Cincinnati Traffic Crash Report

Author: Shams Sadhin
Last Modification: 1/14/2024

Executive Summary

The project focuses on analyzing traffic crashes around Cincinnati using open-source data from the City of Cincinnati website. The dataset contained detailed records of traffic crashes with 3833796 observations and 28 variables. After conducting data cleaning in R, a Tableau dashboard was developed to show key metrics and visualize how different factors contribute to fatal crashes.

Data Preparation involved filtering missing observations, combining duplicate values and creating additional columns essential for better analysis. Outlier Latitude and Longitude values were corrected and columns with excessive missing data were dropped. Following this the cleaned dataset was exported to Tableau.

The Tableau dashboard revealed several important insights. It includes a density map show ing crash frequency across Cincinnati. It includes filters such as crash severity, weather, light conditions and type of person impacted. A key insight from the density map is that highways, intersections and campus areas had large volume of crashes.

Further analysis in R using ggplot explored trend in fatal injuries in regard to age groups, crash manners, road contours, light conditions, and road conditions. The findings showed that young drivers have the highest fatal injury counts, possibly due to risky behaviors. Collision involving pedestrians and curved hilly roads had high risks. In addition, driving during nighttime and wet roads led to higher proportion of fatal injuries.

Topic and Problem Statement

The goal of this project is to conduct exploratory data analysis on the causation of traffic crash in the city of Cincinnati using open-source data from the data.cincinnati.gov. Additionally, a tableau dashboard is created to provide important metrics and visualize how different variables lead to more fatal crashes.

Dataset

The data was obtained from the City of Cincinnati open data source (https://data.cincinnati-oh.gov/safety/Traffic-Crash-Reports-CPD-/rvmt-pkmq/about_data) The data is created by City of Cincinnati Police Department and contains records of all the traffic crashes which is updated daily. The original dataset contains 383796 observations and 28 variables.

Preparing Dataset

Upon inspection of the longitude and latitude values, it was observed that latitude ranged from -84.49 to 842432889.00084 and longitude ranged from 0.00098 to -39085518846.0071, whereas it should be from 38.9 to 39.5 and -85 to -84.2 for Cincinnati.

Upon inspection, variables AGE, GENDER, ADDRESS_X, UNITTYPE were filtered and missing observations were dropped. Columns CRASHDATE, ROADCLASS, CRASHLOCATION, ROADCLASSDESC were dropped as they had significant missing records and not relevant for our analysis.

Duplicate values in several columns (LIGHTCONDITIONSPRIMARY, ROADCONDITIONSPRIMARY) were combined. Additional data cleaning was conducted to clean variable names by removing unnecessary symbols.

A new variable, Age_Group was created for analysis purposes.

Dplyr package in tidyverse was used for the data cleaning process.

```
print(missing_summary)
```

```
                    ADDRESS_X                     LATITUDE_X                    LONGITUDE_X
                           49                             67                             71
                          AGE COMMUNITY_COUNCIL_NEIGHBORHOOD                CPD_NEIGHBORHOOD
                        48229                              0                              0
               SNA_NEIGHBORHOOD                    CRASHDATE                  CRASHLOCATION
                            0                             40                         189775
                 CRASHSEVERITY                CRASHSEVERITYID              DATECRASHREPORTED
                           21                             21                             49
                     DAYOFWEEK                         GENDER                       INJURIES
                           38                          43975                            382
                    INSTANCEID           LIGHTCONDITIONSPRIMARY                  LOCALREPORTNO
                            0                             45                              0
                  MANNEROFCRASH           ROADCONDITIONSPRIMARY                   ROADCONTOUR
                           45                             45                             45
                    ROADSURFACE                     ROADCLASS                  ROADCLASSDESC
                           45                         148843                         148935
                      UNITTYPE                   TYPEOFPERSON                        WEATHER
                           385                            373                             45
                             ZIP
                            7192
```

Data Visualization in Tableau

After cleaning the data, the csv was exported and connected to Tableau. Click on the link to view the interactive dashboard: https://public.tableau.com/views/CincinnatiTrafficCrashDashboard/Dashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

The Tableau dashboard shows important metrics including the Grand Total of crashes and the different types of crashes. The density map shows the region where traffic crashes are clustered and a quick glance shows that there are specific regions where crashes are clustered where people should be cautious. Highways interstate, 4-way crossings and busy streets around campus are primarily those regions. The dashboard also shows crashes by day of week, crash severity, gender, weather and who is involved in the crash (Driver, Occupant or Pedestrian.)

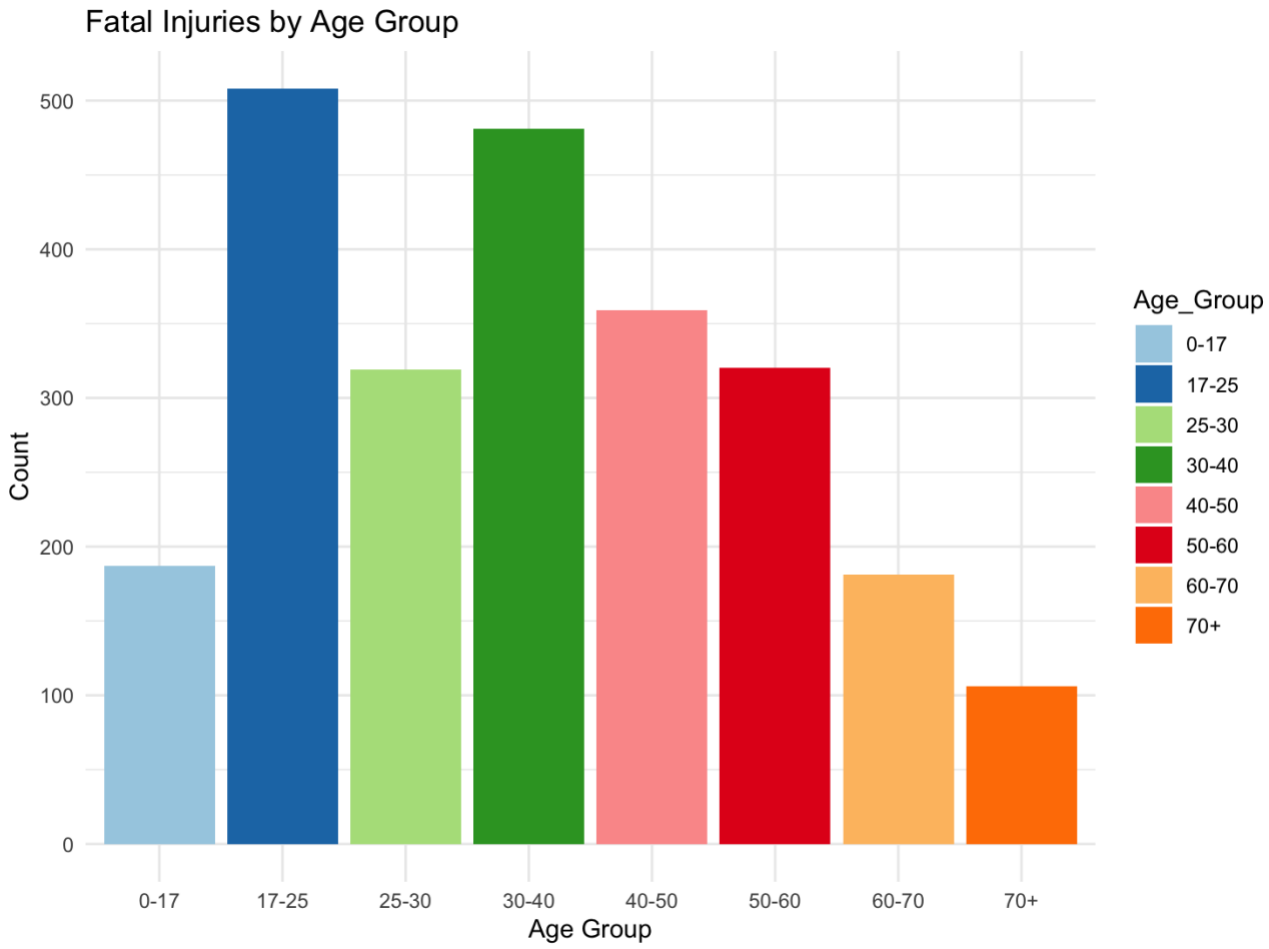
Data Analysis in R

Based on the findings, the analysis focuses on comparing and analyzing the differences between crashes resulting in fatal injuries and those leading to non-fatal injuries. Additionally, the Tableau dashboard highlights that while most traffic crashes involve drivers, some also involve pedestrians. This analysis aims to explore how these types of crashes differ.

The following questions will be addressed using ggplot in R:

1) a. How does Fatal Injuries compare between different age groups
   b. How does Fatal Injuries for Drivers compare against all fatal injuries
2) a. How does Fatal injuries compare between different Manners of Crashes
   b. How does Fatal injuries for Drivers compare between different Manners of Crashes
3) How does Road Contour vary for Fatal and non-Fatal Crashes
4) How does Light Conditions vary for Fatal and non-Fatal Crashes
5) How does Road Conditions vary for Fatal and non-Fatal Crashes

1a)



Observation:

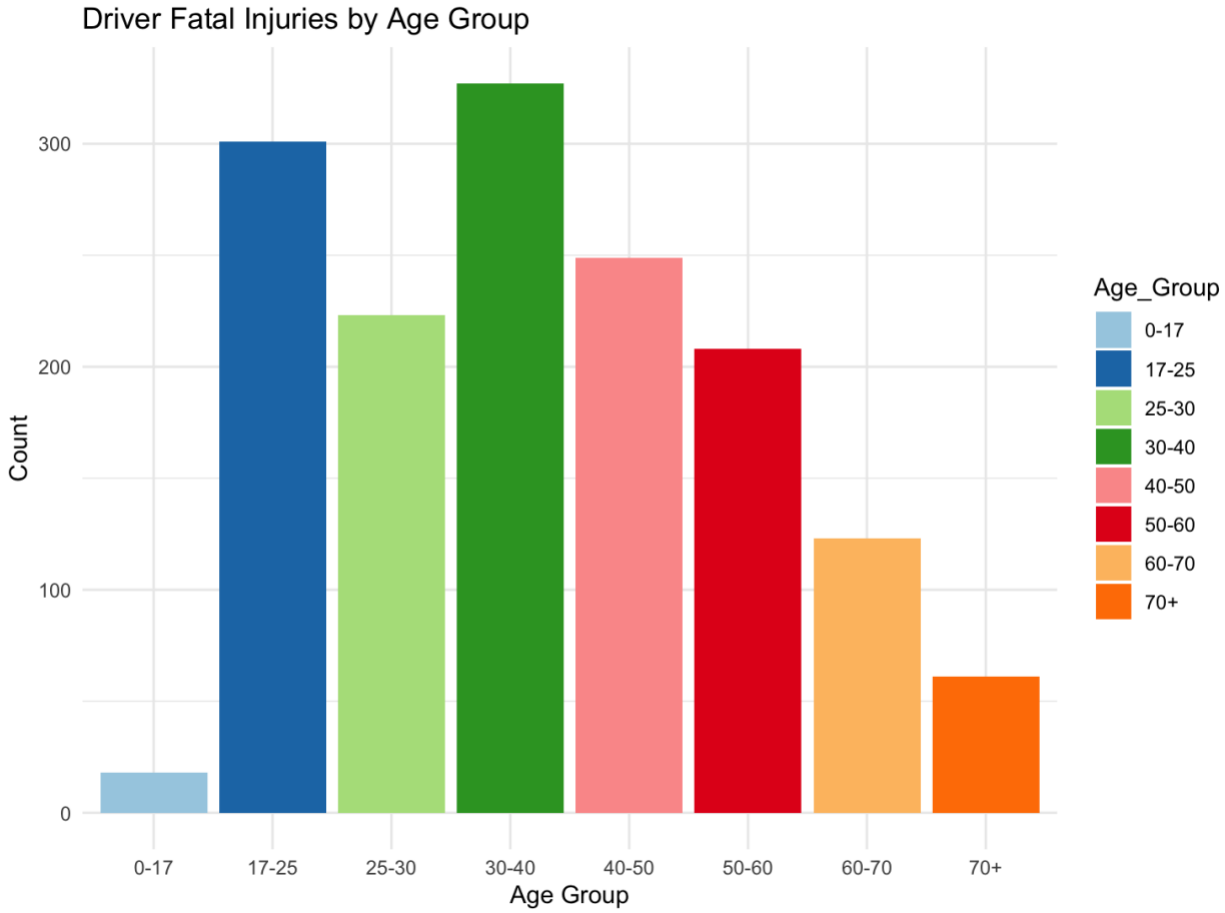
Age 17-25 has the highest count of fatal injuries. This could be because of risky behavior or exposure to dangerous activities.

Ages 25-30 and 30-40 also has high count of fatal injuries, containing the majority of US driving population.

Injuries decreases gradually as age group increases, possibly due to decline in activity levels or increased caution while driving.

Ages 0-17 also has surprisingly high count of fatal injuries and could be due to lack of parental supervision.

1b)



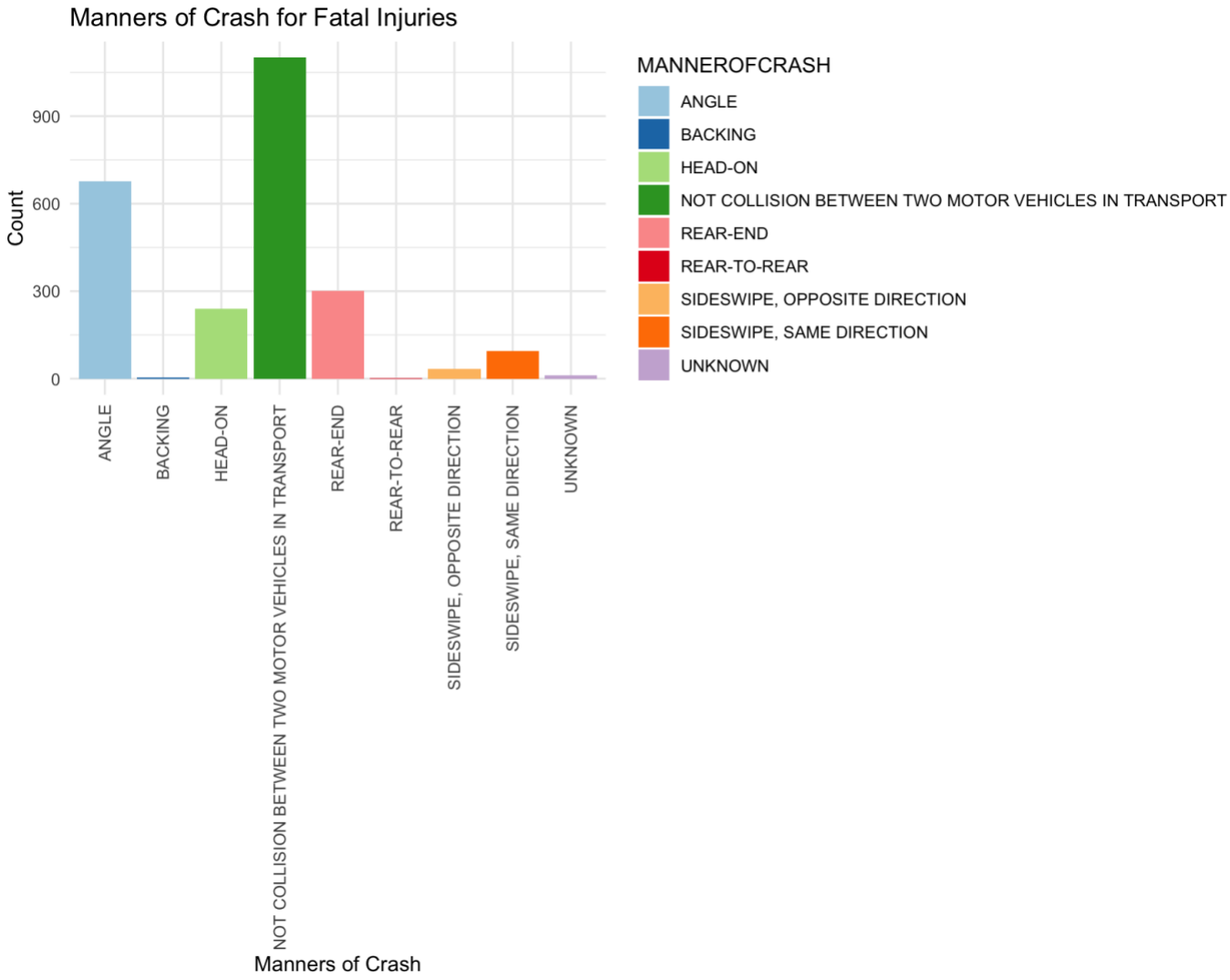
Observation:

Fatalities for Drivers only shows similar trends however age group 30-40 has the highest count, overtaking age 17- 25. This indicates ages 17 to 25 are also victims of fatal injuries as pedestrians. Moreover, there is a decrease in count for age group 0-17, indicating that most of the injuries suffered are pedestrians.

Recommendations:

1) Increase in road signage, safe crosswalks and decrease in speed limits near schools and neighborhoods with kids.
2) Implementation of Mandatory Road Safety Training for Ages 17 to 25 to increase awareness and stay educated even after getting driver's license.

2a)

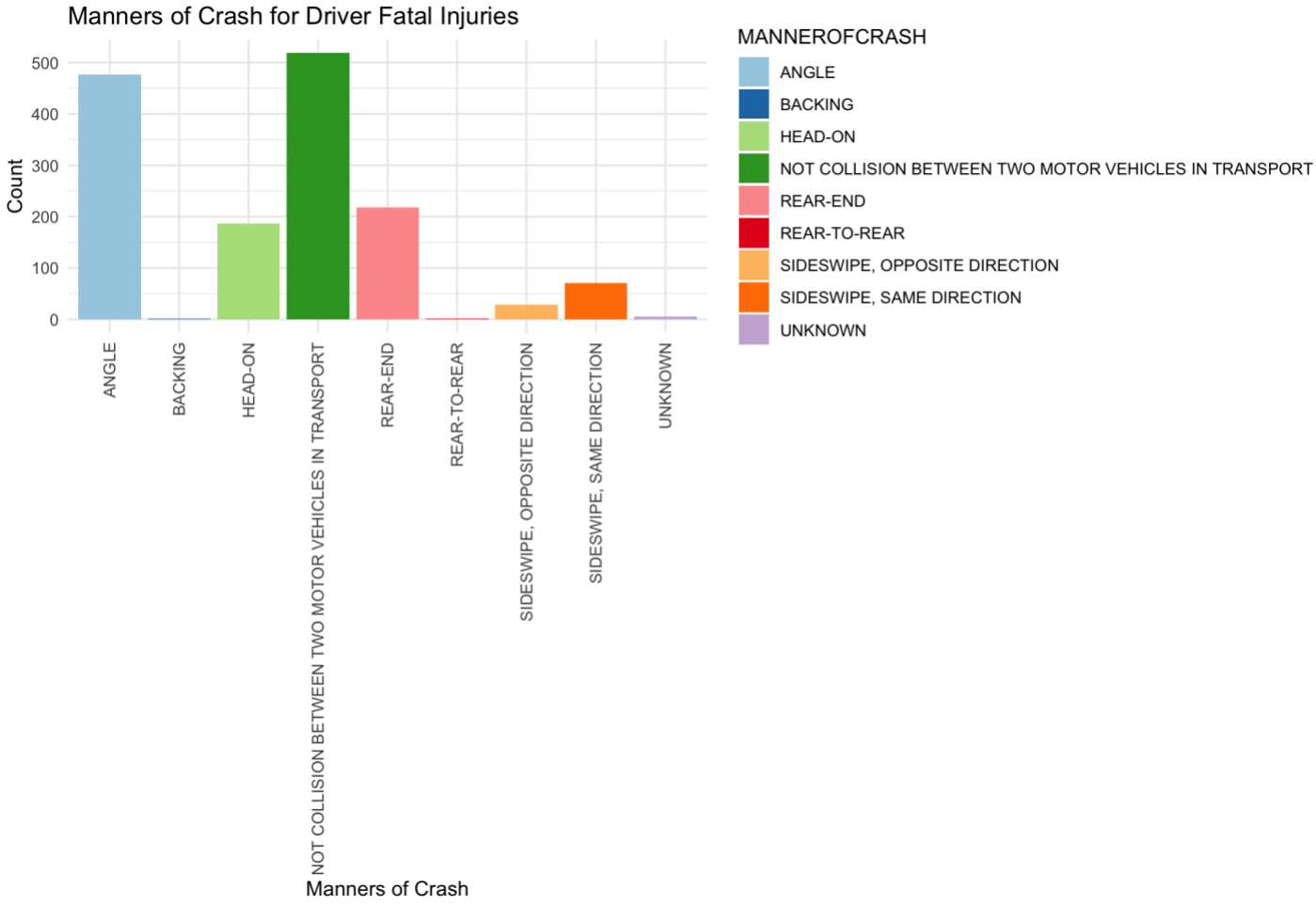


Observation:

The most common manner of crash is not a collision between two motor vehicle transport. This could include car collision with a pedestrian or object in the sidewalk/road.

Angle, Rear-end and Head-on follow up behind with high counts of Fatal injuries.
Sideswipe, Backing and Rear-to-Rear has low counts, likely due to less severe crashes.

2b)



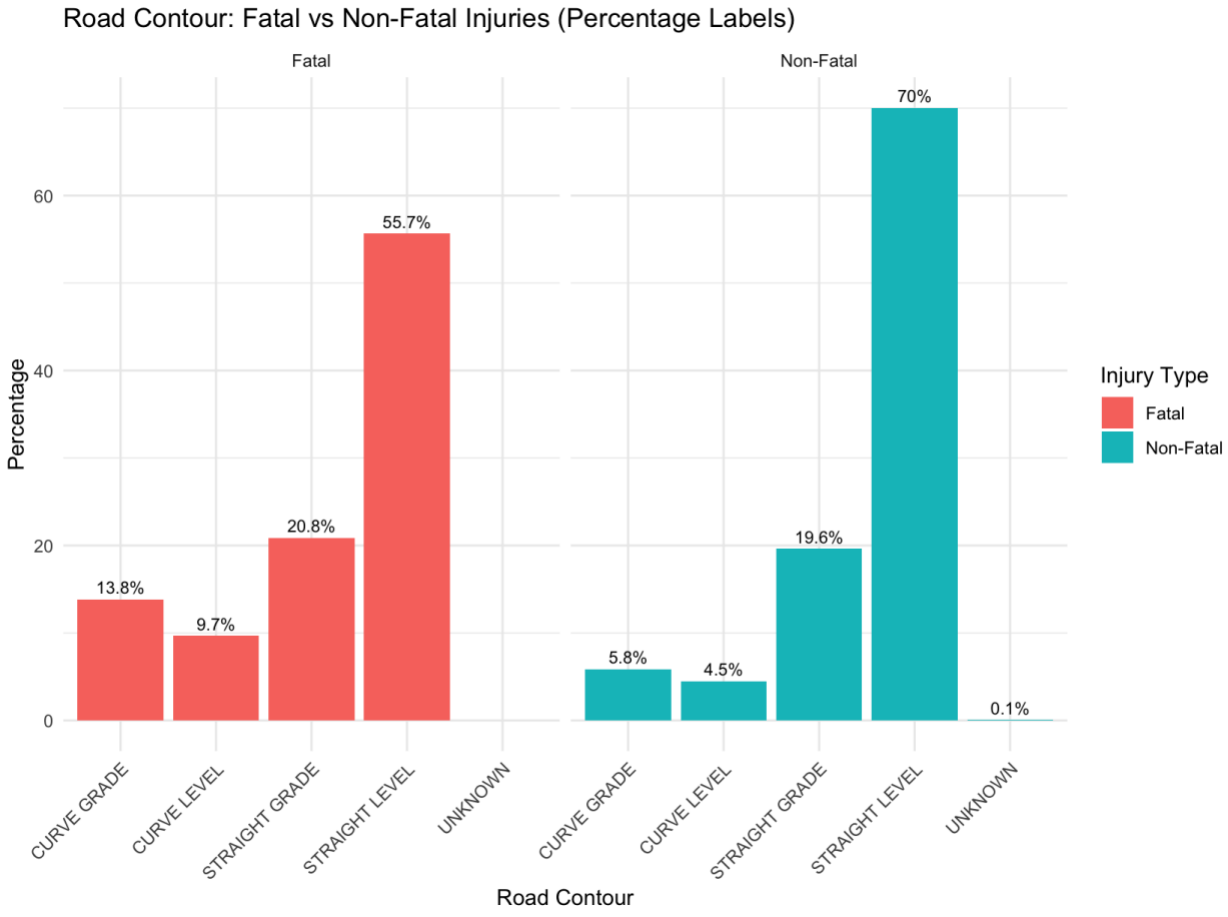
Observation:

There is no significant difference for fatal crashes where drivers are involved.

Recommendations:

1) Improve road infrastructure and visibility in areas where non-vechile collision occurred.

3)



Definitions

"_" Grade = roads having incline or declines (hilly areas)
"_" Level = roads on level surface

Observation:

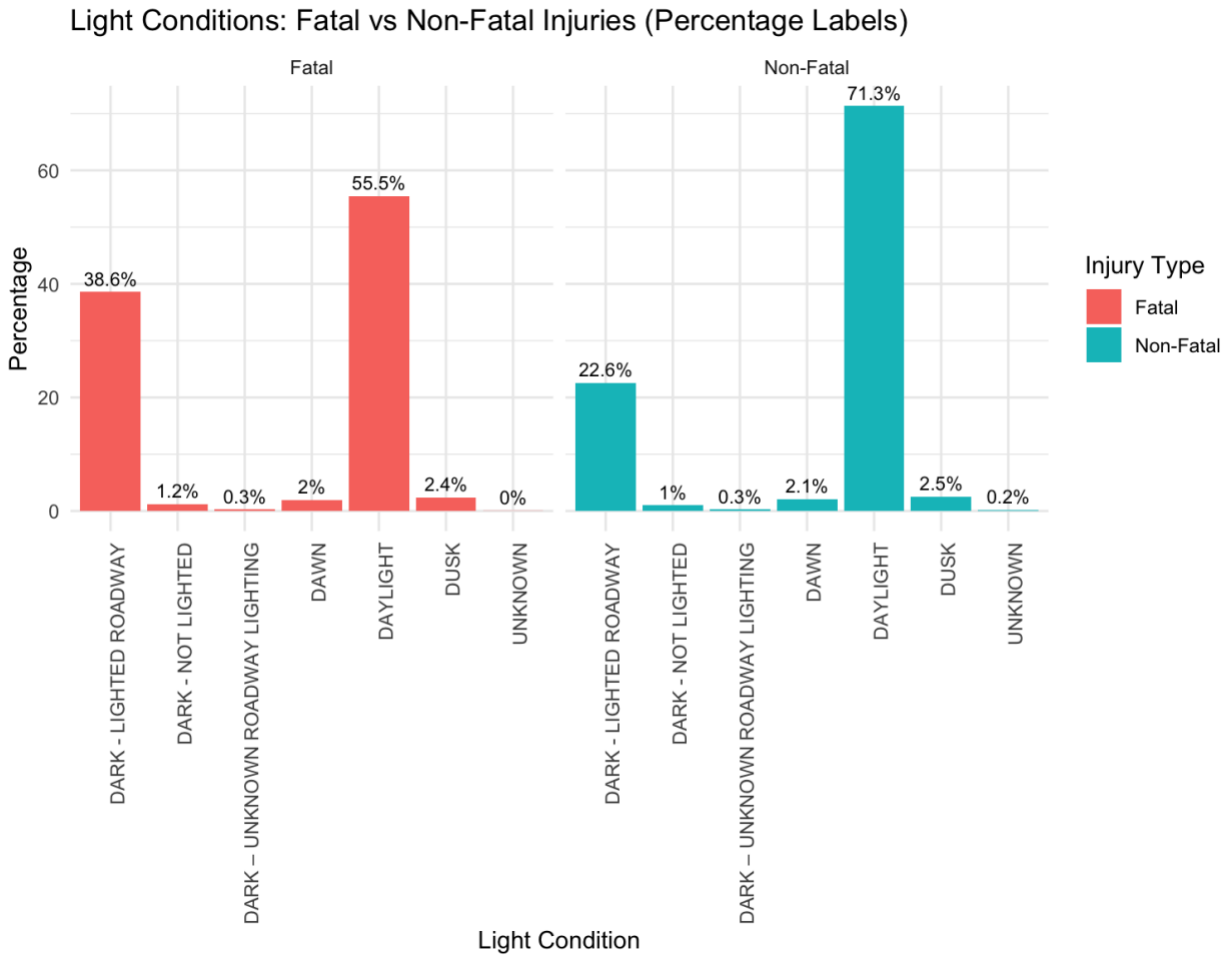
Majority percentage injuries are in straight level roads for both Fatal and non-fatal injuries, which could be due to higher ratio of straight level roads around the city, overconfidence and higher speeds.

There is a comparatively higher percentage of fatal injuries in curvy hilly roads and curvy level roads, indicating curvy roads lead to more fatal injuries

Recommendations:

1) Install speed cameras and lower high-speed limits in danger prone straight roads
2) Add reflective mirrors and better signage on curved roads

4)



Observations:

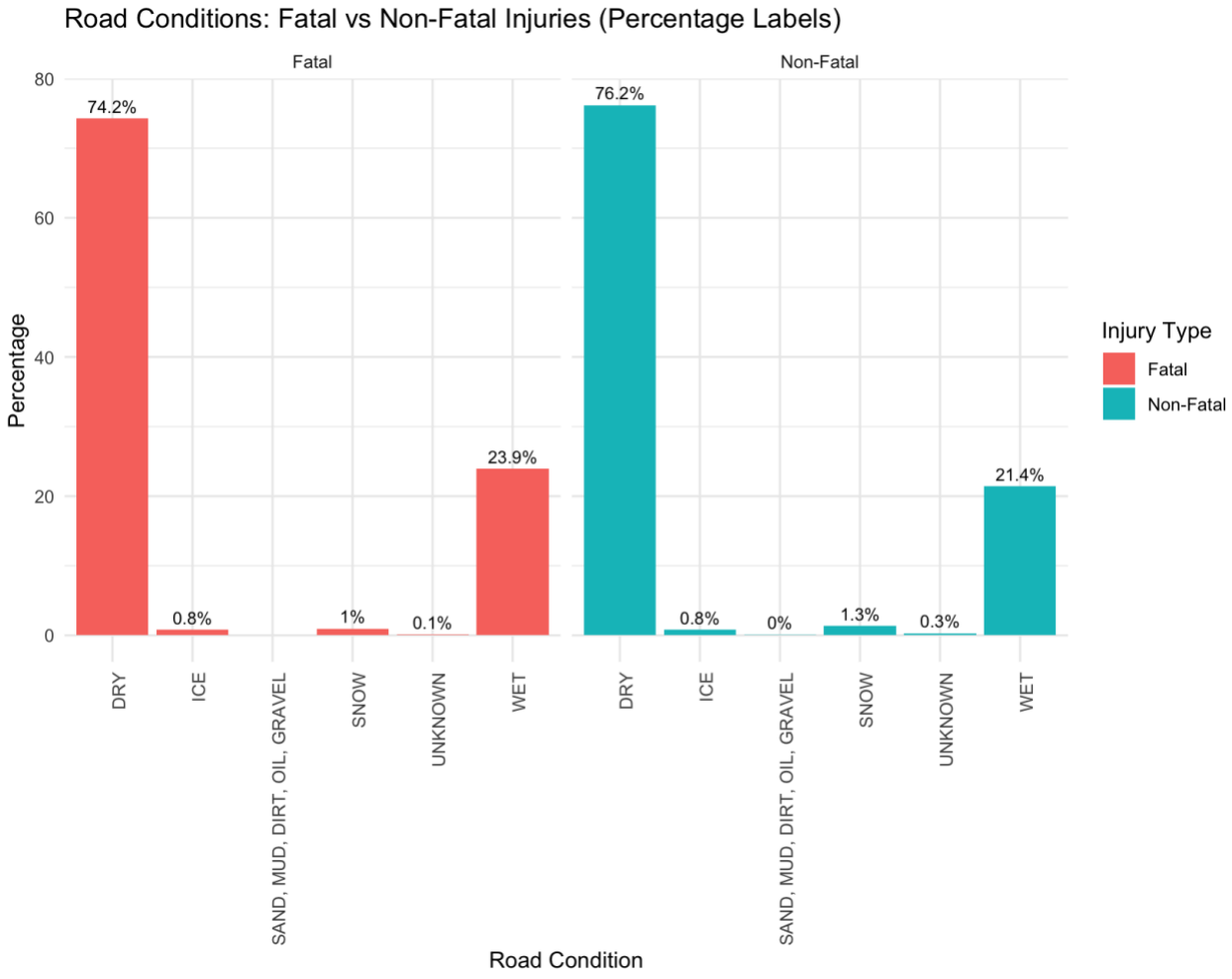
Majority percentage injuries occur during daylight for both fatal and non-fatal injuries, likely due to increased traffic volume and activity during the day.

There is a comparatively higher percentage of fatal injuries on dark-lighted roadways, indicating that nighttime driving, even with lighting, poses a significant risk due to reduced visibility and reaction time

Recommendations:

1) Increase in patrolling and road security during nighttime which might mitigate reckless driving
2) Install reflective lane markings for better visibility during night time

5)



Observations:

Majority percentage injuries occur on dry roads for both fatal (74.2%) and non-fatal (76.2%) cases, likely due to their higher occurrence and increased speeds during optimal conditions.

There is a slightly higher percentage of fatal injuries on wet roads (fatal: 23.9%, non-fatal: 21.4%), indicating that slippery surfaces could increase the risk of deadly accidents.

Ice, snow, and other conditions contribute slightly to fatal injuries, possibly due to less frequent occurrences or more cautious driving in such conditions.

Recommendations:

Implement stricter monitoring on roads to address risks from higher speeds and overconfidence during optimal condition

Conclusion

This project identified key risk factors for traffic crashes in Cincinnati. Fatal crashes often involve young drivers, curvy roads and nighttime conditions.

To reduce these risks, the following recommends are proposed:

1) Increase road signage and safe crosswalks, especially near schools and neighborhoods.
2) Implement mandatory road safety training for ages 17-25 to improve awareness.
3) Improve road infrastructure, including reflective mirrors and better signage on curved roads.
4) Install speed cameras and reduce speed limits in high-risk areas.
5) Enhance nighttime road safety through increased patrolling and reflective lane markings.

This project provides actionable insights to improve traffic safety in Cincinnati and actions needed to decrease fatal traffic crashes.