Deep Breaths:
Mapping the Relationship Between Air Quality and Asthma Across The United States

Maria De Fatima Alejo and Kun Cai
University of Wisconsin-Madison

GEOG574: Geospatial Database Design and Development

8/19/2025

Abstract

Asthma is one of the most common chronic respiratory diseases in the United States with over 28 million people being affected (*NHIS Adult Summary Health Statistics* | *Data* | *Centers for Disease Control and Prevention*, 2021). Poor air quality, often caused by pollutants from industrial activity, vehicle emissions, and other environmental sources, is a known risk factor for respiratory illnesses, including asthma. Utilizing the data that is publicly available thanks to the CDC and EPA we will integrate the datasets using PostgreGIS and then through ArcGISPro visualize how asthma is affecting those in the United States.

Keywords: Asthma, CDC, EPA, Air Quality

Deep Breaths:
Mapping the Relationship Between Air Quality and Asthma Across The United States

Asthma is a chronic respiratory condition that affects millions of individuals in the United States and is a leading cause of emergency room visits, absences from school and work, and an overall decreased quality of life.

The WHO global air quality guidelines (World Health Organization (WHO), 2021) show that exposure to pollutants such as ozone, a particulate matter known as PM2.5, and nitrogen dioxide can lead to worsening asthma symptoms and increase the prevalence of the disease. The severity and distribution of asthma vary across states and regions and with this paper we will be diving into how it is affecting people across the United States of America.

Through the use of GIS software and a PostgreSQL database we will be integrating data from the CDC with environmental data from the EPA and conducting a spatial analysis, revealing geographic patterns, changes, and areas of concern. Through the use of GIS for visualization of these relationships we can learn how public health researchers and policymakers identify areas where poor air quality and high asthma prevalence overlap.

Through this we aim to answer the following questions.

Where is the highest rate of asthma in the U.S., and what is the average air quality in those states?

Are there states with high asthma but good air quality?

Are there states with poor air quality but low asthma rates?

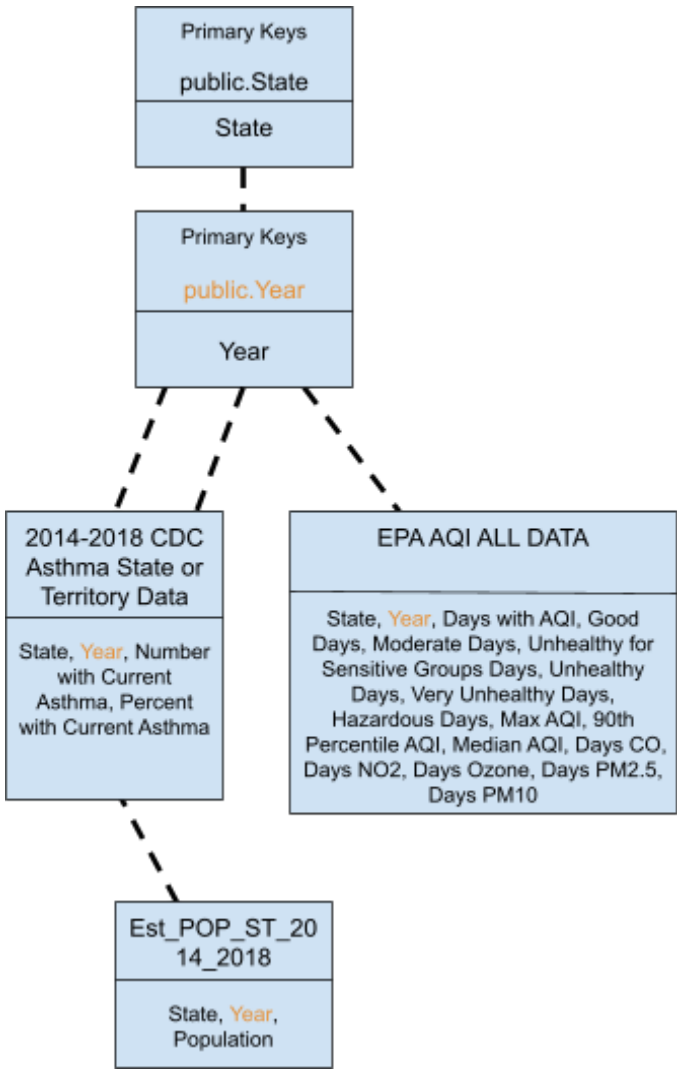
Which states show drastic changes over time in asthma or air quality?

For states with changes, what external factors (policy, environment, population) might explain them?

The database for this project is designed to integrate two main datasets: the CDC's state-level historical asthma prevalence records and the EPA's historical air quality indices along with estimated population totals per state.

To start we decided that the database should be structured around two major factors, the air quality, and the asthma rate of each state. For the air quality table, which we named epa_aqi_all_data, we aggregate records of each state's country to get the average annual measure of air quality, the EPA has a Air Quality Index (AQI) and pollutant types in the public reports as well as degrees of air quality severeness, which we linked to the CDC asthma data by state through a foreign key. For the Asthma Rate table, named cdc_asthma_2014_2018, the CDC captures yearly asthma rate per state so no additional data processing was needed, the data tables provided both total count of asthma afflicted people and percentage in the state overall normalized by the state's population.

The conceptual model diagram is how we formed our foundation to then apply it to the design with PostgreSQL. These tables formed the backbone of our database, and how we were able to run the analytical queries that are shown later in the paper.



Query to create a clean database

```
-- Create a database
-- CREATE DATABASE public_health;
-- \c public_health
```

Query to create reference tables to normalize distinct states and years so fact tables can use foreign keys and we can control data quality

```
DROP TABLE IF EXISTS public."state" CASCADE;
CREATE TABLE public."state" (
  "State" text PRIMARY KEY
);
DROP TABLE IF EXISTS public."year" CASCADE;
CREATE TABLE public."year" (
  "Year" smallint PRIMARY KEY CHECK ("Year" BETWEEN 1900 AND 2100)
);
3) Fact tables (CDC Asthma, EPA AQI, Population)
Purpose: Mirror the cleaned CSV schemas.
-- CDC: % with current asthma per state-year
DROP TABLE IF EXISTS public."cdc_asthma_2014_2018" CASCADE;
CREATE TABLE public."cdc_asthma_2014_2018" (
  "State" text NOT NULL,
  "Year" smallint NOT NULL,
  "Number with Current Asthma" integer,
  "Percent with Current Asthma" numeric(6,3),
  CONSTRAINT cdc_asthma_pk PRIMARY KEY ("State","Year")
);

-- EPA AQI per state-year; includes "Good Days" but NOT any percent column
DROP TABLE IF EXISTS public."epa_aqi_all_data" CASCADE;
CREATE TABLE public."epa_aqi_all_data" (
  "State" text NOT NULL,
  "Year" smallint NOT NULL,
  "Days with AQI" integer,
  "Good Days" integer,
  "Moderate Days" integer,
  "Unhealthy for Sensitive Groups Days" integer,
  "Unhealthy Days" integer,
  "Very Unhealthy Days" integer,
  "Hazardous Days" integer,
  "Max AQI" integer,
  "90th Percentile AQI" integer,
```

```
  "Median AQI" integer,
  "Days CO" integer,
  "Days NO2" integer,
  "Days Ozone" integer,
  "Days PM2.5" integer,
  "Days PM10" integer,
  CONSTRAINT epa_aqi_pk PRIMARY KEY ("State","Year")
);

-- Optional: state population by year
DROP TABLE IF EXISTS public."Est_POP_ST_2014_2018" CASCADE;
CREATE TABLE public."Est_POP_ST_2014_2018" (
  "State" text NOT NULL,
  "Year" smallint NOT NULL,
  "Population" bigint,
  CONSTRAINT pop_pk PRIMARY KEY ("State","Year")
);
```

4) Bulk-load CSVs

Purpose: Import cleaned CDC/EPA/Population CSVs. Adjust file paths to your machine.

```
-- EPA AQI
\copy public."epa_aqi_all_data" ("State","Year","Days with AQI","Good Days","Moderate Days",
"Unhealthy for Sensitive Groups Days","Unhealthy Days","Very Unhealthy Days","Hazardous
Days",
"Max AQI","90th Percentile AQI","Median AQI","Days CO","Days NO2","Days Ozone","Days
PM2.5","Days PM10")
FROM 'C:/Users/Lenovo/Downloads/AQI_cleaned_2014_2018.csv' DELIMITER ',' CSV
HEADER;

-- CDC Asthma
\copy public."cdc_asthma_2014_2018" ("State","Year","Number with Current Asthma","Percent
with Current Asthma")
FROM 'C:/Users/Lenovo/Downloads/Asthma_cleaned.csv' DELIMITER ',' CSV HEADER;

-- Population
\copy public."Est_POP_ST_2014_2018" ("State","Year","Population")
FROM 'C:/Users/Lenovo/Downloads/Population_cleaned.csv' DELIMITER ',' CSV HEADER;
```

5) Seed dimensions & add foreign keys

Purpose: Fill state/year from distinct values across all facts; then enforce referential integrity on facts.

```
-- Populate distinct states
TRUNCATE public."state";
INSERT INTO public."state"("State")
SELECT DISTINCT "State" FROM public."epa_aqi_all_data"
UNION SELECT DISTINCT "State" FROM public."cdc_asthma_2014_2018"
```

```
UNION SELECT DISTINCT "State" FROM public."Est_POP_ST_2014_2018"
ORDER BY 1;

-- Populate distinct years
TRUNCATE public."year";
INSERT INTO public."year"("Year")
SELECT DISTINCT "Year" FROM public."epa_aqi_all_data"
UNION SELECT DISTINCT "Year" FROM public."cdc_asthma_2014_2018"
UNION SELECT DISTINCT "Year" FROM public."Est_POP_ST_2014_2018"
ORDER BY 1;

-- Add FKs
ALTER TABLE public."epa_aqi_all_data"
  ADD CONSTRAINT fk_aqi_state FOREIGN KEY ("State") REFERENCES
public."state"("State"),
  ADD CONSTRAINT fk_aqi_year  FOREIGN KEY ("Year")  REFERENCES
public."year"("Year");

ALTER TABLE public."cdc_asthma_2014_2018"
  ADD CONSTRAINT fk_asthma_state FOREIGN KEY ("State") REFERENCES
public."state"("State"),
  ADD CONSTRAINT fk_asthma_year  FOREIGN KEY ("Year")  REFERENCES
public."year"("Year");

ALTER TABLE public."Est_POP_ST_2014_2018"
  ADD CONSTRAINT fk_pop_state FOREIGN KEY ("State") REFERENCES
public."state"("State"),
  ADD CONSTRAINT fk_pop_year  FOREIGN KEY ("Year")  REFERENCES
public."year"("Year");
```

6) Analytical views (join once, reuse everywhere)

6.1 v_state_year

Purpose: One row per State–Year combining CDC + EPA (+ Population) and deriving good_days_pct from your existing "Good Days" and "Days with AQI". We keep good_days_pct as a computed column here for all downstream queries.

```
DROP VIEW IF EXISTS public."v_state_year";
CREATE OR REPLACE VIEW public."v_state_year" AS
SELECT
  s."State",
  y."Year",
  a."Number with Current Asthma",
  a."Percent with Current Asthma",
  q."Days with AQI",
  q."Good Days",
  q."Moderate Days",
```

```
  q."Unhealthy for Sensitive Groups Days",
  q."Unhealthy Days",
  q."Very Unhealthy Days",
  q."Hazardous Days",
  q."Max AQI",
  q."90th Percentile AQI",
  q."Median AQI",
  q."Days CO",
  q."Days NO2",
  q."Days Ozone",
  q."Days PM2.5",
  q."Days PM10",
  p."Population",
  CASE
    WHEN q."Days with AQI" > 0
      THEN ROUND(q."Good Days"::numeric / q."Days with AQI"::numeric * 100, 2)
  END AS good_days_pct   -- <-- derived from "Good Days" and "Days with AQI"
FROM public."state" s
CROSS JOIN public."year" y
LEFT JOIN public."cdc_asthma_2014_2018" a
  ON a."State" = s."State" AND a."Year" = y."Year"
LEFT JOIN public."epa_aqi_all_data" q
  ON q."State" = s."State" AND q."Year" = y."Year"
LEFT JOIN public."Est_POP_ST_2014_2018" p
  ON p."State" = s."State" AND p."Year" = y."Year";
```

6.2 v_state_year_any

Purpose: Keep only rows where both asthma and AQI metrics are present—this avoids averages or ranks being skewed by missing data.

```
DROP VIEW IF EXISTS public.v_state_year_any;
CREATE OR REPLACE VIEW public.v_state_year_any AS
SELECT
  vs."State",
  vs."Year",
  vs."Percent with Current Asthma" AS asthma_pct,
  vs."Median AQI"                  AS median_aqi,
  vs.good_days_pct                 AS good_days_pct
FROM public."v_state_year" vs
WHERE vs."Year" BETWEEN 2014 AND 2018
  AND vs."Percent with Current Asthma" IS NOT NULL
  AND vs."Median AQI" IS NOT NULL
  AND vs.good_days_pct IS NOT NULL;
```

6.3 v_state_avg_any

Purpose: Compute per-state averages across 2014–2018 for the three key indicators.

```
DROP VIEW IF EXISTS public.v_state_avg_any;
```

```
CREATE OR REPLACE VIEW public.v_state_avg_any AS
SELECT
  "State",
  AVG(asthma_pct)    AS asthma_pct_avg,
  AVG(median_aqi)    AS median_aqi_avg,
  AVG(good_days_pct) AS good_days_pct_avg,
  COUNT(*)           AS n_years
FROM public.v_state_year_any
GROUP BY "State";
```

6.4 v_state_deltas_any

Purpose: For each state, compute endpoint-to-endpoint changes (first vs. last available years in 2014–2018) for asthma %, Median AQI, and Good-Days-% so we can rank “drastic changes”.

```
DROP VIEW IF EXISTS public.v_state_deltas_any;
CREATE OR REPLACE VIEW public.v_state_deltas_any AS
WITH
-- Asthma timeline
asthma_base AS (
  SELECT "State" AS state, "Year" AS year,
         "Percent with Current Asthma"::numeric AS asthma_pct
  FROM public."cdc_asthma_2014_2018"
  WHERE "Year" BETWEEN 2014 AND 2018
),
asthma_first AS (
  SELECT state, year AS asthma_y1, asthma_pct AS asthma_1 FROM (
    SELECT state, year, asthma_pct,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year ASC) AS rn
    FROM asthma_base WHERE asthma_pct IS NOT NULL
  ) t WHERE rn = 1
),
asthma_last AS (
  SELECT state, year AS asthma_y2, asthma_pct AS asthma_2 FROM (
    SELECT state, year, asthma_pct,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year DESC) AS rn
    FROM asthma_base WHERE asthma_pct IS NOT NULL
  ) t WHERE rn = 1
),

-- AQI & Good-Days-% timeline (derived % from "Good Days" / "Days with AQI")
aqi_base AS (
  SELECT
    "State" AS state, "Year" AS year,
    "Median AQI"::numeric AS median_aqi,
    CASE WHEN "Days with AQI"::numeric > 0
```

```
        THEN ("Good Days"::numeric / "Days with AQI"::numeric) * 100
    END AS good_days_pct
  FROM public."epa_aqi_all_data"
  WHERE "Year" BETWEEN 2014 AND 2018
),

-- Median AQI endpoints
aqi_m_first AS (
  SELECT state, year AS aqi_y1, median_aqi AS aqi_1 FROM (
    SELECT state, year, median_aqi,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year ASC) AS rn
    FROM aqi_base WHERE median_aqi IS NOT NULL
  ) t WHERE rn = 1
),
aqi_m_last AS (
  SELECT state, year AS aqi_y2, median_aqi AS aqi_2 FROM (
    SELECT state, year, median_aqi,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year DESC) AS rn
    FROM aqi_base WHERE median_aqi IS NOT NULL
  ) t WHERE rn = 1
),

-- Good-Days-% endpoints
good_first AS (
  SELECT state, year AS good_y1, good_days_pct AS good_1 FROM (
    SELECT state, year, good_days_pct,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year ASC) AS rn
    FROM aqi_base WHERE good_days_pct IS NOT NULL
  ) t WHERE rn = 1
),
good_last AS (
  SELECT state, year AS good_y2, good_days_pct AS good_2 FROM (
    SELECT state, year, good_days_pct,
           ROW_NUMBER() OVER (PARTITION BY state ORDER BY year DESC) AS rn
    FROM aqi_base WHERE good_days_pct IS NOT NULL
  ) t WHERE rn = 1
),

-- State universe
states AS (
  SELECT DISTINCT state FROM asthma_base
  UNION
  SELECT DISTINCT state FROM aqi_base
)
```

```
SELECT
  s.state AS "State",

  -- asthma endpoints & delta
  af.asthma_y1, al.asthma_y2,
  CASE WHEN af.asthma_y1 IS DISTINCT FROM al.asthma_y2
       THEN ROUND(al.asthma_2 - af.asthma_1, 3) END AS asthma_delta,

  -- median AQI endpoints & delta
  mf.aqi_y1, ml.aqi_y2,
  CASE WHEN mf.aqi_y1 IS DISTINCT FROM ml.aqi_y2
       THEN (ml.aqi_2 - mf.aqi_1) END AS median_aqi_delta,

  -- good-days-% endpoints & delta
  gf.good_y1, gl.good_y2,
  CASE WHEN gf.good_y1 IS DISTINCT FROM gl.good_y2
       THEN ROUND(gl.good_2 - gf.good_1, 2) END AS good_days_pct_delta
FROM states s
LEFT JOIN asthma_first af ON s.state = af.state
LEFT JOIN asthma_last  al ON s.state = al.state
LEFT JOIN aqi_m_first  mf ON s.state = mf.state
LEFT JOIN aqi_m_last   ml ON s.state = ml.state
LEFT JOIN good_first   gf ON s.state = gf.state
LEFT JOIN good_last    gl ON s.state = gl.state;
```

7) Queries that answer the 5 research questions

RQ1: “Where is the highest rate of asthma; what is that state’s average air quality?”

Purpose: Rank by average asthma% and read the same row’s average AQI (lower AQI = better air).

```
SELECT "State", n_years,
       ROUND(asthma_pct_avg, 3)    AS asthma_pct_avg,
       ROUND(median_aqi_avg, 3)    AS median_aqi_avg,
       ROUND(good_days_pct_avg, 2) AS good_days_pct_avg
FROM public.v_state_avg_any
ORDER BY asthma_pct_avg DESC
LIMIT 10;  -- set to 1 to get the single highest state
```

RQ2: “Any states with high asthma and good air?”

Purpose: Use robust percentile thresholds: top-25% asthma and bottom-25% median AQI (good air).

```
WITH base AS (
  SELECT * FROM public.v_state_avg_any
),
```

```
th AS (
  SELECT
    percentile_cont(0.75) WITHIN GROUP (ORDER BY asthma_pct_avg) AS asthma_p75,
    percentile_cont(0.25) WITHIN GROUP (ORDER BY median_aqi_avg) AS aqi_p25
  FROM base
)
SELECT b."State", b.n_years,
       ROUND(b.asthma_pct_avg, 3)    AS asthma_pct_avg,
       ROUND(b.median_aqi_avg, 3)    AS median_aqi_avg,
       ROUND(b.good_days_pct_avg, 2) AS good_days_pct_avg
FROM base b CROSS JOIN th t
WHERE b.asthma_pct_avg >= t.asthma_p75
  AND b.median_aqi_avg <= t.aqi_p25
ORDER BY b.asthma_pct_avg DESC, b.median_aqi_avg ASC;
```

RQ3: “Any states with bad air and low asthma?”

Purpose: Opposite tails: bottom-25% asthma and top-25% median AQI (bad air).

```
WITH base AS (
  SELECT * FROM public.v_state_avg_any
),
th AS (
  SELECT
    percentile_cont(0.25) WITHIN GROUP (ORDER BY asthma_pct_avg) AS asthma_p25,
    percentile_cont(0.75) WITHIN GROUP (ORDER BY median_aqi_avg) AS aqi_p75
  FROM base
)
SELECT b."State", b.n_years,
       ROUND(b.asthma_pct_avg, 3)    AS asthma_pct_avg,
       ROUND(b.median_aqi_avg, 3)    AS median_aqi_avg,
       ROUND(b.good_days_pct_avg, 2) AS good_days_pct_avg
FROM base b CROSS JOIN th t
WHERE b.asthma_pct_avg <= t.asthma_p25
  AND b.median_aqi_avg >= t.aqi_p75
ORDER BY b.median_aqi_avg DESC, b.asthma_pct_avg ASC;
```

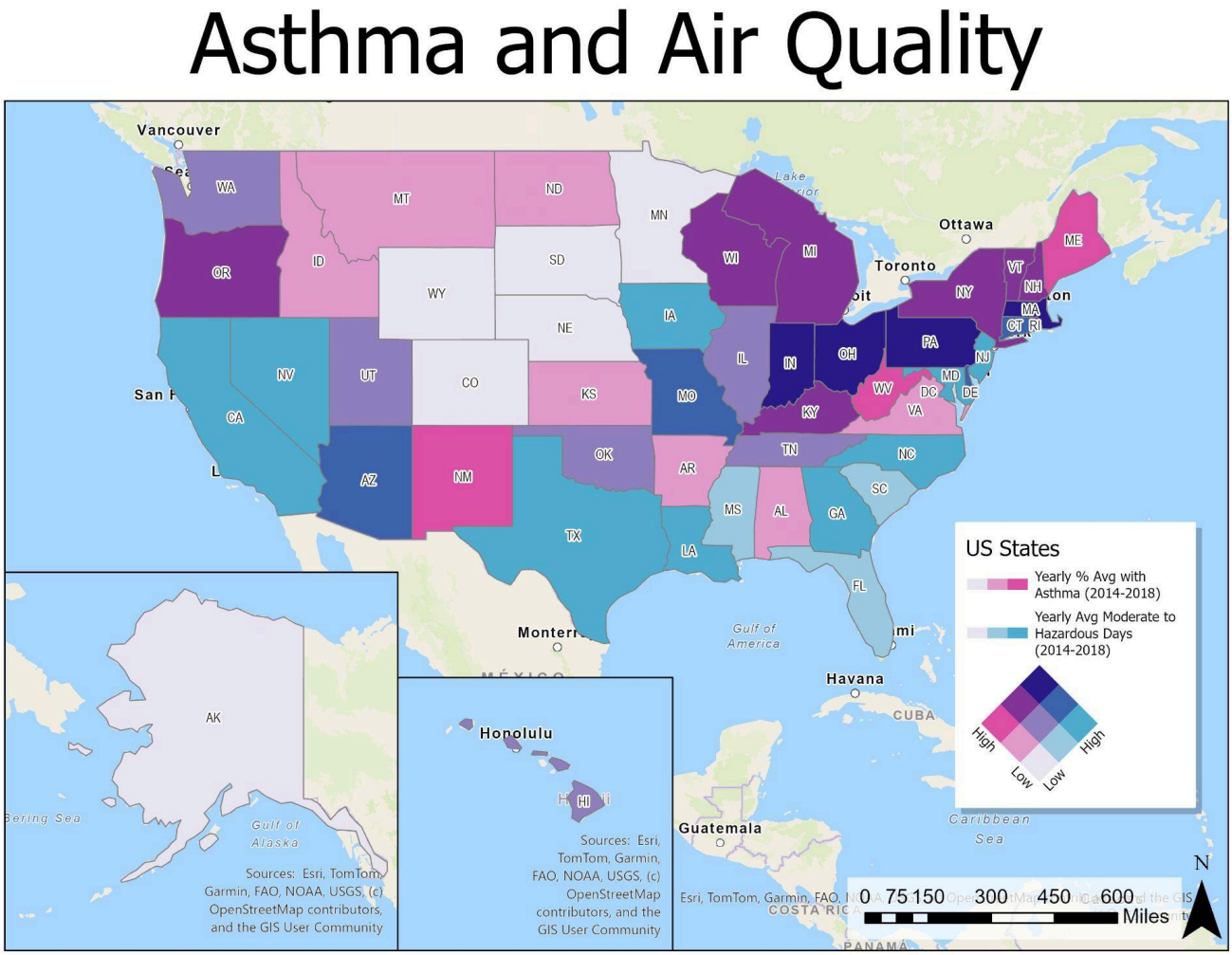
RQ4: “Which states show drastic changes over years (asthma & air quality)?”

Purpose: Two leaderboards using endpoint deltas. Sort by absolute change; keep both metrics present.

```
SELECT "State",
       asthma_y1, asthma_y2,
       ROUND(asthma_delta, 3)      AS asthma_delta,
       aqi_y1, aqi_y2,
       ROUND(median_aqi_delta, 3)  AS median_aqi_delta,
       good_y1, good_y2,
```

```
      ROUND(good_days_pct_delta, 2) AS good_days_pct_delta
FROM public.v_state_deltas_any
WHERE asthma_delta IS NOT NULL
  AND median_aqi_delta IS NOT NULL
ORDER BY ABS(asthma_delta) DESC, ABS(median_aqi_delta) DESC
LIMIT 20;
(b) Biggest AQI swings
SELECT "State",
      aqi_y1, aqi_y2,
      ROUND(median_aqi_delta, 3)  AS median_aqi_delta,
      asthma_y1, asthma_y2,
      ROUND(asthma_delta, 3)      AS asthma_delta,
      good_y1, good_y2,
      ROUND(good_days_pct_delta, 2) AS good_days_pct_delta
FROM public.v_state_deltas_any
WHERE median_aqi_delta IS NOT NULL
  AND asthma_delta IS NOT NULL
ORDER BY ABS(median_aqi_delta) DESC, ABS(asthma_delta) DESC
LIMIT 20;
```

Mapped Results



Through this we were able to answer our initial questions.

Where is the highest rate of asthma, and what's the average air quality there?

We found that the highest average asthma was in West Virginia with 11.8% of the population having asthma, and the average median AQI was around 38.8.

New Hampshire and Maine come in close with 11.62% and an average median AQI of 38.6 for NH and 11.48% and an average median AQI of 36.2.

Are there states with high asthma but good air quality?

Vermont had an average of 11.2% of the population affected by asthma with their average median AQI at 32.6, making them in the bottom 25% of air pollution but top 10% of asthma rates.

Are there states with poor air quality but low asthma rates?

Yes, the top 3 that rated low on asthma but high for air pollution were; California with an average asthma rate of 7.82% and an average median AQI of 52.4. Georgia with only 8.56% of the population having asthma but with an average median AQI of 44.0 and finally Iowa with the asthma rate at 8.5% but an average median AQI of 43.6.

Which states show drastic changes over time (2014–2018)?

For the asthma rate New Hampshire has the most variation throughout the sampled years, a variation of +/- 3 points with Nevada coming in second with an average of +/- 2.5 points.

For the median AQI ranges Puerto Rico averaged a variation of +/- 12 points throughout the sampled years and Rhode Island coming in second with an average of +/- 8 points.

References

NHIS Adult Summary Health Statistics | Data | Centers for Disease Control and Prevention. (2021, November 17). Data.CDC.gov. Retrieved August 17, 2025, from https://data.cdc.gov/d/25m4-6qqq

World Health Organization (WHO). (2021). WHO global air quality guidelines. World Health Organization (WHO). Retrieved August 17, 2025, from https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf