Analysis and Visualization of Road Traffic Accidents in Michigan 2019

Annamalai Balasubramanian, Mageshwaran Anbazhagan, and Sapna Maduraimuthu

Abstract—Vehicle crashes play a significant role in human life and could bring devastating results due to various factors such as: Careless Driving, Intoxicated Driving, Speeding, Tailgating, Wrong-Way Driving, Improper turns, and so on. Traffic accidents in the United States are surging effectively and the rate of road fatalities in the United States is projected to be 12.4 deaths per 100,000 people. In addition to that, 4.4 million individuals had been badly hurt and need medical assistance. This annual increase in road accidents is highly distressing. Out of all the 50 states in the United States, this research mainly focuses on Michigan State to visualize and project the causes of the road accidents in the year 2019 where the raw data was gathered initially from various sources and organized, cleansed, preprocessed, and finally, visualization will be performed to bring out the various findings and to identify how the road accidents can be minimized. With this visualization, specialists can improve by analyzing the historical data and enhance the capabilities in managing the vehicle crash.

Index Terms—Fatalities, Distracted Driving, Fatal Crashes

1 Introduction

According to the World Health Organization, "Every 24 seconds, a life is lost on the road, and it costs countries around 3% of their gross domestic product". Traffic rules and systems are built by the government for its citizens' quick and uninterrupted movement with ease of accessibility. However, if the road regulations are not followed correctly it may lead to unexpected crashes or consequences. Road accidents may also occur due to other factors such as the geometric design of the road, peak hour of the traffic, type of the vehicle, alcohol concentration, and vehicle occupancy. The causes of these factors can be effectively understood using data visualization to implement novel ideas and powerful solutions that can leverage historical accident data to address deficiencies in current road systems. The principal aim of this study is to use technology to help the professionals to visualize the statistical data in-depth without any jeopardy using R. While the world is transforming into a new digital world, there is a huge need for visualizing the data using data science which helps in visualizing the historical accident data that reveals concealed facts. With pictorial representation, it can be easily understood by professionals and traffic law enforcement agencies. Consequently, it will be helpful for the planning engineers and decision-makers to counter the most contributing factors for the accident and devise an effective master plan for the mitigation of accidental cases. This paper helps in examining the factors that contributed to road traffic accidents by using data visualization in Michigan in 2019 by using the past data.

2 Motivation

Every year, approximately 1.3 million people's lives are lost as a result of a road traffic crash, which in turn can cause considerable economic loss to individuals, their families, and to nation as a whole. Despite the coronavirus pandemic in 2019, there have been only a few cars on the road in Michigan. However, there had been a greater number of fatal crashes, according to Michigan State Police. Thus, determining the possible causes or factors that led to the higher number of fatal crashes is significant. This project mainly focuses on the analysis and visualization of road traffic accidents in Michigan, which includes the following tasks:

- Trend analysis of fatal crashes and fatalities in Michigan over the time period of 10 years from 2009 to 2019.
- Comparison of Speed Limit categories.
- Proportion of genders based on their age group.
- Influence of distracted driving.
- Safer day and time to travel.
- Number of Fatal crashes across different counties in Michigan based on the driver's Alcohol consumption.

By understanding and visualizing the above tasks,one would get better insight into the possible tasks that could have led to the huge increase in fatal crashes in Michigan, such that the laws can be further strengthened by the Michigan State Police to reduce the number of fatalities and the number of fatal crashes in the upcoming years.

3 Literature Review

Road Injuries are one of the top 10 causes of death as classified by the World Health Organization in 2019. The Visualization was performed on Road Traffic Accidents data in the Hayatabad area, Pakistan, using Python programming language software that helped the researchers interpret the important information and devise effective prevention and safety measures to avoid the huge number of fatal crashes and fatalities in the future. [7] The visualization consists of four major steps they are: Collection of raw data from various sources such as GPS data, official local department statistics, or the incident logs, preprocessing the raw data by converting the data into temporal properties, making use of the visual symbols and icons like bar and line chart and finally the visualization was created by using Infographics and maps. The research has provided information about the fatal accident time, literacy of the driver, gender, vehicle type, and deaths per year driving out the insight that the most fatal accidents happened in the daytime and the year 2020. Similarly, the research also focuses on the nature of the injury, non-fatal/non-injury accidents, and indicates the major hotspots where the highest number of accidents has taken place using the cartographic representation which was developed by using ArcGIS software. The Infographic visualization used in the study finally gave a detailed summary of the accidents report of Hayatabad from 2009 to 2020, which can be used to extract useful information such that

- *Annamalai Balasubramanian is with Grand Valley State University. E-mail: balasuba@mail.gvsu.edu*
- *Mageshwaran Anbazhagan is with Grand Valley State University. E-mail: anbazham@mail.gvsu.edu*
- *Sapna Maduraimuthu is with Grand Valley State University. E-mail: madurais@mail.gvsu.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

the factors contributing significantly to the accidents in most frequent locations and the time and gender in most of the Road traffic accidents can be identified and investigated.

In another research study, the traffic accident data in Tashkent city from 2005 to 2012 with a total of 2053 observations has been used for visualization using the R programming language. The collected data is processed and cleaned to make sure there are no missing values and duplicate data. It classified road accidents into 9 categories, where collisions and collisions with pedestrian data are considered for analysis and visualization, dropping out other types due to lower frequency. The purpose of the study is to observe the temporal distribution of accidents over the days, days of the week, months, and year [8]. To witness the data allocation of the two major accident types, individual plots are created for the attributes considered. It provides a hint and assistance to reduce the frequency of road accidents over the days, days of the week, months, or years. Bar plots are utilized to support the data to visualize how each accident type compares within itself.

4 Data Source(s)

The data can be found from the following sources:

- National Highway Traffic Safety Administration. [1, 2, 6]
- Distracted Driving Related Crashes in Michigan 2016-2019. [4, p. 7]
- Michigan Traffic Crash Facts. [5]
- 2019 Year End Traffic Crash Statistics by Michigan State Police. [3, p. 6]

The National Highway Traffic Safety Administration website [1, 2, 6] has a list of data sets containing information on aspects of the fatal crashes, fatalities, speed limit ranges, day and time of when the accidents had taken place in Michigan in 2019 that could be useful for the trend analysis over the time span of 10 years, to compare the speed limit categories and to determine the safest day and time to travel.The number of fatal crashes across different counties in Michigan are determined based on the driver's alcohol consumption. [5].The data from distracted driving related Crashes in Michigan 2016-2019 website [4, p. 7] is used to determine the correlation between the driver's distraction and the crash number across the years from 2016 to 2019 in Michigan.

5 Tasks

The different visualizations obtained as a result from the analysis of Road Traffic Accidents would be able to answer the following task-related questions, thus providing context and clarity to the reported fatalities associated with the fatal crashes:

- What is the trend followed for the fatal crashes and fatalities in Michigan over the time of 10 years from 2009 to 2019?
- When checked for the total number of vehicles involved in fatal crashes under each speed limit in Michigan in 2019, how do all the speed limit categories compare with each other?
- For each age group, what is the proportion of genders and the number of accidents aggregated in Michigan in 2019?
- Does the distraction of a driver correlate with the vehicle point of impact on fatal crashes aggregated across the years from 2016 to 2019 in Michigan?
- Based on the total number of road fatalities in Michigan in 2019, how many accidents occurred across each day of the week and time of the day?
- In 2019, how does alcohol intake affect the number of Fatal crash among Michigan counties in terms of road accidents ?

5.1 Trend Analysis of Fatal crash and Fatalities

The first task sets out to explore the trends in the total number of fatal crashes and fatalities in Michigan over the time span of 10 years from 2009 to 2019. The data will be extracted from the National Highway Traffic Safety Administration website [6], which provides the aggregated number of fatalities and fatal crashes over ten years from 2009 to 2019 in Michigan to identify whether the fatal crashes have increased or decreased which could in turn have a possible impact on the pattern of number of fatalities.

5.2 Comparison of Speed Limit Categories

The aggregated number of vehicles involved in fatal crashes in each speed limit (30 mph or less, 35 or 40 mph, 45 or 50 mph, 55 mph, 60 mph or higher, no statutory limit, unknown) category is considered to determine which speed limit category has resulted in higher fatal crashes as an outcome of this task.The data for this task is obtained from the National Highway Traffic Safety Administration [1] from which the speed limit ranges and the land type usage data are attained to identify how the speed limits compare with each other in Michigan in 2019.

5.3 Gender and Age Biasing on Road Accidents

To seek answer to the question related to this task,we are going to find the total number of fatalities under each specific age group and identify whether male or female has contributed towards the highest number of fatalities by analysing the data [3] provided by the Michigan Department of State Police.

5.4 Influence of Distracted Driving

Distracted driving is a deviation which could significantly increase the risk of an accident. This task makes use of the data from the website [4, p. 7] which indicates the drivers involved in the crashes according to the different levels of driver distraction to identify how they are correlated with each other across the years from 2016 to 2019 in Michigan. Also,it indicates the severity of distribution between distracted and non-distracted drivers.

5.5 Day and Time Safe to Travel

The safest day in a week and the time to travel in a day in Michigan in 2019 is determined by the aggregated value of the least number of Fatalities.The data is pulled out from National Highway Traffic Safety Administration [2] from which the aggregated value of number of fatalities are grouped by the time of day and day of the week in Michigan in 2019.

5.6 Number of Fatal crashes across different Counties based on the driver's Alcohol Consumption

In Michigan, the number of fatal crashes reported in 2019 across 83 different counties of Michigan based on the whether the driver has been involved in drinking or not are taken into consideration for this task. The data is gathered from Michigan Traffic crash facts [5] website, in which there are aggregated value of fatal crashes occur ed across different counties in 2019.

6 Visualization Design

Each of the six different tasks will require a different visualization to accomplish its objective.

6.1 Trend Analysis

The first task sets out to inspect the trends in the number of fatal crashes and fatalities aggregated over the years from 2009 to 2019 in Michigan. To accomplish this, a line chart is used for the visualization with two lines plotted on the chart over the same scale thus making it easier to understand the trend followed. The other chart of interest to indicate the trends over time would be the stacked area chart and heat map. The stacked area chart is also used to predict over the time relationships, since the areas are stacked up by bottom to top manner for each category if the data point of the bottom category is greater than the category

stacked above, then it could easily be missed out due to the shading present in the preceding category. This overlapping of the data points can be minimized by reducing the transparency of the area, however it would be hard to differentiate the data points of each category. The heat maps on the other hand would be better to indicate the patterns and variances between multiple categories, but according to Cleveland's rule, the position along a common scale is ranked higher implying that interpreting lines would be easier than that of the area, shading, or color saturation. In a line chart, each point is connected using lines complying with Gestalt's Law of connection where connected points form a group and facilitate in a comparison. Thus, a line chart seemed to be the best option to achieve this task rather than a stacked area chart or a heat map.

There are two primary visual encodings considered for this visualization. First is the format of the point that is plotted to indicate the position and the value. The visualization uses the points to represent the number of fatalities and the number of fatal crashes and the lines connecting all the points to indicate how the line moves through the timeline from 2009 to 2019.A point is a primary geometric primitive and increases the readability of the line chart in general, however the other shapes may be distracting. The shapes are of lesser importance in a line chart, as the lines are focused to determine the trends or patterns. The second visual encoding is the color selection to differentiate each line on the chart. Since each series represents a different group, it is better to use a categorical color scale that aids in the differentiation of the lines easily using colors. The selected colors are red and orange to distinguish the categories and understand the different trends being followed. Red depicts the color of blood seen in the fatalities, and for the fatal crashes orange is chosen which not only differentiates but also complements the color red implying fatal crashes leads to fatalities.

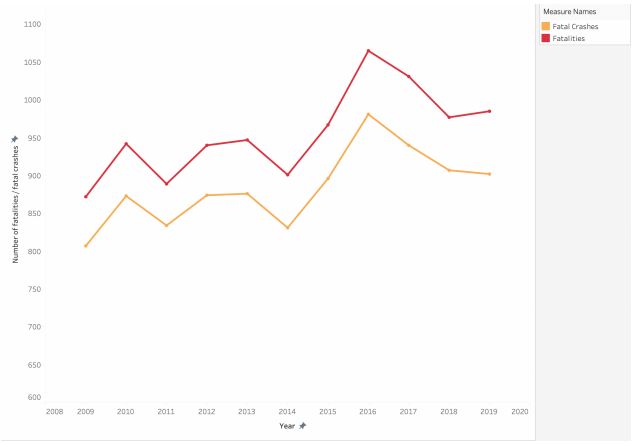

Figure 1: Trend analysis of fatal crashes & fatalities in michigan

6.2 Comparison of Speed Limit Categories

In this task, each speed limit categories are compared against each other to determine the speed in which numerous fatal crashes has occurred. To understand better a visualization involving bar chart is used. The collected data consists of one quantitative and one qualitative data which makes packed bubble chart, stacked bar chart an alternative to help accomplish the task. However, in the packed bubble chart the position of the bubbles is plotted randomly producing little or no information on comparison against each category. The packed bubble chart is good at visualizing the third dimension of the data, since the data lacks the third dimension using a packed bubble chart will not be beneficial. The stacked bar chart aids in comparison of aggregated values across multiple categories grouped together, however the difference between each category can be hard to observe. According to the Cleveland's rule of hierarchy the length of a data can be interpreted easily in comparison to area as in packed bubbled chart. The data is simple and would not require grouping across different categories making the bar chart an ideal choice.

The color would be used a visual encoding for this visualization. Since the data collected is for fatal crashes, it involves fatalities which could be better represented with the color red. Also, to follow the encodings used in the previous task red would be more appropriate to consider. Different shades of red is considered to effectively represent the vehicle count involved in the fatal crashes.

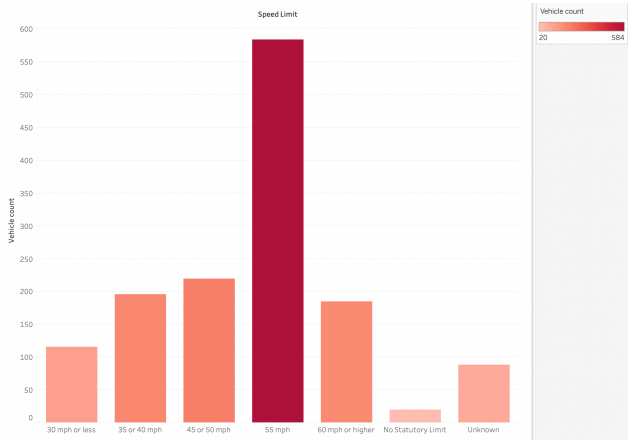

Figure 2: Speed limit categories compared

6.3 Gender and Age Biasing

Figure 3: Gender and age biasing of fatalities

6.4 Influence of Distracted Driving

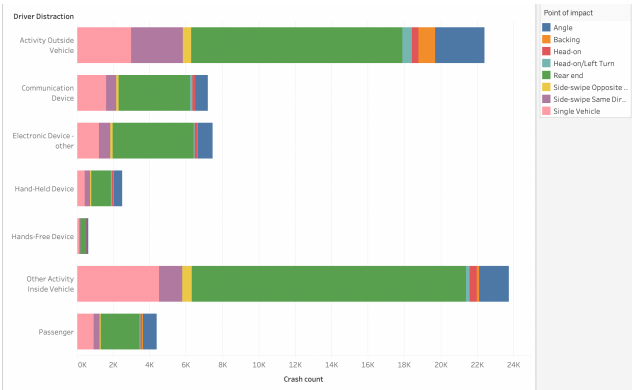

Figure 4: Influence of distracted driving

6.5 Safer day and time to travel

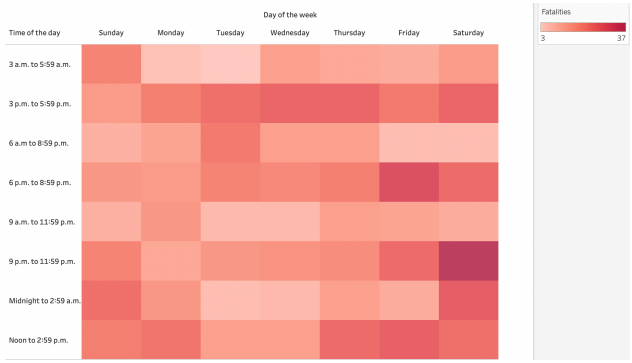

Figure 5: Time and day safe to travel

6.6 Fatal crashes across different counties of Michigan

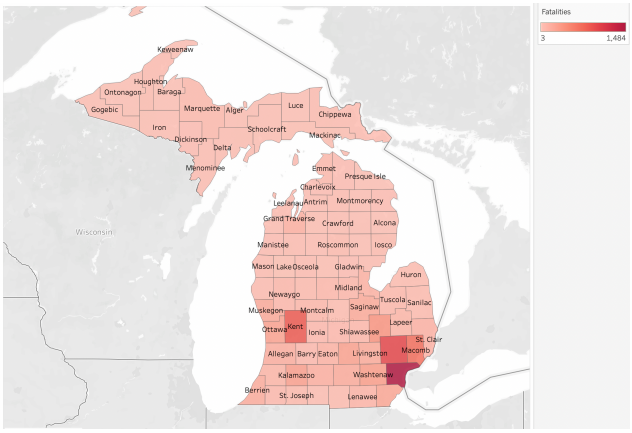

Figure 6: County level fatalities count

REFERENCES

[1] Comparison of speed limit categories:. https://www-fars.nhtsa.dot.gov/Vehicles/VehiclesLocation.aspx, Report Selection Criteria: Vehicles Involved by Speed Limit and Land Use - State: Michigan, Year : 2019.

[2] Day and time safe to travel:. https://www-fars.nhtsa.dot.gov/Crashes/CrashesTime.aspx, Report Selection Criteria: Fatal Crashes by Time of Day and Day of week - State: Michigan, Year : 2019.

[3] Gender and age biasing on road accidents:. https://www.michigan.gov/documents/msp/2019_Year_End_File_for_Web_691956_7.pdf,page:6.

[4] Influence of distracted driving:. https://publications.michigantrafficcrashfacts.org/2019/Distracted_Driving_2019_final_draft.pdf,page:7.

[5] Number of fatal crashes across counties in michigan based on the driver's alcohol consumption:. https://www.michigantrafficcrashfacts.org/querytool/table/0#q1;0;2019;;&p0,7|0,35|0|1000.

[6] Trend analysis of fatal crash and fatalities in michigan:. https://www-fars.nhtsa.dot.gov/Trends/TrendsGeneral.aspx, Report Selection Criteria: Fatal Crashes, 1994 - 2019 State: Michigan, Fatalities and Fatality Rates, 1994 - 2019 State: Michigan.

[7] M. B. A. Rabbani, M. A. Musarat, W. S. Alaloul, A. Maqsoom, H. Bukhari, and W. Rafiq. Road traffic accident data analysis and its visualization. *Civil engineering and architecture*, 9(1):1603–1614, 2021.

[8] J. Sodikov. Road traffic accident data analysis and visualization in r. *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*, (2), 2018.