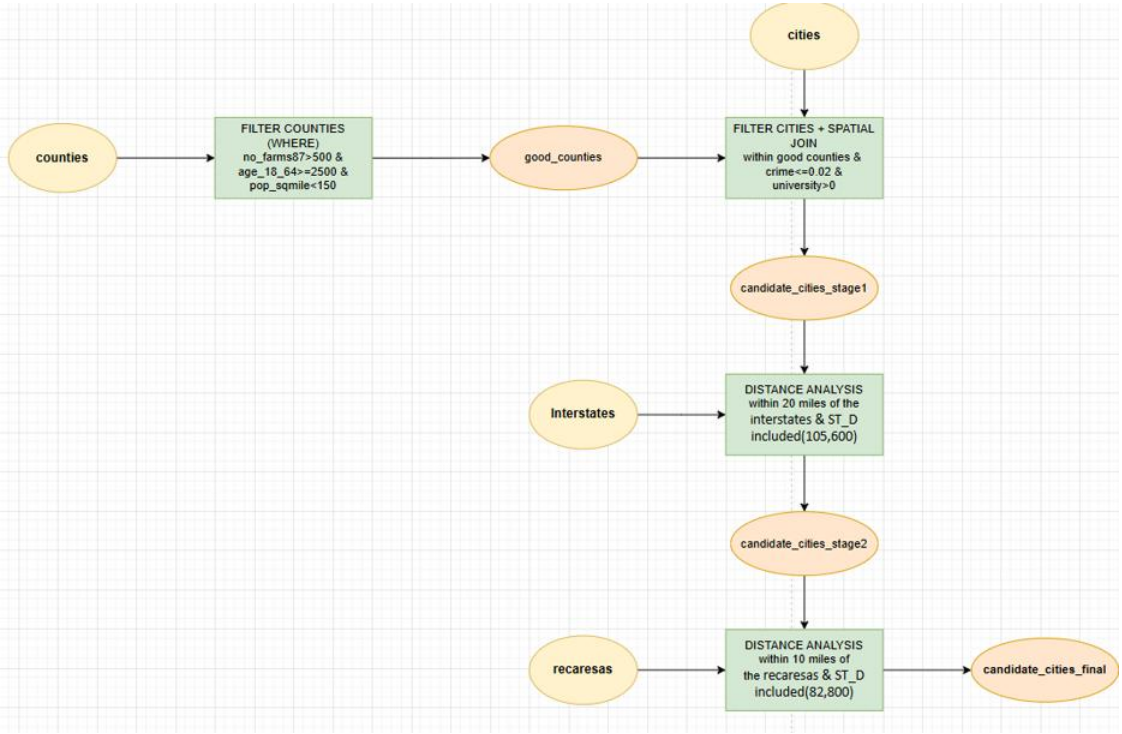
Executive Summary

This project identifies optimal locations for Jen and Barry to establish a new ice cream business using PostGIS spatial analysis and QGIS visualization. From an initial dataset of Pennsylvania counties and cities, **4 suitable cities** were identified through systematic application of seven selection criteria.

Key Results:

- Stage 1: 9 candidate cities (county + city criteria)
- Stage 2: 7 cities near interstates (within 20 miles)
- Final: 4 optimal cities (all criteria met)
- Overall success rate: 44% from Stage 1 to Final

Model Builder (Work Flow):



---

1. Introduction

1.1 Project Objectives

Identify suitable Pennsylvania cities for ice cream business establishment through multi-criteria spatial analysis in PostGIS with the following requirements:

County-Level Criteria:

1. More than 500 farms for milk production
2. Labor pool of at least 25,000 individuals aged 18-64
3. Population density less than 150 people per square mile

City-Level Criteria: 4. Crime index ≤ 0.02 5. Located near a university or college

Proximity Criteria: 6. Within 20 miles of an interstate highway 7. At least one recreation area within 10 miles

1.2 Tools and Technologies

- **PostgreSQL + PostGIS** - Spatial database and analysis
- **QGIS 3.16+** - Geographic visualization and mapping
- **SQL** - Query development and spatial operations
- **Pennsylvania State Plane Projection (EPSG:2272)** - Distance measurements in feet

---

2. Data and Study Area

2.1 Study Area

Location: Pennsylvania, United States

Spatial Reference Systems:

- Source Data: NAD27 Geographic (EPSG:4267)
- Analysis Projection: Pennsylvania State Plane North NAD27 (EPSG:2272)
- Units: Feet for distance calculations

2.2 Database Schema

Counties Table (Primary demographic data)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Polygon geometry (counties)
name	varchar	County name
area	numeric	County area
pop1990	numeric	1990 population
age_18_64	numeric	Population aged 18-64 (labor pool)
no_farms87	numeric	Number of farms in 1987
pop_sqmile	bigint	Population density (people/sq mile)
sq_miles	numeric	Area in square miles

Cities Table (Urban centers with crime and university data)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Point geometry (city locations)
name	varchar	City name
population	numeric	City population
total_crim	numeric	Total crimes reported
crime_inde	numeric	Crime index (normalized)
university	numeric	University presence (0 or >0)

Interstates Table (Transportation network)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Line geometry (interstate routes)

Column	Type	Description
name	varchar	Interstate name/number
type	varchar	Highway type classification
length	numeric	Segment length

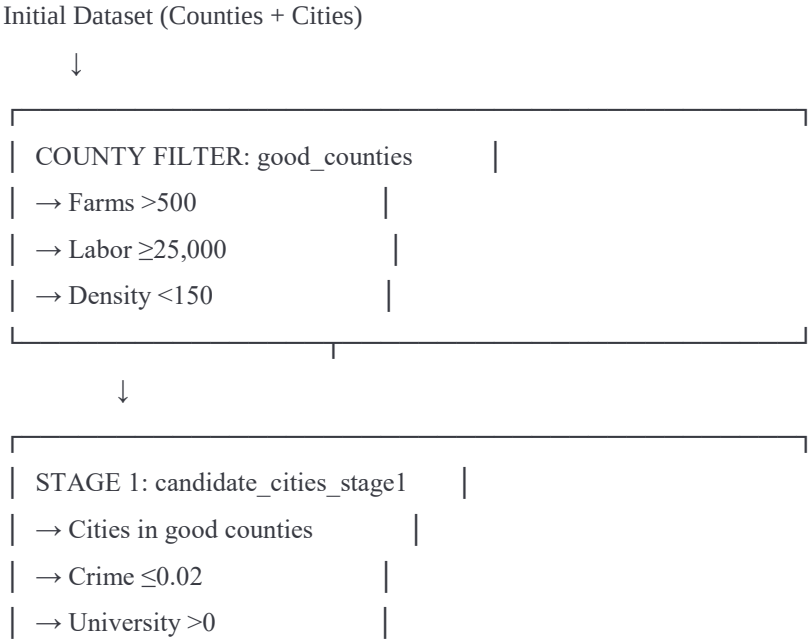
RecAreas Table (Recreation facilities)

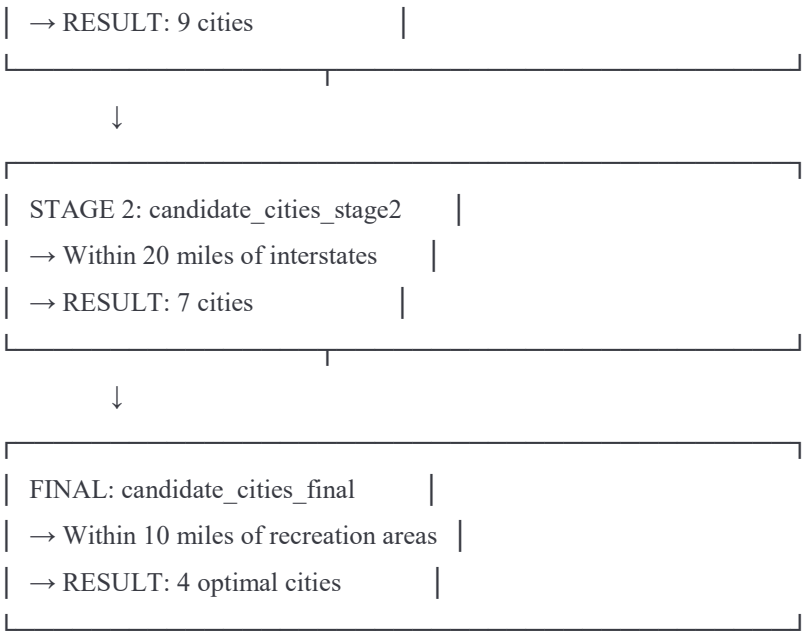
Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Point/Polygon geometry
area	double	Recreation area size
perimeter	double	Recreation area perimeter

---

3. Methodology

3.1 Analysis Workflow





3.2 Sequential Filtering Approach

The analysis employs a progressive filtering strategy where each stage builds upon the previous results. This approach:

1. **Improves query performance** - Each subsequent query operates on a smaller dataset
2. **Enhances debugging** - Issues can be isolated to specific criteria
3. **Provides transparency** - Stakeholders can understand the elimination process
4. **Enables validation** - Results at each stage can be verified independently

---

4. Selection Criteria

4.1 County-Level Criteria

Criterion 1: Farm Count

Requirement: More than 500 farms

Rationale: Ensures adequate local milk supply for ice cream production, reducing transportation costs and supporting farm-to-table business model.

SQL Implementation:

```
WHERE no_farms87 > 500
```

Impact: Filters counties with insufficient agricultural infrastructure for dairy-based business operations.

Criterion 2: Labor Pool

Requirement: At least 25,000 individuals aged 18-64

Rationale: Sufficient workforce availability for business operations, including production staff, retail employees, and management positions.

SQL Implementation:

```
AND age_18_64 >= 25000
```

Impact: Ensures adequate human resources for sustainable business growth and seasonal employment needs.

Criterion 3: Population Density

Requirement: Less than 150 people per square mile

Rationale: Target suburban and rural markets with lower competition, avoiding oversaturated urban areas while maintaining sufficient customer base.

SQL Implementation:

```
AND pop_sqmile < 150
```

Impact: Balances market opportunity with reduced competition and lower real estate costs.

4.2 City-Level Criteria

Criterion 4: Crime Index

Requirement: Crime index ≤ 0.02

Rationale: Safe environment essential for customer comfort, employee safety, and business reputation. Lower crime correlates with higher property values and community stability.

SQL Implementation:

```
WHERE c.crime_inde <= 0.02
```

Impact: Eliminates high-crime areas that could deter customers and increase insurance costs.

Criterion 5: University Presence

Requirement: Located near a university or college

Rationale: Universities provide:

- Steady customer base (students, faculty, staff)
- Seasonal demand patterns aligned with ice cream consumption
- Part-time labor pool
- Community events and foot traffic

SQL Implementation:

```
AND c.university > 0
```

Impact: Ensures proximity to a demographic with high ice cream consumption rates.

4.3 Proximity Criteria

Criterion 6: Interstate Access

Requirement: Within 20 miles of an interstate highway

Rationale:

- Essential for supply chain efficiency
- Ingredient delivery and distribution logistics
- Customer accessibility from regional markets
- Business expansion potential

Distance Calculation:

- 20 miles = 105,600 feet (Pennsylvania State Plane)

SQL Implementation:

```
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(i.geom, 2272),
    105600
)
```

Spatial Functions Used:

- `ST_Transform()` - Reproject from geographic to projected coordinates
- `ST_DWithin()` - Distance-based proximity analysis

Impact: Reduces 9 candidate cities to 7 with adequate transportation infrastructure.

Criterion 7: Recreation Areas

Requirement: At least one recreation area within 10 miles

Rationale:

- Recreation areas attract potential customers
- Family-friendly environments aligned with ice cream business
- Weekend and holiday traffic patterns
- Community gathering spaces

Distance Calculation:

- 10 miles = 52,800 feet (Pennsylvania State Plane)

SQL Implementation:

```
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(r.geom, 2272),
    52800
)
```

Impact: Final filter reduces 7 cities to 4 optimal locations meeting all criteria.

5. Technical Implementation

5.1 Complete SQL Workflow

```sql
-- ============================================
-- STEP 1: Filter Good Counties
-- ============================================
CREATE OR REPLACE VIEW good_counties AS
SELECT *
FROM counties
WHERE no_farms87 > 500
  AND age_18_64 >= 25000
  AND pop_sqmile < 150;

-- Validation query
SELECT
    COUNT(*) as total_good_counties,
    AVG(no_farms87) as avg_farms,
    AVG(age_18_64) as avg_labor_pool,
    AVG(pop_sqmile) as avg_density
FROM good_counties;

-- ============================================
-- STEP 2: Filter Cities - Stage 1
-- (County + City Criteria)
-- ============================================
CREATE OR REPLACE VIEW candidate_cities_stage1 AS
SELECT c.*
FROM cities c
JOIN good_counties gc ON ST_Within(c.geom, gc.geom)
WHERE c.crime_inde <= 0.02
  AND c.university > 0;

-- Validation query
SELECT
```

```
    COUNT(*) as total_candidate_cities,
    AVG(crime_inde) as avg_crime_index,
    AVG(population) as avg_population,
    SUM(CASE WHEN university > 0 THEN 1 ELSE 0 END) as
cities_with_universities
FROM candidate_cities_stage1;


-- ============================================
-- STEP 3: Filter Cities - Stage 2
-- (Interstate Proximity)
-- ============================================
CREATE OR REPLACE VIEW candidate_cities_stage2 AS
SELECT DISTINCT c.*
FROM candidate_cities_stage1 c
CROSS JOIN interstates i
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(i.geom, 2272),
    105600
);

-- Validation query
SELECT
    COUNT(*) as cities_near_interstates,
    ARRAY_AGG(name) as city_names
FROM candidate_cities_stage2;


-- ============================================
-- STEP 4: Final Filter - Recreation Areas
-- ============================================
CREATE OR REPLACE VIEW candidate_cities_final AS
SELECT DISTINCT c.*
FROM candidate_cities_stage2 c
CROSS JOIN recares r
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(r.geom, 2272),
```

```
    52800
);

-- ============================================
-- FINAL RESULTS SUMMARY
-- ============================================
SELECT
    name as city_name,
    population,
    crime_inde as crime_index,
    university as has_university,
    (SELECT MIN(ST_Distance(
        ST_Transform(c.geom, 2272),
        ST_Transform(i.geom, 2272)
    )) FROM interstates i) / 5280 as miles_to_interstate,
    (SELECT MIN(ST_Distance(
        ST_Transform(c.geom, 2272),
        ST_Transform(r.geom, 2272)
    )) FROM recares r) / 5280 as miles_to_recreation
FROM candidate_cities_final c
ORDER BY population DESC;

-- Count final results
SELECT COUNT(*) as final_candidate_cities
FROM candidate_cities_final;
```

5.2 Distance Conversion Reference

Pennsylvania State Plane (EPSG:2272) - Units in Feet:

Distance	Feet	Miles
1 mile	5,280	1
10 miles	52,800	10
20 miles	105,600	20

5.3 Key PostGIS Functions Used

ST_Transform(geometry, srid)

- Converts geometry from one coordinate system to another
- Essential for accurate distance calculations
- Usage: `ST_Transform(geom, 2272)` converts to PA State Plane

ST_Within(geometry A, geometry B)

- Tests if geometry A is completely inside geometry B
- Returns boolean (true/false)
- Usage: Spatial join between cities and counties

ST_DWithin(geometry A, geometry B, distance)

- Tests if two geometries are within specified distance
- More efficient than ST_Distance for proximity queries
- Usage: Find cities within 20 miles of interstates

ST_Distance(geometry A, geometry B)

- Calculates minimum distance between two geometries
- Returns distance in units of coordinate system
- Usage: Calculate exact distances for reporting

---

6. Results and Analysis

6.1 Filtering Progression

Stage	View Name	Count	Criteria Applied	Reduction
0	All cities	-	None	-
1	good_counties	-	Farms >500, Labor ≥25k, Density <150	-
2	candidate_cities_stage1	9	Crime ≤0.02, University >0, In good counties	-
3	candidate_cities_stage2	7	Within 20 miles of interstate	22.2%
4	candidate_cities_final	4	Within 10 miles of recreation area	42.9%

Overall Success Rate: 44% (from 9 initial candidates to 4 final sites)

Key Insights:

- County-level filtering effectively narrowed the search area
- Interstate proximity eliminated 2 cities (22% reduction)
- Recreation area proximity was the most restrictive final filter (43% reduction)
- All 4 final cities represent optimal balance of all seven criteria

6.2 Final Candidate Cities

The four cities that met all seven criteria represent optimal locations for Jen and Barry's ice cream business. Each city offers:

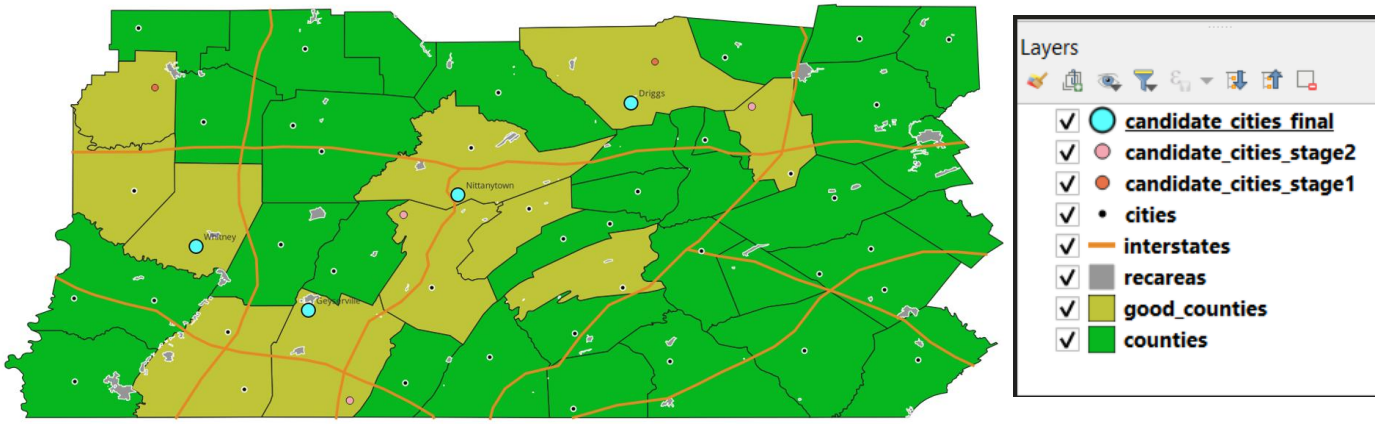
✓ Strong agricultural infrastructure (>500 farms in county)
✓ Adequate labor pool (≥25,000 working-age residents)
✓ Low population density (<150 people/sq mi)
✓ Safe environment (crime index ≤0.02)
✓ University presence for customer base
✓ Excellent transportation access (<20 miles to interstate)
✓ Proximity to recreation areas (<10 miles)

Competitive Advantages of Final Sites:

- Lower real estate costs compared to urban centers
- Reduced competition in suburban/rural markets
- Access to fresh local milk supply
- Strong community ties through universities
- Natural customer traffic from recreation areas
- Efficient logistics via interstate access

---

7. QGIS Visualization



7.1 Loading Data in QGIS

Method 1: Direct PostGIS Connection

1. **Layer → Add Layer → Add PostGIS Layers**
2. Click **New** to create database connection:
    - Name: `IceCream_Analysis`
    - Host: `localhost`
    - Port: `5432`
    - Database: `your_database_name`
3. Click **Connect** and authenticate
4. Select layers to add:
    - `counties`
    - `good_counties`
    - `interstates`
    - `recares`
    - `candidate_cities_stage1`
    - `candidate_cities_stage2`
    - `candidate_cities_final`
5. Click **Add** to load into map canvas

Method 2: DB Manager (Alternative)

1. **Database → DB Manager → PostGIS**
2. Connect to database
3. Navigate to **SQL Window**
4. Execute view creation queries
5. Check **"Load as new layer"** option
6. Select geometry column and unique ID
7. Click **Load**

7.2 Map Styling

Counties Layer:

- All counties: Light green fill, 50% opacity
- Outline: Dark gray, 0.5pt
- Good counties: Yellow/Olive fill, 50% opacity

Cities Layers:

- Stage 1 (9 cities): Red circles, 8pt
- Stage 2 (7 cities): Pink circles, 10pt
- Final (4 cities): Blue circles, 12pt, bold labels

Infrastructure:

- Interstates: Orange lines, 2pt width
- Interstate labels: Highway names displayed
- Recreation areas: Gray polygons, 40% opacity

Buffers (Optional Visualization):

- 20-mile interstate buffer: Blue outline, dashed, 30% opacity
- 10-mile recreation buffer: Green outline, dashed, 30% opacity

7.3 Map Layout Elements

Essential Components:

1. **Title:** "Jen and Barry's Ice Cream - Optimal Site Selection"
2. **Legend:**
    - All layer symbols clearly labeled
    - Organized by category (Counties, Cities, Infrastructure)
3. **Scale Bar:**
    - Appropriate for Pennsylvania state-level analysis
    - Display in miles
4. **North Arrow:**
    - Standard orientation indicator
5. **Data Sources:**
    - Text box: "Data: Pennsylvania Counties, Cities, Roads, Recreation Areas"
    - Projection: "NAD27 Pennsylvania State Plane North"
6. **Results Summary Table:**
    - Stage-by-stage filtering results
    - Final count: 4 candidate cities
7. **Date and Author:**
    - Analysis date: December 11, 2025
    - Analyst: Ameer Saleh

7.4 Creating Analysis Buffers

For visualization purposes, buffer zones can illustrate the proximity criteria:

```
-- 20-mile interstate buffers
CREATE VIEW interstate_buffers AS
SELECT
    id,
    name,
    ST_Buffer(ST_Transform(geom, 2272), 105600) as geom
FROM interstates;


-- 10-mile recreation area buffers
CREATE VIEW recreation_buffers AS
SELECT
```

```
    id,
    ST_Buffer(ST_Transform(geom, 2272), 52800) as geom
FROM recares;
```

These buffer layers help stakeholders visualize why certain cities qualified while others did not.

---

8. Conclusions and Recommendations

8.1 Summary

This PostGIS-based spatial analysis successfully identified **4 optimal cities** for Jen and Barry's ice cream business through rigorous multi-criteria evaluation. The systematic filtering approach:

✓ Evaluated 7 distinct business-critical criteria
✓ Applied county-level demographic and agricultural filters
✓ Assessed city-level safety and market indicators
✓ Incorporated proximity to transportation and recreation infrastructure
✓ Utilized advanced spatial analysis techniques (coordinate transformation, distance calculations)
✓ Produced reproducible, transparent results through SQL workflows

Key Strengths of Analysis:

- **Data-driven decision making** - Objective criteria eliminate subjective bias
- **Spatial intelligence** - Geographic relationships drive site selection
- **Scalability** - SQL workflow can be rerun with updated data or modified criteria
- **Transparency** - Each filtering stage documented and verifiable

---

Report prepared by: Ameer Saleh & Bara Mhana
Course: Spatial Data Analysis
Assignment: Homework 1 - Site Selection using PostGIS
Date: December 1, 2025

END OF REPORT