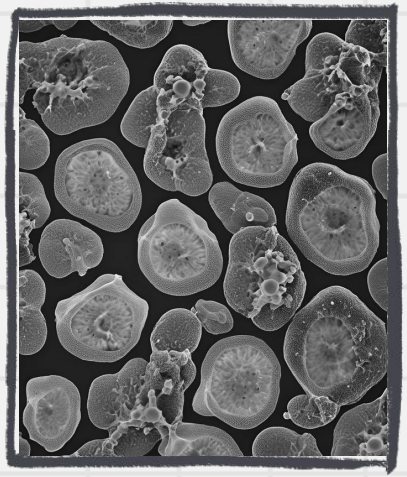
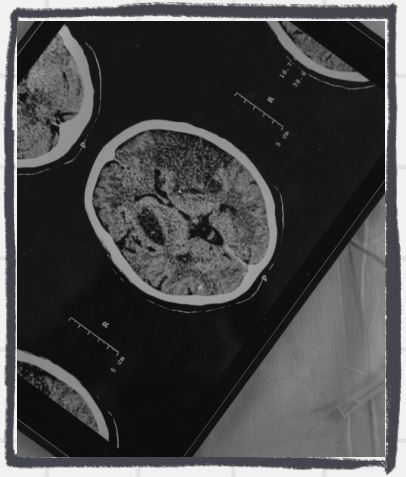
Project 3 Team 11

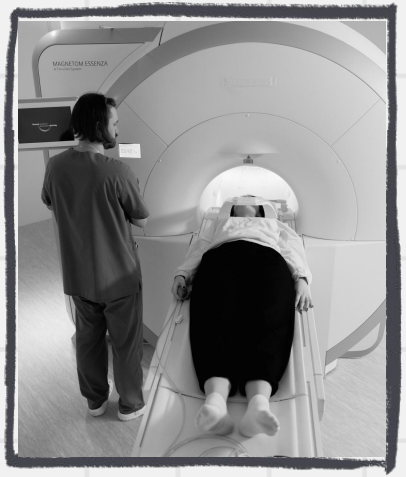
Avery Javier
Daena Wallace
Victoria Mendez

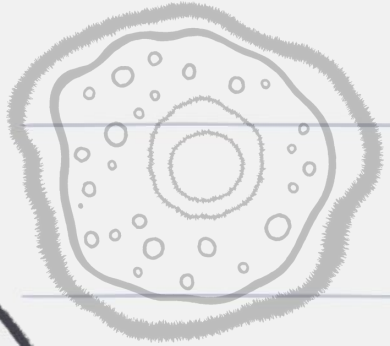
"Healthy Living, Diverse Challenges: Exploring Health Trends by Topic and Region"

Table of Contents

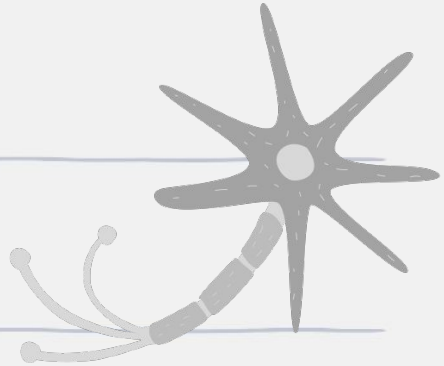








01



Introduction

Purpose of the Project

Explore trends in chronic diseases by analyzing connections between health issues, demographic factors, and lifestyle risks.

Provide actionable insights through interactive data visualizations.

Research Questions

How do health issues and their key measures, including risk factors, compare across topics?

How does smoking status correlate with the prevalence of COPD?

How do Cardiovascular Disease rates differ across gender, age?

Tech Stack

- Backend: Python, SQLite
- Frontend: Dash, Plotly, HTML, Leaflet.js
- Database: SQLite
- Data Source: CDC Chronic Indicators

Implementation

- Dataset preprocessed stored in SQLite
- Dash apps retrieves and filters data directly from the database.

Data Cleaning Summary

[Data Source: CDC Chronic Disease Indicators.]

Loaded Dataset:

Imported a dataset with 34 columns containing health-related indicators across various dimensions.

Dropped Irrelevant Columns:

Removed columns with zero non-null values and columns deemed unnecessary for analysis. This streamlined the dataset by focusing on relevant fields.

Renamed Columns:

Renamed columns for consistency and clarity (e.g., `YearEnd` to `Year`, `LocationDesc` to `Location`, etc.).

Extracted Geolocation:

Parsed `Geolocation` column to separate `Latitude` and `Longitude` values, enabling spatial analysis.

Handled Stratification

Pivoted stratification-related categories (e.g., `Sex`, `Age`, `Race/Ethnicity`) into separate columns for better granularity.

Replaced missing values in these categories with `"Unknown"` to avoid data gaps during analysis.

Removed Duplicates:

Dropped duplicate rows to ensure uniqueness in the dataset.

Dropped Non-Essential Columns:

Further removed columns such as `Overall`, `DataValueAlt`, and others that were redundant or did not contribute to the analysis.

Reorganized Columns:

Rearranged columns to prioritize key dimensions (e.g., `Year`, `Sex`, `Age`, `Topic`, `Value`, etc.), creating a structured dataset ready for visualization and analysis.

02

Data Visualizations and Insights

Key Trends and Insights

Chronic Diseases and Contributing Factors:

Cardiovascular Disease, Chronic Obstructive Pulmonary Disease, Obesity, and Diabetes emerge as significant health burdens, often linked to physical inactivity and poor nutrition.

Mental Health Issues, including depression and frequent distress, are prevalent and interrelated with chronic conditions.

Lifestyle Risks and Preventive Needs:

High rates of **smoking and alcohol consumption** correlate with increased risks of respiratory, cardiovascular, and chronic diseases.

Poor **nutrition and physical inactivity** exacerbate obesity, arthritis, and diabetes prevalence.

Social and Systemic Determinants:

Access to healthcare and preventive services plays a critical role in improving outcomes, particularly in maternal and child health.

Nutrition and Weight Management:

Obesity and low physical activity rates emerge as frequent concerns, with strong correlations to other chronic diseases like diabetes.

Major Correlations

Lifestyle Factors and Chronic Conditions:

Physical Inactivity correlates strongly with **arthritis** and **obesity**, highlighting the need for increased physical activity in prevention efforts.

Smoking rates show a strong positive correlation with **respiratory diseases** (e.g., chronic obstructive pulmonary disease) and **cardiovascular conditions**.

Mental Health and Chronic Illnesses:

Frequent mental distress and **depression** are closely linked to conditions like **obesity**, **diabetes**, and **cardiovascular disease**, emphasizing the interplay between mental and physical health.

Tobacco and Alcohol Impact:

Tobacco use is highly correlated with **lung disease** and **heart disease**, while **alcohol consumption** shows links to **mental health issues** and certain chronic conditions like **liver disease**.

Preventive Services:

Access to preventive healthcare (e.g., immunizations, check-ups) shows moderate correlations with better health outcomes, particularly in managing chronic conditions.

Health Measures Dashboard

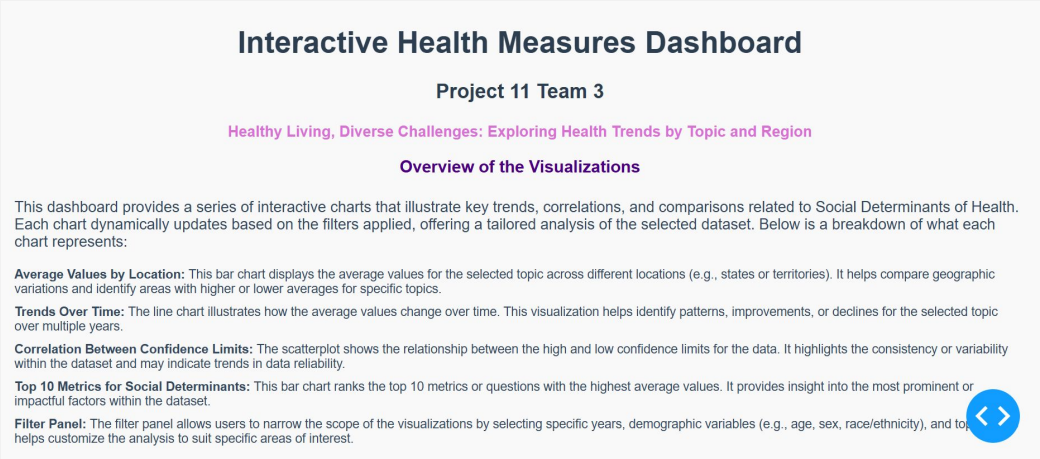
How do health issues and their key measures, including risk factors, compare across topics?

Purpose:

- To explore health data interactively.
- Enable users to identify patterns and insights related to chronic diseases, demographics, and location.

Methods:

- Data stored in SQLite.
- Dashboard built with Dash and Plotly for interactivity.



Interactive Health Measures Dashboard

Project 11 Team 3

Healthy Living, Diverse Challenges: Exploring Health Trends by Topic and Region

Overview of the Visualizations

This dashboard provides a series of interactive charts that illustrate key trends, correlations, and comparisons related to Social Determinants of Health. Each chart dynamically updates based on the filters applied, offering a tailored analysis of the selected dataset. Below is a breakdown of what each chart represents:

Average Values by Location: This bar chart displays the average values for the selected topic across different locations (e.g., states or territories). It helps compare geographic variations and identify areas with higher or lower averages for specific topics.

Trends Over Time: The line chart illustrates how the average values change over time. This visualization helps identify patterns, improvements, or declines for the selected topic over multiple years.

Correlation Between Confidence Limits: The scatterplot shows the relationship between the high and low confidence limits for the data. It highlights the consistency or variability within the dataset and may indicate trends in data reliability.

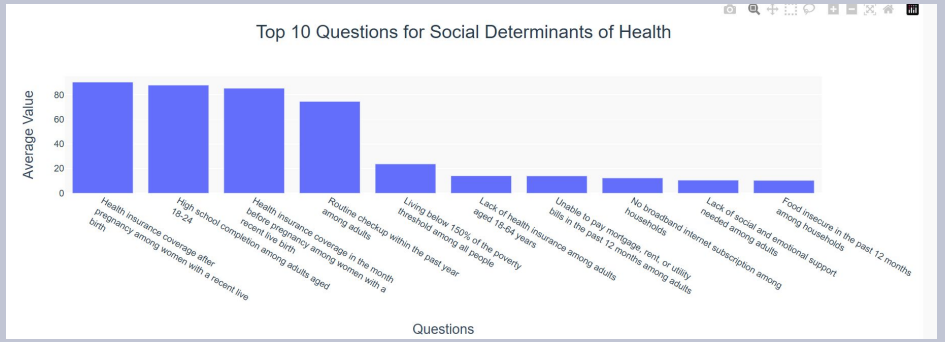
Top 10 Metrics for Social Determinants: This bar chart ranks the top 10 metrics or questions with the highest average values. It provides insight into the most prominent or impactful factors within the dataset.

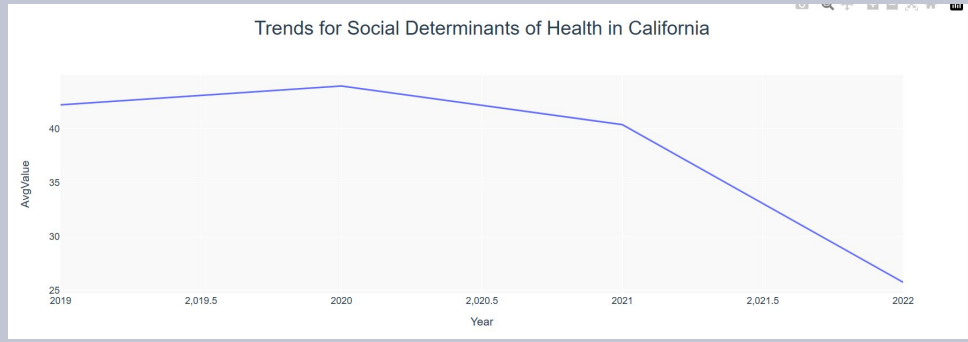
Filter Panel: The filter panel allows users to narrow the scope of the visualizations by selecting specific years, demographic variables (e.g., age, sex, race/ethnicity), and to
helps customize the analysis to suit specific areas of interest.

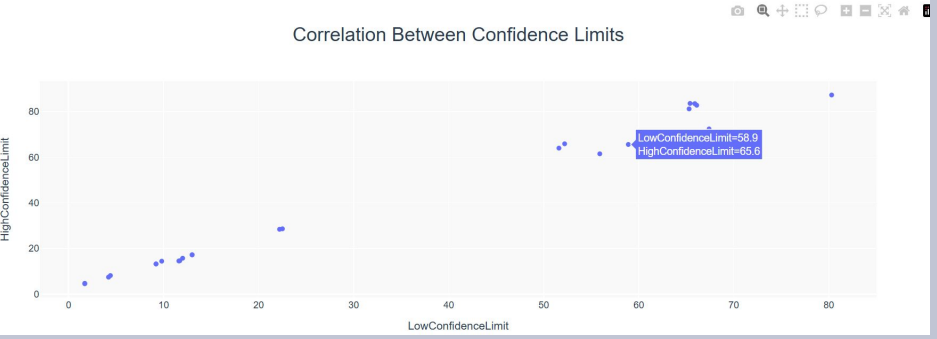
Mixed Value Types in the Dataset

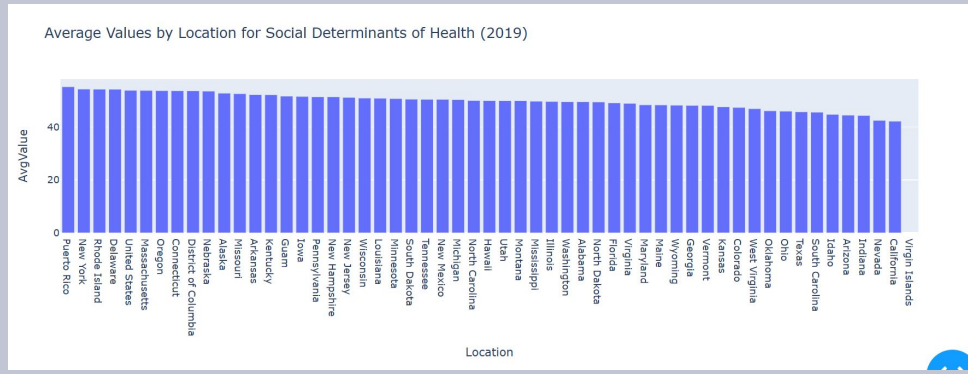
The dataset contains a mix of value types, such as percentages, counts, and rates, depending on the selected topic and metric. The visualizations are designed to dynamically adjust based on the data selected, but users should be mindful that comparisons between different metrics or topics may not always be directly meaningful due to these differences in value types. When in doubt, focus on trends within the same topic or consult the chart descriptions for guidance.

Health Measures Dashboard Visual Samples









Health Measures Dashboard: Features

This dashboard provides a series of interactive charts that illustrate key trends, correlations, and comparisons related to Social Determinants of Health. Each chart dynamically updates based on the filters applied, offering a tailored analysis of the selected dataset. Below is a breakdown of what each chart represents:

Average Values by Location:

This bar chart displays the average values for the selected topic across different locations (e.g., states or territories). It helps compare geographic variations and identify areas with higher or lower averages for specific social determinants.

Trends Over Time:

The line chart illustrates how the average values change over time. This visualization helps identify patterns, improvements, or declines for the selected topic over multiple years.

Correlation Between Confidence Limits:

The scatterplot shows the relationship between the high and low confidence limits for the data. It highlights the consistency or variability within the dataset and may indicate trends in data reliability.

Top 10 Metrics::

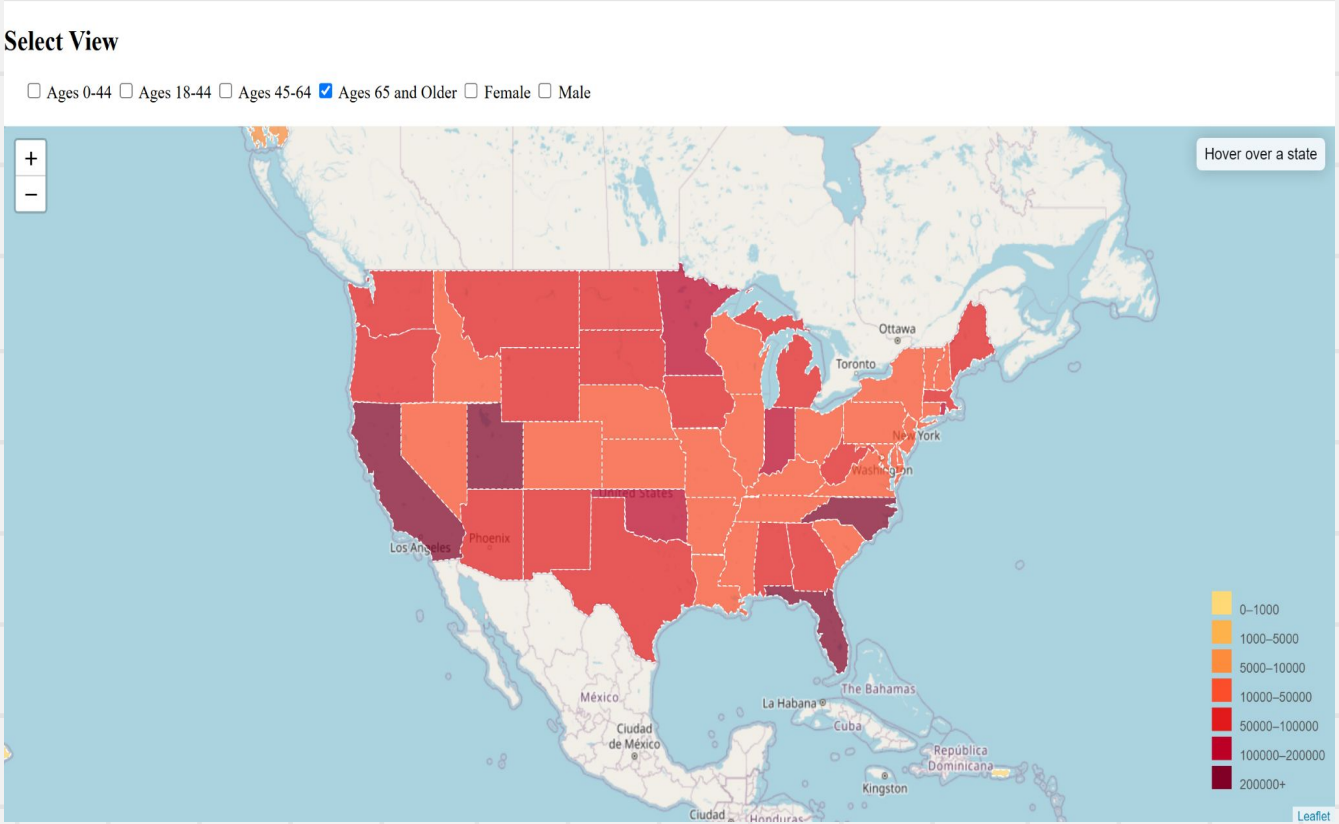
This bar chart ranks the top 10 metrics or questions with the highest average values. It provides insight into the most prominent or impactful factors within the dataset.

Filter Panel:

The filter panel allows users to narrow the scope of the visualizations by selecting specific years, demographic variables (e.g., age, sex, race/ethnicity), and topics. This helps customize the analysis to suit specific areas of interest.

Cardiovascular Disease rates across age groups and genders

- *Cardiovascular Disease prevalence is consistently under 500 for majority of the U.S in age group 18-44, signaling a condition that stems most prevalently in early development and after the age of 44.*
- *Cardiovascular Disease prevalence 5-10% higher in males than females*
- *Age group with the highest cardiovascular disease indications is over the age of 65*



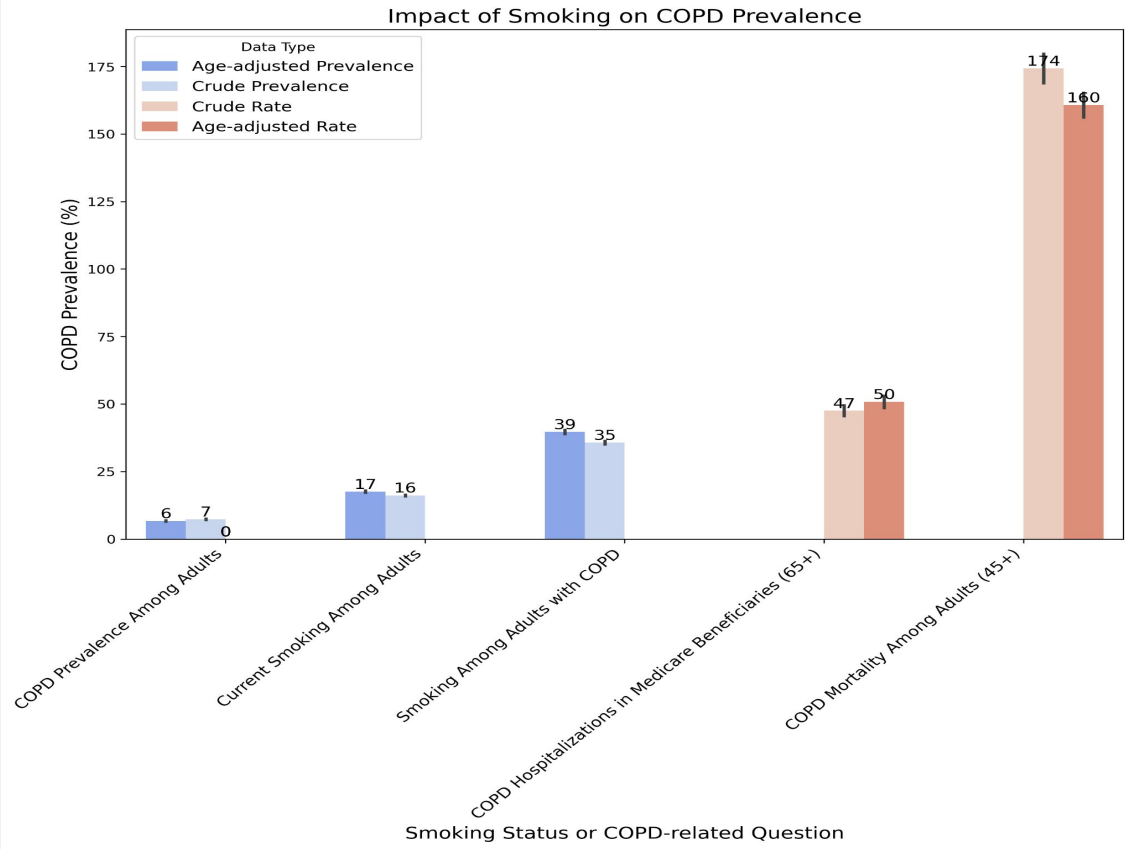
How does smoking affect the risk of developing chronic obstructive pulmonary disease (COPD)?

Overview:

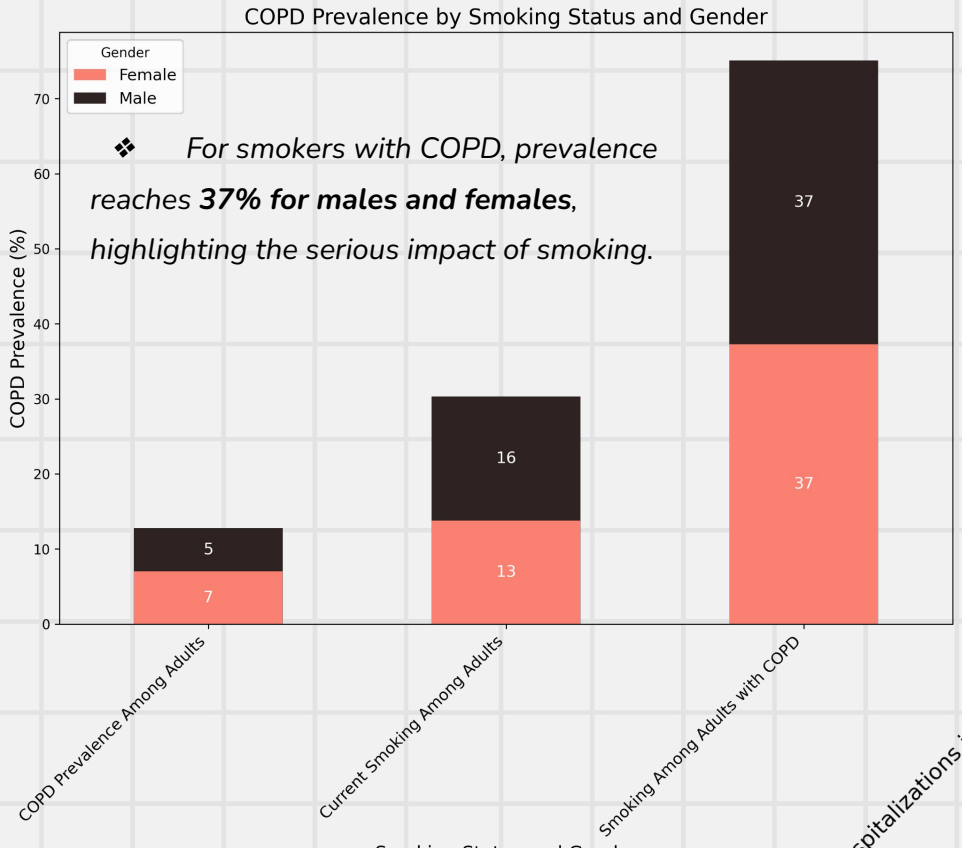
- COPD is a chronic lung condition that causes breathing difficulties.
- Smoking is a primary risk factor for COPD and worsens its progression.
- This analysis examines the impact of smoking on COPD prevalence across different groups.

Key observations:

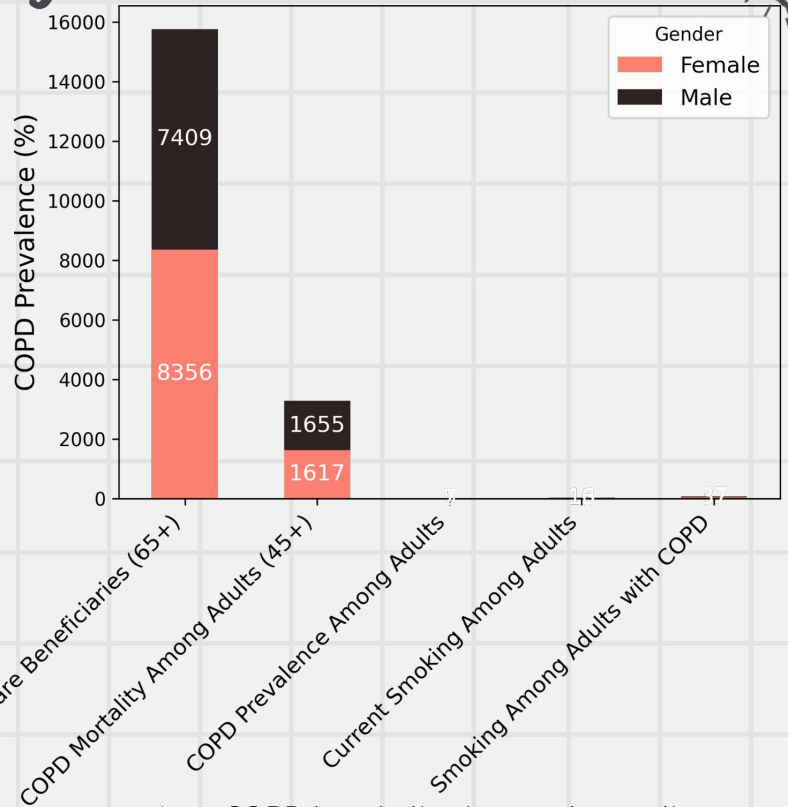
- Higher **hospitalization** and **mortality rates** for individuals aged 65+ indicate a significant impact of COPD in older populations.
- Smoking increases COPD prevalence significantly, as shown in the **"Smoking Among Adults with COPD"** bar.



COPD Prevalence by Smoking Status and Gender



❖ *For smokers with COPD, prevalence reaches **37% for males and females**, highlighting the serious impact of smoking.*



❖ *COPD hospitalizations and mortality are higher for **female smokers** compared to males*

Ethical Considerations in Data Analysis

Data Privacy & Protection

- No **personally identifiable information (PII)** included; data is **aggregated & anonymized** (CDC Chronic Disease Indicators).

Accuracy & Avoiding Misrepresentation

- **Standardized data** to ensure consistency across **age, gender, and location**.
- Visualizations **reflect trends without exaggeration** or misleading comparisons.
- Included **confidence intervals** to highlight **data reliability** and avoid false assumptions.

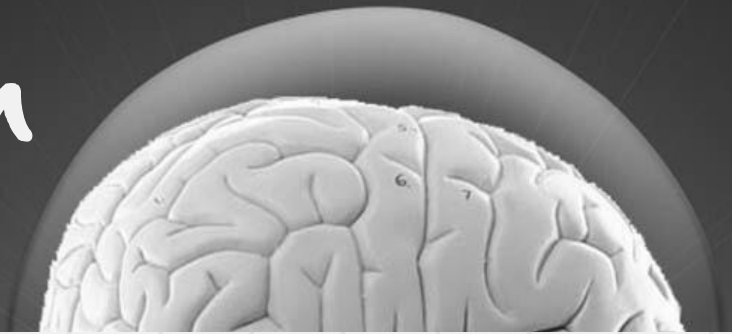
Transparency & Reproducibility

- **Fully documented sources and processes** for data cleaning & visualization.
- Data stored in **SQLite for structured access**, ensuring traceability.
- Jupyter Notebooks outline **step-by-step data processing** for replication. Go to ReadMe

Fair Representation & Limitations

- Highlights **disparities in health outcomes** across demographics.
- Acknowledges **data gaps** and avoids making **causal claims** without evidence.
- Encourages **contextual interpretation** of trends, considering **social determinants**.

Conclusion



Key Insights:

- **Health issues exhibit significant variation across topics**, with clear connections between lifestyle risks and chronic diseases like obesity, diabetes, and cardiovascular disease.
- Smoking is a primary driver of **COPD** risk and progression, disproportionately impacting older adults and women.
- **Cardiovascular disease** rates vary across gender and age, highlighting the importance of targeted interventions for at-risk groups.

Takeaways:

- Chronic diseases are interconnected with social determinants and lifestyle factors.
- Preventive measures and public health campaigns can significantly reduce risks and improve outcomes.

THANK YOU

References:

Data Source: The dataset was sourced from the CDC Chronic Disease Indicators.

Credits

This presentation template is free for everyone to use thanks to the following:

Slides Carnival

for the presentation template

PEXELS & PIXABAY

for the photos