P&I Deaths vs. All Deaths: State & Regional Analysis Using NCHS MSS Data

Data Detectives, a group of students from the HDS_5310_04 - Analytics and Statistical Programming-04 course under the guidance of Dr. Paul Boal, present a comprehensive analysis of Pneumonia and Influenza (P&I) mortality rates in relation to overall death rates across different states and regions of the United States.

Introduction

1

Objective

To explore the association between Pneumonia and Influenza (P&I) mortality rates and overall death rates in different states and regions of the US.

2

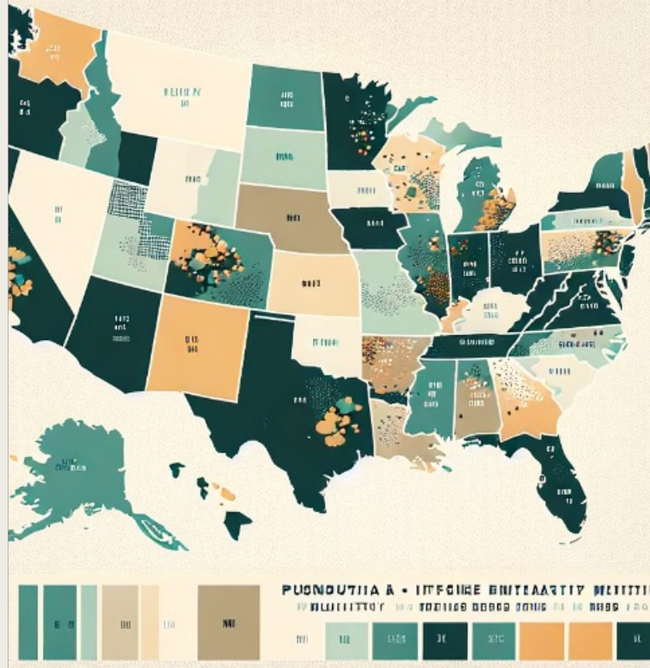
Significance

Understanding these patterns is crucial for effective health prioritization, healthcare resource allocation, and disease surveillance efforts.

3

Importance of State and Regional Analysis

Analyzing differences across states and regions helps identify specific areas with higher risks, guiding targeted interventions and resource distribution. It emphasizes the necessity of regionalized healthcare strategies to effectively manage P&I impacts.



Background and Importance of Research

Why This Research is Crucial

Public Health Impact

P&I are significant contributors to mortality rates; understanding their impact relative to total deaths is essential for evidence-based decision-making in public health and healthcare management.

Policy Making

Insights from this study can influence public health policies by highlighting areas needing urgent healthcare interventions and resource allocation

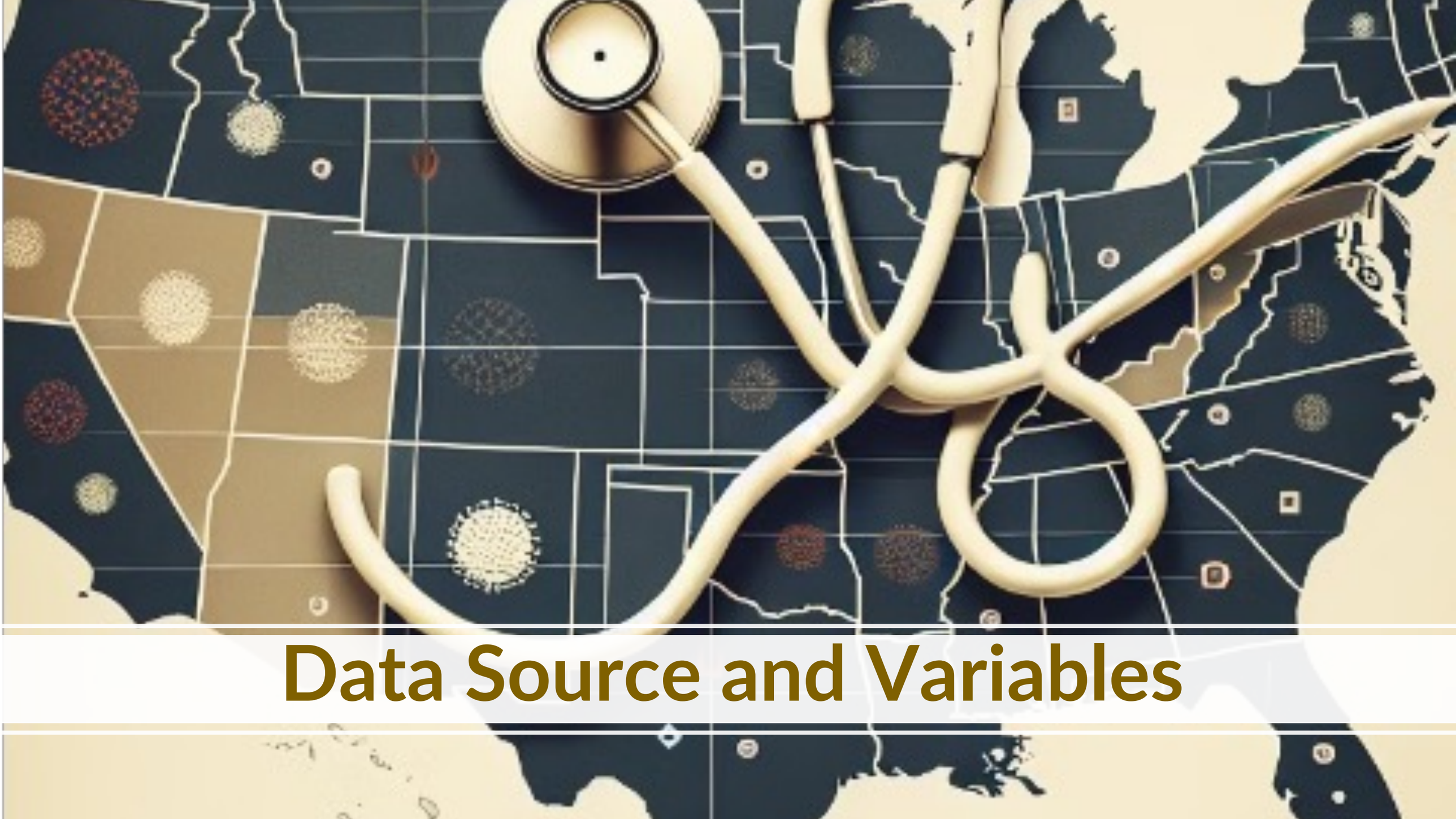
Review of existing studies

Literature Gap

Previous research has primarily focused on isolated studies of P&I mortality without a comparative analysis at state and regional levels.

Study's Contribution

This research seeks to fill existing knowledge gaps by providing a comparative analysis of P&I mortality against overall mortality rates, offering new insights into death trends and disparities within the United States.



Data Source and Variables

Data Source and Variables

1 NCHS Mortality Surveillance System (MSS)

Description

The primary data source for this research is the Mortality Surveillance System provided by the National Center for Health Statistics (NCHS).

Significance

This system offers comprehensive state and regional level mortality data, essential for analyzing P&I deaths across the U.S

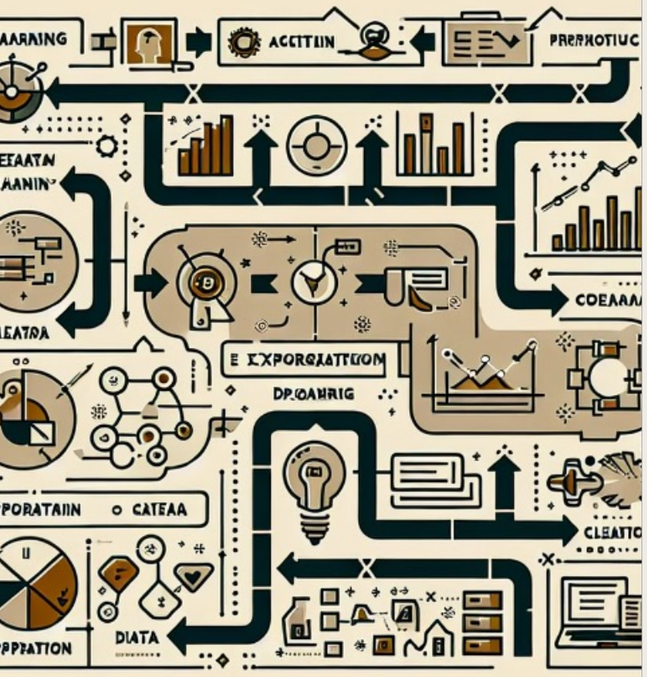
2 Variables in the Study

Dependent Variable

The death rates from Pneumonia and Influenza (P&I), expressed as a percentage of total mortality across different states and regions in the U.S.

Independent Variables

States, Regions, Age-groups, Seasons, and MMWR Years/week, focusing on identifying vulnerable groups and customizing solutions per region and time period.



Methodology - Descriptive Analysis

1

Understanding the Dataset

Initial Examination: Detailed examination of unique values across various columns like 'geoid', 'State', 'age', and 'season' to understand data composition.

Significance: This step is crucial for recognizing substantial diversity in demographic and geographical identifiers, informing further analyses.

2

Data Preparation Techniques

Mode Calculation: A function is defined and applied to calculate the mode for each column to understand the most common occurrences, revealing dominant trends like the most frequently reported state or age category.

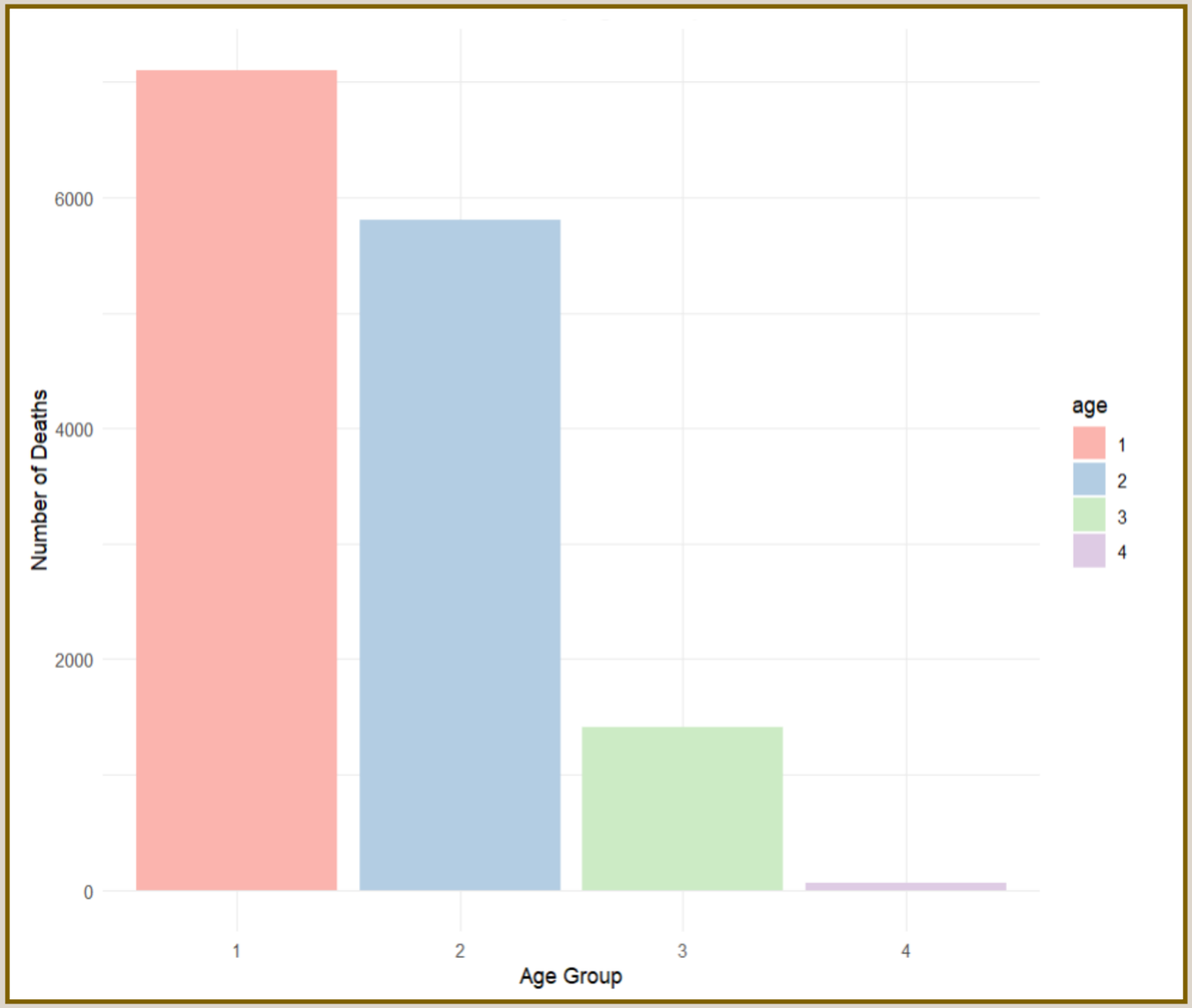
Handling Qualitative Variables: Conversion of qualitative variables into numerical codes to simplify the data structure and enable numerical statistical analysis, thus preventing dataset distortion or noise.

3

Example Insight

Alabama emerges as the most common state in the analysis, providing insights into potential geographical skew or data collection biases.

Deaths from Pneumonia and Influenza by Age Group



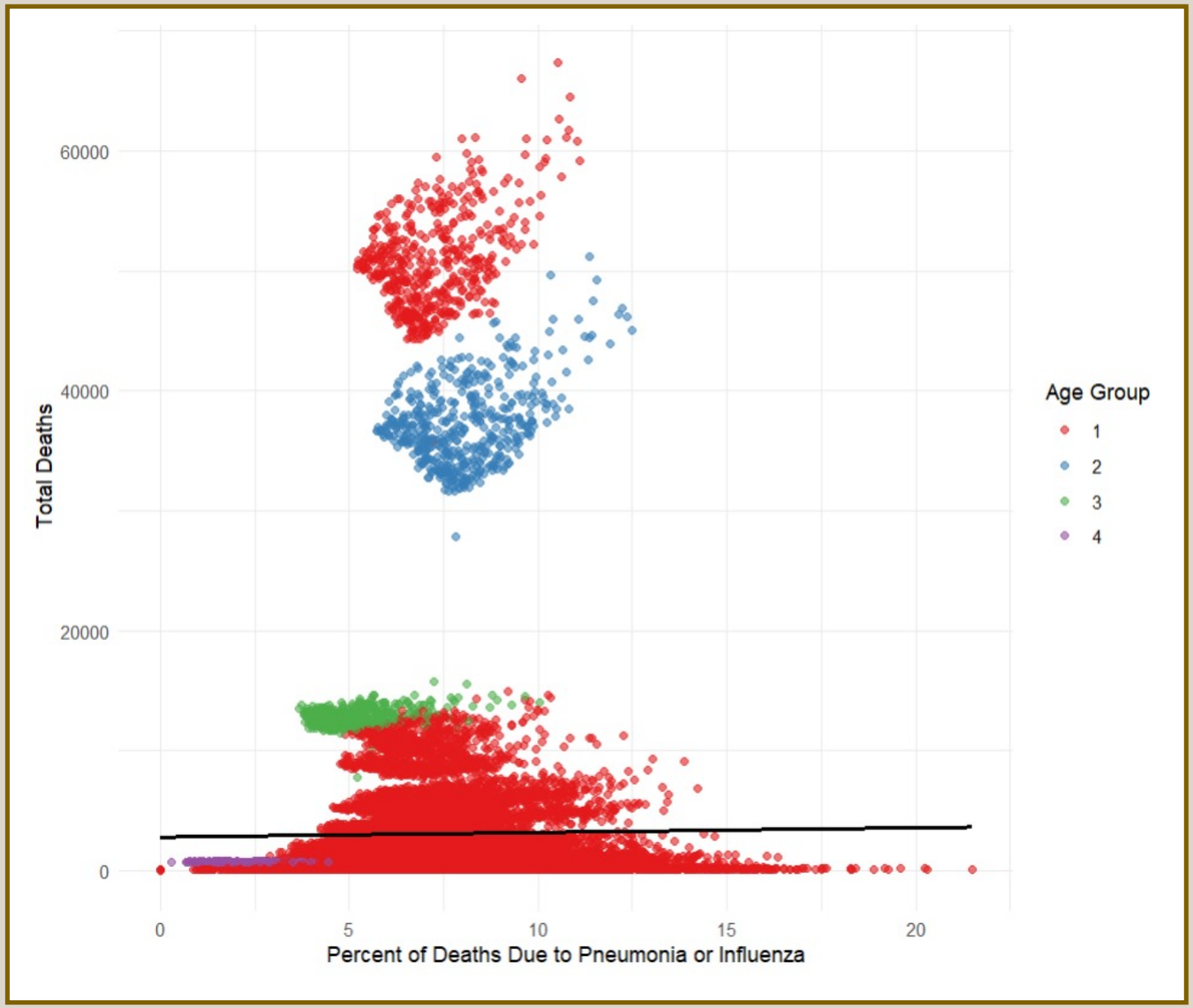
Description

A bar chart showing the number of deaths due to Pneumonia and Influenza across different age groups.

Insight

The youngest age group exhibits the highest number of deaths, possibly indicating increased susceptibility or exposure among this group.

Relationship between P&I Death Percentage and All Deaths by Age



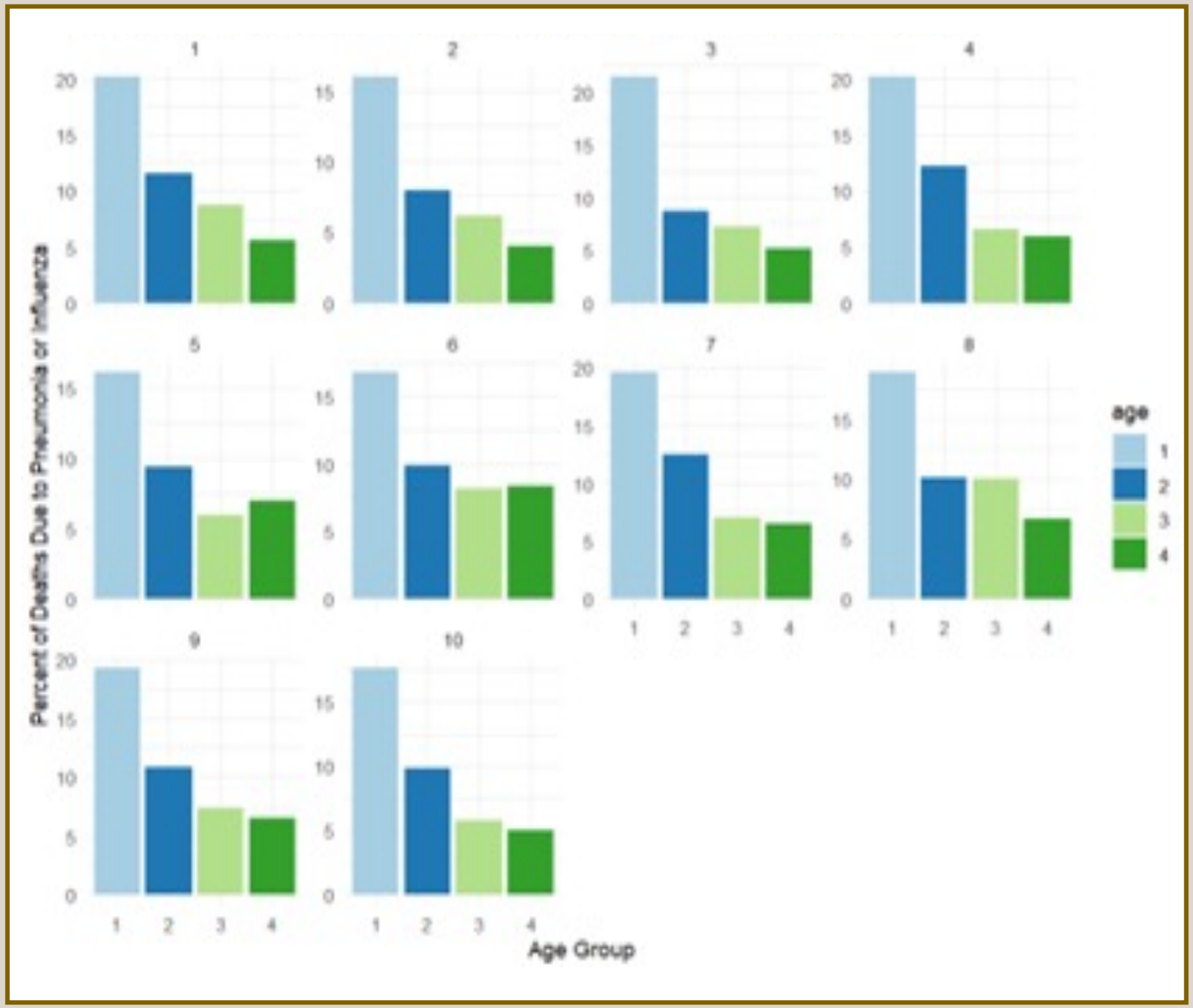
Description

A scatter plot illustrating how the percentage of deaths due to P&I correlates with total deaths across various age groups.

Analysis

Notable variance in younger age groups suggests potential disparities in healthcare access or differential exposure rates.

Percent of Deaths due to Pneumonia or Influenza by Age across Seasons



Description

A grouped bar chart displaying the percentage of deaths due to P&I across different age groups, segmented by season.

Observation

Some seasons show higher percentages, likely correlating with typical flu seasons, highlighting the effectiveness of seasonal healthcare responses.

Statistical Analysis and Findings

Statistical Analysis and Findings

Pearson's Correlation Test

Result
The correlation between P&I deaths and all deaths is very close to 1 (r = 0.9799788).

Interpretation
Indicates a strong positive linear relationship, suggesting that increases in P&I deaths are almost perfectly associated with increases in overall deaths.

Statistical Significance
The p-value (<0.001) strongly supports the predictability of total deaths when P&I deaths are considered.

ANOVA (Analysis of Variance)

Result
Significant effect of state on P&I deaths (F = 195.4, p < 2e-16).

Implication
Demonstrates that the average number of P&I deaths varies significantly by state, suggesting regional disparities possibly due to differences in healthcare access, population density, or state-specific health policies.

Kruskal-Wallis Test

Result
Chi-squared = 29744, df = 51, p-value < 2.2e-16.

Conclusion
Significant differences in the distribution of all deaths across states imply substantial state-based differences in mortality.

Post Hoc Analysis and Diagnostic Checks

TUKEY'S HONEST SIGNIFICANT DIFFERENCE (HSD) TEST

Result
Significant differences in P&I deaths among states identified.

Example
Significant differences in mean deaths between states such as Alaska vs. Alabama.

Implications
These state comparisons highlight the robust and significant regional disparities in P&I mortality, emphasizing the need for state-specific public health strategies.

Post Hoc Analysis and Diagnostic Checks

DIAGNOSTIC CHECKS ON MODEL'S RESIDUALS

SHAPIRO-WILK TEST

Result
W = 0.40255, p-value < 2.2e-16.

Interpretation
Indicates a significant departure from normality, suggesting the residuals do not follow a normal distribution, which is critical as ANOVA assumes normally distributed errors.

Diagnostic Plots

Overview
Include plots of residuals against fitted values and a normal Q-Q plot.

Purpose
Helps detect non-constant variance or outliers, providing visual evidence for the assumptions underlying the ANOVA.

Effect Size Measurement

Welch Two Sample t-test

Objective
Compare the means of P&I death percentages between High-Risk and Low-Risk groups.

Results
Highly significant t-test results indicate a substantial difference in P&I death percentages.

Statistics
|t-value = 207.81, df = 27113, p-value < 0.0001.

Cohen's d for Effect Size

Result
Cohen's d = 2.365715.

Interpretation
This large effect size underscores a marked difference between High-Risk and Low-Risk groups, indicating significant variance in P&I mortality risk.

Practical Implication
The effect size suggests a pronounced disparity that could be targeted through tailored public health interventions in high-risk areas.

Linear Regression Model and Diagnostics

Overview of Linear Regression Model

Description
The model predicts the percentage of deaths due to pneumonia and influenza (P&I) using various predictors from a mortality dataset, split into 80% training and 20% testing sets for validation.

Features Used
The model includes state, region, age groups, seasons, and MMWR weeks as predictors to account for demographic and temporal variations.

Model Diagnostics

Durbin-Watson Test

Result
DW statistic of 1.9832, p-value = 0.0866.

Interpretation
Indicates minimal autocorrelation among residuals, supporting the assumption of independence in the linear regression model.

Breusch-Pagan Test

Result
BP = 7376.9, df = 71, p-value < 2.2e-16.

Interpretation
Strong rejection of homoscedasticity, indicating significant heteroscedasticity, suggesting that the variance of residuals varies with the level of explanatory variables, which might impact the reliability of regression coefficients.

Results, Discussions, and Conclusions

1 Summary of Key Findings

Strong Correlation

The analysis demonstrates a very strong positive correlation (r = 0.9799788) between P&I deaths and overall mortality rates, indicating that P&I deaths are almost perfectly predictive of total mortality variations.

Statistical Significance

This correlation is statistically significant with a p-value < 0.001, underscoring the robustness of the findings.

2 Discussion on Regional Disparities

Variability Across States

Significant disparities in P&I mortality rates across states as evidenced by ANOVA and Tukey's HSD tests, highlighting the impact of regional factors such as healthcare access and population density.

3 Implications for Public Health Policy

The findings advocate for targeted public health measures in regions with higher P&I mortality.

Suggest the need for tailored healthcare strategies and resource allocation to address the identified disparities effectively.

4 Conclusions!

Public Health Relevance

The strong association between P&I deaths and total mortality reinforces the need for enhanced surveillance and intervention strategies during peak seasons or outbreaks.

Future Policy Guidance

These results should guide future policy and healthcare planning to mitigate mortality rates associated with P&I, particularly in high-risk areas.


Thank
you!