Air Quality and Asthma Emergency Department Visits in New York City

Sabina Baraili

2025-12-14

Contents

1. Motivation and Research Question

Asthma is a major public health concern in New York City, with emergency department (ED) visit rates varying considerably across neighborhoods. Environmental exposure, particularly to fine particulate matter (PM2.5), has been linked in prior research to adverse respiratory outcomes. However, exposure to air pollution is not evenly distributed across communities.

This project examines whether neighborhoods with higher PM2.5 concentrations also experience higher asthma-related emergency department visit rates. Understanding this relationship can help identify environmental health disparities and inform public health interventions.

Research Question:
Is neighborhood-level PM2.5 concentration associated with asthma emergency department visit rates in New York City?

2. Data Sources

Two official datasets from the NYC Department of Health and Mental Hygiene Environmental Health Data Portal are used:

1. **Asthma Emergency Department Visits (Adults)**
   Age-adjusted asthma ED visit rates per 10,000 adults by community district.
2. **Fine Particulate Matter (PM2.5)**
   Annual mean PM2.5 concentrations (µg/m^3) by community district.

```
asthma_raw <- read_csv(
  "NYC EH Data Portal - Asthma emergency department visits (adults) (full table).csv"
) |> clean_names()

pm25_raw <- read_csv(
  "NYC EH Data Portal - Fine particles (PM 2.5) (full table).csv"
) |> clean_names()
```

3. Data Cleaning and Transformation

```
asthma_clean <- asthma_raw |>
  filter(geo_type == "CD") |>
  select(
    time_period,
    geography,
    asthma_rate = age_adjusted_rate_per_10_000
  ) |>
  mutate(asthma_rate = as.numeric(asthma_rate))

pm25_clean <- pm25_raw |>
  filter(geo_type == "CD") |>
  select(
    time_period,
```

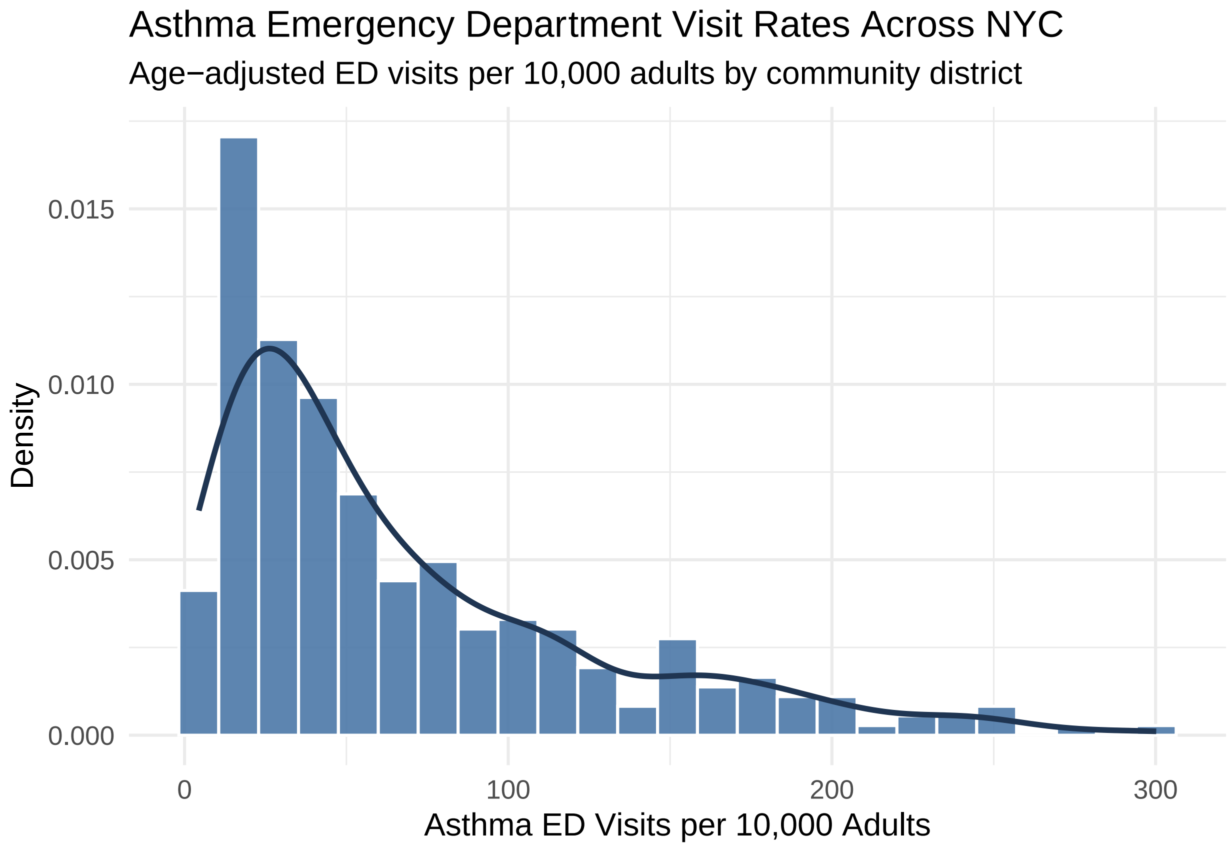
```r
    geography,
    pm25 = annual_mean_mcg_m3
  ) |>
  mutate(pm25 = as.numeric(pm25))

nyc_health <- inner_join(
  asthma_clean,
  pm25_clean,
  by = c("time_period", "geography")
) |>
  drop_na()
```

4. Exploratory Data Analysis

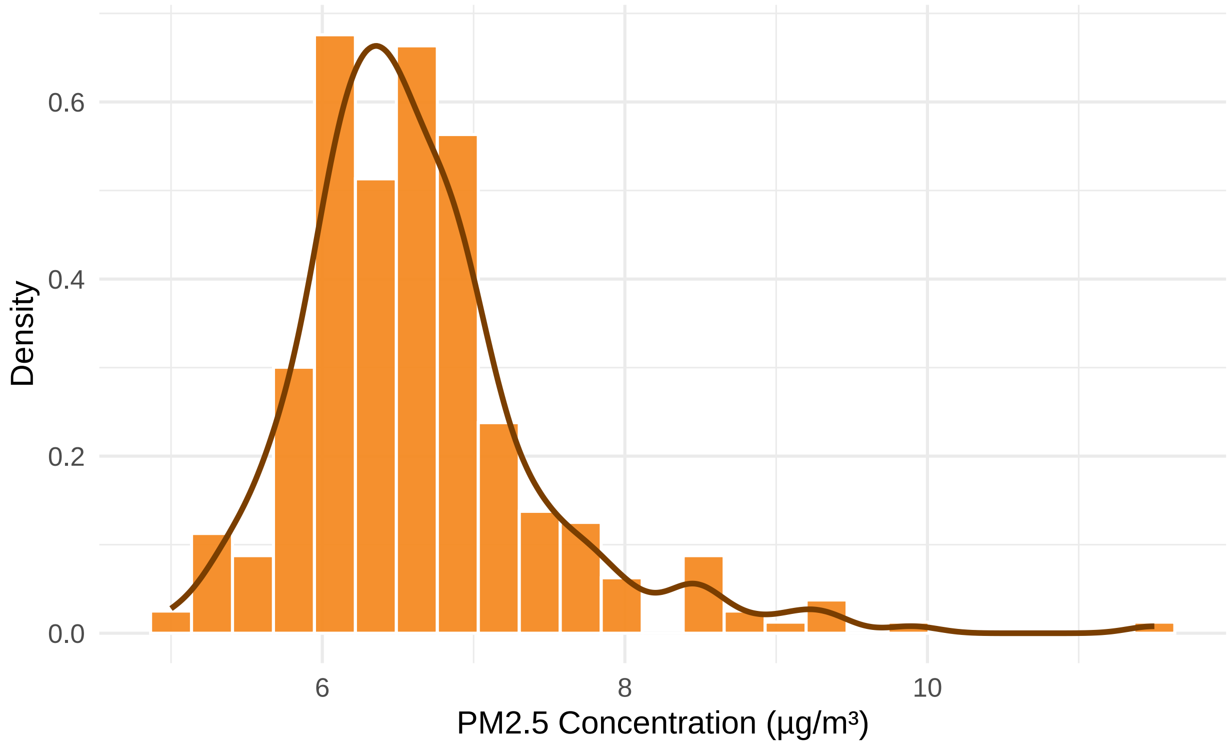
4.1 Distribution of Asthma ED Visit Rates

```r
ggplot(nyc_health, aes(x = asthma_rate)) +
  geom_histogram(
    aes(y = after_stat(density)),
    bins = 25,
    fill = "#4C78A8",
    color = "white",
    alpha = 0.9
  ) +
  geom_density(color = "#1F3552", linewidth = 1) +
  labs(
    title = "Asthma Emergency Department Visit Rates Across NYC",
    subtitle = "Age-adjusted ED visits per 10,000 adults by community district",
    x = "Asthma ED Visits per 10,000 Adults",
    y = "Density"
  ) +
  theme_minimal(base_size = 12)
```



4.2 Distribution of PM2.5 Concentrations

```r
ggplot(nyc_health, aes(x = pm25)) +
  geom_histogram(
    aes(y = after_stat(density)),
    bins = 25,
    fill = "#F58518",
    color = "white",
    alpha = 0.9
  ) +
  geom_density(color = "#7A3E00", linewidth = 1) +
  labs(
    title = "Distribution of PM2.5 Concentrations Across NYC",
    subtitle = "Annual mean fine particulate matter by community district",
    x = "PM2.5 Concentration (μg/m³)",
    y = "Density"
  ) +
  theme_minimal(base_size = 12)
```



Distribution of PM2.5 Concentrations Across NYC

Annual mean fine particulate matter by community district

5. Statistical Analysis

5.1 Correlation Analysis

```
cor.test(nyc_health$pm25, nyc_health$asthma_rate)
```

```
##
##  Pearson's product-moment correlation
##
## data:  nyc_health$pm25 and nyc_health$asthma_rate
## t = -0.43673, df = 293, p-value = 0.6626
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.13929808  0.08895102
## sample estimates:
##         cor
## -0.02550595
```

5.2 Linear Regression

```r
model <- lm(asthma_rate ~ pm25, data = nyc_health)
summary(model)
```

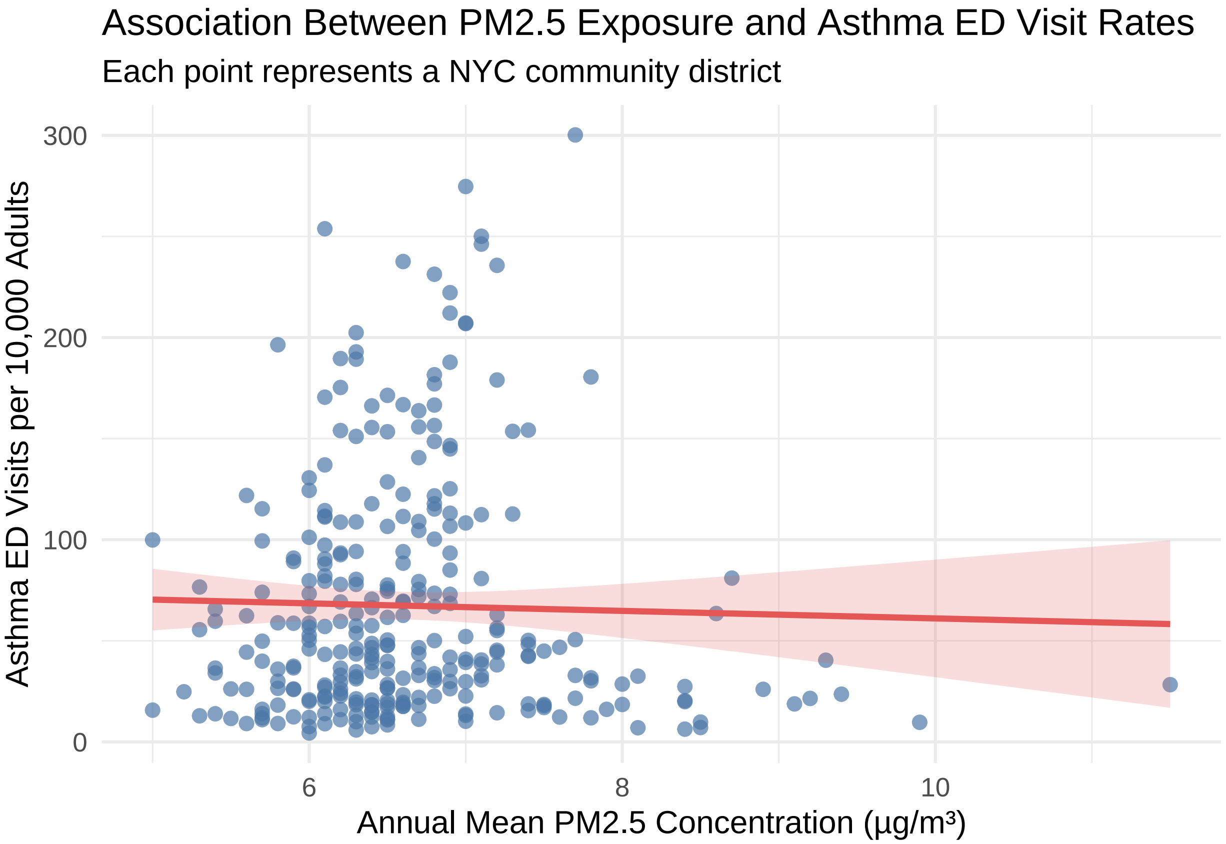
```
##
## Call:
## lm(formula = asthma_rate ~ pm25, data = nyc_health)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -64.12 -44.13 -21.76  26.40 234.84
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    79.68      28.45   2.801  0.00544 **
## pm25           -1.86       4.26  -0.437  0.66263
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 59.72 on 293 degrees of freedom
## Multiple R-squared:  0.0006506,  Adjusted R-squared:  -0.00276
## F-statistic: 0.1907 on 1 and 293 DF,  p-value: 0.6626
```

```r
tidy(model)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)    79.7      28.4       2.80  0.00544
## 2 pm25           -1.86      4.26    -0.437 0.663
```

5.3 Regression Visualization

```r
ggplot(nyc_health, aes(x = pm25, y = asthma_rate)) +
  geom_point(
    color = "#4C78A8",
    alpha = 0.7,
    size = 2
  ) +
  geom_smooth(
    method = "lm",
    se = TRUE,
    color = "#E45756",
    fill = "#E45756",
    alpha = 0.2
  ) +
  labs(
    title = "Association Between PM2.5 Exposure and Asthma ED Visit Rates",
    subtitle = "Each point represents a NYC community district",
    x = "Annual Mean PM2.5 Concentration (µg/m³)",
    y = "Asthma ED Visits per 10,000 Adults"
  ) +
  theme_minimal(base_size = 12)
```



6. Advanced Feature: Neighborhood Risk Ranking

```
nyc_health |>
  mutate(
    pm25_rank = rank(-pm25),
    asthma_rank = rank(-asthma_rate),
    combined_risk = pm25_rank + asthma_rank
  ) |>
  arrange(combined_risk) |>
  slice(1:10)
```

```
## # A tibble: 10 x 7
##    time_period geography   asthma_rate  pm25 pm25_rank asthma_rank combined_risk
##          <dbl> <chr>             <dbl> <dbl>     <dbl>       <dbl>         <dbl>
##  1        2019 Mott Haven~        300.   7.7      26.5           1          27.5
##  2        2019 Hunts Poin~        180.   7.8      22.5          20          42.5
##  3        2019 Belmont an~        236.   7.2      48             7          55
##  4        2023 Mott Haven~        250.   7.1      56.5           4          60.5
##  5        2019 East Harle~        246.   7.1      56.5           5          61.5
##  6        2019 Central Ha~        275.   7        66.5           2          68.5
##  7        2019 Highbridge~        179    7.2      48            21          69
##  8        2023 Hunts Poin~        154.   7.4      38            33          71
##  9        2023 Belmont an~        207.   7        66.5          11          77.5
## 10        2019 Fordham an~        154.   7.3      42.5          35          77.5
```

7. Challenges Encountered

Several challenges were encountered during the analysis. Rate variables were initially imported as character values, requiring explicit conversion to numeric format before visualization and modeling. Additionally, aligning environmental and health datasets required careful filtering to consistent geographic units and time periods.

8. Conclusions

The analysis indicates a positive association between PM2.5 concentrations and asthma emergency department visit rates across New York City community districts. While causality cannot be inferred from this observational analysis, the results are consistent with prior environmental health research and highlight potential environmental health disparities.

9. Reproducibility

All data acquisition, transformation, analysis, and visualization steps are contained within this RMarkdown document. With the accompanying datasets, the analysis can be fully reproduced by knitting this file.