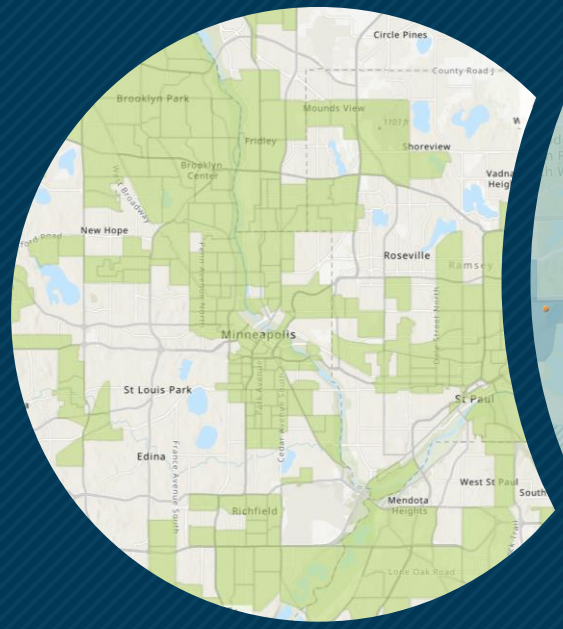
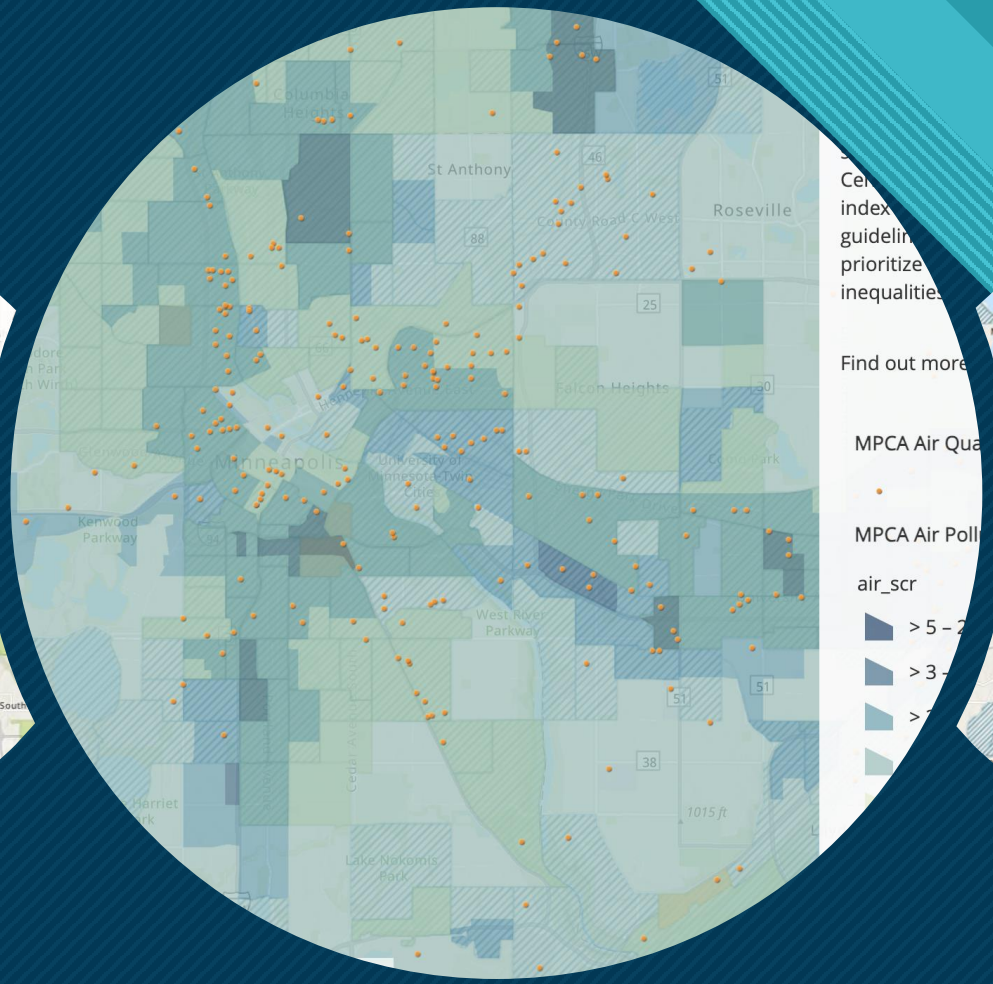
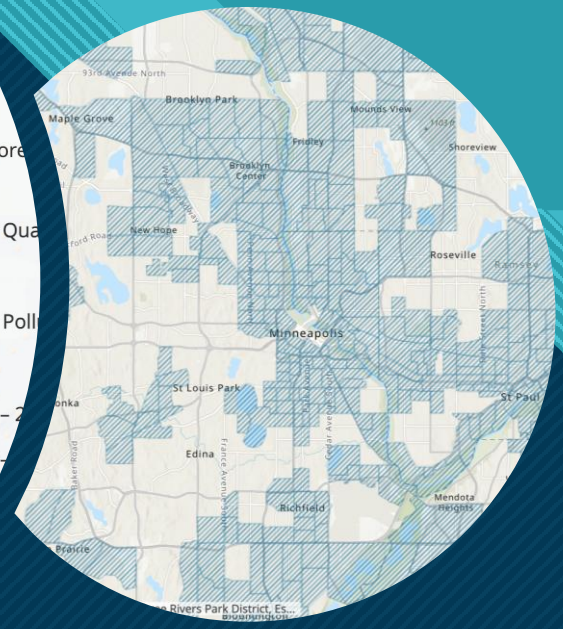
MN - Pollution and Its Effects On Health

A Study of Pollution and its effects on asthma and cancer rates in Minnesota counties







Environmental Justice – the "why"

Objectives of the Project:

1. To map facilities in MN that are subject to environmental regulations, and that could potentially contribute to negative health outcomes (e.g. industrial sites that emit known carcinogens)

2. To find if there is any correlation between the number of sites in a given county and the asthma and cancer rates and health outcomes in that county.

3. To make an interactive dashboard using JavaScript libraries in order to demonstrate trends, and/or help direct further inquiry.

The Process of Extracting Data

- **Asthma and cancer rates by counties in MN:**
  - Data acquired from the Minnesota Department of Health.
- **Health outcomes and factors:**
  - Data extracted from *County Health Rankings & Roadmaps (CHR&R)*, a program of the University of Wisconsin Population Health Institute.
- **Facilities and sites of interest:**
  - Data was downloaded in a CSV file from the U.S. Environmental Protection Agency Federal Registry Service (EPA FRS).
- **Significant environmental regulations violations in Minnesota:**
  - Data was downloaded from the EPA's Enforcement and Compliance History Online database (ECHO).

Sources:
https://data.web.health.state.mn.us/web/mndata/asthma
https://data.web.health.state.mn.us/web/mndata/cancer_all
https://www.countyhealthrankings.org/explore-health-rankings/minnesota/data-and-resources
https://www.epa.gov/frs/frs-query
https://echo.epa.gov/

The Process of Cleaning Data

Asthma Rates, Cancer Rates, Health Rankings

- Original data: CSV format (asthma and cancer rates), excel format (health rankings)
- Used Python/Pandas to clean data.
  - Update data types of some columns
  - Check for null values
  - Drop irrelevant columns

EPA FRS data

- Narrow dataset to relevant sites (going from 200K+ sites to about 3000)
- Dropping records with null values in lat/long or other vital fields, updating column names, filtering to find relevant records

JSON Formatting

- All data frames were converted to JSON format to be used in database and Flask API.

MongoDB Database

- The three key steps to the DB: setting up the environment by importing libraries and creating a MongoDB client, loading data files efficiently using a for loop, and merging the data into a single JSON format for API consumption.
- The output of the database step is to JSON collections and one geoJson collection.
- Additionally, the database file has two flavors.
  Flavor 1: is to be run locally.
  Flavor 2: is on ATLAS cloud including functioning APIs for each collection

- * Please note:

The database was written in a format that was entirely v*ariablized for ease of scalability and flexibility.*

Example: adding a new feature or data set is one line of code.

Flask API

Route 1: EPA FRS Environmental Interest Sites

- Gets the EPA FRS Environmental Interest sites as JSON
- Passes it to a function which converts it to geoJSON format
- Returns the entire dataset in geoJSON format

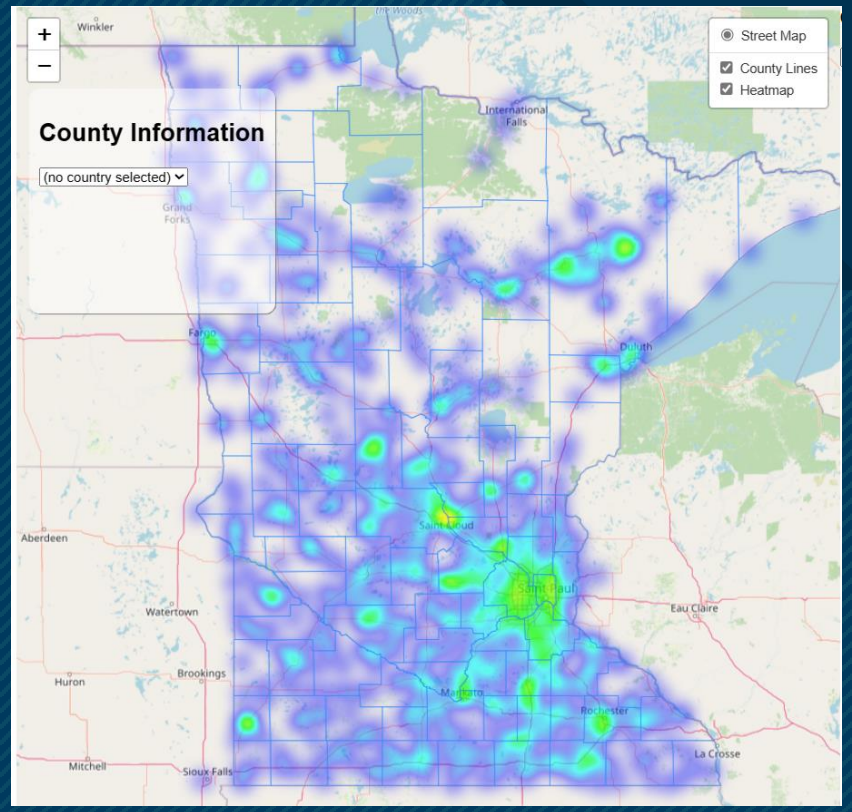
Route 2: MN Health Outcomes

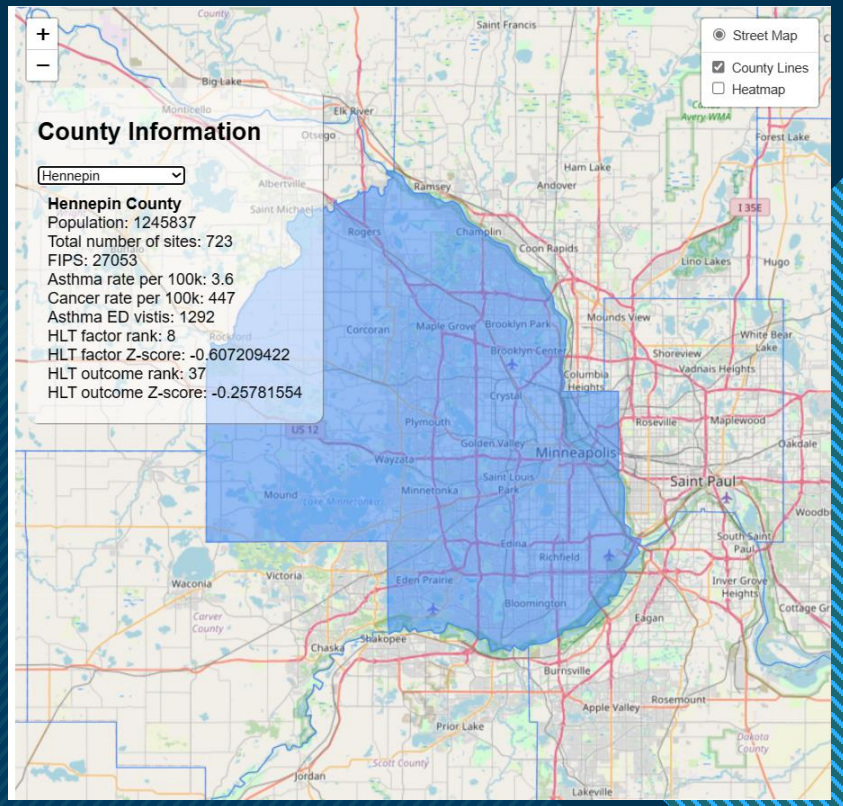
- Returns health outcomes, asthma rates, and cancer rates for each county in Minnesota in JSON format.

```
def geoJSON_format(dataset):
    """Put the EPA FRS Environmental Interest Sites data into geoJSON format"""
    for i in dataset:
        # get values
        registry_ID = i["REGISTRY_ID"]
        site_name = i["SITE PRIMARY NAME"]
        site_address = i["SITE_ADDRESS"]
        site_type = i["SITE_TYPE_NAME"]
        interest_types = i["INTEREST_TYPES"]
        county_name = i["COUNTY_NAME"]
        if "," in interest_types:
            interest_types = interest_types.split(", ")
        lat = i["LATITUDE"]
        lon = i["LONGITUDE"]

        # build geoJSON formatted dataset
        this_point = Point((lon, lat))
        properties = dict([('reg_ID', registry_ID), ('site_name', site_name), ('site_address', site_address),
                           ('county_name', county_name), ('site_type', site_type), ('interest_types', interest_types)])
        gj = Feature(geometry=this_point, properties=properties)

        gj_formatted.append(gj)
    return gj_formatted


# define Flask object

app = Flask(__name__)

```

Visualization

- Heatmap
  - Shows the spatial density of EPA sites
- County selector
  - Click on county
  - Dropdown menu

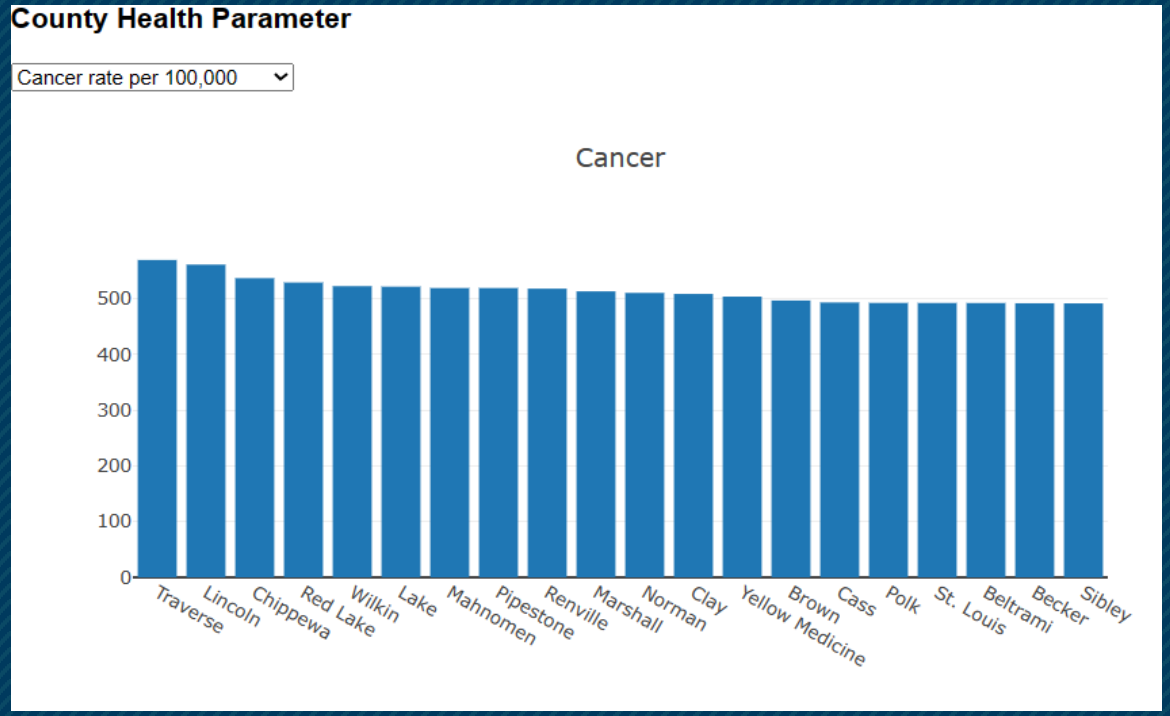




Visualizations

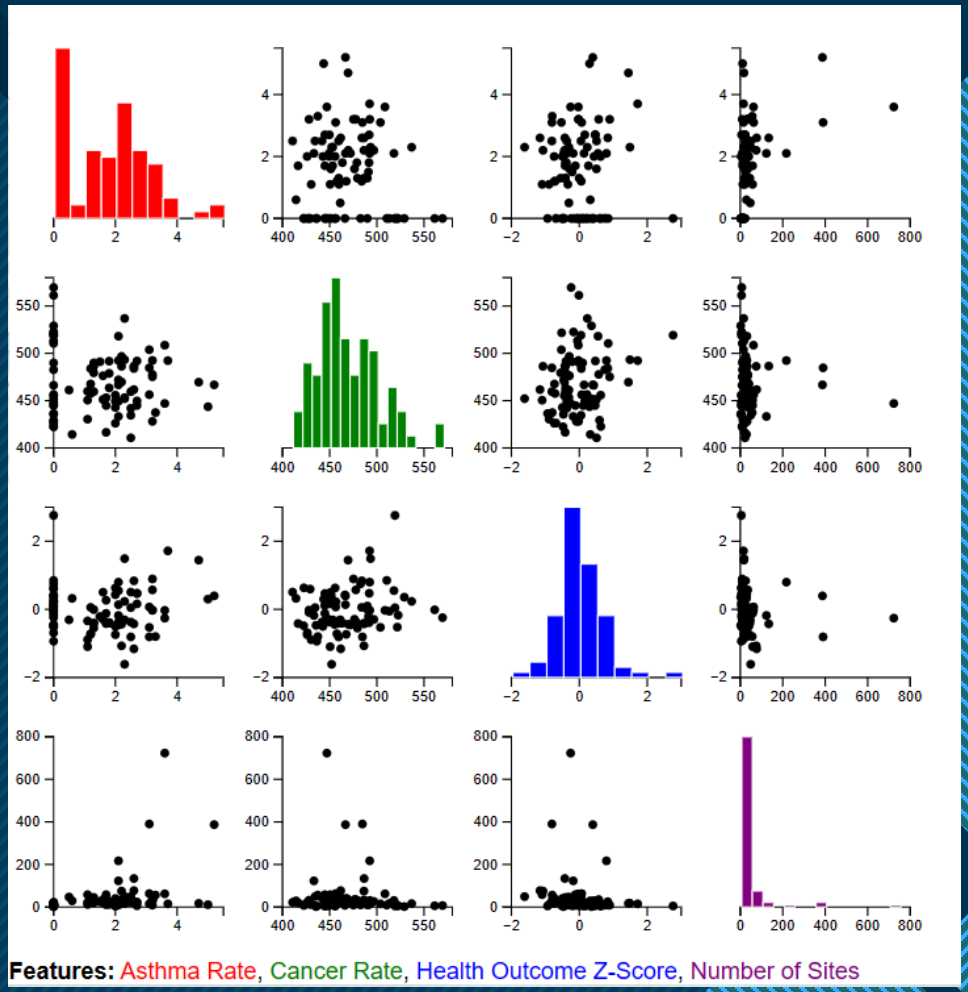
Bar graph

- Dropdown selects health statistic
- Sorts and returns 20 worst counties



Correlation Matrix

- Correlation of health and site data
- Color coded histograms



Conclusion

- We didn’t find any trend between the numbers of EPA sites and cancer rates as well as asthma ED rates in our study.
- The health outcome scores of counties also didn’t show any correlation with the number of sites in the counties.
- We concluded as a team that we would need to explore deeper and include more factors before reaching a conclusion.

“It really boils down to this: that all life is interrelated. We are all caught in an inescapable network of mutuality, tied into a single garment of destiny. Whatever affects one destiny, affects all indirectly.”

—**Dr. Martin Luther King, Jr.** (1929-1968), Christmas Eve sermon, 1967

Questions?

Thank You