California Traffic Collision Data

CIS 4560-01

Group 2

Fausto Vidrio

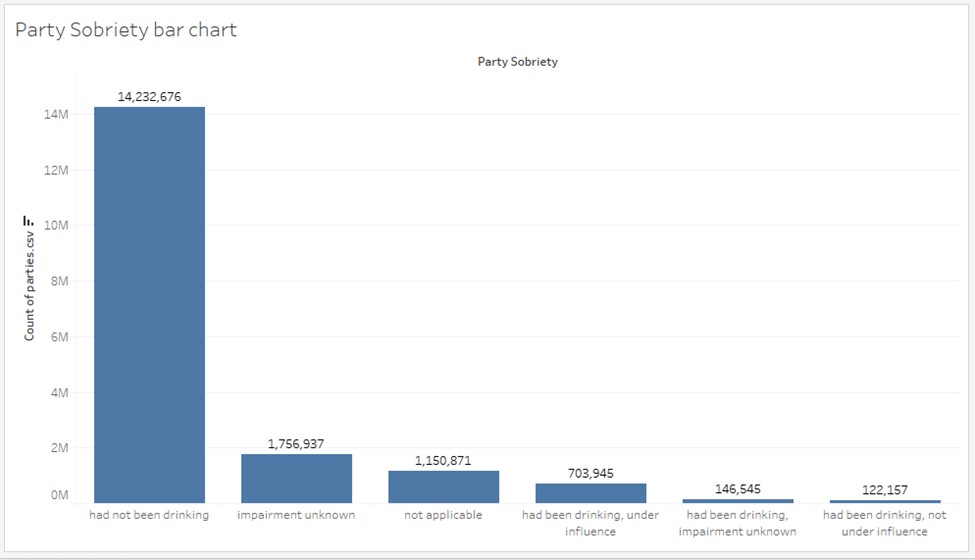
**Abstract:** This document’s purpose is to inform the reader about the findings of our research. Our group had discovered data regarding recently recorded automobile collisions in the state of California.This data was derived from the public website [www.kaggle.com](http://www.kagglr.com) under the data set name “California Traffic Collision Data from SWITRS[[1]](#footnote-0)”.

**Introduction**

In the year of 2021, the state of California had reported 4,161 deaths that were the result of a traffic collision. This was a slight increase from the year 2020. For this reason, our group decided to research and analyze traffic collisions from the state of California that have been recorded from the year 2001 until 2022. In this document, we will discuss some of our findings that will include analyzations of party sobriety, parties at fault, types of victims involved,

**Party Sobriety Analyzation**

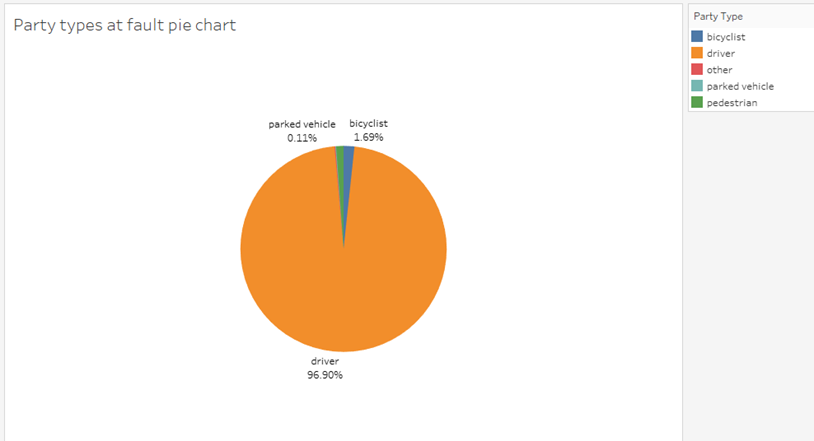
The objective of analyzing the data given from the parties table was to determine if the majority of the parties were under the influence at the time of their collisions. In addition, we also had an interest in determining the percentage contribution of the types of parties that were found at fault for their collisions. After analyzation of the party sobrieties at the time of their collisions, it has been determined that most of the parties involved in collisions were not under the influence which can be seen in the figure (figure 1). The total number of parties that had not been drinking prior to their collision amounted to a total of 14,232,676 parties. This information considers all the recorded collisions throughout every county of California.



***Figure 1****: Bar graph showing the different amount of party sobrieties recorded throughout the state of California. Sobrieties were recorded at or around the time collisions occurred****.***

**Party at Fault Analyzation**

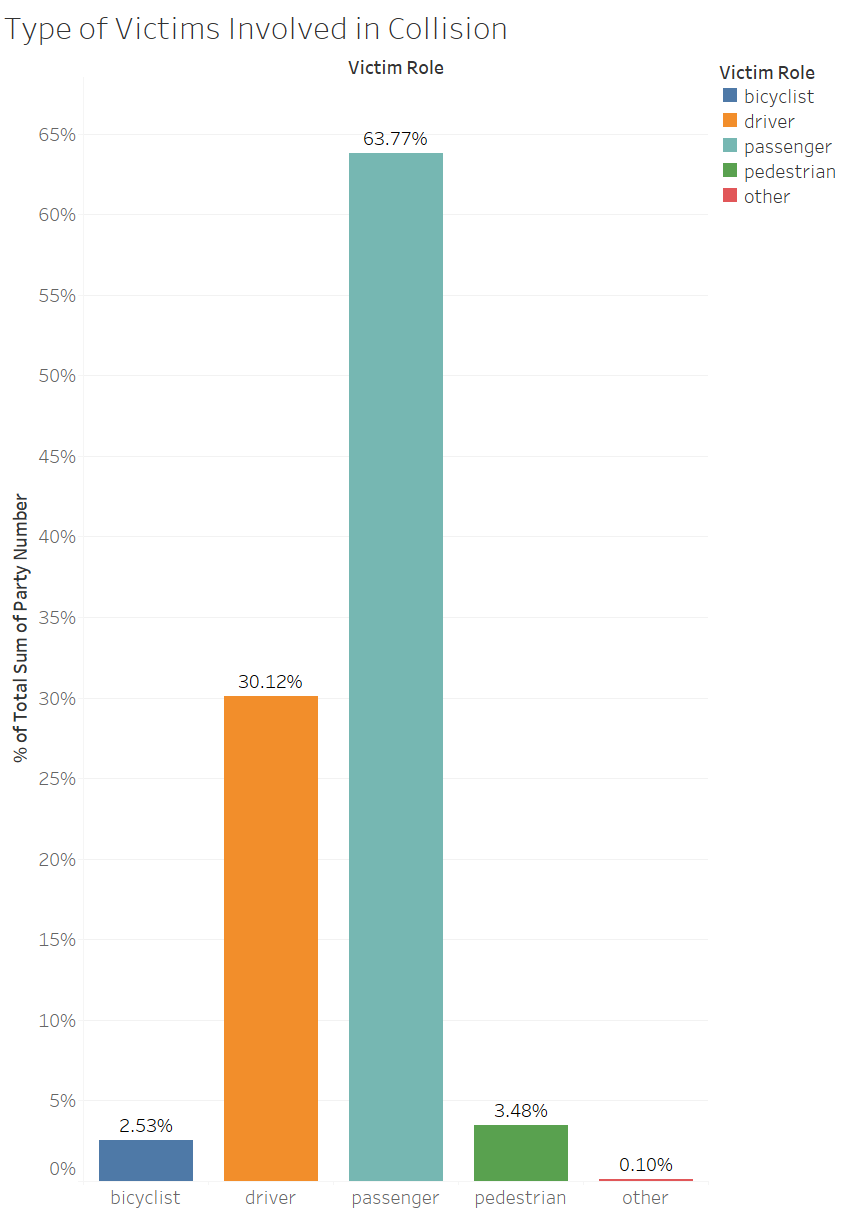
Additionally, there was an interest in analyzing the party types that were found at fault for the collisions they were involved in. Using the data recorded for the parties involved, a pie chart was created to compare the distribution of party types that were found at fault. As can be seen in the pie chart figure (figure 2), the majority of the party types that were found at fault were parties that were driving cars. Drivers were calculated at 96.90% of the party types that were found at fault. Bicyclists were calculated at 1.69%, pedestrians were calculated at 1.15%, and parked vehicles were 0.11%. This information also considers all the recorded collisions throughout every county of California.



***Figure 2:*** *Pie chart showing the distribution of party types that were found at fault of their collisions. Distribution is calculated by percentage of total.*

**Type of Victims Involved in Collision**

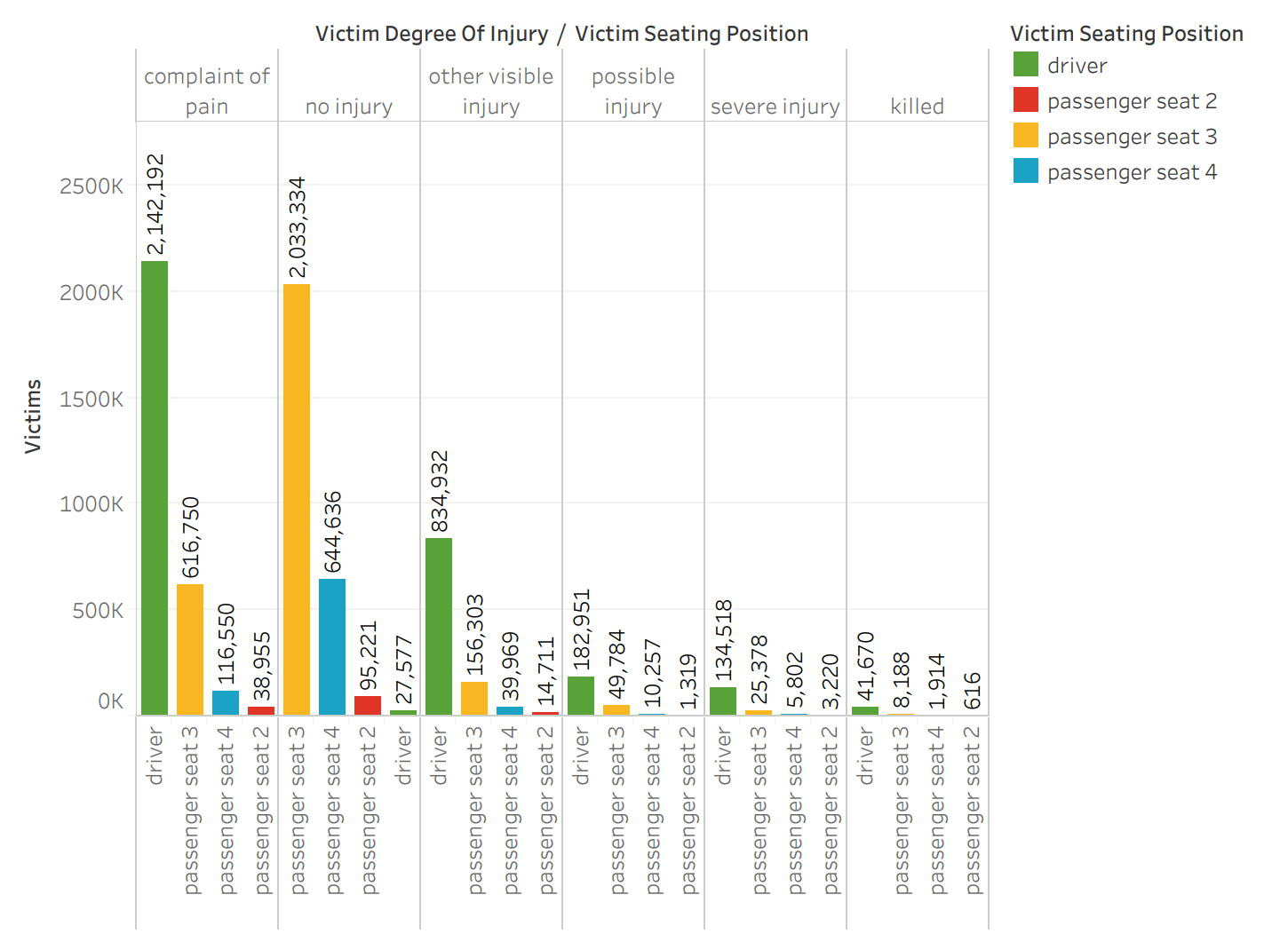
The victim table contains information on the injuries sustained by specific individuals involved in collisions. The dataset includes a total of 1,048,576 ID records, but the number of victims is 1,701,805. Using this table, the victims' roles were analyzed, which included bicyclists, drivers, passengers, pedestrians, and others. The graph below displays the percentage of victims from each category that become victims when involved in a collision. Bicyclists accounted for 2.53% (43,073) of becoming victims, drivers for 30.12% (512,601), passengers for 63.77% (1,085,180), pedestrians for 3.48% (59,277), and others for 0.10% (1,724). Based on the analysis, it can be concluded that passengers are the most frequent participants to become victims when getting into collisions, and following behind passengers are drivers who are victims when involved in collisions.



***Figure 3:*** *This bar graph presents the frequent percentage of participants (bicyclist, driver, passenger, pedestrain, & other) to become victims when involved in a collision.*

**Victims Degree of Injury**

In this data analysis we will focus on the victims table and analyze the degree of injury based on the seating position of the victims. We based this analysis on a seating of a sedan vehicle. From the analysis we can see that the drivers showed less complaint of pain when involved in an incident. Passenger seat 3 has an overwhelmingly amount of no injury reported followed by passenger seat 4. From this analysis we can conclude that the safest seat to be in when involved in an incident is passenger seat 3. At the same time, passenger seat 3 is the second deadliest seat to be in right under the driver. The driver leads in other visible injury and throughout the rest of degrees of injuries. Passenger 3 seems to follow the driver in 2nd place in the rest of injuries.

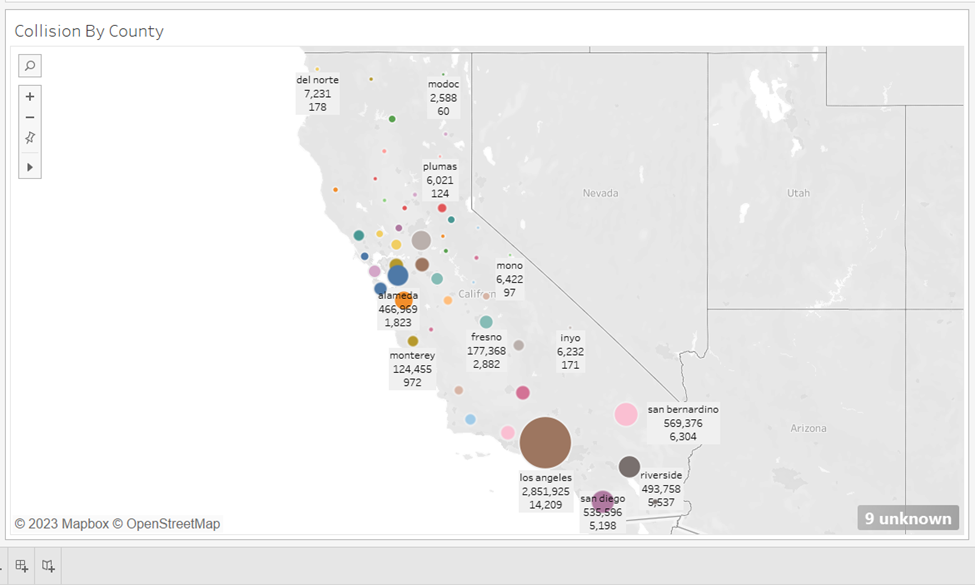


***Figure 4:*** *The bar graph shown above displays the degree of injury the victims sustained from their seating position.*

**Traffic Collision by County Analyzation**

In this analysis we will be comparing the traffic collision count by county in the State of California. With this analysis we will show which counties had the most traffic collisions from 2001 to 2020 based on the traffic collision reported by the CHP.

The figure below, shows a map of the state of California. To show which counties had the most traffic collisions, we used the symbol’s map from Tableau. With the symbol’s map, we were able to add the county location and the collision count to show which county in the state of California had the most traffic collisions. As you can see below, Los Angeles County had the most collisions with a total of 2,851,925. The second county was San Bernardino with 569,376 followed by the county of San Diego with 535,596.

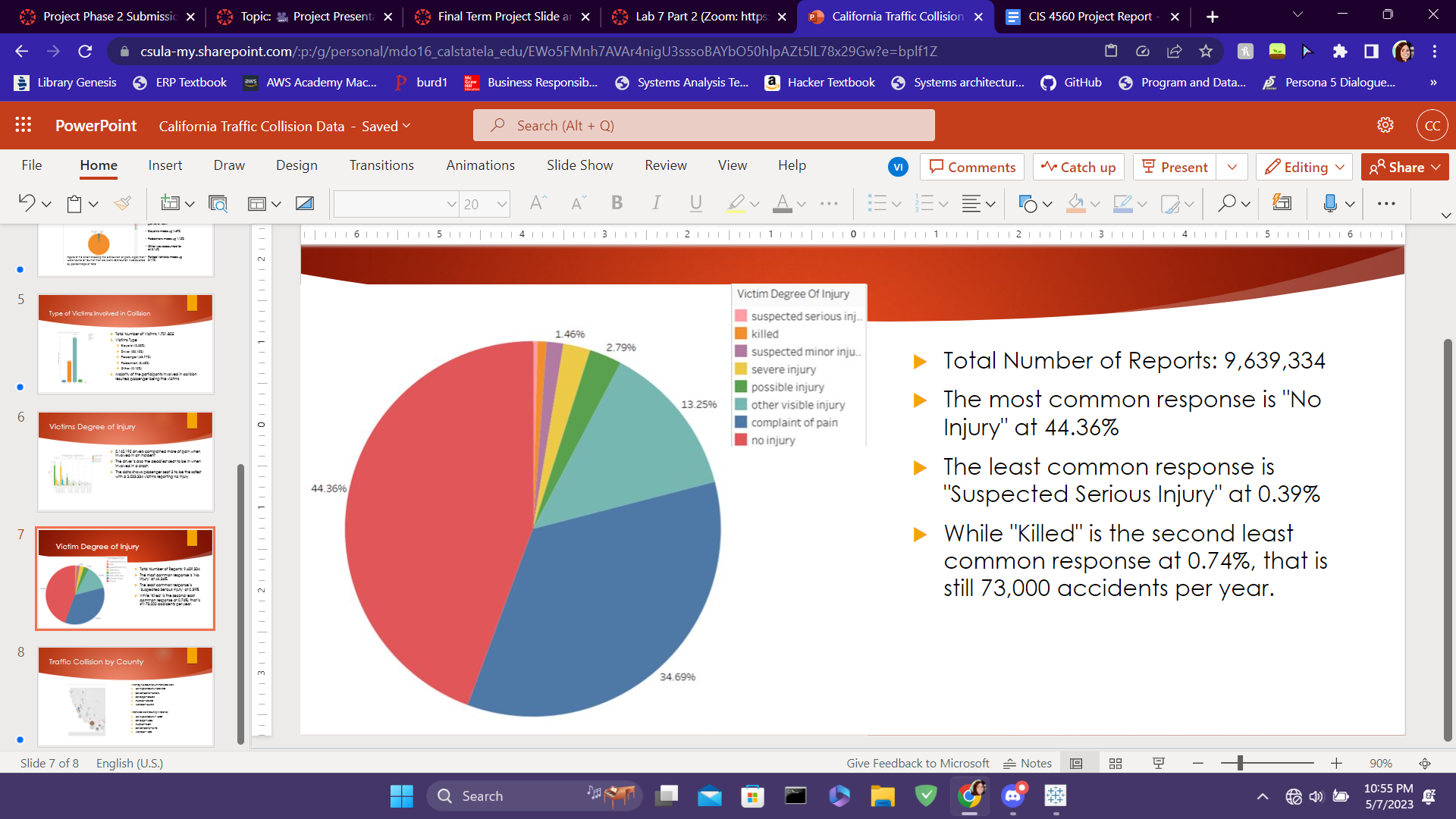
\

***Figure 5:*** *A bubble chart map of California. Each bubble represents a county and bubble size represents the amounts of collisions.*

Once we were able to see which counties had the most traffic collisions, we decided to show the traffic collisions resulting in fatalities. You will find the amount underneath the collision count.

**Victims Degree of Injury by Severity**

In this data analysis, we will focus on the degrees of injury experienced by car accident victims. This chart was based off of roughly 9.6 million reports of California Traffic Data 2021, showing the general distribution of injuries sustained from car accidents. The most common response is “no injury” making up 44.36% of the total. The least common response is “suspected serious injury” making 0.39% of the total. Only about 0.76% of all California car accidents in 2021 were fatal. Learning what types of car accidents result in different injuries is useful to law makers and car manufacturers to prevent these injuries from occurring in the future.

**

***Figure 6:*** *A pie chart showing frequency of degrees of injury from car accidents. Each type of injury is represented by a different color slice of the pie.*

**Works Cited**

“California Traffic Collision Data from SWITRS.” Kaggle, https://www.kaggle.com/datasets/alexgude/california-traffic-collision-data-from-switrs. Accessed 20 March 2023.

Elflein, John. “Motor-vehicle deaths in California 2021.” Statista, 10 June 2022, https://www.statista.com/statistics/675550/motor-vehicle-deaths-california/. Accessed 6 May 2023.

1. SWITRS is the State Wide Integrated Traffic Record System [↑](#footnote-ref-0)