Regional Disparities in U.S. Traffic Accidents: A State-Level Analysis of Incidence Rates

SAS Masters: Urmi Phanse, Ojasvi Vachharajani, Boston University

ABSTRACT

Traffic accidents remain a significant public health concern, with 42,514 fatalities recorded in the U.S. in 2022. This study examines regional and state-level disparities in accident incidence rates using the US Accidents dataset (2016–2023) and population data from the World Population Review. Analysis across 37 states and 500 ZIP codes reveals that California, Florida, Virginia, and Arizona report the highest accident rates, with the South and West regions exhibiting significantly higher rates than the Northeast and Midwest. Statistical modeling confirms substantial geographic variations in accident risks. Findings underscore the need for targeted safety measures, infrastructure improvements, and policy interventions to mitigate traffic fatalities. Future research should incorporate longitudinal data, real-time weather conditions, and additional risk factors to enhance traffic safety strategies.

INTRODUCTION

Traffic-related injuries cause the most deaths among children and young adults aged 5 to 29 years. Traffic accidents kill an estimated 1.19 million people worldwide each year.[1] As of the third quarter of 2021, there were around 284 million automobiles on the road in the United States, making it one of the busiest nations in terms of traffic volume.[2] About 42,514 car crashes resulted in fatalities on American roads in 2022.[3] Road traffic accidents can occur in a variety of ways and for a variety of reasons. Some are under control, such as speeding, tailgating, inappropriate lane changes, fatigued driving, and dangerous behavior. Others are beyond human control, such as mechanical difficulties, bad weather conditions, missing road signs, or poor road conditions.[4,5] It is a serious public health concern as it costs millions of losses in the economy and people's lives. This growing public health crisis results in significant economic losses and thousands of preventable deaths, yet it often goes unrecognized despite its far-reaching consequences.

PROBLEM STATEMENT

Road accidents have resulted in a significant number of fatalities in the United States, including 42,514 in 2022, and traffic-related injuries remain a leading cause of death worldwide. While multiple factors contribute to these accidents, understanding regional and state-level disparities is crucial for identifying the most affected areas and implementing targeted interventions. This study aims to establish a foundation for future research by highlighting the states and regions most impacted, enabling further investigations into contributing factors such as weather conditions, preparedness, consumer behavior, infrastructure, and policies.

DATA PREPARATION

Data sets

The analysis utilizes two datasets: the US Accidents dataset (2016–2023) (ref link) and the population dataset (ref link). The US Accidents dataset provides detailed information on over 7.7 million accidents across the United States, containing 46 variables. The population dataset, sourced from the World Population Review, includes ZIP code-specific demographic details, initially spanning 168,009 records.

Data cleaning and processing

To focus on 2022 data, the accident dataset was filtered using the year variable extracted from the Start_Time field, reducing it to 1,762,452 observations. ZIP codes were standardized to a consistent 5-digit format, and state names in the population dataset were cleaned and converted into standardized state codes for compatibility. We merged the datasets by state code and ZIP code, ensuring each observation represents a unique ZIP code with accident count and population data. This merge limited the final dataset to 500 unique ZIP codes across 37 states, as only ZIP codes present in both datasets were retained.

Accident incidence rates per 100,000 population were calculated at three levels:

- ZIP code level: (accident_count / population) × 100,000
- State level: (total_accidents_state / total_population_state) × 100,000
- National level: (total_accidents_country / total_population_country) × 100,000

The final dataset (cup.rates) includes formatted variables with appropriate labels, enabling comprehensive examination of traffic accident patterns and incorporating demographic insights for risk assessment. All data management and analysis were performed using SAS 9.4, ensuring efficiency and reproducibility in data processing.

STATISTICAL ANALYSIS AND RESULTS

Regional and State-Level Accident Incidence Analysis

To investigate geographic patterns in traffic accident rates, we conducted a multi-level analysis examining both regional and state-specific variations. This approach allowed us to identify high-risk areas and quantify disparities in accident incidence rates across the United States.

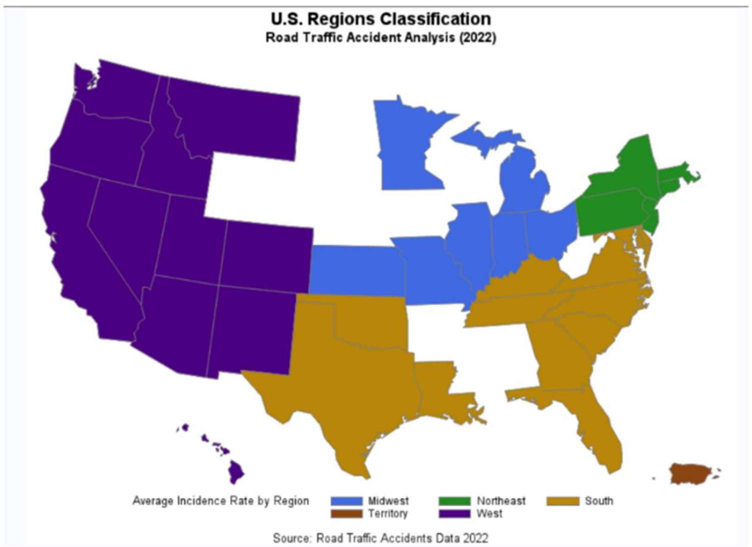

Figure 1: Classification of US map in terms of regions

The PROC GMAP procedure was then used to create a choropleth map, applying a blue gradient color scheme, where darker shades indicate higher accident rates. The choropleth visualization of regional accident incidence rates (Figure 2) revealed pronounced geographic disparities. Using a blue gradient color scheme where darker shades indicate higher accident rates, this map provides a clear visual representation of regional risk patterns. The West region displays the darkest shading, indicating substantially higher accident rates than other regions, particularly compared to the Midwest, which shows the lightest shading.

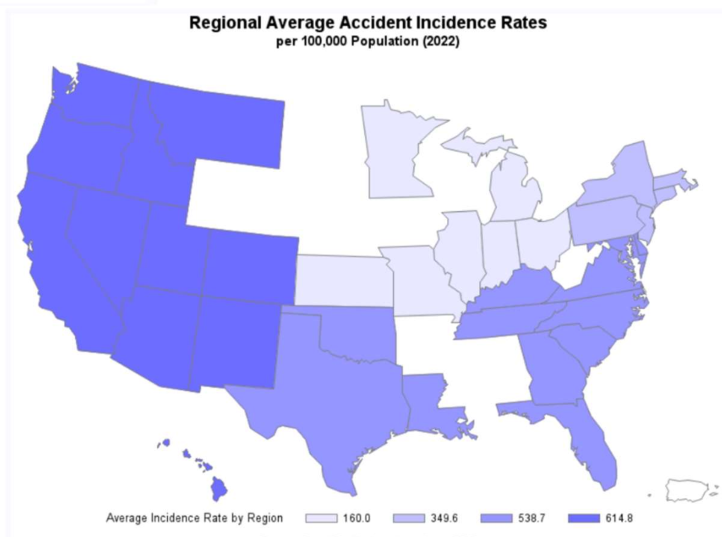

Figure 2: US map depicting Region Wise accident rates

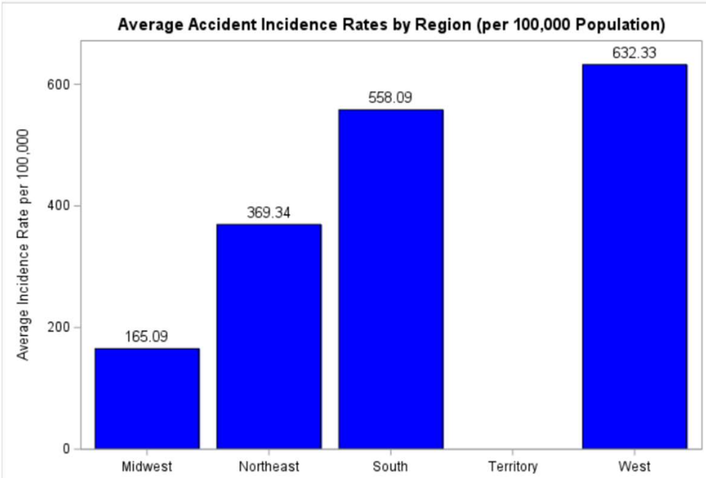

Figure 3: Bar chart for US Regional Accident Rates in 2022

Regional Incidence Rate Comparison

The bar chart comparison of regional accident rates (Figure 3) quantifies these regional disparities. The West region has the highest incidence rate at 632.33 accidents per 100,000 population, followed by the South at 558.09 per 100,000. In contrast, the Midwest has markedly lower rates at only 165.09 per 100,000—nearly four times lower than the West. The Northeast (369.34) falls in the middle range.

State-Level Variation

Examining state-level incidence rates (Figure 4) revealed significant variations within regions. The state-level choropleth map shows that certain states, particularly in the Western region (California) and Southeastern states (Florida, North Carolina, South Carolina, and Virginia), have notably high accident rates, shown in dark blue (between 769.6-1187.6 per 100,000 population).

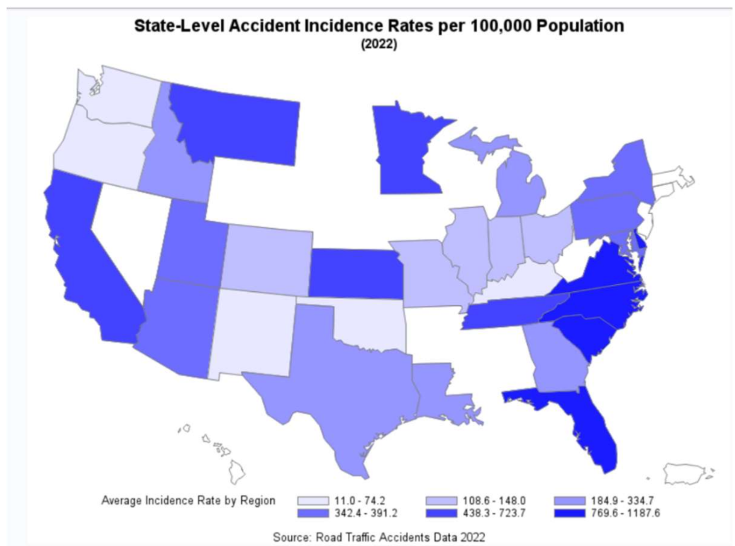

Figure 4: State-Level Accident Incidence Rates Choropleth Map (2022)

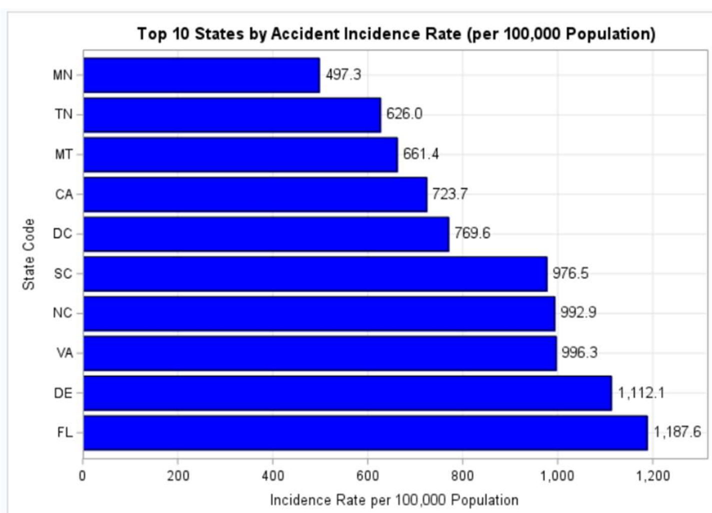

Figure 5: Top 10 States by Accident Incidence Rate (2022)

A horizontal bar chart of the top 10 states by incidence rate (Figure 5) further illustrates these patterns, with Florida showing the highest rate (1187.6 per 100,000), followed by Delaware (1112.1) and Virginia (996.3). This detailed state-level analysis highlights that even within regions, there can be substantial variation in accident risk.

Statistical Modeling with Negative Binomial Regression

To quantify regional differences while accounting for population differences and overdispersion in the accident count data, we conducted negative binomial regression analysis. The Poisson model initially considered showed severe overdispersion (Pearson Chi-Square/DF = 533.11), confirming that the negative binomial model was more appropriate for these data.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.6013	0.1354	-5.8667	-5.3358	1710.49	<.0001
region	Midwest	1	-0.8054	0.2271	-1.2505	-0.3603	12.58	0.0004
region	South	1	0.4128	0.1606	0.0980	0.7276	6.61	0.0102
region	Territory	0	0.0000	0.0000	0.0000	0.0000	.	.
region	West	1	0.5378	0.1641	0.2162	0.8594	10.74	0.0010
Dispersion		1	1.3536	0.0778	1.2094	1.5150		

Table 1: Negative Binomial Regression Results for Regional Analysis

The negative binomial regression results (Table 1 and Figure 6) using the Northeast as the reference region show:

1. **West Region**: 54% higher accident incidence rate compared to the Northeast (IRR = 1.71, p = 0.0010)
2. **South Region**: 41% higher accident incidence rate compared to the Northeast (IRR = 1.51, p = 0.0102)
3. **Midwest Region**: 81% lower accident incidence rate compared to the Northeast (IRR = 0.45, p = 0.0004)

All regional differences were statistically significant ($p < 0.05$), indicating genuine geographic disparities after controlling for population differences. The dispersion parameter (1.3536) was significantly greater than zero, confirming that the negative binomial model appropriately accounted for overdispersion in the data.

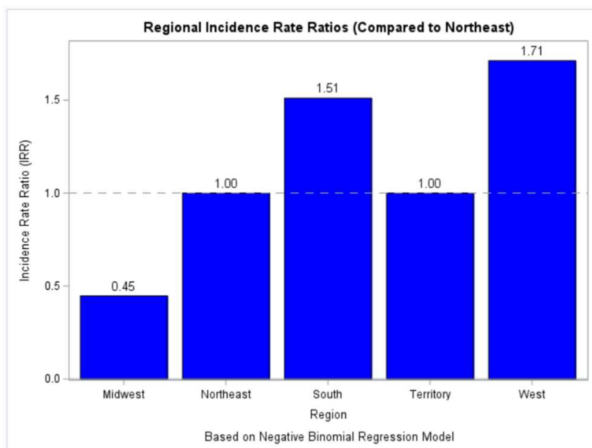

Figure 6: Regional Incidence Rate Ratios Compared to Northeast Region

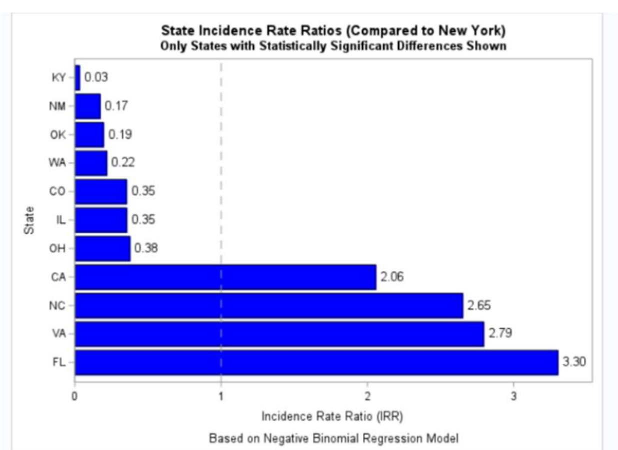

Figure 7: State Incidence Rate Ratios Compared to New York (Statistically Significant States Only)

The state-level negative binomial regression (Figure 7) using New York as the reference state revealed substantial variation in accident risks among states. Florida had the highest relative risk at 3.30 times the rate of New York ($p < 0.001$), followed by North Carolina (2.65 times higher, $p = 0.0031$) and Virginia (2.79 times higher, $p = 0.0018$). Conversely, several states showed significantly lower rates than New York, including Kentucky (97% lower, $p < 0.001$), Washington (78% lower, $p = 0.0018$), and Ohio (62% lower, $p = 0.0088$).

LIMITATIONS

This study has several limitations. Data limitations and reporting biases may lead to underreporting of accidents, as not all incidents are recorded uniformly across states. The exclusion of certain states and territories limits generalizability, particularly for rural areas. Confounding factors, such as driver behavior, road infrastructure, and traffic laws, weather conditions were not accounted for, which may influence accident rates. Despite rigorous data processing, potential merging inconsistencies could slightly impact accuracy. Future research should incorporate real-time weather data, expand geographic coverage, and consider additional risk factors for a more comprehensive understanding of accident trends.

CONCLUSION

The results of our analysis align with previous research, confirming that states such as California,[6] Florida, Virginia, and Arizona[4] experience some of the highest accident rates in the U.S. These states, characterized by high traffic density, urban congestion, and diverse weather conditions, consistently report elevated accident incidence rates.[5] Additionally, regional disparities observed in this study reflect broader trends, with the South and West exhibiting higher accident rates compared to the Northeast and Midwest.

The findings highlight the need for targeted safety measures such as infrastructure improvements, stricter enforcement, and public awareness campaigns. Future research should explore real-time weather impacts and additional risk factors to improve traffic safety strategies nationwide.

REFERENCES

1. Road traffic injuries. Accessed February 21, 2025. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. cycles T text provides general information S assumes no liability for the information given being complete or correct D to varying update, Text SCDM up to DDTR in the. Topic: Road accidents in the United States. Statista. Accessed February 21, 2025. https://www.statista.com/topics/3708/road-accidents-in-the-us/
3. Overview of Motor Vehicle Traffic Crashes in 2022 [Traffic Safety Facts - Research Note]. Accessed February 21, 2025. https://rosap.ntl.bts.gov/view/dot/78044
4. Analysis of the Causes of Car Accidents in the United States of America in 2023: Gauge People Understanding of Data Visualisation. Accessed February 21, 2025. https://arxiv.org/html/2406.17872v1
5. Predicting black ice-related accidents with probabilistic modeling using GIS-based Monte Carlo simulation | PLOS ONE. Accessed February 21, 2025. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0303605
6. Spatial and Temporal Analysis of Road Traffic Accidents in Major Californian Cities Using a Geographic Information System. Accessed February 21, 2025. https://www.mdpi.com/2220-9964/13/5/157

ACKNOWLEDGEMENTS

The authors acknowledge the guidance and support of Dr. William MacLeod, Research Associate Professor in the Department of Global Health at Boston University School of Public Health. His expertise and mentorship were instrumental in shaping this research.

CONTACT INFORMATION

For any questions or suggestions, please contact the authors at:
Urmi Phanse: urmi@bu.edu; Ojasvi Vachharajani: ojasvipv@bu.edu