Analysis of U.S Crash Statistics and the Promise of Emerging Autonomous Driving Technology

CMDA Capstone Final Report

[redacted]

Client: Virginia Tech Transportation Institute
Ride or Die: Qi Xu, Samarth Gupta, Weiting Li

December 11, 2019

1: **Executive Summary**

Tasked with using changepoint detection and visualization packages to proficiently visualize and analyze the trend in fatality statistics in the last two decades, our team spent this whole semester gaining good understanding in learning how to absorb the information and to build a Monte Carlo simulation to predict how automated driving system would impact fatality statistics in the upcoming decade. We begin with exploring the general fatality trend by building time series with parameters like the number of fatalities and normalized fatality parameters e.g., the fatality rate per one hundred million vehicle miles traveled in the last two decades. Then we use changepoint detection to approximate the linear regression to report the significant time points in global statistics. Afterward, we explore trends in rear-end collision, rollover crashes, which are two essential crash types, with changepoint detection again. Ultimately, we want to use the information we gain in the procedures above and Monte-Carlo simulation to predict how many fatalities level 2 automated driving systems (specifically automated braking system and Electronic stability control) could save. Finally, we hope that our project can attract the attention of car developers and invent and improve car technology more specifically.

2: **Problem Statement**

Throughout the semester, our goal is to use two steps to predict whether autonomous driving can effectively reduce the incidence of car accidents in the next 10 years. Through data collection, collation, and modeling throughout the semester, we know that the overall incidence of car accidents in the United States has shown a downward trend in the past 10-20 years. Among all types of car accidents, the trend of individual types does not decrease but increases or shows a constant fluctuation. By analyzing this type of technology, you can improve the technology of related cars, invent and improve them, and it is possible to save more people from the car accident to the greatest extent. A related report states that 73.6% of all accidents are solely the responsibility of the driver and that traffic accidents are listed as the ninth leading cause of death globally [12]. If no action is taken, road traffic injuries are expected to increase in 2030 [13], became the fifth leading cause of death. So first of all, our results are in order to meet the needs of our customers. In addition, we hope that our survey results can attract the attention of all car developers and can be more effective in the development and production of automotive technology so that the fatality of traffic accidents Effectively reduce.

2.1: **Problem Reformulation**

Our original goal of our client is for us to study the impact of smartphones on the incidence of car accidents and apply them to the development of future automotive technology. But during the semester, through our collection and data cleaning, there was not good data sources for us to explicitly explore the influence of smartphone only on fatality statistics. There was not enough information from the data to know exactly at the moment of the accident, what the driver is doing, such as eating, communicating with the co-pilot, and many other action related to distractions.

Therefore, although our initial goal was to conduct our project about how smartphones impact traffic risk, considering the lack of information in the database given and the actual difficulty of collecting data in real life, our client decided to change our topic from that into studying the general trends in the datasets given and conduct descriptive statistics, and use this information to build a successful Monte-Carlo model to predict the impact of automated driving systems on fatality in the future.

3: **Ethical Considerations**

Our project raises quite a few ethical concerns. Perhaps most importantly, we must highlight the underlying assumptions and data limitations as they pertain to our model. All of our models assume that the data being used is accurate, inclusive, and unbiased. However, upon further research into our datasets, it is evident that this is not always the case. For example, the distracted driving data from the Fatality Analysis Reporting System (FARS) actually relies on State Police Accident Reports (PARs) to determine whether or not a distracted driving incident was in-fact 'distracted' or not. There are many problems with this, the foremost being the fact that PARs are inconsistent across different jurisdictions. Different jurisdictions from different states have different guidelines for filling out PARs, which leads to inherent inconsistencies in reporting. Due to these various inconsistencies in PARs, the reported cause and number of distracted-driving related incidents is not always accurate in FARS. These inaccuracies will inherently bleed over into our model and impact the model output as the fundamental data (FARS) behind the model is corrupt. In addition to being inaccurate, I found that the underlying data behind our model could also be inherently biased and un-inclusive. As mentioned above, determining whether or not a driving incident is 'distracted' or not comes largely from PARs. This means that individuals are trusted to truthfully self-report to law enforcement in regards to the nature of their accident. However, research shows that self-reporting of negative behavior is lower than actual occurrence of that negative behavior. In other words, we can infer that the reported presence of distraction during crashes is lower than the actual occurrence. This underreporting skews the FARS data which underlies our model, hence inherently skewing the output of our model and our ability to correctly quantify the number of car crashes that are caused by distractions. As such, it is evident that model assumptions and data limitations play a significant role in our model and must be considered to prevent mischaracterization of model outcomes.

4: **Literature Review**

Although smartphones have made life easier by allowing people to stay connected at all times, they have also posed a serious safety risk in the transport sector. As drivers try to check their text messages, receive calls, respond to emails or perform any other activity on their phones when driving, they get distracted and can easily get involved in an accident. Cell phone distractions have been increasing at alarming rates. For instance, research by Ascone et al has revealed that the risks of using cell phones when a driver is driving are downright startling [3]. Different authors have scrutinized the impacts of using cell phones when driving and came up with different findings.

Research by Horrey & Wickens which scrutinized the impacts of cell phone engagement when driving with an aim of determining the risks associated with the use of cell phones when driving indicated that there were some associated costs. Overall, the research revealed that phone calls posed high risks on driving performance because they distracted drivers to the extent of forgetting that they were driving [4]. The study indicated that the risks were high depending on the reaction time, with minimum risks associated with maintaining their lanes on highways.

A different research study by Patten et al which scrutinized the relationship between the use of mobile phones and road accidents revealed that driver distractions by mobile phones were among the root causes of major road accidents [5]. It also revealed that the content of the conversation in which a driver engaged predicted the likelihood of getting distracted. For instance, when a conversation was difficult and complex, the possibility of such a conversation distracting the driver was also very high. According to Klauer et al, the impacts of driver inattention have been considered among the main causes of increased road crashes [6]. And while this research listed a number of factors that contributed to driver inattention, the use of cell phones was considered as the main distractor because it occupied the mind of the driver.

5: **Project Criteria**

Initially, when our goal was to analyze how the distracted driving-related fatality statistics change in the last two decades, we had planned to use the MICE package to clean our data, and use MLR(multiple linear regression) to analyze it. Then, we wanted to use visualizations in the form of the cartoon to present our results. But since our topic changed from that to how the general driving statistics developed in the last two decades & how the automated driving system would impact on fatalities. And the components reduced into three elements: data collection, data analysis, data visualization. For data cleaning, we do not need to clean our data because data from NHTSA website (which our client prefer) had already been cleaned by their data engineers. Thus, We only need to query for our desired dataset. For data analysis, We ended up spending most of our time applying changepoint analysis [1] and Monte-Carlo simulation [2]. According to these changes, we modified our criteria into:

Criteria	Metric
Accuracy	Average variance on the data we used
Speed	Time(hours) taken to run these programs

Satisfaction of client	Rate by our client from one to ten

Table 1. Criteria for Data Analysis

Criteria	Metric
Minimal	Limit of our final number of graphs is 15.
Informative	Satisfaction from our client
Inclusive	Our visualizations cover 90% of our data

Table 2. Criteria for Data Visualization

6: **Selected Solutions**

As mentioned above, our solution approaches primarily ended up consisting of conducting Change-Point analysis and running Monte-Carlo simulations. Change-Point methods allow us to detect any changes that have occurred in time-series data. These methods determine the number of changes and estimates the time of each change. Furthermore, Change-Point methods provide confidence intervals for each of the time changes. In terms of actually implementing these Change-Point methods, we used various packages in R. These packages included the Chngpt, Cpt, and ChangePoint packages. The Change-Point analysis was particularly useful to us because it allowed us to effectively identify the critical points (years) to focus on in our time-series data. In addition to using Change-Point analysis, we also came up with Monte-Carlo simulations. Monte-Carlo simulations are a broad category of computational algorithms that rely on repeated random sampling to obtain numerical results. More specifically, Monte-Carlo simulations are used to model the probability of different outcomes for a process that cannot easily be predicted due to the presence of random variables. We used Monte-Carlo simulations in conjunction with the Population Attributable Risk Formula to predict the total percentage of crashes that could be eliminated with the increased prevalence and effectiveness of certain automated driving technologies. Specifically, we tried to put a total number on the percentage of crashes that could be reduced with the increased prevalence and effectiveness of automatic braking technology as well as lane-assist technology. To accomplish this, we used the rBeta and rGamma functions (Beta and Gamma distributions) in R to randomly sample values for Beta (effectiveness) and Gamma (prevalence) to run multiple simulations using the Population Attributable Risk Formula. Running these simulations using various randomly sampled values for Beta and Gamma allowed us to identify certain Beta and Gamma values for which we could reduce the greatest percentage of crashes with, which we displayed in the form of a Histogram.

7: **Results**

7.1: **Introduction**

From the appearance of the first portable phone in 1989 to 2019, the technology of smartphones has improved and appeared at an alarming rate. In 2018, the penetration rate of smartphones in the United States was as high as 77%, and it will increase year by year at a rate of 20 million people [11]. The need for communication convenience has also led to a reduction in the safety factor during driving. According to the latest statistics, 62% of distracted driving collisions are caused by drivers driving their minds off the road. In the U.S., a quarter of car accidents are caused by using a mobile phone behind the steering wheel, while the driver is listening Or music, and it can reduce your concentration by 40% while driving [14]. Therefore, many people also believe that the popularity of smartphones has severely affected traffic safety.

According to the National Highway Traffic Safety Administration (NHTSA), traffic accidents are the leading cause of death among young people aged 15-29 worldwide and the 9th leading cause of death among all people. When everyone attributed the leading cause of traffic accidents to the popularity of smartphones, autonomous driving technology attracted widespread attention. Therefore, our team uses the database of the past 10 years to analyze the types of car accidents one by one, hoping to find out the technology areas that need to be improved in the past 10 years. And model it in the future. It is hoped that through the development of automotive technology, more lives can be saved from traffic accidents.

In this report, we will present the research ideas following the mind map below (Figure 1: Mind map for the whole project).

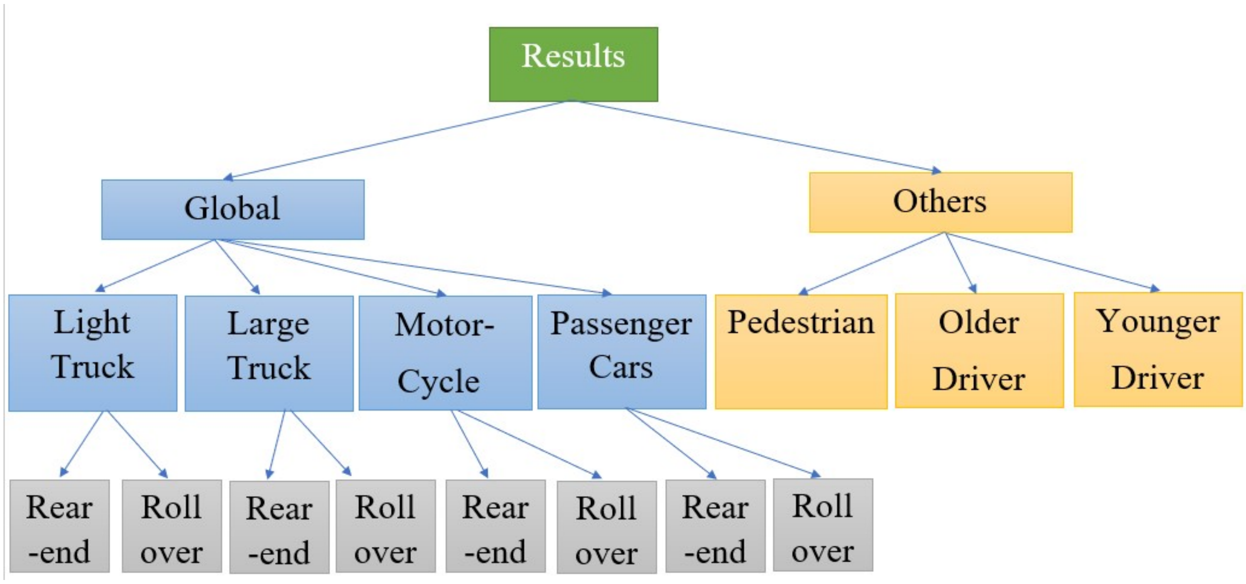


Figure 1: Mind map for the results

7.2: **Data collection**

After we decided that we would use data from the National Highway Traffic Safety Administration (NHTSA), we started to explore options to get comprehensive data that we need. Although NHTSA provides its raw data on its website, NHTSA also has a link to itself-designed query page that allows data researchers to get their desired datasets quickly. With the consent from our client, we began to collect our data in this way. There are a few datasets that we think would be beneficial to our exploration of the general trend:
(1) a dataset that shows general fatality statistics;
(2) a dataset that has information about fatalities related to car types;
(3) a dataset with information about deaths related to specific crash types (rear-end collision, rollover) ;
(4) a dataset with auxiliary data such as involving pedestrians, death related to age groups.

7.3: **Global analysis**

First of all, our team used the database of FARS to visualize the fatality statistics from 2007 to 2017 along with the timeline of popular car technologies(Figure 2: Change in fatalities along with the development of car technology)

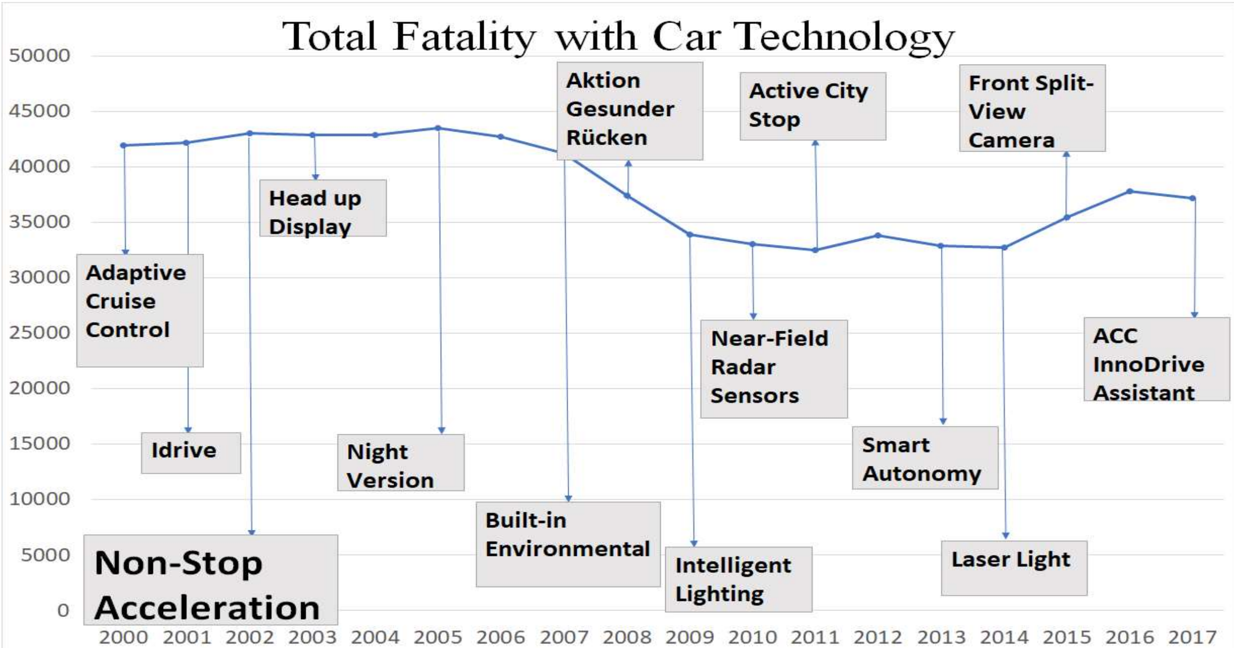


Figure 2: Change in fatalities along with the development of car technology

The technology in Figure 2 only represents the time when the technology appeared and did not represent the time when it has been commonly used. We created this graph because we believe that there is a very likely correlation between innovation in car technology and safety of the car. As time goes by, the widespread application and upgrading of new technologies will take 10 - 20 years to finally have a great impact on our cars' safety. Additionally, it shows us from Figure 2 that though there is no technology that are as important as the inventions of seat belts and airbags, these other small upgrades also make cars more easier and more intelligent for human beings to drive in difficult occasions, such as omnidirectional cameras, which can effectively eliminate the impact of blind spots on driving safety.

Then we try to reassemble and analyze the collected data from a more statistical perspective. We began with using the general dataset, a time series dataset with a series of fatalities related parameters(Fatalities, Resident population, the Fatality Rate Per 100,000 Population, etc). We made a table to show more details on this dataset (Table 3: An overview of the general dataset we enacted):

Year	Fatalities	Resident population (Thousands)	Fatality Rate Per 100,000 Population	Licensed Drivers (Thousands)	Fatality Rate Per 100,000 Licensed Drivers	Registered Motor Vehicles(Thousands)	Fatality Rate Per 100,000 Registered Vehicles	Vehicle Miles Traveled (Billions)	Fatality Rate Per 100,000 VMT

Table 3: An overview of the general dataset we enacted

During this process, we seek to find the global trend firstly by firstly plot some fundamental visualizations in R with packages "ggplot2" [7]. It is worth mentioning, in the beginning, this

dataset only consists of parameters: Year, Fatality (Number of Deaths). After we plot the graph with “ggplot2” package in R. We found that this visualization shows that the global trend of the total number of fatalities has decreased dramatically in 2017 compared to 1994. But we think this result was not statistically significant enough because it is not normalized. Thus, we extract other parameters, which are Resident Population, Licensed Drivers, Registered Motor Vehicles, Vehicle Miles Traveled from another dataset we gained from another query on NHTSA. To get rid of influence from factors e.g, Resident Population, we created normalized parameters, which are the Fatality Rate Per 100,000 Population, the Fatality Rate Per 100,000 Licensed Drivers, the Fatality Rate Per 100,000 Registered Vehicles, the Fatality Rate Per 100,000 VMT.
Here is an example of how we normalized fatality with Resident Population from our original data.

$$\text{Fatality Rate Per 100,000 Population} = \frac{Fatalities}{ResidentPopulation} * 100,000 (1)$$

Formula 1: Fatality Rate Per 100,000 Population

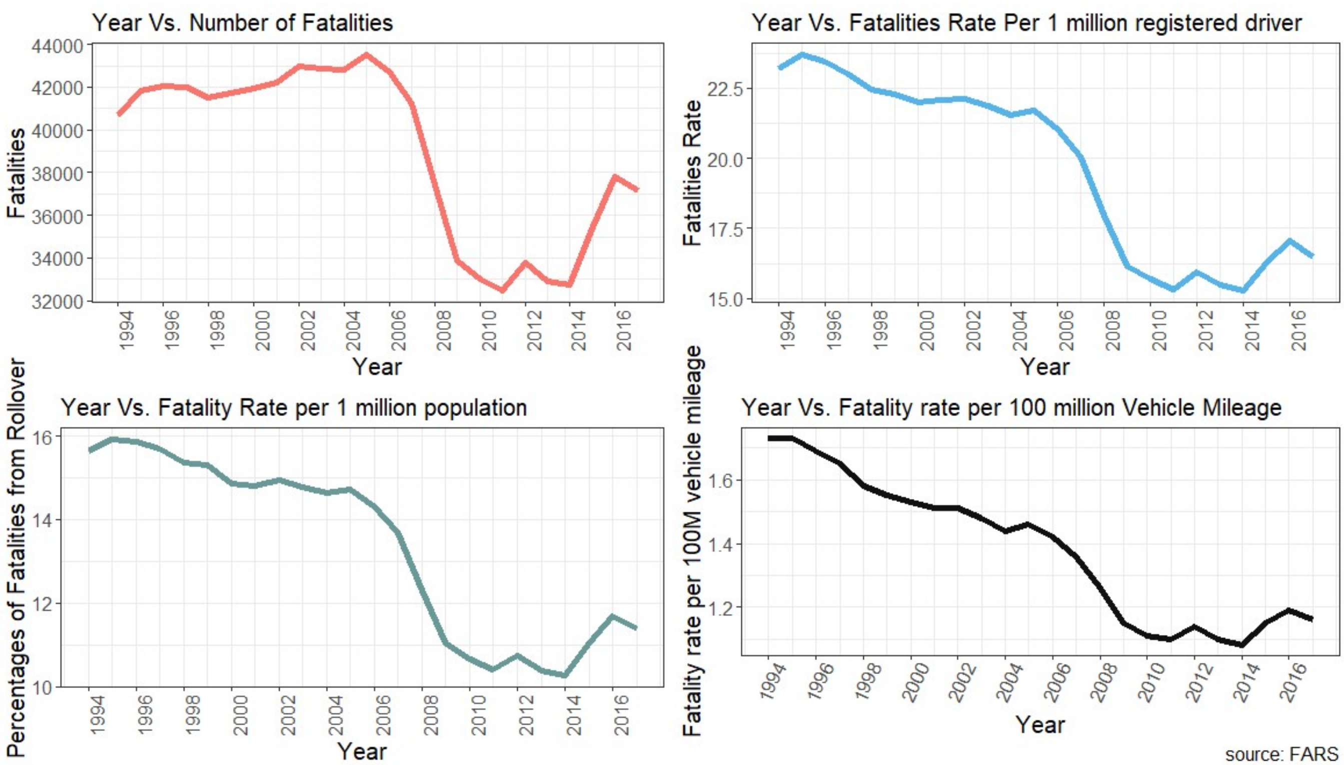


Figure 3: Four Plots of the Global Trends

There are three intriguing observations we made:
(1)As shown in Figure 3, for fatalities Vs. Year (top left plot), the number of fatalities increased since 1994 to 2005. Whereas, for the other three, they had decreased since 1994;
(2)And after that, all plots experienced a similar significant drop in values. And the slope is much sharper compared to earlier;

(3)Around 2010 to 2014, the slopes all experienced obvious sign change (from negative to positive). And Fatalities started to increase slowly in a fluctuating way.
Then, we applied the “chngpt” package [1] in R to further support our discovery from Figure 3. And Figure 4 is the results we gain from that. (Frdriver = Fatality Rate Per one million Registered Drivers; FRPopulation = Fatality Rate Per one million Population; FR100MVMT = Fatality Rate Per 100 Million Vehicle Miles Traveled)

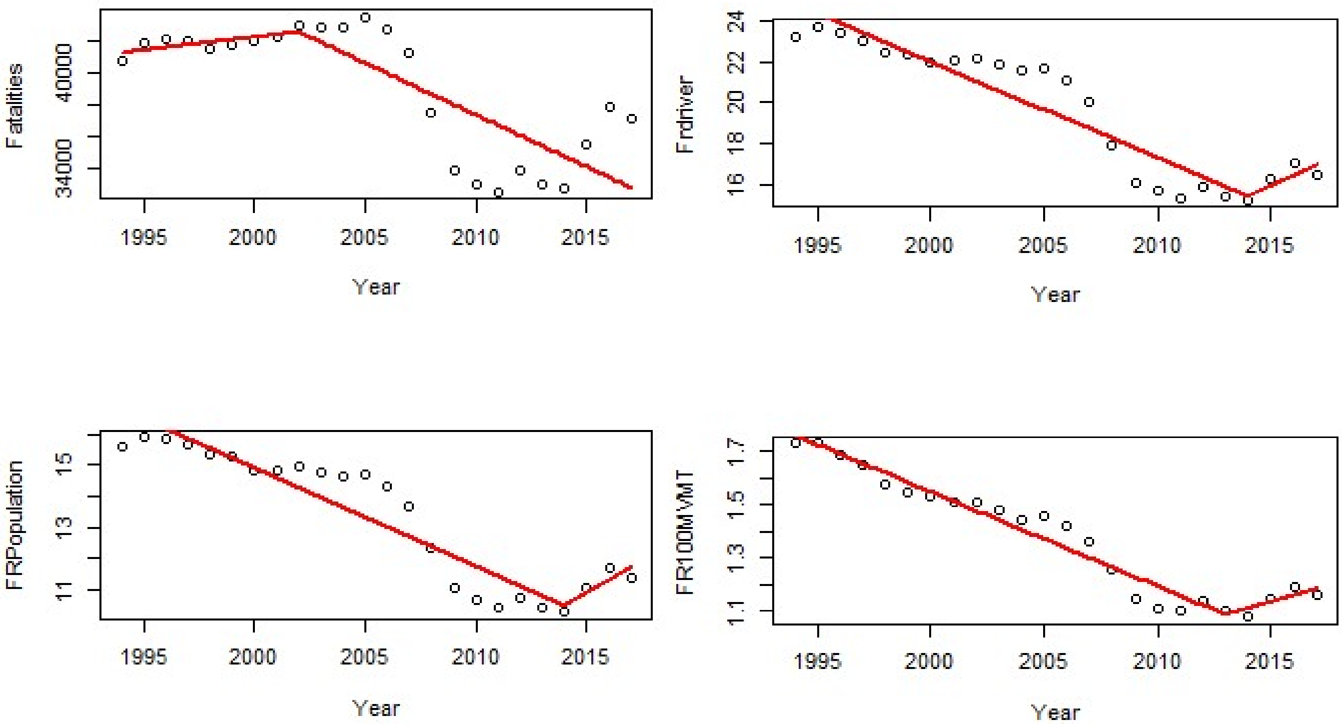


Figure 4: Plots of results from Changepoint analysis

The results we gain from Figure 4 confirmed with our observations earlier. All three plots except for the top left plot (Fatalities vs Year), all have their changepoints in the range between 2013 to 2014. However, for fatalities vs. Year, changepoint analysis tells us that Fatalities, in general, had increased until 2002, and have been decreasing afterward. Out of these four, we prefer the two plots on the right side (Frdriver vs. Year & FR100MVMT vs. Year) because we think the number of registered drivers, number of vehicle miles traveled are the two most important factors that could skew our plots significantly.

From Figure 4, we can't draw any statistical significance, nor can we be sure. All the huge gaps are caused by the emergence of automotive technology. Because of the popularity of automotive technology, it takes about 10-20 years to cover all vehicles, and it starts to show a huge impact from the data. Therefore, we have a second part (7.4: Analysis of Change in Fatality with respect to Car Types), which studies the changes in car types and traffic accidents one by one and draws meaningful results from them.

7.4: Analysis of Change in Fatality with respect to Car Types

7.4.1: Change in fatality composition, 2004 - 2018

As shown above, as the biggest chunk in the total composition, fatalities involved passenger cars owners decreased by 6.38 percent compared to 2004, the beginning year that NHTSA started to collect information about car types. Meanwhile, fatalities involved Light Truck owners also decreased by 2.52 percent. On the contrary, the percentage of fatality involved Large Truck owners, Motorcycles owners, and pedestrian respectively increased by 1.7 percent, 4.48 percent, 0.06 percent accordingly. From the results we observed above, we believe that the reason why although the statistics of global statistics has been decreasing in the last two decades and also the fact that in the meantime many car safety procedures has been added to the cars, only Passenger Car & Light Truck have obvious decrease in percentage of fatality composition. Safety risks for Large truck drivers and motorcyclists are still high.

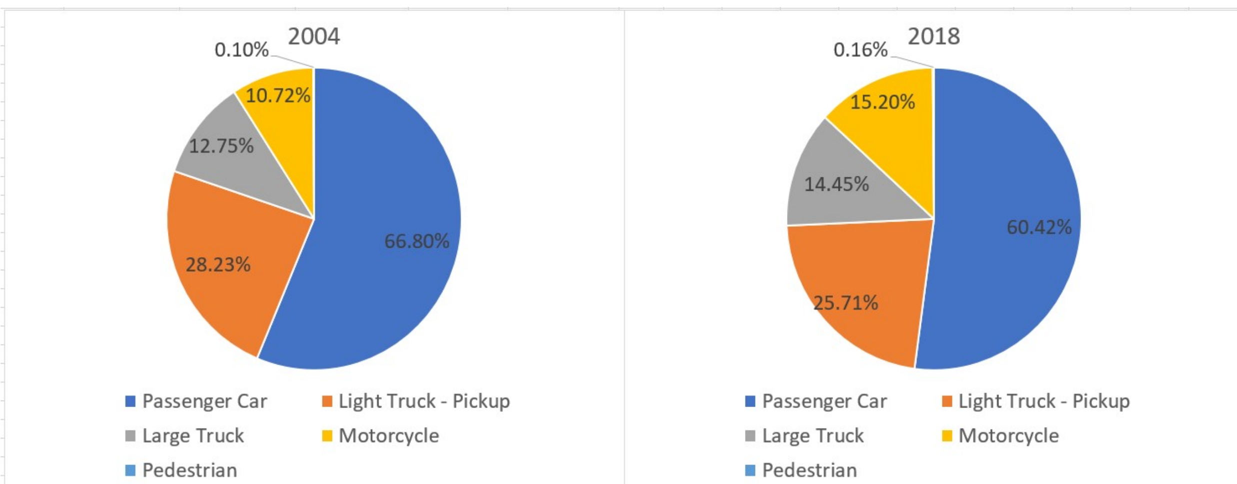


Figure 5: Fatality Composition in total, 2004 & 2018

We all know that compared to ordinary cars, large trucks weigh a few tons more than light trucks. This means that not only is it harder for him to make turning moves, but it is also harder to stop. At the same time, the amount of air driven by it will also make the cars around him more dangerous [8]. Because of its greater operational difficulty and risk factor, on November 1, 2015, the United Nations Economic Commission for Europe (UNECE) announced that it would become a mandatory requirement for new heavy vehicles. This will hopefully make trucks and the cars around them safer. At the same time, in March 2016, the National Highway Traffic Safety Administration (NHTSA) and the Highway Safety Insurance Association announced that 20% of U.S. automakers have agreed to use automatic emergency braking systems by 2022 [9]. Standard on all new cars. This decision will continue to impact statistical data changes over the next 10-20 years.

Similarly, motorcycles are more dangerous, because the driver lacks the protection of the shell, and it is more difficult to maintain a balance between the two wheels. When an accident occurs in the front, the driver is more likely to fall than the car. With the rapid development of automotive technology, fewer technologies can be found to improve motorcycle safety. This is why, over the past 10 years, motorcycle accident rates have increased slightly.

7.4.2: **Change in Fatalities with respect to Car Types, 1994 - 2017**

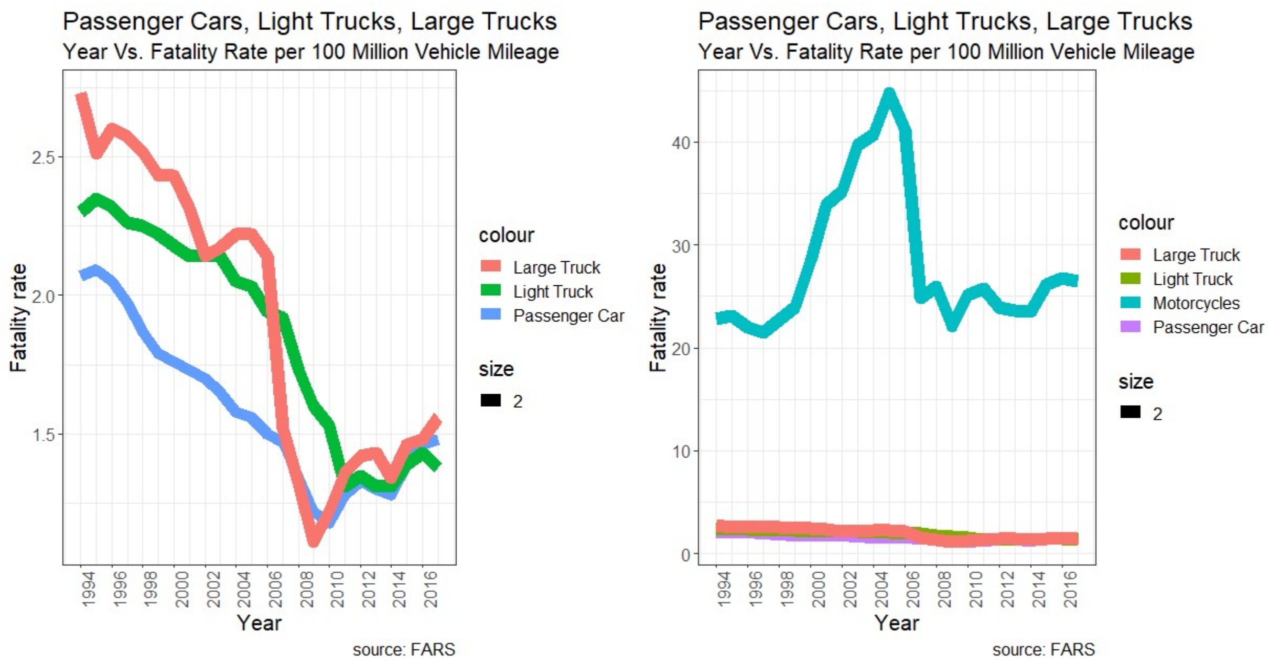


Figure 6: Fatality Composition in Total, 2004 & 2018

The reason why we have two plots instead of one is that the fatality rate of motorcycle is too high so that if we show them in the same plot, we can only observe the slope of the motorcycles. As shown above in Figure 6, the fatality rate 100 million vehicle miles traveled for three different car type: large truck, light truck, and passenger car, show similar pattern of linear line slope as the Fatality vs Year as we show in Figure 3, the left-top plot for global trend. But we also notice that passenger car has comparably the lowest fatality rate. And then the second lowest is the light truck. The one with the highest fatality rate of these three is the large truck.

Intriguingly, only the large truck has a significant drop from 2004 to 2006, while the slope for the passenger car, light truck looks more gradually decreasing. From the plot on the right side, we found that the fatality rate of motorcycles per 100 million VMT is almost 10 to 20 times the rate for the other three car types. And there is also a dramatic drop around 2004 to 2005 for motorcycles too. Until the end, we still have not figured out why there are noticeably similar drops for fatality rate per 1M VMT for large truck and motorcycles. But after consulting which coach and our client, we believe this maybe because of the standard of NHTSA that defines fatality changed around 2004-2006.

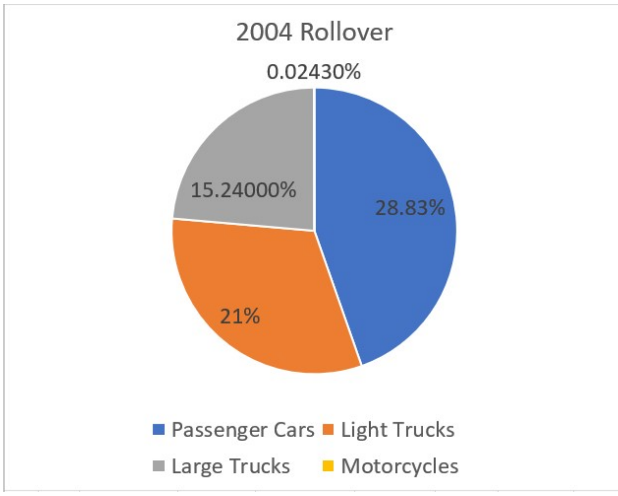
7.4.3: **Change in Rear-end Fatalities of 4 Car Types , 2004 - 2017**

Car Type	2004	2017	change	%change
Passenger Cars	176	146	-30	-17.0%
Light Trucks	165	151	-14	-8.5%
Large Trucks	50	38	-12	-24.0%
Motorcycles	0	14	14	NA

Table 4: Change in Fatalities of Rear-end Crashes from 2004 - 2017

For rear-end crashes (Table 4), large trucks has the highest decrease rate, dropping 24 percent in 2017 compared to it in 2004. This is consistent with our observation is the section 7.4.2, where we discovered that the fatality rate for large trucks decreased the most. Whereas, Light Truck dropped 8.5 percent, and passenger cars dropped 17 percent. Unfortunately, even though we wanted to study the rear-end crash rate involving motorcycles, there were too many years with missing values.

7.4.4: **Change in Rollover Fatalities of 4 Car Types , 2004 - 2017**



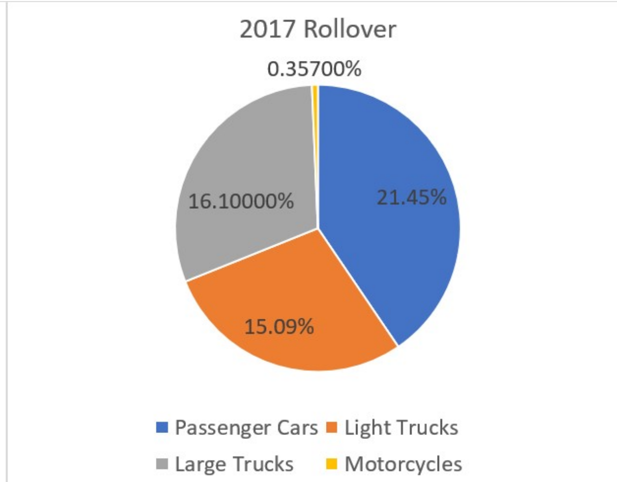


Figure 7: Pie Charts that shows the percentage change in Rollover between from 2004 - 2017

In Figure 7, we calculated the percentage for rollover crashes for all four car types in total rollover crashes. The one that has the most dramatic change is motorcycles. The Fatality percentage for motorcycles in 2017 became 14.69 times the percentage of it in 2004 rollover. However, similar to the trend we found in rear-end crashes, fatality percentage for passenger cars and the Light trucks decreased 7.38 percent, 5.91 percent respectively.

7.5: Comparisons of 10-year (2008 to 2017) Percentage Change of Drivers from different age groups involved in Fatal Crashes

Age Groups	Ten-Year Percentage Difference in Drivers Involved in Fatal Crashes	Ten-Year Percentage Change in Population Estimate	Ten-Year Percentage Change of Number of Licensed Drivers
16-24	-16.4%	+0.3%	-1.5%
25-44	+2.7%	+4.7%	+0.5%
45-64	+9.5%	+7.4%	+7.6%
65+	+29.2%	+31.3%	+34.3%
Total	+3.7%	+7.1%	+7.8%

Table 5: Comparison of Percentage Change of Driver Involved in Fatal Crashes, Percentage Change in Population Estimate, Percentage Change of Licensed Driver in 2008 - 2017

From the table 5 above, we know that there had been a 3.7 percent increase in the ten-year percentage difference in total drivers. Meanwhile, there are also 7.1 percent increase in ten-year percentage change in population, and 7.8 percent increase in difference in number of licensed drivers. Therefore, the speed of the increasing of total number of fatalities is slower than the increasing speed for population or licensed drivers. Interestingly, the number of fatalities involved younger drivers (16-24) compared to 2008 had decreased a significant amount: 16.4 percent, but their population had only increased 0.3 percent. Moreover, the number of licensed young drivers in total had even decreased by 1.5 percent. Whereas, older group of drivers (65 +) had increased about 29.2 percent compared to 2008. But it is worth mentioning that the population of older people had increased about 31.3 percent, and the number of licensed older drivers had increased about 34.3 percent. This is interesting because this discovery is actually contrary to the common-sense that younger drivers compared to older drivers usually are more reckless and less proficient in driving. Therefore, we created Figure 8 to show how the fatality rate of drivers of these two age groups changed from 2004 to 2018. As shown in Figure 8, surprisingly we found that the fatality rate of drivers from both age groups were almost the same back in 2004. But in 2019, the fatality rate of older drivers becomes almost double that for young drivers.

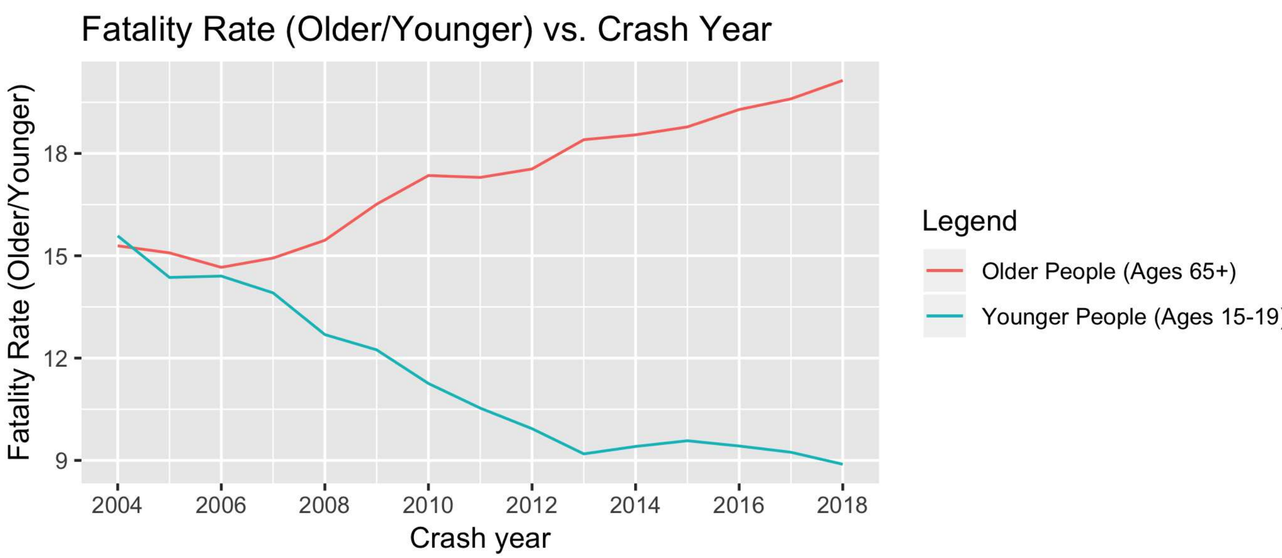
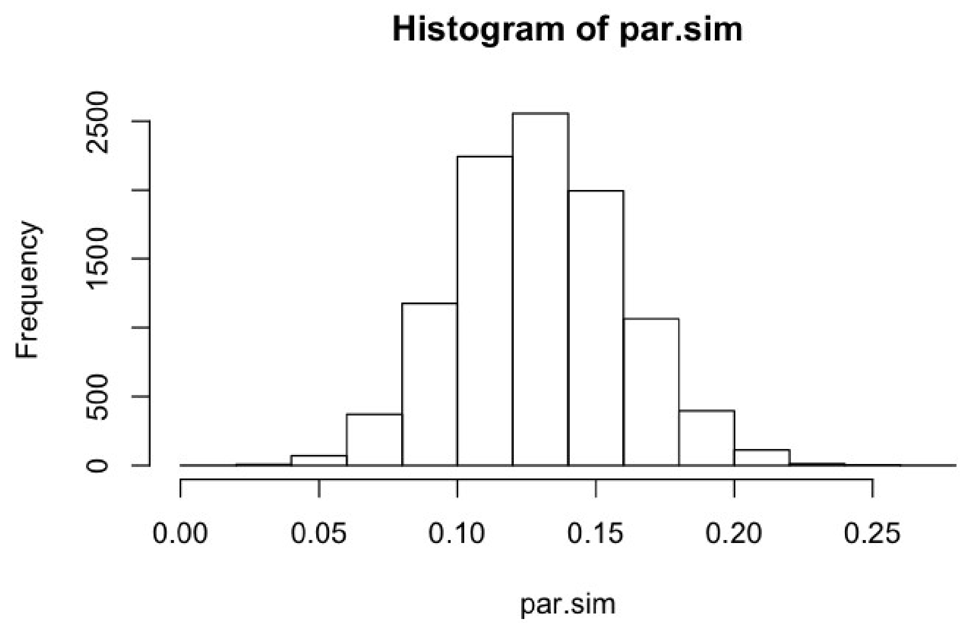


Figure 8: Comparison of fatality rate in total crashes between older/younger group of drivers

7.6: **Monte Carlo Simulation**



0.129552305915188

2.5%	0.0724495384217744
97.5%	0.190304408324019

Figure 9:The Histogram about Monte-Carlo Simulation

Above is a histogram depicting a single Monte-Carlo simulation ran with randomly chosen Beta and Gamma values (Figure 9). Given the Beta (Prevalence) and Gamma (Effectiveness) values selected here, the simulation tells us that we could reduce the total percentage of crashes by 12.96% with Lane-Assist/Automatic Braking technology. Furthermore, this simulation tells us that the confidence interval of the model actually achieving the 12.96% is 95%.

8: **Obstacles**

During the process, we learned how important it is to have frequent back and forth communications with our client. For our team, we spent almost a month on clearing out confusions on our projects in the beginning of this semester. From an overall view, we found that if we had more frequent and effective communications with our client sooner, we would have made a much bigger progress and we could have less confusion when engaging in the project. Also it is important to have a friendly and active information exchange between the team members. In the last few weeks towards the final, we had been having severe team dynamics issues. They were usually about how to split the work fairly, how to get everyone in the team to work collaboratively, and how to finish our work earlier. However, in the end, we have moved past these issues and have been working efficiently and productively as a unified team.

9: **Roles**

According to our original assignment, based on every team member's skill set, Qi would do data collection and data cleaning. Samarth would be in charge of the data analysis with the help of Qi and Weiting. And Weiting would focus on making informative visualizations after the analysis phase. Yet, after the change to our topic asked by our client, our whole team worked together on data collection, data analysis, and data visualization:
1. Samarth focused on Monte-Carlo simulation;
2. Weiting did some descriptive statistics with the general data to show the global trend with Qi.
3. Qi was in charge of making visualizations and gave us feedback on our analytics.

10: **Conclusions and Future Work**

As summarized in our results above, crash statistics have changed significantly in the last decade. Perhaps most importantly, the number of total fatalities in the United States have changed course in trend. From 2004-2011 they were steadily decreasing, but sometime between 2011-2014 they began to increase and have been increasing year-over-year since. While there were many factors that we believe contributed to this change, we believe the primary reason was the relative turnaround in the U.S economy during 2011-2014. Due to an improving economy after the recession in 2008, more people could afford to buy and drive cars which contributed to an increase in the total number of cars on the road. However, it is important to note that the percentage of rear-end collisions has also gone up year-after-year since almost exactly the same time frame (2011-2014), so we can safely say that this may have also contributed to the increase in total fatalities over the last few years.
From our research, it is also evident that the total percentage of pedestrian-related fatalities has increased year-over-year since 2006. Furthermore, it is evident that the percentage of old people (65+) fatalities has steadily *increased* while the percentage of young people (15-19) fatalities has steadily *decreased* in the same time frame. These results are interesting as they seem to contradict the benefits of emerging driver-assist and autonomous driving systems. As demonstrated through our Monte-Carlo simulations, both the prevalence and effectiveness of these systems have been increasing over the course of the last few years. So why are the percentages of certain crash types still increasing? This is an area for future research and analysis as we do not have a conclusive answer to that question yet. Others may be able to garner more insight regarding this as more data from NHTSA becomes readily available about the impact of advanced driving systems. If I were to tackle this problem again, I would attempt

to use more robust time-series analysis tools such as the Autoregressive Integrated Moving Average (ARIMA) and State-space algorithms.

References

[1] Fong, Y., Huang, Y., Gilbert, P. B., & Permar, S. R. (2017). chngpt: threshold regression model estimation and inference. BMC bioinformatics, 18(1), 454.

[2] Binder, K., Heermann, D., Roelofs, L., Mallinckrodt, A. J., & McKay, S. (1993). Monte Carlo simulation in statistical physics. Computers in Physics, 7(2), 156-157.

[3] Ascone, D., Tonja Lindsey, T., & Varghese, C. (2009). *An examination of driver distraction as recorded in NHTSA databases* (No. DOT HS 811 216).The United States. National Highway Traffic Safety Administration.

[4] Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, *48*(1), 196-205.

[5] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

[6] Patten, C. J., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. Accident analysis & prevention, 36(3), 341-350.

[7] Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.

[8] Why Large Trucks Are More Dangerous Than Small Vehicles - Cochran. (2019, October 18). Retrieved from https://cochranfirmhuntsville.com/why-large-trucks-are-more-dangerous/.

[9] Collision avoidance system. (2019, November 19). Retrieved from http://en.wikipedia.org/wiki/Collision_avoidance_system.

[10] VDA. (n.d). Retrieved from http://www.vda.de/en/topics/innovation-and-technology/timeline/timeline-innocations.html#

[11] Holst, A. (2019, November 11). Number of smartphone users worldwide 2014-2020. Retrieved from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.

[12] 裴玉龙, & 马骥. (2003). 道路交通事故道路条件成因分析及预防对策研究. *中国公路学报*, *16*(4), 77-82.

[13] Road Safety Facts. (n.d.). Retrieved from https://www.asirt.org/safe-travel/road-safety-facts/.

[14] Karow, R., Heath, A., Vargas, M., Zack, Natalie, Samantha, … Safety Team. (2019, October 7). 100 Car Accident Statistics for 2019. Retrieved from https://safer-america.com/car-accident-statistics/.