Aayush Acharya, Justyna Szkaradek

Dr. Afenya

MTH 346-01

05-16-2025

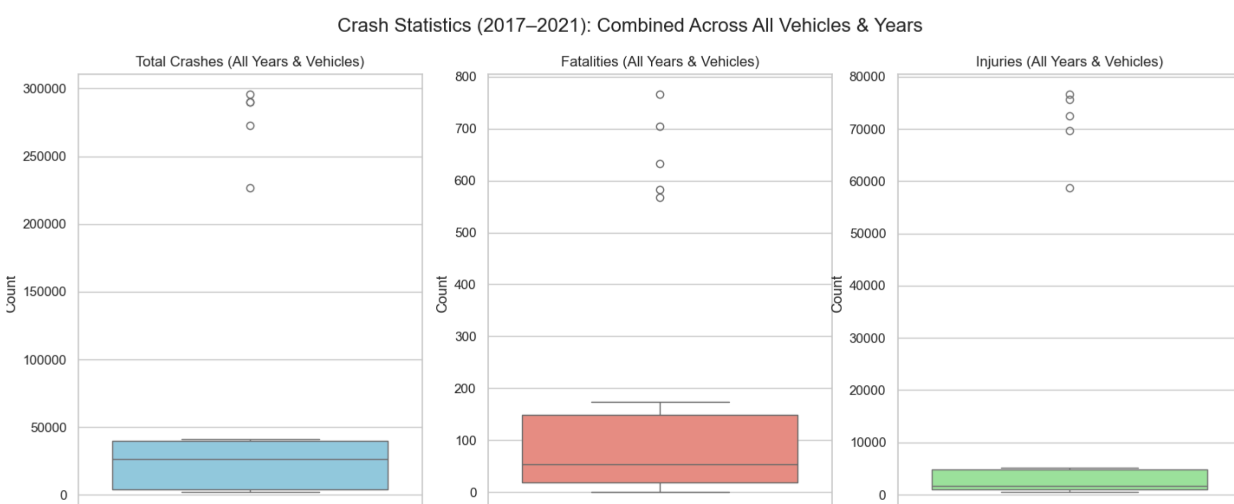
ABSTRACT

Crash Statistics 2017 - 2021

We found our data on the Illinois crash data trends from 2017 to 2021. We collected the data that was given for total crashes, fatalities, and injuries. We didn't have any problems with the data gathering process because we didn't collect the data ourselves. The subject of our investigation was looking at all the trends for total crashes, fatalities, and injuries in Illinois. Our study is relevant because we can identify safety issues, improve road design, and implement effective countermeasures to reduce accidents and fatalities.

VISUALS AND INTERPRETATIONS

Box Plots



These three boxplots summarize the overall distribution of total crashes, fatalities, and injuries across all vehicle types and years from 2017 to 2021. Each plot combines data points from different vehicles and years, allowing us to assess the general spread and detect potential outliers. There are clear extreme outliers for all three categories, most likely corresponding to Passenger Cars, which had significantly higher crash counts than other vehicle types. The boxplots reveal wide variability, moderate skewness, and significant deviations from symmetry in each category, reinforcing our understanding of how extreme values can impact summary statistics and highlighting the importance of visual tools for distribution analysis.

Central Tendencies

Total Crashes Summary

	Mean	Std Dev
Bus	3626.8	942.7
Large Trucks	19091.6	2024.0
Motorcycle	3064.6	234.7
Other Vehicles	34847.6	2480.9
Passenger Car	275005.8	28472.1
Pickup Trucks	38984.8	3206.4

Fatalities Summary

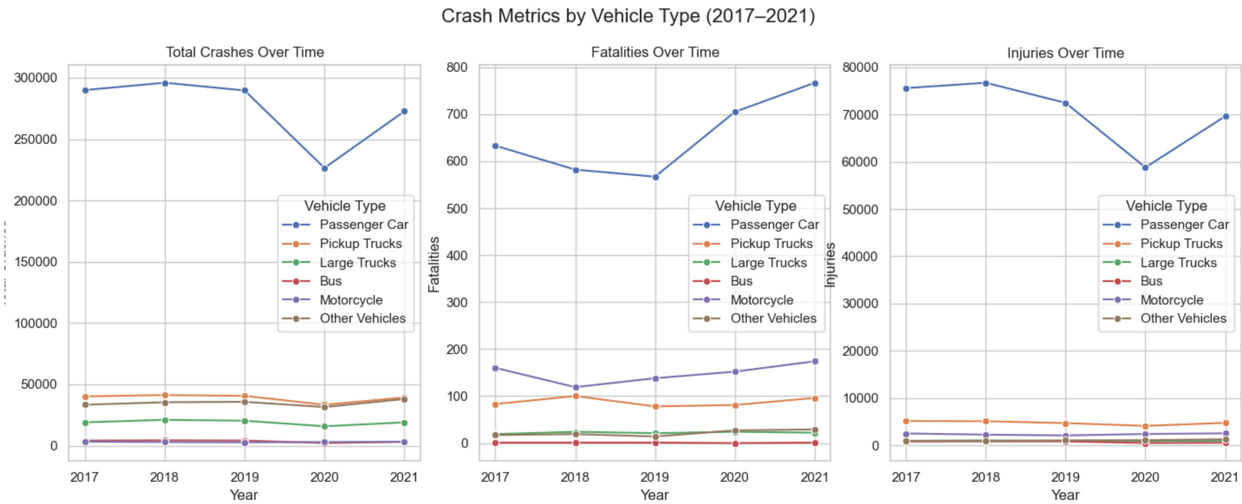
	Mean	Std Dev
Bus	0.8	0.4
Large Trucks	22.0	2.1
Motorcycle	148.6	21.1
Other Vehicles	21.2	6.5
Passenger Car	650.8	84.4
Pickup Trucks	87.6	9.8

Injuries Summary

	Mean	Std Dev
Bus	711.0	178.6
Large Trucks	954.4	65.7
Motorcycle	2365.4	185.2
Other Vehicles	1057.0	152.6
Passenger Car	70641.8	7172.4
Pickup Trucks	4771.8	411.2

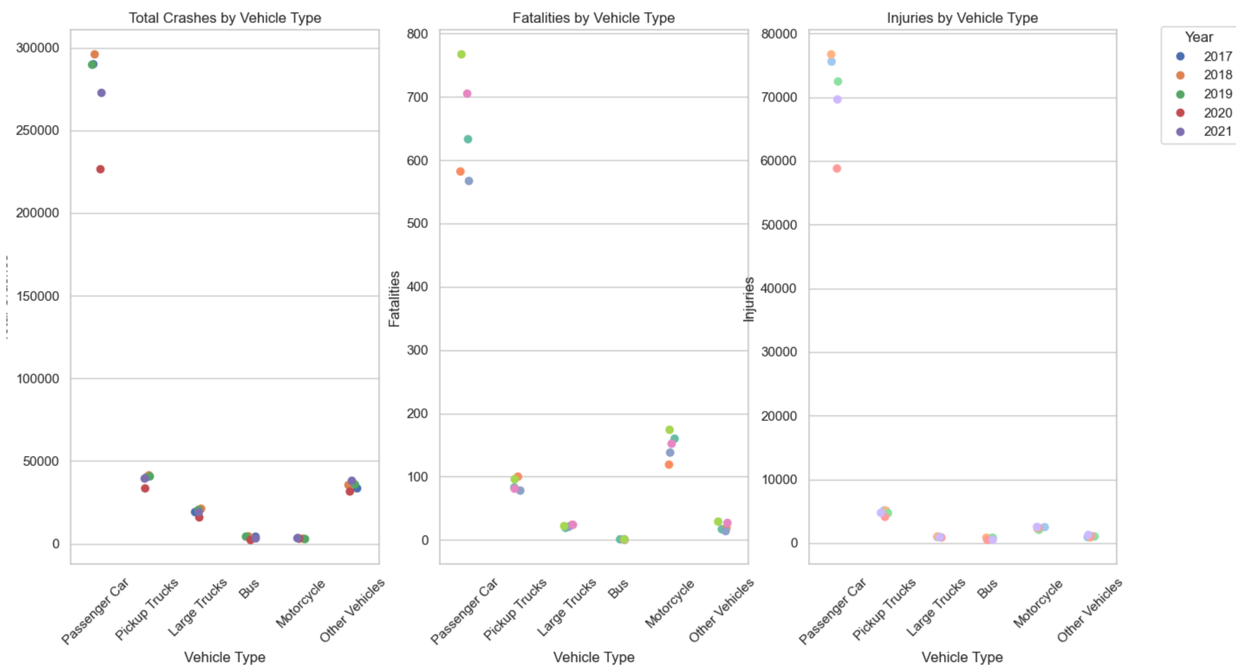
These central tendency summary tables display the mean and standard deviation for total crashes, fatalities, and injuries across all vehicle types from 2017 to 2021. Passenger Cars dominate all three categories, averaging over 275,000 crashes and more than 70,000 injuries annually, far exceeding all other vehicle types, which explains their influence on earlier outlier visuals. The standard deviations are also highest for passenger cars, indicating significant variability in crash and injury counts across years. In contrast, Motorcycles stand out with a disproportionately high mean fatality count (148.6), reflecting their vulnerability despite having fewer total crashes than other vehicle types.

Line Graphs



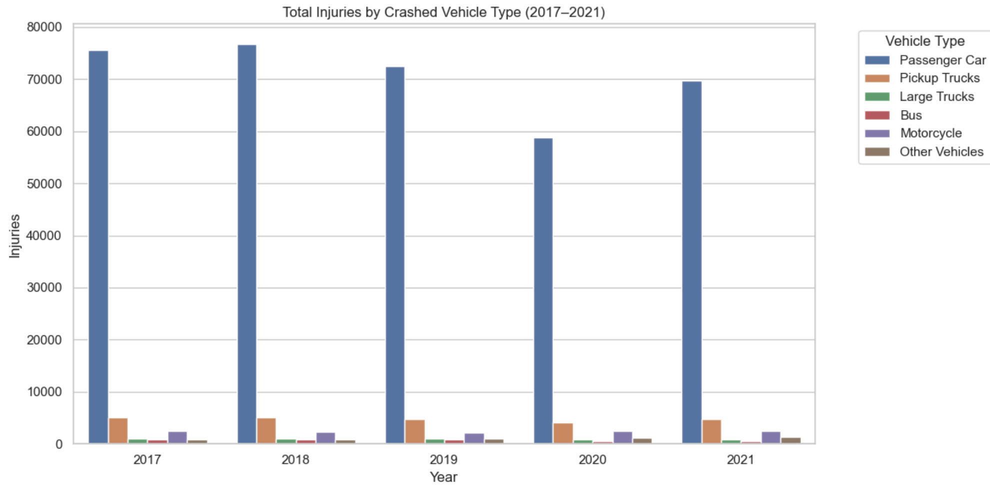
These line graphs depict the year-over-year trends in total crashes, fatalities, and injuries for each vehicle type from 2017 to 2021. Passenger Cars consistently had the highest crash and injury counts, with a sharp decline in 2020 likely due to pandemic-related travel reductions, followed by a rebound in 2021. While other vehicle types maintained relatively stable patterns, Motorcycles showed a noticeable upward trend in fatalities, peaking in 2021 despite moderate crash numbers. This visualization highlights both the dominance of passenger vehicles in overall incidents and the growing fatality concern surrounding motorcycles, reinforcing the importance of temporal analysis in traffic safety studies.

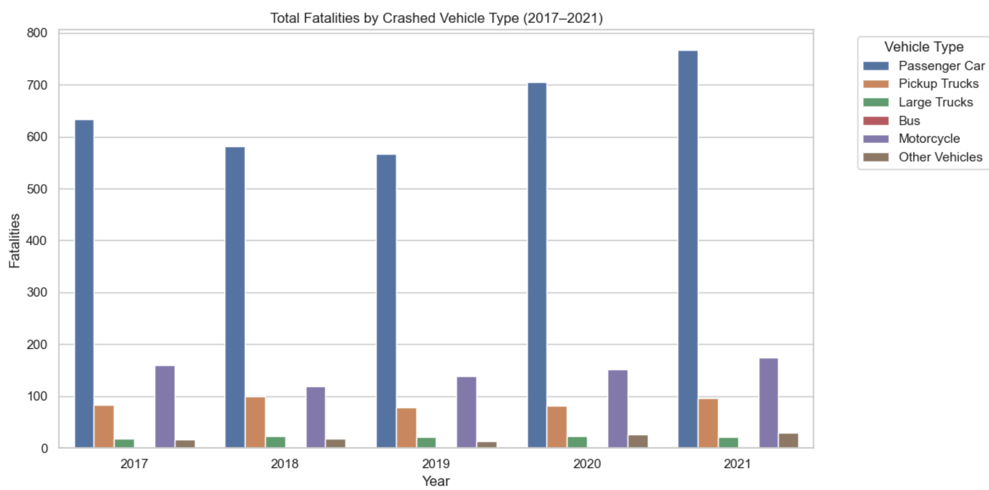
Dot Plots

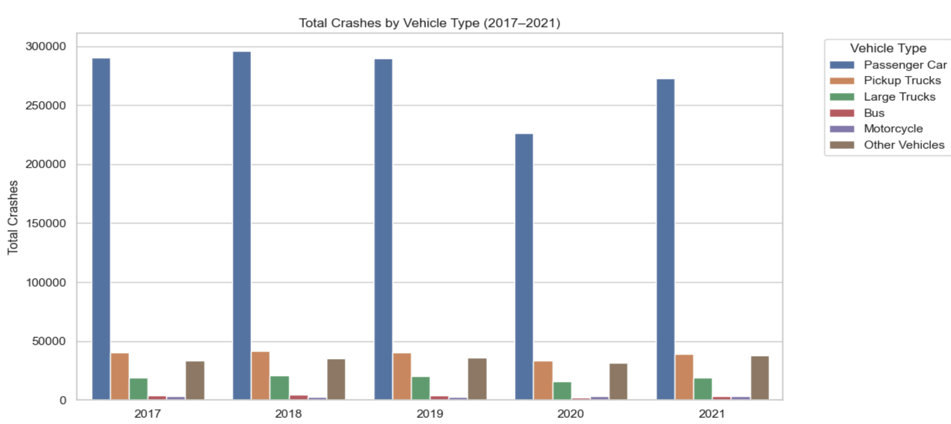


These dot plots illustrate the spread of total crashes, fatalities, and injuries for each vehicle type across all five years, with each colored dot representing a specific year. Passenger Cars again stand out with consistently high crash and injury counts, while Motorcycles exhibit elevated fatalities relative to their crash volume, suggesting a higher risk per incident. The tight clustering of most dots for smaller vehicle categories (like Buses and Large Trucks) indicates relatively stable year-to-year patterns. This visualization reinforces earlier findings and helps us compare both inter-vehicle disparities and intra-vehicle stability or shifts over time.

Bar Graphs

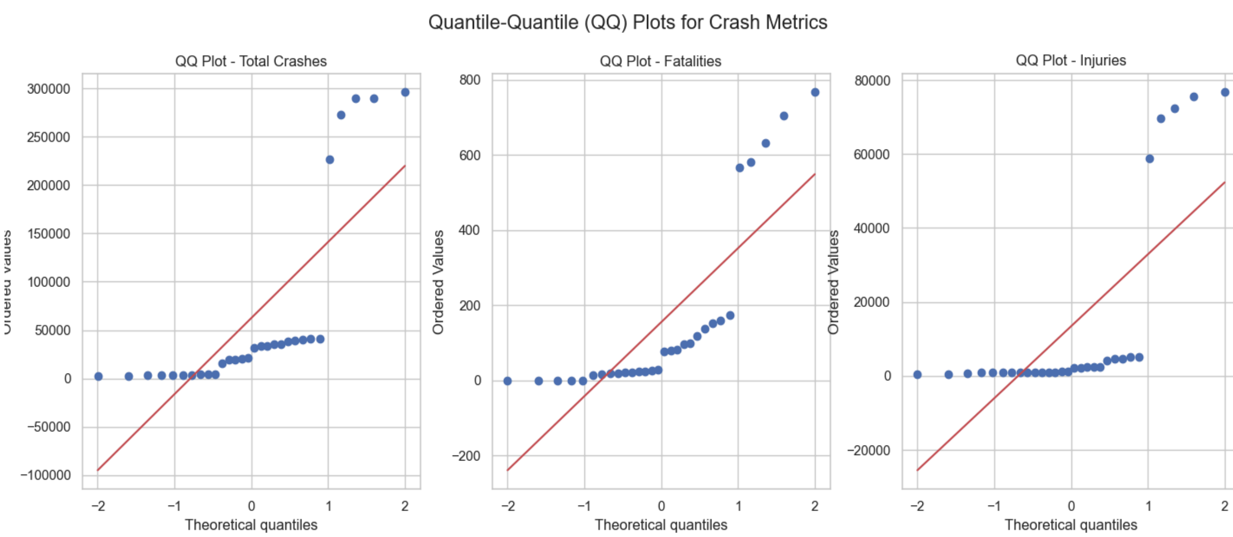

Total Injuries by Crashed Vehicle Type (2017–2021)
Injuries
Year
0
10000
20000
30000
40000
50000
60000
70000
80000
2017
2018
2019
2020
2021
Vehicle Type
Passenger Car
Pickup Trucks
Large Trucks
Bus
Motorcycle
Other Vehicles


Total Fatalities by Crashed Vehicle Type (2017–2021)
Fatalities
Year
0
100
200
300
400
500
600
700
800
2017
2018
2019
2020
2021
Vehicle Type
Passenger Car
Pickup Trucks
Large Trucks
Bus
Motorcycle
Other Vehicles


Total Crashes by Vehicle Type (2017–2021)
Total Crashes
Year
0
50000
100000
150000
200000
250000
300000
2017
2018
2019
2020
2021
Vehicle Type
Passenger Car
Pickup Trucks
Large Trucks
Bus
Motorcycle
Other Vehicles

These bar charts highlight consistent trends in crash outcomes across vehicle types. Passenger Cars dominate all three metrics with significantly higher totals than any other category each year, contributing most to overall crash burden. Motorcycles, while much lower in total crashes and injuries, stand out with a disproportionately high number of fatalities — suggesting higher risk severity. The relative stability in crash patterns among the other vehicle types emphasizes the unique scale and impact of both passenger vehicles and motorcycles in traffic safety outcomes.

Quantile-Quantile Plots



These QQ plots compare the distribution of total crashes, fatalities, and injuries against a theoretical normal distribution to assess normality. In each plot, substantial deviation from the red diagonal line, especially at the upper ends, indicates strong positive skewness and the presence of outliers, with crash and injury data showing the most pronounced divergence. This suggests that these variables are not normally distributed, likely due to extreme values from high-frequency categories like passenger cars.

GENERAL ANALYSIS

The dotplot reveals different aspects of our data distribution. They demonstrate the minimum, first quartile, median, third quartile, and maximum of the data. We also have three different categories representing the fatalities, injuries and total crashes for all vehicles during the years of 2017 through 2021. There are 4/5 extreme outliers in each of our boxplots. These outliers represent the years that had a significantly different amount of crashes than the central tendencies. The data is skewed, it is skewed to the right. All of our bar graphs are unimodal, because passenger cars have the greatest amount of crashes recorded. The data doesn't show any normality because it is not in the form of a bell curve. One interesting thing is that the total crashes, fatalities, and injuries do not align with each other. There is a correlation between total crashes and injuries. Almost all graphs and plots between these two are identical. Total crashes and injuries have a positive correlation with each other because of the variability in the crash severity. There is a hypothesis that we could have tested. Our Ho could be that there is no difference in the mean number of crashes across years. Our H1 could be that there is a statistically significant difference in crash counts across years. The p-values can be obtained by testing if differences in yearly crash, fatalities, and injuries. The p-value could be as low as 0.05. The goodness of fit can be found with the r squared value for linear regression.

CONNECTION TO COURSE CONCEPTS AND LEARNING OUTCOMES

Working with the Illinois crash data from 2017 to 2021 allowed us to apply nearly every major concept from MTH 346 in a real-world context. By calculating measures of central tendency and variation such as the mean and standard deviation for total crashes, fatalities, and injuries across vehicle types, we reinforced our understanding of descriptive statistics covered early in the course. Creating visualizations like histograms, box plots, dotplots, stem-and-leaf plots, and quantile plots helped us explore distribution patterns, identify outliers, and assess normality — directly connecting to our learning on probability distributions and data analysis.

This project also emphasized the value of using statistical software, as we used Python to compute statistics and generate plots efficiently, aligning with the course's goal of integrating technology into data analysis. Through scatterplots and line graphs, we visualized trends and considered correlations across time and vehicle categories, laying the groundwork for understanding regression and hypothesis testing. Overall, working hands-on with this dataset deepened our comprehension of both descriptive and inferential statistics and strengthened our ability to interpret and communicate statistical findings — core goals of the course.

COMMENTS AND INDIVIDUAL ROLES

The software we used was Python including its various libraries such as pandas, matplotlib.pyplot, seaborn, scipy.stats and numpy to create the visuals, and we collected our data from the Illinois Department of Transportation. To complete this project Aayush studied the data and made all the visuals, and Justyna analyzed and explained all the visuals and data.

CONCLUSION

Our analysis of Illinois crash data from 2017 to 2021 revealed clear patterns in crash severity and frequency across different vehicle types. Passenger cars consistently accounted for the highest number of total crashes, injuries, and fatalities, contributing significantly to the state's overall crash burden. While most vehicle types maintained stable trends, motorcycles stood out with disproportionately high fatality rates relative to their crash counts, emphasizing their vulnerability on the road.

Through a variety of visual charts created using Python libraries, including boxplots, dotplots, line graphs, bar charts, and QQ plots, we were able to detect skewed distributions, identify outliers, and observe meaningful trends over time. Our statistical summaries further reinforced these insights, showing wide variability particularly in passenger vehicle data. This project not only helped us apply key statistical methods but also demonstrated how data analysis can inform public safety efforts and transportation policy. Moving forward, such findings could support targeted interventions to reduce crash-related injuries and fatalities in high-risk vehicle categories.

SOURCE:

https://idot.illinois.gov/content/dam/soi/en/web/idot/documents/transportation-system/resources/safety/crash-reports/trends/trends%202017-2021.pdf