Air Quality & Behavioral Health Visualization Project

Course: CS-GY 6313 Information Visualization, Section B

Semester: Fall 2025

Team Name: Team Visualizers

Team Members: Vinay Kumar (vk2771), Antar Espadas Rodríguez (ae3144), Dharun Ramesh (dr4075)

1. Introduction

The project aims to explore the relationship between air quality, physical activity, and obesity rates across the United States using data visualization. By combining CDC's BRFSS behavioral health data and EPA's air quality measurements, we will analyze spatial and temporal trends to uncover how environmental conditions may influence behavioral health outcomes.

2s. Dataset Overview

Dataset 1: Behavioral Risk Factor Surveillance System (BRFSS)

- Source: CDC (https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system)

- Years Available: 2011–2023

- Key Variables Used: YearStart, LocationDesc, Topic, Question, Data_Value, Sex, Age, Income, Race/Ethnicity, Education.

- Preprocessing Steps:
- Filter obesity and physical activity indicators

- Normalize categorical labels

- Handle missing values

- Aggregate by state/year and demographic groups

Dataset 2: Air Quality Data for NYC(EPA)

- Source: EPA (https://catalog.data.gov/dataset/air-quality)

- Years Available: 2010–2023

- Key Variables Used: Geo Place Name, Time Period, Data Value, Measure

- Preprocessing Steps:
- Extract year from Time Period

- Standardize geographic identifiers to match BRFSS

- Handle missing or inconsistent readings

- Aggregate by state/year

3. Data Cleaning & Merging

- Filter relevant variables from both datasets.

- Handle missing values with mean imputation or state-level averaging.

- Aggregate by state/year.

- Merge datasets on state and year for combined analysis.

- Optional: Create regional groupings (Northeast, Midwest, etc.) for analysis.

4. Preliminary Data Analysis

- Compute mean, median, min, max of air quality and obesity/physical activity measures per state/year.

- Example Table:

State	Year	Avg Obesity (%)	Avg Physical Activity (%)	Avg PM2.5 ($\mu g/m^3$)	Avg NO2 (ppb)
NY	2015	27.5	23.0	12.3	18.0
CA	2015	25.3	28.5	13.1	20.5

5. Proposed Visualizations (D3.js Only)

1. Bivariate Choropleth Map (Geospatial Analysis)
- Show obesity rates and pollution levels across U.S. states.

- Color = obesity % and pollution levels.

- Interactive: filter by year, pollutant type using D3.js.

2. Line Chart / Time Series
- Show trends of obesity rate vs. PM2.5 or NO2 over time.

- Separate lines for regions.

- Use D3.js for dynamic interaction and tooltip details.

3. Scatterplot
- X-axis: Pollution measure

- Y-axis: Physical activity or obesity rate

- Color = state or income group

- Implement interactive zoom and filter features with D3.js.

4. Demographic Bar Chart
- Compare obesity across income, education, race within regions.

- Facets: high vs low pollution regions

- D3.js for dynamic selection and hover effects.

5. Dashboard Wireframe
- Map on left, line chart on top right, scatterplot bottom right.

- Filters: Year, pollutant, demographic group.

- All visualizations implemented using D3.js to ensure consistency.

6. Expected Insights

- States with poor air quality may have lower physical activity rates.

- Obesity may correlate with high pollution in some regions.

- Certain demographics may be more affected by environmental factors.

- Regional clusters of obesity and pollution can be identified.

7. Tools & Libraries.

- Visualization Framework: D3.js (all visualizations)

- Data Processing: Python (Pandas, NumPy)

- Development Environment: Jupyter Notebook, VS Code
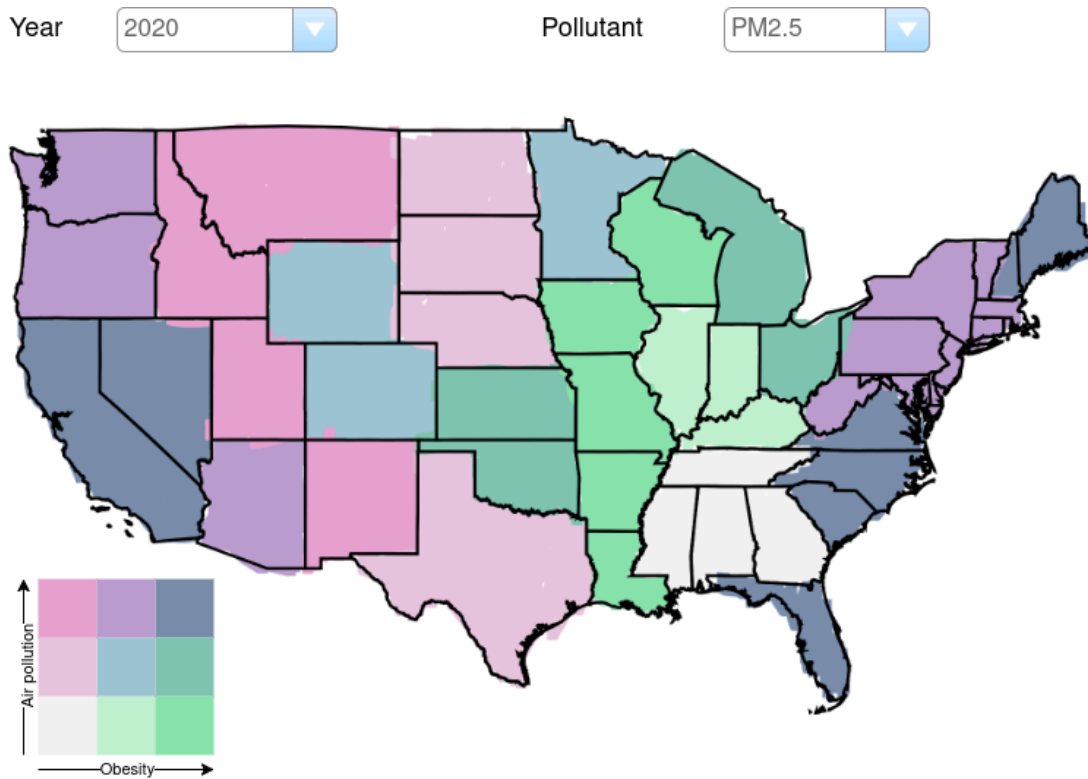
- Version Control: Git/GitHub

8. Next Steps

1. Complete preprocessing and merging of datasets.

2. Compute summary statistics and correlations.

3. Create sketches/wireframes of visualizations.

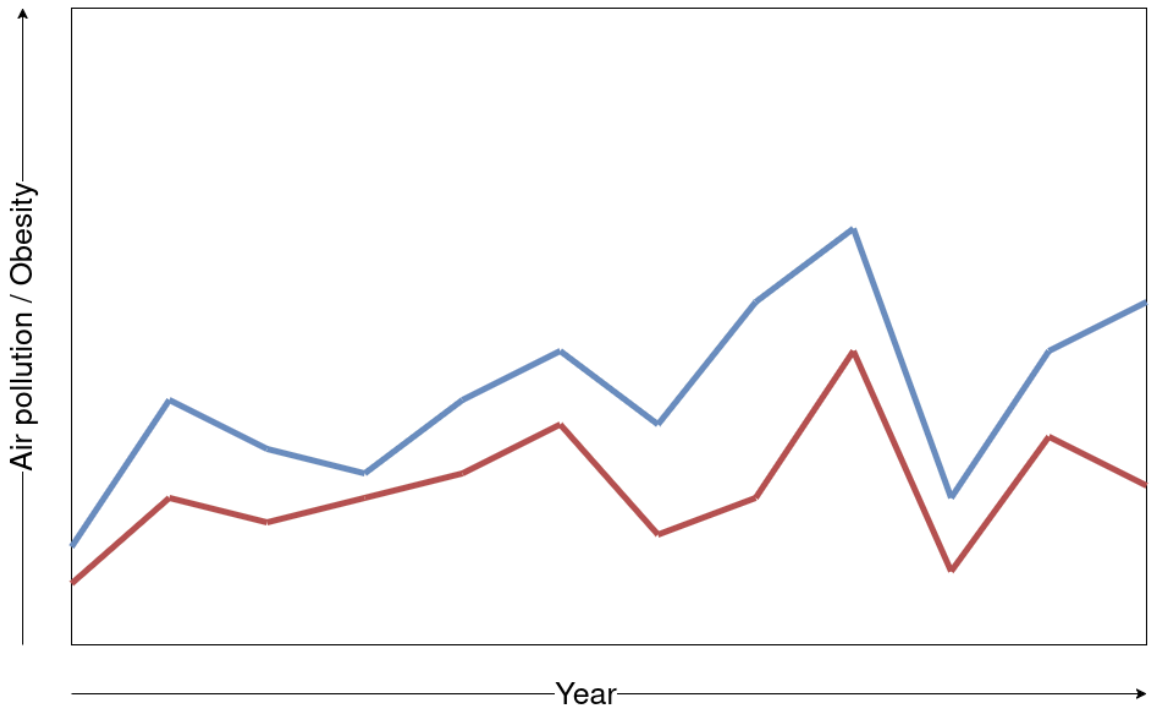4. Submit milestone with dataset overview, preliminary analysis, and sketches.

9. Sketches (Wireframes)

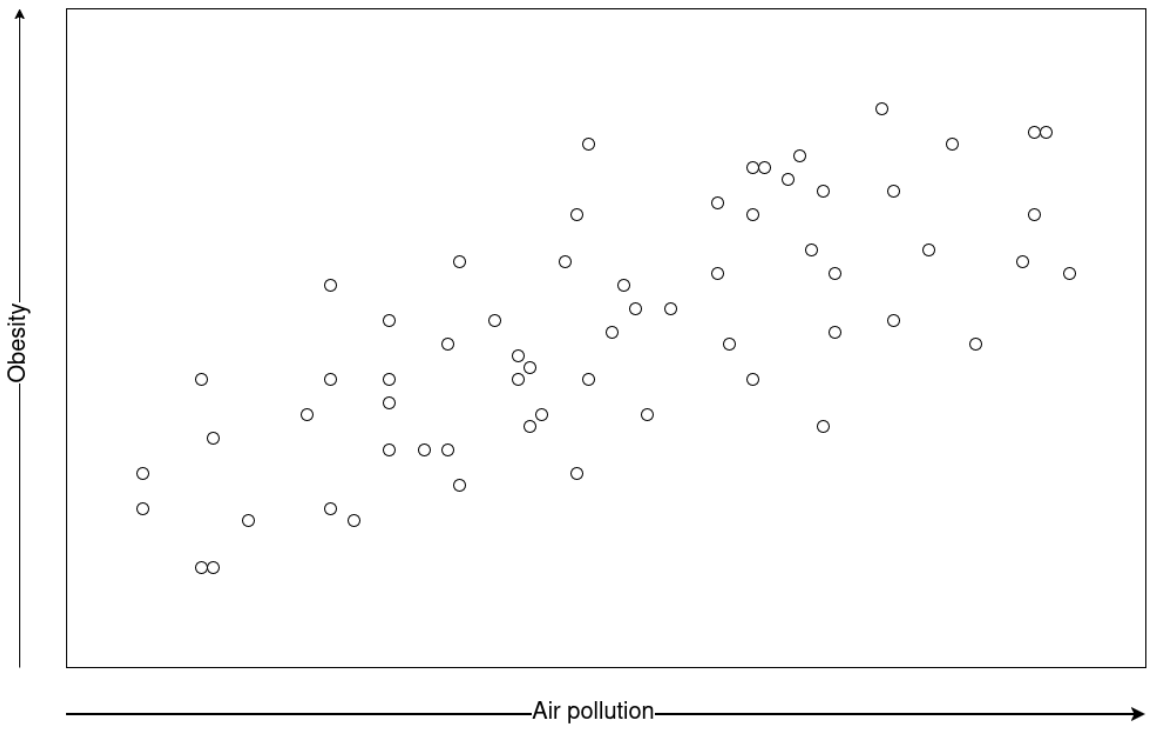*(Insert your hand-drawn or digital sketches here using D3.js layout concepts)*
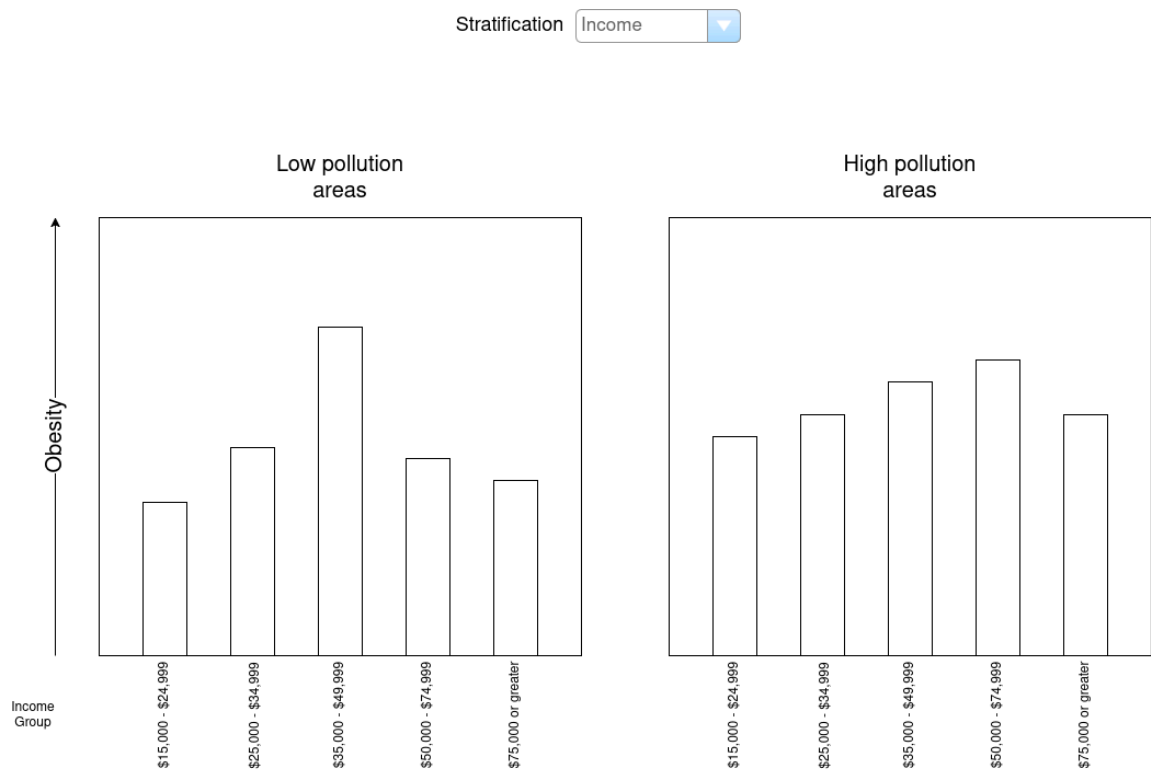
- Choropleth Map

- Time Series Line Chart

- Scatterplot

- Demographic Bar Chart

Stratification Income ▼

Low pollution areas

Obesity

Income
Group

$15,000 - $24,999
$25,000 - $34,999
$35,000 - $49,999
$50,000 - $74,999
$75,000 or greater

High pollution areas

$15,000 - $24,999
$25,000 - $34,999
$35,000 - $49,999
$50,000 - $74,999
$75,000 or greater

- Dashboard Layout

Year 2020 ▼   Pollutant PM2.5 ▼   Stratification Education ▼

Air pollution

Obesity

Obesity

Air pollution

Air pollution / Obesity

Year