Ping Jiang

Prof. Rebin A Muhammad

DATA 110

17 December 2025

<div align="center">Data Final Project Report</div>

Part 1. Introduction

I wanted to do a project that's related to oral health, because it is a critical but often overlooked component of overall public health. Poor oral health is linked to chronic disease, reduced quality of life, and health inequities across populations. For this project, I used data from the CDC National Oral Health Surveillance System (NOHSS) Adult Indicators to explore patterns in adult oral health across the United States. It combines oral health outcomes, preventive behaviors, and geographic coverage, making it ideal for learning and demonstrating skills I've learned from Prof. Muhammad, such as data cleaning, exploratory data analysis (EDA), aggregation, and data visualization. The dataset also allows for meaningful storytelling about public health disparities and long-term health outcomes.

My goal of this project is to use clear Python visualizations to understand trends, distributions, and geographic differences in adult oral health and to communicate these findings effectively to the class.

The body of my report includes:

- Overview of EDA process and key findings

- Detailed discussion of each visualization and the insights they reveal

- A conclusion and possible next steps to explore

- Glossary of terms from the dataset

Part 2. Exploratory Data Analysis (EDA) Overview and Key Findings

My EDA Focused On:

- Understanding the structure of the dataset (columns, indicators, time range)

- Identifying the types of indicators included (behaviors vs. outcomes)

- Examining distributions of values

- Exploring trends over time

- Comparing values across geographic regions

The Steps I Took:

- Checking unique indicators and categories

- Identifying missing or inconsistent values

- Aggregating data by year, state, and indicator where appropriate

- Exploring summary statistics and distributions
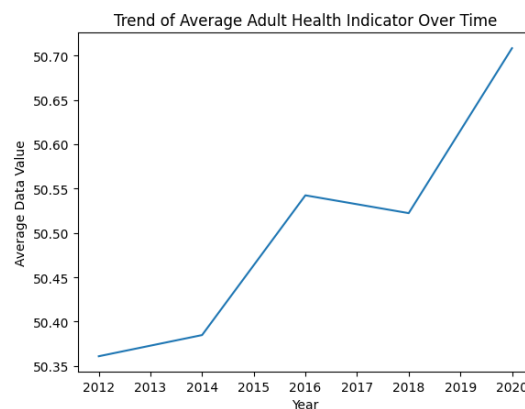
Key EDA Findings

- Multiple Indicators, One Population - All indicators focus on adult oral health, but they measure different concepts, such as dental care utilization and tooth loss.

- Wide Distribution of Values - Indicator values vary widely, indicating substantial differences across states, years, and types of oral health measures.

- Time-Based Changes- Aggregated trends show that oral health indicators change over time rather than remaining static, emphasizing the importance of longitudinal analysis.

- Geographic Disparities - State-level analysis reveals clear differences in oral health outcomes, suggesting unequal access to care and long-term health inequities.

- Missing and Uneven Data - Not all states report all indicators every year, which affects averages and reinforces the need for cautious interpretation.

- Patterns Observed – geographic, age, behavior-outcome variation and trends

    o Southern states generally report higher tooth loss and lower dental visit rates.

    o Northeastern and Western states tend to have better oral health outcomes.

    o This suggests regional socioeconomic and healthcare-access disparities.

    o Seniors (65+) show significantly higher rates of tooth loss.

    o The "lost all natural teeth" indicator has some states with values above 20–30%, highlighting severe disparities.

    o States with higher dental visit rates tend to have lower tooth loss percentages.

    o This supports the narrative: Preventive care is associated with better long-term oral health outcomes.

    o Depending on the state or indicator, slight improvements appear in some locations, but no strong national upward or downward trend is present. This suggests stable but persistent disparities over time.

- Outliers & Anomalies

    o U.S. territories (e.g., Puerto Rico) sometimes have extremely high tooth loss values.

    o Smaller jurisdictions may have wide confidence intervals, indicating less reliable estimates.

    o Some states lack data for early years or single indicators.

    o These anomalies are meaningful—they highlight where public health infrastructure differs significantly.

- Key Indicators Present

  o Adults aged 18+ who have visited a dentist or dental clinic in the past year

    → A measure of preventive dental care utilization.

  o Adults aged 65+ who have lost all natural teeth

    → A marker of severe oral health decline associated with age, access, and

    socioeconomic factors.

  o Adults aged 65+ who have lost six or more natural teeth

    → Indicates moderate oral health deterioration among older adults.

Part 3. Visualizations

Visualization 1: Trend of Average Adult Health Indicator Over Time (Line Chart)

This line chart shows how the average health indicator has changed over time, helping us
see long-term trends.

This line chart shows how the average value of adult oral health indicators changes over

time. By aggregating data annually, this visualization highlights long-term trends and

emphasizes that public health outcomes are influenced by long-term factors such as policy,

access to care, and population aging.

The x-axis represents time, measured in years. Each point along the axis corresponds to a

calendar year for which data is available. The y-axis shows the average Data_Value, which is a

percentage-based measure derived from oral health indicators. It represents the mean percentage across all available indicators and states for a given year.
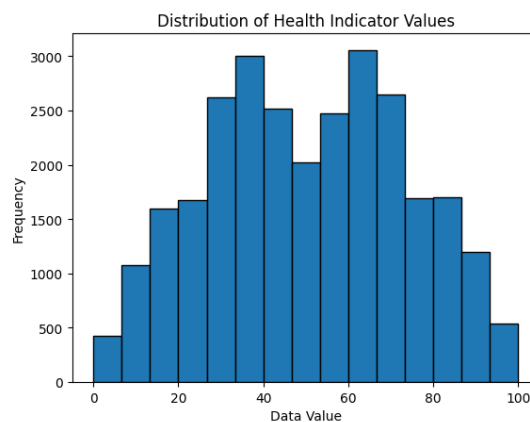
Each point on the line corresponds to one year's average, the general upward trend in the line indicates that, on average, the reported values of adult oral health indicators increase over time. Because this dataset includes both positive behaviors (e.g., dental visits) and negative outcomes (e.g., tooth loss), this upward trend does not necessarily mean oral health is getting worse or better overall—it means that the average measured values are increasing, which must be interpreted in the context of the indicators included.

The downward trend between 2016–2018 likely reflects data-related and contextual factors such as changes in indicator coverage or sampling and survey method changes, rather than a sudden nationwide improvement or decline in oral health.

I learned through this visualization that:

1. Not all dips or increases reflect real-world events

2. Context matters when interpreting time-series data

3. Aggregated data can hide underlying variation

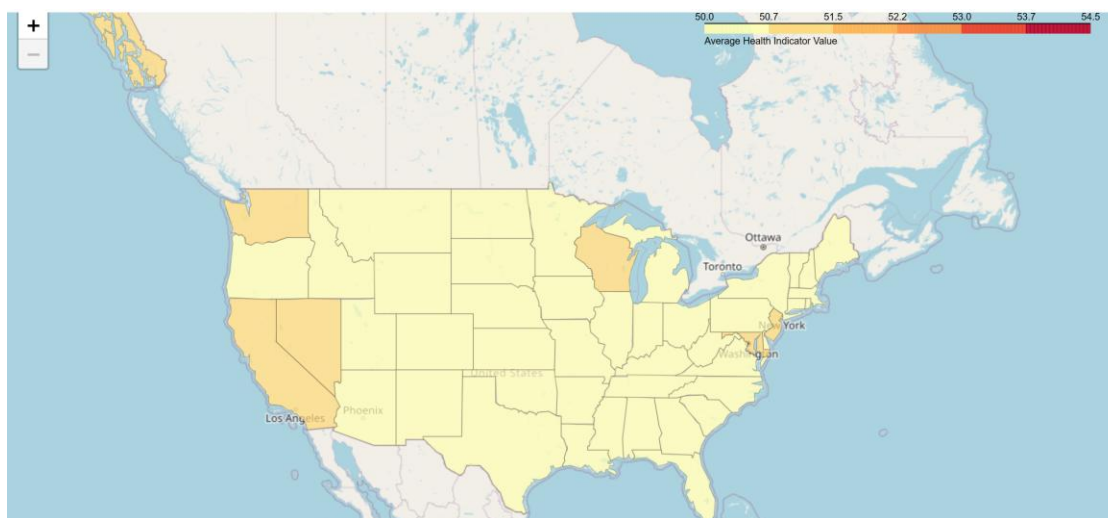Visualization 2: Distribution of Health Indicator Values (Histogram)

This improved histogram uses wider bins and clearer borders to make the distribution of health values easier to interpret and reveal patterns in the data.

This histogram displays the distribution of health indicator values across the entire

dataset, it focuses on how frequently different percentage values occur, providing descriptive

overview of all reported health indicator values, as well as insight into the overall shape, spread,

and concentration of the data.

The x-axis represents Data_Value, each value corresponds to a percentage of the

population experiencing a particular condition or behavior, whereas each bin covers a range of

values (for example, 0–7%, 7–14%, etc., depending on the data). The y-axis represents frequency

that falls within each bin.

I learned that most values cluster within certain ranges, while extreme values are less

common. This reveals variability in oral health measures and demonstrates why averages alone

cannot fully describe the data.

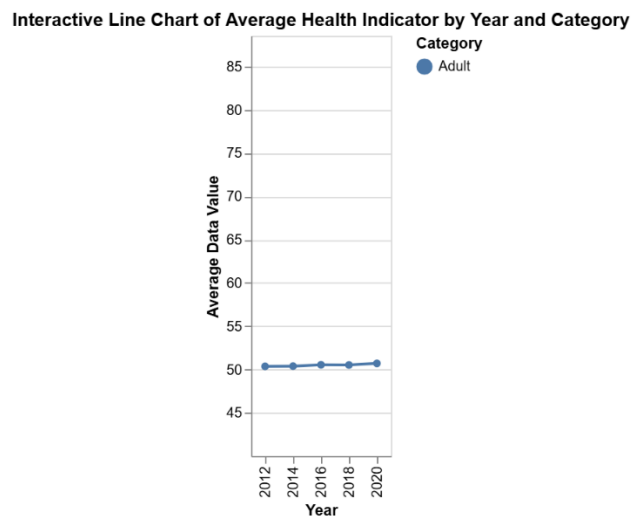Visualization 3: State-Level Tooth Loss Among Adults Aged 65+ (Choropleth Map)

This interactive map visualizes a single oral health indicator - the percentage of adults

aged 65 and older who have lost all natural teeth, aggregated at the state level. Each state is

shaded based on the average of its average health indicator value using a continuous color scale,

expressed as a percentage of the population. It shows geographic differences in complete tooth

loss among these adults. Higher average values mean a larger proportion of people in that state experience all natural teeth loss, and vice versa.

This choropleth map demonstrates that oral health outcomes among U.S. adults are geographically uneven. Certain regions show consistently higher average values, suggesting greater health challenges, while others display more favorable outcomes. The map does not imply causation but highlights where disparities exist, which is the first step toward informed policy and intervention.

I learned that whether higher value is "good" or "bad" depends entirely on the indicator. In this map I'm using a negative indicator, so, higher values indicate poorer oral health.

Visualization 4: Aggregated Adult Oral Health Trends Over Time (Interactive Line Chart)
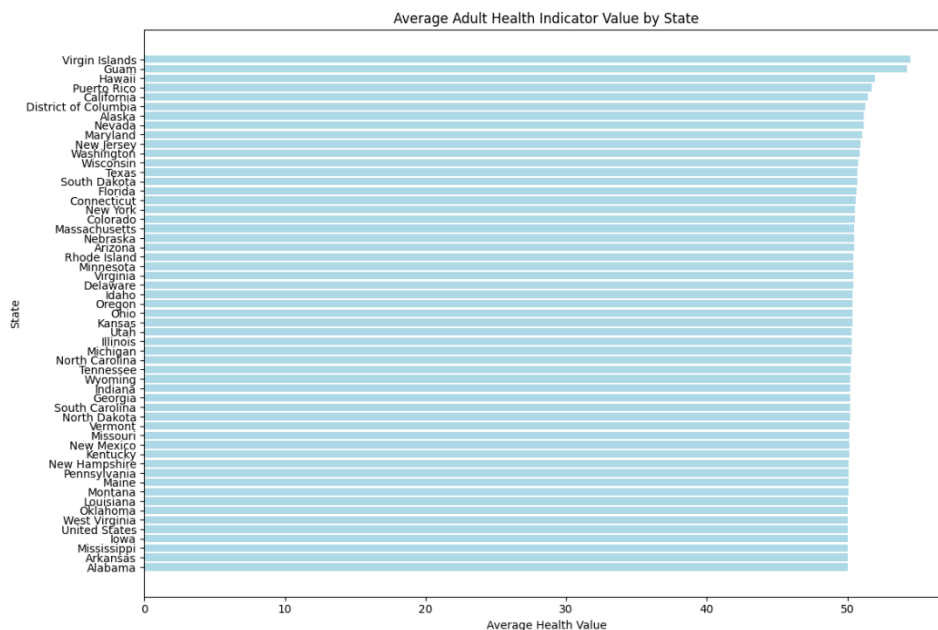

This interactive line chart aggregates all indicators of adult oral health by category into a single annual average; it presents how adult oral health indicators, when considered collectively, have changed over time. Because it uses only one category ("Adult"), all indicator values are combined, resulting in one interactive line with a data point for each year. And because the values are aggregated, the trend should be interpreted as a general summary, not a reflection of any single oral health condition.

The x-axis represents the year, while the y-axis shows the average Data_Value, expressed as a percentage. Each point corresponds to the mean of all adult oral health indicator values reported across states for that year. An upward or downward movement indicates shifts in the average prevalence of measured behaviors and outcomes. Hovering over a point reveals the exact year and average value, so we can explore precise data without overcrowding the visualization.

I realized through this visual the usefulness and limitations of aggregation in public health analysis.

Visualization 5: Average Adult Health Indicator Value by State (Horizontal Bar Chart)

This chart ranks U.S. states by their average health indicator value across all CDC Adult NOHSS measures.
States at the top have lower average values, while those at the bottom show higher health burdens.

This horizontal ranked bar chart compares states based on their average adult oral health indicator values across all measures included in the CDC NOHSS dataset. By aggregating the data at the state level, this chart offers a high-level view of overall health outcomes, allowing us to compare states on a unified scale rather than focusing on individual indicators alone.

The x-axis represents the average Data_Value, expressed as a percentage. Each value is calculated by averaging all adult oral health indicator measurements for that state across available years, the numbers reflect how common the measured behaviors or outcomes are within each state. The y-axis lists U.S. states, ordered from top to bottom based on their average health indicator values. States appearing at the bottom of the chart represent states with lower overall percentages across all indicators, which often suggests lower prevalence of severe oral health problems, whereas states listed toward the top show higher combined percentages, often reflects higher oral health burden, especially from tooth loss indicators. However, because indicators are mixed, this ranking reflects overall oral health signal, not a pure "good vs bad" score.

By sorting states from lowest to highest average values, the chart reveals geographic disparities and addresses a key public health question: Which states are experiencing higher overall health burdens, and which are performing comparatively better? As well as highlighting where health interventions, policy resources, or targeted programs may be most needed.

Part 4. Conclusion and Next Steps

This project has taught me how exploratory data analysis and visualization can be used to uncover meaningful patterns in public health data. The findings show that adult oral health in the U.S. varies significantly across time and geography, with long-term outcomes reflecting historical inequalities in access to dental care.

Key Takeaways

- Oral health indicators vary widely and must be interpreted with context

- Geographic disparities persist, especially among older adults

- Aggregated visuals are powerful but require cautious explanation

Possible Next Steps

- Analyze indicators separately rather than in aggregate

- Explore socioeconomic factors linked to oral health outcomes

- Compare adult oral health trends with healthcare access data

- Perform state-level trend analysis over time

Part 5. Glossary of Key Terms in the Dataset

Term	Definition
Indicator	A measurable oral health condition or behavior (e.g., dental visits, tooth loss).
Data_Value	The primary percentage estimate reported for a state and year.
LocationDesc	Full state or jurisdiction name (e.g., "California").
LocationAbbr	Two-letter state abbreviation used for mapping (e.g., CA, NY).
Year	The year the measurement was collected or estimated.
Sample_Size	Number of survey respondents included in the estimate; affects reliability.
Low_Confidence_Limit / High_Confidence_Limit	Statistical interval showing uncertainty around the Data_Value.
DataSource	Indicates whether data comes from BRFSS or other public health systems.

Term	Definition
Topic / Category	Broad classification of the indicator. For this dataset, the topic is Oral Health.
Stratification	Indicates demographic grouping (not heavily used in this dataset).
State FIPS Code	Numeric code for each state (used for geographic merging in advanced analysis).
Adults Who Visited a Dentist or Dental Clinic in the Past Year	Percentage of adults who reported at least one dental visit in the previous 12 months. Represents preventive care behavior.
Adults with Six or More Teeth Removed	Percentage of adults who have lost six or more permanent teeth due to decay or gum disease. Represents partial tooth loss and long-term oral health outcomes.
Adults Aged 65+ Who Have Lost All Natural Teeth	Percentage of adults aged 65 or older who have lost all natural teeth due to oral disease. Represents severe, cumulative oral health outcomes.
Data_Value	The reported percentage of adults meeting the criteria of a specific oral health indicator.
Category (Adult)	A grouping label indicating that the indicator applies to the adult population.

References

Centers for Disease Control and Prevention. (n.d.). National Oral Health Surveillance System

(NOHSS): Adult indicators. HealthData.gov.

https://healthdata.gov/CDC/NOHSS-Adult-Indicators/7het-pevn/about_data

OpenAI. (2025). ChatGPT (GPT-5.2) [Large language model]. https://chat.openai.com/

I used ChatGPT to assist with improving Python visualization. All coding was reviewed and validated by me.